Iga Krzysik

Individual differences in multilingual acquisition of phonology: A longitudinal study of adolescents

Różnice indywidualne w przyswajaniu fonologii z perspektywy wielojęzyczności: badanie podłużne wśród młodzieży

Rozprawa doktorska napisana
na Wydziale Anglistyki
Uniwersytetu im. Adama Mickiewicza w Poznaniu
pod kierunkiem
dr hab. Magdaleny Wrembel, prof. UAM

Poznań, 2022

ACKNOWLEDGEMENTS

I would like to thank all participants who agreed to take part in this study as well as their parents and school administration. They made the organization of the study possible.

My most sincere thanks go to my supervisor, Prof. Magdalena Wrembel, for her invaluable guidance and support. Thank you for your feedback and advice, extensive comments and enlightening discussions. Thank you also for your continued encouragement and a considerate approach throughout the course of the studies.

Big thank you for all the support to my fellow PhD students at WA. Among them, first and foremost to Halina Lewandowska for being my closest and ever-encouraging companion on the PhD journey. Mateusz Jekiel – thank you for the valuable consultations on statistics, your time and patience.

Thank you to the Multi-Phon team for creating an environment for learning and growth. Big thank you to Christina Nelson for her help with the study materials and consultations concerning L3 German.

Dziękuję rodzicom za niezmienne wsparcie.

Michał, thank you for always being there for me.

OŚWIADCZENIE

Ja, niżej podpisana

Iga Krzysik

przedkładam rozprawę doktorską

pt. Individual differences in multilingual acquisition of phonology: A longitudinal study of adolescents

Różnice indywidualne w przyswajaniu fonologii z perspektywy wielojęzyczności: badanie podłużne wśród młodzieży

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, 6.07.2022

(miejscowość, data)

(czytelny podpis)

Table of contents

List of tables

List of figures

Introduction

This investigation of individual differences in the acquisition of phonology across multiple language systems constitutes a growing subfield of the research on multilingual phonology. Researchers implemented an array of methods and tools to emphasise the influence of individual variability on the process of multilingual phonological acquisition, and consequently, on the perception and production of phonological features. Previously conducted studies explored the role of such factors as executive control (e.g. Aliaga-Garcia et al. 2011, Mora and Darcy 2013, Darcy et al. 2016, Sigmeth and Golin 2018, Krzysik and Wrembel 2019, Sigmeth et al. 2019, Krzysik 2020), language instruction, exposure and use (e.g. Kopečková et al. 2016, Llama and López-Morelos 2016, Wrembel 2014, Wrembel et al. 2019, Wrembel et al. 2020, Nelson 2022) or psychotypology (Nelson et al. 2021), yielding diverse results concerning phonological development. Some of these studies openly acknowledged their approach as belonging within the framework of individual differences, others recognised the role of individual variability and its influence on multilingual phonology. These studies have also reinforced the need for more research exploring the relationship between multilingual phonological perception and production and a host of cognitive and environmental factors.

The research design which seems to be particularly fitting for the investigation of individual differences is a longitudinal study. By conducting multiple observations or testing sessions, the researchers can examine the interaction of several variables over time e.g. L3 instruction and multilingual proficiency or the cumulative effect of multilingual acquisition and the potential interplay of additional factors. Moreover, longitudinal design aligns with the formula of formal language instruction, where the learning outcomes are expected to occur within a given time frame, beyond the scope of cross-sectional design. Despite the listed advantages of longitudinal design, the multilingual phonology studies utilising such a framework still constitute a minority (e.g. Cabrelli Amaro 2013, Kopečková 2015, Kopečková 2016, Kopečková et al. 2019, Wrembel et al. 2019).

The study featured in the present thesis constitutes an attempt at integrating the approaches of the previous works within the field of multilingual phonology. Firstly, it incorporates the assessment of multilingual perception and production; modalities which are rarely explored together in a singular study. Moreover, the battery of tests in the study

incorporates the measures of factors associated with individual differences in the acquisition of multilingual phonology. Additionally, the longitudinal framework is geared towards capturing the development of multilingual phonology. Ultimately, the design of the study constitutes an attempt at extending the scope of the previous research and potentially offering a more multidimensional overview of the process of multilingual development of phonology and the role of individual differences in the aforementioned process.

The thesis aimed to explore the role of individual differences in the process of phonological development of emergent adolescent sequential multilinguals (L1 – Polish, L2 – English, L3 – German), acquiring their L2 and L3 in the context of formal instruction in primary school. The battery of tasks implemented in a longitudinal study included tests of phonological perception and production, inhibitory control, phonological working memory and language proficiency supplemented by measures of psychotypology and language history and use questionnaires. Consequently, through this extended testing battery, the study featured in the thesis attempted to fulfil the objective of assessing the development of phonological perception and production as well as the role of factors potentially underlying individual differences.

The thesis is divided into three parts, namely, an introduction of the theoretical background of the study (Chapters 1-3), the methods and procedures applied in the study (Chapter 4) and a presentation of the results and a discussion of their implications (Chapter 5-6). The theoretical chapters present the issues within the field of multilingual phonology with a focus on the aspects particularly relevant to the tested group. Chapter 1 offers a general overview of multilingual acquisition of phonology and explores such issues as the distinction between the sequential development of L2 and L3, non-native multilingual acquisition or cross-linguistic influence. Moreover, it outlines the current research within the field of multilingual phonology and the relevant models of language acquisition. Chapter 2 discusses the role of individual differences in multilingual acquisition based on the selected factors investigated in the empirical part of the study such as age, inhibitory control, phonological working memory, psychotypology, and the characteristics of language exposure and use. Chapter 3 describes the phonological systems of the tested languages – Polish, English and German. Subsequently, it compares the characteristics of the said systems and discusses their implications for the process of multilingual acquisition. Chapter 4 outlines the research questions with the predictions of

the possible outcomes. Further, it presents the design of the testing battery and offers a detailed description of the applied measures and procedures. Chapter 5 presents the results obtained through the measures of the testing battery. Finally, Chapter 6 offers a discussion and interpretation of the results in relation to the selected aspects of the theoretical background presented in Chapters 1-3. Chapter 6 is followed by conclusions and references.

Chapter 1: Multilingual acquisition of phonology

1.1. Introductory remarks

Multilingualism constitutes a manifestation of an innate propensity for acquiring multiple languages, postulated as a default state of human linguistic capacity (De Angelis 2007: 2). Across the subsequent historical periods, the practice of speaking multiple languages has been a defining hallmark of many societies. Nowadays, multilingualism is gradually becoming an indispensable feature of one's lived experience in the increasingly globalised world. Due to the prevalence of multilingualism, both as an individual and a societal trait, the study of its manifestations and patterns may offer a wide range of answers to questions about mechanisms underlying language acquisition, processing and use.

Multilingual acquisition of phonology constitutes one of the fundamental aspects of multilingualism – its outcomes affect speech processing and production, the efficiency of communication and the societal reception of speakers. However, as a field of study, it is a newly developing and understudied domain, requiring further empirical exploration (Wrembel 2015, Cabrelli Amaro and Wrembel 2019). Consequently, the studies of multilingual phonology draw from the conceptualisations, theories and methods of fields such as psycholinguistics, second language acquisition or other subfields of multilingualism associated with domains such as syntax or morphology. Chapter 1 grounds the topic of multilingual acquisition of phonology within the discussion of the founding issues of the field of multilingualism, namely, the definition of multilingualism, the distinction between L2 and L3 acquisition and cross-linguistic influence. Further, the chapter provides an overview of the body of previous research concerning multilingual phonology. Chapter 1 concludes with the discussion of the models of multilingual acquisition particularly relevant to the studied context.

1.2. Conceptualising multilingualism and multilinguals

Multilingualism is a complex phenomenon, manifested in a variety of contexts and spaces. The definition of multilingualism has frequently been considered against the backdrop of mono- and bilingualism and presented as a special or additional category, implying increased social diversity, individual uniqueness or even suggesting a certain cognitive advantage. Berthele (2021) describes multilingualism as a simultaneously ordinary and extraordinary concept. According to his conceptualisations, the ordinary quality of multilingualism lies in the consideration of the multitude of the existing linguistic styles and dialects, which renders almost anyone multilingual. The extraordinary quality is associated with the assumption that some individuals have more languages at their disposal than others. Drawing on the approaches to multilingualism present in the field, Berthele (2021: 107-109) outlined three scenarios within which multilingualism could be conceptualised by the researchers. The first conceptual scenario is labelled as a postmodern one, abandoning language labels and categories and highlighting the individual and idiosyncratic nature of multilingualism. The second scenario proclaims multilingualism the term multilingualism is too polysemous to be considered useful. Instead, researchers start using other terms, related to one's linguistic repertoire or proficiency. The third scenario retains the term multilingualism and steers towards the investigation of various types of multilingualism, taking into account factors that shape them, with careful attempts at generalisation. The third scenario is aligned with the approach towards conceptualising multilinguals and multilingualism in the current study: detailed characterisation of the investigated group with a reference to a larger, more general classification/classifications. Consequently, the remaining paragraphs of the current section outline the main categories frequently used by linguists to shape the discussions of multilingualism, general definitions of individual multilingualism and a toolbox for characterising and describing multilinguals.

Aronin (2018: 4-7) identified four main dimensions along which multilingualism is frequently discussed: historical, current, individual and societal. The historical dimension is predominantly interested in the instances of past multilingualism, whereas the current dimension focuses on the present context. The individual and societal dimensions are more difficult to separate; the former focuses on language acquisition through the lens of an individual, the latter on the linguistic characteristics and practices

of a given community. The present study is primarily preoccupied with the investigation of current individual multilingualism, while the societal aspect is used to contextualise the participants. Therefore, the following sections will focus primarily on individual multilingualism studied from the current perspective.

Individual multilingualism is commonly defined as the knowledge, the use or the ability to use three or more languages (e.g. Bathia and Ritchie 2013, Maher 2017). This broad definition does not specify the level of knowledge, the frequency of language use or nativeness, which aligns with the current tendencies to account for diverse multilingual experiences, consider them holistically and view them as likely to evolve over one's lifetime (Kemp 2009:19). Consequently, the label *multilingual* would be applicable, among others, to an individual acquiring three languages simultaneously from birth as well as to a formally-instructed learner introduced to their languages sequentially. The definition of multilingualism outlined above may be too broad for research purposes; it can hardly be used as an operational one. However, a move towards a single, uniform and more restrictive formulation may be difficult and should be approached with caution; every research context requires a detailed description of the tested individuals and the characteristics of their linguistic experience (Valdés 2021). Such an approach of contextualisation and description of multilinguals aligns with the most plausible scenario for conceptualising multilingualism offered by Berthele (2021), which, as stated above, favours careful generalisation and in-depth description of the particular subtypes of multilingualism.

The approach following the broad definition of multilingualism and relying on description and classification requires a set of criteria helpful in describing particular types or groups of multilinguals. Festman (2018: 239) proposed categories for classifying various characteristics of individual multilingualism, creating a toolbox helpful in describing multilingual participants. The first subgroup of characteristics associated with the individual multilingual experience included the roles that the languages have in one's repertoire (e.g. first, second and additional language), the current level of proficiency (on a spectrum from near-native to far from native) and the age of acquisition. Further individual criteria included the influence of L1 on L2 and L3 (additive and subtractive), the frequency and recency of language use (equal and unequal), one's switching habits (frequent, infrequent) and the purpose of language use (professional, private). The external criteria encompassed the context and the manner of acquisition (formal,

informal), the status of the acquired languages (high-prestige and low-prestige), typology of the languages (close and distant), and socioeconomic status of an individual. The collection of the criteria proposed by Festman (2018) reflected the wide scope of multilingualism implied by the general definition but also allowed for a detailed description of an individual multilingual profile.

The conceptualisation of multilingualism and multilingual for research purposes constitutes a significant challenge. In their discussion of the criteria for multilingualism, Rothman et al. (2019) concluded that a decision about who qualifies as a multilingual should align with the goals and the context of the study. For instance, if a given study intends to research emerging L3 learners, then a minimal and newly growing proficiency in L3 will satisfy the requirements for applying the term multilingual in such a context. Ultimately, the task of constructing an operational definition of multilingualism and a set of corresponding criteria for classification requires two main steps. First, the assessment of the research goals and the characterisation of the group that will be examined to answer these goals. Then, the construction of detailed criteria would enable the process of qualification of the participants for the study. The later stages described by Kemp (2009) involve the qualification of the tested individuals according to the operational definition and a detailed description of the examined group to facilitate understanding of the principles behind the study design.

1.3. L2 vs L3 acquisition

The quantitative and qualitative differences between bilingual and multilingual language acquisition constitute a fundamental issue in the studies of multilingualism. For years, the investigation of third or additional languages was either ignored or subjected to over-generalisation under the umbrella term of second language acquisition (Hufeisen 2000: 214). Such practices contributed to the development of bilingual bias; an inclination to view multilinguals as bilinguals with additional languages (De Angelis 2007: 15). Such a view of multilingualism has been disputed by numerous researchers, who offered both analytical (e.g. Wrembel 2015, Hammarberg 2018) and empirical evidence (e.g. Leung 2005, Bardel and Falk 2007) supporting the claim about the differences between L2 and L3 acquisition. The primary difference between L2 and L3 acquisition may be reduced to

a claim that in the former case learners rely on their L1, whereas in the latter learners may draw on their experience of acquiring more than one language prior to L3. This simple difference, however, is underpinned by several factors. The main points of the divergence marking the distinction between second language acquisition and L3 acquisition identified by Hufeisen (2000), De Angelis (2007) and Gut et al. (2015) are as follows: the level of metalinguistic awareness, language learning experiences and learning strategies as well as the directionality of cross-linguistic influence.

Metalinguistic awareness is defined as an ability to switch focus between linguistic form and meaning, consider language in abstract terms, reflect on it, play with it and manipulate its structure (Jessner 2006: 42, Jessner 2008: 277). Metalinguistic awareness is forged in the process of language acquisition and further affects the outcomes of the process (Lorenz and Siemund 2019: 84). According to Jessner (2006), metalinguistic awareness in multilinguals falls under the larger term of linguistic awareness along with cross-linguistic awareness – an ability to recognise the connections between the multiple language systems. The dynamic model by Herdina and Jessner (2002) acknowledges that language proficiency – the result of the multilingual speakers' knowledge of language and language use – is largely influenced by metalinguistic and cross-linguistic awareness. Consequently, it might be assumed that the starting point for L2 acquisition is the metalinguistic awareness gained in the process of L1 acquisition, whereas L3 acquisition is approached with the awareness forged in the process of gaining monolingual experience but also through the recognition of cross-linguistic relationships between one's L1 and L2. Therefore, it may be hypothesised that during the process of L3 acquisition the speakers have a more extensive linguistic awareness compared to the onset of L2 acquisition. Such increased multilingual linguistic awareness may demonstrate through improved language monitoring in speech production (Herdina and Jessner 2000), the inclusion of other languages in the process of using or learning another language (Bono 2011) or smoother acquisition of further languages (Festman 2018).

Language learning experience and strategies differ both qualitatively and quantitatively between L2 and L3 acquisition. In sequential language acquisition, which is the predominant context of the current discussion, L2 acquisition is approached without foreign language learning experiences and strategies. L3 acquisition, however, benefits from the experience of learning a previous foreign language, i.e. the L2. The individual acquiring their L3 has already developed different language learning strategies, which

can be extended to the third language, possibly providing a learner with a substantial advantage. Moreover, in the Polish context, textbooks based on the official curriculum for the early stages of L3 instruction underline the possible links with the L2 of the students, which can facilitate the transfer of strategies and experiences. Such facilitation through earlier linguistics experience and activation of already existing strategies would not be possible in L2 acquisition.

L2 and L3 acquisition may also differ in the characteristics of cross-linguistic influence. Directionally of CLI in second language acquisition may be bidirectional and occur between two languages, i.e. the native and target language (Pavlenko 2002), whereas multilingual language acquisition may include multidirectional CLI between all languages in the repertoire of a given speaker (De Angelis 2007). Consequently, the presence of two languages and the addition of L3 increases the complexity of CL1 as compared to the process of L2 acquisition.

From the perspective of phonology, L3 acquisition differs qualitatively and quantitatively from L2 acquisition in the following aspects: the extent of the phonemic repertoire, articulatory knowledge, enhanced metalinguistic awareness and improved perceptual sensitivity (Gut 2010: 21, Wrembel 2015: 33). The acquisition of L2 is approached with a single phonemic repertoire derived from L1, while the acquisition of L3 may already be influenced by two phonemic repertories of L1 and L2. Metalinguistic awareness concerning phonology might be substantially different in the case of L3 learners, who had a chance to develop it already in the process of L2 acquisition. Consequently, the familiarity with the phonemic systems of two languages, the experience of acquiring these systems and metalinguistic awareness forged in the process may improve perceptual sensitivity and facilitate L3 acquisition of phonology. Despite the scarcity of comparative studies examining the qualitative and quantitative differences between L2 and L2 phonological acquisition, several scholars offered an insight into the issues associated with multilingual phonological sensitivity and awareness, using self-reflective measures. Wrembel (2014a) examined adult learners of L3 Polish, who were requested to comment on their earlier phonological performance in their third language. The study found that the participants had a considerable level of metaphonological awareness; they managed to recognise L3 features, self-correct and repair errors. The post-hoc corrections made by the participants had a rather conscious character. Kopečková (2018) investigated a group of L1 German, L2 English and L3 Spanish

multilinguals who demonstrated a certain amount of phonological awareness concerning their errors in L3, manifested in a more intuitive way and lacking metalanguage. Despite the differences in the characteristics of comments and corrections, the results demonstrated that the examined multilinguals managed to develop a certain level of phonological sensitivity and awareness, most likely due to the influence of their linguistic experience.

1.4. Cross-linguistic influence

Cross-linguistic influence (CLI) is a broad term encompassing a host of phenomena described in the literature as transfer, interference or avoidance (Sharwood Smith and Kellerman 1986). CLI describes the mechanisms through which the knowledge of one language affects the learning or use of another in the individual context (Jarvis and Pavlenko 2008: 1, James 2012: 858). The field of second language acquisition has traditionally perceived CLI as a unidirectional influence of L1 on L2 (Pavlenko 2002: 190), however, a revised perspective on directionality in L2 acquisition and the proliferation of L3 research lead to the re-evaluation of this approach. The present section discusses CLI typology and the factors influencing CLI in multilinguals.

Jarvis and Pavlenko (2007: 20-26) offered a thorough overview of the types of cross-linguistic influence. Their typology distinguished 10 types of CLI based on: area of language knowledge/use, directionality, cognitive level, type of knowledge, intentionality, mode, channel, form, manifestation and outcome.

The language knowledge/use type includes types of CLI associated with various linguistic domains such as phonology, orthography, lexis, semantics, morphology, syntax, discourse, pragmatics and sociolinguistics. Phonological CLI, which is of the greatest importance to the current study, may manifest in the characteristics of the perception and production in the languages of one's multilingual repertoire. For instance, Sypiańska (2016a) observed that L1 Polish L2 English and L3 German multilinguals exhibited raising and backing of L1 Polish vowels due to the CLI from L2 English and L2 English and lowering and backing or fronting of L1 Polish vowels as a result of CLI from German. Wrembel et al. (2020) examined the perception of final obstruents by L1 Polish, L2 English and L3 German adolescents. The results indicated that despite 5-6

years of L2 instruction, the participants' perception of the partially devoiced final obstruents in L2 English was substantially affected by the CLI from L1 featuring devoiced final obstruents. These example studies illustrate the premise of phonological CLI – the characteristics of the forming phonological systems or the systems already in place interact and affect the remaining systems in the mind of the speakers.

CLI can be classified into several types based on the notion of directionality; in the context of second language acquisition, Jarvis and Pavlenko (2007: 11) distinguished *forward transfer* (from L1 to L2) and *reverse transfer* (from L2 to L1). However, a simple extension of these terms to L3 would be insufficient due to the complexity of the relations between the languages in a multilingual mind (Wrembel 2015: 48). Therefore, Jarvis and Pavlenko (2007) acknowledged the multilingual context and distinguished *lateral* (L2 to L3 or L3 to L2) and *bidirectional* or *multidirectional* transfer (forward and reverse or lateral transfer between the multiple languages). Such an addition accounted for the different dynamics between the languages of a multilingual, affected by such factors as the order of acquisition, language status or proficiency.

On the cognitive level, CLI can be conceptualised as the transfer of mental representations from one linguistic system to another. Jarvis and Pavlenko (2007) noted that there might be several cognitive levels on which the languages in the speaker's mind interact, even simultaneously. Another type of CLI, closely related to cognitive level CLI, is associated with the type of knowledge; it is based on the distinction between the implicit and explicit types of knowledge popularised in the field of SLA. Jarvis and Pavlenko (2007), following the studies by Fabbro (2002) or Paradis (2004), proposed that this differentiation may be crucial for CLI; it might affect the way in which subsequent languages of the speaker are stored and processed in one's mind. Further, Jarvis and Pavlenko (2007) distinguished CLI motivated by intentionality, conceptualised as a difference between the intentional and unintentional language switches. The former might be a result of a communicative strategy and the latter might be an outcome of cognitive connections between the languages. The types of CLI associated with mode (productive and receptive) and channel (oral-aural/written-manual) form the basis for the distinction between the types of CLI that involve perception and production as well as speech/writing and sign language. Two final types of CLI in the discussed classification refer to manifestation and outcome. Manifestation covers the instances of the overt vs covert CLI; the former refers to the identification of corresponding structures between the languages,

whereas the latter to situations in which the speaker transfers the structures from the source language even though they do not exist in the recipient language. Finally, the notion of the outcome refers to the positive vs negative transfer, namely, the potential distinction between the positive and negative outcomes of CLI in the context of language acquisition. Subsequently, Jarvis and Pavlenko (2007) point out that the 10 types of transfer can be combined, depending on the factors influencing the particular case examined in a given study.

The occurrence and characteristics of CL1 in multilingualism are linked to a host of factors including typology, L2 status, proficiency and recency (Williams and Hammarberg 1998, De Angelis 2007). The assumptions and predictions related to these factors are outlined in the subsequent paragraphs.

Typological relatedness of the languages in the repertoire of a multilingual individual might have a substantial effect on the shape of CLI. Several studies have shown that a typologically closer language is the dominant source of CLI in L3 e.g. Ahukanna et al. (1981), Cenoz (2001), De Angelis and Selinker (2001), Ó Laoire and Singleton (2009) Lindqvist and Bardel (2013), Lindqvist (2019). Some of the listed studies ventured beyond the historical relatedness and structural similarity by referencing the notion of *psychotypology*, which is an individual and subjective perception of language distance. Kellerman (1985), who introduced the term, proposed that L1 might influence L2 when both languages are perceived as close. De Angelis (2007: 23) pointed to the bilingual constraint of this formulation and extended the claim to multilingualism, noticing that multilinguals tend to be most influenced by the languages perceived as close to the target language. The accounts concerning the relationship between psychotypology and CLI are, however, to a large extent speculative, as the majority of the studies have made assumptions about the perception of language distance without directly measuring its impact (Nelson et al. 2021). The role of typology and psychotypology in CLI was explored in several third language acquisition models, predominantly based on syntactic data, such as the Cumulative Enhancement Model (Flynn et al. 2004), Typological Primacy Model (Rothman 2011, 2013, 2015), Linguistic Proximity Model (Westergaard et al. 2016).

The role of L2 status was examined by Williams and Hammarberg (1998) and Hammarberg (2001) in a case study of an adult learner of L3 Swedish with L1 English and L3 German. The collected data revealed different roles of L1 and L2 in the process

of L3 acquisition; L1 English had an "instrumental" role (associated with metalinguistic function), whereas L2 German was a "default supplier" underlying L3 production. Williams and Hammarberg (1998) noted, that depending on the factors such as language proficiency, typology and recency of use, L2 may be activated as a supplier even during the early stages of L3 acquisition, also on the phonological level. These results and assumptions found a reflection in the studies by Bardel and Falk (2007) and Falk and Bardel (2011), who tested the hypothesis about the role of L2 in the area of syntax, with both studies confirming the active role of L2 in L3 acquisition. The obtained results led to the formulation of the L2 Status Factor Model, which postulates that the status of L2 predicts the positive and negative transfer in L3. The authors claim that L2 might be transferred more easily in some situations due to a host of shared situational factors. Especially in the context of formal language instruction, formally acquired L2 and L3 can share circumstantial and cognitive characteristics, which are different in L1 acquisition, such as the age of onset, learning setting, learning strategies, language learning awareness and metalinguistic knowledge/awareness. (Falk et al. 2015: 228). Moreover, the authors of the model relate the role of L2 in the acquisition of the subsequent languages to the organisation of memory, drawing from the declarative/procedural model (Paradis 1994, 2004, 2009). The majority of L1 knowledge (including phonology, morphology and syntax) is sustained by procedural memory, whereas these areas in L2 acquired in the formal instruction context are associated with declarative memory. Such a distinction in cerebral representation may potentially explain the role of the L2 status factor in L3 learning due to their assumed similar organisation in the mind (Falk et al. 2015: 228-229). The assumptions of the declarative/procedural distinction driving the L2 status factor have been challenged, especially in relation to the lack of clarity concerning the potential of structural similarity to override the effect of different mental storage (Rothman 2013). However, the special status of L2 remains a focal point of the studies investigating the CLI in multilingual acquisition.

Language proficiency in both the source and target/recipient languages is positioned as a factor influencing the characteristics and extent of CLI in multilinguals (Odlin and Jarvis 2004, Jarvis and Pavlenko 2007: 201). De Angelis (2007: 33) reaffirmed the earlier claims by Odlin (1989), pointing to caution in approaching proficiency as a factor determining the types of CLI which might occur at different stages of language development. Recency, understood both as the recency in terms of the last acquired

language and the last time a given language was used, was listed by Hammarberg and Hammarberg (1998) and Williams and Hammarberg (2001) as a potential factor affecting CLI. Hammarberg (2001: 23) implied that L2 influence on L3 might be activated more easily if the language has recently been used. Moreover, the results of the reference study suggested that the influence of a more recently acquired language (L2 German) was more abundant than this of L1 English in the early stages of L3 Swedish acquisition. A similar effect was described in other studies, for instance, Shanon (1991), Dewaele (1998) and Freundlich (2016). De Angelis (2007: 36) also noted that contrary to the expectation of many speakers, even the languages which were inactive for a particular amount of time might be visible in CLI.

In the context of multilingual phonology, CLI has been researched in relation to factors outlined above, such as typological distance, L1/L2 status and L2 proficiency (Kopečková 2014: 206). Several studies found evidence for a role of L2 status in phonological production (e.g. Hammarberg and Hammarberg 1993, Wrembel 2009, Llama et al. 2010) or for the L1/L2 combined or hybrid influences (in the case of VOT) (e.g. Wunder 2010, Wrembel 2015, Wrembel et al. 2019). The presence of L2-driven CLI was in some cases associated with the early stages of L3 proficiency (e.g. Hammarberg and Hammarberg 2005, Wrembel 2009). Some studies pointed to a combination of typological, status-related and environmental factors potentially shaping CLI in L3 (e.g. Kopečková 2014, Kopečková et al. 2016).

1.5. Multilingual acquisition of phonology – an overview of research

The field of multilingual phonology acquisition is frequently described as rather an underdeveloped area of study within the broader context of multilingualism (Cabrelli Amaro and Wrembel 2019). However, the body of research which has already been amassed provides a solid foundation for further empirical exploration and highlights certain trends and potential areas of interest. The current section outlines the previous research within the field, focusing on speech production and perception, a longitudinal paradigm and individual differences in multilingual phonology research.

1.5.1. Cross-sectional research on multilingual speech perception

The cross-sectional studies of multilingual phonological perception focused on vowels (Kopečková 2015, Lewandowska and Wrembel 2019) or included multi-feature analysis (Onishi 2016). Kopečková (2015) investigated the perception of Polish and Irish English vowels by a group of L1 Polish, L2 or L3 Irish English children. The results of the perceptual assimilation and goodness of fit tasks revealed that L3 learners performed with greater phonological sensitivity compared to L2 learners. Lewandowska and Wrembel (2019) examined the perception of vowel contrasts in the L3 and L3 of two groups of participants (L1 Polish, L2 English, L3 German and L1 Polish, L2 English, L3 French) with intensive L3 exposure. The L3 French group performed better, despite the lack of potential typological facilitation. The exposure to a typologically close L3 German did not enhance the performance in L2 English and vice-versa. Onishi (2016) combined several categories and investigated the perception of eight Japanese vocalic and consonantal contrasts by L1 Korean, L2 English and L3 Japanese speakers in an ABX discrimination task and a forced-choice task. The participants performed also a forced-choice task based on L2 contrasts. The obtained results suggested that improved L2 perception predicts more accurate L3 perception, however, the effect was restricted to several contrasts. Nevertheless, the overall results implied that phonological perception is not only affected by the previous exposure to particular contrasts but also by the accumulated experience of foreign language learning.

1.5.2. Cross-sectional research on multilingual speech production

The cross-sectional research on multilingual speech production encompasses a wide variety of studies focusing on different features such as vowels, rhotic consonants or voice onset time. The studies featuring multilingual vowel production include, for instance, Gut (2010), Sypiańska (2016b) and Kopečková et al. (2016). The results of the studies are summarised in the following paragraph.

Gut (2010) investigated CLI in vowel reduction and speech rhythm produced by four speakers with different L1, L2, L3 sequences of English and German. The obtained material revealed a CLI from L2 to L3, however, there was no conclusive evidence of the

influence of L1 in L2 and L3. Sypiańska (2016b) examined the acquisition of vowels in L1 Polish, L2 Danish and L3 English adult speakers. The researcher observed a multidirectional CLI; L1 Polish vowels became higher and fronter under the influence of L2 Danish and L3 English and L2 Danish vowels changed their quality to backer affected by L1 Polish. Moreover, the target-like L3 English vowels were affected by combined L1/L2 CLI. Kopečková et al. (2016) focused on the interaction between three vowel inventories in the production of 14 adolescents raised in Germany, with L2 English and 1 year of formal Polish instruction. The participants differed in terms of their Polish language background. The data obtained in delayed repetition tasks in three languages revealed differences in the realisation of individual vowels between the heritage speakers of Polish and those who received the instruction only at school, pointing to the potential role of the language status effect.

The cross-sectional studies focusing on the production of consonants include the examinations of rhotics e.g. Kopečková (2014) and VOT e.g. Tremblay (2007), Llama et al. (2009), Wunder (2010), Wrembel (2011, 2014), Llama and Cardoso (2018). Kopečková (2014) examined the CLI in instructed phonological acquisition, focusing on the production of rhotics by 20 adolescent speakers of L1 German and L2 English during their first year of L3 Spanish instruction. The beginner L3 learners were capable of differentiating between the rhotic sounds in their languages and producing target-like L1 and L2 forms as well as approaching the target-like rhotics in L3 Spanish. Moreover, the CLI seemed to follow the cumulative pattern, as the participants exhibited a balanced influence of their L1 and L2 in L3 speech. Tremblay (2007) examined the VOT production by L1 English, L2 French and beginner L3 Japanese adult learners, who produced corresponding VOT patterns in their L2 and L3. Such results led to two possible explanations: (1) the role of L2 status in early L3 acquisition is dominant, (2) the influence of L2 is a combined L1 and L2 CLI, where L1 has already influenced the production in L2, which subsequently affected L3. Llama et al. 2010 combined the investigation of the L1/L2 status role and typology in CLI in the production of L3 voiceless stops. For both groups of participants, namely, L1 English, L2 French and L1 French, L2 English adult speakers, the production of voiceless stops in L3 Spanish was influenced by their second language. Wunder (2010) examined the acquisition of L3 Spanish voiceless stops by adult L1 German, L2 English speakers with L2 English. The results revealed compromised hybrid VOT values. Wrembel (2011) tested VOT production of L1 Polish, L2 English

and L3 French adult participants in the three languages. The results revealed interlanguage VOT patterns, with hybrid L1/L2 values in L3 VOT, evidencing the influence of combined CLI. In a subsequent study, Wrembel (2014b) extended her testing group to include 64 participants with two language combinations L1 Polish, L2 English and L3 French and L1 Polish, L2 English and L3 German. The obtained results revealed interlanguage patterns; the participants contrasted between VOT durations in three languages. Moreover, L3 VOT revealed hybrid patterns, i.e. intermediate between L1 and L2.

1.5.3. Multilingual production-perception studies

Studies combining production and perception constitute one of the most infrequent categories in multilingual phonology research, however, the exploration of these modalities in the context of a singular study may produce insightful findings concerning the process of multilingual acquisition. Wrembel et al. (2022) outlined four possible scenarios that a relationship between multilingual perception and production may follow. The first scenario, based on L2 speech learning research, proposed that perception precedes production (e.g. Speech Learning Model; Flege 1995, Saito and van Poeteren 2018). The second scenario suggested that perception and production are closely intertwined and develop together (revised Speech Learning Model; Flege and Bohn 2021, Kopečková et al. 2019). The third scenario assumed that accurate production develops before perception, which was linked with the explicit knowledge of problematic sounds and external pressure for accurate pronunciation. The fourth scenario assumes no direct link between the accuracy in speech perception and production. The empirical part of the paper by Wrembel et al. (2022) examined the accuracy of perception and production of rhotics by 24 adolescent learners with two language combinations: L1 Polish, L2 English, L3 German and L1 German, L2 English and L3 Polish. The results revealed a tendency for perception to precede production, which was additionally modulated by proficiency – the effect was more visible in more proficient L2. An earlier study conducted by Zhang (2019) investigated the acquisition of selected affricates and vowels by a group of L1 Catalan or Basque, L2 Spanish, L3 English speakers. The participants performed an ABX task and a non-word repetition task featuring L3 vowels. The obtained results

demonstrated that both L1 Catalan and Basque speakers performed similarly in terms of fricative/affricate and vowel perception, however, Catalan speakers did not demonstrate advantages in the perception and production of contrasts similar to the ones existing in Catalan. Another perception/production study was conducted by Liu and Lin (2021), who tested the perception and production of the word-initial stops in the respective L3s of L1 Mandarin Chinese, L2 English, L3 Russian or Japanese adult speakers. The participants completed an L3 identification task, an L3 reading task and an L2 reading task; the obtained results indicated that the participants could perceive the feature of voicing lead present in both L3s, but experienced difficulties in the production of the L3 voiced stops. No correlation was observed between the perception and production of the voiced stops; the perceptual accuracy of the learners could not predict their VOT production.

1.5.4. Longitudinal research of multilingual phonetics and phonology

Researchers within the field of multilingual acquisition of phonology have also been attentive to the developmental aspect of the researched phenomenon, which resulted in a number of studies utilising a longitudinal framework. A previously mentioned case study by Hammarberg (1993, 2005), Williams and Hammarberg (1997) Hammarberg and Hammarberg (2005) had a longitudinal character; the trilingual informant was observed over the course of 2 years. The study revealed that early acquisition of L3 Swedish was predominantly influenced by L2 German, which switched to L1 English with time and increased L3 proficiency. Another multilingual longitudinal case study was conducted years later by Golin et al. (2019), who focused on adult phonological development in the initial stages of L3 Polish learning by an L1 German speaker with L2 English. The study spanned over 20 recording sessions focusing on the production of /v/-/w/ contrast and vowel reduction over the first six months of L3 learning. The recording showed the connected development of both L2 and L3 was affected by the changing learning environment.

Longitudinal studies involving multiple participants focused on such features as rhotics, /v/-/w/ contrasts or VOT. For instance, Kopečková (2016) investigated the presence of bilingual advantage in L3 rhotic learning in a diverse group of 19 children acquiring Spanish as their L3 in three testing sessions: three and seventh months into L3

learning, and at the end of the third year of instructed L3 learning. The results showed no general bilingual advantage for L3 rhotics acquisition, however, active bilinguals whose repertoire encompassed alveolar trill benefited from their bilingual experience both at the initial and later stages of L3 Spanish acquisition. Wrembel et al. (2019) examined the production of rhotics by 16 adolescent L1 Polish, L2 English and L3 German multilinguals in three testing sessions over one school year. L2 English rhotics were produced at a rather consistent level of accuracy over three testing sessions, whereas the accuracy of L3 German rhotics decreased over time. In terms of CLI, higher proficiency in L2 English, unlike L1 German, was produced with minimal L1 influence over the tested period. L2 influence in L3 German rhotics production started to appear in the later stages of acquisition (T3). Moreover, the results were marked by a great degree of individual variation. Balas et al. (2019) examined rhotics perception in a timed forced choice task performed by 32 sequential adolescent multilinguals (L1 Polish, L2 English, L3 German and L1 German, L2 English, L3 Polish). The task was repeated in two testing sessions with a 4-month break. L1 Polish participants perceived L3 German rhotics with smaller accuracy than the L1 German participants perceived L3 Polish rhotics. Moreover, L1 Polish participants perceived the L3 rhotics with worse accuracy than the L2 rhotics; a similar tendency was observed for L1 German participants. A significant decrease over time over time was observed for the perception of L3 German by L1 Polish participants. Gut et al. (2019) examined the production of [v] / [w] contrast in L1 German/L2 English/L3 Polish adults and children in three testing sessions. The obtained results revealed that the adults had higher accuracy in L2 contrast production at all testing points, whereas the children showed improved accuracy for L3 Polish contrast, however, it was visible only in two testing sessions. The production of VOT in voiceless stops by seven adolescents and seven adults with the same linguistic repertoires of L1 German, L2 English, and L3 Polish was longitudinally examined by Nelson (2022). The production was marked by a great degree of individual variation, however, the patterns that emerged involved significant changes in the L1 and L2 over time for adolescents and relative stability for adults. A longitudinal multi-feature study conducted by Wrembel et al. (2020) focused on rhotics and final obstruent devoicing/voicing. The participants were 13 adolescent L1 Polish, L2 English, L3 German learners, who performed a timed forced-choice goodness task in their L2 and L3 at two testing times. The results revealed that CLI was feature-dependent and differed across L1-, L2- and L3-accented stimuli.

Moreover, the CLI in perceptual development evidenced stability for L2 rhotics, a relative fluctuation for L3 rhotics, and no significant development for final devoicing/voicing. The reaction times did not correlate with accuracy, which demonstrated that processing speed might be quite independent of the degree to which the representations of phonological categories are established.

1.5.5. Research on individual differences in multilingual phonology acquisition

A number of studies investigated multilingual acquisition of phonology in relation to individual differences – traits that distinguish one person from another. In these studies, researchers employ measures of different variables (i.e. cognitive, environmental etc.) to investigate their potential role in phonological perception or production. One of the most frequently investigated areas is the relationship between the selected cognitive variables and individual differences in multilingual phonology acquisition, including inhibition (e.g. Mora and Darcy 2013, Darcy et al. 2016, Sigmeth and Golin 2018, Sigmeth et al. 2019, Krzysik 2020) and phonological working memory (Aliaga-García et al. 2011, Krzysik and Wrembel 2019). The listed studies yielded diverse results – from moderate links between the cognitive factors and multilingual phonology to the lack thereof – but the majority insisted on the necessity to extend the research scope and explore these relationships. An extended review of individual differences in multilingual phonology acquisition research in the areas relevant to the current study such as phonological working memory, inhibitory control, psychotypology, biographical data and language proficiency is presented in Chapter 2.

1.6. Models of multilingual acquisition

The mechanisms behind language acquisition, processing and use are characterised by a high degree of complexity and variation. To organise research outcomes and offer generalisable conceptualisations, the researchers in the field such as first language acquisition, bilingualism or psychotypology develop models outlining these processes. As a fairly young field, multilingualism has relied on the adaptation of the existing

models, but it has also produced several proposals based on multilingual data. Berthele (2021) outlined the conceptual principles which shaped multilingual modelling and distinguished three approaches: formal, holistic-catalytic and no-boundaries. The formal approaches are predominantly rooted within syntactic and morphosyntactic research; they are interested in the examination of the particular linguistic entities (such as syntactic transfer) and outline the clear distinction between the examined languages (evidenced by models of multilingual acquisition proposed by Flynn et al. 2004, Bardel and Falk 2007, Rothman 2015). The holistic-catalytic approaches are not rooted within formal theories and classifications; they offer proficiency-oriented evaluation and examine the interactions between the languages from the perspectives of multicompetence, recognising the unique characteristics of multilingualism (the approach visible in the models proposed by Herdina and Jessner 2002, Hufeisen 2005). The no-boundaries approaches question the role of language counts and clear distinctions between the entities with one's linguistic repertoire in favour of more dynamic and fuzzy multilingual interaction associated with translanguaging (e.g. García and Li Wei 2014). Models grounded in two of these approaches – formal and holistic-catalytic – found their way into the classification of the models proposed by Wrembel (2015: 58-78), which distinguished three main categories:

- classical models - developed in other fields of linguistics and adapted to multilingualism e.g. De Bot's multilingual production model (1992), Green's activation/inhibition model (Green 1986, Green 1998, Green and Wei 2014), Grosjean's language mode (Grosjean 1998).
- models of multilingual speech - based on data and observation rather than pre-existing psycholinguistic models or theories e.g. Hufeisen's Factor Model (Hufeisen 1998, Hufeisen and Marx 2007), Hammarberg's Role-Function Model (Hammarberg 2009).
- third language acquisition models - based on theory and grounded in data; currently mostly of a generative background, supported by morphology and syntax data e.g. Flynn et al.'s Cumulative-Enhancement Model (Flynn et al. 2004), Bardel and Falk's L2 Status Factor Model (Bardel and Falk 2007, Bardel and Falk 2012), Rothman's Typological Primacy Model (2010, 2011, Rothman 2013, Rothman 2015) or the Linguistic Proximity Model (Westergaard et al. 2016).

The discussion of the models – both adapted and developed within the field of multilingualism - will be organised around the classification outlined above as it was created from the perspective of multilingual phonology research. However, as pointed out by Wrembel (2015: 78), no model was specifically tailored to encompass the process of multilingual acquisition of phonology; this claim can be also extended to the current state of the art.

In the absence of the models dedicated solely to multilingual phonology, the subsequent sections overview the models which might be particularly relevant to the current study in the areas of its scope and design, the tested group, and the examined aspects of multilingual phonological development and the role of individual differences. Even though the study does not implicitly set out to test the models, they may serve as a frame of reference for the discussion of the obtained results. The classification of the models follows the groupings outlined above and explains the reasoning behind the selection.

1.6.1. Selected classical models of foreign language acquisition

The classical models of foreign language acquisition were originally developed in neighbouring fields, however, they can potentially be extended to other fields of multilingualism. One of such models which might be particularly relevant to the present study due to its interest in the link between multilingual development and inhibitory control is Green's activation/inhibition model (Green 1986, 1993, 1998, Green and Wei 2014). Another classical model which might also be appropriate to include in the current study is a framework particularly attuned to testing developing phonology, namely, Perceptual Assimilation Model (Best 1995, Best and Tyler 2007).

Green's activation/inhibition model proposes that languages in a bilingual/multilingual mind remain in different states of activation i.e. the selected language is the one currently in use, the active language is not selected, but remains active, and the dormant language is not selected and has low activation. By employing the notions of activation, selection and inhibition Green's model outlined the interaction between the languages in the mind of a speaker. As the notion of inhibition and individual differences is further explored in section 2.3.1 (Chapter 2), a more detailed description of

the model along with the multilingual elaboration proposed by Festman (2008, 2018) is included there.

The Perceptual Assimilation Model (Best 1995, Best and Tyler 2007) offers a conceptualisation of how native language affects the perception of non-native speech in relation to native phonemes. The non-native segments can be assimilated to native-like categories in a number of ways such as assimilation to a native category, assimilation as an unrecognisable sound or assimilation as a non-speech sound. Moreover, PAM predicts the patterns of assimilation of non-native contrasts and outlines possible assimilation types. The first type is *two-category assimilation*; the phonemes of a non-native contrast are assimilated to separate L1 categories and are expected to be successfully discriminated. The second type is *category-goodness difference*; non-native sounds are assimilated to the same category, however, one is perceived as a better fit. Discrimination is anticipated to be very good, depending on the category goodness of each sound. In *single-category assimilation* both sounds are assimilated to the same native category but are perceived as equally fitting or different; their discrimination is expected to be unsuccessful. The *unrecognisable* pattern refers to a situation in which both sounds are categorised within the native phonetic space, but outside of any category. Such assimilation may result in very poor or very good discrimination, depending on the distance from each other and the native categories within the phonetic space. In the *uncategorised vs categorised* pattern, one sound is assimilated to a native category, whereas the other falls outside the native categories; in such a case, discrimination is expected to be very good. The final pattern is *nonassimilable,* where both non-native sounds are assimilated as non-speech sounds. PAM-2 model (Best and Tyler 2007), employs PAM assimilation categories and adjusts them to predict the acquisition of L2 categories during the process of non-native language acquisition. Tyler (2019: 611-12) outlines the potential process of assimilation in L2 learning in the following way: discrimination of non-native contrasts can change as a result of the acquisition of a new category. Prior to L2 learning, a non-native contrast can be perceived as a different L1 category, with an L1 contrast as a basis for discrimination. After the onset of L2 acquisition, a common L1-L2 category may be created for each sound. A pronounced distinction between the L1 and L2 versions of a given contrast may improve the perception of differences; if these contrasts belong to different regions of phonetic space, then separate L1 and L2 categories will be created. More divergent or unusual sounds

may be first established as a potential L2 variant. However, if the L1 and L2 contrasts are similar, then the learners are unlikely to acquire a new phonological category because the already existing L1 contrast serves for the L2 as well. The success in learning L2 contrasts may be modified by numerous factors such as experience, exposure and resulting perceptual skill; more experienced learners are more likely to learn to discriminate with improved accuracy (Best and Tyler 2007). PAM and PAM-2 do not account for multilingual perception, however, a study by Wrembel et. al. 2019 tested to model against multilingual data. The study examined L3 sibilant and vowels perception in L1 German, L2 English and L3 Polish adolescents. The obtained results showed that the participants demonstrated a high perceptual acuteness, attributed to their multilingual language learning experience. New contrasts, non-existent in L1, were not assimilated to L1 counterparts in the process of single-category assimilation, associated with low discrimination of contrasts. The participants used their perceptual sensitivity, developed in the process of multilingual acquisition to bypass this type of assimilation. In the context of PAM-2, the results suggest that L3 learners might categorise novel sounds similarly to advanced L2 learners. In the context of the current study, the framework of PAM and PAM-2 and their extension to L3 might be useful in the explanation and discussion of the results obtained in the multilingual speech perception tests.

1.6.2. Selected models of multilingual speech

The models of multilingual speech are predominantly organised around data and observation, without strong references to theories and frameworks. The conceptualisation which might be particularly applicable in the discussion and explanation of the results obtained in the current study, especially the potential interaction between the L2 and L3 of the participants, has been proposed by De Angelis (2007) and is referred to as *combined CLI*. De Angelis (2007:21-26) points out that the reliance on one source language for multilingual learners is likely to decrease as more linguistic information from the subsequent language acquisition is added. The perception of CLI as a one-to-one relationship between the source and the target language describes bilingualism, however, multilingualism opens more possibilities such as a many-to-one relationship, implying that the target language can be influenced by the accumulated knowledge from multiple

previous languages. Moreover, the many-to-one relationship can have a sequential nature e.g. L1 influences L2, which in turn influences the L3 during the process of language acquisition. Moreover, combined CLI can be modified by the relatedness and formal similarities of the languages in the repertoire of the multilingual individuals. Several studies within the field of multilingual phonology provided evidence for the existence of combined CLI in the area. Among the examples, there is a study by Wrembel (2014b), who tested the production of VOT values in L3 French and L3 German. The L3 VOT values corresponded with the intermediate values between L1 Polish and L2 English VOT, aligning with the assumptions of combined CLI. In another study, Sypiańska (2016b) examined the acquisition of vowels in L1 Polish, L2 Danish and L3 English; the characteristics of the production of L3 vowels were ascribed to the influence of L1 and L2 CLI. Liu and Lin (2021) observed that the perception of stops in L3 Russian or Japanese (short lag vs. voicing lead languages) was influenced by the combined patterns in L1 Mandarin Chinese and L2 English (short lag vs. long lag). The phenomenon of combined CLI and its grounding in L3 phonology research can be helpful in the explanation of the VOT patterns produced by the participants in the current study.

1.6.3. Selected models of third language acquisition

The existing models of third language acquisition are rooted in theory and data; their predictions are tied to theoretical assumptions and supported by evidence. The Cumulative Enhancement Model (Flynn et al. 2004) and the Natural Growth Theory of Acquisition model (Dziubalska-Kołaczyk and Wrembel 2022) were identified as the models, which can be used as frameworks for the discussion of the results obtained in the current study.

The Cumulative Enhancement Model (CEM; Flynn et al. 2004) is a model of third language acquisition based on evidence from syntactic studies. CEM proposes that language learning is a cumulative process; L1 is not the sole source of influence and all languages in the repertoire of a learner can influence the subsequent acquisition. Moreover, CEM does not follow the deficit, negative transfer approach to CLI. Prior language knowledge is envisioned as essentially non-redundant, enhancing or remaining neutral in the property-by-property construction of L3. The assumptions of the CEM

model were supported by the evidence from a study of the acquisition of relative clauses in L1 Kazakh, L2 Russian, L3 English children and adult speakers (Flynn et al. 2004) and by 2 groups of L1 German, L2 English and L1 Hungarian, L2 German, L3 English learners at different proficiency levels (Berkes and Flynn 2012), which supported the claim about the lack of L1 primacy and cumulative influence of linguistic knowledge of previous languages. The strong version of CEM was challenged by Rothman and Cabrelli Amaro (2007, 2010) and Rothman (2011) and their evidence for the non-facilitative influence of prior linguistic knowledge. However, the weaker version of the model, implying that the transfer in L3 may originate from both previously acquired languages might be more generalisable and applicable (Wrembel 2015: 70), possibly accounting for redundancies and non-facilitative influences. Consequently, the weak version of the Cumulative Enhancement Model may serve as a starting point for the discussion about the potential cumulative influence from both the L1 and L2 of the participants in the acquisition of L3 phonology. Moreover, its assumption of the property-by-property construction of L3 fits the longitudinal design of the current study, which (among other aims) intended to observe the potential interaction between the already existing structures of the L2 and the newly developing L3. Additionally, the current study may benefit from considering the obtained data against CEM, which was tested with both children and adult participants.

Natural Growth Theory of Acquisition Dziubalska-Kołaczyk and Wrembel (2022) is an explanatory model grounded within the framework of Natural Phonology and Complexity theory and supported by phonological data. The model is holistic; it attempts to include the role of linguistic input, typology, language universals and learning context. The model proposes three main predictions: (1) the three variables such as L1, Ln and universals have a role in the process of acquisition, which is modified by extralinguistic factors, (2) the age of acquisition influences the complexity of relations between the linguistic variables and extralinguistic factors; (3) articulatory routines and perceptual constraints are main factors driving CLI at the early stages of acquisition (level 1), metalinguistic learning occurs during later stages (level 2). The authors of the NGTA framework translated their predictions to specific, multilingual, data-related hypotheses: (1) low proficiency at early stages of L3 acquisition activates hybrid VOT based on L1 and L2 values, target values emerge with increased proficiency, (2) early stages of L3 acquisition are characterised by L2-accented speech, later L1 becomes prominent, (3)

attainment in L3 is modified by the characteristics of input, training, individual factors and metalinguistic awareness, (4) metalinguistic awareness does not ensure success in the light of universal phonemic difficulties, (5) metalinguistic awareness, the recency of use and order of acquisition reaffirm the role of L2 in L3 acquisition. The VOT production data supported the first and the third of the main predictions and data hypothesis 1. The remaining predictions are still to be substantiated by more evidence. The NGTA model may be particularly relevant in the discussion of the results of the current study for a number of reasons. Firstly, the model is a developmental framework, accounting for the subsequent stages of multilingual language acquisition; such an approach aligns with the developmental character of the current study. Secondly, the predictions of the model were compared against the phonological data of VOT production, which is also one of the data types collected in the current study. These two factors may render the NGTA model particularly applicable in the discussion of the obtained results.

1.7. Summary

Chapter 1 presented an overview of issues associated with the multilingual acquisition of phonetics and phonology, grounding the issue within the main themes present in the research field of multilingualism. The chapter discussed the challenges associated with defining multilingualism and multilingualism as well as their potential classifications. Moreover, Chapter 1 outlined the differences between L2 and L3 acquisition. Furthermore, the chapter overviewed the topic of cross-linguistic influence in the context of multilingualism. Two final sections offered a review of multilingual phonology research and the existing models of acquisition related to the context of the current study.

Chapter 2: Individual differences in multilingual acquisition

2.1. Introductory remarks

Research into individual differences has originated in the field of psychology and attempts to highlight traits or characteristics which distinguish one person from another. From the perspective of skills acquisition and learning, these differences may refer to various aptitudes and abilities as well as more stable characteristics such as age or gender (Sullivan 2009: 252). As far as language acquisition is concerned, the study of individual differences may answer numerous questions associated with the varying degrees of language competence across individuals. A research framework accounting for the factors driving individual differences seems to be particularly suitable to encompass the complexities of multilingual language acquisition; attention to cognitive or environmental influences may offer at least partial explanation for the processes and phenomena observed through testing the language perception and production of multilingual individuals.

In the developing field of multilingual acquisition of phonology, the researchers investigated, among other factors, the individual differences associated with cognitive capacities or language use and exposure (e.g. Lev-Ari and Peperkamp 2013, Mora 2017, Sun and Zhang 2020). The current dissertation employed the research perspective of individual differences to investigate the development of multilingual phonology. Consequently, to provide a background for further empirical investigation, Chapter 2 elaborates on the role of the selected factors potentially underlying individual differences in multilingual phonological development. The selection of the said factors is determined by their relevance to the profile of the participants, i.e. adolescent sequential multilinguals acquiring their L2 and L3 through formal instruction. The subsequent sections focus on the role of age, cognitive factors (inhibitory control and phonological working memory), psychotypology and learning environment.

2.2. The role of age

Age constitutes one of the most crucial factors affecting the process of language learning, which has been extensively examined in the fields of first and second language acquisition. Researchers have identified certain stages of linguistic development linked to age, that might determine (in the case of L1) or potentially shape (in the case of L2) the process of acquisition, deciding about its outcomes. The current thesis is predominantly focused on adolescent sequential multilinguals, therefore, the subsequent section will focus on the role of the age factor for L2 and beyond, with the implications for age-related stages in cognitive development.

Age provides a framework to classify one's bilingual or multilingual profile. *Early bilinguals* start acquiring their L2 between the ages of 0 and 3, whereas *early multilinguals* begin to acquire their L2/L3 between the ages of 0 and 3 up to 6. The onset of L2/L3 acquisition for *late bilinguals* and *multilinguals* occurs after age 6 or beyond childhood (Festman 2018: 237-9). Age is also related to a chronological classification; an individual who started acquiring two or more languages at the same time before the onset of the critical period may be defined as a *simultaneous bi-/multilingual*, whereas a learner who started acquiring their languages one after another can be described as a *sequential bi-/multilingual* (De Angelis 2007, Butler 2013: 112). As partially indicated by the categories of early vs. late bilingual, the age of language acquisition is tied to the concept of a critical period – a hypothesised cut-off point beyond which the acquisition of language becomes virtually impossible (in the case of L1) or challenging (L2 and further) (Lenneberg 1967, Hoff 2013: 46, Hartshorne et al. 2018). Numerous researchers pointed to the existence of the pervasive claims that the earlier the age of language acquisition onset, the higher the later proficiency (Singleton and Lengyel 1995, Kadyamusuma et al. 2018: 275), based predominantly on the evidence from naturalistic and immersion settings. However, a number of studies conducted in a formal education setting showed an advantage of later learners over younger ones in the areas of language learning rate (Muñoz 2006), elicited grammar imitation and perceptive vocabulary (Myles 2017), receptive vocabulary and grammar (De Bot 2014). In the case of the acquisition of phonology, the prevalence of a non-native accent in late second language learners led to the assumption that this modality is particularly vulnerable to the closing gap of a critical period (Oyama 1976, Williams 1979). Such claims, however, were contradicted by a body

of work showing that the relationship between age and language attainment is more complex. For instance, several phonological and morpho-syntactic studies exploring the production of target sounds by L2 learners conducted by Flege et al. (1987, 1999, 1992), offered the results contradicting the decisive role of the early onset in language acquisition. Such a pattern was observed in several later studies comparing adults and children in terms of their phonological perception and production in L2 (e.g. Baker-Smemoe et al. 2008, Kopečková et al. 2019). Flege and McKay (2011) put forward four age-related hypotheses, which referred to several assumptions associated with the role of age in the acquisition of phonology. The first hypothesis is *the maturational constraint hypothesis,* which states that in the course of human maturation mechanisms of language acquisition are becoming less effective. The second hypothesis is *the cognitive development hypothesis* proposes that language learning skills decrease throughout life due to the diminishing cognitive abilities associated with these skills. The third hypothesis was derived from Speech Learning Model (Flege 2003, 2007, 2009) and states that with the development of the L1 systems, the cross-language phonetic interference grows stronger; as a result, the older speakers may be less likely to develop new categories for L2 sounds that are similar to those in L1. The fourth hypothesis is *the input differences between early and late learners hypothesis,* which proposes that foreign language performance is predicted by the amount of input, which varies depending on years of use and frequency of use. Subsequently, it is stated that all hypotheses can predict the differences between the early and late learners and age of acquisition is considered a macro-variable, linked to several underlying variables associated with cognition, maturation or exposure. Consequently, age cannot be treated as an ultimate predictor of the outcomes of bi-/multilingual language acquisition. In the light of this inconclusive evidence, Singleton (2001: 86) proposes to turn to a wider category of age-related factors, which may explain the developmental trajectories and individual differences between speakers.

One of such age-related factors is associated with the stages of cognitive development. Adolescence, which is of particular interest to the current study, is a period of transition between childhood and adulthood. Bunge and Wright (2007) and Cromer et al. (2015) suggest that the neural networks for such cognitive capacities as psychomotor function, attention, working memory and learning are established by the end of childhood, however, they undergo refinement during adolescence and become salient in the

adulthood. In the case of one of the cognitive capacities tested in the current study, namely, inhibitory control, the developmental path is similar. The neural circuits necessary for inhibitory control are present in adolescence, however, the evidence suggests that teens have significantly less control over its mechanisms compared to adults (Aïte et al. 2016, Constantinidis and Luna 2019). Fosco et al. (2019) provided evidence for the acceleration of inhibitory control development during early adolescence. Phonological working memory, another cognitive factor investigated in the study, is also significantly modulated by age; it undergoes a sharp increase in its capacity up to the age of 8-10 to achieve asymptotic levels around the age of 11-12 (Gathercole 1999: 441). Consequently, adolescence may be considered an eventful period in cognitive development, and potentially an interesting stage to examine empirically in the context of individual differences.

As it was outlined above, age has significant implications for language learning. However, especially in the case of bi- and multilingualism, it cannot be treated as an absolute determinant of success in the process of acquisition. However, along with age-related developmental factors, it can still be researched as a potential source of individual differences. Moreover, examination of different age groups, especially the under-researched ones, may inform the general discussion about various stages of language development and provide further material for comparison between the studies.

2.3. The role of cognitive factors

Cognitive abilities, as a result of modulation by a host of neural, genetic and behavioural factors, exhibit a great degree of individual variation (Boogert et al. 2018). Bendall et al. (2016) pointed to the evidence suggesting that individuals exhibit cognitive styles, habitual approaches to tasks and situations related to cognitive processes such as decision making, problem solving, perception or attention. As suggested by Yu and Zellou (2018), cognitive styles might have a substantial influence on the course and outcome of the processes associated with executive function. Moreover, while correlated with the linguistic measurements of phonological development, cognitive styles may highlight individual differences. The current thesis will focus on two cognitive factors partaking in the formation of one's cognitive style, namely, inhibitory control and phonological

working memory. These two subcomponents of executive control are examined in relation to individual differences in the phonological development of multilinguals. The sections below discuss the aspects of inhibitory control and phonological working memory and their potential to highlight individual differences in the development of multilingual phonology through the lens of previous work. Both inhibitory control and phonological working memory were preliminarily presented in two earlier publications by Krzysik and Wrembel (2019) and Krzysik (2020), which also included analyses of small portions of the data gathered in the current study. However, the subsequent sections expand on the topics of inhibitory control and phonological working memory, further enabling an extended analysis and discussion of the results obtained in all three testing sessions.

2.3.1. Inhibitory control

Inhibitory control is frequently conceptualised as an ability to control impulses and prepotent responses as well as to suppress irrelevant information. Along with interference control, working memory and cognitive flexibility, inhibitory control was included as one of the subcomponents of executive function, a group of top-down mental processes required while focusing or paying attention when automatic or intuitive responses are inexpedient or impossible (Miyake et al. 2000, Diamond 2013). The processes underlying inhibitory control can be subdivided into three functionalities such as resistance to distractor interference (associated with the inhibition of distractors), prepotent response inhibition (i.e. the suppression of automatic responses) and resistance to Proactive Interference (i.e. the suppression of memory intrusions from information previously considered as relevant) (Friedman and Miyake 2004). As far as language is concerned, inhibitory control has been positioned as a significant factor underlying mechanisms of language control, with the evidence suggesting the existence of language-specific inhibition mechanisms (Borragan et al. 2018). One of the formative conceptualizations of the inhibitory control mechanism in language was developed by Green (1986, 1993, 1998), where inhibition serves there as a mechanism associated with language activation. The model classifies the languages in a repertoire of the speaker on the basis of their level of activation as *selected* (the highest activation, language currently in use, one at a time),

active (lower activation, not in use), and *dormant* (the lowest activation, not in use) (Green 1986: 215). Language selection requires the inhibition of the active language in one's repertoire; the inhibitory load is directly proportional to the proficiency in the active language. Consequently, linguistic performance may be influenced by one's inhibitory control capacity, which is required to balance out the influence of the active linguistic system or systems. The amount of inhibition required to suppress a language not currently in use may also be related to language proficiency level and the characteristics of language acquisition. Festman (2008, 2018) proposed an elaboration of Green's model, accounting for multilingual language processing. Three (or more) languages can be outlined in three combinations of states proposed in Green's model: selective – active – dormant, selected – active – active and selected – dormant – dormant. Festman (2018) underlined the fundamental distinction between bilingual and multilingual language activation; during language activation, a bilingual speaker inhibits only one language, whereas a multilingual speaker inhibits more languages. Unlike the bilingual processing and production, multilingual processing and production may be subjected to interference from more than one language. Consequently, multilingual processing requires more cognitive resources than bilingual processing. Festman (2018) cautiously suggested that the action of switching between the languages and inhibiting those currently not in use may be a potential source of multilingual advantage.

Previous studies examined the processes associated with inhibitory control in the context of various aspects of language use. By employing separate measures of individual differences in inhibitory control, researchers explored such domains and modalities as auditory comprehension (e.g. Blumenfeld and Marian 2011, Desjardins and Fernandez 2017), vocabulary production and verbal fluency (e.g. Linck et al. 2008, Linck et al. 2009, Borragan et al. 2018, Pino Escobar et al. 2018), language-switching (Costa and Sansebastan 2004, Costa et al. 2006, Liu et al. 2019) as well as the so-called bilingual advantage (Bialystok et al. 2005, Poarch and Bialystok 2015). Despite the wealth of research, the nature of inhibitory control language-related mechanisms is still largely unknown and requires more sensitive measures and further testing. A modest fraction of the existing studies investigated the relationship between the inhibitory control and multilingual phonology. Several studies have proposed that VOT patterns in bilinguals may indicate the extent of language activation and inhibition of non-target plosive lag (e.g. Fricke et al. 2015, Jacobs et al. 2016), however, they did not implement separate

measures of inhibitory control. The focus on individual differences reflected by the inclusion of a separate measure was adopted only by a small number of studies. Levi-Ari and Peperkamp (2013) examined the influence of inhibitory control on L2 perception and production of voiceless stops in late English-French sequential bilinguals. The results of the retrieval-induced inhibition task and a battery of phonological tasks (such as sentence reading, free conversation and phoneme categorization) indicated that weaker inhibitory control was related to less English-like perception and production of VOT in English stops. Stronger inhibitory control was associated with the suppression of the regressive CLI from L2 to L1. Another study that explored phonological perception and production in relation to inhibitory control extended its scope to multilinguals. Mora and Darcy (2013) tested three groups: L1 English learners of Spanish, L1 Spanish learners of English and L1 Catalan-Spanish learners of English. The participants completed a retrieval-induced inhibition task, an ABX perceptual categorization task and a delayed repetition task. Increased inhibitory control led to improved accuracy in perception for the two groups of sequential bilinguals, however, it was not observed for the multilingual speakers who simultaneously acquired Catalan and Spanish. Consequently, inhibitory control mechanisms may function differently in speakers more accustomed to language switching. Two subsequent studies by Darcy et al. (2014, 2016) explored the influence of inhibitory control on L2 phonological perception and production of L1 Spanish/L2 English learners and L1 English/L2 Spanish learners. The results of the first study (2014) demonstrated a relationship between the L2 perception accuracy and stronger inhibitory skill in both groups, however, no clear pattern emerged for L2 perception. The second study (2016) on a larger sample indicated that improved accuracy in L2 perception and production of consonants is related to stronger inhibitory control. Further, a host of studies examined the interaction between inhibitory control and multilingual phonology in formal learning environment. Sigmeth et al. (2019) investigated the role of inhibitory control in the L2 perception/production accuracy of L1 German adolescent learners of L2 English and L3 Polish. The results of an inhibitory control measure correlated significantly with the L2 perception accuracy, however, no relationship with the L2 production was observed. Subsequently, Sigmeth and Golin (2018) examined the L2 perception of adolescent L1 German learners of L2 English and L3 Polish via a forced-choice task and correlated their results with the scores obtained through a modified flanker task. The inhibitory control of the participants improved throughout the

longitudinal study, however, there was no significant relationship with perceptual accuracy. Krzysik (2020) tested L2 and L3 production of adolescent L1 Polish learners of L2 English and L3 German. The inhibitory control scores obtained in a modified flanker task correlated significantly with the overall production of accuracy in of rhotics and VOT in L2 as well as L2 and L3 treated jointly, however, no significant relationship was observed for the scores in L3 alone.

The scores obtained in the described studies reflect the complexity of the processes governing inhibitory control in language. The focus on individual differences has been demonstrated as a suitable paradigm facilitating the exploration of inhibitory control mechanisms in the suppression of irrelevant phonological cues or categories in the perception and production of features.

2.3.2. Phonological working memory

Phonological working memory (PWM, also phonological short term memory) encompasses a set of cognitive processes responsible for the ability to temporarily maintain auditory speech-related information (Fiez 2016: 855). PWM constitutes a component of working memory, a capacity associated with the executive function responsible for temporary storing portions of information and supporting thought processes (Baddeley 2003). Based on the working memory model (Baddeley and Hitch 1974, Baddeley 2003), PWM can be associated with the phonological loop – a subcomponent of the working memory system. The Phonological loop encompasses phonological memory capacity, responsible for maintaining memory traces for a few seconds, and the rehearsal process used to refresh memory traces. The capacity of the described system is limited and substantially affected by word length and similarity. Baddeley et al. (1998) proposed that the primary role of the phonological loop is that of language acquisition; it may serve as a storage of unfamiliar sounds which can subsequently be transferred to long-term storage.

Empirical studies acknowledged the hypothesis of PWM's role in language acquisition, i.e. it has been positioned as a predictor of both monolingual and bilingual spoken language development (e.g. Gathercole and Baddeley 1993, Adams and Gathercole 1995, Adams 1996, Masoura and Gathercole 1999, Leclercq and Majerus

2010 , Pierce et al. 2017, Delcenserie et al. 2021). Moreover, PWM deficits have been proposed as a potential correlate of language impairments such as specific language impairment (Bishop 2002, Archibald and Gathercole 2006, Gribau and Schwarz 2007). In relation to the domain of phonology, PWM has been proposed as a significant predictor in the acquisition of phonological features, possibly facilitating the formation of more accurate representations and realisations (Aliaga-García et al. 2011). The researchers investigating this potential relationship frequently resorted to the examination of individual differences via PWM measures. MacKay et al. (2001) examined the identification of English word-initial and word-final consonants in noise by native speakers of Italian. The scores on PWM measures were found to account for respectively 15% and 8% of the variance in subjects' errors in identifying word-final and word-initial consonants. The results suggested that individual differences in PWM may affect the perception of L2 consonants. Further, a study investigating the perceptual accuracy of Spanish-Catalan EFL learners by Aliaga-García et al. (2011) found that participants with higher PWM scores had also improved perceptual accuracy of English monophthongs. In another study focused on individual differences in L2 acquisition of English, Darcy et al (2015) found that higher PWM storage capacity was related to the development in L2 phonological processing of Korean learners of English. In a study with trilinguals, Krzysik and Wrembel (2019) tested the relationship between PWM and the accuracy in the L2 and L3 phonological production of rhotics and final obstruent devoicing. The performance of L1 Polish, L2 English and L3 German adolescent learners on the PWM measure and delayed repetition tasks revealed a moderately significant relationship between the overall L2 and L3 accuracy and PMW capacity.

Based on the evidence provided by the listed studies, PWM seems to be related to various aspects of bilingual and multilingual language processing and production. Moreover, individual differences in PWM capacity may, to a certain extent, predict language development, also in the domain of bi- or multilingual phonology. However, more empirical evidence is required to map the potential patterns of interaction between PWM capacity levels and phonological accuracy.

2.4. The role of psychotypology

Psychotypology is a concept describing the individual perception of language distance proposed by Kellerman (1977, 1978, 1983) in his studies of cross-linguistic influence in L1 Dutch students of English. In contrast to typological and genetic classifications of languages, psychotypology is based on individual perceived similarity of particular aspects of given languages. Despite its subjective nature, it can still be convergent with the systematic classifications. Kellerman identified psychotypology also as one of the constraints of linguistic transferability; the perceived distance may be related to the likelihood of cross-linguistic influence from L1 to L2. The notion of psychotypology was further explored by Ringbom (1986), who acknowledged its potential role in cross-linguistic influence as a factor compensating for gaps in linguistic knowledge in a given L2, especially in the context of beginners. Moreover, Ringbom (2007) recognised psychotypology, in contrast to actual typology, as fluid and subject to change throughout language learning.

Psychotypology has also been acknowledged as a significant factor in multilingual studies. De Angelis (2007: 23-24) pointed out that the concept of psychotypology was devised on the basis of bilingual studies, therefore, it might not reflect the influence of psychotypology on multiple languages. Furthermore, she elaborated on the existence of various CLI patterns among multilingual learners, which either confirm or disprove the claim that the likelihood of influence is inversely proportional to the perceived language distance. Psychotypology in multilingualism was also approached from a generative perspective in the works of Rothman (2011, 2013, 2015), who proposed the Typological Primacy Model, a framework ascribing no privileged status to either L1 or L3 in the initial state transfer in morphosyntax. TPM implies that either actual typology or psychotypology constitute the main constraint for language transfer. The internal parser uses the hierarchy of cues (the lexicon, phonological cues, functional morphology and syntactic structure) and decides about the proximity of L1 and L2 to L3 and selects which system to transfer, resulting in either positive or negative CLI.

Even though researchers frequently underscored the role of psychotypology in the process of language acquisition, numerous studies have assumed its influence without directly quantifying it as a variable (e.g. De Angelis 2005, Jessner 2006, Singelton and Ó Laoire 2006). Several researchers decided to include a measure of psychotypology in a

form of a questionnaire with questions about the perceived similarity between languages tested in a given study (e.g. Hall et al. 2009, Lindqvist 2015), think-aloud protocols (e.g. Jessner 2006, Wrembel 2015) or magnitude estimation measures focusing on the perceived distance (Xia 2017). The majority of the measures mentioned above focused on a binary comparison between two languages, which might potentially prove insufficient or limiting for multilingual contexts. In order to overcome this constraint, Nelson et al. (2021) introduced a Visualised Language Distance Measure (ViLDiM) aimed at capturing individually perceived language distance in a multidimensional way, avoiding binary comparisons (see Chapter 4 for a detailed description of the measure). Moreover, the aforementioned study conducted by Nelson et al. (2021) was one of the first studies focusing on psychotypology in the context of phonological similarities.

Psychotypology as a measure of one's perception of linguistic distance may offer a unique insight into individual differences in multilingual phonology by pointing at potential differences in linguistic performance related to perceived proximity. Moreover, in the context of a longitudinal study, its fluidity may aptly reflect the undergoing changes in the subsequent stages of multilingual phonology development.

2.5. The role of environmental factors

Multilingual acquisition of phonology may not only be influenced by individual differences in developmental or cognitive factors, but also by the immediate surroundings of a given speaker. Such aspects as the context of language use, the type of instruction, the quantity and quality of language exposure and input or dialectal variation differ across the speakers and the acquisition settings. Consequently, these differences may influence the shape and outcomes of linguistic development. Since the participants in the present study develop in a homogenous linguistic landscape of a medium town in Western Poland and acquire their L2 and L3 at school, the subsections will focus on two environmental factors which appear particularly significant, i.e. foreign language instruction and language exposure.

2.5.1. Formal foreign language instruction

The acquisition of a non-native language outside of its environment is referred to as foreign language instruction (Moeller and Catalano 2015: 327). Among the prerequisites of formal foreign language instruction, The Council of Europe lists the presence of a syllabus, learning goals and learning outcome assessment ("Formal, non-formal and informal learning" 2020). It is delivered predominantly in formalised settings such as schools, universities or language classes. The current subsection characterises foreign language instruction in Polish public primary schools to explain its potential role in the multilingual development of the participants in the current study.

In order to fully understand the implications of foreign language instruction for the current study, it is necessary to outline its characteristics in the Polish context. The number of foreign language classes is specified in the regulation of the Ministry of National Education. The first foreign language is introduced in grade 1 of primary school (age 6-7), with a minimum number of 180 hours[1] until grade 3 (age 8-9). Such a number of hours results in 2 classes per week. From grade 4 (age 9-10) till the end of primary education (grade 8), the number of hours of the first language totals 450, which results in 3 classes per week. Third foreign language is introduced in grade 7 (age 12-13); 120 hours result in 2 classes per week. The learning cycles of the first and second foreign language should respectively result in levels A2 and A1 of the Common European Framework of Reference for Languages (Council of Europe 2001). With regards to teaching modes and aims, early language instruction (grades 1-3) is primarily based on learning through play with a gradual introduction of reading, writing or group work (Szpotowicz 2011). In the later stages of primary school (grades 4-8), foreign language instruction is supposed to equip students with communication skills by improving grammar, vocabulary and other aspects of language proficiency (Niedźwiedź et al. 2017). However, the modes and aims outlined above might be realised differently. The survey conducted by Muszyński et al. (2015) in their study on formal English instruction in schools revealed that grammar is the skill most frequently practised during foreign language classes, followed by vocabulary, reading, and writing, with the smallest amount of time devoted to speaking. Moreover, the results showed that teachers dominate communication in the classroom

[1] „Hour" is understood as a 45-minutes unit.

and use a considerable amount of Polish. Besides the amount of instruction, its content and organisation, Muszyński et al. (2015) pointed to teachers' qualifications, language learning tools, and the number of students in a group as factors that also can impact foreign language learning, and by extension, individual differences between students.

The studies examining multilingual development in the context of formal instruction, especially in a largely monolingual social context, have been rather scarce. Some longitudinal studies to date have produced varied results such as great individual variation (Wrembel et al. 2019) or relative stability for more proficient languages and instability of the newly acquired ones (Balas et al. 2019, Wrembel et al. 2020). Moreover, previous studies in the context of formal instruction focused predominantly on the linguistic development as such, with rare instances of studies investigating the role of individual differences (Sigmeth and Golin 2018). Consequently, further longitudinal studies within such contexts are required to shed more light on the multilingual development in formal language instruction contexts.

Formal foreign language instruction is seemingly uniform and compliant with the rules and outlines of the state-regulated curriculum. However, the variance in its organisation and execution may potentially contribute to the individual differences between learners.

2.5.2. Language exposure

Foreign language exposure, broadly understood as the learner's contact with their Ln, is a factor characterised by a great individual variability. Despite its crucial role in the acquisition of the target language, the role of exposure and its qualitative and quantitative aspects are still subject to debate (Loewen and Reinders 2011: 63). The current subsection will explore the role of quality and quantity of language exposure and its implications for individual differences.

Every speaker receives different amounts of exposure to each language in their repertoire. The quantity of language exposure may not only decide about the success or failure in language acquisition but also about one's language dynamics. Speakers receiving more input in a given language and less in the remaining ones may be defined as unbalanced bi-/multilinguals. Speakers who experience fairly equal exposure to the

languages in their repertoire may be referred to as balanced bi-/multilinguals (Colman 2015: 28, 792, Festman 2018: 237). Furthermore, the outcome of language exposure is greatly individualised; Dewaele (2009: 624) pointed out that the same amounts of language instruction and input may result in different levels of competence. In regards to quality, speakers may be exposed to language from different sources (or learning resources) and of diversified varieties. Taking into account the functional aspects, Maneva (2004) distinguished between active and passive exposure. During active exposure, the speaker participates in communication in a given language, whereas in the state of passive exposure, they assume the role of a listener. Moreover, she distinguishes the types of exposure from the perspective of social interactions as egalitarian (peer to peer) and non-egalitarian (from adult to child). As acknowledged by language exposure and use questionnaires such as The Language History Questionnaire (Li et al. 2006) or The Language Experience and Proficiency Questionnaire (Marian et al. 2007), the quality of language exposure may be also related to the technological device currently in use such as TV, radio or the Internet.

Several studies incorporated language exposure as a factor underlying individual differences between the speakers. The quantitative and qualitative aspect of exposure was included in several studies and provided evidence for the exposure as a factor underlying individual differences (Hurtado et al. 2008, Marchman et al. 2010, Hurtado et al. 2014). An example particularly relevant to the current thesis (due to its relation to formal instruction) comes from a study conducted by De Wilde et al. (2020), who investigated the role of language exposure in the acquisition of English by a group of Dutch children. The results revealed that the most beneficial out-of-school language exposure modes were gaming, use of social media and speaking. The researchers also pointed out a substantial amount of individual differences between the participants.

Language exposure and its quantitative and qualitative characteristics may have a considerable effect on the outcome of the process of language development. Moreover, it may also indicate the source of individual differences between the participants.

2.6. Summary

Chapter 2 presented a host of selected factors potentially underlying individual differences in language development, such as age, inhibitory control and phonological working memory, psychotypology, language instruction and exposure. The review of the studies investigating the role of individual differences, especially in the context of multilingual development, revealed the need for further research. For instance, the studies investigating the role of inhibitory control or phonological working memory in relation to multilingual speech perception and production still constitute a minority in multilingual research. Moreover, there is also a gap in research conducted in the context of formal instruction, which differs significantly from the more naturalistic settings. Additionally, the overview of factors such as inhibitory control, phonological working memory or psychotypology highlighted the complexity of individual differences in language acquisition. The multitude of said factors and their potential interactions may require a more multimodal approach, integrating more than one measure of individual differences in the testing batteries of future studies.

Chapter 3: Phonological systems of Polish, English and German

3.1. Introductory remarks

Chapter 3 outlines the phonological systems of the languages examined in the current longitudinal study, namely, Polish, English and German. Sections 3.2 to 3.4 offer general descriptions of the said systems, focusing on the aspects studied in the experimental part of the thesis such as vowel length, voice onset time (VOT) and rhotic consonants. Section 3.5 presents a comparison of the selected characteristics of Polish, English and German, which are tested in the tasks described in experimental chapters of the thesis.

3.2. Phonological system of Polish

Polish is classified as a Slavic language, belonging to the West Slavic subdivision. Further subgroupings place Polish within a Lechitic group, along with Kashubian and the Sorbian languages. The phonological system of Polish displays a number of characteristics such as a rich consonant inventory or palatalization, which are shared with other Slavic languages (Browne and Ivanow 2019). However, Dziubalska-Kołaczyk and Walczak (2011: 4) identify also its particularly outstanding features such as the preservation of nasalised vowels.

In quantitative terms, the phonological system of Polish encompasses 6 vowels and 31 consonants (Jassem 2003). As outlined in the World Atlas of Language Structures Online (Dryer and Haspelmath 2013), the discrepancy between the number of consonants and vowels in the inventory of Polish produces a moderately high consonant to vowel ratio[2]. As pointed out by Orzechowska (2019: 5), the C/VQ ratio based on Jassem's classification is close to the average of five, however, it may be subject to fluctuation,

[2] The C/VQ ratio is calculated by dividing the number of consonants by the number of vowels existing in a given language; the values of 4.5 or higher but less than 6.5 are granted the label of "moderately high" (Maddieson 2013).

depending on the approach to the palatal consonants and nasalised vowels. The following subsections outline the vowels and consonants in the segmental inventory of Polish.

3.2.1. Polish vowels

The vocalic system of Standard Polish is relatively small and encompasses six oral monophthong vowels /i, ɨ, u, e, a, o/ (Ostaszewska and Tambor 2000: 33, Jassem 2003). Polish features also nasal vowels of a disputed status, however, further discussion regarding this topic is of no relevance to the current thesis, as they are not examined in the empirical part. The segment of Polish high vowels includes /i/ - front high unrounded, /ɨ/ – central high unrounded and /u/. – back high rounded. Mid vowels include mid front unrounded /e/ and mid-back rounded /o/. The low vowel /a/ can be described as central unrounded (Gussmann 2007: 2, Nimz 2016: 80). Figure 1 illustrates the positions of the oral vowels on the IPA quadrilateral.

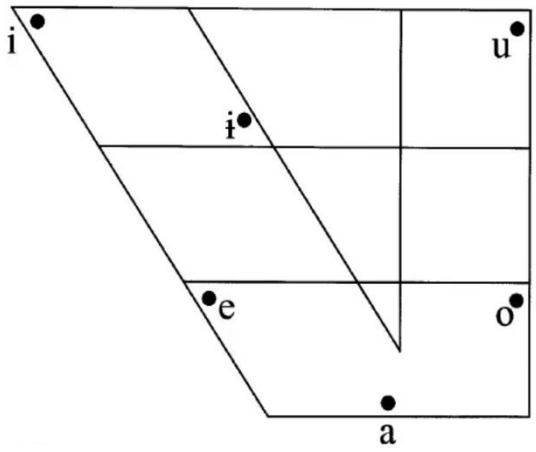


Figure 1. The oral vowels of Polish (Jassem 2003).

Polish vowels display no distinction with regard to tenseness and duration (Balas 2018). As pointed out by Rojczyk and Porzuczek (2012), quantitative vowel reduction in Polish has been considered insignificant, however, there is evidence for reduction in unstressed contexts. These reductions are not identified as phonological, but rather as the artefacts of the processes related to speech kinesthetics (Rojczyk 2012: 63).

3.2.2. Polish consonants

The descriptions of the Standard Polish consonant system have frequently stressed its breadth; the repertoire has been described as rich (Jassem 2003), complex (Gussmann 2007) or large (Orzechowska 2019). According to the classification of the WALS online, based on the data from 563 languages, the Polish consonant inventory falls within the 'moderately large' category, encompassing the sets of 26 up to 33 consonants (Maddieson 2013). The classification proposed by Jassem (2003) lists 31 consonants including 8 plosives, 9 fricatives, 6 affricates, 4 nasals, two approximants, a (post)dental lateral and an alveolar rhotics consonant realised as a trill or flap (the inventory is outlined in Table 1).

Table 1. The consonants of Polish (Jassem 2003).

	Labial	Labiodental	(Post)dental	Alveolar	Alveolo-palatal	Palatal	Velar
Plosive	p b		t d			c ɟ	k g
Fricative		f v	s z	ʃ ʒ	ɕ ʑ		x
Affricate			ts dz	tʃ dʒ	tɕ dʑ		
Nasal	m		n		ɲ		ŋ
Lateral			l				
Flap/Trill				r			
	Front			Back			
Approximant	j (j̃)			w (w̃)			

Due to the focus of the present study, the following paragraphs will concentrate on the phenomenon examined in the experimental part of the thesis, i.e., Voice Onset Time (VOT), a feature in the production of plosives referring to the duration of the period occurring between the release of a plosive and the onset of voicing.

Polish contrasts the prevoicing of /b, d, g/ with voiceless unaspirated or faintly aspirated /p, t, k/. Consequently, voiced plosives exhibit negative VOT values, whereas voiceless plosives are characterised by short, positive values (Keating et al. 1981, Rojczyk 2009). Such characteristics conform to the pattern of prevoiced (voicing lead) vs. short lag VOT, attributed to *true voicing* languages. Mean VOT values for the initial voiced stops attested by Kopczyński (1977:72) ranged around -78 ms for /b/, -72 ms for

/d/, -61 ms for /g/ and 37.5 ms for /p/, 33 ms for /t/ and 49 ms for /g/. Keating et al. (1981) observed mean VOT values of -88.5 ms for /b/, -89,9 ms for /d/ and -66,1 ms for /g/ and 21.5 ms for /p/, 27.9 ms for /t/ and 52.7 ms for /k/. In regards to the consistency of the said values, Malisz and Żygis (2015) established that Polish conforms to the pattern of *true voicing* languages; positive VOT remains stable across the speech rates and prevoicing tends to increase in slower speech.

3.3. Phonological system of English

English is a Germanic language, classified as a member of the West Germanic subdivision, spoken by approximately two billion people across the globe, both as a native and a foreign language (Crystal 2008: 5). The high degree of dialectal variation of English is visible not only in its lexicon but also in the extent of phonological diversity. Since General British[3] is commonly used in educational materials in Poland as well as in the teaching materials of the tested group, the following sections will be devoted to this variety.

The phonological system of English is characterised by a wide array of vowels and a significant number of consonants. The quantitative examination of the said system results in a count of 24 consonants, 12 vowels and 7 diphthongs (Cruttenden 2014). Consequently, the WALS V/VQ ratio of English (Dryer and Haspelmath 2013) is low, which expresses a slight disparity between the number of phonemes in the consonantal and vocalic inventories. Subsections 3.3.1 and 3.3.2 offer an outline of the vowels and consonants in the inventory of English.

3.3.1. English vowels

The inventory of English vowels is of a quite considerable size; the set of 13 monophthongs and 7 diphthongs is classified as large (Maddieson 2013) against the

[3] A term adopted by Cruttenden (2014: 78-81) to avoid the term strongly marked by historical and class associations – RP; also referred to as Southern (Standard) British English.

backdrop of the world's languages. Due to the scope of the current thesis, the section does not elaborate on the characteristics of the diphthongs.

The monophthongs of General British include two high front unrounded vowels /iː/ and /ɪ/, two rounded high back vowels /uː/ and /ʊ/, the rounded mid-open back vowel /ɔː/, the rounded low back /ɒ/, two unrounded mid-central vowels /ɜː/ and /ə/, the unrounded mid-high front vowel /e/, the unrounded mid-open front vowel /æ/, the unrounded mid-open central vowel /ʌ/, and the unrounded low back vowel /aː/ (Roach 2004, Cruttenden 2014: 97, Gut 2009: 62). The monophthongs are also divided into short and long ones (indicated by the 'ː' diacritic). The positions of the listed monophthongs are presented on the quadrilateral below (Figure 2).

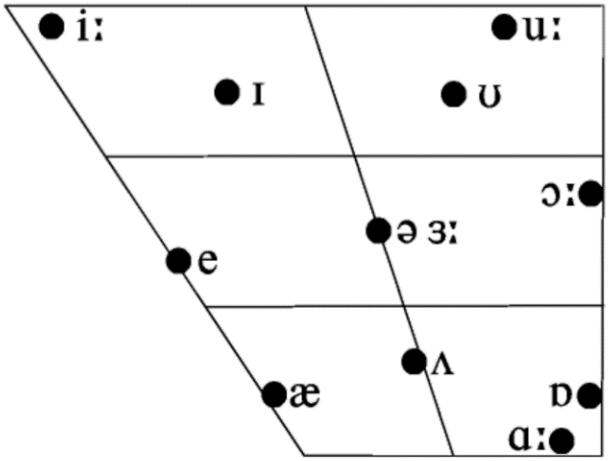


Figure 2. The monophthongs of General British (Roach 2004: 242)

The existence of minimal pairs such as *beat – bit* or *bead – bid* indicates that vowel length in English is phonemic (Gut 2006: 64). However, in the case of /iː/- /ɪ/, /uː/-/ʊ/ and /ɑː/-/ɒ/ pairs, the distinction is based not only on their quantity but also on their different quality (as indicated by their placement in the vowel quadrilateral – Figure 2). The only pair that shares the same vowel quality is /ɜː/ and /ə/, but it differs from the pairs listed above in terms of distribution – /ɜː / occurs only in stressed syllables, whereas /ə/ is found only in unstressed ones (Cruttenden 2014: 100). The quantity of the English vowels is also determined by the characteristics of the subsequent sounds. Vowels are shortened before fortis consonants in the same syllable due to the process of pre-fortis clipping.

When followed by lenis consonants, nasals, /l/ or when occurring word-finally, vowels retain their full length (Collins and Mees 2013: 58).

3.3.2. English consonants

The General British variety of English encompasses 24 consonants, which allows to categorise the size of its inventory as average (Maddieson 2013). The outline of the consonantal system proposed by Roach (2004) lists 6 plosives, two affricates, three nasals, nine fricatives, three approximants and a lateral approximant (presented in Table 2).

Table 2. The consonants of General English (Roach 2004: 240).

	Bilabial	Labiodental	Dental	Alveolar	Post-alveolar	Palatal	Velar	Glottal
Plosive	p b			t d			k g	
Affricate					tʃ dʒ			
Nasal	m			n			ŋ	
Fricative		f v	θ ð	s z	ʃ ʒ			h
Approximant	(w)				r	j	w	
Lateral approximant				l				

English exploits short vs. long lag voicing contrast; voiced plosives /b, d, g / exhibit short lag VOT patterns and voiceless phonemes /p, t, k/ show long lag VOT (Malisz and Żygis 2015). The voiced plosives /b, d, g/ are commonly voiced between two voiced sounds and partially voiced in the initial and final position, in which they can be also voiceless (Ogden 2009: 100). The voiceless plosives are aspirated before stressed vowels, weakly aspirated in front of unstressed vowels and may show no audible release in absolute final positions (Cruttenden 2014: 164). VOT values below 25 ms in stressed single-word utterances can indicate a voiced plosive (Baken and Orlikoff 2000: 275). In a study with British English speakers (Docherty 1992: 145), mean VOT values for voiced plosives in plosive-vowel sequence were experimentally assessed as 18.25 for /b/, 25.64 for /d/ and 30.56 for /g/. In the case of voiceless plosives, Lisker and Abramson (1964) assessed their VOT at 58 ms for /p/, 70 for /t/ and 80 for /k/. VOT patterns in English are

also, to a certain extent, modified by such factors as speech rate, however, with a large degree of individual variation (e.g. Kessinger and Blumstein 1998, Yao 2009).

3.4. Phonological system of German

German is a Germanic language, classified as a member of the West Germanic subdivision, along with such languages as English or Frisian. Spoken by 90 million speakers as a native language and recognised as an official language of Germany, Austria and Switzerland, German is characterised by a considerable level of dialectal variation. The variety predominantly used in the educational context is Standard German or *Hochdeutsch* (Durrell 2003: 11), therefore, the subsequent sections are devoted to this particular variety.

In quantitative terms, the phonological system of German encompasses approximately 24/25 consonants and around 20 vocalic phonemes (including diphthongs and schwa sounds (Boase-Beier and Lodge 2003: 128, Dudenredaktion, Kleiner and Knöbl 2015: 33). Consequently, the WALS V/VQ ratio for Standard German is identified as low (Dryer and Haspelmath 2013). Subsections 3.4.1 and 3.4.2 present the vowels and consonants in the inventory of Standard German.

3.4.1. German vowels

The rich vocalic system of German encompasses around 20 vowels; the disparities in number stem from differences in classification. The inventory includes 15 monophthongs, two unstressed schwa-type sounds and three diphthongs (Dudenredaktion, Kleiner and Knöbl 2015: 33). Figure 3 outlines the diphthongs of Standard German on the quadrilateral.

The inventory of monophthongs, which is of the primary interest to the current study, includes the section of high vowels, namely, the high front unrounded /iː/, the high front rounded /yː/, the near-high unrounded /ɪ/, the near-high front rounded vowel / ʏ/, the near-high near-back rounded vowel /ʊ/ and the high back rounded /uː/. Mid vowels encompass the high-mid front unrounded /eː/, the high-mid front rounded /øː/, the mid

central unrounded /ə/ and the high-mid back rounded /oː/, the low-mid front unrounded /ɛː/, the low-mid front unrounded /ɛ/, the low-mid front rounded /œ/, and the low-mid back rounded /ɔ/. Low vowels include low central unrounded /aː/ and /a/ (Köhler 1999: 87, Nimz 2016: 71).

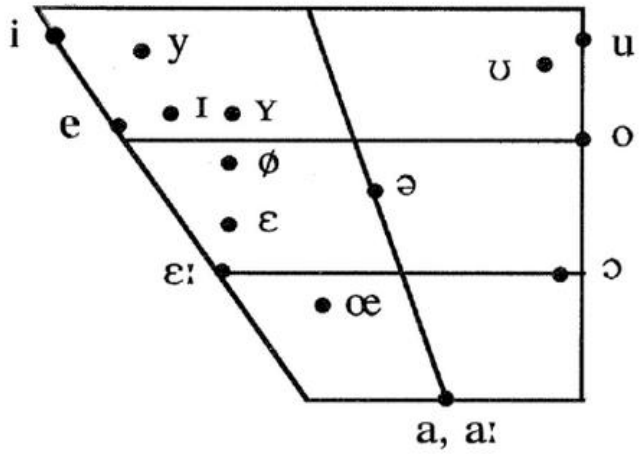


Figure 3. The monophthongs of Standard German (Köhler 1999: 87).

Vowel length in German constitutes a contrastive feature, however, in most cases the difference between short and long vowels is also qualitative. The pair of /aː/ and /a/ differs only in length, whereas the remaining vowel contrasts differ also in their quality (Braunschweiler 1997: 355). Some resources, however, suggest that length is a decisive feature in terms of vowel differentiation (Meinhold and Stock 1982, Köhler 1995:142). Port and O'Dell (1985) reported a difference in vowel length in relation to the immediate neighbourhood of the devoiced word-final obstruent vs. voiceless word-final obstruent, with a longer duration before the former. Braunschweiler (1997) reported a vowel-lengthening effect before the voiced stops for both short and long vowels. However, the length cue has not been proposed as decisive in the distinction between the word pairs with neutralised vs. voiceless obstruent contrast.

3.4.2. German consonants

Based on the classification provided by Kohler (1999, Standard German features 24 consonants, which places its inventory in the category of "average" in terms of size (Maddieson 2013). The table outlined by Kohler (1999) includes seven plosives, three nasals, two trills, ten fricatives, an approximant and a lateral approximant (summarised in Table 3).

Table 3. The consonants of Standard German (adapted from Kohler 1999).

	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Retroflex	Palatal	Velar	Uvular	Pharyngal	Glottal
Plosive	p b			t d				k g			ʔ
Nasal	m			n				ŋ			
Trill				r					ʀ		
Tap or flap									ʁ		h
Fricative		f v		s z	ʃ ʒ		ɕ	x			
Lateral-fricative											
Approximant							j				
Lateral-approximant				l							

German tends to be classified as an aspirating language; voiced plosives /b, d, g/ are produced with a short lag VOT and voiceless plosives /p, t, k/ are produced with a long lag VOT. However, these realisations may change depending on the neighbourhood of the plosives – /b, d, g/ may be voiced in intervocalic positions. Jessen and Ringen (1998) indicated that the dominant pattern for /b, d, g/ in the word-initial intervocalic context is to be voiced, whereas in the utterance initial position - voiceless. When it comes to /p, t, k/, they may be produced with a reduced VOT while preceded by a fricative in the same word or followed by a syllabic nasal (Köhler 1999: 87). The mean VOT values from experimental studies obtained for the voiceless plosives oscillate between 48-59 ms for /p/, around 46 ms for /t/ and between 60-67 ms for /k/ in the context of wordlists, text and sentence reading (Lein et al. 2016: 735).

3.5. Cross-linguistic comparison of the features examined in the study

The selected phonological features examined across Polish, English and German in the experimental part of the study include vowel length and VOT. The following qualities were selected owing to their cross-linguistic contrastive characteristics, and aim to trace the participants' multilingual phonological development over time.

Duration is a significant contrastive feature of vowels across L2 and L3; it can be identified as phonemic. Such properties of the vowels in the two foreign languages (i.e. L2 English and L3 German) form a clear opposition to the system of their L1 Polish, in which vowel quantity is not significant phonemically. Such a contrast may have a potential bearing on the perception and production of the long-short vowel contrasts in the L2 and the L3. Earlier studies, focused predominantly on the L1 Polish learners of L2 English, reveal two possible scenarios concerning perception: duration may be perceived as a valuable cue in discrimination between L2 vowel contrasts (e.g. Bogacka 2004, Rojczyk 2010) or treated as a secondary feature to vowel quality (Balas 2018).

The three languages demonstrate different VOT patterns. Polish displays the contrasts of voicing lead for /b, d, g/ with a short-lag VOT for voiceless /p, t, k/. English is characterised by a contrast of a short lag for voiced plosives and a long lag for voiceless plosives. German represents a similar pattern to English, however, VOT values are shorter. The development of the participants' production may potentially involve L1-like VOT values in L2 and/or L3 plosives, the emergence of hybrid categories constituting an intermediate stage between the non-target and target realisations, or the production of target-like VOT values in L2 and/or L3. Moreover, the potential hybrid values developed for L2 English may resurface in L3 German production as target-like due to shorter VOT lengths in L3 voiceless stops.

A cross-linguistic summary of the features discussed in the current section is available in Table 4.

Table 4. A summary of the L1, L2 and L3 features examined in the study.

language	vowel length	VOT patterns
L1 Polish	non-phonemic	prevoiced vs. short lag
L2 English	Phonemic	short lag vs. long lag
L3 German	Phonemic	short lag vs. long lag

The contrasts listed in Table 4 will be explored in the speech perception and production tasks outlined in Chapter 4 of the current thesis.

Chapter 4: Methods and procedures

4.1. Introductory remarks

Chapter 4 outlines the methodology of the present study, designed to examine the role of individual differences in the phonological acquisition of emergent adolescent sequential multilinguals (L1 – Polish, L2 – English, L3 – German), aged 12-13, over one school year. The study intended to examine the development of the perception and production of selected phonological features, exploring the possible influences of several variables associated with individual differences such as inhibitory control, phonological working memory, language history and use and psychotypology. The research questions formulated in the study concern two main areas – the development of multilingual perception and production and the role of individual differences in the development of multilingual perception and production. The detailed research questions are outlined in section 4.2.

The investigation involved a battery of tasks examining multilingual perception and production, inhibitory control, phonological working memory and psychotypology. These tasks were supplemented by a questionnaire of language history and use and a short task assessing proficiency levels in the L2 and L3 of the participants. The battery was implemented at three testing times within 10 months of the school year 2018/19. The first testing session (T1) took place in late October and early November 2019, the second testing session (T2) spanned the final week of January and the initial week of February, whereas the third testing session (T3) was held between May and June 2019. The present chapter offers a detailed characteristics of the participants (section 4.3) and descriptions of the methodologies and procedures implemented to prepare and conduct the testing sessions in the described study, including the measures of speech perception and production, individual differences and linguistic background (sections 4.4-4.7).

4.2. Research aims and questions

The methodology of the present study was designed to examine the role of individual differences in the phonological acquisition of emergent adolescent sequential multilinguals (aged 12-13), native speakers of Polish. At the time of testing, they were acquiring their L2 (English) for approximately 6 years and started learning their L3 (German), both in the context of formal instruction in a public school system. The study intended to examine the development of the perception and production of the selected phonological features, exploring the possible influences of several variables associated with individual differences such as inhibitory control, phonological working memory, language history and use, and psychotypology.

The research questions formulated in the study can be divided into two groups; the first focuses on the development of multilingual speech perception and production. The second group inquiries about the role of individual differences in the development of multilingual speech perception and production. The following subsections present the research questions and outline the predictions concerning the possible research outcomes.

4.2.1. Research questions – multilingual perception and production

The following section outlines three research questions concerning multilingual speech perception and production, its relation to the contexts of the previous research and the predictions concerning the potential results. The first research question is as follows:

(1) What is the development of L2 perception and production throughout three testing sessions?

As the participants have been acquiring their L2 English for approximately 5 to 6 years; such an extended period of language instruction may indicate that the phonology of the L2 is expected to be rather stable. However, there have been several longitudinal studies investigating the phonetic performance of adolescent sequential multilinguals (L1 Polish, L2 English, L3 German or L1 German, L2 English and L3 Polish), which traced the development of L2 in the context of formal, school instruction, and revealed a

considerable diversity. Concerning perception, researchers have found relative stability of the L2 rhotics (Balas et al. 2019, Wrembel et al. 2022) as well as rhotics and final devoicing (Sigmeth and Golin 2018, Wrembel et al. 2020). Another scenario involved "perceptual confusion", i.e. a significant drop in L2 perceptual accuracy under the influence of L3 (Nelson 2020). In terms of L2 production, the researchers have reported similar stability (Wrembel et al. 2019) or an increase in the production accuracy (Wrembel et al. 2022). However, it is worth noting that the results of the studies mentioned above concern mainly a different group of sounds than the current study, which focuses on vowels, VOT length and final devoicing. Therefore, the predictions are rather general and may differ as a function of individual features.

The following predictions were envisaged for the development of both speech modalities in the second language:

Prediction 1: Overall L2 perception and production accuracy remain stable across the three testing times, some differences may surface based on the individual features. Exposure to L2 may lead to the slight refinement of features over 10 months. L1 remains the most prominent source of influence for the L2 phonology.

Prediction 2: Overall L2 perception accuracy is subjected to the influence of L3 – the "perceptual confusion" is visible at T2, however, it returns to T1 levels in T3. Production accuracy remains stable across the three testing times, some differences may surface based on the individual features.

Prediction 3: The phenomenon of instability under the influence of T3 extends to both L2 perception and production accuracy. The phenomenon is visible in a significant drop in accuracy in T2, however, it returns to T1 levels at T3.

As mentioned previously, the development of individual L2 features, namely, L2 vowel duration and L2 VOT in voiceless stops may follow different paths; some possible scenarios are outlined below:

The predictions concerning vowel duration are as follows: due to the quantity and quality of input in language acquisition through formal instruction (positioned as a significant factor by PAM-L2, Best and Tyler 2007) and the suggested stronger influence of L1 in the later stages of L2 acquisition (NGTA; Dziubalska-Kołaczyk and Wrembel 2022), the participants may produce L2 vowels without the target-like duration or arrive at hybrid forms. If the participants rely primarily on duration in the discrimination of vowels (Bogacka 2004, Rojczyk 2010 and 2011), the results of perception and production

of vowels may align in terms of accuracy rates. If the duration is treated as a secondary cue (Balas 2018), then said results may diverge.

The prediction concerning the L2 VOT duration outlines more than one potential path of development. The analysis of T1 VOT production data (Krzysik 2020) revealed that non-target, L1-driven realisations were the most frequent, followed by hybrid, in-between L1 and L2 realisations. The similarity of L2 voiceless stops to Polish features may lead to the assimilation to existing L1 categories, moreover, the scarcity of input might prevent the participants from refining their production and distinguishing between L1 and L2 stops in terms of VOT (PAM-L2). Additionally, as suggested by the NGTA framework, the L1 CLI dominating the later stages of language acquisition may reinforce non-target VOT. However, this effect might be diminished by the cumulative influence of the exposure to similar VOT patterns in L2 and L3. Consequently, it may lead to an increase in the production of hybrid and target-like forms over time (CEM model; Flynn et al. 2004, Combined CLI; De Angelis 2007).

The second research question regarding multilingual perception and production focused on the third language of the participants:

(2) What is the development of L3 perception and production throughout three testing sessions?

The participants had started their L3 instruction approximately a month before the onset of the first testing session. Such an early stage in the L3 acquisition may result in rather volatile and changeable developmental patterns. Earlier longitudinal studies investigating the development of adolescent sequential multilinguals (L1 Polish, L2 English, L3 German or L1 German, L2 English and L3 Polish) in the context of formal education have shown different paths of L3 development. In terms of perception, the studies reported a slight development over time or the lack of thereof (Balas et al. 2019 for rhotics, Wrembel et al. 2020 for final obstruent devoicing). Nelson (2020) reported a substantial level of L3 perception accuracy for the w/v contrast associated with a "novelty effect" and focus on acoustic cues; a similar effect was found by Wrembel et al. (2020) for rhotics. In terms of production, Wrembel et al. (2019) reported a "u-shaped" developmental pattern with an initial peak in accuracy, whereas Wrembel et al. 2022 found a slight, statistically insignificant increase over time. Again, the results of the

studies mentioned above concern a largely different group of sounds than in the current study. Therefore, the predictions are rather general and may differ as a function of individual features.

Prediction 1: Overall, L3 perception and production remain stable across the three testing times, some differences may surface based on the individual features. Exposure to L3 may lead to the slight refinement of features over 10 months of L3 formal instruction.

Prediction 2: L3 perception is influenced by the novelty effect – reliance on the acoustic cues leads to increased perceptual accuracy at T1, which decreases over time. Production remains stable across the three testing times, some differences may surface based on individual features.

Prediction 3: L3 perception is influenced by the novelty effect – reliance on the acoustic cues leads to increased perceptual accuracy at T1, which decreases over time. L3 production displays a "u-shaped" pattern, with increased accuracy at T1 and a decrease in T2 and T3.

As mentioned previously, the development of individual L3 features may follow different paths; some possible scenarios are outlined below: With regards to vowel duration, the increased sensitivity to acoustic cues during the early stages of acquisition (NGTA: Dziubalska-Kołaczyk and Wrembel 2022) and the "novelty effect" (Nelson 2020) the participants may produce L2 vowels with target-like duration or arrive at hybrid forms. The increased role of L1 may lead to a decrease in accuracy, however, it might be balanced by the influence of L2 (which also uses duration as a significant cue). Similarly to L2, if the participants rely primarily on duration in the discrimination of vowels (Bogacka 2004, Rojczyk 2010 and 2011), the results of perception and production of vowels may align in terms of accuracy rates. If the duration is treated as a secondary cue (Balas 2018), then the results may diverge.

The VOT values for L2 voiceless stops may develop according to a number of different scenarios. Similarly to L2, the analysis of T1 VOT production data (Krzysik 2020) revealed that non-target, L1-driven realisations were the most frequent, followed by hybrid, in-between L1 and L2 realisations. The similarity of L3 voiceless stops to Polish features may lead to the assimilation to existing L1 categories, moreover, the scarcity of input might prevent the participants from refining their production and distinguishing between L1 and L2 stops in terms of VOT (PAM-L2). The counter influence of L2 (which has similar VOT patterns to L2) is rather improbable due to the low rate of

acquisition of the L2 target-like forms by the participants. However, the cumulative influence of the exposure to similar VOT patterns in L2 and L3 may lead to an increase in the production of hybrid and target-like forms over time (CEM model; Flynn et al. 2004, Combined CLI; De Angelis 2007).

The third research question focuses on the relationship between the perception and production in the second language of the participants:

(3) Is there symmetry in the gain between multilingual perception and production over time? Do both modalities co-evolve or is one of the modalities developing faster?

The relationship between the perception and production in a multilingual acquisition may follow several different paths predicted by models (SLM; Flege 1995, SLM-r; Flege and Bohn 2021) and outlined in several studies (e.g. Zhang 2019, Liu and Lin 2021; see Chapter 1 section 1.5.3 for an overview). Wrembel et al. 2022 summarised that there are four possible scenarios: (1) perception exceeds and precedes production in L2 speech learning, (2) perception and production are aligned and co-develop, (3) accurate production precedes perception, (4) there is no direct link between the perception and production. The comprehensive examination of perception and production in the context of multilingual phonology presented in the previously mentioned paper revealed that L1 Polish, L2 English, L3 German and L1 German, L2 English, L3 Polish adolescent learners demonstrated a considerable individual diversity. The patterns that emerged in the study suggested that overall perception surpasses production. Moreover, the L2 perception-production link was more consistent due to a more stable phonological system of the second language compared to L3. When considered individually, the participants demonstrated a wide variety of perception-production scenarios. In L2, they showed tendencies to move from the "perception preceded production" scenario to the opposite and the other way around between the testing times. In L3, the most frequent pattern involved a certain degree of perception accuracy without corresponding accuracy in production (i.e. dissociation).

Consequently, the following predictions concerning the relationship between the multilingual perception and production were envisaged:

Prediction 1: There is an overall link between the perception and production of multilingual participants. L2 perception accuracy exceeds production and the link is more stable compared to L3. Participants demonstrate a degree of variation in the relationship between perception and perception of the testing sessions; for L2 it might be modified by the onset of the acquisition of L3, whereas L3 might be influenced by the early reliance on acoustic cues e.g. increased sensitivity proposed by NGTA (Dziubalska-Kołaczyk and Wrembel 2022), or the 'novelty effect' (Nelson 2020), and later by the growing learning experience.

Prediction 2: There is no overall link between the perception and production of multilingual participants; perceptual accuracy cannot predict production and vice versa. The L2 and L3 perception-production relationships may diverge and are characterised by increased individual variation over time.

4.2.2. Research questions – individual differences in multilingual development

The main research question inquiring about the role of individual differences in multilingual development inquires about the potential ways in which the results of the phonological perception and production tasks interact with the measures of individual differences. The questions referring to the particular measures of individual differences and predictions concerning the research outcomes are outlined in the current section, starting with the question concerning the potential role of inhibitory control:

(1) What is the relationship between the results of phonological perception and production tasks and inhibitory control scores?

Linguistic performance may be influenced by one's inhibitory control capacity, which is required to balance the influence of the active linguistic systems (Inhibition Control Model; Green 1986, 1993, 1998; Festman 2018). Consequently, inhibitory control has been positioned as a predictor of individual differences in speech perception and production. The studies investigating this relationship have generated mixed results. Increased inhibitory control was found to be related to lower error rates and more target-like speech perception (Mora and Darcy 2013, Darcy et al 2014, Darcy et al. 2016) as

well as to speech perception and production (Lev-Ari and Peperkamp). The studies with the groups characterised by a less robust language experience such as adolescent early-stage L3 learners (Sigmeth and Golin 2018, Sigmeth et al. 2019, Sigmeth and Golin 2018 and Krzysik 2020) indicate that the relationship between multilingual development and inhibitory control may be more volatile. A study by Krzysik (2020), employing a portion of the data featured in the current thesis, demonstrated that production accuracy scores correlated with inhibitory control scores. However, the relationship was only visible for L2 and L3 combined scores. When the language accuracy scores were considered separately, only the accuracy scores for L2 were significantly correlated with inhibitory control. These findings indicate that the relationship between multilingual speech and inhibitory control may be modified by language exposure and proficiency.

Predictions: Earlier studies demonstrated that the relationship between inhibitory control and multilingual speech may be modified by language proficiency (Krzysik 2020). The combined scores for production L2 and L3 accuracy may show a relationship with inhibitory control, however, if considered separately, the relationship will be more visible for L2 and less visible for L3. The proficiency in L3 may be too low throughout the testing times to observe a meaningful interaction in this area.

(2) What is the relationship between the results of phonological perception and production tasks and phonological working memory scores?

Phonological working memory has frequently been considered a possible predictor of bi- and multilingual speech development. Previous studies have shown that increased PWM capacity may be linked to improved phonological perception (McKay et al. 2001, Aliaga-García et al. 2011, Darcy et al. 2015) and production (Krzysik and Wrembel 2019). A portion of the data presented in the current study was already analysed in the previously mentioned paper (Krzysik and Wrembel 2019); the obtained results showed a moderate relationship between the increased accuracy in multilingual speech (overall L2 and L3 production scores) and greater PWM capacity.

Prediction: Greater PWM capacity will be linked to improved perception and production accuracy. The participants with greater PWM capacity will score higher on ABX tasks in L2 and L3. Moreover, the participants with increased PWM capacity may produce more target-like forms in the delayed repetition tasks in L2 and L3.

(3) What is the relationship between the results of phonological perception and production tasks and the participants' language proficiency in L2 and L3?

Language proficiency constitutes one of the most frequently measured factors in the studies of multilingual language acquisition, assessed through measurements or reports. In the current study, language proficiency was assessed through short interviews and later rated by independent raters. The analysis will offer insights into the profile of the L2/L3 proficiency of the tested group.

Prediction: due to the longer language learning history, the rating of comprehensibility, overall fluency and overall pronunciation will be higher for L2 than for L3. The frequency of CLI from L1 will be higher in L2 than in L3 due to the increased influence of L1 in the later stages of acquisition. L3 will be characterised by similar levels of L1 and L2 influence due to the cumulative/combined CLI (Flynn et al. 2004, De Angelis 2007). The relationship between proficiency and perception/production may reflect the ratings: high rating in the categories such as comprehensibility, overall fluency and overall pronunciation should correlate with high accuracy in L2/L3 perception and production.

(4) What is the relationship between the results of phonological perception and production tasks and the participants' perception of language distances?

Psychotypology has been conceptualised as a factor potentially affecting multilingual language acquisition and CLI patterns in bi- and multilinguals (e.g. Singelton and Ó Laoire 2006, Hall et al. 2009, Xia 2017). It can be hypothesised that psychotypology aligned with the typological similarities between the languages potentially affects phonological perception and production.

Prediction: The perception of the phonological distances between the L1 Polish, L2 English and L3 German may follow the same pattern as in the study with comparable participants by Nelson et al. (2020), namely, the pair of English-German will be identified as the most similar. The participants who perceive the distances between the L2 and L3 as closer may perform better on the perception and production tasks in L2 and L3. The participants who perceive the pair as closer may be more perceptive when it comes to

noticing similarities between the languages such as L2/L3 phonemic vowel length or short lag vs. long lag VOT, which in turn may result in beneficial CLI.

(5) What is the influence of language exposure, use and language attitudes on the multilingual development of the participants?

In the context of formal instruction, the amount of language exposure and the frequency of use along with the characteristics of the linguistics input have been presented as deciding factors in the development of multilingual speech (e.g. Kopečková 2014).

Prediction: the participants declaring a more extensive exposure to L2 and L3 may evidence more target-like realisations. The L1-dominated exposure may result in non-target like or hybrid realisations in L2 and L3, with L1 as the major source of CLI. Positive language attitudes towards one language may influence the accuracy of perception and production.

4.2.3. Study variables and applied measures

The research questions presented above were investigated through a battery of tasks examining multilingual perception and production, inhibitory control, phonological working memory and psychotypology. These tasks were supplemented by a questionnaire of language history and use and a short task assessing the proficiency level in the L2 and L3 of the participants. Table 5 outlines the variables in the study and the measures used to assess them. The variables associated with individual differences are positioned as predictors (independent variables) of the dependent variables, i.e. speech perception and production in L2 and L3.

Table 5. Study variables and applied measures

Independent variables	Assessment of the independent variables	Dependent variables	Assessment of the dependent variables
Language status of L2 and L3*	Language history interview	Speech perception in L2 and L3	ABX perception task in L2 and L3
		Speech production in L2 and L3	Delayed repetition task in L2 and L3
Individual differences:			
Inhibitory control	flanker task		
Phonological working memory	pseudoword repetition task		
Language exposure and use	language history and use questionnaire		
Psychotypology	VILDIM measure		
Language proficiency	spoken proficiency measure		

*understood as the status of the language as the second or third in one's language learning history.

The framework of the study was longitudinal; the battery of tasks outlined in Table 5 was implemented at three testing times within 10 months of the school year 2018/19. The first testing session (T1) took place in late October and early November 2019. The second testing session (T2) started in the final week of January and lasted into the initial week of February. The third testing session (T3) was held between May and June 2019). Chapter 4 offers a detailed description of the methodologies and procedures implemented to conduct the abovementioned testing sessions.

4.3. Participants

The research plan for the current study intended to recruit the participants of the following profile: adolescent (aged 12-13), sequential multilinguals, acquiring their L2 (English) and L3 (German) through formal instruction in a Polish public school. The selection of this particular group was motivated by two major factors. Firstly, it was affected by the

constraints of the Polish demographics; primary school students constituted one of the few accessible multilingual groups in this rather linguistically homogenous country. Moreover, the characteristics of the Polish public school curriculum ensured a certain level of uniformity among the participants; the majority of the students start their formal L2 instruction in the first form (at the age of 6-7), whereas the L3 is introduced in the seventh form (at the age of 12-13) (see section 2.5.1 for a description of the foreign language education in Poland). Thus, the participants were expected to exhibit comparable degrees of language proficiency in their L2 and L3 and have similar language learning histories. Secondly, the selection of the group was reaffirmed by the scarcity of studies exploring multilingual phonological development of adolescents as opposed to adult or child populations which have been more extensively explored. For a few examples of studies on adolescent participants published to date see e.g. Kopečková 2015, Kopečková et al. 2019, Balas et al. 2019, Wrembel et al. 2020. Therefore, the focus on such a group was reasserted by the possibility to offer new insights and contribute to the further development of the field.

The school approached for the participation in the study was a public primary school, located in one of the western voivodeships of Poland. The research plan was positively evaluated by the AMU ethics committee and the administration of the school; both bodies accepted it for further realisation (the former is available in Appendix 1). The informed consent was obtained both from the parents / legal guardians and the adolescent participants. The school agreed to host the study on their premises; the individual testing sessions for each participant were held separately in a quiet room. The participants were rewarded with small gifts (such as stationery, language learning materials etc.).

The group of participants recruited in the school consisted of 27 adolescent sequential multilinguals, aged 12-13 (mean age=12.41, *SD*=0.43), female to male ratio (12:15). The participants were the students of the 7th form, enrolled in the same class. They had a uniform number of L2 and L3 language classes, namely, 3x45 minutes of English and 2x45 minutes of German per week. The mentioned language classes were taught by the same teachers, non-native speakers of the respective languages with academic degrees in respective languages. All 27 participants took part in three testing sessions. The exclusion of the participants with differing profiles was carried out after the process of data collection was concluded.

The set of exclusion criteria was established after the examination of the data retrieved from the language history and use questionnaires (described in section 4.7.3) administered in T1 and T3 sessions (see the methodology described in section 4.4.3). The criteria for exclusion were as follows: L1 different than Polish, simultaneous bi/multilingualism, different order of L2 and L3 acquisition, the experience of learning L2 and L3 exceeding the years outlined by the school curriculum, knowledge/learning experience concerning other foreign languages, a visit in and L2- or L3-speaking country exceeding two weeks. After the implementation of the exclusion criteria, the target group diminished to 20 participants. One participant was excluded due to their simultaneous bilingualism and their L1 being different from Polish. Another participant was excluded due to their long stay in an L3-speaking country and an extensive learning experience concerning L3. Five participants were excluded due to their earlier exposure to L3 and their L3 instruction exceeding the school curriculum.

4.3.1. The characteristics of the target study group

The final target study group, selected on the basis of the exclusion criteria, included 20 participants, female to male ratio (10:10), all native speakers of Polish. These participants were acquiring English from the onset of primary education and started learning German in the 7th form. Consequently, their respective onset age of acquisition for L2 English was 7, whereas for L3 German it was 12. The use and exposure to the L2 and L3 were restricted to the formal instruction at school. The group had no extensive experience of learning other foreign languages.

The amount of time spent using the L2 and L3 and self-ratings of language proficiency in these languages were derived from the language history and use questionnaires. The overall amount of time spent using L2 and L3 was based on self-estimation made by the participants in three categories: time spent using L2/L3 with (1) family, (2) friends and (3) at school. While performing the ratings the participants had to consider the time spent in all three languages – they could not exceed 100%. The overall amount of time spent using L2 or L3 was based on the means of the three mentioned categories. At T1, the participants spent on average 8.42% of their family/friends/school interactions in L3 and 4.91 in L3. At T2, the participants spent 10.67% of their

family/friends/school interaction in L2, whereas 5.16% in L3. As far as the self-ratings of proficiency are concerned, the participants rated their L2 and L3 language proficiency on a 1-5 scale (1 – very poor to 5 – excellent) in 4 categories: speaking, listening, writing, reading. The overall proficiency self-ratings (featured in Table 7) are combined means of the four mentioned categories. At T1, the participants rated their overall proficiency in L2 (*M*=3.81, *SD*=0.74) higher than in their L3 (*M*=2.43, *SD*=0.70). At T3, the overall proficiency in L2 (*M*=3.80, *SD*=0.69) was again rated higher than in their L3 (*M*=2.58, *SD*=0.84).

The participants' proficiency was also assessed through two short interviews in L2 and L3. The spoken output provided by the participants was then assessed by independent, trained raters on a 10-point rating scale (0 – no verbal output, 1 – very low overall proficiency, 10 – very high overall proficiency) in 6 categories (see section 4.7.5 for more details concerning the rating). The categories of rated proficiency in L2 and L3 (presented in Table 7) were created by combining the scores from three rating categories: comprehensibility, fluency and overall pronunciation. The mean proficiency in L2 during the first testing sessions was *M*=5.57 (*SD*=1.57) and *M*=4.75 (*SD*=2.67) in T2, whereas during the third testing session it was *M*=5.52 (*SD*=1.77). As for L3, it was rated at *M*=4.03 (*SD*=2.57) at T1, *M*=4.49 (*SD*=2.75) in T2 and *M*=4.93 (*SD*=2.30) in T3. The characteristics of the target group and the self-rated and rated proficiency are summarised in Table 6 and Table 7.

Table 6. Demographic information and characteristics of the target study group.

	T1		T2		T3	
		SD		SD		SD
mean age	12.65	0.48	-	-	-	-
AOA* of L2	7	-	-	-	-	-
AOA* of L3	12	-	-	-	-	-
hours of L2 instruction*	3	-	3	-	3	-
hours of L3 instruction*	2	-	2	-	2	-
time spent using L2 (%)	8.42	3.13	-	-	10.67	5.43
time spent using L3 (%)	4.91	3.04	-	-	5.16	3.72

**age of (language learning) onset **per week; hour understood as a 45-minute unit*

Table 7. Self-rated and rated proficiency of the target study group.

	T1		T2		T3	
	M	SD	M	SD	M	SD
self-rated L2 proficiency	3.81	0.74	-*	-	3.80	0.69
self-rated L3 proficiency	2.43	0.70	-*	-	2.58	0.84
rated proficiency in L2**	5.57	1.57	4.72	2.67	5.52	1.77
rated proficiency in L3**	4.03	2.57	4.47	2.49	4.93	2.30

**not assessed at T2 ** rating scale 1-10; 1 – low proficiency, 10 – very high proficiency.*

An extended participant profile and the presentation of the language history and use data, including self-assessment of language proficiency, contexts of language use, language attitudes and language mixing is available in sections 5.5.4 in Chapter 5. Proficiency ratings performed by a group of independent trained raters are available in section 5.5.6 in Chapter 5.

4.4. Testing procedure and the battery of tasks

Testing took place in a quiet room on the premises of the participants' school. Each participant was tested individually for approximately 50 minutes per session. Before the administration of the testing battery, the participants were informed about the characteristics of the tasks and inquired about their consent to participate. The testing battery was divided into the blocks, which are represented in Table 8 along with their distribution across the testing sessions 1, 2 and 3.

Table 8. Tasks administered in the testing battery at T1, T2 and T3

Tasks	Testing session	Tasks
L1 tasks	T1	delayed repetition task, ABX task
L2 tasks	T1, T2, T3	delayed repetition task, ABX task
L3 tasks	T1, T2, T3	delayed repetition task, ABX task
Individual differences		
Phonological working memory	T1, T3	PWM task

Inhibitory control	T1, T3	flanker task
Psychotypology	T1, T3	ViLDiM task
Language proficiency	T1, T2, T3	Rated interview in L2 and L3
Bio data, language history and use	T1, T2	Pen and paper questionnaire

The blocks were counterbalanced across the participants and the testing sessions to avoid any possible effects of order, however, the order of the tasks within the language blocks and the administration of ViLDiM at the end of the session remained the same for each participant. The questionnaire collecting the biographical information as well as the language history and use data was administered on separate days before the start of the session. Due to the logistic constraints and the intention to limit the interference of the experimenter in the education process, the language blocks had to be introduced within a single testing session on the same day. To partially counter this limitation, the experimenter attempted to induce a given language mode (Grosjean 1998) by opening each language block with a short language proficiency interview in the respective language. Moreover, during all testing sessions, the experimenter attempted to create a stress-free atmosphere for the participants and to highlight the enjoyable aspect of taking part in the study. The subsections below outline and describe the tasks of the testing battery implemented during the T1, T2 and T3 sessions.

4.5. Speech perception measure

The development of speech perception was tested by employing an ABX auditory discrimination task which was effectively applied in the previous studies with bilinguals and multilinguals (e.g. Mora 2007, Strange and Schafer 2008). The ABX auditory discrimination task involves a presentation of auditory stimuli in a series of trials. Tokens A and B introduce two contrasting features whereas X is a repetition of either A or B. After listening to the sequence, the participant indicates whether X is the same as A or the same as B using a keyboard or a button box key. The accuracy and reaction time measurements are supposed to reflect the ability to discriminate speech sounds and

possibly indicate one's perceptual sensitivity. Usually, the auditory stimuli of an ABX sequence are presented along with a series of visual guiding cues on a computer screen.

The ABX task in the current study was administered separately in the L2 and L3 of the participants. The tasks in both language versions intended to test the participants' discrimination of the front high vowels based on their quantity and quality. The vowels presented in the task were embedded in a "h_target sound_d" (hVd) framework in one-syllable tokens. The L2 English version of the task involved the vowel contrast of the front high unrounded vowel /ɪ/ and front high unrounded vowel /iː/. The L3 German version of the task involved the contrast of high front unrounded vowels /ɪ/ - /iː/ and high front rounded vowels /ʏ/ - /yː/. The contrasts tested in the ABX tasks in the study are outlined in Table 9.

Table 9. Contrasts in the ABX task in L2 and L3.

Language	examined contrast	contrasts in tokens
L2 English	/ɪ/ - /iː/	/hɪd/ - /hiːd/
L3 German	/ɪ/ - /iː/	/hɪd/ - /hiːd/
	/ʏ/ - /yː/	/h ʏ d/ - /h yːd/

The tokens were recorded by the native speakers of the respective languages and presented through headphones in the randomised order programmed in the experimental software. The tokens in every trial were presented with a 500 ms inter-stimulus interval (ISI). The task in all language versions was prepared and administered in E-Prime. The answers of the participants were collected using a button box. After receiving the verbal instructions supplemented with the visual description of the task on the screen, the participants completed the test trial to familiarise themselves with the procedure. Once the test trial block was completed, the participant had an opportunity to ask for additional clarification. After confirming the familiarity with the procedure, the participant was instructed to proceed to the main block of trials. Stimuli A and B were followed by an inter-stimulus interval of 1000 ms, whereas X was followed by a 3000 ms time window for the participant's response. Right after the presentation of the X trial, the participant had 3000 ms to indicate their answer on the button box. The stimuli were presented in a

randomised order. The presentation of the stimuli in a single trial of the ABX task is presented in Figure 4 (ICI – inter-stimulus interval, ? – response time window).

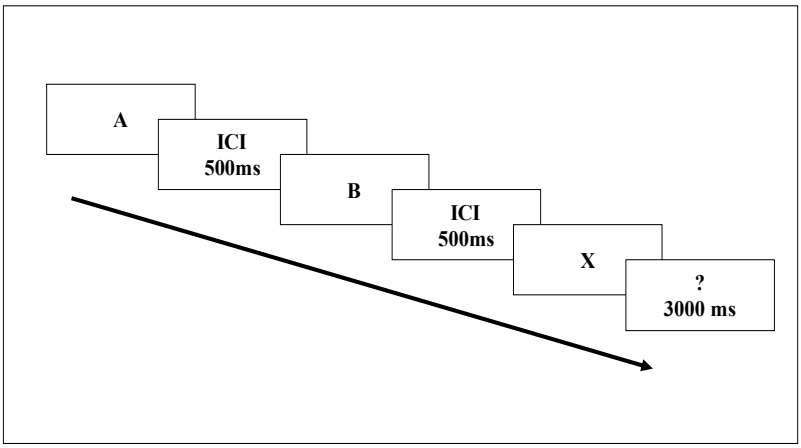


Figure 4. The order of the stimulus presentation in a single trial of the ABX task.

4.6. Speech production measure

The development of multilingual speech production was assessed using a delayed repetition task in L2 and L3 in three testing sessions. The tasks following this model were previously applied in studies on the development of multilingual phonology and rendered to be a suitable paradigm for studies with adolescents with varying degrees of language proficiency (e.g. Kopečková 2016, Balas et al. 2019).

In the current study, the delayed repetition tasks were intended to measure the production of the selected front high vowels, final devoicing and aspirated plosives in the L2 and L3 of the participants. Additionally, the participants were tested on their L1 production to account for possible instances of speech impediments such as rhotacism, which might have influenced the production of the items testing rhotic consonants. Each language was tested separately in the respective block during the session.

The structure of the task followed a uniform pattern for all language versions and it consisted of a set of pre-recorded mini dialogues. Sentence A introduced a real word with a given target feature such as a rhotic consonant, a front high vowel or an aspirated plosive. Sentence B constituted an intervening material aimed at decreasing the risk of direct imitation and short term-recall and allowing for access to the actual representations

of the tested features. The interval between the two sentences was 400 ms, which maintained an impression of a natural conversational pace. Sample delayed repetition sequences for L2 and L3 are outlined below:

(1) A. I say *target word* again.
B. What do you say?

(2) A. Ich sage *target word*, Anna.
B. Was hast du gesagt?

The participants were instructed to listen to the mini dialogue and repeat sentence A after the presentation of the entire two-sentence sequence. All language versions of the task were adjusted to the pace of the participant; after an attempt to reproduce the first sentence of the mini dialogue, the experimenter triggered the transition to the next mini dialogue sequence. The renditions produced by the participants were recorded. The complete list of items used to elicit target features in both language versions of the task is available in Appendix C.

4.7. Measures of individual differences

The study supplemented the tasks tracing phonological developments with measures of individual differences assessing factors such as phonological working memory, inhibitory control, language history and use and psychotypology. A detailed description of the tasks aimed at capturing the possible influence of these variables is provided in the subsequent sections.

4.7.1. Phonological working memory measure

Phonological working memory was measured employing a pseudoword repetition task, obeying the rules of Polish phonotactics. The measure itself was earlier introduced in the study by Krzysik and Wrembel (2019), along with the analysis of a portion of the data

from the T1 testing session. However, unlike the current analysis, it awarded a uniform amount of accuracy points regardless of the word length.

The format and the administration formula of the PWM task was inspired by pseudoword repetition tasks such as Zetotest (Krasowicz-Kupis and Grela-Goryczka 2003) and Test Powtarzania Pseudosłów (Szewczyk et al. 2015). The task featured 29 pseudowords arranged in the order of increasing length (from 2 to 6 syllables), sourced from the Polish Pseudo-words List (Imbir et al. 2015). PPwL's raw dataset of 3240 was generated by an algorithm that substituted letters in the existing Polish words for random letters, taking into account the frequency of letters in given positions. The PPwL word set was subsequently rated by 5 trained judges according to the following rules: a given word is built from the already existing or potential syllables, it is possible to read a given word fluently, it conforms to the rules of Polish spelling, it does not occur in Polish, and does not invoke easy associations with the real words existing in Polish. The rated list was narrowed down to 3023 words of increasing accordance with the rules outlined above. The 2 to 5 syllable words featured in the task were randomly selected from the congruency subset of 1. The 6-syllable words were not included in the PPwL list, therefore 4 additional words were created by combining randomly selected syllables in line with the generation of PPwL items. The 6-syllable items were also permitted phototactically by Polish and not easily associated with any existing Polish word. To validate the task, its scores were correlated with the results of a standardised measure of working memory, namely, a forward digit span task, which resulted in a positive correlation ($r=.665$, $N=25$, $p<.001$) (Krzysik and Wrembel 2019), rendering the pseudoword repetition an acceptable measure of PWM.

The 29 items of the pseudoword repetition task were recorded by a native speaker of Polish and played to the participants via semi-opened headphones in the order of increasing pseudoword length. The participants were requested to repeat each pseudoword right after hearing a short sound signal following the presentation of each item. The answers of the participants were recorded. The complete list of pseudowords is available in Appendix D.

Since the analysis focused on the production of the items based on the L1 phonotactics, it was conducted by a single rater, a native speaker of Polish. The list of pseudowords was divided into two groups: 1. pseudowords ranging from 1 to 3 syllables and 2. 4 to 6 syllables. The realisations of the pseudowords from group 1. were awarded

1 point for each fully target-like rendition and 0 points for non-target like renditions. The realisations of the pseudowords from group 2. were awarded 2 points for fully target like renditions, 1 point for renditions containing 1 incorrectly produced syllable and 0 points for more than 1 incorrectly produced syllable. The maximum number of points possible to award in group 1. was 15, whereas in group 2. it was 28. The points scored for both groups of pseudowords were added to create a general score of phonological working memory.

4.7.2. Inhibitory control measure

The inhibitory control capacity was assessed using a modified domain-general, standardised flanker task (Eriksen and Eriksen 1974), modelled on the version implemented in a study by Poarch and Bialystok (2015). The current version of the task was also employed in an earlier the study by Krzysik (2019). In the task, the participants saw a row of arrows displayed on a computer screen in an E-prime-built testing setup. They were requested to indicate the direction of the middle arrow by pressing one of two assigned buttons on a response box. The task featured four types of trials, namely, baseline, neutral, congruent and incongruent. In the baseline trials, only one arrow appears on the screen, whereas in the neutral trials the arrow is flanked by two diamonds on each side. In the congruent trials all arrows point in the same direction, and the incongruent trials the middle arrow points in the other direction. Figure 5 shows the displays in the respective trials. All blocks (baseline, neutral, congruent and incongruent) were randomised in the experimental software. Every trial was preceded by a 500 ms fixation before the onset of the stimulus. The participants had 3000 ms after the onset of the stimulus to indicate their response. The responses obtained in the incongruent trials requiring the suppression of interference were compared with the ones in which such suppression was not required. They served as a basis for drawing inferences about the participant's inhibitory capacity, with a focus on the mechanism interference suppression mechanism potentially associated with the accuracy in linguistic performance.

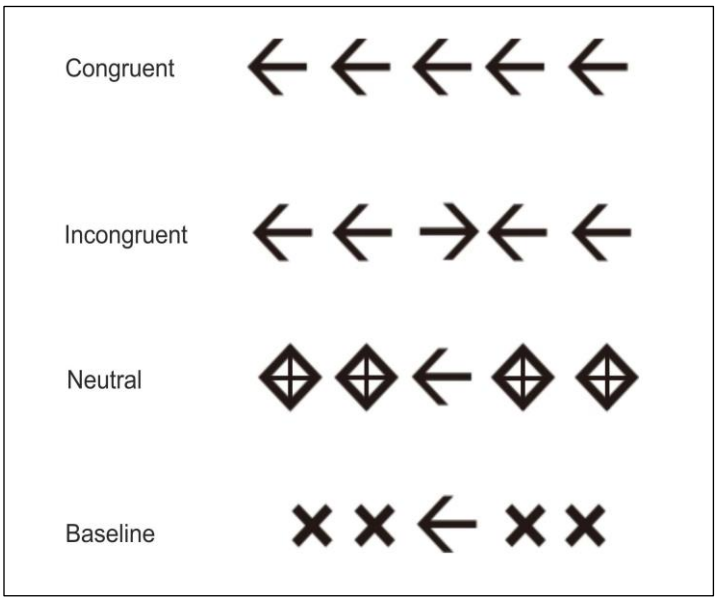


Figure 5. A visual representation of conditions in a flanker task (based on Poarch and Bialystok 2015).

4.7.3. Biographical data, language history and use

The biographical information, language history, exposure and use were collected by means of a questionnaire, based on the frameworks suggested in the Language History Questionnaire (Li et al. 2014) and the Language Experience and Proficiency Questionnaire (Marian et al. 2007). The questionnaire was divided into two parts; the first part inquired about biographical data, whereas the second part focused on the language repertoire, experience and the current use of language. The questionnaire was administered to the participants in the paper form in their L1. The experimenter was present during the administration and provided necessary instructions and assistance. The collected data was archived in an electronic form and used in further analysis. The complete form of the questionnaire translated to English is provided in Appendix E.

4.7.4. Psychotypology measure

The measure of psychotypology employed in the study was the Visual Language Distance Measure (VILDIM) (Nelson et al. 2021). VILDIM makes use of a dotted A3 sheet of paper and a set of transparent circles featuring the names of the languages in the

participants' repertoire (and also blank ones for the additional, potential languages). The participants are instructed to arrange the circles on the sheet of paper to reflect their perception of similarities between particular languages. In the present study, the participants were requested to focus on the phonological similarity between the languages they knew. To avoid possible misunderstandings and adjust the wording to more age-appropriate terminology, the aforementioned phonological similarity was described to the participants as the similarity between the sounds of L1, L2 and L3. After completing the task, the positions of circles are marked on the sheet of paper through a hole in the middle of each transparent circle and signed with the name of the respective language. Figure 6 visualises the dotted sheet of paper with examples of language circles.

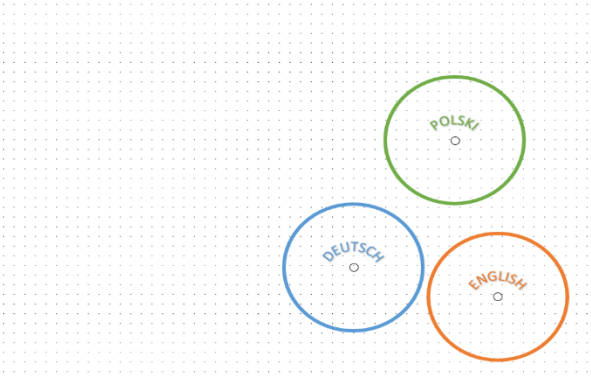


Figure 6. A visualisation of ViLDiM.

The raw values of the distances were obtained by measuring (in centimetres) the distances between the marked positions of the circles. Each distance was normalised by following the formula outlined in Golin et al. (2020) (Equation 1).

Equation 1. Normalisation formula for psychotypology measure (VILDIM).

$$NV_{\mathrm{L1-L2}} = \frac{RV_{\mathrm{L1-L2}}}{RV_{\mathrm{L1-L2}} + RV_{\mathrm{L1-L3}} + RV_{\mathrm{L2-L3}}}$$

NV marks a normalised value of the measured distance (e.g. between L1 and L2) and RV stands for a raw value of the measured distance. The normalised value of each language distance is calculated by dividing the raw value of the said distance by the sum of three raw values of distances collected from a participant. After the normalisation, the distances

can be compared across the group without the potential confounding influence of the individual perception of the distances on a sheet of paper.

4.7.5. Language proficiency

Language proficiency in L2 and L3 was assessed through two short interviews conducted during the testing session. The interviews consisted of 5 questions per language at T1, T2 and T3. The content of the questions was the same in the L2 and L3 of the participants (the questions are available in Appendix F). The analysis of the collected materials had two tiers. The first tier of the analysis, identified as general language proficiency assessment, was conducted by the experimenter. The participants who produced no verbal output were discarded from further analysis and awarded 0 points in each of the subsequent rating categories in the next step of the analysis.

The second tier of the analysis involved proficiency rating by three independent raters per language. The L2 raters had near a native-like knowledge of English and completed English phonetics and phonology and pronunciation courses during their BA studies. Similarly, the L3 raters have a native-like competence in Germans and completed German phonetics and phonology and pronunciation courses during their BA studies. The answers provided by the participants in short interviews were rated on a 10-point rating scale in six categories. The categories were developed in relation to the ones employed in the study by Crowther et al. (2015) and included two subgroups: overall proficiency (comprehensibility, overall fluency, overall pronunciation) and CLI (switch to L1, influence from L1 in pronunciation and influence from L2 or L3 in pronunciation). The category of comprehensibility was understood as the level of difficulty in understanding the speech of a given participant. Overall proficiency was operationalised as a general judgement of the participants' language proficiency, whereas overall pronunciation was a judgement of the participants' pronunciation in the rated language. The category evaluating the participants' switch to L1 involved an assessment of the frequency of the switches. The two final categories focused on the influence on the rated language pronunciation from other languages (L1 and either L2 or L3); the raters were required to assess the frequency of pronunciation-based CLI from the remaining languages in the participants' repertoires. Table 10 presents the six rating categories and the labels of the minimal and maximal scale points.

Table 10. The categories in proficiency rating with rating scale points.

rating category	labels of most distant rating scale points		
comprehensibility	0 – no verbal output	1 – very low comprehensibility, the speech is impossible to understand	10 - very high comprehensibility, the speech is easily understandable
overall fluency	0 – no verbal output	1 – very low overall fluency, issues with formulating simple sentences.	10 – very high overall proficiency, no issues with formulating fluent passages of speech.
overall pronunciation	0 – no verbal output	1 – very frequent mispronunciations, heavy CLI in pronunciation.	10 – no mispronunciations, no CLI from the remaining languages.
switches to L1	0 – no verbal output	1 – no switches to L1	10 – very frequent switches to L1
influence from L1 in pronunciation	0 – no verbal output	1 – no influence from L1	10 – very frequent influence from L2
influence from L2/L3 in pronunciation	0 – no verbal output	1 – no influence from L2	10 – very frequent influence from L3

All raters received extensive instructions concerning the procedure of rating, the implemented rating categories and the premises underlying the applied constructs in the rating. The process of ratings was completed remotely; the raters had online access to the coded and anonymised recordings of the interviews. The raters logged their ratings into the spreadsheets prepared by the experimenter. The recordings were rated in the order of data collection, starting from T1. After completing each session, the raters had no possibility of returning to the data from the previous session and changing their scores.

4.8. Summary

Chapter 4 outlined the methods and procedures applied in the study constituting the empirical part of the thesis. Section 4.2. presented the study aims and research questions. Sections 4.3 demonstrated the general profile of the participants. Section 4.4 described the organisation of the testing sessions and the tasks administered in the testing battery. Further, sections 4.5 and 4.6 presented the speech perception and speech production measures in L2 and L3, assessing the multilingual development of the participants.

Section 4.7 outlined the measures of individual differences, including phonological working memory, inhibitory control, biographical data/language history/language exposure, psychotypology as well as language proficiency.

Chapter 5: Results

5.1. Introductory remarks

The aim of Chapter 5 is to present the results obtained in the testing battery outlined in Chapter 4. Sections 5.1 to 5.3 present the results of the subsequent tests of multilingual phonological perception, production and individual differences across testing sessions. Section 5.4 evaluates the participants' perception and production in terms of accuracy; in the case of production, the values are compared against the native values in L2 and L3 to produce an accuracy score. Section 5.5 presents the results of individual differences tasks. Sections 5.6-5.9 compare the perception and production data with the results of individual differences tasks in order to examine the interactions between the obtained scores. The statistical analyses included in Chapter 5 were performed in IBM SPSS and R software.

5.2. Results of the speech perception tasks

The multilingual speech perception of the participants was examined via the ABX task in L2 and L3 administered through E-Prime software. The phonological features examined in the task included close front vowels. The outcomes of the analysis are outlined by language and featured in the subsections below.

5.2.1. Speech perception task in L2 English

The speech perception task in L2 English focused on the recognition of two close front unrounded vowels /iː/ and /ɪ/ (trials focusing on other features were not included in the current thesis as they exceeded its focus). The vowels were embedded in a "h_target sound_d" (hVd) framework in one-syllable tokens, which were randomised and appeared in every possible combination (ABB, BAA, ABA, BAB). The task measured accuracy in vowel recognition and response time (i.e. reaction times – RT). The process of auditory word recognition requires approximately 200 ms (Wingfield 1996, Payne and Silox

2019), therefore the responses RTs below that number were excluded. The maximum number of points to score in the /ɪ/-/iː/ trials was 16 (1 point per trial); the points scored by the participants in three testing sessions were converted into percentages. At T1, the mean accuracy in perception of /ɪ/-/iː/ contrast reached 48% (SD=11.62). A slight increase was observed for T2 (M=51, SD=9.04). In T3, the mean accuracy noted a slight decrease (M=48, SD=9.04). Table 11 outlines the mean accuracy scores in the perception of L2 contrast in three testing sessions. Figure 7 presents the individual L2 accuracy scores in three testing sessions.

Table 11. Mean accuracy (Acc %) of the L2 perception of /ɪ/-/iː/ contrast in three testing sessions.

Vowel pair	Testing session	N	Acc %	95% CI	SD	Median	Min	Max
/ɪ/-/iː/	T1	20	48	42.996-53.878	11.62	43.75	31.25	68.75
/ɪ/-/iː/	T2	20	51	46.391-54.858	9.04	50.00	37.50	68.75
/ɪ/-/iː/	T3	20	49	44.099-54.650	11.27	50.00	31.25	75.00

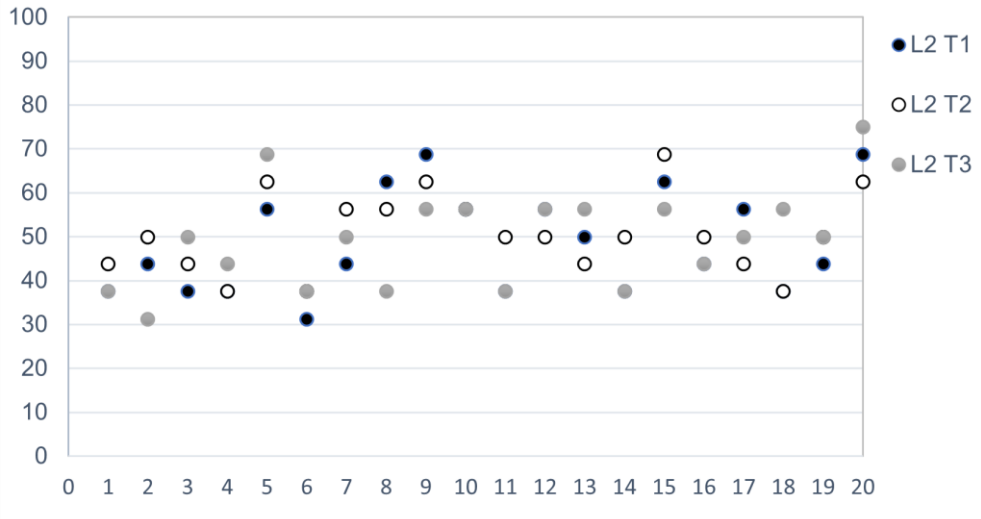


Figure 7. Individual accuracy scores for the perception of /ɪ/-/iː/ contrast in three testing sessions.

Subsequently, the perception accuracy of the L2 /ɪ/-/iː/ contrast obtained in three testing sessions was compared through repeated measures ANOVA. No significant effect of time was found, $F(2, 38)=.540$, $p=.587$, $\eta p2=.028$. Such a result indicates that there was no significant difference between the participants' accuracy scores across the testing sessions. Figure 8 outlines the perception accuracy means in the three testing sessions.

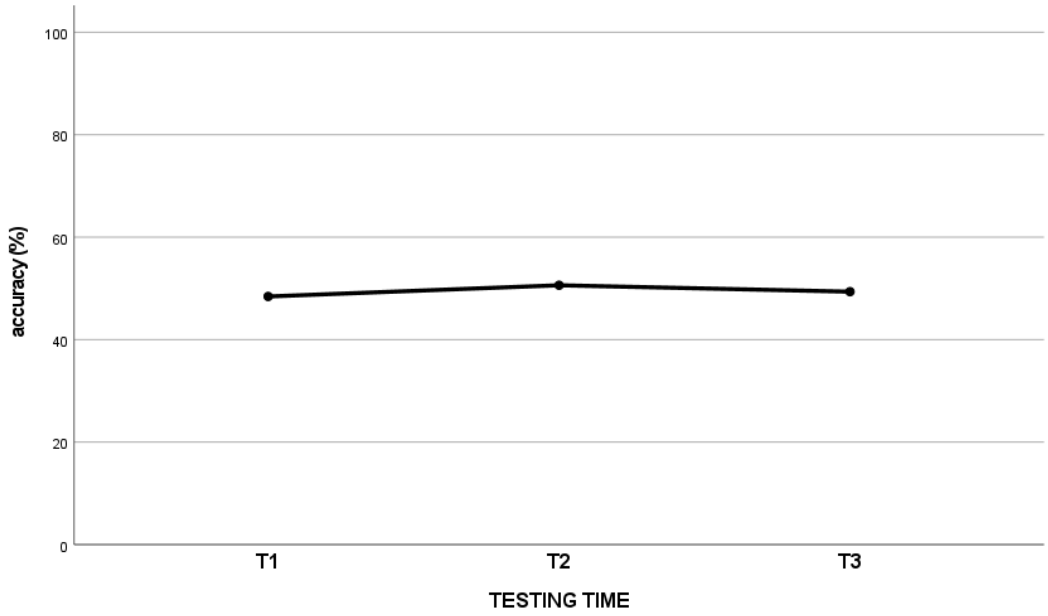


Figure 8. L2 perception accuracy means in T1, T2 and T3.

5.2.2. Speech perception task in L3 German

The speech perception task in L3 German focused on the recognition of two close front unrounded vowels /iː/ and /ɪ/ and two close front rounded vowels /ʏ/ and /yː/ (other types of trials were not included in the study – they were beyond its scope). The vowels were embedded in a "h_target sound_d" (hVd) one-syllable tokens, which were randomised and appeared in different combinations (ABB, BAA, ABA, BAB). Similarly to L2, the RTs below that number were excluded. The maximum number of points to score in the L3 perception task was 16 (1 point per trial; 8 points per contrast); the points scored by the participants in three testing sessions were converted into percentages. At T1, the mean accuracy in perception of /ɪ/-/iː/ contrast approached 63% (*SD*=15.42). A slight increase was noted in T2 (*M*=66, *SD*=12.88). At T3, the mean accuracy further increased (*M*=68, *SD*=17.43). As far as /ʏ/-/yː/ contrast is concerned, the participants' accuracy was generally lower compared to /ɪ/-/iː/ contrast. At T1, the mean accuracy reached 58% (*SD*=16.27); the scores remained comparable in T2 (*M*=59, *SD*=14.55). A small increase in the perceptual accuracy of the /ʏ/-/yː/ contrast occurred in T2 (*M*=62, *SD*=9.48). Table 12 outlines the mean accuracy scores in the perception of L3 contrasts in the three testing

sessions. Figure 9 and Figure 10 present the individual L3 accuracy scores in the three testing sessions for two tested contrasts.

Table 12. Mean accuracy (Acc %) in the perception of L3 /ɪ/-/iː/ and /ʏ/-/yː/ contrasts at T1, T2 and T3.

Vowel pair	Testing session	N	Acc %	95% CI	SD	Median	min	Max
/ɪ/-/iː/	T1	20	63	55.904-70.346	15.42	62.50	37.50	100.00
/ɪ/-/iː/	T2	20	66	60.218-72.282	12.88	62.50	50.00	87.50
/ɪ/-/iː/	T3	20	68	59.967-76.283	17.43	62.50	37.50	100.00
/ʏ/-/yː/	T1	20	58	51.134-66.366	16.27	62.50	37.50	87.50
/ʏ/-/yː/	T2	20	59	52.564-66.186	14.55	62.50	25.00	87.50
/ʏ/-/yː/	T3	20	62	57.434-66.316	9.48	62.50	50.00	75.00

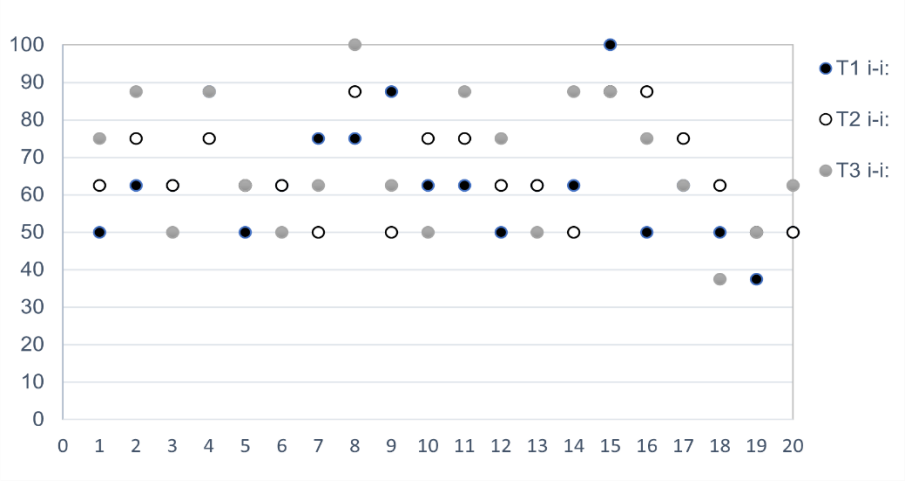


Figure 9. Individual accuracy scores of the recognition of /ɪ/-/iː/ contrast in three testing sessions.

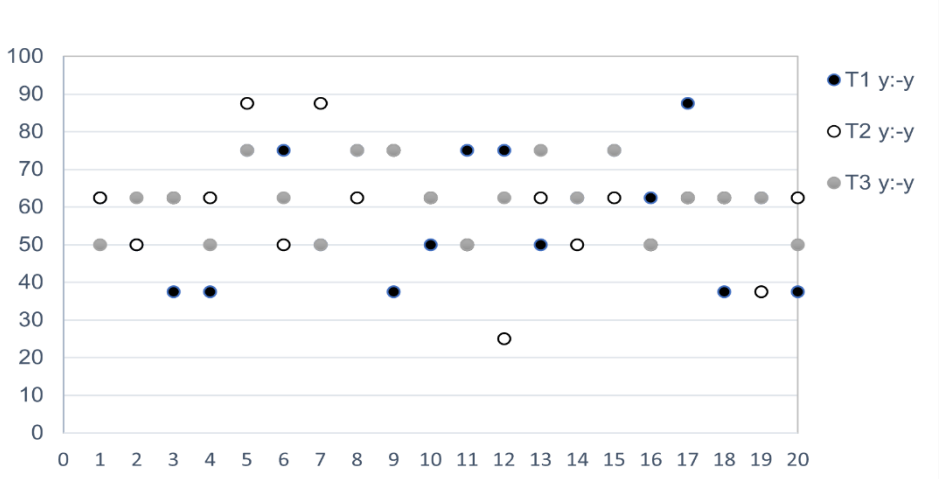


Figure 10. Individual accuracy scores of the recognition of /ʏ/-/yː/ contrast in three testing sessions.

Subsequently, the combined perception accuracy of the two L3 contrasts obtained in the three testing sessions was compared through repeated measures ANOVA. No significant effect of time was found, $F(2, 38)=1.61$, $p=.212$, $\eta p2=.078$. Such a result indicates that there was no significant difference between the participants' perception accuracy scores across the testing sessions. Figure 11 outlines the L3 perception accuracy means in three testing sessions.

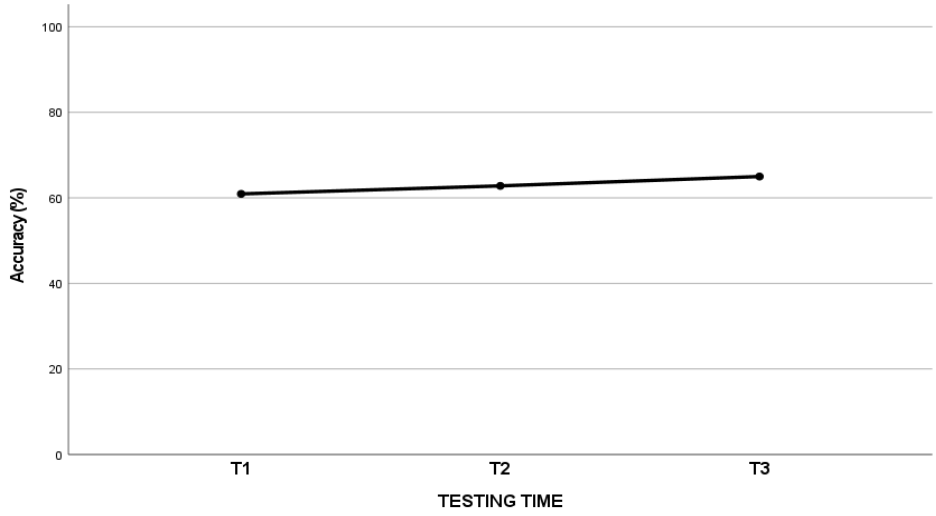


Figure 11. L2 perception accuracy means at T1, T2 and T3.

5.2.3. Summary of the speech perception results

The trials of the ABX task analysed in the current thesis tested the participants' perception of close front unrounded vowels for L2 English and close front rounded and unrounded vowels for L3 German. For both L2 and L3, there was no significant difference in the accuracy of perception over the three testing sessions as it remained stable. The accuracy of L2 English was lower across the three testing sessions compared to the L3 accuracy scores. Table 13 summarises the results of speech perception tasks in L2 and L3.

Table 13. The results of speech perception tasks in L2 and L3

Language	Tested contrast	Change over time
L2 English		
	/ɪ/-/iː/	No significant difference
L3 German		
	/ɪ/-/iː/	No significant difference
	/ʏ/-/yː/	No significant difference

5.3. Results of the speech production tasks

The material recorded in speech production tasks in L2 and L3 was analysed acoustically in PRAAT software (Boersma and Weenink 2022). The phonological features examined in the study included vowel duration and Voice Onset Time (VOT) in voiceless plosives (the latter was partially presented in Krzysik 2020). The outcomes of the analysis are outlined by language and feature in the subsections below.

5.3.1. Vowel duration in L2 English

The production task in L2 English focused on two close front unrounded vowels, namely, /iː/ and /ɪ/. The duration of the vowels was collected through manual segmentation and measurement based on waveforms and spectrograms in PRAAT software (Boersma and Weenink 2022).

At T1, the obtained vowel duration measurements for L2 indicated a distinction between the production of /ɪ/ and /iː/. The vowel duration means (in ms) in the words *wick* and *week* were registered as shortest in the target word set (*M*=115, *SD*=19.25 and *M*=128, *SD*=22.94 respectively), which can be attributed to pre-fortis clipping. The production of /ɪ/ in *bid* and *did* was characterised by comparable duration values (*M*=139, *SD*=18.41 and *M*=136, *SD*=21.05). The duration of /iː/ in *bead* and *deed* had the longest duration values in the target word set (*M*=161, *SD*=28.33 and *M*=151, *SD*=18.14). Table 14 outlines the L2 vowel duration means in the T1 testing session.

Table 14. L2 vowel duration (in ms) at T1.

Target word	vowel	N	M (ms)	95% CI	SD	Median	Min	Max
Wick	/ɪ/	20	115	106.66-124.68	19.25	116.67	74	153
Week	/iː/	20	128	117.57-139.04	22.94	123.71	95	176
Bid	/ɪ/	20	139	130.10-147.33	18.41	144.81	108	168
Bead	/iː/	20	161	161.58-167.19	28.33	167.19	113	207
Did	/ɪ/	20	136	125.76-145.61	21.05	134.18	103	175
Deed	/iː/	20	151	142.69-159.67	18.14	152.68	102	175

At T2, the obtained vowel duration measurements followed a similar pattern to that observed at T1. Once again, the vowel duration means of the words *wick* and *week* were the shortest in the target word set (*M*=114, *SD*=13.31 and *M*=122, *SD*=23.04 respectively). The production of /ɪ/ in *bid* and *did* has similar duration values (*M*=135, *SD*=18.41 and *M*=136, *SD*=21.05) compared to T1. The duration of /iː/ in *bead* and *deed* had again the longest duration values in the target word set (*M*=174, *SD*=26.77 and *M*=158, *SD*=25.72) and were marked by a slight increase compared to T1. Table 15 outlines the L2 vowel duration means in the T2 testing session.

Table 15. L2 vowel duration (in ms) at T2.

Target word	vowel	N	M (ms)	95% CI	SD	Median	Min	Max
Wick	/ɪ/	20	114	107.67-120.13	13.31	110.99	96	158
Week	/iː/	20	122	111.86-133.42	23.04	123.51	90	177
Bid	/ɪ/	20	135	122.97-147.51	26.21	134.66	73	179
Bead	/iː/	20	174	161.30-186.37	26.77	171.29	128	256
Did	/ɪ/	20	134	122.35-146.59	25.89	129.19	102	181
Deed	/iː/	20	158	145.56-169.64	25.72	161.74	107	192

L2 vowel duration at T3 resembled the pattern recognised at T1 and T2. Once again, the vowel duration means of /ɪ/ in the words *wick* and *week* were the shortest in the target word set (*M*=108, *SD*=25.81 and *M*=1, *SD*=22.94) and also the shortest of all testing sessions. The production of /ɪ/ in *bid* and *did* was marked by rather stable values compared to previous testing sessions (*M*=129, *SD*=24.24 and *M*=137, *SD*=27.05). The duration of /iː/ in *bead* and *deed* had the longest duration values in the target word set (*M*=164, *SD*=21.73 and *M*=154, *SD*=27.82). Based on SD values, it might be assumed

that the data gathered in T3 exhibited the most diverse vowel durations among all testing sessions. Moreover, the expanding Max ranges may indicate a development towards target-like vowel duration. Table 16 outlines the L2 vowel duration means in the T3 testing session. Figure 12 presents the vowel duration means over three testing times for each target word.

Table 16. L2 vowel duration (in ms) at T3.

Target word	vowel	N	M (ms)	95% CI	SD	Median	min	Max
Wick	/ɪ/	20	108	96.44-120.60	25.81	108.75	68	151
Week	/iː/	20	120	108.91-130.71	23.29	125.64	86	160
Bid	/ɪ/	20	129	117.42-140.11	24.24	123.22	91	183
Bead	/iː/	20	164	144.59-183.62	21.73	150.57	98	260
Did	/ɪ/	20	137	124.66-150.33	27.42	130.60	103	200
Deed	/iː/	20	154	141.53-167.58	27.82	160.25	105	200

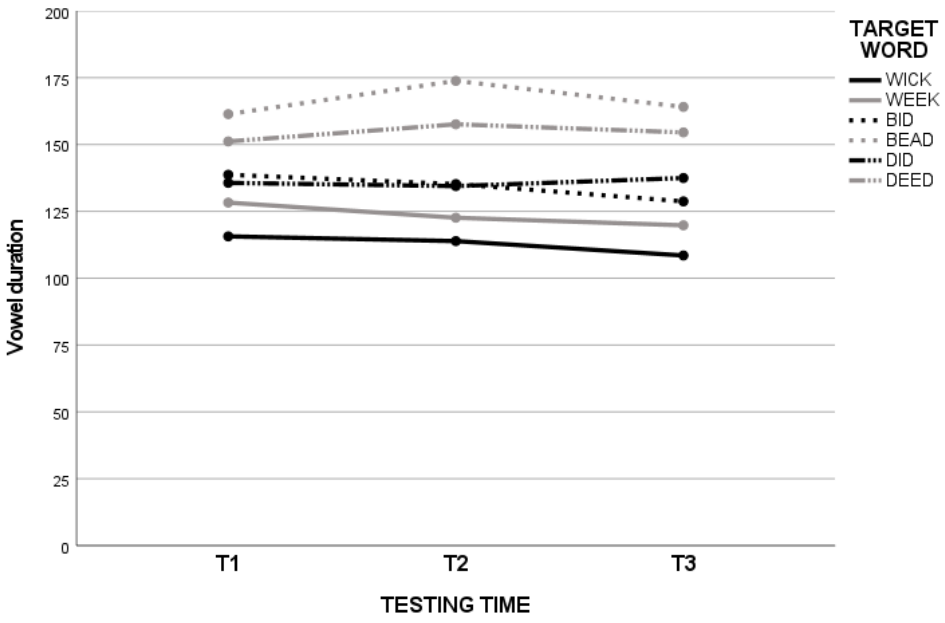


Figure 12. L2 Vowel duration means over three testing times for each target word.

Vowels duration measurements were grouped according to their type as either /ɪ/ or /iː/. The means for grouped vowels revealed the following pattern – in all three testing sessions the mean duration of /ɪ/ vowels was shorter compared to /iː/. Another observed

pattern was the group variability increasing over the testing sessions, indicated by the increasing SD or wider min-max ranges. The means of these two groups in three testing sessions are presented in Table 17. Individual means of /ɪ/ and /iː/ production in three testing sessions are presented in Figure 13 and Figure 14.

Table 17. L2 vowel duration means for /ɪ/ and /iː/ vowels in three testing sessions.

Vowel	testing time	M (ms)	95% CI	SD	Median	min	max
/ɪ/	T1	130	124.84-135.26	11.12	133.00	101	145
/iː/	T1	147	139.36-154.64	16.33	147.50	117	177
/ɪ/	T2	128	120.33-135.47	16.17	128.00	98	150
/iː/	T2	151	143.61-159.19	16.65	150.57	128	178
/ɪ/	T3	125	115.23-134.67	27.42	125.50	94	178
/iː/	T3	146	134.42-157.88	25.06	140.00	98	198

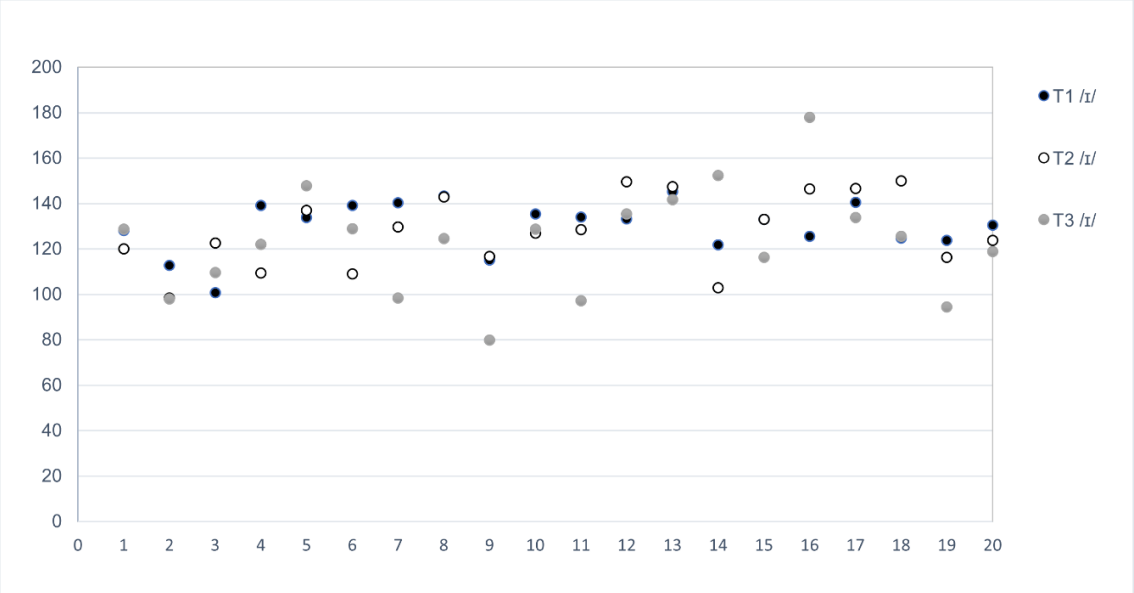


Figure 13. Individual /ɪ/ duration means (in ms) for 20 participants in three testing times.

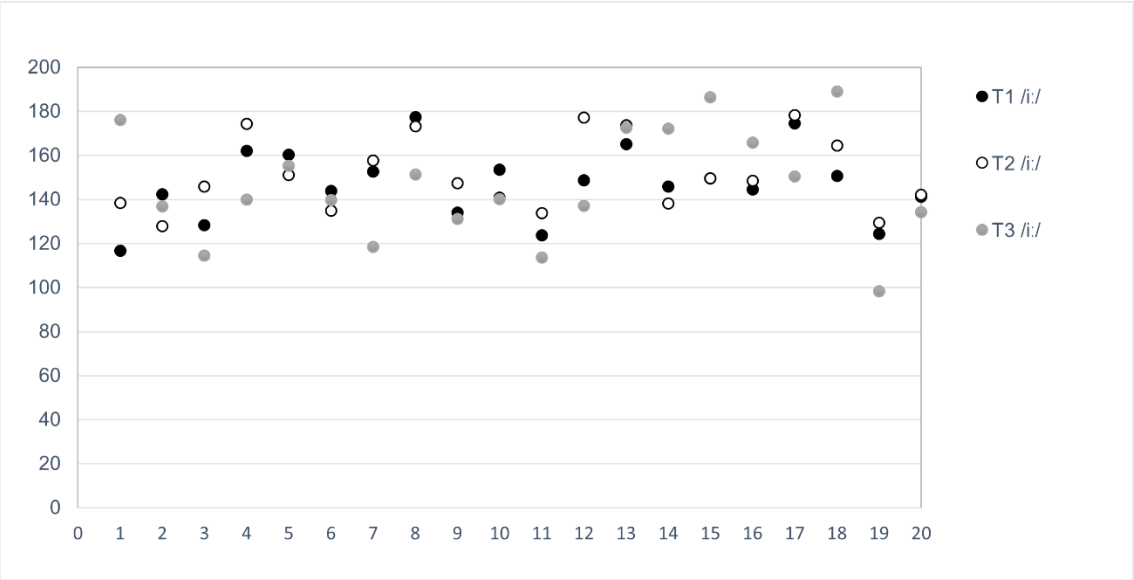


Figure 14. Individual /ɪ/ duration means (in ms) for 20 participants in three testing times.

Subsequently, the vowel duration means for /ɪ/ vowels and /iː/ vowels obtained in the three testing sessions were compared through repeated measures ANOVA. For the condition "time" and "vowel type*time", sphericity was violated as indicated by Mauchly's tests, $X2(2)=7.23$, $p=.027$ and $X2(2)=6.20$, $p=.045$. Therefore, the Greenhouse-Geisser correction was applied. Repeated measures ANOVA revealed a highly significant effect of the vowel type, $F(1, 19)=83.31$, $p<.001$, $\eta p2=.814$. Such a result indicates that the mean vowel durations of /ɪ/ and /iː/ were significantly different. However, there was no significant effect of time, $F(1.50, 28.55)=.57$, $p=.52$, $\eta p2=.02$. and no significant interaction of the vowel type and time, $F(1.55, 29.42)=.99$, $p=.36$, $\eta p2=.05$. Consequently, there was no change in the overall vowel duration over time, as well as no significant change in the duration of vowel types. Figure 15 presents the vowel duration means over three testing sessions.

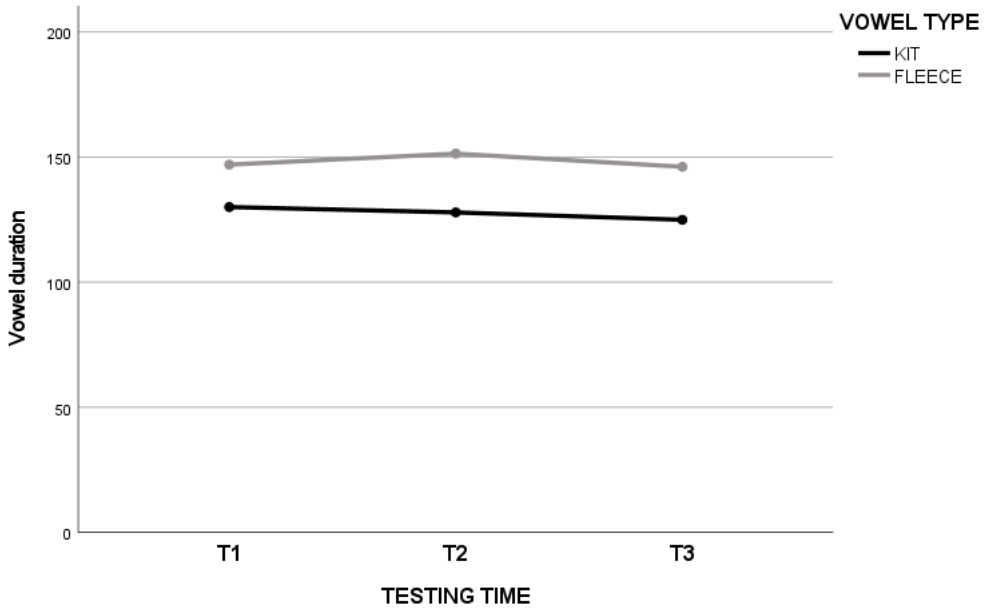


Figure 15. Vowel duration means for /ɪ/ (marked as KIT) and /iː/ (marked as FLEECE) over three testing sessions.

5.3.2. Vowel duration in L3 German

The production task in L3 German focused on two close front unrounded vowels - /iː/ and /ɪ/ - and two close front rounded - /ʏ/ and /yː/. Similarly to L2 vowel analysis, the duration of the vowels was collected via manual segmentation and measurement based on waveforms and spectrograms in PRAAT software (Boersma and Weenink 2022).

At T1, the obtained vowel duration measurements for L3 indicated a distinction between the production of short and long vowels; the former exhibited a shorter and the latter a longer duration. The vowel duration means of /ɪ/ (in ms) in the words *Mitte* and *Bitte* were the shortest in the entire target word set (*M*=73, *SD*=10.12 and *M*=78, *SD*=9.50 respectively), which can be attributed to the influence of the following /t/ consonant, contributing to vowel shortening. A similar phenomenon affected also the production of the long vowel /iː/ in *Miete* (*M*=96, *SD*=13.09), which was considerably shorter than its vowel counterpart in the context of *Linie* (*M*=128, *SD*=20.91). The production of /ʏ/ in *füllen* and *küssen* was characterised by similar duration values (*M*=80, *SD*=9.84 and *M*=113, *SD*=11.68). The duration means of /yː/ in *Lüge* and *Bücher* (*M*=113, *SD*=22.86 and *M*=119, *SD*=22.63) were considerably longer than the ones of /ʏ/. Long vowels were

also characterised by greater SD and wider min-max ranges, which indicates greater variability in the production compared to short vowels. Table 18 outlines the L3 vowel duration means in the T1 testing session.

Table 18. L3 vowel duration (in ms) at T1.

Target word	vowel	N	M (ms)	95% CI	SD	Median	Min	Max
Mitte	/ɪ/	20	73	68.21-77.69	10.12	72.50	54	96
Bitte	/ɪ/	20	78	73.90-82.80	9.50	78.00	60	91
Linie	/iː/	20	128	118.51-138.09	20.91	127.51	92	158
Miete	/iː/	20	96	89.97-102.23	13.09	96.00	77	128
füllen	/ʏ/	20	80	75.14-84.36	9.84	81.00	57	101
küssen	/ʏ/	20	79	73.63-84.57	11.68	78.00	57	102
Lüge	/yː/	20	113	101.90-123.30	22.86	104.00	85	160
Bücher	/yː/	20	119	108.51-129.69	22.63	108.00	96	165

At T2, the obtained vowel duration measurements followed a similar pattern to that observed at T1. The vowel duration means of /ɪ/ in the words *Mitte* and *Bitte* were the shortest in the target word set (*M*=80, *SD*=7.81 *M*=78, *SD*=12.18), along with *füllen* and *küssen* (*M*=78, *SD*=21.02 and *M*=84, *SD*=17.05). The duration of /iː/ in *Miete* and *Linie* retained similar values compared to T1 (*M*=98, *SD*=14.5 and *M*=120, *SD*=18.87), with a slight decrease in the case of the latter target word. The duration means of /yː/ in *Lüge* and *Bücher* were considerably longer than the realisations of /ʏ/ (*M*=117, *SD*=20.80 and *M*=116, *SD*=13.61). Table 19 outlines the L3 vowel duration means in the T2 testing session.

Table 19. L3 vowel duration (in ms) at T2.

Target word	vowel	N	M (ms)	95% CI	SD	Median	Min	Max
Mitte	/ɪ/	20	80	76.14-83.46	7.81	78.50	69	94
Bitte	/ɪ/	20	78	72.30-83.70	12.18	82.50	54	95
Linie	/iː/	20	120	111.62-129.28	18.87	124.00	85	153
Miete	/iː/	20	98	91.92-105.08	14.05	94.50	80	130
füllen	/ʏ/	20	78	68.41-88.09	21.02	82.00	0	109
küssen	/ʏ/	20	84	76.42-92.38	17.05	83.00	50	116
Lüge	/yː/	20	117	108.20-127.00	20.08	109.00	97	167
Bücher	/yː/	20	116	110.13-122.87	13.61	114.00	100	145

At T3, the vowel duration measurements displayed similar regularities to that at T1 and T2. The vowel duration means of /ɪ/ in the words *Mitte* and *Bitte* were the shortest in the target word set (*M*=76, *SD*=7.76 and *M*=75, *SD*=8.49), followed closely by *füllen* and *küssen* (*M*=81, *SD*=6.66 and *M*=75, *SD*=10.62). The duration of /iː/ in *Miete* and *Linie* retained similar values compared to T1 and T2 (*M*=99, *SD*=14.64 and *M*=123, *SD*=17.81). The duration means of /yː/ in *Lüge* and *Bücher* were considerably longer than those registered in the previous testing sessions (*M*=127, *SD*=20.26 and *M*=125, *SD*=18.34). Table 20 outlines the L2 vowel duration means in the T3 testing session. Figure 16 presents the vowel duration means for individual target words at three testing times.

Table 20. L3 vowel duration (in ms) at T3.

Target word	vowel	N	M (ms)	95% CI	SD	Median	Min	Max
Mitte	/ɪ/	20	76	71.96-79.24	7.76	75.00	57	91
Bitte	/ɪ/	20	75	71.33-79.27	8.49	76.00	61	90
Linie	/iː/	20	123	114.66-131.34	17.81	120.00	98	173
Miete	/iː/	20	99	92.15-105.85	14.64	96.50	72	126
füllen	/ʏ/	20	81	77.93-84.17	6.66	82.00	70	92
küssen	/ʏ/	20	75	70.48-80.42	10.62	76.50	54	89
Lüge	/yː/	20	127	117.72-136.68	20.26	129.00	92	174
Bücher	/yː/	20	125	116.77-133.93	18.34	123.00	101	168

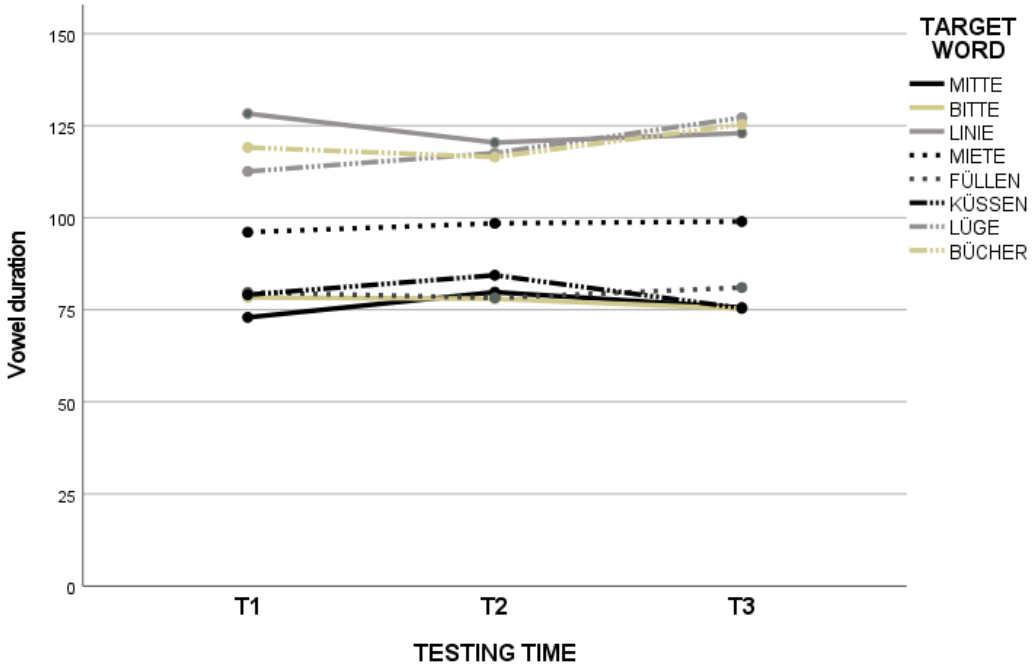


Figure 16. L3 vowel duration means across three testing sessions for individual target words.

Vowels duration measurements were grouped according to their type as either /ɪ/, /ʏ/, /iː/ or /yː/. The means for grouped vowels revealed that in three testing sessions the duration of /ɪ/ and /ʏ/ was shorter compared to /iː/ and /yː/. Moreover, individual variability in the duration of short vowels was smaller compared to long vowels, as indicated by SD and min-max ranges. The mean duration measurements grouped according to the vowel type in three testing sessions are presented in Table 21. Means of four vowels for individual participants are presented in Figure 17, Figure 18, Figure 19 and Figure 20.

Table 21. L3 vowel duration means for four vowel types in three testing sessions.

vowel	testing session	M (ms)	95% CI	SD	Median	min	Max
/ɪ/	T1	72	72.039-79.261	7.71	74.75	62	91
/ʏ/	T1	79	75.143-83.206	8.61	78.75	61	98
/iː/	T1	112	106.561-117.838	12.04	117.00	91	128
/yː/	T1	115	108.850-123.349	16.02	111.00	93	158
/ɪ/	T2	79	75.738-82.061	6.75	78.75	69	90
/ʏ/	T2	81	75.281-87.368	12.91	82.75	51	100
/iː/	T2	109	102.820-116.129	14.21	110.00	86	130
/yː/	T2	117	110.961-123.138	13.00	113.25	104	146

/ɪ/	T3	75	72.093-78.806	7.17	74.50	61	87
/ʏ/	T3	78	75.250-81.242	6.39	77.25	66	90
/iː/	T3	111	104.880-117.119	13.07	111.25	86	135
/yː/	T3	126	118.226-134-323	17.19	123.24	97	171

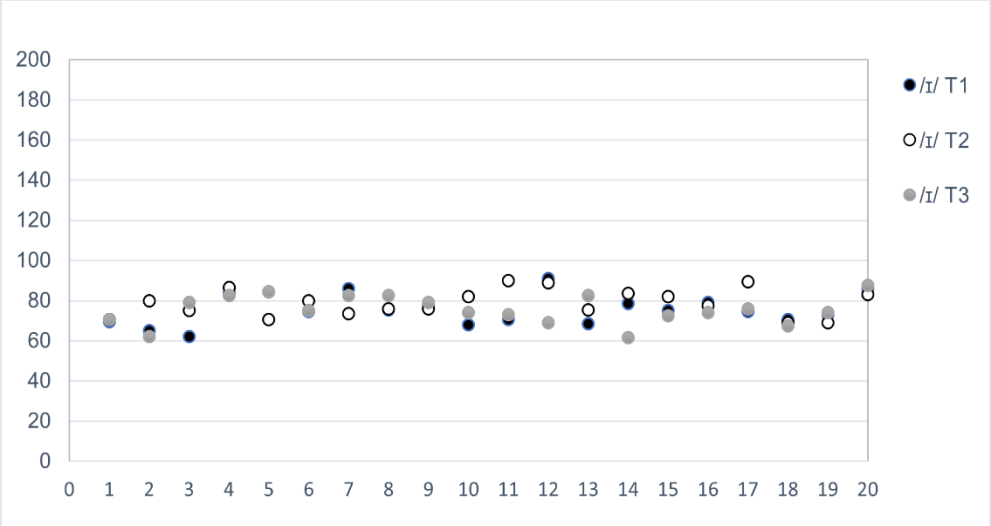


Figure 17. Individual /ɪ/ duration means (in ms) for 20 participants in three testing sessions.

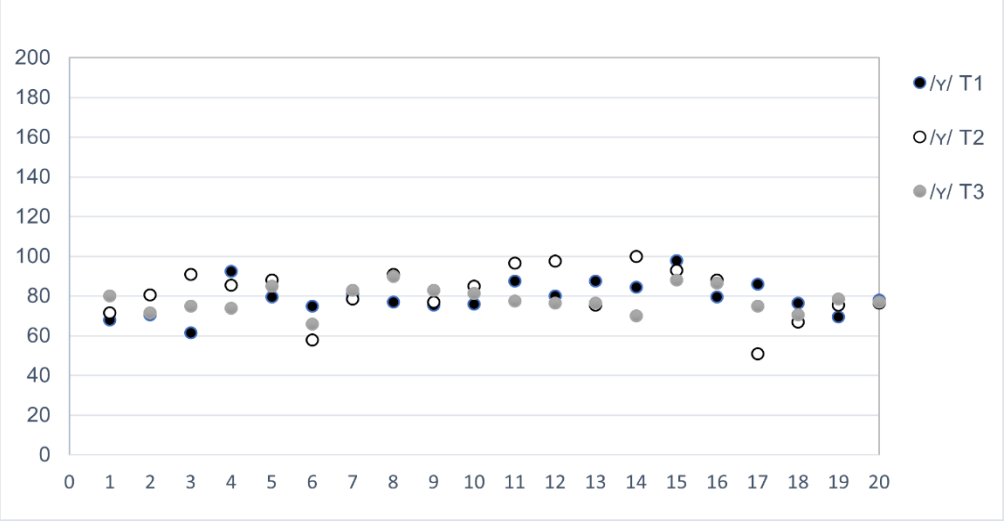


Figure 18. Individual /ʏ/ duration means (in ms) for 20 participants in three testing sessions.

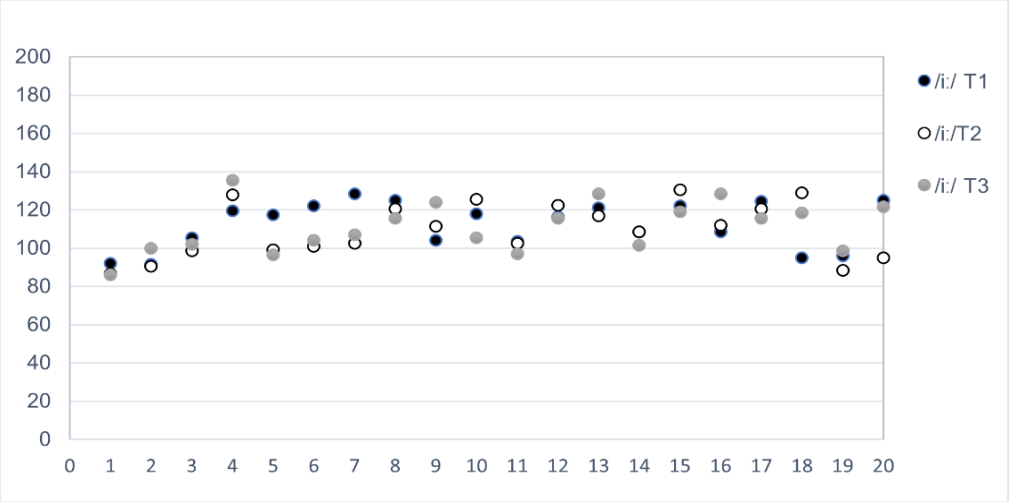


Figure 19. Individual /iː/ duration means (in ms) for 20 participants in three testing sessions.

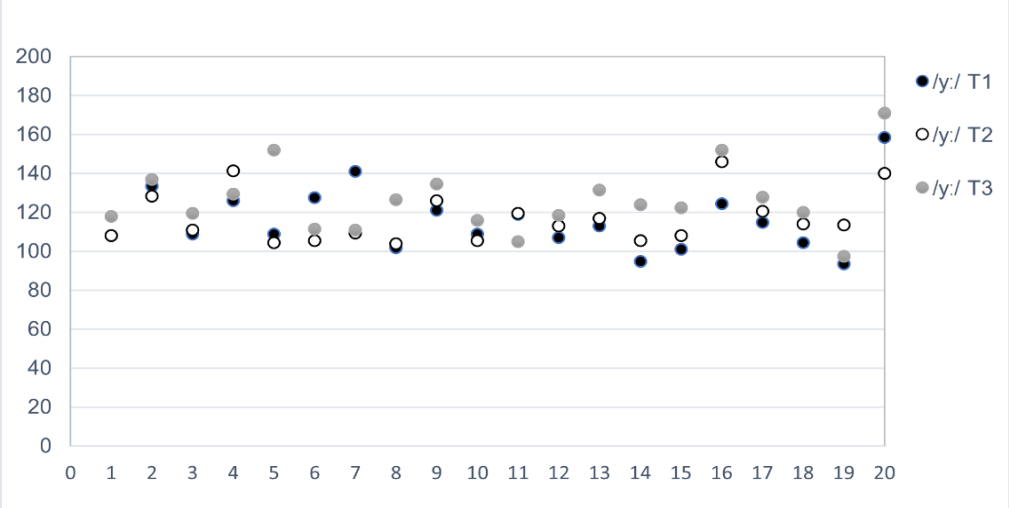


Figure 20. Individual /yː/ duration means (in ms) for 20 participants in three testing sessions.

The vowel duration means of four vowel types in L3 German obtained in three testing sessions were compared through repeated measures ANOVA. For the condition "vowel type", sphericity was violated as indicated by Mauchly's tests, $X^2(5)=18.06$, $p=.002$. Therefore, the Greenhouse-Geisser correction was applied. Repeated measures ANOVA revealed a highly significant effect of vowel type $F(1.97, 57)=142.69$, $p<.001$, $\eta_p^2=.882$, which indicates that the duration means of four vowel types differed significantly. Post-hoc tests with a Bonferroni adjustment for the condition "vowel" revealed that mean duration of the pairs /ɪ/ - /iː/ and /ʏ/ - /yː/ differed significantly at $p<.001$, whereas the duration of the pairs /ɪ/ - /ʏ/ and /iː/ - /yː/ exhibited no such difference

at respectively p=1 and p=.007 Such results of post-hoc tests imply a significant difference between the duration of short vs. long vowels. Moreover, there was a highly significant effect of time*vowel type, $F(6, 114)=3.71$, p <.002, η_p^2=.164. Such a result indicates that a significant change over time was visible for certain vowel types. Post-hoc tests with a Bonferroni adjustment revealed a significant difference between the /yː/ vowel duration means at T1 vs. T3 and T2 vs. T3 at p=.019 and p=.001. There was no significant overall effect of time, $F(2, 38)=.705$, p=.5, η_p^2=.160. Figure 21 presents vowel means in L3 German over three testing times.

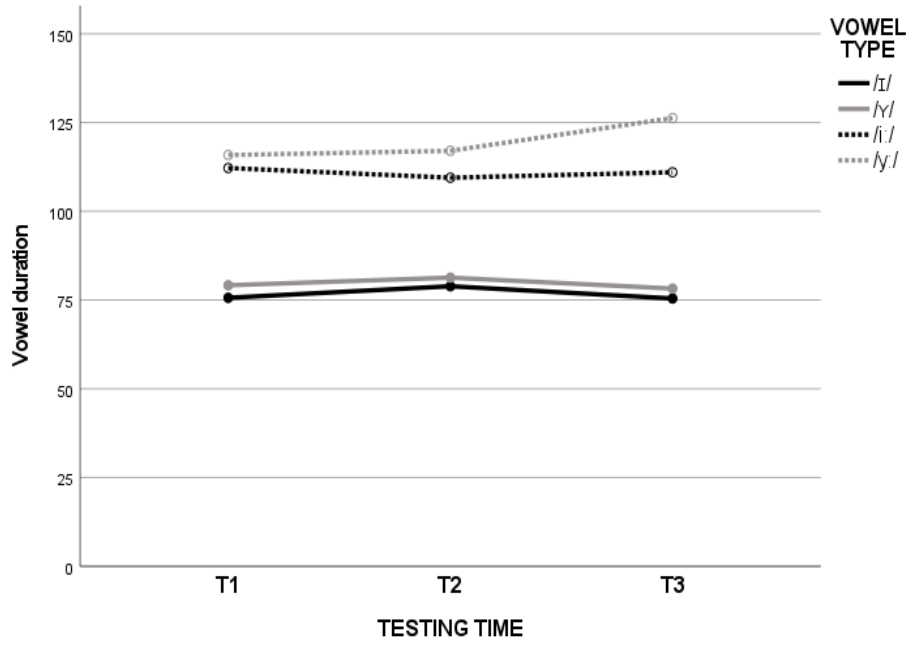


Figure 21. Vowel duration means for /ɪ, ʏ, iː, yː/ in three testing sessions.

5.3.3. VOT in L2 English

The production of VOT in voiceless plosives in L2 English was analysed in Praat software. VOT of initial voiceless plosives /p, t, k/ was measured from the onset of the stop burst to the first voicing pulse of the subsequent vowel. Each language version of the task included 6 VOT tokens (two per initial voiceless plosive). Means presented in the table below include '0' values for mispronunciations of the plosive (unlike the earlier presentation of the T1 portion of the data in the Appendix 2 of Krzysik 2020, which presented means for target-like, hybrid and non-target like realisations).

In the T1 testing session, the overall VOT of /p/ oscillated around 30 ms (*pot* – *M*=30, *SD*=19.27; *party* – *M*=33, *SD*=20.39) whereas /t/ displayed VOT values around 40 ms (*toss* – *M*=38, *SD*=14.68; *target* – *M*=40, *SD*=18.44). The participants produced the longest VOT values in the case of /k/ - around 60 ms (cot – *M*=60, *SD*=14.10; *carpet* – *M*=59, *SD*=24.77). The high SD values and rather wide min-max ranges indicate a considerable level of variability between the participants. Table 22 outlines the L3 VOT means in T1 for individual target words.

Table 22: VOT measurements (in ms) for L2 at T1.

Target word	aspi-rated plosive	N	M (ms)	95% CI	SD	Median	Min	Max
pot	/p/	20	30	20.83-38.87	19.27	27.00	0	73
party	/p/	20	33	23.40-42.50	20.39	27.00	14	91
toss	/t/	20	38	31.12-44.88	14.68	38.50	0	77
target	/t/	20	40	30.97-48.23	18.44	37.50	0	77
cot	/k/	20	60	53.70-66.90	14.10	60.50	41	89
carpet	/k/	20	59	47.61-70.79	24.77	68.00	0	96

A similar pattern was observed at T2; the production of /k/ was characterised by the longest VOT values (*cot* – *M*=71, *SD*=20.73; *carpet* – *M*=71, *SD*=20.90), increased by approximately 10 ms in comparison to T1. The production of /p/ noted a similar increase (*pot* – *M*=41, *SD*=17.56; *party* – *M*=40, *SD*=17.56). In the case of /t/ production, the increase was slightly more pronounced for the word *toss* (*M*=48, *SD*=18.19) than for *target* (*M*=44, *SD*=16.91). The production of /p/ and /t/ exhibited a smaller variability compared to /k/, which was indicated by lower SD values. Table 23 outlines the L2 English VOT means in T2 for individual target words.

Table 23: L2 VOT means (in ms) for L2 at T2.

Target word	aspi-rated plosive	N	M (ms)	95% CI	SD	Median	min	Max
pot	/p/	20	41	32.20-49.90	18.91	36.00	18	87
party	/p/	20	40	31.95-47.45	17.56	38.00	13	71
toss	/t/	20	48	39.53-56.57	18.19	39.00	31	88
target	/t/	20	44	36.08-51.92	16.91	41.00	24	87
cot	/k/	20	71	60.85-80.25	20.73	71.50	16	118
carpet	/k/	20	71	61.57-81.13	20.90	67.00	38	120

At T3, the realisations of /p/ were characterised by the VOT values between T1 and T2 levels (pot – *M*=36, *SD*=21.78; party – *M*=35, *SD*=19.25). VOT means for /t/ exhibited a difference depending on the target word (toss – *M*=41, *SD*=16.73; target – *M*=52, *SD*=12.51); the word *target* was also marked by greater variability. The VOT in the production of /k/ noted a slight decrease compared to T2 (cot – *M*=69, *SD*=23.44; carpet – *M*=67, *SD*=17.72). Table 24 outlines the L2 English VOT means at T3.

Table 24. VOT measurements (in ms) for L2 at T3.

Target word	aspi-rated plosive	N	M (ms)	95% CI	SD	Median	min	Max
pot	/p/	20	36	26.16-46.54	21.78	33.00	0	85
party	/p/	20	35	26.84-44.86	19.25	28.00	11	78
toss	/t/	20	41	34.27-48.33	15.01	37.00	18	75
target	/t/	20	52	41.84-62.66	22.23	49.50	28	112
cot	/k/	20	69	58.94-80.86	23.44	67.50	38	128
carpet	/k/	20	67	58.75-75.35	17.72	64.00	34	115

The production of L2 English VOT over three testing times showed a number of patterns. The production of /p/ noted an increase in VOT values between T1 and T2 and a slight decrease at T3. The VOT means for /t/ showed a slight upward trend from T1 to T3 and a subsequent divergence from T2 to T3 – the word target continued with the initial trend, whereas the word toss noted a decrease. The production of /k/ was characterised by an increase in VOT values between T1 and T3 and a slight decrease at T3. Figure 22 outlines the VOT means for individual target words at the three testing times.

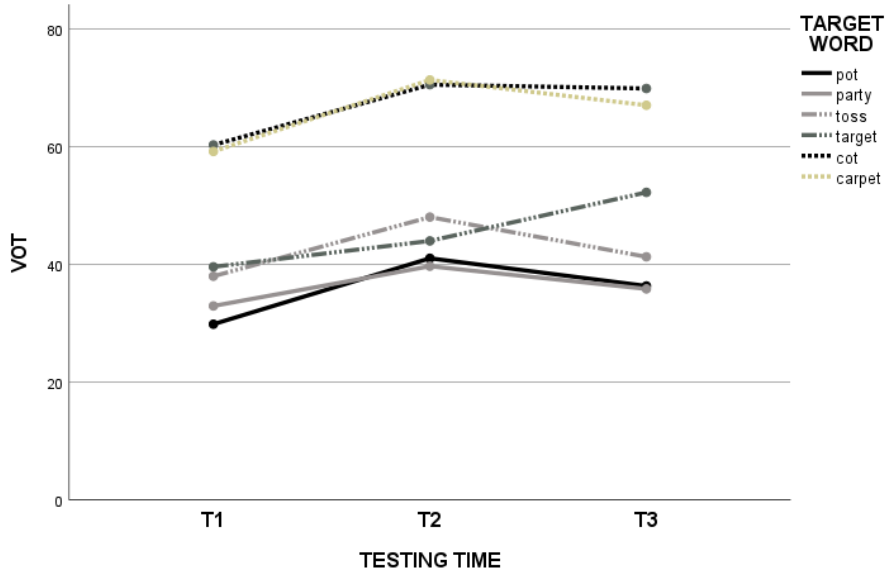


Figure 22. VOT means for the L2 target words at three testing times.

The VOT measurements from individual words were grouped according to the place of articulation as /p/, /t/, /k/; each category was a mean of two realisations containing a given plosive. The production of /k/ was characterised by the longest VOT values, whereas /p/ was marked by the shortest. Individual variability in VOT production was the greatest at T3. The mean VOT measurements grouped according to the aspirated plosive in three testing sessions are presented in Table 25. Individual means for three groups are presented in Figure 23, Figure 24 and Figure 25.

Table 25. Means of VOT measurements per aspirated L2 plosive type (in ms) at T1, T2 and T3.

Aspi-rated plosive	Testing session	N	M (ms)	95% CI	SD	Median	Min	Max
/p/	T1	20	31	23.09-39.71	17.76	25.75	10	82
/t/	T1	20	29	33.77-43.83	10.74	37.50	14	61
/k/	T1	20	60	51.90-67.60	16.78	59.00	27	88
/p/	T2	20	40	33.00-47.75	15.75	38.25	16	70
/t/	T2	20	46	39.19-52.86	14.61	39.25	30	73
/k/	T2	20	71	63.29-78.61	16.37	70.25	33	102
/p/	T3	20	36	28.00-44.20	17.29	34.75	16	82
/t/	T3	20	47	38.61-54.94	17.44	42.25	29	90
/k/	T3	20	68	59.96-76.99	18.19	66.75	40	107

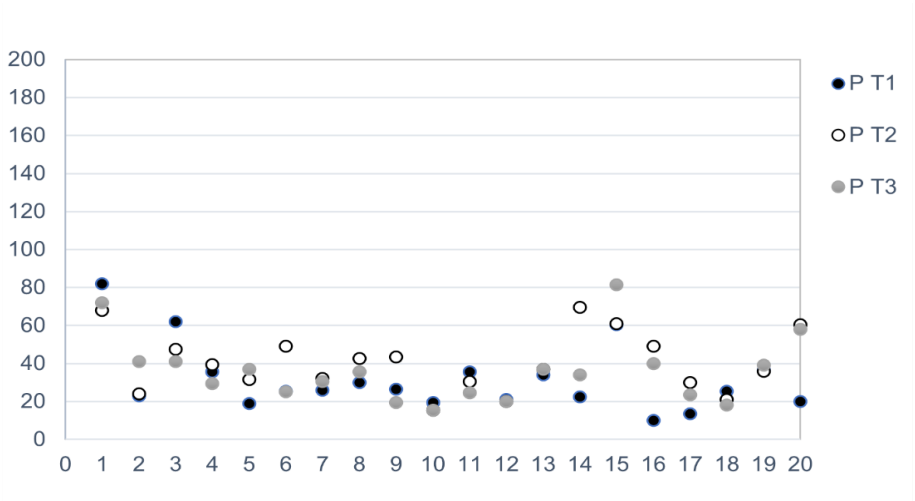


Figure 23. Individual VOT duration means (in ms) for L2 /p/ of 20 participants in three testing sessions.

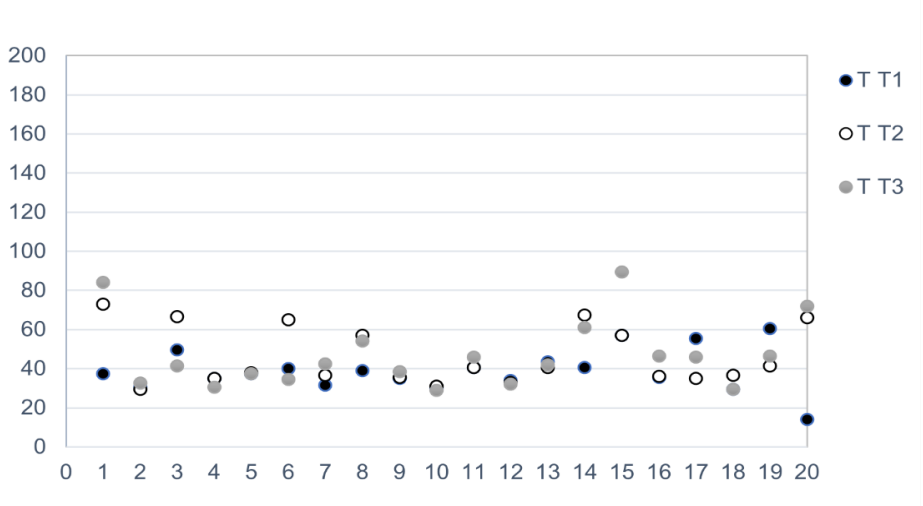


Figure 24. Individual VOT duration means (in ms) for L2 /t/ of 20 participants in three testing sessions.

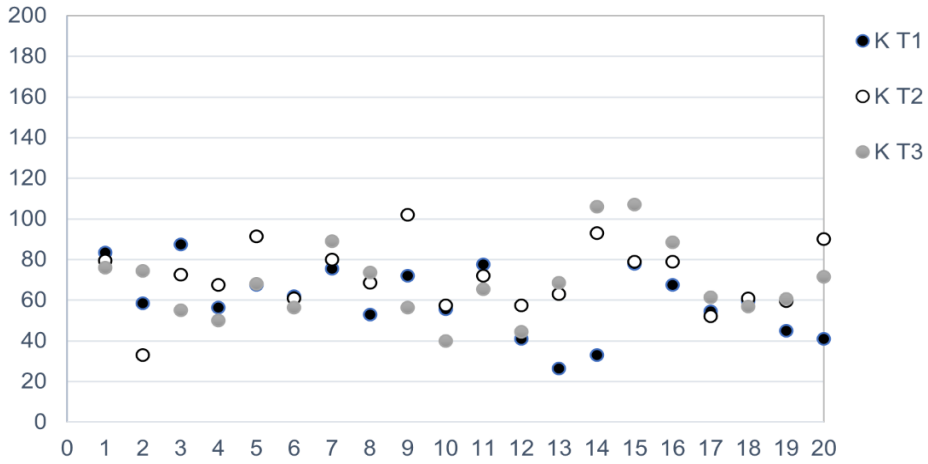


Figure 25. Individual VOT duration means (in ms) for L2 /k/ for 20 participants in three testing sessions.

The VOT duration means obtained in three testing sessions were compared through repeated measures ANOVA. For the condition "aspirated plosive type", sphericity was violated as indicated by Mauchly's tests, $X2(2)=6.14$, $p=.049$, therefore, the Huynh-Feldt correction was applied, whereas the remaining conditions met the assumption of sphericity. Repeated measures ANOVA revealed a significant effect of time, $F(2, 38)=4.65$, $p=.016$, $\eta p2=.197$. Such a result indicates an overall change in VOT production over time. Moreover, there was a highly significant effect of the plosive type $F(1.66, 29.48)=94.35=7$, $p<.001$, $\eta p2=.832$, implying a difference between the VOT duration for each plosive type. However, there was no significant interaction of time*aspirated plosive type, $F(4, 76)=.55$, $p=.69$, $\eta p2=.028$; the VOT durations for aspirated plosive type remained generally stable over the testing sessions. The post-hoc tests with Bonferroni adjustment revealed a significant difference between the testing sessions T1-T2 ($p=.018$) and the pairs of plosives /p/-/t/ ($p=.004$), /p/-/k/ ($p<.001$) and /t/-/k/ ($p<.001$). Figure 26 presents the vowel duration means over three testing sessions.

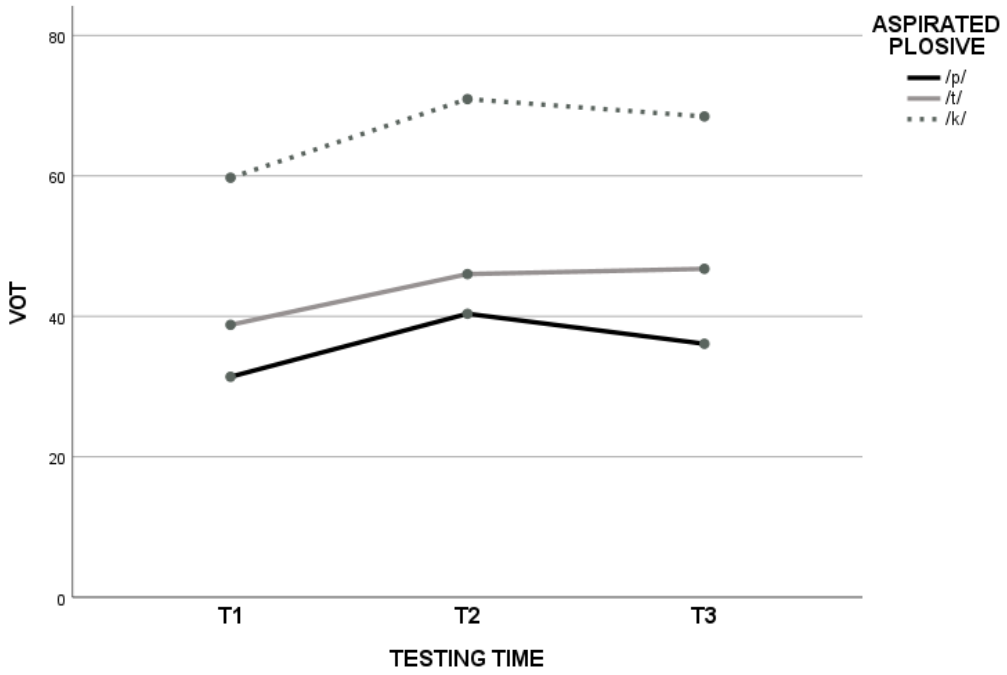


Figure 26. L2 VOT means for each aspirated plosive type at T1, T2 and T3.

5.3.4. VOT in L3 German

The production of VOT in voiceless plosives in L3 German was analysed in Praat software. VOT of initial voiceless plosives /p, t, k/ was measured from the onset of the stop burst to the first voicing pulse of the subsequent vowel. Each language version of the task included 6 VOT tokens (two per each place of articulation (PoA) of the initial voiceless plosives. Means presented in the table below include 0 values for mispronunciations of the plosive in the calculation (unlike the earlier presentation of the T1 portion of the data in the Appendix 2 of Krzysik 2020, which presented means for target-like, hybrid and non-target like realisations).

At T1, the production of L3 VOT was characterised by the following pattern - overall VOT of /p/ and /t/ oscillated around the means of 20-30 ms, which can be contrasted with longer VOT for /k/ - between 40-50 ms. Table 26 outlines the L3 VOT means at T1.

Table 26. VOT measurements (in ms) for L3 at T1 testing session.

Target word	aspi-rated plosive	N	M (ms)	95% CI	SD	Median	Min	Max
Pott	/p/	20	30	24.41-36.29	13.75	29.50	0	54
packe	/p/	20	29	21.67-36.53	15.87	23.50	0	65
toll	/t/	20	20	9.55-30.95	22.85	18.00	0	70
Tante	/t/	20	29	20.44-37.96	18.72	26.00	0	76
Koch	/k/	20	50	39.93-59.67	21.08	50.00	0	88
Kappe	/k/	20	43	31.96-53.14	22.63	46.50	0	98

At T2, the production of /t/ was defined by longer VOT values (*toll* – *M*=51, *SD*=18.76; *Tante* – *M*=43, *SD*=20.43), comparable to the longest in the set, exhibited by /k/ (*Koch* – *M*=63, *SD*=12.33; *Kappe* – *M*=52, *SD*=11.65). The realisations of /p/ retained values comparable to T1 (*Pott* – *M*=29, *SD*=12.94; *packe* – *M*=26, *SD*=11.77). The production of /p/ and /k/ exhibited a smaller variability compared to /t/, which was indicated by lower SD values. The word *toll* showed the greatest increase in VOT since T1 among the entire L3 target word set. Table 27 outlines the L3 VOT means at T2.

Table 27. VOT measurements (in ms) for L3 at T2 testing session..

Target word	aspi-rated plosive	N	M (ms)	95% CI	SD	Median	min	Max
Pott	/p/	20	29	23.04-35.16	12.94	28.00	12	50
packe	/p/	20	26	20.76-31.81	11.77	21.00	12	48
toll	/t/	20	51	42.77-60.33	18.76	49.50	27	89
Tante	/t/	20	43	33.14-52.26	20.43	44.00	0	78
Koch	/k/	20	63	57.18-68.67	12.33	65.50	33	80
Kappe	/k/	20	52	46.19-57.11	11.65	50.50	36	76

At T3 the production of /t/ exhibited similar values to those in T1 (toll – *M*=51, *SD*=16.73; *Tante* – *M*=37, *SD*=12.51), however, the participants were more uniform, as indicated by SD values. The production of /k/ was marked by VOT values comparable to those in T2 (*Koch* – *M*=57, *SD*=12.36; *Kappe* – *M*=47, *SD*=10.62). The realisations of /p/ retained values comparable to T1 and T2 (*Pott* – *M*=27, *SD*=11.74; *packe* – *M*=28, *SD*=12.09). The overall VOT production exhibited smaller variability compared to the previous session, as indicated by lower SD values. Table 28 outlines the L3 VOT means at T3.

Table 28. VOT measurements (in ms) for L3 at T3 testing session.

Target word	aspi-rated plosive	N	M (ms)	95% CI	SD	Median	Min	Max
Pott	/p/	20	27	21.25-32.25	11.74	25.00	11	59
packe	/p/	20	28	22.39-33.71	12.09	24.50	11	48
toll	/t/	20	51	43.17-58.83	16.73	52.00	21	78
Tante	/t/	20	37	31.08-42.82	12.51	37.50	20	61
Koch	/k/	20	57	51.57-63.13	12.36	61.50	32	73
Kappe	/k/	20	47	46.83-56.77	10.62	51.00	33	69

The production of L3 VOT over the three testing times exhibited several patterns. The production of /t/ noted an increase in VOT values between T1 and T2 and a slight decrease or stability at T3. The production of / k/ followed a slight upward trend with a decrease (*Koch*) or a slight increase (*Kappe*) at T3. The VOT means for /t/ remained rather stable over three testing times. Figure 27 outlines the VOT means for individual target words at three testing times.

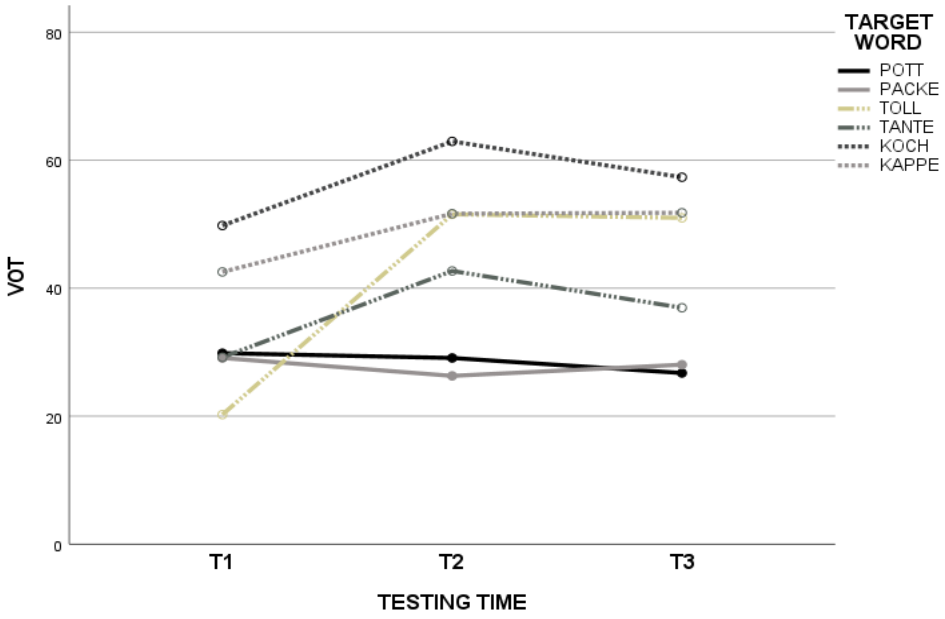


Figure 27. VOT means for the L3 target words in three testing times.

Similarly to the L2 data, the L3 VOT measurements were grouped according to the place of articulation of the aspirated voiceless plosive as /p/, /t/, /k/; each category was a mean of two realisations containing a given plosive. Comparably to L2 data, the production of /k/ was characterised by the longest VOT values, whereas /p/ was marked by the shortest. Individual variability in VOT production was the least pronounced at T3. The mean VOT measurements grouped according to the aspirated plosive in three testing sessions are presented in Table 29. Individual means for three groups are presented in Figure 28, Figure 29 and Figure 30.

Table 29. Means of VOT measurements per L3 aspirated plosive type (in ms) at T1, T2 and T3.

Aspi-rated plosive	Testing session	N	M (ms)	95% CI	SD	Median	Min	Max
/p/	T1	20	29	25.38-33.56	8.73	27.75	17	51
/t/	T1	20	25	24.39-23.25	15.57	23.25	0	56
/k/	T1	20	46	38.75-53.60	15.86	21.00	21	93
/p/	T2	20	28	42.86-32.53	10.38	29.25	13	48
/t/	T2	20	47	39.05-55.19	17.24	44.75	15	78
/k/	T2	20	57	53.05-61.54	9.07	57.50	42	75
/p/	T3	20	23	26.72-27.00	9.52	27.00	13	53
/t/	T3	20	44	37.83-50.12	13.12	45.75	21	66
/k/	T3	20	55	50.53-58.61	8.63	56.00	38	69

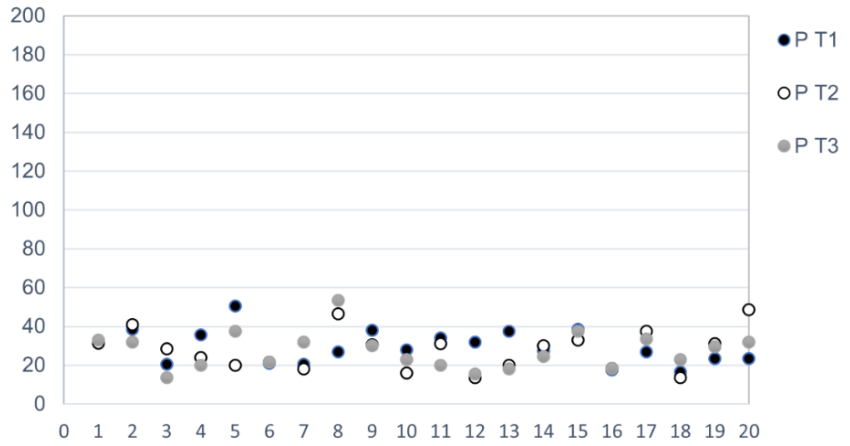


Figure 28. Individual VOT duration means (in ms) of L3 /p/ for 20 participants in three testing sessions.

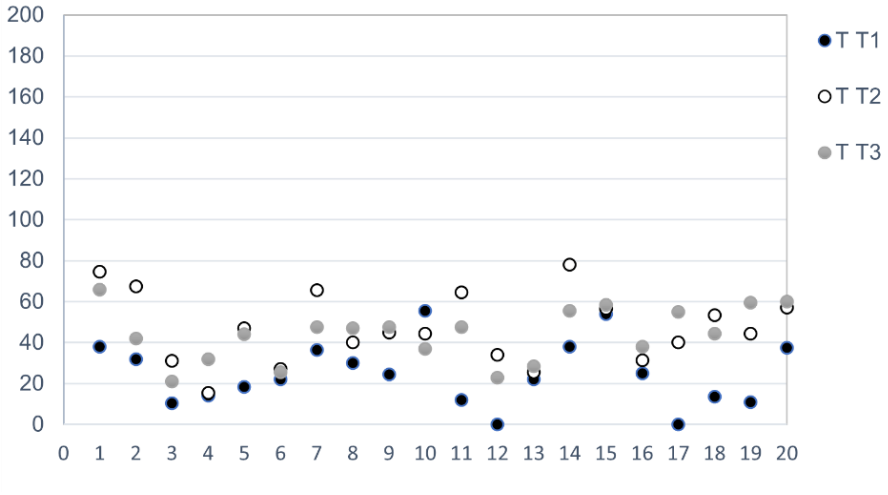


Figure 29. Individual VOT duration means (in ms) of L3 /t/ for 20 participants in three testing sessions.

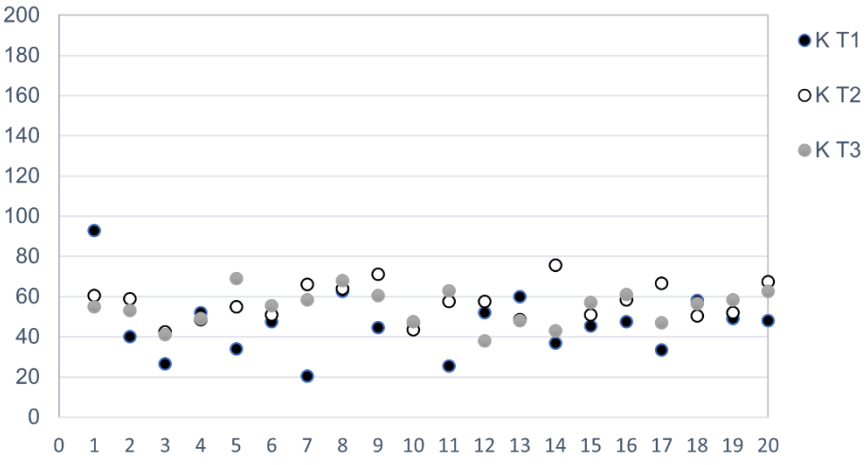


Figure 30. Individual VOT duration means (in ms) of L3 /k/ for 20 participants in three testing sessions.

The VOT duration means obtained in the three testing sessions were compared through repeated measures ANOVA. Repeated measures ANOVA revealed a significant effect of time, $F(2, 38)=13.41$, $p < .001$, $\eta p2=.414$, indicating an overall change in L3 VOT production over the testing sessions. Furthermore, there was a highly significant effect of aspirated plosive type $F(2, 38)=58.27$, $p<.001$, $\eta p2=.754$, which implied a difference in the VOT between the plosive types. Additionally, there was a significant interaction of time*aspirated plosive type, $F(4, 76)=10.18$, $p<.001$, $\eta p2=.349$; the VOT durations by plosive type changed over time. The post-hoc tests with Bonferroni adjustment revealed a significant difference between the testing sessions T1-T2 ($p<.001$) and T1-T3 ($p=.001$) and plosive pairs /p/-/t/ ($p<.001$), /p/-/k/ ($p<.001$) and /t/-/k/ ($p<.001$). Moreover, the post-hoc test with Bonferroni adjustment showed a significant difference between the mean VOT durations over time in the realisations of /t/ (T1-T2, $p<.001$, T1-T3, $p<.001$) and /k/ (T1-T2, $p<.001$; T1-T3, $p<.001$). Figure 31 presents the VOT means over three testing sessions.

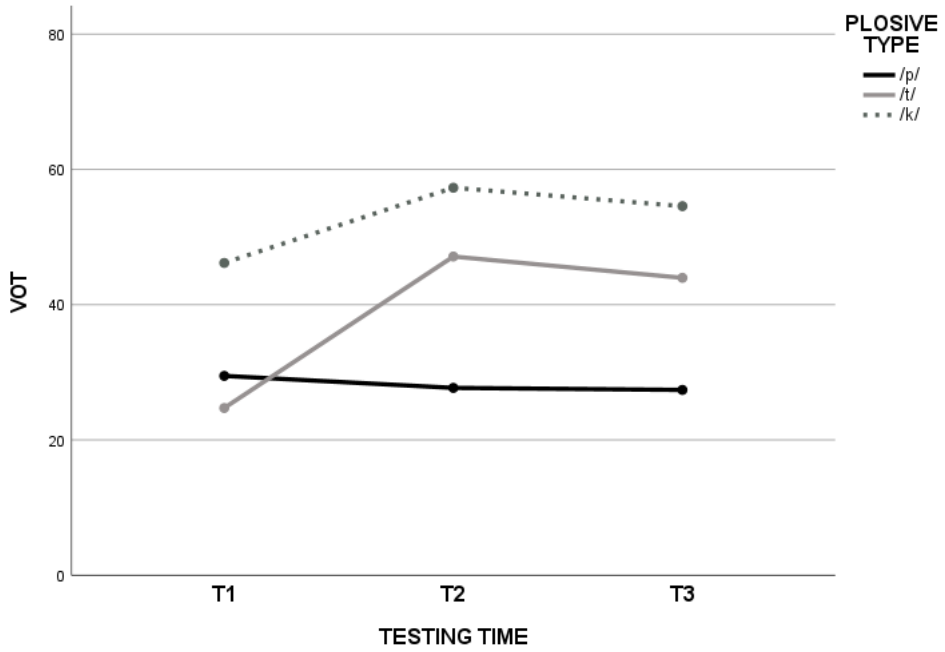


Figure 31. L3 VOT means for each aspirated plosive type at T1, T2 and T3.

5.3.5. Summary of the speech production results

The delayed repetition tasks in L2 and L3 tested the participants' production of vowels (close front unrounded vowels for L2 English; close front rounded and unrounded vowels for L3 German) and aspirated voiceless plosives, which were subsequently analysed through acoustic measurements of vowel duration and VOT. For both L2 and L3, there was a significant difference in the duration of long and short vowels, however, there was no overall change over time. As for L2 English, the duration of the particular vowels did not change as well, whereas in the case of L3 German the change was detected for one vowel, namely /yː/. VOT durations for L3 and L2 showed an overall universal pattern, i.e. velar /k/ was characterised by the longest VOT, followed by alveolar /t/ and bilabial /p/. For L2, there was a significant change over time between the T1-T2 testing sessions, whereas for L3 the difference was significant for T1-T2 and T1-T3. The difference between the plosives was significant for all pairs (/p/-/t/, /p/-/k/ and /t/-/k/) in both languages. In L2, the durations of aspirated plosive by type remained generally stable over the testing sessions. In L3, /t/ and /k/ duration increased significantly between T1 and T2. The summary of the findings is outlined in Table 30.

Table 30. Summary of the speech production results.

Tested feature	Language	Overall findings	Overall change over time	Individual features
Vowel duration	L2	Significant difference between short and long vowels T1, T2, T3	No significant change over time	-
Vowel duration	L2	Significant difference between short and long vowels in T1, T2, T3	No significant change over time	Significant increase of /yː/ duration at T3
VOT	L2	Overall difference between the VOT duration of /p, t, k/	Overall difference between T1-T2	Individual plosive types remained stable
VOT	L3	Overall difference between the VOT duration of /p, t, k/	Overall difference between T1-T2 and T1-T3	Significant increase of /t/ and /k/ duration between T1 and T2

5.4. Multilingual perception and production – the development of the modalities

To evaluate the relationship between the multilingual perception and production of the participants the data collected in speech perception and speech production tasks were transformed to enable comparison. Section 5.4.1 presents the perception (ABX) data in the form of percentages reflecting accuracy rate. Section 5.4.2. outlines the transformation of the production (delayed repetition) data into the scores reflecting the approximation of the L2 or L3 native model. The perception and production data are compared in section 5.4.3. to investigate the pattern of development between the two modalities in the respective languages.

5.4.1. Multilingual perception accuracy

In the case of multilingual perception, the L2 and L3 ABX scores were already transformed into percentages. The section below presents the overall multilingual perception data for L2 and L3. For L2, the data encompasses the accuracy scores from the trials testing the /ɪ/-/iː/ contrast, whereas for L3 the data includes the trials testing the /ɪ/-/iː/ and /ʏ/-/yː/ contrasts combined.

The overall perception accuracy (OPercAcc) can be described as generally higher for L3 than L2. In T1, the accuracy level of L2 reached the mean percentage of *M*=48 (*SD*=11.60) and for L3 it was *M*=61 (SD=10.70). At T2, the perceptual accuracy increased (*M*=51, *SD*=9.04); a similar tendency was observed in L3 (*M*=63, *SD*=9.17). At T3, L2 accuracy returned to levels comparable to T1 (*M*=49, *SD*=11.27), while the L3 accuracy exhibited the highest accuracy (*M*=65, *SD*=9.59). Table 31 presents for OPercAcc in L2 and L3 across the three testing sessions. Figure 32 shows OPercAcc scores in L2 and L3 in three testing sessions, whereas Figure 33 and Figure 34 demonstrate individual perception accuracy scores (%) obtained by the participants in L2 and L3 across the three testing sessions.

Table 31. Overall perception accuracy (OPercAcc) scores for L2 and L3 in three testing sessions.

ABX	Testing session	N	OPercAcc	95% CI	SD	Median	Min	Max
L2	T1	20	48	42.996-53.878	11.62	43.75	31.25	68.75
L3	T1	20	61	55.927-65.948	10.70	62.50	43.75	87.50
L2	T2	20	51	46.391-54.858	9.04	50.00	37.50	68.75
L3	T2	20	63	58.518-67.106	9.17	62.50	43.75	75.00
L2	T3	20	49	44.099-54.650	11.27	50.00	31.25	75.00
L3	T3	20	65	60.508-69.491	9.59	62.50	50.00	87.50

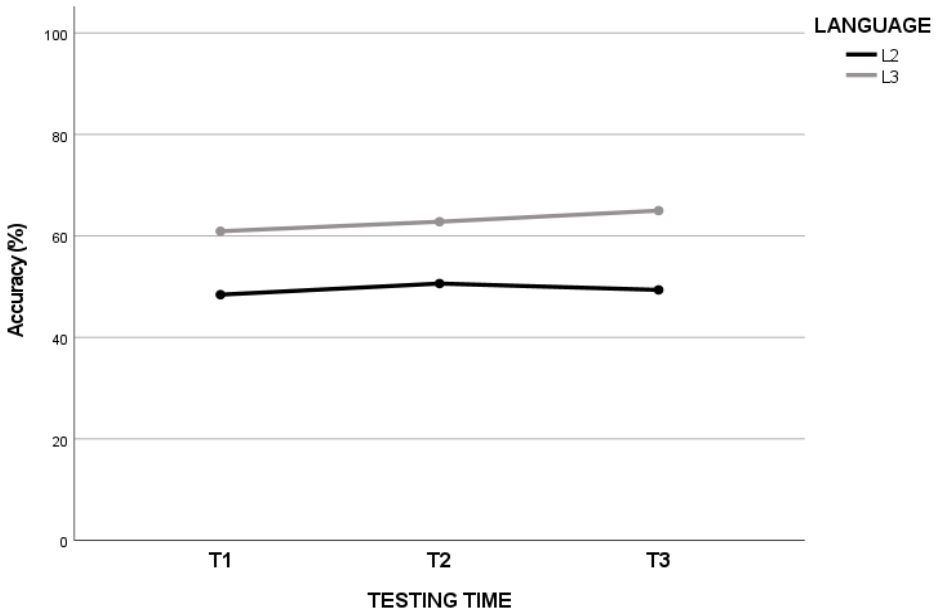


Figure 32. Overall perception accuracy (OPercAcc) means (%) for L2 and L3 in three testing sessions.

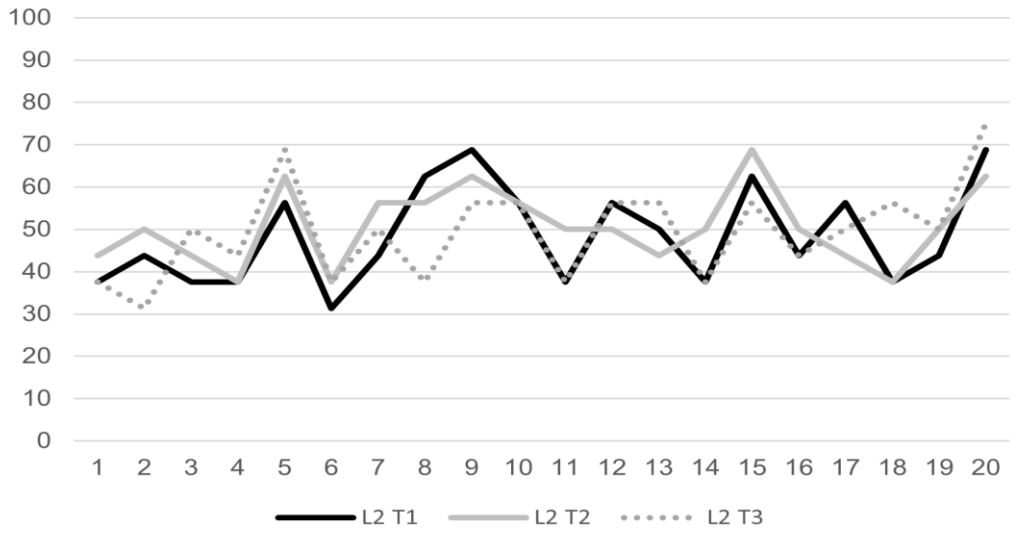


Figure 33. Individual perception accuracy (OPercAcc) (%) for L2 in three testing sessions for 20 participants.

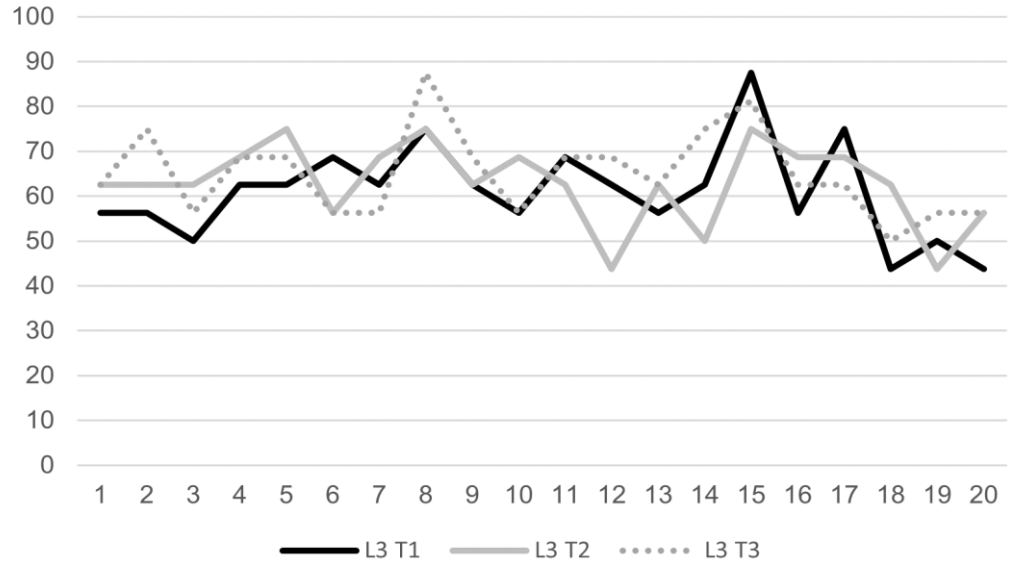


Figure 34. Individual perception accuracy (OPercAcc) (%) for L3 in three testing sessions for 20 participants.

5.4.2. Production accuracy calculations

The production scores posed a greater challenge than those of the perception – the raw measurements of VOT and vowel duration had to be transformed into scores reflecting

the approximation to the L2 or L3 native model. The method of approximation was based on the VOT goodness of fit scale applied by Wrembel (2015), where the levels of the scale corresponded with the distance from the mean VOT of the control group. The scale applied in the present study adopted a similar approach; the levels of the scale indicated the ”goodness of fit”, i.e. how close a participant’s realisation was to the target production. Additionally, the realisation on a particular level was awarded a corresponding score for the accuracy of the production in relation to the L2 target approximation. The scale had 11 levels; the first level encompassed values within the range of +/- 0-9 and indicated a 100% goodness of fit with the target production, whereas the second level corresponded to a +/- 10-19% etc. Every level was associated with accuracy points, starting with 1 point for a realisation fitting the first level and 0 points for the last level, descending by 0.1 points. The scale with the “difference from the target value” ranges, the level of target value approximation (in percentages) and accuracy points on each level is outlined in Table 32.

Table 32. The scale outlining the levels of approximation and accuracy points in relation to target production for L2 and L3.

Difference from the target value	Language	Individual features
+/- 0-9%	100%	1
+/- 10-20%	90%	0.9
+/- 20-29%	80%	0.8
+/- 30-39%	70%	0.7
+/- 40-49%	60%	0.6
+/- 50-49%	50%	0.5
+/- 60-69%	40%	0.4
+/- 70-79%	30%	0.3
+/- 80-89	20%	0.2
+/- 90-99	10%	0.1
+/- 100% or more	0%	0

The scale was applied both in the case of VOT and vowel duration in L2 and L3. The results for individual realisations in decimal fractions were added and transformed into percentages of the total maximum score.

In the case of vowels, the production accuracy (ProdAcc), as calculated with the application of the formula described above), was generally higher for L3 than L2. In T1,

the accuracy level of L2 reached the mean percentage of *M*=72 (SD=5.22) and for L3 it was *M*=86 (SD=3.32). In T2, L2 noted a slight increase in accuracy (*M*=77, *SD*=6.11) whereas L3 was characterised by a small decrease (*M*=83, *SD*=5.33). A reverse trend was observed in T3; L2 accuracy returned to levels comparable to T1, however, it exhibited a larger variation (*M*=73, *SD*=8.41). L3 accuracy noted the highest accuracy levels of all testing sessions (*M*=88, *SD*=3.45). Table 33 presents the vowel ProdAcc scores in L2 and L3 across the three testing sessions. Figure 35 and Figure 36 individual vowel ProdAcc scores obtained by the participants in L2 and L3 across three testing sessions.

Table 33. Vowel production accuracy scores (ProdAcc) for L2 and L3 in three testing sessions.

Lan-guage	Testing time	N	ProdAcc %	95% CI	SD	Median	Min	Max
L2	T1	20	72	72.81-77.69	5.22	76.67	67	87
L2	T2	20	77	73.80-79.53	6.11	76.67	65	87
L2	T3	20	73	68.73-76.69	8.41	71.67	60	85
L3	T1	20	86	83.82-86.93	3.32	86.25	80	93
L3	T2	20	83	80.94-85.93	5.33	83.13	73	94
L3	T3	20	88	86.38-89.62	3.45	87.50	83	96

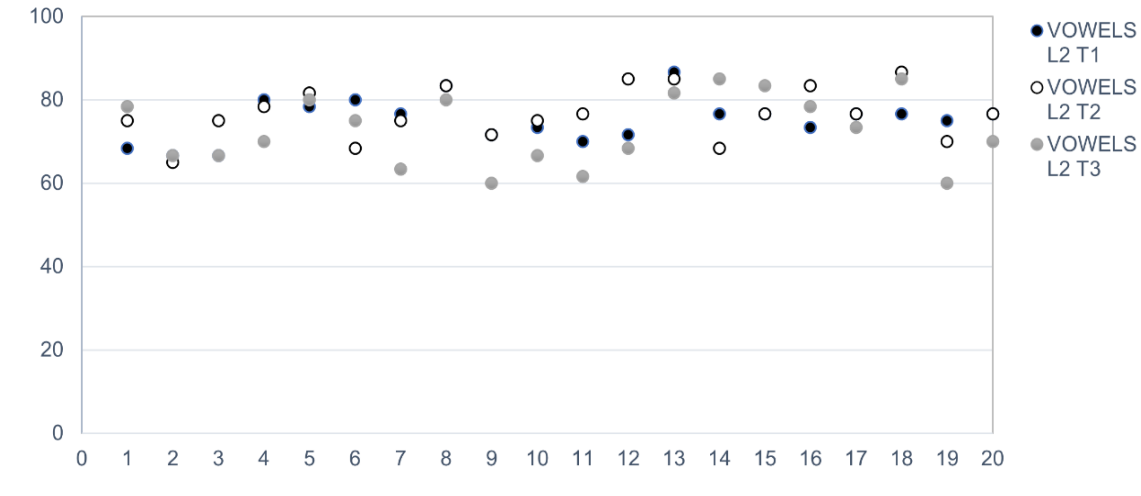


Figure 35. L2 vowel ProdAcc (%) in three testing session.

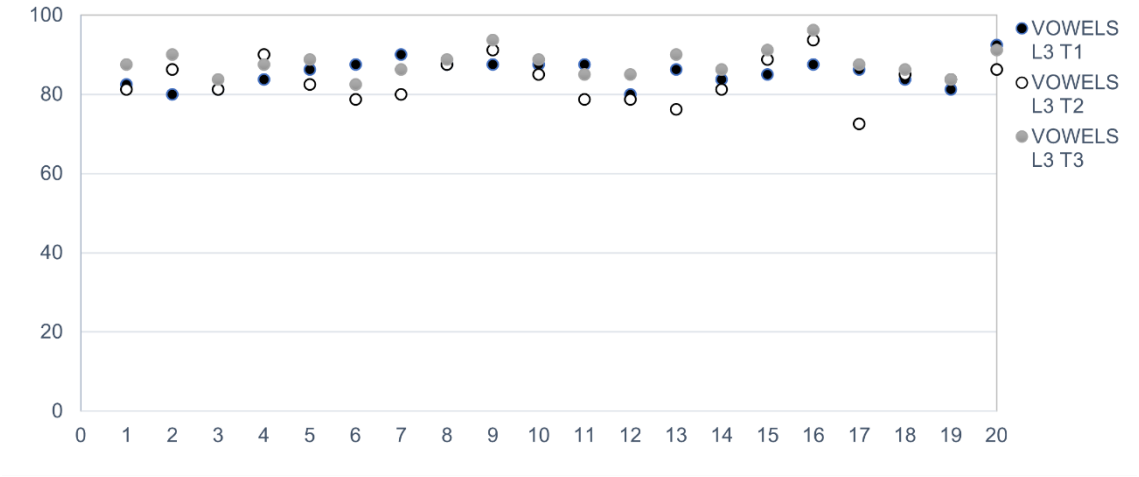


Figure 36. L3 vowel ProdAcc (%) in three testing session.

As for VOT, the ProdAcc scores (calculated with the application of the formula described above) were higher for L2 than L3. At T1, VOT ProdAcc in L2 reached the mean percentage of *M*=57 (SD=10.34), while for L3 it was *M*=46 (SD=9.73). In T2, the accuracy of VOT production increased considerably both for L2 (*M*=67, *SD*=12.83) and L3 (*M*=60, *SD*=12.18). In T2, the production in two languages noted a slight decrease; it was slightly more pronounced in the case of L2 (*M*=62, *SD*=11.39) compared to L3 (*M*=57, *SD*=11.50). Table 34 presents the accuracy scores calculated for VOT production in L2 and L3 across three testing sessions. Figure 37 and Figure 38 outline individual VOT production accuracy scores calculated for the participants in L2 and L3 across three testing sessions.

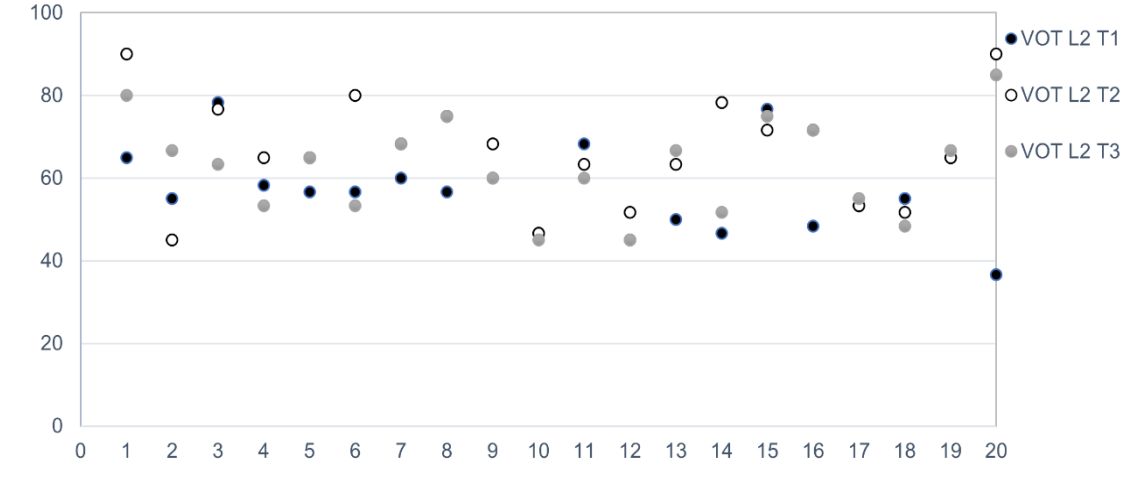


Table 34. VOT production scores for L2 and L3 over three testing sessions.

Lan-guage	Testing time	N	%	95% CI	SD	Median	Min	max
L2	T1	20	57	52.07-61.76	10.34	56.67	37	78
L2	T2	20	67	60.97-73.03	12.83	66.67	45	90
L2	T3	20	62	57.42-68.08	11.39	64.17	45	85
L3	T1	20	46	41.53-50.64	9.73	48.33	27	63
L3	T2	20	60	54.34-65.75	12.18	61.25	42	78
L3	T3	20	57	52.28-63.05	11.50	58.33	37	77

Figure 37. L2 VOT production accuracy (ProdAcc) (%) in three testing session.

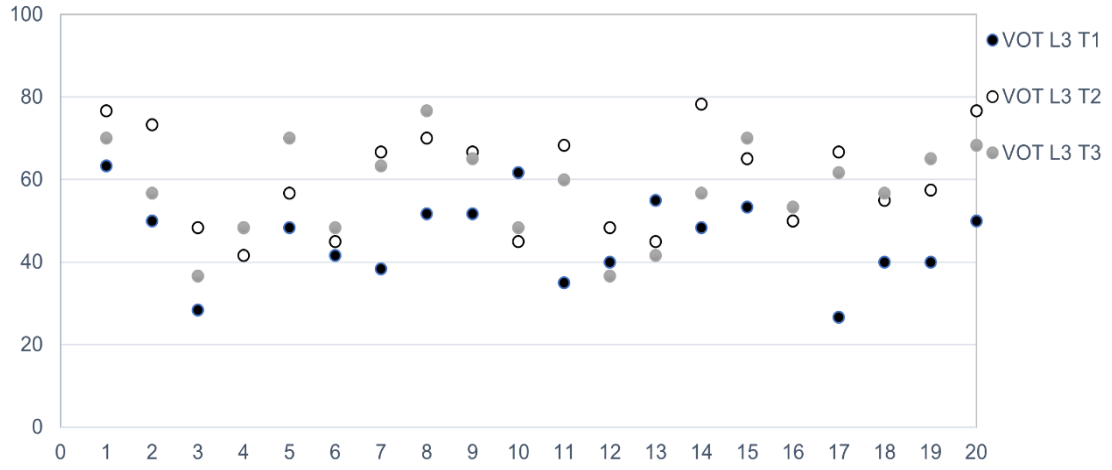


Figure 38. L3 VOT production accuracy (ProdAcc) (%) in three testing session.

The accuracy scores for vowel and VOT production were combined (through mean calculation) to produce overall speech production accuracy (OProdAcc) scores for respective languages in three testing sessions. The results obtained after combining the scores revealed quite a similar accuracy pattern for L2 and L3. At T1, the overall accuracy score (OProdAcc) for both languages was the same (*M*=66), however, L3 was characterised by a slightly wider spread of the scores (*SD*=5.38) compared to L2 (*SD*=5.03). At T2, the OProdAcc score for L2 was slightly larger than for L3 (*M =72, SD*=6.74; *M =71, SD*=6.64). At T3, the L2 OProdAcc score noted a decrease (*M =68, SD*=6.74) and was surpassed by L3 score (*M =73, SD*=6.46). Table 35 presents the overall production accuracy scores in L2 and L3 across three testing sessions. Figure 39 outlines the overall production accuracy scores in L2 and L3 at T1, T2 and T3. Figure 40 and Figure 41 present individual overall production accuracy (OProdAcc) scores in three testing sessions.

Table 35. Overall speech production accuracy scores (%) for L2 and L3 in three testing sessions.

Lan-guage	Testing time	N	OProdAcc	95% CI	SD	Median	min	max
L2	T1	20	66	63.73-68.44	5.03	67.08	57	77
L2	T2	20	72	68.68-74.99	6.74	72.50	55	83
L2	T3	20	68	64.25-71.16	7.38	66.25	56	79
L3	T1	20	66	63.23-68.23	5.38	66.04	55	75
L3	T2	20	71	68.63-74.85	6.64	71.25	61	82
L3	T3	20	73	69.81-75.86	6.46	73.85	60	83

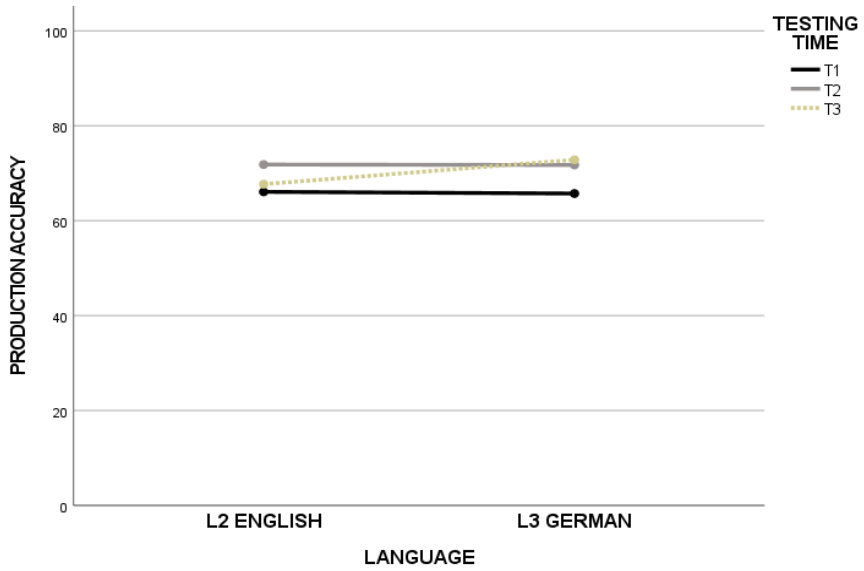


Figure 39. Overall production accuracy means (in percentages) for L2 and L3 in three testing sessions.

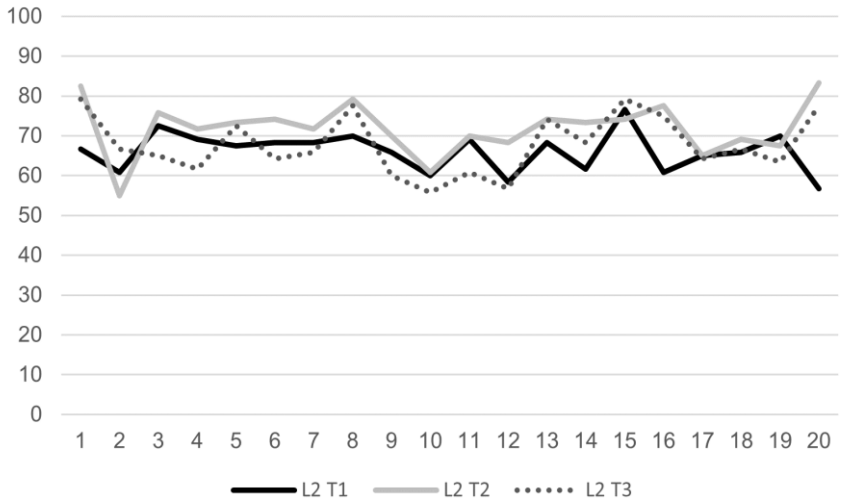


Figure 40. Individual overall production accuracy OProdAcc means for L2 in three testing sessions for 20 participants.

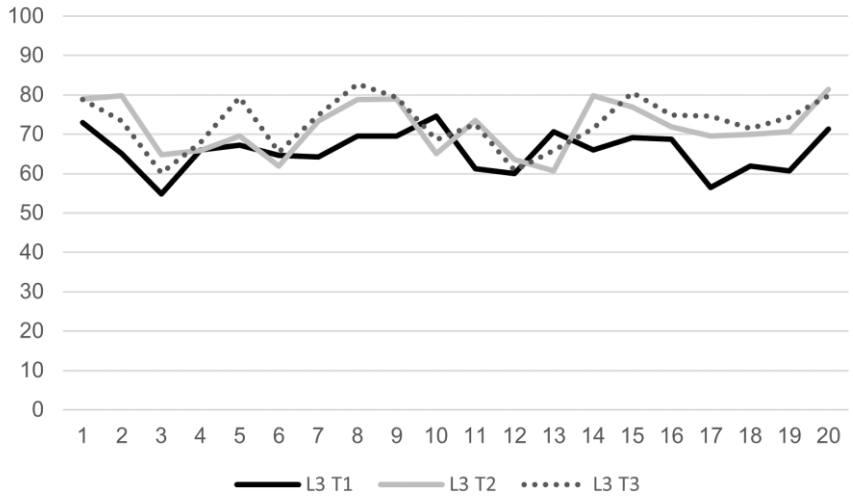


Figure 41. Individual overall production accuracy (OProdAcc) means for L3 in three testing sessions for 20 participants.

5.4.3. The relationship between the perception and production in a multilingual acquisition

To investigate the potential gains between the perception and production over three testing sessions, the overall accuracy scores for both modalities in L2 and L3 (in percentages) were compared.

In L2, the participants showed higher overall production accuracy (OProdAcc) scores compared to the perception (OPercAcc). T1, the difference between the OProdAcc (*M*=66%, *SD*=5.03) and OPercAcc (*M*=48%, *SD*=11.62) amounted to 18 percentage points. In T3, the difference between the OPercAcc (*M*=51, *SD*=9.04) and OProdAcc (*M*=72, *SD*=6.74) increased to 21 percentage points, whereas in T3 the difference between the OPercAcc (*M*=49, *SD*=11.27) and OProdAcc (*M*=68, *SD*=7.38) maintained a similar level of 19 percentage points. It can be concluded, that the difference between the overall accuracy of perception and production in L2 remained rather stable across three testing sessions. Continuing L2 exposure and growing language learning experience might have affected the increase in L2 overall production accuracy in T2, however, such trend did not continue into T3. The L2 perception over three testing sessions was rather stable and fairly unaffected by growing language exposure and experience. Figure 42

outlines the L2 perception (OPercAcc) and L2 production (OProdAcc) accuracy in three testing sessions.

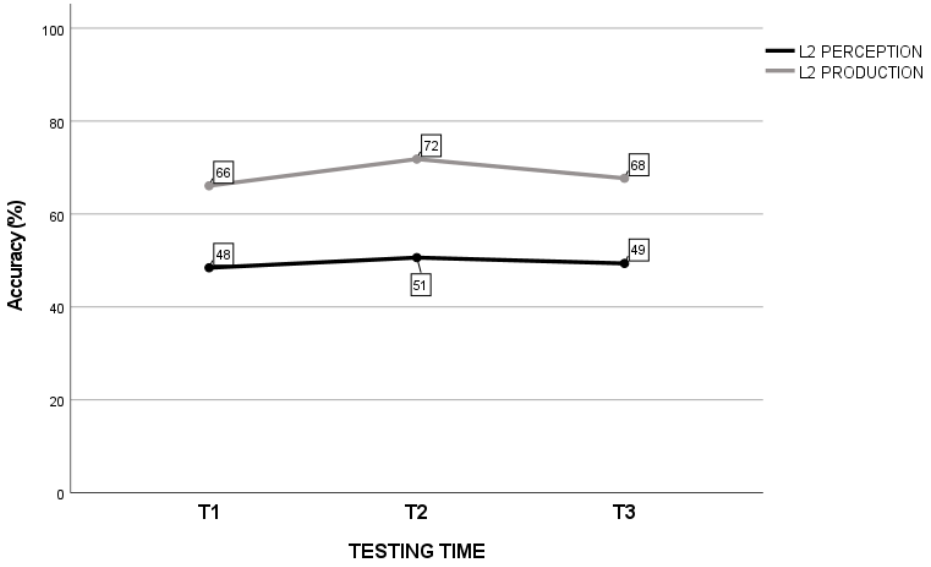


Figure 42. L2 perception (OPercAcc) and L2 production (OProdAcc) accuracy (%) in three testing sessions.

Similarly to L2, the participants showed higher accuracy in their L3 overall perception accuracy (OProdAcc) compared to L3 overall perception accuracy (OPercAcc), however, the differences were less pronounced. In T1, the difference between the OProdAcc (*M*=66%, *SD*=5.03) and OPercAcc (*M*=61, *SD*=11.62) amounted to 5 percentage points. In T3, the difference between the OProdAcc (*M*=72, *SD*=6.74) and OPercAcc (*M*=63, *SD*=9.04) increased to 9 percentage points. In T3, the difference

between the OProdAcc (*M*=73, *SD*=11.27) and OPercAcc (*M*=65, *SD*=7.38) maintained a similar level of 8 percentage points. Consequently, the difference between the accuracy of perception and production in L3 increased slightly from T1 to T2 and remained rather stable from T2 to T3. L3 exposure and growing language learning experience might have had an effect on the slight increase in L3 perception and production accuracy over three testing sessions. Figure 43 presents L3 perception (OPercAcc) and L2 production (OProdAcc) accuracy in three testing sessions.

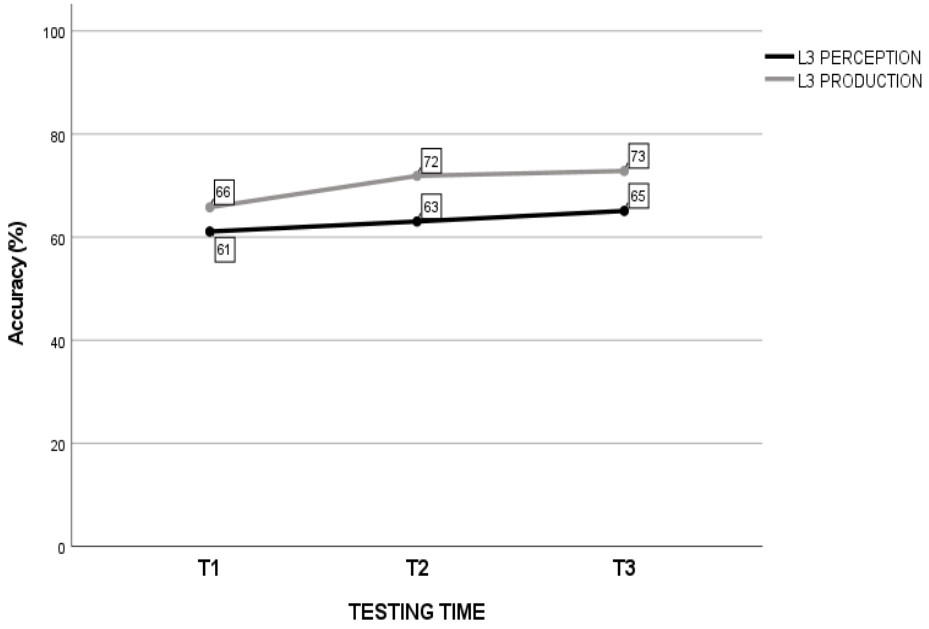


Figure 43. L3 perception (OPercAcc) and L2 production (OProdAcc) accuracy (%) in three testing sessions.

In conclusion, the differences between the overall L2 perception (OPercAcc)_and overall production accuracy (OPercAcc) scores in L2 were more visible at T2, where the production surpassed perception by an average of 19.33 percentage points. In L3 the average difference between the higher OProdAcc and OPercAcc was 7.33 percentage points. The overall accuracy in L2 and L3 perception and production maintained comparable levels (with a slight increase in L3 accuracy in T3). The overall L3 perception, however, surpassed overall L2 perception by an average of 13.66 percentage points across the three testing sessions.

A Pearson correlation coefficient was calculated to assess the strength and the directionality of the relationship between the OPercAcc and OPercAcc in three testing sessions in L2 and L3 respectively. In L2, the correlations were run between the L2 OPercAcc and overall L2 OProdAcc scores in three testing sessions. In T1, there was no correlation between the L2 OPercAcc and L2 OProdAcc scores (r= -.18, p=.43). A similar tendency was observed for T2 (r=.08, p=.71) and T3 (r=.07, p=.75), where no correlation was observed as well. In L3, the correlations did not reveal any significant relationships between the L3 OPercAcc and L3 OProdAcc. In T1, no correlation was observed between the L3 OPercAcc and L3 OProdAcc scores r= -.007, p=.97. A corresponding lack of

relationship was observed for T2 (r=.10, p=.67) and T3 (r=.39, p=.09). It might be therefore concluded that the strength and directionality of the relationships between the perception and production accuracy across three testing sessions were not robust enough to yield significant results. Table 36 summarises the findings concerning L2 and L3 perception and production accuracy.

Table 36. The relationships between the L2/L3 OPercAcc and OProdAcc in three testing sessions.

Language	Overall findings	Directionality of the relationship
L2	OProdAcc preceding OPercAcc	No correlation between the OProdAcc and OPercAcc
L3	OProdAcc preceding OPercAcc	No correlation between the OProdAcc and OPercAcc
L2-L3	L2 OPercAcc lower than L3 OPercAcc L2 ProdAcc comparable to L3 ProdAcc	-

5.5. Results of individual differences tasks

Individual differences (ID) tasks constitute an umbrella term for the measures assessing the additional factors which may be a source of individual variability in participants' multilingual perception and production performance. The group of tasks includes measures of phonological working memory, inhibitory control, and psychotypology. Moreover, the section includes the rating and open-ended questions from the language history and use questionnaire. Additionally, the final section includes the results of language proficiency ratings in L2 and L3. The results obtained in the tasks are presented in a longitudinal perspective, including scores obtained by the participants at T1 and T3.

5.5.1. Results of the phonological working memory task

The recordings obtained from the participants in pseudoword repetition task (PWR task) at T1 (previously presented applying a less detailed analysis in Krzysik and Wrembel

2019) and T3 were analysed according to the steps outlined in section 4.4.1. of Chapter 4. The 29 pseudowords abiding by the rules of Polish phonotactics were divided into two groups according to their length-related difficulty. Group 1 ranged from 1 to 3 syllables, whereas group 2 included words between 4 and 6 syllables. A target-like realisation in the mentioned groups was awarded 1 and 2 points respectively. A partially target-like realisation of a pseudoword in group 2, involving a mistake in one syllable of the word, was awarded 1 point. The maximum number of points possible to score in the task totalled 43. The target-likeness of production in the PWR task was rated by a native speaker of Polish. The points awarded in the process of rating are referred to as phonological working memory score (PWM score).

The results obtained through the rating indicate a slight improvement in PWM scores between the T1 testing session (*M*=30.55, *SD*=5.93) and the T3 testing session (*M*=31.80, *SD*=5.12). The PWM scores in T1 and T3 testing sessions are presented in Table 37. As far as the individual results are concerned, 15 participants showed an increase in their PWM scores between T1 and T3, with the sharpest increase by 6 points. The scores of 5 participants decreased between the testing sessions, with the largest decrease by 6 points. The PWM scores at T1 and T3 are presented in Table 37. The individual scores obtained by the participants in T1 and T3 are presented in Figure 44 .

Table 37. PWM scores in PWR task at T1 and T3

PWM score	N	M	95% CI	SD	median	Min	max
PWM T1	20	30.55	27.77-33.33	5.93	31.50	20	42
PWM T2	20	31.80	29.40-34.20	5.12	32.00	22	43

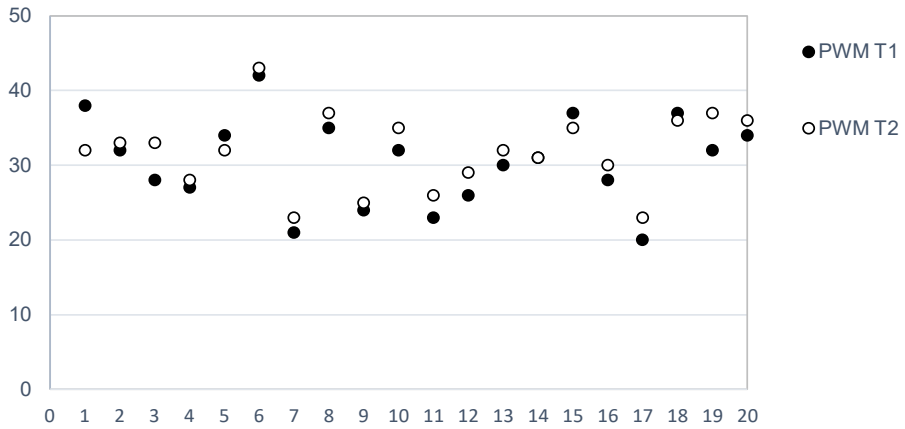


Figure 44. PWM scores at T1 and T3 for individual participants (maximal number of points to score - 43)

The PWM scores from T1 and T3 testing sessions were compared using repeated-measures ANOVA, which revealed a significant difference $F(1, 19)=4.79$, $p = .041$, $\eta p2=.202$. Based on the ANOVA result, it might be inferred that the overall improvement in the PWM scores between T1 and T3 was significant. The mean PWM scores compared over time are presented in Figure 45.

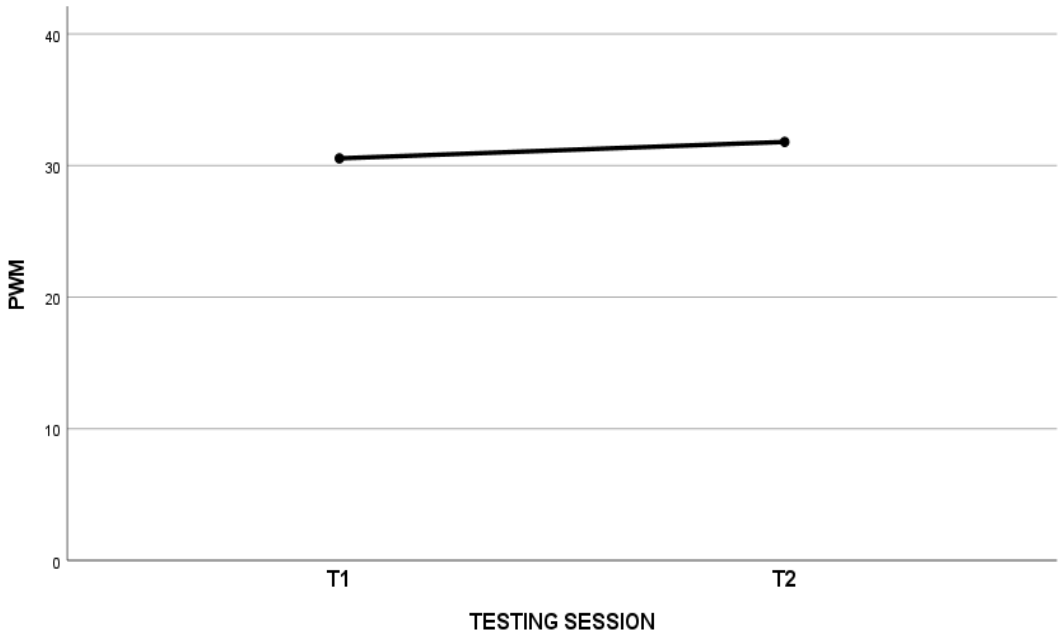


Figure 45. Mean PWM scores over two testing sessions – T1 and T3.

5.5.2. Results of the inhibitory control task

The data gathered through the flanker task were analysed following the procedures applied by Poarch and Bialystok (2015); such a code was also followed in the earlier publication by Krzysik (2020), which featured a T1 portion of the current data. First, the mean reaction time scores and mean accuracy rates[4] were calculated for each flanker task condition (congruent, incongruent, neutral and baseline). However, unlike in the previous studies, the participants were not excluded on the basis of the z-score. Normal distribution

[4] Accuracy in the context of the flanker task is understood as the ability to identify the direction in which the middle arrow was pointing in the experimental display. For details and the visualisation of the task display see section 4.7.2.

was adopted as a criterion, which was met without excluding potential z-score outliers. Further, flanker score, indicating one's degree of inhibitory control, was calculated. To obtain the score, the reaction times from the accurate congruent trials were subtracted from the reaction times from the accurate incongruent trials. Low flanker scores indicate that both the congruent and incongruent trials had a similar cognitive burden on the participant's processing. Consequently, such scores may imply a greater inhibitory control. High flanker scores indicated that incongruent trials required more cognitive resources than congruent ones, indicating a weaker inhibitory control.

The results obtained in T1 for 20 participants present a rather broad range, indicated by the standard deviation (SD=25.56) and minimal (-3) and maximal (89) values. The mean flanker score for T1 was M=53.90, which suggested that the scores for incongruent trials were more challenging that those in the congruent trials. In T3, the flanker score increased along with the variability of the scores (M=59.90, SD=45.33), which were distributed between the frames of the minimal score of -69 and the maximum score of 150. Table 38 presents the flankers scores at T1 and T3.

Table 38. Flanker scores at T1 and T3.

Flanker score	N	M	95% CI	SD	median	Min	max
T1	20	53.90	41.935-65.964	25.56	57.50	-3	89
T2	20	59.00	37.782-81.218	45.33	58.00	-69	150

Flanker scores from T1 and T3 testing sessions were compared using repeated-measures ANOVA; the result of the analysis indicated that there was no significant difference between the flanker scores at both testing times $F(1, 19)=.192$, $p=.666$, $\eta p2=.010$. The results of the repeated-measures ANOVA indicates that the flanker scores obtained by the participants in two testing sessions did not differ; the scores did not improve nor deteriorate significantly over time. Figure 46 presents the flanker score means at T1 and T3.

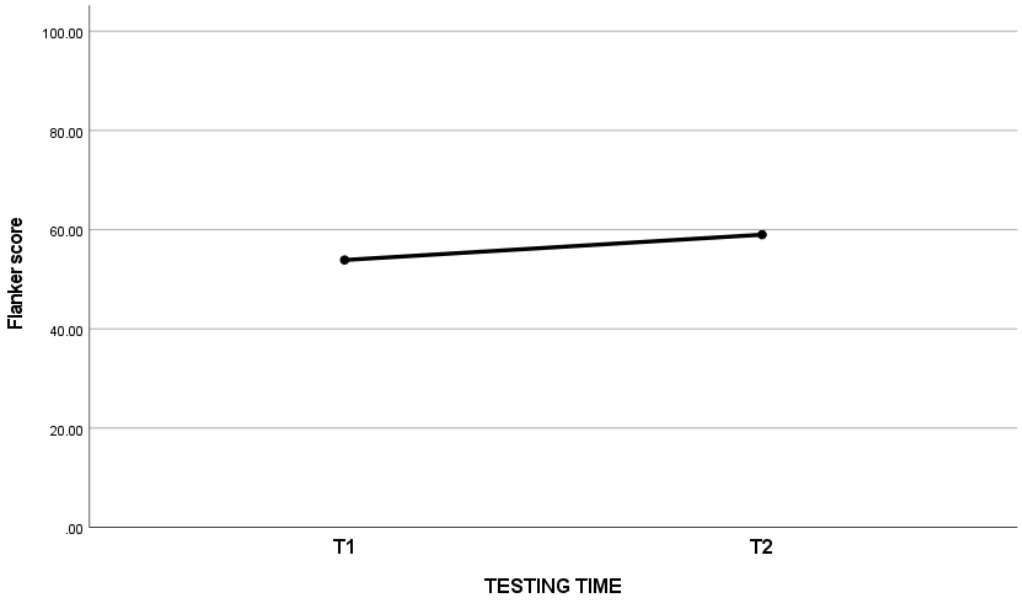


Figure 46. Flanker score means in T1 and T3.

5.5.3. Results of the psychotypology task

The data collected through the VILDIM measure were analysed according to the steps outlined by Nelson et al. (2021) and presented in section 4.7.4. The distances between the marked language circles were measured in centimetres. Subsequently, these raw measurements were transferred to an Excel sheet and transformed into normalised values to eliminate the effect of the student's interpretation of the size of the paper sheet. The normalised measurements obtained in the T1 testing session indicated that the participants perceived the L2 English vs.L3 German pair as the closest languages in their repertoire (*M*=0.28, *SD*=0.09), followed by L1 Polish vs. L2 English (*M*=0.34, *SD*=0.08) and the most distant L1 Polish vs. L3 German (*M*=0.38, *SD*=0.08). The results from the T3 testing session corresponded with the ones from T1; the L2 English vs.L3 German pair was perceived as the closest (*M*=0.26, *SD*=0.07), followed again by L1 Polish vs. L2 English (*M*=0.35, *SD*=0.09) and L1 Polish vs. L3 German (*M*=0.39, *SD*=0.06). The normalised measurements for T1 and T3 testing sessions are presented in Table 39.

Table 39. The normalised measurements of language distance at T1 and T3 (LGD/TT).

LGD/TT	N	M	95% CI	SD	median	min	max
L1-L2/T1	20	0.34	0.306-0.381	0.08	0.35	0.20	0.50
L1-L3/T1	20	0.38	0.339-0.418	0.08	0.38	0.21	0.49
L2-L3/T1	20	0.28	0.234-0.320	0.09	0.26	0.15	0.50
L1-L2/T3	20	0.35	0.303-0.391	0.09	0.33	0.18	0.52
L1-L3/T3	20	0.39	0.362-0.424	0.06	0.39	0.28	0.50
L2-L3/T3	20	0.26	0.225-0.292	0.07	0.28	0.11	0.37

As for PL-ENG language pair, an increase in the perceived distance was observed for 14 participants, whereas a decrease was noted for 6 participants. The most substantial increase was observed in the case of participant 19 (by 39%), while participant 8 noted the largest decrease in the perceived distance (37%). Figure 47 outlines the individual PL-ENG perceived distances (normalised measurements) at T1 and T3.

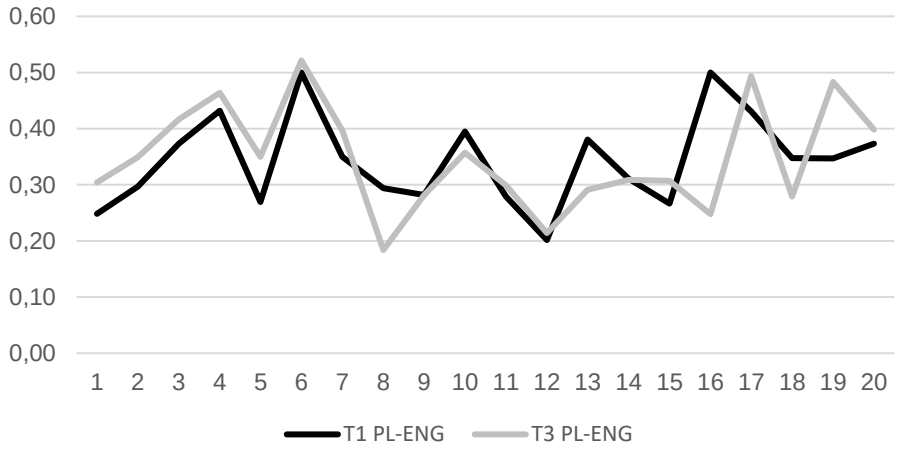


Figure 47. The normalised measurements of PL-ENG language distance at T1 and T3 for individual participants.

In the case of PL-GER language pair, an increase in the perceived distance was observed for 10 participants, whereas a decrease was noted for 10 participants. The most substantial increase was observed in the case of participant 8 (by 139%), closely followed by participant 16 (by 122%). Participant 19 noted the largest decrease in the perceived distance (by 40%). Figure 48 outlines the individual PL-GER perceived distances (normalised measurements) at T1 and T3.

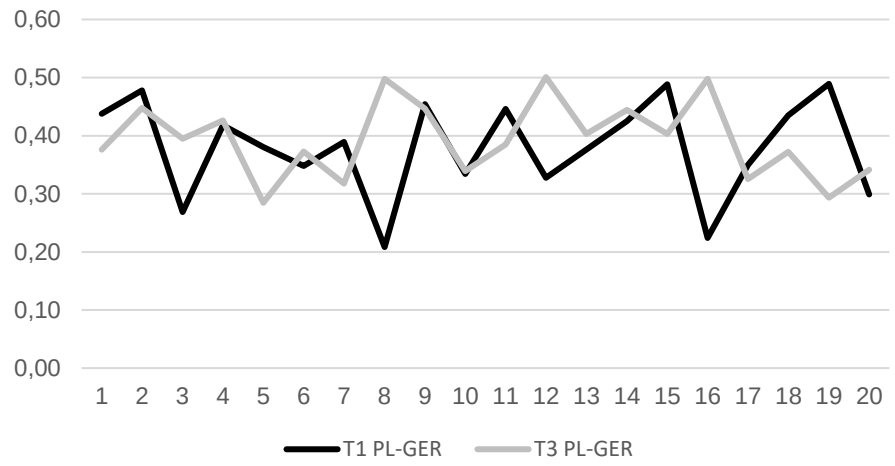


Figure 48. The normalised individual measurements of PL-GER language distance at T1 and T3 for individual participants.

In the case of the ENG-GER language pair, similarly to the previous pair, an increase and a decrease in the perceived distance was observed by 10 participants respectively. The most substantial increase was observed in the case of participant 18 (by 60%). Participant 3 noted the largest decrease in the perceived distance (by 47%). Figure 49 outlines the individual ENG-GER perceived distances (normalised measurements) at T1 and T3.

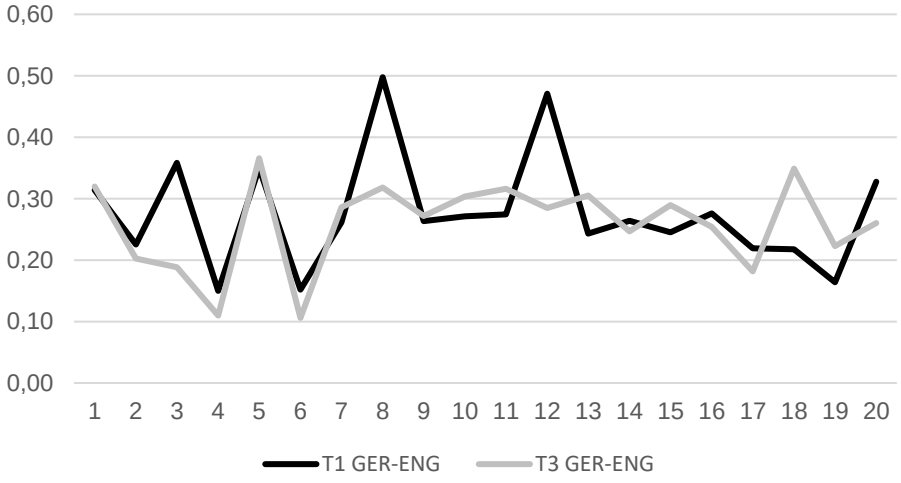


Figure 49. The normalised individual measurements of GER-ENG language distance at T1 and T3 for individual participants.

The relationship between normalised measurements over time was further analysed employing repeated-measures ANOVA. For the condition "language pair distances" and "language pair distances*time", sphericity was violated as indicated by Mauchly's tests, $X^2(2)=6.18$, $p=.039$ and $X^2(2)=6.18$, $p=.045$. Therefore, the Huyn-Feldt correction was applied. Repeated measures ANOVA revealed a highly significant effect of language distance, $F(1.64, 31.25)=11.06$, $p<.001$, $\eta_p^2=.37$. Such a result indicates that the perception of the distances between the languages of the participants was significantly different. However, there was no significant interaction of language distance and time, $F(1.66, 31.65)=.39$, $p=.64$, $\eta_p^2=.02$. Consequently, the perception of distances between each language pair over the testing sessions did not change significantly. The post-hoc test using the Bonferroni correction revealed that the perceived distances between L1 Polish vs. L3 German and L2 English vs. L3 German language pairs were statistically significant in both T1 ($p=0.03$) and T2 ($p=0.001$) testing sessions. Figure 50 shows the perceived mean distances between the language pairs at T1 and T2.

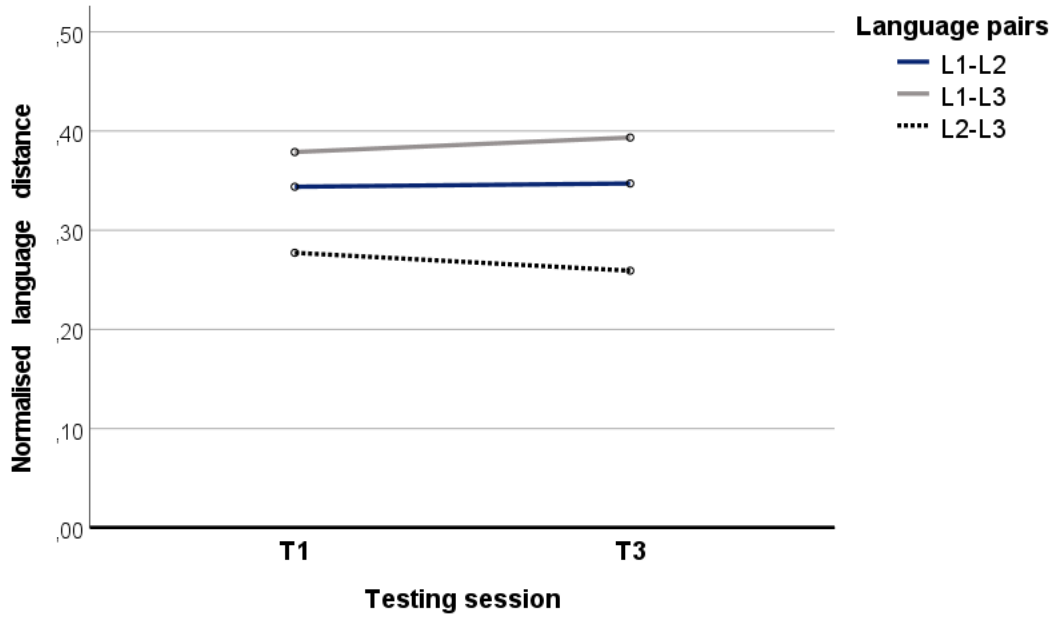


Figure 50. The perceived mean distances between the language pairs (normalised) at T1 and T3.

5.5.4. Language history and use – extended results

The data collected in the language history and use questionnaires was used to characterise the tested group and establish exclusion criteria. A small portion of the data was already included in Chapter 4 (see section 4.2) to outline the general profiles of the participants. The extended results presented in the subsections below offer a more in-depth look into the quantitative and qualitative data gathered in the questionnaires in testing sessions T1 and T3. Section 5.4.4.1 presents the results of the language proficiency self-assessment and the participants' estimations of their frequency of language use. Section 5.4.4.3 features the results of qualitative questions on attitudes towards acquired languages, the perceived role of L1 and L2 in the acquisition of L3 and language mixing.

5.5.4.1. Self-assessment of language proficiency

The first rating question inquired about the self-assessment of language proficiency. Language proficiency was represented with a 5-point scale; 1 implied a very low self-assessment of proficiency and 5 stood for a very high self-assessment of proficiency. The categories in which participants rated their language proficiency included reading, writing, speaking and listening.

As for L2 English, the participants were moderately confident of the language proficiency, with means for the categories ranging between 3.55 and 4.05. Reading and writing were assessed better at T3, whereas the remaining components received higher scores at T1. Table 40 presents the proficiency rating data for L2 at T1 and T3.

Table 40. Self-assessment of language proficiency in L2.

category	N	M	SD	min	max	mode
L2 T1 reading	20	3.80	0.81	2	5	4
L2 T3 reading	20	4.05	0.77	2	5	5
L2 T1 writing	20	3.70	0.90	1	5	4
L2 T3 writing	20	3.55	0.92	2	5	4
L2 T1 speaking	20	4.05	0.74	2	5	4
L2 T3 speaking	20	3.90	0.77	2	5	4

L2 T1 listening	20	3.70	0.95	2	5	3
L2 T3 listening	20	3.75	0.99	2	5	4

The self-assessment of L3 German of the participants resulted in lower scores than those in L2, with a mean between 2.30 and 2.75. The variability between the participants was rather low, as indicated by a low standard deviation. The proficiency in all categories, except for reading, was rated higher at T3, however, the difference between the scores was not substantial. Table 41 presents the proficiency rating data for L3 at T1 and T3.

Table 41. Self-assessment of language proficiency in L3.

	N	M	SD	min	max	mode
L3 T1 reading	20	2.70	0.71	2	4	2
L3 T3 reading	20	2.65	0.79	1	4	2
L3 T1 writing	20	2.30	0.71	1	3	3
L3 T3 writing	20	2.55	1.02	1	5	2
L3 T1 speaking	20	2.50	0.97	1	4	3
L3 T3 speaking	20	2.75	1.22	1	5	3
L3 T1 listening	20	2.25	1.09	1	4	1
L3 T3 listening	20	2.35	1.11	1	5	2

Due to the particular importance of speaking and listening skills to the researched topic of phonological development, the tables with comparisons of self-assessment scores across the languages in these categories are highlighted below. In relation to speaking, the participants were quite consistent in their ratings across T1 and T3 testing sessions; with L2 rated higher than L3. A similar tendency was visible in the self-assessment of listening. Overall, the participants were more confident about their speaking than listening skills, however, the difference did not exceed the average of 0.25 percentage points for L2 and 0.32 percentage points for L3. Figure 51 represents the self-assessment of listening across three languages in both testing sessions, whereas Figure 52 presents the self-assessment of speaking.

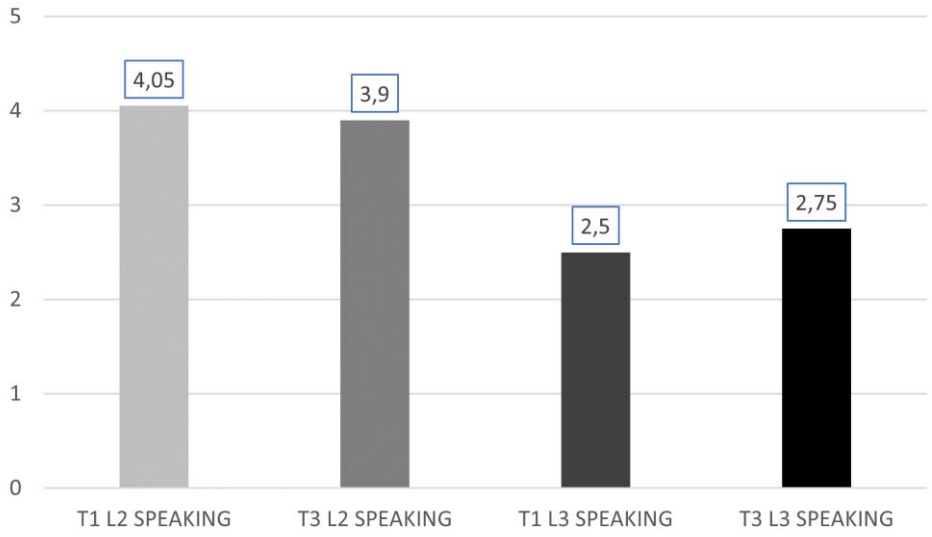


Figure 51. Self-assessment of speaking in L2 and L3 at T1 and T3.

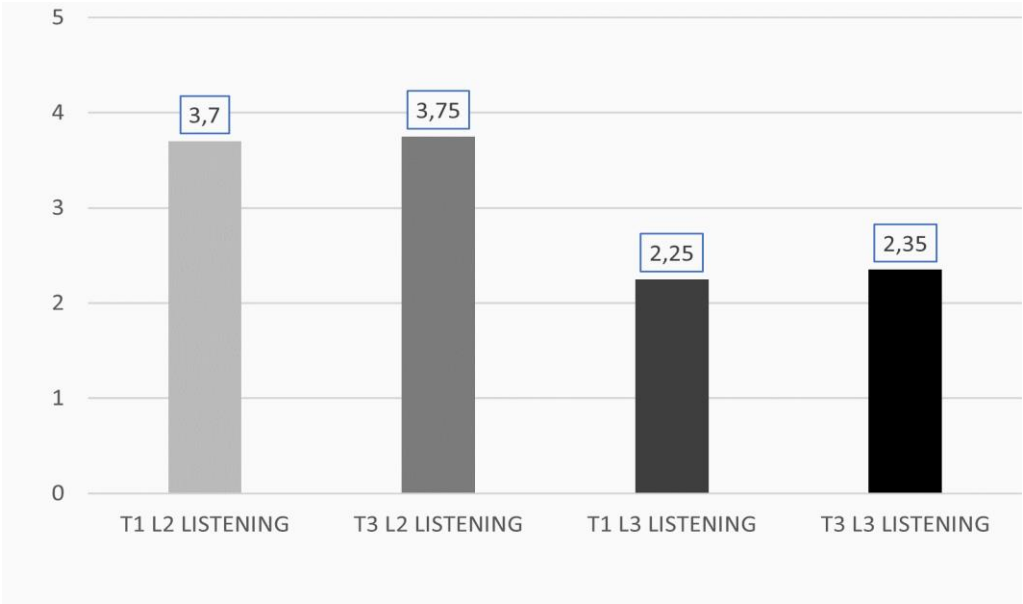


Figure 52. Self-assessment of listening in L2 and L3 at T1 and T3.

5.5.4.2. Self-assessment of language use and exposure

The questionnaire included two questions inquiring about the frequency of language use and exposure. The first question was concerned with the said frequency in different contexts, whereas the second question focused on the medium of use and exposure. Both

questions utilised percentages as a way of illustrating the time devoted to each activity in a given language.

The question about the frequency of language use across different contexts included such settings as family, friends and school. Due to the assumed dominant role of L1 in the lives of the participants, their native language was also included in the comparison. The participants were requested to assess (in percentages) the time spent in their L1, L2 and L3 in the three above mentioned contexts. The time for each context was represented as 100%, therefore, the participants could not exceed this number in their estimates. In T1, the family set was dominated by L1 (*M*=92, *SD*=6.78), followed by L2 (*M*=6.25, *SD*=4.71) and L3 (*M*=1.75, *SD*=3.63). A similar tendency was visible for the use of languages with friends; L1 was the most frequently used language (*M*=94, *SD*=7.35), followed by L2 (*M*=4, *SD*=4.9) and L3 (*M*=2, *SD*=4). The school setting was the most diverse one; L1 was still the most prominent (*M*=73.5, *SD*=11.95), but what was visible was the presence of L2 (*M*=15, *SD*=7.42) and L3 use (*M*=11, *SD*=7.68). Figure 53 presents the described scores.

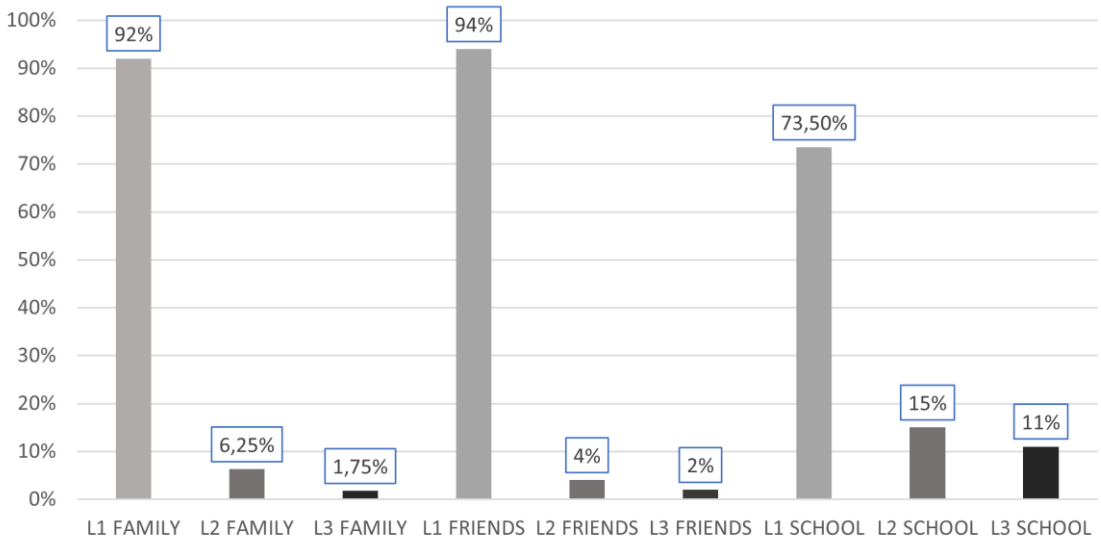


Figure 53. Self-assessment of language use and exposure in the family, friends and school settings at T1.

The patterns observed at T1 were also visible at T3. L1 (*M*=90, *SD*=11.61) dominated the family setting, followed by L2 (*M*=8, *SD*=9.27) and L3 (*M*=2, *SD*=4). In contacts with friends, L1 (*M*=92.5, *SD*=8.87) was the most prominent, whereas L2 (*M*=5.5, *SD*=5.89) and L3 (*M*=2.5, *SD*=4.33) were only slightly more prominent than at T1. The school

setting was again dominated by L1 (*M*=70, *SD*=11.83), with L2 (*M*=18.5, *SD*=7.92) and L3 (*M*=11, *SD*=5.39) with similar prominence to that at T1. Figure 54 presents the described scores.

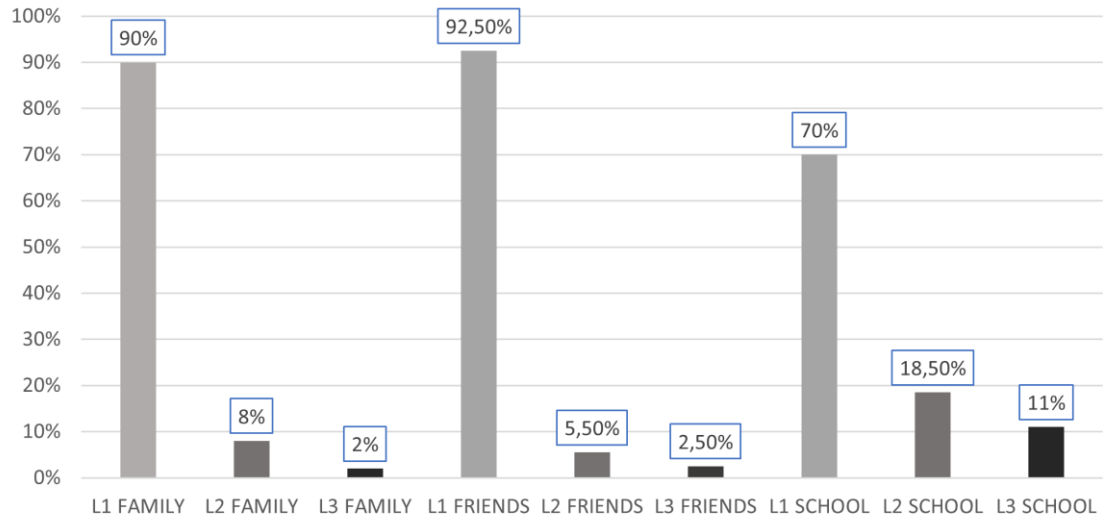


Figure 54. Self-assessment of language use and exposure in the family, friends and school settings at T3.

The self-assessment of the frequency of language use and exposure in the contexts of family, friends and school aligned with the characteristics of the participants' environment. The homogenous, L1-dominant setting was reflected in the high scores of L1. The formal context of L2 and L3 acquisition was also visible in the scores; their highest frequency was associated with the school setting.

The other frequency-related question inquired about the medium of language use and exposure. The participants were requested to assess the time spent in their L1, L2 and L3 on the listed activities in percentages. The time for each category was represented as 100% and the participants were instructed not to exceed this number in their estimates. The categories included (1) radio, TV, YouTube, (2) books, (3) computer games, (4) Internet browsing, (5) Internet communication.

At both T1 and T3, L1 was the language that dominated all the listed categories of use and exposure. On average, the participants estimated that they spent 73.5% of their time devoted to TV/radio/youtube using L1 at T1 and 75.4% in T3. At T1, books were read in L1 for 94% of the time, while in T3 for a comparable 98.5%. Computer games were a special category, where the dominance of L1 was substantially smaller. At T1, 62% of time devoted to gaming was spent using L1 and in T3 it was 64%. Internet

browsing was conducted mostly in L1 with 91.9% of the time at T1 and 90.3% in T3. 83% of the time spent communicating through the Internet was spent using L1, with a similar average of 82% in T3. Figure 55 demonstrates the self-assessment of time devoted to various activities using L1.

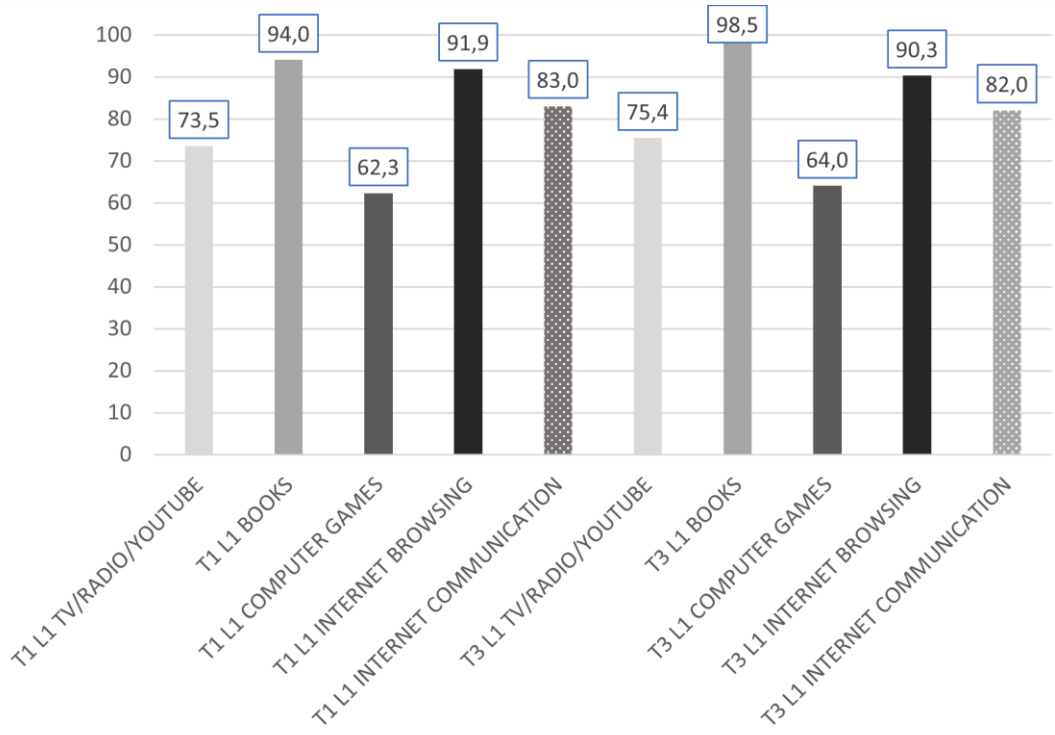


Figure 55. Self-assessment of time devoted to various activities using L1.

The use of L2 was substantially less frequent in the listed contexts when compared to L1. At T1, the participants spent 23.75% of their time devoted to TV/radio/youtube using L2 and 23.85 % in T3. At T1, they read books in L2 for 1.25% of the time, while in T3 for 1.5%. Computer games were quite frequently played in L2, with 38.75% of the time at T1 and 41% at T2. Internet browsing was performed in L2 for 7.1% of the time at T1 and 14.75% at T3. 15.75% of the time spent communicating through the Internet was spent using L1, with a comparable 16.25% at T3. Figure 56 demonstrates the self-assessment of time devoted to various activities using L2.

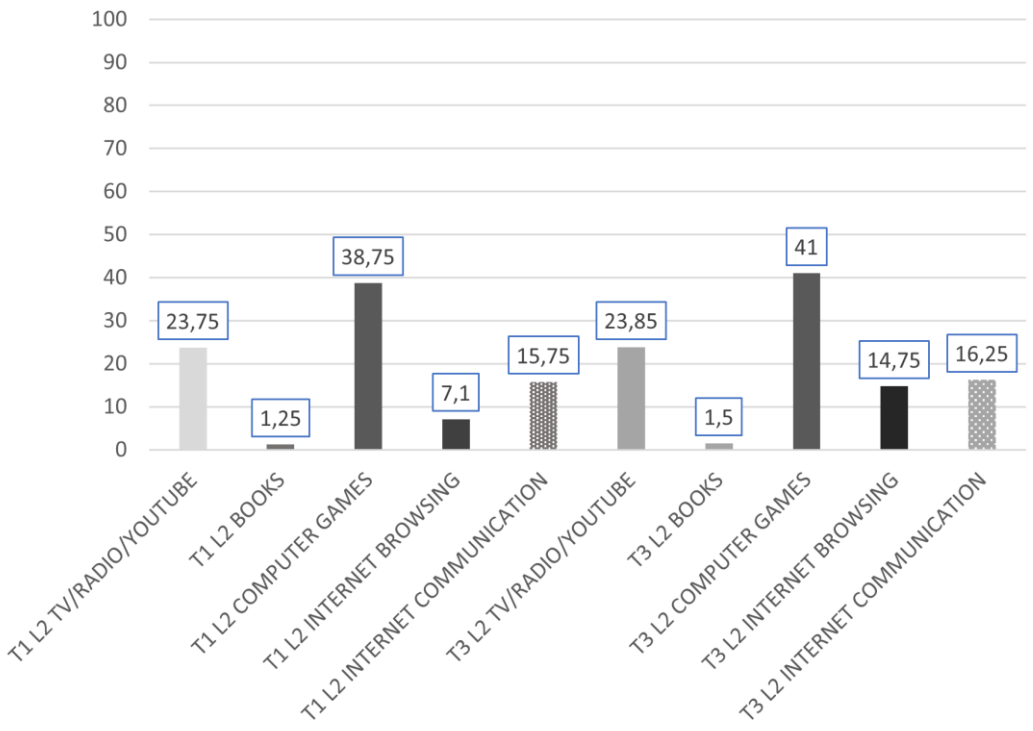


Figure 56. Self-assessment of time spent during various activities using L2.

L3 was characterised by the smallest frequency of use in the listed contexts. At T1, the participants spent 2% of their time devoted to TV/radio/youtube using L3 and 0.75 % in T3. At T1, they read books in L2 for 0.25% of the time, while in T3 for 0%. Computer games were not played in L3 both in T1 and T3. The Internet was browsed in L3 for 0.75% of the time in T1 and 0% in T3. Internet communication through L2 took place 2.25% of the time in T1 and 1.75% in T3.

5.5.5. Language learning attitudes, the role of L1 and L2 in L3 acquisition and language mixing

The language history and use questionnaire included a set of open-ended questions concerning the attitudes towards learning foreign languages, the influence of L1 and L2 on L3 learning, and language mixing. The qualitative data gathered in this part of the questionnaire were analysed for reoccurring themes and the answers were subsequently quantified. The results are presented in the paragraphs below. The number of quantified descriptive answers does not always align with the number of the participants; some

students managed to produce more than one answer and some did provide a descriptive justification of their choice in the questionnaire.

The question about the attitudes towards L2 and L3 inquired whether the participants enjoyed learning foreign languages. In T1, 80% of the participants (16 individuals) expressed their positive attitudes towards language learning, whereas 20% (4 individuals) disliked this activity. Among the participants who decided to justify their attitudes, the positive approach to learning languages was motivated by the enjoyment of learning new things (4 answers), the improved perspectives of work and travel (4 answers), one's interest and personal development (4 answers) and a possibility of befriending someone outside the country of origin (2 answers). The negative attitudes were associated with the difficulty of learning foreign languages (3 answers) and a general dislike of studying (1 answer). Some participants provided more than one justification for their attitude towards learning foreign languages, whereas some did not rationalise their choice (4 participants). In T3, 75% of the participants (15 individuals) expressed their positive attitude towards language learning, whereas 25% (5 individuals) disliked this activity. The participants who explained their positive stance pointed to the enjoyment of learning new things (4 answers), one's interest and personal development (3 answers), improved perspectives of education, work and travel (3 answers), improved perspectives of education, work and travel (2 answers), improved communication with people abroad (2 answers), general usefulness (2) and improved perspectives in general (2 answers). The negative attitudes towards language learning were justified by its inherent difficulty (3 answers), time-consuming nature of the process (1 answer) and a preference for the use of L1 (1 answer). Similarly to T1, some participants provided more than one justification for their attitude towards learning foreign languages, whereas some did not rationalise their choice (2 participants).

A follow-up question inquired whether the participants preferred learning their L2 or L3. In T1, 80% of the participants (16 individuals) pointed to their L2, whereas L3 was mentioned by 20% (4 individuals). Among the reasons justifying L2 as a preferred option, the participants listed the ease of learning (4 answers), longer learning experience and familiarity (3 answers), a general preference for L2 (2 answers), a preference for L2 pronunciation (2 answers), a perception of L2 as a more useful language (2 answers), better learning outcomes in L2 (1 answer) and an interest in L2 and its features (1 answer). The participants who preferred learning L3 admitted that they liked L3 pronunciation and

spelling (2 answers), enjoyed learning L3 because of its novelty (1 answer) and had a general preference for L3 (1 answer). In T3, 95% of the participants (19 individuals) admitted that they prefer learning their L2, whereas L3 was mentioned by 5% (1 individual). The choice of L2 was justified with the ease of learning (8 answers), enjoyment of learning (4 answers), longer learning experience and familiarity (3 answers) and usefulness of L2 (1 answer). The one participant who chose L3 pointed to greater enjoyment of learning.

At T1, the question "Which language is more helpful in learning your L3?" produced a split between the participants; 40% pointed to L1 and 45% to L2. Two participants admitted that none of the two languages helps to learn L2, whereas one individual mentioned that both languages are useful. The participants who decided that L2 is more useful in learning L3 explained their choice by pointing to the following factors: similar words in L2 and L3 (7 answers), a general similarity between L2 and L3 (2 answers), similar pronunciation (1 answer). As the justification for the choice of L1 as the primary aid in learning L3, the participants mentioned that L1 is the language of L3 instruction (2), L1 is used for translation (1 answer), L1 is so different from L3 that it prevents them from mixing (1 answer), and L1 and L3 have a similar level of spelling to pronunciation correspondence (1 answer). The participant who named both L1 and L2 as helpful in learning L3 explained their answer by pointing to similar words across these three languages. In T3, 65% recognised their L2 as more helpful in learning L3 while only 15% pointed to their L1. Three participants claimed that none of their languages helps to learn their L3 and one participant pointed to the beneficial role of both L1 and L3. The participants who selected L2 pointed to similar vocabulary in L2 and L3 (9 answers) and general similarity (3 answers). The participants who chose L1 mentioned its role as a language of instruction (1 answer) and a similar level of spelling to pronunciation correspondence (1 answer). The participant who considered both L1 and L2 as helpful in learning L3 pointed to similar words across these languages.

The question about language mixing rendered the following results in T1 testing sessions: 60% of the participants noticed that they mixed their languages in one conversation/utterance and 40% did not acknowledge such occurrences. The contexts provided by the participants who spotted such mixing included: language classes (5 answers), general communication (4 answers), communication while using the Internet (2 answers), and writing (1 answer). As for T3, 65% of the participants noticed language

mixing in one conversation/utterance and 35% did not observe it. The participants mixed their languages mostly during language classes (3 answers) while using the Internet (2 answers), in conversation (1 answer) while counting (1 answer), while playing computer games (1 answer), in stressful situations (1 answer), when looking for words similar across learned languages (1 answer).

Overall, the majority of the participants in both T1 and T3 expressed a positive attitude towards language learning, with a preference for learning their L2 - English. In T1, L1 Polish and L2 English were assessed as helpful in learning L3 German with almost an equal split of responses. In T2, more participants recognised the helpful role of L2 English. The awareness of language mixing was similar in both testing sessions, with almost two thirds of the participants recognising the presence of this phenomenon in their speaking.

5.5.6. Language proficiency task – ratings

The production data obtained through short interviews in L2 and L3 in three testing sessions were rated by three proficient and trained raters (per language) in a two-tier process (see section 4.7.5 for details concerning the raters' profiles and the rating). First, the participants who did not produce any verbal output were identified and awarded a score of 0 in all categories of the subsequent rating. The verbal output of the remaining participants was rated on a 10-point scale in 6 categories by three raters per language (rating categories are outlined in Chapter 4, section 4.7.5). The following subsections present the rating scores for the participants' production in L2 and L3 in three testing sessions.

5.5.6.1. Language proficiency ratings – L2

Three L2 raters (see 4.7.5 for the raters' profiles) demonstrated a high consistency in their ratings as indicated by Cronbach's alpha; all but two categories (Cronbach's alpha) exceeded α=.7. The low reliability of the category "comprehensibility" (α=.44) in the T3 testing session was corrected through the analysis of item-total statistics. As a result, rater

3 was excluded to improve reliability from this rating category at T3. Similarly, the low reliability of the category "influence from L3 pronunciation" in three testing sessions (T1 α=.49, T2 α=.54, T3 α=.14) was corrected through the analysis of item-total statistics, which indicated that the exclusion of rater 3 would increase the reliability. Consequently, the ratings of rater 3 for the previously mentioned category in three testing sessions were excluded. Table 42 presents the reliability values calculated for L2 ratings in 6 categories at three testing points.

Table 42. Rater consistency (Cronbach's alpha) in L2 proficiency rating categories (T1, T2, T3).

Rating category	Cronbach's alpha (α) T1	Cronbach's alpha (α) T2	Cronbach's alpha (α) T3
comprehensibility	.79	.86	.76
overall proficiency	.94	.92	.73
overall pronunciation	.86	.85	.70
switches to L1	.90	.96	.91
influence from L1 in pronunciation	.75	.79	.79
influence from L3 in pronunciation	.71	.83	.76

At T1, the highest-rated category was comprehensibility (*M*=6.91, *SD*=1.31), with the highest rating of all testing sessions. Overall fluency in L2 was rated as rather moderate, which, to a certain extent, may correspond with the expectations for L2 acquisition in progress (*M*=4.98, *SD*=2.37); overall pronunciation was rated in a similar way (*M*=4.80, *SD*=1.53). As far as the CLI-related categories are concerned, the switches to L1 occurred sporadically (*M*=4.01, *SD*=2.51). The influence from L1 (*M*=4.35, *SD*=1.34) in pronunciation was more visible than the influence from L3 (*M*=2.26, *SD*=.90), however, both can be described as not particularly frequent. Table 43 summarises the proficiency ratings for L2 speech recordings obtained at T1.

Table 43. Proficiency ratings by categories for L2 at T1 (10-point scale).

Rating category	M	95% CI	SD	median	Min	Max
comprehensibility	6.91	6.303-7.529	1.31	7.17	4.67	8.67
overall fluency	4.98	3.873-6.093	2.37	5.67	1.33	8.33
overall pronunciation	4.80	4.079-5.520	1.53	4.66	2.33	8.33
switches to L1	4.01	2.743-5.289	2.71	3.50	1.00	9.67
influence from L1 in pronunciation	4.35	3.718-4.981	1.34	4.33	1.67	6.33

influence from L3 in pronunciation	2.26	1.841-2.692	0.90	2.16	1.00	4.00

At T2, comprehensibility was again the highest-rated category (*M*=6.25, *SD*=3.05), however, it marked a slight decrease in comparison to T1. The category of overall proficiency in L2 noted an increase in rating (*M*=5.38, *SD*=2.61), which to a certain extent reflects the increasing proficiency throughout the process of learning. Overall pronunciation was rated similarly to the T1 rating (*M*=4.88, *SD*=1.67). The switches to L1 were noticeably less frequent (*M*=2.31, *SD*=1.50) than at T1. The influence from L1 (*M*=3.26, *SD*=1.99) in L2 pronunciation was considerably less prevalent compared to T1. The influence from L3 retained similar levels when compared to T1 (*M*=2.21, *SD*=1.28). Table 44 summarises the proficiency ratings for L2 speech recordings obtained at T2.

Table 44. Proficiency ratings by categories for L2 at T2 (10-point scale).

Rating category	M	95% CI	SD	Median	Min	Max
comprehensibility	6.25	5.057-7.442	2.54	7.16	0.00	9.33
overall fluency	5.38	4.160-6.605	2.61	5.83	0.00	8.67
overall pronunciation	4.88	3.883-3.883	2.13	5.33	0.00	9.00
switches to L1	2.31	1.317-3.315	2.13	1.50	0.00	8.33
influence from L1 in pronunciation	3.03	2.251-3.815	1.67	3.33	0.00	6.00
influence from L3 in pronunciation	2.21	1.617-2.816	1.28	1.83	0.00	4.33

At T3, comprehensibility noted a decline (*M*=5.66, *SD*=2.65), with the lowest rating of all sessions. Similarly, overall fluency in L2 was rated lower compared to T1 and T2 (*M*=4.61, *SD*=1.65). Again, the same tendency was visible for overall pronunciation, which noted a decrease in ratings (*M*=4.62, *SD*=1.67) in comparison to T1 and T3. The switches to L1 slightly increased in frequency (*M*=2.55, *SD*=1.79) compared to T2. The influence from L1 (*M*=3.00, *SD*=1.26) in pronunciation was once again more visible than the influence from L3 (*M*=2.21, *SD*=0.96), however, both can be described as rather infrequent. Table 45 summarises the proficiency ratings for L2 speech recordings obtained at T3.

Table 45. Proficiency ratings by categories for L2 at T3 (10-point scale).

Rating category	M	95% CI	SD	median	min	Max
comprehensibility	5.66	4.819-6.513	1.80	6.00	0.00	8.67
overall fluency	4.61	3.840-5.393	1.65	4.33	0.00	7.00
overall pronunciation	4.62	3.830-5.402	1.67	4.33	0.00	7.67
switches to L1	2.55	1.710-3.389	1.79	2.16	0.00	7.00
influence from L1 in pronunciation	3.00	2.409-3.590	1.26	3.00	0.00	4.67
influence from L3 in pronunciation	2.21	1.766-2.667	0.96	2.55	0.00	3.67

5.5.6.2. Language proficiency ratings - L3

Three L3 raters (see 4.7.5 for the raters' profiles) demonstrated a high consistency in their ratings in three testing sessions, as indicated by the reliability values (Cronbach's alpha) computed for all 6 categories, exceeding α=.7 in all cases across the testing session. Table 46 presents the reliability values calculated for ratings in 6 categories at three testing points.

Table 46. Rater consistency (Cronbach's alpha) in L3 proficiency rating categories (T1, T2, T3).

Rating category	Cronbach's alpha (α) T1	Cronbach's alpha (α) T2	Cronbach's alpha (α) T3
comprehensibility	.91	.94	.92
overall proficiency	.89	.92	.93
overall pronunciation	.87	.91	.90
switches to L1	.91	.95	.97
influence from L1 in pronunciation	.78	.88	.86
influence from L3 in pronunciation	.89	.92	.87

At T1, the highest-rated category was comprehensibility (*M*=4.73, *SD*=3.09). Overall proficiency in L3 was rated rather low, which remains in line with the expectations for the early stage of language learning (*M*=3.58, *SD*=2.64); overall pronunciation was rated in a similar way (*M*=3.78, *SD*=2.33). As far as the CLI-related categories are concerned, the switches to L1 were infrequent (*M*=2.86, *SD*=2.51). The influence from L1 (*M*=3.26, *SD*=1.94) in pronunciation was perceived as stronger than

the influence from L2 (*M*=2.28, *SD*=2.84), however, both can be described as not particularly frequent. Table 47 summarises the proficiency ratings for L3 speech recordings obtained at T1.

Table 47. Proficiency ratings by categories for L3 at T1 (10-point scale).

Rating category	M	95% CI	SD	median	min	Max
comprehensibility	4.73	3.287-6.179	3.09	4.83	0.00	9.00
overall proficiency	3.56	2.329-4.803	2.64	3.00	0.00	7.00
overall pronunciation	3.78	2.691-4.875	2.33	4.33	0.00	6.33
switches to L1	2.86	1.692-4.041	2.51	2.16	0.00	8.33
influence from L1 in pronunciation	3.26	2.357-4.175	1.94	3.66	0.00	5.67
influence from L2 in pronunciation	2.28	1.419-3.147	1.84	2.16	0.00	6.33

At T2, comprehensibility was again the highest-rated category (*M*=5.43, *SD*=3.05), marking an increase in comparison to T1. Similarly, the category of perceived overall proficiency in L3 noted an increase in rating (*M*=4.31, *SD*=2.69), which to a certain extent reflects the further stage in L3 instruction. Overall pronunciation received ratings comparable to T1 (*M*=3.76, *SD*=2.13). The switches to L1 were even less frequent (*M*=2.25, *SD*=2.32) than at T1. The perceived influence from L1 (*M*=3.26, *SD*=1.99) in pronunciation was closely comparable to the influence from L2 (*M*=3.25, *SD*=2.25), similarly to T1, both can be described as rather infrequent. Table 48 summarises the L3 proficiency ratings for speech recordings obtained at T2.

Table 48. Proficiency ratings by categories for L3 at T2 (10-point scale).

Rating category	M	95% CI	SD	median	min	Max
comprehensibility	5.43	4.004-6.862	3.05	6.33	0.00	8.67
overall proficiency	4.21	2.954-5.479	2.69	5.00	0.00	7.33
overall pronuncia-tion	3.76	2.765-4.767	2.13	4.66	0.00	7.00
switches to L1	2.25	1.159-3.340	2.32	1.00	0.00	7.67
influence from L1 in pronunciation	3.26	2.332-4.200	1.99	3.67	0.00	6.00
influence from L2 in pronunciation	3.25	2.203-4.330	2.27	3.00	0.00	8.00

At T3, comprehensibility was once again the highest-rated category (*M*=5.75, *SD*=2.65), with the highest ratings of all testing sessions. Likewise, the perceived overall proficiency in L3 was rated higher compared to T1 and T2 (*M*=4.36, *SD*=2.39). Overall pronunciation noted an increase in ratings (*M*=4.68, *SD*=2.39) in comparison to T1 and T3, potentially reflecting the accumulation of knowledge and skill at the later stage of L3 language acquisition. The switches to L1 slightly increased in frequency (*M*=2.53, *SD*=2.44) compared to T2. The influence from L1 (*M*=3.31, *SD*=1.81) in pronunciation was more visible than the influence from L2 (*M*=2.91, *SD*=2.83), however, both can be described as rather infrequent. Table 49 summarises the proficiency ratings for L3 speech recordings obtained in T3.

Table 49. Proficiency ratings by categories for L3 at T3 (10-point scale).

Rating category	M	95% CI	SD	median	min	Max
comprehensibility	5.75	4.505-6.994	2.65	6.50	0.00	8.33
overall proficiency	4.36	2.246-5.487	2.39	5.00	0.00	8.00
overall pronunciation	4.68	3.597-5.769	2.32	5.16	0.00	7.67
switches to L1	2.53	1.387-3.679	2.44	2.00	0.00	9.00
influence from L1 in pronunciation	3.31	2.467-4.165	1.81	3.66	0.00	6.67
influence from L3 in pronunciation	2.91	1.991-3.841	1.97	2.83	0.00	6.00

5.6. Individual differences in multilingual development – results

The results of the phonological perception and production tasks and the scores on individual differences tasks were further analysed through multiple linear regression models to reveal the potential relationships between them. The data prepared for the analysis included overall accuracy scores (OPercAcc and OProdAcc) in L2 and L3 perception and production (outlined in Sections 5.4.1 and 5.4.2) and the scores from individual differences tasks - PWM (see Section 5.5.1), flanker (see 5.5.2) and language proficiency (L2 & L3 Prof) in L2 and L3 (the mean score created from three rated categories – comprehensibility, overall proficiency, and overall pronunciation – see Section 5.5.6). The data analysed through multiple linear regression models came from

T1 and T3 testing sessions since the individual differences tasks were administered only at these testing times.

Multiple linear regression models were built in R software (R Core Team, 2022) to predict the relationship between the L2 and L3 perception and L2 and L3 production overall accuracy scores at T1 and T3, with individual differences tasks scores – PWM scores, flanker scores, and overall proficiency scores in L2 and L3 as fixed effects (R codes available in Appendix G). The results for 20 participants are presented and discussed in the subsections below.

Multiple linear regression models were used to test if PWM scores, flanker scores and L2 and L3 proficiency scores significantly predict the overall speech perception accuracy scores in L2 and L3. As far as L2 in the T1 testing session is concerned, the overall regression was not statistically significant (R^2=0.102, F(3, 16)=0.608, p=.62). None of the individual differences across the task scores significantly predicted the overall L2 speech perception accuracy at T1 (PWM, p=.28; flanker, p=.36; L2 proficiency, p=.35). The overall regression for L2 at T3 was similarly not significant (R^2=0.064, F(3, 16)=0.367, p=.77). Comparably to T1, none of the individual differences task scores significantly predicted the overall L2 speech perception accuracy at T3 (PWM, p=.72; flanker, p=.96; L2 proficiency, p=.31). Linear regressions run for the overall L3 perception accuracy at T1 (R^2=0.032, F(3, 16)=0.181, p=.91) and T3 (R^2=0.135, F(3, 16)=0.831, p=0.49) were not significant. Again, none of the individual differences task scores significantly predicted the overall L3 speech perception accuracy at T1 (PWM, p=.52; flanker, p=.66; L3 proficiency, p=.60) and T3 (PWM, p=.98; flanker, p=.34; L3 proficiency, p=.25). No multicollinearity was observed between the independent variables in the VIF scores (scores for all independent variables were below 1.46). In conclusion, there was no significant correlation between the scores from individual differences across the measured tasks (PWM, flanker, and L2/L3 proficiency) and the overall perception accuracy scores L2/L3 at T1 and T3. The multiple linear regression results predicting the overall speech perception accuracy in L2 and L3 in two testing sessions (T1 and T3) are outlined in Table 50.

Table 50. Multiple linear regression model results predicting speech perception accuracy in L2 and L3.

Language/testing session	Parameter	Estimate	*SE*	Test	*P*
L2 at T1	Intercept	63.15	18.93	3.33	.004
	PWM	-0.62	0.55	-1.11	.28
	Flanker	-0.11	0.12	-0.93	.36
	L2 proficiency	1.83	1.90	0.96	.35
L2 in T3	Intercept	46.59	17.41	2.67	.01
	PWM	-0.21	0.57	-0.36	.72
	Flanker	0.003	0.06	0.05	.96
	L2 proficiency	1.68	1.60	1.05	.31
L3 in T1	Intercept	68.90	18.09	3.80	.001
	PWM	-0.35	0.53	-0.66	.52
	Flanker	-0.05	0.11	-0.43	.66
	L3 proficiency	0.96	1.81	0.53	.60
L3 in T3	Intercept	70.49	14.25	4.94	.001
	PWM	0.01	0.47	0.02	.98
	Flanker	0.05	0.49	0.97	.34
	L3 proficiency	-1.56	1.31	-1.19	.25

Subsequently, multiple linear regression models were used to test if PWM scores, Flanker scores and proficiency scores significantly predict the overall speech production accuracy scores in L2 and L3. The overall regression predicting the overall L2 speech production accuracy at T1 was not statistically significant (R^2=0.121, F(3, 16)=0.735, p=.55). None of the individual differences task scores significantly predicted the overall L2 speech production accuracy at T1 (PWM, p=.91; flanker, p=.64; L2 proficiency, p=.24). The overall regression for L2 at T3 was not significant (R^2=0.083, F(3, 16)=0.486, p=.69). Similarly to T1, none of the individual differences task scores significantly predicted the overall L2 speech production accuracy at T3 (PWM, p=.30; flanker, p=.82; L2 proficiency, p=.99). Linear regressions predicting the overall L3 production accuracy at T1 (R^2=0.208, F(3, 16)=0.404, p=.28) and T3 (R^2=0.024, F(3, 16)=0.134, p=.93) were not significant. Yet again, none of the individual differences task scores significantly predicted the overall L3 speech production accuracy in T1 (PWM, p=.06; flanker, p=.54; L3 proficiency, p=.58) and T3 (PWM, p=.83; flanker, p=.58; L3 proficiency, p=.87). No multicollinearity was observed between the independent variables in the VIF scores (<1.46). It might be concluded that there is no significant correlation between the scores

from individual differences in the measures (PWM score, flanker score and proficiency in L2/L3) and the overall production accuracy in both L2 and L3 at T1 and T2. The multiple linear regression results predicting speech production accuracy in L2 and L3 in two testing sessions (T1 and T3) are outlined in Table 51.

Table 51. Multiple linear regression model results predicting speech production accuracy in L2 and L3.

Language/testing session	Parameter	Estimate	*SE*	Test	*P*
L2 at T1	Intercept	58.39	8.13	7.17	.06
	PWM	0.03	0.24	0.12	.91
	Flanker	0.24	0.05	0.48	.64
	L2 proficiency	1.00	0.82	1.22	.24
L2 at T3	Intercept	54.45	11.29	4.82	>.001
	PWM	0.40	0.37	1.08	.30
	Flanker	0.01	0.04	0.23	.82
	L2 proficiency	0.02	1.04	0.02	.99
L3 at T1	Intercept	51.62	8.29	6.22	>.001
	PWM	0.49	0.24	2.01	.06
	Flanker	0.03	0.05	0.61	.54
	L3 proficiency	0.46	0.83	0.55	.58
L3 at T3	Intercept	74.81	10.28	7.27	>.001
	PWM	-0.07	0.33	-0.21	.83
	Flanker	0.02	0.03	0.56	.58
	L3 proficiency	-0.15	0.95	-0.16	.87

5.7. Multilingual development and psychotypology

The stability of the accuracy in perception and production in both L2 and L3 was mirrored by the stability in the perception of psychotypology. There were, however, individual cases that exhibited substantial changes in their perception of the distances between the L1, L2 and L3. For instance, participant 8 perceived L2 English as 37% closer to L1 Polish at T3 compared to T1; their L2 perception accuracy decreased by 25% percentage points (from 62.5% to 37.5%). The alignment of these results may suggest that the change in the perception of the languages (which are distant typologically in the context of the tested features) negatively impacted the perception accuracy. Moreover, participant 8

perceived L3 German as 139% more distant to L2 Polish at T3 compared to T1; such a change in the perceived distance was aligned with the improved accuracy in L3 German perception by 13 percentage points (from 75% to 88%) and in production by 8% percentage points (from 75% to 83%). In this case, it might be speculated that a small improvement is attributed to an increased perceived distance between the typologically distant languages. However, a different tendency emerged in the case of participant 18, who perceived L1 Polish as almost 122% more distant from L3 German across T1 and T3. Their L3 perception accuracy score decreased by 13 percentage points (from 75% to 62%), whereas their L3 production accuracy increased by 19 percentage points. In this case, only the modality of production was seemingly positively affected by the changing perception of the typologically distant L1 Polish and L3 German. The remaining participants with substantial differences in the perception of language distances listed in section 5.5.3. did not exhibit any considerable changes in the L2 and L3 perception and production accuracy.

5.8. Multilingual development and language proficiency

The combined proficiency score created from three rating categories of language proficiency measure was used in the linear regression model featured in section 5.7. However, the remaining three categories focusing on language switching and CLI were not compared with the multilingual development data. In the current section, the ratings from the aforementioned categories are correlated with the speech production and perception accuracy scores for L2 and L3 at T1 and T3.

In L2, a Pearson correlation coefficient was computed to assess the linear relationship between the rating categories focused on the frequency of switches to L1, influence from L1 on L2 phonology and influence from L3 on L2 phonology and the scores of L2 perception and production accuracy. At T1, a significant relationship was found for the frequency of L1 switches while using L2 and L2 perception accuracy (r=-.56, p=.01), which indicated that the low frequency of L1 switches was moderately correlated with higher accuracy in L2 perception. At T3, the correlations between the categories did not reveal significant relationships between the three rating categories and L2 speech perception and production.

Similarly to L2, in L3 a Pearson correlation coefficient was computed to assess the linear relationship between the rating categories related to the frequency of switches to L1, influence from L1 in L3 phonology and influence from L2 on L3 phonology and the scores of L3 perception and production accuracy. At both T1 and T3, no significant relationship was found for any of the tested variables.

5.9. Multilingual development and language history and use

To compare the measures assessing multilingual development with language history and use, certain sections of the questionnaire such as the self-assessment of language proficiency and language learning attitudes were selected for further evaluation. Additionally, the data obtained from the questions on language mixing and CLI were compared against the proficiency rating scores from the categories concerned with language switches and CLI in pronunciation.

The results of the sections of the questionnaire related to self-assessment of language proficiency were compared to speech perception and production accuracy; their relationship was assessed through the Pearson correlation coefficient. Self-assessment of listening was correlated with speech perception accuracy in both L2 and L3 and self-assessment of speaking were correlated with speech production in L2 and L3. However, no significant relationships were found.

The results of the part of the questionnaire inquiring about language learning attitudes, the role of L1 and L2 in L3 acquisition and language mixing were compared with the accuracy of the perception and production scores as well as the language proficiency rating results. The first question of this part of the questionnaire inquired about the attitude towards foreign language learning; the answers were compared with the perception and production accuracy scores. The participants who expressed a positive attitude towards language learning did not exhibit a vastly improved perception and production accuracy in either their L2 or L2 compared to those who disliked the activity; the average difference between the accuracy scores between these two groups was 9 percentage points (for L2 perception accuracy in T2) and the smallest was 0.8 percentage points (for L3 production in T3). The participants who assessed L1 Polish as more helpful in learning L3 English did not substantially differ in the perception and production

accuracy in L3 from the group who pointed to L2 English. The largest difference was observed for T1 and T3 L3 perception accuracy; the participants who selected L1 Polish as more helpful were respectively 8 percentage points and 5 percentage points less accurate compared to the participants who selected L3.

The results of the questions inquiring about language mixing and CLI were compared with the proficiency rating scores in three categories - "the frequency of switches to L1", "the influence from L1 in target pronunciation", "the influence from L2/L3 in target pronunciation". The participants who noticed they mixed their languages in one conversation did not differ considerably from the participants who did not notice such occurrences in the three previously listed categories of proficiency rating (see section 5.5.6). The overall difference between the two groups in L2 was 0.15 and in L3 it was 0.25 (0-10 rating scale, 0 – no switches/influence / 10 – very frequent switches / heavy influence). The described comparison implies that the awareness of language mixing did not affect the switches to L1/Ln influence of currently used language.

5.10. Summary of the obtained results

Chapter 5 outlined the results of the battery of tests administered in three testing sessions. The speech perception data collected in the ABX tasks in L2 and L3 did not reveal considerable changes over the testing sessions, however, there was a noticeable difference in perception accuracy between the higher scores obtained in L3 and the lower scores in L2. The speech production data collected in the delayed repetition tasks in L2 and L3 showed differences over the testing sessions on the basis of individual features. The perception-production interaction between the languages revealed no correlation, however, the participants were more accurate in L2/L3 production than L2/L3 perception. Further, Chapter 5 outlined the results of the individual differences tasks. Subsequently, the multiple linear regression model built to explain the relationship between perception/production tasks results and the scores on individual differences measures did not reveal any significant relationships. Consequently, it indicates that individual differences measures failed to predict multilingual development in the current study. Further, Chapter 5 discusses the results of the individual differences measures not included in the model in relation to the multilingual development.

Chapter 6: Discussion of the results

6.1. Introductory remarks

The aim of Chapter 6 is to discuss the results obtained in three testing sessions outlined in Chapter 5 in the light of the research questions and predictions outlined in Chapter 4 as well as models of multilingual acquisition and earlier findings. Sections 6.2 discusses the results of multilingual speech perception and production tasks in L2 and L3. Section 6.3 discusses the perception/production results in L2 and L3 in relation to individual differences measures. Limitations of the study and future research directions are discussed in sections 6.4 and 6.5. Chapter 6 ends with the concluding remarks.

6.2. Multilingual speech perception and production

Multilingual perception and production were examined through a battery of tasks described in Chapter 4 (sections 4.4 to 4.7). Speech perception in L2 and L3 was tested in separate ABX auditory discrimination tasks, whereas speech production in L2 and L3 was investigated via delayed repetition tasks. The following subsection includes a discussion of the L2 speech perception and production results and follows with a discussion of the respective results in L3. The subsections include answers to research questions 1-3, namely, (1) What is the development of L2 perception and production throughout three testing sessions?, (2) What is the development of L3 perception and production throughout three testing sessions?, (3) Is there symmetry in the gain between L2 perception and production over time? Is one of the modalities developing faster?

6.2.1. Speech perception and production in L2

Speech perception in L2 English was examined through the ABX task featuring close front unrounded vowels /ɪ/-/iː/. The results in three testing sessions remained stable, with mean accuracy scores oscillating between 48 and 51%, remaining close to the chance

level. Such results indicate that the accurate identification of the tested vowels posed quite a challenge for the participants and remained a difficult task across the three testing sessions, with no significant difference between the perception accuracy scores in T1, T2 and T3.

The production of close front vowels in L2 was analysed acoustically in terms of their duration. The production of /ɪ/ and /iː/ differed significantly; the former was produced as shorter, whereas the latter as longer – this pattern was maintained throughout the three testing sessions. The overall difference between the duration of /ɪ/ and /iː/ was approximately 30 ms. Moreover, the duration of /ɪ/ corresponded with the mean duration of vowels in L1 Polish (75.57 ms before a voiceless stop and 83.11 ms before a voiced stop, Coretta 2019), whereas the production of /iː/ was characterised by hybrid values, exceeding L1 means. Furthermore, no significant change in vowel duration over time was detected; the /ɪ/ and /iː/ retained comparable mean values at T1, T2 and T3. The production of /ɪ/ and /iː/ vowels was influenced by the target words in which they were embedded. In the context of words ending with a fortis consonant (*wick, week*), /ɪ/ and /iː/ were quite close in terms of duration, which can be attributed to the phenomenon of pre-fortis clipping. The remaining realisations in the target words ending with lenis consonants featured a visible distinction in duration between /ɪ/ and /iː/ vowels (*bid, did* vs *bead, deed*). Such an observation may indicate that the participants were sensitive to different types of acoustic cues (vowel duration, pre-fortis clipping) while processing the target forms, which was subsequently reflected in their productions.

Compared to vowel duration, L2 VOT was more prone to changes; there was an overall significant increase in VOT values between the testing sessions T1 and T2 and a relative stability between T2 and T3. The production of VOT in /k/ was characterised by the longest duration among all tested voiceless plosives and differed significantly from the VOT of /p/ and /t/. The overall mean VOT values of /k/ at three testing sessions (T1 – 60, T2 – 71 ms and T3 – 68 ms) approached native-like values outlined by the earlier studies (Lisker and Abramson 1964 – 80 ms, Kopczyński 1977 – 71 ms). There was also significant difference between the VOT values of /p/ and /t/. The production of /p/ oscillated between the L1-values reported by earlier studies (Lisker and Abramson 1964 – 21.5 ms, Kopczyński 1977 – 37 ms), with the means of 31 ms at T1, 40 ms at T2 and 36 ms at T3. A similar tendency was visible for /t/ at T1 (31 ms), however, the VOT at T2 (46 ms) and 47 (ms) was characterised by a slight increase compared to the values

observed for L1 Polish (Lisker and Abramson 1964 – 27.9 ms, Kopczyński 1977 – 33 ms). The target-like production of /k/ accompanied by L1-like realisations of /p/ and /t/ might prompt different explanations. The participants might have, to a certain extent, acquired the L2 values for /k/ and failed to acquire (or might have just started their acquisition in the case of /t/) the L2 values of the other voiceless plosives. However, the observed long VOT durations of L2 velar plosives might have been also influenced by the L1 Polish phenomenon described by Waniek-Klimczak (2011), namely, an increase of /k/ VOT values in L1 Polish among younger speakers.

Research question 1 inquired about the development of L2 perception and production throughout the three testing sessions. Out of three scenarios proposed in Chapter 4 (section 4.2.1), the participants' results correspond with the first one, i.e. the perception and production in L2 remained rather stable across three testing times, with re-emerging differences based on individual features. L2 perception accuracy of /ɪ/ and /iː/ was low and remained at chance level across three testing times. The L2 production of /ɪ/ and /iː/ was characterised by greater accuracy than perception but it also did not change over time. The production of VOT in voiceless plosives exhibited a significant increase in duration between the testing sessions T1 and T2 and slight decrease at T3. The relative stability of the L2 perception and production accuracy over three testing sessions corresponds with the results of the previous studies, which also noted the stability of the L2 acquired already for a longer period in the formal instruction context (Balas et al. 2019, Sigmeth and Golin 2018, Wrembel et al. 2020, Wrembel et al. 2021).

Research question 1 in Chapter 4 was also followed by some predictions concerning the tested features: vowel duration and VOT of voiceless plosives. The low accuracy scores in the perception of /ɪ/ and /iː/ might be associated with the assumption of PAM-L2; the two separate English phonemes were assimilated to the single L1 category. However, the /ɪ/ and /iː/ production data indicate that the participants distinguished between the two L2 vowels, producing the former with L1-like values and the latter with significantly longer, L1-L2 hybrid values. Such a finding might be interpreted as an instance of production accuracy preceding perception accuracy, however, it might be also associated with the nature of the task, i.e. the ABX paradigm might have posed too great a challenge for the adolescent participants. The stronger influence of L1 in the later stages of L2 acquisition predicted by the NGTA framework is reflected by L2 perception data, namely, the participants struggled with distinguishing

between the vowel duration contrasts not present in their L1. However, it might be speculated that L1 CLI did not affect speech production to the same extent, as evidenced by the presence of the distinction between /ɪ/ and /iː/ in terms of vowel duration. Additionally, the nature of the delayed repetition task offers one more potential explanation for the improved production performance; the attention to the target material enforced by the task brought about the heightened attention to acoustic cues, which minimised the influence of the L1 CLI and resulted in more target-like production.

As suggested by PAM-L2 (Best and Tyler 2007), the association of L2 voiceless stops with L1 features induced the assimilation to the existing L1 categories, which had potentially, along with a small amount of target-like input, prevented the category formation and distinguishing between L1 and L2 stops in terms of VOT duration. However, there was an indication that the process of creating the separate categories for L2 voiceless plosives might have been underway, namely, the target like production of the velar plosive /k/ and the instances of hybrid-like VOT of the alveolar plosive /t/. The obtained results correspond with the predictions of the NGTA framework, which suggested that L1 CLI dominating the later stages of language acquisition might have reinforced non-target like VOT, especially in the case of bilabial and alveolar plosives.

6.2.2. Speech perception and production in L3

Speech perception in L3 German was examined through the ABX task featuring close front unrounded vowels /ɪ/-/iː/ as well as close front rounded /ʏ/ and /yː/. The results in three testing sessions remained at a stable level, with mean accuracy for /ɪ/-/iː/ between 63-68% in three testing sessions and 58-62% for /ʏ/ and /yː/. Moreover, the mean accuracy scores indicated that the participants were slightly better at recognising the close front unrounded duration contrasts than close front rounded ones. Additionally, compared to L2 accuracy scores, the participants were better at the discrimination of L3 contrasts across three testing sessions. However, as far as the change over time is concerned, there was no difference between the L3 perception accuracy scores at T1, T2 and T3.

The production of close front vowels in L3 German was analysed acoustically in terms of duration. Similarly to L2 English, the production of long vowels /iː/ and /yː/ was characterised by significantly longer duration compared to short vowels /ɪ/ and /ʏ/. This

tendency was present throughout the three testing sessions, with the overall difference between the short and long vowels at around 35 ms. Similarly to vowel duration in L2, short vowels in L3 were produced with durations close to those in L1 Polish, while the production of long vowels was characterised by hybrid values, exceeding L1 duration means. Overall, a significant change in vowel duration over time was not detected, however, some development was observed for the individual vowel type /yː/; its duration increased over time and differed significantly at T1 vs. T3 and T2 vs. T3. The production of the vowels was influenced by the target words in which they were embedded; the /iː/ vowel before a voiceless consonant (in the word *Miete)* was considerably shorter than /iː/ in the remaining contexts.

L3 German VOT values were characterised by an overall significant increase between the testing sessions T1 and T2 and relative stability between T2 and T3; this significant difference between T1 and T2 mirrored the development observed in L2. Moreover, there was a significant difference between the durations of all tested plosive with different places of articulation, namely, /p, t, k/. The VOT of /p/ remained at the L1-level (T1 – 29 ms, T2 – 28 ms and T3 – 23 ms). The production of VOT in /k/ was characterised by the longest duration among all tested voiceless plosives, followed by /t/ and /k/, remaining in line with universal VOT trends. The overall mean VOT values for /k/ at three testing sessions (T1 – 46 ms, T2 – 57 ms and T3 – 55 ms) were increasing slightly above L1 values, which might be interpreted as the potential early development of hybrid values. The mean VOT values of /t/ were close to L1-like at T1 (25 ms), however, they increased significantly at T2 (47 ms) and T3 (44 ms) and reached target L3 values.

Research question 2 inquired about the development of L3 perception and production in three testing sessions. Out of three scenarios proposed in Chapter 4 (section 4.2.1), the participants' results correspond with scenario 1 and scenario 2. L3 German perception, more accurate than L2 English perception, might have been influenced by the novelty effect, i.e. the reliance on the acoustic cues leads to increased perceptual accuracy in T3 (as predicted by the NGTA framework and reported by Nelson 2020 and Wrembel et al. 2020). The stability of L3 perception accuracy across three testing sessions might indicate that such a novelty effect might have lasted through the first 10 months of L3 learning. The L3 production largely conformed to scenario 1, showing relative stability with differences on the basis of individual features. When it comes to L3 vowels, the

production accuracy of /ɪ/, /iː/ did not change over time. A significant increase in the duration was observed for /yː/, which might be linked to the process of developing a new, separate category for this vowel over time. Similarly to L2, VOT duration in voiceless plosives exhibited a significant increase between the testing sessions T1 and T2 and stability between T2 and T3. The relative stability of the L3 perception and production accuracy over three testing sessions corresponds with the results of the previous studies, which also noted a slight development or stability in the early stages of L3 acquisition (Balas et al. 2019, Wrembel 2020). As far as the "u- shaped" development of the early stages of L3 learning described by Wrembel et al. (2019) is concerned, the participants did not follow a similar route: their L3 perception was characterised by a rather flat trajectory whereas their production exhibited a slight peak at T3.

Research question 2 concerning the development of L3 perception and production was supplemented with some predictions concerning the tested features. The NGTA framework (Dziubalska-Kołaczyk and Wrembel 2022) predicted increased sensitivity to acoustic cues during the early stages of acquisition; thus a greater perception accuracy in L3 compared to L2 might be interpreted as such. As for L3 vowel production, the participants exhibited a combined influence of L1 and L3. They produced L3 short vowels with L1-like vowel duration and long vowels with significantly longer durations, which might be identified as attempts at a target-like production, also visible in L2. Such a cross-linguistic influence pattern corresponds with the one detected by Sypiańska (2016b) in her study of vowel quality, where L3 vowel production was affected by both L1 and L2. Consequently, the joint L1 and L3 influence can be interpreted instance of the combined CLI (De Angelis 2007). Additionally, it corresponds with the framework of the Cumulative Enhancement Model (Flynn et al. 2004), which assumes the cumulative influence from the other languages in the repertoire of the speaker.

Similarly to L2, the analysis of L3 VOT production data revealed that the accuracy of realisations was category-dependent. Again, the similarity of L3 voiceless stops to Polish features may lead to the assimilation to the existing L1 categories, whereas limited input might prevent the participants from developing target-like production (as predicted by PAM-L2). The T1 data were characterised by the greatest amount of L1-driven CLI across all the plosive types. At T2 and T3, there was a significant increase in VOT duration of the voiceless alveolar plosive and the voiceless velar plosive; the former reached L3 values, whereas the latter remained at an L1-like level. The significant

increase in L3 German VOT values for /t/ corresponded with a similar increase in L2 English, which might be interpreted as an instance of CLI. The VOT values for /k/ in L3 were lower than in L2. However, they were high enough to be classified as a potential L1/L3 hybrid. The effect observed for /t/ might be identified as the influence of the L2 (in line with the L2 status factor model, Bardel and Falk 2007). The slightly increased values for /k/ might be a result of the combined L1 and L3 CLI, previously evidenced in VOT-focused studies by Wunder (2010) and Wrembel (2011), however, the observed effect is rather subtle. Alternatively, the increase of VOT in L3 velar plosives might be linked to the L1 phenomena of increasing VOT duration among young speakers described by Waniek-Klimczak (2011). Overall, the presence of shifts in L3 VOT patterns might possibly be attributed to the exposure to the L1 short-lag (potentially slightly increasing in the case of velar plosive) vs. L2/L3 long-lag VOT.

6.2.3. Multilingual perception and production – comparison of the results between the languages and modalities

The results obtained through the perception and production tasks in L2 and L3 were transformed into accuracy scores (see section 5.4) to enable further comparison between the languages and modalities. The current section summarises the findings concerning the accuracy in the between-language perspective as well as between-modalities perspective. Further, the findings are considered against the predictions outlined in Chapter 4.

The comparison between the languages (L2 vs. L3) revealed a greater accuracy of L3 perception compared to L2 perception across three testing sessions. The obtained results may, to a certain extent, correspond with Onishi (2016), who suggested that the improved accuracy of L3 perception may be not only an extension of L2 perception accuracy (not observed in the case of the current data set), but an overall result of accumulated experience of foreign language learning. Consequently, the improved L3 perception may be a result of such a more general, accumulated language experience. Both languages were characterised by a slight peak in accuracy at T2. In the case of vowels, the production accuracy was generally higher for L3 than L2. Such a trend might be associated with the fact that the target durations of tense vowels in L2 were longer than those in L3, consequently, the hybrid values produced in L3 were closer to the target than

those produced in L2. The two languages followed slightly different paths over the testing sessions. At T2, L2 accuracy was characterised by a small decrease whereas in T3 L3 was characterised by a small increase and L2 returned to T1 accuracy levels. In terms of VOT, the trend was reversed – the participants were more accurate in L2. Additionally, both languages were characterised by a slight peak of accuracy at T2.

Further, the overall production and perception scores in the respective languages were compared in order to answer research question 3, inquiring about the symmetry of the gain between the two modalities from the longitudinal perspective. In L2, the overall production was more accurate than perception; with the difference oscillating between 18-21 percentage points across the three testing times. The development curve of L2 perception was rather flat, with a stable accuracy percentage across the testing sessions. The L2 production noted a slight increase at T2 maintained at T3. In L3, the participants also demonstrated higher accuracy in L3 production compared to L3 perception, but the differences were less prominent (between 5 and 9 percentage points across the testing sessions). Both modalities were characterised by a slight increase across three testing sessions. The Pearson correlation coefficient computed to analyse the strength and the directionality of the relationships between the perception and production in L2 and L3 in three testing sessions revealed no significant results. The overall difference between the perception and production accuracy scores was more visible in T2; production surpassed perception by an average of 19.33 percentage points. In L3 a similar tendency was observed, but the distinction was smaller – around 7.33 percentage points.

Research question 3 concerning the symmetry of the gain between the perception and production in L2 and L3 was followed by some predictions concerning the direction of these relationships. Section 4.2.1 outlined four scenarios observed in earlier studies: (1) perception exceeds and precedes production in L2 speech learning, (2) perception and production are aligned and co-develop, (3) accurate production precedes perception, (4) there is no direct link between the perception and production. Based on the perception-production data from multilingual phonology studies (e.g. Liu and Lin 2020, Wrembel et al. 2022), these scenarios were narrowed down to two possible routes: (1) there is an overall link between the perception and production of multilingual participants, L2 perception accuracy exceeds production and the link is more stable compared to L3; L3 perception might initially be characterised by high accuracy due to the influence of the early reliance on acoustic cues (e.g. the mechanism proposed by NGTA, "novelty effect",

Nelson 2020); (2) There is no overall link between the perception and production of multilingual participants. The L2 and L3 perception-production relationships may diverge and are characterised by increased individual variation over time. The overall results of the Pearson correlation aligned with the predictions of the second scenario – there was no significant correlation between the perception and production accuracy in both L2 and L3 (similarly to Liu and Lin 2020). The overall L3 perception was characterised by greater accuracy than L2 perception potentially pointing to the aforementioned "novelty effect" and reliance on the acoustic cues. However, as far as the comparison of the numerical accuracy scores is concerned, the perception was more accurate than production for both languages. The common prediction about the "perception driving production" (Wrembel et al. 2022) did not manifest in the results obtained in the current study. Such an accuracy gap between the perception and production might be explained by the potentially higher demand of the ABX task, both in terms of recognising the contrasts as well as understanding the rules of the task and focusing. The accuracy score discrepancy between the perception and production was greater in the case of L2; such a result may be linked to the assimilation of the tested L2 vowels to a single L1 category (hence low perception accuracy) and production-leading-perception mechanism in the case of the production of L2 features. The smaller perception-production gap in L3 might be related to the previously mentioned heightened attention to the acoustic cues in the early stages of L3 acquisition, which might have affect both perception and production.

6.3. Individual differences in multilingual development of phonology

The results of the multilingual perception and production tasks in L2 and L3 and the scores obtained through individual differences tasks (the flanker task for inhibitory control, pseudoword repetition for phonological working memory) and short proficiency interviews (in L2 and L3) were analysed through linear regression models. The individual difference measures were positioned as the predictors of L2 and L3 perception and production. None of the individual differences variables significantly predicted the accuracy in L2 or L3 production and perception. The closest outcome to approach significance was the positive estimate predicting the relationship between PWM scores and L3 speech production accuracy in T1 (r=0.49, p=.06). The main research question

inquiring about the interaction between the results of the perception/production tasks with measures of individual differences can be, therefore, answered by stating the lack of significant relationship. The scores obtained by the participants in individual differences tasks did not predict the accuracy in L2/L3 perception and production at T1 and T2. The discussion of the overall outcomes of the individual differences measures and the answers to the questions inquiring about the particular individual differences in relation to multilingual development (represented by perception and production accuracy in L2 and L3) are presented in the subsections below.

6.3.1. Inhibitory control and multilingual development

Inhibitory control was tested in a flanker task administered in T1 and T3 testing sessions. The accuracy scores and reaction time measurements obtained in two types of trials – congruent and incongruent – were computed to create the flanker score - the lower the score the stronger the inhibitory control of a participant. Low flanker scores suggest that both the congruent and incongruent trials had a similar cognitive load; such scores may indicate a greater inhibitory control. The participants' results demonstrated that incongruent trials required more time to complete with mean flanker scores of 53.90 (T1) and 59.00 (T3), with no significant change over time.

The prediction formulated in the research questions (section 4.2.2) anticipated that a lower flanker score (indicating greater inhibitory control) would be linked to improved perception and production accuracy. The participants with greater inhibitory control capacity would be capable of inhibiting the influence of other languages in a more successful way, which would result in more accurate perception and production in L2 and L3. Linear regression models build to analyse the relationships between the flanker score and L2/L3 speech perception/ production at T1 and T2 did not reveal any significant relationships. The accuracy of L2 and L3 speech perception measured through ABX tasks was not predicted by the PWM scores. Similarly, more accurate production in L2 and L3 was not associated with increased inhibitory control. Moreover, there was no change in the relationship between the flanker score and L2/L3 perception and production across the testing sessions. The results do not correspond with the previous studies indicating that inhibitory control capacity may be linked to greater accuracy in multilingual

perception and production (Mora and Darcy 2013, Darcy et al 2014, Darcy et al. 2016, Lev-Ari and Peperkamp 2013). Additionally, the results only partially follow the patterns observed for the portion of VOT and rhotics data collected in the T1 testing session of the current study presented in Krzysik (2020). In the aforementioned study, L2 production accuracy moderately correlated with inhibitory control and, similarly to the current results, there was no correlation observed for L3 production accuracy. Moreover, the obtained results correspond with a study conducted with a similar group of adolescent students (Sigmeth and Golin 2018), which found no meaningful relationships. The obtained results may be also considered in the light of Green's activation-inhibition model (1986, 1993, 1998) and its elaboration by Festman (2008, 2018). Green's model predicted that performance in a given language may be influenced by one's inhibitory control capacity, which is required to balance out the influence of the remaining active linguistic system/systems. The lack of relationship between the inhibitory control capacity and speech perception/production in L2 and L3 does not align with this prediction. Festman (2018) suggested that switching between the languages and inhibiting those currently not in use may be a potential source of multilingual advantage. The participants of the current study, possibly due to their low frequency of foreign language use and no opportunity for frequent switching, did not exhibit a visible advantage in their flanker task performance.

6.3.2. Phonological working memory and multilingual development

Phonological working memory (PWM) was tested in a pseudoword repetition task administered in T1 and T3 testing sessions. The obtained recordings were rated in terms of the accuracy of pseudoword production, which generated a phonological working memory score reflecting one's PWM capacity. The participants' scores (see section 5.5.3) indicated mean PWM capacity to reach respectively 71% (T1) and 74% (T3) of the maximum score. Moreover, there was a slight improvement in the PWM capacity over time, which proved to be statistically significant. Such a result might indicate a slight refinement of the PWM between T1 and T3.

The prediction formulated in the research questions for the current study (section 4.2.2) proposed that greater PWM capacity would be linked to improved perception and production accuracy. Consequently, the participants with greater PWM capacity would

score higher on ABX tasks in L2 and L3. Additionally, the participants with increased PWM capacity might produce more target-like forms in the delayed repetition tasks in L2 and L3. Linear regression models built to analyse the relationships between the PWM and L2/L3 speech perception/ production at T1 and T3 did not reveal any significant relationships. The accuracy of L2 and L3 speech perception measured through ABX tasks was not predicted by the PWM scores. A similar outcome was obtained for speech production; more target-like production in L2 and L3 was not associated with increased PWM capacity (understood as a greater PWM score). Moreover, no change in the relationship between the PWM scores and L2/L3 perception and production was observed in the longitudinal perspective. Consequently, the obtained results are not in line with the previous studies indicating that PWM capacity may be linked to improved phonological perception (McKay et al. 2001, Aliaga-García et al. 2011, Darcy et al. 2015) or production (Krzysik and Wrembel 2019; current T1 PWM data correlating moderately with the accuracy of rhotic production).

6.3.3. Language proficiency and multilingual development

Language proficiency in L2 and L3 was measured through short interviews administered at T1 and T3. The recordings of the participants' performance were rated by independent raters in six categories described in section 4.7.6.: comprehensibility, overall fluency, overall perception, the frequency of L1 switches, the influence of L1 on pronunciation, the influence of L2/L3 on pronunciation. The results showed that L2 English was rated higher on comprehensibility in both T1 and T3 compared to L3 German. The influence of L1 on pronunciation, as predicted by the NGTA, was rated as more visible in L2, whereas the influence of L3 was rare. In the case of L3, the influence of both L1 and L2 was on a similar, albeit low level.

To answer the research question inquiring about the relationship between proficiency and speech perception and production in L2 and L3, the combined proficiency score generated from three rating categories (comprehensibility, overall fluency, overall pronunciation) for both L2 and L3 ratings was entered into the linear regression model featured in section 5.6. However, the combined proficiency score in both L2 and L3 failed to significantly predict the perception and production accuracy in the respective

languages. The remaining three categories were focused on CLI: the frequency of switches to L2, the influence of L1 on L2 or L3 phonology and the influence of L2/L3 on L3/L2 phonology were correlated with speech perception and production in the respective languages. The only significant relationship was found in T1 for L2, where the frequency of L1 switches while using L2 correlated with L2 perception accuracy, indicating that the low frequency of L1 switches might be related to higher accuracy in L2 perception. Consequently, language proficiency assessed on the basis of spoken interviews turned out to be a poor predictor of speech perception and production in L2 and L3 in the current study.

6.3.4. Psychotypology and multilingual development

Psychotypology was evaluated with the application of the ViLDiM measure (Nelson et al. 2021) administered in T1 and T3 testing sessions. The research question, as formulated in section 4.2.2., inquired about the relationship between the results of phonological perception and production tasks and the participants' perception of language distances (focusing on phonology). The formulated prediction suggested that the perception of the phonological distances between the L1 Polish, L2 English and L3 German may follow the same pattern as in the study with a comparable group of participants by Nelson et al. (2021), with the pair of L2-L3 identified as the most similar. Moreover, it was speculated that the participants who perceive the distances between the L2 and L3 as closer tend to perform with greater accuracy on the perception tasks in L2 and L3 due to their potentially improved ability to spot regularities between the languages, which can result in enhanced linguistic performance.

At T1, the measurements revealed that the participants perceived the L2 English-L3 German pair as the closest one, followed by L1 Polish-L2 English and L1 Polish- L3 German. At T3, a similar tendency was observed; the participants considered the L2 English-L3 German pair as the closest, followed by L1 Polish- L2 English and L1 Polish-L3 German. The generated results correspond with those obtained by Nelson et al. (2021), who examined a similar group of L1 Polish, L2 English and L3 German adolescents acquiring their foreign languages in the Polish school system. The results might indicate that the participants were largely driven by the language typology. Moreover, the

correspondence with the results obtained by Nelson et al. (2021) might indicate that the common characteristics of the learning environment or teaching materials made the two groups of the participants aware of the distinctions and similarities between these languages.

The relationship between the perception-production accuracy and the perception of language distance was characterised by a similar overall stability; the majority of the participants did not exhibit any meaningful connections in this area. Some interesting interactions arose on an individual basis. For one participant, a decrease in L2 perception accuracy was accompanied by the decrease of the psychotypological distance between L1 Polish and L2 English. Subsequently, the perception of the more typologically distant pair languages as closer might have affected the perception negatively. Another case involved a substantial increase in the perceived distance between L1 Polish and L3 German aligned with a moderately improved perception and production. Yet another individual case indicated a disparity between the perception and production; an increase in the perceived distance between L1 Polish and L3 English corresponded with a decrease in the L3 perception accuracy and L3 production increase. Consequently, the obtained data, despite the instances of individual meaningful interactions, do not confirm the prediction concerning the link between the psychotypology and perception and production accuracy in L2 and L3.

6.3.5. Language history and use and multilingual development

The participants' language history and use was assessed through language history and use questionnaires administered in T1 and T3. The questionnaires inquired about a number of aspects including biographical data, self-assessment of language proficiency, the contexts of language use, language switching and CLI and the attitudes towards language learning. The collected biographical data were used to establish exclusion criteria for the tested group. Subsequently, they were outlined in the section 4.3., presenting the general characteristics of the tested group. The discussion of the results of the remaining aspects of the questionnaire along with the answers to research questions exploring the relationship between language history and use with speech perception and production are outlined below.

The data collected through self-assessment of language proficiency revealed that the participants rated their L2 proficiency in four categories (reading, writing, speaking, listening) as better than in L3 in both T1 and T3. These results are aligned with the rating provided by the independent raters on the basis of short interviews, in which L2 English proficiency was also rated as higher than that in L3 German. Moreover, the participants were more confident about their speaking compared to listening in both L2 and L3; such a finding might be related to the general tendency observed in production-perception accuracy, in which production seemed to precede perception. However, the correlations between the self-assessed listening proficiency and perception and self-assessed speaking proficiency and production in both languages in T1 and T3 were not found to be significant.

The contexts and modes of language use were explored in two questions. The first question investigated three contexts of language use: with family, with friends and at school. All three contexts were largely dominated by L1 Polish, which confirmed the expectations based on the characteristics of the L1-dominated, fairly homogenous study setting (outlined in section 4.3.). The interactions with family and friends were dominated by L1 in more than 90% in both T1 and T3, with approximately 5% of L2 English and below 2% of L3 German. The context in which L2 and L3 were more frequently applied was school; L1 was restricted to approximately 72% of the interaction, followed by around 17% in L2 and 11% in L3 in both testing sessions. Such results once again align with the characteristics of the study setting described in section 4.3., in which the main source of contact with foreign languages was through formal instruction at school. In the context of media through which the participants used and were exposed to the languages in their repertoire, the majority of their interactions such as TV/radio/YouTube watching/listening, book reading, gaming as well as the Internet browsing and communication were L1-dominated. The activities in which L2 had a substantial role were TV/radio/You Tube watching/listening, gaming and the Internet browsing, which occupied between 15% and 40% of the total time devoted to these activities. The interactions in L3 occupied less than 2% of time devoted to these activities. As far as the comparison of the results outlined in the current paragraph with the perception and production accuracy is concerned, the low frequency of L2 and L3 use might have substantially influenced multilingual development. Firstly, low L2 perception accuracy might be associated with low exposure to L2 contrast, which in turn affected the

difficulties in differentiation between the /ɪ/-/iː/ contrast. The L2 exposure in the school context and through the Internet and gaming was not sufficient to meaningfully influence the development of said features. Moreover, the role of the listed media in the development of target-like features might be also restricted by the exposure to non-native speech common in the Internet resources and online gaming. Moreover, the dominance of L1 might have reinforced the L1-based CLI in both speech perception and production, sometimes accompanied by the instances of L2-driven or cumulative CLI observed for particular L3 features (L1-influenced short vowels vs. L2 -influenced long vowels, L1-driven VOT in /p/, a potential L2/L3 hybrid in /k/ vs. partially L2-driven VOT in /t/). The results of the questionnaires revealing the L1-dominated language environment of the participants, paired with the assessment the lack of focus on speaking and pronunciation especially in the early stages of learning in Polish schools (Róg 2020), may potentially explain the low-to-moderate perception and production accuracy scores as well as the comparable accuracy between the L2 and newly acquired L3.

The questions investigating language attitudes revealed that the majority of the participants had a positive attitude towards foreign language learning, with a preference for English. Over time, the participants recognised the helpful role of L2 in learning L3 German, with the initial results indicating a comparable influence of L1 Polish. Almost 2/3 of the participants were aware of language mixing and switching. The positive attitudes towards language learning, with a preference for L2, were not aligned with increased accuracy in L2. However, the overall positive attitude towards language learning and the recognition of usefulness of L3 might have potentially contributed to improved attentiveness towards this language, reflected in heightened accuracy and associated with the "novelty effect" (Nelson 2020). Additionally, the results of the questions inquiring about language mixing and CLI were compared with the proficiency rating scores in three categories - "the frequency of switches to L1", "the influence from L1 in target pronunciation", "the influence from L2/L3 in target pronunciation". It turned out that the mere awareness of language mixing and switching mixing did not substantially affect the frequency of switches to L1/L3/L2.

6.4. Limitations of the study

The present study employed a battery of different tests administered in three testing sessions over 10 months in order to assess the role of individual differences in the multilingual development of adolescents. Despite the modest number of participants and the confinement to three testing sessions, the study resulted in a relative breadth of data. However, the adopted design and the setting were characterised by certain limitations discussed in the paragraphs below.

Firstly, the administration of the testing battery for a single participant was confined to a single testing session, which implies that all languages (L1, L2 and L3) were tested on the same day. Such organisation of the testing sessions was determined by the rules imposed by the school and the limitations of conducting the tests by a single experimenter. The experimenter attempted to induce a language mode through short interviews in a given language before each language block. However, the activation of the remaining languages before the block of tasks in the language tested at the moment might have led to increased CLI between the tested linguistic systems.

Secondly, there were several limitations associated with the applied tests. Both the ABX tasks and the delayed repetition tasks had a limited number of the tested items. Such a testing design choice was to a large extent affected by the organisational constraints of the testing sessions (listed in the previous paragraph), however, they might have influenced the depth of language perception and production assessment, as well as the possibilities concerning a more in-depth analysis.

Moreover, it might be argued that a study in a form of structured tests (as opposed to natural observation) might have influenced the behaviour of the participants. Despite the efforts of the experimenter to introduce herself to the participants before the session and to induce a stress-free atmosphere, some of the participants might have experienced feelings akin to exam stress. Such emotions, in turn, might have had an impact on their linguistic performance. The stress might have especially affected the participants' responses during the proficiency interviews, which were characterised by scarce production in the case of a substantial number of the tested individuals.

The final limitation, which is also associated with testing outside the lab in general, is related to administering the testing battery in a room located in the school of

the participants. Such a setting rendered the organisation of the testing easier, however, it was associated with a certain amount of the outside noise and distraction.

6.5. Future research directions

Future research projects interested in studying the role of individual differences in multilingual development may consider expanding their scope to test larger groups, possibly across different educational contexts or language combinations. Further studies may also benefit from more focused testing of a single phonological feature to obtain more in-depth results. Research teams with greater resources would also benefit from a more comprehensive investigations involving a wider range of phonetic and phonological features and a larger battery of tests. Future studies might consider the implementation of more than one measure of perception and production e.g. a categorisation study and a discrimination study for perception and a delayed repetition and a storytelling exercises involving spontaneous speech. Such tests may provide the researchers with a more diverse perception and production data, potentially offering a greater explanatory potential.

The current study showed that the pace of multilingual development in the context of formal education in a largely monolingual country might be considerably slower than assumed based on the official learning curriculum of Polish schools. As largely evidenced by rather stable results of perception and production tasks as well as low rates of self-reported foreign language use and exposure, the participants' development was characterised by a rather slow pace. Therefore, future studies interested in researching multilingual development in similar social and educational settings may consider spacing their testing sessions over a longer period e.g. two school years. Such a prolonged span of testing may offset the risk of observing no multilingual development in an environment characterised by scarce foreign language exposure and use. Another conclusion related to the participants' multilingual development, especially in L2 English, might be tied to the Polish national curriculum and foreign language teaching, which emphasizes grammar over speaking fluency and pronunciation (Róg 2020). Despite their long L2 studying experience, the participants struggled with speaking, especially during the short language proficiency interviews, which created an impression that it was not a frequent activity

undertaken by the participants, potentially associated with stress. Such a group of participants may largely benefit from separate sessions for L2 and L3 and perhaps additional sessions inducing the desired language mode before the actual testing.

Another possible direction, especially for a study with a larger scope, is to conduct interviews with teachers and parents of the students participating in the study. Foreign language teachers may provide additional information about the characteristics of foreign language teaching and learning, including methods, modes and teaching materials. Additionally, they may provide a sample of their pronunciation model, which can be used to draw conclusions about the role of target-like input of the teacher in the development of multilingual speech perception and production. Furthermore, the input from the parents may supplement the biographical and language history background of the participants.

6.6. Summary

Chapter 6 aimed to interpret the findings generated through the testing battery employed in the current study. The chapter discussed the results of multilingual speech perception and production tasks in L2 and L3 in relation to the research questions and predictions outlined in Chapter 4 as well as selected models of language acquisition. Further, it elaborated on the results of perception/production tasks in L2 and L3 in the context of individual differences measures and the predictions formulated in Chapter 4. The chapter concluded with a discussion of the limitations of the study and future research directions.

Conclusion

The aim of the present dissertation was to investigate the role of individual differences in the multilingual acquisition of phonology by adolescent sequential multilinguals (L1 – Polish, L2 – English, L3 – German), acquiring their L2 and L3 in the context of formal instruction in primary school. A battery of tasks implemented in three testing sessions over 10 months examined perception and production, inhibitory control, phonological working memory and language proficiency, with an additional test of psychotypology and the language history and use questionnaires.

The multilingual development was assessed in three testing sessions through two types of tasks administered in L2 English and L3 German: a perception ABX discrimination task and a production delayed repetition task. The tasks focused on such features as close front vowels duration as well as VOT duration in voiceless plosives in both L2 and L3. The results of the perception tasks in three testing sessions revealed a relative stability, with greater accuracy observed for L3. The production of vowels and VOT was characterised by a slight change over time, largely on the basis of individual features. The overall accuracy of L2 and L3 production reached comparable accuracy levels. No significant correlations between the perception and production in the respective languages were revealed. Overall, production was more accurate than perception in both the second and third languages. As far as within-language comparisons over time are concerned, production preceded perception in terms of accuracy in both L2 and L3, however, the development of the two modalities over time did not correlate significantly. It might be, therefore, concluded that the multilingual development of the participants was characterised by stability. Some instances of significant differences were found only on the basis of individual features.

The individual differences measures administered in two testing sessions (T1 and T3) included: a flanker task investigating inhibitory control, a phonological working memory task based on pseudoword repetition, a visual psychotypology measure, language history and use questionnaires and proficiency interviews. The scores and data obtained through these measures were subsequently compared against the perception-production results in L2 and L3 to uncover any potential link between individual differences and multilingual development.

A part of the data was analysed through multiple linear regression models intending to verify whether selected individual differences measures (i.e. inhibitory control score, phonological working memory score and proficiency scores) can predict the accuracy in perception and production of L2 and L3. The models did not reveal any significant relationships between the individual differences measures and speech perception and production measures. Further, the data obtained through the remaining individual differences measures was compared against the perception-production data in L2 and L3. Psychotypology data, aligned with the actual phonological typology, might have led to the presence of L2-driven CLI in the case of some of the features in L3. The participants' language exposure and use were characterised by the dominance of L1, which might have substantially influenced the low-to-moderate perception and production accuracy, which affected the pace and route of multilingual development.

The characteristics of the L1-dominated setting and the resulting proficiency of the participants, shaped predominantly by the formal instruction, did not allow for the observation of significant relationships between the cognitive measures of individual differences such as inhibitory control and phonological working memory. It might be argued that the language history and use questionnaire was the most significant predictor of multilingual development of adolescent multilinguals, predicting largely L1-driven perception and production and the small or non-existent development over time, affected by the scarce L2 and L3 exposure and use.

Despite the lack of significant relationships between the great majority of the individual differences measures and the slow or stable pace of multilingual development, the current study constitutes a meaningful contribution to the field of multilingual phonology, especially in the context of formal language acquisition in L1-dominated settings and studies with adolescent participants. Future research is needed to continue the investigation of individual differences in multilingual development in the said context, possibly through a more extended span of testing combined with separate sessions for each of the examined languages. Moreover, future research might benefit from the investigation of such factors as the quality of foreign language input received during formal instruction.

Abstract

The aim of this dissertation was to explore the role of individual differences in the process of multilingual acquisition of phonology by adolescents. Individual differences in the acquisition of phonology constitute a growing area within multilingual research. The limited number of studies as well as their divergent results motivate further empirical exploration (Cabrelli Amaro and Wrembel 2019). The following thesis intended to integrate and extend the approaches adopted in previous multilingual phonology research by investigation of more than one factor of individual differences and repeated testing within the framework of a longitudinal study.

The participants of the study featured in the doctoral dissertation were primary school students aged 12-13 with Polish as the first language (L1), learning English (L2) and starting to learn German (L3). The battery of tests in three testing sessions included the tasks investigating phonological perception and production in L2 and L3 and measures of individual difference factors (inhibition control, phonological memory, psychotypology). In addition, it also featured a language proficiency measure and a questionnaire of language history and use. Such a design of the test battery intended to provide the answers to the main research questions posed in the present dissertation: (1) What is the phonological development in the second and third language throughout three testing sessions? (2) What is the role of individual differences in phonological development in the second and third language?

Phonological development was examined in the ABX task (measuring perception) and the delayed repetition task (measuring production) in three testing sessions. The tasks focused on the duration of the selected front vowels and the VOT parameter in voiceless stops in the L2 and L3. The results of the perceptual tasks showed an overall stability in two languages and a greater accuracy in recognizing vowel contrasts in the L3 compared to L2. The production of vowels and VOT in voiceless stops was characterized by comparable accuracy in the L2 and L3 and a slight change throughout the testing sessions, mainly in the case of individual features. Comparisons over time showed that production preceded perception in terms of accuracy in both L2 and L3. Moreover, the development of these two modalities over time did not correlate significantly.

Measures of individual differences in the T1 and T3 testing sessions included: flanker task (inhibition control), pseudoword repetition (phonological memory), psychotypology task, questionnaires of language history and use and a measure of language proficiency in the L2 and L3. The results obtained in the flanker task, the pseudoword task and the language proficiency measure were compared with the results of the perception and production tasks by means of linear regression models. The models did not reveal any significant relationships between the measures of individual differences and the accuracy of perception and production in the L2 and L3. The results obtained through the remaining measures of individual differences were compared with the results of perception and production in the L2 and L3. It has been shown that psychotypology, consistent with the actual phonological typology, might be linked to the presence of cross-linguistic influence. Moreover, the questionnaires demonstrated a strong dominance of the first language in the environment and the daily communication of the participants, which could have had a significant impact on the low to moderate accuracy of perception and production in the L2 and L3 and the overall pace of multilingual development.

Despite the lack of significant relationship between the majority of measures of individual differences and the no substantial changes in multilingual development over time, this study makes an important contribution to the field of multilingual phonology, especially in the context of multilingualism in formal language instruction. The present study, with its constraints and limitations, may be helpful in constructing the methodology of subsequent studies in the aforementioned context by informing the researches about the potential issues associated with the pace of multilingual development in formal instruction contexts and the planning of the testing session.

Streszczenie

Poniższa praca doktorska podejmuje zagadnienie różnic indywidualnych w przyswajaniu fonologii z perspektywy wielojęzyczności u młodzieży. Różnice indywidualne w przyswajaniu fonologii stanowią rozwijającą się dziedzinę badań nad wielojęzycznością. Zarówno ograniczona liczba badań jak i ich często rozbieżne wyniki stanowią motywację do dalszej empirycznej eksploracji (Cabrelli Amaro i Wrembel 2019). Poniższa praca stanowi próbę integracji i rozszerzenia podejść przyjętych we wcześniejszych badaniach z zakresu wielojęzycznej fonologii przez zbadanie więcej niż jednego czynnika różnic indywidualnych i wielokrotne testowanie w formule badania podłużnego.

Uczestnikami badania zawartego w pracy doktorskiej byli uczniowie szkoły podstawowej w wieku 12-13 lat, posługujący się językiem polskim jako pierwszym (L1), kontynuujący naukę języka angielskiego (L2) i rozpoczynający naukę języka niemieckiego (L3). Bateria testów w trzech turach testowych objęła zadania percepcji i produkcji fonologicznej w L2 i L3 oraz miary czynników różnic indywidualnych (kontrola hamowania, pamięć fonologiczna, psychotypologia). Ponadto, zawierała też zadanie mierzące biegłość językową i kwestionariusz dotyczący nauki języków i ich użytkowania. Tak zaprojektowana bateria testowa miała na celu pozyskanie danych, które umożliwią udzielenie odpowiedzi na dwa główne pytania badawcze podstawione w pracy doktorskiej: (1) jak przebiega rozwój fonologiczny w drugim i trzecim języku w trzech rozłożonych w czasie sesjach testowych? (2) jaka jest rola różnic indywidualnych w rozwoju fonologicznym w drugim i trzecim języku?

Rozwój fonologiczny został zbadany poprzez zadanie ABX (mierzące percepcję) oraz zadanie opóźnionego powtarzania (mierzące produkcję) w trzech sesjach testowych. Zadania koncentrowały się na długości wybranych samogłosek przymkniętych przednich i parametrze VOT w bezdźwięcznych spółgłoskach zwartych w L2 i L3. Wyniki zadań percepcyjnych wykazały ogólną stabilność i większą trafność rozpoznawania kontrastów samogłoskowych w L3. Produkcja samogłosek i VOT w bezdźwięcznych spółgłoskach zwartych charakteryzowała się porównywalną poprawnością w L2 i L3 oraz nieznaczną zmianą w czasie, głównie w przypadku pojedynczych testowanych głosek. Porównania w czasie wykazały, że produkcja poprzedzała percepcję pod względem poprawności

zarówno w L2, jak i L3, jednak rozwój tych dwóch modalności w czasie nie korelował znacząco.

Pomiary różnic indywidualnych zastosowane w pierwszej i trzeciej sesji testowej objęły: zadanie flankerów (kontrolę hamowania), powtarzanie pseudosłów (pamięć fonologiczna), pomiar psychotypologii, kwestionariusze historii nauki i użycia języków oraz ocenę biegłości językowej w L2 i L3. Wyniki uzyskane w zadaniu flankerów, teście pseudosłów i zadaniu biegłości językowej zostały zestawione z wynikami zadań percepcji i produkcji w modelach regresji liniowej. Modele nie wykazały żadnych istotnych związków między miarami różnic indywidualnych, a poprawnością percepcji i produkcji w żadnym z języków. Pozostałe dane z miar różnic indywidualnych porównano z wynikami percepcji i produkcji w L2 i L3. Wykazano, że psychotypologia zgodna z rzeczywistą typologią fonologiczną mogła prowadzić do występowania wpływów międzyjęzykowych. Ponadto, kwestionariusze wykazały silną dominacja języka pierwszego w środowisku i komunikacji uczestników badania, co mogło mieć istotny wpływ na niską lub umiarkowaną poprawność percepcji i produkcji w dwóch językach oraz tempo rozwoju wielojęzyczności.

Pomimo braku znaczących związków między zdecydowaną większością miar różnic indywidualnych i powolnym lub stabilnym tempem rozwoju wielojęzyczności, niniejsze badanie stanowi istotny wkład w dziedzinę fonologii wielojęzycznej, szczególnie w odniesieniu do przyswajania drugiego i trzeciego języka w ramach edukacji formalnej, w kontekście zdominowanym przez język pierwszy. Obecne badanie, choć ograniczone przez względy logistyczne, może pomóc w konstruowaniu metodologii kolejnych badań we wspomnianym kontekście językowym i przy odpowiednim dostosowaniu długości testowania do tempa rozwoju wielojęzyczności.

References

Adams, Anne-Marie, and Susan E. Gathercole. 1995. “Phonological working memory and speech production in preschool children”, *Journal of Speech & Hearing Research* 38, 2: 403-14. (doi:10.1044/jshr.3802.403).

Adams, Anne-Marie. 1996. “Phonological working memory and spoken language development in young children”, *The Quarterly Journal of Experimental Psychology Section A* 49, 1: 216-33. (doi:10.1080/713755610).

Ahukanna, Joshua G.W., Nancy J. Lund and J. Ronald Gentile. 1981. “Inter- and Intra-Lingual Interference Effects In Learning a Third Language”, *The Modern Language Journal* 65, 3: 281-287. (doi:10.1111/j.1540-4781.1981.tb00985.x).

Aïte, Ania, Mathieu Cassotti, Adriano Linzarini, Anaïs Osmont, Olivier Houdé, and Grégoire Borst. 2018. “Adolescents’ Inhibitory Control: Keep It Cool or Lose Control“, *Developmental Science 21,* 1: e12491. (doi:10.1111/desc.12491).

Aliaga-García, Cristina, Joan C. Mora, Eva Cerviño-Povedano. 2011. “L2 speech learning in adulthood and phonological short-term memory”, *Poznań Studies in Contemporary Linguistics* 47: 1-c15. (doi:10.2478/psicl-2011-0002).

Archibald, Lisa M. D. and Susan E. Gathercole. 2006. “Short-term and working memory in specific language impairment”, *International Journal of Language & Communication Disorders* 41, 6: 675-93. (doi:10.1080/13682820500442602).

Aronin, Larissa. 2018. “Lecture 1: What is multilingualism”, in: David Singleton and Larissa Aronin (eds.), *Twelve lectures on multilingualism.* Clevendon: Multilingual Matters, 3-35.

Baddeley, Alan D. and Graham Hitch. 1974. “Working Memory”, in: Gordon H. Bower (ed.) *Psychology of Learning and Motivation* 8. Academic Press, 47-89.

Baddeley, Alan, Susan Gathercole, and Costanza Papagno. 1998. “The Phonological Loop as a Language Learning Device. ”, *Psychological Review* 105, 1: 158-73. (doi:10.1037/0033-295X.105.1.158).

Baddeley, Alan. 2003. “Working memory and language: an overview”, *Journal of Communication Disorders* 36, 189-208.

Baken, Ronald. J. and Robert F. Orlikoff. 2000. *Clinical measurement of speech and voice*. (2nd edition). San Diego: Singular Thomson Learning.

Baker Smemoe, Wendy, Pavel Trofimovich, James Flege, Molly Mack, and Randall Halter. 2008. “Child - adult differences in second-language phonological learning: The role of cross-language similarity”, *Language and Speech* 51: 317-42. (doi:10.1177/0023830908099068).

Balas, Anna. 2018. *Non-native vowel perception: The interplay of categories and features.* Poznań: Wydawnictwo Naukowe UAM.

Balas, Anna, Romana Kopečková and Magdalena Wrembel. 2019. “Perception of rhotics by multilingual children”, in: in: Sasha Calhoun, Paola Escudero, Marija Tabain and Paul Warren (eds.), *Proceedings of the 19th International Congress of Phonetic Sciences*. Canberra: Australasian Speech Science and Technology Association Inc, 3725-3729.

Bardel, Camilla and Falk Ylva. 2007. “The role of the L2 in L3 acquisition: The case of Germanic syntax”, *Second Language Research* 23, 4: 459-484. (doi: 10.1177/0267658307080557).

Bardel, Camilla and Ylva Falk. 2012. “The L2 status factor and the declarative/procedural distinction”, in: Jennifer Cabrelli Amaro, Suzanne Flynn and Jason Rothman (eds.), *Third language acquisition in adulthood*. Philadelphia/Amsterdam: John Benjamins Publishing Company, 61-78.

Bhatia, Tej K. and William C. Ritchie. 2013. “Introduction”, in: Bhatia, Tej K. and William C. Ritchie (eds.), Handbook of bilingualism and multilingualism. Malden: Blackwell Publishing, i-xxiii.

Best, Catherine. 1995. “A direct realist view of cross-language speech perception”, in: Winifred Strange (ed.), *Speech perception and linguistic experience: Issues in cross-language research*. Timonium, MD: York Press, 171-204.

Best, Catherine and Michael Tyler. 2007. “Non-native and second language speech perception”, in: Ocke-Schwen Bohn and Murray J. Munro (eds.), *Language experience in second language speech learning*. Amsterdam: John Benjamins, 15-34.

Berthele, Raphael. 2021. “The Extraordinary Ordinary: Re-engineering multilingualism as a natural category”, *Language Learning* 71: 80-120. (doi: 10.1111/lang.12407).

Bialystok, Ellen, Michelle M. Martin and Mythili Viswanathan. 2005. “Bilingualism across the lifespan: The rise and fall of inhibitory control.” *International Journal of Bilingualism* 9, 1: 103-19. (doi:10.1177/13670069050090010701).

Bishop, Dorothy V. M. 2002. “The role of genes in the etiology of specific language impairment”, *Journal of Communication Disorders* 35, 4: 311-28. (doi:10.1016/s0021-9924(02)00087-4).

Blumenfeld, Henrike K. and Viorica Marian. 2011. “Bilingualism influences inhibitory control in auditory comprehension”, *Cognition* 118, 2: 245-57. (doi:10.1016/j.cognition.2010.10.012).

Boersma, Paul and David Weenink. 1992-2022. Praat: doing phonetics by computer. Version 6.2.14. (http://www.praat.org/) (date of access: 24 May 2022).

Bogacka (Balas), Anna. 2004. “On the perception of English high vowels by Polish learners of English.”, in: Evangelia Daskalaki, Napoleon Katsos, Marios Mavrogiorgos and Matthew Reeve (eds.), CamLing 2004: Proceedings of the University of Cambridge Second Postgraduate Conference in Language Research. Cambridge: Cambridge University Press, 43-50.

Bono, Mariana. "Chapter 3: Crosslinguistic interaction and metalinguistic awareness in third language acquisition, in: Gessica De Angelis and Jean-Marc Dewaele (eds.), *New trends in crosslinguistic influence and multilingualism research.* Clevedon: Multilingual Matters, 25-52.

Boase-Beier, Jean and Ken Lodge. 2003. *The German language: A linguistic introduction.* Oxford: Blackwell.

Boogert, Neeltje J., Joah R. Madden, Julie Morand-Ferro and Alex Thornton. 2018. “Measuring and understanding individual differences in cognition”. *Philosophical Transactions of the Royal Society B: Biological Sciences* 373, 1756: 20170280. (https://doi.org/10.1098/rstb.2017.0280).

Borragan, Maria, Angela de Bruin, Clara D Martin, Jon Andoni Duñabeitia. 2018. “Exploring different types of inhibition during bilingual language production”, *Frontiers in Psychology* 9: 1-11. (doi:10.3389/fpsyg.2018.02256).

Braunschweiler, Norbert. 1997. “Integrated cues of voicing and vowel length in German: A production study”, *Language and Speech* 40, 4:353-376.

Browne, Wayles and Vyacheslav Ivanov Vsevolodovich. 2022. “Slavic languages”, in: *Encyclopedia Britannica*. (https://www.britannica.com/topic/Slavic-languages). (date of access: 2 June 2022).

Bunge, Silvia A. and Samantha B. Wright. 2007. “Neurodevelopmental changes in working memory and cognitive control”, *Current Opinion in Neurobiology, Cognitive neuroscience* 17, 2: 243-50. (doi:10.1016/j.conb.2007.02.005).

Butler, Yuko Goto. 2012. “Bilingualism/Multilingualism and Second-Language Acquisition”, in: Bhatia, Tej K. and William C. Ritchie (eds.), *Handbook of bilingualism and multilingualism.* Malden: Blackwell Publishing, 109-136.

Cabrelli Amaro, Jennifer. 2013. Phonological permeability hypothesis: Measuring regressive L3 influence to test L1 and L2 phonological representations. [Unpublished Ph.D. dissertation, University of Florida.]

Cabrelli Amaro, Jennifer and Magdalena Wrembel (eds). 2019. *Advances in the investigation of L3 phonological acquisition*. New York: Routledge.

CEFR. 2001. Common European Framework of Reference for Languages: Learning, Teaching, Assessment. Council of Europe, Education and Languages, (https://rm.coe.int/1680459f97) (date of access: 9 May 2022).

Cenoz, Jasone. 2001. “The effect of linguistic distance, L2 status and age on cross-linguistic influence in third language acquisition”, in: Jasone Cenoz, Britta Hufeisen and Ulrike Jessner (eds.), Cross-linguistic influence in third language acquisition: Psycholinguistic perspectives. Clevedon: Multilingual Matters, 8-20.

Collins, Beverley and Inger M. Mees. 2013. *Practical Phonetics and Phonology: A Resource Book for Students*. (3rd Edition).London & New York: Routledge.

Colman, Andrew M. 2015. A Dictionary of Psychology. Oxford: Oxford University Press.

Constantinidis, Christos, and Beatriz Luna. 2019. “Neural substrates of inhibitory control maturation in adolescence”, *Trends in Neurosciences* 42, 9: 604-616. (doi:10.1016/j.tins.2019.07.004).

Coretta, Stefano. 2019. “An exploratory study of voicing-related differences in vowel duration as compensatory temporal adjustment in Italian and Polish”, *Glossa: a journal of general linguistics* 4, 1: 1-25. (doi:10.5334/gjgl.869).

Costa, Albert and Mikel Santesteban. 2004. “Lexical access in bilingual speech production: evidence from language switching in highly proficient bilinguals and L2 learners”, *Journal of Memory and Language* 50: 491-511. (doi: 10.1016/j.jml.2004.02.002).

Costa, Albert, Mikel Santesteban and Iva Ivanova. 2006. “How do highly proficient bilinguals control their lexicalization process? Inhibitory and language-specific selection mechanisms are both functional”, *Journal of Experimental Psychology: Learning Memory and Cognition* 32: 1057-1074. (doi: 10.1037/0278-7393.32.5.1057).

Cromer, Jason A., Adrian J. Schembri, Brian T. Harel and Paul Maruff. 2015. “The nature and rate of cognitive maturation from late childhood to adulthood”, *Frontiers in Psychology* 6. (doi:10.3389/fpsyg.2015.00704).

Crowther, Dustin, Pavel Trofimovich, Talia Isaacs and Kazuya Saito. 2015. “Does a speaking task affect second languwiage comprehensibility?”, *Modern Language Journal* 99, 1: 80-95. (doi:10.1111/modl.12185).

Cruttenden, Alan. (2014). *Gimson's Pronunciation of English* (8th ed.). Oxon: Routledge. (https://doi.org/10.4324/9780203784969).

Crystal, David. 2008. “Two Thousand Million?”, *English Today* 24, 1: 3-6. (doi:10.1017/S0266078408000023).

Darcy, Isabelle and Joan C. Mora. 2013. Inhibition and phonological processing in a second language. (Paper presented at the conference of European Second Language Association, Amsterdam, 28-31 Aug. 2013).

Darcy, Isabelle, Joan C. Mora & Danielle Daidone. 2016. “Attention control and inhibition influence phonological development in a second language”, *Proceedings of the International Symposium on the Acquisition of Second Language Speech, Concordia Working Papers in Applied Linguistics* 5: 115-129.

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. (doi:10.1016/j.lindif.2015.04.005).

Darcy, Isabelle, Joan C. Mora. and Danielle Daidone. 2016. “The role of inhibitory control in second language phonological processing”, *Language Learning* 66: 741-733. (doi: 10.1111/lang.12161).

Delcenserie, Audrey, Fred Genesee, Natacha Trudeau, and François Champoux. 2021. “The development of phonological memory and language: A multiple groups spproach”, *Journal of Child Language*, 1-40. (doi:10.1017/S0305000920000343).

Desjardins, Jamie L., and Francisco Fernandez. 2018. “Performance on auditory and visual tasks of inhibition in English monolingual and Spanish-English bilingual adults: Do bilinguals have a cognitive advantage? ”, *Journal of Speech, Language, and Hearing Research* 61, 2: 410-419. (doi.org:10.1044/2017_JSLHR-H-17-0160).

Diamond, Adele. 2013. “Executive functions”, *Annual Review of Psychology* 64: 135-68: 135-168. (doi:/10.1146/annurev-psych-113011-143750).

Docherty, Gerard J. 1992. *The Timing of Voicing in British English Obstruents*. New York: De Gruyter Mouton.

Dryer, Matthew S. and Martin Haspelmath (eds.). 2013. The world atlas of language structures online. Munich: Max Planck Digital Library. (http://wals.info) (date of access: 29 May 2022).

De Angelis, Gessica and Larry Selinker. 2001. “Interlanguage transfer and competing linguistic systems”, in: Jasone Cenoz, Britta Hufeisen and Ulrike Jessner (eds.), *Cross-linguistic influence in third language acquisition: Psycholinguistic perspectives*. Clevedon: Multilingual Matters, 42-58.

De Angelis, Gessica. 2005. “Interlanguage transfer of function words”, *Language Learning* 55, 3: 379-414. (doi:10.1111/j.0023-8333.2005.00310.x).

De Angelis, Gessica. 2007. Third or additional language acquisition. Clevendon: Multilingual Matters Ltd.

de Bot, Kees. 1992. “A bilingual production model: Levelt’s speaking model adapted”, Applied Linguistics 13, 1: 1-24. (doi:10.1093/applin/13.1.1).

de Bot, Kees. 2014. “The effectiveness of early foreign language learning in the Netherlands”, *Studies in Second Language Learning and Teaching* 4, 3: 409-418. (doi:10.14746/ssllt.2014.4.3.2).

Dewaele, Jean-Marc. 1998. “Lexical inventions: French interlanguage as L2 versus L3”, *Applied Linguistics* 19, 4: 471-490. (doi:10.1093/applin/19.4.471).

Dewaele, Jean-Marc. 2009. “Individual differences in second language acquisition”, in: William C. Ritchie and Tej K. Bhatia (eds.), *The new handbook of second language acquisition*. Bingley: Emerald Group Publishing Limited, 623-646.

De Wilde, Vanessa, Marc Brysbaert, and June Eyckmans. 2020. “Learning English through out-of-school exposure. Which levels of language proficiency are

attained and which types of input are important?" *Bilingualism: Language and Cognition* 23, 1: 171-85. (doi:10.1017/S1366728918001062).

Durrell, Martin. 2003. *Using German: A guide to contemporary usage*. Cambridge: Cambridge University Press.

Dziubalska-Kołaczyk, Katarzyna and Wrembel, Magdalena. 2022. "Natural Growth Theory of Acquisition (NGTA): Evidence from (mor)phonotactics", in: Sardegna, Veronica and Anna Jarosz (eds): *Theoretical and practical perspectives on English pronunciation teaching and research*. Cham: Springer.

Eriksen, Barbara. A. and Eriksen, Charles. W. 1974. "Effects of noise letters upon identification of a target letter in a non- search task", Perception and Psychophysics 16, 1: 143-149 (doi: 10.3758/BF03203267).

Fabbro, F. 2002. "The neurolinguistics of L2 users", in: Vivian J. Cook (ed.), *Portraits of the L2 user*. Clevedon: Multilingual Matters, 199-218.

Falk, Ylva, and Camilla Bardel. 2011. "Object pronouns in German L3 syntax: Evidence for the L2 Status Factor." *Second Language Research* 27, 1: 59-82, (doi:10.1177/0267658310386647).

Falk, Ylva., Christina Lindqvst and Camilla Bardel. 2015. "The role of L1 explicit metalinguistic knowledge in L3 oral production at the initial state", *Bilingualism: Language and Cognition* 18, 2: 227-235. (doi:10.1017/S1366728913000552).

Festman, Julia. 2008. "Cross-Language Interference during Trilingual Picture-Naming in Single and Mixed Language Conditions", in: Martha Gibson, Britta Hufeisen & Cornelia Personne (eds.), *Mehrsprachigkeit: lernen und lehren / Multilingualism: Learning and Instruction. Selected papers from the L3 conference in Freiburg, Switzerland*. Baltmannsweiler: Schneider Verlag Hohengehren, 110-119.

Festman, Julia. 2018. "Lecture 9: The psycholinguistics of multilingualism", in: David Singleton and Larissa Aronin (eds.), *Twelve Lectures on Multilingualism.* Clevendon: Multilingual Matters, 233-271.

Fiez, Julie A. 2016. "Chapter 68 - Neural Basis of Phonological Short-Term Memory.", in: Gregory Hickok and Steven L. Small (eds.), *Neurobiology of Language*. San Diego: Academic Press. 855-862.

Flege, James E. 1995. "Second-language Speech Learning: Theory, Findings, and Problems", in: Winifred Strange (ed.), *Speech Perception and Linguistic*

Experience: Issues in Cross-language research. Timonium: York Press, 229-273.

Flege, James E. and Ian MacKay. 2011. "What accounts for "Age" effects on overall degree of foreign accent? ", in: Magdalena Wrembel, Małgorzata Kul and Katarzyna Dziubalska-Kołaczyk (eds.), *Achievements and perspectives in the acquisition of second language speech*: *New Sounds 2010*. Vol. 2. Bern, Switzerland: Peter Lang.

Flege, James E. and Ocke-Schwen Bohn. 2021. "The Revised Speech Learning Model (SLM-r)", in: Ratree Wayland (ed.), *Second language speech learning: Theoretical and empirical progress*. Cambridge: Cambridge University Press. 3-83.

Flynn, Suzanne, Claire Foley and Inna Vinnitskaya. 2004. "The Cumulative Enhancement Model for language acquisition: Comparing adults' and children's patterns of development in L1, L2 and L3 acquisition of relative clauses". *The International Journal of Multilingualism* 1, 1: 3-16. (doi:10.1080/14790710408668175).

"Formal, non-formal and informal learning". 2022. (https://www.coe.int/en/web/lang-migrants/formal-non-formal-and-informal-learning) (date of access: 14 May 2022).

Fosco, Whitney D., Larry W. Hawk, Craig R. Colder, Samuel N. Meisel, and Liliana J. Lengua. 2019. "The development of inhibitory control in adolescence and prospective relations with delinquency", *Journal of Adolescence* 76: 37-47. (doi:10.1016/j.adolescence.2019.08.008).

Freundlich, Jerzy. 2016. "Cross-linguistic influence in third-language acquisition: Learning Mandarin Chinese (L3) through the medium of English (L2)", *Państwo i społeczeństwo* 3:11-32.

Fricke, Melinda, Judith F. Kroll and Paola E. Dussias. 2016. "Phonetic variation in bilingual speech: A lens for studying the production-comprehension link.", *Journal of Memory and Language* 89: 110-137. (doi:10.1016/j.jml.2015.10.001).

Friedman, Naomi. P., and Akira Miyake. 2004. "The relations among inhibition and interference control functions: A latent-variable analysis.", *Journal of Experimental Psychology: General* 133, 1: 101-135. (10.1037/0096-3445.133.1.101).

García, Ofelia and Li Wei. 2014. *Translanguaging. Language, bilingualism and education*. London: Palgrave Pivot.

Gathercole, Susan E. and Alan D. Baddeley. 1993. *Working memory and language*. Hillsdale: Lawrence Erlbaum Associates Inc.

Golin, Christina. 2018. Speech perception development in child vs. adult L3 learners. (Paper presented at Young Linguists' Meeting, Poznań 2018, 23-25 Nov. 2018).

Golin, Christina, Romana Kopečková, Ulrike Gut, Wander Lowie. A case study of adult multilingual phonological development in the initial stages of L3 learning. (Paper presented at New Sounds: The 9th International Symposium on the Acquisition of Second Language Speech, 30 Aug.-1 Sept. 2019).

Green, David. 1986. "Control, activation, and resource: A framework and a model for the control of speech in bilinguals", *Brain and Language* 27: 210-223.

Green, David. 1998. "Mental control of the bilingual lexico-semantic system", *Bilingualism: Language and Cognition* 1: 67-81.

Green, David. 1993. "Towards a model of L2 comprehension and production", in: Robert Schreuder & Bert Weltens (eds.), *The bilingual lexicon*. Amsterdam: John Benjamins, 249-277.

Green, David and Li Wei. 2014. "A control process model of code-switching", *Language, Cognition and Neuroscience* 29, 4: 499-511.

Girbau, Dolors, and Richard G. Schwartz. 2007. "Phonological working memory in Spanish-English bilingual children with and without specific language impairment", *Journal of Communication Disorders* 41, 2: 124-145. (doi:10.1016/j.jcomdis.2007.07.001).

Grosjean, François. 1998. "Studying bilinguals: Methodological and conceptual issues". *Bilingualism: Language and Cognition* 1: 131-149. (doi: doi:10.1017/S136672899800025X)

Gussmann, Edmund. 2007. *The phonology of Polish*. Oxford: Oxford University Press.

Gut, Ulrike. 2006. "Unstressed vowels in non-native German", *Proceedings of Speech Prosody*. Dresden.

Gut, Ulrike. 2010. "Cross-linguistic in fluence in L3 phonological acquisition", *International Journal of Multilingualism* 7, 1: 19-38. (doi:10.1080/14790710902972248)

Gut, Ulrike, Robert Fuchs and Eva-Maria Wunder (eds.). 2015. *Universal or diverse paths to English phonology*. Berlin: Mouton De Gruyter.

Hall, Christopher J., Denise Newbrand, Peter Ecke, Ulrike Sperr, Vanessa Marchand. and Lisa Hayes. 2009. “Learners' implicit assumptions about syntactic frames in new l3 words: The role of cognates typological proximity, and L2 status”, *Language Learning*, 59: 153-202. (doi:10.1111/j.1467-9922.2009.00503.x).

Hammarberg, Björn and Britta Hammarberg. 1993. “Articulatory re-setting in the acquisition of new languages”, *Phonum* 2: 61-67.

Hammarberg, Björn. 2001. “Roles of L1 and L2 in L3 production and acquisition”, in: Jasone Cenoz, Britta Hufeisen and Ulrike Jessner (eds.), Cross-linguistic influence in third language acquisition: Psycholinguistic perspectives. Clevedon: Multilingual Matters, 21-41.

Hammarberg, Björn and Britta Hammarberg. 2005. “Re-setting the basis of articulation in the acquisition of new languages: A third-language case study”, in: Björn Hammarberg (ed.), *Processes in Third Language Acquisition*, Edinburgh: Edinburgh University Press, 2022, 74-85.

Hammarberg, Björn. 2009. *Processes in third language acquisition*. Edinburgh: Edinburgh University Press.

Hammarberg, Björn. 2018. “L3, the tertiary language”, in: Andreas Bonnet and Peter Siemund (eds.), *Foreign language education in multilingual classroom*. Amsterdam: John Benjamins Publishing Company, 127-150.

Hartshorne, Joshua K., Joshua B. Tenenbaum and Steven Pinker. 2018. “A critical period for second language acquisition: Evidence from 2/3 million English speakers”, *Cognition* 177: 263-277. (doi: 10.1016/j.cognition.2018.04.007).

Herdina, Philip and Ulrike Jessner. 2000. “The Dynamics of Third Language Acquisition”, in: Jasone Cenoz and Ulrike Jessner (eds.), *English in Europe: The acquisition of a third language*. Clevedon: Multilingual Matters, 84-98.

Herdina, Philip and Jessner, Ulrike. 2002. *A dynamic model of multilingualism: Perspectives of change in psycholinguistics*. Clevendon: Multilingual Matters.

Hoff, Erika. 2013. Language Development. Boston: Cengage Learning.

Hufeisen, Britta. 1998. “L3 - Stand der Forschung - Was bleibt zu tun? ”, in: Britta Hufeisen and Beate Lindemann (eds.), *Tertiärsprachen: Theorien, Modelle, Methoden*. Tübingen: Stauffenburg Verlag.

Hufeisen, Britta. 2000. “A European perspective - Tertiary languages with a focus on German as L3”, in: Judith W. Rosenthal (ed.), *Handbook of undergraduate*

second language education: English as a second language, bilingual, and foreign language instruction for a multilingual world. Mahwah: Erlbaum, 209-226.

Hufeisen, Britta. 2005. "Multilingualism: Linguistic models and related issues", in: Britta Hufeisen and Robert J. Fouser (eds.), *Introductory Readings in L3*. Tübingen: Stauffenburg Verlag 3-45.

Hufeisen, Britta and Marx Nicole (eds.). 2007. *EuroComGerm: Die Sieben Siebe. Germanische Sprachen lesen lernen*. Aachen: Shaker

Hurtado, Nereyda, Virginia A. Marchman, and Anne Fernald. 2008. "Does input influence uptake? Links between maternal talk, processing speed and vocabulary size in Spanish-learning children", *Developmental Science* 11, 6: F31-39. (doi.org/10.1111/j.1467-7687.2008.00768.x).

Hurtado, Nereyda, Theres Grüter, Virginia A. Marchman, And Anne Fernald. 2014. "Relative language exposure, processing efficiency and vocabulary in Spanish-English bilingual toddlers", *Bilingualism: Language and Cognition* 17, 1: 189-202. (doi:10.1017/S136672891300014X).

Imbir, Kamil, Tomasz Spustek and Jarosław Żygierewicz. 2015. "Polish pseudo-words list: dataset of 3023 stimuli with competent judges' ratings", *Frontiers in Psychology* 6. (doi: 10.3389/fpsyg.2015.01395).

Jacobs, April, Melinda Fricke, and Judith F. Kroll. 2016. "Cross-language activation begins during speech planning and extends into second language speech", *Language Learning* 66, 2: 324-53. (doi:10.1111/lang.12148).

James, Mark A. 2012. "Cross-linguistic influence and transfer of learning", in: Norbert M. Seel (ed.), *Encyclopedia of the sciences of learning.* New York: Springer US, 858-861.

Jarvis, Scott and Aneta Pavlenko. 2007. Crosslinguistic influence in language and cognition. New York/London: Routledge

Jassem, Wiktor. 2003. "Polish", *Journal of the International Phonetic Association* 33, 1: 103-107. (doi:10.1017/S0025100303001191).

Jessen, Michael and Catherine Ringen. 2002. "Laryngeal Features in German", *Phonology* 19, 2: 189-218. (doi:10.1017/S0952675702004311).

Jessner, Ulrike. 2006. *Linguistic awareness in multilinguals. English as a third language.* Edinburgh: Edinburgh University Press.

Jessner, Ulrike. 2008. “A DST Model of Multilingualism and the Role of Metalinguistic Awareness”, *The Modern Language Journal* 92: 270-283. (doi: 10.1111/j.1540-4781.2008.00718.x.).

Kadyamusuma, McLoddy, Eve Higby and Obler, Loraine. 2018. “Lecture 10: The Neurolinguistics of Multilingualism”, in: David Singleton and Larissa Aronin (eds.), *Twelve Lectures on Multilingualism.* Clevendon: Multilingual Matters, 271-296.

Keating, Patricia A., Michael J. Mikós and William F. 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. (doi:10.1121/1.387139).

Kellerman, Erik. 1977. “Towards a characterisation of the strategy of transfer in second language learning”, *Interlanguage Studies Bulletin* 2, 1: 58-145.

Kellerman, Eric. 1978. “Giving learners a break: Native language intuitions as a source of prediction about transferability”, *Working Papers on Bilingualism*, 15, 59-92.

Kellerman, Eric.1983. “Now you see it, now you don’t”, in: Susan Gass and Larry Selinker (eds.), *Language transfer in language learning*. Rowley: Newbury House, 112-134.

acquisition. I.S.B., 2.1, 58-146.

Kemp, Charlotte. 2009. “Defining multilingualism”, in: Larissa Aronin and Britta Hufeisen (eds.), *The exploration of multilingualism*. Amsterdam: John Benjamins Company, 11-26.

Kessinger, Rachel H. and Sheila E. Blumstein. 1998. “Effects of speaking rate on voice-onset time and vowel production: Some implications for perception studies”, *Journal of Phonetics* 26, 2: 117-28. (https://doi.org/10.1006/jpho.1997.0069).

Kleiner, Stefan, Ralf Knöbl, Dudenredaktion. 2015. *Duden, das Aussprachewörterbuch*. Dudenverlag.

Kopečková, Romana. 2014. “Crosslinguistic influence in instructed L3 child phonological Acquisition”, in: Mirosław Pawlak and Larissa Aronin (eds), *Essential topics in applied linguistics and multilingualism*: *Second language learning and teaching*. Cham: Springer, 205-224.

Kopečková, Romana. 2015. “Differences in the perception of English vowel sounds by child L2 and L3 learners”, in: Ulrike Gut, Robert Fuchs and Eva-Maria Wunder.

(eds.), *Universal or Diverse Paths to English Phonology?*. Berlin: Mouton De Gruyter, 71-89.

Kopečková, Romana, Marta Marecka, Magdalena Wrembel and Ulrike Gut. 2016. „Interactions between three phonological subsystems of young multilinguals: The influence of language status", *International Journal of Multilingualism* 13, 4: 426-443 (doi:10.1080/14790718.2016.1217603).

Kopečková, Romana. 2016. "The bilingual advantage in L3 learning: A developmental study of rhotic sounds", *International Journal of Multilingualism* 13, 4: 410-425. (doi:10.1080/14790718.2016.1217605).

Kopečková, Romana. 2018. "Exploring metalinguistic awareness in L3 phonological acquisition: The case of young instructed learners of Spanish in Germany", *Language Awareness* 27:153-166. (doi:10.1080/09658416.2018.1432629).

Kopečková, Romana, Christine Dimroth and Ulrike Gut. 2019. "Children's and adults' initial phonological acquisition of a foreign language", *Journal of Second Language Pronunciation* 5, 3: 374-401. (doi:10.1075/jslp.18033.kop).

Kopczyński, Andrzej. 1977. *Polish and American English consonant phonemes: A contrastive study*. Warszawa: PWN.

Köhler, Klaus. 1995. *Einfuhrung in die Phonetik des Deutschen*. (2nd edition). Berlin: Schmidt.

Köhler, Klaus. 1999. *German*, in: *Handbook of the International Phonetic Association*, Cambridge: Cambridge University Press, 86-89.

Künzel, Hermann. 1977. *Signalphonetische untersuchungen deutsch-französischer interferen-zen im bereich der okklusive*. Frankfurt am Main: Peter Lang.

Krasowicz-Kupis, Grażyna and Grela-Goryczka, Renata. 2003. "Badanie pamięci fonologicznej z zastosowaniem Zetotestu" [The assessement of phonological memory through Zetotest], *Annales Universitatis Mariae Curie-Skłodowska, Paedagogia-Psychologia* 16: 58-68.

Krzysik, Iga. 2020. "The relationship between inhibitory control and speech production in young multilinguals", *Yearbook of the Poznan Linguistic Meeting* 6, 1: 59-79. (doi:10.2478/yplm-2020-0006).

Krzysik, Iga and Magdalena Wrembel. 2019. "The relationship between phonological working memory and speech production in young multilinguals", in: Sasha Calhoun, Paola Escudero, Marija Tabain and Paul Warren (eds.), *Proceedings of the*

19th International Congress of Phonetic Sciences. Canberra: Australasian Speech Science and Technology Association Inc, 3730-3734.

Leclercq, Anne-Lise and Steve Majerus. 2010. "Serial-order short-term memory predicts vocabulary development: Evidence from a longitudinal study", *Developmental Psychology* 46, 2: 417-427. (doi:10.1037/a0018540).

Leung, Yan Kit Ingrid. 2005. "L2 vs. L3 initial state: A comparative study of the acquisition o French DPs by Vietnamese monolinguals and Cantonese-English bilinguals". *Bilingualism: Language and Cognition* 8: 39-61. (doi:10.1017/S1366728904002044).

Lein, Tatjana, Tanja Kupisch and Joost van de Weijer. 2016. "Voice Onset Time and global foreign accent in German-French simultaneous bilinguals during adulthood", *International Journal of Bilingualism* 20, 6: 732-49. (doi:10.1177/1367006915589424).

Lenneberg, Eric H. 1967. *Biological Foundations of Language*. New York: Wiley.

Li, Ping, Sara Sepanski and Xiaowei Zhao. 2006. "Language history questionnaire: A web-based interface for bilingual research", *Behavior Research Methods* 38, 2: 202-10. (doi.org/10.3758/BF03192770).

Lindqvist, Christina. 2015. Do learners transfer from the language they perceive as most closely related to the L3? The role of psychotypology for lexical and grammatical crosslinguistic influence in French L3, in: Marijana Kresić, Ulrike Jessner and Gessica De Angelis (eds.), *Crosslinguistic influence and crosslinguistic interaction in multilingual language learning*. London, New York: Bloomsbury, 231-151.

Lindqvist, Christina. 2019. "Didactic challenges in the multilingual classroom. The case of French as a foreign language", in: María Juncal Gutierrez-Mangado, María Martínez Adrián and Francisco Gallardo Del Puerto (eds.), *Cross-linguistic influence: From empirical evidence to classroom practice.* Cham: Springer International Publishing Switzerland, 87-99.

Lisker, Leigh and Arthur S. Abramson. 1964. "A cross-language study of voicing in initial stops: Acoustical measurements, WORD 20, 3: 384-422. (doi:10.1080/00437956.1964.11659830).

Liu, Cong, Chin-Lung Yang, Lu Jiao, John W. Schwieter, Xun Sun, and Ruiming Wang. 2019. 'Training in language switching facilitates bilinguals' monitoring

and inhibitory control', *Frontiers in Psychology* 10. (doi:10.3389/fpsyg.2019.01839).

Liu, Jiaqi and Jiayan Lin. 2021. "A cross-linguistic study of L3 phonological acquisition of stop contrasts.", *SAGE Open*. (doi.org/10.1177/2158244020985510).

Linck, Jared A., Noriko Hoshino, and Judith F. Kroll. 2008. "Cross-language lexical processes and inhibitory control", *The Mental Lexicon* 3, 3: 349-74. (doi:10.1075/ml.3.3.06lin).

Linck, Jared A., Judith F. Kroll, and Gretchen Sunderman. 2009. "Losing access to the native language while immersed in a second language evidence for the role of inhibition in second-language learning", *Psychological Science* 20, 12: 1507-1515. (doi:10.1111/j.1467-9280.2009.02480.x).

Llama, Raquel, Walcir Cardoso and Laura Collins. 2010. "The influence of language distance and language status on the acquisition of L3 phonology", *International Journal of Multilingualism* 7, 1: 39-57. (doi: 10.1080/14790710902972255).

Llama, Raquel and Luz Patricia Lopez-Morelos. 2016. "VOT production by Spanish heritage speakers in a trilingual context", *International Journal of Multilingualism* 13, 4: 444-458. (doi:10.1080/14790718.2016.1217602).

Llama, Raquel, and Walcir Cardoso. 2018. "Revisiting (non-)native influence in vot production: Insights from advanced l3 Spanish", *Languages* 3: 1-18. (doi:10.3390/languages3030030).

Loewen, Shawn and Hayo Reinders. 2011. *Key concepts in second language acquisition*. Basingstoke: Palgrave Macmillan.

Lorenz, Eliane and Peter Siemund. 2019. "Differences in the acquisition and production of English as a foreign language: A study of bilingual and monolingual students in Germany", in: Eva Vetter and Ulrike Jessner (eds.), *International Research on Multilingualism: Breaking with the Monolingual Perspective*. Cham: Springer International Publishing, 81-102.

MacKay, Ian R.A., Diane Meador and James Emil Flege. 2001. "The identification of English consonants by native speakers of Italian", *Phonetica* 58, 1-2: 103-25. (doi.org/10.1159/000028490).

Maddieson, Ian. 2013. "Consonant Inventories", in: Matthew S. Dryer and Martin Haspelmath (eds.), *The world atlas of language structures online*. Lipzig: Max

Planck Institute for Evolutionary Anthropology. (http://wals.info/chapter/1). (Accessed on 2022-06-02.).

Maher, John C. 2017. *Multilingualism: A very short introduction*. Oxford: Oxford University Press.

Malisz, Zofia, and Marzena Żygis. 2015. "Voicing in Polish: Interactions with Lexical Stress and Focus.", in: The Scottish Consortium for ICPhS 2015 (ed.), *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow: The University of Glasgow.

Maneva, Blagovesta. 2004. "Maman, je suis polyglotte!: A case study of multilingual language acquisition from 0 to 5 years", *International Journal of Multilingualism* 1, 2: 109-22. (doi:10.1080/14790710408668182).

Marchman, Virginia A., Anne Fernald and Nereyda Hurtado. 2010. "How vocabulary size in two languages relates to efficiency in spoken word recognition by young Spanish-English bilinguals", *Journal of Child Language* 37, 4: 817-40. (doi: 10.1017/S0305000909990055).

Marian, Viorica, Henrike K. Blumenfeld and Margarita Kaushanskaya. 2007. "The language experience and proficiency questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals", *Journal of Speech, Language, and Hearing Research: JSLHR* 50 4: 940-67. (doi:10.1044/1092-4388(2007/067)).

Masoura, Elvira V. and Susan E. Gathercole. 1999. "Phonological Short-Term Memory and Foreign Language Learning", *International Journal of Psychology* 34, 5-6: 383-88. (doi:10.1080/002075999399738).

Meinhold, Gottfried And Eberhard Stock. 1982. *Phonologie Der Deutschen Gegenwartssprache*. Leipzig : Bibliographisches Institut.

Miyake, Akira, Naomi P. Friedman, Michael J. Emerson, Alexander H. Witzki, Amy Howerter, Tor D. Wager. 2000. "The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis.", *Cognitive Psychology, 41, 1*: 49-100. (doi:10.1006/cogp.1999.0734_).

Miyake, Akira and Naomi P. Friedman. 2012. "The nature and organization of individual differences in executive functions: Four general conclusions.", *Current Directions in Psychological Science*, 21, 1: 8-14. (doi:10.1177/0963721411429458).

Moeller, Aleidine J. and Theresa Catalano. 2015. “Foreign language teaching and learning”, in: James D. Wright (ed.), *International encyclopedia of the social & behavioral sciences*. Amsterdam: Elsevier, 327-332.

Mora, Joan C. 2007. “Methodological issues in assessing L2 perceptual phonological competence”, *Proceedings of the PTLC 2007 Phonetics Teaching and Learning Conference*. London: London: Department of Phonetics and Linguistics, University College London.

Mora, Joan. C. and Isabelle Darcy. 2013. Inhibition and phonological processing in a second language. (Paper presented at the 23rd Annual Conference of the European Second Language Association EUROSLA 23, 28-31 Aug. 2013).

Mora, Joan C. 2017. Assessing cross-linguistic influence in L3 phonology through language switching tasks: the role of L1 dominance and individual differences in attention and inhibitory control. (Paper presented at Workshop on Multilingual Language Acquisition, Processing and Use, Poznań, 6-7 May 2017).

Muñoz, Carmen. 2006. “Chapter 1 .The Effects of Age on Foreign Language Learning: The BAF Project”, in: Carmen Muñoz (ed.), *Age and the rate of foreign language learning*. Bristol: Multilingual Matters, 1-40.

Muszyński, Marek, Dorota Campfield, and Magdalena Szpotowicz. 2015. „Język angielski w szkole podstawowej–proces i efekty nauczania. Wyniki podłużnego badania efektywności nauczania języka angielskiego 2011-2014”[English in Primary School-the Process and Learning Outcomes. Results of Longitudinal Research on the Effectiveness of English Language Learning 2011-2014], Warsaw: Wydawnictwo IBE.

Myles, Florence. 2017. “Learning foreign languages in primary schools: is younger better? ”, *Languages, Society & Policy.* (doi:10.17863/CAM.9806).

Nelson, Christina, Iga Krzysik, Halina Lewandowska and Magdalena Wrembel. 2021. “Multilingual learners’ perceptions of cross-linguistic distances: a proposal for a visual psychotypological measure”, *Language Awareness* 30, 2: 176-194, (doi: 10.1080/09658416.2021.1897132).

Nelson, Christina. 2022. "Do a learner’s background languages change with increasing exposure to l3? Comparing the multilingual phonological development of adolescents and adults" *Languages* 7, 2: 78. (doi:10.3390/languages7020078).

Niedźwiedź, Katarzyna, Jolanta Sochaczewska-Kuleta, and Dorota Wosińska. 2017. „Program nauczania języka angielskiego dla II etapu edukacyjnego w klasach IV – VIII szkoły podstawowej" [English language teaching curriculum for the second educational stage, grades IV-VII, primary school]. Warsaw: Nowa Era.

Nimz, Katharina. 2016. *Sound perception and production in a foreign language: Die rolle der orthographie in der fremdsprachlichen lautperzeption und produktion*. Postdam: Universitätsverlag Potsdam.

Lev-Ari, Shiri and Sharon Peperkamp. 2013. "Low inhibitory skill leads to non-native perception and production in bilinguals' native language". *Journal of Phonetics*, 41, 5: 320-331. (doi: 10.1016/j.wocn.2013.06.002).

Ó Laoire, Muiris and David Singleton. 2009. "The role of prior knowledge in L3 learning and use", in: Larissa Aronin and Britta Hufeisen (eds.), *The exploration of multilingualism*. Amsterdam: John Benjamins Publishing Company, 79-102.

Odlin, Terence. 1989. *Language transfer: Cross-linguistic influence in language learning*. Cambridge: Cambridge University Press.

Odlin, Trence and Scott Jarvis. 2004. "Same Source, Different Outcomes: A Study of Swedish Influence on the Acquisition of English in Finland", *International Journal of Multilingualism* 1, 2: 123-140. (doi:10.1080/14790710408668183).

Ogden, Richard. 2011. "An Introduction to English Phonetics", *Phonetica* 68, 1-2: 111-112. (https://doi.org/10.1159/000328775).

Onishi, Hiromi. 2016. "The effects of L2 experience on L3 perception", *International Journal of Multilingualism 13*, 4: 459-475. (doi:10.1080/14790718.2016.1217604)

Orzechowska, Paula. 2019. Complexity in Polish phonotactics: On features, weights, rankings and preferences. Singapore: Springer.

Ostaszewska, Danuta and Jolanta Tambor. 2000. Fonetyka i fonologia współczesnego języka polskiego [Phonetics and phonology of Modern Polish]. Warszawa: Wydawnictwo Naukowe PWN.

Oyama, Susan. 1976. "A sensitive period for the acquisition of a nonnative phonological system", *Journal of Psycholinguistic Research* 5, 3: 261-83. (doi.org/10.1007/BF01067377).

Paradis, Michael. 1994. “Neurolinguistic aspects of implicit and explicit memory: implications for bilingualism”, in: Nick Ellis (ed.), *Implicit and explicit learning of second languages*. London: Academic Press, 393-41.

Paradis, Michael. 2004. *A neurolinguistic theory of bilingualism*. Amsterdam: John Benjamins Publishing Company.

Paradis, Michel 2009. *Declarative and procedural determinants of second languages*. Amsterdam: John Benjamins Publishing Company.

Pavlenko, Aneta. 2002. “Bidirectional transfer”, *Applied Linguistics,* 23, 190-214. (doi:10.1093/applin/23.2.190).

Payne, Brennan R. and Jack W. Silcox. 2019. “Chapter Seven - Aging, Context Processing, and Comprehension”, in: Kara D. Federmeier (ed.), *Psychology of Learning and Motivation* 71: 215-264. Academic Press. (doi:10.1016/bs.plm.2019.07.001).

Pierce, Lara J., Fred Genesee, Audrey Delcenserie and Gary Morgan. 2017. “Variations in phonological working memory: Linking early language experiences and language learning outcomes”, *Applied Psycholinguistics* 38, 6: 1265-1300. (doi:10.1017/S0142716417000236).

Pino Escobar, Gloria, Marina Kalashnikova and Paola Escudero. 2018. “Vocabulary matters! The relationship between verbal fluency and measures of inhibitory control in monolingual and bilingual children”, *Journal of Experimental Child Psychology* 170: 177-89. (doi:10.1016/j.jecp.2018.01.012).

Poarch, Gregory J. and Ellen Bialystok. 2015. “Bilingualism as a model for multitasking”, *Developmental Review* 1, 35: 113-124. (doi:10.1016/j.dr.2014.12.003).

R Core Team. 2017. R: A language and environment for statistical computing. (https://www.R-project.org/) (date of access: 21 March 2022).

Ringbom, Håkan. 1986. “Crosslinguistic influence and the foreign language learning process”, in: Eric Kellerman and Michael Sharwood-Smith (eds.), *Crosslinguistic Influence in Second Language Acquisition*, Oxford: Pergamon Press: 150-162.

Ringbom, Håkan. 2006. *Cross-linguistic similarity in foreign language learning*. Bristol: Multilingual Matters. (doi:10.21832/9781853599361).

Roach, Peter. 2004.“British English: Received Pronunciation.”, *Journal of the International Phonetic Association* 34, 2: 239-45. (doi:10.1017/S0025100304001768).

Rojczyk, Arkadiusz. 2009. „Parametr VOT w języku polskim i angielskim. Badanie percepcji" [VOT in Polish and English: A perception study.], *LingVaria* 4: 29-47.

Rojczyk, Arkadiusz. 2010. "Production and perception of vowel 'ash' by Polish learners of English, in: Katarzarzyna Dziubalska-Kołaczyk, Magdalena Wrembel and Małgorzata Kul (eds.), *Proceedings from the Sixth International Symposium on the Acquisition of Second Language Speech.*

Rojczyk, Arkadiusz and Andrzej Porzuczek. 2012. "Vowel reduction in English and Polish. General tendencies and individual variation", in: Ewa Piechurska-Kuciel and Liliana Piasecka (eds.), *Variability and stability in foreign and second language learning contexts.* Vol. 2. Newcastle upon Tyne: Cambridge Scholars Publishing, 207-227.

Rojczyk, Arkadiusz. 2013. "Vowel quality and duration in as a cue to word stress for non-native listeners: Polish listeners' perception of stress in English", in: Ewa Waniek-Klimczak and Linda Shockey (eds.), *Teaching and researching in english accents in native and non-native speakers*. Berlin Heidelberg: Springer, 59-72.

Rothman, Jason and Jennifer Cabrelli Amaro. 2010. "What variables condition syntactic transfer?: A look at the L3 initial state", *Second Language Research* 26, 2: 189-218.

Rothman, Jason. 2011. "L3 syntactic transfer selectivity and typological determinacy: the typological primacy model", *Second Language Research* 27, 1: 107-127. (doi: 10.1177/0267658310386439).

Rothman, Jason. 2013. "Cognitive economy, non-redundancy and typological primacy in L3 acquisition: Evidence from initial stages of L3 Romance", in: Sergio Baauw, Frank Dirjkoningen, Manuela Pinto and Luisa Meroni (eds.), *Romance languages and linguistic theory 2011,* Amsterdam: John Benjamins Publishing Company, 217-247.

Rothman, Jason. 2015. "Linguistic and cognitive motivations for the Typological Primacy Model (TPM) of third language (L3) transfer: Timing of acquisition and proficiency considered", *Bilingualism: Language and Cognition* 18, 2: 179-190. (doi:10.1017/S136672891300059X).

Rothman, Jason, Jorge González Alonso, Eloi Puig-Mayenco. 2019. *Third language acquisition and linguistic transfer*. Cambridge: Cambridge University Press.

Róg, Mirosław Łukasz. 2000. „Problemy w nauczaniu i uczeniu się języka angielskiego w klasach I–III szkoły podstawowej w opiniach nauczycieli" [Problems in teaching and learning English in grades 1-3 in the opinions of teachers], *Pedagogika wczesnoszkolna i szkolna* 8: 22-36.

Saito, Kazuya and Kim van Poeteren. 2018. "The perception–production link revisited: the case of Japanese learners' English /ɹ/ performance". *International Journal of Applied Linguistics* 28: 3-17. (doi:10.1111/ijal.12175).

Shanon, Benny. 1991. "Faulty language selection in polyglots", *Language and Cognitive Processes* 6 :4: 339-350. (doi: 10.1080/01690969108406947).

Sharwood Smith, Michael and Erik Kellerman. 1986. *Crosslinguistic influence in second language acquisition*. New York: Pergamon Press.

Sigmeth, Kristy and Christina Golin. 2018. Executive function and phonological perception in young L3 learners. (Paper presented at Young Linguists' Meeting, Poznań 2018, 23-25 Nov. 2018).

Sigmeth, Kristy, Romana Kopečková and Gregory Poarch. 2019. Phonological acquisition and inhibitory control in L2 English learners. (Paper presented Conference on Multilingualism, Universiteit Leiden, 1-3 Sept. 2019).

Singleton, David M. and Zsolt Lengyel. 1995. *The age factor in second language acquisition: A critical look at the critical period hypothesis*. Clevedon: Multilingual Matters.

Singleton, David. 2001. "Age and second language acquisition.", *Annual Review of Applied Linguistics* 21: 77-89. (doi:10.1017/S0267190501000058).

Sullivan, Larry E. 2009. "Individual differences", in: *The SAGE glossary of the social and behavioral sciences*. Vol. 1. New York: SAGE Publications, 252-252.

Sun, Peijian Paul and Lawrence Jun Zhang. 2020. "A Multidimensional Perspective on Individual Differences in Multilingual Learners' L2 Chinese Speech Production", *Frontiers in Psychology* 11. (doi: 10.3389/fpsyg.2020.00059).

Sypiańska, Jolanta. 2016a. "L1 vowels of multilinguals: The applicability of SLM in multilingualism", *Research in Language* 14, 1:79-94. (doi:10.1515/rela-2016-0003).

Sypiańska, Jolanta. 2016b. “Multilingual acquisition of vowels in L1 Polish, L2 Danish and L3 English”, *International Journal of Multilingualism*, 13, 4: 476-495. (doi:10.1080/14790718.2016.1217606).

Szewczyk, Jakub, Magdalena Smoczyńska, Ewa Haman, Magdalena Łuniewska, Magdalena Kochańska and Joanna Załupska. 2015. *Test powtarzania pseudosłów TPP [Pseudoword Repetition Test TPP]*. Warszawa: Instytut Badań Edukacyjnych.

Szpotowicz, Małgorzata. 2011. "Lekcje języka obcego w nauczaniu wczesnoszkolnym"[Foreign language classes in early school education], *Nauczanie wczesnoszkolne* 1: 4-11.

Tremblay, Marie-Claude. 2007. L2 influence on L3 pronunciation: Native-like VOT in the L3 Japanese of English-French bilinguals. (Paper presented at the Satellite Workshop of ICPhS XVI, Freiburg, Germany, 3-4 Aug. 2007).

Valdés, Guadalupe. 2021. Multilingualism (https://www.linguisticsociety.org/resource/multilingualism) (date of access: 18 May 2021).

Waniek-Klimczak, Ewa. 2011. *Aspiration in Polish: A sound change in progress*?, in: Mirosław Pawlak and Jakub Bielak (eds.), *New perspectives in language, discourse and translation studies*. Berlin: Springer-Verlag. 3-12.

Westergaard, Marit, Natalia Mitrofanova, Roksolana Mykhaylyk, and Yulia Rodina. 2016. “Crosslinguistic influence in the acquisition of a third language: The Linguistic Proximity Model.”, *International Journal of Bilingualism* 21, 6: 666-82. (doi:10.1177/1367006916648859).

Williams, Lee. 1979. “The modification of speech perception and production in second-language learning”, *Perception and Psychophysics* 26: 95-104. (doi:10.3758/BF03208301).

Williams, Sarah and Björn Hammarberg. 1998. “Language switches in L3 production: Implications for a polyglot speaking model”, *Applied Linguistics*, 19: 295-333. (doi:10.1093/applin/19.3.295).

Wingfield Andrew. “Cognitive factors in auditory performance: context, speed of processing, and constraints of memory”, *Journal of American Academy of Audiology* 7, 3:175-82.

Wrembel, Magdalena. 2014a. “Metaphonological awareness in multilinguals; a case of L3 Polish”, *Language Awareness* 24, 1: 60-83. (doi:10.1080/09658416.2014.890209).

Wrembel, Magdalena. 2014b. “VOT Patterns in the Acquisition of Third Language Phonology”, *Concordia Papers in Applied Linguistics* (COPAL), 5: 751-771.

Wrembel, Magdalena. 2015. *In search of a new perspective: Cross-linguistic influence in the acquisition of third language phonology*. Poznań: Wydawnictwo Naukowe UAM.

Wrembel Magdalena, Ulrike Gut, Iga Krzysik, Halina Lewandowska, Anna Balas. 2019. “Acquisition of rhotics by multilingual children”, in: Sasha Calhoun, Paola Escudero, Marija Tabain and Paul Warren (eds.), *Proceedings of the 19th International Congress of Phonetic Sciences*. Canberra: Australasian Speech Science and Technology Association Inc, 954-958.

Wrembel Magdalena, Ulrike Gut, Romana Kopečková, Anna Balas. 2020. “Cross-Linguistic Interactions in Third Language Acquisition: Evidence from Multi-Feature Analysis of Speech Perception”, *Languages*, 5, 4: 1-21. (doi:10.3390/languages5040052).

Wrembel, Magdalena, Ulrike Gut, Romana Kopečková and Anna Balas. 2022. ’’The relationship between the perception and production of L2 and L3 rhotics in young multilinguals: An exploratory cross-linguistic study”, *International Journal of Multilingualism*. (DOI:10.1080/14790718.2022.2036158).

Wunder, Eva-Maria. 2010. “Phonological cross-linguistic influence in third or additional language acquisition”. *Proceedings of 6th New Sounds* 2010, Poznań.

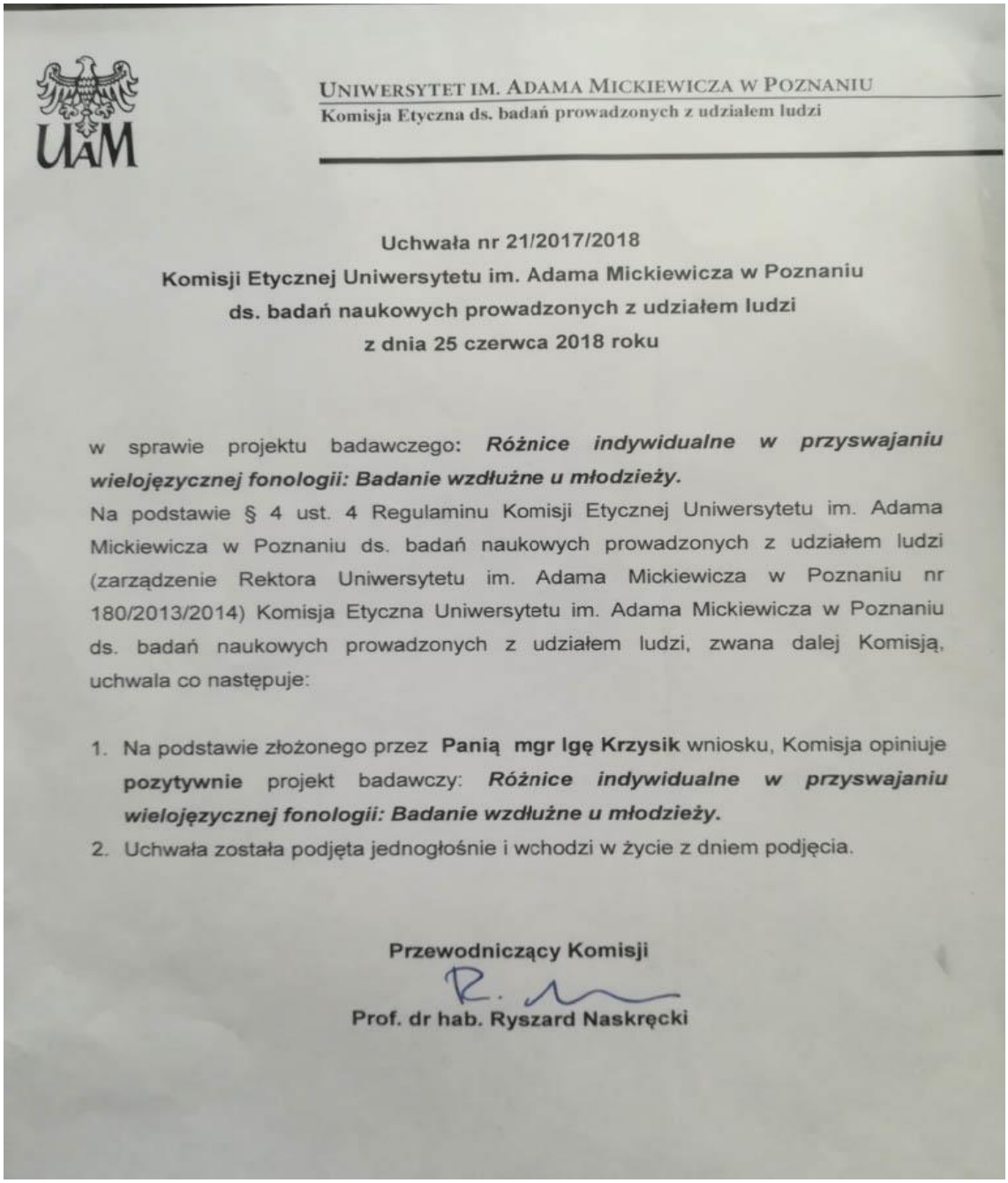
Xia, Chris M. 2017. “Psychotypology of Chinese learners of English and its influence on the acquisition of metaphorical expressions: An online study”, *Cambridge Occasional Papers in Linguistics* 101: 237-255.

Yao, Yao. 2007. “Closure duration and VOT of word-initial voiceless plosives in English in spontaneous connected speech”, *UC Berkeley Phonology Lab Annual Reports 3*. (doi:10.5070/P71HS7H769.)

Yu, Alan C. L. and Georgia Zellou. 2019. “Individual Differences in Language Processing: Phonology”, *Annual Review of Linguistics* 5, 1: 131-50. (doi:10.1146/annurev-linguistics-011516-033815).

Zhang, Jennifer. 2019. “Feature-specific advantages in L3 phonological acquisition”, in: Sasha Calhoun, Paola Escudero, Marija Tabain and Paul Warren (eds.), *Proceedings of the 19th International Congress of Phonetic Sciences*. Canberra: Australasian Speech Science and Technology Association Inc, 3740-3744.

Appendix A: The approval of the AMU ethics committee



UNIWERSYTET IM. ADAMA MICKIEWICZA W POZNANIU
Komisja Etyczna ds. badań prowadzonych z udziałem ludzi

Uchwała nr 21/2017/2018
Komisji Etycznej Uniwersytetu im. Adama Mickiewicza w Poznaniu
ds. badań naukowych prowadzonych z udziałem ludzi
z dnia 25 czerwca 2018 roku

w sprawie projektu badawczego: ***Różnice indywidualne w przyswajaniu wielojęzycznej fonologii: Badanie wzdłużne u młodzieży.***
Na podstawie § 4 ust. 4 Regulaminu Komisji Etycznej Uniwersytetu im. Adama Mickiewicza w Poznaniu ds. badań naukowych prowadzonych z udziałem ludzi (zarządzenie Rektora Uniwersytetu im. Adama Mickiewicza w Poznaniu nr 180/2013/2014) Komisja Etyczna Uniwersytetu im. Adama Mickiewicza w Poznaniu ds. badań naukowych prowadzonych z udziałem ludzi, zwana dalej Komisją, uchwala co następuje:

1. Na podstawie złożonego przez **Panią mgr Igę Krzysik** wniosku, Komisja opiniuje **pozytywnie** projekt badawczy: ***Różnice indywidualne w przyswajaniu wielojęzycznej fonologii: Badanie wzdłużne u młodzieży.***
2. Uchwała została podjęta jednogłośnie i wchodzi w życie z dniem podjęcia.

Przewodniczący Komisji
Prof. dr hab. Ryszard Naskręcki

Appendix B: Informed consent forms (in Polish) for the parents/caregivers and children

Informed consent form for the schoolchildren (the participants) in Polish

FORMULARZ ZGODY UCZNIA / UCZENNICY NA UDZIAŁ W BADANIU JĘZYKOWYM

Wyrażam zgodę na udział w 45 minutowym, anonimowym i dobrowolnym badaniu językowym w dniu _____ przeprowadzonym przez Igę Krzysik, doktorantkę Wydziału Anglistyki Uniwersytetu im. Adama Mickiewicza w Poznaniu. Podczas badania wypełnię kwestionariusz dotyczący nauki i użycia języków obcych, wykonam krótkie testy językowe (w j. polskim, angielskim i niemieckim) oraz test uwagi. Przebieg badania został mi szczegółowo objaśniony i jest dla mnie zrozumiały. Wyniki testów i uzyskane nagrania nie będą podstawą do oceny z żadnego przedmiotu. Rozumiem, że mogę wycofać się z badania w każdej chwili, bez podawania przyczyny.

_____ _____
podpis osoby badanej klasa

_____
podpis osoby przeprowadzającej badanie

Informed consent form for the parents/caregivers of the participants

OŚWIADCZENIE RODZICÓW / OPIEKUNÓW USTAWOWYCH O ZGODZIE NA UDZIAŁ W BADANIU I PRZETWARZANIU DANYCH OSOBOWYCH

Tytuł projektu badawczego: Różnice indywidualne w przyswajaniu wielojęzycznej fonologii: Badanie wzdłużne u młodzieży.

Wykonawca projektu badawczego: mgr Iga Krzysik, doktorantka na Wydziale Anglistyki, Uniwersytet im. Adama Mickiewicza w Poznaniu. Praca realizowana jest pod kierunkiem prof. UAM dr hab. Magdaleny Wrembel.

Kontakt z wykonawcą projektu badawczego: ikrzysik@wa.amu.edu.pl

Imię i nazwisko
dziecka .....
Data
urodzenia ……
……
Imię i nazwisko osoby składającej
oświadczenie .....

Stosunek prawny od osoby uczestniczącej w
badaniu ……...

Niniejszym oświadczam, że:

1. zostałam/em szczegółowo poinformowana/ny o celu zamierzonych badań i sposobie ich przeprowadzenia. Rozumiem na czym mają one polegać i do czego potrzebna jest moja zgoda.
2. zostałam/em poinformowana/ny, że mogę odmówić zgody na udział dziecka w badaniach lub cofnąć ją w każdej chwili bez podawania przyczyny, także podczas wykonywania badań.
3. zostałam/em poinformowana/ny, że dziecko może wycofać się udziału w projekcie badawczym w każdej chwili, bez podawania przyczyny, także podczas wykonywania badań.
4. zostałam/em poinformowana/y o tym, że badaczka właściwie zabezpieczy powierzone jej informacje zawierające dane osobowe.

Wyrażam pełną i świadomą zgodę na udział w badaniach oraz na przetwarzanie danych osobowych dla celów niezbędnych dla przeprowadzenia badania. Wyrażona zgoda może zostać cofnięta poprzez kontakt z badaczką. *

…….....……..
miejscowość, data

.....
podpis rodzica (opiekuna)

.....
podpis osoby przeprowadzającej badanie

*Zgodnie z rozporządzeniem Parlamentu Europejskiego i Rady (UE) 2016/679 z 27 kwietnia 2016 r. w sprawie ochrony osób fizycznych w związku z przetwarzaniem danych osobowych i w sprawie swobodnego przepływu takich danych oraz uchylenia dyrektywy 95/46/WE (ogólne rozporządzenie o ochronie danych) (Dz. Urz. UE L nr 119, str. 1) oraz ustawą z 29 sierpnia 1997 r. o ochronie danych osobowych (tekst jedn.: Dz. U. z 2016 r. poz. 922).

Appendix C: Speech production tokens in L2 and L3

Tested feature	L2 English	L3 German
VOT	pot	Pott
	party	packe
	cot	Tante
	carpet	toll
	toss	Kappe
	target	Koch
vowel length	wick	Mitte
	week	bitte
	bid	Miete
	bead	Linie
	did	L*ü*ge
	deed	Bücher
		küssen
		f*ü*llen

Appendix D: Pseudoword tokens

Number of syllables	Pseudoword
1	szyl
	łeć
	pnuk
	kłan
	gmut
2	pykon
	stynał
	bagół
	ośręcz
	łątka
3	pilichek
	odlarki
	retolnik
	głanyca
	gorwiwaść
4	słumoniny
	rożdziejenie
	wcisąwanie
	lęgiczybość
	cnekotada
5	owyganalność
	kontralowaria
	kantnolowania
	ukatowanonie
	zrenaksownie
6	ksopiliwaniewaść
	słubagorwinienie
	stymugowaniny
	kołąpykojenie

Appendix E: Language history questionnaire

The English translation of the language history and use questionnaire used in the task.

Language questionnaire date _____ code _____

Age	
Gender	
The country of birth	

Language learning history

1. What is your first language?

_____

2. List the languages that you know/learn in the order in which you started acquiring them (starting from your native language).

_____

3. How long have you been learning the listed languages? Indicate the number of years/months.

_____

4. Have you spent more than two weeks in a country where one of the languages listed is spoken? (if yes, please write the country of your stay and the language used).

_____

5. How would you rate your skills in the languages you know. Indicate your rating in the table below:

SCALE **1** – very poor **2** – poor **3** – average **4** – good **5** – very good

Language	READING	WRITING	SPEAKING	LISTENING
POLISH				
ENGLISH				
GERMAN				
OTHER ……				

6. How often do you use the languages in these situations situation? The percentage of time allocated to individual activities in each category should total 100%.

WITH FAMILY

	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
POLISH											
ENGLISH											
GERMAN											
OTHER …….											

WITH FRIENDS

	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
POLISH											
ENGLISH											
GERMAN											

OTHER …….											

AT SCHOOL

	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
POLISH											
ENGLISH											
GERMAN											
OTHER …….											

7. How often do you use individual languages to perform the following activities?

SCALE 1 – never **2** – once a month **3** – once a week **4** – a few times a week **5** – everyday

POLISH	1	2	3	4	5
Watch TV / listen to the radio					
Read books					
Play computer games					
Surf the Internet					
Communicate via the Internet					

ANGIELSKI	1	2	3	4	5
Watch TV / listen to the radio					
Read books					

Play computer games					
Surf the Internet					
Communicate via the Internet					

GERMAN	1	2	3	4	5
Watch TV / listen to the radio					
Read books					
Play computer games					
Surf the Internet					
Communicate via the Internet					

OTHER ……	1	2	3	4	5
Watch TV / listen to the radio					
Read books					
Play computer games					
Surf the Internet					
Communicate via the Internet					

8. Does your knowledge of Polish and English help you learn German? If so, explain how.

_____

9. Do you often mix your languages while speaking?

_____

9b. (If the answered “yes” in the question above) How often and in what situations does it happen?

_____

Appendix F: Language proficiency measure - questions

Questions in L2 English	Questions in L3 German
How old are you?	Wie alt bist du?
What is your name?	Wie heißt du?
Do you like learning English? Why?	Magst du Deutsch lernen? Warum?
What is your favourite school subject?	Was ist dein Lieblingsfach in der Schule?
What are your hobbies?	Was sind deine Hobbies?

Appendix G: Multiple regression analysis – R code

```
# run multiple regression analysis

# L2 models
L2_T1_PERC_Model <- lm(T1_L2_PERC ~ T1_FLANKER + T1_PWM + T1_PROFI, data = L2)
L2_T3_PERC_Model <- lm(T3_L2_PERC ~ T3_FLANKER + T3_PWM + T3_PROFI, data = L2)
L2_T1_PROD_Model <- lm(T1_L2_PROD ~ T1_FLANKER + T1_PWM + T1_PROFI, data = L2)
L2_T3_PROD_Model <- lm(T3_L2_PROD ~ T3_FLANKER + T3_PWM + T3_PROFI, data = L2)

# L3 models
L3_T1_PERC_Model <- lm(T1_L3_PERC ~ T1_FLANKER + T1_PWM + T1_PROFI, data = L3)
L3_T3_PERC_Model <- lm(T3_L3_PERC ~ T3_FLANKER + T3_PWM + T3_PROFI, data = L3)
L3_T1_PROD_Model <- lm(T1_L3_PROD ~ T1_FLANKER + T1_PWM + T1_PROFI, data = L3)
L3_T3_PROD_Model <- lm(T3_L3_PROD ~ T3_FLANKER + T3_PWM + T3_PROFI, data = L3)

# create table with results
summary(L2_T1_PERC_Model)
summary(L2_T3_PERC_Model)
summary(L2_T1_PROD_Model)
summary(L2_T3_PROD_Model)
summary(L3_T1_PERC_Model)
summary(L3_T3_PERC_Model)
summary(L3_T1_PROD_Model)
summary(L3_T3_PROD_Model)

# checking multicollinearity for independent variables
car::vif(L2_T1_PERC_Model)
car::vif(L2_T3_PERC_Model)
car::vif(L2_T1_PROD_Model)
car::vif(L2_T3_PROD_Model)
car::vif(L3_T1_PERC_Model)
car::vif(L3_T3_PERC_Model)
car::vif(L3_T1_PROD_Model)
```