Zuzanna Ewa Cal

150024

Voice onset time in multilingual speech: Production and perception-based perspectives

Parametr voice onset time w mowie wielojęzycznej na podstawie badań produkcji i percepcji

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

Poznań, 2026

ACKNOWLEDGEMENTS

...what a ride, huh? Honestly, there were plenty of days when I truly thought I'd never be writing these words, let alone actually finishing this thing. If you had told me at the start how this journey would go, I'm not sure I would have believed I'd make it. But I did, and I wouldn't be holding this finished thesis if it weren't for the people around me. This is my chance to say thank you.

First and foremost, I would like to thank my supervisor, Professor Magdalena Wrembel, for trusting me and allowing me to be a part of something bigger. I am deeply grateful for the scholarship funding through the project, which not only financed my research but also gave me invaluable experience I wouldn't have gained anywhere else. Most of all, thank you for your constant support. Whenever I was stuck down a rabbit hole, you helped me zoom out, and your confidence in this work kept me going, even at times when my own had completely vanished.

I extend this gratitude to the entire CLIMAD and ADIM project teams. Thank you for the teamwork, feedback, all the brainstorming and your unwavering reliability during data collection. It was an absolute pleasure working alongside all of you.

I would also like to thank my colleagues and fellow PhD students at the Faculty of English for their support. I am especially grateful to Iryna and Justyna, with whom I could share the daily chaos and existential crises, knowing I would be understood like nowhere else.

Heartfelt thanks go to Jola, for sharing with me her love of language, guiding me through my very first steps in academia and for consistently believing in me and motivating me to keep going.

I owe a lot to my family, friends, and my partner, Łukasz, for always backing me up, giving me space when I needed it and graciously avoiding the forbidden question of "How is your PhD going?".

Special credit goes to my dog, Soma, who contributed absolutely nothing to the methodology but everything to my mental health, enforcing outdoor breaks and sleeping beside me through every stressful deadline.

And finally, a quiet nod to myself. For showing up on the days when nothing worked, rewriting the parts that hurt to delete and simply refusing to quit. We did it!

OŚWIADCZENIE

Ja, niżej podpisana

Zuzanna Ewa Cal

przedkładam rozprawę doktorską

pt. *Voice onset time in multilingual speech: Production and perception-based perspectives*

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,

(miejscowość, data) (czytelny podpis)

Table of contents

TABLE OF CONTENTS ..... **IV**

LIST OF TABLES ..... **VI**

LIST OF PUBLICATIONS ..... **VII**

FUNDING ..... **VIII**

LIST OF ABBREVIATIONS ..... **IX**

PART 1 : INTRODUCTION **1**

1.1. THEORETICAL BACKGROUND 3

1.1.1. Cross-linguistic influence *3*

1.1.2. Models of Third Language Acquisition *5*

1.1.3. Voice onset time *7*

PART 2 : RESEARCH **10**

2.1. RESEARCH OBJECTIVES 10

2.2. PUBLICATION 1 (CAL & WREMBEL, 2025): DEVELOPMENTAL ACQUISITION OF STOPS BY MULTILINGUAL SPEAKERS 12

2.3. PUBLICATION 2 (CAL, 2026A): CROSSING THE BOUNDARIES: PERCEPTION OF VOICE ONSET TIME IN WORD-INITIAL STOPS BY MULTILINGUAL LEARNERS 14

2.4. PUBLICATION 3 (CAL, 2026B): IS L2 THE LEAST STABLE SYSTEM? THE DYNAMICS OF VOT TRILINGUAL PRODUCTION IN POLISH-ENGLISH-NORWEGIAN READ AND SPONTANEOUS SPEECH 16

PART 3 : GENERAL DISCUSSION **20**

3.1. PRODUCTION DATA 20

3.1.1. VOT across speech modes *21*

3.1.2. Asymmetry between voiceless and voiced stops *22*

3.2. PERCEPTION DATA 23

3.3. INFORMING THEORETICAL MODELS OF MULTILINGUAL ACQUISITION 25

3.4. INDIVIDUAL VARIATION 28

3.5. IMPLICATIONS 28

3.6. LIMITATIONS AND FUTURE RESEARCH DIRECTIONS 29

CONCLUSION 31

ABSTRACT 33

STRESZCZENIE 35

REFERENCES 37

APPENDIX A: PUBLICATION 1 (CAL & WREMBEL, 2025): AUTHOR CONTRIBUTION STATEMENT (POLISH) 45

APPENDIX B: PUBLICATION 1 (CAL & WREMBEL, 2025): AUTHOR CONTRIBUTION STATEMENT (ENGLISH) 46

APPENDIX C: PUBLICATION 2 (CAL, 2026A): AUTHOR CONTRIBUTION STATEMENT (POLISH) 47

APPENDIX D: PUBLICATION 2 (CAL, 2026A): AUTHOR CONTRIBUTION STATEMENT (ENGLISH) 48

APPENDIX E: PUBLICATION 3 (CAL, 2026B): AUTHOR CONTRIBUTION STATEMENT (POLISH) 49

APPENDIX F: PUBLICATION 3 (CAL, 2026B): AUTHOR CONTRIBUTION STATEMENT (ENGLISH) 50

APPENDIX G: PUBLICATION 1 (CAL & WREMBEL, 2025) 51

APPENDIX H: PUBLICATION 2 (CAL, 2026A) 75

APPENDIX I: PUBLICATION 3 (CAL, 2026B) 100

5:4211820347

List of tables

Table 1 VOT values (in ms) for voiced and voiceless stops in Polish, English and Norwegian measured in word-initial stressed syllables. 8

6:1038730515

List of Publications

The present PhD thesis comprises three thematically related Publications:

Publication 1 (Cal & Wrembel, 2025)

Cal, Zuzanna and Magdalena Wrembel. 2025. "Developmental acquisition of stops by multilingual speakers", *International Journal of Multilingualism* 22, 1: 13–35.

- doi: https://doi.org/10.1080/14790718.2024.2363777

Publication 2 (Cal, 2026a)

Cal, Zuzanna. 2026a. "Crossing the boundaries: Perception of voice onset time in word-initial stops by multilingual learners", *Language and Speech*. (Online first publication)

- doi: https://doi.org/10.1177/00238309251414959
- Supplementary materials: https://osf.io/m7a43/

Publication 3 (Cal, 2026b)

Cal, Zuzanna. 2026b. "Is L2 the least stable system? The dynamics of VOT trilingual production in Polish-English-Norwegian read and spontaneous speech", *Languages* 11, 9: 1–28.

- doi: https://doi.org/10.3390/languages11090182
- Supplementary materials: https://osf.io/2jucr/

7:6176750357

Funding

The research presented in this PhD thesis and Publications 1-3 (Cal & Wrembel, 2025, Cal 2026a, 2026b) was supported financially by a research grant from the National Science Centre (Poland) – OPUS 2020/37/B/HS2/00617, titled "Cross-linguistic influence in multilingualism across domains: Phonology and syntax (CLIMAD)," whose Principal Investigator was prof. UAM dr hab. Magdalena Wrembel. The PhD candidate received a doctoral scholarship as part of this project.

8:3717438083

List of abbreviations

CEM	Cumulative-Enhancement Model
CLI	cross-linguistic influence
L1	first (native) language
L2	second language
L3	third language
Ln	additional foreign language
LPM	Linguistic Proximity Model
NGTA	Natural Growth Theory of Acquisition
PPH	Phonological Permeability Hypothesis
T1	testing time 1
T2	testing time 2
T3	testing time 3
TLA	third language acquisition
TPM	Typological Primacy Model
VOT	voice onset time

9:6950491112

Part 1: Introduction

Everyone speaks with an accent and nowhere is this phonetic reality more pervasive than in multilingual speech. As famously posited by Grosjean (1989:1), "the bilingual is not two monolinguals in one person." Rather, the language systems of a bi-/multilingual speaker co-exist and influence one another, leading to more or less noticeable shifts in one's phonetic realisations of all spoken languages, whether it is an L1, L2 or Ln. These linguistic interactions are, therefore, an intrinsic property of multilingual speech. This phenomenon, widely known as cross-linguistic influence (CLI), has received increased attention since the 1980s (cf. Sharwood-Smith 1983; Kellerman 1984). However, while earlier studies on bilingualism have shown that CLI operates as a bidirectional influence between L1 and L2, the situation becomes more complex when a third or additional language is added to the system. With the majority of the world population being bi- or multilingual (Grosjean, 2024), it seems vital to deepen our understanding of the underlying mechanisms of CLI in speech production and perception.

CLI in multilingual speech has been examined from various perspectives, focusing both on the production of individual sounds, including vowels (e.g., Sypiańska, 2016), nasals (Lin, 2025; Zhu & Wang, 2025), approximants (e.g., Sypiańska, 2025), rhotics (Kopečková et al., 2025), obstruent voicing (e.g., Kopečková et al., 2025; Patience et al., 2026), as well as of suprasegmental features (e.g., Gut, 2010; Sypiańska & Wrembel, 2026). Stops have also been widely investigated, primarily through the acoustic measure of voice onset time (VOT), which allows for the examination of phonetic interactions across languages (e.g., Llama & Cardoso, 2018; Amengual, 2021; Geiss et al., 2022; Wei, 2025; Zeng, 2026). The greatest number of previous investigations is centred on establishing the source of CLI in third language (L3) voiceless stops. The results are, however, inconclusive. While some studies identify the L1 as the primary source of CLI (e.g., Bandeira & Zimmer, 2012; Llama & Cardoso, 2018), others point to L2 as a predominant source of influence on L3 (e.g., Tremblay, 2007; Llama et al., 2010; Wei, 2025). Nonetheless, the majority of investigations suggests a combined effect of both L1 and L2 on L3 VOT values (e.g., Wunder, 2010; Wrembel, 2011, 2014, 2015; Sypiańska, 2017; Zhang, 2022). Furthermore, there is also some evidence of regressive (i.e. where a newly acquired language influences previously learnt ones) CLI occurring in voiceless stops,

10:6592466982

which leads to changes in already established languages (e.g., Nelson, 2022; Stoehr et al. 2024; Xu, 2025).

Compared to the substantial body of research on voiceless stops, the voiced series has received relatively little attention. The gap persists despite evidence that voiced and voiceless stops are affected differently by CLI, with the former being more prone to phonetic variation (e.g., Zając, 2015; Geiss et al., 2022; Wojtkowiak, 2022; Zhang et al., 2023) – a pattern that was also observed in Publications 1 and 3 of the present dissertation (Cal & Wrembel, 2025; Cal, 2026b). Comparably, little attention has also been devoted to more holistic analyses including not only the comparison of three languages, but also the two voicing series.

Interestingly, CLI is not only attributed to speech production, but it also has been attested in speech perception. However, there is a notable paucity of studies examining stop perception in multilinguals with existing investigations offering equivocal findings. On the one hand, several findings document the presence of CLI in perception, typically manifested as merged perceptual categories or boundary shifts (e.g., J. Liu et al., 2019; Stan, 2024). On the other hand, there is also evidence of separate perceptual systems (e.g., Z. Liu, 2016; Cal & Wrembel, 2023).

All things considered, the previous research undoubtedly provides important insights into the field of VOT acquisition by multilinguals. However, the scope of the existing investigations is rather narrow, as they tend to focus only on a fraction of variables rather than providing a bigger picture on the dynamics and patterns of CLI, both in perception and production of speech. Hence, this PhD dissertation is an attempt to bridge these gaps by employing a more holistic approach to the acquisition of stops by multilingual learners. This project aims to depart from isolated investigations by bringing together a series of three Publications intended to capture the dynamic nature of CLI. The thesis, thus, investigates multilingual stop acquisition across the speakers' entire linguistic repertoire (L1 Polish, L2 English and L3 Norwegian) and across all three places of articulation (bilabial, coronal and velar) in both voicing series. The overarching objective of this PhD dissertation is to determine how CLI manifests itself and evolves across stop perception and production by multilingual speakers. To achieve this, three interconnected Publications were prepared: (1) a longitudinal study tracking the development of VOT production over time; (2) a cross-sectional study examining stop perception and category boundary

11:6367713574

shifts; and (3) a cross-sectional production study comparing VOT patterns across both controlled (read) and more ecologically valid (spontaneous) speech styles.

The structure of this PhD thesis is as follows. First of all, a section on theoretical background, in which the empirical studies are grounded, is provided. It offers definitions of the concepts that are essential for this project. The second part is devoted to research conducted within this PhD dissertation. It starts with the presentation of research objectives, which is followed by detailed outlines of Publication 1 (Cal & Wrembel, 2025), Publication 2 (Cal, 2026a) and Publication 3 (Cal, 2026b). Part 3 provides a general discussion of the key findings of the three papers.

1.1. Theoretical background

This section introduces the key concepts discussed in the present dissertation. Firstly, it explains the phenomenon of CLI with a particular focus on its occurrence in multilingualism and speech. Attention then turns to models of third language acquisition (TLA) and how they conceptualise CLI. Finally, voice onset time is defined, and the languages of interest are introduced in terms of their laryngeal contrasts.

1.1.1. Cross-linguistic influence

Multilingualism is characterized by the constant interplay between linguistic systems of a speaker. Although this process often occurs unobtrusively in a casual conversation, it constitutes a key mechanism of non-native speech. This phenomenon is formally conceptualized as cross-linguistic influence (CLI) and was defined by Sharwood-Smith (1994: 198) as "the influence of the mother tongue on the learner's performance in and/or development of a given target language; by extension, it also means the influence of any 'other language' known to the learner on that target language." Historically, CLI research was predominantly focused on the bidirectional influence between a native and a foreign language (i.e., between L1 and L2). However, implementing a third or additional language (L3/Ln) to the speaker's repertoire exponentially increases complexity, creating a greater number of potential influences (De Angelis, 2007).

12:8817363514

CLI can be classified based on an array of factors (Jarvis & Pavlenko, 2008). In terms of its outcome, it is commonly dichotomized into facilitative (positive) CLI, which promotes target-like renditions, and non-facilitative (negative) CLI, resulting in non-target-like productions or perceptual shifts. Furthermore, CLI can vary in terms of intentionality or a linguistic subfield it appears in. Central to this dissertation, however, is the distinction based on directionality. Gass and Selinker (2001) distinguish between the forward (L1→L2) and backward (L2→L1) transfer, which has been attested in second language acquisition (SLA) research. While this classification is useful for describing the interplay of two languages, with more additional languages it might not be sufficient. Therefore, Jarvis & Pavlenko (2008) put forward the idea of lateral and bi-/multidirectional transfer. The former refers to any influence from a non-native language to another non-native language, while the latter means the simultaneous influence between two or more languages from the multilinguals' repertoire (e.g., L1 ↔ L2, L2 ↔ L3). Yet, the most ubiquitous terminology in multilingual research seems to classify CLI as progressive (L1/L2→L3), where previously learnt languages affect a new language, and regressive (L3→L1/L2), which assumes the influence of a new language on the existing linguistic repertoire (Cabrelli Amaro, 2013). Furthermore, De Angelis (2007) also distinguishes combined CLI, which assumes a many-to-one influence, that is the simultaneous influence stemming from more than language on a target one.

CLI in speech production and perception is frequently framed within the Speech Learning Model (SLM; Flege, 1995) and its revised version (SLM-r; Flege & Bohn, 2021) as the continuous interaction of sound categories in a single phonetic space. According to the model, when non-native sounds are perceived as phonetically similar to the existing, established categories, the formation of a new category for this sound is inhibited by equivalence classification. Consequently, this leads to category assimilation, where shared categories yield intermediate values across production and perception. It thereby accounts for bidirectional CLI. Only when the sounds are sufficiently different from the already known targets, can learners bypass equivalence classification and establish an independent phonetic category. Conversely, even when a new phonetic category is successfully established, the need to maintain a systemic contrast may trigger category dissimilation, causing existing representations to move away from their targets.

13:6530228280

1.1.2. Models of Third Language Acquisition

There have been attempts to provide an explanatory model of how CLI functions based on a number of factors that could play a role in modulating the process of third language acquisition (TLA). L3 models aim to answer the question of what the source of CLI in the L3 is and to provide an explanation of this process. Although most of them focus solely on the L3, there are also some frameworks that account for all languages in multilingual speakers' repertoires. The five most prominent frameworks for modelling L3 acquisition include the Cumulative-Enhancement Model (Flynn et al., 2004), the L2 Status Factor Model (Bardel & Falk, 2007, 2012), the Typological Primacy Model (Rothman, 2011, 2015), the Linguistic Proximity Model (Westergaard et al., 2017), and the Scalpel Model (Slabakova, 2017). This section briefly outlines each model and further introduces the Phonological Permeability Hypothesis (PPH; Cabrelli Amaro & Rothman, 2010) and the Natural Growth Theory of Acquisition (NGTA; Dziubalska-Kołaczyk & Wrembel, 2022, 2024) focusing specifically on speech acquisition, given that the five preceding frameworks are primarily grounded in (morpho)syntactic rather than phonological CLI.

The Cumulative-Enhancement Model (CEM) was proposed by Flynn et al. (2004) as a result of a study that aimed to establish if L1 had a privileged role in the acquisition of a foreign languages, or whether other non-native languages could influence this process. This model proposes the cumulative nature of language acquisition and influence on the L3. In other words, all the previously known languages cumulate and can enhance the process of learning an additional one. Moreover, according to the model, the previously acquired languages can only have a positive (facilitative) or neutral, but never negative, influence on a newly acquired language.

The L2 Status Factor Model (Bardel & Falk, 2007, 2012) assumes that the second language is the main source of influence on the L3 at the initial stages of acquisition, regardless of any typological and structural similarities between L2 and L3. The rationale behind this claim is that there is a greater cognitive similarity between L2 and L3 than between L1 and L3. Thus, speakers might rely on L2 more, as it is, like L3, a foreign language. What is more, Bardel & Falk (2012) also stress the role of declarative memory and metalinguistic awareness which are associated with foreign language learning. This model was found to be useful in explaining CLI in phonetics as well, and there are studies

14:6286298263

that provide evidence for L2 being the primary source of CLI in L3 (e.g., Llama et al., 2010; Wrembel, 2010; Marinho Ribeiro, 2021).

According to the Typological Primacy Model (TPM; Rothman, 2011, 2015), the initial state of L3 acquisition is governed by an internal parser that assesses structural proximity across previously acquired linguistic systems, once sufficient L3 input has been provided. The proximity is assessed based on linguistic input across a hierarchy of linguistic cues: the lexicon, followed by phonology and phonotactics, functional morphology and, finally, syntactic structures. Then, the language that is the most similar to L3 becomes the source of influence. Hence, both L1 and L2 can influence L3. Because transfer is conditioned by holistic typological similarity rather than property-by-property alignment, the TPM accounts for both facilitative and non-facilitative effects (Cabrelli Amaro and Rothman 2015).

In contrast to the wholesale transfer proposed by the TPM (Rothman, 2011, 2015), the Linguistic Proximity Model (LPM; Westergaard et al., 2017) opposes the idea of an exclusive source of influence. Grounded in a property-by-property architecture, the LPM posits that the learner's cognitive parser evaluates structural similarity at the level of individual linguistic features, rather than the entire systems. As a result, both the L1 and L2 remain simultaneously accessible as supplier languages, allowing CLI to operate selectively and yielding either facilitative or non-facilitative outcomes depending on the property.

Similarly, the Scalpel Model (Slabakova, 2017) conceptualizes transfer as highly selective, using the metaphor of a scalpel to illustrate the precise extraction from prior languages. While acknowledging that structural proximity may trigger initial CLI, Slabakova (2017) argues that the process is also governed by a number of modulating variables, such as property complexity, frequency, salience, and explicit knowledge. Consequently, both the Scalpel Model (Slabakova, 2017), as well as the LPM (Westergaard et al., 2017) shift from a global typological transfer towards more dynamic, feature-specific interactions across linguistic systems.

One framework that has been built around phonetic data is the Phonological Permeability Hypothesis (PPH; Cabrelli Amaro & Rothman, 2010). Focusing on regressive CLI, it posits that L1 and L2 are fundamentally different in terms of their stability. It is claimed that even the most proficient and native-like L2 is still less stable, and therefore more vulnerable to CLI, than L1 (Cabrelli Amaro, 2017). Consequently, the framework

15:1159129303

predicts CLI stemming from L3 to L2 as the proficiency in the L3 develops. While it does not preclude possible effects on the L1, it expects that L2 is more likely to be affected first.

To address speech development alongside existing (morpho)syntactic frameworks, the NGTA (Dziubalska-Kołaczyk & Wrembel, 2022, 2024) conceptualises multilingual phonology as a dynamic and emergent system. It unites the phonetic constraints of Natural Phonology (Donegan & Stampe, 2009) with the non-linear, usage-based principles of Complexity Theory (Kretzschmar, 2015), projecting commensurate developmental pathways rather than mutually exclusive models. Development operates across two processing levels; that is, an automatic articulatory routines (Level 1) and conscious metalinguistic control (Level 2). Moreover, it is governed by a hierarchy of linguistic factors (L1, Ln, typology) and environmental variables (input, proficiency, age). Given its specific orientation toward multilingual speech and gradient individual differences, NGTA provides the theoretical framework for the empirical investigation in Publication 3 (Cal, 2026b).

In summary, the discussed frameworks are largely restricted in scope to progressive CLI from L1 and/or L2 on the developing L3 and frequently examine one language in isolation. Yet, multilingual speech is inherently dynamic, characterized by continuous mutual interaction across the entire linguistic repertoire. What is more, existing models, except for the Scalpel Model (Slabakova, 2017) and the NGTA (Dziubalska-Kołaczyk & Wrembel, 2022, 2024), rarely take into account external factors that could shape the process of CLI. Most crucially, because these theories were originally developed for (morpho)syntax, their direct applicability to phonetic data remains limited (e.g., Llama et al., 2010; Wrembel, 2010; Marinho Ribeiro, 2021). To develop a comprehensive account of multilingual speech, the field needs theoretical tools that could capture continuous acoustic variation across all languages of a speaker and account for extralinguistic factors that might affect the process of CLI.

1.1.3. Voice onset time

Voice onset time (VOT) functions as the defining feature of stop consonants and is crucial to the empirical investigations in this dissertation. This section defines the parameter,

16:2181482887

stresses its applicability for exploring CLI, and provides general characteristics of laryngeal systems of the languages of interest to the current dissertation.

VOT is defined as the time interval between the release of a stop consonant and the onset of vocal fold vibration of a subsequent vowel (Lisker & Abramson, 1964). Given its acoustic robustness, VOT is one of the most extensively used acoustic metrics for investigating L2 speech (Hansen Edwards & Zampini, 2008). Across languages, stop voicing contrasts are organized into three primary phonetic categories; that is, voicing lead (negative VOT), short-lag (0 - ~30 ms), and long-lag (aspirated; >35 ms) (Lisker & Abramson, 1964). Being a continuous variable, measured in milliseconds, VOT provides a highly sensitive metric that is capable of capturing small shifts in category representations. In multilingual acquisition, it allows to detect even subtle phonetic drift, category assimilation or bidirectional interactions.

The languages under scrutiny in the present dissertation include Polish, English and Norwegian. These languages are complimentary in terms of their laryngeal features, which enables a systematic comparison of CLI and acquisition patterns. Table 1 presents VOT values for voiced and voiceless stops in the three languages, provided in the literature for word initial stress syllables.

Table 1 VOT values (in ms) for voiced and voiceless stops in Polish, English and Norwegian measured in word-initial stressed syllables.

Language	p	t	k	b	d	g
Polish (Keating et al., 1981)	21.5	27.9	52.7	−88.2	−89.9	−66.1
English (Lisker and Abramson, 1964)	58	70	80	1	5	21
Norwegian (Ringen and van Dommelen, 2013)	50	52	53	−80	−72	−73

Firstly, Polish is a true voicing language characterised by prevoicing in the voiced series of stops and short-lag VOT in voiceless stops (e.g., Keating et al., 1981; Jassem, 2003). The rate of prevoicing in Polish voiced stops is consistently high. For instance, Sypiańska (2021) documented 100% prevoiced tokens in native speakers of Polish, while Wojtkowiak (2022) observed that fully prevoiced items accounted for over 95% of all realisations.

Secondly, English is an aspirating language in which /b, d, g/ are typically realized as voiceless unaspirated (or partially voiced) and /p, t, k/ are aspirated (e.g., Docherty, 1992; Lisker & Abramson, 1964). It is worth noting that while prevoicing in English voiced stops has been observed, its prevalence varies considerably across the literature.

17:4403771183

For example, Docherty (1992) identified prevoicing in only 7% of tokens, whereas Lisker and Abramson (1964) observed it in approximately 20%, and Westbury (1979) reported rates as high as 57% among Texan speakers. Nevertheless, a fundamental difference remains between English and the other two languages, as prevoicing is not an obligatory cue for the voicing contrast in English.

Finally, regarding Norwegian, Czarnecki (2016) argues that the phonological contrast between /p, t, k/ and /b, d, g/ relies primarily on the presence versus absence of aspiration, which might suggest that it is also an aspirating language. However, phonetic evidence provided by Halvorsen (1998) and Ringen and van Dommelen (2013) reveals that voicing lead is present in voiced stops in approximately 80% and 40% of tokens, respectively. Consequently, in the present dissertation, Norwegian is categorized as a language exhibiting both prevoicing and aspiration, thereby aligning with Polish in the realization of voiced stops and with English in the realization of voiceless stops.

18:1082323960

Part 2: Research

This section introduces the core research of this dissertation. It opens with a description of primary research objectives, which is then followed by the presentation of the three peer-reviewed articles that constitute this thesis.

2.1. Research objectives

VOT has been one of the key acoustic metrics in examining CLI in multilingual speech acquisition. A substantial body of literature demonstrates that multilingual speakers frequently exhibit CLI during both production (e.g., Wrembel, 2015; Llama & Lopez-Morelos, 2016; Sypiańska, 2017; Gabriel et al., 2018; Geiss et al., 2022; Zhang, 2022; Zhang et al., 2023) and perception (e.g., Z. Liu, 2016; J. Liu et al., 2019; J. Liu and Lin, 2021; Stan, 2024; Zeng, 2024) of stop consonants. However, its magnitude and presence seem to vary across the findings and there is still no consensus how exactly CLI manifests itself in stop production across all languages of multilingual speakers.

Earlier L3 production research predominantly examined the voiceless series of stops (e.g., Wrembel, 2015; Llama & Lopez-Morelos, 2016; Sypiańska, 2017; Amengual, 2021), showing either progressive and/or regressive CLI or separate VOT patterns for the three languages (e.g. Sypiańska, 2017; Amengual, 2021). On the other hand, more recent studies have begun to extend this line of inquiry to the voiced stop counterparts, providing a more holistic perspective the VOT acquisition patterns in multilinguals (e.g., Gabriel et al., 2018; Geiss et al., 2022; Zhang, 2022; Zhang et al., 2023). The results of these investigations show an asymmetry between the two voicing series, with more CLI attested in the voiced stops and more accurate production of the voiceless series. However, further research is still needed on various language repertoires to confirm the previous findings and provide a broader perspective that would consider both voicing series and all languages involved simultaneously.

The acquisition of plosives by multilinguals has also been examined in perception, yet to a lesser extent. Investigations into VOT perception are usually focused on establishing the perceptual boundary location between homorganic voiced and voiceless stops.

19:6117630129

The results of existing investigations mainly point to CLI in perception (e.g., J. Liu et al., 2019; Stan, 2024), resulting in merged perceptual categories or boundary shifts; however, some studies also showed evidence of multiple perceptual systems (e.g, Z. Liu, 2016; Cal & Wrembel, 2023). It can, thus, be concluded that the area of VOT perception by multilinguals remains rather underexplored and there are mixed results regarding the manifestation of CLI in multilingual perception.

All things considered, previous research undoubtedly provides important insights into the field of VOT acquisition by multilinguals. However, the scope of existing investigations remains limited by several research gaps. Firstly, most L3 studies prioritise voiceless stops, leaving the voiced counterparts underexplored. Secondly, the predominant reliance on cross-sectional designs creates an empirical gap concerning how the entire trilingual system restructures and interacts over time during the initial stages of L3 development. Thirdly, while production has received increasing attention, the perception of voicing distinction in multilinguals remains underresearched, leading to limited understanding of how perceptual categories across all languages interact. Finally, prior work has relied predominantly on highly controlled read speech, overlooking whether CLI patterns persist in more ecologically valid spontaneous speech, and how extralinguistic factors modulate this process.

Hence, the present PhD dissertation aims to bridge these empirical gaps by employing a holistic approach to the acquisition of stops by multilingual learners across their entire linguistic repertoire (L1 Polish, L2 English, L3 Norwegian), across all three places of articulation of stops (bilabial, coronal, velar), and in both voicing series. The overarching objective of this PhD dissertation is to determine how CLI manifests itself and evolves across stop perception and production in multilingual speakers. To address this aim, the dissertation pursues three specific research objectives, operationalised across three thematically connected Publications:

- **Objective 1** is to bridge the gap in longitudinal research on multilingual speech by tracing the developmental acquisition of VOT in voiceless and voiced stops across three testing times during the first year of L3 learning, determining whether novice learners maintain separate language-specific categories over time or experience CLI across their three languages. This objective is addressed in Publication 1 (Cal & Wrembel, 2025).

20:7473457746

- **Objective 2** addresses the paucity of research on multilingual preception by determining how CLI affects perceptual boundary locations between voiced and voiceless stops in multilingual speakers. This objective is pursued in Publication 2 (Cal, 2026a).
- **Objective 3** is to assess CLI in VOT production and addresses the ecological validity gap by comparing read and spontaneous speech produced by multilingual speakers, additionally exploring the influence of external variables on the acquisition of stops. This objective is explored in Publication 3 (Cal, 2026b).

Each of the papers provides further detailed research questions, which are tackled in the sections that follow and in the publications themselves.

2.2. Publication 1 (Cal & Wrembel, 2025): Developmental acquisition of stops by multilingual speakers

The aim of Publication 1 (Cal & Wrembel, 2025) was to analyse the production of plosive sounds by L1 Polish – L2 English – L3 Norwegian speakers with particular focus on tracing the development of VOT acquisition in voiceless and voiced stops (/p, t, k/ and /b, d, g/) in all three languages (Polish, English and Norwegian) produced by novice L3 learners throughout their first year of learning the third language. In other words, the goal was to explore possible cross-linguistic interactions in the oral productions of multilingual speakers with a focus on both progressive and regressive CLI.

The research questions guiding Publication 1 (Cal & Wrembel, 2025) were as follows:

- **RQ1:** What sources of CLI can be traced in the VOT patterns in the three languages?
- **RQ2:** How does the acquisition of VOT in trilinguals change over time in the respective languages?
- **RQ3:** Do voiced and voiceless plosives exhibit similar trends across languages in the multilinguals' repertoire?

To address the research questions, a group of 12 L1 Polish – L2 English – L3 Norwegian speakers (mean age = 20) was recruited. The participants were first-year students of Norwegian Studies at two Polish universities. (in Szczecin and Poznań). They

21:5135581370

were intermediate (around B2) learners of L2 English and beginner (A1) learners of L3 Norwegian.

The study was a longitudinal investigation consisting of three separate testing times. The first testing time (T1) was conducted after 8 weeks of learning the L3. The second testing time (T2) was carried out after the first semester of learning the L3, i.e. after approximately 5 months. The third testing time (T3) took place after two semesters, that is, after 7 months of learning L3 Norwegian. In each session, the participants were recorded reading word lists with stop sounds (voiced and voiceless) placed in stressed onset positions followed by a vowel in all three languages. The stimuli included isolated one-syllable words with stop sounds (fortis and lenis) placed in stressed onset positions followed by a vowel. The words with fortis and lenis stops constituted minimal pairs that were elicited in random order. The stimuli were presented in a form of PowerPoint presentations and the recordings for each language block were conducted on different days to exclude a potential language-mixing effect. The obtained audio files were force-aligned using WebMAUS (Kisler et al., 2017) and VOT boundaries were manually corrected in Praat (Boersma & Weenink, 2021). The data was analysed using a logistic mixed-effects regression model in R employing a partially Baesian method within the blme package (Chung et al., 2013) to avoid the singularity issue. The model specification was as follows: VOT ~ Sound*Language*-Session + (Session|Participant) + (Session|Item).

The results showed that L1 Polish voiceless stops consistently exhibited the shortest VOT durations, while L2 English voiceless stops yielded the longest values, and L3 Norwegian values were in between the two. In the voiced series, L1 Polish maintained the longest prevoicing, whereas L2 English exhibited the shortest values. Furthermore, the descriptive data showed that VOT durations generally changed over time. Specifically, voiceless stops lengthened, while prevoicing shortened with time across all three languages. In addition, the number of unvoiced lenis tokens increased considerably over time in L2 English and L3 Norwegian, but not in L1 Polish.

The statistical analysis revealed language-specific VOT patterns in all three languages, which means that the VOT production patterns differed significantly across the language systems and that CLI was overall not attested in the dataset. However, looking from a broader perspective, a change in VOT production patterns over the three testing times was well visible, especially with regards to the voiced series of stops, which turned out to be produced with fewer prevoiced items in all three languages as the L3 learning

22:8498736595

progressed. What is more, the study also showed the differences in the behaviour of voiced and voiceless stops, with the former being affected by CLI to a greater extent. This finding validates the existence of the voiced-voiceless stop asymmetry previously largely attested in L2 speech research (e.g. Zając, 2015; Schwartz, 2020; Wojtkowiak, 2022). Additionally, the study also revealed a considerable individual variation, especially in the case of the voiced stops.

Summarising, Publication 1 (Cal & Wrembel, 2025) demonstrated that VOT production in multilinguals follows distinct patterns across the three languages. However, its longitudinal acquisition is a dynamic process influenced by cross-linguistic interactions within the language system, affecting the two series of stops differently. The publication constitutes an important contribution to the growing body of research on multilingual speech, as it offers crucial insights into the development of speech over time. Moreover, it is one of the first longitudinal investigations in the field that provide a holistic perspective, exploring of all three language systems of the speakers in two voicing series of stops. See Appendix G for the complete manuscript.

2.3. Publication 2 (Cal, 2026a): Crossing the boundaries: Perception of voice onset time in word-initial stops by multilingual learners

The primary goal of Publication 2 (Cal, 2026a) was to scrutinise the perception of word-initial voiced and voiceless stops in the three languages of L1 Polish – L2 English – L3 Norwegian speakers. Of particular interest was to establish whether the perceptual boundary locations between voiced and voiceless plosives (measured by VOT in ms) for the trilingual speakers are subject to CLI and to explore the directionality of the CLI process. What is more, the study also offered a comparison with native speakers' norms for the three languages as well as an individual variation analysis.

The following research questions were posited in Publication 2 (Cal, 2026a):

- **RQ1:** Is VOT perception in multilinguals subject to the process of CLI? In what way is it different from native speakers' perception?
- **RQ2:** What is the directionality of CLI in the VOT patterns of L1 Polish – L2 English – L3 Norwegian speakers? Is its presence dependent on the PoA of stops?
- **RQ3:** How do VOT boundary locations vary for individual participants?

To answer the research questions, the following procedure was established. First of all, VOT continua were prepared separately for each language and place of articulation. They ranged from -100 to 100 ms, consisted of 21 steps (each of 10 ms) and were created using a Praat script (Winn, 2020). Then, the tokens were used in a two-alternative forced-choice (2AFC) task, in which the participants were presented with one random word from the continuum and asked whether they heard a voiced or voiceless consonant at the beginning. Every token appeared five times in a randomized order. The experiment was conducted in three separate language blocks. Each one lasted around 15min and was preceded by a language mode introduction.

The participants included 30 L1 Polish – L2 English – L3 Norwegian speakers (aged 21) studying at the Norwegian Studies programme at two Polish universities (in Szczecin and Poznań). Their English proficiency ranged from B1 to C2, while Norwegian proficiency was in the range of A2 to B1+. Their exposure to Norwegian was extensive, as they were taking part in practical Norwegian classes, including speaking, writing, listening, and pronunciation. The experiment was also carried out on three groups of control speakers of respective languages; that is, on 21 L1 Polish (mean age = 22), 18 L1 English (mean age = 24) and 10 L1 Norwegian (mean age = 25) functional monolinguals.

The perceptual boundary between voiced and voiceless stops was determined in the place where half of the participants' responses indicated a voiced sound and the other half the voiceless one. The statistical analysis was performed using a Bayesian mixed-effects logistic regression model, fitted using the brms package (Bürkner, 2017) in R. The model predicted binary response outcomes (0 indicating a voiced stop and 1 indicating a voiceless stop) using a Bernoulli likelihood with a logit link to assess the effects of VOT, language, place of articulation and group, as well as all possible interactions among these predictors. Random effects included by-participant varying intercepts, as well as correlated slopes for VOT, place of articulation, and their interaction. The model specification was as follows: resp ~ VOT_scaled * language * PoA * group + (1+VOT_scaled*PoA | participant).

The results indicated credible differences between multilingual and control groups in the /b-p/ continuum, where the multilinguals responded with a greater number of voiced stop choices than the controls in L2 English and L3 Norwegian. It means that, in these two languages, their perceptual boundary between the /b/ and /p/ sounds was placed earlier in the continuum than in the case of controls. On the contrary, no differences were

24:3042630949

found between multilinguals and controls in Polish. Taking into consideration the performance of the multilingual group only, the findings pointed to credible differences across the languages in the /b-p/ and /d-t/ continua; that is, the probability of categorising VOT stimuli as voiced was the highest for Polish, followed by English and Norwegian; however, the magnitude of the effect was greater in the /d-t/ continuum. On the contrary, no credible differences were found across the three languages in the /g-k/ continuum. The study also uncovered considerable inter- and intra-individual variation among the participants with patterns of VOT categorization, confirming the complex nature of CLI.

To sum up, Publication 2 (Cal, 2026a) demonstrated that multilingual speakers experience CLI also in perception; however, it turned out that its nature is rather complex, and the directionality and magnitude depend on the place of articulation of stops. That is, in bilabial stops, the participants were found to undergo an L1-induced CLI; in coronal stops, they maintained separate perceptual systems; whereas in velar stops, they demonstrated an L3-driven merged perceptual category. The full text of the paper can be found in Appendix H.

2.4. Publication 3 (Cal, 2026b): Is L2 the least stable system? The dynamics of VOT trilingual production in Polish-English-Norwegian read and spontaneous speech

Publication 3 (Cal, 2026b) set out to investigate the production of voiceless and voiced stops in both read and spontaneous speech in the three languages of L1 Polish – L2 English – L3 Norwegian speakers, as well as native controls of the respective languages. The specific objective of the study was to analyse CLI across the three languages of multilingual speakers and to establish its potential sources and directionality. What is more, the paper also aimed to integrate production data with extralinguistic factors, including the level of proficiency, exposure, frequency of use and age of acquisition to provide a deeper understanding of the dynamics of CLI in three phonological systems. Additionally, the study also provided baseline data for VOT in Polish, English, and Norwegian spoken by functional monolinguals.

Publication 3 (Cal, 2026b) strived to answer the following research questions:

25:1146540523

- **RQ1.** Do multilingual speakers of L1 Polish – L2 English – L3 Norwegian exhibit CLI in the production of stops? Are there any differences in their VOT production compared to native control speakers?
- **RQ2.** How is CLI in VOT production of stops manifested across the three languages of L1 Polish – L2 English – L3 Norwegian speakers? Is it dependent on the place of articulation of stops and/or voicing?
- **RQ3.** What is the difference in VOT produced in read vs. spontaneous speech? Are CLI patterns attested to an equal extent in the read speech reflected in spontaneous speech?
- **RQ4.** What is the impact of extralinguistic factors on the production of stops by L1 Polish – L2 English – L3 Norwegian? Which factors (and to what extent) affect the presence or ab-sence, as well as the degree of CLI?

To resolve the questions driving this study, two speech tasks were conducted for Polish, English, and Norwegian. The first task was a reading task of 60 sentences per language The stimuli targeted word-initial stop sounds in a stressed onset position that was controlled for vocalic context. The target words were embedded into structurally similar sentences. The second task elicited spontaneous speech and involved speaking for approximately 2 minutes on a designated topic. To prevent language mixing, sessions occurred on different days, each preceded by a 3-minute video retelling task to set the language mode.

The study was conducted on a group of 37 L1 Polish – L2 English – L3 Norwegian speakers, aged 21, who were students or graduates of Norwegian Studies at two Polish universities (in Poznań and Szczecin). On average, the participants were B2 speakers of L2 English and B1 speakers of L3 Norwegian. The experiment also involved three control groups consisting of native speakers of the three languages; that is, there were 20 L1 Polish speakers (mean age = 22), 18 L1 English speakers (mean age = 24), and 12 L1 Norwegian speakers (mean age = 25). The experimental group completed the tasks in all three languages, while the control groups completed them only in their native language.

The results revealed distinct acoustic patterns across the three languages. In L1 Polish, voiceless stops were produced with the shortest VOTs and were highly comparable between the multilinguals and native controls, while the voiced stops exhibited shorter prevoicing in the multilingual group compared to the target. In L3 Norwegian, the multilinguals produced voiceless stops with slightly shorter aspiration than native controls,

26:4481849795

whereas their voiced stops displayed somewhat longer prevoicing and a higher probability of prevoicing realisation compared to native Norwegians. Meanwhile, the most substantial deviations were observed in L2 English. That is, multilinguals produced voiceless stops with considerably shorter aspiration and voiced stops with heavy prevoicing, starkly contrasting with the near-zero VOTs of native English controls. Similar relational patterns were observed in the spontaneous speech data, though with generally shorter VOT in voiceless stops and greater variability across all categories.

The results from read speech analysis demonstrated that the multilinguals significantly differed from controls in their production of VOT in L2 English, while in L1 Polish and L3 Norwegian they seemed to produce the stops in a more target-like fashion in both voiced and voiceless series of stops. This suggests a great deal of influence stemming from L1 on L2 in both series of stops, which could have been further enhanced by a newly learnt L3. The findings also revealed the asymmetry between the two series of stops in terms of their susceptibility to CLI. It was found that the voiced series is more prone to undergo assimilative CLI, resulting in a difficult-to-suppress prevoicing in all three languages, while the voiceless stops were easier to acquire on target.

In spontaneous speech, multilinguals were able to produce voiceless stops with longer VOTs in all three languages compared to the read speech, which could be attributed to the speech rate. At the same time, in the voiced series, the participants maintained prevoicing of a similar length in all three languages. This finding could be caused by reduced phonetic control, which is also inherent to spontaneous speech.

Additionally, the data was interpreted within the framework of NGTA (Dziubalska-Kołaczyk & Wrembel, 2022, 2024). The analysis pointed to the dynamicity and complexity of the multilingual system in a specific acquisition situation, which was modulated by a hierarchy of extralinguistic factors. The predictors of VOT acquisition depended heavily on the voicing series. For the more stable voiceless stops, language remained the primary predictor, followed by L2 age of acquisition, L3 proficiency, and speech mode. Conversely, for the highly permeable voiced stops, L2 proficiency, L2 age of acquisition, and the frequency of L3 use emerged as the most important factors.

Overall, Publication 3 (Cal, 2026b) provided a comprehensive investigation into CLI in VOT production by L1 Polish – L2 English – L3 Norwegian speakers. The study evaluated how stop production differs across groups, languages and speech styles, showing that, in this particular linguistic repertoire, non-facilitative CLI was attested in the

27:1123721691

case of L2 English, the voiced series of stops and spontaneous speech. Furthermore, the research integrated key extralinguistic factors, such as proficiency, frequency of use, and exposure to map the dynamics of CLI in multilingual phonological systems. The full paper can be found in Appendix I.

28:1157648047

Part 3: General discussion

The third part of this dissertation provides a general discussion of the selected key findings from Publication 1 (Cal & Wrembel, 2025), Publication 2 (Cal 2026a) and Publication 3 (Cal 2026b). The results obtained from the production and perception studies are juxtaposed and the findings from the longitudinal and cross-sectional studies are compared with additional focus on speech modes, acquisition models as well as individual variation. Thereafter, the section provides implications, limitations, and future research directions.

3.1. Production data

Production data were analysed within Publication 1 (Cal & Wrembel, 2025) and Publication 3 (Cal 2026b). The former provided a longitudinal account, while the latter was a cross-sectional study. The findings from both articles demonstrate that multilingual phonetic acquisition cannot be reduced to simple unidirectional cross-linguistic influence (CLI). Instead, multilinguals actively maintain separate phonetic systems across their three languages, while simultaneously exhibiting more targeted and feature-specific cross-linguistic interactions.

In Publication 1 (Cal & Wrembel, 2025), learners displayed statistically distinct VOT patterns across L1 Polish – L2 English – L3 Norwegian, which corroborated previous evidence that multilinguals maintain separate linguistic systems when it comes to the production of stops (Z. Liu, 2016; Amengual, 2021; Geiss et al., 2022). However, it was also revealed that these distinct systems shifted concurrently over time (between T1 and T3), which was especially visible in the shortening of prevoicing across all three languages. This points to the emergence of a holistic "global language entity" (cf. Sypiańska, 2017), in which adding an L3 induces a dynamic change affecting the multilinguals' system as a whole.

Juxtaposing this longitudinal development with the cross-sectional findings of Publication 3 (Cal, 2026b) provides a comprehensive picture of CLI directionality, particularly regarding the vulnerability of L2 English. In Publication 3 (Cal, 2026b),

29:4761311658

participants successfully produced short-lag stops in L1 Polish and near-native long-lag stops in L3 Norwegian (which did not differ significantly from native Norwegian controls). In contrast, their L2 English VOT yielded intermediate values. While it was significantly longer than L1 Polish VOT, it still remained too short to reach native English targets. This pattern points to a complex, bidirectional dynamic where L2 English is simultaneously constrained by progressive L1-driven CLI and reshaped by regressive L3-driven CLI.

This interpretation gains strong longitudinal support from Publication 1 (Cal & Wrembel, 2025). Using T1 data as a baseline proxy for L1 Polish-L2 English bilinguals prior to substantial L3 acquisition, Publication 1 (Cal & Wrembel, 2025) demonstrated that learners already possessed relatively high L2 English VOT values. Crucially, their L2 English VOT durations lengthened further across subsequent sessions (i.e., T2 and T3) alongside intensive Norwegian instruction, despite limited communicative exposure to English. This points to a cumulative effect of L3 learning, where multilinguals likely detect the phonetic similarity between two shared long-lag systems (English and Norwegian) and assimilate VOT values across them, or conversely, struggle to fully reach long-lag targets due to deeply rooted L1 short-lag motor routines (Bandeira & Zimmer, 2012; Llama & Cardoso, 2018). The prevalence of CLI in L2 English might be caused by the lack of phonetic training in L2 English, as the multilinguals attended Norwegian pronunciation classes in their Norwegian studies programme, but did not receive any explicit pronunciation training in L2 English, which could have led to smaller awareness of the English pronunciation rules.

3.1.1. VOT across speech modes

One of the main aims of Publication 3 (Cal, 2026b) was to establish any potential differences in VOT production between read and spontaneous speech. The results demonstrated that phonetic CLI and category distinctiveness were modulated by speech mode. While read speech fostered more monitored, target-like articulations, spontaneous speech triggered greater acoustic variability and heightened vulnerability to CLI. Participants successfully preserved the primary phonological contrast between short-lag L1 Polish VOT vs. long-lag L2 English and L3 Norwegian in both tasks; however, spontaneous

30:8405461673

production yielded significantly longer VOTs in voiceless stops as well as the persistence of prevoicing across all three languages. Although, the participants produced relatively longer and, what might seem like, more target-like VOT values in their L2 English and L3 Norwegian voiceless stops, this acoustic lengthening could partly stem from reduced cognitive control (Jerotijević Tišma, 2019) and lower speech rate, which is a characteristic of less proficient, newly acquired languages. As communicative and processing demands increase in spontaneous interaction, attentional control over fine-grained laryngeal timing diminishes, resulting in a bidirectional CLI.

At the same time, this interaction between task type and laryngeal timing remains complex and potentially confounded by speech rate. Because native control speakers communicating casually in their L1 typically exhibit higher speech rates this naturally leads to shorter VOT values. The longer VOT durations observed in the multilingual participants' L2 and L3 may partially reflect a slower rate of speech rather than category expansion per se. Consequently, more empirical data from a larger pool of participants as well as diverse and more controlled speaking tasks, which systematically measure speech rate, are still needed.

3.1.2. Asymmetry between voiceless and voiced stops

Both Publication 1 (Cal & Wrembel, 2025) and Publication 3 (Cal, 2026b) attested an asymmetry between the voicing series in the production of stops. Across both studies, voiceless plosives exhibited higher stability and more target-like VOT realisations across languages, whereas voiced plosives were characterised by prevalent CLI, and longitudinal shifts. Specifically, in Publication 1 (Cal & Wrembel, 2025), the longitudinal investigation revealed significant compression of prevoicing and a growing proportion of unvoiced tokens in voiced stops across all three languages between T1 and T3, while the voiceless series exhibited more stability over time. Similarly, in Publication 3 (Cal, 2026b), group comparisons against native controls demonstrated that non-target-like realisations were predominantly concentrated in the voiced series, most notably in L2 English, where learners produced non-nativelike prevoicing. These results are in line with the findings from the second language acquisition research attesting an asymmetrical

31:1010449295

behaviour of the two series of stops (e.g., Wojtkowiak, 2022; Schwartz et al., 2020; Bless, 2015).

There are several factors that could account for this asymmetry. The first one is a possible greater perceptual salience of voiceless stops. Aspiration noise in fortis plosives is a robust and perceptually prominent feature. Therefore, the contrast between short-lag and long-lag stops can be readily noticed by learners, which facilitates the establishment of separate phonetic categories (Flege, 1995; Flege & Bohn, 2021). On the contrary, voicing lead in lenis stops has lower perceptual salience and it could be more difficult for learners to detect subtle differences between prevoicing and short lag. This could lead to the lenis series being less monitored and thus more susceptible to category assimilation.

Secondly, the production of prevoicing is more demanding, as it requires a specific articulatory gesture that involves laryngeal voicing and an oral closure (Zhang, 2022; Zhang et al., 2023). Because of that, speakers may experience greater difficulty establishing a negative VOT category or controlling it consciously. Therefore, the participants struggled to suppress prevoicing in L2 English, resulting in the persistence of prevoicing habits stemming from L1 Polish. What is more, from the perspective of phonology, as described within the Onset Prominence theory (cf. Schwartz, 2020), prevoicing is considered to be a low-level phonetic detail, while aspiration is an easier-to-notice phonological feature.

Finally, it is important to also consider pronunciation instruction. The participants from both Publication 1 and 3 received explicit pronunciation instruction exclusively in L3 Norwegian, which distinguishes between prevoicing in voiced stops and aspiration in the voiceless ones. Therefore, no explicit instruction was given on how to suppress the voicing in L2 English, leaving the English voiced stops vulnerable to CLI from both L1 Polish and L3 Norwegian.

3.2. Perception data

Publication 2 (Cal, 2026a) aimed to complete the picture by dealing with stop perception. The findings demonstrated that VOT perception in multilinguals is subject to CLI with both its presence and directionality strongly conditioned by the place of articulation and cross-language similarity. Statistical analyses revealed credible differences between the

32:6673926476

multilingual experimental group and native control speakers, indicating that multilinguals perceive stop consonants differently than controls. Descriptive statistics showed that the multilingual group tended to place their perceptual boundaries across languages closer together than the control groups, pointing to an overall assimilation of categories. Further pairwise comparisons demonstrated that these group differences were driven by the place of articulation, revealing context-specific instances of CLI rather than a uniform transfer across the language inventory.

Importantly, cross-language similarity emerged as a primary factor affecting perceptual boundary locations, which provides further empirical confirmation for the key tenets of the Speech Learning Model (SLM; Flege, 1995; Flege & Bohn, 2021). Rather than showing a uniform pattern across the whole stop inventory, in the case of L1 Polish – L2 English – L3 Norwegian participants, each place of articulation is characterised by a varying degree of cross-language similarity and perceptual salience. In line with the SLM, these differences determine whether sounds are perceived as similar and undergo equivalence classification into a single merged category, or whether they are perceived as distinct, leading to the formation of separate phonetic categories across the bilabial, coronal, and velar continua.

In the bilabial continuum, the multilinguals maintained Polish-like identification patterns in their three languages, which suggests considerable L1-driven influence on the remaining foreign languages. Because the acoustic and articulatory features of bilabial stops are very similar across languages, listeners might not have heard the fine-grained differences between them. In accordance with the SLM (Flege, 1995; Flege & Bohn, 2021), this high cross-language similarity leads to equivalence classification, resulting in a single, merged perceptual category for the /p-b/ continuum across the three languages which is governed by L1 Polish-induced progressive CLI.

The opposite pattern can be noticed in the coronal continuum, which has a lower cross-linguistic similarity and higher perceptual salience, due to the articulatory differences. Coronal stops are dental in Polish (Dukiewicz & Sawicka, 1995), alveolar in English (Roach, 2009), and highly variable in Norwegian (Kristoffersen, 2000). These differences could have prevented equivalence classification and allowed the participants to establish separate, language-specific perceptual categories. This tendency was clearly visible in the data, as the /t-d/ identification patterns of multilinguals did not differ credibly from native control baselines in any language, showing native-like category separation.

33:6767352591

Yet another tendency emerged in the velar continuum, where the multilinguals relied on longer VOT values for stop identification across all three languages compared to control baselines. Velar stops naturally display longer VOT durations across the three languages, which increases their cross-language similarity in the fortis series. Rather than preserving native-like L1 boundaries for the /k-g/ continuum, prolonged exposure to long-lag L2 and L3 stops could have resulted in a merged perceptual category driven, this time, by regressive, cumulative L2- and L3-induced CLI as well as an attested L1 drift.

3.3. Informing theoretical models of multilingual acquisition

One of the main goals of Publication 3 (Cal, 2026b) was to further test the applicability of the Natural Growth Theory of Acquisition NGTA (Dziubalska-Kołaczyk & Wrembel, 2022, 2024) to multilingual speech acquisition. Moreover, it seemed vital to evaluate other theoretical frameworks proposed in the L3 domain in terms of their relevance to predict CLI behaviour in multilingual speech. The empirical data stemming from all three publications (Cal & Wrembel, 2025; Cal, 2026a; Cal, 2026b) could serve to provide some new insights into the predictive power of the existing third language acquisition models.

Overall, the results seem to fail to support models positing the wholesale transfer, such as the Typological Primacy Model TPM (Rothman, 2010, 2015). As it was attested in the current data, the participants maintained distinct, language-specific VOT categories in voiceless stops production and coronal perception, while simultaneously exhibiting hybrid realisations of the voiced series, as well as perceptual L1-driven CLI in bilabials, and L3-driven CLI in velars. It can, therefore, be stated that phonetic acquisition is property-dependent, selective, and conditioned by linguistic factors.

Furthermore, the results do not support models that predict only facilitative CLI, such as the Cumulative-Enhancement Model (CEM; Flynn et al., 2004). As observed in L2 English productions, participants largely maintained non-target-like realizations of voiced stops. Although this outcome may have been modulated by the cumulative effect of learning two languages with prevoicing, it clearly demonstrates the occurrence of non-facilitative CLI.

When examining voiceless stops, the results offer compelling evidence for the privileged role of the L2 status factor (Bardel & Falk, 2007, 2012). Because L1 Polish

34:4661166158

relies on short-lag VOT and lacks aspiration, L2 English could serve as the primary supplier for long-lag aspiration in L3 Norwegian. This demonstrates that learners actively relied on, previously acquired, longer VOT in their L2, rather than their native language when establishing aspirated categories. However, it cannot be unequivocally stated if this influence reflects complete, wholesale transfer of the L2 phonological system, as the participants already exhibited non-target-like prevoicing in their L2 English productions. It is, thus, impossible to determine the exact source of prevoicing in L3 Norwegian. Specifically, whether it was transferred directly from the L1 Polish, or whether it was supplied via the L2 English system, which had already undergone persistent L1 influence.

From a different perspective, production data might also align well with the models that assume property-by property transfer, such as the Linguistic Proximity Model (LPM; Westergaard et al., 2017) and the Scalpel Model (Slabakova, 2017), with the latter also positing the conditioning influence of external variables. It cannot be ruled out that the results show selective transfer of specific phonetic features across the repertoire. That is, prevoicing in Norwegian voiced stops could have been reinforced by L1 Polish, whereas aspiration in voiceless stops could be influenced by L2 English. In addition, the Random Forest analysis in Publication 3 (Cal, 2026b) confirmed that this process is modulated by extralinguistic factors such as exposure and language use, directly supporting the Scalpel Model (Slabakova, 2017). At the same time, Publication 2 (Cal, 2026a) revealed the importance of cross-linguistic similarity in multilingual perception posited within the SLM (Flege, 1995; Flege & Bohn, 2021).

These frameworks, however, focus on progressive CLI into the L3 and do not consider the dynamicity of the multilingual system as a whole. With respect to regressive CLI, the results reported in Publication 3 (Cal, 2026b) only partially corroborate the Phonological Permeability Hypothesis (PPH; Cabrelli Amaro & Rothman, 2010). Although L2 English proved to be the most susceptible to CLI, especially regarding voiced stops, this permeability was largely attributable to L1 rather than L3. This L1 influence was also visible in T1 of Publication 1 (Cal & Wrembel, 2025), where participants produced prevoicing in voiced stops at the very start of their L3 Norwegian exposure. Consequently, these findings highlight the need to look at the bigger picture. Rather than focusing on unidirectional effects, researchers must account for continuous interactions across the entire multilingual repertoire.

35:1622375397

The present data also lend strong empirical support to the Natural Growth Theory of Acquisition (NGTA; Dziubalska-Kołaczyk & Wrembel, 2022, 2024). The theory posits that cross-linguistic interactions are dynamic, continuous, and conditioned by a wide array of linguistic and extralinguistic factors. In line with this perspective, the CLI patterns observed across Polish, English, and Norwegian cannot be accounted for by static transfer models relying on a single source-recipient relationship. Instead, the considerable individual variability substantiates the NGTA's view that phonetic systems organically evolve and co-exist within a multilingual repertoire.

Concluding, the existing models of TLA appear not to be fully sufficient to account for the process of multilingual speech development, if they focus exclusively on linguistic factors and progressive CLI. Instead, the field requires a more comprehensive framework that would account for extralinguistic variables (such as proficiency, age of acquisition, and frequency of use, among others) alongside cross-language similarity, while fully accommodating the dynamics of multilingual development.

Another important reason why (morpho)syntactic frameworks struggle to account for phonetic data is the nature of the data itself. While syntactic research usually operates on discrete, categorical structures, phonetic data are, more often than not, continuous. Therefore, it can be subject to subtle cross-linguistic acoustic interactions that categorical models cannot easily capture, such as the shortening or lengthening of aspiration. Yet, this does not come without a cost, which is that such small shifts are extremely difficult to categorise into a single CLI source-recipient relationship, especially if found across all language systems. The ideal model of multilingual speech acquisition would need to operate more globally, accounting for all languages of a speaker, their similarities, but also a variety of external factors guiding the process of speech acquisitions by multilinguals. So far, the Speech Learning Model (SLM; Flege, 1995) and its revision (Flege & Bohn, 2021) seem to effectively predict CLI patterns in multilingual speech based on perceived cross-language similarity. The NGTA (Dziubalska-Kołaczyk & Wrembel, 2022, 2024) can also be useful if we want to broaden the perspective to extralinguistic variables and their impact on CLI. However, more analyses are still needed to determine the importance of certain linguistic and extralinguistic factors in shaping multilingual speech.

36:7437983764

3.4. Individual variation

Individual variation in the acquisition of stops was assessed in Publication 1 (Cal & Wrembel, 2025) and Publication 2 (Cal 2026a). The articles provide data from a longitudinal investigation of stop production and a cross-sectional study on stop perception, respectively. Besides the group tendencies, the analysis of individual profiles revealed substantial inter- and intra-speaker variation that highlights the non-linear, dynamic nature of multilingualism (Gut et al., 2023; Wrembel et al., 2022).

The longitudinal production study in Publication 1 (Cal & Wrembel, 2025) uncovered different individual developmental paths. In the voiceless series, some participants maintained separate patterns across their three languages, whereas others assimilated L2 English and L3 Norwegian, simultaneously separating L1 Polish. In the voiced series, individual developmental trajectories between T1 and T3 included non-linear upturns and downturns, with the greatest shifts frequently occurring at T2. In the perception study, presented in Publication 2 (Cal, 2026a), individual variation data also revealed distinct CLI profiles rather than a uniform group pattern. While certain speakers demonstrated a complete L1-L2-L3 category overlap across continua, others exhibited isolated L1 phonetic drift (cf. Chang, 2019), distinct L2-L3 merger separate from the L1, or L1-L3 influence.

These individual patterns corroborate that multilingual acquisition is dynamic and variable, uncovering the need to integrate individual differences into models of multilingual speech acquisition. The substantial variability attested in VOT production and perception could be shaped by individual differences in L2/L3 proficiency, frequency of use, language aptitude, and auditory sensitivity, among others. Therefore, it seems vital to include such external factors into the investigations and theoretical reasonings.

3.5. Implications

Research on single speech sounds can be seen as rather abstract by those outside the field, often prompting questions such as “What does it really mean?” or “Why does it matter?”. While the purpose of the research might seem obvious to phoneticians, for non-

37:5872642895

professionals it might not be so. Nevertheless, the findings of this PhD dissertation carry both theoretical and practical value.

For phoneticians, it provides much-needed up-to-date baseline VOT data for native speakers of Polish, English, and Norwegian, while also offering empirical data on non-native production and perception. Furthermore, these results underline the need to expand research into new, understudied language combinations and move toward a more holistic framework that takes into account cross-language similarity, individual variation, and the influence of extralinguistic factors within the dynamics of multilingualism.

For language teachers and learners, the findings have clear pedagogical implications. The research identified which languages and sounds are more vulnerable to CLI, and consequently, more likely to contribute to accented speech. Bringing these findings into the pronunciation classroom can raise phonetic awareness and help teachers design more effective pronunciation practice tailored to multilingual learners.

Finally, for the broader public, this thesis can serve as a tool to raise awareness of the prevalence and nature of multilingualism. With the majority of the world's population speaking at least two languages (Grosjean, 2024), scrutinizing cross-linguistic influence can show that non-canonical phonetic realisations are not errors, but natural manifestations of multi-competence. In essence, everyone speaks with an accent and nowhere is this phonetic reality more pervasive than in multilingual speech.

3.6. Limitations and future research directions

Notwithstanding the contributions presented in the preceding sections, the current project was not free from limitations. The following section outlines methodological constraints encountered across the three publications, pairing each one with recommendations for future research.

The primary limitation across the studies concerns the pool of participants, especially in Publication 1 (Cal & Wrembel, 2025). The longitudinal design led to a reduced number of participants due to a considerable drop-out rate after T1, which constrained the statistical power needed for fitting the models with three-way interactions. To address this issue, future research should recruit larger initial cohorts to prevent the participant loss linked to longitudinal studies. Similarly, Publication 3 (Cal, 2026b) would have

38:1137831684

benefited from a larger sample size for both the multilingual group and the native controls to ensure that the number of participants matched for specific extralinguistic factors is more comparable.

The second limitation relates to the elicitation tasks and speech modes. While read speech lists with word-initial stressed onsets provided high phonetic control in Publication 1 (Cal & Wrembel, 2025), they lacked the ecological validity of natural speech. Furthermore, when comparing sentence reading and spontaneous speech in Publication 3 (Cal, 2026b), the dataset suffered from an unbalanced number of observations across the two speech modes, alongside potential speech rate differences that were not extensively measured. Future research should implement more structured elicitation tasks, such as picture-based storytelling, that combine ecological validity with targeted elicitations of specific sounds. Future designs should also balance the number of tokens between speech modes and control for speaking rate across tasks.

Yet another source of weakness in this dissertation concerns the perception study in Publication 2 (Cal, 2026a). The experiment would benefit from a greater control over stimuli preparation to avoid a potential influence of secondary voicing cues on the perception of VOT. Even though the issue was addressed by adjusting F0 values, further attention to this aspect would likely enhance the validity and interpretability of future finding. What is more, the investigation could also consider the variability in participants' L2 and L3 proficiency levels, especially when it comes to the individual variation across the participants. Other individual differences could also be taken into consideration together with the development of perception trajectory over time.

Another limitation involves the composition of the control groups. While native controls were tested for Polish, English, and Norwegian across the studies, Publication 3 (Cal, 2026b) could have further benefited from an L1 Polish – L2 English bilingual control group. Including such participants in future studies would help isolate whether the observed CLI in L2 English voiced stops stems strictly from L1 Polish or is reinforced by L3 Norwegian.

Finally, in future research the empirical scope should be expanded beyond the current triad of L1 Polish–L2 English–L3 Norwegian to include speakers of different linguistic repertoires and other sounds, thereby testing the generalisability of the findings presented in this PhD dissertation.

39:5170184738

Conclusion

The present PhD dissertation aimed to provide a comprehensive analysis of stop production and perception by multilingual speakers of L1 Polish – L2 English – L3 Norwegian. To this end, a longitudinal study on VOT production (Publication 1; Cal & Wrembel, 2025) and two cross-sectional studies, one on VOT perception (Publication 2; Cal 2026a) and one devoted to VOT production (Publication 3; Cal 2026b), were conducted. Overall, the findings demonstrate that multilingual phonetic acquisition is neither static nor governed by wholesale transfer from a single language system. Instead, all languages of trilingual speakers operate as a dynamic, interconnected system, where cross-linguistic interactions are conditioned by voicing series, place of articulation, speech mode, similarity, and individual variation.

In the production domain, the findings from Publication 1 (Cal & Wrembel, 2025) and Publication 3 (Cal, 2026b) show that while learners maintain separate, language-specific VOT patterns across their three systems, these systems undergo subtle, continuous changes. The longitudinal investigation (Cal & Wrembel, 2025) demonstrated that the acquisition of L3 triggers a shortening and, in some cases, complete loss of prevoicing across all three languages. This entails the adaptation of a global multilingual repertoire rather than an isolated addition of the third language to the system. Furthermore, both Publication 1 (Cal & Wrembel, 2025) and Publication 3 (Cal, 2026b) revealed an asymmetry between the two voicing series. That is, voiceless stops proved easier to acquire and exhibited more stable, target-like realisations, whereas voiced stops were more prone to CLI. This susceptibility to CLI in voiced series was particularly attested in L2 English, where learners experienced difficulties suppressing prevoicing due to the combined progressive CLI from L1 Polish and the reinforcing presence of prevoicing in newly learnt L3 Norwegian.

In the perception domain, Publication 2 (Cal, 2026a) expanded these results by revealing that perceptual boundary locations between voiced and voiceless stops are likewise subject to CLI, but in a manner determined by the place of articulation. That is, bilabial stops were dominated by L1-induced CLI, coronal stops were kept separate across languages while velar stops demonstrated an L3-driven merged perceptual category characterized by the perceptual boundary placed later in the continuum.

40:6173914724

Additionally, Publication 1 and 2 pointed to substantial individual variation, which confirms that varying perception and production patterns across individuals constitute an inherent property of multilingual perception and production.

The novelty of this PhD dissertation lies in its holistic approach to multilingual speech acquisition. First of all, it adds up to a very limited body of longitudinal research on multilingual acquisition of stops. The unique contribution of this work is in its comprehensive scope, as it tracks the developmental trajectories of both voicing series across all three places of articulation (bilabial, coronal, and velar), all three languages spoken by the participants (L1 Polish, L2 English, L3 Norwegian), and three testing times. By scrutinising the development of stops, the study provides new empirical evidence of how language systems evolve and interact over time. Secondly, the thesis extends the investigation to VOT perception, which is a lesser explored domain in speech acquisition, especially from a multilingual perspective. The study offers a thorough investigation of all three places of articulation across three languages in comparison with native controls for each language. Furthermore, it extends the existing research by comparing stop production in two speech modes and sheds more light on individual variation. All in all, the current dissertation addresses research gaps in the extant literature and contributes to a better understanding of the dynamics of cross-linguistic interactions in VOT acquisition in trilingual speakers.

41:6992664297

Abstract

With the majority of the world's population speaking two or more languages (Grosjean, 2024), multilingualism has become a norm rather than exception. Thus, it seems vital to explore how, in expanding linguistic repertoire, co-existing language systems interact. Voice onset time (VOT) has served as one of the primary acoustic metrics to examine cross-linguistic interactions during stop consonant acquisition (e.g., Amengual, 2021; Llama & Lopez-Morelos, 2016; Sypiańska, 2017; Wrembel, 2015). However, current understanding of how cross-linguistic influence (CLI) manifests across a trilingual repertoire remains limited by several empirical gaps. First of all, prior L3 acquisition studies have focused overwhelmingly on voiceless stops, despite growing evidence of an asymmetry between voicing series, where voiced stops show markedly higher susceptibility to phonetic variation and non-target-like CLI (e.g., Gabriel et al., 2018; Geiss et al., 2022; Zhang et al., 2023). Additionally, longitudinal data capturing how phonetic systems restructure over time are scarce and the area of stop perception across multilingual categories remains underexplored (J. Liu et al., 2019; Z. Liu, 2016). Moreover, the majority of investigations continues to rely on more controlled read speech at the expense of spontaneous, ecologically valid production.

To address these gaps, this PhD dissertation adopted a holistic approach to investigate the acquisition of stop consonants across the speakers' entire linguistic repertoire (L1 Polish, L2 English, and L3 Norwegian). The thesis aimed to examine both production and perception across all three places of articulation (bilabial, coronal, and velar) in both voicing series through three interconnected publications.

Publication 1 (Cal & Wrembel, 2025) tracked the longitudinal development of VOT production during the first year of L3 learning, determining whether novice learners maintain separate language-specific categories or experience progressive and regressive CLI. The results showed that VOT production in multilinguals follows distinct patterns across the three languages. However, its longitudinal development is dynamic and influenced by cross-linguistic interactions, affecting the two voicing series of stops differently.

Publication 2 (Cal, 2026a) investigated multilingual perception, assessing how CLI affects perceptual category boundary locations between voiced and voiceless stops. The findings demonstrated that multilingual speakers also experience CLI in perception,

42:4156317243

but that its directionality and magnitude depend on place of articulation. Bilabial stops underwent L1-induced CLI, coronal stops maintained separate perceptual systems, and velar stops exhibited an L3-driven merged category.

Publication 3 (Cal, 2026b) compared VOT realisations across read speech and spontaneous speech, while also evaluating the role of extralinguistic factors. The study revealed non-facilitative CLI in L2 English. What is more it pointed to the asymmetrical behaviour of the two voicing series of stops, showing that the voiced series is more vulnerable to CLI than the voiceless one. It also attested more variability in spontaneous speech. The research integrated key extralinguistic factors, such as proficiency, frequency of use, and exposure to map the dynamics of CLI in multilingual phonological systems.

All in all, the findings of the present PhD project provide new insights into research on multilingual speech acquisition by offering a holistic and comprehensive analysis of stop acquisition in a single linguistic repertoire.

43:4180164766

Streszczenie

W obliczu faktu, że większość populacji świata posługuje się co najmniej dwoma językami (Grosjean, 2024), wielojęzyczność stała się raczej normą niż wyjątkiem. Z tego względu kluczowe wydaje się zbadanie, w jaki sposób współistniejące systemy językowe wchodzą ze sobą w interakcje w ramach rozwijającego się repertuaru językowego. Parametr voice onset time (VOT) stanowi jeden z podstawowych parametrów akustycznych wykorzystywanych do badania wpływów międzyjęzykowych podczas akwizycji spółgłosek zwartych (np. Amengual, 2021; Llama & Lopez-Morelos, 2016; Sypiańska, 2017; Wrembel, 2015). Jednakże obecny stan wiedzy na temat przejawów wpływów międzyjęzykowych (ang. cross-linguistic influence, CLI) w repertuarze trójjęzycznym wciąż ograniczony jest kilkoma lukami empirycznymi. Przede wszystkim, wcześniejsze badania nad akwizycją trzeciego języka (L3) koncentrowały się przeważnie na spółgłoskach bezdźwięcznych, mimo rosnącej liczby dowodów na asymetrię między obiema seriami dźwięczności, w której to spółgłoski dźwięczne wykazują znacznie większą podatność na wariancję fonetyczną oraz nienatywny CLI (np. Gabriel i in., 2018; Geiss i in., 2022; Zhang i in., 2023). Ponadto, brakuje badań wzdłużnych, które uchwyciłyby restrukturyzację systemów fonetycznych w czasie, a obszar percepcji spółgłosek zwartych w kontekście wielojęzycznym pozostaje słabo zbadany (J. Liu i in., 2019; Z. Liu, 2016). Co więcej, większość analiz wciąż opiera się na wysoce kontrolowanej mowie czytanej kosztem spontanicznej produkcji, która jest bardziej trafna ekologicznie.

Aby wypełnić te luki, w niniejszej rozprawie doktorskiej zastosowano podejście holistyczne do badania akwizycji spółgłosek zwartych w całym repertuarze językowym osób wielojęzycznych (L1 polski, L2 angielski i L3 norweski). Celem pracy było zbadanie zarówno produkcji, jak i percepcji mowy we wszystkich trzech miejscach artykulacji (wargowym, przedniojęzykowym i tylnojęzykowym) w obu seriach dźwięczności w ramach trzech powiązanych ze sobą publikacji.

Publikacja 1 (Cal & Wrembel, 2025) śledziła rozwój produkcji VOT w ujęciu podłużnym podczas pierwszego roku nauki L3, ustalając, czy początkujący uczący się utrzymują odrębne, specyficzne dla danego języka kategorie, czy też doświadczają progresywnego i regresywnego CLI. Wyniki wykazały, że produkcja VOT u osób wielojęzycznych przebiega różnie w każdym z trzech języków. Niemniej jednak, jej rozwój w

44:1424409692

ujęciu czasowym ma charakter dynamiczny i podlega oddziaływaniom międzyjęzykowym, wpływając odmiennie na obie serie dźwięczności spółgłosek zwartych.

Publikacja 2 (Cal, 2026a) badała percepcję wielojęzyczną, oceniając, jak CLI wpływa na lokalizację granic kategorii percepcyjnych między dźwięcznymi i bezdźwięcznymi spółgłoskami zwartymi. Wyniki dowiodły, że osoby wielojęzyczne doświadczają CLI również w percepcji, lecz jego kierunkowość oraz skala zależą od miejsca artykulacji. Zwarte spółgłoski wargowe ulegały wpływowi inicjowanemu przez L1, w zwartych spółgłoskach przedniojęzykowych zachowano odrębne systemy percepcyjne, natomiast zwarte spółgłoski tylnojęzykowe wykazały wspólną kategorię percepcyjną uwarunkowaną przez L3.

Publikacja 3 (Cal, 2026b) porównywała realizacje VOT w mowie czytanej i spontanicznej, oceniając jednocześnie rolę czynników pozajęzykowych. Badanie ujawniło obecność niefacylitatywnego CLI w L2 angielskim. Ponadto wskazało na asymetryczne zachowanie dwóch serii dźwięczności, pokazując, że seria dźwięczna jest bardziej podatna na CLI niż bezdźwięczna. Odnotowano również większą wariancję w mowie spontanicznej. W pracy zintegrowano kluczowe czynniki pozajęzykowe, takie jak poziom biegłości, częstotliwość użycia oraz ekspozycja na język, aby zobrazować dynamikę CLI w wielojęzycznych systemach fonologicznych.

Podsumowując, wnioski z niniejszego projektu doktorskiego wnoszą nowatorski wkład w badania nad wielojęzyczną akwizycją mowy, oferując holistyczną i wyczerpującą analizę akwizycji spółgłosek zwartych w ramach jednego repertuaru językowego.

45:9528783418

References

Amengual, M. (2021). The acoustic realization of language-specific phonological categories despite dynamic cross-linguistic influence in bilingual and trilingual speech. *Journal of the Acoustical Society of America*, *149*(2), 1271–1284. https://doi.org/10.1121/10.0003559

Bandeira, M. H. T., & Zimmer, M. C. (2012). The dynamics of interlanguage transfer of VOT patterns in multilingual children. *Linguagem and Ensino*, *15*(2), 341–364.

Bardel, C., & Falk, Y. (2007). The role of the second language in third language acquisition: The case of Germanic syntax. *Second Language Research*, *23*(4), 459–484. https://doi.org/10.1177/0267658307080557

Bardel, C., & Falk, Y. (2012). The L2 status factor and the declarative/procedural distinction. In J. Cabrelli Amaro, S. Flynn, & J. Rothman (Eds.), *Third language acquisition in adulthood* (pp. 61–78). John Benjamins.

Bless, M. (2015). *Interference in early acquisition Dutch-English bilinguals: A phonetic examination of voice onset time in Dutch and English bilabial plosives* [Unpublished master's thesis]. University of Amsterdam.

Boersma, P., & Weenink, D. (2021). Praat: Doing phonetics by computer (Version 6.1.50) [Computer software]. http://praat.org

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

Cabrelli Amaro, J. (2013). *The Phonological Permeability Hypothesis: Measuring regressive L3 influence to test L2 phonological representations* [Unpublished doctoral dissertation]. University of Florida.

Cabrelli Amaro, J. (2017). Testing the phonological permeability hypothesis: L3 phonological effects on L1 versus L2 systems. *International Journal of Bilingualism*, *21*(6), 698–717. https://doi.org/10.1177/1367006916637287

Cabrelli Amaro, J., & Rothman, J. (2010). On L3 acquisition and phonological permeability: A new test case for debates on the mental representation of non-native phonological systems. *International Review of Applied Linguistics in Language Teaching*, *48*(2–3), 275–296.

46:4827389313

Cal, Z. (2026a). Crossing the boundaries: Perception of voice onset time in word-initial stops by multilingual learners. *Language and Speech*. Advance online publication. https://doi.org/10.1177/00238309251414959

Cal, Z. (2026b). Is L2 the least stable system? The dynamics of VOT trilingual production in Polish-English-Norwegian read and spontaneous speech. *Languages*, *11*(9), Article 182. https://doi.org/10.3390/languages11090182

Cal, Z., & Wrembel, M. (2023). Perception of voicing distinction in syllable-initial stops by multilingual speakers. In R. Skarnitzl & J. Volin (Eds.), *Proceedings of the 20th International Congress of Phonetic Sciences–ICPhS 2023* (pp. 451–455). International Phonetic Association.

Cal, Z., & Wrembel, M. (2025). Developmental acquisition of stops by multilingual speakers. *International Journal of Multilingualism*, *22*(1), 13–35. https://doi.org/10.1080/14790718.2024.2363777

Chang, C. B. (2019). Phonetic drift. In M. S. Schmid & B. Köpke (Eds.), *The Oxford Handbook of Language Attrition* (pp. 191-203). Oxford, UK: Oxford University Press.

Chung, Y., Rabe-Hesketh, S., Dorie, V., Gelman, A., & Liu, J. (2013). A nondegenerate penalized likelihood estimator for variance parameters in multilevel models. *Psychometrika*, *78*(4), 685–709. https://doi.org/10.1007/s11336-013-9328-2

Czarnecki, P. (2016). *The phonology of quantity in Icelandic and Norwegian*. Wydział Neofilologii UAM w Poznaniu.

De Angelis, G. (2007). *Third or Additional Language Acquisition*. Bristol, Blue Ridge Summit: Multilingual Matters. https://doi.org/10.21832/9781847690050

Docherty, G. J. (1992). *The timing of voicing in British English obstruents*. De Gruyter.

Donegan, P. J., & Stampe, D. (2009). Hypotheses of natural phonology. *Poznan Studies in Contemporary Linguistics*, *45*(1), 3–31.

Dukiewicz, L., & Sawicka, I. (1995). *Gramatyka współczesnego języka polskiego: Fonetyka i fonologia* [Grammar of the contemporary Polish language: Phonetics and phonology] (H. Wróbel, Ed.). Wydawnictwo Instytutu Języka Polskiego PAN.

Dziubalska-Kołaczyk, K., & Wrembel, M. (2022). Natural growth theory of acquisition (NGTA): Evidence from (mor)phonotactics. In V. Sardegna & A. Jarosz (Eds.),

47:3593834704

Theoretical and practical developments in English speech assessment, research, and training (pp. 281–298). Springer Nature.

Dziubalska-Kołaczyk, K., & Wrembel, M. (2024). A revised natural growth theory of acquisition: Evidence from L3 phonology. In E. Babatsouli (Ed.), *Multilingual acquisition and learning: An ecosystemic view to diversity*. John Benjamins.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–276). York Press.

Flege, J. E., & Bohn, O.-S. (2021). The Revised Speech Learning Model (SLM-r). In R. Wayland (Ed.), *Second Language Speech Learning: Theoretical and Empirical Progress* (pp. 3–83). chapter, Cambridge: Cambridge University Press.

Gabriel, C., Krause, M., & Dittmers, T. (2018). VOT production in multilingual learners of French as a foreign language: Cross-linguistic influence from the heritage languages Russian and Turkish. *Revue Française de Linguistique Appliquée*, *23*(1), 59–72. https://doi.org/10.3917/rfla.231.0059

Gass, S., & Selinker, L. (2001). *Second language acquisition: An introductory course* (2nd ed.). Lawrence Erlbaum Associates.

Geiss, M., Gumbsheimer, S., Lloyd-Smith, A., Schmid, S., & Kupisch, T. (2022). Voice onset time in multilingual speakers: Italian heritage speakers in Germany with L3 English. *Studies in Second Language Acquisition*, *44*(2), 435–459. https://doi.org/10.1017/S0272263121000280

Grosjean F. (1989). Neurolinguists, beware! The bilingual is not two monolinguals in one person. *Brain and language*, *36*(1), 3–15. https://doi.org/10.1016/0093-934x(89)90048-5

Grosjean, F. (2024). The Statistics of Bilingualism. In F. Grosjean (Ed.), *On Bilinguals and Bilingualism* (pp. 138–147). Cambridge University Press.

Gut, U. (2010). Cross-linguistic influence in L3 phonological acquisition. *International Journal of Multilingualism*, *7*(1), 19–38. https://doi.org/10.1080/14790710902972248

Halvorsen, B. (1998). *Timing relations in Norwegian stops* [Unpublished doctoral dissertation]. University of Bergen.

Hansen Edwards, J. & Zampini, M. (2008). *Phonology and Second Language Acquisition*. John Benjamins Publishing Company. https://doi.org/10.1075/sibil.36

48:3883878633

Jassem, W. (2003). Polish. *Journal of the International Phonetic Association*, *33*(1), 103–107. https://doi.org/10.1017/S0025100303001191

Jarvis, S., & Pavlenko, A. (2007). *Crosslinguistic influence in language and cognition*. Routledge.

Jerotijević Tišma, D. M. (2019). The effect of speech style on VOT values in Serbian-English interphonology. *Srpski Jezik*, *24*(1), 337–355.

Keating, P. A., Mikoś, M. J., & Ganong, W. F., III. (1981). A cross-language study of range of VOT in the perception of stop consonant voicing. *The Journal of the Acoustical Society of America*, *70*(5), 1260–1271. https://doi.org/10.1121/1.387139

Kellerman, E. (1984). The empirical evidence for the influence of the L1 in interlanguage. In A. Davies, C. Criper, & A. P. R. Howatt (Eds.), *Interlanguage* (pp. 98–122). Edinburgh University Press.

Kisler, T., Reichel, U. D., & Schiel, F. (2017). Multilingual processing of speech via web services. *Computer Speech & Language*, *45*, 326–347. https://doi.org/10.1016/j.csl.2017.01.005

Kopečková, R., Gut, U., Wrembel, M., & Balas, A. (2023). Phonological cross-linguistic influence at the initial stages of L3 acquisition. *Second Language Research*, *39*(4), 1107-1131. https://doi.org/10.1177/02676583221123994

Kretzschmar, W. A., Jr. (2015). *Language and complex systems*. Cambridge University Press.

Kristoffersen, G. (2000). *The phonology of Norwegian*. Oxford University Press.

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops. *Word*, *20*(3), 384–422. https://doi.org/10.1080/00437956.1964.11659830

Lin, Y. (2025). Sound change and cross-linguistic influence among multilingual speakers: evidence from /n/-/l/ merger in Cantonese. *International Journal of Multilingualism*, 1–19. https://doi.org/10.1080/14790718.2025.2604152

Liu, J., & Lin, J. (2021). A cross-linguistic study of L3 phonological acquisition of stop contrasts. *SAGE Open*, *11*(1). https://doi.org/10.1177/2158244020985510

Liu, J., Zeng, T., & Lu, X. (2019). Challenges in multi-language pronunciation teaching: A cross-linguistic study of Chinese students' perception of voiced and voiceless stops. In Y. Zheng & X. Gao (Eds.), *Círculo de Lingüística Aplicada a la Comunicación* (Vol. 79, pp. 99–118). https://doi.org/10.5209/clac.65652

49:3640710909

Liu, Z. (2016). *Exploring cross-linguistic influence: Perception and production of L1, L2 and L3 bilabial stops by Mandarin Chinese speakers* [Master's thesis, Universitat Autònoma de Barcelona]. Dipòsit Digital de Documents de la UAB.

Llama, R., & Cardoso, W. (2018). Revisiting (non-)native influence in VOT production: Insights from advanced L3 Spanish. *Languages*, *3*(3), 30.

Llama, R., Cardoso, W., & Collins, L. (2010). The influence of language distance and language status on the acquisition of L3 phonology. *International Journal of Multilingualism*, *7*(1), 39–57. https://doi.org/10.1080/14790710902972255

Llama, R., & López-Morelos, L. P. (2016). VOT production by Spanish heritage speakers in a trilingual context. *International Journal of Multilingualism*, *13*(4), 444–458. https://doi.org/10.1080/14790718.2016.1217602

Marinho Ribeiro, D. (2021). *Third language acquisition: A study of unstressed vowel reduction* [Doctoral dissertation, Purdue University]. Purdue University Graduate School. https://doi.org/10.25394/PGS.14515668

Nelson, C. (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. https://doi.org/10.3390/languages7020078

Patience, M., Qian, W., & Steele, J. (2026). Assessing L1 Mandarin and L2 English influence on the L3 production of French obstruent coda voicing. *Second Language Research*, *42*(3), 423-448. https://doi.org/10.1177/02676583251336582

Ringen, C., & van Dommelen, W. A. (2013). Quantity and laryngeal contrasts in Norwegian. *Journal of Phonetics*, *41*(6), 479–490. https://doi.org/10.1016/j.wocn.2013.09.001

Roach, P. (2009). *English phonetics and phonology: A practical course* (4th ed.). Cambridge University Press.

Rothman, J. (2011). L3 syntactic transfer selectivity and typological determinacy: The Typological Primacy Model. *Second Language Research*, *27*(1), 107–127.

Rothman, J. (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.

Schwartz, G. (2020). Asymmetrical cross-language phonetic interaction: Phonological implications. *Linguistic Approaches to Bilingualism*, *12*(2), 103–132. https://doi.org/10.1075/lab.19092.sch

50:6023772515

Schwartz, G., Wojtkowiak, E., & Dzierla, J. (2020). Laryngeal phonology and asymmetrical cross-language phonetic influence. In M. Wrembel, A. Kiełkiewicz-Janowiak, & P. Gąsiorowski (Eds.), *Approaches to the study of sound structure and speech: Interdisciplinary work in honour of Katarzyna Dziubalska-Kołaczyk* (pp. 316–326). Routledge.

Sharwood-Smith, M. (1983). On first language loss in the second language acquirer: problem of transfer. In S. Gass, & L Selinker (Eds.) *Language Transfer in Language Learning* (pp. 222-231). Rowley, MA: Newbury House.

Sharwood-Smith, M. (1994). *Second language acquisition: Theoretical foundations.* London: Longman.

Slabakova, R. (2017). The scalpel model of third language acquisition. *International Journal of Bilingualism*, *21*(6), 651–665.

Stan, I. (2024). A cross-linguistic study of L3 speech perception of voiced and voiceless stops. *Journal of Experimental Phonetics*, *33*, 137–152. https://doi.org/10.1344/efe-2024-33-137-152

Stoehr, A., Jevtović, M., de Bruin, A., & Martin, C. D. (2024). Phonetic and lexical crosslinguistic influence in early Spanish–Basque–English trilinguals. *Language Learning, 74*(2), 332–364. https://doi.org/10.1111/lang.12598

Sypiańska, J. (2016). Multilingual acquisition of vowels in L1 Polish, L2 Danish and L3 English. *International Journal of Multilingualism*, *13*(4), 476–495. https://doi.org/10.1080/14790718.2016.1217606

Sypiańska, J. (2017). *Cross-linguistic influence in bilinguals and multilinguals*. Wydawnictwo Naukowe UAM.

Sypiańska, J. (2021). Production of voice onset time (VOT) by senior Polish learners of English. *Open Linguistics*, *7*(1), 316–330. https://doi.org/10.1515/opli-2021-0016

Sypiańska, J. (2025). The pronunciation of the L3 Polish labiovelar approximant by Ukrainian-Russian bilinguals: The interplay between cross-linguistic influence from the background languages and L3 phonotactics. *Linguistik Online*, *134*(2), 51-66. https://doi.org/10.13092/lo.134.12178

Sypiańska, J., & Wrembel, M. (2026). Multilingual speech rhythm: A comprehensive investigation into rhythm metrics across three languages and two speech modes.

51:4052620342

Journal of Phonetics, *117*, Article 101498. https://doi.org/10.1016/j.wocn.2026.101498

Tremblay, M. C. (2007, August 6–10). *L2 influence on L3 pronunciation: Native-like VOT in the L3 Japanese of English-French bilinguals* [Paper presentation]. The Satellite Workshop of ICPHS XVI, Freiburg, Germany.

Wei, Y. (2025). Stop production of L1 Chinese - L2 English - L3 Spanish learners: The source and direction of crosslinguistic influence. *Journal of Experimental Phonetics*, *34*, 21–40. https://doi.org/10.1344/efe-2025-34-21-40

Westbury, J. R. (1979). *Aspects of the temporal control of voicing in consonant clusters in English* [Doctoral dissertation, University of Texas at Austin].

Westergaard, M., Mitrofanova, N., Mykhaylyk, R., & Rodina, Y. (2017). Crosslinguistic influence in the acquisition of a third language: The Linguistic Proximity Model. *International Journal of Bilingualism*, *21*(6), 666–682.

Winn, M. B. (2020). Manipulation of voice onset time in speech stimuli: A tutorial and flexible Praat script. *The Journal of the Acoustical Society of America*, *147*(2), 852. https://doi.org/10.1121/10.0000692

Wojtkowiak, E. (2022). *L1 phonetic drift in the speech of Polish learners of English: Phonological implications* [Unpublished doctoral dissertation]. Adam Mickiewicz University.

Wrembel, M. (2010). L2-accented speech in L3 production. *International Journal of Multilingualism*, *7*(1), 75–90.

Wrembel, M. (2011). Cross-linguistic influence in third language acquisition of voice onset time. In W.-S. Lee & E. Zee (Eds.), *Proceedings of the 17th International Congress of Phonetic Sciences* (pp. 2157–2160). City University of Hong Kong.

Wrembel, M. (2014). VOT patterns in the acquisition of third language phonology. *Concordia Working Papers in Applied Linguistics*, *5*, 751–771.

Wrembel, M. (2015). *In search of a new perspective: Cross-linguistic influence in the acquisition of third language phonology*. Wydawnictwo Naukowe UAM.

Wunder, E. M. (2010). Phonological cross-linguistic influence in third or additional language acquisition. In M. Wrembel, M. Kul, & K. Dziubalska-Kołaczyk (Eds.), *Achievements and perspectives in the acquisition of second language speech: Proceedings of the 6th international conference on native and non-native phonology* (New Sounds 2010, pp. 566–571). Peter Lang.

52:3873548566

Xu, B. (2025). Verifying the Phonological Permeability Hypothesis – L3 French's regressive impact on L2 English aspirated consonants. *Communications in Humanities Research*, *55*, 243–256. https://doi.org/10.54254/2753-7064/2024.22507

Zając, M. (2015). *Phonetic convergence in the speech of Polish learners of English* [Unpublished doctoral dissertation]. University of Łódź.

Zeng, M. (2024, May 17–19). *Mandarin trilinguals' perception of L3 Japanese stops* [Poster presentation]. L3 Workshop: Multilingual Language Acquisition, Processing and Use, Tokyo, Japan.

Zeng, M. (2026). Cross-linguistic phonetic interference: The production and perception of L3 Japanese word-initial stops by Mandarin trilinguals. *Journal of Japanese Linguistics*, *42*(1), 27-56. https://doi.org/10.1515/jjl-2026-2004

Zhang, L. (2022). ¿Pollo o Bollo? Acquisition of voicing contrast by Chinese learners of L3 Spanish. *Hispania*, *105*(4), 595–613. https://doi.org/10.1353/hpn.2022.0118

Zhang, L., Morales-Front, A., & Sanz, C. (2023). Dynamics in cross-linguistic influence: Evidence from VOT development of Mandarin-English bilingual learners of L3 Spanish. In M. Brown, S. Flynn, & É. Fernández-Berkes (Eds.), *L3 development after the initial state* (pp. 121–149). John Benjamins. https://doi.org/10.1075/sibil.65.06zha

Zhu, Y., & Wang, J. (2025). What drives L3 English nasal acquisition? evidence of cross-linguistic influence from L1 and L2 in Sinitic speakers. *International Journal of Multilingualism*, 1–26. https://doi.org/10.1080/14790718.2025.2604625

53:3162679954

Appendix A: Publication 1 (Cal & Wrembel, 2025): Author contribution statement (Polish)

Poznań, 04.09.2026

OŚWIADCZENIE O WKŁADZIE AUTORÓW I AUTOREK

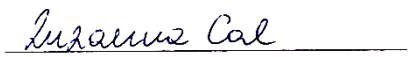
W odniesieniu do publikacji:

Cal, Zuzanna and Magdalena Wrembel. 2025. "Developmental acquisition of stops by multilingual speakers", *International Journal of Multilingualism* *22*(1): 13–35. https://doi.org/10.1080/14790718.2024.2363777.

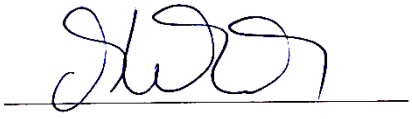
Autorki powyższej publikacji niniejszym deklarują swój następujący wkład w powstanie artykułu:

- **Zuzanna Cal** (80%): konceptualizacja, zebranie danych, analiza i wizualizacja danych, przygotowanie oryginalnego tekstu manuskryptu, recenzja i redagowanie tekstu manuskryptu;
- **Magdalena Wrembel** (20%): konceptualizacja, opieka merytoryczna nad przeprowadzaniem badania, recenzja i redagowanie tekstu manuskryptu.

Autorki

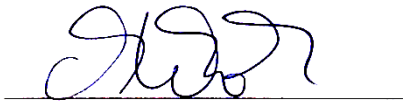


mgr Zuzanna Cal



prof. UAM dr hab. Magdalena Wrembel

Promotorka Pracy Doktorskiej



prof. UAM dr hab. Magdalena Wrembel

54:4314571859

Appendix B: Publication 1 (Cal & Wrembel, 2025): Author contribution statement (English)

Poznań, 04 September 2026

AUTHOR CONTRIBUTION STATEMENT

In reference to the publication:

Cal, Zuzanna and Magdalena Wrembel. 2025. "Developmental acquisition of stops by multilingual speakers", *International Journal of Multilingualism* 22(1): 13–35. https://doi.org/10.1080/14790718.2024.2363777.

The co-authors of the above publication hereby declare that they have contributed to the research article in the following way:

- **Zuzanna Cal** (80%): conceptualisation, data collection, data analysis and visualisation, original draft preparation, review and editing of the manuscript;
- **Magdalena Wrembel** (20%): conceptualisation, supervision over the conduct of the study, review and editing of the manuscript.

Authors

Zuzanna Cal

mgr Zuzanna Cal

prof. UAM dr hab. Magdalena Wrembel

PhD Supervisor

prof. UAM dr hab. Magdalena Wrembel

55:8664758901

Appendix C: Publication 2 (Cal, 2026a): Author contribution statement (Polish)

Poznań, 04.09.2026

OŚWIADCZENIE O WKŁADZIE AUTORÓW I AUTOREK

W odniesieniu do publikacji:

Cal, Zuzanna. 2026. "Crossing the boundaries. Perception of voice onset time in word-initial stops by multilingual learners", *Language and Speech* *0*(0). 1–24. https://doi.org/10.1177/00238309251414959

Autorka powyższej publikacji niniejszym deklaruje swój następujący wkład w powstanie artykułu:

- **Zuzanna Cal** (100%): konceptualizacja, zebranie danych, analiza i wizualizacja danych, przygotowanie oryginalnego tekstu manuskryptu, recenzja i redagowanie tekstu manuskryptu.

Autorka

Zuzanna Cal

mgr Zuzanna Cal

Promotorka Pracy Doktorskiej

prof. UAM dr hab. Magdalena Wrembel

56:7134643240

Appendix D: Publication 2 (Cal, 2026a): Author contribution statement (English)

Poznań, 04 September 2026

AUTHOR CONTRIBUTION STATEMENT

In reference to the publication

Cal, Zuzanna. 2026. "Crossing the boundaries: Perception of voice onset time in word-initial stops by multilingual learners", *Language and Speech* 0(0) 1–24. https://doi.org/10.1177/00238309251414959

The author of the above publication hereby declares that they have contributed to the research article in the following way

- **Zuzanna Cal** (100%): conceptualisation, data collection, data analysis and visualisation, original draft preparation, review and editing of the manuscript,

Authors

Zuzanna Cal

mgr Zuzanna Cal

PhD Supervisor

prof. UAM dr hab. Magdalena Wrembel

57:3075212399

Appendix E: Publication 3 (Cal, 2026b): Author contribution statement (Polish)

Poznań, 04 09 2026

OŚWIADCZENIE O WKŁADZIE AUTORÓW I AUTOREK

W odniesieniu do publikacji:

Cal, Zuzanna. 2026. "Is the L2 the least stable system? The dynamics of VOT trilingual production in Polish-English-Norwegian read and spontaneous speech", *Languages 11*(9): 1–28. https://doi.org/10.3390/languages11090182

Autorka powyższej publikacji niniejszym deklaruje swój następujący wkład w powstanie artykułu:

- **Zuzanna Cal** (100%): konceptualizacja, zebranie danych, analiza i wizualizacja danych, przygotowanie oryginalnego tekstu manuskryptu, recenzja i redagowanie tekstu manuskryptu

Autorka

Zuzanna Cal

mgr Zuzanna Cal

Promotorka Pracy Doktorskiej

prof. UAM dr hab. Magdalena Wrembel

58:7992963285

Appendix F: Publication 3 (Cal, 2026b): Author contribution statement (English)

Poznań, 04 September 2026

AUTHOR CONTRIBUTION STATEMENT

In reference to the publication

Cal, Zuzanna. 2026. "Is the L2 the least stable system? The dynamics of VOT trilingual production in Polish-English-Norwegian read and spontaneous speech", *Languages 11*(9): 1–28. https://doi.org/10.3390/languages11090182

The author of the above publication hereby declares that they have contributed to the research article in the following way:

- **Zuzanna Cal** (100%): conceptualisation, data collection, data analysis and visualisation, original draft preparation, review and editing of the manuscript,

Authors

Zuzanna Cal

mgr Zuzanna Cal

PhD Supervisor

prof. UAM dr hab. Magdalena Wrembel

59:6236418769

Appendix G: Publication 1 (Cal & Wrembel, 2025)

Cal, Zuzanna and Magdalena Wrembel. 2025. “Developmental acquisition of stops by multilingual speakers”, *International Journal of Multilingualism 22*(1): 13–35. https://doi.org/10.1080/14790718.2024.2363777.

60:1062872450

INTERNATIONAL JOURNAL OF MULTILINGUALISM
2025, VOL. 22, NO. 1, 13–35
https://doi.org/10.1080/14790718.2024.2363777

Routledge
Taylor & Francis Group

OPEN ACCESS Check for updates

Developmental acquisition of stops by multilingual speakers

Zuzanna Cal and Magdalena Wrembel

Faculty of English, Adam Mickiewicz University, Poznań, Poland

ABSTRACT

The study aimed to analyse the production of plosive sounds by trilingual speakers of L1 Polish – L2 English – L3 Norwegian. The main objective was to trace any changes in the VOT production of fortis and lenis stops in all three languages (Polish, English and Norwegian) produced by novice L3 learners with a focus on possible cross-linguistic interactions. The investigation was conducted across three testing times throughout the first year of L3 learning. The results show that participants maintained separate VOT patterns in all three languages. What is more, VOT production patterns changed over time, but only in the voiced series of stops. The study also provides evidence for asymmetry between voiced and voiceless stops, as the two series turned out to be affected by CLI in a different manner.

ARTICLE HISTORY
Received 29 June 2023
Accepted 30 May 2024

KEYWORDS
Cross-linguistic influence; speech acquisition; VOT; multilingualism; third language acquisition of phonology

Introduction

Research on cross-linguistic influence (CLI) in multilinguals has been in its prime for a few decades now. However, compared to other language domains, investigations into CLI in phonological acquisition have not been so comprehensive. What is more, the majority of studies in this domain seem to be centred on CLI from previously learnt languages on the L3, whereas any potential impact on the L1 or L2 has been less studied. From a phonological perspective, the acquisition of stops by foreign language learners has been a rather popular research agenda for some time now. Nevertheless, the L3 literature to date has focused mostly on the voiceless series of stops (e.g. Amengual, 2021; Llama & López-Morelos, 2016; Sypiańska, 2013; Wrembel, 2015). More recent investigations also include the voiced counterparts (e.g. Gabriel et al., 2018; Geiss et al., 2022), but the body of literature on this matter is still rather scarce. What is more, the area that remains largely unexplored concerns the developmental aspects of the acquisition of stops by multilingual learners over time, especially when it comes to a holistic approach to investigating interactions between all the languages in a speaker's repertoire (but see Nelson, 2022 for a notable exception). Hence, the current study aims to fill these gaps by analysing the production of plosives by L1 Polish – L2 English – L3

CONTACT Zuzanna Cal zuzanna.cal@amu.edu.pl Faculty of English, Adam Mickiewicz University, ul. Grunwaldzka 6, 60-780 Poznan, Poland

© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

Norwegian speakers at the early stages of learning their L3 throughout three testing times. The main objective is to explore whether early-stage multilingual learners keep their categories apart while they advance in their L3 proficiency and to trace the development of voice onset time (VOT) acquisition in all three languages in both series of stops (/p,t,k/ and /b,d,g/).

This paper is structured in the following manner. It begins with a discussion of the state of the art in VOT production research in the context of L2 and L3 acquisition. Further, the VOT parameter is discussed from the point of view of the three languages under investigation, namely Polish, English and Norwegian. The following sections feature the research design and methodology description. Finally, the results are presented and discussed with reference to the extant literature on multilingualism.

VOT in bi-/multilinguals[1]

VOT is a feature of plosive consonants that is a measure of the time between the stop release and the voicing onset (Lisker & Abramson, 1964). This parameter is said to be the most salient cue that allows to distinguish between voiced (/b, d, g/) and voiceless (/p, t, k/) plosives and, what is more, it is considered to be one of the most commonly researched phonetic features in the L2 speech literature (Hansen Edwards & Zampini, 2008). Lisker and Abramson (1964) distinguish three main types of VOT, namely (1) voicing lead or prevoicing in which voicing starts before the release of a plosive; (2) short-lag VOT in which voicing begins with or just after the release of a stop (VOT duration between 0 and 35 ms); (3) long-lag VOT in which voicing begins later after a stop release (VOT durations longer than 35 ms).

VOT has been widely researched in the last decades, being a parameter relatively comparable across languages and easy to measure. Not only has it been measured by phoneticians to make comprehensive descriptions of specific languages, but its analysis has also been employed in research on foreign language acquisition to investigate possible cross-linguistic interactions during the process of L2 and L3 acquisition.

From the SLA perspective, a number of studies resorted to VOT for an analysis of cross-linguistic influence. In one of his seminal studies, Flege (1987) examined VOT durations of /t/ produced by L1 English speakers who were proficient in L2 French. The obtained results indicated that VOT values for L2 French /t/ were in between the target values for English and French, suggesting an occurrence of CLI between the two languages. Similar trends were also attested for speakers of different linguistic repertoires (cf. Flege, 1991; Flege & Eefting, 1987; Flege & Hillenbrand, 1984; Hazan & Boulakia, 1993; Thornburgh & Ryalls, 1998; Williams, 1977). What is more, this trend was also observed for reverse settings. In Flege's (1987) study, L1 French speakers who resided in the L2 English-speaking country produced L1 French /t/ with VOT values in between those of English and French, giving the evidence for the existence of L1 drift, which refers to phonetic changes in the native language (L1) caused by recent experience in another language (L2) (Chang 2019).[2]

More recent studies on VOT acquisition in bilinguals, however, tend to provide more comprehensive analyses, as they focus on both series of stops more often than multilingual studies. Moreover, they point to an interesting asymmetry between voiced and voiceless stops in which the voiced stops[3] seem to be more prone to CLI than the voiceless

ones (e.g. Antoniou et al., 2010; Bless, 2015; Chionidou & Nicolaidis, 2015; Herd et al., 2015; Newlin-Łukowicz, 2014; Schwartz et al., 2020; Sučková, 2018). As far as research on Polish speakers is concerned, Zając (2015) found that they experienced problems suppressing pre-voicing in voiced stops but managed to converge with aspirated voiceless plosives. This asymmetry between voiced and voiceless series of stops was also observed more recently by Schwartz (2020) who examined VOT production in speakers' L1 Polish and L2 English. Experiment 1 focused on L1 phonetic drift and included two groups of L2 English speakers – learners at B2 level of proficiency who underwent an extensive phonetic training in L2 and L1 Polish quasi-monolinguals who had approximately A1-A2 level of proficiency in English. The findings revealed that VOT values for both groups were almost identical for voiceless stops; however, the more proficient group exhibited shorter pre-voicing in the voiced series and fewer prevoiced items in general. Experiment 2 analysed speakers' L2 English performance across two groups – Polish learners of L2 English at B1 level of proficiency and staff members in the English department at C2 level of proficiency in English, both of whom were phonetically trained. Again, the findings revealed an asymmetry in the realisation of the two series of stops. The VOT values for both groups were almost identical for voiceless stops; however, the more proficient group exhibited less and shorter pre-voicing in the voiced series. However, between-group differences were not found to be significant. The author interpreted the findings based on the framework of the Speech Learning Model (SLM; Flege, 1995), which posits three-way similarity classification between pairs of sounds into identical, similar and different with varied degrees of the resulting equivalence classification. It was suggested that the bidirectional CLI observed in the voiced stops may be indicative of these sounds being treated as equivalents in Polish and English, which means that the voiced stops in these languages were perceived as similar by the participants. It is worth noting that this tendency has also been reported for longitudinal data on L1 drift and L2 production. Wojtkowiak (2022) analysed L1 and L2 productions of word-initial stops by three groups of L1 Polish L2 English speakers and a group of Polish functional monolinguals. L1 drift results showed that voiceless stops were rather unaffected by the intensive acquisition of L2 English, while the voiced category turned out to be more vulnerable to the influence of the foreign language. What is more, a comparison with L2 data revealed that participants must have formed a new category for voiceless stops in English that was separate from the Polish ones. On the other hand, voiced plosives were found to be more affected by the L2, which may mean that the sounds in the two languages were perceptually assimilated to a single, shared category (Wojtkowiak, 2022). The author accounts for the finding claiming that lenis stops are more vulnerable to CLI than the fortis ones due to their different phonological status (see Schwartz, 2020). This claim is embedded in the Onset Prominence framework, according to which aspiration in voiceless stops is a phonological feature ([fortis]), while pre-voicing is considered to be a phonetic detail.

Compared to the extensive body of literature on VOT acquisitionin L2, to date,this parameter has received considerably less attention from a multilingual perspective. Yet, with multilingualism being the norm in today's world, it seems vital to examine more linguistically diverse groups of participants. There have been attempts to apply the explanatory frameworks put forth in the context of L3 morphosyntax to account for L3 phonological acquisition (see, e.g. Kopečková et al., 2023 for an overview); however, in the present contribution we refer only to the L2 status model by Bardel and Falk (2007), as the most

applicable one for the interpretation of the current data. On the whole, the existing L3 studies in this area seem to be rather limited, focusing mostly on the main source of CLI in the third language and pointing to L2 influence, combined L1-L2 effect or providing mixed results (e.g. Llama & López-Morelos, 2016; Tremblay, 2007). For example, Llama et al. (2010) analysed L3 Spanish spoken by two mirror groups, i.e. L1 English L2 French and L1 French L2 English speakers to determine whether it is L2 status or linguistic typology that playsx a role in shaping VOT patterns in voiceless stops. The results demonstrated that the L2 status was the main predictor of L3 VOT values; however, there was also some evidencepointing to the interaction of L1 and L2 influence. Mixed results were also obtained by Wunder (2010), who analysed voiceless stops produced by L1 German L2 English L3 Spanish speakers performing a reading task. The findings suggested either an L1 influence or a combined CLI stemming from both L1 and L2. In fact, in most cases, the data showed hybrid VOT values in L3 Spanish that could not be classified as either L1 or L2 influence. Wrembel (2011) studied voiceless stops production by L1 Polish, L2 English, L3 French speakers, demonstrating that L2 English VOT was produced on a target, but L3 values were in-between those of Polish and English. These data suggest the co-existence of L2 status effect and L1 influence thus attesting combined CLI. In a further study Wrembel (2015) revealed similar trends showing that L3 might be influenced by both the first and the second language. It was also concluded that this influence is gradual, structure-dependent and can only partially be accounted for by the existing models of L3 acquisition. Gabriel et al. (2016) compared the L3 French VOT patterns of German/Mandarin-Chinese multilingual learners to two control groups of monolingual learners and found no significant differences; neither in the production of /t, d/ nor the learners' perceived accent. The authors interpret these results as the lack of confirmation of any potential advantage of multilinguals when acquiring the phonology of an additional foreign language. However, the source of L1-based transfer was hard to determine since the selected background languages (German and Mandarin-Chinese) align in terms VOT target values. A more recent study by Amengual (2021), however, demonstrated a slightly different trend in terms of possible cross-linguistic interactions. The author examined the production of /k/ by two groups of participants: English-Japanese bilinguals and L1 Spanish, L2 English, L3 Japanese speakers. The findings revealed that both the bilingual and trilingual speakers were able to maintain language specific VOT categories, however, only the latter showed significant differences between VOT values across languages.

VOT in multilinguals has also been investigated from a more holistic point of view, providing evidence also for regressive CLI. For example, Sypiańska (2013) investigated the production of word-initial voiceless stops in L1 Polish L2 Danish L3 English speakers. She found combined CLI stemming from L1 and L2 to L3 VOT values, but also longer VOT values in L1 and L2, which pointed to the possibility of regressive CLI from the L3 to L1 and L2. Another important contribution was recently made by Nelson (2022), who investigated regressive phonological CLI in voiceless stops produced by adolescents and adults speaking L1 German, L2 English, L3 Polish. The study explored the development of all three languages over four testing times. The author observed significant changes to both L1 and L2 over time in the adolescent group and relatively stable L1 and L2 systems in the adults group suggesting that the stability of background languages might depend on age, at least at a group level.

Surprisingly, only a handful of studies to date have investigated the voiced category of stops in L3 in a systematic way. From what has been established, however, it can be assumed that a similar asymmetry, as attested in the L2 data, will also be observed in L3. Liu (2016) investigated VOT in bilabial stops /p/ and /b/ in L1 Mandarin – L2 English participants compared to L1 Mandarin – L2 English – L3 Spanish speakers. L3 learning turned out to influence the productions of the voiced stop in L1, as well as voiced and voiceless stops in L2. What is more, it was revealed that L3 learners maintained three separate systems for the production of the voiceless stop and one system for the production of L1 and L2 voiced stops, but a different one for L3. Gabriel et al. (2018) examined word-initial fortis and lenis stops produced in L3 French by multilingual learners speaking Russian or Turkish as a heritage language alongside German, their dominant language. The results showed that compared with monolingual data, multilinguals produced voiceless stops closer to the target, but it was not the case for the voiced stops. It is explained that the learners might perceive aspiration as more salient than the presence or absence of prevoicing. On the other hand, slightly different results were reported by Geiss et al. (2022) who analysed voiced and voiceless stops produced by Italian-German heritage speakers with L3 English. The findings pointed to separate VOT patterns in both series of stops in all three languages.

Table 1 summarises the literature review regarding the multilingual acquisition of VOT. All in all, current evidence showsan asymmetry in the acquisition of the two series of stops in both L2 and L3 learning contexts. However, the L3 studies have remained quite narrow in focus so far dealing mostly with voiceless plosives and not taking into consideration all three languages and directions of CLI, nor the developmental aspects of L3 acquisition. To the best of our knowledge, no study to date has tackled all of the aforementioned aspects jointly. Therefore, the present contribution aims to bridge this gap and offer new insights by exploring longitudinally the acquisition of VOT in voiceless and voiced stops by

Table 1. Summary of studies investigating multilingual acquisition of VOT.

Investigated stops	Investigated type of CLI	Reference(s)	Relevant results
/*p*, t, k/	Progressive CLI	Tremblay (2007); Llama and López-Morelos (2016); Wrembel (2011); Wrembel (2015)	– Combined L1-L2 effect on L3 or mixed results
		Llama et al. (2010)	– L2 status was the main predictor of L3 VOT values – Combined L1-L2 effect on L3
		Wunder (2010)	– L1 or a combined L1-L2 influence on L3
/k/		Amengual (2021)	– language specific VOT categories maintained
/*p*, t, k/	Progressive and regressive CLI	Sypiańska (2013)	– combined L1-L2 effect on L3 – regressive CLI from L3 to L1 and L2
	Regressive CLI	Nelson (2022) – longitudinal investigation	– significant changes to both L1 and L2 over time in adolescents – relatively stable L1 and L2 systems in adults
/*p*, b/	Progressive and regressive CLI	Liu (2016)	– voiceless series: L3 influence on both L1 and L2, separate language systems maintained – voiced series: L3->L2 influence, one L1-L2 system, L3 stops kept separate
/*p*, k, b, g/	Progressive CLI	Gabriel et al. (2018)	– voiceless series produced closer to the target than the voiced one
	Progressive and regressive CLI	Geiss et al. (2022)	– separate VOT patterns in both series of stops in all three languages

trilinguals in all of their three languages in order to trace CLI patterns over the first year of learning a new L3.

VOT in Polish, English and Norwegian

As far as the languages under investigation are concerned, they differ considerably in terms of their VOT characteristics. Polish is a true voicing language, which means that it has prevoicing in voiced stops (/b, d, g/) and short-lag VOT in voiceless stops (/p, t, k/) (e.g. Jassem, 2003; Keating et al., 1981). The occurrence of prevoicing in Polish voiced stops is high, for example, Sypiańska (2021) reported 100% of prevoiced items in Polish control group, while in Wojtkowiak (2022) fully prevoiced items constituted over 95% of the tokens. English, on the other hand, is an aspirating language that has short-lag VOT invoiced stops (/b, d, g/) and long-lag VOT in voiceless plosives (/p, t, k/) (e.g. Docherty, 1992; Lisker & Abramson, 1964). It is important to note that some sources report prevoicing in English voiced stops. However, its occurrence is rather limited, for example Docherty (1992) reports prevoicing in 7% of the items, most of which came from one speaker. For this reason, we still treat English as exceptional for its prevoicing status compared to the remaining two languages. Considering Norwegian, this language has not received as much attention in the literature in terms of VOT as English and Polish. Nevertheless, Czarnecki (2016) claims that the purely phonological difference between /p, t, k/ and /b, d, g/ pertains to aspiration or lack thereof, which might suggest that it is an aspirating language. However, the measurements taken by Halvorsen (1998) as well as Ringen and van Dommelen (2013) suggest the presence of voicing lead in the lenis stops in about 80% and 40% of the tokens, respectively, as well as aspiration in the fortis series. Based on that, in the current paper Norwegian is treated as a language with both prevoicing and aspiration.

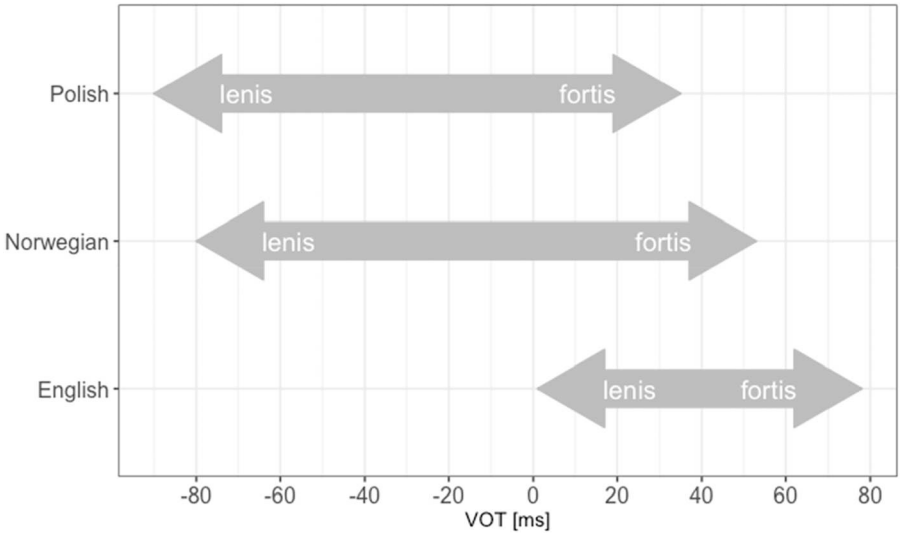


Figure 1. Comparison of VOT values of stops in Polish, English and Norwegian.

Figure 1 shows approximate VOT ranges of stops in the three languages under investigation. Yet, it is important to note that due to differences in methodologies, i.e. the durations having been measured in various contexts and by different researchers, the cross-linguistic comparison should be approached with some caution. For this reason, the values presented in Figure 1 reflect general VOT durational tendencies for the plosive sounds in the three languages generated from the overview of previous literature. Furthermore, specific baseline values are not provided as reference points in the subsequent analysis as participants investigated longitudinally are treated as their own controls (cf. e.g. Cabrelli Amaro, 2012).

A potential learning scenario for L1 Polish L2 English L3 Norwegian would be as follows: as native speakers of a voicing language (such as Polish), learners acquiring an aspirating language (such as English) as L2 would need to readjust their phonetic realisation of fortis/lenis distinction in such a way that short-lag voiceless stops would be realised as long-lag and, on the other hand, in voiced stops, prevoicing would need to be generally suppressed (with a potential caveat that English and Norwegian voiced stops are sometimes realised inconsistently in this regard). While learning a language with both prevoicing and aspiration (as in the case of Norwegian) as L3, learners would need to rely on their L1 to maintain prevoicing in the voiced stops, and on their L2 to realise long-lag voiceless stops.

The current study

Aim

The aim of this study is to analyse the production of plosive sounds by L1 Polish – L2 English – L3 Norwegian speakers. In particular, the goal is to trace the development of VOT acquisition in voiceless and voiced stops (/p, t, k/ and /b, d, g/) of all three languages (Polish, English and Norwegian) produced by novice L3 learners throughout their first year of learning the third language across three testing times. In each session, the participants were recorded reading word lists with stop sounds (voiced and voiceless) placed in stressed onset positions followed by a vowel in all three languages. The goal was to explore possible cross-linguistic interactions in the oral productions of multilingual speakers with a focus on both progressive and regressive CLI.

RQs and hypotheses

The following research questions were posed in order to investigate VOT development and possible cross-linguistic interactions in L3 learners:

RQ1: What sources of CLI can be traced in the VOT patterns in the three languages?

In relation to the first research question, predictions were formed based on certain factors that were attested in the literature for determining types of CLI. From the perspective of second language acquisition research, the most prevalent source of influence was ascribed to the first language. L1 status cannot, thus, be excluded from the current predictions, as it is assumed to rely on well-established neuromotor articulatory routines in the first language. Taking into consideration typological

(genetic) proximity, English and Norwegian are Germanic languages, thus we would expect more CLI between L2 and L3 than either of them and Polish which is a Slavic language. If we allow for the L2 status effect (see e.g. Bardel & Falk, 2007), more CLI could be predicted between L2 and L3, due to similar (formal) settings and routes of acquisition as foreign languages, as opposed to the naturalistically acquired L1. On the other hand, based on purely phonological similarity between the stops in the three languages, we could expect a twofold scenario. Since the voiced series /b, d, g/ exhibits closer similarity between L1 Polish and L3 Norwegian, we predict more CLI to occur between L1 and L3 in this category of stops. Moreover, as the voiceless series /p, t, k/ demonstrate closer proximity between L2 English and L3 Norwegian, bidirectional CLI should be attested between L2 and L3. Additionally, taking into account the assumptions of the Phonological Permeability Hypothesis (PPH, Cabrelli & Rothman, 2010), according to which the L1 system is more stable than the L2 that is acquired later in life, we do not expect major changes to affect the L1, but we foresee more CLI from L3 to L2.

RQ2: How does the acquisition of VOT in trilinguals change over time in the respective languages?

With regard to the second research question, we divided our predictions according to the three languages under investigation. In L3 Norwegian, we foresee the change in VOT values towards more target-like values, as a function of increasing language proficiency in this emerging phonological system as well as the frequency of exposure to the L3 input (cf. Dziubalska-Kołaczyk & Wrembel, 2024 forthcoming). On the other hand, L2 English is the language that the participants are already quite proficient in, thus it constitutes a more established system. Consequently, we predict that this language may be less prone to changes (cf. Wrembel et al., 2022), so that participants will produce stops with VOT close to the target English values at T1 and that these values will not change considerably over time. If any, the change would be observed in voiced stops, as this category is more affected by CLI even in proficient learners (e.g. Schwartz, 2020). As far as L1 Polish is concerned, there may be a potential L1 drift effect as a result of the L2/L3 frequency of use (cf. Chang, 2013; Wojtkowiak, 2022). Overall predictions of the degree of development in the three languages under investigation stem from their varying status, as L1, L2, L3. Because of that we expect the greatest changes towards target-like realisation in the L3 (due to intense exposure and sensitivity to change); moderate changes in the L2 (due to its relative stability and the L2 status which predicts some impact of the newly learnt L3), and little or no changes in the L1 (due to its being acquired the earliest as the first language).

RQ3: Do voiced and voiceless plosives exhibit similar trends across languages in the multilinguals' repertoire?

With respect to the third research question, based on previous research (e.g. Cal & Wrembel, 2022 Wojtkowiak, 2022;), we predict separate, language-specific patterns of VOT acquisition in the voiceless stops. That is, we expect participants to maintain VOT values in their three languages specific to each respective language, which means short-lag VOT in their L1 and long-lag VOT in L2 and L3. Moreover, the voiced series /b, d, g/ are hypothesised to be more vulnerable to CLI than voiceless /p, t, k/ due to their

Table 2. Summary of CLI predictions for fortis and lenis stops.

	CLI predictions	
Effect	Fortis stops	Lenis stops
L1 status	L1 → L3	L1 → L3
L2 status	L2 → L3	L2 → L3
Generic typology	L2 ⇄ L3	L2 ⇄ L3
PPH	L3 → L2	L3 → L2
Phonological similarity	L2 ⇄ L3	L1 ⇄ L3
Voiceless-voiced asymmetry	L1 ≠ L2 ≠ L3	L1 = L2 = L3

different phonological status according to the Onset Prominence framework and predictions (see Schwartz, 2020).

All the predicted scenarios of CLI patterns are summarised in Table 2.

Methods

Participants

The participants of this longitudinal study included 12 L1 Polish speakers, who had been learning English as their L2 and had recently started learning L3 Norwegian extensively as part of a Norwegian Studies programme at two Polish universities. The original pool of participants was larger ($N = 24$), yet the number was reduced over the period of investigation (due to drop-out and various exclusion criteria), thus we report only those participants with a complete data set for all the testing times. The group consisted of 8 females and 4 males, whose mean age was 19.9 years old ($SD = 0.93$).

The participants were asked to fill in the Language History Questionnaire (LHQ; Li et al., 2020) in order to obtain information about their language learning history and use. Their mean age of onset in L2 English and L3 Norwegian was 8.2 ($SD = 4.2$) and 19 ($SD = 1$) years, respectively. They were exposed to Norwegian extensively by taking part in Practical Norwegian classes as part of their study programme at a university. The undergraduate programme focused on literature, linguistics, history and practical knowledge of Norwegian.[4] The Practical Norwegian classes included speaking, writing, listening as well as pronunciation. They were conducted nearly entirely in Norwegian by Polish university lecturers highly proficient in Norwegian, speaking what can be broadly described as Eastern Norwegian. At the first testing time, the participants had received approximately 80 h of intensive Norwegian instruction, followed by 195 h at the second testing time, and 300 at the third testing time. Before they attended the Norwegian programme, the participants had had limited, almost none, contact with L3 Norwegian, as declared in the LHQ. As far as L2 English is concerned, they reported exposure through social media and other media platforms, especially streaming services. Their study programme did not involve any English classes during the first year of studies. Additionally, they claimed not to be proficient in or use any other foreign languages apart from those under investigation at the time of the study.

Except for the LHQ, the participants also completed proficiency tests in both of their foreign languages. Their proficiency in English was measured by means of the LexTALE

Table 3. Participants' profile data.

	Mean	SD
Age	19.9	0.93
AoA L2	8.2	4.2
AoA L3	19	1
Years of learning L2	11.9	4.5
Years of learning L3	0.3	1.1
L2 proficiency	B2	
L3 proficiency	A1	

test[5] (Lemhöfer & Broersma, 2012). Norwegian proficiency was measured using an adapted version of the Norwegian placement test that assessed proficiency up to the A2 level. Based on the grading criteria the participants were classified as intermediate (approximately B2 level) learners of L2 English and beginner (A1) learners of L3 Norwegian. Table 3 summarises the participants' profile data.

Procedure

All in all, the participants took part in three testing times that included three separate sessions for all three languages under investigation. The first testing time (T1) was conducted after 8 weeks of learning the L3. The second testing time (T2) was carried out after the first semester of learning the L3, i.e. after approximately 5 months. The third testing time (T3) took place after two semesters, that is, after 7 months of learning L3 Norwegian.

Among a battery of production and perception tests, the participants performed reading tasks of three word lists separately for each language. The stimuli included isolated one-syllable words with stop sounds (fortis and lenis) placed in stressed onset positions followed by a vowel. The words with fortis and lenis stops constituted minimal pairs that were presented in random order. All in all, the participants read three language lists with 30 token words for Polish (5 per stop), 48 for English (8 per stop) and 66 for Norwegian (11 per stop). The number of selected tokens depended on possible vowel contexts. The participants were presented with the stimuli on a computer screen and were asked to read the words as they appeared on the screen. Recording sessions were carried out in the language blocks on separate consecutive days to avoid language mixing. Firstly, the participants took part in the Norwegian block, then English and, at the end, Polish. The order of experimental tasks remained unchanged across the blocks. Recordings were conducted in a sound-treated booth or a quiet room using Shure SM-35 unidirectional cardioid head-worn condenser microphone and Marantz PMD620 recording device. The speech rate of individual participants was controlled in a twofold manner: each individual token item was placed at a separate slide and the experimenter controlled for the displaying the slides at a steady pace of delivery. The procedure remained unchanged across all three testing times.

The obtained recordings were then force-aligned using WebMAUS (Kisler et al., 2017), and VOT boundaries were manually corrected in Praat (Boersma & Weenink, 2021). The measurements for voiced stops included prevoicing which was annotated from the first noticeable pulse of voicing to the release of the closure. The burst and release noise were not included. In the voiceless stops, the measurements focused on the duration of positive VOT, which was marked from the release of the closure until the onset of

Table 4. Mean VOT values (in ms) and standard deviation (in brackets) for each stop consonant across three languages and testing times.

Polish	*p*	t	k	b	d	g
T1	46.8 (17.7)	48.7 (15.0)	72.0 (21.7)	−103.3 (30.8)	−98.3 (32.5)	−86.2 (33.0)
T2	45.0 (17.9)	55.0 (18.3)	71.5 (18.5)	−99.3 (33.2)	−88.2 (26.0)	−77.0 (25.4)
T3	46.9 (19.3)	52.5 (20.4)	73.6 (19.4)	−93.1 (32.1)	−82.5 (23.3)	−74.5 (29.2)
English						
	p	t	k	b	d	g
T1	55.4 (19.0)	69.6 (24.6)	82.7 (22.1)	−102.6 (31.1)	−87.5 (31.1)	−82.7 (32.4)
T2	59.9 (23.2)	64.3 (25.3)	88.2 (22.0)	−86.5 (30.1)	−85.4 (27.4)	−75.9 (39.6)
T3	58.4 (20.2)	66.8 (24.4)	88.5 (22.6)	−82.3 (29.4)	−80.5 (30.7)	−72.6 (30.1)
Norwegian	*p*	t	k	b	d	g
T1	45.4 (21.5)	65.0 (24.3)	79.3 (24.1)	−104.7 (42.5)	−98.9 (38.4)	−85.3 (35.9)
T2	49.5 (25.7)	67.3 (24.6)	73.0 (25.2)	−85.8 (32.2)	−90.5 (32.3)	−81.8 (30.6)
T3	53.6 (25.1)	69.2 (27.2)	78.6 (24.9)	−81.2 (28.1)	−87.6 (30.5)	−68.5 (28.6)

the subsequent vowel. All boundaries were placed at zero-crossings. The VOT durations were extracted with the use of a Praat script (Lennes, 2002).

Results

This section presents the results of VOT measurements in word-initial fortis and lenis stops (/p, t, k/ and /b, d, g/) in the three languages under investigation (L1 Polish, L2 English, L3 Norwegian) across three testing times (T1, T2, T3). Firstly, overall descriptive statistics are provided. Next, VOT measurements for the three languages are reported. Thereafter, the obtained results are compared across the three testing times. After that, a comparative analysis of the VOT values in the voiced and voiceless stops is presented. Finally, an analysis of individual trajectories of VOT development is offered.

Out of the total 5400 tokens, 528 were discarded from the analysis due to them being realised without prevoicing or mispronounced. Thus, the total number of tokens, analysed for VOT durations, was 4872. Items lacking prevoicing are discussed later in this section. Table 4 shows mean VOT values (in ms) for each stop consonant across the three languages and testing times. As can be seen, VOT values of voiceless stops have overall the shortest lag in L1 Polish and the longest lag in L2 English, while the values for L3 Norwegian stops seem to fall in-between the two. In the voiced stops, L1 Polish exhibited the longest pre-voicing values in /b/ and /g/ and L3 Norwegian in /d/, while the shortest prevoicing was found in L2 English.

A logistic mixed-effects regression model was run in R (R Core Team, 2024) using a partially Baesian method within blme package (Chung et al., 2013) to avoid singularity. VOT

Table 5. Fixed effects generated in the statistical analysis.

model term	df1	df2	F.ratio	*p*-value
Sound	5	126.91	2355.93	<.001
Session	2	11.43	8.90	.005
Language	2	130.29	15.49	<.001
Session:Sound	10	124.14	3.71	<.001
Language:Sound	10	126.12	1.19	.302
Session:Language	4	128.10	2.06	.090
Session:Language:Sound	20	122.57	1.13	.333

values were included as the dependent variable and Sound, Language and Session were added as fixed factors. The model also included interaction between Sound, Language and Session as well as by-participant and by-item random slopes for Session. *Post-hoc* analyses were carried out with the use of the emmeans package (Lenth, 2023) by applying Tukey adjustments to compare a family of three estimates. This package was also used to calculate effect sizes. The model specification was as follows: VOT ~ Sound*Language*-Session + (Session|Participant) + (Session|Item). Obtained fixed effects are presented in Table 5. The output of the model is provided in the Appendix.

The results of the analysis pointed to statistically significant main effects of Sound ($F = 2355.93$, $p < .001$), Session ($F = 8.90$, $p = .005$) and Language ($F = 15.49$, $p < .001$), along with a Sound by Session interaction ($F = 2.06$, $p < .001$).

A post hoc analysis for language revealed statistically significant differences between all three languages, that is: L2 English – L3 Norwegian ($z = 3.011$, $p = .007$), L2 English – L1 Polish ($z = 5.555$, $p < .001$) and L3 Norwegian – L1 Polish ($z = 3.265$, $p = .003$).

Mean VOT values generated across the three testing times are visualised in Figure 2. The data obtained from the first testing time (T1)shows that VOT values in L3 Norwegian voiceless stops fall in-between those of L1 Polish and L2 English, while in the voiced series, they seem to approximate L1 Polish values. The values for L2 English are the shortest. In

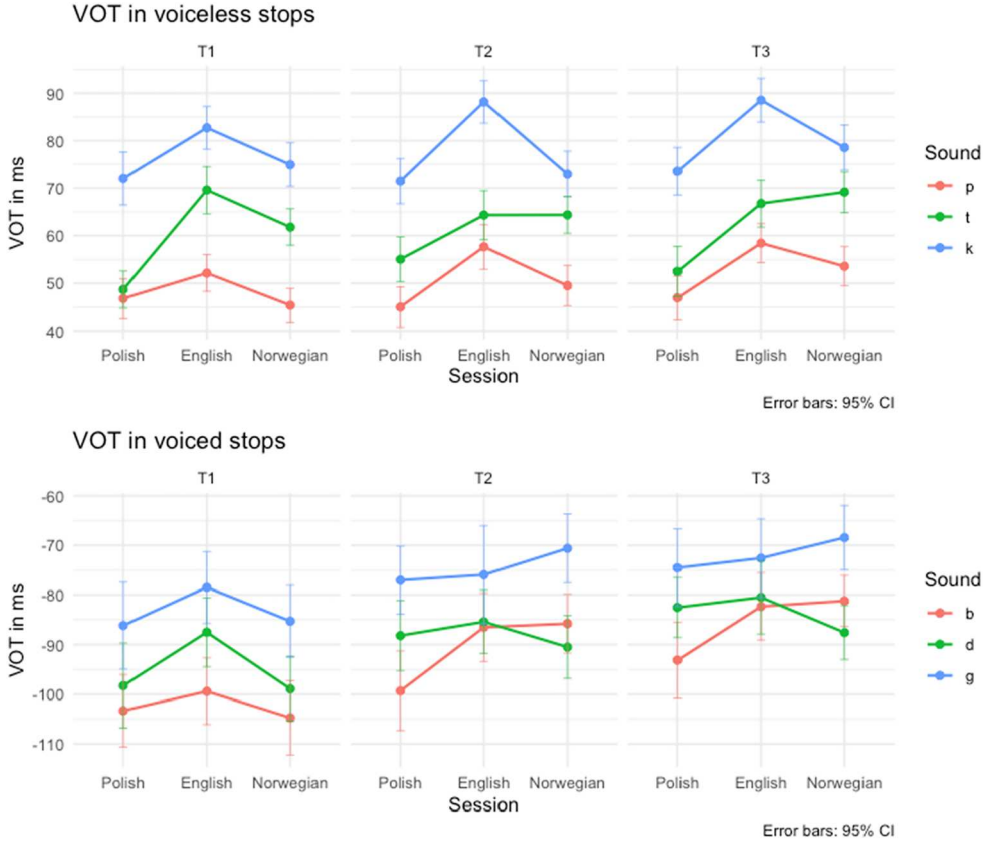


Figure 2. Mean VOT values of voiceless and voiced plosives in L1 Polish, L2 English, L3 Norwegian across the three testing times (T1-T3).

the second testing time (T2), L2 English values turned out to be the longest in /p/ and /k/, while L1 Polish and L3 Norwegian VOT values appear rather comparable. For /t/, L2 English and L3 Norwegian approximate one another, while Polish VOT remains the shortest. For the bilabial voiced plosive /b/, L1 Polish values turned out to be the longest, while in the rest of the sounds they are rather comparable. As far as the third testing time (T3) is concerned, it can clearly be observed that VOT values are considerably shorter in Polish than in L3 Norwegian and L2 English, which exhibits the longest lag for voiceless stops. The values generated for voiced stops seem to be rather comparable, with L1 Polish yielding the longest mean VOT in /b/ and L3 Norwegian in /d/. Pairwise comparisons across testing sessions pointed to statistically significant differences only between T1 and T3 ($z = -4.082$, $p < .001$).

Figure 3 shows mean VOT values of voiceless and voiced plosives produced across the three testing times in L1 Polish, L2 English, L3 Norwegian. What can be seen is that, overall, in the voiceless series of stops, there is a tendency for VOT to increase with time, which is especially visible in L2 English and L3 Norwegian /p/ and /k/. On the other hand, in the voiced series, there is a decrease in prevoicing values with time in

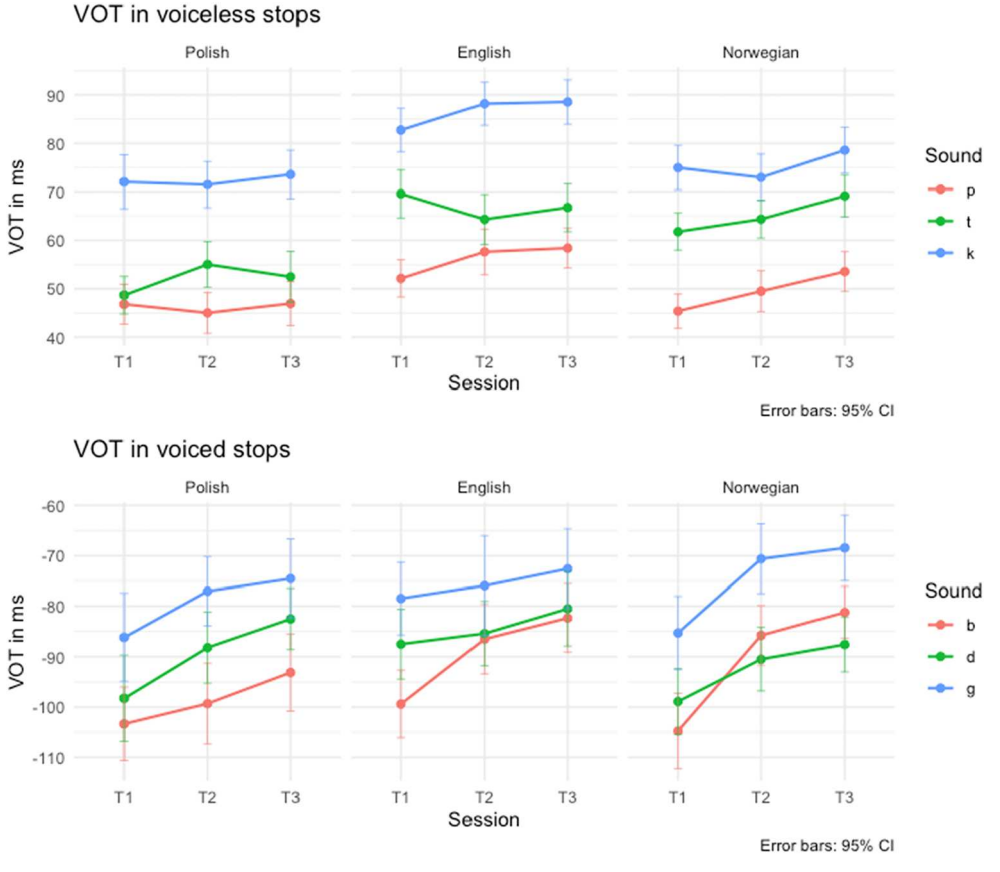


Figure 3. Mean VOT values of voiceless and voiced plosives in three testing times (T1-T3) across the three languages (L1 Polish, L2 English, L3 Norwegian).

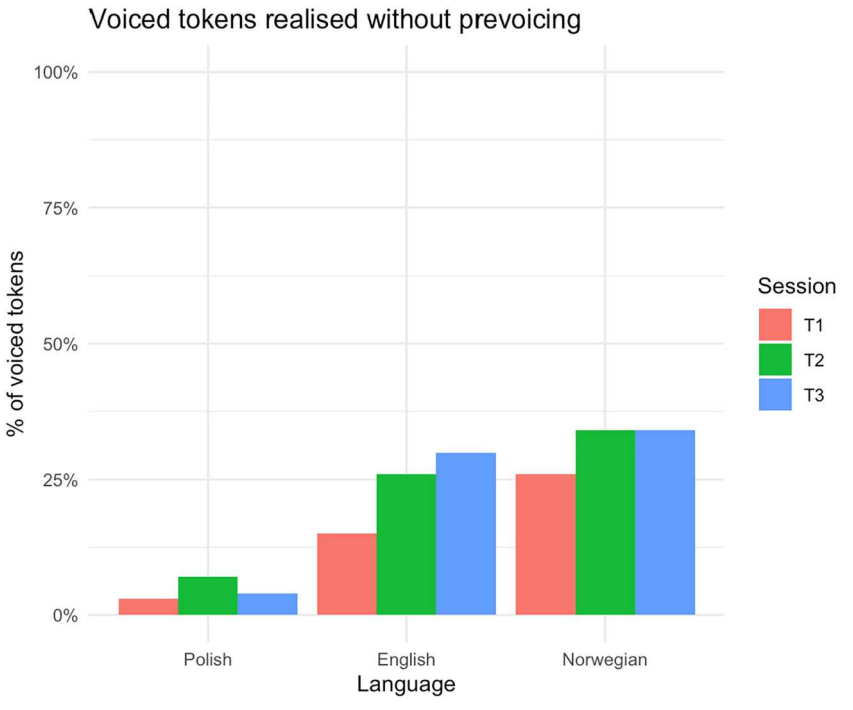


Figure 4. Percentage of unvoiced realisations of lenis stops /b, d, g/.

all three languages. What is more, as shown in Figure 4, the number of unvoiced items increased considerably with time in L2 English and L3 Norwegian, but not in L1 Polish.

A post hoc analysis of the interaction between sounds and testing sessions showed that only the voiced series of stops differed significantly across three testing times. More specifically, the differences between T1 and T2 were found for /b/ ($z = -3.389$, $p = .002$) and /g/ ($z = -2.662$, $p = .021$), while the difference between T1 and T3 was attested for all three voiced stops, namely /b/ ($z = -5.911 < .001$), /d/ ($z = -3.681$, $p < .001$) and /g/ ($z = -3.908$, $p = <.001$). This suggests that VOT values in these sounds generally increased over time, that is, prevoicing became shorter.

To provide a more nuanced understanding of the process of acquisition, we zoomed into individual variation by analysing separate trajectories of development, on the one hand, to see interactions between the three languages in particular individuals and, on the other hand, to observe the development over time. Figures 5 and 6 demonstrate individual trajectories of VOT development of voiceless and voiced stops across the three testing times in the three languages under investigation. As can be seen, the plots point to a considerable individual variation, yet some stable patterns can be identified based on the data. Taking into account the relationships between languages, in the voiceless series some participants maintain more unified VOT values across languages (i.e. participants 41, 58 and 59), while others produce more distinct VOT values in their three languages (i.e. participants 47 and 63). What can also be noticed is that in some of the cases, foreign languages (L2 English and L3 Norwegian) are produced in a similar manner that is divergent from L1 Polish. This is visible, for

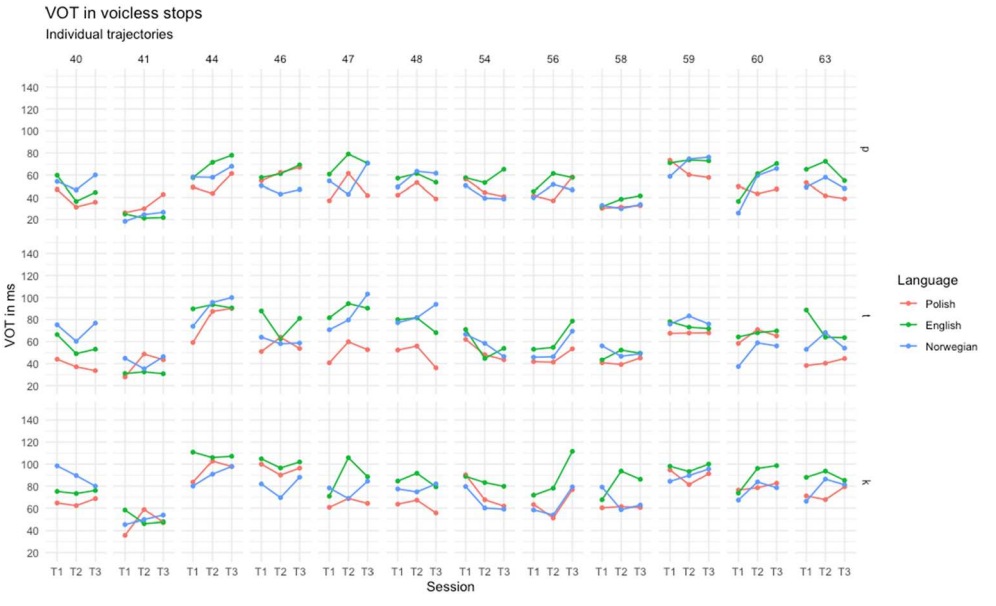


Figure 5. Individual trajectories of VOT development of voiceless stops across three testing times in the three languages.

example, in participants 46 (in /p/ and /k/) or 60 (in /p/). The voiced series appears to be more diverse, but still similar patterns of VOT trajectories can be identified. That is, unified VOT values across languages can be observed, for example in participants 47, 48

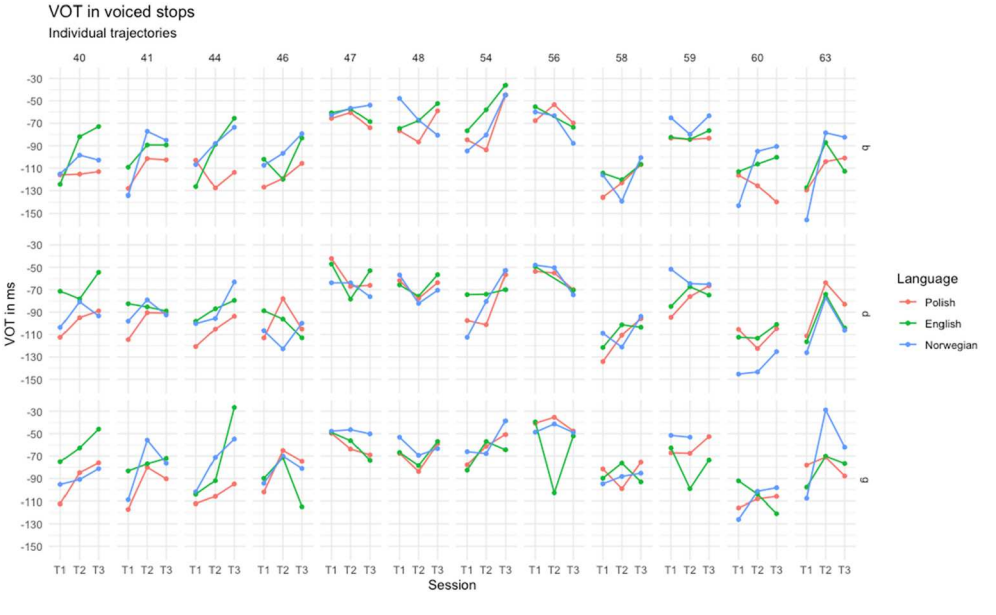


Figure 6. Individual trajectories of VOT development of voiced stops across three testing times in the three languages.

(in /d/ and /g/), 56 (in /d/) or 63 (in /d/), while more distinct VOT values in the three languages can be seen in participants 40, 41 or 60. Participant 44 maintained different VOT values for their foreign languages compared to L1 Polish. As far as the development over time is concerned, in the voiceless series some of the participants experienced an upward trend in the VOT values increasing with time in all three languages (e.g. participant 44 in /p/ and /t/, 56 in /t/ and /k/) or in the foreign ones (e.g. participant 60 in /p/). On the contrary, a downward trend in VOT durations was also attested (e.g. participant 54). Interestingly, a number of participants experienced the greatest downturn or upturn at T2.

Discussion

The aim of this paper was to analyse the production of plosive sounds by L1 Polish, L2 English, L3 Norwegian speakers and potential patterns of cross-linguistic interactions over time. Our particular objective was to trace the development of VOT acquisition in voiceless and voiced stops (/p, t, k/ and /b, d, g/) in all three languages (Polish, English and Norwegian) produced by novice L3 learners throughout their first year of learning of the third language across three testing times. To this end, three research questions were posed.

The first research question dealt with sources of CLI that could be traced for VOT patterns in the three languages. Predictions were formed based on selected factors that were attested in the literature for determining CLI directionality (cf. Puig-Mayenco et al. 2020 for a systematic overview of morphosyntactic transfer studies in third language acquisition), including typological proximity, L2 status factor (see e.g. Bardel & Falk, 2007), phonological similarity between the language systems and PPH (Cabrelli & Rothman, 2010). The present results do not confirm any of the predictions made with regards to potential sources of CLI. In fact, the VOT production patterns differed significantly across the three languages, thus indicating that multilingual learners regard the three systems as distinct. This finding is in line with previous research (Amengual, 2021; Geiss et al., 2022; Liu, 2016) which also attested separate language systems. Although Amengual's study focused only on a velar plosive /k/, it clearly showed that language-specific VOT categories were maintained in multilingual learners to a higher degree than in bilinguals. What is more, the study by Geiss et al. (2022) that analysed the production of voiced and voiceless stop sounds also indicated separate VOT patterns in both series of stops in all three languages pointing to the possibility of bilingual advantage. In turn, our participants were rather advanced in their second language (English) which might have led to a different realisation of VOT in their newly acquired L3.

The second question in this study sought to determine how VOT patterns in trilinguals change over time. We expected VOT durations in L3 Norwegian stops to change towards more target-like values, as a function of increasing language proficiency in this emerging phonological system as well as the frequency of exposure to the L3 input. We did not foresee any major changes in the L2 English stops because of the participants' rather advanced proficiency in this language, but if any, we expected more change to affect the lenis series, which would be in line with the existing literature (e.g. Gabriel et al., 2018; Schwartz, 2020; Schwartz et al., 2020; Wojtkowiak, 2022). For L1 Polish we predicted a potential L1 drift effect as a result of the L2/L3 frequency of use (e.g. Wojtkowiak, 2022).

The results, however, did not point to an interaction between Language and Session. Thus, we cannot explicitly state how particular languages were affected over time, yet we witness a change in progress between T1 and T3 in the overall system of multilingual learners. On the whole, we attested differences in VOT realisation between T1 and T3, which can be interpreted as indicating a potential influence of an additional language (i.e. L3) being added to the multilingual system in the VOT production patterns. Although we found that VOT values are significantly different in all three languages, it can be stated that the whole multilingual system changed with time in one direction, especially in terms of voiced stops, where we observed shortening of prevoicing across times. Adding Norwegian as a third language into the system might have caused that the participants associated it with L2 English (a voicing language) due to their typological proximity and as a result of the L2 status effect. This, in turn, could have led to the decrease of prevoicing in all three languages, as a gradual influence of L3 due to increasing exposure to this language with time. The overall decrease in the length of prevoicing shows that the multilingual system develops from L1 Polish-like values, towards more L2 English-like realisations. This is also visible in the number of unvoiced items (i.e. without prevoicing) which increased considerably with time, especially in L2 English and L3 Norwegian, suggesting more English-like realisation of the two languages. This attested dynamic change affecting the multilinguals' system as a whole might be interpreted as a potential manifestation of a global language entity, as proposed by Sypiańska (2017). The author postulates the existence of a global language system treating it as one entity, thus the dynamic shifts in system may impact all its component languages in a similar manner. In our case the influence is holistic and tacit, departing from a typical cross-linguistic interaction frequently reported in L3 phonological acquisition research.

The third research question aimed to determine whether fortis and lenis plosives exhibit similar trends across the languages in the multilinguals' repertoire. Based on previous research outcomes (e.g. Cal & Wrembel, 2022; Wojtkowiak, 2022), we expected to observe separate, language-specific patterns of VOT acquisition in the voiceless stops and more CLI in the voiced series. The present findings support this prediction, but only to some extent. Indeed, the interaction between sound and session pointed to significant differences between sessions in all three voiced sounds (/b, d, g/), but not in the voiceless ones (/p, t, k/). It was shown that the voiced plosives changed significantly with time, by means of shortening of prevoicing across the three sessions, while there were no significant changes in the voiceless series. What is more, the number of unvoiced items decreased with time. Previous literature also attested similar asymmetry (e.g. Gabriel et al., 2018 Schwartz, 2020; Schwartz et al., 2020; Wojtkowiak, 2022), where both voiceless and voiced series were affected by CLI in a different manner. This might be because prevoicing is a less salient feature than aspiration and, therefore, the participants did not hear the difference between voiced stops in the three languages. Consequently, these sounds tend to be less stable and more prone to changes during the process of L3 acquisition.

Notwithstanding the novelty of the contribution, there are some limitations that need to be mentioned. First of all, the major limitation of this study is a reduced pool of participants, due to a considerable drop-out rate in the original number at T1, which was the effect of the longitudinal investigation. A low number of participants, inevitably linked with longitudinal investigations in L3 acquisition, is also a limitation when it comes to fitting statistical models that include a three-way interaction. Another limitation was

related to the task type as the participants read a list of words consisting of stop consonants in the stressed onset position followed by a vowel, as it has been the most frequent practice in previous related studies. However, free speech elicitation tasks might reflect the actual production performance with more ecological validity, yet lesser control over the target tokens.

Conclusion

The present investigation analysed the production of plosive sounds by L1 Polish, L2 English, L3 Norwegian speakers by tracing the development of VOT acquisition in voiceless and voiced stops in all three languages produced by novice L3 learners throughout their first year of learning of the third language. The paper investigated cross-linguistic interactions in their oral productions and the developmental trajectory of L1, L2 and L3 speech over three testing times. To the best of our knowledge, it is the first such longitudinal study investigating the voiced and voiceless stops in all three languages over time. The study revealed that the participants maintained language-specific VOT patterns in all three languages. What is more, it showed a change in VOT production patterns over time in the voiced series of stops, namely shortening of or even loosing prevoicing. Last but not least, the study confirmed previous findings regarding the asymmetry between voiced and voiceless stops, as both of the series turned out to be affected by CLI in a different manner.

Notes

1. In this paper, the authors use the terms bilinguals and L2 learners, as well as multilinguals and L3 learners interchangeably.
2. In this paper, exposure to L3 Norwegian that lasted no more than 9 months is still considered by authors to be a recent exposure to the foreign language.
3. As suggested by the reviewer, to make the terminology more consistent, we tried to use the voiceless/voiced distinction to refer to phonetic voicing, and fortis/lenis to refer to phonological voicing.
4. Theoretical subjects, like Norwegian phonetics and phonology, Norwegian literature and introduction to linguistics, were taught in Polish. The pronunciation course covering both segmental and suprasegmental features was supplemented by a theoretical course in phonetics and phonology (30h/semester). The model accent taught in these classes was Eastern Norwegian.
5. We used LexTALE as a proxy to proficiency being aware that it should be used with caution (cf. Puig-Mayenco et al., 2023). It is important to stress that proficiency is not used as a variable in the study.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by Narodowe Centrum Nauki [grant number UMO-2020/37/B/HS2/00617].

ORCID

Zuzanna Cal http://orcid.org/0000-0002-1053-8440
Magdalena Wrembel http://orcid.org/0000-0001-7804-3199

References

Amengual, M. (2021). The acoustic realization of language-specific phonological categories despite dynamic cross-linguistic influence in bilingual and trilingual speech. *Journal of the Acoustical Society of America*, *149*(2), 1271–1284. https://doi.org/10.1121/10.0003559

Antoniou, M., Best, C. T., Tyler, M. D., & Kroos, C. (2010). Language context elicits native-like stop voicing in early bilinguals' productions in both L1 and L2. *Journal of Phonetics*, *38*(4), 640–653. https://doi.org/10.1016j.wocn.2010.09.005

Bardel, C., & Falk, Y. (2007). The role of the second language in third language acquisition: the case of Germanic syntax. *Second Language Research*, *23*, 459–484. https://doi.org/10.1177/0267658307080557

Bless, M. (2015). *Interference in early acquisition Dutch-English bilinguals: A phonetic examination of voice onset time in Dutch and English bilabial plosives*. [Unpublished master's thesis]. University of Amsterdam.

Boersma, P., & Weenink, D. (2021). Praat: Doing phonetics by computer [Computer software]. Version 6.1.50, retrieved 22 Jan 2021 from http://praat.org.

Cabrelli, J., & Rothman, J. (2010). On L3 acquisition and phonological permeability: A new test case for debates on the mental representation of Non-native phonological systems. *International Review of Applied Linguistics in Language Teaching*, *48*(2–3), 275–296.

Cabrelli Amaro, J. (2012). L3 Phonology: An understudied domain. In J. Cabrelli Amaro, S. Flynn, & J. Rothman (Eds.), *Third language acquisition in adulthood* (pp. 33–60). John Benjamins.

Cal, Z., & Wrembel, M. (2022). *Exploring the patterns of cross-linguistic influence in the acquisition of stops by L1 Polish – L2 English – L3 Norwegian speakers [Paper presentation]*. The 10th International Symposium on the Acquisition of Second Language Speech (New sounds 2022), Barcelona, Spain.

Chang, C. B. (2013). A novelty effect in the phonetic drift of the native language. *Journal of Phonetics*, *41*(6), 520–533. https://doi.org/10.1016/j.wocn.2013.09.006

Chang, C. B. (2019). Phonetic drift. In M. S. Schmid & B. Köpke (Eds.), *The Oxford handbook of language attrition* (pp. 190–203). Oxford University Press.

Chionidou, A., & Nicolaidis, K. (2015). *Voice onset time in bilingual Greek-German children*. Proceedings of the 18th International Congress of Phonetic Sciences. Glasgow, UK.

Chung, Y., Rabe-Hesketh, S., Dorie, V., Gelman, A., & Liu, J. (2013). A nondegenerate penalized likelihood estimator for variance parameters in multilevel models. *Psychometrika*, *78*(4), 685–709. https://doi.org/10.1007/s11336-013-9328-2

Czarnecki, P. (2016). *The Phonology of Quantity in Icelandic and Norwegian*. WydziałNeofilologii UAM w Poznaniu.

Docherty, G. (1992). *The timing of voicing in British English obstruents*. de Gruyer.

Dziubalska-Kołaczyk, K., & Wrembel, M. (2024, forthcoming). A revised natural growth theory of acquisition: Evidence from L3 phonology. In E. Babatsouli (Ed.), *Multilingual acquisition and learning: An ecosystemic view to diversity*. John Benjamins.

Flege, J. E. (1987). The production of 'new' and 'similar' phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics*, *15*(1), 47–65. https://doi.org/10.1016/S0095-4470(19)30537-6

Flege, J. E. (1991). Age of learning affects the authenticity of voice-onset time (VOT) in stop consonants produced in a second language. *Journal of the Acoustical Society of America*, *89*(1), 395–411. https://doi.org/10.1121/1.400473

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–276). York Press.

Flege, J. E., & Eefting, W. (1987). Cross-language switching in stop consonant perception and production by Dutch speakers of English. *Speech Communication*, *6*(3), 185–202. https://doi.org/10.1016/0167-6393(87)90025-2

Flege, J. E., & Hillenbrand, J. (1984). Limits on pronunciation accuracy in adult foreign language speech production. *Journal of the Acoustical Society of America*, *76*(3), 708–721. https://doi.org/10.1121/1.391257

Gabriel, C., Krause, M., & Dittmers, T. (2018). VOT production in multilingual learners of French as a foreign language: Cross-linguistic influence from the heritage languages Russian and Turkish. *Revue française de linguistiqueappliquée*, *1*(1), 59–72. https://doi.org/10.3917/rfla.231.0059

Gabriel, C., Kupisch, T., & Seoudy, J. (2016). Vot in French as a foreign language: A production and PerceptionStudy with mono- and multilingual learners (German/mandarin-Chinese). In C. Gabriel, T. Kupisch, & J. Seoudy (Eds.), *Shs Web of conferences* (Vol. 27, pp. 1–14). EDP Sciences.

Geiss, M., Gumbsheimer, S., Lloyd-Smith, A., Schmid, S., & Kupisch, T. (2022). Voice onset time in multilingual speakers: Italian heritage speakers in Germany with L3 English. *Studies in Second Language Acquisition*, *44*(2), 435–459. https://doi.org/10.1017/S0272263121000280

Halvorsen, B. (1998). *Timing relations in Norwegian stops.* [Unpublished doctoral dissertation]. University of Bergen.

Hansen Edwards, J. G., & Zampini, M. L. (Eds) (2008). *Phonology and second language Acquisition*. John Benjamins.

Hazan, V. L., & Boulakia, G. (1993). Perception and production of a voicing contrast by French-English bilinguals. *Language and Speech*, *36*(1), 17–38. https://doi.org/10.1177/002383099303600102

Herd, W., Walden, R., Knight, W., & Alexander, S. (2015). Phonetic drift in a first language dominant environment. *Proceedings of Meetings on Acoustics*, *23*(1), 060005. https://doi.org/10.1121/2.0000100

Jassem, W. (2003). Polish. *Journal of the International Phonetic Association*, *33*(1), 103–107. https://doi.org/10.1017/S0025100303001191

Keating, P., Mikoś, M., & Ganong III, W. (1981). A cross-language study of range of VOT in the perception of stop consonant voicing. *Journal of the Acoustical Society of America*, *70*(5), 1260–1271. https://doi.org/10.1121/1.387139

Kisler, T., Reichel, U. D., & Schiel, F. (2017). Multilingual processing of speech via web services. *Computer Speech & Language*, *45*, 326–347. https://doi.org/10.1016/j.csl.2017.01.005

Kopečková, R., Gut, U., Wrembel, M., & Balas, A. (2023). Phonological cross-linguistic influence at the initial stages of L3 acquisition. *Second Language Research*, *39*(4), 1107–1131. https://doi.org/10.1177/02676583221123994

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid Lexical Test for advanced learners of English. *Behavior Research Methods*, *44*(2), 325–343. https://doi.org/10.3758/s13428-011-0146-0

Lennes, M. (2002). Praat script.

Lenth, R. V. (2023). emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.8.8.

Li, P., Zhang, F., Yu, A., & Zhao, X. (2020). Language History Questionnaire (LHQ3): An enhanced tool for assessing multilingual experience. *Bilingualism: Language and Cognition*, *23*(5), 938–944. https://doi.org/10.1017/S1366728918001153

Lisker, L., & Abramson, A. (1964). A cross-language study of voicing in initial stops. *Word*, *20*(3), 384–422. https://doi.org/10.1080/00437956.1964.11659830

Liu, Z. (2016). *Exploring cross-linguistic influence: Perception and production of L1, L2 and L3 bilabial stops by mandarin Chinese speakers*. [Master's thesis. UniversitatAutònoma de Barcelona].

Llama, R., Cardoso, W., & Collins, L. (2010). The influence of language distance and language status on the acquisition of L3 phonology. *International Journal of Multilingualism*, *7*(1), 39–57. https://doi.org/10.1080/14790710902972255

Llama, R., & López-Morelos, L. P. (2016). VOT production by Spanish heritage speakers in a trilingual context. *International Journal of Multilingualism*, *13*(4), 444–458. https://doi.org/10.1080/14790718.2016.1217602

Nelson, C. (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. https://doi.org/10.3390/languages7020078

Newlin-Łukowicz, L. (2014). From interference to transfer in language contact: Variation in voice onset time. *Language Variation and Change*, *26*(3), 359–385. https://doi.org/10.1017/S0954394514000167

Puig-Mayenco, E., Chaouch-Orozco, A., Liu, H., & Martín-Villena, F. (2023). The LexTALE as a measure of L2 global proficiency: A cautionary tale based on a partial replication of Lemhöfer and Broersma (2012). *Linguistic Approaches to Bilingualism*, *13*(3), 299–314. https://doi.org/10.1075/lab.22048.pui

Puig-Mayenco, E., González Alonso, J., & Rothman, J. (2020). A systematic review of transfer studies in third language acquisition. *Second Language Research*, *36*(1), 31–64. https://doi.org/10.1177/0267658318809147

R Core Team. (2024). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Ringen, C., & van Dommelen, W. A. (2013). Quantity and laryngeal contrasts in Norwegian. *Journal of Phonetics*, *41*(6), 479–490. https://doi.org/10.1016/j.wocn.2013.09.001

Schwartz, G. (2020). Asymmetrical cross-language phonetic interaction: Phonological implications. *Linguistic Approaches to Bilingualism*, *12*(2), 103–132. https://doi.org/10.1075/lab.19092.sch

Schwartz, G., Wojtkowiak, E., & Dzierla, J. (2020). Laryngeal phonology and asymmetrical cross-language phonetic influence. In M. Wrembel, A. Kiełkiewicz-Janowiak, & P. Gąsiorowski (Eds.), *Approaches to the study of sound structure and speech: interdisciplinary work in honour of Katarzyna Dziubalska-Kołaczyk* (pp. 316–326). Routledge.

Sučková, M. (2018). *First language attrition and maintenance in the accents of native speakers of English living in the Czech Republic* [Paper presentation]. The 12th conference on native and non-native accents of English (Accents 2018), Łódź, Poland.

Sypiańska, J. (2013). *Quantity and quality of language use and L1 attrition of Polish due to L2 Danish and L3 English*. [Unpublished doctoral dissertation]. Adam Mickiewicz University, Poznań.

Sypiańska, J. (2013). *Quantity and quality of language use and L1 attrition of Polish due to L2 Danish and L3 English* [Unpublished doctoral dissertation]. Adam Mickiewicz University.

Sypiańska, J. (2017). *Cross-linguistic influence in bilinguals and multilinguals*. Wydawnictwo Naukowe UAM.

Sypiańska, J. (2021). Production of voice onset time (VOT) by senior Polish learners of English. *Open Linguistics*, *7*(*1*), 316–330. http://doi.org/10.1515/opli-2021-0016

Thornburgh, D. F., & Ryalls, J. F. (1998). Voice onset time in Spanish-English bilinguals: early versus late learners of English. *Journal of Communication Disorders*, *31*(3), 215–229. https://doi.org/10.1016/S0021-9924(97)00053-1

Tremblay, M. C. (2007). *L2 influence on L3 pronunciation: Native-like VOT in the L3 Japanese of English-French bilinguals*. [Paper presentation]. The Satellite Workshop of ICPhS XVI, Freiburg, Germany.

Williams, L. (1977). The voicing contrast in Spanish. *Journal of Phonetics*, *5*(2), 169–184. https://doi.org/10.1016/S0095-4470(19)31127-1

Wojtkowiak, E. (2022). *L1 phonetic drift in the speech of Polish learners of English: Phonological implications.* [Unpublished doctoral dissertation]. Adam Mickiewicz University, Poznań.

Wrembel, M. (2011). Cross-linguistic influence in third language acquisition of voice onset time. In Wai-Sum Lee, & Eric Zee (Eds.), *Proceedings of the 17th International Congress of Phonetic Sciences* (pp. 2157–2160). City University of Hong Kong.

Wrembel, M. (2015). *In search of a new perspective: Cross-linguistic influence in the acquisition of third language phonology*. Poznań: Wydawnictwo Naukowe UAM.

Wrembel, M., Gut, U., Kopečková, R., & Balas, A. (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*, *21*(1), 92–111.

Wunder, E. M. (2010). Phonological cross-linguistic influence in third or additional language acquisition. *Proceedings of 6th New Sounds*, *2010*, 566–571.

Zając, M. (2015). *Phonetic convergence in the speech of Polish learners of English*. [Unpublished doctoral dissertation]. University of Łódź.

Appendix. Model output

Fixed effects				
	Estimate	Std. Error	t value	*p*-value
(Intercept)	−98.54	5.64	−17.47	<.001
TimeT2	12.78	4.69	2.72	.007
TimeT3	17.43	4.34	4.02	<.001
LanguageNorwegian	−4.94	4.50	−1.10	.272
LanguagePolish	−4.69	5.24	−0.90	.370
Soundd	12.63	4.90	2.58	.010
Soundg	21.28	4.98	4.28	<.001
Soundk	181.26	4.85	37.38	<.001
Soundp	150.69	4.86	31.04	<.001
Soundt	168.12	4.85	34.67	<.001
TimeT2:LanguageNorwegian	4.72	5.09	0.93	.354
TimeT3:LanguageNorwegian	6.18	5.01	1.23	.218
TimeT2:LanguagePolish	−8.91	5.81	−1.53	.125
TimeT3:LanguagePolish	−7.43	5.70	−1.30	.192
TimeT2:Soundd	−11.51	5.64	−2.04	.041
TimeT3:Soundd	−11.02	5.61	−1.96	.050
TimeT2:Soundg	−8.72	5.76	−1.51	.130
TimeT3:Soundg	−10.55	5.76	−1.83	.067
TimeT2:Soundk	−7.35	5.37	−1.37	.171
TimeT3:Soundk	−11.62	5.28	−2.20	.028
TimeT2:Soundp	−7.29	5.38	−1.36	.176
TimeT3:Soundp	−11.17	5.29	−2.11	.035
TimeT2:Soundt	−18.10	5.38	−3.36	.001
TimeT3:Soundt	−20.25	5.28	−3.84	<.001
LanguageNorwegian:Soundd	−5.67	6.16	−0.92	.358
LanguagePolish:Soundd	−7.37	7.60	−0.97	.332
LanguageNorwegian:Soundg	−1.72	6.60	−0.26	.795
LanguagePolish:Soundg	−4.37	7.66	−0.57	.568
LanguageNorwegian:Soundk	−2.92	6.45	−0.45	.651
LanguagePolish:Soundk	−5.99	7.55	−0.79	.427
LanguageNorwegian:Soundp	−1.66	6.23	−0.27	.790
LanguagePolish:Soundp	−0.67	7.35	−0.09	.928
LanguageNorwegian:Soundt	−2.78	6.21	−0.45	.655
LanguagePolish:Soundt	−16.22	7.55	−2.15	.032
TimeT2:LanguageNorwegian:Soundd	1.94	7.19	0.27	.788
TimeT3:LanguageNorwegian:Soundd	−1.83	7.12	−0.26	.797
TimeT2:LanguagePolish:Soundd	18.10	8.50	2.13	.033
TimeT3:LanguagePolish:Soundd	16.53	8.33	1.99	.047
TimeT2:LanguageNorwegian:Soundg	6.10	7.61	0.80	.423
TimeT3:LanguageNorwegian:Soundg	4.13	7.56	0.55	.585
TimeT2:LanguagePolish:Soundg	14.10	8.59	1.64	.101
TimeT3:LanguagePolish:Soundg	12.71	8.47	1.50	.133
TimeT2:LanguageNorwegian:Soundk	−12.04	7.13	−1.69	.091
TimeT3:LanguageNorwegian:Soundk	−8.27	6.99	−1.18	.237
TimeT2:LanguagePolish:Soundk	2.92	8.25	0.35	.724
TimeT3:LanguagePolish:Soundk	3.16	8.08	0.39	.696
TimeT2:LanguageNorwegian:Soundp	−6.01	6.92	−0.87	.386
TimeT3:LanguageNorwegian:Soundp	−4.14	6.79	−0.61	.542
TimeT2:LanguagePolish:Soundp	1.63	8.04	0.20	.839
TimeT3:LanguagePolish:Soundp	1.31	7.87	0.17	.867
TimeT2:LanguageNorwegian:Soundt	3.08	6.88	0.45	.654
TimeT3:LanguageNorwegian:Soundt	3.94	6.75	0.58	.559
TimeT2:LanguagePolish:Soundt	20.58	8.26	2.49	.013
TimeT3:LanguagePolish:Soundt	14.06	8.08	1.74	.082

(*Continued*)

Continued.

		Fixed effects			
		Estimate	Std. Error	t value	*p*-value
Random effects					
Groups	**Name**	**Variance**	**Std.Dev.**	**Corr**	
Item	(Intercept)	42.683	6.533		
	TimeT2	9.333	3.055	0.15	
	TimeT3	4.587	2.142	−0.16	0.68
Participant	(Intercept)	235.833	15.357		
	TimeT2	76.470	8.745	−0.65	
	TimeT3	42.779	6.541	−0.63	0.72
Residual		578.007	24.042		
Number of obs: 4869, groups: Item, 150; Partitipants: 12					
Scaled residuals					
Min	**1Q**	**Median**	**3Q**	**Max**	**Min**
−7.5979	−0.6132	−0.0141	0.6286	4.2962	−7.5979

Appendix H: Publication 2 (Cal, 2026a)

Cal, Zuzanna. 2026. “Crossing the boundaries: Perception of voice onset time in word-initial stops by multilingual learners”, *Language and Speech 0*(0): 1–24. https://doi.org/10.1177/00238309251414959.

84:1110403117

Check for updates

Original Article

Language *and* Speech

Crossing the Boundaries: Perception of Voice Onset Time in Word-Initial Stops by Multilingual Learners

Language *and* Speech
1–24
© The Author(s) 2026
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/00238309251414959
journals.sagepub.com/home/las
Sage

Zuzanna Cal
Faculty of English, Adam Mickiewicz University, Poland

Abstract

The perception of voice onset time (VOT) in word-initial stops has been widely studied in the second language, although it remains underexplored from a multilingual perspective. Previous limited research on multilingual VOT perception offers inconsistent results, with studies suggesting either cross-linguistic interactions or language-specific VOT patterns. This study aims to consolidate these findings by investigating VOT perception in trilingual speakers of L1 Polish, L2 English, and L3 Norwegian, as well as examining whether perceptual boundary locations between voiced and voiceless stops in L3 learners differ from those of native speakers and are influenced by L1, L2, and L3 interactions. To this end, VOT continua were created for three places of articulation (labial, coronal, velar) and three languages (Polish, English, Norwegian). Participants completed a two-alternative forced-choice task separately for each language. The results indicate that trilingual speakers experience cross-linguistic interactions, particularly based on the place of articulation, highlighting the complex nature of VOT perception in multilingual acquisition.

Keywords

Third language acquisition, speech perception, multilingual perception, cross-linguistic influence, voice onset time

Introduction

The perception of voice onset time (VOT) in word-initial stops has been quite widely studied by researchers exploring second language acquisition (e.g., Aliaga-García & Mora, 2009; Gorba, 2019; Gorba & Cebrian, 2021; Pan et al., 2022; Rojczyk, 2010), although it remains largely underexplored from a multilingual perspective. With more languages added to the repertoire, speech perception might turn out to be a complex phenomenon. It is so especially due to an interplay of multiple language systems as well as potentially overlapping phonetic categories, which leads to cross-linguistic influence (CLI). Furthermore, in the light of ongoing discussions on multilingual

Corresponding author:
Zuzanna Cal, Faculty of English, Adam Mickiewicz University, Grunwaldzka 6, Poznań 60-780, Poland.
Email: zuzanna.cal@amu.edu.pl

advantage in the perception of novel contrasts, it seems vital to further inform the extant theories and research.

The existing body of literature on VOT perception in multilinguals provides rather inconsistent results. Although some of the studies provide evidence for more than one perceptual system or language-specific VOT patterns (Cal & Wrembel, 2023; Z. Liu, 2016), others report cross-linguistic interactions resulting in overlapping rather than discrete or separate systems (J. Liu et al., 2019; Stan, 2024; Zeng, 2024). What is more, the analyses offered in these studies tend to be focused on one place of articulation (PoA) and are conducted on speakers with varying language repertoires. It might, therefore, be difficult to come up with a unified conclusion on how CLI behaves in VOT perception in multilinguals. This article aims to contribute to the growing body of research on multilingual perception by attempting to consolidate the previous findings and further inform the theoretical models of non-native speech perception. The goal of the study is, thus, to explore VOT perception in all three languages of trilingual speakers of L1 Polish, L2 English, and L3 Norwegian as well as in all three places of articulation (bilabial, coronal, and velar). In particular, the study will examine whether perceptual boundary locations between voiced and voiceless stops in L3 learners differ from those of native speakers and to what extent they are influenced by L1, L2, and L3 interactions.

This article has been organized in the following manner: first, it provides a brief overview of speech perception research in multilinguals, focusing specifically on cross-linguistic relationships. Then, it offers a review of previously conducted studies on VOT perception, both from the perspective of L2 and L3 learners. Thereafter, the study is described together with its methodology and findings, which is followed by the discussion of the results as well as conclusions.

2 CLI in multilingual perception

Although research on CLI in bilingual perception is extensive and well-established (e.g., Aliaga-García & Mora, 2009; Gorba, 2019; Gorba & Cebrian, 2021; Pan et al., 2022; Rojczyk, 2010), interest in multilingual perception has been growing in recent years. Although it remains less explored, leaving gaps in understanding how multiple phonological systems interact.

The majority of multilingual studies to date have focused on the role of previously acquired languages in the perception of a newly acquired language, that is, L3. It has been postulated that the knowledge of multiple languages can generally enhance perception in L3, because of the so-called multilingual advantage (Kopečková, 2015). It is because multilinguals are thought to have greater phonological awareness and cognitive flexibility, due to a prior knowledge of two (or more) languages, which allows them to differentiate between the sounds more easily and spot more subtle differences between the languages (Gut, 2010). This tendency has been confirmed by a number of studies which found that multilingual learners perceived novel contrasts in L3 better than L2 learners (e.g., Antoniou et al., 2015; Balas, 2018; Enomoto, 1994; Kopečková, 2015; Tremblay & Sabourin, 2012). A similar tendency was revealed also in the study by Wrembel et al. (2019) on the perception of sibilant pairs. It was found that multilingual learners assimilated L3 sounds to both L1 and L2 categories, with a preference for the latter, and, what is more, beginner L3 learners were more likely to perceive subtle acoustic differences in novel phonological contrasts. In this light, multilingual learners experience an advantage that allows them to maintain language-specific perceptual patterns and experience little to no CLI. On the contrary, Patihis et al. (2013) provided evidence suggesting that multilingual advantage only plays a role when L3 is a language with similar phonemic categories to previously learned ones. In their study on Korean stop discrimination, they found that learners of languages dissimilar to Korean in terms of stop consonants experienced more difficulty with the task.

Even though the number of studies centered specifically on L3 perceptual acquisition seems to be growing, research devoted to regressive CLI is rather scarce. Existing results show that learners tend to keep their language systems apart, and L1 as well as L2 are not affected by L3 acquisition. This tendency was observed, for example, by Cabreli Amaro (2017), who examined the acquisition of L3 Brazilian Portuguese by L1 English-L2 Spanish and L1 Spanish-L2 English bilinguals. The results did not show any L1 and L2 vulnerability to L3 influence. Similarly, Balas et al. (2019) in their study on rhotic acquisition by multilingual children found that the participants maintained a stable, optimal state of L2 rhotic perception, achieving high accuracy in the task. However, quite an opposite tendency was established by Sypiańska and Cal (2022) who provided evidence for perceptual drift in the perception of an open-mid front unrounded vowel /ɛ/ by L1 Polish-L2 English-L3 Spanish speakers divided into three groups with different Spanish exposure (Polish residents exposed to Spanish through instruction, Polish immigrants to Spain exposed to Spanish through immersion, and a control group with no exposure to Spanish). It was demonstrated that the group with the highest L3 exposure experienced a boundary shift toward a higher F1 value, suggesting the possibility of L1 drift.

Examination of individual variation patterns in multilingual perception also seems to be lacking in the literature. To the best of our knowledge, there is only one perceptual study on multilinguals that addresses this topic directly. Gut et al. (2023) analyzed individual variation in the perception of /v-w/ contrast, rhotics, final obstruent (de)voicing, and vowel reduction. Their results pointed to great variety across phonetic features except for rhotics, for which rather consistent patterns were obtained. Wrembel et al. (2022) also provided some insights into individual variation; however, their research focused on the perception–production interface. Nevertheless, their findings provide some evidence for rather high individual variation in the perception of rhotic sounds.

To conclude, the existing literature on multilingual perception points to interesting tendencies suggesting that multilingual speakers tend to keep their languages separate, leading to the absence of CLI, especially when it comes to languages that share similar phonetic features. However, the studies regarding regressive CLI provide rather inconclusive results, suggesting either no effect of L3 on the already learned languages or L1 drift effect in more experienced learners. More research is still needed to clarify how languages interact at different levels of exposure, as well as to understand how the degree of participants' unfamiliarity with phonetic contrasts affects these relationships.

2.1 VOT perception in bi-/multilingual acquisition

VOT perception has been explored from various angles in second language acquisition research. A large amount of this work explores whether bilinguals maintain separate phonetic systems for each language or whether they experience CLI, which results in shifted or merged perceptual categories. Across studies, there is no clear consensus on whether, and to what extent, the two languages are affected by CLI. Some early papers point to separate VOT perceptual systems in L1 and L2 (e.g., Williams, 1977). However, in a more recent study, early bilinguals were shown to be able to integrate both language systems in a common phonetic space depending on the language context (Antoniou et al., 2012), at the same time also maintaining largely separate categories for each language while making phonological judgments. On the contrary, some studies show clear L1 influence on L2 perception. For instance, it was found that L1 Spanish-L2 English bilinguals categorize short-lag tokens as voiceless more often than native English listeners (e.g., Flege & Eefting, 1987; Montanari et al., 2023). Together, these findings point to CLI as a dynamic phenomenon, sensitive to language dominance and context.

These results align with the most prominent models of L2 speech perception, that is, the Speech Learning Model (SLM; Flege, 1995) and the revised Speech Learning Model (SLM-r, Flege & Bohn, 2021), the Second Language Linguistic Perception Model (L2LP; Escudero, 2005), and the Perceptual Assimilation Model (PAM-L2; Best & Tyler, 2007), which are based on the claim that an L2 learners may experience CLI stemming from their native language. The SLM (Flege, 1995), which is the only framework accounting for regressive CLI, proposes that L2 learner's perception is influenced by how similar or different L2 sounds are to those in the learner's L1. It predicts that when L2 sounds are perceived as distinct from L1 sounds, they are easier to acquire, while similar sounds are more challenging due to interference from the already established L1 phonetic categories because of equivalence classification. Consequently, a merged category is formed between L1 and L2, leading to cross-linguistic interactions.

The studies on regressive CLI from L2 to L1 report a varying degree of CLI in bilingual speakers, from no L1 drift (Sypiańska & Cal, 2023) to L1 drift being attested (Gorba, 2019), even after a 1.5-hr exposure to a new language (Kellogg & Chang, 2023). Recently, however, the main focus of L2 investigations into L2 stop perception seems to be placed mostly on the role of various factors, such as the age of acquisition (e.g., Pan et al., 2022; Stölten, 2013), L2 phonetic training (e.g., Aliaga-García & Mora, 2009), and L2 exposure (e.g., Gorba, 2019; Gorba & Cebrian, 2021).

When it comes to the population of L1 Polish-L2 English speakers, which constitute two out of three languages investigated in this article (cf. Section 3.1), the research specifically examining VOT perception in this population is rather scarce. Rojczyk (2010) examined the perception of English /k/ in beginner and advanced L1 Polish learners compared with L1 English controls. The study found that even beginner learners exhibit some sensitivity to the English voicing boundary within the positive VOT range, indicating relatively early perceptual attunement. Advanced learners approached near-categorical perception of the English contrast. However, their categorization peak remained lower than that of native English speakers, who consistently placed the boundary at higher VOT values. Importantly, the study shows that Polish learners never fully aligned with the native-like perceptual pattern, suggesting persistent L1 influence despite increased L2 proficiency. This supports the view that L2 perceptual categories may remain partially distinct from L1, even in advanced bilinguals. The study also examined L2 influence on L1 by analyzing learners' categorization of the stimuli with 0 ms VOT, yet no L2-L1 effect was found in the data. In line with these findings, a more recent study by Sypiańska and Cal (2023) on VOT perception by L1 Polish-L2 English speakers found no evidence of L1 drift in perception, contrary to what has been observed in production. The authors conclude that there may be a perception–production mismatch, indicating that L1 drift could manifest primarily at the articulatory level without affecting underlying phonological representations.

As far as methodology is concerned, the existing studies examine VOT perception using a variety of tasks; yet, the most prominent one in the literature is a two-alternative forced-choice identification task of VOT continua (e.g., Aliaga-García & Mora, 2009; Gorba, 2019; Gorba & Cebrian, 2021; Pan et al., 2022; Rojczyk, 2010), where participants are presented with a sound stimulus and then asked to choose which word they heard out of the two responses presented to them. Some studies also include an AX (e.g., McKelvie-Sebileau & Davis, 2014) and AXB discrimination task and its variations (e.g., Alves & Zimmer, 2015), as well as a go/no-go task (e.g., García-Sierra et al., 2021). The collected data are usually analyzed in terms of boundary and slope, as well as the accuracy of responses.

Although a considerable body of research on VOT perception has been dedicated to second language acquisition, with analyses approaching it from various perspectives, the literature addressing VOT perception within a multilingual context remains relatively limited. The existing studies generally point to cross-linguistic interactions taking place in multilingual systems, but as they

focus on different languages and their statuses (L3 vs. all three languages), places of articulation, and measurements taken, it is difficult to draw far-reaching conclusions based on their findings.

The first available study that deals with VOT perception by multilingual learners is by Z. Liu (2016) who analyzed perception and production of bilabial stops by L1 Mandarin Chinese, L2 English, and L3 Spanish speakers in all three languages. The analysis focused on accuracy and boundary locations obtained from a two-alternative forced-choice task. It was shown that L3 learners maintained more than one perceptual system compared with L2 learners. The results also indicated that long VOT in L1 and/or L2 lengthened the boundary in L3, which is an indication of CLI. As far as regressive CLI is concerned, the findings pointed to L1 drift effect, but there was no influence on L2. Contrary results were obtained by J. Liu et al. (2019), who analyzed the effect of language learning on the VOT perception of bilabial stops in L1 Mandarin Chinese, L2 English, and L3 Spanish speakers. The findings did not provide evidence for separate perceptual systems in multilinguals, as the obtained values for the perceptual boundary locations were similar across the three languages. The study showed influence of the L1 (or L2) on the L3 but also provided some evidence for regressive CLI, as VOT boundaries for L1 Mandarin Chinese were much lower than in the control group. Stan (2024), in their study on L1 English-L2 Spanish and L1 Spanish-L2 English perceptual acquisition of L3 French bilabial stops, also found evidence for progressive influence stemming from L1/L2 or only L1 on L3. However, this study did not observe any regressive CLI. Similar results were obtained by Zeng (2024), who found evidence for both L1 Mandarin and L2 English influence on L3 Japanese stop perception.

Yet another outcome was generated in the study by Cal and Wrembel (2023), who examined VOT perception by initial learners of L3 Norwegian with the L1 Polish-L2 English background and focused on all three places of articulation of stops, namely bilabial, coronal, and velar. Three separate perceptual systems were found for the bilabial stops, whereas for the coronal stops, an L1-to-L3 influence was attested. In the velar stops, a potential L1-to-L2 influence was reported. It was concluded that at the onset of L3 acquisition, the multilinguals maintained patterns of VOT categorization specific to a language and that possible cross-linguistic interactions depended on the PoA. This outcome may have been influenced by the varying perceptual salience associated with different places of articulation, as well as by differences in the VOT continua ranges used in the study, which varied across languages and places of articulation, potentially contributing to a range effect (Keating et al., 1981). The current research builds upon this study by broadening the scope of investigation to include not only initial learners but also a more diverse population with varying proficiency levels. It further improves upon the earlier design through the inclusion of a control group and the use of unified VOT continua across languages and places of articulation, which aims to enhance comparability.

In turn, J. Liu and Lin (2021) investigated the perception and production of L3 stop contrasts by L1 Mandarin Chinese, L2 English, and L3 Japanese or Russian speakers. An identification task was conducted to investigate the accuracy in perception of voiceless and voiced stops at three places of articulation in L3 Japanese or Russian. It was found that L3 learners were less accurate in perceiving L3 stops than control speakers (native Japanese and Russian speakers). However, the learners' accuracy was higher for voiced stops in L3 Japanese and L3 Russian than for voiceless stops, as prevoicing was a novel phonological feature for them. Nevertheless, for the voiceless stops, the scores were higher with the increase of VOT values.

Recapitulating, only a handful of studies to date have examined VOT perception by multilingual learners in a systematic way. The results undoubtedly shed some light on CLI directionality; however, more research is needed to provide a deeper understanding of the processes that take place in multilingual minds. This article attempts to add up to this growing body of literature by

analyzing VOT perception by trilingual learners in all their languages, focusing on all three places of articulation.

3 Current study

3.1 Aim

The aim of this study is to explore how trilingual speakers of L1 Polish, L2 English, and L3 Norwegian perceive word-initial voiced and voiceless stops in their three languages. More specifically, the objective is to determine whether the obtained perceptual boundary locations between voiced and voiceless plosives (measured by VOT in ms) for the trilingual speakers are subject to CLI and to assess the directionality of the CLI process. Moreover, the study seeks to obtain data that would allow to establish whether the boundaries in the trilingual group differ from the values found in the native speakers of the respective languages. In addition, the analysis offers some insights into the individual variation across the participants.

The languages under investigation are L1 Polish, L2 English, and L3 Norwegian. These three languages were selected specifically because they exhibit complementary patterns in VOT of their stop consonants, allowing for a systematic comparison of CLI and acquisition patterns. First, Polish is a true voicing language that has prevoicing in the voiced series of stops and short-lag VOT in voiceless stops (e.g., Jassem, 2003; Keating et al., 1981). English, on the contrary, is an aspirating language in which /b, d, g/ are partially voiced and /p, t, k/ are aspirated (e.g., Docherty, 1992; Lisker & Abramson, 1964). It is worth noting that some studies have observed prevoicing in English voiced stops; however, its presence varies across investigations. For instance, Docherty (1992) found prevoicing in only 7% of the cases, while Lisker and Abramson (1964) observed prevoicing in approximately 20% of English voiced stops and Westbury (1979) reported rates as high as 57% among speakers from Texas. Nevertheless, a systematic difference remains between English and the other two languages, as voicing is not an obligatory feature of the contrast in English. It is therefore treated here as exceptional in its prevoicing status. When it comes to Norwegian, the voiced stops are prevoiced, similarly to the Polish ones, whereas its voiceless stops are aspirated and resemble those in English (e.g., Halvorsen, 1998, Ringen & van Dommelen, 2013). Table 1 presents VOT values for voiced and voiceless stops in the three investigated languages, provided in the literature for word-initial stressed syllables. Figure 1 shows approximate VOT ranges of stops in the three languages under investigation based on the literature.

3.2 Research questions and hypotheses

The study strives to answer the following research questions:

RQ1. Is VOT perception in multilinguals subject to the process of CLI? In what way is it different from native speakers' perception?

H1. The first hypothesis stems from the results of previous multilingual studies on VOT perception (Cal & Wrembel, 2023; Z. Liu, 2016; J. Liu et al., 2019), which showed cross-linguistic interactions between languages in the multilinguals' repertoire. Thus, CLI effects are also predicted to be present in the current data. More specifically, it is expected that the trilinguals will yield different values for the perceptual boundary locations than native speakers of the respective languages, especially in their L2 English and L3 Norwegian.

Table 1. VOT Values (in ms) for Voiced and Voiceless Stops in Polish, English, and Norwegian Measured in Word-Initial Stressed Syllables.

	p	t	k	B	d	g
Polish (Keating et al., 1981)	21.5	27.9	52.7	−88.2	−89.9	−66.1
English (Lisker & Abramson, 1964)	58	70	80	1	5	21
Norwegian (Ringen & van Dommelen, 2013)	50	52	53	−80	−72	−73

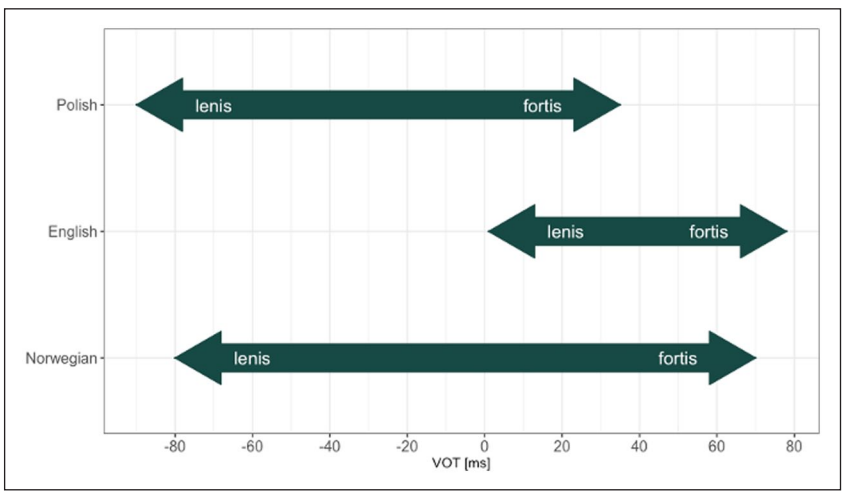


Figure 1. Comparison of approximate VOT ranges of stops in Polish, English, and Norwegian.

RQ2. What is the directionality of CLI in the VOT patterns of L1 Polish, L2 English, and L3 Norwegian speakers? Is its presence dependent on the PoA of stops?

H2. The second prediction is that all three languages will interact, but the interaction will manifest itself differently depending on a language pairing and PoA. Therefore, the following subhypotheses were proposed:

H2.1. More interactions are foreseen to take place between L2 English and L3 Norwegian, because of their typological proximity, a similar setting of acquisition (i.e., via classroom instruction), as well as the assumed foreign language effect. What is more, as aspiration is generally considered to be a more salient feature than prevoicing, and it was shown to be susceptible to CLI in the literature (e.g., Sypiańska, 2021), it is expected that L3 learners might perceive L3 Norwegian stops as overall more similar to the L2 English ones, due to aspiration being present in the two languages. At the same time, L1 Polish stops might be perceived as distinct from English and Norwegian. According to the SLM (Flege, 1995), the learners will form two separate categories for the stops—one for L1 Polish stops and the second one for L2 English and L3 Norwegian stops. An L2-to-L3 influence would result in similar boundary locations in the two languages. This scenario is visualized in Figure 2.

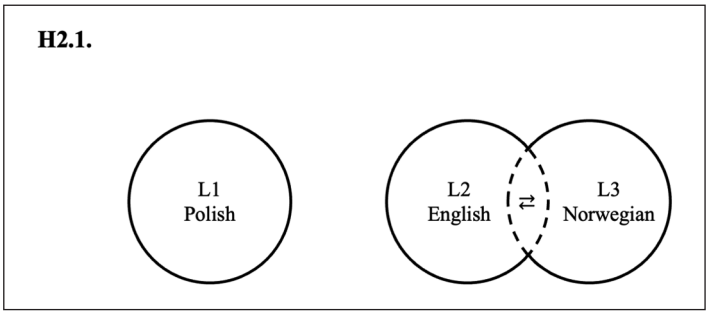


Figure 2. Visualization of subhypothesis 2.1.

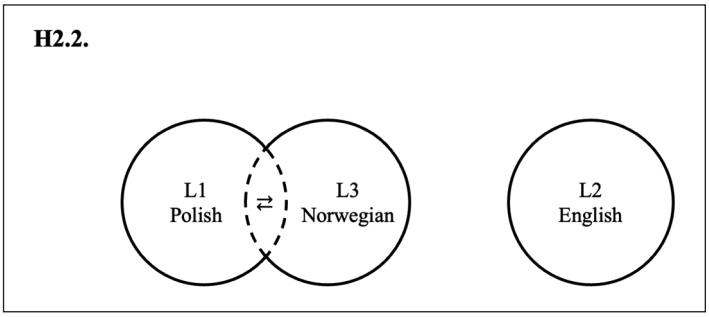


Figure 3. Visualization of subhypothesis 2.2.

H2.2. An alternative, although less plausible, scenario would be that the learners associate L3 Norwegian stops with L1 Polish ones because of the presence of prevoicing. In this case, L1 Polish and L3 Norwegian stops would be treated as equivalent, and they would form a single category; thus, as a result, they would influence one another, resulting in higher VOT boundaries in L1 Polish leading to L1 drift and lower boundaries in L3 Norwegian. On the contrary, L2 English would remain separate with the VOT boundary locations close to the target. This scenario is visualized in Figure 3.

H2.3. With regard to the presence of CLI depending on the PoA, it is hypothesized that CLI will affect all three places of articulation of stops, but to a different extent. More CLI is expected to be found in stops that have the highest VOT, as their VOT values are the most similar across languages, leading to equivalence classification. Therefore, according to the SLM (Flege, 1995), in these stops, the participants are more likely to form a single category for their three languages. It is, thus, expected that the boundaries will be more unified the further back the PoA. This subhypothesis is visualized in Figure 4.

RQ3. How do VOT boundary locations vary for individual participants?

H3. It is predicted that the patterns of VOT boundary locations across languages and places of articulation will differ for each participant (as opposed to mean values) and will reveal considerable intra- and inter-individual differences. These predictions are limited due to the lack of previous investigations into individual variation in multilingual perception—to the best of our

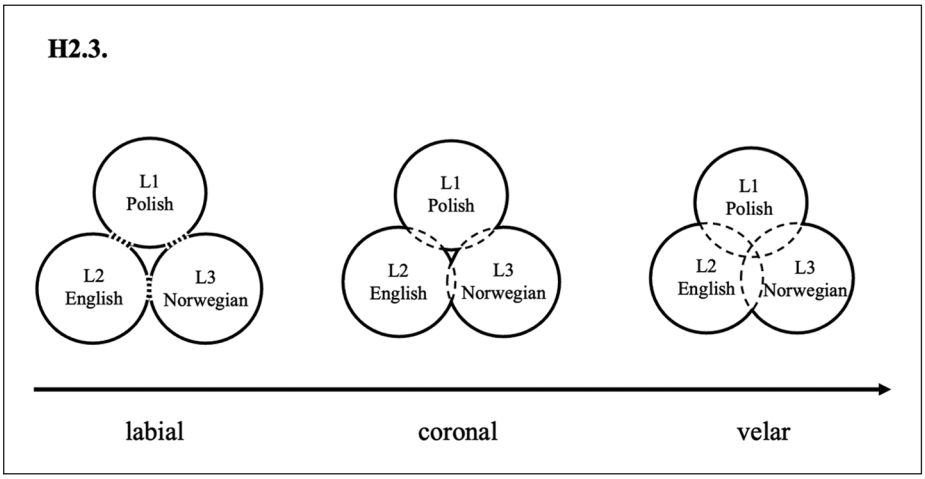


Figure 4. Visualization of subhypothesis 2.3.

knowledge, there have been very few studies so far that focus on individual variation that deal with features other than VOT (Gut et al., 2023; Wrembel et al., 2022).

3.3 Participants

Participants in this study involved an experimental group and three control groups. The former group consisted of trilingual speakers who were the main subjects of this experiment, while the latter included functionally monolingual native speakers of the three languages under investigation. All participants received gift cards for their participation in the study, and an ethics committee approval was obtained prior to the study.

3.3.1 Experimental group. The experimental group included 30 L1 Polish, L2 English, and L3 Norwegian speakers. They featured 24 females, 5 males, and 1 nonbinary person, whose mean age was 21.4 ($SD = 2.1$). The participants were L1 Polish speakers, who had been learning English as their L2. In addition, they had been learning L3 Norwegian extensively as part of the Norwegian Studies program at two Polish universities. The group was diverse in terms of proficiency levels in their L2 and L3, which were tested by means of placement tests in English and Norwegian. Their English proficiency ranged from B1 to C2, whereas Norwegian proficiency was in the range of A2 to B1+, according to the Common European Framework of Reference for Languages. There were 10 1BA students, 10 2BA students, and 10 students at the 3BA level or above. They were exposed to Norwegian extensively by taking part in practical Norwegian classes, including speaking, writing, listening, and pronunciation that were carried out as part of their study program at the university and were taught by Polish university lecturers highly proficient in Norwegian. To obtain information about their language learning history and use, the participants were asked to fill in the Language History Questionnaire (Li et al., 2020). There were 12 participants who reported previous instruction in German, 7 participants knew some French, and 4 of them had learnt Russian before. Yet, at the time of the study, they claimed not to be proficient in or use any of these languages.

Table 2. Profiles of the Investigated Groups of Participants.

Group	Investigated languages	Mean age (*SD*)	Place of residence	Number of participants
Experimental	L1 Polish, L2 English, L3 Norwegian	21.4 (2.1)	Poznań and Szczecin (Poland)	30
Control	L1 Polish	21.8 (2.2)	Poznań (Poland)	21
Control	L1 English	23.6 (3.8)	Bangor (UK)	18
Control	L1 Norwegian	25 (3)	Tromsø (Norway)	10

3.3.2 Control groups. The control groups consisted of functionally monolingual native speakers of Polish, English, and Norwegian, respectively. There were 21 L1 Polish speakers, 18 L1 English speakers, and 10 L1 Norwegian speakers.

L1 Polish controls included native speakers of Polish who, despite having learned English or other languages (such as German, French, or Spanish) at school, claimed not to have contact with any foreign languages on a daily basis. There were 19 females and 2 males residing in Poznań (Poland). The mean age of the Polish group was 21.8 (*SD*=2.2).

There were 18 L1 English controls, who included 11 females and 7 males residing in Bangor (UK), whose mean age was 23.6 (*SD*=3.8). Some of the participants reported limited proficiency in Romance languages (French, Spanish, or Italian) and Welsh but claimed not to have a regular contact with them at the time of the study.

L1 Norwegian controls consisted of 9 females and 4 males, aged 25 (*SD*=3), who resided in Tromsø (Norway). All of them were advanced (C1-C2 level) speakers of English who started learning the language approximately at the age of 6.6 (*SD*=2.7). It is important to note that, because of high English proficiency being a norm among the Norwegian society (cf. Education First, 2023), we decided to include speakers of such a linguistic profile in the experiment. A monolingual group would have been rather difficult to find and probably would not reflect the profile of the experimental group that comprised young and educated individuals. The participants also reported some knowledge of other Scandinavian languages (Swedish and Danish), as well as limited previous experience in German, Spanish, and French.

Table 2 summarizes the profiles of the investigated groups of participants.

3.4 Procedure

The study design involved the preparation of VOT continua with the use of a Praat script (Winn, 2020) based on one-syllable minimal pairs with word-initial stop sounds /p,b,t,d,k,g/. The token words included *pas-bas, tam-dam, kas-gaz* (Eng. *belt-bass, there-give* [first person singular future tense of *dać*], *cash register* [genitive plural of *kasa*]-*gas*) for Polish, *pun-bun*, *ton-done*, *come-gum* for English, and *par-bar*, *ta-da*, *kar-gard* (Eng. *pair-bare*, *take-there*, *guy-farm*) for Norwegian and were recorded in carrier phrases by native speakers of the respective languages. Nine continua were prepared—three per language (Polish, English, Norwegian) and per PoA (labial, coronal, velar) using the Praat script and the description provided by Winn (2020). The continua ranged from −100 to 100 ms and consisted of 21 steps, each of 10 ms. The script (see Winn, 2020, for a more detailed description) is designed to generate natural-sounding speech stimuli while allowing precise control over VOT and related acoustic parameters, such as F0. Continuum endpoints were based on two recorded tokens differing in the voicing of an initial stop consonant. Aspiration was

manipulated by selecting the aspirated portion from the voiceless endpoint, while prevoicing was created by extracting the onset of the vowel from the voiced token, modifying its amplitude envelope, applying low-pass filtering, and attenuating it to resemble a voicing murmur. This component was generated artificially through the filtering procedure implemented in the script. The F0 of the prevoicing segment was controlled to blend smoothly with the vowel onset; that is, the F0 contour of the prevoicing portion was adjusted to match the onset of the vowel so that there was no perceptual discontinuity at the transition. In addition, as F0 is known to be a secondary cue to voicing (e.g., Dmitrieva et al., 2015), it was automatically adjusted at each VOT step of the continua based on values extracted from the original recordings. Thus, the continuum endpoints for F0 differed across languages to reflect language-specific phonetic patterns observed in the recorded stimuli.

The obtained tokens were used to design a two-alternative forced-choice (2AFC) task in which the participants were presented with one random word from the continuum and asked whether they heard a voiced or voiceless consonant at the beginning. The decisions were made using two keys on a keyboard—A and L, for words appearing on the left and right parts of the screen, respectively. The experiment was designed in PsychoPy (Peirce et al., 2019) and carried out in three separate language blocks. Every token was presented five times in a randomized order. Each language block lasted around 15 min and was preceded by a language mode introduction, which involved watching a short video in the respective language. Before the actual task, the participants were presented with a set of instructions in the language of the experiment and took part in a practice session that consisted of five trials containing stop-initial words that differed from the ones used in the study. The experiment was conducted in a quiet room on a laptop, and the participants were instructed to wear headphones. The obtained data included identification responses for each step in the continuum, resulting in 945 data points per participant.

4 Results

This section sheds light on the results of the perceptual experiment. It begins with descriptive statistics that provide information about the VOT boundary placement across the three languages and continua in the experimental and control groups. The next section includes findings from the statistical analysis that give insights into possible cross-linguistic interactions. Finally, the section on individual variation across the participants is offered. The dataset and code used for the analysis are available online under the following link: https://osf.io/m7a43/.

4.1 Descriptive statistics

Identification responses of voiced sounds (/b/, /d/, /g/) in the three continua were used to fit a logistic curve that allowed to calculate the perceptual boundary location for VOT. The VOT boundary was placed at the 50% crossover point in the curve; that is, in the place where half of the participants' responses indicated a voiced sound and the other half the voiceless one. This procedure was done in RStudio (RStudio Team, 2019, version 2024.04.2) using the quickpsy package (Linares & López-Moliner, 2016). Figure 5 visualizes the curves fitted for identification scores for voiceless stops, together with mean VOT boundaries marked with a vertical line. The explicit values of mean VOT boundaries across continua, languages, and groups can be found in Table 3.

Taking into consideration the experimental group, the /b-p/ continuum yielded the lowest values (around 0 ms) for the perceptual boundary location. The lowest, below-zero, boundary was found in L1 Polish, whereas L3 Norwegian turned out to have the highest boundary between /b/ and /p/. In the /d-t/ continuum, the lowest value for the perceptual boundary location could be observed in

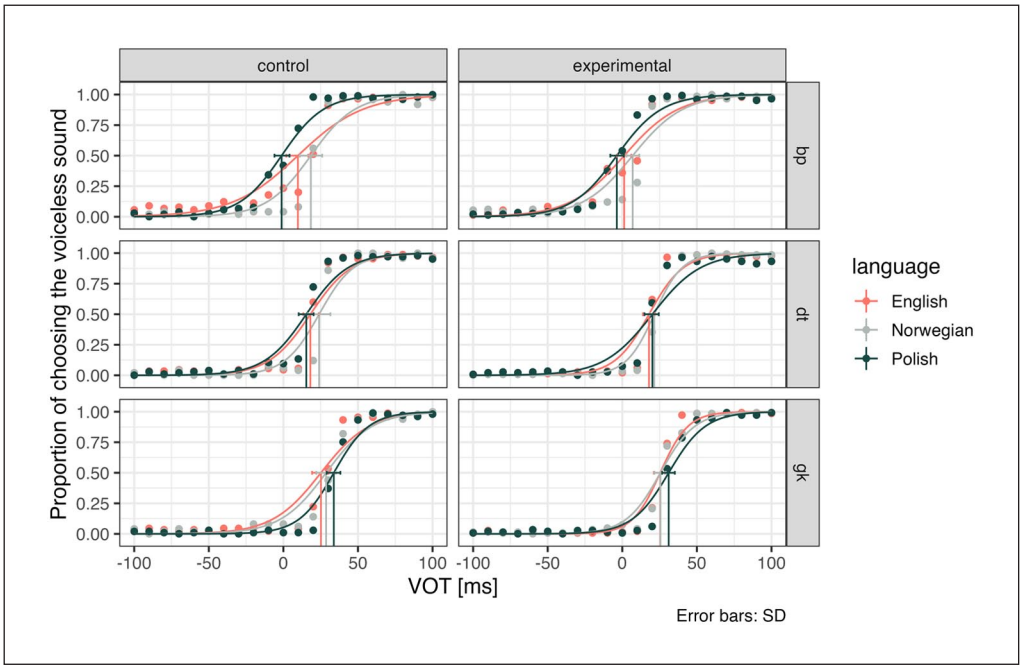


Figure 5. Logistic curves fitted for mean identification responses and VOT boundaries, separately for each language, continuum, and group.

Table 3. Mean Perceptual Boundary Locations for VOT Across Continua, Languages and Groups.

Continuum	Language	Group	*n*	*M*	*SD*
b-p	English	Experimental	30	1.125	14.604
b-p	Norwegian	Experimental	30	6.603	13.409
b-p	Polish	Experimental	30	−3.732	10.653
d-t	English	Experimental	30	17.930	5.089
d-t	Norwegian	Experimental	30	21.455	5.441
d-t	Polish	Experimental	30	20.458	10.520
g-k	English	Experimental	30	25.475	5.198
g-k	Norwegian	Experimental	30	25.605	7.311
g-k	Polish	Experimental	30	31.184	7.727
b-p	English	Control	18	9.456	14.970
b-p	Norwegian	Control	10	18.461	7.805
b-p	Polish	Control	21	−1.281	7.965
d-t	English	Control	18	18.309	4.749
d-t	Norwegian	Control	10	24.006	2.611
d-t	Polish	Control	21	15.419	7.791
g-k	English	Control	18	25.100	9.607
g-k	Norwegian	Control	10	28.729	13.023
g-k	Polish	Control	21	33.956	6.114

L2 English, whereas the highest was in L3 Norwegian. Finally, the boundary between /g/ and /k/ was the lowest for L3 Norwegian and the highest for L1 Polish.

As far as the control group is concerned, the perceptual boundary locations between voiced and voiceless stops in Polish turned out to be the lowest out of the three languages in the /b-p/ and /d-t/ continua. At the same time, the Polish controls yielded the highest, out of the three languages, boundary in the /g-k/ continuum. English values were in-between those found in Polish and Norwegian controls in /b-p/ and /d-t/ continua and the lowest for the /g-k/ continuum. Norwegian controls appeared to place the boundary between voiced and voiceless stops the highest in the /b-p/ and /d-t/ continua, whereas the boundary in the /g-k/ continuum was almost identical to the English one.

Compared with the control group, it can be seen that the boundaries in the experimental group are closer together irrespective of the continua. The biggest differences between the groups can be noticed in the Norwegian /b-p/ continuum, where the mean perceptual boundary location is considerably higher in the control groups, as well as in the Polish /d-t/ continuum, where the mean boundaries are greater in the experimental group.

4.2 Statistical modeling

To address the research questions, a Bayesian mixed-effects logistic regression model was fitted using the brms package (Bürkner, 2017) in R. The model used the default "treatment" or "dummy" coding scheme in R for the model's variables. It predicted binary response outcomes (coded as 0 indicating a voiced stop and 1 indicating a voiceless stop) using a Bernoulli likelihood with a logit link to assess the effects of VOT (scaled), language (Polish, English, and Norwegian), PoA (labial, coronal, and velar), and group (experimental vs. control), including all possible interactions among these predictors. Random effects included by-participant varying intercepts, as well as correlated slopes for VOT_scaled, PoA, and their interaction, specified as (1 + VOT_scaled * PoA | participant). All continuous predictors were standardized prior to analysis, using the scale() function in R. The model employed weakly informative priors to facilitate convergence and prevent overfitting. Specifically, the prior for the VOT_scaled coefficient was set to Normal(0, 2), other fixed effects were assigned Normal(0, 2), the intercept followed a Normal(0, 2) prior. Bayesian estimation was performed using Hamiltonian Monte Carlo with four chains, each run for 4,000 iterations (2,000 warm-up). Convergence was assessed via $\hat{R}$ statistics (all ≈ 1.00) and effective sample sizes. Pairwise comparisons were obtained using the emmeans package (Lenth, 2023). The outcome from the model can be found in Supplementary Files. Following standard Bayesian inference criteria, an effect was considered credible when the 95% highest posterior density (HPD) interval excluded zero.

4.2.1 Across-group comparisons. The first research question was whether multilinguals experience CLI in VOT perception, which should be visible in comparison with the control groups. Figure 6 depicts an estimated probability of categorizing a stop as voiceless (response = 1) across the VOT continuum in two groups (control vs. experimental), plotted separately for three languages and three places of articulation.

The multilingual group showed a tendency for more voiceless responses overall, implying a perceptual boundary shifted to a later point on the VOT continuum. This general difference compared with the control groups, however, was not credible (estimate = −0.25, 95% HPD = [−0.65, 0.16]). Credible evidence of CLI was, nonetheless, found in subsequent pairwise comparisons, where the direction and magnitude of the effect varied by language and PoA.

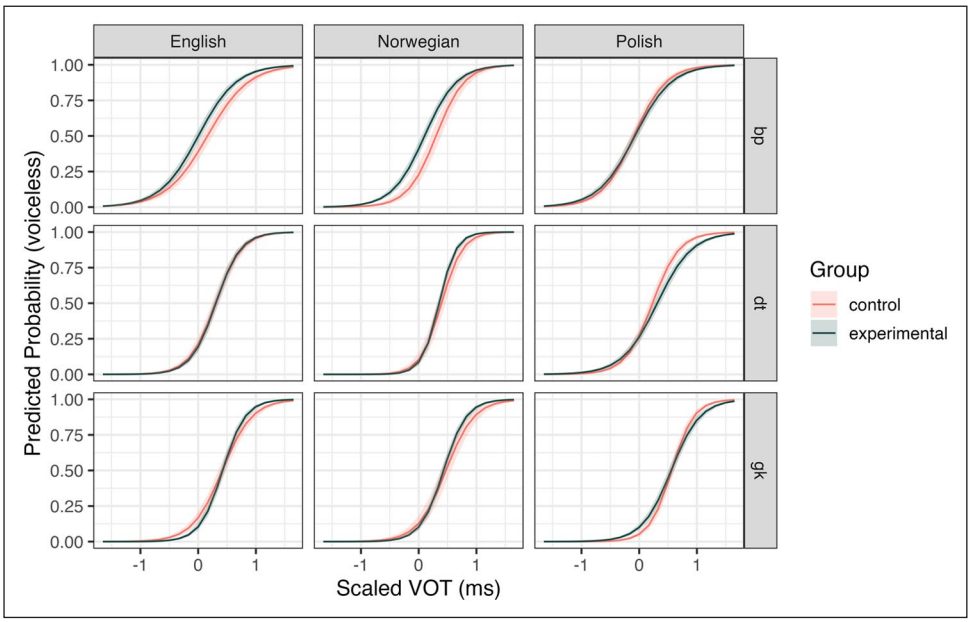


Figure 6. Estimated probability of categorizing a stop as voiceless across the VOT continuum in two groups (control vs. experimental) plotted separately for three languages and PoAs.

As far as the /b-p/ continuum is concerned, in English, the multilingual group responded with credibly higher voiced stop choices than the controls (estimate = −0.53, 95% HPD [−1.00, −0.10]), suggesting the boundary placement earlier in the continuum and toward the Polish values. A similar, even stronger, pattern can be observed in Norwegian (estimate = −1.01, 95% HPD = [−1.62, −0.36]). In Polish, however, no credible difference between the two groups was found (estimate = 0.03, 95% HPD = [−0.42, 0.51]), indicating that multilinguals approximate native Polish-like perception in this case.

In the /d-t/ and /g-k/ continua, on the contrary, neither of the contrasts across the three languages shows credible differences between groups, as their HPD intervals include 0.

4.2.2 Multilingual group. The second research question specifically concerns the performance of the multilingual group. RQ2 asked about the directionality of CLI in the linguistic repertoire of the multilinguals and the role of the PoA in VOT perception by multilingual speakers. To investigate that, the model-estimated logit-transformed probabilities of choosing a voiceless stop across three languages and three PoAs in the multilingual group were examined. A Bayesian estimated marginal means analysis revealed patterns of CLI that varied depending on both language and PoA. Figure 7 presents the predicted probability of categorizing a stop as voiceless (response = 1) across the VOT continuum separately for three languages and places of articulation.

In the /b-p/ continuum, the estimated log-odds of a voiceless response differed credibly across the three languages. The results for Polish (emmean = 0.14) show a higher probability of categorizing VOT stimuli as voiced, followed by English (emmean = −0.15) and Norwegian (emmean = −0.54). Pairwise comparisons revealed credible differences between English and Norwegian (estimate = 0.39, 95% HPD = [0.22, 0.56]), between English and Polish (estimate = −0.28, 95% HPD = [−0.45, −0.11]), as well as between Norwegian and Polish (estimate = −0.67, 95% HPD = [−0.84, −0.49]).

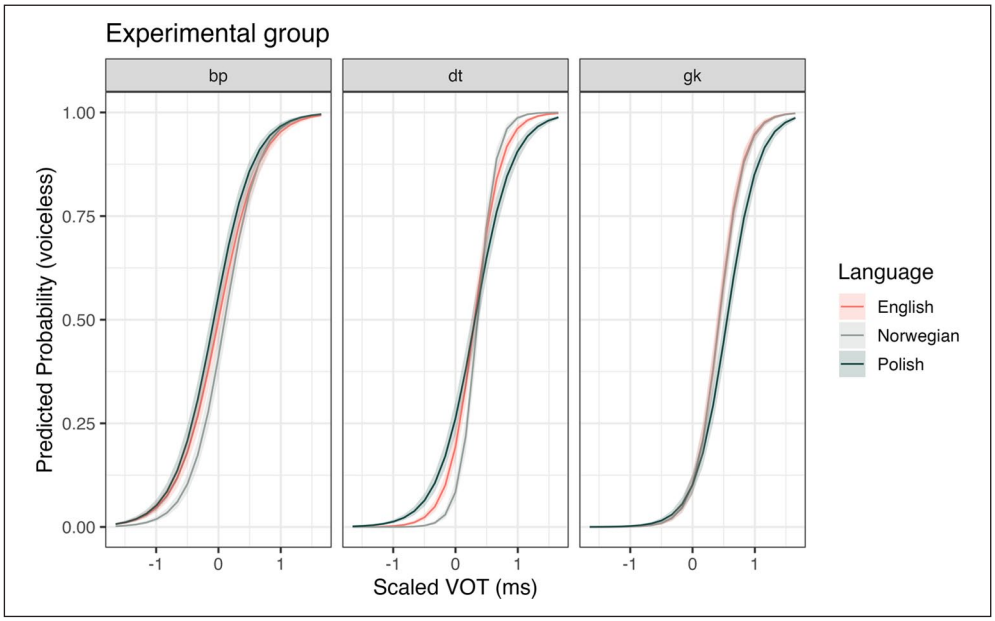


Figure 7. Predicted probability of categorizing a stop as voiceless across the VOT continuum separately for three languages and places of articulation in multilingual group.

A similar pattern was observed for the categorization of the /d-t/ continuum, although the magnitude of differences was greater. Polish, again, showed the highest probability of voiced stop categorization (emmean = −1.42), followed by English (emmean = −1.75) and Norwegian (emmean = −2.61). All pairwise contrasts were credibly different, that is, English versus Norwegian (estimate = 0.86, 95% HPD = [0.58, 1.18]), English versus Polish (estimate = −0.32, 95% HPD = [−0.53, −0.11]), and Norwegian versus Polish (estimate = −1.19, 95% HPD = [−1.48, −0.91]).

In contrast, no credible differences were found among the three languages for the /g-k/ continuum. Posterior estimates minimal differences across English (emmean = −2.88), Norwegian (emmean = −2.91), and Polish (emmean = −3.09), with all contrasts yielding HPD intervals that included zero.

4.3 Individual variation

The results presented so far have been focused mainly on means and their comparisons. Although they are quite telling, they still do not reflect individual differences that may arise for each participant, which was the focus of RQ3. Figure 8 shows VOT boundaries for every participant in the three languages and continua. Even though identifying specific patterns that would fit several participants might pose a challenge, an attempt has been made to classify the trends into some categories that would reflect possible scenarios of CLI. They include the following observations featuring participants' codes that illustrate particular scenarios:

- **Language-specific patterns of VOT categorization in the /b-p/ continuum, while the boundaries in the other two continua overlap**—the /b-p/ differs across languages by at least ~10 ms (i.e., one step of the continuum), this pattern reflects the results based on

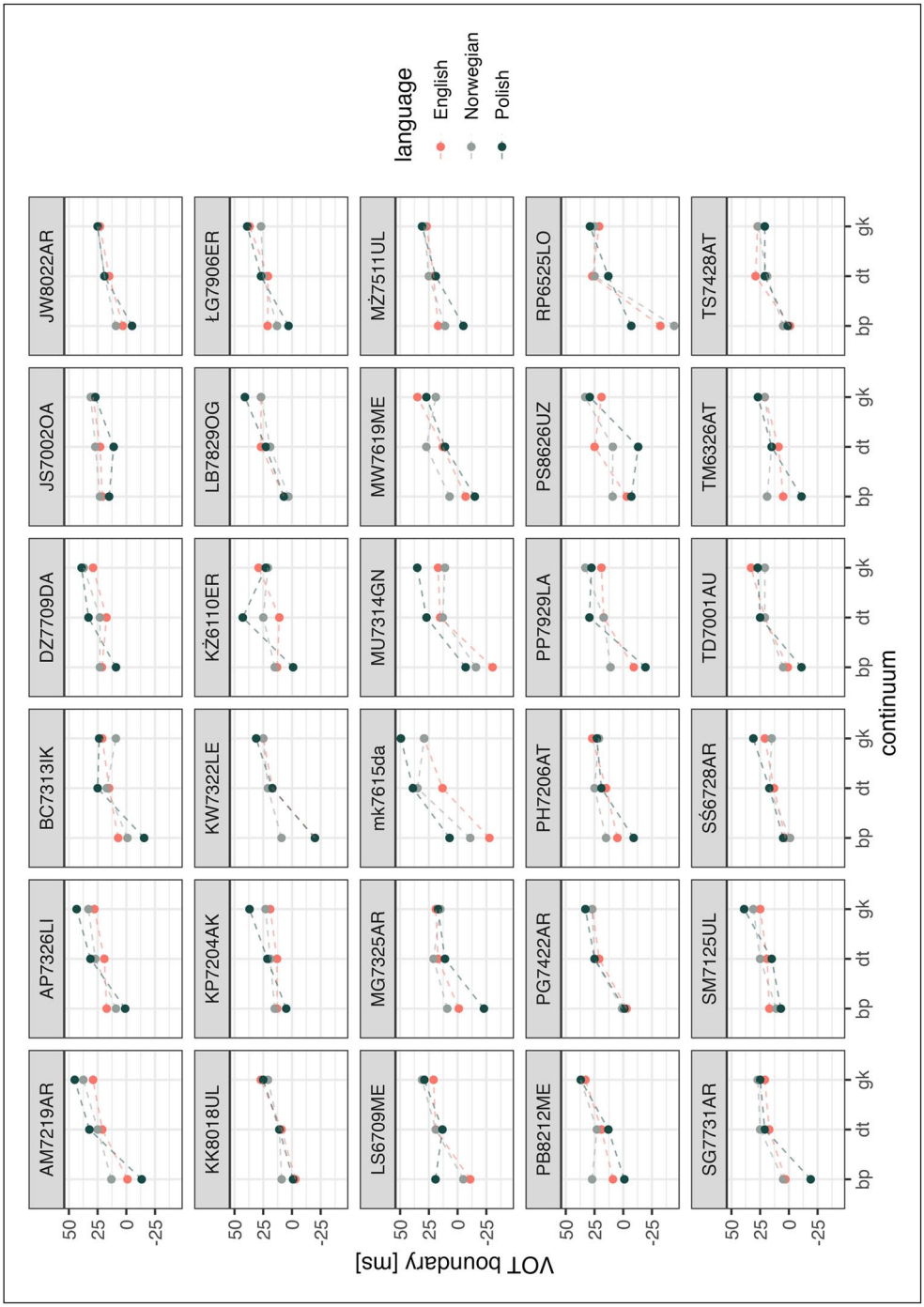


Figure 8. Individual variation plots for VOT perceptual boundary across the three languages and places of articulation.

average boundary locations: MG7325AR, MŻ7511UL PB8212ME, PH7206AT, PP7929LA, RP6525LO, and TM6326AT.

- **L1-L2-L3 overlap**—VOT boundaries in the three continua and languages are close to one another (i.e., the difference across the three languages is smaller than 10 ms): JW8022AR, KK8018UL, PG7422AR, SG7731AR, SM7125UL, and TD7001AU.
- **L1 drift**—VOT boundaries for L1 Polish exceed those in L2 English and L3 Norwegian:
  - In /b-p/ and /d-t/[1]: MK7615DA and MU7314GN.
  - Only in /b-p/: RP6525LO and SŚ6728AR.
  - Only in /d-t/: AM7219AR, PG7422AR, PP7929LA, and TD7001AU.
- **L2-L3 overlap**—VOT boundaries for L2 English and L3 Norwegian are comparable and different from L1 Polish (this pattern does not apply to all three continua):
  - In /b-p/ and /d-t/: LS6709ME and TD7001AU.
  - In /b-p/ and /g-k/: KP7204AK.
  - Only in /b-p/: KŻ6110ER and SG7731AR.
  - Only in /d-t/: BC7313IK, JS7002OA, MG7325AR, MU7314GN, PB8212ME, PP7929LA, and RP6525LO.
  - Only in /g-k/: KW7322LE, LB7829OG, MK7615DA, TM6326AT, and TS7428AT.
- **L1-L2 overlap**—VOT boundaries for L1 Polish and L2 English are comparable and different from L3 Norwegian (this pattern does not apply to all three continua):
  - Only in /b-p/: KK8018UL, KW7322LE, and PS8626UZ.
  - Only in /d-t/: MW7619ME and MŻ7511UL.
  - Only in /g-k/: BC7313IK and ŁG7906ER.
- **L1-L3 overlap**—VOT boundaries for L1 Polish and L3 Norwegian are comparable and different from L2 English, but the differences are much smaller than in the case of other patterns (this pattern does not apply to all three continua):
  - Only in /d-t/: MK7615DA and TS7428AT.
  - Only in /g-k/: DZ7709DA, LS6709ME, PP7929LA, and PS8626UZ.

5 Discussion

The aim of this study was to examine the perception of word-initial fortis and lenis stops by trilingual speakers of L1 Polish, L2 English, and L3 Norwegian across their three languages. The first research question was whether VOT perception in multilinguals is subject to CLI and in what way it might differ from native speaker perception. It was expected to find cross-linguistic interactions between languages in the multilinguals' repertoire and, especially, different values for the perceptual boundary locations in the multilingual experimental group versus control native speakers, particularly in their L2 English and L3 Norwegian.

Statistical testing revealed credible differences between the two groups, suggesting that multilinguals perceive VOT differently than the controls, which might possibly be due to CLI in trilingual participants. This tendency is also evident in the descriptive statistics, where it appears that the multilingual group might have placed their perceptual boundaries across languages closer together than the control groups. Pairwise comparisons showed that group differences in the perception of the VOT continua varied depending on the language and PoA, revealing context-specific instances of CLI rather than a uniform transfer. That is, differences between the two groups were attested only in the /b-p/ and /g-k/ continua, while the identification of the /d-t/ continuum by the multilingual group did not differ credibly from the controls.

In the /b-p/ continuum, the multilinguals maintained Polish-like identification patterns in their three languages, which suggests considerable L1 Polish influence on the remaining languages.

A possible explanation for this is that it might be challenging for the participants to notice the differences between languages in the /b-p/ continuum. In the process of equivalence classification proposed in the SLM (Flege, 1995), these sounds might be perceived by the multilinguals as similar, leading them to form a single perceptual category for the bilabial stops in all three languages, and, as a result, they experience CLI in this continuum stemming from L1 Polish.

On the contrary, the lack of differences across the two groups in the /d-t/ continuum and, therefore, native-like categorization patterns of stops in the three languages, may result from greater perceptual salience. Coronal stops differ notably across the three languages: they are dental in Polish (Dukiewicz & Sawicka, 1995), alveolar in English (Roach, 2009), and highly variable in Norwegian (Kristoffersen, 2000). This articulatory variation may make the across-language contrasts more perceptible to the participants, thereby leading to the formation of separate phonetic categories for the stops in the three languages.

The results for the /g-k/ continuum also point to the CLI effect. However, in this case, participants appeared to rely more on longer VOT in the stop identification, possibly due to the relatively long VOT values of velar fortis stops in all three languages. No credible group differences were observed between the multilinguals' identification patterns and the native norms in all three languages, which might suggest that the multilingual group maintained native-like categorization patterns in this continuum. However, taking into account relatively small differences between category boundaries in the controls' identification patterns and rather long VOT in velar fortis stops in the three languages, it might also be possible that the participants, in fact, perceive velar stops the same way in all three languages. Exposure to languages with long-lag VOT in velar stops may have reinforced participants' sensitivity to longer VOT durations, causing a shift in their perceptual boundaries toward the voiceless end of the continuum. Therefore, there is a possibility that the multilinguals possess a merged perceptual category of velar stops for all three languages. Across-language comparisons might provide more insight into this category.

The second research question focused on the directionality of CLI across the languages of L1 Polish-L2 English-L3 Norwegian speakers and how its presence is modulated by different places of articulation of stops. It was hypothesized that all three languages would interact, but to a different extent, depending on several factors such as typological proximity, settings of acquisition, foreign language effect, or cross-linguistic similarity across the three languages. Although the descriptive statistics showed relatively small differences between languages in all three places of articulation, the statistical analysis revealed interesting patterns with credible differences found in the /b-p/ and /d-t/ continua.

First, even though credible differences between the three languages have been attested in the identification patterns of the /b-p/ continuum, the magnitude of differences was rather small, and, in addition, the boundary locations turned out to be placed early in the continuum. This finding, together with group differences in English and Norwegian described above, suggests a heavy reliance on L1 Polish when it comes to bilabial stop perception. A possible explanation is that the multilinguals might not have noticed the difference between /b-p/ continua in the three languages just yet and, therefore, perceive them as similar in line with the equivalence classification (Flege, 1995). As the boundaries in the three languages are around 0 ms, it can be concluded that there is no L1 drift, but rather L2 English and L3 Norwegian are influenced by L1 Polish, whose boundary is considered to be lower in the continuum. This finding, although rather unexpected in our study, aligns with the main predictions of the L2 speech models, which assume a general reliance on L1, especially at the initial stages of L2 acquisition, as well as with previous findings regarding bilabial stop perception by multilinguals (J. Liu et al., 2019; Stan, 2024; Zeng, 2024).

In the /d-t/ continuum, the across-language differences in the multilingual group turned out to be larger than for the bilabial continuum, and, as mentioned above, there were no

credible differences between control and multilingual groups. Because of that it can be stated that multilinguals managed to maintain separate language systems in coronal stops due to possible greater perceptual salience stemming from the variability in the PoA of these stops across the three languages under investigation.

With regard to the /g-k/ continuum, however, we did not observe any credible differences across the languages. This outcome suggests the previously mentioned possibility of maintaining one merged category for velar stops in the three languages. However, unlike in the predictions, it was L1 Polish that was possibly the most affected by this process, compared with L2 English and L3 Norwegian, as voiceless stop identification was the highest for multilinguals in this language. Nevertheless, VOT in the Polish voiceless velar stop is rather long compared with the other places of articulation, and adding L3 Norwegian, which is yet another language with a long-lag VOT, to the system might have resulted in shifting the perceptual boundary toward the voiceless part of the continuum, possibly leading to the L1 drift effect. L2 English also seemed to be affected by CLI, which may be attributed to the closer typological relationship between English and Norwegian compared with Polish, as the former two languages come from the same family of Germanic languages.

Recapitulating, our results partially confirm the first hypothesis, that is, they show that trilingual speakers do experience CLI in their perception of stop continua, but its presence and directionality depend on the PoA of a stop. The second hypothesis was also confirmed only partially. Overall, the general prediction that all languages will interact can be corroborated; however, we did not find an unequivocal confirmation of any of the predicted scenarios. We found evidence of CLI in both directions, namely L1-induced CLI emerged in the perception of bilabial stops, which is consistent with previous findings (J. Liu et al., 2019; Stan, 2024; Zeng, 2024) and thereby confirms the H2.2 scenario. However, in the case of this study, both foreign languages were affected by CLI, not only L3, which might point to the lower stability of the L2 English system in multilinguals. At the same time, L3-induced CLI was observed in velar stops. This outcome, on the contrary, confirms the H2.1 scenario. Yet, in this case, again, both L1 Polish and L2 English were affected by CLI, suggesting possible L1 drift effect caused by an extensive L2 and L3 exposure. In contrast, the perception of coronal stops was not affected and the participants maintained separate language systems. The H2.3 scenario was, thus, also partially confirmed, as each PoA was affected by CLI in a different manner. However, the magnitude of this effect differed from the predictions, and we did not find the expected gradual increase in the amount of CLI with respect to the further back the PoA. It can be, thus, concluded that the patterns of CLI vary across the three places of articulation and seem to be conditioned by cross-linguistic similarity, perceptual salience, and typological relations between languages, thus indirectly testifying to the complex nature of phonological acquisition in multilinguals that calls for more nuanced predictions and interpretations.

The third research question addressed individual variation among participants in terms of VOT perceptual boundary placement. The general prediction was to expect considerable intra- and inter-individual differences across the participants. The obtained data pointed to the foreseen variability among the multilinguals and revealed some interesting patterns of boundary placement across the languages and continua, which seems to be consistent with the findings reported by Wrembel et al. (2022) and Gut et al. (2023). This observation could be attributed to individual perceptual differences that the participants exhibit in the group, which may differ across individuals. The factors that might play a role in this study concern especially some variation in the level of L2 and L3 proficiency, but also the actual frequency of use of particular languages as well as language aptitude or auditory sensitivity. It seems important to further scrutinize these factors in future studies to better understand the observed variability and the role that individual differences play in perceptual boundary location. In the current contribution, we identified potential individual patterns, in

line with previous rare studies (e.g., Wrembel et al. 2022), in the attempt to integrate this approach more firmly into L3 acquisition research as advocated by some researchers (e.g., Gut et al. 2023).

Despite the valuable insights provided by this study, it is not without limitations. The study would benefit from a larger sample size, both in terms of the multilingual group and the controls. Moreover, the stimuli design needs to avoid a potential influence of secondary voicing cues on the perception of VOT across the three languages under investigation. Although the present study addressed this issue by adjusting F0 values, further attention to this aspect would likely enhance the validity and interpretability of future findings. Further investigations should focus on the acquisition of stops in different places of articulation perceived by speakers of various linguistic repertoires to corroborate the findings provided herein. What is more, it could also account for the variability in participants' L2 and L3 proficiency levels, especially when it comes to the individual learners' outcomes. Other individual differences could also be taken into consideration as well as the development of perception trajectory over time.

6 Conclusion

This study aimed to explore how trilingual speakers of L1 Polish, L2 English, and L3 Norwegian perceive word-initial fortis and lenis stops in their three languages. In particular, the goal was to establish whether the obtained perceptual boundary locations between voiced and voiceless plosives for the trilingual speakers are subject to CLI, and what the directionality of this process is. The novelty of the study is twofold; it provided a deeper insight into the field of multilingual perception by presenting the analysis of all three places of articulation of stops in all three languages of the speakers together with the native controls of the respective languages. In addition, the study offered an overview of individual variation, which is still a rare practice in the field.

All in all, the results demonstrated that L1 Polish-L2 English-L3 Norwegian speakers experience CLI, but its nature was rather complex, and the directionality and magnitude depended on the PoA of stops. That is, it was found that in bilabial stops, the participants undergo an L1-induced CLI; in coronal stops, they maintain separate perceptual systems; whereas in velar stops, they demonstrate an L3-driven merged perceptual category. The analysis of individual variation revealed various patterns of VOT categorization, confirming the complex nature of CLI, at the same time highlighting the need for future research.

Finally, the study has the potential to further inform the existing models of foreign language speech acquisition by expanding the discussion to the third language acquisition perspective and providing new evidence for a more nuanced formulation of predictions related to the nature and directionality of CLI in multilingual speakers.

Funding

The author disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the research project OPUS-19-HS 2020/37/B/HS2/00617 funded by the National Science Centre.

Ethical considerations

This study was approved by the Ethics Committee for Research Involving Human Participants at Adam Mickiewicz University in Poznań (resolution no. 12/2020/2021) on July 05, 2021.

Consent to participate

All participants provided written informed consent prior to participating.

ORCID iD

Zuzanna Cal https://orcid.org/0000-0002-1053-8440

Data availability

The data sets generated and analyzed during the current study are available in the OSF repository under the following link: https://osf.io/m7a43/.

Supplemental material

Supplemental material for this article is available online.

Note

1. The /g-k/ continuum was not taken into account here, because the values obtained from the Polish control group for /g-k/ already exceeded those for English and Norwegian.

References

Aliaga-García, C., & Mora, J. C. (2009). Assessing the effects of phonetic training on L2 sound perception and production. In M. A. Watkins, A. S. Rauber, & B. O. Baptista (Eds.), *Recent research in second language phonetics/phonology* (pp. 2–31). Cambridge Scholars Publishing.

Alves, U. K., & Zimmer, M. C. (2015). Perception and production of English VOT patterns by Brazilian learners: The role of multiple acoustic cues in a DST perspective. *Alfa: Revista de Linguística*, *59*(1), 157–180. https://doi.org/10.1590/1981-5794-1502-7

Antoniou, M., Liang, E., Ettlinger, M., & Wong, P. (2015). The bilingual advantage in phonetic learning. *Bilingualism: Language and Cognition*, *18*, 683–695. https://doi.org/10.1017/S1366728914000777

Antoniou, M., Tyler, M. D., & Best, C. T. (2012). Two ways to listen: Do L2-dominant bilinguals perceive stop voicing according to language mode? *Journal of Phonetics*, *40*(4), 582–594.

Balas, A. (2018). *Non-native vowel perception: The interplay of categories and features*. Adam Mickiewicz University Press.

Balas, A., Kopečková, R., & Wrembel, M. (2019). Perception of rhotics by multilingual children. In S. Calhoun, P. Escudero, M. Tabain, & P. Warren (Eds.), *Proceedings of the 19th International Congress of Phonetic Sciences (ICPHS), Melbourne, Australia 2019* (pp. 3725–3729). International Phonetic Association.

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In M. Munro & O.-S. Bohn (Eds.), *Second language speech learning* (pp. 13–34). John Benjamins.

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

Cabreli Amaro, J. (2017). Testing the phonological permeability hypothesis: L3 phonological effects on L1 versus L2 systems. *International Journal of Bilingualism*, *21*(6), 698–717. https://doi.org/10.1177/1367006916637287

Cal, Z., & Wrembel, M. (2023). Perception of voicing distinction in syllable-initial stops by multilingual speakers. In R. Skarnitzl & J. Volín (Eds.), *Proceedings of the 20th International Congress of Phonetic Sciences–ICPhS 2023* (pp. 451–455). International Phonetic Association.

Dmitrieva, O., Llanos, F., Shultz, A. A., & Francis, A. L. (2015). Phonological status, not voice onset time, determines the acoustic realization of onset f0 as a secondary voicing cue in Spanish and English. *Journal of Phonetics*, *49*, 77–95. https://doi.org/10.1016/j.wocn.2014.12.002

Docherty, G. J. (1992). *The timing of voicing in British English obstruents*. De Gruyter Mouton.

Dukiewicz, L., & Sawicka, I. (1995). *Gramatyka współczesnego języka polskiego: Fonetyka i fonologia* [Grammar of the contemporary Polish language: Phonetics and phonology] (H. Wróbel, Ed.). Wydawnictwo Instytutu Języka Polskiego PAN.

Education First. (2023). *EF English proficiency index 2023*. https://www.ef.com/assetscdn/WIBIwq6RdJvcD9bc8RMd/cefcom-epi-site/reports/2023/ef-epi-2023-english.pdf

Enomoto, K. (1994). L2 perceptual acquisition: The effect of multilingual linguistic experience on the perception of a "less novel" contrast. *Edinburgh Working Papers in Applied Linguistics*, *5*, 15–29.

Escudero, P. (2005). *Linguistic perception and second language acquisition: Explaining the attainment of optimal phonological categorization* [Doctoral dissertation, Utrecht University]. LOT Dissertation Series.

Flege, J. E. (1995). Second-language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issue in cross-language research* (pp. 229–273). York Press .

Flege, J. E., & Bohn, O.-S. (2021). The revised Speech Learning Model (SLM-r). In R. Wayland (Ed.), *Second language speech learning: Theoretical and empirical progress* (pp. 3–83). Cambridge University Press. https://doi.org/10.1017/9781108886901.002

Flege, J. E., & Eefting, W. (1987). The production and perception of English stops by Spanish speakers of English. *Journal of Phonetics*, *15*(1), 67–83. https://doi.org/10.1016/S0095-4470(19)30538-8.

García-Sierra, A., Schifano, E., Duncan, G. M., & Fish, M. S. (2021). An analysis of the perception of stop consonants in bilinguals and monolinguals in different phonetic contexts: A range-based language cueing approach. *Attention, Perception, & Psychophysics*, *83*, 1878–1896. https://doi.org/10.3758/s13414-020-02183-z

Gorba, C. (2019). Bidirectional influence on L1 Spanish and L2 English stop perception: The role of L2 experience. *The Journal of the Acoustical Society of America*, *145*(6), EL587–EL592. https://doi.org/10.1121/1.5113808

Gorba, C., & Cebrian, J. (2021). The role of L2 experience in L1 and L2 perception and production of voiceless stops by English learners of Spanish. *Journal of Phonetics*, *88*, Article 101094. https://doi.org/10.1016/j.wocn.2021.101094

Gut, U. (2010). Cross-linguistic influence in L3 phonological acquisition. *International Journal of Multilingualism*, *7*(1), 19–38. https://doi.org/10.1080/14790710902972248

Gut, U., Kopečková, R., & Nelson, C. (2023). *Phonetics and phonology in multilingual language development*. Cambridge University Press. https://doi.org/10.1017/9781108992527

Halvorsen, B. (1998). *Timing relations in Norwegian stops* [Unpublished doctoral dissertation]. University of Bergen.

Jassem, W. (2003). Polish. *Journal of the International Phonetic Association*, *33*(1), 103–107. https://doi.org/10.1017/S0025100303001191

Keating, P., Mikos, M., & Ganong, W. F. III. (1981). A cross-language study of range of VOT in the perception of stop consonant voicing. *The Journal of the Acoustical Society of America*, *70*(5), 1260–1271.

Kellogg, J., & Chang, C. B. (2023). Exploring the onset of phonetic drift in voice onset time perception. *Languages*, *8*(1), Article 78. https://doi.org/10.3390/languages8010078

Kopečková, R. (2015). Differences in the perception of English vowel sounds by child L2 and L3 learners. In U. Gut, R. Fuchs, & E.-M. Wunder (Eds.), *Universal or diverse paths to English phonology* (pp. 71–90). De Gruyter Mouton. https://doi.org/10.1515/9783110346084-005

Kristoffersen, G. (2000). *The phonology of Norwegian*. Oxford University Press.

Lenth, R. V. (2023). *emmeans: Estimated marginal means, aka least-squares means* (Version 1.8.8) [R package]. https://CRAN.R-project.org/package=emmeans

Li, P., Zhang, F., Yu, A., & Zhao, X. (2020). Language History Questionnaire (LHQ3): An enhanced tool for assessing multilingual experience. *Bilingualism: Language and Cognition*, *23*, 938–944. https://doi.org/10.1017/S1366728918001153

Linares, D., & López-Moliner, J. (2016). Quickpsy: An R package to fit psychometric functions for multiple groups. *The R Journal*, *8*(1), 122–131. https://doi.org/10.32614/RJ-2016-008

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops. *Word*, *20*(3), 384–422.

Liu, J., & Lin, J. (2021). A cross-linguistic study of L3 phonological acquisition of stop contrasts. *SAGE Open*, *11*(1). https://doi.org/10.1177/2158244020985510

Liu, J., Zeng, T., & Lu, X. (2019). Challenges in multi-language pronunciation teaching: A cross-linguistic study of Chinese students' perception of voiced and voiceless stops. In Y. Zheng & X. Gao (Eds.), *Círculo de Lingüística Aplicada a la Comunicación* (Vol. 79, pp. 99–118). https://doi.org/10.5209/clac.65652

Liu, Z. (2016). *Exploring cross-linguistic influence: Perception and production of L1, L2 and L3 bilabial stops by Mandarin Chinese speakers* [Master's thesis, Universitat Autònoma de Barcelona]. Dipòsit Digital de Documents de la UAB. https://ddd.uab.cat/record/167902

McKelvie-Sebileau, P., & Davis, C. (2014). Discrimination of foreign language speech contrasts by English monolinguals and French/English bilinguals. *The Journal of the Acoustical Society of America*, *135*(5), 3025–3035. https://doi.org/10.1121/1.4870701

Montanari, S., Steffman, J., Mayr, R., Flores, P., & Feeney, C. (2023). Differential perception of vowels and stop consonants in Spanish-English bilingual preschoolers. In R. Skarnitzl & J. Volín (Eds.), *Proceedings of the 20th International Congress of Phonetic Sciences* (pp. 42–46). Article 86 Guarant International. https://guarant.cz/icphs2023/86.pdf

Pan, L., Ke, H., & Styles, S. J. (2022). Early linguistic experience shapes bilingual adults' hearing for phonemes in both languages. *Scientific Reports*, *12*(1), 4703. https://doi.org/10.1038/s41598-022-08557-7

Patihis, L., Oh, J. S., & Mogilner, T. (2013). Phoneme discrimination of an unrelated language: Evidence for a narrow transfer but not a broad-based bilingual advantage. *International Journal of Bilingualism*, *19*(1), 3–16. https://doi.org/10.1177/1367006913476768

Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Ringen, C., & van Dommelen, W. A. (2013). Quantity and laryngeal contrasts in Norwegian. *Journal of Phonetics*, *41*(6), 479–490. https://doi.org/10.1016/j.wocn.2013.09.001

Roach, P. (2009). *English phonetics and phonology: A practical course* (4th ed.). Cambridge University Press.

Rojczyk, A. (2010). *Temporal and spectral parameters in perception of the voicing contrast in English and Polish*. Wydawnictwo Uniwersytetu Śląskiego.

RStudio Team. (2019). *RStudio: Integrated development for R* [Computer software]. RStudio, Inc. http://www.rstudio.com/

Stan, I. (2024). A cross-linguistic study of L3 speech perception of voiced and voiceless stops. *Journal of Experimental Phonetics*, *33*, 137–152. https://doi.org/10.1344/efe-2024-33-137-152

Stölten, K. (2013). *The effects of age of onset on VOT in L2 acquisition and L1 attrition: A study of the speech production and perception of advanced Spanish-Swedish bilinguals* [Doctoral dissertation]. Stockholm University. https://www.diva-portal.org/smash/get/diva2:657142/fulltext01.pdf

Sypiańska, J. (2021). Production of voice onset time (VOT) by senior Polish learners of English. *Open Linguistics*, *7*, 316–330. https://doi.org/10.1515/opli-2021-0016

Sypiańska, J., & Cal, Z. (2022). Perceptual drift in L1 phonetic categories in multilinguals. In A. Jarosz & A. A. Sardegna (Eds.), *Theoretical and practical developments in English speech assessment, research, and training* (pp. 299–313). Springer. https://doi.org/10.1007/978-3-030-98218-8_17

Sypiańska, J., & Cal, Z. (2023). Perception-production mismatch in L1 drift. In R. Skarnitzl & J. Volín (Eds.), *Proceedings of the 20th International Congress of Phonetic Sciences–ICPhS 2023* (pp. 2855–2859). International Phonetic Association.

Tremblay, M.-C., & Sabourin, L. (2012). Comparing behavioral discrimination and learning abilities in monolinguals, bilinguals and multilinguals. *Journal of the Acoustical Society of America*, *132*, 3465–3474. https://doi.org/10.1121/1.4756955

Westbury, J. R. (1979). *Aspects of the temporal control of voicing in consonant clusters in English* [Doctoral dissertation, University of Texas at Austin].

Williams, L. S. (1977). The perception of stop consonant voicing by Spanish-English bilinguals. *Perception & Psychophysics*, *21*, 289–297. https://doi.org/10.3758/BF03199477.

Winn, M. B. (2020). Manipulation of voice onset time in speech stimuli: A tutorial and flexible Praat script. *The Journal of the Acoustical Society of America*, *147*(2), 852. https://doi.org/10.1121/10.0000692

Wrembel, M., Gut, U., Kopečková, R., & Balas, A. (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*, *21*(1), 92–111. https://doi.org/10.1080/14790718.2022.2036158

Wrembel, M., Marecka, M., & Kopečková, R. (2019). Extending perceptual assimilation model to L3 phonological acquisition. *International Journal of Multilingualism*, *16*(4), 513–533. https://doi.org/10.1080/14790718.2019.1583233

Zeng, M. (2024, May 17–19). *Mandarin trilinguals' perception of L3 Japanese stops* [Poster presentation]. L3 Workshop: Multilingual Language Acquisition, Processing and Use, Tokyo, Japan.

Appendix I: Publication 3 (Cal, 2026b)

Cal, Zuzanna. 2026. “Is L2 the least stable system? The dynamics of VOT trilingual production in Polish-English-Norwegian read and spontaneous speech”, *Languages 11*(9): 1–28. https://doi.org/10.3390/languages11090182.

109:1784007404

languages

MDPI

Article

Is L2 the Least Stable System? The Dynamics of VOT Trilingual Production in Polish–English–Norwegian Read and Spontaneous Speech

Zuzanna Cal

Department of Contemporary English and Multilingualism, School of Languages and Literatures, Faculty of English, Adam Mickiewicz University, 60-780 Poznań, Poland; zuzanna.cal@amu.edu.pl

Abstract

This study analyses the production of voiceless and voiced plosives (/p, t, k, b, d, g/) in both read and spontaneous speech produced by L1 Polish, L2 English, L3 Norwegian multilingual learners and native control speakers. The goal is to examine whether L2/L3 learners maintain distinct phonological categories in all their languages or exhibit cross-linguistic influence, as well as to identify which voicing series and places of articulation are most susceptible to the interference. Moreover, the study seeks to integrate acoustic data with extralinguistic factors, such as proficiency, exposure, frequency of use and age of acquisition, to provide a deeper understanding of the dynamic coexistence of three phonological systems both in controlled and more authentic environments. The study reveals significant L1 and L3 influence on L2 stop production, especially in the voiced series of stops, and more target-like L1 and L3 production.

Keywords: voice onset time; cross-linguistic influence; multilingualism; read speech; spontaneous speech; speech modes

Check for updates

Academic Editor: Elena Babatsouli

Received: 26 April 2026
Revised: 20 August 2026
Accepted: 24 August 2026
Published: 31 August 2026

Copyright: © 2026 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

The acquisition of stop consonants by multilingual learners has been the focus of increasing attention in recent L3 speech acquisition research (e.g., Sypiańska, 2013; Wrembel, 2015; Llama & López-Morelos, 2016; Gabriel et al., 2018; Amengual, 2021; Geiss et al., 2022; Nelson, 2022; Zhang et al., 2023; Cal & Wrembel, 2025). Stop production, quantified primarily using voice onset time (VOT), provides a robust acoustic metric for comparing interaction across multiple phonetic systems of a speaker. While L3 phonological research has predominantly examined the voiceless series of stops (e.g., Sypiańska, 2013; Wrembel, 2015; Llama & López-Morelos, 2016; Amengual, 2021), more recent studies have extended this inquiry to the voiced counterparts (e.g., Gabriel et al., 2018; Geiss et al., 2022). Nonetheless, the area remains rather underexplored, as the current literature seems to offer several different perspectives regarding how a multilingual repertoire interacts. There is an ongoing debate concerning the source and directionality of cross-linguistic influence (CLI) during L3 acquisition. The extant literature offers conflicting evidence that points alternatively to L1 dominance (e.g., Bandeira & Zimmer, 2012; Llama & Cardoso, 2018), L2 dominance (e.g., Llama et al., 2010; Tremblay, 2007), or a combined effect of both prior languages (e.g., Sypiańska, 2017; Wrembel, 2014; Zhang, 2022). Furthermore, while some holistic studies document progressive and regressive CLI across the three languages (e.g., Liu, 2016; Sypiańska, 2017), others find evidence of separate phonetic systems (e.g.,

Amengual, 2021; Geiss et al., 2022). Consequently, further empirical work is required to clarify these contradictory findings, particularly from a perspective that accounts for both voicing series and all languages within the speaker's repertoire.

Adding to this complexity, VOT and CLI patterns are also realised differently depending on speech styles and task types, with spontaneous speech data yielding considerably more variability than read utterances (e.g., Baran et al., 1977; Yao, 2009). However, phonetic CLI has rarely been examined simultaneously across the contrasting speech modalities. The present contribution, thus, aims to address this gap and lay the foundations for further research by examining the trilingual repertoire of L1 Polish, L2 English and L3 Norwegian. These languages were chosen due to their complementary VOT patterns, allowing for a systematic analysis of CLI and language acquisition patterns. As the third instalment in a research series exploring this specific trilingual repertoire, the current study builds directly upon the previous work that includes a longitudinal investigation of VOT production (Cal & Wrembel, 2025) and a subsequent study on VOT perception (Cal, 2026). For this reason, this paper will additionally evaluate whether the longitudinal acquisition and CLI patterns observed in the first study hold robustly across a new cohort and different experimental conditions.

The paper is structured in the following manner. Firstly, the research on VOT acquisition and cross-linguistic influence is described, specifically delving into elicitation tasks and speech modes used across studies, together with their effect on VOT variability and CLI patterns. Secondly, attention is drawn to the existing models of third language acquisition with a particular focus on the Natural Growth Theory of Acquisition (NGTA; Dziubalska-Kołaczyk & Wrembel, 2022, 2024) and the rationale behind embedding this work in the NGTA framework. The next part of the paper is devoted to the methods used in the study; that is, the aim, research questions and hypotheses, followed by an in-depth description of the participant pool and procedure. This is succeeded by the discussion of the results and conclusions.

2. CLI in VOT Production in Multilinguals

The body of research on VOT production in multilinguals has been growing quite extensively in the past two decades. Undoubtedly, VOT is an easily measurable and ubiquitous parameter that offers reliable and relatively fast across-language analyses. Therefore, it has been widely chosen to provide understanding of how CLI functions in multilingual speakers. While bilingual research offers comprehensive insights into phonetic interactions between L1 and L2, the extant multilingual studies usually solely focus on the L3 acquisition (e.g., Bandeira & Zimmer, 2012; Llama & Cardoso, 2018; Llama et al., 2010; Tremblay, 2007; Wrembel, 2011, 2014, 2015); however, there also exists an increasing amount of work on regressive CLI (e.g., Sypiańska, 2013; Nelson, 2022), as well as some research dealing with all three languages simultaneously (e.g., Cal & Wrembel, 2025; Geiss et al., 2022; Amengual, 2021; Liu, 2016). Moreover, the primary focus of the research to date has been voiceless stops, leaving the domain of the voiced counterparts and the comparison thereof rather underexplored. Yet, adopting such a holistic perspective is highly warranted, as previous research has pointed to an interesting asymmetry in the acquisition of the two voicing series, both in bilingual (e.g., Zając, 2015; Schwartz et al., 2020; Sypiańska, 2021; Wojtkowiak, 2022) and multilingual speakers (e.g., Cal & Wrembel, 2025, Zhang et al., 2023; Nelson, 2022).

Existing studies on VOT acquisition are frequently based on the Speech Learning Model (SLM; Flege, 1995) and the revised Speech Learning Model (SLM-r; Flege & Bohn, 2021) to explain cross-linguistic interactions. Grounded in the premise that L2 learners experience phonetic influence from their native language, this framework also accounts for

 https://doi.org/10.3390/languages11090182

regressive CLI. The model proposes that L2 sound learning is dependent on how learners perceive the distance between L2 and L1 sounds. That is, distinct L2 sounds are more easily acquired because they do not fall within the equivalence classification. Conversely, similar sounds pose a greater challenge, as they trigger equivalence classification with established L1 phonetic categories, resulting in a merged category that leads to bidirectional cross-linguistic interference. A vast body of research, within the framework of second language acquisition, has used these claims to investigate VOT production in the two languages. As far as the population of L1 Polish-L2 English bilinguals is concerned, previous research shows an asymmetry between voiced and voiceless series of stops (e.g., Zając, 2015; Schwartz et al., 2020; Sypiańska, 2021; Wojtkowiak, 2022). That is to say, the results show that the participants experience problems suppressing prevoicing in voiced stops in the L2, but manage to produce voiceless plosives with aspiration. The pattern was reversed in the case of L1, where the bilinguals produced fewer prevoiced items and maintained the Polish-like voiceless stops. In this sense, it can be concluded that the voiced sounds in the two languages were perceptually assimilated to a single, shared category, and therefore, were subject to CLI. It is generally believed that aspiration is a more salient feature that learners pay attention to and acquire more easily and early on (Sypiańska, 2021), while prevoicing is hardly noticeable and more difficult to suppress. Phonologically speaking, it is claimed that fortis and lenis stops have different status, meaning that aspiration in voiceless stops is a phonological feature, while prevoicing is said to be a phonetic detail (cf. Schwartz et al., 2020).

The analysis of CLI in multilingual speakers is devoted to determining the possible source of influence on the third language; that is, whether it is L1, L2, or the two combined that shape the development of the L3. The existing data points to mixed results. First of all, some investigations identify the L1 as the main source of CLI (e.g., Bandeira & Zimmer, 2012; Llama & Cardoso, 2018). On the other hand, some studies also point to L2 as a predominant source of CLI on L3 (e.g., Llama et al., 2010; Tremblay, 2007). The greatest number of investigations, however, suggest the combined effect of L1 and L2 on L3 VOT values (e.g., Zhang, 2022; Sypiańska, 2017; Wrembel, 2011, 2014, 2015; Wunder, 2010). For example, Wrembel (2014) analysed VOT production of L1 Polish, L2 English and L3 French multilinguals and found that L3 French values for /p, t, k/ were in between L1 and L2 mean VOT.

Recent investigations tend to employ a more holistic, multilingual approach towards CLI that not only seeks any traces of L3 influence, but also explores a potential impact on L2, as well as L1 drift or attrition (e.g., Cal & Wrembel, 2025; Zhang et al., 2023; Geiss et al., 2022; Nelson, 2022; Zhang, 2022; Amengual, 2021; Sypiańska, 2017; Liu, 2016). Yet, contradictory findings were established in this approach as well. On the flip side, some studies reported progressive and/or regressive CLI. For example, Sypiańska (2017), in her study on L1 Polish, L2 Danish, L3 English speakers, revealed that not only did the multilinguals experience progressive CLI towards their L3 English, but the acquisition of L3 also led to the shifts in their L1 and L2. Similar results for regressive CLI were obtained by Liu (2016) for voiceless stops, and Nelson (2022) in younger learners. Conversely, there are also investigations that show separate VOT patterns for the three languages, meaning that CLI did not occur in the data. For instance, Geiss et al. (2022), in their study on heritage speakers of Italian speaking German as a majority language and English as a third language, found evidence for three separate language systems. Similar findings were also reported by Amengual (2021) for the voiceless velar plosive /k/. Yet another outcome was obtained by Liu (2016) suggesting a separate L3 system and interaction between L1 and L2 for the voiced series.

https://doi.org/10.3390/languages11090182

The research described so far predominantly focuses on VOT production in voiceless stops. Only more recent investigations have dealt with voiced stops and their production in the multilingual context. The results point to the aforementioned asymmetry, with voiced stops being more susceptible to CLI and the voiceless series being produced more accurately. This was observed, for example, by Cal and Wrembel (2025) in their longitudinal study investigating L1 Polish, L2 English, L3 Norwegian speakers. This language triad is characterised by a complementary distribution of VOT values: Polish is a true voicing language, English is an aspirating language and Norwegian combines prevoiced stops, like Polish, and aspirated voiceless stops, as in English. The research showed different patterns of VOT production and its asymmetric development across the two voicing series of plosives. That is, in voiceless stops, multilinguals maintained separate patterns for the three languages that did not change significantly over time, whereas in the voiced plosives significant changes with time were observed by means of shorter prevoicing and an increased number of unvoiced items across the sessions. These findings confirmed previously stated conclusions from the second language acquisition research referring to a different behaviour of the two series of stops (e.g., Wojtkowiak, 2022; Schwartz et al., 2020; Bless, 2015).

Some studies also employed a longitudinal design to explore the development of plosive acquisition in multilinguals (e.g., Cal & Wrembel, 2025; Zhang et al., 2023; Nelson, 2022). Both Cal and Wrembel (2025) and Zhang et al. (2023) showed the asymmetrical behaviour of the two series of stops, confirming that it is the voiced series that is more susceptible to changes over time and, as a result, CLI. On the other hand, Nelson (2022) focused on regressive CLI in the production of voiceless stops only; however, the results pointed to an increasing influence of L3 on L1 and L2 with time in adult learners.

To sum up, the interest in CLI in VOT production has developed significantly in recent years. The earlier investigations focused mostly on L3 acquisition of voiceless plosives, which laid the groundwork for more complex and holistic approaches concerning the multilingual system as a whole; that is, expanding the analyses into voiced stops and employing longitudinal designs.

3. Speech Modes in VOT Production Research

VOT has been shown to be influenced by speech styles, or, in other words, the task type. Building on the work of Labov (1972), speech styles are often conceptualised as a continuum defined by the speaker's level of self-monitoring. While this spectrum ranges from highly controlled word lists to casual conversation, spontaneous speech is further influenced by situational context and communicative intent (Beckman, 1997). Speech modes are also characterised by their degree of spontaneity and preparation (Fujisaki, 1997; Swerts et al., 1996). Based on this, Koch (2008) grouped speech styles into three categories; that is, a reading category consisting of single utterances, read speech, and spontaneous speech. Experimental data shows that in comparison to read speech, spontaneous modality is characterised by reduced spectral space (Nakamura et al., 2008).

It was observed that VOT has considerably more variability when it is produced in spontaneous speech, rather than from read utterances (e.g., Stuart-Smith et al., 2015; Yao, 2009; Syrdal, 1996; Baran et al., 1977). What is more, the rate of speech, together with word frequency and other contextual factors, is believed to have an effect on VOT patterns (Yao, 2009). The existing literature, including the comparison of speech modes in bilinguals, reports shorter VOT values in less formal tasks including spontaneous speech compared to the citation forms of elicitation (e.g., Jerotijević Tišma, 2019; Gósy, 2001); that is, when speech becomes faster and more spontaneous, VOT gets shorter.

https://doi.org/10.3390/languages11090182

The comparison of different speech modes in multilingual research on VOT acquisition remains rather unexplored, as the studies usually seem to be focused on one task only. The most prominently used tasks include: word reading elicited in isolation (e.g., Cal & Wrembel, 2025; Zhang et al., 2023; Zhang, 2022; Llama & Cardoso, 2018; Wrembel, 2014; Llama et al., 2010), words in a carrier sentence (Bondarenko et al., 2022; Amengual et al., 2019) or a picture-naming task (e.g., Dittmers et al., 2017), as well as text reading (e.g., Zhang et al., 2023). In turn, Geiss et al. (2022) and Sypiańska (2017) examined spontaneous speech by means of picture-cued storytelling and structured interviews, respectively. Yet, the results of the two studies seem rather contradictory, with the former reporting language-specific patterns of VOT realisation and the latter pointing to combined CLI. This inconsistency of the findings makes research on VOT production in spontaneous speech worth further exploration.

4. L3 Speech Acquisition Modelling

CLI can be analysed from different perspectives. A considerable body of literature explores the process of acquisition and development of an L3. More specifically, it aims to find any potential sources of influence on this language and factors that might affect this process. The need for systematising and predicting the development of the L3 led to the establishment of certain models that take into account various linguistic and psycholinguistic factors that might play an important role in the process of acquiring a third language, such as language typology, psychotypology, L2 status, among others. The five prominent frameworks for modelling L3 acquisition include the Cumulative-Enhancement Model (Flynn et al., 2004), which assumes that CLI is strictly cumulative and facilitative, alongside models that allow for both facilitative and non-facilitative influence, such as the L2 Status Factor Model (Bardel & Falk, 2007, 2012), the Typological Primacy Model (Rothman, 2011, 2015), the Linguistic Proximity Model (Westergaard et al., 2017), and the Scalpel Model (Slabakova, 2017). However, the assumptions of these models, namely that L3 acquisition is driven by factors such as structural proximity, typological similarity, or L2 status, are largely based on morphosyntactic and lexical data but not phonological ones. Although some attempts have been made to apply these models to phonetic data, predominantly focusing on the L2 Status (e.g., Llama et al., 2010; Wrembel, 2010), further investigations fail to corroborate these assumptions (e.g., Llama & Cardoso, 2018; Wrembel, 2011, 2014, 2015). Thus, these models might be of limited use while establishing specific predictions regarding L3 speech.

A framework that aims to reconciliate the aforementioned models and take the speech domain into consideration is the Natural Growth Theory of Acquisition (NGTA; Dziubalska-Kołaczyk & Wrembel, 2022, 2024). This theory offers a holistic approach that explains multilingual phonological development as a gradual, dynamic emergence of language systems. It integrates the functionalist, preference-based approach of Natural Phonology, which views phonological processes as reflecting real, phonetically motivated constraints on speaker perception and production (Donegan & Stampe, 2009; Stampe, 1979), with the non-linear, self-organising principles of Complexity Theory, which treats language as a dynamic system where stable patterns and generalisations emerge post-factum from usage and experience (Kretzschmar, 2015; Mitchell, 2009). The NGTA theory takes into account the dynamics and individual nature of the acquisition process as well as its dependence on a variety of external variables that may facilitate or inhibit acquisition. This approach defines multiple non-competing scenarios that are viewed as commensurate pathways reflecting the complexity of language acquisition.

According to the NGTA, the acquisition process is believed to depend on a wide range of relevant linguistic and extralinguistic variables which are hierarchised based on their

https://doi.org/10.3390/languages11090182

importance in the learning process. Central to the NGTA is the distinction between two levels of processing: Level 1, which is automatic and involves articulatory routines and phonetic constraints, and Level 2, which is conscious and manifested through metalinguistic awareness. The linguistic factors include the first language and languages other than L1 (including L2, L3, and further languages together with their language-specific systems and typological properties), as well as preferability generalisations (both deductive and inductive). At the same time, the external variables, also referred to "the acquisition situation" comprise the stages of acquisition, frequency of input and use, age of acquisition, proficiency level, metalinguistic awareness and individual factors. Figure 1 illustrates the proposed influential variables (after Dziubalska-Kołaczyk & Wrembel, 2022).

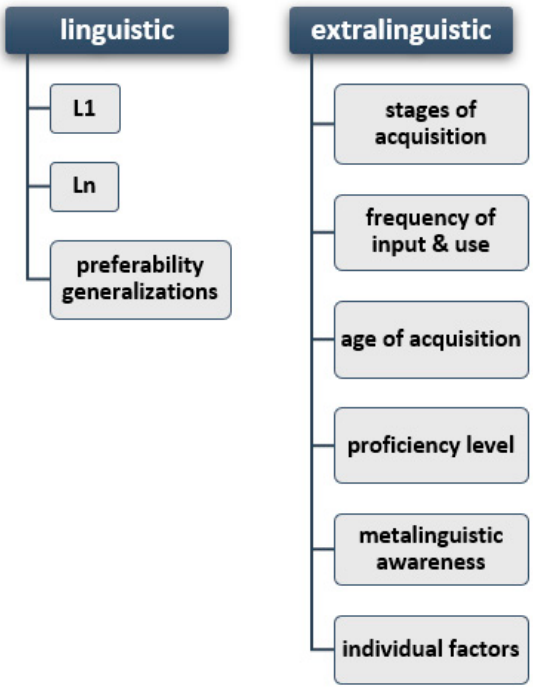


Figure 1. Linguistic and extralinguistic variables in NGTA (Dziubalska-Kołaczyk & Wrembel, 2022).

As the current investigation is centred on speech acquisition specifically and aims to explore the dynamics and complexity of the process of multilingual acquisition, this framework was chosen to serve as the theoretical embedding of the study and to provide a bigger picture of how the process of acquisition is modulated by external factors in an attempt to provide their hierarchy of importance in this particular acquisition situation.

5. Current Study

5.1. Aim

The aim of this study is to investigate how trilingual speakers produce stop consonants in their languages across two speech modalities. Specifically, the goal is to explore potential cross-linguistic interactions across all three languages of L1 Polish, L2 English, L3 Norwegian speakers. To this end, the study analyses VOT in both voiceless (/p, t, k/) and voiced (/b, d, g/) stops to determine whether the learners maintain separate phonological categories or whether CLI emerges, and if so, which voicing series, languages and places of articulation are the most susceptible to this process. The measurements are conducted in both read and spontaneous speech, comparing the results across the speakers' languages and against control values obtained from native speakers of the respective languages. The study also aims at analysing extralinguistic factors that might condition VOT acquisition as

https://doi.org/10.3390/languages11090182

well as the presence, absence and the degree of CLI using the NGTA (Dziubalska-Kołaczyk & Wrembel, 2022, 2024) as the framework of reference.

The languages under investigation, namely Polish, English and Norwegian, are complementary in terms of the characteristics of their stop consonants in word-initial stressed onset position. Polish is a true voicing language that has prevoicing in the voiced series of stops and short-lag VOT in voiceless stops (e.g., Keating et al., 1981; Jassem, 2003). English, on the other hand, is an aspirating language in which /b, d, g/ are, in principle, partially voiced and /p, t, k/ are aspirated (e.g., Lisker & Abramson, 1964; Docherty, 1992). In Norwegian the voiced stops are prevoiced, similarly to the Polish ones, whereas the voiceless stops are aspirated and resemble those in English (e.g., Halvorsen, 1998; Ringen & van Dommelen, 2013). Figure 2 depicts approximate VOT ranges of stops in the three languages under investigation based on the literature. For the sake of comparison, this study will provide reference values for VOT in each language taken from native speakers performing the same tasks as the experimental group.

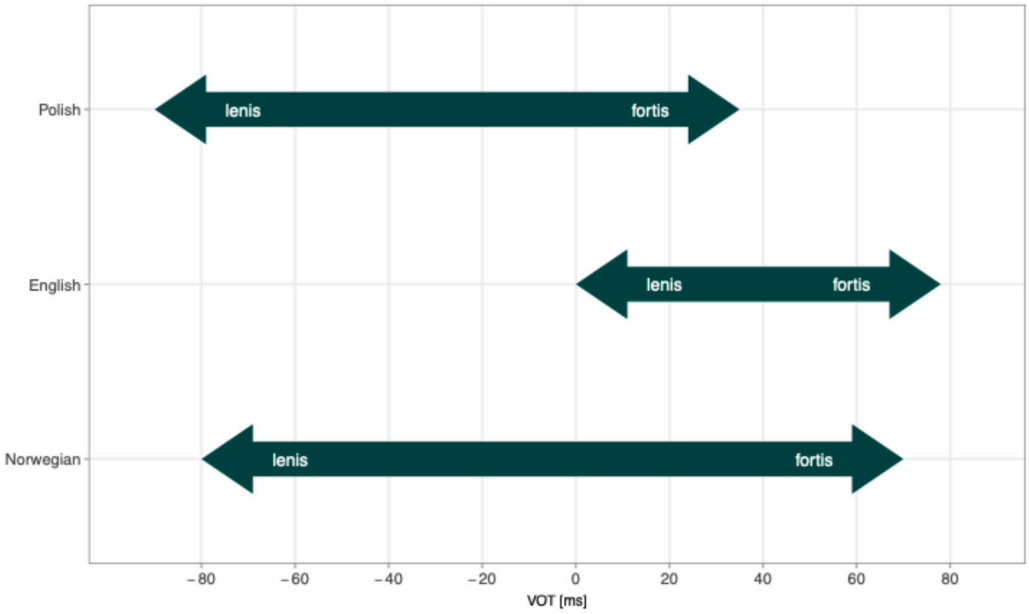


Figure 2. Reference values for VOT in Polish, English and Norwegian (based on Keating et al., 1981, Lisker & Abramson, 1964, and Ringen & van Dommelen, 2013, respectively). Specific values, measured for control speakers of the three languages, are presented in Sections 6.1 and 6.2 for read and spontaneous speech, respectively.

5.2. Research Questions and Hypotheses

The study is guided by the following research questions and corresponding hypotheses:

RQ1. Do multilingual speakers of L1 Polish—L2 English—L3 Norwegian exhibit CLI in the production of stops? Are there any differences in their VOT production compared to native control speakers?

H1. *Based on previous research on VOT acquisition in multilinguals (Sypiańska, 2013; Wrembel, 2015; Geiss et al., 2022; Cal & Wrembel, 2025), we predict CLI to take place in the L1 Polish—L2 English—L3 Norwegian speakers as well. This would be manifested in the differences in VOT values across the multilingual group and native controls.*

RQ2. How is CLI in VOT production of stops manifested across the three languages of L1 Polish—L2 English—L3 Norwegian speakers? Is it dependent on the place of articulation of stops and/or voicing?

https://doi.org/10.3390/languages11090182

H2. *It is expected that all three languages in the repertoire of multilingual speakers will interact, but the interaction will be manifested differently depending on the language pair, place of articulation and voicing series.*

Predictions based on language pairing

We foresee more interaction taking place between L2 English and L3 Norwegian, because of their typological proximity, a similar setting of acquisition (i.e., via classroom instruction), as well as the foreign language effect (e.g., Hammarberg & Hammarberg, 2009). These factors would cause our participants to rely more on L2 values as a reference point while acquiring the L3 and, therefore, they would experience progressive CLI from L2 to L3. Similarly, if the L2 was treated as similar to the L3, regressive CLI would take place from L3 to L2. This would be reflected by longer VOTs in voiceless stops in both L2 and L3, as well as similar realisation of voiced stops in the two languages, across these two systems, with the presence of prevoicing modulated by the directionality of CLI.

Predictions based on voicing series

Based on previous research (e.g., Gabriel et al., 2018; Geiss et al., 2022; Cal & Wrembel, 2025), we hypothesise that more CLI will be attested in the voiced series of stops, while VOT values in the voiceless stops will manifest themselves as more target-like. This prediction rests on the premise that aspiration is a more salient, phonological feature that is more easily acquired early on (Sypiańska, 2021), whereas prevoicing behaves as a less noticeable phonetic detail that is harder to suppress (cf. Schwartz et al., 2020). Therefore, similar realisation of voiced stops in the three languages is anticipated, with the presence of prevoicing dependent on the directionality of CLI. We also expect to find facilitative CLI between L2 English and L3 Norwegian regarding the production of voiceless stops, and between L1 Polish and L3 Norwegian when it comes to the voiced counterparts.

Predictions based on places of articulation

It is predicted that CLI will be present in all three places of articulation of stops, but to a different extent. We expect to observe more CLI in voiceless stops that have higher VOT and voiced stops with the shortest prevoicing, such as the velar stops /k/ and /g/, as their VOT values are the most similar across languages, leading to equivalence classification. Therefore, according to the SLM (Flege, 1995), in these stops which are not sufficiently dissimilar the participants are more likely to form a single category for their three languages. We also hypothesise that we will find the least amount of CLI in coronal stops, as their place of articulation differs considerably across the three languages; that is, /t/ and /d/ are dental in Polish (Dukiewicz & Sawicka, 1995), alveolar in English (Roach, 2009) and their place of articulation varies in Norwegian (Kristoffersen, 2000). This might trigger dissimilatory CLI and, consequently, the formation of separate categories for coronal stops.

RQ3. What is the difference in VOT produced in read vs. spontaneous speech? Are CLI patterns attested to an equal extent in the read speech reflected in spontaneous speech?

H3. *Based on the existing literature, we predict more carefully monitored speech in read mode. As a result, we expect the read speech to be characterised by more accurate VOT production and longer VOT in L2 English and L3 Norwegian voiceless stops, as well as fewer prevoiced stops in L2 English compared to L1 Polish and L3 Norwegian. We also expect to see less individual variation, as the recording process is strictly timed. Conversely, in spontaneous speech we foresee finding shorter VOT in L2 English and L3 Norwegian voiceless stops (e.g., Gósy, 2001; Jerotijević Tišma, 2019), but also more prevoiced items in L2 English. Furthermore, we expect VOT values to be highly dependent on speech rate, resulting in more variation across participants (e.g., Syrdal, 1996;*

https://doi.org/10.3390/languages11090182

Stuart-Smith et al., 2015). Consequently, we predict that more CLI will be found in spontaneous speech, as it is less controlled than the read speech (Jerotijević Tišma, 2019).

RQ4. What is the impact of extralinguistic factors on the production of stops by L1 Polish, L2 English, and L3 Norwegian speakers? Which factors (and to what extent) affect the presence or absence, as well as the degree of CLI?

H4. *Based on the NGTA assumptions (Dziubalska-Kołaczyk & Wrembel, 2022, 2024), the following extralinguistic factors are expected to influence the process of VOT acquisition in multilinguals in this specific acquisition situation to varying extents:*

- L2 and L3 proficiency—The level of proficiency in both L2 and L3 was shown to affect speech acquisition (e.g., Cal & Sypiańska, 2020; Sypiańska & Cal, 2020); thus, it is predicted that higher levels of proficiency in L2 and L3 will trigger more accurate target-like VOT productions and language-specific category formation.
- Mode of speech (read vs. spontaneous)—The highly monitored nature of a read speech task is expected to minimise phonetic variation and, as a result, yield more target-like productions. Conversely, unmonitored spontaneous speech is predicted to exhibit a significantly higher rate of non-target-like productions, including assimilatory CLI.
- Frequency of use of L1, L2, L3—We foresee more target-like VOT production in the three languages with higher frequency of use of the respective languages.
- Amount of exposure to L2 and L3—Higher exposure to L2 and L3 is hypothesised to be a better predictor of more target VOT production in these languages.
- Age of acquisition of L2 and L3—The lower the age of acquisition (AoA), the more probable target-like VOT production. Within the NGTA framework, a younger age of acquisition is associated with less complex cross-linguistic interactions, which limits the multi-directional CLI (both assimilatory and dissimilatory) from affecting the target categories.

Based on our data and the NGTA assumptions, we aim to propose a hierarchy of predictors valid in this specific acquisition situation (cf. Dziubalska-Kołaczyk & Wrembel, 2024).

5.3. Participants

The present study involved 37 L1 Polish, L2 English, L3 Norwegian multilingual learners and, in total, 50 native controls of the three languages. To collect the data regarding their language history and use, the participants completed the Language History Questionnaire (Li et al., 2020). The experimental group additionally completed standardised proficiency tests in L2 English and L3 Norwegian, while L1 Polish and L1 Norwegian controls were asked to fill in a proficiency test in L2 English. We used adapted versions of the English and Norwegian written placement tests that assessed proficiency up to the C1 and B2 levels, respectively. The participants received gift cards for their involvement in the study, and the ethics committee approval was obtained prior to the study. The sections below provide detailed profiles of the two groups.

5.3.1. Experimental Group

There were 37 L1 Polish–L2 English–L3 Norwegian speakers in the experimental group, aged 21.28 (sd = 1.98). The group included 33 females, 3 males and 1 nonbinary person. All speakers were students or graduates of Norwegian studies at two Polish universities (in Poznań and Szczecin). The pool consisted of 15 first-year students, 12 second-year students and 10 students at the third-year level or above. The proficiency test results classified them on average as B2 (upper-intermediate) speakers of English (mean score = 77.3%, sd = 13.9%) and B1 (intermediate) speakers of Norwegian (mean score = 82.9%, sd = 16.3%),

https://doi.org/10.3390/languages11090182

according to the Common European Framework of Reference for Languages. All of the participants attended Norwegian practical courses including speaking, grammar, writing and pronunciation; additionally, their curriculum included a general English course oriented towards integrated skills. The participants reported limited knowledge of other foreign languages, particularly German (N = 15) and French (N = 5), learnt by means of prior formal instruction. A few students claimed to have lived in a foreign speaking country (England, Norway and Germany) for a period of 3–12 months. Detailed biodata information about the experimental group (i.e., data regarding their foreign language exposure and use) can be found in the OSF repository[1].

5.3.2. Control Groups

The study involved 3 subgroups of control speakers that consisted of native speakers of Polish, English and Norwegian respectively. There were 20 L1 Polish speakers, 18 L1 English speakers and 12 L1 Norwegian speakers. We aimed to match the profiles of the control speakers to our experimental group in terms of their age and education. Thus, all three subgroups comprise young and educated individuals, most of whom were students or graduates of non-linguistic majors.

The L1 Polish controls included native speakers of Polish, consisting of 18 females and 2 males. Their mean age was 21.77 (sd = 2.2). They all reported having learnt English or other foreign languages at school, but they claimed not to have a regular contact with any of the foreign languages.

There were 18 L1 English controls residing in Bangor (UK), whose mean age was 23.6 (sd = 3.8). The group included 11 females and 7 males. Some of the participants indicated limited proficiency in Romance languages (French, Spanish or Italian) and Welsh, but at the time of the study they reported not using foreign languages on a daily basis.

The L1 Norwegian controls consisted of 9 females and 3 males, who were aged 25 (sd = 3). They resided in Tromsø (Norway). All of them were advanced (C1-C2 level) speakers of English who started learning the language approximately at the age of 6.6 (sd = 2.7). We decided to include speakers of such a linguistic profile in the experiment, as a monolingual group would have been rather difficult to find due to the relatively high level of English proficiency among the Norwegian population.

5.4. Procedure

In this study the experimental group performed two production tasks including read and spontaneous speech separately for Polish, English and Norwegian. Firstly, they took part in a reading task, in which they were presented with sentences on a computer screen, one sentence per slide, and asked to read them in a natural manner. The slides were changed at a steady, controlled pace. The sentences included target words containing word-initial stops /p, t, k, b, d, g/ in a stressed onset position that was controlled for vocalic context. There were 60 sentences per language, each list targeted voiced and voiceless stops and the three places of articulation in an equal number. Thus, there were 10 target words per sound. The list of sentences for each language can be found in OSF repository (see Note 1 above). The second task targeted the production of spontaneous speech. In this task, the participants were asked to talk for about 2 min on a given topic that was varied across the three languages. In the Polish block, the participants talked about their childhood in Poland, in the English block the topic was summer vacation abroad, while in Norwegian they were asked to talk about free time in Norway. To avoid a language mixing effect, the language sessions were conducted on different days. Additionally, a specific language mode was introduced before each session, by means of a task that required them to watch a short 3 min video in a particular language and then to retell the story in this

 https://doi.org/10.3390/languages11090182

language. The control speakers participated in the same tasks, but in one language block only, corresponding to their native language.

The recordings were made using a Shure SM-35 unidirectional cardioid head-worn condenser microphone and a Marantz PMD620 recording device with the microphone positioned approximately 6 cm away from the participant's mouth.

The recordings were then transcribed with the use of an ASR tool and force-aligned using WebMAUS (Kisler et al., 2017). Based on the aligned TextGrid files, a Python (version 3.12.4) script was used to find target words and sentences. In the spontaneous speech recordings, the script targeted only word-initial voiced and voiceless stops that were followed by a vowel and preceded by a word ending with a vowel. After that, VOT boundaries in the targeted sounds were manually corrected in Praat (Boersma & Weenink, 2023). Measurements for voiced stops captured the entire prevoicing interval, extending from the initial pulse of voicing up to the point of articulatory release. Neither the burst nor release noise was included in this interval. For voiceless stops, the measurements included the positive VOT interval from the release burst to the onset of the subsequent vowel. In all cases, the acoustic markers were placed at the zero-crossing points of the waveform. The tokens were verified to include only the items with full acoustic closure. Any stops exhibiting lenition or non-standard articulation were discarded from the analysis. Finally, the VOT durations (in milliseconds) were extracted using another Python script. This procedure was chosen, as it allowed us to extract as many VOT durations as possible in the quickest possible time, yet we still faced some data loss due to speakers' mispronunciation or forced-alignment issues.

We obtained a total of 8462 tokens from read speech, within which 5845 tokens were produced by the experimental group and 2617 by the control group. For the spontaneous modality, the dataset yielded 2088 tokens, out of which the experimental group accounted for 911 tokens, while the control group contributed 1177 tokens. A relatively smaller number of tokens was obtained from the spontaneous speech recordings, due to the uncontrolled nature of the task. As the task for spontaneous speech was not experimentally controlled for specific target items, several participants did not produce any stop-initial words within the required phonological contexts. Therefore, the study includes data only from a subset of participants from the experimental group (N = 27) and all of the control speakers.

6. Results

Having obtained the VOT durations, we employed Linear Mixed-Effects Models (LMMs) and Generalised Linear Mixed-Effects Models (GLMMs), as implemented in the lme4 package (Bates et al., 2015) in R. The models were run separately for the two speech modes and included the full factorial interaction between the fixed effects. Additionally, speech rate operationalised as the number of actually pronounced phonemes per second (see van Dommelen, 2006) was included as a covariate in the model concerning spontaneous speech. The code for the analysis as well as the models' outputs can be found in the OSF repository (see Note 1 above).

The duration analysis was conducted to measure the effect of sound, language and group on VOT duration (ms). The dependent variable was scaled using the inverse hyperbolic sine (IHS) transformation prior to the analysis using the asinh() function in R, in order to improve residual normality. Random effects included participant and token. The following formula was implemented to run the model: VOT_scaled ~ sound * language * group + (1 | participant) + (1 | token).

The prevoicing analysis was done using a Logistic GLMM with a binomial distribution. The model tested the effects that sound, language and group have on the presence or absence of prevoicing (coded as 1 = presence, 0 = absence thereof), with participant and

https://doi.org/10.3390/languages11090182

token as random effects. The model formula was as follows: prevoicing ~ sound * language * group + (1 | participant) + (1 | token). To avoid convergence issues, the model was specified with the following control parameters: control = glmerControl(optimizer = "optimx", optCtrl = list(method = "L-BFGS-B", maxit = 10,000)). Table 1 summarises the variables and factors that were tested in the two types of models.

Table 1. Factors and variables used in the statistical analyses.

Model	Dependent Variable	Scale	Fixed Effects	Random Effects
Model 1: VOT duration	VOT_scaled_ihs (in milliseconds)	Continuous (crosses zero)	Sound (/p, t, k, b, d, g/); Language (Polish, English, Norwegian); Group (Experimental, Control)	Random intercept of participant; Random intercept of token word
Model 2: Prevoicing	Prevoicing presence/absence	Binary (0 = absent, 1 = present)	Sound (/b, d, g/ subset); Language (Polish, English, Norwegian); Group (Experimental, Control)	Random intercept of participant; Random intercept of token word

Furthermore, to analyse the effect of extralinguistic factors on VOT duration, as provided in RQ4, we conducted a Random Forest Regression using the randomForest package (Liaw & Wiener, 2002) in RStudio (version 2026.01.0+392; Posit Team, 2025). VOT duration (IHS-transformed) was used as the outcome variable, while predictor variables included: L2 proficiency (%), L3 proficiency (%), exposure to L3 (years), speech mode (read vs. spontaneous), AoA in L2 and L3 (years), and frequency of use in L1, L2, L3 (7-point Likert scale). The model formula was as follows: VOT_ms ~ sound + language + participant + voicing + L2_prof + L3_prof + exposure_L3 + speechmode + AoA_L2 + AoA_L3 + frequency_L2 + frequency_L3. Alongside the Random Forest analysis, LMMs were run with the same predictors as a validation step to check for linear relationships.

6.1. Read Speech

The first objective of this study was to determine whether multilingual speakers of L1 Polish, L2 English, L3 Norwegian exhibit CLI in their production of stops (RQ1), which could be assessed in comparison to control values. Table 2 presents mean VOT values and standard deviations for both the multilingual group and the control speakers across all three languages, separately for each stop. As can be seen in Polish, the greater differences across the two groups can be observed in the voiced series of stops, where multilingual participants produced stops with shorter prevoicing than baseline values obtained from control speakers of L1 Polish. When it comes to English, the multilingual group produced voiceless stops with considerably shorter VOT, and voiced stops with much longer prevoicing than L1 English controls. As far as Norwegian is concerned, VOT durations in the multilingual group were shorter for /p/ and /t/, and slightly longer for the three voiced stops compared to the control group.

The results of the statistical analysis revealed significant main effects of Sound ($F(5, 167.71) = 3628.56$, $p < 0.0001$), Language ($F(2, 222.09) = 149.08$, $p < 0.0001$), and Group ($F(1, 83.93) = 57.57$, $p < 0.0001$). Crucially, the analysis showed a significant three-way interaction between Sound, Language and Group ($F(10, 8210.07) = 58.70$, $p < 0.0001$), which indicates that the differences in VOT production between the multilingual and control groups were not uniform across languages or places of articulation. To further explore these interactions, post hoc pairwise comparisons were performed using the emmeans package (Lenth & Piaskowski, 2026) in R. Figure 3 depicts the distribution of VOT values for voiceless stops /p, t, k/ across the three languages and two speaker groups. Although the distribution

 https://doi.org/10.3390/languages11090182

of Polish VOT values seems comparable across the two groups, suggesting highly stable phonetic targets for voiceless stops, the multilinguals showed wider and more diffuse distributions in English and Norwegian with their peak lower than in the controls. The post hoc analysis revealed significant differences between the two groups in English /p/ ($z = 2.95$, $p = 0.003$) and English /t/ ($z = 3.66$, $p = 0.0003$).

Table 2. Descriptive statistics for VOT (in ms) in read speech across two groups.

Language	Sound	Mean VOT (*sd*) Multilingual Group	Mean VOT (*sd*) Control Groups
Polish	p	23.59 (12.7)	22.33 (9.22)
	t	28.24 (11.93)	26.66 (11.17)
	k	42.15 (13.75)	39.08 (10.88)
	b	−66.06 (19.21)	−73.01 (24.07)
	d	−53.98 (17.75)	−62.17 (20.87)
	g	−46.71 (18.73)	−55.24 (19.21)
English	p	45.14 (25.98)	59.69 (19.61)
	t	50.54 (25.85)	75.29 (21.04)
	k	69.5 (26.29)	73.97 (21.43)
	b	−61.67 (34.73)	−6.67 (17.86)
	d	−47.08 (30.66)	−8.19 (16.82)
	g	−36.37 (32.15)	−7.39 (14.27)
Norwegian	p	42.52 (28.18)	48.2 (17.42)
	t	41.5 (21.61)	50.02 (17.35)
	k	57.95 (22.1)	55.67 (14.98)
	b	−66.53 (30.41)	−60.02 (26.83)
	d	−48.09 (27.97)	−41.14 (27.85)
	g	−50.47 (30.27)	−43.39 (32.59)

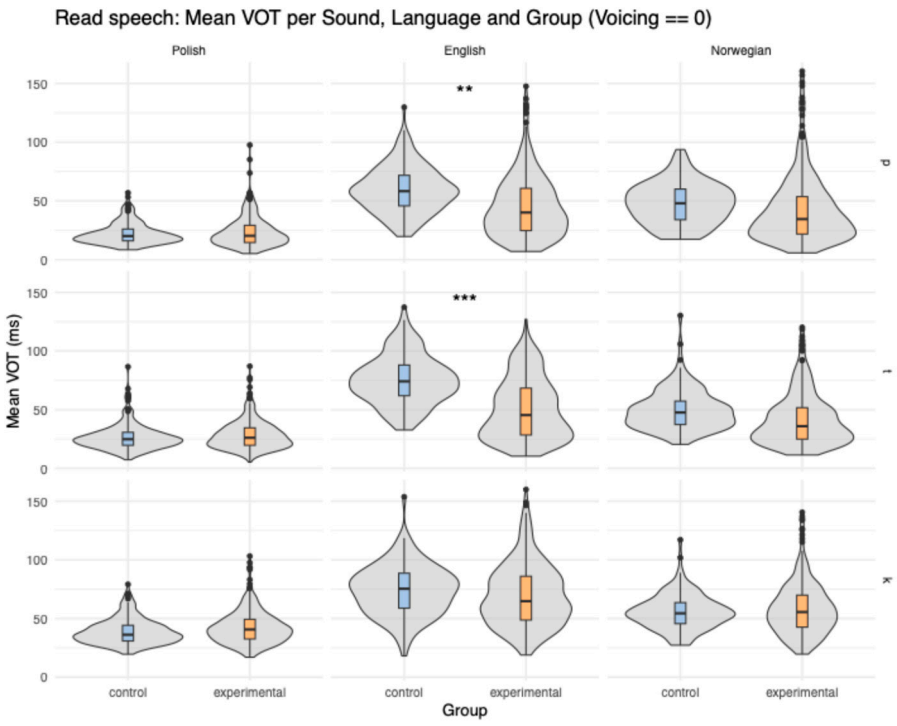


Figure 3. VOT for voiceless stops (/p, t, k/) in read speech across Polish, English and Norwegian for the control and multilingual (experimental) groups. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Figure 4 shows corresponding across-group comparisons for the voiced stops. The greatest difference between the two groups can be noticed in English, where the multilin-

https://doi.org/10.3390/languages11090182

gual group exhibited a wide range of negative VOT values which are not present in the control data. The difference was significant for all three voiced stops, that is, /b/ ($z = 24.49$, $p < 0.0001$), /d/ ($z = 18.90$, $p < 0.0001$) and /g/ ($z = 14.39$, $p < 0.0001$). Furthermore, multilingual speakers produced the Polish stops with shorter prevoicing; however, the difference reached significance only for the velar stop /g/ ($z = -2.71$, $p = 0.007$). In Norwegian, on the other hand, the multilingual group produced the voiced stops with slightly longer prevoicing than the control group. This difference turned out to be significant for /d/ ($z = 3.33$, $p = 0.0009$) and /g/ ($z = 3.44$, $p = 0.0006$).

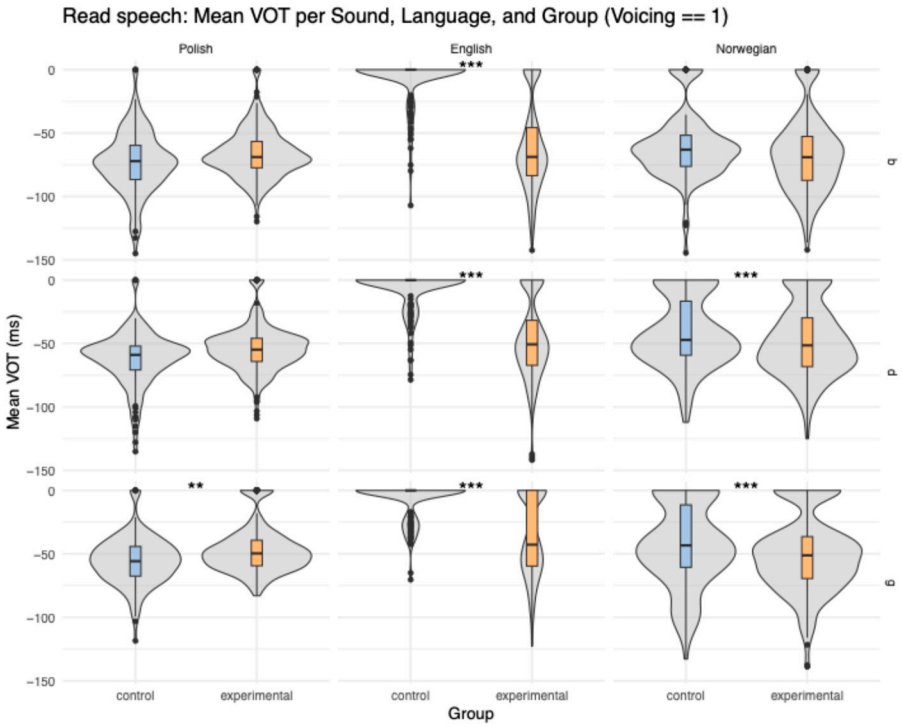


Figure 4. VOT for voiced stops (/b, d, g/) in read speech across Polish, English and Norwegian for the control and multilingual (experimental) groups. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Given the substantial variability and the high frequency of unvoiced realisations observed in the voiced stop categories (see Figure 4), we decided to conduct a categorical analysis of the pre-voicing presence versus absence in the data to determine whether CLI in multilingual speakers manifests itself as a shift in duration or a categorical change in the voicing lead. The GLMM analysis revealed significant main effects of language ($\chi^2(2) = 111.43$, $p < 0.0001$) and group ($\chi^2(1) = 14.89$, $p = 0.0001$). These effects were modulated by a significant three-way interaction between sound*language*group ($\chi^2(1) = 16.48$, $p = 0.002$), which indicates that not only did the probability of producing a voicing lead depend on language, but it also varied significantly by the place of articulation and group. Figure 5 presents predicted probabilities of producing prevoicing in the three languages across the two groups under investigation. As the figure demonstrates, the most notable differences across the groups can be found for English, where the control group was significantly less likely to produce prevoicing in all three places of articulation of stops compared to multilingual speakers (i.e., /b/ ($z = -9.07$, $p < 0.0001$), /d/ ($z = -7.11$, $p < 0.0001$), /g/ ($z = -5.03$, $p < 0.0001$)). In Norwegian, the multilingual group turned out to produce voiced stops with a slightly higher probability of prevoicing compared to the controls, but the difference was significant only for /d/ ($z = -2.34$, $p = 0.02$). When it comes to Polish, the probability of producing prevoicing was the highest in the two groups, reaching almost 100%. No statistically significant differences across the groups were found.

https://doi.org/10.3390/languages11090182

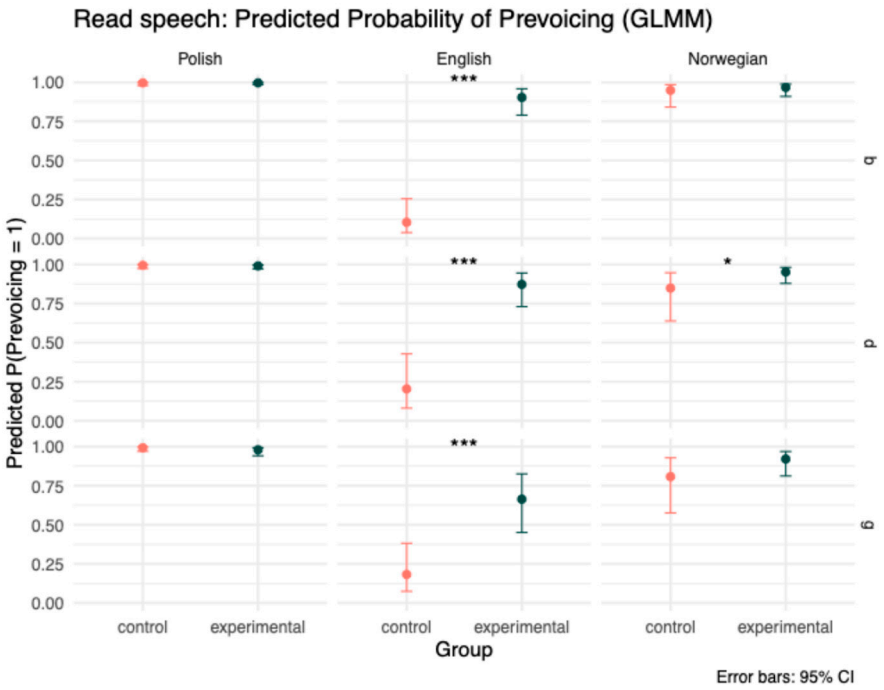


Figure 5. Predicted probability of producing prevoicing in voiced stops in the three languages across two groups—control and multilingual (experimental). * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

The second research question dealt with the directionality of CLI in VOT production of stops across the three languages; thus, it concerns the multilingual group's performance only. Figure 6 shows mean VOT for each stop across the three languages in the multilingual group. In voiceless stops, it can be observed that the multilingual group produced all three plosives with the shortest VOT value in L1 Polish, while the longest values were found in L2 English. L3 Norwegian VOT was found to be in between those of L1 and L2. Post hoc across-language comparisons revealed statistically significant differences between mean VOT durations in L1 Polish and L2 English in all three places of articulation (i.e., /p/: $t = -3.23$, $p = 0.004$; /t/: $t = -2.95$, $p = 0.009$; /k/: $t = -2.52$, $p = 0.03$). Additionally, for /p/, a statistically significant difference was found between L1 Polish and L3 Norwegian ($t = -2.66$, $p = 0.02$). As far as the voiced series of stops is concerned, the mean length of prevoicing seems to be rather homogenous across the three languages; however, a greater variation in length, but also the presence or absence of prevoicing can be seen in L2 English and L3 Norwegian. The post hoc analysis pointed to significant differences between the VOT realisations in L1 Polish and L2 English in all three stops, that is, /b/ ($t = -3.58$, $p = 0.001$), /d/ ($t = -4.2$, $p = 0.0001$), and /g/ ($t = -7.015$, $p < 0.0001$). Moreover, statistically significant differences were also found between VOT in L1 Polish and L3 Norwegian /d/ ($t = -2.47$, $p = 0.04$), as well as between L2 English and L3 Norwegian /g/ ($t = 5.99$, $p < 0.0001$).

Figure 7 shows predicted probabilities of prevoicing realisation in the three languages and places of articulation of voiced stops for the multilingual group. As shown on the graph, the probability of producing prevoicing is the lowest in L2 English, followed by L3 Norwegian and L1 Polish, which reached almost 100% probability of producing stops with a voicing lead. Statistically significant differences in the probability of prevoicing realisation were found between L1 Polish and L2 English in all three places of articulation of stops, that is, /b/ ($t = 4.34$, $p < 0.0001$), /d/ ($t = 4.03$, $p = 0.0002$), and /g/ ($t = 4.82$, $p < 0.0001$). What is more, significant differences were also found between L1 Polish and L3 Norwegian realisation of voiced stops in /b/ ($t = 2.72$, $p = 0.02$) and /d/ ($t = 2.42$, $p = 0.04$).

https://doi.org/10.3390/languages11090182

Another statistically significant difference was also revealed between the L2 English and L3 Norwegian in the velar stop /g/ ($t = -2.9$, $p < 0.0001$).

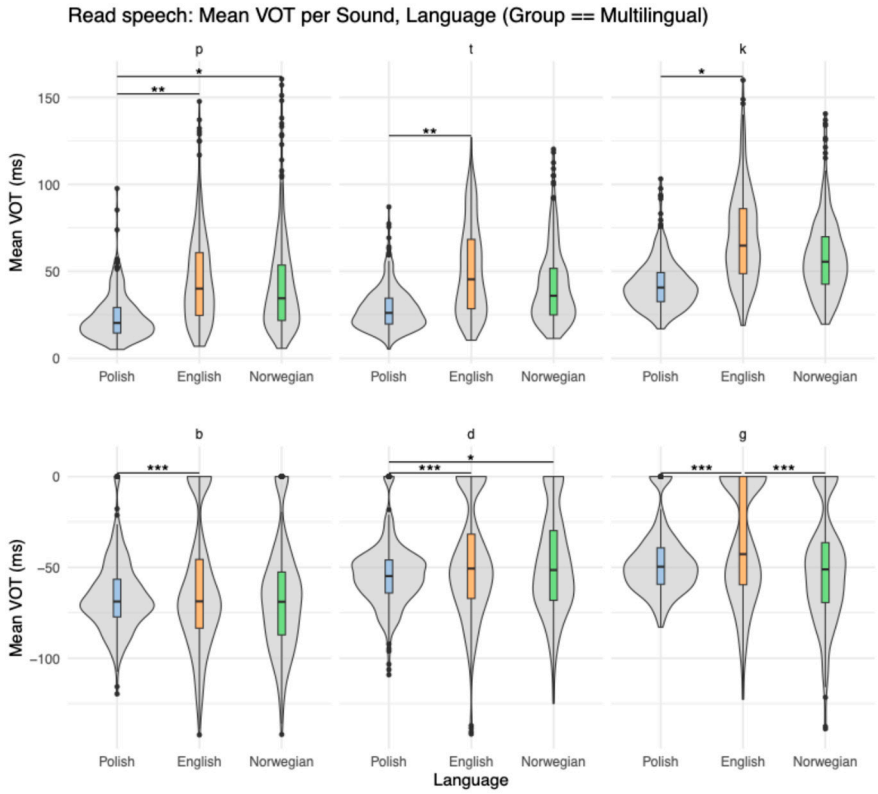


Figure 6. Mean VOT for voiceless and voiced stops across L1 Polish, L2 English, L3 Norwegian in the multilingual group. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

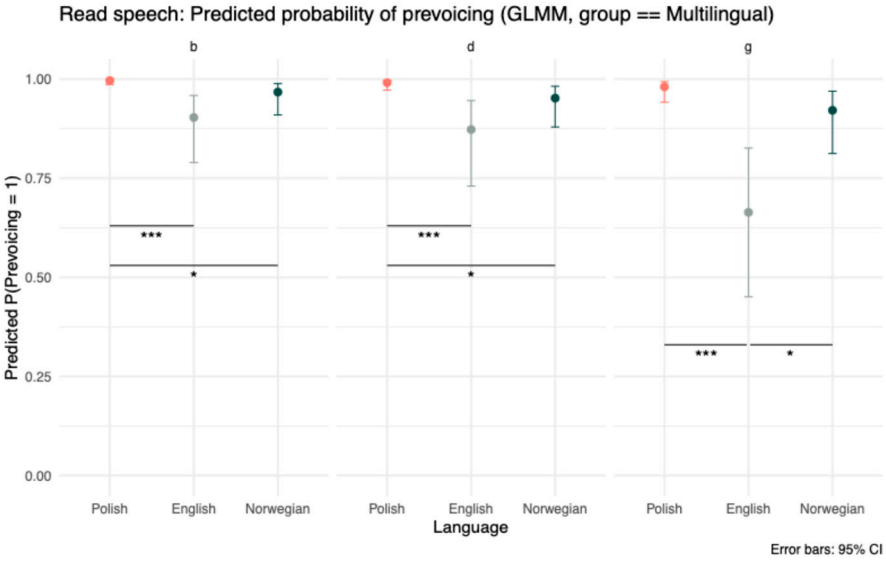


Figure 7. Predicted probability of producing prevoicing in multilingual group shown separately for three stops and languages. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

6.2. Spontaneous Speech

The third research question investigated whether the CLI patterns observed in the read speech data were mirrored in spontaneous production. To ensure comparability, the spontaneous speech data were subjected to the same statistical modelling and categorical analysis as previously described for the read speech. Descriptive statistics for mean VOT

https://doi.org/10.3390/languages11090182

values and their corresponding standard deviations across all stop categories, languages and groups, specifically for spontaneous speech, are summarised in Table 3. Compared to the read speech, shorter VOT values can be observed in the spontaneous speech in both the experimental and control groups. Differences between the two groups are attested especially in English in both series of stops; however, they appear smaller than those in the read speech data. Moreover, slightly greater variation in the data can also be observed in both groups in spontaneous speech compared to read speech.

Table 3. Descriptive statistics for VOT (in ms) in spontaneous speech across the two groups.

Language	Sound	Mean VOT (*sd*) Multilingual Group	Mean VOT (*sd*) Control Groups
Polish	p	29.42 (13.93)	32.82 (14.85)
	t	27.47 (12.1)	27.84 (10.45)
	k	47.2 (13.46)	51.82 (15.09)
	b	−60.99 (23.88)	−63.44 (24.54)
	d	−55.32 (29.31)	−52.76 (21.87)
	g	−41.26 (29.55)	−58.34 (39.4)
English	p	49.58 (26.94)	56.71 (23.26)
	t	64.66 (32.1)	73.68 (25.06)
	k	55.44 (25.87)	61.56 (22.77)
	b	−70.9 (28.58)	−8.9 (20.99)
	d	−50.72 (26.25)	−4.03 (11.84)
	g	−54.5 (31.19)	−4.37 (13.16)
Norwegian	p	45.08 (17.29)	42.06 (16.65)
	t	61.07 (26.79)	55.85 (21)
	k	59.62 (24.71)	54.89 (18.2)
	b	−68.9 (26.61)	−57.44 (29.97)
	d	−47.98 (34.66)	−43.94 (24.31)
	g	−53.54 (37.1)	−38.09 (21.69)

Statistical analysis revealed significant main effects of sound ($F(2, 245.89) = 2722.98$, $p < 0.0001$), language ($F(2, 226.22) = 65.57$, $p < 0.0001$) and group ($F(1, 111.63) = 44.31$, $p < 0.0001$), together with a significant three-way interaction between the three variables ($F(10, 1347.13) = 26.75$, $p < 0.0001$). Figures 8 and 9 show mean VOT for voiceless and voiced stops, respectively, produced in spontaneous speech. In voiceless stops, the most substantial differences between the groups were found for English. VOT productions in the other two languages are rather comparable across the two groups. Post hoc comparisons did not reveal any statistically significant differences between control and multilingual speakers. As far as the voiced series of stops is concerned, the most notable differences between the groups can be seen in English, in which the control speakers produced very few tokens preceded by prevoicing, compared to the multilinguals. Post hoc comparisons showed significant differences between the two groups in English in all three places of articulation, that is, /b/ ($t = 16.78$, $p < 0.0001$), /d/ ($t = 15.06$, $p < 0.0001$), and /g/ ($t = 14.95$, $p < 0.0001$). Another statistically significant difference was attested for the Norwegian coronal stop /d/ ($t = -2.09$, $p = 0.04$), where multilinguals produced the stop with slightly longer voicing lag on average, but also with higher variation.

With regard to across-language comparisons in the multilingual group, it can be observed that the participants produce L1 Polish voiceless stops with the shortest VOT, while L2 English and L3 Norwegian exhibit rather comparable, longer values. In the case of the voiced stops, however, the prevoicing length seems to be quite similar across the languages. These comparisons are presented in Figure 10. Post hoc testing pointed to significant differences between the languages in coronal stops in both voiceless (/t/, English–Polish:

https://doi.org/10.3390/languages11090182

$t = 3.83$, $p = 0.0004$, Norwegian–Polish: $t = 3.25$, $p = 0.004$) and voiced series (/d/, English–Norwegian: $t = -3.14$, $p = 0.005$, Norwegian–Polish: $t = 2.86$, $p = 0.01$), as well as in the velar stop /g/ (English–Norwegian: $t = -2.4$, $p = 0.04$). The categorical analysis of prevoicing was also performed, but it did not point to any statistically significant differences in the probability of prevoicing realisation in the three languages. The multilingual participants produced voiced stops in all three languages with a very high probability of prevoicing.

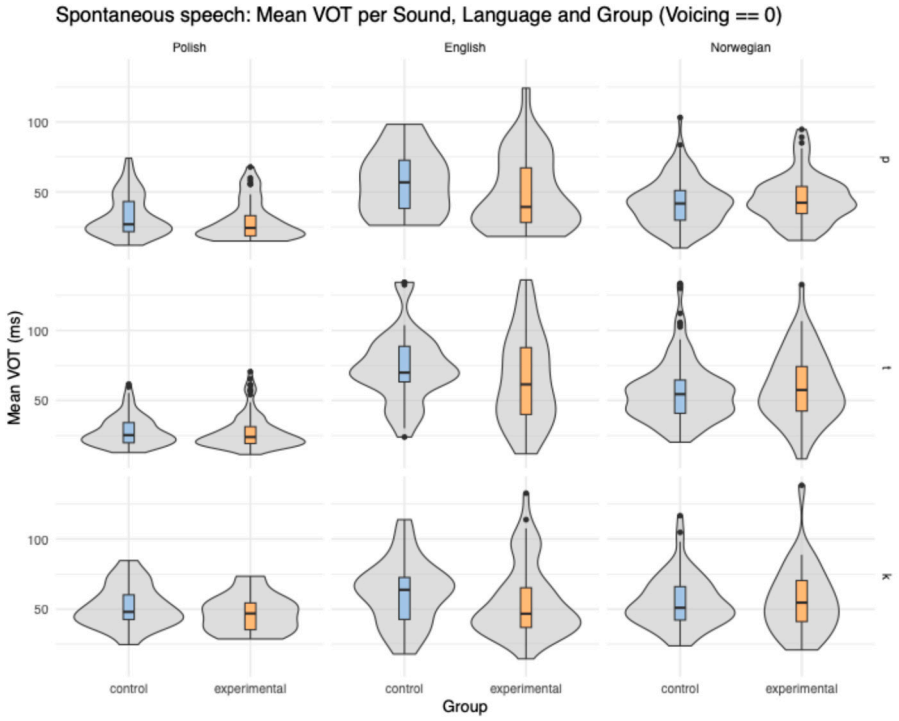


Figure 8. VOT in voiceless stops /p, t, k/, produced in spontaneous speech across two groups—control and multilingual (experimental).

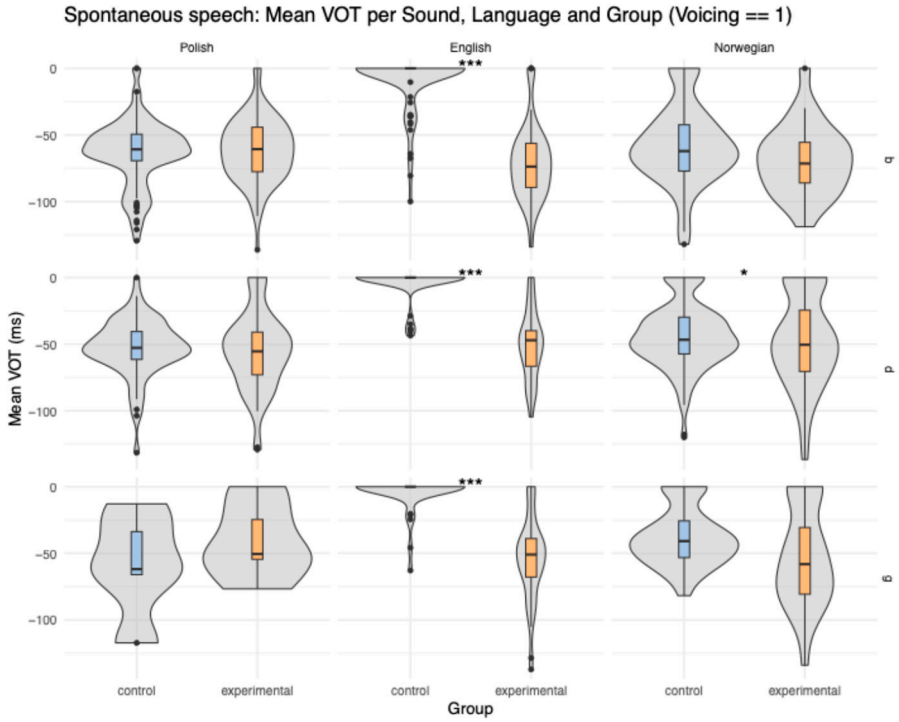


Figure 9. VOT in voiced stops /b, d, g/, produced in spontaneous speech across two groups—control and multilingual (experimental). * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

https://doi.org/10.3390/languages11090182

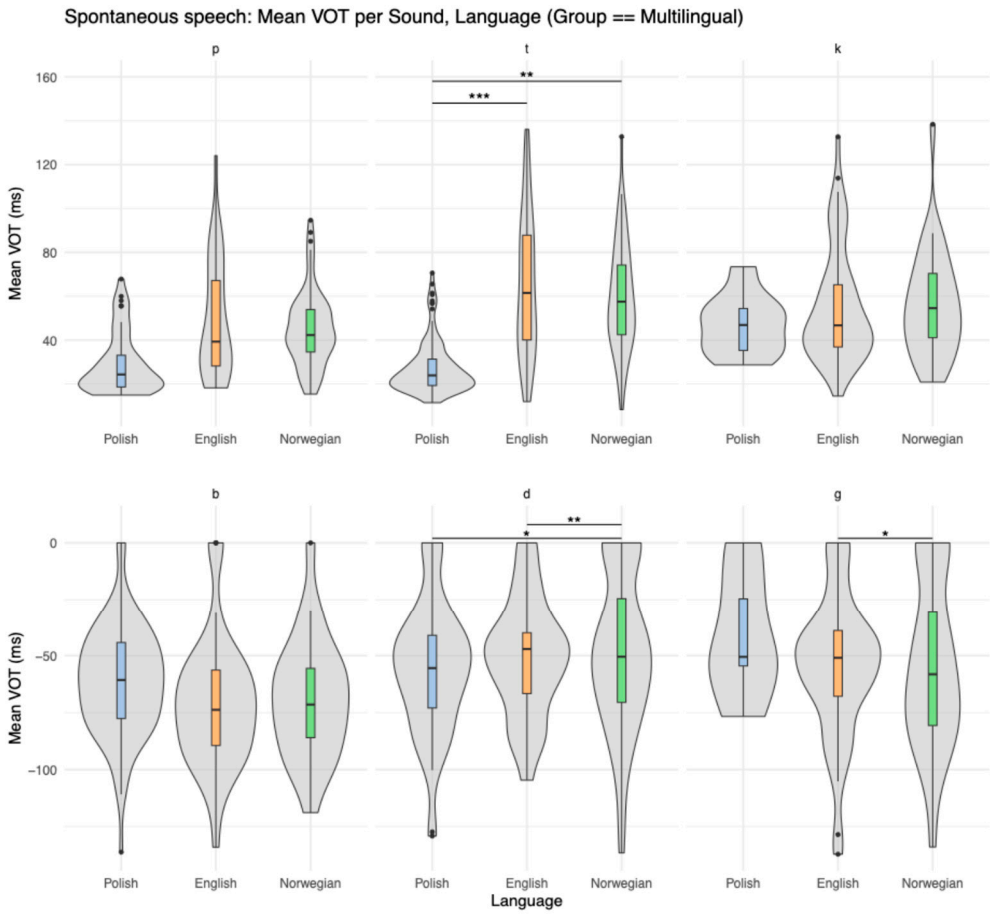


Figure 10. Mean VOT for voiceless and voiced stops across L1 Polish, L2 English, L3 Norwegian in the multilingual group produced in spontaneous speech. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

Finally, Figure 11 shows the comparative distributions of VOT for voiceless and voiced plosives across Polish, English and Norwegian between read and spontaneous speech modes in the multilingual group. The data reveals a divergence in VOT between the speech modes across all three languages. For the voiceless stops, spontaneous speech consistently showed higher variation compared to the read condition, as evidenced by the increased vertical elongation of the violin plots. While the mean VOT for English and Norwegian voiceless stops remained relatively stable across the modes, the spontaneous speech yielded a higher frequency of more extreme values. In contrast, the voiced stops exhibited a distinct trend towards the reduction in prevoicing, particularly within the Polish data. Overall, while read speech maintained a more unified production of stops, spontaneous speech was characterised by a broader range of values and a general trend towards the shortening of prevoicing.

https://doi.org/10.3390/languages11090182

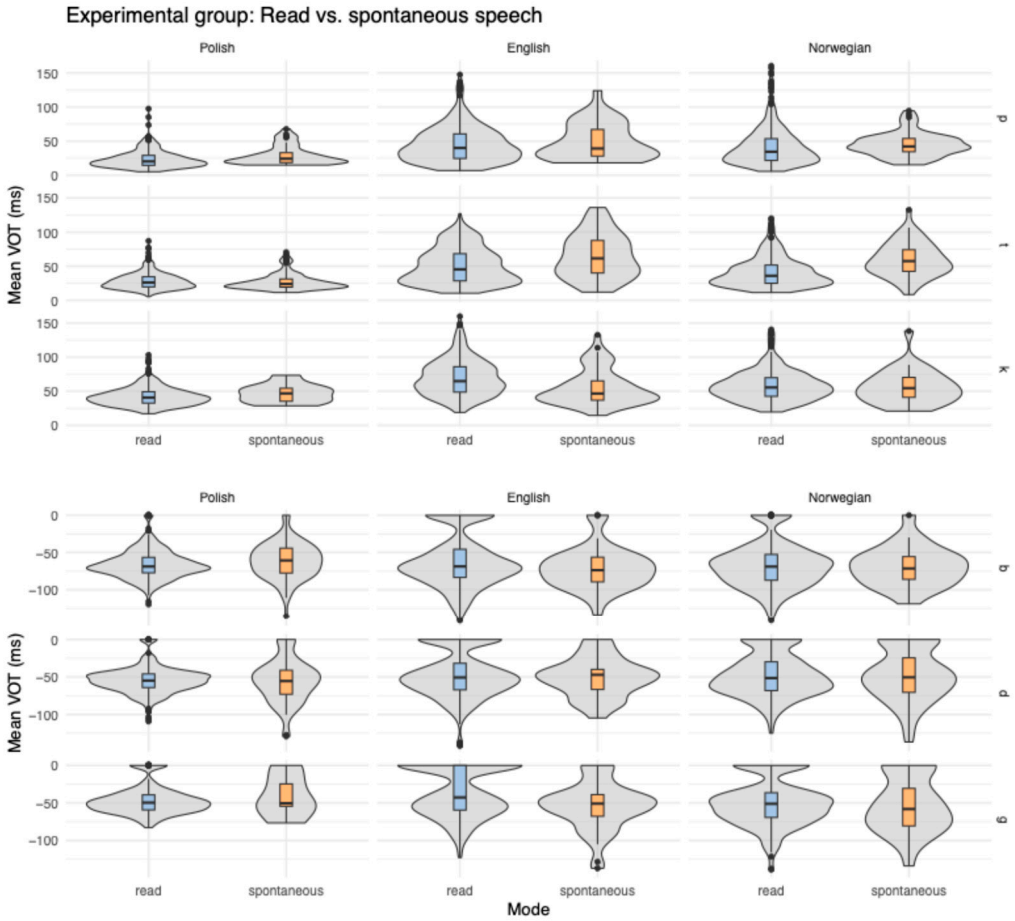


Figure 11. Distribution of VOT in the multilingual (experimental) group by language, stop sound, and speech mode.

6.3. Random Forests Analysis of Extralinguistic Factors

A Random Forest regressor was trained to predict VOT duration (ms) using 12 linguistic and extralinguistic predictors, i.e., language (L1 Polish, L2 English, L3 Norwegian), sound (/p, t, k/, /b, d, g/), speech mode (read, spontaneous), L2 and L3 proficiency (%), L2 and L3 age of acquisition (AoA; in years), as well as L1, L2 and L3 frequency of use and L3 exposure (years). The model for voiced stops accounted for 30.16% of the variance, while the voiceless model accounted for 50.32%, indicating a moderate fit for the data.

The variable importance hierarchy is presented in Figure 12. It was revealed that Language and Sound were the most influential factors for both stop series. However, while Language was a primary predictor for voiceless stops (IncMSE = 239.16%), its influence was substantially reduced for voiced stops (IncMSE = 97.47%). This suggests that language-specific realisations are more strictly maintained in the voiceless category.

For voiced stops, L2 proficiency (IncMSE = 78.28%), L2 age of acquisition (IncMSE = 69.06%) and the frequency of L3 use (IncMSE = 62.39%) turned out to be among the top five predictors, while in the voiceless series the main predictors included L2 age of acquisition (IncMSE = 91.22%), L3 proficiency (IncMSE = 72.16%) and language mode (IncMSE = 71%).

In the LMMs, however, the only variable that turned out to be significant was speech mode ($\chi^2(1) = 16.649$, $p < 0.0001$) in voiceless stops. It was revealed that, on average, voiceless stops were produced with significantly longer VOT in the spontaneous speech mode (average VOT = 46.1 ms, *SD* = 25.6), compared to read speech (average VOT = 44 ms, *SD* = 25.1). Figure 11 visualises the differences between speech modes separately for language and place of articulation.

https://doi.org/10.3390/languages11090182

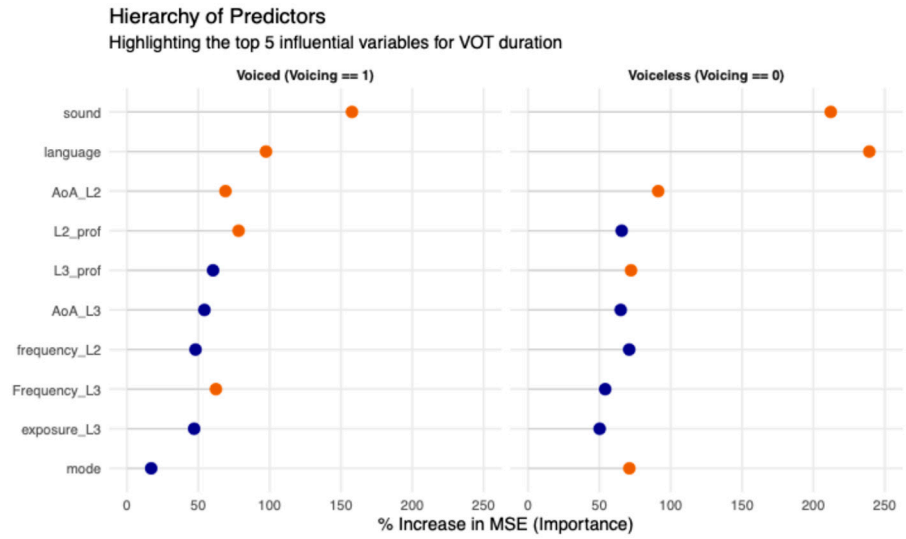


Figure 12. The variable importance hierarchy for voiced and voiceless stops. The top 5 influential variables are highlighted in red.

7. Discussion

This study aimed to investigate how trilingual speakers produce stop consonants in their languages across two speech modalities. The overarching goal was to explore potential cross-linguistic interactions across all three languages of multilingual speakers, as well as selected extralinguistic factors that might condition VOT acquisition as well as the presence, absence and the amount of CLI.

The first research question enquired whether multilingual speakers of L1 Polish—L2 English—L3 Norwegian exhibit CLI in the production of stops, which would be manifested in the differences in VOT values compared to control speakers. It was hypothesised that CLI would indeed manifest across the multilingual group. The results revealed existing differences between the groups, thus corroborating the hypothesis; however, the magnitude of the effect, as well as its presence, varied with respect to the language, place of articulation and voicing. In both series of stops, the most notable differences between the groups were attested in L2 English, in which multilinguals produced more Polish-like VOT, which was shorter in the voiceless series and prevoiced in the voiced one. This finding points to L1-driven influence on L2 English. In turn, in L1 Polish and L3 Norwegian, the voiceless stops were produced in a target-like manner, suggesting separate patterns of VOT realisation. In the voiced series, the L1 Polish velar stop /g/ was realised by the multilinguals with shorter prevoicing, which could indicate the influence of L2 English; however, as the probability of prevoicing realisation in this sound almost reached a ceiling effect, it might as well have been the effect of coarticulation or the rate of speech. Furthermore, Norwegian /d/ was found to have longer and more probable prevoicing, suggesting, again, a possible L1 Polish-driven influence. This effect might be attributed to the varying place of articulation of Norwegian /d/ (Kristoffersen, 2000), which could have been assimilated to the L1 Polish sound by the speakers. To sum up, the across-group comparisons showed a potential L1 Polish effect on both series of stops in L2 English as well as the L3 Norwegian voiced plosives, which confirms earlier findings from the research on CLI in L3 (e.g., Llama & Cardoso, 2018; Bandeira & Zimmer, 2012).

The second research question was devoted to the directionality of CLI in VOT production of stops across the three languages of L1 Polish—L2 English—L3 Norwegian speakers and its dependence on the place of articulation of stops and/or voicing. It was predicted that all three languages would interact, but the interaction would be mediated by a given language pair, the place of articulation and voicing series. In the voiceless stops,

 https://doi.org/10.3390/languages11090182

the participants managed to produce target-like L1 Polish stops, but their VOT was longer in L2 English and L3 Norwegian compared to the L1. Although in L3 Norwegian their pronunciation was close to a target, meaning it did not significantly differ from the control group, in English their VOT was significantly longer than in L1 Polish, but still too short to reach the target realisation of native English speakers. This across-language comparison adds up to the picture, creating the possibility of not only L1 Polish, but also L3 Norwegian influence on the L2 English realisations. To support this claim, we juxtaposed the current findings with those from a previous investigation examining stop production in speakers with the same linguistic repertoire (Cal & Wrembel, 2025). Cal and Wrembel (2025) provide a longitudinal perspective on VOT acquisition across three testing times, with the first one (T1) acting as a proxy for a control group condition by providing VOT data from L1 Polish—L2 English bilinguals at the very beginning of learning L3 Norwegian. The participants demonstrated relatively high VOT values in L2 English voiceless stops at the beginning of L3 learning at T1. Interestingly, the L2 English VOT durations increased even more with time (at T2 and T3), most likely as the result of intensive learning of L3 Norwegian, while the exposure to L2 English was limited. This effect may be interpreted as a cumulative effect of learning L3. Thus, we might assume a similar trend in this investigation. What is more, it is possible that the multilinguals noticed the similarity between L2 and L3 stops and therefore assimilated values in the two languages, or, conversely, that they are not able to reach the target realisation, because of being accustomed to short-lag stops in their L1. Both outcomes have been already observed in the literature (cf., e.g., Llama & Cardoso, 2018; Bandeira & Zimmer, 2012; Zając, 2015 for L1-driven CLI; Aoki & Nishihara, 2013 for L3-to-L2 CLI). The prevalence of CLI in L2 English might be caused by the lack of phonetic training in L2 English, as the multilinguals attended Norwegian pronunciation classes in their Norwegian studies programme, but did not receive any explicit pronunciation training in L2 English, which could have led to smaller awareness of English pronunciation rules. As far as the voiced series of stops is concerned, our results showed an extensive use of prevoicing in all three places of articulation and in all three languages, suggesting a target-like realisation of L1 Polish and L3 Norwegian and facilitative CLI between the two. However, in L2 English the prevoicing was significantly shorter and, what is more, the probability of its realisation was also significantly lower. This finding suggests that the participants may have subconsciously noticed the difference between voiced stops in English vs. the remaining two languages, but knowing two prevoicing languages and not having L2 pronunciation instruction might make it difficult for them to lose this feature completely. Interestingly, we also found that the participants are less likely to produce prevoicing in L3 Norwegian compared to L1 Polish, which might suggest a possible L2 English influence, which has been shown in some previous studies (e.g., Llama et al., 2010; Tremblay, 2007). The asymmetry between the two series of stops in terms of their susceptibility to CLI could be attributed to a greater perceptual salience of voiceless stops. It is easier to notice the difference between short-lag and long-lag stops than between prevoicing and lack thereof in the voiced series. Moreover, the participants received phonetic instruction only in L3 Norwegian, which distinguishes between prevoicing in voiced stops and aspiration in the voiceless ones, therefore no instruction was given on how to suppress the voicing. What is more, previous research (Zhang, 2022; Zhang et al., 2023) showed an opposite pattern, that is, the speakers experienced greater difficulty establishing a negative VOT category (i.e., prevoicing) than adjusting long-lag VOT to shorter positive VOT values. The production of prevoicing, thus, seems to be more subconscious and demanding, as it allows a specific articulatory gesture involving laryngeal voicing and an oral closure. From a phonological standpoint, prevoicing is also considered to be a phonetic detail, while aspiration is an easier-to-notice phonological feature (Schwartz et al., 2020). Summarising,

https://doi.org/10.3390/languages11090182

the results regarding RQ2 pointed to a possible effect of both L1 and L3 on L2 English in the two series of stops. Moreover, in voiced stops, the possibility of L2-driven CLI on L3 Norwegian was attested, especially in /b/ and /d/.

In the third research question we sought to establish any potential differences in VOT production between read and spontaneous speech and to test whether CLI patterns remain consistent across both speaking modes. The hypothesis posited that read speech would result in more careful, target-like, and less variable VOT productions, whereas spontaneous speech would be characterised by more variation and higher levels of CLI due to the nature of the task. We confirmed the prediction only partially, as we found that the participants managed to distinguish between short- and long-lag VOT in L1 Polish vs. L2 English and L3 Norwegian, which could suggest more target-like realisations of voiceless stops. As in the previous research question, relatively more CLI was attested in the voiced series of stops, compared to the voiceless counterparts. We observed high occurrence of prevoicing in all three languages even though L2 English turned out to be produced with significantly shorter voicing lag. This might result from a lower level of control that the participants have over their pronunciation while producing spontaneous speech, which was also shown by Jerotijević Tišma (2019). What is more, the multilinguals managed to produce voiceless stops close to the target, or sometimes with VOT longer than in controls. This outcome could have also been boosted by two foreign languages sharing the same feature. However, it is also important to notice that longer VOT values are found specifically in L3 Norwegian, which is the language that the participants know for the shortest time and largely at the lowest level. This could have led to the production of speech at a lower rate and, therefore, caused the elongation of VOTs. At the same time, control speakers speaking in their native languages could have a higher speech rate, leading to VOT compression in casual speech. Therefore, further research would be needed to investigate the influence of speech rate on VOT production in multilinguals.

The fourth research question explored the impact of extralinguistic factors on the presence and degree of CLI in the production of stops by trilinguals. Based on the NGTA framework, it was hypothesised that higher L2 and L3 proficiency, increased frequency of use, greater exposure, and a lower AoA in the L2 and L3 would all correlate with more target-like VOT productions and a reduction in CLI. The results of the Random Forest analysis provide partial support for these hypotheses. The model showed that the influence of extralinguistic factors is not uniform across the voicing series of stops. That is, for voiceless stops, a higher importance of language indicates that language-specific patterns of stop realisations are more strictly maintained in the voiceless category. However, the inclusion of L2 AoA and L3 proficiency in the top five predictors confirms that the target-likeness of aspiration is indeed modulated by the speaker's developmental history, as predicted by the NGTA. Interestingly, the findings also point to the high ranking of speech mode, which suggests that the realisation of voiceless stops is also task dependent. This was demonstrated in relation to RQ3, where the learners produced voiceless stops with longer VOT in the spontaneous task. The output from the LMM pointed only to the significant relation between speech mode and VOT production with longer VOTs found in spontaneous speech. However, this finding could have been masked by potential lower speech rate in the production of foreign languages. In contrast, the voiced stop model offers a different picture. As the relative weight of language dropped, the importance of extralinguistic variables increased. Specifically, L2 proficiency, L2 AoA and L3 frequency of use turned out to affect the VOT duration the most. This suggests that for voiced stops, which are often more prone to CLI, the speaker's individual linguistic profile becomes the dominant force in speech production. The fact that L3 frequency of use was a top-five predictor only for voiced stops suggests that L3 Norwegian, together with L1 Polish as

https://doi.org/10.3390/languages11090182

both are characterised by a voicing lead, may exert an effect on the realisation of voiced stops, potentially leading to increased CLI. These data also point to the dynamicity of the multilingual system specified in the NGTA (Dziubalska-Kołaczyk & Wrembel, 2022, 2024), which is modulated by a hierarchy of external factors, as discussed earlier. To synthesise how these variables prioritise themselves within the present study's context, the resulting hierarchy of external factors is visualised in Figure 13. Although these factors largely lack a significant linear relationship with VOT duration, the Random Forest analysis captured their subtle non-linear interactions across the global dataset, proving that individual variation in multilingual speakers still shapes the overall acquisition of VOT.

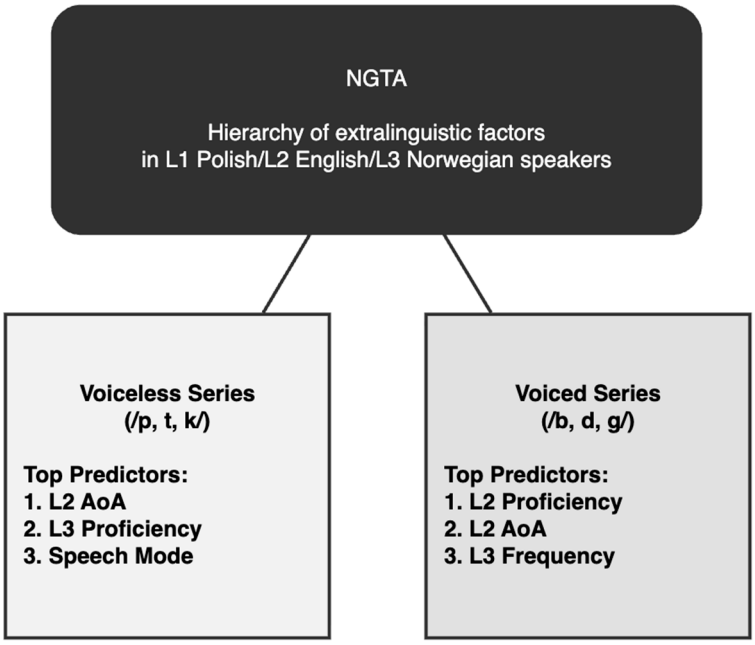


Figure 13. A proposed hierarchy of external factors for the "acquisition situation" in trilingual stop production.

This study has some possible limitations. First of all, it would benefit from a more balanced number of observations across the two speech modes, which would allow for more comprehensive comparisons and greater statistical power. The elicitation task for spontaneous speech could then be changed to, for example, a picture storytelling task targeting more precisely the words and sounds of interest. Both read and spontaneous speech measurements could also be supplemented by more extensive rate of speech comparisons. The pool of control speakers could be enriched with L1 Polish L2 English speakers, which could help isolate the source of CLI. What is more, a future cross-sectional study could extend the participant pool to include a greater number of participants matched for specific extralinguistic factors, such as level of proficiency, frequency of use, amount of exposure, among others. It could also look at individual variation among these speakers.

8. Conclusions

The aim of this study was to explore the production of stop consonants by trilingual speakers across two modalities of speech. What is more, the goal was to investigate potential cross-linguistic interactions across all three languages of L1 Polish, L2 English, L3 Norwegian speakers as well as selected extralinguistic factors that might condition VOT acquisition as well as the presence, absence and the amount of CLI. The investigation provides new insights into the research on CLI, as it offers a holistic approach to the dynamics and interactions between all three languages and places of articulation in two voicing se-

https://doi.org/10.3390/languages11090182

ries. Additionally, the study provides benchmark measurements taken from native control speakers of particular languages performing the same tasks as the experimental group, which allows for greater reliability of the comparisons. All the above-mentioned procedures and practices have been rarely applied in the field to date.

The results of the study demonstrated a great deal of influence stemming from L1 on L2 in both series of stops, which was further enhanced by a newly learnt L3. The findings also point to the asymmetry between the two series of stops in terms of their susceptibility to CLI. It was found that the voiced series is more prone to undergo assimilative CLI, while the voiceless stops are easier to acquire. What is more, it was shown that in spontaneous speech, multilinguals can produce voiceless stops with longer VOTs in all three languages, while they experience problems with suppressing prevoicing in the voiced series. The data also pointed to the dynamicity and complexity of the multilingual system in a specific acquisition situation, as specified in the NGTA (Dziubalska-Kołaczyk & Wrembel, 2022, 2024), which is modulated by a hierarchy of extralinguistic factors.

Supplementary Materials: Supplemental material for this article is available online through OSF: https://osf.io/2jucr/overview (accessed on 25 April 2026).

Funding: This research was funded by the National Science Centre grant number OPUS-19-HS 2020/37/B/HS2/00617.o.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee for Research Involving Human Participants at Adam Mickiewicz University in Poznań (resolution no. 12/2020/2021) on 5 July 2021.

Informed Consent Statement: Written informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The original data presented in the study are openly available in OSF at https://osf.io/2jucr/overview (accessed on 25 April 2026).

Conflicts of Interest: No conflicts of interest were declared. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

VOT	voice onset time
CLI	Cross-linguistic influence
NGTA	Natural Growth Theory of Acquisition
AoA	Age of acquisition

Note

[1] The OSF repository (Supplementary Materials) can be accessed through the following link: https://osf.io/2jucr/overview (accessed on 25 April 2026).

References

Amengual, M. (2021). The acoustic realization of language-specific phonological categories despite dynamic cross-linguistic influence in bilingual and trilingual speech. *Journal of the Acoustical Society of America*, *149*(2), 1271–1284. [CrossRef] [PubMed]

Amengual, M., Meredith, L., & Panelli, T. (2019, August 5–9). *Static and dynamic phonetic interactions in the L2 and L3 acquisition of Japanese velar voiceless stops*. Proceedings of the 19th International Congress of Phonetic Sciences (pp. 1312–1316), Melbourne, VIC, Australia.

Aoki, R., & Nishihara, F. (2013). Sound feature interference between two second languages: An expansion of the feature hypothesis to the multilingual situation in SLA. In C. Cathcart, I.-H. Chen, G. Finley, S. Kang, C. S. Sandy, & E. Stickles (Eds.), *Proceedings of the 37th annual meeting of the Berkeley linguistics society* (pp. 18–32). Berkeley Linguistics Society.

Bandeira, M. H. T., & Zimmer, M. C. (2012). The dynamics of interlanguage transfer of VOT patterns in multilingual children. *Linguagem and Ensino*, *15*(2), 341–364.

Baran, J. A., Laufer, M. Z., & Daniloff, R. (1977). Phonological contrastivity in conversation: A comparative study of voice onset time. *Journal of Phonetics*, *5*(4), 339–350. [CrossRef]

Bardel, C., & Falk, Y. (2007). The role of the second language in third language acquisition: The case of Germanic syntax. *Second Language Research*, *23*(4), 459–484. [CrossRef]

Bardel, C., & Falk, Y. (2012). The L2 status factor and the declarative/procedural distinction. In J. Cabrelli Amaro, S. Flynn, & J. Rothman (Eds.), *Third language acquisition in adulthood* (pp. 61–78). John Benjamins.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. [CrossRef]

Beckman, M. E. (1997). A typology of spontaneous speech. In Y. Sagisaka, N. Campbell, & N. Higuchi (Eds.), *Computing prosody: Computational models for processing spontaneous speech* (pp. 7–26). Springer.

Bless, M. (2015). *Interference in early acquisition Dutch-English bilinguals: A phonetic examination of voice onset time in Dutch and English bilabial plosives* [Unpublished master's thesis]. University of Amsterdam.

Boersma, P., & Weenink, D. (2023). *Praat: Doing phonetics by computer* (Version 6.3.14) [Computer software]. Available online: https://praat.org (accessed on 15 August 2023).

Bondarenko, M., Butera, B., & Rao, R. (2022). Voice onset time (VOT) of L3 Spanish/ptk/by multilingual heritage speakers of Ukrainian and Polish. *EuroAmerican Journal of Applied Linguistics and Languages*, *9*(1), 28–51. [CrossRef]

Cal, Z. (2026). *Crossing the Boundaries: Perception of Voice Onset Time in Word-Initial Stops by Multilingual Learners*. Language and Speech. [CrossRef] [PubMed]

Cal, Z., & Sypiańska, J. (2020). The interaction of L2 and L3 levels of proficiency in third language acquisition. *Poznan Studies in Contemporary Linguistics*, *56*(4), 577–603. [CrossRef]

Cal, Z., & Wrembel, M. (2025). Developmental acquisition of stops by multilingual speakers. *International Journal of Multilingualism*, *22*(1), 13–35. [CrossRef]

Dittmers, T., Gabriel, C., Krause, M., & Topal, S. (2017, September 28–29). *The production of voiceless stops in multilingual learners of English, French, and Russian: Positive transfer from the heritage languages?* 13th Conference on Phonetics and Phonology in the German-Speaking Countries, Berlin, Germany.

Docherty, G. J. (1992). *The timing of voicing in British English obstruents*. De Gruyter.

Donegan, P. J., & Stampe, D. (2009). Hypotheses of natural phonology. *Poznan Studies in Contemporary Linguistics*, *45*(1), 3–31. [CrossRef]

Dukiewicz, L., & Sawicka, I. (1995). *Gramatyka współczesnego języka polskiego: Fonetyka i fonologia* [Grammar of the contemporary Polish language: Phonetics and phonology] (H. Wróbel, Ed.). Wydawnictwo Instytutu Języka Polskiego PAN.

Dziubalska-Kołaczyk, K., & Wrembel, M. (2022). Natural growth theory of acquisition (NGTA): Evidence from (mor)phonotactics. In V. Sardegna, & A. Jarosz (Eds.), *Theoretical and practical developments in English speech assessment, research, and training* (pp. 281–298). Springer Nature.

Dziubalska-Kołaczyk, K., & Wrembel, M. (2024). A revised natural growth theory of acquisition: Evidence from L3 phonology. In E. Babatsouli (Ed.), *Multilingual acquisition and learning: An ecosystemic view to diversity*. John Benjamins.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–276). York Press.

Flege, J. E., & Bohn, O.-S. (2021). The revised Speech Learning Model (SLM-r). In R. Wayland (Ed.), *Second language speech learning: Theoretical and empirical progress* (pp. 3–83). Cambridge University Press. [CrossRef]

Flynn, S., Foley, C., & Vinnitskaya, I. (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. *International Journal of Multilingualism*, *1*(1), 3–16. [CrossRef]

Fujisaki, H. (1997). Prosody, models, and spontaneous speech. In Y. Sagisaka, N. Campbell, & N. Higuchi (Eds.), *Computing prosody: Computational models for processing spontaneous speech* (pp. 27–42). Springer.

Gabriel, C., Krause, M., & Dittmers, T. (2018). VOT production in multilingual learners of French as a foreign language: Cross-linguistic influence from the heritage languages Russian and Turkish. *Revue Française de Linguistique Appliquée*, *23*(1), 59–72. [CrossRef]

Geiss, M., Gumbsheimer, S., Lloyd-Smith, A., Schmid, S., & Kupisch, T. (2022). Voice onset time in multilingual speakers: Italian heritage speakers in Germany with L3 English. *Studies in Second Language Acquisition*, *44*(2), 435–459. [CrossRef]

Gósy, M. (2001). The VOT of the Hungarian voiceless plosives in words and in spontaneous speech. *International Journal of Speech Technology*, *4*, 75–85. [CrossRef]

Halvorsen, B. (1998). *Timing relations in Norwegian stops* [Unpublished doctoral dissertation]. University of Bergen.

Hammarberg, B., & Hammarberg, B. (2009). Re-setting the basis of articulation in the acquisition of new languages: A third language case study. In B. Hammarberg (Ed.), *Processes in third language acquisition* (pp. 74–85). Edinburgh University Press.

Jassem, W. (2003). Polish. *Journal of the International Phonetic Association*, *33*(1), 103–107. [CrossRef]

Jerotijević Tišma, D. M. (2019). The effect of speech style on VOT values in Serbian-English interphonology. *Srpski Jezik*, *24*(1), 337–355. [CrossRef]

Keating, P. A., Mikoś, M. J., & Ganong, W. F. (1981). A cross-language study of range of VOT in the perception of stop consonant voicing. *The Journal of the Acoustical Society of America*, *70*(5), 1260–1271. [CrossRef]

Kisler, T., Reichel, U. D., & Schiel, F. (2017). Multilingual processing of speech via web services. *Computer Speech & Language*, *45*, 326–347. [CrossRef]

Koch, K. A. (2008, May 31–June 2). *Spontaneous speech, lab speech, and effects on intonation: Some useful findings for fieldworkers (and laboratory phonologists)*. Proceedings of the 2008 Annual Conference of the Canadian Linguistic Association, Vancouver, BC, Canada.

Kretzschmar, W. A., Jr. (2015). *Language and complex systems*. Cambridge University Press. [CrossRef]

Kristoffersen, G. (2000). *The phonology of Norwegian*. Oxford University Press.

Labov, W. (1972). The isolation of contextual styles. In W. Labov (Ed.), *Sociolinguistic patterns* (pp. 70–109). Blackwell.

Lenth, R., & Piaskowski, J. (2026). *emmeans: Estimated marginal means, aka least-squares means* (R package version 2.0.5). Available online: https://rvlenth.github.io/emmeans/ (accessed on 15 August 2023).

Li, P., Zhang, F., Yu, A., & Zhao, X. (2020). Language History Questionnaire (LHQ3): An enhanced tool for assessing multilingual experience. *Bilingualism: Language and Cognition*, *23*(5), 938–944. [CrossRef]

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News*, *2*(3), 18–22.

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops. *Word*, *20*(3), 384–422. [CrossRef]

Liu, Z. (2016). *Exploring cross-linguistic influence: Perception and production of L1, L2 and L3 bilabial stops by Mandarin Chinese speakers* [Unpublished master's thesis]. Universitat Autònoma de Barcelona.

Llama, R., & Cardoso, W. (2018). Revisiting (non-)native influence in VOT production: Insights from advanced L3 Spanish. *Languages*, *3*(3), 30. [CrossRef]

Llama, R., Cardoso, W., & Collins, L. (2010). The influence of language distance and language status on the acquisition of L3 phonology. *International Journal of Multilingualism*, *7*(1), 39–57. [CrossRef]

Llama, R., & López-Morelos, L. P. (2016). VOT production by Spanish heritage speakers in a trilingual context. *International Journal of Multilingualism*, *13*(4), 444–458. [CrossRef]

Mitchell, M. (2009). *Complexity: A guided tour*. Oxford University Press.

Nakamura, M., Iwano, K., & Furui, S. (2008). Differences between acoustic characteristics of spontaneous and read speech and their effects on speech recognition performance. *Computer Speech & Language*, *22*(2), 171–184. [CrossRef]

Nelson, C. (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. [CrossRef]

Posit Team. (2025). *RStudio: Integrated development environment for R* (Version 2026.01.0+392) [Computer software]. Available online: http://www.posit.co/ (accessed on 10 January 2026).

Ringen, C., & van Dommelen, W. A. (2013). Quantity and laryngeal contrasts in Norwegian. *Journal of Phonetics*, *41*(6), 479–490. [CrossRef]

Roach, P. (2009). *English phonetics and phonology: A practical course* (4th ed.). Cambridge University Press.

Rothman, J. (2011). L3 syntactic transfer selectivity and typological determinacy: The Typological Primacy Model. *Second Language Research*, *27*(1), 107–127. [CrossRef]

Rothman, J. (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. [CrossRef]

Schwartz, G., Wojtkowiak, E., & Dzierla, J. (2020). Laryngeal phonology and asymmetrical cross-language phonetic influence. In M. Wrembel, A. Kiełkiewicz-Janowiak, & P. Gąsiorowski (Eds.), *Approaches to the study of sound structure and speech: Interdisciplinary work in honour of Katarzyna Dziubalska-Kołaczyk* (pp. 316–326). Routledge.

Slabakova, R. (2017). The scalpel model of third language acquisition. *International Journal of Bilingualism*, *21*(6), 651–665. [CrossRef]

Stampe, D. (1979). *A dissertation on natural phonology*. Garland.

Stuart-Smith, J., Sonderegger, M., Rathcke, T., & Macdonald, R. (2015). The private life of stops: VOT in a real-time corpus of spontaneous Glaswegian. *Laboratory Phonology*, *6*(3–4), 505–549. [CrossRef]

Swerts, M., Strangert, E., & Heldner, M. (1996). F0 declination in read-aloud and spontaneous speech. *Proceedings of the 4th International Conference on Spoken Language Processing (ICSLP)*, *3*, 1501–1504. [CrossRef]

Sypiańska, J. (2013). *Quantity and quality of language use and L1 attrition of Polish due to L2 Danish and L3 English* [Unpublished doctoral dissertation]. Adam Mickiewicz University.

Sypiańska, J. (2017). *Cross-linguistic influence in bilinguals and multilinguals*. Wydawnictwo Naukowe UAM.

Sypiańska, J. (2021). Production of voice onset time (VOT) by senior Polish learners of English. *Open Linguistics*, *7*(1), 316–330. [CrossRef]

Sypiańska, J., & Cal, Z. (2020). The influence of level of proficiency in the L2 and L3 on the production of the L3 Spanish apico-alveolar sibilant. *Research in Language*, *18*(1), 89–107. [CrossRef]

Syrdal, A. K. (1996). Acoustic variability in spontaneous conversational speech of American English talkers. In *Proceedings of the 4th international conference on spoken language processing (ICSLP)* (pp. 438–441). IEEE. [CrossRef]

Tremblay, M. C. (2007, August 6–10). *L2 influence on L3 pronunciation: Native-like VOT in the L3 Japanese of English-French bilinguals* [Paper presentation]. The Satellite Workshop of ICPhS XVI, Freiburg, Germany.

van Dommelen, W. A. (2006). Quantification of speech rhythm in Norwegian as a second language. *Lund University Department of Linguistics and Phonetics Working Papers*, *52*, 33–36.

Westergaard, M., Mitrofanova, N., Mykhaylyk, R., & Rodina, Y. (2017). Crosslinguistic influence in the acquisition of a third language: The Linguistic Proximity Model. *International Journal of Bilingualism*, *21*(6), 666–682. [CrossRef]

Wojtkowiak, E. (2022). *L1 phonetic drift in the speech of Polish learners of English: Phonological implications* [Unpublished doctoral dissertation]. Adam Mickiewicz University.

Wrembel, M. (2010). L2-accented speech in L3 production. *International Journal of Multilingualism*, *7*(1), 75–90. [CrossRef]

Wrembel, M. (2011). Cross-linguistic influence in third language acquisition of voice onset time. In W.-S. Lee, & E. Zee (Eds.), *Proceedings of the 17th international congress of phonetic sciences* (pp. 2157–2160). City University of Hong Kong.

Wrembel, M. (2014). VOT patterns in the acquisition of third language phonology. *Concordia Working Papers in Applied Linguistics*, *5*, 751–771.

Wrembel, M. (2015). *In search of a new perspective: Cross-linguistic influence in the acquisition of third language phonology*. Wydawnictwo Naukowe UAM.

Wunder, E. M. (2010). Phonological cross-linguistic influence in third or additional language acquisition. In M. Wrembel, M. Kul, & K. Dziubalska-Kołaczyk (Eds.), *Achievements and perspectives in the acquisition of second language speech: Proceedings of the 6th international conference on native and non-native phonology* (New Sounds 2010, pp. 566–571). Peter Lang.

Yao, Y. (2009). Understanding VOT variation in spontaneous speech. *UC Berkeley PhonLab Annual Report*, *5*, 312–335. [CrossRef]

Zając, M. (2015). *Phonetic convergence in the speech of Polish learners of English* [Unpublished doctoral dissertation]. University of Łódź.

Zhang, L. (2022). ¿Pollo o Bollo? Acquisition of voicing contrast by Chinese learners of L3 Spanish. *Hispania*, *105*(4), 595–613. [CrossRef]

Zhang, L., Morales-Front, A., & Sanz, C. (2023). Dynamics in cross-linguistic influence: Evidence from VOT development of Mandarin-English bilingual learners of L3 Spanish. In M. Brown, S. Flynn, & É. Fernández-Berkes (Eds.), *L3 development after the initial state* (pp. 121–149). John Benjamins. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.