Ewelina Wojtkowiak

373767

L1 phonetic drift in the speech of Polish learners of English: phonological implications

Dryft fonetyczny w języku ojczystym u Polaków uczących się języka angielskiego: implikacje fonologiczne.

Praca doktorska napisana
na Wydziale Anglistyki
Uniwersytetu im. Adama Mickiewicza w Poznaniu
pod kierunkiem prof. dr hab. Katarzyny Dziubalskiej-Kołaczyk i
prof. dr hab. Geoffreya Schwartza.

Poznań, 2022

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my two advisors who made this work possible:

To Prof. Katarzyna Dziubalska-Kołaczyk, with whom I have had the honour of working for the last eight years. Thank you for your endless support and sharing your knowledge, experience, and views on linguistics with me. You have helped me broaden my perspective on a great deal of issues and I appreciate each and every one of the countless discussions we have had.

To Prof. Geoffrey Schwartz for teaching me acoustics, and showing me what in speech is linear and what is not, but perhaps most importantly, sharing a vision of phonological representation that offers real hope for describing the phonetics-phonology relationship. Thank you for providing me with an opportunity to work in your projects, one of which helped me formulate the research questions explored in this dissertation.

My words of thanks go to Prof. Charles Chang from Boston University who offered his expertise and evaluated an early version of this project. His comments, advice, and experience helped me tremendously. Thank you. I would also like to extend my gratitude to Prof. Eugeniusz Cyran from Katolicki Uniwersytet Lubelski for his valuable comments whenever he has had the chance to listen to me.

I am also grateful to those closest to me: my family for their patience and my dearest friends for their support and encouragement: Małgorzata Dębińska, Joanna Maciąg, and Zuzanna Cal – thank you! To Bartosz Brzoza for the company and the support. To my PhD support group which kept me sane: Agata Janicka, Halina Lewandowska, Iga Krzysik, and Ewa Olszewska. To Dr Tomasz Paciorkowski for the peer pressure.

Finally, I would like to thank my participants – without your willingness to spend hours in the recording studio, this thesis literally wouldn't have been written.

Poznań, dnia 14.04.2022

OŚWIADCZENIE

Ja, niżej podpisana

Ewelina Wojtkowiak

przedkładam rozprawę doktorską

pt. *Dryft fonetyczny w języku ojczystym u Polaków uczących się języka angielskiego: implikacje fonologiczne.*

na Uniwersytecie im. Adama Mickiewicza w Poznaniu

i oświadczam,

że napisałam ją samodzielnie.

Oznacza to, że przy pisaniu pracy, poza niezbędnymi konsultacjami, nie korzystałam z pomocy innych osób, a w szczególności nie zlecałam opracowania rozprawy lub jej istotnych części innym osobom, ani nie odpisywałam tej rozprawy lub jej istotnych części od innych osób.

Jednocześnie przyjmuję do wiadomości, że gdyby powyższe oświadczenie okazało się nieprawdziwe, decyzja o wydaniu mi dyplomu zostanie cofnięta.

Poznań, 14.04.2022

(miejscowość, data)

(czytelny podpis)

Table of contents

List of tables

List of figures

Abbreviations

- **AMU** – Adam Mickiewicz University, where the study was carried out;
- **CEFR** – Common European Framework of Reference;
- **Group 1** – 1BA students, taking part in the longitudinal study;
- **1BA** – the first year of the Bachelor of Arts programme;
- **Group 2** – 2BA students, recorded after two years of intensive phonetic training;
- **2BA** – the second year of the Bachelor of Arts programme;
- **Group 3** – 3BA students, recorded one year after finishing their phonetic training, but still using English on daily basis during their classes;
- **3BA** – the third (and last) year of the Bachelor of Arts programme;
- **Group 4** – quasi-monoliguals, who served as a comparison group;
- **OP** – Onset Prominence (Schwartz 2016 et seq)
- **P-I** – phrase-initial position, following rising intonantion and a comma in the spelling;
- **P-M** – phrase-medial position, not following any prosodic break;
- **SPE** – Sound Pattern of English (Chomsky and Halle 1968)
- **T1** – first recording session, taking place in October, within the first two weeks of the start of the academic year;
- **T2** – second recording session, taking place in February, after one semester of training;
- **T3** – third recording session, taking place in June, towards the end of the academic year, before the start of the summer exam session;
- **Type 1** – fully pre-voiced /bdg/ item;
- **Type 2** – an instance of /bdg/ produced with a break in pre-voicing;
- **Type 3** – an unvoiced realisation of /bdg/, “canonical”, English-like production;
- **U-I** – utterance-initial position, following a full stop;

Introduction

The present dissertation is devoted to the topic of L2-induced phonetic drift in L1. In general, the L1 (i.e. one's native language) interference in L2 (i.e. one's second language) is a relatively well-studied subject. Even naïve speakers intuitively realise that non-native talkers might display traces of a "foreign accent" in their L2, which may indicate where they are from. Far less attention has been paid to whether L2 learning can influence L1 pronunciation, although hints that this is possible date back to 1987, when James E. Flege explored this idea in his works on L2 speech. The research into non-pathological, short term changes that L1 might undergo under the influence of recent exposure to L2 started with Charles B. Chang's series of experiments on American speakers of Korean in 2010s. Chang coined the term "phonetic drift" (at least, in its modern understanding) to refer to those changes. Ever since then, more and more language pairs have been studied. With respect to Polish and English, the two languages which are the main subject of the present work, while there exist some studies pertaining to them, there are also problems with them: they either focus on too few categories (e.g. Waniek-Klimczak 2011), are merely cross-sectional (e.g. Schwartz 2020), or the results thereof might be influenced by other languages (e.g. Wrembel 2015). The study presented in this dissertation attempts to fill these gaps. It encompasses a number of categories (i.e. voiced and voiceless initial plosives as well as vowels in citation forms and sentence lists), it is longitudinal (i.e. with pre-tests and post-tests as well as the inclusion of monolingual controls, thanks to which we are able to trace the patterns of phonetic drift in time), and focuses on bilingual speakers who are not fluent in any other language. With 70 speakers in total, the phonetic studies analyse over 10 thousand tokens. Furthermore, aside from presenting the phonetic data, the thesis evaluates phonological theories and their predictions as far as the magnitude of phonetic drift in consonants and vowels is concerned. The dissertation comprises eight chapters, out of which five are theoretical and three empirical.

Chapter 1 serves as the basic introduction to research into L2 speech, focusing on selected models which have been devised to date and whose common main aim is to predict the ease and/or difficulty with which speakers are able to acquire a new language from the perspective of pronunciation. Aside from overviewing theories of L2 acquisition, it also introduces the model whose predictions serve as the basis in the formulation of hypotheses underlining the present dissertation, namely the Speech Learning Model (SLM; Flege 1995). One of the main principles found in the SLM is the so-called "equivalence classification". Its definition indicates two possible scenarios that can occur when a speaker is learning their second language. They either find the new L2 sound different enough from their L1 phonetic inventory that a completely new category is formed; this entails success in both perception and production of that sound. Conversely, if the new L2 sound is deemed to be equivalent to a sound belonging to an already existing L1 phonetic category, the two are predicted to be merged together and the acoustic details differentiating them to be tuned out. As a result, the L2 pronunciation of that sound is supposed to carry traces of foreign accent. The SLM assumes the effects of equivalence classification to be bi-directional, which, as a corollary, suggests that L2 sounds exert influence on the L1 categories with which they were merged.

In this way, the SLM recognises the existence of phonetic drift in L1, triggered by L2 acquisition. Chapter 2 begins with a discussion that consolidates terminological difficulties in the descriptions of phonetic drift – that is, it describes how different it is from competing terms used in the literature to describe L2 influence on L1, such as *backward transfer*, *cross-linguistic interaction*, or *phonetic/phonological attrition*. After that, it presents a selection of phonetic experiments which have been conducted on different language pairs, focused on consonants, vowels, and on perception. Most importantly, it highlights the fact that phonetic drift effects have been attested both in naturalistic L2 learning (that is, L2-immersion environments) as well as in L1-dominant contexts.

The following two chapters narrow the perspective down somewhat and discuss the issues related to the phonetics and phonology of the language pair under investigation in the present work, namely Polish and English. They provide a selection of studies that investigated the success with which Polish learners acquire English contrasts. Chapter 3 focuses on laryngeal phonology. First of all it compares and contrasts the phonetics of laryngeal contrasts in Polish and English. Then it describes four different phonological theories which strive to account for two-way laryngeal systems, such as Polish and English, namely Feature Theory (Chomsky and Halle 1968), Laryngeal Realism (Honeybone 2005), Laryngeal Relativism (Cyran 2011, 2014), and Onset Prominence (Schwartz 2017). An overview of the major tenets of each framework is presented, simultaneously highlighting some underlying issues thereof (e.g. the units used or the omnipresent segmental bias). In this chapter we see that each theory makes predictions about the direction of phonetic drift in L1; these predictions differ depending on what the given theory sees as equivalent.

Chapter 4 focuses on vowels. Once again we start with the discussion of the phonetics of Polish and English vocalic inventories, alongside some phonetic studies that looked at both perception and production of English vowels by Polish learners. Then we look at three phonological theories and the ways in which they represent vowels (Feature Theory, Element Theory (Harris 1991), and Onset Prominence). Contrary to what was the case for laryngeal phonology, in the case of vowels it is only Onset Prominence that makes explicit claims about the behaviour of vowels in second language acquisition, in particular in the context of phonetic drift.

Chapter 5 introduces the study conducted for the purposes of the present dissertation. It describes the methodology of the experiment, including the materials used in the longitudinal study, the participants, and procedures. It highlights some of the methodological improvements over the phonetic studies presented in Chapter 2. The crucial research questions that the work sets out to investigate are whether phonetic drift can be attested in the productions of Polish students enrolled in the English Programme at a university who undergo intensive phonetic instruction in their L2 and what the phonological implications of the data we gather are.

Chapters 6 and 7 are devoted to the results. The former thoroughly describes two production studies on L1 Polish, focusing on both initial stop consonants and vowels in #CV contexts. It presents data from two tasks: word reading and sentence reading. Chapter 7 zooms in on the word reading task and juxtaposes the L1 Polish results with L2 productions at corresponding recording sessions in order to investigate the possible links between L1 phonetic drift and L2 acquisition.

Finally, Chapter 8 summarises the findings and places them within a larger body of phonological theory. It aims to present the phonological implications of the data obtained in the acoustic experiments and assesses the predictive powers of the frameworks presented in Chapters 3 and 4.

Chapter 1: L2 speech: theories and models

1.1. The aim of the chapter: theories of non-native language acquisition

Before we begin the exploration of the questions concerning the influence that second language might exert upon one's native tongue, as has been mentioned in the Introduction, we must take a step back and look at the acquisition of L2 speech. Over the years, a number of theories of non-native language acquisition have been proposed. These L2 speech models are important insofar as they provide the researchers with predictions as to how the process of foreign language learning will progress and offer testable hypotheses, on the basis of which specific research questions may be formulated and subsequently explored in empirical studies. In this chapter we will focus mainly on the predictions with respect to L2 speech perception and production. A selection of most well-known models is presented below. Most important assumptions and predictions of each are discussed and empirical findings supporting or undermining those are presented. Finally, one model in particular – namely the *Speech Learning Model* (Flege 1995) – is chosen as one of the theoretical basis of the empirical research questions investigated in the present dissertation. This choice is explained in section 1.1.6. The chapter ends with a general discussion.

1.1.1. Critical period hypothesis

First proposed by Penfield and Roberts (1959), the critical period hypothesis gained more notoriety after the publication of Lenneberg's (1967) work on *Biological Foundations of Language*. In his book, Lenneberg lists a number of arguments speaking in favour of age limitations associated with language acquisition. He points out that the ability to develop language diminishes after a certain age, citing examples of feral children (1967: 141) and discusses the links between age and the likelihood of recovering from aphasia, stressing that full retention of language is much more likely in the case of children than adults (1967: 142ff).

This hypothesis was subsequently extended to theories of second language acquisition theories. Lenneberg commented that while it is possible to learn a foreign language after the age of twenty, "the incidence of 'language-learning blocks' rapidly increases after puberty" (1967: 176). Furthermore, language acquisition from mere immersion in a given language is said to no longer be possible (1967: 176). Finally, although one should be able to learn to *communicate* in a foreign language, losing a foreign accent after puberty is difficult, if not impossible (1967: 176). He supports his hypothesis with examples from other critical biological periods (1967: 162ff).

The critical period hypothesis has long been the centre of a heated debate amongst linguists, as these claims did not find support in a number of the empirical studies that followed. First of all, the claim that children acquiring a foreign language will easily lose foreign accent has been called into question. A longitudinal study conducted by Flege et al. (2006), they found that native Korean children who immigrated to the United States between the ages of 6 and 14, spoke with a detectable foreign accent, and the degree of accentedness did not appear to decrease over time. Even though the children seemed to outperform the native Korean adult group, which was also studied, a confounding issue was that the children reported much more L2 input, which might have skewed the results to a certain extent. They spoke English to friends to a larger extent and used it at home much more often than the adults did.

Second of all, there have been numerous studies showing that speaking without a foreign accent after acquiring a foreign language later in life is indeed possible. Munro et al. (1996) studied the productions of English vowels by 240 Italian residents of Canada, whose age of arrival ranged from 2 to 23 years. In a foreign accent rating experiment, they found that the range of age of arrival effects on the presence of a foreign accent ranged from 7.5 to 15.8 years old. They also found substantial differences between particular vowels, with some vowels – such as [ɛ] or [æ] – being easier to acquire even for those arriving to Canada later in their life. More than half of this group produced those two vowels with native-like accuracy. In turn, Flege et al. (1995) investigated the success rate of acquiring English [ɹ] and [l] by Japanese native speakers, sounds which are missing from Japanese inventory. They found that experienced Japanese speakers of English were able to accurately master English liquids and their productions fell within the mean rating range of native English speakers' productions. Such results were also obtained by Bongaerts et al. (1997), who asked native English speakers to rate English productions of advanced Dutch-English late bilinguals. They found that the ratings of a number of Dutch speakers fell within the range of native English controls'. They highlighted the fact that learning contexts and the speaker's characteristics might surpass the difficulties associated with starting to learn the language later in life. A similar comment has been made by Kuhl (2000), who notes that learners do not lose the ability to perceive non-native contrasts and minimising the influence of memory, together with phonetic training, might boost the adult speakers' performance.

These findings, taken together, constitute strong evidence against a strict interpretation of the critical period hypothesis. Nonetheless, age effects, on the whole, are found – it is likely that the earlier one starts acquiring the language, the more success they can achieve.

1.1.2. Contrastive analysis hypothesis

With respect to second language acquisition, contrastive analysis approach (Lado 1957) postulated that if an L2 sound is similar to a sound existing in one's L1, it will be learnt relatively easily. In this instance the L1 is helpful in mastering a foreign sound, which is referred to as *positive transfer* (Benson 2002). In turn, L2 sounds which are much different to the ones in a speaker's L1 inventory, will be more problematic, which means that in this case L1 plays a detrimental role. This is an example of *negative transfer* (Benson 2002). Therefore, this hypothesis suggests that the main cause of problems with L2 learning is the interference from L1 (cf. Werinreich 1953). It is necessary to point out that contrastive analysis does not limit itself only to the acquisition of sound systems, but may be extended to other linguistic levels, i.e. lexicon or syntax.

Contrastive analysis has been challenged by the findings of a number of experiments. In one of them, Flege (1987) investigated the productions of English native learners of French, with various degree of proficiency in French and compared their realisations of two French vowels, [u] and [y] with a group of French monolinguals. If contrastive analysis hypothesis were to be borne out, English-French bilinguals should realise the vowel [u] with native-like accuracy, since [u] is present in English vocalic inventory. In turn, since [y] is not a native English sound, it should not be easy for English speakers to acquire, and therefore should not approximate French norms. The results turned out to show the exact opposite. The vowel [y] was realised with F2 frequencies only slightly lower than French monolinguals' productions, with only the least proficient group displaying significant differences. However, all groups produced French [u] with more English-like F2 values, significantly different from French norms. These results were true even for very advanced English-French bilinguals who had been living in Paris for more than twelve years.

An influential critique of contrastive analysis came from Wardhaugh (1970). He pointed to the fact that a comparison of L1 and L2 on the phonemic level is not enough to predict all problems that the learner might display (cf. Weinreich 1953). As such Lado's claims can be seen as an oversimplification. Wardhaugh offered an alternative, which is a weaker version of contrastive analysis. Instead of foreseeing what would cause difficulties for an L2 speaker and what would not, he proposed to first look at evidence. That is, we should assess possible problems that the learners are displaying, such as translation errors, foreign accentedness, and only on the basis of the observed issues should we make reference to the two systems (1970: 7f).

1.1.3. The Perceptual Assimilation Model and PAM-L2

In its original form, the Perceptual Assimilation Model (henceforth: PAM; Best 1995) focused on the perception of non-native contrasts (rather than individual phones or segments) by naïve listeners. The direct realist view of speech, on which PAM is built, postulates that phonetics and phonology are very much articulation oriented (Best 1995: 182). As such, perception of non-native contrasts is based upon the listeners' judgements as to how similar or different a given foreign sound is to a gesture used to produce a native segment (Best 1995: 193). Therefore it is the proximity of articulatory acts which are claimed to predict the perceptual assimilation of non-native and native contrasts (Best 1995: 194). In other words, the speaker perceives a non-native contrast in relation (and an approximation) to the native gesture they deem to be the closest to what they hear in a foreign language. Some foreign sounds may be so different to the gestures employed by the listener's native language that they will not be assimilated to any particular native category or even to speech as such (Best 1995: 194).

PAM postulates that, on the whole, non-native contrasts should be treated in the same manner and only some of them should pose perceptual difficulty (Best 1995: 191).

There are six patterns predicting the accuracy with which non-native contrasts will be perceived; for clarity, they are summarised in Table 1.

Table 1. Patterns in the accuracy of non-native speech perception according to PAM (after Best 1995: 195).

Assimilation pattern	Abbreviation	Description	Predicted degree of discrimination
Two-category	TC type	each non-native segment assimilated to a different native category	excellent
Single-category	SC type	both non-native segments assimilated to the same native category	poor
Category-goodness	CG type	both non-native segments assimilated to the same native category, but one is a better match	moderate to very good
Uncategorisable	UU type	both non-native segment not assimilated to any native category	poor to very good (depending on the proximity to native categories)
Uncategorised vs. categorised	UC type	one non-native segment assimilated to a native category, but not the other one	very good
Nonassimilable	NA type	categorised as non-speech sounds	good to very good

There have been a number of empirical studies conducted to test PAM's hypotheses, however they will not be reported on in this work as they fall out of the scope of the present thesis. This stems from the fact that, as has been mentioned at the beginning of this section, PAM is concerned with how naïve listeners perceive foreign contrasts while this thesis is focused on the phonetic realisation of individual phones.

PAM was subsequently extended (Best and Tyler 2007) to account for the fact that similarly to how adults perceive language differently than infants, bilingual speakers perceive L2 and L1 contrasts differently than monolinguals of both those languages. As a result, PAM-L2 was created. The major tenet of PAM-L2 is that it is concerned with "natural communicative situations" (2007: 18), as opposed to foreign language acquisition in a classroom context, since the latter includes formal instruction which does not entail L2 usage in more natural communicative contexts. It highlights the fact that non-native L2 teachers, or L2 teachers who speak dialects which differ from the L2 standard, can affect L2 perception. Therefore, the authors distinguish between "true" L2 listeners, exposed to L2 as their predominant language of input, and those that acquire it during foreign language classes. In other words, then, they distinguish between L2-immersion and L1-dominant contexts. Their sole focus is on the former group.

Another key aspect of PAM-L2 is that it disposes of any mental representations in the traditional meaning thereof (Best and Tyler 2007: 25). Essentially, rather than forming representations of categories, listeners learn to observe variants of articulatory gestures employed by the speakers in L1 and L2. Furthermore, PAM-L2 distinguishes between phonetic and phonological levels, in addition to gestural level (Best and Tyler 2007: 25). The latter is defined as the level of phonological categories, which generally refer to phonemes, as they indicate the minimal units necessary to express lexical contrasts in a language. The former is concerned with allophonic variation, be it positional or stemming from the speaker's dialect usage. They are all articulatory in nature. The differences between L1 and L2 at any of those three levels might affect the accuracy of perceiving contrast by a listener.

Phonetic and phonological levels are claimed to interact; an L2 sound might be assimilated to an L1 category at the phonological level, but the two might be kept separate at the phonetic level (Best and Tyler 2007: 26). In the case of an overlap of L1 and L2 phonetic categories, L2 learners might seek to dissimilate the categories between the two languages, which might lead to "a global shift of all related L1 and L2 categories at the phonetic level" (2007: 27).

Even though PAM-L2 is generally grounded within a gestural framework and focuses on the way in which articulation of L2 sounds is perceived by a listener, it does not directly address the issue of L2-speech production. As a corollary, possible effects of L2 on L1 are discussed – albeit briefly – also merely in relation to L1 perception. We will return to this point in section 1.1.6.2.

1.1.4. Second Language Linguistic Perception model

Perception is also crucial for another model of L2 speech, namely the Second Language Linguistic Perception (henceforth: L2LP; Escudero 2005, van Leussen and Escudero 2015). This computational model is grounded within Bidirectional Phonetics and Phonology framework (also referred to as BiPhon; Boersma 1998 et seq.), which is implemented within Optimality Theory (OT; Prince and Smolensky 1993).

Contrarily to PAM, which focuses on naïve listeners, the main distinguishing aspect of this model is that it projects the entire trajectory of speech perception: from naïve non-native to native-like (2015: 2). In light of the *optimal perception hypothesis* (cf. Escudero 2005), this model conjectures that at the very start of second language acquisition, the state of L2 is the result of L1 acquisition (van Leussen and Escudero 2015: 3). In other words, at the initial stages what is available to the learner is whatever L1 phonology they have. From there, the authors predict three scenarios: *new*, *similar*, and *subset*. These scenarios are analogous to PAM's SC-type, TC-type, and UC-type respectively and as such will not be elaborated on at present. The central assumption is the so-called *full copying* hypothesis, which states that L2 learning starts on the basis of a copy of their L1 grammar. As the learning progresses and the exposure to L2 becomes more robust, the copy adjusts so as to better account for the L2 contrasts (van Leussen and Escudero 2015: 5). The advantage of L2LP over PAM appears to be the ability to point out prospective difficulties that the learner might face, which transcend merely perceptual problems; namely, L2LP focuses not only on discrimination, but also on how the learner utilises the newly acquired L2 categories to expand their mental L2 lexicon (2015: 3).

Despite its similarities to PAM in some respects, as has been shown above, L2LP does not reject the notion of mental representations. On the contrary, L2LP differentiates between four levels of representations and connections (2015: 5), namely (from bottom up) acoustic, phonetic, phonemic, and lexical. The learners update their L2 lexical representations in the course of L2 learning via the links connecting the three levels below it. The connections strengthen and weaken, in accordance with the richness of the L2 input that the learners are receiving, in order to finally arrive at a representation that best matches the intended form.

It is not, however, clear to what extent the two grammars influence one another once the learner reaches the final step of foreign language acquisition. Since *full copying* results in the formation of a copy of the L1 grammar which later becomes attuned to the needs of an L2, the relationship and mutual influence between the two in advanced bilinguals could be predicted to be minimal.

1.1.5. Natural Growth Theory of Acquisition

The Natural Growth Theory of Acquisition (henceforth: NGTA; Dziubalska-Kołaczyk 2016, Dziubalska-Kołaczyk and Wrembel 2019; Dziubalska-Kołaczyk and Wrembe; 2022) is an example of a more holistic approach to the acquisition of language, and in this respect it differs from the theories which have so far been discussed. It strives to provide the big picture of language learning, instead of focusing on "elicited details", and is open to interdisciplinarity. It stems from Natural Phonology (e.g. Stampe 1979; Donegan and Stampe 2009; Dziubalska-Kołaczyk 2002) and is enhanced by Complexity Theory (Kretzschmar 2015). The theory aims to account for both first language acquisition as well as the acquisition of second and subsequent languages but has a potential to be extended to model other aspects of linguistic structures' learning. For the purposes of time and space, we will focus here mainly on the assumptions made with regard to the SLA, though it remains unclear whether the theory makes any reference to possible effects of just the second language learning on one's native language.

Natural phonology explicitly distinguishes between automatic first language acquisition and a more controlled second language learning, which may – but does not have to – result in successful mastering of a foreign language, provided "all the conditioning acquisition factors are favourable" (Dziubalska-Kołaczyk 1990: 65). The learner of a second language is different from a child acquiring their first language also with respect to the motivation, the fact that the former already possesses one language in their mind, and also their attitude towards learning a new language (Dziubalska-Kołaczyk 1990: 66).

Dziubalska-Kołaczyk (1990) proposed a model of second language phonology acquisition, embedded within Natural Linguistics. It is illustrated in Fig. 1 below and sketches the interplay of foreign input, instruction, frequency, and universal phonetic processes.

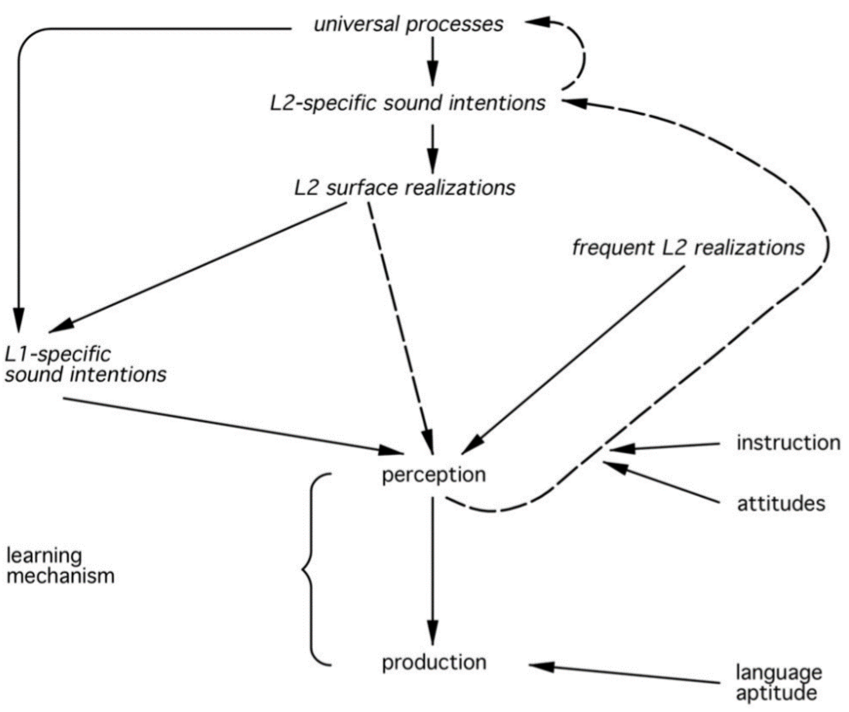


Fig. 1. A model of acquisition of second language phonology (after Dziubalska-Kołaczyk 1990), employed and developed further in NGTA.

The learner perceives the non-native input on the phonemic level, with phonemes in Natural Phonology being understood as sound intentions. They arrive at the correct phonemic forms by means of training their perception of the phonetic realisations of non-native contrasts and imitating the input that they receive. Once they successfully attain the foreign sound intentions, the production thereof should resemble that of native speakers of the target language (Dziubalska-Kołaczyk 1990: 65f.).

NGTA develops the insights coming from Natural Phonology and complex systems (Kretzschmar and Dziubalska-Kołaczyk *submitted*), the latter of which argue in favour of emergentism and frequentist perspective, in order to account for the omnipresence of multilingualism. It aims to stress the complexity of multilingual acquisition and as such, it takes into consideration both linguistic (e.g. L1, L2 and subsequent languages, as well as preferability generalisations, which are both inductive and deductive in nature) and extralinguistic variables (e.g. stage of acquisition, age, frequency of input and use, proficiency, metalinguistic awareness, and individual factors) (Dziubalska-Kołaczyk and Wrembel 2019) present in the process of acquisition. The former group always affects the process but the extent to which they exert their influence may vary, as it is governed by the variables from the latter group. So far, the main tenets of the theory have been applied to explain data coming from L3 learning and morphonotactics (Dziubalska-Kołaczyk and Wrembel 2022); there is, however, a need for its further development and experimental support, as the theory was created after the aforementioned data from L3 and morphonotactics had already been obtained and analysed. Nonetheless, as noted by the authors, "[the] theory provides a potential to build models of acquisition showing dynamic emergence of structures with respect to selected extralinguistic variables and individual paths of acquisition. Finally, it entails that a multilingual mind is more than a sum of monolingual minds" (Dziubalska-Kołaczyk and Wrembel 2022). In particular, it highlights the individual differences in the learners' process of acquisition, even when they have similar linguistic backgrounds, not focusing solely on age effects, which sets it apart from other theories discussed in this chapter and constitutes a promising new outlook on bilingual and multilingual language learning.

1.1.6. The Speech Learning Model

The Speech Learning Model (henceforth: the SLM; Flege 1995 et seq.) has been developed in order to investigate whether there exist sounds in learners' L2s which they are unable to correctly produce as well as to which extent the correct perception of those sounds influences their actual realisation (Flege 2005). The core idea behind the creation of the SLM was to study the phenomenon of foreign accent, and in particular the effects of age upon successful acquisition of non-native sounds (Flege 1995: 237). It is mainly focused on proficient bilinguals, rather than beginners or naïve listeners, which sets it apart from the models discussed previously (e.g. PAM or LPL2).

Flege (1995: 239) enumerates a number of postulates and hypotheses put forward by the SLM. The first assumption is that the ability of learning a new language does not disappear with aging, and hence new categories might be formed at any point in one's life. The effects of age are conceived of differently than what we saw in critical period hypothesis (cf. section 1.1.1). According to the SLM, the age at which one begins learning a new language will affect the accurate production in that the earlier one begins, the less phonetic information is necessary for a new phonetic category to be formed (1995: 264). In other words, late bilinguals need more phonetic distance between two sounds for them to form a new category and keep it separate from an already existing L1 one. This might also stem from the fact that older learners' L1 categories are fully developed, while for children they are less stable and hence "represent weaker 'attractors'", which makes the fusion of an L1 and L2 category less likely (Flege 2005: 112). Interestingly, the question as to whether age effects – aside from phonetic similarity – may also be observed in perception was left unanswered (Flege 1995: 268). Crucially, even though age effects are postulated to exert influence on the successful category formation in L2, adult speakers are argued to be capable of a similar degree of success (Flege 2005: 112) – as has been mentioned, the ability of category formation is not lost as one gets older. This goes in line with e.g. Bongaerts et al. (1997) cited in section on the CPH.

The phonetic categories which a child forms in the process of language acquisition are not set in stone – they can change and evolve throughout one's life span. Bilinguals will try to keep L1 and L2 phonetic categories separate (Flege 1995: 239).

A pertinent question that might be asked at this point is what Flege means when he talks about "phonetic categories". After all, it appears to be an oxymoron on the face of it – phonetics is usually associated with gradience, while categories are on the whole associated with the discreteness of phonology. Indeed, the SLM – as opposed to for instance PAM and PAM-L2 – does not reject categories as mental representations of speech sounds. Flege defines a phonetic category as "perceptual in nature, (…) stored in long-term memory and (…) intermediate in degree of abstractness to phonological codes – phonemes – that differentiate words in the mental lexicon [and] the sensory input obtained via the ears and eyes" (Flege 2016: 22). These categories are language specific and they encompass all physical characteristics which are necessary in order for children to differentiate between two sounds of one language as well as to distinguish between pairs of sounds from two different languages, even when they are denoted by a single IPA symbol (Flege 2016: 23). In the process of the development of these categories, the speaker becomes either more sensitive to the distinctiveness of two categories or, on the contrary, fully desensitised to the differences between them; in the latter scenario they are judged as instances of the same category (Flege 2016: 22). What is important is that in the SLM perception guides production[1]. That means that it is the level of phonetic categories that influences how the speaker implements the rules they observed in the process of language learning in their actual realisations. This is illustrated in Fig. 2 (after Flege 2016: 26).

[1] The revised version of the SLM has somewhat different assumptions about the perception-production relationship (cf. section 1.1.6.3).

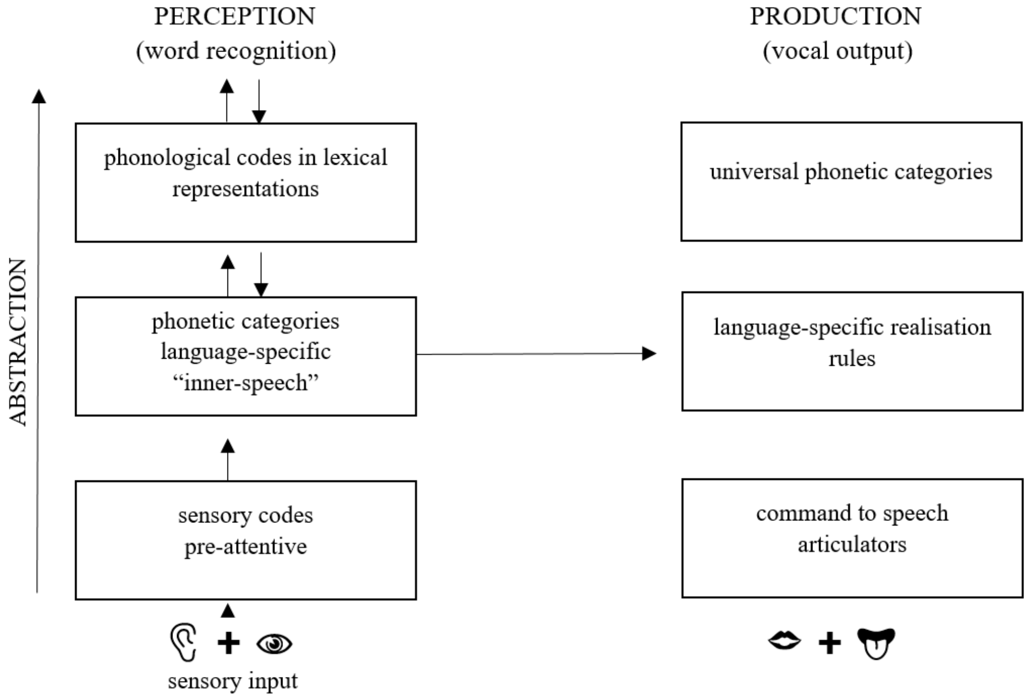


Fig. 2. The link between perception and production in SLM (after Flege 2016: 26).

The claim concerning the link between perception and production was rejected by, for example, Bever (1981), He argued that the "psychogrammar" which serves as the bridge between the perception and production is lost once the L1 is fully acquired and can only be maintained by constant learning of new languages. Flege (1999) reports on a number of studies that looked at the correlation between perception and production and yielded moderate outcome (i.e. r=0.50). He suggested a few explanations as to why the correlation has not been found to be stronger. First of all, he noted that speakers might be able to correctly perceive an L2 sound and create a new phonetic category but this ability might either not be extrapolated to production at all, or the implementation of the perceived contrast might take time. That points to an advantage of longitudinal studies over cross-sectional ones: a speaker might be unable to accurately produce a sound at T1, but the result might be different at T2. In other words, "accurate perception does not entail accurate production whereas accurate production requires accurate perception" (Flege 2016: 31).

Furthermore, phonetic categories, as opposed to much more abstract phonemes, are claimed to develop more slowly (Flege 2016: 43) and to frequently be more numerous than the phonemic inventory (Flege 2016: 63). The reason for this slow pace of development of phonetic categories is that children need to receive sufficient amount of input in order to become aware of all the physical characteristics of a given category, which will in turn enable them correct differentiation between sounds even in "not-ideal listening conditions" (Flege 2016: 46). The rate with which children acquire the categories might depend upon the complexity of the physical make-up of a given category and thus might be language-dependent. Since phonetic categories are less abstract than phonemes, they are argued to be accessible to the speakers, which is of particular importance for L2 learners (Flege 2016: 52). In one of his studies, Flege investigated the learners' perception of their own foreign accent in L2 (1988). He investigated Taiwanese learners of English who differed with respect to the length of residence in the United States (1.2 years vs. 5.1 years). While the degree of the foreign accent in productions of both groups was relatively similar, their awareness of it increased proportionally to their L2 experience. These findings suggest that with experience, the learners are able to access more and more phonetically relevant information, which boosts their perception and perhaps, eventually, also production.

1.1.6.1. Equivalence classification

One of the most relevant principles of the SLM is the notion of *equivalence classification*. It is defined as "a basic cognitive mechanism which permits humans to perceive constant categories in the face of the inherent sensory variability in the many physical exemplars which may instantiate a category" (Flege 1987: 49). In other words, this mechanism aids children in perceiving the different realisations of the same sounds of their ambient language as being instances of the same category, even when they are produced by different speakers (Flege 1987: 50). This assumption is similar to the key postulates of exemplar-based models (e.g. Bybee 2013).

When a person starts learning their second language they already have their L1 categories attuned to L1 norms; yet, similarly to children acquiring language, L2 learners might also be inclined to assess L2 sounds as being exemplars of their L1 categories, since these are the only ones that they possess (Flege 1987: 50). It is hypothesised, then, that L2 learners should have fewer problems with correctly perceiving a completely new sound as opposed to a sound which is similar to an existing L1 category. In the latter scenario, those two sounds will be deemed as equivalent – the L2 sound will be treated as another realisation of the L1 category and the speaker will tune out the acoustic details which distinguish the two; as a result, the two categories will be fused together and the creation of a new category will be blocked. However, if the new L2 sound is different enough, equivalence classification is not predicted and, what follows, a formation of a new L2 category is expected (Flege 1987: 48; Flege 1995: 239). Recall that those assumptions contradict the predictions of the contrastive analysis, discussed in section 1.1.2. It is worth noting that the study which we discussed therein (i.e. Flege 1987), whose aim was to impugn the conjectures of that model, goes exactly in line with the assumptions that equivalence classification entails. Indeed, English speakers appeared to acquire a new sound – French [y] – without much trouble, yet French [u] seemed to merge with their English [u] category with respect to the F2 values, as they were perceived as equivalent.

1.1.6.2. Why the SLM?

The SLM is undoubtedly a model with a great predictive power, whose hypotheses have been extensively tested and supported by empirical investigations. It is also chosen as the model whose theoretical tenets serve as the basis for the research questions and hypotheses put forward in the present work, to which we will return in section 5.2. One might ask: why exactly is this thesis based on the SLM? This subsection aims to explain this choice.

Aside from the fact that, as has been already mentioned, the SLM – to the best of my knowledge – is one of few models that explicitly addresses the link between perception and production, it also makes direct predictions regarding the bi-directionality of the interaction between L1 and L2. In other words, it allows us to formulate hypotheses regarding the influence which second language acquisition may exert on one’s L1. Admittedly, within PAM-L2 a reference is made to the fact that insufficient L1 input in an L2-immersion environment might result in changes in L1 perception, but the overall predictions pertaining to those changes and, crucially, whether or not any effects on production might be expected, are rather inconclusive (Best and Tyler 2007).

Not only does the SLM not limit itself to learners immersed in L2-dominant communities, it makes two precise predictions on how L2 learning might influence L1 speech production. These predictions are based upon the basic principle of equivalence classification, which has been described in section 1.1.6.1. Anecdotally, this observation was made by Flege in 1979 when he was still writing his PhD dissertation and his friends made a remark that after getting back from West Africa after years of their absence, their American peers made fun of their English accent (Flege 2005: 75). Recall that the SLM postulates that in the process of L2 learning, the relative ease – and accuracy – of perception and production of an L2 sound depends upon whether or not a new category will be formed. Conversely, "L1 and L2 sounds that are perceptually linked to one another (diphones) come to resemble one another in production" (Flege 1995: 241). Therefore, if two sounds are perceived as equivalent, not only will this result in discernible accentedness in L2 productions of that diphone, but it may simultaneously also lead to it diverging from monolingual L1 norms. When the two sounds are assigned to two distinct categories, less interaction is predicted and the chance of foreign accent in L2 as well as L2-induced effects on L1 decreases. An early phonetic study conducted by Flege (1987) might be referenced at this point to illustrate this.

He studied the productions of forty two women who were divided into four groups, according to their linguistic background. Two of those groups are of particular interest to the present argument, namely the group of Americans who had been living in Paris for ca. 11.7 years and who were advanced English-French bilinguals, married to native French speakers and a group of French subjects, who were proficient French-English bilinguals, and who had been living in Chicago for ca. 12.2 years. Those speakers were contrasted with English and French monolingual control groups respectively. Flege (1987: 55) found that the American experienced speakers of French produced /t/ with VOT values shorter, relative to the English monolingual control group (49ms vs. 77ms; contrast estimate: 28ms), which made their English /t/ productions more French-like. In turn, the French living in Chicago produced their French /t/ in a more English-like manner, with VOT values longer than the French monolinguals' productions (51ms vs. 33ms; contrast estimate: 18ms). Similar effects of L2 influence on L1 were not present in the productions of less proficient speakers. Notably, the effects observed in this study are quite possibly examples of phonetic attrition, rather than phonetic drift (cf. section 2.2), however what is of importance here is the presence and the explicit comments made on the L2 influence on L1.

An interesting prediction is made by SLM with respect to L2 effects on L1 vocalic inventories. If a new category is not established for an L2 vowel, then the speakers are predicted to produce both the L1 and L2 vowels (merged into a diphone) with intermediate formant values. Flege illustrates this on the example of a Spanish speaker who classifies Spanish [a] and English [æ] as equivalent – they are predicted to produce Spanish [a] with F2 values which are too high for Spanish monolingual norms, and English [æ] with F2 values which are too low, relative to English monolinguals' productions (1995: 243). If a new category is established, two scenarios are possible: either there will be no interaction between the L1 and L2 sounds, or the effects will be dissimilatory. That is, the Spanish [a] will be produced as lower, in order to maintain contrast between the L1 and L2 (1995: 244). Flege notes that "one gap in literature is the virtual absence of work examining bilingual's production of L1 vowels" (1995: 244), which has since, thankfully, been remedied.

The idea of the mutual bidirectional interaction, be it assimilating or deflecting, between L1 and L2 sounds has another valid implication. Namely, it suggests that L1 and L2 sounds exist in a single common phonological space (Flege 1995: 242). Note that this is a salient divergence from the assumptions indirectly made by for instance L2LP (cf. section 1.1.4). Since the present work sets out to investigate the influence of English on the speech of Polish learners of English, the SLM is the model whose assumptions are most relevant. We are going to assume that it is equivalence classification that lies at the root of the effects we will be exploring.

1.1.6.3. The revised Speech Learning Model (the SLM-r)

Recently, Flege and Bohn (2021) proposed a revised version of the SLM (henceforth: SLM-r), wherein some of the postulates are re-formulated on the basis of the research into second language acquisition conducted in the last twenty-five years, while some are left intact.

Similarly to the original version, the SLM-r is mostly concerned with the way in which sequential bilinguals perceive and produce non-native sounds. The main objective is no longer to assess whether or not advanced L2 speakers attain native-like accuracy in these two areas, as they do not receive input that could be compared with that of the native speakers of the language in question. Furthermore, the SLM-r no longer assumes that production lags behind perception, but rather that the two co-evolve and does not pose any a-priori predictions regarding the effects of age on successful acquisition of L2 phonetic categories. In other words, it is possible to proficiently master non-native speech sounds regardless of the age at which one begins learning that language.

While the core ideas upon the formation of phonetic categories in L2 learning and equivalence classification remain mostly intact, some new hypotheses are introduced. First of all, the SLM-r recognises the fact that while a new phonetic category formed in the process of L2 acquisition might be different from the category of the same sound formed by a native speaker of a given language due to the fact that no two languages use the same features in an identical manner. However, the *full-access* hypothesis (Flege 2005) assumes that during the course of L2 acquisition those non-native features are available to the learner, therefore there is no limit to the accuracy with which the category is going to be acquired. In other words, the perceptual cues which make up new L2 categories or L1-L2 merged diphones are weighted by the learner, and with sufficient input the learner is able to accurately categorise the non-native sounds and learn to produce them. The relative success depends on how easy it is for a learner to perceive the difference between the new L2 sound and the closest L1 sound as well as how precisely defined the latter is in their inventory. This entails another important change that SLM-r brings about, namely the individual factor. Essentially, we cannot predict that all speakers of the same language learning the same L2 will obtain the same level of proficiency in sound production. Nonetheless, "L1 and L2 phonetic subsystems remain malleable across the life span, responding to variation in the phonetic input that has been received, even recent input" (Flege and Bohn 2021: 63), implying that we cannot assume that the learner has reached their "end state" in L2 learning unless they remove themselves from the learning environment altogether.

In sum, the most relevant new assumptions that the SLM-r presents is the focus on individual variation and disposing of the "age" factor entirely.

1.2. Discussion

This chapter has presented selected models of second language acquisition. We have seen some of the older theories, whose assumptions have been contested by empirical findings as well as more recent models. Those ranged from models focused solely on perception of non-native contrasts, to those taking into account links between perception and production, finally to those which either model the entire path of second language acquisition and strive to predict the obstacles that learners might face or aim to look at acquisition more holistically, in order to reconcile second language acquisition with the multilingual reality of the present world.

The chapter has paid special attention to one model in particular, namely the SLM, so as to defend choosing it as the model on which some of the research questions investigated in the present thesis will be based.

Before moving on to discussing empirical findings associated with L2 influence on L1, it is necessary to touch upon some other issues connected with L2 learning, not necessarily explicitly addressed by any of the L2 speech models, as well as to review some literature which explores the nature of L2 learning in a formal context. The two following subsections are devoted to those topics.

1.2.1. Other issues in L2 speech acquisition

As can be noted on the basis of the discussion presented in this chapter, most research regarding L2 speech is informed by the L2 speech models, which are concerned with the relationships between native and non-native contrasts. Bohn (2020), however, stresses that cross-language phonetic interactions do not account for all L2 speech learning problems (2020: 171). In fact, there exist "universal phonetic biases" which are independent of the learner's L1 (2020: 175) as well as generalised sensitivities which are linked to L1 experience but which are not directly addressed by L2 speech models.

First of such biases has been formulated as a *desensitisation hypothesis* (Bohn 1995: 294f.) which states that if a learner is not sensitive to certain spectral differences that distinguish L2 vowels, and this lack of sensitivity stems from the fact that these differences are not employed by their L1, they will try to identify the vowels on the basis of their duration. Crucially, the duration cue will be utilised regardless of whether or not temporal cues are phonologically contrastive in the learner's L1. Therefore, this phenomenon is independent of the characteristics of a learner's L1 and appears to be universal. Desensitisation hypothesis has been supported by a number of empirical studies; for example Balas [Bogacka] 2004 showed that L1 Polish learners of English relied strongly on vowel quantity when asked to discriminate between English high vowels (/iː/-/ɪ/ and /uː/-/ʊ/), even though length is not a contrastive feature in neither Polish nor English.

The second bias comes from *Natural Referent Vowel* framework (Polka and Bohn 2011), which highlights the special status of peripheral vowels as "natural referents (…) possibly because of their acoustic properties which [are] describe[d by some sources] as formant focalization" (Bohn 2020: 176). Essentially, certain asymmetries in the successful discrimination of foreign vowel might not be the result of the L1 background, that is the learner's difficulty of perceiving certain vowels due to their L1 vocalic inventory, but rather of the favoured peripherality. For instance, Polka and Bohn (2011) studied L1 Danish learners of English on their perception of the /ʌ/-/ɒ/ contrast in English (which is not present in Danish) and showed that not only are peripheral vowels easier to discriminate but also that the direction of the presentation of tokens matters. It was shown that going from less to more peripheral (i.e. from /ʌ/ to /ɒ/) yielded higher accuracy of discrimination (69.2%) than when going from more to less peripheral (50.5%, which is discrimination at chance level). This trend has also been observed in infant perception of vowels in the process of L1 acquisition (cf. Polka and Bohn 1996).

Finally, a number of studies have shown that some generalised sensitivities based upon one's L1 may be of help in L2 speech perception. In particular, two perception experiments are worth mentioning. Bohn and Best (2012) tested the perception of English approximants /r, j, l, w/ by L1 Danish and L1 Danish and German lack /w/ in their consonant inventories, but both have very rich vowel systems. It was shown that both L2 groups outperformed L1 English speakers in the /w/-/j/ discrimination task (Bohn and Best 2012: 125ff.). The authors note that both Danish and German use lip rounding as a contrastive feature in their vocalic inventories – both languages discriminate between /i(ː)/ and /y(ː)/, while for English lip rounding is only a secondary cue. It was thus suggested that attending to this contrastive feature with respect to vowels simultaneously boosts the performance in consonant discrimination as the sensitivity to lip rounding in L1 is generalised onto foreign contrasts as well. Pajak and Levy (2014) show that this is true also for consonants. They demonstrate that sensitivity to contrasts based on a particular contrastive feature in L1 is enhanced with respect to non-native contrasts in different categories, which are based on the same feature. They investigated the perception of consonantal length in Polish by Korean, Vietnamese, Cantonese, and Mandarin speakers. It was shown that Korean (with both long vowels and consonants) and Vietnamese (with long vowels) outperformed Cantonese (with length being a secondary cue) and Mandarin (with no length contrasts) participants. Crucially, Vietnamese speakers did better than Cantonese and Mandarin ones, even though length is not exploited with respect to consonantal contrasts in this language. This suggests that the sensitivity was again generalised onto other L2 categories.

Therefore, it is important, when studying L2 acquisition in speech production and perception, to consider these aforementioned universal phonetic biases as well as the L1 generalised sensitivities, which both contribute to the results that our studies might obtain.

1.2.2. Individual differences

It has been demonstrated that the success with which learners acquire their second language depends on many extralinguistic factors (already mentioned in section 1.1.5 in the discussion on the NGTA framework, which is a holistic theory of (any) language acquisition), such as aptitude, motivation (Dörnyei 2005) but also working memory, processing speed, or attention control (Darcy et al. 2015). Therefore, it appears that looking at group performance only might be quite a myopic approach as it does not tell us to what extent a given L2 contrast has been mastered in individual cases.

To provide an example of how focusing on groups only might hide some results from us, Díaz et al. (2012) conducted a perception study on (late) Dutch learners of English and their ability to perceive the /æ/-/e/ contrast in several tasks. Group results indicated that overall non-native speakers performed worse than native speakers of English, which is quite an intuitive result to obtain. However, once the authors looked at individual learner's performance it was found that in the categorisation task almost 50% of the Dutch participants attained native-like success rate. Smaller percentages of L2 learners also scored within native norms in the identification and lexical decision tasks. The authors concluded that the performance depended on the nature of the task, but looking at individual scores was what truly gave support to that claim. These results also supported the findings previously described in section 1.1.2, going against the critical age hypothesis as it was observed that even late bilinguals – at least some of them – were able to acquire native-like proficiency with regard to the perception of this particular non-native vocalic contrast.

In yet another case, Kim et al. (2018) investigated individual differences in a longitudinal study of perceptual cue weighting of two English vowel contrasts: /i/-/ɪ/ and /æ/-/e/ by Korean children and adults during their first year of living in Canada. At group level, they found that the learners relied on duration cues more at the onset of their study and on spectral cues later on. They also observed that the relevant acoustic cues were learnt faster in the case of the /i/-/ɪ/ contrast. However, once they looked at individual learners' performance, they noticed that some participants relied on spectral cues more even at T1. They stressed that "[these differences] are not merely random variability in the learner's response patterns, but are systematically associated with the developmental trajectories of individual learners" (Kim et al. 2018: 17).

1.2.3. Phonetic variability in native and non-native speech

On top of differences across L2 learners, some authors have suggested that, "nonnative speakers also tend to be more variable in their pronunciation than native speakers (Wade et al. 2007), such that their realizations of particular segments often vary from moment to moment" (Witteman et al. 2014: 515). In other words, individual speakers might be observed to produce their L2 sounds with a lot of variation with respect to the L2 targets even within one task or one stretch of speech. As noted by Vaughn et al. (2019) while such claims have been made, there is a relative paucity of actual phonetic studies investigating the question of non-native speech variability. Xie and Jaeger (2020) theorise that this dearth of research on intra-talker variability might stem from the fact that such studies require a lot of data from the participants, which in turn is time-consuming as far as phonetic annotation of these data goes. Therefore, studying non-native speech variability is quite a difficult task in itself.

Vaughn et al. (2019) replicated the methodology of two most prominent studies – by Wade et al. 2007 and Jongman and Wade 2007 – which put forward the most explicit claims about non-native speech being more variable than native productions. They studied the realisations of stops and vowels of English and Mandarin learners of Japanese and compared them with L1 Japanese speakers. The authors found that native Japanese speakers employ a wider array of lenited realisations of stop consonants than non-native speakers. There was some L1 interference in the productions of the VOT of voiced and voiceless stops in Japanese in the productions of English and Mandarin speakers. Mandarin learners were shown to be more variable in their voiceless Japanese stops' realisation, while Japanese speakers showcased more variability in the voiced category. As far as vowels are concerned, the spectral variability of both L2 groups fell within the native norms. In discussing their results, Vaughn et al. notice that the amount of variability in non-native speech might hinge upon the language pair. Wade et al. (2007) investigated Spanish learners of English (fewer vowels → more vowels) while Vaughn et al. English/Mandarin learners of Japanese (more vowels → fewer vowels). It might be the case that variability is more present when the learners need to map fewer vowels onto a larger vocalic inventory. Variability may also stem from exposure – on the whole, Japanese speakers might flap to a lesser extent than native American speakers simply due to the fact that they are less exposed to the American accent of English. Xie and Jaeger in their study of word-final stop and vowel realisation of English in the speech L1 English and L1 Mandarin learners of L2 English yielded similar results, namely there was no evidence of the general tendency for non-native speech to be more variable. Therefore, while intuitively we might assume that L2 speakers will present less stability in their L2 speech categories, the actual phonetic evidence in favour of this claim is rather inconclusive.

1.2.4. L2 learning in a classroom context

Recall that in section 1.1.3 it was mentioned that the authors of PAM-L2 were not interested in the second language acquisition outside of immersion environment. Best and Tyler (2007) pointed out that in a classroom setting, the target language might be scarcely used and that the instructors are very likely to speak the target language with a foreign accent, which might skew the learners' perception of some contrasts. Therefore, under this view, instructed second language learning does not present itself as an area worth of investigation nor worthy of the application of PAM-L2 predictions.

On the other hand, studies which aim at investigating both formal and naturalistic second language learning note that learners undergoing formal instruction may reach higher degree of proficiency within a shorter time frame and ultimately reach comparable levels of fluency (Ellis 2008). Rahimpour and Salimi (2010) in their study of grammaticality judgements further argue that "formal instruction of the language (…) result[s] in improvement in L2 learner accuracy and provide[s] a favourable condition for L2 acquisition" (2010:1745), with Norris and Ortega (2005) providing evidence that learners receiving explicit instruction might outperform those receiving implicit types.

In particular, it is intriguing to investigate the issue of the role of formal phonetic instruction and its effects on the learners' performance. There have been a number of studies conducted whose main aim was to determine to what degree phonetic training improves the learners' L2 pronunciation (e.g. Champagne-Muzar et al. 1993; Derwing et al. 1997; Couper 2006). While in general the effects of phonetic instruction seem to be positive, the studies stress the methodological diversity of instruction. Production training in particular, especially when it is supplemented with regular language classes, has been shown to be very beneficial as far as the improvement of students' pronunciation is concerned (e.g. Mildner and Tomić 2007). In turn, Balas (2018a, b) showed that Polish students, who had been subject to formal English education in Polish schools, present good perception of English vowel contrasts at the very onset of their University education in English. Subsequent formal pronunciation training boosts their production skills considerably but does not really improve the perception thereof any further.

These considerations are important for the purposes of the present thesis in two ways. First of all, when discussing the phonetic studies concerned with phonetic drift we will distinguish between immersive and non-immersive (i.e. L1-dominant) environments and whether they have any bearing on the magnitude of the drift effects found in the speakers' productions. Second of all, it will be shown that – according to the findings of the results of the studies conducted within this PhD project – formal phonetic instruction appears to magnify the drift effects. We will return to this point in due time.

Chapter 2: Phonetic data on L1 drift

2.1. The aim of the chapter

Chapter 1 discussed the theories of L2 speech acquisition and explained the choice of the SLM as the main model chosen for the purposes of the present dissertation. This chapter, in turn, reports on phonetic studies which have investigated the effects of L2-induced phonetic drift in L1, which is the main focus of the thesis. First, it provides an overview of terms which are used in the literature to describe the influence that L2 can have on L1, explaining the differences between them and justifying the choice of the term "phonetic drift" as the leading concept describing the phenomenon that this thesis aims to investigate. Then, it presents a number of studies concerned with both consonantal and vocalic drift, distinguishing between L2-immersion and L1-dominant contexts, before summarising which features appear to be most susceptible to it.

2.2. Bidirectional effects and phonetic drift: problems with terminology

A crucial point highlighted by the SLM, which is to a large extent also the starting point of the present thesis, is the fact that not only does L1 affect L2 but that the effects are bidirectional. As a result, L2 may also exert influence on L1 productions. While the claim is relatively straight-forward, there are some problems with the terminology associated with those effects.

Note that the title of the present work uses *phonetic drift* as the term that best describes L2's effects on L1 which are of interest from the perspective of the research question undertaken herein. Nonetheless, many other terms have been used: *L1 attrition*, *reverse transfer*, *backward transfer* or simply *cross-linguistic influence*, to name a few. When one analyses what each of those terms means and entails, one may note that sometimes – even though the terms are often used interchangeably – they actually look at slightly different phenomena or lack precision in pinpointing what they indicate.

Let us start with defining the term employed to describe the research presented in this thesis. *Phonetic drift*, even though mentioned as early as 1921 by Edward Sapir, has been brought into the forefront of attention by Charles B. Chang (2010, 2012, 2013, 2018, 2019). His definition states that phonetic drift refers to the changes in L1 phonetics which are "short term (…) , both in early- and late-onset bilinguals, which are attributable to *recent* L2 experience (for example, because the change or divergence from L1 norms coincides with concomitant L2 exposure)" (Chang 2019: 3). Notice that by referring to those changes as *phonetic* drift, we stress the fact that they are somewhat independent of the changes resulting from cross-linguistic interaction at other linguistic levels, such as lexical, syntactic, or semantic (Chang 2012: 249). Another important aspect of phonetic drift is that while L2-induced changes might still persist once the L2 instruction ceases but the learner stays within L2 environment (Chang 2019), they may be reversible once the learner becomes immersed in L1 environment, without robust L2 input (cf. Stolberg and Münch 2010). Furthermore, they are claimed to result from "recent L2 experience", which stresses the fact that the contact with L2 is being maintained but also that it targets both elementary and more proficient bilinguals.

Now let us turn to the competing terms in order to relate them to the aforementioned phonetic drift and highlight the similarities, as well as the differences, so as to defend the stance that this term, and not the other ones, is best suited for the purposes of the present dissertation.

Beginning with *attrition*, one of the most widely-accepted definitions thereof has been provided by Köpke and Schmid (2004: 5) which states that it is "the nonpathological decrease in proficiency in a language that had previously been acquired by an individual, i.e. intragenerational loss". This implies that the process of attrition takes place over an individual's lifetime. Since attrition as such can entail deterioration of proficiency at various linguistic levels, and those levels might be affected by attrition to different degrees (Köpke and Schmid 2004), let us focus on *phonetic* and *phonological attrition*, as we are mainly concerned with those two branches of linguistics. Phonetic attrition, first explicitly studied by Major (1992), is defined in terms of "long-term changes in the native pronunciation of an individual as a result of the acquisition of a new dialect or language acquired post-adolescence" (de Leeuw 2018: 3). Therefore, it is not limited to L2 influence on one's L1, but also encompasses the influence of a new dialect of the same language. Indeed, for example American English expatriates living in the United Kingdom showed slightly lower numbers of flapped realisations of intervocalic /t/, claimed to be the standard positional allophone of /t/ in this context for American English (Shockey 1984). However, while de Leeuw cites this study as an example of an investigation into phonetic attrition, Shockey herself uses the term of *accommodation* in the very title of her article. Indeed, accommodation has been attested to occur amongst speakers of the same language (often referred to as *phonetic convergence*) wherein a speaker converges with the phonetic norms of their interlocutor (cf. Communication Accommodation Theory (CAT): Giles et al. 1991, Shepard et al. 2001). Accommodation is most common among the speakers of the same language and can target various acoustic parameters, such as vowel quality and quantity (e.g. Babel 2012), VOT duration (e.g. Rojczyk 2012), or pitch (e.g. Babel and Bulatov 2012).

In turn, phonological attrition is concerned with "representational changes to the phonemic organisation of the native language" (de Leeuw 2018: 5). Therefore, the crucial difference between phonetic and phonological attrition is, simply put, the difference between changes in performance vs. competence in classical Chomskyan terms (cf. Chomsky 1965). Phonological attrition entails changes in rule applications or any other structural changes at the phonemic level.

As can be seen, teasing apart phonetic attrition from phonetic drift might at first glance appear to be problematic. However, some differences between the two concepts are rather striking. In particular, the attention should be drawn to the following aspects. First of all, phonetic attrition describes the changes in L1 production under the influence of a new language or a new dialect. Therefore, it does not truly distinguish between L2-induced effects and the effects of accommodation. Chang, in turn, is very clear that phonetic drift does not result from accommodation:

> [N]either the strong sociolinguistic motivation to accommodate nor the lexical overlap supporting within-language convergence are present in the case of between-language convergence. While there are clear social reasons why L1 talkers might come to speak more or less like other members of the same speech community (...) it is difficult to imagine similar social motivations for L2 learners to modify their L1 representations with respect to an L2. (...) Moreover, in the case of L1 speech accommodation there are several ways in which L1 tokens may be connected to each other that are not available in the case of L1–L2 phonological interaction in most models of bilingual speech processing and production (Chang 2012: 250).

In other words, while it is to be expected for Shockey's (1984) American speakers to accommodate to the British residents living in their shared community, it is difficult to imagine a Polish learner of English to accommodate their *Polish* productions under the influence of English, as it lacks certain social and sociolinguistic motivations. Conversely, phonetic convergence of L2 learners' productions to the L2 norms has been attested: Zając (2015) found that Polish speakers of English accommodate their L2 /p, t, k/ (but not /b, d, g/) productions to the English norms while Tobin et al. (2017) found accommodation effects in the productions of Spanish-English bilinguals – the speakers converged phonetically to the VOT norms of the ambient language in their L2 (i.e. their /p, t, k/ productions were more English-like or more Spanish-like, depending on their immersion context), but their Spanish VOT was not affected. These findings provide support to Chang's argument about the lack of motivations for speakers to accommodate their L1 to the norms of L2.

Another diverging characteristic is that phonetic attrition corresponds to deterioration of speakers' L1 productions, thus it bears negative connotations. It implies that the speaker has lost fluency in their production or acquired foreign accent in their native language. De Leeuw (2018) provides anecdotal examples of immigrants visiting their homeland after years spent abroad with detectable L2 influences, sometimes unable to accurately produce certain L1 sounds. Recall that Flege himself noted that he first became interested in L2 influence on L1 when he heard his friends' "funny" English productions after they had spent years in West Africa (Flege 2005; cf. Section 1.1.6.2). On the contrary, phonetic drift refers to changes, which do not point to some regression of L1 production, but can be attributed to recent L2 experience (Chang 2012: 266). Chang, again anecdotally, notes that the productions of his participants did not present noticeable accentedness, nor did the speakers experience any decline in fluency (2012: 266). Therefore, phonetic drift lacks the negative implications that phonetic attrition entails. Nonetheless, it has been noted that phonetic drift may result in phonetic attrition at some point (Chang 2019); when the speaker immerses themself in L2 environment and spends years without L1 input, or is a very proficient speaker of L2, spending the majority of time immersed in their second language, over time the changes may become more permanent and not easily reversible. Therefore, phonetic drift may be the starting point leading towards attrition (both phonetic and phonological) of one's L1. We will return to this in section 8.3.

Now let us briefly consider the remaining terms enumerated above. The problem with the term *transfer* (be it positive, negative, backward or reverse) lies in the fact that it is inherently segment-oriented, at least as far as L2 pronunciation is concerned. It assumes that one phoneme is transferred from one language into another, which to a number of phonological theories which dismiss the concept of a phoneme is problematic (cf. Section 3.4.1). As early as in 1953 Weinreich noted that phonemes as such were incommensurable.

In turn, *cross-linguistic influence* is more of an umbrella term; similar to transfer, it implies that the changes occur at various levels (lexical, semantic, syntactic, etc.), which is not the case with phonetic drift, and as such is much broader and much less precise.

The reason why we need to take into account the differences between those terms (especially the divergences between phonetic drift and phonetic attrition) is that sometimes phonetic experiments use the wrong term when describing a study that the authors were conducting. Therefore, it is important for us to be able to differentiate between those concepts so as to avoid introducing confusion when discussing previous studies concerned with phonetic drift. Hence, the following sections will provide a selection of experiments dealing with phonetic drift in Chang's understanding thereof; phonetic attrition, as such, will not be considered[2].

2.3. L2-induced phonetic drift in L1

Phonetic drift in L1 is a phenomenon that has been studied in a number of languages and in different conditions. In this section, we will review the relevant phonetic studies, looking at both changes influencing consonants and vowels. The discussion will span studies carried out in immersion environments as well as those which were conducted in L1-dominant settings. The methodological choices made by various authors will be discussed only briefly in the following sections[3], as their main focus is placed on the results obtained by the various authors.

---

[2] We will return to some studies regarding attrition in section 8.3, when we focus on phonological implications thereof.
[3] Those will be discussed and assessed in more detail in section 5.8.

2.3.1. Consonantal drift

As far as the studies concerned with the effects of phonetic drift on the L1 consonants are concerned, we can distinguish two major conditions: L2-immersion and L1-dominant environments. In the former, the common denominator amongst the studies conducted is the fact that the speakers are immersed in their foreign language, which means that they are living abroad. As a result, the L1 input – albeit still present – is to a certain extent limited. In the latter, the speakers are living in their homelands, as a result of which, they are receiving more L1 input than in the case of the speakers immersed in L2 contexts. The studies reviewed in this section will follow this division and discuss experiments conducted in either condition separately.

There have been numerous phonetic studies which attempted to investigate phonetic drift effects in L2-immersion environments. In a longitudinal study on English-Korean[4] bilinguals, Chang (2012) investigated the productions of 19 speakers of American English participating in a language course in Korea who were novice learners of this language. Their English input was limited to the conversations amongst themselves outside of language classes. They were tested on both their English and Korean stop productions in a word-list reading task, with the instructions given in English. The results showed that the effects of learning Korean on English lenis productions were limited to the parameter of f_0 (as a cue to laryngeal contrast, measured at the onset of the vowel following the plosive) in the speech of female participants. The most salient changes occurred between Week 1 and 2 and Week 3 and 4. Decidedly more drift effects, both with respect to VOT and f_0, was observed in the voiceless series, in particular between Week 1 and Week 2 and Week 4 and Week 5. The VOT durations of /p, t, k/ lengthened to approximate Korean norms[5]; similarly, the f_0 of English voiceless stops drifted upwards to approach Korean norms.

[4] It is necessary to stress that English has a two-way laryngeal system (voiceless aspirated vs. unvoiced lenis stops), whereas Korean distinguishes three series of stops (plain – oftentimes referred to as 'lax' or 'lenis', with low f_0; tense – also referred to as 'fortis' or 'hard', with high f_0; and aspirated stops, with the highest f_0; Cho et al. 2002). Therefore, aside from studying VOT, Chang included f_0 at the onset of the vowel (with fundamental frequency increasing going from lenis to plain to aspirated), as depending solely on VOT would have been insufficient (Chang 2012: 253).

[5] Chang reports that the VOT norms of female speakers for Korean aspirated stops are ca. 90ms, compared to English apirates which yield ca. 62ms (Chang 2012: 253).

In another study, Chang (2013) compared the results obtained from the 19 inexperienced American learners of Korean (cf. Chang 2012) with 11 experienced American learners of Korean who completed the same language programme in South Korea as the other group. The main objective of this study was to see whether the degree of phonetic drift would be comparable or whether linguistic experience played a role in how much L1 was affected by learning L2. His hypothesis was that the so-called "novelty effect", operationalised in terms of salient changes evoked by novel – rather than familiar – stimuli, would induce more drift; in turn, although more advanced learners had more cumulative experience with L2 Korean, through formal study, extended stays in Korea, or their heritage, the lack of the novelty effect entailed less robust effects of drift. The learning context, seeing as the participants were enrolled in the same language course as in Chang (2012), will not be repeated here. The procedures and materials were also the same. As far as the results are concerned, over the course of six weeks, no significant changes were found for English voiced stops with respect to VOT duration. For the voiceless series, English VOT was lengthened towards Korean monolingual norms for both learner groups, but the effects were much more noticeable for the inexperienced learners (the difference of ca. 15ms). The effects on f_0 were also less robust for the experienced learners. Chang explains his findings with relation to the purported novelty effect in cross-linguistic interaction:

> "[G]roup differences found in this study suggest that phonetic drift following from recent L2 experience is reduced as a consequence of familiarization with the L2. To be specific, repeated exposure to an L2 over the course of learning and re-learning habituates an L2 learner to the phonetic properties of the L2, such that further L2 experience becomes less auditorily novel and, therefore, less perceptually salient; because less salient experience is encoded less robustly, this then leads to less cross-over influence of the recent L2 experience on the L1" (2013: 529).

Two things are worth noting based on Chang's research. First of all, Chang found less pronounced effects on L1 in the productions of more advanced bilinguals, which goes against Flege (1987), who found more equivalence classification effects in the productions of more advanced bilinguals, immersed in L2-dominant communities. Of course, when taking into account the discussion regarding terminology (cf. Section 2.2), we must stress that the specificity of Flege's experiments leads us to assume that, in fact, he studied the effects of phonetic attrition, rather than drift. Secondly, while there may be less drift in the speech of advanced learners, it is not clear whether pronunciation training might influence the degree to which L1 productions change. In other words, language programmes often encapsulate various tasks: reading comprehension, vocabulary learning, speaking, and writing. Undergoing intensive phonetic instruction and training the students to perceive and produce L2 sounds accurately might be novel to them enough to exert influence on L1 productions. We will return to this issue in the subsequent chapters, as the present thesis deals with this particular population, which is one of the differences between Chang's studies and the ones reported in this dissertation (cf. Chapter 5 and following).

Not all authors have observed L1 drift in interactions between Korean and English. Kim (2019) looked at the productions of 13 Koreans who came to live in the United States after puberty. The production study revealed no apparent effects of L2 English on the VOT of the speakers' L1. No correlation with exposure nor proficiency was observed. Overall, the drift effects were described as "marginal" and "linguistically unimportant"[6].

Turning to the interactions between English and Japanese, Harada (2003) studied stop productions of three distinct groups: Japanese monolinguals living in Japan, English monolinguals living in the United States, and Japanese-English bilinguals living in the US, but speaking exclusively Japanese at homes. In a picture-naming task, it was found that Japanese-English bilinguals differed from Japanese monolinguals in their Japanese productions, with the VOT for the bilinguals being longer, relative to monolingual norms. The bilingual group's L2 productions did not differ from English monolingual control group's realisations. One of the limitations of the study is that the bilingual group comprised both heritage speakers as well as Japanese immigrants who had spent some time in Japanese schools. The length of residence in the US was not provided.

In yet another study, Sučková (2018) investigated the English stop realisations of British and American expatriates who had been living in Czechia (mean length of residence: 9.8 years) but still extensively used their L1 as English teachers, at their homes, and in international corporations. She found significantly higher percentage of prevoiced, Czech-like realisations in the expatriates' productions (ca. 55% of all productions) in comparison to English monolingual controls. The voiceless series appeared to be more maintained, as no effects of Czech on /p, t, k/ English productions were observed.

[6] Once again, however, as observed by Chang (2012), since Korean is a three-way laryngeal system whereas in English the contrast is only between two series of stops, taking into consideration VOT only (as opposed to, e.g. f_0) might be insufficient to observe any effects.

Similar results were also obtained by Dokovova (2015), who studied 10 Bulgarian expatriates living in Scotland (mean length of residence: 9.4 years) and compared their productions with a functionally monolingual control group of 10 Bulgarians living in Bulgaria. While no effects of English as the ambient language was found on the voiceless series, pre-voicing was much shorter for the expatriates group. Moreover, she found some effects of age, as the drift effects were much more robust in the speech of older speakers (mean age: 58 years old) in her population sample.

Phonetic drift in L1 has also been observed in case studies looking at individual bilingual speakers. Podlipský et al. (2020) studied two sisters who were English-Czech simultaneous bilinguals upon their arrival in Czechia and towards the end of their three-week stay there. More effects of recent exposure to Czech were found for the pre-voiced series, with more instances of pre-voiced realisations of /b, d, ɡ/ in both languages after the immersion period. Furthermore, the pre-voiced series was found to be less consistent – there was a lot of variability in the speakers' productions, with some of the words produced with positive VOT values and some with pre-voicing. At both testing times, the VOT values of English and Czech voiceless stops were kept separate (with English values longer than Czech values). A similar case study was conducted by Sancier and Fowler (1997) who studied a Brazilian Portuguese-English bilingual female who travelled back and forth between Brazil and the United States. The VOT values of her L1 productions drifted towards the norms of her ambient language (i.e. the /p, t, k/ values were longer when she stayed in the US and reversed back to Brazilian Portuguese monolingual norms when she went back to Brazil). This difference was perceptible by L1 speakers of Brazilian Portuguese norms, who noted that she produced her stop consonants with additional puffing noise after coming back from the US. This example is particularly interesting when one recalls the discussion on drift vs. attrition – here, although the foreign accent in her L1 productions was observable by others (which is an attribute associated with phonetic attrition), it was the result of recent exposure to her L2 and not a permanent effect (a feature associated with phonetic drift).

A relatively large number of studies set in an L1-dominant environment have also been conducted, once again mostly investigating the behaviour of stop consonants. Herd et al. (2015) studied Southern American English speakers learning L2 Spanish in the United States. It was a cross-sectional experiment investigating speakers at various levels of proficiency: beginners, intermediate students, advanced learners, as well as graduates who completed their MA degree in Spanish. On the whole, they observed no Group effects on the English /p, t, k/ productions, while significant phonetic drift effects on pre-voicing values of /b, d, g/ were found. Proficiency played a big role in the results: Spanish graduates had more pre-voicing than advanced learners, who in turn pre-voiced more tokens than intermediate and elementary students, respectively. However, a lot of individual variation was also observed. Some speakers produced only unvoiced items, some only pre-voiced items, and some a combination of both. Therefore, students seemed to have little control over the phonetic realisation of voiced consonants. The authors explicitly note that "it is notable that such a large number of stops were pre-voiced, with even the beginning learners of Spanish pre-voicing 44%" (2015: 7), however they also point out that, on the whole, pre-voicing is attested as a feature of Southern American English (e.g. Morris and Hunnicut 2016). Therefore, it is unclear whether the observation of Spanish influencing English at very early stages is not – in this particular case – an instance of a Type 1 error[7]. The readers also are not informed whether or not all participants underwent similar Spanish instruction and how much focus was placed on pronunciation, as well as whether or not the participants had classes with native speakers of Spanish. Nonetheless, worth noting is the fact that the voiced series of stops was more susceptible to drift than the voiceless one.

In turn, Schuhmann and Huffman (2015), who also studied the speech of L1 American English learners attending a Spanish course, reported on a longitudinal study which spanned a few months and in which explicit phonetic instruction was present. Their results also point to much variation among the learners. Overall, they found no effects on English voiced stops (while in Spanish productions there was a big increase in the numbers of pre-voiced tokens which almost approached significance between Weeks 6 and 10), but they noticed a slight decrease in VOT positive values for /p, t, k/. For the whole group, the lowering almost approached significance but the estimated differences was about 4ms, which is below the Just Noticeable Difference threshold. For three speakers this lowering was significant, but the VOT values still fell within the English range. In general, although they observed some effects of drift, their results were very preliminary, as they studied the productions of only five speakers. Additionally, they recorded English

[7] We will return to this issue in Section 5.8.

productions immediately after Spanish ones, which might have led to some language mixing effects (Grosjean 2004)[8]. Nonetheless, a more thorough investigation of this particular group of speakers could have been beneficial, as the results on which the authors reported were exactly opposite to what Herd et al. (2015) found. The authors conclude that the degree of a difference between L1 phones and L2 phones is not sufficient for predicting automatic drift effects.

Yusa et al. (2010) investigated the voiceless stop productions of Japanese children (N=107) at their early stages of learning English (i.e. at the age of 4, when children are first exposed to English classes at pre-school). The participants were subdivided into three distinct groups on the basis of the amount of English instruction they were receiving at the time. The picture naming task conducted after 20 months of English classes revealed that the two groups for whom English lessons were most frequent not only kept the VOT values for Japanese and English stops separate but also produced Japanese stops with VOT values significantly shorter relative to monolingual norms. The drift effects were thus dissimilatory and the speakers' main objective was ensuring the phonetic contrast between Japanese and English voiceless stops.

In another study, Osborne (2016) studied Brazilian Portuguese learners of English (N=36) living in Brazil and compared their /p, b, k, g/ productions with the productions of Portuguese monolingual speakers (N=36). She found dissimilatory drift effects in the voiced series of stops – both /b/ and /g/ were produced with longer negative VOT values relative to monolingual norms (the difference for /b/ was 13ms and 20ms for /g/). While no drift in either direction was observed for /p/, there was some assimilatory drift found for /k/, with the VOT values longer for bilingual speakers – hence, more English-like – than for the monolinguals. The degree of drift in /k/, however, was relatively small (ca. 7ms). Nonetheless, the difference in the behaviour of /p/ vs. /k/ is interesting. We know that in general VOT values for bilabials are much shorter than for velars (Lisker and Abramson 1964). We also have seen that bilinguals sometimes produce stops in their L2 with VOT values intermediate between their L1 and L2 (e.g. Flege 1995). Osborne noticed that while /p/ in her participants' L1 and L2 productions was kept separate, no difference was found in their L1 and L2 /k/ productions. Since /k/ VOT values in true-voice languages are on the whole longer, it is my understanding that perhaps some perceptual

[8] Additional problems with their methodology will be discussed in Section 5.8.

linkage between the slightly aspirated /k/ in L2 English and Brazilian Portuguese unaspirated /k/ did occur. Hence, the assimilatory drift towards the English norms for the velar stop but not for the bilabial one, for which the VOT difference is decidedly more robust.

Polish has received moderate amount of attention. Waniek-Klimczak (2011) administered a small-scale phonetic study to compare the realisation of voiceless plosives /p/ and /k/ after high and low vowels by Polish proficient learners of English and quasi-monolingual native speakers of Polish. She observed lengthening of VOT in both bilabial and velar context by the former group, however, due to the limitations of the study (a very small number of tokens and participants; a limited dataset which did not include the voiceless series; speech style – careful, emphatic speech elicited in the experiment) no conclusions could be drawn. Two other studies by Wrembel (2011)[9] and Sypiańska (2017)[10] also dealt with the voiceless series only. To the best of my knowledge, only Schwartz (2020) investigated both series of stops. He studied 18 native Polish speakers enrolled in an English-Russian programme at a Polish university and compared their productions with a group of Polish quasi-monolinguals. The students have undergone an intensive English phonetics course (both theoretical and practical) before their participation in his study. He observed more effects of drift in the voiced series of stops – both with respect to pre-voicing duration (with the values being shorter for the student group) as well as with the number of unvoiced, English-like productions (higher numbers of such tokens in the students' productions). As far as the voiceless series goes, no significant differences between the students' and monolinguals' productions were observed.

Table 2 summarises the studies discussed in this section, highlighting methodology differences and the obtained results.

[9] This study is not described in more detail as Wrembel (2011) investigated the effects of L2 English *and* L3 French on L1 Polish; as she was interested in the effects of multilingualism rather than bilingualism, the study falls outside of the scope of the present thesis.

[10] Sypiańska's (2017) study is not elaborated on here as it focused on L1 attrition and the effects of multilingualism (L2 English, L3 Danish) rather than bilingualism. Similarly to Wrembel, it falls outside of the scope of the present thesis.

Table 2. A summary of phonetic drift studies concerned with consonants.

Settings	L1	L2	Method	Results	References
L2-immersion	English	Korean	word lists	more drift observed for English voiceless plosives, for both VOT and f_0	Chang (2012)
	English	Korean	word lists	replicated the results of Chang 2012; more drift in the productions of inexperienced (rather than experienced) learners	Chang (2013)
	Korean	English	sentence list	little effects of L2 on L1; “marginal and unimportant”	Kim (2019)
	English	Czech	word lists	voiceless stops in both lgs kept separate; more pre-voicing after recent exposure to Czech in /bdg/; the voiced series displaying more variability	Podlipský et al. (2020)
	Japanese	English	picture-naming	L1 /ptk/ productions realised with longer, English-like values	Harada (2003)
	English	Czech	word lists	more pre-voicing in L1 English productions; no shortening of /ptk/ found	Sučková (2018)
	Bulgarian	English	word list	Drift effects in /bdg/ only; drift more robust in the speech of older speakers	Dokovova (2015)
L1-dominant	English	Spanish	word lists	effects of Group for /bdg/, but not /ptk/; effects of proficiency;	Herd et al. (2015)
	English	Spanish	word lists	slight decrease in /ptk/ VOT duration (below 4ms); some pre-voicing found in /bdg/ (almost reaching significance)	Schuhmann and Huffman (2015)
	Japanese	English	picture naming	Japanese stops produced with VOT values shorter than monolingual norms; dissimilatory drift in the productions of children with highest degree of exposure to L2	Yusa et al. (2010)

	Brazilian Portu-guese	English	delayed repeti-tion	dissimilatory drift found in /bg/; some as-similatory drift in /k/ but not /p/	Osborne (2016)
	Polish	English	sentence list and emphatic productions	some lengthening of VOT in /p/ and /k/ by more proficient learn-ers	Waniek-Klimczak (2011)
	Polish	English	word lists	drift effects in the voiced series (shorter pre-voicing, more un-voiced productions) but not in the voiceless se-ries	Schwartz (2020)

2.3.2. Vocalic drift

The effects of phonetic drift are not by any means limited to consonants only. There have been a number of experiments investigating the changes in one's L1 vocalic inventory under the influence of recent exposure to L2. Similarly to the way in which the previous section was organised, the overview of studies on vowel systems will be divided into two parts, namely phonetic drift in an L2-immersion and L1-dominant environments.

Beginning with the former, Chang (2012) investigated the productions of 19 American English speakers enrolled in a language course in South Korea (i.e. novice learners of Korean). The English vowels were couched in a CVC context, with C_1 being [h] and C_2 being a coronal stop (or a velar stop in one instance). The vowels investigated were FLEECE, KIT, FACE, DRESS, TRAP, GOAT, GOOSE, FOOT, STRUT, and THOUGHT[11]. While the differences were subtle, English vowels were found to drift to approximate Korean norms, with respect to vowel height (i.e. F1) and but not advancement (i.e. F2). For a few male speakers the changes in F1 were not significant. Moreover, the drift effects were observed to be systemic, thus targeting the entire vocalic inventory rather than individual segments. In a follow-up study, where Chang (2013) compared 11 experienced and 19 inexperienced American learners of Korean, he found that the former group did

[11] Since various scholars employ various transcription symbols for English vowels (sometimes choosing an allophone that is most common in a given accent), the decision was made to use Wells' lexical set (2010) to describe English vowels throughout the present thesis. Each vowel named a representative keyword which introduces more clarity into the discussions. For clarity purposes, the complete list of Wellsian key words is available in Appendix 6 to the present dissertation.

not show any significant effects of drift with respect to vowel height, in contrast to inexperienced learners. F2 values for English vowels for both groups remained unaffected by L2 Korean.

Lang and Davidson (2017) studied 11 American English speakers with no prior exposure to French enrolled in a student exchange programme and beginning their one semester at an American university in Paris. They were compared with 11 Paris residents – American speakers who were instructors (or friends thereof) at that university. Similarly to Chang's participants, the American participants were tested after 6 weeks of French instruction and their productions were compared with pre-test data obtained from them. The stimuli consisted of CVC target words containing the following English vowels: FLEECE, KIT, DRESS, TRAP, BATH, STRUTM LOT, FOOT, GOOSE, subsequently embedded in sentences. In contrast to Chang's (2013) results, no change between the inexperienced learners of French was found between Week 1 and Week 6. The authors theorise that the contradictory results might be due to the differences in the compactness and crowdedness of the French vowel space in comparison to what we see in Korean, respectively. Furthermore, they hint at the amount of French input – the relatively small number of classes in French that they had been attending might have been insufficient for French to exert any influence on their English. Interestingly, changes in both F1 and F2 (both systemic and for individual vowels) were observed for the instructors and their friends (i.e. speakers immersed in French to a greater extent and for longer).

Some studies additionally pointed to the effects of age on the degree of drift effects. Oh et al. (2011) studied the vowel productions of 32 Japanese-English bilinguals (16 adults and 16 children) living in Texas, USA for a year and compared them with a control group of monolingual English speakers from the same area. The length of residence in the USA for the bilingual group was ca. 4 months at T1 and 1.6 years at T2. By means of a picture naming task the authors obtained acoustic data which showed that while no changes over time were detected in the adult learners' production of Japanese vowels, Japanese children showed drift effects for the vowels /i/ and /a/. They were most robust for the former, possibly due to its merger with the English FLEECE vowel. Importantly, it was the group of the children that improved most in their L2 productions, while no changes in the adults' productions of English vowels between the two testing times were found. This suggests that the broader the experience in the L2, the more salient its influence on the L1. Similar results were gathered by Baker and Trofimovich (2005) who investigated the vowels of early and late Korean-English bilinguals (N=20 and N=20 respectively). The groups were further divided into subgroups on the basis of their length of residence in the USA. While no effects of English on Korean vowels were found for older late bilinguals, young early bilinguals with extensive exposure to English (and who had been living in the USA longer, for ca. 7 years) produced their Korean vowels /i, u, ɛ/ higher in the vowel space, and with /u/ being less advanced, relative to monolingual norms.

A number of studies on L1 vocalic drift was concerned with the vowel productions of speakers living in L1-dominant environment. Recall that Chang (2013) showed that inexperienced learners tended to show more salient L2 influences in their L1 speech due to the so-called novelty effect. However, for instance, Yang and Fox (2017) studied three distinct groups: Mandarin monolingual (N=15), Mandarin-English advanced bilingual (N=8, aged: 5-6, born and raised in Ohio, USA), and novice Mandarin-English (N=7, living in Ohio, USA for ca. 6 months) bilingual children. They were tested on word lists of 10 Mandarin disyllabic words. They observed that as far as Mandarin vowels (that is: /i, y, a, u, ɤ/) are concerned, for all vowels but /u/ the distance between monolingual norms and bilingual productions was greater for the group of children who were more proficient in English. Furthermore, interestingly, the more experienced bilingual group showed greater formant movement relative to less proficient and the monolingual groups. It was observed that in particular Mandarin vowel /a/ was raised towards the norms of English BATH and TRAP, while Mandarin /i/ showed more dynamic formant movements, observable in the diphthongisation of the English FLEECE vowel[12].

Similarly, Mora and Nadeu (2012) studied the productions of Catalan-Spanish bilinguals and the extent to which extensive exposure to L2 Spanish (in particular its one vowel /e/) influences the correct realisation of the Catalan /e/-/ɛ/ contrast. The authors observed that the group of speakers with high weekly usage of Catalan produced the difference between this particular vowel pair more robustly (in minimal pairs, cognates, and

[12] Interestingly, the vowel internal spectral movement (e.g. Nearey and Assmann 1986) has been shown to be a phonological property of English vowels and the cross-linguistic differences in formant movement are treated as systemic in nature (e.g. Schwartz 2018). As such, the L2-induced VISC effects on L1 productions could be perceived as structural changes and as such be attributed to, perhaps, early stages of attrition. This would go in line with the fact that these effects were observed in more proficient children whose exposure to Mandarin was limited to the input they received from their parents at home. Even though the border between phonetic drift and attrition is blurry, this study is of importance for our phonological discussion to which we will return in Chapter 8.

noncognates alike) than the speakers with higher degree of exposure to Spanish and lower weekly usage of Catalan. The influence of Spanish /e/ was most saliently observable in the production of Catalan-Spanish cognates in the latter group. Therefore, greater experience in Spanish contributed to more noticeable changes in L1 Catalan production.

The relationship between drift effects and cognates was also inspected by Yao and Chang (2016) who looked at the contact-induced behaviour of the Shanghainese vowel /ɛ/ in the productions of 24 Shanghainese-Mandarin bilinguals. It was hypothesised that the Mandarin lexical set containing the vowel /ej/ influences the quality of L1 Shanghainese /ɛ/, in particular in cognates. In order to check this hypothesis, a sentence reading task and an audio translation task were designed. Indeed, it was found that in a set of cognates which are the Shanghainese counterparts of the Mandarin /ej/ lexical set, the vowel /ɛ/ drifted towards /e/ (or even /ej/, with traces of diphthongisation). The drift effects were especially salient in the productions of younger speakers (who are more bilingual than older speakers due to stricter policies regarding Mandarin as the official language of China in the recent decades) and in the bilingual mode (i.e. in the translation task, when Mandarin was also activated).

In yet another study, Guion (2003) compared early, mid, and late Quichua-Spanish bilinguals (N=20) and the effects of Spanish vowels on their L1 Quichua /i, a, u/. The results of a delayed repetition task revealed that the speakers who successfully acquired L2 vowels showed more noticeable drift effects in vowel height. Since it was the early bilinguals who produced their L2 Spanish vowels in a fashion that best resembled native Spanish norms, they were the group with most distinct differences in their native vowel system. The changes were global and targeted the entire vocalic system rather than individual segments.

Kartushina et al. (2016) additionally looked at whether articulatory training in L2 has any effects on L1 vowel production. In order to do so, 20 monolingual L1 French speakers were trained on two non-native vowels: Danish /ɔ/ and Russian /ɨ/. The three training sessions, including pre-test and post-test production sessions, consisted of L2 vowel repetition tasks. The participants received visual feedback during the training. The investigators were interested to see whether the Danish vowel would influence its closest French counterpart /o/, and similarly, whether Russian /ɨ/ would affect any of the French front vowels /i, y, ø/. While no global change of the French /o/ was found, in individual cases, the F1 and F2 of this vowel appeared to drift between pre- and post-test recording sessions. French /ø/ moved closer to Russian /ɨ/ after the training, while the effect was marginal for /y/ and not detected for /i/. Overall, it was shown that even brief phonetic training on novice learners triggered the occurrence of phonetic drift in L1 vowel formants.

To the best of my knowledge, Polish has not been subject to any studies related to potential vocalic drift effects[13].

Table 3 provides a summary of the phonetic studies investigating drift effects in vowels.

Table 3. A summary of phonetic drift studies concerned with vowels.

Settings	L1	L2	Method	Results	References
L2-immersion	(American) English	Korean	word list	systemic assimilatory drift towards Korean norms in vowel height	Chang (2012)
	(American) English	Korean	word list	experienced learners show less drift effects than inexperienced learners in vowel height	Chang (2013)
	Japanese	English	picture naming	age effects found: more drift found for young speakers; more progress in L2 entails more phonetic drift	Oh et al. (2011)
	Korean	English	picture naming	Korean vowels produced as higher (and /u/ as fronter) relative to monolingual norms but only in the case of young early bilinguals	Baker and Trofimovich (2005)
	(American) English	French	sentence list	lack of drift effects for inexperienced learners; drift in both F1 and F2 towards French norms in the case of more immersed speakers	Lang and Davidson (2017)
L1-dominant	Mandarin	English	word list	greater distance between monolingual norms and bilingual productions for more proficient speakers; drift observed for both	Yang and Fox (2017)

[13] But see Sypiańska (2016) on the effects of multilingualism on L1 Polish.

				static and dynamic aspects of Mandarin vowels	
	Catalan	Spanish	word lists	[e]-[ɛ] contrast maintained more successfully by speakers with less exposure to Spanish; Spanish influence most robust in cognates	Mora and Nadeu (2012)
	Shanghainese	Mandarin	sentence list and auditory translation	contact-induced drift of L1 [ɛ] in a specific set of cognates; effects of age and language mode observed	Yao and Chang (2016)
	Quichua	Spanish	delayed repetition	most successful L2 learners show most robust drift effects; changes found to be systemic	Guion (2003)
	French	Danish, Russian	word list	articulatory training on novice learners affected the L1 vowel qualities (both globally and within subjects)	Kartushina et al. (2016)

2.3.3. Perceptual drift

Drift effects in perception have not been investigated quite so extensively as in production, nonetheless there exist a few of studies that looked at possible effects of learning L2 on one's perception of L1 categories. As in the previous two section, we will discuss studies conducted in L2-immersion settings and L1-dominant environments.

Starting with L2-immersion, Tice and Woodley (2012) reported on perceptual drift occurring over the span of an intensive summer language course attended by American learners of French in Paris. The 5 speakers had no previous experience with French, therefore the authors were interested to see whether the novelty effect (cf. Chang 2013) would play a role in their perception of VOT continua. The drift in perception turned out to manifest itself within the first few weeks of the language course, with the biggest perceptual shift occurring between Week 3 and 4. When compared with the control group of American monolinguals, the boundary between voiced-voiceless stops oscillated within 35ms for the controls and 30ms for the English-French novice learners. Therefore, the perceptual boundary shifted towards French-like VOT norms quite dynamically, even though the degree of that shift was not extreme.

Osborne (2016) conducted a similar experiment, also concerned with perception of VOT continua, but in an L1-dominant environment. In a study supplementing her production experiment, she investigated the perceptual effects of drift in Brazilian Portuguese-English bilinguals. In a two-alternative forced choice task, she found that while the perception of Portuguese initial stops /p/-/b/ was not affected by learning English to a statistically significant degree, overall Brazilian Portuguese monolingual speakers tended to perceive /p/ at a lower VOT value than L2 English learners.

In turn, Dmitrieva (2019) looked at the differences in cue weighting in the perception of medial and final voicing by 37 advanced Russian-English bilinguals living in the United States and compared them with English and Russian monolingual controls. She found no difference between English monolinguals and Russian-English bilinguals in the English mode with respect to the cues to which they attended in the perception of medial and final voicing in English. On the other hand, the bilinguals' performance in the Russian mode was significantly different when compared with Russian monolinguals in that the two groups – even though Russian was the L1 for all of them – attended to different cues. In particular, for the bilingual group vowel duration was a more important cue for final voicing than glottal pulsing, while for monolinguals the situation was reversed.

In a cross-sectional study, Gorba (2019) studied the perception of VOT continua of voiceless-voiced bilabial stops by Spanish-English bilinguals at different levels of proficiency, based upon their length of residence in the L2 country. The groups included students of English who had never been to an English-speaking country, who had spent one semester/one year in an English-speaking country as part of a student exchange programme, and who had been living in an English-speaking country for ca. 4 years. The results of a forced choice identification task revealed that all groups, regardless of their experience in English, showed greater VOT boundaries in comparison to Spanish monolingual controls. Therefore, the boundary shift towards English norms was observed; nonetheless, the difference was statistically significant only for the most experienced Spanish-English group.

As far as the L1-dominant setting is concerned, as an extension of their production study, Mora and Nadeu (2012) checked the perception accuracy of Catalan-Spanish bilinguals of the vowels /e/-/ɛ/ on a 10-step continuum. While no effects of L2 Spanish on categorisation was observed, the reaction times for the speakers with extensive exposure to Spanish were significantly slower than for the speakers receiving more Catalan input. This finding suggests that the processing of the native contrast might be hindered by the influence of L2.

Finally, Cabrelli et al. (2019) examined perceptual shifts at a suprasegmental level. They explored the perception of illusory vowels[14] in Brazilian Portuguese. The authors were interested to see whether experience in L2 English, with different phonotactics, would cause perceptual shifts in the speakers' L1. They found that learners immersed in the L2 environment were able to overcome the phonotactic restrictions of Brazilian Portuguese and that the correct perception of coda stops, devoid of the inserted illusory vowel, was correlated with the level of proficiency in English. That is, the speakers with better perception of coda stops in English were able to more successfully extend it to their L1. Nonetheless, the perceptual shifts and the bilingual advantage over monolinguals was observable in all participants, regardless of whether or not their L2 proficiency was target-like.

Yet again, Polish has not been investigated with respect to possible perceptual shifts due to L2 learning[15]. Table 4 below summarises the methods and the results of the perception studies discussed in this section.

Table 4. A summary of perceptual drift studies.

Settings	L1	L2	Method	Results	References
L2-immersion	(American) English	French	phoneme categorisation	perceptual shift of VOT continua towards L2 norms	Tice and Woodley (2012)
	Russian	English	forced choice identification	bilinguals attend to different cues for voicing in their L1 than monolinguals	Dmitrieva (2019)

[14] Defined in terms of "[perception of a] vowel that is not present in the input in order for illicit input to conform to [the speakers'] L1's well-formedness restrictions" (Cabrelli et al. 2019: 56). In Brazilian Portuguese the only consonant allowed to occur in the coda position is /s/. Whenever a stop appears in the coda, an illusory vowel with /i/-like quality is inserted and as a result, the stop is shifted to the onset of the syllable. Cabrelli et al. illustrate this by means of the word "apto", produced and perceived by Brazilians as /ap/i/to/ (2019: 56).

[15] Sypiańska and Cal (2019) conducted a small-scale study related to perceptual shifts in L1 Polish under the influence of L2 English and L3 Spanish, but since their main focus was on the effects of multilingualism rather than bilingualism, the study falls outside of the scope of the present thesis.

	Spanish	English	forced choice identification	perceptual shift of VOT continua towards L2 norms	Gorba (2019)
	Brazilian Portu-guese	English	ABX discrimi-nation	perceptual shift as a re-sult of L2 phonotactics	Cabrelli et al. (2019)
L1-dominant	Catalan	Spanish	AXB speeded categorisation	RTs slower for more experienced L2 Span-ish speakers	Mora and Nadeu (2012)
	Brazilian Portu-guese	English	Two-way forced choice identification	no effects of drift on L2 English learners	Osborne (2016)

2.4. Which features are most susceptible to phonetic drift?

On the basis of the overview regarding phonetic studies exploring the effects of drift, we may observe that phonetic drift targets a number of acoustic correlates. This section summarises the findings.

In the case of consonants, in languages with two-way laryngeal systems it is VOT that is most susceptible to undergoing changes under the influence of second language acquisition.

Furthermore, in the studies wherein the language pair differed with respect to the laryngeal contrasts (e.g. Korean and English), f_0, that is pitch or fundamental frequency, at the onset of the vowel was also shown to be subject to drift effects. VOT on its own appeared to be inadequate.

As far as vowels are concerned, drift effects were observed more often in F1, therefore vowel height appeared to be more of a target, although F2 – i.e. vowel advancement – also showed some changes in a few studies. Chang (2019) notes that drift in vowels might occur in various ways and there is no one pattern to be expected. Instead, it appears that the direction of vocalic drift depends upon the acoustic differences between the vocalic inventories of one's L1 and L2.

Phonetic drift was also observed in suprasegmentals. Aside from pitch, which has been mentioned above, we have also seen the effects of L2 on coda perception (Cabrelli et al. 2019). Some studies indicate that drift has been attested in intonation contours (Mennen 2004; de Leeuw et al. 2012)[16], however it remains unclear whether drift in suprasegmentals operates in similar ways to what we see in the case of segmental features (Chang 2019).

2.5. Discussion

This chapter has provided an overview of studies concerned with L2-induced phonetic drift in L1. It has discussed a number of relevant experiments from various language pairs and enumerated the acoustic correlates which seem to drift to the greatest degree. Upon discussing both the terminological discrepancies and reviewing the relevant background literature, a few issues appear to be worth noting.

One of the major contributions of Chang's series of studies was the observation that phonetic drift occurs in a very dynamic way and very early on in the process of second language acquisition, in particular when the learners are immersed in L2-dominant communities. These findings were attributed to the so-called novelty effect. On the other hand, however, we have seen studies which showed that language experience correlates with the degree to which L1 drifts. We have also been shown that training (whether it is just pronunciation instruction or more specific articulatory training; cf. Schuhmann and Huffman 2015 and Kartushina et al. 2016, respectively) also contributes to the observed phonetic drift effects. As has been mentioned before, it is unclear whether we can hypothesise that training actually triggers the occurrence of novelty-effect. The learners are trained to perceive acoustic differences of which they could have been unaware before instruction, and subsequently L2 exerts more influence on their L1, resulting in more extensive drift.

Nonetheless, phonetic drift effects were attested also in the productions of the learners staying in L1-dominant environment, both with respect to vowels and consonants. At the same time, we also have studies where drift is absent in the speech of the learners staying in L2-dominant countries (e.g. Kim 2019; Lin and Davidson 2017). It is, then, safe to assume that the environment in which the learner is immersed is of secondary importance.

[16] These two studies have not been reviewed in this chapter as they fall out of the scope of the present thesis, since they do not deal with consonantal or vocalic drift. It is, however, worth noting that drift is not limited to segmentals only.

Another relevant observation to be made is the fact that the discussed studies appear to suggest that drift effects encompass entire natural classes of sounds (e.g. Chang 2012; Guion 2003), rather than individual segments (cf. Flege 1987). Therefore, phonetic drift is a systemic phenomenon. The direction in which the effects go, however, seems to be dependent on the acoustic differences of a given language pair.

Finally, an interesting asymmetry in the behaviour of the two series of stops may be noticed (cf. Schwartz 2020). Upon inspection of the results from studies concerned with drift in the realisation of laryngeal contrasts, we may conclude that while drift effects are detectable in both voiced and voiceless stops, decidedly more interaction between L2 and L1 occurs in the former. Chapter 3 looks at different theories concerned with laryngeal phonology that could explain the special status of the voiced series and account for the asymmetry in L2-induced phonetic drift.

Chapter 3: The phonetics and phonology of laryngeal contrasts in Polish and English

3.1. The aim of the chapter

In the revision of the phonetic studies associated with phonetic drift we have seen that the parameters targeted in the case of consonantal drift pertain primarily to the acoustic correlates of voicing: VOT (both positive and negative) as well as fundamental frequency. The asymmetries that we observe in the behaviour of stop consonants in various languages lead us to the question as to how the current phonological theories can account for the phonetic data we have. In particular, we are interested in the languages with two-way laryngeal contrasts.

The term Voice Onset Time (VOT; Lisker and Abramson 1964) as an acoustic parameter of interest has been appearing relatively often in the previous chapters. It remains one of the fundamental characteristics of stop consonants scrutinised in cross-linguistic investigations as it is relatively easy to measure and languages of the world reveal striking differences in how they employ it. Languages which contrast two series of stops are generally divided into two types: *aspiration* and *true-voice* languages (Iverson and Salmons 1995). The former type, which encompasses Germanic languages such as English or German, distinguishes between unvoiced (or weakly/passively voiced) and voiceless aspirated stops. The latter type, which includes Romance and Slavic languages such as French or Polish, contrasts truly voiced series (usually characterised by negative VOT, also referred to as pre-voicing or voicing lead, triggered by vocal fold vibration during closure) with voiceless unaspirated series. A question, then, arises. Namely, to what extent should these phonetic facts be incorporated into phonological representations? Should the VOT typology, which divides languages into those two categories, have any bearing on laryngeal phonology?

The rest of the chapter provides an overview of three different phonological theories: feature theory, laryngeal realism, and laryngeal relativism and the ways in which they represent laryngeal contrasts in aspiration and voicing languages. Then some shortcomings of those theories are discussed, before an alternative framework – Onset Prominence (Schwartz 2016 *et seq.*) – is presented. The chapter ends with a formulation of predictions that all those theories make for the behaviour of L1 stop consonants of Polish learners of English in the process of L2 acquisition. Those predictions will be tested in the empirical part of the present dissertation.

3.2. The phonetics of Polish and English laryngeal systems

While Polish and English have already been mentioned numerous times in the previous chapters and some data on those two languages have been provided, the following sections aim at recapitulating the most relevant information on their laryngeal contrasts and vowel inventories by comparing and contrasting them. Additionally, they provide overviews of studies which investigated the acquisition of English laryngeal and vocalic contrasts by Polish learners. This is of interest insofar as it provides background on L2 acquisition of the language in question (that is, English) by the population studied in the present thesis (that is, Polish-English bilinguals).

If you recall our considerations Section 3.1, languages which contrast two series of stops are commonly subdivided into the so-called “aspiration” and “true-voice” languages (Iverson and Salmons 1995)[17]. English falls into the former category, contrasting voiceless aspirates with unvoiced lenis stops, while Polish is an example of the latter, having plain voiceless stops and pre-voiced ones in its inventory.

For illustration purposes, Fig. 3 shows the waveforms of four plosive-initial words, two from each series, in Polish and English respectively.

[17] As this thesis focuses on word-initial contrasts, the laryngeal reality of Polish and English in other contexts (word-medial voice assimilations, final obstruent devoicing, stop release, etc.) will not be discussed.

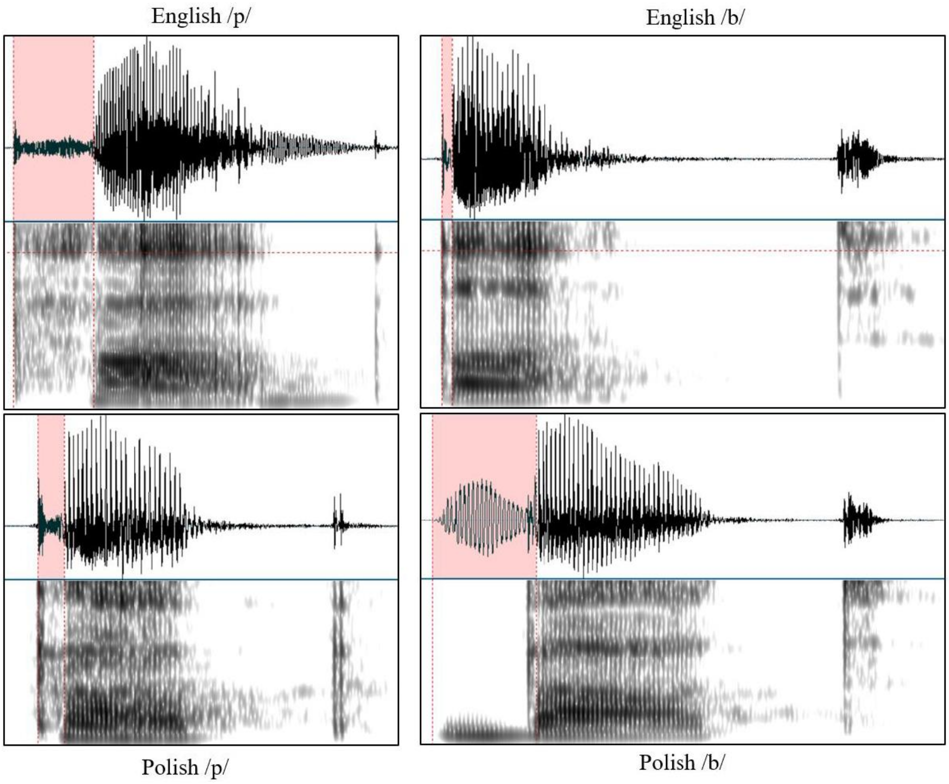


Fig. 3. VOT patterns in word-initial plosives in Polish and English.

The VOT of the initial plosive in the word *pub* [pʰʌb̥] (top left) is 113ms, compared to 28ms in the word *pub* [pap] 'a pub' (bottom left) produced by a Polish speaker. In the voiced series, on the other hand, we can observe positive VOT of 10ms in the English word *but* [b̥ʌ̆t] (top right), while the Polish word *bat* 'a whip' [bat] (bottom right) displays negative VOT values (i.e. voicing-lead) of -113ms, followed by a burst and some accompanying noise, which indicates the release of the closure.

Keating (1984) observes that the way in which plain voiceless [p, t, k] differ from unvoiced lenis [b̥, d̥, g̊] remains unclear[18]. Note, however, that the VOT of [p] in Polish *pub* is in fact slightly longer than the positive VOT in English *but*, even though both these words begin with a bilabial plosives and are followed by a non-high vowel. At the same time, some studies, report that while the difference in VOT duration between plain voiceless stops in Polish and unvoiced lenis stops in English is relatively salient indeed, it still entails a fair amount of confusion in discrimination studies (e.g. Kopczyński 1977). While VOT might prove to be insufficient a parameter to discriminate between the two, Schwartz and Arndt (2018) found that while Polish speakers display the highest accuracy

[18] Recall that Laryngeal Realism assumes they are unmarked and, thus, phonologically identical.

when asked to distinguish between pre-voiced /b, d, g/ and voiceless /p, t, k/ (96.8%), they also do well with regards to differentiating between pre-voiced and unvoiced /b, d, g/ tokens (75.3%). They also seem to attend to other phonetic cues in order to maintain laryngeal contrast, in particular to F1 at the onset of the vowel (Schwartz and Wojtkowiak 2018). Nonetheless, on the whole Polish speakers try to avoid confusing plain voiceless and unvoiced stops as in her study of initial laryngeal voicing in Polish, Keating (1979) noted that there were overall remarkably few unvoiced realisation present in the productions of Polish monolingual speakers (see also Schwartz 2020). On the other hand, pre-voicing has been attested for some accents of English (in particular Southern American English; e.g. Morris and Hunnicut 2016) and it appears to be present also in hyperspeech.

To provide more data on the VOT measurements from the two languages, Table 5 summarises the results of a few phonetic studies on Polish and English. First let us look at the Polish values. Keating (1981) analysed Polish VOT in citation forms of real words in the productions of five speakers. Malisz and Żygis (2015) collected recordings from 17 monolingual speakers living in major Polish cities. When it comes to English, Lisker and Abramson (1964) examined citation forms in both series of stops from four speakers of English[19]. In turn, Dochtery (1992) recorded five male Southern British English speakers, with the keywords in a CVC context, in both citation forms and carrier phrases. The table includes also measurements from Polish and American English obtained by Kopczyński (1977) and reported in Zając (2015). On the basis of the available data, we can clearly observe that the difference in the way in which Polish and English employ the parameter of VOT is rather striking. Therefore, it provides a ground for comparison.

Table 5. VOT measurements for Polish and English.

Language	p / t / k	b / d / g	Source
Polish	37.5 / 33 / 49	-78 / -72 / -61	Kopczyński (1977)
	21 / 28 / 53	-88 / -90 / -66	Keating (1981)
	18 / 24 / 35	-57 / -46 / -49	Malisz and Żygis (2015)
English	58 / 70 / 80	1 / 5 / 21	Lisker and Abramson (1964)
	82.5 / 84 / 71	18 / 14 / 31	Kopczyński (1977)

[19] Interestingly, one of the speakers recorded by Lisker and Abramson was responsible for 95% of pre-voiced realisations of the voiced series, indicating that if pre-voicing is present in the speech of a given participant, they are going to produce nearly all their tokens this way, without randomly choosing between pre-voiced and unvoiced realisations.

	42 / 64 / 62	15 / 21 / 27	Dochtery (1992)

Recently, Lindsey (2019: 56) claimed that contemporary British English appears to have even longer aspiration than what was attested a few decades ago, resulting in a much more robust contrast between its voiceless series when compared with the VOT values of languages with short-lag realisations.

3.2.1. The acquisition of initial laryngeal contrasts by Polish learners of English: an overview of selected phonetic studies

When it comes to the phonetic studies concerned with the acquisition of English laryngeal contrast, there is a tendency for the voiced series of stops to be generally understudied. Recall that this was also the case in the studies concerned with phonetic drift effects in L1 Polish (cf. section 2.3.1).

In the studies investigating aspiration in the speech of Polish learners, Waniek-Klimczak (2005) investigated the productions of /p, t, k/ by early and late Polish-English bilinguals and noted that both groups tended to produce values intermediate between Polish and English norms. This finding is in accord with what Flege (1995) predicts and might also point to the existence of interlanguage (Selikner 1972) in their productions. In studies that looked at both series of stops, Zając (2015) showed that Polish speakers converged phonetically with aspirated /p, t, k/, but had problems with pre-voicing supersession in their L2. Similar results were presented by Schwartz and Dzierla (2017), who showed high pre-voicing rates in the speech of very proficient speakers of English, with only ca. 40% of success rate in producing unvoiced, English-like items. These findings were replicated by Schwartz (2020), who compared two groups of speakers: first year students of English and PhD students and Professors from the same institution. He found that while, on the whole, both students and PhD candidates/professors produced voiceless plosives with relatively long VOT (ca. 65ms), pre-voicing was still prevalent in both groups' productions, with only slight decrease in duration for the latter group (ca. -81.1ms in students' realisations vs. -62.2ms in professors'). Importantly, the successful pre-voicing suppression oscillated within chance level – students produced English lenis tokens with full, Polish-like, pre-voicing 58.6% of the time, while professors in 50.3% of the cases.

Furthermore, Dzierla (2019) showed that perceptual training does not appear to affect the rates of pre-voicing suppression in the productions of Polish students of English. One surprising finding of his study was, however, that the voiceless category underwent improvement, even though it was not trained. In other words, while training did not help his participants lose pre-voicing in L2, more native-like aspiration was actually acquired without training.

In perception experiments, the relative weight of VOT as a cue at distinguishing English laryngeal contrasts has been questioned. Rojczyk (2011) found that Polish learners of English are not consistent at locating the boundary between aspirated and unaspirated stops in perception experiments. Aperliński (2012) further found that early learners of English attend to f_0 at vowel onset to a greater extent than to VOT as such when they are tasked with identifying laryngeal categories.

For clarity, the abovementioned studies are summarised in Table 6.

Table 6. Summary of the results of the studies investigating the acquisition of L2 English laryngeal contrasts by Polish learners (in both production and perception).

Type	Series	Task	Results	Source
production	/p, t, k/	word lists	VOT values intermediate between Polish and English	Waniek-Klimczak (2005)
	/p, t, k/ /b, d, g/	word lists	phonetic convergence towards aspirated plosives, but lack of pre-voicing suppression	Zając (2015)
	/p, t, k/ /b, d, g/	word lists	lack of pre-voicing suppression even in advanced learners	Schwartz and Dzierla (2017)
	/p, t, k/ /b, d, g/	word lists	successful acquisition of aspiration but problems with pre-voicing suppression	Schwartz (2020)
	/p, t, k/ /b, d, g/	word lists	perceptual training does not reduce pre-voicing rates but improves aspiration rates	Dzierla (2019)
perception	/p, t, k/	forced-choice identification	lack of consistency in locating the boundary between aspirated and unaspirated stops	Rojczyk (2011)
	/p, t, k/ /b, d, g/	forced-choice identification	F0 as a more relevant cue in category identification than VOT	Aperliński (2012)

All in all, nonetheless, it appears that Polish learners of English have decidedly fewer problems with perception and subsequent production of aspirated English stops. They seem to be more salient and different than what they find in their native language, which leads to more easily obtained success in production. On the contrary, they appear to have less control over pre-voicing.

Interestingly, the asymmetry in acquisition success between L2 English lenis and fortis voiceless plosives has been observed for learners of other languages with a pre-voicing vs. short-lag opposition, including Spanish (Zampini 1998) and Dutch (Simon 2009). These facts, combined with the L1 drift studies described in the preceding chapter, suggest that something in the phonology of laryngeal contrasts may be determining patterns of cross-language phonetic interaction.

3.3. Laryngeal phonology and two-way systems

Having discussed the phonetics of Polish and English laryngeal systems and reviewed a selection of studies concerned with the success rate with which Polish learners acquire the English laryngeal contrast, let us now turn to how phonological theories represent two-way systems. We are going to focus our discussion on traditional feature theory (i.e. SPE-based), Laryngeal Realism and Laryngeal Relativism, and the Onset Prominence framework.

3.3.1. Distinctive features theory and the SPE tradition

The distinctive feature theory can be traced back to the Prague School of Phonology[20]. This linguistic circle is represented most notably by Roman Jakobson (1896-1982) and Nikolaj S. Trubetzkoy (1890-1938). It is difficult to unambiguously state which one of the two first hinted at the idea that the phoneme might not be the smallest contrastive unit in the sound structure of the language. Although in his seminal work, *Grundzüge der Phonologie* 'Principles of Phonology' ([1938] 1962), Trubetzkoy mentions the fact that

[20] Parts of this section have been taken from my unpublished Master's thesis (Wojtkowiak 2016).

phonemes might be made up of smaller elements – which logically implies the existence of features (Anderson 1985: 118) – as noted by Honeybone (2005), this comment might be Jakobson's influence, as it was Jakobson who was responsible for the publication of *Grundzüge* after Trubetzkoy's death[21]. Nonetheless, in Trubetzkoy's theory of oppositions[22] we already see the beginnings of binary contrasts on the basis of which we can distinguish between speech sounds.

Upon analysing Trubetzkoy's natural classes of opposition, Jakobson concluded that most of them are, in fact, bilateral; those which are not, can be further subdivided so as to form bilateral oppositions (Anderson 1985: 117). He proposed to analyse all phonological oppositions from the perspective of binary features. The development of devices such as the spectrograph allowed for an inclusion of acoustic correlates of speech sounds to his theory (Tobin 2006: 173). The proposed set of features was published in the book *Preliminaries to speech analysis: the distinctive features and their correlates*, co-authored by Roman Jakobson, Morris Halle, and Gunnar Fant (1951). There, not only did the authors further develop the initial set of features introduced earlier by Jakobson but also provided detailed definitions and examples of sounds – not exclusively from English – bearing particular features. Distinctive features were defined as "the ultimate distinctive elements of language since no one of them can be broken down into smaller linguistic units" (Jakobson et al. 1951: 3). Similarly to what could be found in Jakobson's previous works, they noted that a phoneme was made up of these features, which put together formed a "concurrent bundle" (Jakobson et al. 1951: 3). The set of features was finite and so was the set of rules that arranged them into phonemes; the process of word-formation, which in basic terms could be defined as arranging phonemes into sequences, was also thought to be governed by a limited array of rules (Jakobson et al. 1951: 4).

When it comes to the laryngeal features conceived of in that work[23], features [voiced] and [voiceless] (Jakobson et al. 1951: 26) were said to be extremely common in consonantal oppositions in European languages and their definitions are relatively

[21] In fact, in his definition of the phoneme, Trubetzkoy states that, "[t]he phoneme (…) cannot be analyzed into still smaller distinctive (phonological) units" ([1938] 1962: 41). Therefore, it seems plausible that some comments pertaining to the existence of features in Trubetzkoy's work are indeed Jakobson's.
[22] Trubetzkoy's natural classes of oppositions included three major classes for vowels (quality, resonance, and prosody) and four classes for consonants (localisation, manner of articulation, timbre, and intensity). A more thorough summary of the natural classes of oppositions will not be provided here, as it is not relevant to the present work, but can be found e.g. in Tobin (2006).
[23] As they are not the main focus of the chapter, we will not focus on other features proposed by Jakobson et al. (1951).

straightforward. The authors distinguished between *buzz* phonemes, i.e. /d, b, z, v/, and *hiss* phonemes, i.e. /t, p, s, f/, which were the voiceless counterparts of the *buzz* ones. Additionally, two features which are worth mentioning are the features [tense] and [lax]. Acoustically, the pair was distinguished on the basis of the behaviour of a segment on the spectrogram. Tense consonants, called also strong or fortes, would be much longer; tense stops would also show a stronger burst than their lax counterparts, called weak or lenes. Therefore, [tense] equalled aspiration. An important issue that the authors discussed was the relationship between tenseness and laxness and voicing and voicelessness (Jakobson et al. 1951: 38). Many languages, in particular the Germanic family, such as English or German, employ the tenseness feature as their primary one with voicing being redundant. As a result, as observed by L.G. Jones (1956) in his study, when some portion of the compression stage in the production of plosives /p, t, k/ was cut out, the listeners perceived these sounds as /b, d, g/; contrarily, speakers whose native languages employed the voicing feature as the primary one, such as Polish, would still hear /p, t, k/, as in their case tenseness is redundant (Jakobson et al. 1951: 38). In a very small number of languages both oppositions are attended to, most commonly in plosives; in this case, voiced stops form one class and aspiration is used to distinguish between tense and lax segments, a distinction found, for example, in Eastern Armenian (Ladefoged and Maddieson 1996: 67). Indic languages are said to employ aspiration to distinguish also between tense and lax voiced stops, but these languages are somewhat exceptional (Ladefoged and Maddieson 1996: 69)[24].

Chomsky and Halle's *The Sound Pattern of English* (henceforth: *SPE*) published in 1968 is one of the groundbreaking works in phonological theory, with Anderson describing it as "by far the most comprehensive presentation and exemplification of the theory of generative phonology to appear up to that point (or since, for that matter)" (1985: 328). The work builds upon the feature theory developed by Jakobson et al. (1951), but introduces a number of changes, especially with respect to the treatment of articulatory vs. acoustic correlates of distinctive features.

The definition of distinctive features read, "[features are] the minimal elements of which phonetic, lexical, and phonological transcriptions are composed, by combination and concatenation" (Chomsky and Halle 1968: 64). Phonemes, in the traditional sense of

[24] These issues will not be elaborated on in the present thesis as the main focus is placed upon two-way laryngeal contrasts.

letter-sized units, were on the other hand treated as "abbreviations for feature bundles, introduced for ease of printing and reading, but without systematic import" (Chomsky and Halle 1968: 64) and were not mentioned throughout the work. The authors suggested that underlying representations of utterances be represented as a linear structure (Chomsky and Halle 1968: 296). A string of speech sounds would be shown in a sequential way, as columns, each of which corresponded with a single segment.

These columns contained matrices of plus- or minus- valued distinctive features. Rules of grammar would then transform the underlying representation into its final phonetic form; these rules were allowed to either add or delete columns, as well as change the specification of each matrix. Features introduced in SPE are innate and phonetic in nature, and they are meant to represent "independently controllable aspects of the speech event" (Chomsky and Halle 1968: 298). Features were assigned parallel articulatory gestures to describe them and the authors admitted that the reason why they ceased to rely on the acoustics, therefore no longer following Jakobsonian tradition, was for fear of their work becoming too lengthy.

Laryngeal features were part of the so-called *source features*[25]. The distinction between voiceless and voiced segments was based upon the binary [±voice]. The difference between [+voice] and [-voice] segments boiled down to the presence or absence of vocal fold vibration, respectively, in the articulation of the segment. In order to distinguish between aspirated and unaspirated segments, the feature [heightened subglottal pressure] was introduced. This feature, however, has been claimed to be rather controversial. Dixit and Shipp (1985) analysed Hindi stop consonants and showed that there is no systematic correlation between subglottal air pressure and voicing and/or aspiration. The feature that proved to be more successful among both phoneticians and phonologists was [±spread glottis], introduced by Halle and Stevens (1971). However, this feature is not used to account for aspiration in English in their work, as it is claimed not to be distinctive for English obstruents. Furthermore, Ridouane (2006) notes that a spread glottis may be implemented in many ways and enhanced glottal opening does not unambiguously entail aspiration.

Therefore, bearing in mind the questionable status of [heightened subglottal pressure], in *SPE* tradition the actual phonetic realisation of the stop is of no interest to the

[25] Once again, other features included in *SPE* will not be discussed, as they fall beyond the scope of this chapter.

phonological specification; VOT typology appears to play no role in the way in which the stops are represented. While the differences between pre-voiced and unvoiced stops as well as between plain voiceless and aspirated stops do lie in their phonetic realisation but not in the phonological underlying representation of the voicing contrasts. As a result, both pre-voiced and unvoiced stops are represented as [+voice], while plain voiceless and aspirated stops as [-voice]. Table 7 summarises the SPE representation of laryngeal contrasts in aspiration and voicing languages on the examples of English and Polish.

Table 7. Binary representation of voice on the example of English and Polish.

Language	Series	Specification
English	/p, t, k/	[-voice]
	/b, d, g/	[+voice]
Polish	/p, t, k/	[-voice]
	/b, d, g/	[+voice]

Such an approach to phonological analyses is not limited to Chomsky and Halle's work only; an adherence to a binary set of features with respect to laryngeal considerations can also be found in e.g. Rubach (1996) or Keating (1984).

3.3.2. Laryngeal Realism

The tradition of Laryngeal Realism (Honeybone 2005) has become a widely accepted approach to representing two-way laryngeal contrasts. While the name was first coined by Patrick Honeybone in 2005, similar descriptions related to laryngeal phonology (albeit using different phonological primes) can be found also in the works of John Harris (1994) and Linda Lombardi (1991, 1994), which came before him.

Laryngeal Realism and the way in which it represents the languages with two-way contrasts is strikingly different from SPE. Before we continue, let us compare the representations used by those three authors which are shown in Table 8.

Table 8. A comparison of privative approaches to voicing contrasts on the example of English and Polish.

Language	Series	Lombardi (1991)	Harris (1994)	Honeybone (2005)[26]
English	/p, t, k/	[aspiration]	{H}	|spread|
	/b, d, g/	Ø	Ø	Ø
Polish	/p, t, k/	Ø	Ø	Ø
	/b, d, g/	[voice]	{L}	|voice|

Upon inspection, there are several facts that emerge. First of all, it is clear that all of the authors claim virtually the same thing: in both types of languages one series receives a phonological specification while the other is left unspecified. Furthermore, while Harris employs elements[27] as the fundamental units on which he bases his assumptions, the difference between Lombardi and Honeybone essentially boils down to the names given to the features. The question as to which tactic is better is yet to be agreed upon; while on the whole features relate to abstract phonological entities, Harris (1994: 96) argues that elements are superior in that they are less segment-oriented[28]. This is to say that features on their own are not phonetically interpretable; it is a segment, which is made up of an amalgamation of features, that is the smallest unit that can be pronounced. In turn, elements can be interpreted both on their own and when they combine into bigger units. As a result, they are pronounceable at every stage of derivation which lowers the risk of overgeneralisations. In that sense, Laryngeal Realism is, to a large extent, independent of any particular theory of specifications and allows for a certain degree of freedom in the choice of features or elements.

Second of all, the monovalent nature of features and elements within Laryngeal Realism allows for a certain asymmetry. Namely, in binary approaches, since both members of a pair are specified for a feature (either plus- or minus-valued), they are expected to behave similarly. Here we see that one member of a contrast is unspecified. In aspiration languages, it is the voiced series that lacks phonological specification, whereas the voiceless series bears the feature [aspiration], |spread|, or the element {H}. In turn, in voicing languages, the voiceless series is left unspecified, whereas the voiced series is

[26] The vertical slashes that Honeybone uses for his features are taken from Dependency Phonology (2005: 326)

[27] Elements as units of representation will be elaborated on in section 4.3.2.

[28] One additional problem with Honeybone's usage of |spread| for aspirated segments is that, as has been mentioned, Ridouane (2006) observed that "spread glottis" does not always entail aspiration. While Honeybone argues that his features are abstract and their phonetic interpretation must be carefully considered (2005: 325), the name might be actually very misleading.

represented by means of the feature [voice], |voice|, or the element {L}[29]. Therefore, the VOT typology is reflected in the phonological representations, as opposed to what was the case for SPE-based analyses. We will now focus on some of the implications which can be read off the proposed representations, while simultaneously analysing evidence undermining some of the claims made by Laryngeal Realism.

The asymmetry in feature assignment, in its logical extension, points to the issue of markedness in laryngeal systems. Markedness, introduced first by Trubetzkoy (1939), was originally understood as specification for phonological distinction, and has later been developed to express some degree of complexity, difficulty, or even abnormality (Haspelmath 2013). In general, this term is a source of confusion in linguistics; as noted by Hume (2011) since Trubetzkoy's times, its meaning has become much broader. In phonology, markedness is represented by formal devices (e.g. feature values, diacritics, constraints) or relational mechanisms (e.g. ordering, constraint ranking, quantitative measures) (Hume 2011: 6). On the basis of this, it can be concluded that the category specified for a feature is to be considered marked.

Therefore, bearing feature specifications in mind, the unvoiced lenis series in English and the voiceless unaspirated series in Polish are usually assumed to be unmarked. Phonologically, they are the same; the unvoiced lenis might be subject to passive (or automatic) voicing in voicing-conducive environments[30]. As a corollary, they are supposed to be perceived as more natural and frequent (following Greenberg's (1987) association of markedness with relative typological cross-linguistic frequency). Indeed, the unaspirated voiceless stop is assumed by many to be the maximally unmarked for laryngeal features (Maddieson 1984). Vaux and Samuels (2005) present a review of the arguments made in favour of this claim and, at the same time, provide empirical evidence *against* those arguments (395ff.). Let us have a look at some of them.

First of all, one of the supposed advantages of Laryngeal Realism is how neatly it accounts for neutralisation processes. These are claimed to be a case of laryngeal delinking, which results in a plain voiceless stop (i.e. the unmarked category). However, Vaux

[29] Languages with more contrasts, such as Thai (three-way) or Hindi (four-way) use both features and allow them to combine.
[30] Beckman et al. (2013) studied the amount of intervocalic voicing in Russian and German; they conclude that the consistent closure voicing in Russian vs. inconsistent voicing in German provide a diagnostic that in the latter case, the voiced stops are not specified for any feature.

and Samuels (2005: 397ff.) show that in some languages (e.g. Tlingit) neutralisation results in a voiceless aspirate and it may even entail some fortition (e.g. final devoicing in German).

Another claim tackled in that work is the apparent typological omnipresence of plain voiceless stops. According to Maddieson (1984) this type of stops can be found in 98% of world's languages (as opposed to 29% with voiceless aspirates). Interestingly, Vaux and Samuels highlight the possibility of this argument being based upon wrong criteria, for instance the languages' grammars as opposed to recordings (cf. Szigetvári 1996). The authors subsequently enumerate many studies which point to the fact that the languages described by Maddieson as having plain voiceless stops in their inventories actually show the existence of voiceless aspirates instead.

Yet another argument pertains to the apparent ease of articulation of plain voiceless stops. This also seems not to be corroborated; when it comes to controlling the articulatory gestures, it is easier when the speaker is required to fit into a relatively long VOT time window rather than make sure they meet the strict criteria set for plain voiceless stops (2005: 401). Subsequently, while plain voiceless stops are argued to be less susceptible to reduction processes, the authors argue that fast speech appears not to be any indicative of whether or not something is more or less marked. On the contrary, casual speech processes involving reductions have been shown to target plain voiceless stops all the same, sometimes entailing an increase of their degree of markedness (2005: 402).

Moreover, Vaux and Samuels review a list of ca. fifteen studies which have shown that children appear to acquire aspirated stops more easily, as opposed to the idea that plain voiceless stops are the ones which are acquired first. This, again, stems from the fact that aspirates require less articulatory precision, thus constituting less of a problem to children (2005: 403f). For clarity, Table 9 summarises Vaux and Samuel's (2005) arguments included in their work.

Table 9. Arguments in favour of the status of plain voiceless stops as maximally unmarked juxtaposed with Vaux and Samuel's (2005) data suggesting otherwise.

Argument	Vaux and Samuel's rebuttal
neutralisation processes result in plain voiceless stops	in some languages it results in voiceless aspirates
plain voiceless stops are typologically most common	yes, but Maddieson's (1984) data are not infallible

easy to articulate	actually require more articulatory precision than voiceless aspirates
do not undergo reduction	are targeted by fast speech processes; casual speech may result in an increase of markedness
children acquire them first	they require more articulatory precision; ca. 15 studies have shown that aspirates are acquired first

All in all, we may observe that the imputation of the unmarked status of the plain voiceless stop found in Laryngeal Realism may be unfounded. Haspelmath (2013) goes as far as to suggest that linguists may dispose of the term *markedness* altogether.

Having looked at the issue of markedness and its relation to plain voiceless stops, let us go back to the asymmetrical feature assignment in Laryngeal Realism. Another conclusion which may be drawn from the proposed representations of two-way laryngeal systems is the nature of assimilation in aspiration and voicing languages. Indeed, the asymmetry in feature specifications allows the proponents of this approach to elegantly explain the fact that regressive voicing assimilation appears to be absent in aspiration languages. This stems from the fact that [voice] is not phonologically active in those; there is no feature that can spread which would entail assimilatory processes.

One other finding which does not seem to be reflected in the typology purported by Laryngeal Realism is the case of Swedish. Helgason and Ringen (2008) provide evidence that this language contrasts pre-voiced stops with voiceless aspirates. This constitutes a challenge to phonological analyses which incorporate those phonetic details into their representations, as it goes against privativity since representing it would require overspecification.

While Laryngeal Realism approach appears to have dominated the literature, undoubtedly providing a number of advantages, such as directly encoding the VOT typology into phonological representations and elegantly expressing numerous laryngeal phenomena, it has not managed to avoid encountering issues, which have been presented in this section.

3.3.3. Laryngeal dimensions and gestures

Drawing on the works of Lombardi (1991) and Harris (1994) – therefore on the assumptions of Laryngeal Realism-like approach – Avery and Idsardi (2001) offer an approach whose feature system is much more developed[31]. Fig. 4 represents the basics of their theoretical model.

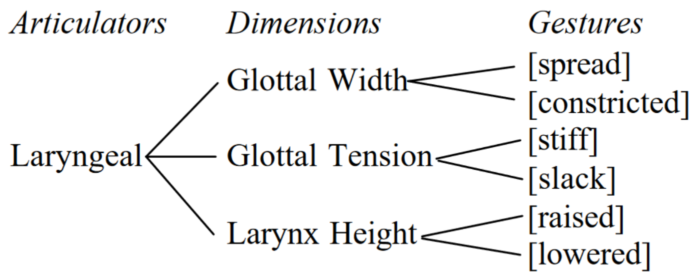


Fig. 4. Avery and Idsardi's model of laryngeal contrasts' representations (2001: 2)

In their approach, phonology is economical, therefore underspecified to a large extent. On the other hand, their phonetic representations are *over*-specified, with more than enough information regarding contrast maintenance. With respect to Fig. 4, gestures play a limited role in phonology and are postulated to be singular articulatory actions. They are unary, as a muscle group can be either activated or not. Gestures form opposing pairs, controlled by dimensions. Dimensions (i.e. Glottal Width, Glottal Tension, and Larynx Height) are claimed to be the initial stage of phonology and are the ones responsible for contrast, yet are not pronounceable by themselves; they are also the primary interference between phonetics and phonology (Avery and Idsardi 2001: 6).

Similarly to Lombardi's and Harris' works, contrast in two way systems is represented as Ø vs. marked, therefore in a pair, one segment in an obstruent pair is always specified for a dimension. Table 10 illustrates the representations of English and Polish from the perspective of Avery and Idsardi's model (2001: 6)[32].

[31] Due to the subject matter of the present chapter we limit the discussion to their approach to laryngeal contrast; however, see the appendix in Avery and Idsardi (2001) for an overview to their inventory of distinctive features.
[32] What is noteworthy, their approach successfully handles also three- and four-way contrasts (found in Thai and Hindi, respectively).

Table 10. The representations of voicing contrasts on the example of English and Polish from the perspective of Avery and Idsardi's approach.

Series	Polish	English
/p, t, k/	Ø	Glottal Width
/b, d, g/	Glottal Tension	Ø

The dimensions are mapped onto the phonetic level by means of the process referred to as *completion* (Avery and Idsardi 2001: 7). Completion is distinguished from enhancement; while in the former we only supplement the dimension with the necessary gesture, the latter involves an addition of an extra dimension, leading to over-differentiation of contrast (Avery and Idsardi 2001: 7). Since enhancement results in a Ø/X contrast becoming a Y/X contrast, the Swedish riddle, which Laryngeal Realism was unable to account for, is elegantly explained. The presence of pre-voicing in that language does not stem from overspecification of any sort, but rather is an example of enhancement.

In general, Avery and Idsardi's model is still very linear and segment-oriented, as it assigns the purported dimensions to segments by means of association lines. This is shown in Fig. 5 on the example of the underlying representations of plain voiceless, aspirated, and pre-voiced stops (recall, that similarly to Laryngeal Realism, it is assumed that plain voiceless and unvoiced lenis stops are phonologically identical).

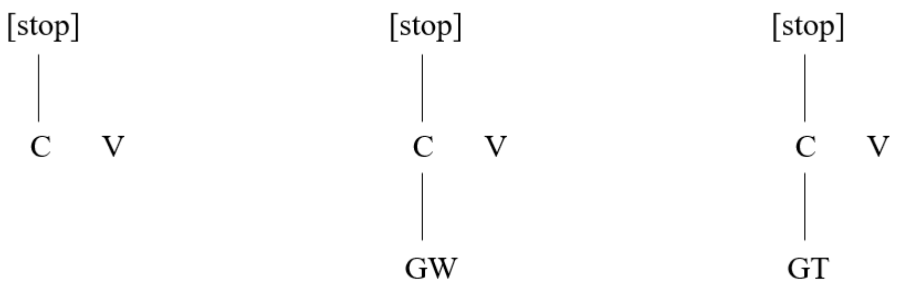


Fig. 5. Representations of the initial laryngeal contrast in Avery and Idsardi's model: plain voiceless (on the left), voiceless aspirated (in the middle), and pre-voiced (on the right).

The model shows, however, that employing articulatory parameters can contribute to very similar analyses as the ones shown with regard to Laryngeal Realism, which was more acoustically-oriented.

3.3.4. Laryngeal Relativism

One of the serious issues that Laryngeal Realism encounters is accounting for *sandhi*-phenomena in voicing languages. One such phenomenon is found in Polish. In certain dialects[33], aside from regular voicing agreement found in Standard Polish[34], word-final obstruents in pre-sonorant environment are also voiced. An example is given in Table 11 below.

Table 11. *Sandhi*-voicing in Standard and dialectal Polish.

Example	Standard Polish realisation	Poznań-Kraków realisation
kot Basi 'Basia's cat'	[kɔd baɕi]	[kɔd baɕi]
sok malinowy 'rasberry juice'	[sɔk malinɔvɨ]	[sɔg malinɔvɨ]
kot Igora 'Igor's cat'	[kɔt igɔra]	[kɔd igɔra]

Before Cyran, this *sandhi* phenomenon in Polish was regarded as simply another example of regressive voice assimilation (Bethin 1984; Gussmann 1992; Rubach 1996). In those analyses, sonorants were ascribed a [voice] feature which would spread onto the preceding obstruent in *sandhi* dialects but only across a word boundary. The problem, however, is that vowels and sonorants cannot be phonologically specified for [voice], even though they are typically voiced[35]. A phonetic argument against the treatment of *sandhi*-voicing as an instance of regressive voice assimilation came from Strycharczuk (2012). She showed that, on the whole, the word final obstruents were voiced to a larger extent when they preceded a voiced obstruent rather than a sonorant. Furthermore, it was shown that underlyingly voiced consonants were more susceptible to *sandhi*-voicing.

On the whole, Cyran is generally against the phonetic bias found in Laryngeal Realism. He stresses that incorporating every phonetic detail that we find in our data into phonological representations does not entail a better understanding of the issues at hand

[33] To be exact, in Poznań-Kraków Polish, associated with dialectal areas spanning Greater and Lesser Poland (Wielkopolska and Małopolska regions respectively).
[34] In Standard Polish, obstruents always agree in voice, and it is C_2 that determines voicing of the whole cluster. Therefore in *sandhi* contexts, voiced C_2 induces voicing of the word-final obstruent which precedes it. More details about Polish laryngeal phonology will be provided in section 3.2.
[35] Rice (1993) put forward a proposal which distinguished between two different [voice] features, out of which one was employed to specify obstruents and the other – sonorants.

(Cyran 2020). Therefore, while phonetics may shed light on what type of laryngeal contrasts we observe in a specific language, it cannot provide any insights into its phonology (Cyran 2020). Cyran stresses the fact that linguists often mistakenly confuse mere phonetics with actual phonological operations (2011). Therefore, while he does incorporate phonetics into his phonological representations to a certain extent, the actual relationship between phonetics and phonology is in his view largely arbitrary (Cyran 2011). It is this arbitrariness that appears to be the driving force behind the approach he proposes, called Laryngeal Relativism. Essentially, the relationship between the phonological categories |L| and |H| and the phonetic realisation thereof is claimed to be relative, so long as sufficient phonetic distance between the two series is maintained (both perceptual and articulatory). It is reminiscent of Kaye's *Phonological Epistemological Principle* which reads, "the only source of phonological knowledge is phonological behaviour" (Kaye 2005: 283).

As a result of the arbitrary relationship between phonology and phonetics, in his approach Cyran distinguishes between three types of voicing: spontaneous (cf. Westbury and Keating 1986), active (cf. Harris 2009), and passive (cf. Kohler 1984). Spontaneous voicing is mostly associated with vowels and sonorants, but oftentimes obstruents might be subject to it. It occurs mostly in environments conducive to voicing (e.g. in an intervocalic position) or used for contrast enhancement. Active voicing is the interpretation of the feature [voice] or |L|, whereas passive voicing concerns the unmarked series in an aspiration language. Passive voicing is disallowed in voicing languages, as differentiating between actively and passively (or: weakly) voiced obstruents would be impossible, in both perception and production. He posits different laryngeal systems in *sandhi*-voicing and non-*sandhi*-voicing dialects, despite the fact that word-internally they show the same behaviour.

Therefore, going back to the problem of *sandhi*-voicing in dialectal Polish (cf. Table 11), Cyran's (2011) analysis thereof is innovative as the phenomenon does not stem from any arbitrary rule but rather is characteristic of the laryngeal system. This constitutes a significant break for laryngeal phonology of Polish. He postulates that Standard Polish is an example of a typical L-system. The active |L| category in the voiced-initial obstruent spreads onto the preceding voiceless obstruent in the process of regressive voice assimilation. Such a process in the case of a sequence of a voiceless obstruent followed by a sonorant spanning a word boundary is impossible in an L-system, as the element |L| cannot be licenced in this context. This is shown in Fig. 6.

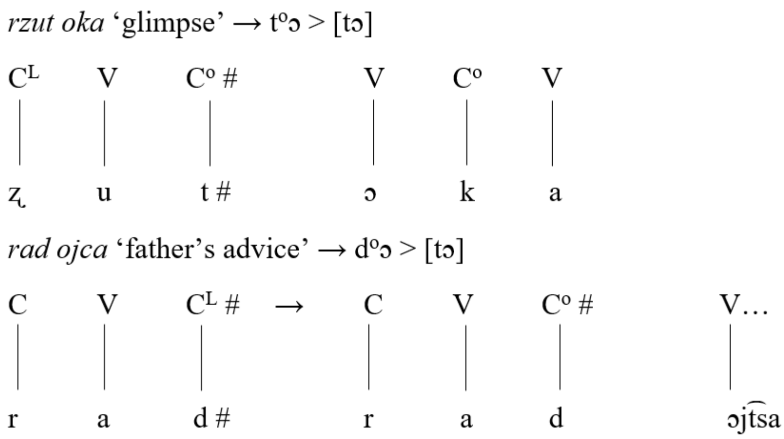


Fig. 6. Cyran’s representation of the lack of *sandhi* voicing process in Standard Polish (adapted from Cyran 2011: 75).

Cyran postulates that Poznań-Kraków Polish is actually an example of an H-system, which due to the arbitrariness of the relationship between phonology and phonetics in Laryngeal Relativism, results in a system that is strikingly akin to an L-system, in that there is no aspiration and full voicing. However, the full voicing is referred to as “enhanced passive voicing” (2011: 69). Essentially, voicing in this dialect is not an interpretation of an active feature [voice] (or, |L|), but rather a phonetic realisation of an unmarked member of the contrast.

Enhanced passive voicing requires a spontaneously voiced segment to follow, which explains the presence of pre-voicing in word-initial position and lack thereof word-finally in dialectal Polish. Word-final devoicing in Standard Polish is treated as delaryngealisation (i.e. delinking of the |L| element word-finally), whereas in dialectal Polish it is the by-product of the absence of a voiced segment that follows. In the case of word-final voiceless obstruents in voicing dialects (which, as in other H-systems, are specified for |H|, as opposed to Standard Polish where this series is unmarked), the |H| element is lost. This allows enhanced passive voicing to take place in pre-sonorant and pre-vocalic positions. *Sandhi* voicing in dialectal Polish is formalised in Fig. 7.

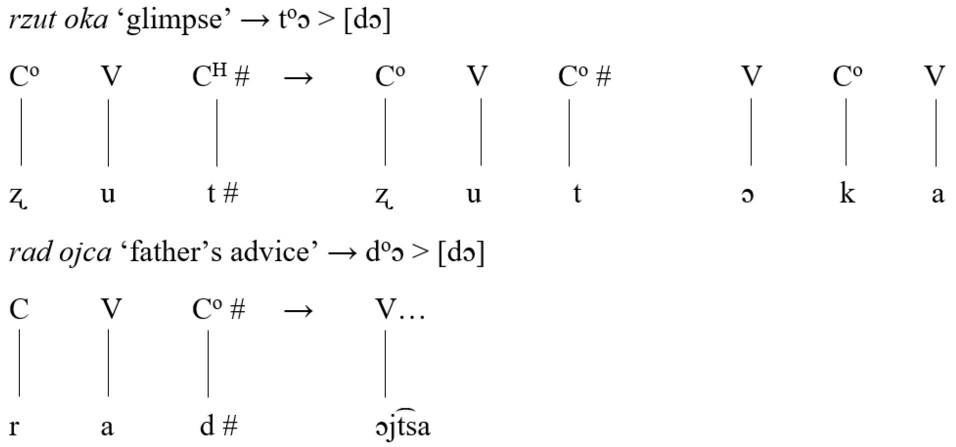


Fig. 7. Cyran's representation of the presence of *sandhi* voicing in dialectal Polish (adapted from Cyran 2011: 76).

In other words, in Standard Polish the obstruent preceding a sonorant is realised as voiceless as L-delinking cannot be undone; in dialectal Polish it must be realised as voiced if the H-element is lost (Cyran 2011: 76). Therefore, what results in identical surface phonetic facts – as both dialects on the face of it to behave similarly in all contexts aside from *sandhi* – is due to different phonological operations. Cyran assumes that one language can actually have two different systems.

Quite possibly the most crucial observation that we get in Cyran's analyses is the following: "negative VOT is no guarantee of the presence of feature |voice| or element |L|" (2011: 77). In other words, he claims that voicing need not be phonological. We will see this observation taken to its logical conclusion in section 3.5.2 (cf. Schwartz 2020). Interestingly, the relativist perspective allows Cyran to also solve the riddle posed by Swedish. In his approach, the presence of pre-voicing in this H-system is just the phonetic observation that we make, as voicing is not indicative of there being an active |L|. Recall that the relationship between phonology and phonetics is for him arbitrary, and the main objective is to maintain adequate contrast between the categories. For some reason, "goes for maximal dispersion rather than sufficient phonetic distance" (2017: 502), but that, in itself, is not a problem for Laryngeal Relativism.

3.4. Unresolved issues with respect to laryngeal phonology

In the previous sections we have discussed some of the more prominent theories of accounting for two-way laryngeal systems. While some problems associated with them have already been discussed (in particular the issue of markedness), two other prominent quandaries arise. The following sections will discuss the apparent overreliance on segmental representation of speech as well as the long-standing debate on feature values.

3.4.1. Overreliance on segments and linear representations

The problem of segmental representation of speech and using phonemes in phonological analyses is not new. As early as in 1953, Weinreich notes that phonemes are not commensurable and we cannot make comparisons between languages based on those entities. Port (2007) argues that using alphabetic letter-sized units which are arranged linearly is intuitive to those who know Roman alphabet. As such, it is an example of Eurocentric bias and “lifelong training using letters to represent speech” (2007: 505), despite the fact that the link between sequentially-ordered letter-sized units and the overlapping aspects of continuous speech appears to be missing.

At first glance, one would think that an important break came with the introduction of features. While they did allow us to observe near-minimal pair contrasts between speech sounds, they were still heavily segment-based. Even though for Chomsky and Halle (1968) phonemes were nothing but a “shortcut” for feature bundles, those feature matrices were still arranged horizontally in a sequential manner, with no overlaps. Laryngeal Realism has not brought about much change either; the features – or elements – are still assigned to specific segments. Honeybone (2005) goes even a step further; he comes up with a new set of diacritics used in transcription which are supposed to indicate the difference between plain voiceless (and unvoiced lenis), voiceless aspirates, and voiced series, given in Table 12.

Table 12. Honeybone's new symbols for segments in two-way systems (2005: 332).

Letter	Polish symbol	English symbol
<b>	[b]	[p˚]
<d>	[d]	[t˚]
<g>	[g]	[k˚]
<p>	[p˚]	[p^h]
<t>	[t˚]	[t^h]
<k>	[k˚]	[k^h]

Introducing changes in transcription in the 21st century once again highlights the apparent attachment of many phonologists to segmental representation of speech. It is worth noting that Bloomfield, as early as in 1933, expressed an opinion against incorporating such details into transcription (1933: 84).

An interesting perspective on this issue is provided by D. Robert Ladd (2011). He recognises the fact that segmental representations found in IPA charts – seemingly universally accepted as basic phonological units – are useful insofar as describing general properties of languages but he stresses that they fail short of accounting for individual speech acts (2011: 365). Among other things, Ladd provides an example from Kera, the language spoken in Chad, where VOT (a segmental feature) is shown to pattern with tone (a suprasegmental feature). Essentially, the relationship between the two appears to be automatic. If VOT patterns with tone, the former is not entirely segmental. In turn, if tone patterns with VOT, the former is not entirely suprasegmental. The Kera problem gives us further evidence that our basic understanding of the distinction between what is segmental and what is suprasegmental might be in need of refinement. All these issues stem from the unanimous acceptance of segmental representation as a default in phonology.

It appears, then, that the discussion on segmental representation of speech boils down to the long-standing debate on the interface between phonetics and phonology and to what extent those two are intertwined, in particular what the relevant units in phonology are. Nevertheless, despite some claims disregarding the necessity of conducting phonetic studies in order to validate phonological assumptions (e.g. Substance-free phonology; Hale and Reiss 2000), acoustic experiments have been shown to shed new light on some of the impressionistic assumptions made by phonologists and improve phonological analyses of various processes. As noted by Ohala (1990), if phonological representations fail to refer to phonetic research, they may fail to accurately encapsulate linguistic phenomena. While some progress in this respect has been made in recent years, "phonetics as a motivating force for phonology remains controversial" (Dziubalska-Kołaczyk 2012). At the same time, however, phonologists appear to fall into the trap of idealisation of phones as units underlying phonological theories and it is to these idealisations that modern phonetic research constitutes a growing problem (Ladd 2011).

3.4.2. Binary vs. unary approach to laryngeal features

In our discussion of traditional accounts of two-way laryngeal systems (e.g. SPE; cf. section 3.3.1) juxtaposed with the realistic approaches (e.g. Laryngeal Realism; cf. 3.3.2) we observed that the choice of feature values has serious implications on the predicted behaviour of the contrasts (symmetric in binary vs. asymmetric in unary accounts). Among other things, one of the main arguments put forward against binary accounts of voicing is that a minus-valued feature – i.e. the absence of voicing conceived of in terms of [-voice] – cannot be phonologically active. However, there are data on the basis of which we can make an opposite claim.

We have already seen that certain dialectal phenomena pose problems to privative analyses of voicing (e.g. Cyran 2014; cf. section 3.3.4). On the same note, Rubach (1996) argues that voicing phenomena in Polish, such as final obstruent devoicing or progressive devoicing (as well as Poznań-Kraków voicing) are better accounted for only if one accepts the binary nature of [voice]. Wetzels and Mascaró (2001) attempt to further show that voicing is, indeed, a binary feature on the basis of a typological overview of laryngeal facts from different languages (or certain dialects thereof).

In Yorkshire English (2001: 227) all voiced obstruents undergo devoicing before a voiceless consonant across word boundaries. This is shown in (1):

(1) *bed-time* → [ˈbettaɪm]
headquarters → [ˌhetˈkwɔːtəz]
frogspawn → [ˈfrɒkspɔːn]

However there is no voicing in the same context. An example that illustrates this is given in (2):

(2) *white book* → [ˌwaɪt ˈbʊk]

Notably, this neutralisation process is so thorough that the underlying /d/ might oftentimes be realised as a glottal stop, which is a well-known positional allophone of /t/.

In Parisian French (2001: 227f.), a dialect with no final obstruent devoicing, in word-internal codas shows obligatory regressive devoicing or optional regressive voicing. As a result the following morpheme-internal clusters are attested: voiceless+voiceless (e.g. *distinctif* [distɛ̃**kt**if] 'distinctive') voiced+voiced (*anecdote* [anɛ**gd**ɔt] 'an anecdote'), voiceless+voiced (*anecdote* [anɛ**kd**ɔt] 'an anecdote')[36], but voiced+voiceless is disallowed, as [-voice] spreading is mandatory.

Another interesting example comes from Ya:thê (2001: 228), the language spoken in the Pernambuco region of Brazil, which contrasts voiceless aspirated, plain voiceless, and voiced series. In Ya:thê consonant clusters are found in word-initial positions, while word-internally they are the by-product a lexical rule of vowel deletion. While there is no voice spreading (e.g. [**sd**adaka] 'spider', [**pd**ãneka] 'to slip', [**fd**esea] 'fog'; Wetzels and Mascaró 2001: 230), voiceless obstruents have been shown to regressively devoice voiced obstruents. Therefore, not only is voicelessness phonologically active, it is contrastive with aspiration.

The authors further note that Lombardi (1996), one of the representatives of privative treatment of voicing, changed her opinions on the matter. Recall that in her original analyses, voicelessness played no role in phonology (1991, 1994; cf. section 3.3.2). Later (1996), she assumed that [voice] is privative lexically, but binary post-lexically. As a rebuttal, Wetzels and Mascaró provide a number of examples of lexical activity of [-voice].

More recently, Bennett and Rose (2017) provided an analysis of a dissimilation process in Moro, the language spoken in the Nuba Mountains in Sudan, in which [-voice] is phonologically active. This is at odds with privative accounts of two-way laryngeal systems, as Moro is an example of a voicing language, wherein voicelessness is unmarked. Only by employing binary feature specifications to laryngeal features in Moro does the dissimilatory process easily explained – simply put, it is a process whose main objective is to avoid correspondence relationship between two consonants (cf. Surface

[36] Notice that the word 'anecdote' can have two different realisations as the regressive voice assimilation in Parisian French is optional.

Correspondence Theory of Dissimilation; Bennet 2015). Dissimilation in this case must tackle features that exist, therefore [-voice] must be present in Moro.

This section has shown that there still remain a lot of unsolved issues within the scope of laryngeal phonology – from taking the segment for granted, disregarding phonetic research, to evidence in favour of phonologically active plain voiceless stops.

3.5. Onset Prominence representational model

So far, we have discussed the major approaches to denoting laryngeal contrasts in two-way laryngeal systems. However, we have also observed a number of issues associated with those theories. First of all, there is evidence that plain voiceless stops might not necessarily be the maximally unmarked segment. Second of all, not only have we had suggestions pointing to the fact that voicing as such might not be phonological, but it also has been postulated that [spread glottis] or {H} may occur in languages traditionally assigned to the voicing category (Cyran 2014, van der Hulst 2015). Moreover, pre-voicing has been referred to as enhanced passive voicing, thus being treated as part of phonetics.

VOT appears to be insufficient a parameter to describe the phonetic reality of laryngeal contrasts. Nonetheless, the differences with respect to VOT across languages are quite salient and categorical. Therefore, there is a necessity for a theory which would account for different realisations of voicelessness (in light of the evidence supporting the view that voicelessness might be phonologically active in voicing languages, as discussed in Section 3.4.2), and which would at the same time not fall victim to linearity.

An alternative approach to representing laryngeal contrasts which does just that is Onset Prominence (hereafter: OP; Schwartz 2016 *et seq*.) OP is a fully-fledged representational environment which accounts for all sorts of phonological phenomena[37]. In addition, its representations make important predictions with respect to second language acquisition and phonetic drift. Nonetheless, before we proceed with discussing OP's views

[37] Thus far the list comprises onsetless syllables and 'empty onsets' (Schwartz 2013); opacity in Polish palatalization alternations (Schwartz 2013); phonotactics and syllabification (Schwartz 2015); complications in Irish mutation (Schwartz and Anderson 2015); relative stability of vowel systems (Schwartz 2016, inter alia); release of coda stops (Schwartz et al. 2014); prosodic structure (Schwartz 2016).

on laryngeal phonology, it is important that the rationale behind the framework is presented, alongside an overview of the model's main tenets[38].

3.5.1. OP: preliminaries

Before presenting the formal representations postulated by OP, it is crucial to first discuss place it within phonological theory as such. Schwartz (2016: 40) highlights three main meta-theoretical assumptions of the framework. First of all, there is no phonetic component that would be the intermediate stage between phonological representations and phonetic output, as was the case for Chomsky and Halle (1968). Rather, in a view that resembles that of Natural Phonology (Donegan and Stampe 1979), categorical representations, encoding certain non-contrastive properties, are mapped directly onto speech. As noted by Donegan (2002: 79), "speech perception is categorical, down to the level of phonetic (pronounceable) representation." Therefore, in OP some non-contrastive properties can be accounted for by means of formal representations. Second of all, it is assumed that the phonetics-phonology relationship is driven by perception. As explicitly stated by Schwartz, in OP "it is speech perception, rather than production, that drives the formation of phonological categories" (2020: 7). This view can be traced to Ohala (1981), who stressed that it is the listener, rather than the speaker, who is the instigator of any phonetic sound change. Namely, the listener faces many challenges when they perceive speech (e.g. ambiguities stemming from production, such as vowel distortions due to the consonants adjacent to them) and is subsequently tasked with reconstructing what they heard and producing it by themselves. If they fail to do it correctly and their mispronunciation is copied by other listeners, this may result in a sound change. In other words, then, perception takes precedence over production, as for any change to take place, it must first be perceived by the listener. Finally, phonological primes emerge on language-specific basis, with the exception of one universal: a hierarchical CV structure[39], based on a stop+vowel sequence.

[38] The most comprehensive overview of the model can be found in Schwartz (2016). In the present work, we limit ourselves only to the most crucial information regarding the framework.
[39] The CV structure (albeit not in a hierarchical manner like here) has been also employed as the basic unit by other frameworks, e.g. CVCV (Scheer 2004).

The stop+vowel CV sequence in OP is the most basic unit, from which all representations are abstracted. The representational hierarchy is thus shown in Fig. 8.

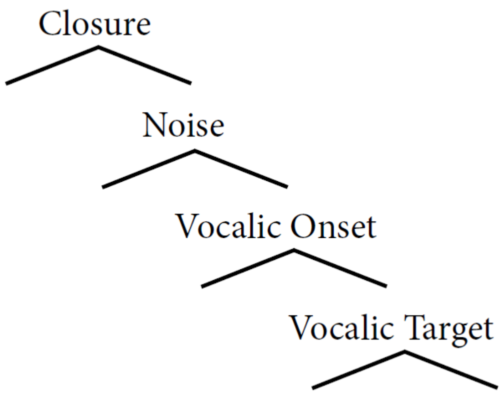


Fig. 8. OP representational hierarchy (after Schwartz 2016: 41)

Each layer of the tree is derived from an identifiable acoustic landmark in a CV unit (cf. Stevens 2002). In this way, a CV sequence has an internal structure that is independent of its segmental makeup. This is shown in the right-most tree given in Fig. 10. The level of C(losure) corresponds to the hold phase in the process of stop production. The subsequent layer, N(oise) is the release noise of the stop that follows the burst. The next layer, V(ocalic) O(nset) is the initial portion of the vowel, characterised by formant movement associated with vowel transitions. Finally, V(ocalic) T(arget) corresponds to the portion of the vowel where the formants reach their maximal values and are relatively stable.

A spectrogram of the English onset /gɑː/ is shown in Fig. 9. The annotations thereof illustrate the phonetic events which build the CV tree representation in Figure Fig. 10. The portion between the release noise and the stable portion of the vowel corresponds to the VO[40] level and is supposed to indicate the formant movement at the onset of the vowel.

[40] The affiliation of the VO level with a consonant or a vowel depends on a language and has important implications. Since this example comes from English, where VO has consonantal affiliations, it is included in the consonant. More on the status of the VO can be found for example in Schwartz 2016.

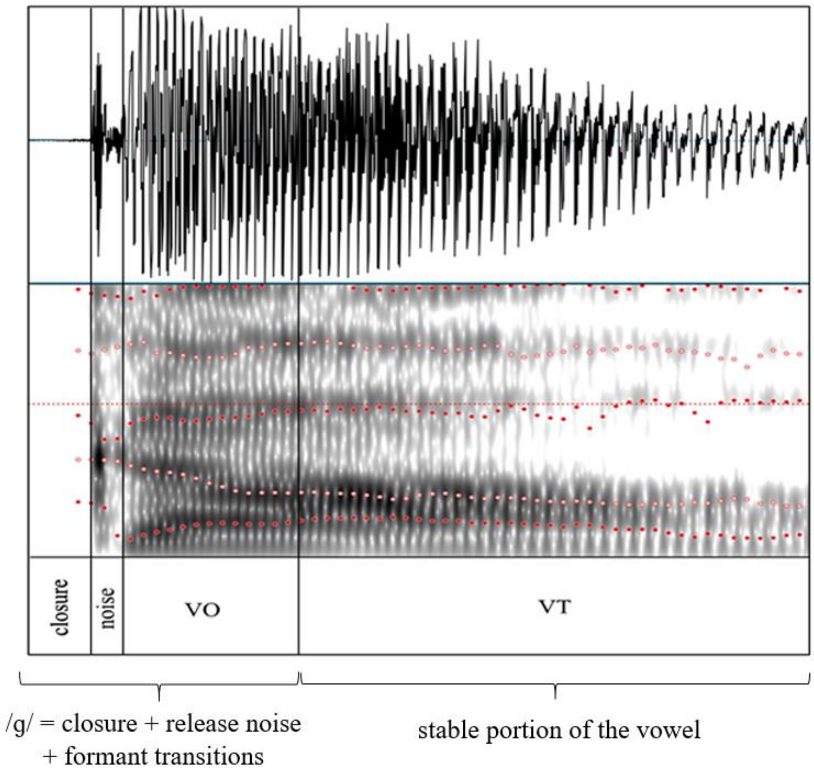


Fig. 9. A spectrogram of the English onset “gɑː”, illustrating the acoustic landmarks which make up the universal CV structure in OP.

A pertinent question that might arise at this point is the rationale behind employing a CV sequence as the basic unit. Previously, Golston and van der Hulst (1999) proposed that lexical representations are actually syllabified in the lexicon, and as a corollary segmental representations are unnecessary. They argued that the syllable can be used to encode most major class and manner features and as such, is “the state of affairs” rather than a derivational process. Typologically, virtually every attested language makes use of the syllable (Maddieson 2013) and a syllable-free representation appears unlikely (but see Dziubalska-Kołaczyk 2002 *et seq.*, for whom the syllable is epiphenomenal). The authors also make use of the prosodic structure in the stead of segmental representations, as there is no evidence against underlying prosodic structures available. In their proposal, stricture is not encoded featurally, but rather structurally. This goes in line with OP’s views; additionally, from the typological perspective the CV structure is the most common type of syllable structure (Maddieson 2013).

Returning to OP representations, the framework conjectures that manner of articulation is structural. While the general issues concerning the role of phonetics in phonology are still to a large extent unresolved, in his revision of the physical realisation of speech sounds, Schwartz (2014) observes that the phonetics of manner of articulation is phonological in nature. He suggests that listeners are more sensitive to gradience in place of articulation and voicing than in manner. In other words, manner appears to be more categorical. Furthermore, manner is cued by the relation between two distinct scales: amplitude and time, making the perception of manner more cognitively demanding. In turn, place and voicing do not require the listener to rely on more than one scale; the former is limited to spectral properties, the latter – temporal. This means that both are contingent on one single cue. On the basis of these observations, in OP "segments" are derived from the universal CV structure (cf. Fig. 10) and manner of articulation is encoded as active binary nodes.

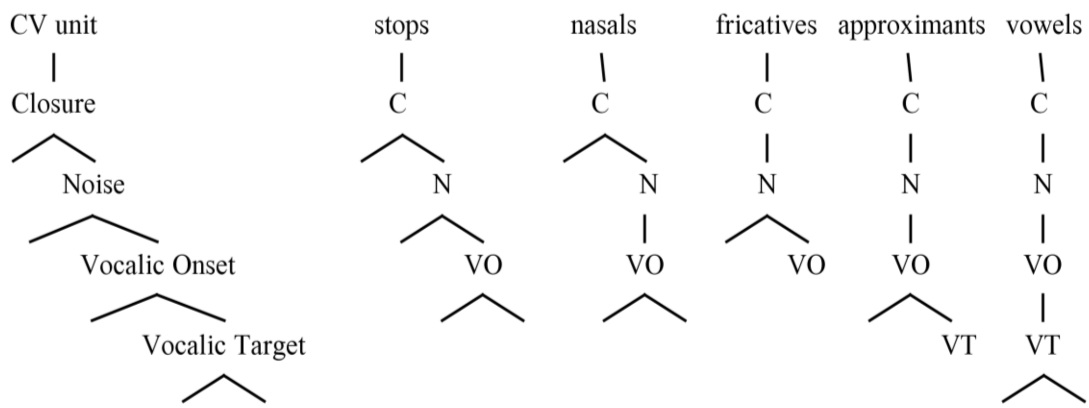


Fig. 10. Onset Prominence hierarchy (right-most tree) and the 'segments' extracted thence (after Schwartz 2016).

As can be observed, stops are characterised as having all layers as they are part of the universal CV structure, therefore all nodes are active. Nasals lack release noise; fricatives lack closure but have release. Approximants, in turn, lack both Closure and Noise, while the only active Node in Vowels is the VT node. Prosodic constituents are formed by means of joining the individual trees together[41]. Place of articulation and voicing specifications, in turn, are attached to structural nodes and as such belong to the domain of phonetics-phonology interface (Schwartz 2014: 357).

[41] The most basic mechanism responsible for this is "absorption", defined as a repair strategy which joins a sonorant with the preceding consonant in order to form the universal CV structure, in accordance with the Minimal Constituent Constraint, which says that a well-formed prosodic constituent must have active nodes above and below the VT level (Schwartz 2016: 43). There are also different mechanisms which join

One of the most important aspects of OP is that it unites segmental representations and prosodic constituents by means of disposing of association lines, as a result creating one representational system. This idea can be traced back to Pöchtrager (2006), who in his model referred to as GP 2.0, postulated that phonological domains are organised in a hierarchical manner, similarly to syntactic domains in minimalist syntax. In GP 2.0 association lines are replaced with m-command, as association lines require a model with an ample inventory of elements (Pöchtrager 2006: 82). OP employs Pöchtrager's insights into the role of association lines and develops this idea further. Essentially, the Prosodic Hierarchy, which includes a number of domains, such as – moving from smallest to largest – the syllable, foot, prosodic word, clitic group, phonological phrase, intonational phrase, and utterance (Nespor and Vogel 1986), is traditionally thought of as being part of the Universal Grammar. As such is an innate property of all languages, therefore being imposed upon them. It is the association lines that link it to lower-level segmental representations and can be seen as a boundary between segmental and suprasegmental features. However, when we look at certain processes which occur at prosodic boundaries, such as liaison in French, vowel glottalisation in German, as well as a number of prosodically-conditioned phonetic variation in the phonetic realisation of segments (in languages such as Korean, Dutch, or English), we will notice that association lines must be disposed of if one wants to explain those processes (Schwartz 2016: 38). Instead, segmental realisations are included in the prosodic structure. Furthermore, as there is evidence that suggests that even such basic prosodic domains as the prosodic word is an emergent phenomenon (Schiering et al. 2010)[42], OP questions the universal status of the Prosodic Hierarchy, instead proposing the way in which prosodic constituents develop independently in different languages (cf. Schwartz 2016; Wojtkowiak and Schwartz 2022).

As has already been mentioned, as there are no association lines, place specifications are assigned directly to structural nodes. This is shown in Fig. 11. [Place] is assigned to the level of Closure and subsequently 'trickles' down onto the Noise and VO levels (Schwartz 2016: 45). Trickling is defined as salient phonetic carry-over stemming from feature assignment (Schwartz 2016: 45). In other words, the [place] feature attached to Closure is still perceptible in noise spectrum or formant transitions. Trickling is blocked

the individual trees to form bigger units (i.e. submersion, adjunction, promotion), which will not be discussed herein as they fall beyond the scope of the present thesis (but see Schwartz 2016).

[42] There are also theories which dispose of domains such as the syllable, e.g. Dziubalska-Kołaczyk 2002.

by assigning another feature at a given level (most commonly associated with vowel quality at VT level).

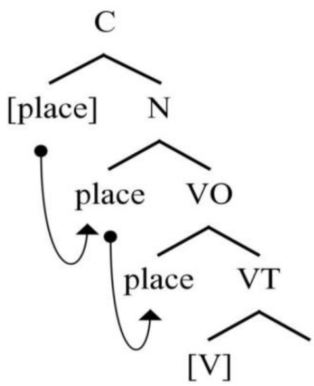


Fig. 11. Place of articulation assignment in OP (after Schwartz 2016)

Notice that the usage of square brackets contributes to the distinction between attached and trickled features. This is important, as for example, it makes it possible for us to distinguish between a spirantised bilabial stop (resulting in a bilabial fricative) and a labiodental fricative (Schwartz 2016: 45), illustrated in Fig. 12. The difference between the two fricatives lies exactly in the type of a feature that is assigned at the Noise level. If in the process of spirantisation, the active Closure node is lost, what is left is only the trickled place specification. In contrast, a labiodental fricative has its [labial] feature attached directly to the Noise node.

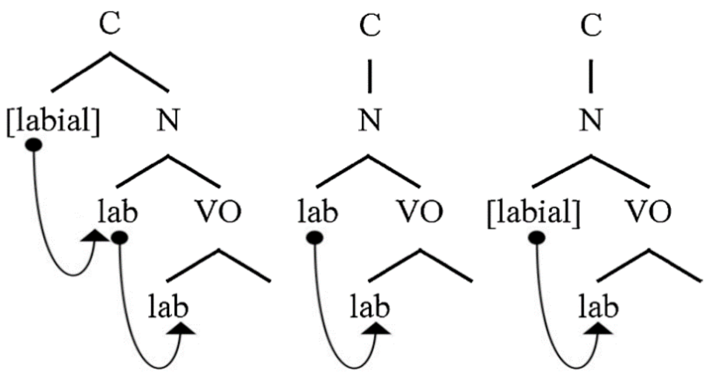


Fig. 12. The representations of a bilabial stop (left-most tree), spirantised bilabial stop (middle tree), and a labiodental fricative (right-most tree) in OP (adapted from Schwartz 2016: 45).

Therefore, not only does trickling allow us to elegantly explain phenomena such as spirantisation or vowel fronting in coronal environments (Schwartz 2016: 48), it also makes it unnecessary for any place-spreading processes to take place.

Dividing the articulation of segments into intermediate steps is not necessarily a new idea. In Aperture Theory (Steriade 1993), stops (as well as affricates) were made up of sequences of closures and releases (thus creating two aperture positions), as opposed to continuants, which were treated as single units. The release of a stop was equalled with the aperture position of an approximant. The three postulated aperture positions are given in (3) (after Steriade 1993: 402).

(3) A_0 – oral closure; minimal aperture;
A_{MAX} – maximal degree of oral aperture;
A_{f} – aperture of enough turbulent airflow for a fricative to be produced;

These aperture positions subsequently combine to create segments. As a result, then, a plain voiceless stop is represented as A_0A_{MAX} (Steriade 1993: 402), its unreleased variant would simply be A_0, whereas an affricate would be represented as A_0A_{f}. As can be observed, while the plosive is divided into two stages, the representations is still very much conceived of as a horizontally ordered linear string.

More recently, Q theory has been proposed (Inkelas and Shih 2013; Shih and Inkelas 2014), wherein segments are again split into three *q* subsegments, i.e. Q(q^1, q^2, q^3), which are supposed to indicate the transition into, target, and transition out of a vowel or a consonant. Q theory is thereby equipped to account for various contour segments, including diphthongs, contour tones, plosives with secondary articulation gestures, pre- and post-nasalised segments, as well as laryngealisation gestures (Inkelas and Shih 2016). The proposed representation of plain voiceless and aspirated series is given in (4) (after Inkelas and Shih 2016: 3).

(4) k (k, k, k)
k^h (k, k, h)

The difference between Aperture Theory and Q theory lies in the number of subsegments postulated by the two frameworks (two in Aperture Theory, three in Q theory) as well as in the fact that the latter subdivides both consonants and vowels into subsegments. However, the basic similarity between the two – that is, ordering the subsegments as a linear string – is maintained. Aside from the linear nature of these two frameworks, in both of them the relationship between phonological primes (i.e. aperture positions and *q* subsegments, respectively) and observable phonetic effects is less straight forward than in OP.

As shown in this subsection, OP offers phonological categorical representations which are enriched Having discussed the necessary preliminaries, we may now move on to the description of laryngeal phonology from the perspective of OP.

3.5.2. Laryngeal phonology within OP

The representation of laryngeal contrasts in OP aims to reconcile the Laryngeal Realism's way of encoding the VOT typology into phonological representation with the idea that [-voice] may actually be phonologically active (cf. 3.4.2). On the whole, the representations of two-way systems[43] depend upon the presence or absence of the feature [fortis] in the structures as well as on the level at which this feature is assigned.

First let us look at voiceless stops. Recall that in OP, three distinct levels are at one's disposal: C(losure) level, N(oise) level, and VO-level. However, if we chose to attach the feature [fortis] at Noise node, the trickling of [place] would be blocked (cf. Section 3.5.1 on trickling). If place of assimilation were to be released only at the level of Closure, it would not be perceptible (Schwartz 2020). Therefore, the two possibilities left are assigning [fortis] at Closure or at VO, shown in Fig. 13.

[43] Since this thesis focuses on word-initial contrasts, we are not going to discuss the ways in which OP deals with word-medial (including, for instance, voice assimilation processes) or word-final positions (including the resolution of final obstruent devoicing). However, these issues have been described elsewhere. Cf. Schwartz (2020) and Wojtkowiak and Schwartz (2018).

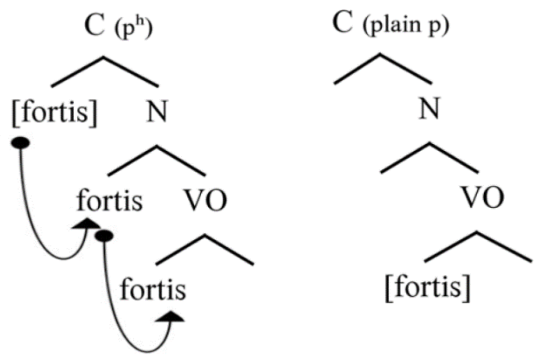


Fig. 13. The representation of an aspirated stop (left-most tree) and a plain voiceless stop (right-most tree) (adapted from Schwartz 2017).

What we can see in Fig. 13 is the aspirated variant of a voiceless stop, found in English, wherein [fortis] is assigned at Closure level and then trickles down onto Noise and VO, resulting in perceptible aspiration. In contrast, in plain voiceless stops, found in Polish, [fortis] is assigned at the VO level, resulting in a relatively short-lag VOT.

Turning to the voiced series, before we discuss the representations of unvoiced and pre-voiced stops, we must first address the issue of what voicing is in OP. The most crucial point to be made is that there is that OP allows to represent voicing languages without resorting to the actual feature [voice]; [voice] is not a phonological entity in this framework. This has already been hinted at by Cyran (2014; cf. Section 3.3.4), but OP takes his observations to their logical conclusion. The immediate question to be asked at this point pertains to what voicing is, if it has no bearing on phonology. Essentially, it is conceived of as part of the "carrier" acoustic signal, described by Traunmüller in his Modulation Theory (1994). The carrier is conceived of as a schwa-like vocoid with evenly spaced formant structures. As part of the acoustic background, the carrier signal encapsulates the linguistic message, including extra-linguistic background information, e.g. the speaker's age, sex, or mood. It is by definition voiced. Carrier modulation, then, refers to departures from the carrier. Therefore, while voicelessness represents a more salient departure from the acoustic signal, relative to voicedness, the latter of which *being* the carrier itself. Only phonological primes can modulate the carrier, therefore [fortis] appears to be a better candidate for phonological status (Schwartz 2017).

Having said that, let us then look at the OP representations of two series of stops in aspiration and voicing languages, respectively. They are given in Fig. 14.

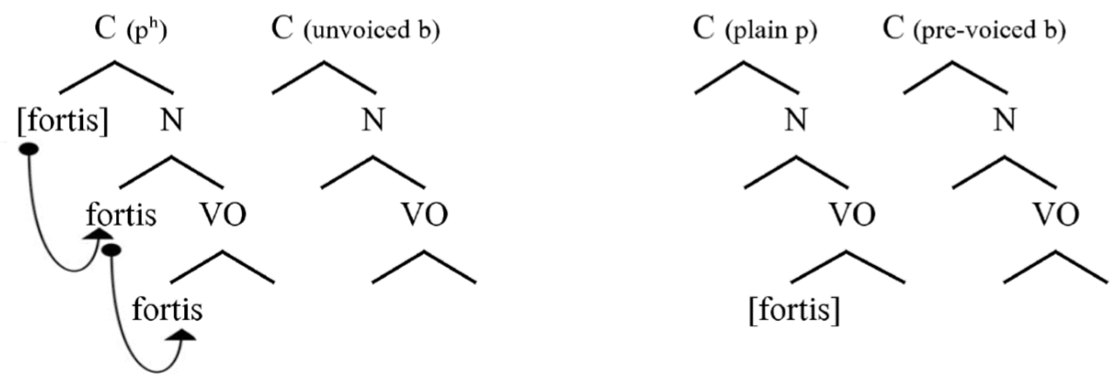


Fig. 14. OP representations of voicing contrast in aspiration (two left-most trees) and voicing (two right-most trees) languages (adapted from Schwartz 2017).

What we may notice is that the only phonological feature necessary to account for two-way laryngeal contrasts is [fortis]. The phonetic realisation of this feature – resulting in either aspiration or short-lag VOT – is contingent on the level at which the feature is assigned, which has already been explained when Fig. 13 was discussed. In turn, both unvoiced lenis and pre-voiced stop are phonologically identical and are not specified for any feature. Pre-voicing as such is of little phonological importance and is regarded as a phonetic detail (or, a manifestation of the carrier).

This approach is strikingly different to what we saw in Laryngeal Realism, for which the presence of pre-voicing was crucial as the phonological representations relied on it. Pre-voicing, in general, is evidenced to have little phonological importance. First of all, no languages contrast two laryngeal series on the basis of the length of pre-voicing the way they do with positive VOT (Schwartz 2020). Furthermore, pre-voicing has been shown to be a less stable cue to voice contrasts in voicing languages than in aspiration languages (e.g. Keating 1979; Schwartz and Arndt 2018; Schwartz et al. 2019). The general results of the phonetic studies discussed in section 2.3.1 in relation to consonantal phonetic drift also highlighted the higher likelihood of pre-voiced and unvoiced consonants to be classified as equivalent. If these two types of realisations are perceptually the same, pre-voicing is not really of interest to phonology.

The representations proposed by OP elegantly account for situations where we find both pre-voicing and aspiration. Such scenarios were attested in Swedish (Beckman 2011), Turkish (Petrova et al. 2006), Southern American English (Hunnicutt and Morris 2016), and Standard German (Hamman and Seinhorst 2016). Explaining this by means of the representations provided by Laryngeal Realism could result in overspecification;

combing the features responsible for voicing and aspiration is characteristic of laryngeal systems with three- and four-way contrasts. In Laryngeal Relativism, we saw that Cyran managed to avoid the issues with which Laryngeal Realism failed short of accounting for and claimed they were not a problem for his theory due to the arbitrary relationship between phonetics and phonology. Cyran saw the co-existence of pre-voicing and aspiration in one system as accidental. In OP these problems are non-existent since the explanation falls out directly from the representations; the appearance of pre-voicing in aspiration languages is simply a manifestation of the carrier, which could stem from the need of contrast enhancement or cross-language influence (the latter of which being a possibility for Southern American English due to its proximity to Mexican Spanish, a voicing language).

There is empirical evidence which yields support to the idea of employing only the feature [fortis] to express laryngeal contrasts in languages with two series of stops. It has been observed that voicing affects pitch (i.e. f_0) at vowel onset insofar as pitch was raised following a voiceless consonant (Ohde 1984). With respect to what was proposed by Laryngeal Realism, such effects should be found in aspiration languages as there is an active [aspiration]/{H}/|spread| feature that could influence pitch. This has been corroborated by Hanson (2009), who found it to be true for American English, an aspiration language. Accepting Laryngeal Realism's representations, pitch should not be affected in voicing languages as voiceless stops are left unspecified as there is no active phonological feature to raise pitch; we could, in turn, expect the active [voice] feature to lower it. However, using nasals as a baseline, Kirby and Ladd (2016) and Schwartz et al. (2019) studied voicing languages – French, Italian, and Polish – and their results replicated those of Hanson's for American English. Pitch lowering can be claimed to be a universal process. In OP representations pitch lowering is expected in both types of languages, as [fortis] is an active feature in aspiration and voicing languages alike.

Finally, is worth noting that OP representations assume that unvoiced lenis and plain voiceless stops are not phonologically identical, as was the case for what Laryngeal Realism proposed. The difference between the two might be observable both in terms of VOT values, but also on the effects that assigning the feature [fortis] in plain voiceless stops might exert on pitch and F1 at vowels onset in vowels adjacent to the unaspirated voiceless consonant. This prediction will be explored shortly.

3.6. Predictions for phonetic drift

This chapter has aimed at providing an overview of the most relevant theories which strive to account for two-way laryngeal systems. We have looked at traditional feature theory, dating back to Chomsky and Halle's works (1968) as well as more recent theories, such as Laryngeal Realism (Lombardi 1991, 1994; Harris 1994; Honeybone 2005) and Laryngeal Relativism (Cyran 2014). We have subsequently discussed some problematic issues that these theories encounter, such as a bias towards linear, segmental representations and the problems in feature specifications. Finally, the chapter presented an alternative approach to the representations of laryngeal phonology, Onset Prominence (Schwartz 2016 *et seq*.), and discussed its assumptions.

Going back to the theory and phonetic studies presented in earlier chapters, recall that we assume that phonetic drift effects stem from equivalence classification (Flege 1995). If a new category is created because the new L2 phone is perceptually different enough from an already existing L1 sound, interaction is predicted to be minimal. However, if the L2 and L1 sounds are deemed to be perceptually the same, phonetic drift is predicted. Therefore, drift will be expected to occur between the series which are phonologically equivalent.

Looking at the theories we have reviewed, we can conjecture that the magnitude of L2-induced phonetic drift in L1 will be contingent on what a particular theory treats as equivalent. Table 13 presents a summary of the representations proposed by the three most prominent theories of laryngeal phonology, provides feature specifications for Polish and English (as these two languages are of main interest to the present thesis), and the expected drift magnitude.

Table 13. Predictions for drift degree from the perspective of the three most prominent laryngeal theories.

Theory	Specifications for Polish	Specifications for English	Degree of drift
Feature Theory (Chomsky and Halle 1968)	[+voice] [-voice]	[+voice] [-voice]	drift predicted in both series of stops
Laryngeal Realism (Honeybone 2005; Harris 1994; Lombardi 1991)	[voice] for /b, d, g/ Ø for /p, t, k/	Ø for /b, d, g/ [spread glottis], {H} for /p t, k/	drift predicted in neither series of stops
Onset Prominence (Schwartz 2016)	Ø for /b, d, g/ [fortis] for /p, t, k/ at VO-level	Ø for /b, d, g/ [fortis] for /p, t, k/ at C-level	drift predicted for the voiced series; minimal or none for the voiceless series

As can be seen, traditional feature theory predicts drift for both series of stops, as they are phonologically identical in both languages. Laryngeal Realism, in turn, expects no drift effects to occur in either series, as their specifications are completely different. Finally, in OP, more salient effects of drift are expected to occur in the voiced series, as both prevoiced and unvoiced stops are represented in the same way. Some effects (albeit minimal) could be also found in the voiceless series, as an aspirated stop and a plain voiceless stop both have the [fortis] specification at the VO level (in Polish it is assigned therein, whereas in English it trickles onto both Noise and VO levels). However, a robust asymmetry is still predicted.

The subsequent chapters (beginning with Chapter 5) will present a longitudinal empirical study which has been designed in order to assess the degree to which these theories are correct in what they predict for the phonetic drift effects, thereby providing empirical evidence in favour of or against those predictions. Before then, however, let us discuss the phonology of vowel systems in Polish and English and their relation to the main subject of the thesis.

Chapter 4: The phonetics and phonology of Polish and English vowel systems

4.1. The aim of the chapter

Having discussed laryngeal contrasts, from the perspective of both phonetics and phonology, it is now time to turn to vowels. We have seen in section 2.3.2 that L2-induced phonetic drift in L1 has been found in both vowel production and vowel perception.

The aim of this chapter is two-fold. First of all, the vocalic systems of Polish and English are described. After comparing and contrasting the vowel systems, we then discuss selected studies which investigated the success with which Polish learners of English acquire vowels. The rest of the chapter focuses on what phonological theories – including traditional feature theory and Element Theory – tell us about the make-up of vowels and whether or not they make any predictions with respect to how the principle of equivalence classification might influence cross-linguistic interaction of vowel sounds. Similarly to what we saw in Chapter 3, we will also discuss an alternative approach to representing vowels, which moves away from linear ordering and in favour of a more hierarchical approach, that is OP (Schwartz 2016 *et seq.*).

4.2. The phonetics of Polish and English vowel systems

Polish is a language with a relatively simple vocalic system, compared to its consonantal inventory. Jassem (2003) distinguishes 6 oral vowels, depicted in the vowel chart in Fig. 15. There are also two nasalised vowels, described by graphemes <ę> and <ą>, however their status is disputed and as such they will not be described herein[44].

[44] But see e.g. Bloch-Rozmej (1997) for discussion.

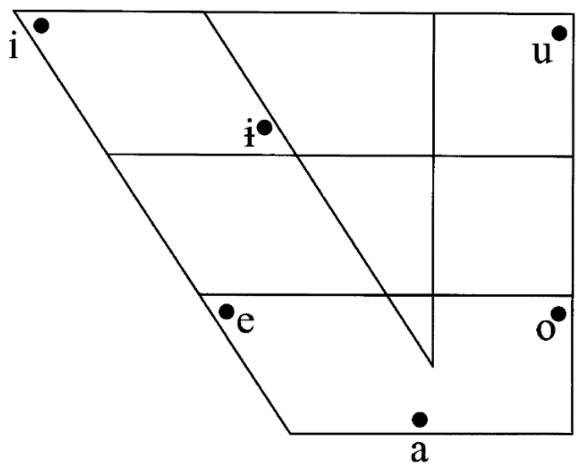


Fig. 15. Standard Polish vowel quadrilateral (after Jassem 2003: 105)

The vowel /ɨ/ is subject to some distributional restrictions: it does not appear word-initially and may not follow palatal and velar consonants (with the exception of the velar fricative /x/). As a result, there is some debate on its phonemic status, with some authors claiming it is an allophone of /i/ (e.g. Feldstein and Franks 2002). Nonetheless, most phonological analyses treat it as a phoneme (e.g. Jassem 2003, Gussmann 2007), with the phonological arguments boosted by the fact that the vowel has some psychological reality for Polish native speakers (Rydzewski 2017). Finally, most accounts seem to prefer the symbols /ɛ/ and /ɔ/ for the open-mid vowels (as opposed to /e/ and /o/ as used by Jassem in his vowel chart).

According to most descriptions, there is little contextual variation within vowels in Polish. However, /ɛ/ and /a/ are raised in palatal context. The effects are particularly robust for /ɛ/ and when the palatal sound precedes the vowel (cf. Weckwerth and Balas 2020: 344). Therefore, the vowel in *sienny* "hay" (adj.) [ɕenːɨ] is not the same as in *senny* "sleepy" [sɛnːɨ]. There is no phonological reduction of unstressed vowels (Crosswhite 2001). Rojczyk (2019) demonstrated that vowels in stressed vs. unstressed syllables differ significantly both in terms of quality and duration. Stressed vowels seem to be articulated at more peripheral points of the vowel space.

On the whole, there are very few studies providing acoustic measurements of Polish vowels. Aside from a study by Strycharczuk and Jurgec (2008), to my knowledge the most recent spectral measurements of Polish vowels can be found in Weckwerth and Balas (2020). The former experiment investigated prosodic effects on the acoustic realisation of Polish vowels in the production of six speakers of Standard Polish (aged 20-25).

The conditions which were studied were monosyllables (CVC) and trisyllabic words (CVCVCV) with pre-stressed, stressed, and post-stressed contexts. The results are shown in Fig. 16.

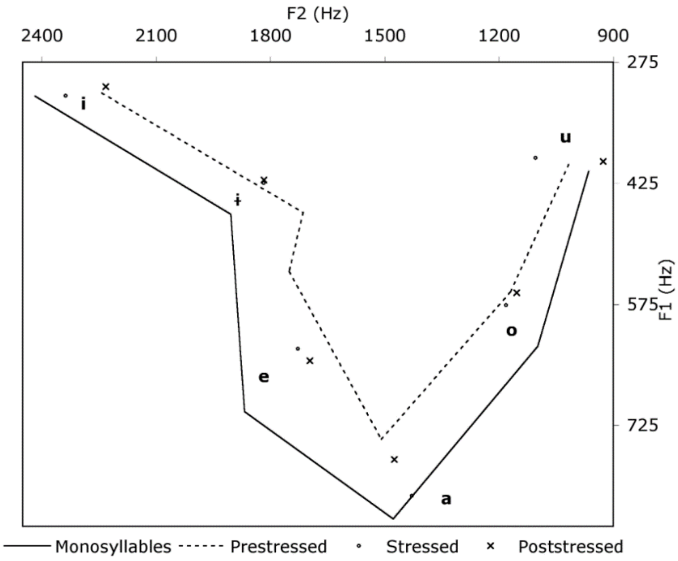


Fig. 16. Average F1 and F2 of Standard Polish vowels sorted for prosodic position (after Strycharczuk and Jurgec 2008: 4).

The authors observed that vowels in monosyllables are the most peripheral. Their results indicate that the vowel /ɨ/ is relatively front (as opposed to, for instance, Rubach's (1984) claims that it is [back]), while /e/ is low, thus the symbol [ɛ] to transcribe it appears to be, indeed, more fitting (Strycharczuk and Jurgec 2008: 5). No consistent differences were shown for pre- vs. post-stressed positions.

The more recent study mentioned before (i.e. Weckwerth and Balas 2020) investigated the productions of ten female speakers of Standard Polish (aged 21-38). Their results indicate that the vowel /ɨ/ would best be described as near-high front, rather than high central, as it is placed directly above /ɛ/ in their data. Fig. 17 illustrates the token clouds for all six Polish monophthongs, with the measurements given in raw Hertz.

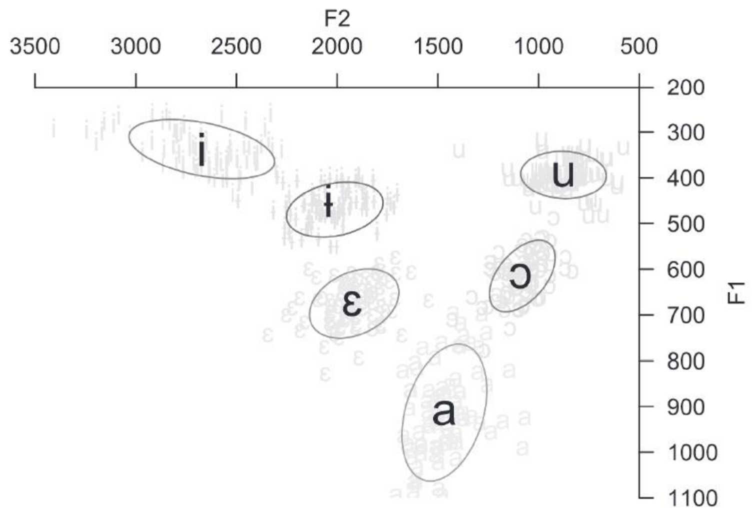


Fig. 17. Token clouds for six Polish vowels (in raw Hertz) (after Weckwerth and Balas 2020: 343)

With respect to the English vocalic inventory, we will consider two accents, Standard Southern British English and General American, as these are the two models that the English programme students who participated in the main study described in the present dissertation can choose to study. Both are considered to be good targets for foreign language teaching as they are examples of accents devoid of regional characteristics (e.g. Cruttenden 2014; Wells 1982).

The first one of the two, Standard Southern British English (also referred to as Received Pronunciation, RP, or BBC English[45]) is the one which has undergone the biggest change over time, with Lindsey (2012) noting that "the [standard] RP vowel system – still enthroned in our dictionary transcriptions after 60 glorious years – strikes contemporary Brits as pretty funny". Therefore, it is necessary for us to compare what the RP vowel system was originally thought to be and what it is nowadays. Fig. 18 represents the traditional vowel quadrilateral, with the symbols as chosen by Gimson in his initial ([1962] 2014) descriptions of this accent, translated onto the vowel chart by Roach (2004).

[45] Due to the fact that many names are used with reference to this particular accent, the name "RP" will be used henceforth.

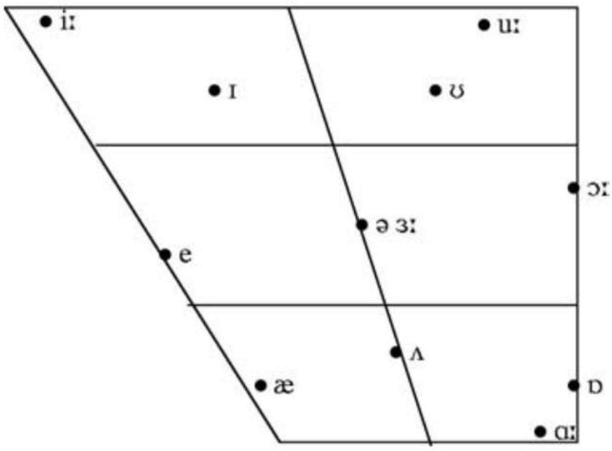


Fig. 18. Standard Southern British English (i.e. RP) vowel quadrilateral – monophthongs (after Roach 2004: 242)

As can be seen, RP has a relatively large vocalic inventory, with 12 “short” and “long” monophthongs. It must be stressed, however, that the length marks are somewhat redundant, as both traditionally long and short vowels are subject to processes such as pre-fortis clipping, which influences the length of the vowel. Therefore, it might be the case that a “short” vowel is longer than a “long” vowel in certain environments (Roach 2004: 241; Ladefoged 2001: 85)[46]. It is also worth stressing that RP is a non-rhotic accent, with both BATH-TRAP and FOOT-STRUT splits (Cruttenden 2014).

Aside from the 12 monophthongs illustrated above, there are three centring and five closing diphthongs attested in traditional RP. They are shown in Fig. 19 and Fig. 20 respectively (Roach 2004).

[46] As a result, the length diacritics will not be used when describing the vowels of English. Moreover, we will refer mostly to the Wellsian lexical sets throughout the thesis, as discussed previously in Chapter 2. The lexical sets are shown in Appendix 6.

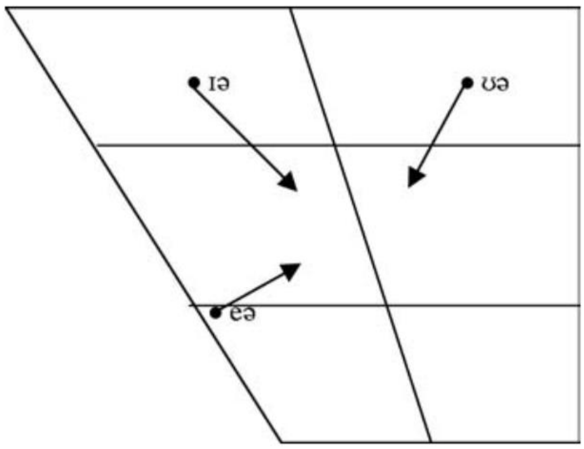


Fig. 19. Standard Southern British English (i.e. RP) vowel quadrilateral – centring diphthongs (after Roach 2004: 242)

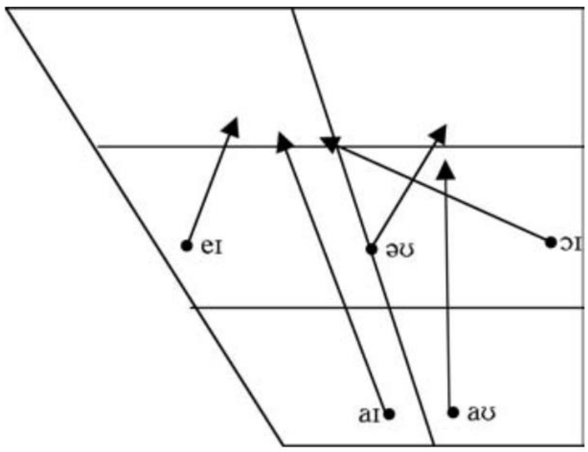


Fig. 20. Standard Southern British English (i.e. RP) vowel quadrilateral – closing diphthongs (after Roach 2004: 242)

As noted by phoneticians in recent years, the vowels – both monophthongs and diphthongs – have changed considerably in the speech of younger speakers of the accent (Lindsey 2019; Cruttenden 2014). First of all, the FLEECE /i/ vowel is described as a potential diphthong (Collins and Mees 2013; Lindsey 2019), given its diphthongal quality, in particular in word-final positions (unless subject to pre-fortis clipping, where it remains relatively pure). Before final /l/, both FLEECE and GOOSE /u/ develop a centring offglide (e.g. in *feel* [fiəɫ] and *rule* [ruəɫ]). Moreover, the GOOSE vowel is pronounced as relatively central (Collins and Mees 2013), oftentimes transcribed as /ʉ/ (Lindsey 2019). Similarly, the FOOT vowel has been claimed to become more centralised, with Lindsey (2019: 27) describing it as a close-mid central rounded vowel /ɵ/. Both GOOSE and FOOT retain their back qualities only when they precede a velarized [ɫ]. Furthermore, the TRAP /æ/ vowel has got significantly lower, resembling the quality of the Cardinal Vowel #4 and has been transcribed as /a/ (Cruttenden 2014: 97; Lindsey 2019: 19). The centring SQARE /eə/ diphthong has been monophthongised and nowadays /ɛː/ is used to denote it (Lindsey 2019: 19). In general, centring diphthongs appear to be diminishing; the NEAR vowel is often produced with a monophthongal quality of the KIT vowel (albeit slightly longer) whereas CURE is being replaced by /ɔː/ or /ɵː/ (Lindsey 2019: 48f.). However, the process of centring diphthongs smoothing is most stable for SQUARE, which is why it is the only vowel which has been universally recognised as a monophthong (Collins and Mees 2013).

Given the recent changes, Cruttenden (2014: 330) provided an illustration of the vowel chart which consolidates the acceptable areas within which every monophthong in RP might be released by L2 learners. It is given in Fig. 21.

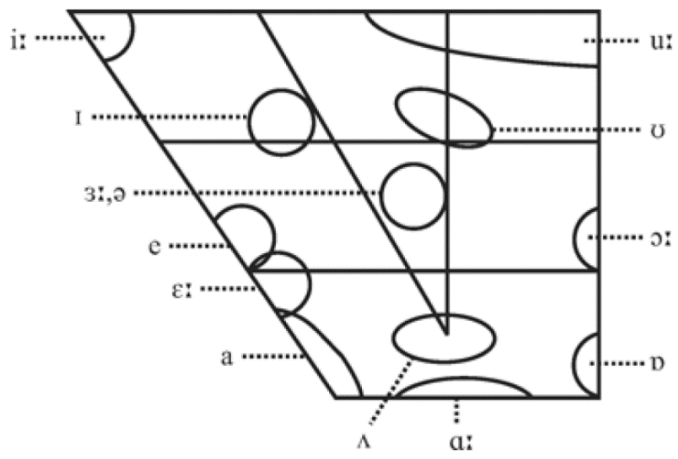


Fig. 21. Acceptable areas of the production of Standard Southern British English monophthongs (after Cruttenden 2014: 330).

All these shifts – and acceptable production areas – must be taken into account because as the model accents change, the pronunciation training syllabi follow. Therefore, when we compare L2 productions to the prototypical vowel quadrilaterals of the target, it is of utmost importance that we follow the latest and most current data available.

The second model accent of interest is General American. There are fewer monophthings and diphthongs in this accent, compared to RP; it lacks centring diphthongs, while the vowels LOT-PALM have merged (Wells 1982: 473). It is also a rhotic accent, which means that some r-colouring can be observable on vowel quality in certain environments (e.g. obliterating the quality distinction between FLEECE and KIT in pre-/r/ contexts; Carley and Mees 2020: 133) and there is not BATH-TRAP split. The monophthongs of General American are shown in Fig. 22 (after Wells 1982: 486).

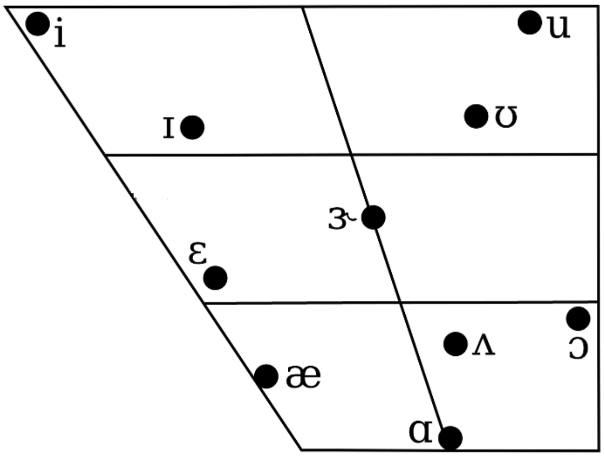


Fig. 22. General American vowel quadrilateral – monophthongs (after Wells 1982: 486)

In the most recent descriptions of General American, the THOUGHT /ɔ/ vowel is treated as optional, in contexts where other speakers would opt for PALM /ɑ/ (a phenomenon also referred to as a lack of the COT/CAUGHT split) (Carley and Mees 2020: 125). Similarly to RP, FLEECE /i/ and GOOSE /u/ are subject to breaking when they precede the word-final /l/ and are released with a schwa-like off-glide (Carley and Mees 2020: 135). While General American does not exhibit centring diphthongs, the TRAP vowel (overall much higher than in RP) undergoes the process of tensing in pre-nasal and pre-velar contexts and is pronounced as [eə] in words such as *can* or *bag* (Mielke et al. 2017), resembling traditional quality of SQUARE. While the DRESS vowel is represented by means of a different symbol in comparison to RP (i.e. /ɛ/ in General American, /e/ in RP), the difference in its openness is not that striking (Wells 1982: 485).

There are five closing diphthongs in General American, which more or less correspond to what we saw in RP, illustrated in Fig. 23. The most salient difference is the starting quality of the GOAT diphthong, transcribed as /oʊ/; the more British-like pronunciation, /əʊ/, is becoming increasingly more popular amongst younger American speakers (Collins and Mees 2013: 159). There are no centring diphthongs; they are realised as sequences of /ir/, /ɛr/, and /ur/ instead (Carley and Mees 2020: 133)

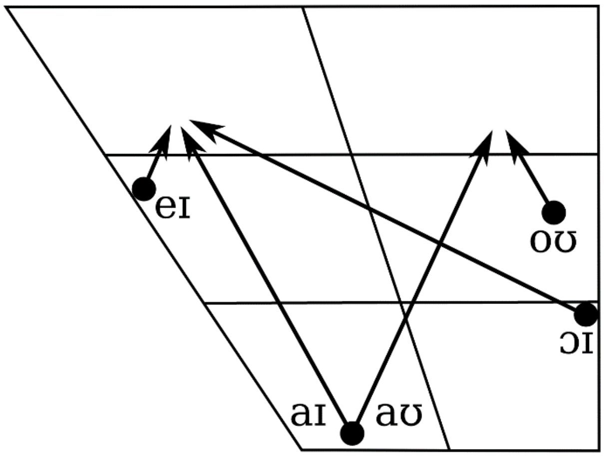


Fig. 23. General American vowel quadrilateral – closing diphthongs (after Wells 1982: 486)

Having discussed the vowel inventories of Polish and English, let us have a look at phonetic studies investigating the success with which Polish learners acquire English vowels.

4.2.1. The acquisition of vowel contrasts by Polish learners of English: an overview of selected phonetic studies

In general, Polish learners of English have been shown to have difficulties with successful acquisition of vocalic contrasts in their L2 English. This might result from the fact that the English vowel chart is relatively crowded in comparison to Polish vocalic inventory. This section provides an overview of selected phonetic studies which assessed the acquisition of various English vowel contrasts by Polish speakers[47].

First, let us look at production. Rojczyk (2010a) investigated the pronunciation of the TRAP vowel by 43 Polish students, enrolled in the English programme at the University of Silesia. Previous research has shown that this particular vowel has a tendency to be assimilated to either Polish /ɛ/ or /a/, and the direction thereof might stem from spelling influences (Gonet et al. 2010) or personal preferences (Sobkowiak 2003). The target words in a [bVt] frame were couched in a carrier sentence and each was repeated twice.

[47] The studies discussed here focus of static measurements of formant values. A series of experiments on dynamic aspects of vowels was conducted by Schwartz et al., and these will be discussed in section 4.3.3, after the concept of vowel inherent spectral changes has been introduced, as it is tightly linked to OP's assumptions.

The L2 data compiled from the participants were compared to their Polish productions. It was observed that the TRAP vowel was overall realised with a partial assimilation with their Polish /a/, while its distance to /ɛ/ was quite substantial. An additional finding was that the realisations of TRAP overlap with STRUT – both are assimilated to Polish /a/ and are distinguished by means of temporal rather than spectral cues (with TRAP being significantly longer). This goes in line with the presence of a universal phonetic bias towards turning to temporal cues when the spectral differences are not sufficient (Bohn 2020; see section 1.2.1).

The realisation of TRAP, alongside STRUT and DRESS, was also studied by Weckwerth (2011). He examined the productions of 106 Polish first year students of English at a Polish university, but did not compare them to the participants' Polish productions to examine possible assimilation effects. He found that while both STRUT and DRESS were well-separated, the realisation of the TRAP vowel was subject to a lot of variability. While some tokens revealed a formation of a TRAP category, a lot of realisations were distributed within the DRESS category and within the STRUT category. A fourth possibility was the so-called bi-modal system, wherein the TRAP instances were realised within both DRESS and STRUT and only one third of them fell outside of the scope of those two vowels. His results, then, point to a lack of stability of the TRAP vowel in the L2 productions of Polish learners of English.

In turn, the vowels KIT and DRESS in the productions of 43 third year students of the English programme were studied by Rojczyk (2010b). The vowels were embedded in [bVt] frames and were recorded alongside Polish productions. While a separate category for the KIT vowel was found in the students' productions, located between Polish /i/ and /ɨ/, the DRESS vowel was classified as equivalent to Polish /ɛ/ and no distinction between the two was observed.

Bogacka [Balas] et al. (2006) tested the pronunciation of the COMMA (schwa) vowel in the productions of 13 pre-intermediate learners of English. They found that the realisation of the vowel is influenced by context; in initial and final positions it was assimilated to Polish /ɛ/ whereas in word-medial positions it was realised as something akin to Polish /ɨ/. There was also a lot of individual variation.

When it comes to perception studies, Balas (2018) examined English vowel perception by 35 advanced Polish learners acquiring their L2 in a formal setting. In the categorisation task, she found salient effects if L1 phonology on L2 vowel perception; in velar frames, the KIT vowel was categorised as /i/ in 80% of the cases. This stems from the fact that /ɨ/ is disallowed from occurring after velar stops, showing that perception of non-native contrasts might be influenced by phonotactics. It was shown that English low vowels posed a problem for Polish learners, with categorising them as Polish /a/ but with different goodness ratings. They had also difficulties with correctly perceiving the fronted realisations of GOOSE and FOOT in alveolar contexts. Interestingly, Balas's study did not yield support to Natural Referent Vowel hypothesis (discussed in section 1.2.1). It appears that advanced L2 learners, with substantial experience, do not adhere to this principle, in contrast to naïve listeners and speakers with low proficiency in their L2.

Rojczyk (2010a) conducted a perception experiment which supplemented his production study and tested the TRAP/STRUT contrast on 17 advanced learners of English. An identification task revealed that the participants by and large failed to perceive the spectral differences between these two vowels, instead relying heavily on duration, which as you may recall is a universal phonetic bias in L2 research (cf. Bohn 2020; section 1.2.1).

Kaźmierski (2009) studied the perception of FLEECE-KIT contrast by Polish high school students with limited experience with English. This pair notoriously poses major problems for Polish speakers, even though Polish, too, has two high front vowels, namely /i/ and /ɨ/. In a L1-assimilation test, Kaźmierski found ceiling effects, with his subjects consistently differentiating between the two vocalic categories. Therefore, while Polish speakers might show lack of success in the production of this particular contrasts, they appear to correctly perceive it.

Table 14 summarises the results of the studies described above. While in general we may observe a lot of individual variation, it appears that sometimes Polish learners of English base the contrasts on temporal rather than spectral cues; furthermore, DRESS-/ɛ/ and TRAP-STRUT-/a/ seem to be the contrast that are the most difficult ones for the learners to acquire.

Table 14. Summary of the results of the studies investigating the acquisition of L2 English vowels by Polish learners (in both production and perception).

Type	Stimuli	Task	Results	Source
production	TRAP STRUT	word-list	partial assimilation to Polish /a/ rather than /ɛ/; the contrast between TRAP and STRUT based on temporal cues;	Rojczyk (2010a)
	TRAP DRESS STRUT	word-lists; read text	STRUT and DRESS produced as separate categories; a lot of variation for TRAP	Weckwerth (2011)
	COMMA (schwa)	word-list	a lot of individual variation and context-dependent variability – assimilation to Polish /ɨ/ or /ɛ/	Bogacka [Balas] et al. (2006)
	KIT DRESS	word-lists	a new category formed for KIT, but complete assimilation of DRESS to Polish /ɛ/	Rojczyk (2010b)
perception	English monophthongs	categorical discrimination; L1 assimilation; (dis-)similarities rating	low vowels and vowels fronted in alveolar contexts most problematic; influences of L1 phonotactics; no effects of NRV	Balas (2018)
	TRAP STRUT	self-paced identification	temporal cues more important than spectral cues	Rojczyk (2010a)
	FLEECE KIT	L1-assimilation	Successful discrimination of the vowel contrast	Kaźmierski (2009)

4.3. The phonology of vowel systems from the perspective of major theories

The following sections will zoom in on the phonological representations of vowels (in particular of Polish and English) from the perspective of three major phonological theories – traditional feature theory, Element Theory, and Onset Prominence. The descriptions of vowel representations proposed by all those theories will be followed by a discussion on whether any implications for L2 speech can be observed.

4.3.1. Distinctive features and the SPE model

Seeing as a brief introduction to the distinctive feature theory (Jakobson et al. 1951) and the Sound Pattern of English (SPE; Chomsky and Halle 1968) and their major tenets has already been given in section 3.3.1, we will move directly to what can be found in those works as far as vowels are concerned.

Recall that in Jakobsonian tradition, distinctive features were defined in terms of the acoustics. The natural class of vowels was described using the feature [vocalic], and subsequently more oppositions were introduced in order to differentiate between the speech sounds belonging to that class. The first of this opposition, [compact] vs. [diffuse] (Jakobson et al. 1951: 27f.), was part of a bigger group of resonance features and as such was based upon the number of dominating formants in the spectrogram picture of a given sound. Vowels described as [compact] had one main formant visible, whereas the ones described by means of [diffuse] were characterised as more than one dominating or a few non-central ones. In other words, the closer F1 and F2, the more compact the vowel. This made it possible to distinguish such pairs as /e – iː/, /æ – ʌ/, /ɒ – ʊ/.

Another feature opposition – tonality features – was based on the second formant (Jakobson et al. 1951: 30). The closer the second formant was to the first one, the more grave the sound was; if the second formant was closer to the second or higher formants, the sound was described as [acute]. From the articulatory perspective, [acute] sounds – i.e. front vowels, characterised by a large distance between F1 and F2 – are produced with less cavity space. In turn, [grave] vowels – i.e. back vowels, wherein F1 and F2 are close – have larger articulatory space. The second formant, and most importantly its apparent interaction with lip rounding, served as grounds for another feature opposition: [flat] vs. [plain] (Jakobson et al. 1951: 31). If the second formant was lowered due to labialisation, the vowel was described as [flat]. Vowels which were articulated with spread lips would be in turn ascribed the feature [plain].

The proposed set of features introduced by Jakobson and colleagues has never truly made it into mainstream phonology but, as has already been mentioned, heavily influenced the creation of SPE, even if the latter moved significantly away from the tenets found in Jakobson et al.'s work.

The features describing vowels in SPE had their basis in articulation and were plus- or minus-valued (therefore they indicated both a presence and an absence of a given feature, and the absence was implied to be phonologically relevant). The two-dimensional vowel chart was encapsulated by means of [high] vs. [low] and [front] vs. [back]. They were concerned mostly with the position of the tongue in the process of the vowel's articulation. A curious ambiguity arises when we focus on the definitions of features [±high] and [±low] (Chomsky and Halle 1968: 304f.). A [+high] vowel is a vowel during which production the tongue raises from its neutral position; conversely a [+low] sound is a vowel which is produced with a tongue lower than its neutral placement. However, when the position of the tongue remains unchanged, the sound could be described as both [-high] and [-low]. With respect to tongue advancement, [+back] was used when the tongue retracted from its neutral position, while [-back] entailed no changes in tongue retraction (Chomsky and Halle 1968: 305). Features [+rounded] and [-rounded] were relatively straight-forward in their definitions; [+rounded] entailed labialisation whereas [-rounded] did not (Chomsky and Halle 1968: 309). Finally, a [±tense] binary opposition covered the distinction between short ([-tense]) and long ([+tense]) vowels (Chomsky and Halle 1968: 325).

It has been shown that the feature inventory proposed by SPE is not able to account for all attested vowel systems of the world. An example comes from Vietnamese (Sampson 1970: 598); features [high] and [low] allow us to account for only three contrasts, while Vietnamese has a four-way contrast in its inventory of front vowels. Another issue is also the linearly-ordered feature matrices and how well they represent diphthongs. Sampson (1970: 601) points out that, "in a diphthong or a triphthong, the feature-values of vocalicness are separately specified for each vowel-segment recognized (...) [b]ut if the systematic phonetic output is supposed to model the commands given to the speech apparatus, this is surely unrealistic" (Sampson 1970: 601). Indeed, the transition from one quality to the other in a diphthong is so smooth, it is difficult to imagine treating them as separate segments, and assigning them features in a linear sequential order does just that.

4.3.2. Element Theory

Elements were introduced by Kaye, Lowenstamm, and Vergnaud in 1985. Their main goal was to resolve problems with which the feature theory was faced (Backley 2012: 58).

The definition of elements reads that they are phonological "autonomous, independently pronounceable" phonological units (Kaye et al. 1985: 306) which combine together to make up a speech segment. We have already briefly touched upon the issues associated with the choice of units used in phonological representation (cf. section 3.3.2), thus drawing attention to the fact that elements are different from features. Interestingly, elements were initially defined in terms of distinctive features, essentially being amalgams thereof (Kaye et al. 1985: 306). The three main elements were represented as shown in (1):

(1) $$\text{I}=\begin{pmatrix}-round\\-back\\+high\\-ATR\\-low\end{pmatrix} \qquad \text{U}=\begin{pmatrix}+round\\+back\\+high\\-ATR\\-low\end{pmatrix} \qquad \text{A}=\begin{pmatrix}-round\\+back\\-high\\-ATR\\+low\end{pmatrix}$$

As can be observed, the feature matrices are reminiscent of the SPE tradition; however, instead of corresponding to segments, they spell out elements. By using features to define elements, the link between phonological representation and phonetics (be it acoustic or articulatory, the latter one characteristic of the SPE tradition) was preserved. As the theory developed, the elements moved away from being defined in terms of phonetics and became conceived of as cognitive units placed on the level of phonology (Backley 2011: 61). As a result the boundary between representation and physical realisation became more noticeable. The physical reality of elements lies in the acoustics; however, there is a relative freedom of articulation of particular segments so long as the acoustic patterns associated with modelling the representation of a given element into speech signal remain intact (Backley 2012: 60).

The proposed inventory of elements tends to vary[48]; the main elements (i.e. {A, I, U}, already mentioned in (1)) correspond to the three peripheral vowels – namely /a, i, u/ – and as pronounceable as such (Harris 1994: 97). The elements are free to merge together, thus creating additional vowel possibilities; this is shown in (2) (after Harris 1994: 97).

(2)	Simplex	Compound:
	{A} = /a/	{A, I} = /e/
	{I} = /i/	{A, U} = /o/
	{U} = /u/	{U, I} = /y/

[48] As the chapter is concerned with vowels only, we will limit our discussion to elements associated solely with vowels. For a more thorough discussion regarding the inventory of elements, cf. e.g. Harris (1994), Gussmann (2007), or Backley (2012).

The fact that the elements can merge into compounds entails the existence of a dependency relation between them (Harris 1994: 105). In other words, one of the elements in a compound is the head (underlined in the representations) while the other – a dependent (Harris 1994: 105). It is said that “heads make a greater contribution to an overall expression than do dependents, both acoustically and phonologically” (Backley 2012: 64). It is thanks to headedness that we are able to distinguish between /o/ (defined in terms of {U, A}) and /ɒ/ (defined as {U, A}). As noted by Backley (2012: 65), some variants of Element Theory actually allow some compounds to be non-headed, which in turn results in creating more expressions. This is shown in (3) on the example of back vowels /o, ɒ, ɔ/ (taken from Backley 2012: 65).

(3)	Obligatory headedness:	Non-obligatory headedness:
	{U, A} = either [o] or [ɔ]	{U, A} = [o]
	{U, A} = either [ɔ] or [ɒ]	{U, A} = [ɒ]
		{U, A} = [ɔ]

The question whether or not non-headed expressions are allowed remains unsolved[49]. Backley (2012: 65) points out that some claims are made that instead of using non-headed expressions, a neutral headed element {@} should be chosen, with Harris describing it as “a blank canvas” (1994: 109). This element is also used to define the COMMA (or, schwa) vowel /ə/, as it is the most neutral vowel and contributes to representing vowel reduction processes in which it replaces the dependents in compound expressions (Harris 1994: 111).

On the basis of what has been discussed so far we can notice that elements are unary. Similarly to what was the case with representing laryngeal contrasts, the choice of unary features entails the issue of markedness. That is, for example by choosing the element {I} to represent front vowels we make the decision that [+front] is somewhat more marked than [+back] and, as a corollary, backness is not phonologically active (Backley 2012: 60). These choices are made on the basis of cross-linguistic observations of the behaviour of speech sounds (Backley 2012: 61), however – as we have seen before – they are not always entirely correct (cf. Vaux and Samuels 2005).

[49] As the issue of headedness as such falls beyond the scope of the present paper, it will not be discussed further in much detail.

In contrast to SPE-based representations, the phonological make-up of vowels within Element Theory is more complex, even though its inventory itself is small. As the present thesis is focusing on Polish and English, we will now consider the representations of these two languages' vocalic systems. Let us start with Polish, given in (4) (based on Harris 1994; Backley 2012; and Gussmann 2007).

(4) {$\underline{\text{A}}$} = /a/
{A, $\underline{\text{I}}$} = /ɛ/
{$\underline{\text{I}}$} = /i/
{U, A} = /ɔ/
{$\underline{\text{U}}$} = /u/
{I} = /ɨ/

In order to account for Polish system, the representations shown in (4) allow for non-obligatory headedness. This is to distinguish between /o/ and /ɔ/ (Backley 2012: 65) and between /i/ and /ɨ/. The former is of less importance, as these two vowels are not contrastive in Polish, so in general – if we assume that representations and their phonetic reality are kept separate – we could easily adopt {$\underline{\text{U}}$, A} instead. In the latter case, however, it is necessary for us to be able to distinguish /i/ from /ɨ/; Gussmann (2007: 55) claims that the differences lies in headedness, with /i/ being headed and /ɨ/ non-headed.

Turning to English, the representations of its vocalic system are given in (5) (based on Harris 1994).

(5) {$\underline{\text{I}}$} = /i/ (FLEECE)
{I, $\underline{\text{@}}$} = /ɪ/ (KIT)
{A, $\underline{\text{U}}$} = /ɔ/ (THOUGHT)
{$\underline{\text{A}}$, U} = /ɒ/ (LOT)
{$\underline{\text{A}}$} = /ɑ/ (PALM)
{$\underline{\text{U}}$} = /u/ (GOOSE)
{U, $\underline{\text{@}}$} = /ʊ/ (FOOT)
{A, I, $\underline{\text{@}}$} = /e/ (DRESS)
{I, A} = /æ/ (TRAP)

{<u>@</u>} = /ʌ/ (STRUT)

{@} = /ə/ (COMMA)

In the case of traditionally “long” vowels (but see section 4.2), including diphthongs, the elements are associated with two skeletal positions (Harris 1994: 103).

While Backley (2012: 65) notes that Element Theory and Government Phonology are generally intertwined, which can be noticed when discussing the issues of headedness or non-linear representations (e.g. Harris 1994, where the elements are linked to skeletal positions via association lines), he stresses that elements as such can be freely used by other theoretical frameworks.

4.3.3. Onset Prominence

Onset Prominence (Schwartz 2016 *et seq.*) has already been introduced in our discussion on laryngeal phonology in Chapter 3. As you may recall, this framework makes use of a universal CV sequence and its representations are connected with the phonetic events associated with the production of a stop+vowel sequence. This most basic hierarchical CV structure from which ‘segmental’ configurations might be extracted is shown once again in Fig. 24.

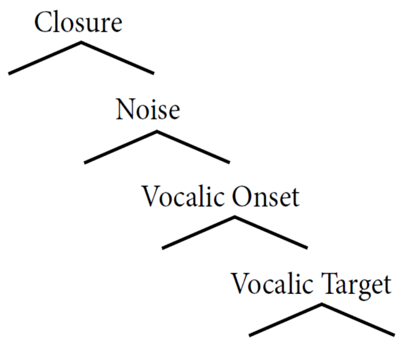


Fig. 24. The universal CV sequence in OP (after Schwartz 2016: 41)

While it is the two bottom layers of the hierarchy that are associated with the production of the vowel, the VO node is somewhat special. Namely, it is both characterised by periodicity and formant structure, characteristic of what we see in vowels, but it also bears important acoustic information about the consonant that precedes it (Schwartz 2016: 46). Therefore languages are tasked with the decision of whether the VO node is included in their vocalic representations or if its affiliation remains consonantal (Schwartz 2016: 46). Therefore the CV sequence can be parsed in two distinct ways, illustrated in Fig. 25 (after Schwartz and Kaźmierski 2019: 6).

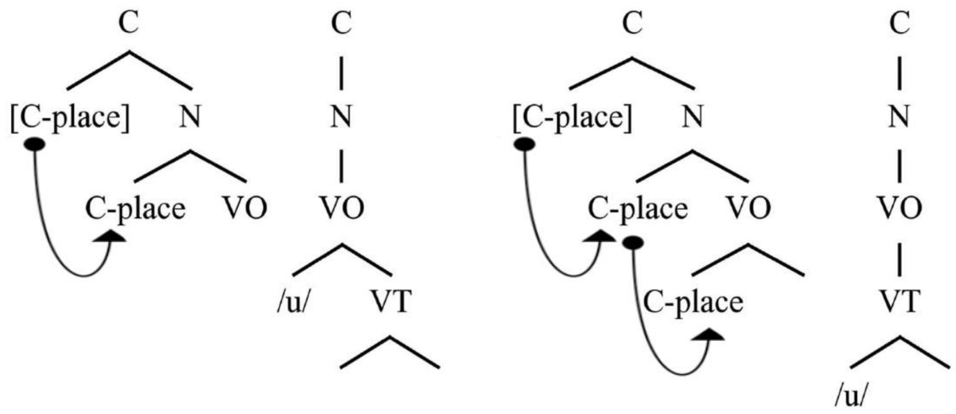


Fig. 25. Two possible parses of the CV sequence in OP: with vocalic (left-most pair) and consonantal (right-most pair) VO affiliation (after Schwartz and Kaźmierski 2019: 6)

With regards to the VO, it is claimed that it is a phonological parameter, present in the representations (Schwartz 2021), as its affiliation constitutes categorical differences between languages. In both of these possibilities, the place of articulation of the consonant is assigned at the Closure level and trickles down onto the Noise node; when the VO is part of the vowel representation (left-most trees), the vowel quality is assigned at this level, therefore blocking the process of trickling of place specifications. This results in purer vowel quality and less formant movements in the initial portion of the vowel, associated with the co-articulatory effects in the CV sequence production. On the other hand, when the VO affiliation is consonantal (right-most trees), the place specifications trickle onto both the Noise and the VO nodes. The vowel quality is assigned only at the VT level. This entails more salient formant movements (e.g. potential diphthongisation) and more robust consonantal effects on the initial portion of the vowel.

As a result, a very explicit claim is made about the possibility of cross-linguistic differences in vowel production. The proposed representations of vowels in Polish and English spawned a number of interesting hypotheses about L2 vowel acquisition (with regards to both perception and production) by Polish learners, investigated by Schwartz and his team in one of his projects. The research focused on the principle of Vowel Inherent Spectral Change (henceforth: VISC; e.g. Nearey and Assmann 1986; Morrison and Assmann 2013). Putting it simply, using two-dimensional charts to describe and compare vowels might not be sufficient, as the dynamic characteristics of vowels – e.g. vowel internal formant movement – contribute to important clues about the behaviour of vocalic systems, alongside their traditional descriptions (Schwartz 2021: 2). VISC has been attested for English (e.g. Fox and Jacewicz 2009), but with the relative paucity of data from other languages, Schwartz's project attempted to fill that gap and assess to what extent VISC is language-specific or universal, and whether or not it has any phonological implications. OP assumes that formant movement should be greater in English due to VO's consonantal association. In Polish, wherein VO is associated with vowels, VISC should be observable to a lesser degree, with the quality of the vowel being more stable. In general, their findings revealed that as proficiency increases, the formant movements in Polish learners' productions change – advanced speakers tend to show more dramatic movement patterns relative to those whose proficiency oscillates within the intermediate level, thus the former group is closer to native-like targets (Schwartz et al. 2016). The crucial portion of the vowel which requires some temporal reorganisation is the 20%-40% in the vowel duration (Schwartz and Kaźmierski 2021). Some effects of VISC on vowel perception have also been attested (Schwartz and Dzierla 2018).

It is worth noting that in the representations shown in Fig. 25 attached to the VO or VT nodes (depending on the structure which hinges upon the choice of a language) is a segmental symbol of /u/. However, as has been stated, segments are derivative from the hierarchical structures and as such as not units employed by OP analyses. Therefore the 'segments' attached to the trees might be conceived of as shorthand for labial specifications (Schwartz 2017). In general, while the choice of elements as basic units denoting vowel quality appears to be favoured, featural specifications in themselves are irrelevant to the framework.

Returning to the main issue, namely the possible predictions that vowel representations might have with respect to phonetic drift, assuming that it is the principle of equivalence classification that lies at the core of possible L1 phonetic drift effects, the representations postulated by OP would let us to predict two scenarios, hinging upon which level the cross-linguistic similarity which leads to cross-linguistic influence targets: VO or VT. If it targets the VO level, that is the initial portion of the vowel associated with formant movement, very little phonetic drift in the productions of Polish learners of English. The parses of the CV sequence are different in both languages so the learners would be expected to acquire the new structures, and subsequently form new categories, with no effects exerted on their L1. Notice, however, that both Polish and English vowels are phonologically identical at one level, that is VT. If it is the VT level that is targeted by equivalence classification, some L2 influence on L1 vowels might be observed with respect to F1 and F2, that is vowel quality.

4.4. Summary

The previous sections have discussed the representations of vowels within three prominent phonological theories: SPE, Element Theory, and Onset Prominence. All of these theories have been previously discussed to account for laryngeal systems of Polish and English (cf. Chapter 3)[50].

In general, no claims when it comes to the relative success or lack thereof in the process of L2 vowel acquisition is offered by the SPE and Element Theory frameworks. These two theories appear not to be focused on exploring cross-linguistic similarities and differences, which results in them not contrasting languages with one another but rather describing the sound inventories of various languages. From the perspective of the SLM model, only Onset Prominence offers testable hypotheses with regards to both L2 acquisition and L2-induced phonetic drift, as its representations indicate cross linguistic differences between languages. These predictions will be assessed in the empirical study that follows.

[50] Recall that Harris' (1994) representations of two-way laryngeal contrast, where he used elements {H} and {L}, was in line with the theory of Laryngeal Realism.

Chapter 5: Empirical studies: research questions, methods, and analyses

5.1. The aim of the chapter

The previous chapters have dealt with the theoretical basis of the research undertaken in the present thesis: they have described the main model of second language acquisition to which we adhere herein (cf. Chapter 1), provided empirical data regarding the research problem – that is, L1 phonetic drift – from other languages (cf. Chapter 2), and finally discussed phonological theories aimed at predicting and accounting for the results of the studies (cf. Chapter 3 and 4).

This chapter discusses at length the methodology of the longitudinal (in both real and apparent time) acoustic study whose main aim was to investigate the L2-induced phonetic drift effects in the productions of Polish advanced learners of English.

With "phonetic drift" being our main construct of interest, its effects are operationalised as changes in the phonetic realisation of L1 Polish stops and vowels under the influence of L2 English.

With respect to consonants, the main parameter of interest VOT (negative for voiced stops, positive for voiceless stops), f_0 (i.e. pitch or fundamental frequency) and F1 at the onset of the vowel following the plosive. As far as vowels are concerned, the phonetic drift effects pertain to changes in formant values: F1 (i.e. vowel height) and F2 (i.e. vowel advancement).

At the same time, we also study the trajectory of changes in our participants' English productions which result from the training that they were undergoing. These are operationalised in the same way as was the case for Polish: we are focusing on the VOT of stop consonants as well as changes in the vowels' formant values.

The explanation for the choice of the abovementioned acoustic parameters will be provided in section 5.9, where the details regarding acoustic annotation will be discussed in more depth as well.

The following sections enumerate the research questions and hypotheses, describe the materials which were used and provide information about the study groups as well as discuss the procedure, the process of acoustic annotation, and the statistical analyses.

5.2. Research questions and hypotheses

Our previous considerations, including phonetic studies conducted on L1 phonetic drift Polish (Schwartz 2020), have led us to the formulation of research questions which the present dissertation attempts to answer. These are given below.

(1) Will L1 Polish students majoring in English display L2 pronunciation training-induced phonetic drift in their Polish productions?
(2) Will the possible phonetic drift effects be different depending on the year of study? Will the degree of phonetic drift be less observable, seeing as the participants are advanced, rather than novice, learners?
(3) Will the phonetic drift effects be contingent on the principle of equivalence classification?
(4) Will the success with which English laryngeal contrasts as well as vowels are acquired influence the amount of phonetic drift observed?

Previous phonetic studies, presented in Chapter 2, which dealt with other language pairs have allowed us to formulate some hypotheses related to the research questions.

For Research Question #1, the following hypothesis is postulated: Polish learners of English undergoing phonetic training in L2 are going to display phonetic drift effects in their L1 productions. This is claimed bearing in mind previous research on Polish (Schwartz 2020), as well as studies conducted on similar populations and in similar contexts on other languages. Phonetic drift should be observable regardless of the fact that the learners stay in an L1-dominant environment.

When it comes to Research Question #2, it is hypothesised that the amount of phonetic drift found will depend upon the year of study. Considering that our participants undergo training during the first and second year of university, the effects should be most salient after two years of continuous pronunciation training, with less robust effects when the students no longer undergo phonetic instruction (i.e. the third year of study) and at the very onset of their pronunciation instruction[51]. It is also predicted that the fact that our sample consists of advanced learners of English, with vast L2 experience, will have little

[51] Detailed explanation of what pronunciation training that the participants underwent entails can be found in section 5.4.

bearing on the degree of phonetic drift in their L1. Even though Chang (2013) did claim that it was the "novelty effect" that was most responsible for this phenomenon to occur, other studies and the work done on Polish appear to point to the fact that phonetic drift should be noticeable also in the productions of students with higher proficiency. This might stem from the fact that University students are subject to phonetic instruction and "discover" the L2 contrasts sometimes for the first time during the phonetic course, which makes the L2 vowels and consonants novel enough for drift to take place.

As far as Research Question #3 is concerned, the effects of phonetic drift are hypothesised to depend upon the principle of equivalence classification. The effects will differ across the two categories under investigation, that is in consonants and in vowels. For consonants, as has already been mentioned, the degree of drift will hinge upon what is phonologically equivalent in the case of Polish-English laryngeal contrast. While theories differ in this respect, it has been noticed across languages that the voiced series of stops displays higher susceptibility to cross-language interactions. This is also indicated by previous research on Polish (Schwartz 2020). Such a prediction is made by the Onset Prominence representational framework; if the results obtained in the present experiment show similar asymmetry, empirical support will be yielded to the representations postulated by OP.

On the other hand, recall that equivalence classification operates slightly differently with respect to vowels (cf. section 1.1.6.2). Namely, if a new phonetic category emerges in the inventory, two scenarios are deemed plausible: either there will be no interaction between L1 and L2 vowels or the effects will be dissimilatory. That is, the L1 vowel will move away from canonical monolingual norms in order to maximise the contrast between the newly-established L2 category. Conversely, if the L1 and L2 vowels are classified as identical, a diphone that merges the L1 and L2 phones together will be produced with intermediate formant values. In general, when we compare Polish and English vocalic inventories (cf. section 4.2) and evaluate previous research on vowels (cf. section 4.2.1), it might be difficult to predict what kind of behaviour the L2 English vowels will display. Therefore, while it is difficult to formulate hypotheses for specific vowels, it is predicted that at the very onset of the pronunciation training English vowels will be assimilated to L1 categories (and might be substituted with L1 vowels in L2 productions). Over the course of the training, especially after two years, new L2 vowel categories should be formed and the overlap between L1 and L2 productions should be less observable. At this point, L1 Polish formant values of the vowels might move away from the monolingual norms.

At the same time, and this relates to Research Questions #3 and #4, alongside Polish productions, we aimed to monitor the students' progress in their L2 pronunciation. Therefore, by recording the data sets in both languages we could trace the trajectory with which they acquire laryngeal contrasts and the vocalic inventory in order to determine whether or not new phonetic categories are established and equivalence classification effects present. For Research Question #4 in particular it is interesting to see the interaction between English and Polish. We hypothesise that phonetic drift effects should be more observable in the case of L1 categories which prove to be "problematic" to students. In other words, these features which are deemed as equivalent, and hence, difficult to acquire, should be more susceptible to cross-language interaction.

The next section will detail the methodology undertaken in order to test the abovementioned hypotheses.

5.3. Methodology

It is important to stress one thing before we proceed. While the study of consonants (and any changes with respect to the acoustic parameters associated with them) had originally been envisioned to be the main focus of the present dissertation, the decision to include the vowels was made after the data have been collected. For this reason, the studies on vowels can be thought of as pilot studies, mostly descriptive and exploratory in nature.

The empirical experiments on which the dissertation reports comprise two different types of production tasks. In the first one, the participants were asked to read two word lists: one in Polish and one in English, while the second task involved reading carefully controlled sentence lists, also in both these languages. The word list task was chosen in order to try and replicate the previous studies on phonetic drift that have been reported on in Chapter 2. Experiments in both L2-immersion environment and L1-dominant environment alike employed word lists as their main datasets, with only a handful using carrier sentences. The sentence list task was chosen in order to assess whether the phonetic drift effects are present also in longer utterances. Spontaneous speech would have been ideal, however such studies entail a number of issues. Measuring both VOT and vowel quality is decidedly easier in controlled contexts. Sentence lists, while still an example of a laboratory speech sample, constitute an intermediate state between a simple word reading list and spontaneous speech.

5.4. Participants

The participants that have taken part in the present can be divided into two major groups: the main group of students taking part in the longitudinal study (henceforth Group 1) and the comparison group, comprising three smaller subgroups: 2BA (henceforth Group 2), 3BA students (henceforth Group 3)[52], as well as Polish "quasi-monolinguals" (henceforth Group 4).

The choice of the population of Polish English majors is by no means accidental. This particular group is somewhat special as their motivation to acquire native-like proficiency in English is assumed to be high – they are studying to use English professionally at work, be it teaching, translation, or other endeavours.

As will be shown, the groups are treated as homogenous groups. They were all enrolled in the same university-level courses, and while they did come from various parts of Poland, it will be evident in the report on their pre-test questionnaires than none of them claimed to speak any regional dialects. This is to be expected as regional features of speech are less common amongst educated speakers.

None of the participants were paid for taking part in the experiment, except for a subset of the participants from Group 4. Additionally, it is the Faculty's requirement for 1BA students to spend at least three hours in laboratories in order to receive credit for the year – taking part in all six recording sessions that they underwent for my experiment added up to these three required hours. The following subsections describe the characteristics of each of these groups in detail.

[52] The abbreviation "BA" here refers to the year and level of study; that means that 1BA students are first-year students of the lower Bachelor of Arts level of studies, 2BA are second-year students, and so forth.

5.4.1. Group 1: 1BA students of English

Group 1 consisted of twenty students of English who started their University education at Adam Mickiewicz University in Poznań in the same year that the longitudinal phonetic study was conducted. The original number of students that were recorded was thirty five. However, the number was subsequently reduced to twenty due to numerous reasons. First of all, the drop-out rate during the first year of studies is rather high. There were participants who did not take part in all of the recording sessions and for that reason they were excluded from the study. Moreover, there were students who reported a high L3 proficiency: two of them in French and two in German. Since they took their advanced *Matura*[53] exam in those languages, it can be assumed that their fluency in those languages oscillated at about B2 level, according to *CEFR*[54]. In order to avoid any L3-induced effects on their L1 and L2 productions, those four students were also not included in the final group of analysed speakers.

The final set of participants comprising Group 1 were all native speakers of Polish, born and raised in an L1-dominant country, whose both parents were also native speakers of Polish. The participants (N=20) were aged 19-20 (median age: 19). There were sixteen females and four males[55]. Fourteen of them were enrolled in British groups and six in American ones. Prior to taking part in the first recording session, they were asked to fill out a short questionnaire regarding their demographics and some language background data (cf. Appendix 5: Questionnaire). Importantly, none of the students had spent more than three consecutive weeks in an English-speaking country, which would entail spending a significant amount of time in an L2-dominant environment. Thirteen speakers reported briefly visiting English-speaking countries (the United Kingdom, Ireland, the

[53] *Matura* (officially: *egzamin maturalny*) in Polish educational system is a series of standardised exams sat upon the completion of high school, evaluated by independent examiners (*Centralna Komisja Egzaminacyjna – CKE*). These exams are not obligatory but must be taken in order to apply for courses in Universities. The compulsory, "basic" level includes written exams in Polish, mathematics, and a modern foreign language (of the pupils' choosing); the "advanced" level (at least one subject) also needs to be taken. The *Matura* also includes two oral exams: in Polish and in a modern foreign language. In order to be accepted into the AMU Faculty of English's full time *English programme*, the candidate must pass the advanced level Polish and English exams with very good marks (source: https://cke.gov.pl/en/egzamin-maturalny/egzamin-w-nowej-formule/ and my own experience).

[54] CEFR (Common European Framework of Reference) refers to a guideline which helps assess linguistic proficiency in SLA. The six levels, going from "beginner" to "native-like proficiency" are: A1, A2, B1, B2, C1, and C2.

[55] It was impossible to get a group with equal numbers of males and females due to the availability of participants.

United States of America, and Canada), but these visits varied in length between four days and two weeks at a time.

What is also important is that none of the participants had undergone any form of a specialised phonetic training before starting the English programme at our University. When asked about pronunciation classes in high school or private tutoring, they reported that teachers did not pay too much attention to their productions. Three participants mentioned that they did learn about the "th"-contrast in school. Furthermore, no participant had ever had any classes with a native speaker of English before, either in school or during private English classes.

The participants did not complete any sort of a standardised test which would assess their knowledge of English. It was assumed that since all of them were required to pass the advanced *Matura* exam in written and oral English in order to be accepted into the English programme, they all could be described as B2-level (i.e. upper-intermediate) speakers of English (in terms of *CEFR*). According to the data gathered in the questionnaires, they started learning English on average at the age of 6.5 years old (median age: 6). In high school, they attended on average 5.35 English lessons per week (median: 6, with one lesson lasting 45 minutes, which is standard for Polish schools). Two participants mentioned that they attended private tutoring prior to their *Matura* exams (60 minutes per week in both cases) and the tutoring was conducted by native Polish teachers of English.

When it comes to the demographic data, fifteen students originally came from the Wielkopolska region where our University is located. Two of them were from the Dolny Śląsk region, and three from Mazowsze. These details are illustrated in Fig. 26. Out of the fifteen participants from Wielkopolska, seven of them reported familiarity with the Poznań dialect and claimed using it at home. The Poznań-Kraków dialect of Polish is described as a dialect displaying *sandhi*-voicing phenomena (e.g. Gussmann 1992, Rubach 1996; Cyran 2014; Wojtkowiak and Schwartz 2018), but these should bear no effects upon the phonetic realisation of word-initial stops or vowels; the latter are described in the literature as being well-representative of Standard Polish (e.g. Weckwerth and Balas 2020). Nonetheless, when asked to elaborate on their usage of the dialect, those participants pointed to some lexical differences, such as borrowings from German, which are characteristic of Poznań speech, rather than any phonetic ones. Usage of no other dialect was reported.

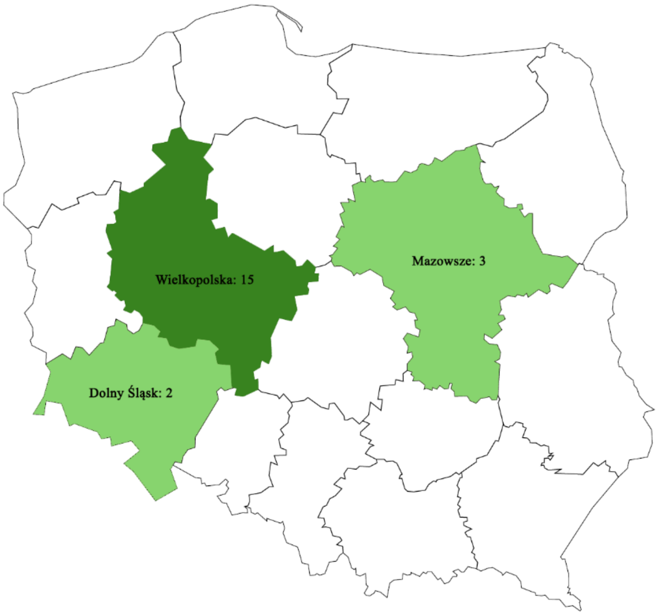


Fig. 26. The areas of origin of the twenty 1BA participants[56]

The participants were also asked about their command of other foreign languages. As has already been mentioned, four speakers with very high L3 fluency were excluded from the study. Fig. 27 illustrates other languages which were mentioned by the participants. When asked about more details, they mostly mentioned being taught them at school for varying periods of time, but they claimed not to be fluent in communication. It needs to be stressed that foreign language classes in Polish schools are concerned mostly with reading and listening comprehension, as well as with grammar and vocabulary practice and are limited, on average, to a maximum of two meetings per week (thus, ninety minutes in total). Two students reported taking their basic *Matura* exam in German and three in Spanish.

[56] The map template used with permission of Piotr Walczak; source: https://www.bajkidoczytania.pl/mapa-polski-wojewodztwa (date of access: 1 March 2019).

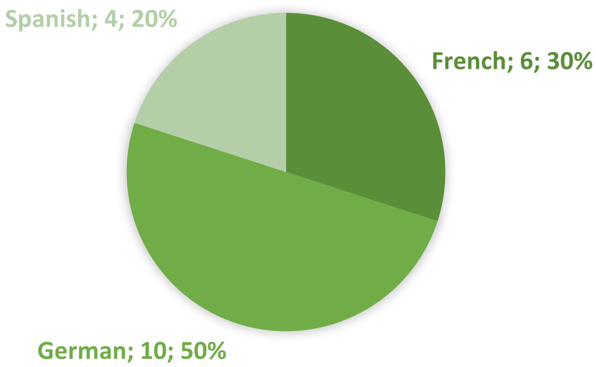


Fig. 27. Command of other languages as reported by Group 1 participants.

Finally, we can now turn to the details of their University education, with special focus on the actual phonetic training. Production training, especially when it is supplemented with regular language classes, has been shown to be very beneficial as far as the improvement of students' pronunciation is concerned (e.g. Mildner and Tomić 2007). The 1BA students of the English programme at our University attend two types of phonetic courses. The first one is *EFL Practical Phonetics* class which takes place twice a week for ninety minutes. This course is mainly focused on the successful acquisition of segmental phonetics, where the students do plenty of drilling exercises. The second one is *English Phonetics and Phonology*, which is a more theoretical course and covers a variety of issues connected with phonetics and phonology, including articulatory phonetics, acoustic phonetics, suprasegmentals (such as stress, intonation, and rhythm), as well as the key differences between Polish and English phonologies. This course includes a seminar which takes place once a week for ninety minutes, and is additionally supplemented by a weekly forty-five-minute lecture.

A question about the specifics of the training may arise at this point. As far as the consonants are concerned, we are interested especially in the laryngeal contrasts in Polish and in English. One may ask, then, how *precise* the teachers who are conducting the phonetic training are with respect to pre-voicing, lack of pre-voicing as well as the differences between plain voiceless and aspirated stops. Most of the focus is decisively placed on the latter – that is, both practical phonetics and *English Phonetics and Phonology* teachers stress the difference between aspirates and plain voiceless realisations of stop consonants. There are numerous drilling exercises that teach aspiration; additionally, the students are required to provide the teachers with short recordings of texts from textbooks on pronunciation and one of the aspects evaluated is whether or not they employ aspiration[57]. In the theoretical courses, the students are shown spectrograms that display the differences between both series of stops in Polish and English[58]. When these contrasts are demonstrated, however, the students' report that they have no problem with perceiving aspiration, but they do not hear any contrast between pre-voiced and unvoiced stops[59] (but see Schwartz and Arndt 2018).

Aside from that, they attend Practical English classes with native and non-native speakers of English as well as some standard courses concerned with English and American history, literature, and culture. All classes are conducted in English.

One last remark that needs to be made with respect to this group is that in the second semester the students choose their obligatory "foreign language" classes. Out of the twenty participants included in my analysis, 10 of them chose A2 German, 4 A1 French, and 6 A1 Spanish. Each of these foreign language courses comprised 15 weekly meetings for 90 minutes. Considering the fact that the participants all chose beginner levels and that the courses focus – similarly to the foreign language courses in high school – on reading and writing abilities as well as on expanding the vocabulary and learning grammar of a given language, they should not have any bearing on the results of the phonetic studies in which they were involved for the purposes of the present thesis.

5.4.2. Groups 2 and 3: 2BA and 3BA students of English

Group 2 was made up of fifteen students of the second year of the English programme at our University. The basic requirements for them to be able to take part in the experiment was that they had to be native speakers of Polish, whose parents were also born and raised

---

[57] Here the evaluation of whether or not the student has mastered aspiration is auditory. All of our teachers of phonetics are trained phoneticians from Poland.

[58] The teachers conducting the theoretical classes are trained phoneticians and phonologists, most of the time actively conducting research themselves (aside from one American teacher, all of them are of Polish origin).

[59] This remark is based on my own students' comments on these contrasts elicited from them in class. I have been teaching *English Phonetics and Phonology* over the course of my PhD programme to both British and American groups.

in Poland. They were aged 20-21 (median age: 20), and 11 of them were female while 4 were male. 7 of them were enrolled in British groups and 8 in American ones.

At the time of their recordings, they had completed two full years of phonetic training. While in their first year the amount of training adds up to 5.25 hours per week (3 hours of practical phonetic and 2.25 hours of theoretical phonetics), in the second year the training is limited to 90 minutes per week. The students attend the *Practical phonetics* class, where – having mastered the segmentals in the previous year – they focus on casual speech processes (including assimilation, deletion, and linking processes) as well as sentence rhythm and proper intonation. They are required to submit a number of recordings on the basis of which they are evaluated by trained phoneticians.

Similarly to Group 1, Group 2 were not asked to perform any sort of a standardised test in order to measure their command of English. It was assumed that having passed all of the required courses after their first year, they all presented a similar level of English, in accordance with the requirements set by our Faculty, that is B2/C1 (upper-intermediate/advanced)[60].

As far as the demographics are concerned, once again most of this group came from Wielkopolska (N=10), two speakers were from the Dolny Śląsk region and from the zachodniopomorskie voivodship, while one speaker came from Mazowsze. The details are given in Fig. 28.

[60] 1BA Practical English exam requirements: https://ects.amu.edu.pl/en/courses/view?prz_kod=15-PNJA-1BA-12.

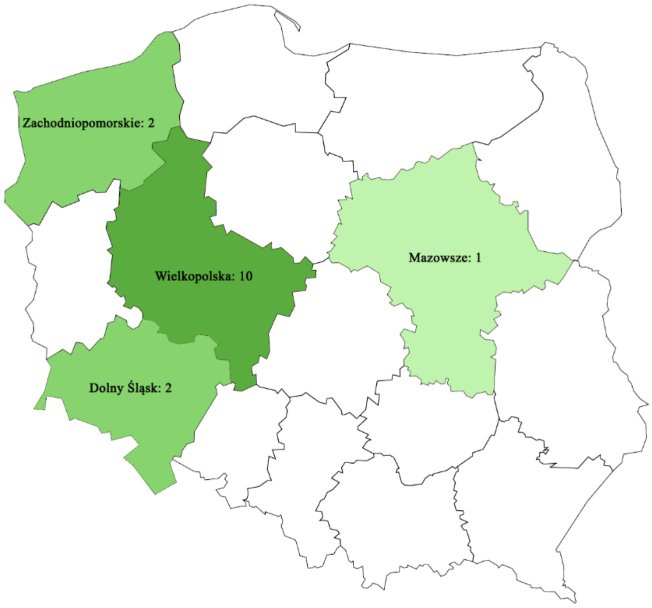


Fig. 28. The areas of origin of the fifteen 2BA participants.

Similarly to the speakers from this region who were included in Group 1, eight speakers from Group 2 reported being familiar with the Poznań dialect. They mentioned using dialectal vocabulary at home and they were familiar with the specific high tone melody associated with the dialect (with some reporting that this was observable in their speech to such an extent that their university groupmates from outside of Wielkopolska noticed it in their productions). However, this is predicted to have no bearing on the results obtained herein, as it is mostly noticeable in spontaneous speech. The author (a speaker of this particular dialect herself) did not notice Poznań-like intonation patterns in the productions of those speakers in the process of speech annotation.

On average, the speakers from Group 2 started learning English at the age of 7.10 (median age: 7) and attended 5.25 English classes per week in high school (median: 5, with one class lasting 45 minutes). They unanimously claimed that they had not had classes with native speakers of English prior to starting their university education and that correct pronunciation was not truly discussed while they were at school. One participant mentioned (similarly to what we found in Group 1's questionnaires) that their teacher told them about the "th" sound in English but it had been a one-time occurrence and the students had not been taught how to produce it, but rather had been informed that these sounds do not exist in Polish.

As for spending time in English-speaking countries, the participants from Group 2 reported that they had visited the United Kingdom and Ireland, but none of them had spent more than three consecutive weeks there.

The students included in Group 2 were also attending foreign language classes, whose nature was already discussed in Section 5.4.1. The details are given in Fig. 29.

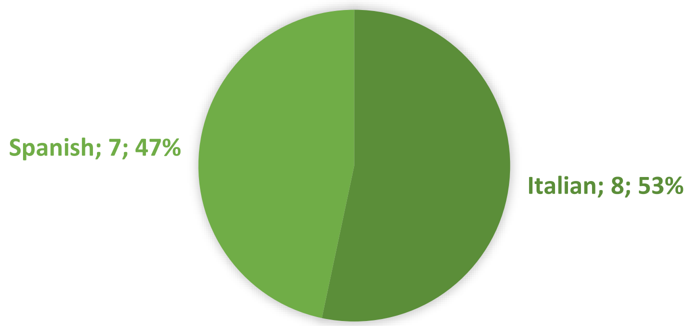


Fig. 29. The foreign language courses attended by 2BA participants.

Even though at that point their fluency in these languages oscillated between A2 (elementary) and B1 (intermediate) levels according to *CEFR*, they were not being phonetically trained in either Spanish or Italian. Furthermore, both Spanish and Italian contrast prevoiced stops with plain voiceless stops (e.g. Cho et al. 2018; Balogné-Bérces and Huszthy 2018), so from the laryngeal perspective they resemble Polish. Their vocalic inventories are rather scarce as well: there are five vowels in Spanish (Ladefoged and Johnson 2010) and seven in Italian (Rogers and d'Arcangeli 2004), which again is similar to what we see in Polish (that is, six vowels). Therefore, the chance that these courses could affect the results is relatively slim. The speakers did not report speaking any other languages aside from the ones they were learning in their classes at university.

Group 3, in turn, consisted of fifteen students of the third year of the English programme. They were all aged 21-22 (median age: 21) and 14 of them were female, with one male. 12 of them were enrolled in British groups and 3 in American ones.

At the time of the recordings, the students had completed two years of pronunciation training (the exact same one as Groups 1 and 2), with the last pronunciation class taking place exactly twelve months prior to their participation in the study. In other words, the students from this group had spent their academic year still taking all their classes in

English but without any explicit phonetic instruction. It is assumed that they had acquired all segmental and suprasegmental features of English pronunciation, as all of them had passed their Practical English exams the previous June. Moreover, as they were about to sit their third year Practical English exam within weeks of the recording session, they must have been consciously taking care of the pronunciation model to which they adhered. Nonetheless, they constituted an interesting group for comparison – they allowed us to assess how stable the new phonetic categories that they (possibly) established over the course of the training were and whether there were any changes now that their pronunciation was not evaluated on regular basis.

As far as the demographics go, the students from Group 3 came from the western part of the country, with seven of them from the zachodniopomorskie region, four from Wielkopolska, and 2 from both pomorskie and dolnośląskie voivodships. The details are shown in Fig. 30.

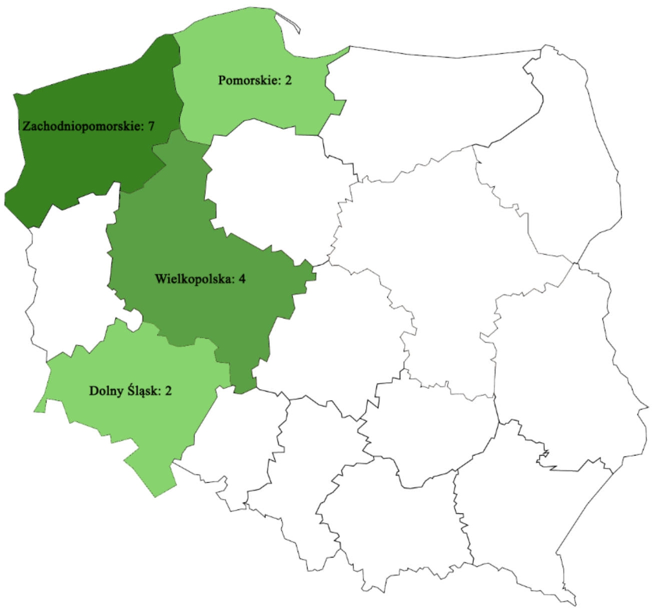


Fig. 30. The areas of origin of the fifteen 3BA participants.

While the usage of no other dialect was reported, the four participants from Wielkopolska, as was the case for Groups 1 and 2, mentioned being familiar with the Poznań dialect and two reported actively using it at their home. When asked to elaborate, the students mentioned mostly different lexicon and swapping sequences of a stop and a fricative, as in *trzy* ‘three’ [tʂɨ] and *drzewo* ‘a tree’ [dʐɛvɔ], with alveolar affricates ([t͡ʂɨ] and [d͡ʐɛvɔ] respectively), which is a common feature in this dialect (Kaźmierski et al. 2019). As a result the words *czy* 'if' and *trzy* 'three', which are not homophones in Standard Polish, would be produced identically (as [t͡ʂɨ]) by those speakers. Seeing as the realisation of affricates or consonant clusters was not of interest to the present thesis, it was judged that this feature did not constitute a problem and those participants were included in the study.

As was the case with Group 1 and 2, we did not conduct any sort of a standardised L2 evaluation on the basis of which the speakers were to be included in the analysis. It was assumed that having passed their exams the previous year, in particular their Practical English one, they all were comparatively proficient in English, at this point reaching the level of C1/C2 (advanced/near-native proficiency) according to *CEFR*.

On average, the participants from this group started learning English at 7.46 years old (median age: 7) and attended ca. 5.5 English classes per week back in high school (median: 5, with one class lasting 45 minutes). As was the case with Groups 1 and 2, these students had not had classes with native speakers at school, not had they attended any sort of private tutoring. They visited countries such as the United Kingdom, Ireland, and Canada, but had not stayed there for longer than three consecutive weeks.

During their third academic year the students from Group 3 attended one last semester (i.e. 15 classes, 90 minutes each) of their foreign language course, mentioned previously. The course ended in February, four months before the recording sessions for the purposes of the present study. By February, their proficiency in the chosen foreign language was supposed to reach the B2-level (upper-intermediate), but the students self-assessed their proficiency as lower. The languages chosen by the students in Group 3 are given in Fig. 31.

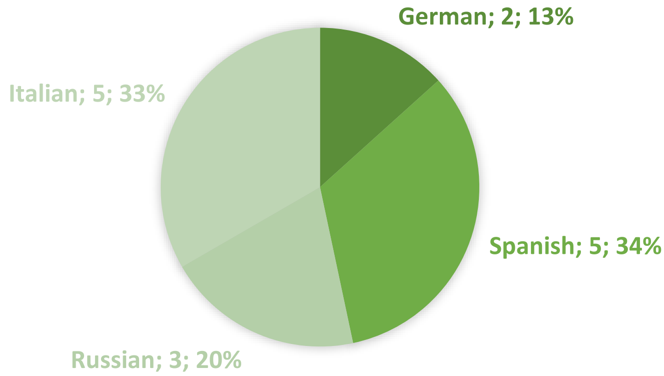


Fig. 31. The foreign language courses attended by 3BA participants.

Once again, we do not predict these foreign language courses to exert much influence on the students' performance in English or in Polish. Italian and Spanish have been mentioned already; Russian is a voicing language with a relatively small (N=6) vocalic inventory (Jones and Trofimov 1923), resembling Polish in this respect. German is an aspiration language (Iverson and Salmons 1995) with admittedly larger vocalic inventory, however, once again we must stress that the nature of the foreign language seminars at our institution focuses mostly on grammar, vocabulary, reading, and writing skills, whereas pronunciation training in those languages is extremely slight. Aside from the foreign language courses, three participants reported the knowledge of Dutch at A1 level (basic vocabulary and grammar constructions).

What is also worth an explanation is why there were only fifteen people included in both Groups 2 and 3. The reason for that is that while 1BA students are required to spend three hours per year participating in experiments run by the doctoral students and the staff of our Faculty, 2BA and 3BA students are not. As a result, it is very difficult to encourage them to take their free time and take part in various studies. Given the fact that the participants still produced a significant number of words and sentences in Polish and English, it is believed that they still constitute a valid comparison group that can tell us something about whether or not the drift effects are permanent and how stable are the L2 categories at the very end of phonetic training as well as one year after their last phonetics class.

5.4.3. Group 4: Polish quasi-monolinguals

The last group were Polish quasi-monolinguals, which served as a control group. Two things need explaining before we proceed.

First of all, notice that this control group is described as a group of quasi-monolingual participants. This is because it would be extremely difficult to find a true group of monolinguals, that is people who have never been taught any other language, as foreign language classes in Poland are obligatory, starting as early as in primary school. Major (1992: 201) argued that not fully-mastered L2 should not have much bearing on L1, not limiting to pronunciation, but also to morphosyntax, lexicon, or semantics[61]. Pavlenko and Jarvis (2002: 209) claim that it takes about three years of thorough interaction in the foreign language for the L2 to influence one's native language command (possibly hinting at language attrition effects). Chang (2018) however, draws the attention to a general overreliance on the "nativeness", that is not providing relevant language background descriptions and assuming that a person who has not fully mastered an L2, or is described as being a marginal bilingual, displays linguistic behaviour that is unaffected by the omnipresent multilingualism (2018: 109). In other words, a control group who is not fully monolingual is acceptable as long as we are explicit about their linguistic background. Bearing this in mind, our quasi-monolingual group comprised speakers who *do* have a history of learning foreign languages in school, but they claim not to be fluent in any of them, at least as far as production is concerned. They could, then, be described as "functional monolinguals", in a sense that they "have notable difficulty categorizing and discriminating many phonetic contrasts from unfamiliar languages that are not used to distinguish lexical items in their own language, including both nonnative consonants and vowels" by Best and Tyler (2007).

The second comment which is necessary is that even though all speakers are placed in one category of "the controls", there are, in fact, two control groups. There are two reasons for that: one is more pragmatic, the other has theoretical merits. The former is that these speakers were recorded for two different projects. The latter is that one of these groups read a Polish word list while the other one read a Polish sentence list. Therefore, word lists produced by the Students will be compared to the productions of the

[61] But see Chang (2012); it was a major observation that changes in pronunciation occur at very early stages of L2 learning.

“monolingual” speakers who also read word lists, while sentences will be compared with the group that also produced the sentences. It has been found that the VOT of Polish stops is slightly longer when these are produced in citation forms than when they are couched in longer prosodic constituents (e.g. Schwartz 2020; Malisz and Żygis 2015). For this reason, we do not want to mix these two tasks.

When it comes to the first group of controls[62], that is those speakers who read the word list, they consisted of fifteen native speakers of Polish who were enrolled in an elementary English course taught by the author of the thesis at a private language school in Poznań. Their level of English was quite low (A0/A1, according to *CEFR*). Their English vocabulary was extremely limited and they were familiar most commonly only with the most ubiquitous English phrases, such as introducing themselves or saying where they came from; some of them had had English classes at school but claimed never to have reached a level of proficiency that would allow them to hold a conversation in English, which was why they decided to attend that particular elementary course. They also reported not knowing any other foreign language. The group comprised females aged 17-38 (median age: 25). They all came from the Wielkopolska region and have resided in Poznań their whole lives.

The second control group[63], whose Polish sentence productions were analysed, comprised twenty native Polish speakers, 14 females and 6 males, who were aged 18-29 (median age: 24) and were paid for their participation in the study. They came from various parts of the country but all of them have resided in Poznań for a number of years at the time of the recordings. While they have reported some experience with foreign languages – mostly English and German – which they had studied in school, they claimed not to be fluent in any of them.

The participants from either one of the groups have reported no speech or hearing impairments and – crucially – had never undergone any form of phonetic training. As can be seen, then, even though these two “monolingual” controls were not, in fact, the same

[62] Originally recorded for the purpose of a project on modulation in laryngeal features financed by the National Polish Science Centre (Narodowe Centrum Nauki, grant number UMO-2016/21/B/HS2/00610; Principal Investigator: prof. Geoffrey Schwartz). This group served as controls in an initial study, described in Schwartz (2020), which served as a motivation for carrying out the longitudinal experiment described here.

[63] Originally recorded for the purpose of a project on prosody-segment interactions, carried out by the author of the present dissertation and financed by the National Polish Science Centre (Narodowe Centrum Nauki, grant number 2015/19/N/HS2/03395); cf. Wojtkowiak and Schwartz 2019; Wojtkowiak 2020; Wojtkowiak and Schwartz 2022.

group of speakers, they are similar enough with respect to their language backgrounds and as a result they were deemed appropriate for employing them as controls in the present study.

5.5. Pronunciation training in the case of Polish English majors

In the previous sections the theme of phonetic training has been mentioned quite a few times. While the main goal of the present thesis is to study its effects on Polish productions instrumentally, the students' motivation to acquire English-like pronunciation and be diligent as far as their phonetic courses are concerned is quite important. There exist a few studies which analysed their attitudes towards pronunciation training at the University level.

For instance, the students from Adam Mickiewicz University in Poznań, examined by Sobkowiak (2002) overwhelmingly (67% of respondents, to be exact) disagreed with a statement claiming that good pronunciation outweighed grammar or lexical skills in L2. At the same time, 48% of them claimed to practice good pronunciation in their free time while 75% expressed a wish to attend more pronunciation classes at university.

In a survey in the Łódź area, Waniek-Klimczak (1997) found that while the respondents were concerned with their pronunciation and expressed the desire to sound close to native-like norms, it was not the most important goal in their L2 acquisition. Fluency, confidence, and communication were valued more than losing foreign accent.

Cieślicka and Rojczyk (2017) conducted a study which compared fifth year students of English at the University of Silesia with speakers with lower proficiency in English and the way in which they evaluate their pronunciation skills. They observed that, despite two-year long phonetic training in the case of English majors, their confidence in their pronunciation skills did not differ significantly from less proficient speakers. They also displayed self-reported desire to change their pronunciation. The participants were subsequently exposed to their own recordings, on the basis of which they were asked to assess their pronunciation again. It was found that the post-exposure ratings did not affect their self-evaluation of their skills. Pronunciation was often mentioned in the qualitative part of their study and the students appeared to pay attention to the way they sounded and claimed that they wished for their productions not to display a Polish accent.

Another survey study of this kind was done by Waniek-Klimczak et al. (2015). Their questionnaires related to the attitudes of English majors towards their own pronunciation were filled out by students from two big Polish universities: the University of Łódź and the University of Silesia. It was observed that the respondents felt very strongly about attaining native-like pronunciation; they also felt that their pronunciation skills were somewhat lacking. Therefore, the authors conclude that the acquisition of pronunciation over the course of their studies is perceived as relatively difficult a task. Going deeper into their participants' profiles, the authors observed that female students appeared to assess their skills more critically relative to male students. Females were also observed to feel more strongly about correctness. What is also interesting is that there was a difference in the attitudes towards Polish features in the students' productions which hinged upon the level of studies[64]. The students further down the path of their university education seemed to be less concerned with eradicating all traces of Polish accent in their pronunciation. Therefore, while pronunciation was still important to older students, they were less concerned with sounding like a native speaker.

What is also quite intriguing is the English majors' attitudes with respect to the models of pronunciation taught at Universities: British and American. The former has been described by students from Adam Mickiewicz University in Poznań as 'unspoilt', 'academic', 'classy', 'charismatic' but also 'ridiculous', 'stiff', and 'old fashioned' (Janicka et al. 2005). In turn, the latter has been claimed to be 'relaxed', 'business-like', and 'neutral', albeit 'primitive' and 'careless' (Janicka et al. 2005). A little over a decade later, when asked how they feel while speaking their accent, British English graduates from the same institution described themselves as 'professional', 'educated', 'empowered', and 'high class', but also 'pretentious', 'uncomfortable', 'embarrassed', and 'inadequate' (Krzysik and Lewandowska 2017). Meanwhile, their American English colleagues claimed to feel overwhelmingly 'cool' and 'comfortable' and did not associate their accent with any negative connotations that the British English group mentioned (Krzysik and Lewandowska 2017).

[64] That is, between the BA level, that is the initial three years of studies vs. the MA level, the subsequent two years, resulting in obtaining the degree of a Master of Arts.

5.6. Materials

This section will present the materials used in the present study. These can be divided into a word list and a sentence list in both Polish and English. The details are given below.

What is worth noting is that while for the purposes of the present thesis it was the laryngeal contrast between /p, t, k/ and /b, d, ɡ/ that was of utmost importance, both the word lists and the sentence lists contain a number of fillers, which were not chosen at random. Fricative-initial, affricate-initial, and cluster-initial tokens, on the whole, also represent interesting differences between Polish and English. For time and space reasons they are not analysed in this work, however they constitute an exciting avenue for future research.

5.6.1. Word lists

The Polish word list consisted of 121 words in total, out of which 54 were included in the analysis with the rest of them being fillers. The /b, d, ɡ/ and /p, t, k/[65] items were mono- and di-syllabic words, followed by vowels [a, ɛ, ɔ]. The dataset included 17 labial, 19 coronal, and 18 velar onsets. 29 of them were voiceless while 25 of them – voiced. With a total of 65 speakers (out of whom 20 were recorded three times), a database of 5670 Polish recorded items was comprised.

The English word list consisted of 131 items mono- and di-syllabic words in total, with 55 of them included in the analysis. 21 of them were labial, 18 alveolar, and 16 velar. There were 27 voiced and 28 voiceless onsets. The initial consonants were followed by non-high vowels: KIT, DRESS, TRAP, STRUT, LOT, BATH, PRICE, and FACE. A number of words from this language pair were the so-called “false-friends”, that is words which look or sound similar to the speakers’ native language lexical items, but their meaning might be different. Since the English word list was read by Groups 1, 2, and 3 only, the 50 speakers recorded 4950 items in total.

All words used in the study, including fillers, can be found in Appendices 2 and 4.

[65] The names “plosives” and “stops” are used interchangeably hereafter.

5.6.2. Sentences

As far as the sentence lists are concerned, the Polish dataset comprised a list of twenty four disyllabic Polish words, starting with /p, t, k, b, d, g/, which were followed by three front vowels: /a, ɛ, ɨ/. It is important to note that the vowel /ɨ/ is subject to some distributional restrictions in Polish and does not occur after velar plosives. As far as the voiceless series is concerned, the combination of */kɨ/ was therefore not included in the materials. In the case of the voiced velar, we used the word gyros ([gɨrɔs], 'gyros'), which is not a native Polish word but with which Polish speakers are familiar due to its omnipresence.

The words were counterbalanced for place of articulation (8 labial, 8 coronal, and 7 velar targets) and voicing (11 voiced and 12 voiceless). They were subsequently embedded in meaningful sentences, carefully controlled for the number of syllables that preceded and followed them. The target words were always preceded by the vowel [ɨ], thereby creating a context conducive to voicing (Cyran 2011). The sentences were devised in a way which would allow the target words to be placed at three distinct prosodic positions: utterance-initial, phrase-initial (i.e. utterance-medial), and phrase-medial.

Following other studies which dealt with different prosodic positions (e.g. Cho and McQueen 2005; Cho and Keating 2001) it was assumed that the utterance-initial position will be elicited by means of a full stop that preceded the target word. Similarly, a comma was supposed to trigger a weaker prosodic boundary, thus placing the target word in a phrase-initial position. A relatively good indicator of the difference between these two positions was the intonation contour on the vowel preceding the target word; in the case of the utterance-initial position it was usually a fall, whereas in the phrase-initial position the contour was rising, implying that more information is to follow. The third position, that is the phrase-medial context, was elicited by no punctuation marks preceding the target word, with the intonation contours relatively flat.

To illustrate all this, an example is given in (1), with the target word *bochen* 'a loaf' bolded and underlined.

(1) a) Codziennie go wypiekamy. <u>**Bochen**</u> żytniego chleba jest smaczny.
*We bake it every day. A <u>**loaf**</u> of rye bread is tasty.*
SENTENCE TYPE: UTTERANCE-INITIAL.

b) Zawsze gdy go wypiekamy, **<u>bochen</u>** pszennego chleba jest chrupki.

*Whenever we bake it, the **<u>loaf</u>** of wheat bread is crunchy.*

SENTENCE TYPE: PHRASE-INITIAL.

c) Od pokoleń wypiekamy **<u>bochen</u>** wiejskiego chleba na mące.

*For generations, we've been baking a **<u>loaf</u>** of rustic bread.*

SENTENCE TYPE: PHRASE-MEDIAL.

It cannot be ruled out that the target words located next to a stronger prosodic boundary (such as in the utterance-initial position) will be subject to some domain-initial lengthening of VOT (Kuzla and Ernestus 2011). Prosodic position will, therefore, be included as a predictor in the analysis. Assuming that some domain-initial lengthening will be observed (though previous studies on Polish show that there is little contrast with respect to the acoustics of segments following a prosodic break; cf. Wojtkowiak and Schwartz 2021, Wojtkowiak 2020), contrast should be observed between stronger boundaries and the phrase-medial position, wherein there is no prosodic boundary adjacent to the target word. The phrase-medial positions could be conceived of as a baseline. This way we will not confuse domain-initial lengthening with possible phonetic drift effect, which could also result in differences with respect to VOT, thereby lowering the risk of committing Type 1 error.

In total the sentence list consisted of 119 sentences (including fillers: fricative-, vowel-, and s+stop cluster initial words), out of which 69 ((6 plosives × 4 vowel contexts × 3 prosodic positions) – 3 [*kɨ] = 69) were included in the analysis. With 70 speakers we obtained 4830 recorded productions.

The English sentence list was devised to resemble the Polish one as closely as possible. The target words were counterbalanced for place of articulation (8 labial, 8 coronal, and 8 velar) and voicing (12 voiced onsets and 12 voiceless ones). They were followed by vowels KIT, TRAP, DRESS, STRUT. Nonetheless, it must be stressed that first year students, even though they are fluent in English by the time they enter University, are quite reluctant to produce longer utterances in their L2, especially when they know that they are being recorded. In order to make the task as easy as possible, an attempt was made for the sentences to be relatively simple. While the prosodic position was still controlled for and the target words were placed in similar contexts (i.e. utterance-initially, phrase-initially, and phrase-medially), they were not always preceded by a vowel, as was the case for the Polish list. As a result, an additional predictor which would need to be included in the analysis will be the preceding context (i.e. conducive or not conducive to voicing). The number of syllables preceding and following them was also not kept constant.

Once again, to illustrate this, an example of a target word *guessing* placed in three distinct prosodic positions is shown in (2):

(2) a) You should know this by now. **Guessing** shouldn't even be an option.
SENTENCE TYPE: UTTERANCE-INITIAL.
b) If you don't know the answer, **guessing** might be your only choice.
SENTENCE TYPE: PHRASE-INITIAL.
c) It's not a surprise that people started **guessing** which answer was correct.
SENTENCE TYPE: PHRASE-MEDIAL.

There were 131 sentences in the dataset (including vowel-, fricative-, and s+stop cluster initial fillers), out of which 72 (6 plosives × 4 vowel contexts × 3 prosodic positions) were included in the analysis. With 55 speakers we obtained 3960 recorded sentences in total.

All sentences used in the study, including fillers, can be found in Appendices 1 and 3.

5.7. Procedure

Group 1 were the only that have participated in the longitudinal experiment. They were tested three times during the first year of their University education, with six recording sessions in total. The sessions were held in October (henceforth T1), February (henceforth T2), and June (henceforth T3). The first batch of recordings was made within the first two weeks of October. Therefore, it can be assumed that the session was held early enough for the effects of phonetic training not to have affected the students' productions yet. The second session was held in February, during the winter exam session and once again took two weeks in total. At the time of the recordings, Group 1 were just about to start their second semester of studies. The last recording session was held in early June, towards the very end of the academic year, after about eight months of phonetic training. In order to avoid language mixing effects (Grosjean 2004), each language was recorded separately and participants never recorded both languages on the same day. Moreover, the whole procedure (i.e. welcoming the participant, giving instructions, chit-chat) was conducted in the language about to be recorded – either in Polish or in English

Groups 2 and 3 were recorded once, in late June, towards or during the summer exam session. Similarly to Group 1, the recordings for each language were made on separate occasions.

In the case of all three groups described above, both the sentence list and the word list were recorded during the same meeting, with a short (usually five minute) break in between the recordings. The recordings for each language lasted around twenty five minutes in total. They were made in a sound-attuned booth at Adam Mickiewicz University, directly onto laptop, using a condenser microphone and a USB interface. The words and sentences were shown to the participants on a monitor via PowerPoint slides in a pseudo-randomised order (the same for each participant), with the pace controlled by the researcher sitting in another room.

As far as Group 4 are concerned, as has been mentioned in section 5.4.3, the word list and the sentence list were elicited from two different groups of people; therefore the slight procedural differences in the case of this group stem solely from the fact that the controls were originally parts of two different projects.. In the former case, while the items were presented to the participants by means of PowerPoint slides, in a pseudo-randomised order (the same for each speaker), the recordings were made in a quiet classroom in the private language school in Poznań which they were attending. They were recorded directly onto laptop, using a head-mounted microphone and a USB interface. In the latter case, the recordings were held at Adam Mickiewicz University and the procedure was similar to Groups 1-3, with one slight difference, namely the order of the sentences was randomised for each participant. Furthermore, this subset of recordings was made by the former Principal Investigator working on the project on prosody-segment interactions in Polish.

Aside from the questionnaire about their linguistic backgrounds, which has already been mentioned, the participants also signed informed consents before taking part in the study – they were given very little information about the study itself (aside from the fact that it was a phonetic study and that recorded data were to be collected) so as not to reveal the purpose of the experiment. The procedure was explained – the participants were told that they were to read common sentence and word lists in Polish and English, at their own pace, as if they were reading them to a friend sitting next to them. The participants were informed that at any stage of the research project they could make the decision to withdraw their contribution and that only the person conducting the analysis would have access to the .wav files collected. The recordings from the sessions were later stored on an encrypted external drive without Internet connection. Each participant was given a number associated with their name and the recordings were given the names in the following manner: participant_session_language_task (e.g. 001_October_Polish_sentences; 202_2BA_English_words). The numbers corresponding to the participants' names were written on the consent forms so that only the student and the experimenter would be able to identify them.

5.8. Methodological improvements over existing studies

At this point, I would like to point out some of the methodological improvements that have been introduced in the present study, relative to the existing studies on phonetic drift in an L1-dominant environment (in particular, Herd et al. 2015 and Shuchmann and Huffman 2015, due to the similarity of these studies to mine). Some of these may have briefly been mentioned in the previous sections of the thesis, but may be in need of reiteration before we move onto the discussion of the results.

First of all, the study described here is a longitudinal study. Such a study can be more methodologically challenging, especially with regards to the number of speakers participating in it. As has already been mentioned, there were thirty five speakers originally recorded in Group 1 at T1, however the final sample comprised twenty speakers only. The remaining fifteen either dropped out or decided to resign from their participation in the experiment for personal reasons unknown to the investigator. On the other hand, despite lower numbers of attendees, a longitudinal study allows us to get a bigger picture of what happens over the course of L2 phonetic training in real time and as a result gain more insight into the data with which we are working.

Furthermore, the groups under investigation are more homogenous. All of the students (i.e. Group 1 and the two comparison groups: 2 and 3) have undergone the same type of theoretical and practical phonetic training. In the case of Herd et al.'s (2015) study, the participants took part in language classes but the authors do not provide an explanation of what these classes entailed and how much focus was placed on pronunciation training. Schuhmann and Huffman (2015) were more explicit in what their phonetic training looked like, however a big disadvantage was the number of participants (i.e. only five people) included in the study, as well as the fact that they recorded both language samples (i.e. Spanish and English) consecutively during the same recording session. In particular, it is not clear which language was used when instructions regarding the experiment were given, especially since both languages were recorded at the same time. While on the whole, the research on the influence of language mode on phonetic realisations has brought inconclusive results, Antoniou et al. (2010, 2011) found that in the case of Greek-English bilinguals, monolingual language mode induced VOT values that fell within the monolingual norms of each language, but in a code-switching task, L2 values were affected by the L1. Furthermore, Amengual (2018) found that Spanish-English heritage speakers and L2-Spanish late learners produce less target-like laterals when in a bilingual language mode. To avoid such risks, in the case of the present study, as has been mentioned before, the recording sessions for Polish and English were held on separate days and the entire experiment was conducted exclusively in the language which was being tested.

Furthermore, both of those other studies relied on words lists that included multiple repetitions of individual words. Winter (2015) argues against using multiple repetitions in phonetic studies, claiming that "averaging" over repetitions of the same word – as was the case in Herd et al.'s and Schuhmann and Hufman's studies – might become "a biased estimator of the phonetic target" (2015: 1). Kello et al. (2008) show that variation over multiple repetitions is not normally distributed. Hence, in the present study repetitions were avoided. The target words in sentences were identical in order for us to be able to investigate the effects of prosodic positioning on segmental phonetics, however, the meaningful sentences in which they were couched differed.

The first recording session in which Group 1 participated was held very early in October. This provides more internal validity to the study, as we can compare the students' productions at the very onset of their phonetic training, before the effects of this training can influence them. As a result, we reduce the risk of committing Type 1 error, which would entail observing a difference where there is none. This might have been the case with Herd et al.'s (2015) study; as has been already discussed in sections 2.3.1 and

3.5.2, we know that pre-voicing is attested in the productions of Southern American English speakers', therefore we cannot be sure that the observed pre-voicing is there due to the exposure to Spanish. Hence, the presence of a control group in the present study is also beneficial. We can compare the students' productions to the "monolingual" norms at every stage of their training.

Finally, thanks to including English recordings we may also trace the trajectory of the acquisition of segments. Taking into account that we are studying the productions of groups at various stages of their University education, who use their L2's on daily basis, this will allow us to also see how stable the categories are – we can investigate whether the acquisition of the 'canonical', English categories is permanent, or whether we may observe some asymmetries in their behaviour as well, including them as predictors in our statistical analyses.

5.9. Acoustic annotation

The acoustic annotations were made exclusively by hand using Praat (version 6.1.50; Boersma and Weenink 2021) by the author of the thesis. Manual annotation requires the researcher to spend a great deal of time looking at acoustic displays, much more so than forced alignment with manual correction. This provides a better and more insightful perspective on the acoustic richness of speech.

Let us begin with the word lists, where the annotation was identical for both languages. They were annotated using three tiers. Tier 1, named "segment" was used in order to mark all phonetic events associated with the production of stop-vowel sequences. Tier 2, named "type", was used to categorise VOT according to four types, about which we will speak shortly. Tier 3, in turn, was named "vowel" and was used to mark the duration and the quality of the vowel only.

Voiced plosives were subdivided into three groups according to their type. "Type" was operationalised as three distinct ways in which a voiced stop could be produced, on the basis of previous research on Polish (Schwartz 2020). "Type 1" group encapsulated standard, pre-voiced realisation of a voiced plosive, characterised by vocal fold vibration during the closure phase. The measurements on Tier 1 ("segment") included pre-voicing which was measured from the first observable pulse of voicing until the release of the closure, excluding burst and the release noise. The following vowel, also included on Tier 1, was measured from the onset of voicing associated with vowel production (thus, excluding burst and release noise) until the point in which F2 and F3 were no longer visible. Tier 2 ("type") included the length of pre-voicing, while Tier 3 ("vowel") measured the length of the vowel only. The decision to include Tier 3, even though the vowel was also labelled as such on Tier 1, stemmed from the attempt to make subsequent analyses (in particular, using the Praat scripts for the purposes of formant values' extraction) easier[66]. This will be explained shortly. An example of a Type 1 voiced plosive is given in Fig. 32.

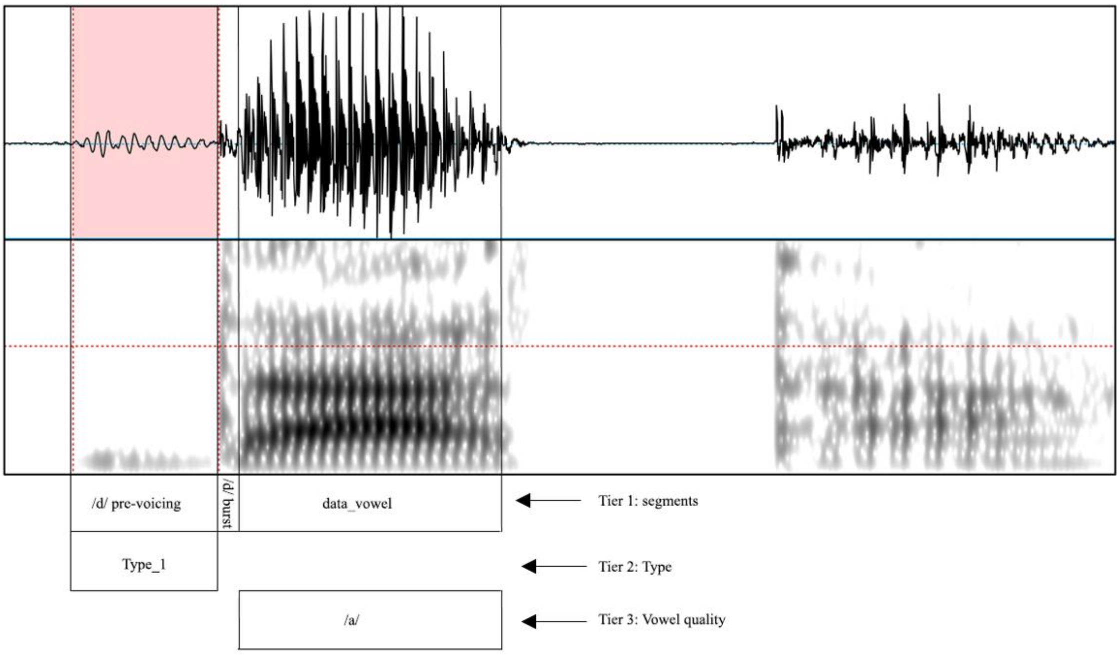


Fig. 32. An example of the Polish word *data* ('a date') with a Type 1 initial plosive. Tier 1 corresponds to the duration of individual phonetic events associated with the CV sequence (pre-voicing, release noise associated with burst, and vowel), Tier 2 denotes the type of voicing realisation, while Tier 3 measures vowel duration only.

"Type 2" realisation was a type of production wherein there was a noticeable break in negative VOT, preceding the release of the closure. In those cases, the annotations on Tier 1 looked as follows. First, we measured the vocal fold vibration (i.e. pre-voicing) from the first pulse of laryngeal activity until the break. The following break, a voiceless portion of virtual silence, was subsequently labelled as such, and was measured from the

[66] In Polish the vowel on Tier 3 was labelled using a corresponding alphabetic symbol, whereas in English – with its Welssian keyword.

cessation of laryngeal activity until the burst and noise, indicating the release of the closure. The vowel was measured similarly to what was the case for Type 1 plosives. Importantly, Tier 2 (“type”) subsumed both pre-voicing and the subsequent break. As a result we could easily measure those two events separately (by reading the measurements off Tier 1) as well as collectively (by reading the information off Tier 2). This will allow us to then see how much breaks in pre-voicing influenced the overall length of the negative VOT in Type 2 productions. For clarity, an example of a Type 2 plosive is shown in Fig. 33.

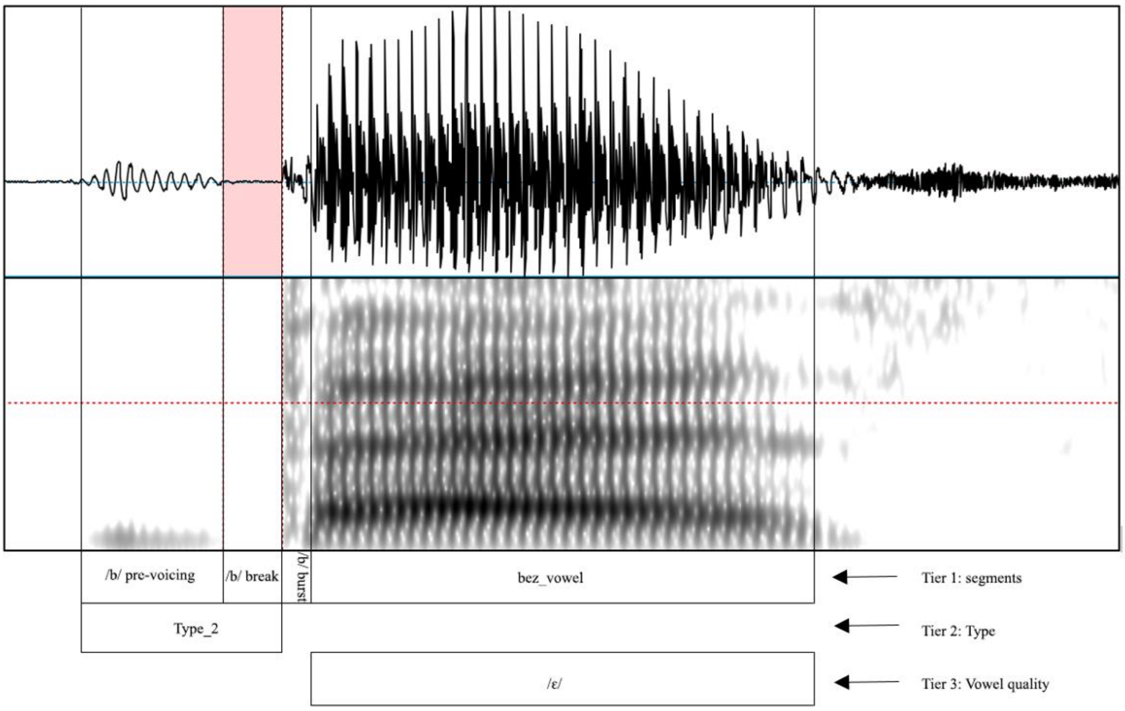


Fig. 33. An example of the Polish word *bez* (‘lilac’) with a Type 2 initial plosive. Tier 1 corresponds to the duration of individual phonetic events associated with the CV sequence (pre-voicing, pre-voicing break, release noise associated with burst, and vowel), Tier 2 denotes the type of voicing realisation (encompassing both pre-voicing and the break), while Tier 3 measures vowel duration only.

Finally, “Type 3” was an instance of an English-like production, where pre-voicing is missing and the voiced stop is realised with short positive VOT. In this case, the VOT was measured from the release of the closure until the onset of voicing associated with the vowel and labelled, “unvoiced” on Tier 1 (“segment”). The vowel was measured in a standard way, as described for Type 1 and Type 2 productions and also marked on Tier 1. Tier 2 (“type”) included the VOT of the stop as it was marked on Tier 1. Tier 3 included the vowel only. This is shown in Fig. 34.

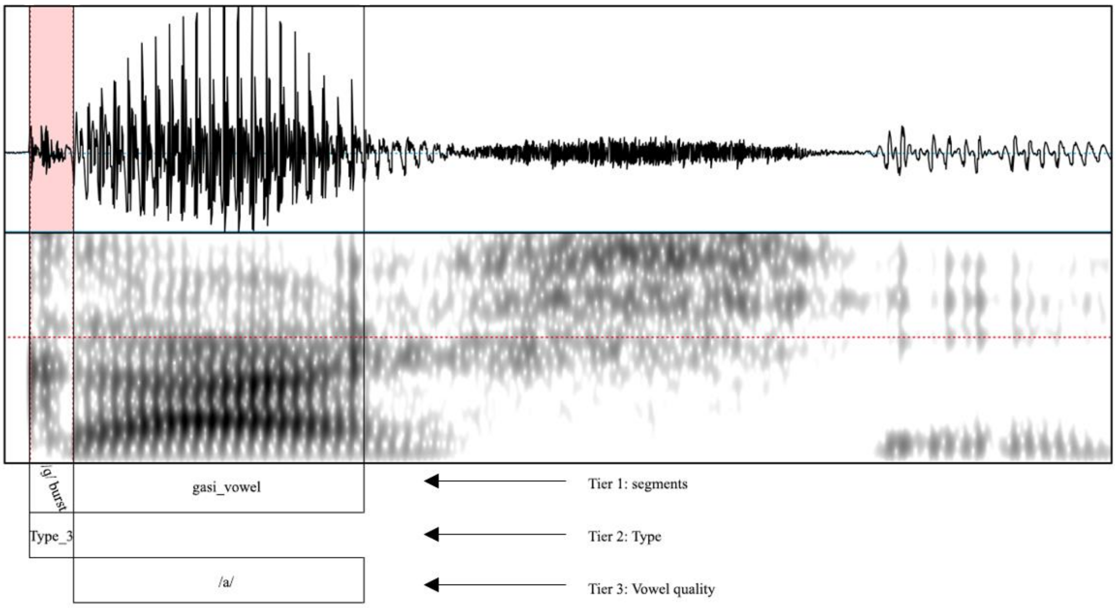


Fig. 34. An example of the Polish word *gasi* (‘to extinguish’, 3rd.sg.) with a Type 3 initial plosive. Tier 1 corresponds to the duration of individual phonetic events associated with the CV sequence (release noise associated with burst, and vowel), Tier 2 denotes the type of voicing realisation, while Tier 3 measures vowel duration only.

When it comes to voiceless consonants, the duration of positive VOT was marked, counted from the release of the closure until the pulse of phonation associated with the onset of the vowel, visible on the spectrogram. An example is given in Fig. 35. On Tier 2, this category was marked as ‘Type 4’, while the vowel on Tier 3 was marked in the same way as has been already described for the voiced category.

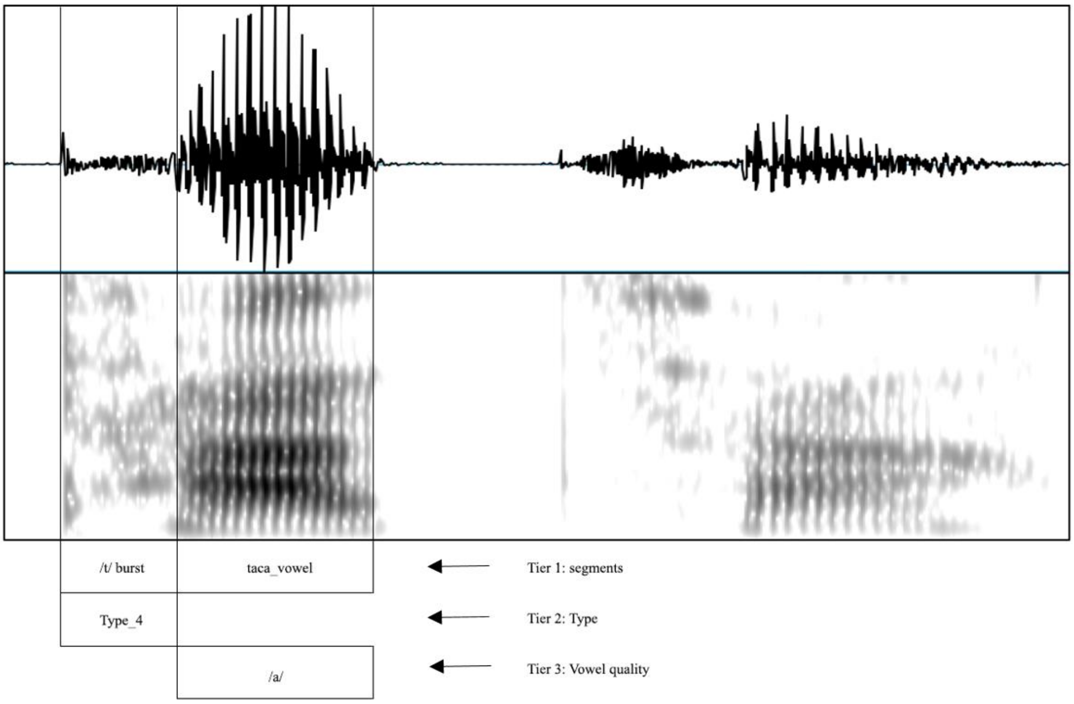


Fig. 35 An example of the Polish word *taca* ('a tray') with a Type 4 initial plosive. Tier 1 corresponds to the duration of individual phonetic events associated with the CV sequence (release noise, and vowel), Tier 2 denotes the type of voicing realisation, while Tier 3 measures vowel duration only.

To reiterate, the inclusion of three separate tiers made the subsequent analysis easier and clearer. Thanks to Tier 1 ("segment"), we obtained the measurements of each phonetic event associated with the production of stop+vowel sequences, and by labelling them with the spelling of a given word, we were able to gather information about the place of articulation of the plosive, which could then be included in the statistical analysis. This was especially important for obtaining f_0 and F1 measurements from the onset of the vowel's production, which is influenced by e.g. voicing of the preceding consonant. Tier 2 not only calculated the total length of the plosives articulated with a break in negative VOT but also allowed us to then count the sheer numbers of different types of stop production by different groups and at different stages of the phonetic training. Tier 3, which included only vowels, allowed us to obtain F1 and F2 measurements from the medial portions of the vowels, when the analysis does not necessitate including other predictors, such as place of articulation or voicing of the preceding consonant. It might have been redundant, as the same could be read off Tier 1, but it made the data easier to deal with.

Using Praat scripts, the following measurements were acquired. The first script provided the length of positive (for voiceless stops) and negative (for voiced stops) VOT (in milliseconds). The second script was used to divide each target vowel on Tier 3 into five equal parts. Subsequently, we looked at two of the five parts. f_0 (in Bark) and F1 (F1-f_0; Bark normalised) were obtained from the measurements taken from the initial portion of the vowel (first 20% of its duration) to study laryngeal effects on vowel onset. These values were matched with the labels extracted from Tier 2, thanks to which it was possible to assign onset type to the vowel acoustic data.

Additionally, the values of F1 (F1-f_0; Bark normalised) and F2 (F3-F2; Bark normalised) from the middle 20% of the vowel were taken to be included in the analysis. We chose to focus our attention on two Polish (i.e. /a/ and /ɛ/) and three English (TRAP, DRESS, STRUT) vowels. This decision was made for practical purposes: the present thesis overviews many different parameters and it was impossible to include everything. The remaining recordings might be analysed in the future, to further investigate whatever findings we observe in the present thesis. The choice of those five vowels might be further supported by the fact that – as has been shown in section 4.2.1 – the English vowels TRAP, STRUT, and DRESS constitute a challenge for Polish learners and can interact with Polish /a/ and /ɛ/ quite robustly.

At this point, the choice of vowel normalisation requires some explanation. The Bark difference measures used in the Syrdal and Gopal normalisation method have been found to better reflect the auditory properties of phonological categories associated with vowels, such as height and backness, than raw values of single formants (see Chistovich et al. 1979; Homeke and Diehl (1994)). While Syrdal and Gopal's normalisation has been observed to be slightly less successful than vowel-extrinsic methods in classification studies dealing with sociolinguistic variation, the present study does not focus on speech categorisation. For the purposes of the study, the Bark-difference normalisation method allows us to control for speaker-dependent differences (due to the speakers' sex, the size of their oral cavity, etc.), while yielding a single-measure suitable for cross-speaker comparison.

As could be observed in the presentation of the materials used for the present study, when it comes to the sentence lists the target words were placed at different positions within a word: utterance initial, phrase initial (i.e. utterance medial), and phrase medial. In order to ensure that the prosodic positions are categorised correctly, the following steps were taken. The recordings obtained from the control group of quasi-monolinguals were listened to by two phonetically-trained native speakers of Polish (fully crossed design; Hallgren 2012), one of whom being the author of the thesis[67]. The annotators marked them as correct (i.e. the speaker produced the sentence they were supposed to with regards to prosodic position) or incorrect (i.e. the speaker did not produce the expected prosodic boundary or some other mistake was made). The Cohen's Kappa Coefficient (Cohen 1968) agreement between the raters was calculated and the results showed that agreement regarding the prosodic position were 'almost perfect' (κ=0.97, following Landis and Koch 1977). The tokens which yielded disagreement between the two raters were discarded. The excluded sentences were also those wherein the sentences were produced with errors, unnatural production of the prosodic boundary, or hesitations. When it came to the recordings from Groups 1, 2, and 3 the same two annotators independently cross-referenced a portion of the recordings after they have been made with the sentence list of the desired realisations. First, we pooled all recordings from all recording sessions into two groups: Polish and English, forgoing the division into the years of study. Out of these two, 15% of speakers from each language (N=13[68]). Having listened to and marked the recordings from both groups (897 Polish and 936 English sentences in total), Cohen's kappa coefficient was again calculated. The choice was again binary: either *correct* or *incorrect* (i.e. an erroneous sentence, mispronunciation of the target word, an unnatural production, hesitations, etc.). For Polish, the agreement between the raters was 'almost perfect' (κ=0.98). For English the results reached the level of 'substantial agreement' (κ=0.76). Upon inspection, it was found that the most problematic was one English sentence (i.e. *'Some animals are natural enemies, penguins and seals being an example'*, with some speakers producing with a strong prosodic boundary, that is placing it utterance-initially). After excluding it completely from the dataset, the calculated agreement rose to 'almost perfect' (κ=0.93). The remaining sentences from both languages were subsequently annotated by the author of the thesis only, with the inclusion of all remaining tiers.

As far as the rest of annotations go, the VOT (both positive and negative) was measured and categorised in the exact same way as was the case for the word lists. Therefore Tier 1 comprised the measurements of all phonetic events (pre-voicing, breaks, re-

[67] This was done first as the recordings from the group were gathered earlier than the recordings of the study groups.

[68] Group 1: 60 speakers (20 at T1, 20 at T2, 20 at T3), Group 2: 15 speakers, Group 3: 15 speakers – 90 speakers in total; 15% out of 90 is 13.5, rounded down to 13.

lease bursts, and the vowel) while Tier 2 encompassed the entire duration of VOT (including possible breaks in negative VOT but excluding release burst), with the label corresponding to VOT type (that is, fully pre-voiced, partially pre-voiced, unvoiced, or voiceless). Tier 3 included the vowel only. An additional, fourth tier, was introduced in order to account for speech rate. There the duration of the entire utterance was measured, with the label corresponding to the spelling of the sentence with the inclusion of the number of syllables of the given sentence. Examples, including all four tiers, are given in Fig. 36 and Fig. 37.

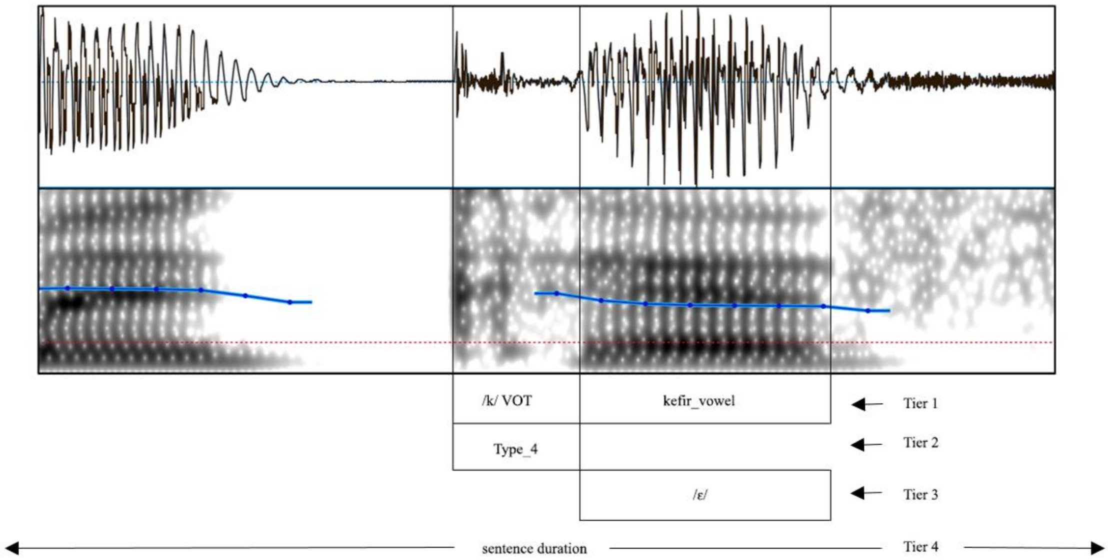


Fig. 36. An example of the Polish sentence, zoomed in on the keyword *kefir* ('kefir) with a Type 4 initial plosive. Tier 1 corresponds to the duration of individual phonetic events associated with the CV sequence (release noise and vowel), Tier 2 denotes the type of voicing realisation, while Tier 3 measures vowel duration only. Tier 4 denotes the duration of the entire sentence.

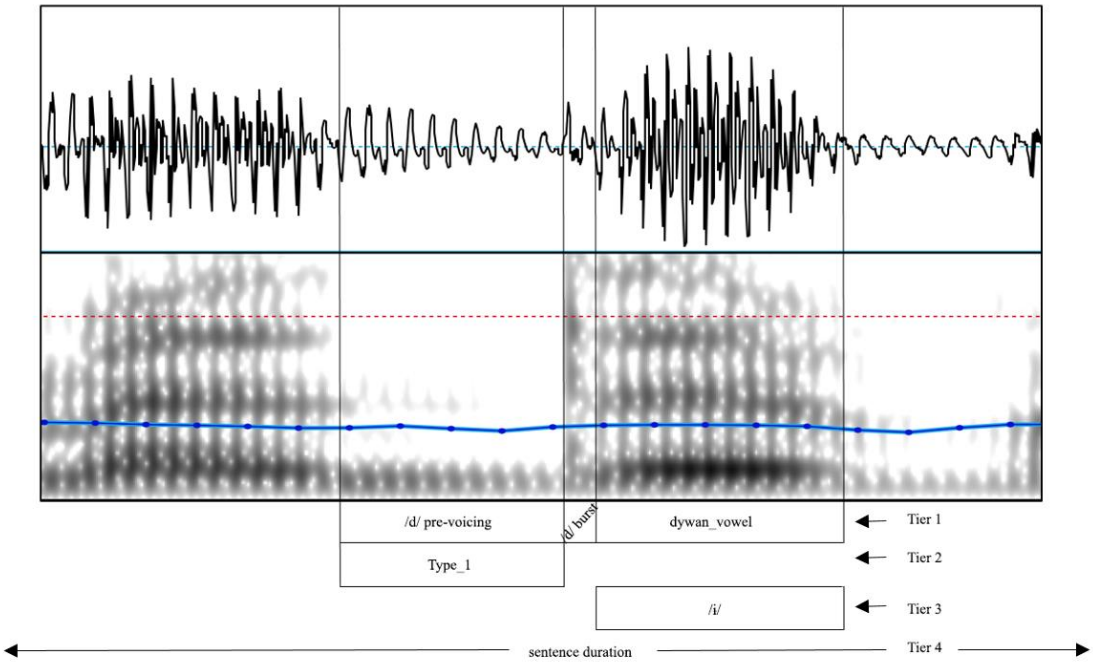


Fig. 37. An example of the Polish sentence, zoomed in on the keyword *dywan* (‘carpet’) with a Type 1 initial plosive. Tier 1 corresponds to the duration of individual phonetic events associated with the CV sequence (pre-voicing, release noise, and vowel), Tier 2 denotes the type of voicing realisation, while Tier 3 measures vowel duration only. Tier 4 denotes the duration of the entire sentence.

Both these figures depict sentences wherein the keyword is placed in a phrase-medial position (i.e. with no prosodic boundary preceding it). The intonation contour is included, which was used as additional help in judging whether or not a correct sentence was produced (a falling contour on the preceding vowel implied a strong prosodic boundary and an utterance-initial position of the keyword; a raising contour on the preceding vowel implied that more information was coming, therefore the keyword was placed in a phrase-initial position; a flat contour indicated a lack of a prosodic boundary). Previous studies dealing with prosodic positions used similar methodology (e.g. Cho and McQueen 2005; Cho and Keating 2001).

One additional comment to be made regards the categorisation problem between Type 2 (i.e. partially pre-voiced) and Type 3 (i.e. unvoiced) tokens in phrase-medial sentences. Since virtually no breaks before the production of the target word (due to a lack of any prosodic boundary) were made on the whole, it was sometimes difficult to distinguish those two types of VOT. In general, when there was a break in voicing in the closure portion, a decision had to be made whether this is an instance of a partially pre-voiced item or whether the vocal folds’ activity during closure is a case of voicing bleed (cf. Davidson 2016) from the preceding vowel and the target word is in fact unvoiced. In order to be consistent, an arbitrary cut-off point was decided on: if the break in vocal fold vibration was 50 milliseconds or longer, the voiced target word was categorised as Type 3 token. If it was 49 milliseconds or shorter, it was assigned into the Type 2 category. An example of a token with a 55-millisecond break (i.e. categorised as Type 3) is shown in Fig. 38.

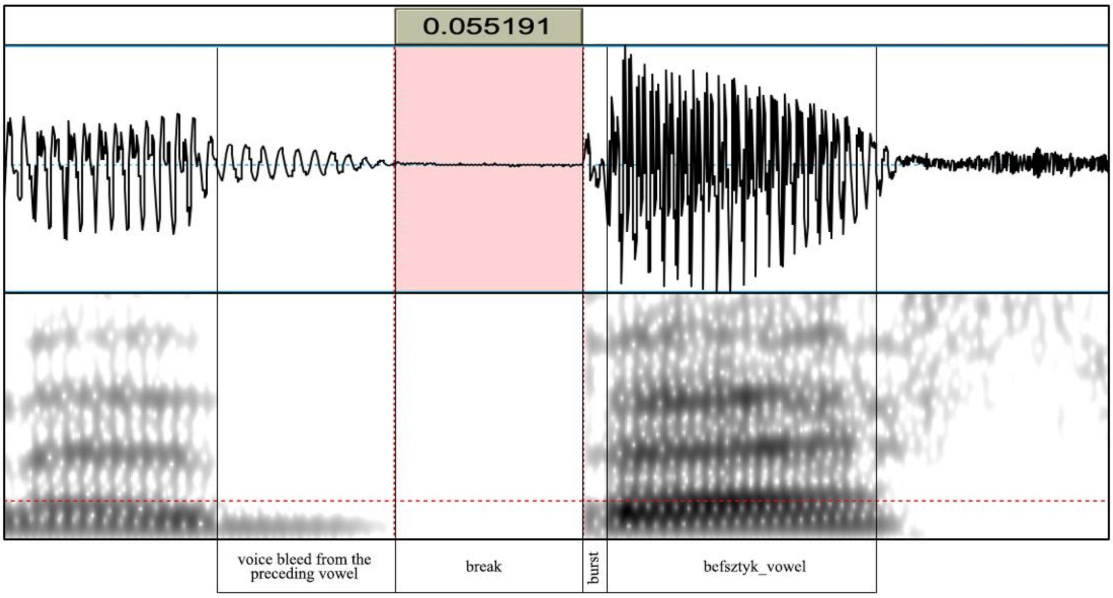


Fig. 38. An example of a phrase-medial keyword *befsztyk* (‘beefsteak’) in the sentence production task, categorised as Type 3 token. Notice that the break in voicing, preceding release burst, exceeds 50ms.

5.10. Statistical analyses

This section will describe the details of the statistical analyses run after the collection and subsequent annotation and extraction of the acoustic data. The analyses were done in SPSS v. 27.1 (IBM Corporation 2020).

Let us begin with laryngeal contrasts. As far as word lists are concerned, we focused on three parameters: VOT, f_0, and F1. These were our dependent variables for which separate Generalised Linear Mixed Models were run. For VOT, the interaction between (recording) session*type*place (of articulation) was the main predictor of interest (i.e. the fixed factor), while Speaker and Item were the random factors. Place of articulation was included in the analysis as the dataset was not perfectly counterbalanced with respect to this parameter. An additional Binary Logistic Regression analysis was run for

Type 2 (i.e. partially voiced) Type 3 (i.e. unvoiced) of voiced plosive realisation. This variable was coded in a binary fashion (*yes* if there was no pre-voicing present and the token was produced in an English-like fashion and *no* if there was negative VOT present, regardless of the fact whether it was full or partial). Session was the main predictor in this analysis while Speaker and Item were the random factors. For the dependent variables of f0 and F1, session*vowel*type was the main interaction of interest, while Speaker and Item were the random factors.

In the case of the comparison of VOT in Polish and English, to be reported in Chapter 7, the dependent variable was VOT (positive of /p, t, k/ and Type 1 negative of /b, d, g/), with Language*Session being the main interaction, and Speaker and Item the random factors. The interaction Language*Session*POA as the main predictor will also be described.

Table 15 summarises the dependent variables analysed in the case of word lists, including information about fixed and random factors.

Table 15. A summary of all parameters included in the statistical analyses, including main predictors and random factors.

Task	Dependent variable	Scale	Fixed factors	Random factors
Word list	VOT (positive and negative)	milliseconds	Place of Articulation*Type*Session	Speaker, Item
	VOT Type (for Types 2 and 3)	yes/no	Session	Speaker, Item
	VOT (positive and Type 1 negative)	milliseconds	Place of Articulation*Language*session	Speaker, Item
	f0 (first 20% of the vowel)	Bark	Vowel*Type	Speaker, Item
	F1 (first 20% of the vowel)	Bark	Vowel*Type	Speaker, Item
	F1 (middle 20% of the vowel)	Bark	Vowel*Session	Speaker, Item
	F2 (middle 20% of the vowel)	Bark	Vowel*Session	Speaker, Item
Sentence list	VOT (positive and negative)	milliseconds	Type*Session, speech rate	Speaker, Item
	VOT type	yes/no	Session	Speaker, Item
	f0 (first 20% of the vowel)	Hertz (Hz)	Vowel*Type	Speaker, Item
	F1 (first 20% of the vowel)	Bark	Vowel*Type	Speaker, Item

	F1 (middle 20% of the vowel)	Bark	Vowel*Session	Speaker, Item
	F2 (middle 20% of the vowel)	Bark	Vowel*Session	Speaker, Item

5.11. Limitations

Before we proceed to the results, it is important to acknowledge some limitations of the present stud, even though an attempt was made for the study to be as thorough as possible.

The main points regarding the limitations to be made are associated with the analyses presented in Chapter 7 and are related to the L2 English data and how they compare with L1 Polish. The first is the lack of a control group, made up of L1 English speakers. Such a group was, unfortunately, not included in our analyses as finding a homogenous group of L1 English speakers in Poznań (where the study was conducted) was impossible to recruit. The English native speakers who live in Poznań come from very different parts of various English-speaking countries and are, by and large, quite proficient in Polish. The latter fact would immediately disqualify them from taking part in the study, as true-voice languages have been found to exert influence on aspiration languages (e.g. Podlipský 2020; Sucková 2018; Dokovova 2015), similarly to the reverse situation (which has been the main focus of the dissertation). One could possibly envision using the recordings available in the form of various corpora (e.g. the Buckeye corpus of spoken American English or The International Corpus of English – spoken General British), but these use spontaneous speech; both tasks our participants were subject to included simple reading tasks, rather than spontaneous interactions. Therefore, this choice would have also been rather suboptimal. Gathering the data from English controls remains, therefore, a possible task for the future.

Second of all, the main issue with regards to the vowel productions (once again, relating to the L1-L2 comparison in Chapter 7), which were the main focus of section 7.3, is the difference in the numbers of particular vowel contexts. While an attempt was made to only include non-high vowels, as they are less likely to influence VOT values, there were significant discrepancies in the numbers of the words containing the vowels in question: that is, while there were many instances of STRUT and DRESS vowels, TRAP and BATH were extremely underrepresented. This stems from the fact that the idea to compare the vowels alongside consonants was born after the first recording session had already finished. Due to time constraints, pausing the experiment and waiting another year to start it with a word list which would be better suited was out of the question. In principle, however, the sentence list is much more counterbalanced and still awaits a proper analysis. It is hoped that the results regarding the comparison of Polish and English sentence lists will be published after the present thesis has been defended.

Another issue related to vowels pertains to the choice of measurements taken to assess vowel quality as well as the model itself. The measurements of height and advancement were the averages taken from the middle portion of the vowel (i.e. the middle 20%). Previous research has shown (e.g. Schwartz 2020 on *Vowel Inherent Spectral Change*; see also Jibson 2021) that choosing static points, as opposed to dynamic ranges is suboptimal for assessing vowel changes. VISC, in general, is a rather new area of study for which the acoustic measurements and statistical analyses (e.g. Generalised Additive Mixed Modelling which allow for smooth functions to be incorporated into regression; Rupert et al. 2003) are still being refined. It is hoped that in the future, the section on vowels will be re-analysed and enriched with more detail.

Nevertheless, it is hoped that the thesis has presented the acoustic data gathered thus far and the phonological implications thereof in a way that the reader will find interesting and insightful, despite how the work that still remains.

Chapter 6: The results – L1

6.1. The aim of the chapter

Having discussed both the theoretical background as well as the methodology of the empirical study conducted for the purposes of this dissertation, it is now time to discuss the results. As has been shown in the previous sections, there is a vast number of acoustic parameters and groups that are to be compared. Therefore, to ensure that the presentation of the obtained results is clear, it will be organised as follows.

This chapter focuses solely on L1 Polish data. The results are divided into the discussion according to the acoustic parameters under consideration, that is VOT, f_0, F1 at the onset of the vowel and F1 and F2 over the entire duration of the vowel. These main sections are subdivided into subsections devoted to task type, namely word and sentence lists. For each parameter, we will first discuss the results of Group 1, as they were the group that took part in the longitudinal portion of the study and compare them with monolinguals. After that, we will include Group 2 and Group 3 as comparison groups[69].

6.2. Laryngeal contrast in L1 Polish

This section describes the results of the study with VOT as the dependent variable. First we present group results collapsed across place of articulation, after which we look at the effects of POA.

6.2.1. VOT of the voiceless series in Polish

In this section we discuss the voiceless series only, that is /p, t, k/, in both the word reading and sentence reading task.

[69] This is done purely for the purposes of the clarity of presentation; all groups were included in one analysis.

6.2.1.1. Word reading task

On the whole, 1715 items were analysed from Group 1, that is the first year students (576 at T1, 572 at T2, and 567 at T3) and compared with 359 tokens obtained from the Monolingual group (i.e. Group 4). Fig. 39 below illustrates the mean durations of the voiceless consonants, collapsed across place of articulation.

Fig. 39. Mean VOT of /p, t, k/ in Group 1's productions across the three testing times compared with Group 4 (word-reading task).

As can be observed, at T1 the mean duration of /p, t, k/ was 41.64 ms (SD=17) and did not differ much from the monolingual norms, which was 41.70 ms (SD=16). At T2, the mean VOT underwent slight lengthening, to 43.65 ms (SD=17), before shortening to 41.75 ms (SD=17) at T3.

The results of the Generalised Linear Mixed Models with VOT as the dependent variable, Group*Session as the main interaction and Speaker and Item as random factors are shown in Table 16. The table describes pairwise comparisons, as it is the comparison between the sessions that is of main interest to us. Full coefficient tables are included in Appendix 7. The significant contrasts are bolded.

Table 16. The results of the GLMM with VOT of /p, t, k/ as the dependent variable of Group 1 (pairwise comparisons; word-reading task).

Session	Contrast estimate	Std. Error	t	Significance
T1 vs. T2	**-2.103**	**.653**	**-3.218**	**.001**
T1 vs. T3	-.514	.655	-.784	.433
T2 vs. T3	**1.589**	**.656**	**2.422**	**.015**
T1 vs. Monolinguals	.607	3.249	.187	.852
T2 vs. Monolinguals	2.710	3.249	.834	.404
T3 vs. Monolinguals	1.120	3.250	.345	.730

What can be observed is that the lengthening at T2 (p=.001) as well as the subsequent shortening (p=.015) was significant; no difference was observed between T1 and T3. Interestingly, at no point (even at T2) did the VOT values of voiceless plosives produced by the students undergoing the phonetic training diverge from monolingual norms. Therefore, the pronunciation training in L2 did not seem to have an effect of L1 /p, t, k/, seeing as the differences between the VOT mean values were relatively minute (ca. 2 ms).

Going deeper, Fig. 40 shows the visual representation of mean VOT values in the productions of Group 1 and Group 4, sorted for place of articulation.

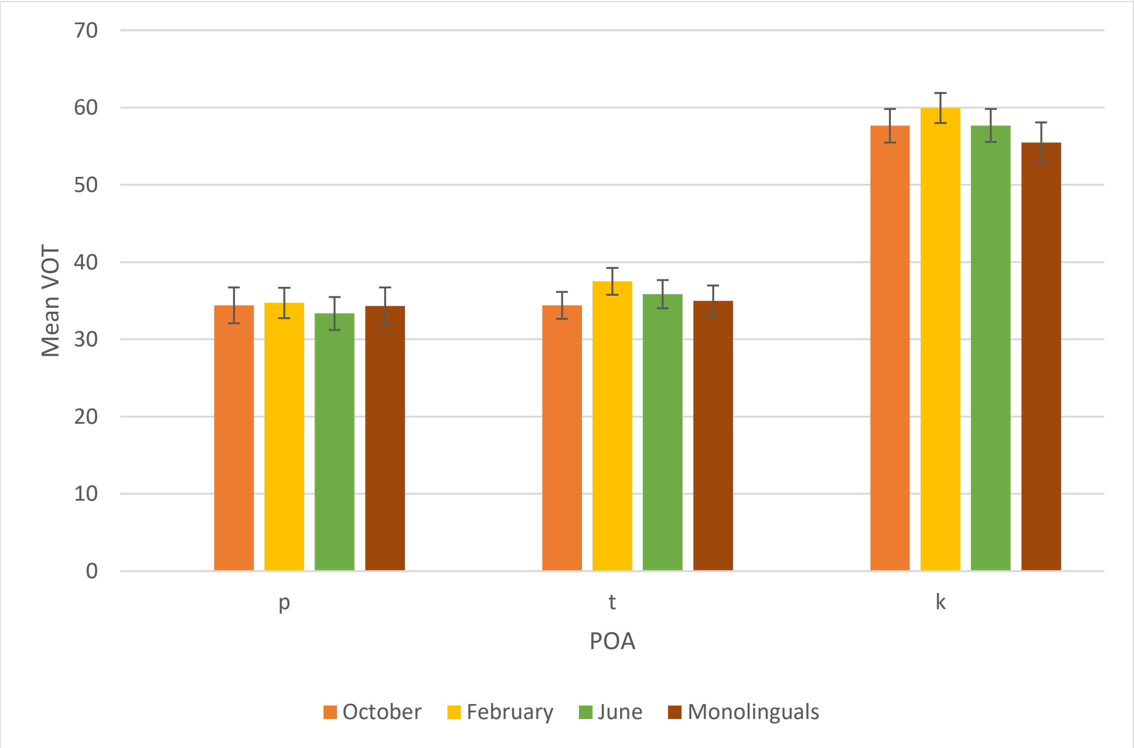


Fig. 40. Mean VOT of /p, t, k/ in Group 1's productions across the three testing times compared with Group 4, sorted for place of articulation (word-reading task).

Table 17 provides the number of productions of each segment, alongside the mean VOT value and Standard Deviations.

Table 17. Mean VOT duration of /p, t, k/ in Group 1 and 4's productions alongside Standard Deviation, sorted for place of articulation (word-reading task).

Place of articulation	Session/Group	N	Mean (ms)	Std. Deviation
/p/	T1	178	34.39	15.71
	T2	174	34.72	13.23
	T3	177	33.36	14.34
	Monolinguals	105	34.33	12.35
/t/	T1	219	34.41	13.01
	T2	220	37.50	13.12
	T3	217	35.87	13.64
	Monolinguals	133	34.97	11.57
/k/	T1	179	57.66	13.01
	T2	178	59.95	13.24
	T3	173	57.69	14.19
	Monolinguals	121	55.48	14.55

As Table 17 illustrates, the VOT values of /p/ and /t/ are relatively comparable; the mean durations of /k/ are noticeably longer, which is to be expected for the velar place of articulation (Maddieson 1997; Cho and Ladefoged 1999). The results of the GLMM model for the interaction between VOT and place of articulation are given in Table 18. The significant differences are bolded and the full coefficient table is provided in Appendix 7.

Table 18. The results of the GLMM with VOT of /p, t, k/*place of articulation as the dependent variable (pairwise comparisons: Groups 1 and 4; word-reading task).

Place of articulation	Session/Group	Contrast estimate	Std. Error	t	Significance
/p/	T1 vs. T2	-.450	1.180	-.381	.703
	T1 vs. T3	.672	1.175	.572	.567
	T2 vs. T3	1.122	1.182	.949	.343
	T1 vs. Monolinguals	-.018	3.434	-.005	.996
	T2 vs. Monolinguals	.431	3.436	.126	.900
	T3 vs. Monolinguals	-.691	3.435	-.201	.841
/t/	**T1 vs. T2**	**-3.097**	**1.057**	**-2.931**	**.003**
	T1 vs. T3	-1.435	1.060	-1.354	.176
	T2 vs. T3	1.662	1.059	1.569	.117
	T1 vs. Monolinguals	-.325	3.377	-.096	.923
	T2 vs. Monolinguals	2.772	3.376	.821	.412
	T3 vs. Monolinguals	1.110	3.377	.329	.742

/k/	**T1 vs. T2**	**-2.499**	**1.172**	**-2.133**	**.033**
	T1 vs. T3	-.573	1.181	-.486	.627
	T2 vs. T3	1.926	1.182	1.629	.103
	T1 vs. Monolinguals	2.177	3.408	.639	.523
	T2 vs. Monolinguals	4.676	3.409	1.372	.170
	T3 vs. Monolinguals	2.750	3.412	.806	.420

The statistical analysis revealed significant differences only in two instances: both /t/ (*p*=.003) and /k/ (*p*=.033) got longer between T1 and T2. However, this lengthening was slight and the VOT values were not significantly different from the monolingual norms. It appears, then, that the first year students undergoing pronunciation training in L2 did not drift towards the L2 norms (that is, the mean VOT of their Polish productions did not got closer to English aspirated /p, t, k/) at any point.

Having discussed the results of Group 1 and compared them with L1 Polish monolingual norms, let us now include the two additional comparison groups – Group 2 (i.e. the one finishing a two-year long phonetic training) and Group 3 (i.e. the group which had finished their training a year prior to the recording session). There were 851 items included in the analysis in total (426 from Group 2 and 425 from Group 3). The mean VOT values of all groups are shown in Fig. 41[70].

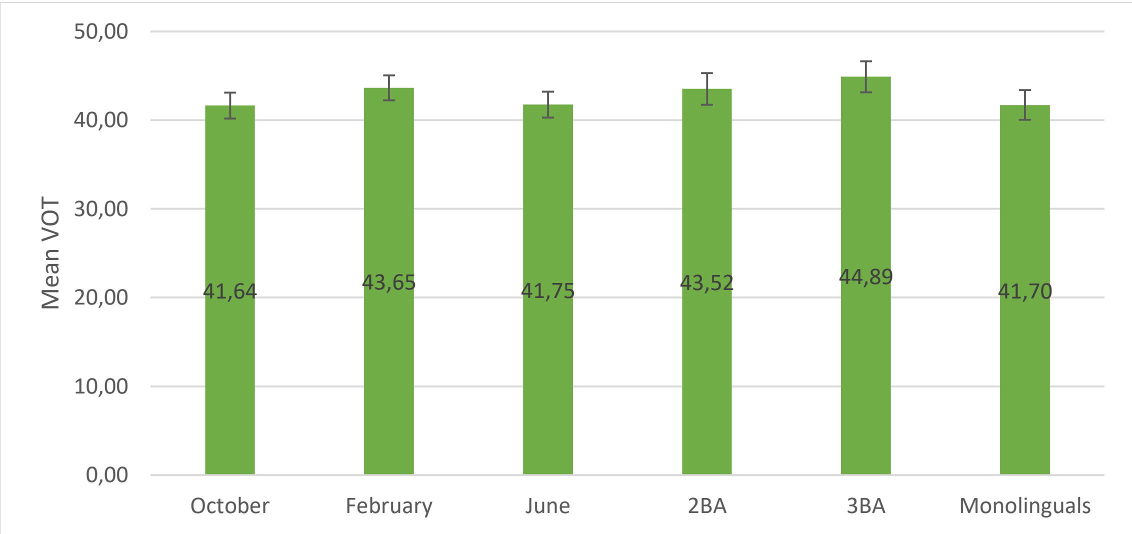


Fig. 41. Mean VOT of /p, t, k/ of students' productions (all groups) compared with Group 4 (word-reading task).

[70] In the Figure Group 2 is described as 2BA, while Group 3 as 3BA. The acronym stands for the year of study and the programme (in this case: Bachelor of Arts = BA). This will be used in all the figures to follow in order for the reader not to get too confused with which group attends which year of study.

The differences between the groups are yet again relatively minimal. Group 2's average oscillated at 43.52 ms (SD=18), while Group 4's at 44.89 ms (SD=18). The details of the statistical analysis are given in Table 19. Here we compare Groups 2 and 3 with the groups described previously. Pairwise comparisons follow.

Table 19. The results of the GLMM with VOT of /p, t, k/ as the dependent variable (pairwise comparisons between Group 2 and 3 and the rest; word-reading task).

Session	Contrast estimate	Std. Error	t	Significance
Group 2 vs. T1	2.135	3.240	.659	.510
Group 2 vs. T2	.032	3.241	.010	.992
Group 2 vs. T3	1.621	3.241	.500	.617
Group 2 vs. Monolinguals	2.742	3.473	.790	.430
Group 3 vs. T1	3.060	3.241	.944	.345
Group 3 vs. T2	.957	3.241	.295	.768
Group 3 vs. T3	2.546	3.241	.786	.432
Group 3 vs. Group 2	.925	3.465	.267	.790
Group 3 vs. Monolinguals	3.667	3.473	1.056	.291

On the basis of the table we see that with the results collapsed across three places of articulation, no significant contrasts between Groups 2 and 3 and the rest were present. When we subdivide the mean durations according to POA, the results look similar. This is illustrated in Fig. 42.

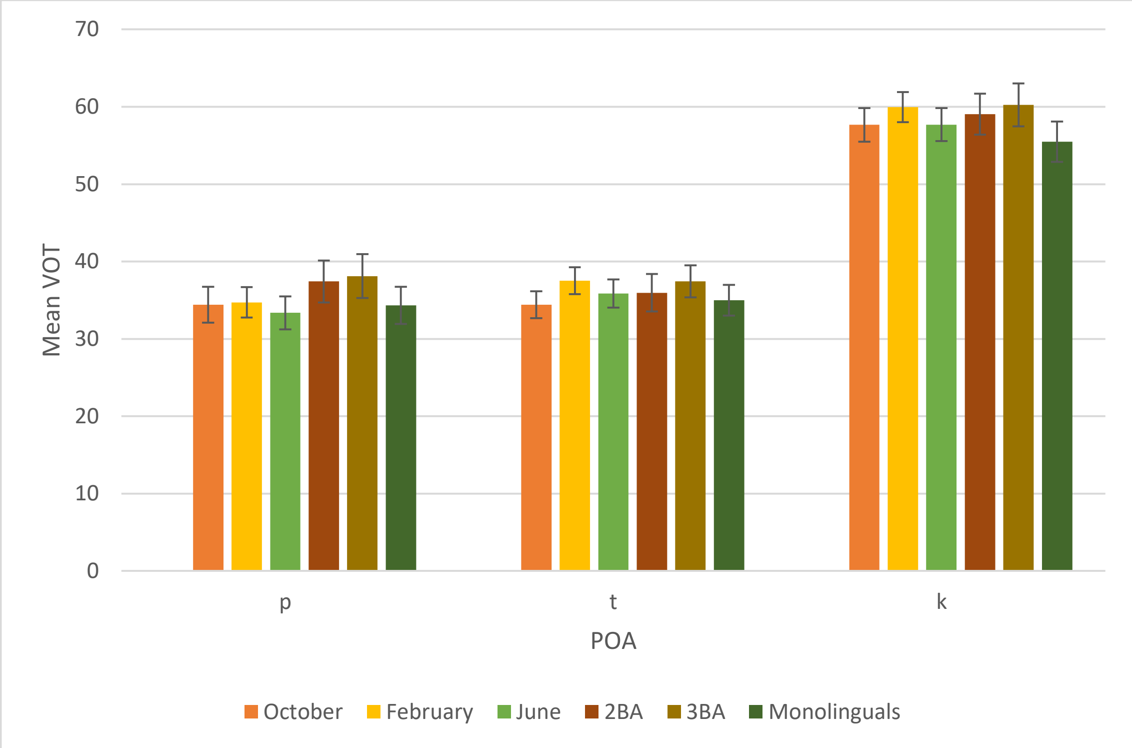


Fig. 42. Mean VOT of /p, t, k/ of all Groups, sorted for place of articulation (word-reading task).

Table 20 provides the results in more details, including the number of observations and Standard Deviations. As Table 17 has already shown these measurements for Groups 1 and 4, here we limit ourselves to Group 2 and 3 only.

Table 20. Mean VOT duration of /p, t, k/ in Group 2 and 3's productions alongside Standard Deviation, sorted for place of articulation (word-reading task).

Place of articulation	Session/Group	N	Mean (ms)	Std. Deviation
/p/	Group 2	132	37.42	15.71
	Group 3	129	38.11	16.27
/t/	Group 2	163	35.95	15.61
	Group 3	161	37.43	13.30
/k/	Group 2	131	59.06	15.29
	Group 3	135	60.23	16.31

Finally, Table 21 presents the results of the GLMM model, with the VOT*POA interaction as the main predictor of interest. We again compare Groups 2 and 3 with the ones already described. Pairwise comparisons follow.

Table 21. The results of the GLMM with VOT of /p, t, k/ as the dependent variable and place of articulation*type as the main interaction (pairwise comparisons: Groups 2 and 3 vs. Groups 1 and 4; word-reading task).

Place of articulation	Session/Group	Contrast estimate	Std. Error	t	Significance
/p/	Group 2 vs. T1	2.967	3.396	.874	.382
	Group 2 vs. T2	2.517	3.398	.741	.459
	Group 2 vs. T3	3.639	3.987	1.071	.284
	Group 2 vs. Monolinguals	2.948	3.667	.804	.421
	Group 3 vs. T1	3.801	3.400	1.118	.264
	Group 3 vs. T2	3.351	3.402	.985	.325
	Group 3 vs. T3	4.473	3.400	1.316	.188
	Group 3 vs. Group 2	.834	3.635	.229	.819
	Group 3 vs. Monolinguals	3.783	3.670	1.031	.303
/t/	Group 2 vs. T1	1.659	3.351	.495	.621
	Group 2 vs. T2	-1.438	3.350	-.429	.668
	Group 2 vs. T3	.224	3.352	.067	.947
	Group 2 vs. Monolinguals	1.134	3.607	.370	.711
	Group 3 vs. T1	2.812	3.352	.839	.402
	Group 3 vs. T2	-.285	3.352	-.085	.932
	Group 3 vs. T3	1.377	3.353	.411	.681
	Group 3 vs. Group 2	1.153	3.584	.322	.748
	Group 3 vs. Monolinguals	2.487	3.608	.689	.491
/k/	Group 2 vs. T1	1.895	3.397	.558	.577
	Group 2 vs. T2	-.604	3.397	-.178	.859
	Group 2 vs. T3	1.322	3.400	.389	.697
	Group 2 vs. Monolinguals	4.072	3.645	1.117	.264
	Group 3 vs. T1	2.647	3.392	.780	.435
	Group 3 vs. T2	.148	3.393	.044	.965
	Group 3 vs. T3	2.074	3.396	.611	.541
	Group 3 vs. Group 2	.752	3.630	.207	.836
	Group 3 vs. Monolinguals	4.824	3.641	1.325	.185

No significant differences were observed. It can be concluded that when it comes to the voiceless series in the word reading task, overall little to no phonetic drift effects have been found as far as the parameter of VOT is concerned.

6.2.1.2. Sentence reading task

Now let us turn to the sentence list. One will notice that, on the whole, the mean VOT values in this task were shorter than what we saw in citation forms' productions. It has been shown that mono/di syllabic words induce longer VOT values, relative to longer utterances (Yu et al. 2015). In general, the longer the stretch of speech, the shorter the VOT (Lisker and Abramson 1964).

We collected 1876 items from Group 1, that is our first year students, (616 at T1, 629 at T2, and 631 at T3). They were subsequently compared with 626 tokens obtained from the group of quasi-monolinguals (i.e. Group 4). As you may recall, we distinguished between three different prosodic positions: utterance-initial, phrase-initial (i.e. utterance-medial), and phrase medial. Fig. 43 illustrates the mean averages of the positive VOT sorted for session and the target word's position within the sentence for Groups 1 and 4[71].

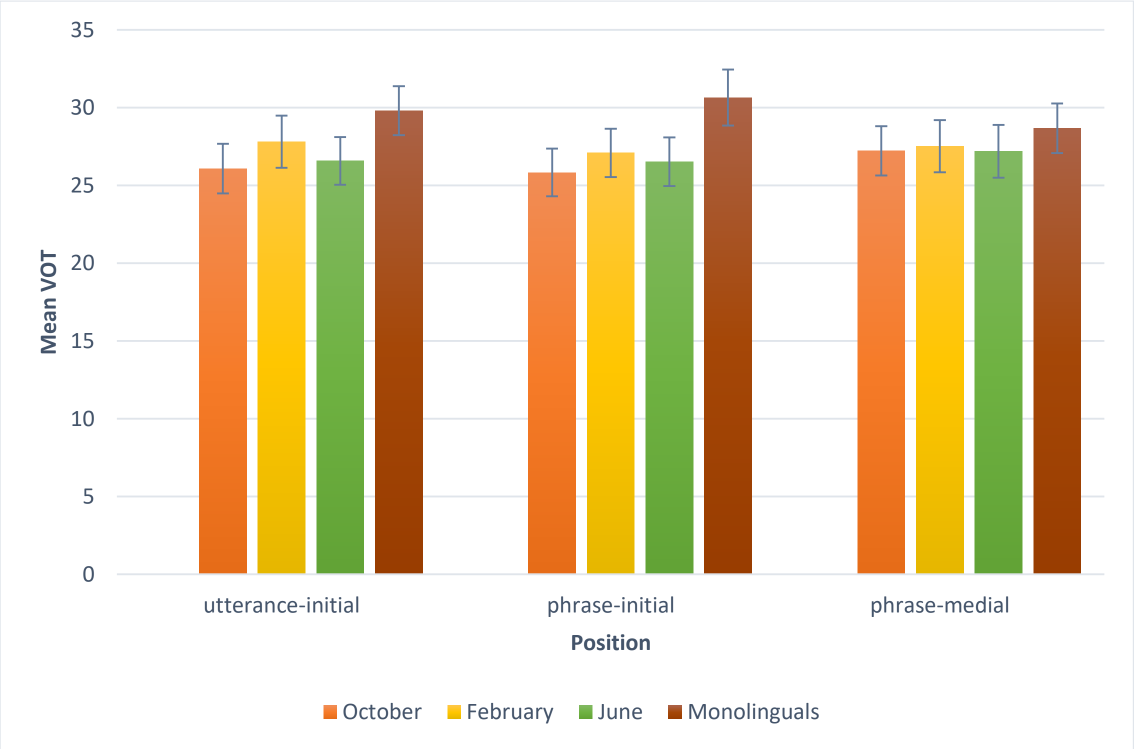


Fig. 43. Mean VOT of /p, t, k/-initial target words in sentence production task, sorted for session and prosodic position (Groups 1 and 4 only; sentence-reading task).

The specific averages are given in Table 22, alongside the specific numbers of tokens and Standard Deviations.

[71] Seeing as the sentence list was counterbalanced for place of articulation and in the word list task there were no interesting interactions observed, we do not subdivide the items according to their POAs. In general, the expected pattern was observed – the shortest VOT was yielded by labials and the longest by velars.

Table 22. Mean VOT of /p, t, k/-initial target words in sentence production task with Standard Deviations, sorted for session and prosodic position (Groups 1 and 4 only; sentence-reading task).

Prosodic position	Session/Group	N	Mean (ms)	Std. Deviation
utterance-initial	T1	195	26.09	11.28
	T2	200	27.81	12.02
	T3	198	26.58	10.95
	Monolinguals	212	29.81	11.61
phrase-initial	T1	210	25.83	11.27
	T2	218	27.10	11.64
	T3	217	26.53	11.74
	Monolinguals	212	30.65	13.34
phrase-medial	T1	211	27.23	11.67
	T2	211	27.52	12.31
	T3	216	27.19	12.64
	Monolinguals	214	28.69	11.84

There are two main things which we can see in Fig. 43 and Table 22. First of all, there does not appear to be too much variation across the three prosodic positions – the VOT values are comparable and we do not observe any domain-initial lengthening of this parameter in Polish productions. Second of all, when we focus on the three testing times and compare the VOT averages we will notice that on the whole the VOT values of Group 4's productions were in fact longer than the Students' at any of the recording sessions. The general trend of VOT being the longest, which we saw in the previous task, appears to be present as well.

A GLMM analysis was run with VOT as the dependent variable and the interaction between Session*Prosodic_position as the main predictor. Since VOT is a temporal cue, we included Speech rate as a control variable, and Speaker and Item as random factors. The results of pairwise comparisons between Groups 1 and 4 are shown in Table 23[72]. The significant contrasts are bolded. A full coefficient table is available in Appendix 7.

[72] Similarly to the word list task, this is done for clarity purposes only. All groups were included in the same statystical analysis.

Table 23. The results of the GLMM with VOT of /p, t, k/ as the dependent variable and Session*Prosodic_position as the main predictor (pairwise comparisons: Groups 1 and 4; sentence-reading task).

Prosodic position	Session/Group	Contrast estimate	Std. Error	t	Significance
utterance-initial	T1 vs. T2	-1.140	.691	-1.650	.099
	T1 vs. T3	-.389	.695	-.560	.575
	T2 vs. T3	.751	.690	1.087	.277
	T1 vs. Monolinguals	**-4.306**	**1.449**	**-2.972**	**.003**
	T2 vs. Monolinguals	**-3.166**	**1.446**	**-2.190**	**.029**
	T3 vs. Monolinguals	**-3.917**	**1.449**	**-2.703**	**.007**
phrase-initial	T1 vs. T2	-1.057	.664	-1.593	.111
	T1 vs. T3	-.641	.666	-.962	.336
	T2 vs. T3	.416	.660	.630	.529
	T1 vs. Monolinguals	**-4.130**	**1.446**	**-2.855**	**.004**
	T2 vs. Monolinguals	**-3.073**	**1.442**	**-2.131**	**.033**
	T3 vs. Monolinguals	**-3.489**	**1.444**	**-2.416**	**.016**
phrase-medial	T1 vs. T2	-.318	.668	-.476	.634
	T1 vs. T3	.331	.667	.497	.619
	T2 vs. T3	.649	.667	.973	.331
	T1 vs. Monolinguals	-.765	1.456	-.525	.600
	T2 vs. Monolinguals	-.447	1.455	-.307	.759
	T3 vs. Monolinguals	-1.096	1.458	-.751	.452

The statistical analysis revealed that, in general, we do not observe effects of phonetic training in the productions of first year students – at all testing times the differences in mean VOT values, when present, were small in magnitude and not statistically significant. However, in utterance-initial and phrase-initial positions we see that the monolinguals' produced VOT significantly longer than the students at any testing time. It appears that the small domain-initial lengthening effects were present in the speech of quasi-monolinguals' but not students'. These domain-initial lengthening effects are not expected in phrase-medial positions (as no prosodic break precedes the target word) and as a result, the VOT values of Group 4 did not differ from the students'. The differences were nonetheless relatively small. Additionally, speech rate proved to be significant (*p*=.009).

Now let us have a look at what happens when we include our two comparison groups: Group 2 (i.e. second year students; 458 tokens) and Group 3 (i.e. third year students; 465 tokens). Fig. 44 shows the mean VOT values of all four groups.

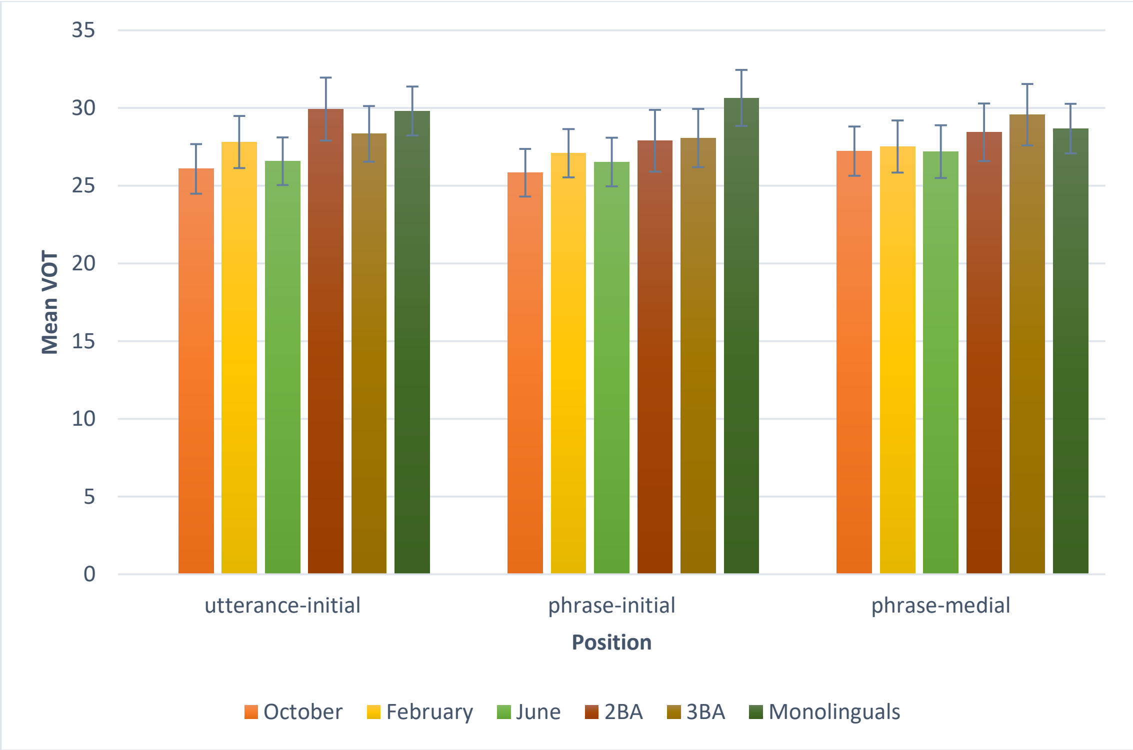


Fig. 44. Mean VOT of /p, t, k/-initial target words in sentence production task, sorted for session and prosodic position (all groups; sentence-reading task).

The average VOT values obtained by Groups 2 and 3 are given in Table 24. The means of these groups in this task were once again shorter than what was observed as far as citation forms go.

Table 24. Mean VOT of /p, t, k/-initial target words in sentence production task with Standard Deviations, sorted for session and prosodic position (Groups 2 and 3 only; sentence-reading task).

Prosodic position	Session/Group	N	Mean (ms)	Std. Deviation
utterance-initial	Group 2	143	29.93	12.31
	Group 3	147	28.34	10.98
phrase-initial	Group 2	153	27.90	12.44
	Group 3	159	28.07	11.90
phrase-medial	Group 2	162	28.45	11.93
	Group 3	159	29.58	12.61

Immediately one can notice that the differences between the two groups across all three prosodic positions are minute. Table 25 provides the results of the GLMM analysis, the details of which have already been provided. Pairwise comparisons follow, with the significant contrasts in bold.

Table 25. The results of the GLMM with VOT of /p, t, k/ as the dependent variable and Session*Prosodic_position as the main predictor (pairwise comparisons: Groups 2 and 3 vs. the rest; sentence-reading task).

Prosodic position	Session/Group	Contrast estimate	Std. Error	t	Significance
utterance-initial	**Group 2 vs. T1**	**4.155**	**1.573**	**2.642**	**.008**
	Group 2 vs. T2	3.015	1.571	1.920	.055
	Group 2 vs. T3	**3.766**	**1.572**	**2.395**	**.017**
	Group 2 vs. Monolinguals	-.151	1.568	-.097	.923
	Group 3 vs. T1	2.246	1.570	1.431	.153
	Group 3 vs. T2	1.106	1.568	.706	.480
	Group 3 vs. T3	1.857	1.569	.706	.480
	Group 3 vs. Group 2	-1.909	1.680	-1.136	.256
	Group 3 vs. Monolinguals	-2.060	1.565	-1.317	.188
phrase-initial	Group 2 vs. T1	2.605	1.561	1.669	.095
	Group 2 vs. T2	1.549	1.558	.994	.320
	Group 2 vs. T3	1.965	1.560	1.260	.208
	Group 2 vs. Monolinguals	-1.524	1.567	-.972	.331
	Group 3 vs. T1	2.188	1.557	1,405	.160
	Group 3 vs. T2	1.131	1.554	.728	.467
	Group 3 vs. T3	1.547	1.555	.994	.320
	Group 3 vs. Group 2	-.418	1.667	-.251	.802
	Group 3 vs. Monolinguals	-1.942	1.562	-1.243	.214
phrase-medial	Group 2 vs. T1	1.104	1.555	.710	.478
	Group 2 vs. T2	.786	1.555	.505	.613
	Group 2 vs. T3	1.435	1.554	.924	.356
	Group 2 vs. Monolinguals	.339	1.575	.215	.830
	Group 3 vs. T1	2.209	1.557	.215	.830
	Group 3 vs. T2	1.891	1.557	1.215	.224
	Group 3 vs. T3	2.540	1.556	1.633	.103
	Group 3 vs. Group 2	1.105	1.661	.665	.506
	Group 3 vs. Monolinguals	1.445	1.577	.916	.360

All in all, significant contrasts were found only in the utterance-initial voiceless plosives between Group 2 and Group 1 at T1 (*p*=.008) and T3 (*p*=.017). Therefore, in general the voiceless stops produced in longer utterances in the speech of second and third year students do not diverge from monolingual norms.

To sum up the results of /p, t, k/-initial target words in sentence reading task, we may state that while some contrasts were detected, the effects of phonetic training on L1 stop production were scarce. We will return to these results and compare them with the voiced series in section 6.2.3.

6.2.2. Realisation of voiced plosives in Polish

Recall that in section 5.9 we distinguished between three distinct types of voiced series' realisation, namely Type 1 (i.e. fully pre-voiced), Type 2 (i.e. partially pre-voiced), and Type 3 (i.e. unvoiced). The length of negative VOT from the possible perspective of L2-induced phonetic drift will be discussed for fully pre-voiced (that is, Type 1) tokens only; with Type 2[73] and Type 3 we will focus on the likelihood of their occurrence and whether or not this can be seen as evidence of phonetic drift.

6.2.2.1. Word reading task

Let us start with Type 1 tokens in the word reading task. In total, there were 1034 Type 1 productions analysed for Group 1 (393 at T1, 323 at T2, and 318 at T3) and compared with 344 obtained from the group of Monolinguals (that is, Group 4). The mean duration of pre-voicing for each testing time is visualised in Fig. 45, collapsed across the places of articulation.

[73] In general, the patterns observed in Type 2 tokens mirror Type 1, but their occurrence (as will be shown shortly) is not evidence of drift in itself. For this reason, reporting on them in this chapter does not contribute to the discussion. However, comparing Type 1 and Type 2 tokens has interesting implications for phonological theory – this will be shown and explained in more detail in section 8.2.

Fig. 45. Mean negative VOT of /b, d, g/ in Group 1's productions across the three testing times compared with Group 4 (Type 1 only; word-reading task).

Similarly to what was the case for the voiceless series, we might notice a slight lengthening of pre-voicing at T2 – while the mean at T1 is -80.37 ms (SD=22), at T2 the average is -83.82 (SD=23). At T3, the duration of pre-voicing is shorter than both at T1 and T3 with -78.87 ms (SD=24). Notice that at all three testing times the mean values are quite saliently shorter in comparison to monolingual norms, that is -92.05 ms (SD=26). The GLMM results comparing Groups 1 and 4 are given in Table 26. Pairwise comparisons follow, with significant contrast in bold. The full coefficient table might be found in Appendix 7.

Table 26. The results of the GLMM with the negative VOT of /b, d, g/ (Type 1) as the dependent variable (pairwise comparisons; word-reading task).

Session	Contrast estimate	Std. Error	t	Significance
T1 vs. T2	**3.762**	**1.546**	**2.434**	**.015**
T1 vs. T3	-1.240	1.553	-.798	.425
T2 vs. T3	**-5.002**	**1.626**	**-3.077**	**.002**
T1 vs. Monolinguals	**11.485**	**4.926**	**2.332**	**.02**
T2 vs. Monolinguals	7.723	4.953	1.559	.119
T3 vs. Monolinguals	**12.725**	**4.954**	**2.569**	**.01**

The lengthening of pre-voicing at T2 was significant – there was a contrast between T1 and T2 (*p*=.015) as well as T2 and T3 (*p*=.002), at which point pre-voicing underwent shortening. Interestingly, while when compared with the monolingual norms, the students' pre-voicing values were significantly shorter at the onset of phonetic training (*p*=.02) as well as towards the end of the academic year (*p*=.01), the lengthening that was observed at T2 resulted in the contrast between the students and the monolinguals to disappear.

The specific changes over time within Group 1, across the three places of articulation, compared with the monolingual group, are illustrated in Fig. 46.

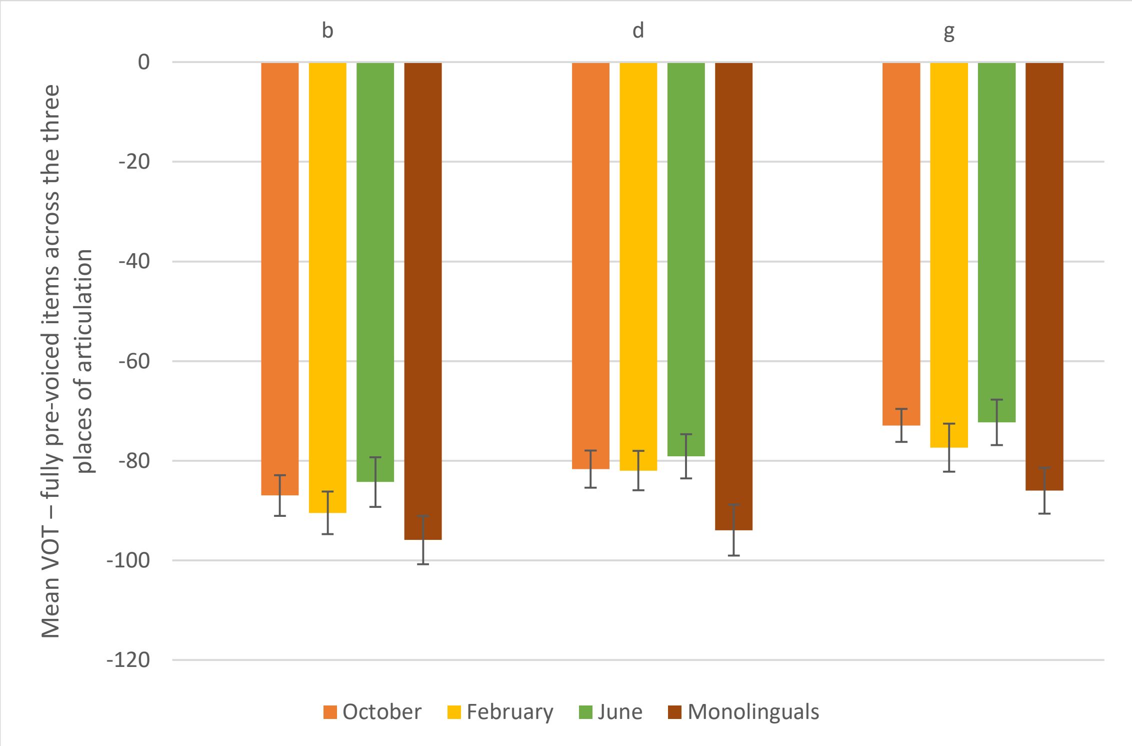


Fig. 46. Mean negative VOT of fully pre-voiced /b, d, g/ in Group 1's productions across the three testing times compared with Group 4, sorted for place of articulation (word-reading task).

In order to provide more details, Table 27 presents the number of tokens obtained for each place of articulation, as well as the mean pre-voicing duration and Standard Deviations. Recall that the numbers are given for Type 1 tokens only at this point.

Table 27. Mean VOT duration of /b, d, g/ in Group 1 and 4's productions alongside Standard Deviation, sorted for place of articulation (Type 1 only; word-reading task).

Place of articulation	Session/Group	N	Mean (ms)	Std. Deviation
/b/	T1	128	-86.96	23.30
	T2	123	-90.44	24.00
	T3	116	-84.26	26.97
	Monolinguals	119	-95.88	26.76
/d/	T1	130	-81.65	21.40
	T2	103	-81.96	20.31
	T3	103	-79.12	22.67
	Monolinguals	114	-93.93	27.52
/g/	T1	135	-72.89	19.47
	T2	97	-77.36	23.84
	T3	99	-72.30	22.94
	Monolinguals	111	-86.00	24.47

The results of the GLMM model with the interaction of session*POA as the main predictor of interest are shown in Table 28. Pairwise comparisons follow, with the significant contrasts bolded. Full coefficient table, as always, is provided in Appendix 7.

Table 28. The results of the GLMM with VOT of /b, d, g/ as the dependent variable and place of articulation*type as the main interaction (Type 1 only; pairwise comparisons: Groups 1 and 4; word-reading task).

Place of articulation	Session/Group	Contrast estimate	Std. Error	t	Significance
/b/	T1 vs. T2	3.877	2.581	1.502	.133
	T1 vs. T3	-1.982	2.622	-.756	.450
	T2 vs. T3	**-5.859**	**2.646**	**-2.214**	**.027**
	T1 vs. Monolinguals	8.590	5.395	1.592	.111
	T2 vs. Monolinguals	4.713	5.408	.871	.384
	T3 vs. Monolinguals	10.572	5.426	1.948	.051
/d/	T1 vs. T2	1.706	2.701	.632	.528
	T1 vs. T3	-2.621	2.701	-.970	.332
	T2 vs. T3	-4.326	2.851	-1.518	.129
	T1 vs. Monolinguals	**12.858**	**5.403**	**2.380**	**.017**
	T2 vs. Monolinguals	11.153	5.484	2.034	.042
	T3 vs. Monolinguals	**15.479**	**5.484**	**2.822**	**.005**
/g/	T1 vs. T2	5.638	2.727	2.067	.039
	T1 vs. T3	.791	2.711	.292	.770
	T2 vs. T3	-4.847	2.930	-1.654	.098
	T1 vs. Monolinguals	**13.368**	**5.400**	**2.476**	**.013**
	T2 vs. Monolinguals	7.729	5.518	1.401	.161
	T3 vs. Monolinguals	**12.577**	**5.508**	**2.129**	**.022**

On the basis of the analysis we can see that as far as fully pre-voiced /b/ is concerned, the only statistically significant change occurred between February and June, when pre-voicing got shorter (*p*=.027), however that shortening failed to yield a significant contrast between T3 and Monolinguals (*p*=.051). On the whole, the duration of pre-voicing of /b/ did not differ between any of the testing times and the control group.

In the case of fully pre-voiced /d/, while no significant changes occurred across the testing times, each testing session of the 1BA group differed from the durations obtained from the monolinguals. That is, already at T1 pre-voicing values of the coronal stop were shorter than what we see in monolingual norms, and this trend persisted throughout the duration of the training.

As for the velar place of articulation, there was a significant lengthening of full pre-voicing between T1 and T2 (*p*=.039), resulting in an obliteration of contrast between the students and the monolinguals in February (*p*=.161). Aside from this, the students' generally differed from the monolingual group, with pre-voicing values shorter both in October (*p*=.013) and June (*p*=.022).

Now let us look at full pre-voicing values when we include the two remaining comparison groups, that is Groups 2 (2BA) and 3 (3BA). In sum, there were 202 Type 1 tokens obtained from Group 2 and 195 from Group 3. Fig. 47 illustrates the mean pre-voicing durations for each group, collapsed across places of articulation.

Fig. 47. Mean negative VOT of fully pre-voiced /b, d, g/ of students' productions (all groups) compared with Group 4 (word-reading task).

At first glance we can see that Group 2 appears to showcase shortening of negative VOT, with mean duration of -74.42 ms (SD=23), whereas in the case of Group 3 the average length of pre-voicing is very close to the monolingual norms, reaching -91.42 ms (SD=28) (cf. Schwartz 2020; Chang 2013 on how phonetic drift effects may cease once the explicit instruction stops). The details of the statistical analysis are given in Table 29. The results of the GLMM model for Type 1 tokens here compare Groups 2 and 3 with the remaining ones which have been discussed above. Pairwise comparisons follow, with significant contrasts in bold.

Table 29. The results of the GLMM with negative VOT of /b, d, g/ (Type 1) as the dependent variable (pairwise comparisons between Groups 2 and 3 and the rest; word-reading task).

Session	Contrast estimate	Std. Error	t	Significance
Group 2 vs. T1	**6.668**	**5.023**	**1.328**	**.01**
Group 2 vs. T2	**10.430**	**5.049**	**2.066**	**.039**
Group 2 vs. T3	5.428	5.050	1.075	.283
Group 2 vs. Monolinguals	**18.153**	**5.340**	**3.399**	**.001**
Group 3 vs. T1	**-11.888**	**5.043**	**-2.357**	**.018**
Group 3 vs. T2	-8.126	5.070	-1.603	.109
Group 3 vs. T3	**-13.128**	**5.071**	**-2.589**	**.01**
Group 3 vs. Group 2	**-18.556**	**5.450**	**-3.405**	**.001**
Group 3 vs. Monolinguals	-.403	5.361	-.075	.940

Indeed, the mean duration of negative VOT in Type 1 voiced plosives in Group 2 productions appears to be significantly shorter than what was observed for Group 1 at both T1 (*p*=.01) and T2 (*p*=.039). There is no contrast between T3 and Group 2 (*p*=.283), however the difference between Group 2's productions and the monolingual norms is maintained (*p*=.001). On the other hand, there is no contrast between Group 3 and Group 1 at T2 (that is, February recording session; *p*=.109) as well as the monolingual norms (*p*=.940). Therefore while the shortening of pre-voicing appears to peak at 2BA, after one year of no pronunciation training the negative VOT values seem to return to norms.

When we sort the items for place of articulation, the mean durations of negative VOT look as illustrated in Fig. 48.

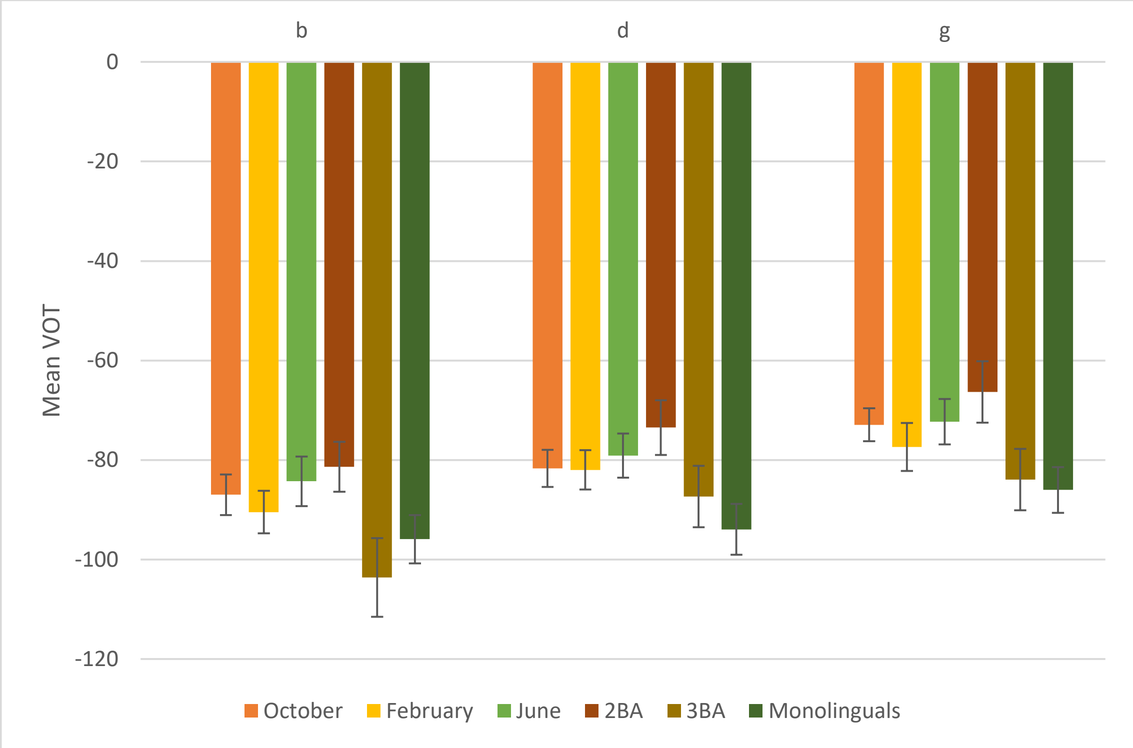


Fig. 48. Mean negative VOT of fully pre-voiced /b, d, g/ of all groups, sorted for place of articulation (word-reading task).

The detailed measurements, including the number of productions analysed, mean duration as well as Standard Deviation are given in Table 30.

Table 30. Mean VOT duration of /b, d, g/ in Group 2 and 3's productions, alongside Standard Deviation, sorted for place of articulation (Type 1 only; word-reading task).

Place of articulation	Session/Group	N	Mean (ms)	Std. Deviation
/b/	Group 2	78	-81.36	22.20
	Group 3	62	-103.60	31.15
/d/	Group 2	65	-79.12	22.67
	Group 3	71	-87.33	25.99
/g/	Group 2	59	-66.28	23.70
	Group 3	62	-83.89	24.25

The statistical analysis yielded the following results. Table 31 presents the GLMM model with the negative VOT (Type 1) and Session*POA interaction as the main predictor; the data compare Groups 2 and 3 with Groups 1 and 4 presented before. The significant contrasts of the pairwise comparisons are bolded.

Table 31. The results of the GLMM with VOT of /b, d, g/ as the dependent variable and place of articulation*type as the main interaction (Type 1 only; pairwise comparisons: Groups 2 and 3 against the rest; word-reading task).

Place of articulation	Session/Group	Contrast estimate	Std. Error	t	Significance
/b/	Group 2 vs. T1	3.674	5.570	.660	.510
	Group 2 vs. T2	7.551	5.583	1.353	.176
	Group 2 vs. T3	1.692	5.601	.302	.763
	Group 2 vs. Monolinguals	**12.264**	**5.870**	**2.089**	**.037**
	Group 3 vs. T1	**-16.796**	**5.711**	**-2.941**	**.003**
	Group 3 vs. T2	**-12.919**	**5.725**	**-2.257**	**.024**
	Group 3 vs. T3	**-19.778**	**5.744**	**-3.269**	**.001**
	Group 3 vs. Group 2	**-20.470**	**6.168**	**-3.319**	**.001**
	Group 3 vs. Monolinguals	-8.206	6.015	-1.364	.173
/d/	Group 2 vs. T1	9.681	5.672	1.707	.088
	Group 2 vs. T2	**11.387**	**5.748**	**1.981**	**.048**
	Group 2 vs. T3	7.060	5.749	1.228	.220
	Group 2 vs. Monolinguals	**22.539**	**5.986**	**3.765**	**.000**
	Group 3 vs. T1	-6.446	5.619	-1.147	.251
	Group 3 vs. T2	-4.740	5.697	-.832	.405
	Group 3 vs. T3	-0.067	5.697	-1.592	.112
	Group 3 vs. Group 2	**-16.127**	**6.181**	**-2.609**	**.009**
	Group 3 vs. Monolinguals	6.412	5.936	1.080	.280
/g/	Group 2 vs. T1	7.424	5.722	1.297	.195
	Group 2 vs. T2	**13.062**	**5.835**	**2.239**	**.025**
	Group 2 vs. T3	8.215	5.824	1.410	.159
	Group 2 vs. Monolinguals	**20.792**	**6.049**	**3.437**	**.001**
	Group 3 vs. T1	**-13.146**	**5.699**	**-2.307**	**.021**
	Group 3 vs. T2	-7.508	5.812	-1.292	.197
	Group 3 vs. T3	**-12.355**	**5.802**	**-2.129**	**.033**
	Group 3 vs. Group 2	**-20.570**	**6.318**	**-3.256**	**.001**
	Group 3 vs. Monolinguals	.222	6.028	.037	.971

We can see that in the case of labial stops, Groups 1 and 2 do not really differ. There is however a significant contrast between Group 2 and Monolinguals (*p*=.037) as well as between Groups 2 and 3 (*p*=.001). Additionally, the difference between Group 3 and Group 1 at all testing times has turned out to be significant; however, no contrast was found between Group 3 and Monolinguals (*p*=.137).

In the coronal place of articulation only three significant contrasts were found: between Group 2 and Group 3 (*p*=.009), Group 2 and Group 4 (*p*=.000) as well as Group 2 and Group 1 in February, but only just (*p*=.048).

Finally, the velar stop was produced differently by Group 2 when compared with Group 3 (*p*=.001), Group 4 (*p*=.001), and Group 1 in February (*p*=.025). Group 3, who displayed generally longer pre-voicing, differed from Group 1 in both October (*p*=.021) and June (*p*=.033).

Overall, in general we see that at all places of articulation the trend was similar: the negative shortening underwent progressive shortening throughout the two years of pronunciation training (excluding the lengthening observed unanimously in February), peaking in the productions of 2BA students, and then – after the training has been over for exactly a year – the values of pre-voicing get back to fall within monolingual norms.

Having discussed Type 1 realisations, let us now turn to Type 2, illustrated in section 5.9 (in particular, recall Fig. 33). In general, partially pre-voiced items, though they were omnipresent in the analysis, were less numerous than Type 1s. However, as has been pointed out before, we wanted to distinguish between those two types of realisations as treating them as identical has proven to skew the results before (e.g. Wojtkowiak and Schwartz 2022). Fig. 49 presents percentages of Type 2 realisations across all testing times[74].

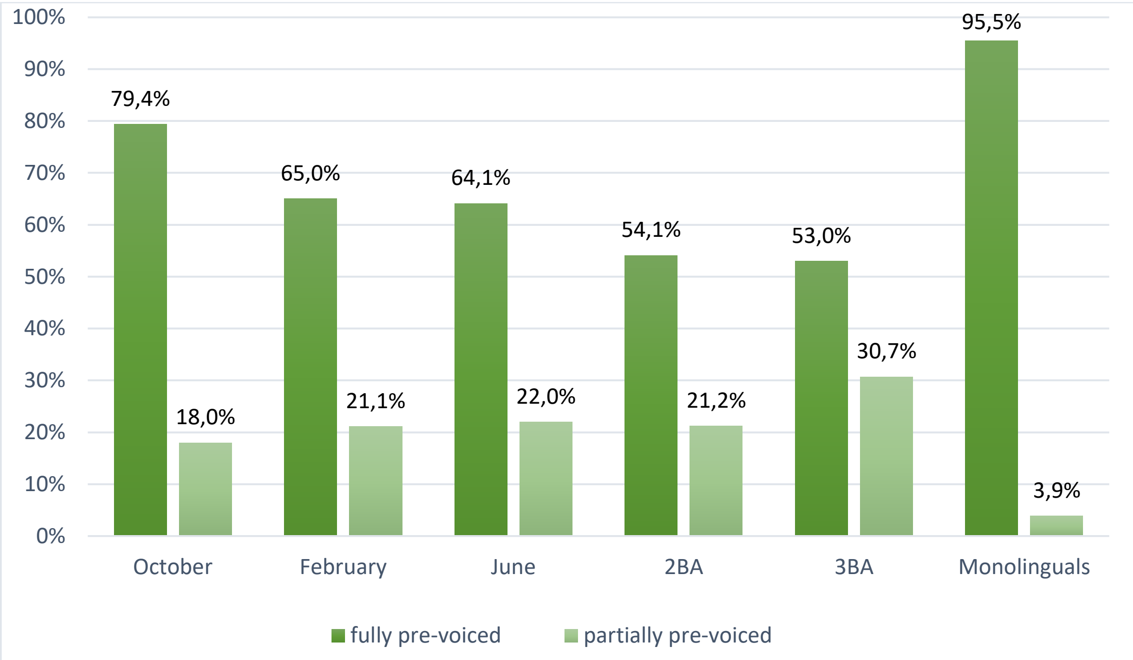


Fig. 49. Percentage of Type 1 (i.e. fully pre-voiced) realisations juxtaposed with Type 2 (i.e. partially pre-voiced) productions across all groups (word-reading task).

[74] The percentages do not add up to 100% due to Type 3 realisations (that is, unvoiced plosives) being excluded.

A binary logistic regression with Type 2 as the dependent variable (coded as *yes/no*; the break being either present or absent), Session as the main fixed factor, and Speaker and Item as random factors revealed the following. Significant contrast was observed between the control group of monolinguals and the first year students at T2 (t=2.093; p=.03), T3 (t=2.176; p=.03), and the third year students (t=2.309; p=.02). Amongst the students' groups, we found differences also between T1 and T2 (t=2.066; p=.03) and T1 and T3 (t=2.270, p=0.2). No other contrasts were found[75]. For the time being, though some differences have been attested, we are reticent to ascribe them to phonetic drift; we will see in section 6.2.2.2 that partially pre-voiced realisations in the production of sentences are quite common, also amongst monolingual speakers.

The last type of voiced plosives' realisation to be considered are Type 3 tokens, which were produced with English-like positive VOT, thus lacking vocal fold vibration during the closure phase (cf. Fig. 34). Here we were not interested in the duration of pre-voicing (as there was none) but rather with the percentages of this type of realisations across all groups and testing times, similarly to what was the case for Type 2 items. This is shown in Fig. 50.

Fig. 50. The percentage of pre-voiced (both fully and partially) vs. unvoiced realisation of /b, d, g/ across all testing times and groups in Polish (word-reading task).

[75] The full coefficient table is available in Appendix 7.

As we can notice, the group of monolinguals which has little experience with English produced the voiced plosives overwhelmingly with the presence of pre-voicing (2 out of 360 tokens). In the longitudinal data gathered from Group 1, we can see that 2.6% of their /b, d, g/ productions lack pre-voicing already at T1 (13 out of 495), but the numbers sky-rocket when the phonetic training truly commences, reaching 13.9% at both T2 (69 out of 497) and T3 (69 out of 496). Type 3 realisations appear to peak in Group 2's productions at 24.7% of unvoiced realisations (92 out of 373) and seem to remain present in the students' productions even after the pronunciation training has ended, with 16.3% Type 3 realisations found in the speech of Group 3 participants (60 out of 368). Therefore, while pre-voicing values (in both Type 1 and Type 2 tokens) found in Group 3 showed the trend of returning to monolingual norms after one year of no explicit phonetic instruction, the presence of unvoiced productions was preserved.

The results of Binary Logistic Regression with Type 3 (binary choice: *yes* or *no*) as the main predictor and Session as the fixed factor (with Speaker and Item as random factors) are presented in Table 32. Pairwise comparisons follow and the significant results are bolded. The full coefficient table can be found in Appendix 7.

Table 32. The results of Binary Logistic Regression with Type_3 realisation of voiced plosives in Polish as the dependent variable (pairwise comparisons; word-reading task).

Session	Contrast estimate	Std. Error	t	Significance
T1 vs. T2	**.089**	**.024**	**3.622**	**.000**
T1 vs. T3	**.089**	**.025**	**3.625**	**.000**
T2 vs. T3	.000	.018	.022	.982
T1 vs. Group 2	**.198**	**.054**	**3.644**	**.000**
T2 vs. Group 2	.109	.059	1.860	.063
T3 vs. Group 2	.109	.059	1.852	.064
T1 vs. Group 3	**.104**	**.037**	**2.855**	**.004**
T2 vs. Group 3	.016	.044	.354	.724
T3 vs. Group 3	.015	.044	.344	.731
Group 2 vs. Group 3	-.094	.063	-1.490	.136
T1 vs. Monolinguals	-.013	.008	-1.663	.096
T2 vs. Monolinguals	**-.102**	**.028**	**-3.612**	**.000**
T3 vs. Monolinguals	**-.102**	**.028**	**-.3.615**	**.000**
Group 2 vs. Monolinguals	**-.211**	**.054**	**-3.905**	**<.001**
Group 3 vs. Monolinguals	**-.117**	**.036**	**-3.226**	**.001**

In general, there was no statistical difference found between the group of quasi-monolinguals and the students upon their commencement of phonetic training in English (*p*=.096). There was a difference between the students' performance at T1 and T2 (*p*=.000) as well as T1 and T3 (thus, the start vs. the end of their first year of phonetic instruction; *p*=.000) but not between T2 and T3 (*p*=.982). 1BA students at T1 had also significantly fewer unvoiced realisations than students in 2BA (*p*=.000) and 3BA (*p*=.004). Additionally, focusing on the group of monolinguals, significant difference was found between them and students at T2 (*p*=.000), T3 (*p*=.000), as well as Group 2 (*p*<.001) and Group 3 (*p*=.001). No other contrasts were found.

To sum up, as far as the parameter of VOT in plosives in words in citation forms goes, we observe that Polish learners of English display more L2-induced drift effects in the voiced series relative to the voiceless one. This is visible not only in the average duration of pre-voicing under the influence of phonetic instruction in L2 but also in the absolute numbers of English-like unvoiced realisations in Polish productions of our participants. Let us now turn to other acoustic correlates of the voicing contrast in Polish.

6.2.2.2. Sentence reading task

Starting, once again, with Polish Type 1 tokens (i.e. fully pre-voiced ones) we analysed 1673 items from Group 1, that is the first-year students (577 at T1, 551 at T2, 545 at T3) and compared them with 575 items obtained from Group 4 (that is, the quasi-monolinguals). Fig. 51 illustrates the mean pre-voicing values sorted for Group/Session and prosodic position.

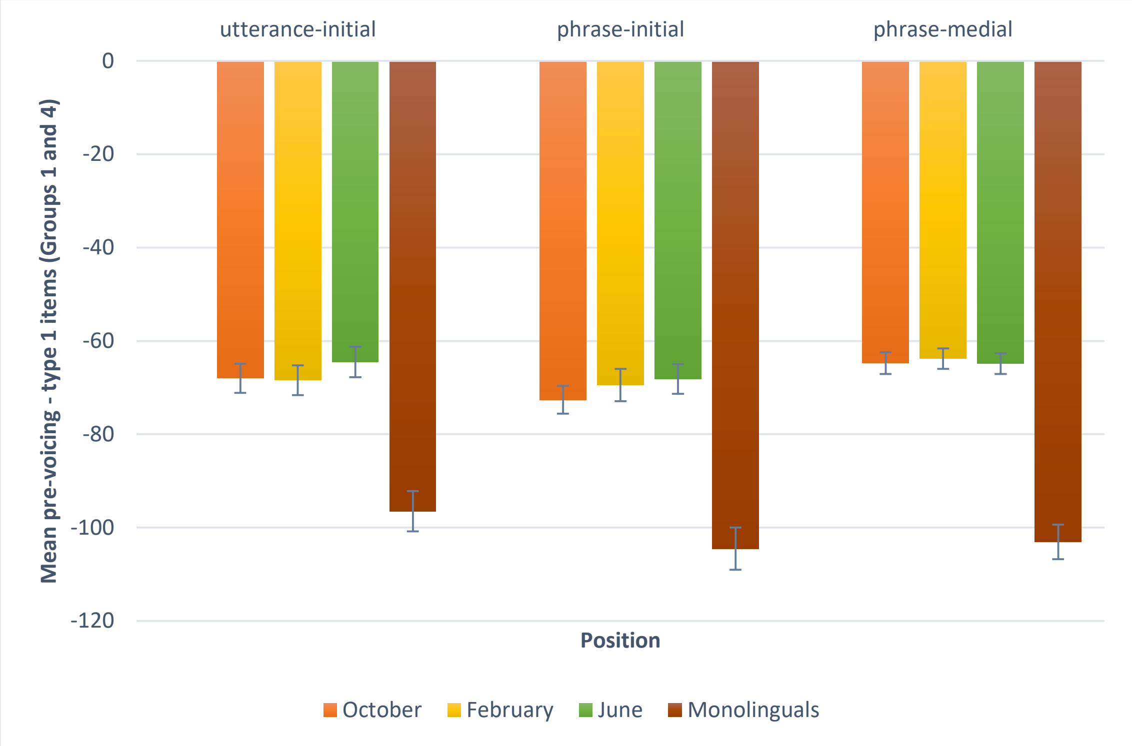


Fig. 51. Mean VOT of /b, d, g/-initial target words in sentence production task, sorted for session and prosodic position (Type 1; Groups 1 and 4 only; sentence-reading task).

Immediately one can notice the striking difference in pre-voicing duration between students' and quasi-monolinguals' productions. Table 33 provides the specific mean values alongside standard deviations.

Table 33. Mean VOT of /b, d, g/-initial target words in sentence production task with Standard Deviations, sorted for session and prosodic position (Groups 1 and 4 only; sentence-reading task).

Prosodic position	Session/Group	N	Mean (ms)	Std. Deviation
utterance-initial	T1	180	-67.99	21.23
	T2	199	-68.41	22.82
	T3	181	-64.53	22.17
	Monolinguals	194	-96.49	30.42
phrase-initial	T1	189	-72.62	20.73
	T2	165	-69.42	22.48
	T3	177	-68.14	21.54
	Monolinguals	186	-104.51	31.43
phrase-medial	T1	208	-64.75	16.92
	T2	187	-63.82	15.16
	T3	187	-64.85	15.43
	Monolinguals	195	-103.09	26.23

It appears that the effects of the phonetic training in L2 on pre-voicing in L1 are rather moderate – the differences oscillate at 2-3 ms. However, the general contrast between pre-voicing in those with higher English proficiency and the quasi-monolinguals are much greater – ca. 30 ms, and this difference is present from the very start of their pronunciation instruction[76].

A GLMM analysis was run with VOT as the dependent variable and the three-way interaction between Session*Prosodic_position*Type as the main predictor of interest. Speech rate was included as a control variable, and Speaker and Item as random factors. The results of pairwise comparisons between Groups 1 and 4 for Type 1 items are presented in Table 34[77]. Pairwise comparisons follow, with the significant differences bolded. A full coefficient table can be found in Appendix 7.

Table 34. The results of the GLMM with VOT of /b, d, g/ as the dependent variable and Session*Prosodic_position*Type as the main predictor (pairwise comparisons: Groups 1 and 4; Type 1 only; sentence-reading task).

Prosodic position	Session/Group	Contrast estimate	Std. Error	t	Significance
utterance-initial	T1 vs. T2	1.125	2.063	.546	.585
	T1 vs. T3	-2.380	2.111	-1.127	.260
	T2 vs. T3	-3.505	2.058	-1.703	.089
	T1 vs. Monolinguals	**29.979**	**3.468**	**8.645**	**.000**
	T2 vs. Monolinguals	**28.853**	**3.433**	**8.405**	**.000**
	T3 vs. Monolinguals	**32.359**	**3.465**	**9.339**	**.000**
phrase-initial	T1 vs. T2	-3.558	2.135	-1.667	.096
	T1 vs. T3	-5.272	2.096	-2.515	.012
	T2 vs. T3	-1.714	2.168	-.790	.429
	T1 vs. Monolinguals	**32.025**	**3.479**	**9.204**	**.000**
	T2 vs. Monolinguals	**35.583**	**3.521**	**10.106**	**.000**
	T3 vs. Monolinguals	**37.297**	**3.498**	**10.663**	**.000**
phrase-medial	T1 vs. T2	-.859	2.019	-.425	.671
	T1 vs. T3	.550	2.023	.272	.786
	T2 vs. T3	1.409	2.075	.679	.497
	T1 vs. Monolinguals	**38.343**	**3.477**	**11.028**	**.000**
	T2 vs. Monolinguals	**39.201**	**3.509**	**11.173**	**.000**
	T3 vs. Monolinguals	**37.792**	**3.533**	**10.697**	**.000**

[76] We will get back to this observation in section 6.4.
[77] As per usual, this is done for presentation purposes only; all Types and Groups were part of one analysis.

Aside from the significant difference between Group 1 and 4, the shortening of pre-voicing across the three testing times in Group 1's productions observed in word reading task is less salient; the only contrast that turned out to be significant was the difference between T1 and T3 in the phrase-initial position (*p*=.012).

Our comparison groups – Groups 2 (second-year students) and 3 (third-year students) – provided 369 and 379 additional items respectively. Fig. 52 presents mean values of Type 1 tokens from all groups and sessions, sorted for the position within the sentence.

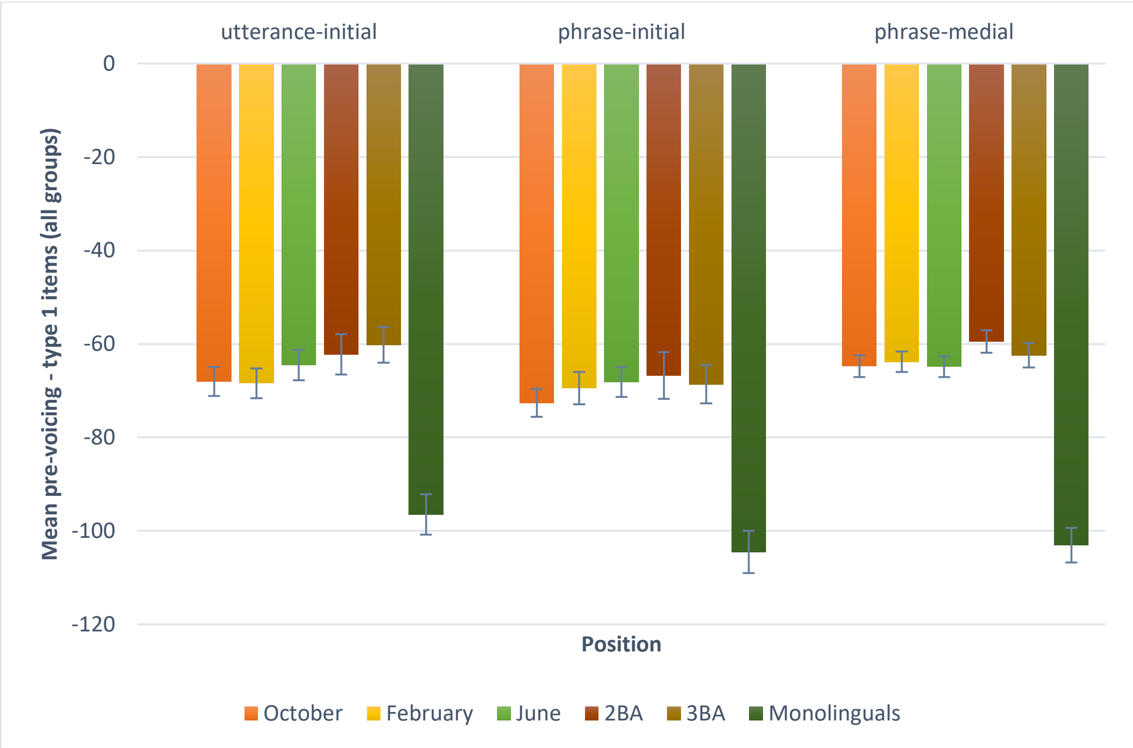


Fig. 52. Mean VOT of /b, d, g/-initial target words in sentence production task, sorted for session and prosodic position (Type 1; all groups; sentence-reading task).

Visual inspection of the bar graphs indicates that once again, the shortest pre-voicing is found in Group 2 (i.e. the second year students' productions). This is corroborated when we look at the averages shown in Table 35, alongside Standard Deviations.

Table 35. Mean VOT of /b, d, g/-initial target words in sentence production task with Standard Deviations, sorted for session and prosodic position (Groups 2 and 3 only; sentence-reading task).

Prosodic position	Session/Group	N	Mean (ms)	Std. Deviation
utterance-initial	Group 2	125	-62.22	24.54
	Group 3	132	-60.21	22.03
phrase-initial	Group 2	109	-66.73	26.35
	Group 3	112	-68.65	21.74
phrase-medial	Group 2	135	-59.47	14.16
	Group 3	132	-62.44	15.16

While the differences are not big, it does appear that pre-voicing got progressively shorter as the training progressed, peaking at 2BA. The results of the pairwise comparisons taken from the GLMM model already described before and focused on Groups 2 and 3 are given in Table 36.

Table 36. The results of the GLMM with VOT of /b, d, g/ as the dependent variable and Session*Prosodic_position as the main predictor (Type 1; pairwise comparisons: Groups 2 and 3 vs. the rest; sentence-reading task).

Prosodic position	Session/Group	Contrast estimate	Std. Error	t	Significance
utterance-initial	Group 2 vs. T1	5.184	3.808	1.361	.173
	Group 2 vs. T2	6.309	3.780	1.669	.095
	Group 2 vs. T3	2.804	3.807	.737	.461
	Group 2 vs. Monolinguals	**35.163**	**3.794**	**9.267**	**.000**
	Group 3 vs. T1	6.480	3.783	1.713	.087
	Group 3 vs. T2	**7.605**	**3.753**	**2.026**	**.043**
	Group 3 vs. T3	4.100	3.781	1.084	.278
	Group 3 vs. Group 2	1.296	4.084	.317	.751
	Group 3 vs. Monolinguals	**36.459**	**3.765**	**9.684**	**.000**
phrase-initial	Group 2 vs. T1	6.874	3.858	1.782	.075
	Group 2 vs. T2	3.315	4.212	.834	.404
	Group 2 vs. T3	1.602	3.880	.413	.680
	Group 2 vs. Monolinguals	**38.898**	**3.892**	**9.994**	**.000**
	Group 3 vs. T1	3.361	3.843	.875	.382
	Group 3 vs. T2	-.198	3.884	-.051	.959
	Group 3 vs. T3	-1.912	3.868	-.495	.621
	Group 3 vs. Group 2	-3.513	4.212	-.834	.404
	Group 3 vs. Monolinguals	**35.385**	**3.868**	**9.148**	**.000**
phrase-medial	Group 2 vs. T1	4.421	3.737	1.183	.237
	Group 2 vs. T2	3.562	3.767	.946	.344
	Group 2 vs. T3	4.971	3.762	1.321	.186
	Group 2 vs. Monolinguals	**42.764**	**3.850**	**11.108**	**.000**
	Group 3 vs. T1	1.247	3.739	.334	.739
	Group 3 vs. T2	.389	3.769	.103	.918
	Group 3 vs. T3	1.798	3.766	.477	.633

Group 3 vs. Group 2	-3.174	4.044	-.785	.433
Group 3 vs. Monolinguals	**39.590**	**3.843**	**10.301**	**.000**

As can be observed, the only significant contrasts concerned the groups of students and the quasi-monolinguals. Nonetheless, the general trend of (admittedly, very modest in this task) pre-voicing shortening over the course of pronunciation instruction which is most visible at 2BA persisted and replicated the trends we have seen in the previous task.

A considerable number of voiced-initial target words were realised as Type 2 tokens (i.e. as partially pre-voiced items with a break in pre-voicing up to 50 ms). This is shown in Fig. 53, where we can see the percentages of Type 1 vs. Type 2 realisations across all groups and sessions[78].

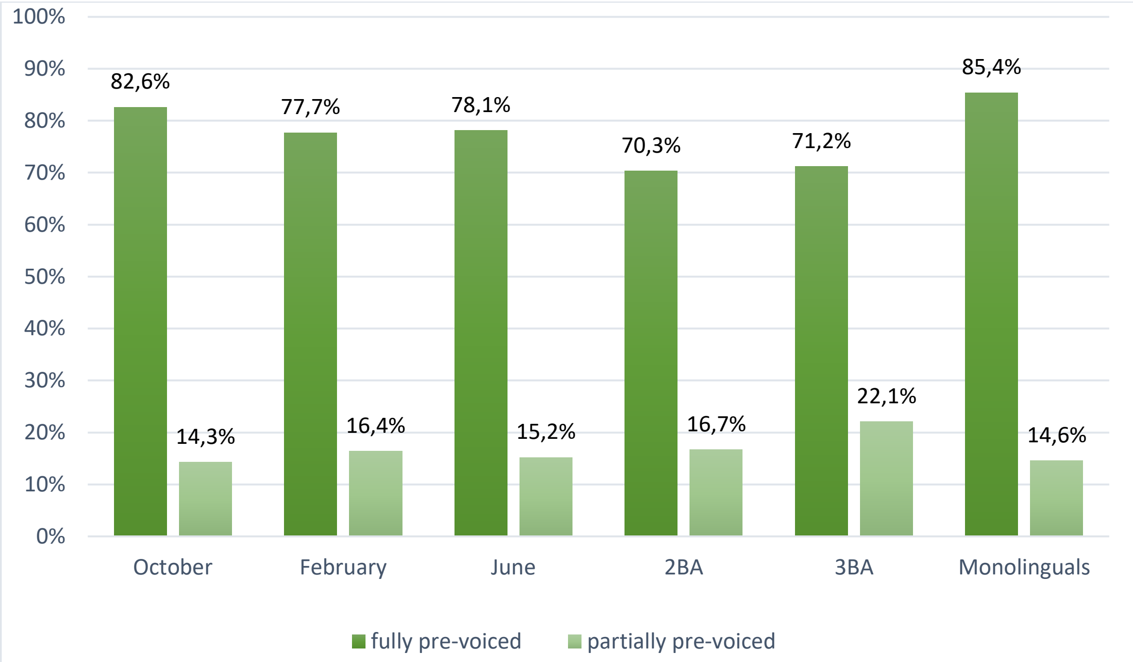


Fig. 53. The percentage of Type 1 vs. Type 2 realisations across groups and sessions in the sentence reading task.

Fig. 53 shows that at T1, the first year students and the group of quasi monolinguals display comparable numbers of Type 2 items (specifically N=100 in the former and N=98 in the latter). At T2 (N=117) and T3 (N=107) there is a very small increase, which appears

[78] The percentages do not add up to 100% as we excluded Type 3 tokens (i.e. unvoiced ones).

to continue through both Group 2 (N=88) and Group 3 (N=117)[79]. In total there were 627 Type 2 items in the sentence reading task.

A binary logistic regression was run for Type 2 tokens with the dependent variable coded as *yes* or *no*, Session as the main predictor, and Speaker and Item as random factors. The results thereof revealed that no differences were found between any of the groups[80]. Therefore, the presence or absence of a break in pre-voicing, though some contrasts between the groups were found in the previous task (i.e. word reading) is unlikely to be the result of phonetic drift effects in L1 Polish associated with phonetic training in L2.

However, interesting data were obtained for the Type 3 (i.e. unvoiced) realisations. These were expected to be less common in this task relative to the word reading task, as the target words were placed in an intervocalic context which is quite conducive to voicing in general. Nonetheless, although perhaps the percentages were less substantial, the "English-like" realisations were very much present, as shown in Fig. 54.

Fig. 54. The percentage of pre-voiced (Type 1 and 2) vs. unvoiced realisations of /b, d, g/ in the sentence reading task.

[79] The raw numbers of Type 2 productions in Group 2 and Group 3 are lower due to fewer participants in those two groups.
[80] The full coefficient table can be found in Appendix 7.

The first thing to be observed is that Group 4 produced exclusively pre-voiced target words; Type 3 realisations were absent. In comparison, already at T1 3.1% (N=22) voiced-initial items were produced without pre-voicing. At T2 it was 5.9% (N=41) while at T3 – 6.7% (N=45). A considerable increase can be noticed in Group 2's productions with 13% of target words produced without pre-voicing (N=68). Group 3, in turn, presented a small decrease and the percentage of English-like realisations was equal to what we have seen in Group 3 at T3, that is 6.7% (N=35). Therefore, even without explicit phonetic instruction Type 3 tokens persisted in their speech. On the whole, out of 211 unvoiced tokens, 98 were produced utterance-initially, 91 phrase-initially, and only 22 phrase-medially. This is quite intuitive as recall that in phrase-medial position there is not prosodic break preceding the target word which means that shorter breaks are made and voicing is more likely to be maintained.

A Binary Logistic Regression was run in order to determine whether those increases were significant. The dependent variable was a binary (*yes/no*) choice between pre-voicing being present or not. Session and Prosodic_position were included as the main predictors of interest, while Speaker and Item were the random factors in the analysis. The results of the model are given in Table 37; the pairwise comparisons are concerned with session only, as position did not turn out to be a significant predictor (*p*=.988). The significant contrasts are bolded and the full coefficient table is provided in Appendix 7.

Table 37. The results of Binary Logistic Regression with Type_3 realisation of voiced plosives in Polish as the dependent variable (pairwise comparisons; sentence-reading task).

Session	Contrast estimate	Std. Error	t	Significance
T1 vs. T2	.015	.008	1.933	.053
T1 vs. T3	**.019**	**.009**	**2.185**	**.029**
T2 vs. T3	.004	.008	.529	.597
T1 vs. Group 2	**.052**	**.025**	**2.046**	**.041**
T2 vs. Group 2	.037	.027	1.377	.168
T3 vs. Group 2	.032	.027	1.199	.230
T1 vs. Group 3	.010	.013	.787	.431
T2 vs. Group 3	-.005	.015	-.339	.735
T3 vs. Group 3	-.009	.016	-.577	.564
Group 2 vs. Group 3	-.042	.027	-1.558	.119
T1 vs. Monolinguals	**-.017**	**.006**	**-2.646**	**.008**
T2 vs. Monolinguals	**-.032**	**.011**	**-2.915**	**.004**
T3 vs. Monolinguals	**-.037**	**.012**	**-2.961**	**.003**

Group 2 vs. Monolinguals	**-.069**	**.025**	**-2.766**	**.006**
Group 3 vs. Monolignuals	**-.027**	**.011**	**-2.382**	**.017**

It appears that the number of unvoiced realisations significantly increases between October and June (that is, T1 and T3; *p*=.029) and the contrast is also observed between students in October and after their second year of phonetic instruction (Group 1 at T1 and Group 2; *p*=.041). Moreover, students across all recording sessions differ significantly from the group of monolinguals.

To recap, when it comes to pre-voicing in the sentence reading task we observed mostly the salient difference in duration of vocal fold vibrations between the students and the monolinguals, less so within the three students' groups. The sheer numbers of breaks in pre-voicing were unlikely to be evidence of phonetic drift. However, as far as Type 3 is concerned, once again, there was a robust difference between the students' and the monolinguals' performance in the absolute numbers of unvoiced realisations. The contrast was also visible between the students' at the onset and towards the end of the pronunciation training, indicating that the formal instruction could exert influence on their performance in L1 sentence reading task.

6.2.3. Comparison of voiceless vs. voiced plosives

The two production studies described in the previous sections have described the differences in the duration of Polish VOT (both positive and negative) across four different groups in order to assess the behaviour of the two series of stops in the productions of students undergoing intensive phonetic training in L2 and compare them with monolingual norms.

Let us compare the results with what has been reported for Polish (cf. Table 5). It appears that as far as the voiceless plosives go, in the word reading task[81] the averages are slightly higher than those of Keating (1981) and Malisz and Żygis (2015) and are somewhat comparable to Kopczyński's (1977). With respect to the voiced series, the pre-voicing values in the speech of our speakers (in particular the monolingual group) are longer than what was attested previously. Table 38 provides a comparison of the previous

[81] We focus on the word reading task as POA was not distinguished in the sentence reading task.

studies on Polish with the inclusion of the averages obtained from the groups evaluated in the present thesis.

Table 38. Polish VOT measures comparing previous findings with the data presented in the thesis.

p / t / k	b / d / g	Source
37.5 / 33 / 49	-78 / -72 / -61	Kopczyński (1977)
21 / 28 / 53	-88 / -90 / -66	Keating (1981)
18 / 24 / 35	-57 / -46 / -49	Malisz and Żygis (2015)
	THE PRESENT THESIS	
34/34/57	-86/-81/-72	Group 1 at T1 (words)
34/37/59	-90/-81/-77	Group 1 at T2 (words)
33/35/57	-84/-79/-72	Group 1 at T3 (words)
37/35/59	-81/-79/-66	Group 2 (words)
38/37/60	-103/-87/-83	Group 3 (words)
34/34/55	-95/-93/-86	Group 4 (words)

From the perspective of phonetic drift, we observed asymmetrical behaviour of the two series. It appeared that the voiceless category was somewhat more stable and less susceptible to the influence of L2 phonetic training relative to the voiced series, where we observed progressive shortening of pre-voicing. Fig. 55 juxtaposes the averages to illustrate this.

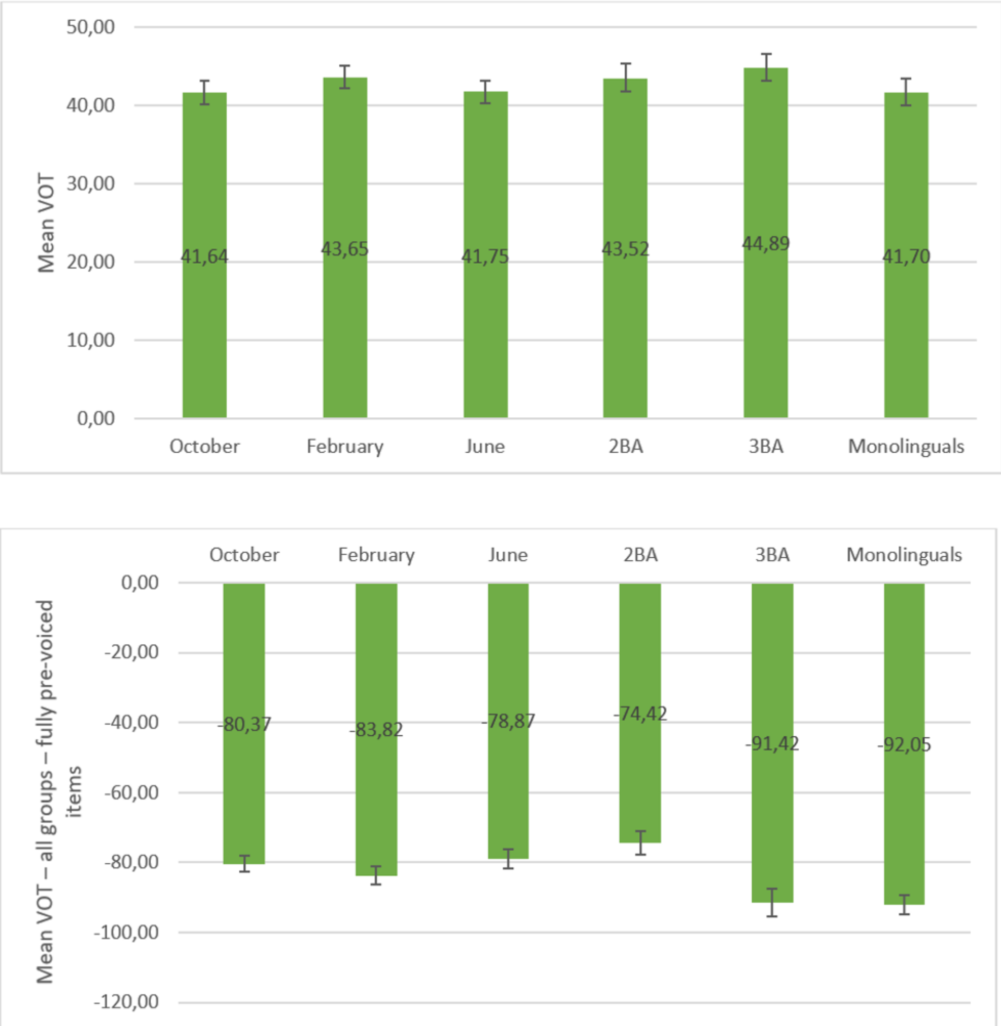


Fig. 55. A comparison of the contrast magnitude between the groups in the voiceless (top) and voiced (bottom) category in the word reading task in L1 Polish.

As can be observed, while the difference between the students at the onset of phonetic training (T1, that is October of the first year) and the offset thereof (2BA, that is the end of the second year) in the voiceless category is ca. 2 ms, the contrast between those two in the voiced category is ca. 6 ms. While this may not seem like too much of a difference, more striking is the comparison between the students at the end of the phonetic training and the monolinguals. In the voiceless category the difference is ca. 2 ms while in the voiced one – ca. 18 ms. Similarly, even before the training begins (i.e. T1), there is a more significant difference in pre-voicing between the students and monolinguals in the voiced series (ca. 12 ms) than in the voiceless one (less than 1 ms).

Recall also that not only does the sheer pre-voicing duration differ, the type of pre-voicing realisation changes significantly as well (cf. Fig. 50). As the training progressed, we noted more and more unvoiced realisations of Polish voiced plosives (peaking at nearly 25% in the second-year students' productions), which is not typical for Polish.

Similar results were observed in the sentence reading task. Fig. 56, Fig. 57, and Fig. 58 compare the changes between voiceless and voiced categories across the groups sorted for prosodic position: utterance-initial (Fig. 56), phrase-initial (i.e. utterance-medial; Fig. 57), and phrase-medial (Fig. 58), respectively.

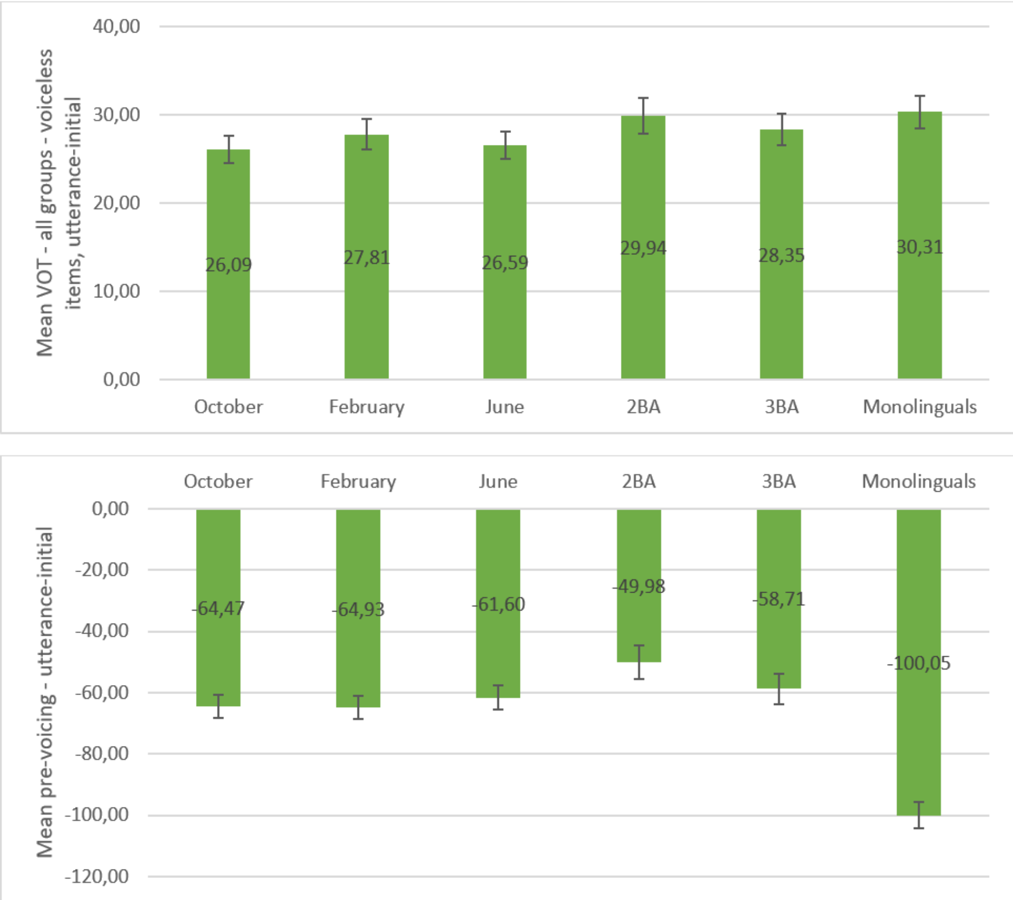


Fig. 56. A comparison of the contrast magnitude between the groups in the voiceless (top) and voiced (bottom) category in the sentence reading task in L1 Polish (utterance-initial position).

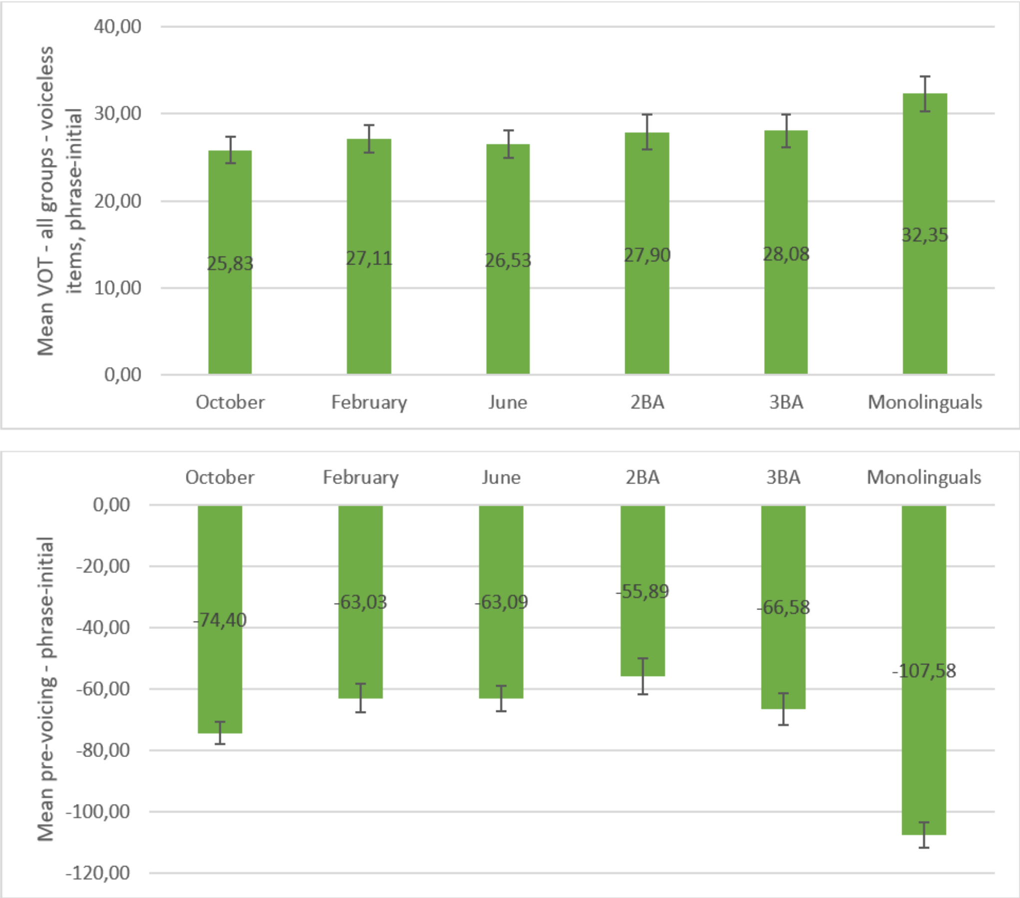


Fig. 57. A comparison of the contrast magnitude between the groups in the voiceless (top) and voiced (bottom) category in the sentence reading task in L1 Polish (phrase-initial position).

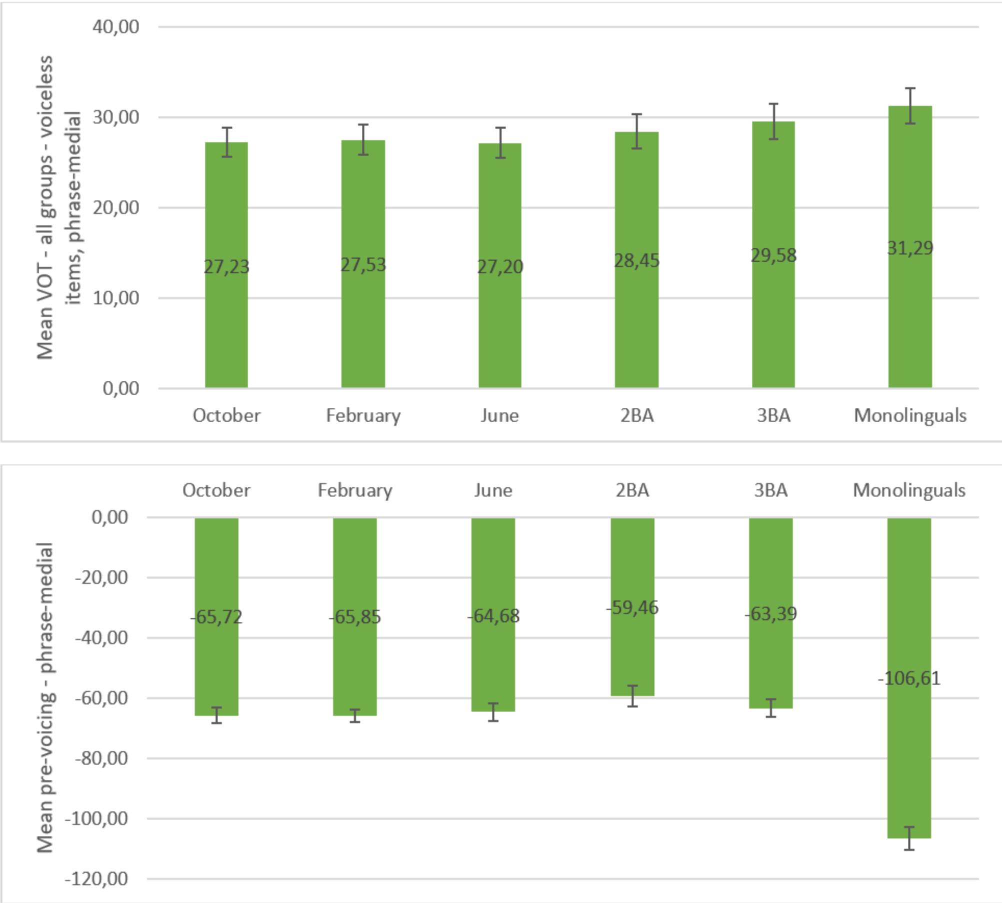


Fig. 58. A comparison of the contrast magnitude between the groups in the voiceless (top) and voiced (bottom) category in the sentence reading task in L1 Polish (phrase-medial position).

Notice that the trend in pre-voicing shortening across all three prosodic positions mirrors the effects observed in the word reading task in that the shortest pre-voicing is found in the productions of second-year students. The differences in pre-voicing durations between students and monolinguals are also rather striking. Less salient changes are observed in the voiceless series, with this category, again, displaying more stability.

With respect to type of voicing realisation, despite the conducive intervocalic position, a number of tokens was realised without pre-voicing, in an English-like fashion (cf. Fig. 54), once again mirroring the patterns observed in the word reading task.

From the perspective of the SLM and the principle of equivalence classification, it may be conjectured that the voiced category, since it displays more variability across the groups and sessions and diverges from the monolingual norms to a greater extent, is deemed as being equivalent in L1 and L2. On the other hand, it appears that, since the voiceless category is less affected by the L2 phonetic training, it is not classified as equivalent; instead, a separate category for the voiceless category must have been formed. We will explore this to a greater detail in section 7.2.

6.2.4. Other laryngeal cues

The parameter of VOT is by no means the only correlate to voicing. It has been shown that voiceless onsets exert influence on the f_0 (Ohde 1984) as well as F1 (Stevens and Klatt 1974) at the onset of the following vowel.

Before we describe the results of the phonetic study thereof, let the rationale behind including this section in the thesis be stated. If you recall the phonological theories described in Chapter 3, neither Feature Theory nor Laryngeal Realism go beyond the parameter of VOT (with feature theory not being at all concerned with the actual phonetic implementation of the phonological representations). In the way in which laryngeal contrast in Polish is represented in Laryngeal Realism, pre-voicing is in fact of utmost importance; the feature [voice] is virtually equated with the presence thereof in the phonetics. How is the contrast maintained if pre-voicing is absent, as we have seen in the data obtained from our students? Therefore, while we do not really ask any phonetic drift-related questions in the present section, as no theories make any predictions with respect to cross-linguistic interaction as far as f_0 and F1 at vowel onset are concerned, the main issue to be explored is whether or not the drift effects in the voiced series have any effects on contrast maintenance in Polish. To the best of my knowledge, this question has not been explored aside from Schwartz et al. (2019). Therefore the data to be presented will be a contribution to the phonetic data we possess on Polish laryngeal contrasts.

Furthermore, if we do assume that phonological features exert influence on f0 and pitch, the three theories described in Chapter 3 make different predictions as far as this influence presents itself. In feature theory, polarisation effects can be expected (Keating 1984), with [+voice] lowering pitch and F1 and [-voice] raising them. In Laryngeal Realism we would expect [voice] to lower pitch and F1, and the unspecified voiceless category to be unaffected and comparable with baseline segments. In turn, in Onset Prominence (cf. Fig. 14), the presence of the feature [fortis] at the VO level in the voiceless category suggests that pitch and F1 would be raised after a voiceless onset; since the voiced category is unspecified for phonological features, pitch and F1 should remain unaffected. Therefore, it is clear that the theories make contradictory predictions and assessing those predictions empirically presents a viable way of evaluating those laryngeal theories.

6.2.4.1. L1 Polish: f_0 (pitch)

As mentioned in the introduction to this section, we were interested to see whether the effects would be there, even without pre-voicing present. In other words, looking at pitch after different types of onsets allowed us to check if our students maintained the laryngeal contrast between the voiced and voiceless categories in the event of there being no vocal fold activity in the closure phase.

In order to investigate this, we pooled together the productions of Groups 1 (at all testing times), 2, and 3 (Group 4 was excluded as there were only two cases of unvoiced realisations) and sorted them for onset type (pre-voiced, unvoiced, and voiceless) and compared them with nasal-initial items, which served as a baseline (as they are claimed not to affect pitch; Kirby and Ladd 2018). The nasal items initially served as a subset of fillers included in the word list task. They were mono- and disyllabic real Polish words, with bilabial and alveolar place of articulation, followed by the vowel /a/.

In total, 4783 tokens were included in the analysis; exclusions involved cases where the scripts were unable to calculate pitch or when the measurements were unreliable. The latter case was most prominent for the vowel /ɔ/ and a decision was made to exclude this vowel from the analysis entirely. Out of the 4783 items, 3822 measurements were taken from the onset of /a/ and 961 from the onset of /ɛ/. We looked at the first 20% of the vowel (starting at the onset of voicing following the release noise associated with the burst portion of the preceding consonant) and the averages were given in Bark, in order to normalise the data.

Visual inspection of the line graph given in Fig. 59 points to a conclusion that pitch was raised following a voiceless onset, but pre-voiced and unvoiced onsets appeared not to be affected, relative to the nasal baseline. Similar effects have been observed in French and Italian (cf. Kirby and Ladd 2018).

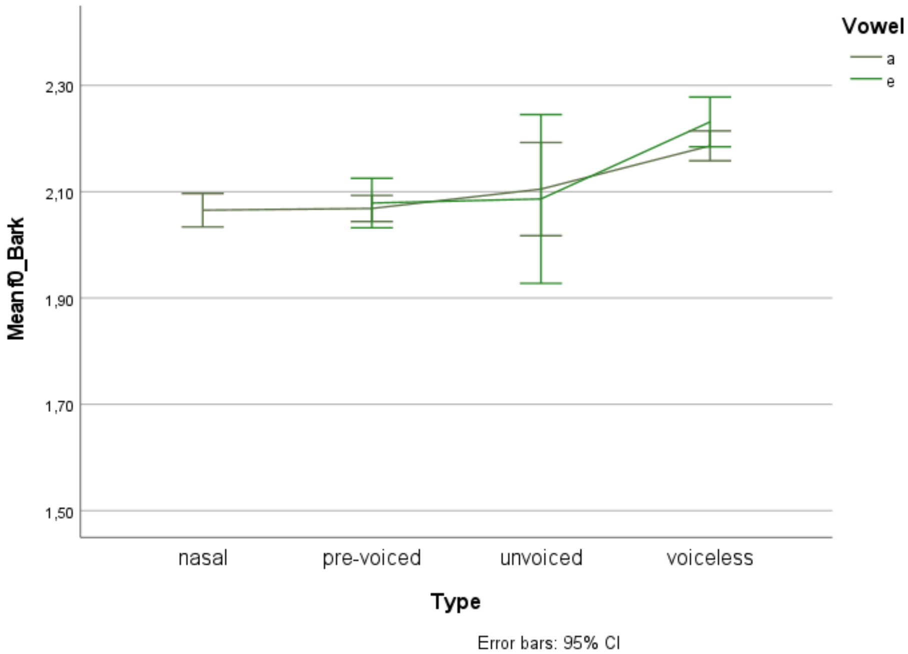


Fig. 59. Pitch (in Bark; an average taken from the first 20% of the vowel) following nasal (baseline), pre-voiced, unvoiced, and voiceless onsets, sorted for vowel quality (word reading task).

Table 39 provides the means, sorted for onset type and vowel quality.

Table 39. Mean pitch (in Bark) according to onset type and vowel quality (word reading task).

Vowel	Onset type	N	Mean (in Bark)	Std. Deviation
/a/	nasal	867	2.06	.47
	pre-voiced	1376	2.06	.46
	unvoiced	215	2.10	.65
	voiceless	1364	2.18	.52
/ɛ/	nasal	--	--	--
	pre-voiced	384	2.07	.46
	unvoiced	64	2.08	.63
	voiceless	513	2.23	.53

The results of a GLMM analysis yield support to the initial observations and are shown in Table 40. Pitch was the dependent variable, while the interaction of Type*Vowel was the main predictor of interest, Speaker and Item were included as random factors. The full coefficient table is given in Appendix 7; here we present pairwise comparisons, with the significant contrasts bolded.

Table 40. The results of a GLMM analysis with f0 at the onset of the vowel in Polish, sorted for Type of consonantal onset and vowel quality (pairwise comparisons; word reading task).

Vowel	Type	Contrast estimate	Std. Error	t	Significance
/a/	Type 1 vs. Type 3	-.044	.024	-1.818	.069
	Type 1 vs. Type 4	**-.120**	**.012**	**-9.840**	**.000**
	Type 3 vs. Type 4	**-.077**	**.024**	**-3.205**	**.001**
	Nasal vs. Type 1	.000	.014	.030	.976
	Nasal vs. Type 3	-.043	.025	-1.746	.081
	Nasal vs. Type 4	**-.120**	**.014**	**-8.620**	**.000**
/ɛ/	Type 1 vs. Type 3	-.038	.044	-.873	.383
	Type 1 vs. Type 4	**-.150**	**.022**	**-6.980**	**<.001**
	Type 3 vs. Type 4	**-.112**	**.043**	**-2.622**	**.009**

As can be seen, contrast between the voiceless and voiced categories was maintained regardless of whether or not pre-voicing was present.

Turning to the sentence reading task, the keywords couched in the longer sentences have been subject to the same analysis. Once again we pooled the productions gathered from all three groups (aside from Group 4, that is monolinguals, as there have been no Type 3 realisations of the voiced category present in the corpus). Unfortunately, our dataset lacked the nasal baseline here, for which reason we limited ourselves to the comparison of pre-voiced[82], unvoiced, and voiceless onsets only. A previous study on Polish has shown minimal effects of prosodic position on vowel quality (Wojtkowiak 2020) and as such, position was not included as a predictor in the following analysis.

Now let us turn to the sentence reading task. In general it yielded 4830 vowel tokens in the #CV position. However, on top of erroneous productions, we also excluded the vowel /ɔ/, once again due to problems with formant extraction. As opposed to the word reading task, the dataset in this task provided a number of words containing the vowel /ɨ/ (N=1311), in addition to /a/ (N=1536) and /ɛ/ (N=1571). All in all, 4418 items were included in the analysis, out of which 2183 were pre-voiced, 160 were unvoiced, and 2075 were voiceless. The first 20% of the vowel (starting at the onset of voicing following the release noise associated with the burst portion of the preceding consonant) was investigated and the averages were given in Bark, in order to normalise the data. Fig. 60 provides a line graph with the mean values of f_0, sorted for the vowel quality and onset type.

[82] Recall that here we do not distinguish between Type 1 and Type 2 tokens and we treat them as belonging to a “pre-voiced” category.

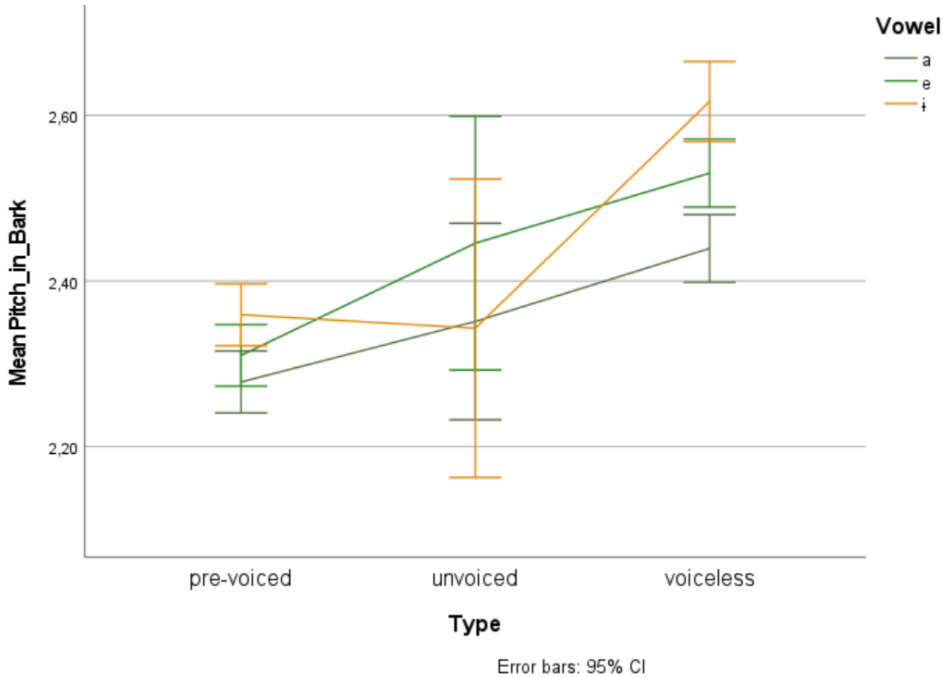


Fig. 60. Pitch (in Bark; an average taken from the first 20% of the vowel) following, pre-voiced, unvoiced, and voiceless onsets, sorted for vowel quality (sentence reading task).

To supplement the above figure, mean values of pitch sorted for onset type and vowel quality are given in Table 41.

Table 41. Mean pitch (in Bark) according to onset type and vowel quality (sentence reading task).

Vowel	Onset type	N	Mean (in Bark)	Std. Deviation
/a/	pre-voiced	709	2.27	.50
	unvoiced	66	2.35	.48
	voiceless	761	2.43	.57
/ɛ/	pre-voiced	736	2.31	.51
	unvoiced	46	2.44	.51
	voiceless	789	2.53	.58
/ɨ/	pre-voiced	738	2.35	.51
	unvoiced	48	2.34	.62
	voiceless	525	2.61	.56

Overall, it appears that, indeed, pitch was generally raised after voiceless onsets, although in the case of the vowel /ɛ/, there is a hierarchy suggesting that it was also slightly higher after unvoiced onsets relative to pre-voiced ones. A GLMM model was run with Pitch as the dependent variable and the interaction between Type*Vowel as the main predictor, and Speaker and Item as random factors. The details of a pairwise comparison are shown in Table 42, with significant contrasts in bold. A full coefficient table is provided in Appendix 7.

Table 42. The results of the GLMM model with pitch as the dependent variable and Type*Vowel as the main interaction (pairwise comparisons; sentence reading task).

Vowel	Type	Contrast estimate	Std. Error	t	Significance
/a/	Type 1 vs. Type 3	.058	.032	1.820	.069
	Type 1 vs. Type 4	**-.150**	**.013**	**-11.896**	**.000**
	Type 3 vs. Type 4	**-.208**	**.032**	**-6.573**	**<.001**
/ɛ/	Type 1 vs. Type 3	.058	.037	1.540	.124
	Type 1 vs. Type 4	**-.209**	**.012**	**-16.876**	**.000**
	Type 3 vs. Type 4	**-.267**	**.037**	**-7.155**	**<.001**
/ɨ/	Type 1 vs. Type 3	-.011	.037	-.288	.774
	Type 1 vs. Type 4	**-.246**	**.014**	**-17.801**	**.000**
	Type 3 vs. Type 4	**-.235**	**.037**	**-6.332**	**<.001**

The results of the statistical analysis showed that pitch is, once again, a reliable acoustic cue for voicing contrasts; pitch is significantly raised after voiceless contrasts for all vowel qualities, while the difference between pre-voiced and unvoiced is insignificant (even in /ɛ/, where the general means could suggest otherwise).

6.2.4.2. L1 Polish: F1 (at vowel onset)

The second vowel-based laryngeal cue of interest was F1. The first formant has been shown to be affected by the laryngeal category of the preceding category, with voiceless onsets shortening the F1 transition, resulting in F1 of non-high vowels being raised (Stevens and Klatt 1974).

Once again the students' productions were pooled together and subdivided into three groups according to pre-voicing type (Type 1, Type 3, and Type 4)[83]. Out of the 3916 items, 2955 measurements were taken from the onset of /a/ and 961 from the onset of /ɛ/. We looked at the first 20% of the vowel (starting at the onset of voicing following the release noise associated with the burst portion of the preceding consonant) and the F1

[83] Here we excluded nasal onsets as they did not serve as a control category.

averages were given in Bark (F1-f_0), in order to normalise the data. A visual representation of the data is provided in Fig. 61.

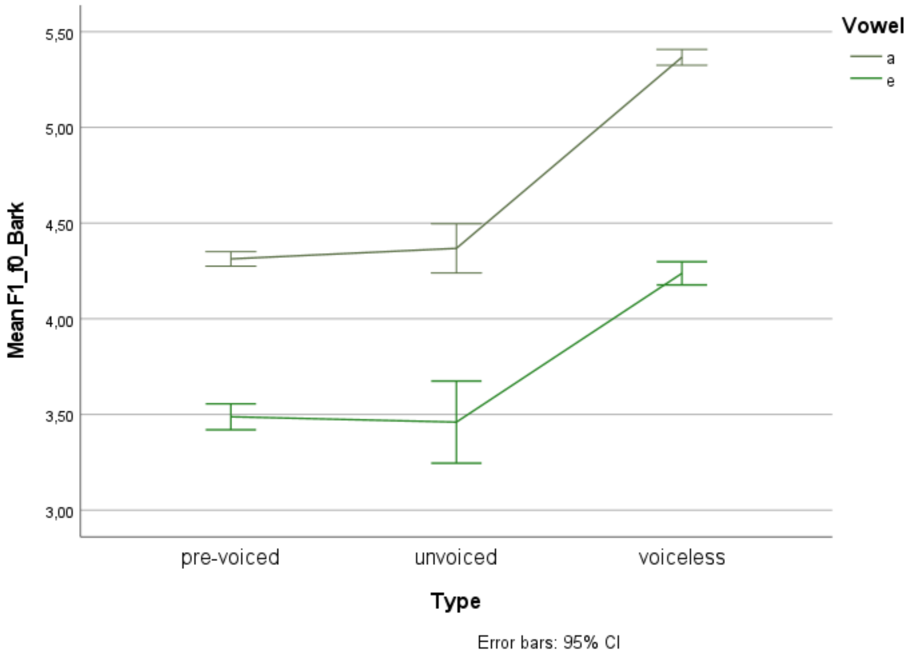


Fig. 61. Mean F1 (F1-f_0; in Bark; an average taken from the first 20% of the vowel) following pre-voiced, unvoiced, and voiceless onsets, sorted for vowel quality (word reading task).

As can be observed, while the averages after pre-voiced and unvoiced onsets are quite comparable, for both vowels F1 is raised after a voiceless stop[84]. The specific mean values are given in Table 43.

Table 43. Mean F1 (F1-f_0; Bark normalised) sorted for onset type and vowel quality (word reading task).

Vowel	Onset type	N	Mean (in Bark)	Std. Deviation
/a/	pre-voiced	1376	4.31	.71
	unvoiced	215	4.36	.95
	voiceless	1364	5.36	.78
/ɛ/	pre-voiced	384	3.48	.67
	unvoiced	64	3.46	.85
	voiceless	513	4.23	.69

[84] Recall that the mean values are F1-f0; the higher the difference between the first formant and the fundamental frequency, the lower the vowel.

The results of a GLMM analysis appear to validate the visual observations and are given in Table 44. In this model, F1 was the dependent variable, while the interaction of Type*Vowel was the main predictor of interest, and Speaker and Item were random factors. The full coefficient table is provided in Appendix 7; here we present pairwise comparisons, with the significant contrasts in bold.

Table 44. The results of a GLMM analysis with F1 at the onset of the vowel in Polish, sorted for Type of consonantal onset and vowel quality (pairwise comparisons; word reading task).

Vowel	Type	Contrast estimate	Std. Error	t	Significance
/a/	Type 1 vs. Type 3	.017	.050	.339	.734
	Type 1 vs. Type 4	**-1.053**	**.025**	**-41.337**	**.000**
	Type 3 vs. Type 4	**-1.070**	**.050**	**-21.405**	**.000**
/ɛ/	Type 1 vs. Type 3	.049	.091	.534	.594
	Type 1 vs. Type 4	**-.755**	**.045**	**-16.780**	**.000**
	Type 3 vs. Type 4	**-.803**	**.090**	**-8.975**	**.000**

The contrast with respect to F1 (F1-f_0) was maintained between voiced and voiceless onsets irrespective of whether or not pre-voicing was present (p=.000 for all cases). No difference was found between Types 1 and 3, that is between pre-voiced and unvoiced onsets. Therefore, the laryngeal contrast between voiceless and voiced categories was robustly maintained even with pre-voicing being absent.

Moving on to the sentence reading task, F1 (F1-f_0; Bark normalised) mean values from the target words were obtained in the same way as what was done for the word reading task; we looked at the first 20% of the vowel. Here, due to the specificity of the dataset (already mentioned in section 6.2.4.1), three vowels were included in the analysis: /a/ (N=1536), /ɛ/ (N=1571), and /ɨ/ (N=1311). Out of the 4418 tokens in total, 2183 were following a pre-voiced onset, 160 an unvoiced onset, and 2075 a voiceless one. Fig. 62 shows the average values, sorted for onset type and vowel quality.

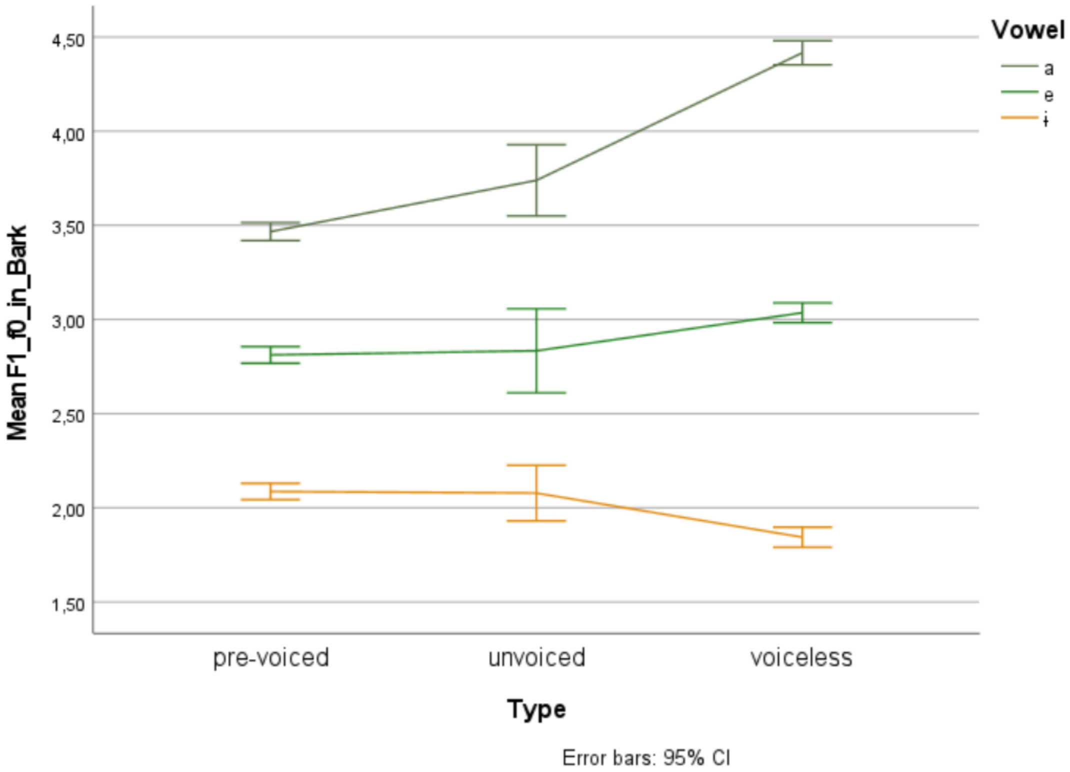


Fig. 62. Mean F1 (F1-f_0; in Bark; an average taken from the first 20% of the vowel) following pre-voiced, unvoiced, and voiceless onsets, sorted for vowel quality (sentence reading task).

We can immediately see that the line graph differs from what we have seen in the previous task; while the general tendency of the F1 being raised after voiceless onsets is followed to a certain extent by both /a/ and /ɛ/, for /ɨ/ it appears to be lowered. While this lowering is quite surprising, raising effect should not be expected as it is not a non-high vowel. Table 45 provides the detailed mean values for each vowel quality and onset type.

Table 45. Mean F1 (F1-f_0; in Bark; an average taken from the first 20% of the vowel) following pre-voiced, unvoiced, and voiceless onsets, sorted for vowel quality (sentence reading task).

Vowel	Onset type	N	Mean (in Bark)	Std. Deviation
/a/	pre-voiced	709	3.46	.64
	unvoiced	66	3.73	.76
	voiceless	761	4.41	.89
/ɛ/	pre-voiced	736	2.81	.61
	unvoiced	46	2.83	.75
	voiceless	789	3.03	.75
/ɨ/	pre-voiced	738	2.08	.59
	unvoiced	48	2.07	.50
	voiceless	525	1.84	.62

Indeed, both /a/ and /ɛ/ show differences in F1 means in the expected direction, but in the case of former the differences are quite robust; F1 appears to be raised already after the unvoiced onset. The difference between pre-voiced and unvoiced onsets in the case of /ɛ/ is not quite salient, but neither is the difference between unvoiced and voiceless ones in this vowel. In the case of /ɨ/, F1 after pre-voiced and unvoiced onsets is comparable, but after the voiceless one it is much lower (that is, the vowel is raised).

A GLMM model was run with F1 (F1-f_0) as the dependent variable, Type*Vowel interaction as the fixed factor, and Speaker and Item as random factors. The results of pairwise comparisons are given in Table 46. Significant contrasts are shown in bold while the full coefficient table is provided in Appendix 7.

Table 46. The results of a GLMM analysis with F1 at the onset of the vowel in Polish, sorted for the type of consonantal onset and vowel quality (pairwise comparisons; sentence reading task).

Vowel	Type	Contrast estimate	Std. Error	t	Significance
/a/	**Type 1 vs. Type 3**	**-.254**	**.079**	**-3.199**	**.001**
	Type 1 vs. Type 4	**-.951**	**.031**	**-30.215**	**.000**
	Type 3 vs. Type 4	**-.697**	**.079**	**-8.837**	**.000**
/ɛ/	Type 1 vs. Type 3	-.062	.093	-.662	.508
	Type 1 vs. Type 4	**-.227**	**.031**	**-7.363**	**<.001**
	Type 3 vs. Type 4	-.166	.093	-1.784	.074
/ɨ/	Type 1 vs. Type 3	.051	.092	.552	.581
	Type 1 vs. Type 4	**.243**	**.034**	**7.055**	**<.001**
	Type 3 vs. Type 4	**.192**	**.092**	**2.078**	**.011**

The results are different from the previous task. First of all, a significant contrast is maintained between all onset types for /a/; interestingly, pre-voiced and unvoiced onsets yield different F1 values (p=.001). In the case of /ɛ/, in turn, there is no contrast between pre-voiced and unvoiced items (p=.508), but the difference is also absent from the unvoiced vs. voiceless pair (p=.074). Only the vowel /ɨ/ follows the pattern we have seen for both fundamental frequency in both tasks (cf. section 6.2.4.1) as well as in the results from the word reading task investigating the F1 cue, presented in this subsection, wherein while there is no contrast between pre-voiced and unvoiced onsets (p=.581), the contrast is maintained between either of the two and the voiceless ones (p<.001 and p=.011 respectively). It might be then the case that it is pitch that is more reliable a cue as far as voicing distinction in Polish is concerned.

We will return to these results in the discussion in section 6.4.2.

6.3. L1: Vowels

The analysis of L1 words focused on two vowels /a/ and /ɛ/, whose formant values were extracted from the middle portion of each production thereof. The measurements are given in Bark differences. We analysed the tokens from all groups taking part in the study; possible exclusions were made in two cases: either when the script was unable to provide us with formant measures (as it would be impossible to obtain the average F1 or F2 values from the specific point of the vowel and then transform those measures into Bark) or in the case of /aj/ and /ej/ sequences (due to problems with segmentation in the case of a sequence of a vowel followed by a glide).

6.3.1. Word reading task

Before we discuss both height and advancement separately, let us visually inspect whether or not, in general, there are any changes that occur over the span of the three years that our students spend enrolled in the English programme. In order to do that, both /a/ and /ɛ/ were plotted on F1/F2 planes in R (R Core Team 2018), using the `ggplot2()` package (Wickham 2016). We included all data points, with the ellipses corresponding to the distance of one standard deviation from the mean F1 (F1-f_0) and F2 (F3-F2) for each of the groups included in the analysis.

First, let us look at /ɛ/, whose scatter plot is shown in Fig. 63.

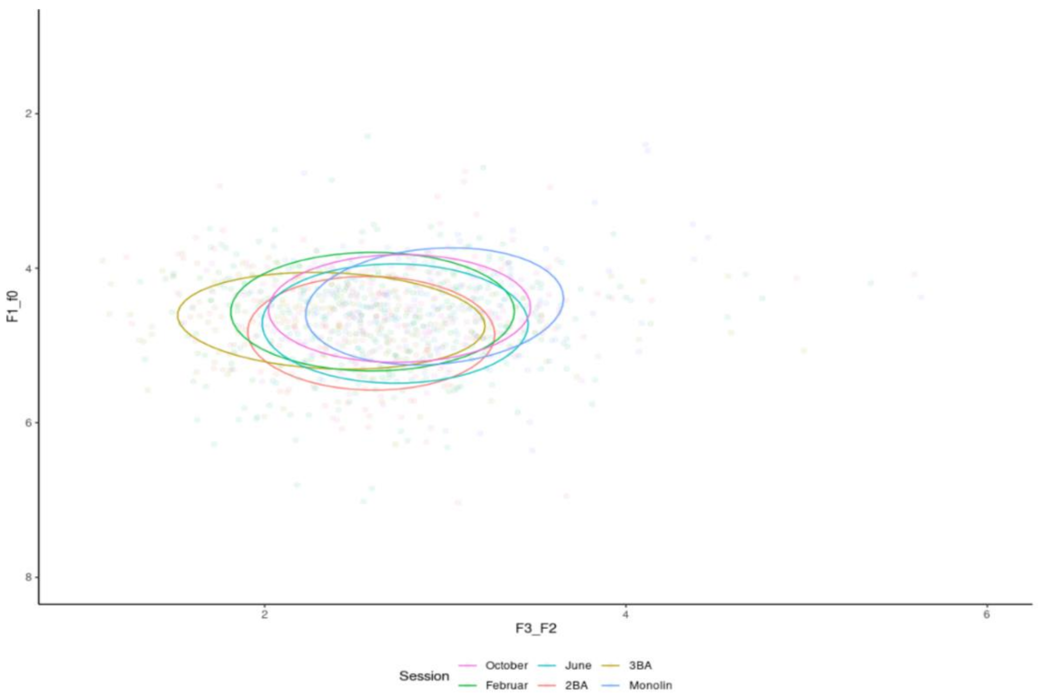


Fig. 63. Polish /ɛ/ in the word reading task plotted on the F1/F2 plane, with the ellipses corresponding to the distance of one standard deviation from the mean, sorted for group (and session); word reading task.

On the whole, we can see that the quality of /ɛ/ appears to undergo some changes as the phonetic training progresses. For monolinguals (in blue) /ɛ/ is slightly more central; less so for students at T1 (in pink), at T2 (in green), and in June (in teal). In the productions of 2BA speakers (i.e. Group 2, in red) it is both fronter and lower, with most front realisations observed in the speech of 3BA (i.e. Group 3, in yellow).

Changes are also present in the realisation of /a/. Let us look at its scatter plot in Fig. 64.

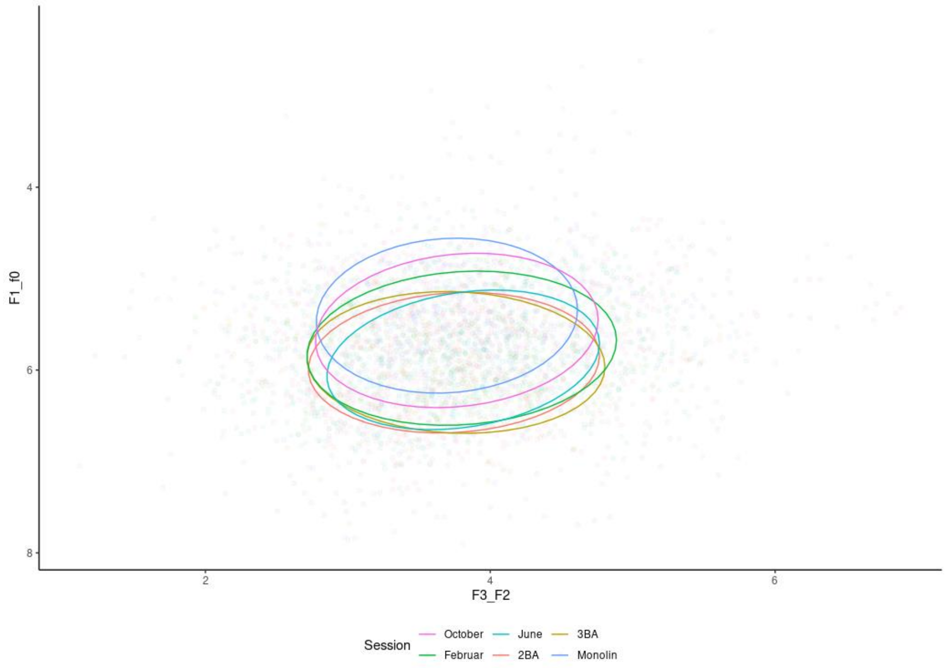


Fig. 64. Polish /a/ in the word reading task plotted on the F1/F2 plane, with the ellipses corresponding to the distance of one standard deviation from the mean, sorted for group (and session); word reading task.

At first glance we may notice that while there is little difference across the groups with respect to vowel advancement and the vowel is relatively central for all speakers, the height of /a/ appears to undergo more changes. The most effects of drift can be seen in the first year of phonetic training: we can see that the vowel moves away from the Monolingual norms (in blue) already at T1 (in pink), and then gets progressively lower at T2 (in green) and T3 (in teal); then the effects seem to slow down. Its quality for T3, 2BA (i.e. Group 2, in red) and 3BA (i.e. Group 3, in yellow) is by and large quite comparable.

Before we look at each formant in detail, we may tentatively state that phonetic drift triggered by intense phonetic training appears to target the vowels of Polish learners of English; the vowel space shows signs of expansion with some lowering of the vowel /a/ and fronting of /ɛ/. As is the case with drift, however (mentioned by Chang 2012), the changes are relatively small and seem to remain inaudible to listeners (at least to one native speaker of Polish, the author of the thesis)[85].

[85] As opposed to the changes associated with phonetic/phonological attrition; cf. Section 2.2.

6.3.1.1. Vowel height

This subsection describes the results of the analyses with F1 as the dependent variable. Mean F1 values (Bark normalised, F1-f_0 from the middle 20% of the vowel) to measure vowel height were obtained for each of the two vowels. In total, 4179 tokens were analysed, out of which /ɛ/ comprised 990 items and /a/ – 3189 items. In what follows, we will discuss each vowel individually.

Let us begin with the higher of the two vowels, that is /ɛ/. Out of the 990 obtained productions, Group 1 provided 584 items (194 at T1, 196 at T2, and 194 at T3), Group 2 – 147 items, Group 3 – 144, while Group 4 – 115. The mean F1 values for each group are given in Table 47.

Table 47. Mean values of F1 (F1-f_0; in Bark) for /ɛ/ in the word reading task, sorted for Group/Session; word reading task.

Vowel	Session/Group	N	Mean (Bark)	Std. Deviation
/ɛ/	Group 1: T1	194	4.50	.55
	Group 1: T2	196	4.57	.62
	Group 1: T3	194	4.74	.62
	Group 2: 2BA	147	4.85	.59
	Group 3: 3BA	144	4.70	.48
	Group 4: Monolinguals	115	4.47	.64

The changes in the height of the vowel /ɛ/ are subsequently visualised in the line graph in Fig. 65. The *y* axis has been reversed for the reader to notice more clearly the differences with respect to height.

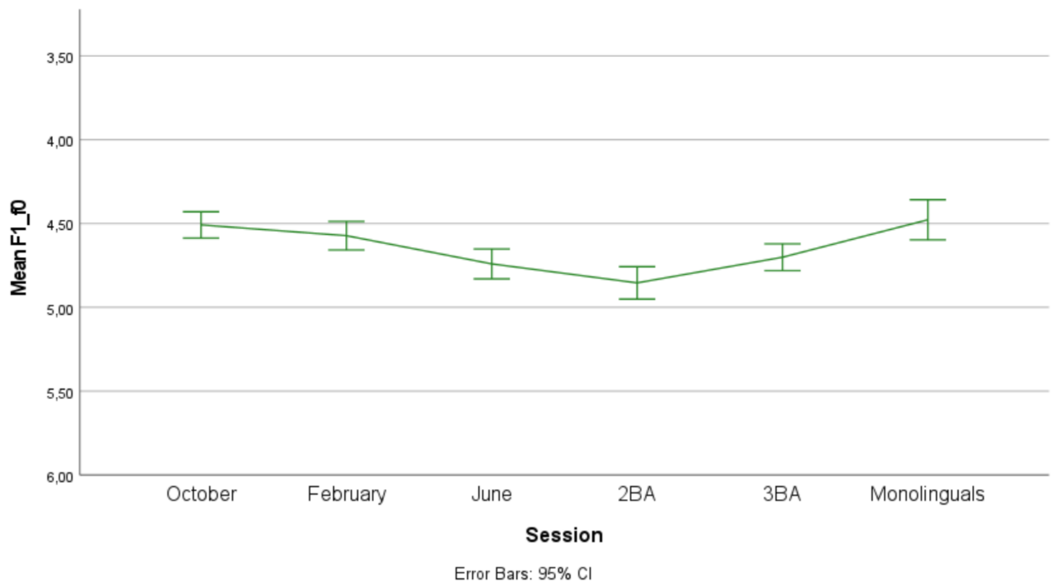


Fig. 65. Mean F1 (F1-f_0; Bark normalised) values of the vowel /ɛ/ in the word reading task, sorted for Group/Session; word reading task.

While the differences are very small, we can observe some lowering taking place, with the F1-f_0 difference getting bigger, over the course of the phonetic instruction in L2. In order to see whether those differences were statistically significant we looked at the results of the GLMM model with F1 (F1-f_0) as the dependent variable and the interaction of Vowel*Session as the main predictor of interest, and Speaker and Vowel as random factors. Table 48 reports on the model for /ɛ/ only; pairwise comparisons are shown and the significant differences are bolded. The full coefficient table is available in Appendix 7.

Table 48. The results of the GLMM model with F1 (F1-f0) of /ɛ/ as the dependent variable (pairwise comparisons; word reading task).

Session	Contrast estimate	Std. Error	t	Significance
T1 vs. T2	-.058	.049	-1.174	.241
T1 vs. T3	**-.223**	**.050**	**-4.487**	**<.000**
T2 vs. T3	**-.164**	**.049**	**-3.325**	**.001**
T1 vs. Group 2	**-.324**	**.147**	**-2.212**	**.027**
T2 vs. Group 2	-.266	.147	-1.816	.069
T3 vs. Group 2	-.102	.147	-.695	.487
T1 vs. Group 3	-.179	.147	-1.221	.222
T2 vs. Group 3	-121	.147	-.826	.409
T3 vs. Group 3	.043	.147	.294	.768
Group 2 vs. Group 3	.145	.157	.925	.355
T1 vs. Monolinguals	.043	.148	.291	.771

T2 vs. Monolinguals	.101	.148	.682	.495
T3 vs. Monolinguals	.266	.148	1.792	.073
Group 2 vs. Monolinguals	**.368**	**.158**	**2.323**	**.02**
Group 3 vs. Monolinguals	.222	.158	1.404	.160

With respect to height, the analysis revealed that there was a significant difference between 1BA students between T1 and T3 (*p*<.001) and T2 and T3 (*p*=.001). T1 was also different from 2BA students, that is Group 2 (*p*=.027), while Group 2 differed from the monolingual controls (*p*=.02). No other contrasts turned out to be significant. Therefore the most visible effects of phonetic drift on the height of /ɛ/ were present in after eight months and two years of pronunciation instruction, with the vowel getting slightly lower.

Now let us turn to the vowel /a/. Out of the 3189 obtained productions, Group 1 provided 1854 items (618 at T1, 612 at T2, and 624 at T3), Group 2 – 467 items, Group 3 – 468, while Group 4 – 400. The mean F1 values for each group are given in Table 49.

Table 49. Mean values of F1 (F1-f0; in Bark) of /a/, sorted for Group/Session; word reading task.

Vowel	Session/Group	N	Mean (Bark)	Std. Deviation
/a/	Group 1: T1	618	5.58	.64
	Group 1: T2	612	5.75	.65
	Group 1: T3	624	5.88	.57
	Group 2: 2BA	467	5.91	.57
	Group 3: 3BA	468	5.88	.61
	Group 4: Monolinguals	400	5.36	.75

Fig. 66 traces the changes in the mean height values of the vowel /a/ over time and compares them to the monolingual norms. Once again, the *y* axis has been inverted for ease of interpretation.

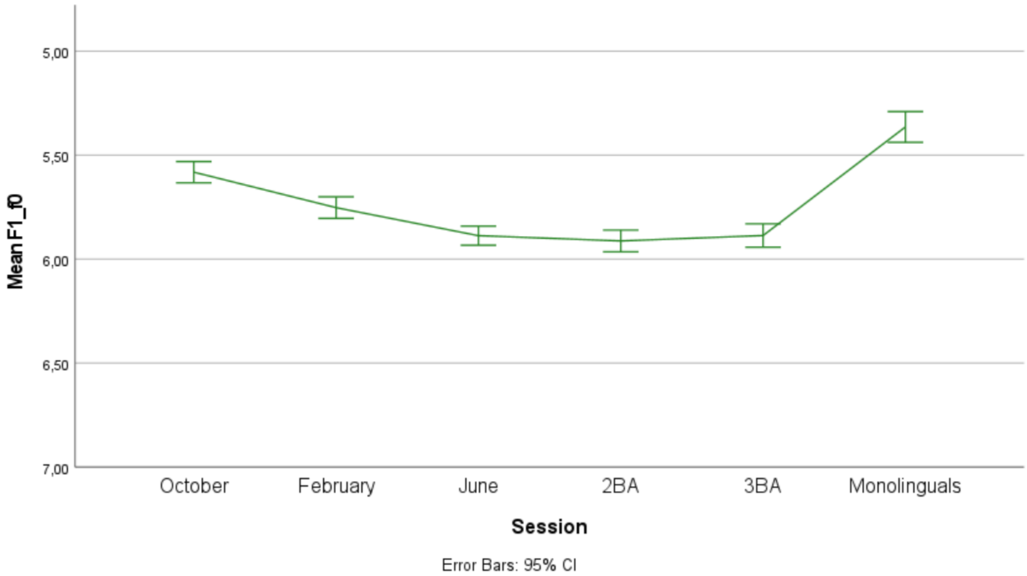


Fig. 66. Mean F1 (F1-f0; Bark normalised) values of the vowel /a/, sorted for Group/Session; word reading task.

Again, as we can see the vowel appears to get progressively lower (with the F1-f_0 difference getting bigger) over the course of training. The results of the GLMM model with F1 (F1-f0) as the dependent variable and the interaction of Vowel*Session as the main predictor of interest, and Speaker and Vowel as random factors are shown in Table 50 for /a/ only. Pairwise comparisons are given and the significant differences are bolded.

Table 50. The results of the GLMM model with F1 (F1-f0) of /a/ as the dependent variable (pairwise comparisons; word reading task).

Session	Contrast estimate	Std. Error	t	Significance
T1 vs. T2	**-.169**	**.028**	**-6.051**	**<.001**
T1 vs. T3	**-.299**	**.028**	**-10.798**	**.000**
T2 vs. T3	**-.131**	**.028**	**-4.707**	**<.001**
T1 vs. Group 2	**-.322**	**.140**	**-2.306**	**.021**
T2 vs. Group 2	-.154	.140	-1.100	.271
T3 vs. Group 2	-.023	.140	-.165	.869
T1 vs. Group 3	**-.292**	**.140**	**-2.084**	**.037**
T2 vs. Group 3	-.123	.140	-.879	.380
T3 vs. Group 3	.008	.140	.057	.955
Group 2 vs. Group 3	.031	.149	.207	.836
T1 vs. Monolinguals	.232	.140	1.654	.098
T2 vs. Monolinguals	**.400**	**.140**	**2.857**	**.004**
T3 vs. Monolinguals	**.531**	**.140**	**3.791**	**.000**
Group 2 vs. Monolinguals	**.554**	**.150**	**3.700**	**.000**
Group 3 vs. Monolinguals	**.523**	**.150**	**3.494**	**.000**

In the case of Group 1, the progressive lowering of /a/ turned out to be significant across all testing times: T1 vs. T2 (p<.001), T2 vs. T3 (p<.001), as well as T1 and T3 (p=.000). 1BA students at T1 differed also from Group 2 (p=.021) and Group 3 (p=.037). Additionally, aside from 1BA students at T1 (p=.098), all groups were significantly different from the monolingual controls.

Overall, we observed differences in vowel height which appear to be signs of vocalic phonetic drift induced by L2 phonetic training. The following section looks at vowel advancement.

6.3.1.2. Vowel advancement

Now let us look at vowel frontness/backness with F2 as the dependent variable. Mean F2 values (Bark normalised, F3-F2 from the middle 20% of the vowel) to measure vowel height were gathered for each of the two vowels. In total, 4179 tokens were analysed, out of which /ɛ/ comprised 990 items and /a/ – 3189 items. Similarly to what was the case for F1, we will discuss each vowel individually.

As far as /ɛ/ is concerned, the mean F2 values for each group are given in Table Table 51.

Table 51. Mean values of F2 (F3-F2 in Bark) of /ɛ/, sorted for Group/Session; word reading task.

Vowel	Session/Group	N	Mean (Bark)	Std. Deviation
/ɛ/	Group 1: T1	194	2.74	.60
	Group 1: T2	196	2.61	.63
	Group 1: T3	194	2.69	.59
	Group 2: 2BA	147	2.59	.51
	Group 3: 3BA	144	2.41	.69
	Group 4: Monolinguals	115	3.00	.60

A visual representation of the changes with respect to the advancement of /ɛ/ is shown in Fig. 67. The *x* and *y* axes have been swapped for the reader to clearly see the changes in fronting across the groups.

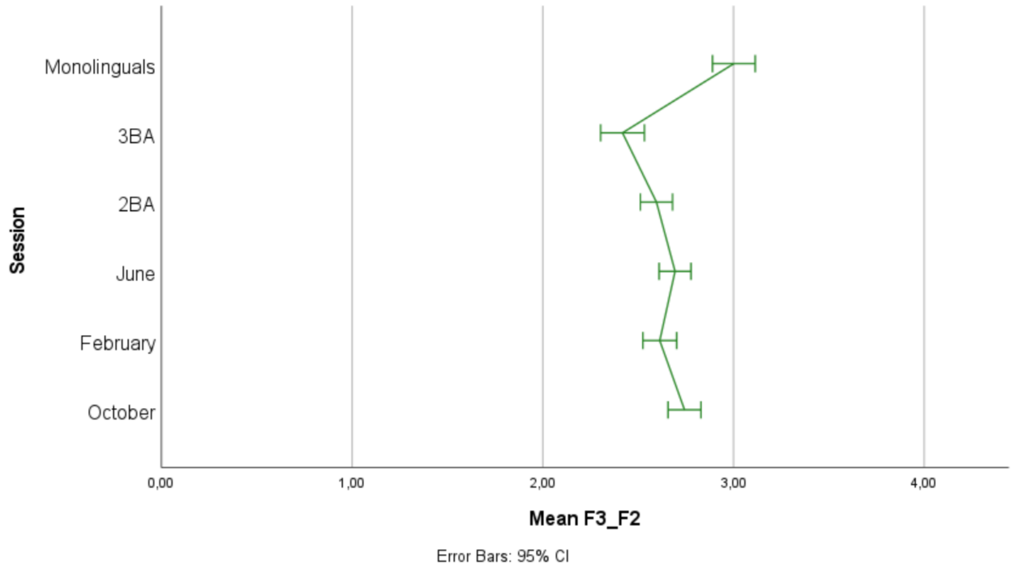


Fig. 67. Mean F2 (F3-F2; Bark normalised) values of the vowel /ɛ/, sorted for Group/Session; word reading task.

As can be observed, /ɛ/ appears to move towards more front realisations as the phonetic instruction progresses, peaking in the productions of 3BA's (i.e. Group 3) students. In order to see whether these effects were significant, a GLMM model with F2 (F3-F2) as the dependent variable and the interaction of Vowel*Session as the main predictor and Speaker and Vowel as random factors was run. The results for thereof (for /ɛ/ only) are shown in Table 52. Pairwise comparisons are given with the significant differences given in bold. Full coefficient table can be found in Appendix 7.

Table 52. The results of the GLMM model with F2 (F3-F2) of /ɛ/ as the dependent variable (pairwise comparisons; word reading task).

Session	Contrast estimate	Std. Error	t	Significance
T1 vs. T2	**.126**	**.062**	**2.042**	**.041**
T1 vs. T3	.054	.062	.871	.384
T2 vs. T3	-.072	.062	-1.169	.243
T1 vs. Group 2	.144	.167	.864	.388
T2 vs. Group 2	.018	.167	.109	.913
T3 vs. Group 2	.090	.167	.541	.588
T1 vs. Group 3	**.331**	**.167**	**1.981**	**.048**
T2 vs. Group 3	.205	.167	1.227	.220
T3 vs. Group 3	.277	.167	1.659	.097
Group 2 vs. Group 3	.187	.179	1.046	.295
T1 vs. Monolinguals	-.258	.169	-1.525	.127
T2 vs. Monolinguals	**-.384**	**.169**	**-2.271**	**.023**
T3 vs. Monolinguals	-.312	.169	-1.844	.065

Group 2 vs. Monolinguals	**-.402**	**.180**	**-2.228**	**.026**
Group 3 vs. Monolinguals	**-.589**	**.181**	**-3.261**	**.001**

The statistical analysis revealed that Group 1 at T1 differed significantly from both Group 1 at T2 (*p*=.041) and Group 3 (*p*=.048), but only just. In turn the productions of the monolingual group were significantly less front than Group 1 at T2 (*p*=.023), and both Group 2 (*p*=.026) and Group 3 (*p*=.001). No other contrasts turned out to be significant.

Decidedly less variation was present in the central vowel /a/. The mean F2 values for each group are given in Table 53.

Table 53. Mean values of F2 (F3-F2 in Bark) of /a/, sorted for Group/Session; word reading task.

Vowel	Session/Group	N	Mean (Bark)	Std. Deviation
/a/	Group 1: T1	618	3.77	.75
	Group 1: T2	612	3.80	.82
	Group 1: T3	624	3.79	.73
	Group 2: 2BA	467	3.81	.84
	Group 3: 3BA	468	3.74	.81
	Group 4: Monolinguals	400	3.75	.73

As can be seen, the differences in F3-F2 distance are very small in magnitude, indicating very little change in the advancement of /a/. This is illustrated also in Fig. 68. Once again, the *x* and *y* axes have been swapped to better illustrate the changes in fronting.

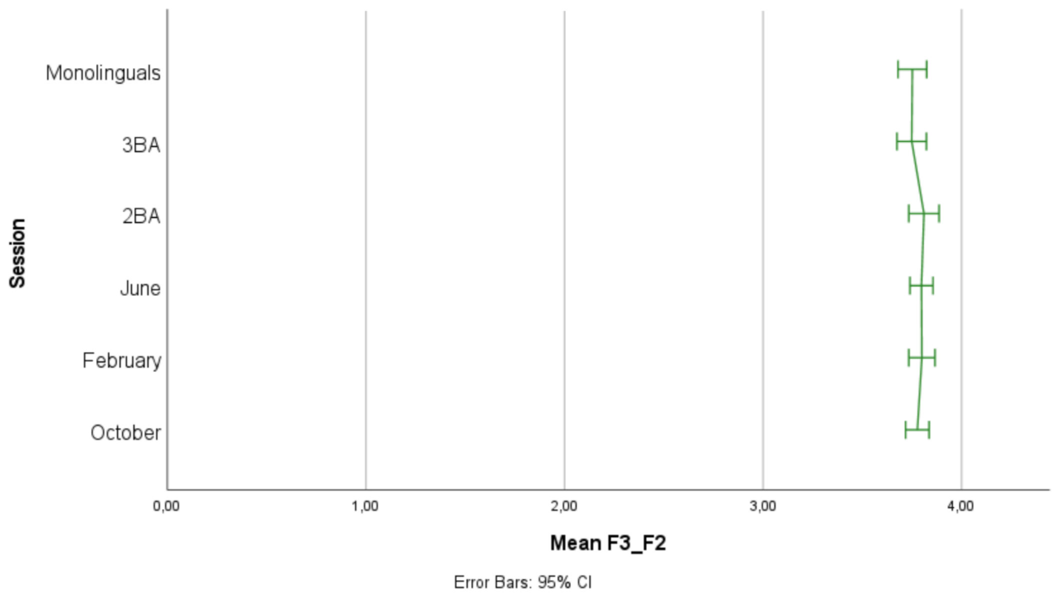


Fig. 68. Mean F2 (F3-F2; Bark normalised) values of the vowel /a/, sorted for Group/Session; word reading task.

Indeed, the results of the GLMM model with F2 (F3-F2) as the dependent variable, Vowel*Session as the fixed factor, and Speaker and Item as random factors indicate that the differences were not significant. Table 54 provides details of the pairwise comparisons.

Table 54. The results of the GLMM model with F2 (F3-F2) of /a/ as the dependent variable (pairwise comparisons; word reading task).

Session	Contrast estimate	Std. Error	t	Significance
T1 vs. T2	-.020	.035	-.581	.561
T1 vs. T3	-.012	.035	-.333	.739
T2 vs. T3	.009	.035	.251	.802
T1 vs. Group 2	-.029	.158	-.182	.855
T2 vs. Group 2	-.009	.158	-.054	.957
T3 vs. Group 2	-.017	.158	-.109	.913
T1 vs. Group 3	.032	.158	.204	.838
T2 vs. Group 3	.052	.158	.332	.740
T3 vs. Group 3	.044	.158	.277	.782
Group 2 vs. Group 3	.061	.168	.362	.718
T1 vs. Monolinguals	.021	.158	.136	.892
T2 vs. Monolinguals	.042	.158	.264	.792
T3 vs. Monolinguals	.033	.158	.209	.835
Group 2 vs. Monolinguals	.50	.169	.297	.766
Group 3 vs. Monolinguals	-.011	.169	-.063	.949

Overall we found some evidence for phonetic drift effects triggered by L2 phonetic training in the speech of our participants with respect to vowel advancement for /ɛ/ but not for /a/. We will come back to these findings in the general discussion of the results (cf. sections 6.4.3 and 8.1).

6.3.2. Sentence reading task

In the case of the second task, once again the two vowels – /a/ and /ɛ/ – were plotted on F1/F2 planes in R (R Core Team 2018), using the `ggplot2()` package (Wickham 2016). Although the dataset did include two other vowels, namely /ɨ/ and /ɔ/, these were excluded from the analysis. The former was not considered as we did not want to include another vocalic category which we could not compare with what was found in the previous task.

As has been mentioned, we did not differentiate between the three prosodic positions in which the vowels were placed as previous research has shown negligible effects of position on vowel quality in Polish (Wojtkowiak 2020). The latter was excluded on similar grounds as to what was the case for the word lists, that is problems with mean formant measurements extraction.

In our visual inspection of the results, all data points obtained were mapped, with the ellipses corresponding to the distance of one standard deviation from the mean F1 (F1-f_0) and F2 (F3-F2) for each of the groups and sessions included in the present study.

Let us begin with the vowel /ɛ/. Fig. 69 presents the average formant values thereof.

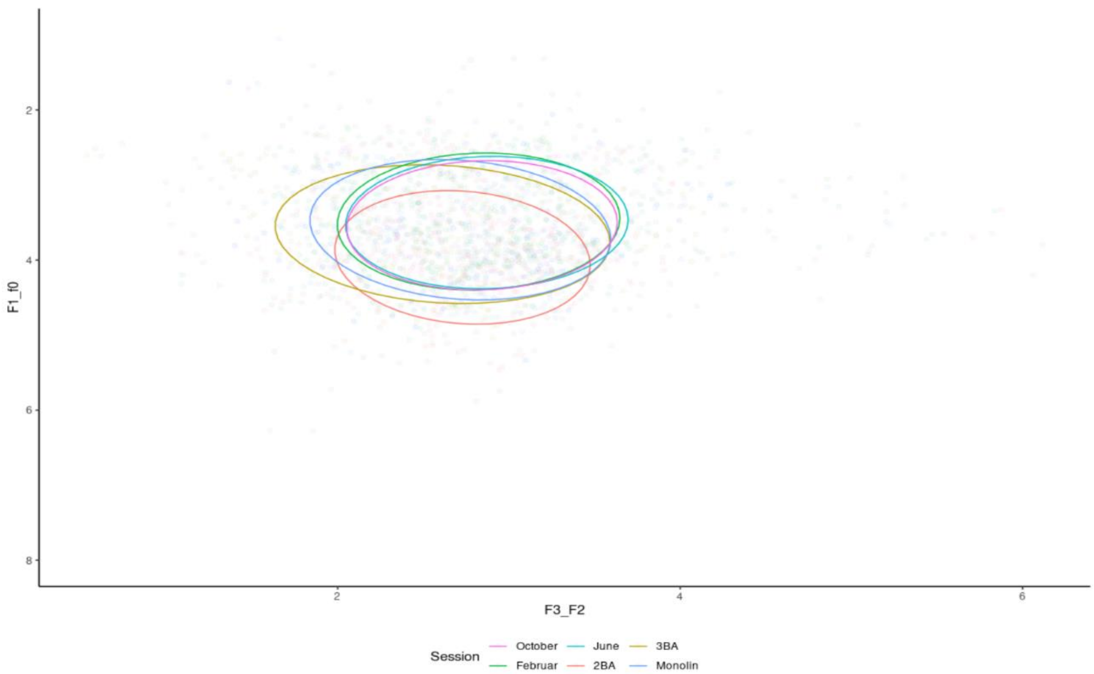


Fig. 69. Polish /ɛ/ in the sentence reading task plotted on the F1/F2 plane, with the ellipses corresponding to the distance of one standard deviation from the mean, sorted for group (and session).

The figure shows that with respect to vowel height there is not much variation across the groups: T1 (in pink), T2 (in green), T3 (in teal) as well as Group 3 (in yellow) and Group 4 (that is, the monolingual controls in blue) are relatively comparable. We can see that Group 2 (in red) appears to produce /ɛ/ as slightly lower than the rest. With respect to advancement, once again the groups are quite uniform, however it appears that both the monolinguals and – in particular – Group 3 realise the /ɛ/ vowel as more front than the remaining groups. On the whole, though, there appears to be less group differences present than what was observed in the previous task.

As far as /a/ is concerned, Fig. 70 illustrates the data, once again sorted for groups/sessions.

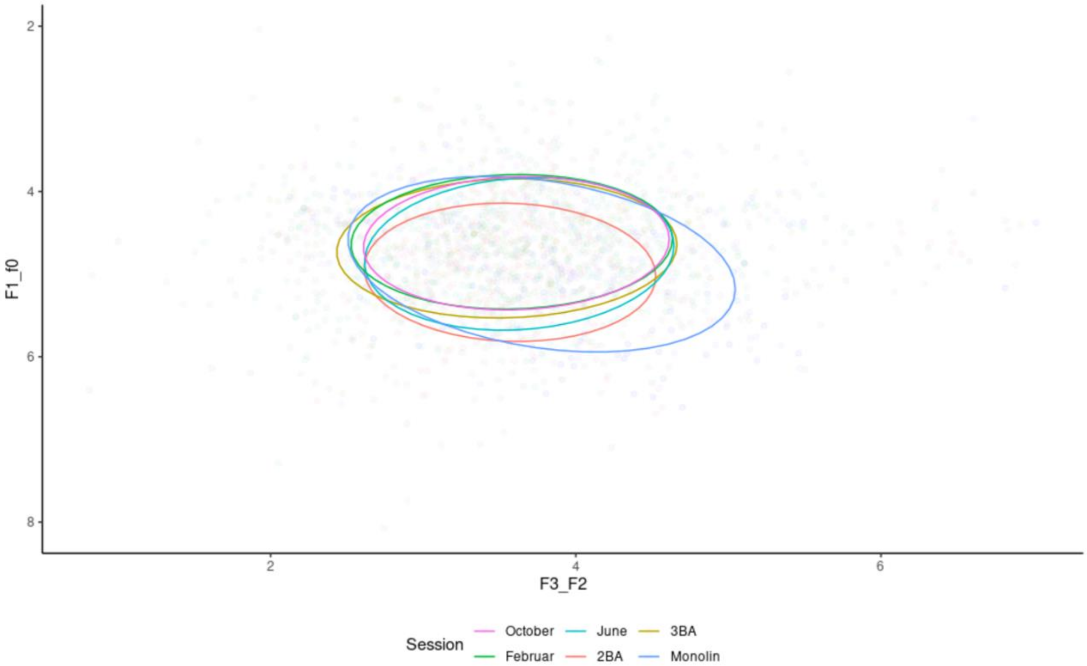


Fig. 70. Polish /a/ in the sentence reading task plotted on the F1/F2 plane, with the ellipses corresponding to the distance of one standard deviation from the mean, sorted for group (and session).

Overall, there appears to be less variability across the students' groups, as their productions are all comparable, except for – once again – Group 2 (in red), whose /a/ is produced as slightly lower. There are no robust differences as far as advancement is considered, however the monolinguals' ellipsis encompasses quite a bit of space, indicating some variation in vowel frontness, which was not really present in the previous task.

The following subsections will discuss the two dimensions – i.e. height and advancement – for each vowel separately in order to investigate the possible L1 drift effects.

6.3.2.1. Vowel height

As far as the results associated with vowel height are concerned, the mean F1 values (Bark normalised, F1-f_0 from the middle 20% of the vowel) were gathered for each of the two vowels. In total, 3893 tokens were analysed: 1954 of /ɛ/ and 1939 of /a/.

Let us begin similarly to how we approached this topic in the previous task and discuss each vowel individually. Out of the 1954 obtained productions of /ɛ/, Group 1 provided 1066 items (352 at T1, 359 at T2, and 355 at T3), Group 2 – 259 items, Group 3 – 263, while Group 4 – 366. The average F1 values for each group are provided in Table 55.

Table 55. Mean values of F1 (F1-f_0; in Bark) for /ɛ/ in the sentence reading task, sorted for Group/Session.

Vowel	Session/Group	N	Mean (Bark)	Std. Deviation
/ɛ/	Group 1: T1	352	3.48	.64
	Group 1: T2	359	3.46	.80
	Group 1: T3	355	3.49	.66
	Group 2: 2BA	259	3.93	.68
	Group 3: 3BA	263	3.62	.72
	Group 4: Monolinguals	366	3.55	.70

The changes in the height of /ɛ/ are illustrated in Fig. 71. Recall, that the higher the difference between F1 and f_0, the lower the vowel; the *y* axis has been nonetheless reversed for ease of interpretation.

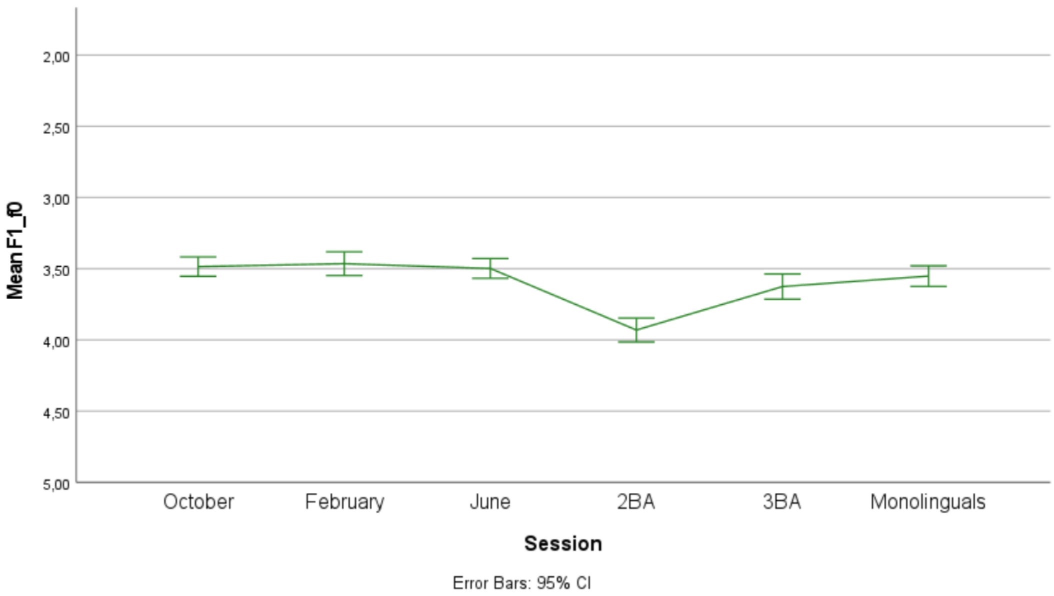


Fig. 71. Mean F1 (F1-f_0; Bark normalised) values of the vowel /ɛ/ in the sentence reading task, sorted for Group/Session.

While the differences in height were notably less robust than what was observed in the previous task, Group 2 (i.e. second year students) again showed most salient changes, relative to the monolingual norms. The results of a GLMM model with F1 (F1-f0) as the dependent variable, the interaction between Session*Vowel as the main predictor[86] and Speaker and Item as random factors are presented in Table 56. The significant differences in pairwise comparisons are bolded while the full coefficient table is given in Appendix 7.

Table 56. The results of the GLMM model with F1 (F1-f0) of /ɛ/ as the dependent variable (pairwise comparisons) in the sentence reading task.

Session	Contrast estimate	Std. Error	t	Significance
T1 vs. T2	.015	.042	.351	.726
T1 vs. T3	-.016	.042	-.378	.705
T2 vs. T3	-.031	.042	-.351	.726
T1 vs. Group 2	**-.444**	**.155**	**-2.863**	**.004**
T2 vs. Group 2	**-.459**	**.155**	**-2.960**	**.003**
T3 vs. Group 2	**-.428**	**.155**	**-2.759**	**.006**
T1 vs. Group 3	-.140	.155	-.900	.368
T2 vs. Group 3	-.154	.155	-.997	.319
T3 vs. Group 3	-.124	.155	-.797	.425

[86] Both /ɛ/ and /a/ were included in the same statistical analyses but are reported on separately for clarity purposes.

Group 2 vs. Group 3	.304	.166	1.836	.066
T1 vs. Monolinguals	-.073	.170	-.429	.668
T2 vs. Monolinguals	-.088	.170	-.516	.606
T3 vs. Monolinguals	-.057	.170	-.334	.738
Group 2 vs. Monolinguals	**.371**	**.180**	**2.061**	**.039**
Group 3 vs. Monolinguals	.067	.180	.370	.711

The results indicate that, indeed, Group 2's productions were the only ones that differed significantly from the monolingual norms (*p*=.039); they were also different from Group 1's productions at all testing times (*p*=.004, *p*=.003, and *p*=.006, respectively) but not from Group 3's (*p*=.066). No other contrasts were detected.

With /a/, out of all items gathered, Group 1 provided 1043 items (344 at T1, 350 at T2, and 349 at T3), Group 2 – 261 items, Group 3 – 263, while Group 4 – 372. The average F1 values for each group are provided in Table 57.

Table 57. Mean values of F1 (F1-f_0; in Bark) for /a/ in the sentence reading task, sorted for Group/Session.

Vowel	Session/Group	N	Mean (Bark)	Std. Deviation
/a/	Group 1: T1	344	4.63	.64
	Group 1: T2	350	4.61	.65
	Group 1: T3	349	4.75	.71
	Group 2: 2BA	261	4.98	.66
	Group 3: 3BA	263	4.70	.68
	Group 4: Monolinguals	372	4.85	.79

These mean values are illustrated in Fig. 72. Once again, the *y* axis has been reversed for clarity.

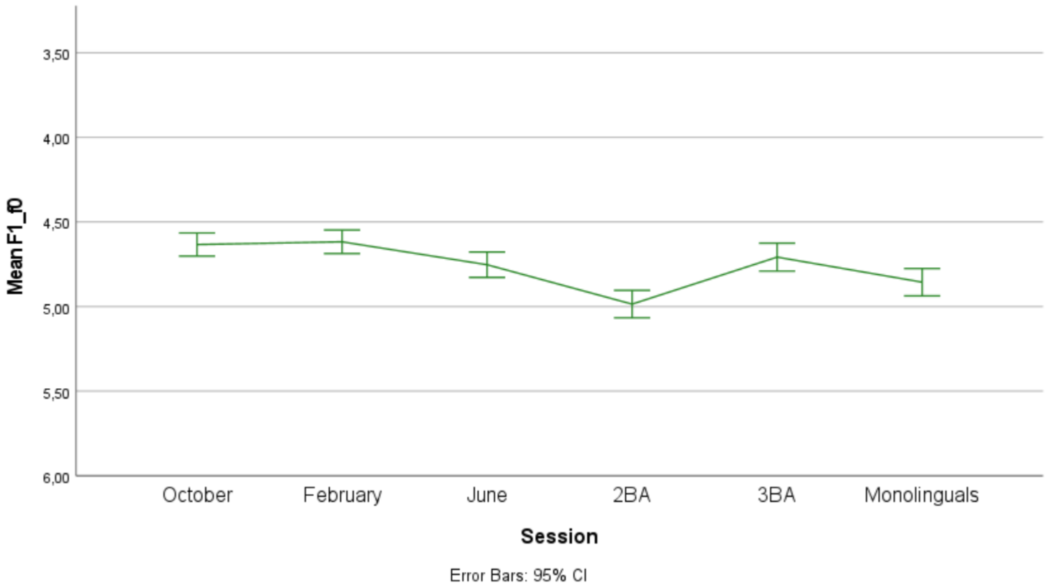


Fig. 72. Mean F1 (F1-f_0; Bark normalised) values of the vowel /a/ in the sentence reading task, sorted for Group/Session.

Fig. 72 shows that there was some vowel lowering between the groups, which started around T2 and peaked in Group 2's production, for whom the difference between F1-f_0 is the largest. After that point, the vowel got slightly raised again, returning to the initial values in Group 3's data, i.e. the group who were no longer undergoing any phonetic instruction.

Table 58 presents the results of the GLMM model (with the specifications the same as for /ɛ/ before.

Table 58. The results of the GLMM model with F1 (F1-f0) of /a/ as the dependent variable (pairwise comparisons) in the sentence reading task.

Session	Contrast estimate	Std. Error	t	Significance
T1 vs. T2	.012	.043	.285	.775
T1 vs. T3	**-.122**	**.043**	**-2.842**	**.005**
T2 vs. T3	**-.134**	**.043**	**-3.140**	**.002**
T1 vs. Group 2	**-.360**	**.155**	**-2.322**	**.02**
T2 vs. Group 2	**-.372**	**.155**	**-2.402**	**.016**
T3 vs. Group 2	-.238	.155	-1.536	.125
T1 vs. Group 3	-.080	.155	-.515	.607
T2 vs. Group 3	-.092	.155	-.594	.553
T3 vs. Group 3	.042	.155	.272	.786
Group 2 vs. Group 3	.280	.166	1.692	.091
T1 vs. Monolinguals	-.217	.170	-1.279	.201
T2 vs. Monolinguals	-.229	.170	-1.351	.177

T3 vs. Monolinguals	-.095	.170	-.561	.575
Group 2 vs. Monolinguals	.143	.180	.797	.426
Group 3 vs. Monolinguals	-.137	.180	-.765	.445

The statistical analysis shows that indeed, while no difference was found between T1 and T2, /a/ was lower at T3 relative to both T2 (p=.002) and T1 (p=.005). Group 2's realisation was also lower than both T2 (p=.016) and T1 (p=.02), but not T3 (p=.125). No other contrasts were detected. Interestingly, at no point any divergence from the monolingual norms was found.

6.3.2.2. Vowel advancement

Finally, turning to vowel frontness/backness with F2 as the dependent variable, mean F2 values (Bark normalised, F3-F2 from the middle 20% of the vowel) to measure vowel height were obtained for each of the two vowels. 3893 tokens were analysed: 1954 of /ɛ/ and 1939 of /a/. Seeing as the measurements were taken from the exact same vowels as what was presented for height, the numbers of tokens obtained from each group will not be repeated here (they are included in the tables below).

Beginning with /ɛ/, Table 59 presents mean values of F2 (F3-F2 in Bark) alongside standard deviations.

Table 59. Mean values of F2 (F3-F2; in Bark) for /ɛ/ in the sentence reading task, sorted for Group/Session.

Vowel	Session/Group	N	Mean (Bark)	Std. Deviation
/ɛ/	Group 1: T1	352	2.86	.66
	Group 1: T2	359	2.85	.69
	Group 1: T3	355	2.86	.65
	Group 2: 2BA	259	2.73	.64
	Group 3: 3BA	263	2.62	.78
	Group 4: Monolinguals	366	2.73	.67

The visual representation of the mean values in Table 59 is shown in Fig. 73. Recall that the smaller the difference between F3-F2, the higher the F2, and the fronter the vowel; the *x* and *y* axes have been swapped to better illustrate the changes in frontness.

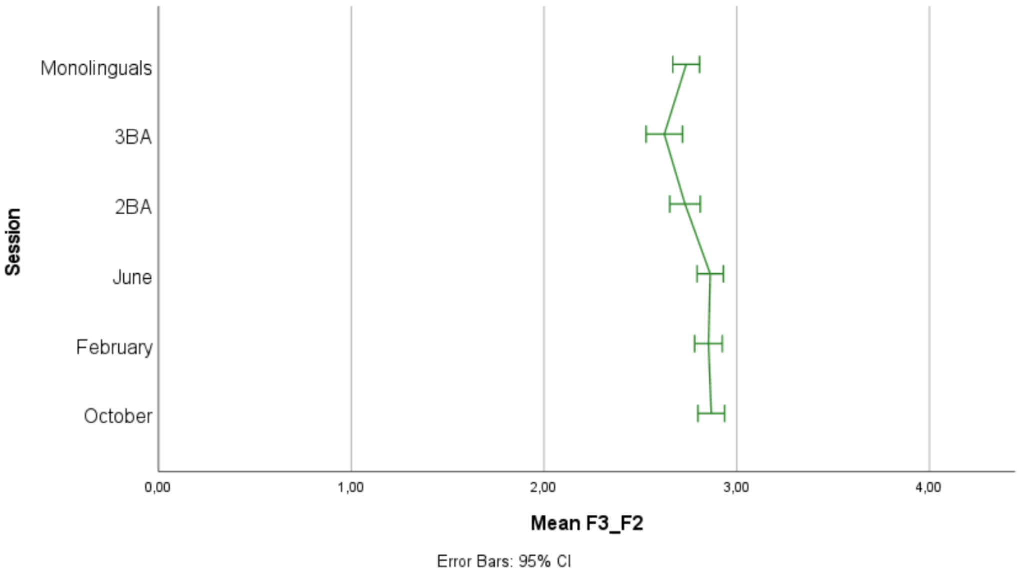


Fig. 73. Mean F2 (F3-F2; Bark normalised) values of the vowel /ɛ/ in the sentence reading task, sorted for Group/Session.

In Fig. 73 we see that there was some fronting of /ɛ/ taking place, which is most visible between T3 and Group 2, and Group 2 and Group 3. Both these groups' vowels appear to be slightly more front than the monolingual norms. a GLMM model with F2 (F3-F2) of /ɛ/ as the dependent variable and the interaction of Vowel*Session as the main predictor and Speaker and Vowel as random factors was run. The results for thereof (for /ɛ/ only) are shown in Table 60. Pairwise comparisons are given with the significant differences given in bold. Full coefficient table can be found in Appendix 7.

Table 60. The results of the GLMM model with F2 (F3-F2) of /ɛ/ as the dependent variable (pairwise comparisons) in the sentence reading task.

Session	Contrast estimate	Std. Error	t	Significance
T1 vs. T2	.013	.050	.256	.798
T1 vs. T3	.002	.050	.045	.964
T2 vs. T3	-.010	.049	-.211	.833
T1 vs. Group 2	.147	.152	.970	.332
T2 vs. Group 2	.134	.151	.887	.375
T3 vs. Group 2	.145	.151	.956	.339
T1 vs. Group 3	.240	.151	1.587	.113
T2 vs. Group 3	.228	.151	1.504	.133
T3 vs. Group 3	.238	.151	1.572	.116
Group 2 vs. Group 3	.093	.162	.576	.565
T1 vs. Monolinguals	.100	.235	.423	.672
T2 vs. Monolinguals	.087	.235	.370	.712
T3 vs. Monolinguals	.097	.235	.414	.679

Group 2 vs. Monolinguals	-.047	.242	-.196	.845
Group 3 vs. Monolinguals	-.141	.242	-.581	.561

The analysis revealed that, in fact, no significant changes in /ɛ/ advancement were observed.

Finally, Table 61 presents mean values of F2 (F3-F2 in Bark) for /a/.

Table 61. Mean values of F2 (F3-F2; in Bark) for /a/ in the sentence reading task, sorted for Group/Session.

Vowel	Session/Group	N	Mean (Bark)	Std. Deviation
/a/	Group 1: T1	344	3.61	.79
	Group 1: T2	350	3.57	.84
	Group 1: T3	349	3.62	.79
	Group 2: 2BA	261	3.63	.83
	Group 3: 3BA	263	3.58	.89
	Group 4: Monolinguals	372	3.84	.95

The obtained averages for the advancement of /a/ in each group are plotted in Fig. 74. The *x* and *y* axes have been swapped for ease of interpretation.

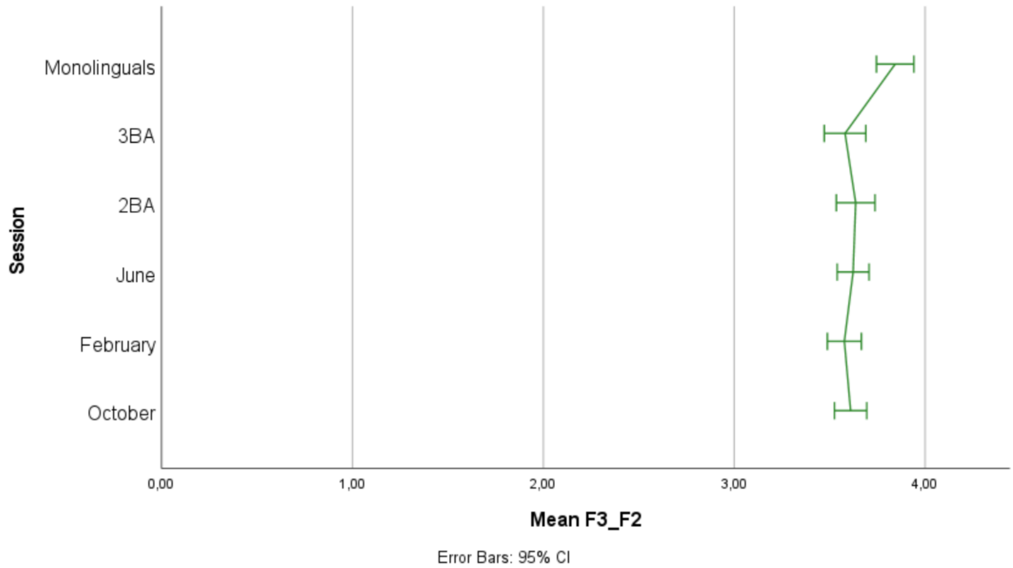


Fig. 74. Mean F2 (F3-F2; Bark normalised) values of the vowel /a/ in the sentence reading task, sorted for Group/Session.

By and large, the changes between the groups as far as /a/ frontness are concerned are even more negligible than what was the case for /ɛ/. The one notable difference is the contrast between the students and the quasi-monolinguals, with /a/ being more central for the latter. The results of the GLMM model with F2 (F3-F2) of /a/ as the dependent variable (and the rest of the details identical as in /ɛ/, seeing as both vowels were subject to the same analysis) are given in Table 62. Pairwise comparisons follow.

Table 62. The results of the GLMM model with F2 (F3-F2) of /a/ as the dependent variable (pairwise comparisons) in the sentence reading task.

Session	Contrast estimate	Std. Error	t	Significance
T1 vs. T2	.041	.050	.826	.409
T1 vs. T3	-.009	.050	-.172	.863
T2 vs. T3	-.050	.050	-1.001	.317
T1 vs. Group 2	-.017	.152	-.113	.910
T2 vs. Group 2	-.059	.152	-.387	.699
T3 vs. Group 2	-.009	.152	-.057	.955
T1 vs. Group 3	.036	.152	.236	.814
T2 vs. Group 3	-.006	.151	-.038	,970
T3 vs. Group 3	.044	.151	.293	.770
Group 2 vs. Group 3	.053	.162	.327	.744
T1 vs. Monolinguals	-.268	.233	-1.149	.251
T2 vs. Monolinguals	-.309	.233	-1.326	.185
T3 vs. Monolinguals	-.259	.233	-1.112	.266
Group 2 vs. Monolinguals	-.251	.240	-1.044	.297
Group 3 vs. Monolinguals	-.304	.240	-1.264	.206

Similarly to /ɛ/, no significant contrasts were observed.

6.4. Discussion of the results in L1

The present phonetic study has investigated the effects of undergoing intensive phonetic training in English by Polish undergraduate students on the acoustics of their L1. The main objective thereof was to conduct an empirical acoustic experiment which would attest the existence of L1 drift in the productions of the chosen population and second of all, test the phonological hypotheses put forward by selected frameworks. In order to do that, we have looked at both initial consonants and the vowels that followed them and evaluated changes between the sessions and groups by means of assessing pairwise comparisons. Table 63 presents the general results of the two production studies on Polish.

Table 63. General summary of the two production studies on Polish.

Task	Voiceless stops	Voiced stops	Vowels
word reading	no drift effects	drift in length of pre-voicing and the likelihood of the occurrence of unvoiced items	vowel space expansion
sentence reading	no drift effects	drift in length of pre-voicing and the likelihood of the occurrence of unvoiced items	vowel space expansion

The sections below summarise the findings in some more detail and couch them within the larger body of empirical data on L2-induced phonetic drift.

6.4.1. VOT

Recall that we looked at VOT values of voiceless (/p, t, k/) and voiced (/b, d, g/) series of stops in L1 Polish in two different tasks: word list reading and sentence list reading. Together these two production studies have analysed the VOT values of 12,773 consonants across the two tasks and all groups and sessions. Following Table 63, we will discuss the voiceless series first, after which we will proceed to the voiced category (which will be discussed according to pre-voicing type).

As far as the voiceless series is concerned, the word-reading task provided data from 65 speakers overall who were divided into four groups: first year students (Group 1, N=20, tested three times over the course of the entire academic year), second year students (Group 2, N=15), third year students (Group 3, N=15), and the quasi-monolingual group of controls (Group 4, N=15). Overall 2925 items were analysed.

Group results of /p, t, k/, with the items collapsed across place of articulation, indicated that there was not much L2-induced phonetic drift to be observed. The VOT of all groups oscillated at ca. 41 ms and any changes in the duration across the testing times fell within 2-3 ms, which is a very minute difference (cf. Fig. 41). In general, the VOT duration in this task was relatively long when one considers that Polish voiceless consonants (cf. Table 38) are supposed to be realised with a short-lag voice onset time. This might stem from the specificity of the task, with citation forms yielding more domain-initial lengthening effects when produced in isolation, which has been attested by previous scholars (e.g. Lisker and Abramson 1964).

The only significant difference we observed happened in the productions of Group 1: VOT got longer between T1 and T2 and then subsequently shorter between T2 and T3 (essentially getting back to the values found at T1 as there was no difference between T1 and T3). We will return to this special case of T2 shortly, as it will also be important in the discussion of the voiced series' results.

When we split the results according to the place of articulation of each plosive, in the voiceless series the differences between sessions and groups were once again minimal. The VOT duration of /p/ and /t/ was quite comparable and in general longer than what is expected in Polish as it reached ca. 35 ms. The VOT of /k/ was decidedly longer, with values reaching 55-60 ms on average. The only significant difference found across the three places of articulation occurred between T1 and T2 in Group 1's productions in the case of both /t/ and /k/, with VOT being lengthened at T2. This lengthening, however, was not robust as the difference between T1 and T2 for /t/ and /k/ was ca. 3 ms and 2 ms, respectively. Importantly, no difference between the students' and monolinguals' productions was observed.

As far as the sentence reading task in this laryngeal category goes, productions from 70 speakers were considered (20 first year students or Group 1; 15 second year students or Group 2; 15 third year students or Group 3; and 20 quasi-monolingual controls or Group 4). There were 3425 voiceless stops included in the analysis which were placed at three different prosodic positions: utterance-initial, phrase-initial (i.e. utterance-medial, and phrase-medial. Place of articulation was not included as a factor in the analysis as the dataset was counterbalanced for place and voicing of the consonant.

While not much variability with respect to prosodic positions was observed, in utterance-initial and phrase-initial contexts the students from Group 1 produced significantly shorter VOT relative to the monolingual norms at all three testing times. No such differences between students and monolinguals were observed in the phrase-medial context, nor in the productions of Group 2 and Group 3. Group 2 produced slightly longer VOT than Group 1 students at T1 and T3 (but not at T2) in the utterance position. Despite these small disparities, no consistent effects of L2 on L1 VOT duration were observed; VOT did not move in the direction of aspirated, English-like stops. On the contrary, in the cases where the contrast between students and monolinguals was found, the students' VOT was shorter than that of the controls. As has been mentioned in the descriptions of the statistic results, it may be suggested that very small effects of domain-initial lengthening were observed in the speech of the quasi-monolinguals but not the students; these effects were nonetheless almost negligible, as the contrast between phrase-medial and phrase-initial/utterance-initial positions in the monolingual productions did not exceed 2 ms. An interesting observation to be made, however, is the length of VOT in this task compared to the previous one; VOT was decidedly shorter in sentence productions.

All in all, the voiceless series has proven to be quite resistant to L2-induced phonetic drift; even when some significant contrasts were detected by the statistical models, the sheer differences in VOT duration between groups, sessions, and prosodic positions were very small. No consistent patterns were found in this category. As has been already mentioned in section 6.2.3, the fact that Polish /p, t, k/ remained unaffected by the phonetic training in L2 that our students were undergoing, from the perspective of SLM (Flege 1995) this suggests that a new category has been formed for the aspirated English variants [p^h, t^h, k^h] and, as a result, no cross-linguistic interaction between these two categories occurred.

With regard to the voiced series, a decision was made to subdivide the productions into three categories. Type 1 comprised tokens realised with full pre-voicing, without any breaks preceding the release of the closure (cf. Fig. 32). This type was the most common type of realisation overall. Type 2 consisted of items produced with partial pre-voicing, wherein pre-voicing was interrupted by a break suggestive of no vocal folds' vibration right before the release burst (cf. Fig. 33), though the actual duration of this type of pre-voicing was not included in the results section; rather, we focussed on the likelihood of its occurrence and whether or not it is suggestive of drift effects. Finally, the Type 3 category encompassed all /b, d, ɡ/ tokens realised without pre-voicing, with positive VOT, associated with canonical English productions of unvoiced/lenis items (cf. Fig. 34). It was assumed that Type 3 realisation found in Polish, alongside any variation with respect to pre-voicing duration overall, would be indicative of L2-induced phonetic drift taking place.

Let us first look at, perhaps, the most interesting finding, that is the way in which Polish initial voiced plosives were realised in the productions of our students. Fig. 75 illustrates the percentages of each of the three types of voicing, sorted for groups (adapted from Fig. 49 and Fig. 50) in the word reading task.

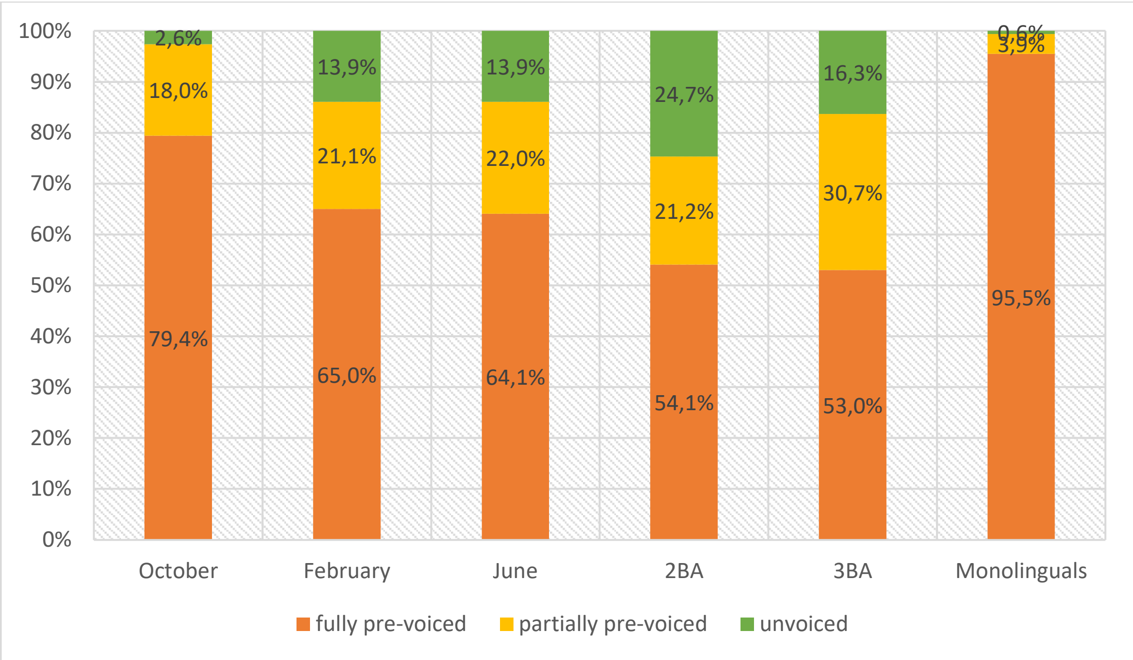


Fig. 75. Mean percentages of type of voicing realisation in Polish, sorted for groups and sessions (word reading task only).

Notice that as time progressed, even though the canonical Polish realisation (i.e. Type 1) remained the most numerous, the numbers of both Type 2 and Type 3 tokens grew. Even after the training stopped, our third-year students still displayed significant numbers of unvoiced, English-like realisations.

This particular finding regarding Type 3 tokens was especially interesting – while at the onset of the training the percentage of unvoiced tokens in the data from students was comparable with the control group's results, such tokens grew more and more common as the training progressed, peaking in Group 2's productions where they constituted almost a quarter of all voiced consonants produced by that group in the word reading task (305 items in total). The fact that the highest number of Type 3 realisations was found in Group 2 follows the general tendency for the L2-induced phonetic drift effects to be most visible in the productions of the group undergoing the second year of their pronunciation instruction, which we have seen in relation to pre-voicing duration (i.e. it was on the whole the shortest in the data from this group). Interestingly, while Group 3 – the group who have finished their phonetic training already – appeared to display pre-voicing values which were the most similar to the norms set by the quasi-monolinguals, as far as the occurrence of Type 3 tokens is concerned, they were very much present in their productions, constituting ca. 16% of all tokens. This seems to suggest that while pre-voicing as such returns back to normal after the instruction is over, some drift effects persist.

Perhaps even more remarkable are the effects we observe in the sentence reading task; the intervocalic context in which the key words are couched is very conducive to voicing in general. Fig. 76 presents the percentages of the three different type of voicing realisation in this task, sorted for groups and sessions (adapted from Fig. 53 and Fig. 54).

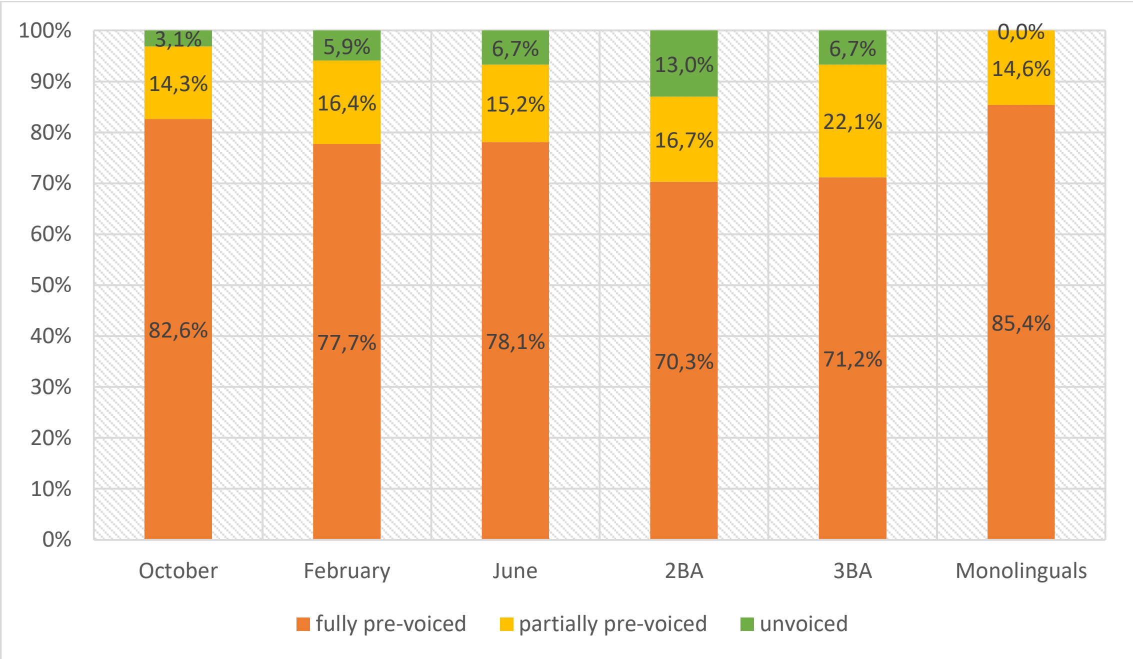


Fig. 76. Mean percentages of type of voicing realisation in Polish, sorted for groups and sessions (sentence reading task only).

And yet, while no voiced consonants are produced without pre-voicing in the data from the quasi-monolinguals, in the case of students we once again observe the progressive increase in Type 3 tokens frequency, which once more peaked in Group 2 data (with 13% of tokens in total realised without pre-voicing) and persisted even after training was over, with 6.7% of unvoiced items in the productions of Group 3. The numbers (211 tokens in total) are indeed lower than what was the case in the previous task, but here the environment is less favourable. Therefore, the fact that the general tendency is maintained appears to be evidence for phonetic drift effects in L1.

On the other hand, we postulate that the presence of absence of Type 2 tokens (i.e. those with breaks in pre-voicing) is not connected to L2 influence; while the statistical analysis revealed some differences between groups in the word reading task, no such contrasts were found in the sentence reading task. Breaks in pre-voicing were as common in the productions of monolinguals as they were in the speech of bilinguals. Therefore, we do not ascribe those to be the effect of L2 phonetic training.

Moving on to the length of pre-voicing, Type 1 productions comprised 1775 tokens in total and were, on the whole, the most common way in which Polish voiced plosives were produced. On the whole, Groups 1 (at all three testing times) and 2 – therefore the two groups of students who were undergoing phonetic instruction – showed pre-voicing values shorter than Groups 3 (i.e. students with no pronunciation training in their curriculum at the time of the recording) and 4 (i.e. the monolinguals); cf. Fig. 47. These findings suggest that L2 phonetic training may exert influence on the pre-voicing duration; once the instruction ceases, the values return to the monolingual norms. This goes in line with Chang (2012); he suggested that the re-immersion in L1 environment may reverse the effects of drift; this was also the case in Schwartz (2020). In this case the situation is slightly different as our students are immersed in L1 environment over the entire duration of the phonetic training, however in their third year they do not take active part in any type of pronunciation training. It appears that extensive usage of English in a professional setting might not be enough for it to influence the L1 acoustic features in third-year students' productions. Within the two groups affected by the L2 phonetic training, we observe progressive shortening of pre-voicing in Polish which peaks in Group 2's productions – the difference between Group 1 at T1 (therefore the students at the onset of the training) and Group 2 (the students at the offset of the training) is ca. 6 ms. While 6 ms might not seem like too big of a contrast, the difference between Group 2 and Group 4 is ca. 18 ms. Such disparities were not present in the voiceless series.

When looking at the specific places of articulation, in general the expected pattern was found, namely the labial and velar places of articulation yielded respectively the longest and shortest pre-voicing durations, with the coronals falling somewhere in between. In the case of both /b/ and /g/ the slight lengthening of pre-voicing at T2 was observable. It was significant for those two consonants in the following manner: the shortening of pre-voicing in the case of /b/ between T2 and T3, and the lengthening of pre-voicing of /g/ between T1 and T2 both presented $p<.05$ values. As far as the contrast between students and monolinguals is concerned, no such difference was found for /b/; in the case of /d/ the contrast was found between every testing times and the control group, while in the case of /g/ the difference was not found only between T2 and the controls. At every place of articulation Group 2's pre-voicing duration was the shortest, which goes in line with what we saw in the group-level results. Group 2's pre-voicing was consistently shorter than the monolingual norms for all three places of articulation; it was also consistently shorter than that of Group 3. Group 3, in turn, did not differ from the monolingual norms; the labial place of articulation showed in fact longer pre-voicing in the productions of third year students relative to the quasi-monolinguals. It can be, thus, claimed at phonetic training in L2 affects L1 pre-voiced consonants in a very similar manner, regardless of the place of articulation at which a given plosive is produced.

Turning to the sentence reading task in the subset of fully pre-voiced, i.e. Type 1 tokens, 2996 tokens in total were included in the analysis. The most striking observation to be made as far as the students' productions are concerned is that while the general trend that took place in the word reading task was also present in this task, though the differences between sessions were very small nonetheless present in every prosodic position. The pre-voicing shortening, once again, peaked in second year students' productions, with the third year students producing slightly longer pre-voicing, comparable to Group 1's (excluding the utterance-initial position, where Group 3 presented the shortest pre-voicing duration). Perhaps more interesting is the comparison of the students with the quasi monolinguals; the sheer difference between these two groups was very salient and fell between 30-40 ms. The differences between the students' groups were not significant in this task, since as has been mentioned, the pre-voicing values differed to quite a small extent. Nonetheless, each group of students (including Group 1 at all three testing times) was significantly different from the controls. Therefore, while the pattern replicating the word reading task could be seen, the most reliable difference was that of a general group comparison, which was consistent across all three prosodic positions.

Let us place the results obtained from L1 Polish speakers of L2 English within the larger body of phonetic literature discussed at length in Chapter 2. It appears that as far as the phonetic drift effects in the two-way laryngeal system of Polish are concerned, an asymmetry in the behaviour of the consonants is found with respect to VOT values. Namely, the voiceless category remains more stable across all groups and places of articulation and is affected by L2 phonetic instruction to a minimal extent. In turn, the voiced series seems to be more likely to be influenced by L2 pronunciation instruction. Not only do we find progressive pre-voicing shortening in both fully pre-voiced and partially pre-voiced plosives which is the most striking in the group that is at the offset of a two-year-long training, but we also observe progressive increase in the numbers of unvoiced, English-like realisations of Polish voiced consonants, which are canonically pre-voiced. In the phonetic drift studies which investigated both series of stops, we find similar asymmetry. These pairs include, e.g. Czech-English (Podlipský et al. 2020, Sučková 2018), Bulgarian-English (Dokovova 2015), English-Spanish (Herd et al. 2015). Thus, the present study replicates the results of previously described language pairs.

A question which may arise at this point is the specificity of the T2. The results showed a general tendency for the VOT to undergo slight lengthening – in both laryngeal categories – in this specific recording session. It might be the case that the reason for this slight change in the unexpected direction (positive VOT getting longer alongside pre-voicing lengthening) might stem from the fact that the recordings took place over the course of a winter break in-between the winter and spring semester. Therefore, it can be theorised that the break in phonetic instruction resulted in the students going for maximum dispersion between the two categories to maintain contrast between them in the paucity of pronunciation training. It would then be suggestive of the very unstable nature of phonetic drift effects, which point to how susceptible it is to any changes with respect to the environment in which the speakers find themselves. In section 7.2 we will see what happens at this testing time with their L2 productions and how it relates to the findings presented in this chapter.

6.4.2. Vowel-based laryngeal cues

Taking into consideration that VOT is only one cue associated with voicing, the second acoustic parameter of interest was f_0 (i.e. pitch) at the onset of the vowel; the pitch of the vowel has been found to undergo raising after a voiceless consonant (Ohde 1984). The measurement was the average pitch over the first 20% of the duration of the vowel. The values were given in Bark. Taken together both tasks analysed 9201 items.

The word reading task focused on two vowels: /a/ and /ɛ/ and looked at 4783 tokens from all students' groups. We differentiated between four types of onsets: voiced (i.e. Type 1 and Type 2), unvoiced (i.e. Type 3), voiceless, and nasal (i.e. our baseline as nasals are claimed to influence pitch in a neutral way). In general, the expected pattern was found, witch pitch being significantly raised after voiceless onsets in the case of both vowels. Importantly, no difference was observed between voiced and unvoiced tokens. Therefore, despite the absence of pre-voicing in the case of Type 3 items, they still performed as the voiced category.

The sentence reading corroborated the findings of the word reading task. The vowels were included in that analysis, namely /ɨ/, /a/, and /ɛ/, adding up to 4418 items in total. The types of onsets were the same as in the previous task, but the group of nasal onsets as controls was absent. Once again, the statistical significance showed that while no difference was found between pre-voiced and unvoiced onsets, pitch was raised after voiceless ones in the case of all three vowels, though with respect to raw mean values the hierarchy was less clear than in the first task.

The results go in line with other phonetic experiments from languages belonging to both aspiration and true-voice groups. Similar findings have been attested for e.g. Italian and French (true-voice languages; Kirby and Ladd 2018) and American English (an aspiration language; Hanson 2009); the fact that pitch is raised after voiceless onsets appears to be a universal process in languages with two-way laryngeal contrasts.

The second vowel-based laryngeal cue of interest was F1, once again taken from the first 20% of the vowel and calculated by means of the difference between F1 and f_0 and the value was the average obtained from that portion of the vowel, given in Bark. 8334 items were considered in total. Nasals were not included as a baseline as they could influence vowel formant trackers to some degree, due to carry-over nasality.

In the word reading task 3916 tokens in total were analysed. We compared three different types of onsets: pre-voiced (i.e. Type 1 and Type 2), unvoiced (i.e. Type 3), and voiceless. The results with F1 as the dependent variable quite reliably replicated those of f_0. While no difference was found between pre-voiced and unvoiced onsets, both of these types differed from voiceless ones in that the F1 was raised after voiceless consonants.

Therefore in this case, F1 was a credible cue for the laryngeal reality of the consonant, regardless of the presence or absence of pre-voicing.

F1 appeared to be slightly less reliable in the sentence reading task. 4418 tokens were analysed split into categories depending on the vowel quality (/a, ɛ, ɨ/) and onset type. The results were slightly more blurry; in the case of the /a/ vowel, contrast was maintained between all onset types. That is, pre-voiced and unvoiced were kept separate (though only to a small extent), and F1 after the voiceless onset was robustly higher. With respect to /ɛ/, contrast was maintained only between pre-voiced and voiceless onsets, while no contrast was found between both unvoiced and voiceless and pre-voiced and unvoiced. In the case of /ɨ/, in turn, pre-voiced and unvoiced tokens were not distinguished, but there was contrast between both pre-voiced and unvoiced and the voiceless onsets. While this is the expect pattern that seems to best replicate the results of the word-reading task, the direction was different. F1 was lowered after the voiceless onset. The fact that /ɨ/ did not follow the general tendency we observed for both pitch and F1 might stem from the fact that this is the only high vowel considered herein. F1 is low for high vowels, therefore the raising effect does not apply.

While the inclusion of pitch and F1 at vowel onset is not strictly connected with the subject matter of the present dissertation, that is, phonetic drift (i.e. the results obtained by this part are not related to cross-linguistic interaction), it suggests that the L2-induced drift effects that we *do* find (i.e. an increased in unvoiced realisations of voiced plosives in Polish under the influence of L2 phonetic training), do not affect the success with which the laryngeal contrast in Polish is maintained.

On the whole, these findings – aside from what has been mentioned in the previous paragraph – have other interesting phonological implications, which will be discussed in section 8.2.

6.4.3. Vowels

The final part of the study was concerned with possible phonetic drift effects in the acoustics of vowels (namely vowel height and vowel advancement). Two non-high unrounded vowels were considered, that is /a/ and /ɛ/. They were first plotted on F1/F2 planes for each group and session to visually inspect the possible changes over time. After that, the two acoustic parameters were considered separately for each vowel. In the two tasks 8072 productions were analysed in total: 4179 in the word reading task and 3893 in the sentence reading task. On the whole, the effects of phonetic drift in the vowel productions were quite modest and appeared to have been task-dependent. Drift has occurred to a decidedly lesser degree in the sentence reading task. Importantly, however, some changes have been observed and by and large they seem to – rather than moving towards any English targets – be representative of vowel space expansion. In the SLM, Flege (1995) predicted two scenarios (cf. 1.1.6.2). If the L1 and L2 vowels are classified as equivalent, the L1 vowel is likely to be produced with intermediate formant values (not quite L1-like, but not L2-like either). If the L2 category is different enough and a new category is formed, no interaction or dissimilatory interaction are equally both likely. In our case, the movement observed in the vowels /a/ and /ɛ/ does not necessarily stem from the need to maintain contrast between L1 and L2 categories. Rather, it appears that, since both languages exist in common phonological space, the vowel space is progressively expanded in order for the new L2 categories to fit. Once again, however, it needs to be stated that the study of vowels was rather exploratory and descriptive in nature and we do not wish to make any strong claims. Nonetheless, we have looked at a significant number of tokens, therefore some conclusions can be drawn from our data.

Let us now summarise the results. As far as /ɛ/ is concerned, in the first task the L2-induced changes targeted advancement, rather than height, to a larger extent. Nonetheless, we did observe a hierarchy forming, though the differences were slight. Group 1 at the onset of the training and quasi-monolinguals produced comparable vowels with respect to height, but Group 1 at T3, Group 2, and Group 3 presenting slightly lower realisation. /ɛ/ was lowest in the Group 2's data and it was the only group that differed from monolinguals; there was also contrast found between Group 2 and Group 1 at T1, as well as between T1 and T3 and T2 and T3. The results were more robust with regard to vowel frontness. Monolingual controls produced slightly more central /ɛ/ and differed from Group 1 at T2, Group 2, and Group 3. Additionally, significant contrast was found between T1 and T2 in Group 1's productions as well as between T1 and Group 3. In general the vowel space can be claimed to have been expanding, seeing as Group 3's realisation is decidedly more front than that of the controls.

In turn in the sentence reading task, the only group that displayed any differences in /ɛ/ height was Group 2 – their productions were lower than those of students' in Group

1 as well as the controls'. They did not diverge from the realisations of Group 3's students. When it comes to advancement, contrary to the previous task, no changes were observed.

Turning to /a/, the first task appeared to target its height to a greater extent relative to advancement. The most dramatic vowel lowering occurred over the first year of phonetic training – while there was no significant difference in vowel height between the control group and the students at the onset of pronunciation instruction, there was significant lowering between T1 and T2 as well as T2 and T3. After that the process slowed down - there was no difference between T2 and either Group 2 or 3, as well as no contrast between T3 and either Group 2 and 3. Importantly, however, both groups 2 and 3 presented significant difference relative to monolingual norms. Once again the lowering of the vowel might suggest an expansion of the vowel space. We have seen this in the previous vowel as well, but in the other dimension. No significant contrasts between the groups was attested with respect to /a/ frontness.

In the second task, there appeared to be some vowel lowering taking place in Group 1 between T1 and T3 as well As T2 and T3. Group 2's productions were once again the lowest and differed significantly from first year students at both T1 and T2. However, no group diverged from monolingual norms with respect to vowel height. Similarly, no effects of phonetic instruction on vowel advancement were observed in the sentences.

The phonological implications as well as the relation of the phonetic experiment to the research questions will be elaborated on in Chapter 8.

Chapter 7: The results – L2 performance and possible links to L1 drift.

7.1. The aim of the chapter

Chapter 6 has presented the results of a longitudinal study of Polish students of English in whose productions we looked at the effects of intensive pronunciation training in L2 on their L1 realisations, with the focus placed on the influence of L2 on their L1 laryngeal contrasts and vowel formants.

According to the principles of the SLM, all sound inventories of the languages one acquires exist in a common phonological space. Recall that over the course of data collection the participants were recorded in both of their languages. The reason for that was to be able to trace the trajectory of the acquisition of consonant and vowel contrasts and be able to check whether the success with which one does in their pronunciation class has any bearing on the extent of phonetic drift in their L1. From the perspective of equivalence classification (which extends onto the phonological theories presented in Chapters 3 and 4), investigating the contrast between the categories in L2 and L1 is equally important insofar as it allows us to say something about whether the L2 categories are indeed new or whether they merge with the existing L1 structures. The more contrast we find between the participants' Polish and English data, the more separate the categories may be claimed to be.

In this chapter we will present both group and individual results obtained from our participants in English which will be subsequently compared with their L1 performance. For pragmatic reasons, we will limit ourselves to English word lists, as these were easier to control for any confounding variables. Looking at English sentences will remain a task for the future.

The chapter will present the results on two acoustic parameters, namely VOT (positive and negative, in milliseconds) and vowel formants (F1 and F2 over the middle 20% of the vowel, Bark normalised), and will be subdivided into group and individual results[87]. We will not look at f_0 and F1 at the onset of the vowel as these two acoustic cues and the effects that fortisness exerts on them have been shown to be similar both in aspiration and voicing languages. Hence, not much variation is expected for which reason they will be excluded. The chapter will end with a discussion of the findings.

7.2. VOT: L2 data juxtaposed with L1 results

The first acoustic parameter to be discussed is VOT. In Chapter 6 we have seen that there is little L2-induced phonetic drift effects in the voiceless series of stops; on the contrary, the voiced series appears to be more susceptible to equivalence classification. We have seen that there is a substantial amount of pre-voicing in the L2 data (possibly due to L1 Polish), as well as numerous unvoiced realisation of the voiced category in both languages, which suggests bi-directional effects in the L2 English acquisition.

It has been suggested, then, that plain voiceless and aspirated stops constitute two separate categories (hence the interaction between the two is less prevalent) while both pre-voiced and unvoiced stops belong to the same category. As a corollary, when comparing L1 and L2 stops' VOTs, we would expect the magnitude of contrast between the voiceless categories to be more salient, relative to the voiced category, where the boundary should be blurrier. The following sections will present the results of this comparison – group results and individual differences will be discussed.

7.2.1. Group results

This section provides an overview of the comparison between L1 and L2 stops in the productions of the three groups of students: first year (i.e. Group 1; at three testing times), second year (Group 2), and third year (Group 3).

Since in the latter part of this section we will contrast the VOT durational differences between groups and the two languages, let us begin by assessing the differences in

[87] Note that the individual differences discussed in this chapter will be purely descriptive in nature. They will not offer statistical analyses of the obtained results. This will also remain a task for the future. It is hoped, however, that the reader will still find those findings interesting.

the realisation of the voiced series, as this is perhaps one of the most interesting observations to be made at this point. Fig. 77 presents the percentages of Type 1 (i.e. fully pre-voiced), Type 2 (i.e. partially pre-voiced), and Type 3 (i.e. unvoiced) realisations in English, sorted for session.

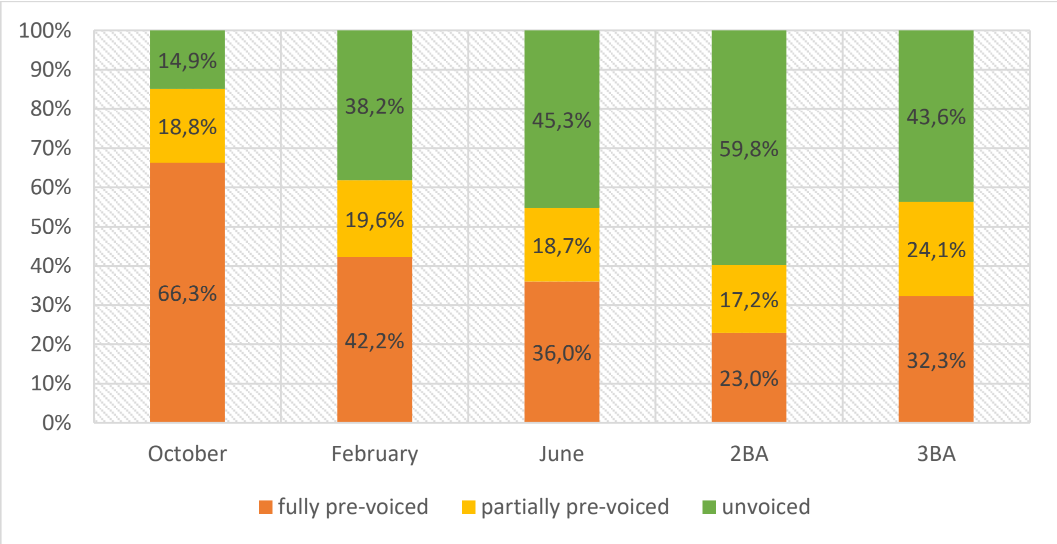


Fig. 77. Percentages of the three types of voiced category realisation in English, sorted for session.

Let us have a look at a similar figure – Fig. 78 – but one that focuses on types of realisation in Polish[88].

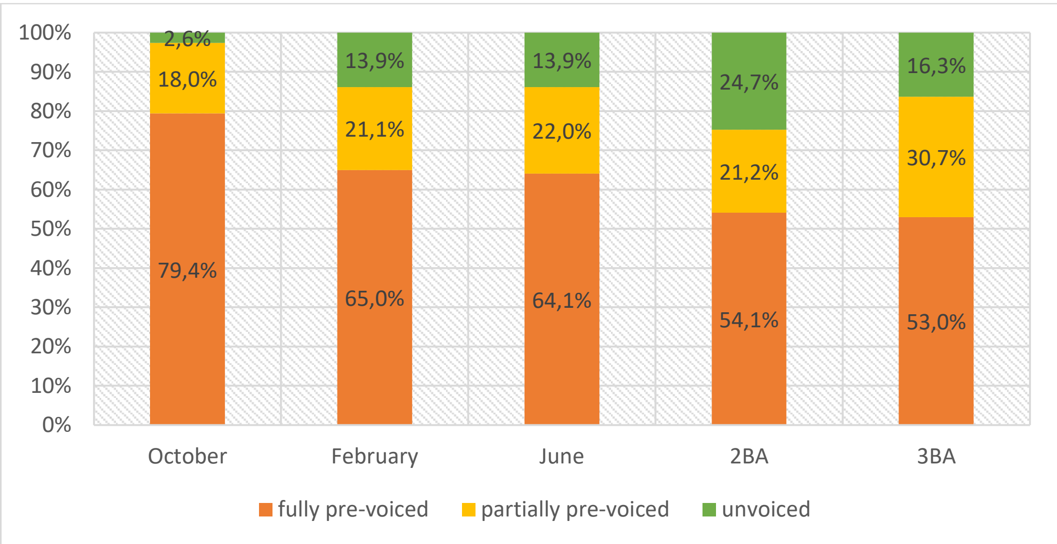


Fig. 78. Percentages of the three types of voiced category realisation in Polish, sorted for session.

[88] This figure has already been shown in section 6.4.1. It is included here to enable a comparison of the two languages.

The data in Fig. 77 and Fig. 78 are interesting insofar as we can observe that the unvoiced, English-like realisations are more common in L2 productions. This is quite intuitive in and of itself – L1 drift is triggered by L2 acquisition. It can be claimed that the more of correct English realisations we see in L2, the more phonetic drift is likely to be observed in L1. What might be surprising is that already at T1, our students displayed over 14% of correct, unvoiced realisations in L2. Alongside the fact that (as will be shown) aspiration had already been acquired at T1 as well, it might be claimed that we have missed the beginning of the interaction between the two languages. Nonetheless, as can be seen, phonetic training appears to boost those effects, as they are much stronger after two years of explicit instruction (i.e. in Group 2, our second-year students) relative to what was the case at T1. Without it, phonetic drift might have taken place to a much lesser extent. We will return to these observations shortly, when we look at individual performance of our students in section 7.2.2.

Now let us compare the duration of VOT (positive in the case of voiceless and negative in the case of voiced[89]) In section 6.2 we saw that, in general, the Polish VOT values of voiceless plosives remain relatively stable across all groups and testing times and we observe very little effects of phonetic drift. An interesting question thus arising was whether or not our participants actually acquire aspiration, typical of English. In order to check this, the L1 data were supplemented by the results from the L2 word reading task. At the same time, we wanted to see whether our students would produce English stops in a Polish-like manner, or whether pre-voicing would be longer or shorter, as well as whether it would be influenced by phonetic instruction which then carries over onto their Polish productions (Chapter 6 presented data suggesting that indeed, Polish pre-voicing changes under the influence of English). Additionally, the lack of contrast with respect to pre-voicing could nicely show the bi-directionality of equivalence classification. Essentially, we have already seen that L2 phonetic instruction results in an increase of unvoiced, English-like productions in Polish. Long, Polish-like negative VOT in English could be a sign of L1 interference. Thus, the cross-linguistic interaction of the category deemed as phonologically identical would be clearly visible.

[89] Once again, Type 1 only.

Fig. 79 compares the voiceless category in Polish and English (top figure; 2494 English words and 2566 Polish words were included in the analysis) with the voiced category (bottom; 1431 tokens from Polish and 973 tokens from English).

Fig. 79. Mean VOT values of /p, t, k/ (top) and /b, d, g/ (bottom) sorted for Group and Language.

Starting with the voiceless series, Fig. 79 shows that already at T1 (that is, in October) our participants showed a difference between /p, t, k/ in Polish and in English. Their L2 mean at this testing time was 58.59 ms (SD=24.54), therefore the magnitude of contrast reached ca. 16 ms. At T2 the mean L2 value got slightly longer – 61.77 ms (SD=25.41), with the difference between L1 and L2 being ca. 18 ms. At T3, the average VOT was 60.19 ms (SD=24.22) and the contrast between L1 and L2 again reached ca. 18 ms. The differences between Polish and English are even more visible in Groups 2 and 3, our second- and third year students respectively. Group 2's L2 average was 75.69 ms (SD=24.42), which yielded ca. 32 ms of a contrast between L1 and L2. Group 3 was similar; the mean VOT in English was 73.82 (SD=26.10), and the difference between the two languages totalled ca. 28 ms.

Turning to the voiced series, we can notice that the differences between Polish and English are not substantial. At T1, the very onset of the pronunciation training, pre-voicing in English reached -88.48 ms (SD=25.65), with the contrast between L1 and L2 being ca. 8 ms. At T2 the difference was even smaller – only ca. 1 ms, with the duration in English being -84.21 ms (SD=24.79). At T3, after eight months of instruction, the contrast between Polish and English oscillated within 5 ms and English yielded -83.47 ms (SD=26.62). Group 2 yielded the average of -79.41 ms (SD=25.12) and the contrast reached ca. 5 ms. In Group 3's productions, English pre-voicing was shorter than what we observed in Polish by ca. 3 ms (contrarily to the previous sessions and groups), with the mean being -88.59 ms (SD=26.23). The group data revealed that pre-voicing in both languages remained comparable.

A GLMM model with VOT as the dependent variable, Session*Language*Voicing_Type as the main interaction of interest, and Speaker and Item as random factors revealed that the contrast in the voiceless series between the two languages was statistically significant at every session (*p*=.000). Thus, it appears that aspiration was such salient a cue that, to a certain extent, students who are proficient in English contrast it with plain voicelessness in their native language before the explicit phonetic instruction even begins. The contrast is maintained throughout the training, with the largest magnitude visible after the second year, and it is still present even after the pronunciation instruction had stopped. With respect to the voiced series, although the differences were by and large small, the analysis revealed that the difference at T1, where English pre-voicing turned out to be longer, was significant (*t*=3.245; *p*=.001). The second significant contrast was the longer pre-voicing in Polish in Group 3's data, but only just (*t*=1.969; *p*=.049). The contrasts at the remaining sessions did not reach statistical significance[90].

A question that may arise at this point is whether or not the contrast is maintained across all places of articulation. Recall that, for example in Osborne (2016; cf. section

[90] See Appendix 7 for the full coefficient table alongside pairwise comparisons.

2.3.1) plain voiceless /p/ in Brazilian Portuguese and aspirated /p/ in English were dissimilated, whereas /k/ in both languages underwent some assimilation, possibly due to a relatively long VOT of voiceless velars in general which blurs the boundary between short-lag and long-lag VOT. Fig. 80 thus presents the mean VOT values sorted for Group/Session, language, and place of articulation for both series of stops.

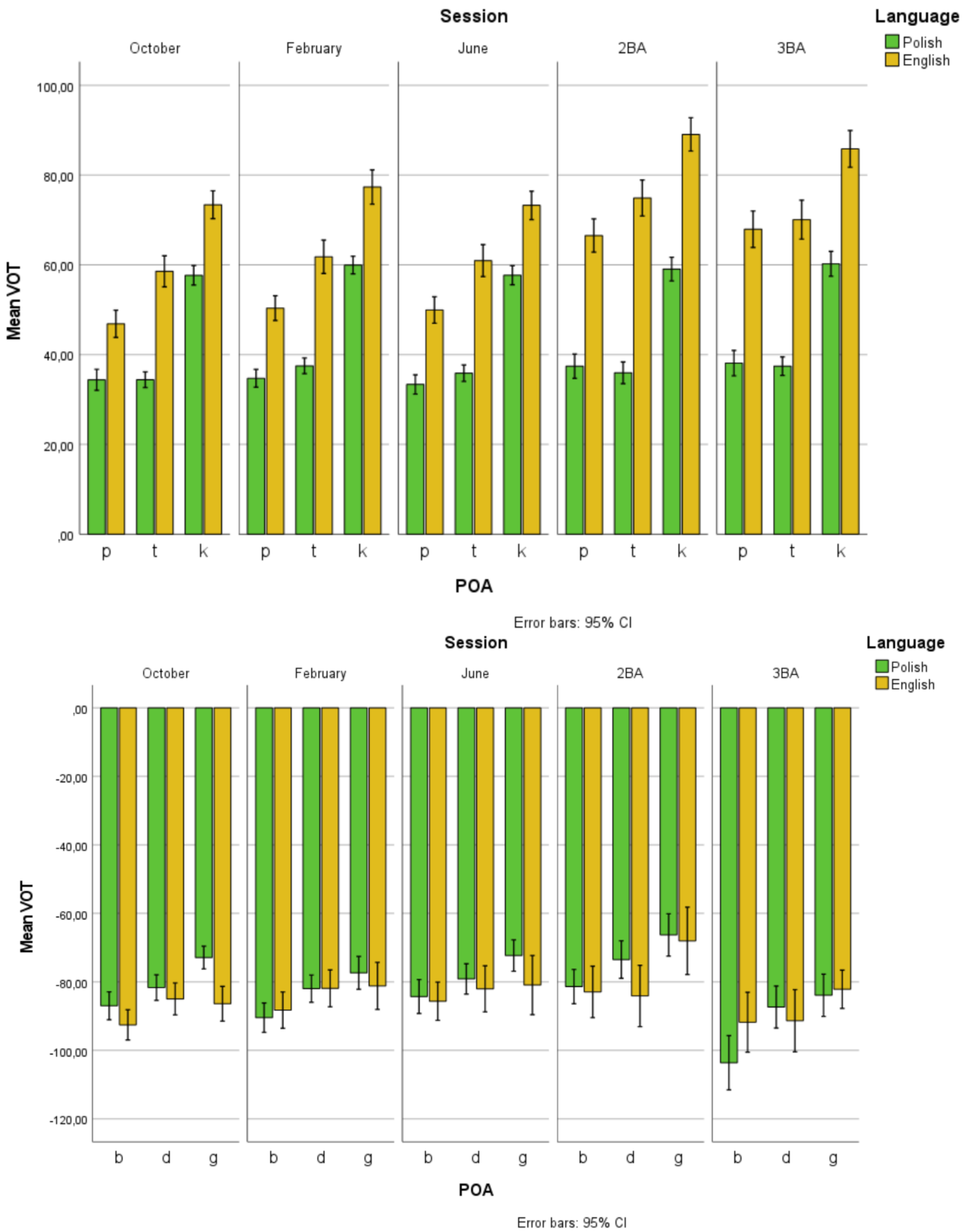


Fig. 80. Mean VOT of /p, t, k/ in Polish and English, sorted for Group/Session and place of articulation.

On the basis of Fig. 80, it appears that, indeed, contrast between plain voiceless stops in Polish (green) and aspirated stops in English (yellow) is kept across all places of articulation. Notice also that while the VOT of /p/ and /t/ in Polish are quite similar, the mean values in English show a more robust hierarchy across the places of articulation, going from shortest (labial) to longest (velar). As far as /b, d, g/ goes, we can see that, on the whole, the general expected pattern is also followed – labials yield the longest and velar the shortest pre-voicing (cf. Maddieson 1996). The differences between Polish and English are minute, in contrast to the voiceless series.

Table 64 presents the specific VOT values for all stops, including the Polish data (repeated here for ease of interpretation).

Table 64. Mean VOT values of /p, t, k/-initial words in Polish and English alongside Standard Deviations, sorted for Group/Session, place of articulation, and language.

POA	Session	Language	N	Mean	SD
/p/	T1	Polish	178	34.39	15.71
		English	207	46.87	22.07
	T2	Polish	174	24.72	13.23
		English	215	50.36	20.35
	T3	Polish	177	33.36	14.43
		English	215	49.95	21.94
	Group 2	Polish	132	37.42	15.71
		English	161	66.53	23.80
	Group 3	Polish	129	38.11	16.27
		English	164	67.94	26.38
/t/	T1	Polish	219	34.41	13.01
		English	184	58.56	23.75
	T2	Polish	220	37.50	13.12
		English	175	61.80	24.96
	T3	Polish	217	35.87	13.64
		English	181	60.95	24.34
	Group 2	Polish	163	35.95	15.61
		English	132	74.90	23.26
	Group 3	Polish	161	37.43	13.30
		English	132	70.08	25.24
/k/	T1	Polish	179	57.66	14.72
		English	164	73.04	20.08
	T2	Polish	178	59.95	13.24
		English	157	77.35	24.23
	T3	Polish	173	57.69	14.19
		English	158	73.26	20.17
	Group 2	Polish	131	59.06	15.29

		English	118	89.07	20.32
	Group 3	Polish	135	60.23	16.31
		English	121	85.84	22.66

In turn, Table 65 presents more details on both languages (the data for Polish is repeated here for ease of interpretation) of the voiced series of stops, once again including the number of tokens, mean durations, and Standard Deviations.

Table 65. Mean VOT values of /b, d, g/-initial words in Polish and English alongside Standard Deviations, sorted for Group/Session, place of articulation, and language (Type 1).

POA	Session	Language	N	Mean	SD
/b/	T1	Polish	128	-86.96	23.30
		English	144	-92.54	26.71
	T2	Polish	123	-90.44	24.00
		English	87	-88.23	24.91
	T3	Polish	116	-84.26	26.97
		English	90	-85.63	26.67
	Group 2	Polish	78	-81.36	22.20
		English	52	-83.91	27.00
	Group 3	Polish	62	-103.60	31.15
		English	50	-91.76	30.80
/d/	T1	Polish	130	-81/65	21.40
		English	114	-84.96	25.07
	T2	Polish	103	-81.96	20.31
		English	72	-81.87	22.90
	T3	Polish	103	-79.12	22.67
		English	55	-82.02	24.89
	Group 2	Polish	65	-73.47	22.12
		English	17	-84.09	17.40
	Group 3	Polish	71	-87.33	25.99
		English	36	-91.32	26.82
/g/	T1	Polish	135	-72.89	19.47
		English	88	-86.37	23.93
	T2	Polish	97	-77.36	23.84
		English	60	-81.17	26.59
	T3	Polish	99	-72.30	22.94
		English	45	-80.17	28.76
	Group 2	Polish	59	-66.28	23.70
		English	23	-68.02	22.73
	Group 3	Polish	62	-83.89	24.25
		English	40	-82.13	17.46

A GLMM model with VOT as the dependent variable and an extended interaction of POA*Language*Session*Voicing_Type as the main predictor shows that the differences between Polish and English are maintained across the board. Table 66 presents the detailed pairwise contrasts between Polish and English (the baseline is Polish; the significant contrasts are bolded), sorted for session and place of articulation[91].

Table 66. The results of the GLMM model with VOT of /p, t, k/ as the dependent variable and POA*Language*Session*Voicing_Type as the main interaction (pairwise comparisons; baseline: Polish).

Session	POA	Contrast estimate	Std. Error	t	Significance
October	**/p/**	**-13.461**	**1.723**	**-7.815**	**<.001**
	/t/	**-25.606**	**1.801**	**-14.216**	**.000**
	/k/	**-22.531**	**2.347**	**-9.600**	**.000**
February	**/p/**	**-16.586**	**1.721**	**-9.639**	**.000**
	/t/	**-26.057**	**1.827**	**-14.264**	**.000**
	/k/	**-24.401**	**2.369**	**-10.301**	**.000**
June	**/p/**	**-17.348**	**1.713**	**-10.128**	**.000**
	/t/	**-26.521**	**1.812**	**-14.633**	**.000**
	/k/	**-22.305**	**2.369**	**-9.416**	**.000**
2BA	**/p/**	**-29.823**	**1.972**	**-15.124**	**.000**
	/t/	**-40.001**	**2.065**	**-19.366**	**.000**
	/k/	**-36.354**	**2.590**	**-14.037**	**.000**
3BA	**/p/**	**-30.422**	**1.977**	**-15.390**	**.000**
	/t/	**-34.512**	**2.072**	**-16.658**	**.000**
	/k/	**-32.319**	**2.570**	**-12.574**	**.000**
October	**/b/**	**-5.604**	**2.781**	**-2.015**	**.044**
	/d/	-4.159	2.970	1.400	.162
	/g/	**-13.618**	**3.315**	**-4.108**	**<.001**
February	/b/	3.130	3.196	.979	.328
	/d/	1.838	3.511	.524	.601
	/g/	-2.783	3.884	-.726	.474
June	/b/	.561	3.199	.175	.861
	/d/	-2.357	3.811	-.619	.536
	/g/	-6.542	4.245	-1.541	.123
2BA	/b/	.571	4.122	.138	.890
	/d/	-3.251	6.261	-.519	.604
	/g/	2.375	5.691	.417	.677
3BA	**/b/**	**14.335**	**4.284**	**3.346**	**.001**
	/d/	-3.416	4.719	-.724	.469
	/g/	-.609	4.750	-.128	.898

[91] In this case the full coefficient table for /p, t, k/ indicated no significant differences; however contrasts between languages was found in pairwise comparisons and is reported in Table 66. The coefficient table for the voiced series can be found in Appendix 7.

As can be seen, not only do our participants keep the two voiceless categories separate throughout the entire study, the differences are also quite salient – they range between 16 ms up to 40 ms. On the other hand, in the voiced series contrast between Polish and English was observed only at T1 for labials (*p*=.044) and velars (*p*<.001) and in Group 2 for labials only (*p*=.001).

7.2.2. Individual differences

Sections 1.2.2 and 1.2.3 mentioned that as far as L2 productions are concerned, there is a lot of variability across the speakers as well as *within* speakers. This section focuses on the former. Essentially, we were interested to see whether the group results mirror individual speakers' performance. Going back to the principle of equivalence classification and what we have already seen in Chapter 6 and in the previous section of the present Chapter 7, we may expect to see a clearer boundary between Polish and English voiceless stops as it appears that two distinct categories have been formed for plain voiceless and voiceless aspirated plosives, as decidedly less drift overall has been observed in /p, t, k/. With regards to the voiced series, we may expect more variability in pre-voicing duration.

Let us begin with T1, which was the first recording session in which the students took part at the very start of their phonetic training. The following drop-line graphs illustrate the difference between Polish and English mean VOT of /p, t, k/- and /b, d, g/-initial words in order to see to what extent they are kept separate. Fig. 81 illustrates the magnitude of contrast in the voiceless series, while Fig. 82 displays the difference in the voiced series.

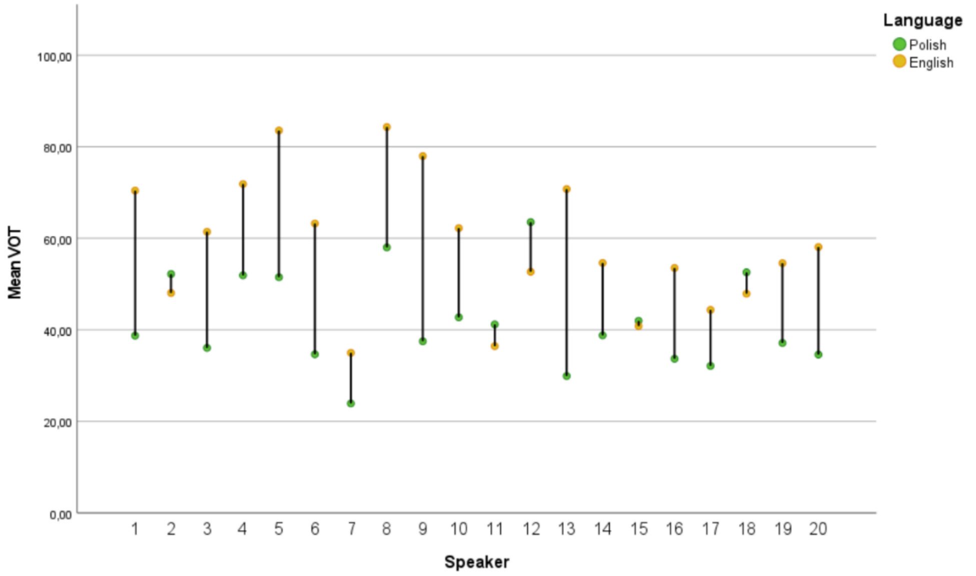


Fig. 81. Individual differences in contrast magnitude between mean VOT of /p, t, k/ in Polish (green) and English (yellow) – Group 1 at T1.

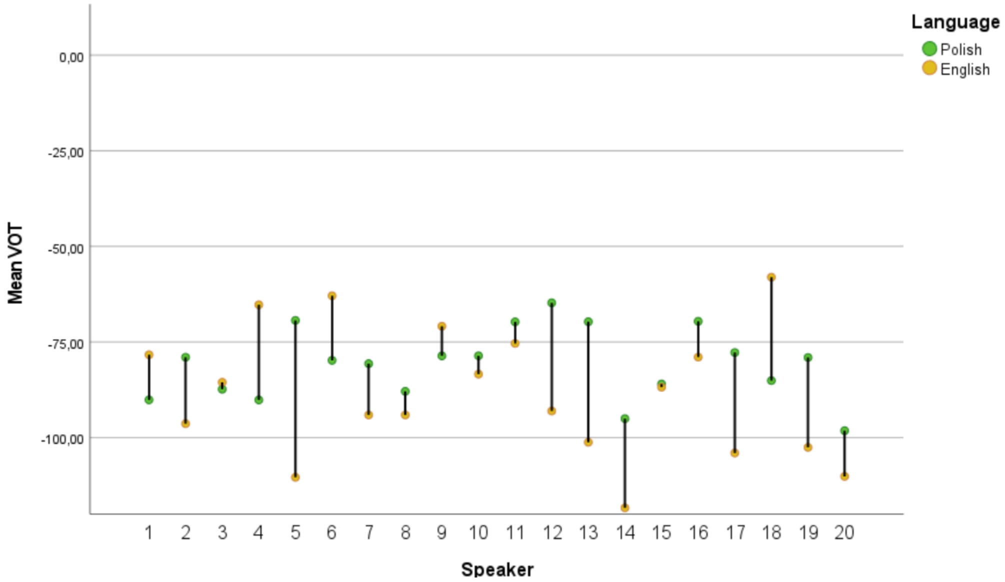


Fig. 82. Individual differences in contrast magnitude between mean VOT of Type 1 /b, d, g/ in Polish (green) and English (yellow) – Group 1 at T1.

We can observe that, indeed, a significant number of students appears to distinguish between plain voiceless /p, t, k/ in Polish and their aspirated counterparts in English (in particular S1, S3, S4, S5, S6, S8, S9, S10, S13, S14, S16, S17, S19, and S20). For all those students Polish mean VOT is saliently lower relative to their English mean VOT value. In the case of S7, while they seem to keep their categories separate, the overall VOTs (for both languages) are shorter than what we see in the rest of the cases. There is also a group of students for whom Polish VOT averages are in fact longer than their L2 means (S2, S11, S12, S15, and S18). S15 does not seem to distinguish between the two categories of stops at all – their mean VOT values overlap.

In turn, as far as the voiced series is concerned, there is a lot of variation present. First of all, for the majority of speakers English yields longer pre-voicing than Polish (S2, S5, S7, S8, S10, S11, S12, S13, S14, S16, S17, S19, S20). Recall that already at T1, the first year students displayed English-like aspiration in their productions. Such long pre-voicing in their voiced category might suggest that they attempt to make the voiced and voiceless categories maximally different. For a few speakers (S1, S4, S6, S9, S18) it is Polish that yields longer pre-voicing. For two participants the difference is virtually non-existent (S3 and S15). Overall, the difference in pre-voicing length is salient for only a couple of speakers (S4, S5, S12, S13, S15, S18); for the rest the magnitude is not that robust. Notice also that while 20-30 ms of a difference in VOT in the voiceless category is perceptually much more audible than differences in pre-voicing duration.

Fig. 83 shows the mean VOT values in Polish and English of the same group of students at T2 (that is, in February), after ca. three months of phonetic instruction in the voiceless series while Fig. 84 is concerned with the differences in pre-voicing across the two languages.

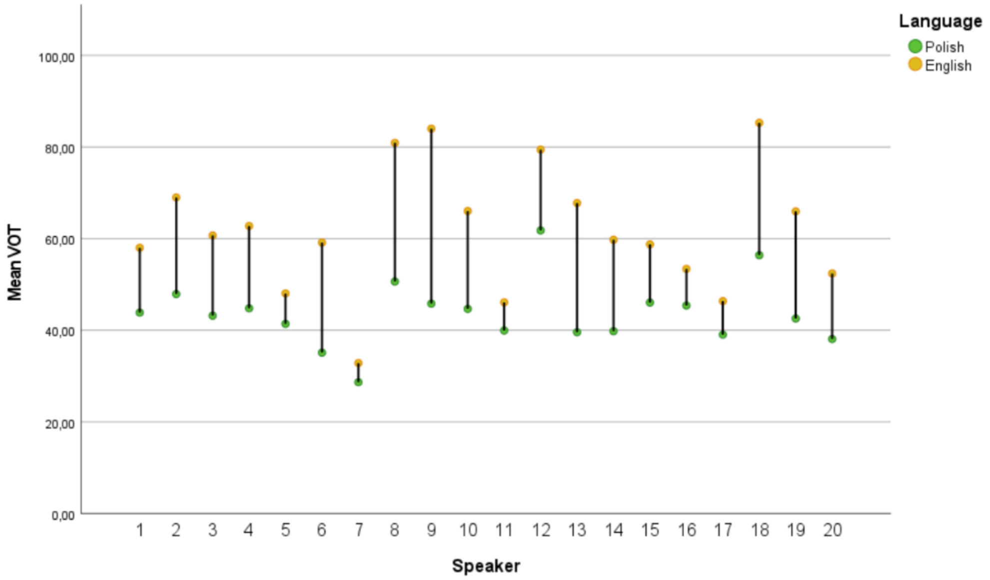


Fig. 83. Individual differences in contrast magnitude between mean VOT of /p, t, k/ in Polish (green) and English (yellow) – Group 1 at T2.

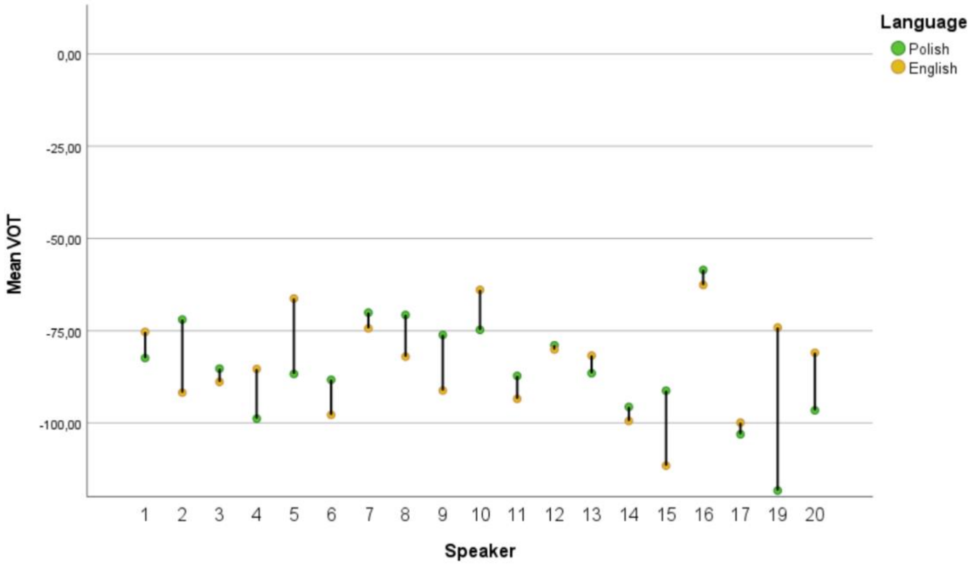


Fig. 84. Individual differences in contrast magnitude between mean VOT of Type 1 /b, d, g/ in Polish (green) and English (yellow) – Group 1 at T2.

The most interesting observation one can make on the basis of Fig. 83 is that after three months of phonetic instruction, the mean English VOT of /p, t, k/ in the productions of all those speakers for whom the average was lower than their Polish VOT had undergone lengthening. In other words, over the course of the first semester the speakers' productions changed to an extent that L2 VOTs are on the whole longer than what we see in their L1. They cannot be claimed to have acquired aspiration just yet, however all 20 participants do appear to strive to keep their categories separate. S7 is again the speaker whose mean VOTs are on average the shortest, in both languages. Interestingly, the magnitude of contrast between L1 and L2 in the productions of S5, S16, S17, and S20 appears to be smaller than what we saw at T1.

In the case of the voiced series, it appears that the difference in pre-voicing duration got less observable – overall, aside for S19 – the contrast between pre-voicing in Polish and English got smaller. Notably, it seems that in comparison to T1, pre-voicing on the whole got shorter – only in the case of three speakers (S15, S17, and S19) it surpassed -100 ms, for either language. For many speakers, however, English still yielded in general longer values (S2, S6, S7, S8, S9, S11, S15).

The final recording session, at T3, took place after eight months of L2 pronunciation training (that is, in June). Fig. 85 displays the magnitude of contrast between mean VOTs in Polish and English for Group 1 at the end of their first academic year of phonetic training of the voiceless series, while Fig. 86 displays the magnitude of contrast with respect to pre-voicing.

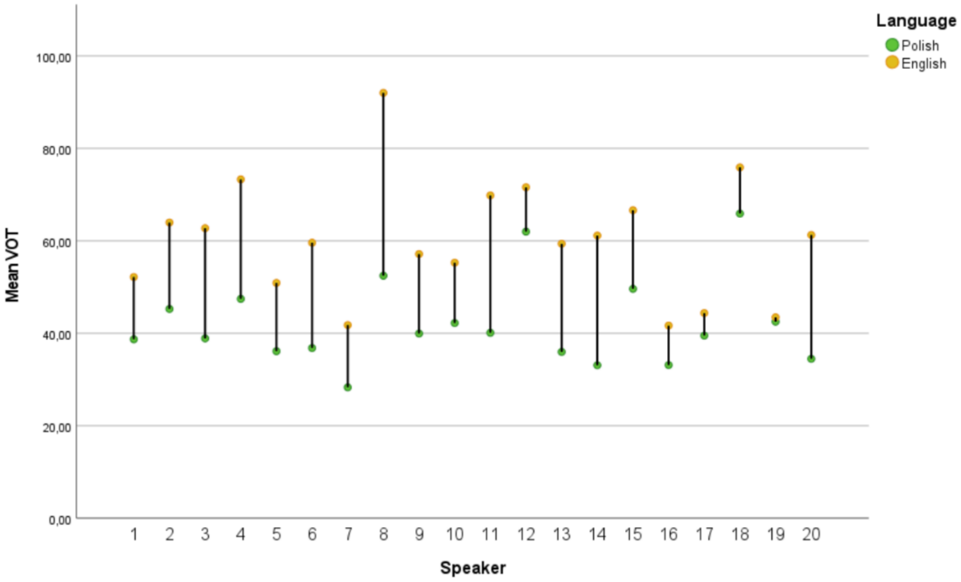


Fig. 85. Individual differences in contrast magnitude between mean VOT of /p, t, k/ in Polish (green) and English (yellow) – Group 1 at T3.

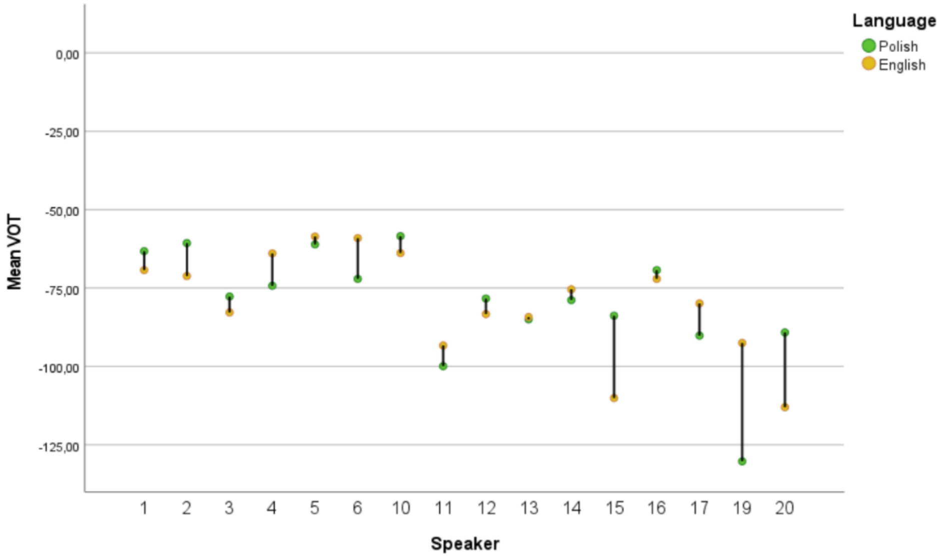


Fig. 86. Individual differences in contrast magnitude between mean VOT of Type 1 /b, d, g/ in Polish (green) and English (yellow) – Group 1 at T3.

Again, for all speakers mean Polish VOT of /p, t, k/ is shorter than English, meaning that they attempt to keep their L1 voiceless stops and L2 aspirated stops separate. An exception to this pattern is S19 for whom these two categories are indistinguishable – a curious case, since they showed contrast at both T1 and T2. We can also see that S7 had made a lot of progress and their mean English VOT had got overall much longer than what was the case at T1. S8 appears to be the person who was the most successful one at approaching native English norms in their L2 speech. The rest of the speakers (aside from S19) can be claimed to adhere to the observation made by Flege (1995), who mentions that most commonly L2 speakers produce intermediate VOT values which fall somewhere between L1 and L2 monolingual norms. While the students may not have reached English monolingual values, the important thing is that aspiration was strong enough a cue for them to create a new category for the English /p, t, k/.

As far as the voiced series goes, at T3, interestingly, there have been speakers who no longer produced Type 1 tokens in either of their language (that is, either in their Polish or in their English data there were no Type 1 productions). For this reason S7, S8, S9, and S18 are missing from Fig. 86. We can also see that for many of the participants pre-voicing in both languages underwent shortening – for S1, S2, S4, S5, S6, S7, and S16 it does not surpass -75 ms, while for S3, S11, S12, S13, S14, and S17 it is not longer than -100 ms. Therefore pre-voicing shortening might be claimed to have occurred in the productions of the majority of our participants over the course of their first year of studies.

Recall that Group 2, the second-year students, whose participants had been going through two years of pronunciation training at the time of the recordings displayed overall the largest effects of phonetic drift. Fig. 87 and Fig. 88 shows the individual differences in contrast magnitude of the voiceless and voiced series, respectively, in Polish and English for this group.

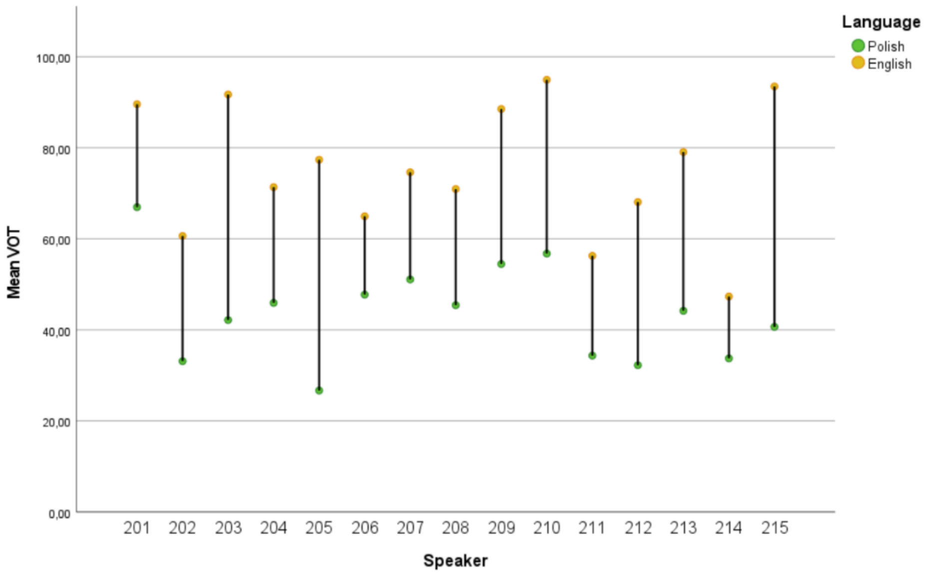


Fig. 87. Individual differences in contrast magnitude between mean VOT of /p, t, k/ in Polish (green) and English (yellow) – Group 2.

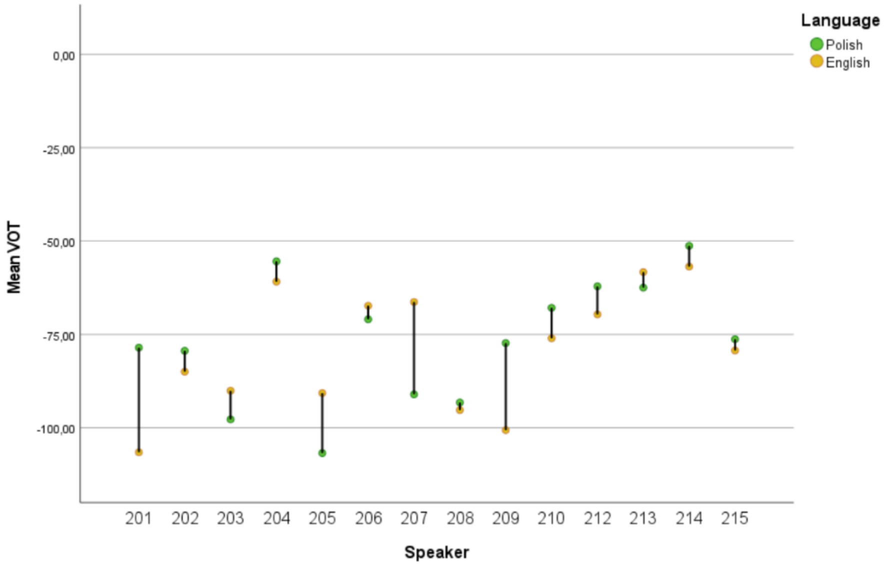


Fig. 88. Individual differences in contrast magnitude between mean VOT of Type 1 /b, d, g/ in Polish (green) and English (yellow) – Group 2.

When it comes to Fig. 87, notice that in comparison with Group 1, the magnitude of contrast between mean VOTs in Polish and English in Group 2's productions is saliently larger. While their Polish scores are comparable with Group 1, aside from S214, Group 2's speakers produce longer English VOT, especially S203, S209, S210, and S215. Again, Group 2, similarly to Group 1, appear to keep their /p, t, k/ realisations in both of their languages distinct, though it appears that an additional year of phonetic training makes this distinction even more robust.

Turning to the voiced series, as can be seen, S211 is excluded – there were no Type 1 tokens in their Polish. As for the remaining fourteen participants, we can see that for the majority both Polish and English pre-voicing values did not exceed -75 ms (S204, S206, S210, S212, S213, S214), and only two speakers displayed pre-voicing of over -100 ms (S201 and S205). There was again a lot of variability – for some speakers it was English that yielded longer averages while for some it was Polish. Nonetheless, aside from S201, S207, and S209, the magnitude of contrast was not great.

Finally Fig. 89 and Fig. 90 deal with Group 3 – the third year students who had gone through two years of L2 phonetic instruction, but the training had ended one year prior to the recording session.

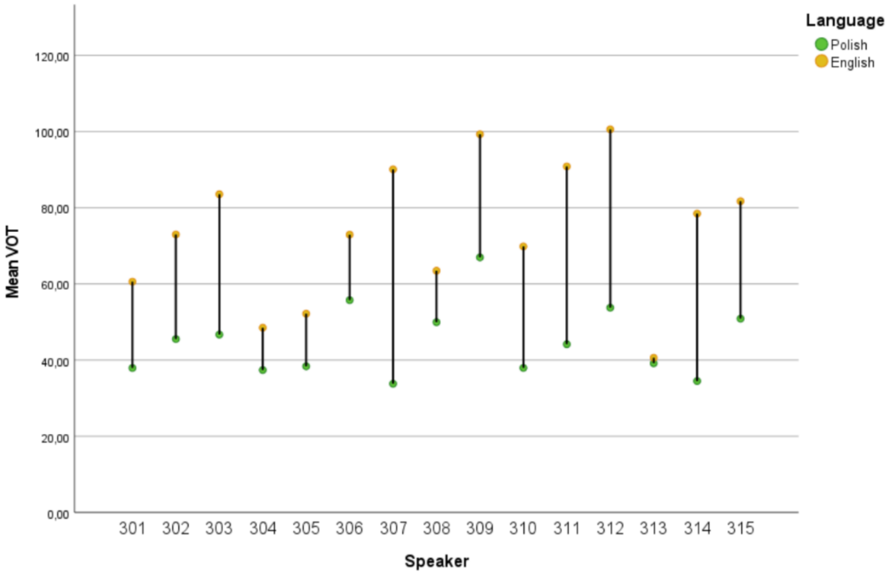


Fig. 89. Individual differences in contrast magnitude between mean VOT of /p, t, k/ in Polish (green) and English (yellow) – Group 3.

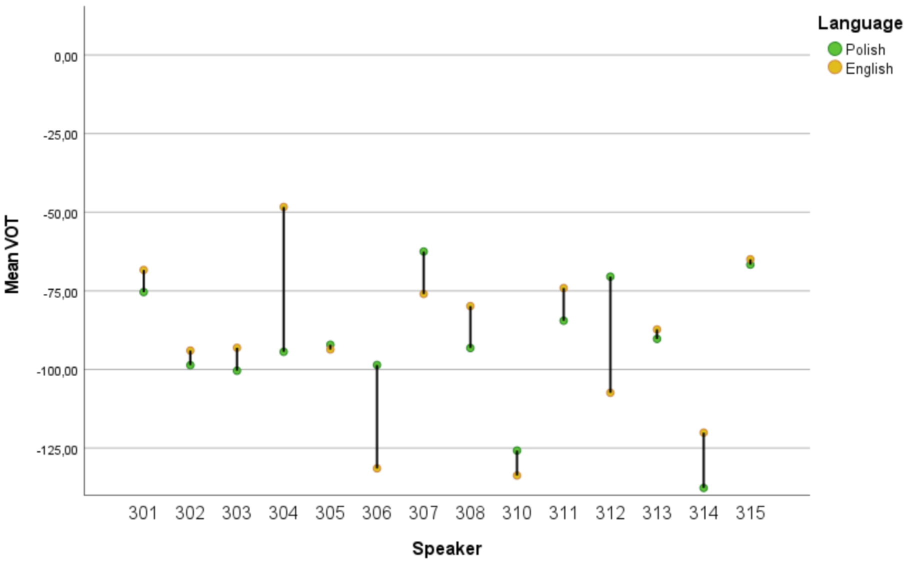


Fig. 90. Individual differences in contrast magnitude between mean VOT of Type 1 /b, d, g/ in Polish (green) and English (yellow) – Group 3.

In general, Group 3 seem to maintain the contrast between the voiceless series in Polish and English (in particular S301, S302 S303, S307, S309, S310, S311, S312, S314, and S315). The difference is quite small in the productions of S304, S305, S306, and S308, and is virtually absent in S313. Therefore, it can be claimed that even after the instruction had ended, the majority of the L2 English speakers taking part in the study are able to maintain aspiration in their productions.

As far as the voiced series goes, one speaker (S309) had to be excluded as they did not have Type 1 realisations in their Polish productions. In the case of four speakers, the magnitude of contrast is rather striking. S304's English pre-voicing is shorter than -50 ms in general, while their Polish pre-voicing on average reaches nearly -90 ms. S314, whose contrast between L1 and L2 is less salient, also is characterised by a relatively long pre-voicing, with Polish values surpassing -125 ms. In the case of S306 and S312 the difference between English and Polish is robust, but it is English that yields longer pre-voicing, with both speakers' mean durations being longer than -100 ms. For the rest of the participants in this group the differences are rather minimal, and even negligible in the case of S305, S313, and S315.

All in all, indeed, in light of the SPE model, the L1 Polish speakers appear to have managed to create a new category for the voiceless aspirates. As there was little perceptual linkage of the two in general, we did not find effects of L2-induced phonetic drift in the data discussed in Chapter 6.

Finally, let us look at the individual differences with respect to the sheer numbers of Type 3 realisations in the speech of our participants. This was done as we were looking for parallel behaviour between the two languages. In Fig. 77 and Fig. 78 in the previous section it has been shown that overall these type of productions was more common in English – when we looked at group results, it appeared that the more common Type 3 tokens are in English, the more drift we observe in Polish.

Here we were interested to see whether the percentage of Type 3 realisations in English of the individual speakers had any bearing on how often they occurred in their L1 speech data. Before we discuss this, let us have a look at the general patterns across the groups and sessions. In order to see the individual differences in our data, we counted the number of Type 3 productions in the data taken from each speaker in either language and calculated to what percentage of all of that speaker's voiced-initial items this type of productions contributed.

In Group 1 at T1 we found 14.9% of Type 3 tokens in English, relative to 2.6% in Polish. Fig. 91 shows the percentages of this type of tokens in the productions of those twenty speakers at the onset of their phonetic instruction in English.

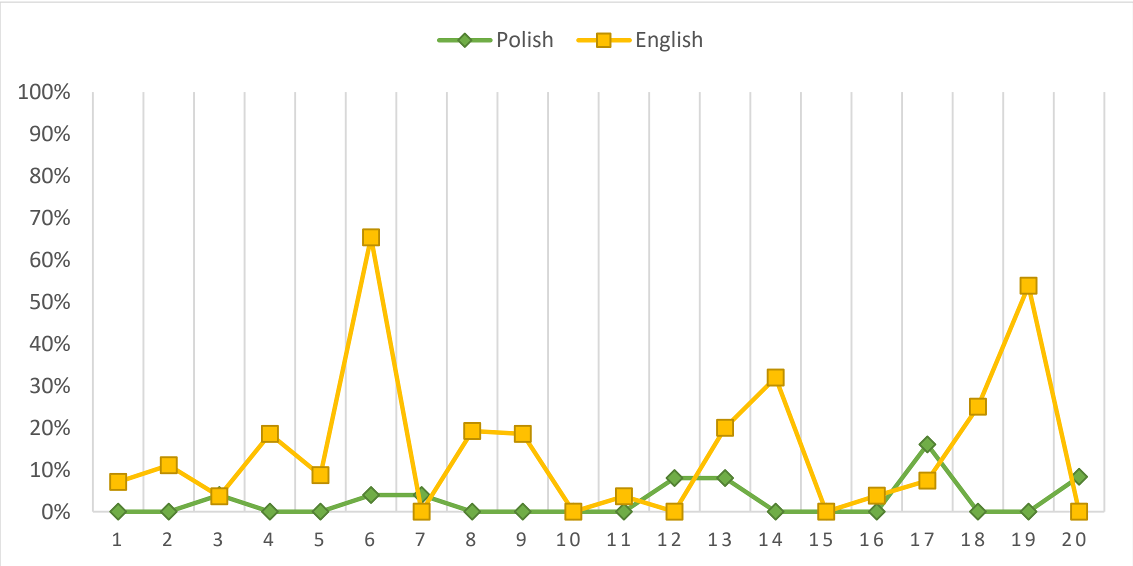


Fig. 91. Individual differences in the percentages of unvoiced realisation of /b, d, g/ in Polish (green) and English (yellow) – Group 1 at T1.

The first observation to be made is that, indeed, there is a lot of variability across our participants. In general, Type 3 tokens are more numerous in English. For S6 they contribute to nearly 70% of all items; they are equally common also in S19's data (ca. 55%). For S4, S8, S9, S13, S14, S18 these tokens constitute ca. 20%-30% of all voiced-initial words. There are also speakers (S7, S10, S12, S15, S20) for whom Type 3 items are absent. Interestingly, however, although there are no unvoiced initial plosives in L2 productions in the speech of S7, S12, and S20, we find such cases in their Polish data (between 5%-10% of all words). Nonetheless, at this recording session, such realisation is relatively uncommon in Polish (0% for S1, S2, S4, S5, S8, S9, S10, S11, S14, S15, S16, S18, and S19, that is, the majority of the speakers). This was visible in the group results as well.

A substantial change occurs between T1 and T2, that is three months into the pronunciation training – the Type 3 productions skyrocket in both languages, to 38.2% in English and 13.9% in Polish. Fig. 92 illustrates these numbers for individual speakers.

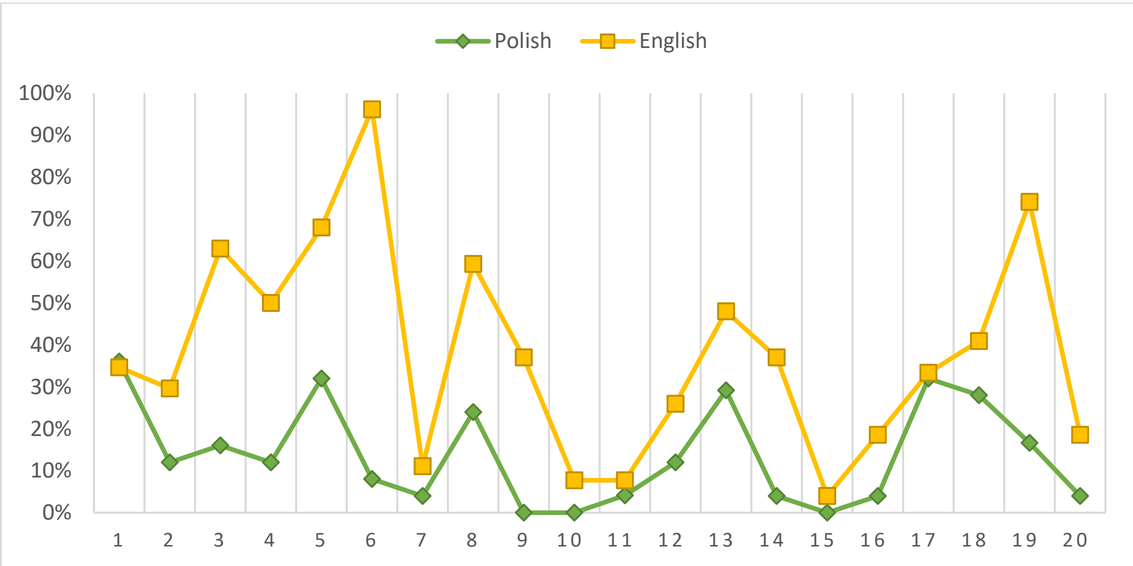


Fig. 92. Individual differences in the percentages of unvoiced realisation of /b, d, g/ in Polish (green) and English (yellow) – Group 1 at T2.

The change is particularly salient in English, however it is still quite substantial also in Polish. The two speakers, S6 and S19, who displayed the most Type 3 productions in English at T1, lead the way also at T2. Especially striking is S6, for whom unvoiced realisations reach 96% of all cases in English. On the whole, for the majority of the participants the canonical English productions oscillate within 40%-60%. Most importantly, there are no speakers for whom they are absent from their L2. Polish displayed quite an interesting behaviour as well; out of thirteen speakers who had had no Type 3 realisations at T1, at T2 there were only three such speakers, namely S9, S10, and S15. S1, S5, S12, S17, and S18 had between 30%-40% of such productions in their data. S1 is particularly interesting, as they had gone from 0% at T1 to almost 40% at T2. We can therefore see that significant changes in L2 led to robust differences in L1 as well, but these changes occurred on individual basis.

At T3, as far as group results are concerned, there were 45.3% of Type 3 productions in English and 13.9% in Polish. Individual data are given in Fig. 93.

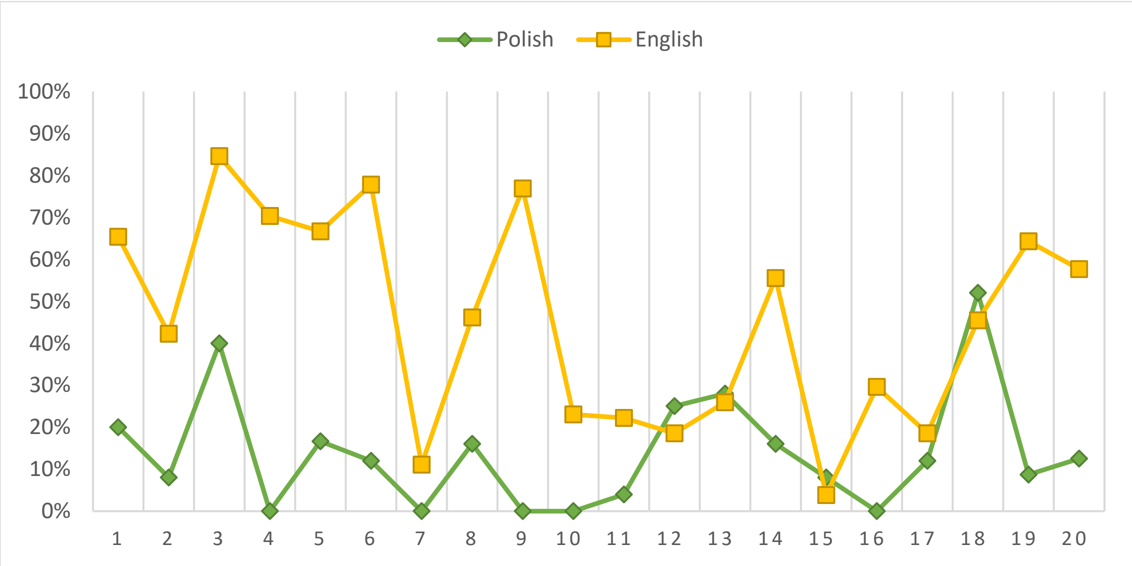


Fig. 93. Individual differences in the percentages of unvoiced realisation of /b, d, g/ in Polish (green) and English (yellow) – Group 1 at T3.

Eight speakers (S1, S3, S4, S5, S6, S9, S19 and S20) displayed over 60% of unvoiced realisations in their L2 data. Only two speakers (S7 and S15) showed 10% or less of such productions; they were present in the data from all twenty speakers. While they were five speakers for whom we did not find Type 3 tokens in their L1 (S4, S7, S9, S10, and S16), there were a few speakers who had more unvoiced items in their L1 than their L2 (S12, S12, S15, S17, and S18). Interestingly, some speakers with significant percentages of unvoiced items in L2 were found not to produce them in L1 (e.g. S4 and S9). Therefore, again, we can observe a lot of individual variation and interaction between the two languages across speakers.

Recall that it was Group 2 for whom we observed the most amount of L2-induced phonetic drift in L1; group results showed up to 59.8% of Type 3 realisations in English and 24.7% in Polish. Individual data are shown in Fig. 94.

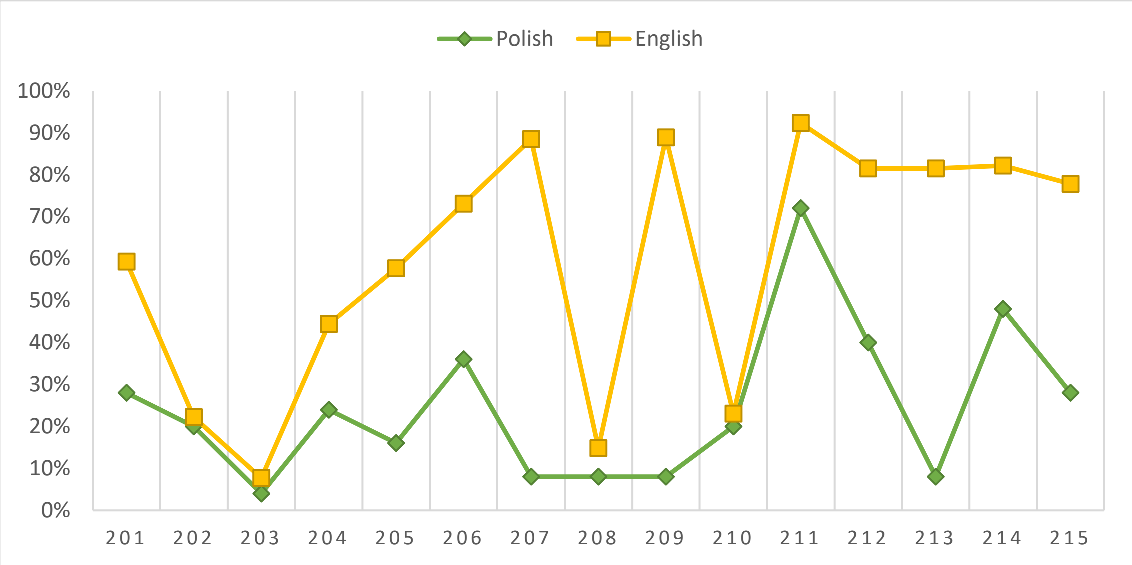


Fig. 94. Individual differences in the percentages of unvoiced realisation of /b, d, g/ in Polish (green) and English (yellow) – Group 2.

Indeed, after two years of intensive phonetic training, the majority (i.e. ten) of our participants from this group displayed over 60% of "correct" English realisations of voiced-initial plosives (S201, S205, S206, S207, S209, S211, S212, S213, S214, S215), out of which for six speakers the numbers were as high as over 80%. None of the participants had no Type 3 realisations. Interestingly, the same was true for Polish – all participants had between 4% up to 72% unvoiced tokens in their L1 data. For all of them, however, it was English that yielded higher percentages of such items relative to Polish.

Finally, in Group 3 we have seen 43.6% Type 3 tokens in English and 16.3% in Polish, suggesting that despite there not being pronunciation classes in their curriculum anymore and the pre-voicing values going back to resemble monolingual norms, Type 3 tokens persisted in their data, although to a lesser extent than what was found for Group 2. Fig. 95 illustrates the individual performance of our participants.

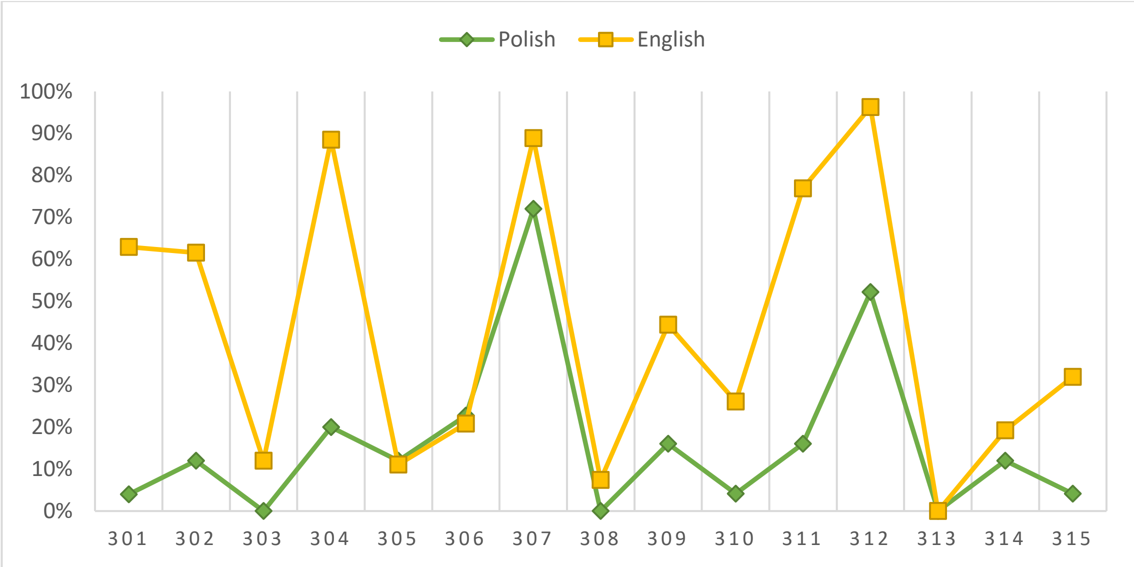


Fig. 95. Individual differences in the percentages of unvoiced realisation of /b, d, g/ in Polish (green) and English (yellow) – Group 3.

There were only four speakers for whom Type 3 realisations constituted more than 60% of all their productions (S301, S302, S304, S307, S311, and S312). For many of them, there were in fact less than 20% of canonical English realisations in their L2 data (i.e. S303, S305, S206, S308, and S313), which is a robust decrease from Group 2 results. In the case of S313, there are no Type 3 tokens in either their English or their Polish. Therefore while still present, Type 3 items are not found universally in Group 3's productions. As far as Polish is concerned, only two speakers displayed high percentages of unvoiced tokens in their L1 (i.e. S307 and S312); for the rest the numbers were lower. For six speakers there were less than 10% of such realisations in their speech (S301, S303, S308, S310, S313, and S315).

We calculated the percentage of Type 3 realisations in English for each of the participant and included it as a predictor in a Binary Logistic Regression analysis with Type 3 in Polish as the dependent variable, Session as the fixed factor, and Speaker and Item as random factors. The analysis revealed that the percentage of unvoiced realisations in L2 is a significant predictor of their occurrence in L1 (t=4.183; p<.001).

Let us turn to other acoustic variables of interest. We will return to the results concerning VOT in section 7.4.1 for a discussion.

7.3. Vowels: L2 data juxtaposed with L1 results

The acoustic parameters associated with vowels which we investigated were the two formants – F1 (F1-f_0, Bark normalised) and F2 (F3-F2, Bark normalised) – indicating vowel height and advancement. Recall that in Chapter 6 (section 6.3.1) we saw that the vowel space appears to be expanding as the phonetic training in L2 progresses. By including the L2 data in the analysis, we wanted to see whether the height and advancement of the students' L2 vowels have any effects on what happens to their L1 vowels over the course of the training.

Before we look at the findings, it is important to make some methodological comments regarding the upcoming analyses. Recall that our participants were enrolled in British and American groups; while these two accents do not really differ with respect to stop consonants, the differences in their vowel inventories are striking (cf. section 4.2). Unfortunately, it was impossible to control for the accent and make the groups equal due to the participants' availability. Seeing as it could introduce some bias into the results, it was decided to include the more numerous group and look at the students acquiring the British English model. As a result, the population sample included in the analysis got significantly smaller (Group 1 N=14, Group 2 N=7, Group 3 N=12).

The next comment regards the vowels included in the study. Our L1 data focused on the two Polish vowels: /a/ and /ɛ/. By looking at the vowel inventories of the two languages and drawing from the other studies concerned with the Polish-English language pair (cf. section 4.2.1) we have chosen the following English ones: DRESS and TRAP as a possible competitor to /ɛ/, and TRAP, BATH, and STRUT as possible competitors to /a/. As with the VOT data, we decided to look exclusively at the word reading task for the time being.

The following section will look at group results, where we will look at the vowel space and see how the L1 and L2 vowels coexist there at respective testing times. After that we will look at individual differences displayed by our participants.

7.3.1. Group results

As has been mentioned in section 6.3.1 for L1 Polish we obtained 3664 items in total; however, seeing as American English speakers were not included in this particular analysis, the number of tokens in total was subsequently reduced to 2499 (cf. Table 67 for details). In addition to L1 data, the L2 English word reading task allowed us to gather 3133 (General British) items in total (after excluding mispronunciations, mostly stemming from spelling biases, e.g. pronouncing the word *ton* as [tɒn] or Polish-induced mistakes involving words similar in the participants' L1 and L2, e.g. pronouncing the word *penguin* as [piŋgwɪn], after Polish [pingvin]. The details are given in Table 67.

Table 67. Sheer numbers of recorded vowel tokens in Polish and General British sorted for vowel quality and Group/Session.

Group	/a/	/ɛ/	TRAP	STRUT	DRESS	BATH
Group 1, T1	435	138	40	308	348	24
Group 1, T2	435	140	32	320	345	32
Group 1, T3	442	138	32	322	345	32
Group 2	217	68	14	156	167	14
Group 3	372	114	24	271	283	24
TOTAL	1901	598	142	1377	1488	126

Similarly to what was the case for the L1 tasks, in investigating the interactions between L1 and L2 vowels we plotted them on F1/F2 planes in R (R Core Team 2018), using the `ggplot2()` package (Wickham 2016) for each group and session. Notice that, overall, the ellipses corresponding to L2 vowels are much wider than the ones associated with L1 productions, indicating that there might be more variability in individual vowels' production in the second language (cf. section 1.2.3).

Starting with Group 1, Fig. 96 illustrates the vowel space for our students at the onset of their pronunciation training (i.e. T1) in October.

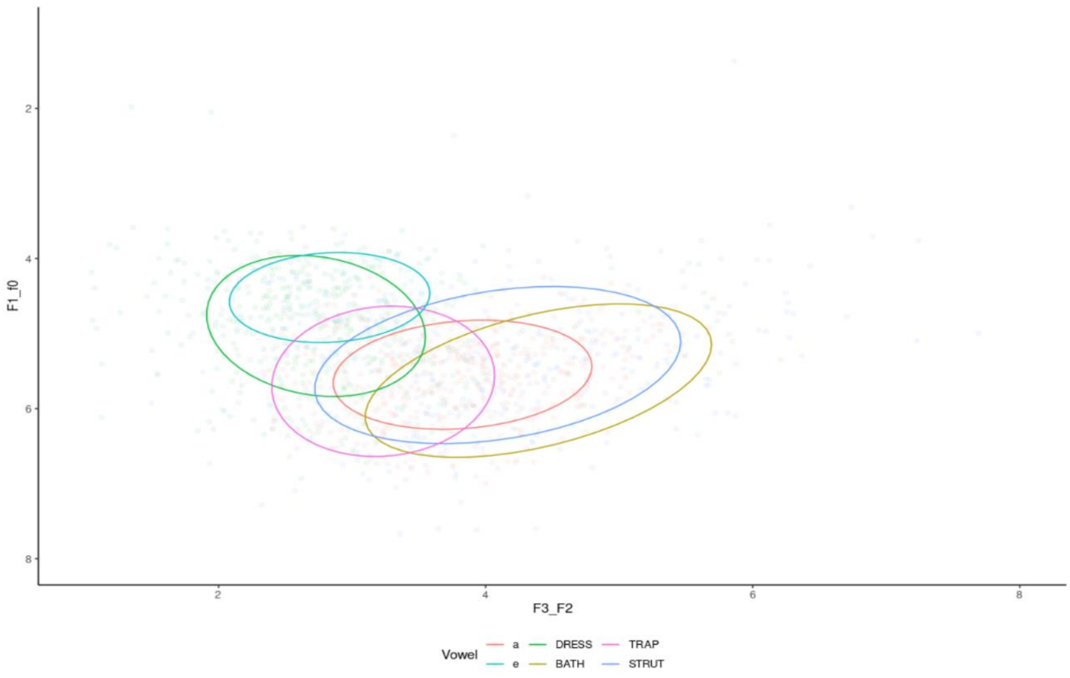


Fig. 96. L1 and L2 vowels in the word reading task at T1 plotted on the F1/F2 plane, with the ellipses corresponding to the distance of one standard deviation from the mean, sorted for vowel quality.

As can be observed, the vowel space is extremely crowded and the L1/L2 vowels (but also L2 vowels themselves) overlap to a great extent. Polish /ɛ/ (in teal) and English DRESS (in green) appear to be extremely similar; there is more variability with respect to the DRESS vowel, as the ellipses is much larger, and it reaches much lower in the vowel space than /ɛ/ does. This goes against the phonetic description of the monolingual English norms, as the DRESS vowel there is supposed to be higher than Polish /ɛ/ (indicated also by the IPA symbol used to denote the English vowel, which is /e/). TRAP (in pink) overlaps with the Polish /ɛ/ less, but it does imbricate with the L2 DRESS as produced by the students. As far as /a/ (in red) is concerned, all three L2 English vowels – BATH (in yellow), TRAP (in pink), and STRUT (in blue)– appear extremely similar in their qualities to it. In particular, while TRAP is slightly fronter, BATH and STRUT overlap almost completely, with BATH being on average only a little more back than Polish /a/.

Fig. 97 depicts the same set of L1 and L2 vowels after three months of phonetic training (i.e. at T2, in February) in Group 1's productions.

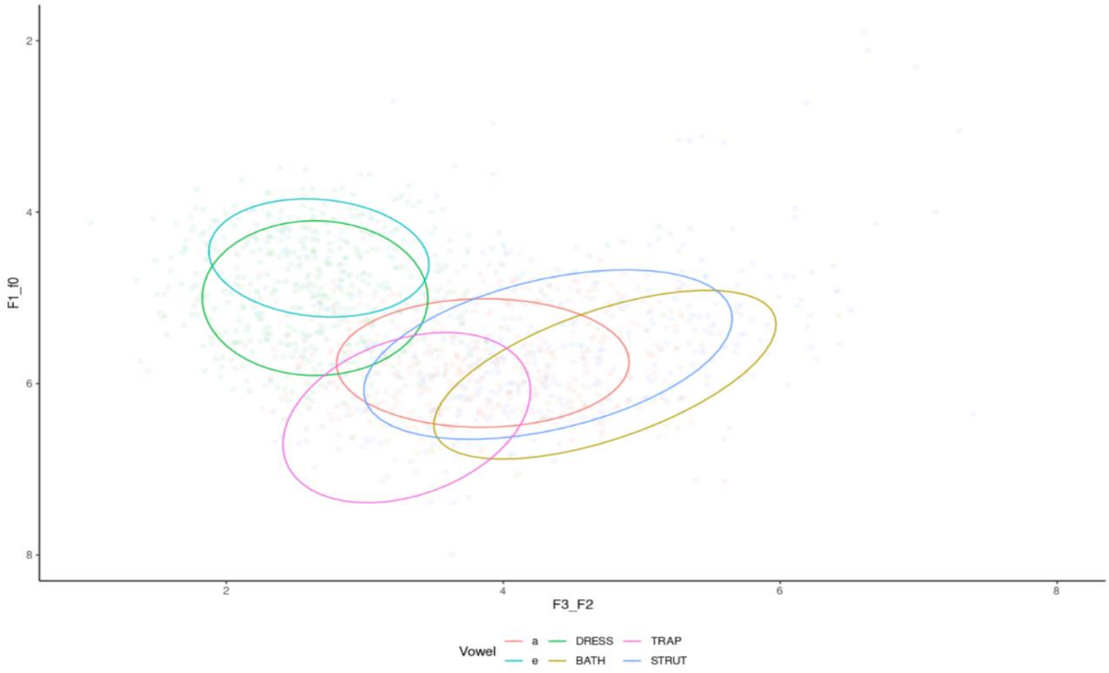


Fig. 97. L1 and L2 vowels in the word reading task at T2 plotted on the F1/F2 plane, with the ellipses corresponding to the distance of one standard deviation from the mean, sorted for vowel quality.

It appears than by and large the vowel space got bigger, at least with respect to height. Polish /ɛ/ and DRESS are still similar and DRESS is still somewhat lower than /ɛ/. With the TRAP vowel undergoing robust lowering between T1 and T2, TRAP and DRESS seem to be quite distinct and do not really overlap; Polish /ɛ/ and TRAP in particular are quite distant in their qualities. With respect to /a/ and its L2 competitors, both STRUT and BATH (the former to a greater extent) are very similar in their qualities, though both L2 vowels extend to somewhat higher qualities relative to /a/. TRAP, in turn, while still overlapping with /a/, reaches much lower qualities.

The results of L1 and L2 production tasks at the last recording session (i.e. T3, in June) are shown in Fig. 98.

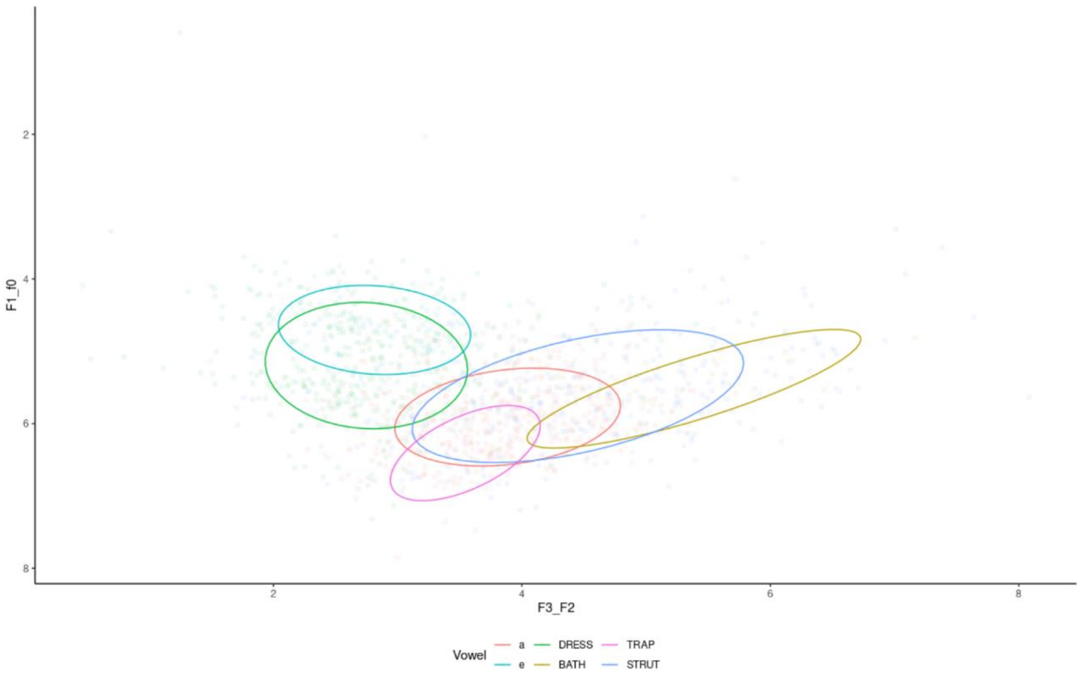


Fig. 98. L1 and L2 vowels in the word reading task at T3 plotted on the F1/F2 plane, with the ellipses corresponding to the distance of one standard deviation from the mean, sorted for vowel quality.

Immediately we can see that the ellipses corresponding to the L2 vowels (in particular to TRAP and BATH) are much smaller, even though the scale is constant; it could indicate less variability in the realisation of those vowels. Notice that BATH extends further back at T3, thereby overlapping to a much lesser extent with the Polish /a/. There is not much change with respect to STRUT; it does appear more back than /a/ but still encompasses /a/ in its entirety. TRAP appears quite similar to /a/ as well, if slightly lower. Not much difference between T3 and T2 can be seen with respect to /ɛ/ and DRESS; the contrast between TRAP and DRESS, however, is much more salient relative to what we have previously observed.

Turning to Group 2, our second year students, Fig. 99 illustrates the data we have gathered[92].

[92] Recall that this group in particular had quite a few American English learners who were excluded from this part of analysis. Thus the investigated sample of recordings is much smaller. Limitations of the present research has been discussed at length in section 5.11.

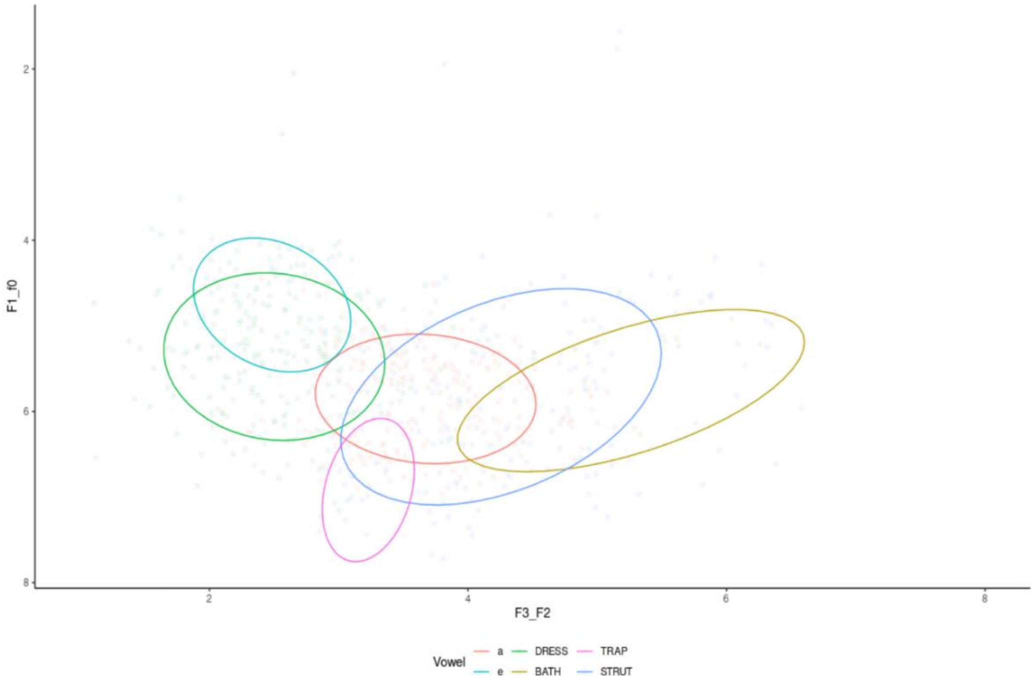


Fig. 99. L1 and L2 vowels in the word reading task in Group 2 on the F1/F2 plane, with the ellipses corresponding to the distance of one standard deviation from the mean, sorted for vowel quality.

Group 2's vowel space appears to be quite expanded. There is not much difference between Group 2 and Group 1 (at all testing times) with regards to /ɛ/ and DRESS – the two vowels are kept close together, with the L2 vowel again being lower than the L1 one. /ɛ/ and DRESS are saliently separate from TRAP. In turn /a/ and STRUT, again, overlap greatly; /a/ and BATH are relatively apart. TRAP reaches qualities of both /a/ and STRUT, but is overall much lower than either one of those.

Finally, Fig. 100 shows the results obtained from Group 3, our third year students, no longer undergoing pronunciation training.

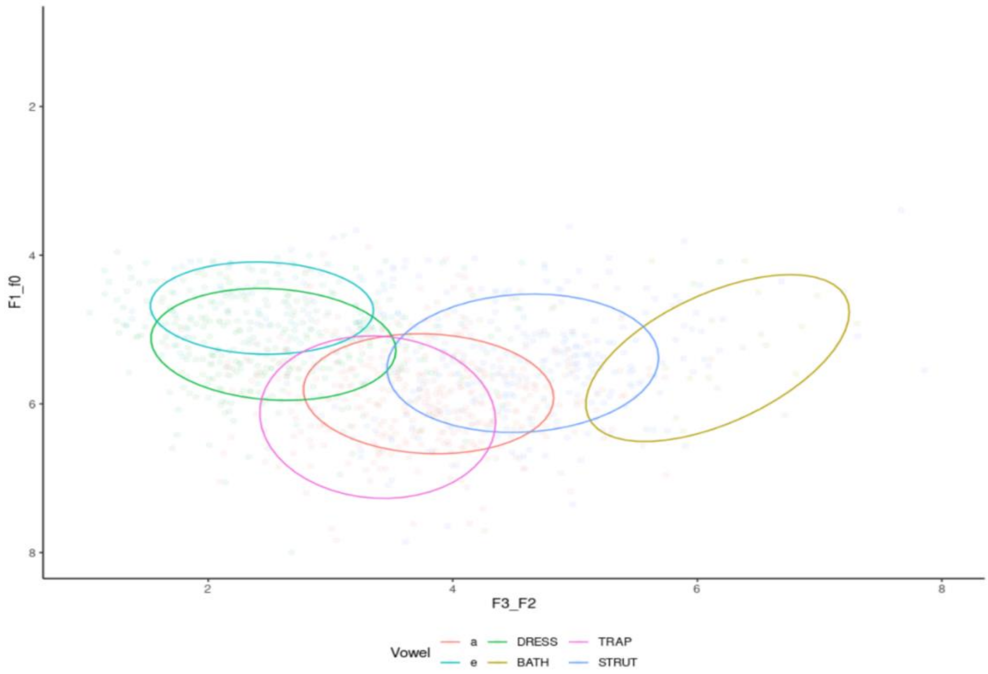


Fig. 100. L1 and L2 vowels in the word reading task in Group 3 plotted on the F1/F2 plane, with the ellipses corresponding to the distance of one standard deviation from the mean, sorted for vowel quality.

Overall the vowel space appears as expanded as what was the case for Group 2. /ɛ/ and DRESS are similar – especially in advancement – but the latter is, again, slightly lower than the Polish vowel. Interestingly, as opposed to Group 2 (and Group 1 at T3, and to an extent at T2) the TRAP and DRESS vowels overlap in some of the cases, indicating that without pronunciation training the former has undergone some rising, landing somewhere in between Polish /a/ and /ɛ/. Another difference we can observe in Group 3's data is the fact that /a/, STRUT, and BATH are kept relatively separate; /a/ and BATH do not overlap at all, there is also almost no overlap between STRUT and BATH either. /a/ and STRUT do imbricate to an extent, but less so than in the productions of the previous groups.

The three main points that one can gather from the figures presented in this subsections are as follows. First of all, there is a lot of variability in the L2 productions of the students participating in the current experiment. Second of all, in general, taking into account the assumptions made by the equivalence classification principle, it seems that it is the pairs /ɛ/-DRESS and /a/-STRUT that are deemed as equivalent to the greatest extent by our students. Hence, it is hypothesised that these two L2 vowels' acoustic characteristics may be significant contributors to the L2-induced phonetic drift in L1 that we have described in section 6.3.1. Last but not least, on the whole the phonetic space seems to have expanded over the course of the phonetic training.

The next section will look at how L2 vowel height and advancement might influence L1 vowels. Before that, however, we will discuss individual differences displayed by our participants.

7.3.2. Individual differences

In section 7.2.2 we have seen that the parameter of VOT has been subject to a lot of individual variation in the speech of our students; similar effects are expected in the case of vowel height and advancement. The following subsections will discuss these two dimensions separately (as was the case in the results section on vowels in Chapter 6, i.e. section 6.3) and look at the two Polish vowels: /ɛ/ and /a/ compared with their competitors mentioned in the discussion of group results, that is DRESS and TRAP for the former and STRUT, BATH, and TRAP for the latter.

7.3.2.1. Vowel height

Fig. 101 presents the height of three vowels: L1 /ɛ/ (light green and triangles), L2 DRESS (dark green and circles) and L2 TRAP (yellow and crosses) in individual speakers' performance. Note that the higher the F1-f_0 mean difference, the higher the first formant, hence the lower the vowel. For ease of interpretation, the *x* axis has been reversed.

Fig. 101. Individual differences with respect to vowel height in L1 /ɛ/ and its L2 competitors TRAP and DRESS across groups.

A visual inspection tells us that at the onset of the pronunciation training, the TRAP vowel was generally lower than either DRESS or /ɛ/ for the majority of the participants, which is consistent with the trends we observe now in modern RP (cf. section 4.2). TRAP and DRESS were extremely close with respect to height for five out of fourteen speakers (i.e. S6, S8, S9, S11, S14). For three of them this corresponded with a very low DRESS vowel (i.e. S6, S9, S11), while the remaining two were characterised with a relatively high TRAP (i.e. S8, S14). Aside from three speakers (i.e. S4, S6, S9), for whom their DRESS vowel was generally lower than /ɛ/, for the majority of the speakers this difference was minimal, hence these two vowels were impossible to disentangle.

At T2 the first observable difference is that for all speakers (though for four of them, i.e. S3, S4, S5, S6, to a lesser extent) the TRAP vowel is decidedly lower than either DRESS or /ɛ/. As far as the contrast between DRESS and /ɛ/ is concerned, although there was some difference for four speakers (i.e. S4, S5, S6, and S9), by and large these two vowels were at this session realised as virtually identical.

At T3, TRAP continued to be produced as lower than DRESS, thus the contrast between these two vowels was maintained by all participants. Some progress was also observed for the mean values of DRESS relative to /ɛ/; these two categories appeared to be at least slightly distinguished between by six out of fourteen speakers (i.e. S3, S4, S5, S6, S8, and S9). However, while they were not produced as identical, the contrast was counter-intuitive – the L2 DRESS was overall *lower* for all those students who produced them as different. This goes against the phonetic data which we find in the literature where Polish /ɛ/ is lower than the English DRESS vowel. Therefore, while some contrast between the two was found, Polish learners of English did not move in the direction of the English targets.

In Group 2, i.e. the second year students, our limited sample of participants successfully distinguished the two front vowels TRAP and DRESS, with the former decidedly lower (although for S206 they were rather close in height). Once again, the contrast between DRESS and /ɛ/ was extremely slight, in particular for speakers S209, S211, S212, for whom they were nearly identical. No speaker showed the desired differentiation of height with regard to DRESS and /ɛ/, namely DRESS was still equal to or lower than /ɛ/.

In Group 3, while for all speakers the TRAP vowel was still lower than DRESS, some showed some effects of TRAP raising (i.e. S304, S310, S312, S315). While we cannot judge whether this is the result of ceasing phonetic training or just a characteristics of the speech of some of these speakers (as we possess no pre-test data of those speakers), we might observe that TRAP was similarly raised in the data obtained from Group 1 at T3. Therefore we may suggest that phonetic training might be responsible for lowering TRAP to the RP norms attested by the literature - once this ends, for some speakers the vowel might get raised back to the /ɛ/-like height. With regard to the interaction between DRESS and /ɛ/, it was similar to what we have seen in the previous groups; for five out of twelve speakers these two were indistinguishable (i.e. S306, S311, S312, S313, and S315), and DRESS was on the whole lower than /ɛ/.

A more complex image emerges when we look at the height of /a/ (in light green with squares) alongside its main L2 competitors, that is TRAP (in yellow with crosses), STRUT (in brown with crosses), and BATH (in orange with vertical lines). All these three vowels are quite similar to Polish /a/ and might be claimed to be similar to it by a naïve Polish listener. Fig. 102 shows the individual differences in height of these four vowels. According to the phonetic literature discussed in section 4.2, /a/ and BATH should be the lowest (and comparable), followed by TRAP, with STRUT the highest out of the four; for ease of interpretation, the *x* axis has been reversed.

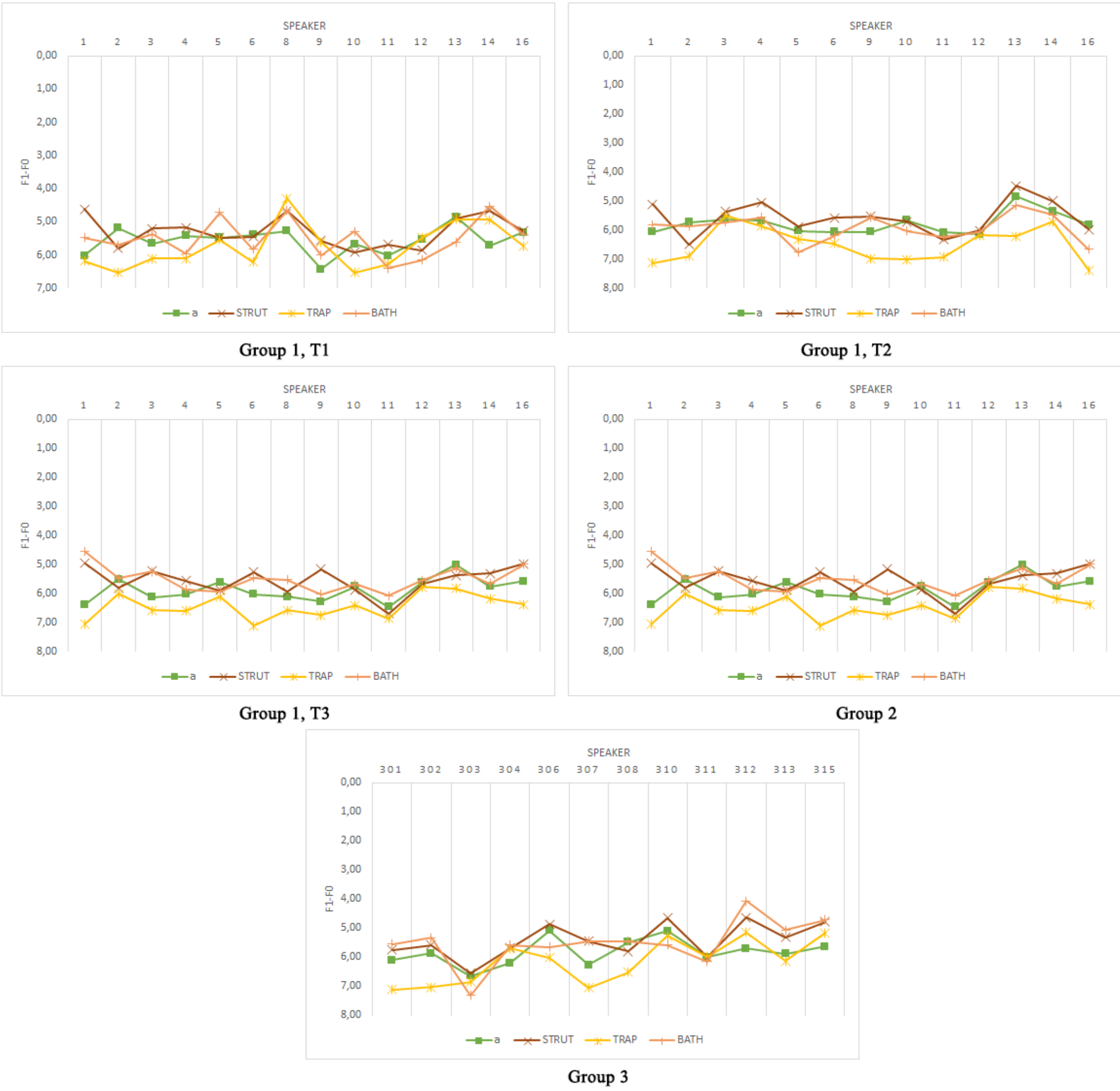


Fig. 102. Individual differences with respect to vowel height in L1 /a/ and its L2 competitors STRUT, TRAP, and BATH across groups.

Immediately we can notice that there was a lot of variability in the productions of individual speakers. At T1 some trends that could be observed include TRAP being the lowest for seven out of fourteen speakers (i.e. S1, S2, S3, S5, S6, S10, and S16). Polish /a/ and STRUT were comparable with regards to height for the majority of the participants, with the differences being very small for S2, S3, S4, S5, S6, S10, S11, S12, S13, and S16. Interestingly, in many cases two or more vowels completely overlapped. For instance for S2, S14, and S15 STRUT and BATH were identical; for S5, S12, S13 /a/ and TRAP were the same; for S5, S8, and S10 TRAP and STRUT were equal. Therefore, it appears that at T1 speakers tended not to be able to distinguish between the openness of these four vowels.

At T2 the desired height hierarchy appears to have been reached by S3. STRUT is the highest for seven speakers (i.e. S1, S3, S4, S5, S6, S12, and S14). For the majority of participants TRAP continued to constitute the lowest vowel (S2, S2, S9, S10, S11, S13, and S16). There was still plenty of confusion with respect to height present; for speakers, e.g. S10, S11, S12, or S14, the vowels were virtually indistinguishable.

At T3 TRAP remained the lowest vowel for all participants; for S2, S5, and S12 it overlapped in quality with other vowels (with BATH, STRUT, and all of them, respectively). BATH appeared to have undergone some raising, relative to the previous recording sessions; for nine speakers (i.e. S1, S2, S3, S6, S8, S9, S10, S11, and S16) it was now higher than Polish /a/, and for four of the nine it was higher than STRUT (i.e. S1, S8, S11, and S16). Overall, once again, there was a lot of individual variation and the differences with regard to height were extremely small for the majority of speakers. After one year of phonetic training it cannot be said that the group consistently differentiated between them.

In Group 2 productions, similarly to Group 1, TRAP seemed to be realised as the lowest for all speakers, though S201 and S211 merged it with their BATH vowel. While TRAP was visibly lower than the rest especially for four speakers, i.e. S212, S213, S214, and S215, the rest of the vowels overlapped to such an extent that the differences between them were minute.

Group 3 is not much different from Groups 1 and 2 discussed thus far. We see a lot of overlap with respect to vowel height – for all of the speakers in this sample two or more vowels were practically identical.

The following section describes similar issues but connected with vowel advancement. We will return to the data described herein in section 7.4.2.

7.3.2.2. Vowel advancement

Fig. 103 illustrates the Polish vowel /ɛ/ (in light green with triangles) and its two main competitors, that is DRESS (in dark green with circles) and TRAP (in yellow with crosses), now concentrating on their advancement. Note that the parameter of F2, which corresponds to vowel frontness/backness, is given here in Bark and presented as the difference between F3-F2. This means that the smaller the difference, the higher the F2, and the fronter the vowel; for ease of interpretation the *x* and *y* axes have been transposed.

Fig. 103. Individual differences with respect to vowel advancement in L1 /ɛ/ and its L2 competitors DRESS and TRAP across groups.

In the data obtained from Group 1 at T1 we can see that there was a lot of variation with respect to the advancement of these three vowels. It appears that, overall, /ɛ/ and DRESS were realised extremely close to one another in the productions of eleven out of fourteen speakers (i.e. S1, S2, S3, S4, S6, S9, S10, S11, S12, S14, and S16). The TRAP vowel seemed to be realised slightly less front in comparison with DRESS and /ɛ/, at least in the productions of eight speakers (i.e. S3, S5, S8, S9, S10, S12, S14, and S16), however the difference was robust only in the case of S3.

At T2 the two main observations to be made are as follows. /ɛ/ and DRESS got closer to each other with regards to their advancement after the first few months of phonetic training, and this observation held for all fourteen speakers. Second of all, compared to T1, TRAP appeared to be consistently less front than /ɛ/ and DRESS; this was especially salient in the productions of S4 and S10. Interestingly, at T1 this was the most visible in the data from S3; at T2 all three vowels were almost completely merged at the frontness/backness continuum.

With respect to the relationship between /ɛ/ and DRESS, not much changed between T2 and T3. The contrast between the two vowels – if it existed, which it did not for seven out of fourteen speakers – was negligible. The difference between the advancement of TRAP compared with both /ɛ/ and DRESS appeared to have become more solid. It was decidedly less front for all but one speaker, that is S1.

Group 2's data were extremely similar to what was observed for Group 1 at T3. That is, both /ɛ/ and DRESS overlapped almost completely for the majority of the speakers; the degree of contrast was extremely slight. TRAP was consistently less front for all seven speakers.

Group 3, who were no longer undergoing pronunciation instruction, appeared to present similar degree of variation as to what was the case for Group 1 at the onset of the training. Some speakers (i.e. S301, S302, S303, S306, S310, S311, S312, and S315) displayed patterns which we observed in Group 2's and Group 3 at T3's productions, that is lack of contrast between /ɛ/ and DRESS, with TRAP realised more centrally that the former two vowels. However, for some speakers TRAP and DRESS and/or /ɛ/ were equally advanced (i.e. S304, S307, S308) and there was a lot of variation between them with regard to how front or back the three vowels were. For instance, the difference between S304 and S307 was almost two Bark.

Finally, Fig. 104 illustrates the individual differences with respect to the advancement of the Polish /a/ (in light green with squares) and its main competitors, that is STRUT (in brown with crosses), TRAP (in yellow with crosses), and BATH (in orange with vertical lines). We expect the picture to be more complex here, because while overall the differences in height, rather than advancement, are generally more salient for the /ɛ/-DRESS-TRAP vowels, advancement is more robust for /a/-STRUT-TRAP-BATH, with TRAP being more front, and BATH decidedly more back than both /a/ and STRUT. For ease of interpretation, the *x* and *y* axes have been transposed.

Fig. 104. Individual differences with respect to vowel advancement in L1 /a/ and its L2 competitors STRUT, TRAP, and BATH across groups.

As expected, at T1 Group 1 presented a lot of variability. In general, TRAP was indeed fronter than their Polish /a/ (for every speaker except for S3). Polish /a/ overlapped mostly with STRUT (in the case of five out of fourteen speakers, that is S2, S10, S11, S12, and S16), though on the whole, STRUT was less front. BATH was generally the most back vowel, though for five speakers (i.e. S1, S5, S8, S12, and S16) it appeared merged with Polish /a/. STRUT and BATH were very close of five speakers (i.e. S4, S9, S12, S13, S16).

At T2 the desired order of advancement began to emerge and was observed for ten out of fourteen speakers (there were overlaps in the case of S3, S4, S9 and S16). However, the differences were sometimes extremely small. This can be noticed especially with respect to the contrast between STRUT and BATH in the productions of S3, S4, S5, S6, and S9 or STRUT and /a/ in the case of S2, S10, S11, S12, S13, S14, and S16.

While the general pattern observed at T2 remained the same at T3, namely the students followed the sequence of advancement, there were still some overlaps. These targeted especially /a/ and TRAP (merged or nearly merged by six out of the fourteen speakers, i.e. S2, S8, S10, S11, S12, and S13) and /a/ and STRUT (five out of the fourteen speakers, i.e. S2, S10, S11, S12, and S13). Five speakers merged more than two vowels (i.e. S2, S10, S11, S12, and S13).

Group 2 also adhered to the desired order of advancement, however in some cases, once again, the differences were very minute. STRUT and BATH were nearly identical in the productions of three speakers (i.e. S206, S211, and S212), while /a/ and TRAP in the data from four of them (i.e. S206, S209, S213, and S215). Speaker 214 did not truly differentiate between /a/, STRUT, and TRAP, although their BATH was saliently more back.

In turn Group 3, whose data did not really differ from the previously discussed groups, merged mostly /a/ with TRAP, which were realised identically in the recordings from S301, S311, S312, and S315 and nearly identically from S310. BATH was the most back vowel for all participants from that group, though it was very close to STRUT for S310, while STRUT and /a/ were very close for S306 and S315.

This subsection showed that, similar to what was the case for height, vowel advancement in L2 is also subject to a lot of individual differences. The results will be discussed in section 7.4.2.

7.4. Discussion of the results

This chapter has attempted to juxtapose the L1 results from the word reading task described in detail in Chapter 6 with the participants' L2 performance in an analogical task in English in order to assess the relation that these two languages have. The comparison has been limited to three acoustic parameters: VOT (positive and negative, in milliseconds), the percentages of the three types of voicing realisation, and vowel quality (F1-f_0

and F3-F2 as an average taken from the middle 20% of the vowel, given in Bark). This section will summarise the findings. A more general discussion, enriched by theoretical issues, will be given in section 8.1.

7.4.1. VOT

In discussing the parameter of VOT, we once again focused on the asymmetrical behaviour of the voiced and voiceless series.

Let us begin with the voiceless series and the comparison of Polish and English. 5060 tokens were considered: 2494 from English and 2566 from Polish. The group level results of voiceless plosives in the initial position show that L1 plain voiceless stops and L2 aspirated stops were kept separate across all recording sessions, as early as T1. The difference between Polish and English reached ca. 17 ms at the onset of the training, while towards the end of the two-year phonetic instruction the contrast magnitude was ca. 32 ms. Importantly, the two languages were saliently kept apart even in Group 3's productions, in the absence of phonetic training, with the difference being ca. 29 ms.

Contrast was maintained across all places of articulation, and similarly to the group level results, while it was present already at the start of the training, it was the most robust in Group 2's productions and remains intact even after the training has ended.

These results suggest that aspiration is relatively easy to acquire by Polish speakers of English, which has previously been observed (e.g. Schwartz 2020). So much so, that they keep their voiceless categories in L1 and L2 separate without any explicit pronunciation instruction, already at T1 (cf. Fig. 79). The training that they undergo, however, appears to boost their performance, as English VOT is elongated (in particular in Group 2's and 3's data, with the averages across all places of articulation being longer relative to Group 1 at T1). Therefore, the instruction helps the students produce more target-like aspiration while at the same time not exerting any influence on their L1 VOT duration in the voiceless category.

A visual inspection of individual differences with regard to maintaining contrast between the two languages appears to corroborate the group level results. The majority of students successfully distinguished the two categories already at T1 (14 out of 20 speakers). This was also observed at T2, T3, peaking in Group 2's productions, for whom the contrast magnitude was the most robust. Nonetheless, to a great extent, the differentiation of plain voiceless and aspirated stops persisted even after the training had ended and was also found in Group 3's production (in all but one speaker's).

Taken together, it appears that the students who participated in the present experiment found aspiration easy enough to acquire and from the SLM perspective we may gather that a new category was formed. This resulted in a relatively good separation of the L1 and L2 sound, with L2 VOT values falling somewhere between L1 and L2 monolingual norms[93] and L1 VOT values being unaffected by L2 pronunciation training.

On the other hand, it appears that the voiced series is characterizsd by equivalence classification between the two languages. In general, Type 1 realisation of voiced series in both languages appeared to be prevalent. Therefore, we observed L1 interference in L2 students' productions. At group level, the differences between Polish and English were negligible. Very minute shortening of negative VOT appeared to have taken place, with the pattern replicating what was found for Polish (i.e. progressive shortening, peaking at year two of phonetic instruction, then lengthening when the training was over; cf. section 6.2), with one small difference – the T2 lengthening did not take place in English. Therefore, phonetic instruction appeared to affect both languages in the same way. Pre-voicing differences between English and Polish turned out to be significant at the beginning of training (with English values being longer) and in Group 3's data (with Polish yielding longer means).

On the whole, across all places of articulation a great deal of variation with regard to the duration of pre-voicing between the two languages was observed. Significant differences between Polish and English (as far as pre-voicing duration is concerned) were found only in the case of labial and velar stops at T1 and labial stops in Group 3's productions. In the majority of cases it was English that yielded longer values, and with there

[93] Seeing as there was no English monolingual control group to whom we could compare our students' productions, it is difficult to say whether or not the L2 speakers were *successful* in their acquisition of aspiration, i.e. whether they acquired native-like accuracy. The fact that L2 speakers produce L2 sounds with intermediate VOT is predicted by SLM. We may postulate that a new category has been formed, as all groups consistently retained contrast between their L1 and L2 categories. However, it is possible that some equivalence classification of that newly formed L2 category with actual English monolingual norms has taken place insofar as the most important thing for our speakers was to keep the category apart from their L1 voiceless stops by making VOT longer; the question as to whether or not it was long enough was secondary. L2 learners could have tuned out the details to a certain extent and deemed their "longer-than-L1-Polish-values" as equivalent to English targets.

being aspiration present in the students' productions, we may see that when pre-voicing was present the speakers opted for maximal dispersion in their L2 productions.

Individual differences revealed that contrast magnitude for the majority of speakers was less salient in voiced stops than what was the case for aspiration juxtaposed with plain voicelessness. Furthermore, as the training progressed, the contrast appeared to become less observable, with many students presenting virtually no difference between the two languages by T3. In Group 2's productions, small differences between the two languages are also more common, whereas four speakers from Group 3 appear to be comparable with the state of the students at the very onset of the phonetic instruction, that is with the differences being quite striking. Interestingly, for two of them Polish yielded much longer values while for the other two - it was English. In general, however, differences in pre-voicing length were quite imperceptible, at least to the ear of the author of the thesis.

From the perspective of the SLM (Flege 1995), the L2 English data, when contrasted with the L1 Polish results, provides evidence in favour of the assumption that the learners form a new category for the voiceless aspirated stops and as a result, less cross-linguistic interaction is observed therein. On the other hand, both L1 and L2 voiced stops appear to be deemed as equivalent, which leads to both more L1 interference in L2 (observed as a significant number of English unvoiced stops realised with pre-voicing) and L2-induced phonetic drift in L1 (observed as a significant shortening of pre-voicing in L1 as the training progresses and an increase in unvoiced realisations of Polish stops).

And indeed, perhaps the most interesting result was the actual comparison of the percentages of unvoiced realisations in L2 English and L1 Polish. On the whole, the numbers were much higher in English (T1=14.9% vs 2.6%; T2=38.2% vs. 13.9%; T3=45.3% vs. 13.9%; Group 2=59.8% vs. 24.7%; Group 3=43.6% vs. 16.3%). Therefore, it may be suggested that it is, indeed, English that exerts influence on Polish – the higher the number of unvoiced realisations in the L2, the higher the numbers in the L1. Furthermore, with the numbers growing in sequential manner, it seems that phonetic training is actually quite successful at teaching unvoiced realisations, which persist even after it has ended, however the success rate is less stable than what was the case for aspiration – while in Group 2' productions nearly 60% of tokens was realised without pre-voicing, after one year with no explicit instruction the numbers were down to ca. 44%.

Overall, the voiced category showed a lot more variability compared to the voiceless series. There appeared more variability at group level with respect to pre-voicing duration, the patterns of contrast magnitude at speaker level were also less stable. Moreover, although we have seen fair amount of success with respect to the acquisition of unvoiced, English-like, realisation of the voiced category, once the training was over the success rate dropped significantly. Therefore, once again, it may be stated that, similarly to L1 results, an asymmetry in the behaviour of the two series of stops was found, with more cross-linguistic interaction taking place in the voiced category.

7.4.2. Vowels

When the results of L2-induced phonetic drift effects in vowel productions in the word reading task were being discussed in section 6.3, we saw that, while such effects were present, they were generally quite modest and generally they entailed movement towards more peripheral positions, suggesting vowel space expansion.

The analysis included in this chapter targeted a much smaller and less counterbalanced sample, as it was limited to the participants enrolled in groups acquiring the General British model. Therefore any of the results obtained herein are merely a description of what happened in our sample of students, and as such cannot be extrapolated onto a bigger population. They may be conceived of as a pilot study, whose main aim was exploratory. We are reticent to make any strong claims on the basis of the data presented herein.

By including the L2 English vowels in our analysis and plotting them alongside L1 Polish ones, we did get a bigger picture of the changes. One of the most important observations to be made that the phonetic space employed in vowel productions got expanded once the L2 speakers started undergoing phonetic training. Therefore, it might be claimed that the L1 vowels did not necessarily move in the direction of L2 targets; since, from the SLM perspective, they all exist in a common phonological space, the space gets bigger so that the new L2 categories being acquired can be accommodated.

Indeed, the group level results showed that at the onset of the training the vowels appear difficult to disentangle and they were crowded together in a very small fragment of the vowel space. As the training progressed, the phonetic space appeared to expand.

Both /ɛ/ and DRESS moved towards more front positions, TRAP got lower, while BATH started moving back. This progressed across the recording sessions insofar as to making the difference between the phonetic space at T1 and in Group 3's productions look strikingly different. As far as L2 target acquisition is concerned, fair amounts of overlap between the vowels were still present, even in Group 2's and 3's data.

At the individual level, we saw that once the phonetic training started, TRAP was realised as much lower, in comparison to both DRESS and /ɛ/, though once the training ceased, it got very close to DRESS for some speakers, resembling the state of the matter from T1. As far as height is concerned, DRESS and /ɛ/ were virtually identical for most speakers, with DRESS being only slightly lower for them. As far as vowel advancement is concerned, the picture was more complex, though once again we did see the effects of training, in particular at T3 and in Group 2, for whom TRAP was much more advanced in comparison to DRESS and /ɛ/. This contrast became less robust in Group 3's productions, suggesting that in terms of both height and advancement TRAP moved back towards DRESS and /ɛ/ for many speakers. Similarly to what was the case for height, the differences in vowel advancement with regards to /ɛ/ and DRESS were negligible. The two vowels were nearly indistinguishable for the majority of speakers, regardless of the recording session.

With regard to /a/ and its interaction with the L2 vowels (STRUT, TRAP, and BATH), the picture that emerged was rather complex, in particular with regard to height. All four vowels did not really differ in this dimension; for four speakers from Group 2 TRAP was slightly lower than the other three. However, while in Group 1's data at all three testing times there was a lot of variability, Group 2 and Group 3 displayed (in the case of most speakers from that group) a clear sequence of the vowels, indicating that over the course of the training they did indeed get more distinguishable.

While these results encompass but a small sample of the participants who took part in the study, they do indicate two main things. First of all, they highlight the direction of L2-induced phonetic drift – rather than moving towards English targets, adhering to the SLM's predictions, the L1 vowels move to more extreme points in the vowel space, therefore causing its expansion. Second of all, there is some cross-linguistic interaction present in the data, with a lot of overlap between the vowels.

Chapter 8: Implications and questions for further research

8.1. General discussion and questions for further research

A detailed discussion of the phonetic study undertaken for the purposes of the present dissertation has already been provided in sections 6.4 (for the L1 results) and 7.4 (for the L2 results alongside the L1 data). Here we will focus on how the provided data allow us to answer the research questions enumerated in section 5.2.

The first research question read, "[w]ill L1 Polish students majoring in English display L2 pronunciation training-induced phonetic drift in their Polish productions?" The phonetic findings of the study conducted for the purposes of the present dissertation indicate that the answer to this question is *yes*, in part. Indeed, we have seen that both in the category of stop consonants and vowels some changes in the acoustic make-up of the sounds under investigation were displayed. In the case of consonants, the Polish productions appeared to move towards English-like values but only in the voiced series – prevoicing got progressively shorter and the numbers of unvoiced realisations increased. The voiceless category, in turn, remained stable across all sessions and groups and did not display any L2 influence as it did not diverge from monolingual norms. As far as vowels are concerned, we have observed progressive expansion of the vocalic space as the training continued. Therefore, phonetic drift in L1 induced by pronunciation training in L2 appears to have been attested. Importantly, the effects were noticeable in both tasks alike.

The second research question asked whether, "(…) the possible phonetic drift effects be different depending on the year of study? Will the degree of phonetic drift be less observable, seeing as the participants are advanced, rather than novice, learners?" This stemmed from the fact that Chang (2013) observed that phonetic drift is a phenomenon that is most prevalent in the learners who are being exposed to the L2 for the first time. While studies have seen phonetic drift in more advanced learners, it was a possibility that phonetic drift would be very minute in the productions of speakers who had been learning English for many years prior to their admission into the English programme at university. The present acoustic studies have found quite salient evidence in favour of phonetic drift even in very advanced learners, though it appears that the Polish productions of our speakers had diverged from monolingual norms earlier than the onset of the phonetic instruction. There are two pieces of data that are suggestive of this. First of all, our speakers kept their L1 and L2 voiceless categories separate from the very start. Therefore, aspiration might be claimed to be such an easy feature for Polish speakers to master that they form a new phonetic category for aspirated stops without explicit training or classes with native speakers. Second of all, pre-voicing was by and large shorter in the speech of our students at T1 as well. Phonetic training, however, boosted the numbers of unvoiced realisations in their productions, which were overall absent up until the start of explicit instruction. Vowel space expansion also appears to have started after the training had commenced, as the most significant divergences from monolingual norms were found beginning with T2.

A different population of interest would be the former students of the English programme, who have graduated and still work with English professionally. We saw that in Group 3, who no longer participated in phonetic training, some effects of drift persisted but in some respects they returned to monolingual norms. It would be interesting to see whether or not they would still be there (e.g. Type 3 items) after a few years since completing the same pronunciation instruction as the one described here.

The third research question was concerned with whether, "(…) the phonetic drift effects [will] be contingent on the principle of equivalence classification?" The answer to this question is *yes*, in part. We could observe the effects of equivalence classification in particular with the stop consonants – while the voiceless series were kept apart and the cross-language interaction between them was minimal (i.e. two separate categories were at play), the voiced series in both languages appeared to be classified as identical in both languages and was subject to more salient phonetic drift effects. As far as vowels are concerned, it appears that /ɛ/ and DRESS were classified as equivalent, though TRAP was a strong competitor as well, especially right at the onset of the instruction and when it had ended (i.e. in Group 3's data). With /a/, it appeared that both STRUT and TRAP were the vowels that overlapped with the Polish category to the greatest extent, with BATH moving towards a more back position once the training started. Therefore, some equivalence classification effects in vowel production were observed as well.

The last research question read, "[w]ill the success with which English laryngeal contrasts as well as vowels are acquired influence the amount of phonetic drift observed?" The inspection of the comparison of L1 and L2 data seems to suggest that the answer to this question is also *yes*, in part. With regards to the consonants, the L2 voiceless category of stops which has been acquired rather successfully by the vast majority of our participants did not exert much influence on the L1 voiceless stops production. We have seen that the participants kept the two categories separate and as a result, the cross-linguistic influence between them was minimal. With the L2 voiced category, we have observed more L1 interference in the L2 data, with a lot of Polish-like pre-voicing in the English data, with the success rate of unvoiced tokens acquisition being less stable. As a result, with the bi-directional influence in place, phonetic drift effects in the L1 data were also much more dramatic, with pre-voicing getting progressively shorter and a lot of unvoiced realisations of Polish voiced plosives (even in an intervocalic context in the sentence reading task, an environment very conducive to voicing) being found. With regards to vowels, we have seen that successful acquisition of some vowels (in particular, the /ɛ/ vs. TRAP and /a/ vs. BATH contrasts) allows them to move away from the Polish categories and possibly exert less influence on the behaviour of the L1 vowels; in turn, with DRESS and /ɛ/ and /a/ and STRUT being hard to distinguish, we may expect that these two L2 vowels were influencing the Polish categories to a greater extent. These issues, however, require more research in the future.

There remains a great deal of recordings which have been made for the purposes of the present dissertation yet have not been included in it (e.g. L2 sentence reading task or other onsets aside from stop consonants; individual differences in more detail) and still require work, which constitute an exciting avenue for further research on this population.

8.2. Phonological implications of the results

An important aim of the present thesis was to ascertain that the acoustic data presented here would not be devoid of phonological merit. Therefore, both Chapters 3 and 4, aside from providing phonetic descriptions of the consonantal and vocalic inventories of Polish and English included some phonological theories dealing with laryngeal contrasts and vowel systems as well. The phonetic experiment designed for the purposes of the present thesis was meant to test the theoretical predictions put forward by those frameworks. This section will recall these once again, this time in order to see how well they were able to account for what the empirical data tell us.

Let us begin with the laryngeal theories which were the main topic of Chapter 3. Three main theories described there were: feature theory (Chomsky and Halle 1968), Laryngeal Realism (Honeybone 2005), and Onset Prominence (OP; Schwartz 2016 *et seq*.). We assumed that it is equivalence classification that is the source of both L1 interference and L2-induced phonetic drift; therefore, with Polish and English in mind, phonetic drift is expected to be more dramatic in the series which are deemed to be phonologically equivalent in both those languages. Therefore, we must first recall what each of those theories sees as equivalent.

In feature theory phonetic drift is predicted in both series of stops. Regardless of the actual phonetic implementation of voicedness/voicelessness, /b, d, g/ in both Polish and English are ascribed the feature [+voice] and /p, t, k/ bear the feature [-voice]. Therefore, irrespective of the phonetic differences between both pre-voiced and unvoiced as well as plain voiceless and voiceless aspirated stops, both series are equivalent in both languages. As a result, phonetic drift is predicted to occur in both voiced and voiceless categories to a comparable degree.

Laryngeal Realism, in turn, differs a lot in both its attitude towards including the phonetic detail in its representations as well as the choice of features. Recall that, in contrast to traditional feature theory, the specifications postulated by Laryngeal Realism relied on unary features or elements. Pre-voiced stops in Polish were specified in terms of [voice]/{L} while unvoiced stops in English were unspecified; on the other hand, Polish plain voiceless stops were unspecified whereas aspirated stops in English were described in terms of the feature |spread glottis|/[aspiration]/{H}. Since neither series was identical, neither series was predicted to induce equivalence classification resulting in phonetic drift. In other words, equivalence classification could not take place as nothing could be deemed phonologically equivalent.

Finally, Onset Prominence moves away from segment-oriented linear representations, opting for hierarchical trees with the CV sequence being the universal (cf. Fig. 8). As far as the laryngeal representation go, the voiced series (regardless of whether or not pre-voicing is present) is identical in both languages (cf. Fig. 14); the difference between plain voiceless and voiceless aspirated stops lies at the level at which we assign the feature [fortis] (cf. Fig. 13). In the case of English aspirates, [fortis] is assigned at the Closure level and then trickles down onto Noise and VO, resulting in aspiration. In the case of Polish plain voiceless stops, [fortis] is assigned at the VO level, yielding short, positive

VOT. Feature [voice] is not employed in the phonological account of languages with two-way laryngeal stops; [fortis] is the only feature that is necessary. Voicing, in turn, is conceived of as the carrier signal (cf. Modulation Theory; Traunmüller 1994) and as such, is not part of phonology. Looking back at the representations given in Fig. 14, we see that more drift is predicted for the voiced series, with those being identical in both languages (that is, susceptible to more interaction). The voiceless categories is similar only at the VO level (but different at the levels of Closure and Noise, hence susceptible to only minimal interaction). As a result, an asymmetrical behaviour of the two series is expected.

For clarity purposes, we repeat Table 13 below, to once again summarise the predictions made by the three frameworks in question. Hence, both the representations and the expected degree of drift is presented in Table 68.

Table 68. Summary of the representations and degree of L2-induced phonetic drift in L1 from the perspective of three laryngeal theories.

Theory	Specifications for Polish	Specifications for English	Degree of drift
Feature Theory (Chomsky and Halle 1968)	[+voice] [-voice]	[+voice] [-voice]	drift predicted in both series of stops
Laryngeal Realism (Honeybone 2005; Harris 1994; Lombardi 1991)	[voice] for /b, d, g/ Ø for /p, t, k/	Ø for /b, d, g/ [spread glottis], {H} for /p t, k/	drift predicted in neither series of stops
Onset Prominence (Schwartz 2016)	Ø for /b, d, g/ [fortis] for /p, t, k/ at VO-level	Ø for /b, d, g/ [fortis] for /p, t, k/ at C-level	drift predicted for the voiced series; minimal or none for the voiceless series

The results of the experiment provide support to the last framework, that is to OP. In both tasks the asymmetry predicted by the OP laryngeal typology was attested. One of the crucial points of the OP proposal is that it postulates that there is no feature [voice] in either English or Polish, a claim which was hinted at previously by Cyran (2014). Moreover, thanks to moving away from linear representations and instead postulating hierarchical structures, it allows us to account for the VOT typology with just one feature [fortis].

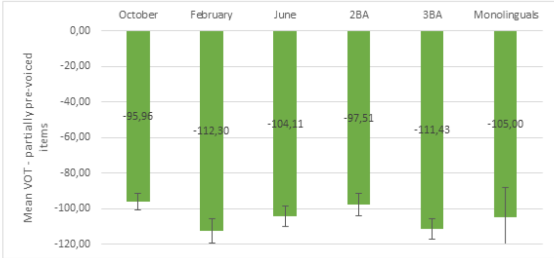
Admittedly, pre-voicing, in general, is a very well-established cue for voiced stops. In Keating (1981) we find a claim stating that,

> "(…) [in Polish] because the voicing contrast is between voicing lead and voicing lag, the only information needed to identify a stop as voiced or voiceless is whether the VOT value is positive or negative, not its numerical value. This is not the case for English, where (…) the two categories are much closer together along the VOT continuum on the positive side, and much more precise information is required for phonetic categorization" (1981: 1263).

Given the fact that pre-voicing is so crucial for distinguishing the two laryngeal categories in true-voice languages, the employment of the feature [voice] to represent voiced categories is sometimes taken for granted. However, the results of the present study indicate that the absence of pre-voicing does not mean that a voiceless item is going to be produced. It has been shown that even without pre-voicing, Polish speakers maintain contrast between the voiced and voiceless categories by means of other cues (i.e. f_0 and F1 at vowel onset, with the former being slightly more reliable). Perception studies corroborate these findings – lack of pre-voicing does not hinder the perceptibility of the contrast between the two categories (Schwartz and Arndt 2018; Schwartz et al. 2019). This is also what we can predict from the hierarchical representations proposed by OP (cf. Fig. 14). Notice that in Polish (and true-voice languages in general), the feature [fortis] is assigned at the VO level (i.e. the onset of the vowel). Therefore, we may expect that the phonological feature exerts influence on that part of the vowel (i.e. pitch and F1 is raised after a voiceless onset, due to there being a phonological feature). No such effects are expected after a voiced onset, as there are no features to trigger them. The OP representations, once again, predict contrast maintenance, without resorting to any feature [voice].

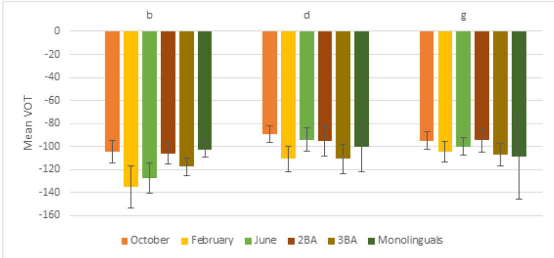
The common occurrence of voicing breaks in negative VOT also might be seen as a problem to the postulates in favour of [voice] being phonological. Recall that in the two production studies we distinguished between Type 1 (i.e. fully pre-voiced; cf. Fig. 32) and Type 2 (i.e. partially pre-voiced; cf. Fig. 33) tokens. Type 2 items were not included in our discussion on phonetic drift as they mirrored patterns found in Type 1 items (and as such were not interesting, especially since their occurrence as such has been found not to be evidence of L2-induced phonetic drift), yet they could have skewed the result, had both Type 1 and Type 2 tokens been treated as equal. Nonetheless, juxtaposing the two types is quite intriguing. Fig. 105 presents a comparison in negative VOT duration between Type 1 and Type 2 tokens in the word reading task (including general group-level results and across all three places of articulation).

fully pre-voiced (word reading task)



partialy pre-voiced (word reading task)

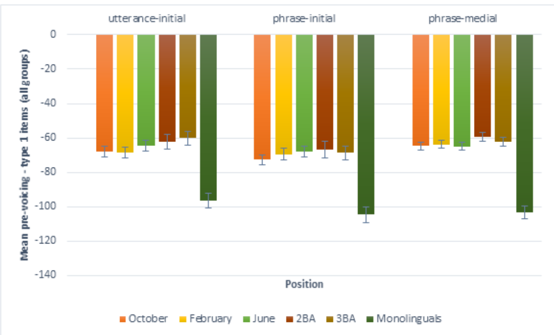
fully pre-voiced (sorted for POA; word reading task)



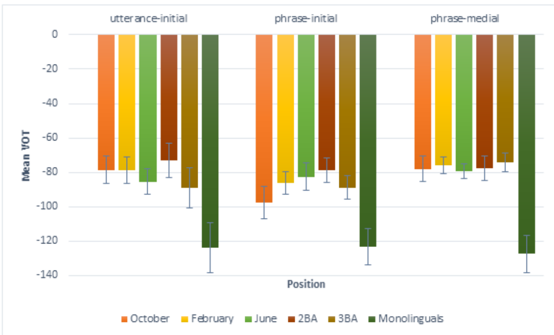
partially pre-voiced (sorted for POA; word reading task)

Fig. 105. A comparison of Type 1 and Type 2 tokens with respect to negative VOT duration in the word reading task (L1 Polish)

As can be observed, the presence of breaks in negative VOT contributes to longer durations overall, and this holds true for all groups as well as for all places of articulation. Similar results were present in the sentence reading task as well. This is illustrated in Fig. 106.



fully pre-voiced (sentence reading task)



partially pre-voiced (sentence reading task)

Fig. 106. A comparison of Type 1 and Type 2 tokens with respect to negative VOT duration in the sentence reading task (L1 Polish)

In the sentence reading task we can see that across all groups and all prosodic positions, negative VOT is decidedly longer in the tokens containing a break. Finally, the same pattern was found when we compared Type 1 and Type 2 tokens in L2 English.

fully pre-voiced (word reading task, L1+L2) partially pre-voiced (word reading task, L1+L2)

Fig. 107. A comparison of Type 1 and Type 2 tokens with respect to negative VOT duration in the word reading task (L1 Polish +L2 English)

The relationship between voicing breaks and negative VOT measures poses a challenge to phonological accounts that posit a feature [voice] in Polish. As noted by Schwartz (2020), one hand, since the breaks contribute to longer negative VOT duration as such, this could imply fortition of the voiced segments associated with overall longer segment duration. On the other hand, the fact that there *is* a break in voicing entails what might be thought of as weakening of [voice], implying lenition. OP avoids this apparent contradiction by assuming that voicing is not phonological. From the phonetic perspective, the breaks do not constitute a problem, as speech aerodynamics contribute to the vocal fold vibration ceasing right before the release of the closure. Indeed, Davidson found that even in supportive environment (such as what was the case for our sentence reading task, where the keywords were preceded and followed by a vowel, therefore being placed in an intervocalic position conducive to voicing), "[voicing] may not necessarily take optimal advantage of the surrounding phonation" (2016: 49). From Davidson's perspective, the voicing we found in the sentence reading task was, most probably, voicing 'bleed' from the preceding vowel (Davidson 2016: 43). In other words, 'bleed' is just a phonetic effect of periodicity from the vowel carrying over into the closure stage of the stop. Indeed, recall that a relevant methodological problem was making the decision whether voicing in phrase-medial position belongs to the voiced stop or is the voicing bleed from the preceding vowel (cf. Fig. 38).

We have also seen that, in general, the variability in pre-voicing duration was much more dramatic compared to what was the case in the voiceless category (cf. Fig. 82, Fig. 84, Fig. 86, Fig. 88, Fig. 90, where the individual results showcase how much the speakers differed with regards to pre-voicing values). As noted by Schwartz (2021), “[g]reater variability along a given phonetic dimension is associated with a lower likelihood that the feature in question has *phonological* status”. In our data, the voiceless categories in both languages appeared to be more stable, which from the perspective of OP indicates that they are specified for a phonological feature. With pre-voicing being a phonetic detail, less stability is expected.

The results obtained as far as pitch is concerned, also appear to be more compatible with the OP proposals, compared to the rest. We have seen that it is postulated that the presence of a phonological feature might exert influence on the fundamental frequency at the onset of the vowel following a stop (Ohde 1984). Therefore, when it comes to the predictions, in feature theory, since both voiced and voiceless categories are specified for a feature, a symmetrical polarisation effects should be expected – voiced onsets should lower pitch while voiceless ones should raise it. In turn, Laryngeal Realism predicts that pitch raising after voiceless onsets in aspiration languages is due to the feature [spread glottis] (or the element {H}) specification in the voiceless series. By logical extension, then, in voicing languages we should find the opposite results (i.e. the [voice]/{L} specification in the voiced series should lower pitch). However, as demonstrated by Kirby and Ladd (2018) as well as the present dissertation, the effects on pitch appear to be universal, regardless of the language type. No lowering of pitch was observed. This is what OP predicts – the effects on pitch are due to the fortis specification at the VO level, which is present in both aspiration and true-voice languages. Since the voiced category in either language is not specified for a feature, no effects on pitch are to be found. Therefore, OP appears to more successfully predict the phonetic effects that can be found in both aspiration and true-voice languages; they have something in common which is encoded in phonology.

With respect to vowels, there is less to say, as neither feature theory nor Element Theory offer any testable predictions with regard to cross-language interaction between vowels. OP, in turn, provides two different structures, which depend on the VO level affiliation (that is, there are two possibilities: either a consonantal VO affiliation, or a vocalic one; cf. Fig. 25). English and Polish vowels in #CV sequences opt for opposing settings[94] in this respect and the vocalic structures are similar only at the VT level. Recall that in the case of voiced stops, the entire CV structure was identical for both Polish and

[94] As noted by Schwartz (2016: 47) the affiliation may vary within individual languages, as some evolutionary relics may persist. For a discussion, see Schwartz 2016.

English /b, d, g/ and we observed quite robust L2-induced effects in that category. The voiceless series was similar only at the VO level, and the effects were minimal (cf. Fig. 14). Thus, in a way, the situation with regards to vowels can be compared to that of voiceless stops: here we also have only one level of the structure that is identical between those two languages. Hence, minimal interaction between vowels in Polish and English is expected. The results presented here appear to adhere to these claims: while some effects of L2 phonetic training were found, they were not as striking as what was attested in voiced consonants but slightly more salient than what was the case for voiceless stops. The expansion of vowel space was observed for both /a/ and /ɛ/ and it closely mirrored the patterns observed in voiced stops (with the magnitude of the effects peaking in second-year students' productions).

To conclude this part of the discussion, phonetic research helps us evaluate and provide empirical support to the representations postulated by phonology. In particular, the results of the present thesis allowed us be able to more clearly observe the boundary between phonetics and phonology. In particular, the empirical studies conducted herein, but also found in e.g. Schwartz 2020, has allowed us support the claim that voicing lies in the realm of phonetics rather than phonology. In turn, phonetics needs these representations in order to find areas worth investigating. The theoretical works of Cyran (2014) who postulated that [voice] need not be phonological or Schwartz (2017) who proposed that two-way laryngeal contrast be represented without the feature [voice] were what really draw attention to the equivalent structures for /b, d, g/ in Polish and English. Therefore, it was the phonological proposals put forward by those authors that triggered the execution of the studies discussed in this work. Without those proposals, the formulation of the research questions and hypotheses tested in the present dissertation would not have been possible.

8.3. Phonetic drift versus phonetic and phonological attrition

In section 2.2 we have seen that, overall, there is a lot of confusion with regard to the distinction between phonetic drift and phonetic/phonological attrition. To recap, Table 69 summarises the main differences that have been discussed which allow us to disentangle the two phenomena.

Table 69. The main differences between phonetic drift and phonetic/phonological attrition

L2-induced phonetic drift in L1	L2-induced phonetic/phonological attrition in L1
no apparent decrease in proficiency in L1	apparent decrease in proficiency in “1
short-term changes due to recent exposure to L2	long-term changes due to long-term exposure to L2 or a new dialect of L1
no effects on fluency	deterioration of fluency = negative connotations
not noticeable to listeners (i.e. requires instrumental methods to be observed)	noticeable by listeners (=“foreign accent” in L1)
reversible once the exposure to L2 ceases	non-reversible, permanent changes

Phonetic drift might be thought of to be the starting point, eventually resulting in phonetic and phonological attrition. Possibly one of the first studies explicitly addressing the issue of first language attrition was conducted by Major (1992), with a very apt title *Losing English as your first language*, which stressed the negative connotations associated with attrition (as mentioned in section 2.2)[95]. He studied five speakers of American English who had been living in Brazil for a significant number of years on their productions of /p, t, k/ in English and Portuguese in formal and casual speech. The most robust effects of L2 Brazilian Portuguese on L1 English, with the speakers losing aspiration and producing VOT values moving in the direction of the short-lag L2 (but not quite as short, with the VOT falling somewhere in between the two categories) has been observed in casual speech. Major assumed, then, that the effects of attrition are first observed in casual speech and afterwards target formal productions (by this logic, seeing as we have observed phonetic drift in L1 in formal tasks, it might be even more pronounced in casual speech, which may be an exciting avenue for future research). Unfortunately, Major did not look at the voiced category, despite having apparently recorded some words as they were included in the appendix.

A very interesting case study was conducted by Mayr et al. (2012). It investigated two monozygotic Dutch twins (for both of whom English was their L2), with one of them living in the Netherlands her whole life and the other one living in the UK with her English-speaking husband for thirty years. The twin who spent 30 years in an L2-dominant environment, when compared with her sister’s productions, has been shown to realise the voiceless plosives (in both Dutch and English) with values intermediate to both her L1

[95] But see Flege 1987 (discussed in section 1.1.6, and hence not repeated here) and his study of Americans living in Paris and the French living in Chicago. “Attrition” as such was not explicitly addressed there.

and L2, that is longer than her sister's Dutch VOT, but shorter than what is usually reported for English. Her voiced category displayed native-like pre-voicing and remained unaffected by attrition (in line with the proposal presented in section 8.2)[96].

Lev-Ari and Peperkamp (2013) studied voiceless stops productions of English residents living in France (with their length of residence varying from 4 to 49 years, median=17 years) and they found that in both a sentence reading task and in a free conversation, English VOT durations resembled French monolingual norms in that the VOT was shortened. Furthermore, they observed that the poorer the inhibitory skills of a given speaker, the greater the influence on French on their English. Since French is an example of a voicing language, whose /p, t, k/ structures resemble what we observed for Polish, the fact that it influenced English entails a structural change in the representations of English voiceless stops. Sadly, the authors did not look at what happens to the voiced series' realisations in the speech of this population.

There exist studies showing that speakers sometimes opt for maximum dispersion – they once again move away from monolingual norm but not in the direction towards L2 but opposite to it. Thus, for instance in Flege and Eefting (1987) advanced L1 Dutch speakers of L2 English produced their native /t/ with VOT values which were shorter than monolingual norms, as a result enlarging the contrast magnitude between L1 and L2. Similarly, de Leeuw et al. (2011) showed that out of ten L1 German speakers who lived in Canada for over 30 years, two showed effects of "overshooting" the monolingual norms observed as far as tonal alignment of pre-nuclear rise is concerned inasmuch as making it more saliently dissimilar from the L2 norm, relative to what was found for L1 Germans living in Germany. As noted by the authors, in this case the attrition effects "resemble neither the L1 nor the L2" (de Leeuw et al. 2012: 113). A pertinent question is whether or not we can speak about attrition if, instead of diverging from monolingual norms, they become monolingual to an extreme extent. This could, rather, be an example of hypercorrection and an effort made in order to maintain monolingual norms or some sort of polarisation effects (Keating 1984).

On the whole, L1 attrition is not the focal point of the present thesis. Rather, the point of the present section was to present some research done on this topic and show

[96] Her vowels were also compared to her sister's and it was found that she produced her Dutch vowels with unusually high F1. In fact, they were much higher than what was observed by Chang as far as phonetic drift effects are concerned, once again stressing the difference between short- and long-term changes associated with those two different phenomena.

how distinct the effects that we see here are from the phonetic drift studies. Having illustrated this, let us consider the differences between the two phenomena on the example of Polish and English, bearing in mind the results of the production studies the present thesis provided.

The leading assumption of the thesis was that what lies at the root of L2-induced phonetic drift is the principle of equivalence classification (Flege 1995). The L1 and L2 categories which are phonologically different will be subject to minimal cross-linguistic interaction, whereas L1 and L2 categories deemed to be phonologically identical will be susceptible to significantly more cross-linguistic influence which is expected to be bi-directional. Therefore, we will see both L1 interference in L2 and L2 influence in L1 within those equivalent categories. We have also seen that with regards to languages with two-way laryngeal series, the behaviour of initial laryngeal series is asymmetrical (with Polish described in the present work but also cf. e.g. Podlipský et al. 2020, Sučková 2018, Dokovova 2015). The voiced category is much more likely to be subject to phonetic drift effects, relative to the voiceless series. In the case of Polish, we have observed these effects in both progressive shortening of pre-voicing averages and in an increase of English-like, unvoiced realisations of Polish /b, d, g/. In turn, /p, t, k/ was virtually unaffected by L2 pronunciation instruction. Moreover, even though the effects were measureable with the help of acoustic software, the changes were too minute to be spotted by the naked ear (at least, this was the impression of the author of the present thesis, a native Polish speaker herself).

In our phonological considerations in section 8.2 we have also concluded that out of the three phonological theories proposing laryngeal typology of two-way systems, it is OP that best predicts the empirical findings of the two production studies. Let Fig. 108 remind us the representations postulated by this framework for Polish (two right-most trees) and English (two left-most trees).

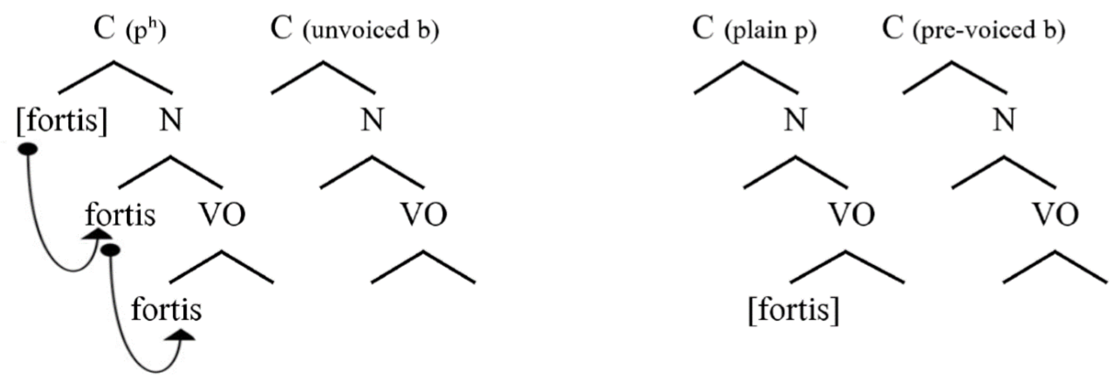


Fig. 108. OP representations of voicing contrast in aspiration (two left-most trees) and voicing (two right-most trees) languages (after Schwartz 2016).

Assuming, once again, that phonetic drift targets the identical structures, OP predicts more dramatic drift for the voiced series and minimal or no drift effects in the voiceless series (as it is similar only at the VO level). What happens in the event of attrition, then?

A proposal of a formalisation within the OP framework can be put forward. It is suggested that attrition results in the new category (L2) replacing an existing L1 category. As a corollary, it is going to go beyond the principle of equivalence classification; it is not going to target the representations which are identical in the two languages, as swapping them would make no difference. Rather, the L2 category formed in the process of L2 acquisition is going to substitute the L1 category, and the latter will be lost. In turn, the structures which are the same in L1 and L2 may continue to interact freely.

Let us, again, take the example of the Polish-English pair to illustrate what is predicted. Fig. 109 illustrates a proposed formal representation of attrition.

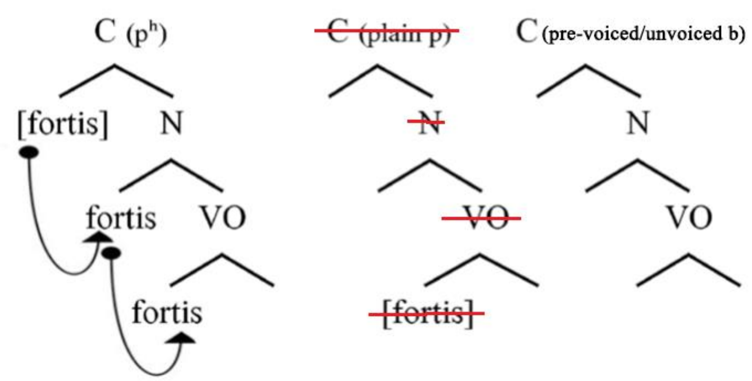


Fig. 109. The formal representation of phonetic attrition in consonants, with the new L2 category of an aspirated stop (left-most tree) replacing the L1 category of a plain voiceless one (middle tree) within OP.

As can be seen, due to attrition taking place, the existing L1 category which was realised as a plain voiceless stop is lost and subsequently replaced by an L2 category of a voiceless aspirate acquired in the process of L2 acquisition. Once kept separate, due to either substantial long-term exposure to L2 or lack of contact with L1, the "non-native" category takes place of the "native" one[97]. As a result, Polish speakers of English with attrition will display aspiration in their L1 productions (though the VOT values might not be as long as the L2 targets, they will still be heard as different from the monolingual norms). Additionally, such speakers can (but do not have to) produce the voiced L1 series without pre-voicing – in general, both scenarios are possible (i.e. the presence or absence of pre-voicing in their speech) as the structures are identical in both languages and pre-voicing is conceived of as a phonetic detail in OP.

A similar scenario is envisioned for the vowels. In the case of the two production studies we have observed some vowel space expansion studies but, overall, the magnitude of the changes was not substantial. This still goes in line with the structures proposed by OP; both Polish and English vowels are equivalent at the VT level, therefore some effects of drift are expected and they, indeed, were found.

From the perspective of OP, in the process of attrition the status of the VO affiliation changes. In the case of Polish speakers of English that would entail a replacement of the L1 structure (left-most tree) with the L2 structure (right-most tree), therefore resulting in VO changing the affiliation from vocalic to consonantal. This is illustrated in Fig. 110.

[97] It should be noted that this formalisation is an initial proposal that should be explored further. An alternative possibility would be to introduce a mechanism, essentially being the exact opposite of trickling (cf. section 3.5), which would entail some sort of "feature climbing". The legitimacy of introducing such a mechanism needs further work; therefore while, possibly, more viable options can be explored within OP, for the purposes of the present thesis we postulate a replacement of structures resulting from attrition.

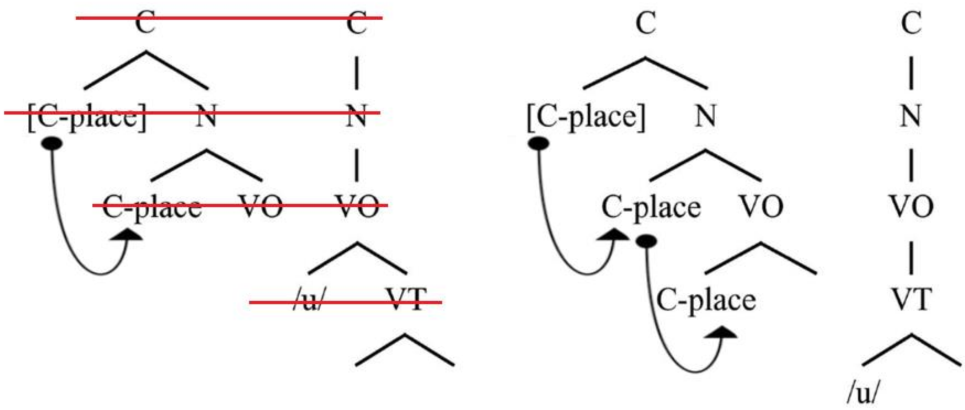


Fig. 110. The formal representation of phonetic attrition in vowels, with the new L2 representation of consonantal VO affiliation replacing the L1 consonantal one within OP.

What this implies is that Polish speakers of L2 English who display attrition, could be found to have the effects of vowel inherent spectral change in their L1 Polish. VISC was briefly mentioned in section 4.3.3. Essentially, it has been shown that English displays much more dynamic formant movements compared to Polish (Schwartz 2016). These discrepancies stem from structural differences between English and Polish, which are associated with VO affiliation. If we assume that such a structural change (i.e. the change of VO affiliation) occurs in the event of attrition, we may expect VISC to develop in the speaker's L1 Polish[98]. Such speakers may thus display more formant movement in their L1 Polish vowels, with more diphthongal qualities thereof, as the CV transition (due to the consonantal affiliation of VO) is somewhat slowed down in English (Schwartz 2021). With phonetic drift, such dramatic effects are not predicted – and have not been attested in our data.

The proposed structures open new avenues of investigation as they offer testable hypotheses to be explored. For instance, assuming that the difference between phonetic drift and attrition really lies at how monolinguals perceive the L2 learners' L1 productions, a perception study using the production data obtained for the purposes of this thesis as stimuli could be interesting. No audible divergence from the monolingual norms

[98] In theory, attrition could target any of the structures which are different in a given language pair. For instance, a notable difference between Polish and English in OP is the presence of submersion process in the latter (Schwartz 2016). Submersion is responsible for, among other things, an optional lack of release of the final stops in English. If the process found its way to a Polish speaker's L1 phonological representations, it could lead that speaker to display unreleased stops in Polish. Anecdotal evidence shows this is found in the L1 productions of the author of this thesis.

should be noticed by naïve speakers in such a perception experiment. Additionally, ideally, we would like to compare the students' productions to the data coming from a population with the possibility of phonetic/phonological attrition taking place. Both an acoustic study of such a population as well as subsequent perception tests comparing those two groups would be worth conducting.

The investigation of the phenomena of phonetic drift and phonetic/phonological attrition allows us also to further develop the theoretical implications of the SLM (and the SLM-r) model. It lets us assess what happens to the newly formed L2 categories and the existing L1 ones in the long run; if we assume that attrition is the end-product of phonetic drift, it appears that the speakers strive not to have too many competing categories in their common phonological space. While, on the whole, the L1 plain voiceless stops and L2 aspirated stops are evaluated as different enough, with excessive L2 input, the L2 category may oust the L1 one, as the latter is used to a lesser extent. Since the L1 pre-voiced and L2 unvoiced stops are seen as equivalent anyway, and therefore form one category, no such competition occurs, so no attrition effects take place therein. This is especially relevant to the SLM-r, where less focus on the end-state result of the L2 acquisition is placed.

Hopefully, all findings presented in the thesis, alongside the theoretical implications thereof, contribute to the state of knowledge regarding L2-induced phonetic drift and serve as the basis for further exploration of the issues associated both with drift and attrition.

Conclusion

This research project aimed at investigating the issue of L2-induced phonetic drift in L1 by conducting a phonetic experiment and subsequently studying the phonological implications thereof. In other words, the main objective was to describe the phenomenon in question, assess to what extent it is predicted by selected phonological frameworks, design a study which would provide empirical data, and see whether or not any of the frameworks is able to account for the results.

Before the research questions were addressed, the L2 acquisition model was presented. The choice of the *Speech Learning Model* was warranted as it is, to the best of my knowledge, the only model which explicitly states that since the two languages that the bilingual speakers have exist in the common phonological space, the interaction between them is bi-directional. Before the background on L2 influence on L1 was presented, it was necessary to ascertain that the discrepancies in terminology were sorted out. Therefore, an important section of the first two chapters of the dissertation consisted of a discussion delineating the differences between phonetic drift and phonetic attrition, among other things.

Having presented phonetic studies dealing with other language pairs, it became clear that there was a dearth of research concerning L1 Polish-L2 English. While some studies had been conducted, they appeared to either deal not strictly with bilingualism (but, for instance, with multilingualism; cf. Sypiańska 2017; Wrembel 2011) or focused on one category only (cf. Sypiańska 2017, Wrembel 2011). It was of utmost importance to include both consonants (including voiceless and voiced series) and vowels in order to obtain the fullest picture of possible drift effects. While the study in itself was an interesting subject to investigate, I believe that it would have been incomplete without a thorough phonological (that is, more theoretical) discussion. This is why before the study was presented, two chapters had been presented which dealt with laryngeal phonology and the phonology of vowels. There we considered the way L2 speech and cross-linguistic interaction between languages is handled, and what predictions are made with regard to the behaviour of L1 under the influence of L2. We looked at the laryngeal and vocalic typology from the perspective of the SLM's principle of *equivalence classification* which states that less interaction takes place between categories which are perceived as different.

Those deemed to be similar are subject to both L1 interference in L2, which can result in foreign accent, and L2 influence on L1, which can result in phonetic drift. This happens because speakers tend to tune out the acoustic details and merge the two categories into a diaphone.

The thesis consisted of eight chapters in total: five theoretical and three empirical. Chapter 1 introduced selected L2 speech models and discussed the main issues of L2 speech acquisition. Chapter 2 provided background information on the main concept of interest: phonetic drift (Chang 2012), operationalised as changes in the acoustics of L1 sounds under the influence of learning L2. Chapter 3 and Chapter 4 looked at the phonologies of laryngeal contrast and vowels respectively. The research questions and main hypotheses, alongside the methodology of the acoustic phonetic study undertaken for the purposes of the present thesis were presented in Chapter 5. Chapter 6 presented the L1 findings, focusing on four main acoustic parameters: VOT (positive and negative, with the latter subdivided into three distinct types), pitch and F1 at vowel onset, and vowel quality (F1 and F2 separately). These parameters were presented for two distinct tasks: word reading and sentence reading. Chapter 7 juxtaposed the L1 results of the word reading task with L2 data from an identical task, performed by the same group of participants, limiting the discussion to VOT and vowel quality. It allowed us to see the interaction between the two languages. Finally, Chapter 8 provided a general discussion of the results, together with the final evaluation of the phonological frameworks from Chapters 3 and 4. Additionally, this chapter put forward a proposal of an attempt at formalising the difference between phonetic drift and phonetic attrition from the perspective of the framework that best accounted for the empirical data presented in the thesis which opens up an exciting avenues for future investigation of the subject.

The general conclusion of this project is that phonetic drift effects are observable in the speech of advanced Polish learners of English and appear to be boosted by explicit pronunciation instruction. Furthermore, while in the consonants we observe an asymmetry in the behaviour of the two series of stops (with the voiced series being much more susceptible to cross-language interaction, while the voiceless series seems more stable), the effects in vowels seem to be an example of an expansion of the phonetic space. Out of all phonological frameworks presented in the theoretical part, one theory in particular appears to stand out and be best suited for predicting the phonetic behaviour of both vowels and consonants, namely Onset Prominence.

SUMMARY

The effects that one's native language can exert on the pronunciation in one's second language have been studied quite thoroughly. The opposite, however, that is the influence that one's second language may have on the native productions has only been brought to the forefront of attention in the last decade or so, beginning with a series of studies authored by Charles Chang (2010 *et seq*.). He coined the term *phonetic drift* to refer to short-term changes in the pronunciation of one's native language resulting from recent exposure to a new language. Ever since then, more and more language pairs have been investigated, with the purpose of learning which acoustic parameters, and to what extent, are subject to drift. Overall, while the phonetic data have been described, not many authors have attempted to couch their research in theoretical discussions. Phonological frameworks should, after all, be able to predict the possible drift effects.

The present thesis aims to fill this niche. Aside from the experimental part, which describes the effects of intensive phonetic training in L2 English of L1 Polish university students, it describes some of the major phonological theories, with particular focus placed on whether or not they provide any predictions about the degree of phonetic drift that we may expect to find in the speakers' Polish productions. The starting point is the principle of *equivalence classification* purported by the Speech Learning Model (Flege 1995), which assumes that the source of bi-directional interaction between two languages is which categories the learners deem as equivalent in the particular language pair.

The phonetic experiment investigated the initial stop consonant (both voiced and voiceless) and vowel productions in #CV sequences in the speech of L1 Polish learners of L2 English in both real and apparent time and in two different tasks: word reading and sentence reading. The data obtained from the bilinguals were compared with a control group of functional Polish monolinguals. The results indicate that, indeed, phonetic drift is attested in the productions of this population. In the case of L1 stop consonants, the voiceless category remained relatively unaffected, whereas the voiced series seemed susceptible to the non-native influence. When compared with the students' L2 data, a new category for the English voiceless stops appeared to have been formed and was kept separate from the L1 Polish voiceless category. On the contrary, the voiced series in both languages seemed to have been classified as equivalent and as a result displayed a great

deal of cross-linguistic interaction. In the case of L1 vowels, we observed a phonetic space expansion, with the L1 targets moving to more peripheral positions in the vowel space. These results appeared to replicated the findings obtained by the authors concerned with other language pairs.

Upon theoretical reflection, the asymmetry in the behaviour of L1 stop consonants and the expansion of the vowel space was accounted for by only one of the phonological frameworks, that is Onset Prominence (Schwartz 2016 *et seq*.). Not only does this representational environment predict the be degree of drift, but it also allows us to formally predict what happens at the further stages of phonetic drift, that is phonetic attrition.

STRESZCZENIE

Wpływy, jakie język ojczysty może wywierać na wymowę w języku drugim zostały na przestrzeni lat zbadane dość dogłębnie. Sytuacja odwrotna, to jest wpływ języka drugiego na język ojczysty stały się przedmiotem badań dopiero w ostatniej dekadzie. Zapoczątkowała to seria badań autorstwa Charlesa Changa (2010). To on zdefiniował pojęcie dryftu fonetycznego, który to opisuje krótkotrwałe efekty dostrzegalne w wymowie języka ojczystego, wynikające z niedawnej ekspozycji na język drugi. Od tego czasu coraz więcej par językowych zostało zbadanych, a ich celem było odkrycie parametrów akustycznych, które podlegają dryftowi fonetycznemu, oraz określenie do jakiego stopnia się to dzieje. Ogólnie rzecz ujmując, o ile dane fonetyczne zostały przedstawione, niewielu autorów próbowało je umieścić w szerszej dyskusji teoretycznej. Teorie fonologiczne powinny być w stanie przewidywać takie efekty.

Niniejsza praca doktorska stara się wypełnić tę lukę. Poza częścią eksperymentalną, która bada efekty intensywnego treningu fonetycznego w języku angielskim, w którym to uczestniczą polscy studenci, praca opisuje również jedne z najważniejszych teorii fonologicznej, ze szczególną uwagą na to czy i na ile przewidują prawdopodobieństwo wystąpienia dryftu. Punktem wyjścia jest dla nas zasada równorzędnej klasyfikacji (ang. *equivalence classification*; Flege 1995), która to zakłada, że źródłem dwukierunkowej interakcji pomiędzy dwoma językami jest które kategorie zostaną uznane przez uczących się za identyczne w danej parze języków.

Eksperyment fonetyczny tu opisany badał początkowe spółgłoski zwarto-wybuchowe (dźwięczne i bezdźwięczne) oraz samogłoski w sekwencjach #CV, czyli znajdujące się po nagłosie spółgłoskowym w czasie rzeczywistym i pozornym w dwóch różnych zadaniach: czytaniu listy wyrazów i listy zdań. Dane, które udało nam się zgromadzić od studentów zostały porównane z kontrolną grupą polskich osób jednojęzycznych. Wyniki eksperymentu wykazały, że efekty dryftu fonetycznego u studentów przechodzących trening fonetyczny w języku angielskim są dostrzegalne. Jeśli chodzi o spółgłoski zwarto-wybuchowe, kategoria bezdźwięczna pozostała relatywnie niezmieniona, podczas gdy kategoria dźwięczna była tą bardziej narażoną na wpływy międzyjęzykowe. Po porównaniu danych z języka polskiego z danymi z języka obcego w tej samej populacji, można było zauważyć, że kategoria spółgłosek bezdźwięcznych z języku angielskim utworzyła

nową kategorię, odseparowaną od swojego odpowiednika w języku polskim. W odróżnieniu do kategorii bezdźwięcznej, spółgłoski dźwięczne w obu językach były sklasyfikowane jako identyczne, przez co wykazywały duży stopień interakcji. Jeśli chodzi o samogłoski, dostrzec można było ekspansję przestrzeni samogłoskowej, przez co polskie samogłoski przesunęły się na bardziej peryferyjne pozycje. Te wyniki w dużej mierze replikowały wyniki zaobserwowane w innych parach językowych.

Asymetria w zachowaniu się polskich spółgłosek językowych i ekspansja przestrzeni samogłoskowej idzie w parze z przewidywaniami teorii *Onset Prominence* (Schwartz 2016 *et seq*.). Teoria ta nie tylko przewiduje stopień dryftu, ale pozwala nam również formalnie reprezentować nasze przewidywania dotyczące tego, co stanie się w kolejnych fazach nabywania języka obcego, kiedy to dojdzie do fonetycznej atrycji.

References

Amengual, Mark. 2018. "Asymmetrical interlingual influence in the production of Spanish and English laterals as a result of competing activation in bilingual language processing", *Journal of Phonetics* 69: 12-28.

Anderson, Stephen R. 1985. *Phonology in the twentieth century. Theories of rules and theories of representations.* Chicago: University of Chicago Press.

Antoniou, Mark, Catherine T. Best, Michael D. Tyler, and Christian Kroos. 2010. "Language context elicits native-like stop voicing in early bilinguals' productions in both L1 and L2". *Journal of Phonetics* 38, 4: 640-653.

Antoniou, Mark, Catherine T. Best, Michael D. Tyler, and Christian Kroos. 2011. "Inter-language interference in VOT production by L2-dominant bilinguals: Asymmetries in phonetic code-switching", *Journal of Phonetics* 39, 4: 558-570.

Aperliński, Grzegorz. 2012. Is VOT enough? (Paper presented at The international conference on native and non-native accents of English (Accents 2012), Łódź, 12-14 December 2012.).

Avery Peter and William Idsardi. 2001. "Laryngeal dimensions, completion and enhancement", in: Hall, T. Alan (ed.). *Distinctive feature theory*. Berlin-New York: Mouton de Gruyter, 41-70.

Babel, M. 2012a. "Evidence for phonetic and social selectivity in spontaneous phonetic imitation", *Journal of Phonetics* 40, 1: 177-189.

Babel, M. and D. Bulatov. 2012b. "The role of fundamental frequency in phonetic accommodation", *Language and Speech* 55, 2: 231-248.

Backley, Phillip. 2011. *An introduction to Element Theory*. Edinburgh: Edinburgh University Press.

Backley, Philip. 2012. "Variation in Element Theory", *Linguistic Variation* 12, 1: 57-102.

Baker, Wendy and Pavel Trofimovich. 2005. "Interaction of native- and second-language vowel system(s) in early and late bilinguals", *Language and Speech* 48, 1: 1-27.

Balas [Bogacka], Anna. 2004. On the perception of English hugh vowels by Polish learners of English. *Preceedings of the University of Cambridge second postgraduate* *conference in language research*, 43-50. [https://repozytorium.amu.edu.pl/bitstream/10593/4684/1/Bogacka%20Balas%20perception%20of%20English%20high%20vowels%202004.pdf] (date of access: 25 May 2020).

Balas [Bogacka], Anna, Geoffrey Schwartz, Monika Połczyńska-Fiszer, Paulina Zydorowicz, and Paula Orzechowska. 2006. "The production and perception of schwa in second language acquisition: the case of Polish learners of English", in: Dziubalska-Kołaczyk, Katarzyna (ed.). *IFAtuation: A life in IFA: A festschrift for Professor Jacek Fisiak on the occasion of his 70th birthday*. Poznań: Wydawnictwo Naukowe.

Balas, Anna. 2018a. "English vowel perception by Polish advanced learners of English", *Canadian Journal of Linguistics* 63, 3: 1-30.

Balas, Anna. 2018b. *Non-native vowel perception: the interplay of categories and features*. Poznań: Wydawnictwo Naukowe Uniwersytetu im. Adama Mickiewicza.

Balogné Bérces, Katalin and Bálint Huszthy. 2018. "Laryngeal Relativism predicts Italian", *Yearbook of Poznań. Linguistic Meeting* 47: 153-177.

Beckman, Jill, Michael Essen, and Catherine Ringen. 2013. "Evidence for laryngeal features: Aspirating vs. true-voice languages", *Journal of Lingu*istics 49, 2: 259-284.

Beckman, Jill, Pétur Helgason, Bob McMurray, and Catherine Ringen. 2011. "Rate effects on Swedish VOT: evidence for phonological overspecification", *Journal of Phonetics* 39: 39-49.

Bennet, Wm. G. and Sharon Rose. 2017. "Moro voicelessness dissimilation and binary [voice]", *Phonology* 34: 473-505.

Bennett, Wm. G. 2015. *The phonology of consonants: harmony, dissimilation, and correspondence*. Cambridge: Cambridge University Press.

Benson, Cathy. 2002. "Transfer / Cross-linguistic influence", *ELT Journal* 56: 68-70.

Best, Catherine. 1995. "A direct realist view of cross-language speech perception", in: Strange, Winifred (ed.). *Speech perception and linguistic experience: issues in cross-language research*. Timonium, MD: York Press, 171-204.

Best, Catherine and Michael Tyler. 2007. "Nonnative and Second-language Speech Perception: Commonalities and Complementarities," in: Munro, Murray and Ocke-Schwen Bohn (eds.). *Second language* speech *learning – the role of language experience in speech perception and production*. Amsterdam: John Benjamins, 13-34.

Bethin, Christina. 1984. "Voicing assimilation in Polish", *International Journal of Slavic Linguistics and Poetics* 29: 17-32.

Bever, Thomas. 1981. "Normal acquisition processes explain the critical period for language learning", in: Diller, Karl C. (ed.), *Individual differences and universals in language learning aptitude*. Rowley: Newbury House Publishers, Inc, 176-199.

Bloch-Rozmej, Anna. 1997. Nasal vowels in Polish. (http://www.ifa.amu.edu.pl/psicl/files/32/08Bloch-Rozmiej.pdf) (date of access: 25 Jan 2021).

Bloomfield, Leonard. [1935] 1996. *Language*. Delhi: Motilal Banarsidass Publishers Private Limited.

Boersma, Paul. 1998. Functional phonology: formalizing the interactions between articulatory and perceptual drives. [Unpublished Phd dissertation, University of Amsterdam.].

Boersma Paul & David Weenink. 2021. *Praat: Doing phonetics by computer* [Computer software]. Version 6.1.50, retrieved 22 Jan 2021 from http://praat.org

Bohn, Ocke-Schwen and Catherine T. Best. 2012. "Native-language phonetic and phonological influences on perception of American English approximants by Danish and German listeners", *Journal of Phonetics* 40: 109-128.

Bohn, Ocke-Schwen. 1995. "Cross-language speech perception in adults: L1 transfer doesn't tell it all", in: Strange, Winifred (ed.), *Speech perception and linguistic experience: theoretical and methodological issues in cross-language speech research*. Timonium: York Press, 275-300.

Bohn, Ocke-Schwen. 2020. "Cross-language phonetic relationships account for most, but not all L2 speech learning problems", in: Wrembel, Magdalena, Agnieszka Kiełkiewicz-Janowiak, and Piotr Gąsiorowski (eds.), *Approaches to the study of sound structure and speech. Interdisciplinary work in honour of Katarzyna Dziubalska-Kołaczyk*. New York/Oxon: Routledge, 171-184.

Bongaerts, Theo, Chantal van Summeren, Brigitte Planken, and Erik Schils. 1997. "Age and ultimate attainment in the pronunciation of a foreign language", *Studies in Second Language Acquisition* 19, 4: 447-465.

Bybee, Joan L. 2013. "Usage-based theory and exemplar representation", in: Hoffman, Thomas and Graeme Trousdale (eds.). *The Oxford Handbook of Construction Grammar*. Oxford: Oxford University Press, 49-69.

Cabrelli, Jennifer, Alicia Luque, and Irene Finestrat-Martinez. 2019. “Influence of L2 English phonotactics in L1 Brazilian Portuguese illusory vowel perception”, *Journal of Phonetics* 73: 55-69.

Carley, Peter and Inger Mees. 2020. *American English phonetics and pronunciation practice*. London: Routledge.

Champagne-Muzar, Cécile, Eta Schneiderman, and Johanne Bourdages. 1993. “Second language accent: the role of the pedagogical environment”, *IRAL: International Review of Applied Linguistics in Language Teaching* 31: 143

Chang, Charles B. 2010. First language phonetic drift during second language acquisition. [Unpublished PhD thesis, University of California, Berkeley.].

Chang, Charles B. 2012. “Rapid and multifaceted effects of second-language learning on first-language speech production”, *Journal of Phonetics* 40, 2: 249-268.

Chang, Charles B. 2013. “A novelty effect in the phonetic drift of the native language”, *Journal of Phonetics* 41: 520–533.

Chang, Charles B. 2019a. “Phonetic drift”, in: Schmid, Monika. S. and Barbara Köpke (eds.). *The Oxford Handbook of Language Attrition* Oxford, UK: Oxford University Press, 191-203.

Chang, Charles B. 2019b. “Language change and linguistic inquiry in a world of multicompetence: sustained phonetic drift and its implications for behavioural linguistic research”, *Journal of Phonetics* 74: 96-113.

Chistovich, Ludmilla A. and Valentina V. Lublinskaya. 1979. “The ‘center of gravity’ effect in vowel spectra and critical distance between the formants: psychoacoustical study of the perception of vowel-like stimuli”, *Hearing Research* 1, 3: 185-195.

Cho, Taehong and Patricia A. Keating. 2001. Articulatory and acoustic studies on domain-initial strengthening in Korean, *Journal of Phonetics* 29: 155-190.

Cho, Taehong and Peter Ladefoged. 1999. “Variation and universals in VOT: evidence from 18 languages”, *Journal of Phonetics* 27: 207-229.

Cho, Taehong and James M. McQueen 2005. “Prosodic influences on consonant production in Dutch: effects of prosodic boundaries, phrasal accent and lexical stress”, *Journal of Phonetics* 33: 121-157

Cho, Taehong, D. H. Walen, and Gerard Docherty. 2018. “Voice onset time and beyond: exploring laryngeal contrasts in 19 languages”, *Journal of Phonetics* 72: 52-65.

Cho, Taehong, Sun-Ah Jun, and Peter Ladefoged. 2002. "Acoustic and aerodynamic correlates of Korean stops and fricatives", *Journal of Phonetics* 20: 193-228.

Chomsky, Noam. 1965. "Persistent topics in linguistic theory", *Diogenes* 13: 13-20.

Chomsky, Noam and Morris Halle. 1968. *The Sound Pattern of English*. New York: Harper & Row.

Cieślicka, Ewa and Arkadiusz Rojczyk. 2017. "Self-reported vs. self-rated pronunciation in a non-native language", *Theory and Practice of Second Language Acquisition* 3: 69-86.

Cohen, Jacob. 1968. "Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit", *Psychological Bulletin* 70, 4: 213-220.

Collins, Beverly and Inger Mees. 2013. *Practical Phonetics and Phonology, A Resource Book for Students. 2nd ed*. London: Routledge.

Common European Framework of Reference for Languages: Learning, Teaching, Assessment. 2011. "Council of Europe Education and Languages". (date of access: 1 May 2019). https://www.coe.int-/EN/web/common-european-frameworkreference-languages/

Couper, G. (2006). "The short and long-term effects of pronunciation instruction", *Prospect* 21: 46-66.

Crosswhite, K., 2001. *Vowel reduction in Optimality Theory*. Routledge, New York.

Cruttenden, Alan. 2014. *Gimson's Pronunciation of English*. London: Arnold.

Cyran, Eugeniusz. 2011. "Laryngeal realism and laryngeal relativism: two voicing systems in Polish?", *Studies in Polish Linguistics* 6: 45-80.

Cyran, Eugeniusz. 2014. *Between phonology and phonetics – Polish voicing*. Berlin: Mouton de Gruyter.

Cyran, Eugeniusz. 2017. "'Voice' languages with no [voice]? Some consequences of Laryngeal Relativism", *Acta Liguistica Academica* 64, 4: 477-511.

Cyran, Eugeniusz. 2020. "Boundaries and typological variation in laryngeal phonology", in: Wrembel, Magdalena, Agnieszka Kiełkiewicz-Janowiak, and Piotr Gąsiorowski (eds.). *Approaches to the study of sound structure and speech: interdisciplinary work in honour of Katarzyna Dziubalska-Kołaczyk*. New York/Oxon: Routledge, 159-171.

Darcy, Isabelle, Hanyong Park, and Chung-Lin Yang. 2015. "Individual differences in L2 acquisition of English phonology: The relation between cognitive abilities and phonological processing", Learning and Individual Differences 40: 63-72.

Davidson, Lisa. 2016. "Variability in the implementation of voicing in American English obstruents", *Journal of Phonetics*, 54: 35-50.

de Leeuw, Esther, Monika S. Schmid, and Ineke Mennen. 2010. "The effects of contact on native language pronunciation in an L2 migrant setting", *Bilingualism: Language and Cognition*, 13: 33-40.

de Leeuw, Esther. 2018. "Phonetic attrition", in: Schmid, Monika. S. and Barbara Köpke (eds.). *The Oxford Handbook of Language Attrition* Oxford, UK: Oxford University Press. (https://www.researchgate.net/publication/323581364_Chapter_18_Phonetic_Attrition) (date of access: 10 Sep 2020.).

de Leeuw, Esther, Ineke Mennen, and James M. Scobbie. 2012. "Singing a different tune in your native language: first language attrition of prosody", *International Journal of Bilingualism* 16, 1: 101-116.

Derwing, Tracey, Murray Munro, and Grace Wiebe. 1997. "Pronunciation instruction for fossilized learners. can it help?", *Applied Language Learning* 8: 217-235.

Díaz, Begoña, Holger Mitterer, Mirjam Broersma, and Núria Sebastián-Gallés. "Individual differences in late bilinguals' L2 phonological processes: From acoustic-phonetic analysis to lexical access", *Learning and Individual Differences* 22, 6: 680-689.

Dixit, R.Prakash and Thomas Shipp. 1985. "Study of subglottal air pressure during Hindi stop consonants", *Phonetica* 42: 53-77.

Dmitrieva, Olga. 2019. "Transferring perceptual cue-weighting from second language into first language: cues to voicing in Russian speakers of English", *Journal of Phonetics* 73: 18-143.

Docherty, G. (1992). *The timing of voicing in British English obstruents*. Berlin: de Gruyer.

Dokovova, Maria. 2015. Bulgarian voicing contrast: at home and abroad. [Unpublished MA thesis, University of Edinburgh].

Donegan, Patricia. 2002. "Phonological processes and phonetic rules", in: Dziubalska-Kołaczyk, Katarzyna and Jarosław Weckwerth (eds.). *Future challenges for natural linguistics*. Munich: Lincom. 57-81.

Donegan, Patricia and David Stampe. 1979. "The study of Natural Phonology", in: Dinnsen, Daniel (ed). *Current approaches to phonological theory*. Bloomington: Indiana University Press, 126-173.

Donegan, Patricia. J. and David Stampe. 2009. "Hypotheses of natural phonology", *Poznań Studies in Contemporary Linguistics* 45, 1: 3-31.

Dörnyei, Zoltán. 2005. *The Psychology of the Language Learner. Individual Differences in Second Language Acquisition*. Routledge: New York.

Dzierla, Jerzy and Geoffrey Schwartz. 2017. Pre-voicing suppression in the speech of Polish learners of English. (Paper presented at Approaches to Phonology and Phonetics (APAP2017), Lublin, 23-25 June 2017.).

Dzierla, Jerzy. 2019. The effects of auditory training on the perception and production of English word-initial laryngeal contrasts by Poles (Paper presented at Approaches to Phonology and Phonetics (APAP2019), Lublin, 21-23 June 2019.).

Dziubalska-Kołaczyk, Katarzyna. 1990. *A theory of second language acquisition within the framework of Natural Phonology*. Poznań: Adam Mickiewicz University Press.

Dziubalska-Kołaczyk, Katarzyna. 2002. *Beats-and-Binding Phonology*. Frankfurt am Main: Peter Lang.

Dziubalska-Kołaczyk, Katarzyna. 2012. "Modern Natural Phonology and phonetics", in: Cyran, Eugeniusz, Henryk Kardela and B. Szymanek (eds.). *Sound, Structure and Sense. Studies in Memory of Edmund Gussmann* Lublin: Wydawnictwo KUL, 199-210.

Dziubalska-Kołaczyk, Katarzyna. 2019. Natural Growth Theory of Acquisition (NGTA): Evidence from (mor)phonotactics. (Paper presented at 3rd ISMBS 2019, Chania, 27-29 August 2019.).

Dziubalska-Kołaczyk, Katarzyna and Magdalena Wrembel. 2017. Natural Growth Model: Explaining third language phonological acquisition. (Paper presented at Societas Linguistica Europaea (SLE 2017), Zürich, 10-13 September 2017.).

Dziubalska-Kołaczyk, Katarzyna and Magdalena Wrembel. 2019. Natural Growth Theory of Acquisition. (Paper presented at The 9th International Symposium on the Acquisition of Second Language Speech, Tokyo, 30 August - 1 September 2019).

Dziubalska-Kołaczyk, Katarzyna and Magdalena Wrembel. 2022. "Natural Growth Theory of Acquisition: A theory of multilingual acquisition of speech", in: Jarosz, Anna and Veronica Serdagna (eds.), *Theoretical and Practical Developments in English Speech Assessment, Research, and Training*. Berlin: Springer.

Ellis, Rod. 2008. "Are classroom and naturalisitic acquisition the same?: a study of the classroom acquisition of German word order rules". *Studies in Second Language Acquisition* 11, 3: 305-328.

Escudero, Paola. 2005. *Linguistic perception and second language acquisition: explaining the attainment of optimal phonological categorization*. LOT Dissertation Series 113: Utrecht University.

Feldstein, Ron F. and Steven Franks. 2002. *Polish*. München: LINCOM Europa.

Flege, James E. 1987. "The production of "new" and "similar" phones in a foreign language: evidence for the effect of equivalence classification", *Journal of Phonetics* 14: 47-65.

Flege, James. 1995. "Second language speech learning: theory, findings, and problems", in: Strange, Winifred (ed.). *Speech Perception and Linguistic Experience: Issues in Cross-language Research*. Timonium, MD: York Press, 233-276.

Flege, James E. 1999. "The relation between L2 production and perception", *Proceedings of the 14th International Congress of Phonetic Sciences in San Francisco*.

Flege, James E. 2005. Origins and development of the Speech Learning Model. (Paper presented at the 1st ASA workshop on L2 speech learning, Vancouver, 14-15 April 2005.).

Flege, James E. 2016. The role of phonetic category formation in second language speech acquisition. (Paper presented at 8th International Conference on Second Language Speech, Aarhus 10-12 June 2016.). (http://www.jimflege.com/files/Flege_Aarhus_talk_20162.pdf) (Date of Access: 3 June 2020).

Flege, James E. and Wendy Eefting. 1987. "Cross-language switching in stop consonant perception and production by Dutch speakers of English", *Speech Communication* 6, 3: 185-202.

Flege, James E., Naoyuki Takagi, and Virginia Mann. 1995. "Japanese adults can learn to produce English /ɹ/ and /l/ accurately", *Language and Speech* 38, 1: 25-55.

Flege, James E., David Birdsong, Ellen Bialystok, Molly Mack, Hyekyung Sung, and Kimiko Tsukada. 2006. "Degree of foreign accent in English sentences produced by Korean children and adults", *Journal of Phonetics* 34: 153-175.

Flege, James E. and Ocke-Schwen Bohn. 2021. "The revised Speech Learning Model (SLM-r)", in: Wayland, Ratree (ed.), Second language speech learning: Theoretical and empirical progress. Cambridge: Cambridge University Press, 3-83.

Fox, Robert and Ewa Jacewicz. 2009. "Cross-dialectal Variation in Formant Dynamics of American English Vowels", *Journal of the Acoustical Society of America* 126: 2603-2618

Giles, Howard, Joustine Coupland, and N Coupland. 1991. "Accommodation Theory: communication, context, and consequence", in Giles, Howard, Justine Coupland and N. Coupland (eds.). *Contexts of Accommodation*. New York: Cambridge University Press.

Golston, Chris and Harry van der Hulst. 1999. "Stricture is structure", in: Hermans, Ben and Marc van Oostendorp (eds.), *The derivational residue in phonological Optimality Theory*. Amsterdam: John Benjamins, 153-173.

Gonet, Wiktor, Jolanta Szpyra-Kozłowska and Radosław Święciński. 2010. "Clashes with Ashes", in: Waniek-Klimczak, Ewa (ed.). *Issues in accents of English 2: variability and norm*. Cambridge: Cambridge Scholars' Publishing, 213-232.

Gorba, Celia. 2019. "Bidirectional influence on L1 Spanish and L2 English stop perception: the role of L2 experience", *The Journal of the Acoustical Society of America* 145, 6: 1-7.

Greenberg, Joseph. 1987. "The present status of markedness theory: a reply to Scheffler", *Journal of Anthropological Research* 43, 4: 367-374.

Grosjean, François. 2004. "Studying bilinguals: Methodological and conceptual issues", *Bilingualism: Language and Cognition* 1: 131-149.

Guion, Susan G. 2003. "The vowel systems of Quichua-Spanish bilinguals. Age of acquisition effects on the mutual influence of the first and second languages", *Phonetica* 60: 98-128.

Gussmann Edmund. 1992. "Resyllabification and delinking: the case of Polish voicing", *Linguistic Inquiry* 23: 29-56.

Gussmann, Edmund. 2007. *The Phonology of Polish*. Oxford: Oxford University Press.

Hale, Mark and Charles Reiss. 2000. "Substance abuse and dysfunctionalism: current trends in phonology". *Linguistic Inquiry* 31: 157-169.

Halle, Morris and Kenneth Stevens. 1971. "A note on laryngeal features", *MIT Quarterly Progress Report* 101: 198-212.

Hallgren, Kevin A. 2012. “Computing inter-rater reliability for observational data: an overview and tutorial”, *Tutororials in quantitative methods for psychology* 8, 1: 23-34.

Hamann, Sielke and Klaas Seinhorst. 2016. Prevoicing in Standard German plosives: implications for phonological representations? (Paper presented at the 13th Old World Conference in Phonology, Budapest, 13-16 January 2016).

Hanson, Helen M. 2009. “Effects of obstruent consonants on fundamental frequency at vowel onset in English”, *The Journal of the Acoustical Society of America* 125, 1: 425-441.

Harada, Tetsuo. 2003. “L2 influence on L1 speech in the production of VOT”. ICPHS Barcelona.

Harris, John. 1994. *English Sound Structure*. Oxford: Blackwell Publishers Ltd.

Harris, John. 2009. “Why final obstruent devoicing is weakening”, in: Nasukawa, Kunya and Paul Backley (eds.), *Strength relations in phonology*. Berlin: Mouton de Gruyter, 9-46.

Haspelmath, Martin. 2013. “Against markedness (and what to replace it with)”, *Journal of Linguistics* 42: 25-70.

Helgason P and Catherine Ringen. 2008. “Voicing and aspiration in Swedish stops”, *Journal of Phonetics* 36, 4: 607-628.

Herd, Wendy, Robin L, Walden, Whitney L. Knight, and Savana Alexander. 2015. “Phonetic drift in a first language dominant environment”, *Proceedings of Meetings on Acoustics* 23.

Hoemeke, Kathryn A. and Randy L. Diehl. 1994. “Perception of vowel height: the role of F1–F0 distance”, *The Journal of the Acoustical Society of America* 96, 2: 661-674.

Honeybone, Patrick. 2005. “Diachronic evidence in segmental phonology: the case of obstruent laryngeal specifications”, in: van Oostendorp , Marc and Jeroen van der Weijer (eds.), *The internal organization of phonological systems*. Berlin: Mouton de Gruyter. 319-354.

Hume, Elizabeth. 2011. “Markedness”, in: van Oostendorp, Mark, C. Ewen & Elizabeth Hume (eds.). *The Blackwell Companion to Phonology*. Hoboken, NJ: Wiley-Blackwell, 79-106.

Hunnicutt, Leigh and Paul A. Morris. 2016. “Prevoicing and Aspiration in Southern American English”, *University of Pennsylvania working papers in linguistics* 22, 1: 1-12.

IBM Corporation. 2020. *IBM SPSS Statistics for Windows, Version 27.1*. Armonk, New York: IBM Corp.

Inkelas, Sharon and Stephanie Shih. 2013. Contour segments and tones in (sub)segmental agreement by correspondence. (Paper presented at the 21st Manchester Phonology Meeting (MFM2012), 23-25 May 2013.).

Inkelas, Sharon and Stephanie Shih. 2016. Re-representing phonology: consequences of Q theory. (http://linguistics.berkeley.edu/~inkelas/Papers/InkelasShihNELS.pdf) (date of access: 17 Dec 2020).

Iverson, G. K., Salmons, J. C. 1995. “Aspiration and laryngeal representation in Germanic”, *Phonology* 12, 3: 369-396.

Jakobson, Roman, Morris Halle, and Gunnar M. Fant. 1951. *Preliminaries to speech analysis: the distinctive features and their correlates*. Cambridge: MIT Press.

Janicka, Katarzyna, Małgorzata Kul, and Jarosław Weckwerth. 2005. “Polish students' attitudes to native English accents”, in: Dziubalska-Kołaczyk, Katarzyna and Joanna Przedlacka (eds.). *English pronunciation models: a changing scene*. Bern: Peter Lang, 251-292.

Jassem, Wiktor. 2003. “Polish”, *Journal of the International Phonetic Association* 33, 1: 103-107

Jibson, Jonathan. 2021. “Assessing merged status with Pillai scores based on dynamic formant contours”, *Proceedings of the Linguistic Society of America* 6, 1: 203.

Jones, L.G. 1956. “English consonantal distribution”, in: Halle, Morris, Hugh Mclean and Horace Lunt (eds.), *For Roman Jakobson: essays on the occasion of his sixtieth birthday, 11 October 1956*, The Hague: Mouton, 245-253.

Jones, Daniel and Michael V. Trofimov. 1923. *The pronunciation of Russian*. Cambridge: Cambridge University Press.

Jongman, Allard and Travis Wade. 2007.“Acoustic variability and perceptual learning”, in Bohn Ocke-Schwen and Murray Munro (eds.), *Language experience in second language speech learning: in honor of James Emil Flege*. Amsterdam, John Benjamins Publishing Company, 135-150.

Kartushina, Natalia, Alexis Hervais-Adelman, Ulrich Hans Fauenfelder, and Narly Golestani. 2016. “Mutual influences between native and non-native vowels in production: evidence from short-term visual articulatory feedback training”, *Journal of Phonetics* 57: 21-39.

Kaye, Jonathan. 2005. “GP, I'll have to put your flat feet on the ground”, in: Broekhuis, Hans, Norbert Corver, Riny Huybregts, Ursula Kleinhenz, and Jan Koster (eds.), *Organizing grammar. Studies in honor of Henk van Riemsdijk*. Berlin: Mouton de Gruyter: 283-288.

Kaye, Jonathan, Jean Lowenstamm, and Jean-Roger Vergnaud. 1985. “The internal structure of phonological elements: a theory of charm and government”, *Phonology Yearbook* 2: 305-328.

Kaźmierski , Kamil. 2009. “On the role of perception in the acquisition of the *peach - pitch* contrast by Polish learners of English”. [Unpublished MA thesis. Adam Mickiewicz University].

Kaźmierski, Kamil, Małgorzata Kul, and Paulina Zydorowicz. 2019. “Educated Poznań speech 30 years later”, *Studia Linguistica Universitatis Iagellonicae Cracoviensis* 136, 4: 245-264.

Keating, Patricia. A. 1979. “A phonetic study of a voicing contrast in Polish” [Upublished PhD dissertation. Brown University.].

Keating, Patricia, Michael Mikos, and William Ganong. 1981. A cross-language study of range of voice onset time in the perception of initial stop voicing, *Journal of the Acoustical Society of America* 70: 1261-1271

Keating, Patricia A. 1984. “Phonetic and phonological representation of stop voicing”, *Language* 60: 286-319

Kello, Christopher P., Gregory G. Anderson, John G. Holden, Guy C. Van Orden. 2008. “The pervasiveness of 1/f scaling in speech reflects the metastable basis of cognition”, *Cognitive Sciences* 32, 7: 1217-1231.

Kim, Donghyun, Meghan Clayards, and Heather Goad. 2018 “A longitudinal study of individual differences in the acquisition of new vowel contrasts”, *Journal of Phonetics* 67: 1-20.

Kim, Mi-Ryoung. 2019. “A study of L1 phonetic drift in the voice onset times of Korean learners of English with long L2 exposure”, *Phonetics and Speech Sciences* 11, 4: 35-43.

Kirby, James and D. Robert Ladd. 2016. “Effects of obstruent voicing on vowel f_0: evidence from ‘true voicing’ languages”, *Journal of the Acoustical Society of America* 140: 2400-2411.

Kirby, James and D. Robert Ladd. 2018. “Effects of obstruent voicing on vowel f_0: implications for laryngeal realism”, *Yearbook of Poznań Linguistic Meeting* 4: 213-235.

Kohler, Klaus 1984. “Phonetic explanation in phonology: the feature fortis/lenis”, *Phonetica* 41: 150-174.

Kopczyński, A. 1977. *Polish and American English consonant phonemes: a contrastive study*. Warszawa: Państwowe Wydawnictwo Naukowe.

Köpke, Barbara and Monika S. Schmid. 2004. “Language attrition: the next phase”, in: Schmid, Monika S., Barbara Köpke, Merel Keijzer, and Lina Weilemar (eds.), *First language attrition: interdisciplinary perspectives on methodological issues*. Amsterdam: John Benjamins, 1-43.

Kretzschmar, Jr, William. 2015. *Language and Complex Systems*. Cambridge: Cabridge University Press.

Kretzschmar, Jr, William and Katarzyna Dziubalska-Kołaczyk. *Submitted*. “Complex natural systems for language”, *Language Sciences*.

Krzysik, Iga and Halina Lewandowska. 2017. “English pronunciation training through the eyes of university graduates”, *Research in Language* 15, 3: 299-312.

Kuhl, Patricia. 2000. “A new view of language acquisition”, in: Cozzarelli, Nicholas R. (ed.) *Preceedings of the national academy of sciences of the United States of America* 97, 22: 11850-11857. (http://www.ncbi.nlm.nih.gov/pmc/articles/PMC34178/) (date of access: 1 Feb. 2016)

Kuzla, Claudia and Mirjam Ernestus. 2011. “Prosodic conditioning of phonetic detail in German plosives”, *Journal of Phonetics* 39: 143-155.

Ladd, D. Robert. 2011. “Phonetics in phonology”, in: Goldsmith, John A., Jason Riggle, and Alan C. L. Yu (eds.), *The handbook of phonological theory* (2nd edition). Malden: Blackwell Publishing, 348-373.

Ladefoged, Peter and Ian Maddieson. 1996. *The sounds of the world’s languages*. Oxford: Blackwell.

Ladefoged, Peter and Keith Johnson. 2010. *A course in phonetics*. Wadsworth Publishing Company.

Ladefoged, Peter. 2001. *Vowels and consonants: an introduction to the sounds of languages*. Oxford: Blackwell.

Lado, Robert. 1957. *Linguistics across cultures; applied linguistics for language teachers*. Ann Arbor: University of Michigan Press.

Landis J. Richard and Gary G. Koch. 1977. "The measurement of observer agreement for categorical data", *Biometrics* 33, 1: 159-174.

Lang, Benjamin and Lisa Davidson. 2017. "Effects of exposure and vowel space distribution on phonetic drift: evidence from American English learners of French". *Language and Speech*: 1-31.

Lenneberg, Eric. 1967. *Biological foundations of language*. New York: Wiley.

Lev-Ari, Shiri and Sharon Peperkamp. 2013. "Low inhibitory skills lead to non-native perception and production in bilinguals' native language", *Journal of Phonetics* 41: 320-331.

Lindsey, Geoffrey. 2012. Funny, old vowels. (Blog post 3. Jun 2012) (https://www.englishspeechservices.com/blog/funny-old-vowels/) (date of access: 1 Jun. 2021).

Lindsey, Geoffrey. 2019. *English after RP: standard British pronunciation today*. Chem: Palgrave Macmillan.

Lisker, Leigh and Arthur S. Abramson. 1964. "A cross-language study of voicing in initial stops: acoustical measurements", *Word* 20, 384-422.

Lombardi, Linda. 1991. Laryngeal features and laryngeal neutralization. [Unpublished PhD dissertation. University of Massachusetts.].

Lombardi, Linda. 1994. *Laryngeal Features and Laryngeal Neutralization*. London: Routledge.

Lombardi, Linda. 1996. "Postlexical rules and the status of privative features", *Phonology* 13, 1: 1-39.

Maddieson, Ian. 1984. *Patterns of sounds*. Cambridge: Cambridge University Press.

Maddieson, Ian. 1997. "Phonetic universals", in: W. Hardcastle & J. Laver (eds.). *The Handbook of Phonetic Sciences*. Oxford: Blackwell, 619-639.

Maddieson, Ian. 2013. "Syllable structure", in: Dryer, Matthew S. & Haspelmath, Martin (eds.) *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at http://wals.info/chapter/12) (date of access 14 Dec 2020).

Major, Roy C. 1992. "Losing English as a first language", *The modern language journal* 76 (2): 190-208.

Malisz, Zofia and Marzena Żygis. 2015. "Voicing in Polish: interactions with lexical stress and focus", *Proceedings of the 18th ICPhS in Glasgow*.

Mayr, Robert, Sacha Price, and Ineke Mennen. "First language attrition in the speech of Dutch-English billinguals: the case of monozygotic twin sisters", *Billingualism: language and cognition* 15, 4: 687-700.

Mennen, Ineke. 2004. "Bi-directional interference in the intonation of Dutch speakers of Greek", *Journal of Phonetics* 32, 4: 543-563.

Mielke, Jeff, Christopher Carigan, and Erik R. Thomas. 2017. "The articulatory dynamics of pre-velar and pre-nasal /æ/-raising in English: an ultrasound study", *Journal of the Acoustical Society of America* 142, 1: 332-349.

Mildner, Vesna and Tomić, Diana. 2007. Effects of phonetic speech training on the pronunciation of vowels in a foreign language. *Proceedings of the 16th International Congress of Phonetic Sciences, Saarbruecken*. 1665-1668.

Mora, Joan C. and Marianna Nadeu. 2012. "L2 effects on the perception and production of a native vowel contrast in early bilinguals", *International Journal f Bilingualism* 16, 4: 484-500.

Morrison, Geoffrey S., and Peter Assmann. 2013. *Vowel Inherent Spectral Change*. Berlin: Springer.

Munro, Murray J., James E. Flege, and Ian R. A. MacKay. 1996. "The effects of age of second language learning on the production of English vowels", *Applied Psycholinguistics* 17: 313-334.

Nearey, Terrance and Peter Assmann. 1986. "Modeling the role of vowel inherent spectral change in vowel identification", *Journal of the Acoustical Society of America* 80: 1297-1308.

Nespor, Marina and Irene Vogel 1986. *Prosodic phonology*. Dordrecht: Foris Publications.

Norris, John N. and Lourdes Ortega. 2001. "Does Type of Instruction Make a Difference? Substantive Findings From a Meta-analytic Review", *Language learning. A journal of research in language studies* 51: 157-213

Oh, Grace E., Susan Guion-Anderson, Katsura Aoyama, James E. Flge, Reiko Akahane-Yamada and Tsuneo Yamada. 2011. "A one-year longitudinal study of English and Japanese vowel production by Japanese adults and children in an English-speaking setting", *Journal of Phonetics* 39: 156-167.

Ohala, John J. 1981. "The listener as a source of sound change", in: Masek, Carrie S., Roberta A. Hendrick, and Mary Frances Miller (eds.), *Papers from the parasession on language and behaviour. Chicago Linguistic Society*. Chicago: Chicago Linguistic Society, 178-203.

Ohala, John J. 1990. "There is not interface between phonology and phonetics: a personal view", *Journal of Phonetics* 18: 153-171.

Ohde, Ralph, N. 1984. "Fundamental frequency as an acoustic correlate of stop consonant voicing", *The Journal of the Acoustical Society of America* 75: 224–230.

Osborne, Denise Maria. 2016. The acquisition of fine phonetic detail in a foreign language: perception and production of stops in L2 English and L1 Portuguese. [Unpublished PhD dissertation. University of Arizona.].

Pajak, Bozena and Roger Levy. 2014. "The role of abstraction in non-native speech perception", *Journal of Phonetics* 46: 147-160.

Pavlenko, Aneta and Scott Jarvis. 2002. "Bidirectional transfer", *Applied Linguistics* 23, 2: 190-214.

Penfield, Wilder and L. Roberts. 1959. *Speech and brain mechanisms*. Princeton: Princeton University Press.

Petrova, Olga, Rosemary Plapp, Catherine Ringen and Szilárd Szentgyörgyi. 2006. "Voice and aspiration: Evidence from Russian, Hungarian, German, Swedish, and Turkish", *The Linguistic Review* 23, 1: 1-35.

Pöchtrager, Markus. 2006. *The structure of length*. [Unpublished PhD Dissertation. University of Vienna.].

Podlipský, Václav Jonáš, Šárka Šimáčková, and Kateřina Chládková. 2020. "Phonetic drift reveals interconnected phonological representations in simultaneous bilinguals: a case study of English and Czech stop consonants", *International Journal of Bilingualism* 6, 3: 271-289.

Polka, Linda and Ocke-Schwen Bohn. 1996. "A cross-language comparison of vowel perception in English-learning and German-leraning infants", *Journal of the Acoustical Society of America* 100, 1: 577-592.

Polka, Linda and Ocke-Schwen Bohn. 2011. “Natural referent vowels guide the development of vowel perception”, *Journal of the Acoustical Society of America* 117 (4): 2398.

Port, Robert F. 2007. “The problems of speech patterns in time”, in: Gaskell, Gareth M. (ed.), *The Oxford handbook of psycholinguistics*. Oxford: Oxford University Press, 503-514.

Prince, Alan and Paul Smolensky. 1993. *Optimality Theory: constraint interaction in generative grammar*. Oxford: Blackwell.

R Core Team. 2018. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna. https://www.R-project.org

Rahimpour, Massoud and Asghar Salimi. 2010. “The impact of explicit instruction on foreign language learners’ performance”, *Procedia Social and Behavioral Sciences*, 1740-1746.

Rice, Keren D. (1993). “A reexamination of the feature [sonorant]”, *Language* 69 (2): 308-344.

Ridouane, Rachid. “On the feature [spread glottis]”. (http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.504.3088&rep=rep1&type=pdf) (date of access: 11 February 2021).

Roach, Peter. 2004. “British English: Received Pronunciation”, *Journal of the International Phonetic Association* 34, 2: 239-245.

Rogers, Derek and Luciana d'Arcangeli. 2004. “Illustrations of the IPA: Italian”, *Journal of the International Phonetic Association* 34. 117-121.

Rojczyk, Arkadiusz. 2010a. “Production and perception of vowel ‘ash’ by Polish learners of English”, in: Dziubalska-Kołaczyk, Katarzyna, Magdalena Wrembel, and Małgorzata Kul (eds.). *Proceedings from the Sixth International Symposium on the Acquisition of Second Language Speech*.

Rojczyk, Arkadiusz. 2010b. “Forming new vowel categories in second language speech: The case of Polish learners' production of English /ɪ/ and /e/”, *Research in Language* 8: 85-97.

Rojczyk, Arkadiusz. 2011. “Perception of the English Voice Onset Time continuum by Polish learners”, in: Arabski, J., Wojtaszek, A. (eds.). *The Acquisition of L2 Phonology*. Bristol, Buffalo, Toronto: Multilingual Matters.

Rojczyk, Arkadiusz. 2012. Phonetic and phonological mode in second-language speech: VOT imitation. (Paper presented at the EuroSLA22 – 22nd Annual Conference of the European Second Language Association, Poznań, 5-8 September 2012.).

Rojczyk, Arkadiusz. 2019. Quality and duration of unstressed vowels in Polish. Lingua 217: 80-89.

Rubach, Jerzy. 1984. *Cyclic and Lexical Phonology: The Structure of Polish*. Dordrecht: Foris.

Rubach, Jerzy. 1996. "Nonsyllabic analysis of voice assimilation in Polish", *Linguistic Inquiry* 27: 69-110.

Rupert, David, MP Wand, and Raymond J. Carroll. 2003. *Semiparametric regression*. Cambridge: Cambridge University Press.

Rydzewski, Paweł. 2017. "A Polish argument for the underlying status of [ɨ]", *Studies in Polish Linguistics* 11 (4). 111-131.

Sampson, Geoffrey. 1980. *Schools of linguistics*. Stanford: Stanford University Press.

Sancier M and C Fowler. 1997. "Gestural drift in a bilingual speaker of Brazilian Portuguese and English", *Journal of Phonetics* 25, 421-437.

Sapir, Edward. 1921. *Language: An introduction to the study of speech*. New York: Harcourt, Brace and Co.

Scheer, Tobias. 2004. *A lateral theory of phonology, vol. 1: What is CVCV, and why should it be?* Berlin: Mouton de Greuter.

Schiering, René, Balthasar Bickel, and Kristine A. Hildrebrandt. 2010. "The prosodic words is not universal but emergent", *Journal of Linguistics* 46: 657-709.

Schuhmann, Katharina S. and Marie K. Huffman. 2015. "L1 drift and L2 category formation in second language learning", *Proceedings of the 18th International Congress of Phonetic Sciences, Glasgow*. (https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0850.pdf) (date of access: 2 September 2020).

Schwartz, Geoffrey. 2013. "A representational parameter for onsetless syllables", *Journal of Linguistics* 49, 3: 613-646.

Schwartz, Geoffrey. 2014. "Which phonetics is phonological?", in: Szpyra-Kozłowska, Jolanta, Ewa Guz, Piotr Steinbrich and Radosław Święciński (eds.). *Recent developments in applied* phonetics. Lublin: Wydawnictwo KUL, 345-371.

Schwartz, Geoffrey. 2015. “Who needs a nucleus? Tashlhiyt Berber syllabification in the Onset Prominence representational environment”, *Poznań Studies in Contemporary Linguistics* 51, 2: 247-290.

Schwartz, Geoffrey. 2016. “On the Evolution of Prosodic Boundaries – Parameter Settings for Polish and English”, *Lingua* 171: 37-74.

Schwartz, Geoffrey. 2017. “Formalizing modulation and the emergence of phonological heads”, *Glossa: a journal of general linguistics* 2, 1: 1-20.

Schwartz, Geoffrey. 2020. “Asymmetrical cross-language interaction – phonological implications”, *Linguistic Approaches to Bilingualism* 12, 2: 103-132.

Schwartz, Geoffrey. 2021. Apart, but where’s the boundary? (Paper presented at the 2nd International Conference on Approaches to Phonology and Phonetics (APAP2021), Lublin, 25-26.06.2021.).

Schwartz, Geoffrey, Anna Balas, and Arkadiusz Rojczyk. 2014. “Stop release in Polish English – implications for Prosodic Constituency”, *Research in Language* 12, 2: 131-144.

Schwartz, Geoffrey and Cormac Anderson. 2015. Consonant mutation and onset structure in the history of Irish. (Paper presented at the 2nd International Conference on Approaches to Phonology and Phonetics (APAP2015), Lublin, 26-28.06.2015.).

Schwartz, Geoffrey and Ewelina Wojtkowiak. 2017. Asymmetries in L2-induced phonetic drift in Polish. (Paper presented at the 11th conference on native and non-native accents of English (Accents 2017), Łódź, 30 November-2 December 2017.).

Schwartz, Geoffrey and Daria Arndt. 2018. “Laryngeal realism vs. modulation theory – evidence from VOT discrimination in Polish”, *Language Sciences* 69: 98-112.

Schwartz, Geoffrey, and Jerzy Dzierla. 2018. “Polish listeners’ perception of vowel inherent spectral change in L2 English.” *Poznań Studies in Contemporary Linguistics* 54 (2): 307-332

Schwartz, Geoffrey, Jerzy Dzierla and Ewelina Wojtkowiak. 2018. Why is aspiration easier to produce for L2 learners than lenis stops without pre-voicing? (Paper presented at the 12th conference on native and non-native accents of English (Accents 2018), 29 November-1 December 2018.).

Schwartz, Geoffrey and Kamil Kaźmierski. 2019. "Vowel dynamics in the acquisition of L2 English – an acoustic study of L1 Polish learners", *Language Acquisition* 27, 3: 227-254.

Schwartz, Geoffrey, Ewelina Wojtkowiak, and Bartosz Brzoza. 2019. "Beyond VOT in the Polish laryngeal contrast", *Proceedings of the 19th International Congress of Phonetic Sciences. Melbourne, Australia*. (https://icphs2019.org/icphs2019-full-papers/pdf/full-paper_256.pdf) (date of access: 21 June 2021).

Schwartz, Geoffrey, Ewelina Wojtkowiak, and Jerzy Dzierla. 2020. "Laryngeal phonology and asymmetrical cross-language phonetic influence", in: Wrembel, Magdalena, Agnieszka Kiełkiewicz-Janowiak, and Piotr Gąsiorowski (eds.). *Approaches to the study of sound structure and speech: interdisciplinary work in honour of Katarzyna Dziubalska-Kołaczyk*. New York/Oxon: Routledge, 316-326.

Selinker, Larry. 1972. "Interlanguage", *International Review of Applied Linguistics in Language Teaching* 10: 209-241.

Shepard, Carolyn. A., Giles, H., & Le Poire, Beth. 2001. "Communication accommodation theory", in: W. P. Robinson & H. Giles (eds.), *The new handbook of language and social psychology*. New York: Wiley, 33-56.

Shih, Stephanie and Sharon Inkelas. 2014. "A subsegmental correspondence approah to contour tone (dis)harmony patterns", in: Kingston, John, Claire Moore-Cantwell, Joe Pater, and Amanda Rysling (eds.), *Proceedings of the 2013 Meeting on Phonology*. Washington, DC: Linguistic Society of America.

Shockey, Linda. 1984. "All in a flap: long-term accommodation in phonology", *International Journal of the Sociology of Language* 46: 87-95.

Simon, Ellen. 2009. "Acquiring a new second language contrast: an analysis of the English laryngeal system of first language Dutch speakers", *Second Language Research* 25, 3: 377-408.

Sobkowiak, Włodzimierz. 2002. "English speech in Polish eyes: What university students think about English pronunciation teaching and learning", *Accents and speech in teaching English phonetics and phonology* 177-193.

Sobkowiak, Włodzimierz. 2003. *English phonetics for Poles*. Poznań: Bene Nati.

Stampe, David. 1979. *A Dissertation on Natural Phonology*. New York: Garland.

Steriade, Donca. 1993. "Closure, release, and nasal contours", *Phonetics and Phonology* 5: 401-470.

Stevens, Kenneth and Dennis Klatt. 1974. “Role of formant transitions in the voiced-voiceless distinction for stops”, T*he Journal of the Acoustical Society of America* 55: 653–659.

Stolberg, Doris and Alexandra Münch. 2010. “Die Muttersprache vergisst man nicht - or do you? A case study in L1 attrition and its (partial) reversal”, *Bilingualism* 13, 19-31.

Strycharczuk, Patrycja. 2012. Phonetics-phonology Interactions in Pre-Sonorant Voicing. [Unpublished PhD dissertation. University of Manchester.].

Strycharczuk, Patrycja and Peter Jurgec. 2008. Prosodic influences on formant frequencies of Polish vowels. (Paper presented at The 3rd Newcastle Postgraduate Conference in Theoretical and Applied Linguistics, Newcastle, 9 Jul 2008). [https://personalpages.manchester.ac.uk/staff/patrycja.strycharczuk/CV_files/2008.7%20Newcastle%20Handout%20w_%20Pata%20v.2.pdf] [date of access: 21 Jan 2021]

Sučková, Magda. 2018. First language attrition and maintenance in the accents of native speakers of English living in the Czech Republic. (Paper presented at the 12th conference on native and non-native accents of English (Accents 2018), Łódź, 29 November-1 December 2018.).

Sypiańska, Jolanta. 2016. “L1 vowels of multilinguals: the applicability of SLM in multilingualism”, *Research in Language* 14, 1:79-94.

Sypiańska, Jolanta. 2017. *Cross-linguistic influence in bilinguals and multilinguals*. Poznań: Wydawnictwo Naukowe UAM.

Sypiańska, Jolanta and Zuzanna Cal. 2019. Perceptual drift in L1 phonetic categories in multilinguals. (Paper presented at the 13th conference on native and non-native accents of English (Accents 2019), Łódź, 6-8 December 2019.).

Syrdal Ann K., and Hundrai S. Gopal. 1986. “A perceptual model of vowel recognition based on the auditory representation of American English vowels”, *Journal of the Acoustical Society of America* 79, 4: 1086–1100

Szigetvári, Péter. 1996. “Laryngeal contrasts and problematic representations”, *The Even Yearbook* 2: 97-110.

Tice, Marisa and Melinda Woodley. 2012. “Paguettes & Bastries: novice French learners show shifts in native phoneme boundaries”, *UC Berkeley PhonLab Annual Repor*t (https://escholarship.org/uc/item/9h18t4rz) (retrieved: 20 October 2020).

Tobin, Stephen J., Hosung Nam and Carol A. Fowler. 2017. “Phonetic drift in Spanish-English bilinguals: experiment and self-organizing model”, *Journal of Phonetics* 65: 45-59.

Tobin, Yishai. 2006. “Structuralist Phonology: Prague School”, in: Brown, Keith (ed.), *Encyclopedia of language & linguistics*. Vol. 12 (2nd edition.) Oxford: Elsevier, 170-177.

Traunmüller, Helmut. 1994. “Conventional, biological, and environmental factors in speech communication: a modulation theory”, *Phonetica* 51: 170-183.

Trubetzkoy, Nikolai Sergeyevich. [1939] 1962. *Grundzüge der Phonologie* [Principles of phonology]. (Translated by Christiane A. M. Baltaxe). Berkley: University of California Press.

van der Hulst, Harry. 2015. “The laryngeal class in RcvP and voice phenomena in Dutch”, in: Caspers, J., Y. Chen, W. Heeren, J. Pacilly, N. Schiller & E. van Zanten (eds.), *Above and beyond the segments*. Amsterdam: John Benjamins Publishing Company, 323-349

van Leussen, Jan-Willem and Paola Escudero. 2005. “Learning to perceive and recognize a second language: the L2LP model revised”, *Frontiers in Psychology* 6 (1000).

Vaughn, Charlotte, Melissa Baese-Berk, and Kaori Idemaru. 2019. “Re-examining phonetic variability in native and non-native speech”, *Phonetica* 76: 327-358.

Vaux, Bert and Bridget Diane Samuels. 2005. “Laryngeal markedness and aspiration”, *Phonology* 22, 3: 395-436.

Wade, Travis, Allard Jongman, and Joan Sereno. 2007. “Effects of acoustic variability in the perceptual learning of non-native- accented speech sounds”, *Phonetica* 64: 122-144.

Waniek-Klimczak, Ewa. 1997. “Context for teaching English phonetics and phonology at Polish universities and colleges: a survey”, in: Waniek-Klimczak, Ewa and Patrick J. Melia (eds.) *Accents and Speech in Teaching English Phonetics and Phonology: EFL Perspective*. Frankfurt: Peter Lang, 5-17.

Waniek-Klimczak, Ewa. 2005. *Temporal parameters in second language speech: an applied linguistic phonetics approach*. Łódź: Wydawnictwo Uniwersytetu Łódzkiego.

Waniek-Klimczak, Ewa. 2011. “Aspiration in Polish: a sound change in progress?”, in: Pawlak, Mirosław and Jakub Bielak (eds.). *New perspectives in language, discourse, and translation studies*. Berlin: Springer.

Waniek Klimczak, Ewa, Arkadiusz Rojczyk, and Andrzej Porzuczek. 2015. “‘Polglish’ in Polish eyes: what English studies majors think about their pronunciation in English”, *Teaching and researching the pronunciation of English*, 23-34.

Wardaugh, Ronald. 1970. The contrastive analysis hypothesis. (Paper presented at the 4th TFSOL Convention in San Francisco, San Francisco, 18-21 March 1970.).

Weckwerth, Jarosław. 2011. “English TRAP vowel in advanced polish learners: variation and system typology”, *Proceedings of the 17th International Congress of Phonetic Sciences*, 2110-2113.

Weckwerth, Jarosław and Anna Balas. 2020. “Selected aspects of Polish vowel formants”, in: Wrembel, Magdalena, Agnieszka Kiełkiewicz-Janowiak, and Piotr Gąsiorowski (eds.). *Approaches to the study of sound structure and speech: interdisciplinary work in honour of Katarzyna Dziubalska-Kołaczyk*. New York/Oxon: Routledge, 338-349.

Weinreich, Uriel. 1953. *Languages in contact: findings and problems*. New York: Linguistic Circle of New York.

Wells, John C. 1982. *Accents of English: Volume 3*. Cambridge: Cambridge University Press.

Wells, John. 2010. Lexical sets. (Blog post, 1 Feb 2010). (http://phonetic-blog.blogspot.com/2010/02/lexical-sets.html) (date of access: 8 October 2020).

Westbury, John and Patricia Keating. 1986. “On the naturalness of stop consonant voicing”, *Journal of Linguistics* 22, 1: 145-166.

Wetzels, W. Leo and Joan Mascaró. 2001. “The typology of voicing and devoicing”, *Language* 77, 2: 207-244.

Wickham, Hadley. 2016. *ggplot2. Elegant graphics for data analysis*. New York: Springer.

Winter, Bodo. 2015. “The other n: the role of repetitions and items in the design of phonetic experiments”, *Proceedings of the 18th International Congress of Phonetic Sciences, Glasgow*.

Witteman, Marijt J., Andrea Weber, and James M. McQueen. 2014. “Tolerance for inconsistency in foreign-accented speech”, *Psychonom Bull Rev* 21: 512-519.

Wojtkowiak, Ewelina. 2016. Features in phonology. [Unpublished MA thesis, Adam Mickiewicz University.].

Wojtkowiak, Ewelina. 2020. “Prosody-segment interaction in the acoustics of Polish front vowels”, *Studies in Polish Linguistics* 15, 3: 151-175.

Wojtkowiak, Ewelina and Geoffrey Schwartz. 2018. “*Sandhi*-voicing in dialectal Polish – prosodic implications”, *Studies in Polish Linguistics* 13, 2: 123-143.

Wojtkowiak, Ewelina and Geoffrey Schwartz. 2022. “The acoustics of plosives and the formation of prosodic structure in Polish”, *to appear* in *Folia Linguistica*.

Wrembel, Magdalena. 2011. “Cross-linguistic influence in third language acquisition of voice onset time”, *Proceedings of the 17th International Congress of Phonetic Sciences*: 2157-2160.

Xie, Xin and T. Florian Jaeger. 2020. “Comparing non-native and native speech: Are L2 productions more variable?”, *Journal of the Acoustical Society of America* 147: 3322.

Yang, Jing and Robert A. Fox. 2017. “L1-L2 interactions of vowel systems in young bilingual Mandarin-English children”, *Journal of Phonetics* 65: 60-76.

Yao, Yao and Charles B. Chang. 2016. “On the cognitive basis of contact-induced sound change: vowel merger reversal in Shanghainese”, *Language* 92, 2: 433-467.

Yu, Vickie, Luc De Nil, and Elizabeth Pang. 2015. “Effects of age, sex and syllable number on voice onset time: Evidence from children’s voiceless aspirated stops”, *Language and Speech*, 58, 2: 152-167

Yusa, Noriaki, Kuniya Nasukawa, Masatoshi Koizumi, Kim Jungho, Naoki Kimura, and Kensuke Emura. 2010. “Unexpected effects of the second language on the first”, in: Dziubalska-Kołaczyk, Katarzyna, Magdalena Wrembel, and Małgorzata Kul. *Proceedings of the 6th International Symposium on the Acquisition of Second Language Speech, New Sounds 2010*, 580-584.

Zając, Magdalena. 2015. Phonetic convergence in the speech of Polish learners of English. [Unpublished doctoral dissertation. University of Łódź.].

Zampini, Massimiliano. 1998. “The relationship between the production and perception of L2 Spanish stops”, *Texas Papers in Foreign Language Education*, 3, 3: 85-100.

Appendix 1: The complete list of Polish sentences

The Polish sentence list is presented below; U-I indicates an utterance-initial position; P-I a phrase-initial position, while P-M a phrase medial one. Fillers are not included.

TARGET WORD	VOWEL	TYPE	SENTENCE	ENGLISH TRANSLATION
bawić [bavit͡ɕ] 'to play/to entertain' (inf.)	[a]	U-I	W tym momencie się uczymy. **Bawić** możemy się po szkole.	Right now we're studying. **Playing** can wait till after school.
		P-I	Jeśli teraz się uczymy, **bawić** będziemy się po lekcjach.	If we're studying right now, we'll **play** after school.
		P-M	Na tym szkoleniu uczymy **bawić** publiczność dowcipami.	At this training we're learning how to **entertain** the audience with some jokes.
bywać [bɨvat͡ɕ] 'being' (ger.)	[ɨ]	U-I	Tak dziś postanowiliśmy. **Bywać** w kinie będziemy w piątki.	That's what we decided on today. **Being** in the cinema will be a Friday tradition.
		P-I	Choć tak postanowiliśmy, **bywać** w domu częściej jest trudno.	Even though that's what we decided on, **being** at home more often is difficult.
		P-M	Zgodnie postanowiliśmy **bywać** w pracy także w soboty.	We all agreed on **being** at work also on Saturdays.
bochen [bɔxɛn] 'a loaf' (nom. sg.)	[ɔ]	U-I	Codziennie go wypiekamy. **Bochen** żytniego chleba jest smaczny.	We bake it every day. A **loaf** of rye bread is tasty.
		P-I	Zawsze gdy go wypiekamy, **bochen** pszennego chleba jest chrupki.	Whenever we bake it, the **loaf** of wheat bread is crunchy.
		P-M	Od pokoleń wypiekamy **bochen** wiejskiego chleba na mące.	For generations, we've been baking a **loaf** of rustic bread.
befsztyk [bɛfʂtɨk] 'a beef-stake' (nom. sg.)	[ɛ]	U-I	Nieraz go tu zamawiamy. **Befsztyk** wołowy to ich specjalność.	We often order this. The **beefsteak** is their specialty.
		P-I	Zawsze gdy go zamawiamy, **befsztyk** tatarski jest wysmażony.	Every time we order it, the **beefsteak** is well-done.
		P-M	Tradycyjnie zamawiamy **befsztyk** wieprzowy z ziemniakami.	As per usual we order a **beefsteak** with potatoes.
dawać [davat͡ɕ] 'to give' (inf.)	[a]	U-I	Naprawdę to uwielbiamy. **Dawać** prezent jest bardzo przyjemnie.	We really love this. **Giving** [someone] a gift is very nice.
		P-I	Jakkolwiek to uwielbiamy, **dawać** kwiatów nie można codziennie.	However much we love this, **giving** flowers every day is unseemly.

		P-M	My najbardziej uwielbiamy **dawać** książki z okazji urodzin.	We love **giving** people books for birthdays the most.
dywan [dɨvan] 'a carpet' (nom. sg.)	[ɨ]	U-I	Niezbyt często go trzepiemy. **Dywan** zwykle tylko odkurzamy.	We don't beat it very often. The **carpet** is usually just vacuumed.
		P-I	Zawsze kiedy go trzepiemy, **dywan** kurzy się na wszystkie strony.	Every time we beat it, the **carpet** spews out dust everywhere.
		P-M	Dwa razy w roku trzepiemy **dywan** małą trzepaczką z wikliny.	Twice a year we beat the **carpet** using a small wicker carpet-beater.
dowód [dɔvut] 'proof' (nom. sg.)	[ɔ]	U-I	Niezwłocznie go przedstawimy. **Dowód** wpłaty wyjaśni sprawę.	We'll present it at once. The **proof** of payment will explain everything.
		P-I	Gdy tylko go przedstawimy, **dowód** zbrodni obciąży sprawcę.	As soon as we present it, the **proof** will incriminate the perpetrator.
		P-M	Na rozprawie przedstawimy **dowód** winy oskarżonego.	During the trail we'll present the **proof** that incriminates the perpetrator.
defekt [dɛfɛkt] 'a glitch' (nom. sg.)	[ɛ]	U-I	W warsztacie go usuniemy. **Defekt** silnika nie jest poważny.	We'll fix this in the shop. The **glitch** in the engine isn't serious.
		P-I	Jeśli go nie usuniemy, **defekt** czujnika wywoła spięcie.	If we don't fix it, the **glitch** in the sensor will cause short circuit.
		P-M	Na przeglądzie usuniemy **defekt** przewodu paliwowego.	During the inspection will fix the **glitch** in the fuel line.
gafy [gafɨ] ' blunder' (nom. pl.)	[a]	U-I	Niekiedy je popełniamy. **Gafy** czasem nam się przydarzają.	We make them from time to time. **Blunders** happen every so often.
		P-I	Kiedy już je popełniamy, **gafy** warto po prostu przemilczeć.	If we actually make them, **blunders** are best left unmentioned.
		P-M	Nieświadomie popełniamy **gafy** podczas spotkań towarzyskich.	Unbeknownst to us we make **blunders** during social occasions.
gyros [gɨrɔs] 'gyros' (nom. sg.)	[ɨ]	U-I	Specjalnie go przyrządzamy. **Gyros** z papryką jest wyśmienity.	We're making it specifically for this occasion. **Gyros** with pepper is delicious.
		P-I	Zawsze gdy go przyrządzamy, **gyros** z indyka smakuje gościom.	Every time we make it, turkey **gyros** is appreciated by the guests.
		P-M	W każdy wtorek przyrządzamy **gyros** z frytkami i warzywami.	Every Tuesday we make **gyros** with chips and veggies.
gofry [gɔfrɨ] 'a waffle' (nom. pl.)	[ɔ]	U-I	Od jutra je sprzedajemy. **Gofry** z jogurtem to u nas nowość.	We're starting the sale tomorrow. **Waffles** with yogurt are our novelty.
		P-I	Od kiedy je sprzedajemy, **gofry** z kokosem robią furorę.	Ever since we started selling them, **waffles** with coconut have been a sensation.

		P-M	W naszym punkcie sprzedajemy **gofry** z polewą czekoladową.	Here we sell **waffles** with chocolate sauce.
geny [gɛnɨ] 'a gene' (nom. pl.)	[ɛ]	U-I	Zawodowo je badamy. **Geny** roślin to nasza działka.	It's our job to study them. The **genes** of plants is are our specialty.
		P-I	Dlatego iż je badamy, **geny** ptaków nie są nam obce.	Since we study them, the **genes** of birds are familiar to us.
		P-M	W naszym ośrodku badamy **geny** ludzi chorych na raka.	In this centre we study the **genes** of cancer patients.
pawła [pavwa] 'Paul' (gen. sg.)	[a]	U-I	Wszyscy bardzo go lubimy. **Pawła** Lisa znamy od gimnazjum.	We all like him very much. **Paul** Lis [we've] known since middle school.
		P-I	Chociaż bardzo go lubimy, **Pawła** Wilka nie cierpi nasz sąsiad.	Even though we like him very much, **Paul** Wilk is not liked by our neighbour.
		P-M	Chyba najbardziej lubimy **Pawła** Sowę za jego dowcipy.	It's likely that our favourite thing about **Paul** Sowa is his sense of humour.
pychę [pɨxɛw̃] 'pride' (acc. sg.)	[ɨ]	U-I	Stanowczo je potępiamy. **Pychę** i butę trzeba piętnować.	We strongly condemn this. **Pride** and arrogance has to be stigmatised.
		P-I	Jakkolwiek je potępiamy, **pychę** i zazdrość trudno jest zwalczyć.	Although we condemn them, **pride** and envy is difficult to fight.
		P-M	Jednoznacznie potępiamy **pychę** i zawiść w każdym człowieku.	We unequivocally condemn **pride** and envy in every person.
powód [pɔvut] 'a cause' (nom. sg.)	[ɔ]	U-I	Niebawem go ustalimy. **Powód** śmierci wreszcie będzie znany.	We'll establish it soon. The **cause** of death will be known.
		P-I	Kiedy już go ustalimy, **powód** zgonu będzie oczywisty.	Once we establish it, the **cause** of death will be obvious.
		P-M	W krótkim czasie ustalimy **powód** złego stanu gospodarki.	In a very short time we'll establish the **cause** of the poor state of the economy.
pewność [pɛvnɔɕt͡ɕ] 'certainty' (nom. sg.)	[ɛ]	U-I	Wnikliwie ją oceniamy. **Pewność** teorii trzeba potwierdzić.	We're assessing it carefully. The **certainty** of this theory needs to be verified.
		P-I	Ilekroć ją oceniamy, **pewność** założeń nie jest wysoka.	Whenever we evaluate it, the **certainty** of these assumptions is not very high.
		P-M	W przybliżeniu oceniamy **pewność** istnienia życia w kosmosie.	We're roughly assessing the **certainty** of extraterrestrial life.
taflę [taflɛw̃] 'a tile' (acc. sg.)	[a]	U-I	Starannie ją polerujemy. **Taflę** szklaną należy wygładzić.	We're polishing it carefully. The glass **tile** needs smoothing.
		P-I	Zanim ją polerujemy, **taflę** lustra najpierw oczyszczamy.	Before we polish it, the glass **tile** needs to be cleaned.
		P-M	W przerwie meczu polerujemy **taflę** lodu całego lodowiska.	During the break we polish the **tile** of the entire ice rink.

tyfus [tɨfus] ‘typhus’ (nom. sg.)	[ɨ]	U-I	Dziś skutecznie go leczymy. **Tyfus** został już opanowany.	Nowadays we successfully treat it. **Typhus** has been contained.
		P-I	Choć skutecznie go leczymy, **tyfus** nadal budzi przerażenie.	Although we can successfully treat it, **typhus** still evokes fear.
		P-M	W dzisiejszych czasach leczymy **tyfus** dzięki antybiotykowi.	Nowadays we treat **typhus** with antibiotics.
towar [tɔvar] ‘goods’ (nom. sg.)	[ɔ]	U-I	Stale u nich kupujemy. **Towar** z przemytu jest bardzo tani.	We always buy from them. The **goods** that come from smugglers are very cheap.
		P-I	Ilekroć go kupujemy, **towar** z importu jest uszkodzony.	Whenever we buy them, the [imported] **goods** are damaged.
		P-M	Na bieżąco kupujemy **towar** z hurtowni po niskich cenach.	We buy cheap **goods** from wholesalers on regular basis.
teflon [tɛflɔn] ‘teflon’ (nom. sg.)	[ɛ]	U-I	W tym celu go stosujemy. **Teflon** zmniejsza przywieranie potraw.	This is why we use it. **Teflon** reduces the risk of the food sticking to the pan.
		P-I	W miarę jak go stosujemy, **teflon** łatwiej da się zarysować.	In the cause of its utilisation, **teflon** runs the risk of getting scratched.
		P-M	Przy produkcji stosujemy **teflon** jako powłokę patelni.	In the process of production we use **teflon** as the pan’s coating.
kawior [kaviɔr]	[a]	U-I	Niezbyt często go jadamy. **Kawior** z jesiotra jest rarytasem.	We rarely eat it. Sturgeon **caviar** is a rarity.
		P-I	Zawsze kiedy go jadamy, **kawior** z łososia nieźle smakuje.	Whenever we eat it, salmon **caviar** tastes good.
		P-M	Podczas bankietów jadamy **kawior** z mintaja jako przystawkę.	At banquets we eat pollock fish **caviar** as a starter.
kęsy [kɛw̃sɨ] ‘a bite’ (nom. pl.)	[ɛ]	U-I	Powoli je przeżuwamy. **Kęsy** chleba popijamy wodą.	We chew them slowly. **Bites** of bread are washed down with water.
		P-I	W chwili gdy je przeżuwamy, **kęsy** ciasta się rozpływają.	As we chew them, the **bites** of pie melt in our mouths.
		P-M	Bez pośpiechu przeżuwamy **kęsy** pizzy i rozmawiamy.	We chew the **bites** of pizza without rush when talking.
kowno [kɔvnɔ] ‘Kaunas’ (nom. sg.)	[ɔ]	U-I	Cały piątek je zwiedzamy. **Kowno** było stolicą Litwy.	We have been seeing the sights all of Friday. **Kaunas** used to be the capital of Lithuania.
		P-I	Zawsze kiedy je zwiedzamy, **Kowno** robi na nas wrażenie.	Every time we see the sights, **Kaunas** makes an impression.
		P-M	Według programu zwiedzamy **Kowno** zaraz po meczu w Wilnie.	According to the programme we’re sightseeing **Kaunas** right after the game in Vilnius.
kefir [kɛfir] ‘kefir’ (nom. sg.)	[ɛ]	U-I	Nawet chętnie go pijemy. **Kefir** smakuje nie najgorzej.	We drink it relatively willingly. **Kefir** doesn’t taste that bad.
		P-I	Zawsze kiedy go pijemy, **kefir** mieszamy z owocami.	Whenever we drink it, we mix **kefir** with fruit.
		P-M	Każdego ranka pijemy **kefir** smakowy do śniadania.	Every day we drink flavoured **kefir** for breakfast.

Appendix 2: Polish word list

The words are ordered alphabetically, followed by their IPA transcription, the translation into English, and their grammatical form. An asterisk (*) indicates a filler word. The words are given in the alphabetical order; they were pseudo-randomised (the same for each participant) when the study was conducted.

1. *bać*: /batɕ/, ‘to fear’ (inf.)
2. *bak*: /bak/, ‘a tank’ (nom. sg.)
3. *bal*: /bal/, ‘a ball’ (nom. sg.)
4. *bar*: /bar/, ‘a bar’ (nom. sg.)
5. *bas*: /bas/, ‘bass’ (nom. sg.)
6. *bat*: /bat/, ‘a whip’ (nom. sg.)
7. *bel*: /bɛl/, ‘a log’ (gen. pl.)
8. *bez*: /bɛs/, ‘a lilac’ (nom. sg.)
9. *blat**: /blat/, ‘a counter’ (nom. sg.)
10. *blok**: /blɔk/, ‘a block’ (nom. sg.)
11. *bluz**: /blus/, “blouses’ (gen. pl.)
12. *brat**: /brat/, ‘a brother’ (nom. sg.)
13. *daj*: /daj/, ‘to give’ (imp.)
14. *dam*: /dam/, ‘ladies’ (gen. pl.)
15. *dane*: /danɛ/, ‘data’ (nom. pl.)
16. *dar*: /dar/, ‘a gift’ (nom. sg.)
17. *data*: /data/, ‘date’ (nom. sg.)
18. *demo*: /dɛmɔ/, ‘a demo’ (nom. sg.)
19. *doba*: /dɔba/, ‘24 hours’ (nom. sg.)
20. *dom*: /dɔm/, ‘a house’ (nom. sg.)
21. *drap**: /drap/, ‘to scratch’ (imp.)
22. *dres**: /drɛs/, ‘a tracksuit’ (nom. sg.)
23. *drops**: /drɔps/, ‘a candy’ (nom. sg.)
24. *dryl**: /drɨl/, ‘a training’ (nom. sg.)
25. *flesz**: /flɛʂ/, ‘flesh’ (nom. sg.)
26. *fluid**: /fluit/, ‘a foundation’ (nom. sg.)

27. *frajer**: /frajɛr/, ‘a nincompoop’ (nom. sg.)
28. *front**: /frɔnt/, ‘front’ (nom. sg.)
29. *gad*: /gat/, ‘a reptile’ (nom. sg.)
30. *gafa*: /gafa/, ‘a blunder’ (nom. sg.)
31. *gaj*: /gaj/, ‘a grove’ (nom. sg.)
32. *gam*: /gam/, ‘a range’ (gen. pl.)
33. *gapa*: /gapa/, ‘a giddy goat’ (nom. sg.)
34. *gasi*: /gaɕi/, ‘to quench’ (3rd. sg.)
35. *gaz*: /gas/, ‘gas’ (nom. sg.)
36. *gen*: /gɛn/, ‘a gene’ (nom. sg.)
37. *gest*: /gɛst/, ‘a gesture’ (nom. sg.)
38. *glob**: /glɔp/, ‘a globe’ (nom. sg.)
39. *gram**: /gram/, ‘a gram’ (nom. sg.)
40. *grant**: /grant/, ‘a grant’ (nom. sg.)
41. *grill**: /gril/, ‘a grill’ (nom. sg.)
42. *grób**: /grup/, ‘a tombstone’ (nom. sg.)
43. *kac*: /kats/, ‘a hangover’ (nom. sg.)
44. *kara*: /kara/, ‘a punishment’ (nom. sg.)
45. *kat*: /kat/, ‘an executioner’ (nom. sg.)
46. *kawa*: /kawa/, ‘coffee’ (nom. sg.)
47. *kaźń*: /kaʑɲ/, ‘an execution’ (nom. sg.)
48. *keks*: /kɛks/, ‘a fruitcake’ (nom. sg.)
49. *klacz**: /klatʂ/, ‘a mare’ (nom. sg.)
50. *klej**: /klɛj/, ‘glue’ (nom. sg.)
51. *kler**: /klɛr/, ‘clergy’, (nom. sg.)
52. *koc*: /kɔts/, ‘a blanket’ (nom. sg.)
53. *kos*: /kɔs/, ‘a blackbird’ (nom. sg.)
54. *kot*: /kɔt/, ‘a cat’ (nom. sg.)
55. *krab**: /krap/, ‘a crab’ (nom. sg.)
56. *kraj**: /kraj/, ‘a country’ (nom. sg.)
57. *mak*: /mak/, ‘a poppy’ (nom. sg.)
58. *mam*: /mam/, ‘I have’ (1st. sg.)
59. *mapa*: /mapa/, ‘a map’ (nom. sg.)

60. *mara*: /mara/, ‘a ghoul’ (nom. sg.)
61. *mas*: /mas/, ‘masses’ (gen. pl.)
62. *mata*: /mata/, ‘a mat’ (nom. sg.)
63. *maź*: /maɕ/, ‘grease’ (nom. sg.)
64. *moc*: /mɔts/, ‘power’ (nom. sg.)
65. *molo*: /mɔlɔ/, ‘a pier’ (nom. sg.)
66. *mop*: /mɔp/, ‘a mop’ (nom. sg.)
67. *nać*: /natɕ/, ‘leaves’ (nom. sg.)
68. *nam*: /nam/, ‘us’ (dat.)
69. *nawa*: /nava/, ‘a nave’ (nom. sg.)
70. *noc*: /nɔts/, ‘night time’ (nom. sg.)
71. *noga*: /nɔga/, ‘a leg’ (nom. sg.)
72. *nos*: /nɔs/, ‘a nose’ (nom. sg.)
73. *pak*: /pak/, ‘a pack’ (aug.; gen. pl.)
74. *pan*: /pan/, ‘Sir’ (nom. sg.)
75. *para*: /para/, ‘a couple’ (nom. sg.)
76. *pas*: /pas/, ‘a belt’ (nom. sg.)
77. *pejs*: /pɛjs/, ‘payot’ (nom. sg.)
78. *pet*: /pɛt/, ‘coll. a cigarette butt’ (nom. sg.)
79. *plan**: /plan/, ‘a plan’ (nom. sg.)
80. *polo*: /pɔlɔ/, ‘polo’
81. *pop*: /pɔp/, ‘pop music’ (nom. sg.)
82. *prom**: /prɔm/, ‘a ferry’ (nom. sg.)
83. *pub*: /pap/, ‘a pub’ (nom. sg.)
84. *skan**: /skan/, ‘a scan’ (nom. sg.)
85. *skocz**: /skɔtʂ/, ‘to jump’ (imp.)
86. *skup**: /skup/, ‘to focus’ (imp.)
87. *slang**: /slaŋk/, ‘slang language’ (nom. sg.)
88. *slogan**: /slɔgan/, ‘a slogan’ (nom. sg.)
89. *smak**: /smak/ ‘a taste’ (nom. sg.)
90. *smok**: /smɔk/, ‘a dragon’ (nom. sg.)
91. *SPA**: /spa/, ‘a SPA resort’ (nom. sg.)
92. *spam**: /spam/, ‘spam’ (nom. sg.)

93. *spić**: /spitɕ/, ‘to get drunk’ (inf.)
94. *spiker**: /spikɛr/, ‘a speaker’, (nom. sg.)
95. *spinacz**: /spinatʂ/, ‘a clip’ (nom. sg.)
96. *sport**: /spɔrt/, ‘a sport’ (nom. sg.)
97. *stadium**: /stadjum/, ‘a stage’ (nom. sg.)
98. *stan**: /stan/, ‘a state’ (nom. sg.)
99. *starczy**: /startʂɨ/, ‘senile’ (adj.)
100. *step**: /stɛp/, ‘step aerobics’ (nom. sg.)
101. *stok**: /stɔk/, ‘a slope’ (nom. sg.)
102. *stół**: /stuw/, ‘a table’ (nom. sg.)
103. *stop**: /stɔp/, ‘a stop’ (nom. sg.)
104. *stopper**: /stɔpɛr/, ‘a stop watch’ (nom. sg.)
105. *studio**: /studjɔ/, ‘a studio’ (nom. sg.)
106. *styl**: /stɨl/, ‘style’ (nom. sg.)
107. *tabu*: /tabu/, ‘a taboo’ (nom. sg.)
108. *taca*: /tat͡sa/, ‘a tray’ (nom. sg.)
109. *tak*: /tak/, ‘yes’
110. *tara*: /tara/, ‘tare’ (nom. sg.)
111. *targ*: /tark/, ‘a market’ (nom. sg.)
112. *tas*: /tas/, ‘to shuffle [the cards]’ (acc. sg.)
113. *tek*: /tɛk/, ‘a briefcase’ (aug.; gen. pl.)
114. *ten*: /tɛn/, ‘this’ (dem.)
115. *test*: /tɛst/, ‘a test’ (nom. sg.)
116. *toga*: /tɔga/, ‘a robe/gown [worn by solicitors]’ (nom. sg.)
117. *ton*: /tɔn/, ‘tone’ (nom. sg.)
118. *traktor**: /traktɔr/, ‘a tractor’ (nom. sg.)
119. *trend**: /trɛnt/, ‘a trend’ (nom. sg.)
120. *trup**: /trup/, ‘a corpse’ (nom. sg.)
121. *tryb**: /trɨp/, ‘mode’ (nom. sg.)

Appendix 3: The complete list of English sentences

The sentences are ordered as follows: a. Utterance-initial; b. Phrase-initial (utterance-medial), and c. Phrase-medial. Fillers are not included.

CONSO-NANT:	/p/	KEY-WORD:	SENTENCE:
VOWEL:	[ɪ]	*pity*	a. *I don't want you to feel that way.* ***<u>Pity</u>*** *is definitely not needed.* b. *Out of all the things you can feel,* ***<u>pity</u>*** *must be the worst feeling.* c. *It isn't out of character for her to throw a* ***<u>pity</u>*** *party for herself.*
	[æ]	*packing*	a. *I don't want you to feel that way.* ***<u>Pity</u>*** *is definitely not needed.* b. *Although I love travelling,* ***<u>packing</u>*** *still is my least favourite thing to do.* c. *Personally I don't think* ***<u>packing</u>*** *an hour before leaving is a good idea.*
	[e]	*penguins*	a. *These are hard times for animals.* ***<u>Penguins</u>*** *might soon be extinct.* b. *Some animals are natural enemies,* ***<u>penguins</u>*** *and seals being an example.* c. *I'm hoping to spot some* ***<u>penguins</u>*** *today at the Zoo.*
	[ʌ]	*pubbing*	a. b. *While I don't like going to clubs,* ***<u>pubbing</u>*** *is my favourite pastime.* c. *I don't think we should go* ***<u>pubbing</u>*** *on a Monday night.*

CONSO-NANT:	/b/	KEY-WORD:	SENTENCE:
VOWEL:	[ɪ]	*billiards*	a. *It is quite popular.* ***<u>Billiards</u>*** *is mostly played in bars.* b. *When we don't know what to do on Friday,* ***<u>bil-liards</u>*** *is always a good idea.* c. *It is really easy to play* ***<u>billiards</u>*** *and lose a lot of money.*
	[æ]	*battles*	a. *It didn't matter who was going to die.* ***<u>Battles</u>*** *were fought to be won.* b. *With so many people wanting power,* ***<u>battles</u>*** *were difficult to avoid.*

			c. *There are very different kinds of **battles** that people fight nowadays.*
	[e]	*benches*	a. *It isn't clear why people do it. **Benches** in parks are often demolished.* b. *If the chairs are occupied, **benches** are also available for the audience.* c. *Senior citizens can sit on **benches** in all parks in our city.*
	[ʌ]	*butter*	a. *You know I'm a vegan. **Butter** is not something I eat.* b. *Should you prefer it, **butter** can be used instead.* c. *Vegan diet replaces **butter** with margarine.*

CONSONANT:	/t/	KEYWORD:	SENTENCE:
VOWEL:	[ɪ]	*tickets*	a. *It is an amazing show. **Tickets** are already sold out.* b. *Before you begin your journey, **tickets** must be validated.* c. *It's a pity we forgot our **tickets** and couldn't see the play.*
	[æ]	*talent*	a. *Both go pretty much hand in hand. **Talent** and luck are equally important.* b. *While it's very important, **talent** is sometimes not enough.* c. *I think there are too many **talent** shows on TV nowadays.*
	[e]	*tennis*	a. *I watch many disciplines. **Tennis** is definitely my favourite.* b. *Although they look similar, **tennis** and badminton are very different.* c. *I don't know how to play **tennis** but I like watching others play.*
	[ʌ]	*tunnels*	a. *When I am driving, **tunnels** are quite terrifying.* b. *I thought I wasn't claustrophobic. **Tunnels** proved me wrong.* c. *I actually really loved the **tunnels** that I visited in Naples.*

CONSONANT:	/d/	KEYWORD:	SENTENCE:
VOWEL:	[ɪ]	*dinner*	a. *Some people say it's breakfast. **Dinner** is equally important, though.* b. *If you call me when you leave work, **dinner** will be ready by the time you're home.* c. *I don't know what you want for **dinner** but I can think of something.*

	[æ]	*damage*	a. *You don't have to apologise to me.* ***Damage*** *has already been done.* b. *In some kinds of accidents,* ***damage*** *can be mostly internal.* c. *We can't really say how much* ***damage*** *the rains have caused yet.*
	[e]	*dentists*	a. *The new toothpaste is good.* ***Dentists*** *really recommend it.* b. *While I dislike doctors in general,* ***dentists*** *are absolutely awful.* c. *I really don't like those* ***dentists*** *who are very mean to you.*
	[ʌ]	*dubbing*	a. *I myself prefer subtitles.* ***Dubbing*** *is very annoying.* b. *When the actors are good,* ***dubbing*** *can actually be cool.* c. *There are people who prefer* ***dubbing*** *but I'm not one of them.*

CONSONANT:	/k/	KEY-WORD:	SENTENCE:
VOWEL:	[ɪ]	*kitchens*	a. *We want more people to spend time there.* ***Kitchens*** *should be quite roomy.* b. *For people to enjoy cooking,* ***kitchens*** *should have big windows.* c. *The atmosphere in some* ***kitchens*** *is really quite dramatic.*
	[æ]	*camera*	a. *You dream of hearing these words. "****Camera****, action", that's my biggest goal.* b. *According to some people,* ***camera*** *adds five pounds.* c. *For my birthday I got a new* ***camera*** *and I couldn't be happier.*
	[e]	*kelvin*	a. *I understand Celcius.* ***Kelvin*** *is more difficult.* b. *While not used by the general public,* ***Kelvin*** *is important for physicists.* c. *We needed to convert it from* ***Kelvin*** *into Celcius to understand it.*
	[ʌ]	*culture*	a. *It's not only the legends or the food.* ***Culture*** *includes also the language.* b. *Reading about it is one thing,* ***culture*** *must be experienced, though.* c. *I enjoy getting to know the* ***culture*** *of the country I'm going to visit.*

CONSONANT:	/g/	KEY-WORD:	SENTENCE:
VOWEL:	[ɪ]	*giving*	a. *You can stop doing it.* ***Giving*** *me a hard time is boring.*

			b. *While employers forget about it,* ***giving*** *your workers a break is important.* c. *I truly believe that* ***giving*** *is better than receiving.*
	[æ]	*gather*	a. *I'm going to tell you a tale.* ***Gather*** *round children and we'll begin.* b. *Should it be necessary,* ***gather*** *more people that can help us.* c. *I think people are starting to* ***gather*** *near the crime scene.*
	[e]	*guessing*	a. *You should know this by now.* ***Guessing*** *shouldn't even be an option.* b. *If you don't know the answer,* ***guessing*** *might be your only choice.* c. *It's not a surprise that people started* ***guessing*** *which one of the answers was correct.*
	[ʌ]	*gunshots*	a. *Not many people survive this.* ***Gunshots*** *are very often fatal.* b. *In some foreign countries,* ***gunshots*** *can be heard daily.* c. *What we just heard was either* ***gunshots*** *or fireworks outside.*

Appendix 4: English word list

The words are ordered alphabetically. An asterisk (*) indicates a filler word.

1. *back*
2. *bar*
3. *beg*
4. *bell*
5. *Ben*
6. *bet*
7. *blues**
8. *brat**
9. *buck*
10. *bug*
11. *bus*
12. *but*
13. *claire**
14. *clay**
15. *clutch**
16. *come*
17. *crab**
18. *cry**
19. *cub*
20. *cup*
21. *cut*
22. *dead*
23. *debt*
24. *deck*
25. *Depp*
26. *die*
27. *does*
28. *done*
29. *drab**
30. *dress**
31. *drill**
32. *drops**
33. *duck*
34. *dutch*
35. *flesh**
36. *fluid**
37. *frank**
38. *front**
39. *fryer**
40. *get*
41. *gets*
42. *glob**
43. *gram**
44. *grant**
45. *grill**
46. *group**
47. *guess*
48. *guest*
49. *gull*
50. *gum*
51. *gun*
52. *gut*
53. *keg*
54. *kelp*
55. *Kent*
56. *kept*
57. *meant**
58. *Meg**
59. *men**
60. *mess**
61. *met**
62. *money**
63. *much**
64. *muck**
65. *mud**
66. *mug**
67. *mum**
68. *neck**
69. *nest**
70. *net**
71. *nets**
72. *none**
73. *nub**
74. *null**
75. *nun**
76. *nut**
77. *pace*
78. *pass*
79. *peck*
80. *peg*
81. *pen*
82. *pet*
83. *plan**
84. *pop*

85. *prom**
86. *pub*
87. *puck*
88. *pug*
89. *pup*
90. *scan**
91. *scoop**
92. *scotch**
93. *shrank**
94. *slang**
95. *sleepy**
96. *slogan**
97. *smack**
98. *smart**
99. *smock**
100. *spa**
101. *spam**
102. *speaker**
103. *speech**
104. *spend**
105. *spinach**
106. *sport**
107. *stadium**
108. *stan**
109. *starch**
110. *step**
111. *still**
112. *stock**
113. *stool**
114. *stop**
115. *stopper**
116. *studio**
117. *tech*
118. *Ted*
119. *tell*
120. *ten*
121. *test*
122. *ton*
123. *tractor**
124. *trap**
125. *trend**
126. *trick**
127. *trim**
128. *trip**
129. *troupe**
130. *tub*
131. *tuck*
132. *tug*

Appendix 5: Questionnaire

8.4. Polish version given to the participants

KWESTIONARIUSZ OSOBOWY

Imię i nazwisko:

Grupa:

Nr uczestnika:

Informacje dotyczące języka polskiego:

1. Z jakiego regionu Polski Pan/Pani pochodzi? ____________________________________
2. Czy może Pan/Pani powiedzieć, że w Pana/Pani stronach rodzinnych/w Pana/Pani rodzinie używa się gwary? TAK/NIE
   a. jeśli TAK, to jakiej? ____________________________
   b. jeśli TAK, jak Pan/Pani to zauważa? (inne słownictwo, gramatyka, wymowa?) _______
3. Czy któreś z Pana/Pani rodziców pochodzi zza granicy? TAK/NIE
   a. jeśli TAK, skąd? ____________________________

Informacje dotyczące nauki języka angielskiego:

1. Ile miał/a Pan/Pani lat, gdy zaczął/zaczęła się Pan/Pani uczyć języka angielskiego? _______
2. Ile lekcji języka angielskiego tygodniowo miał Pan/Pani w liceum? _________
3. Czy lekcje języka angielskiego w liceum poruszały zagadnienia związane z wymową? (Np. nauka transkrypcji fonetycznej, warsztaty poświęcone poprawnej wymowie danych dźwięków, osobne zajęcia z fonetyki) TAK/NIE
   a. jeśli TAK, to jaką formę one przyjmowały? ______________________________
4. Czy odwiedził/a Pan/Pani któreś z państw, gdzie angielski jest językiem urzędowym? TAK/NIE
   a. jeśli TAK, to jaki(e)? __________________________________
   b. ile czasu spędził/a Pan/Pani w krajach angielskojęzycznych? (razem, w tygodniach/miesiącach) ______________________________
   c. ile trwał najdłuższy wyjazd? __________________________

Informacje dotyczące innych języków:

1. Czy zna Pan/Pani inne języki? TAK/NIE
   a. jeśli TAK, to jakie? (wymienić) ____
   b. który z tych języków jest wg. Pana/Pani najsilniejszy? (w którym jest Pan/Pani najpłynniejszy/a, oprócz angielskiego) ______________________________
   c. Czy zdawał/a Pan/Pani z któregoś z nich maturę? ____________________________

d. Na jaki lektorat języka obcego w ramach studiów na filologii angielskiej Pan/Pani uczęszcza? __________

i. Poziom? _______

2. Czy wyjeżdżał/a Pan/Pani do krajów nieanglojęzycznych na dłużej niż 3 tygodnie?

TAK/NIE

a. jeśli TAK, jaki był to kraj? _________________________________

Dziękuję!

DANE OSOBOWE PANA/PANI I INFORMACJE, KTÓRE ZOSTAŁY TU PODANE SŁUŻĄ TYLKO I WYŁĄCZNIE CELOM NAUKOWYM. BĘDĄ DOBRZE PRZECHOWYWANE A WGLĄD DO NICH BĘDZIE MIAŁA TYLKO OSOBA PRZEPROWADZAJĄCA OBECNE BADANIE.

8.5. English translation

QUESTIONNAIRE

Name:

Group:

Participant number:

Information regarding Polish:

1. Which area of Poland do you come from? ______

2. Would you say that you use a dialect at home, in interactions with your family? YES/NO

a) if YES, then which dialect? ___

b) if YES, what area do you notive this in? (lexicon, grammar, pronunciation?) ____

3. Is any of your parents from abroad? YES/NO

a) if YES, where are they from? ______

Information regarding your English acquisition:

1. How old were you when you first started learning English? ____

2. How many English classes per week did you have in high school? ____

3. Did you have any pronunciation training at school? (e.g. learning transcription, workshops on how to properly produce English sounds, separate phonetics classes?) YES/NO

a) if YES, what form did these take? ____

4. Have you ever visited any English-speaking countries? YES/NO

a) if YES, which ones? _____

b) how much time did you spend in English speaking countries (in total, in weeks or months) _____________

c) how long was your longest stay in an English-speaking country? ______

Information regarding other languages:

1. Do you know any other languages? YES/NO

a) if YES, which ones? (enumerate them) _______

b) which one of these languages do you think is your strongest one (aside from English) _____

c) did you take your Matura exam in any of them? _____

d) which foreign language class are you taking currently? _____

i) what is its level? ____

2. Have you ever stayed in non-English speaking countries for longer than three consecutive weeks? YES/NO

a) if YES, which country was that?

Thank you!

YOUR PERSONAL INFORMATION PROVIDED IN THIS QUESTIONNAIRE SERVE SOLELY SCIENTIFIC PURPOSE, WILL BE KEPT ANONUMOUS AND WILL BE VIEWED ONLY BY THE PERSON CONDUCTING THE PRESENT STUDY.

Appendix 6: Wellsian lexical sets

Below please find John Wells' lexical sets (after Wells 1982, 2010) for the two English accents models (General British and General American) which are the targets in the phonetic training our Faculty's students undergo.

Keyword	Symbol (General British)	Symbol (General American)	Example words
KIT	[ɪ]	[ɪ]	sit, dip
FLEECE	[iː]	[i]	seat, deep
DRESS	[e]	[e]	set, bed
TRAP	[æ]	[æ]	got,
LOT	[ɒ]	[ɑ]	stop, odd
STRUT	[ʌ]	[ʌ]	but, mud
FOOT	[ʊ]	[ʊ]	good, book
BATH	[ɑː]	[æ]	laugh
NURSE	[ɜː]	[ɝ]	curse, girl
FACE	[eɪ]	[eɪ]	base, maze
PALM	[ɑː]	[ɑ]	balm, father
THOUGHT	[ɔː]	[ɔ]	taught, caught
GOAT	[əʊ]	[oʊ]	soap, coal
GOOSE	[uː]	[uː]	shoot, tooth
PRICE	[aɪ]	[aɪ]	like, write
CHOICE	[ɔɪ]	[ɔɪ]	boy, noise
MOUTH	[aʊ]	[aʊ]	out, house
NEAR*	[ɪə]	[ɪɚ]/[ɪr]	beer, here
SQUARE*	[eə]	[eɚ]/[er]	care, share
START	[ɑː]	[ɑr]	bar, part
NORTH	[ɔː]	[ɔr]	war, storm
FORCE	[ɔː]	[or]	more, horse
CURE*	[ʊə]	[ʊɚ]/[ʊr]	pure, tour
happY	[i]	[i]	lucky, ready
commA	[ə]	[ə]	visa, panda
lettER	[ə]	[ɚ]	paper, teacher

Appendix 7: GLMM results – full coefficient tables

Below please find the full coefficient tables of the GLMM models described in Chapters 6 and 7.

WORD READING: L1 POLISH.
Target: VOT of /p, t, k/; fixed effect: session/group. The Intercept is Monolinguals (i.e. Group 4).

					95% Confidence Interval	
Model term	Coefficient	Std. Error	t	Significance	Lower	Upper
Intercept	40.306	4.1351	9.747	.000	32.198	48.414
Group 1: T1	.607	3.2490	.187	.852	-5.764	6.977
Group 1: T2	2.710	3.2492	.834	.404	-3.661	9.081
Group 1: T3	1.120	3.2496	.345	.730	-5.251	7.492
Group 2	2.742	3.4727	.790	.430	-4.067	9.551
Group 3	3.667	3.4727	1.056	.291	-3.142	10.476

WORD READING: L1 POLISH.
Target: VOT of /p, t, k/; main predictor: place of articulation*session/group. The Intercept is Monolinguals*/t/ (i.e. Group 4).

					95% Confidence Interval	
Model term	Coefficient	Std. Error	t	Significance	Lower	Upper
Intercept	35.692	3.3579	10.629	.000	29.108	42.276
Group 1: T1*p	-1.409	4.4446	-.317	.751	-10.124	7.306
Group 1: T1*t	-.325	3.3766	-.096	.923	-6.946	6.296
Group 1: T1*k	19.904	4.5762	4.349	.000	10.931	28.877
Group 1: T2*p	-.960	4.4469	-.216	.829	-9.679	7.760
Group 1: T2*t	2.772	3.3762	.821	.412	-3.848	9.392
Group 1: T2*k	22.403	4.5766	4.895	.000	13.429	31.376
Group 1: T3*p	-2.081	4.4449	-.468	.640	-10.797	6.634
Group 1: T3*t	1.110	3.3773	.329	.742	-5.512	7.732
Group 1: T3*k	20.477	4.5787	4.472	.000	11.499	29.455
Group 2*p	1.558	4.6276	.337	.736	-7.516	10.631
Group 2*t	1.334	3.6069	.370	.711	-5.738	8.407
Group 2*k	21.799	4.7549	4.585	.000	12.476	31.122
Group 3*p	2.392	4.6312	.516	.606	-6.689	11.473
Group 3*t	2.487	3.6083	.689	.491	-4.588	9.562
Group 3*k	22.551	4.7521	4.746	.000	13.233	31.869
Group 4*p	-1.391	3.2211	-.432	.666	-7.707	4.925
Group 4*k	17.727	3.3681	5.263	.000	11.123	24.331

WORD READING: L1 POLISH.

Target: VOT of /b, d, g/; main interaction: voicing_type*session/group. The intercept is monolinguals*Type_3 (Type_2 and Type_3 rows are not included as these were not reported on in the thesis).

					95% Confidence Interval	
Model term	Coefficient	Std. Error	t	Significance	Lower	Upper
Intercept	22.436	20.9638	1.595	.111	-7.672	74.544
Group 1: T1*Type 1	-101.31	15.5417	-6.519	<.001	-131.768	-70.895
Group 1: T2*Type 1	-105.072	15.5495	-6.757	<.001	-135.786	-74.581
Group 1: T3*Type 1	-100.070	15.5503	-6.435	<.001	-130.563	-69.578
Group 2*Type 1	-94.642	15.6774	-6.037	<.001	-125.384	-63.901
Group 3*Type 1	-113.199	15.6863	-7.216	<.001	-141.958	-82.440
Group 4*Type 1	-112.796	14.8405	-7.601	<.001	-141896	-83.695

WORD READING: L1 POLISH.

Target: VOT of /b, d, g/; main interaction: voicing_type*POA*session/group. The intercept is monolinguals*Type_3*d (Type_2 and Type_3 rows are not included as these were not reported on in the thesis).

					95% Confidence Interval	
Model term	Coefficient	Std. Error	t	Significance	Lower	Upper
Intercept	48.122	32.5157	1.480	.139	-15.638	111.882
Group 1: T1*b*Type_1	-135.978	47.2262	-2.879	.004	-228.584	-43.372
Group 1: T1*d* Type_1	-100.433	21.4338	-4.686	<.001	-142.462	-58.403
Group 1: T1*g* Type_1	-122.683	47.2249	-2.598	.009	-215.287	-30.080
Group 1: T2*b* Type_1	-139.855	47.2280	-2.961	.003	-232.464	-47.246
Group 1: T2*d *Type_1	-102.138	21.4537	-4.761	<.001	-144.207	-60.070
Group 1: T2*g* Type_1	-128.322	47.2386	-2.716	.007	-220.952	-35.691
Group 1: T3*b* Type_1	-133.996	47.2305	-2.837	.005	-226.610	-41.381
Group 1: T3*d* Type_1	-97.812	21.4557	-4.559	<.001	-139.884	-55.739
Group 1: T3*g* Type_1	-123.474	47.2382	-2.614	.009	-216.104	-30.845
Group 2*b* Type_1	-132.304	47.2841	-2.798	.005	-225.023	-39.584

Group 2*d* Type_1	-90.752	21.5899	-4.203	<.001	-133.087	-48.416
Group 2*g* Type_1	-115.259	-47.3037	-2.437	.015	-208.017	-22.501
Group 3*b* Type_1	-152.774	47.2991	-3.230	.001	-245.523	-60.025
Group 3*d* Type_1	-106.879	21.5781	-4.953	<.001	-149.191	-64.566
Group 3*g* Type_1	-135.829	47.3008	-2.872	.004	-228.581	-43.077
Group 4*b* Type_1	-144.568	46.9949	-3.076	.002	-236.720	-52.415
Group 4*d* Type_1	-113.291	20.9198	-5.415	<.001	-154.313	-72.269
Group 4*g* Type_1	-136.051	46.9928	-2.895	.004	-228.199	-43.903

WORD READING: L1 POLISH.
Target: Type_3 (yes/no); fixed factor: session/group. The intercept is Monolinguals.

					95% Confidence Interval	
Model term	Coefficient	Std. Error	t	Significance	Lower	Upper
Intercept	5,269	.7735	6.812	<.001	3.752	6.786
Group 1: T1	-1.283	.8556	-1.500	.134	-2.961	.395
Group 1: T2	3.147	.8175	-3.849	<.001	-4.750	-1.544
Group 1: T3	-3.151	.8175	-3.854	<.001	-4.754	-1.548
Group 2	-3.981	.8269	-4.814	<.001	-5.602	-2.359
Group 3	-3.300	.8343	-3.955	<.001	-4.936	-1.664

SENTENCE READING: L1 POLISH.
Target: VOT of /p, t, k/; main interaction: Prosodic_position*group/session, speech rate. The intercept is Monolinguals*phrase_medial.

					95% Confidence Interval	
Model term	Coefficient	Std. Error	t	Significance	Lower	Upper
Intercept	32.644	2.4380	13.390	.000	27.864	37.424
Group 1 T1*U-I	-3.079	1.4481	-2.126	.034	-5.919	-.240
Group 1 T1*P-I	-2.428	1.4468	-1.678	.093	-5.265	.408
Group 1 T1*P-M	-.765	1.4561	-.525	.600	-3.619	2.090
Group 1 T2*U-I	-1.940	1.4453	-1.342	.180	-4.774	.894
Group 1 T2*P-I	-1.372	1.4422	-.951	.342	-4.199	1.456
Group 1 T2*P-M	-.447	1.4552	-.307	.759	-3.300	2.406
Group 1 T3*U-I	-2.690	1.4483	-1.858	.063	-5.530	.149
Group 1 T3*P-I	-1.788	1.4448	-1.237	.216	-4.620	1.045
Group 1 T3*P-M	-1.096	1.4583	-.751	.452	-3.955	1.763
Group 2*U-I	1.075	1.5676	.686	.493	-1.998	4.149
Group 2*P-I	.177	1.5680	.113	.910	-2.897	3.251

Group 2*P-M	.339	1.5754	.215	.830	-2.750	3.428
Group 3*U-I	-.834	1.5642	-.533	.594	-3.900	2.233
Group 3*P-I	-.241	1.5629	-.154	.878	-3.305	2.823
Group 3*P-M	1.445	1.5766	.916	.360	-1.647	4.536
Group 4*U-I	1.227	.6647	1.846	.65	-.077	2.530
Group 4*P-I	1.701	.6622	2.569	.10	.403	3.000
Speech rate	-.620	.2381	-2.606	.009	-1.087	-.154

SENTENCE READING: L1 POLISH.

Target: VOT of /p, t, k/; main interaction: Prosodic_position*group/session*Voicing_type, speech rate. The intercept is Monolinguals*phrase_medial*Type_3 (Type_2 and Type_3 rows are not included as these were not reported on in the thesis).

					95% Confidence Interval	
Model term	Coefficient	Std. Error	t	Significance	Lower	Upper
Intercept	16.257	11.3371	1.434	.152	-5.971	38.484
Group 1 T1*U-I	-97.551	10.7218	-8.539	.000	-112.572	-70.530
Group 1 T1*P-I	-91.833	10.7144	-9.131	.000	-118.839	-76.826
Group 1 T1*P-M	-90.188	10.7028	-8.427	.000	-111.171	-69.204
Group 1 T2*U-I	-92.677	10.7153	-8.649	.000	-113.685	-71.668
Group 1 T2*P-I	-94.274	10.7294	-8.786	.000	-115.310	-73.238
Group 1 T2*P-M	-89.329	10.7126	-8.339	.000	-110.332	-68.326
Group 1 T3*U-I	-89.172	10.7231	-8.316	<.001	-110.195	-68.148
Group 1 T3*P-I	-92.560	10.7218	-8.633	.000	-113.581	-71.539
Group 1 T3*P-M	-90.738	10.7145	-8.469	.000	-111.745	-69.731
Group 2*U-I	-86.368	10.8312	-7.974	<.001	-106.978	-64.556
Group 2*P-I	-90.959	10.8488	-8.384	.000	-112.229	-69.689
Group 2*P-M	-85.767	10.8187	-7.928	<.001	-106.978	-64.556
Group 3*U-I	-85.071	10.3528	-8.217	<.001	-105.369	-64.774
Group 3*P-I	-94.472	10.3860	-9.096	.000	-114.835	-74.109
Group 3*P-M	-89.940	10.3589	-8.586	.000	-109.250	-68.631
Group 4*U-I	-121.530	10.7240	-11.332	.000	-142.556	-100.504
Group 4*P-I	-129.857	10.7307	-12.101	.000	-150.896	-108.819
Speech rate	-128.530	10.7240	-11.985	.000	-149.556	-107.505

SENTENCE READING: L1 POLISH

Target: Type 3 (yes/no); fixed factor: session/group. The intercept is Monolinguals.

					95% Confidence Interval	
Model term	Coefficient	Std. Error	t	Significance	Lower	Upper
Intercept	19.930	40.19	.050	.960	-768	807
Group 1: T1	-14.737	40.19	-.037	.971	-802	773
Group 1: T2	-15.389	40.19	-.038	.969	-803	772
Group 1: T3	-15.513	40.19	-.039	.969	-803	772
Group 2	-16.181	40.19	-.040	.968	-804	772
Group 3	-15.208	40.19	-.038	.970	-803	772

WORD READING: L1 POLISH

Target: Type 2 (yes/no); fixed factor: Group/session. The intercept is Monolinguals.

					95% Confidence Interval	
Model term	Coefficient	Std. Error	t	Significance	Lower	Upper
Intercept	3.941	.5702	6.912	<.001	2.823	5.060
Group 1: T1	-1.030	.6837	-1.506	.123	-2.370	.311
Group 1: T2	-1.510	.6799	-2.221	.026	-2.843	-.177
Group 1: T3	-1.583	.6788	-2.332	.020	-2.914	-.252
Group 2	-1.437	.7093	-2.027	.043	-2.828	-.047
Group 3	-2.014	.7022	-2.866	.004	-3.392	-.636

SENTENCE READING: L1 POLISH

Type 2 (yes/no); fixed factor: Group/session.

					95% Confidence Interval	
Model term	Coefficient	Std. Error	t	Significance	Lower	Upper
Intercept	2.354	.3763	6.254	<.001	1.616	3.091
Group 1: T1	.098	.4875	.201	-.858	-.858	1.054
Group 1: T2	-.131	.4853	-.269	-1.082	-1.082	.821
Group 1: T3	-.026	.4859	-.053	-.978	-.978	.927
Group 2	-.162	.5140	-.315	-1.170	-1.170	.846
Group 3	-5.29	.5103	-1.036	-1.529	-1.529	.472

WORD READING: L1 Polish

Target: pitch; main interaction Vocing_type*Vowel_quality. The intercept is nasal onsets.

					95% Confidence Interval	
Model term	Coefficient	Std. Error	t	Significance	Lower	Upper
Intercept	2.072	.3609	5.741	<.001	1.365	2.780
Type_1*a	.000	.0139	-.03	.976	-.028	.027
Type_1*ε	.013	.5043	.025	.980	-.976	1.001
Type_3*a	.043	.0247	1.746	.081	-.005	.092
Type_3* ε	.051	.5056	.010	.920	-.940	1.042
Type_4*a	.120	.0139	8.620	.000	.092	.147
Type_4* ε	.163	.5042	.324	.746	-.825	1.152

WORD READING: L1 Polish

Target: F1 (f1-f0); main interaction: Voicing_type*Vowel_quality. The intercept is Type_4* ε

					95% Confidence Interval	
Model term	Coefficient	Std. Error	t	Significance	Lower	Upper
Intercept	4.299	.5357	8.024	<.001	3.249	5.349
Type_1*a	.078	.7535	.104	.918	-1.399	1.555
Type_1*ε	-.755	.0450	-16.780	.000	-.843	-.666
Type_3*a	.061	.7547	.081	.936	-1.419	1.541

Type_3* ɛ	-.803	.0895	-8.975	.000	-.979	-.628
Type_4*a	1.131	.7535	1.501	.133	-.346	2.608

SENTENCE READING: L1 Polish

Target: pitch; main interaction: Voicing_type*Vowel_quality. The intercept is Type_4*ɨ.

					95% Confidence Interval	
Model term	Coefficient	Std. Error	t	Significance	Lower	Upper
Intercept	2.612	.1567	16.663	.000	2.304	2.919
Type_1*a	-.343	.2064	-1.662	.097	-.748	.062
Type_1*ɛ	-.307	.2064	-1.490	.136	-.712	.097
Type_1*ɨ	-.246	.0138	-17.801	.000	-.273	-.219
Type_3*a	-.401	.2084	-1.925	.054	-.810	.007
Type_3* ɛ	-.365	.2093	-1.744	.081	-.776	.045
Type_3*ɨ	-.235	.0372	-6.332	<.001	-.308	-.163
Type_4*a	-.193	.2063	-.934	.350	-.597	.212
Type_4* ɛ	-.098	.2063	-.476	.634	-.503	.306

SENTENCE READING: L1 POLISH

Target: F1 (F1-f0); main interaction: Voicing_type*Vowel_quality

					95% Confidence Interval	
Model term	Coefficient	Std. Error	t	Significance	Lower	Upper
Intercept	1.926	.0812	23.723	.000	1.767	2.085
Type_1*a	1.626	.0853	19.066	.000	1.459	1.793
Type_1*ɛ	.965	.0852	11.325	.000	.798	1.132
Type_1*ɨ	.243	.0344	7.055	<.001	.175	.310
Type_3*a	1.880	.1118	16.817	.000	1.661	2.099
Type_3* ɛ	1.026	.1221	8.404	.000	.787	1.266
Type_3*ɨ	.192	.0925	2.078	.038	.011	.374
Type_4*a	2.577	.0851	30.292	.000	2.410	2.744
Type_4* ɛ	1.192	.0850	14.029	.000	1.025	1.359

WORD READING: L1 Polish

Target: F1 (f1-f0; middle 20% of the vowel); main interaction: Vowel_quality*session/group

					95% Confidence Interval	
Model term	Coefficient	Std. Error	t	Significance	Lower	Upper
Intercept	5.481	22270.01	.000	1.00	-43655	43666
Group 1: T1*a	.112	22270.01	.000	1.00	-43660	43661
Group 1: T1* ɛ	-.954	22270.01	.000	1.00	-43662	43660
Group 1: T2*a	.280	22270.01	.000	1.00	-43662	43660
Group 1 T2* ɛ	-.896	22270.01	.000	1.00	-43660	43660
Group 1: T3*a	.411	22270.01	.000	1.00	-43661	43661

Group 1 T3* ε	-.731	22270.01	.000	1.00	-43660	43660
Group 2*a	.434	22270.01	.000	1.00	-43661	43661
Group 2* ε	-.630	22270.01	.000	1.00	-43660	43660
Group 3*a	.403	22270.01	.000	1.00	-43660	43661
Group 3* ε	-.775	22270.01	.000	1.00	-43661	43660
Group 4*a	-.120	22270.01	.000	1.00	-43662	43660
Group 4* ε	-.997	22270.01	.000	1.00	-43661	43660

WORD READING: L1 Polish

Target: F2 (F3-F2; middle 20% of the vowel); main interaction: Vowel_quality*session/group. The intercept is Monolinguals*ɛ.

					95% Confidence Interval	
Model term	Coefficient	Std. Error	t	Significance	Lower	Upper
Intercept	3.004	.5453	5.510	<.001	1.935	4.073
Group 1: T1*a	.776	.7673	1.011	.312	-.729	2.280
Group 1: T1* ε	-.258	.1691	-1.525	.127	-.590	.074
Group 1: T2*a	.796	.7673	1.037	.300	-.708	2.300
Group 1 T2* ε	-.384	.1691	-2.271	.023	-.715	-.052
Group 1: T3*a	.787	.7673	1.026	.305	-.717	2.291
Group 1 T3* ε	-.312	.1691	-1.844	.065	-.643	.020
Group 2*a	.804	.7696	1.045	.296	-.704	2.313
Group 2* ε	-.402	.1805	-2.228	.026	-.765	-.048
Group 3*a	.743	.7696	.966	.334	-.765	2.252
Group 3* ε	-.589	.1806	-3.261	.001	-.943	-.235
Group 4*a	.754	.7521	1.003	.316	-.720	2.229

SENTENCE READING: L1 POLISH

Target: F1 (F1-f0; middle 20% of the vowel); main interaction: Vowel_quality*group/session. The intercept is Monolinguals*ε.

					95% Confidence Interval	
Model term	Coefficient	Std. Error	t	Significance	Lower	Upper
Intercept	3.560	.1143	31.142	.000	3.336	3.784
Group 1: T1*a	1.071	.1702	6.291	<.001	.737	1.404
Group 1: T1* ε	-.073	.1701	-.429	.668	-.407	.261
Group 1: T2*a	1.058	.1701	6.221	<.001	.725	1.392
Group 1 T2* ε	-.088	.1701	-.516	.606	-.421	.246
Group 1: T3*a	1.193	.1702	7.009	<.001	.859	1.526

Group 1 T3* ε	-.057	.1701	-.334	.738	-.390	.277
Group 2*a	1.431	.1800	7.950	<.001	1.078	1.784
Group 2* ε	.371	.1800	2.061	.039	.018	.724
Group 3*a	1.150	.1799	6.393	<.001	.798	1.503
Group 3* ε	.067	.1799	.370	.711	-.286	.419
Group 4*a	1.288	.0849	15.161	.000	1.121	1.454

SENTENCE READING: L1 POLISH

Target: F2 (F3-F2; middle 20% of the vowel); main interaction: Vowel_quality*group/session. The intercept is Monolinguals*ε.

					95% Confidence Interval	
Model term	Coefficient	Std. Error	t	Significance	Lower	Upper
Intercept	2.768	.1473	18.787	.000	2.479	3.056
Group 1: T1*a	.849	.2353	3.607	<.001	.387	1.310
Group 1: T1* ε	.100	.2352	.423	.672	-.362	.561
Group 1: T2*a	.807	.2352	3.431	<.001	.346	1.268
Group 1 T2* ε	.087	.2352	.370	.712	-.374	.548
Group 1: T3*a	.857	.2352	3.644	<.001	.393	1.318
Group 1 T3* ε	.097	.2352	.414	.679	-.364	.558
Group 2*a	.866	.2421	3.576	<.001	.391	1.340
Group 2* ε	-.047	.2421	-.196	.845	-.522	.427
Group 3*a	.813	.2421	3.357	<.001	.338	1.287
Group 3* ε	-.141	.2421	-.581	.561	-.615	.334
Group 4*a	1.116	.1590	7.023	<.001	.805	1.428

WORD READING: L1 POLISH VS. L2 ENGLISH

Target: VOT of /p, t, k/; main interaction: Session/Group*Language. The intercept is Group 3*English.

					95% Confidence Interval	
Model term	Coefficient	Std. Error	t	Significance	Lower	Upper
Intercept	78.442	4.5234	17.341	.000	69.574	87.309
Group 1: T1*Polish	-34.923	3.8956	-8.895	.000	-42.560	-27.286
Group 1: T1* English	-15.420	3.8724	-3.982	<.001	-23.011	-7.828
Group 1: T2*Polish	-32.861	3.8964	-8.434	.000	-40.499	-25.222
Group 1 T2* English	-11.773	3.8733	-3.040	.002	-19.367	-4.180
Group 1: T3*Polish	-34.476	3.8962	-8.849	.000	-42.115	-26.838

Group 1 T3* English	-13.450	3.8725	-3.473	<.001	-40.990	-24.671
Group 2*Polish	-32.831	4.1620	-7.888	<.001	-40.990	-24.671
Group 2* English	1.683	4.1408	.407	.684	-6.435	9.801
Group 3*Polish	-31.817	1.2292	-25.884	.000	-34.227	-29.407

WORD READING: L1 POLISH VS. L2 ENGLISH

Target: VOT of /b, d, g/ (Type 1); main interaction: Group/Session*Language. The intercept is Group 3*English.

					95% Confidence Interval	
Model term	Coefficient	Std. Error	t	Significance	Lower	Upper
Intercept	-88.366	4.2312	-20.885	.000	-96.663	-80.069
Group 1: T1*Polish	7.285	5.0835	1.433	.152	-2.683	17.254
Group 1: T1* English	1.190	5.0251	.237	.813	-8.664	11.044
Group 1: T2*Polish	4.484	5.1117	.877	.380	-5.540	14.507
Group 1 T2* English	6.906	5.1121	1.351	.177	-3.118	16.931
Group 1: T3*Polish	9.136	5.1173	1.785	.074	-.899	19.171
Group 1 T3* English	8.519	5.1510	1.654	.098	-1.582	18.620
Group 2*Polish	13.222	5.4672	2.419	.016	2.502	23.943
Group 2* English	13.856	5.7182	2.423	.015	2.642	25.069
Group 3*Polish	-5.469	2.7782	-1.969	.049	-10.917	-.021

WORD READING: L1 POLISH VS. L2 ENGLISH

Target: VOT of /b, d, g/; main interaction: Session/Group*Language*POA. The intercept is Monolinguals*English*/d/

					95% Confidence Interval	
Model term	Coefficient	Std. Error	t	Significance	Lower	Upper
Intercept	-87.271	5.0575	-17.256	.000	-97.189	-77.354
Group 1: T1*Polish*b	.050	6.2047	.008	.994	-12.117	12.218
Group 1: T1*Polish*d	5.864	6.1853	.948	.343	-6.265	17.994
Group 1: T1*Polish*g	14.298	6.1739	2.316	.021	2.191	26.405
Group 1: T1* English*b	-5.553	6.1502	-.903	.367	-17.614	6.507

Group 1: T1* English*d	1.705	6.2139	.274	.784	-10.480	13.891
Group 1: T1* English*g	.680	6.1916	.110	.913	-11.462	12.821
Group 1: T2*Polish*b	-3.189	6.2199	-.513	.608	-15.386	9.008
Group 1: T2*Polish*d	5.286	6.2682	.843	.399	-7.005	17.578
Group 1: T2*Polish*g	9.888	6.2941	1.571	.116	-2.454	22.331
Group 1 T2* English*b	-.059	6.3366	-.009	.993	-12.484	12.367
Group 1 T2* English*d	7.124	6.4183	1.110	.267	-5.456	19.667
Group 1 T2* English*g	7.106	6.4059	1.109	.267	-5.456	19.667
Group 1: T3*Polish*b	2.675	6.2415	.429	.668	-9.565	14.914
Group 1: T3*Polish*d	9.295	6.2682	1.483	.138	-2.998	21.586
Group 1: T3*Polish*g	13.999	6.2927	2.225	.026	1.659	26.339
Group 1 T3* English*b	3.236	6.3223	.512	.609	-9.162	15.633
Group 1 T3* English*d	6.937	6.5873	1.053	.292	-5.981	19.854
Group 1 T3* English*g	7.457	6.6176	1.127	.260	-5.520	20.434
Group 2*Polish*b	3.495	6.6269	.527	.598	-9.500	16.491
Group 2*Polish*d	14.652	6.7229	2.179	.029	1.469	27.835
Group 2*Polish*g	21.116	6.7735	3.117	.002	7.833	34.398
Group 2* English*b	4.066	6.8859	.590	.555	-9.437	17.569
Group 2* English*d	11.401	8.2764	1.378	.168	-4.2829	27.631
Group 2* English*g	23.491	7.6459	3.072	.002	8.497	38.484
Group 3*Polish*b	-17.583	4.7873	-3.673	<.001	-26.891	-8.195
Group 3*Polish*d	-1.647	4.6711	-353	.724	-10.807	7.513
Group 3*Polish*g	.609	4.7496	.128	.898	-8.705	9.923
Group 3*English*b	-3.248	5.2949	-.956	.339	-15.446	5.320
Group 3*English*d	-5.063	5.2949	-.956	.339	-15.446	5.320