Beyond transfer?
The acquisition of an L3 phonology by Turkish-German bilinguals

Inaugural-Dissertation

zur Erlangung der Doktorwürde der
Philosophischen Fakultät der
Julius-Maximilians-Universität Würzburg

Vorgelegt von
Christina Domene Moreno
aus Würzburg

Würzburg
2021

Erstgutachter:
Prof. Dr. Barış Kabak

Zweitgutachterin:
Prof. Dr. Tanja Kupisch

Tag des Kolloquiums:
26.06.2020

Deutsche Zusammenfassung

Der Erwerb des phonologischen Systems einer Drittsprache wurde bisher vor allem vor dem Hintergrund zweier Fragen betrachtet: (a) Welche Sprache oder Sprachen werden in die Zielsprache transferiert und (b) welche Rolle spielen inner- und außersprachliche Faktoren dabei? Hier sind grundsätzlich drei Szenarien denkbar, welche sich auch in den im Forschungsfeld vorherrschenden theoretischen Ansätzen wiederfinden. Bei einem bilingualen Sprecher, also einem Sprecher mit zwei Hintergrundsprachen, ist erstens der ausschließliche Transfer einer der Hintergrundsprachen möglich, wobei für die Vorhersage der Sprache, die am wahrscheinlichsten transferiert wird, verschiedene Parameter vorgeschlagen wurden, wie beispielsweise das typologische Verhältnis der Hintergrund- und Zielsprachen untereinander oder der Status einer der beiden Hintergrundsprachen als Erst- bzw. Zweitsprache. Weiterhin ist Transfer aus beiden Sprachen denkbar, bei dem ausschließlich „nützliche", das heißt den Zielspracherwerb unterstützende Strukturen transferiert werden. Drittens kann gemischter Transfer, also der Transfer verschiedener Strukturen, aus beiden Hintergrundsprachen angenommen werden, der allerdings nicht unbedingt positiv ist, sondern stattdessen von vielfältigen Faktoren konditioniert wird. In anderen Worten wird hier die Möglichkeit angenommen, Strukturen aus beiden Hintergrundsprachen in die Zielsprache zu transferieren. Dieser letzte Ansatz entspricht am besten den bisherigen Forschungsergebnissen des Felds: Während sich bei Betrachtung verschiedener Sprachkombinationen und verschiedener (und verschiedenartiger) phonologischer Strukturen für sich genommen kein einheitliches Muster erkennen lässt, wird deutlich, dass gemischter Transfer möglich ist – dies kommt vor allem bei Studien zu graduellen phonetischen Phänomenen wie konkreter Vokalqualität oder Voice Onset Time zum Ausdruck. Bisherige Studien haben sich auf einzelne phonologische oder phonetische Phänomene beschränkt, anhand derer die entsprechenden Ansätze überprüft wurden. Bisher gibt es keine systematischen Untersuchungen mehrerer verschiedener Phänomene an denselben Sprechern. Da allerdings davon ausgegangen werden kann, dass phonologische Phänomene unterschiedliche strukturelle Eigenschaften haben, ist genau dies nötig. Um also die Mechanismen des gleichzeitigen Transfers mehrerer Sprachen auf eine Zielsprache, und somit auch die Interaktion der Hintergrundsprachen in Multilingualen besser zu verstehen, müssen mehrere phonologische Strukturen an denselben Sprechergruppen und mit denselben Methoden untersucht werden. Diese Forschungslücke will die vorliegende Arbeit schließen.

Hierfür wurde der Erwerb des Lautsystems der Drittsprache Englisch durch türkisch-deutsch bilinguale Lerner untersucht. Es wurden phonologische Strukturen des Englischen zur Untersuchung ausgewählt, die entweder im Deutschen (aber nicht im Türkischen), im Türkischen (aber nicht im Deutschen), oder in keiner der beiden Hintergrundsprachen vorkommen. Diese Unterteilung sollte dazu dienen den potentiellen Einfluss der jeweiligen Hintergrundsprache auf die Zielsprache identifizieren zu können.

In zwei Studien wurden jeweils die Perzeption und die Produktion der obengenannten Strukturen an türkisch-deutsch Bilingualen sowie an einer monolingual deutschen Kontrollgruppe getestet. Studie 1 testete junge Lerner des Englischen in den ersten Lernjahren, während Studie 2 Studierende der Englischen Sprach- und Literaturwissenschaft testete. Zusätzlich wurden extralinguistische Faktoren und biographische Details erhoben. In den beiden Studien sollte beantwortet werden, ob sich bilinguale und monolinguale Testpersonen in der Perzeption und/oder der Produktion der Zielstrukturen unterscheiden, ob diese Unterschiede durch Transfer aus den Hintergrundsprachen zu erklären sind und, falls dies nicht durchgehend der Fall ist, welche anderen strukturellen oder extralinguistischen Faktoren herangezogen werden können, um die Ergebnisse zu erklären. Zudem sollte geprüft werden, inwieweit Perzeption und Produktion zusammenhängen und welchen Effekt fortschreitender Spracherwerb hat.

Die Ergebnisse der beiden Studien dieser Arbeit bestätigen die Vermutung, die anhand der Kombination aus verschiedenen vorausgehenden Studien getroffen wurden: Transfer findet ausgehend von beiden Hintergrundsprachen statt, er kann positiv und negativ sein, und er wird von strukturellen Faktoren konditioniert. So zeigte sich im Vergleich ein deutlicher Einfluss – sowohl positiv wie auch negativ – des Türkischen auf die Perzeption und Produktion des Englischen bei den bilingualen Sprechern, aber ebenso ein Einfluss des Deutschen. Weiterhin konnte der gleichzeitige Einfluss beider Hintergrundsprachen auf einzelne Strukturen belegt werden. Perzeption und Produktion stehen bei keiner der Sprechergruppen in einer direkten Korrelation, allerdings kann anhand der Ergebnisse Perzeption als einer der Faktoren angenommen werden, der einen Einfluss auf die Produktion der zielsprachlichen Struktur hat. Als weitere Faktoren konnten der Grad der Markiertheit der Struktur, ihre artikulatorische Komplexität sowie das Fehlen eines artikulatorischen Ankers identifiziert werden. Eine Verschiebung der Transferquellen fand zwischen den beiden Studien, also bedingt durch fortgeschrittenen Erwerb, nicht statt. Es konnte allerdings gezeigt werden, dass Nachteile, die aufgrund der Hintergrundsprachen den Sprechergruppen entstanden, in der älteren Gruppe häufig ausgeglichen werden konnten.

Die Erklärungsansätze für das Autreten von CLI-Effekten waren abhängig von den Eigenschaften der einzelnen abgeprüften Strukturen und konnten nicht systematisch verallgemeinert werden. Dies bestätigte die wenig einheitlichen Einzelergebnisse vorhergehender Studien für die Perzeption und Produktion mehrerer verschiedener phonologischer Strukturen durch dieselben Sprechergruppen. So konnten methodische oder gruppeninterne Gründe für die sich widersprechenden Ergebnisse erstmals ausgeschlossen werden. Im vorletzten Kapitel dieser Arbeit werden darauf aufbauend alternative Gründe für die Ergebnisse erarbeitet. Hierfür wird das Konzept des komplexen verschachtelten (*complex interlaced*) eingeführt, das chaostheoretische Ansätze mit der Annahme einer hierarchischen Bit-Struktur verbindet.

Die vorliegende Arbeit trägt zu aktuellen Debatten im Bereich der Drittsprachforschung insofern bei, als dass sie als erste mehrere phonologische „Einheiten" (bits), also Phoneme, Allophone, Prozesse etc., an den gleichen Sprechern testet. Dadurch wurde deutlich, dass die konkrete Struktur phonologischer Systeme in den Mittelpunkt der Forschung zum Drittspracherwerb rücken muss, damit vorhandene Muster erkennbar werden können. Zudem wird vorgeschlagen, verstärkt auf Forschungsergebnisse und Theorien aus anderen Bereichen, wie beispielsweise dem Sprachwandel, der Varietätenlinguistik und der Kontaktlinguistik zurückzugreifen, um letztendlich Forschung am Drittspracherwerb für den theoretischen Erkenntnisgewinn nutzbar zu machen.

For my mother.

Acknowledgements

It is done.

My sincerest thanks go to my first supervisor Barış Kabak, who has been an inspiring teacher, great mentor, and wonderful colleague and friend. Without your guidance and support, and your fierce belief in me and my academic abilities, I highly doubt I would ever have taken the path that led me here. Thank you for introducing me to the world of academia in general and linguistics in particular, for providing feedback and advice whenever I needed it, for your extensive and productive comments on my dissertation drafts, and especially for much needed last-minute support. Thank you for teaching me to question myself every step of the way, for showing me how to ask questions and how to find answers. And, finally, thank you for all the basement discussions about linguistics, about music, about food, and about life in general, for tea-making and sushi-eating.

I would also like to thank Tanja Kupisch, my second supervisor, for agreeing to be on my committee rather late "in the game". Thank you for your valuable advice and fruitful feedback (I am in serious awe of the incredible speed with which you replied to every single one of my emails) and input at various conferences and for cheering me on. I also owe my deepest gratitude to Carolin Biewer, who has given me the opportunity to work at her department throughout my PhD process: Thank you for your unwavering trust, for fighting for me and for all of us, and thank you for being a collaborator rather than a rival in this harsh reality that academia can be.

I've been extremely fortunate to work in a team of wonderfully and excitingly crazy people, with whom I could be myself in the most entertaining ways. I would not have been able to complete this dissertation without our adventures and without you: Lisa – thank you for always lending an ear when I needed to rant, thank you for proofreading the entire dissertation (some parts more than once…), for discussing, for listening, and for reality-checking, and for doing an incredible job at helping me stay sane during the last round of editing and formatting. Ninja – thank you for your invaluable feedback throughout, for your encouragement, for your comments on my drafts and on my thoughts, and for all those weekend coffees, research dinners, and Old English sessions, especially in the last few weeks of my writing process. Having you around every day to do fun stuff and also to listen to all my weird thoughts and doubts was truly invaluable. Patrick – thank you for your friendship. You are one of the few people whose presence I could deal with – and was indeed thankful for – even when I was in hyper-data-dissertation-defense-craziness mode. Reasons for thanking you include but are by far not limited to your excellent, entertaining, and enormously encouraging comments on my dissertation drafts (and on life), and agreeing to discuss my ideas with me again and again and again and again. Marie – thank you for being the "big sister in the basement", for teaching me about statistics in general and R in particular, and for all those sushis and adventures we shared in the earlier stages of this endeavor. Thank you also for drawing the pictures for my first study. Raphael – I'm sure it can't have been easy to share an office with me during much of the last stage of my dissertation journey, but I'd like to thank you for making this time fun, for filling the office with even more dinosaurs, laryngeeeees, and instant coffee, and for sharing your research with me and letting me (over)share mine with you. Thank you also for proofreading particularly the more technical parts of this dissertation. I can't wait to collaborate on our future fun research projects! Zeyu – thank you for being part of this journey. You are one of the people I can be most creative with, and you are just an excellent writing buddy. Thanks for cheering me on at any time and for being the brilliant researcher and wonderful friend that you are. I would also like to thank Marlene Keßler, Emma Stier, Janne Lorenzen, Regina Grund as well as the wonderful MA students I've had the opportunity to work with for inspiring me and for countless chats and coffees, as well as Elisabeth Reber, Elisabeth Chowanetz, Kathrin Lang, Hümeyra Uzunkaya, and Matthias Krebs for valuable discussions about my dissertation as well as the fun times we've had. I owe big thank-yous to Karin Kernahan and Sonja Kuhn as well as Fabienne Dietl and Christina Schäfer, who have helped me through the paperwork and bureaucracy of this endeavor and who have always been up for an office chat, which oftentimes has made working much more fun. Thanks also to everyone else I have been fortunate enough to share my workplace with. I would particularly like to mention Zeno Ackermann and Sandra Ellena, who – in addition to just being great people to be around – have provided constant cheers, encouragement, valuable advice, and smiles. Sandra, thank you also for agreeing to be the third examiner at my defense.

Thanks to all those wonderful people I've met at conferences and workshops. Please forgive me for only listing some of you here: Ulrike Gut, Romana Kopeckova, Kristy Sigmeth, Christina Golin, Magdalena Wrembel, Halina Lewandowska, Iga Krzysik, Annika Lloyd-Smith, Jennifer Cabrelli, Thorsten Piske, and everybody in the L3 crowd for all the invaluable feedback, discussions, and – yes – fun times!

I could never have written this dissertation without my young participants in the schools I visited as well as the students at Würzburg University for patiently doing all the tasks I gave them. I would particularly like to thank Gülaziz Kizilay at Hulda-Pankok-Gesamtschule Düsseldorf for agreeing to let me collect data at his school and for taking care of all the logistics involved. Special thanks also go to Konrad Maiwald who organized two full days of testing at St.Ursula-Schule in Duderstad. He and Gwynne Maiwald also generously opened their doors and welcomed me into their home for those two days. Additional thanks go to Gudrun Reiners from Gustav-Walle-Schule in Würzburg for organizing pilot testing sessions with some of her students.

I owe the deepest thanks to my wonderful friends and family. I know that I have been absent a lot during these last years, and I would like to thank you all for your patience, love, and support. To my wonderful friend Robert – thank you for always being there, for discussing, for cheering, for providing a different angle, for traveling, for being silly and for being serious, for believing in me without end and for simply being the stable ground in my life. Without you, this project would have been much greyer. Thank you also for recording my audio stimuli and for always being ready to provide technical feedback and assistance with whatever I was doing. Larissa – thanks for being the beautiful and generous person that you are, for being curious about my work at all times and for pushing me forward by asking "out-of-the-box" questions. Thank you also for patiently lending your voice for my audio stimuli. Stefan, Raimund, Claudia – thank you for pulling me away from my desk on so many occasions and for being so much fun to be with. Our adventures have always given me the energy and motivation to go back to work and finish it too. Julia – thank you for being who you are: My world is so much more colorful with you in it! Learning Arabic while working on my dissertation has often kept me sane. This is why I would like to thank all my Arabic teachers and especially Gamiela Safiya for being an incredible teacher and for countless conversations about languages and multilingualism.

An meinen Vater: Ohne deine Unterstützung, deinen Glauben und dein Vertrauen in mich hätte ich dieses Projekt nicht stemmen können. Dafür bin ich dir unendlich dankbar. And to my mum: I am eternally sad that you are not here with me to share all this. Thank you for making me the person I am today. Thank you for your love, your beauty, and your encouragement. I miss you so much every day. This is for you.

Table of contents

List of figures

List of tables

List of Abbreviations

AIC	Akaike information criterion
AMTB	Attitude/Motivation Test Battery
BA	Bilingual advantage
BLP	Bilingual Language Profile
CBT	Combined beneficial transfer
CC	Consonant cluster
CEFR	Common European Framework of Reference for Language
CEM	Cumulative Enhancement Model
CLI	Cross-linguistic influence
CPH	Critical period hypothesis
CTOPP-2	Comprehensive Test of Phonological Processing
FT	Feature type
fta	Full trial accuracy
GA	General American
HL	Heritage language
HS	Heritage speaker
ISI	Inter-stimulus interval
L1	First language
L2	Second language
L2SFM	L2 Status Factor Model
L3	Third language
LPC	Linear predictive coding
LPM	Linguistic Proximity Model
PAM	Perceptual Assimilation Model

PG	Phonological grammar
PWM	Phonological working memory
RP	Received Pronunciation
SDT	Signal detection theory
SLA	Second language acquisition
SLM	Speech Learning Model
SLT	Single language transfer
SM	Scalpel Model
SPM	Stratal Permutation Model
TL	Target language
TPM	Typological Primacy Model
UG	Universal grammar
VOT	Voice onset time

1 Introduction

1.1 Third language (L3) phonology – a not so new field of study

The last decade has seen numerous papers and volumes published on matters in L3 phonology, meetings of researchers both at specific workshops, and during special sessions at large conferences, as well as large-scale research projects. This has led to a pool of empirical and theoretical work conducted by an ever-growing network of researchers in the field – circumstances which have paved the way for interesting new questions that are formulated to address the complexity of multilingual language acquisition as well as the theoretical demands of phonological acquisition. L3 phonology, thus, is now far from being the "understudied domain" it was a decade ago. Increasingly, theories, methods, and findings from neighboring fields are taken into consideration, such as cognitive linguistics (e.g., Krzysik & Wrembel, 2019), multilingualism research (e.g., Kopečková, 2019), and L2 phonology (e.g., Wrembel, Marecka, & Kopečková, 2019). This (re-)connects L3 research to areas that the subject has, in fact, always been part of, which allows a shift in focus from a very centralized micro perspective with the acquisition of an L3 phonology at its center towards a more holistic macro perspective that regards the acquisition of a third language as a tool to learn more about general mechanisms of language acquisition, the structure and, more broadly, the nature of language.

Following Ortega (2013), the field second language acquisition (SLA) is concerned with "the human capacity to learn languages other than the first, during late childhood, adolescence or adulthood, and once the first language or languages have been acquired." (Ortega, 2013, pp. 1–2) In this vein, the study of L3 acquisition can be considered to be not so much research in a completely distinct field but rather a methodological utility that provides opportunities to detect phenomena in SLA that do not usually surface when looking at only two language systems. More specifically, investigating how two (or more) background systems surface in the L3 can enable us to observe the way language systems interact in multilingual speakers under varying conditions. In other words, the interaction of language-specific and universal factors in the learner's mind can potentially become visible through the output in the target language. Thus, studying the way an L3 is acquired may open a window into the processes of language acquisition as well as multilingual language use.

While different methods and approaches have been used in the field, some relatively robust and consistent trends have emerged as research progressed:

Trends in L3 research

(a) The most prominent question asked has been one about the source of transfer into the L3 (Cabrelli Amaro & Iverson, 2018, p. 738) as well as the source(s) and directionality of lateral transfer, a term coined by Jarvis and Pavlenko (2010) to denote transfer from any available language onto any other available language in the multilingual system (see, for instance, Wunder, 2015). In connection to this, identifying the factors that condition the source of transfer has been an additional question.
(b) To gain insights into the matter and answer these questions, L3 learners have either been compared to their monolingual peers, or to other bilingual groups. Here, a mirror-image setup has often been used to tease apart the influence of structural and personal factors.
(c) The subjects studied in most studies have been consecutive bilinguals, i.e., learners of an L3 with a clearly identifiable second language (L2) that was acquired after the critical period.
(d) With few exceptions, studies have used cross-sectional designs.
(e) Most studies have either used very holistic diagnostics like foreign accentedness as a proxy for language transfer or provided detailed close-ups of the phonological system by looking at single features in relative isolation.[1]

[1] In fact, various studies report results on more than one phonological feature, but, even in those cases, results for these individual features are rarely seen in connection or relation to each other.

Crucially, neither of the diagnostics in (e) can give insights into the complexity of the multilingual system and thus on multilingual language acquisition on their own. This can probably be remedied either by using both methods and then relating and combining their results. However, another possible approach is one that examines larger portions of the phonological grammar while at the same time focussing on the individual entities in that larger portion. This latter solution is the one taken in this dissertation.

In order to answer questions about transfer behavior in multilingual learners, German monolingual and Turkish-German bilingual young learners (in Study 1) and adult speakers (in Study 2) of English are tested on the perception and production of a set of phonological features. It is crucial to test both perception and production to uncover a potential psychoacoustic effect on non-native speech production and to gain further insight into the general interaction between speech perception and speech production. Additionally, extra-linguistic variables are assessed to identify potential predictors of the degree of success in both domains. The studies ultimately aim at answering (a) whether there is a difference between monolinguals and bilinguals in their perception and/or their production of English phonological contrasts that are acquired as an additional language by both the bilinguals and the monolinguals, (b) whether potential differences can be explained by cross-linguistic influence (CLI) from Turkish, (c) whether any of the extra-linguistic factors predict perception or production, and (c) how transfer scenarios differ across age groups (young vs. old).

The data is analyzed using a bit-by-bit approach, an approach that considers phonological entities like features, properties, or processes, i.e., bits, as the domain of CLI rather than full phonological grammars. Finally, the results of the empirical studies in this dissertation and the previous literature are accounted for in the light of interlaced complex systems.

1.2 Structure of this dissertation

This dissertation consists of 7 chapters, which are organized as follows: Chapter 2 introduces the most prevalent and well-established models in L2 phonology, Flege's *Speech Learning Model* (*SLM*, Flege, 1987) and Best's *Perceptual Assimilation Model* (*PAM*, Best, McRoberts, & Goodell, 2001), discusses conceptual and practical differences between the study of second language vs. additional language acquisition and defines the terminology used throughout this dissertation. Afterward, an overview is given of the research that has been conducted in L3 phonology, which can roughly be sorted into three categories: (a) Early studies that focused on exploring the most prevalent errors made by speakers or learners of an L3 using a mostly qualitative approach, (b) studies concerned with foreign accentedness in multilinguals, and (c) the production and perception of discreet phonological features by L3 speakers. The chapter ends by arguing for the necessity to study multiple features at a time in order to tease apart linguistic, feature-specific as well as typological effects, and by highlighting relevant previous studies.

Chapter 3 introduces the phonological features tested in the empirical part of this dissertation by first giving a brief overview of the phonological systems of Turkish, German, and English, the three languages under scrutiny. Next, the choice of phonological features, grouped according to their presence or absence in the three languages, is motivated, and their linguistic characteristics are briefly explained. Lastly, the chapter provides an explication of the general research objectives taking into consideration the empirical research and models reviewed in Chapter 2.2.

Chapters 4 and 5 contain the two empirical studies conducted for this dissertation. Study 1 (Chapter 4) tests the perception and production of the selected phonological features in young monolingual and young bilingual learners of English, whereas Study 2 (Chapter 5) tests the same phenomena in adult monolingual and bilingual highly proficient users of English.

Finally, Chapter 6 discusses and interprets the results of both studies in light of previous findings. Based on the individual discussions in Chapters 4 and 5, a new approach to L3 phonology is proposed that shifts the main focus away from L3 acquisition and puts phonology in the center instead. After this, a framework is suggested that combines the notion of phonological grammars as complex interlaced systems and a bit-by-bit approach to CLI. Chapter 7 summarizes and concludes the dissertation.

2 The phonological acquisition of an additional language

This chapter first presents some prerequisites by detailing two of the most prevalent models of Second Language Phonology, introducing the concepts of CLI and transfer, and presenting some terminological discussions and choices.[2] Section 2.2 discusses the research conducted in the field of L3 phonology and presents the prevailing models in the field.

2.1 Prerequisites

2.1.1 Speech Learning Model and Perceptual Assimilation Model

The fact that an L2 phonology is influenced by an First Language (L1) phonology has long been known. Perhaps one of the most prominent manifestations of this influence is perceivable foreign accent due to non-native-like production on the segmental and suprasegmental level. Different factors have been identified that determine the degree to which the L1 exerts an influence on the L2, among them cognitive, individual, social, and linguistic factors, that I will refer to throughout this dissertation. Concerning linguistic factors, two widely tested models in the study of Second Language Phonology are, as already mentioned in Chapter 1, the Speech Learning Model (SLM, Flege, 1987) and the Perceptual Assimilation Model (PAM-L2[3], Best & Tyler, 2007). While not specifically tested in the studies presented in this dissertation, knowledge of the basic processes described in both models can be used to

[2] Since terms are often used with different meanings by different authors, I adjust the terminology accordingly (i.e., I use the terminology laid out in this chapter even if it means using different terms from those used by the authors of individual studies) in order to achieve a degree of consistency. This is done for all terms defined in the course of this chapter.

[3] The original model by Best (1995) studied the perception of non-native speech sounds by naive listeners. Best and Tyler (2007) slightly modified this original model to make predictions about L2 speakers' perception of non-native speech sounds.

interpret the findings related to transfer in studies on additional language acquisition in general and the present study in particular. Both models are concerned with the language-internal factors that determine the degree of L1 on L2 influence. Specifically, both models use (degrees) of differences and similarities, both phonetically and phonologically between the L1 and the target language (TL) to make predictions about the difficulty a learner will encounter when faced with a TL sound (Best & Tyler, 2007, p. 22). The researchers in both pieces of work prioritize speech perception over production, and while they acknowledge that the relationship between L2 perception and L2 production is not straightforward, they also assume some type of connection between the two. Flege (1991, p. 264), for example, accepts that in "speech learning, perceptual representations derived from experience with phones in the ambient language guide development of the motor programs needed to implement phonemic categories" (Flege, 1991, p. 264).

Flege (1995, p. 239) assumes phonetic categories are formed in the acquisition of an L1 and that the mechanisms used for category formation remain active and are also used for non-native language acquisition. Both L1 and L2 categories exist in one common linguistic space, and a multilingual speaker aims to keep apart L1 and L2 categories in that space. As a consequence, L1 categories remain flexible across the lifespan since they may be modified either to incorporate new L2 sounds or to allow new L2 categories to be formed. Based on his empirical work in which he tested non-native speakers on various potentially difficult TL contrasts, Flege developed a number of hypotheses to predict the likelihood of a new L2 category being formed in acquisition as opposed to an incoming L2 phone being categorized with an existing L1 category: A new category can only be formed if the learner perceives a phonetic difference between the L1 sound and the L2 sound, although the cues for this differentiation do not necessary overlap with the cues used by native speakers of that L2. Furthermore, he hypothesized that category formation is more likely, the more distinct the L2 sound is from the closest equivalent L1 sound. If the L2 sound is perceived as very similar to an equivalent L1 sound, it will be categorized in perception and production with the L1 sound. Flege calls this process *equivalence classification*. Even in cases where the learner forms a new category, this category may differ from the same category in a speaker of the same language as an L1. This happens when the learner needs to distinguish the newly formed category from an existing one (in which case the category will have to shift) or when different phonetic cues are used by the L1 speaker and the L2 learner. This set of hypotheses leads to several concrete predictions for L2 speech sounds. L2 speech sounds are assumed to be either *new sounds* or *similar sounds* (Flege, 1991, p. 251). New sounds are the ones that do not have an equivalent sound in the L1, while similar sounds do. Thus, based on Flege's model, new sounds are more likely to form a new phonetic category than similar sounds, which, in turn, means that they are acquired more easily and "produced more authentically" (Flege, 1991, p. 267) by the learner.[4]

While the PAM-L2, like the SLM, assumes cross-linguistic similarity to be a crucial predictor for the successful perception of TL sounds, the pivotal parameter in PAM-L2 is articulatory rather than acoustic similarity (Best & Tyler, 2007, p. 22). Furthermore, PAM-L2 takes into consideration not only phonetic but also functional equivalence of speech sounds. In the model, contrasting L2 categories are categorized according to their relationship with L1 categories (Best, 1995, p. 195): Two L2 categories may be assimilated to two existing L1 categories (*Two Category Assimilation*), in which case discrimination will be excellent, they may be perceived as two equally good exemplars of one L1 category (*Single Category Assimilation*) resulting in poor discrimination, or they may be perceived as exemplars of the same L1 category where one is a more prototypical exemplar than the other (*Category*

[4] The model also includes hypotheses and predictions about age effects on new category formation: Flege assumes that the formation of a new category becomes less likely with growing age. This, according to him, affects only similar sounds, not new sounds (Flege, 1991, p. 282).

Goodness Assimilation), which results in intermediate discrimination. If one member of the L2 category pairs is uncategorized, i.e., perceived as not similar to any existing L1 sound while the second member is assimilated to an existing category (*Uncategorized-Categorized*), discrimination is predicted to be very good, whereas any degree of success in discrimination (i.e., poor to excellent) can be expected if both members are uncategorized (*Uncategorized-Uncategorized*). These categories, while developed in the context of the naïve learner assumed in the original PAM, can also be applied to a non-naïve L2 learner in a slightly rephrased version that does, however, not touch upon the conceptual characteristics (see Best & Tyler, 2007, pp. 28–31 for more details).

The central premise of both models briefly described in this chapter is that the native language exerts an influence on the non-native language. They ultimately share the same goal: predicting perceptual difficulty in novel contrasts primarily based on the L1 of the respective speaker. It is, however, not only the degree of difficulty that matters in non-native phonology, but also the concrete (acoustic) effect the L1 has on the L2, or, in other words, the way L1 structures are transferred onto the L2 system. This transfer, and underlying CLI will be introduced in the next section.

2.1.2 Unpacking transfer and cross-linguistic influence (CLI)

One of the main differences between acquiring[5] a native language and acquiring a non-native language – apart from cognitive differences and differences in the way the language is

---

[5] Often a terminological distinction is made between *learning* and *acquiring* a language, whereby learning takes place in a formal and acquisition in a naturalistic setting. I will refrain from making this distinction for reasons of readability, but, more so, due to theoretical considerations: While some aspects of language competency depend on the specific setting, most notably metalinguistic awareness, motivation, and language attitude, the main processes can be assumed to be comparable in both cases. The differences found in the two scenarios can be captured without invoking the acquisition vs. learning dichotomy and by instead referring to the individual factors that are connected to it. Furthermore, as argued in Schmidt's (1990) *Noticing Hypothesis*, a feature must be noticed in order to be acquired,

learned – is the presence of a previous language in the L2 learner. This influence, often called CLI, can, under certain circumstances, result in transfer from one language to another.[6] While transfer has first been observed from the native to the TL, any language in the multilingual system can, in fact, be the target of transfer, and any language or languages can be the source. When an L2 is acquired, L1 structures can be transferred from the L1 onto the L2 or vice versa. In simple terrms, transfer onto a non-native language can be positive or negative: In positive transfer, the source language and the target language share a structure, or structures that are perceived by the speaker to be similar (Ortega, 2013, p. 53), and the speaker makes use of this source language structure in the TL. Negative transfer is the result of (perceived) structural differences between the source and the target language, where the source language structure is falsely applied in the TL. Transfer is, however, an abstract concept and thus can occur in an indirect manner: The learner may, for instance, hypothesize that the TL must be different from L1, or they may overgeneralize the occurrence of certain structures or features in the TL. Ortega (2013, p. 53) explains that transfer "is about tendencies and probabilities", inside which the learner constantly balances the assumption that a target language structure will be similar to the L1, and the assumption that structures should be sufficiently different since they belong to different languages (Ortega, 2013, p. 53). Furthermore, Ortega (2013, pp. 39–42) notes that transfer does not always surface as such, but can instead result in the avoidance of certain structures, as well as in their under- or overuse. Going back to the dichotomy introduced above,

---

which, according to the author, involves a certain degree of consciousness. Thus, assuming acquisition to be fully unconscious as opposed to the conscious process of learning seems unrealistic. Moreover, in language training as practiced in schools and in language courses nowadays, the focus is not so much on explicitly learning the rules of a language, but rather on enabling students to internalize language structures. Thus, in this dissertation, the two terms will be used interchangeably, and instead reference will be made – if necessary – to the distinction between *naturalistic* and *formal*.

[6] The use of the term *transfer* has been criticized lately as lacking neutrality. I do not share the emotionally based dislike of the term *transfer* as one that conveys a negative bias, and in fact consider *transfer* and *CLI* to be conceptually different (see above). Thus, both terms will be used in this to refer to their respective concepts.

the distinction between positive and negative transfer is not always a straightforward one when there is more than one background language involved. This is the case especially when the two background languages are chosen to test the influence on one particular feature. When the L1 matches the target language and the L2 does not, the target-like perception or production of the target feature can be attributed to positive transfer from the L1 or to lack of negative transfer from the L2, irrespective of whether the L1 matches the L2 or not.

In generative frameworks, one assumes a learner who, to some degree, has access to universal grammar (UG), a fundamental language system that is innate and that is shared by all humans (White, 1990). The universal principles that are part of UG are then parameterized to ultimately form the grammar (or set of specific parameters) of the language that is being acquired. This, in its most untainted form, is thought to be the case for the acquisition of an L1 (White, 2013, p. 2), but there is ample evidence that UG is still active in L2 acquisition,[7] which has been shown by various studies pointing towards the influence of universal factors irrespective of the L2 being acquired or the L1 of the speaker. In fact, there is a plethora of both empirical studies and theoretical discussions trying to establish the degree of access to UG that a learner of a non-native language has. In a theoretical approach, Birdsong (1990) adapts and applies the arguments for UG in the acquisition of an L1 to SLA. Empirical studies have shown that universally unmarked structures emerge under certain conditions in intermediate acquisition stages even in cases where neither the L1 nor the target language motivate the use of the specific structure (see, for instance, Broselow, Chen, & Wang, 1998). These discussions are illustrated, for instance, in Schachter (1988), Schwartz and Sprouse

[7] There are three major stances to the degree of access to UG in SLA: a) The learner of an L2 does not have access to the UG (i.e., access is lost once the L1 is acquired). b) The learner of an L2 has full access to UG, c) The L2 learner has access to UG, but this access is mediated by the L1 (White 1990). For summaries of the different views on access to UG in SLA, see also Epstein, Flynn, and Martohardjono (1996).

(1996), and White (2003). I will refrain from delving deeper into this discussion and instead assume a potential influence of universal factors like markedness and by cognitive and physical properties like acoustic or articulatory complexity, for instance (see also Sönning, 2017).

To sum up, while effects are exerted by language universals in both native and non-native acquisition, in non-native language acquisition, there is an additional effect of the specific background languages and their interaction in the learner’s mind, irrespective of whether the subject of study is a monolingual or a multilingual. When studying multilingual learners in contrast to monolingual ones, one is faced with a large number of potential language biographies due to a greater number of languages in the speakers’ and learners’ systems, and thus a higher number of language combinations. In terms of transfer, this gives rise to more transfer scenarios and linguistic permutations, where any language in the speakers’ system can act as the source as well as the target of transfer. What is more, there may be multiple transfer sources active at the same time as well as combinations of particular languages that may ultimately be integrated into a fused/combined/intermediate structure which can then be transferred. At the same time, the terminology also inevitably becomes more diverse and more complex in the field of L3 and multilingual language acquisition. This is why it is necessary to establish the nomenclature that will be used in this dissertation, also touching upon (at least in part) current terminological debates.

2.1.3 L3 acquisition and bilingualism: terminological considerations

Two major dichotomies in bilingualism are based in the chronology of language acquisition, i.e., the time frame in which acquisition takes place. In this respect, *early bilingualism*, sometimes also referred to as *child bilingualism*, is the outcome of the acquisition of two languages, both of which are acquired before reaching the critical period,[8] while a *late bilingual* acquires their L2 after this cut-off point (see, for instance, Rothman, 2009). While this distinction revolves around the absolute age of onset, another distinction can be made between *simultaneous* and *sequential* or *successive bilingualism* (as used, for instance, by Amengual, 2019) considering the order of acquisition of the different languages. A simultaneous bilingual acquires both their languages at the same time as two L1s, whereas a sequential bilingual acquires their second language when the acquisition of the L1 is well on the way or already completed. Consequently, a late bilingual will necessarily have gone through consecutive language acquisition of a first and a second[9] language, whereas an early bilingual may either have gone through simultaneous or sequential acquisition. The languages of the early consecutive bilingual are sometimes referred to as *1L1* and *2L1*, taking into consideration both the chronology as well the status of both languages as non-L2s, as the onset of acquisition for both of them is set before the critical period and the learners/speakers are thus often "indistinguishable from monolingual L1 speakers in production" (Isbell, p. 61).

[8] Lenneberg's (1967) *Critical Period Hypothesis (CPH)*, suggests that there is a critical period for first language acquisition, after which the acquisition of native competency in a language becomes more difficult. He suggested that this critical period ends at the onset of puberty where, according to him, the brain matures significantly and thus changes its functionality (Lenneberg, 1967). Even though Lenneberg did not suggest that his critical period would imply a "sudden" cut-off point that makes native attainment impossible, some authors criticize exactly that and prefer to talk about a *sensitive* rather than a critical period (see Long 1990 for an overview and discussion). Other authors, such as Patkowski (1980), use the term *critical* for the acquisition of a first language and the term *sensitive* for the acquisition of a non-native language. Since Lenneberg, a large number of studies have been conducted to determine both the validity of the CPH and the age at which a hypothesized cut-off point occurs and researchers have continuously found evidence for the existence of a critical/sensitive period placing this period at an age somewhere before puberty. Furthermore, different periods have been identified for different linguistic levels. For phonology, for instance, some assume the critical age to be at about 5-6 years of age (see, for instance Flege, 1991, p. 270).

[9] *Second language* and *foreign language* are often used to refer to the outcome of acquisition through different processes or in different environment. Specifically, a second language is assumed to be a language that is acquired in a society in which it is a crucial part of everyday life, whereas a foreign language is usually acquired in a formal setting and is not used in the society where it is learned. Since all speakers tested in the empirical studies in this dissertation are learning or have learned their target language in the same setting and environment, i.e., in a secondary school setting, I will not make a distinction between the two terms in this dissertation and instead refer to both as *L2*.

Furthermore, languages beyond the L1 can be acquired either in a naturalistic or in a formal (i.e., instructed) setting (Ross, Yoshinaga, & Sasaki, 2002, p. 296). Research has shown that the outcome of instructed acquisition potentially leads to higher metalinguistic awareness as well as better learning strategies, which may, in turn, influence the way any subsequent language will be acquired. Thus, the distinction is crucial for research in L3 acquisition. A bilingual who has acquired their L2 in a formal setting might have a general advantage when acquiring another language in the same setting due to their heightened metalinguistic knowledge, while this is not necessarily the case (or at least not for the same reasons) in a simultaneous bilingual who has acquired the L2 naturally without formal instruction.

Heritage speakers (HSs)[10] constitute a particular group as they are situated somewhat between simultaneous and consecutive bilinguals in terms of language chronology: According to Montrul (2015), the typical HS is a second-generation immigrant and starts with acquiring the family language spoken at home by both parents (which is a language other than the majority language spoken in the learner's country of residence) as an L1 or *Heritage Language (HL)*, and later starts acquiring the societal language once they enter formal education. If the child has older siblings who are already in kindergarten or school and is thus surrounded by the societal language outside the home, there might be some input from the societal language before formal education, which, however, can be considered minimal since the family language is still the primary means of communication in the home. Language dominance, i.e., the relative degree of language use or the relative proficiency in two (or more) languages, typically shifts

[10] Across the literature, there are two different ways to define a heritage speaker. In one definition, the term describes "current adult learners of a second language (L2) that is a language of familial heritage" (Kupisch and Rothman, 2018, p. 567), i.e., HSs that are in the process of re-learning their HL in adulthood, potentially after an extended time of not using it. According to another definition, it is used to describe "naturalistic bilingual acquirers/learners of an HL" (Kupisch and Rothman, 2018, p. 567), i.e., children acquiring their HL naturally in the home environment. In this dissertation, the term heritage speaker will refer to the latter definition.

during the early years of the heritage speaker starting with the onset of acquisition of the societal language, which gradually becomes more dominant than the home language (Valdés, 2005, p. 417). Section 4.1.3.4.3 gives more details about the operationalization of language dominance utilized in this dissertation (see also Treffers-Daller, 2019, for further reading). This process has been claimed to result in incomplete acquisition of the L1, a claim which has been repeatedly refuted by the literature, however. For detailed discussions about the concept of incomplete acquisition in heritage speakers, see, for instance, Cabo and Rothman (2012), and Kupisch & Rothman (2018). While heritage speakers are a potentially heterogeneous group of speakers (with different levels of proficiency in their languages, as well as different linguistic and biographical backgrounds), Lohndal, Rothman, Kupisch, and Westergaard (2019) sum up the basic profile of a heritage speaker using the following aspects: They are “minority language speakers in a majority speaker environment (Lohndal et al., 2019, p. 4), they are bilingual, and “by the time they are adults, [they] tend to be dominant (i.e., more proficient) in the language of their larger national community” (Lohndal et al., 2019, p. 4).

Some authors furthermore make a distinction between *active* and *passive bilingualism* (see, for instance, Sharwood Smith, 2017). According to this distinction, a passive bilingual is fully functional in one of their languages, but only partially functional in the other in that they cannot produce the language at a sufficient level, while an active bilingual can use both languages productively as well as receptively. In this dissertation, I will use the term *bilingual* to refer to active bilingualism unless stated otherwise. In some cases, a heritage speaker may shift the role of his languages to such a degree as to essentially become a passive bilingual with the home language assuming the role of the passive language. It is important to note at this point that the language configurations in a speaker’s biography are, as stated earlier, not fixed, so in a bilingual, dominance relations may change multiple times across the lifespan, as may competencies (for further reading, see, for instance, Gathercole, 2016 or Unsworth, 2016). For example, a heritage speaker may only learn to write the heritage language later in life, after learning to read and write in the societal language, or they may speak the language initially, but lose most of their ability to produce the language later in life when it gradually loses importance to the speaker, while still retaining the ability to comprehend the language. In cases like this, where competencies disperse differently across the languages of a speaker, Maher (2017, p. 63) uses the term *asymmetrical bilingualism*.

Like the terms for different types of bilingualism, those for the various languages acquired by different speakers in different acquisition scenarios are not used consistently by different authors. The use of *L1* to denote the native language of a monolingual speaker or late learner of an additional language, for instance, is well established. The term *L2* is sometimes used broadly to refer to any language acquired after the L1 (i.e., any additional, non-native language, even in cases where other non-native languages have been acquired before), while others use the term to denote specifically the second language, i.e., the language acquired immediately after the first language.[11] In HSs, the terms *HL* and *L1* are often used interchangeably, with the former focusing on the status the language has in a society and, possibly, the dominance situation, and the latter focusing on chronology. In the same context, the L2 of a HS can either be considered the chronological L2, i.e., the societal language acquired after the HL or an additional L1, i.e., an additional language acquired before the critical period. I will use the former and thus refer to a HS's family language as their HL or L1, and to the societal language as their L2.

Generally, in cases where a speaker can function in more than two languages, the term *multilingualism* will be used. It is important to remember that the use of static terminology to

[11] Outside the field of language acquisition and psycholinguistics, a different usage of these terms can be found, as is the case, for instance in the fields of sociolinguistics and language teaching. These different uses of established terms will, however, not be discussed here.

refer to dynamic concepts like those found in multilingual language acquisition will always require a certain amount of compromise. With this in mind, most language biographies found in bilinguals can be covered using combinations of the above terms, especially since acquisition scenarios and language combinations become more varied as more languages are added, resulting in potentially highly diverse language biographies. The terminological and theoretical considerations detailed in this section lead to a classification of languages that will be summarized in Table 1.

Table 1: Terms and their use in this dissertation

Term	Definition
L1	The (chronologically) first language acquired by a monolingual or a late sequential bilingual
L2	• The (chronologically) second language acquired by a late sequential bilingual • The societal language of a HS that is acquired after the L1
1L1/2L1	The (chronologically) first or second language respectively acquired by an early sequential bilingual
HL (heritage language)	The home language of a HS
L3	Any additional language acquired by any type of bilingual or multilingual
TL (target language)	The language being acquired/under scrutiny
Background language	Any language in the learner's system other than the TL
Monolingual	A speaker that can function in one and only one language
Bilingual	A speaker that can function in two languages
Multilingual	A speaker that can function in more than two languages

2.2 Previous research in L3 phonology

The acquisition of an L3 as distinct from the acquisition of an L2 has gained importance as an object of study during recent decades. While the scenario was first studied in the context of (morpho-)syntactic or lexical acquisition, research on L3 phonology has recently been gaining momentum with numerous projects in process, both large- and small-scale. L3 phonologists initially ventured to adapt existing L3 syntax theories, models, and methods, but attempts are now being made to modify methodological approaches and formulate hypotheses that are more firmly based in what is known about L2 phonology (see, for instance, Kopečková, 2019, who uses a Dynamic Systems approach to track language development in multilingual learners, or Wrembel et al., 2019, who attempt to extend the PAM-L2 to phonological L3 acquisition).

This chapter focusses on studies about L3 phonology instead of exhaustively covering syntax and the lexicon as well, unless studies and findings are deemed directly or indirectly relevant to research on L3 phonology. Thus, outside of phonology, only those studies are reviewed that directly led to the formulation of L3 models which, in turn, were also adapted for phonology. The main reason for this decision is that methods, findings, and - as I will show later - models and concepts developed for or as a consequence of research on other fields, cannot readily be transferred to phonological acquisition and it seems prudent – at least for the time being – to keep the domains separate in the model formation stage before potentially relating the different domains to each other again in future research.

As has been established in Section 2.1.1, the learner deals with two major factors in language acquisition: non-language-specific universals (assumed to be active in both L1 and L2 acquisition) and language-specific influence, i.e., structural or surface transfer (only active in L2 acquisition). These factors are likely to be conditioned by social (e.g., identity, attitude) and individual (e.g., memory, aptitude) factors, with interactions between the two occurring. These same factors can also be assumed to be at play when acquiring an L3 (or any additional language, for that matter). Universal factors and tendencies should not play a different role in L3 than in L2 acquisition, or, in other words, UG, if indeed still accessible to the L2 learner, should also be so for the L3 learner. Indeed, studies designed to examine the most likely source(s) of transfer in a set of languages have found effects that could be attributed to the influence of universals. Louriz (2007), for instance, used an Optimality Theory approach to analyze the acquisition of English word-level stress by speakers of L1 Moroccan Arabic and L2 French. She found that her results were best explained by an interplay of universal factors and language-specific misinterpretations of the L3 input. Gut (2010), when conducting a case study on the production of speech rhythm and vowel reduction in four multilingual learners of L2 German/L3 English or L2 English/L2 German with different L1s (Polish, Russian, Hungarian, Spanish) found that, while both vowel reduction and speech rhythm in the L2 and the L3 of these four multilinguals differed significantly from control values obtained from native speakers, these differences could not be traced back to the respective L1s, i.e., they could not be attributed to transfer. Furthermore, all four speakers produced distinctly different patterns in English and in German, respectively, irrespective of which of the two was acquired first. Gut takes these findings to mean that the three languages do not exert a strong influence on each other on the level of rhythm and vowel reduction, and her results can thus, according to her, be interpreted as evidence that it is "language-inherent properties [of the L3] which determine vowel reduction" (Gut, 2010, p. 33), which Cabrelli Amaro interprets as the influence of universals (Cabrelli Amaro, 2012, p. 44).

In terms of the role of UG, then, L3 acquisition as a process cannot be considered fundamentally different from L2 acquisition. The main difference between acquiring an L2 and acquiring an L3 is the number of individual languages in, and thus the complexity of the learner's language system. In other words, where in L2 acquisition potential language-specific influence can only be exerted from the native language, multiple languages can potentially be the source of transfer in the multilingual learner. These languages have different linguistic, i.e., language-specific properties, but also speaker-specific attributes, i.e., there might be a difference in degree of dominance, chronology of acquisition, attitudinal aspects, etc. Researching the acquisition of an L3 as well as the roles of and interaction between different attributes can contribute to what is known about the interaction between different languages in a speaker's mind by studying the way structures are transferred to the L3. The following sections deal with different approaches to this endeavor, by first explaining the early L3 syntax models, and then tracing the lines of inquiry taken in the subfield of L3 phonology.

2.2.1 Early L3 models

In order to gain insights into the interaction of different languages in a multilingual's mind, researchers dealing with the acquisition of an L3 have focused mostly on identifying the source of transfer into the L3, which – very broadly – can be the L1, the L2, or both L1 and L2, [12] while the preferred source can be conditioned by various language-specific and extralinguistic factors. In the context of L3 syntax, three models were developed to predict the most likely source of transfer on the TL in L3 acquisition. These models, which have also been redeployed frequently in L3 phonology, are the Typological Primacy Model (TPM; Rothman, Iverson, & Judy, 2010), the L2 Status Factor Model (L2SFM; Bardel & Falk, 2007), and the Cumulative Enhancement Model (CEM; Flynn, Foley, & Vinnitskaya, 2004). Rothman et al. (2010) proposed their model based on a study of a group of L1 Italian/L2 English learners of L3 Spanish and a group of L1 English/L2 Spanish learners of L3 Portuguese. These speakers were tested on their acquisition of adjectival interpretation conditioned by noun raising, a semantic distinction absent from English, but present in Italian, Spanish, and Portuguese. The study revealed that both speaker groups seemed to make use of their background languages

[12] The same can be done for larger numbers of languages. Restricting the number to two background languages is done for the sake of simplicity (in terms of theoretical explanations) and for practical reasons (in terms of empirical inquiries).

and transferred the distinction positively, irrespective of which of their background languages (L1 or L2) employed the feature. Based on these findings as well as on findings from previous studies (e.g., Judy, Guijarro-Fuentes, & Rothman, 2008, Cabrelli Amaro & Rothman, 2010), Rothman et al. proposed that typological proximity plays the decisive role in conditioning transfer on the L3 irrespective of the status of the background languages, given that typological relations between the languages make such a distinction possible.[13]

Bardel & Falk (2007), in contrast, found that, in their study, it was the L2 rather than the L1 that exerted an influence on the L3 in sequential bilinguals. They tested a small set of learners with different L1s and L2s acquiring L3 Swedish or Dutch on their acquisition of placement of negation in the initial stage of L3 acquisition and found that the status of a language as an L2 will override typological proximity. They went so far as to conclude that "in L3 acquisition, the L2 acts as a filter, making the L1 inaccessible" (Bardel & Falk, 2012, p. 480). Later, they formulated their L2SFM (Bardel & Falk, 2012), which still sees L2 status as the main conditioning factor for the choice of transfer source, but which additionally allows for decisive roles of other factors such as typological proximity or extra-linguistic factors, such as motivation and age, in determining the preferred source of transfer in L3 acquisition.

Flynn et al. (2004) formulated their *Cumulative Enhancement Model* (CEM) based on a study on the acquisition of relative clauses in child and adult L2/L3 learners of English with L1 Kazakh/L2 Russian, all sequential bilinguals, and compared the results to comparable studies in L2 learners of English. They found that the L1 did not play a privileged role as opposed to the L2 in L3 acquisition but that, instead, speakers could draw from experience with

[13] In a language grouping with a similar typological distance between languages, predictions based on the TPM are difficult, if not impossible. This may be the case if both distances are very small, as would be the case in three closely related languages like Swedish, Danish, and Norwegian, or if they were similarly large, as, for instance, in a language grouping like Basque, Korean, and Farsi.

a linguistic structure from either of their background languages and transfer the respective structure positively. However, in speakers who were acquiring L2 Russian and L3 English simultaneously, the authors did not find an influence from the L2 onto the L3. Thus, they concluded that a learner needs to have achieved a certain degree of proficiency in the L2 before it can exert an influence on the L3.

Summing up, the TPM predicts exclusive transfer from the background language that is typologically closer to the TL, the L2SFM predicts exclusive transfer from the L2, and the CEM is concerned with sequential bilinguals acquiring an L3 and predicts potential transfer from both background languages as long as transfer is beneficial. All three of these models lack strong predictive power, however, and they are restrictive as they presuppose very specific conditions: The L2SFM, for instance, requires a particular type of bilingual whose background languages can be unambiguously identified as L1 and L2 respectively. The TPM assumes a clear typological categorization into "closely related" and "less closely related" while not indicating to what extent this distinction should be based on genetic relationships of the languages concerned or on the overlap of specific linguistic features. The CEM, while the most suitable of the three for this particular study, is generally problematic due to its focus on the binary distinction of beneficial vs. non-beneficial transfer – a distinction that cannot always be easily made. Consequently, no single one of the models in their original form is appropriate for the analysis of the speakers in the empirical studies conducted in the course of this dissertation, which is also true for a (rather large) number of subsequent studies that have claimed to test them. While I will discuss potential influence from different transfer sources in the speakers' language system, I will refrain from later invoking these models directly, whereas they will find their way, in modified forms, into the potential scenarios presented in Section 3.3. In the following sections, I will outline the paths that have been taken in previous research, starting with early case studies and progressing to the mostly quantitative research that informs the use of theoretical approaches in L3 phonology.

2.2.2 Early exploratory studies in L3 phonology

The earliest studies on L3 acquisition were predominantly exploratory and qualitative. As such, they usually focused on multiple linguistic phenomena, either sticking to one linguistic domain (and, for instance, for phonological acquisition, examining various phonemes) or even extending investigations to multiple domains or the full language system. The first studies were conducted using an error analysis approach, i.e., the identification and listing of all occurring errors in several language samples and seeking to explain and, if possible, categorize them according to the background language to which they can be attributed. In a longitudinal case study, Chamot (1973) performed such an error analysis on the speech of a ten year old French-Spanish bilingual who had just moved to the USA and was thus acquiring English when the study was conducted. Before the move, French had been the boy's home language and Spanish the societal language. During the course of nine months, the boy's natural speech was recorded regularly and compared to the standard spoken by the American caregiver; items were marked as "errors" when they did not match this standard. Chamot found that both languages could be sources of negative transfer,[14] and this negative transfer could be "intensified" (Chamot, 1973, p. 249) if the feature that was transferred occurred in both background languages.

Bentahila (1982) used a slightly different approach to determining transfer errors in L3 phonology by analyzing the teacher statements to collect typical student errors from Moroccan

[14] Error analysis, concerned with errors that are made by a speaker, crucially does not see potential errors that are not made and it does not account for other learner phenomena such as avoidance.

learners of English (L1 Moroccan Arabic, L2 French). He furthermore collected learner errors in the English of students from Kuwait and Iraq - both countries where French is not spoken – as well as Algeria, which has a bilingual background similar to that of Morocco. He found that the mistakes made by Moroccan learners could often be traced back to their L2 French phonological system, even when their Arabic would have provided a source for positive transfer. He attributes this to the learners' perception of the L2 French and the L3 English as more similar to each other than to Arabic, both in terms of structure and lexicon, and due to the writing systems used in the languages.

Chumbow (1981) equally found evidence for phonological transfer from the L2 rather than the L1 in L3 acquisition in his study using an error analysis approach, but, in contrast to Bentahila (1982), he interprets this as the result of a shift of the unique role of the mother tongue, i.e., he claims that the dominant role the L1 is supposed to play can be passed over to one of the other languages in the speaker's mind which then becomes the structure providing language. In fact, Chumbow (1981) proposes a set of variables whose interplay determines the language that is the recipient of this *mother tongue effect*. Specifically, he names language dominance, proficiency in each background language, degree of standardization (both spoken and orthographic), the role of either background language as medium of instruction in general and in learning the foreign language, the perception of ethical and cultural distance from and attitude to the background languages, typological distance, and the speech produced by the language teacher or target speaker (if the target produces speech that deviates from the standard, error analysis might misidentify certain features as "errors"). He attempted to quantify the influence of each of these factors using a scoring system by which he then tried to predict the most likely transfer source. While this assumes the phonological system of each language as a whole (i.e., one of them will be more likely to transfer) and thus does not take into account the potential differences between the various features, this was the first – and, as of now, to the best of my knowledge, the last – time that such a quantification method has been attempted in this particular field, and it is particularly noteworthy that this promising approach has not been taken up again in later studies where methods other than error analysis could have been included.

A more global approach was taken by the series of seminal studies conducted by Hammarberg and colleagues (Hammarberg & Williams, 1993; Hammarberg, 1996; Hammarberg, 2001; Hammarberg, 2005). In their longitudinal case study, the authors monitored and analyzed the language production of an L1 English/L2 German learner of Swedish as an L3 in terms of phonology, lexicon, and morphology. Initially using accentedness judgments to determine perceived foreign accent source, Hammarberg and Williams (1993) found that the L2 was the predominant source of transfer in L3 productions in the initial stages of acquisition (1 week after onset), while this status shifted to the L1 as L3 proficiency increased (as exemplified with a second test 1 year later). In addition, they uncovered a task effect: The second test was complemented with an imitation task, in which the subject repeated small 3-6 word segments of a text after a native speaker of Swedish, and then reading a second passage on her own. The imitation passage was perceived as English-accented, while the self-read passage was perceived as German-accented, which they explained with the potentially different roles the two background languages play in L3 production. According to their interpretation, the influence of the L1 is the result of an unconscious process, while influence from the L2 stems from conscious operations.

Overall, these studies managed to provide valuable initial insights into the processes involved in L3 acquisition, and unearthed the existence of multiple transfer sources that were shown to be affected by factors such as language status (i.e., L1 vs. L2), stage of acquisition, and task, and they also indicated that more than one language could be transferred at the same time, both negatively and positively. This way, they paved the way for later studies, although they were not yet able – due to the methods employed - to characterize the phenomena and processes more systematically.

Subsequently, research in the field mostly took one of two paths to tackle the issues and questions raised by these early studies: Some researchers, following the Hammarberg studies, started to examine more systematically how the background languages surfaced in TL productions by investigating perceived foreign accent in the L3. Others increasingly started to focus on individual phenomena and features in great detail, which, in phonological L3 research, mostly meant intricate acoustic measurements and analyses, as well as – although to a lesser degree – perceptual studies. These two paths will be unpacked in the following two sections, 2.2.3 and 2.2.4.

2.2.3 Global accentedness

Building on Hammarberg's and Williams' (1993) case study (see 2.2.2), Wrembel (2010, 2012, 2013) ran three accentedness studies using comparable methods and speech samples collected from people who mainly differed in their language biographies. In her first study (2010), judges were asked to identify the L1 of the read and free speech produced by learners of English with L1 Polish and L2 German. The speakers differed according to their proficiency level (beginner, elementary, pre-intermediate, intermediate). In total, 33.3% of the speech samples were correctly identified as being produced by native speakers of Polish, whereas the ratio of correct identification increased in higher proficiency levels (beginner: 20%, elementary: 8%, pre-intermediate: 50%, intermediate: 71%). Incorrect identification as German amounted to 53% in the combined beginner/elementary, and 17% in the pre-intermediate/intermediate group. The author interpreted this as evidence for a combined influence from both background languages, whereby the influence of the L2 was stronger in earlier acquisition phases and then gradually decreased, while at the same time the influence of the L1 increased, which corroborates the findings from Hammarberg & Williams (1993). The author remarked, however, that this effect did not necessarily have to stem from language status, as suggested by Hammarberg, since the L2 was also the language that was typologically closer to the target language, which made it impossible to determine with certainty where the effect stemmed from.

In order to eliminate this confound, Wrembel (2012, 2013) conducted two follow-up studies with speech samples from L1 Polish/L2 French/L3 English and L1 Polish/L2 English/L3 French, respectively, using an approach similar to that in Wrembel (2010). The author found that, in the first study (2012), Polish was correctly identified as L1 at a rate of 63%, while the rate of identification as L1 French was 16%, and there was no significant effect of proficiency level. The author interpreted this as evidence for a generally stronger influence of the L1 Polish on the speakers' English while acknowledging a certain degree of L2 influence and thus mixed transfer. In the second scenario (Wrembel, 2013), the results showed a predominant effect of the L1 on the perception of accentedness in speech produced by speakers at elementary proficiency level (speech identified as L1 accented: 53%, speech identified as L2 accented: 28%), whereas this difference could not be found in the speech of intermediate level learners. Thus, the findings from the follow-up study could not confirm the shift from the L2 to the L1 as the primary transfer source found in Wrembel (2010). Instead, the influence of the L1 Polish was stronger than that of L2 French at all proficiency levels, while the L1 French was the primary transfer source as opposed to L2 Polish only at low proficiency level. These results seem to contradict findings from Hammarberg and Williams (1993), although it has to be noted that the speakers categorized into elementary level had reached stage A2 (CEFR),[15]

[15] The *Common European Framework of Reference for Languages* (Council of Europe, 2001) is a reference instrument for language proficiency and forms the basis for curricula and teaching materials in Europe and also in other continents. It describes six proficiency levels, A1, A2, B1, B2, C1, and C2,

which is well above the initial stages tested in the Hammarberg study. Taken together, the results from Wrembel's studies point towards an interplay between factors such as language status, language proficiency, and language typology.

Chang (2015) ran a study on perceived global foreign accent in the TL Korean of L1 English monolinguals vs. L1 English/L2 Japanese bilinguals and reported results that, as he claimed, contradict Wrembel's study. In his study, native Korean judges were asked to identify the speakers' dominant accent (either English or Japanese) and to rate this judgment on a scale ranging from "definitely English" (corresponding to a score of 0) to "definitely Japanese" (corresponding to a score of 9). The midpoint of the scale was meant to indicate that the speaker was not perceived as strongly accented in either direction. Chang's results showed that the bilinguals generally received higher scores than the monolinguals, or, in other words, they were rated as "more Japanese", indicating an influence of the L2 on the speakers' L3 Korean. When, additionally, the speakers were categorized as beginner, intermediate, or advanced learners of Korean, respectively, results revealed a general tendency for advanced learners to be perceived as less foreign-accented (i.e., these speech samples received scores closer to 5) than their lower-proficiency counterparts. While this is not particularly surprising, the author took this result as evidence for a gradual decline of L2 influence as observed by Hammarberg and Williams (1993), while refuting the expected increase of L1 influence. However, mixing the notions of source and degree of transfer in one scale conflates two very distinct concepts, and it presupposes the presence of one distinct transfer source by not providing an option for mixed transfer from both languages, which can be considered a serious methodological problem in the light of the question that was supposed to be answered by Chang's study.

---

where A1 is the lowest and C2 is the highest level of proficiency. The framework was designed by the Council of Europe to ensure comparable proficiency standards across different languages (Council of Europe, 2019).

In a similar vein, Lloyd-Smith, Gyllstad, and Kupisch (2017) examined perceived foreign accent in the L3 English of 18 Turkish HSs with L2 German as well as 5 monolingual Turkish speakers and 5 monolingual German speakers, each with L2 English, and five native speakers of English. Speech samples produced by these participants were judged for foreign accentedness by 20 native speaker English teachers who also identified the L1 of the speaker in each speech sample as either Turkish, German, English, or “none of these” (in which case they were asked to indicate the alternative L1). The bilinguals were identified as German-accented at a rate of 60%, and as Turkish-accented at a rate of 18%, whereas the L1 German speech samples were identified correctly at a rate of 80% and the L1 Turkish speakers at a rate of 47%. The accuracy rate for Turkish monolinguals is suspiciously low, which can be best explained by the judges’ lack of familiarity with Turkish accented as opposed to German-accented English (all judges had lived and taught in Germany for a considerable time, a shortcoming acknowledged by the authors). However, and more interestingly, the bilinguals were generally rated more variably, or, in other words, German monolinguals were rated as more clearly German, and Turkish monolinguals as more clearly Turkish than the bilinguals, which can be interpreted as mixed transfer from the L1 and the L2, while the influence from the speakers’ Turkish was shown to be lower than that from their German.

When accentedness ratings are concerned, it is particularly important to note that often it is not clear which elements of speech the judges use to make their choice. So, for instance, in a language setup as seen in Wrembel (2010), judges might predominantly rely on Voice Onset Time (VOT)[16] to categorize the speech samples they hear, and thus judge the TL as

[16] VOT is the time that passes between the release of a stop consonant and the onset of the following vowel sound (i.e., the onset of voicing), or in other words, the duration of the burst that directly follows the opening of the oral cavity during the stop release. For more information, see, for instance, Halle, Hughes, and Radley (1957).

Polish-accented when VOT is relatively low, and as German if VOT is relatively high. Assuming a scenario in which different features can be influenced by different background languages, vowel quality, for instance, may acoustically be influenced more by the speaker's German, irrespective of their VOT range. In other words, if judges heavily rely on one feature or one particular set of features, global influence may be masked by this bias in certain combinations. While judges are sometimes asked to name or list the specific aspects that, to them, make the speech sample sound accented in a certain way, answers might not always be precise enough to allow for proper categorization, and even if they are, this approach does not necessarily help in controlling for (perceptual) bias. In Wrembel (2013), for instance, the judges were asked to name the features that led them to the perception of the individual speech samples as accented by specific L1s. While the author does not clarify whether the judges were provided with possible response options that they were asked to choose from, or whether this was an open question the answers to which the authors clustered into categories after the fact, they conclude that "the most frequently identified features of a foreign accent" (Wrembel, 2013, p. 42) included vowels, consonants, intonation, rhythm, nasality, etc. Even though the judgments were made by phonetically trained participants, the categorizations are very broad and do not reveal, for instance, whether, in vowels, it was vowel quality or duration that was perceived as accented, or which aspects of intonation and rhythm gave the impression of non-nativeness. In order to remedy this, a look at the phonological and phonetic features in the multilinguals' speech can serve as complementing evidence. Thus, the following section will discuss those studies that examine discrete features.

2.2.4 Single feature acoustic studies

Most research on production in L3 phonology that uses acoustic measurements as a tool of analysis has dealt with either VOT or vowel quality. This is due to two reasons: Firstly, both

VOT and vowel quality are acoustically relatively straightforward, with simple durational measurements for VOT and measurements of formant frequency for vowel quality. Secondly, and this is most likely also a result of the ease with which they can be analyzed, they have also been a major object of investigation in L2 phonology (as noted, for instance, by Hansen Edwards & Zampini, 2008), which makes possible a relatively direct comparison between many studies in the two areas.

Tremblay (2007), for instance, analyzed VOT values of L1 English, L2 French speakers in their L3 Japanese and reported a positive influence of their L2 resulting in native-like values in Japanese, allowing for a potential influence of the L1 on the L2, which results in combined rather than exclusive L2 transfer on the L3. Based on these findings, Llama, Cardoso, and Collins (2010) tested two groups of Spanish learners on their production of voiceless stops in onset position. One group contained speakers of L1 English and L2 French, and the other one contained speakers of L1 French and L2 English. The authors found that, in their study, L2 status was more powerful in determining transfer source than typological distance was. Importantly, however, they used a binary operationalization of aspiration in their analysis (with presence of aspiration corresponding to *English-influenced* and absence of aspiration corresponding to *French-influenced*) instead of taking into consideration absolute VOT values in their final analyses. This would have obscured any potential hybrid effect. However, the authors do report that their acoustic analyses indicate an influence of L1 VOT on the production of L2 VOT in both speaker groups, which, in turn, and in line with Tremblay (2007), points to combined transfer from a hybrid background system. These findings could be consolidated by Wrembel (2011), who tested bilinguals with L1 Polish and L2 English on their VOT patterns in L3 French. She found that the bilinguals make use of merged VOT values from their background languages when producing the L3, and she additionally reports “universal effects of place of articulation on VOT length” (Wrembel, 2011, p. 4). This effect of mixed VOT values from the background languages as the source of transfer was also found by Sypiańska (2013) as well as Wrembel (2015).

Llama & Lopez-Morelos (2016) found an effect of language dominance on the target-like production of stop aspiration by HS of Spanish with L2 English learning L3 French. According to the authors, Spanish and French native-like values are usually below 35ms (i.e., short-lag unaspirated), English values can exceed 60ms (i.e., long-lag aspirated). They analyzed VOT values for all three languages, and their results show that the speakers produced native-like values in their Spanish as well as their English, while they transferred values from their dominant language English in the production of their L3 French. In a similar study that also tested VOT, Llama & Cardoso (2018) reported more prominent transfer from the L1 as opposed to the L2 onto the L3. They explained the conflicting findings in their study and in previous VOT studies with a potential influence of L3 proficiency. Amengual, Meredith, and Panelli (2019) additionally uncovered an effect of language mode on the VOT values produced by multilingual speakers in their study on speakers of L1 Spanish-L2 English-L3 Japanese and L1 English-L2 Japanese. They tested the production of both/all three languages using reading tasks in two different types of sessions: a monolingual session (where the languages were strictly divided into different blocks) and a bilingual session (where all languages appeared in the same block). The authors measured VOT for the participants' productions of the voiceless velar stop /k/ in each of their respective languages and found that the speakers maintained distinctive VOT productions for each of their languages. However, they also discovered a significant interaction of language and speaker group. Specifically, they found that in the productions in the bilingual tasks differed significantly from the production in the monolingual tasks and that VOT length converged across languages in the bilingual as opposed to the monolingual task. Thus, the author concluded that multilinguals do indeed form language-specific VOT categories, but that their concrete phonetic realization is affected by transfer when speakers are in "multilingual mode".

The intermediate VOT values that are usually found in multilingual speakers reflect the gradient nature of the phenomenon. Similarly, vowel quality is not binary but gradient, and, based on the findings from most of the VOT studies, intermediate formant values for vowels in the vowel system(s) of the multilingual can be expected. Blank and Zimmer (2009) tested vowel duration and vowel quality produced by one L1 Brazilian Portugues, L2 French learner of L3 English, and they indeed found hybrid values, i.e., values (both quality and duration) between the L1 and the L2 in the speakers L3. Similarly, Sypiańska (2016) found evidence for combined transfer when examining L3 English vowels produced by L1 Polish/L2 Danish speakers. Additionally, she found that all three of the speakers' languages exerted an influence on all other languages, which she takes to be evidence for a global language system in the multilingual mind "whose characteristics depend on the characteristics of its component languages" (Sypiańska, 2016, p. 478). In Llisterri and Poch-Olivé's (1987) study on L3 phonetics, the authors tested the production of L3 French (Experiment 1) and L3 English (Experiment 2) vowels by Catalan-Castilian bilinguals and compared their productions to those produced by Catalan monolinguals.[17] Assuming a higher degree of overlap between different vowels in non-native as opposed to native speakers of any given language (which translates into more robust category boundaries in native speakers) they found that the bilinguals in Experiment 1 behaved like non-native speakers in both their L2 and their L3. Experiment 2 showed that English schwa was produced with the same acoustic characteristics as Catalan schwa in the bilinguals, as well as positive transfer in vowel contrasts that occur in Catalan but

[17] The authors of this study examine the production of both vowels and fricatives. However, they do not put the results of these two analyses in relation to each other, which is why I will treat the vowel section of the study as a separate study.

not in Castillian. They interpret these results as a lack of evidence for an L2 influence in the production of L3 vowels and instead attribute all deviations from the target norm to interference from the speakers' L1. Missaglia (2010), on the other hand, tested the acquisition of L3 English vowels by young German-Italian early bilinguals and found a beneficial effect of both languages on the production of vowel quality. Based on the results of her study, Missaglia additionally proposed that the language that is typologically closer to the target language can serve as a bridge that ultimately helps in the acquisition process.

While studies on VOT and vowel quality still form the majority of research in L3 phonology, some studies have considered other features. Llisterri and Martinez-Dauden (1990), for instance, tested the production of laterals by Spanish-Catalan bilinguals learning French. They found a tentative positive effect of the speakers' L1 Spanish, which enabled the speakers to produce a target-like version of the French non-velarized lateral. However, the participants tested in their study were particularly heterogeneous in terms of both language dominance and proficiency. Furthermore, the learners produced the unmarked version of the lateral sound pair – which just happened to coincide with the target-like production. This further calls into question the interpretation of the authors' study as predominant influence from the L1.

In a particularly interesting study, Benrabah (1991) found that the most likely transfer source was determined by the complexity of the system. He tested Algerian Arabic-French bilinguals on their production of English consonants and vowels. The participants transferred consonants from their Arabic, which has a more complex consonant system than French, and vowels from their French, whose vowel system is more complex than that of Arabic. Benrabah concluded that it was the more complex system that was most likely to be transferred in L3 acquisition.

Kopečková, Gut, and Golin (2019) studied the production of /w/ vs. /v/ by two groups (one group of adults, one group of children) of L1 German, L2 English learners of L3 Polish to find out about the development of production accuracy. Unsurprisingly, the adult participants outperformed the children in the target-like production of the L2 English /v/-/w/ contrast. However, the children performed better in the production of Polish /v/-/w/ at the early stages (the first ten weeks) of learning. Notably, these initial analyses were done using impressionistic judgments. The authors explained the unexpected results by the concrete acoustics used by the children in the production of the Polish fricatives: While both groups used the first formant (F2) to distinguish the two sounds, only the children additionally used friction (as measured as harmonics-to-noise ratio). It is unclear from this study, however, why the children used different strategies than the adults in their L3 Polish, and this may actually be caused by something other than acquisition or a change in transfer source caused by multilingual language development. More specifically, different learning strategies employed for the learners' Polish lessons or different varieties (or non-native accents) used by the teacher may also explain the results.

2.2.5 Perception studies

Perception studies usually lag behind production studies in research on language acquisition, and L3 phonology surely is no exception to this. However, while there are still only few studies on the perception of an L3 phonological system, those that do exist have yielded some noteworthy results. In their study on the perception of rhotics by multilingual children in the early stages of L3 acquisition, for instance, Balas, Kopečková, and Wrembel (2019) tested L1 German – L2 English learners of L3 Polish, and L1 Polish – L2 English learners of L3 German using a forced-choice goodness task. They found that both groups performed at ceiling level in their L2 English. However, the German learners of Polish scored significantly higher on the perception of the Polish rhotic (alveolar trills) than the Polish learners of German on the perception of the German rhotic (uvular fricative). They took this to be an effect of markedness, as they claim the German uvular fricative to be more marked than the Polish alveolar trill. However, it should be noted here that numerous German varieties employ trills instead of the standard uvular fricative, and that the authors do not report the German variety spoken by the German learners of Polish. Thus, the markedness effect claimed in this study, while attractive, must be seen as tentative.

Onishi (2016) examined the potential influence of L2 experience on the perception of L3 contrasts by testing speakers of L1 Korean, L2 English, and L3 Japanese in a number of perception tests. Control groups were L1 English learners of L2 Japanese, and L1 Korean learners of L2 Japanese. The participants were tested on the perception of both their L2 and their L3. The author could show a link between L2 perception and L3 perception, albeit only for some contrasts. Onishi concluded beneficial transfer from both of the background languages, as well as increased global perceptual sensitivity in the L3 learners as opposed to the L2 learners. These findings are challenged, however, by studies like that conducted by Patihis, Oh, and Mogilner (2013), who tested monolingual English, bilingual English-Spanish, bilingual English-Armenian, and trilingual English-Spanish-Armenian speakers on the discrimination of phonemes of Korean, a language to which they had not had any previous exposure. They found language-specific effects of positive transfer, but no evidence for a broad bilingual advantage. The same advantage due to a language-specific rather than a general bilingual experience was observed by Kopečková (2016) in her study on the production of English rhotics by speakers of various background languages.

Gallardo del Puerto (2007) explored the influence of what he calls *level of bilingualism* on the discrimination of L3 English vowels and consonants by Basque-Spanish bilinguals. In his study, level of bilingualism refers to how balanced a bilingual is, and this degree of balance is operationalized as the participants’ use of Basque. Del Puerto tested three age groups (7, 8, and 11 years old) of speakers with high use of Basque and those with low use of Basque each.

In contrast to his predictions, the results revealed no effect of level of bilingualism on the perception in any of the age groups, which he attributed to the additional influence of factors like "the specific linguistic aspect under study, interlinguistic distance and type of schooling" (Gallardo del Puerto, 2007, p. 13).

Lewandowska and Wrembel (2019) tested the discrimination of vowel length by L1 Polish, L2 English speakers of either L3 French or L3 German. They focused on the influence the L3 had on the speakers' L2 and thus expected a facilitating effect of L3 German (which, like English, does employ phonemic vowel length), but not L3 French (which does not). Contrary to their expectations, they did not find a significant group difference in the discrimination of L2 English vowels, which they interpreted as a challenge for the typologically based L3 models. It is necessary to note, however, that the participants were at B2 level (upper-intermediate) in English, so a potential beneficial effect of the chronological L3, which was less proficient than the L2, may just not have surfaced yet. Additionally, it is not entirely clear whether the study design allowed for differences in voice quality (as additional cues to vowel length) or whether these differences were eliminated to arrive at more "untainted" results. So, while the authors did not find evidence for beneficial transfer from the L3 onto the L2, this study leaves many questions unanswered.

One of the few studies in the field that tested both perception and production was conducted by Zhang (2019), who tested the perception and production of English fricatives and vowels by Basque-Spanish and Catalan-Spanish bilinguals. While she found no group differences in the perception of either type of sound, Catalan speakers outperformed their Basque counterparts in the production of [ɛ]. Other sounds that could have been transferred positively, for instance, unstressed [ə], were not produced significantly differently across groups, which, according to the author, indicates that features can be, but are not necessarily transferred positively from a background language even if they are available.

In summary, the results obtained in the last decades, using various approaches on various types of speakers with different language biographies and different language combinations, give evidence for multiple potential transfer scenarios: Transfer has been revealed from the L1, from the L2, or a combination of both, but no coherent explanation or groups of explanations, which would need to account for all these scenarios, has been proposed so far.

2.2.6 L3 syntax and the L3 models focusing on complexity

A similar trend can be seen in research on the L3 of syntax, where results have also been pointing towards more complex transfer scenarios than those described by the three “classical” models (see Puig-Mayenco, González Alonso, & Rothman, 2018 for a very comprehensive and systematic account). As a consequence, two more recent models developed in the domain of syntactic acquisition have been used to account for this complexity: the Scalpel Model (SM; Slabakova, 2016) and the Linguistic Proximity Model (LPM; Westergaard, Mitrofanova, Mykhaylyk, & Rodina, 2016), both of which consider transfer beyond the initial state. The SM assumes that neither one of the background languages is inherently more likely to be the source of transfer, that there is no wholesale transfer at the initial stages, that transfer is conditioned by various factors (linguistic, experiential, typology, etc.), and that different types of structures will provoke different types and degrees of transfer (Slabakova, 2016). In a similar vein, the LPM proposes property-by-property instead of wholesale transfer and allows for facilitative as well as non-facilitative transfer from either or both of the background languages. Westergaard et al. (2016) provided empirical evidence for their claim: In an acceptability judgment study on word order in English on two groups of young monolingual (Norwegian and Russian, respectively) and bilingual (Norwegian-Russian) children, Westergaard et al. (2016) showed that the Norwegian monolinguals (as opposed to Russian monolinguals and the bilinguals) tended to over-accept ungrammatical inversion in English that would be acceptable as the default V2 order in Norwegian, but ungrammatical in Russian. However, the monolingual Russian participants outperformed the bilinguals on the grammatical sentences. According to the authors, this shows that facilitative and non-facilitative CLI can be active at the same time, and that it can stem from abstract rather than concrete structures. It needs to be noted that neither of the models makes any predictions as to which factors influence transfer, and how this applies to different features or structures. While observing and interpreting the results of L3 acquisition studies in general and the respective studies in particular in a way that clearly differs from earlier attempts in allowing for more diverse processes and results in L3 acquisition, the two approaches fail at their attempt to be predictive across a large number of structures and not falsifiable.

2.2.7 A new path: Multi-feature/multi-domain studies

An important additional observation when examining phonological transfer in particular, and when assuming language transfer that originates from either or both of the background languages, is the sheer variety of phonological elements found in different types of phenomena. For instance, one cannot necessarily assume that transfer works the same way in segments as it does in prosody, or that an allophonic distribution is transferred the same way as the articulatory gestures needed in the production of a novel L3 speech sound. Instead, different parameters can be expected to play a role in determining the degree of transfer. In other words, the overall transferability of a phenomenon (whether it is a contrastive segment, a phonological process, a prosodic structure, or something else) as conditioned by acoustic, articulatory, cognitive, and universal factors as well as by the factors previously identified (dominance, typological relatedness, language mode, etc.) has to be considered and weighed in order to allow predictions for the degree and source of transfer in a multilingual. Thus, all background languages can be potential sources of transfer accordingly. In order to gain an insight into the nature and interaction of these factors, studies are needed that examine multiple linguistic features in the same speakers using the same methods. Additionally, it is crucial to examine both the production and perception of the same features by the same speakers since transfer might work differently in the two domains. Furthermore, examining perception as well as production may reveal the role of psychoacoustic factors in the production of a non-native phonology.

Thus, I argue that one of the reasons for the conflicting results in the reviewed studies is the reliance on the analysis of individual phonological features and, thus, the disregard for different types of acoustic, articulatory, and phonological properties inherent in these phonological features. An awareness and understanding of just these properties and their influence on the transfer scenarios observes in L3 phonology is precisely what is needed if a comprehensive, predictive model is of L3 transfer is to be achieved. In order to identify these properties, the behavior of different linguistic items in the same speakers or speaker groups has to be observed and systematically analyzed.

One of the most recent approaches to multi-feature analyses is a suggestion made by Domene Moreno (2019) and Domene Moreno and Kabak (to appear) who propose multi-level and multi-dimensional bit-by-bit transfer.[18] Domene Moreno and Kabak (to appear) analyzed a subset of the speaker data used in Study 2 of this dissertation (6 adult Turkish-German bilingual users of L3 English and 9 German monolingual users of L2 English) on their speech rhythm in English and German by measuring and calculating a number of canonical duration and pitch based rhythm metrics. Their calculations revealed significant differences between the

[18] Similar suggestions have been made for L3 syntax, for instance by Westergaard et al. (2016), who assumes property-by-property transfer on the level of the micro-cue (see Westergaard, 2009 for a detailed account).

bilinguals and the monolinguals in the duration based, but not in the pitch based rhythm metrics, and the differences surfaced in more of metrics in the participants' English than in the participants' German productions, but crucially yielding the same directions of differences across languages. These differences matched the predictions made based on language transfer (i.e., the bilinguals produced more Turkish-like values) for some, but not all the metrics. The authors proposed that the tested rhythm metrics act as individual bits that are affected in different ways and thus interpreted their results as bit-by-bit language transfer.

The resulting bit-by-bit model assumes that the phonological grammar (PG) is made up of individual elements called bits. These bits are located on different levels and in different domains of the full grammar. More specifically, PG contains a finite number of discrete and non-overlapping bits. Each of these individual bits, in turn, contains a finite number of discrete and non-overlapping bits. Thus there is a vertical system of higher-level and lower-level bits (that can but do not have to contain each other) as well as a horizontal system of bit structures. The affecting factors and the interaction of horizontally aligned bits are subject to the same general principles of interaction and change, irrespective of the vertical level on which they are situated. In other words, the bits that make up the rhythmic structure of speech will be affected individually by factors such as language transfer, and the same is true for the bits that constitute other subsystems, like, for instance, the vowel system (i.e., the individual vowels). The model also predicts that similar processes should occur both in levels below the vowel, i.e., in the individual features of each of the vowels.

To test this proposal for another subsystem of PG, Domene Moreno (2019) measured the formant structure of L1 Turkish, L2 German, L3 English and L1 German, L2 English vowels of the bilingual and the monolingual participants from the same dataset and found that both groups exhibit similar vowel spaces both in their German and in their English, whereas slight but relatively robust (given the expected inter- and intra-speaker variability in non-native speech production) differences in the speakers' high back and low back vowels could be identified. This outcome was explained with the influence from the bilinguals' L1 Turkish. Furthermore, she found that deviations in the size and position of the vowel space could be explained by the influence of the smaller L2 Turkish vowel space as produced by the bilinguals. She took these findings as further evidence for bit-by-bit transfer with both the vowel space as a whole and the individual vowels as separate entities acting as individual bits, albeit on different levels.

In summary, research in L3 phonology has developed as a research field from Second Language Acquisition in general alongside L3 morphosyntactic studies, feeding on models and theoretical approaches from both. While different methodological and conceptual approaches (holistic and qualitative, indirect – as in accentedness studies, or direct, as in phonetic and psychoacoustic studies) have been used over the decades to test both production and perception, L3 phonological research as a whole has mostly focused on determining the (most likely) source(s) of transfer. Conflicting results have lead to the interim conclusion that discrete transfer sources are highly unlikely, necessitating more fine-grained study designs and analyses.

3 The present studies: Phonological acquisition in monolingual and bilingual learners of English

Here, the bit-by-bit approach is applied to an additional level of the phonological grammar, the segmental inventory as a unit, which can be considered a higher-level unit (as compared to the vowel inventory) with a particularly complex structure: One can assume the bits that make up the vowel inventory, for instance, to be relatively homogenous when it comes to their overall makeup – the different vowels can be described using the same distinctive features, such as backness or rounding, for instance – and thus the influence they are subject to can possibly be traced relatively straightforwardly. However, the factors that play a role in the different bits of the segment layer (which includes both contrastive and non-contrastive elements as well as elements that are subject to or that are in itself phonotactic constraints) is likely to be harder to identify and model. Several phonological structures were chosen that particularly lend themselves to the analysis in the Turkish-German-English context as they are potentially subject to different types and degrees of CLI. In the following chapter, German, Turkish, and English, the three languages that are part of the studies in this dissertation are introduced and briefly compared, after which the specific features selected for the empirical part of this dissertation are presented.

In this dissertation, I test the acquisition of English by Heritage Speakers of Turkish who grew up, went through formal education and reside in Germany. In line with the terminology that has been established in Chapter 2.1, Turkish is called their *Heritage Language* (HL), German their *Second Language* (L2), and English their *Target Language* (TL) or *Third Language* (L3). HSs are a useful group to study since they form a relatively homogeneous group (or at least provide relatively homogeneous subgroups of reasonable size) among multilinguals. They acquire both their HL and their L1 relatively early in life (pre-critical period), and many of them are regular users of both their languages (see Chapter 2.1.3). The language combination Turkish-German-English was chosen for two reasons: Due to the Gastarbeiter-movement in the 1960s and 1970s, during which tens of thousands of Turkish workers emigrated to Germany (Höhne, Linden, Seils, & Wiebel, 2014, pp. 3–7), Germany is now home to a large community of Turkish families of second- and third-generation immigrants. Since the first-generation immigrants were usually working-class and kept to themselves (socially), they often learned little German or none at all. Thus, their children, while growing up and schooled in Germany, usually did not lose their Turkish since it has remained the primary family language and often the only language of communication with their parents. Consequently, both languages can be expected to be active in the Turkish-German bilinguals from such families when they start learning English in school.

3.1 Typology

Along with languages such as Uzbek, Azerbaijani, Tatar, and Turkmen (Johanson, 2001, p. 5), to name just a few, Turkish belongs to the Turkic language family, whose members are structurally similar to each other despite considerable linguistic changes in individual languages (Johanson, 2006, p. 81). According to Kornfilt (2009), for instance, all Turkic languages are agglutinative, and Turkish also shares its basic word order pattern, S-O-V, with most other Turkic languages. Another common feature in this language family, shared by all Turkic languages except Uzbek (Pereltsvaig, 2017, p. 156), is stem-based vowel harmony, where suffixes will harmonize with the stem to which they are attached.

Both English and German, on the other hand, are West Germanic languages. Like Turkic languages, Germanic languages share a number of structural features due to their common ancestry, which include, for instance, “the introduction of verb-second word order”

(Harbert, 2007, p. 6), “the introduction of definite articles [and] the reduction of the system of inflectional tenses to a simple contrast between non-past and past” (Harbert, 2007, p. 7). However, English has been claimed to be the “typological outlier” by Harbert (2007, p. 13), since, in many cases, it deviates from the other members of the family, having lost a large number of its Germanic structures.[19]

Summing up, in an overall genetic typological account, German and English are assumed to be closely related and thus similar, whereas Turkish would be structurally distinct from the other two.[20] However, the presence and absence of (the same) phonological features or phonetic characteristics of a particular feature cannot necessarily be predicted by genetic relationships alone. The interdental fricative /θ/, for instance, is part of the phoneme inventory of English (West-Germanic), Modern Standard Arabic (Semitic), and Spanish (Romance), but not in German (West-Germanic), Modern Hebrew (Semitic) and Italian (Romance). While a speaker of Spanish can, in theory, make use of their L1 system when acquiring the interdental fricative in English, a native speaker of German cannot, even though their L1 is more closely related to the TL. The same is true for certain vowels in Turkish, German, and English. While German and English are the two languages of the triplet that are most closely related, Turkish and German both have the vowels /œ/ and /y/, while English does not. In other words, the overall genetic relationship between languages cannot serve to reliably predict transfer. Instead, it is necessary to categorize individual bits or structures according to their presence/absence or their behavior in the background languages and the target language. The phonological features chosen as testing grounds for the two studies in this dissertation on precisely this basis are introduced in the next section.

[19] For a detailed account and discussion, see McWhorter (2002).
[20] The TPM (see Chapter 2.2.1) would, then, predict transfer from German but not from Turkish in L3 English.

3.2 Phonological bits: a cross-linguistic comparison

In order to tease apart the influence of potential transfer sources, it was deemed useful to choose phonological structures from English that fit into three distinct *feature types* (FT) phonology: *FT-G* contains structures that are present in German, but not in Turkish, *FT-T* contains structures that are present in Turkish, but not in German, and *FT-E* contains structures that are present in neither of the two background languages. This means that FT-G is susceptible to transfer from German onto English (available to both monolingual and bilingual speakers in this dissertation), FT-T is amenable to transfer from Turkish onto English (available only to the HSs), and FT-E allows transfer from neither of the two background languages. This categorization, then, will provide a basis for predicting various potential transfer scenarios. Before introducing transfer scenarios based on the literature review given in 2.2 and making predictions based on these scenarios, the next section will provide some background information about the features tested in this study, concerning cross-linguistic differences, but also touching upon controversies in the literature. Table 2 shows a full list of the structures classified into their respective FTs.

Table 2: Phonological features grouped into feature types

Feature	German	Turkish	FT
initial CC	✓	✗	FT-G
phonemic vowel length	✓	✗	FT-G
/l/ - /ɫ/	✗	✓	FT-T
voiced coda	✗	✓	FT-T
/oʊ/	✗	✗	FT-E
/θ/	✗	✗	FT-E
/æ/	✗	✗	FT-E

3.2.1 FT-G: German as a potential transfer source

Word-initial consonant clusters

In general, complex onsets, i.e., syllables that contain more than one consonant segment, are more marked than simplex onsets (Topintzi, 2011, p. 1285),[21] evidenced by the fact that syllables with simplex onsets are generally preferred over syllables with complex onsets (Blevins, 2006, p. 334). Furthermore, a language that allows complex onsets will, by implication, also allow simplex onsets (Kager, 2010, p. 96).

Word-initial consonant clusters do not occur in Turkish. In loan words from other languages like English that include initial consonant clusters, Turkish speakers break up the cluster by adding an epenthetic vowel, English *train*, for instance, is loaned into Turkish as *tren*, and is frequently produced [tiren] by Turkish speakers (Clements & Sezer, 1982; Kabak, 2011). This constraint on syllable structure also frequently surfaces in Turkish learners acquiring English as the disallowed L2 syllable structure is resolved through vowel epenthesis in learner language as well (Yavaş & Altan, 2016, p. 3).

On the other hand, just like in Germanic languages in general, consonant clusters in onset position are allowed and frequently occur in German (van Oostendorp, 2019). While generally, the same types of clusters are allowed in both English and German, in that a similar degree of complexity occurs, and onset structure broadly follows sonority based restrictions, "English is the least permissive of all Germanic languages" (van Oostendorp, 2019). According to König and Gast (2012, pp. 38–39), complex onsets that are permitted in German but not in English are the combination of plosive and nasal (/kn/, /gn/, /pn/), which was historically present in English, but which have been reduced to /n/ in modern English, as well as the

[21] Complex onsets are treated as actual clusters in this dissertation. Theoretical approaches that assume an underlying CV structure in all syllables (see, for instance, Lowenstamm 1996) will be disregarded.

combination of plosive and fricative /ps/. English and German behave differently when it comes to the combination of fricatives and liquids or nasals, where some combinations are partially complementary in the two languages (/ʃl/ in German vs. /sl/ in English), whereas the languages converge in allowing /ʃr/ but not /sr/ (König & Gast, 2012, pp. 38–39).

Contrastive vowel duration

Contrastive vowel duration is relatively frequent across the languages of the world, and much more frequent than other durational oppositions, such as consonant duration (Gordon, 2016, p. 51). English and German share the feature of phonemic vowel length. The phonological reality of the categorical vowel length distinction, however, is disputed since they usually co-occur along with acoustically different vowel quality (i.e., the long version of a vowel and the short version of the same vowel do not share the same formant structure), a contrasts which, based on articulatory accounts, is captured by the classification of vowels as either tense (“long”) or lax (“short”). While this classification is well motivated by both variationist and diachronic accounts (Durand, 2005, p. 85), and stances on the usefulness of either categorization for phonological analyses are up for discussion, the physical existence of contrastive vowel duration can hardly be disputed (see Russ, 2010, p. 92 for German). At the very least, it has to be acknowledged that vowel length has a similar status in German and English, irrespective of whether it coincides with a physical difference in vowel quality.

Van der Hulst and van de Weijer (1991, p. 18) sum up that, in Turkish, “the vowel length opposition is marginal”. In other words, while a vowel length distinction is realized phonetically, for instance as compensatory lengthening that results in so-called *secondary long vowels* (Kabak, 2004, p. 363), it is not contrastive. Compensatory lengthening affects vowels preceding a dropped /h/, /y/, or /v/ in syllable-final position as well as those that precede an underlying /ɣ/ (see Sezer, 1986 and Kabak, 2007 for detailed accounts). The alternation furthermore occurs in short-long alternations in some lexical items (Göksel & Kerslake, 2005, p. 19), and additionally in a number of Persian and Arabic loanwords that retain the original long vowel in the target language (Göksel & Kerslake, 2005, p. 12). Thus, the distinction between short vs. long vowels is not relevant for lexical access for a speaker of Turkish, while it is for a speaker of German. In a study on vowel discrimination, Nimz (2015) tested Turkish and German young adults to determine whether contrasts in duration or in quality were the main source of problems for Turkish learners of German. She found that if vowels differed only in length, Turkish learners did not perform worse in a discrimination task than German native speakers. They did, however, perform worse than their German peers on the quality distinction in what Nimz calls the "u"-group, which she attributes to the fact that Turkish /u/ is equally far away from German /ʊ/ and /uː/, which, in turn, leads to equivalence classification and thus poor discrimination. Nimz's study gives evidence for the fact that it is not a pure length distinction that leads to learner mistakes in Turkish learners acquiring a language that employs contrastive vowel length, i.e., quantity, but at least a combination of both quantity and quality cues to the contrast.

3.2.2 FT-T: Turkish as a potential transfer source

Laterals

The velarized lateral [ɫ] has been shown to be relatively more marked than its palatal equivalent [l], evidenced by the fact that it is less common in the languages of the world and by its instability in synchronic and diachronic variation (Sönning, 2017, pp. 99–100). In English, both these variants of the lateral /l/ are present, the palatal "clear" one and the velarized "dark" one, which, traditionally, are presumed to be in complementary distribution, with the clear [l] occurring before vowels and /j/, as well as a devoiced variant [l̥], after voiceless plosives and fricatives, and the dark allophone occurring elsewhere (see, for instance, Cruttenden, 2014, p. 217). Acoustic analyses have revealed, however, that the distinction between the two allophones is gradient rather than categorical (see, for instance, Sproat & Fujimura, 1993, and Hayes, 2007). More recent empirical evidence has shown that categorical distribution, in fact, co-occurs with gradient phonetic effects. Specifically, ultrasound studies, have revealed that the degree of velarization is conditioned not only by the phonological but also by the phonetic environment the segment occurs in, as well as by morphosyntax (see, for instance, Lee-Kim, Davidson, & Hwang, 2013 for American English, and Turton, 2017 for different dialects of British English). Crucially, however, velarization is position-sensitive (as explained above), and the expected phonetic realizations of the two variants, as well as the acoustic differences between them, are conditioned by the variety it occurs in.

In Turkish, which also employs both a palatal, i.e., clear [l] and a velarized, i.e., dark [ɫ], the distribution of the two laterals is conditioned by its position as well as by the vocalic environment in which it occurs: In word-initial position, the lateral is realized by its clear allophone (Clements & Sezer, 1982, pp. 236–238). In other positions, the lateral is conditioned by the quality of the adjacent vowel and surfaces as [l] in front vowel contexts and as dark [ɫ] in back vowel contexts. The occurrence of the clear lateral in back vowel contexts is possible in some loanwords (Göksel & Kerslake, 2005, p. 9), which yield a few minimal pairs, e.g., [sol] "musical note" vs. [soɫ] "left".

So, while both English and Turkish make use of both a clear and a dark allophone of /l/, their distribution is different, which means that, while speakers of both languages, actually physically produce this sound and thus share the mastery of the articulatory gestures involved, they differ when it comes to the distribution of /l/, which may translate into different perceptual patterns across speakers of different L1s. German does not make use of two different lateral allophones but uses the clear variant in all contexts.[22]

Voiced final obstruents

Voiced obstruents are generally more marked than their unvoiced counterparts, even more so in syllable- or word-final position (Edge, 1991, p. 378). In German, the voicing opposition is categorically neutralized in final position. In fact, the devoicing of voiced obstruents in word- and morpheme-final position is one of the most stereotypical and most robust learner mistakes made by German learners of English, where a voicing opposition in obstruents generally exists in all syllable positions. Whether or not the voicing opposition is retained in English depends on the event following the segment in question: If followed by a pause or an unvoiced segment, the syllable-final obstruent is highly likely to be partially or fully devoiced, whereas voicing is more likely to be retained before a voiced segment, i.e., a vowel or a voiced consonant. However, even in the case of phonetic devoicing, "the affected segment is still perceived and interpreted as the voiced member of a pair of obstruents" (Brockhaus, 1995, p. 3). In German, on the other hand, final obstruents are categorically devoiced in all environments, i.e., irrespective of whether the word is followed by a vowel segment, a consonant segment, or a pause (Wiese, 1996, pp. 200–206). This is a very robust process that also holds for loan words (e.g., English /ʤɒb/ surfaces as [ʤɒp] when borrowed into German (Russ, 2010, p. 171), and these devoiced consonants are perceived as voiceless by speakers of German (Brockhaus, 1995, p. 3). In Turkish, voiced obstruents are possible in

[22] Note that this is the case in Standard German as well as most non-standard varieties of German. A small number of varieties, one of which is the variety spoken in the area around Cologne (Kohler 1995), have the dark rather than the clear lateral. This issue, especially the way it pertains to the studies conducted for this dissertation, will be briefly discussed in Footnote 30 in section 4.1.1.

word- and syllable-final position. However, similarly to German, in some loan words, word-final voiced plosives are devoiced. Arabic *kitab*, for instance, becomes *kitap*[23] when borrowed with a voicing alternation when suffixes are added. While native speakers have been found by Becker, Ketrez, and Nevins (2011) to apply this voicing, or laryngeal alternation, to nonce words (crucially conditioned by lexical statistical information), which suggests that the laryngeal alternation found in numerous lexical items in Turkish is, in fact, productive, there are several cases of final voiced obstruents in Turkish words (e.g., *ad* = name, *ab* = water, *otuz* = thirty) and in loan words (e.g., *katalog*, *lig* = league). Thus, while phonetic devoicing processes exist in English, Turkish, and German, they are exceptionless only in German, and should thus also be more robust and persistent in language acquisition. Summing up, both Turkish and German (like English) allow voiced obstruents in general, whereas their occurrence in final position is absolutely restricted in German (i.e., they never occur, irrespective of lexical item or context) but permissible in Turkish.

3.2.3 FT-E: No potential transfer source

The diphthong /oʊ/

Diphthongs are defined as a vowel glide that moves from a source to a target sound. While, typologically, complex nuclei (long monophthongs and diphthongs) are more marked than simple monophthongs, long monophthongs and diphthongs are not in an implicational relationship (Gordon, 2016, p. 108).

The German phoneme inventory has three diphthongs, i.e., vocalic sounds in which "the position of the tongue changes during the articulation" (Russ, 2010, p. 35): [aɪ], [aʊ], and

[23] Note that the devoicing process is sometimes reflected in the orthography, where, for instance, <b> alternates with <p>. This may lead to a higher degree of awareness for this specific process in literate speakers.

[ɔɪ], whereas Turkish does not know any diphthongs in its vocalic system (see Göksel & Kerslake, 2005). Standard British English has five closing diphthongs (Cruttenden, 2014, pp. 140–149) and some centering diphthongs. In the latter case, the target sound of the glide is the vocalized variant of an underlying rhotic. The English diphthong [oʊ] constitutes a compelling case as its phonetic reality, i.e., this particular combination of sounds, is found neither in Turkish nor German, while diphthongs in general are phonologically possible in German, but not in Turkish.

Voiceless interdental fricative

The interdental fricative is both cross-linguistically rare (Maddieson, 2013) and relatively difficult to articulate. Even in those languages that do employ the interdental fricative, this sound is often substituted in non-standard varieties, and it is commonly acquired relatively late by L1 speakers of the respective languages (see Kabak, 2019, pp. 228–230).

The English voiceless interdental fricative /θ/ occurs in neither German nor Turkish and can be classified as a "new sound" according to Flege's Speech Learning Model (Flege, 2007). A common strategy to deal with this type of sound is substituting it with sounds that are similar in some way (acoustically, articulatorily, etc.) and that occur in the native language of the learner. Substitution patterns across languages differ, however, even in cases where the same potential substitutions are found in the different phonological systems. This has been explained by different feature weights found in the same segment-class across different languages, which then leads to different substitution patterns.[24] While German is considered a primarily [s]-substituting language, Turkish speakers of English are mostly found to substitute the interdental fricative with [t]. A third potential substitution is [f], which is the sound that is acoustically

[24] For a detailed analysis, see Brannen (2011).

closest to the target feature (Johnson, 2012, p. 121), which is, thus the most difficult to discriminate from the interdental fricative (Kabak, 2019, p. 239) and which, for that reason, is also found as a substitute both in L1 acquisition and in some native varieties of English. This differential substitution, then, can often be used as an L1 shibboleth in non-native accented English.

<u>*The vowel /æ/*</u>

The vowel inventories of English, German, and Turkish are different from each other both when it comes to overall shape and as well as crowdedness, i.e., the number of discrete segments in the system.

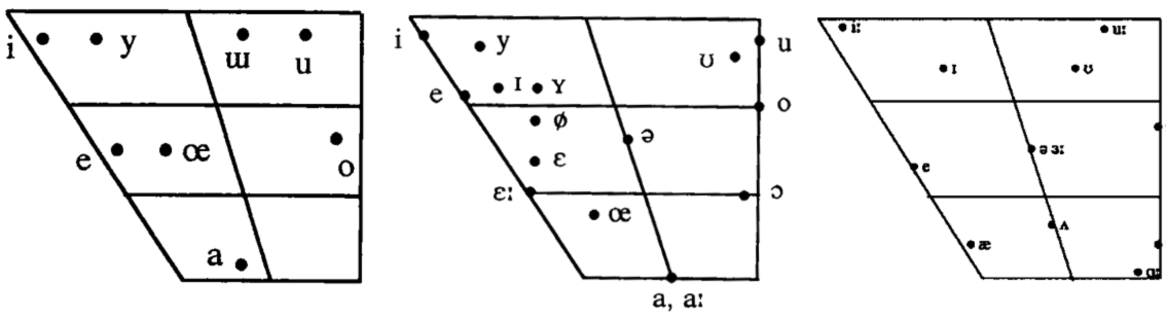

Figure 1: Vowel spaces of Turkish (Zimmer & Orgun, 1999, p. 155), German (Kohler, 1999, p. 87), and English Received Pronunciation (Roach, 1999, p. 242)

Figure 1 shows the vowel charts for Turkish, Standard German, and English Received Pronunciation. It can be seen that German and English employ vowel systems that are relatively wide, with German being closer to the extremes than English. Turkish, on the other hand, employs an overall more central system, of which large parts are not populated. English, while still having a system that is less crowded than that of German, exhibits a relatively even distribution across the whole vowel space.

Crucially, neither German nor Turkish have a vowel that is close to /æ/, which is lower than the closest equivalents in the two languages. However, since there is an equivalent for the specific vowel to be mapped onto, it can be classified, according to Flege's (1993) terminology, as a similar sound. Consequently, it can be expected to be particularly challenging to acquire by speakers of both background languages. Specifically, German speakers are known to substitute /æ/ with their native [ɛ], which can also be observed in the phonological adaptation of English loanwords in German.[25] A comparable substitution for /æ/ is also known for Turkish learners of English, especially apparent through the adaption of English loan words in Turkish (where 'patchwork' is adapted to Turkish as 'peçvörk', for instance). The grapheme <e> is realized as [ɛ] or [e], with a high degree of both regional and idiosyncratic variation.

Vowels, in general, can be expected to be especially difficult to acquire for two additional reasons: Firstly, if a vowel [V] is produced too far from the ideal vowel /V/ by an L2 speaker of a language, the production might be produced as non-target-like. However, if the acoustic parameters of the vowel [V] do not fall into the parameters of a different vowel phoneme in the L2, this mispronunciation may not trigger problems in terms of comprehensibility, and thus the speaker may not repair the non-target-like vowel in their L2 system. While this reasoning can generally apply to all phonemes, the effect is more likely to be active in vowels than in most consonants since the concrete acoustic characteristics of vowels are more gradient than those of consonants. Furthermore, there is no articulatory constriction involved in the production of vowel sounds. To be more precise, articulators do not touch, and thus there is little haptic information, which can be made use of in the acquisition of stops, for instance. Both these characteristics make it more challenging to produce a target-like version of a vowel sound.

[25] According to Hausmann (2006), orthographic integration into the native system has, in fact, been exceedingly slow in very recent loans, i.e., English orthography is retained for a relatively long period.

3.3 Research objectives and transfer scenarios

As shown in Chapter 2.2, studies on the acquisition of additional languages by monolingual and multilingual learners have produced highly diverse results that sometimes seem to contradict each other. Note that this is most likely due to the following reasons: Usually, only individual features are tested, which disregards the complexity of phonological grammars and the diversity of the phonological bits therein (see Chapter 2.2). Furthermore, it is often a specific model that is being tested using research designs that are not fit to test any of the other models. This dissertation will attempt to remedy these two methodological issues by using the seven phonological features detailed earlier in this chapter. The objective is to identify the transfer scenario at play by following a number of steps. First, the predictions made by the classical models used in the research of L3 Acquisition will be tested. Specifically, this will be done by collapsing multiple or slightly modifying these existing models. The first possible scenario is **single language transfer (SLT)**, whereby only one of the background languages is activated and thus available for transfer. Both negative and positive transfer are possible. This collapses the predictions of the single-language models proposed in the literature, e.g., the TPM (Rothman et al., 2010), the L2SFM (Bardel & Falk, 2012), as well as potential exclusive transfer from the L1. There can be exclusive transfer either from the 2L1 (German) or the HL (Turkish) in the bilingual speakers. If the exclusive transfer source is the bilinguals' L2, one would not expect a difference between Turkish-German bilinguals and German monolinguals in either of the feature types since they would all rely solely on their shared background language, German. If, on the other hand, the HL language is the exclusive transfer source, a difference in FT-G would be expected, where the monolinguals would outperform the bilinguals, as well as a difference in FT-T, where bilinguals would outperform the monolinguals.

The second potential scenario is **combined beneficial transfer (CBT)** as proposed by Flynn et al. (2004) in their CEM. Both background languages are available for transfer, transfer takes place on the level of the individual feature (bit-by-bit transfer), and, in case of competing input from the background languages, transfer will be beneficial. In this case or, in other words, if the speaker makes use of what is helpful to them in their language system, both groups would be expected to perform equally well in FT-G since they will both be able to draw on their German. The bilinguals will outperform the monolinguals in FT-T since only they benefit from their Turkish. Neither of the two scenarios make any predictions about FT-E since they only consider transfer of complete features. Thus, since in order to perceive and produce features from FT-E, the speakers can draw neither on full features from German nor from Turkish, no difference is expected across groups. Instead, any difference found in FT-E may be due to extralinguistic factors or universal tendencies. Furthermore, if the bilinguals outperform their monolingual peers in the perception or production of FT-E, this could also be due to a more general bilingual advantage in the acquisition of an L3 phonology.

The final scenario is **bit-by-bit transfer**. Here, both background languages are available for transfer, transfer takes place on a bit-by-bit level, and can be positive or negative. Presence/absence and the source of transfer are crucially conditioned by language-specific as well as universal properties (phonemic vs. allophonic, functional load, markedness, etc.). In case of competing information from the two systems, different properties are weighed against each other to determine the outcome. This draws on the ideas proposed by more recent models, i.e., the SM (Slabakova, 2016) and the LPM (Westergaard et al., 2016) as well as on the ideas developed in Section 2.2.7. In this third scenario, there will be a difference between monolinguals and bilinguals. This difference, however, will be based on language-specific patterning of the feature in either or both of the speakers’ background languages. Whether or not a feature is perceived or produced differently across groups will depend on language-specific and universal factors that may interact with each other additively or subtractively, and that are weighed against each other. As a result, there will be no straightforward categorization according to feature type.

Crucially, however, CLI is not only the straightforward copying of L1 patterns onto a target language but also includes more subtle manifestations of CLI in an L3 phonology. To uncover such effects, it is necessary to analyze multiple bits that differ not only on a binary dimension as *present* or *absent* from one or both of the background languages, but also on different parameters like their phonemic status in the languages, their functional load, etc. To sum up, the phonological grammar of the L3 learner is examined as a system by scrutinizing a subset that is as diverse as possible while at the same time considering the individual bits as single entities. This way, I aim to trace both larger- and smaller-scale patterns. Since the perception and the production of a phonological bit may exhibit different degrees of sensitivity to transfer effects, it will be necessary to test both domains for all bits.

In addition, a potential influence of extra-linguistic factors is tested. Specifically, these factors will be Phonological Working Memory (PWM) and motivation in Study 1 (Chapter 4) and language dominance as well as L1 and HL proficiency in Study 2 (Chapter 5) to see whether and how CLI is modulated by these. If, as has been claimed, bilinguals enjoy a general advantage over monolinguals due to cognitive advantages, the bilinguals in the studies of this dissertation would be expected to outperform the monolinguals in connection with higher PWM. Motivation (Dörnyei & Czisér, 1998) and language dominance (Montrul, 2010) have been shown to be related to successful acquisition, which is why the potential effect of these variables as well as the effect of proficiency in the background languages are tested here. In addition, the age factor is compared across the two studies.

The next two chapters comprise two empirical studies. In Study 1, young prepubescent learners of English with HL Turkish and L2 German were tested on their perception and production of the phonological features introduced above, and their results were compared to a matched L1 German group acquiring English as their L2. In Study 2, two groups with the same language background were tested on the same features, but they were adult speakers of their L2 (for the monolinguals) or L3 (for the bilinguals) English. After the analyses, the results of the two studies are discussed and compared to each other. This comparison is meant to shed light on potential developmental tendencies in L3 acquisition, and differences between the way background languages influence the TL at different stages of acquisition.

4 Study 1: Monolingual and bilingual young learners of English

4.1 Methods

In the following sections, I first introduce the participants of this study and then formulate the research questions that are to be answered in order to test the hypotheses posed in Chapter 3.3. Next, the materials and procedures used to answer the research questions are introduced in detail, and the variables tested are defined. This subchapter closes with an explanation of the tools and methods used to code and analyze the data.

4.1.1 Participants

Two groups of pre-pubescent children were tested: *Group 1* (bilingual)[26] included 12 pupils at a German public high school who had grown up with Turkish parents. Their HL was Turkish, and they had started acquiring German as their L2 when they entered formal education. None of the participants in Group 1 knew any additional languages apart from Turkish and German. In all cases, both parents were Turkish L1 speakers, some without any knowledge of German, some with basic knowledge of German, and some with advanced to native-like proficiency. However, irrespective of the parents' proficiency, the participants reported using Turkish whenever talking to them. While Turkish was the primary home language for all the

---

[26] Henceforth in Study 1, the term *Group 1* will always refer to the bilinguals and the term *Group 2* will always refer to the monolinguals. When referring to individual participants, I will use the participant code, which consists of a B or an M (for bilingual and monolingual, respectively), followed by a K (for kids, meaning they are participants in Study 1 rather than Study 2) and finally ending in a sequential number. Note that in Group 1, numbers 01 and 02 referred to bilinguals who participated in a test run of the experiments and who were thus excluded from the final sample. Participants in Group 1 start with BK03, and participants in Group 2 start with MK01.

participants, some of them used a mix of German and Turkish with their siblings.[27] As such, Turkish can be regarded as their HL (as defined by Montrul, 2015, see Chapter 2.1.3 for a detailed review on terminology and categorization), and I will henceforth refer to their Turkish as their HL and the speakers as HSs. At the time of testing, the HSs were in the process of shifting their dominant language from Turkish to German since life outside the home domain, i.e., social contacts at school, in sports or music clubs, etc., were gradually becoming more important in the participants' lives. Thus, the role of German as a language of communication and social connections was gaining importance for these speakers.[28] The subjects in Group 1[29] were between 10 and 13 years old (mean age=11.5, sd=0.8), and they all attended grade 5 (n=5) or 6 (n=7) at the same public high school (Gesamtschule) in Düsseldorf, North Rhine-Westphalia, where they participated in regular English lessons. There were both male (n=4) and female (n=8) participants in the sample, and all of them were right-handed.

Group 2 (monolingual), the control group, included 12 pupils at a public high school (Gesamtschule) in Duderstadt, Lower Saxony. They were between 10 and 13 years of age (mean age = 11.33, sd = 0.94) and attended grades 5 (n=3), 6 (n=5), or 7 (n=4). There were both male (n=3) and female (n=8) participants, and all but two subjects were right-handed. All subjects in the control group had grown up with only German in their homes, i.e., their parents were native speakers of German, and English was the first foreign language they were acquiring. None of the participants in the control group had any knowledge of Turkish or any other additional language.

---

[27] All this background information was assessed in informal conversations before and after the testing sessions, as well as between tasks. These conversations were not audio recorded, but the interviewer took notes regarding the general biographical and linguistic background of the participants.
[28] This information was assessed during informal conversations before, between, and after the tasks, and it fits the language profile usually found in HSs (see 2.1.3 for more details).
[29] See Appendix A.1 for detailed speaker profiles.

The two groups came from different areas whose native dialects can also be considered somewhat different from each other in terms of accent: The dialect spoken in Lower Saxony is phonologically and phonetically close to Standard German, while dialects spoken in North Rhine-Westphalia employ features that are very distinct from the standard German variety.[30] However, during the informal conversations that took place before, between, and after the tests, the researcher assessed the participants' spoken German and found all of them to speak accents close to standard German. The participants from North Rhine-Westphalia (i.e., the bilinguals in this study) might still be more familiar with the dark variant of the lateral than the participants from Lower Saxony, which might have an effect, particularly in the perception test. Such potential dialectal differences will be considered in the discussion.

4.1.2 Specific research questions

As outlined in Chapter 2.1.2, the languages in a speaker's mind are known to exert an influence on each other, both in production and perception, and both in the acquisition process and in fully acquired systems.[31] However, the exact mechanisms at play in the interaction between the languages in the multilingual's mind, as well as the factors that modulate an initial intrinsic interaction are not yet fully understood. It is believed that the covert interactions can be made visible, at least in part, by looking at the output of an L3, and the type as well as the

[30] This was a concern especially for the lateral distinction: In the dialect spoken in the Cologne area of North Rhine-Westphalia, a darker version of the lateral is generally used. (Kohler, 1995, p. 164) Note that there is little descriptive, let alone empirical literature on synchronic phonological variation in German. Most claims are based on anecdotal evidence. In any case, no deviation from the standard variety was perceptible in the participants' German as assessed by the experimenter. Furthermore, Cologne (the specific area for which the velarization of laterals is attested) is, in fact, in a dialect region distinct from Düsseldorf, where the study was conducted (see Russ, 2010, p. 204).

[31] Both native and non-native systems can be affected by CLI. Here, only the influence on the non-native languages will be an issue. For a detailed overview over the influence that the languages in the multilingual system (can) exert over a native system – a process usually referred to as *language attrition* – see, for instance, Stoehr, Benders, van Hell, and Fikkert (2017) and Yilmaz and Schmid (2018).

degree of influence that the background languages exert on the TL. Thus, in order to understand the shape of and mechanisms active in the multilingual language system(s), the respective influence of the background languages on the TL needs to be identified and explained. In order to do so for the Turkish-German HSs in this study, I ask the following research questions:

(1) Do young Turkish-German HSs differ from their monolingual peers in the perception and/or the production of phonological bits if their background languages supply conflicting transfer information, i.e., if a bit can be transferred positively from one, but negatively from the other of the two languages? What happens if neither of the background languages is a potential source of transfer (as it is foreign to both background grammars)?

(2) Can these potential differences in perception and/or production be explained by SLT or CBT?

(3) Are production and/or perception modulated by cognitive (phonological working memory) and/or affective (motivation/attitude) variables, and is there an interaction of Group with any of these secondary variables?

(4) Is there a correlation between perception and production in general or in individual features?

4.1.3 Material and procedure

4.1.3.1 General setting

Both groups were tested in their respective schools in quiet office or library rooms. The tests took place on two consecutive days for Group 2 and on three days in two consecutive weeks for Group 1. Each subject was tested individually, and the testing procedure took about

45 to 60 minutes for each of the participants, including small breaks between tasks. All tests were administered by a German native speaker in German[32] in the following order:

1. Perception task: AX discrimination
2. Production task: delayed repetition
3. Phonological working memory I: nonsense word repetition
4. German and Turkish samples: picture naming
5. Phonological working memory II: digit span task

The test items in the AX task (nonsense words, see below) and in the delayed repetition task (real English words forming novel compounds, see below) were English, while the words that had to be repeated in task 3 were nonsense words following German phonology. In task 4, real German and Turkish words (in that order) were elicited in a picture-naming task. Since task 5 did not require the subjects to make oral utterances, but rather to indicate what they remembered by typing numbers on the computer keyboard, they were not restricted to using a single one of their languages if they employed articulatory rehearsal (vocal or sub-vocal) in order to solve the task (which is the strategy expected in this type of task to keep the items in the working memory, see Baddeley, 1992).[33] Thus, by administering the tasks in this particular

[32] While there might be recency effects or effects of internal phonetic convergence (similar to the effect found in code-switching situations, see, for instance, Bullock, 2009; Toribio, Bullock, Botero, and Davis, 2005), it was not possible to test the subjects with English as language of instruction due to their low proficiency. Since the subjects were tested in a German (rather than Turkish) speaking environment (and an environment that would trigger the use of their L2 rather than of their HL), the additional effect of the German interlocutor can, however, be considered negligible.

[33] The fact that the speakers can choose the language for articulatory rehearsal may in fact skew the results due to the word length effect (see Baddeley, 1992): Memory span for a word and the time it takes to utter that word are inversely correlated. This means that if the items that have to be remembered are longer in language A than in language B, all other things being equal (proficiency in both languages, language dominance, recency, etc.), a speaker who chooses language B will have an advantage over a speaker that chooses language A. In this study, however, no participant used Turkish for articulatory rehearsal instead of German (the participants were asked about their choice after all tasks were finished), which also made possible the comparison within and across groups.

order, the individual languages the participants had to access were kept steady for the longest possible time frames.

4.1.3.2 Perception: discrimination task

In order to test the subject's perception of the test features, an AX discrimination task was administered. To that end, 14 English pseudowords were designed, 2 for each test feature. They were all paired with an equivalent that included the expected learner mistakes. Each of the items for the interdental fricative was paired with three such equivalents since there were three expected substitutions attested in the literature ([f], [t], and [s]; for a detailed discussion see 3.2.3). For a list of the pseudowords, see Table 3.

Table 3: Pseudoword patterns for perception experiment

Feature	Stimuli
initial CC[34]	/ʃrɪs/ - /ʃɪrɪs/; /krʊs/ - /kərʊs/
phonemic vowel length	/nɑ:f/ - /nʌf/; /jɑ:f/ - /jʌf/
/l/ - /ɫ/	/lʊp/ - /ɫʊp/; /tʊl/ - /tʊɫ/
voiced final obstruent	/niːb/ - /niːp/; /viːd/ - /viːt/
Interdental	/pəθiːt/ - /pətiːt/ - /pəsiːt/ - /pəfiːt/ /kəθip/ - /kətip/ - /kəsip/ - /kəfip/
/æ/	/pæf/ - /pɛf/; /tæs/ - /tɛs/
/oʊ/	/loʊn/ - /loːn/; /roʊn/ - /roːn/

All pseudowords and their counterparts were recorded by a female native speaker of Standard British English. Each of the items was recorded eight times, and from those eight

[34] Both clusters used in the pseudowords are licit in English as well as German (see Chapter 3.2.1).

recordings, four were chosen.[35] The items were then checked in Praat: Specifically in the cluster pairs, the relative duration of the epenthetic schwa was measured and manually adjusted, if necessary.

Using Praat, the final items, preceded by 4 item pairs for the training round, were then presented to the participants in the following way: The participant was seated comfortably in front of a computer screen wearing Philips SBC AH1000 headphones. They heard an item through the headphones either paired with a different instance of the same item or paired with an instance of its equivalent item. The items were presented with an inter-stimulus interval (ISI) of 1500ms[36]. Participants were then prompted by two symbols as well as the words “Ist das das gleiche Wort?” on the computer screen to indicate whether they thought the items they had heard were the same or different by pressing the respective key on the computer keyboard (S for same, L for different). Once they had pressed a key, the next item pair was presented in the same way. All possible permutations per item type were played (AA, AB, BA, BB) and each of the permutations contained a different instance of the item, i.e., the A in AB was a different recording of A than the A in BA, and both same-permutations consisted of two different tokens. This was different only for the interdental fricatives, where each test item had three counterparts to be paired with (see above), making for 12 stimuli pairs for each item of the interdental fricative. This way, there was a total of 72 stimuli pairs with three repetitions for each of them. The stimuli pairs were presented in a randomized order in 6 trial blocks of

[35] In the case of the lateral tokens, re-recording some of the items was necessary as the two allophones were not distinct enough and the speaker had difficulties producing the clear allophone in the context that requires the dark allophone in her variety and vice versa. The re-recordings were done by the same speaker in the same room using the same equipment in the week immediately following the initial recordings.

[36] Long ISIs have been shown to trigger phonemic rather than phonetic discrimination, whereas lower ISIs enable the listener to rely on auditory processing (Werker and Logan, 1985). Since this experiment is meant to test phonological acquisition, which crucially includes the phonemic categorization of non-native speech sounds, a relatively high ISI of 1500ms was chosen.

36, between which the participants could take a break. Both +/- correct and reaction time was recorded in Praat and later extracted into an Excel file where the data was further analyzed (see below).

4.1.3.3 Production: delayed repetition task

In order to elicit English speech from the participants, a delayed repetition task was designed. The subjects were required to combine two English words to form a novel compound. Four compounds were designed for each of the phonological features to be tested, which made for a total of 28 compounds. The feature in question always appeared in the first constituent of the compound, and thus, each compound included one test item. Where possible,[37] the number of syllables, syllable length, and syllable complexity were controlled for. Twenty five of the words were nouns, 2 were adjectives, and 1 was a numeral. In addition to the compounds containing one of the test items, there were another 4 word pairs for a training round that were not included in the analyses.

All items used to test initial consonant clusters included clusters consisting of a stop followed by a liquid, and they were all monosyllabic. They all had CCVC structures except for one item, which was structured CCV:.

The test words for vowel length were all monosyllabic and were structured CV:C. Three of the test words for the lateral were monosyllabic, and one was disyllabic. The lateral always appeared in coda position and thus triggered the production of the dark allophone of the lateral.

[37] To ensure that the test words were items that were known to the participants, I consulted the vocabulary sections of various commonly used school books. This approach worked for all test items that were eventually chosen (i.e., the participants indeed knew the words) except for *thumb,* which most of the participants turned out not to be familiar with. (The impact this might have had on some of the results will be discussed in the discussion section.) As a direct consequence of this requirement, however, it was not always possible to entirely control for all structural characteristics of the test words.

The test items for obstruent voicing in word-final position were all monosyllabic, 3 CVC and 1 CCVC. The obstruents tested were /b/, /d/, and /g/. They were all followed by onset-less words, the most likely environment for native speakers of English to produce full voicing in the obstruent.

Two of the test words for the unvoiced interdental fricative contained /θ/ in initial position and two of them in final position, and all of them were monosyllabic.

The test words for the low front vowel /æ/ and for the diphthong /oʊ/ were also monosyllables, all CæC and CoʊC respectively, except for one test word for /oʊ/ which was structured CCoʊ.[38]

Using a Neumann TLM102 microphone, a Focusrite Scarlett 18i8 Interface, and the recording and editing software ProTools, audio recordings of all items for use in the experiment were spoken by a 42-years old female native speaker of Standard British English.

The test was administered via PowerPoint on a laptop computer. The participants were shown a drawing of the first compound constituent and simultaneously, via Philips SBC AH1000 headphones, listened to the audio recording of that constituent twice, followed by the same procedure for the second part of the compound. They then were shown a drawing that combined the two concepts. They would, for instance, see the picture of a cross and hear the word *cross* twice, then see a picture of a lamp and hear the word *lamp* twice and finally see a drawing of a lamp shaped like a cross and simultaneously hear the native speaker say *What is this?* They were then required to produce the full compound, for instance *cross lamp*, twice (see experimental sequence in Figure 2). After they had produced the full compound, the experimenter manually triggered the start of the next sequence for the next compound using a remote control. This was done manually to give the participant as much time as they needed to

[38] For a full list of the test words, see Appendix B.1.

produce the required items. There was a training round with four word pairs that did not include

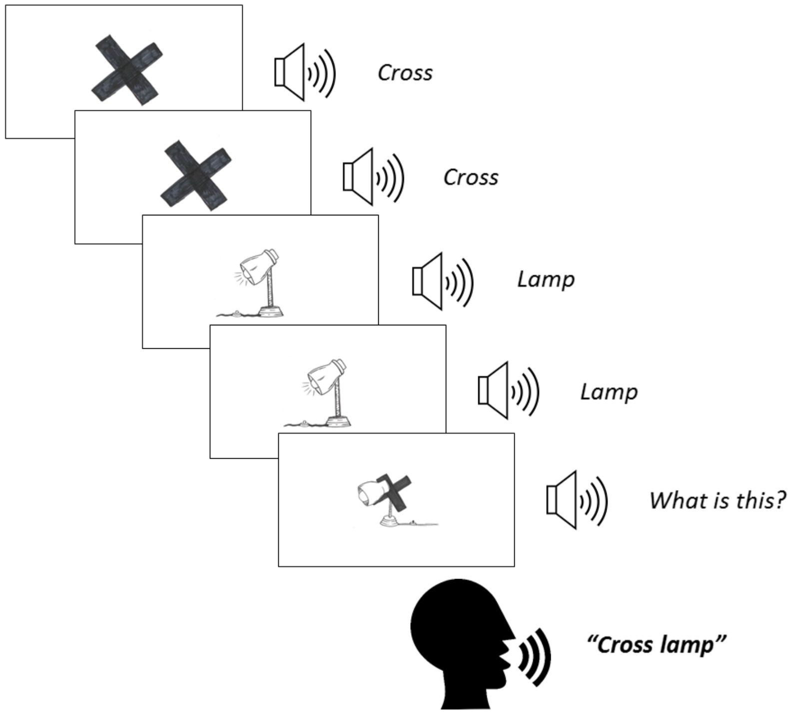

Figure 2: Setup of the delayed repetition task

the experimental features, during which the experimenter corrected responses that were not in line with the expected response, and during which the participants could pause and ask for help or clarifications at any time. This option was also given to the participants between the training round and the experiment, whereas no response feedback was given during the trial rounds.

The delayed repetition task using compounds was chosen since the participants were at the initial stages of language acquisition and thus they could not be expected to access lexical items quickly or reliably enough by themselves even if they were familiar with them. However, it was essential to avoid the participants merely parroting the words they had heard. So, for one thing, participants had to combine individual items to form a compound (which made the task more complex and cognitively more demanding), and their phonological loop was interrupted auditorily by the question "What is this?".

The participants were recorded using the internal microphones of an Olympus LS-11 audio recorder, which was placed before them on a soft piece of cloth. Recordings were made in wav-format at a sampling rate of 44.1 kHz and a bit depth of 16 bit.

4.1.3.4 Secondary variables

In addition to the primary linguistic tasks, a battery was administered that consisted of various additional tasks and questionnaires. Cognitive characteristics, such as working memory, executive functions, or metalinguistic awareness, have been widely discussed as linked to both bilingualism and success in language acquisition. Specifically, bilinguals have often been assumed to have higher cognitive abilities (see, for instance, Bialystok & Craik, 2010, or Adesope, Lavin, Thompson, & Ungerleider, 2010 for a meta-analysis), which has been claimed to translate into what researchers have called *Bilingual Advantage* (BA) in language acquisition in general (see, for instance, Kaushanskaya & Marian, 2009) and the acquisition of a non-native phonology in particular (e.g., Spinu, Hwang, & Lohmann, 2016). However, these advantages have recently come under scrutiny due to a lack of effect of bilingualism on the outcome of cognitive tasks (see, for instance, Gathercole et al., 2014; Paap & Greenberg, 2013). There is now reasonable doubt about a clear, straightforward impact of bilingualism on cognitive functions (Paap, Johnson, & Sawi, 2015), and the BA found in various studies is claimed to be a mixture of different variables that positively influence language acquisition (active video game playing, education, musical training, and aerobic exercise among others, see Valian, 2015), heightened cognitive abilities in bilinguals being only one of them (see, for instance, Valian, 2016).

Motivation has also been shown to be a factor that affects the process and the outcome of language acquisition: Dörnyei and Csizér (1998) go so far as to call it "one of the most important factors that determine the rate and success of L2 attainment" (Dörnyei & Csizér, 1998, p. 203), and the attitudes multilinguals have towards their languages are thought to influence their speech production in the respective language (see Moyer, 2013, pp. 70–71 for a brief discussion in the context of foreign accentedness), and may thus also determine the source of transfer into the L3 or the concrete realizations of phonological features in the L3. To determine these potential effects, and to eliminate a positive effect of assumed bilingual traits like enhanced working memory or motivation, the participants were tested on their phonological working memory via experimental testing, and on their motivation and attitude as well as their general language use in all their languages via questionnaires. Furthermore, German and – in the case of the bilinguals – Turkish speech samples were collected (see Section 4.1.3.4.4 for more information).

4.1.3.4.1 Motivation and attitude

In order to measure the participants' motivation and attitude, a shortened, adapted, and translated version of Gardner's (1985) *Attitude/Motivation Test Battery* (AMTB) was administered. Gardner's original test battery was designed for testing English-speaking students learning French in school in Canada. It is based on his *socio-cultural approach* to motivation and attitude, which goes beyond individual motivation as a driving force behind the learning process and instead claims that motivation may be driven by social and cultural factors. Specifically, it is assumed that a person will be more motivated to learn a foreign language if they want to adapt to the target culture. This particular approach to motivation was chosen for this study since it fits well with the multilingual and multicultural settings of the test subjects.[39] The original AMTB consists of a total of 134 statements and questions that are assessed or

[39] Note that the questionnaire was not used to assess attitudes towards the target culture/language but towards the two background cultures/languages, which makes the sociocultural component particularly relevant here.

answered using 7-point Likert scales, single choice and semantic differential assessments, and that are used to compute composite scores for *Integrativeness* (i.e., the integrative or social reasons to learn the L2), *Motivation* (i.e., individual motivation), *Attitudes* towards the learning situation, and an *Attitude/Motivation Index* (including the three subsequent scores as well as a score for classroom anxiety). While this test battery has been widely used in research on SLA in general as well as repeatedly validated by multiple research teams for different language combinations (see, for instance, Atay & Kurt, 2010 for Turkish learners of English; Cocca, Pérez García, Zamaripa, Demetriou, & Cocca, 2017 for Mexican learners of English), it must be noted that the original version is rather lengthy. Since the participants of this study were children, the test was shortened considerably: In the questionnaire for the bilinguals, I used two parts of the battery that were meant to measure motivation or attitude, respectively. The attitude subset contained statements about the participants' attitude towards the Turkish language, people and culture (n=12), their attitude towards the German language, people and culture (n=8), and their attitude towards foreign languages and language learning in general (n=10), whereas the motivation subset contained statements about the participants' motivation for learning English (n=10), their English language anxiety (n=5) and parental encouragement (n=10). The monolinguals received the same battery of statements, excluding the 12 statements about the attitude towards Turkish. Since the scores for parental encouragement in many cases did not fit the overall motivational and attitudinal profile of the participants, they were asked at the end of the experimental session why they had given specific answers. For instance, an otherwise highly motivated participant with high interest in English rated parental encouragement very low, which, according to her, was because her parents did not *have to* encourage her any further to study and learn English. Since these types of reply were relatively frequent and could potentially skew the results (i.e., lower the motivational scores of highly motivated participants), it was decided that the 10 items on parental encouragement would not count towards the total score.

The items were taken from the original AMTB and translated into German; some statements were additionally adapted to fit the cultural and linguistic background of the participants in this study. They were randomized using Excel and presented to the participants, who were asked to rate each item on a 6-point Likert scale (1 meaning “I do not agree at all” and 6 meaning “I completely agree”).[40] The scores for items that were phrased negatively were reversed and, subsequently, all scores for motivation items and all scores for attitude items were summed up to get a motivation score and two attitudinal scores (one for Turkish and one for German) respectively. The participants received the questionnaire in advance through their teacher and filled it in at school during afternoon lessons.

4.1.3.4.2 Phonological working memory (PWM)

Based on the *Phonological Memory Composite* of Wagner, Torgensen, Rahotte, and Pearson’s (2013) *Comprehensive Test of Phonological Processing* (CTOPP-2), phonological working memory was assessed using a combination of a digit span test and a nonword repetition task. The original CTOPP-2 was developed to assess the general phonological skills of individuals between 4 and 24 years of age relative to their peers. In addition to phonological

---

[40] There has been much debate about the most sensible number of response options in this type of test that are largely concerned with whether to choose an odd or an even number of response options. An odd number of options (mostly either 5 or 7) allows the participants to choose the neutral (middle) answer. This neutral response has been claimed to not be in between the two adjacent options but really a separate response category in itself that cannot readily computed into a final score alongside the more straightforward options in the scale (see Simms, Zelazny, Williams, and Bernstein, 2019 for a detailed discussion and empirical assessment). Thus, an even rather than the original odd number of options was chosen to force participants to make a choice towards the positive or the negative end of the response spectrum. A number of 6 options has been shown by Simms et al. (2019) to be the ideal number of options – large enough to provide fine-grained results, but small enough to enable respondents to grasp the response categories.

memory, it also assesses phonological awareness and rapid naming using various subtests. Subtests in the composites other than phonological memory were disregarded.

The digit span task was administered using a downloadable Java-based application (Stone & Towse, 2015) in the Tatool architecture (Bastian, Locher, & Ruflin, 2013). The participant was seated in front of the computer screen and was presented with strings of numbers (between 1 and 99). The strings were between n=2 and n=7 numbers long, starting with strings of two numbers and finishing at seven numbers, three trials for each n/string length. In each trial, the numbers appeared on the screen separately. After each string of numbers, the participant was prompted to type the numbers into the computer in the right order, one at a time. Feedback was given for each number, but trials continued even after the participant made mistakes.

The output file was used to calculate the *full trial accuracy score* (fta-score) in the following way: If a string of n digits was remembered correctly, it received a score of n. Since there were three trials for each n, the maximum score for each n was 3n (e.g., the maximum score for the strings of 4 numbers was 4x3=12, making for a total maximum score of 81. This way, longer trials carried higher weight relative to shorter trials, which reflects the higher cognitive demand as digit count increases.

The second half of the battery for testing PWM was a nonword repetition task. The items for the task were all German nonwords. There were items of one to four syllables in length, 10 for each, making for a total of 40 nonwords. The structures of the items can be seen in Table 4. For a full list, see Appendix B.3.

Table 4: Patterns for nonword repetition task

Syllable count	Structure
1	CVC, CV:C
2	‘CV.CVC, ‘CV:.CVC
3	CV.‘CV.CV
4	CV.CV.‘CV:.CV

All items were recorded by a female German native speaker. During the experiment, these recorded items were presented to the participants auditorily via headphones, and participants were asked to repeat, as accurately as possible, what they heard. The participants’ responses were recorded using the same recorder as in the other tasks. In the analysis of the responses, each item that was repeated correctly immediately received 2 points, each item that was repeated incorrectly but then repaired received one point, and each item that was not repeated correctly received 0 points. Thus, the maximum number of points that could be achieved in this task was 80. Finally, the fta-score from the digit span task and the score from the nonword repetition task were summed up for each participant, yielding an overall PWM score.

4.1.3.4.3 Language use

Language use[41] for all three languages was assessed using a short, self-compiled questionnaire that was aimed at assessing the use of German, Turkish, and English in different

[41] The relative amount of language use is often used as a proxy for the abstract concept of language dominance (see Treffers-Daller, 2019 for review and discussion). This, however, fails to capture the complex nature of dominance (which, in addition to very straightforward use, such as the amount to which a language is spoken and consumed, must also include “softer” factors such as attitude towards a language and a culture, as well as the practical and emotional importance a language has for the speaker), which is why I will assume what has been measured by this questionnaire to indicate language use only.

modes (active/passive) and different domains (at school/at home, with friends/with parents). There were three questions for active and passive language use each that the participants answered for all of the three languages[42] by ticking one of four options: often, sometimes, rarely, never. Each answer of *often* was worth 3 points, *sometimes* received 2 points, *rarely* received 1 point, and *never* was worth 0 points. The points were added up for each of the languages individually, yielding one score for English use, one score for German use, and one score for Turkish use.

4.1.3.4.4 German and Turkish language samples

In order to assess participants' behavior in their L1 German and, in the case of the bilinguals, their HL Turkish, the participants performed a short picture naming task in German (monolinguals) or German and Turkish (bilinguals). The eight German items contained one of the features from FT-G (i.e., initial CC or phonemic vowel length, n=4 each), and the Turkish items all contained one of the features from FT-T (i.e., lateral or final voiced obstruent, n=4 each). The items had to be words that children could be expected to be familiar with, and they were words that were concrete enough to be easily drawn.

The instructions for the task were given in German by the experimenter. Just as in the primary tasks, the participants were audio-recorded using the same recorder placed in front of them. Participants were presented with picture cards and were supposed to name what they saw in the picture. If they did not recognize the item in a drawing or did not know the word for it, they were allowed to skip the card. The cards were presented in the same order for each

[42] Both bilinguals and monolinguals answered the questions for all three languages. This was done to make sure that none of the monolinguals had any input of Turkish.

participant: first the German ones, then the Turkish ones.[43] During the picture naming, the experimenter did not talk to the participants, but, when a participant was stuck, helped by pointing to the part of the picture that showed the target word.

All participants were able to name every German picture (although some of the pictures did not always trigger the intended target word, e.g., some of the participants produced *Mund* instead of *Lippe*). The Turkish items were slightly more problematic: Not all participants were familiar with all the Turkish target items, and some of them thus had to skip a substantial number of pictures. However, every participant managed to produce at least one target item per feature type in both German and Turkish.

4.1.4 Data analysis

4.1.4.1 Perception

Perception was operationalized as the d-prime score *d'*.[44] The d-prime score is a metric based on signal detection theory and frequently used in discrimination tasks in order to obtain a score for overall discrimination accuracy. This score was chosen over percent correct in order to account for potential response biases in the participants. Percent correct as a measure of perceptual performance is based on the assumption that listeners will respond randomly with *same* or *different* whenever they are not sure about whether there is a difference between the tokens. This assumption might, however, not hold for all listeners. A participant might, for

[43] Since the instructions were given in German, this was chosen as the most sensible order to allow the participants to remain in German mode for as long as possible before switching to the expected Turkish responses.
[44] d' is the most frequently used measure for performance in discrimination tasks. Its validity relies on two assumptions: "(1) The signal and noise variable are both normal, and (2) the signal and noise distributions have the same standard deviation" (Stanislaw and Todorov, 1999, p. 140). These assumptions cannot be assessed in terms of a yes-no-paradigm, which is one of the reasons why researches have resorted to the non-parametric measure A' instead. However, it should be noted that this measure also relies on two basic mathematical assumptions (see Smith, 1995 for a more detailed mathematical explanations), which calls into question the advantage of A' over d'.

instance, always respond with *different* when they are not sure, which will result in 100% correct in the different pairs, even when actual discrimination is in fact poor. This is why it is advisable to take the participant's response bias into account, e.g. their overall tendency to respond *same* or *different*. To resolve this bias, Signal Detection Theory (SDT) is used to quantify the ability of participants to distinguish the presence of a stimulus (i.e. signal) from a lack of presence of the stimulus (i.e., noise). In simple yes/no tasks (*yes* if stimulus present, *no* if stimulus is not present), a signal trial is called *hit* when answered with yes and a *miss* when answered with no. A noise trial is called *false alarm* in case of a yes-response and a *correct reject* in case of a no-response. (See Stanislaw & Todorov, 1999 for a detailed technical description and evaluation.)

The classification of responses can be seen in Table 5.

Table 5: d' matrix for yes-no discrimination task

	Response: yes (same)	Response: no (different)
Stimuli: yes (same)	HIT	MISS
Stimuli: no (different)	FALSE ALARM	CORRECT REJECT

d' is calculated in the following way: The probabilities for hits p(H) and false alarms p(F) are computed by dividing the total number of hits H by the sum of the total number of H and the total number of misses M, and equally dividing the total number of false alarms F by the sum of F and the total number of correct rejects C. These probabilities are then used to compute two z-scores, z(H) and z(F). d' is calculated using the following formula, where N_{inv} is the inverted normal distribution:

$$d' = 2 \times N_{inv}\left(0.5\left(1 + \sqrt{2\left(N_{inv}\left(\frac{z(H) - z(F)}{2}\right) - 1\right)}\right)\right)$$

Mathematically, the d'-score cannot be computed with hit or false alarm rates of 0, as the resulting z-score would then be $-\infty$ or $+\infty$. One possible solution that is commonly used for this problem is slightly adjusting the extreme rates upwards by 0.5/n (Macmillan & Kaplan, 1985), which is the method that was also adopted here.[45]

For each participant, d-prime scores were calculated separately for each of the items, e.g., *paf* and *taz* were the two items for testing the vowel /æ/, so there were two d-prime scores for the /æ/-contrast for each participant. The d-prime scores for each of the features, in this case the perception of the vowel /æ/, was then determined by calculating the arithmetic mean. This approach was chosen rather than directly subsuming d' for contrasts in order to make sure that a potential item effect could be detected during analysis. Generally, a higher d-prime score represents higher discrimination accuracy. The minimum d-prime score was 0 (i.e., random discrimination), and the maximum d-prime score in this experiment was 3.499821.

4.1.4.2 Production: phonetic and acoustic analyses

Overall production accuracy was operationalized as percent native-like production. To obtain these scores, each item was judged auditorily by two phonetically trained judges and categorized as native-like (yielding a score of 100) or not native-like (yielding a score of 0). Inter-rater reliability was assessed by calculating Cohen's Kappa, and there was substantial

[45] Researchers have claimed that this problem can also be solved by resorting to the non-parametric measure A' instead of d' (see, for instance, Stanislaw and Todorov, 1999, p. 143). However, as Smith (1995) notes, these measures cannot, in themselves, deal with such extreme response rates, which also "require special considerations" (Smith, 1995, p. 85) when calculating A'.

inter-rater agreement, as evidenced by a Cohen's Kappa of κ = 0.742.[46] Arithmetic means were then calculated from these accuracy scores for each of the contrasts to yield percent correct for each of the contrasts.

Furthermore, additional analyses were performed for the interdental fricative, the laterals, and the vowel /æ/, as they potentially offer insights into phenomena that go beyond the +/- native-like categories: Since there were three different potential substitutions that could be expected to occur in the data (e.g., [f], [t], [s], see Chapter 3.2.3 for details), and, crucially, since these substitutions have been shown to be language-specific and differ across the background languages studied in this dissertation, the actually substituted sound was coded for each non-native-like production of the interdental fricative. These productions were categorized as *f*, *t*, or *s*, or as *other*, respectively, and percentages were calculated for each of the potential substitutions.

While there is a +/- native-like dimension to lateral allophones, the allophony is, in fact, gradient (see Chapter 3.2.2 for more details on laterals in English.) So, to assess the concrete production of the lateral allophones in the participants' English, all laterals that occurred in the test words were measured. Since the test words were designed to target the dark allophone of /l/, light /l/s that occurred in non-target positions in the target items or filler words were also

---

[46] Whenever two raters assess the same data and are not sure about the answer, they are very likely to obtain a certain degree of agreement by chance. This type of chance agreement cannot be distinguished from reliable agreement after conscious decisions when only employing a raw agreement rate. Thus, actual agreement rates will usually be overestimated. Cohen's Kappa, a form of a correlation coefficient frequently used to assess inter-rater agreement, takes into account this chance agreement. Cohen's Kappa is calculated using the following formula:
$$\kappa = \frac{\sum_{j=1}^{s} h_{jj} - \sum_{j=1}^{s} h_{.j} \times h_{j.}}{1 - \sum_{j=1}^{s} h_{.j} \times h_{j.}} = \frac{P_0 - P_e}{1 - P_e},$$
where s is the number of categories, h_{jj} is the relative frequency of choice jj by both raters, and $h_{.j}$ and $h_{j.}$ are the relative frequencies of choice jj by rater 1 and rater 2, respectively (McHugh, 2012; Wirtz, 2019) Cohen's Kappa is usually a positive value between 0 and 1, where higher numbers signify a higher degree of inter-rater agreement. Negative numbers (between 0 and -1) are theoretically possible, but highly unlikely, as they actually signify active inter-rater disagreement which is usually not to be expected.

considered. To obtain these values, all lateral segments were manually marked in Praat by visual and, if necessary, auditory inspection. Lateral onsets were identified by a steep rise in F3. A semi-automated Praat script based on the LPC (Linear Predictive Coding) algorithm was then used to measure F1, F2, and F3 for each lateral at the lowest F2 dip, following Epsy-Wilson (1992, pp. 747–748). This measuring point is considered to be more resistant to inter-speaker variation as well as variation in the phonetic environment in which the lateral occurs. For a detailed discussion of the choice of measuring point in laterals, see Himmel (2019, pp. 160–161). The values obtained were then normalized via the Bark Difference Method (Traunmüller, 1997) using the *Vowel* package (Kendall & Thomas, 2018) in R to yield Z1 (the normalized equivalent of F1) and Z2 (the normalized equivalent of F2) values, which are measured in Bark (instead of Hertz).[47]

As explained in detail in 3.2.2, the two variants of /l/ were not presumed to be discrete categories, but rather extremes on a gradient scale. As such, the degree of velarization of /l/, $V_{/l/}$, was operationalized as the difference between (normalized) Z1 and Z2. Thus, the following formula expresses the degree of velarization:

$$V_{/l/} = Z2 - Z1$$

Dark and clear laterals are mainly distinguished acoustically by lower F2 (or Z2) in dark and higher F2 (or Z2) in clear versions (Rodrigues, Martins, Silva, & Jesus, 2019), which can be seen in Figure . Thus, smaller $V_{/l/}$ was taken to indicate a stronger degree of velarization (since Z2 was lower and therefore the difference between Z1 and Z2 was smaller the more velarized or "darker" the lateral), while higher $V_{/l/}$ was taken to mean a weaker degree of

[47] The Bark Difference Method was chosen since it is a vowel-intrinsic method, i.e., it does not require the measurement of all vowels in the speaker's system but instead normalizes acoustic values using the parameters in single vowels, specifically the differences between Z1, Z2, and Z3 (Flynn and Foulkes, 2011).

velarization, i.e., a "lighter" lateral. Figure shows the productions of *ill* and *lip* by a female native speaker of British English with the raw spectrogram at the top and the extracted formants at the bottom. The lower F2 values expected in the dark can clearly be seen in these extractions.

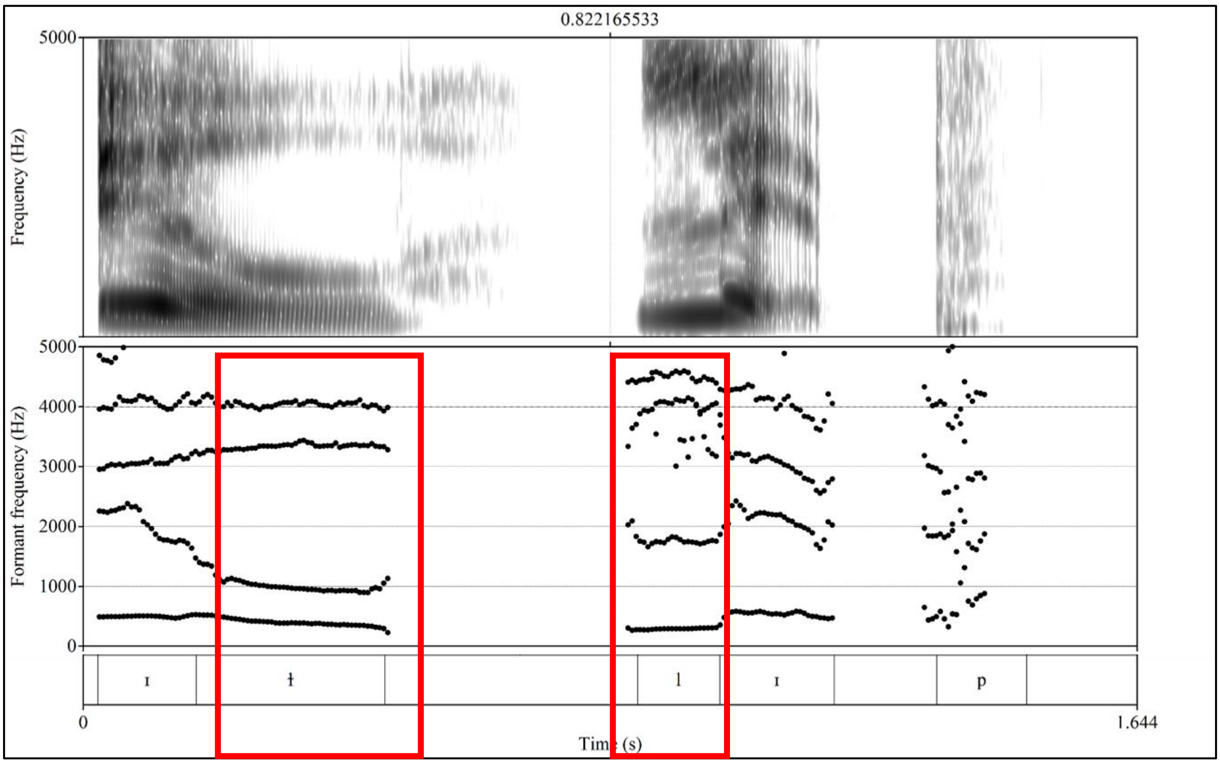

Figure 3: Spectrogram of dark vs. clear lateral production

Furthermore, in order to assess the concrete vowel quality produced by the participants, F1, F2, and F3 of all instances of the vowel /æ/ were measured for both groups. The obtained values were normalized using the Bark Difference Method (see above). In addition to the measurements conducted on the participants' productions of /æ/, measurements were taken of the same vowel in the stimuli as produced by the British native speaker to be able to assess how a potential difference would relate to the native speaker target that was the direct perceptual model during the experiment. To identify a potential influence from Turkish, it was necessary to measure the productions of the same vowel by native speakers of Turkish without German in their language system. To that end, female speakers from Turkey were selected from the Speech Accent Archive (Weinberger, 2015), and one English item (*bags*) was measured per speaker. After listening to all the female speakers with L1 Turkish and excluding those whose formants could not be measured due to insufficient audio quality,[48] three samples (turkish5, turkish11, turkish23) were measured and analyzed in the same way as the primary data.

4.1.4.3 Secondary variables

As presented in Section 4.1.3.4, various secondary variables were considered in addition to the primary tasks, perception, and production. Table 6 shows the name of each variable as well as its operationalization and concrete scoring system employed for analyzing the data (see 4.1.3.4 for more detailed explanations).

Table 6: Secondary variables and their operationalizations

Variable	Operationalization
PHONOLOGICAL WORKING MEMORY	PWM score Higher score ≙ higher working memory
MOTIVATION	Motivation index Higher score ≙ higher motivation
ATTITUDE (Turkish, German, language learning)	Attitude indices Higher attitude index ≙ more positive attitude
LANGUAGE USE (Turkish, German)	Language use indices Higher language use index ≙ higher degree of use

[48] Note that audio files are available for download from the Speech Accent Archive only in MP3 format, which Praat cannot handle. Thus, the MP3 files were converted to WAV files, but the original quality could, of course, not be recovered, which is why most of the files were not usable for measurements. In addition, the recording quality in some of the files was also low with a high amount of noise, which also made formant measuring impossible in many cases.

4.1.4.4 Statistical analyses

To test the predictive power of the different variables, linear mixed-effects models were fitted in R (R Core Team, 2018) using the lme4 package (Bates, Mächler, Bolker, & Walker, 2015). Linear mixed-effects models are linear regression models that, in addition to the fixed effects (or independent variables), include random effects, i.e., variables that have not been actively controlled for but that may exert an influence on the dependent variables, i.e., the "'noise' that should be filtered out" (Levshina, 2015, p. 192). In experimental linguistic studies, these are very often participants/speakers and (lexical) items.

Two primary analyses (M1 and M2) for both perception and production were conducted as part of the present study. Random factors in both M1 and M2 were PARTICIPANT and WORD. The potential predictors tested for in M1 were GROUP, FEATURE, GRADE, PWM, and MOTIVATION, as well as interactions with FEATURE. In the M2 models, FEATURE was replaced with FEATURE TYPE (i.e., FT-G, FT-T, FT-E, see Chapter 3.2 for more details). The two variables could not be tested in one model due to collinearity (since the distinction according to FEATURE TYPE crucially includes the distinction according to FEATURE).

Ideal models were determined by stepwise model reduction: First, an initial model was built which contained all potential predictors (i.e., independent variables, n=x) as well as the random effects. In each step, the exclusion of each variable was tested, which led to a number of potential models all containing n-1 variables. These potential models were then compared to the original model using a chi-square test. Potential variables were eliminated based on the p-value obtained in the model comparisons: Variables that did not significantly improve model fit, i.e., those with $p>.05$, were eliminated. If more than one variable resulted in $p>.05$, the AIC value from the output was inspected. A lower AIC is assumed to correspond to a better fit, and thus, of those models with $p>.05$ in a step of comparisons, the variable whose elimination resulted in the lowest AIC was eliminated. This was done until no variables could be eliminated (i.e., the elimination of a variable from the model did not result in a better fit). The resulting model was considered the best model, and it is thus the model reported. All the factors that remained in this final model were considered to be significant predictors of the dependent variable. In addition to the main models, different models were fitted whose initial structure contained fewer variables. The structure of smaller initial models will be specified in the respective passages. Model reductions were carried out using the same procedure as described above.

Pairwise comparisons using lsmeans (Lenth, 2016), analyses of effect size Cohen's d using the effsize package (Torchiano, 2016), and Pearson's correlations using the Hmisc package (Harrell, 2019) in R were run on the ideal models for post-hoc analyses to test the differences between individual factor levels. Linearity, homoskedasticity and normality, the main assumptions in all linear regression analyses and thus also for linear mixed-effects models, were confirmed by visual inspection of the respective residual plots.

4.2 Results

The following sections contain further participant by-group comparisons based on PWM- and motivation-scores, the two extra-linguistic variables collected from all participants (Section 4.2.1). Global perception and production results are presented in Sections 4.2.2.1 and 0, followed by the analysis of some of the individual bits.

4.2.1 Descriptive statistics: participant profiles

Figure 4 shows the scores achieved in the PWM test by all participants, where higher numbers indicate a more powerful phonological working memory. The monolinguals are on the right, the bilinguals on the left. There is individual variation in both groups, but no extreme values can be observed. Overall, there seems to be a tendency for the monolinguals to achieve slightly higher values. More specifically, the two highest scores were achieved by participants in Group 2 (monolingual), and only participants from Group 1 (bilingual) received scores below 40 (4 participants). However, this trend is not extreme (especially when considering the scoring method used and the consequent meaning of these minute differences), which is also confirmed by a t-test that showed no significant difference between the two groups ($t = 1.2908$, $df = 21.323$, $p = 0.2106$).

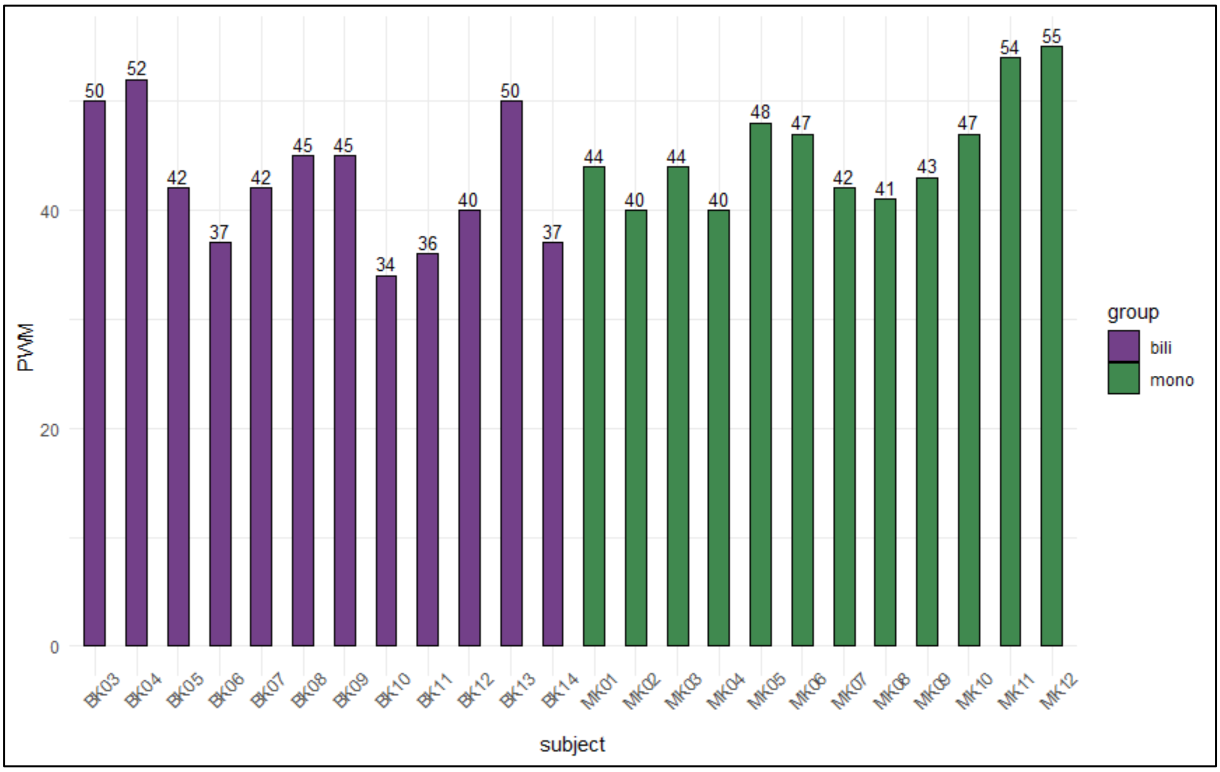

Figure 4: PWM scores for individual participants

Figure 5 shows the motivation scores of all participants obtained using the AMTB. Again, the scores of the monolinguals can be seen on the right, those of the bilinguals on the left. While motivation scores are overall more dispersed across participants than PWM scores, no group difference seems apparent, which is, again, confirmed by a t-test comparison ($t = 0.89776$, $df = 21.639$, $p = 0.3792$).

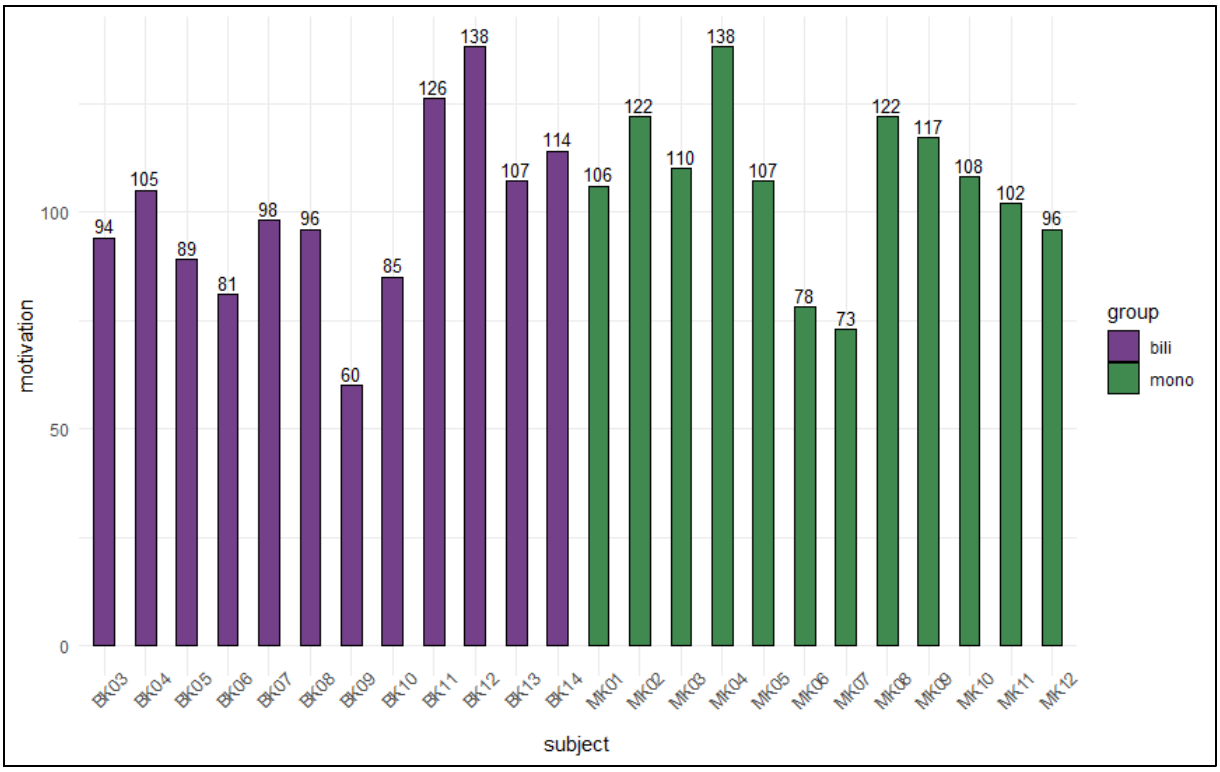

Figure 5: Motivation scores for individual participants

It needs to be noted that, overall, the participants of this study differed only slightly concerning PWM. Specifically, PWM ranges from 34 to 55 with a mean PWM of 43.96. The absence of significant variation also applies to the case of motivation. While motivation scores vary more across participants than PWM scores, there is still a high degree of similarity with a range from 60 to 138 and a mean of 103. Since participation in this study was voluntary, it is highly likely that only those students took part that had a reasonable level of motivation towards languages, language learning, and English in particular. Participants, then, are relatively homogeneous across and within groups when it comes to both motivation and PWM.

4.2.2 Inferential statistics: Production and perception

For both perception and production, one model (M1) was fitted to include FEATURE as a variable, and a second model (M2) was fitted to include FEATURE TYPE instead. Linearity, homoskedasticity, and normality were tested visually for all models using residual plots, and all models were shown to conform with the necessary assumptions, i.e., there were no deviations from linearity, homoskedasticity, or normality. Significant main effects and interactions were determined by pair-wise model comparisons (see Section 4.1.4.4 for a detailed description).

4.2.2.1 Global results: perception

The final model (Table 7) obtained via model comparison shows a main effect of GROUP and FEATURE as well as a significant interaction between the two in perceptual accuracy. During the model comparisons, PWM, GRADE, and MOTIVATION were eliminated as their presence did not significantly improve the model. This means that neither of the three factors significantly predict d’ in the present sample.

Table 7: Linear mixed-effects model M1 for perception

Random effects:					
Groups	Name		Variance	Std. Dev.	
Speaker	(Intercept)		0.1272	0.3566	
Word	(Intercept)		0.1671	0.4088	
Number of obs: 432, Groups: Speaker, 24; Word, 18					
Fixed effects:					
	Estimate	Std.Error	df	t value	Pr(>|t|)
(Intercept)	3.1702	0.3475	27.2944	9.124	8.85e-10***
Group (bili)	-0.7857	0.2727	150.5228	-2.881	0.00454**
Feature (CC)	-1.0091	0.4693	23.1179	-2.150	0.04225*
Feature (lateral)	-2.2147	0.4693	23.1179	-4.719	9.27e-05***
Feature (ou)	-0.9232	0.4693	23.1179	-1.967	0.06128.
Feature (th)	-0.9646	0.3832	23.1179	-2.517	0.01921*
Feature (voicing)	-1.2099	0.4693	23.1179	-2.578	0.01678*

Feature (V length)	-0.9732	0.4693	23.1179	-2.074	0.04944*
Group (bili) : Feature (CC)	0.3141	0.3261	390.8073	0.963	0.33609
Group (bili) : Feature (lateral)	0.8435	0.3261	390.8073	2.586	0.01006*
Group (bili) : Feature (ou)	0.0016	0.3261	390.8073	0.005	0.99608
Group (bili) : Feature (th)	0.7704	0.2663	390.8073	2.893	0.00403**
Group (bili) : Feature (voicing)	-0.0355	0.3261	390.8073	-0.109	0.91333
Group (bili) : Feature (V length)	-0.1215	0.3261	390.8073	-0.372	0.70976

Post-hoc t-tests with Bonferroni corrections were run on the model. When collapsing both groups, d' was significantly lower for the lateral distinction than for the /æ/-distinction (p=0.0449, df=28.61, t=3.214), with all other features spread between these two. Table 8 shows the perception hierarchy of both groups taken together, while Table 9 and Table 10 show the mean perception scores of the monolinguals and the bilinguals, respectively.

Table 8: Perception hierarchy for all participants

Feature	Feature type	d'-score
/æ/	FT-E	2.78
/θ/	FT-E	2.21
CC	FT-G	1.93
/oʊ/	FT-E	1.86
vowel length	FT-G	1.75
voicing	FT-T	1.56
lateral	FT-T	0.99

Table 9: Monolinguals' perception hierarchy

Feature	Feature type	d'-score
/æ/	FT-E	3.13
/oʊ/	FT-E	2.21
/θ/	FT-E	2.17
vowel length	FT-G	2.16
CC	FT-G	2.12
voicing	FT-T	1.92
lateral	FT-T	0.92

Table 10: Bilinguals' perception hierarchy

Feature	Feature type	d'-score
/æ/	FT-E	2.44
/θ/	FT-E	2.24
CC	FT-G	1.74
/oʊ/	FT-E	1.51
vowel length	FT-G	1.34
voicing	FT-T	1.19
lateral	FT-T	1.06

Crucially, when separating the two groups, perception scores for the lateral, the feature with the lowest d'-score, are significantly lower than that for /æ/, the feature with the highest d'-score, only in the monolingual group ($p=0.0084$, $df=36.97$, $t=3.809$), but not in the bilinguals ($p=0.2451$, $df=36.97$, $t=2.358$). Furthermore, the monolinguals performed significantly better than the bilinguals in the perception of /æ/ (p=0.0186, df=173.01, t=2.377), /oʊ/ (p=0.0188, df=173.01, t.ratio=2.371), voicing (p=0.0134, df= 173.01, t=2.498), and vowel length (p=0.0058, df=173.01, t=2.791).

The hierarchies suggest a possible effect of FEATURE TYPE on perceptual accuracy, with the features from FT-E perceived particularly well and those from FT-T particularly weakly by both the monolinguals and the bilinguals. Indeed, multiple comparisons of linear mixed-effects models containing FEATURE TYPE instead of FEATURE as an independent variable (see Table 11 for the ideal model) reveal a significant main effect of FEATURE TYPE as well as GROUP, but no significant interaction.

Table 11: Linear mixed-effects model M2 for perception

Random effects:			
Groups	Name	Variance	Std. Dev.
Speaker	(Intercept)	0.1253	0.3540
Word	(Intercept)	0.2343	0.4841
Number of obs:432, Groups: Speaker, 24; Word, 18			

Fixed effects:					
	Estimate	Std.Error	df	t value	Pr(>|t|)
(Intercept)	2.4540	0.1957	31.3466	12.541	9.36e-14***
Group (bili)	-0.4175	0.1649	23.5649	-2.532	0.01844*
FT-G	-0.4108	0.3032	17.9278	-1.355	0.19227
FT-T	-0.9782	0.3032	17.9278	-3.226	0.00471**

Post-hoc t-tests reveal that features in FT-E are perceived significantly more accurately than those in FT-T (p=0.0193, df=21.11, t=2.967). Features in FT-G are intermediate, but the differences between FT-G and FT-E and FT-T, respectively, do not come out as significant, which holds for both monolinguals and bilinguals. There is no interaction between GROUP and

FEATURE TYPE. In other words, while all speakers perceive contrasts with different accuracy as a function of FEATURE TYPE (i.e., FEATURE TYPE is a predictor of perceptual accuracy), this effect is unrelated to whether the speaker is monolingual or bilingual.

Figure 6, Figure 7, and Figure 8 show d'-scores by FEATURE and GROUP ordered according to FEATURE TYPE. In addition to the differences in the perception of individual features reported earlier, it is especially apparent here that there are group differences in features from each of the feature types, and there is thus no consistent grouping according to feature type. Furthermore, if there is a group difference for a feature in d', it is always the monolinguals that outperform the bilinguals, and never the other way around.

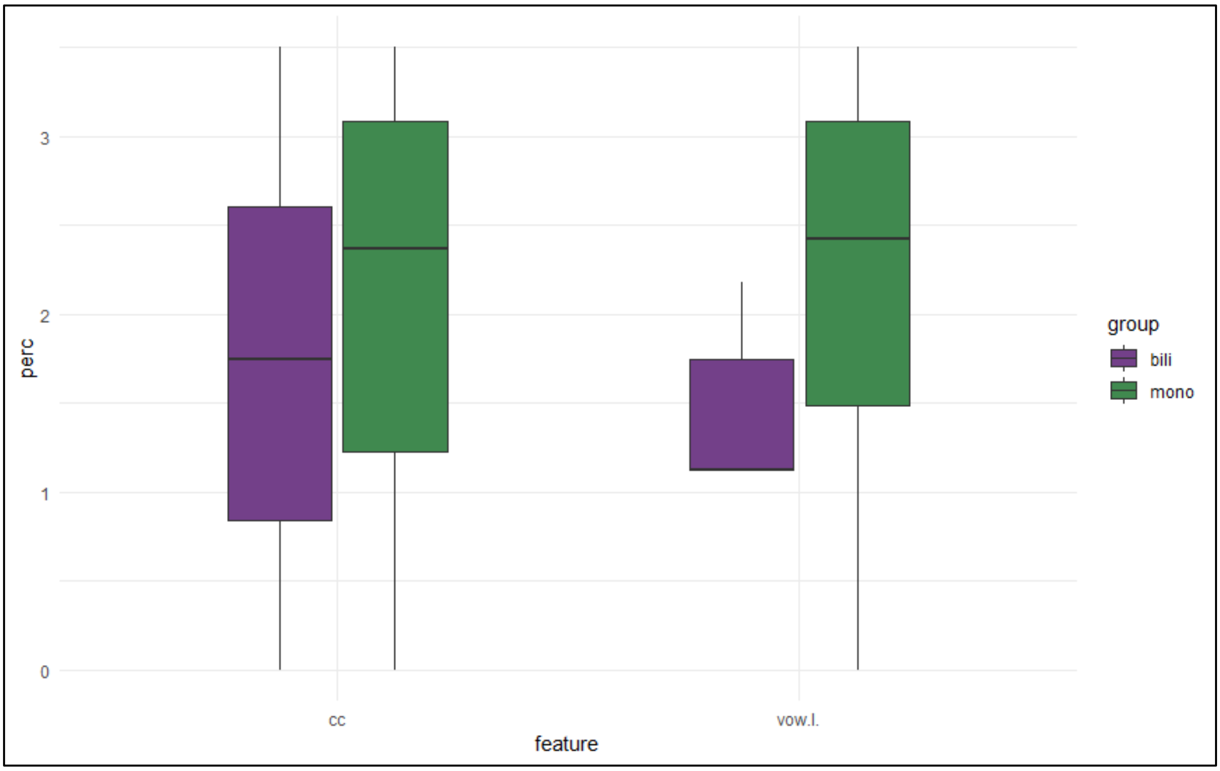

Figure 6: Perception accuracy (d') FT-G

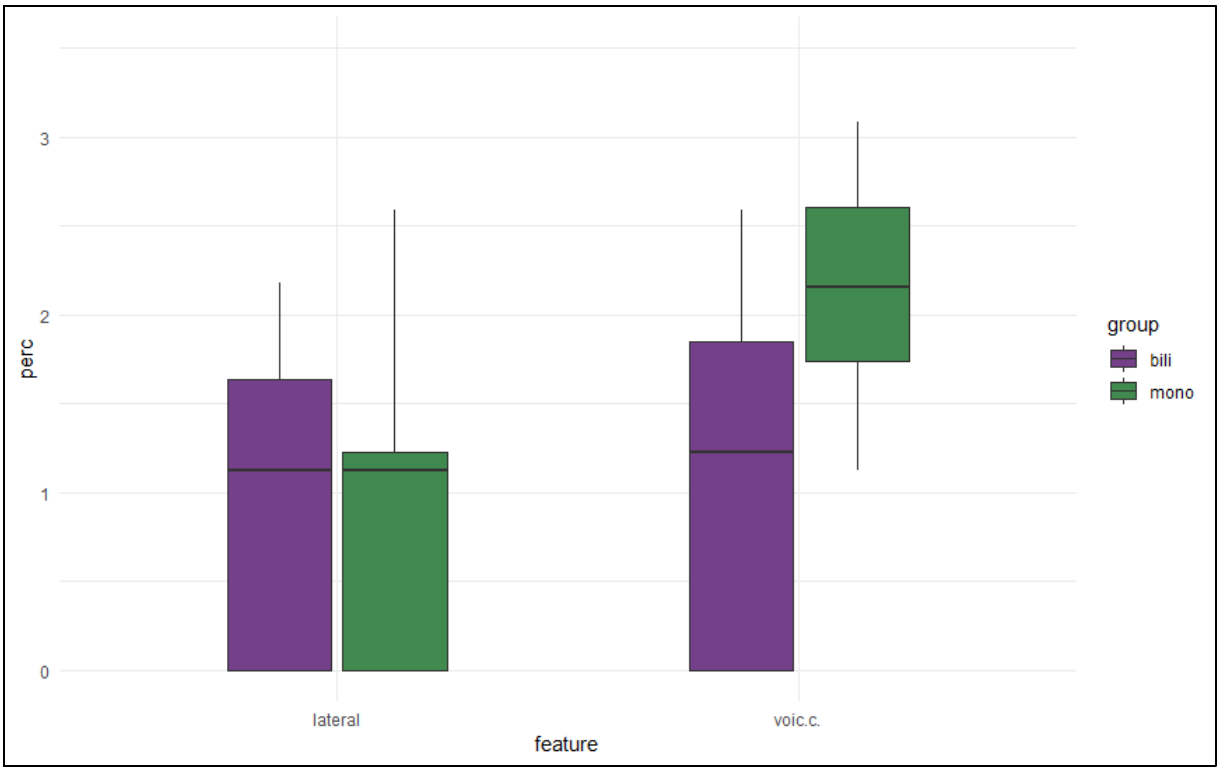

Figure 7: Perception accuracy (d') FT-T

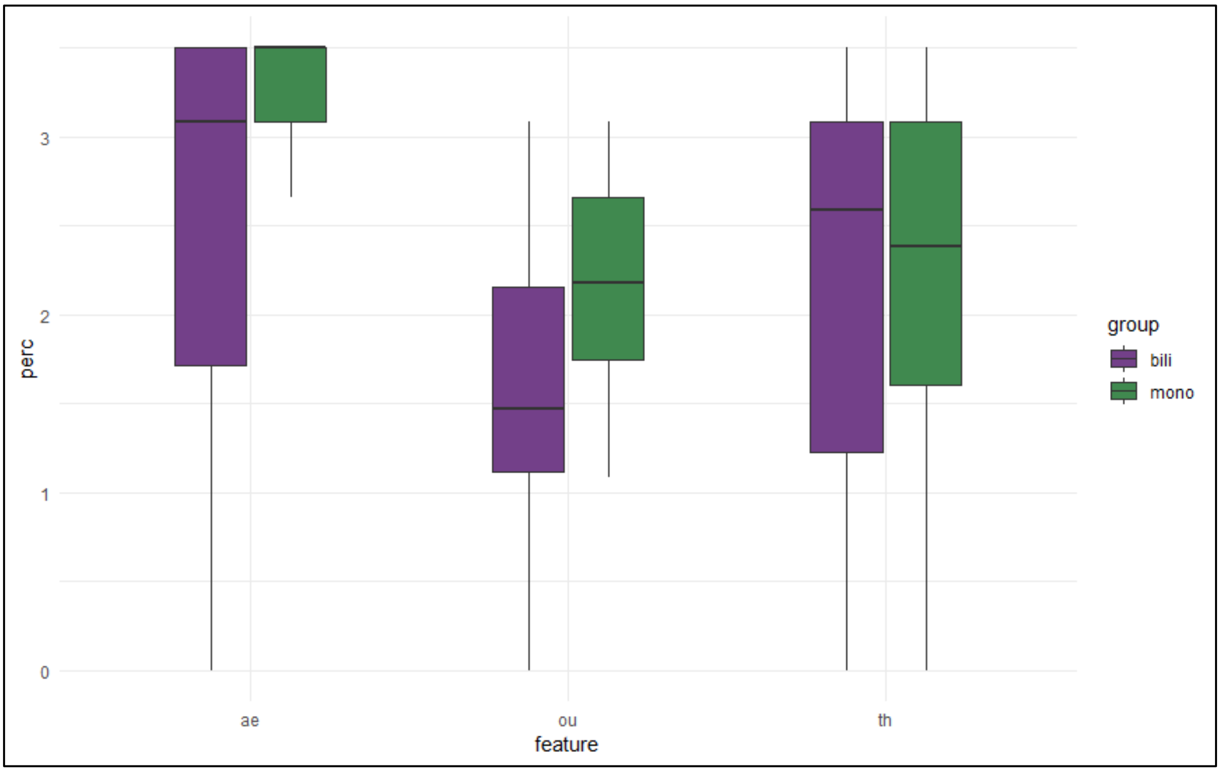

Figure 8: Perception accuracy (d') FT-E

Furthermore, neither PWM or MOTIVATION nor GRADE were significant predictors of perception accuracy when the model contained FEATURE or FEATURE TYPE as one of the independent variables. Table 12 shows the results of the model comparisons.

Table 12: Statistical results for secondary variables in perception

	p	Chisq	Df
PWM (M1)	0.2732	1.2006	1
Motivation (M1)	0.2984	1.0815	1
Grade (M1)	0.7197	0.6577	2
PWM (M2)	0.274	1.1966	1
Motivation (M2)	0.2991	1.078	1
Grade (M2)	0.7205	0.6556	2

4.2.2.2 Global results: production

A linear mixed-effects model was fit for production in the same way as for perception, with GROUP, FEATURE, GRADE, PWM, and MOTIVATION as fixed, and WORD and SPEAKER as random effects in the initial model. As described above, percent native-like production was used as an operationalization of production accuracy for the global analysis. The final model, which can be seen in table Table 13, contains an interaction of FEATURE and GROUP as well as a main effect of FEATURE.

Table 13: Linear mixed-effects model M1 for production

Random effects:			
Groups	Name	Variance	Std. Dev.
Speaker	(Intercept)	51.33	7.165
Word	(Intercept)	10.39	3.223
Number of obs: 672, Groups: Speaker, 24; Word, 28			

Fixed effects:					
	Estimate	Std.Error	df	t value	Pr(>|t|)
(Intercept)	4.167	5.342	90.57	0.780	0.437
Group (mono)	4.167	7.203	293	0.578	0.563
Feature (CC)	87.5	6.966	87.94	12.562	<2e-16***
Feature (lateral)	43.75	6.966	87.94	6.281	1.25e-08***
Feature (ou)	85.42	6.966	87.94	12.262	<2e-16***
Feature (th)	37.50	6.966	87.94	53.84	6.00e-07***
Feature (voicing)	35.42	6.966	87.94	5.084	2.06e-06***
Feature (V length)	91.67	6.966	87.94	13.160	<2e-16***
Group (mono) : Feature (CC)	4.167	9.309	620.1	0.448	0.655
Group (mono) : Feature (lateral)	-14.58	9.309	620.1	-1.567	0.118
Group (mono) : Feature (ou)	6.25	9.309	620.1	0.671	0.502
Group (mono) : Feature (th)	6.25[49]	9.309	620.1	0.671	0.502
Group (mono) : Feature (voicing)	-43.75	9.309	620.1	-4.700	3.21e-06***
Group (mono) : Feature (V length)	1.089e-13[50]	9.309	620.1	0.000	1

[49] Note that in both /oʊ/ and /θ/, the bilingual group exhibits five non-target-like productions more than the monolingual group, which is why the numbers for both features are the same.

[50] Both monolinguals and bilinguals perform exceptionally well in the production of phonemic vowel length. Indeed, the monolinguals performed target-like in all instances, while the bilinguals performed target-like in all but two instances (*food*, with a possible lexical effect of *food* vs. *foot*, and *moon*; note that this is the same vowel in both cases).

Model comparisons showed a significant main effect of FEATURE and post-hoc t-tests run on the model revealed that the features cluster into three distinct groups: those with very high production accuracy (vowel length, CC, /oʊ/), those with intermediate production accuracy (/θ/ and lateral) and those with very low production accuracy (voicing and /æ/). Note, however, that this clustering does not match the clustering according to feature type predicted in Chapter 3.3. Table 14 shows the production hierarchy for both groups. Lines in the table denote a significant difference between adjacent features.

Table 14: Production hierarchy for all participants

Feature	Feature type	% target-like (ls.sq.mean)
vowel length	FT-G	98.26
CC	FT-G	96.17
/oʊ/	FT-E	95.13
/θ/	FT-E	47.22
lateral	FT-T	43.05
voicing	FT-T	20.13
/æ/	FT-E	6.59

In addition, the model comparisons revealed a significant interaction of FEATURE and GROUP. The following two tables (Table 15 and

Table 16) show the production hierarchies for the monolinguals and the bilinguals, respectively.

Table 15: Monolinguals' production hierarchy

Feature	Feature type	% target-like (ls.sq.mean)
vowel length	FT-G	99.22
CC	FT-G	99.22
/oʊ/	FT-E	99.22
/θ/	FT-E	51.3
lateral	FT-T	36.72
/æ/	FT-E	7.55
voicing	FT-T	-0.78[51]

Table 16: Bilinguals' production hierarchy

Feature	Feature type	% target-like (ls.sq.mean)
vowel length	FT-G	97.3
CC	FT-G	93.13
/oʊ/	FT-E	91.05
lateral	FT-T	49.38
/θ/	FT-T	43.13
voicing	FT-T	41.05
/æ/	FT-E	5.63

Very clearly, the overall hierarchy remains stable across groups. It is especially apparent in the top and bottom sections of the tables, however, that effects, both positive and

[51] This is not the arithmetic, but the least square mean, which explains the negative value.

negative, are more extreme in the monolinguals than in the bilinguals, which suggests that the monolinguals do not fluctuate as much as the bilinguals across instances in achieving target-like production. Furthermore, the bilinguals have a higher score in the production of word-final voiced obstruent (41.05% vs. -0.78% in the monolingual productions), which is confirmed by post-hoc t-tests (p<.0001, df=297.83, t=-5.54).

While the effect of feature type clustering seems to be much less pronounced here than in the perception test, there is also a visible tendency of feature types clustering together in production, albeit in a different direction. In M2 for production (see Table 17), there is, indeed, a main effect of FEATURE TYPE as well as a significant interaction of FEATURE TYPE and GROUP.

Table 17: Linear mixed-effects model M2 for production

<table>
<tr><td colspan="6">Random effects:</td></tr>
<tr><td>Groups</td><td colspan="2">Name</td><td colspan="2">Variance</td><td>Std. Dev.</td></tr>
<tr><td>Speaker</td><td colspan="2">(Intercept)</td><td colspan="2">53.62</td><td>7.323</td></tr>
<tr><td>Word</td><td colspan="2">(Intercept)</td><td colspan="2">609.76</td><td>24.693</td></tr>
<tr><td colspan="6">Number of obs: 672, Groups: Speaker, 24; Word, 28</td></tr>
<tr><td colspan="6">Fixed effects:</td></tr>
<tr><td></td><td>Estimate</td><td>Std.Error</td><td>df</td><td>t value</td><td>Pr(>|t|)</td></tr>
<tr><td>(Intercept)</td><td>45.139</td><td>7.913</td><td>36.073</td><td>5.704</td><td>1.72e-06***</td></tr>
<tr><td>Group (mono)</td><td>8.333</td><td>4.860</td><td>55.354</td><td>1.715</td><td>0.091994.</td></tr>
<tr><td>FT-G</td><td>48.611</td><td>12.058</td><td>31.884</td><td>4.032</td><td>0.000322***</td></tr>
<tr><td>FT-T</td><td>-1.389</td><td>12.058</td><td>31.884</td><td>-0.155</td><td>0.909019</td></tr>
<tr><td>Group (mono) : FT-G</td><td>-2.083</td><td>6.058</td><td>620.873</td><td>-0.344</td><td>0.731060</td></tr>
<tr><td>Group (mono) : FT-T</td><td>-33.333</td><td>6.058</td><td>620.873</td><td>-5.502</td><td>5.50e-08***</td></tr>
</table>

Figure 9, Figure 10, and Figure 11 show production accuracy across groups and features clustered by feature type. As was the case in perception, the monolinguals often tend to outperform the bilinguals even when the difference does not reach statistical significance. However, this is not the case in FT-T, where the bilinguals reach higher success rates in both contrasts, of which one, i.e., voicing in final position ($p<.0001$, $df=279.83$, $t=-5.54$), is, in fact, the only contrast that is produced significantly differently across groups.

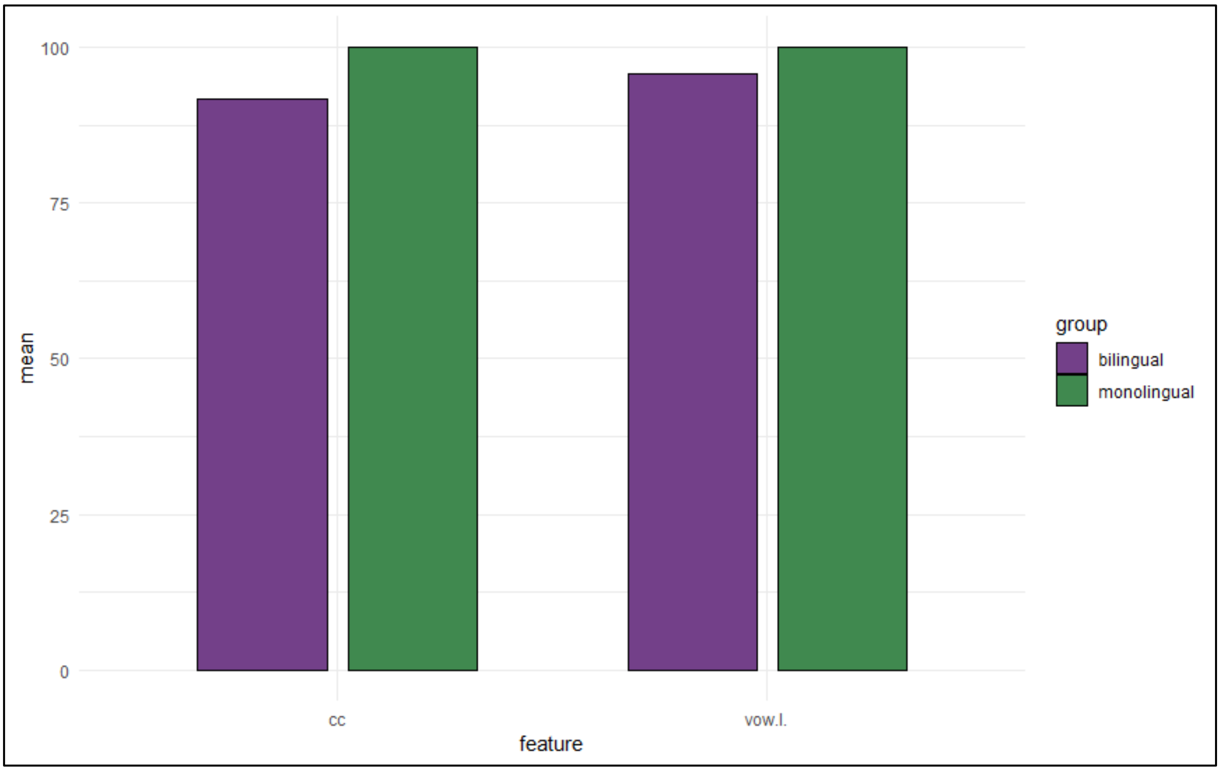

Figure 9: Production accuracy (% correct) FT-G

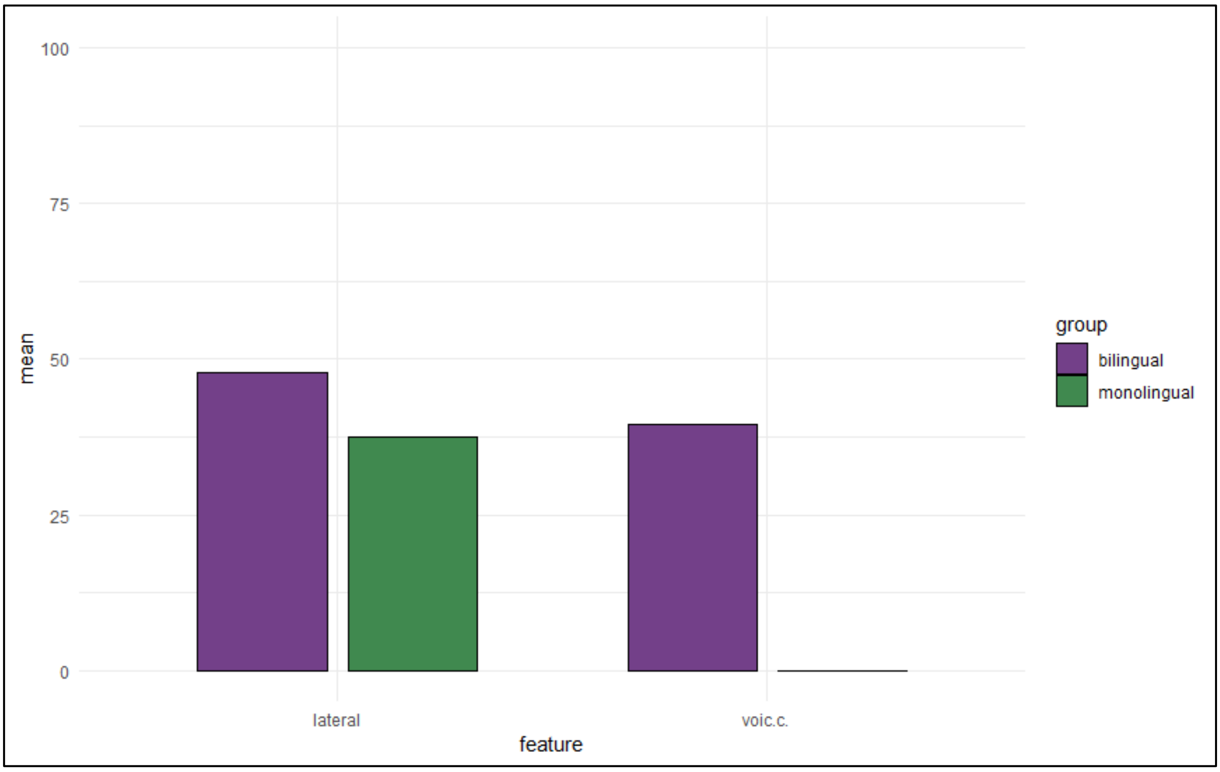

Figure 10: Production accuracy (% correct) FT-T

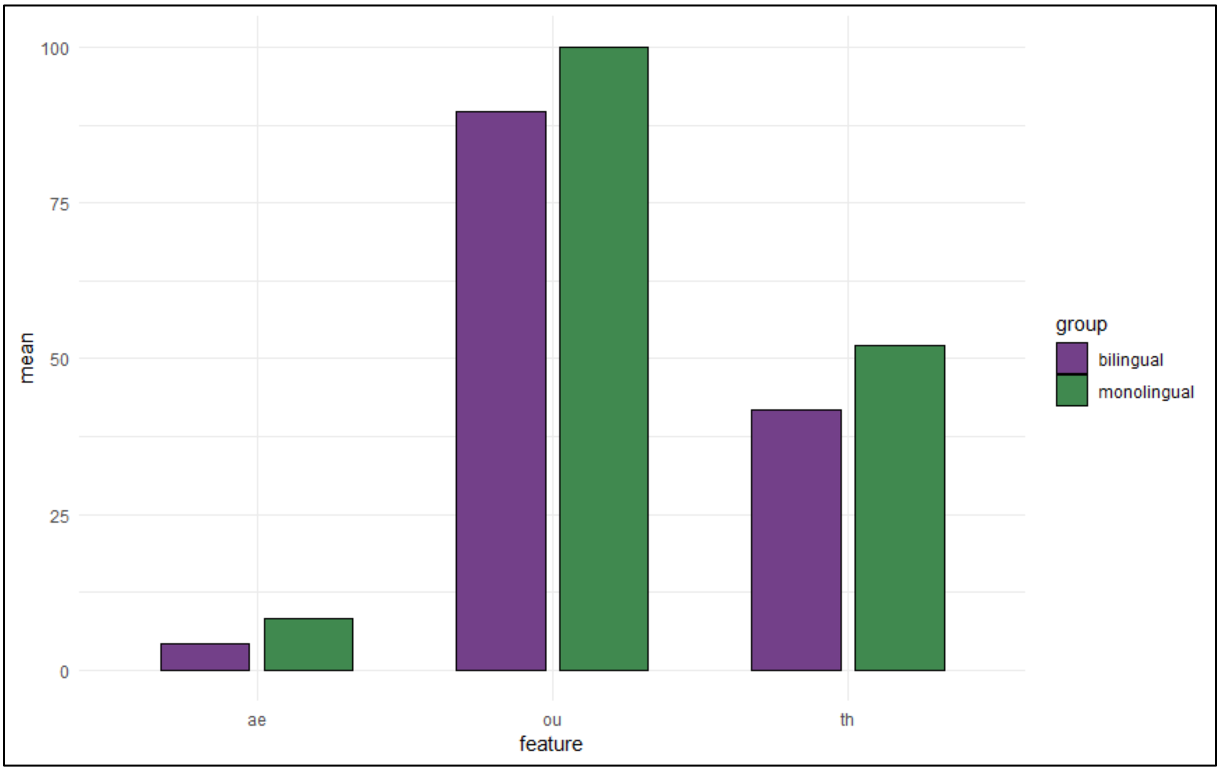

Figure 11: Production accuracy (% correct) FT-E

As was the case in the perception analyses, PWM, MOTIVATION, and GRADE were eliminated from the ideal model via model comparisons in both M1 and M2, which indicates they do not predict production accuracy for either of the groups, neither for individual features nor for feature types.

4.2.2.3 Perception and production: on the interface

When comparing the perception and production hierarchies both in general and across speaker groups, no directly transparent parallels can be seen. In fact, there seem to be pronounced differences between the production and the perception hierarchies obtained in previous sections. For instance, while the /æ/-/ɛ/ contrast was particularly well perceived by all participants, /æ/ was usually produced non-target-like (i.e., it was usually substituted with [ɛ]), which results in very weak scores for production accuracy. To test this statistically and to further test for a direct link between perception and production in the other features, Pearson's correlations[52] were run both on the full dataset and each of the individual features. As both perception and production hierarchies were highly similar across groups, both speaker groups were combined for the correlations to obtain more data points and thus higher statistical validity. The test showed no significant global correlation (i.e., with all features included) between perception and production in the participants (cor=-0.0781, t=-1.01, df=166, p=0.314). This is also the case for each feature individually (see Table 18). In summary, performance in perception and production by the participants of the current study are not correlated, or, in other words, they are not linked to each other straightforwardly.

[52] Since the data in this study has been shown to be normally distributed and the assumed relationship is linear, Pearson Product-Moment correlation was used rather than Spearman Rank correlation (see Schober, Boer, and Schwarte, 2018).

Table 18: Correlations PERCEPTION:PRODUCTION

feature	cor	t	df	p
/æ/	0.14	0.64	22	0.53
CC	0.02	-0.09	22	0.92
lateral	0.37	1.88	22	0.07
/oʊ/	0.25	1.2	22	0.24
/θ/	0.37	1.88	22	0.07
voicing	0.37	-1.89	22	0.07
vowel length	-0.02	-0.11	22	0.91
global	-0.08	-1.01	166	0.31

4.2.2.4 Further analyses concerning laterals

Research in non-native language acquisition not only deals with the binary distinction of whether a feature gets transferred or does not, but also investigates patterns, i.e., the way in which CLI leads to systematic differences that reflect mental representations. Since the analysis of global perception accuracy did not consider the potential difference in perception as a function of phonological context (i.e., the target allophone that was expected to occur in the respective position) for the lateral contrast, an additional model was fitted that included CONTEXT as an independent variable. The random factor WORD as excluded from the model due to its collinearity with CONTEXT. Furthermore, the secondary variables were not included in the model since they had already been shown not to predict perception or production in the global analyses. The final model can be seen in Table 19.

Table 19: Linear mixed-effects model for discrimination of laterals by context

Random effects:					
Groups	Name		Variance		Std. Dev.
Speaker	(Intercept)		0.1619		0.4024
Number of obs: 48, Groups: Speaker, 24					
Fixed effects:					
	Estimate	Std.Error	df	t value	Pr(>|t|)
(Intercept)	0.5673	0.2009	43.1720	2.825	0.00714**
Context (final)	0.7765	0.2317	24.0000	1.623	1.86e-12***
Group (bili)	-0.2456	0.3950	25.4385	-0.622	0.11186
Context (final) : Group (bili)	-0.8063	0.3277	24.0000	-2.460	0.02146*

Model comparisons reveal a main effect of CONTEXT on the perception accuracy of the laterals. Furthermore, there is a significant interaction between CONTEXT and GROUP. Post-hoc t-tests show that the monolinguals are significantly better at discriminating dark vs. clear laterals in final (d'=1.19) as opposed to initial (d'=0.41) position (p=0.0035, df=26.18, t.ratio=-3.208) while the bilinguals reach very similar scores (final: d'=0.95, initial: d'= 0.99) in both contexts (p=0.9027, df=26.18, t.ratio=0.123). This interaction can also be seen in figure Figure 12.

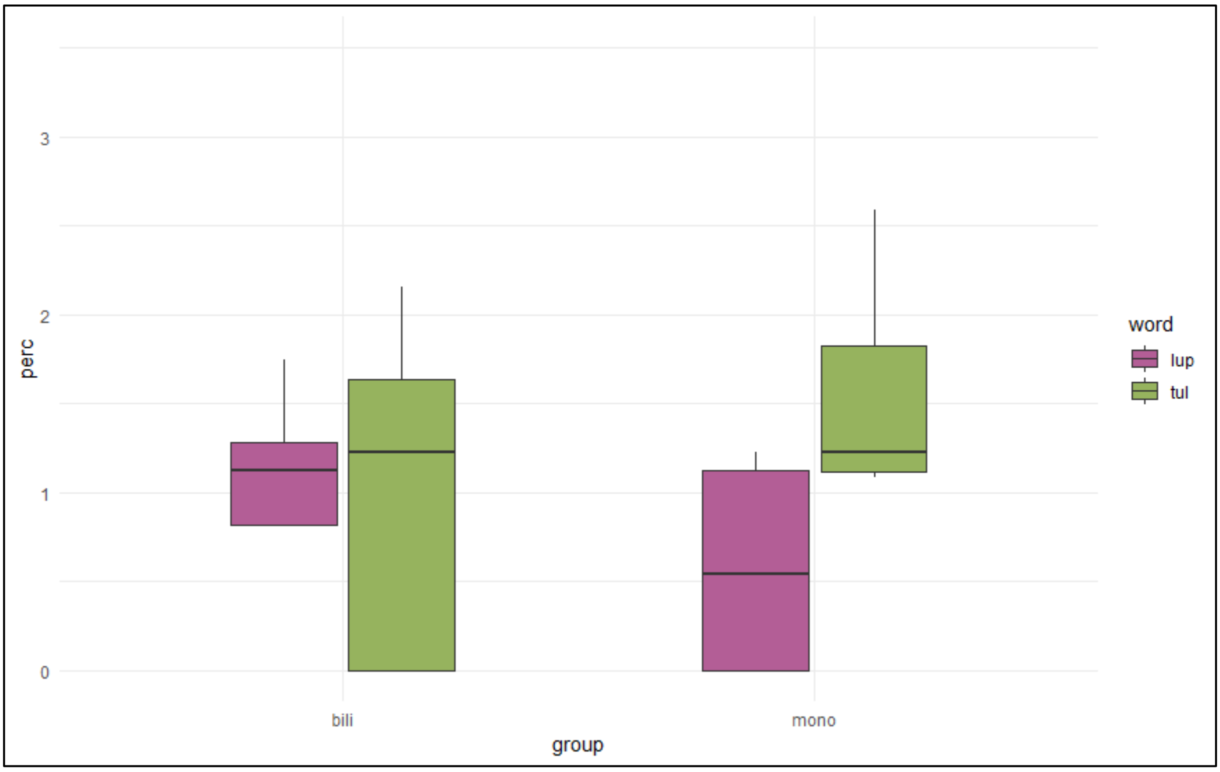

Figure 12: Discrimination accuracy of laterals by context

To determine whether this group difference regarding the laterals also holds for production, formants were measured (see Section 4.1.4.2 for details on the measurement and normalization procedures) for both allophonic contexts to see whether speakers produce the laterals distinct as a function of position (i.e., whether they were producing one variant as darker and the other one as clearer), and whether this was happening as a function of group. Gradient measurements were chosen rather than a binary clear-dark distinction since the former are considered the most faithful representation of this specific kind of allophony: Whereas the degree of difference varies as a function of the variety under scrutiny (i.e., the lateral in clear environments is generally produced as darker in General American than in Received Pronunciation, which diminishes the acoustic difference between the clear and the dark variant), speakers of both varieties indeed distinguish between the two allophones to some degree (Recasens, 2012, p. 371). A model was fitted with Z2-Z1 as the dependent variable, CONTEXT and GROUP as independent variables, and SPEAKER as a random factor (see table Table 20).

Again, WORD was excluded as a random factor due to its collinearity with CONTEXT, and the secondary variables were not added to the model as they were not significant in any of the global analyses.

Table 20: Linear mixed-effects model acoustic differences in laterals across context and group

Random effects:					
Groups	Name	Variance	Std. Dev.		
Speaker	(Intercept)	0.8369	0.9148		
Number of obs: 475, Groups: Speaker, 24					
Fixed effects:					
	Estimate	Std.Error	df	t value	Pr(>|t|)
(Intercept)	5.4845	0.2794	25.4611	19.633	<2e-16***
Context (initial)	1.2329	0.1701	451.4077	7.247	1.86e-12***
Group (bili)	-0.2456	0.3950	25.4385	-0.622	0.539651
Context (initial) : Group (bili)	0.8861	0.2432	451.4425	3.625	0.000322***

Model comparisons revealed a main effect of CONTEXT as well as a significant interaction of CONTEXT and GROUP. Post-hoc t-tests further showed that, overall, Z2-Z1 was lower in final than in initial position, which is in line with the expected target, where laterals are produced as relatively clear initial and as relatively dark in final position. Both groups make an articulatory distinction between initial and final position, but this distinction is made more strongly by the bilinguals than the monolinguals, as evidenced in Cohen’s d, a measure of effect size, which is higher in the bilinguals’ productions ($|d|_{mono}$=.84, $|d|_{bili}$=.94). This difference can be attributed mostly to the difference in initial position: While in final position, speakers produce comparable values, i.e., the production of the dark laterals is consistent across groups, the bilingual speakers produce higher, i.e., clearer, values than the monolinguals in initial position. This way, the formant values for initial vs. final context are more distant and, crucially, more distinct in the bilingual than in the monolingual speakers, as can be seen in the lack of overlap in the bilingual productions. Table 21 and Figure 13: Bilingual and monolingual productions of the laterals across context give an overview of the acoustic results of lateral production.

Table 21: Formant values (Bark) of laterals and statistical differences by context

group	context:ini	context:fin	p	df	t.ratio	Cohen's |d|
mono	6.717	5.484	<.0001	453.44	-7.231	0.84
bili	7.353	5.239	<.0001	453.44	-12.137	0.94

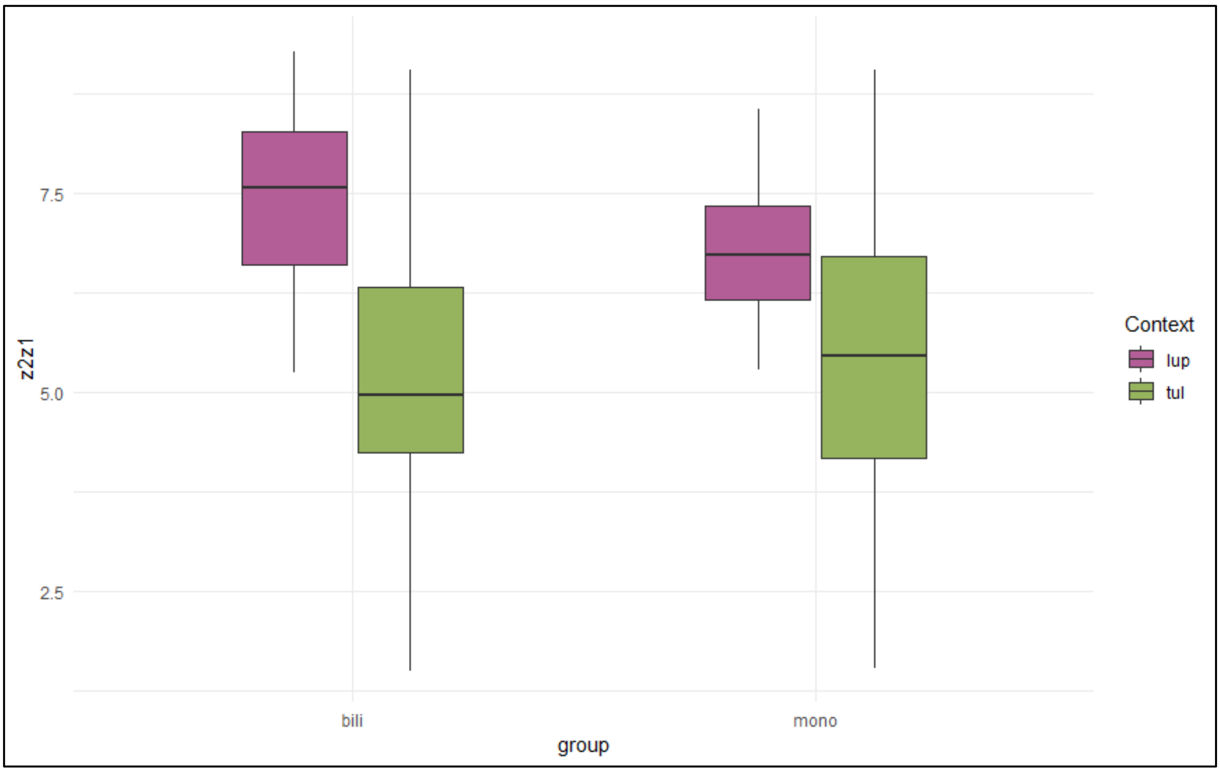

Figure 13: Bilingual and monolingual productions of the laterals across context

4.2.2.5 Further analyses of /æ/

Perception and production of the vowel /æ/ have been shown to diverge significantly in that the distinction between /æ/ and /ɛ/ seems to be very easy to perceive, but a target-like production poses a problem for both groups. Furthermore, production is particularly weak in both groups. Thus, this particular English vowel (or possibly vowels in general, although this has not been tested in this study) provides an ideal test case that could help understand the relationship between perception and production as well as shed further light on the influence of the background languages on the TL. This is why, in addition to the impressionistic judgments, F1, F2, and F3 were measured of all instances of the vowel for both groups. The obtained values were normalized using the Bark Difference Method (see Section 4.1.4.2 on lateral measurements for more details). After excluding one participant (BK13) whose formants could not be measured since his voice was creaky due to a sore throat, two linear mixed-effects models (see Table 22 and

Table 23) were fitted with Z3-Z1 (corresponds to F1) and Z3-Z2 (corresponds to F2) respectively as dependent variables, Group, Grade, PWM and Motivation as independent variables, and Speaker as random variable. The models revealed a significant main effect of Group on vowel height (Z3-Z1) but not on vowel backness (Z3-Z2).

Table 22: Linear mixed-effects model for vowel height (Z3-Z1) by group

<table>
<tr><td colspan="6">Random effects:</td></tr>
<tr><td>Groups</td><td colspan="2">Name</td><td colspan="2">Variance</td><td>Std. Dev.</td></tr>
<tr><td>Speaker</td><td colspan="2">(Intercept)</td><td colspan="2">0.6436</td><td>0.8022</td></tr>
<tr><td colspan="6">Number of obs:</td></tr>
<tr><td colspan="6">Fixed effects:</td></tr>
<tr><td></td><td>Estimate</td><td>Std.Error</td><td>df</td><td>t value</td><td>Pr(>|t|)</td></tr>
<tr><td>(Intercept)</td><td>9.4511</td><td>0.2274</td><td>31.6495</td><td>41.554</td><td><2e-16***</td></tr>
<tr><td>Group (mono)</td><td>-1.2476</td><td>0.2278</td><td>138.3646</td><td>-5.477</td><td>1.98e-07***</td></tr>
</table>

Table 23: Linear mixed-effects model for vowel backness by group

Random effects:					
Groups	Name	Variance	Std. Dev.		
Speaker	(Intercept)	0.1564	0.3955		
Number of obs:					
Fixed effects:					
	Estimate	Std.Error	df	t value	Pr(>|t|)
(Intercept)	2.4399	0.1421	31.5987	17.173	<2e-16***
Group (mono)	-0.1188	0.1716	74.1062	-0.692	0.491

Figure 14 shows that, specifically, the bilinguals' productions of /æ/ are produced significantly higher than those of the monolinguals. In addition, the monolinguals' productions are dispersed more widely than those of the bilinguals.

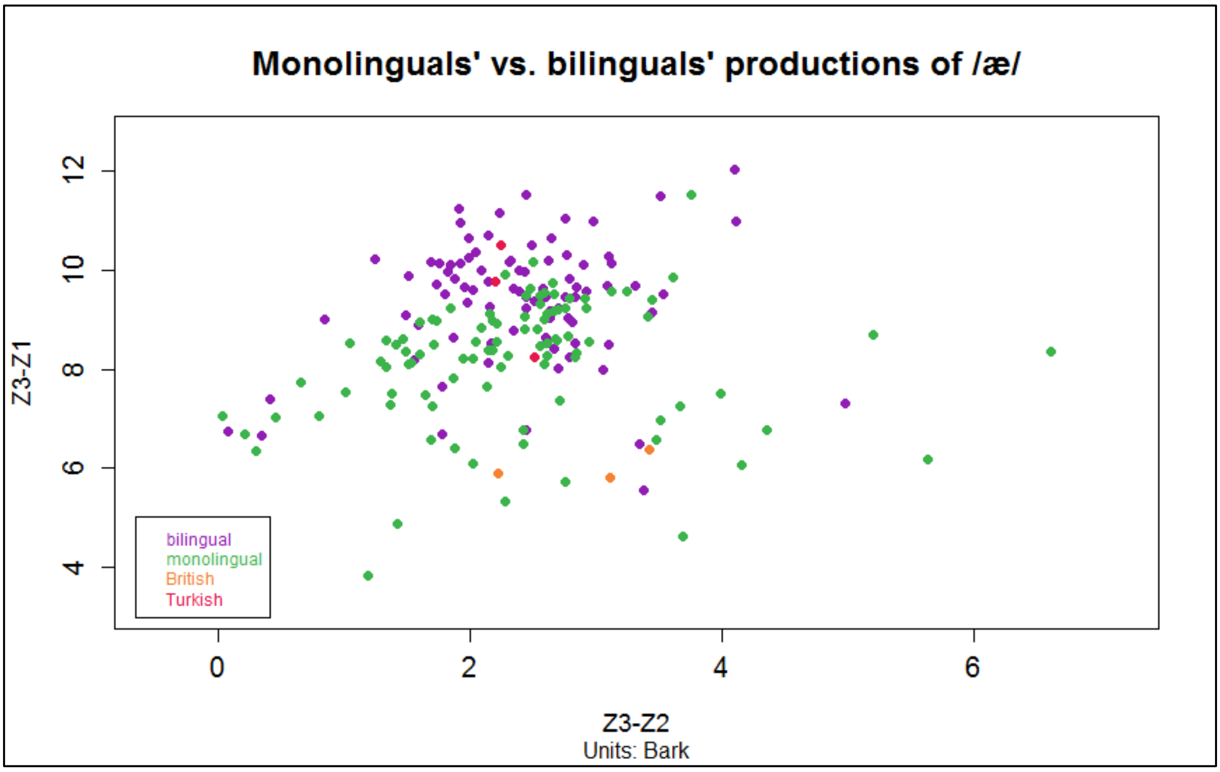

Figure 14: Monolingual and bilingual productions of /æ/

An additional analysis of the /æ/ tokens produced by the native speaker in the stimuli used in the perception task revealed that the monolinguals' productions are closer to the native speaker values whereas the bilinguals are relatively high and thus relatively close to the Turkish-accented English /æ/ that was measured additionally (see Figure 14).

4.2.2.6 Differential substitution of the interdental fricative in perception and production

Since various substitutions for the interdental fricative have been reported for speakers of different languages, and especially since different default substitutions can be expected from L1 speakers of Turkish (who are likely to substitute [t]; Yildiz, 2006) and L1 speakers of German (who can be expected to substitute [s]; Weinberger, 1996, p. 269), this section teases apart these different substitution patterns in the perception and the production of interdental fricatives. A linear mixed-effects model was fitted with SUBSTITUTION and GROUP as independent variables (see Table 24 for the final model) to see whether the two groups encounter different degrees of difficulty when they have to discriminate the interdental fricative and its potential substitutions.

Table 24: Linear mixed effects model for the perception of /θ/ by substitution

Random effects:					
Groups	Name		Variance	Std. Dev.	
Speaker	(Intercept)		0.2250	0.4744	
Number of obs: 144, Groups: Speaker, 24					
Fixed effects:					
	Estimate	Std.Error	df	t value	Pr(>|t|)
(Intercept)	1.2787	0.1358	50.6918	9.416	1.02e-12***
Substitution (s)	1.2365	0.1346	120.0000	9.183	1.51e-15***
Substitution (t)	1.5214	0.1346	120.0000	11.299	<2e-16***

Model comparisons revealed a main effect of SUBSTITUTION. Post-hoc pairwise comparisons showed that the perception of the /f/-/θ/ contrast (d'=1.28) was significantly weaker than that of both the /s/-/θ/ (d'=2.52; p<.0001, df=122.03, t.ratio=-9.106) and the /t/-/θ/ contrast (d'=2.8; p<.0001, df=122.03, t.ratio=-11.205). There is, however, no significant interaction of SUBSTITUTION and GROUP, i.e., both groups discriminated the contrasts equally well. This can also be seen in Figure 15, which shows an almost identical substitution profile in both groups.

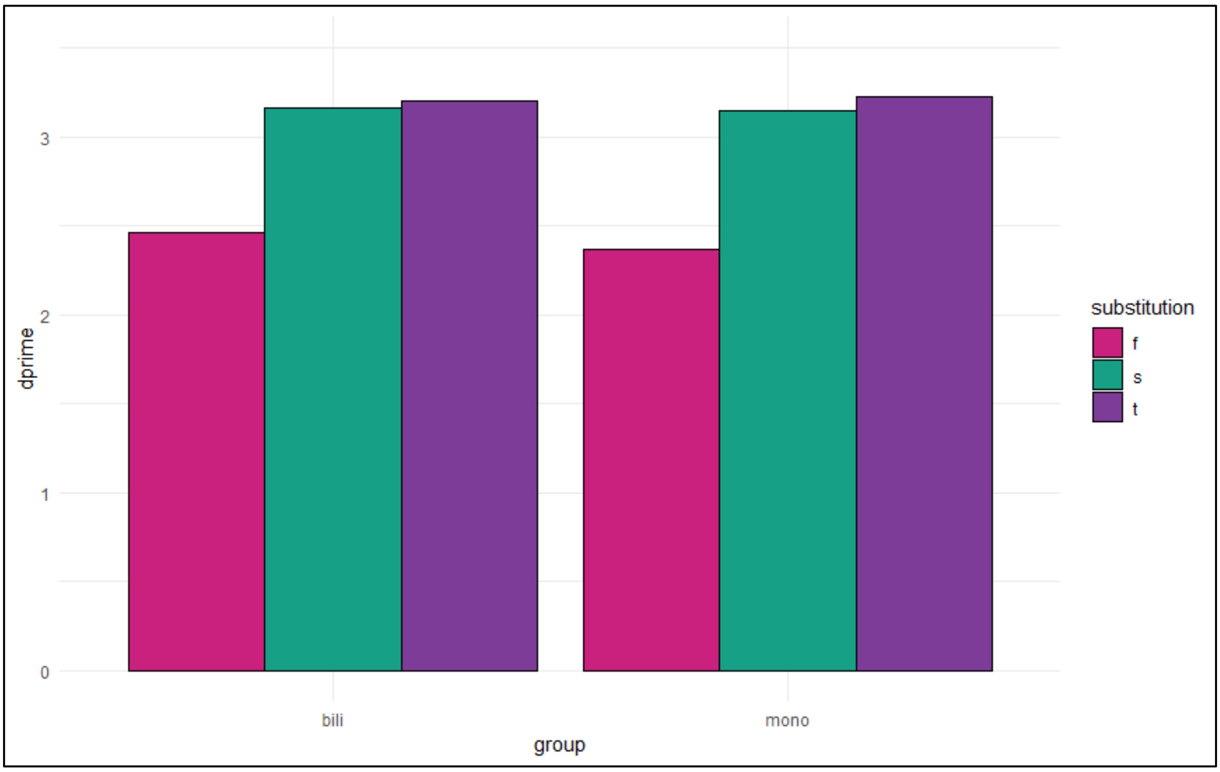

Figure 15: Discrimination accuracy by substitution and group

While speakers from both groups behaved very similarly in the perception of the interdental fricative, they did exhibit different substitution patterns in production. Since there were not enough instances of interdental fricatives to run an inferential model, only descriptive statistics are reported here. Figure 16 shows the preferred substitution patterns of the monolingual vs. that of the bilingual participants.

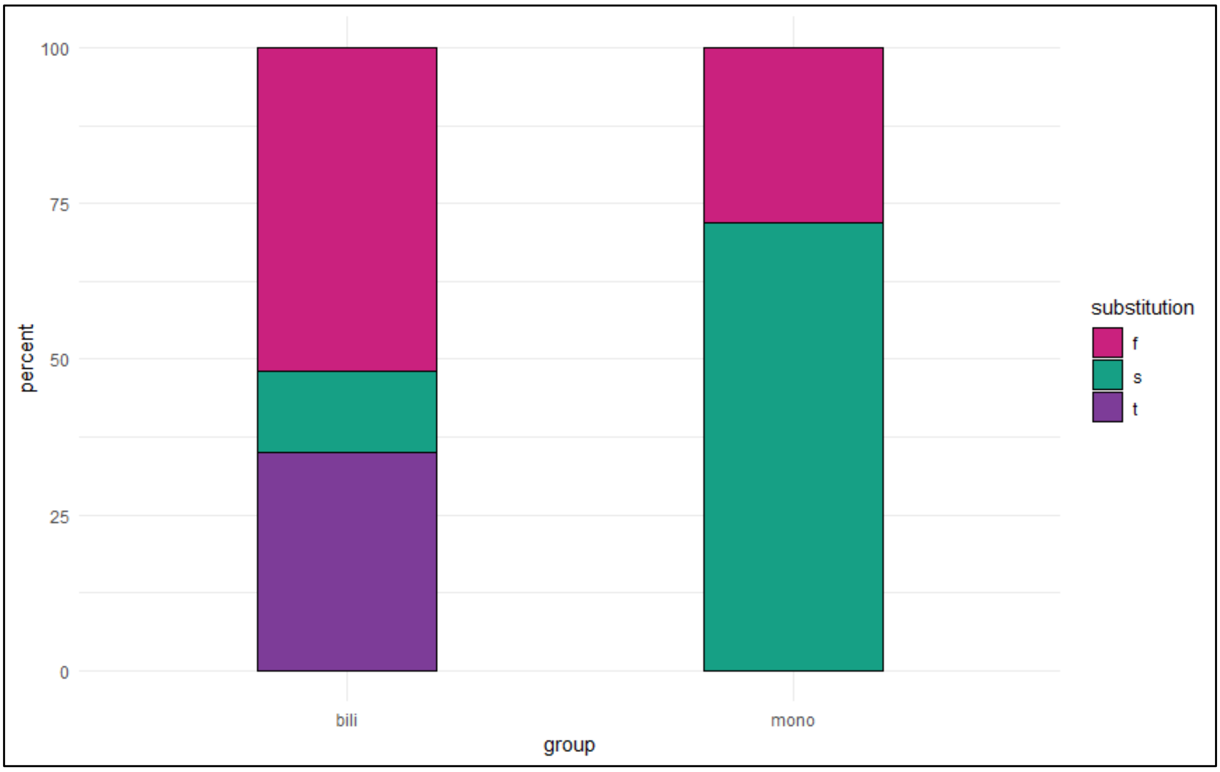

Figure 16: Substitution patterns in production by group

As can be seen, the monolinguals preferred to substitute the voiceless interdental fricative with [s] and also produced [f] in some instances[53]. Both these substitution options were also realized by the bilinguals, but they produced only a very small amount of substitutions using [s], whereas more than half of the instances were [f] substitutions. In addition, the bilinguals used [t] as a substitution for the interdental fricative.

4.2.2.7 Secondary variables

As has been reported earlier, none of the tested secondary variables had a significant effect on either perception or production of any of the features. To nevertheless explore further a potential relationship between the secondary variables in the bilinguals, this relationship was

[53] It should be noted here that the [f]-substitution in the monolinguals only involved one item, the word *thumb*. See the discussion section for more details.

tested using Pearson's correlations. The factors under scrutiny were TURKISH USE, GERMAN USE, ENGLISH USE, PWM, MOTIVATION, ATTITUDE towards German, and ATTITUDE towards Turkish. The analyses revealed a negative correlation between TURKISH USE and PWM (r=-0.82) as well as a positive correlation between ENGLISH USE and MOTIVATION (r=0.84). Subsequent visual inspection of the data, however, showed a rather substantial scatter of data points for TURKISH USE by PWM (see Figure 17), whereas motivation and English use indeed seem to be correlated (Figure 18).

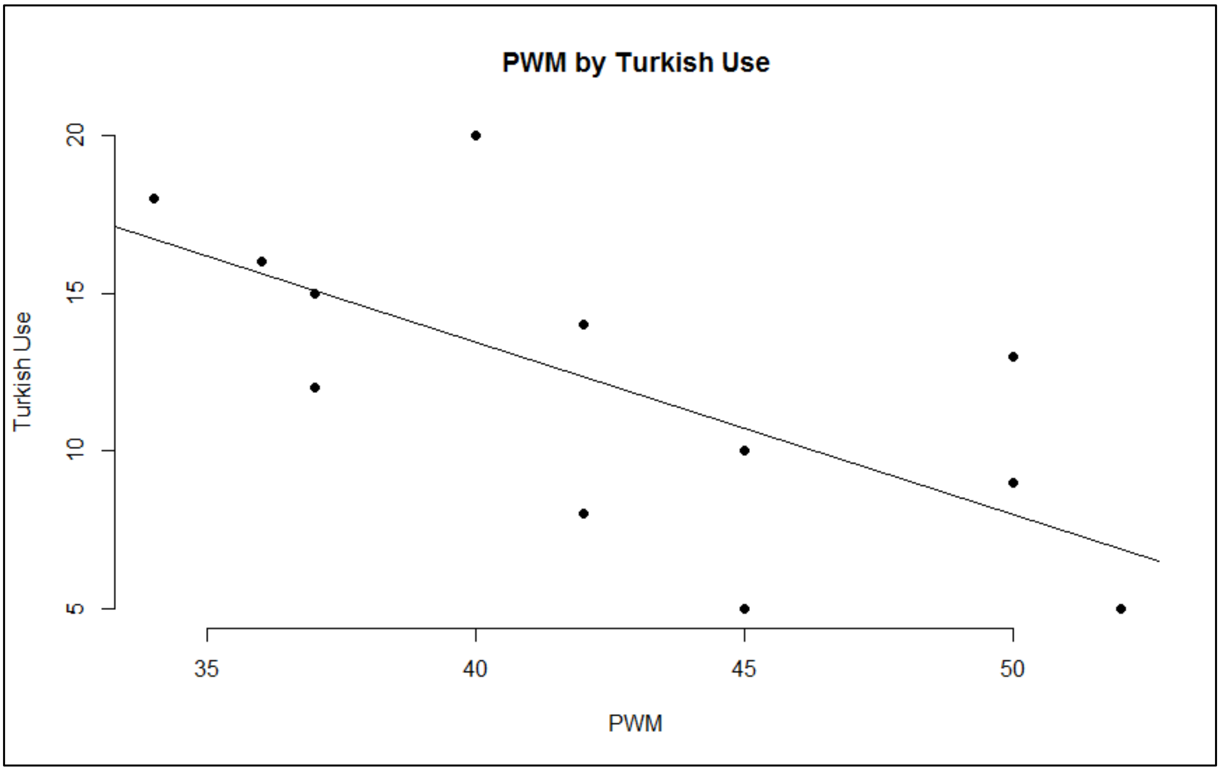

Figure 17: PWM by Turkish use

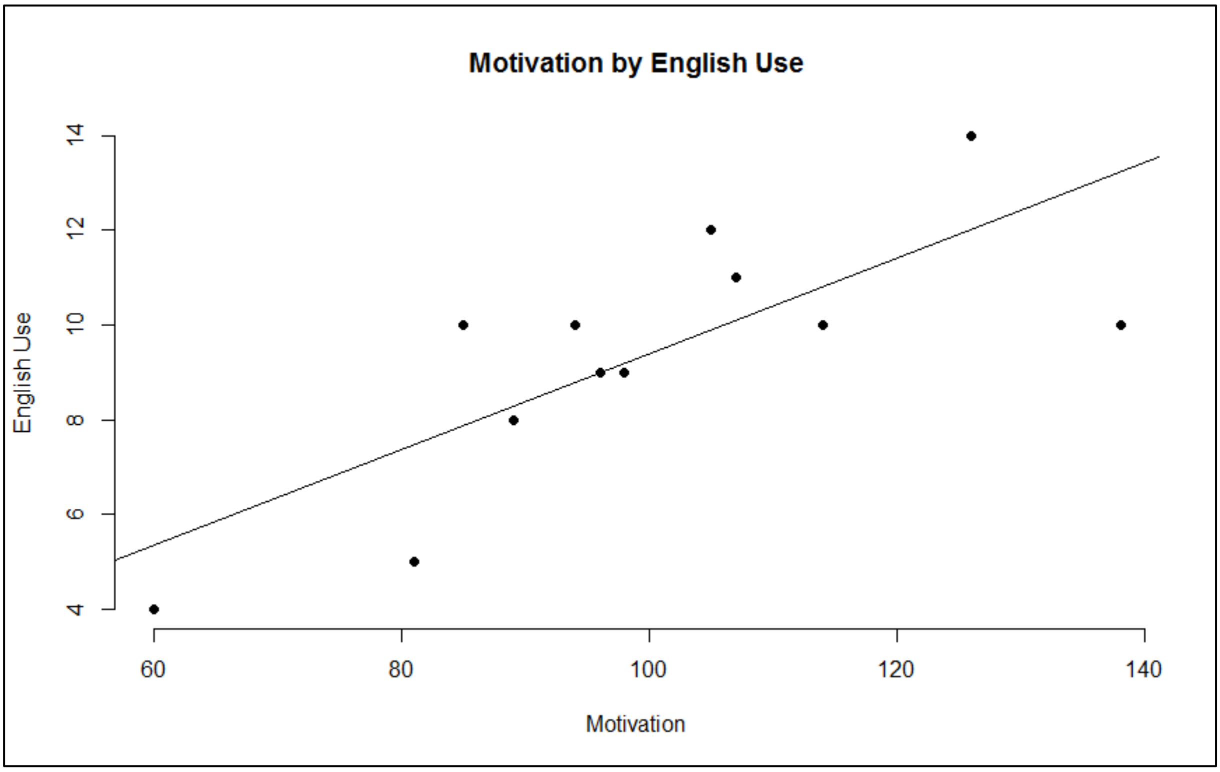

Figure 18: Motivation by English use

4.3 Discussion

4.3.1 Global discussion

In this study, I asked whether young Turkish/German HSs learning English differed from their monolingual peers in the perception of phonological features if their background languages supplied conflicting transfer information, i.e., if a feature could be transferred positively from one, but negatively from the other of the two languages. Secondly, I asked what would happen if neither of the background languages was a potential source of transfer, or, in other words, if the feature or contrast existed in neither of the background systems. In order to answer these questions, I chose phonological bits that would lend themselves to positive transfer from one and negative transfer from the other background language (Turkish and German, respectively), as well as bits of English phonology that were novel both for speakers of Turkish and speakers of German. Participants who were in their first or second year of formal

English education and between 10 and 13 years old were tested on the perception as well as the production of all these features to determine the effect of language-specific, universal, and extralinguistic factors.

First, it is important to note that neither PWM, nor MOTIVATION, or any other secondary variable such as AGE or GENDER, for that matter, were significant predictors of the perception or the production of the target features, and no significant correlation was found between motivation or PWM and the production or perception of any of the features. This lack of any effect can be attributed to the selection of test groups since the participants were well-balanced along the dimensions tested.

While there were pronounced group differences in the perception and at least some group differences in the production of the bits, the participants' performance finds no straightforward explanation through the initial 3-way grouping into feature types in either perception or production. Specifically, the monolinguals outperformed the bilinguals in the perception of phonemic vowel length (FT-G), voicing in coda position (FT-T), and /æ/ and /θ/ (both FT-E), whereas the bilinguals performed better than the monolinguals in the production of voicing in coda position. Group differences occurred in all of the feature types in perception and only in one of the clusters in production. It seems then, probably unsurprisingly, that the tested factors are further constrained by properties of the linguistic system and specific domains therein. A number of these properties are identified and examined in the remainder of this chapter.

Going back to the scenarios presented in Chapter 3.3, SLT predicted transfer from only one of the background languages in the speaker. In the case of the bilinguals, this could be either German (SLT:L2) or Turkish (SLT:HL). CBT predicted positive transfer wherever possible, meaning that features that are potentially supplied by one of the background languages would be used in the L3 by the multilingual learners. According to this second scenario, positive bit-by-bit transfer could thus be expected. Table 25 and

Table *26* list the concrete predictions made by SLT:L2, SLT:HL, and CBT for the individual features in perception and production, respectively. As stated in Chapter 3.2, consonant clusters and vowel length are in FT-G, voicing and laterals are in FT-T, and /æ/, /θ/, and /oʊ/ are in FT-E. Predictions were made as to whether bilinguals and monolinguals were expected to differ from each other, and if so, in which way. No specific predictions were made as to how native-like the speakers were going to be since the predictions are based exclusively on CLI effects, which is, however, not expected to be the only factor active in language learning. Instead, extralinguistic and universal factors were also expected to play a role. Thus, the mere presence of a bit in the background language(s) that allows positive transfer did not mean that the feature would necessarily be perceived or produced particularly target-like, or that it would happen to a degree that results in a successful native-like perception or production. Instead, the differences across groups were expected to give an insight into the effect of the background languages on L3 performance.

The results are coded in the following way: If a result matches the prediction from SLT:L2, it is marked in bold, if it matches the prediction from SLT:HL, it is marked in italics, and if it matches the predictions made by CBT, it is underlined. Combinations are possible. Predictions are marked in bold in columns 2-4 when the outcome is consistent with these claims.

Table 25: Predicted and actual results in the ***perception*** *of individual features*

Feature	SLT:L2	SLT:HL	CBT	Results
CC	**M=B**	M>B	**M=B**	**M=B**
Vowel length	M=B	**M>B**	M=B	*M>B*
Voicing	M=B	M<B	M<B	M>B
Laterals	**M=B**	M<B	M<B	**M=B**
/æ/	M=B	M=B	M=B	M>B
/θ/	**M=B**	**M=B**	**M=B**	***M=B***
/oʊ/	M=B	M=B	M=B	M>B

Table 26: Predicted and actual results in the ***production*** *of individual features*

Feature	SLT:L2	SLT:HL	CBT	Results
CC	**M=B**	M>B	**M=B**	**M=B**
Vowel length	**M=B**	M>B	**M=B**	**M=B**
Voicing	M=B	**M<B**	**M<B**	*M<B*
Laterals	**M=B**	M<B	M<B	**M=B**
/æ/	**M=B**	**M=B**	**M=B**	***M=B***
/θ/	**M=B**	**M=B**	**M=B**	***M=B***
/oʊ/	**M=B**	**M=B**	**M=B**	***M=B***

Neither SLT (both as SLT:L2 and SLT:HL) nor CBT fully manage to explain the results in either perception or production, which clearly encourages the application of multi-bit analyses: If, in this study, transfer scenarios had only been tested for vowel length and laterals, for instance, one might have been misled to assume that the HL influences L3 perception both negatively and positively, while, in production, it is the L1 German that exerts primary influence. However, this is clearly not the case for all of the bits, or systematically for all bits of any one of the feature types. Thus, a bit-by-bit explanation is needed to account for the results of this study as there is evidence for transfer from both background languages which cannot be explained using the categorization into feature types. In the following section of this chapter, I thus discuss the results bit by bit for both perception and production, and, afterwards, give an account of potential linguistic and extra-linguistic factors that are responsible for determining the most likely source of CLI in third language phonology.

4.3.2 Discussion of individual features

4.3.2.1 Initial consonant clusters

Both groups performed equally well in the discrimination of initial consonant cluster. In their production, both monolinguals and bilinguals produced a particularly high amount of target-like instances, and their performances did not significantly differ from each other. In monolingual L1 Turkish speakers growing up in Turkey, a lower percentage of target-like productions in their L2 English would be expected since clusters are likely to be broken by a (harmonized) epenthetic vowel (see, for instance, Yavaş, 1980; Clements and Sezer, 1982; more information in Chapter 3.2.1). Since the bilinguals in this study did not exhibit this behavior, the primary influence on their productions of CCs in English was most likely their L1 German rather than their HL Turkish. This may be because, here, vowel epenthesis is a repair process rather than an integral part of the Turkish phonological system and thus does not surface in the bilinguals' L3 English, especially given potential positive transfer from German. In other words, the negative transfer of epenthesis that is based in the repair strategy for a phonotactic constraint can be overridden relatively easily by another language in the language system that does not have this constraint. This, in turn, resulted in target-like consonant cluster productions in both monolingual and bilingual learners.

4.3.2.2 Vowel length

The monolinguals outperformed the bilinguals in the perception of the vowel length distinction. So, given that the Turkish phonological system does not employ contrastive vowel length distinctions, Turkish seems to have exerted a negative influence on the L3 performance in the bilingual learners. In the production of vowel length, however, just like in the production of CCs, both monolinguals and bilinguals performed on a par and exceptionally well although Turkish native speakers were expected, and are in fact reported (Rolffs, 2005)[54] to have difficulties with the target-like production of long vs. short vowels in an L2 that employs phonemic vowel length (see Chapter 3.2.1 for more information). HL influence seems to have been strong enough in the participants of this study to surface in the perception of the contrast. It should be noted that previous research has repeatedly shown that durational contrasts in vowels are relatively easy for L2 learners even if their L1 does not employ phonemic vowel duration (see, for instance, Cebrian, 2006, Altmann, Berger, & Braun, 2012). This ease with which vowel length contrasts are acquired is however only reflected in the production results of the present study, where the lack of the contrast in Turkish resulted in difficulties for the bilinguals in the perception but not in production.

[54] In her perception study, Nimz (2015) did not find an effect of the lack of phonemic vowel length in Turkish in the learner's L2 German, which she explains by the phonetic experience the Turkish speakers have due to their use of secondary lengthening as well as by the fact that vowel duration contrasts are generally easy to perceive.

4.3.2.3 Coda voicing

In the voicing of final coda position, the bilinguals performed significantly better than the monolinguals in production, which can easily be attributed to positive transfer to the L3 English from their HL Turkish, where voiced obstruents in coda position occur. However, in the perception of voicing in coda position, the monolinguals clearly outperformed the bilinguals, which both seems counterintuitive at first and contradicts the results obtained in the production part of this study. This apparent advantage for the monolingual participants might, however, have a different reason: Vowel duration is known to be a robust cue for the voicing contrast both in English (Port & Dalby, 1982) and in German (Port & O'Dell, 1985) where voiced obstruents are preceded by longer vowels than their unvoiced counterparts. Port and O'Dell (1985) actually showed that even in word-final position, where the voicing contrast is neutralized in German, speakers produce minute durational differences between vowels preceding underlyingly voiced (longer) and underlyingly voiceless (shorter) consonants. In a perception experiment conducted in the course of the same study, Port & O'Dell found that native speakers of German perform above chance level when discriminating word-final consonants based on this minute length difference in the preceding vowel. Since the present study found an advantage in the perception of vowel duration for the monolinguals, the monolinguals may have based their discriminatory decisions on vowel length as opposed to actual voicing as their primary cue, which, in turn, would result in their superior performance when compared to the bilinguals, which can also be supported by the fact that the bilinguals did not perform well in the perception of vowel length.

4.3.2.4 Laterals

Both in the perception and the production of laterals, the monolinguals and the bilinguals performed on a par, which suggests a lack of influence from Turkish, but leaves room for potential negative transfer from German. While overall scores were similar, the two groups differed in the distribution of the two allophonic variants of the lateral, both in perception and production. The monolinguals were significantly less accurate at discriminating the dark and the clear lateral in initial position than in final position, while the bilinguals discriminated equally well in both positions. This can be explained in different ways: Generally, participants may generally be more sucessful at perceiving the difference between the two different allophones in coda vs. onset position due to the vocalic cues that precede the coda, but not the onset. Furthermore, when the monolinguals start acquiring lateral allophony in English, they are likely to become aware at some point that English laterals differ drastically from German laterals in final but much less so in onset position. Thus, they will be more used to listening out for this type of difference in final than in onset position, where they would not expect a difference to begin with and are thus not as sensitive towards it. The bilinguals, on the other hand, have lateral allophony in their language system coming from Turkish, where the distribution, however, differs from that of English (see Chapter 3.2.2 for more details). Thus, they are familiar with both variants, but they might not be sensitive to the distinction as a function of the position in English that the respective allophone occurs in. They have to deviate from the pattern in one of their background languages, i.e., Turkish, in both positions and are thus equally sensitive to the distinction irrespective of position.

A group difference also surfaced when examining the target-like production of lateral allophones by position. Formant measurements showed that both groups clearly distinguished acoustically between the two allophones, but the distinction was stronger in the productions of the bilinguals. Specifically, the two groups produced similar values for Z2-Z1 in the dark variant but differed in the clear variant, where the bilinguals exhibited even higher values than the monolinguals and thus made a greater distinction between the two allophones. There was no overlap between the productions of the clear vs. the dark lateral in the bilinguals, whereas Z2-Z1 values did overlap across variants in the monolinguals' productions.

The way both groups treated the laterals in perception and production shows that different mechanisms are employed for both domains. In the perception of the target contrast, phonological distribution, and, crucially, informational load, are determining factors for the listeners: If they are used to perceiving a distinction in a particular environment, they will be able to perceive the same distinction in the same environment without major problems. Similarly, if listeners are used to paying attention to deviations from their native language only in a certain environment x but not in a different environment y, the distinction will be generally easier for them to detect in environment x than in environment y. In production as opposed to perception, on the other hand, the bilinguals may additionally benefit from the fact that the articulatory gesture required for the production of the dark lateral is already available to them from their Turkish. They can use this gesture and transfer it to their L3 English, which makes their production more target-like once the allophonic distribution is in place, which translates into the two lateral variants being more distinct than in the monolingual speakers, who are likely to struggle with this new gesture.

4.3.2.5 The vowel /æ/

In the perception of the /æ/-/ɛ/ contrast, the monolinguals outperformed the bilinguals, overall discrimination was excellent in the monolinguals and good, but more variable, in the bilinguals. The German vowel system is generally more crowded than the Turkish vowel system. Crucially, two German front vowels can potentially be perceived as equivalents for English /ɛ/ and /æ/, which are overall lower, but their relative distance from each other is similar to the relative distance between the two German vowels. Thus, German monolinguals can rely on the presence of a height distinction in their L1 front vowels and apply it to their English, and they can thus successfully discriminate between the two English vowels. The bilinguals seemed to use their Turkish rather than their German as a blueprint for vowel perception:[55] They only have one potential equivalent vowel, /e/, that can serve as a perceptual magnet for both English vowels in question. Thus, they are more likely to mismatch the stimuli on the same general vowel space, which, in turn, resulted in a lower overall d'-value. It is important to note at this point that, while significantly lower than in the monolinguals, the d'-value achieved by the bilinguals was still relatively high, which might be due to the German in their system. In other words, the results of the perception part of the experiment point to a measurable influence of the HL in Group 1 for the vowel /æ/, although the bilinguals may still benefit from their German. For one thing, then, simultaneous negative and positive influence from the background languages is possible in L3 perception and may result in higher variability. However, the fact that the less crowded system exerts an influence in addition to the more crowded one points towards a partial connection of the two systems: The two vowel systems seem to exist in parallel, on top of each other, but can be accessed separately.

The percentage of target-like productions of the vowel /æ/, on the other hand, was not significantly different across groups: Both monolinguals and bilinguals performed particularly poorly on the production of the /æ/-/ɛ/ contrast. The immediate acoustic difference between the two vowel sounds seems to be salient enough for non-native speakers to perceive even at this early stage, while producing the target sound faithfully seems to be much harder. This difficulty can be attributed to the articulatory properties of vowel segments in general: They are usually produced without any constriction, i.e., without the articulators touching. This lack of touch

---

[55] These assumptions are based on two distinct vowel systems in the bilinguals' Turkish and German where an apparent influence from Turkish-like structures comes from the Turkish system. It is not possible here to tease apart an influence from the bilinguals' Turkish from a potential influence from a Turkish-influenced German or an influence from the multi-ethnolect *Kiezdeutsch* as spoken by some young migrants in Germany.

also means a lack of haptic information and thus an increased difficulty when trying to fine-tune the articulation. In other words: Since there is no haptic anchor to facilitate articulation in vowels, their target-like production is particularly difficult, even when discrimination is accurate (and even easy).

While there was no significant effect of group in the categorical production of /æ/, acoustic vowel measurements showed that concrete productions did indeed differ across groups and that, specifically, bilinguals produced the target vowel higher than the monolinguals, but converged with the monolingual productions on the backness dimension. In fact, measurements of the concrete vowels produced by the native speaker in the experimental stimuli showed that the monolinguals were closer to the target height than the bilinguals. Additional control measurements of Turkish accented English (produced by monolingual L1 Turkish, L2 English speakers) revealed that the bilinguals tended to converge with monolingual native speakers of Turkish, giving more evidence for an influence of the participants' Turkish on their production of the target sound.

4.3.2.6 Voiceless interdental fricative

No group difference was found for the perception or the target-like production of the interdental fricative, with high success in perception and intermediate success in production. Moreover, the investigation of the perception of the interdental fricative and its individual substitutions revealed a pattern that was almost identical in the monolinguals and in the bilinguals: The contrast between /θ/ and /f/ was the hardest to perceive, followed by the contrast between /θ/ and /s/, and /θ/ and /t/. This rather unsurprising hierarchy can easily be explained by the acoustic properties of the segments in question: The two fricatives /θ/ and /f/ share acoustic properties that are notoriously difficult to distinguish without any additional cues (Kabak, 2019), for instance, visual cues or lexical context. A group difference did, however, surface in the production of the items with interdental fricatives, or more specifically, in the substitution patterns exhibited by the two groups. The German control participants mostly substituted [s] for /θ/, which is the pattern typically expected in native speakers of German. Native speakers of Turkish, on the other hand, are known to substitute the interdental fricative with [t]. The bilingual participants of this study used both [s] and [t] as well as [f] to substitute /θ/. This last substitution can be considered to be the universal, acoustically natural substitution for /θ/, and it is also found in a large number of native standard and non-standard varieties of English. Thus, it can be concluded that the bilinguals' substitution patterns emerged from both their L2 and their HL grammars, and that the interaction of the two grammars prompted the emergence of the universal pattern, which was not the case in the German participants, who only resorted to one supply grammar.[56]

4.3.2.7 The diphthong /oʊ/

No group difference was found in the production of the diphthong /oʊ/, with both groups performing well. In the perception of the same contrast, on the other hand, the monolinguals outperformed the bilinguals. While the specific segment /oʊ/ is not present in either German or Turkish, German phonology contains other diphthongs, whereas there are no diphthongs at all in Turkish. This shows that the bilingual participants have likely been influenced negatively by their Turkish, especially in the task that involved only acoustic cues, i.e., in the discrimination of the diphthong vs. the equivalent monophthong [o]. This difference could not be observed in the production task since both visual (i.e., lexical) and contextual cues

[56] In fact, some of the monolingual participants did substitute [f] for the interdental fricative. This was, however, restricted to one single lexical item, *thumb*. When asked about this specific lexical item after completing the tasks, the respective participants stated that they were not familiar with the word. Thus, they were probably trying to mimic the word they heard based on the acoustic information alone which probably led them to misidentify the interdental segment and produce the acoustically close labiodental segment instead.

were present that the bilinguals could make use of. Therefore, the bilinguals indeed faced a disadvantage in the perception of segments that do not occur in their HL when they had to rely on acoustic cues alone, but, in contrast to what would have been expected of monolingual native speakers of Turkish, they could make use of additional information due to their substantial experience with German diphthongs.

4.3.2.8 Perception and production: the interface

No correlation was found between the perception and the production of the features tested. In other words, if a contrast was perceived particularly well, this did not necessarily mean that production was also relatively target-like. This is in line with a multitude of previous studies: Bailey and Haggard (1973), for instance, tested the perception and production of voicing contrasts in initial stops and found no significant correlation between perception on production. Similarly, Ainsworth and Paliwal (1984) found no significant correlations when she tested the perception and production of English /w, r, l, j/. When taking into account a developmental perspective, studies have, in fact, found evidence for a nonsimultaneous development of the two domains. Usually, production is found to lag behind perception (see, for instance, Aoyama, Flege, Guion, Akahane-Yamada, & Yamada, 2004 for the /r/-/l/ contrast in Japanese learners of English, or Baker & Trofimovich, 2006 for vowels in Korean learners of English). Yet other studies have shown that the production of a contrast is sometimes more accurate than its perception. This was the case, for instance in Sheldon and Strange (1982), who tested the English /r/-/l/ contrast as perceived and produced by Japanese learners. While some type of link between the perception and the production of speech sounds cannot be disregarded, the literature shows that, ultimately, little is known about the the overall makeup of this link.

In the light of the disparity of previous findings it makes sense to assume a more multi-layered connection between perception and production and, specifically, that the perception of contrasts is only one factor that plays a role in the native-like production of speech sounds (see, for instance, Colantoni & Steele, 2008). Others are, for instance, articulatory difficulty or visual salience, or the type of segment the speech sound in question is. For instance, a consonant sound might be relatively difficult to discriminate acoustically when its acoustic structure is very similar to its counterpart (this is the case, for instance, in /θ/ vs. [f]), while it might be relatively easy to produce since its articulation is additionally based on a visual cue (again, in the case of /θ/ vs. [f], the difference in the position of the tongue, the teeth, and the lips are very salient visually).[57] In other words, the perception and production of speech sounds are not unrelated, but individual perceptual salience as operationalized by discrimination accuracy is merely one factor that potentially determines successful production.

In cases of conflicting transfer information from the background languages, e.g., in the case diphthong /oʊ/, where Turkish is a potential source of negative and German is a potential source of positive transfer, negative transfer from the HL could be observed in perception, but not in production. This was likely due to the lexical cues available to the participants in the production task. Pure segmental discrimination is thus more prone to an observable influence from the negative transfer source than context-bound production.

---

[57] Of course, in natural speech and in natural situations, speech is usually accompanied by the visual image of the speaker. However, most auditory discrimination protocols, including the one used in this study, only include auditory stimuli, highly restricting the cues that the participants can use to reach their decisions.

4.3.3 Determining factors in CLI

Generally, the feature type (i.e., the patterning according to potential transfer sources, see Chapter 3.2 for a detailed explanation) does not have a direct, straightforward relation to how the influence of the languages in the multilingual's mind surfaces in perception or production. Transfer, then, does not occur on the level of the language system as a whole but rather on the level of the individual feature. It has been shown that phonological **markedness** as well as articulatory difficulty are two factors that determine whether or not a potential negative influence from the HL on the L3 is active or not. Specifically, if a phonological feature in the L3 is relatively marked or if it is **articulatorily complex**, potential negative transfer from the background language will be active.[58] Furthermore, certain types of features will be challenging to produce and thus prone to negative transfer from the background language due to their lack of an **articulatory anchor**, which is particularly true for vowels, which are produced without constriction in the oral cavity and thus cannot be easily anchored haptically. Moreover, the slightly unexpected results for the production of initial consonant clusters and vowel length, as well as the results for the lateral distinction have shown that it is necessary to know as precisely as possible how a contrast works, i.e., one ideally needs to know about its acoustic, articulatory, but also phonological anatomy, and about what (if any) the functional load of the contrast is in order to make accurate predictions for transfer scenarios. Furthermore, the phonological system of the background language may have a more indirect, abstract influence on the TL that cannot be captured by the mere copying of background language structures onto the TL.

[58] This idea as such is, of course, not new and, for markedness, has already been formulated in a similar way by Eckman as early as 1977 as the *Markedness Differential Hypothesis*.

To sum up, the interaction of language-specific and universal factors determines (un-)successful perception and the production of target sounds by multilingual learners, whereby universal factors function as switches that condition transfer. It needs to be noted, however, that in the speakers of this study, who are still in the early stages of language acquisition, the interaction of elements in their multilingual repertoire may not yet be stable enough to produce results that support one of the one-dimensional models. Note also that the speakers tested in this study were in the process of dramatically changing their language system in terms of dominance: Their HL was losing prominence in their daily lives as their focus shifted away from the family towards their peers as their main point of social reference. This ongoing shift may have been one of the reasons for this combined language system. It is thus necessary to additionally test the initial hypotheses on data produced by more proficient L3 users with similar language profiles that may exhibit a more stable multilingual language system. This approach will give further evidence for the development of CLI in a language system and provide a basis for creating a working model that factors in a developmental perspective in addition to properties and characteristics of the target sounds. If, as I claim, universal factors modulate transfer (triggered by language-specific factors), a difference beyond the expectable effect of ongoing acquisition would point towards a changing influence (or an activation or deactivation) of these universal factors in the course of the acquisition process. This question will be explored in the next chapter by testing adult speakers of English whose language background is comparable to the participants of this study and by subsequently comparing the results of both studies.

5 Study 2: Highly proficient monolingual and bilingual L2 users of English

The results of Study 1 revealed bit-by-bit transfer in young bilingual learners of L3 English in the early stages of language learning. This transfer has been shown to be conditioned not only by the combinations of various background languages and the way their features relate to each other but also by feature-intrinsic (non-language specific) phonetic and phonological properties. Specifically, Study 1 could show that the bilinguals' production of their L3 English is influenced by their HL Turkish, and that whether or not this influence surfaces as transfer is conditioned by universal factors such as markedness and articulatory difficulty as well as by the type of contrast (specifically evidenced by the vowel-ness of a sound that is characterized by the absence of a constriction during articulation and thus the lack of a haptic anchor). In addition, the modality of psycholinguistic testing plays a role as perception and production do not align in revealing a clear picture of transfer.

The second study of this dissertation will explore the same phenomena as Study 1 in a learner group that is comparable biographically (i.e., the subjects will be HSs of Turkish with German as their societal language and English as their L3), but older (post-adolescent) and highly proficient users of their L3. By studying this additional sample in a quasi-longitudinal design, I hope to uncover general developmental effects as observed already in Hammarberg's seminal study (see Chapter 2.2.2 for a detailed review). Based on the studies discussed in Chapter 2.2, the older participants might exhibit different transfer patterns: Following the Hammarberg studies (see Chapter 2.2.2), the influence of the L1 on the target language may grow as acquisition progresses, whereas the influence of the L2 may decrease. Unmarked universals may surface less in the highly proficient speakers of English than in the younger participants of Study 1 due to the advanced state of acquisition in the former. However, based on Louriz (2007), universal factors are still expected to exert an influence on the perception and production of L2/L3 speech sounds.

5.1 Methods

The L3 users of English will be tested both in perception and production of English sounds using methods that tap the same mechanisms and representations as do the methods used in Study 1. Thus, the results of this study are expected to be comparable to those of Study 1. It was necessary, however, to adapt the tasks according to the advanced proficiency of the participants.[59] Furthermore, since neither PWM nor motivation as tested in Study 1 were significant factors in either the perception or the production in neither of the features, these were not tested in Study 2.[60] In the following sections, the participants in this study will be introduced, and the experimental design as well as the methods of analysis will be explained.

5.1.1 Participants

In Study 2 of this dissertation, two groups of highly proficient and frequent users of English were tested. At the time of testing, the speakers in both groups were students of English Language and Literature at the University of Würzburg at the time of testing, a study program whose courses are predominantly held in English. *Group 1* (n=12) were HSs (heritage speakers, see Chapter 2.1.1 for more details) of Turkish who had all grown up in Germany and attended German formal education from the age of around 3. For all these participants, Turkish was the primary family language while growing up, and it still is, while German is the language spoken outside the family context. As such, all participants from Group 1 are German dominant, both

[59] The way the tasks compare across studies will be explained in the respective passages.

[60] It might have been interesting to see whether an effect of these two factors emerges in more advanced speakers of L3 English. However, since the groups are, again, highly similar to each other in that they are, for instance, likely to exhibit a similar degree of motivation because they all study English at university level, I chose to introduce different factors to be able to cover a larger ground.

as self-reported and as quantified through their *Bilingual Language Profile* (BLP, see Section 5.1.4.3 for more detailed information on the method of assessment and the speaker profiles). They were between 18 and 32 years of age (mean=24.25, sd=2.65).[61] *Group 2* consisted of 12 native speakers of German who had grown up monolingual and were between 20 and 29 years old (mean=22.92, sd=3.43). Participants from both groups had started acquiring English as their first foreign language (L3 for Group 1, L2 for Group 2) between grade 2 and 4 in primary school. Figure 19 shows the age of the individual participants by Group.

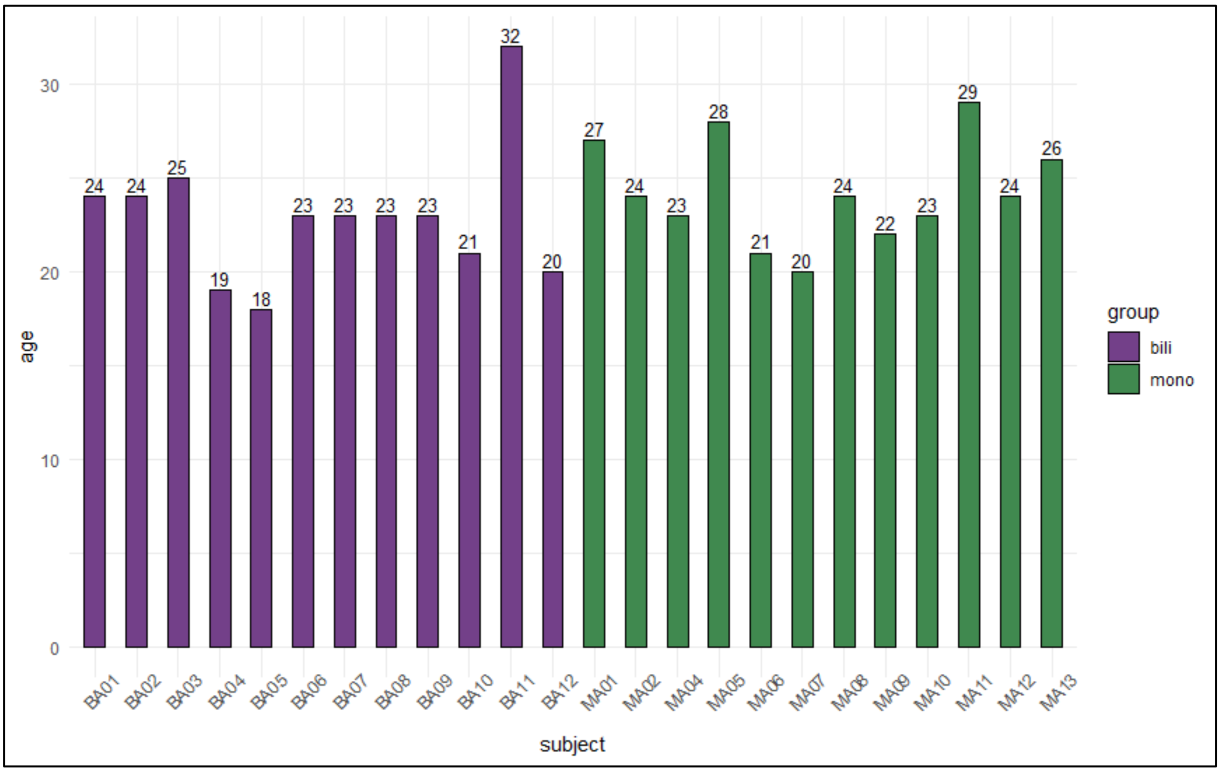

Figure 19: Age of individual participants

Additional languages in the participants' systems are listed in Table 27. Except for one bilingual speaker (BA01), all participants reported having knowledge of at least one additional language. In most cases, at least one Romance language per speaker was reported, mostly

[61] All background information was collected either by means of a questionnaire as part of the BLP (see Section 5.1.3 for more details) or during the sociolinguistic interviews conducted in the testing session.

French, but, in some cases, Spanish or Italian. Five of the participants had learned Latin either in school or at university as part of their teacher training, whereas only one of them (BA05) reported Latin[62] as their only additional language. Non-romance languages spoken by at least one of the participants are Mandarin Chinese, Arabic, and Korean. BA12 was exceptional in being an avid and highly interested language learner who reported having at least a basic knowledge of French, Italian, Korean, Mandarin Chinese, and Arabic in addition to her HL Turkish and her L1 German.

Table 27: Language backgrounds of individual participants

Participant	Additional languages
MA01	French
MA02	French, Latin, Spanish
MA04	Latin, Italian
MA05	French, Latin
MA06	French, Spanish
MA07	French, Italian
MA08	French, Italian
MA09	French, Spanish, Mandarin Chinese
MA10	French
MA11	French
MA12	French, Italian
MA13	French

[62] Latin, of course, is unlikely to have any influence on the phonological acquisition of any other language since it is learned in schools merely to be read and translated.

BA01	-
BA02	French
BA03	Arabic
BA04	Latin, French
BA05	Latin
BA06	Latin, Spanish
BA07	Spanish
BA08	French, Spanish, Latin
BA09	French, Arabic
BA10	Italian
BA11	Spanish
BA12	French, Italian, Korean, Mandarin Chinese, Arabic

As far as their biographical and language background is concerned, the participants in Study 2 can be considered to be an older subset of the participants in Study 1. Thus, even though they are all highly educated university students, which cannot be known of the younger participants, the study follows a quasi-longitudinal design.

5.1.2 Specific research questions

Building on the results from Study 1, I aimed to study how transfer outcomes differ as a function of growing proficiency and age.[63] In order to allow a comparison between the outcomes in Study 1 and the outcomes in Study 2, it is necessary to match the analyses as

[63] *Proficieny* and *age* cannot be teased apart due to the way this study has been designed: Since English instruction starts at a very similar time in all German schools, proficiency is ideally in a direct linear relationship with age since age corresponds to years of instruction.

closely as possible, which is why I asked the following questions that diverge only slightly from the questions in Study 1, specifically concerning the secondary variables (see above):

(1) Do Turkish/German adult HSs differ from their monolingual peers in the perception of phonological features if their background languages supply conflicting transfer information? What happens if neither of the background languages is a potential source of transfer?

(2) Can these differences in perception and production be explained by SLT, CBT, or bit-by-bit transfer?

(3) Are production and perception modulated by proficiency in any one of the speakers' languages, or, in case of the HSs, by language dominance?

(4) Is there a significant correlation between perception and production in general or for individual features?

In addition, I asked whether the factors that were found to exert an influence on L3 perception and production of the young learners do so in the adults as well and how (if at all) this influence changes with growing proficiency/age.

5.1.3 Material and procedure

5.1.3.1 General setting

Individual subjects from both groups were tested in a quiet university office between August 2017 and May 2018. Each subject was tested once, and the complete testing session took about 75 minutes for monolingual participants and about 90 minutes for bilingual participants. The tests were administered in the following order:

1. Perception task: AXB test
2. Production I: English reading task
3. Production II: Interview
4. Production III: German reading task
5. Production IV: Turkish reading text (only for bilinguals)

6. German cloze test
7. Turkish cloze test (only for bilinguals)
8. Biographical questionnaire (including BLP in the bilingual group)

The language of instruction was English for tasks 1 to 3 and German for tasks 4 to 8. The interviewer was a native speaker of German with native-like pronunciation in English. So, except for the two Turkish tasks,[64] the language of instruction matched the language of the task.

5.1.3.2 Perception: AXB discrimination

In order to assess perception accuracy, an ABX discrimination protocol was employed. This specific test was chosen for the highly proficient users of English because it increases cognitive load and thus is more likely to provoke underlying errors to surface while still tapping the same mechanisms as the AX protocol in Study 1. The items used in this task were identical to those used in the perception part of Study 1 (see Chapter 4.1.3.2).

Subjects were seated in front of a computer screen, and the test was administered using a Praat script.[65] I used the pseudowords that had been designed for the discrimination task in Study 1, so there were two test items for each of the features. Like in Study 1, each of the test items was paired with one item that included the expected deviation, with the exception of the interdental fricative, which was paired with three such equivalents since three possible substitutions are attested in the literature (for detailed information, see 3.2.3). The audio stimuli were presented via Philips SBC AH1000 headphones, as triplets with an ISI of 1500ms and a

[64] Since the interlocutor did not speak Turkish, German was chosen as the language of instruction. However, this reflects well the multilingual reality of the HSs since they are used to producing their Turkish in a non-Turkish environment and to switching languages (e.g., they might have to switch from German to Turkish when getting a phone call from a family member).
[65] A basic AXB script was retrieved from the Praat manual ("ExperimentMFC 3.2. An AXB discrimination experiment," 2016) and then modified according to the requirements of this study.

combination of a sine wave and white noise (1500ms) between items.[66] After the stimulus was presented, the participant was prompted to respond *A* or *B* on the keyboard – *A* if item A and item X were the same, and *B* if item X and item B were the same. Once the response had been given, the next stimulus was presented. Before the trials, there was a training phase of 8 stimulus triplets different from the test items. During this trial phase, subjects could ask questions and request clarification, whereas no feedback was given during the trial phase. Each of the test items was presented in 4 permutations: AB-A, AB-B, BA-A, BA-B, “A” representing the target-like production of the item and “B” representing the expected non-native production. Note that the “same” pair in each stimulus triplet always contained two different instances of the same item rather than two iterations of the same recording. Each of these triplets was repeated three times, making for a total of 216 stimuli that were presented to the participants in randomized order in 10 blocks of 20 and 1 final block of 16. Participants had the option of taking short breaks between blocks.

5.1.3.3 Production

For the production tasks, participants were seated at a desk, and speech was recorded using an Olympus Linear PCM Recorder LS-11. The audio data was recorded in wav-format (44.1 kHz, 16bit). Participants were assumed to be potentially less target-like and more variable in free speech than in carefully read speech. So, in order to assess such a potential task effect, two tasks were recorded. The first was a reading text to supply material comparable to the spoken material in Study 1 as well as a sufficient amount of tokens to analyze for each of the target features. The second task was a free interview to elicit spontaneous speech.

[66] An ISI of 1500ms (as opposed to a shorter ISI) was chosen since it triggers phonemic discrimination (see footnote 36, Werker and Logan, 1985). The combination of a sine wave and white noise was used as an inter-stimulus distractor.

5.1.3.3.1 Scripted speech: read text

The English reading text (see Appendix D.1 for full text) comprised a total of 1711 words. It was written to contain the target features /æ/ (n=56), initial consonant clusters (n=95), voiceless interdental fricatives (n=32)[67], laterals (n=222, final position n=77, initial position n=145), /oʊ/ (n=61), voicing in coda position (n=145), and contrastive vowel length (n=59). The test items were constructed to cover the largest possible range of phonological environments for the target feature to occur in. Participants were asked to read the text out loud on first seeing it. They were not given a chance to practice reading the text before the recording started. The experimenter was in the room during the reading tasks[68], but participants did not get any feedback on pronunciation errors. If, however, they noticed an error they had made, they were allowed to repair it. In these cases, the repaired utterance was the one considered for analysis. The length of the read text was between 9’4’’ and 12’43’’ minutes (mean=10’56’’).

To get speech samples in the participants’ background languages, they were presented with a German translation of “The North Wind and the Sun” (Appendix D.2). The bilingual participants additionally received a Turkish translation of the same text (see Appendix D.3) for full texts). This particular text was chosen since it is reasonably short and entertaining, it has been translated into a multitude of languages, and it has been used as a standard demonstration text for the phonology of different languages by the International Phonetic Association (International Phonetic Association, 1999). The German translation comprises 108 words, and

[67] The total number of interdental fricatives in the text was n=122, but voiced instances were excluded from analysis for better comparability with the other tasks as well as with Study 1, where only voiceless interdental fricatives were elicited.
[68] Notes were taken during the recordings of non-target-like productions and other irregularities.

the Turkish version comprises 71 words. As with the English reading task, participants were asked to read the text out loud, and their productions were recorded.

5.1.3.3.2 Free speech: interview

Furthermore, a semi-structured interview was conducted in English by the experimenter who asked questions about the biographical and linguistic background of the participants and their families, about attitudes (with a focus on bilingual identity in the bilingual group) and about their education and plans for the future (see the list of guiding questions in Appendix E). The data was collected both to examine the production of free speech as opposed to scripted speech and thus to uncover potential task differences and to gather additional background information about the participants. Interviews took between 5’24’’ and 17’33’’ minutes (mean=8’50’’) but were generally longer in Group 1 due to additional questions about their early bilingualism as well as their migrant background. All interviews were transcribed orthographically, and instances of all occurring test features were identified and coded in Praat for further analysis.

5.1.3.4 Secondary variables

5.1.3.4.1 L1 proficiency: c-tests

Proficiency in the L1 and the HL was assessed using c-tests (German and Turkish for the bilinguals and German only for the monolinguals). C-tests, first proposed by Klein-Braley and Raatz (1984), are texts in which parts of words have been removed and have to be filled in by the participant. The c-test was developed from the classical *cloze test*, in which complete words are removed and have to be filled in.[69] The German c-test was taken and adapted from Schmid and Dusseldorp (2010) and consisted of 4 short individual texts of between 80 and 100 words with 20 gaps (partial words) in each text, making for a total of 80 gaps that had to be filled. The first sentence of each text was given in full, and after that, every other word was clipped by half or, if the word had an uneven number of letters n, by (n+1)/2. The Turkish c-test was taken from Montrul (1997) and consisted of one single text that included 30 gaps. The gaps had to be filled with complete words rather than with parts of words due to the agglutinative nature of Turkish and the high likelihood of only grammatical morphemes being deleted when cutting off the second half of the word. In fact, the reliability of c-tests has been disputed over the last decades due to this very fact (see Ulusoy, 2008 for an empirical evaluation and discussion). I used them despite these disadvantages since it is the simplest way to assess general holistic proficiency and the ability to use morphosyntactic as well as contextual information to use a language without adding additional tasks.

Proficiency scores were calculated as the number of gaps filled in with the correct solution (or an alternative judged as acceptable by a linguistically trained native speaker of German or Turkish, respectively). The total proficiency score for German was 80, and the total proficiency score for Turkish was 30. The difference in the total score is not problematic since proficiency in Turkish and German were not compared to each other directly but rather used separately as a potential factor across participants. Thus, no normalization of this data took place.

[69] The cloze test has been shown to be problematic in practice since solutions are often ambiguous and scoring may thus often be subjective. See Beinborn, Zesch, and Gurevych (2014) for a discussion an empirical assessment of different proficiency tests.

5.1.3.4.2 Language dominance: Bilingual Language Profile

In order to assess language dominance, or, more precisely, the degree of language balance, in the bilinguals, the BLP was administered to all bilingual participants. The BLP conceptualizes language dominance as a gradient, relativistic measure by taking into account different factors such as exposure, language attitudes, proficiency, linguistic environment, processing ability and combining them to obtain a composite score for each of the languages, which are then combined to form a score of relative language dominance in bilingual speakers ("Need for the BLP"). As such, the instrument captures the multi-faceted nature of language dominance, and it furthermore provides a continuous score that allows for fine-grained assessment. The original BLP questionnaire (Birdsong, Gertken, & Amengual, 2012) was translated into German and adapted for Turkish-German bilingual speakers by the author, using the template provided on the BLP website. The questionnaire consists of four parts assessing (1) language background, (2) language use, (3) proficiency, and (4) language attitude. Each of the parts receives an individual module score based on the answers given, and both languages are scored separately so that each participant receives four raw module scores for German and four raw module scores for Turkish. These individual scores are then normalized by multiplying them with the factors suggested by the BLP manual in order to make sure that the parts receive equal weighting (see Appendix F for scoring and calculation directions). Finally, the scores for Turkish are subtracted from the score for German to obtain the final BLP score. Scores between -218 and +218 are possible with values close to 0, suggesting a high bilingual balance and values close to the extreme suggesting strong dominance in either direction. In the calculations used in this study, negative numbers signify Turkish dominance while positive numbers suggest German dominance.

In addition to assessing the BLP scores, participants were asked about their attitudes towards the languages they speak and about their general language background. In some cases, the answers given in the interview and the scores obtained through the BLP match, while in other cases they seem to be contradictory. In other words, no clearly identifiable pattern links the participants' free statements from their interviews and the scores calculated from the responses given in the BLP questionnaire. Taken together, this shows that the BLP scores indeed quantify a concept that does not straightforwardly reflect the participants' self-professed language attitudes, but seems to touch upon more deeply ingrained, more complex structures of use, dominance, and attitudes.

In its original form, the BLP questionnaire also contains a short section to assess biographical data (name, age, sex, place of residence, highest level of formal education). Since this study is specifically concerned with the interplay of various languages in the multilingual mind, it was necessary to extend the questionnaire to include questions on additional languages spoken and on the parents' language background, as well as on the time spent abroad. This extended biographical questionnaire was administered to both the bilingual and the monolingual group.

5.1.4 Data analysis

5.1.4.1 Perception

Just as in the perception task of Study 1, perception accuracy was operationalized as d'. Note, however, that the coding was slightly different since instead of a simple yes-no task in which the participant has to decide whether the stimulus in question is signal or noise (or, in other words, whether the two items in the stimulus are the same or different), participants in Study 2 were asked to perform what Stanislaw and Todorov (1999, p. 141) call a forced-choice task. In such a forced-choice task, the participants are presented with one signal stimulus and one or more noise stimuli.

Here, participants were presented with a sequence of items ABX. Item X was the target item, i.e., the item the other two items had to be compared to. One of the two remaining items, either A or B, matched X. The participant was asked to choose either A or B. In order to avoid a sequential bias, the d'-matrix (see Table 28) was adapted to accommodate the slightly different data structure.[70]

Table 28: d' matrix for forced-choice task

	Response: yes (A matches X)	Response: no (B matches X)
Stimuli: yes (A matches X)	HIT	MISS
Stimuli: no (B matches X)	FALSE ALARM	CORRECT REJECT

Whenever A was the same as the target stimulus X, this was counted as a stimulus (or "yes") response, and whenever B was the same as X, this counted as noise (or "no"). Thus, when a stimulus was correctly identified (i.e., when the participant indicated that A=X and this was correct), the response was counted as a Hit. This, as can be seen in the above table, was done similarly for all other possible stimuli-response combinations.

Finally, d'-scores for all features were calculated for each participant. Like in Study 1, three individual d'-scores were calculated for the interdental contrast to be able to tease apart the discrimination of the interdental versus its three most common substitutions ([f], [t], [s]). The lowest possible d'-score in this task was 0, and the highest was 3.499821.

[70] Stanislaw and Todorov (1999, p. 143) suggest to success rate (%) as an operationalization of sensitivity in forced choice tasks as the responses are not expected to be biased. However, since the items are structured in such a way that one of the possible choices is (temporally) closer to the target item than the other, it was deemed necessary to take into account the possibility of biased responses, which is why d'-prime was chosen as a more precise measure for this task as well.

5.1.4.2 Production

Production ability was operationalized as success rate (%). For this, each token of each feature that occurred in the reading text was coded as either target-like or non-target-like by two phonetically trained judges, on whose basis mean rates were calculated. To analyze free speech, all interviews were transcribed orthographically, and potential occurrences of the target features in the participants' turns were identified and then rated and coded as +/- target-like. For the interdentals, substitutions were additionally noted down and mean rates calculated for all non-target-like productions. As stated earlier, substitutions were not only identified auditorily. Instead, notes about articulation taken by the experimenter during the interviews were consulted, and Praat spectrograms were inspected visually. Like in Study 1, the items were judged by two phonetically trained raters. Agreement analyses using Cohen's Kappa (see Chapter 4.1.4.2 for a more detailed account) showed an inter-rater agreement of $\kappa = 0.82$, which is considered near perfect agreement.

In addition to the auditory judgments, monolinguals' and bilinguals' productions of the vowel /æ/ were measured. The first four measurable instances of /æ/ in the English reading passage were manually segmented and then measured (F1, F2, F3) for each participant using a semi-automated script. Measured formant values were then normalized using the Bark normalization method (Traunmüller, 1997). For more information on the procedure and technical specifications of the normalization process, see Chapter 4.1.4.2.

5.1.4.3 Secondary variables

Detailed descriptions of the elicitation and scoring methods used for the secondary variables of this study, i.e., for German and Turkish proficiency and language dominance, can be read in Section 5.1.3.4. The following table (Table 29) gives an overview of all secondary variables with their concrete operationalizations as used in the analyses in this study.

Table 29: Secondary variables and their operationalizations

Variable	Operationalization
TURKISH PROFICIENCY	Proficiency score Higher score ≙ higher proficiency
GERMAN PROFICIENCY	Proficiency score Higher score ≙ higher proficiency
BLP-SCORE, INDIVIDUAL (TUR, GER)	BLP Tur/BLP Ger Higher score ≙ higher degree of use
BLP-SCORE, COMBINED	BLP score Closer to 0 ≙ more balanced Positive ≙ German dominant Negative ≙ Turkish dominant

5.1.4.4 Statistical analyses

As was the case in Study 1, the effect of the variables on the perception and production of English phonological features in the adult L3 users was tested using linear mixed-effects models fitted with the lme4 package (Bates et al., 2015) in R (R Core Team, 2018). Main models were fitted for perception and production[71] each, one containing FEATURE (M1) and

[71] The reading task was considered the main production task since it provides a sufficient amount of controlled items for each feature and since it is more comparable to the task in Study 1 than the free speech task is. This is why the main models are run on the data obtained from the read speech, while further additional analyses will be conducted that include both tasks. These analyses will be explained further in the relevant sections of this chapter.

one containing FEATURE TYPE (M2) as one of the independent variables. Further independent variables in the original models were GROUP, AGE, and GERMAN PROFICIENCY, and SPEAKER and WORD were added to the model as random factors. The ideal models were determined using a top-down model reduction approach (for a detailed description of model reduction, see Chapter 4.1.4.4). Additional models were fitted on subsets of the data – the effect of language dominance on perception and production was only tested for the bilingual participants, for instance – in the same way. These additional models will be described in more detail in the relevant results sections. Linearity, homoskedasticity, and normality were tested visually (using residual plots) for all models, and all models were shown to conform with the necessary assumptions, i.e., there were no deviations from linearity, homoskedasticity, or normality. Significant main effects and interactions were determined by pair-wise model comparisons. Post-hoc tests performed on the data were pairwise comparisons with Bonferroni corrections using the lsmeans package (Lenth, 2016) as well as Pearson's correlations using the Hmisc package (Harrell, 2019).

5.2 Results

5.2.1 Descriptive statistics: participant profiles

Dominance scores of all 12 participants in Group are positive, meaning that they are German-dominant, albeit to different degrees. The mean BLP score observed in the participants of this study is 45.32 (sd = 25.29). The highest BLP score was obtained by A2005 (BLP = 84.37), and the lowest score was obtained by A2012 (BLP = 6.72). Figure 20 shows the BLP scores for all participants.

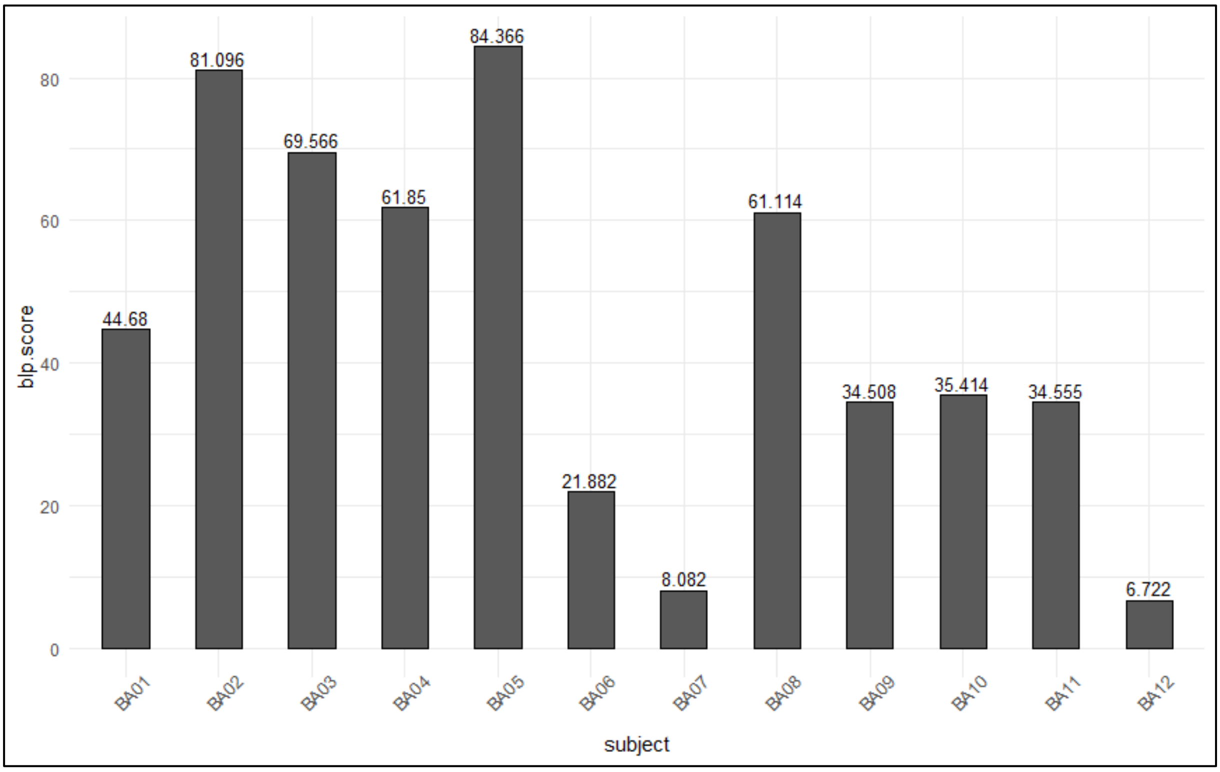

Figure 20: BLP scores of individual participants

5.2.2 Inferential statistics: Perception and production

5.2.2.1 Global results: perception

For performance on the perception task, linear mixed-effects modeling (see Table 30 for the final model) revealed a main effect of FEATURE on perception, but no effect of GROUP, AGE, or GERMAN PROFICIENCY. No significant interactions were shown that affect perception in the present data, which means that bilinguals and monolinguals perform similarly with overall performance modulated by feature type.

Table 30: Linear mixed-effects model M1 for perception

<table>
<tr><td colspan="6">Random effects:</td></tr>
<tr><td>Groups</td><td>Name</td><td colspan="2">Variance</td><td colspan="2">Std. Dev.</td></tr>
<tr><td>Speaker</td><td>(Intercept)</td><td colspan="2">0.05898</td><td colspan="2">0.2429</td></tr>
<tr><td>Word</td><td>(Intercept)</td><td colspan="2">0.06349</td><td colspan="2">0.2520</td></tr>
<tr><td colspan="6">Number of obs: 432, Groups: Speaker, 24; Word, 18</td></tr>
<tr><td colspan="6">Fixed effects:</td></tr>
<tr><td></td><td>Estimate</td><td>Std.Error</td><td>df</td><td>t value</td><td>Pr(>|t|)</td></tr>
<tr><td>(Intercept)</td><td>3.22656</td><td>0.2044</td><td>20.0194</td><td>15.787</td><td>9.18e-13***</td></tr>
<tr><td>Feature (CC)</td><td>-0.3144</td><td>0.2804</td><td>17.8663</td><td>-1.121</td><td>0.277</td></tr>
<tr><td>Feature (th)</td><td>-0.2976</td><td>0.2290</td><td>17.8663</td><td>-1.300</td><td>0.210</td></tr>
<tr><td>Feature (lateral)</td><td>-1.9800</td><td>0.2804</td><td>17.8663</td><td>-7.061</td><td>1.44e-06***</td></tr>
<tr><td>Feature (V length)</td><td>-0.0443</td><td>0.2804</td><td>17.8663</td><td>-0.158</td><td>0.876</td></tr>
<tr><td>Feature (ou)</td><td>-0.6398</td><td>0.2804</td><td>17.8663</td><td>-2.282</td><td>0.035*</td></tr>
<tr><td>Feature (voicing)</td><td>-0.4172</td><td>0.2804</td><td>17.8663</td><td>-1.488</td><td>0.154</td></tr>
</table>

Post-hoc comparisons with Bonferroni corrections show that the lateral is perceived significantly more poorly than all the other features. It is important to note, however, that the maximum d'-score in this task was 3.499821, so discrimination for all contrasts except the lateral is strong to intermediate to begin with. Table 31 shows the scores for all participants (combining both groups).[72]

[72] No by-group hierarchy is given here due to the lack of group difference.

Table 31: Perception hierarchy for all participants

Feature	Feature type	d'-score
/æ/	FT-E	3.23
vowel length	FT-E	3.18
/θ/	FT-G	2.93
CC	FT-E	2.91
voicing	FT-G	2.81
/oʊ/	FT-T	2.59
lateral	FT-T	1.25

The results for M2 (which contains FEATURE TYPE instead of FEATURE) show similar results (see Table 32): There is a main effect of FEATURE TYPE, but no significant effect of GROUP, AGE, or GERMAN PROFICIENCY, and there are no significant interactions.

Table 32: Linear mixed-effects model M2 for perception

<table>
<tr><td colspan="6">Random effects:</td></tr>
<tr><td>Groups</td><td>Name</td><td colspan="2">Variance</td><td colspan="2">Std. Dev.</td></tr>
<tr><td>Speaker</td><td>(Intercept)</td><td colspan="2">0.05992</td><td colspan="2">0.2448</td></tr>
<tr><td>Word</td><td>(Intercept)</td><td colspan="2">0.22659</td><td colspan="2">0.4760</td></tr>
<tr><td colspan="6">Number of obs: 432, Groups: Speaker, 24; Word, 18</td></tr>
<tr><td colspan="6">Fixed effects:</td></tr>
<tr><td></td><td>Estimate</td><td>Std.Error</td><td>df</td><td>t value</td><td>Pr(>|t|)</td></tr>
<tr><td>(Intercept)</td><td>2.9201</td><td>0.1633</td><td>21.5171</td><td>17.881</td><td>2.14e-14***</td></tr>
<tr><td>FT-G</td><td>0.1271</td><td>0.2909</td><td>17.9741</td><td>0.437</td><td>0.66723</td></tr>
<tr><td>FT-T</td><td>-0.8921</td><td>0.2909</td><td>17.9741</td><td>-3.067</td><td>0.00665**</td></tr>
</table>

Post-hoc comparisons show that the perception of features from FT-T is significantly weaker than that of both FT-G (p=0.035, df=21.31, t=2.688) and FT-E (p=0.0268, df=21.31, t=2.688), which is probably due to the very low d'-scores for the lateral as shown above. There is no significant difference between the perception of FT-G and the perception of FT-E contrasts (p=0.9156, df=21.31, t=-0.401). Figure 21, Figure 22, and Figure 23 show d'-scores by FEATURE and GROUP ordered by FEATURE TYPE. These figures show very clearly that both groups perform very similarly in the perception of the tested contrasts.

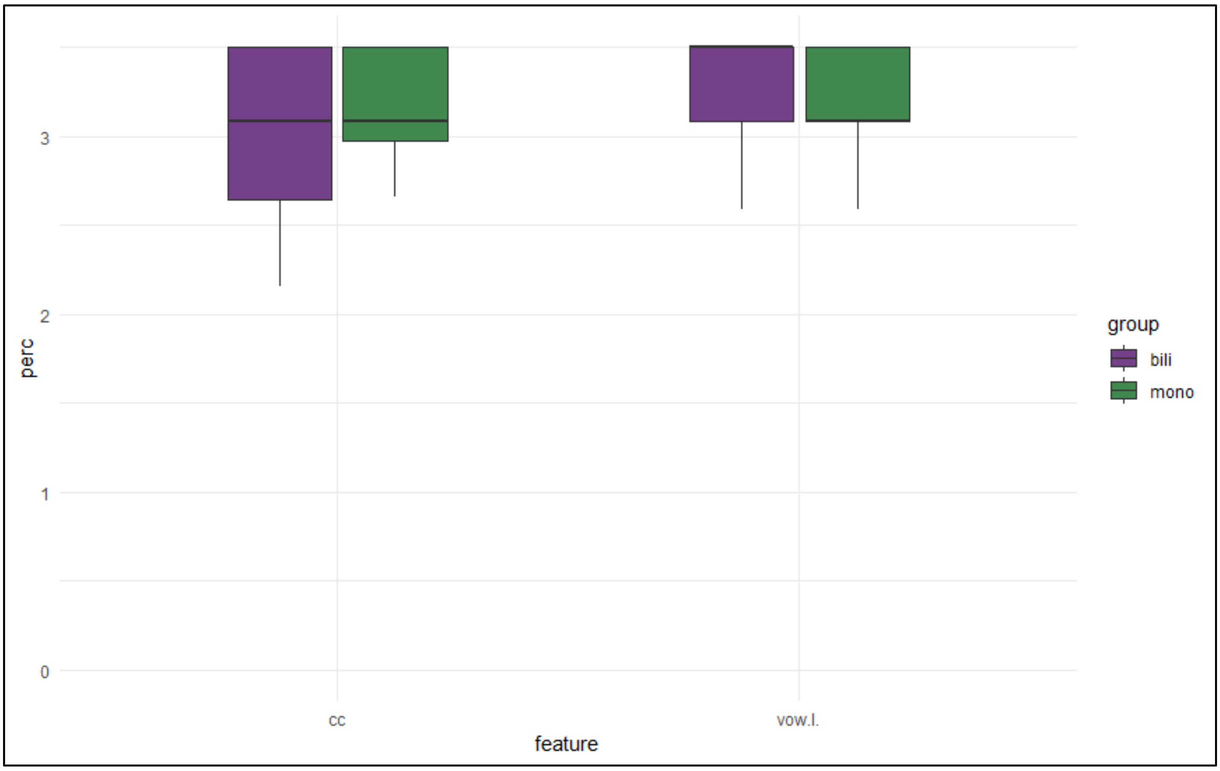

Figure 21: Perception accuracy (%correct) FT-G

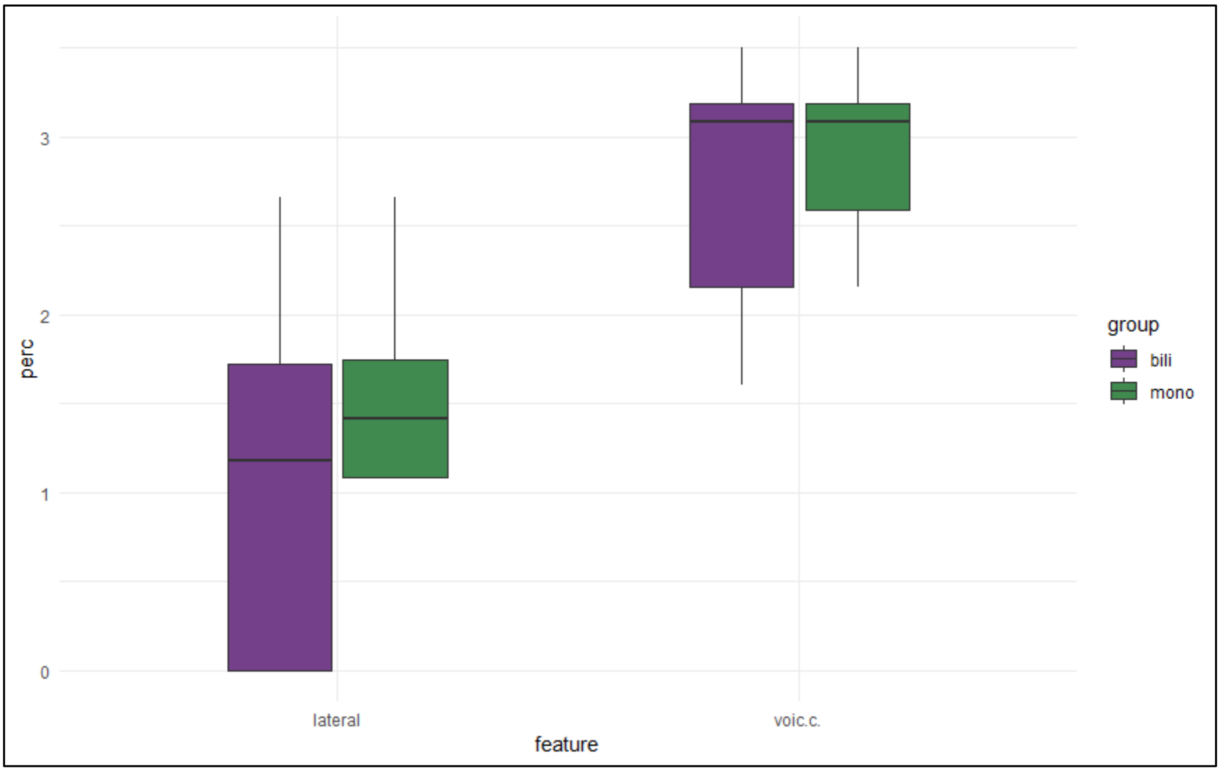

Figure 22: Perception accuracy (%correct) FT-T

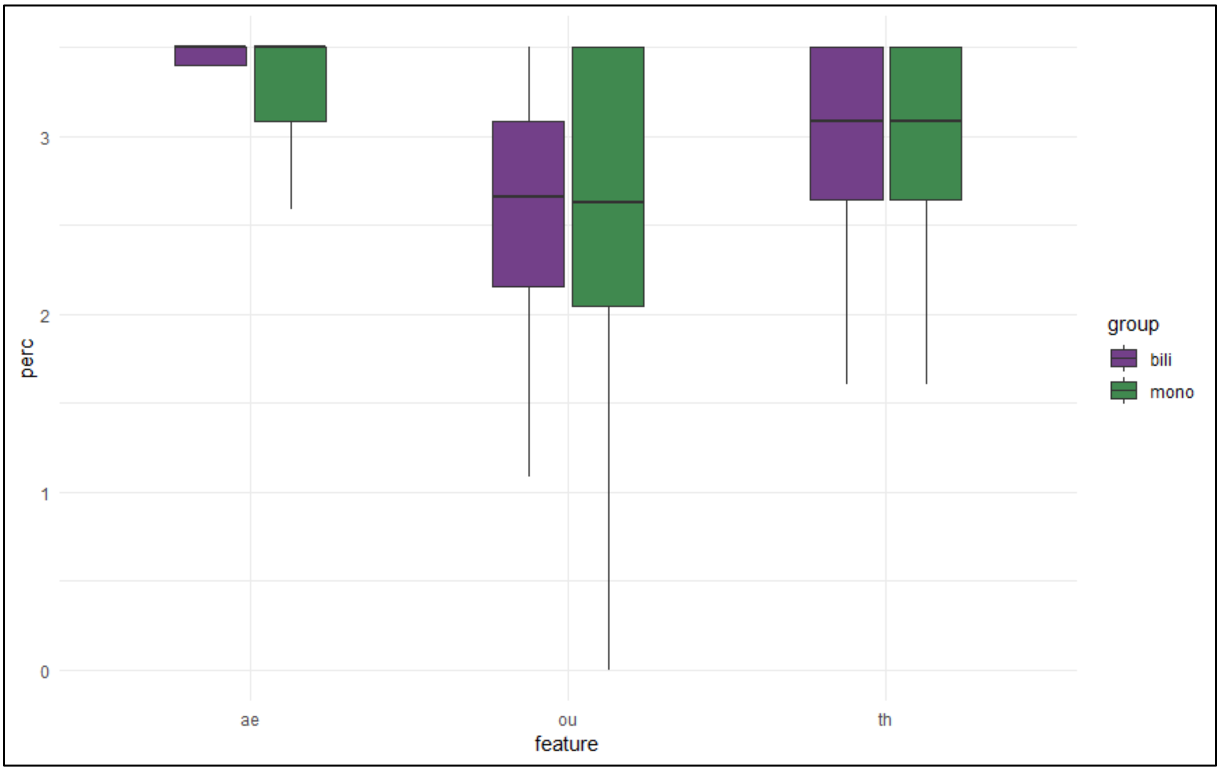

Figure 23: Perception accuracy (%correct) FT-E

An additional model was fitted using only the bilingual data to test for a potential effect of TURKISH PROFICIENCY, GERMAN PROFICIENCY, TURKISH USE, or BLP SCORE on the perception of any or all of the contrasts by the speakers in Group 1. However, top-down model comparisons revealed no main effect of any of these secondary variables, and no interactions between FEATURE and any of the variables. Instead, the final model only contained FEATURE as a predictor, as can be seen in Table 33.

Table 33: Linear mixed-effects model effect of TURKISH PROFICIENCY, GERMAN PROFICIENCY, TURKISH USE, and BLP SCORE on bilingual perception

Random effects:					
Groups	Name	Variance	Std. Dev.		
Speaker	(Intercept)	0.34082	0.2203		
Word	(Intercept)	0.09993	0.3161		
Number of obs: 216, Groups: Speaker, 12; Word, 18					
Fixed effects:					
	Estimate	Std.Error	df	t value	Pr(>|t|)
(Intercept)	3.249	0.261	12.335	12.441	<2.4e-08***
Feature (CC)	-0.298	0.358	11	-0.833	0.422758
Feature (th)	-0.306	0.293	11	-1.045	0.318431
Feature (lateral)	-2.112	0.358	11	-5.895	0.000104***
Feature (ou)	-0.609	0.358	11	-1.7	0.117203
Feature (voicing)	-0.483	0.358	11	-1.348	0.204706
Feature (V length)	0.023	0.358	11	-0.065	0.949615

No post-hoc tests were performed since the model only contained a subset of the data, and no interaction was shown earlier that contained GROUP. Thus, the feature hierarchies produced by the bilingual speakers alone cannot give any additional information.

5.2.2.2 Global results: production

As explained in Section 5.1.3.3 of this chapter, the production part of this study contained two main tasks, a reading task, and a free speech. The results from the reading task, the primary production task, will be analyzed first. The linear mixed-effects model with GROUP, FEATURE, AGE, and GERMAN PROFICIENCY as independent, PRODUCTION ACCURACY as dependent and SPEAKER and WORD as random variables[73] (see Table 34 for the final model) reveals a main effect of FEATURE and GROUP as well as a significant interaction of FEATURE and GROUP.

Table 34: Linear mixed-effects model M1 for production (reading)

Random effects:					
Groups	Name		Variance	Std. Dev.	
Speaker	(Intercept)		52.86	7.271	
Word	(Intercept)		143.17	11.965	
Number of obs: 15611, Groups: Speaker, 24; Word, 357					
Fixed effects:					
	Estimate	Std.Error	df	t value	Pr(>|t|)
(Intercept)	47.941	2.731	64.157	17.556	<2e-16***
Group (bili)	-16.157	3.489	42.858	-4.632	3.37e-05***
Feature (CC)	53.962	2.027	9026.978	26.628	<2e-16***
Feature (th)	37.788	3.197	2734.067	11.819	<2e-16***
Feature (lateral)	25.192	1.837	7858.028	13.715	<2e-16***
Feature (ou)	55.023	2.393	6411.741	22.996	<2e-16***
Feature (voicing)	-11.815	1.968	7844.190	-6.005	2.00e-09***

[73] Note that BLP SCORE, TURKISH PROFICIENCY, and TURKISH USE were not included in the model since these are only available for the bilingual participants and thus cannot be tested in a model that combines both groups.

Feature (V length)	47.535	2.417	5702.841	19.668	<2e-16***
Group (bili) : Feature (CC)	16.149	2.309	15201.244	6.995	2.76e-12***
Group (bili) : Feature (th)	14.054	3.098	15201.035	4.536	5.78e-06***
Group (bili) : Feature (lateral)	24.604	2.053	15201.849	11.982	<2e-16***
Group (bili) : Feature (ou)	16.141	2.532	15201.443	6.375	1.88e-10***
Group (bili) : Feature (voicing)	14.139	2.205	15204.282	6.411	1.48e-10***
Group (bili) : Feature (V length)	14.045	2.557	15201.666	5.492	4.03e-08***

Post-hoc comparisons revealed that the bilinguals outperformed the monolinguals in the production of the laterals (p=0.0184, Df=inf, z=-2.357), while the monolinguals performed better than the bilinguals in the production of /æ/ (p<.0001, Df=inf, z=4.233). Table 35 and Table 36 show the production hierarchies for the monolinguals and the bilinguals, respectively.

Table 35: Monolinguals' production hierarchy (reading)

Feature	Feature type	% target-like
/oʊ/	FT-E	103.05
CC	FT-G	101.98
vowel length	FT-G	95.56
/θ/	FT-E	85.81
lateral	FT-T	**73.21**
/æ/	FT-E	**48.02**
voicing	FT-T	36.21

Table 36: Bilinguals' production hierarchy (reading)

Feature	Feature type	% target-like
/oʊ/	FT-E	102.87
CC	FT-G	101.81
vowel length	FT-G	93.28
/θ/	FT-E	83.54
lateral	FT-T	**81.49**
voicing	FT-E	34.03
/æ/	FT-T	**31.7**

When comparing the two production hierarchies, the only difference can be observed in the bottom part of the hierarchy, where the bilinguals perform equally poorly on voicing and on /æ/, while the monolinguals perform poorly only on the voicing contrast. It needs to be noted that overall performance is strong in both groups for all contrasts except for the vowel /æ/ and the voiced final obstruent, both of which are produced target-like in less than 50% of all instances by both groups.

In M2 (see Table 37 for the final model), a significant effect is revealed for FEATURE TYPE. The productions of all feature types differ from each other significantly: The rate of successful production is highest for FT-G (105.99), followed by FT-E (70.1) and FT-T (61.73). This can be easily attributed to the fact that laterals and voicing, both features of FT-T, are produced relatively poorly by both groups. Neither GROUP nor AGE or GERMAN PROFICIENCY predict production accuracy in the reading task. There is, however, a significant interaction between GROUP and FEATURE TYPE.

Table 37: Linear mixed-effects model M2 for production (reading)

Random effects:					
Groups	Name	Variance	Std. Dev.		
Speaker	(Intercept)	50.53	7.108		
Word	(Intercept)	445.24	35.599		
Number of obs: 15611, Groups: Speaker, 24; Word, 357					
Fixed effects:					
	Estimate	Std.Error	df	t value	Pr(>|t|)
(Intercept)	73.334	2.620	58.246	27.995	<2e-16***
Group (bili)	-6.502	3.137	30.228	-2.072	0.04686*
FT-G	29.116	1.677	12133.403	17.360	<2e-16***
FT-T	-13.904	1.424	12242.376	-9.765	<2e-16***
Group (bili):FT-G	5.675	1.668	15224.027	3.402	0.00067***
Group (bili):FT-T	11.094	1.426	15224.953	7.781	7.65e-15***

Post-hoc comparisons revealed that the production hierarchy by feature type differs across groups in the following way:

monolinguals: FT-G (102.52) > FT-E (73.41) > FT-T (59.5)

bilinguals: FT-G (101.55) > FT-E (66.76) = FT-T (63.95)

In other words, bilinguals produce features of FT-E and FT-T at a similar level, whereas monolinguals perform more poorly in FT-T than in FT-E. There is no difference between the two groups' in FT-G. Figure 24, Figure 25, and Figure 26 illustrate production performance by group, contrast, and feature type.

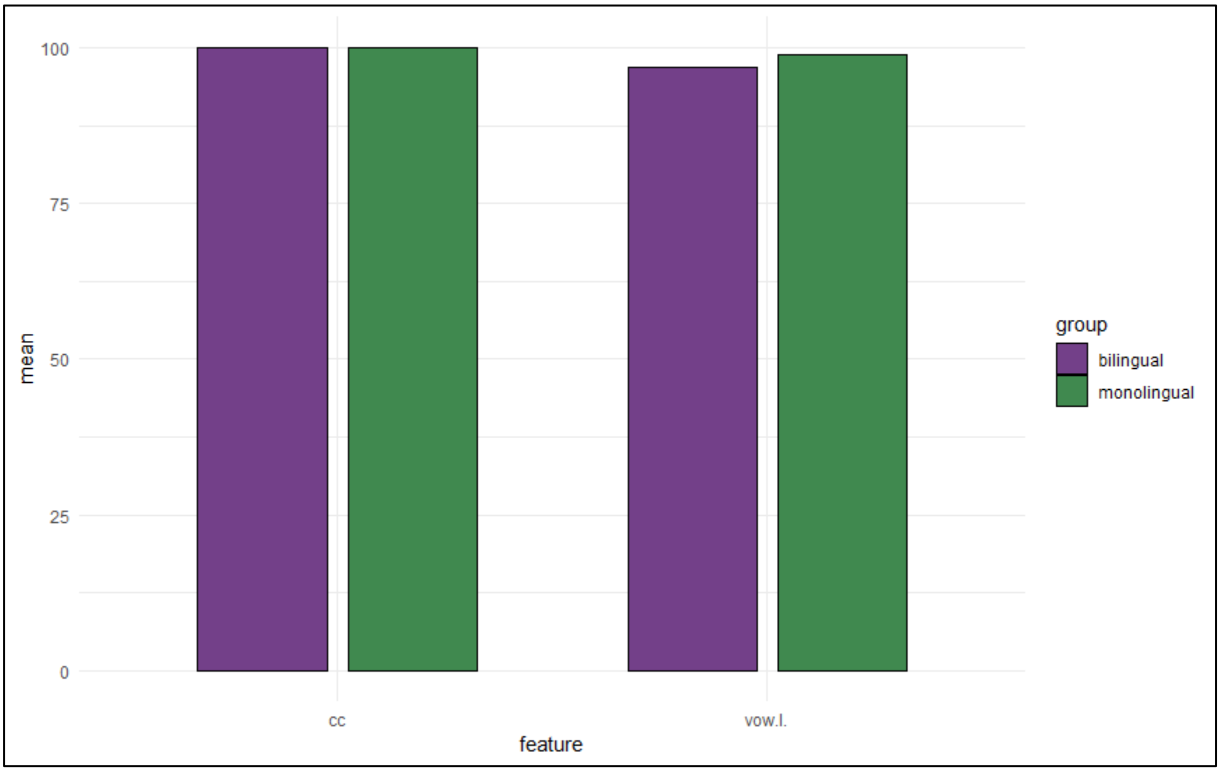

Figure 24: Production accuracy (% correct) FT-G

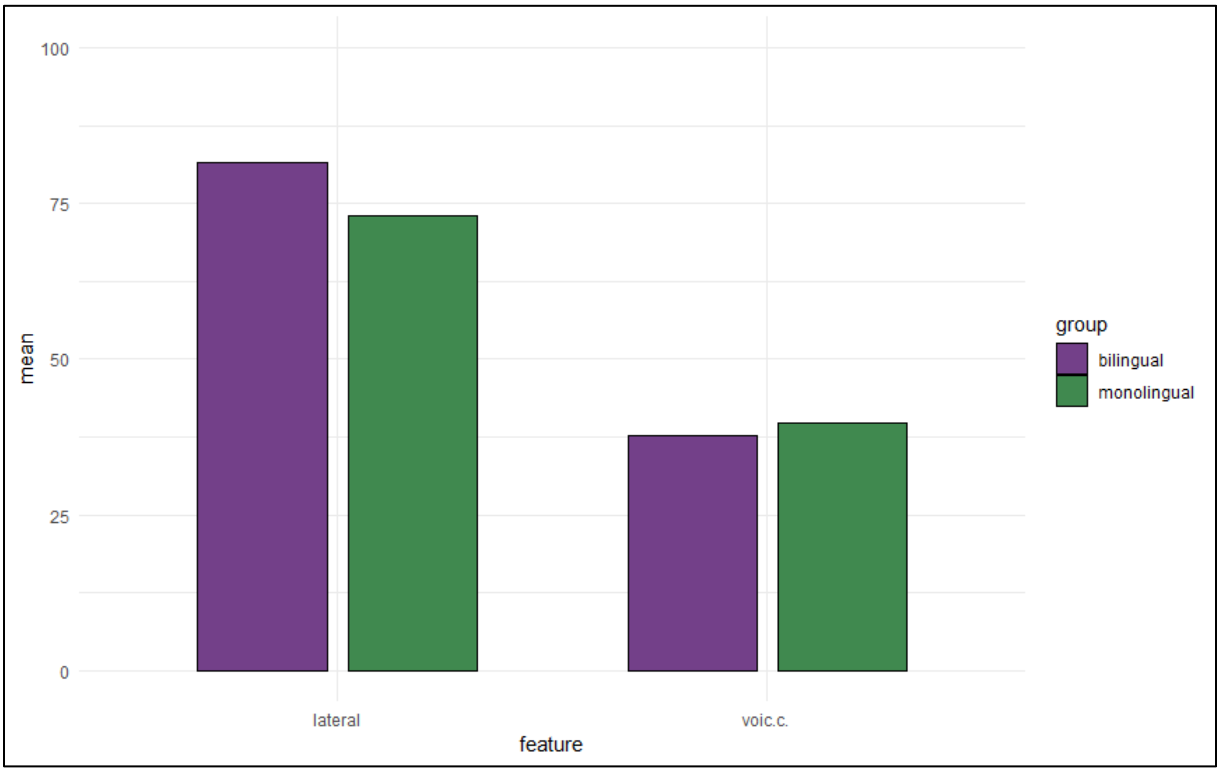

Figure 25: Production accuracy (% correct) FT-T

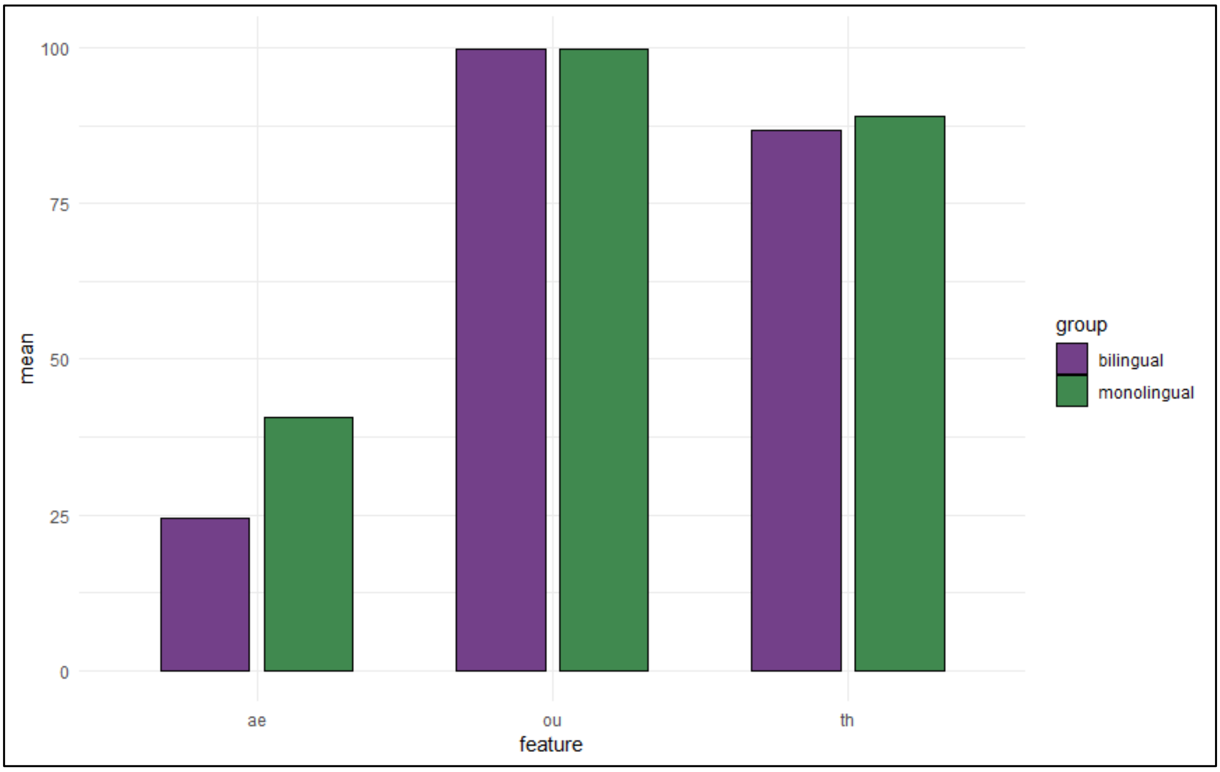

Figure 26: Production accuracy (% correct) FT-E

The data from the free speech task was analyzed in the same way as that of the reading text. The linear mixed-effects model with GROUP, FEATURE, AGE, and GERMAN PROFICIENCY as independent, and WORD and SPEAKER as random variables (see Table 38 for final model) revealed a main effect of FEATURE, but no other main effects and no interactions between any of the variables.

Table 38: Linear mixed-effects model M1 for production (free speech)

Random effects:					
Groups	Name		Variance	Std. Dev.	
Speaker	(Intercept)		29.03	5.388	
Word	(Intercept)		110.89	10.530	
Number of obs: 6104, Groups: Speaker, 24; Word, 738					
Fixed effects:					
	Estimate	Std.Error	df	t value	Pr(>|t|)
(Intercept)	41.982	2.251	243.199	18.653	<2e-16***

Feature (CC)	57.399	2.367	1853.142	24.246	<2e-16***
Feature (th)	45.452	3.524	731.498	12.896	<2e-16***
Feature (lateral)	25.354	2.126	2256.087	12.396	<2e-16***
Feature (ou)	55.095	2.859	974.602	19.272	<2e-16***
Feature (voicing)	-8.991	2.337	2190.066	-3.847	2.00e-09***

The production hierarchy from the free speech task can be seen in Table 39. Lines between two features denote statistically significant differences. Please note that vowel duration is missing since it was not analyzed in the free speech task due to a lack of controlled conditions in the material. Thus, in all analyses involving the free speech task, vowel duration will not be considered.

Table 39: Production hierarchy all participants (free speech)

Feature	Feature type	% target-like
CC	FT-G	99.43
/oʊ/	FT-E	97.15
/θ/	FT-E	87.51
lateral	FT-T	68.39
/æ/	FT-E	42.05
voicing	FT-T	33.07

When the features are grouped according to FEATURE TYPE in M2 (see Table 40 for final model), FEATURE TYPE is shown to be the only variable tested to have a significant effect on production in free speech.

Table 40: Linear mixed-effects model M2 for production (free speech)

<table>
<tr><td colspan="6">Random effects:</td></tr>
<tr><td>Groups</td><td>Name</td><td colspan="2">Variance</td><td colspan="2">Std. Dev.</td></tr>
<tr><td>Speaker</td><td>(Intercept)</td><td colspan="2">26.83</td><td colspan="2">5.18</td></tr>
<tr><td>Word</td><td>(Intercept)</td><td colspan="2">324.07</td><td colspan="2">18.00</td></tr>
<tr><td colspan="6">Number of obs: 15611, Groups: Speaker, 24; Word, 357</td></tr>
<tr><td colspan="6">Fixed effects:</td></tr>
<tr><td></td><td>Estimate</td><td>Std.Error</td><td>df</td><td>t value</td><td>Pr(>|t|)</td></tr>
<tr><td>(Intercept)</td><td>66.3541</td><td>2.4682</td><td>107.7970</td><td>26.884</td><td><2e-16***</td></tr>
<tr><td>Group (bili)</td><td>0.1361</td><td>2.8936</td><td>51.7312</td><td>0.047</td><td>0.96266</td></tr>
<tr><td>FT-G</td><td>35.0514</td><td>2.7619</td><td>4643.1396</td><td>12.691</td><td><2e-16***</td></tr>
<tr><td>FT-T</td><td>-6.4760</td><td>2.1763</td><td>4996.4249</td><td>-2.976</td><td>0.00294**</td></tr>
<tr><td>Group (bili):FT-G</td><td>-0.4554</td><td>3.1351</td><td>6037.9568</td><td>-0.145</td><td>0.8845</td></tr>
<tr><td>Group (bili):FT-T</td><td>-6.6964</td><td>2.4155</td><td>5988.6619</td><td>-2.772</td><td>0.00558**</td></tr>
</table>

As obtained by pairwise comparisons, the production hierarchy for feature types is FT-G (101.3) > FT-E (66.75) > FT-T (56.44). As opposed to the test on M1, there is a significant interaction between Feature type and Group in M2. Here, while all three feature types are produced significantly differently from each other by the bilinguals resulting in the following production hierarchies:

monolinguals: FT-G (101.94) > FT-E (65.58) = FT-T (59.12)

bilinguals: FT-G (101.94) > FT-E (67.32) > FT-T (54.14)

The monolinguals produce contrasts in FT-T at the same level as those in FT-E, while all feature types are produced differently by the bilinguals. Notice that this is the opposite trend from that observed in the reading task.

In order to reveal a potential task effect, both tasks were then put into a single model to test for interactions between Group, Feature, and Task in all possible combinations. A model was fitted with Group, Feature, Task, Age, and German proficiency as independent, and

SPEAKER and WORD as random variables (see Table 41 for the final model). Main effects were found for FEATURE and TASK, but not for any of the other variables. The main effect for FEATURE is not surprising since this main effect has also been found in both tasks separately, and, thus, it will not be reported here in detail since the comparisons on the individual results are deemed to be more fine-grained.

Table 41: Linear mixed-effects model M1 for both production tasks

<table>
<tr><td colspan="6">Random effects:</td></tr>
<tr><td>Groups</td><td colspan="2">Name</td><td colspan="2">Variance</td><td>Std. Dev.</td></tr>
<tr><td>Speaker</td><td colspan="2">(Intercept)</td><td colspan="2">51.67</td><td>7.188</td></tr>
<tr><td>Word</td><td colspan="2">(Intercept)</td><td colspan="2">141.9</td><td>11.912</td></tr>
<tr><td colspan="6">Number of obs: 21101, Groups: Speaker, 24; Word, 992</td></tr>
<tr><td colspan="6">Fixed effects:</td></tr>
<tr><td></td><td>Estimate</td><td>Std.Error</td><td>df</td><td>t value</td><td>Pr(>|t|)</td></tr>
<tr><td>(Intercept)</td><td>42.913</td><td>3.266</td><td>138.167</td><td>13.139</td><td><2e-16***</td></tr>
<tr><td>Group (bili)</td><td>-1.886</td><td>4.256</td><td>99.871</td><td>-0.443</td><td>0.6587</td></tr>
<tr><td>Feature (CC)</td><td>57.073</td><td>3.04</td><td>16414.707</td><td>1.877</td><td><2e-16***</td></tr>
<tr><td>Feature (th)</td><td>43.381</td><td>4.164</td><td>10471.6</td><td>10.417</td><td><2e-16***</td></tr>
<tr><td>Feature (lateral)</td><td>26.516</td><td>2.757</td><td>17887.265</td><td>9.616</td><td><2e-16***</td></tr>
<tr><td>Feature (ou)</td><td>55.124</td><td>3.507</td><td>14359.176</td><td>15.72</td><td><2e-16***</td></tr>
<tr><td>Feature (voicing)</td><td>-7.181</td><td>3.116</td><td>16505.08</td><td>-2.305</td><td>0.02118***</td></tr>
<tr><td>Task (reading)</td><td>4.347</td><td>3.026</td><td>13156.273</td><td>1.436</td><td>0.1509</td></tr>
<tr><td>Group (bili) : Feature (CC)</td><td>2.094</td><td>3.796</td><td>20993.406</td><td>0.551</td><td>0.58131</td></tr>
<tr><td>Group (bili) : Feature (th)</td><td>-1.308</td><td>4.843</td><td>20963.398</td><td>-0.27</td><td>0.78710</td></tr>
<tr><td>Group (bili) : Feature (lateral)</td><td>-2.123</td><td>3.457</td><td>21049.117</td><td>-0.614</td><td>0.539</td></tr>
<tr><td>Group (bili) : Feature (ou)</td><td>2.395</td><td>4.042</td><td>21078.111</td><td>0.593</td><td>0.55344</td></tr>
</table>

Group (bili) : Feature (voicing)	-3.788	3.846	20996.25	-0.985	0.32462
Group (bili) : Task (reading)	-14.296	3.643	21076.264	-3.924	8.72e-05***
Feature (CC) : Task (reading	2.739	3.636	16394.765	-0.753	0.45121
Feature (th) : Task (reading)	-9.276	4.594	15415.377	-2.019	0.04347*
Feature (lateral) : Task (reading)	0.422	3.294	16915.411	0.128	0.89815
Feature (ou) : Task (reading)	-0.24	4.053	16489.29	-0.059	0.95277
Feature (voicing) : Task (reading)	-3.893	3.683	15657.128	-1.057	0.29064
Group (bili) : Feature (CC) : Task (Reading)	14.083	4.515	21075.792	3.119	0.00182**
Group (bili) : Feature (th) : Task (reading)	21.879	5.545	21080.165	3.946	7.97e-05***
Group (bili) : Feature (lateral) : Task (reading)	26.759	4.084	21061.317	6.552	5.80e-11
Group (bili) : Feature (ou) : Task (reading)	13.763	4.581	21018.997	2.837	0.00455**
Group (bili) : Feature (voicing) : Task (reading)	18.097	4.499	21077.948	4.022	5.78e-05***

Significant interactions involving TASK were found for TASK and FEATURE as well as TASK and GROUP, and there is a three-way interaction between TASK, GROUP, and FEATURE.

Taking together both groups and all features, the performance in the reading task was significantly better than the performance in the free speech task (p=0.0056, Df=Inf, z.ratio=-2.773, estimate=-2.46). However, while the bilinguals exhibit an overall higher degree of target-like productions (p=0.0023, Df=Inf, z.ratio=-3.047, estimate=-3.19), there is no significant effect of task on overall production in the monolinguals (p=0.133, Df=Inf, z.ratio=1.502, estimate=-1.73).

The lateral is performed more target-like in read than in free speech (p<.0001, df=inf, z.ratio=-9.113) by both monolinguals and bilinguals, which can also be seen in Figure 27 and Figure 28. Moreover, the bilinguals outperform the monolinguals in their production of the lateral. The bilinguals performed better in their productions of /æ/ in free speech than in read speech, while the monolinguals perform the same in both tasks. The monolinguals outperform the bilinguals in their productions of /æ/ in read but not in free speech. There is no GROUP-by-FEATURE effect in the free speech task.

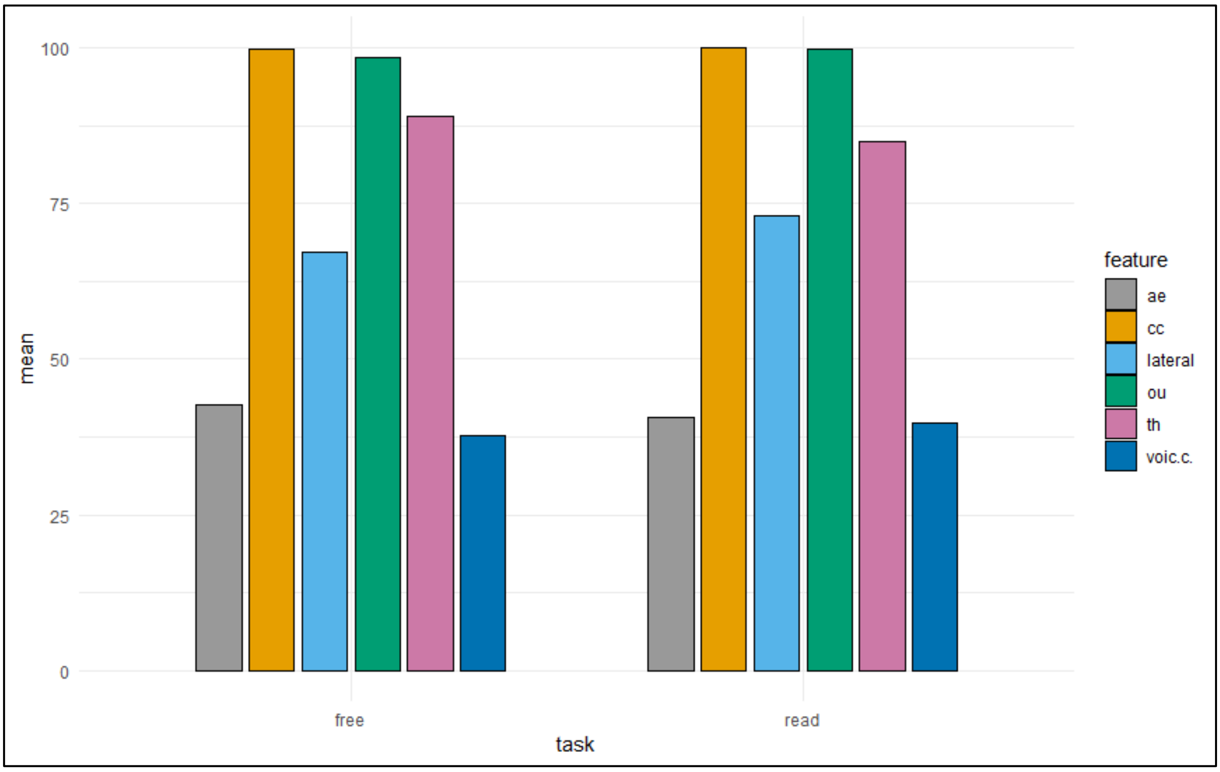

Figure 27: Monolinguals' production accuracy across tasks

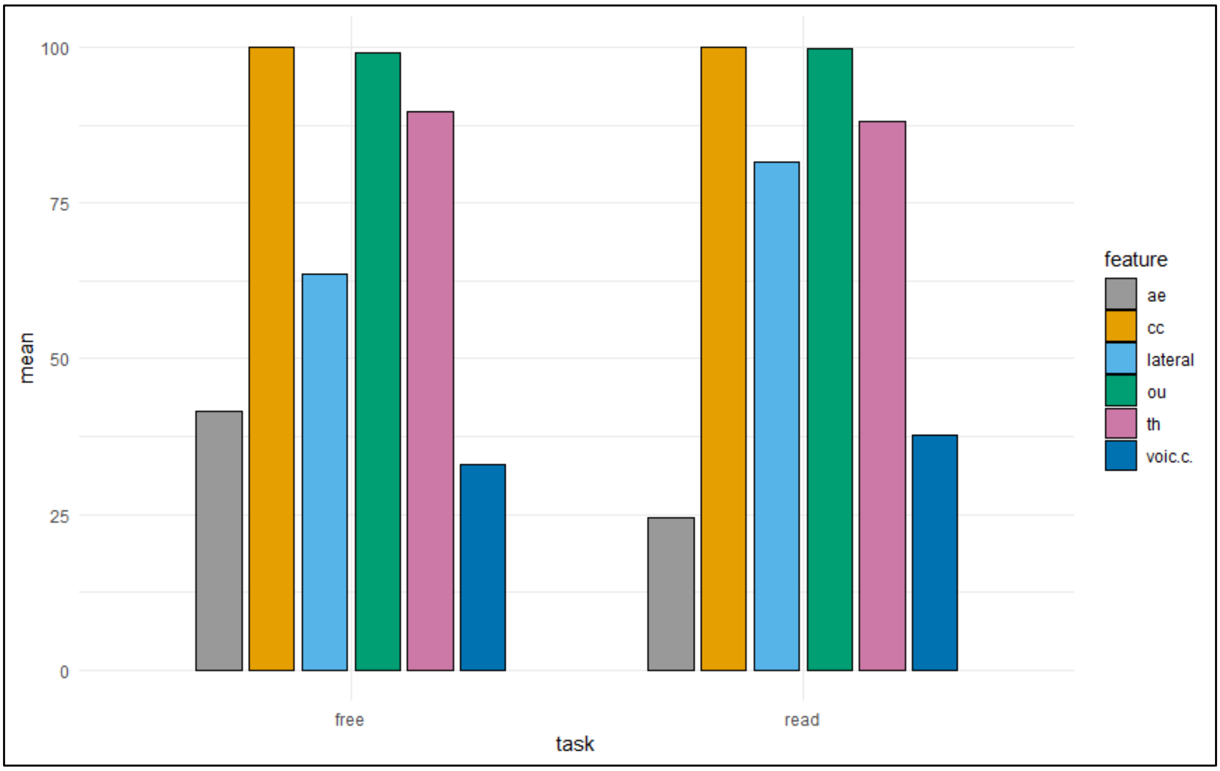

Figure 28: Bilinguals' production accuracy across tasks

5.2.2.3 Perception and production: on the interface

In Study 1, no correlation was found between the perception and production of the participants' L3 English. To see whether an effect actually emerges in the older, more proficient speakers, a potential correlation, both global and for individual features, was also tested for in this study. The results of the Pearson correlations can be seen in Table 42.

Table 42: Correlations perception:production

feature	cor	t	df	p
/æ/	0.26	1.24	22	0.23
CC[74]	-			
lateral	-0.17 (-0.45)	-0.76 (-2.37)	20 (22)	0.46 (0.03)
/oʊ/	0.27	1.3	22	0.21
/θ/	0.26	1.3	22	0.21
voicing	0.06	0.29	22	0.77
vowel length	0.12	0.58	22	0.57
global	-0.07	-0.94	166	0.35

For the laterals using the full dataset, a significant negative correlation was detected (see numbers in brackets in Table 42), indicating an inverse relationship of the perception and production of the lateral. However, in a visual inspection of the scatterplot for this correlation (see Figure 29), two outliers were identified that skewed the otherwise widely scattered data.

[74] A correlation coefficient could not be calculated for CC since production was at 100% in all speakers.

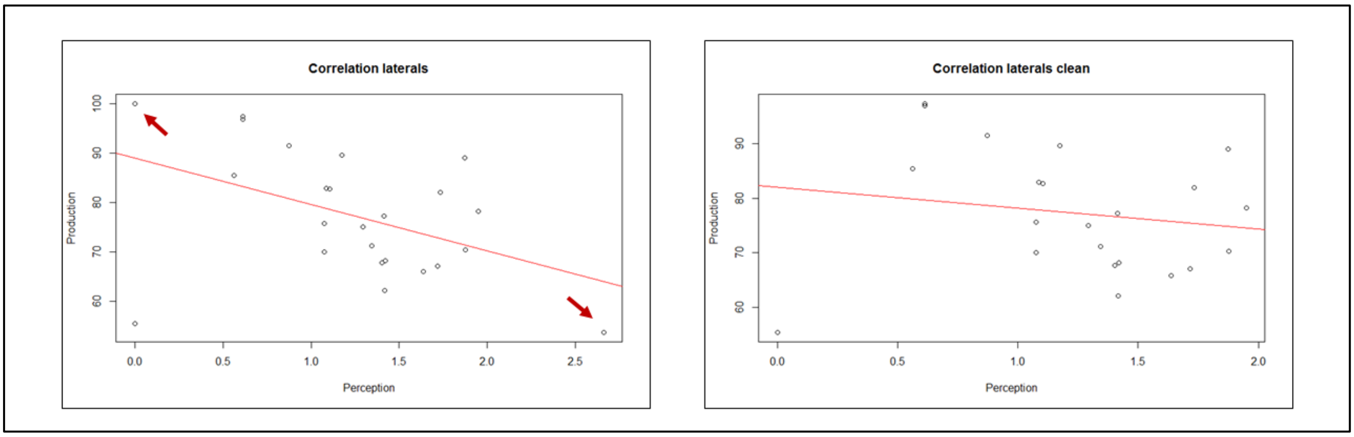

Figure 29: Perception-production correlation of laterals with (left) and without (right) outliers

After removing these two outliers from the calculations, the current numbers (i.e., the numbers in Table 42 that are not in brackets) were obtained for the lateral correlation. The subsequent computations excluding the outliers revealed no significant correlations between perception and production in any of the features or the full dataset.

5.2.2.4 Further analyses concerning laterals

As the global analysis of the data did not distinguish the perception or production of the lateral in different phonological contexts, i.e., in positions that are expected to trigger different allophones, additional analyses were performed to reveal whether context plays a role in variable lateral perception and production. To that end, linear mixed-effects models were fitted with GROUP, CONTEXT, AGE, and GERMAN PROFICIENCY as predictor variables, and SPEAKER as a random factor. Table 43 shows the final model for perception.

Table 43: Linear mixed-effects model for discrimination of laterals by context

<table>
<tr><td colspan="6">Random effects:</td></tr>
<tr><td>Groups</td><td colspan="2">Name</td><td colspan="2">Variance</td><td>Std. Dev.</td></tr>
<tr><td>Speaker</td><td colspan="2">(Intercept)</td><td colspan="2">0.1222</td><td>0.3496</td></tr>
<tr><td colspan="6">Number of obs: 48, Groups: Speaker, 24</td></tr>
<tr><td colspan="6">Fixed effects:</td></tr>
<tr><td></td><td>Estimate</td><td>Std.Error</td><td>df</td><td>t value</td><td>Pr(>|t|)</td></tr>
<tr><td>(Intercept)</td><td>1.4973</td><td>0.1579</td><td>46.0759</td><td>9.484</td><td><2.1e-12***</td></tr>
<tr><td>Context (ini)</td><td>-0.5014</td><td>0.1991</td><td>24.0000</td><td>-2.518</td><td>0.0189*</td></tr>
</table>

Pairwise model comparisons revealed a significant main effect of Context, but not of any of the other variables. Post-hoc comparisons show that the lateral contrast was perceived significantly better in final than in initial position (p=0.0209, df=25.04, t=2.465). There are no significant interactions between any of the variables. However, Figure 30 reveals another trend: While the lateral in final position is perceived with a very similar degree of accuracy across groups, the monolinguals perform better than the bilinguals at discriminating the laterals in initial position. In fact, while the monolinguals have a similar rate of success in both positions, the bilinguals have a harder time discriminating the two lateral allophones in initial than in final position, which can be seen in the downward scattering of the results in the bilingual data, in spite of a relatively similar mean across groups for the perception in initial position.

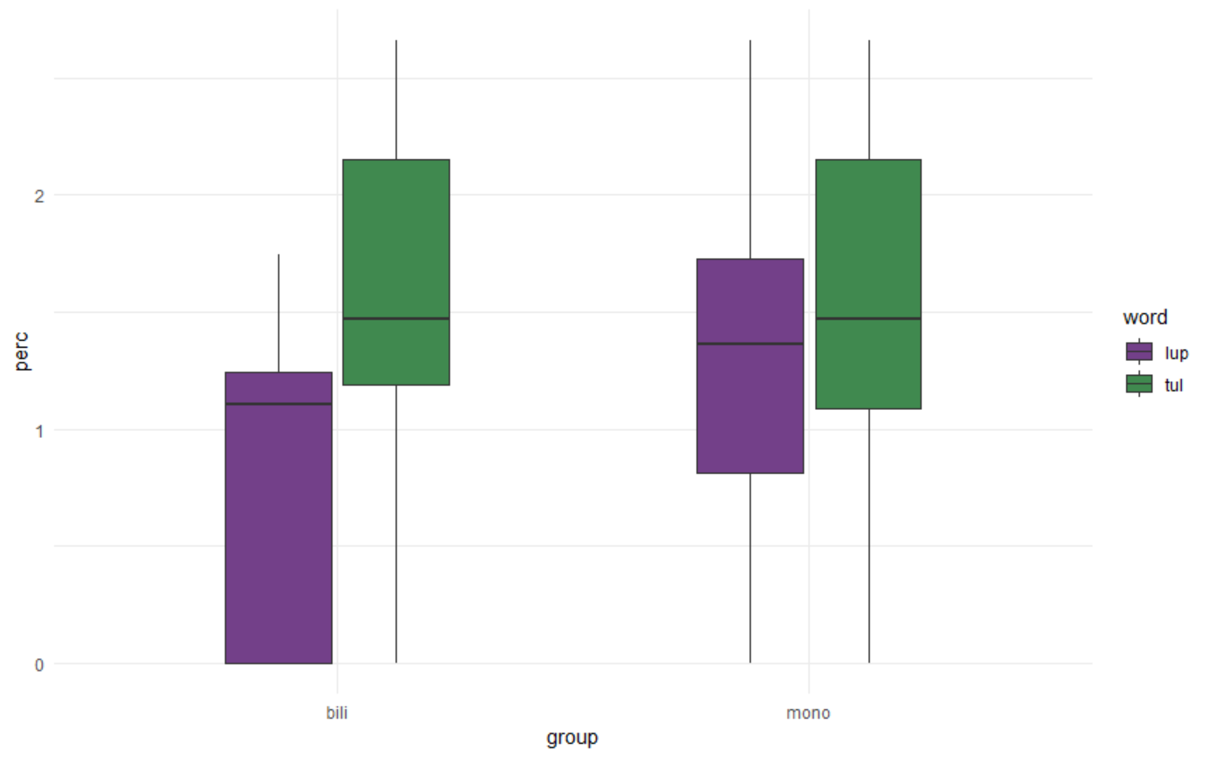

Figure 30: Discrimination accuracy of laterals by context

Equivalent models were then fitted for the production of laterals. The reading task, rather than the free speech task, was used for the primary analysis, which ensured a more controlled dataset. Just as in the previous analysis, the model contained GROUP, CONTEXT, AGE, and GERMAN PROFICIENCY as predictor variables, and SPEAKER as a random factor. The final model (see Table 44) contained GROUP and CONTEXT as significant predictors of target-like production, as well as a significant interaction between GROUP and CONTEXT.

Table 44: Linear mixed-effects model for target-like production of laterals by context

Random effects:			
Groups	Name	Variance	Std. Dev.
Speaker	(Intercept)	132.4	11.51
Word	(Intercept)	115.2	11.51
Number of obs: 5290, Groups: Speaker, 24; Word, 170			

Fixed effects:					
	Estimate	Std.Error	df	t value	Pr(>|t|)
(Intercept)	58.861	3.811	37.475	15.444	<2e-16***
Group (bili)	20.791	5.023	28.474	4.139	0.000282***
Context (ini)	20.843	2.308	300.568	9.030	<2e-16***
Group (bili):Context (ini)	-18.771	2.193	5094.725	-8.559	<2e-16***

As can be seen in Figure 31, the lateral was generally produced more target-like in initial than in final position (p<.0001, df=inf, z=-5.654). This difference is, however, only significant in the monolingual group, who do indeed a reach higher accuracy rate in initial than in final position (p<.0001, df=inf, z=-9.03), while there is no difference in the rate of accurate productions between the two contexts in the bilingual group (p=0.3697, df=inf, z=-0.897). This difference is connected to the bilinguals' performance in the production of the lateral in final position, where they reached significantly higher accuracy rates than the monolinguals (p=0.0002, df=inf, z=-3.781), which levels out the differences found in the monolingual group.

The same analysis was also performed on the free speech task, and the results did not differ from those obtained in the reading task: There is a significant main effect of CONTEXT as well as a significant interaction of GROUP and CONTEXT. Both groups perform equally well on the laterals in initial context (p=0.1, df=308.06, t=1.637), while the bilinguals outperform the monolinguals in final laterals (p<.0001, df=464.23, t=-5.85).

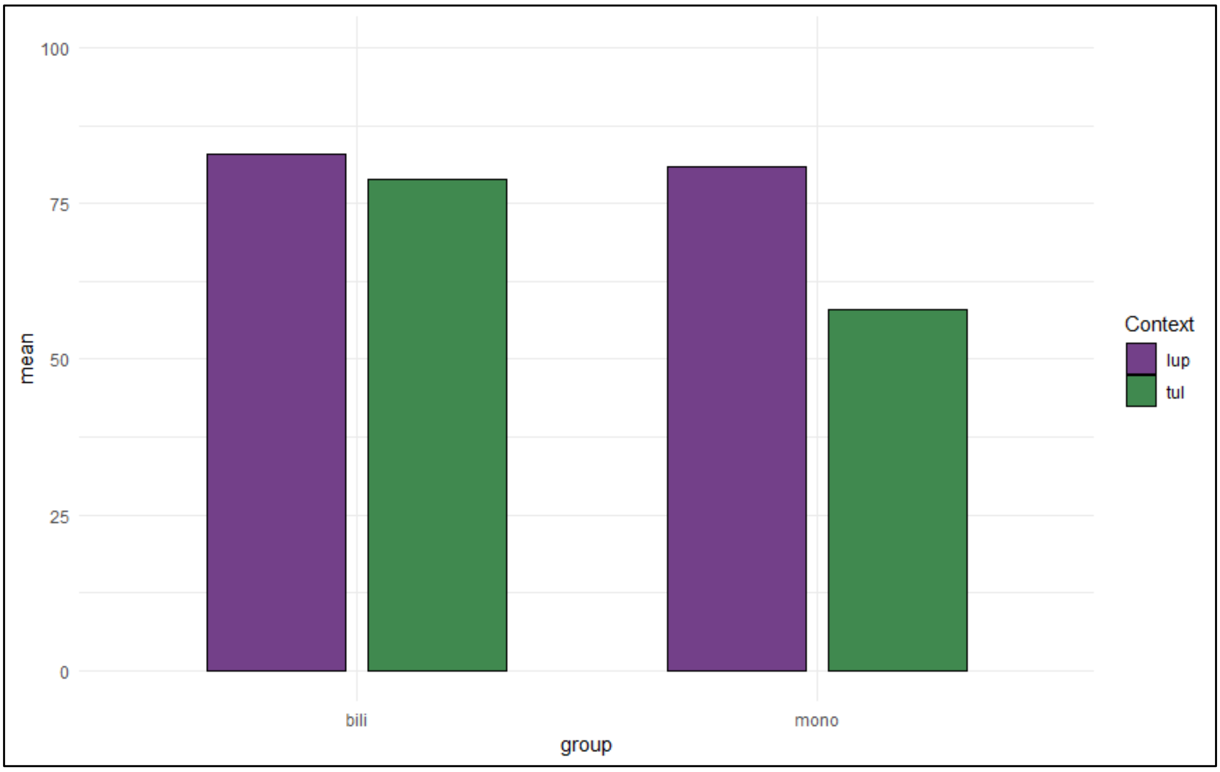

Figure 31: Production accuracy of laterals by context

5.2.2.5 Further analyses of /æ/

In order to assess the concrete productions of the vowel /æ/ across groups, F1, F2, and F3 of the first four instances of the vowel in the reading task were measured for all participants. Additionally, all instances of the Turkish vowel /e/ were measured in the bilinguals' Turkish reading text. Two simple linear models were fitted with GROUP as the potential predictor and with no random factors. Dependent variables were Z3-Z1 (the normalized equivalent of F1, i.e., vowel height) and Z3-Z2 (the normalized equivalent of F2, i.e., vowel backness) respectively.

Table 45: Linear regression model for vowel height by group

Coefficients: (Z3-Z1)				
	Estimate	Std.Error	t value	Pr(>|t|)
(Intercept)	8.606	0.184	46.73	<2e-16***
Group (mono)	-0.215	0.269	-0.8	0.427

Table 46: Linear regression model for vowel backness by group

Coefficients: (Z3-Z2)				
	Estimate	Std.Error	t value	Pr(>|t|)
(Intercept)	3.183	0.16	19.838	<2e-16***
Group (mono)	-0.223	0.234	-0.954	0.344

Table 45 and Table 46 show that neither height (Z3-Z1) nor backness (Z3-Z2) were significantly predicted by GROUP. This can also be seen in Figure 32, which shows British native speaker values (orange, measured in the test stimuli of the production task in Study 1) and the Turkish vowel /e/ as produced by the speakers in the bilingual group (pink), in addition to the bilinguals' English productions in purple and the monolinguals' English productions in green.

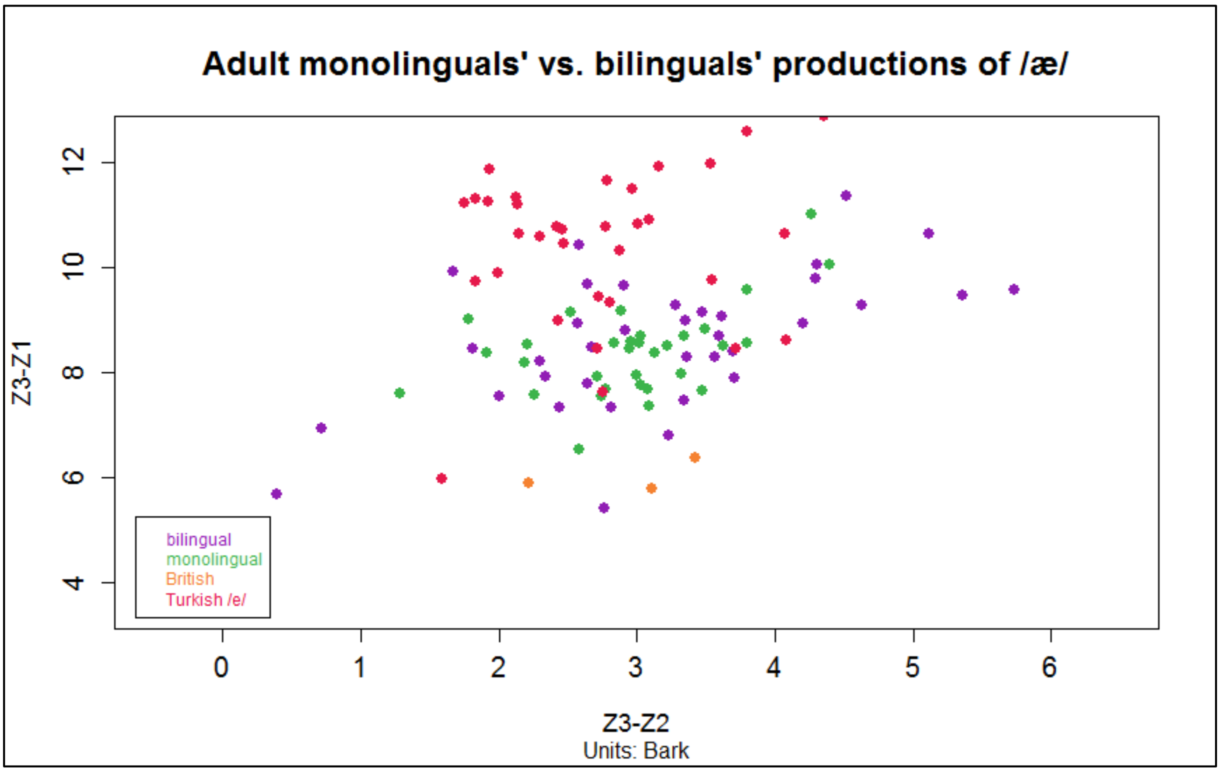

Figure 32: Monolingual and bilingual productions of /æ/

The English values of both groups are higher than the values of the English native speaker target, but they are lower than the native speaker productions in Turkish. The latter is also evidenced in two simple linear models with Z3-Z2 and Z2-Z1 as dependent and LANGUAGE as independent variable.[75] In these two models, the productions of Group 1 and Group 2 were collapsed since, before, no group difference was shown. As can be seen in Table 47 and Table 48, Turkish /e/ is produced significantly higher than English /æ/ by the bilinguals, whereas there is no significant difference in backness.

Table 47: Linear regression model for vowel height in Turkish /e/ vs. English /æ/

Coefficients: (Z3-Z1)				
	Estimate	Std.Error	t value	Pr(>|t|)
(Intercept)	8.606	0.232	37.065	<2e-16***
Language (Tur)	1.847	0.331	-5.586	4.28e-07***

Table 48: Linear regression model for vowel backness in Turkish /e/ vs. English /æ/

Coefficients: (Z3-Z2)				
	Estimate	Std.Error	t value	Pr(>|t|)
(Intercept)	3.183	0.191	16.68	<2e-16***
Language (Tur)	-0.264	0.272	-0.97	0.335

5.2.2.6 Differential substitution of the interdental fricative in perception and production

While the interdental fricative is a novel contrast for both the monolinguals and the bilinguals, the literature on interdental substitutions by language learners has shown differences in substitution patterns depending on the speaker's language background. Study 1 revealed group differences in the production, but not the perception of the interdental fricative by early-

[75] Adding the native English items to the model was not possible, since only three instances were available.

stage learners of English, pointing to an effect of the HL in the bilingual productions (see Chapter 4.3 for details). To determine the effect of SUBSTITUTION on the perception of the interdental fricative by the adult participants, i.e., by the participants in this study, a linear mixed-effects model was fitted, which revealed a main effect of SUBSTITUTION, but no main effect of GROUP and no interaction of SUBSTITUTION and GROUP (see Table 49).

Table 49: Linear mixed-effects model for the perception of /θ/ by substitution

Random effects:					
Groups	Name		Variance		Std. Dev.
Speaker	(Intercept)		0.4236		0.2058
Number of obs: 144, Groups: Speaker, 24					
Fixed effects:					
	Estimate	Std.Error	df	t value	Pr(>|t|)
(Intercept)	2.4172	0.0916	85.8314	26.38	<2e-16***
Substitution (s)	0.7362	0.1152	120.0000	6.393	3.25e-09***
Substitution (t)	0.7991	0.1152	120.0000	6.938	2.16e-10***

Post-hoc comparisons with Bonferroni corrections showed that the /θ/-/f/ contrast was perceived significantly more poorly (d’=2.42) than both /θ/-/s/ (d’=3.15; p<.0001, df=122.03, t=-6.339) and /θ/-/t/ (d’=3.216; p<.0001, df=122.03, t=-6.880). There was no significant difference between the discrimination of /θ/-/s/ and /θ/-/t/ (p=.8512, df=122.3, t=.541). In addition to the lack of an interaction with GROUP, a visual inspection of the perception patterns (see Figure 33) clearly shows the strong convergence of monolinguals and bilinguals in the discrimination of the interdental fricative, which, in turn, matches the pattern found in the younger learners.

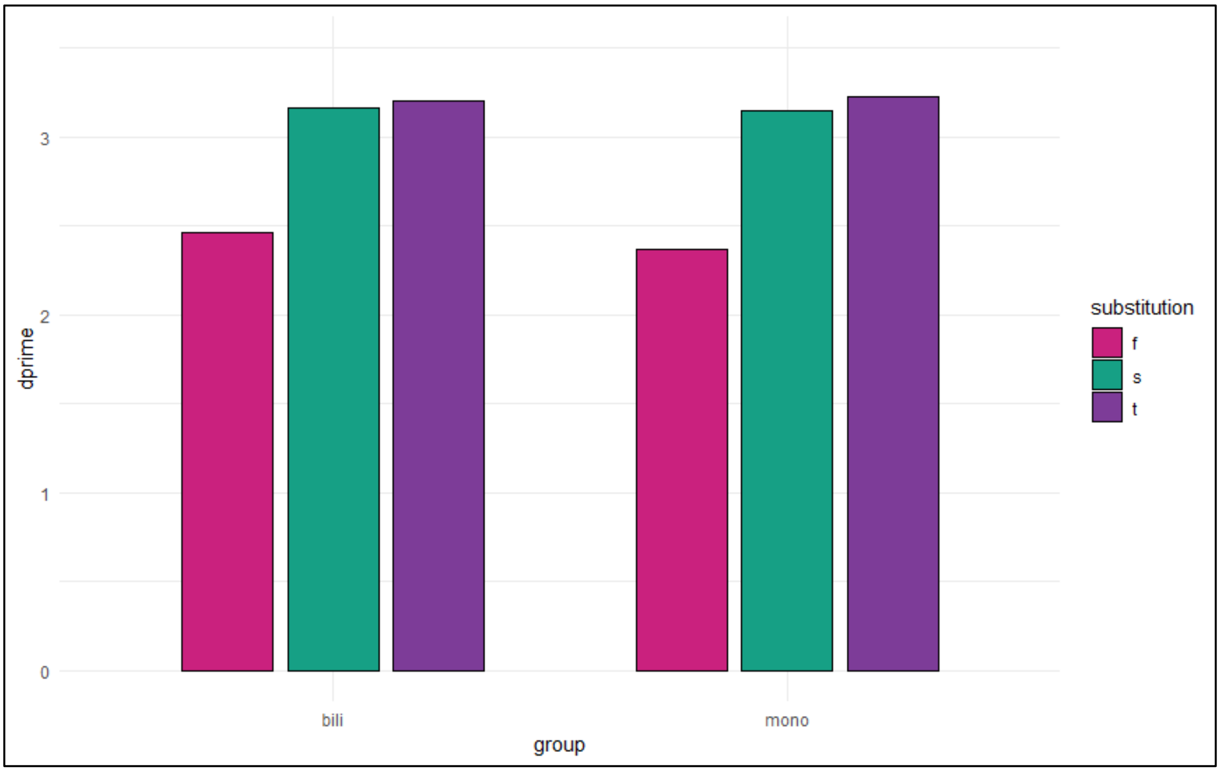

Figure 33: Discrimination accuracy by substitution and group

To examine the substitution patterns in the productions of the adult speakers of L3 English, non-target-like productions of the unvoiced interdental fricative were coded for substitution, categorized as *f, s, t*, or *other,* and counted. Since not enough instances of the unvoiced interdental fricative occurred in the free speech tasks, this was only done for the reading task. When the substitution was not clear from the auditory inspection of the audio recording, the notes taken on the day of the experiment[76] as well as the spectrogram were consulted. Figure 34 shows the relative distribution of interdental substitutions by group. In both groups, [s] is the most common substitution, with 46% in the bilinguals and 58% in the monolinguals. 32% of the bilinguals' substitutions were [t], while this was only found in 10.5% of the cases in the monolinguals (a number that translates into only four items in total). A

[76] While the participant was reading the text, the experimenter listened and watched carefully for interdental substitutions and took notes based on both the auditory and the visual information about the articulation.

further 16% of interdental fricatives in the bilinguals and 10.5% in the monolinguals were substituted with the acoustic default [f], while the remaining 6% in the bilinguals and 21% in the monolinguals were various substitutions (e.g., [ʃ], [ð]). The main difference, then, lies in the stop-substitution, which was more common in the bilingual group and which corresponds to the substitution pattern expected from monolingual L1 speakers of Turkish speaking English as an L2 (see Chapter 3.2.3).

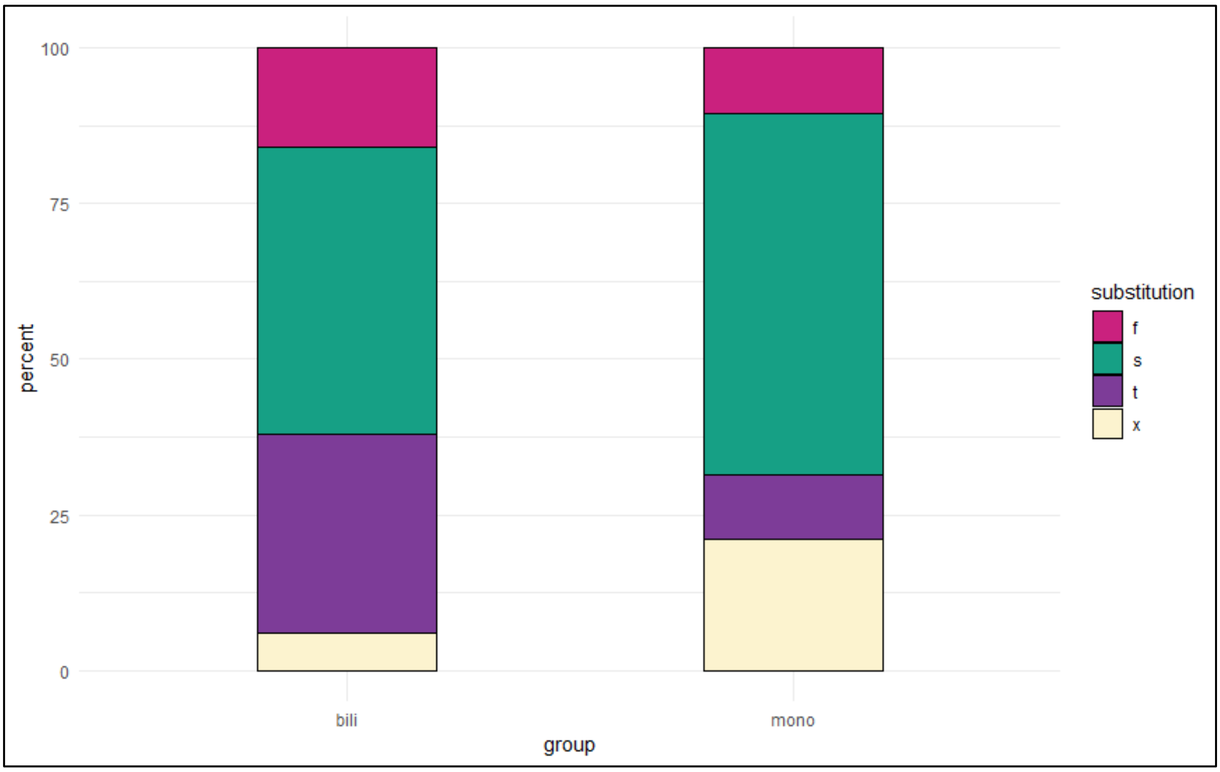

Figure 34: Substitution patterns in production by group

5.2.2.7 Secondary variables

As shown in section 5.2.2, none of the secondary variables (TURKISH PROFICIENCY, GERMAN PROFICIENCY, BLP-SCORE, AGE) had any effect on the monolinguals' or bilinguals' perception or production of any of the features tested. However, when looking at the scores obtained for the use of German and the use of Turkish in the BLP questionnaire separately, as seen in Figure 35, those participants with particularly high scores for German seem to score relatively low on Turkish.

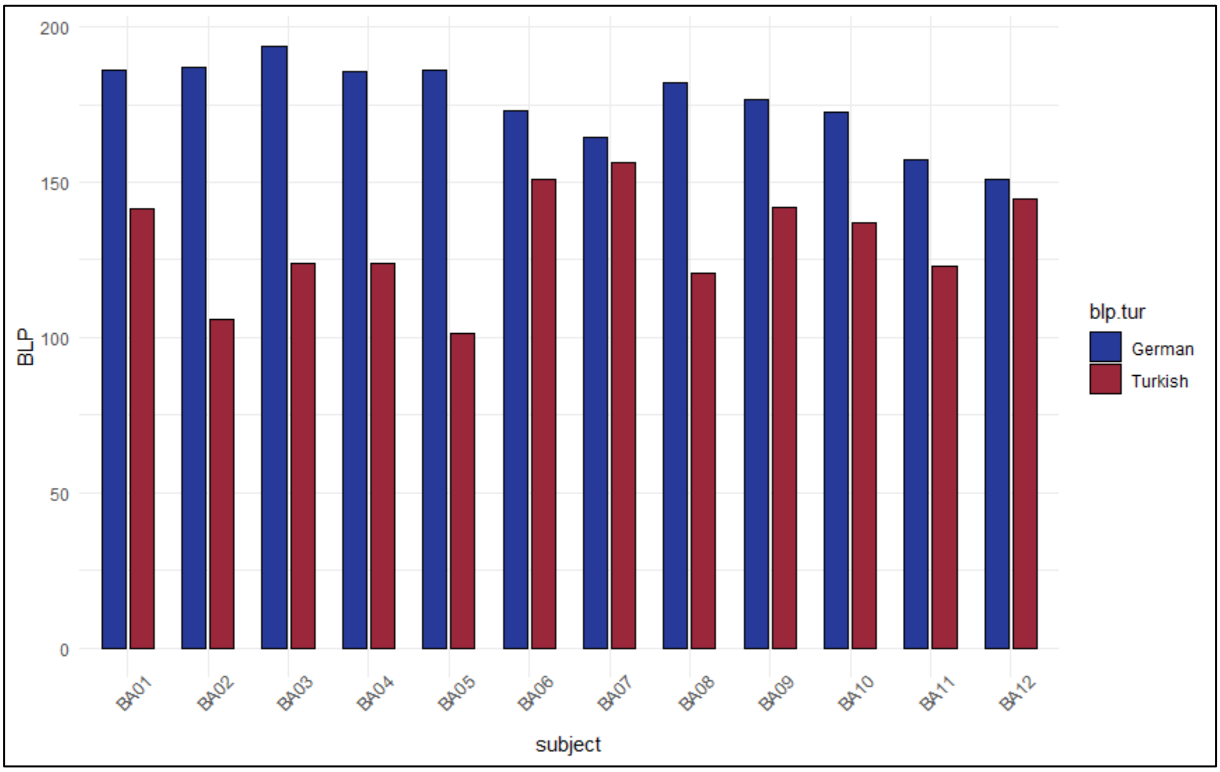

Figure 35: BLP scores for German vs. Turkish by participant

Pearson's correlations did, however, not come out as significant for this relationship (cor=-0.51, t=-1.89, df=10, p=0.09). An estimation of non-linear correlation using the nlcor (Ranja & Najari, 2019) function in R failed due to the small size of the dataset (only 12 data points per variable for each language, one each for the 12 participants).

5.3 Discussion

After examining and attempting to explain the (non-)convergence patterns across two groups (one monolingual and one bilingual) of pre-pubescent learners of the TL English in Chapter 4, similar experiments and analyses were run on two additional groups on non-native speakers of English, this time highly proficient young adults, to uncover how CLI effects would change due to growing proficiency. Here, as in Study 1, one group was monolingual (German), and the other group was bilingual (Turkish-German) young adults. I first use the results presented in Section 5.3.1 of this chapter to answer the research questions posed in Section 5.1.2. After this, each of the bits is examined individually, and their perception and production by the two groups are discussed, both in isolation and relative to the findings of Study 1.

5.3.1 Global discussion

I first asked whether there was a group difference in the perception and the production of the tested target sounds. Indeed, differences across groups were revealed in both perception and production, albeit crucially depending on the respective bit. More specifically, the monolinguals outperformed the bilinguals in the production of the vowel /æ/, whereas bilinguals exhibited a (slightly, but significantly) higher rate of target-like productions for the lateral. This indicates that there was no clear patterning according to feature type (as categorized in Chapter 3.2), which was particularly evident in FT-T: While both groups converged in the perception and production of the voiced final obstruent (whose rate of target-like production was particularly low in both groups), the differences uncovered for the lateral were qualitatively different in the two domains.[77] In the bits from FT-G, participants in both groups exhibited excellent performance, and in FT-T, performance varied depending on the bit in question.

To see whether the results are consistent with any of the scenarios introduced in Chapter 3.3, the predictions made in the scenarios are compared with the perception (Table 50) and production (Table 51) results of this study.[78]

---

[77] The lateral, of course, constitutes a special case since, in the general analyses in this study, both variants and their distributional patterns are collapsed into single scores. The results of the specific lateral analyses are discussed in Section 5.3.2.4.

[78] Just like in the discussion of Study 1, the coding of the results is as follows: If a result matches the prediction from H1:L2, it is marked in bold, if it matches the prediction from H1:HL, it is marked in italics, and if it matches the predictions made by H2, it is underlined. Combinations are possible.

Table 50: Predicted and actual results in the ***perception*** *of individual features*

Feature	SLT:L2	SLT:HL	CBT	Results
CC	**M=B**	M>B	**M=B**	**M=B**
Vowel length	**M=B**	M>B	**M=B**	**M=B**
Voicing	**M=B**	M<B	M<B	**M=B**
Laterals	**M=B**	M<B	M<B	**M=B**
/æ/	**M=B**	**M=B**	**M=B**	***M=B***
/θ/	**M=B**	**M=B**	**M=B**	***M=B***
/oʊ/	**M=B**	**M=B**	**M=B**	***M=B***

Table 51: Predicted and actual results in the ***production*** *of individual features*

Feature	SLT:L2	SLT:HL	CBT	Results
CC	**M=B**	M>B	**M=B**	**M=B**
Vowel length	**M=B**	M>B	**M=B**	**M=B**
Voicing	M=B	**M<B**	M<B	**M=B**
Laterals	M=B	**M<B**	**M<B**	*M<B*
/æ/	M=B	M=B	M=B	M>B
/θ/	**M=B**	**M=B**	**M=B**	***M=B***
/oʊ/	**M=B**	**M=B**	**M=B**	***M=B***

In perception, the results of almost all bits are consistent with SLT:L2, according to which only one of the languages is the source of potential CLI. Here, it is the L2 or societal language German that seems to have been exerting that influence. However, it is essential, particularly for the lateral, to take a closer look at the feature's allophonic distribution and the way monolinguals and bilinguals handled the different allophones, which is done in Section 5.3.2.4. Furthermore, a different picture emerges for the production data: Here, both SLT:L2 and CBT can explain a large amount of the data. The two bits SLT:L2 cannot account for are the vowel /æ/ and the lateral, and the results do not match the predictions of CBT for the voiced final obstruents and the vowel /æ/. However, when taking into account the acoustic structure of the vowel productions for this one feature, different results are obtained: In fact, as has been shown in Section 5.2.2.5, monolinguals and bilinguals pattern into one single group.[79] Thus, when considering these acoustic rather than the impressionistic analyses, the production results for the vowel are consistent with both SLT:L2 and CBT.

It is apparent, however, that, ultimately, no single one of these two scenarios successfully predicts transfer in the bilinguals tested in this study. Keeping in mind that the scenarios are based on existing L3 models, it needs to be noted that this perfectly reflects the results from previous studies (see Chapter 2.2): Just as multiple studies testing multiple models have been obtaining conflicting findings, the results of this multi-feature study cannot be explained by a simplex transfer scenario. Instead, results are consistent with a scenario that assumes bit-by-bit transfer from any or both of the two background languages. Furthermore, transfer is conditioned by language-specific, universal, and extra-linguistic factors. In this study, neither proficiency nor, in the bilinguals, language dominance, significantly predicted perception or production of any of the phonological features, which is in line with the findings of, for instance, Gallardo del Puerto (2007). Furthermore, when compared to the reading task, only minute, and somewhat surprising, differences emerged in the free speech task, which ruled

[79] This discrepancy is most likely the result of the way in which the judges perceive the participants' productions: The judges may use different mapping strategies when deciding whether or not a sound is produced target-like. Specifically, they may use additional cues for their decisions that are not captured by the formant analyses, e.g., overall voice quality, whether or not neighboring segments were produced target-like, prosody, etc.

out a strong task effect as a predictor of successful production. To identify the linguistic and universal factors at play in CLI, each bit is discussed separately in the following section. Results are also compared to the results of Study 1 (Chapter 4).

5.3.2 Discussion of individual features

5.3.2.1 Initial consonant clusters

The participants of Study 2 performed particularly successfully in both the perception and production of initial consonant clusters with almost perfect scores obtained in both domains, and there was no inter-group difference. Notably, while initial consonant clusters are relatively marked and are frequently subject to vowel epenthesis (see Chapter 3.2.1 for more details), German, a background language for both groups, does not have a constraint against them. As vowel epenthesis is a repair strategy used in languages that do not allow consonant clusters rather than part of the phonological grammar of Turkish, the bilinguals can make use of the German in their language system and thus override the potentially detrimental effect from Turkish.

In both studies, monolinguals and bilinguals performed equally well in the perception as well as the production of initial consonant clusters. Since participants already performed at ceiling level at a relatively low degree of proficiency in Study 1, the fact that there was no development from Study 1 to Study 2 is not surprising.

5.3.2.2 Vowel length

The auditory discrimination of long and short vowels, as well as the target-like production of the contrast in this study, was at ceiling level in both groups. While this was also the case for the perception in the young participants of Study 1, the young monolinguals outperformed their bilingual peers in the production of the vowel length contrast, which was explained by negative transfer from Turkish in Chapter 4.3.2.2. The fact that no such difference was observed in the production of the adult speakers could, in theory, be interpreted as a change of transfer source with the influence from the HL diminishing either due to due to a systematic shift in dominance (see Chapter 2.1.3) or due to increasing L3 proficiency. However, the fact that no negative transfer be observed in the adult speakers combined with the ceiling performance points to convergence due to acquisition rather than such a radical shift of transfer source.

5.3.2.3 Coda voicing

Both adult groups in Study 2 performed on par and at high accuracy levels in the perception of voiced vs. voiceless coda obstruents. Similarly, there was no group difference in production accuracy, but the overall amount of target-like productions was below 50%. In Study 1, by contrast, the bilinguals outperformed the monolinguals in the production of voiced codas. Crucially, the monolinguals neutralized the voicing contrast categorically, while the bilinguals produced the target-like voiced counterpart in slightly less than 50% of all instances. Taken together, this indicates progress in the acquisition of the target sound, through which the monolinguals have now “caught up” with their bilingual peers in production. It should furthermore be remembered that the perception results could not be interpreted in light of coda voicing due to the likely confound with vowel duration (see Chapter 4.3.2.3 for details), and it can thus be speculated that both perception and production were subject to this acquisition effect.

5.3.2.4 Laterals

In Study 2, there was no group difference in the perception of laterals in initial vs. final position (i.e., across allophone contexts), but the bilinguals performed significantly better than the monolinguals in the overall production of the contrast. On the other hand, no significant group differences were found in either perception or production in Study 1. It could be argued that this later group difference was caused by positive influence from the bilinguals' HL Turkish considering that the Turkish lateral distribution, while conditioned by different factors than the distribution in RP, surfaces as a subset of the RP distribution. Thus, to untangle potential transfer effects in connection to this allophonic distribution, it was important to examine the two allophones (clear and dark /l/) separately. When discriminating the clear and dark variant of the lateral, the bilinguals in Study 1 reached significantly higher d'-scores in final than in initial position, while there was no significant effect of position on perception in the monolinguals.[80] This effect could not be found in the participants of Study 2, where both groups performed better in final than initial position. This convergence in the adult participants, as opposed to the children, might be caused by the development of economic strategies to reduce cognitive load when faced with perceptual allophone distinction: The two systems feed the learner two competing strategies when it comes to the clear/dark /l/ distinction: On the one hand, they can use their HL system (Turkish) to perceive and produce the sound that does not occur in their L2 German. This, on the other hand, makes necessary an additional process: They have to unlearn, a distributional rule taken from their Turkish along with the concrete phonetics of the sound, and they then have to learn the new distribution in English, or, in other words, they have to modify an existing distributional rule in their language system. The monolinguals use their L1 German system to deal with English laterals. They have one of the variants, the clear variant of /l/, in their L1 system already, and only need to acquire a new allophonic rule

[80] The group differences in Study 1 could be explained by the influence of the Turkish lateral distribution on the mechanisms employed by the bilingual participants to distinguish the two speech sounds (see Chapter 3.2.2 for a detailed discussion).

that allows them to choose the right allophone.[81] Thus, they are more sensitive to differences in those positions where they expect deviations to occur, i.e., in final position. So, while the bilingual kids in the early acquisition stage seem to still make use of both their systems when it comes to this particular allophonic distribution, they resort to their more economical L2 German system when they become more proficient and thus use the same perceptual strategies as their monolingual peers.

In the production part of Study 2, the analyses revealed a main effect of position, with laterals in initial position being produced significantly more target-like than laterals in final position. There was also a significant interaction of position and group: No significant difference was found in the target-like production with regard to position in the bilingual group, whereas the monolinguals exhibited significantly more target-like productions in initial than in final position. Thus, while neither of the groups was 100% target-like in either of the allophones (i.e., the allophonic distribution was not produced in a stable fashion), the monolinguals performed significantly worse when encountering the dark allophone, which they, in contrast to the bilinguals, do not have in their language system.[82] This revealed that the phonetic absence/presence of a speech sound, and with it the (un-)familiarity with the articulatory gestures involved in production, are critical in production but not necessarily in perception.

5.3.2.5 The vowel /æ/

As shown in Study 1, the bilingual children differed from their monolingual peers in the production of the low front vowel /æ/. Specifically, their /æ/ was higher than that of the

---

[81] They also have to acquire the unfamiliar articulatory gesture, but this is not considered a primary factor for speech perception in the framework used here.

[82] Note that the results of Study 1 and Study 2 cannot be compared directly as in the former the laterals were analyzed acoustically, while in the latter, their production was operationalized as +/- target-like.

monolinguals and thus closer to the Turkish equivalent. They seemed to make use of their HL vowel system as an (additional) reference map for the production of the vowel that was unfamiliar to them. This, however, was not the case for the participants of Study 2. While auditory judgments of +/- native-like production did indeed show a group difference for the production of /æ/ (where the monolinguals outperformed the bilinguals), measurements and statistical tests revealed that the bilingual adults differed from the monolingual adults in neither height nor backness. This discrepancy might be based in a methodological issue: In the auditory judgments, judges may have listened not only for vowel quality, or even only for the vowel, but instead (unconsciously) have taken into account other parameters of the acoustics of /æ/, or the accuracy of surrounding segments. In the acoustic analyses, the bilinguals' productions of Turkish /e/, the closest equivalent available on which to map the target sound, are significantly higher than their productions of the English target sound /æ/. This is evidence for a diminished influence of the bilinguals' HL Turkish. It needs to be noted, however, that, even though the adult HSs produced the vowel lower than their young counterparts and thus closer to English /æ/, neither of the groups achieved native-like performance.[83] In summary, the young bilingual learners' vowel production exhibited a clear influence from Turkish, whereas this influence was diminished in the highly proficient adult bilinguals.

5.3.2.6 The voiceless interdental fricative

Just like in Study 1, there was neither a main effect of group nor an interaction between group and substitution in the perception of the interdental fricative in Study 2. This means that

[83] It is not possible to determine beyond any doubt whether the bilingual speakers approach the target norm more closely irrespective of their German vowel system or whether they are approaching the target norm and, "on the way there" happen to be in a similar area as the monolinguals. However, in both cases, the influence of the Turkish vowel system on the production of the L3 target is reduced.

both the monolinguals and the bilinguals discriminated the contrasts equally well. In fact, d'-scores in both groups were high, i.e., discrimination was excellent. There was a main effect of substitution in the full dataset, and pairwise comparisons revealed that the perception of the /f/-/θ/ contrast was discriminated less accurately than both /s/-/θ/ and /t/-/θ/ by both groups. These differences in perceptual ability by substitution occurred due to the acoustic similarity of the labial and the interdental fricative.

Both monolinguals and bilinguals were very successful in the production of interdental fricatives, and no group differences were found. While the participants of Study 1 also performed on a par in the production of this sound, their performance was overall slightly weaker. This difference across age groups is best explained by the successful acquisition of the target sound. Substitution patterns in Study 2 differed across groups. The most striking difference was the relatively large number of [t]-substitutions used by the bilinguals as opposed to the monolinguals, which revealed an influence of Turkish in the HSs, a trend which could also be observed in the young learners of Study 1. This influence was not exclusive, as evidenced in the use of other substitutions, including [s]. In the production of the interdental fricative, then, both Turkish and German exerted an influence, which, however, did not surface as a difference in (un-)successful production of the target segment but instead as a different substitution strategy in cases where target-like production was not achieved.

5.3.2.7 The diphthong /oʊ/

Bilinguals and monolinguals performed on a par in both the perception and production of the diphthong /oʊ/, achieving results at ceiling level. The group difference in the perception of the same segment by the participants of Study 1, where the bilinguals reached significantly lower d'-scores than the monolinguals, was thus evened out in the older participants. This was most likely the result of successful acquisition, during which the bilinguals could compensate for their initial disadvantage caused by their Turkish.

5.3.2.8 Perception and production: the interface

The correlation tests confirmed the results found in Study 1: There once more was no significant correlation between perception and production for any of the bits. Only in the lateral were perception and production significantly correlated, which may, however, have been due only to two outliers. Even if one does accept the existence of a significant correlation, it is necessary to remember that the analysis collapses the performance in both positions, which makes the results less meaningful. All in all, there was no clear indication of a direct correlation between perception and production in either of the studies. This, along with the results found by other studies that have tested the perception-production link (see Chapter 4.3.2.8) suggests a higher complexity of the perception-production interface. As already proposed in Chapter 4.3, I assume that perception and production ability are not organized parallelly but rather hierarchically. More specifically, I assume that perceptual (discrimination) accuracy is only one of the factors that, combined with and triggered by others, predicts target-like production. This connection, as well as the parameters that condition the activation of perceptual accuracy as a predictor for production, and the other factors at play in the perception and production of an L3 phonology are summarized and discussed in Chapter 6.

6 Re-framing L3 phonology

6.1 A shift of perspective

The analyses conducted as part of this dissertation have shown the following for the set of bits that were tested: The two groups, Turkish-German bilinguals and German monolinguals, behaved differently in their perception and in their production of English phonological bits (i.e., phonological entities, structures, and phenomena), albeit to different degrees. Furthermore, both positive and negative influence from the background languages were observed. As such, the results were not consistent with either of the hypotheses stated in Chapter 3.3: The findings can neither be explained by single language transfer from either one of the background languages (SLT, see Chapters 4.3.1 and 5.3.1) nor by combined beneficial transfer (CBT). Instead, both languages have been shown to be active in the multilingual speakers and exerted an influence on the perception and production of the TL system.

Table 52 shows the perception results of both studies for each bit as well as their respective feature type (FT, see Chapter 3.2). Predictions are based on the initial grouping into FTs, and specifically on the presence and absence of the bits in the background languages. The last column gives the most likely explanation for group differences (or the lack thereof). Note that full accounts of these explanations can be found in the discussion sections of Chapters 4 and 5. The overall results suggest that CLI and the interactions of the background languages had different consequences for different bits, many of which could only be explained by invoking characteristics that are specific to their phonological make-up. Similar effects were observed in the production of the phonological bits, an overview of which can be found in Table 53.

Table 52: Results for L3 perception

Bits	FT	Predictions	Study 1	Study 2	Comments
CC	FT-G	M>B	M=B (good)	M=B (strong)	The tested contrast reflects a repair strategy that is used by speakers of languages in which initial clusters are illicit. It is not part of the phonological grammar. The repaired version of the bit is not expected to be difficult to discriminate from the target-like structure, especially since one of the background languages allows it in its native phonology.
vowel length	FT-G	M>B	M>B	M=B (strong)	The young learners exhibit the expected behaviour in the discrimination of vowel length. It is sufficiently acquired by the adult participants.
lateral	FT-T	M<B	M=B (weak)	M=B (medium)	Since the distribution of the lateral allophones in Turkish is different from that in English, the bilinguals do not benefit from their HL in the overall perception of the lateral distinction.
voicing	FT-T	M<B	M>B	M=B (strong)	The stimuli were not controlled for vowel duration. Duration is likely to have been the voicing cue used by the monolinguals, resulting in better performance by the monolinguals than by the bilinguals.
/oʊ/	FT-E	M=B	M>B	M=B (strong)	The absence of diphthongs in general is detrimental to successful discrimination rather than the absence of the specific segment.
/θ/	FT-E	M=B	M=B (strong)	M=B (strong)	Both groups perform equally well, and their discrimination of the interdental fricative and the three potential substitutions can be explained by the acoustic characteristics of the segments.
/æ/	FT-E	M=B	M>B (strong)	M=B (strong)	The difference in perception accuracy can be explained by the different vowel spaces in Turkish, German, and English.

Table 53: Results for L3 production

Bits	FT	Predictions	Study 1	Study 2	Comments
CC	FT-G	M>B	M=B (strong)	M=B (strong)	The articulatory gestures involved in producing word-initial consonant clusters is very frequent in German and corresponds closely to the English counterpart. This is why the bilinguals can rely on their German to produce the structure faithfully.
vowel length	FT-G	M>B	M=B (strong)	M=B (strong)	Previous studies have shown that the durational contrast in vowels is acquired with relative ease (as opposed to durational contrasts in consonants, for instance). This is why the bilingual participants can make use of their L2 German when producing the contrast.
lateral	FT-T	M<B	M=B (medium)	M<B	In the early acquisition stage, the bilinguals do not have an advantage due to their Turkish, since they have to acquire an allophonic distribution that is different from the one in their HL system. The bilingual adults, having acquired the distributional rules, can then benefit from the corresponding gesture in their language system.
voicing	FT-T	M<B	M<B	M=B (medium)	As predicted, the young bilinguals outperform their monolingual peers.
/oʊ/	FT-E	M=B	M=B (strong)	M=B (strong)	Contextual cues aided the bilinguals in producing the diphthong at the same level as their monolingual peers.
/θ/	FT-E	M=B	M=B	M=B	The interdental fricative was produced with the same level of accuracy by both groups. However, both bilingual groups exhibited different substitutions than their monolingual counterparts.
/æ/	FT-E	M=B	M=B (weak)	M>B	The differences in the adult speakers is likely related to the additional vowel system in the bilinguals.

Group differences were not only found on the dimension of target-like production and successful perception, but furthermore in concrete phonetic differences between groups for some of the bits. It could be shown, for instance, that the production of the vowel /æ/ was affected by both background languages, but the influence of the HL Turkish seemed to decrease in the adult participants of Study 2 as opposed to the young participants of Study 1. The substitution patterns for the interdental fricative also revealed a mixed influence of both background languages in the bilinguals, which was, however, the same for both the adult and the young groups.

Furthermore, there was no one-to-one correspondence or straightforward correlation of the results in perception and production of any of the bits in either of the studies. From the perspective of *perception leads production* (as assumed, for instance, by Flege, 1991, see Chapter 2.1.1), this is counterintuitive at first glance. However, there is more to speech production than speech perception, and merely being able to discriminate does not guarantee accurate production (see Chapter4.3.2.8). Instead, perception is only one of a multitude of aspects that affect the (successful) production of a speech sound.

When comparing the results of Study 1 and Study 2, no systematic differences were uncovered that would indicate a radical shift of transfer source or mechanism. Instead, the participants of Study 2 exhibited an overall progress in their acquisition of the target structures, which is evidenced by a levelling out of between-group differences. For instance, the monolingual learners had drawn level with their bilingual peers in their successful perception of the voiced final obstruent, and the bilinguals in Study 2 had caught up with the monolinguals in the perception of phonemic vowel length and thus cancelled out the group difference found in Study 1.

One of the main assumptions of each of the models that the hypotheses (see Chapters 2.2.1 and 3.3) are based on is uniform transfer. This transfer is expected to pattern either according to the status of the background languages as units or based on the presence or absence of features in the background languages. More specifically, in the acquisition of an L3, the source of transfer is either assumed to be one single language onto the target language or instead both languages, but crucially conditioned by flat parameters of whether or not a feature in the background language aids the acquisition of the TL feature. However, neither of the two hypotheses presented in Chapter 3.3 could account for all the findings of the two studies conducted in the course of this dissertation. It is necessary, then, to assume transfer to become observable on a bit-by-bit level. In other words, the perception and production of different phonological bits are not uniform – they are conditioned by factors that are specific to the bit in question. This is what adaptions of the more recent models (SLM and LPM, see Chapter 2.2.6) have attempted to account for by allowing for transfer from both (i.e., all) languages that can be positive or negative and that is potentially conditioned by a variety of factors. However, by allowing for all these options, these models fail to predict anything and ultimately cannot be falsified.

The core of the problem is two-fold. Firstly, it lies specifically in the subject matter under scrutiny: It has not been useful to frame phonological CLI in terms of the binary oppositions that form the basis for the model predictions since transfer behaviour cannot be predicted by the presence or absence of a feature in one or both of the background languages. This has been shown in previous studies on single phonological features which have continuously yielded conflicting results. The present studies confirmed these conflicting results while additionally strengthening the argument by eliminating the confounding factors of speaker- and group-internal variables by testing multiple bits in the same speaker groups. Since phonological structures cannot readily be described as discrete, coherent, one-dimensional units, the particular binary opposition of presence or absence of a phonological bit is even less meaningful in this context. Instead, a comprehensive description or analysis of a phonological bit will necessarily have to include phonetic and mental attributes, such as the acoustic structures of sounds and their articulatory characteristics, as well as their function(s) in the respective languages and their mental representation in the speakers. To sum up, there is more to the make-up of a phonological bit both underlyingly and on the surface. This issue is further elaborated in Section 6.2. In fact, the need to invoke the term "bit" as an umbrella term for phonological entities, structures, and phenomena that exhibit effects of CLI, and, which more specifically, can be the target of transfer, reveals the particular complexity of phonological grammars.

Secondly, research in L3 phonology has, so far, mostly dealt with identifying the most likely source of transfer out of all the background languages of the learner of an additional language. However, it has become clear that all background language systems are generally available to and active in the learner. This specific finding needs to become one of the assumptions for further research: The languages in the multilingual system all interact, they are all predictors in CLI. Thus, the search for the source of transfer and the reason for the choice of the particular source becomes meaningless, which makes it necessary to re-frame the initial question. Instead of asking whether a phonological structure is transferred from a particular language and why this transfer surfaces, the aim should be to understand the workings of the interactions that we already know to be there. These two outcomes lead to the following desiderata:

(1) The specific, particularly complex and interlaced make-up of phonological grammars, one of the core reasons for the conflicting results in L3 phonology, has to be acknowledged and factored in when studying an L3 phonological acquisition.

(2) The interaction of all languages in the multilingual mind have to be a basic assumption rather than the subject of study.

This triggers a shift of focus: It cannot be informative to look at L3 phonology in isolation. In fact, the same phenomena and mechanisms that have been observed in L3 phonological acquisition are also found in research on synchronic variation as well as language change. In other words, L3 phonological acquisition is not a discrete phenomenon in its own right with a specific set of rules and patterns. Instead, it is one of a number of domains that can equally inform us about how phonological systems behave. Language change, synchronic variation, language acquisition (L1, L2, and L3) – all those are different parts of the same elephant. Each of them can be studied in its own right, but always keeping in mind what we know about the other parts. Only by doing that can we learn about the animal *phonology* as a whole in all its facets, which should ultimately be the aim of research in L3 phonology.

6.2 Interlaced complex systems in a bit-by-bit frame

The previous section has argued that instead of focussing on L3 acquisition and searching for the most likely source of transfer onto an L3 phonology based on cross-linguistic comparisons, it is necessary to take a step back and re-focus on the actual subject matter in question, which is the structure of phonological systems and their emergence in domains, such as language acquisition. Even when only considering a single snapshot that merely displays a static picture of the system, a phonological grammar is highly complex. It consists of entities on multiple levels such as discrete segments or prosodic structures. These entities, in turn, have multiple characteristics themselves. Depending on the theoretical framework as well as the specific focus, segments can, for instance, be described using distinctive features, acoustic properties, or the articulatory gestures necessary to produce a sound. Their make-up furthermore includes distributional patterns, both structural (where is a certain structure most likely to occur? what are its context-dependent surface forms?) and functional (which contrasts are expressed using the structure in question? how high is its functional load? how frequent is it overall in the system?). Furthermore, physical and distributional properties do not necessarily reflect the way in which phonological segments are represented in the speaker's mind, and mental representations of two similar surface forms can differ substantially, for instance in terms of mapping strategies or the use of phonological cues. In addition, there is an extralinguistic aspect insofar as phonological structures can also be used as socio-linguistic markers on various social dimensions. This simplified static account is not supposed to be a comprehensive compilation of the structural characteristics of phonological systems, but is instead meant to illustrate the complexity of phonological grammars.

However, phonological systems are not, in fact, static but dynamic in nature. Their elements do not exist in isolation, but in relation to and interaction with each other. This dynamism concerns both the individual bits and larger units (up to the whole grammar). Vowels in a vowel system, for instance, will “react” to the shift of other vowels in the same system potentially triggering a chain shift that ultimately aims at keeping the individual items distinct.[84] Interactions are also possible between entities that are not part of the same core unit. For instance, tone distinctions have been argued to emerge from voicing of onset consonants via a different voice quality (Bybee, 2015, p. 64). The phonological grammar of only one individual language as a whole can thus be considered to be a complex system. When more than one grammar interact in situations such as language acquisition, multilingualism, or language contact, multiple of these complex systems influence each other. The degree to which this happens must depend on the stability of the system. It can be assumed that a more stable system is subject to more subtle effects than a less stable one since the categories and the boundaries between the entities are more robust in the former. However, the general principle of mutual influence and of interaction remains the same.

[84] This is a blatantly simplified account of the process, which is, however, sufficient for the argument posed here.

The idea of phonology as a complex system is not a new idea by far. In fact, the notion of language as a self-organizing complex system emerged as early as the 1980s in a paper by Lindblom and Maddieson (1988), only slightly later than the general notion of complex systems (Coupé, Marsico, & Pellegrino, 2017). In fact, the theory of Complex Dynamic Systems has been proposed for L3 phonology before by Kopečková (2019). In her approach, however, she focuses on phonological acquisition as her core research subject: She models the development of language interactions as evidenced by fluctuating transfer behaviour with the language systems as the main agents. She assumes that the language system of a multilingual incorporates various subsystems (i.e., the individual languages) and that, in the acquisition process of a new language, all subsystems of the larger system interact witch each other. By examining shifting accuracy rates in the segments under scrutiny (/v/ vs. /w/), she aims at answering how adding an additional language to an existing language system influences the languages already in the system.

The core assumption of languages as complex systems and Kopečková's (2019) adaptation of it to L3 research is plausible. However, only one of the two necessary shifts developed in the previous section has been realized by her. As I have argued above, the approach of aiming to learn about L3 phonology by looking through the L3 "lens" is unlikely to reveal coherent patterns, which has, so far, been the case in that line of L3 research. Crucially, it is necessary to instead view the findings from the studies of this dissertation, as well as the findings from other studies in L3 phonology (see Chapter 2.2), in the framework of complex systems while focussing on phonological processes rather than L3 acquisition. The failure to do so explains why such multitude and diversity of explanations was necessary to account for the group differences revealed in the studies of this dissertation: The bits under scrutiny differed immensely, which can be illustrated by comparing the English lateral and English /æ/. Like English, Turkish has two allophones of the lateral which are acoustically and articulatorily close to their English counterparts (though by no means identical). Their distribution pattern is, however, different from that of English. German, on the other hand, does not have the dark allophone of the lateral, which is one variant in English, and instead uses the clear lateral in all positions. The vowel /æ/, on the contrary, does not occur at all in Turkish or in German. However, both languages have equivalents of the sound which are sufficiently close for the incoming L3 sound to be mapped on. Those equivalents differ in both languages (albeit only slightly) and, more crucially, they are part of different vowel systems. Thus, a comparison of the effects that the addition of a new language to an existing system has on such vastly different phonological structures must be difficult at least, and it is not surprising that only few coherent patterns emerge, especially when the structural differences are not considered in the first place. In fact, the comparisons made in the field of L3 phonology cannot be likened to the comparison of apples and pears. A more fitting image would be the comparison between a wasp sitting on an oak on a mountain and a hazelnut growing on a tree in an orchard near the sea. The wasp and the hazelnut have roughly the same size and this may lead to the assumption that they are both subject to the same general forces. If both trees are felled, the wasp can be observed rising into the air while the hazelnut will most likely fall to the ground. Potential culprits of their different behaviour could be air pressure or humidity. However, instead, it should have been the make-up of the objects – the wasp is animate, has wings, and can fly while the hazelnut is inanimate and attached to the tree – that would have led to a more meaningful explanation and thus reproducable results.

To further our knowledge about phonology by using L3 acquisition, I instead propose a framework of complex systems that integrates the bit-by-bit frame introduced in Chapter 2.2.7. Any phonological unit – irrespective of its characteristics, make-up, or level – is considered a bit. The full phonological grammar of a language is a bit, and so are systems of phonological processes (e.g., lenition, neutralization), prosodic structures (e.g., prominence patterns), or segments (e.g., the vowel system), but also the entities that make up those bits, e.g., each individual vowel phoneme. Below the phoneme bit, each phonological feature (e.g., *voicing*, *tenseness*, *anterior*) again constitutes a discrete bit. A simplified schematic illustration of this bit system can be seen in Figure 36.

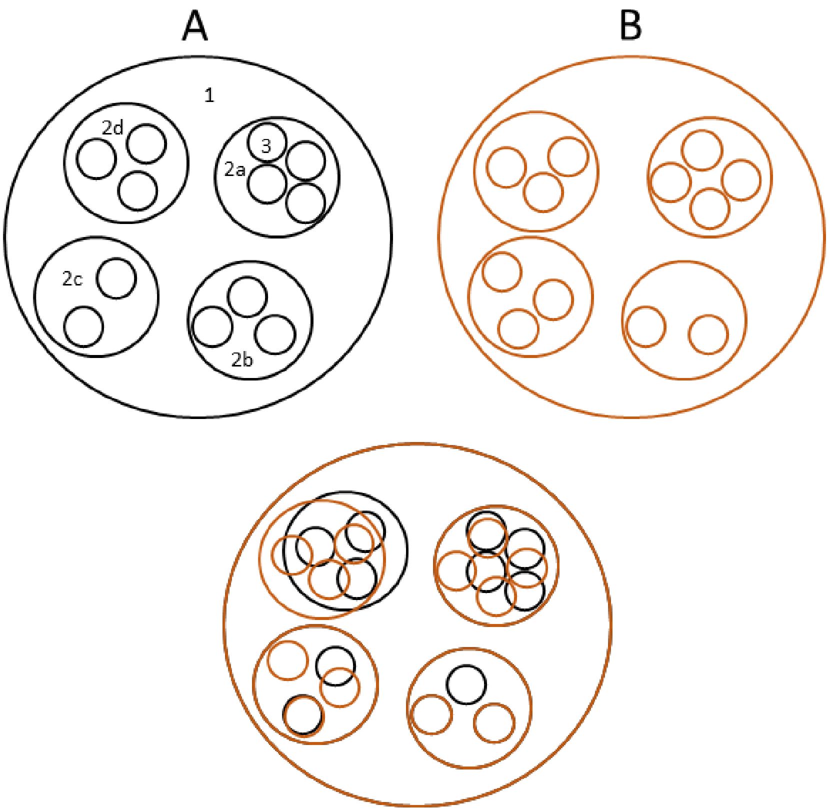

Figure 36: The bit model for multiple languages

Figure 36 depicts a number of hierarchically ordered circles, each of which signifies one discrete bit. Circles of the same size are bits on the same level that are, in turn, contained in higher-level bits and themselves contain lower-level bits. The largest circles in the figure, level-1 bits, stand for the full phonological grammars of two different language systems A and B. In general, each bit can potentially affect every other bit, irrespective of their levels. While the upper part of the figure shows the two separate language systems, the lower part shows a combination of the two systems as it would, for instance, be found in a bilingual. The two systems are relatively similar. On level 2, only one of the bits is in a noticably different position (i.e., the bit differs across languages). All other level-2 bits correspond to each other. However, if the level-3 bits that form the make-up of those on level 2 are taken into consideration, it can be observed that they do, in fact, differ across language systems. Based on the assumption that all bits potentially affect each other, these differences can impact level-1 and level-2 bits. Thus, the interaction of bits whose make-up is different across language systems will affect the outcome in different ways than that of bits whose make-up is the same. These effects can only be predicted reliably if, ideally, the full bit-structure of all the phonological grammars is known and taken into account.[85]

As has been shown above, the bit-by-bit approach can shed light on the interaction of multiple phonological grammars in a speaker's mind by examining L3 learners. More specifically, by studying the output of multilingual systems under the assumption of complex interlaced systems, phonological patterns can be revealed. These patterns, in turn, can inform the researcher about the individual languages under scrutiny – for instance about the language-specific weighting of cues for mental representations of phonological structures – by factoring in what is known about processes of language contact, language acquisition, and language change. In summary, one should not use phonology to understand L3 acquisition, but instead use L3 acquisition as a source of evidence that can inform about phonological processes and language-specific structures.

[85] Depicting the multilingual system as a combination of the two monolingual systems is, of course, a simplification. Instead, it is generally necessary to take into account that the systems change by interacting.

7 Conclusions

In this dissertation, two studies were conducted to investigate CLI effects in Turkish-German bilingual and German monolingual learners of English, both in perception and production. The aim was to test the assumptions of existing L3 models – collapsed into two competing hypotheses – by uncovering differences between monolinguals and bilinguals in their perception and production of TL sounds of three different feature types (FTs): those that are either present in German (but not in Turkish), in Turkish (but not in German) or in neither of the two background languages. This was meant to identify the transfer source as well as the quality of transfer (positive or negative). In addition, extralinguistic variables were tested for their potential effect on the perception and production of the TL structures in either or both of the groups. Finally, potential developmental effects on CLI behaviour were studied employing a quasi-longitudinal design: The participants of Study 1 were young learners of English at the beginning of secondary school, and those of Study 2 were university students who were highly proficient and frequent users of English.

The results of the two studies could confirm neither of the two hypotheses and showed that successful perception and target-like production of TL structures did not pattern according to their FT. However, by comparing bilinguals with monolinguals in the perception and production of English, it could be confirmed that both languages in a bilingual system are available for transfer and that transfer can be positive or negative. Furthermore, no direct relationship could be shown between perception and production, which was taken to suggest that the two are in a hierarchical rather than a mutual relationship. As a consequence, successful perception was considered to constitute a separate entity conditioned by CLI effects and universals as well as one of the parameters that, under certain conditions, exert an influence on production. The developmental analysis revealed fewer group differences in the adult participants of Study 2 than in the young participants in Study 1. These results could be attributed to ongoing acquisition.

Overall, the results could be explained by invoking specific phonological and phonetic characteristics of the target structures as well as the equivalent structures in the background languages. Consequently, explanations were different for each structure. Additional acoustic and pattern analyses revealed fine-grained group differences that could be attributed to the interplay of the background languages.

Based on the theoretical overview and the empirical part, an argument was put forward that encourages a paradigm shift in the study of L3 phonology and entails the following main points:

1. The need for the diverse explanations to account for the results obtained for the different phonological structures tested in this dissertation corresponds to the large variety of differing and often contradictory results yielded by previous studies. By testing the same speaker groups on both the perception and production of multiple phonological structures, potential methodological effects due to different approaches, speaker groups, or language combinations could be ruled out. Thus, the reason for diverging results is likely to be found elsewhere. Furthermore, this same requirement reveals the complexity of the domain, which is not usually taken into account in L3 studies, where phonological structures are mostly treated as one-dimensional, discrete entities that pattern cross-linguistically according to their presence or absence in the languages involved. However, phonological grammars need to be viewed as interlaced systems whose components are structured hierarchically and are mutually dependent.

2. The interaction of all languages in a multilingual speaker has been shown repeatedly and thus, it should be treated as a basic assumption. Finding the transfer source in L3 acquisition in any given scenarios can thus not be informative in and of itself. Instead, observing CLI in multilinguals and at the same time factoring in what is known from other areas that are concerned with phonological contact and development can enrich our knowledge of phonological processes.

In summary, it is necessary to shift the focus and change the questions that are asked when studying the acquisition of L3 phonologies. Future studies on L3 systems should test well-defined phonological structures that – based on their physical make-up and psycholinguistic properties – lend themselves to meaningful comparisons. Findings from such comparisons can then be integrated into a more comprehensive description of contact phenomena in phonological systems in general. Furthermore, systematic and fine-grained analyses of unexpected CLI effects in multilingual systems can be used as an additional tool to uncover structural and psycholinguistic properties that cannot be observed directly: Differences in CLI effects observed in learners of an L3 with different language backgrounds that cannot be accounted for by the visible characteristics of the structures under scrutiny, could, for instance, serve as evidence for different degrees of importance of phonetic cues or different mental representations. By participating in this paradigm shift suggested by the analyses performed in this dissertation and consequently altering the assumptions and the questions that are, as of yet, prevalent in the field, L3 phonological research can contribute to larger, more general issues and thus go beyond transfer.

References

Adesope, O. O., Lavin, T., Thompson, T., & Ungerleider, C. (2010). A Systematic Review and Meta-Analysis of the Cognitive Correlates of Bilingualism. *Review of Educational Research*, *80*(2), 207–245.

Ainsworth, W. A., & Paliwal, K. K. (1984). Correlation between the production and perception of the English glides /w, r, l, j/. *Journal of Phonetics*, *12*(3), 237–243.

Altmann, H., Berger, I., & Braun, B. (2012). Asymmetries in the perception of non-native consonantal and vocalic length contrasts. *Second Language Research*, *28*(4), 387–413.

Amengual, M. (2019). Type of early bilingualism and its effect on the acoustic realization of allophonic variants: Early sequential and simultaneous bilinguals. *International Journal of Bilingualism*, *23*(5), 954–970.

Amengual, M., Meredith, L., & Panelli, T. (2019). Static and dynamic phonetic interactions in the L2 and L3 acquisition of Japanese velar voiceless stops. In S. Calhoun, P. Escudero, M. Tabain, & P. Warren (Eds.), *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019* (pp. 964–968). Canberra: Australasian Speech Science and Technology Association Inc.

Aoyama, K., Flege, J. E., Guion, S. G., Akahane-Yamada, R., & Yamada, T. (2004). Perceived phonetic dissimilarity and L2 speech learning: The case of Japanese /r/ and English /l/ and /r/. *Journal of Phonetics*, *32*(2), 233–250.

Atay, D., & Kurt, G. (2010). The socio-educational model of second language acquisition: The Turkish context. *Procedia - Social and Behavioral Sciences*, *2*(2), 3088–3093.

Baddeley, A. (1992). Working memory. *Science (New York, N.Y.)*, *255*(5044), 556–559.

Bailey, P. J., & Haggard, M. P. (1973). Perception and production: Some correlations on voicing of an initial stop. *Language and Speech*, *16*(3), 189–195.

Baker, W., & Trofimovich, P. (2006). Perceptual paths to accurate production of L2 vowels: The role of individual differences. *IRAL – International Review of Applied Linguistics in Language Teaching*, *44*(3), 297.

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, Melbourne, Australia 2019 (pp. 3725–3729). Canberra: Australasian Speech Science and Technology Association Inc.

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.

Bardel, C., & Falk, Y. (2012). The L2 status factor and the declarative/procedural distinction. In J. Cabrelli Amaro, S. Flynn, & J. Rothman (Eds.), *Studies in Bilingualism. Third Language Acquisition in Adulthood* (Vol. 46, pp. 61–78). Amsterdam: John Benjamins Publishing Company.

Bastian, C. C. von, Locher, A., & Ruflin, M. (2013). Tatool: A Java-based open-source programming framework for psychological studies. *Behavior Research Methods*, *45*(1), 108–115.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1).

Becker, M., Ketrez, N., & Nevins, A. (2011). The Surfeit of the Stimulus: Analytic Biases Filter Lexical Statistics in Turkish Laryngeal Alternations. *Language*, *87*(1), 84–125.

Beinborn, L., Zesch, T., & Gurevych, I. (2014). Predicting the difficulty of language proficiency tests. *Transactions of the Association for Computational Linguistics*, *2*, 517–529.

Benrabah, M. (1991). Learning English segments with two languages. In *Actes du XIIème Congrès International des Sciences Phonétiques* (pp. 334–337). Aix-en-Provence: Université de Provence, Service des Publications.

Bentahila, A. (1982). The influence of the L2 on the learning of L3. In A. Benhallam (Chair), *2nd Spring Conference,* Fes.

Best, C. T., McRoberts, G. W., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *The Journal of the Acoustical Society of America*, *109*(2), 775–794.

Best, C. T. (1995). A direct realist perspective on cross-language speech perception. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-language Research* (pp. 167–200). Timonium MD: York Press.

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In M. J. Munro & O.-S. Bohn (Eds.), *Language learning and language teaching. Language experience in second language speech learning: In honor of James Emil Flege* (pp. 13–45). Amsterdam, Philadelphia: J. Benjamins Pub.

Bialystok, E., & Craik, F. I. M. (2010). Cognitive and Linguistic Processing in the Bilingual Mind. *Current Directions in Psychological Science*, *19*(1), 19–23.

Birdsong, D., Gertken, L. M., & Amengual, M. (2012). *Bilingual Language Profile: An Easy-to-Use Instrument to Assess Bilingualism*. Austin: University of Texas. Retrieved from https://sites.la.utexas.edu/bilingual/

Birdsong, D. (1990). Universal Grammar and Second Language Acquisition Theory. *Studies in Second Language Acquisition*, *12*(3), 331–340.

Blank, C. A., & Zimmer, M. C. (2009). A transferência fonético-fonológica L2 (francês) – L3 (inglês): Um estudo de caso. *Revista de Estudos da Linguagem*, *17*(1).

Blevins, J. (2006). Syllable typology. In K. Brown (Ed.), *Encyclopedia of language & linguistics* (2nd ed., pp. 333–337). Amsterdam: Elsevier.

Brannen, K. (2011). *The Perception and Production of Interdental Fricatives in Second Language Acquisition* (Dissertation). McGill University, Ann Arbor.

Brockhaus, W. (1995). *Final Devoicing in the Phonology of German*. *Linguistische Arbeiten: Vol. 336*. Berlin: De Gruyter. Retrieved from http://www.degruyter.com/search?f_0=isbnissn&q_0=9783110966060&searchTitles=true

Broselow, E., Chen, S.-I., & Wang, C. (1998). The emergence of the unmarked in Second Language Phonology. *Studies in Second Language Acquisition*, *20*(2), 261–280.

Bullock, B. E. (2009). Phonetic reflexes of code-switching. In B. E. Bullock & A. J. Toribio (Eds.), *The Cambridge Handbook of Linguistic Code-switching* (pp. 163–181). Cambridge: Cambridge University Press.

Bybee, J. L. (2015). *Language change*. *Cambridge textbooks in linguistics*. Cambridge: Cambridge University Press.

Cabo, D. P. Y., & Rothman, J. (2012). The (Il)Logical Problem of Heritage Speaker Bilingualism and Incomplete Acquisition. *Applied Linguistics*, *33*(4), 450–455.

Cabrelli Amaro, J. (2012). L3 Phonology: An understudied domain. In J. Cabrelli Amaro, S. Flynn, & J. Rothman (Eds.), *Studies in Bilingualism. Third Language Acquisition in Adulthood* (pp. 33–60). Amsterdam: John Benjamins Publishing Company.

Cabrelli Amaro, J., & Iverson, M. (2018). Third Language Acquisition. In K. L. Geeslin (Ed.), *The Cambridge Handbook of Spanish Linguistics* (pp. 737–757). Cambridge University Press.

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. *IRAL - International Review of Applied Linguistics in Language Teaching*, *48*(2-3), 107.

Cebrian, J. (2006). Experience and the use of non-native duration in L2 vowel categorization. *Journal of Phonetics, 34*(3), 372–387.

Chamot, A. U. (1973). Phonological Problems in Learning English as a Third Language. *International Review of Applied Linguistics in Language Teaching*, *11*(3), 243–250.

Chang, S.-E. (2015, August). *Degree and direction of foreign accent in L2 and L3 Korean speech*. ICPhS XVIII, Glasgow.

Chumbow, B. S. (1981). The Mother Tongue Hypothesis in a Multilingual Setting. In J.-G. Savard & L. Laforge (Eds.): *Vol. 16. Travaux du Centre International de Recherche sur le Bilinguisme A, Études, Actes du 5e Congrès de l'Association Internationale de Linguistique Appliquée: Montréal, août/August, [20. - 26.] 1978* (pp. 42–55). Québec: Presses de l'Univ. Laval.

Clements, G. N., & Sezer, E. (1982). Vowel and Consonant Disharmony in Turkish. In H. van der Hulst & N. Smith (Eds.), *The structure of phonological representations* (pp. 213–255). Dordrecht: Foris.

Cocca, M., Pérez García, J. A., Zamaripa, J. I., Demetriou, Y., & Cocca, A. (2017). Psychometric parameters of the attitude/motivation test battery instrument in a mexican environment. *Journal of Sport Psychology*, *26*(2), 149–155.

Colantoni, L., & Steele, J. (2008). Integrating articulatory constraints into models of second language phonological acquisition. *Applied Psycholinguistics*, *29*(3), 489–534.

Council of Europe. (2001). *Common European Framework of References for Languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.

Council of Europe (2019). Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR). Retrieved from https://www.coe.int/en/web/common-european-framework-reference-languages

Coupé, C., Marsico, E., & Pellegrino, F. (2017). To What Extent Are Phonological Inventories Complex Systems? In S. Mufwene, C. Coupe, & F. Pellegrino (Eds.), *Complexity in Language* (pp. 135–164). Cambridge: Cambridge University Press.

Cruttenden, A. (2014). *Gimson's pronunciation of English* (8. edition). London, New York: Routledge Taylor & Francis Group.

Domene Moreno, C. (2019, June). *L1-L2 interaction in L3 acquisition - why we need multi-feature analyses*. L3 Workshop: Multilingual Language Acquisition, Processing and Use, Konstanz.

Domene Moreno, C., & Kabak, B. (to appear). Monolingual-bilingual (non-)convergence in L3 rhythm. In R. Fuchs (Ed.), *Prosody, Phonology and Phonetics. Speech Rhythm in L1, L2 and Learner Varieties of English.* New York: Springer.

Dörnyei, Z., & Csizér, K. (1998). Ten commandments for motivating language learners: Results of an empirical study. *Language Teaching Research*, *2*(3), 203–229.

Durand, J. (2005). Tense/lax, the vowel system of English and phonological theory. In P. Carr, J. Durand, & C. J. Ewen (Eds.), *Headhood, elements, specification and contrastivity: Phonological papers in honour of John Anderson* (pp. 77–97). Amsterdam: John Benjamins Pub.

Eckman, F. R. (1977). Markedness and the contrastive analysis hypothesis. *Language Learning*, *27*(2), 315–330.

Edge, B. A. (1991). The Production of Word-Final Voiced Obstruents in English by L1 Speakers of Japanese and Cantonese. *Studies in Second Language Acquisition*, *13*(3), 377–393.

Epstein, S. D., Flynn, S., & Martohardjono, G. (1996). Second language acquisition: Theoretical and experimental issues in contemporary research. *Behavioral and Brain Sciences*, *19*(4), 677–714.

Epsy-Wilson, C. Y. (1992). Acoustic measures for linguistic features distinguishing the semivowels/wjrl/in American English. *The Journal of the Acoustical Society of America*, *92*(2 Pt 1), 736–757.

ExperimentMFC 3.2. An AXB discrimination experiment (2016). Retrieved from http://www.fon.hum.uva.nl/praat/manual/ExperimentMFC_3_2__An_AXB_discrimination_experiment.html

Flege, J. E. (1987). The production of ‘new’ and ‘similar’ phones in a foreign language: Evidence for the effect of equivalence classification, *15*, 47–65.

Flege, J. E. (1991). Perception and production: the relevance of phonetic input to L2 phonological learning. In C. A. Ferguson & T. Huebner (Eds.), *Language acquisition & language disorders: Crosscurrents in second language acquisition and linguistic theories* (pp. 249–289). Amsterdam, Philadelphia: John Benjamins Pub. Co.

Flege, J. E. (1993). Production and Perception of a Novel, Second-language Phonetic Contrast. *The Journal of the Acoustical Society of America*, *93*(3), 1589–1608.

Flege, J. E. (1995). Second Language Speech Learning: Theories, Findings, Problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–277). Baltimore: York Press.

Flege, J. E. (2007). Language contact in bilingualism: Phonetic system interactions. In J. Cole & J. I. Hualde (Eds.), *Phonology and phonetics: Vol. 4,3. Laboratory phonology 9* (pp. 353–382). Berlin: Mouton de Gruyter.

Flynn, N., & Foulkes, P. (2011). Comparing vowel formant normalization methods. In *Online Proceedings of the XIIIth International Congress of Phonetic Sciences* (pp. 683–686). Hong Kong. Retrieved from https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2011/OnlineProceedings/RegularSession/Flynn/Flynn.pdf

Flynn, S., Foley, C., & Vinnitskaya, I. (2004). The Cumulative-Enhancement Model for Language Acquisition: Comparing Adults' and Children's Patterns of Development in First, Second and Third Language Acquisition of Relative Clauses. *International Journal of Multilingualism*, *1*(1), 3–16.

Gallardo del Puerto, F. (2007). Is L3 Phonological Competence Affected by the Learner's Level of Bilingualism? *International Journal of Multilingualism*, *4*(1), 1–16.

Gardner, R. C. (1985). *Social psychology and second language learning: The role of attitudes and motivation*. London: Arnold.

Gathercole, V. C. M. (2016). Factors moderating proficiency in bilingual speakers. In E. Nicoladis & S. Montanari (Eds.), *Bilingualism across the lifespan: Factors moderating language proficiency* (pp. 123–140).

Gathercole, V. C. M., Thomas, E. M., Kennedy, I., Prys, C., Young, N., Viñas Guasch, N., Jones, L. (2014). Does language dominance affect cognitive performance in bilinguals? Lifespan evidence from preschoolers through older adults on card sorting, Simon, and metalinguistic tasks. *Frontiers in Psychology*, *5*, 11.

Göksel, A., & Kerslake, C. (2005). *Turkish: A Comprehensive Grammar*. New York, London: Routledge.

Gordon, M. K. (2016). *Phonological Typology*. Oxford: Oxford University Press.

Gut, U. (2010). Cross-linguistic influence in L3 phonological acquisition. *International Journal of Multilingualism*, *7*(1), 19–38.

Halle, M., Hughes, G. W., & Radley, J.-P. A. (1957). Acoustic properties of stop consonants. *The Journal of the Acoustical Society of America*, *29*, 107–116.

Hammarberg, B. (1996). Conditions on Transfer in Phonology. In A. R. James & J. Leather (Eds.), *Second-language speech: Structure and process* (pp. 161–180). Berlin, New York: Mouton de Gruyter.

Hammarberg, B. (2001). Roles of L1 and L2 in L3 Production and Acquisition. In B. Hufeisen, J. Cenoz, & U. Jessner (Eds.), *Cross-linguistic influence in third language acquisition: Psycholinguistic perspectives* (pp. 21–41). Clevedon, UK, Buffalo: Multilingual Matters.

Hammarberg, B. (2005). *Processes in Third Language Acquisition*: Edinburgh University Press.

Hammarberg, B., & Williams, S. (1993). A study of third language acquisition. In B. Hammarberg (Chair), *Stockholm-Abo Conference*, Stockholm.

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

Harbert, W. (2007). *The Germanic languages*. *Cambridge language surveys*. Cambridge: Cambridge University Press. Retrieved from https://doi.org/10.1017/CBO9780511755071

Harrell, F. E. (2019). Hmisc: Harrell Miscellaneous. Retrieved from https://CRAN.R-project.org/package=Hmisc

Hausmann, D. (2006). "downgeloaded" und "geforwardet" - Sprecherverhalten in morphologischen Zweifelsfällen am Beispiel des Sprachgebrauch im Internet. *Arbeitspapiere (N.F.)*, *50*.

Hayes, B. P. (2007). Gradient well-formedness in Optimality Theory. In J. Dekkers, F. R. H. van der Leeuw, & J. M. van de Weijer (Eds.), *Oxford linguistics. Optimality theory: Phonology, syntax, and acquisition* (pp. 88–120). Oxford: Oxford University Press.

Himmel, M. C. (2018). *Phonetic and phonological variability in the L1 and L2 of late bilinguals: The case of /r/ and /l/* (Dissertation). Julius-Maximillians-Universität, Würzburg.

Höhne, J., Linden, B., Seils, E., & Wiebel, A. (2014). *Die Gastarbeiter: Geschichte und aktuelle soziale Lage*. Retrieved from Wirtschafts- und Sozialwissenschaftliches Institut website: https://www.boeckler.de/pdf/p_wsi_report_16_2014.pdf

Isbell, D. The perception-production link in L2 phonology. *MSU Working Papers in Second Language Studies*, *7*, 57–67.

Jarvis, S., & Pavlenko, A. (2010). *Cross-linguistic influence in language and cognition*. New York, NY: Routledge.

Johanson, L. (2001). *Discoveries on the Turkic Linguistic Map*. *Skrifter / Svenska Forskningsinstitutet i Istanbul: Vol. 5*. Stockholm: Svenska Forskningsinstitutet i Istanbul.

Johanson, L. (2006). The history of Turkic. In L. Johanson (Ed.), *The Turkic languages* (pp. 81–125). London: Routledge.

Johnson, K. (2012). *Acoustic and auditory phonetics* (3. ed.). Malden, Mass.: Wiley-Blackwell. Retrieved from http://lib.myilibrary.com/detail.asp?ID=317822

Judy, T., Guijarro-Fuentes, P., & Rothman, J. (2008). Adult Accessibility to L2 Representational Features: Evidence from the Spanish DP. In M. A. Bowles, M. Foote, S. Perpinán, & R. Bhatt (Eds.), *Selected proceedings of the 2007 Second Language Research Forum* (pp. 1–21). Somerville, MA: Cascadilla Press.

Kabak, B. (2004). Acquiring phonology is not acquiring inventories but contrasts: The loss of Turkic and Korean primary long vowels. *Linguistic Typology*, *8*(3).

Kabak, B. (2019). A dynamic equational approach to sound patterns in language change and second language acquisition: The (un)stability of English dental fricatives illustrated. In R. Moehlig Falke & B. Busse (Eds.), *Patterns in Language and Linguistics: New Perspectives on a Ubiquitous Concept* (pp. 228–254). Berlin: Mouton de Gruyter.

Kabak, B. (2007). Hiatus resolution in Turkish: An underspecification account. *Lingua*, *117*(8), 1378–1411.

Kabak, B. (2011). Turkish vowel harmony. In M. van Oostendorp, C. J. Ewen, E. Hume, & K. Rice (Eds.), *The Blackwell Companion to Phonology*. Oxford, UK: John Wiley & Sons, Ltd.

Kager, R. (2010). *Optimality theory*. *Cambridge textbooks in linguistics*. Cambridge: Cambridge Univ. Press.

Kaushanskaya, M., & Marian, V. (2009). The bilingual advantage in novel word learning. *Psychonomic Bulletin & Review*, *16*(4), 705–710.

Kendall, T., & Thomas, E. R. (2018). Vowels: Vowel manipulation, Normalization, and Plotting. Retrieved from https://CRAN.R-project.org/package=vowels

Klein-Braley, C., & Raatz, U. (1984). A survey of research on the C-Test. *Language Testing*, *1*(2), 134–146.

Kohler, K. J. (1995). *Einführung in die Phonetik des Deutschen*. Berlin: Schmidt.

Kohler, K. J. (1999). German. In International Phonetic Association. (1999). *Handbook of the International Phonetic Association: A guide to the use of the international phonetic alphabet*. Cambridge: Univ. Press.

König, E., & Gast, V. (2012). *Understanding English-German contrasts.* Berlin: Schmidt.

Kopečková, R. (2016). The bilingual advantage in L3 learning: A developmental study of rhotic sounds. *International Journal of Multilingualism*, *13*(4), 410–425.

Kopečková, R. (2019, May). *The dynamics of multilingual phonological acquisition in the initial stages of L3 learning.* Talk given at the Department of English, Münster.

Kopečková, R., Gut, U., & Golin, C. (2019). Acquisition of the /v/-/w/ contrast by L1 German children and adults. In S. Calhoun, P. Escudero, M. Tabain, & P. Warren (Eds.), *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne,* *Australia 2019* (pp. 3745–3749). Canberra: Australasian Speech Science and Technology Association Inc.

Kornfilt, J. (2009). Turkish and the Turkic languages. In B. Comrie (Ed.), *The world's major languages* (2nd ed., pp. 519–544). New York: Routledge.

Krzysik, I., & Wrembel, M. (2019, June). *The relationship between speech production and inhibitory control in young multilinguals.* L3 Workshop: Multilingual Language Acquisition, Processing and Use, Konstanz.

Kupisch, T., & Rothman, J. (2018). Terminology matters! Why difference is not incompleteness and how early child bilinguals are heritage speakers. *International Journal of Bilingualism*, *22*(5), 564–582.

Lee-Kim, S.-I., Davidson, L., & Hwang, S. (2013). Morphological effects on the darkness of English intervocalic /l/. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, *4*(2).

Lenneberg, E. H. (1967). *Biological foundations of language*. New York: Wiley.

Lenth, R. V. (2016). Least-Squares Means: The R Package lsmeans. *Journal of Statistical Software*, *69*(1), 1–33.

Levshina, N. (2015). *How to do linguistics with R: Data exploration and statistical analysis*. Amsterdam, Philadelphia: John Benjamins Publishing Company.

Lewandowska, H., & Wrembel, M. (2019). The mutual influence of vowel length perception in the second and third language. In S. Calhoun, P. Escudero, M. Tabain, & P. Warren (Eds.), *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019* (pp. 3750–3753). Canberra: Australasian Speech Science and Technology Association Inc.

Lindblom, B., & Maddieson, I. (1988). Phonetic Universals in Consonant Systems. In L. M. Hyman & C. N. Li (Eds.), *Language, speech and mind: Studies in honour of Victoria A. Fromkin* (pp. 62–78). London: Routledge.

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.

Llisterri, J., & Martinez-Dauden, G. (1990, April). *Phonetic interference in bilingual speakers learning a third language: the production of lateral consonants*. Meeting of the World Congress of Applied Linguistics, Thessaloniki.

Llisterri, J., & Poch-Olivé, D. (1987). Phonetic interference in bilinguals' learning of a third language. In *XIth ICPHS: the Eleventh International Congress of Phonetic Sciences,* Tallin, Estonia.

Lloyd-Smith, A., Gyllstad, H., & Kupisch, T. (2017). Transfer into L3 English. *Linguistic Approaches to Bilingualism*, *7*(2), 131–162.

Lohndal, T., Rothman, J., Kupisch, T., & Westergaard, M. (2019). Heritage language acquisition: What it reveals and why it is important for formal linguistic theories. *Language and Linguistics Compass*, *36*(3), 165.

Long, M. H. (1990). Maturational Constraints on Language Development. *Studies in Second Language Acquisition*, *12*(3), 251–285.

Louriz, N. (2007). Alignment in L3 phonology, *18/19*, 129–160.

Lowenstamm, J. (1996). CV as the only syllable type. In J. Durand & B. Laks (Eds.), *Current Trends in Phonology: Models and Methods* (pp. 419–442). Manchester: University of Salford Publications.

Macmillan, N. A., & Kaplan, H. L. (1985). Detection theory analysis of group data: Estimating sensitivity from average hit and false-alarm rates. *Psychological Bulletin*, *98*(1), 185–199.

Maddieson, I. (2013). Presence of uncommon consonants. In M. S. Dryer & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online.* Leipzig: Max Planck Institute for Evolutionary Anthropology. Retrieved from (https://wals.info/chapter 19

Maher, J. C. (2017). *Multilingualism: A very short introduction*. Oxford: Oxford University Press.

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica*, *22*(3), 276–282.

McWhorter, J. H. (2002). What happened to English? *Diachronica*, *19*(2), 217–272.

Missaglia, F. (2010). The acquisition of L3 English vowels by infant German–Italian bilinguals. *International Journal of Multilingualism*, *7*(1), 58–74.

Montrul, S. A. (1997). *Transitivity Alternations in Second Language Acquisition: A Cross-linguistic Study of English, Spanish and Turkish* (Dissertation). McGill University.

Montrul, S. A. (2010). Dominant language transfer in adult second language learners and heritage speakers. *Second Language Research*, *26*(3), 293–327. https://doi.org/10.1177/0267658310365768

Montrul, S. A. (2015). Bilingualism and the Heritage Language Speaker. In T. K. Bhatia & W. C. Ritchie (Eds.), *The handbook of bilingualism and multilingualism* (pp. 168–189). Chichester: Wiley-Blackwell.

Moyer, A. (2013). *Foreign accent: The phenomenon of non-native speech*. Cambridge: Cambridge University Press.

Need for the BLP. Retrieved from https://sites.la.utexas.edu/bilingual/need-for-the-blp/

Nimz, K. (2015). The Perception of Vowel Quality and Quantity by Turkish Learners of German as a Foreign Language. In E. Raimy & C. E. Cairns (Eds.), *The Segment in Phonetics and Phonology* (pp. 253–266). Chichester, UK: John Wiley & Sons, Ltd.

Onishi, H. (2016). The effects of L2 experience on L3 perception. *International Journal of Multilingualism*, *13*(4), 459–475.

Ortega, L. (2013). *Understanding second language acquisition*. London, New York: Routledge Taylor & Francis Group.

Paap, K. R., & Greenberg, Z. I. (2013). There is no coherent evidence for a bilingual advantage in executive processing. *Cognitive Psychology*, *66*(2), 232–258.

Paap, K. R., Johnson, H. A., & Sawi, O. (2015). Bilingual advantages in executive functioning either do not exist or are restricted to very specific and undetermined circumstances. *Cortex*, *69*, 265–278.

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.

Patkowski, M. S. (1980). The sensitive period for the acquisition of syntax in a second language. *Language Learning*, *30*(2), 449–468.

Pereltsvaig, A. (2017). *Languages of the world: An introduction*. Cambridge, United Kingdom, New York, NY, Port Melbourne, Australia, Delhi, India, Singapore: Cambridge University Press.

Port, R. F., & Dalby, J. (1982). Consonant/vowel ratio as a cue for voicing in English. *Perception & Psychophysics*, *32*(2), 141–152.

Port, R. F., & O'Dell, M. L. (1985). Neutralization of syllable-final voicing in German. *Journal of Phonetics*, *13*(4), 455–471.

Puig-Mayenco, E., González Alonso, J., & Rothman, J. (2018). A systematic review of transfer studies in third language acquisition. *Second Language Research*, *59*, 026765831880914.

R Core Team. (2018). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Ranja, C., & Najari, V. (2019). nlcor: Compute Nonlinear Correlations.

Recansens, D. (2012). A cross-language acoustic study of initial and final allophones of /l/. *Speech Communication*, *54*(3), 368-383.

Roach, P. (1999). British English: Received Pronunciation. *Journal of the International Phonetic Association*, *34*(2), 239–245.

Rodrigues, S., Martins, F., Silva, S., & Jesus, L. M. T. (2019). /l/ velarisation as a continuum. *PloS One*, *14*(3), e0213392.

Rolffs, S. (2005). Türkisch. In U. Hirschfeld (Ed.), *Phonetik International: von Afrikaans bis Zulu.* Waldsteinberg: Popp.

Ross, S., Yoshinaga, N., & Sasaki, M. (2002). Aptitude-exposure interaction effects on Wh-movement violation detection by pre-and-post-critical period Japanese bilinguals. In P. Robinson (Ed.), *Individual Differences and Instructed Language Learning* (Vol. 2, pp. 267–299). Amsterdam: John Benjamins Publishing Company.

Rothman, J. (2009). Understanding the nature and outcomes of early bilingualism: Romance languages as heritage languages. *International Journal of Bilingualism*, *13*(2), 155–163.

Rothman, J., Iverson, M., & Judy, T. (2010). L3 syntactic transfer selectivity and typological determinacy: The typological primacy model. *Second Language Research*, *27*(1), 107–127.

Russ, C. V. J. (2010). *The sounds of German*. Cambridge: Cambridge University Press.

Schachter, J. (1988). Second Language Acquisition and Its Relationship to Universal Grammar. *Applied Linguistics*, *9*(3), 219–235.

Schmid, M. S., & Dusseldorp, E. (2010). Quantitative analyses in a multivariate study of language attrition: The impact of extralinguistic factors. *Second Language Research*, *26*(1), 125–160.

Schmidt, R. W. (1990). The Role of Consciousness in Second Language Learning. *Applied Linguistics*, *11*(2), 129–158.

Schober, P., Boer, C., & Schwarte, L. A. (2018). Correlation Coefficients: Appropriate Use and Interpretation. *Anesthesia and Analgesia*, *126*(5), 1763–1768.

Schwartz, B. D., & Sprouse, R. A. (1996). L2 cognitive states and the Full Transfer/Full Access model. *Second Language Research*, *12*(1), 40–72.

Scoring and interpreting the results (2019). Retrieved from https://sites.la.utexas.edu/bilingual/scoring-and-interpreting-the-results/

Sezer, E. (1986). An autosegmental analysis of compensatory lengthening in Turkish. In L. Wetzels & E. Sezer (Eds.), *Studies in compensatory lengthening* (pp. 227–250). Dordrecht: Foris Publ.

Sharwood Smith, M. (2017). Active–passive bilingualism and the functional distance between L1 and L2 as viewed within one unifying theoretical framework. In A. Ardila, A. B. Cieślicka, R. R. Heredia, & M. Roselli (Eds.), *Psychology of Bilingualism: The Cognitive and Emotional World of Bilinguals* (pp. 157–185). Cham: Springer.

Sheldon, A., & Strange, W. (1982). The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception. *Applied Psycholinguistics*, *3*(3), 243–261.

Simms, L. J., Zelazny, K., Williams, T. F., & Bernstein, L. (2019). Does the number of response options matter? Psychometric perspectives using personality questionnaire data. *Psychological Assessment*, *31*(4), 557–566.

Slabakova, R. (2016). The scalpel model of third language acquisition. *International Journal of Bilingualism*, *21*(6), 651–665.

Smith, W. D. (1995). Clarification of Sensitivity Measure A'. *Journal of Mathematical Psychology*, *39*(1), 82–89.

Sönning, L. (2017). *Phonological variation in German learner English* (Dissertation). Otto-Friedrich-Universität Bamberg, Bamberg.

Spinu, L. E., Hwang, J., & Lohmann, R. (2016). Is there a bilingual advantage in phonetic and phonological acquisition? The initial learning of word-final coronal stop realization in a novel accent of English. *International Journal of Bilingualism*, *22*(3), 350–370.

Sproat, R., & Fujimura, O. (1993). Allophonic variation in English /l/ and its implications for phonetic implementation. *Journal of Phonetics*, *21*, 291–311.

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, *31*(1), 137–149.

Stoehr, A., Benders, T., van Hell, J. G., & Fikkert, P. (2017). Second language attainment and first language attrition: The case of VOT in immersed Dutch-German late bilinguals. *Second Language Research*, *33*(4), 483–518.

Stone, J. M., & Towse, J. N. (2015). A Working Memory Test Battery: Java-Based Collection of Seven Working Memory Tasks. *Journal of Open Research Software*, *3*(2), 92.

Sypiańska, J. (2013). *Quantity and quality of language use and L1 attrition of Polish due to L2 Danish and L3 English* (Dissertation). Adam Mickiewicz University, Poznan.

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

Topintzi, N. (2011). Onsets. In M. van Oostendorp, C. J. Ewen, E. Hume, & K. Rice (Eds.), *The Blackwell Companion to Phonology* (pp. 1285–1308). Oxford, UK: John Wiley & Sons, Ltd.

Torchiano, M. (2016). Effsize - A Package For Efficient Effect Size Computation: Zenodo.

Toribio, A. J., Bullock, B. E., Botero, C. G., & Davis, K. A. (2005). Perseverative Phonetic Effects in Bilingual Code-Switching. In R. Gess & E. J. Rubin (Eds.), *Theoretical and*

Experimental Approaches to Romance Linguistics (Vol. 272, pp. 291–306). Amsterdam: John Benjamins Publishing Company.

Traunmüller, H. (1997). Auditory scales of frequency representation. Retrieved from http://www.ling.su.se/staff/hartmut/bark.htm

Treffers-Daller, J. (2019). What Defines Language Dominance in Bilinguals? *Annual Review of Linguistics*, *5*(1), 375–393.

Tremblay, M.-C. (2007, August). *L2 influence on L3 pronunciation: Native-like VOT in the L3 Japanese of English-French bilinguals*. L3 phonology workshop, Freiburg.

Turton, D. (2017). Categorical or gradient? An ultrasound investigation of /l/-darkening and vocalization in varieties of English. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, *8*(1), 497.

Ulusoy, M. (2008). Evaluating sixth graders' reading levels with different cloze test formats. *Eurasian Journal of Educational Research*. (31), 115–133.

Unsworth, S. (2016). Quantity and quality of language input in bilingual language development. In E. Nicoladis & S. Montanari (Eds.), *Bilingualism across the lifespan: Factors moderating language proficiency* (pp. 103–122).

Valdés, G. (2005). Bilingualism, Heritage Language Learners, and SLA Research: Opportunities Lost or Seized? *The Modern Language Journal*, *89*(3), 410–426.

Valian, V. (2015). Bilingualism and cognition. *Bilingualism: Language and Cognition*, *18*(1), 3–24.

Valian, V. (2016). Bilingualism and Executive Function. *Linguistic Approaches to Bilingualism*, *6*(5), 565–574.

Van der Hulst, H., & van de Weijer, J. (1991). Topics in Turkish phonology. In H. E. Boeschoten (Ed.), *Turkish linguistics today* (pp. 11–59). Leiden, New York, København, Köln: Brill.

Van Oostendorp, M. (2019). Germanic syllable structure. In M. T. Putnam & B. R. Page (Eds.), *The Cambridge Handbook of Germanic Linguistics.* Cambridge: Cambridge University Press.

Wagner, R. K., Torgensen, J. K., Rahotte, C. A., & Pearson, N. A. (2013). *Comprehensive test of phonological processing - Second edition (CTOPP-2).* Austin: PRO-ED.

Weinberger, S. (1996). Minimal segments in second language phonology. In A. R. James & J. Leather (Eds.), *Second-language speech: Structure and process* (pp. 263–312). Berlin, New York: Mouton de Gruyter.

Weinberger, S. (2015). Speech Accent Archive. Retrieved from ttp://accent.gmu.edu

Werker, J. F., & Logan, J. S. (1985). Cross-language evidence for three factors in speech perception. *Perception & Psychophysics, 37*(1), 35–44.

Westergaard, M. (2009). *The acquisition of word order: Micro-cues, information structure, and economy*. *Linguistik aktuell: Vol. 145*. Amsterdam: Benjamins.

Westergaard, M., Mitrofanova, N., Mykhaylyk, R., & Rodina, Y. (2016). Cross-linguistic influence in the acquisition of a third language: The Linguistic Proximity Model. *International Journal of Bilingualism*, *21*(6), 666–682.

White, L. (1990). Second Language Acquisition and Universal Grammar. *Studies in Second Language Acquisition*, *12*(2), 121–133.

White, L. (2003). On the nature of interlanguage representation: Universal Grammar in the Second Language. In C. J. Doughty & M. H. Long (Eds.), *The Handbook of Second Language Acquisition* (pp. 19–42). Oxford, UK: Blackwell Publishing Ltd.

White, L. (2013). *Second language acquisition and universal grammar*. Cambridge: Cambridge University Press.

Wiese, R. (1996). *The phonology of German*. Oxford: Oxford University Press

Wirtz, M. A. (2019). Cohens Kappa. In M. A. Wirtz (Ed.), *Dorsch - Lexikon der Psychologie.* Retrieved from https://portal.hogrefe.com/dorsch/cohens-kappa/

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

Wrembel, M. (2011, August). *Cross-linguistic influence in third language acquisition of voice onset time*. ICPhS XVII, Hong Kong.

Wrembel, M. (2012). Foreign accentedness in third language acquisition. In J. Cabrelli Amaro, S. Flynn, & J. Rothman (Eds.), *Studies in Bilingualism. Third Language Acquisition in Adulthood* (pp. 281–310). Amsterdam: John Benjamins Publishing Company.

Wrembel, M. (2013). Foreign Accent Ratings in Third Language Acquisition: The Case of L3 French. In E. Waniek-Klimczak & L. R. Shockey (Eds.), *Second Language Learning and Teaching. Teaching and Researching English Accents in Native and Non-native Speakers* (pp. 31–47). Berlin, Heidelberg: Springer Berlin Heidelberg.

Wrembel, M. (2015). Cross-linguistic influence in second vs. third language acquisition of phonology. In E.-M. Wunder, R. Fuchs, & U. Gut (Eds.), *Universal or diverse paths to English phonology* (pp. 41–70). Berlin: De Gruyter Mouton.

Wrembel, M., Marecka, M., & Kopečková, R. (2019). Extending perceptual assimilation model to L3 phonological acquisition. *International Journal of Multilingualism*, *5*, 1–21.

Wunder, E.-M. (2015). *On the Hunt for Lateral Phonological Cross-linguistic Influence in Third or Additional Language Acquisition* (Dissertation). Westfälische Wilhelms-Universität, Münster.

Yavaş, M. (1980). Some pilot experiments on Turkish vowel harmony. *Paper in Linguistics*, *13*(3), 543–562.

Yavaş, M., & Altan, A. (2016). Vowel epenthesis in the acquisition of English /s/-clusters by Turkish speakers. In M. Güven, D. Akar, B. Öztürk, & M. Kelepir (Eds.), *Exploring the Turkish Linguistic Landscape* (Vol. 175, pp. 3–20). Amsterdam: John Benjamins Publishing Company.

Yildiz, Y. (2006). The acquisition of English interdentals by Turkish learners: explaining age effects in L2 phonology. *Essex Graduate Papers in Language and Linguistics*, *8*, 144–165.

Yilmaz, G., & Schmid, M. S. (2018). First language attrition and bilingualism. In D. Miller, F. Bayram, J. Rothman, & L. Serratrice (Eds.), *Bilingual cognition and language: The state of the science across its subfields* (pp. 225–250). Amsterdam, Philadelphia: John Benjamins Publishing Company.

Zhang, J. (2019). Feature-specific advantages in L3 phonological acquisition. In S. Calhoun, P. Escudero, M. Tabain, & P. Warren (Eds.), *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019* (pp. 3740–3744). Canberra: Australasian Speech Science and Technology Association Inc.

Zimmer, K. & Orgun, O. (1999). Turkish. In International Phonetic Association. (1999). *Handbook of the International Phonetic Association: A guide to the use of the international phonetic alphabet*. Cambridge: Univ. Press.

Appendices

Appendix A: Speaker profiles

Appendix A.1 – Study 1

Participant ID	Age	Gender	Handedness	Grade
MK01	11	f	r	6
MK02	11	f	r	6
MK03	12	f	r	6
MK04	10	f	r	5
MK05	12	f	r	7
MK06	13	f	r	6
MK07	11	f	r	6
MK08	10	f	r	5
MK09	10	m	l	5
MK10	12	f	r	7
MK11	12	m	r	7
MK12	12	m	l	7
BK03	11	m	r	5
BK04	10	f	r	5
BK05	12	f	r	5
BK06	11	m	r	6
BK07	13	f	r	6
BK08	12	m	r	6
BK09	12	f	r	6
BK10	11	f	r	6
BK11	12	m	r	6
BK12	11	f	r	5
BK13	11	f	r	5
BK15	12	f	r	6

Appendix A.2 – Study 2

Participant ID	Age	Gender	Additional languages[86]	English abroad[87]	German home	Turkish home
MA01	27	m	Fr	+	Eichstätt	-
MA02	24	m	Fr, Lt, Sp	+	Peißenberg	-
MA04	23	f	Lt, It	-	Schweinfurt	-
MA05	28	m	Fr, Lt	+	Kronach	-
MA06	21	f	Fr, Sp	+	Erlenbach a. M.	-
MA07	20	f	Fr, It	-	Schwäbisch Hall	-
MA08	24	f	Fr, It	+	Flieden	-
MA09	22	m	Fr, Sp, MC	+	Ebersberg	-
MA10	23	f	Fr	+	Aschaffenburg	-
MA11	29	f	Fr	-	Lauffen a. N.	-
MA12	24	f	Fr, It	-	Ansbach	-
MA13	26	f	Fr	+	Wertheim	-
BA01	24	f	-	-	Aschaffenburg	Kayseri
BA02	24	f	Fr	-	Aschaffenburg	Gaziantep
BA03	25	m	Ar	-	Schützingen	Isparta
BA04	19	f	Lt, Fr	-	Marktheidenfeld	Konya
BA05	18	f	Lt	-	Prien am Chiems.	Bursa
BA06	23	f	Sp	-	Schwäbisch Hall	Sinop
BA07	23	f	Sp	+	Erlenbach a. M.	Gaziantep
BA08	23	f	Fr, Sp, Lt	+	Köln	Gümüşhane
BA09	23	f	Fr, Ar	-	Erlenbach a. M.	Istanbul
BA10	21	f	It	-	Würzburg	Elaziğ
BA11	32	m	Sp	-	Kitzingen	İzmir
BA12	20	f	Fr, It, Kr, MC, Ar	-	Krumbach	Uşak

[86] Fr = French, Lt = Latin, Sp = Spanish, It = Italian, MC = Mandarin Chinese, Ar = Arabic, Kr = Korean
[87] More than two months in an English speaking country

Appendix B: Test items for Study 1

Appendix B.1 – Compounds

Feature	Test compounds
Consonant cluster	**cross** lamp **glass** lip **bread** puzzle **tree** mouse
Vowel length	**food** bike **moon** fall **sheep** tree **bean** house
Lateral	**ill** frog **ball** fish **towel** dress **pill** box
Voiced coda	**bed** apple **club** owl **bag** eye **sad** orange
GOAT	**boat** pie **snow** spider **nose** bath **coat** fish
Interdental fricative	**mouth** light **thumb** watch **tooth** worm **three** bush
TRAP	**back** drum **cat** fruit **pan** shower **cap** shop
- (items for training round)	paper window child car pen doll bear party

Appendix B.2 – German and Turkish reference words

Feature	Words
Consonant clusters (German)	Brief (letter) Treppe (chairs) Glocke (bell) Kreis (circle)
Vowel length (German)	Lippe (lip) Sieb (sieve) Stuhl (chair) Puppe (doll)
Lateral allophones (Turkish)	kol (arm) pul (stamp) el (hand) lokum (Turkish delight)
Voiced coda	makyaj (make up) ad (name) garaj (garage) göz (eye)

Appendix B.3 – Nonsense word list for PWM test

1-syllable	2-syllable	3-syllable	4-syllable
nʊks	ˈzɛsɐ	zʊˈga:lʊ	dɔkɔˈza:nʊ
lɛn	ˈnanə	lɪˈzu:ta	rʊzaˈti:nɔ
hɔl	ˈtifʊ	zɔˈva:di	pɔʃʊˈna:kə
nɛs	ˈpo:ri	pɪˈno:ta	tɛmɪˈça:nə
ʃɪm	ˈko:na	paˈlʊχɔ	lʊnɔˈka:ri
vɔp	ˈro:mʊ	vaˈni:mʊ	zalɔˈpa:nʊ
lu:n	ˈba:vi	tɔˈpe:gə	nɪlɪˈro:na
tsy:k	ˈlʊtɔ	dʊˈlu:ra	bazɔˈru:kə
ze:f	ˈhu:də	kiˈne:fa	fɪnuˈpo:gʊ
ga:t	ˈfi:ma	fɪˈço:li	rapaˈnu:gə

Appendix C: Questionnaires

Appendix C.1 – Biographical questionnaire, Study 1

Fragebogen

Name: ____

Alter: ____

Geschlecht:____

Muttersprache(n): ____

Geburtsort: ____

Wohnort: ____

Schule: ____

Eltern (Name/Alter/Nationalität): ____

____

____

Geschwister (Name/Alter/Geburtsort): ____

____

____

Welche Sprache(n) sprichst du normalerweise...

... zu Hause? ____

... in der Schule? ____

... mit deinen Freunden? ____________________________________

In welcher Sprache...

... siehst du Filme/Serien? ______________________________

... liest du? ____

... ist die Musik, die du hörst? __________________________

Appendix C.2 – Consent form, Study 1

Einwilligung zur Teilnahme an einer wissenschaftlichen Studie

Versuchleiter

Christina Domene Moreno M.A.

Büroadresse: Julius-Maximilians-Universität Würzburg

Philosophische Fakultät

Neuphilologisches Institut

Lehrstuhl für Englische Sprachwissenschaft

Am Hubland

97074 Würzburg

Email: christina.domene-moreno@uni-wuerzburg.de

Erklärung der Studie

Wir möchten untersuchen, wie ein- und zweisprachige Menschen eine weitere Sprache erlernen. Insgesamt werden ca. 24 Personen teilnehmen. Ihr Kind wird gebeten, eine Reihe von Aufgaben auszuführen: Unterscheidung von Wörtern am PC, Nachsprechen von zusammengesetzen Wörtern, Erinnern von Zahlenreihen, Ausfüllen zweier Fragebögen. Während diesen Aufgaben wird Ihr Kind aufgenommen (nur Ton). Alle Daten werden zur weiteren Analyse aufgezeichnet.

Risiken

Es gibt keine Risiken durch die Teilnahme an der Studie.

Bezahlung und Vorteile

Sie erhalten keine Bezahlung für die Teilnahme an der Studie. Es gibt keinerlei Vorteile, die Sie durch die Teilnahme an der Studie erhalten.

Vertraulichkeit

Jegliche Information, die gesammelt wird, wird vertraulich behandelt und ausschließlich zum Zweck der Forschung genutzt. Teile der Aufnahmen können zu Demonstrationszwecken der linguistischen Analyse in linguistischen Kursen, Präsentationen auf Konferenzen oder in Forschungsberichten veröffentlicht werden. Die Ergebnisse der Studie können bei Konferenzen oder in Publikationen veröffentlicht werden.

Die Teilnehmer können anonym oder unter einem Pseudonym an der Studie teilnehmen, und werden während der Aufnahme nicht nach Ihrem Namen gefragt. Wenn Sie es wünschen, wird der Name Ihres Kindes nicht mit den Daten in Verbindung gebracht und nicht gespeichert. Es wird jedem Teilnehmer stattdessen ein anonymisierter Code zugeordnet. Selbst wenn Ihnen Anonymität nicht wichtig ist und Ihr Kind seinen Namen angeben möchten, so wird der Name dennoch niemals in schriftlichen oder mündlichen Veröffentlichungen der Studie genannt, und niemals öffentlich mit den gesammelten Daten in Verbindung gebracht.

Freiwilligkeit der Teilnahme

Die Teilnahme an dieser Studie ist strikt freiwillig. Wenn Ihr Kind sich aus irgendwelchen Gründen unwohl fühlt und das Experiment abbrechen möchte, soll er/sie bitte den Versuchsleiter informieren. Ihr Kind darf jederzeit ohne Angabe von Gründen die Teilnahme am Experiment beenden. Die Entscheidung, die Teilnahme zu beenden, zieht keinerlei negative Konsequenzen nach sich.

Ergebnisse und weitere Fragen

Wenn Sie es wünschen, kontaktieren wir Sie nach Beendigung der Experimente und erklären Ihnen die Ergebnisse. Wenn Sie weitere Fragen oder Probleme nach dem Teilnahmetermin haben, zögern Sie nicht uns unter oben angegebener Email-Adresse zu kontaktieren.

Stellungnahme des Versuchsleiters

Ich habe dem Teilnehmer die Versuchsreihe erklärt. Ich habe alle Fragen beantwortet und, wenn nötig, Übersetzungen von Fachbegriffen und Konzepten vorgenommen.

Name ______________________________

Datum ___________________ Unterschrift ______________________________

Zustimmung des Erziehungsberechtigten

Ich habe die Informationen in diesem Zustimmungsformular gelesen. Mein Kind nimmt freiwillig an der Studie teil.

Name ______________________________

Datum ___________________ Unterschrift ______________________________

Appendix C.3 – Consent form, Study 2

Consent to participate in research

Contact:

Christina Domene Moreno

christina.domene-moreno@uni-wuerzburg.de

Explanation of the study

I am interested in the process of acquisition of a phonological system by monolingual and bilingual learners. During the experiment, you will do a perception task on a computer, read texts in different languages, and fill in a questionnaire. You will also be briefly interviewed in English. All the tests involving production will be recorded (audio only).

Risks of participation

There are no risks associated with participation in this study.

Payment and benefits

You will not be paid for participating in this study. There are no direct benefits that you can expect to receive from participation in this study.

Confidentiality

All the information collected will be confidential and will only be used for linguistic research. You may participate anonymously or under a pseudonym, and will not be asked your name during the interview. The results from the study might be presented at scientific conferences and published in scientific journals and books. If data is published, your name will not be used.

Voluntary Participation

Participating in this study is strictly voluntary. You may refuse to answer any question you do not want to answer without giving a reason. You may demand the deletion of audio sections you feel uncomfortable with. You are free to stop at any time without any reason, and will not be penalized for choosing to end the interview.

Findings and further questions

If you would like me to, I will contact you to tell you about the results after the study has been concluded. If you have any further questions or inquiries after your session, do not hesitate to contact me via the email address indicated above.

Researcher's statement

I have fully explained this study to the participant. I have answered all questions and, if necessary, provided explanations of key terms and concepts.

Name

Date

Signature

Participant's consent

I have read and understood the information provided in this consent form. I voluntarily agree to participate in this study.

Name

Date

Signature

Appendix C.4 – Bilingual Language Profile

I. Biographische Information

Alter: ______

Geschlecht: ❑ männlich ❑ weiblich

Geburtsort: ______________________________

Derzeitiger Wohnort: ______________________________

Muttersprache(n): ______________________________

Welche weiteren Sprachen sprechen Sie bzw. haben Sie gelernt, und in welchem Alter haben Sie begonnen, diese zu lernen?

Sprache: ____________________ Alter: ____________

Sprache: ____________________ Alter: ____________

Sprache: ____________________ Alter: ____________

Sprache: ____________________ Alter: ____________

Sprache: ____________________ Alter: ____________

Höchster Abschluss: ❑ kein Abschluss ❑ (Qualifizierender) Hauptschulabschluss ❑ Mittlere Reife

❑ (Fach-)Abitur ❑ B.A., B.S. (oder vergleichbares) ❑ M.A., M.Sc.

❑ Promotion ❑ anderer Abschluss: ____________

Geburtsort der Mutter: ______________________________

Muttersprache der Mutter: ______________________________

Geburtsort des Vaters: ______________________________

Muttersprache des Vaters: ______________________________

Haben Sie längere Zeit (mehr als zwei Monate) in einem türkischsprachigen Land verbracht? Falls ja, wie lange?

❑ ja, von ________ bis __________ ❑ nein

Haben Sie längere Zeit (mehr als zwei Monate) im englischsprachigen Ausland verbracht? Falls ja, wann? Wo?

Ort: ______________________________ von: ________ bis: __________

Ort: ______________________________ von: ________ bis: __________

Ort: ______________________________ von: ________ bis: __________

II. Sprachlicher Hintergrund

Bitte beantworten Sie die folgenden Fragen zu Ihrer Sprachbiographie, indem Sie das entsprechende Kästchen ankreuzen.

1. Mit welchem Alter haben Sie **angefangen**, die folgenden Sprachen zu lernen?

Deutsch

❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑
Seit Geburt	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20+

Türkisch

❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑
Seit Geburt	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20+

2. Ab welchem Alter haben Sie sich damit **wohlgefühlt**, die folgenden Sprachen zu benutzen?

Deutsch

❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20+	noch nicht

Türkisch

❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20+	noch nicht

3. Wie viele Jahre waren die folgenden Sprachen **Unterrichtssprachen (Geschichte, Mathematik etc.)** (Grundschule bis Universität)?

Deutsch

❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20+

Türkisch

❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20+

4. Wie viele Jahre haben Sie in einem **Land/einer Region** verbracht, in der die folgenden Sprachen gesprochen werden?

Deutsch

❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20+

Türkisch

❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20+

5. Wie viele Jahre haben Sie in einem **familiären Umfeld** verbracht, in der die folgenden Sprachen gesprochen wurden?

Deutsch

❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20+

Türkisch

❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20+

6. Wie viele Jahre haben Sie in einer **Arbeitsumgebung** verbracht, in der die folgenden Sprachen gesprochen wurden?

Deutsch

❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20+

Türkisch

❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20+

III. Sprachgebrauch

Bitte beantworten Sie die folgenden Fragen zu Ihrer Verwendung von Sprachen, indem Sie das entsprechende Kästchen ankreuzen. Die Verwendung aller Sprachen zusammen sollte in jeder Frage 100% ergeben.

7. Zu welchem Anteil verwenden Sie in einer durchschnittlichen Woche die folgenden Sprachen **mit Freunden**?

	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
Deutsch	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑
Türkisch	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑
Andere Sprachen	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑

8. Zu welchem Anteil verwenden Sie in einer durchschnittlichen Woche die folgenden Sprachen **mit der Familie**?

	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
Deutsch	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑
Türkisch	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑
Andere Sprachen	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑

9. Zu welchem Anteil verwenden Sie in einer durchschnittlichen Woche die folgenden Sprachen **in der Uni/bei der Arbeit**?

	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
Deutsch	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑
Türkisch	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑
Andere Sprachen	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑

10. Zu welchem Anteil verwenden Sie die folgenden Sprachen **in Selbstgesprächen**?

	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
Deutsch	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑
Türkisch	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑
Andere Sprachen	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑

11. Zu welchem Anteil **zählen** Sie in den folgenden Sprachen?

	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
Deutsch	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑
Türkisch	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑
Andere Sprachen	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑	❑

IV. Sprachkenntnisse

*Bitte bewerten Sie Ihre Sprachkenntnisse auf einer Skala von 0 bis 6 (**0 = gar nicht gut, 6 = sehr gut**).*

12. Wie gut **sprechen** Sie...

	❑	❑	❑	❑	❑	❑	❑
a. … **Deutsch**?	0	1	2	3	4	5	6
b. … **Türkisch**?	0	1	2	3	4	5	6

13. Wie gut **verstehen** Sie gesprochenes…

	❑	❑	❑	❑	❑	❑	❑
a. … **Deutsch**?	0	1	2	3	4	5	6
b. … **Türkisch**?	0	1	2	3	4	5	6

14. Wie gut können Sie Texte in folgenden Sprachen **lesen**?

	❑	❑	❑	❑	❑	❑	❑
a. **Deutsch**	0	1	2	3	4	5	6
b. **Türkisch**	0	1	2	3	4	5	6

15. Wie gut können Sie Texte in folgenden Sprachen **schreiben**?

	❑	❑	❑	❑	❑	❑	❑
a. **Deutsch**	0	1	2	3	4	5	6
b. **Türkisch**	0	1	2	3	4	5	6

V. Spracheinstellungen

Bitte bewerten Sie die folgenden Aussagen auf einer Skala von 0 bis 6 (0 = stimme gar nicht zu, 6 = stimme völlig zu).

16. Ich fühle mich wie ich selbst, wenn ich…

a.	… **Deutsch spreche.**	❑ 0	❑ 1	❑ 2	❑ 3	❑ 4	❑ 5	❑ 6
b.	… **Türkisch spreche.**	❑ 0	❑ 1	❑ 2	❑ 3	❑ 4	❑ 5	❑ 6

17. Ich identifiziere mich mit der...

a.	… **deutschsprachigen** Kultur.	❑ 0	❑ 1	❑ 2	❑ 3	❑ 4	❑ 5	❑ 6
b.	… **türkischsprachigen** Kultur.	❑ 0	❑ 1	❑ 2	❑ 3	❑ 4	❑ 5	❑ 6

18. Es ist mir wichtig **wie ein Muttersprachler**…

a.	… **Deutsch** zu verwenden.	❑ 0	❑ 1	❑ 2	❑ 3	❑ 4	❑ 5	❑ 6
b.	… **Türkisch** zu verwenden.	❑ 0	❑ 1	❑ 2	❑ 3	❑ 4	❑ 5	❑ 6

19. Ich möchte, dass **andere denken**, ich sei ein…

a.	… **deutscher** Muttersprachler.	❑ 0	❑ 1	❑ 2	❑ 3	❑ 4	❑ 5	❑ 6
b.	… **türkischer** Muttersprachler.	❑ 0	❑ 1	❑ 2	❑ 3	❑ 4	❑ 5	❑ 6

Appendix C.5 – Motivation and attitude questionnaire (adapted from Gardner, 1985)

Du wirst jetzt einige Aussagen lesen. Bitte kreuze für jede Aussage an, wie sehr du ihr zustimmst.

1 = Dieser Aussage stimme ich gar nicht zu.
2 = Dieser Aussage stimme ich kaum zu.
3 = Dieser Aussage stimme ich ein bisschen zu.
4 = Dieser Aussage stimme zu.
5 = Dieser Aussage stimme ich sehr zu.
6 = Dieser Aussage stimme ich vollkommen zu.

Einige der Sätze klingen sehr ähnlich. Bitte lies sie trotzdem genau und beantworte sie sorgfältig.

Es ist wichtig für jeden Menschen in Deutschland, fließend Deutsch zu sprechen	1	2	3	4	5	6
Es ist wichtig, dass Menschen, die in Deutschland leben, so oft wie möglich Deutsch sprechen.	1	2	3	4	5	6
Ich würde meine Zeit lieber mit anderen Schulfächern verbringen als mit Englisch.	1	2	3	4	5	6
Ich schaue ausländische Filme lieber in der Originalsprache an als in der synchronisierten Fassung.	1	2	3	4	5	6
Meine Eltern versuchen, mir mit Englisch zu helfen.	1	2	3	4	5	6
Es macht mir wirklich viel Spaß, Englisch zu lernen.	1	2	3	4	5	6
Ich bin stolz darauf, Türkisch zu sprechen.[88]	1	2	3	4	5	6
Meine Eltern denken, dass ich so lange wie möglich Englisch lernen sollte.	1	2	3	4	5	6

[88] Statements in italics were only included in the version for the bilingual participants.

Meine Eltern denken, dass ich mehr Zeit für Englisch aufbringen sollte.	1	2	3	4	5	6
Wenn es möglich wäre, würde ich nur Türkisch sprechen.	1	2	3	4	5	6
Englisch zu lernen ist eine Zeitverschwendung.	1	2	3	4	5	6
Ich schäme mich, wenn ich in der Englischstunde Antworten geben muss.	1	2	3	4	5	6
Wenn ich einen Aufenthalt in einem anderen Land planen würde, würde ich mich sehr anstrengen, die Sprache zu lernen, auch wenn ich dort mit *Türkisch*/Deutsch zurechtkäme.	1	2	3	4	5	6
Ich werde nervös, wenn ich im Englischunterricht spreche.	1	2	3	4	5	6
Ich wünsche mir oft, Bücher in einer anderen Sprache lesen zu können.	1	2	3	4	5	6
Eine Fremdsprache zu lernen ist eine schöne Erfahrung.	1	2	3	4	5	6
Wenn es möglich wäre, würde ich nur Deutsch sprechen.	1	2	3	4	5	6
Meine Eltern haben betont, wie wichtig Englisch für mich nach der Schule sein wird.	1	2	3	4	5	6
Ich spreche Deutsch, so oft ich kann.	1	2	3	4	5	6
Ich finde, dass Englisch lernen langweilig ist.	1	2	3	4	5	6
Türken sind sehr freundlich, warmherzig und kreativ.	1	2	3	4	5	6
Ich liebe es, Englisch zu lernen.	1	2	3	4	5	6
Türken sind vertrauenswürdig und verlässlich.	1	2	3	4	5	6
Ich finde es schön, Menschen zu treffen, die eine andere Sprache sprechen, und mich mit ihnen zu unterhalten.	1	2	3	4	5	6
Meine Eltern denken, dass es gerade in unserer modernen Welt wichtig für mich ist, Englisch zu lernen.	1	2	3	4	5	6
Ich wünscht, ich könnte noch eine weitere Sprache fließend sprechen.	1	2	3	4	5	6

Meine Eltern denken, dass ich wirklich versuchen sollte, Englisch zu lernen.	1	2	3	4	5	6
Englisch ist ein wichtiger Teil der Schule.	1	2	3	4	5	6
Ich würde auch dann eine Fremdsprache an der Schule lernen, wenn es nicht vorgeschrieben wäre.	1	2	3	4	5	6
Meine Eltern drängen darauf, dass ich meine/n Lehrer/in um Hilfe bitte, wenn ich in Englisch etwas nicht verstehe.	1	2	3	4	5	6
Es ist wichtig für Menschen in Deutschland, eine Fremdsprache zu lernen.	1	2	3	4	5	6
Deutsche sind vertrauenswürdig und verlässlich.	1	2	3	4	5	6
Wenn ich ein fremdes Land besuchen würde, würde ich gerne die Landessprache sprechen.	1	2	3	4	5	6
Ich erzähle anderen nicht gerne, dass ich Türkisch spreche.	1	2	3	4	5	6
Ich bin stolz darauf, Türke zu sein.	1	2	3	4	5	6
Türken in Deutschland sollten sich sehr bemühen, die deutsche Sprache zu lernen.	1	2	3	4	5	6
Ich spreche Türkisch so oft ich kann.	1	2	3	4	5	6
Es ist wichtig, dass Türken in Deutschland ihre Sprache bewahren.	1	2	3	4	5	6
Ich habe Angst, dass meine Mitschüler mich auslachen, wenn ich Englisch spreche.	1	2	3	4	5	6
Meine Eltern interessieren sich sehr für alles, was mit meinem Englischunterricht zu tun hat.	1	2	3	4	5	6
Ich spreche gerne Türkisch.	1	2	3	4	5	6
Wenn ich mit der Schule fertig bin, werde ich Englisch komplett aufgeben, weil es mich nicht interessiert.	1	2	3	4	5	6
Es ist wichtig, dass Türken in Deutschland ihre Kultur/Bräuche bewahren.	1	2	3	4	5	6
Ich bin mir nie ganz sicher, wenn ich in der Englischstunde spreche.	1	2	3	4	5	6
Meine Eltern ermuntern mich sehr dazu, Englisch zu lernen.	1	2	3	4	5	6

Ich habe vor, so viel Englisch zu lernen wie möglich.	1	2	3	4	5	6
Deutsche sind sehr freundlich, warmherzig und kreativ.	1	2	3	4	5	6
Jedes Kind mit türkischer Herkuft sollte fließend Türkisch sprechen.	1	2	3	4	5	6
Englisch zu lernen ist wirklich großartig.	1	2	3	4	5	6
Ich habe immer das Gefühl, dass meine Mitschüler besser Englisch sprechen als ich.	1	2	3	4	5	6
Ich spreche gerne Deutsch.	1	2	3	4	5	6
Ich hasse Englisch.	1	2	3	4	5	6
Die meisten Türken sind so freundlich und angenehm, dass Deutschland sich glücklich schätzen kann, sie zu haben.	1	2	3	4	5	6
Ich würde gerne wirklich viele Fremdsprachen lernen.	1	2	3	4	5	6
Meine Eltern ermuntern mich dazu, so viel wie möglich Englisch zu sprechen.	1	2	3	4	5	6

Appendix D: Reading texts for Study 2

Appendix D.1 – English reading text

One day Catherine and Rose, two recent Caltech graduates who had been friends ever since they had been little children, decided to go on a three week hiking trip to the Laprig hills together. They met at Rose's place to make travel plans: to book their hotel room, decide on various places they wanted to visit for their breakfasts, lunches and dinners and find out about the rangers' warnings they had to heed. Rose had prepared the most wonderful meal - potato soup, cooked plums with root vegetables and cranberry sauce, and apple ice cream with sunflower seeds for dessert - and she had put a full bottle of expensive red wine into the fridge. Rose was extremely excited about their holiday and wanted this evening of planning to be perfect. They used to go to the theater together every Thursday, but the two young women hadn't really spent quality time with each other for the last three years since they both had extremely busy schedules. At six o'clock sharp, Catherine rang the doorbell, and Rose opened her front door. After they had shooed away a rook from the front porch, the two friends hugged happily and went inside, where Catherine hung up her bag on the coat rack. Rose asked her to take a seat. Doing as she had been bid, Catherine sat down at the clean wooden table, and Rose went into the kitchen to fetch drinks. While Catherine was sitting in the living room on her own contemplating the neat yellow pattern on Rose's wallpaper, she heard the strangest noise. It sounded as if a panther cub was trying to roar, and she could not place it anywhere: It might have originated outside the window just as well as right under the plush sofa that she could just see from where she was sitting. For a moment, Catherine was worried, but then Rose came back with two crystal glasses and the wine, so Catherine forgot about the incident. When the wine had been poured, and some candles had been lit, the two women started chatting about various things that had happened in the preceding week: Rose's pet hamster had caught the flu, her milkman had turned vegan and thus quit his job to become a roadside pumpkin salesman, and she had planted new angel wing begonias in her garden, which already had fresh shoots, and which she said she was looking forward to showing Catherine once they had started to bloom. Catherine's week had been equally eventful: On Wednesday, she had left the house to find that her car had probably been re-varnished overnight and that it now was a bright shade of green! She had briefly contemplated calling the police but then decided that she actually quite liked the new colour and that she was going to accept this unexpected, unusual present. At that point of Catherine's story, there was a "bing" from the kitchen, and Rose got up to get the food that had been warming up on the stove. She put it down on the table along with plates and the relevant authentic cutlery: asparagus forks, table spoons, as well as butter knives. After she had served her friend, she put some food on her own plate and started to eat. Catherine followed her lead. “Wow, this could feed an army,” she exclaimed. Catherine enjoyed the food Rose had prepared lovingly, and after only a short time they had finished everything, including dessert. The two friends cleared away together before they sat down on the sofa. Rose had already laid out the computer, the reading lamp and various maps and leaflets, as well as a list of tips written up by a cool neighbour. Noticing this, Catherine sighed heavily. “I think I should tell you something now,” she said. Rose looked at her curiously. “Is there a problem?” she asked. “Well, not really. Do you remember Frank?” Rose nodded. “We’ve been dating for a couple of months now.” At that point, Rose furrowed her brows in disbelief. “Why didn’t you tell me?” she asked quietly. “I’m your friend. I need to know these things! You should have kept me in the loop!” Catherine didn’t know how to respond to that. She tried to put her hand on Rose’s shoulder, but Rose brushed her off forcefully. “You know, Rose, we were both so busy with our exams, and with everything that was going on that I didn’t feel I wanted to bother you with this. I didn’t know that it would become serious. Up until last week, it might have just gone away, it was just a crush.” Rose looked surprised. “Why? What happened last week?” Again, Catherine smiled. “Well, last week I found out that I am pregnant!” There was total silence. Rose stared at her friend blankly and bit her lip. Catherine’s smile took a while to fade, but fade it did. “Rose, this is also why I need to talk to you now rather than later. I don’t really want to go away for as long as we had planned, especially not to hike. I’d rather just make it a weekend and maybe do some sightseeing and chatting by the pool instead of walking around in the heat for days.” Rose got up and started walking towards the kitchen again. Just before she would have left the room, she turned around and threw her wine glass to the floor. “You can’t be serious! This cannot be true!” she yelled angrily, her eyes seemingly aflame, “How could you do this to me? First you lie to me for months, and now you are cancelling our holiday – the one that we both have been looking forward to for years! Cat, we haven’t really talked to each other since we started studying. I feel like I barely know you anymore! Are you even fit to have a baby?” Then she burst into tears. Catherine hesitated for a second, but then got up and walked up to Rose. She tried to look into Rose’s eyes, but Rose evaded her friend’s gaze. When Catherine tried to draw her close to her and hug her, Rose pushed her away violently, and Catherine crashed to the floor and hit her head on the sofa. She felt the back of her head for blood, but it was dry. And then it happened again: that strange noise that Catherine had heard earlier, and that she had already forgotten about. And this time, it didn’t go away but persisted. “What is that?” asked Catherine weakly. “I don’t know,” whispered Rose, “but it sounds like an extremely high pitched lawnmower. Hm, this doesn't make sense. Why would there be a lawnmower in my living room?" Catherine crawled towards Rose, got up, her heart beating fast. She looked around in agony and fear. "I think it's moving. It's moving towards us, Rose." The two women moved closer together. Rose gripped Catherine's shoulder. "Oh, this is bad! What do we do?" she asked. "Let's hide!" was Catherine's short reply. Catherine got back on her feet, and they walked slowly towards a wood panel door that led to the basement. Rose reached for the handle and opened the door. "The light is broken," she whispered. Catherine merely nodded. They walked through the door onto the stairwell that lay right behind it, and Rose locked the door behind them. It smelled musky and damp, and everything was pitch black. The noise was still there, but it seemed to be stationary now. The two friends were holding on to each other in terror. Death seemed to be looming. After some minutes had passed, Rose clasped Catherine's hand and whispered: "Cat, I am so sorry. I was unfair and cruel. Of course I am happy for you. And of course we can make it a weekend instead of that long hike. If we make it out of this trap alive, that is…" An hour passed. Catherine and Rose barely spoke during that time, getting more and more frightened. Beads of sweat began rolling down their faces. Suddenly the doorbell rang. The women didn't move. The doorbell rang again, followed by frantic knocking. Finally, there was the sound of a key turning in the lock. Rose was shivering, Catherine started to cry again. Now, certainly, all was lost. And that poor unborn child! They heard footsteps coming closer. The handle moved. Both women closed their eyes and held their breaths. They knew that it was over. "Rose? Cat? What are you doing?" asked a familiar female voice. Rose opened her eyes and found herself looking into her mother's puzzled face. "Oh, mum! Thank god it's you! But… do you hear this noise? Something is wrong. And it was following us around earlier. Something was moving INSIDE the house! That's why we hid." Rose's mother stepped back and listened. "Calm down, you fools! Do you mean this whizzing, buzzing noise?" she asked. Rose nodded, and her mother started to laugh. "It sounds a lot like your old toy monkey, the pretty one that plays the drums. I brought it here in a box last week, along with some other toys that I had found in the attic and that I didn't just want to get rid of. Remember? I think you put it under the sofa and said you were going to have a look once you find time." She walked into the living room and towards the sofa, where she knelt down and pulled a cardboard box out from underneath it. Catherine and Rose followed her. Rose's mum lifted the lid off the box, and the three women looked inside. There, lying on the top of a pile of books, knitting needles, stuffed animals, and dolls, was the monkey, drumming away. Catherine stared at Rose, then they both started to laugh. "Rose, is that the monkey we used to play with when we were children?" Rose nodded: "Yes, I’d forgotten all about the box.” She picked up the monkey and then looked at Catherine. “This used to be so precious to me! You should have it, for the baby. From her aunt Rose, her mum’s best friend!” Catherine smiled, took the monkey and hugged Rose. “Thank you. Let’s never fight again!”

Appendix D.2 – Der Nordwind und die Sonne

Einst stritten sich Nordwind und Sonne, wer von ihnen beiden wohl der Stärkere wäre, als ein Wanderer, der in einen warmen Mantel gehüllt war, des Weges daherkam. Sie wurden einig, dass derjenige für den Stärkeren gelten sollte, der den Wanderer zwingen würde, seinen Mantel abzunehmen. Der Nordwind blies mit aller Macht, aber je mehr er blies, desto fester hüllte sich der Wanderer in seinen Mantel ein. Endlich gab der Nordwind den Kampf auf. Nun erwärmte die Sonne die Luft mit ihren freundlichen Strahlen, und schon nach wenigen Augenblicken zog der Wanderer seinen Mantel aus. Da musste der Nordwind zugeben, dass die Sonne von ihnen beiden der Stärkere war.

Appendix D.3 - Kuzey rüzgarı ve Güneş

Kuzey rüzgarı ve Güneş hangimiz daha güçlü diye tartışıyorlarmış. Tam o sırada kalın paltolu bir yolcu çıkagelmiş. Kim yolcunun paltosunu çıkartmasına ilk önce neden olursa, onun diğerinden daha güçlü olduğunun kanıtlanacağına karar vermişler. Kuzey Rüzgarı olabildiğince esmiş, ama o estikçe yolcu paltosuna daha çok sarınmış. En sonunda Kuzey Rüzgarı vazgeçmiş. Güneş sıcak sıcak parlamaya başlamış, ve yolcu hemen paltosunu çıkarıvermiş. Böylece Kuzey Rüzgarı, Güneşin daha güçlü olduğunu itiraf etmek zorunda kalmış.

Appendix E: Interview guiding questions for Study 2

Domain	Questions[89]
Personal and family	Where and when were you born? Did you grow up with your parents? Where were your parents born? Did you grow up with any siblings? If so, are they older or younger than you are? *Were your parents born in Turkey? Where in Turkey?* *Can you tell me about the place your family comes from?*
Language	What languages did you speak in your childhood? *When did you start speaking German?* *Which of the languages is more important to you?*
Traveling	Have you spent time abroad? Would you like to live abroad at some point in your life? If so, where would you like to live and why? *Do you regularly spend time in Turkey?* *Would you like to move to Turkey?*
Culture and identity	*Do you sometimes feel like you are living in two cultures? How does this show?* *Do you have more Turkish or more German friends?*
Studies	Why did you decide to study English? Are you going to become a teacher? Was English your favorite subject in school? Do you speak English outside university?

[89] Questions in italics were only given to the bilingual participants.

Appendix F: C-tests in German and Turkish

Turkish text:

Mehmet cuma akşamı işten döndü. Maaş dağıtım günü olmasına rağmen, s__________(1) bile değildi. Oturup faturaları ödedikten, k__________(2) parasını ayırdıktan, arabasının yakıt m__________(3) hesapladıktan ve banka tasarruf hesabına a__________(4) bir para koyduktan sonra, rahat b__________(5) yaşam için geriye bir şey kalmayacağını ç__________(6) iyi biliyordu.

Sevdiği lokantaya gitmeyi i__________(7), fakat keyfi yerinde değildi. Dairesinde b__________(8) süre dolaştı ve bir sandviç y__________(9). Kafası, kısa bir süre maddi k__________(10) takıldı. Arabasını alıp, geziye çıkmaya k__________(11) verdi. Belli bir istikameti olmamasına r__________(12), ikamet ettiği yerden uzaklaşmayı arzu e__________(13). Sakin bir köy yoluna saptı. m__________(14) güzelliği moralini düzeltti. Arabayı kullanırken, d__________(15) daldı ve kendisinin arsa sahibi o__________(16), maddi durumunun düzeldiğini göz önüne g__________(17). Mehmet bu hayali uzun z__________(18) kurmaktaydı; fakat, hiçbir zaman o__________(19) gerçekleştirmek için adım atmamıştı. Hayal kurmaya devam etti. Kafasında, kendi mahsulünü t__________(20) olumlu ve olumsuz taraflarını tartışıyordu. e__________(21) çatısındaki güneş enerji sistemi ile h__________(22) evini hem de suyunu ısıtabileceğini tasavvur e__________(23). Aynı zamanda, sebze dolu tarlaları o__________(24), ve sebzeleri konserve edip, kışları y__________(25) imkanını göz önüne getiriyordu. Mahsul ç__________(26) olduğu takdirde, fazlasını satıp, ek g__________(27) ile tarım makineleri alabileceğini düşündü.

S__________(28) Mehmet hayal dünyasından çıktı ve y__________(29) sesle güldü. Acaba bu düşünceleri g__________(30) gerçekleştirebilir miyim diye düşündü.

German texts:

Text 1:

Die Geschichte der Kernspaltung reicht zurück in das frühe 19. Jahrhundert. In d____________________ Folgejahren leg____________________ Chemiker d____________________ Grundstein f____________________ den mode____________________ Atombegriff. S____________________ erkannten, da____________________ die chemi____________________ Elemente a____________________ Teilchen aufg____________________ sind, d____________________ untereinander völ____________________ gleichartig reag____________________, sich jed____________________ von and____________________ Elementen unters____________________. 1871 erschien d____________________ erste tabell____________________ Aufstellung d____________________ Eigenschaften al____________________ bekannten Elem____________________, das Periode____________________.

Text 2:

Eine Wünschelrute ist ein gegabelter Zweig, ursprünglich meist vom Haselnussstrauch, später verwe____________________ man au____________________ ähnliche Instr____________________ aus untersch____________________ Materialien. S____________________ dient d____________________ so gena____________________ Rutengänger, ei____________________ Person, d____________________ für si____________________ eine beso____________________ Begabung bean____________________t, als Hilfs____________________ zum Auff____________________ von unterir____________________ »Reizzonen«, z____________________ Beispiel Wasse____________________, Erdölvorkommen od____________________ Erzlagerstätten.

Text 3:

Sicherheitshinweise

Bedienungsanleitung bitte vollständig vor Inbetriebnahme des Bügeleisens durchlesen und aufbewahren.

Reparaturen an Elektro____________________ dürfen n____________________ von Fachk____________________ durchgeführt wer____________________. Durch unsach____________________ Reparaturen kön____________________ erhebliche Gefa____________________ für d____________________ Benutzer entst____________________.

Wird d____________________ Gerät zwecken____________________ oder fal____________________ bedient, ka____________________ keine Haf____________________ für dad____________________ verursachte Sch____________________ übernommen wer____________________.

Das Ge____________________ wurde v____________________ uns sicherheitstechnisch geprüft.

Text 4:

Schon in ältester Zeit haben die Menschen den Himmel beobachtet. Je stä____________________ frühe Kult____________________ von d____________________ Natur abhä____________________ waren, de____________________ näher l____________________ es f____________________ sie, a____________________ den o____________________ periodischen Ersche____________________ der Na____________________ und d____________________ Sternenhimmels besti____________________ Faktoren abzul____________________, die i____________________ tägliches Le____________________ beeinflussten.

I____________________ Verlauf d____________________ Entwicklung d____________________ mensch____________________ Zivilisation verl____________________ diese natür____________________ Zyklen im____________________ mehr a____________________ Bedeutung.

Appendix G: BLP manual, excerpt (scoring system)

Each question response in the BLP is scalar and associated with a certain point value. Scores that can be calculated include:

- Language particular score for each module
- Global score for each language
- Composite dominance index

Step 1 : Module Scores

Within each module, tally the point totals for each language separately based on the following guidelines. This will yield the language particular score for each module.

Language History
- 6 questions: each worth between 0 and 20
- Each item is worth the numerical value given in the response, with three exceptions:
    - The first two items are scored in the reverse: A "20" response is worth 0, a "19" is worth 1, and so on
    - Phrasal responses "Since birth" and "For as long as I can remember" are worth 20 points
    - "Not yet" is worth 0 points

Language Use
- 5 questions: each worth between 0 and 10
- Each item is worth the numerical value given in the response

Language Proficiency
- 4 questions: each worth between 0 and 6
- Each item is worth the numerical value given in the response

Language Attitudes
- 4 questions: each worth between 0 and 6
- Each item is worth the numerical value given in the response

Step 2 : Global Language Scores

To ensure that each module receives equal weighting in the global language score, multiply the score for each module (each language calculated separately) by a factor of:

Language History	0.454
Language Use	1.09
Language Proficiency	2.27
Language Attitudes	2.27

Add the new module totals together to yield a global score for each language. Total points possible is 218.

Step 3 : Dominance Score

To obtain the language dominance index, subtract one language total from the other to render a dominance score that ranges from -218 to +218. A score near zero indicates balanced bilingualism and more positive or more negative scores reflect respective language dominance.

(Taken from "Scoring and interpreting the results," 2019)