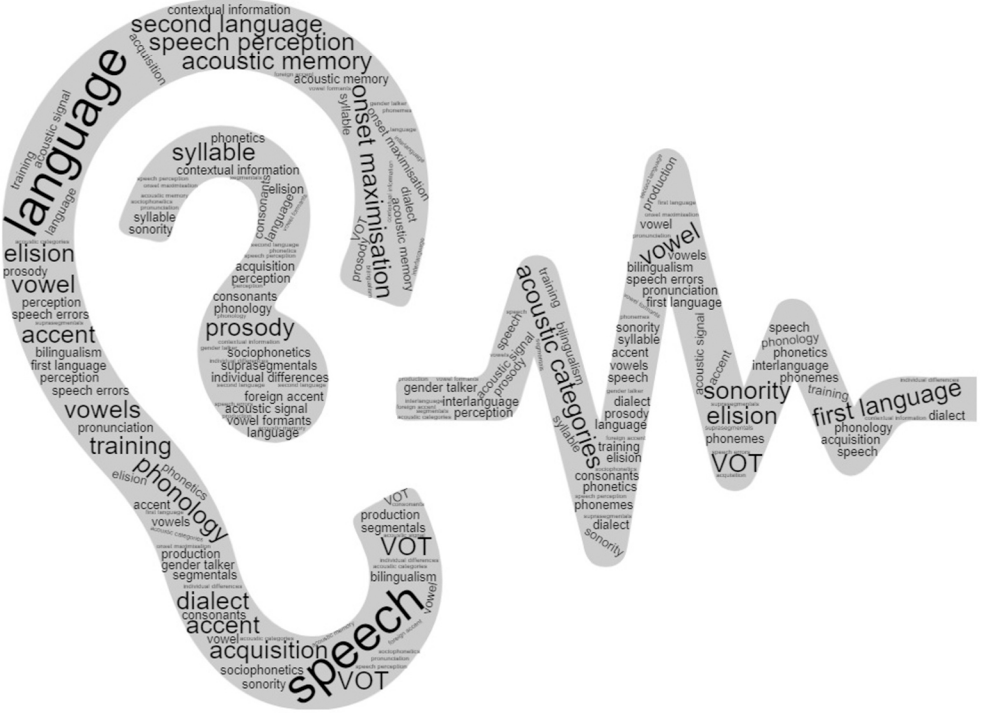
Speech Perception and Production in L2



Edited by

Elena Kkese

Speech Perception and Production in L2

Speech Perception and Production in L2

Edited by

Elena Kkese

Cambridge
Scholars
Publishing

Speech Perception and Production in L2

Edited by Elena Kkese

This book first published 2022

Cambridge Scholars Publishing

Lady Stephenson Library, Newcastle upon Tyne, NE6 2PA, UK

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN (10): 1-5275-8144-6
ISBN (13): 978-1-5275-8144-9

TABLE OF CONTENTS

Contributors vii

Introduction xiii
Elena Kkese

Part I: Theoretical Perspectives of Speech Perception and Production in L2

Chapter One 2
Research-to-Practice: Language Science as a Guide to Structured Literacy, Speech Perception, and Production
Sandie Barrie Blackley

Chapter Two 20
Discussing Bilingualism based on Phonetic Elements Attested in Speech Production
Niki Efstathopoulou

Chapter Three 34
Facilitation of L2 Instruction with Data from Auditory-Visual Speech Perception Research
Doğu Erdener

Chapter Four 56
The Arcadian Greek Syllable
Photini Coutsougera

Chapter Five 78
Speech Errors Greeks make in English: A Tutorial
Elena Babatsouli

Part II: Practical Perspectives of Speech Perception and Production in L2

Chapter Six 106
Realizations of /d/ in Spanish Past Participle -ar Verbs: A Sociophonetic Analysis between Native Speakers and L2 Learners
Giulia Cortiana and Olga Tararova

Chapter Seven..... 125
Perceptual Training Effects on the Perception and Production of L2 English Vowels by L1 Greek-Cypriot Learners
Dimitra Dimitriou

Chapter Eight..... 152
Effect of Talker Gender Information on Speech Perception in L2 English
Elena Kkese and Sviatlana Karpava

Chapter Nine..... 173
L2 Phone Perception on a Syllabic Level might be Driven by L1 Transfer Mechanisms and/or Universal Patterns of Phone Acquisition
Elena Kkese and George Georgiou

Chapter Ten 199
Factors Affecting Accent Acquisition among Female American Immigrants to Israel
Efrat Last, Shira Matalon, Nitzan Bechar Marom, Noa Vittenberg, Sapir Shemesh and Noam Amir

Chapter Eleven 236
The Role of Acoustic Memory and Phonological Short-Term Memory in Non-Native Perception: Which Memory Matters More and When?
Vita V. Kogan

CONTRIBUTORS

Editor

Elena KKESE obtained her first degree in English and was awarded both MA and PhD degrees in linguistics from the University of Cyprus (BA in English Language and Literature; MA in Applied Linguistics; PhD in Linguistics). She has been teaching in secondary and tertiary level education since 2004. Her experience involves teaching courses in phonetics, phonology, sociolinguistics, clinical linguistics and language pathology, contrastive linguistics, methodology in TESOL (Teaching English to Speakers of Other Languages), morphology, general linguistics as well as general English and academic writing courses. Elena's research interests include the L2 perception and production difficulties emphasising on phonology, spelling, and writing. She is the author of *Identifying Plosives in L2 English: The Case of L1 Cypriot Greek Speakers* (2016) and of *L2 Writing Assessment: The Neglected Skill of Spelling* (2020), and has published numerous journal articles and book chapters on the areas of L2 phonology, spelling, and writing.

Contributors

Noam AMIR completed his doctoral studies in electrical engineering, on the topic of acoustic wave propagation in waveguides with varying cross section, better known as trumpets. After completing a postdoc in the Laboratoire d'Acoustique of Universite du Maine, in Le Mans, France, in 2000 he became a full-time staff member of the Department of Communication Disorders, at Tel Aviv University. For the past 30 years Dr. Amir has been conducting research at the juncture of physics, perception and production of speech and music. His research spans a diverse range of subjects such as segmental properties of Hebrew and Arabic, linguistic and affective prosody, accent acquisition, characteristics of the singing voice, psychoacoustic abilities of musicians and laymen and more. Dr. Amir has published more than 40 full length papers in the scientific literature and has participated in numerous conferences and workshops. He has four children who are all pursuing wonderfully impractical careers as professional musicians. Dr. Amir practices meditation regularly and is an avid photographer.

Elena BABATSOULI is an Assistant Professor and the Ben Blanco/BORSF Endowed Professor in Communicative Disorders in the College of Liberal Arts at the University of Louisiana at Lafayette, an editor of the *Journal of Monolingual and Bilingual Speech* (Equinox), and founder of the International Symposium on Monolingual and Bilingual Speech. She has a BA in English (Royal Holloway, University of London), an MA in Languages and Business (London Southbank University), and a PhD in Linguistics (University of Crete). Her research interests are in child monolingual and bilingual acquisition, atypical speech (SSDs), SLA, culturally responsive practices in speech pathology, phonetics, phonology, psycholinguistics, assessment measures and quantitative methods. She has over 30 peer-reviewed publications, and she has edited five books, two journal special issues, and three conference proceedings.

Sandie BARRIE BLACKLEY is a speech-language pathologist, a Fellow of the American Speech-Language-Hearing Association, past president of the North Carolina Speech, Hearing, Language Association and recipient of their two clinical awards, the Public Service Award and Clinical Services Award. She also served two terms on the North Carolina Board of Examiners for Speech-Language Pathologists & Audiologists. Sandie taught structured literacy courses and supervised practicum for speech-language pathology graduate students at the University of North Carolina Greensboro, where she coordinated a five year, federally funded personnel preparation grant. In 2009, Sandie and her business partner, Chad Myers co-founded Mind InFormation, Inc. and Lexercise, a teletherapy platform that provides structured literacy services for thousands of students across the English-speaking world.

Giulia CORTIANA is a fourth-year Ph.D. candidate, who specializes in phonetics and acoustic analyses, in Hispanic Linguistics at Western University, London Ontario, Canada. She holds a BA and a MA from the University of Bergamo, Italy, and a MA in Hispanic Studies from the University of British Columbia, Vancouver, Canada. Giulia currently works as Teaching Assistant at the Department of Modern Languages and Cultures, as well as a Research Assistant for several ongoing studies at the Brain and Mind Institute, at Western University. Giulia's Ph.D. dissertation focuses on third language acquisition and the effect of orthography, by including both languages with a Romanized alphabet, such as English, Italian, and Spanish, as well as languages with a non-Romanized alphabet, such as Farsi and Japanese. Giulia has presented her findings both at national and international conferences, and she has also published several articles in renowned journals.

Photini COUTSOUGERA received her PhD from the School of Linguistics & Applied Language Studies, University of Reading, UK in 2002. She held sessional lectureship and teaching assistant appointments at the University of Reading, UK (1997-2004), a temporary lectureship appointment at the University of Oxford, UK (2002-2004), a tenure track lectureship appointment at the University of Cyprus (2004-2011), and is currently an Assistant Professor, Teaching Stream, LTA at the University of Toronto. Her research interests lie in the areas of the phonology of Greek dialects, L1 acquisition and lexicography.

Dimitra DIMITRIOU graduated from Lancaster University, where she completed her BA (Hons) in Linguistics and English Language (2014) and her MA in English Language and Literary Studies (2015). She is currently a PhD student in the School of Language and Global Studies at the University of Central Lancashire, researching Greek-Cypriot learners' perception and production of L2 English vowels. She is also an Associate Lecturer in the School of Business and Management at the University of Central Lancashire, teaching English phonology and morphology as well as Academic Writing. Her research interests include L2 phonetic and phonological acquisition, speech production and perception, and EFL teaching and learning. In combination with her academic work, she has been working as a Business Support Officer at a Court since 2015, while also engaging in translation, proofreading and editing, and teaching.

Pagona-Niki EFSTATHOPOULOU is a Greek teacher in the Model Junior High School of Patras, a school affiliated with the University of Patras. She worked as a researcher in Simon Fraser University, Canada, where she completed her MA. Her MA Thesis centers on how bilingualism affected the speech of Greek/English bilinguals in Vancouver area, Canada by studying their voiceless stops in both languages and their accented speech in English. Also, she has taught "Phonetics" and "Bilingualism" in the Department of Speech and Language Therapy in the University of Patras. She teaches for the past 15 years, while she specializes in teaching Greek as a second language to foreigners and immigrants. Her interests include phonetics and bilingualism.

Doğu ERDENER is an associate professor in psychology. He completed his PhD thesis on the development of auditory-visual speech perception in English-speaking children with Professor Denis Burnham at MARCS Institute for Brain, Behaviour and Development at Western Sydney University in Australia. During his graduate studies at WSU, he also worked as a tutor and guest lecturer at the School of Psychology. Following a

temporary research officer position at Charles Sturt University, he took up a faculty post at Middle East Technical University's Cyprus campus where he is currently the head of the psychology program. His research interests revolve around auditory-visual speech processing and its applications. After his doctoral work on the developmental aspects, he became interested in applied aspects particularly in second language acquisition domain. More recently, he has been working on how auditory and visual sources of speech are integrated in clinical populations, specifically in individuals with bipolar disorder.

George GEORGIOU is a faculty member at the Department of Languages and Literature of the University of Nicosia and he is the coordinator of the Cyprus Linguistics and the Humanities Research Group. He has previously worked as a Postdoctoral Researcher and Assistant Lecturer in Linguistics at RUDN University, as a Postdoctoral Researcher in Speech-Language Pathology at the Cyprus University of Technology, and as an Associate Lecturer in Linguistics at the University of Central Lancashire. He has also served as the Head of RUDN University Phonetic Lab and Principal Investigator on speech acquisition projects. He has published more than 22 papers in high-impact international standing journals. He has been awarded two times the Research Recognition Award by the University of Nicosia and he was won three national and international research grants.

Sviatlana KARPAVA is a Lecturer in Applied Linguistics/TESOL and Linguistics Section Coordinator (2020-present) at the Department of English Studies, University of Cyprus. She is also a Co-coordinator of the Testing, Teaching and Translation Lab (2019-present). She is a Management Committee Member and WG5 Co-Leader of the "European Family Support Network Cost Action. A bottom-up, evidence-based and multidisciplinary approach" (2019-2023). She is the Principal Investigator of the two internally funded (University of Cyprus) research projects: MIGDISCY-Migration, Discourse, Cyprus and Heritage Language Maintenance, Variation, and Change (2021-2023). Dr Karpava is the author of the book: Vulnerable Domains for Cross-Linguistic Influence in L2 Acquisition of Greek (2015) Frankfurt am Main: Peter Lang. She has also presented her research at numerous international conferences and published her research work in various peer-reviewed journals. Her area of research is applied linguistics, syntax, semantics and pragmatics, phonetics and phonology, speech perception and production, orthography, first and second language acquisition, bilingualism, multilingualism, sociolinguistics, teaching and education, critical digital literacy; heritage language use, maintenance and

transmission, language loss, shift and attrition, family language policy, home literacy environment and intercultural communication.

Vita KOGAN is a Lecturer of Linguistics at Queen Mary University of London, where she teaches modules on phonetics and bilingualism. Previously Vita served as an Assistant Professor at the Defense Language Institute and coordinated and taught language programs at the Middlebury Institute of International Studies, Concordia College, and the University of Pittsburgh (USA). Vita has received her Ph.D. from the University of Barcelona and M.Sc. from the University of Edinburgh. Vita's research concerns the cognitive underpinnings of second language acquisition – in particular, individual differences in the acquisition of L2 phonetics and phonology. Her latest project investigates the contribution of phonological and acoustic memories to the task of L2 speech perception and how these cognitive abilities contribute differently at the various stages of the acquisitional trajectory. As an active language teacher, Vita is also interested in pedagogical applications of L2 speech research to the language classroom.

Efrat LAST, MA, has been a Speech and Language Therapist since 2003 and is currently working at her privately owned clinic, as well as doing student supervision. Efrat specializes in pediatric speech therapy, focusing on articulation, language, stuttering and voice disorders. Efrat's Masters thesis on accent acquisition was inspired by her childhood as a diplomat's daughter who lived in 5 countries and was exposed to various languages and accents. Efrat is married and the mother of four girls and a boy.

Shira MATALON completed her B.A. in Communication Disorders studies at Tel Aviv University in 2018. She is now an M.A. student in the same department, performing research in the area of disfluency. In her clinical work as a speech-language pathologist, she treats children and teenagers for issues with language, communication and fluency. She served for three years as an officer in the IDF infantry. Ms. Matalon has travelled around the world to follow her passion for beautiful scenery and unique wildlife. She lives with her husband who is a data scientist; their two-year-old son who fascinates her with his speech development – and with everything else he does; and his older four-legged brother, a large lovable mixed-breed rescue dog.

Olga TARAROVA (PhD, University of Toronto, Canada) is an Assistant Professor at the University of Western Ontario, Canada, where she teaches both at the graduate and undergraduate levels courses in *SLA and Third*

Language Acquisition, Introduction to Linguistics, Comparative Analysis of Spanish and English, Sociolinguistics, Heritage Speakers and Bilingualism, among others. During her Masters and PhD, she worked with a heritage Italian community, Chipilo in Mexico who has preserved their language for over 130 years, so she has extensive knowledge in heritage languages, issues concerning language maintenance and loss, grammar of heritage speakers. Her interests include sociolinguistics, languages in contact, heritage speakers, bilingualism, SLA, and methodology and teaching. Her current projects include 1) feature transfer in L3 from typologically different languages (Mandarin, Russian, English and Spanish) and 2) implementation of hybrid courses. In the past 5 years, she has published in 8 journals and presented at over 30 conferences, both locally and abroad.

INTRODUCTION

ELENA KKESE

One of the difficulties of second language (L2) users refers to phonological processing as L2 sounds are often difficult to be both perceived and produced. Some sounds may cause more problems than others while nativelike performance is very rarely achieved. Based on most speech perception and production theories, there is a relatively straightforward relationship between the two modalities, which enables them to share representations and processes. Speech perception starts with auditory processing of words; these sounds are then mapped onto phonetic and phonological representations, lexical representations, and semantic representations. Speech production, on the other hand, is described in reverse as the semantic representation needs to be accessed first, followed by the lexical representation, and then the sound structure before the word is produced using articulators. This volume examines the interaction between speech perception and production in L2 presenting evidence based on theoretical and empirical research on the linguistic situations in different countries -the United States of America, Greece, Northern Cyprus, Canada, the Republic of Cyprus, Israel, and Spain.

The book comprises eleven chapters, of which four focus on speech perception and production in L2 in Cyprus, a linguistically unique island in the Eastern Mediterranean Sea, which despite its minute size, its linguistic situation is of special interest due to its multilingual and multicultural character. The official languages of the Republic of Cyprus are Greek and Turkish while Cypriot-Greek is the vernacular or low variety; English, the language of the island's former British colonisers, is also widely used (Kkese, 2020). Further, in Northern Cyprus, the official language is Turkish while English is also used and the vernacular is Cypriot Turkish (Zheng, 2015).

The chapters in this volume are grouped under two themes. Five papers in Part 1 discuss theoretical issues related to speech perception and production in L2 while six papers in Part 2 are based on research conducted in different

linguistic L2 contexts concentrating on issues related to speech perception and production.

The five papers in Part 1 discuss how speech perception and production in L2 could affect different skills in the L2 classroom context such as structured literacy (when it comes to reading and spelling instruction), pronunciation, and generally bilingualism. It could further influence clinical contexts, where aptitude in linguistic diversity is interwoven with learning or therapy success. Even though speech perception and production have been characterised as categorical processes, these are affected by the speaker's face and gestures as well as the actual sound of the speech (Massaro, 1998).

In the first paper of this volume, a veteran speech-language pathologist and literacy expert describes how language science guides a best-practices approach to literacy instruction for all students, including students learning an L2 and students with language processing difficulties like dyslexia. In her paper *Research-to-Practice: Language Science as a Guide to Structured Literacy, Speech Perception, and Production*, Sandie Barrie Blackley delineates essential elements of a research-informed and scalable structured literacy approach. The chapter provides a checklist of features to look for in a science-backed, effective, and efficient word structure curriculum and concludes with a challenge for the future.

The second paper focuses on the phonetic elements influenced by bilingualism, as seen in various studies on bilingual speakers. Pagona-Niki Efstathopoulou writes about the relation of L1 and L2 by studying bilingual speakers having as L1 some of languages spoken in the Mediterranean Sea in her paper *Discussing Bilingualism Based on Phonetic Elements Attested in Speech Production*. The focus is on how bilingualism affects segmental and/or suprasegmental phonetic features in the L1 or in L2 speech production.

The author of the third chapter examines the multimodal nature of speech perception, namely, auditory-visual perception. In *Facilitation of L2 Instruction with Data from Auditory-Visual Speech Perception Research*, Doğu Erdener handles L2 as auditory-visual speech perception over three dimensions: (1) L2's relevance to cross-language speech perception theories and how they are revised from an auditory-visual perspective; (2a) how auditory-visual speech mediates the relationship between L2 speech perception and production; (2b) via the L1-L2 orthographic relationship. His paper emphasizes that knowledge from auditory-only speech perception and auditory-visual speech perception research can assist L2 instruction.

In her chapter *The Arcadian Greek Syllable*, Photini Coutsougera examines the syllable of Arcadian Greek in an effort to understand its prosody in more depth and compare it with the Standard Modern Greek (SMG) syllable. In her paper, she investigates Arcadian Greek, which is a nonstandard variety of the L1 SMG; until recently it was thought to resemble SMG too closely to be worthy of study or research (Newton, 1972: 14). In this context, the Arcadian Greek syllable is examined and analyzed on the basis of (a) Kahn's (1976) syllabification model based on the onset maximization principle and (b) sonority requirements and the Sonority Sequencing Principle. Based on the comparison, the onset maximization model accounts for the syllabification of Arcadian Greek more insightfully as it elegantly avoids assigning obstruent codas (with one or two rare exceptions) to the Arcadian Greek syllable.

The author of the final chapter in Part 1 of the book examines speech errors adult native speakers of Standard Modern Greek make when they speak in L2 English. In her paper *Speech Errors Greeks make in English: A Tutorial,* Elena Babatsouli provides a review of literature on L2 English speech production by native Greek speakers. The chapter continues with a comprehensive delineation of the speech errors occurring in individual sounds (segments) and their combinations across prosodic contexts (phonotactics) that native Greek speakers make when speaking English as an L2, providing examples from original data collected over the past ten years as well as on available published data to identify the contexts of such interlanguage errors. The practical significance of this is to provide a useful yardstick for use in L2 instruction settings, either in classrooms or for independent use, and in clinical contexts, where aptitude in linguistic diversity is interwoven with learning or therapy success.

Part 2 of this volume contains six chapters on research relating to speech perception and production in L2 in different linguistic contexts -Canada, Israel, the Republic of Cyprus, and Spain.

The first paper of this part is a sociophonetic pilot study comparing the phenomenon of complete elision in Spanish past participle *-ar* verbs among Latin American Spanish native speakers and English-speaking Spanish L2 learners. In their paper, *The Italian Palatal Lateral and Italian Palatal Nasal: How does English-Speaking Leaners' Production Differ from the Native Speakers?* Giulia Cortiana and Olga Tararova examine the L2 production of the Italian palatal lateral /ʎ/ and the palatal nasal /ɲ/. The stimuli included 48 Italian target words, and an acoustic analysis of 52 tokens was conducted. The findings of this study are important because they

show that L2 learners have difficulty acquiring the phonetic parameters of some L2 sounds.

The Republic of Cyprus is the context of the seventh, eighth and ninth papers. In the seventh paper, High Variability Phonetic Training (HVPT) is proposed as a way to provide naturalistic exposure to L2 learners in order to facilitate the formation of L2 phonetic categories. In her paper *Perceptual Training Effects on the Perception and Production of L2 English Vowels by L1 Greek-Cypriot Learners,* Dimitra Dimitriou examines the effect of eight sessions of High Variability Phonetic Training (HVPT) on the perception and production of English vowels by adult Greek-Cypriot (intermediate) learners, in the absence of any production training. The paper also examines the generalisability of any acquired knowledge to new speakers and contexts.

The learning of English in the Republic of Cyprus at the tertiary level is a topic that has been understudied and it is of particular interest that the focus of the eighth paper in this section throws light on the *Effect of Talker Gender Information on Speech Perception in L2 English.* In this chapter Elena Kkese and Sviatlana Karpava examine how talker gender could influence L2 speech processing, like phoneme categorisation of English consonants and vowels by Cypriot-Greek speakers. The findings indicate that performance is significantly better regarding vowels when the talker is female but when the talker is male, performance is significantly better concerning consonants.

In their chapter *L2 Phone Perception on a Syllabic Level might be Driven by L1 Transfer Mechanisms and/or Universal Patterns of Phone Acquisition,* Elena Kkese and George Georgiou examine how contrastive L2 English plosives on syllable-level are perceived by Cypriot-Greek speakers and how the use of L2 affects these perceptions. The data suggest syllables are important in segmental perception of spoken L2 English, while L1 transfer and universality of phone perception may be influential factors.

The penultimate paper in Part 2 examines the factors influencing the perceived foreign accent of American immigrants to Israel. In their paper *Perception and Production of Hebrew Accent among Female American Immigrants to Israel*, Efrat Last et al. found a nonlinear effect of age of immigration, and an additional factor of time spent daily speaking Hebrew. The acoustic features best correlated with accent were F2 of /u/ and /e/, F1 of /e/, and the VOT of unvoiced plosives. Finally, they found that immigrants had difficulties in acquiring the native /l/, as compared to /r/.

This volume concludes with *The Role of Acoustic Memory and Phonological Short-Term Memory in Non-Native Perception: Which Memory Matters More and When?* Vita Kogan discusses phonological short-term memory (PSTM) and acoustic memory (AM) in their relation to non-native speech perception. PSTM is thought to facilitate the development of target-like cue-weighting and is an important factor in the acquisition of L2 categories (Cerviño-Povedano & Mora, 2015; MacKay, Meador & Flege, 2001) while AM is believed to be responsible for storing acoustic information at a pre-categorical level, and it also contributes to the task of L2 speech perception and learning (Safronova, 2016). In her study, monolingual Spanish speakers were asked to discriminate between two members of an unfamiliar (L0) Russian contrast /i - ɨ/ that does not exist in Spanish. The results indicate that naïve listeners rely on AM and not on PSTM when discriminating an unfamiliar vocalic contrast. L2 learners seem to employ both PSTM and AM in the same task.

References

Cerviño-Povedano, E., and Mora, J. C. 2015. Spanish EFLs' Identification of /i:-I/ and Phonological Short-term Memory. In *Proceedings of the 32nd International Conference of the Spanish Association of Applied Linguistics (AESLA): Language Industries and Social Change. Procedia – Social and Behavioral Sciences* 173, edited by N. Jiménez Carra, E. Calvo, N. Fernández-Quesada, A.M. López Márquez, and A. Stender, 18-23.

Kahn, D. 1976. *Syllable-Based Generalizations in English Phonology*. MIT, Cambridge, Mass.

Kkese, E. 2020. The role of teacher code-switching and its functions in the L2 English classroom: insights into CG students' attitudes at tertiary level education. *Journal of Mediterranean Studies, 29*(1): 95-120.

Massaro, D. W. 1998. Perceiving talking faces: From speech perception to a behavioral principle. Cambridge, MA: MIT Press.

MacKay, I. R., Meador, D., and Flege, J. E. 2001. The Identification of English Consonants by Native Speakers of Italian. *Phonetica* 58(1-2), 103-125.

Newton, B. 1972. *The Generative Interpretation of Dialect: A Study of Modern Greek Phonology*. Cambridge: CUP.

Safronova, E. 2016. *The Role of Cognitive Ability in the Acquisition of Second Language perceptual Phonological Competence*. Unpublished Doctoral Dissertation. Unpublished Doctoral Dissertation. University of Barcelona, Barcelona, Spain. Retrieved from https://www.tesisenred.net/bitstream

Zheng, Y. 2015. Family Language Policy of an English-Turkish Bilingual Family in Northern Cyprus: A Case Study (Master's thesis). Eastern Mediterranean University EMU, Northern Cyprus.

PART I:

THEORETICAL PERSPECTIVES OF SPEECH PERCEPTION AND PRODUCTION IN L2

CHAPTER ONE

RESEARCH-TO-PRACTICE: LANGUAGE SCIENCE AS A GUIDE TO STRUCTURED LITERACY, SPEECH PERCEPTION, AND PRODUCTION

SANDIE BARRIE BLACKLEY

Abstract

Spoken language underpins literacy. Computational modeling has described how reading skills grow, how spoken words trigger learning, and how novel printed words are learned through a combination of context (semantic and syntactic knowledge) and phonological recoding. Across languages and individuals, science is illuminating how beginning readers become proficient readers, leading to some specific, best-practice guidance for teaching students to read and spell. The chapter provides a checklist of features to look for in a science-backed, effective, and efficient word structure curriculum and concludes with a challenge for the future.

Keywords: reading, spelling, literacy, statistical learning, reading science, simple view of reading

1. Introduction

Education and literacy rates have grown steadily worldwide over the past century, enabling rapid economic and social progress and fostering health and wellbeing (Roser & Ortiz-Ospina, 2016). While the definition of literacy is broad, reading, the oral interpretation of written language, is central to all types of literacy and is a major driver of economic and social well-being worldwide (Castles, Rastle, & Nation, 2018). At the same time, there are now many more bilingual or multilingual individuals in the world than there are monolingual individuals, as well as many more children who

are educated through a later-acquired language than there are children educated exclusively through their first language (Tucker, 1999). English is the most common second language in the world and serves as the world's lingua franca (Paulesu et al., 2021). Yet, in the United States alone there are almost five million public school students whose English language proficiency is limited (National Center for Educational Statistics, 2021).

Reading and writing is a cognitive feat unique to humans (and, recently, the computers humans build). As Grainger and Holcomb (2009) as well as Grainger and Ziegler (2011) have pointed out, reading requires extraordinarily complex, multidimensional processing:

> ".... visual word recognition is a mixture of two worlds: one whose main dimension is space – the world of visual objects; and the other whose main dimension is time – the world of spoken language. Skilled readers might therefore have learned to capitalize on this particularity, using structure in space in order to optimize the mapping of an orthographic form onto semantics, and using structure in time in order to optimize the mapping of an orthographic form onto phonology." (Grainger & Ziegler, 2011, p.4).

2. The Science of Reading

The Simple View of Reading (Gough & Tunmer, 1986; Hoover & Gough, 1990; Hoover & Tunmer, 2018; Tunmer & Hoover, 2019; Hoover & Tunmer, 2020) is a theoretical model for how people learn to read, with strong, consensus research support (Petscher et al., 2020) and applicability across languages, cultures, and time. It involves two main, interdependent processes: 1) word recognition processes and 2) oral language comprehension processes. The famous Reading Rope infographic (Scarborough, 2001) illustrates these two processes and how they interact (see the infographic on the webpage of the International Dyslexia Association: Scarborough's Reading Rope: A Groundbreaking Infographic). In this chapter, we will focus on word recognition processes.

In alphabetic systems, the phonemes of the language are represented by letters or groups of letters. In these languages, regardless of the consistency of their spelling-sound correspondences (orthographic depth), phoneme isolation and phoneme-grapheme association trigger the acquisition of the alphabetic principle, which underpins the word recognition process described in the Simple View model.

The ultimate goal of learning to read is to comprehend what is read, and to do that, students must understand how their language's writing system

codes spoken language. Even when reading silently, there is good evidence that reading involves the activation of phonology (Brysbaert, 2019). Since the 1970s, cognitive scientists have described with increasing specificity how beginning readers in alphabetic languages become proficient readers using a dual route to word identification: 1) a sub-lexical route that requires using letter–sound relationships to identify the pronunciation of a novel printed word and 2) a lexical route in which words are automatically recognized, without requiring phonological recoding (Coltheart, 2005). Connectionist models hypothesize that three interrelated domains make word reading possible:

- phonological (pronunciation)
- orthographic (spelling)
- semantic (meaning)

Computational modeling has further described how reading skills grow, how spoken words trigger learning, and how novel printed words are learned through a combination of context (semantic and syntactic knowledge) and phonological recoding (Pritchard et al., 2018). The idea that learning to read involves learning statistical mappings between spelling, sound, and meaning is well-established in reading science (Arciuli, 2018; Qi et al., 2019; Treiman & Kessler, 2021), though not in education (Seidenberg, 2017).

The extent to which readers rely on each route may vary as a function of the orthographic consistency of a language (see Marinelli et al., 2020; Schmalz et al., 2015). Fluent word recognition requires more years in languages like English that have less transparent, more complex orthographies (e.g., Seymour et al., 2003), but when decoding is taken into account similar developmental trajectories are found across orthographies (Frost, 2012; Megherbi et al., 2018). For English, the importance of direct, systematic instruction focused on speech sounds and letter symbols and, specifically, on letter-sound associations (i.e., phonics) has been repeatedly demonstrated. For a review, see Castles, Rastle, and Nation (2018).

The way with which graphemes and phonemes are associated during reading acquisition also shapes the development of the reading system (Ziegler et al., 2005; Ziegler et al., 2020). Research addressing the core mechanisms involved in cross-linguistic transfer of print-to-sound mappings is in its infancy, but recent research (Paulesu et al., 2021) suggests that for high-proficient bilingual Italian-English adult readers, learning the relatively consistent orthography (Italian) first establishes more efficient sets of

orthographic representations for reading and, hence, is beneficial. However, it should be noted that in this otherwise well-controlled study, one thing that was not controlled was the instructional approach used to teach the research subjects the foundations of reading as children (e.g., a phonics versus a whole-language approach).

While word recognition is achieved with unconscious ease by most people, a significant percentage of people struggle with it. The most common reason for reading problems in otherwise cognitively normal individuals is dyslexia, a difficulty related to word recognition processes as opposed to language comprehension processes. Shaywitz (1998) and Shaywitz et al. (1999) noted that dyslexia is perhaps the most common neurobehavioral disorder affecting children, with prevalence rates as high as 17.5 percent.

In what they referred to as the Lexical Quality Hypothesis (LQH), Perfetti (2007) and Perfetti and Hart (2002) postulated that less skilled readers have difficulty with reading comprehension due to inaccurate and laborious word reading caused by weak (poor quality) connections among the three domains of phonology, orthography and meaning. Across orthographies and educational contexts, difficulty with phoneme isolation and categorization tends to be predictive of reading problems, especially in less than transparent orthographies (Cook et al., 2016; Richlan, 2014; Solheim et al., 2021). Phonological awareness and specifically phoneme isolation is a foundational skill for reading acquisition (Georgiou et al., 2021; Milankov et al., 2021), even in non-alphabetic orthographies (Ho et al., 2011). Even in languages that are not alphabetic such as Japanese, the reading acquisition process can be disrupted when a student has difficulty with phoneme isolation (Hoeft et al., 2015). What exactly drives the relationship between reading ability and phonemic skills is not entirely clear (O'Brien et al., 2018). For example, the difficulty may be related to the accessibility of phonological representations or to difficulties with discrimination or perception (Boets et al., 2013; Virtala et al., 2020). In any case, one of the strongest predictors of reading difficulties across languages is actually rapid automatized naming (RAN) (Araújo et al., 2020; Vataja et al., 2021), which is viewed as a measure of the efficiency of phonological retrieval.

Further, Nora et al. (2021) found that students at genetic risk for dyslexia had underspecified, left-hemispheric lexical phonological representations and poorer long-term explicit memory for novel word forms. Since students with word reading difficulties like dyslexia typically avoid reading (Haft et al., 2019), there has been speculation that their weak phonological skills may be a result and not a cause of their reading difficulties. However, in a

longitudinal study using structural magnetic resonance imaging techniques, Kuhl et al. (2020) found that compared to controls, children with dyslexia had significantly lower phonological processing skills and altered functional and structural connectivity within a left-hemispheric network, both before and after literacy instruction, supporting the conclusion that dyslexia originates from an atypical maturation of the speech network. They further noted, "...individuals with dyslexia hyperactivate the ventral premotor cortex during speech processing, presumably because they have to rely more strongly on articulation strategies to compensate for faulty encoding of phonemes…" (p.8).

It is clear that specific guidance for literacy instruction is relative to a number of factors, including the orthography of the language being learned, the reading subskill (i.e., word identification vs. spelling) being taught, and student factors such as bilingualism or dyslexia (Petscher et al., 2020). Nevertheless, across languages and individuals, some general principles pertain to the practice of teaching word identification and spelling.

3. The Practice: Language Science as a Guide to Word Reading and Spelling Instruction

Seidenberg et al. (2020) identify six challenges to connecting reading science and practice:

1. Pursue cross-disciplinary collaborations.
2. Work toward a new science of teaching.
3. Avoid a narrow focus on phonics.
4. Invest in early learning.
5. Develop a science of reading that applies to all readers.
6. Examine existing systems of learning. (p. 9)

As suggested by the sixth item in the list above, it is important to recognize the overarching principles related to teaching and learning in general. Dehaene (2020) has summarized these basic principles in his "Four Pillars of Learning":

- Attention
- Active engagement
- Error feedback
- Consolidation (through sleep)

With these four overarching principles in mind, we will consider the science-backed principles for teaching students to read words and spell proficiently.

Words are central in both oral communication and written communication. Whether communicating orally or in writing, knowing the meaning of a word is essential. But unlike spoken language, learning to read requires developing mental representations for internal (sublexical) word structures. As reviewed above, the mental representations needed for proficient word reading fall in three domains:

- phonological (pronunciation)
- orthographic (spelling)
- semantic (meaning)

These domains are not separate. They interrelate and intertwine (e.g., Llompart, 2021). Further, much of the knowledge in these domains tends to be unconscious. Examples from English include:

- The past tense suffix is always spelled -ed. When it is added to a base word that ends with a voiceless consonant, the suffix is pronounced /t/. When it is added to a base word that ends with a voiced consonant, it is pronounced /d/ unless the base word ends with a /t/ or /d/, in which case it is pronounced /əd/ or /ɪd/.
- In the word *sign*, the <g> is not pronounced, but it is pronounced in the affixed word *signal*.
- In the noun *record* the first syllable is stressed: /ˈrɛ.kərd/. But in the verb *record,* the second is stressed: /rɪ.ˈkɔrd/.

The instructional process is designed to raise the student's awareness of each mental representation, illustrate how it connects to mental representations in the other domains and then, through adequate practice, allow for its automatic and, ultimately, unconscious use. This process is complex and takes a period of years but, as described above, providing instruction using research-backed principles can greatly improve reading and spelling proficiency.

Following is a checklist of features to look for in an effective and efficient word structure curriculum:

- Uses a connectionist approach that identifies the elements or structures in multiple domains and connects them to one another (e.g., Afacan et al., 2017; Donegan & Wanzek, 2021; Ramus et al., 2013; Rayner

et al., 2001; Saletta et al., 2016; Seidenberg, 2005; Ziegler et al., 2020). For example, predictable patterns for pronouncing letters are explained (e.g., the letter <c> is pronounced /k/ as in *call, cup, common* unless it comes before <e>, <i>, or <y>, in which case it is pronounced /s/ as in *cell, city, cycle*. Pronunciation and spelling patterns are connected to meaning through morphology (e.g., the words *city*, *civic* and *citizen* have a common base element).

- Begins early, with a focus on the oral language underpinnings of literacy in preschool (Byers-Heinlein & Lew-Williams, 2013; Jin et al., 2020; Lyster et al., 2020; Scarborough, 2001).
- Begins with a small set of the most consistent and frequent elements (Vousden et al., 2011). Build systematically from simple word structure elements to more complex elements (Petscher et al., 2020). For example, a beginning step is teaching phonemes that are both frequent and that have consistent spellings, such as /s/, /l, /t/, /æ/. A later step in teaching the multiple ways to spell the tense (aka, long) vowel sounds (e.g., m<u>ai</u>l, m<u>a</u>l<u>e</u>, m<u>ay</u>) should be introduced as a later step.
- Provides phonemic awareness practice as a regular part of foundational lessons, even when teaching a consistent orthography (Georgiou et al., 2021). Establishes clear and precise mental models of phonemes with special attention to those that are difficult for the individual to separate or perceive (Chládková & Paillereau, 2020; Liberman, 1973; Liberman et al., 1990).
- Uses a speech-to-print framework (Moats, 2020). In a speech-to-print framework, the teacher begins with a speech sound (e.g., /k/) and after being sure the student can isolate and differentiate it from similar sounds, s/he introduces the way(s) it is spelled (e.g., <c>, <k>, <ck>, etc.). In contrast, a print-to-speech framework begins in the opposite direction, showing the student a letter (e.g., <c>) and then telling him/her how it is pronounced (e.g., /k/). Of course, sound-letter associations are bidirectional, but starting with the speech sound rather than the written symbol provides an opportunity to focus solely on the speech element (e.g., a phoneme), to contrast it to similar elements (e.g., /b/ versus /p/ or /ɛ/ versus /ɪ/).
- Provides enough instruction and practice in writing letters, using a movement pathway approach, to allow consistent, automatic, and effortless letter writing for spelling words (e.g., Feng et al., 2019; Pritchard et al., 2020). Each letter formation has its own mental representation.

- Provides enough instruction and practice with letter-sound association to allow consistent, automatic, and effortless word identification (Petscher et al., 2020). Each letter-sound association is a mental representation. There are networks of letter-sound and sound-letter associations. Each time a new spelling pattern is introduced, the previous patterns are reviewed.
- Provides direct and explicit instruction and practice in how to spell words using predictable orthotactic patterns (Castles, Rastle, & Nation, 2018; Treiman & Kessler, 2019). Not all patterns are equal in their frequency of occurrence nor in their potential to improve performance, so it is important to target the most useful patterns. Taking one of the most complex examples from English orthography, the patterns for spelling /k/ would be introduced over time from basic to advanced:
  - /k/ is spelled <c> before an -a-, -o- or -u- (Examples: *cat, cot, cute*).
  - /k/ is spelled <k> before an -i-, -e- or -y- (Examples: *kit, key, sky*)
  - /k/ is spelled <ck> after a one letter short vowel sound at the end of a one syllable base word (Examples: *back, lick*).
  - /k/ is usually spelled <c> at the end of a multisyllable base word (Examples: *attic, panic*).
  - /k/ is part of a consonant blend /ks/, spelled <x> after a one letter short vowel at the end of a base word (Examples: *wax, tax*).
  - /k/ is spelled <q> in the <qu> spelling pattern before a vowel (Examples: *quiz, squash*).
  - /k/ is spelled <c> when it is the first sound in a consonant blend (Examples: *clip, creek, pact*).
  - /k/ is spelled <k> when it is the second sound in a consonant blend (Examples: *skate, tank, walk*).
  - /k/ is spelled <ch> in words that come from Greek (Examples: *Christmas, choir*).
- Provides direct and explicit instruction and practice with morphology to support vocabulary growth, word use and spelling (Yadav & Yadav, 2021).
- Provides direct and explicit instruction and practice to support vocabulary growth, connecting pronunciation and spelling to meaning (Austin et al., 2021; Baxter et al., 2021; Petscher et al., 2020).
- Provides direct and explicit instruction and practice with sentences, reading sentences fluently and with expression, combining sentences

with (e.g., Walter et al., 2021) and writing sentences with correct spelling and sentence conventions (Graham & Alves, 2021).
- Provides opportunities to read connected text aloud, with feedback and opportunities for repeated practice (Stevens et al., 2017; Therrien, 2004).
- Provides immediate error correction in all practice exercises (Ericsson & Pool, 2016; Roberts et al., 2008; Torgesen et al., 2001).
- Uses a scope and sequence and a set of materials to facilitate teaching an explicit and systematic connectionist approach with cumulative review (Saletta et al., 2016; Spear-Swerling, 2019).
- Provides for differentiated instruction and practice. Students with reading disorders need more explicit instruction and more practice (Kaganovich et al., 2021; Saletta et al., 2016; Treiman, 2018; Ziegler et al., 2005; Ziegler et al., 2020). Students being educated via a L2 (second language) benefit from contrasting mental representations in their L1 (first language) with those in their L2 (Kkese, 2020; Lindner et al., 2021; Llompart, 2021; Ziegler, 2020).

Together, these features describe the structured literacy approach to reading and spelling instruction (Spear-Swerling, 2019). Traditionally, structured literacy (a.k.a., Orton-Gillingham) guidelines included the use of "multisensory" methods (Birsh, 2011). However, there has never been much research supporting this guidance, in part because it is very difficult to separate the effects of multisensory input from the effects of simply better engaging attention. In a small study focused specifically on the question of the value that multisensory input adds over and above the use of explicit, systematic phonics instruction, Schlesinger and Gray (2017) demonstrated that multisensory intervention did not provide an advantage over the structured intervention for participants with typical development or those with dyslexia. While we do not yet fully understand if and how multisensory input operates to impact outcomes, recent brain imaging research has consistently revealed that learning to read and write requires the interconnection, sometimes referred to as "functional connection", of the brain centers responsible for processing various components (e.g., speech sounds, letter symbols, word meanings). These centers are connected through focused practice.

4. Summary

The checklist above can be summarized in three main elements:

- *A curriculum* that includes the multiple components of word structure with an explicit, systematic and cumulative but flexible scope and sequence;
- *A teacher* who knows the language's word structure elements and can actively engage, coach and differentiate lessons for students;
- *A system* that facilitates the delivery of direct, explicit, engaging lessons coordinated with a platform that provides brief, daily practice with immediate feedback and that automatically collects key data.

While the proposed checklist potentially addresses the challenges listed by Seidenberg et al. (2020), implementing such a program is a formidable task. The real challenge for the future will be creating systems to manage this complexity.

References

Afacan, K., Wilkerson, K.L. and Ruppar, A.L. 2017. Multicomponent Reading Interventions for Students with Intellectual Disability. *Remedial and Special Education* 39(4), 229 - 242. https://doi.org/10.1177/0741932517702444

Araújo, S., Huettig, F. and Meyer, A. 2020. What Underlies the Deficit in Rapid Automatized Naming (RAN) in Adults with Dyslexia? Evidence from Eye Movements, *Scientific Studies of Reading*. Advance online publication: https://doi.org/10.1080/10888438.2020.1867863

Arciuli, J. 2018. Reading as statistical learning. *Language, Speech, and Hearing Services in Schools* 49(3S), 634–643. https://doi.org/10.1044/2018_LSHSS-STLT1-17-0135

Austin, C.R., Vaughn, S., Clemens, N.H., Pustejovsky, J.E. and Boucher, A.N. 2021. The Relative Effects of Instruction Linking Word Reading and Word Meaning Compared to Word Reading Instruction Alone on the Accuracy, Fluency, and Word Meaning Knowledge of 4th-5th Grade Students With Dyslexia, *Scientific Studies of Reading*. Advance online publication. https://doi.org/10.1080/10888438.2021.1947294

Baxter, P., Droop, M., van den Hurk, M., Bekkering, H., Dijkstra, T. and Leoné, F. 2021. Contrasting Similar Words Facilitates Second Language Vocabulary Learning in Children by Sharpening Lexical Representations. *Frontiers in Psychology* 12, 688160.

https://doi.org/10.3389/fpsyg.2021.688160
Birsh, J. R. 2011. *Multisensory Teaching of Basic Language Skills* (3rd ed.). Brookes.
Boets, B., Op de Beeck, H. P., Vandermosten, M., Scott, S. K., Gillebert, C. R., Mantini, D., Bulthé, J., Sunaert, S., Wouters, J., and Ghesquière, P. 2013. Intact but less accessible phonetic representations in adults with dyslexia. *Science* 342(6163), 1251–1254. https://doi.org/10.1126/science.1244333
Byers-Heinlein, K., and Lew-Williams, C. 2013. Bilingualism in the Early Years: What the Science Says. *LEARNing landscapes* 7(1), 95–112. https://doi.org/10.36510/learnland.v7i1.632
Brysbaert, M. 2019. How many words do we read per minute? A review and meta-analysis of reading rate. *Journal of Memory and Language*, 109, Article 104047. Advance online publication. https://doi.org/10.1016/j.jml.2019.104047
Castles, A., Rastle, K., and Nation, K. 2018. Ending the Reading Wars: Reading Acquisition From Novice to Expert. *Psychological Science in the Public Interest* 19(1), 5–51. https://doi.org/10.1177/1529100618772271
Chládková, K. and Paillereau, N. 2020. The What and When of Universal Perception: A Review of Early Speech Sound Acquisition. *Language Learning* 7(4), 1136-1182. https://doi.org/10.1111/lang.12422
Coltheart, M. 2005. Modeling reading: The dual-route approach. In M. J. Snowling and C. Hulme (Eds.), *The science of reading: A handbook* (pp. 6–23). Blackwell.
Cook, S. V., Pandža, N. B., Lancaster, A. K., and Gor, K. 2016. Fuzzy Nonnative Phonological Representations Lead to Fuzzy Form-to-Meaning Mappings. *Frontiers in Psychology* 7, 1345. https://doi.org/10.3389/fpsyg.2016.01345
Dehaene, S. 2020. *How We Learn: Why brains learn better than any machine...for now.* Viking.
Donegan, R.E. and Wanzek, J. 2021. Effects of reading interventions implemented for upper elementary struggling readers: A look at recent research. *Reading and Writing*. https://doi.org/10.1007/s11145-021-10123-y
Ericsson, A. and Pool, R. 2016. *Peak: Secrets from the New Science of Expertise*. Houghton Mifflin Harcourt.
Feng, L., Lindner, A., Ji, X.R., Joshi, R.M. 2019. The roles of handwriting and keyboarding in writing: A meta-analytic review. *Reading and Writing* 32, 33–63. https://doi.org/10.1007/s11145-017-9749-x

Frost, R. 2012. Towards a universal model of reading. *Behavioral and Brain Sciences* 35(5), 263-279. https://doi.org/10.1017/S0140525X11001841

Georgiou, G., Inoue, T., Papadopoulos, T., and Parrila, R. 2021. Examining the growth trajectories and cognitive predictors of reading in a consistent orthography: Evidence from a 10-year longitudinal study. *Applied Psycholinguistics,* 1-25. https://doi.org/10.1017/S0142716421000321

Gough, P.B. and Tunmer, W.E. 1986. Decoding, reading, and reading disability. *Remedial and Special Education* 7, 6–10. https://doi.org/10.1177/074193258600700104

Graham, S. and Alves, R.A. 2021. Research and teaching writing. *Reading and Writing*. Advance online publication. https://doi.org/10.1007/s11145-021-10188-9

Grainger, J., and Holcomb, P. J. 2009. Watching the Word Go by: On the Time-course of Component Processes in Visual Word Recognition. *Language and Linguistics Compass* 3(1), 128–156. https://doi.org/10.1111/j.1749-818X.2008.00121.x

Grainger, J., and Ziegler, J. C. 2011. A dual-route approach to orthographic processing. *Frontiers in Psychology* 2, 45. https://doi.org/10.3389/fpsyg.2011.00054

Haft, S.L., Duong, P.H., Ho, T.C., Hendren, R.L., and Hoeft, F. 2019. Anxiety and Attentional Bias in Children with Specific Learning Disorders. *Journal of Abnormal Child Psychology* 47, 487–497. https://doi.org/10.1007/s10802-018-0458-y

Ho, C. S., Leung, M. T., and Cheung, H. 2011. Early difficulties of Chinese preschoolers at familial risk for dyslexia: deficits in oral language, phonological processing skills, and print-related skills. *Dyslexia* 17(2), 143–164. https://doi.org/10.1002/dys.429

Hoeft, F., McCardle, P., and Pugh, K. 2015. The Myths and Truths of Dyslexia in Different Writing Systems. Published online at https://dyslexiaida.org/. Retrieved June 12, 2021 from https://dyslexiaida.org/the-myths-and-truths-of-dyslexia/

Hoover, W. A., and Gough, P. B. 1990. The Simple View of Reading. *Reading and Writing: An Interdisciplinary Journal* 2, 127-160. http://dx.doi.org/10.1007/BF00401799

Hoover, W.A. and Tunmer, W.E. 2020. *The Cognitive Foundations of Reading and Its Acquisition: A Framework with Applications Connecting Teaching and Learning*. Springer.

Hoover, W.A. and Tunmer, W.E. 2018. The Simple View of Reading: Three Assessments of Its *Adequacy. Remedial and Special Education* 39(5), *304–312.* https://doi.org/10.1177/0741932518773154

Jin, F., Schjølberg, S., Wang, M.V., Eadie, P., Nes, R. B., Røysamb, E. and Tambs, K. 2020. Predicting Literacy Skills at 8 Years From Preschool Language Trajectories: A Population-Based Cohort Study. *Journal of Speech and Hearing Research* 63(8), 2752-2762. https://doi.org/10.1044/2020_JSLHR-19-00286

Kaganovich, N., Schumaker, J., and Christ, S. 2021. Impaired Audiovisual Representation of Phonemes in Children with Developmental Language Disorder. *Brain Sciences* 11(4), 507. https://doi.org/10.3390/brainsci11040507

Kkese, E. 2020. Phonological Awareness and Literacy in L2: Sensitivity to Phonological Awareness and Phoneme-Grapheme Correspondences in L2 English in Neokleous, G., Krulatz, A. and Farrelly, R. *Handbook of Research on Cultivating Literacy in Diverse and Multilingual Classrooms*. IGI Global. https://doi.org/10.4018/978-1-7998-2722-1

Kuhl, U., Neef, N. E., Kraft, I., Schaadt, G., Dörr, L., Brauer, J., Czepezauer, I., Müller, B., Wilcke, A., Kirsten, H., Emmrich, F., Boltze, J., Friederici, A. D., and Skeide, M. A. 2020. The emergence of dyslexia in the developing brain. *NeuroImage* 211, 116633. https://doi.org/10.1016/j.neuroimage.2020.116633

Liberman, I. Y. 1973. Segmentation of the spoken word and reading acquisition. *Bulletin of the Orton Society* 23, 65-77. https://doi.org/10.1007/BF02653842

Liberman, I. Y., Shankweiler, D., and Liberman, A. M. 1990. The Alphabetic Principle and Learning to Read. In D. Shankweiler and I. Y. Liberman (Eds.), *Phonology and Reading Disability: Solving the Reading Puzzle*. (1989). *International Academy for Research in Learning Disability Monograph Series*. University of Michigan Press.

Lindner, A. L., Wijekumar, K., and Joshi, R. M. 2021. English Spelling Performance in Writing Samples Among Spanish-Speaking ELLs. *Journal of Learning Disabilities*. Advance online publication. https://doi.org/10.1177/0022219420982995

Llompart M. 2021. Lexical and Phonetic Influences on the Phonolexical Encoding of Difficult Second-Language Contrasts: Insights from Nonword Rejection. *Frontiers in Psychology 12*, 659852. https://doi.org/10.3389/fpsyg.2021.659852

Lyster, S.H., Snowling, M.J., Hulme, C, and Lervåg, A.O. 2020. Preschool phonological, morphological, and semantic skills explain it all: following reading development through a 9-year period. *Journal of Research in Reading* 44 (1), 175-188.

Marinelli, C.V., Zoccolotti, P. and Romani, C. 2020. The Ability to Learn New Written Words is Modulated by Language Orthographic Consistency. *PLoS ONE* 15(2): e0228129. https://doi.org/10.1371/journal.pone.0228129

Megherbi, H., Elbro, C., Oakhill, J., Segui, J., and New, B. 2018. The emergence of automaticity in reading: Effects of orthographic depth and word decoding ability on an adjusted Stroop measure. *Journal of Experimental Child Psychology* 166, 652–663. https://doi.org/10.1016/j.jecp.2017.09.016

Milankov, V., Golubović, S., Krstić, T., and Golubović, Š. 2021. Phonological Awareness as the Foundation of Reading Acquisition in Students Reading in Transparent Orthography. *International Journal of Environmental Research and Public Health* 18(10), 5440. https://doi.org/10.3390/ijerph18105440

Moats, L.C. 2020. Speech to Print: Language essentials for teachers (3rd ed.). Brookes.

National Center for Education Statistics., National Assessment of Educational Progress. Educational Testing Service., and United States. 2021. Condition of Education (NCES 2021-144). Table 6. p.14. National Center for Education Statistics, Office of Educational Research and Improvement, U.S. Dept. of Education. https://nces.ed.gov/pubs2021/2021144.pdf

Nora, A., Renvall, H., Ronimus, M., Kere, J., Lyytinen, H., and Salmelin, R. 2021. Children at risk for dyslexia show deficient left-hemispheric memory representations for new spoken word forms. *NeuroImage* 229(1), 117739. https://doi.org/10.1016/j.neuroimage.2021.117739

O'Brien, G.E., McCloy, D.R., Kubota, E.C. and Yeatman, J.D. 2018. Reading ability and phoneme categorization. *Scientific Reports* 8, 16842. https://doi.org/10.1038/s41598-018-34823-8

Paulesu, E., Bonandrini, R., Zapparoli, L., Rupani, C., Mapelli, C., Tassini, F., Schenone, P., Bottini, G., Perry, C. and Zorzi, M. 2021. Effects of Orthographic Consistency on Bilingual Reading: Human and Computer Simulation Data. *Brain Science* 11, 878. https://doi.org/10.3390/brainsci11070878

Perfetti, C. 2007. Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading* 11(4), 357–383. https://doi.org/10.1080/10888430701530730

Perfetti, C. A., and Hart, L. 2002. The lexical quality hypothesis. Precursors of Functional Literacy 11, 67–86. https://doi.org/10.1075/swll.11.14per

Petscher, Y., Cabell, S. Q., Catts, H. W., Compton, D. L., Foorman, B. R., Hart, S. A., Lonigan, C. J., Phillips, B. M., Schatschneider, C., Steacy,

L. M., Terry, N. P., and Wagner, R. K. 2020. How the Science of Reading Informs 21st-Century Education. *Reading Research Quarterly 55*(Suppl 1), S267–S282. https://doi.org/10.1002/rrq.352

Pritchard, S. C., Coltheart, M., Marinus, E., and Castles, A. 2018. A Computational Model of the Self-Teaching Hypothesis Based on the Dual-Route Cascaded Model of Reading. *Cognitive Science* 42(3), 722–770. https://doi.org/10.1111/cogs.12571

Pritchard, V.E., Malone, S.A. and Hulme, C. 2020. Early Handwriting Ability Predicts the Growth of Children's Spelling, but Not Reading Skills, *Scientific Studies of Reading* 25(4), 304-318. https://doi.org/10.1080/10888438.2020.1778705

Qi, Z., Sanchez Araujo, Y., Georgan, W. C., Gabrieli, J. D. E., and Arciuli, J. 2019. Hearing matters more than seeing: A cross-modality study of statistical learning and reading ability. *Scientific Studies of Reading* 23(1), 101–115. https://doi.org/10.1080/10888438.2018.1485680

Ramus, F., Marshall, C.R., Rosen, S. and van der Lily, H.K.J. 2013. Phonological deficits in specific language impairment and developmental dyslexia: towards a multidimensional model, *Brain* 136 (2), 630–645. https://doi.org/10.1093/brain/aws356

Rayner, K., Foorman, B.R., Perfetti, C.A., Pesetsky, D. and Seidenberg, M.S. 2001. How Psychological Science Informs the Teaching of Reading. *Psychological Science in the Public Interest* 2(2), 31-74. https://doi.org/10.1111/1529-1006.00004

Richlan, F. 2014. Functional Neuroanatomy of Developmental Dyslexia: the role of orthographic depth. *Frontiers in Human Neuroscience* 8, 347. http://doi.org/10.3389/fnhum.2014.00347

Roberts, G., Torgesen, J.K., Boardman, A., and Scammacca, N. 2008. Evidence-based strategies for reading instruction of older students with learning disabilities. *Learning Disabilities Research & Practice* 23(2), 63–69. http://dx.doi.org/10.1111/j.1540-5826.2008.00264.x

Roser, M. and Ortiz-Ospina, E. 2016. Literacy. Published online at OurWorldInData.org. Retrieved June 12, 2021 from 'https://ourworldindata.org/literacy'

Saletta, M., Goffman, L., and Hogan, T. P. 2016. Orthography and Modality Influence Speech Production in Adults and Children. *Journal of Speech, Language, and Hearing Research: JSLHR* 59(6), 1421–1435. https://doi.org/10.1044/2016_JSLHR-L-15-0242

Scarborough, H. S. 2001. Connecting Early Language and Literacy to Later Reading (Dis)abilities: Evidence, theory, and practice. In S. Neuman and D. Dickinson (Eds.), *Handbook for Research in Early Literacy* (pp. 97–110). Guilford Press.

Schlesinger, N.W., and Gray, S. 2017. The Impact of Multisensory Instruction on Learning Letter Names and Sounds, Word Reading, and Spelling. *Annals of Dyslexia* 67(3), 219-258. https://doi.org/10.1007/s11881-017-0140-z

Schmalz, X., Marinus, E., Coltheart, M. and Castles, A. 2015. Getting to the bottom of orthographic depth. *Psychonomic Bulletin & Review* 22, 1614–1629. https://doi.org/10.3758/s13423-015-0835-2

Seidenberg, M. S. 2005. Connectionist Models of Word Reading. *Current Directions in Psychological Science* 14(5), 238–242. https://doi.org/10.1111/j.0963-7214.2005.00372.x

Seidenberg. M. 2017. *Language at the Speed of Sight: How We Read, Why So Many Can't, and What Can Be Done About It*. Basic Books.

Seidenberg, M.S., Borkenhagen, M.C. and Kearns, D.M. 2020. Lost in Translation? Challenges in Connecting Reading Science and Educational Practice. *Reading Research Quarterly* 55 (S1), *S119-S130. https://doi.org/10.1002/rrq.341*

Seymour, P. H., Aro, M., and Erskine, J. M. 2003. Foundation Literacy in European Orthographies. *British Journal of Psychology* 94, 143–174. https://doi.org/10.1348/000712603321661859

Shaywitz, S. E. 1998. Current concepts: Dyslexia. *The New England Journal of Medicine* 338(5), 307–312. https://doi.org/10.1056/NEJM199801293380507

Shaywitz, S. E., Fletcher, J. M., Holahan, J. M., Shneider, A. E., Marchione, K. E., Stuebing, K. K., Francis, D. J., Pugh, K. R., and Shaywitz, B. A. 1999. Persistence of dyslexia: the Connecticut Longitudinal Study at adolescence. *Pediatrics* 104(6), 1351–1359. https://doi.org/10.1542/peds.104.6.1351

Solheim, O. J., Torppa, M., Uppstad, P. H., and Lerkkanen, M.K. 2021. Screening for Slow Reading Acquisition in Norway and Finland: a Quest for Context Specific Predictors. *Scandinavian Journal of Educational Research 65*(4), 584-600. https://doi.org/10.1080/00313831.2020.1739130

Spear-Swerling, L. 2019. Structured Literacy and Typical Literacy Practices: Understanding Differences to Create Instructional Opportunities. *Teaching Exceptional Children* 51(3), 201–211. https://doi.org/10.1177/0040059917750160

Stevens, E. A., Walker, M. A., and Vaughn, S. 2017. The effects of reading fluency interventions on the reading fluency and reading comprehension performance of elementary students with learning disabilities: A synthesis of the research from 2001 to 2014. *Journal of Learning Disabilities* 50(5), 576–590.

https://doi.org/10.1177/0022219416638028
Therrien, W. J. 2004. Fluency and Comprehension Gains as a Result of Repeated Reading: A Meta-Analysis. *Remedial and Special Education* 25(4), 252–261. https://doi.org/10.1177/07419325040250040801
Torgesen, J.K., Alexander, A.W., Wagner, R.K., Rashotte, C.A., Voeller, K., Conway, T., and Rose, E. 2001. Intensive remedial instruction for children with severe reading disabilities: Immediate and long-term outcomes from two instructional approaches. *Journal of Learning Disabilities* 34, 33–58. https://doi.org/10.1177/002221940103400104
Treiman, R.2018. Statistical Learning and Spelling. *Language Speech Hearing Services in the Schools* 49(3S), 644-652. https://doi.org/10.1044/2018_LSHSS-STLT1-17-0122
Treiman, R., and Kessler, B. 2019. Development of context-sensitive pronunciation in reading: The case of ‹c› and ‹g›. *Journal of Experimental Child Psychology 182*, 114–125. https://doi.org/10.1016/j.jecp.2019.02.001
Treiman, R. and Kessler, B. 2021. Statistical Learning in Word Reading and Spelling across Languages and Writing Systems, *Scientific Studies of Reading* 25(4). https://doi.org/10.1080/10888438.2021.1920951
Tucker, G. R. 1999. A Global Perspective on Bilingualism and Bilingual Education. ERIC Publications. Retrieved June 26, 2021 from https://eric.ed.gov/?id=ED435168
Tunmer, W. and Hoover, W.A. 2019. The cognitive foundations of learning to read: a framework for preventing and remediating reading difficulties, *Australian Journal of Learning Difficulties* 24(1), 75-93. https://doi.org/10.1080/19404158.2019.161
Vataja, P., Lerkkanen, M.-K., Aro, M., Westerholm, J., Risberg, A.-K., and Salmi, P. 2021. The Predictors of Literacy Skills among Monolingual and Bilingual Finnish-Swedish Children During First Grade. *Scandinavian Journal of Educational Research*. Advance online publication. https://doi.org/10.1080/00313831.2021.1942191
Virtala, P., Talola, S., Partanen, E. and Kujala, T. 2020. Poor neural and perceptual phoneme discrimination during acoustic variation in dyslexia. *Scientific Reports* 10, 8646. https://doi.org/10.1038/s41598-020-65490-3
Vousden, J.I., Ellefson, M.R., Solity, J., and Chater, N. 2011. Simplifying reading: Applying the simplicity principle to reading. *Cognitive Science* 35(1), 34–78. https://doi.org/10.1111/j.1551-6709.2010.01134.x
Yadav, M. and Yadav, M. 2021. An Overview on the Impacts and Strategies of Morphology in English Language Teaching to the L2 Learners.

International Journal of Research 8(3), 311-331. Retrieved July 14, 2021 from https://journals.pen2print.org/index.php/ijr/article/view/20487

Walter, K., Dockrell, J. and Connelly, V. A sentence-combining intervention for struggling writers: response to intervention. *Reading and Writing*. 2021. https://doi.org/10.1007/s11145-021-10135-8

Ziegler, J. C., Pech-Georgel, C., George, F., Alario, F. X., and Lorenzi, C. 2005. Deficits in speech perception predict language learning impairment. *Proceedings of the National Academy of Sciences of the United States of America* 102(39), 14110–14115. https://doi.org/10.1073/pnas.0504446102

Ziegler, J. C., Perry, C., and Zorzi, M. 2020. Learning to Read and Dyslexia: From Theory to Intervention Through Personalized Computational Models. *Current Directions in Psychological Science 29*(3), 293–300. https://doi.org/10.1177/0963721420915873

CHAPTER TWO

DISCUSSING BILINGUALISM BASED ON PHONETIC ELEMENTS ATTESTED IN SPEECH PRODUCTION

NIKI EFSTATHOPOULOU

Abstract

This chapter starts with some basic definitions concerning the phenomenon of bilingualism, while focusing on the diversity of the phenomenon, since there are many different kinds of bilinguals. Then, some important factors affecting the speech of bilinguals in both their first language (L1) and their second language (L2) are mentioned. With regard to the theories attempting to explain how bilingualism affects the phonetics of the speech produced by bilingual speakers, a short reference is made to the Perceptual Assimilation Model (PAM) suggested by Best and to the Speech Learning Model (SLM) proposed by Flege. The last and most extended part of this chapter presents studies discussing the relation of bilingualism to L1 and L2 phonetics. The main question discussed is whether the speech of bilinguals is phonetically alike to L1 native speakers' speech or to L2 native speakers' speech. If none of the above is true, then bilinguals' speech is the proof that an interlanguage is created; a shared system with phonetic elements that cannot be found neither in L1 nor in L2. Lastly, studies about the advantages and/or disadvantages of bilingualism for an individual are presented. The writer concludes that bilingualism appears to offer many advantages to a person's life, contrary to what it was believed in the past based on the balance theory.

Keywords: bilingualism, L1, L2, phonetics, interlanguage

1. Attempting to define "bilingualism"

Bilingualism refers to the situation of a person speaking two or more languages. In other words, a bilingual speaker has good production and perception of each language in all linguistic levels (i.e. phonology, phonetics, morphology, syntax, etc.). However, as it can be deduced, it is difficult for scientists to define what exactly "good" means. There is a discussion on when a bilingual is like a native speaker in his/her first language (L1) and/or in his second language (L2). When the bilingual speaker has native-like production and perception in both the L1 and L2, the term "perfect bilingual" or "balanced bilingual" is used (Corvalán & Treffers-Daller, 2015: 95). When this is not the case, the bilingual is mentioned as unbalanced as a result of uneven bilingual development (De Houwer, 2021: 21-22).

Another common distinction among bilinguals is based on when they first acquired the L2. Bilinguals are divided to "early and late bilinguals" based on whether have acquired the L2 as children or later on in their lives (De Houwer, 2021: 22-25). As Marian and Hayakawa (2020: 4) discuss, the question that emerges is "how bilingual are you?", that is, when someone is considered a better bilingual than another bilingual person. To answer this question, it should be acknowledged that bilingualism is a phenomenon with several aspects and many different forms. For example, someone can learn both languages from an early age and he/she is able to speak both of them in formal social environments (work, home, written speech, etc.) and in informal social environments (family, friends, spoken speech, etc.). Another individual may come into contact with his L2 in his early adulthood as a university student; he may have spent lot of time writing, reading literature in L2 and studying formally the language without being in contact with the native L2 speakers. While both situations described are situations of bilingualism, these individuals are not identical bilinguals. Thus, Marian and Hayakawa (2020: 28-31) conclude that scientists cannot really characterize the first or the second type of bilingual speaker as better bilingual.

Therefore, it is important for describing the phenomenon of bilingualism to acknowledge its diversity and complexity. Factors such as age of acquisition, manner of acquisition, level of proficiency, language use, language identity and language switching should be taken into consideration when examining bilingualism (Marian & Hayakawa, 2020: 4-7). In other words, a bilingual's perception and production in both L1 and L2 depends on the above factors. To summarize the impact of these factors, the earlier

someone acquires L2, the better he/she performs at this language, while he also discriminates between L1 and L2 performance. So, generally speaking, bilinguals who were first exposed and learned a language as children are "better" bilinguals in comparison to the ones who got exposed to the L2 later on, for example as adults. In phonetics' level, a "better" performance usually means that the bilingual produces unaccented speech in the L2. Moreover, teaching a language appears to have "worse" results on bilinguals' performance than acquiring the language in a native-like manner. The level of proficiency in a language is difficult to tackle in detail and in an objective manner. This fact is clearly shown in language exams, where the examiners try to assess a student's overall language proficiency in a short period of time with various tests and methods. Moreover, the term "language use" refers to how L1 and L2 are used by the bilingual speaker. For example, some bilinguals use the L1 only at home or with friends, thus in informal environments, while the L2 is used in their work environment or out-of-home environments, that is, it is their societal language. As a result, in real life, there are few balanced bilinguals. This suggests that language use "shapes" the type of bilingual an individual is.

Another important aspect of bilingualism is related to language identity. Some questions that may be formed may be "Is L1/heritage language and the country where it is spoken considered to be superior compared to L2?", "has L2 a higher identification status compared to L1?" The answers to these questions have a strong impact on an individual's bilingualism, since language is significant for a speaker's identity. For example, the accented L2 English speech production is negatively marked for a speaker with Spanish as L1 since Spanish has a lower language status compared to English, but it is positively marked when the speaker's L1 is French given its equal language status to English). Lastly, bilingual speech production is strongly related to linguistic phenomena such as language or code switching; that is, when the bilingual person freely alternates between L1 and L2 in his/her speech.

To conclude, bilingualism is a state of flux, constantly changing and it is a phenomenon affected not only by merely linguistic factors, but also social ones (Piske et al., 2001: 191-215).

2. Theories associating bilingualism and phonetics

There are several theories on how bilingualism affects speech production and perception in both the L1 and L2. Since this chapter focuses on

bilingualism's effects on phonetics, the focus is on the known linguistic theories that can accommodate the phonetic phenomena related to bilingualism: a) the Perceptual Assimilation Model (PAM) proposed by Best and b) the Speech Learning Model (SLM) suggested by Flege.

a) Perceptual Assimilation Model: PAM was proposed as an account for the discrimination of non-native sounds by adult speakers (Best, 1991, 1993a, 1993b, 1995, 2001, 775-780, Antoniou et al., 2011: 560). This theory follows a more psychological/developmental point of view on bilingualism, focusing on the differences between children and adult speakers. It examines perception and discrimination of non-native sounds - not only L2 sounds- based on the phonological space of the L1. When a speaker listens to a non-native sound/phone, an assimilation process is generated in relation to L1 sounds. Thus, two phones can be identified and perceived as either:

1) a good match for two distinct native categories (TC=two category assimilation), 2) a good assimilation of both to a single L1 phoneme (SC=single category assimilation), 3) a better match for the one and a less good for the other to the same native phoneme (CG=category goodness) or 4) non-assimilable, non-speech sounds (NA). The combinations possible are Uncategorized-Categorized (UC) or Uncategorized- Uncategorized (UU). This model describes how sound categories are formed in L2 acquisition, but it does not directly address speech production of those categories.

To shortly explain PAM sounds' classification for a speaker with Greek as L1, /p/ and /b/ are TC for the equivalent English L2 consonant sounds. Similarly, while listening to L2 English vowel sounds /i/ and /ɪ/, a Greek speaker assimilates to SC in the L1 category of Greek /i/. To present an example of CG assimilation, imagine a French/Greek bilingual. The Greek vowel sound /u/ is the closest and better choice for the French /u/, but it is a worse choice for the French vowel sound /y/ to assimilate. Also, for a Greek native speaker, click sounds found in Khoisan languages (i.e. labial ʘ, dental ǀ, palato-alveolar ǂ) are usually perceived and identified as NA.

b) The Speech Learning Model was developed by Flege (1987a: 9-20, 1987b: 47-54, 1991a, 1992, Flege et al., 2003: 468-472) after many studies suggesting that late bilinguals usually do not produce L2 sounds authentically; in everyday life, native speakers perceive this inability of L2 speakers as foreign accented speech. Flege explained this situation as "the new wine in old bottles" phenomenon, meaning that the L1 crucially

influences the L2 acquisition and production. He also used the metaphor of applying a new color of paint on a wall already painted; the first coat of paint (L1) has an impact on how the new colour of paint will turn out (L2 production) because of the influence of the underlying colour.

With regard to SLM, speech production is organized at the phonemic category, phonetic category and sensory-motor levels (Flege, 1991). For late bilinguals, this organization means that they identify similar sounds under one phonetic category but still detect the sounds' acoustic differences. Then, they manage to produce those sounds differently somehow by using different realization rules. An important notion within SLM is the "equivalence classification" that exists in both L1 and L2. Equivalence classification allows native speakers of a language to perceive all the instances of a phone under one phonetic category, despite their phonetic differences. However, equivalence classification works as an impediment for L2 learners, because it prevents them from categorizing a sound under one L2 phonetic category when this sound is similar to an existing L1 category. Flege suggested that L2 learners may form a new category including phonetic characteristics of both L1 and L2 relevant sounds.

Thus, Flege (1987a, 1987b) taxinomized L2 language sounds based on their relation to L1 language sounds in "identical", "new" or "similar". When they are "identical", a sound is the same in L1 and L2, so the L2 learner does not need to create a new category, but he/she simply produces the L2 sound authentically. For example, the English /i/ as L2 sound can be subcategorized under the same L1 category of the Greek /i/. When the L2 sound is "new" compared to L1 sounds, the L2 learner creates a new phonetic category and produces the L2 sound according to the phonetic characteristics of this category. For example, the French /y/ does not exist in the L1 phonetic inventory of a Greek speaker. However, the production of "similar" sounds is a challenge for an L2 learner. "Similar" sounds differ systematically acoustically and phonetically, but they are similar enough so as to use the same IPA symbol. This is the case for voiceless stops /p t k/ in Greek and English. They do have the same place of articulation, but they differ in regard to aspiration: Greek voiceless stops /p t k/ are short-lag unaspirated, while English ones are long-lag aspirated /p^h t^h k^h/ (Efstathopoulou, 2007: 4-6).

Since many studies have shown that some interlanguage phenomena could not be accommodated based on SLM, Flege revised his model to SLM-r. SLM-r is based on studies on Japanese/English speakers regarding the liquid sounds (Flege & Bohn, 2021: 23-31, 64-66). When native Japanese

speakers are first exposed to English, they categorize both English /r/ and /l/ sounds under the L1 category of Japanese /R/. These sounds are distinct in L2/English, but undistinguished in L1/Japanese. According to Flege, when a speaker is exposed to an L2 language, he/she maps all L2 sounds to the L1 sound categories. This is a subconscious and automatic cognitive process called "interlingual identification".

However, all the instances of sounds that cannot be categorized under L1 sound categories are "stored" under a new category that Flege names "equivalence class". For native Japanese speakers, this new phonetic category will combine the phonetic properties of the English /l r/ and Japanese /R/ to which they have been exposed during their lives. Eventually, this category will modify how the Japanese /R/ is produced and perceived. Therefore, according to Flege, all the mechanisms and processes needed for native language (L1) acquisition remain accessible for use in L2 learning during someone's life, while the phonetic categories making up the L1 and L2 phonetic subsystems interact with one another dynamically and can be updated (Flege & Bohn, 2021: 3-83).

3. Studies on how bilingualism affects speech production

Flege et al. (1996) carried out a study involving 240 Italian/English bilinguals focusing on their productions of initial consonants, including the voiceless stops /p/ and /t/, because the production of these consonants differs between the two languages. A written word list was placed in front the subjects of the study. Then, they heard a sentence "___is the next word" and they had to use this word in the carried phrase "Now I say____". Some of the English words used as stimuli were: pick /pɪk/, peak /pi:k/, tack /tæk/, tag /tæg/, they /ðeɪ/, then /ðɛn/, thought /θɔ:t/, thief /θi:f/. The results of the Italian/English bilingual subjects were compared to those of a control group of 24 native English speakers. The analysis of the selected data showed that L2 learners produced compromise VOT (voice onset time) values for English, significantly shorter than those of native English speakers. Interestingly enough, language related factors such as AOL (age of learning) accounted only for 50% of the variance in the data, thus, more investigation is needed.

In Flege's study (1991a) on Spanish/English bilinguals and their VOT productions of /t/, participants were divided into two groups; early and late bilinguals. Their VOT productions were compared to these of two monolingual groups, English and Spanish. The native Spanish subjects read Spanish words at the end of the Spanish carrier phrase "Tengo un

___". The stimuli were divided to words where /t/ was followed by /i/ (tigre, tipo, tiro, timbre) and those followed by/ɛ/ (tema, termo, templo, texto). The English monolinguals and the Spanish/English bilinguals read English words with /i/or /e/ (t-bone, teapot, teabag, t-shirt, teller, temple, textbook, tempo) at the end of the English carrier phrase "Take a ___". Similar stimuli were used for the second experiment, however, they were produced in three different conditions; sentence, phrase and word. The results of both experiments showed that early bilinguals produced /t/ authentically in both the L1 and L2. On the other hand, late bilinguals' productions proved that their L1 was not affected, but this was not the case for their L2, since the bilinguals produced shorter English VOT values than monolingual English speakers. However, Flege and Eefting (1986) compared the VOT values of Spanish /t/ in the speech of Spanish monolinguals and Spanish/English bilingual speakers. They showed that the bilinguals' values were significantly longer than those of monolinguals, so L1 was also affected by the bilingual status of the speakers. Thus, a bidirectional interference hypothesis is supported (Flege & Davidian, 1985).

Efstathopoulou (2007) carried out a study that examined bilingualism's phonetic impact on the speech of 20 Greek/English late bilinguals. They produced voiceless stops /p t k/ in both Greek and English real words in a controlled environment as in the sentence "I said___again" in English and "Ipa___pali" in Greek; the stimuli included all the possible voiceless stop-vowel environments in the first syllable. For example, some Greek stimuli were /ˊkita/ 'κοίτα', /ˊpote/ 'πότε', /ˊpezo/ 'παίζω', while some English stimuli were 'coping' /ˈkoʊpɪŋ/, 'kettle' /ˈkɛtəl/, 'tanning' /ˈtænɪŋ/, etc. An accent-rating experiment was also performed; the participants were recorded when reading specific sentences in English. Four native English speakers were asked to rate these recordings based on a 9-point scale. Many factors were taken into consideration such as Age of Arrival (AOA) and Length of Residence (LOR) in Canada, education level, estimated everyday use of L1/Greek, estimated proficiency in L1/Greek and L2/English and so on.

The bilinguals overall produced significantly different VOTs for Greek and English. These results supported Flege's notion of language learning mechanisms. Such results support the idea that phonetic ability remains intact throughout an individual's life-span. Both L1 and L2 systems of bilinguals appear to be different than those of monolinguals. They distinguished voiceless stops in L1 and L2 but crucially not in all vowel environments; this suggested that bilinguals created a new phonetic

category that does not exist in L1 or L2 (Efstathopoulou, 2007: 69). Thus, a theory of interlanguage, a shared system of bilinguals, seems to be supported as the SLM-r suggested later on (Flege & Bohn, 2021: 64-66).

Discussing bilingualism's advantages, Antoniou (2019) presents some research in favour of this view. The research examined bilingualism's influence on executive functions, cognitive aging and brain plasticity of bilingual speakers. A bilingual person is able to speak two or more languages, thus, he benefits from his ability to interact with many different people, associate with different cultures and/or countries, have multiple work and study opportunities.

Despite this obvious advantage, bilingualism is considered to have a positive status in a person's life. During the 20th century, the majority of scientists used to refer to bilingualism as a "problem of the bilingual child" (Smith, 1923), while many studies presented low scores in verbal tests that examined the cognitive abilities of the bilingual children (Darcy, 1953). To summarize, the hypothesis behind these studies insisted that bilingualism confuses children and constitutes an impediment in their linguistic and educational evolution in comparison to a monolingual child. Some scientists even speak about "mental retardation" of bilingual children (Goodenough, 1926). For Antoniou (2019), these studies are obviously flawed, since factors like age, socioeconomic status, and degree of bilingualism of the participants were not taken into account. Moreover, he underlines the fact that such approaches focus on the immigrant or refugee status of children's parents in the country where the L2 was spoken. This is why such opinions are still expressed by pedagogues who try to convince immigrant parents to minimize the use of heritage language with their children, since their L2 is considered to be of higher status than their L1. L1 is considered the one to blame for the derogatory term of "semi-linguistic" that is used to describe the speech of bilinguals (Antoniou, 2019: 396; Τριάρχη-Herrmann, 2000: 73-81)

A well-known hypothesis is the balance theory, which is described as an analogy with balloons. That is, L1 is expanding like a balloon in a child's brain occupying all the space available. However, when there are two languages, an L1 and an L2, it is like there are two balloons in a child's brain. Thus, when two languages are available, the mental space seems to be divided into two parts. The theory implies that no language is fully acquired by a bilingual child, while the two languages are "impaired" and somehow "less" than a monolingual's child. Therefore, the two languages stay separated within the human brain (Cummins, 1980, 1981).

Eventually, Peal and Lambert's (1962) study took into consideration the methodological shortcoming of previous research on bilingual children indicating that bilingual children could outperform monolingual ones in non-verbal, but also verbal tests in the participant's L1. Thus, bilinguals presented to have a "suboptimal cognitive development". Since then, a considerable amount has been carried out that proves a general "bilingual advantage" in many aspects, such as executive function, metalinguistic awareness, phonetic perception, flexibility, thinking, and even delay of dementia (Gollan et al., 2011; Alladi et al., 2013).

Antoniou (2019) admits that how bilingualism works within the human brain and its mechanisms have yet to be discovered. There are more studies to be done on this field, but he concludes that the constant struggle to balance efficiently between the two languages and process all this cognitive workload makes bilinguals to have an advantage (Antoniou et al., 2013). He refers to studies proposing that bilingualism improves executive functions; "cognitive processes that control behavior in service of attaining goals" (Antoniou, 2019: 397). These abilities are changing during a person's life because they are subject to decline due to aging and changes because of cognitive maturation. They are also influenced by short and long-term experiences. Since a bilingual's everyday life involves "constant monitoring, inhibition, selection, and planning" (Antoniou, 2019: 397), it is deduced that there are some effects of bilingualism on executive functions, even though they are not yet clearly defined.

Children and older bilinguals perform better than monolinguals in different memory tasks, while advantages of bilingualism have been found even in 18-month old babies and 7-month-olds (Brito & Barr, 2012; Kovács & Mehler, 2009). Bilingualism affects cognition before children even speak. It may help to a smoother and health cognitive aging process, as studies on bilingual elderly mention (Bak et al., 2014; Perquin et al., 2013). Antoniou (2019: 403) found that a bilingual's cognitive ability predicts cognitive ability better than age, age at immigration, education, or gender. Even serious neurological problems such as dementia are under investigation as studies on elderly with dementia or Alzheimer suggest that their bilingual status delays the incidence of Alzheimer (Bialystok et al., 2007).

Lastly, there are studies that examine brain plasticity, which refers to changes that bilingualism imposes to the human brain due to experience. Those changes affect future brain function (Luk et al., 2011; Abutalebi et al., 2015). What is important is that bilinguals are found to compensate for

the deterioration in brain structure and function due to aging. Also, neuroscience gives plenty of evidence suggesting that bilingualism alters human brain to a better/improved direction in comparison to monolinguals' brain. Antoniou (2019) concludes that bilingualism functions as a cognitive stimulation, while it also engages a larger brain network in order for a bilingual individual to succeed when having to constantly use two or more languages.

Antoniou et al. (2015: 683-686) carried out a study trying to examine the following known facts about bilinguals: a) bilinguals learn a third language easier than a monolingual, b) bilinguals learn easier a language related to the ones they already know, and c) some phonetic features are universally more difficult to acquire than others. As a result, artificial words were created in order to test these facts comparing a sample of bilingual Mandarin–English speakers with a sample of English speaking monolinguals. In the three experiments carried out, the bilinguals outperformed monolinguals. These results point to an advantage of bilinguals over monolinguals when it comes to phonetic learning. Lastly, two studies on Greek/English early bilinguals (Antoniou et al., 2012, 2013; Abu-Rabia & Sanitsky, 2010) showed that bilinguals discriminate phonetic categories better than monolingual speakers. Thus, bilinguals' phonetic learning may rely on bilingual's prior phonetic experience.

4. Conclusion

Bilingualism is a common but complex phenomenon. As it was discussed, it affects both the L1 and L2 language system leading to interlanguage phenomena in bilinguals' speech. While the advantages of bilingualism are still debated, more research needs to be carried out on how neurological and linguistic mechanisms work within a bilingual's brain.

References

Abu-Rabia, S., & Sanitsky, E. 2010. Advantages of Bilinguals Over Monolinguals in Learning a Third Language. *Bilingual Research Journal* 33(2), 173-199. DOI: 10.1080/15235882.2010.502797

Abutalebi, J., Guidi, L., Borsa, V., Canini, M., & Della Rosa, P. 2015. Bilingualism provides a neural reserve for aging populations. *Neuropsychologia* 69, 201–10.

Alladi, S., Bak, T., Duggirala, V., Surampudi, B., & Shailaja, M. 2013. Bilingualism delays age at onset of dementia, independent of education and immigration status. *Neurology* 81, 1938–44.

Antoniou, M. 2015. The bilingual advantage in phonetic learning. *Bilingualism: Language and Cognition* 18(4), 683–695. DOI: 10.1017/S1366728914000777

Antoniou, M. 2019. The Advantages of Bilingualism Debate. *Annual Review of Linguistics* 5(1), 395-415. DOI: https://doi.org/10.1146/annurev-linguistics-011718-011820

Antoniou, M., Best, C. T., Tyler, M. D., & Kroos, C. 2011. Inter-language interference in VOT production by L2-dominant bilinguals: Asymmetries in phonetic code-switching. *Journal of Phonetics* 39(4), 558–570. DOI: 10.1016/j.wocn.2011.03.001

Antoniou, M., Best, C. T., & Tyler, M. D. 2013. Focusing the lens of language experience: Perception of Ma'di stops by Greek and English bilinguals and monolinguals. *Journal of the Acoustical Society of America* 133, 2397–2411. DOI:10.1121/1.4792358

Antoniou, M., Gunasekera, M. G., Wong, P. C.M. 2013. Foreign language training as cognitive therapy for age-related cognitive decline: A hypothesis for future research, *Neuroscience & Biobehavioral Reviews* 37(10), 2689-2698. DOI: https://doi.org/10.1016/j.neubiorev.2013.09.004.

Antoniou, M., Liang, E., Ettlinger, M., & Wong, P. 2015. The bilingual advantage in phonetic learning. *Bilingualism: Language and Cognition* 18(4), 683-695. DOI:10.1017/S1366728914000777

Antoniou, M., Tyler, M. D., & Best, C. T. 2012. Two ways to listen: Do L2-dominant bilinguals perceive stop voicing according to language mode? *Journal of Phonetics* 40, 582–594. DOI: 10.1016/j.wocn.2012.05.005

Best, C. 1991. The emergence of Native-language phonological influences in infants: a perceptual assimilation model. *Haskins Laboratories status report on speech research*, July-December 107-108, 1-30.

Best, C. 1993a. Emergence of language-specific constraints in perception of non-native speech: a window on early phonological development. In *Developmental Neurocognition: Speech and face processing in the first year of life* edited by B. de Boysson-Bardies, S. de Schonen, P. Jusczyk, P. MacNeilage, & J. Morton, 289-304. Netherlands: Kluwer Academic Publishers.

Best, C. 1993b. Learning to perceive the sound pattern of English. *Haskins Laboratories status report on speech research*, April-June 114, 31-79.

Best, C. 1995. A Direct Realistic View of Cross-Language Speech Perception. In *Speech Perception and Linguistic Experience: Issues in Cross-Language Research* edited by W. Strange, 171-204, Baltimore: York Press.

Best, C. 2001. Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *Journal of Acoustical Society of America* 109, 775-794.

Bialystok, E., Craik, F., & Freedman, M. 2007. Bilingualism as a protection against the onset of symptoms of dementia. *Neuropsychologia* 45, 459–64.

Brito, N., & Barr, R. 2012. Influence of bilingualism on memory generalization during infancy. *Developmental Science* 15, 812–16.

Cummins, J. 1980. *The Construct of Language Proficiency in Bilingual Education.* In *Current issues in bilingual education: Georgetown University Round Table on Languages and Linguistics* edited by J. E Alatis, 81-103.

Cummins, J. 1981. *Bilingualism and Minority Language Children.* Toronto: The Ontario Institute for Studies in Education.

Darcy, N. T. 1953. A review of the literature on the effects of bilingualism upon the measurement of intelligence. *The Journal of Genetic Psychology* 82, 21–57.

De Houwer, A. 2021. *Bilingual Development in Childhood (Elements in Child Development).* Cambridge: Cambridge University Press. DOI: https://doi.org/10.1017/9781108866002

Gollan, T., Salmon, D., Montoya, R., & Galasko, D. 2011. Degree of bilingualism predicts age of diagnosis of Alzheimer's disease in low-education but not in highly educated Hispanics. *Neuropsychologia* 49, 3826–30.

Goodenough, F. L. 1926. Racial differences in the intelligence of school children. *Journal of Experimental Psychology* 9, 388–97.

Flege, J. E. 1987a. Effects of Equivalence Classification on the Production of Foreign Language Speech Sounds. In *Sound Patterns in Second Language Acquisition* edited by A. James & J. Leather, 9-39. Dordrecht-Holland/Providence-USA: Foris Publications.

Flege, J. E. 1987b. The production of "new" and "similar" phones in a foreign language: evidence for the effect of equivalence classification. *Journal of Phonetics* 15, 47-65.

Flege, J. E. 1991a. Second Language Speech Learning Theory, Findings, and Problems. In *Speech Perception and Linguistic Experiment: Theoretical and Methodological Issues in Cross-Language Speech Research* edited by W. Strange, 233-277. Timonium, MD: York Press.

Flege, J. E. 1991b. Age of learning affects the authenticity of voice onset time (VOT) in stop consonants produced in a second language. *Journal of Acoustical Society of America* 89, 395-411.

Flege, J. E. 1992. Speech Learning in a Second Language. In *Phonological Development: Model, research, implications* edited by C. A. Ferguson, L. Menn & C. Stoel-Gammon, 565-604. Timonium, MD: York Press.

Flege, J., & Bohn, O. 2021. *The Revised Speech Learning Model (SLM-r).* In *Theoretical and Empirical Progress* edited by R. Wayland, 3-83. Cambridge University Press. DOI: https://doi.org/10.1017/9781108886901.002

Flege, J E., & Davidian, R. 1985. Transfer and developmental processes in adult foreign speech production. *Journal of Applied Psycholinguistic Research* 5, 323-347.

Flege, J. E., & Eefting, W. 1986. Linguistic and Developmental Effects on the Production and Perception of Stop Consonants. *Phonetica* 43, 155-171.

Flege, J. E., Munro, M., & MacKay, I. R. A. 1996. Factors Affecting the Production of Word-initial Consonants in a Second Language. In *Second Language Acquisition and Linguistic Variation* edited by R. Bayley & D. Preston, 47-73. Amsterdam, Philadelphia: John Benjamins Publishing Company.

Flege, J. E., Schirru, C., & MacKay, I. 2003. Interaction between the native and second language phonetic subsystems. *Speech Communication* 40, 467-491.

Kovács, A. M., & Mehler, J. 2009. Cognitive gains in 7-month-old bilingual infants. *Proceedings of the National Academy of Sciences of the United States of America* 106, 6556–60.

Luk, G., Bialystok, E., Craik, F., & Grady, C. 2011. Lifelong bilingualism maintains white matter integrity in older adults. *Journal of Neuroscience* 31, 16808–13.

Marian, V., & Hayakawa, S. 2020. Measuring Bilingualism: The Quest for a "Bilingualism Quotient". *Applied Psycholinguistics* 42(2), 527-548. DOI: https://doi.org/10.1017/S0142716420000533

Peal, E., & Lambert, W. E. 1962. The relation of bilingualism to intelligence. *Psychological Monographs: General and Applied* 76, 1–23.

Piske, T., MacKay, I., & Flege, J. E. 2001. Factors affecting degree of foreign accent in an L2: a review. *Journal of Phonetics* 29, 191-215. DOI: 10.006/jpho.2001.0134

Silva-Corvalán, C., & Treffers-Daller, J. (eds.), 2015. *Language Dominance in Bilinguals: Issues of Measurement and Operationalization*. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9781107375345

Smith, F. 1923. Bilingualism and mental development. *British Journal of Psychology* 13, 271–82.

Τριάρχη-Herrmann, Β. 2000. *Η διγλωσσία στην παιδική ηλικία: Μια ψυχογλωσσική προσέγγιση*. Αθήνα: Gutenberg.

CHAPTER THREE

FACILITATION OF L2 INSTRUCTION WITH DATA FROM AUDITORY-VISUAL SPEECH PERCEPTION RESEARCH

DOĞU ERDENER

Abstract

Speech perception does not just happen in an auditory way but in a visual way as well – we "hear" lip and mouth movements when we have face-to-face communication. Auditory-visual speech perception has been adding a solid dimension to our understanding of speech perception. This new way of looking at speech processing paved the way for contemplating mainstream theories on speech perception to calculate the role of visual aspects of speech. Visual speech information has instantiations on several levels, e.g., physiological, but more importantly, auditory-visual speech processing research has applications in several domains. One such domain over which research has been cumulating in the past two decades is L2 acquisition. In this chapter, the basics of auditory-visual speech perception is presented. Following this, how research in this relatively new area can profitably be used in improving and developing new ways of L2 instruction is described in detail. As well, based on empirical data, thoughts on under what conditions and with what type of languages L2 orthography can beneficially be used – with and without visual speech information – to assist foreign language instruction.

Keywords: audiovisual speech perception, L2 instruction, language acquisition, speech perception, speech production

1. Introduction

One challenge in developing effective foreign language teaching systems and curricula is posed by the paucity of the use of knowledge obtained from psycholinguistic research in language instruction. In other words, even though there is a growing body of research on how humans process language input and how they produce language on the basis of the available perceptual information, the practitioners have still yet to be aware of that knowledge and put it into practical use. This, in fact, applies to pretty much all facets of psycholinguistic enterprise, i.e., phonological, semantic, syntactic, rhetoric, etc. as well as applied. This chapter will focus on the paucity of the link between psycholinguistic research and language instruction hence how it can profitably be used in the domain of foreign language (which will be used interchangeably with "L2", hereafter) instruction.

This chapter will first look at the conventional models of speech perception and then will switch focus to the more recently acknowledged multimodal nature of speech perception, namely, auditory-visual speech perception. Even though, recently, an account of how auditory-visual speech perception research can profitably be used and put into practice in L2 teaching domain was given (Erdener, 2020), this chapter will expand beyond that position and present a broader perspective by highlighting how the traditional models and theories of speech perception should be revised factoring in what we call visual speech information in the form of orofacial – lip and mouth – movements. Following this psycholinguistic focus, the chapter will handle how both knowledge from auditory-only speech perception and auditory-visual speech perception research can be implemented into L2 instruction. This ever-growing knowledge coming from psycholinguistics has instantiations for L2 instruction in a plethora of domains, e.g., pronunciation, comprehension, perception, etc.

Until recently, speech perception has been handled exclusively in an auditory domain. However, while a small number of isolated studies had already shown the facilitative effect of visual speech information in the form of lip and face movements in impoverished listening conditions (e.g., Sumby & Pollack, 1954), ever since the discovery of what is called the *McGurk effect / illusion* (McGurk & MacDonald, 1976), we have come to appreciate this phenomenon as a multimodal, and specifically, as an auditory-visual process. The most significant contribution of the McGurk effect was that we now see that visual speech information is evidently used in clear, uninterrupted listening conditions. This finding has theoretical and

practical implications. The main theoretical point is that the major theories of speech perception until recently endeavoured to explain speech perception from solely an auditory perspective. Of course, speech perception is *primarily* an auditory event and does work in an auditory-only fashion as anybody can anecdotally observe in the case of people who were born with congenital blindness or, simply, when we talk on the phone. However, many of us are engaged in conversations in a face-to-face manner nowadays including – and at an increased rate amid the COVID-19 pandemic at the time of writing this chapter – in online contexts for various reasons – personal, business, education, entertainment and so forth. So, this theoretical issue necessitates that all mainstream theories of speech perception be revised in the context of auditory-visual speech. Furthermore, this theoretical concern is organically tied to practical concerns. After all, science is there to help us understand natural phenomena such as speech perception; however, another major purpose of science is to put discoveries and knowledge into practical use. There are a plethora of uses of auditory-visual speech perception and processing research in the domains of engineering, education, speech therapy, health, etc. A thorough understanding of how humans integrate auditory and visual speech information calls for an interdisciplinary approach to auditory-visual speech perception research and that is what has been happening particularly in the past four decades. The applied area of focus in this chapter is L2 instruction. Before resuming the main theme of the chapter, let us glean over the basics of and mainstream theories in speech perception. This will be followed by the basic physiological aspects of auditory-visual speech perception and will be followed by some key developmental and cross-language issues. Cross-language domain is the junction point before switching to how auditory-visual speech perception research from the lab can be applied to the L2 instruction domain.

1.1 Speech Perception: basics and development from conventional evidence

The original problem posed by research in speech perception since 1950s is how human process and decipher the seemingly single-pieced speech signal which cannot be discerned into discreet units such as phonemes or words (Figure 1). As Licklider (1952) put forward, speech perception research has three dimensions: its physiology, the elements that make up the speech signal and the deciphering of the meaning. Despite this assertion by Licklider nearly 70 years ago, we can confidently say that the enterprise of understanding the nature of speech perception more or less sits on these

three pillars. Hence, a concise snapshot of the state of auditory-visual speech perception research will be presented on the basis of these dimensions before moving onto applications in the foreign language instruction domain. But first, let us have a quick glance at some of the conventional theories of speech perception with a developmental perspective.

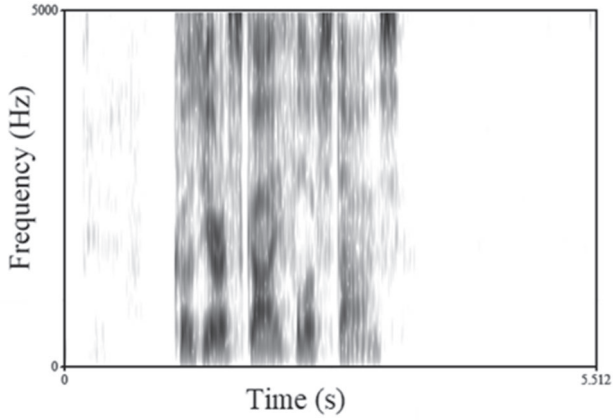


Figure 1. The spectrogram of the utterance "*Because you are a sky full of stars*" spoken by the author of this chapter – a non-native speaker of English. The dark areas represent high energy concentrations denoting vowel productions. Aside from this distinction, continuous speech has hardly discernible components allowing us to tell apart lexical items, i.e., words. How we discern words in running speech has been one of the mind-boggling enterprises of speech perception research since the 1950s.

Understanding the phenomenon of speech perception in its entirety necessitates a multidimensional – hence a multidisciplinary – approach. The conventional theories of speech perception have usually taken a developmental perspective in their endeavour to explain the phenomenon of speech perception and its development. The legitimacy of this approach has been backed up by the very nature of the speech perception process itself as well. For instance, Burnham (2003) explained this via a heuristic in which he suggests that speech perception developmentally (and as a perceptual process) kicks off as an acoustic/phonetic process in the first six months of age. Newborns' acoustically-based, language-general capacity which is characterized by the ability to discriminate most, if not all contrasts from every language in the world presented to them gradually becomes sensitive to the contrasts in the native language (L1, hereafter) environment. This

language-specific tuning is marked by phonologically irrelevant contrasts being attenuated by means of a phonologically based perceptual reorganisation. Hence, in the first year of life, speech perception development follows a pattern from an acoustically-based, language-general fashion to a more phonologically-based, language-specific processing paving the way for the infrastructure that allows for the discrimination of phonologically relevant contrasts that in turn help with semantic contrasts and vocabulary acquisition (for an earlier review see Burnham, 1986). While the developmental pattern remains the same (i.e., from acoustically-based language-general to phonologically-based, language-specific), different linguistic elements go through different chronological points in time such that this trend occurs earliest for vowels sometime between 4 and 6 months of age (Kuhl et al., 1992), for lexical tones around between 6 and 9 months of age (Mattock & Burnham, 2006) and consonants around 7 to 11 months of age (Best, McRoberts, & Goodell, 2001).

Some models of speech perception and development focus on, particularly after the aforementioned period of first year of life, how the phonological space is reorganized in an attempt to explain the ways through which L2 (non-native) phonemes are processed (Best, 1994; Tyler, 2019). During the process that a newborn undergoes in the first 12 months, native phonemes become *phonologically relevant* and the phonetically based perception of non-native phonemes become attenuated. This allows for the formation of the L1 phonological repertoire that leads to the processing of non-native phonemes through what we may refer to as a phonologically based L1 filter. In adulthood, this leads to a sort of "hearing with an accent", hence, speaking with an accent when acquiring an L2 especially after puberty which is presumed to be associated with the end of neural plasticity (Lenneberg, 1967; Vanhove, 2013). Such *hearing with an accent* can simply be exemplified by what can be called "minimal pairs" an example of which is the well-known /r/-/l/ phonemic pair (as in the English words ***r***ake vs ***l***ake). This pair is infamously a significant perceptual challenge for adult native speakers of Japanese in the L2 context (Goto, 1971; Aoyama et al., 2004) unlike what would be predicted for the Japanese infants younger than one year of age. By adulthood both of these phonemes are perceived as the apico-alveolar tap phoneme, /ɾ/ (Aoyama, Flege, Guion, Akahane-Yamada, & Yamada, 2004). However, research has shown that /r/ vs /l/ and other phonemically irrelevant contrasts can be taught to learners to discern them via overt perceptual and productive training by conventional (Aoyama, et al., 2004; Flege, Aoyama, & Bohn, 2021) as well as auditory-visual methods (Massaro & Light, 2003; Shinohara, 2021) a point to which we will return in due course.

While there are theories that were advanced specifically to address L2 acquisition in adulthood such as the Speech Learning Model (SLM) (Flege, 1995; for the revised version, SLM-r, please see Flege, Aoyama, & Bohn, 2021), it stands clear that looking at the L2 acquisiton process from a developmental and multidimensional perspective adds a wealth of information to our theoretical understanding of the issues. This, in turn, has important implications that are of paramount importance to the development of data-based approaches to L2 instruction. Such data is available in both conventional auditory-only and more recently in auditory-visual forms. So let us now see how the available data – as well as future studies – can contribute to the goal of integrating auditory-visual speech perception to the practice of L2 instruction. The critical domain of research here is the *cross-language* studies in auditory-visual speech perception. This is the very domain through which research in auditory-visual speech perception links with L2 instruction.

2. Auditory-Visual Speech Perception

2.1 Basics of auditory-visual speech perception

The legitimacy of making the great shift in speech perception research from the conventional auditory perspective to the auditory-visual one is attested not just by the evergrowing body of behavioural research but also by means of physiological studies. Some of the cortical activities associated with speech perception are also active not just in auditory-visual speech perception but also in visual-only speech conditions, namely, lipreading. This is interesting, because in the absence of auditory speech information, one would instantly assume that seeing speech would only activate the occipital lobe – or the visual cortex, the primary region of brain associated with visual perception. However, as demonstrated by Calvert and her colleagues, when we observe face movements during lipreading, the areas of the brain that are normally active while attending to auditory speech also become activated (e.g., angular gyrus) (Calvert et al., 1997). Further to this, cortical studies also suggest that the visual speech information, when available, interacts and integrates with the auditory input on the cortical level (Jones & Callan, 2003) affecting the eventual percept as we see in the case of the McGurk effect (McGurk & MacDonald, 1976). Such results on both behavioural and cortical levels show us that visual speech information has an indispensible role in auditory-visual speech perception. Its apparent function of clarifying the auditory speech input and, as elucidated below, its usefulness in practical contexts warrants the scrutiny of visual speech

information in L2 instruction. The role of visual speech information surfaces in two domains of auditory-visual speech perception research: cross-language and L2, to the former of which, we now turn our attention.

2.2 Auditory-visual speech perception: cross-language issues

Currently, we know that visual speech information fuses with the heard speech on word (Erdener & Burnham, 2005) and sentence contexts (Sams et al., 1998). This basic knowledge is important as we would like to ensure that the potential use of visual speech information in L2 instruction is evident in all these contexts and not just limited to syllabic level on which most research seems to be based. Following a series of studies, in what she referred to as the *foreign speaker effect*, Sekiyama asserted that when we watch and listen to a talker who – presumably – speaks a foreign language, the magnitude of visual speech influence is greater than when the same individuals watch and listen to a talker speaking their native language. This seems to be the case irrespective of the L1 background (Sekiyama & Tohkura, 1993). However, though, the extent to which visual speech information is used in L2 acquisition appears to be contingent upon the L1-L2 phonological overlapping or coherence. Flege, Munro and Fox (1994) found that, when comparing the vowel pairs /a/-/ɛ/, /a/-/ɪ/, /a/-/ʌ/ and /a/-/æ/ – all of which are phonemically contrastive in English – native English speakers perceived a lot more phonetically-based differences than their English-experienced and non-experienced Spanish-speaking counterparts. In other words, native English speakers were able to discriminate all these contrasts as two distinct phonemes, whereas in a significant number of trials, both experienced and inexperienced native Spanish-speaking learners of English as L2 tended to perceive these pairs as same or variants of a single phoneme, e.g., /a/-/ʌ/ and /a/-/æ/. This and several other studies have shown that the degree of L1-L2 phonological similarity works rather in a more convoluted way than straightforward. When two phonemes are similar across the respective L1 and L2, then the amount of confusion as reported by Flege et al. (1995) would increase as the two phonemes would map onto a single category as is the case with the non-Japanese /r/-/l/ contrast in which case both phonemes map onto the phoneme /ɾ/ (Flege, Takagi, & Mann, 1995). Encompassed in his SLM, Flege elucidates the factors that mediate the phonological relationship between L1 and L2. Some of these are the relative use of L1 and L2 (Flege, 1995; Flege, Frieda, & Nozawa, 1997; Munro, Flege, & Mackay, 1996; Guion, Flege, & Loftin, 2000), age of acquisition (Flege, Munro, & MacKay, 1995), and training (Flege, Takagi, & Mann, 1995). According to James E. Flege, all these factors affect the

way L2 speech production/accent is perceived by native speakers of an L1 – mostly English in his studies. In the meantime, it would be helpful to refer to a recent revision of the SLM by Flege and his colleagues (Flege, Aoyama, & Bohn, 2021).

The "perception" challenge or what was referred to as "hearing with accent" above in the process of L2 acquisition is of utmost importance as it most likely originates from the phonological reorganization in the first year of life at the end of which the L1 phonological repertoire starts to act as a perceptual filter. That, in turn, affects the way we hear and interpret on all levels of processing: phonological, perception, etc. This perceptual filtering process presumably eventuates in an accented speech, which, in some cases affects the legibility in L2 speech. The timeline of the development of this perceptual hence production bias is a matter of debate – first year of life to adolescence as predicted by Lenneberg's (1967) Critical Period Hypothesis (CPH). Turning our attention back to the role of the L1-L2 interaction as extensively studied by James E. Flege in the 1980s and 1990s, one wonders about the role of visual speech information in L2 acquisition – both naturalistic and formal instruction ways. The first known study on the effect of visual speech information coupled with L2 experience was conducted by Werker, Frost and McGurk (1992). In this study, they tested a group of native Quebecois-French speakers with four different levels of English language proficiency: beginning, intermediate, advanced and English-French early bilinguals who grew up speaking both languages on native levels. Also, a group of English-speaking monolinguals were recruited as the control group. They used typical McGurk-style stimuli one of which featured the dental-fricative /ð/ (e.g., ***th***at), a native phoneme in English but not in French. When featured in the visual portion of the McGurk stimuli, its influence on the resultant percept turned out to be contingent upon the level of experience in English: the greater the experience, the bigger the influence such that the perception of the L2-native phonemes increased as L2 proficiency rose (Figure 2). Therefore, the extent to which visual speech information is profitably used in L2 appears to be a function of level of familiarity and experience with the target L2. This is the crux of benefit of the research in auditory-visual speech perception for L2 instruction. Let us now look at the existing studies on the potential use(fulness) of visual speech information.

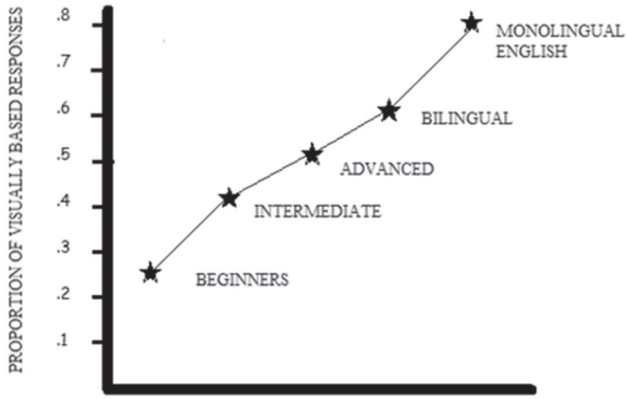


Figure 2. The hypothetical representation of the data from Werker, Frost and McGurk (1992). The proportion of visually-based responses to the McGurk type stimuli in which the visual portion is the fricative-dental /ð/, an English-native phoneme that is not a part of French phonological repertoire. The greater the naturalistic L2 experience with /ð/, the more the influence of visual speech information.

2.3 Auditory-visual speech perception: issues pertaining to L2 instruction

We already stated that speech perception should be and recently has been considered as a multimodal process and that conventional models of speech perception should be revised from this perspective as visual speech information provides perceivers with additional articulatory information. For instance, when we inspect L2 perception studies based on Best's (1994) Perceptual Assimilation Model (PAM), we see that this and other influential models (e.g., the SLM by Flege), contemplate non-native phonological processing in an L2 immersion context (Tyler, 2019). The theoretical strength of PAM (nowadays, PAM-L2), over the years, evolved from being a theoretical model based on experimental work to a more applied orientation to elucidate the L2 acquisition process (Faris, Best, & Tyler, 2016). More recently, the model was tested in the territory of auditory-visual speech perception (Fenwick et al., 2017). The inclusion and adaptation of visual speech information to the conventional speech perception theories such as PAM-L2 is important, as such enterprise will help us lay the foundations of both research and practice of how auditory-

visual speech information will and can be implemented in L2 classrooms. In a nutshell, the revised PAM-L2 makes four assumptions as to how a non-native (or L2) phoneme is assimilated into the L1 phonological space: (a) as a good, acceptable, but not native-like instantiation of the target phoneme; (b) as a speech element; however, not belonging to the native category; (c) as speech but perceived as ambiguously belonging to one or more phonemic category; (d) assimilated as non-speech (Tyler, 2021; Faris et al., 2016; Best, McRoberts, & Sithole, 1988). On the basis of pairwise comparisons, the model makes six further predictions in which two different phonemes are perceived in one of these ways: (a) both phonemes are percieved to belong to two different categories; (b) the two phoemes are percieved as two versions of the same phonemic category; (c) the phonemes are perceived as two different instances of the same category; (d) both are perceived as different whereby one of the pair is perceived as a native phoneme whereas the other remains uncategorised within the native phonological space; (e) both phonemes remain uncategorised within the native phonoogical space; (f) both remain unassimilable. Recently, in a test of PAM-L2 from an auditory-visual speech perspective, Fenwick and her colleagues investigated whether inclusion of visual speech information had a facilitative effect on the discrimination, as opposed to assimilation, of non-native consonants. To test this, native Australian English-speaking monolinguals were presented with Sindhi[1] consonants in auditory-only, visual-only and auditory-visual conditions and asked to rate each target phoneme in terms of how much they sounded like Australian English and if so, how good it sounded in auditory-visual, auditory-only and visual-only conditions. They found that visually based assimilation of L2 phonemes were based on, partly, to the degree of the salience and resemblence of L2 phonemes to L1 phonemes (Fenwick et al., 2017). This supports suggestions by such models as SLM that the greater the resemblance between L1 and L2 variants of a phoneme (e.g., L2 English /æ/ vs L1 Turkish /e/) the more difficult difficult it will be for a new category for the target L2 to be formed (Flege, 1995). However, if the non-native phoneme is distinctly different or sometimes perceived as non-speech, as is the case with Zulu clicks (Best, McRoberts, & Sithole, 1988), then a perceptual categorisation is made relatively easier (Fenwick, Best, Davis, & Tyler, 2017; Flege, Aoyama, & Bohn, 2021).

Intriguingly, the ecological salience – hence the practical use – of visual speech information is more discernable in the L2 context than in L1. For

[1] An Indo-Aryan language predominantly spoken in the Sindhi province of Pakistan and the adjacent areas in India.

instance, in what Sekiyama and her colleagues had referred to as the *foreign speaker effect*, there is implicitly more use of and benefit from visual speech information (Sekiyama & Tohkura, 1993). The abovementioned Werker et al. (1992) study further added to this by demonstrating that the visual salience and the perception of non-native phonemes is a function of the magnitude of L2 experience such that L2 phonemes become more visually discernable with increasing L2 experience. In one sense, the original McGurk effect study revealed this as well such that an alternative visually-influenced response to the typical McGurk stimulus (auditory [ba] + visual [ga]) was "*tha*" as well as "da" (McGurk & MacDonald, 1976). The "tha" response appears to be special to the English L1 acquisition experience whereas it was never the case in other languages contexts wherein the McGurk effect was tested, e.g. Turkish (Erdener, 2015) unless significant experience was evident with English as an L2 (Werker, Frost & McGurk, 1992). While some languages such as Kurdish show relatively more robust visual speech influence, e.g., vs Persian (Asiaee, Kivanani, & Nourbakhsh, 2016), such inter-language differences have yet to be explored.

So far, we mentioned studies that tapped on perceptual aspects within the scope of speech perception. In the past couple of decades, a growing body of research investigated the role of visual speech information and auditory-visual speech integration in L2 with a practical, applied perspective. In an earleir study, Ortega-Llebaria, Faulkner and Hazan (2001) investigated the way native Spanish speakers learning English as an L2 perceived English consonants that are both native and non-native to Spanish. There are a few peculiar phonological differences between Spanish and English with respect to their consonantal repertoires. For instance, native speakers of Spanish usually confuse English-native but Spanish-non-native phonemes such as the English /t/ and /d/ in perception tasks and they usually tend to assimilate both to the Spanish /t/ category. Another example is that the English-native contrast /ð/- /d/ is perceived to be same by native speakers of Spanish with little or no experience in English language and assimilated into the Spanish phonemic category of /d/. It is possible to observe the same phenomenon in other languages and dialects such as Cypriot Greek in phonemic contrasts such as the consonant pairs /p/ vs /b/, /t/ vs /d/, /k/ vs /g/, /θ/-/ð/. Or vowel contrasts such as /æ/-/e/, /ɒ/-/ɔ:/, /u:/-/ʊ/, /ɑ:/-/æ/ (Kkese & Karpava, 2021).

Research so far has shown that perceptual training and raising awareness of these peculiarities do help with more accurate perception hence production. Ortega-Llebaria et al. (2001) tested a group of native Spanish speakers on a series of English-native consonants and vowels in both auditory and auditory-visual training conditions. The target phonemes were presented

using an interactive conversational agent. Following the perceptual training phase of the study, the participants' productions of these target phonemes were analysed. The results showed that auditory-visual training led to significantly reduced number of errors in consonant production. The advantage of auditory-visual training was more apparent when contrasted against the auditory-only form of training. Furthermore, in two different studies, Hazan and her colleagues recruited native speakers of Spanish and Japanese who were learning English as an L2. They presented their participants with visual speech information by means of an auditory-visual training mode in order to see whether their perception – hence production – of the English-native bilabial and labio-dental phonemes, /p/, /b/, and /v/, would improve. The visual speech information was presented in visual-only and audio-visual conditions and they were contrasted against an audio-only control condition. In both studies with native speakers of Japanese and Spanish, a statistically significant advantage for the auditory-visual training condition over the auditory-only one was found and this was especially the case for those relatively difficult L2 contrasts for the speakers of these L1s (Hazan et al., 2005; Hazan et al., 2006).

An example of such challenging contrasts to be discerned by L2 learners is the frequently confused, challenging-for-Spanish-speakers one: /p/ vs /b/. Another notoriously challenging contrast is the /r/ vs /l/ that exists contrastively in several languages such as English, Greek or Turkish but not – and especially – in Japanese (Goto, 1971; Aoyama et al., 2004). As mentioned above, both phonemes are, if we use the PAM-L2 terminology (Tyler, 2019), assimilated by the Japanese flap phoneme /ɾ/. In the light of what we know from developmental studies, we can comfortably predict that this contrast can easily be discriminated by newborn Japanese infants on an acoustic/phonetic basis, native Japanese speakers beyond the first year of life cannot (Burnham, 1986). Following the initial report of the perception of the /r/-/l/ contrast by native speakers of Japanese by Goto (1971), several studies analysed the mechanism of this perceptual and productive difficulty but few looked at how to tackle the perceptual challenge which is a significant one in the applied context of L2 acquisition. One such study was conducted by Aoyama et al. (2004) in which they longitudinally tested child and adult native speakers of Japanese on the English-native /r/ vs /l/, /r/ vs /w/ and other English-native contrasts that are known to be "difficult" for Japanese learners of English. They predicted, on the basis of Flege's SLM, that an early L2 acquisition would not be suppressed by early-forming L1 categories and L2 categories would be formed. This would not be the case for adults whose L2 acquisition began later in life so their perception would be filtered through the well-established L1 phonemic categories hence

leading to an accented, or in some cases, unintelligible production. In the perception experiment, the participants were tested and trained using the odd-one-out method and then they were re-tested afterwards. The results showed native Japanese-speaking children showed improvement on the perception of these difficult contrasts, whereas their adult counterparts did not. The production experiment revaled similar results that children's intellgibility scores significantly improved over the year that seperated the two testing sessions (Aoyama et al., 2004). However, similar studies seem to have produced mixed results. For instance, after testing native English-speakers of five different age groups learning Dutch in the Netherlands, no significant differences with respect to pronunciation were observed and both adults' and older children's vocabulary and sentence production scores were high (Snow & Hoefnagel-Höhle, 1978). Though such differences may be explained by cognitive-developmental factors, sticking with the theme of this chapter, the Aoyama et al.'s (2014) results beg the following question: what if visual speech information is incorporated into such training? Recently, this question was put to test by Shinohara (2021). Over several trials, Shinohara presented a large group of child and adult monolingual speakers of Japanese with several minimal difference word pairs contrasting the /r/ and /l/ phonemes which were controlled for with respect to formant adjustments, particularly, F3 which was set at 2639 Hz for /r/ and 3328 Hz for /l/. These values set the phonemic boundaries. The training stimuli were presented showing the full face of the speaker, and, auditory-only and visual-only training conditions were also added. Subsequent to the training session, participants were given tests of identification and discrimination. According to the results, both adults and children showed statistically significant improvements in both identification and discrimination tasks. Furthermore, the visual-only condition resulted in better performance relative to the auditory-only condition – as has been demonstrated in several other studies earlier on. These results lend support to previous studies that the provision of visual speech information enhances perception (Hazan et al., 2006; Hardison, 2003); hence, it should improve production as well (Erdener & Burnham, 2005). The results reported by Shinohara (2021) also suggest a small, but a technically significant point. When designing foreign language instruction materials, the design should allow the end user – teachers – to make certain adjustments to create variants of the target phonemes and help students to *perceptually* discern the minimal pair differences such as the /r-l/ for Japanese learners (Massaro & Light, 2003), and anecdotally, the /d-ð/ contrast for Turkish and the /ʃ-s/ contrast for mainland Greek (but not Cypriot) learners of English.

As the end of this chapter nears, I would like to talk about another aspect of foreign language instruction: L1-L2 orthographic coherence. The next section visits issues that pertain to how and under what conditions orthographic information may come in handy and when it doesn't – both coupled with and separate from the visual speech information.

2.4 Non-visual aids to deciphering L2 speech: orthography and others

So far the nature, development and the potential use of visual speech information in the form of orofacial movements in L2 instruction have been discussed with reference to the available empirical evidence. Along with the *seen* speech, our speech perception system seems to be able to make use of other sources of information that are not visual. One such source is the sensation of touch. Plant et al. studied the role of touch in speech perception in a case study of a male native speaker of Swedish (GS) who, at the time of study, had been lipreading visually and haptically for over 45 years. GS also had learned English as an L2 via these two methods. His haptic-visual lipreading method was tested for voicing as well as several manners and places of articulation. Results generally showed that when GS placed his hand on the neck of the speaker, his accuracy scores rose significantly compared to just visual lipreading (Plant, Gnosspelius, & Levitt, 2000).

Another way of delivering accurate L2 pronunciation could be rendering the way teachers speak a salient method of instruction. What is referred to as *teacherese* is an example of just that. Teacherese is a form of hyper-articulated speech style that presents itself with exaggerated articulatory movements. The interlocutor – mostly teachers – engages in teacherese during instruction in a classroom setting (Håkansson, 1987). Hyperarticulated speech characterised by visually very noticeable articulatory movements enhances the visual discernability of speech input and leads to a clearer perception (Lees & Burnham, 2005). A noticeable use of hyperarticulated speech with awareness on the side of both instructors and students would presumably be helpful in instilling the L2 material to hear and speak, hence improving perception and production.

The existing evidence warrants the scrutiny of the facilitative and inhibitory effects of the L2 orthography, especially the L1-L2 orthographic coherence. Native speakers of languages with transparent orthographies (characterised by consistent phoneme-grapheme correspondences) such as Turkish, perform phonological awareness tasks better than those with an opaque orthographic background, such as English (Oktay & Aktan, 2002). In a test

of the effect of L1-L2 coherence on word learning in L2, Escudero and her colleagues presented native speakers of Spanish with pairs of pseudowords differing on one phonological element (e.g., a vowel) and pairs differing on multiple elements over two training and testing conditions: auditory and auditory-orthographic. The results revealed that when the target stimuli were orthographically congruent, the overall performance was better than when the target items were incongruent and previous experience in Dutch improved performance (Escudero, Simon, & Mulak, 2013). Erdener and Burnham (2005) tested the effect of L1-L2 congruence together with visual speech information on production using Spanish (transparent orthography) and Irish Gaelic (opaque orthography) pseudowords. They recruited monolingual speakers of Turkish and Australian English, languages with transparent and opaque orthographies, respectively[2]. The participants were tested in four experimental conditions: auditory-only; auditory-visual, in which the target stimuli were presented with the talker's face; auditory-visual-orthographic; and auditory-orthographic. The task was to repeat the orally reproduce stimuli. It was found that when the L1 and L2 orthographies were congruent, i.e., the "learners" were Turkish speakers and the target language was Spanish, there were significantly fewer errors than when the target language was Irish – an unconventionally opaque orthography – in the conditions featuring orthographic information (i.e., auditory-visual-orthographic and auditory-orthographic).

Further to supporting the previous studies on the advantage of providing visual speech information, the orthographic information was handy when both L1 and L2 were orthographically transparent languages, allowing for deciphering the phonological structure of the target pseudowords. These results suggest that when native speakers of languages with transparent orthographies (e.g., Italian) learn languages with comparably transparent orthographies (e.g., Turkish), the sound-letter correspondences become more salient and thus can be exploited for instructional purposes whereas when the L1 of a learner has an opaque orthography (e.g., English), then

[2] It should be noted here that English language has an unusually opaque orthography both in terms of sound-to-letter and letter-to-sound correspondences, namely feedforward and feedback consistencies (for an elucidation of these concepts, see (Ziegler, Jacobs, & Stone, 1996). Even though Irish orthography may be conceived as a transparent one, the letter-to-sound correspondences deviate from the conventional forms, e.g., the letter clusters "mh" and "bh" correspond to the phoneme /v/.

teaching from print, at least in early stages of instruction, may not be as fruitful (Erdener & Burnham, 2005; Escudero, Simon, & Mulak, 2013).

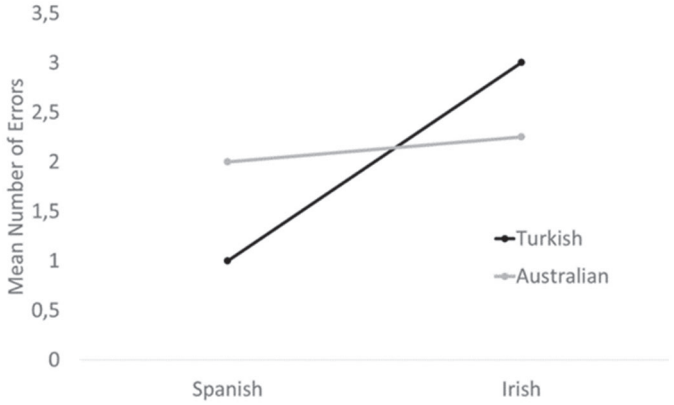


Figure 3. The correct score mean values Spanish and Irish production tasks by monolingual Turkish and Australian speakers in auditory-visual-orthogtraphic and auditory-orthographic experimental conditions (from Erdener & Burnham, 2005)

3. Concluding Remarks

This chapter endeavoured to present the existing evidence from psycholinguistics on auditory-visual speech perception, its nature and development, and how this knowledge can profitably be used in the context of L2 instruction. Even though ever since the late 1990s, a significant body of research has been accumulating, there remains a great deal of research to be done. A glean over the literature would reveal that an overwhelming majority of studies are still dominated by the context of European languages, mostly, English, and to some extent other languages such as Spanish. The existing evidence is in need of further work by means of two strategies. First, conventional speech perception models tapping on L2 matters, such as PAM-L2 (Best & Tyler, 2007), should continue to be tested in the context of other languages. Partially serving that strategy, Kömürcü and Yıldız (2011) tested child native speakers of Turkish on how they perceived a series of English consonants using PAM as the theoretical framework. However, more studies with an array of age groups are required from different languages to test these powerful models factoring in visual speech information (Fenwick et al., 2017). This will constitute the second

strategy that needs to be embraced and more languages – along with their orthographies – must be scrutinised as both L1s and L2s.

Parallel to what is suggested above, the training of teachers must feature curricula that incorporate knowledge from psycholinguistic laboratories more frequently; thus more interdisciplinary work should be encouraged and conducted. Not only will this increase the volume of knowledge and widen our perspectives as L2 researchers, but it will also raise awareness among L2 instructors of the existing literature on general speech perception as well as auditory-visual speech perception; hence more theoretical and practical materials to use in the classroom. Such tools will surely be needed more as we are drawn towards teaching and learning via online methods. The foreseen and suggested interdisciplinary work is surely not limited to psychologists and foreign language instruction researchers. On the contrary, it also extends to a broader audience that will incorporate those who are already working in different domains of auditory-visual speech processing, such as engineers, linguists, etc. Nevertheless, ever since the early 2000s, research in speech perception has grown, and branched into areas of application – a specific one being L2 instruction necessitating the use of technology. One such example is the interactive animated agent named "Baldi" developed by Massaro and his colleagues (Massaro, 1987; Massaro & Palmer, 1998) as specifically exemplified by an attempt to tackle the infamous /r-l/ contrast challenge by the native speakers of Japanese (Massaro & Light, 2003). Large scale projects have been underway to develop ways of making practical use of knowledge from auditory-visual speech processing research; e.g., the international and interdisciplinary *Thinking Head* project (Herath et al., 2010).

References

Aoyama, Katsura, James Emil Flege, Susan G. Guion, Reiko Akahane-Yamada, and Tsuneo Yamada. 2004. "Perceived phonetic dissimilarity and L2 speech learning:the case of Japanese /r/ and English /l/ and /r/." *Journal of Phonetics* 32: 233-250. doi:10.1016/S0095-4470(03)00036-6.

Asiaee, Maral, Nina Hosseini Kivanani, and Mandana Nourbakhsh. 2016. "A Comparative Study of McGurk Effect in Persian and Kermanshahi Kurdish Subjects." *Language Related Research* 7 (3): 2-18.

Best, Catherine T. 1994. *Learning to Perceive the Sound Pattern of English.* Vol. 9, in *Advances in Infancy Research*, by Carolyn Rovee-Collier and Lewis P. Lipsitt, 217-305. Norwood, NJ: Ablex.

Best, Catherine T., and Michael D. Tyler. 2007. "Nonnative and second-languagespeech perception: Commonalities and complementaries. ." In *Language experience in second language speech learning: In honor of James Emil Flege*, by O.-S. Bohn and M. J. Munro, 13-34. Amsterdam, The Netherlands: John Benjamins.

Best, Catherine T., Gerald W. McRoberts, and Elizabeth Goodell. 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. doi:10.1121/1.1332378.

Best, Catherine T., Gerald W. McRoberts, and Nomathemba Sithole. 1988. "Examination of Perceptual Reorganization for Nonnative Speech Contrasts: Zulu Click Discrimination by English-Speaking Adults and Infants." *Journal of Experimental Psychology Human Perception & Performance* 14 (3): 345-360. doi:10.1037//0096-1523.14.3.345.

Burnham, Denis K. 1986. "Developmental loss of speech perception: Exposure to and experience with a first language." *Applied Psycholinguistics* 7 (3): 207-239. doi:https://doi.org/10.1017/S0142716400007542.

Burnham, Denis. 2003. "Language specific speech perception and the onset of reading." *Reading and Writing: An Interdisciplinary Journal* 16: 579-609.

Burnham, Denis. 2003. "Language specific speech perception and the onset of reading." *Reading and Writing: An Interdisciplinary Journal* 16: 579-609.

Calvert, Gemma A., Edward T. Bullmore, Michael J. Brammer, Ruth Campbell, Steven C. R. Williams, Philip K. McGuire, Peter W. R. Woodruff, Susan D. Iversen, and Anthony S. David. 1997. "Activation of auditory cortex during silent lipreading." *Science* 276: 593-596.

Erdener, Doğu. 2020. "Second Language Instruction: Extrapolating From Auditory-Visual Speech Perception Research." In *Handbook of Research on Bilingual and Intercultural Education.*, by M.E. Gómez-Parra and C.A.H. Abril, 105-124. IGI Global. doi:10.4018/978-1-7998-2588-3.

Erdener, Doğu. 2015. "Türkçede McGurk İllüzyonu / The McGurk Illusion in Turkish." *Turkish Journal of Psychology* 30: 19-27.

Erdener, Doğu, and Denis Burnham. 2005. "The Role of Audiovisual Speech and Orthographic Information in Nonnative Speech Production." *Language Learning* 55 (2): 191-228. doi:10.1111/j.0023-8333.2005.00303.x.

Escudero, Paola, Ellen Simon, and Karen E. Mulak. 2013. “Learning words in a new languıage: orthography doesn't always help.” *Bilingualism: Language and Cognition* 1-12.

Faris, Mona M., Catherine T. Best, and Michael D. Tyler. 2016. “An examination of the different ways tha tnon-native phones may be perceptually assimilated as uncategorized.” *Journal of the Acoustical Society of America* EL1-EL5. doi:10.1121/1.4939608.

Fenwick, Sarah E., Catherine T. Best, Chris Davis, and Michael D. Tyler. 2017. “The influence of auditory-visual speech and clear speech on cross-language perceptual assimilation.” *Speech Communication* 92: 114-124. doi:10.1016/j.specom.2017.06.001.

Flege, James Emil. 1995. “Second Language Speech Learning Theory, Findings, and Problems.” In *Speech Perception and Linguislic Experience: Issues in Cross-Language Research*, by Winifred Strange, 233-277. Timonium, MD: York Press.

Flege, James Emil, Elaina M. Frieda, and Takeshi Nozawa. 1997. “Amount of native-language (L1) use af fects the pronunciation of an L2.” *Journal of Phonetics* 25: 169-186.

Flege, James Emil, Katsura Aoyama, and Ocke-Schwen Bohn. 2021. “The Revised Speech Learning Model (SLM-r) Applied.” In *Second Language Speech Learning: theoretical and empirical progress*, by Ratree Wayland, 84-118. Cambridge: Cambridge University Press. doi:10.1017/9781108886901.003.

Flege, James Emil, Murray J. Munro, and Robert Allen Fox. 1994. “Auditory and categorical effects on cross-language vowel perception.” *The Journal of Acoustical Society of Amedc* 95 (6): 3623-3641.

Flege, James Emil, Murray Munro, and Ian MacKay. 1995. “Factors affecting degree of perceived foreign accent in a second language.” *Journal of the Acoustical Society of America* 97: 3125-3134.

Flege, James Emil, Naoyuki Takagi, and Virginia Anne Mann. 1995. “Japanese Adults can Learn to Produce English /R/ and /l/ Accurately.” *Language and Speech* 38 (1): 25-55. doi:10.1177/002383099503800102.

Goto, Hiromu. 1971. “Auditory perception by normal Japanese adults of the sounds “L” and “R”.” *Neuropsychologia* 9 (3): 317-323. doi:10.1016/0028-3932(71)90027-3.

Guion, Susan G., James Emil Flege, and Jonathan D. Loftin. 2000. “The effect of L1 use on production in Quichua - Spanish bilinguals.” *Journal of Phonetics* 28: 27-42. doi:10.1006/jpho.2000.0104.

Håkansson, Gisella. 1987. *Teacher Talk: How Teachers Modify their Speech when Addressing Learners of Swedish as a Second Language.* Lund, Sweden: Lund University Press.

Hardison, Debra. 2003. "Acquisition of second-language speech: Effects of visual cues, context, and talker variability." *Applied Psycholinguistics* 24: 495-522. doi:10.1017.S0142716403000250.

Hazan, Valerie., A. Sennema, A. Faulkner, M.Ortega-Llebaria, M Iba, and H. Chunge. 2006. "The use of visual cues in the perception of non-native consonant contrasts." *Journal of the Acoustical Society of America* 119 (3): 1740-1751.

Hazan, Valerie., Anke Sennema, Marta Ortega-Llebaria, Midori Iba, Andrew Faulkner. 2005. "The effect of audiovisual perceptual training on the perception and production of consonants by Japanese learners of English." *Speech Communication* 47: 360-378.

Herath, D. C., Christian Kroos, Catherine J. Stevens, Lawrence Cavedon, and Prashan Premaratne. 2010. "Thinking Head: Towards Human Centred Robotics." *Proceedings of the 11th International Conference on Control, Automation, Robotics and Vision (ICARCV2010).* Singapore. 2042-2047. doi:10.1109/ICARCV.2010.5707899.

Jones, Jeffery A., and Daniel E. Callan. 2003. "Brain activity during audiovisual speech perception: An fMRI study of the McGurk effect." *Cognitive Neuroscience and Neuropsychology* 14 (8): 1129-1133. doi:10.1097/01.wnr.0000074343.81633.2a.

Kkese, Elena and Sviatlana Karpava, S. 2021. "Challenges in the perception of L2 English phonemes by native speakers of Cypriot Greek." *Journal of Monolingual and Bilingual Speech* 3(1):1–39. doi: 0.1558/jmbs.15362.

Kömürcü, Buket., and Yasemin Yıldız. 2011. "An assessment of the auditory speech perceptual skills of young Turkish learners of English: perceptual assimilation model." *Procedia – Social and Behavioral Sciences* 15: 3511-3515.

Kuhl, Patricia K., Karen A. Williams, Francisco Lacerda, Kenneth N. Stevens, and Björn Lindblom. 1992. "Linguistic experience alters phonetic perception in infants by 6 months of age." *Science* 255 (5044): 606-608. doi:10.1126/science.1736364.

Lees, Nicole, and Denis Burnham. 2005. "Facilitating speech detection in style!: the effect of visual speaking style on the detection of speech in noise." In *Proceedings of the Auditory-Visual Speech Processing International Conference*, by E. Vatikiotis-Bateson, D. Burnham and S.Fels, 23-28. Vancouver, BC, Canada: Causal Productions.

Lenneberg, Eric H. 1967. *Biological foundations of language.* New York: Wiley.

Licklider, J.C.R. 1952. "On the Process of Speech Perception." *The Journal of the Acoustical Society of America* 24: 590. doi:https://doi.org/10.1121/1.1906938.

Massaro, Dom W., and Joanna Light. 2003. "Read my tongue movements: bimodal learning to perceive and produce non-native speech /r/ and /l/." In *Proceedings of Eurospeech (Interspeech), 8th European Conference on Speech Communication and Technology*, 1-4. Geneva, Switzerland.

Massaro, Dominic W. 1987. *Speech Perception By Ear and Eye: A Paradigm for Psychological Inquiry 1987.* New York, NY: Lawrence Erlbaum Associates, Inc.

Massaro, Dominic W., and Stephen E. Palmer. 1998. *Perceiving Talking Faces: From Speech Perception to a Behavioral Principle.* Cambridge, MA: MIT Press.

Mattock, Karen, and Denis Burnham. 2006. "Chinese and English Infants' Tone Perception: Evidence for Perceptual Reorganization." *Infancy* 10 (3): 241-265. doi:https://doi.org/10.1207/s15327078in1003_3.

McGurk, Harry, and John MacDonald. 1976. "Hearing lips and seeing voices." *Nature* 264: 746-748.

Munro, Murray J., James Emil Flege, and Ian R.A. Mackay. 1996. "The effects of age of secondlanguage learning on the productionof English vowels." *Applied Psycholinguistics* 17: 313-334.

Oktay, Ayla and Ebru Aktan. 2002. "A cross-linguistic comparison of phonological awareness and word recognition in Turkish and English." *International Journal of Early Years Education* 10: 37-48.

Ortega-Llebaria, Marta, A, Andrew Faulkner, and Valerie Hazan. 2001. "Auditory-visual L2 speech perception: effects of visual cues and acoustic-phonetic context for Spanish learners of English." *Speech, Hearing and Language: work in progress* 13: 39-51.

Plant, Geoff, Johan Gnosspelius, and Harry Levitt. 2000. "The Use of Tactile Supplements in Lipreading Swedish and English: A Single-Subject Study." *Journal of Speech, Language, and Hearing Research* 43: 172-183. doi:10.1044/jslhr.4301.172.

Sams, Mikko, Petri Manninen, Veiko Surakka, Pia Helin, and Ritta Kättö. 1988. "McGurk effect in Finnish syllables, isolated words, and words in sentences: Effects of word meaning and sentence context." *Speech Communication* 26 (1-2): 75-87.

Sekiyama, Kaoru, and Yoichi Tohkura. 1993. "Cross-language differences in the influence of visual cues in speech perception." *Journal of Phonetics* 21: 427-444.

Shinohara, Yasuaki. 2021. "Audiovisual English /r/−/l/ identification training for Japanese-speaking adults and children." *Journal of Speech, Language, and Hearing Research* 1-10. doi:10.1044/2021_JSLHR-20-00506.

Snow, Catherine E., and Maria Hoefnagel-Höhle. 1978. "The critical period for language acquisition: Evidence from second language learning." *Child Development* 49: 1114-1128.

Sumby, William H., and Irwin Pollack. 1954. "Visual contribution to speech intelligibility in noise." *Journal of the Acoustical Society of America* 26: 212-215. doi:https://doi.org/10.1121/1.1907309.

Tyler, Michael Douglas. 2019. "PAM-L2 and Phonological Category Acquisition in the Foreign Language Classroom." In *A Sound Approach to Language Matters: In Honor of Ocke-Schwen Bohn*, by A. M. Nyvad, M. Hejná, A. Højen, A. B. Jespersen and M. H. Sørensen, 607-630. Aarhus, Denmark: Aarhus University. doi:10.7146/aul.322.218.

Tyler, Michael Douglas. 2021. "Phonetic and phonological influences on the discirmination of non-native phones." In *Second language speech learning: theoretical and empirical progress*, by R. Wayland, 157-174. Cambridge, UK: Cambridge University Press.

Vanhove, Jan. 2013. "The Critical Period Hypothesis in Second Language Acquisition: A Statistical Critique and a Reanalysis." *PLOS Onde* e69172. doi:10.1371/journal.pone.0069172.

Werker, Janet F., Peter E. Frost, and Harry McGurk. 1992. "La langue et les levres:Cross-language influences on bimodal speech perception." *Canadian Journal of Psychology* 46: 551-568. doi: 10.1037/h0084331.

Ziegler, Johannes C., Arthur M. Jacobs, and Gregory O. Stone. 1996. "Statistical analysis of the bidirectional inconsistency of spelling and sound in French." *Behavior Research Methods, Instruments, & Computers* 28 (4): 504-515.

CHAPTER FOUR

THE ARCADIAN GREEK SYLLABLE

PHOTINI COUTSOUGERA

Abstract

This paper seeks to uncover the patterns that lie at the base of the syllable of Arcadian Greek in an effort to understand its prosody in more depth. An inescapable comparison with the Standard Modern Greek syllable is also attempted. At first glance, the Arcadian Greek syllable looks quite different from that of Standard Modern Greek, an open-syllable language par excellence, as portrayed in the literature. This is because the Arcadian Greek high vowels lose their syllabicity when unstressed and undergo different phonological processes, thus yielding interesting consonantal sequences and a variety of codas both word-finally and word-internally. However, underlying onsets and codas behave and pattern differently from derived, non-underlying ones and, in order to determine their well-formedness, it was necessary that the two were treated separately. The Arcadian Greek syllable is examined and analyzed on the basis of (a) Kahn's (1976) syllabification model based on the onset maximization principle and (b) sonority requirements and the Sonority Sequencing Principle.

Keywords: Arcadian Greek, Greek dialects, syllable, sonority, onset maximization

1. Introduction

Arcadian Greek is a variety of Modern Greek spoken in the region of Arcadia, located in the centre of the Peloponnese, Greece. In the present study, we use the term Arcadian Greek to refer specifically to the variety spoken in the mountainous Northeastern part of Arcadia. Until recently, the prevailing view had been that Standard Modern Greek (henceforth SMG) was predominantly based on Peloponnesian varieties (among others, Newton,

1972; Babiniotis, 1979, 1998; Kontossopoulos, 2001/1994; Browning, 1995; Tonnet, 1995; Horrocks, 2010), which led to the belief that the latter were resembling SMG too closely to be worthy of study or research (Newton, 1972: 14). A notable exception is the work of Pantelidis (2001, 2007, 2015). In absence of studies on the wider region of Arcadia (see Coutsougera, 2019), we will be using the term Arcadian Greek broadly, although in awareness that there may well be a degree of regional variation as far as this variety is concerned. This remains to be investigated.

Arcadian Greek, as all nonstandard varieties of Greek, has undergone considerable levelling, phonologically and morphologically as well as lexically, for well over a century. This dialect would have still been intact at around the end of the 19th century and the beginning of the 20th century, when it would have been spoken primarily or exclusively by monodialectal speakers (according to Trudgill, 2003, regional dialects were still fully vibrant all over Greece at around that time). However, it is still preserved in the speech of older people who have lived in Arcadia their entire lives and almost all of whom are bidialectal in Arcadian Greek and SMG. It is this phonology that the present study is interested in and not that of levelled Arcadian Greek currently spoken by younger people.

2. The vowels and consonants of Arcadian Greek

Arcadian Greek and SMG share the same five-vowel system /i e a o u/ (for SMG vowels, see Joseph & Philippaki-Warburton, 1987; Mackridge, 1985; Arvaniti, 1999; Sfakianaki, 2002). It is possible that minor acoustic[1] differences between the vowels of SMG and Arcadian Greek might exist, even though these differences are not auditorily distinct.

With respect to its consonantal system, Arcadian Greek differs from SMG in that it possesses at least phonetic palato-alveolar affricates: *tʃ, dʒ*[2]. More specifically, the velars /k/ and /g/ are fronted into [tʃ] and [dʒ] before the front vowels /i/ and /e/ (ex. a-d). Additionally, [tʃ] and [dʒ] can surface as allophones of /ts/ and /tz/ before unstressed [i] in morpheme boundaries (where /i/ may be prevocalic or word-final: /ts-iV/, /dz-iV/ or /ts-i#/, /dz-

[1] See Fourakis et al. (1999) for an account of the acoustic characteristics of the Greek vowels.

[2] It is questionable whether these are palato-alveolar or alveolo-palatal, tɕ/dʑ (see Manolessou & Pantelidis, 2013), an issue which must be resolved with the help of electropalatography.

i#/) (ex. e-h):

(1)

	Underlying form	Surface form	Gloss	Surface form	Gloss
a.	/ka**k**-i/	[kaˈ**tʃ**i]	‘bad.FEM.’	cf. [kaˈ**k**os]	‘bad.MASC.’
b.	/mermi**g**-a/	[merˈmi**dʒ**a]	‘ants’	cf. [mermiˈ**g**oɲ]	‘big ant’
c.	/ka**k**-es/	[kaˈ**tʃ**es]	‘bad.FEM.PL.’	cf. [kaˈ**k**os]	‘bad.MASC.’
d.	/efe**g**-e/	[ˈefe**dʒ**e]	‘it was shining’	cf. [ˈfe**g**o]	‘I shine’

	Underlying form	Surface form	Gloss	Surface form	Gloss
e.	/pe**ts**-ia/	[peˈ**tʃ**a]	‘rags’	cf. [peˈ**ts**i]	‘rag’
f.	/nera**dz**-ia/	[neraˈ**dʒ**a]	‘bergamot tree’	cf. [neraˈ**tz**ula]	‘small bergamot tree’
g.	/kori**ts**-i/	[koˈri**tʃ**]	‘girl’	cf. [koˈri**ts**aros]	‘big girl’
h.	/nera**dz**-i/	[neˈra**dʒ**]	‘bergamot’	cf. [neraˈ**tz**ula]	‘small bergamot tree’

At the same time, *tʃ* and *dʒ* also appear in loanwords in environments other than /i/ and /e/ (e.g. [ˈtʃupa] ‘girl, lass’, [ˈdʒami] ‘pane’ etc.). However, in the vast majority of cases, these are phonetic occurrences derived from underlying /k/, /g/ or /ts/, /dz/. In an effort to identify relevant minimal pairs, we realized that there is a scarcity of them: [kaˈtsula] ‘cat’ vs. [kaˈ tʃula] ‘hood’ was the only one we could find, while we were unable to identify any examples involving the pair dz/dʒ. Despite their distribution, it is still uncertain as to whether *tʃ* and *dʒ* are fully established phonemes of Arcadian Greek.

In Arcadian Greek - as well as in SMG - apart from /k/ and /g/, the remaining dorsals, /x/ and /ɣ/, are also palatalized into [ç] and [ʝ] before the front vowels /i/ and /e/. Additionally, all non-rhotic coronals, namely the sibilants /s/, /z/ and the sonorants /l/, /n/ get palatalized (softened) into [ʃ], [ʒ] and [ʎ], [ɲ] respectively before the front vowel /i/, where /i/ is fully absorbed by the coronal (for Arcadian Greek high vowels, see Coutsougera, 2019). Another point in which SMG and Arcadian Greek differ is that [ʃ] and [ʒ], as surface forms of /s/ and /z/, do not feature in the former. Even though *ʃ* and *ʒ* may also appear in a small number of loanwords in environments other than /i/, we will not consider them to be candidates for phonemic status. The same applies to *ʎ, ɲ* in SMG, in which they are not awarded phonemic status. For a thorough account of the complex topic of palatalization in Greek, see

Baltazani and Topintzi (2012) and Topintzi and Baltazani (2016).

Finally, this system has no phonemic diphthongal vowels, although a plethora of so-called 'light diphthongs' surface as a result of semivocalization. These are phonetic manifestations of underlying /Vi/ vowel sequences, where the high front vowel /i/ gets semivocalized and forms a light diphthong with the vowel preceding it. This applies intermorphemically, across morpheme boundaries and across word boundaries (e.g. /le-i/ > [ˈlej] 'say.3SG.', /voiθ-os/ > [vojˈθos] 'helper', /me iðe/ > [ˈmejðe] 's/he saw me', etc.). However, there are several examples (all non-SMG) in which a light diphthong appears not to have been derived from /i/, or others for which semivocalization is not a synchronic process as they are already recorded with a light diphthong in the *Lexicon of Medieval Greek Demotic Literature 1100–1669*[3] by Kriaras (e.g. [lejmoˈɲa] 'lemon tree', [ˈklajma] 'crying' etc.).

3. The Arcadian Greek syllable

As Malikouti-Drachman (2001) points out in her state-of-the-art article on Greek Phonology, published in the Journal of Greek Linguistics, the view that Greek historically evolved from a closed syllable language into an open syllable one was prevalent in the 1970s and 1980s (see Setatos, 1974,1987). Later analyses[4] of the Greek syllable (notably Malikouti-Drachman, 1984; Kappa, 1996, 1997a, 1997b) also portrayed Greek as a language that showed a strong preference for open syllables either on the basis of maximal onset requirements or sonority requirements. Gratsouni and Topintzi (2020) have recently challenged this view. More specifically, they questioned the syllabification of word-medial obstruent+nasal clusters as tautosyllabic in SMG - and therefore assigned to the syllabic onset - by appealing to native speakers' intuitions who indeed showed a preference to assigning them to two adjacent syllables and thus spanning the syllable boundary (e.g. [riθ.ˈmos] as opposed to [ri.ˈθmos] 'rhythm').

In the present study, we will provide an analysis of the internal structure of the Arcadian Greek syllable from the perspective of onset maximization and

[3] In the Dictionary of Northeastern Arcadian Greek (Coutsougera, forthcoming), over one third of its 4000 entries are also present in E. Kriaras's (1969-2016) medieval dictionary with the same or different meaning and/or different phonological, morphological or syntactic features. This is the reason why Kriaras's dictionary is frequently referred to in this study.

[4] See also Karzi (2000) for an alternative account of the SMG syllable.

sonority. To that end, we will study the syllable margins of Arcadian Greek, primarily focusing on the syllabification of its biconsonantal sequences[5] which present the greatest interest. As far as simple onsets are concerned, any segment from the consonant inventory of the dialect can occupy this position.

The grid below displays all legal word-initial and word-medial biconsonantal sequences attested in Arcadian Greek. The consonants displayed vertically are combined with those across the top of the grid and the resulting sequence is displayed in the relevant individual cell. The grid also essentially demonstrates the consonantal system of Arcadian Greek, which comprises a series of voiced and voiceless stops, fricatives, alveolar affricates, nasals, liquids and a palatal glide. Palato-alveolar affricates are not included for the reasons explained in section 2 above. Following a recent analysis of Topintzi and Baltazani (2016), in which the palatal glide has been granted phonemic status in SMG, we have also included /j/ in our consonant inventory of Arcadian Greek, as our view is that the same premises that apply to SMG also apply to Arcadian Greek.

[5] We will be using the term ‘sequence’ to refer to a string of unsyllabified consonants and the term ‘cluster’ to refer to syllabified ones.

(2)

first consonant \ second consonant	p	b	t	d	K	g	f	v	θ	ð	s	Z	x	⊗	ts	dz	m	n	l	r	j
p											**ps**								**pl**	**pr**	*pç*
b																			**bl**	**br**	*bʝ*
t																				**tr**	*tç*
d																				**dr**	*dʝ*
k											**ks**								**kl**	**kr**	
g																			**gl**	**gr**	
f			**ft**		Fk								*fk*					fn	**fl**	**fr**	*fç*
v										**vð**				**vɣ**					**vl**	**vr**	*vʝ*
θ																			*xl*	**θr**	*θç*
ð																				**ðr**	*ðʝ*
s	**sp**		**st**		**sk**		**sf**		*st*				*sk*		*stʃ*						
z		**zb**				**zg**		**zv**						**zɣ**			**zm**				

x			**xt**						*xt*									xn	**xl**	**xr**	
⊗										**ɣð**								**ɣn**	**ɣl**	**ɣr**	
ts																					
dz																					
m								*b/v*										mn?			*mj*
n									nθ?	*d*		*Rz*									
l	ʎp	ʎb	ʎt	ʎd	ʎk	?	*rf*	*rv* *rb*	*rθ*	*r*					ʎts	?	ʎm *rm*				
r	rp	rb	rt	rd	Rk	rg	rf	rv	rθ	rð	rs	Rz	rx	rɣ	rts	rdz	rm	rn	rl		*rj*
j																					

The consonantal sequences in bold are attested both word-initially and word-medially. The sequences in regular print are exclusively word-medial, while those in italics are non- underlying. Finally, the ones in smaller print are output sequences/segments which surface in place of illegal ones which have been banned. Sequences with a question mark next to them are contestable, while the question mark alone denotes a possible accidental gap.

3.1 Syllabification operations: Onsets

In an effort to compute the syllabification of Arcadian Greek, we will first consider the maximal onset principle (MOP), following Kahn's (1976) bottom-up syllabification algorithm. We will then consider a top-down templatic approach to syllabification (Selkirk, 1982) by filling in the slots of a template by virtue of sonority requirements.

The former is a rule-based approach which builds nuclei, onsets and codas (only if a language permits codas in its syllables) with the help of structure building rules. This approach prioritizes onsets over codas by ordering the onset rule before the coda one. As a result, an intervocalic consonant will always be assigned to the onset. The rule applies iteratively by scanning all intervocalic consonants from right to left and assigning each one of them to the onset of the syllable to the right as long as the resulting cluster is also found word-initially. That way word structure is mirrored in our syllables.

To that end, we first need to establish the consonantal sequences which are attested in Arcadian Greek. Then we will attempt to syllabify them. A first observation to be made is that none of the consonantal sequences displayed in the grid are attested exclusively word-initially, namely they are not restricted to the periphery of the word. Word-initial consonantal sequences which do not feature word-medially are often syllabified in separate syllables, with the word-peripheral consonant often being considered extrasyllabic or extraprosodic (Appendix) and therefore not a part of the onset. This does not apply to our case. On the contrary, all word-initial consonantal sequences are also attested word-medially while, on the other hand, there are word-medial ones which are not attested word-initially. In other words, word-initial sequences are a subset of word-medial ones.

The fact that word-initial sequences are a subset of word-medial ones would, at least at first glance, favour Kahn's syllabification model. Since our syllable onsets mirror word onsets, MOP does not get violated and accounts for their syllabification quite successfully. According to this model, all sequences in bold would have to be syllabified tautosyllabically (i.e. as onsets of the same syllable) and all word-medial sequences would have to be syllabified heterosyllabically (i.e. as coda and onset of two adjacent syllables). If we look at the make-up of the tautosyllabic clusters in the second last and third last columns versus that of the heterosyllabic ones in the bottom second last and third last row of the grid, one notices that they largely mirror each other. Essentially, all tautosyllabic clusters are obstruent+liquid, while all heterosyllabic ones are liquid+obstruent or

liquid+sonorant (liquid+anything goes). The rhotic /r/ is clearly the preferred word-medial coda in the dialect and some underlying lateral+C sequences also surface as [r]+C (e.g. [vorˈvos] ‘bulb’ cf. SMG [volˈvos], etc.). Homorganic obstruent+lateral clusters are disallowed (i.e. *tl, *dl, *θl, *ðl).

Observing now the remaining columns and rows, we can see that fricatives, as first members of clusters, behave differently from other cluster-first consonants in that they combine with other obstruents (i.e. stops and other fricatives). Obviously, they also combine with liquids just like all other cluster-first consonants. However, all these obstruent+obstruent sequences are unproblematic as they are also attested word-initially and are therefore assigned to the onset position. They would be problematic for this model if we had to split them up and assign obstruent codas, which are not optimal for Arcadian Greek, as we will see in section 3.2 below. The sequences whose second part is atypically a nasal, i.e. *fn, xn, ɣn*, *zm* and *mn*, require further discussion. To begin with, the sequences *ɣn*[6] and *zm* are tautosyllabic as they feature in the word-initial onset position (e.g. [ˈɣnora] ‘recognition’, [ˈɣnomi] ‘opinion’ etc. or [ˈzmiɣo] ‘mix’ etc.). The members of these clusters agree in voicing and are also dissimilated in terms of continuancy; two requirements that SMG obstruent clusters must follow (see Drachman and Malikouti-Drachman, 1997). The sequence *mn*, where two nasals are combined, is often simplified word-medially (e.g. [skaˈmi] vs. SMG [skaˈmni] ‘stool’), even though not always (e.g. [ˈʎimna] ‘lake’, cf. SMG [ˈlimni]). It is nevertheless infrequent and mainly found in SMG onsets of words which are widely used in the dialect as well (e.g. [mniˈmosino] ‘memorial’, etc.). The simplifications show that *mn* is not ideal as a cluster, but the impact of SMG may suggest that it is best assigned in the syllabic onset. Similarly, *xn* is often simplified word-medially (e.g. [ˈrino], cf. SMG [ˈrixno] ‘drop’, [ˈftçano] ‘make, do’, cf. SMG [ˈftçaxno] etc.), even though not always (e.g. [aˈxnari] ‘trace, footprint’, etc.). However, the fact that *xn* is also found in word-initial onsets, would point towards tautosyllabicity in word-medial position. Finally, *fn*, which is never word-initial, would have to be syllabified in two separate syllables (e.g. [ˈaksaf.na] ‘suddenly’). An obstruent coda is not ideal for the dialect (see section 3.2 below) but we are left with no choice.

[6] Note that words starting with *ɣn* have variants with an epenthesized /e/ in the medieval dictionary (e.g. [eɣnora] ‘recognition’).

Finally, the problematic sequence *nθ* is found specifically in the word /anθos/ > [an.ˈθos] ‘flower’ and its many derivatives but gets simplified in other examples (e.g. [ˈaθropos] vs. SMG [ˈanθropos] ‘person’ etc.). So, our assumption would be that this sequence became permitted at some point in time due to a relaxation of the relevant phonotactic constraint[7] as a result of SMG influence. As such, it will not be treated as a true Arcadian sequence.

Lastly, we can observe that, within the fricatives, the sibilant /s/ is the most active of all in cluster participation, as it combines with stops and other fricatives as a cluster-first member as well as with /p/ and /s/ as a cluster-second member (the latter being the only case where a stop is combined with another obstruent). We would therefore agree that the phonotactic behaviour of /s/ is idiosyncratic.

Furthermore, the strong preference of this dialect for the onset position of its syllables to be occupied is also demonstrated by the fact that hiatus is avoided at all costs, both within the word and across word boundaries:

Within the word, hiatus is resolved with the help of:

(3)

a.	palatalization		[katoˈmirjo]	< /katomirio/ ‘million’
b.	palatalization and fusion	[ˈʎo]	< /lio/	‘untie’
c.	deletion	[ˈtro]	< /troo/	‘eat’
d.	epenthesis		[kriˈjeno]	< /krieno/ ‘be cold’

Even though we will not be looking outside the word in this study, across word-boundaries, hiatus is resolved with the help of:

(4)

a.	semivocalization/fortition	[jareˈti]	< /i areti/	‘the.Areti (name)’
b.	coalescence		[ˈpokama]	< /pu ekama/ ‘that I did’
c.	deletion	[ˈtaloɣo]	< /to aloɣo/	‘the horse’

[7] In the medieval dictionary of E. Kriaras (1969-2016), this cluster seems to be systematically simplified via deletion.

The onset position of the Arcadian Greek syllable, as the SMG one, can be occupied by maximum three consonants, the make-up of which is always s+voiceless stop+sonorant. As these are found both word-initially and word-medially, they can be syllabified in the same syllable:

(5)

Word-initially:

a.	[ˈstreɣo]	‘come true’
b.	[skroˈpao]	‘spill’
c.	[ˈsproxno]	‘push’
d.	[ˈskɲipa]	‘small mosquito’ < /sknipa/

Word-medially:

a.	[ˈastrama]	‘thunder’
b.	[aˈsprizo]	‘whiten’ etc.

At this point, it is important to clarify that the onset clusters that we have described and analyzed so far are all *underlying* clusters (in terms of notation, they are all in bold type). However, there are clusters occupying the onset position of the Arcadian syllable, which do not conform to the description we have just put forth. However, these clusters are *not* underlying but rather reflexes of underlying /Ci/ sequences, where the high vowel /i/ loses its syllabicity and gets strengthened into a fricative before another vowel so that hiatus is avoided (*pç, tç, bj, dj, fç, vj, θç ðj, mj, rj* - see last column of grid). Both in Arcadian Greek and SMG (see Baltazani et al., 2016, among others), strengthening or fortition of /i/ typically applies in morpheme boundaries. As a result, these clusters are products or outcomes of morphological operations:

(6)

	Underlying form	**Surface form**	**Gloss**	***Surface form***
	Plural forms of neuter nouns ending in -i			*Singular forms*
a.	/pap-ia/	[pa**pç**á]	‘ducks’	[pap-í]
b.	/ɣat-ia/	[ɣa**tç**á]	‘kittens’	[ɣat-í]
c.	/kub-ia /	[ku**bj**á]	‘buttons’	[kub-í]
d.	/ɣiurd-ia /	[jur**dj**á]	‘traditional women’s coats’	[jurd-í]

e.	/karf-ia/	[kar**fç**á]	‘nails’	[karf-í]
f.	/kluv-ia/	[klu**vj**á]	‘cages’	[kluv-í]
g.	/spaθ-ia/	[spa**θç**á]	‘swords’	[spaθ-í]
h.	/peð-ia/	[pe**ðj**á]	‘children’	[peð-í]
i.	/psom-ia/	[pso**mj**á]	‘breads’	[psom-í]
j.	/mer-ia/	[mer**j**á]	‘thighs’	[mer-í]

These clusters deviate from our analysis in that they are the only ones in which both obstruents and sonorants can combine with other obstruents to their right. Essentially, any voiceless consonant can combine with *ç* and any voiced one with *j* except for those with which /i/ is fused without leaving its trace behind (i.e. k, g, s, z, x, ɣ, ts, dz – see grid). Much less frequently, these clusters appear intermorphemically (e.g. [ˈvjazume] ‘make haste’, [ˈpçato] ‘plate’, etc.), in words that historically included an underlying /i/ in that place or in loanwords. Additionally, the reflexes of /i/, i.e. the strengthened *ç* or *j*, also contribute to the formation of onsets consisting of three obstruents in a row: *ftç, xtç, ɣðj* and *vðj*. This happens when the underlying string /CCiV/ yields [CCCV]: e.g. /aftia/ > [aˈftça] ‘ears’, /ðixtia/ > [ˈðixtça] ‘nets’, etc. These triconsonantal sequences would be problematic as far as their syllabification is concerned. The sequence *ftç*, for example, could be classified as a tautosyllabic cluster as it is attested word-initially in a handful of examples (e.g. /ftiano/ > [ˈftçano] ‘make, do’). However, as none of the remaining ones (i.e. *xtç ɣðj* or *vðj*) is attested word-initially, they would have to be syllabified heterosyllabically. Nevertheless, by spanning the syllable boundary, potentially problematic word-medial codas are yielded (see the analysis of codas in section 3.2 below). Less often, the underlying string /CCiV/ may also yield obstruent+sonorant+obstruent clusters, such as *prj, brj, ɣrj,* etc. (e.g. /priali/ > [ˈprjaʎi] ‘day after tomorrow’, etc.), which are also attested at the beginning of the word.

Let us now attempt a sonority account of the Arcadian onset. The notion of sonority in conjunction with the internal organization of the syllable is widely discussed in the literature (among others, Hooper, 1976; Steriade, 1982; Selkirk, 1984; Zec, 1995; Clements, 1990; Parker, 2002, 2011). The Sonority Sequencing Principle or Sonority Sequencing Generalization (SSP or SSG) organizes the syllable phonotactically in terms of sonority. According to it, the peak of sonority is always the nucleus of the syllable, with the consonants to its left rising in sonority and those to its right falling in sonority. Sounds are hierarchically ranked according to how sonorous they are by means of a sonority scale which dictates their exact sonority values.

If we revisit the grid in (2) above and now study the consonantal sequences, it features from the point of view of sonority, we will notice immediately that the mirroring sequences that we also noted in our onset maximization account of the Arcadian syllable, i.e. obstruent+liquid vs. liquid+obstruent, are of rising and of falling sonority respectively. The obstruent+liquid sequences are therefore optimal candidates for the onset position of the Arcadian Greek syllable as they typify a maximum sonority distance. Numerically, the former is also the largest type of clusters to be occupying the onset position in this dialect. Their mirror images, on the other hand, namely liquid+obstruent sequences, would naturally have to be syllabified heterosyllabically as they fall in sonority. This, so far, makes quite an elegant account.

Nevertheless, turning to obstruent+obstruent sequences, namely fricative+fricative and fricative+stop ones, we notice that their sonority profile violate SSP as the former are of identical sonority while the latter are of falling sonority. However, it is well known in the literature that several languages permit sonority plateaus or violate SSP (see Vaux & Wolfe, 2009 for a discussion of SSP violations). As a result, sonority is not central to their syllabification. What's more, according to Morelli (2003), fricative+stop clusters constitute the unmarked type among obstruent clusters cross-linguistically. Among fricative clusters, /s/+stop ones - where sibilants are more sonorous than stops - are particularly interesting since several languages (including English), which normally obey the SSP, violate it with sC clusters[8]. In fact, these are the most frequent violations to SSP cross-linguistically (see Parker, 2011). Wright (2004) argues that sC clusters pattern idiosyncratically in languages due to the fact that the perceptual acoustic cues of the sibilant [s] stand out and an intervening low sonority consonant in the onset of the same syllable does not therefore affect its perceptibility. What is also interesting is that sibilants do not combine with sonorants despite the fact that they would exhibit an ideal sonority profile.

All non-underlying sequences (*pç, tç, bj, etc.* - see last column of grid) would make problematic onsets as they are sonority reversals and do not, therefore, meet sonority requirements.

[8] See also Goad (2011) for a coda+onset analysis of sC clusters following from cross-linguistic observations stemming from the syllable contact approach.

3.2 Syllabification operations: Codas

Let us now turn our attention to the coda position of the Arcadian Greek syllable and see how this may differ from that of SMG. After prioritizing onsets and analyzing their internal make-up, we now need to examine the nature of the codas which we left behind as we were syllabifying sequences that spanned the syllable boundary.

SMG permits only coronals in its coda, namely /s/ and /n/ word-finally and the liquids /r/ and /l/ (as well as the rarer nasal /n/) word-medially. Even rarer would be the case of word-medial three- or four-consonant sequences, some of which are also found in morpheme boundaries, where an obstruent may have to be syllabified in the coda (see Malikouti-Drachman, 2001).

A first general statement that we could make is that underlying Arcadian Greek codas are non-branching. Word-finally, the only underlying segment permitted as coda is /s/, which is also a morphological marker as it marks gender, case and/or number: e.g. [ˈtopos] 'place, location', [ˈelatus] 'fir tree. ACC.PL.', etc. The coronal nasal /n/, which is a permissible coda in SMG along with /s/, is disallowed from this position in Arcadian Greek via an obligatory epenthesis of /e/ word-finally: e.g. /exun/ → [ˈexune] 'they have'; /peðion/ → [peˈðjone] 'children.GEN.PL.'; /ɣlikon/ → [ɣʎiˈkone] 'sweet.MASC.ACC.SG.', etc.

Following onset maximization requirements, sonorant+obstruent word-medial sequences would have to be syllabified heterosyllabically as they are not attested word-initially. On the other hand, as we have already seen, obstruent+obstruent word-medial sequences are syllabified tautosyllabically as they are attested word-initially. Thus, we can only encounter a sonorant segment as a legal coda, either a rhotic or a lateral one but not a nasal. (Please note that /s/, in accordance with onset maximization requirements, would be syllabified in the onset and not in the coda position word-medially). While /r/ is the most frequent coda in the word-medial coda position (ex. 7a-c), laterals follow suit but are somewhat peculiar: they never surface as coronals but as palatals. This is because the majority of them are yielded by fusion with a following unstressed [i] (see 8 below), and, consequently, they get resyllabified from the onset to the coda position of the preceding syllable (e.g. [ˈfiʎ.pas] < /filipas/ 'Philip'). What is interesting, though, is that even laterals in lateral+C sequences in loanwords (not very many in number) get palatalized (ex. 7d-f):

(7)

a.	[tar.naˈrizo]	‘sway’
b.	[kar.teˈro]	‘await/tend’
c.	[ˈemor.fos]	‘beautiful.MASC.’
d.	[ˈzveʎ.tos]	‘svelte.MASC.’
e.	[ˈvoʎ.ta]	‘stroll’
f.	[paʎ.ˈto]	‘coat’

(Examples 7d-f feature a coronal [l] in SMG)

If we look at the grid again, we will notice that, in effect, nasals do not combine with other consonants to form clusters (with the exception of ɣn, zm and the rarer xn, mn, discussed in 3.1 above). Note that the row corresponding to *n* (and *m*) is empty in the grid. This means that /n/ is not found next to another consonant word-initially or medially[9]. Essentially, nasals do not precede other consonants and are very rare or even contestable as cluster-second members in Arcadian Greek. It would be worth-mentioning that, in SMG, there is a series of word-medial obstruent+nasal sequences (tn, fn, vn, ðn, km, vm, θm, ðm, xm, ɣm) which are problematic as far as their syllabification is concerned as they would be classified as heterosyllabic if we were to follow the onset maximization principle, since they are not attested word-initially, whereas they would have to be classified as tautosyllabic if we were to follow sonority requirements. These clusters do not posit a problem for Arcadian Greek as they are simplified by epenthesis, metathesis or deletion: e.g. [vaθiˈmos] vs. SMG [vaθmós] ‘degree, grade’, [arˈθimos] vs. SMG [ariθmós] ‘number’, [ˈalama] vs. SMG [ˈalaɣma] ‘change’ etc.

In symmetry with the word-final coda position, nasals are disallowed from the word-medial coda position as well. Word-initial or word-medial obstruent+nasal sequences, such as *pn*, *kn*, *θn*, *tm* and often *xn*, are also simplified: e.g. [kapiˈnos] vs. SMG [kaˈpnos] ‘smoke’, [piˈɲiɣo] vs. SMG [ˈpniɣo] ‘drown’, [ˈrino] vs. SMG [ˈrixno] ‘drop’, [tsikuˈɲiða] vs. SMG [tsuˈkniða] ‘nettle’. It is obvious that all three processes, namely deletion, epenthesis and metathesis, conspire to ban an obstruent coda from a word-

[9] Historically, the nasal environment in the consonantal sequences nt, mp, nk gave rise to the voiced stops of Modern Greek. Hence the reason why these sequences are not attested. Moreover, nasal+fricative sequences, such as *mf, nð, mv, nθ,* etc., are also simplified in Arcadian Greek.

medial syllable. Two exceptions to the word-medial sonorant coda are (a) the sequence *fn* (see section 3.1 above) and (b) the triconsonantal *xtr* sequence which we would also have to syllabify in separate syllables, i.e. as x.tr, since it is not attested at the beginning of words: e.g. [ˈsfalax.tro] 'type of thorn', [ˈplex.tra] 'knitter/bunch', etc. Nevertheless, it is not frequent.

As can be seen, while, on the one hand, our syllabic onsets perfectly mirror our word onsets, there is no symmetry between word-final and word-medial codas. After determining where the syllable boundary is in word-medial sequences, we notice that we must assign a plethora of sonorants in the coda position. These, at the same time, do not occur word-finally. Moreover, the word-medial coda position can be occupied by a far larger number of consonants than the word-final one. As a result, we have not succeeded in reducing well-formed words to sequences of well-formed syllables as our syllables are not perfect mirrors of our words (see also Goldsmith, 2011).

From the perspective of sonority, word-medial sequences of descending sonority, such as sonorant+obstruent, would have to be syllabified heterosyllabically. The same goes for word-medial obstruent+obstruent sequences which form either sonority plateaus or descend in sonority and, thus, violate SSP. These clusters are unproblematic from an onset maximization perspective as they would make perfect onsets in the middle of the word given that they also appear at the beginning of the word. However, they are indeed problematic for a sonority account. They cannot be assigned in the onset because they are sonority reversals or plateaus, nor can they be split up between the coda and the onset of two adjacent syllables since the dialect, as we have already demonstrated, clearly disfavours obstruents as codas. As we saw earlier, it summons different processes, such as epenthesis, metathesis and deletion, to simplify obstruent-first clusters so as to avoid an obstruent in its word-medial coda. As we have also seen, it even disfavours nasals word-finally by epenthesizing /e/. It would therefore be counterintuitive, on the one hand, to assign obstruents in the coda and, on the other hand, to ban them from it. Moreover, allowing a series of obstruent codas in word-medial position is also problematic because they are restricted to the middle of the word.

The codas that we have described and analysed so far are *underlying*. As with onsets, there are also *non-underlying* codas in the Arcadian Greek syllable which do not conform to the analysis we have just proposed. These are (a) word-final or word-medial obstruent codas and (b) branching word-final codas.

The non-underlying consonants that can be found in the word-final position are (a) palatalized dorsals and coronals which are yielded by palatalizaton and fusion with a following unstressed [i] and which in turn gets resyllabified to the coda on the left (see 8a) and (b) consonants which have undergone secondary palatalization or labialization (see 8b). The former are found word-medially as well[10] (see 8a).

(8a)

a.	[ˈfiʎ.pas] [ˈfi.li.pos])	<	/filipas/	‘Philip.NOM.’(cf. SMG
b.	[kuˈveʎ]	<	/kuveli/	‘beehive’
c.	[aftocíɲ.to]	<	/aftokinito/	‘car’
d.	[eleɲ]	<	/eleni/	‘Helen’
e.	[ˈvriʃ]	<	/vrisi/	‘water tap’
f.	[ˈxaʃmo]	<	/xasimo/	‘loss’
g.	[kafuˈriʒ]	<	/kafurizi/	‘it started snowing’
h.	[ˈviʒras]	<	/viziras/	‘traditional children’s game’
i.	[luˈmaɟ]	<	/lumaki/	‘young tree’
j.	[sfaˈraʤ]	<	/sfaragi/	‘asparagus’
k.	[ˈmac]	<	/maki/	‘Mike’
l.	[ˈmaç]	<	/maxi/	‘battle’
m.	[ˈruçnos]	<	/ruxinos/	‘of fabric’
n.	[ˈmaj]	<	/mayi/	‘magicians’
o.	[aˈnajra]	<	/anayira/	‘around’

(8b: the following are proper names - the masculine ones are in non-nominative cases)

a.	[ˈpopʲ]	<	/popi/	‘Popi’, short for ‘Penelope or Calliope’
b.	[ˈbabʲ]	<	/babi/	‘Babis’
c.	[ˈfotʲ]	<	/foti/	‘Photius’
d.	[vaˈladʲ]	<	/valadi/	‘Valantis’
e.	[ˈʃifʲ]	<	/sifi/	‘Sifis’
f.	[ˈstaθʲ]	<	/staθi/	‘Stathis’

[10] In /C_1iC_2/ context (where C_1 is a labial or coronal (p, f, v / t, θ, ð) which cannot palatalize and fuse with [i]) [i] could get devoiced, reduced or even elided. The location of stress and the voicing of the surrounding sounds could also play a role. We will not concern ourselves with such (potential) derived consonantal sequences as only an acoustic analysis would shed light on the status of the intervocalic vowel with accuracy. This would be a future step.

g.	[ˈfivʲ]	<	/fivi/	‘Phoebe’
h.	[ðjoˈmiðʲ]	<	/ðiomiði/	‘Diomedes’
i.	[soˈtirʲ]	<	/sotiri/	‘Sotiris’
j.	[ˈmimʲ]	<	/mimi/	‘Mimis’, short for ‘Demetrius’

Furthermore, a smaller number of branching word-final codas, which do not fit into the dialect’s possible core syllables, are generated when unstressed [i] and [u] get deleted word-finally triggering secondary palatalization or labialization (respectively) in the preceding consonant. All branching codas are outputs of underlying /(C)CCi/ or /(C)CCu/ strings and rise in sonority: e.g. [ðiˈmitrʲ] < / ðimitri/ ‘Dimitri.GEN./ACC.’, [ˈkastrʷ] < /kastru/ ‘castle.GEN.’. It is worth noticing that, at the same time, secondary palatalization or labialization causes resyllabification which can in turn yield codas of falling sonority word-finally: e.g. [mar̲tʲ] < /marti/ ‘March.GEN./ACC.’ (cf. SMG [ˈmar.ti]) – the mirror image of the coda of [ðiˈmitr̲ʲ]. Evidence in support of extrasyllabicity of these clusters (see Vaux & Wolfe, 2009) is not compelling as they do not count towards weight and are not required by the morphology.

The above examples, generated after the high vowels lose their syllabicity in unstressed positions and undergo different phonological processes, invariably feature more complex codas, either a wider variety of them, as essentially any consonant can now occupy this position, or even branching codas. However, cross-linguistically, we would normally expect a smaller variety of consonants to be permitted in the coda than in the onset of the syllable of a language. At the same time, we normally expect fewer consonant clusters to be permitted in the coda than in the onset of a language. This is indeed true for the *underlying* onsets and codas of Arcadian Greek. The non-underlying onsets and codas, however, are clearly aberrant and diverge from the norm or a typical pattern of behaviour.

Moreover, the goal of phonological processes such as epenthesis or deletion is typically to turn marked syllables into unmarked ones. In our case, all the non-underlying onsets and codas are infinitely more marked, since a greater variety of consonants is now permitted as codas and their complexity is explicitly increased. Consequently, it is necessary that non-underlying onsets and codas are treated separately from underlying ones in Arcadian Greek. In this study, they are treated as mere surface outputs or reflexes of underlying structure at a phonetic level.

4. Conclusions

This study attempted an analysis of the patterns that lie at the base of the syllable of Arcadian Greek. In doing so, we inevitably compared the Arcadian Greek syllable with the SMG one; a comparison which showed that the two differ considerably, with the former exhibiting less complexity in its underlying syllabic margins than the latter, while, on the other hand, exhibiting significant more complexity in its non-underlying ones. What is interesting is that we had to treat underlying syllabic margins separately from non-underlying ones, with the latter being handled as surface outputs or reflexes of underlying structure at a phonetic level. It would not otherwise be possible to reconcile having, on the one hand, a language which prefers to have an onset rather than be onsetless or summons different phonological processes to ban obstruents from its codas word-medially (via cluster simplification), or even sonorants word-finally (via epenthesis) and, on the other hand, one that happily allows obstruent or branching codas.

The Arcadian Greek syllable was analyzed using (a) Kahn's (1976) syllabification model based on the onset maximization principle and (b) a sonority model. The onset maximization model accounted for the syllabification of Arcadian Greek more insightfully as it elegantly avoided assigning obstruent codas (with one or two rare exceptions) to the Arcadian Greek syllable, even though it did not fully succeed in creating syllables perfectly mirroring words. Sonority, on the other hand, was less successful in that. This should be hardly surprising given that Arcadian Greek is a dialect which allows a series of obstruent+obstruent clusters (namely sonority reversals and sonority plateaus) both word-initially and word-medially and has unequivocally shown that it disfavours obstruents in its coda position.

References

Arvaniti, A. 1999. Illustrations of the IPA: Standard Modern Greek. *Journal of the International Phonetic Association* 29(02), 167–172.

Babiniotis, G. 1979. A Linguistic Approach to the Language Question in Greece. *Byzantine and Modern Greek Studies*, 1–16. Oxford: Blackwell.

Babiniotis, G. 1998. *Συνοπτική Ιστορία της Ελληνικής Γλώσσας [A Concise History of the Greek Language].* Athens: Ellinika Grammata.

Baltazani, M. and Topintzi, N. 2012. On some Phonetic and Phonological Properties of the Greek Glide. In *Selected Papers of the 10th International Conference of Greek Linguistics,* edited by Z. Gavriilidou,

A. Efthymiou, E. Thomadaki, and P. Kambakis-Vougiouklis, 153–166. Komotini: Democritus University of Thrace.
Baltazani, M, Kainada, E., Revithiadou, A., and Topintzi, N. 2016. Vocoid-driven Processes: Palatalization and Glide Hardening in Greek and its Dialects. *Glossa: A Journal of General Linguistics* 1(1) 23, 1–28.
Browning, R. 1995. *Medieval and Modern Greek*. Second Edition. Cambridge: CUP.
Clements, G. N. 1990. The role of the sonority cycle in core syllabification. In *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*, edited by J. Kingston and M. E. Beckman, 283–333. Cambridge: CUP.
Coutsougera, P. 2019. Palatalization, fortition and deletion in the high vowels of Arcadian Greek. *Toronto Working Papers in Linguistics* 41 (1). https://doi.org/10.33137/twpl.v41i1.32771
Coutsougera, P. forthcoming. *Λεξικό της Βορειοανατολικής Αρκαδίας [Dictionary of Northeastern Arcadia]*. Athens: Patakis Publications.
Drachman, G. and Malikouti-Drachman A. 1997. Dissimilation in Cypriot Greek: Competing Analyses. *Studies in Greek Linguistics* 17, 42–57.
Fourakis, M., Botinis, A. and Katsaiti, M. 1999. Acoustic Characteristics of Greek Vowels. *Phonetica* 56, 28–43.
Goad, H. 2011. The Representation of sC Clusters. In *The Blackwell Companion to Phonology*, edited by M. van Oostendorp, C. Ewen, E. Hume and K. Rice, 898-923. Oxford: Wiley-Blackwell. https://doi.org/10.1002/9781444335262.wbctp0038
Goldsmith, J. A. 2011. Syllables. In *The Handbook of Phonological Theory*, edited by J.A. Goldsmith, J. Riggle and A.C.L. Yu, 164-196. Oxford: Wiley-Blackwell.
Gratsouni, D. and Topintzi. N. 2020. On The Syllabification of ON (Obstruent-Nasal) Clusters in Greek. *Studies in Greek Linguistics* 40, 105-113.
Hooper, J. B. 1976. *An Introduction to Natural Generative Phonology*. New York: Academic Press.
Horrocks, G. 2010. *Greek: A History of the Language and its Speakers.* Second Edition. Chichester.
Joseph, B. and Philippaki-Warburton, I. 1987. *Modern Greek.* London: Croom Helm.
Kahn, D. 1976. *Syllable-Based Generalizations in English Phonology*. MIT, Cambridge, Mass.
Kappa, I. 1996. Syllabification in Standard Modern Greek. *Studies in Greek Linguistics* 16, 83–91.
Kappa, I. 1997a. A First Approach to the Optimal Codas in Modern Greek.

Proceedings of the Second International Conference on Greek Linguistics, edited by G. Drachman, A. Malikouti-Drachman, J. Fykias and S. Klidi, 25–33. Graz: W. Neugebauer.

Kappa, I. 1997b. Συλλαβική Δομή της Ν.Ε. στη Θεωρία του Βέλτιστου [Syllabic Structure of Modern Greek in OT]. *Studies in Greek Linguistics* 17, 72–84.

Karzi, V. 2000. *Θεωρίες για την Συλλαβή: Μια Πρόταση για τα Εναλλασόμενα Συμφωνικά Συμπλέγματα της Νέας Ελληνικής. [Theories on the Syllable: A Proposal for the Alternating Consonant Clusters of Modern Greek]*. MA Thesis. University of Athens.

Kontossopoulos, N. 2001/1994. *Διάλεκτοι και Ιδιώματα της Νέας Ελληνικής [Dialects and Varieties of Modern Greek]*. Athens: Gregoris Publications.

Kriaras, E. 1969–2016. *Λεξικό της Μεσαιωνικής Ελληνικής Δημώδους Γραμματείας 1100–1669 [Lexicon of Medieval Greek Demotic Literature 1100–1669]*, 20 vols. Thessaloniki.

Mackridge, P. 1985. *The Modern Greek Language.* Oxford: Oxford University Press.

Malikouti-Drachman, A. 1984. Syllables in Modern Greek. *Phonologica* 181–187. London: CUP.

Malikouti-Drachman, A. 2001. Greek Phonology: A Contemporary Perspective. *Journal of Greek Linguistics*, 187–243.

Manolessou, I. and Pantelidis, N. 2013. Velar Fronting in Modern Greek Dialects. In *Proceedings of the 5th International Conference of Modern Greek Dialects and Linguistic Theory*, edited by M. Janse, B. Joseph, A. Ralli, and M. Bagriacik, 272–286.

Morelli, F. 2003. The Relative Harmony of /s+stop/ Onsets: Obstruent Clusters and the Sonority Sequencing Principle. In *The Syllable in Optimality Theory*, edited by C. Féry and R. van de Vijver, 356–371. CUP.

Newton, B. 1972. *The Generative Interpretation of Dialect: A Study of Modern Greek Phonology*. Cambridge: CUP.

Pantelidis, N. 2001. Πελοποννησιακός Ιδιωματικός Λόγος και Κοινή Νεοελληνική [Peloponnesian Dialectal Speech and Standard Greek]. *Studies in Greek Linguistics* 21: 550–561.

Pantelidis, N. 2007. Κοινή Δημοτική: Παρατηρήσεις στη Διαδικασία Διαμόρφωσής της [Common Demotic: Observations in the Process of its Development]. *Studies in Greek Linguistics* 27: 337–347.

Pantelidis, N. 2015. Σχετικά με το Ιδιωματικό Υπόβαθρο της Κοινής Ελληνικής [On the Dialectal Background of Modem Greek]. In *Οι Γεωγραφικές και Κοινωνικές Ποικιλίες της Νέας Ελληνικής και η*

Παρουσία τους στην Εκπαίδευση [The Geographical and Social Varieties of Modern Greek and their Presence in Education], edited by. J. Androulakis, 41–50. Tyrnavos.

Parker, S. 2002. *Quantifying the Sonority Hierarchy*. Ph.D. Dissertation. University of Massachusetts, Amherst.

Parker, S. 2011. Sonority. In *The Blackwell Companion to Phonology*, edited by M. van Oostendorp, C. J. Ewen, E. Hume and K. Rice, 1160–1184. Oxford, UK: Wiley-Blackwell.

Selkirk, E. 1984. On the Major Class Features and Syllable Theory. In *Language Sound Structure: Studies in Phonology Dedicated to Morris Halle by his Teachers and Students,* edited by M. Aronoff and R. T. Oerhle,107–113. Cambridge, MA: MIT Press.

Setatos, M. 1974. *Φωνολογία της Κοινής Νεοελληνικής [Phonology of Standard Modern Greek]*. Athens: Papazisis.

Setatos, M. 1987. Απ' τα Αρχαία στα Νέα Ελληνικά: Φωνολογικές Αλλαγές [From Ancient Greek to Modern Greek: Phonological Changes]. *Studies in Greek Linguistics* 8, 187–193.

Sfakianaki, A. 2002. The Acoustic Characteristics of Greek Vowels Produced by Adults and Children. In *Selected Papers on Theoretical and Applied Linguistics from the 14th International Symposium*, edited by M. Makri-Tsilipakou, 383–394. Thessaloniki: Aristotle University.

Steriade, D. 1982. *Greek Prosodies and the Nature of Syllabification*. Ph.D. dissertation, MIT. Published 1990, New York: Garland.

Tonnet, H. 1995. *Ιστορία της Νέας Ελληνικής Γλώσσας [The History of the Modern Greek Language].* Athens.

Topintzi, N. and Baltazani, M. 2016. The Third Dream: Aspects of the Morphophonology of the Greek Glide and Palatals. *Linguistics Unlimited* 1.

Trudgill, P. 2003. Modern Greek Dialects: A Preliminary Classification. *Journal of Greek Linguistics* 4, 45–64.

Vaux, B. and Wolfe, A. 2009. The Appendix. In *Contemporary Views on Architecture and Representations in Phonology*, edited by E. Raimy and C. Cairns, 101–143. Cambridge: MIT Press.

Wright, R. 2004. A Review of Perceptual Cues and Cue Robustness. In *Phonetically Based Phonology*, edited by B. Hayes, D. Steriade and R. Kirchner, 34-57. Cambridge: CUP.

Zec, D. 1995. Sonority Constraints on Syllable Structure. *Phonology* 12, 85–129.

CHAPTER FIVE

SPEECH ERRORS GREEKS MAKE IN ENGLISH: A TUTORIAL

ELENA BABATSOULI

Abstract

Extensive literature on second language speech perception and production increasingly documents second language phonology across a variety of first and second (foreign) language combinations. Little is known on the speech errors adult native speakers of standard modern Greek make when they speak English. The current contribution utilizes original data collected over the past ten years to identify the contexts of such interlanguage errors. The linguistic evidence documents mismatches at the segmental (vowel and consonant) and prosodic (syllable and word) levels that are primarily instigated by transfer of L1 structures and the effect of transparent L1 orthographic readings, though universal, prototypical, and developmental processes are also noted. The chapter is advanced as a tutorial that uses actual example archetype-words from the dataset to typify patterns of Greek speakers' accent in English and aims to narrow the gap between theory, research findings, and actual learners – native Greek speakers interested in improving their pronunciation in English as a second/foreign language. The practical significance of this is to provide a useful yardstick for use in L2 instruction settings, either in classrooms or for independent use, and in clinical contexts, where aptitude in linguistic diversity is interwoven with learning or therapy success.

Keywords: speech errors, interlanguage, L1 Greek, L2 English, pronunciation

1. Introduction

Can you say, 'the sixth sick sheik's sixth sheep's sick'? If you are Greek, this may be challenging, sounding like 'the six sik six six sip's sik'. English is the most widely spoken language in the world; in fact, in every four English speakers, three are non-native (Babatsouli, 2019a). As the current volume aptly demonstrates, abundant literature on second language (L2) speech investigates the causes of non-native pronunciation in terms of perception and production, seeking to provide effective guidelines that inform L2 instruction. While intelligibility is vital, nativelikeness is less crucial especially in non-prescriptivist circles that include renowned dialectologists (Trudgill, 2008), linguists (Jusczyk, 2000) and sociolinguists (Fishman, 1976), since "accent serves as a powerful symbol of ethnicity" (Leather & James, 1991, 310). The practical significance of this is in instructional and clinical contexts, where aptitude in linguistic diversity is interwoven with learning or therapy success (e.g. Babatsouli, 2021). The present contribution builds on limited and intermittent literature informing the field of second language acquisition in the past 40 years on production patterns that characterize native Greek speakers' accent in English.

The chapter is thematically divided into *themes* and *evidence*. *Themes* begins by succinctly describing fundamental tenets guiding our understanding of second language speech, and then outlines existing literature examining L2 English in speakers whose first language (L1) is Greek. Next, the phonological/phonetic systems of English and Greek are compared, highlighting similarities and differences between them. While the contrastive analysis approach to comparing English and Greek phonology is not without precedent (Babatsouli, 2019b, Koutsoudas & Koutsoudas, 1985), translating knowledge into practice, and research into policy, is admittedly slow and arduous (Archibald, 2019). *Evidence* espouses an applied stance to encapsulating accent, aiming to narrow the gap between theory, research findings and the non-academic layman i.e., Greek learners interested in improving their pronunciation in English. To this end, example-productions from actual data are used to illustrate Greek speakers' accent in English. It is noted that the study elucidates L2 English accent focusing on native speakers of the standard variety, known as Standard Modern Greek (SMG).

The learning of speech sound systems is known as phonological acquisition. Foreign accent has been called "global" (Major, 2001: 18) because errors occur in individual sounds (segments) and their combinations across prosodic contexts (phonotactics). A generic theme in phonological acquisition relates to non-linearity (Rice & Avery, 1995), in that sounds, syllables, and

words (phonological units) have underlying mental representations that are interconnected in a geometric hierarchy of levels from *top* (phrase) to *bottom* (features of sounds). In this spirit, the present tutorial, elucidates accent in a non-linear approach, based on evidence illustrated by examples drawn from original data.

2. The acquisition of second language speech

Learning L2 speech involves non-linguistic and linguistic factors, rooted in not-fully transparent phenomena in physiology, perception, cognitive processing, alongside social and affective factors. As in L1 and child bilingual acquisition (*protolanguage*), acquiring L2 in adulthood (*interlanguage*) is further subject to global interrelations between the L1 and L2 sound systems, and dynamic influences spanning acquisition time longitudinally (Babatsouli & Ingram, 2018). Paramount is the "filtering" of non-native sounds through the native "sieve" (Trubetzkoy, 1939/1958), and an inherent predisposition for psycholinguistic *transfer* (Major, 2008), meaning that previously acquired knowledge (one's native speech sounds) influences subsequently acquired skill (the production of L2 sounds). These are the foundations of foreign accent.

2.1 Non-linguistic constraints

L2 learner constraints may be maturational, individual, and societal. A critical period of human cognitive and physical maturation (Lenneberg, 1967) assumes an upper age limit (before puberty) that is detrimental to our fundamental capacity to effortlessly acquire novel phonological contrasts. Infants have an advantage over older children and adults in learning L2 pronunciation, though there is no unequivocal consensus that adults cannot master L2 sounds (Millard & Yavaş, 2018). Steered by perception, the production of L2 speech in adults is impacted by already formed articulatory habits, since native language sounds are largely mastered by age 4;0 cross-linguistically (Babatsouli, 2019c).

Alongside age and maturational limitations, personal and social factors facilitate or inhibit native-like attainment. Such dynamics include *oral and auditory aptitude* that is subject to sensorineural sensitivity and developmental variation, *motivation*, and *personality traits*, like self-esteem and ego, *sex/gender*, with findings showing that females outperform males, and *social acceptance* and *distance*, which tampers on how well non-native speech is accepted by native speakers, and how this empowers the non-

native speaker to feel at ease among them (Leather and James, 1991). Other non-linguistic constraints involve age of speech onset (AoO) in the L2, length of L2 learning and its reported use (LoL/U), length of residence (LoR) in countries where L2 is spoken (Babatsouli and Ingram, 2018), and proficiency maintenance (Blake et al., 2021). Differences in exposure between naturalistic versus instructional settings further regulate pronunciation outcomes, as does the influence of orthography (Georgiou, 2021).

2.2 L2 perception and production

Though L2 speech *can* be unaccented, it more often *is* than *isn't*. Speech perception (phonemic awareness) critically guides production (articulation), and both drive speech learning. Languages share universal properties, but each is distinctly different regarding grammars, sound inventories, sound combinations, and sound distributions. At age 1;0, infants' perception of non-native sounds becomes L1-bound; all non-native sounds are mapped within the established L1 perceptual space based on a magnet-like theory: similar L2 phonological categories are attracted by L1 perceptual prototypes; different, thus more challenging, L2 sounds require cognitive reorganization of the speaker's perceptual space and creation of new archetypal categories (Kuhl, 1991, Werker & Tees, 1984). Perception is optimal when L2 phonological categories match on a one-to-one basis into distinct L1 categories, while when there is a mismatch between L2 and L1 categories on a many-to-one or one-to-many fashion (Best & Tyler, 2007), "accented perception" is more likely (Strange, 1995: 2).

Nevertheless, while ultimate attainment (when L2 ceases to improve) is subject to perceptual sensitivity, motoric skill may not follow suit without articulatory adjustment and practice. Reversely, the ability to articulate greatly dissimilar L2 sounds, thus creating novel phonetic categories, also contributes to the creation of corresponding perceptual categories in the learning process. When L2 sounds are similar (not identical) to L1 sounds, they are cognitively classified with respective L1 equivalents, assimilating both categories into one; this supports two-way interference between the languages in both perception and production (Flege, 1995). Erroneous input on learners' targeted perceptual and articulatory "templates" (Leather & James, 1991: 314) is common in L2 instructional settings. On practical grounds, L2 pronunciation training can be enhanced by sound discrimination training. Thus, it is advisable that learners are exposed to appropriate sound contrasts early on rather than on working to reverse the effects of erroneous input on their perceptual and articulatory templates.

2.3 Developmental aspects of L2

L2 accented speech resembles in part child developmental speech. Its development longitudinally exhibits similar phenomena with those in L1, like universals (common patterns), markedness (certain targeted constructs are less frequent to acquire and more difficult) and prototypical processes that steer errors. Just like learning to speak in monolingual contexts takes several years (minimum of 4), mastery of L2 phonology accordingly spans a long time, subject to individual variation that is contingent on the combination of factors at play outlined above. Nevertheless, the developmental path in L2 is slower, it typically takes longer, ultimate attainment is not guaranteed, and universal processes, prevalent early on, subsequently play a smaller role than transfer, which shrinks progressively as the L2 gradually advances towards nativelikeness; interestingly, segmental transfer recedes earlier than suprasegmental (stress) transfer (Major, 2001). An illustration of the dynamic nature of development is that L2 errors may show variability across utterances produced on separate occasions and times (De Bot et al., 2007). The different ways in which LI may influence L2 learning also relies on interactions with the phonetic environment i.e., combinations of sounds in syllables/words/phrases and sound distributions in the language, and on how difficult (marked) L2 sounds are based on systematic comparisons of the L1 grammar, the L2 target grammar, and markedness relations outlined in universal grammar of language as a human capacity (Eckman, 1977).

3. The acquisition of L2 English by Greek adult learners

How monolingual Greek learners speak English as a second, or foreign, language is increasingly being investigated in the past few decades in terms of both perception and production. This section summarizes related literature incorporating studies of L2-English speech in learners with standard Modern Greek (MG) and Cypriot Greek (CG) as L1. Koutsoudas and Koutsoudas (1985), an early work tackling the topic, assumes a theoretical approach and outlines phoneme similarities and differences in the two languages within the contrastive analysis framework, providing insights on likely challenges in teaching English pronunciation to native Greek speakers.

Perception literature is exemplified early on by Lengeris and Hazan (2010) who investigate how phonetic training in learning L2 vowels relates to L1 vowel processing ability and frequency discrimination acuity. Findings showed that high variability training had better results than low-variability training, but that listeners' category representation did not change. Also,

participants with better frequency discrimination acuity for synthetic vowels were better at identifying natural L2 vowels both before and after training. Giannakopoulou et al. (2013) involves Greek adult (20-30 years) and child (7-8 years) speakers of English, studying training effects on L1/L2 cue weighting to improve vowel identification and discrimination. Significant difference is documented compared to native standards, and training results in more pronounced improvement for children than adults, due to brain placticity. Lengeris and Nicolaidis (2016) studied L2 identification and production of English consonants by Greek speakers, reporting higher identification accuracy scores in quiet than noise settings. There was performance variability in the perception and production of consonant accuracy across modalities.

Kkese (2016) is a comprehensive study on L2 perception examining how plosive voicing contrasts in L2 English are acquired by CG college students, that focuses on different factors influencing plosive identification and production. Perception is found better for voiceless plosives which, in turn, positively affected production; voiced plosives in the study were perceived as prenasalized. Kkese and Petinou (2017) elaborate this further focusing on gaps between phonetic cues and phonological contrasts and assigning poor perceptual performance of voiced plosives on their distinct phonetic-phonological differences in the two languages. Kkese and Karpava (2021) explored discrimination difficulties CG listeners had with L2 English vowels and consonants, looking at both linguistic and non-linguistic factors. Listeners' discrimination was better for consonants than vowels and non-linguistic factors (age, gender, years of L2 instruction, and experience) had variable effects on speech perception. Acoustic cues, like voicing and vowel length, were not found to be as relevant for CG listeners as previously attested for native-English listeners.

In terms of production, Babatsouli and Kappa (2010, 2011) present an early experimental study documenting adults' transfer of Modern Greek sounds into English and discuss implications for teaching. While phonemic transfer into an L2 is well supported cross-linguistically (Krashen, 1987), fewer studies report L1 allophonic transfer. The study reports Greek palatals substituting English velar and glottal fricative consonants. Transfer was shown consistent, irrespective of English proficiency level, suggesting complexity in shedding allophonic rules in L2 production. Following phonetic and phonological awareness training, re-testing confirmed benefits for L2 pronunciation with positive implications for L2 instruction. Babatsouli (2019) reports the following outputs as L2 English input in a bilingual child's Greek-English development: /k/→[c], /g/→[ɟ], /ɹ/→[ɾ],

/j/→[j], /h/→[x. ç].

Two more studies have investigated L2 English sound production by Greeks, both involving the English rhotic, /ɹ/, produced as the Greek tap, /ɾ/. Harper et al. (2016) focused on the pharyngeal gesture in English /ɹ/ using real-time MRI recordings from one L1 French/L2 English and one L1 Greek/L2 English speaker. Neither participant produced native-like laryngeal constrictions, while both produced the targeted rhotic as for their L1 targets -- in the case of the L1 Greek/L2 English speaker with no pharyngeal constriction at all. Dimitriou (2018) explored English rhotic production in word-initial and intervocalic positions, and in Cr and rC clusters. The informants were two CG-speaking groups who attended English- or Greek-speaking schools. Though no group achieved nativelikeness in terms of neither duration nor formant frequencies, the study showed better rhotic performance in learners from English-speaking schools.

Much scarcer is literature on Greek orthography effects on L2 English pronunciation. Coutsougera (2007) reviews relevant literature and compares Greek and English orthographic systems. Georgiou (2021) investigated L1 Greek orthography effects on the perception of English vowel contrasts using an auditory and orthography test. The auditory test showed better discrimination and it was concluded that L2 auditory processing is subject to L1 orthography influence. The transparent nature of grapheme-phoneme correspondence in Greek negatively influences L2 pronunciation of English sounds that are represented by more opaque English orthographic convention.

4. Comparing Greek and English

Though the languages share speech sounds[1], including interdentals /θ, ð/ that are universally rare (Ladefoged & Maddieson, 1996), their consonant/vowel inventories and phonotactics differ in several respects.

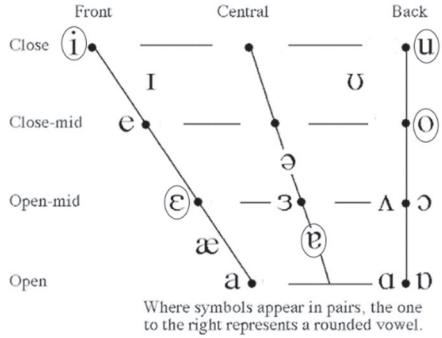


Figure 1. Greek and English vowel chart

Figure 1 charts Greek and English vowel inventories that differ in size and quality. Greek (circled) only has five vowels, two diphthongs, /a͡i/ /e͡i/, and permits vowel sequences, e.g. /o.o/, /e.i/ (Babatsouli, 2020). There are some differences in the vocalic inventories between Standard British English (SBE) and General American English (GAE). Their common monophthongs are: short /ɪ, e, æ, ʌ, ʊ/, long /i/ /ɛ/ /ɜ/ /ɑ/ /u/ /ɔ/, and weak /ə/. SBE also has /ɒ/ while GAE has /a/ and rhotic monophthongs /ɝ/ /ɚ/. Common diphthongs in the dialects are: /e͡ɪ/ /a͡ɪ/ /ɔ͡ɪ/ /a͡ʊ/ /ɪ͡ə/ /ʊ͡ə /, while specific to SBE is /ə͡ʊ/ and to GAE are /o͡ʊ/ and the rhotic diphthongs /ɪ͡ɚ/ /ɑ͡ɚ/ /ɛ͡ɚ/ /ʊ͡ɚ/. Among the diphthongs, /a͡ɪ/ /ɔ͡ɪ/ /a͡ʊ/ are phonemic, i.e. they cannot be produced as monophthongs without the word changing meaning, e.g. /ɔ͡ɪ/→[ɔ], e.g. *joy* /dʒɔ͡ɪ/→[dʒɔ͡ɪ] but not *[dʒɔ] because *jaw* /dʒɔ/→[dʒɔ]. However, the non-phonemic diphthongs are allophonically realized as monophthongs, e.g. *boat* /bo͡ʊt/→[bo͡ʊt]~[bot], *bake* /be͡ɪk/→[be͡ɪk]~[bek]. Both varieties also have triphthongs, as in ***fire*** /a͡ɪə/, ***player*** /eɪə/, ***employer*** /ɔ͡ɪə/, ***hour*** /aʊə/. SBE further includes /əʊə/ (as in *mower*), and GAE comprises respective rhotic variants /aɪɚ/ /eɪɚ/ /ɔ͡ɪɚ/ /aʊɚ/ /oʊɚ/.

Greek has no phonemic tense/lax, short/long distinction as in English, e.g. /i:/-/ɪ/, /u:/-/ʊ/, though vowels are slightly centered in unstressed syllables, and /i:/ /u:/ are allophonically laxed ([ɪ] [ʊ]) before voiceless consonants and utterance-finally in casual speech, also subject to within-speaker variation (Koutsoudas & Koutsoudas, 1985: 212). Greek vowels lengthen in stressed syllables, but not as prominently as English vowels before voiced consonants. English vowels are shorter before voiceless consonants, and also rhotacized in General American before the rhotic in syllable-final position, so: stressed /ɝ/: ***air**plane* [ˈɝˌpleɪn], and unstressed /ɚ/: *teach**er***

[ˈtʰiʧɚ], though not in British English, e.g. *airplane* [ˈeəˌpleɪn], *teacher* [ˈtʰiʧə]. In similar contexts, Greek has a clear vowel-rhotic segmental distinction, e.g. [**ɐɾ**. çi] 'beginning'. Importantly, English vowels in unstressed syllables are phonemically reduced to /ə/ and, less commonly, to /ɪ/ (e.g. *roses* [ˈroʊzɪz]), but Greek vowels maintain full quality (even in the phonetically laxed contexts), e.g. English *album* /ˈæl.bəm/, Greek /ˈɐl.b**u**m/. This is due to rhythm differences since Greek is syllable-timed (each syllable produced at approximately equal amounts of time), while English is stressed-timed (stressed syllables produced at approximately regular intervals but unstressed ones shortened). So, for example, *parameter*, in Greek /pɐ.ˈɾɐ.mɛ.tr**o**s/ will be [pə**.**ˈɹæ.mə**.**tɚ].

Table 1. Greek and English consonants

OBSTRUENTS

	Plosives	**Fricatives**	**Affricates**
Common	p b t d k g	f v θ ð s z	
Greek		x ɣ	ʦ ʣ
Allophones	ᵐp ᵐb ⁿt ⁿd ᵑg c ɟ	ç j	
English		ʃ ʒ h	ʧ ʤ
Allophones	pʰ tʰ kʰ		

SONORANTS

	Nasals	**laterals**	**rhotics**	**glides**
Common	m n	l		
Greek			ɾ	
Allophones	n̥	ʎ		
English			ɹ	j w
Allophones	m̥ n̥	l̥ ɬ	ɹ̥	

Table 1 groups common and language-specific consonants separately. Even among common consonants, there are phonetic differences, like voice onset time (see Kkese, 2016). Regarding differences, Greek has alveolar affricates, /ʦ/ /ʣ/ and velar fricatives, /x/ /ɣ/, while English has postalveolar affricates/fricatives, /ʧ/ /ʤ/ /ʃ/ /ʒ/, and a glottal fricative, /h/, instead. In Greek, the rhotic is a tap, /ɾ/, while in English it is an approximant, /ɹ/. English also has two glides, /j/ /w/, which are sometimes realized in vowel contexts in Greek, e.g. /ɣ**ai**ðɐros/→[ɣ**a**ʲðɐros] 'donkey', /frɐ.**u**lɐ/→[frɐʷlɐ] 'strawberry'. English /ŋ/ appears at the end of words, *sing* [sɪŋ] while, in Greek, /n/ assimilates to [ŋ] before /k/ /x/ /g/ /ɣ/.

The following Greek allophones are productive: /s/ becomes [z] before voiced consonants, e.g. /**sɣ**u.ˈɾɐ/→[**zɣ**u.ˈɾɐ] ‘curly ones (noun, neutral)’; voiced plosives /b/ /d/ /g/ are prenasalized [ᵐp] [ᵐb] [ⁿt] [ⁿd] [ᵑg], subject to dialectal, gender and generational variation, e.g. /d/→[ⁿd] [ɐvoˈkɐⁿdo] ‘avocado’; velars /k/ /g/ /x/ /ɣ/ are palatalized to [c] [ɟ] [ç] [j] before /i/ /ɛ/, and before /i/ followed by a vowel; /n/ and /l/ are palatalized to [ɲ] and [ʎ], respectively, before /i/ followed a vowel and, in some dialects, before /i/ or /ɛ/ (Babatsouli, 2019). In English, voiceless stops /p/ /t/ /k/ in stressed syllables are aspirated (produced with a burst of air) [pʰ] [tʰ] [kʰ], e.g. *a**pp**ear* (see Kkese, 2016 for VOT specifications). Nasals /m/ /n/ and liquids /l/ /ɹ/ are syllabified, [m̩] [n̩] [l̩] [ɹ̩], following obstruents in unstressed syllables, e.g. *bo**ttl**e* /ˈbɑ**tə**l/→[ˈbɑ**tl̩**]; this gives them vowel quality. A *dark* (velarized) allophone [ɬ] appears for /l/, when it follows a vowel or in syllable codas before a consonant, CV**C.**C, e.g. *pal* [pʰæ**ɬ**], *ki**l**t* [kɪ**ɬ**t], *el**b**ow* [ɛ**ɬ**.boʊ]; /s/ /z/ are alveopalatalized to [ʃ] [ʒ], respectively before /j/, e.g. *ble**ss** **y**ou* /blɛ**s**.**j**u/→ [blɛ.**ʃ**u]; plosives /t/ /d/ are replaced by [ɾ] between vowels when the first vowel is stressed, e.g. *ha**t**ed* [heɪ**ɾ**əd], *beauty* [ˈbju**ɾ**i].

Phonotactic differences between the languages involve differences in syllable and word structures. Monosyllables are more common and more complex in English, while in Greek they are few and mostly function words. Word-finally, Greek words can only end in /n/ or /s/, unless the word is a loanword, whereas English has a much richer consonantal inventory in that position. Clusters may comprise up to four consonants preceding or following a vowel in English, while Greek only has them word-initially and across syllables word-medially. For cluster types permitted in the languages, see Babatsouli and Geronikou (submitted) and McLeod et al. (2001). Both primary and secondary stress is productive in the languages.

5. Patterns of Greek speakers’ accent in English

This section first describes the sources of data and then presents an analysis of errors documented in the datasets employed. The chapter upholds that interlanguage productions are anticipated aspects of speech development (Babatsouli & Ingram, 2018) and, though called ‘errors’, they are considered normal realizations.

5.1 Method

Data used in the analysis and examples come from the author’s observations of adult speakers’ productions in actual conversations and from television

and radio broadcasts, collected in notes in the past ten years. Participants were Greek adults (>18 years) speaking standard modern Greek and were of various proficiency levels in English. Words comprising mismatches were transcribed in IPA symbols during initial note taking. Due to the nature of data collection, background information on AoO, LoL, and LoR is not available. The examples provided here are archetypes characteristic of the prosodic context of mismatches also noted in other words. The chapter is advanced as an evidence-based qualitative portrayal of L1Greek/L2English accent with the focus being on errors, though native-like productions across contexts were not absent from the available dataset.

5.2 Results

Tables 2 and 3 in the Appendix summarize errors for vowels and consonants, respectively. The production patterns are discussed next in terms of vowel mismatches, consonant mismatches, and prosodic influences.

5.2.1 Vowels errors

Among vowel mismatches, the five vowel Greek system is predominant. Specifically, the patterns are:

i) *English vowel distinctions are neutralized.*

Specifically, /ɔ:/-/ɒ/→[o], /ɑ:/-/a/-/æ/→[ɐ], /ɛ/-/ɜ:/-/æ/→[ɛ], /u:/-/ʊ/→[u], /i:/-/ɪ/→[i]. Thus, there is no contrast in pairs like: h*a*t/h*u*t, sh*i*p/sh*ee*p, b*i*tch/b*ea*ch, c*o*t/c*au*ght, *good*/*goose*.

ii) *Lax vowels are substituted by tense vowels*.

Examples are: *c**a**t* (/æ/→[ɐ]), *pl**ai**t* (/æ/→[ɛ]), *vill**a**ge* (/ɪ/→[ɐ]), *pr**e**fer* (/ɪ/→[ɛ]), *c**ou**ntry* (/ʌ/→[ɐ]~[u), *n**ou**rish* (/ʌ/→[ɛ]~[u).

iii) Monophthongs and diphthong onglides are lengthened in stressed contexts.

Examples are: *s**a***[ɐ:]*lmon* (/æ/→[ɐ:]), *l**o**ng* (/ɒ /→[o:]), *c**u**shion* (/ʊ/→[o:]), *b**o***[o:]*som* (/ʊ/→[o:]), *t**a**pe* (/eɪ/→[ɛ:.i]), *st**o**ne* (/əʊ/-/oʊ/

→[o:.u]), *d**ou**gh* (/əʊ/-/oʊ/→[ɐ:.u]).

iv) Reduced English vowels in unstressed syllables are produced as full vowels.

Unstressed /ɪ/ becomes [i]. Schwa, /ə/, is replaced by full vowels, commonly guided by the transparent grapheme-phoneme correspondence in Greek, e.g. ***a***[ɐ]*bove, oth****e***[ɛ]*r, fam****i***[i]*ly, parr****o***[o]*t, Sat****u***[u]*rday*.

v) Most diphthongs and triphthongs are produced as vowel hiatus, but some diphthongs are monophthongized.

Examples of hiatus are: *cash****ie****r* (/ɪə/-/ɪɚ//→[i:.ɛ]), *v****ou****cher* (/aʊ/→[o.u]),

and of monophthongs are: *c****a****ke* (/eɪ/→[ɛ]), *m****y****self* (/aɪ/→[ɐ], *s****ew***(/əʊ/-/oʊ/→[u]), *v****ou****cher* (/aʊ/→[u]).

vi) Rhotic vowels, e.g. /ɝ / /ɚ/ are produced as VC$_r$.

The same holds in the context of rhotic diphthongs and triphthongs, e.g. /eɚ/ /aɪɚ/, that are produced as hiatus followed by [ɾ] (V.VC$_r$, V.V.VC$_r$), e.g. ***flower***[ɐ.u.ɛɾ]. The [ɐ.u] is sometimes glided, ***flower***[**ɐ^w.ɛɾ**].

vii) /i/ is glided when preceding a vowel, as for contexts similar to /i/+V hiatus that triggers palatalized consonant productions in Greek.

Specifically, /i/ becomes [j], e.g. *comed****i***[j]*an, di*[j]*abolic, Louis****i***[j]*ana*, but not without exception, e.g. *mi****nia***[ni.ɐ]*ture*, *disc****ui***[u.i]*t*.

viii) English vowel digraphs, representing monophthongs, are often realized as vowel hiatus.

Examples are: *dr****aw***/ɔː/→[o.u]*er, for****ei***/ə/→[ɛ.i]*gn, surg****eo***/ə/→[ɛ.o]*n, cupb****oa***/ə/→[o.ɐ]*rd*, but exceptions are also evidenced, e.g. ***ea***/ɜ/→[ɛ]*rly.*

ix) There is influence instigated by grapheme-phoneme correspondences in Greek.

An /ɔː/→[u] pronunciation of English <ou>, e.g. ***pou***[u]*r,* and an /æ/→[ɛ] pronunciation of <ai>, <ae>, e.g. *pl****ai***[ɛ]*t*, are common, influenced by Greek <ου> and <αι>, respectively.

x) Vowel assimilation to another vowel in the word is less common, but notable.

An example is: *spe*[ɛ]*c****i***[ɛ]*fic*.

xi) Overgeneralization of an English rule in erroneous context.

Examples are: *c**u**lture, res**u**lt*, likely influenced by /ɔ:/→[o] in e.g. *c**a**ll, **a**lmost*,

xii) Renditions influenced by transparent approach to readings of orthography.

Examples are: *l**i**ngerie* (/æ/→[i]), *min**u**te* (/ɪ/→[u]), *qu**a**lify* (/ɒ/→[ɐ]), *c**ou**gh* (/ɒ/→[o]), *c**o**me* (/ʌ/→[o]), *Jan**ua**ry* (/ə/→[u.ɐ]), *cupb**oa**rd*, (/ə /→[o.ɐ]), ***a***[ɐ]*lmost*, and lastly,

xiii) *Consonant addition as lexical transfer*.

An example is: *dinos**au***[ɐv]***r*** from Greek /dinos**a**.**v**ros/

5.2.2 Consonant errors

Among consonant mismatches, Greek sound productions are predominant. Specifically, the patterns are summarized next, with targeted English allophones discussed separately.

i. English-specific sounds are produced as similar Greek consonants sounds or they are simplified, evidence of development.

Thus, obstruent productions are: /ʃ/→[s, z], /ʒ/→[z, s], /h/→[x, ç], /ʧ/→[ts, t, s], /ʤ/→[dz, d, z], /ŋ/→[n, ŋk, ŋg], and the rhotic approximant becomes a tap, /ɹ/→[ɾ]. Examples are: ***s***[s]*ugar, ca**sh***[z]*mere, massa**ge***[z], ***h***[x]*ave, act*[ts]*ual, **g***[z]*iraffe.* Palatal productions are discussed next. The glides become consonantal, e.g. ***y**ellow*(/j/→[ȷ]). Exceptions are noted for /w/ produced as a velar fricative followed by vowel prosthesis, /w/→[ɣu]~[xu] in e.g. ***w***[ɣu]*ord, **wh***[xu]*ich*. Among the patterns produced, also notable are simplifications documented for *affricates* (notably, either to a Greek-specific affricate, or to a component plosive or fricative) /ʧ/→[t, s], /ʤ/→[d, z], the *rhotic*, /ɹ/→[v, w] (e.g. *so**rr***[w]*y, **r***[ɣu]*abbit*), and *semi-vowel* /w/ vocalized to [u], ***o***[u.ˈɐ]*ne, **qu***[u]*ick*.

ii. *Palatal productions (transfers) are triggered by contexts similar to the Greek allophonic palatalization rule.*

Examples are: ***c***[c]*ake*, *etti**q***[c]*uette*, ***h***[ç]*istory*, *ha**g***[ʝ]*iography*, *inge**n***[ɲ]*ious*, *par**l***[ʎ]*iament*.

iii. *Rare productions influenced by English orthography,* e.g. ***j***[dz]*oj*[dz]*oba.*

iv. */s/ voicing in the context of a following voiced consonant*, e.g. ***s***[z]*mall, ca****sh***[z]*mere.*

v. *Consonant additions in Greek loanwords that are silent but orthographically transparent in English,*

e.g. ***p***[p]*salm*, *su****b***[b]*tle*, *apos****t***[t]*le*, *phle****g***[g]*m*, *hym****n***[n], *as****th***[θ]*ma*, ***m***[m]*neumonic*, ***p***[p]*neumonia*.

vi. *prenasalization of voiced plosives,*

e.g. la**d**[^{n}d], lu**gg**[^{ŋ}g]age.

vii. *Lexically influenced transfers*,

e.g. *pres****b***[v]*terian*, ***ch***[x]*aos*, *epo****ch***[x], *stoma****ch***[Ø]~[x], *phle****g***[g]~[ɣ]*m*, ha**g**[ɟ]iography, *catalo****g***[ɣ], *diaphra****g***[g]~[ɣ]*m*, ***x***[ks]*enophobia*.

vii) *Orthographically instigated additions of silent English graphemes,*

e.g. *recei****p***[p]*t*, *com****b***[b], *Chris****t***[t]*mas*, *mor****t***[t]*gage*, *san****d***[d]*witch*, ***k***[k]*nob*, *sc*[c]*ene*, ***g***[g]~[ɣ]*nome*, *as****th***[θ]*ma*, ***h***[x]*onest*, *ca****l***[l]*m*, *autum****n***[n]. The production in *s****w***[ɣu]*ord* is interesting in that silent <w> is produced by the respective substitution in targeted contexts.

ix) *Word-final deletions,*

e.g. *bathta****b***[Ø], *instinc****t***[Ø], *gingerbrea****d***[Ø], *hedgeho****g***[Ø].

x) *Word-final devoicing,*

e.g. *cu****b***[p], *insi****d****e*[t], *fi****ve***[f], *kni****ves***[fs], *syrin****g***[k], *massa****ge***[s], *spoon****s***[s].

Targeted English-specific allophones are produced as follows:

i. *Voiceless plosives are unaspirated,*

e.g. ***p****et* (p^{h}→[p]), ***t****ea* (t^{h}→[t]), ***c****at* (t^{h}→[t]), ***c****at* (k^{h}→[k]).

ii. *The velarized lateral, [ɬ], is produced as a clear lateral [l].*

e.g. *ba**ll*** (e.g. [bol]*, fe**ll**, a**l**batross*. It is notable that target-like productions are also evidenced in Greek speakers of Northern dialects where a velarized allophone also exists (e.g. Trudgill, 2003).

iii. *Syllabic consonants show most variation.*

In terms of (a) *transfer*, a similar consonant is produced for targeted syllabic consonant, [l̩]→[l], e.g. *bo**ttl**e*[tl]; (b) *development*, vowel epenthesis for targeted syllabic consonants, e.g. *a**bl**e*[bol]*, bo**ttl**e*[tɛl]*, bu**ckl**e*[kol], vocalic productions, and deletion of targeted syllabic consonant, e.g. *Lid**l***[∅].

iv. *Intervocalic allophone tap is non-existent in the dataset; substitutions are influenced by orthography*, i.e. plosive productions for targeted allophonic tap: /t/, /d/→[ɾ], e.g. *bu**tt***[t]*er, la**dd***[d]*er*.

5.2.3 Prosodic influences

Rhythm differences between the languages, alongside with the influence of orthographic readings, are causes of several mismatches at the syllable and word levels. Most notable are silent vowels that are orthographically transparent in English, realized as full vowels in Greek L2-English productions. This adds syllables to the word (see Table 2), e.g. *logic**a**lly* (/ˈlɑ.ʤɪ.kli/→[ˈlo.ʤɪ.**kɐ**.li]), *music**a**lly* (/ˈmju.zɪ.kli/→[mɲu.zɪ.**kɐ**.li])*, veg**e**table* (/ˈvɛ.ʤtə.bəl/→[ˈvɛ.**ʤɛ**.tɐbl]~[ˈvɛ.**ʤɛ**.tɐ.bol], *Wedn**e**sday* (/ˈwɛn.zdi/→[**ɣu**.ˈɛn.zdɛ.**i**]), *bus**i**ness* (/ˈbɪz.nəs/→[ˈbi.**zi**.nɛs]), *parl**i**ament* (/ˈpɑr.lə.mənt/ →[ˈpɐɾ.li.**ɐ**.mɛnt]). Lengthening also occurs when schwa and other monophthongs are substituted by a full vowel, and a diphthong or vowel hiatus, respectively, e.g. *p**y***[i]*jama* (/pəˈʤɑːmə/→[pi.ˈ**ʤɐ**.**mɐ**]), *fire* [fɐi.ɛɾ]. This is not to the exclusion of shorter timing units produced, as well, like monophthongizations in e.g. *c**a**ke* (/k**eɪ**k/→[ˈcɛk]). Other mismatches affecting syllable complexity involve variable cluster production patterns (e.g. Babatsouli & Sotiropoulos, 2018), such as reductions, member deletions, vowel epenthesis and substitutions, e.g. *pla**st**ic* (/ˈplae**st**ɪk/→[ˈplɐs**Ø**ik], ***t**rain* (/tɹeɪn/→[ˈts**Ø**ɛ.in], *glamour* (/ˈglæmɚ/→ [g**ɐ**.ˈlɐ.moɾ]. A comprehensive analysis of cluster productions, however, is beyond the scope of this chapter.

6. Overview

The dataset has provided evidence of mismatches at the segmental and prosodic levels that were primarily instigated by transfer of L1 structures, though universal, prototypical, and developmental processes are also documented.

Transfer is observed at the segmental, syllable, lexical and orthographic levels. Starting with segments, because Greek has a poorer vocalic inventory than English, English-specific monophthong contrasts are neutralized and produced as similar L1 monophthongs. Noticeable lengthening of stressed vowels in Greek (Arvaniti, 2007) is also characteristic of the L2 English monophthong and onglide productions evidenced. English-specific diphthongs and triphthongs are produced either as vowel hiatus (Babatsouli, 2020) of two or three members in the sequence respectively, also involving substitutions by similar Greek vowels. Transfer of similar Greek consonant sounds is also common for English-specific sounds, both targeted phonemes and their allophones. The transfers are characterized by Greek-specific phones, or allophones (i.e. palatals) in contexts likely to trigger the respective allophonic rule (i.e. /i/ and following vowel).

Interestingly, developmental patterns involving unmarked forms, like stopping, vocalizations, simplifications, mostly pertain to universally marked segments, like the rhotic and affricates (Ladefoged & Maddieson, 1996), but also marked allophones, as in the case of the velarized lateral, and marked prosodic structures, like syllabic consonants and clusters (e.g. Anderson, 1983). It is also interesting that the developmental patterns in these L2 productions for targeted *affricates*, the *rhotic*, and the labio-velar *semi-vowel*, /w/, (with the exception of /ɹ/→[ɣ]) are normative developmental substitutions L1 English acquisition (e.g. Smit et al., 1990) and in Greek-English child bilingualism (Babatsouli, 2015, 2019). Among the patterns produced, also notable are simplifications documented for *affricates* (notably, either to a Greek-specific affricate, or to a component plosive or fricative) /ʧ/→[t, s], /ʤ/→[d, z], the *rhotic*, /ɹ/→[v, w] (e.g. *so**rr***[w]*y, **r***[ɣu]*abbit*), and *semi-vowel* /w/ vocalized to [u], ***o***[u.ˈɐ]*ne, q**u***[u]*ick*.

Universal processes are also documented in the dataset like word-final deletions and devoicing (e.g., Flege & Davidian, 1984), as well as, known stages in cluster acquisition (e.g. Babatsouli & Sotiropoulos, 2018). While devoicing occurs irrespective of word length or complexity, word-final deletions predominate in complex words, like compounds and those comprising clusters (e.g. *bathta**b***[∅], *instinc**t***[∅]), or words that may be less

common in the language or, likely, less familiar to the L2 speaker (e.g. *hedgehog*[Ø]). Lastly, L2 English realizations are further subject to the influence of a transparent approach to orthographic readings, as is the case in the Greek language (e.g. Georgiou, 2021). This is clearly an across-the-board pattern instigating non-targeted addition of consonants, monophthongs, diphthongs, and syllables. While non-linguistic factors cannot be accounted for in the analysis of productions explored here, the effect of L1 production and perceptual practices is clearly paramount in the interlanguage evidenced. Greek L2 English speech is characterized by negative interference from L1 to L2 in terms of sounds, words, and the orthography, as well as by prototypical, universal, and developmental processes, and it is subject to dynamic processes guiding production variability (De Bot et al., 2007). The practical outcome of such interlanguage is a foreign accent that may or may not affect intelligibility, which depends on the extent of the errors and individual speaker variation.

The present chapter aimed to provide a rudimentary guide for identifying possible contexts of interlanguage errors in Greek speakers of English as an L2, and it has elucidated observed patterns and tendencies by providing apposite example archetypes that serve as a model for errors in similar contexts. The adopted approach is hoped to prove useful across different frameworks: i) to aid in teaching pronunciation within instructional contexts, ii) to serve as a self-contained instruction manual that may prompt further interest in the subject for individual learner use, and iii) as a much-needed benchmark of anticipated Greek L2-English productions to be used in clinical assessment and intervention protocols across the globe. Whether native-like competence and pronunciation can be achieved depends on how well the novel perceptual and articulatory skills are mastered and applied, but also on the degree to which speakers of the L2 can move beyond their familiar L1 language practices. As Koutsoudas and Koutsoudas (1985: 224) note:

> "…, we not only believe that the problem of learning a new set of language habits is much less difficult than that of breaking old habits, but also hold that the greater the number of old habits, the greater will be the difficulty of mastering the desired pronunciation."

Conclusively, it ought to be emphasized that language learning success and native-like performance and aptitude rely on a complex nexus of cognition, motoric skills, personality, and affective factors alongside variation in learning styles and variation in learning strategies (Ehram et al., 2003).

Note

1 An IPA chart with audio files of sounds can be found at: IPA Chart with Sounds International Phonetic Alphabet Sounds and IPA Online - Consonants (pulmonic) (ncl.ac.uk). For the articulation of sounds, see Sounds of Speech (uiowa.edu)

References

Anderson, J.I. 1983. Syllable Simplification in the Speech of Second Language Learners. *Interlanguage Studies Bulletin* 7(1), 4-36.

Archibald, J. 2019. A Unified Model of Mono- and Bilingual Intelligibility: Psycholinguistics Meets Pedagogy. *Journal of Monolingual and Bilingual Speech* 1(1), 8-31. https://doi.org/10.1177/13670069040080010301.

Arvaniti, A. 2007. Greek Phonetics: The State of the Art. *Journal of Greek Linguistics* 8, 97-208. https://doi.org/10.1075/jgl.8.08arv.

Babatsouli, E. 2015. Technologies for the Study of Speech: Review and an Application. *Themes in Science and Technology Education* 8 (1), 17-32.

Babatsouli, E. 2019a. Multilingualism. In *The SAGE Encyclopedia of Human Communication Sciences and Disorders*, edited by Jack. S. Damico and Martin J. Ball. Thousand Oaks: SAGE Publications.

Babatsouli, E. 2019b. Melodies of Child Greek-English Phonological Interference. *Revista de Estudos da Língua(gem)* 17(2), 209-236. https://doi:10.22481/el.v17i2.5347.

Babatsouli, E. 2019c. A Phonological Assessment Test for Child Greek. *Clinical Linguistics and Phonetics* 33(7), 601-607. https://doi: 10.1080/02699206.2019.1569164.

Babatsouli, E. 2020. Separation of Vowel Sequences by Consonant Addition in a Child's Bilingual Development. *Journal of Monolingual and Bilingual Speech* 2(2), 290-314. https://doi.org/10.1558/ jmbs.18848.

Babatsouli, E. 2021. Diversity Considerations in Speech and Language Disorders: A Focus on Training. In *Handbook of Language and Speech Disorders*, edited by J.S. Damico, N. Müller, and M.J. Ball, 2nd edition, 33-52. Hoboken: Wiley-Blackwell.

Babatsouli, E., and Geronikou, E. (submitted). Phonological Delay of Segmental Sequences in a Child's Greek.

Babatsouli, E., and Sotiropoulos, D. 2018. A Measure for Cluster Proximity (MCP) in Child Speech. *Clinical Linguistics and Phonetics*, 32(12), 1071-1089. https://doi:10.1080/02699206.2018.1510982.

Babatsouli, E. and Ingram, D. (eds.). 2018. *Phonology in Protolanguage and Interlanguage*. Sheffield, UK: Equinox Publishing.

Babatsouli E., and Kappa, I. 2010. Transfer of Greek palatals in L2 English.

In *Proccedings of the 6tn International Symposium on the Acquisition of Second Language Speech – New Sounds 2010,* edited by K. Dziubalska-Kołaczyk, M. Wrembel, and M. Kul. Poznań, Poland.

Babatsouli, E., and Kappa, I. 2011. Transfer of Greek palatals in L2 English: Practice and Theory. In *Achievements and Perspectives in the Acquisition of Second Language Speech: New Sounds 2010*, edited by Magdalena Wrembel, Malgorzata Kul and Katarzyna Dziubalska-Kołaczyk, 27-42. Frankfurt: Peter Lang.

Blake, H.L., Verdon, S., and McLeod, S. 2021. Multilingual University Students' Perceived English Proficiency, Intelligibility and Participation. *Journal of Monolingual and Bilingual Speech* 3(1), 146-170. https://doi.org/10.1558/jmbs.18179.

Chomsky, N., and Halle, M. 1968. *The Sound Pattern of English*. New York: Harper & Row.

Coutsougera, P. 2007. The Impact of Orthography on the Acquisition of L2 Phonology: Inferring the Wrong Phonology from Print. In *Proceedings of the Phonetics Teaching and Learning Conference 2007*, 1-5. London: University College London.

De Bot, K., Lowie, W., and Verspoor, M.H. 2007. A Dynamic Systems Theory Approach to Second Language Acquisition. *Bilingualism: Language & Cognition* 10, 7-21.

Dimitriou, D. 2018. L2 Acquisition and Production of the English Rhotic by L1 Greek-Cypriot Speakers: The Effect of L1 Articulatory Routines and Phonetic Context. *Philologia* 16(16), 45-64. https://doi:10.18485 /philologia.2018.16.16.3.

Eckman, F.R. 1977. Markedness and the Contrastive Analysis Hypothesis. *Language Learning* 27(2), 315- 330. https://doi.org/ 10.1111/j.1467-1770.1977.tb00124.x

Ehrman, M.E., Leaver, B.L., and Oxford, R.L. 2003. A Brief Overview of Individual Differences in Second Language Learning. *System* 31(3), 313-330. https://doi.org/10.1016/S0346-251X(03)00045-9.

Fishman, J. 1976. *Bilingual Education: An International Sociological Perspective*. Rowley: Newbury House.

Flege, J. 1995. Second Language Speech Learning: Theory, Findings, and Problems. In *Speech perception and linguistic experience: Issues in cross language research*, edited by W. Strange, 233-78. Baltimore: York Press.

Flege, J.E., and Davidian, R.D. 1984. Transfer and Developmental Processes in Adult Foreign Language Speech Production. *Applied Psycholinguistics* 5, 323-347. https://doi:10.1017/S014271640000 521X.

Georgiou, G.P. 2021. How Do Speakers of a Language with a Transparent Orthographic System Perceive the L2 Vowels of a Language with an

Opaque Orthographic System? An Analysis through a Battery of Behavioral Tests. *Languages* 6, 118. https://doi.org/10.3390/languages60 30118.

Giannakopoulou, A., Uther, M., and Ylinen, S. 2013. Enhanced Plasticity in Spoken Language Acquisition for Child Learners: Evidence from Phonetic Training Studies in Child and Adult Learners of English. *Child Language Teaching and Therapy* 29(2), 201-218. https://doi.org/10.1177/026565 9012467473

Harper, S., Goldstein, L., and Narayanan, S. 2016. L2 Acquisition and Production of the English Rhotic Pharyngeal Gesture. *Interspeech*, September 8-12, 2016, San Francisco, USA. http://dx.doi.org/10.21437/ Interspeech.2016-658.

Jusczyk, P. 2000. *The Discovery of Spoken Language*, 2nd edition. Cambridge: MIT Press.

Kkese, E. 2016. *Identifying Plosives in L2 English: The Case of L1 Cypriot Greek Speakers*. Bern: Peter Lang.

Kkese, E., and Karpava, S. 2021. Challenges in the Perception of L2 English Phonemes by Native Speakers of Cypriot Greek. *Journal of Monolingual and Bilingual Speech* 3(1), 1-39. https://doi.org/10.1558/jmbs.15362.

Kkese, E., and Petinou, K. 2017. Perception Abilities of L1 Cypriot Greek Listeners - Types of Errors Involving Plosive Consonants in L2 English. *Journal of Psycholinguist Research* 46(1), 1-25. https://doi:10.1007/s109 36-016-9417-3.

Koutsoudas, A., and Koutsadas, O. 1985. A Contrastive Analysis of the Segmental phonemes of Greek and English. In *Second Language Learning, Contrastive Analysis, Error Analysis, and Related Aspects*, edited by Robinett, B. and Schachter, J., 211-230. Ann Arbor: Michigan University Press.

Krashen, S.D. 1987. *Principles and Practice in Second Language Acquisition*. New York: Prentice Hall.

Kuhl, P. 1991. Human Adults and Human Infants Show a 'Perceptual Magnet Effect' for the Prototypes of Speech Categories, Monkeys Do Not. *Perception and Psychophysics* 50(2), 93-107. https://doi.org/10.3758/ bf03212211.

Ladefoged, P., and Maddieson, I. 1996. *The Sounds of the World's Languages*. Oxford: Blackwell.

Leather, J., and James, A. 1991. The Acquisition of Second Language Speech. *Studies in Second Language Acquisition* 13 (3), 305-341.

Lengeris, A., and Hazan, V. 2010. The Effect of Native Vowel Processing Ability and Frequency Discrimination Acuity on the Phonetic Training

of English Vowels for Native Speakers of Greek. *The Journal of the Acoustical Society of America* 128, 3757. https://doi.org/10.1121/1.3506351

Lengeris, A., and Nicolaidis, K. 2016. The Identification and Production of English Consonants by Greek speakers, *Selected Papers of the 21st International Symposium on Theoretical and Applied Linguistics* (ISTAL 21), 224-238. https://doi.org/10.26262/istal.v21i0.5227.

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

Major, R.C. 2001. *Foreign Accent: The Ontogeny and Phylogeny of Second Language Phonology*. Hillsdale, NJ: Lawrence Erlbaum.

Major, R.C. 2008. Transfer in Second Language Phonology. In *Phonology and Second Language Acquisition*, edited by J.G.H. Edwards and M.L. Zampini, 63-94, Philadelphia/Amsterdam: John Benjamins.

McLeod, S., van Dorn, J., and Reed, V. 2001. Normal Acquisition of Consonant Clusters. *American Journal of Speech-Language Pathology* 10 (2), 99-110. https://doi.org/10.1044/1058-0360(2001/011).

Millard, K., and Yavaş, M. 2018. Schwa Productions in Spanish-English Bilingual Adults. In *Phonology in Protolanguage and Interlanguage*, edited by E. Babatsouli and D. Ingram, 201-230. Sheffield: Equinox Publishing.

Rice, K., and Avery, P. 1995. Variability in a Deterministic Model of Language Acquisition: A Theory of Segmental Elaboration. In *Phonological Acquisition and Phonological Theory*, edited by J.A. Archibald, 23-42. Hillsdale: Lawrence Erlbaum.

Smit, A.B., Hand, L., Freilinger, J.J., Bernthal, J.E., & Bird, A. 1990. The Iowa Articulation Norms Project and its Nebraska Replication. *Journal of Speech and Hearing Disorders* 55, 779-798. https://doi:10.1044/jshd.5504.779.

Trubetzkoy, N.S. 1939/1958. *Fundamentals of Phonology*. Göttingen: Travaux du Cercle Linguistique.

Trudgill, P. 2003. Modern Greek Dialects. *Journal of Greek Linguistics* 4, 45-64. https://doi:10.1075/jgl.4.04tru.

Trudgill, P. 2008. Native Speaker Segmental Phonological Models and the English Lingua Franca Core. In *English Pronunciation Models: A Changing Scene,* edited by K. Dziubalska-Kołaczyk and J. Przedlacka, 77-98. Wien: Peter Lang.

Werker, J., and Tees, R. 1984. Cross-language Speech Perception: Evidence for Perceptual Reorganization During the First Year of Life. *Infant Behavior and Development* 7(1), 49-63. https://doi.org/10.1016/s0163-6383(84) 80022-3.

Appendix

Table 2. L1Greek/L2English vowel errors

L1 English	L2 English	Example archetypes
Monophthongs		
Short vowels		
/æ/	[ɐ]~[ɐ:] [ɛ]~[ɛ:] [i]~[i:]	***ca***[ɐ]*t, s**a**[ɐ:]lmon,* *pl**ai**[ɛ]t, G**ae**[ɛ]lic,* *l**i**[i]ngerie*
/ɛ/	[ɛ] [ɐ] [ɛ.o]	*p**e**t, necess**a**ry, s**ai**d, s**ay**s,* ***a**ny, fr**ie**nd, j**eo**[ɛ.o]pard*
/ɪ/	[i] [ɐ] [u] [ɛ] [ɛ.o] [ɛ.i]	*p**i**t, cr**y**stal, **e**nough, forehe**a**d,* *w**o**[u]men, min**u**[u]te, vill**a**[ɐ]ge,* *pr**e**[ɛ]fer, spec**i**[ɛ]fic,* *pig**eo**[ɛ.o]n, Wedn**e**sd**ay**[ɛ.i]*
/ɒ/	[o]~[o:] [ɐ] [o.u]	*l**o**[o:]ng, h**o**[o]nest, sw**a**[ɐ]llow,* *qu**a**[ɐ]lify, c**ou**[ɐ]~[o]gh, y**a**[ɐ]cht,* ***Au**stralia, ackn**ow**[o.u]ledge*
/ʌ/	[ɐ] [ɛ] [u] [o]	***o**ther, d**oe**s, fl**oo**d,* *c**ou**[ɐ]~[u]ntry,* *n**ou**[ɛ]~[u]rish,* *ab**o**[o]ve, c**o**[o]me, l**o**[o]~[ɐ]ve,* *c**u**[o]lture, res**u**[o]lt*
/ʊ/	[u]~[u:] [o]~[o:]	*p**u**t, c**ou**ld, c**u**shion, b**o**[o]som*
/ə/	[ɐ] [ɛ] [i] [o] [u] [Ø] deletion [ɛ.i] [ɛ.o] [u.i] [u.ɐ] [i.ɐ] [o.ɐ] [i.u] [o.u]	***a**[ɐ]bove, pleas**a**[ɐ]nt,* *chronologic**a**[ɐ]l,* *probl**e**[ɛ]m, oth**e**[ɛ]r, cert**ai**[ɛ]n,* *acr**e**[Ø],* *p**y**[i]jama, fam**i**[i]ly,* *f**o**[o]rget, forw**a**[o]~[ɐ]rd,* ***o**[o]bscure, parr**o**[o]t, creat**o**[o]r,* *Aug**u**[u]~[ɐ]st,* *Sat**u**[ɛ]~[u]rday, car**ou**[u]sel,* *fail**u**[u]re, fab**u**[u]lus,* *mortg**a**[ɛ.i]]ge,* *for**ei**[ɛ.i]gn, surg**eo**[ɛ.o]n,* *bisc**ui**[u.i]t, circ**ui**[u.i]t* *Janua[u.ɐ]ry, min**ia**[i.ɐ]ture,* *cupb**oa**[o.ɐ]rd, anx**iou**[i.u]s,* *sorr**ow**[o.u]ful*

Long vowels		
/ɑː/	[ɐ] [ɐː]	*f**a**rther, f**a**ther, baz**aa**r, ball**a**de, moust**a**che*
/ɔː/	[o] [ɐ] [u] [ɐ.u]~[ɐː.u] [o.u]~[oː.u] C-addition	*f**o***[o]*r*, ***au***[o]*tumn*, *c**a***[oː]*ll*, ***a***[o]~[ɐ]*lmost, w**a***[oː]*ltz c**au**se, ch**a**lk, c**ou**rse, p**a***[ɐ]*lsy, **a***[ɐ]*ltar, w**a***[ɐ]*rd, p**ou***[u]*r*, ***au***[ɐ.u]*rora, l**aw***[o.u]*s*, ***aw***[o]~[o.u]*ful, dr**aw***[o.u]*er*, abr**oa***[o.u]*d, dinosa**u***[v]*r*
/ɜː/-/ɝ/	[ɛː] [ɛɾ] [ɐ] [iː] [oː] [u] [u.i]	***ea**rly, l**ea**rn, pref**e**r, b**u***[ɛɾ]*rial*, l**u***[ɐ]*rk*, f**i***[iː]*r*, wh**i***[iː]*rl*, m**y***[iː]*rtle*, w**o***[oː]*rth*, , c**olo***[o]*nel*, f**u***[uː]*r*, b**ou***[uː]*rbon*, j**ou***[uː]*rnal*, sq**ui***[u.i]*rrel*
/iː/	[ɛ.i] [i.ɐ] [i.ɛ.i] Gk palatalization	*counterf**ei***[ɛ.i]*t*, voll**ey***[ɛ.i], Disn**ey**, d**ea***[i.ɐ]*l*, st**ea***[i.ɐ]*l*, qu**ay***[i.ɛ.i], com**ed**i*[j]*an*, d**i***[j]*abolic*, Louis**i***[j]*ana*
/uː/	[o] [u.i]	*g**r**o**u**p*, ***ou**zo*, *can**oe***[o], *adi**eu***[o] ***sui***[u.i]*t*
Diphthongs		
/eɪ/	[ɛ] [ɛ.i]~[ɛː.i] [ɐ.u]	*th**ey**, d**ay**, m**a**jor, freight, t**a**pe, prev**ai**l, ph**a**se, communic**a**te, c**a***[ɛ]*ke, g**au***[ɐ.u]*ge*
/aɪ/	/ɐ/ /ɐi/ /i/	*m**y***[ɐ]~[ɐi]*/self, soc**ie**ty, qu**i**te, beh**i**nd, as**y***[i]*lum, cy***[i]***press*
/əʊ/ /oʊ/	[o.u] [oː.u] [ɐ.u] [o] [u]	*epis**o**de* [o.u], *st**o**ne*[oː.u], ***o**we, d**ou***[ɐː.u]*gh, f**oe***[o], *d**o***[o]*n't, w**o***[o]*n't, plat**eau***[o], *bur**eau***[o], *s**ew***[u]
/ɔɪ/	[o] [o.u] [o.i]~[oː.i] [o.u]~[oː]	*r**o***[o]~[o.u]*yal*, ***oi**l, b**oy**, ch**oi**ce, l**aw***[o.u]*yer*
/aʊ/	[u] [ɐ.u] [oː.u]	*dev**ou***[u]*r*, ***ou**r*[ɐ.u], *t**ow***[ɐ.u]~[ɐ^{w}]*n*, *fl**ou***[ɐ.u]~[ɐ^{w}]*r, v**ou***[oː.u]~[u]*cher* b**ou***[o.u]*gh, h**ow***[o.u]*l*

/eə/-/eɚ/	[ɐ] [ɛ] [ɛ.ɐ]~[ɛ:.ɐ]	*v**a**[ɐ]rious, hil**a**[ɐ]rious,* *f**ai**[ɛ]ry, y**ea**[ɛ]h,* ***ai**[ɛ.ɐ]r, c**a**re, ch**ai**r, t**ea**r, th**ei**rs*
/ɪə/-/ɪɚ/	[i] [i:.ɛ] [i.ɐ] [i:.ɐr] [i.o]	*ch**ee**[i]r, sn**ee**[i]r, chasm**e**[i]re,* *p**ie**[i.ɛ]r, cash**ie**[i:.ɛ]r,* ***ea**r, car**eer**, r**ea**lly,* *fash**io**[i.o]n, extens**io**[i.o]n*
/ʊə/	[u.e]	*cr**uel**, j**ewel**,*
Triphthongs		
/aɪə/-/aɪɚ/	[ɐi.ɛr]	***fire**, ch**oir**, l**ier***
/eɪə/-/eɪɚ/	[ɛ.i.ɛr]	***player**, l**ayer**, m**ayor***
/ɔ͡ɪə/~/ɔ͡ɪɚ/	[o.i.ɛr]	***lawyer**, empl**oyer**, distr**oyer***
/əʊə/~/oʊɚ/	[o.i.ɛr]	***mower**, g**oer**, s**lower***
/aʊə/-/aʊɚ/	[ɐ.u.ɛr]~[ɐ^{w}.ɛr]	*f**low**[ɐ.u.ɛr]~[ɐ^{w}.ɛr]er,* *h**ou**[ɐ.u]~[ɐ^{w}]r*
Silent vowels		
<a> <e> <i>		*logic**a**[ɐ]lly, music**a**[ɐ]lly,* *physic**a**[ɐ]lly, veg**e**[ɛ]ta[ɐ]ble,* *Wedn**e**[ɛ]sday,* *bus**i**[i]ness, parl**i**[i]ament*

Table 3. L1Greek/L2English consonant errors

	L1 English	L2 English	Example archetypes
Singletons			
plosives	/p/ [p^{h}] silent <p>	deaspiration addition	***p**[p]et, a**pp**[p]ear,* ***p**[p]salm, **p**[p]neumonia* *de**b**[p]t, cou**p**[p],* *cu**p**[p]board,* *ras**p**[p]berry, recei**p**[p]t*
	/b/ silent <b>	final devoicing/deletions additions Gk prenasalization [v]	*cu**b**[p], bathta**b**[ø]* *bom**b**[b], lam**b**[b],* *com**b**[b], su**b**[b]tle,* *a**b**[^{m}b]out* *pres**b**[v]terian*
	/t/ [t^{h}] [ɾ] silent <t>	deaspiration final deletion addition	***t**[t]ea,* *deposi**t**[ø], effor**t**[ø],* *instinc**t**[ø], foresigh**t**[ø],* *significan**t**[ø]* *bu**tt**[t]er, la**dd**[d]er,* *Chris**t**[t]mas,* *apos**t**[t]le,* *mor**t**[t]gage,* *mois**t**[t]en, so**ft**[t]en*

	/d/ silent<d>	final devoicing/deletions addition prenasalization	*cried*[t], *inside*[t], *gingerbread*[Ø] *sand*[d]*witch,* *hand*[d]*kerchief,* *Wed*[d]*nesday,* *avocad*[ⁿd]*o, lad*[ⁿd]
	/k/ [kʰ] silent<k> silent<c>	deaspiration word-final deletion [x] [ts] Gk palatalization addition	***c***[k]*at, acc*[k]*ustomed* ***ch***[x]*arisma,* ***ch***[x]*aos,* *stoma**ch***[Ø]~[x], *epo**ch***[x], *zu**cch***[ts]*ini, bro**ch***[ts], ***c***[c]*ake, s**k***[c]*etch,* ***ch***[c]*emical, or**ch***[c]*id,* *etti**q***[c]*uette,* ***k***[k]*nuckle, **k***[k]*nob,* *s**c***[c]*ene*
	/g/ silent<g>	final devoicing/deletions [ɣ] addition prenasalization Gk palatalization	*catalo**g***[k]~[ɣ], *hedgeho**g***[Ø], ***g***[g]*naw,* *diaphra**g***[g]~[ɣ]*m,* *phle**g***[g]~[ɣ]*m,* ***g***[g]*narl, **g***[g]~[ɣ]*nome,* *si**g***[ɣ]*nificant,* *resi**g***[ɣ]*n,* *lu**gg***[ⁿg]*age,* *Hilfi**g***[ⁿg]*er,* *hag*[ɻ]*iography*
fricatives	/f/	-	-
	/v/	final devoicing	*fi**ve***[f], *remo**ve***[f],
	/s/ silent<s>	voicing addition	***s***[z]*mall, cas**h***[z]*mere,* *ai**s***[s]*le*
	/z/	final devoicing [ts] [ks]	*leave**s***[s], *knive**s***[s], *clo**se***[s], *hou**se***[s], ***ts***[ts]*ar,* ***x***[ks]*enophobia,* *an**x***[ks]*iety*
	/ʃ/	[s] [z] [ts]	***s***[s]*ugar, fa**sh**ion, fi**sh**,* *por**t**ion, deli**c**ious,* *admi**ss**ion,* *cas**h***[z]*mere,* ***ch***[ts]*andelier*
	/ʒ/	[z] final devoicing	*vi**s***[z]*ion, plea**s***[z]*ure,* *bei**ge***[z], *massa**ge***[z] ~[s]

	/θ/ silent <th>	addition	*as**th**[θ]ma*
	/ð/	final devoicing	*loa**the**[θ], soo**the**[θ]*
	/h/ silent <h>	[x] [dz] Gk palatalization addition	***h**[x]ave, **wh**[x]o,* *per**h**[x]aps,* ***j**[dz]oj[dz]oba,* ***h**[ç]istory, be**h**[ç]ave,* *ad**h**[ç]ere, faj[j]ita* ***h**[x]otel, **h**[x]onest*
nasals	/m/ [m̩]	deletion vocalization vowel epenthesis addition	*bo**mb**[Øb],* ***m**[m]neumonic*
	/n/ [n̩] silent <n>	deletion Gk palatalization vocalization vowel epenthesis addition	*television**[Ø],* *o**n**[ɲ]ion, Jan[ɲ]uary,* *inge**n**[ɲ]ious, m[ɲ]usic,* *kitte**n**[tn]~[tɛn]~[tɛ],* *butto**n**[tn]~[ton]~[tɐ],* *colum**n**[n], hym**n**[n]* *autum**n**[n]*
	/ŋ/	[n] [ng] [nk]	*ba**ng**, lu**ng**, to**ngue**,* *readi**ng**, lo**nger***
liquids	/l/ [ɬ] [l̩] silent <l>	[l] vocalization vowel epenthesis palatalization deletion addition	*ba**ll**[l], fe**ll**, a**l**batross,* *a**ble**[bl]~[bol]~[bo],* *bo**ttle**[tl]~[tɛl]~[tɛ],* *bu**ckle**[kl]~[kol]~[ko]* *Lid**l**[Ø],* *portfo**l**[ʎ]io,* *par**l**[ʎ]iament,* *a**l**[l]mond, ha**l**[l]f,* *yo**l**[l]k, ca**l**[l]m*
	/ɹ/ /ɻ/	final deletion [ɾ] [w] [ɣu] vowel epenthesis vocalization	*dinosau**r**[Ø],ai**r**[ɾ],* ***r**[ɾ]iver, so**rr**[w]y,* ***r**[ɣu]abbit, butte**r**[ɛɾ],* *hist**ory**[ori]*

affricates	/tʃ/ /tj/→[tʃ]	[ts] [t] [tç] [k] [t]	***ch****urch, du****ch****ess,* *ac****t***[ts]*ual, con****c***[ts]*erto,* *ri****ght***[ts]*eous,* *ques****tio*** [tsi.o]*n,* *ques****t***[ti.o]~[tç]*ion,* *sandwi****tch***[ts], ***t***[tç]*une,* *T****ue***[ç]*sday, int*[ç]*uition,* ***t***[t]*ulip, spat*[t]*ula,* *whi****ch***[t], ***ch***[k]*ore*], *ar****ch***[k]
	/dʒ/	[dz] [d] [z] [g] [dj] palatalization [Ø]	*ma****j***[dz]*or, j****ar,*** *ve****g****etable, sur****g****eon,* ***j****u****dg****e, e****d****ucation,* ***d***[d]*uet,* ***g***[z]*iraffe,* *syrin****g***[g]~[k], ***d***[dj]~*uring* *pla****g***[ɟ]*iarism*
glides	/j/ /ju/	Gk palatalization [i] [Ø]	***y***[j]*ellow, Lafa****y***[j]*et,* ***u***[j]*nion, v****i***[j]*ew,* ***d***[j]*uring,* *f****ew***[ç], *f****u***[ç]*ture,* *p****u***[ç]*re, fai****l***[ʎ]*ure,* *var****i***[i]*able, comed****i***[i]*an,* *on****i***[Ø]*on, purs****ue***[Ø], *h****i***[Ø]*est*, a*nno****y***[Ø]*ance*
	/w/ silent <w>	[ɣu] [xu] [u.ɐ] addition	***w***[ɣu]*ord,* ***w****alk, je****w****el,* ***wh***[xu]*ich,* ***o***[u.ɐ]*ne, q****u***[u]*ick,* *lang****u***[u]*age,* *c****o***[u]*iffure, c****h****oir* *s****w***[ɣu]*ord*

PART II:

PRACTICAL PERSPECTIVES OF SPEECH PERCEPTION AND PRODUCTION IN L2

CHAPTER SIX

REALIZATIONS OF /D/ IN SPANISH PAST PARTICIPLE -AR VERBS: A SOCIOPHONETIC ANALYSIS BETWEEN NATIVE SPEAKERS AND L2 LEARNERS

GIULIA CORTIANA AND OLGA TARAROVA

Abstract

The present sociophonetic study aimed to compare the phenomenon of complete elision in Spanish past participle *-ar* verbs among Latin American Spanish native speakers and English-speaking Spanish L2 learners. This study examined seven Latin American Spanish native speakers (NS) and six English-speaking learners of Spanish (NNS) as second language (L2), by using a preference forced choice (T3) and a reading task (T4). Stimuli included 20 Spanish past participle *-ar* verbs, each three-syllable stimulus was used as an adjective or as a verb alternatively across T3 and T4. Our dependent variable was the complete elision, and our independent variables were sex, age, and country of origin as well as country where NNS learnt Spanish. 780 tokens were collected, 420 from the NS group and 360 from the NNS group. Results showed that NS speakers produced a complete elision at a rate of 35%, while NNS produced a complete elision at a lower rate of 13%. Step-up/step-down multivariable analyses were carried out with Goldvarb statistical software. Our results indicate that sex and age had a significant effect across the NS group; the country where NNS learnt Spanish had a significant effect among NNS. When NNS did not produce complete elision, they produced 1.39 times longer approximants [ð] than their NS counterparts. Our results also seemed to suggest that NNS produced a third allophone, the fricative [ð], at a rate of 41%. Furthermore, our dataset demonstrated a strong positive correlation (R^2 = .55) between NNS' dominance score and their rate of approximant [ð] productions. Our

results seemed to suggest that the more dominant a NNS was, the more native-like their productions were.

Keywords: *Spanish, sociophonetic, complete elision, L2 production, past participle -ar verbs*

1. Introduction

This study focuses on production of the Spanish intervocalic voiced dental plosive /d/ in past participle verbs that belong to the first verbal conjugation group (e.g., *bailado*, [bai.ˈla.**ð**o], 'danced'), among Spanish native speakers and English-speaking learners of Spanish as a second language. When the /d/ is found in the aforementioned linguistic context, the phoneme /d/ presents 2 allophones: an approximant [ð] (e.g., *bailado*, [bai.ˈla.**ð**o], 'danced'), and a complete elision [Ø] (e.g., *bailado*, [bai.ˈla.o], 'danced'). This phenomenon is known to occur with high frequency in many dialects and varieties of Iberian (among others, Molina Martos, 2001; Gómez Molina, Molina Martos, & Paredes García, 2012) and Latin American Spanish (Navarro, 1982; Hualde, 2005; Colantoni & Marinescu, 2010; Radu, 2014; Solon, Linford, and Geeslin; 2018). However, most previous studies have mostly focused on native speakers of Iberian Spanish. Moreover, very few studies have contrasted the production of native speakers with learners of Spanish as second language (Zampini, 1994; Rogers & Alvord, 2014; Solon, Linford, & Geeslin; 2018). Therefore, the present pilot study is a preliminary comparative sociophonetic analysis between native speakers of Latin American Spanish and English-speaking Spanish learners in their production of the voiced dental plosive /d/ in past participle -ar verbs (e.g., *bailado*, [bai.ˈla.**ð**o], 'danced').

This paper aims to answer three major research questions:

a. Do native speakers of Spanish produce more complete elision [Ø] than their L2 counterparts?
b. If NNS participants do not produce a complete elision [Ø], how do their productions differ acoustically from those of their native counterparts?
c. Which demographic variables, such as sex, age, and country of learning, have the greatest effect on both groups?

2. Literature Review

2.1 /d/ realizations among Spanish native speakers

Navarro's (1982) investigation seemed to confirm that both the approximant [δ] and the complete elision [Ø] are phenomena that also appear in Cuban Spanish. While listing Cuban Spanish sounds that could suffer from a 'relaxation' process, Navarro included the voiced dental plosive /d/. According to her investigation (1982), the Cuban Spanish approximant [δ] occurs in all phonetic contexts, except for the absolute initial position, after the nasal /n/ or after the lateral /l/. Navarro (1982) seemed to confirm the complete elision of /d/ in past participle endings, at a higher rate among -ar verbs and a lower rate among -er and -ir verbs. Navarro's (1982) investigation was purely theoretical and, therefore, lacked information regarding whether these phenomena do happen among Cuban Spanish speakers.

Radu (2014) tested four Colombian speakers and found that the complete elision appears at a lower rate than it does in other Spanish varieties. Her four Colombian speakers completed two reading tasks and two spontaneous tasks. Radu carried out acoustic analyses with PRAAT and measured duration to properly classify the sounds. Her results showed that it was not possible to account for complete elision in Colombian Spanish. However, her paper lacked the necessary information regarding gender and origin of participants, as well as information regarding the stimuli she used. Nevertheless, her results seemed to agree with Hualde (2005), who suggested that the Colombian Spanish voiced dental stop /d/ can also occur in the intervocalic position.

2.2 /d/ realizations among L2 Spanish leaners

Zampini (1994) investigated the production of Spanish voiced stops /b, d, g/ among five Spanish, from Colombia or Puerto Rico, NS and 32 English-speaking university students of Spanish. Her results showed that English speakers produced the approximant [δ] at a lower rate because of articulatory difficulties, because [δ] is an allophone of /d/ in Spanish, whereas it does not exist in English.

Rogers and Alvord (2014) studied three Spanish NS and eight adult learners of Spanish, four of whom had not spent time in a Spanish-speaking country (AB), and four of whom had (UL). Both groups completed a background questionnaire, the "Survey of Motivational Intensity", and two oral speaking tasks. Rogers and Alvord collected 423 tokens from the NNS

group and 83 from the NS group. They carry out acoustic analyses in PRAAT, as for duration of /d/ and its allophones, and oral constricting. Their results suggest that, regardless of their language knowledge level, L2 learners hardly show the same parameters as their native counterparts.

Solon, Linford, and Geeslin (2018) studied the production of /d/ among L3 Spanish native speakers and 13 Spanish L2 learners. Their NS group was composed of seven males and six females, from several Spanish-speaking countries, all residing in the US, while the NNS group was composed of nine males and four females, who were either graduate students of Spanish or teaching Spanish at the university. Both groups performed three tasks: a background questionnaire, a proficiency test, and a digitally recorded sociolinguistic interview. They collected a total of 3.612 tokens, 2.168 from the NS group and 1.444 from the NNS group. Their results showed that NS produced a complete elision [Ø] at a rate of 44.5%, while NNS produced a complete elision [Ø] at a rate of 18%. They also found that the preceding vowel had a significant effect (p <.001) in the production of a complete elision [Ø], for both groups. Overall, their results regarding the NNS group seemed to suggest that participants who had spent a significant amount of time in a Spanish-speaking country produced more complete elisions than participants who had not, which was similar to native-like realizations.

2.3 Social factors – sex, age, country of learning

Molina Martos (2001) and Gómez Molina et al. (2012) investigated participants from Madrid and Valencia, Spain, and found that women are more conservative; that is, they produce more approximant [δ] than their male counterparts, which also agrees with sociolinguistic results of Labov's (1975). Labov (1975) argued that women produced conservative, yet prestigious forms; and that is, the approximant [δ] is considered more prestigious than the complete elision.

Also, Molina Martos (2001) and Gómez Molina et al. (2012) seemed to suggest that younger speakers (< 35 y.o.) produced complete elisions at a higher rate than older speakers, which also correlates with Labov (1975). More recently, Rogers and Alvord (2014)'s research also suggested how participants from a Spanish-speaking country (AB) showed measurements closer to their native counterparts.

2.4 Linguistic factors and parameters for acoustic analyses

Previous research (Escudero, 2001; Ruiz-Peña, Sevilla & Rafat, 2017) suggested that vowel length is a speech production feature that some L2 learners tend to acquire with time and practice, while other studies (e.g., Cordero et al., 2018) showed that some L2 learners have difficulty acquiring length in a target-like manner.

Colantoni and Marinescu (2010) studied six male Argentinian Spanish speakers performing sociolinguistic interviews. They collected a minimum of 20 tokens per stop consonant per participant, and acoustically analyzed them in PRAAT, measuring duration. Their results showed that duration could be considered a valuable measurement in phonetics when comparing NS and NNS productions. They also found that the complete elision is more favored if the /d/ is preceded by /a/ or /o/, and that the /d/ is more likely to suffer a complete elision if it is found in an unstressed syllable, which is consistent with the aforementioned studies. Moreover, Molina Martos (2001) and Gómez Molina et al. (2012) study showed how the linguistic context that most favors the complete elision is when /d/ is preceded by the central front open vowel /a/.

3. Research questions and hypotheses

The first research question asked if Spanish native speakers (NS) produced more complete elision than their L2 English-speaking learners (NNS). Based on Navarro (1982), Zampini (1994), Radu (2014), and Solon, Linford, and Geeslin (2018), our first hypothesis (H1) predicted that NS will produce more complete elisions [Ø] than their NNS counterparts.

The second research question examined how NNS productions acoustically differed from NS productions, if NNS participants did not produce a complete elision [Ø]. Based on Rogers and Alvord (2014), our second hypothesis (H2) predicted that NNS would not show parameters equal to their native counterparts. Specifically, it was foreseen that NNS production would be longer.

Finally, the third research question asked which demographic variable, among sex, age, and country of learning, had the greatest effect. Based on Labov (1975), Molina Martos (2001) and Gómez Molina et al. (2012), our third hypothesis (H3) anticipated that male speakers and young speakers would produce more complete elisions. Based on Rogers and Alvord

(2014), it was also expected that, among English-speaking leaners of Spanish, the country of learning would present a greater effect.

4. Methodology

4.1 Participants

The present study included two groups of participants: seven native speakers of Latin American Spanish (NS) and six English-speaking learners of Spanish (NNS).

Among the NS group, three participants were from Colombia, and four from Cuba; four were female speakers (two from Colombia, and two from Cuba), and three were male speakers (one from Colombia and two from Cuba). Participants were labelled according to the initial of their country of origin (C stands for Colombia and B stands for Cuba), to a chronological number of participation in the study, and to their declared sex in the questionnaire (F if female or M if male). For example, participants C_01_M was the first male participant from Colombia, while participant B_02_F was the second female participant from Cuba. The NS' age range was from 18 to 42 (mean 31), and participants were grouped according to their age; that is, younger than 25 years old, between 26 and 35 years old, and older than 36 years old.

The NNS group included two participants from the United States, one from Canada, one from the United Arab Emirates, one from Kenya, and one from Jamaica. In the NNS group, three participants were female, and three were male. Participants were labelled according to their order in participating in the project, as to their declared sex in the questionnaire (F if female or M if male). For example, L_01_F was the first female Spanish L2 English-speaking participant, while L_02_M was the second male Spanish L2 English-speaking participant. The age range is from 20 to 32 (mean 26) and participants were grouped according to their age; that is, younger than 25 years old, and between 26 and 35 years old. Three NNS participants learnt Spanish in a Spanish-speaking country (one in Dominican Republic, and two in Spain), while three learnt Spanish in an English-speaking country (Canada, Jamaica, and United Kingdom). Participants from both groups were undergraduate and graduate students at the University of Western Ontario, London, from different departments and programs.

4.2 Tasks

Participants were required to complete four tasks. First, they completed a language background questionnaire, adapted from Tararova (2016). Then, they completed the Bilingual Language Profile (Birdsong, Gertken, & Amengual, 2012). A preference forced choice task (T3) followed. During that task, participants were asked to read 19 sentences aloud, followed by a related question and the three possible answers. They were also required to choose an answer. However, since this is a phonetic study, the choice of answer was not taken into consideration while performing the analyses. Finally, they completed a reading task (T4) that included 19 sentences. All the sentences provided in Task 3, the preference forced choice task, and in Task 4, the reading task, included of only one phoneme /d/, the target sound. Overall, participants completed the whole study in 25 minutes (SD = 6 minutes).

4.3 Stimuli

The stimuli were 20 *-ar* verbs, conjugated as past participles (e.g., bailado [bai. la.**ð**o], 'danced'). Each stimulus was used once as verb and once as adjective, alternatively across Task 3 and Task 4. Each stimulus had three syllables, and the target sound /d/ was always found in an unstressed position. The positioning of the stimuli was randomly distributed across Task 3 and Task 4, and so was their positioning within the sentence (i.e., initial, middle, and final position).

4.4 Variables

The present study uses Goldvarb as software for statistical analyses. As the program only allows for two variables at a time, the 'step-up/step-down' multivariable analysis was carried out with either the presence, regardless of the allophone, or the omission (i.e., complete elision) of the Spanish /d/. Therefore, the complete elision [Ø] and the presence of any allophone of /d/ are the dependent variables. The following independent variables were also taken into consideration: sex, age, country (of origin for NS, and of learning for NNS).

5. Procedure and Data Analyses

A total of 780 tokens were collected, 420 from the NS group and 360 from the NNS group. Seventeen tokens were excluded from the present study

when participants mispronounced the target word (*bailando*, [bai.ˈlan.**ð**o], 'dancing', gerund instead of a participle), and the intended phonetic context was lost. Each participant was recorded individually. From the whole *.wav* file, the stimuli were extracted and labeled according to the word that presented the target sound /d/. Acoustic analyses were then carried out in PRAAT (Boersma & Weenink, 2019), and each target sound was classified as an approximant [ð], a complete elision [Ø] or 'other', according to duration and spectrograms. Data were later inserted in an Excel file. Data were run in Goldvarb.

6. Results

A total of 763 tokens (780 minus 17 tokens) were acoustically analyzed using PRAAT (Boersma and Weenink, 2019). The percentage of complete elisions, the percentage of approximants, and the corresponding duration were measured for each participant first, and later overall as a group.

6.1 Results regarding the Spanish native speakers' group

A total of 418 tokens were collected from the NS group, and acoustically analyzed using PRAAT (Boersma & Weenink, 2019). Table 1 presents a summary of the data collected among the NS group regarding the production of complete elision per participant. As Table 1 shows, NS participants overall produced complete elision at a rate of 35% (N= 148/418).

Table 1. Percentage of complete elision of /d/ per participant in the NS group.

	PARTICIPANT'S CODE							AVERAGE
	C_01_M	C_01_F	B_01_M	B_02_M	B_01_F	B_02_F	C_02_F	
% (N= /tokens) COMPLETE ELISION OF /D/	43% (N= 26/60)	28% (N= 17/60)	52% (N= 31/60)	32% (N= 19/59)	27% (N= 16/60)	22% (N= 13/59)	43% (N= 26/60)	**35% (N= 148/418)**

According to multivariate analysis, carried out by a logistic regression with complete elision as the dependent variable, sex and age had a significant effect, as shown in Table 2.

Table 2. Step-up/step-down multivariable analysis across the NS group.

Corrected mean			0.0120
Log likelihood			-264.066
Total N			418
	FW	%ND	N/Total
Sex			
Male	**.64**	42.5	76/179
Female	.39	30.1	72/239
Range	*25*		
Age			
Younger Group (>25)	.47	32.2	19/59
Middle Group (26-35)	**.67**	36	86/239
Older Group (<36)	.29	35.8	43/120
Range	*38*		
Country			
Colombia	[.46]	38.3	69/180
Cuba	[.53]	33.2	79/238

Table 2 shows the distribution of the complete elision variable with respect to each variant of the independent variables. Male participants produced the highest rate of complete elision (FW .64), while female participants produced complete elisions at a lower rate. Participants included in the 'middle group' produced complete elisions at a higher rate (FW .67), in comparison to the older group who produced complete elisions at a lower rate (FW .29). The country of origin had no significant effect.

6.2 Results regarding the English-speaking learners of Spanish group

A total of 345 tokens were collected from the NNS group, and acoustically analyzed using PRAAT (Boersma & Weenink, 2019). Table 3 presents a summary of the data collected among the NNS group regarding the production of complete elisions. As Table 3 shows, NNS participants overall produced complete elision at a rate of 13 (N= 45/345), which is higher compared to their native speaker counterparts.

Table 3. Percentage of complete elision of /d/, per participant in the NNS group.

	PARTICIPANT'S CODE						AVERAGE
	L_01_F	L_01_M	L_02_F	L_03_F	L_02_M	L_04_M	
% (N= /tokens) COMPLETE ELISION OF /D/	8% (N= 5/60)	2% (N=1 /60)	14% (N= 8/58)	18% (N= 10/55)	16% (N= 9/56)	21% (N= 12/56)	**13% (N= 45/345)**

According to multivariate analysis, carried out by logistic regression with complete elision as the dependent variable, the country of learning had significant effect, as shown in Table 4.

Table 4. Step-up/step-down multivariable analysis across the NNS group.

Corrected mean			5.5094
Log likelihood			-128.437
Total N			345
	FW	%ND	N/Total
Sex			
Male	[.51]	12.8	22/172
Female	[.50]	13.3	23/173
Age			
Younger Group (>25)	[.44]	16.0	27/169
Middle Group (26-35)	[.57]	10.2	18/176
Country			
Spanish-speaking	.38	7.9	14/178
English-speaking	**.62**	18.6	31/167
Range	*24*		

Table 4 shows the distribution of the complete elision variable with respect to each variant of the independent variables. Participants who have not spent time in a Spanish-speaking country produced the highest rate of complete elisions (FW .62), while participants who have spent time in a Spanish-speaking country produced complete elisions at a lower rate. Sex and age had no significant effect.

6.3 Comparative analysis of results

With regards to the first research questions as to whether native speakers of Spanish produce more [Ø] than their L2 counterparts, results showed that the NS group indeed produced more complete [Ø] elisions than the NNS group. Figure 1 shows that NS participants produced a compete elision [Ø] at a rate of 35%, while the NNS participants' rate was just 13%.

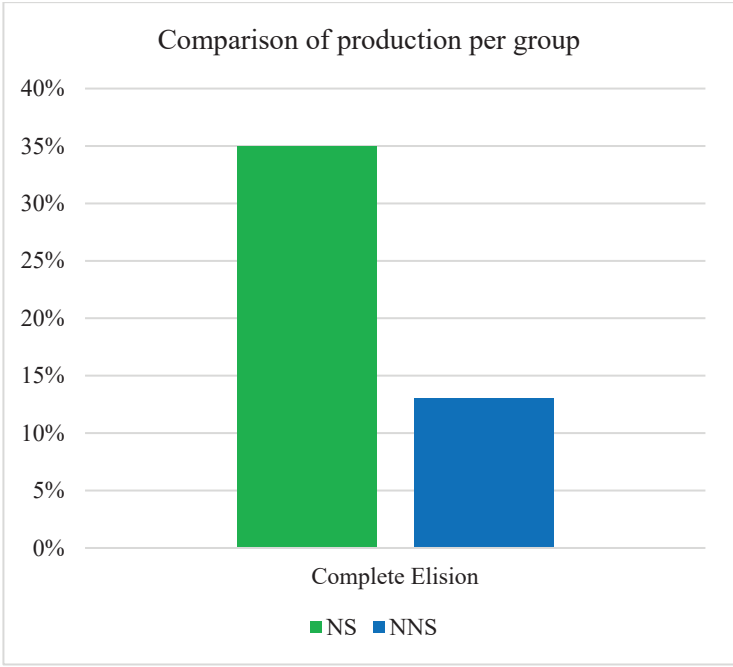


Figure 1. Percentage of complete elision of /d/ in both groups.

As for the second research question, regarding how NNS productions differ acoustically from those of their native counterparts when the NNS participants do not produce a complete elision [Ø], results show that NNS Spanish approximants are on average 1.39 time longer than those of their native counterparts, as shown in Table 5.

Table 5. Mean duration of approximant productions in both groups.

Group	Mean approximant duration
NS	327.42ms (SD= 76.13)
NNS	456.34ms (SD= 41.13)

With regards to the third research question, as to which demographic variable, such as sex, age, and country, has the greatest effect on both groups, results show that, among the NS group, sex and age are the most significant variables, and, among the NNS group, the country of learning is the most significant variable, as shown in Table 6.

Table 6. FW factors and significance in both groups.

NS participants

	FW	***%ND***	***N/Total***
Sex			
Male	**.64**	42.5	76/179
Female	.39	30.1	72/239
Range	*.25*		
Age			
Younger Group (>25)	.47	32.2	19/59
Middle Group (26-35)	**.67**	36	86/239
Older Group (<36)	.29	35.8	43/120
Range	*.38*		

NNS participants

	FW	***%ND***	***N/Total***
Country			
Spanish-speaking	.38	7.9	14/178
English-speaking	**.62**	18.6	31/167
Range	*24*		

7. Discussion

The first hypothesis predicted that NS would produce more complete elisions [Ø] than their NNS counterparts. This hypothesis has been confirmed, as NS participants produced complete elisions [Ø] at a higher rate than NNS participants. Results align with previous studies (Navarro, 1982; Zampini, 1994; Radu, 2014; Solon, Linford, & Geeslin, 2018).

The second hypothesis predicted that if NNS participants did not produce a complete elision [Ø], their production would not show parameters equal to their native counterparts. This hypothesis has also been confirmed, as NNS productions of the allophone [ð] are 1.39 longer than NS productions, on average. These results concur with previous research (Rogers & Alvord, 2014).

The third hypothesis predicted that, among Spanish native speakers, male speakers and young speakers would produce more complete elisions, and also that, among English-speaking learners of Spanish, the country of learning would present a greater effect. This hypothesis has been partially confirmed. Molina Martos' (2001) and Gómez Molina's et al. (2012) results show that native speakers of Spanish younger than 35 produced more complete elisions [Ø]. However, in the present study, the 'middle age' group, that included participants between 26 and 35 y.o., seems to have produced the most complete elisions [Ø], while younger speakers (<25 y.o.) produced significantly fewer complete elisions. These data correlated with Labov (1975), Molina Martos (2001), and Gómez Molina et al. (2012). In the NS group, male speakers did indeed produce more complete elisions [Ø] than their female counterparts, which also aligned with Labov (1975) as the complete elision is not the most prestigious realization. In the NNS group, the country of learning was a significant factor (FW .62, *p*. 0.005). These results contradict previous research (Solon, Linford, & Geeslin, 2018), which suggested that participants who have spent a significant amount of time in a Spanish-speaking country produce more complete elisions than participant who have not. Nevertheless, the present study combined all NNS Spanish-speaking countries into one category, regardless of whether the complete elision [Ø] is a phonological phenomenon that appears, and at what rate it might appear, in particular countries. Moreover, the present study did not take into consideration whether the NNS participants' Spanish professors came from countries where the complete elision [Ø] is a phonological phenomenon that occurs. It could be that their professors' varieties of Spanish played a role in these results. However, further research is needed in order to verify this hypothesis.

Results also indicate that [ð] productions of English-speaking learners were longer than their native counterparts. Even if some previous research suggested that length is a feature that learners tend to acquire (Escudero, 2001; Ruiz-Peña & Rafat, 2017), other studies showed that learners have difficulty acquiring length in a target-like manner (e.g., Cordero et al., 2017). It is recommended that further research is carried out regarding the

duration of approximant productions in Spanish native speakers and English-speaking learners of Spanish as second language.

What this study also found is that NNS participants produced a third allophone of /d/; that is, the fricative [ð] at a rate of 41% (N= 142/345). While our results aligned with previous studies (Zampini, 1994; Rafat, 2011), they could be explained by exploring the correlation between NNS Spanish proficiency and their [ð] productions. During the analyses, a preliminary correlation between the NNS participants' dominance score obtained by Task 2, the Bilingual Language Profile, and the percentage of [δ] productions was carried out. As shown in Figure 3, the R^2 equalled .55, which meant that 74% of the variation in the production of approximant could be accounted for by the participants' language dominance score. The standardized correlation coefficient 0.74 indicated a strong positive correlation between language dominance and production of approximant. Results seemed to suggest that the more Spanish dominant a NNS participant was, the more native-like their [δ] productions were.

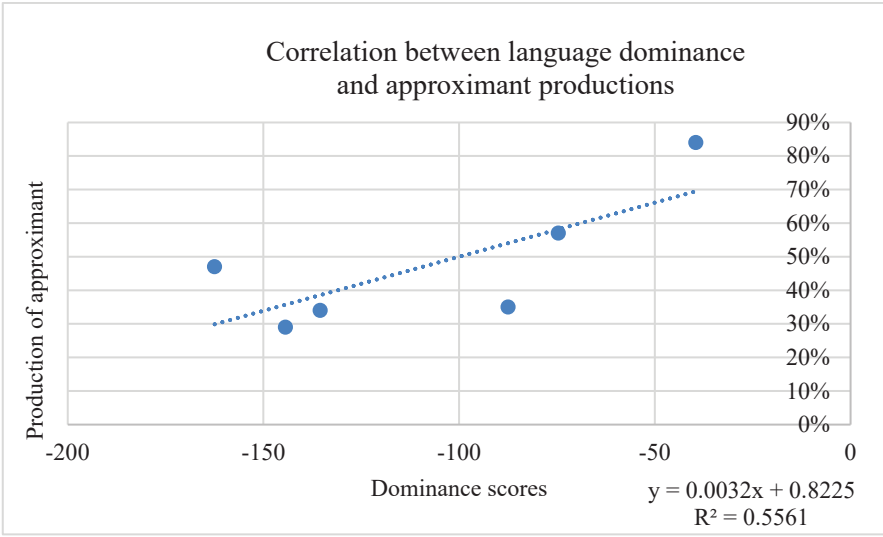


Figure 3. Correlation between NNS participants' dominance and their [δ] productions.

8. Conclusion

The present study aimed to be a preliminary comparative sociophonetic analysis between native speakers of Latin American Spanish and English-speaking Spanish learners regarding the production of the voiced dental plosive /d/ in past participle *-ar* verbs (e.g., *bailado*, [bai.ˈla.**ð**o], 'danced').

The goal of this study was to compare the production of complete elisions [Ø]. Seven native speakers of Latin American Spanish (NS) and six English-speaking learners of Spanish (NNS) completed four tasks, including a background questionnaire, the BLP, a preference forced choice (T3), and a reading task (T4). Our results showed that NS produced complete elisions [Ø] at a higher rate than NNS, that sex and age were significant factors within the NS group, and that the country of learning was a significant factor for the NNS group. Our results also suggested that NNS' [ð] productions were 1.39 longer than those of their NS counterparts. It also seemed that there was a strong positive correlation ($R^2 = .55$) between NNS participants; that is, between language dominance and production of approximants [ð]. Finally, our results interestingly suggested that NNS participants produced a third allophone, the fricative [ð].

However, the present study presents some limitations. First, in the NNS group both undergraduate and graduate students were included but their number was not equal. Further research should include an equal number of learners of Spanish, depending on their level of proficiency, in order to verify if there is a correlation between NNS participants' scores and their [ð] productions. Another limitation is the presence of just one male Colombian Spanish native speaker in the NS group. Further studies should include more male Colombian Spanish speakers, to have an equal number of male native speakers from different Spanish-speaking countries taken into consideration. Also, it must be pointed out that Task 3 and Task 4 were created from scratch by one of the authors: the goal of presenting just one /d/ in the target word/position, sometimes made the sentence sound not 100% native like, as some Spanish-speaking participants observed after completing the study.

The next steps for this study include presenting individual variations, as well as carrying out step-up/step-down multivariable analyses regarding the [ð] productions, in order to have clearer results regarding a comparison between native speakers of Spanish and English-speaking learners of Spanish. Moreover, based on our results regarding the correlation between proficiency and [ð] productions by English-speaking Spanish leaners, it is

recommended that further research is carried out regarding the correlation between NNS participants' dominance and their [ð] productions.

References

Birdsong, D., Gertken, L. M. and Amengual, M. 2012. *Bilingual Language Profile: An Easy-to-Use Instrument to Assess Bilingualism.* COERLL, University of Texas at Austin. Web.

Boersma, P. and Weenink, D. 2019. *Praat: doing phonetics by computer /Computer program*/.

Colantoni, L. and Marinescu, I. 2010. The Scope of Stop Weakening in Argentine Spanish. *Proceedings of the 4th Conference on Laboratory Approaches to Spanish Phonology*, *2010,* 100-114. Somerville, MA: Cascadilla Proceedings Project.

Cordero Varona, D., Ruiz-Pena, E., Sierra, E., Stevenson, R. and Rafat, Y. 2018. Second dialect and second language imitation of geminates by Colombian Spanish speakers. *Proceedings of the International Symposium on Monolingual and Bilingual Speech, 2015*, 99-106. Chania, GR: Institute of Monolingual and Bilingual Speech.

Escudero, P. 2001. The role of the input in the development of L1 and L2 sound contrasts: language specific cue weighting for vowels. *The 25th annual Boston University conference on language development Proceedings, 2001*, 250-261. Somerville, MA: Cascadilla Proceedings Project.

Gómez Molina, J. R., Molina Martos, I. and Paredes García, F. 2012. La variación de [d] intervocálica en comunidades de habla conservadoras de España. *ALFAL International Congress Proceedings*, *2012*, 1975-1986. Alcalá: ES: University of Alcalá.

Hualde, J. I. 2005. *The Sounds of Spanish*. Cambridge, UK: Cambridge University Press.

Labov, W. 1975. *Sociolinguistic Patterns*. Philadelphia: PA: University of Pennsylvania Press.

Molina Martos, I. 2001. Geografía y estratificación social de un cambio fonético: la -d en español peninsular. *Verba 28*, 81-99. Santiago de Compostela, ES: University of Santiago de Compostela.

Navarro, M. (1982). The Relaxing of the Consonants /b,d,g,k,h,c,y,f,s/ in Cuban Spanish. Deseret Language and Linguistic Society Symposium: 8(1), Article 7. https://scholarsarchive.byu.edu/dlls/vol8/iss1/7

Radu, M. 2014. Weakening of intervocalic stops in Colombian Spanish. *Paper presented at the 18th Hispanic Linguistics Symposium, 2014*, in West Lafayette, Indiana, US.

Rafat, Y. 2011. Orthography-induced Transfer in the Production of Novice Adult English-speaking Learners of Spanish. (Unpublished Doctoral Dissertation, University of Toronto, Canada).

Rogers, B. and Alvord, S. 2014. The gradience of spirantization: Factors affecting L2 production of intervocalic Spanish [β, ð, ɣ]. *Spanish in Context 14*, 402-424. Amsterdam: NL: John Benjamin Publishing Company.

Ruiz-Peña, E., Sevilla, D., and Rafat, Y. 2017. The acquisition of second dialect speech: an acoustic examination of the production of Ecuadorian Spanish assibilated rhotics by Andalusian speakers of Spanish. In *Phonology in Protolanguage and Interlanguage*, edited by E. Babatsouli and D. Ingram, 286-315. Sheffield, UK: Equinox Publishing.

Solon, M., Linford, B. and Geeslin, K. L. 2018. Acquisition of Sociophonetic variation: Intervocalic /d/ reduction in native and non-native Spanish. *Revista española de lingüística aplicada* 31(1), 309-344. Amsterdam: NL: John Benjamin Publishing Company.

Tararova, O. 2016. Negative doubling in the Italo-Mexican community of Chipilo, Mexico. *Revista Española de lingüística aplicada* 29(2), 582-612. Amsterdam: NL: John Benjamin Publishing Company.

Zampini, M. L. 1994. The Role of Native Language Transfer and Task Formality in the Acquisition of Spanish Spirantization *77*, 470-481. Hispania Madrid: ES: Editorial CSIC.

CHAPTER SEVEN

PERCEPTUAL TRAINING EFFECTS ON THE PERCEPTION AND PRODUCTION OF L2 ENGLISH VOWELS BY L1 GREEK-CYPRIOT LEARNERS

DIMITRA DIMITRIOU

Abstract

Segmental perception and production pose considerable difficulties for L2 learners as an effect of the influence of the L1. High Variability Phonetic Training (HVPT) has been proposed as a way to provide naturalistic exposure to L2 learners in order to facilitate the formation of L2 phonetic categories. This study examined the effects of HVPT on Greek-Cypriot learners' perception and production of L2 English contrastive vowels. Participants' perceptual and production performance was assessed before and after exposure to 11 English vowels over 8 training sessions. The HVPT paradigm used was effective in improving perceptual performance, which reflected durational results in production, with nativelike tendencies in /iː/-/ɪ/ and /uː/-/ʊ/, and difficulties in /ɒ/-/ɔː/, /ɑː/-/æ/-/ʌ/ and /e/-/ɜː/. This alignment was not extended to spectral characteristics in learners' productions: /ʊ/, /uː/, /ɑː/ and /iː/ were the least nativelike, while /ɒ/, /ɔː/ and /æ/ were the most nativelike. These results are not entirely consistent with previous studies, especially in the case of /e/-/ɜː/, most likely due to methodological differences, especially in relation to tasks, groups of participants and sample size. However, they are promising, as they suggest that improvement is possible if enriched training paradigms are used.

Keywords: HVPT, pronunciation training, vowel perception, vowel production, Greek-Cypriot learners, L1-L2 interaction

1. Introduction

The perception and production of second language (L2) phonetic categories is arguably one of the most challenging tasks L2 learners face (e.g. Iverson et al., 2003; Flege & MacKay, 2004; Lengeris, 2009; Dimitriou, 2018), especially when the sounds of an L2 contrast are not contrastive in the learners' first language (L1). This influence of the L1 on L2 segmental acquisition is reflected in current models of L2 speech perception and production, which attempt to predict difficulties in L2 segmental learning and explain how the L1 and L2 phonetic subsystems of learners interact. Two of the most widely tested models are the Speech Learning Model (SLM) (Flege, 1995; 2002) and the extension of the Perceptual Assimilation Model (PAM) (Best, 1995) aimed at L2 learners, i.e., the PAM-L2 (Best & Tyler, 2007), both arguing that the mechanisms used by children in language learning remain available to adults learning an L2.

The SLM focuses on speech production, arguing that new category formation is inhibited when an L2 sound is perceived as similar to an L1 sound, but also incorporates perceptual processes, suggesting that production performance is hindered by insufficient perceptual ability (Flege, 1995; 2003). As such, production can be improved if perception is enhanced, which can be achieved with sufficient exposure to native-speaker input (Flege, 1995; 2003). However, segmental perception is not always perfectly aligned with segmental production, as learners may not update their articulatory habits even when accurate perception of a sound is achieved (Flege, 1999). L1 interference and adult learners' perception of non-native phonemes is also addressed by the PAM-L2, which argues that upon encountering a new sound, learners attempt to assimilate it to existing L1 perceptual categories, especially when it is perceived to be similar to an L1 phoneme (Best, 1995; Best & Tyler, 2007). The PAM-L2 provides six assimilation types of contrastive non-native sounds, which enable predictions for the degree of L2 discrimination accuracy. These are Two-Category Assimilation, Category-Goodness Difference, Single-Category Assimilation, Uncategorised-Uncategorised, Uncategorised-Categorised, and Non-assimilated (detailed in Best, 1995; Best & Tyler, 2007).

Based on the assumptions of these models, Greek-Cypriot (CG) learners are likely to face difficulties in the perception and production of L2 English vowels due to the differences between the L1 and L2 vowel inventories. CG is a dialect of Standard Modern Greek (SMG) spoken mainly on the island of Cyprus by approximately 1 million speakers (Arvaniti, 2006; Simaki et al., 2015). The dialect shares the same five vowel qualities as SMG

(/i,e,a,o,u/) (Arvaniti, 2010), although the vowels of the two varieties have "some small but noticeable differences in terms of the position they occupy in the vowel space" (Georgiou, 2019: 4). Standard Southern British English (SSBE), on the other hand, has tense-lax vowel distinctions, with five tense (/iː,uː,ɜː,ɔː,ɑː/) and six lax (/ɪ,ʊ,e,æ,ʌ,ɒ/) vowels (Deterding, 1997; Roach, 2004).

Previous studies show that distinguishing between the members of English vowel contrasts is problematic for SMG and CG learners, since their L1 has a smaller vowel inventory and does not employ spectral or durational cues to signal phonological contrast (e.g. Lengeris, 2009; Lengeris & Hazan, 2010; Georgiou, 2019). Given that L2 learners are less accustomed to attending to acoustic cues that are not employed in their L1 (Escudero, 2005; Kkese & Petinou, 2017), CG learners are likely to perceive two contrasting L2 vowels as exemplars of the same category, and produce them without durational or spectral differences, merging them in a single, L1-based category (Flege, 2002). Similar patterns have been observed for Spanish learners of English, whose L1 only has five vowels (Cebrian, 2006; Iverson & Evans, 2009).

Georgiou (2019), for example, examined the perception of English vowels by native CG speakers and found that these learners assimilate more than one L2 vowel to a single L1 category, as they occupy an area of the vowel space that is filled by a single L1 category. The assimilation of English vowel contrasts to CG categories was found to be as follows: English /ɪ/ and /iː/ were assimilated to CG /i/, English /e/ and /ɜː/ were assimilated to CG /e/, English /æ/, /ʌ/ and /ɑː/ were assimilated to CG /a/, English /ɒ/ and /ɔː/ were assimilated to CG /o/ and English /ʊ/ and /uː/ were assimilated to CG /u/. This is consistent with previous studies investigating the perception of vowels by learners whose L1 involves a smaller vowel inventory than their L2 (e.g. Escudero, 2005; Aliaga-Garcia & Mora, 2009; Carlet & Cebrian, 2014 for Catalan and Spanish learners; Lengeris, 2009 for SMG learners).

Importantly, as suggested by the SLM and previous studies (e.g. Flege & Liu, 2001; Flege & MacKay, 2004; Flege, 2009; Thomson, 2018), these difficulties can be overcome through exposure to high-quality nativelike input. However, exposure to such input is problematic in EFL contexts, where the L2 is mostly acquired through formal instruction by non-native teachers, sometimes with noticeable foreign accents. In addition, despite the increasing evidence that L2 pronunciation instruction can be beneficial for learners' L2 speech performance (see Derwing et al., 2014; Krzonowski et al., 2015, among others), it remains a highly neglected aspect of language

teaching (Piske et al., 2001; Derwing & Munro, 2005; Kyprianou, 2007; Papachristou, 2011; Lengeris, 2018; Georgiou, 2019). This means that learners have limited opportunities to encounter nativelike input, and are therefore prevented from attuning their L1 phonetic categories to the acoustic properties of the L2 sounds (Georgiou, 2019).

In order to address this issue, and reflecting the importance assigned to the quality of the L2 experience, phonetic training has been proposed as a way to compensate for the lack of naturalistic exposure required for the formation of L2 phonetic categories in EFL classrooms, by providing targeted input and exposure to numerous samples of the stimuli in various phonetic contexts and from multiple speakers (Sakai & Moorman, 2018; Thomson, 2018). The High Variability Phonetic Training (HVPT) approach has been found to improve the perception of both consonants (Lively et al., 1993; Pruitt et al., 2006; Carlet & Cebrian, 2014) and vowels (Wang & Munro, 2004; Nishi & Kewley-Port, 2007; 2008; Iverson & Evans, 2009; Iverson et al., 2012; Carlet & Cebrian, 2014), and to help learners improve their production without explicit pronunciation instruction (e.g. Bradlow et al., 1997; Lengeris & Hazan, 2010; Thomson, 2011).

The effectiveness of HVPT on SMG learners' perception and production of English vowels was examined by Lengeris (2009; 2018) and Lengeris and Hazan (2010), who found that production improvement was achieved as a result of perception training. Lengeris (2009) reports that SMG trainees learned to differentiate the English vowels spectrally instead of using their five L1 vowel qualities for the production of English vowels, and improved their durational distinctions, approximating target durations. Lengeris and Hazan (2010) found that HVPT improved the identification of L2 vowels, and that learning was transferred to L2 vowel production as well, since learners initially used their five L1 categories in their L2 productions, but there was considerably less overlap between English vowels after the training.

Based on the findings of previous studies as outlined above, this study aims to add to the existing literature by incorporating this new group of learners, focusing on the largely neglected EFL context. The Research Questions (RQs) that this paper aimed to address are the following:

1. Does HVPT help adult CG learners improve their perception of contrastive L2 English vowels?
2. How effective is HVPT in improving CG learner's productions of L2 English vowels?

The perceptual patterns reported in Georgiou (2019) enable predictions for the production of English vowels by CG learners, who are expected to use their five L1 vowels for the production of the L2 vowels similarly to their perceptual patterns. Georgiou (2019) reports a Category-Goodness Difference or an Uncategorised-Categorised assimilation pattern for the English vowel contrasts /ɪ/-/iː/ and /e/-/ɜː/, and a Single-Category assimilation pattern for the English vowel contrasts /æ/-/ʌ/ and /ɒ/-/ɔː/. Assuming that better discrimination will lead to better production, and based on the assumptions of the PAM-L2 that Single-Category contrasts are the most difficult to discriminate, it is expected that CG learners of English will find it more difficult to accurately produce the vowels /æ/-/ʌ/ and /ɒ/-/ɔː/.

2. Methodology

2.1 Participants

8 native English (NE) speakers and 5 native Greek-Cypriot (CG) learners took part in this study. CG participants (4 female, 1 male) were 18- to 28-year-old students of various disciplines at an English-speaking university in Cyprus (intermediate to advanced learners of English). The NE speakers (5 female, 3 male) were 19- to 28-year-old students of various disciplines at an English-speaking university in the UK. The target variety was Standard Southern British English (SSBE) as in previous studies (e.g. Lengeris, 2009; Giannakopoulou et al., 2013; Krzonowski et al., 2015, among others). All participants were volunteers and the procedure adhered to all ethical standards. None of the participants reported having a speech or hearing impairment.

2.2 Procedure

NE speakers completed two wordlist-reading tasks, while CG participants completed a pre-test, eight perceptual training sessions, and an immediate post-test. The procedure for CG participants is outlined in Table 1 below. All participants completed a linguistic background questionnaire before the tasks, providing information about their linguistic and educational background, in order to ensure that the two groups are as homogeneous as possible. Participants were encouraged to take breaks between the perceptual and production tasks and halfway through the stimuli in each task, which were taken as needed.

Table 1. Data collection procedure (CG participants)

Phase	Tasks
Pre-test	1. Linguistic Background Questionnaire 2. Perceptual Test: Forced-Choice Identification Task (FCID) without feedback 3. Production Task: Elicited imitation task 4. Production Task: Cypriot-Greek wordlist reading
Training (8 sessions)	HVPT: FCID with feedback
Post-test	1. Perceptual Test: FCID without feedback 2. Production Task: Elicited imitation task

2.3 Stimuli

2.3.1 Natural Stimuli

The stimuli used in the study are given in Table 2. Although real words were preferred when available, both real and non-words (indicated by asterisks) were included (similarly to Bohn & Bundgaard-Nielsen, 2008; Mayr & Escudero, 2010) in order to ensure that the vowels appear in matching environments in both languages (Di Paolo et al., 2011), and direct the attention of the trainees only on phonetic form and not meaning (Thomson, 2011). The use of a disyllabic CVCV rather than a monosyllabic CVC structure for the CG stimuli was made to ensure that the stimuli appear in phonotactically permitted sequences, so as to limit the difficulty that prohibited structures may entail for learners (Thomson, 2011). Only the stressed vowels were analysed in the CG stimuli.

Table 2. English (adapted from Bohn & Bundgaard-Nielsen, 2008; Lengeris, 2009; Mayr & Escudero, 2010) and CG stimuli (adapted from Papachristou, 2011)

English stimuli			CG stimuli	
Vowel	**bVt**	**gVt**	**Vowel**	**bVtV**
ɪ	bit	git	**i**	*/'bita/
i:	beat	*gheat	**e**	*/'beta/
e	bet	get	**a**	*/'bata/
ɜ:	*burt	*gert	**o**	/'bota/
ɑ:	*bart	*gart	**u**	*/'buta/
æ	bat	gat		
ʌ	but	gut		
ɒ	bot	got		
ɔ:	bought	*gort		
ʊ	butch	good		
u:	boot	*gould		

NE speakers produced the English stimuli and fillers in the carrier phrases "He said…and left" (Recording 1) and "The next word is…" (Recording 2), which were visually presented in random order using PowerPoint. The target words produced by NE speakers in Recording 1 were isolated and imported into the TP software (Teste/Treinamento de Percepção)–Version 3.1 (Rauber et al., 2012) for the preparation of the training sessions and the perceptual tests completed by CG learners. In order to minimise variation as a result of the linguistic background of NE participants and ensure that the vowels were produced in the target variety (SSBE), two NE raters trained in phonetics assessed their quality, eliminating stimuli that were unclear due to recording or voice quality, and stimuli in which the vowel could not be classified as the target SSBE vowel. A total of 35 stimuli were excluded during this process.

2.3.2. Synthetic Stimuli

Since it was previously observed that L2 learners of English tend to inappropriately rely on vowel quantity rather than quality to distinguish the members of contrastive vowels (e.g. Wang & Munro, 2004; Cebrian, 2006; Aliaga-Garcia & Mora, 2009; Lengeris, 2009; Thomson, 2011), synthetic

stimuli were also used in the training sessions, in order to draw trainees' attention away from vowel length and towards vowel quality. A Praat script (Boersma & Weenik, 2013) was used to create a 5-step duration continuum (60ms, 120ms, 180ms, 240ms, 300ms) for each target vowel using the original recordings. All synthetic stimuli were checked for naturalness, and distorted items were excluded.

2.4 Tasks

2.4.1 Perceptual tasks

The TP software (Rauber et al., 2012) was used for the perceptual pre-test and post-test. CG participants completed a forced-choice identification (FCID) task without feedback, including 165 natural stimuli presented in random order. Participants heard the stimuli in isolation through headphones (up to 9 times), and then clicked on the label containing the word they heard on a computer screen, choosing one of 11 options as shown in Figure 1 below. Each stimulus was presented with its corresponding label in advance, and participants were instructed to read all labels before starting the task. Twenty randomly selected items were included at the beginning as a practice session.

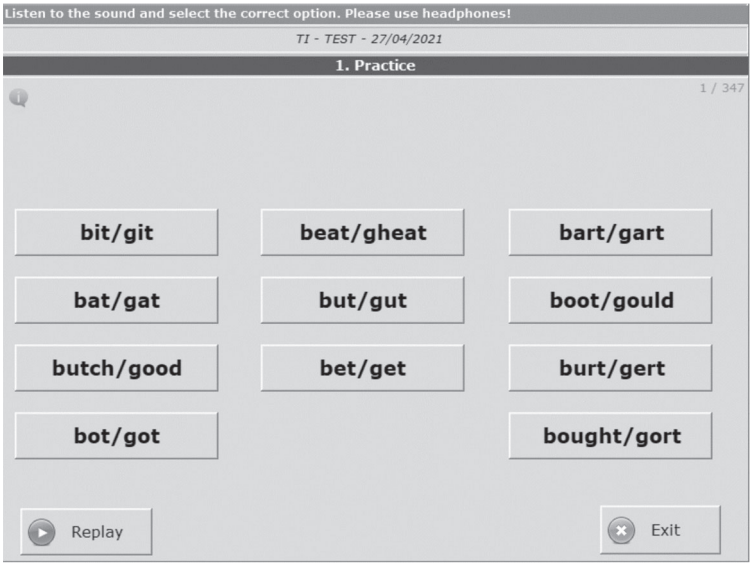


Figure 1. Example of FCID test

2.4.2 Production tasks

CG participants produced all stimuli in an elicited imitation task, in order to eliminate orthographic effects as reported in previous studies (e.g. Giannakopoulou et al., 2013; 2017), given that SMG orthography is transparent whilst English is not. CG participants listened to the target words presented in random order in the carrier phrase "The next word is…" (Recording 2 of NE speakers) via headphones, and repeated them in a new carrier phrase, "Now I say…for you". CG speakers also produced the CG stimuli (Table 1) in a wordlist reading task, in normal speaking rate, in the carrier phrase "Είπεν του…τζι' έφυε" (/ˈipendu ... ˈdʒefie/, "He said to him…and left") presented in random order in PowerPoint. The stimuli for the elicited imitation task in the post-test were randomised again to limit familiarity effects.

Production data were collected using a Zoom H1 audio recorder (sampling rate 44.1 kHz). The data collection process took place at quiet university rooms or labs in the UK and Cyprus.

2.4.3 Perceptual Training

The training consisted of FCID tasks with immediate feedback, which is arguably the most effective paradigm used in HVPT (Lengeris, 2018; Sakai & Moorman, 2018; Thomson, 2018). Participants were not prompted in any way to produce the target sounds during the training.

The training stimuli (Recording 1 of NE participants) were limited to two phonetic contexts so as to avoid confusing and overwhelming the participants, while at the same time maintaining some phonetic variation (Thomson, 2011). Learners listened to the stimuli up to 9 times at a comfortable volume using headphones and clicked on the labelled button containing the word they heard from a set of options. The feedback provided for correct and incorrect responses is shown in Figures 2 and 3, respectively.

Trainees completed eight training sessions over three weeks, on a flexible schedule. A maximum of one session per day was completed and the researcher observed the completion of each session in person, or via online video-calling software in the case of one participant. Each training session contained 220 natural and synthetic stimuli (20 stimuli/vowel) from a single speaker and lasted 17 minutes on average (range: 9-30 min). Although each session included the whole set of target vowels in both contexts, these were divided into 5 training blocks, each focusing on only one vowel contrast (/ɪ-i:/, /e-ɜ:/, /ɑ:-æ-ʌ/, /ɒ-ɔ:/ and /ʊ-u:/), similarly to Wang and Munro (2004).

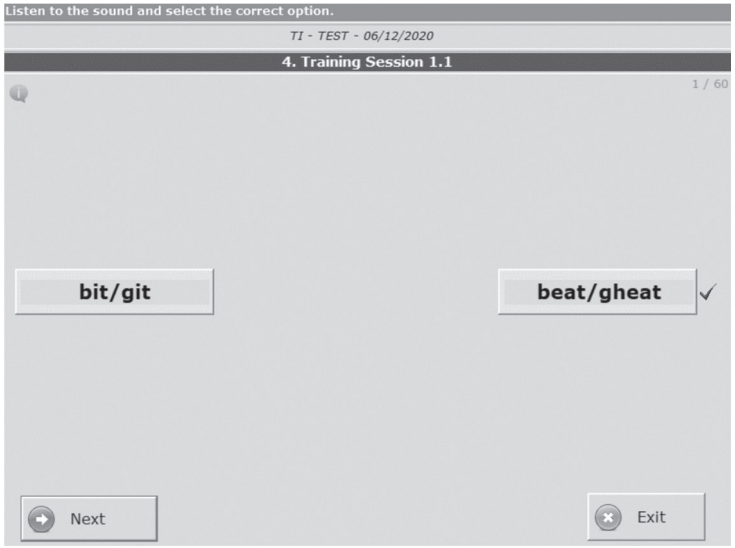


Figure 2. Example of FCID task with feedback (correct response)

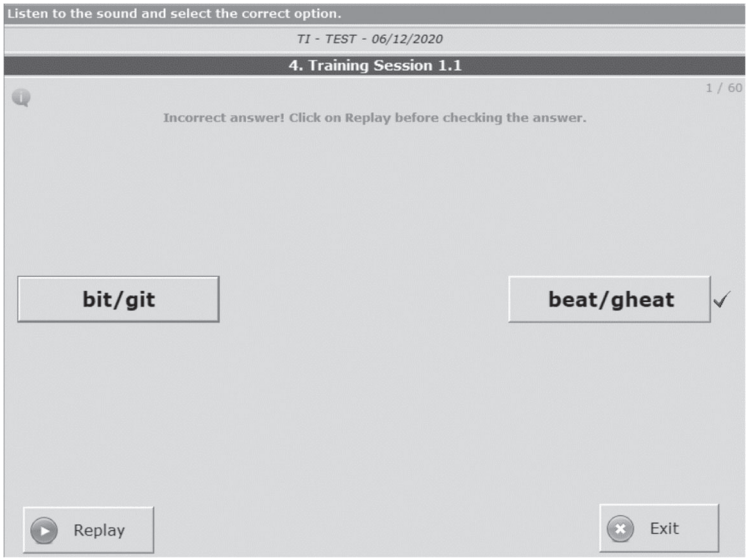


Figure 3. Example of FCID task with feedback (incorrect response). Participants were prompted to listen to the stimulus again and choose the correct response which was revealed by a green tick.

This grouping was determined by the perceptual mapping of L2 English vowels onto CG vowels (Georgiou, 2019), and corresponds to the CG vowels /i, e, a, o, u/, respectively. These blocks were presented in random order in each session, but the order was the same for all participants.

2.5 Analyses

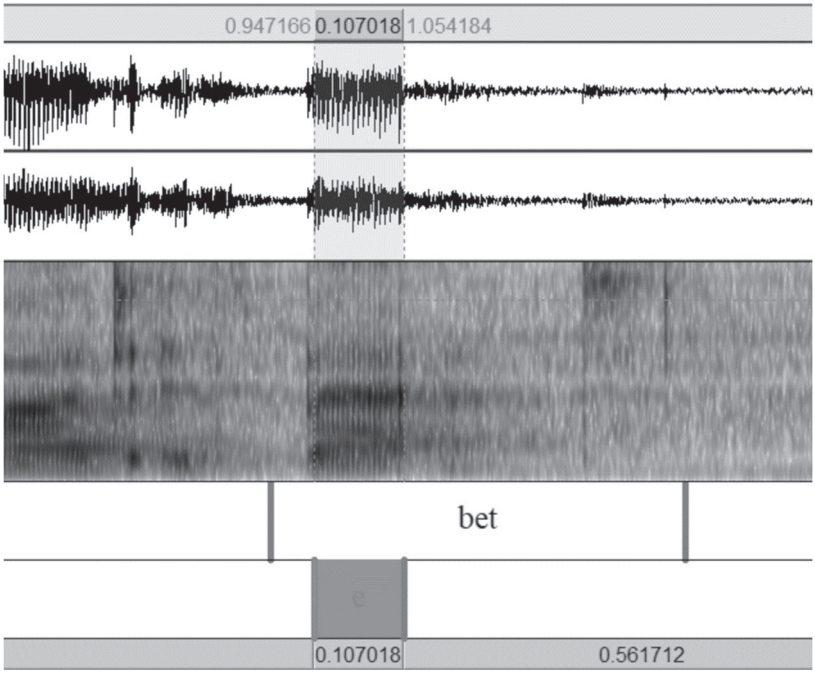


Figure 4. Example of segmentation of target vowels

Perceptual performance was measured through %-correct identification scores. The productions of NE speakers and CG learners were acoustically analysed using Praat (Boersma & Weenik, 2013) to identify similarities and differences. Duration and formant values (F1 and F2 at mid-point) were measured from the onset to the offset of periodic energy in F2 in the spectrogram, which were identified acoustically and based on spectral characteristics (Figure 4). The F1 (high-low dimension) and F2 (front-back dimension) are particularly important as they enable the plotting of vowels in the vowel space. Formant frequency analysis was performed using the Formant (burg) function in Praat, with the standard settings for male

(5000Hz) and female (5500Hz) participants. Formant values were normalised using the Lobanov method (Clopper, 2009; Flynn, 2011). Due to the small number of data collected for each vowel, no statistical analyses were conducted. However, some tendencies can be observed.

3. Results

3.1 Perception

Figure 5 shows the mean %-correct identification score across all target vowels for each participant in the pre-test and post-test, demonstrating a clear improvement in the perceptual performance of all participants.

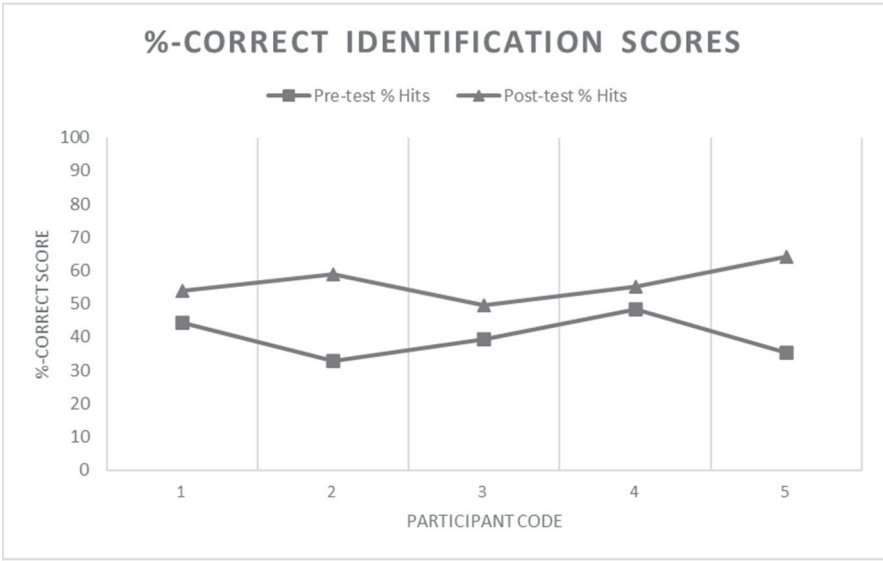


Figure 5. Mean perceptual identification score of each participant at pre- and post-test FCID tasks

Figure 6 illustrates participants' mean perceptual performance in each target vowel at Time 1 (T1) and Time 2 (T2) showing that participants were better able to identify all target vowels, except for /ʌ/, at T2. The largest improvement was observed in the vowels /u:/ and /i:/. The most difficult vowels were /ɜ:/, /ɑ:/ and /ɔ:/ at T1, and /ʌ/, /ɜ:/, and /ɪ/ at T2. After exposure to the training, correct identification increased by 21.25% in /ɜ:/, 21.42% in

/ɑ:/ and 23.75% in /ɔ:/, even though only the latter was correctly identified at above-chance level (>50%) at T2.

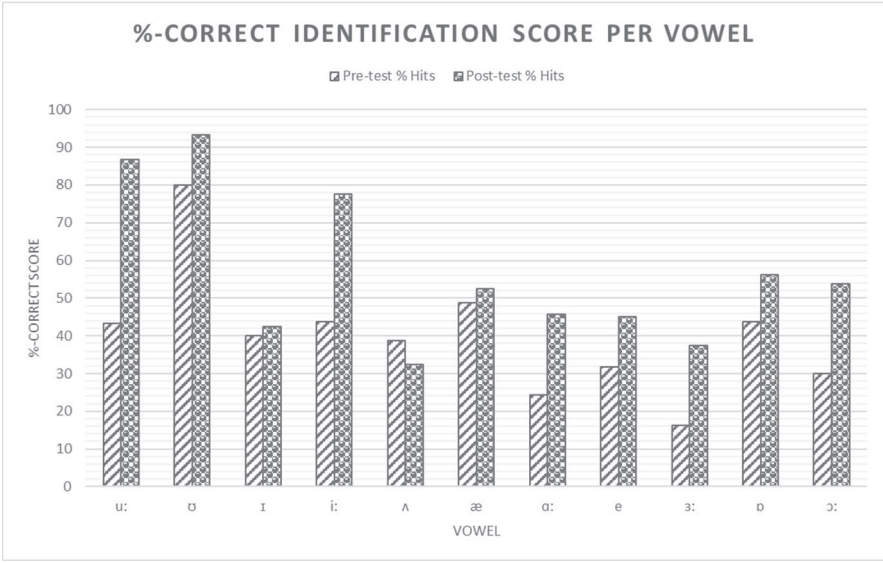


Figure 6. Mean perceptual identification scores for each target vowel at pre- and post-test FCID tasks.

3.2 Production

3.2.1 Duration

Figure 7 shows the duration range of each target vowel produced by the CG learners at T1 and T2, and by the NE speakers. Table 3 shows the mean duration of each vowel produced by CG speakers at T1 and T2, and by the NE speakers, as well as the mean duration of corresponding CG vowels. As can be seen, only /ʊ/ and /æ/ were produced with similar mean durations to NE durations at T1, and only /i:/, /ʊ/ and /æ/ approximate NE durations at T2. Nativelike tendencies were observed for the contrasts /i:/-/ɪ/ and /u:/-/ʊ/, where the long vowel in the pair became longer and the short, shorter, although durations remained different compared to NE. In the case of /u:/ and /ʊ/, the difference is more prominent, as the two vowels were produced with very similar durations at T1, but after the training, they had considerably different durations both between them, and between T1 to T2.

NE mean durations for /e/-/ɜ:/ show that the two vowels are different by almost 100ms on average, which is not the case for the productions of CG learners at either T1 or T2. However, learners' mean duration of the vowels in this contrast decrease for the short vowel and slightly increase for the long vowel at T2, resulting in a further divergence between the mean durations of the two L2 vowels.

Learners did not improve their mean durations for the /ɒ/-/ɔ:/ contrast, which had a short-long distinction of approximately 28.9ms initially but was reduced at T2 as opposed to nativelike patterns. For the /ɑ:/-/æ/-/ʌ/ contrast, it is evident that CG speakers produced the three vowels with a similar duration at T1, as opposed to NE speakers who had three distinct duration means for the three vowels. At T2, CG learners produced the vowels with different mean durations, lengthening /ɑ:/ although not to the extend produced by NE speakers and maintaining /æ/ similar to T1 and to NE speaker duration. The vowel /ʌ/ was produced with longer instead of shorter mean duration at T2.

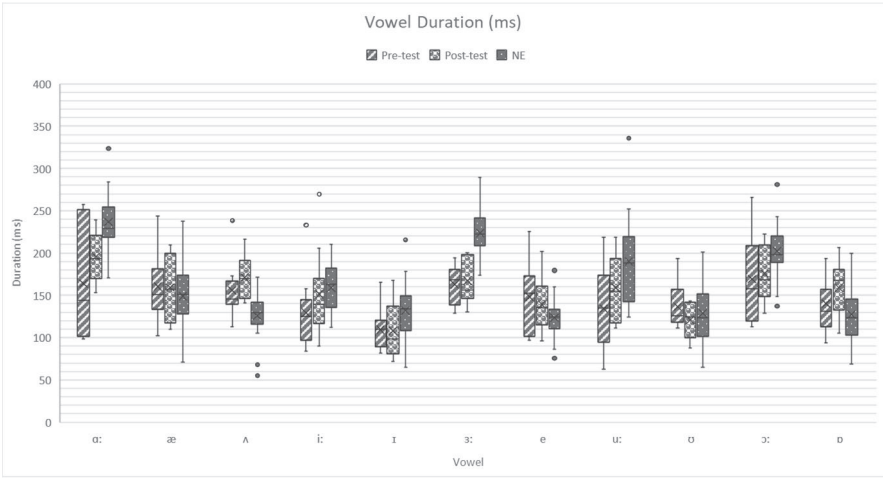


Figure 7. Range of duration of each target vowel produced by CG learners at T1 and T2, and by NE speakers

Table 3. Mean duration (in ms) of each vowel

	Pre-test	Post-test	NE speakers	CG Vowels
ɑ:	163.89 (SD 66.69)	195.94 (SD 29.18)	237.10 (SD 33.49)	171.17 (SD 39.84)
æ	160.56 (SD 45.8)	159.33 (SD 41.42)	149.09 (SD 39.1)	
ʌ	155.88 (SD 37.26)	171.3 (SD 25.3)	126.09 (SD 29.71)	
i:	129.37 (SD 42.97)	151.59 (SD 51.94)	159.16 (SD 28.63)	116.66 (SD 20.67)
ɪ	109.4 (SD 24.28)	108.06 (SD 34.13)	132.19 (SD 33.74)	
e	148.52 (SD 42.62)	139.17 (SD 31.91)	123.79 (SD 27.01)	140.13 (SD 29.83)
ɜ:	162.94 (SD 23.68)	167.24 (SD 28.29)	223.72 (SD 28.54)	
ʊ	136.73 (SD 27.54)	120.97 (SD 21.28)	129.14 (SD 37.07)	123.89 (SD 26.45)
u:	134.43 (SD 50.85)	159.11 (SD 39.19)	190.05 (SD 56.55)	
ɒ	140.31 (SD 42.01)	158.86 (SD 30.71)	127.39 (SD 31.22)	156.06 (SD 44.53)
ɔ:	169.18 (SD 53.62)	175.95 (SD 32.26)	202.25 (SD 32.46)	

3.2.2 Formant Values

The target vowels produced by NE speakers and CG learners were plotted in the vowel space to visualise their position and enable comparisons. Figure 8 below shows the mean, normalised F1xF2 values for each target vowel produced by CG learners at T1 and T2, the vowels produced by NE speakers, and the CG vowels.

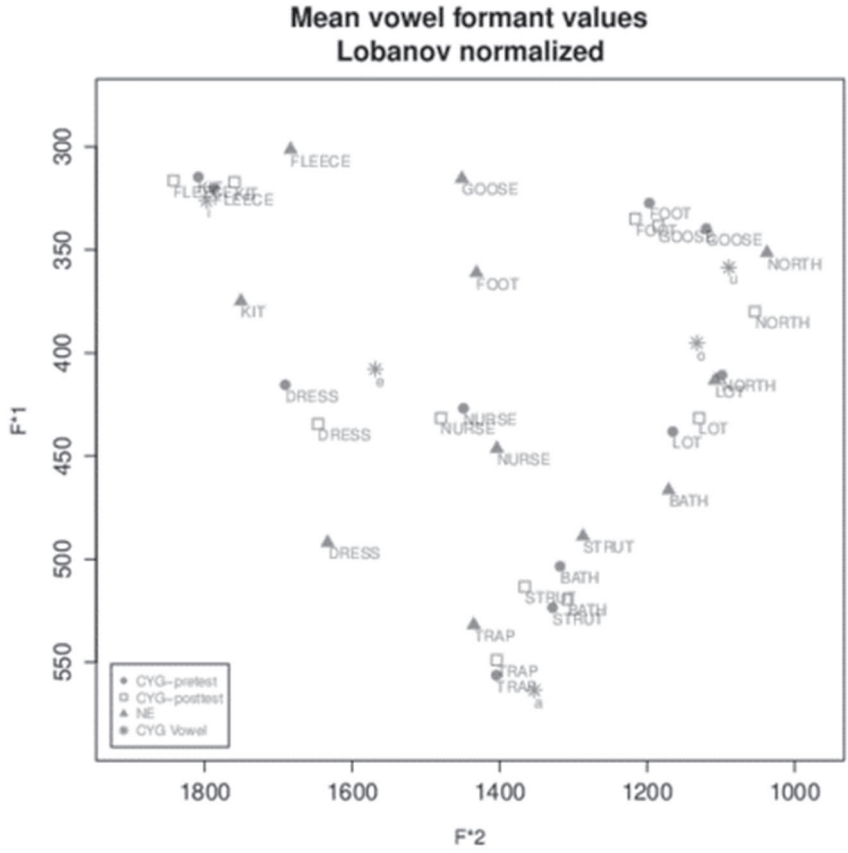


Figure 8. F1xF2 plot of mean formant values (Lobanov-normalised) of English vowels at pre-test (red), post-test (blue), and as produced by NE speakers (green), and of CG vowels (yellow)

In order to examine the effects of the training, the Euclidean distance (ED) in Hz between a) the two vowels in an L2 contrast as produced by the CG learners at T1 and T2, and b) the CG and NE productions at pre-test and post-test were calculated using IBM Corp. SPSS (Version 27.0) and are presented in Tables 4 and 5, respectively. Mean Lobanov-normalised F1 and F2 values were used to calculate ED.

Table 4. ED (in Hz) between the vowels of each L2 contrast produced by the CG learners at T1 and T2

Vowel Contrast	Pre-test (T1)	Post-test (T2)
/ʊ/-/uː/	78.1	31.1
/ɒ/-/ɔː/	72.2	91.3
/e/-/ɜː/	242.3	166
/i̞ː/-/ɪ/	21.6	82
/ɑː/-/æ/	100.5	100.6
/ɑː/-/ʌ/	22.4	58.3
/æ/-/ʌ/	82.5	52.3

Table 4 shows that the vowels in the contrasts /ɒ/-/ɔː/, /iː/-/ɪ/ and /ɑː/-/ʌ/ become more distant at T2, indicating that the trainees were better able to produce the two vowels in each of these contrasts with different tongue configurations after the training. On the other hand, the vowels in the contrasts /ʊ/-/uː/, /e/-/ɜː/ and /æ/-/ʌ/ become more merged at T2, while the distance between the vowels /ɑː/ and /æ/ remains very similar at T1 and T2.

In order to assess the direction of the changes, i.e., whether learners' productions approximate or diverge from NE productions, the ED between NE speakers' and CG learners' productions of each target vowel at T1 and T2 are given in Table 5.

Table 5. ED (in Hz) between vowels produced by CG learners and vowels produced by NE speakers at T1 and T2

Target Vowel	Pre-test (T1)	Post-test (T2)
/ʊ/	236.5	216.6
/u:/	331.9	266.9
/ɒ/	62.2	28.3
/ɔ:/	85.6	33.6
/e/	95.6	59.4
/ɜ:/	48.8	77.3
/i̯:/	106.7	159.7
/ɪ/	83.5	58.7
/ɑ:/	151.6	146.5
/æ/	39.2	35.4
/ʌ/	53.9	82.6

A comparison of the EDs in Table 5 indicates that there is a decrease in the distance between CG productions and NE productions at T2 compared to T1 for most target vowels (/ʊ/, /u:/, /ɒ/, /ɔ:/, /e/, /ɪ/, /ɑ:/ and /æ/). The remaining vowels (/ɜ:/, /i:/ and /ʌ/) move further away from NE values. The most nativelike productions at T1 were found in /æ/, /ɜ:/ and /ʌ/, and at T2 in /ɒ/, /ɔ:/ and /æ/. The vowels /ʊ/, /u:/, /ɑ:/ and /i:/ were the most distant to NE productions both at T1 and T2.

Figure 9 shows the F1xF2 position of CG vowels in the vowel space in relation to L2 English vowel productions at T1 and T2, and the vowel space for English vowels as produced by the NE speakers. The figure clearly demonstrates that while there is a differentiation between contrastive vowels in NE speakers' productions, CG learners produce these vowels with considerable overlap corresponding to the positions for the CG vowels. The most noteworthy observations are in relation to the /e/-/ɜ:/ contrast, where the two vowels were differentiated at T1 but had some overlap at T2, and to the /ɒ/-/ɔ:/ contrast, the vowels of which had considerably more overlap at T1 compared to T2.

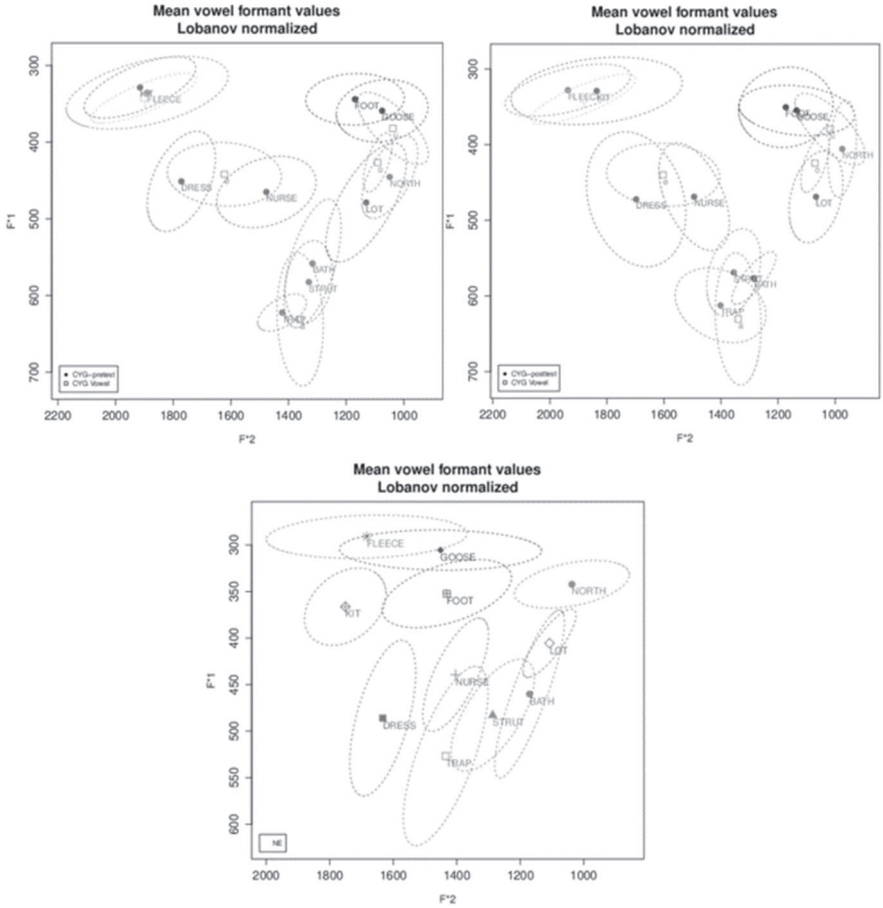


Figure 9. F1xF2 position of CG vowels in relation to L2 vowel productions at T1 (top-left) and T2 (top-right), and vowel space for English vowels produced by NE speakers (bottom)

4. Discussion

This pilot study aimed to address two RQs: whether HVPT can help adult CG learners improve their perception of contrastive L2 English vowels, and how effective it can be in improving CG learner's productions of L2 English vowels.

To address RQ 1, our results support that the HVPT paradigm used was effective in improving perceptual performance, similarly to previous studies on vowels (e.g. Wang & Munro, 2004; Nishi & Kewley-Port, 2007; 2008; Iverson & Evans, 2009; Iverson et al., 2012; Carlet & Cebrian, 2014), although with great variation in learners' performance for different target vowels. The largest improvement was seen in the vowels /uː/ and /iː/, while the vowels /ɜː/, /ɑː/ and /ɔː/ were the most difficult for these learners at T1, and /ʌ/, /ɜː/, and /ɪ/ at T2. Interestingly, the stimuli for the vowels /ɜː/, /ɑː/ and /ɔː/ include an elided /ɹ/, which could be a source of perceptual difficulty for these learners that seems to be overcome after the training. Distinguishing between the vowels in the contrasts /e/-/ɜː/ and /ɑː/-/æ/-/ʌ/ was evidently the most difficult task for learners.

These results are, to some extent, consistent with Lengeris (2009) and Georgiou (2019), who examined the perceptual assimilation patterns of SMG and CG learners, respectively. Based on their findings, it was expected that these learners would have difficulties in perceiving the vowels in the contrasts /ɑː/-/æ/-/ʌ/ and /ɒ/-/ɔ/:, but not /e/-/ɜː/. This inconsistency might be due to the difference in tasks or groups of participants in the present study compared to previous studies; this study examined CG adult learners, whereas Lengeris (2009) and Georgiou (2019) examined SMG adult learners and CG children, respectively. The small sample size may have also affected the results of the present study.

In terms of duration, production performance reflects perceptual performance: nativelike tendencies were observed for the contrasts /iː/-/ɪ/ and /uː/-/ʊ/, although durations remained different compared to NE values. In addition, the most difficult contrastive vowels for these learners in terms of duration were /ɒ/-/ɔː/ and /ɑː/-/æ/-/ʌ/, which also posed difficulties for learners in perception.

This alignment of perceptual and durational results, however, is not extended to spectral characteristics in the productions of learners. The vowels /ʊ/, /uː/, /ɑː/ and /iː/ were the most distant to NE productions, while the most nativelike productions at T2 were achieved in /ɒ/, /ɔː/ and /æ/, which were more difficult for learners in perception and duration. This is also inconsistent with our predictions based on the perceptual assimilation patterns reported in Georgiou (2019), according to which CG learners would have the most difficulties in producing the vowels /æ/-/ʌ/ and /ɒ/-/ɔː/. In addition, while the vowels /e/ and /ɜː/ occupied different positions in the vowel space at T1, they overlapped after the training, which was unexpected. In combination with the difficulties that this contrast posed to

learners in perception and duration, it seems worthwhile to examine it in more depth.

The alignment between perceptual and durational performance as opposed to spectral characteristics, in combination with the fact that the training stimuli varied in duration meaning that learners could not rely on durational cues in perception, indicates that learners may have used durational cues in their production as a means to distinguish between newly perceived differences in L2 vowels. This may be because, according to Bohn (1995), duration is a language-independent, salient cue that is easily accessible. Previous research has shown that L2 learners are sensitive to durational cues, even when their L1 does not use them contrastively (e.g. Cebrian, 2006; Lengeris, 2009; Lengeris & Hazan, 2010). For example, whereas English speakers rely primarily on spectral cues in identifying their L1 vowels, Catalan and Spanish learners were previously found to over-rely on durational cues in producing L2 English vowels, even though vowel duration is not used contrastively in their L1 (Cebrian, 2006; Aliaga-Garcia & Mora, 2009).

Overall, CG learners produced the L2 vowels without the required durational or spectral differences, merging them in a single, L1-based category, both at T1 and T2, as anticipated. However, the training was at least partly effective, in that it led to changes in all vowels in respect of either duration, spectral characteristics or both, although the statistical significance of these changes cannot be ascertained due to the small sample size. While learners modified their productions of L2 vowels at T2, they failed to reach nativelike productions in both durational and spectral characteristics. This indicates that learners became aware of the differences between contrastive L2 vowels, although the training was not adequate to enable them to develop new categories for L2 vowels. The lack of improvement in spectral characteristics may also mean that learners need additional time to update their articulatory habits to match perception.

The results of this study are nevertheless promising, as they suggest that better results can be achieved if future studies include more intensive and enriched training paradigms, providing exposure to a greater variety of stimuli. Given that a positive effect of the perceptual training on both perception and production gains was previously observed for learners of different L1 backgrounds (e.g. Bradlow et al., 1997; Lengeris, 2009; 2018; Lengeris & Hazan, 2010; Thomson, 2011), this study offers additional insights in the domain of L2 learning and teaching, as it demonstrates that time spent training in one modality can be beneficial for both modalities.

5. Limitations

This pilot study is part of a larger project and therefore suffers from several limitations. Firstly, the small sample size and limited number of data collected did not allow us to conduct analyses that would carry statistical power. Future studies with larger data sets can examine the effects of TIME, VOWEL and the TIME*VOWEL interaction on duration, F1 and F2 in a mixed model analysis, to reach more robust conclusions. The validity of the study could also be improved by the inclusion of a control group, to ascertain whether any improvements are directly related to the intervention.

Another limitation of this study lies in the stimuli used, which were chosen so that the target vowels appear in matching environments (Di Paolo et al., 2011) and that the attention of the trainees is directed only on phonetic form and not meaning (Thomson, 2011). This means that there was no control over word frequency, which can be examined in future studies. In addition, the stimuli in the /ʊ/-/u:/ contrast enabled participants to rely on phonetic environment to identify the target vowel, as indicated by the near-ceiling performance of participants in these vowels. Importantly, production results suggest that the most improved duration occurred in the vowels in this contrast, despite the cues that facilitated their identification in perception tests. Therefore, despite its effect on perceptual performance, this limitation does not seem to affect participants' ability to develop their sensitivity to this secondary cue. However, future studies may need to consider modifying or excluding such stimuli.

Finally, although it was not the aim of this study to examine the benefits of HVPT in the generalisation of knowledge to new speakers and contexts and its long-term retention, as argued in previous studies (e.g. Lively et al., 1993; Thomson, 2011; Carlet & Cebrian, 2014; Qian et al., 2018), it would be beneficial to investigate them in future studies, to gain an in-depth knowledge on the effectiveness of HVPT on these learners.

References

Aliaga-Garcia, C., & Mora, J. C. (2009). Assessing the effects of phonetic training on L2 sound perception and production. In M. A. Watkins, A. S. Rauber, & B. O. Baptista (Eds.), *Recent Research in Second Language Phonetics/Phonology: Perception and Production* (pp. 2-31). Newcastle upon Tyne: Cambridge Scholars.

Arvaniti, A. (2006). Linguistic practices in Cyprus and the emergence of Cypriot Standard Greek. *San Diego Linguistic Papers*, 2, 1-24.

Arvaniti, A. (2010). A (brief) review of Cypriot phonetics and phonology. In A. Voskos, D. Goutsos, & A. Moser (Eds.), *The Greek Language in Cyprus from Antiquity to Today* (pp. 107-124). Athens: University of Athens.

Best, C. T. (1995). A direct realist view of cross-language speech perception: New directions in research and theory. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues* (pp. 171-204). Baltimore: York Press.

Best, C. T., & Tyler, M. (2007). Non-native and second-language speech perception: Commonalities and complementarities. In O.-S. Bohn, & M. J. Munro (Eds.), *Second Language Speech Learning: In Honor of James Emil Flege* (pp. 13-34). Amsterdam/Philadelphia: John Benjamins.

Boersma, P., & Weenink, D. (2013). Praat: doing phonetics by computer (Version 6.0.52) [Computer software]. Retrieved from http://www.praat.org/

Bohn, O.-S. (1995). Cross-language speech perception in adults: First language transfer doesn't tell it all. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-language Research* (pp. 279-304). Baltimore: York Press.

Bohn, O.-S., & Bundgaard-Nielsen, R. L. (2008). Second language speech learning with diverse inputs. In T. Piske, & M. Young-Scholten (Eds.), *Input Matters in SLA* (pp. 207-218). Bristol: Multilingual Matters.

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. I. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *The Journal of the Acoustical Society of America*, 101(4), 2299-2310.

Carlet, A., & Cebrian, J. (2014). Training Catalan speakers to identify L2 consonants and vowels: A short-term high variability training study. *Proceedings of the International Symposium on the Acquisition of Second Language Speech*, 5, 85-98.

Cebrian, J. (2006). Experience and the use of duration in the categorization of L2 vowels. *Journal of Phonetics*, 34, 372-387.

Clopper, C. G. (2009). Computational methods for normalizing acoustic vowel data for talker differences. *Language and Linguistics Compass*, 3(6), 1430-1442.

Derwing, T. M., & Munro, M. J. (2005). Second language accent and pronunciation teaching: A research-based approach. *TESOL Quarterly*, 39(3), 379-397.

Derwing, T. M., Munro, M. J., Foote, J. A., Waugh, E., & Fleming, J. (2014). Opening the window on comprehensible pronunciation after 19 years: A workplace training study. *Language Learning*, 64(3), 526-548.

Deterding, D. (1997). The formants of monophthong vowels in Standard Southern British English pronunciation. *Journal of the International Phonetic Association*, 27(1-2), 47-55.

Di Paolo, M., Yaeger-Dror, M., & Wassink, A. B. (2011). Analyzing vowels. In M. Di Paolo & M. Yaeger-Dror (Eds.), *Sociophonetics: A Student's Guide* (pp. 87-106). London: Routledge.

Dimitriou, D. (2018). L2 acquisition and production of the English rhotic by L1 Greek-Cypriot speakers: The effect of L1 articulatory routines and phonetic context. *Philologia*, 16(1), 45-64.

Escudero, P. (2005). *Linguistic Perception and Second Language Acquisition: Explaining the Attainment of Optimal Phonological Categorization* (Doctoral dissertation). Retrieved from http://www.fon.hum.uva.nl/archive/2005/PaolaEscuderoPhD2005.pdf

Flege, J. E. (1995). Second language speech learning: theory, findings and problems. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues* (pp. 233-277). Baltimore: York Press.

Flege, J. E. (1999). The relation between L2 production and perception. In J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, & A. Bailey (Eds.), *Proceedings of the 14th International Congress of Phonetics Sciences: Vol. 2* (pp. 1273-1276). University of California, Berkeley.

Flege, J. E. (2002). Interactions between the native and second-language phonetic systems. In P. Burmeister (Ed.), *An Integrated View of Language Development: Papers in Honor of Henning Wode* (pp. 217-243). Trier: Wissenschaftlicher Verlag.

Flege, J. E. (2003). Assessing constraints on second-language segmental production and perception. In A. Meyer, & N. Schiller (Eds.), *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities* (pp. 319-355). Berlin: Mouton de Gruyter.

Flege, J. E. (2009). Give input a chance!. In T. Piske, & M. Young-Scholten (Eds.), *Input Matters in SLA* (pp. 175-190). Bristol: Multilingual Matters.

Flege, J. E., & Liu, S. (2001). The effect of experience on adults' acquisition of a second language. *Studies in Second Language Acquisition*, 23(4), 527-552.

Flege, J. E., & MacKay, I. R. A. (2004). Perceiving vowels in a second language. *Studies in Second Language Acquisition*, 26(1), 1-34.

Flynn, N. (2011). Comparing vowel formant normalisation procedures. *York Papers in Linguistics*, 2(11), 1-28.

Georgiou, G. P. (2019). Bit and beat are heard as the same: Mapping the vowel perceptual patterns of Greek-English bilingual children. *Language Sciences*, 72, 1-12.

Giannakopoulou A., Uther M., & Ylinen S. (2013). Enhanced plasticity in spoken language acquisition for child learners: evidence from phonetic training studies in child and adult learners of English. *Child Language Teaching and Therapy*, 29(2), 201-218.

Giannakopoulou, A., Brown, H., Clayards, M., & Wonnacott, E. (2017). High or low? Comparing high and low-variability phonetic training in adult and child second language learners. *PeerJ*, 5, e3209.

IBM Corp. (Released 2020). IBM SPSS Statistics for Windows (Version 27.0). Armonk, NY: IBM Corp

Iverson, P., & Evans, B. G. (2009). Learning English vowels with different first-language vowel systems II: Auditory training for native Spanish and German speakers. *The Journal of the Acoustical Society of America*, 126(2), 866-877.

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y. I., Kettermann, A., & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, 87(1), B47-B57.

Iverson, P., Pinet, M., & Evans, B. G. (2012). Auditory training for experienced and inexperienced second-language learners: Native French speakers learning English vowels. *Applied Psycholinguistics*, 33(1), 145-160.

Kkese, E., & Petinou, K. (2017). Perception abilities of L1 Cypriot Greek listeners – Types of errors involving plosive consonants in L2 English. *Journal of Psycholinguistic Research*, 46(1), 1-25.

Krzonowski, J., Ferragne, E., & Pellegrino, F. (2015). Effects of perception and production trainings on the production of English vowels by French native learners. In A. Botinis (Ed.), *Proceedings of the 6th Tutorial and Research Workshop on Experimental Linguistics* (pp. 37-40). Athens, Greece.

Kyprianou, M. (2007). Teaching pronunciation in Cyprus. *Proceedings of the Phonetics Teaching and Learning Conference 2007*. UCL, London, UK.

Lengeris, A. (2009). *Individual Differences in Second-language Vowel Learning* (Doctoral dissertation). Retrieved from https://discovery.ucl.ac.uk/id/eprint/19029/1/19029.pdf

Lengeris, A. (2018). Computer-based auditory training improves second-language vowel production in spontaneous speech. *The Journal of the Acoustical Society of America*, 144(3), EL165-EL171.

Lengeris, A., & Hazan, V. (2010). The effect of native vowel processing ability and frequency discrimination acuity on the phonetic training of English vowels for native speakers of Greek. *The Journal of the Acoustical Society of America*, 128(6), 3757-3768.

Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *The Journal of the Acoustical Society of America*, 94(3), 1242-1255.

Mayr, R., & Escudero, P. (2010). Explaining individual variation in L2 perception: Rounded vowels in English learners of German. *Bilingualism: Language and Cognition*, 13(3), 279-297.

Nishi, K., & Kewley-Port, D. (2007). Training Japanese listeners to perceive American English vowels: Influence of training sets. *Journal of Speech, Language, and Hearing Research*, 50(6), 1496-1509.

Nishi, K., & Kewley-Port, D. (2008). Nonnative speech perception training using vowel subsets: Effects of vowels in sets and order of training. *Journal of Speech, Language, and Hearing Research*, 51(6), 1480-1493.

Papachristou, V. (2011). Explicit vs. implicit Pronunciation Teaching to Greek children: The case of the acquisition of English vowels. *Selected Papers on Theoretical and Applied Linguistics*, 19, 371-381.

Piske, T., MacKay, I. R. A., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics*, 29(2), 191-215.

Pruitt, J. S., Jenkins, J. J., & Strange, W. (2006). Training the perception of Hindi dental and retroflex stops by native speakers of American English and Japanese. *The Journal of the Acoustical Society of America*, 119(3), 1684-1696.

Qian, M., Chukharev-Hudilainen, E., & Levis, J. (2018). A system for adaptive high-variability segmental perceptual training: Implementation, effectiveness, transfer. *Language Learning and Technology*, 22(1), 69-96.

Rauber, A., Rato, A., Kluge, D., & Santos, G. (2012). TP (Perception Testing/Training) (Version 3.1) [Computer software]. Retrieved from http://www.worken.com.br/tp/tp_install.html

Roach, P. (2004). British English: Received Pronunciation. *Journal of the International Phonetic Association*, 34(2), 239-245.

Sakai, M., & Moorman, C. (2018). Can perception training improve the production of second language phonemes? A meta-analytic review of 25 years of perception training research. *Applied Psycholinguistics*, 39(1), 187-224.

Simaki, V., Mporas, I., Kondyli, M., & Megalooikonomou, V. (2015). Acoustic differences between Standard Greek and Cypriot Greek. In A. Ronzhin, T. Potapova, & N. Fakotakis (Eds.), *Booklet of Papers from the 17th International Conference of Speech and Computer (SPECOM 2015)* (pp. 71-78). Athens, Greece.

Thomson, R. I. (2011). Computer assisted pronunciation training: Targeting second language vowel perception improves pronunciation. *CALICO Journal*, 28(3), 744-765.

Thomson, R. I. (2018). High variability [pronunciation] training (HVPT): A proven technique about which every language teacher and learner ought to know. *Journal of Second Language Pronunciation*, 4(2), 208-231.

Wang, X., & Munro, M. J. (2004). Computer-based training for learning English vowel contrasts. *System*, 32(4), 539-552.

CHAPTER EIGHT

EFFECT OF TALKER GENDER INFORMATION ON SPEECH PERCEPTION IN L2 ENGLISH

ELENA KKESE AND SVIATLANA KARPAVA

Abstract

As adult listeners, perceiving speech and understanding the message(s) conveyed is an everyday skill, which requires little effort in typically developing population. Nonetheless, speech perception is a demanding perceptual task due to acoustic variability. Listeners need to apprehend several spatial and indexical factors conveyed by the acoustic speech signal besides the linguistic message of the talker. This task becomes even more demanding in the case of L2 perception since factors related to L2 proficiency, frequency, and context of use may be involved. Nonetheless, most of the previous work focused on the effect of acoustic cues while fewer studies have concentrated on the effect of contextual information. The present study examines how talker gender could influence L2 speech processing, like phoneme categorisation of English consonants and vowels by Cypriot-Greek speakers. The participants were 70 Cypriot-Greek university students with sufficient knowledge of English as an L2. They completed a word dictation task in which they heard a word and were asked to respond based on what word they heard by writing the word in a given scoresheet. The task was designed on word level as the aim was to eliminate any semantic or syntactic information from the input while the target words were pronounced by a male and a female talker. The findings indicated that participants performed significantly better regarding vowels when the talker was female but this changed when the talker was male with performance being significantly better concerning consonants. Together, the data suggest that talker gender information may affect L2 speech perception.

Keywords: speech perception; phonetic information; contextual information; talker gender; VOT; vowel formants

1. Introduction

Speech perception can be defined as the process by which speech acoustic cues are mapped onto linguistic messages such as phonemes, syllables, words, and phrases (Kkese, 2016). Acoustic cues (VOT, formant frequencies, spectral mean) are responsible for distinguishing phonemes, which suggests that listeners have to encode continuous acoustic cues, map the acoustic cues to phoneme categories, and access lexical entries to be able to perceive speech. Nonetheless, speech perception constitutes a challenging categorisation problem since the acoustic cues are affected by several factors while input tends to be variable and ambiguous. Factors such as speaking rate (Kessinger & Blumstein, 1997), neighbouring phonetic material (Nearey & Rochet, 1994; Repp, 1982), and talker identity (Allen et al., 2003) are related to contextual information. Listeners have to integrate information provided through acoustic cues and context in order to perceive speech. What is surprising, though, is that most listeners manage to perceive speech easily and accurately irrespectively of the context and/or the talker. Listeners seem to use a predictive coding (Gagnepain et al., 2012) that allows them to guess what it will be heard and perceive speech based on these predictions.

Turning to acoustic cues, these refer to a continuous acoustic dimension that leads to the differentiation of two phonemes. Distinguishing phonemes depends on multiple acoustic cues, which arrive over time. As a result, acoustic cues provide direct information for distinguishing phonemes or words. An acoustic cue like Voice Onset Time (VOTs) is crucial for categorising a plosive consonant as voiced **/b d g/** or voiceless **/p t k/**. VOT refers to the time between the release of the articulators and the vibrations of the vocal folds (onset of voicing). Voiced plosives have shorter VOTs (0-15ms) compared to voiceless plosives (40-60 ms) (Lisker & Abramson, 1964). Nonetheless, a clear boundary between the two categories does not seem to exist as there is considerable variability in the VOTs of voiced and voiceless plosive consonants. This suggests that a 20msec VOT could indicate a **/b/** or **/p/** in certain cases. On the other hand, the acoustic cues for vowels are formant frequencies (Peterson & Barney, 1952). These refer to the resonant frequencies of the vocal tract and are responsible for which vowel a listener can hear as well as for the differences in quality among different periodic sounds (Williams, 2013).

Contextual information also facilitates the interpretation of acoustic cues but does not distinguish between two phonemes. Several factors can serve as contextual information for different phonetic contrasts. These may

involve speaking rate (Summerfield, 1981; Toscano & McMurray, 2015), neighbouring phonemes (Mann, 1980), acoustic or visual information about the talker (Johnson et al., 1999; McMurray & Jongman, 2011), or acoustic information expressing emotion (Murray & Arnott, 1993). Contextual information could be used by listeners while trying to interpret ambiguous cues and to account for variability in speech. Findings from numerous previous studies have suggested that listeners use talker information to influence categorisation of phonemes (Johnson et al., 1999; Repp & Mann, 1980; Strand, 1999; Strand & Johnson, 1996).

The effect of talker gender constitutes an example of how contextual information could affect acoustic cues. Differences between male and female talkers' voices cause considerable variation in acoustic cues; this variation is the result of differences in the size and shape of the articulators (Hillenbrand et al., 1995; Peterson & Barney, 1952) as well as sociophonetic reasons (Strand & Johnson, 1996). Two acoustic cues that vary with talker gender information refer to formant frequencies for vowels and VOT for consonants. Several studies have explored the effects of the frequencies of the first two formants (Cole et al., 2020; Johnson et al., 1999; Peterson & Barney, 1952; Strand & Johnson, 1996). By analysing vowel perception and specifically the formant frequencies of vowels, Peterson and Barney (1952) suggested that there is great variability between vowel categories in terms of frequencies of the first two formants as well as differences in the size and shape of the vowel space used by genders. Specifically, formant frequencies for vowels spoken by women were higher compared to men indicating differences in the size of the vocal tract for men and women. Hillenbrand et al. (1995) in a reduplication of Peterson and Barney's study found similar results. The only difference involved the vowels **/æ/** and **/ɛ/** for which Peterson and Barney found that **/æ/** had lower first and second formant values than **/ɛ/** but Hillenbrand et al. found a lower second formant value for **/æ/** and a similar first formant value for **/ɛ/**. It was argued, though, that the differences in the two studies could be attributed to dialectal differences between vowels in Hillenbrand et al.'s study. Cole et al.'s study (2010) enabled a further investigation of formant frequencies of vowels indicating that talker gender information could account for approximately 80% of the variance in the first formant and 40% of the second formant when examining the vowels **/ʌ/** and **/ɛ/**. Strand and Johnson (1996) stressed the importance of sociophonetic reasons when it comes to variation in acoustic cues.

According to the researchers, phonetic differences such as spectral means may be due to sociophonetic variability. Women tend to have higher

spectral means compared to men for frication sounds; this is responsible for distinguishing fricative place of articulation in sibilant fricatives. Using spectral mean as a cue to fricative place of articulation implies that listeners need to compensate for this variability. Phonetic differences were also the focus of Johnson et al.'s (1999) study; by varying a phonetic cue along a continuous dimension, the researchers examined whether listeners could shift category boundaries along that dimension as a function of context. In this study, vowel quality was investigated along a first formant continuum from *hood* to *hud* because of talker gender information. The results suggest that a shift in category boundaries was feasible since men had a lower boundary and, as a result, made more *hud* responses. The abovementioned studies indicate that both acoustic cues and phonological features may be affected by talker gender information; nonetheless, listeners are able to compensate with variability found in talkers' voices.

VOT values could also vary because of talker gender information. A study conducted by Swartz (1992) focused on the production of voiced and voiceless plosive consonants suggesting that female speakers generally produce longer VOTs compared to male speakers. Applying this to plosive consonants suggests that a VOT of 30 msec could be a **/b/** when pronounced by a female but a **/p/** when a male speaker would pronounce it. Studies by Ryalls et al. (1997) as well as Whiteside and Irving (1998) found similar results. However, Morris et al. (2008) examining talkers producing syllables in isolation suggested no effect. Toscano (2011) indicated that listeners' VOT category boundaries change depending on whether a word was pronounced by a male or female talker. Listeners could categorise a consonant as voiceless if it was spoken by a male talker. Using talker information may help listeners disambiguate the voiced-voiceless plosive consonants more effectively rather than relying solely on the VOT cue. Listeners may use a variety of factors to predict the talker's voice such as the preceding sentence context (Nearey, 1989), cues in the same syllable (Holt et al., 2000), or expectations set up by visual information (Johnson et al., 1999). Non-linguistic sounds can further affect speech perception mirroring the effects of talker (Laing et al., 2012). This implies a possible early auditory basis for categorisation; the auditory system needs to handle variability in the speech signal while it integrates acoustic information over time.

2. Compensating with talker gender information

Contextual information could influence speech perception and, as a result, categorisation of phonemes leading to contextual variability. However, early speech perception research neglected the importance of context proposing that there is sufficient invariance in the acoustic speech signal to enable distinguishing between phonemes depending solely on acoustic cues (Blumstein & Stevens, 1979; Stevens & Blumstein, 1978). As it has been indicated, though, invariant cues are not present for all phonological distinctions (Lindblom, 1996; McMurray & Jongman, 2011; Ohala, 1996), even though there are some exceptions (i.e., spectral shape as a cue for place of articulation). With reference to talker gender information, this seems to be contextual for phoneme categorisation, but it is also important for identifying the talker.

The exemplar theory or episodic memory (Palmeri et al., 1993) tries to explain how contextual information is used during speech perception. According to the exemplar theory, what really matters are the exemplars in memory, which consist of both contextual and acoustic cue information within the same memory code. Speech perception is more than matching cues to boundaries since it refers to a more holistic matching problem. As a result, contextual information could be used as a direct cue to phoneme categorisation. For instance, categorising **/p/** and **/b/** does not happen due to VOT but it depends on how well **/b/** matches other **/b/**s in memory. Talker gender information can be used to distinguish between phonemes; therefore, when listeners hear a female voice, they tend to interpret this as **/b/**. This means that talker gender information would be encoded along the acoustic cues.

Auditory contrast (Assgari & Stilp, 2015; Holt et al., 2000; Kluender et al., 2003) constitutes another attempt to account for how contextual information is used by the auditory system. Specifically, incoming speech is encoded in relation to prior acoustic speech signal. The auditory system perceptually increases the difference between incoming speech and prior acoustic speech signal. For instance, listeners categorise a vowel as having a higher first formant frequency when the range of the first formant frequency in a preceding sentence is shifted down (Assgari & Stilp, 2015). Likewise, listeners categorise a vowel as having a lower first formant frequency when the range of the first formant frequency in a preceding sentence is shifted up. Concerning talker gender information, Johnson et al.'s (1999) study focusing on how visual information affects phoneme perception indicated that the listeners' responses were based on which face was seen. Listeners

had to respond on what word they heard from a continuum between the words *hood* and *hud;* the results suggest that when listeners saw a male face, they were more likely to categorise the word as *hud*.

The Expectation-relative cues theory suggests that individuals form expectations about perceptual events and encode new information based on the differences between input and expectations. Comparing expectations and perceptual events leads to faster processing since it involves fewer steps and accounts for variability in the input. Concerning speech, theoretical accounts including Computing CUes Relative to Expectations [C-CuRE] (McMurray & Jongman, 2011) have tried to address the problem of integrating contextual information. McMurray and Jongman (2011) proposed a computational model that was based on a comparison of what types of information a computational model needed to categorise speech similarly to human listeners. Categorisation of speech sounds was made possible by using a set of acoustic cues. The findings indicated that the model could reach listener-level accuracy and match the listeners' pattern of errors. In this manner, contextual information was used to build expectations for specific cue values. This suggest that the auditory system uses talker gender information to create expectations for certain cue values and uses this information to encode incoming acoustic cues based on the difference between the acoustic cue and an expected value for that talker. For instance, a lower first formant would be predicted in the case the listener knew that the talker was male while VOT would be encoded slightly higher compared to VOT spoken by a female speaker. Listeners seem to depend on contextual information for phoneme categorisation.

The Shifting boundaries theory also provides an explanation for the importance of contextual information supporting that speech cues are encoded in their raw form while contextual information merely shifts phoneme category boundaries. Contextual information does not influence the earliest cue encoding level of processing; it rather influences later phoneme categorisation (that is, later levels of processing). Smits (2001a,b) argues that talker gender could affect phoneme categorisation but not cue encoding. According to this view, listeners shift their category between **/p/** or **/b/**; when the speaker is female, they tend to perceive **/b/** instead of **/p/**; this is because female speakers produce longer VOTs compared to male speakers.

The abovementioned theories provide a better understanding of how contextual information is encoded and processed into abstract categories. Several studies have been conducted, which support these theories;

however, no study could account for all available data. This is partly because measuring early sensory processes using behavioural measures is quite challenging.

3. Speech perception in CG

Speech perception involving the perceptual mapping from the highly variable acoustic speech signal to a linguistic representation such as phonemes, syllables, or words is an area that has not been extensively investigated in CG (Cypriot-Greek). Categorical perception suggests that discriminable speech sounds can be assigned to functionally equivalent classes while listeners may face several difficulties discriminating between sounds that they cannot identify as belonging to different categories. Given that acoustic cue(s) are available to all listeners, these difficulties are attributed to the fact that listeners are not skilled at attending to the acoustic cue(s) that can lead to the reliable identification of speech sounds. This is particularly seen in L2 (second language) listeners trying to discriminate the members of an L2 contrast.

Previous empirical studies have examined speech perception of plosive consonants to existing voicing contrasts in Standard Modern Greek (SMG) and CG and the corresponding characteristics in VOT by pre-schoolers (Okalidou et al., 2010, Okalidou et al., 2002). Results indicated a variation of voicing patterns for children than adults, in both SMG and CG VOT contrasts. Specifically, CG children presented a developmental lag in the acquisition of voicing contrasts compared to their Greek counterparts. Speech perception difficulties of CG listeners were further investigated in relation to L2 English. CG has a simple system of five vowels involving the monophthongs **/i e a o u/** but a rather complex system of fifty-one consonants including **/p p ʰ b t t ʰ d c c ʰ ɟ k k ʰ**: **g ts ʧ ʧ**ː **ʤ f f**ː **v v**ː **θ θ**ː **ð ð**ː **s s**ː **z z**ː **ʃ ʃ**ː **ʒ ʒ**ː **ç ç**ː **j j**ː **x x**ː **ɣ ɣ**ː **m m**ː **n n**ː **ɲ ŋ l l**ː **ʎ ɾ r r/** (Kkese, 2020a). Consonants in CG can be distinguished into voiced and voiceless; furthermore, plosive consonants can be distinguished by the absence or presence of aspiration resulting in a three-way contrast: voiced unaspirated plosives (i.e., *band* [ˈᵐ ba.ⁿta]) when preceded by a nasal consonant, voiceless, unaspirated plosives (i.e., *always* [ˈpa.ⁿta]), and voiceless aspirated plosives (i.e., *panda* [ˈpʰa.ⁿta]) plosives (Botinis et al., 2004; Kkese, 2016; Okalidou et al., 2010). A further difference refers to the use of singleton and geminate consonants in CG (Arvaniti, 1999, 2007, 2010; Kkese, 2016; Kkese & Karpava, 2019; Lengeris, 2009). Most consonants in CG can appear either as single or geminate in word-initial and -medial

positions but not in word-final position; this does not occur in English. On the other hand, English includes a more complex system of twenty vowels, out of which twelve are the monophthongs **/i: ɪ ɛ æ u: ʊ ɔ: ɒ ɑ: ʌ ɜ: ə/** and eight the diphthongs **/eɪ aɪ ɔɪ ɪə ɛə ʊə aʊ əʊ/** (Kkese, 2016, 2020a; Ladefoged, 2006). The consonantal system is simpler compared to CG since English involves only twenty-four consonants. These are **/p b t d k g f v Θ ð s z ʃ ʒ h ʧ ʤ m n ŋ l ɹ j w/**, which could be distinguished into voiceless and voiced. Aspiration is not a contrastive feature in English; this suggests that there is only a two-way contrast: voiced unaspirated (i.e., *bang*) and voiceless unaspirated consonants (i.e., *pang*).

CG listeners of L2 English experience difficulties with L2 voiced plosive consonants **/b d g/** (Kkese, 2016; Kkese & Petinou, 2017a,b) since these are absent or are realised differently in CG and English. Kkese (2016) investigated L2 difficulties with voiced plosive consonants at the word and sentence level. Performance was significantly better with reference to voiceless plosive consonants compared to their voiced counterparts at the word-initial position (i.e., *pacing-basing, towering-dowering, crammer-grammar*), word-medial position (i.e., *calipers-calibers, sighting-siding, lacquered-laggard*), and word-final position (i.e., *tripe-tribe, squat-squad, broke-brogue*). Kkese and Petinou (2017a) examined the types of errors involving plosive consonants indicating that performance was significantly better in the voiceless plosive category. CG listeners of L2 English were able to perceive voiced plosives but they treated such instances as a /nasal+voiced plosive/ sequence. Furthermore, Kkese and Petinou (2017b) investigated possible factors affecting plosive identification providing an indication of the rank order for the examined factors. Specifically, syllable position was identified as having the greatest influence on plosive identification, followed by voicing, word position, and place of articulation. In another perceptual study focusing on vowels and consonants in L2 English, CG listeners performed better when dealing with vowels compared to consonants (Kkese & Karpava, 2019). Based on the findings, most difficulties with vowels involved vowel length while plosive and fricative consonants appeared to the most difficult consonants for perception. The abovementioned studies suggest that the difficulties with consonants in L2 English may be the outcome of VOT cues. In English, voiced plosives are characterised by short lag VOT while their voiceless counterparts fall within the long lag VOT continuum. The same phonetic contrast is used in CG to differentiate between single and geminate voiceless plosives. In a study that is considered to be the first comparative sociophonetic research on SMG and CG and the first report on the acoustic structure of CG vowels, Themistocleous and Logotheti (2016) suggested significant effects of stress

on vowel formants, vowel intrinsic duration, and intrinsic fundamental frequency (F0). Effects of stress on vowel formants indicated that both SMG and CG stressed vowels are more peripheral than unstressed, unstressed vowels are more central and raised compared to stressed, and CG unstressed vowels are lower than SMG unstressed vowels. Concerning intrinsic duration, unstressed vowels in CG were found to be substantially longer than unstressed SMG vowels while as for F0, high vowels had higher F0 than middle and low vowels.

Moreover, heavy reliance on grapheme-phoneme correspondences may be another source of L2 speech perception difficulties for CG listeners. CGs' ability to spell gets in the way leading them to draw on the graphemes in a word rather than the phonemes when confronted with phonological awareness tasks. This is attributed to the L1, which is considerably much more transparent compared to L2 English. Studies by Kkese (2010, 2018, 2020a,b) have indicated that most of the spelling errors produced by L2 English users of CG are the outcome of transfer from L1.

The abovementioned studies emphasise the importance of acoustic cues while they examine speech perception on a phoneme, word, and utterance level. This stresses the need for contextual information studies. Consequently, the present study seems to be the first to examine how contextual information and specifically talker gender information is used during speech perception of L2 English by CG listeners. As a result, for this study, the following research questions will be investigated:

1. Will talker gender effects be observed for either L2 English consonants and/or vowels?
2. How well will CG listeners perform to L2 English consonants and vowels at the word level?

4. Methodology

4.1. Participants

Seventy undergraduate students *(n* = 30 females and 40 males) between the ages of 17-27 *(M_{age}*= 19.8 years) took part in this study attending an English-speaking private university. The language of instruction in which the participants were exposed was English even though the latter was not their major. However, participants had to attend an English compulsory subject, which took place two times per week, that is, four hours in total. The participants' proficiency levels varied and their years of studying

English ranged from 5 to 18 (M_{years} = 6.5). A brief questionnaire was distributed before the word dictation test to ensure homogeneity of the sample. The purpose of the questionnaire was to ensure that the group shared the same characteristics, that is, educational level and socio-economic status since the aim was to eliminate inter-group differences. Based on the data, participants were young adults with a CG linguistic background, and normal speech and hearing. Most of them reported using English in their everyday life (n=65) and having travelled to English-speaking countries (n=42). Non-probability convenience sampling was used since participants were selected based on their possession on targeted characteristics and availability. Participation was on a completely voluntary basis and students were ensured about the confidentiality of their personal details.

4.2. Stimuli

For the present study, a word dictation task was used to test L2 English vowel and consonant perception and written production of sounds by L1 CG students, their word recognition skills, and accuracy of decoding acoustic speech signal into words. Using words was thought to be more appropriate for any semantic and/or syntactic information from the input to be eliminated. A total of 120 target words were compiled (60 for consonants, 60 for vowels) that were tested in six sessions. Each session involved 20 target words; ten target words aimed at consonants and ten at vowels. There were 10 conditions for consonant sounds (6 test items each): **[ð]**, **[z]**, **[θ]**, **[v]**, **[d]**, **[ŋ]**, **[h]**, **[b]**, **[g]**, **[ɹ]** and 10 conditions for vowel sounds (6 test items each): **[æ]**, **[ɜ:]**, **[ɔ:]**, **[i:]**, **[u:]**, **[ɑ:]**, **[e]**, **[ʌ]**, **[ə]**, **[ʊ]**.

All stimuli were pronounced by a monolingual male or a female speaker of British English. Using bilingual speakers could affect various acoustic parameters such as VOT. The audio pronunciations were taken from the Macmillan online dictionary. For the 60 target words referring to consonants, 32 involved words pronounced by the male speaker and 28 by the female speaker; for vowels, 31 target words were pronounced by the male speaker and 29 by the female speaker. Since we intended to examine the effect of talker gender information on the learners' L2 perception, the target words were randomised per voice of the presenter.

4.3. Procedure

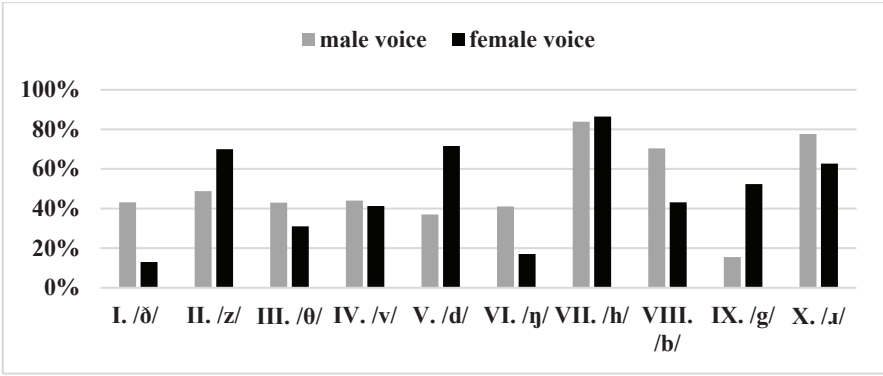
Collection of samples took place in one academic semester during class time. The word dictation tests focusing on L2 consonants and vowels were distributed in six different sessions with one-week interval. Each session lasted 20-30 minutes and it was done in groups of 15-20 students. First, the background questionnaire and the first word dictation test were distributed and selected during the same session. With reference to the background questionnaire, its purpose was to ensure homogeneity of the population sample, while it focused on the years of studying English as well as the setting and frequency of English use. A small pilot study involving five students before the word dictation sessions provided evidence that the testing stimuli were appropriately representative of each category/condition tested. The findings of this small case preliminary study were not included in the results.

Word dictation tests involved participants providing the orthographic input, which matched the acoustic input. Participants had to listen to stimuli for two times from Dolby, theatre sound system in class and respond accordingly while recording on a given scoresheet their responses. Between each repetition, there was 5-second pause; between words, the recording was paused until students finished writing the words down.

The classroom where the sessions were conducted involved a small computer lab. Participants were asked to sit at a separate desk, which supported a computer and listen to the each of the tasks through headphones (listening volume was set at 75dB). In this way, factors such as noise, talking to each other or copying each other's answers were prevented and uniform subject performance was ensured; no feedback was provided to the participants.

5. Results

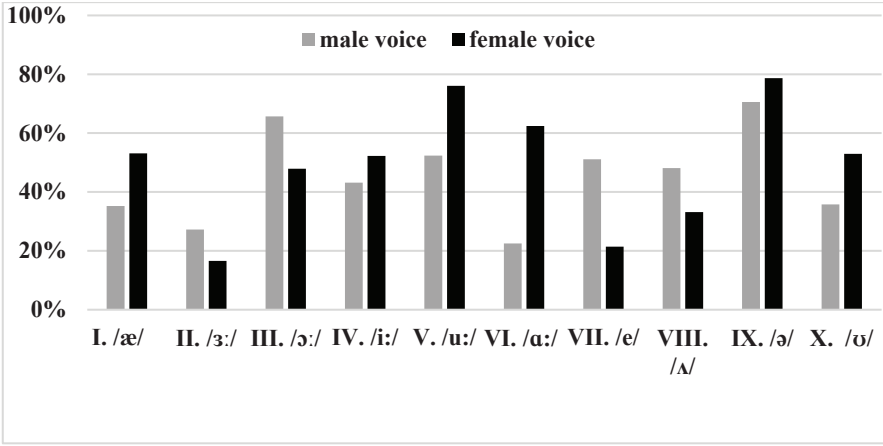
The results indicate that for the target words involving consonants, there was no clear preference relative to the gender of the talker. However, when words were spoken by a male talker, more target-like perception and written production was found for the consonants **/ð/** (43.14% vs. 13%), **/θ/** (43% vs. 31%), **/v/** (44% vs. 41.25%), **/ŋ/** (41% vs. 17%), **/b/** (70.38% vs. 43.11%), **/ɹ/** (77.67% vs. 62.73%). On the other hand, words spoken by a female talker elicited more target-like perception and written production for the consonants **/z/** (48.76% vs. 70%), **/d/** (37% vs. 71.55%), **/h/** (83.89% vs. 86.46%), **/g/** (15.47% vs. 52.38%). Graph 1 summarises these findings.



Graph 1: Target consonant production: female vs. male voice of the word presentation

According to paired sample t-test statistical analysis, there is a statistically significant difference between male and female voice regarding perception and production of the following consonants **/ð/** *(t(69)= -4.853, p=.000)*, **/z/** *(t(69)=2.846, p=.005)*, **/θ/** *(t(69)= -5.338, p=.000)*, **/v/** *(t(69)= -4.493, p=.000)*, **/d/** *(t(69)=3.996, p=.000)*, **/ŋ/** *(t(69)= -4.421, p=.000)*, **/h/** *(t(69)=7.810, p=.000)*, **/b/** *(t(69)= -5.946, p=.000)*, **/g/** *(t(69)=8.011, p=.000)*, but not of the **/ɹ/** consonant *(t(69)=.286, p=.775)*.

With reference to speech perception and written production of vowels, the results suggest that participants had more target-like production of vowels in the words spoken by a male talker for **/ɜ:/** (27.22% vs. 16.58%), **/ɔ:/** (65.67% vs. 47.89%), **/e/** (51.08% vs. 21.42%), and **/ʌ/** (48.09% vs. 33.14%). Words spoken by a female talker elicited significantly better performance when the vowels **/æ/** (35.24% vs. 53.11%), **/i:/** (43.14% vs. 52.22%), **/u:/** (52.34% vs. 76.05%), **/ɑ:/** (22.51% vs. 62.43%), **/ə/** (70.55% vs. 78.69%), and **/ʊ/** (35.76% vs. 52.93%) were involved. Graph 2 presents the results for vowels.



Graph 2: Target vowel production: female vs. male voice of the word presentation

According to paired sample t-test statistical analysis, there is a statistically significant difference between male and female voice regarding perception and production of the following vowels **/æ/** *(t(69)=3.374, p=.001)*, **/ɜ:/** *(t(69)= -3.125, p=.003)*, **/u:/** (*t(69)=5.146, p=.000)*, *(t(69)=8.276, p=.000)*, **/ɑ:/** *(t(69)=8.276, p=.000)*, **/e/** *(t(69)=-3.107, p=.002)*, **/ʌ/** (t(69)= -3.893, p=.000), **/ə/** *(t(69)=5.934, p=.000)*, **/ʊ/** *(t(69)=3.919, p=.000)*, but not for the following vowels **/ɔ:/** *(t(69)=1.895, p=.061)* and **/i:/** *(t(69)= -.096, p=.924)*.

Comparing consonants and vowels, the findings suggest that participants performed nearly the same on vowel and consonant sounds in the dictation task with a slightly better performance on consonants. According to the paired samples t-test, there is a statistically significant difference between target and non-target production, for both consonants and vowels *(t(69)= -9.958, p=.000)*; between target and non-target production of vowels *(t(69)= -8.398, p=.000)*; between target and non-target production of consonants *(t(69)= -11.271, p=.000)*; and between target production of consonants and vowels *(t(69)= 6.492, p= .000)*.

Talker gender information seems to affect the perception and written production of consonants and vowels in L2 English. Specifically, it was found that female voice elicits more target written production of vowels since vowels had 49.45% correct discrimination responses compared to 45.16% obtained for the male voice. On the other hand, male voice elicited more target written production of consonants since consonants had 50.46%

correct discrimination responses; the female voice elicited 49%. The paired samples t-test suggests that there is a statistically significant difference between the overall target written production of vowels in the words read by the female and/or male talker *(t(69)= 4.142, p=.000)*, but not in the case of consonants *((t(69)=.627, p=.533)*, see Table 1.

Table 1: Speech perception and written production of consonants and vowels

Overall	**Vowels**		**Consonants**	
target	48.67%		49.81%	
non-target	51.33%		50.19%	
Voice	**Male**	**Female**	**male**	**female**
target	45.16%	49.45%	50.46%	49%
non-target	54.84%	50.56%	49.54%	51%

According to Pearson correlation, age is correlated with the target production of consonants in words read by the female talker *(r(70)= -.250*, p=.037)*; visits to English-speaking countries is an important factor for the target production of vowels in words presented by male voice *(r(70)= -.351**, p=.002)*; target consonant production in words presented by female voice *(r(70)= -.244*, p=.042)* and target consonant production in words presented by male voice *((70)= -.349**, p=.003)*.

6. Discussion

The present study investigated the effect of talker gender information on the perception of contrastive English consonants and vowels by CG listeners. Also, the study examined the most challenging categories of both consonants and vowels in L2 English. The participants of the study completed a word dictation task in which they were asked to identify the correct word containing the target consonant or vowel from the auditory stimulus.

The results suggest that talker gender information influences how acoustic cues are encoded. This is consistent with previous findings, which support that listeners could compensate for the variability of the acoustic speech signal across different types of context such as talker gender information to be able to categorise speech sounds successfully (Johnson et al., 1999; McMurray & Jongman, 2011; Repp & Mann, 1980; Strand, 1999; Strand &

Johnson, 1996). However, the most significant finding was that words pronounced by a female talker elicited more target production of vowels; the words pronounced by a male talker elicited more target production of consonants. This suggests that variation between male and female talkers occurs because of differences in vowel formants and VOT. In the effort to account for the participants' different performance in terms of consonants and vowels in L2 English, four reasons seem to merit special attention as briefly described below.

a) Vowel formants are a spectral cue, whereas VOT is a temporal cue. Spectral cues are more heavily influenced by talker due to the differences in the vocal tract. In the case of vowels, this refers to the frequencies of the first and second formant.
b) Vowel differences are defined by more acoustic cues (2-5 formants) compared to consonants, which are defined by VOT that differs along one linear dimension.
c) Gender differences in vowels result from differences in pitch and formant frequencies, which are influenced by gender. Gender differences in consonants, however, result from VOTs that depend on the use of aspiration. Even though women talk with more aspiration (Klatt & Klatt, 1990), this feature is not necessarily related to gender.
d) Vowels consist of talker gender information throughout the relevant sounds via pitch and formants but talker gender information for consonants can only be accessed before or after VOT.

Several theories of speech perception have attempted to account for contextual information such as talker gender information and how it affects acoustic cue encoding. According to the shifting boundaries theory, talker gender information does not influence acoustic cue encoding but it affects only the later stages of processing as when it shifts the boundaries between phoneme categories (Smits, 2001a,b). This implies that the cue encoding processing cannot be influenced by the preceding context. Similarly, the expectation-relative cues theory proposes that cues are encoded based on their raw values and only then talker gender information and other contextual information can recode the cue values based on the context (McMurray & Jongman, 2011). Nonetheless, this theory assumes that other modalities could influence the re-coding. In the present study, no other modality has been included, which suggests that this theory may not be able to account for the results. The auditory contrast theory assumes that acoustic cues are compared to immediate previous information during encoding (Assgari & Stilp, 2015; Holt et al., 2000). However, previous information

has to be from the same modality, which again seems to be problematic for the present study since the target sounds could be at word-initial position. On the other hand, talker gender information may be encoded along the acoustic cues as an integral part of the exemplars in the exemplar theory (Palmeri et al., 1993). This theory, though, uses talker gender information as an additional cue for categorising phonemes instead of influencing cue encoding. As a result, none of the theories has suggested multiple mechanisms to talker variability. Predictive coding may provide a more appropriate alternative supporting that the auditory system could employ acoustic cues based on talker gender to deal with variability of the acoustic speech signal, but this may not be the only way since previous knowledge could also contribute to this compensation (Gagnepain et al., 2012).

With respect to L2 phonemes, it was found that CG listeners encountered difficulties with plosive, fricative, and nasal consonants and particularly with **/g ð θ ŋ/**. These findings agree with Kkese and Karpava (2019), even though the experimental tasks were different. For the present study, participants had largely substitution and no production errors regarding consonants. The substitution errors were based on the similarity or contrast of voice feature and manner of articulation. When it comes to vowels, most of the participants' difficulties were manifested in short and long vowels such as /ɜ: **ɑ: i: æ ʌ ʊ**/. This may be attributed to the differences between English and CG phonological and writing systems since in CG, there is no long-short vowel distinction. Non-target perception and production of vowels was characterised by no production and substitution errors based on the similarity of [±back], [±round] and duration features. Lengeris (2009) also found that the vowel length feature causes perceptual difficulties to L2 users of English.

7. Conclusion

The study provided novel findings which suggest how contextual information such as talker gender information influence phoneme categorisation in L2 English by CG listeners. Nonetheless, several limitations have to be acknowledged. For instance, the acoustic speech signal always includes simultaneous talker gender information and phonetic information; therefore, isolating talker gender information from the acoustic speech signal can be quite difficult. Furthermore, other talker-specific factors such as dialectal differences and/or differences in accent may influence L2 phoneme perception. In order to compensate for this, only British English talkers were included for the present study. Studies comparing CG/SMG and English

and/or other L1 backgrounds could also provide interesting findings about speech perception. Future studies could also investigate whether talker gender information is used immediately by the auditory system or is stored in memory to be integrated with other cues when they arrive. This will enable a close insight into the use of talker gender information as an acoustic cue or as context. In the case of VOT, talker information should not be used before the arrival of the VOT as it can serve to bias the listener's use of VOT (direct use as a cue); on the contrary, by using talker information after VOT is encoded, implies that this is stored for later and is used as context. Whereas VOT needs to be relatively encoded, vowels do not since they contain robust talker information such as fundamental frequency and formants, which always co-exist. A final point refers to visual information about talker's gender and whether it influences acoustic cue encoding. Generally, vowels are thought to be easier to distinguish than consonants since they are produced with a different oral cavity shape. Examining how visual gender information affects phoneme categorisation could provide a useful insight into information processing since the two types of information are processed in different areas of the cortex. Exposing the participants to visual stimuli such as videos of the target words while listening to gender-neutral audio stimuli could provide a more reliable way to measure the effect of visual gender information on acoustic cue encoding.

References

Allen, J. S., Miller, J. L., & DeSteno, D. (2003). Individual talker differences in voice-onset-time. *Journal of the Acoustical Society of America, 113*(1), 544 –552.

Arvaniti, A. (1999). Greek voiced stops: prosody, syllabification, underlying representations or selection of the optimal? *Proceedings of the 3rd International Linguistics Conference for the Greek Language*, Athens, Ellinika Grammata, 383–390.

Arvaniti, A. (2007). Greek phonetics: The state of the art. *Journal of Greek Linguistics*, *8*(1), 97–208.

Arvaniti, A. (2010). A (brief) overview of the phonetics and phonology of Cypriot Greek. Proceedings of the conference on *The Greek language in Cyprus from antiquity to today*, University of Athens.

Assgari, A. A., & Stilp, C. E. (2015). Talker information influences spectral contrast effects in speech categorization. *Journal of the Acoustical Society of America, 138*(5), 3023–3032.

Blumstein, S. E., & Stevens, K. N. (1979). Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics

of stop consonants. *Journal of the Acoustical Society of America, 66*(4), 1001–1017.

Botinis, A., Christofi, M., Themistocleous, C., & Kyprianou, A. (2004). Duration correlates of stop consonants in CG. *Proceedings, FONETIK 2004*, Dept. of Linguistics, Stockholm University.

Cole, J., Linebaugh, G., Munson, C. M., & McMurray, B. (2010). Unmasking the acoustic effects of vowel-to-vowel coarticulation: A statistical modeling approach. *Journal of Phonetics, 38*(2), 167–184.

Gagnepain, P., Henson, R. N., & Davis, M. H. (2012). Temporal predictive codes for spoken words in auditory cortex. *Current Biology: CB, 22*(7), 615–621.

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America, 97*(5), 3099–3111.

Holt, L. L., Lotto, A. J., & Kluender, K. R. (2000). Neighboring spectral content influences vowel identification. *Journal of the Acoustical Society of America, 108*(2), 710–722.

Johnson, K., Strand, E. A., & D'Imperio, M. (1999). Auditory–visual integration of talker gender in vowel perception. *Journal of Phonetics, 27*(4), 359–384.

Kessinger, R. H., & Blumstein, S. E. (1997). Effects of speaking rate on voice-onset time in Thai, French, and English. *Journal of Phonetics, 25*(2), 143–168.

Kkese, E. (2010). Issues in L2 acquisition of spelling: When spelling tests are simply not enough. In P. Pavlou (ed.), *Research on English as a Foreign Language in Cyprus, 1,* 223-257. Nicosia: University of Nicosia Press.

Kkese, E. (2016). Identifying plosives in L2 English: The case of L1 Cypriot Greek speakers. *Linguistic Insights: Studies in Language and Communication, li217*. Switzerland: Peter Lang AG.

Kkese, E. (2018). Assessing L2 writing in the absence of scoring procedures: Construction of rating scales in a Cypriot-Greek EFL in-class context. *Journal of English Education, 3*(2), 46–56.

Kkese, E. (2020a). *L2 Writing Assessment: The Neglected Skill of Spelling*. Cambridge Scholars Publishing.

Kkese, E. (2020b). Phonological awareness and literacy in L2: Sensitivity to phonological awareness and phoneme-grapheme correspondences in L2 English. In G. Neokleous, A. Krulatz, & R. Farrelly (Eds.), *Handbook of Research on Cultivating Literacy in Diverse and Multilingual Classrooms* (pp. 62–81). IGI Global Press.

Kkese, E., & Petinou, K. (2017a). Perception abilities of L1 Cypriot Greek listeners – Types of errors involving plosive consonants in L2 English. *Journal of Psycholinguistic Research, 46*(1), 1–25.

Kkese, E., & Petinou, K. (2017b). Factors affecting the perception of plosives in second language English by Cypriot-Greek listeners. In E. Babatsouli (Ed.), *Proceedings of the International Symposium on Monolingual and Bilingual Speech 2017* (pp. 162–167).

Kkese, E., & Karpava, S. (2019). Applying the Native Language Magnet Theory to an L2 setting: Insights into the CGR adult perception of L2 English. In E. Babatsouli (Ed.), *Proceedings of the International Symposium on Monolingual and Bilingual Speech* 2019 (pp. 67–74).

Klatt, D. H., & Klatt, L. C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *Journal of the Acoustical Society of America, 87*(2), 820–857.

Kluender, K. R., Coady, J. A., & Kiefte, M. (2003). Sensitivity to change in perception of speech. *Speech Communication, 41*(1), 59–69.

Ladefoged, P. (2006). *A Course in Phonetics* (5th ed.). Boston: Thomson/Wadsworth.

Lengeris, A. (2009). Perceptual assimilation and L2 learning: Evidence from the perception of Southern British English vowels by native speakers of Greek and Japanese. *Phonetica, 66*(3), 169–187.

Laing, E. J. C., Liu, R., Lotto, A. J., & Holt, L. L. (2012). Tuned with a tune: Talker normalization via general auditory processes. *Frontiers in Psychology, 3*, 1–9.

Lengeris, A. (2009). Perceptual assimilation and L2 learning: Evidence from the perception of southern British English vowels by native speakers of Greek and Japanese. *Phonetica*, *66*(3), 169–187.

Lindblom, B. (1996). Role of articulation in speech perception: Clues from production. *Journal of the Acoustical Society of America, 99*(3), 1683–1692.

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

Mann, V. A. (1980). Influence of preceding liquid on stop consonant perception. *Perception and Psychophysics, 28*(5)*,* 407–412.

McMurray, B., & Jongman, A. (2011). What information is necessary for speech categorization? Harnessing variability in the speech signal by integrating cues computed relative to expectations. *Psychological Review, 118*(2), 219–246.

Morris, R. J., McCrea, C. R., & Herring, K. D. (2008). Voice onset time differences between adult males and females: Isolated syllables. *Journal of Phonetics, 36*(2)*,* 304–317.

Murray, I. R., & Arnott, J. L. (1993). Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. *Journal of the Acoustical Society of America, 93*(2), 1097–108.

Nearey, T. M. (1989). Static, dynamic, and relational properties in vowel perception. *Journal of the Acoustical Society of America, 85*(5), 2088–2113.

Nearey, T. M., & Rochet, B. L. (1994). Effects of place of articulation and vowel context on VOT production and perception for French and English stops. *Journal of the International Phonetic Association, 24*(1), 1–18.

Ohala, J. J. (1996). Speech perception is hearing sounds, not tongues. *Journal of the Acoustical Society of America, 99*(3), 1718–1725.

Okalidou, A., Petinou, K., Theodorou, E., & Karasimou, E. (2002). Development of voicing contrasts in stop sounds in Modern Greek and CG children. *Proceedings of the 5th Colloque International De Linguistique Grecque*, Sorbonne, Recearch de Linguistique Creque, Sorbonne, Universite Rene Descartes- Paris 5, Paris, France.

Okalidou, A., Petinou, K., Theodorou, E., & Karasimou, E. (2010). Development of voice onset time in standard-Greek and Cypriot-Greek-speaking preschoolers. *Clinical Linguistics and Phonetics, 24*(7), 503–519.

Palmeri, T. J., Goldinger, S. D., & Pisoni, D. B. (1993). Episodic encoding of voice attributes and ecognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*(2), 309–328.

Peterson, G., & Barney, H. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America, 24*(2), 175–184.

Repp, B. H. (1982). Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception. *Psychological Bulletin, 92*(1), 81–110.

Repp, B. H., & Mann, V. A. (1980). Influence of vocalic context on perception of the [ʃ]-[s] distinction. *Perception and Psychophysics, 28*(3), 213–228.

Ryalls, J., Zipprer, A., & Baldauff, P. (1997). A preliminary investigation of the effects of gender and race on voice onset time. *Journal of Speech, Language, and Hearing Research, 40*(30), 642–645.

Smits, R. (2001a). Evidence for hierarchical categorization of coarticulated phonemes. *Journal of Experimental Psychology, 27*(5), 1145–1162.

Smits, R. (2001b). Hierarchical categorization of coarticulated phonemes: A theoretical analysis. *Perception and Psychophysics, 63*(7), 1109–1139.

Stevens, K. N., & Blumstein, S. E. (1978). Invariant cues for place of articulation in stop consonants. *Journal of the Acoustical Society of America, 64*(5), 1358–1368.

Strand, E. A. (1999). Uncovering the role of gender stereotypes in speech perception. *Journal of Language and Social Psychology, 18*(1), 86–100.

Strand, E. A., & Johnson, K. (1996). Gradient and visual speaker normalization in the perception of fricatives. In D. Gibbon (Ed.), *Natural Language Speech Processing and Speech Technology* (pp. 14–26). Berlin: Mouton de Gruyter.

Summerfield, Q. (1981). Articulatory rate and perceptual constancy in phonetic perception. *Journal of Experimental Psychology*: *Human Perception and Performance, 7*(5), 1074–1095.

Swartz, B. L. (1992). Gender difference in voice onset time. *Perceptual and Motor Skills, 75*(3), 983–992.

Themistocleous, C., & Logotheti, A. (2016). Standard Modern Greek and Cypriot Greek vowels: a sociophonetic study. In A. Ralli, N. Koutsoukos, S. Bompolas (Eds.), *Proceedings of the 6th International Conference on Modern Greek Dialects and Linguistic Theory* (MGDLT6), September 25-28, 2014, pages 177–183. University of Patras.

Toscano, J. C. (2011). *Perceiving speech in context: Compensation for contextual variability during acoustic cue encoding and categorization*. PhD thesis, University of Iowa.

Toscano, J. C., & McMurray, B. (2015). The time-course of speaking rate compensation: effects of sentential rate and vowel length on voicing judgments. *Language, Cognition, and Neuroscience, 30*(5), 529–543.

Whiteside, S. P., & Irving, C. J. (1998). Speakers' sex differences in voice onset time: A study of isolated word production. *Perceptual and Motor Skills, 86*(2)*,* 651–654.

Williams, D. (2013). *Cross-language acoustic and perceptual similarity of vowels: The role of listeners' native accents*. PhD thesis, University of Sheffield.

CHAPTER NINE

L2 PHONE PERCEPTION ON A SYLLABIC LEVEL MIGHT BE DRIVEN BY L1 TRANSFER MECHANISMS AND/OR UNIVERSAL PATTERNS OF PHONE ACQUISITION

ELENA KKESE AND GEORGE GEORGIOU

Abstract

The present study examines how contrastive L2 English plosives on syllable-level are perceived by Cypriot Greek speakers and how the use of L2 affects these perceptions. The participants were 30 university students with sufficient knowledge of English as an L2. They completed two forced-choice discrimination tasks in which they had to discriminate three different types of English plosives contrasted in voicing in both closed (CVC) and open syllables (VC); plosive position was also taken into consideration. The findings indicated that participants yielded poor to moderate discrimination accuracy over the L2 plosive contrasts under investigation. Also, they performed significantly better in the open-syllable task, while they were proceeding to more errors in terms of plosives in the coda position in closed syllables. L2 use did not have an effect on the discrimination of contrastive English plosives. Together, the data suggest that syllables have an important role in the segmental perception of spoken L2 English, while L1 transfer and universality of phone perception on a syllabic level may be influential factors.

Keywords: speech perception; phonemes; open syllables; closed syllables; plosive consonants

1. Introduction

Speech perception refers to the mapping of auditory input signals into linguistic information. Nonetheless, this mapping may be quite challenging since it involves transforming the auditory signal into several levels of language structure such as phonemes, syllables, morphemes, words, phrases, and sentences. Further, the spoken input is a continuous and noisy signal (Hurmalainen, Gemmeke, & Virtanen 2012). Speakers do not provide adequate reliable cues to the boundaries of units of meaning and effects such as coarticulation and phonemic context can further affect speech segments such as phonemes. Even so, listeners can naturally extract meaning from speech identifying the speech sounds (phonemes), recognising the words they make up (lexicon), and understanding the message they produce (semantics) by the sentences they form (syntax) (Kkese, 2016).

Trying to understand the segmentation units that are involved in speech perception and facilitate access to lexical representations of word form, the two sublexical units of phonemes and syllables seem to be of particular importance and have been extensively studied (Cutler et al., 1986; Dumay, Frauenfelder, & Content 2002; Zwitserlood, 2004). A phoneme constitutes the smallest linguistic unit an utterance can be broken into; it is, therefore, the smallest set of perceptual units, decreasing substantially the parts of the speech wave that must be tested, making the process faster. Syllables, on the other hand, can be viewed as combinations of phonemes; therefore, they involve a larger set of perceptual units and a larger candidate set. A syllable must have a centre, called nucleus or peak, which has to be a vowel (i.e., *I* [aɪ]); depending on language-specific constraints, the syllable could have an onset, which is the initial part of the syllable that consists of either one or more consonants (i.e., *buy* [baɪ]) and a coda, which is the final part of the syllable involving either one or more consonants (i.e., *bite* [baɪt]). As a result, the auditory context can benefit syllables since phonemes may be affected by the characteristics of the neighbouring phonemes rendering a sound very difficult to be identified without reference to its context.

With reference to phoneme-sized representations, several studies support the role of phonemes in speech perception (Bien et al., 2009; Bien & Zwitserlood, 2013; Hickok & Poeppel, 2007; Kkese, 2016; Kkese & Petinou, 2017a,b; Kkese & Lokhtina, 2017; Kkese & Karpava, 2019; Georgiou, 2018, 2020; Obleser & Eisner, 2009). In these studies, the emphasis is to identify the segmental content associated with the corresponding auditory signal. Word-recognition models suggest that phonemes are enough for lexical access and selection when these are continuously mapped

onto the lexical word-form representations, irrespectively of where the words begin or end (Christiansen & Chater, 2001; McClelland & Elman, 1986). Spoken-word recognition models, on the other hand, suggest that word onsets are important for lexical access (Marslen-Wilson, 1987; Norris & McQueen, 2008). Several studies indicated that potential onsets may be signalled by transitional probabilities between consecutive segments (Saffran, Aslin & Newport, 1996), syllabic boundaries (Content, Kearns, & Frauenfelder, 2001; Cutler et al., 2001; Zwitserlood, 2003, 2004), syllable duration and stress (Tyler & Cutler, 2009; Langus et al., 2012), phrase-boundary cues (Christophe et al., 2004) or a combination of these factors (Mattys, White & Melhorn, 2005).

On the other hand, syllables are thought as suprasegmental structures that provide information for phone recognition (Hallé et al., 1998; Massaro & Cohen, 1983; Pitt, 1998) and language parsing (McQueen, 1998; Vroomen & Gelder, 1999). The important role of syllables in speech perception has been stressed in studies employing monitoring paradigms (Mehler et al., 1981). Listeners are generally faster monitoring for syllable-sized targets than phoneme-sized targets (Svain & Bever, 1970; Segui, Frauenfelder, & Mehler, 1981). This suggests that listeners syllabify the speech signal and, therefore, syllable-based representations enable lexicon access (Bradley, Sánchez-Casas, & García-Albea, 1993). Nonetheless, monitoring paradigms have provided mixed evidence for English (Bradley, Sánchez-Casas, & García-Albea, 1993; Cutler et al., 1986).

Syllables may not be as effective in prelexical processing across languages. This happens since languages differ in terms of syllable boundaries and the extent these are clear and unambiguous, as well as in the regularity of syllable structure (Bradley, Sánchez-Casas, & García-Albea, 1993). With reference to English, this involves segments that are part of two syllables at once (ambisyllabic segments). Examples could be intervocalic consonants preceding an unstressed vowel. Bradley, Sánchez-Casas and García-Albea (1993), though, argue that ambisyllabicity may cause difficulties for syllabic segmentation since syllable boundaries are not clear. Cutler, Dahan and van Donselaar (1997) argue that listeners use the rhythmic structure of their language in the effort to segment the speech signal. In English, listeners use stress as a segmentation cue given that the language is a stress-timed language and syllable boundaries are not clear. Examples could be the words *speeches* [ˈspiːtʃəs] and *battle* [ˈbætəɬ], in which both [spiːtʃ] and [tʃəs] as well as [bæt] and [təɬ] are permissible English syllables; [spiː] and [əs] or [bæ] and [əɬ] are less satisfactory. As a result, /ʧ/ and /t/ in [ˈspiː[tʃ]əs] and [ˈbæ[t]əɬ] seem to be part of both the first and second syllables

(Anderson & Jones, 1974; Kahn, 1976). Nonetheless, in syllable-timed languages such as Standard Modern Greek (SMG) or Spanish, syllabification is clear since there are syllabic boundaries. Syllable-boundary information seems to be a cue that facilitates the parsing of speech input for lexical access (Content, Kearns, & Frauenfelder, 2001; Cutler et al., 2001; Dumay, Frauenfelder, & Content, 2002; Zwitserlood, 2003, 2004). Examples could be the words *λόγοι* [ˈlo.ji] and *discursos* [dis.ˈkuɾ.sos] for *speeches* as well as *μάχη* [ˈma.çi] and *batalla* [ba.ˈta.ja] for *battle* in SMG and Spanish respectively.

2. Studies on the segmental and suprasegmental level

Different models for speech learning examined the connection between second language (L2) perception/production and the first language (L1) phonological system by employing data from the segmental level such as consonants and vowels (Best, 1995; Best & Tyler, 2007; Flege, 1995, 2003, 2007; Georgiou & Kkese, 2020; Kuhl, 2004). The Speech Learning Model (Flege, 1995, 2003, 2007) suggests that adult L2 speakers acquire the consonants and vowels of the target language by relying on phonetic similarity to L1. As a result, L2 production will be facilitated and/or hindered according to how accurately the L2 sounds are perceived (MacKay et al., 2001). According to the model, L2 learners irrespectively of age will maintain their ability to associate their L2 production to long-term memory representations for L2 segments. Likewise, the Perceptual Assimilation Model and its later extension, the Perceptual Assimilation Model-L2, (Best, 1995; Best & Tyler, 2007) suggest that L2 listeners assimilate L2 sounds to L1 categories depending on their similarities and degrees of learning difficulty. Specifically, L2 contrasts could be a Two Category, Category Goodness, Single Category, Uncategorised-Categorised, and Uncategorised-Uncategorised contrast (see Best & Tyler, 2007). According to Best and Strange (1992), the ability to discriminate L2 contrasts may improve due to L2 experience. A third model refers to the Native Language Magnet/Neural Commitment (Kuhl, 2004), which includes a neural-level specification and explanation by adhering to both developmental and cross-linguistic perspectives. According to this model, language acquisition is seen as an adaptive computational process aiming to extract the abstract speech categories and higher-order linguistic structures (Kuhl, 2004).

Whereas the three theoretical models on speech learning indicate that segmental differences between L1 and L2 considerably affect L2 perception and production at a syllable-initial position, differences in L1 and L2

syllable structures seem to be equally important (Anderson, 1983; Broselow & Xu, 2004; Cebrian, 2007; Flege, 1989; Hayes-Harb et al., 2008; Rogers & Dalby, 2005; Stockman & Pluut, 1999; Tsukada et al., 2004). For instance, Cypriot-Greek (CG) allows no syllable-final consonants other than /s/ and /n/ (Kkese, 2016). Similarly, Mandarin Chinese does not allow syllable-final consonants other than /n/ and /ŋ/ (Duanmu, 2000). This implies that L2 perception in English is more accurate at the syllable-initial rather than the syllable-final position (Flege, 1989) while more production errors are expected in the syllable-final than the -initial position (Bent, Bradlow & Smith, 2007; Rogers & Dalby, 2005). Rogers and Dalby (2005) also suggest that most difficulties involve the word-final position since Mandarin users of L2 English perceive considerably better the phonemic contrasts in word-initial position. The studies by Eckman (1981) and Flege (1995) further support the importance of L2 syllable structure in the L1 and of phonotactics on L2 perception and production.

These findings point to the fact that difficulties in syllable-final position in L2 perception and production of consonants may be attributed to a general pattern of universal structural asymmetry favouring the syllable-initial onset because of accessibility, stability, and learnability (Carlisle, 2001; Nam, Goldstein & Saltzman, 2009; Ohala, 1996). If the asymmetry could be universally applied, then it could even occur across L1 learners, irrespectively of the language experience. Further, if this asymmetry is an important force when it comes to L2 phonetic learning, then it could overrule the influence of acoustic, phonetic, and phonemic factors and apply to all the categories of consonants.

3. Phonological systems of CG and English

A comparison between CG and English phonemes and syllable structures seems mandatory at this point to understand the consonant and vowel systems as well as the syllable structures of the two linguistic varieties. At the individual phoneme level, English has forty-four phonemes; twenty-four phonemes stand for consonants and twenty phonemes stand for vowels. In contrast, CG has approximately fifty-six phonemes; fifty-one phonemes account for consonants and five for vowels (Kkese 2020a). With reference to English consonants, these involve the phonemes /p b t d k g f v ϴ ð s z ʃ ʒ h ʧ ʤ m n ŋ l ɹ j w/ and can be distinguished into voiceless and voiced as in *thigh* [θaɪ] and *thy* [ðaɪ]. The same occurs in CG since the phonemes for the consonants /p pʰ b t tʰ d c cʰ ɟ k kʰ g ts ʧ ʧː ʤ f fː v vː θ θː ð ðː s sː z zː ʃ ʃː ʒ ʒː ç çː j jː x xː ɣ ɣː m mː n nː ɲ ŋ l lː ʎ ɾ r/ can be distinguished into

voiced and voiceless as in the CG words for *canopy* [ˈθɔ.lɔs] and *deceit* [ˈðɔ.lɔs]. With reference to plosive consonants in CG, these can further be distinguished by the absence or presence of aspiration resulting in a three-way contrast: voiced unaspirated plosives (i.e., *ball* [ˈᵐba.la]) when preceded by a nasal consonant, voiceless, unaspirated plosives (i.e., *hay ball* [ˈpa.la]), and voiceless aspirated plosives (i.e., *cleaver* [ˈpʰa.la]) plosives (Botinis et al., 2004; Kkese, 2016; Okalidou et al., 2010) (see Table 1). Nonetheless, in English, aspiration is not a contrastive feature, as a result, there is no lexical distinction between aspirated and unaspirated plosives (i.e., spin - pin) even though there is phonetic difference (Fung, 2004). Another important difference refers to the use of singleton and geminate consonants in CG (Arvaniti, 1999, 2007, 2010; Kkese, 2016; Kkese & Karpava, 2019; Lengeris, 2009). Specifically, most consonants in CG can appear either as single or geminate in word-initial and -medial positions but not in word-final position; this does not occur in English.

Table 1: **IPA charts for CG and English plosive consonants (adapted from Kkese, 2020b)**

	Bilabial	Alveolar	Velar
CG	ᵐb p pʰ:	ⁿd t tʰ:	ᵑg k kʰ:
English	b p	d t	g k

Concerning vowels, English includes twelve monophthongs /iː/, /ɪ/, /ɛ/, /æ/, /uː/, /ʊ/, /ɔː/, /ɒ/, /ɑː/, /ʌ/, /ɜː/, /ə/ and eight diphthongs /eɪ/, /aɪ/, /ɔɪ/, /ɪə/, /ɛə/, /ʊə/, /aʊ/, /əʊ/ (see Georgiou, 2019a,b; Kkese, 2016, 2020a; Ladefoged, 2006). CG on the other hand, consists of only five phonemes, which are similar to the vowels of SMG; these are the monophthongs /i, ɛ, ɐ, ɔ, u/ while no diphthongs are involved (see Arvaniti, 1999; Georgiou, 2018; Kkese, 2016, 2020a). /i/ is of special interest since it can be affected by neighbouring phonemes. Examples may be the word *one* (fem.) pronounced as [mɲɐ] when the vowel is between /m/ and another vowel in the same syllable or [mi.ɐ] when there is a syllable boundary; the same is observed for the word *ears*, which could be [ɐf.ˈcɐ] or [ɐf.ˈtiɐ], and *eyes*, which could be [ˈmɐ.ϴcɐ] or [ˈmɐ.tiɐ]; in both words /i/ precedes another vowel in the same syllable.

Consonants and vowels constitute the segmental level and together they form syllables. The latter constitute an important phonological unit across languages, which are perceived via segmental cues and phonotactics. In

both CG and English, segments can be organised into CV-type syllables containing a consonant and a vowel as in *yes* [nɛ] and *no* [nəʊ]; V-type syllables containing an empty onset as in *there* [ɛ.ˈci] and *open* [ˈəʊ.pən]; also, branching onsets are acceptable given that certain phonotactic constraints are applied on the form that the branching onsets may take based on the Sonority Hierarchy (Carr, 1999; Kkese, 2016) (see Table 2).

Table 2: **Phonotactics constraints in CG and English**

	CG	English
single onsets	[kɐ.ˈlɔs] (good)	[nəʊ] (no)
2-member onsets	[ˈflɔ.ɣɐ] (flame)	[snəʊ] (snow)
3-member onsets	[strɐ.ˈtɔs] (army)	[stɹɔː] (straw)
4-member onsets	[ɛ.kstrɐ.ˈtiɐ] (campaign)	----
single codas	[ɔ.ˈrɛ.ɔs] (handsome)	[iːt] (eat)
2-member codas	[tʰːɛst] (test)	[iːst] (east)
3-member codas	[tʰːɐŋks] (tanks)	[ɑːskt] (asked)
4-member codas	----	[tʰɛksts] (texts)

Both languages allow closed syllables (Figure 1a, b) such as *light* [fɔs] and *fought* [fɔːt], *garden* ['ci.pɔs] and [ˈgɑː.dən) as well as open syllables (Figure 2a, b) such as *yes* [nɛ] and *key* [kʰiː], *very* [pɔ.'lːi] and ['vɛ.ɹi]. In CG, however, most words end with an open syllable (a vowel); the only coda consonants in word-final position are the coronal segments /s n/ as in *young* [mi.ˈʦis], *mob/hare* [lɐ.ˈɔs] and *mountain* [vu.ˈnɔn], *hand* [ˈʃɛ.rin]. In case other coda consonants appear in word-final position, this involves borrowings of foreign origin as in *computer* [kɔ.ˈᵐbju.tɛr], *golf* [ᵑgɔlf], or *tanks* [tʰːɐŋks]. Word-internal codas such as *purse/bag* ['tsɐn.tɐ] are also present in the dialect given that these are weaker than the following onsets (Kappa, 2002). On the other hand, words in English can end with a vowel *key* [kʰiː], a consonant *yes* [jɛs], two consonants *tent* [tɛnt] (sequence of plosives, sonarants, coronals), three consonants *sixth* [sɪksθ] (usually /t, s, θ/), or four consonants *sixths* [sɪksθs] (usually /s, t, d/).

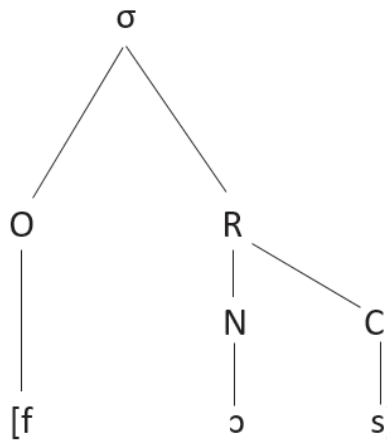


Figure 1a. Closed syllables in CG

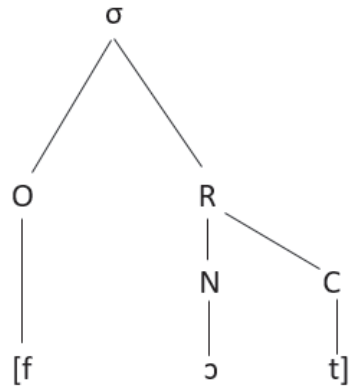


Figure 1b. Closed syllables in English

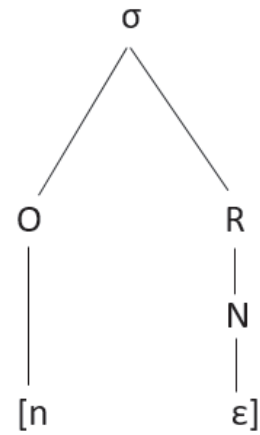


Figure 2a. Open syllables in CG

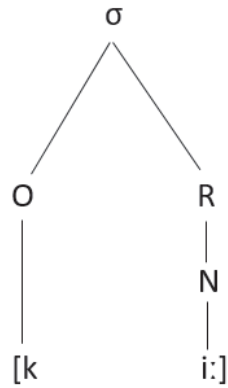


Figure 2b. Open syllables in English

Consequently, the present study seems to be the first to test phoneme perception on a syllable level and specifically how syllable structure differences between the L1 CG and L2 English affect the perception of plosive consonants in the L2. This allows a close insight into the effect of L1 transfer based on the difference of plosive consonant inventories and syllable structures of CG and English, as well as L2 language experience on non-native perception of L2 English phonemes. Even though adult L2 perception and production studies indicate a positional asymmetry that

could be attributed to differences in L1 and L2 syllable structures in a number of languages (Anderson, 1983; Broselow & Xu, 2004; Cebrian, 2007; Flege, 1989; Hayes-Harb et al., 2008; Rogers & Dalby, 2005; Stockman & Pluut, 1999; Tsukada et al., 2004), phonetic research comparing L1 CG and L2 English is limited. For example, research dedicated to phoneme perception refers to plosive consonant perception on a word (Kkese, 2016; Kkese & Petinou, 2017a,b) or utterance level (Kkese, 2016), as well as consonant and vowel perception on a word level (Georgiou, 2019a,b; Kkese & Karpava, 2019). As a result, for this study, the following research questions will be investigated:

1. To what extent can CG speakers discriminate L2 English plosives contrasted in voicing in both closed and open syllables?
2. What is the effect of consonant position (onset/coda for closed, and onset/medial for open syllables) on the discrimination of L2 English contrastive plosives?
3. How does the use of English in the speakers' everyday life affect the discrimination of L2 English plosive contrasts?

We hypothesise that there will be some difficulties in the discrimination of plosive contrasts in both types of syllables considering the findings of previous studies (e.g., Kkese, 2016). Also, we predict that coda consonants will be distinguished with more difficulty than the initial consonants since consonants in coda position are less frequent in the speakers' L1. Finally, we assume that individuals who use English in their everyday life will be able to discriminate plosive contrasts more accurately than individuals who do not use English. This is because evidence from the literature designates L2 use as an important predictor of L2 accent acquisition (Piske et al., 2001).

4. Methodology

4.1. Participants

Thirty active undergraduate students (n = 28 females and 2 males) between the ages of 17-23 (M_{age}= 18.47 years) took part in this study attending a Greek-speaking public university. The language of instruction in which the participants were exposed was SMG, while they had to attend an English compulsory subject, which took place two times per week, that is, three hours in total. The participants' proficiency levels varied and their years of studying English ranged from six to 13 (M_{years} = 9). A brief questionnaire was distributed before the perceptual tasks to ensure homogeneity of the

sample. The purpose of the questionnaire was to ensure that the group shared the same characteristics, that is, educational level and socio-economic status since the aim was to eliminate inter-group differences. Based on the data, participants were young adults with a CG linguistic background, and normal speech and hearing. Since we intended to examine the effect of English use on the learners' L2 perceptions, we divided the population into two groups, according to their responses to the questionnaire. The groups consisted of individuals who were using English in their everyday life (n=20) and individuals who were not using English (n=10). Non-probability convenience sampling was used since participants were selected based on their possession on targeted characteristics and availability. Participation was on a completely voluntary basis and students were ensured about the confidentiality of their personal details.

4.2. Stimuli

For the present study, two auditory tasks were used; the first was focusing on open syllables and the second on closed syllables. A total of 120 target syllables for each task were compiled that were arranged in 60 pairs focusing on plosive consonants. 16 pairs were included for each category of plosive consonants, that is, the voiceless/voiced bilabial /p b/, the alveolar /t d/, and the velar /k g/. Distractors were also used and made up 12 of the syllable pairs; specifically, two distractors were used for every ten presentations. These involved fricative and affricate consonants differing in terms of voicing; fricative consonants included the labiodental /f/ and /v/, dental /θ/ and /ð/ as well as alveolar /s/ and /z/; affricate consonants included the palato-alveolar consonants /ʧ/ and /ʤ/. The phonemic contrasts emphasising on plosive consonants in L2 were selected because of their relative difficulty for CG speakers (Kkese, 2016; Kkese & Petinou, 2017a,b). Plosive consonants are present in both the L1 and L2, but their phonetic identifications vary among the two languages. Even though English is widely used on the island (Kkese & Lokhtina, 2017), it is a corollary question to examine perception as a function of sound category and syllabic position.

Concerning the open syllable perception task (CV), plosive consonants were contrasted in syllable-initial and word-medial position. Examples involved the syllables *[paː, aː.**paː**, baː, aː.**baː**, taː, aː.**taː**, daː, aː.**daː**, kaː, aː.**kaː**, gaː, aː.**gaː**]*. For the closed syllable task (CVC), plosive consonants were contrasted in syllable-initial and -final position. Examples involved the closed syllables *[baː**p**, paː**b**, taː**d**, daː**t**, kaː**g**, gaː**k**, **b**aː**p**, **p**aː**b**, **t**aː**d**, **d**aː**t**, **k**aː**g**, **g**aː**k**]*. The selection of the vowel /ɑː/ was because this is the widest

vowel requiring the speaker to concave and lower the tongue allowing a moderately wider vocal tract. In VCV words, the target consonants were embedded in word context instead of syllables since it was impossible to have them in medial intervocalic position in a single syllable. To avoid perceived ambisyllabicity, CV syllables in words were spoken with a stress pattern.

All stimuli were spoken by a monolingual 33-year-old male speaker of Received Pronunciation (RP). Using a bilingual speaker could affect various acoustic parameters such as Voice Onset Time (VOT). With reference to RP, this was used since this accent is free of regional features and is the one typically used with non-native speakers of L2 English even though it is not widely employed in Britain (Romaine, 2000; Trudgill, 1974; Wells, 1982). The speaker was recorded at a 44.1 kHz sample, reading the pairs from a printed list while speaking into a microphone that fed directly into the sound card (IDT High Definition Audio CODEC) of a laptop computer.

4.3. Procedure

Collection of samples took place in one academic semester during class time. The two perceptual tasks focusing on open and closed syllables were distributed in two different sessions with a two weeks interval. First, the background questionnaire and the open syllable perception task were distributed and selected during the same session. With reference to the background questionnaire, its purpose was to ensure homogeneity of the population sample, while it focused on the years of studying English as well as the setting and frequency of English use. After two weeks, the closed syllable perception task was distributed and collected. A small pilot study involving three CG users of L2 English before the perceptual tasks ensured the naturalness and prototypicality of pronunciation. The findings of this small case preliminary study were not included in the results.

Both tasks were administered as a two-alternative force choice task with a circling response mode. The syllables were presented auditorily in two fully randomised blocks to exclude any systematic patterning while the reason syllables were used to eliminate any semantic information from the input (absence of semantic information) as it may have occurred in the case of words or a conversation. Participants had to listen to stimuli two times from the PC loudspeakers and respond accordingly while response order and stimuli were counterbalanced. For example, when they listened to the stimulus [pɑː], they were asked to circle either the written responses <p> or <b>. Also, when listening to closed syllables (e.g., [bɑːp]), they were told

to focus on a specific consonant position each time (e.g., on the final position), since the closed syllables included two plosives. Time interval and the number of the pairs included in the presentation of stimuli helped students not to lose track of the correspondences between the spoken sound and the printed form.

The classroom where the perception tasks were conducted involved a small computer lab. Participants were asked to sit at a separate desk, which supported a computer and listen to each of the tasks through headphones (listening volume was set at 75dB). In this way, factors such as noise, talking to each other or copying each other's answers were prevented and uniform subject performance was ensured. For stimulus extraction and editing, the software package Praat (Boersma & Weenink, 2018) was used. The syllables were presented in an alternating order at an interstimulus interval of two seconds and an inter-trial interval of five seconds in order to limit the effect memory mechanisms could have on the results, particularly as far as auditory sensory memory (Nees, 2016), while no feedback was provided to the participants.

5. Results

The results indicated that the /k-g/, /p-b/, and /t-d/ contrasts in closed syllables had 51.8, 47.8, and 49.2% correct discrimination responses respectively, while the same contrasts in open syllables had 75.6, 76, and 74.9% correct discrimination responses (see Figure 3). Also, in closed syllables, learners who were using English in their daily life correctly discriminated the three contrasts with 48.6% accuracy, while the discrimination accuracy of learners who were not using English correctly was 51.6%. With respect to open syllables, learners who were using English had 75% correct responses, whereas learners who were not using English had 76.6% correct responses (see Figure 4). Also, there were 77.2% correct responses in the onset position for open syllables and 74.9% correct responses in the medial position. In closed syllables, there were 61.9% correct responses in the onset position but only 42.2% correct responses in the coda position.

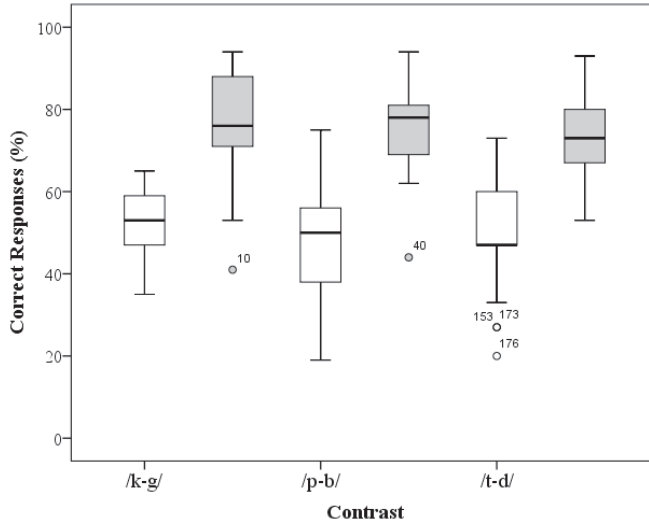


Figure 3: Percentage of correct discrimination responses of the three consonantal contrasts in closed syllables (white) and open syllables (grey)

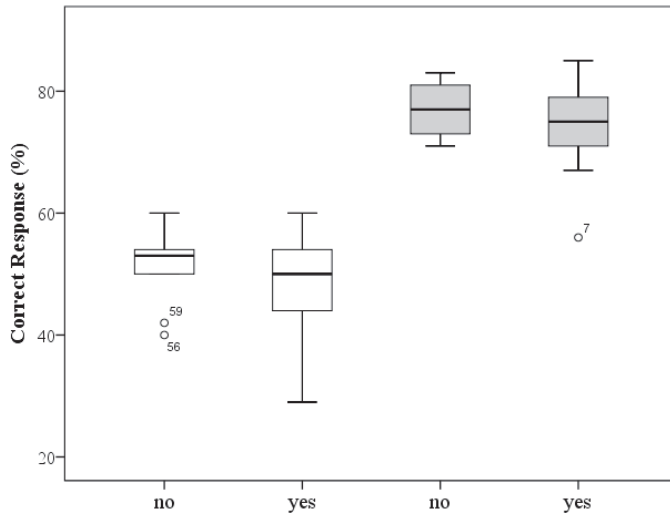


Figure 4: Percentage of correct discrimination responses of consonantal contrasts for learners who were not using and learners who were using English in their daily life. Closed syllables are represented in white and open syllables in grey.

To determine the effect of syllable type, consonantal contrast, use of English, and consonant position on the perception of English consonants, we employed a *binomial logistic mixed effect* model using the *lme4* package in R (Bates et al., 2015). *Response* (correct/incorrect) was the binomial dependent variable, whereas *syllable* (open/close), *contrast* (/b/-/p/, /t/-/d/, and /k/-/g/), *English use* (yes/no), *position.closed* (onset/coda), *position.open* (onset/medial), and *syllable* × *contrast* interaction were the fixed factors. *Subject* was modelled as a random effect.

The following results emerged from the analysis:

i. contrasts did not have significant differences with each other, yielding that they could be equally discriminated,
ii. there was not any contrast × syllable interaction,
iii. use of English did not have any significant effect on the discrimination of syllables in both closed and open syllables,
iv. there was a significant effect of position on consonant discrimination accuracy in closed syllables; this indicates that consonants in onset position could be more accurately discriminated than consonants in coda positions,
v. there was a significant effect of syllable on the discrimination of contrasts, yielding that contrasts in open syllables could be more accurately discriminated than contrasts in closed syllables.

Table 3 presents the results of the mixed model.

Table 3: **Results of the binomial logistic mixed model with fixed and random slopes**

Random effects:					
Groups	Name	Variance	Std.Dev.		
Subject	(Intercept)	6.204e-14	2.491e-07		
Number of obs: 2880, groups: subject, 30					
Fixed effects:					
	Estimate	Std.Error	z value	Pr(>|z|)	
(Intercept)	0.14403	0.10787	1.335	0.182	
contrastp_b	-0.17080	0.12734	-1.341	0.180	
contrastt_d	-0.10624	0.12943	-0.821	0.412	
syllableOpen	1.06548	0.13607	7.830	4.86e-15	***
english.useYes	-0.10481	0.08769	-1.195	0.232	
position.closedCoda	0.11435	0.12405	1.456	5.65e-12	***
position.openMedial	0.12345	0.14568	0.876	0.179	
contrastp_b:syllableOpen	0.19022	0.19573	0.972	0.331	
contrastt_d:syllableOpen	0.06336	0.19807	0.320	0.749	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					

6. Discussion

The present study investigated the perception of contrastive English plosive consonants in both closed and open syllables by CG speakers. Also, the study examined the effect of consonant position on these perceptions. Finally, we assessed how the use of English in the speakers' everyday life shapes their perception patterns with respect to the discrimination of English plosives. The participants of the study completed two-alterative forced-choice discrimination tasks in which they were asked to identify the correct consonant from the auditory stimulus.

The results suggested that all three contrasts in both closed and open syllables could be discriminated in the same manner with poor to moderate discrimination accuracy. This is consistent with previous findings which support the difficulty of CG listeners to perceive English plosives (Kkese, 2016; Kkese & Petinou, 2017a,b). This emerges since the short-long lad phonetic contrast used to distinguish voiced and voiceless plosives in English is used to discriminate single and geminate voiceless plosives in

CG. However, the most significant finding was that consonants in open syllables could be better discriminated than consonants in closed syllables by the L2 learners. Although both CG and English allow closed CVC and open CV syllables, the latter is an *unmarked* feature in CG and, therefore, in SMG in comparison to closed syllables – which is a marked feature – since it is more frequent. This might explain the learners' ease in discriminating consonants found in open rather than closed syllables. Better performance in CV context has also been reported in other studies with L2 listeners (Golestani and Pallier 2007), adults (Redford & Diehl, 1999; Kochetov, 2004), infants (Jusczyk, Goodman, & Baumann, 1999; Zamuner, 2006), and auditory neurophysiology data from animals and humans (Greenberg, 2005). These findings suggest a strong influence of syllable structures in the L1 when it comes to L2 phonetic learning favouring a language-universal preference of CV structure.

With respect to the effect of consonant position in closed and open syllables, it was found that onset consonants in closed syllables could be more accurately discriminated than final consonants. This corroborates earlier findings which support that more production errors are expected for syllable-final than syllable-initial phones (Bent, Bradlow, Smith, 2007; Rogers & Dalby, 2005). Three explanations can be provided for these findings. First, there might be a universal structural asymmetry that favours better discrimination of syllable-initial consonants rather than syllable-final consonants, as it was previously proposed (see Carlisle, 2001; Nam, Goldstein & Saltzman, 2009; Ohala, 1996). Second, this difference might have emerged due to L1 influence. Specifically, the listeners' L1 – Cypriot Greek – does not allow other consonants apart from /s n/ in word-final position; with the exception of some borrowed words. Nonetheless, our stimuli consisted of syllable-final plosives such /k, g, b/ and this can explain, to some extent, the difficulty of listeners when discriminating plosives in the coda position. Similar difficulty was reported for Mandarin learners of English due to L1 restrictions (Rogers & Dalby, 2005). Third, it is believed that plosive consonants in word-initial and –medial positions could be more easily identified compared to consonants in word-final position. This is because in onset position, plosives may provide more salient acoustic information (Kkese, 2016; Kkese & Petinou, 2017a,b). In contrast, plosives – specifically voiced plosives – in word-final positions often lose their voiceness and they are heard as devoiced; thus, final [b] might be pronounced as [p].

Overall, consonants in syllable onset are produced with greater articulatory effort and precision as well as longer duration when compared to the

consonants in syllable coda (Greenberg, 2005). This implies that a language-universal explanation may not be supported since a variety of factors interact for both perception and production. These may involve the acoustic, articulatory, perceptual, and co-articulatory information of each category of phonemes in the onset and coda syllable-position as well as how the interphonology of the two languages, that is the L1 and L2, accommodates the similarities and differences at the syllabic level. Nonetheless for the present study, using VC and CVC stimuli that were clearly articulated in an onset/coda syllable-position without using a semantic/syntactic context could facilitate the predictability of the target plosive consonants while acoustic and articulatory differences could have been perceived with more difficulty.

The results of the discrimination task can also be interpreted from the viewpoint of several speech models. The Native Language Model would suggest that the difficulties in the discrimination of L2 English plosives by CG listeners is attributed to the fact that the English plosives were perceived as similar to L1 "prototypes", that is, phonetic representations of listeners' L1 categories. The Speech Learning Model would suggest that L2 learners merge L2 English voiced and voiceless plosives with a single phonetic category of their L1. This happens because both L1 and L2 phones are located in a common phonological space and the two systems interact with each other. The Perceptual Assimilation Model-L2 would assume that speakers categorise two contrastive English plosives in a single phonological category of their L1 and, therefore, the discrimination of these L2 consonants ranges from poor to moderate. Both the Speech Learning Model and the Perceptual Assimilation Model support that L2 learners should develop new categories for the English plosive consonants in order to accurately discriminate them. This development can be enhanced through further experience in the L2, which can be acquired, for example, through phonetic training, use of the target language, acquisition of qualitative and quantitative L2 input, etc. The present study supplements the three theoretical models of L2 speech perception by emphasising the importance of syllable structure in learning the L2 plosive consonants.

In regard to the effect of English use on the speakers' phone perceptual patterns, the results demonstrated that individuals who use English in their everyday life have similar L2 English plosive discrimination accuracy in both closed and open syllables to individuals who do not use English in their daily communication. We assume that this factor alone cannot contribute to better phone perception. Flege, Frieda and Nozawa (1997) supported that a positive effect of L2 use on the decrease of foreign accent is associated with

age of exposure in the L2. The authors found that Italian-English bilinguals with early exposure in the L2, who were using their L1 more frequently, had a stronger foreign accent speech than the other bilinguals who were rarely using their L2. Piske, MacKay and Flege (2001) found similar results for Italian-English bilinguals with late age of exposure in the L2. However, Piske et al. (2002) added that not only L2 use and age of exposure shape the acquisition of L2 phones, but also L2 input quality and quantity. They stated that Italian/English bilinguals who were frequently using their L1 were likely exposed to more Italian accented English speech compared to the others who were rarely using their L1. The age of formal exposure in English for the participants of the present study varied and possibly these differences affected the results. Therefore, L2 use should be combined with other factors in order to have a positive effect on the L2 learners' perceptual abilities.

7. Conclusions

The study provided novel findings which suggest how L2 English plosives contrasted in voicing are perceived by CG listeners in closed and open syllables. However, we recognise some limitations. For example, we did not take into account the effect of learners' L2 proficiency level or the effect of L2 vocabulary size on the discrimination of L2 English plosives; these factors might be crucial in explaining the listeners' L2 perceptual patterns. Future studies should consider the aforementioned factors and extend this study by investigating other consonants as well beyond plosives. Further, investigating the processing of syllables in different languages could also provide a useful insight to speech perception and production research. The present study emphasised only VC and CVC structures of CG and English but through the theoretical framework, the widely varying structures of the two linguistic varieties has become evident. In this context, cross-language research can help understanding whether a certain effect is the result of acoustic and/or lexical factors.

References

Anderson, J., & Jones, C. (1974). Three theses concerning phonological representations. *Journal of Linguistics, 10,* I-26.

Arvaniti, A. (1999). Illustrations of the IPA: Modern Greek. *Journal of the International Phonetic Association, 19*(2), 167–172.

Arvaniti, A. (2007). Greek Phonetics: The State of the Art. *Journal of Greek Linguistics, 8,* 97–208.

Arvaniti, A. (2010). A (brief) overview of the phonetics and phonology of Cypriot Greek. *Proceedings of the conference on The Greek language in Cyprus from antiquity to today,* University of Athens. http://idiom.ucsd.edu/~arvaniti/ ArvanitiCY.pdf

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1 - 48.

Bent, T., Bradlow, A. R., & Smith, B. L. (2007). Segmental errors in different word positions and their effects on intelligibility of non-native speech: All's well that begins well. In O.-S. Bohn, & M. J. Munro (Eds.), S*econd-language speech learning: The role of language experience in speech perception and production: A festschrift in honour of James E. Flege* (pp. 331-347). Amsterdam: John Benjamins.

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. 171-204). Timonium, MD: York Press.

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: commonalities and complementarities. In O.-S. Bohn, & M. Munro (Eds), *Second-Language Speech Learning: The Role of Language Experience in Speech Perception and Production. A Festschrift in Honour of James E. Flege* (pp. 13-34). Amsterdam: John Benjamins.

Bien, H., Lagemann, L., Dobel, C., & Zwitserlood, P. (2009). Implicit and explicit categorization of speech sounds—dissociating behavioural and neurophysiological data. *European Journal of Neuroscience, 30*, 339–346.

Bien, H., & Zwitserlood, P. (2013). Processing nasals with and without consecutive context phonemes: evidence from explicit categorization and the N100. *Frontiers in Psychology, 4*: 21.

Boersma, P., & Weenink, D. (2018). *Praat: Doing Phonetics by Computer [Computer program] 6.0.37*. http://www.praat.org/

Botinis, A., Christofi, M., Themistocleous, C., & Kyprianou, A. (2004). Duration correlates of stop consonants in CG. *Proceedings, FONETIK 2004, Dept. of Linguistics, Stockholm University*.

Bradley, D. C., Sánchez-Casas, R. M., & García-Albea, J. E. (1993). The status of the syllable in the perception of Spanish and English. *Language and Cognitive Processes, 8*(2), 197–233.

Broselow, E., & Xu, Z. (2004). Differential difficulty in the acquisition of second language phonology. *International Journal of English Studies, 4*, 135–163.

Carlisle, R. S. (2001). Syllable structure universals and second language acquisition. *International Journal of English Studies, 1,* 1–19.

Cebrian, J. (2007). Interaction of L2 phonotactics and L1 syllable structure in L2 vowel Production. *New Sounds 2007: Proceedings of the Fifth International Symposium on the Acquisition of Second Language Speech* (Florianópolis: Federal University of Santa Catarina), 107–113.

Christiansen, M. H., & Chater, N. (2001). Connectionist psycholinguistics: capturing the empirical data. *Trends in Cognitive Sciences, 5*(2), 82–88.

Christophe, A., Peperkamp, S., Pallier, C., Block, E., & Mehler, J. (2004). Phonological phrase boundaries constrain lexical access I. Adult data. *Journal of Memory and Language, 51*(4), 523–547.

Content, A., Kearns, R. K., & Frauenfelder, U. H. (2001). Boundaries versus onsets in syllabic segmentation. *Journal of Memory and Language, 45*, 177–199.

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1986). The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language, 25*, 385–400.

Cutler, A., Dahan, D., & van Donselaar, W. (1997). Prosody in the comprehension of spoken language: a literature review. *Language and Speech, 40*, 141–201.

Cutler, A., McQueen, J. M., Norris, D., & Somejuan, A. (2001). The roll of the silly ball. In E. Dupoux (Ed.), *Language, brain, and cognitive development: Essays in honor of Jacques Mehler* (pp. 181–194). Cambridge, MA: MIT Press.

Duanmu, S. (2000). *The Phonology of Standard Chinese.* Oxford: Oxford University Press.

Dumay, N., Frauenfelder, U. H., & Content, A. (2002). The role of the syllable in lexical segmentation in French: word-spotting data. *Brain and Language, 81*(1-3), 144–161.

Eckman, F. (1981). On the naturalness of interlanguage phonological rules. *Language Learning, 31*(1), 195–216.

Flege, J. E. (1989). Chinese subjects' perception of the word-final English /t//d/ contrast: performance before and after training. *Journal of Acoustical Society of America, 86*, 1684–1697.

Flege, J. E. (1995). Second language speech learning: Theory, findings and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233-277). Baltimore: York Press.

Flege, J. (2003). Assessing constraints on second-language segmental production and perception. In A. Meyer, & N. Schiller (Eds.), *Phonetics*

and Phonology in Language Comprehension and Production, Differences and Similarities (pp. 319–355). Berlin: Mouton de Gruyter.

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

Flege, J. E., Frieda, E. M., & Nozawa, T. (1997). Amount of native-language (L1) use affects the pronunciation of an L2. *Journal of Phonetics, 25*(2), 169–186.

Fung, A. (2004). L2 Phonology of Cantonese Speakers of English: Voicing and Aspiration Contrast of Stops in Onset and Coda. *Doctoral dissertation,* University of Calgary. http://www.scribd.com/doc/18925261/L2-Phonology-of-Cantonese-speakers-of-English-voicingand-aspiration-of-voicing-and-stops-in-onset-and-coda

Georgiou, G. P. (2018). Discrimination of L2 Greek vowel contrasts: Evidence from learners with Arabic L1 background. Speech Communication, 102, 68-77.

Georgiou, G. P. (2019a). 'Bit' and 'beat' are heard as the same: Mapping the vowel perceptual patterns of Greek-English bilingual children. Language Sciences, 72, 1-12.

Georgiou, G. P. (2019b). Cue-reliance and vowel perceptual patterns of Cypriot Greek children who learn English. Journal of Experimental Phonetics, 28, 229-253.

Georgiou, G. P. (2020). Discrimination of Uncategorized-Categorized and Uncategorized-Uncategorized Greek consonantal contrasts by Russian speakers. Topics in Linguistics, 21(1), 74-82.

Georgiou, G. P. & Kkese, E. (2020). Naturalistic exposure to an unknown language alone does not lead to attunement to its phonetic invariants. Journal of Experimental Phonetics, 29, 125-152.

Golestani, N., & Pallier, C. (2007). Anatomical correlates of foreign speech sound production. *Cerebral Cortex, 17*(4), 929–934.

Greenberg, S. (2005). A multi-tier framework for understanding spoken language. In S. Greenberg, & and W. Ainsworth (Eds.), *Listening to Speech: An Auditory Perspective*, pp. 411-33. Hillsdale,NJ: Erlburm.

Hallé, P., Seguí, J., Frauenfelder, U., & Meunier, C. (1998). The processing of illegal consonant clusters: A case of perceptual assimilation? *Journal of Experimental Psychology: Human Perception and Performance, 24*, 592-608.

Hayes-Harb, R., Smith, B. L., Bent, T., & Bradlow, A. R. (2008). The interlanguage speech intelligibility benefit for native speakers of

Mandarin: production and perception of English word-final voicing contrasts. *Journal of Phonetics, 36(4)*, 664–679.

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience, 8*, 393–402.

Hurmalainen, A., Gemmeke, J. F., & Virtanen, T. (2012). Detection, separation and recognition of speech from continuous signals using spectral factorisation. Proceedings of the 20th European Signal Processing Conference (EUSIPCO), Bucharest, 2012, pp. 2649-2653.

Jusczyk, P. W., Goodman, M. B., & Baumann, A. (1999). Nine-month-old's attention to sound similarities in syllables. *Journal of Memory and Language, 40*(1), 62–82.

Kahn, D. (1976) Syllable-based Generalisations in English Phonology. *Doctoral dissertation,* MIT, Cambridge, MA.

Kappa, I. (2002). On the Acquisition of Syllable Structure in Greek. *Journal of Greek Linguistics, 3*, 1–52.

Kkese, E. (2016). Identifying plosives in L2 English: the case of L1 Cypriot Greek speakers. Linguistic Insights: Studies in language and communication, li217. Switzerland: Peter Lang AG.

Kkese, E. & Lokhtina, I. (2017). Insights into the Cypriot-Greek Attitudes toward Multilingualism and Multiculturalism in Cyprus. Journal of Mediterranean Studies, 26(2), 227-246.

Kkese, E., & Petinou, K. (2017a). Perception Abilities of L1 Cypriot Greek Listeners – Types of Errors involving Plosive Consonants in L2 English. Journal of Psycholinguistic Research, 46(1): 1-25.

Kkese, E., & Petinou, K. (2017b). Factors affecting the perception of plosives in second language English by Cypriot-Greek listeners. In E. Babatsouli (ed.), Proceedings of the International Symposium on Monolingual and Bilingual Speech 2017 (pp. 162-167). ISBN: 978-618-82351-1-3. URL: http://ismbs.eu/publications-2017.

Kkese, E., & Karpava, S. (2019). Applying the Native Language Magnet Theory to an L2 setting: Insights into the CGR adult perception of L2 English. In E. Babatsouli (ed.), Proceedings of the International Symposium on Monolingual and Bilingual Speech 2019 (pp. 67-74). ISBN: 978-618-82351-3-7. URL: http://ismbs.eu/publications-2019.

Kkese, E. (2020a). Phonological Awareness and Literacy in L2: Sensitivity to Phonological Awareness and Phoneme-Grapheme Correspondences in L2 English. In G. Neokleous, A. Krulatz, & R. Farrelly (eds.), Handbook of Research on Cultivating Literacy in Diverse and Multilingual Classrooms, 62-81. IGI Global Press.

Kkese, E. (2020a). *L2 writing assessment: The neglected skill of spelling*. Cambridge Scholars Publishing.

Kkese, E. (2020b). Phonological awareness and literacy in L2: Sensitivity to phonological awareness and phoneme-grapheme correspondences in L2 English. In G. Neokleous, A. Krulatz, & R. Farrelly (Eds.), Handbook of research on cultivating literacy in diverse and multilingual classrooms. IGI Global Press.

Kochetov, A. (2004). Perception of place and secondary articulation contrasts in different syllable positions: language-particular and language-independent asymmetries. *Language and Speech, 47*(4), 351–382.

Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. *Nature Reviews Neuroscience, 5*, 831–843.

Ladefoged, P. (2006). *A course in phonetics* (5th edition). Boston: Thomson/Wadsworth.

Langus, A., Marchetto, E., Bion, R., & Nespor, M. (2012). Can prosody be used to discover hierarchical structure in continuous speech? *Journal of Memory and Language, 66*, 285–306.

Lengeris, A. (2009). Perceptual assimilation and L2 learning: Evidence from the perception of southern British English vowels by native speakers of Greek and Japanese. *Phonetica*, *66*(3), 169–187.

Mattys, S. L., White, L., & Melhorn, J. F. (2005). Integration of multiple speech segmentation cues: a hierarchical framework. *Journal of Experimental Psychology*: *General, 134*(4), 477–500.

MacKay, I. R., Flege, J. E., Piske, T., & Schirru, C. (2001). Category restructuring during second-language speech acquisition. *Journal of Acoustical Society of America, 110,* 516–528.

Marslen-Wilson, W. (1987). Functional parallelism in spoken word-recognition. *Cognition, 25,* 71-102.

Massaro, D. W., & Cohen, M. M. (1983). Phonological context in speech perception. *Perception & Psychophysics, 34*(4), 338-348.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18*, 1-86.

McQueen, J. M. (1998). Segmentation of continuous speech using phonotactics. *Journal of Memory and Language, 39*, 21-46.

Mehler, J., Dommergues, J., Frauenfelder, U., & Seglji, J. (1981) The syllable's role in speech segmentation. *Journal of Verbal Learning and Verbal Behavior, 20*, 298-305.

Nam, H., Goldstein, L., & Saltzman, E. (2009). Self-organization of Syllable Structure: A Coupled Oscillator Model. In F. Pellegrino, E. Marsico, I. Chitoran, & C. Coupé (Eds.), *Approaches to Phonological Complexity* (pp. 299–328). Mouton de Gruyter: New York.

Nees, M. A. (2016). Have we forgotten auditory sensory memory? Retention intervals in studies of nonverbal auditory working memory. *Frontiers in psychology, 7,* 1892.

Norris, D., & McQueen, J. M. (2008). Shortlist B: a Bayesian model of continuous speech recognition. *Psychological Review, 115,* 357–395.

Obleser, J., & Eisner, F. (2009). Pre-lexical abstraction of speech in the auditory cortex. *Trends in Cognitive Sciences, 13*(1)*,* 14–19.

Ohala, J. J. (1996). Speech perception is hearing sounds, not tongues. *Journal of Acoustical Society of America, 99,* 1718–1725.

Okalidou, A., Petinou, K., Theodorou, E., & Karasimou, E. (2010). Development of voice onset time in standard-Greek and Cypriot-Greek-speaking preschoolers. *Clinical Linguistics & Phonetics, 24(*7), 503–519.

Piske, T., MacKay, I. R. A., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics, 29*, 191–215.

Piske, T., Flege, J. E., MacKay, I. R. A., & Meador, D. (2002). The production of English vowels by fluent early and late Italian-English bilinguals. Phonetica, 59(1), 49-71.

Pitt, M. A. (1998). Phonological processes and the perception of phonotactically illegal consonant clusters. *Perception & psychophysics, 60*(6), 941-951.

Redford, M. A., & Diehl, R. L. (1999). The relative perceptual distinctiveness of initial and final consonants in CVC syllables. *Journal of the Acoustical Society of America, 106*(3), 1555–1565.

Rogers, C. L., & Dalby, J. (2005). Forced-choice analysis of segmental production by Chinese-accented English speakers. *Journal of Speech, Language and Hearing Research, 48*(2), 306–322.

Romaine, S. (2000). *Language in society: An introduction to Sociolinguistics* (2nd ed). Oxford: Oxford University Press.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8month-old infants. *Science, 274*, 1926–1928

Segui, J., Frauenfelder, U., & Mehler, J. (1981). Phoneme monitoring, syllable monitoring and lexical access. *British Journal of Psychology, 72*(4), 471-477.

Stockman, I. J., & Pluut, E. (1999). Segment composition as a factor in the syllabification errors of second-language speakers. *Language Learning, 49,* 185–209.

Savin, H. B., & Bever, T. G. (1970). The nonperceptual reality of the phoneme. *Journal of Verbal Learning & Verbal Behavior, 9*(3), 295–302.

Trudgill, P. (1974). *The Social Differentiation of English in Norwich.* Cambridge: Cambridge University Press.

Tsukada, K., Birdsong, D., Mack, M., Sung, H. Y., Bialystok, E., & Flege, J. (2004). Release bursts in English word-final voiceless stops produced by native English and Korean adults and children. *Phonetica, 61*, 67–83.

Tyler, M., & Cutler, A. (2009). Cross-language differences in cue use for speech segmentation. *Journal of Acoustical Society of America, 126*, 367–376.

Vroomen, J., & Gelder (de), B. (1999). Lexical access of resyllabified words: evidence from phoneme monitoring. *Memory and Cognition 27*(3), 413-421.

Wells, J. C. (1982). *Accents of English: Volume 3: Beyond the British Isles*. Cambridge University Press.

Zamuner, T. S. (2006). Sensitivity to word-final phonotactics in 9- to 16-month-old infants. *Infancy 10*, 77–95.

Zwitserlood, P. (2003). The internal structure of words: Consequences for listening and speaking. In N. O. Schiller, & A. S. Meyer (Eds.), *Phonetics and phonology in language comprehension and production* (pp. 79-114). Berlin, Germany: Walter de Gruyter GmbH & Co.

Zwitserlood, P. (2004). Sublexical and morphological information in speech processing. *Brain and Language, 90,* 368–377.

Chapter Ten

Factors Affecting Accent Acquisition among Female American Immigrants to Israel

Efrat Last, Shira Matalon, Nitzan Bechar Marom, Noa Vittenberg, Sapir Shemesh and Noam Amir

Abstract

This study explored accent acquisition among American females who immigrated to Israel at various ages. Two control groups were assembled, composed of nine native Hebrew speakers and nine very recent immigrants. The study group was comprised of 30 American women between the ages of 17-42, who had arrived in Israel between the ages of 2-21. Members of the control group answered questions detailing the factors which may have affected their accent acquisition, such as age of arrival, motivation, percentage of time speaking Hebrew and several more. The subjects were recorded uttering various speech materials, including written sentences, a passage, and a free dialogue. The accents were rated perceptually by a panel of judges, and the speech material was analyzed acoustically. The perceptual results demonstrated the presence of a critical/sensitive period for acquiring a native accent, that was found to be between ages 6-12, although outliers were found on both ends. The "age of arrival" and the "use of Hebrew with friends" variables were the main variables that predicted accent acquisition. The acoustic features best correlated with accent were F2 of /u/ and /e/, F1and F2 of /e/, and the VOT of unvoiced plosives. Finally, we found that immigrants had difficulties in acquiring the native /l/, as compared to /r/. Further research should examine the implementation of a vowel-based accent learning program which would be available to new immigrants to Israel.

Keywords: Second language acquisition, foreign accent, Hebrew, L2, immigration, American English

1. Introduction

An accent, by definition, contains all the elements of prosodic and phonetic precision required to convey meaning, as well as being a "fluid, contextualized expression of our personal and social identity" (Moyer, 2013: 10). An accent contains the sounds, melodies and rhythms of speech, however, beyond the acoustics of speech it bears much symbolic value as it conveys much more than the literal meaning, whether the speaker likes it or not (Moyer, 2013). Acquisition of accent can have a strong effect on social integration of immigrants in a new society. It has been found that listeners generally disfavour nonstandard accents (Gluszek & Dovidio, 2010). Furthermore, speakers with a foreign accent may be judged as less competent, less intelligent, less educated and less trustworthy than L1 (native language) accents in certain contexts (Gluszek & Dovidio, 2010; Lev-Ari & Keysar, 2010) and have been shown to suffer negative legal, monetary, political and social consequences (Pantos & Perkins, 2013). Having a foreign accent can also influence one's self-perception, which in turn can affect the adjustment to a new environment and society (Dewaele & McCloskey, 2014).

For over forty years, foreign accent has been the focus of research in several related fields, such as linguistics, speech development and neurology. The classic articles by Asher & Garcia (1969) and Scovel (1969) were amongst the first to shed light on the question of accent acquisition and the optimal age to learn a second language. Since then, many studies have explored the phenomenon of perceived foreign accent (Piske, MacKay & Flege, 2001) and have yielded different and often opposing results. Several factors have been suggested to explain the ability to acquire a native accent when learning an L2 (a second language). The most prominent of them is Age of Arrival (AoA). This is the age at which the second language is learned and appears to be the most discussed factor. Four different patterns have been observed in the literature, relating AoA and accent acquisition: "critical/sensitive period", "linear decline", "sigmoid decline" and "constant ability" (Munro & Mann, 2005).

The "*critical/sensitive period*" is linked to the critical period hypothesis (CPH), which states that there are maturational constraints on the time during which an L1 can be acquired (Lenneberg, Chomsky & Marx, 1967; Scovel, 1969). Lenneberg et al. (1967) further claimed that if a language is

acquired after puberty, some aspects of it can be learned, however, full mastery cannot be achieved. There have been many findings regarding the critical or sensitive turning points. The lowest age of the range where a native accent would no longer be fully acquired was found to be 3 years, whilst the oldest age at which a native accent could still be acquired was found to be 15 (Flege & Fletcher, 1992; Flege et al., 1995; Thompson, 1991).

The "*linear decline*" pattern suggests a single, monotonic, negative slope implying a diminishing ability to acquire the L2 throughout the entire human lifespan (Munro & Mann, 2005).

The "*sigmoid decline*" refers to a progressively shifting slope and can be described as having a brief initial period when the L2 accent will be native, a longer period of linear decreasing of the native L2 accent and a final period of maximal deflection from the native accent (Munro & Mann, 2005). "*Constant ability*" reflects a rather surprising set of results, implying an equal ability to acquire the L2 accent at any age. That is, there is no real difference between younger and older ages of arrival, and should such a difference exist, it would be attributed to possible input differences and not to learner ability (Bongaerts, 1999; Bongaerts, van Summeren, Planken & Schils, 1997; Moyer, 1999).

As to the reasons for why the age of arrival would have effects on the ability to acquire a native accent, the traditional explanation is biologically based and suggests maturational constraints such as loss of neural plasticity which causes the mechanisms required to produce the L2 not to operate as effectively (Lennenberg, 1967; Bongaerts et al., 1997; Flege, Schirru, & Mackay, 2003). A second explanation relates to the reduced amount of phonetic input when acquiring a second language at a later stage (Flege et al., 2003) whilst a third explanation suggests that the L1 and L2 systems influence each other more the later the L2 is acquired (Flege et al., 2003; Ng, Chen & Sadaka, 2008; Piske et al., 2001). Therefore, the Speech Learning Model (SLM) proposes that the capacity to learn a new language remains the same throughout the life span (Flege et al., 2002). The model further explains that "the phonetic elements making up the L1and L2 phonetic subsystems of a bilingual exist in a "common phonological place", and so will necessarily influence one another" (Flege 1995, 1999, 2002).

Several further factors appearing in the literature are as follows:

- Length of Residence (LoR): The LoR refers to the number of years spent living in a place where the predominant language is the L2. There have been discrepancies in the findings across studies regarding the importance of the LoR in terms of pronunciation accuracy, in part since LoR only provides a rough index of the overall L2 experience as shown in a meta-analysis by Piske et al. (2001).
- Formal instruction: This has also been investigated as a possible predictor of L2 pronunciation. In the field of language instruction, the role of pronunciation has varied from being the focus in audio-lingual programmes to having no role in more traditional grammar-based methods (Pardede, 2010). Moyer (1999) found that adults who had received suprasegmental (especially intonation and stress) and segmental training in English obtained ratings that were closer to the range of accent ratings obtained for native speakers. This led her to propose phonological instruction, specifically on sentence and discourse levels, in foreign language learning programs. This corresponds with Bongaerts' (1999) claim that one of the main factors that contributed to the success of certain individuals in attaining a native-like accent was intensive training in L2 perception and L2 production.
- Ability to mimic: Another interesting angle that has been researched is the speaker's personal ability to *mimic* unfamiliar speech sounds. This ability has been suggested as an important learning mechanism, which can provide a basis for acquiring languages and musical systems (Reiterer, Hu, Sumathi & Singh, 2013). Thompson (1991) found that the ability to mimic facilitated the acquisition of accurate pronunciation. The ability to mimic has also been found to be a good tool in pronunciation training programs. Higher success rates in achieving native-like accents were achieved when training to mimic people with a native accent as well as training to mimic one's own voice digitally adjusted to have a native accent (Felps, Bortfeld & Gutierrez-Osuna, 2009).
- Motivation: Many studies have found *motivation* to have some influence on the degree of foreign accent (Flege, Munro & MacKay, 1995; Flege, 1999; Granena & Long, 2012; Moyer, 1999; Polat, 2011). In the different studies, motivation was defined in relation to different aspects, such as "integrative motivation", "concern for L2 pronunciation" and "instrumental/professional motivation". For example, Moyer (1999) found that professional motivation accounted for 41% of the variance regarding accent acquisition.

- Language use: The investigation of *language use*, which is defined as the amount of time spent speaking the L2 with native speakers, as a predictor of foreign accent has yielded contradicting results. Some studies (Flege & Fletcher, 1992; Thompson, 1991) have not found a significant correlation between the two while others (Flege et al., 1995; Flege et al., 1999) found that it accounted for a small percentage (8%) of the variance.

1.1 Foreign accent research methodologies

It is evident that the many different researchers have arrived at a wide range of findings and drawn different and occasionally contradicting conclusions. This could be attributed to the differences in research methodologies, such as subject population, elicitation methods and rating techniques (Piske et al., 2001).

1.1.1 Subject population

Regarding *subject population*, the differences include individual external factors such as variation in age, different target languages (L2) and native languages (L1), as well as individual internal factors such as language aptitude and motivation. Most studies have examined English as the target language (Piske et al., 2001), while the native languages spoken by the participants have been more diverse, for example, Dutch (Pinget, Bosker, Quené & De Jong, 2014), Turkish (Ng et al., 2008; Pinget et al., 2014), Kurdish (Polat, 2011), Korean and Chinese (Fledge et al., 1999; Huang & Jun, 2015) and others. Each of these languages has its own unique accent which could affect the accent acquisition differently, due to the phonetic elements making up the L1 and L2 influencing one another (Flege 1995, 1999, 2002). Investigating other target languages could possibly shed more light on the importance of the individual internal factors versus the external ones as it may highlight one or more common factors which lead to successful native accent acquisition regardless of what the L1 or L2 are. The current study has indeed chosen a different target language (Hebrew), whilst the L1 is English.

1.1.2 Elicitation techniques

Elicitation techniques have also varied in different studies including uttering words (Baker, 2010), sentences (Asher & Garcia, 1969), words and sentences (Arslan & Hansen, 1997), direct repetition techniques (Piske et al., 2001), and story retelling (Munoz, 2014). Thompson (1991) found that

read speech was judged as having more of a foreign accent than spontaneous speech. This has been explained as being caused by the differences in amount of formal education that immigrants receive in the L2, leading to a lower level of reading ability (Piske et al., 2001). Conversely, spontaneous speech may be judged to be stronger foreign accented that read speech because of grammatical errors that the speaker may make. Munro and Mann (2005) found that words and sentences were judged as less foreign accented in comparison to paragraphs and picture descriptions. Therefore, on the one hand, read material has the advantage of less grammatical errors but the disadvantage of being dependant on reading ability on the other. Spontaneous speech has the advantage of being able to choose the words one is comfortable using, however, may be more prone to grammatical errors. It would seem that when asked to rate an accent, the grammatical errors would play a lesser role in the decision, whilst the level of the speaker's comfort when articulating "self-chosen" would play a greater role. The contradicting results make it difficult to determine which would be the stronger factor relating to the accent rating.

1.1.3 Rating techniques – listener evaluation

Rating techniques have varied and developed over the years. The most common technique cited is listener evaluation (Asher & Garcia, 1969; Baker, 2010; Derwing & Munro, 2013; Piske et al., 2001). It is important to note that even when listener evaluation is used, there can be many differences such as number of listeners, listener experience (naive or professional) and evaluation scales. Despite these differences, the use of listener evaluation is the most widely used form of accent rating, which may be because the listeners can subconsciously integrate all the acoustic factors (whether temporal or spectral) in a way that is difficult to implement in acoustic analysis.

1.1.4 Acoustic and phonological analysis

Acoustic analysis is another form of evaluation used to assess accent accuracy (Barker & Turner, 2015; Felps et al., 2009; Floccia, Butler, Girard & Goslin, 2009; Lee & Iverson, 2012; Netelenbos, Li & Rosen, 2015; Ng et al., 2008; Poretta & Tucker, 2012; Porretta, Kyröläinen & Tucker, 2015). A foreign accent can be defined as "deviations from the expected acoustic (e.g. formants) and prosodic (e.g. intonation, duration, and rate) norms of a language" (Felps et al., 2009: 922). Several features of the speech sample can be chosen for analysis. These include temporal features (word-final stop closure duration, voice onset time (VOT), average voicing duration and

average word duration), frequency characteristics (analysis of the formant frequencies), and intonation, which is a more temporal property despite it being a measure of fundamental frequency. The focus of the current study is primarily on vowel analysis.

The number and types of vowels in different languages' phonemic inventory differ greatly, which raises many questions when researching L2 accent acquisition (Zampini, 2008). Vowels may be classified by the location of the tongue, i.e. in the front or back of the mouth, by the height of the tongue within the mouth and by lip rounding. In general, non-native speakers produce different tongue movements compared to native speakers. This is because they "automatically produce different sounds based on learned tongue movements of their native language." (Arslan & Hansen, 1997: 15). It is widely accepted that the information necessary for vowel recognition is mainly carried in the first two formants, F1 and F2 (Amir, Tzenker, Amir & Rosenhouse, 2012; Hillenbrand, Getty, Clark & Wheeler, 1995; Pols, Van der Kamp, & Plomp, 1969). Arslan and Hansen (1997) interestingly add that F2 and F3 carry more important information when investigating accent differences due to the different, non-native tongue movements.

Concerning the vowel systems examined in the present study, in American English (AE) there are 11 monophthong vowels that can be represented by /i, I, e, ɛ, æ, ʌ, u, ʊ, o, ɔ, a/ using the International Phonetic Alphabet (IPA). The Israeli Hebrew vowel system consists of only five vowels: /i, e, a, o, u/, with no central "schwa" vowel (Amir et al., 2012; Most, Amir & Tobin, 2000).

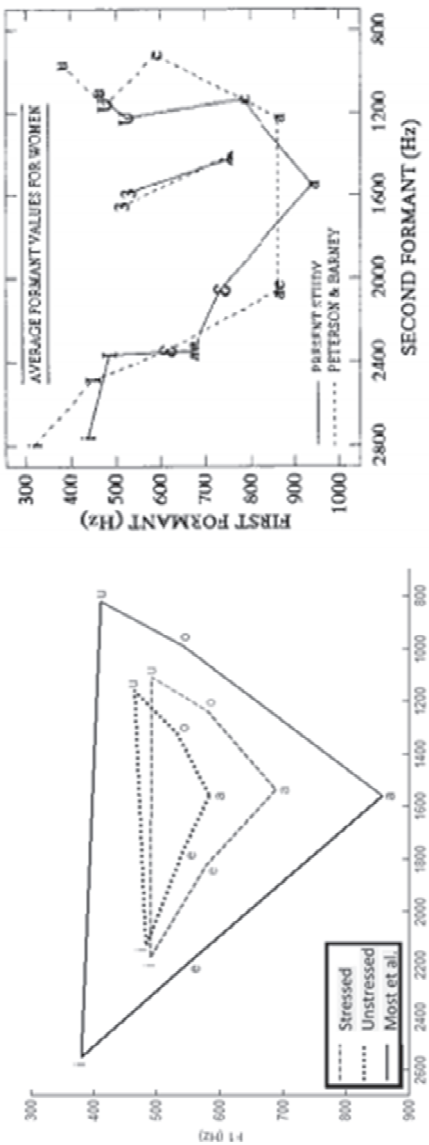


Figure 1. Comparison of Hebrew formant values (from Amir et al., 2012) and American English formant values (from Hillebrand, Getty, Clark & Wheeler, 1995).

Figure 2 highlights the differences between the Hebrew and English vowel systems in the formant values of each vowel and in the number of vowels, indicating that major differences between the two accents may be found in the vowel qualities. Indeed, studies focussing on vowel characteristics have yielded interesting results. For example, Floccia et al. (2009) found that the space between formants differed between accents; Lee and Iverson (2012) as well as Lee, Guion and Harada (2006) found significant differences between F1 and F2 when compared to native norms. Ng et al. (2008) found significant differences between the F1 and F2 of close vowels (vowels that are produced in the L1 and L2) as well as in distant vowels (vowels that are only produced in the L2). However, the Euclidean distance that was used to measure the separation between the corresponding vowels was not found to be significantly different. These results provide evidence of the importance of the spectral qualities of vowels in differentiating accents across different languages.

Previous studies indicate that differences in VOT can also be found between Hebrew and English. Raphael et al. (1995) found that for Hebrew speakers, the mean VOT duration of /p/ was 28.5ms whereas for AE speakers, this was found to be around 51ms (Chodroff & Wilson, 2017; Schoonmaker-Gates, 2015; Lisker & Abramson, 1964). Furthermore, studies have shown that VOT can influence the perception of accent (Gonzalez-Bueno, 1997). As opposed to vowels, laymen are not usually acquainted with this parameter, therefore it is of interest to examine whether native-like values are acquired by immigrants without being aware of it.

1.1.5 Relating acoustic parameters and accent ratings

The studies in this area appear to be few and far between, though their more consistent results highlight the importance of vowels to the accent ratings. These studies have focused on the acoustic parameters that would assist in predicting/rating a foreign accent. Porretta and Tucker (2012) found that F1 values that are far from native speaker norms may provide a strong cue that a speaker has a foreign accent. Mccullough and Clopper (2016) researched the acoustic information that assisted listeners in classifying different groups of accents in English based on different L1 backgrounds. They suggested that the deviation from native norms in temporal and spectral properties of vowels, as well as in VOT, predicted the listeners' classification of accent. Chan, Hall and Assgari (2016) manipulated formant values as well as vowel duration to determine their role in the perception of a foreign accent. They too found that vowel formant frequencies are important cues to foreign accent. As mentioned above, the relative consistency of the

results presented regarding the importance of spectral qualities when predicting or rating a foreign accent may provide the key to achieving more unified results in this field.

1.2 Rationale and objectives

We live in a world where contact between L1 and L2 speakers is growing rapidly due to increased worldwide migration and corporate globalisation (Pantos & Perkins, 2013). Israel is a centre for both phenomena: it is well known for constantly welcoming new immigrants from around the world as well being a global high-tech centre. Therefore, the question of L2 acquisition and particularly accent acquisition, is of great value and consequence to Israeli society, specifically regarding immigrants who come to settle in Israel permanently. The immigrants to Israel vary greatly in their AoA, ranging from in-utero to elderly. They also vary in many of the other factors outlined above, namely, LoR, language use, personal aptitude and others. Despite the obvious importance of this topic to Israeli society, insufficient research has been devoted to exploring its many aspects. Better understanding of the factors that affect accent acquisition amongst immigrants to Israel may allow us to enable a smoother transition into Israeli society.

The purpose of this study is to shed light on the factors affecting accent acquisition amongst American females who immigrated to Israel at various ages. This will be done from two perspectives – perceptual and acoustic, which will provide a wide range of information regarding the topic at hand. The study aims to incorporate the various factors affecting accent acquisition, along with the more dominant AoA factor, to more fully understand the individual influences which may play a part in the attainment of a native accent. Furthermore, the study aims to map out the acoustic information upon which a listener may determine the degree of the foreign accent. It is important to note, however, that only a subset of acoustic properties will be investigated. For vowels, the focus will be on their spectral qualities, i.e. formants. Unvoiced plosives will be examined for VOT.

2. Method

2.1 Participants

The participants in the study group were 30 American-born females of a similar socioeconomic status, most living in large Israeli cities, between the ages of 18-44; their L1 was AE. None of the participants reported hearing loss or speech impediments. We chose to study a single gender to avoid dealing with the formant frequency differences between males and females. Due to the many variables involved (namely AoA, LoR, etc), we chose this to be a group of adult women who had immigrated to Israel at different ages. This enables correlation analyses rather than the comparison of two groups – one which had arrived at an early age and one which had arrived at a later age. This is a widely accepted method in this field of research (Flege, 1991; Moyer, 1999; Piske et al., 2001; Thompson, 1991). Table 1 presents mean values for several background variables: Age, AoA, LoR, and Years of Education in Hebrew (YeH).

Table 1. Test group demographics: averages, SDs, ranges

	N	Min	Max	Average	SD
Age	30	18	44	29.58	7.12
AoA	30	2	21	12.33	5.69
LoR	30	6	38	16.38	7.39
YeH	30	1	16	8.18	4.77

The control groups consisted of nine Israeli females whose L1 is Hebrew and nine American females who had arrived in Israel within the last year and therefore had a strong and undiminished American accent, of a similar age range and socioeconomic status as the test group.

The participants volunteered to participate in the research and were recruited through word of mouth as well as social networks. The research design and the questionnaires received approval from the institutional review board.

2.2 Procedure

2.2.1 Recording Procedure and Speech Materials

The subjects were tested individually and recorded in a quiet room. Recordings were carried out using a high-quality head-mounted microphone

(Audio Technica AT892), with an external soundcard (Centrance Micport Pro) using Audacity software (Audacity Team, 2000, Version 1.2.5) on a laptop (Lenovo G550). They were recorded at a sampling rate of 44.1KHz, 16 bits per sample.

The speech materials recorded by the test group included the following:

Sentence reading- Two sets of sentences were read by the subjects. One set included five /pVp/ nonsense words (Most et al., 2000). The other included 15 sentences with real words containing the Hebrew vowels- 3 words per vowel. In total, 30 sentences were recorded (Appendix 1a).

Passage reading- The subjects were asked to read a passage containing the vowels and consonants of the Hebrew language (from Amir & Levine-Yundof, 2013) (Appendix 1b).

Spontaneous speech- The subjects were asked to describe their activities on the day of the recording and were encouraged to speak for about one minute.

The Israeli and American control groups recorded only the 15 non-word sentences and the 15 real word sentences.

2.2.3 Questionnaire

The 30 participants in the test group filled out a questionnaire regarding different factors, which may have affected their accent acquisition (Piske et al., 2001). These questions include:

1. Age of arrival to Israel (AoA)
2. Length of residence in Israel (LoR)
3. Years of Education in Hebrew (YeH)
4. Ability to mimic (MiM)- Subjects assessed their ability to imitate speech sounds on a 5-point scale (based on Thompson, 1991).
5. Motivation (MoT). Subjects rated the importance to themselves of having a good accent on a 5-point scale ranging from “very important” to “not important at all”.
6. Use of Hebrew – The subjects estimated the percent of time they speak Hebrew at work (PHW), at home (PHH), and with friends (PHF) by circling one of the following percentages: 0%, 10%, 30%, 50%, 70%, 90% and 100%.
7. Self-rating - The participants were also asked to self-rate their accent on a 5-point scale, ranging from 1- “not great” to 5- “excellent”.

2.2.4 Perceptual test - listener rating of accent

Ten recordings were taken from each participant, nine read sentences and one sentence extracted from the spontaneous conversation. Five sentences were extracted from two members of each control group, providing ten sentences from each control group. In total, 320 sentences were presented, ten from each participant and ten from each control group.

A group of judges was recruited for the study consisting of ten female Speech and Language Therapists. All the judges were 30-35 years old, born in Israel, whose mother tongue is Hebrew and with no known hearing or speech and language disorders. These women volunteered to participate in the research and were recruited through word of mouth as well as social networks.

The judges listened to and rated each recording on a Lenovo Yoga 3 or Dell Vostro 5470 laptop, using custom software created for this study running under MATLAB (version 7.10, the Mathworks). After hearing each of the 320 sentences using supra-aural earphones, model JBL E30, each judge was asked to rate the speaker's accent on a scale of 1-4: 1-Israeli accent 2-weak American accent 3-Medium American accent 4- Strong American accent. Reliability between the judges' ratings on the listening materials was tested by calculating Cronbach's alpha, which indicated a high reliability: 0.943 (n=10).

2.2.5 Acoustic measures

The first two formant frequencies (F1 and F2) were calculated for the target vowels in the non-word and meaningful-word utterances for the test group and both control groups. A total of 1,419 vowels was analysed – 263 of the Israeli control group, 266 of the American control group and 890 of the test group. This was carried out by first annotating these vowels using PRAAT software (version 6.0.2.8, Boersma & Weenik), then calculating the formant frequencies with custom written Matlab software, using the LPC (Linear Prediction Coding) algorithm. Formant values were first calculated automatically for all vowels, after down sampling to 11 KHz, and using a filter order of 10. The results were then verified manually using a custom written Graphic User Interface (GUI) written also in Matlab. Wherever a visual disagreement was found between the calculated formant tracks and the spectrogram, filter order was changed manually to obtain the best agreement.

Several further features were derived from the formant frequencies to be explored for correlation with the AoA and with the listener ratings:

a. F1/F2 polygon area: This is simply the area of the polygon outlined by the five vowels. Two values were calculated per speaker, once using the mean formant values for nonsense words, and once using the mean formant values for meaningful words.
b. Euclidean distances: calculated for each vowel of each speaker, twice: once for nonsense words and once for meaningful words. This is the Euclidean distance in the F1/F2 plane between the point representing the uttered vowel, and the point obtained from the mean of the American control group.

Finally, VOT for the first /p/ in each of the nonsense /pVp/ words was measured manually, also using Praat software. These values were averaged over the three repetitions of the same /pVp/ word and then analyzed separately for each of the five vowels.

2.2.6 Perceptually assessed acoustic features

Two consonants that are notoriously different in Hebrew and English, even to the untrained ear, and known to be challenging for immigrants learning Hebrew are /l/ and /r/. Therefore, three instances each of /l/ and /r/ in different positions (beginning, middle and end of word) were tagged and rated perceptually by an expert speech therapist (Author EL) on a scale of 1 (native Israeli) – 3 (American). Since these consonants, especially the rhotic, are well known characteristics of an American accent in Hebrew, it is possible that immigrants would invest a large effort in acquiring them in order to integrate better into society.

3. Results

3.1 Perceptual Analysis

Perceptual ratings of accent (which was rated on a scale of 1-4) were first analyzed for the two control groups, as summarized in Table 2. For the Israeli control group, the average accent rating was 1.48 and for the American control group, the average accent rating was 3.94, as expected. For the study group, the average accent evaluation was 2.4, falling between the two control groups.

A one-way between-subjects ANOVA was conducted to compare the three groups (the study group and both control groups). A significant main effect was found ($F(2,317) = 30.22$, $p<0.001$). Contrast analysis revealed significant differences between each group and the two others ($p<0.05$).

Table 2. Min, Max, Mean and SD of the accent ratings for the control groups and the study groups

Speakers	N	Min.	Max.	Mean	SD
Native Israeli control group	9	1	3	1.48	.56
Study group	30	1	4	2.40	.97
American control group	9	3	4	3.94	.24

Pearson correlation between the perceptual accent ratings and the AoA was $r=0.616$ ($p<0.001$). A scatter plot of the accent ratings and AoA of the test group participants is presented in Figure 2.

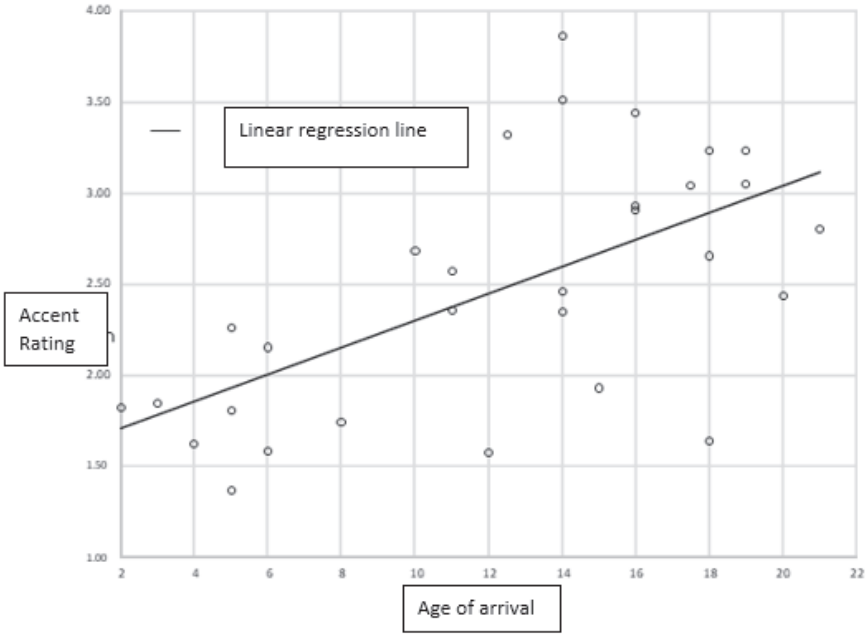


Figure 2. Linear regression line depicting the relationship between the AoA and the accent rating.

Interestingly, the speakers' self-rating was found to have a high Pearson correlation with the accent ratings ($r=-0.744$, $p<0.001$ respectively). The Pearson correlation is negative because the listeners gave more points for a stronger American accent whilst the self-ratings gave higher scores for a more Israeli accent.

To further examine the relationship of the AoA to the accent ratings, two additional variables were calculated:

1. "AoA group" – which divides the study group alone into sub-groups based on their AoA. The first sub-group included speakers whose AoA was 1-3, the second included speakers whose AoA was 4-6 and so on. The ranges were chosen both by examining the data as well as based on previous research which suggested that the critical age may appear at age 6 or 12 (Long, 2005), or ranges between 3-15 (Flege, Munro & MacKay, 1995).

2. "Accent group" – which divides the study group into sub-groups based on their accent rating. The group with the strongest Israeli accent included the speakers whose accent rating was between 1-1.5; the next group included speakers whose accent rating was 1.5-2, and so on. These ranges were chosen by examining the data, and by the fact that the average accent rating within the Israeli control group was around 1.5 and that 90% of the native Israelis were given an accent rating between 1-2. Therefore, the two accent groups "1-1.5" and "1.5-2" will be considered as having an Israeli accent.

Table 3. Number of participants in each group (AoA and accent rating).

		Accent group						
		1-1.5	1.5-2	2-2.5	2.5-3	3-3.5	3.5-4	**Total**
AoA group	1-3	0	2	0	0	0	0	**2**
	4-6	1	3	2	0	0	0	**6**
	7-9	0	1	0	0	0	0	**1**
	10-12	0	1	1	2	0	0	**4**
	13-15	0	1	2	0	1	2	**6**
	16-18	0	1	0	3	3	0	7
	19-21	0	0	1	1	2	0	**4**
Total		**1**	**9**	**6**	**6**	**6**	**2**	**30**

In Table 3, we can see the following:

- Out of the 10 immigrants in both Israeli accent groups, 8 immigrated prior to age 12 and only 2 thereafter.
- Out of the 17 immigrants who arrived after the age of 12, only 2 were found to have an Israeli accent. 8 speakers from this group were in the "3-4" rating group, which indicates a strong American accent.
- Out of the 12 immigrants who arrived prior to age 12, 8 were in the Israeli accent group and none of them were in the "3-4" rating group.

Table 4 explores the relation between age and accent; specifically, age of 12 as the "critical age". The study group was split into two groups: one of

those whose AoA was under 12 (n=13) and the other whose AoA was over 12 (n=17). As we can see, the Pearson correlation of the accent ratings and the AoA was not significant in either group. The only significant Pearson correlation found was with the subjects' self-rating of their accent in both groups: AoA 1-12 (r=-0.766, p=0.002) and AoA 13-21 (r=-0.609, p=0.009). This finding was predictable considering the aforementioned high Pearson correlation found between the accent ratings and the self-ratings.

Table 4. Pearson correlations between accent rating and background variables for those who arrived before and after age 12.

Background variable	AoA before 12 (N=13)		AoA after 12 (N=17)	
	R	P	r	P
AoA	0.447	0.126	-0.194	0.455
YEH	-0.404	0.171	-0.180	0.489
LOR	-0.232	0.126	-0.081	0.757
MOT	-0.186	0.544	-0.324	0.205
PHW	-0.070	0.839	-0.102	0.717
PHH	-0.210	0.491	-0.459	0.064
PHF	-0.428	0.145	-0.440	0.077
Self rating	-0.766*	0.002	-0.609*	0.009

These findings may indicate a non-linear CPH (critical period hypothesis) pattern, and therefore, a non-linear trend line was calculated using Least Squares Fit (as seen in Figure 8) between the AoA and sigmoid (S-curve), as described by equation (1). A significant correlation was found (r=0.667, p<0.001) indicating a non-linear pattern of accent acquisition. It is also noteworthy that for this regression, the coefficient obtained (0.667) was slightly higher than for the linear correlation (0.616).

$$y = 1.777 + \frac{1.064}{1 + e^{-0.785(x-10.073)}}$$

In this formula, x is the AoA and y is the perceived foreign accent for each speaker.

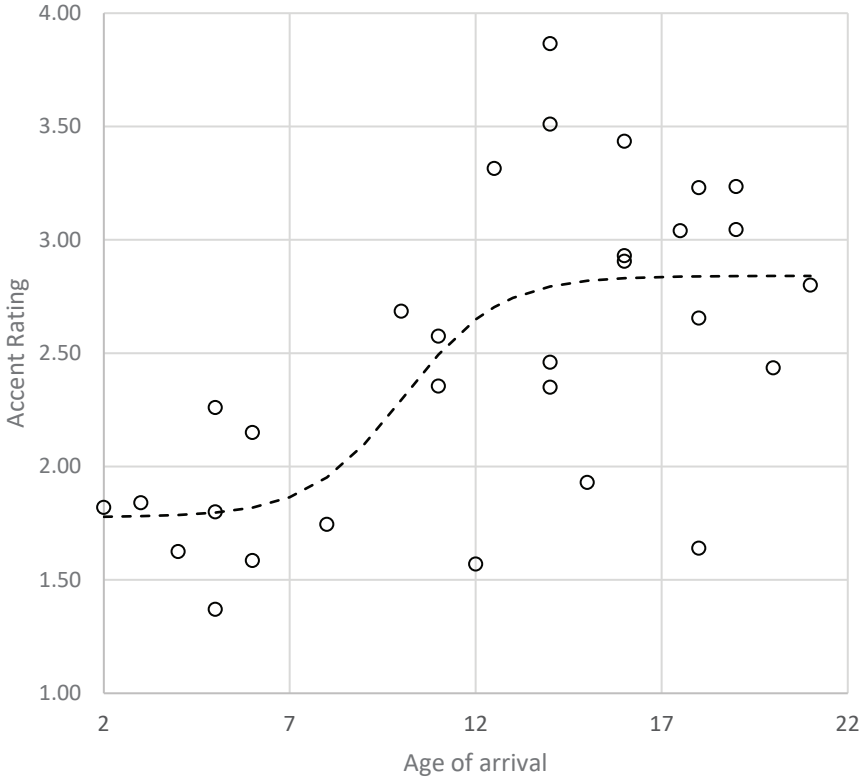


Figure 3. Non-linear regression line depicting the relationship between the AoA and the accent rating.

3.1.1 The influence of additional factors

A stepwise regression model was used to predict the average accent rating based on all the independent variables. The model received included only the AoA and PHF variables. The AoA alone explained 45% of the variance ($F(1,24) = 19.832$, $p<0.001$) and when adding the PHF variable, the model explained 56% of the variance ($F(2,23) = 14.316$, $p<0.001$). A scatter plot relating accent ratings and PHF, along with a regression line, is presented in Figure 4.

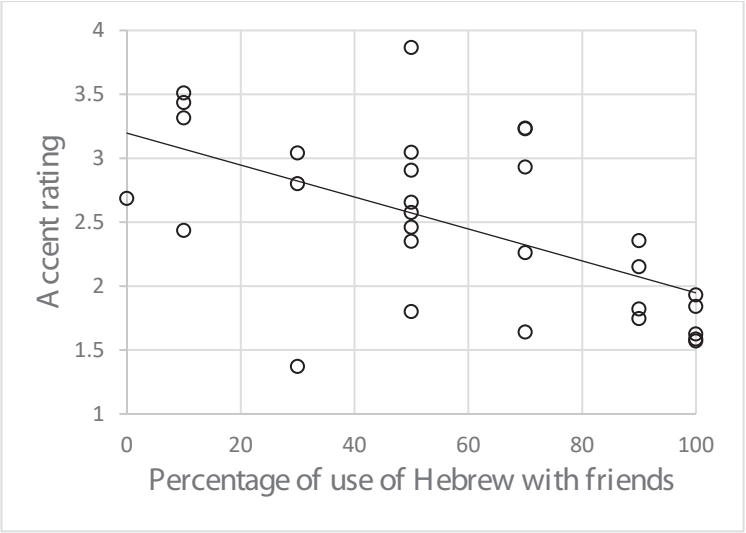


Figure 4. Scatter plot of accent ratings and PHF.

3.1.2 Outliers

Due to the small number of outliers found (n=2), we are only able to describe the findings qualitatively. The first outlier had an AoA of 15 and had an average accent rating of 1.93, indicating a good L2 accent. Her motivation score was 5 (the highest option) and her percentage of Hebrew use variable was high – 90% at work (PHW), 70% at home (PHH) and 100% with friends (PHF). The second outlier had an AoA of 18 and had an average accent rating of 1.64, also indicating a good L2 accent. Her motivation score was 5 as well, although her percentage of Hebrew use variable were lower than the first outlier (PHW = 50%, PHH=70%, PHF = 70%).

3.2 Acoustic Analysis

3.2.1 Vowels

In order to obtain baseline values, we first compared F1 and F2 of the control groups for each vowel. A visual overview of the mean formant spaces for these groups can be seen in 10a and 10b, separately for words and non-words. The red polygon in each diagram represents the Israeli control group, the green polygon represents the American control group and the black polygon provides a reference polygon taken from Most et al.

(2000), which employed the same set of non-words used in the present study. We can see from the nonword diagram that the Israeli test group results are in good agreement with Most et al. (2000), supporting the validity of the Israeli control group and its recordings. The American control group polygon appears to be of a smaller area than that of the Israeli, although the difference between the groups appears larger for the meaningful words. The differences also appear to be greater in certain vowels, namely the /e/ and /u/ vowels. It is also interesting to note the overall difference between meaningful words and nonwords, an aspect that will be addressed in the discussion.

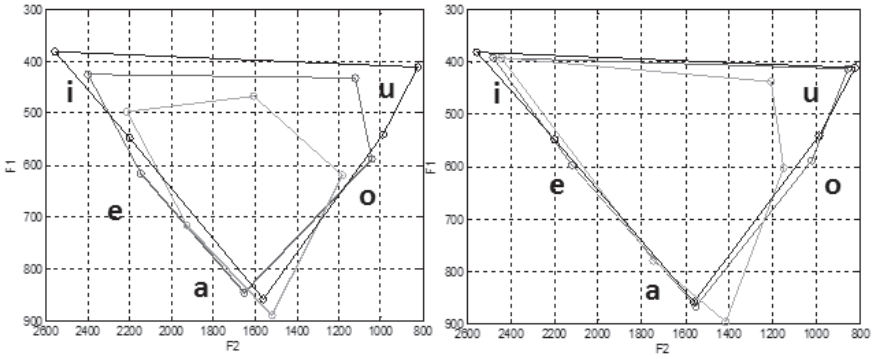


Figure 5. Comparing control groups – meaningful words (10a – on the left) & nonwords (10b-on the right). Green- American control group; Red- Israeli control group; Black- Most et al. (2000).

To further examine the acoustic differences between the control groups, a multivariate ANOVA analysis with factors group, vowel and formant, was carried out. A logarithmic scale was chosen, as it is more closely related to human perception (Hillenbrand & Gayvert, 1993; Hillenbrand et al., 1995) and is therefore more commonly used when examining formants. The ANOVA yielded significant main effects for Group ($F=7.52$, $p<.0145$), Vowel ($F=297.96$, $p<.0001$) and Formant ($F=6411.59$, $p<.0001$). Several interactions were tested, for nonwords and words – vowel * group; vowel * formant; vowel * formant * group. These interactions were significant on all accounts ($p<0.0001$ for all interactions), which led to the examination of each vowel separately. Results for this analysis are shown in Table 5. For the nonwords (/pip/, /pep/, /pap/, /pop/, /pup/), differences in F1 were found significant for the vowel /e/ alone. Differences in F2 were found to be significant for the /e,a,o,u/ vowels.

Within the meaningful words, significant differences were found for F2 in all five vowels. No significant differences were found in F1.

Table 5. **Averages, SD, F and P for the F1, F2 and F3 of the nonwords and words in the control groups**

Nonwords	Israeli control group		American control group		F(1,16)	P
	M	SD	M	SD		
F1 i	389.05	47.47	389.05	27.82	0.00	0.96
F1 e	602.56	43.095	776.25	94.72	42.20	**<0.0001**
F1 a	870.96	41.05	891.25	85.99	0.79	0.39
F1 o	588.84	42.11	602.56	58.13	0.32	0.58
F1 u	407.38	49.71	407.38	39.30	1.55	0.23
F2 i	2454.71	236.83	2454.71	236.83	0.13	0.72
F2 e	2137.96	100.76	1737.80	124.29	64.77	**<0.0001**
F2 a	1548.82	72.99	1412.54	66.57	17.14	**<0.0008**
F2 o	1023.29	48.23	1148.15	82.12	22.58	**<0.0002**
F2 u	851.14	103.85	1202.26	210.27	24.32	**<0.0002**
Words						
F2 i	2398.83	113.05	2187.76	211.07	7.42	**<0.015**
F2 e	2137.96	49.80	1905.46	136.28	15.91	**<0.0011**
F2 a	1659.59	78.21	1513.56	71.33	14.80	**<0.0014**
F2 o	1047.13	74.89	1174.90	84.03	18.25	**<0.0006**
F2 u	1122.02	108.25	1584.89	113.35	76.46	**<0.0001**

A visual overview of the formant differences between the control groups and the test group, within the words and the nonwords, can be seen in Figure 12. The circles are a scatter plot of the test group values for the F1 and F2 for each vowel. As above, the red polygon in each diagram represents the Israeli control group, the green polygon represents the American control group and the black polygon provides us with the Most et al. (2000) reference polygon.

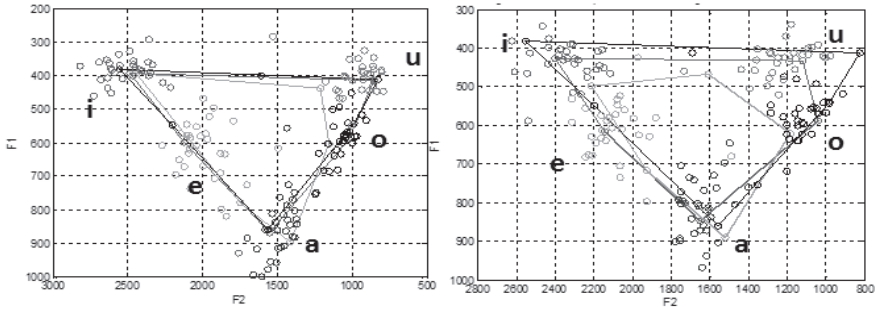


Figure 6. Control group polygons and scatter of test group for nonwords (12a) and words (12b). Green- American control group; Red- Israeli control group; Black- Most et al. (2000).

Tables 6 and 7 show the Pearson correlations between the formant frequencies and the accent ratings for the nonwords (Table 6) and for the meaningful words (Table 7). Table 8 shows the derived acoustic variables and their Pearson correlations with the accent ratings for both nonwords and meaningful words. Several significant correlations were found:

- F1 and F2 of the /e/ vowel and F2 of the /u/ vowel in the nonwords.
- the Euclidean distance of /pep/, /pup/ non-words, and the /e/ vowel for meaningful words.

Thus, the significant findings were all found in the parameters related to the /e/ and /u/ vowels.

Table 6. Mean, SD, r and p for F1 and F2 and their Pearson correlations with the accent ratings for nonwords

	M	SD	Pearson correlation	
			R	p
F1 i	384.58	21.14	0.22	0.23
F2 i	2523.77	78.13	0.02	0.89
F3 i	2996.76	107.79	0.04	0.80
F1 e	623.39	33.47	0.56	**0.001**
F2 e	2011.98	71.14	-0.42	**0.019**
F3 e	2906.96	79.12	-0.24	0.19
F1 a	861.92	39.58	0.11	0.56

F2 a	1479.32	50.38	0.14	0.46
F3 a	2753.15	79.55	0.04	0.79
F1 o	590.23	30.13	0.26	0.15
F2 o	1069.72	45.27	0.23	0.21
F3 o	2778.94	70.91	-0.26	0.15
F1 u	410.17	28.64	0.14	0.45
F2 u	938.13	72.54	0.39	**0.03**
F3 u	2777.08	88.22	-0.15	0.41

Table 7. Mean, SD, r and p for the acoustic variables and their Pearson correlation with the accent ratings for the meaningful words, N=30

	M	**SD**	**Pearson correlation**	
			p	r
F1 i	433.36	36.73	0.30	0.10
F2 i	2357.09	146.36	-0.10	0.59
F3 i	3004.35	76.98	-0.16	0.39
F1 e	609.98	56.24	0.24	0.20
F2 e	2108.12	136.12	-0.31	0.08
F3 e	3049.47	88.38	-0.13	0.47
F1 a	822.54	42.44	-0.02	0.89
F2 a	1653.54	69.97	-0.09	0.61
F3 a	2741.79	168.71	0.04	0.83
F1 o	585.15	43.20	0.27	0.14
F2 o	1132.06	70.26	0.31	0.09
F3 o	2801.91	150.90	-0.06	0.72
F1 u	432.18	37.51	0.13	0.47
F2 u	1211.86	105.27	0.30	0.10
F3 u	2764.250	122.17	-0.11	0.54

Table 8. Mean, SD, r and p for the additional acoustic variables and their Pearson correlation to the AoA and the accent rating for both nonwords and meaningful words, N=30

	M	SD	Pearson correlation	
			r	p
Areas P	**389079.02**	**96567.76**	**-0.01**	**0.94**
Areas W	287996.38	78178.29	-0.26	0.15
eucP1	147.96	93.96	-0.04	0.83
eucP2	322.05	118.35	-0.58	**0.0007**
eucP3	128.12	70.37	0.07	0.68
eucP4	115.77	55.53	-0.20	0.28
eucP5	277.59	102.73	-0.38	**0.03**
eucW1	186.36	129.27	-0.24	0.19
eucW2	232.58	85.77	-0.38	**0.03**
eucW3	177.96	63.51	-0.00	0.99
eucW4	110.51	68.30	-0.11	0.54
eucW5	399.04	122.02	-0.31	0.08

3.2.2 VOT

A two-way ANOVA was conducted with a between-subject factor of Group (three groups: the study group and the two control groups) and a within-subject factor of Vowel, i.e. the vowel following the initial /p/ (these were /i/, /e/, /a/, /o/, /u/). Means and STDs are shown in Figure 7. A significant main effect was found for Group, ($F(2,44) = 21.60$, $p < 0.001$) as well as for Vowel ($F(4,176) = 3.64$, $p < 0.05$). A significant interaction was also found between Group and Vowel ($F(2,44) = 3.05$, $p < 0.05$).

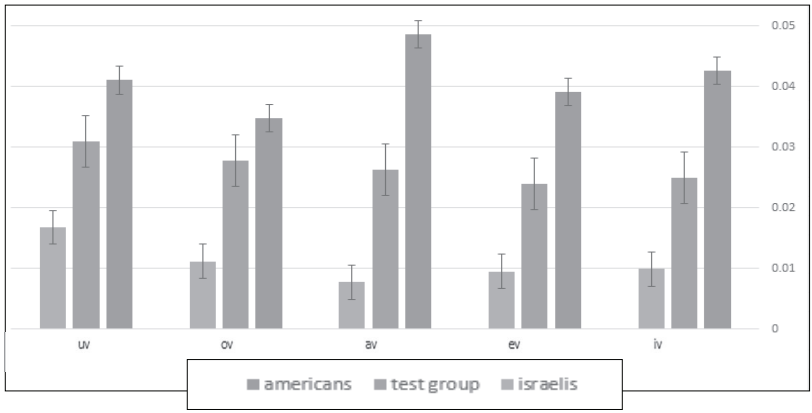


Figure 7. Average VOT duration between the different groups.

Specific comparisons were made using a Tukey post-hoc test, with significant differences found between all vowels and all groups ($p < 0.001$), except for the vowel / o / which was not significantly different between the study group and the American control group.

Next, within the experimental group the Pearson correlation between VOT duration and accent rating was calculated; the results are in Table 9. It can be seen that a moderate positive correlation was found for all vowels. It can be concluded that immigrants do tend to acquire a native VOT duration as their accent improves.

Table 9: Pearson correlations between VOT of /p/ followed by the various vowels, and the accent ratings.

Vowel	r	p
/i/	0.426	0.019
/e/	0.376	0.04
/a/	0.441	0.015
/o/	0.434	0.017
/u/	0.415	0.023

3.2.3 /l/ and /r/

To assess the specific effect of accent in /l/ and /r/, the Pearson correlation was calculated between the expert evaluation of these two consonants and

the accent ratings. Only the presence of American /l/s correlated significantly with the accent ratings, at a moderate level (r=0.57, p=0.0008).

4. Discussion

The purpose of this study was to examine the factors that affect accent acquisition amongst female American immigrants to Israel. The study also attempted to map out significant acoustic information upon which listeners determine the degree of the American accent. This was accomplished through two complementary analyses – perceptual and acoustic. The perceptual analysis was carried by having a panel of judges rate the accents of the two control groups (American and Israeli) and the study group, which consisting of women who had immigrated to Israel at various ages. The acoustic analysis included the examination of vowel formants, VOT, and expert assessment of pronunciation of the /l/ and /r/ consonants. This study is unique in the sense that most of the related body of research was carried out on acquisition of native American English accents by immigrants *to* the USA, whereas here we studied the opposite – acquisition of native Israeli accent by immigrants *from* the USA.

The main findings of the perceptual analysis were that there is a critical/sensitive period between the ages of 6-12 for the acquisition of a native accent and that the PHF (percentage of use of Hebrew with friends) plays an important role in the attainment of a native accent. From the acoustic point of view, the importance of the vowels /e/ and /u/ stood out in the formant analysis as well as VOT and the pronunciation of /l/ as substantial factors affecting the accent ratings. These finding are discussed below in detail.

4.1 The Perceptual Perspective

Initial analysis of the perceived accent ratings revealed clear differences in mean ratings between each of the three groups. This gave a good indication that the three groups were indeed representative. However, our main objective in this respect was to examine the nature of the relationship between the AoA and the accent rating as well as the influence of the other factors.

As expected, and in agreement with past research (Abu-Rabia & Iliyan, 2011; Piske et al., 2001), the AoA was in fact the strongest predictor of pronunciation accuracy. However, since there are several different theories regarding the nature of the relationship between AoA and accent, we

examined this in detail. Based on previous work (Flege & Fletcher, 1992; Flege et al., 1995; Thompson, 1991) it was decided to examine age 12 as a possible critical age. It was found that 80% of the women whose accent was rated as native Israeli immigrated before age 12. None of the immigrants who arrived before age 12 were rated as having a strong American accent. Of those who immigrated after age 12, only 17% were found to have an Israeli accent. However, upon further investigation, it was found that the correlation to the AoA variable only remains significant down to age 6. These findings indicated a non-linear acquisition pattern, which led to explore a non-linear trend line, the sigmoid pattern of acquisition. This supports the critical/sensitive period hypothesis. It is difficult to point out a narrow critical period. The findings point approximately at age 6 as the possible starting point and age 12 as the ending point, which parallels previous research (Fledge, Munro & MacKay, 1995; Long, 2005).

It is important to note that whilst our findings indicate that immigration prior to the critical/sensitive period provides an obvious advantage in the accent acquisition, it is not a guarantee of success. Similarly, immigration after the critical/sensitive period will most likely result in a less native accent, however, achieving a native accent even after the critical/sensitive period is not impossible.

Regarding other factors influencing accent acquisition, it was initially thought that the motivation (MoT) variable would be found to have a strong and independent correlation to the foreign accent level. This was based on the impression received during our meetings with the participants, as those with un-characteristically good Israeli accents testified that they were highly motivated to speak with a native accent for various reasons (professional or social). Indeed, significant correlations were found between motivation, years of education in Hebrew (YeH), use of Hebrew with friends (PHF) and the accent rating. However, when analysing these correlations considering the AoA variable, only the PHF variable was found to have a significant independent correlation with the accent rating. This was also the only variable chosen by the stepwise regression model along with the AoA. Equating correlation with causality, however, is questionable. On the one hand, this finding may be due to the possible tendency of those with a more native accent to socialize more with native Israelis or alternatively it may indicate that those who socialize more with native Israelis may acquire a better accent. Nevertheless, it seems safe to suggest that socializing with native Israelis as much as possible is likely to assist in acquiring a more native accent.

As expected, and in accordance with previous research (Moyer, 1999; Poretta & Tucker, 2012) we also found a small number of outliers, which could only be analyzed qualitatively. In the case of women with late AoA having good Israeli accent ratings, we noted high motivation scores, and middle to high percentages of Hebrew use scores (the highest of these being use of Hebrew with friends). These observations are in line with the finding that the percentage of Hebrew use with friends plays a significant role in developing a native accent. The motivation variable seems to play a larger role in the case of the outliers, however, due to the small number of outliers this could not be verified statistically in the current study.

Interestingly, there were also two outliers on the other end of the scale: women who arrived in Israel between the ages of three and six and yet did not achieve a native accent rating. Upon closer investigation, one of them did not have a high motivation score nor did she have a high percentage of Hebrew use, which is in line with the other findings. However, the second had a high motivation score and very high percentage of Hebrew use. This may point to the presence of individual differences between the participants and merits further investigation.

4.2 The Acoustic Perspective

Regarding the acoustic analysis, the findings for the control groups were as follows:

- The main differentiator between vowel pronunciation between the two control groups, for most vowels, was F2. Differences in F1 were only found for the /e/ vowel in the nonwords.
- There were less formant differences between vowels in the non-words *vs.* the real words. A possible explanation is that when articulating real words, speakers use their regular speaking habits, as opposed to articulating nonsense words.
- VOT duration for /p/ was approximately twice as long for the American control group in comparison to the Israeli control group.

Regarding the test group, several variables were found to correlate with the accent ratings, highlighting their importance. Within these variables, both formants of /e/ and F2 of /u/ stood out as being highly important. These results are supported by the control group results stated above. However, unlike the control groups, within the test group there were more differences in the nonwords than the words. Possibly, women who have already been living in Israel for many years have more practice articulating common

Hebrew words and can therefore pronounce them with a better Israeli accent than the nonwords.

We hypothesized that within the subjective acoustic data, the ratings of the /l/ and /r/ consonants would correlate with the listener evaluations. However, only the American /l/ variable was found to correlate with the accent rating. We find it surprising that the American /r/, commonly acknowledged as a characteristic property of the American accent, was not found to correlate with the accent rating. A possible explanation is that due to this stereotype, immigrants make a conscious effort to learn how to pronounce the Hebrew /r/ rather quickly, as opposed to pronunciation of /l/ and the various vowels.

Finally, VOT of the /p/ consonant (the only consonant studied in this context) was found to be moderately correlated to the accent ratings. We find this particularly interesting, since VOT is not a parameter that is well-known to laymen, nor as easily perceived as the other acoustic parameters studied here. Nevertheless, it appears that speakers adopt a more Israeli VOT as their accent improves.

4.3 Limitations and directions for future research

Our study group was comprised of women participants alone due to the formant differences between men and women. Further research should examine whether these factors generalize to men as well.

Outliers were found in this study on both ends of the scale– those who acquired a native accent despite arriving after the critical period and those who arrived before the critical period but did not achieve a native accent. Due to the small number of participants, only a very small number of outliers was found. It would be very interesting to further investigate this phenomenon on a larger scale, testing the speakers for particular psychoacoustic abilities.

These results of this study give a strong indication that accent reduction programs formulated for American immigrants (old and new) should include exercises with a particular focus on the /e/ and /u/ vowels, as well as teaching the Israeli /l/. A longitudinal study on the effectiveness of such focused training would be very interesting.

Whilst the emotional impact of having a foreign accent was not incorporated into this study, its importance was highlighted throughout our informal

discussions with the participants. During the recording sessions, many of the participants expressed their frustration at having a foreign accent. They felt that it hindered their ability to negotiate, to be successful at job interviews or meetings and that they were constantly treated as foreign by various people they encountered. In Israel, a school for learning Hebrew is available for those who immigrate known as "ulpan". Currently, pronunciation lessons are not an integral part of the "ulpan" system. The data presented in this research, along with further research required to build and implement an accent learning program, which would be available to all new immigrants, would be of great importance to the integration of new immigrants into Israeli society. We hope that the information gathered in this research may in turn allow new immigrants to Israel to acquire a native accent more easily and hence assist in their adjustment to their new living environment.

References

Abu-Rabia, S. and Iliyan, S. 2011. Factors affecting accent acquisition: The case of Russian immigrants in Israel. *Reading Matrix: An International Online Journal* 2011 (4), 160-170.

Amir, N., Tzenker, O., Amir, O. and Rosenhouse, J. 2012. Quantifying Vowel Characteristics in Hebrew and Arabic. In *Afeka Conference for Speech processing*.

Amir, O. and Levine-Yundof, R. 2013. Listeners' attitude toward people with dysphonia. *Journal of Voice* 27 (4), 524-e1.

Arslan, L. M. and Hansen. J. H. 1997. A study of temporal features and frequency characteristics in American English foreign accent. *The Journal of the Acoustical Society of America*, 102 (1).

Asher, J. J. and Garcia, R. 1969. The optimal age to learn a foreign language. *The Modern Language Journal* 53 (5), 334-341.

Baker, W. 2010. Effects of age and experience on the production of English word-final stop by Korean speakers. *Bilingualism: Language and Cognition* 13 (03), 263-278.

Barker, B. A. and Turner, L. M. 2015. Influences of foreign accent on preschoolers' word recognition and story comprehension. *Applied Psycholinguistics* 36 (5), 1111-113.

Bongaerts, T., Planken, B. and Schils, E. 1995. Can late starters attain a native accent in a foreign language? A test of the critical period hypothesis. *The Age Factor in Second Language Acquisition*, 30-50.

Bongaerts, T., Van Summeren, C., Planken, B. and Schils, E. 1997. Age and ultimate attainment in the pronunciation of a foreign language. *Studies in Second Language Acquisition* 19 (04), 447-465.

Bongaerts, T. 1999. Ultimate attainment in L2 pronunciation: the case of very advanced late L2 learners. In *Second Language Acquisition and the Critical Period Hypothesis*, edited by D. Birdsong. Routledge.

Chan, K. Y., Hall, M. D. and Assgari, A. A. 2017. The role of vowel formant frequencies and duration in the perception of foreign accent. *Journal of Cognitive Psychology* 29 (1), 23-34.

Chodroff, E. and Wilson, C. 2017. Structure in talker-specific phonetic realization: Covariation of stop consonant VOT in American English, *Journal of Phonetics* 61, 30-47.

Derwing, T. M. and Munro, M. J. 2013. The Development of L2 Oral Language Skills in Two L1 Groups: A 7-Year Study. *Language Learning* 63 (2), 163-185.

Dewaele, J. M. and McCloskey, J. 2015. Attitudes towards foreign accents among adult multilingual language users. *Journal of Multilingual and Multicultural Development* 36 (3), 221-238.

Felps, D., Bortfeld, H. and Gutierrez-Osuna, R. 2009. Foreign accent conversion in computer assisted pronunciation training. *Speech Communication* 51 (10), 920-932.

Flege, J. E. and Fletcher, K. L. 1992. Talker and listener effects on degree of perceived foreign accent. *The Journal of the Acoustical Society of America* 91 (1), 370-389.

Flege, J. E., Munro, M. J. and MacKay, I. R. 1995. Factors affecting strength of perceived foreign accent in a second language. *The Journal of the Acoustical Society of America* 97 (5), 3125-3134.

Flege, J. E., Schirru, C. and MacKay, I. R. 2003. Interaction between the native and second language phonetic subsystems. *Speech Communication* 40 (4), 467-491.

Floccia, C., Butler, J., Girard, F. and Goslin, J. 2009. Categorization of regional and foreign accent n 5-to 7-year-old British children. *International Journal of Behavioral Development* 33 (4), 366-375.

Gluszek, A. and Dovidio, J. F. 2010. Speaking with a nonnative accent: Perceptions of bias, communication difficulties, and belonging in the United States. *Journal of Language and Social Psychology* 29 (2), 224-234.

González-Bueno (1997). Voice-onset-time in the perception of foreign accent by native listeners of Spanish**.** *IRAL - International Review of Applied Linguistics in Language Teaching* 35 (4), 251-268.

Granena, G. and Long, M. H. 2013. Age of onset, length of residence, language aptitude, and ultimate L2 attainment in three linguistic domains. *Second Language Research* 29 (3), 311-343.

Hillenbrand, J., Getty, L. A., Clark, M. J. and Wheeler, K. 1995. Acoustic characteristics of American English vowels. *The Journal of the Acoustical society of America* 97 (5), 3099-3111.

Huang B. H. and Jun, S. 2015. Rater differences in the assessment of foreign accents. *Studies in Second Language Acquisition* 77 (3), 1-28.

Last, E. 2018. Factors affecting accent acquisition among female American immigrants to Israel, MA thesis, Tel Aviv University.

Lee, B., Guion, S. G. and Harada, T. 2006. Acoustic analysis of the production of unstressed English vowels by early and late Korean and Japanese bilinguals. *Studies in Second Language Acquisition* 28 (3), 487-513.

Lee, S. A. S. and Iverson, G. K. 2012. Vowel category formation in Korean–English bilingual children. *Journal of Speech, Language, and Hearing Research* 55 (5), 1449-1462.

Lenneberg, E. H., Chomsky, N. and Marx, O. 1967. *Biological foundations of Language* (Vol. 68). New York: Wiley.

Lev-Ari, S. and Keysar, B. 2010. Why don't we believe non-native speakers? The influence of accent on credibility. *Journal of Experimental Social Psychology* 46 (6), 1093-1096.

Lisker, L. and A.S. Abramson. 1964. A Cross-Language Study of Voicing in Initial Stops: Acoustical Measurements, *Word* 20, 384-422.

Long, M. 2005. Problems with supposed counter-evidence to the critical period hypothesis, *International Review of Applied Linguistics IN Language Teaching* 43 (4), 287-317.

McCullough, E. A. and Clopper, C. G. 2016. Perceptual subcategories within non-native English. *Journal of Phonetics* 55, 19-37.

Most, T., Amir, O. and Tobin, Y. 2000. The Hebrew vowel system: Raw and normalized acoustic data. *Language and Speech* 43 (3), 295-308.

Moyer, A. 1999. Ultimate attainment in L2 phonology. *Studies in Second Language Acquisition* 21 (01), 81-108.

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

Muñoz, C. 2014. Contrasting effects of starting age and input on the oral performance of foreign language learners. *Applied Linguistics* 35 (4), 463-482.

Munro, M. and Mann, V. 2005. Age of immersion as a predictor of foreign accent. *Applied Psycholinguistics* 26 (3), 311-341.

Netelenbos, N., Li, F. and Rosen, N. 2016. Stop consonant production of French immersion students in Western Canada: A study of voice onset time. *International Journal of Bilingualism* 20 (3), 346-357.

Ng, M. L., Chen, Y. and Sadaka, J. 2008. Vowel features in Turkish accented English. *International Journal of Speech-Language Pathology* 10 (6), 404-413.

Pardede, P. 2010. The Role of Pronunciation in a Foreign Language Program. *Universitas Kristen Indonesia. Retrieved from* https://parlindunganpardede.wordpress.com/2010/10/07/349.

Pantos, A. J. and Perkins, A. W. 2013. Measuring implicit and explicit attitudes toward foreign accented speech. *Journal of Language and Social Psychology* 32 (1), 3-20.

Pinget, A. F., Bosker, H. R., Quené, H. and de Jong, N. H. 2014. Native speakers' perceptions of fluency and accent in L2 speech. *Language Testing* 31 (3), 349-365.

Piske, T., MacKay, I. R., and Flege, J. E. 2001. Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics* 29 (2), 191-215.

Polat, N. 2011. Gender differences in motivation and L2 accent attainment: An investigation of young Kurdish learners of Turkish. *Language Learning Journal* 39 (1), 19-41.

Pols, L. C., Van der Kamp, L. T. and Plomp, R. 1969. Perceptual and physical space of vowel sounds. *The Journal of the Acoustical Society of America* 46 (2B), 458-467.

Porretta, V. and Tucker, B. V. 2012. Predicting accentedness: Acoustic measurements of Chinese-accented English. *Canadian Acoustics* 40 (3), 34-35.

Porretta, V., Kyröläinen, A. J. and Tucker, B. V. 2015. Perceived foreign accentedness: Acoustic distances and lexical properties. *Attention, Perception and Psychophysics* 77 (7), 2438.

Raphael, L., Tobin, Y., Faber, A., Most, T., Kollia, H. and Milstein, D. 1995. Intermediate values of voice onset time. *Producing Speech: Contemporary Issues*, 117-127.

Reiterer, S. M., Hu, X., Sumathi, T. A., and Singh, N. C. 2013. Are you a good mimic? Neuro-acoustic signatures for speech imitation ability. *Frontiers in Psychology* 4, 1-13.

Schoonmaker-Gates, E. 2015. On Voice-onset Time as a Cue to Foreign Accent in Spanish: Native and Nonnative Perceptions. *Hispania* 98 (4), 779-791.

Scovel, T. 1969. Foreign Accents, language acquisition and cerebral dominance. *Language Learning* 19 (3-4), 245-253.

Thompson, I. 1991. Foreign Accents Revisited: The English Pronunciation of Russian Immigrants*. *Language Learning* 41 (2), 177-204.

Zampini, M. 2008. L2 speech production research. *Phonology and Second Language Acquisition*, 219-249.

Appendix 1a.: Reading materials- sentences

נא לקרא את המשפטים הבאים:
1. המורה כתב פיפ על הלוח.
המורה כתב פפ על הלוח.
המורה כתב פפ על הלוח.
המורה כתב פופ על הלוח.
המורה כתב פופ על הלוח.

2. הילד צייר פיל על הדף.
הילדה שמה סיר על הראש.
האב כתב "גיר" על הדף.

3. האמא נתנה שק לילדה.
אילנית שמה קש בתוך הכוס.
ראובן צייר ים על הדף.

4. שמעון בן שנה.
חוקרים מצאו גן הגורם למחלה.
איתן הדליק נר בחג החנוכה.

5. אורית ציירה קוף על הנייר.
מיכל החזיקה מוט כבד וגדול.
הילד הכניס חול לתוך הדלי.

6. איילה טפסה על צוק גבוה מאד.
חיים אכל שום לבן וחריף.
סבתא ספרה על שוק מפורסם באמריקה.

II. קטע קריאה – דף נפרד
III. ספרי לי על עצמך, מה עשית היום

Appendix 1b.: The "Thousand Islands" Hebrew Reading Passage (Amir & Levine-Yundof, 2013)

The "Thousand Islands" Hebrew Reading Passage

אלף האיים

רוטב "אלף האיים" הוא רוטב אותו מכינים ממיונז, קטשופ ומגוון של ירקות חתוכים. הרוטב יכול להכיל בצלים, מלפפונים חמוצים, פלפלים, זיתים וביצה קשה. ניתן להשתמש ברוטב כממרח לכריך או כרוטב לסלט. בשנים האחרונות הפך הרוטב למקובל מאוד ברשתות המזון המהיר. את הרוטב מוסיפים כמעט לכל מאכל, מהמבורגר ועד לסלט.

רוטב "אלף האיים" נקרא על שם "ממלכת אלף האיים". זו קבוצת איים קטנים באוקיאנוס ההודי. שלושה רבעים מהאזור מכוסים בים. בשטח היבשתי של הממלכה קיימים אלף וחמש מאות איים, שרק מאתיים מהם מיושבים. האיים הללו הם ראשי הרים תת-מימיים אשר מבצבצים מעל פני הים. באיים אין הרים או גבעות, ולכן הממלכה נחשבת למדינה השטוחה בעולם.

"אלף האיים" היא מדינה עם דגל, חוקים, בתי-ספר ומשטרה. תושבי האיים מנותקים מהעולם וחיים בכפרים קטנים ובעיירות. בכל אי ניתן למצוא כפר או עיירה אחת בלבד. מרבית האיים עשירים בצמחיה טרופית ובעצי קוקוס. ניתן להקיף כל אי תוך עשר דקות, בהליכה איטית על חוף הים הלבן. האיים מאפשרים נופש מיוחד, רגוע או פעיל, על רקע של חופים לבנים, שוניות עשירות ושמיים כחולים.

האזור בו ממוקמת "ממלכת אלף האיים" הוא גן-עדן לאוהבי ים. האיים הם אתר הנופש המושלם עבור הצוללן החובב. בכל צלילה ניתן לראות דגים מעניינים ותמנונים אשר חיים בין אלמוגים מרהיבים. לעיתים קרובות ניתן להבחין גם בכרישים ובצבי ים. במהלך הנופש שוהים הצוללנים על ספינה, וכאשר הם רוצים לצלול, הם עוברים לסירת עץ מיוחדת. הספינות משמשות גם כבתי מלון צפים.

ספנים ותיקים מספרים כי המעבר בין האיים הקטנים צר ומסוכן. לכן ספינות רבות עלו על שרטון בסמוך לאיים. מעברים אלו נחשבו למלכודות מוות לספינות שניסו למצוא מסתור מפני סערות. בשולי המעברים הצרים קיימות מערות קטנות וגדולות. מכיוון שהמערות הללו מוגנות מפני זרמי הים, הן עשירות במיוחד באלמוגים ובדגים נדירים.

Chapter Eleven

The Role of Acoustic Memory and Phonological Short-Term Memory in Non-Native Perception: Which Memory Matters More and When?

Vita V. Kogan

Abstract

Previous research has shown that phonological short-term memory (PSTM) is closely connected to non-native perceptual abilities. PSTM facilitates the development of target-like cue-weighting and is an important factor in the acquisition of L2 (second language) categories (Cerviño-Povedano & Mora, 2015; MacKay, Meador & Flege, 2001). Another important cognitive ability for non-native perception is acoustic memory (AM). Largely under-researched, AM is believed to be responsible for storing acoustic information at a pre-categorical level, and it also contributes to the task of L2 speech perception and learning (Safronova, 2016). In our study, monolingual Spanish speakers were asked to discriminate between two members of an unfamiliar (L0) Russian contrast /i - ɨ/ that does not exist in Spanish. Their PSTM was measured with a non-word recognition task and AM with a target sound recognition task. Mixed-effects regression analysis showed that it was AM capacity and not PSTM capacity that influenced L0 perception. In a follow-up study with the same design, Spanish L2 learners of Russian relied on both PSTM and AM to discriminate between Russian /i/ and /ɨ/. It seems that naive listeners use auditory-based pre-categorical judgments when discriminating between two members of a novel contrast, and L2 learners use both – acoustic and phonological information in the same task. The surprising finding was that naive listeners who never learned Russian before managed the L0 perceptual task quite well. This finding goes

along with the studies that demonstrate that monolingual adults have life-long access to cross-language phonetic differences (Flege & Hammond, 1982; Grenon, 2006).

Keywords: L0/L2 speech perception, phonological memory, acoustic memory, individual differences, phonetic categories

1. Introduction

It is widely acknowledged that adult L2 learners experience difficulty when acquiring non-native sound systems. The situation, in which a learner has mastered other L2 subsystems (e.g., syntax, morphology), but still struggles perceiving and producing novel non-native sounds is rather common and granted phonetic ability a special status in second language acquisition research (e.g., Dogil & Reiterer, 2009; Neufeld, 1987). Yet some individuals manage to achieve a near-native phonetic and phonological proficiency (Moyer, 2014). Even when only limited L2 input is available, cases have been reported of learners demonstrating exceptional outcomes in L2 phonetics and phonology (Bongaerts et al., 1997; Ioup et al., 1994). Factors such as native language (L1) background, age of onset of L2 learning, and the quality and quantity of L2 exposure constitute a stable source of inter-learner variability. Research on the role of aptitude in L2 acquisition has shown that also a number of cognitive abilities and personality-based features influence L2 phonetic and phonological abilities (Cerviño-Povedano & Mora, 2015; Darcy, Mora, & Daidone, 2016; Darcy, Park, & Yang, 2015; Hu & Reiterer, 2009; Isaacs & Trofimovich, 2010). The term *phonetic talent* first appeared in Jilka's research (Jilka et al., 2008; Jilka, 2009) who defined the concept broadly as an individual ability to acquire L2 phonetics and phonology and clearly separated it from other linguistic abilities. The notion that phonetic talent or phonetic ability is an independent subcomponent of language aptitude is not new: half a century ago, Carroll (1981) coined the term *phonemic coding ability* that he defined as the distinct innate ability to recognize new sounds and to store them in long-term memory.

Based on recent research, among most prominent cognitive candidates that play a role in L2 phonetic and phonological acquisition are phonological short-term memory (Cerviño-Povedano & Mora, 2015; Inceoglu, 2019; Lengeris & Nicolaidis, 2014), acoustic memory (Safronova & Mora, 2012), attentional control (Darcy, Park, & Yang, 2015; Safronova & Mora, 2013), inhibition (Darcy, Mora, & Daidone, 2016), and musical ability (Coumel, Christiner & Reiterer, 2019). In regard to the latter, some researchers

suggest that musical ability can be further connected to a number of cognitive abilities, one of which is phonological short-term memory (Williamson, Baddeley, & Hitch, 2010).

In this chapter, we are particularly interested in non-native perception. Therefore, all references to L2 phonetic and phonological acquisition should be understood as the acquisition of L2 perception as opposed to L2 production or both. There are several reasons why we are interested in L2 perception. First, L2 perception has to be in place to ensure accurate production. Even though there is no consensus in the field on such a sequence of L2 sound acquisition, we believe that difficulties with L2 sound production have a perceptual basis, i.e. learners have a "perceptual foreign accent" (Escudero, 2005). Once L2 perceptual representations have formed, they are used as targets for articulatory movements (Tourville & Guenther, 2011). Thus, better perception assumes better production. Flege's Second Language Model (SLM: 1995, 2002) also suggested that the accuracy of L2 segmental perception places an upper limit on the accuracy with which L2 sounds are produced. Recently, Flege and Bohn (2021) reviewed this hypothesis and proposed a more complex relationship between L2 perception and production emphasizing their coevolving nature, i.e., there is a bidirectional and co-equal link between the two without precedence (also see Chao, Ochoa, & Daliri, 2019 and Reiterer et al., 2013) for similar observations). Another reason to study L2 perception is its ethereal nature that causes methodological difficulties with measuring perception consistently. For this reason, research on individual differences in L2 perception is still limited and rather mixed.

1.1. Phonological short-term memory

Phonological short-term memory (PSTM) has received noticeable attention in second language acquisition. First appeared as a subcomponent of Baddeley's working memory model and called the phonological loop (Figure 1: Baddeley, 1986), PSTM is the ability to hold phonological material in the short-term store for about two seconds and to generate phonological representations (e.g., phonetic categories) that later can be transferred to long-term memory.

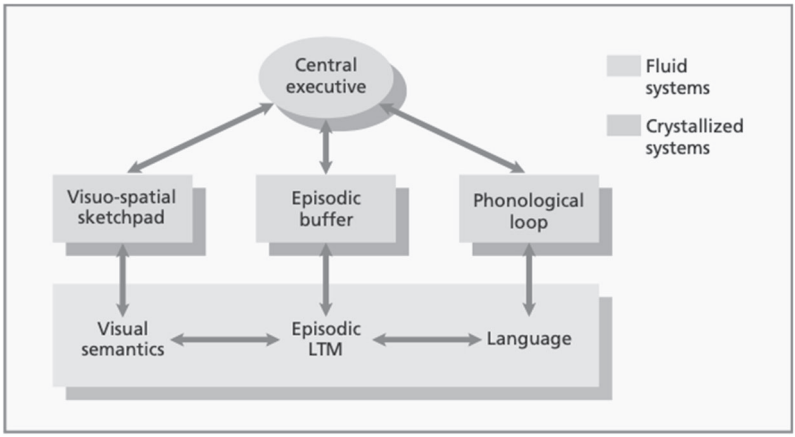


Figure 1. The Baddeley (2000) version of the multicomponent working memory, reproduced from Baddeley, Eysenck, and Anderson (2015).

Baddeley and other researchers observed that PSTM seems to play a central role in first and second language acquisition (L1: Baddeley, 1986; L2: Brooks, Kempe, & Donachie, 2009; Hummel, 2009; Kormos & Sáfár, 2008a). For example, Brooks, Kempe, and Donachie (2009) suggested that PSTM capacity could be a better predictor of the overall L2 attainment than working memory capacity, non-verbal intelligence, or prior language learning experience. In regard to the L2 segmental processing, PSTM has been shown to influence the perception of L2 consonants (Lengeris & Nicolaidis, 2014; MacKay, Meador & Flege, 2001) and vowels (Cerviño-Povedano & Mora, 2011; Cerviño-Povedano & Mora, 2015; Inceoglu, 2019) by facilitating the development of native-like cue-weighting and, thus, the long-term representations of L2 phonetic segments. Lengeris and Nicolaidis (2014) demonstrated that even in adverse listening conditions when two types of noise were introduced, difficult L2 English consonants were identified better by Greek listeners with greater PSTM.

1.2. Acoustic memory

In contrast, acoustic memory (AM) has not been researched much in its relations to second language acquisition and specifically to L2 perception. Most of the research done on AM comes from the field of cognitive psychology and culminated in the 60s and 70s. Sometimes referred to as echoic memory (Neisser, 1967), a sensory buffer store (Atkinson & Shiffrin,

1968), pre-categorical acoustic storage (Crowder & Morton, 1969), or auditory memory (Fujisaki & Kawashima, 1970; Cowan & Morse, 1986), AM is operationalized as an individual's memory capacity to temporarily store non-verbal speech-like acoustic information at a pre-categorical level, i.e. before phonological encoding. Several studies have shown that AM might play a role in within-category vowel discrimination (Cowan & Morse, 1986; Safronova, 2016), although the findings are mixed. Safronova and Mora (2012) tested a group of Spanish/Catalan L2 learners on their ability to perceive the English tense-lax vowel contrast /i:- ɪ/ and did not find the relationship between AM and the discrimination ability. However, they observed that AM measures did not correlate with PSTM measures suggesting that different cognitive substrates underlay these skills. Ghaffarvand Mokari and Werner (2019) confirmed this observation: in their study, there was no significant relationship between PSTM and AM; they also did not find an effect of AM on L2 phonetic training. On the other hand, Safronova (2016) found the connection between AM and L2 perception of cross-language phonetic distance suggesting that listeners do benefit from greater AM. Safronova (2016) suggested that AM provides the potential for processing a larger amount of L2 speech input, which facilitates the detection of fine acoustic-phonetic differences between sounds, even if they are two instances of the same L1 category (e.g., the English /i:/ and /ɪ/ are two instances of the Spanish category /i/).

Thus, it seems that the language perception mechanism consists of at least two perceptual levels: the sensory level and the phonetic category level (Figure 2). Most of the L2 speech models would include both levels (e.g., Flege & Bohn, 2021; Grenon, 2006; Ramus et al., 2010), suggesting that allophones are encoded separately at the early stage of speech processing, and the efficiency of this process might depend on AM capacity.

Because beginner language learners do not have robust phonetic categories for the novel L2 sounds they encounter, AM might be the only perceptual channel they rely on at the onset of L2 acquisition. Considering how important AM at this acquisitional stage, the lack of research on the contribution of AM to L2 sound processing is rather intriguing, especially in comparison with how much research is produced in relation to L2 and PSTM. A better understanding of cognitive abilities involved at each stage of L2 speech acquisition might explain conflicting results in PSTM/AM studies, not to mention that linking various cognitive abilities with SLA developmental stages has important pedagogical implications (e.g., different types of instruction and feedback, see Skehan (2015) and R. Ellis (2012) for relevant discussion).

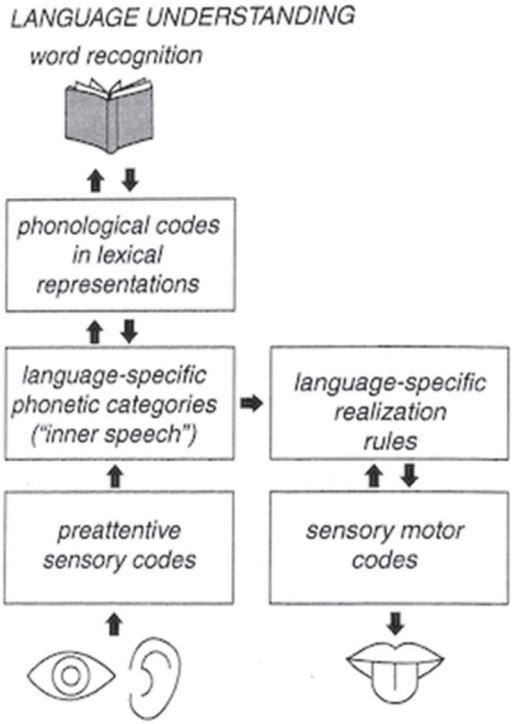


Figure 2. The perception-production model assumed by SLM (Flege, 1995), reproduced from Flege and Bohn (2021). Here pre-attentive sensory codes will be processed by acoustic memory and language-specific phonetic categories by phonological short-term memory.

1.3. Present study

The present study compares two groups of Spanish speakers: naïve perceivers with no previous training in Russian (the L0 group) and A1.2 beginner learners of Russian (the L2 group). Both groups were tested in terms of their perceptual sensitivity for a novel Russian contrast /i - ɨ/. According to Perceptual Assimilation Model (PAM: Best, 1993, 1994; PAM-L2: Best & Tyler, 2007), non-native contrasts present various degrees of challenge to perceivers. Depending on the assimilation pattern, novel sounds can be assigned to one or two native categories. The intermediate state is also possible: when one member of the contrast is perceived as a better exemplar of a native category and the other member – as a deviant

exemplar. The latter is the case with the Russian contrast /i - ɨ/ perceived by naïve Spanish speakers as two instances of their native category /i/ but with various degrees of deviation. Whereas Spanish /i/ is very similar to Russian /i/, Russian / ɨ/ does not have a close analogy in Spanish and is likely to be perceived as a bad variant of Spanish /i/). Two members of the Russian contrast /i - ɨ/ are very close acoustically and the only feature that differentiates them from each other is [+-back] (see Figure 3).

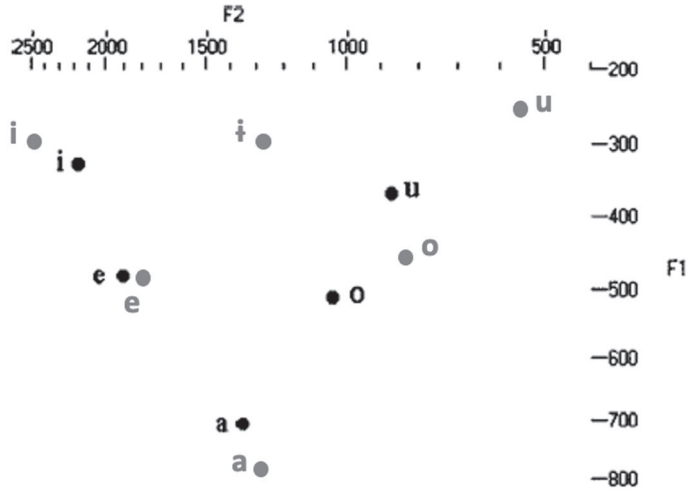


Figure 3. The vowel inventory of Spanish (in black: Martínez-Celdrán, Fernández-Planas, & Carrera-Sabaté, 2003) and Russian (in red: Jones, 1953; Yanushevskaya, & Bunčić, 2015).

The prediction is that the L0 group will rely on AM in discriminating between /i/ and /ɨ/ and the L2 group – who had prior experience with both vowels and began developing phonetic categories for them – will employ both AM and PSTM. Thus, our hypothesis is that PSTM will only facilitate the perception of the L2 contrast at the later level of proficiency but not at the very onset of acquisition (L0/naïve perception).

2. Methodology

2.1. Online recruiting and testing

The data collection was administered via *PsyToolkit*, an online platform for designing, programming, and running experiments. L2 Spanish learners of Russian were recruited with the help of the Pushkin State Russian Language Institute in Barcelona, Spain. To recruit monolingual Spanish speakers, we used *CloudResearch* (formally known as *TurkPrime*), a crowdsourcing data acquisition service. *CloudResearch* uses proprietary algorithms to collect information across time and to check for consistency; participants who provide inconsistent information are red-flagged and disqualified from taking part in future studies. *CloudResearch* also uses cross-demographic comparisons to validate the feasibility of online data collection approaches. Multiple data sources are leveraged to provide a broad global network of respondents. By using digital fingerprinting technology *CloudResearch* prevents duplicates within the same study or between studies and across sample sources. To achieve sample representativeness, the sample demographics are matched to the proportion that is reported in the latest census in terms of the distribution of gender, age, ethnicity, region, and income.

2.2. Participants

Spanish monolingual participants. Thirty-one participants (female = 8; *M*age = 30.84, *SD* = 5) were European Spanish functional monolinguals from central Spain without any prior knowledge or significant exposure (friends and family, traveling, media, etc.) to Russian. According to Best and Tyler (2007), functional monolinguals are raised in monolingual homes without learning another language before attending school. They have a basic knowledge of one foreign language, usually, English, learned in school (i.e., basic classroom instruction and grammar), have not resided in a non-Spanish-speaking country for more than a month, and use only Spanish in their daily lives. All participants were screened for knowledge of languages other than Spanish, by being presented with appropriate questions multiple times. Eleven participants reported being unable to hold a conversation in *any* foreign language. Twenty participants reported basic knowledge of English that did not surpass the A1-A2 levels of proficiency. Other languages that the participants indicated as being "familiar with" included (here in the alphabetic order): Basque, Catalan, French, Galician, Italian and Valencian. One participant mentioned studying German and

Japanese "briefly, long ago" and another participant reported a "very rudimentary" knowledge of Romanian.

Spanish L2 learners. Twenty-seven participants (female = 11; *M*age = 26.8, *SD* = 6.05) were European Spanish learners of Russian from Barcelona, Spain. None of the participants were Spanish-Catalan bilinguals or had the above-intermediate (B1 or higher) level of proficiency in any foreign language, including Russian. Seventeen participants reported speaking other languages besides Spanish and Russian, namely (here in the alphabetic order): English, French, Galician, and Italian. One participant reported speaking five foreign languages: Danish, French, Italian, Swedish, and German, although none of these languages were at the level higher than A2.

Spanish L2 learners have completed one semester (44 hours) of Russian language instruction at the Pushkin State Russian Language Institute in Barcelona and identified their proficiency level as advanced A1 or A1.2. At this level, learners can understand and produce short messages on familiar topics and activities (Council of Europe. Council for Cultural Co-operation. Education Committee. Modern Languages Division, 2001).

Russian participants. The study recruited a group of 16 native Russian speakers (female = 10, *M*age = 24.2, *SD* = 3.14) to provide baseline data for the Russian vowel discrimination task, recruited and screened online, with the help of several social networks, principally Facebook. As opposed to two previous groups of participants, the Russian participants were not compensated for their time. The data from the Russian speakers were compared to the Spanish speakers' performance to ensure that the discrimination task worked appropriately.

2.3. General procedure

Participants were informed about the purpose, procedures, and risks of the study as well as its potential benefits. They voluntarily agreed to participate, signed a consent form, and received monetary compensation for their participation (except for the Russian group).

The experimental design consisted of a single testing session conducted in Spanish or Russian (for the Russian native speakers only) on the same day. After a short screening session to confirm normal hearing and expected language proficiency levels (e.g., no advanced L2 proficiency for Spanish monolinguals), each participant received a URL to the experiment's website

hosted on the *PsyToolkit* platform. On entering the website, participants were taken to an information sheet and informed-consent screens. They were also required to declare the model of technology/device they were using to perform the tasks. The tasks could be only taken on a stationary computer and not on a mobile device, the set of headphones was a prerequisite. Before beginning the audio portion of the experiment, participants were prompted to check the sounds in the right and the left channels of their headphones.

Next, participants took a questionnaire that consisted of demographic information and language history. Then, the tasks followed with multiple opportunities to take breaks. At the end of the experiment, Spanish participants received an automatically-generated individual code to present to a researcher to receive monetary compensation for their participation.

2.4. Instruments

Questionnaire. To collect demographic information and language history a questionnaire was administered. The questionnaire consisted of 11 questions related to participants' age, gender, birthplace, and abilities to speak various languages.

Serial nonword recognition task. PSTM capacity has a limited span, which can be measured by a serial recall of numbers or words. In this study, PSTM was assessed through a serial nonword recognition task developed by Cerviño-Povedano and Mora (2011). The only difference between their task and the present one is that we used Spanish nonwords instead of Catalan nonwords. Compared to a nonword *repetition* task, a serial nonword *recognition* task might be a better measure of PSTM since there is no speaking involved, and therefore, the articulatory constraints are avoided altogether.

Participants heard 24 pairs of nonword strings increasing in length and decided whether the order of the nonwords in the sequences was the same or different. The nonwords were developed using *Syllabarium* corporus that consists of statistics for Spanish syllables (Duñabeitia et al., 2010). Only the CVC syllables of high frequency were selected, equal to or greater than 1000 on the frequency index. The selected 160 syllables were recorded at the Phonetic Laboratory at the University of Barcelona by a native female speaker of European Spanish as a part of a carrier sentence *Yo digo _ / Yo digo _ una vez.* (I said _ / I said _ again). The recorded CVC nonwords were extracted from the sentences and processed in Praat (Boersma & Weenink,

2013) to normalize for amplitude (70dB) and remove alternating current (AC). The nonwords were normalized for length, which was determined by taking the average length of all nonword stimuli (650ms). The best tokens were selected based on auditory judgments and acoustic measurements. Every trial contained a variety of vowels and consonantal contexts.

Example of the same sequence:

bul tad som fes sil
bul tad som fes sil

Example of a different sequence (the 3rd and the 4th nonwords are switched):

bul tad ***som fes*** *sil*
bul tad ***fes som*** *sil*

In this task, a weighted score was obtained by assigning five, six, and seven points to the correct responses of five-, six- and seven-item sequences with a maximum score of 144 points.

Target sound recognition task. To test participants' ability to hold acoustic properties of speech in AM, we used a target sound recognition task (Li, Cowan, & Saults, 2013; Safronova & Mora, 2012). Participants had to listen to a sequence of sounds and a target sound presented after the sequence and to decide whether the target sound belonged to the sequence. The stimuli consisted of 101 Spanish CV syllables from *Syllabarium* corpora recorded in the same manner as the nonwords in the serial nonword recognition task. The CV syllables were normalized for length and manipulated through frequency rotation – speech rotation (Scott et al., 2009). This technique preserves the acoustic complexity of the stimuli while making it impossible to encode phonologically.

The stimuli were presented in three blocks of two-, three- and four-item sequences. Each sequence was followed by 3000ms silence and a target sound afterward, which could be the same or different from the sequence items. Each block contained eight trials in randomized order. The weighted score was computed by assigning scores of two, three and four points to the correct responses of two-, three-, and four-item sequences, respectively, with a maximum score of 72 points.

Example of the same trial:

Sound 1 – Sound 2 *Sound 1*

Example of a different trial:

Sound 1 – Sound 2 *Sound 3*

Rated dissimilarity task. A rated dissimilarity task was designed to assess participants' perceptual sensitivity in differentiating Russian /i - ɨ/. A rated dissimilarity task has been widely used in previous L2 speech research to measure non-native perception (e.g., Flege, Munro & Fox, 1994; Cebrian, Mora, & Aliaga-Garcia, 2011).

In addition to the target contrast /i - ɨ/, the following distractors were randomly included in the task: Russian /a - ɨ/, /o - a/, /i - e/, /e - o/, /a - u/, /ɨ - e/, /o - u/, and /u - ɨ/. Two female and two male native speakers of Russian recorded vowel contrasts in the /bVt/ context as part of a carrier sentence *Я сказал(а) _ / Я сказал(а) _ опять.* (I said _ / I said _ again). The selected tokens were extracted from the sentences and processed in *Praat* to normalize for amplitude (70dB) and remove AC. The tokens were normalized for length, which was determined by taking the average length of all tokens (360ms). The best tokens per speaker were selected based on auditory judgments and acoustic measurements.

The stimuli were organized into trials where each token A and X were spoken by the same or different individual(s) and presented with an inter-stimulus interval of 700ms. Each sound category was represented by at least five tokens to activate categorical perception in experienced listeners (L2 Spanish learners and Russian listeners). Alternating speakers and tokens ensured a more categorical nature of the task (Flege, 2003). That being said, naïve listeners (Spanish monolinguals) could still identify differences between tokens using the psychical (acoustic) properties of the sounds. To avoid the ordering effects, such as peripherality (Polka & Bohn, 1996) and nativeness – discrimination is facilitated when a native-like sound goes first (Ettlinger & Johnson, 2009), the token order within a pair was counterbalanced.

Participants were told that they would hear two vowels that might (or might not) sound familiar and that their task was to decide if the vowels sounded the same or different. Participants were asked to assess the difference between two vowels by marking the degree of mismatch on the intuitive rating scale (Jilka, 2009): the representation of the difference between two sounds corresponded to the least-to-most intensity of the red color (Figure 4).

Figure 4. The intuitive rating scale is reproduced from Jilka (2009).

Participants were instructed to respond by clicking a mouse or a touch panel alongside the scale and given 3000ms to respond. The next trial came immediately after the response or after 3000ms if no response was given. The experiment block began after participants had completed the practice block. The five test blocks consisted of eight change and eight no-change trials in a randomized order. Eight distractor trials were randomly included in each block.

The following measures were collected during this task: average numerical response from 1-10 to all instances of the target contrast and the standard deviation. For example, a score of 1 would represent poor discrimination of the contrast and a score of 10 perfect discrimination. The standard deviation would indicate the degree of consistency, with which a particular participant responded (smaller standard deviation indicating greater consistency).

3. Results

From the summary of performance across all tasks (Table 1) one can see that Spanish participants nearly excelled at the rated dissimilarity task (perceiving the difference between Russian /i/ and /ɨ/): the mean score for monolingual speakers was 8.24 (*SD* = 1.42) and for L2 learners was 6.92 (SD = 2.80): the difference between two groups was not statistically significant (an independent one-tailed Wilcoxon test: $W = 514$, $p = .07$). As we described earlier, in the rated dissimilarity task participants had to assess the difference between Russian /i/ and /ɨ/ by marking the degree of mismatch on an intuitive rating scale. During statistical analysis, for computational purposes, the intuitive scale was divided into ten equal sectors with 1 representing poor discrimination and 10 representing perfect discrimination. For each participant, we calculated an average response value across all instances of the target contrast /i- ɨ/.

Table 1: ***Summary of performance across all tasks***

Group	Task	Max score poss.	N	Min	Max	Mean	Median	SD
Spanish monolinguals	PSTM: *Serial Nonword Recognition Task*	144	31	43.00	137.00	91.03	89.00	23.54
	AM: *Target Sound Recognition Task*	72	31	31.00	66.00	52.71	54.00	8.93
	Non-native perceptual sensitivity: *Rated Dissimilarity Task*	10	31	5.16	10.00	8.24	8.78	1.42
Spanish L2 learners	PSTM: *Serial Nonword Recognition Task*	144	27	84.00	138.00	109.20	99.00	20.07
	AM: *Target Sound Recognition Task*	72	27	39.00	69.00	47.70	47.00	7.59
	Non-native perceptual sensitivity: *Rated Dissimilarity Task*	10	27	2.23	9.94	6.92	8.13	2.80

Even though the Spanish participants' performance on the rated dissimilarity task was rather high, the Russian controls did much better on the same task with the mean score of 9.86 (*SD* = 0.57. The difference between the two groups was significant (an independent one-tailed Wilcoxon test: $W = 923$, $p < .001$).

The preliminary correlational analysis suggested that there was no significant relationship between *PSTM* and *AM* in neither Spanish monolingual group (Pearson's $r = -0.075$, $p = .69$) nor L2 learners ($r = 0.12$, $p = 0.56$).

3.1. Spanish monolingual participants

Multiple regression analysis was used to test if *PSTM* and *AM* significantly predicted participants' ratings of the distance between two non-native sounds, *Non-native Perceptual Sensitivity*. In the first model, the predictors were *AM*, *PSTM*, *Age, Gender, Exposure to Russian*, *Phonetic training*, and the number of *L2s* studied in the past. This model explained 57% of the variance in *Non-native Perceptual Sensitivity* ($R^2 = 0.57$, $F(7,23) = 4.67$, $p < .002$). The effect of *PSTM* did not reach significance ($p = .52$), while *AM* contributed to *Non-native Perceptual Sensitivity* at the trend level ($\beta = 0.04$, $p = .09$) (Figure 5).

Other significant contributors were *Exposure to Russian* ($\beta = -2.35$, $p = .0009$), *L2s* studied ($\beta = -2.23$, $p = .02$), and *Age* ($\beta = .09$, $p = .04$). Previous *Phonetic training* boosted *Non-native Perceptual Sensitivity* by 0.84 at the trend level ($p = .099$). There was no effect of *Gender* on *Non-native Perceptual Sensitivity* ($p = .14$).

When non-significant predictors (*PSTM* and *Gender*) were taken out of the model, *AM* contributed to *Non-native Perceptual Sensitivity* at the trend level ($p = .056$). *Phonetic training* was no longer significant, while other predictors (*Exposure to Russian, Age, and L2s*) continued playing a role. Lastly, without *Phonetic training*, it was *AM* and *Exposure to Russian*, but not *L2s* studied and *Age*, that had a significant effect on *Non-native Perceptual Sensitivity.* This model explained 51% of the variance ($R^2 = 0.51$, $F(4,26) = 6.62$, $p = .0008$) and was better-fit than the parsimonious sub-model with only two predictors *AM* and *Exposure to Russian* (AIC = 98.8 vs AIC for the sub-model = 100).

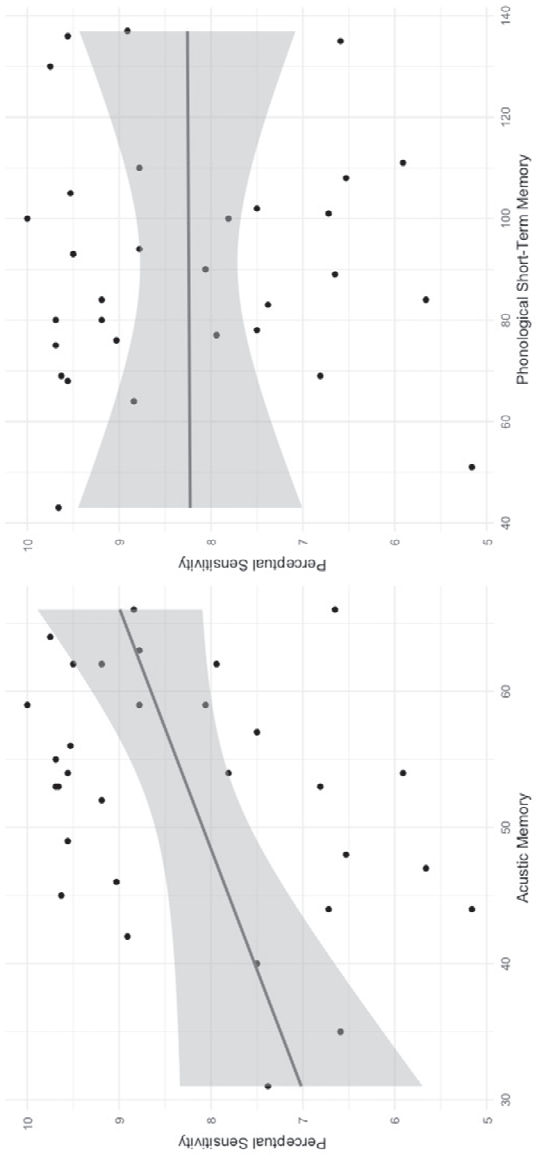


Figure 5. The relationship between *AM* and *Non-native Perceptual Sensitivity* on the left and *PSTM* and *Non-native Perceptual Sensitivity* on the right in Spanish monolinguals.

In our previous research (Kogan, 2020), we observed total monolinguals relying on *AM* more in a non-native perceptual task than monolinguals with L2 experience (even if this experience did not result in the above-beginner level of proficiency). However, in the present study, there was no interaction between *AM* and *L2*s, i.e. both total monolinguals and experienced monolinguals relied on *AM* in the same way.

The visual tests of the assumptions showed some deviation from normality. To confirm if the same predictors would remain significant, a beta regression model was run. This type of a generalized linear regression does not assume normality and linearity of residuals. A beta regression model also works well with a rating scale as a dependent variable (*Non-native Perceptual Sensitivity* in our case). All predictors came significant or at a trend level of significance with pseudo-R^2 equating to 0.51 (Table 2).

Table 2: ***The beta regression model coefficients with Non-native Perceptual Sensitivity as a response***

Predictor	β	SE	z-value	*p*-value
Acoustic Memory	0.006550	0.002481	2.640	0.008286**
Age	0.008258	0.004422	1.868	0.061811.
L2s studied	0.157677	0.083519	1.888	0.059038.
Exposure to Russian	-0.265286	0.070650	-3.755	0.000173***

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

3.2. Spanish L2 learners

The visual inspection of the relationships between *AM* and *Perceptual Sensitivity* and *PSTM* and *Perceptual Sensitivity* in Spanish L2 learners showed a positive trend for both types of memory (Figure 6). It seems like Spanish L2 learners rely on both *AM* and *PSTM* in determining the distance between two non-native sounds. To see if these relationships are significant, a series of multiple linear regressions were fitted with *AM*, *PSTM*, *Age, Gender, Exposure to Russian* (outside the class), *Phonetic training*, and the number of *L2s* studied as independent variables. This model explained 62% of the variance in *Non-native Perceptual Sensitivity* ($R^2 = 0.62$, $F(7,19) = 4.34$, $p < .005$). The effect of both *AM* and *PSTM* reached significance (correspondingly $\beta = .12$, $p = .046$ and $\beta = .06$, $p = .047$). *Gender* also contributed significantly with women perceiving greater distance between two non-native sounds in comparison to men ($\beta = 3.21$, $p = .046$). Further regression analysis revealed that men benefit more from improved *AM* in

comparison to women (β = 0.38, p = .007) (Figure 7), no such interaction was detected between *Gender* and *PSTM* (p = .27).

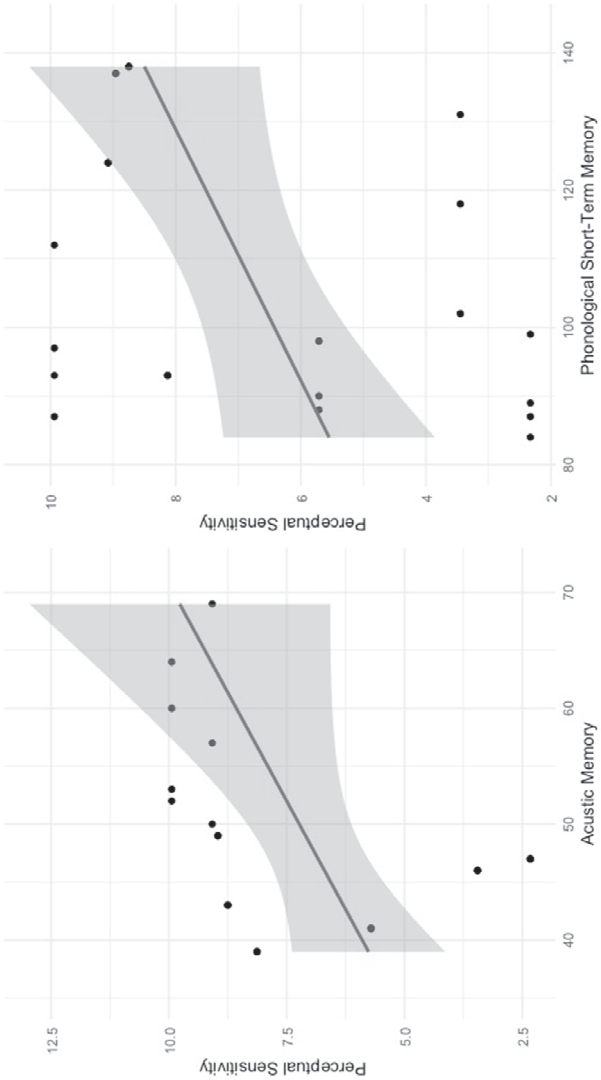


Figure 6. The relationship between *AM* and *Non-native Perceptual Sensitivity* on the left and *PSTM* and *Non-native Perceptual Sensitivity* on the right in Spanish L2 learners.

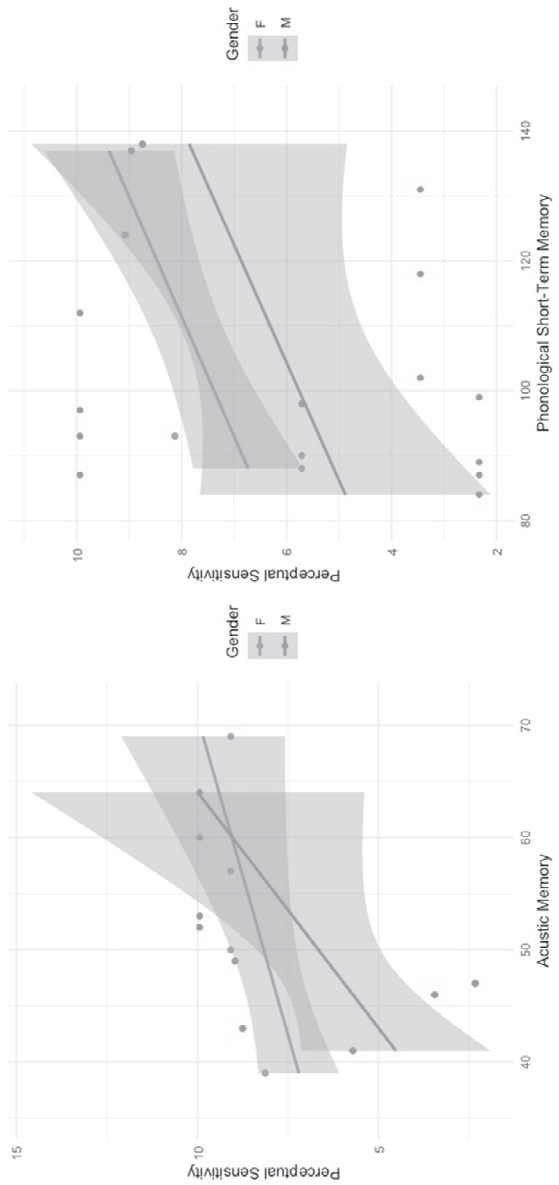


Figure 7. The contribution of *Gender* to the relationships between *Non-native Perceptual Sensitivity* and *AM*.

Interestingly, neither *Age* nor *L2s* studied (other foreign languages besides Russian) and *Exposure to Russian* outside the class reached the significance level for Spanish L2 learners as they did for Spanish monolinguals. Overall, both types of memory declined with age across the entire Spanish sample (monolinguals and L2 learners) but this trend was not significant (note that the participants older than 45 years old were excluded from the study following the assumption that audial perception declines with age).

A beta regression confirmed our best model with *PSTM*, *Gender*, and interaction between *AM* and *Gender* as predictors (pseudo-R^2 =0.56, $p <$ .001; Table 3).

Table 3: ***The beta regression model coefficients with Non-native Perceptual Sensitivity as a response***

Predictor	β	SE	z-value	*p*-value
PSTM	0.013419	0.003270	4.103	< 0.001***
Gender: Men	-3.351770	0.830557	-4.036	< 0.001***
AM & Gender: Men	0.063859	0.016999	3.757	< 0.001***

Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1

4. Discussion

Recent research has shown that PSTM and AM contribute to listeners' success in processing non-native sounds (Cerviño-Povedano & Mora, 2015, Ghaffarvand Mokari & Werner, 2019, Inceoglu, 2019, Lengeris & Nicolaidis, 2014, Safronova, 2016). Non-native perception is a challenging task: listeners have to handle novel sounds that do not exist in their native inventory. Prior L1 experience interferes with this process, which causes perceptual accent and impedes L2 sound acquisition (Best, 1995; Escudero, 2005; Flege, 1995; Kuhl, 2000). In the present study, Spanish listeners performed relatively well in perceiving the difference between Russian /i/ and /ɨ/ – most of the listeners, both monolingual speakers of Spanish and L2 Spanish learners of Russian, were able to perceive the acoustic distance between the sounds, i.e. they perceived them as two distinct sounds as opposed to two instances of the same sound. Despite the near-excellent performance, Russian native speakers demonstrated significantly better results when tested on the same task. The observed difference suggests that the Spanish participants' perception was influenced by the differences between their native and Russian phonological systems — this observation is supported by current L2 speech perceptual models (Best & Tyler, 2007;

Flege, 1995; van Leussen & Escudero, 2015). On the other hand, it seems that the ability to perceive subtle acoustic differences between non-native sounds is still accessible later in life. In other words, underlying sensory-perceptual mechanisms have not been permanently modified or lost due to prior (L1) linguistic experience and might pave the way to L2 category formation (Flege & Bohn, 2021).

An alternative explanation of the ceiling effect observed in our study is that the Russian /ɨ/ sound that does not have a counterpart in Spanish phonology has the acoustic properties salient enough for somewhat effortless perception. Flege (1987, 1992) observed that listeners identify "new" L2 sounds better than the sounds acoustically "similar" to their L1s. That being said, previous research on the non-native perception of Russian /ɨ/ by speakers of the languages where this sound does not exist has shown that listeners struggle to differentiate between Russian /i/ and /ɨ/ (Andreyushina, 2014; Shutova & Orekhova, 2018). Spanish learners are not an exception: there is a strong tendency to assimilate Russian /ɨ/ to either Spanish /i/ or /e/ (Klimova et al., 2017). The ceiling effect might be also explained by the nature of the task – listening to the sounds embedded in syllables is easier than listening to the same sounds in a more complex context (Best & Tyler, 2007).

In the present study, we did not find a connection between PSTM and AM indicating that they are autonomous concepts. This observation confirms previous research that has shown that two types of memory contribute to non-native perception independently from each other (Ghaffarvand Mokari & Werner, 2019, Safronova & Mora, 2012). Joseph and colleagues (2015) suggested that PSTM (and not AM) is more readily used when a greater cognitive load is present. When in their perceptual experiment Joseph and colleagues (2015) increased the number of phonemes to be maintained in the memory (increased load) participants relayed more on abstract categorical representations and not on sensory continuous properties of the stimuli. In our study, all listeners perceived sounds embedded in CVC syllables – a relatively manageable cognitive load. Thus, it came to no surprise that all listeners employed AM in this task (L2 learners also relied on PSTM in addition to AM). It would be plausible to assume that when attending to more complex natural-like stimuli (streams of connected speech), L2 learners will exclusively rely on PSTM and not AM. Investigating the relationships between PSTM and AM and their connection to the task of L2 perception remains a fruitful direction for future research.

4.1. Monolingual Spanish speakers

In the task of non-native sound perception monolingual Spanish speakers relied on AM, but not on PSTM. It means monolingual listeners relied on their ability to detect a physical (auditory) difference between a pair of stimuli, rather than judging the two sounds as instances of different categories. Acoustic memory supports a perceptual system in encoding acoustic detail at the pre-categorical level when it is necessary to recognize differences between the sounds which are not instances of two distinctive categories. Since none of our monolingual listeners were proficient in Russian, their vowel inventory lacked a phonetic category /ɨ/, which made it difficult to discriminate /ɨ/ vs /i/ based on the category-relevant dimensions.

Monolingual listeners who were exposed to Russian in their day-to-day environment did significantly worse on the task of non-native perception. This seemingly counterintuitive result replicates our previous findings where true monolinguals outperformed functional monolinguals on the same task (Kogan, 2020). It means that at the onset of language acquisition previous exposure to foreign languages might have a temporarily deteriorative effect on L2 perception. For example, in her computational simulations Lev-Ari (2017) demonstrated that at the early stage of L2 exposure, there is a tendency towards forming large sound categories, each compromised of several non-native categories. Basically, naïve listeners tend to assign novel sounds to their existing native phonetic categories, which results in enlarged native categories. Larger native categories might hinder the perception of non-native sounds (Kogan, 2020). Fortunately, this is just a temporary stage of L2 phonological acquisition. Lev-Ari (2017) showed that later, with the increased amount of L2 input, large native categories are split into smaller L2 categories improving L2 perception. In this sense, when compared to functional monolinguals (who had previous exposure to one or several L2s), true monolinguals (no previous L2 exposure) have a perceptual advantage since true monolinguals' native phonetic categories are more compact and not “contaminated” with diverse sound variants.

4.2. L2 Spanish learners

Spanish learners of Russian relied both on AM and PSTM in discriminating between Russian /i/ and /ɨ/. It means that their judgments were supported by both the ability to detect an auditory difference between the sounds and the ability to recognize /i/ and /ɨ/ as instances of two different phonetic categories, which is a more abstract level of processing (Joseph et al., 2015).

PSTM supports categorical perception of sounds meaning that listeners with greater PSTM are capable to cope with a higher degree of sound variability: production by various speakers, contextual variability, etc. Being able to differentiate between different L2 sound categories as opposed to between different L2 sound instances or variants is at the heart of L2 speech perception. Mastering L2 phonetic categories in perception facilitates overall L2 listening and supports L2 production (e.g., Flege & Bohn, 2021).

In our study, the L2 learners' gender came out to be a contributing factor to the task of non-native perception. Female listeners outperformed male listeners being able to recognize the difference between Russian /i/ and /ɨ/ significantly better. Previous research on the contribution of gender to the task of L2 perception is rather limited. It is generally believed that women are more successful language users than men (Maitland et al., 2004; Payne & Lynn, 2011; Weiss et al., 2003). Yet, it is still unclear at which aspects of language use (L1 and L2) females outperform males, and what role gender plays in L2 perception. It is possible that females demonstrate better L2 perception simply due to the possibility that female L2 learners get more L2 exposure than male L2 learners. This possibility is connected to gender-specific socialization and expectation patterns with women socializing more on average and listening more in conversations than men (e.g., Legato, 2005).

Interestingly, that for the task of non-native perception, male L2 learners in our study relied more on AM than female L2 learners. Previous research has shown that in comparison to females males are better at mimicking unfamiliar languages (Reiterer et al., 2011; Wucherer & Reiterer, 2018) – the ability that is heavily connected to AM.

4.3. General discussion

This study demonstrates that various cognitive abilities, namely various types of memory might be important at various stages of L2 acquisitional trajectory: whereas PSTM is likely to support L2 learners, it is AM that matters more for naïve listeners and total beginners. While PSTM has been researched extensively in its connection to L2 acquisition, it is still poorly understood how AM contributes to the tasks of L2 perception and learning. It seems that AM is important for the initial detection of fine acoustic differences between novel sounds (allophonic sensitivity) that take place at the onset of L2 acquisition and with naïve listeners. Total beginners and naïve listeners rely on AM and not PSTM when are first exposed to non-native phonology. In this regard, AM is particularly important as greater

AM might trigger different sound assimilation patterns and help form novel phonetic categories.

The relevant question is, of course, if AM can be trained. Incorporating AM training into the language classroom might prepare a beginner learner for consequent stages of perceptual learning when more categorical abstract processing takes place and the emphasis is moved towards detecting invariance. Perrachione and colleagues (2011) reported that high-variability phonetic training, the type of training that supports category formation, only benefits learners with enhanced perceptual abilities – in their study it was the perceptual ability to differentiate between various types of the pitch in unfamiliar Mandarin Chinese. Perrachione and colleagues (2011) do not directly employ the term "acoustic memory" but it is most likely that AM was employed in discriminating between novel pitch contours (a pre-categorical level of processing). Participants with poor perceptual abilities for pitch were disproportionately impaired in a high-variability training environment. In contrast, participants with enhanced perceptual abilities benefited greatly from high-variability training. It might be beneficial to train AM first, before employing high-variability training that focuses on PSTM training.

AM training might constitute an important step in preparing learners for the task of categorical perception. Because AM is responsible for processing non-linguistic stimuli, it is possible that AM training would be largely non-linguistic in nature: e.g., musical training, sound symbolism, and synesthetic activities (Christiner & Reiterer, 2015; Laroy, 1995; Wrembel, 2010). Further exploration of both the connection between AM and L2 perceptual learning and the associated with this connection pedagogical applications are critical directions for future research.

References

Andreyushina, E. A. 2014. Philological Sciences. Issues of Theory and Practice. Part 2. *Gramota Publ.* 11, 27–30.

Atkinson, R. C., and Shiffrin, R. M. 1968. Human memory: A proposed system and its control processes. In *Psychology of Learning and Motivation: Advances in Research and Theory*, edited by R.C. Atkinson and R. M. Shiffrin, 89-195. New York: Academic Press.

Baddeley, A. D. 1986. *Working memory.* Oxford, England: Oxford University Press.

Baddeley, A. D. 2000. The Episodic Buffer: A New Component of Working Memory? *Trends in Cognitive Sciences* 4(11), 417-423.

Baddeley, A. D., Eysenck, M. W., and Anderson, M. C. 2015. *Memory*. Hoboken: Taylor and Francis.

Best, C. T. 1995. A Direct Realist View of Cross-language Speech Perception. *Speech Perception and Linguistic Experience*, 171-206.

Best, C. T. 1994. The Emergence of Native-language Phonological Influences in Infants: A Perceptual Assimilation Model. *The Development of Speech Perception: The Transition from Speech Sounds to Spoken Words* 167(224), 233-277.

Best, C. T. 1993. Emergence of Language-specific Constraints in Perception of Non-native Speech: A Window on Early Phonological Development. In *Developmental neurocognition: Speech and Face Processing in the First Year of Life*, edited by B. de Boysson-Bardies, S. de Schonen, P. Jusczyk., P. McNeilage P., and L. Morton, 289-304. Springer, Dordrecht.

Best, C. T., and Tyler, M. D. 2007. Non-native and Second-language Speech Perception: Commonalities and Complementarities. In *Second language speech learning: The role of language experience in speech perception and production*, edited by M. J. Munro, and O.-S. Bohn, 13-34. Amsterdam: John Benjamins.

Boersma, P., and Weenink, D. 2013. Praat: Doing Phonetics by Computer [Computer Program]. Version 5.3. 51. *Online: http://www. praat. org*

Bongaerts, T., van Summeren, C., Planken, B., and Schils, E. 1997. Age and Ultimate Attainment in the Pronunciation of a Foreign Language. *Studies in Second Language Acquisition*, 447-465.

Brooks, P. J., Kempe, V., and Donachie, A. 2009, November. Individual Differences in Adult Foreign Language Learning. Paper presented at *the 50th Annual Meeting of The Psychonomic Society*, Boston, MA.

Carroll, J. B. 1981. Twenty-five Years of Research on Foreign Language Aptitude. In *Individual Differences, and Universals in Language Learning Aptitude*, edited by K.C. Diller, 119-154. Rowley, MA: Newbury House.

Cebrian, J., Mora, J. C., and Aliaga-Garcia, C. 2011. Assessing Cross-linguistic Similarity by Means of Rated Discrimination and Perceptual Assimilation Tasks. In *Achievements and Perspectives in the Acquisition of Second Language Speech: New Sounds 2010*, *Volume I*, edited by K. Dziubalska-Kołaczyk, M. Wrembel, and M. Kul, 41-52. Frankfurt am Main: Peter Lang.

Cerviño-Povedano, E., and Mora, J. C. 2015. Spanish EFLs' Identification of /i:-I/ and Phonological Short-term Memory. In *Proceedings of the 32nd International Conference of the Spanish Association of Applied Linguistics (AESLA): Language Industries and Social Change. Procedia*

– *Social and Behavioral Sciences* 173, edited by N. Jiménez Carra, E. Calvo, N. Fernández-Quesada, A.M. López Márquez, and A. Stender, 18-23.

Cerviño-Povedano, E., and Mora, J. C. 2011. Investigating Catalan Learners of English Over-reliance on Duration: Vowel Cue Weighting and Phonological Short-term memory. In *Achievements and Perspectives in the Acquisition of Second Language Speech: New Sounds 2010* edited by K. Dziubalska-Kołaczyk, M. Wrembel, and M. Kul, 53-64. Bern: Peter Lang.

Chao, S. C., Ochoa, D., and Daliri, A. 2019. Production Variability and Categorical Perception of Vowels are Strongly Linked. *Frontiers in Human Neuroscience* 13, 96.

Christiner, M., and Reiterer, S. M. 2015. A Mozart is not a Pavarotti: Singers Outperform Instrumentalists on Foreign Accent Imitation. *Frontiers in Human Neuroscience* 9, 482.

Coumel, M., Christiner, M., and Reiterer, S. M. 2019. Second Language Accent Faking Ability Depends on Musical Abilities, not on Working Memory. *Frontiers in Psychology* 10.

Council of Europe. 2001. *Common European Framework of Reference for Languages: Learning, Teaching, Assessment.* Cambridge, U.K: Press Syndicate of the University of Cambridge.

Cowan, N., and Morse, P. A. 1986. The Use of Auditory and Phonetic Memory in Vowel Discrimination. *The Journal of the Acoustical Society of America* 79(2), 500–507.

Crowder, R. G., and Morton, J. 1969. Precategorical Acoustic Storage (PAS). *Perception & Psychophysics* 5*(6),* 365-373.

Darcy, I., Mora, J. C., and Daidone, D. 2016. The Role of Inhibitory Control in Second Language Phonological Processing. *Language Learning* 66*(4),* 741-773.

Darcy, I., Park, H., and Yang, C. L. 2015. Individual Differences in L2 Acquisition of English Phonology: The Relation Between Cognitive Abilities and Phonological Processing. *Learning and Individual Differences* 40*,* 63-72.

Dogil, G., and Reiterer, S. M. 2009. *Language Talent and Brain Activity (Vol. 1).* Berlin and New York: Walter de Gruyter.

Duñabeitia, J. A., Cholin, J., Corral, J., Perea, M., and Carreiras, M. 2010. SYLLABARIUM: An Online Application for Deriving Complete Statistics for Basque and Spanish Orthographic Syllables. *Behavior Research Methods* 42(1), 118-125.

Ellis, R. 2012. *Language Teaching Research and Language Pedagogy*. John Wiley & Sons.

Escudero, P. 2005. *Linguistic Perception and Second Language Acquisition: Explaining the Attainment of Optimal Phonological Categorization*. LOT Dissertation Series 113, Utrecht University.

Ettlinger, M., and Johnson, K. 2009. Vowel Discrimination by English, French and Turkish Speakers: Evidence for an Exemplar-based Approach to Speech Perception. *Phonetica* 66(4), 222-242.

Flege, J. E. 1987. The Production of "New" and "Similar" Phones in a Foreign Language: Evidence for the Effect of Equivalence Classification. *Journal of phonetics* 15(1), 47-65.

Flege, J. E. 1992. The Intelligibility of English Vowels Spoken by British and Dutch Talkers. *Intelligibility in Speech Disorders: Theory, Measurement, and Management* 1, 157-232.

Flege, J. E. 1995. Second Language Speech Learning: Theory, Findings, and Problems. In *Speech perception and linguistic experience: Issues in cross-language research*, edited by W. Strange, 233-277. Baltimore, MA: York Press.

Flege, J. E. 2003. Assessing Constraints on Second-language Segmental Production and Perception. *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities* 6, 319-355.

Flege, J. E., and Bohn, O. S. 2021. The Revised Speech Learning Model (SLM-r). In *Second Language Speech Learning: Theoretical and Empirical Progress*, edited by S. Chen, 3-83. Cambridge University Press.

Flege, J. E., and Hammond, R. M. 1982. Mimicry of Non-distinctive Phonetic Differences Between Language Varieties. *Studies in Second Language Acquisition* 5(1), 1-17.

Flege, J. E., Munro, M. J., and Fox, R. A. 1994. Auditory and Categorical Effects on Cross-language Vowel Perception. *The Journal of the Acoustical Society of America* 95(6), 3623-3641.

Fujisaki, H., and Kawashima, T. 1970. On the Modes and Mechanisms of Perception of Speech sounds. *The Journal of the Acoustical Society of America* 47(1A), 57-57.

Ghaffarvand Mokari, P., and Werner, S. 2019. On the Role of Cognitive Abilities in Second Language Vowel Learning. *Language and Speech* 62(2), 260-280.

Grenon, I. 2006. Adults Still Have Direct Access to UG: Evidence from the Perception of a Non-native Feature Contrast. In *Proceedings of the 8th Generative Approaches to Second Language Acquisition Conference (GASLA 2006)*, edited by M. G. O'Brien, C. Shea, and J. Archibald, 51-62.

Hu, X., and Reiterer, S. 2009. Personality and Pronunciation Talent in Second Language Acquisition. In *Language Talent and Brain Activity*, edited by G. Dogil, and S. Reiterer, 97-130. Berlin and New York: Mouton de Gruyter.

Hummel, K. M. 2009. Aptitude, Phonological Memory, and Second Language Proficiency in Nonnovice Adult Learners. *Applied Psycholinguistics* 30(02), 225-249.

Inceoglu, S. 2019. Individual Differences in L2 Speech Perception: The Role of Phonological Memory and Lipreading Ability. *The Modern Language Journal* 103(4), 782-799.

Ioup, G., Boustagui, E., El Tigi, M., and Moselle, M. 1994. Reexamining the Critical Period Hypothesis: A Case Study of Successful Adult SLA in a Naturalistic Environment. *Studies in Second Language Acquisition* 16(1), 73-98.

Isaacs, T., and Trofimovich, P. 2010. Falling on Sensitive Ears? The Influence of Musical Ability on Extreme Raters' Judgments of L2 Pronunciation. *TESOL Quarterly* 44(2), 375-386.

Jilka, M., Anufryk, V., Baumotte, H., Lewandowska, N., Rota, G., and Reiterer, S. 2008. Assessing Individual Talent in Second Language Production and Perception. In *New Sounds 2007: Proceedings of the Fifth International Symposium on the Acquisition of Second Language Speech. Florianópolis, Federal University of Santa Catarina*, edited by A. S. Rauber, M. A. Watkins, and B. O. Baptist, 224-239.

Jilka, M. 2009. Assessment of Phonetic Ability. In *Language talent and brain activity*, edited by G. Dogil, and S. Reiterer, 17-66. Berlin: Mouton De Gruyter.

Jones, L. G. 1953. The Vowels of English and Russian: An Acoustic Comparison. *Word* 9*(*4), 354-361.

Joseph, S., Iverson, P., Manohar, S., Fox, Z., Scott, S. K., and Husain, M. 2015. Precision of Working Memory for Speech Sounds. *The Quarterly Journal of Experimental Psychology* 68(10), 2022-2040.

Klimova, U. A., Yurchenko, N.V., Cherkashina, O.M., and Kulik, S.S. 2017. Comparative Analysis of the Phonological Systems of the Russian and Spanish Languages (for the Purpose of Teaching the Russian pronunciation to Spanish-speaking Students). *Kazan Pedagogical Journal* 6(125).

Kogan, V. V. 2020. *The Effect of First Language Perception on the Discrimination of a Non-native Vowel Contrast: Investigating Individual Differences*. Unpublished Doctoral Dissertation. University of Barcelona, Barcelona, Spain. https://www.tesisenred.net/bitstream

Kormos, J., and Sáfár, A. 2008a. Phonological Short-term Memory, Working Memory and Foreign-language Performance in Intensive Language Learning. *Bilingualism: Language and Cognition* 11(02), 261-271.

Kuhl, P. K. 2000. A New View of Language Acquisition. In *Proceedings of the National Academy of Sciences 97*(22), 11850-11857.

Laroy, C. 1995. *Pronunciation*. Oxford University Press.

Legato, M. 2005. *Why Men Never Remember and Women Never Forget*. New York: Rodale.

Lengeris, A., and Nicolaidis, K. 2014. English Consonant Confusions by Greek Listeners in Quiet and Noise and the Role of Phonological Short-term Memory. In *Fifteenth Annual Conference of the International Speech Communication Association*, edited by H. Sak, A. Senior, and F. Beaufays.

van Leussen, V., and Escudero, P. 2015. Learning to Perceive and Recognize a Second Language: the L2LP Model Revised. *Frontiers in Psychology* 6, 1000.

Lev-Ari, S. 2017. Talking to Fewer People Leads to Having More Malleable Linguistic Representations. *PloS one* 12(8), e0183593.

Li, D., Cowan, N., and Saults, J. S. 2013. Estimating Working Memory Capacity for Lists of Nonverbal Sounds. *Attention, Perception, & Psychophysics* 75(1), 145-160.

MacKay, I. R., Meador, D., and Flege, J. E. 2001. The Identification of English Consonants by Native Speakers of Italian. *Phonetica* 58(1-2), 103-125.

Maitland, S. B., Herlitz, A., Nyberg, L., Bäckman, L., and Nilsson, L. G. 2004. Selective Sex Differences in Declarative Memory. *Memory & Cognition* 32(7), 1160-1169.

Martínez-Celdrán, E., Fernández-Planas, A. M., and Carrera-Sabaté, J. 2003. Castilian Spanish. *Journal of the International Phonetic Association* 33(2), 255-259.

Moyer, A. 2014. Exceptional Outcomes in L2 Phonology: The Critical Factors of Learner Engagement and Self-regulation. *Applied Linguistics* 35(4), 418-440.

Neisser, U. 1967. *Cognitive Psychology.* New York: Appleton-Century-Crofts.

Neufeld, G. 1987. On the Acquisition of Prosodic and Articulatory Features in Adult Language Learning. In *Interlanguage Phonology*, edited by G. Ioup, and S. Weinberger, 321–332. Cambridge, MA: Newbury House.

Payne, T. W., and Lynn, R. 2011. Sex Differences in Second Language Comprehension. *Personality and Individual Differences* 50(3), 434-436.

Perrachione, T. K., Lee, J., Ha, L. Y., and Wong, P. C. 2011. Learning a Novel Phonological Contrast Depends on Interactions between Individual Differences and Training Paradigm Design. *The Journal of the Acoustical Society of America* 130(1), 461-472.

Polka, L., and Bohn, O. S. 1996. A Cross-language Comparison of Vowel Perception in English-learning and German-learning Infants. *The Journal of the Acoustical Society of America* 100(1), 577-592.

Ramus, F., Peperkamp, S., Christophe, A., Jacquemot, C., Kouider, S., and Dupoux, E. 2010. A Psycholinguistic Perspective on the Acquisition of Phonology. *Laboratory Phonology* 10(3), 311-340.

Reiterer, S. M., Hu, X., Sumathi, T. A., and Singh, N. C. 2013. Are You a Good Mimic? Neuro-acoustic Signatures for Speech Imitation Ability. *Frontiers in Psychology* 4, 782.

Reiterer, S. M., Hu, X., Erb, M., Rota, G., Nardo, D., Grodd, W., Winkler, S., and Ackermann, H. 2011. Individual Differences in Audio-vocal Speech Imitation Aptitude in Late Bilinguals: Functional Neuro-imaging and Brain Morphology. *Frontiers in Psychology* 2, 271.

Safronova, E. 2016. *The Role of Cognitive Ability in the Acquisition of Second Language perceptual Phonological Competence*. Unpublished Doctoral Dissertation. Unpublished Doctoral Dissertation. University of Barcelona, Barcelona, Spain. Retrieved from https://www.tesisenred.net/bitstream

Safronova, E., and Mora, J. C. 2013. Attention Control in L2 Phonological Acquisition. In *Applied linguistics in the age of globalization*, edited by A. Llanes Baró, L. Astrid Ciro, L. Gallego Balsá, and R. M. Mateus Serra, 109-114. Lleida: Edicions de la Universitat de Lleida.

Safronova, E., and Mora, J. C. 2012. Acoustic and Phonological Memory in L2 Vowel Perception. In *Proceedings of 22nd EUROSLA*, 384-390.

Scott, S. K., Rosen, S., Beaman, C. P., Davis, J. P., and Wise, R. J. S. 2009. The Neural Processing of Masked Speech: Evidence for Different Mechanisms in the Left and Right Temporal Lobes. *Journal of Acoustic Society of America* 125(3), 1737-1743.

Shutova, M. N., and Orekhova, I. A. 2018. Phonetics in Teaching Russian as a Foreign Language. *Russian Language Studies* 16(3), 261-278.

Skehan, P. 2015. Foreign Language Aptitude and its Relationship with Grammar: A Critical Overview. *Applied Linguistics* 36(3), 367–384.

Tourville, J. A., and Guenther, F. H. 2011. The DIVA Model: A Neural Theory of Speech Acquisition and Production. *Language and Cognitive Processes* 26(7), 952-981.

Weiss, E. M., Kemmler, G., Deisenhammer, E. A., Fleischhacker, W. W., and Delazer, M. 2003. Sex Differences in Cognitive Functions. *Personality and Individual Differences* 35(4), 863-875.

Williamson, V. J., Baddeley, A. D., and Hitch, G. J. 2010. Musicians' and Nonmusicians' Short-term Memory for Verbal and Musical Sequences: Comparing Phonological Similarity and Pitch Proximity. *Memory & Cognition* 38*(*2), 163-175.

Wrembel, M. 2010. Sound Symbolism in Foreign Language Phonological Acquisition. *Research in Language* 8, 1–14.

Wucherer, B. V., and Reiterer, S. M. 2018. Language is a Girlie Thing, isn't it? A Psycholinguistic Exploration of the L2 Gender Gap. *International Journal of Bilingual Education and Bilingualism* 21(1), 118-134.

Yanushevskaya, I., and Bunčić, D. 2015. Russian. *Journal of the International Phonetic Association* 45(2), 221-228.