The effects of English proficiency on the processing of Bulgarian-accented English by Bulgarian-English bilinguals

MARIE DOKOVOVA

A thesis submitted in partial fulfilment

of the requirements for the degree of

Doctor of Philosophy

QUEEN MARGARET UNIVERSITY

2019

Declaration

I confirm that the thesis submitted is my own work and that appropriate credit has been given where reference has been made to the work of others.

Marie Dokovova 30 July 2019

Acknowledgements

When I was starting this course, I was warned by many that the PhD time can be an excruciating experience. Fortunately for me, I was surrounded by people who didn't let that happen and I want to thank them all.

Firstly, I want to thank my supervisors Robin Lickley and Jim Scobbie, who make the best supervisory team and guided me through the treacherous waters of decision-making without imposing their point of view. Thank you for always being so kind and understanding and turning our meetings into something to look forward to!

Thank you to Steve, for always responding to my bizarre technical requests and for always being there to talk to. Special thanks to Anna, Klara, Jocelynne, Velina, Gijsbert Stoet and my generous participants for making my studies possible! I am also eternally grateful to Bodo Winter, Márton Sóskuthy and Timo Roettger for teaching me everything I know about statistics.

They say it takes a village to raise a child, but in my case, it took a labful of people and lots of additional support to raise this PhD student. I can’t imagine being able to get through without the friendships and the supportive atmosphere in the Speech and Hearing department of QMU, the Speech Lab and all the visiting students, my PhD Flat family, and my dance, art, and cow buddies, and Monika. I’ll completely use up my word limit if I were to list you all. Plus, if you’re reading this, you know who you are (except Monika maybe). Thank you for distracting me, for dancing, painting, taking, and laughing with me, for feeding me and my soul, and for not asking me how the PhD is going!

Finally, thank you to my family and Michael, for always being there for me and doing all of the above and more.

Abstract

This dissertation explores the potential benefit of listening to and with one's first-language accent, as suggested by the Interspeech Intelligibility Benefit Hypothesis (ISIB). Previous studies have not consistently supported this hypothesis. According to major second language learning theories, the listener's second language proficiency determines the extent to which the listener relies on their first language phonetics. Hence, this thesis provides a novel approach by focusing on the role of English proficiency in the understanding of Bulgarian-accented English for Bulgarian-English bilinguals.

The first experiment investigated whether evoking the listeners' L1 Bulgarian phonetics would improve the speed of processing Bulgarian-accented English words, compared to Standard British English words, and vice versa. Listeners with lower English proficiency processed Bulgarian-accented English faster than SBE, while high proficiency listeners tended to have an advantage with SBE over Bulgarian accent.

The second experiment measured the accuracy and reaction times (RT) in a lexical decision task with single-word stimuli produced by two L1 English speakers and two Bulgarian-English bilinguals. Listeners with high proficiency in English responded slower and less accurately to Bulgarian-accented speech compared to L1 English speech and compared to lower proficiency listeners. These accent preferences were also supported by the listener's RT adaptation across the first experimental block.

A follow-up investigation compared the results of L1 UK English listeners to the bilingual listeners with the highest proficiency in English. The L1 English listeners and the bilinguals processed both accents with similar speed, accuracy and adaptation patterns, showing no advantage or disadvantage for the bilinguals.

These studies support existing models of second language phonetics. Higher proficiency in L2 is associated with lesser reliance on L1 phonetics during speech processing. In addition, the listeners with the highest English proficiency had no advantage when understanding Bulgarian-accented English compared to L1 English listeners, contrary to ISIB.

Keywords:

Bulgarian-English bilinguals, bilingual speech processing, L2 phonetic development, lexical decision, proficiency

Dedication

To my fellow Bulgarians, living in the UK.

Chapter 1 Table of Contents

List of Figures

List of Tables

Nomenclature

ANOVA	analysis of variance
ASP	automatic selective perception
BA	Bulgarian accent
BIMOLA	model of bilingual lexical access
BNC	British National Corpus
CI	confidence interval
DPH	Different Processes Hypothesis
EEG	electroencephalogram
EFL	English as a foreign language
ERP	event related potential
fo	fundamental frequency
F1	first formant
F2	second formant
FA	foreign accent
FAS	foreign accentedness score
GAMM	generalised additive mixed model
Hz	hertz
IAF	Ideal Adapter Framework
ISIB-L	Interspeech Intelligibility Benefit for listeners
ISIB-T	Interspeech Intelligibility Benefit for talkers
L1	first language
L2	second language
MMN	mismatch negativity

ms	milliseconds
NA	native accent
PAM	Perceptual Assimilation Model
PDH	Perceptual Distance Hypothesis
RLS	reverse linguistic stereotyping
RP	Received Pronunciation
RT	reaction time
SD	standard deviation
SLM	Speech Learning Model
SPR	selective perception routine
SBE	Standard British English
TOEFL iBT	Test of English as a Foreign Language internet-based test
VOT	voice onset time

Introduction

Speech perception depends on a multitude of factors, such as the listening environment, the speaker, and the listener. All three components are instrumental for successful communication. Though a lot of knowledge has been amassed for the independent components, there still is no complete and universal model of speech processing that can account for the behaviour of all listeners in all possible environments and with all possible types of speech.

One of the reasons for that is that existing models of speech processing only focus on a few aspects at a time of the components listed above for the purpose of clarity. One of the components that many models of speech processing assume as stable, is the listener themselves. Models of speech processing are usually designed with a specific listener in mind: often a native listener of a language with typical cognitive development. Of course, models have been devised for listeners with different backgrounds (Alcántara et al., 2004; Golden et al., 2015; Jepsen, 2010) but there is no unifying approach to speech processing that combines all possible listeners. It is argued here that general models of speech perception can be challenged and improved by taking the perspective of less typical listeners.

One of the listener groups that needs to be better understood is that of non-native listeners, here defined as those who have first acquired a different language from the one they are trying to understand. Though there are models targeting such listeners, they often do not take into account the fact that bi- (or multi-) linguals might display completely different behaviours as a result of their proficiency in their second language. In addition, most models of L2 listening do not make explicit predictions about how L2 listeners cope with accent variability. Hence, there are few studies investigating how multilinguals process different accents in their non-native language (henceforth L2).

Non-native speakers and listeners are a frequent subject of psycholinguistic and phonetic investigation. The literature review in the following chapters reveals that there are two well-developed branches of psycholinguistics, concerning L2 users: processing of L2 accent vs. native accent by native listeners; and native speech processing by L2 listeners.

There is a gap in the literature regarding the crossover of these two sub-branches: how do non-native listeners process foreign accented speech? Specifically, how do non-native listeners process L2 speech produced by other non-native speakers with a similar background as them? This is the central question addressed by this dissertation.

This dissertation investigates how L2 listeners process the words pronounced by speakers of a similar linguistic background as them and who speak with their shared L1 accent. Current models of speech perception do not make specific predictions whether L1-accented L2 speech would be processed more easily and efficiently (i.e., with a processing benefit) by L2 listeners of varying proficiency compared to other accents, particularly prestige native accents. A processing benefit could result from a phonetic similarity with their native language. However, a phonetic similarity to L1 could also impair processing due to the accent's divergence from the standard L2 variety taught in formal environments. It has also not been widely investigated whether the phonetic similarity between L1 and L1-accented L2 can facilitate the processing of L2 vocabulary. These will be the main questions pursued in this thesis.

These questions will be explored from the perspective of bilingual Bulgarian residents in the UK. According to the Office for National Statistics (Blake, 2018), from July 2017 to June 2018 there were 93,000 people (CI +/- 14,000) of a Bulgarian nationality residing in the UK. Of them only about 3000 were born in the UK, suggesting that the majority are first generation immigrants, likely to speak with Bulgarian-accented English. The number of Bulgarians in the UK has been steadily increasing since 2007 when Bulgaria joined the EU and 2014 when the UK lifted the restrictions on the Bulgarians' right to work. Possibly due to the recency of this immigration wave, there are not yet phonetic or psycholinguistic studies known to the author involving Bulgarian-English bilinguals and Bulgarian-accented English.

By assessing this population's responses to Bulgarian-accented English compared to Standard British English, this dissertation will not only fill gaps in general research of speech perception, but it will also be relevant for the practical issues late bilinguals might face when adapting to a new country of residence. Standard British English (SBE, also known as RP or SSBE) has been chosen as a baseline comparison variety because of its assumed salience to Bulgarian L2 learners of English, specifically, and to residents within the UK generally.

This dissertation starts with a literature review, which expands over four chapters. Chapter 1 introduces different approaches of conceptualising general speech processing, putting an emphasis on exemplar approaches. Chapter 2 focuses on models designed for L2 processing and tries to relate them to the general models discussed in Chapter 1. Chapter 3 discusses the benefits and challenges of listening to a foreign accent. Chapter 4 focuses on the phonetic, and sociolinguistics characteristics of Bulgarian accent in English. Together this discussion leads to the overview of experiments laid out in Chapter 5. Chapters 6 to 8 are dedicated to the empirical testing of the hypotheses. Lastly, Chapter 9 revisits the hypotheses and questions raised in the literature review and synthesises the main conclusions and opportunities for future research.

Chapter 1 General models of speech processing

The problem of how listeners process spoken language is incredibly complex, but, as mentioned in the introduction, it can be stripped down to three main ingredients: the characteristics of the acoustic signal that is being processed relative to the speaker, the characteristics of the listeners (language experience and psychophysiological state), and the environment in which the listening occurs.

Due to the complexity of the process and the large number of factors that can affect it, most models and frameworks, trying to account for language processing, only focus on a limited number of factors at a time. As the creation of a model or a framework calls for simplicity and predictive power (which distinguishes a model from a description), gaps in existing models should not be treated as flaws, but rather as opportunities for development of competing or complementing models. This is the spirit with which the following literature review will be presented.

The present study mainly focuses on two of the aspects mentioned above: the characteristics of the acoustic input (Bulgarian speech, Bulgarian-accented English speech and native Standard British English speech) and the characteristics of the listeners (here, the linguistic experience of Bulgarian –English bilinguals, and native English listeners).

There is a lot of disagreement in the study of human perception about the stages and basic elements involved in speech processing. Probably the most wide-spread, but certainly not unchallenged view (cf., Coleman, 2002; Hawkins and Smith, 2001; Wade and Möbius, 2010), is that during processing speech is segmented into phoneme-sized units (Marslen-Wilson and Welsh, 1978; McClelland and Elman, 1986; Norris and McQueen, 2008). A phoneme is usually considered to be the smallest segment of sound which, when changed, can lead to a difference in meaning in a word (although see Ladd, 2006). It is generally agreed that a listener must be able to analyse the continuous acoustic signal into some form of units that allow the activation of the correct lexical representation, which is stored in the memory (Stevens, 2002). Most models reviewed in this dissertation adhere to this basic structure of speech perception, although there are substantial differences between the scope, architecture and basic elements they consider.

1.1. Traditional models

Traditional views of speech perception assume that minimal segments are decoded one after the other from the input (Halle, 1985). It has been posited that the human perceptual system has feature detectors (Abbs and Sussman, 1971) that can decode the acoustic stream into different features, characteristic of each phoneme, called distinctive features (Jakobson et al., 1952). A bundle of relevant features activates a phoneme and a group of phonemes activate a relevant neighbourhood of words that share these phonemes (Luce and Pisoni, 1998; Marslen-Wilson and Welsh, 1978). As more and more phonemes are decoded, the neighbourhood of candidates becomes smaller until only the right word candidate is available for selection. There is usually little discussion regarding the listener characteristics and their background.

In very broad terms, this is the structure of the Cohort model (Marslen-Wilson and Welsh, 1978), TRACE model (McClelland and Elman, 1986), Neighbourhood model (Luce and Pisoni, 1998), Merge (Norris, 1999), Shortlist (Norris and McQueen, 2008), inter alia. While these models differ in many important ways, which are beyond the scope of this discussion, they all share the assumption that the incoming speech is segmented into phoneme-sized units as it unfolds. This also suggests that a lot of acoustic information which is not directly necessary for the identification of a segment (i.e., which is not a distinctive feature of a phoneme) does not affect the process of word recognition but may provide other types of information. In practice, this type of "redundant" information could indicate the identity of the speaker, their physical state and even intonation. These models assume that all adult speakers of a language have stored the relevant lexicon and phonemic system of their native language in their memory and as long as the input contains the minimal necessary acoustic information to identify the message, the listeners should be able to decode the input. These models do not consider that some adult speakers are illiterate and have low phonological awareness but can still be successful communicators (Loureiro et al., 2004; Morais et al., 1979).

Of course, this is a very brief summary of the traditional approach to speech processing, which cannot do it full justice. Recent takes on this approach have a remarkable resemblance with aspects of the episodic and predictive models discussed in the following sections. For instance, Calabrese (2012) postulates the importance of forward top-down prediction in speech processing, the embodied nature of phonemic features, the long-term availability of detailed echoic memory ("where faithful auditory representations of acoustic inputs are stored" (Calabrese, 2012, p. 356) and the fact that different speech decoding procedures might be involved in different situations. Nevertheless, Calabrese takes the position that the purpose of "exponents", or the acoustic input, is to transmit its lexical meaning, and predictably this model insists on the lack of effect of sociolinguistic and idiosyncratic phonetic detail on speech comprehension.

Despite this position, Calabrese's model is notable for representing a "traditional" approach in speech processing and also for considering the point of view of listeners who are not expert in the type of speech they are hearing. This is achieved by discussing how foreign words are processed, produced and changed phonetically to become integrated in the listeners' L1. One of the phenomena related to the adaptation of loanwords is the "phonological illusion" effect. In his paper Calabrese's (2012) argues that the existence of phonological illusions, in which listeners report hearing a phoneme that was not acoustically present (Peperkamp and Dupoux, 2003) is strong evidence that listeners apply top-down phonological representations in the process of analysing unfamiliar speech signal. A specific neurological mechanism for this type of phenomena will be discussed in Section 1.3.2 and 1.3.3.

In Calabrese's (2012) model phonological features are considered to be the basic component of speech processing and phonemes are just bundles of features. However, unlike some other models, which use features as entirely abstract concepts, he defines features as the correlation between quantal acoustic phenomena and the articulatory patterns that may have produced them. In order to account for the listening of both expert listeners, accustomed to their preferred L1 variety and naïve listeners in the process of borrowing a loanword, Calabrese (2012) borrows the concepts of "phonological" and "phonetic" listening from Werker and Logan (1985). These two concepts will be further explained in Section 2.2.2.

Calabrese (2012) adds a novel layer to the traditional models of speech processing and accounts for situations where the speech signal is less predictable (e.g., novel combinations of familiar words or an unfamiliar accent). Thus, he marks a welcome departure from most influential traditional models, which are concerned only with expert native listeners.

1.2. Exemplar models

Alternatively, it has been proposed that incoming speech is not parsed down through such a strict hierarchy, which requires the discarding of a lot of phonetic detail. Exemplar models of speech perception assume that speech is not necessarily segmented into phonemes but that larger units of sound (e.g., syllables, whole words, intonational phrases, or individual phonetic cues) directly activate memory traces of speech stored in whole with all their detail (Coleman, 2002; Goldinger, 1998, 2000; Hawkins and Smith, 2001; Johnson, 1997; Semon, 1923). The exact details of the processing units depend on the listening task and the listener's experience.

Indeed, there is a lot of empirical support that listeners remember specific instances of voices and make use of fine phonetic information when processing speech (Goldinger, 1996; Shockley et al., 2004, inter alia). For example, if listeners were not affected by fine phonetic detail, they would process speech of the same language at the same speed as long as it contains all the distinctive acoustic detail of segments. However, this has been disproved empirically (Clopper, 2017; Kim, 2016). The evidence suggests that listeners perceptually adjust to the idiosyncrasies of individual speakers and that switching between speakers of the same variety and gender slows down speech perception (Mullennix et al., 1989). Further evidence for this point can be found in Section 3.3 in the discussion of the processing of different sociolects.

While the more traditional psycholinguistic models of perception focus on the ability of listeners to perceive invariant categories, such as features and phonemes, within a variable signal, episodic models can make predictions about the effects of category-independent phonetic details in the input. The main advantage of episodic modes is also the main drawback: on the surface such models might struggle to explain why speakers generalise past knowledge on neologisms for which they have no stored episodic memories. For example, Pierrehumbert (2016) suggests that American speakers would produce the neologism "macket" with [ɾ] in some contexts (such as "My *macket* is all wet now"). A counterargument is that in the case of neologism production, generalisations can be supported by the spelling or familiar components smaller than a word. In addition, this argumentation, as well as the evidence of phoneme-like segments from speech errors (Fromkin, 1971), might suggest some compositionality within word representations for production but they do not directly address the units of speech processing. Nevertheless, some models of speech perception, often called hybrid models, try to combine the advantages of both types of models (Cai et al., 2017; Johnson, 1997; Jusczyk, 1993; Kleinschmidt and Jaeger, 2015a; Pierrehumbert, 2002; Wade and Möbius, 2010).

One way of addressing compositionality in perception and production is offered by Pierrehumbert (2002)'s hybrid model. It assumes that there are traces of categories within categories, where a trace is a memory of an exemplar of that specific category. For instance, the word 'night' might have exemplars stored as sub-components of the phrase 'good night' and the nucleus of 'night' might cluster with traces of other nuclei, for example from 'fight' or 'light'. How central an exemplar is, depends on a combination of factors, such as how frequently it is encountered or whether it is the first exemplar of that category to be encountered (primacy).

Depending on the listening task some of these categories can play a smaller or a bigger role in processing and can be updated or not (Kraljic et al., 2008).

One of the predictions that arise from episodic models is that categories rich with exemplars (such as highly frequent words) would be easier to recognise but also harder to adjust in the face of novel input of sufficient sociolinguistic importance compared to categories with fewer exemplars. Based on the general support of this hypothesis (Goldinger, 1998; Kraljic et al., 2008; Nielsen, 2011), another prediction can be made, namely that listeners who have more experience with a specific variety and its phonetic system would process that variety faster than listeners with less experience, but would be slower to adjust the boundaries of the categories that drive their speech processing (Lev-Ari, 2017). A detailed explanation on the mechanism of perceptual recalibration will be discussed in Section 1.3.1. These predictions can be tested in the context of L2 processing, where listeners with higher proficiency, being more experienced, would be expected to process speech faster and more correctly than lower proficiency listeners but may also need time to adjust to an unusual accent.

Related to this is the "perceptual magnet" effect observed by Kuhl (1991). Exemplars closer to the prototype are harder to discriminate than exemplars at the fringes of categories. This phenomenon implies that examples of speech tend to naturally cluster together as categories, within which some examples are easier to process than others. If exemplars at the edges of a category are more easily discriminable, there is also a higher chance that they are misperceived as examples of a different categories and inhibit speech processing. Hence processing examples of speech closer to the prototypes is expected to be easier and more precise than processing speech at the boundaries of a category. For instance, a foreign accent would constitute of many examples of sounds that fall at the fringes of native listeners' representations and thus lead to problems with intelligibility. This concept underlies the bilingual Perceptual Assimilation Model (Best and Tyler, 2007) and the Ideal Adapter Framework (Kleinschmidt and Jaeger, 2015a), discussed in the following sections.

An episodic approach to speech processing can benefit from a parallel to the philosophical work on the organisation of concepts. The context model proposed by Medin and Schaffer (1978) originally aimed to address ambiguity in concept representations, can be also be applied for phonetic representations of speech. Their model involves some level of abstraction (such as a summary description), which is based on observation, is not innate and can be updated in light of novel evidence. Exemplar models need not be extreme, like the Proximity model (Reed, 1972), which assumes that only instances are stored in memory without any level of abstraction or analysis. According to Medin and Schaffer (1978) categories have several focal exemplars based on primacy. Novel input is scored (metaphorically) based on its resemblance to the focal exemplars and the category can be reevaluated, keeping only high-scoring exemplars. Despite the importance of early examples, this mechanism can allow for a radical change in the prototypes, given a gradual but radical change in the input (for instance emigration of the listener). This might account for how learning and updating categories can happen in the face of novel evidence.

To summarise, episodic models of speech processing assume that the listeners' prior experience with the phonetic variability found in the input directly influences their speech processing. The exemplar models mentioned here do not directly address the problem of listening to multiple languages and the extent to which one or more prototypes of the same declarative category can be maintained in parallel. This issue will be brought up again in Chapter 2.

1.3. Dynamic approaches of speech processing

Of the two approaches to speech processing discussed so far, the exemplar models provide a more dynamic perspective, as they choose the presumably ever-changing listener experience to be at the core of speech processing phenomena. However, the models mentioned in the last section do not make specific predictions about the moment-to-moment adaptation to the speech signal. In order to gain a full understanding of how listeners develop multilingual speech processing systems it is also important to account for the effect of acoustic input on their listening. This is the reason why this section focusses on two dynamic approaches to speech adaptation the Ideal Adapter Framework (Kleinschmidt and Jaeger, 2015a) and the Thousand Brains Intelligence Model (Hawkins et al., 2019).

Before continuing, it is important to state how the concept of phonetic representations will be treated. This dissertation takes a dynamic and embodied approach. In Section 1.3.2 it will be argued that phonetic representations are interlinked with representations of other modalities and have a physiological reality in the brain. In this dissertation the word representation is used to stand for connective links between neurons that would tend to be (pre-) activated together in response to external stimuli. For instance, a phonetic representation of a word is the implicit expectation a listener has of what that word might sound like. This expectation can be preactivated based on the information the listener has about the speaker, the listening environment and perhaps the sentence they are currently hearing.

The present definition of phonetic representation puts the concept of prediction, expectation and episodic memories from past experiences at its core. This approach is in line with the Ideal Adapted Framework introduced in the following section.

1.3.1. Ideal Adapter Framework (IAF)

An example of a predictive approach to speech processing is the Ideal Adapter Framework proposed by Kleinschmidt and Jaeger (2015a). It is based on the Bayesian theorem, but unlike most models mentioned in this thesis it focuses on the perception and weighting of individual acoustic cues, with respect to a predicted phoneme, as opposed to word recognition. The Bayesian theorem states that the posterior probability of an outcome is proportional to the prior probability of that outcome multiplied by the likelihood that what is being observed is, in fact, the outcome (Dienes, 2008). Kleinschmidt and Jaeger (2015a) argue that this formulation is an accurate representation of how listeners adapt to within and cross-speaker phonetic variation, which will be explained further down.

The main advantages of the Ideal Adapter Framework are that it accounts for the dynamic nature of phonetic representations and provides a specific mechanism of combining prior experience with the speaker-specific acoustic cue distributions. Its disadvantage is that it appears to adopt a fairly traditional segmental position by assuming that speech processing is a serial decoding of phoneme-sized units.

IAF is based on an ideal adapter listener engaged in an optimal listening process. Optimal listening assumes that the listeners use all the evidence available to them consistently, to determine the posterior probabilities of different interpretations of the input. The interpretation with the highest probability is the winner, but new evidence can be added at any point, leading to reinterpretation. In order to form these posterior probabilities, listeners use their prior experiences with the statistical distributions of phonetic units or frequencies of words, forming the prior factor in the formula. The likelihood factor is formed by using both the average and the spread of the distributions of phonetic cues, in the context of structured knowledge about different speakers, accents and contextual information (Kleinschmidt and Jaeger, 2015a).

As an example, the model allows to estimate the probability that a native English listener will interpret the word "sit" (i.e., form a high posterior probability for that interpretation), given the acoustic input [sɪʔ] in different syntactic contexts. This posterior probability for "sit" is proportional to the likelihood that [sɪʔ] is heard when "sit" is intended (based on the listener's expectation that the sounds [sɪʔ] can stand for /sɪt/) multiplied by the probability of hearing the word "sit" in that context (using their knowledge of English word frequencies, syntax and pragmatics).

Posterior probability (of "sit") ~ Likelihood (of hearing [sɪʔ] when "sit" is intended) * Prior probability (of speaker intending "sit")

Using this formula, Kleinschmidt and Jaeger (2015a) propose two ways in which adaptation to phonetic variation can happen. They are called phonetic recalibration and selective adaptation, or positive and negative after-effects. Recalibration, or positive after-effect, is observed when, after hearing a category of sounds with slightly unusual phonetic cues, listeners expand their category boundaries to include the less representative examples of that category (demonstrated in Norris, 2003). For example, after hearing a few examples of dental [d] a listener might expand their perceptual category that normally includes only alveolar [d] and accept the dental exemplars. Selective adaptation, or negative after-effects, describes the process where, after prolonged exposure to a similar examples of a sound (e.g., many dental [d] and no alveolar exemplars), listeners shift their perceptual category boundary, so that it includes fewer examples (i.e., some extreme alveolar [d] exemplars might be excluded) and is narrowed around the exemplars of the sound they had just heard (Vroomen et al., 2007). For example, after a prolonged exposure to approximately 100 ms VOT of [t] English listeners, who would initially expect normal distribution around 70 ms will narrow their expected mean and standard deviation of VOT values closer to 100 ms value and might even shift their phoneme boundary between /d/ and /t/ to a higher value.

In order to explain these two phenomena, Kleinschmidt and Jaeger (2015a) assume that listeners are aware of both the mean and the variance of a specific cue distribution. As a result, if the listeners hear one or two occurrences of an ambiguous sound, the resulting input for that phonetic cue would have both an unusual mean and an unusual variance compared to the internal representation, which would lead to positive after-effects (recalibration, expansion of boundaries). However, if listeners are exposed to the same unusual cue for a prolonged period, they are expected to experience recalibration, followed by selective adaptation (i.e., initial expansion to include the new exemplar followed by negative after-effect or shrinking of boundaries around the new example). This would occur because the variance for that cue lowers, as the listeners hear the same input again. These are the effects observed in Vroomen et al. (2007) and Kleinschmidt et al. (2012). This, Kleinschmidt and Jaeger (2015a) take as evidence that listeners are aware of the underlying distributions of phonetic cues. These changes in the expected variability of phonetic cues affect the likelihood factor of the Bayesian formula.

Unfortunately, the authors do not elaborate on how this system can be used for second language processing. However, there are several aspects of the models that can be useful in conceptualising how listening in a non-native language may operate. First, IAF makes provisions for listening to unusual, or hard to predict phonetic input. Second, IAF allows for the listeners' expectations to be strongly or weakly biased in what input they may expect.

As stated earlier, for optimal listening to occur the listener must have a representation of the intended message, associated with its acoustic input, otherwise there may be a breakdown in communication. A listener must know that in some cases /t/ is realised as a [ʔ] or they will fail to process the message if the context is not sufficiently biasing. The model predicts that short-term adaptation (within a single session of listening) will be hampered in cases when listeners are confronted by a speaker whose acoustic cues range falls far outside the cue distributions the listener is used to. This has been demonstrated in cases of American English listeners failing to adapt to prevoiced stops (Sumner, 2011) or incongruent voice onset time and f0 cues (Idemaru and Holt, 2011). Hence, a second language listener, or another novice listener who is hearing unexpected phonetic input which does not automatically link to a representation, may need to rely a lot more on context to reconstruct the message, hence taking more time and perhaps having to start over (cf., Song and Iverson, 2018).

As mentioned above, the model makes provisions for listeners who have small and larger experiences with mental models of different voices, speakers and accents. Kleinschmidt and Jaeger (2015a) refer to them as narrow and flexible listeners. Narrow listeners choose between fewer interpretations of the input than flexible listeners, due to their past experiences and knowledge or because they are in a more biasing situation. In IAF listeners have multiple speech models (in this context a model can correspond to any level of categorisation, such as word, morpheme, syllable, phone) available to them a priori. In IAF the listeners have a limited mass of probability that they can assign to each model (all posterior probabilities have to add up to one). Thus, more flexible listeners who have a greater diversity of a priori models, can assign less confidence to each at the start of listening because the same input could support multiple models. Listeners with more limited model inventories would assign higher probabilities to each starting model because they have fewer options. If the new acoustic input matches any of the existing models in the inventories of the flexible and the narrow listeners, the narrow listeners will need less additional input in order to reach the minimal amount of confidence to select the right model. As the flexible listeners are starting with a lower probability assigned to the right model (assuming they have the right model) by virtue of having a greater choice, they will require more input in order to achieve the sufficient level of confidence for that model. Hence, there is a payoff for being a more flexible listener but also there are some benefits to forming stereotypes about the speakers one is exposed to, in order to quickly narrow down their expected variability. If the observed input does not match the models available in a narrow selection with higher probability, then listeners will make a mistake at first and then need more input to evoke a model with low probability or start creating a completely novel model.

Taking this into account, if it is assumed that the main difference between lower and higher proficiency L2 listeners is that the latter have more experience with the L2, then it would be expected that they are generally the more flexible listeners compared to lower proficiency listeners. Higher proficiency listeners by definition would have richer vocabulary and more experience with English speaking voices compared to lower proficiency listeners. As the model is underspecified regarding bilinguals, it is unclear whether bilinguals, regardless of their proficiency, would have access to L1-like phonetic variation and hence be more flexible than a Bulgarian or English monolingual. This question is discussed in Chapter 2. There is a caveat, however. Both flexible and narrow listeners may experience perceptual recalibration and selective adaptation in after exposure to biasing input. Hence, in any specific situation their ability to interpret incoming input can be modulated by what they had just heard.

As it stands, it may be predicted that lower proficiency listeners would be narrow and more specialised to their L1 phonetic variation, due to less experience with L2. Higher proficiency L2 listeners would be more flexible than the lower proficiency listeners, due to their richer experiences with the L2 and assumed maintained access to their knowledge of L1 phonetic variability. Following from that, highly proficient L2 listeners would be more flexible than monolingual speakers of their second language.

Though at first this might make the flexible (higher proficiency) listeners slower than the narrow (lower proficiency) listeners, after some exposure to the speaker both groups can experience selective adaptation to a stored Bulgarian-accented or native English speaker model. Having richer and more precise models of English speech than the lower proficiency listeners would allow them to achieve overall faster and more accurate performance after some initial adaptation.

Although, not designed to account for L2 speech processing, IAF provides a detailed and specific mechanism of how fine phonetic detail can affect understanding for listeners of different levels of experience. On the basis of this mechanism, it was hypothesised that higher L2 proficiency would have an overall faster and more accurate responses with both Bulgarian accented and

Standard British English speech, despite initial slower responses compared to lower proficiency listeners.

1.3.2. Thousand Brains Intelligence Model (TBIM)

A recent neuroscientific model of perception and intelligence by Hawkins et al. (2019) proposes a new idea that stands in opposition to the traditional hierarchical models of processing, underlying most currently existing processing theories, including some speech theories mentioned earlier, such as Cohort (Marslen-Wilson and Zwitserlood, 1989), TRACE (McClelland and Elman, 1986), and Calabrese (2012). Though the Thousand Brains Theory of Intelligence has not been applied in the context of speech perception yet, it will be discussed here as it has the potential to make a valuable contribution in the reconceptualisation of speech processing. Its main advantages are its integration of different modalities, its ability to explain the effects of different levels of attention and give precise accounts of why processing can be slowed down.

The main goal of the model is to capture the workings of the neocortex, basing its architecture on the fact that the neocortex has the same neural structure no matter at which region it is sampled (Hawkins et al., 2019). The neocortex consists of layers of cells in a hierarchical organisation. The layers at the bottom are involved in the processing of basic sensory input (e.g., the edges of an object, or potentially acoustic landmarks as described by Stevens, 2002) and as the input is passed upwards across the layers it becomes more abstract. However, unlike the predictions of some of the traditional models, TBIM assumes that all the layers are connected and work simultaneously.

Hawkins et al. (2019) propose that as sensory processing unfolds a person builds up multiple models of the perceived object, which is the highlight of their model. For instance, if a person is perceiving a specific poodle, they would have multiple complete models of the dog based on each of their visual, auditory, tactile and olfactory receptors with relevant information, as opposed to joining all the information from these receptors into a one unified model. If any of these sources of sensory information is missing, the perceiver could recreate it based on the input from the other modalities and memories within the missing modality; so if they see the curly fur they evoke multiple tactile models of what it would feel like at different receptors. In addition, the sensory information evokes multiple models of related but ultimately unselected objects, such as cat or sheep. Long-range connections between all levels of all these models help select the final winning candidate. Associating complete models of abstract categories with limited input allows the organism to combine input from different senses and make predictions about the sensory information which is missing. This way having only seen a set of objects, a person is later able to recognise them by touch only. Hence, TBIM can account for the interaction of different modalities during processing, which will be discussed in the following section.

The highly interactive structure differs from the concept representation model by Medin and Schaffer (1978) that was mentioned in Section 1.2. In TBIM each of the thousands of models of one object from different sensory inputs are integrated at all levels, not just the most abstract level. According to TBIM, during processing the neocortex creates a large number competing models of the object of perception and then a winning candidate is picked. In must be noted that although the winning candidate is picked based on multiple models, there can (but does not have to) be a superordinate declarative abstract concept to be assigned to each of the models based on partial sensations (e.g., a person does not need to know that a poodle is a separate dog breed to be able to perceive it).

In order to explain different levels of efficiency in processing, Hawkins et al. (2019) use a coffee cup as their canonical example. They propose that the multiple coffee cup models are represented in the neocortex via multiple cortical columns. Cortical columns are defined as an abstract unit of cortical region that contains cells with varying functionality. In practice, the cells from a cortical "column" must be connected but do not have to be located close to each other. The word column is used just for convenience. The activation of a specific cortical column pre-activates (predicts) other coffee-cup related columns, which means they can be accessed faster when congruous input becomes available. If the correct column was not predicted (e.g., if the perceiver was biased to predict vase- instead of cup-related sensory input), then a new column has to be activated from scratch (hence more slowly), to capture the perceived cup-related sensory input. If learning has occurred in that instance, then that new column should be predicted in future exposure to the same input (Hawkins et al., 2017).

One of the key features of the model for which there is less empirical evidence is the sparseness of representations. It is hypothesised that unique combinations of cells participate in the representation of one object (abstract or real), with very little overlap. This also allows the perceiver to quickly and efficiently reach the final abstract concept based on a minimal amount of input. If the perceiver is already strongly biased to perceive either a basketball or a football, they should be able to distinguish between the two based on only one touch of the finger. If there is little overlap between the cells representing different objects, then only a small number of neocortex cells from a specific pattern need to be activated in order to reach the correct model. It has been demonstrated that during silent reading of English and Mandarin two discrete patterns of cortical activation are observed in the processing of two languages by the same readers (Xu et al., 2017). This might be taken as weak evidence that lexical representations between English and Mandarin as well as the recognition of their respective writing systems have somewhat sparse representations.

Overall, the model predicts that a perceiver can compensate across modalities when perceiving an object, in addition to combining different senses and multiple to achieve more efficient recognition. Also, it provides a concrete mechanism (via cortical column activation) that can describe learning and perceiving expected and unexpected input.

1.3.3. Applying TBIM to speech processing

This section will combine the strengths of TBIM with aspects of other processing models mentioned earlier to account for speech processing in general, with a focus on processing an unfamiliar accent by more and less experienced listeners.

The aspects of the Thousand Brains Intelligence Model (TBIM) described so far can readily be applied to more abstract objects of representation, such as mathematics or word meanings, labelled by sequences of sounds. This model accounts for converging sensory streams and can account for the fact that both visual and auditory information can affect speech processing (Masapollo et al., 2017; McGurk and MacDonald, 1976). It might also explain the visually evoked auditory response (vEAR or 'visual ear') phenomenon in which a high proportion of participants report hearing a sound or feeling a vibration when observing silent videos of collisions or flashing lights (Fassnidge and Freeman, 2018). If the visual input triggers an already established multisensory model of a collision or flashing siren, then they might be able to vividly predict and "hear" the accompanying sound even though it is not present in the auditory input. This might also explain the illusory vowel phenomenon observed by Peperkamp and Dupoux (2003). They report that in the process of loanword adaptation listeners of Japanese report hearing a vowel in an otherwise "illegal" consonant cluster in Japanese. This "illusion" occurs even in the absence of any vowel-like input. In addition, Moos et al. (2014) report that gradient associations between vowels and colour can be observed in both synaesthetic and non-synaesthetic populations, suggesting that there is a cross-sensory aspect to the encoding speech sounds in memory. The embodied nature of speech processing has even been used to improve the comprehension of speech in noise through tactile sensory feedback (Cieśla et al., 2019).

Another sensory source of information in speech representations might come from the articulatory gestures involved in speech production, which is the main object of perception and representation according to the direct realist theory (Fowler, 1996) and the Perceptual Assimilation Model (Best, 1994; Best et al., 2001). There is evidence that articulatory movement is involved in speech processing. For instance, Bruderer et al. (2015) found that 6-month old infants are better able to discriminate non-native phonetic contrasts when they did not have teething toys in their mouths. While this study probably captured the process of forming rich multi-sensory models of the sounds (that had not been heard by the infants before), as opposed to discrimination using a complete model, it demonstrates that articulatory gestures are also an important part in the representation of speech sounds.

TBIM addresses the mechanism of speech processing when encountering an unfamiliar accent. If listeners have less experience with a particular variety, and its expected phonetic realisation, they would be faced with less familiar combinations of cues that they can use to access the right phonetic representation. At the same time, they would be building new representations of that speakers' productions. The active process of learning an idiolect might slow down speech processing initially, because unpredicted cortical columns will have to be activated, which is less efficient than reaching preactivated columns. If the neural representations of concepts are sparse (i.e., have little overlap), then even highly disrupted acoustic signal, missing the majority of distinctive features should still lead to the correct representation (Ernestus, Baayen, and Schreuder, 2002; Mitterer and McQueen, 2009).

The way that TBIM accounts for noise in the input (unpredicted variation, such as a new accent) is by positing that it activates cortical columns that were not predicted and which in turn might lead to further incorrect predictions, which are punished (Hawkins et al., 2017). According to the model, column activation should be fastest when the column is predicted, and the input matches the column at least partially. If the input fails to activate neurons in the predicted column(s), the activation of unpredicted columns is slower and could impair the correct recognition of a given object by virtue of increasing the number of possible models that need to be considered. Inefficiencies arise because TBIM assumes that when completely unexpected input is heard it activates a less used column at random. All the cells in the new column get activated, in a so-called “burst”, otherwise, if a cell is correctly predicted, only the correct cells within the column become active (Hawkins et al., 2017).

For an inexperienced native listener, foreign accented speech can present a large amount of small or large deviations from a language and a variety they would otherwise be familiar with. Hence such speech can present numerous opportunities for “bursts” of new cortical columns, which would make the processing of this speech more inefficient and slower, at least initially. However, if the foreign-accented speech presents L1-like phonetic variation and is heard by listeners who are accustomed to this variation, there is a possibility that they will process it fluently.

In addition, as suggested in the previous section, a listener who is familiar with a specific accent and a type of speaker, could understand their speech by paying less attention to it. This means that they can attend to fewer phonetic cues in order to recognise the intended meaning, thus making their listening efficient. This would be possible because their representations of a given familiar accent or speaker are so rich that even few cues would be sufficient to sparsely activate the correct representation. This is equivalent to the tactile example in the previous section when a sufficiently biased perceiver can distinguish between a basketball and a football with the touch of one finger. However, a listener who is less experienced with an accent will need to use more attentional resources and sample more acoustic input in order to correctly recognise the intended meaning. This would make their auditory processing more inefficient due to the higher attentional load required.

Overall, while there is a lot of scope for development, the TBIM can be used to make predictions about speech processing. For example, engaging more resources during auditory processing (perhaps using more focused attention) and given a previous familiarity with the input, a person should achieve faster processing than a listeners who is using less resources or a listener who cannot effectively predict the upcoming input. This concept will be explored again in the discussion of the Automatic Selective Perception model in Section 2.3.3

1.4. Summary

The theories and models of speech processing, discussed so far, represent a great variety of schools of thought. The two main extremes were represented by classical perception models and the exemplar models. The traditional models assume that listeners identify abstract pre-lexical features at an early stage of listening and use them to gradually identify more superordinate abstract elements, assuming that all speakers of a language share the same abstract inventory of features. The exemplar models focus on the listeners' experience with different types of speech and their ability to make predictions based on that. The Ideal Adapter Model (IAF) uses a Bayesian approach to evaluate the likelihood of acoustic evidence during speech processing. Although focusing more on individual sound perception than whole-word or phrase processing, IAF provides the concepts of flexible and narrow listeners and perceptual recalibration and selective adaptation. These concepts allow for predictions regarding the listening behaviours of less experienced listeners. The Thousand Brains Intelligence Model (TBIM) proposes a specific and original architecture of how multiple sensory sources contribute to the ultimate identification of a single abstract concept, which is consistent with the IAF and exemplar approaches, and is applicable for speech processing. In turn, TBIM gives neurological validity of the exemplar speech processing models.

One of the main challenges for the models listed so far is that they need to be applicable to a wide range of listening situations, including bilingual perception. This is not an easy task, because additional languages require the development of an additional lexicon and phonetic cue identification preferences, which raises the question how the multiple systems interact. As they stand, the models discussed so far are completely agnostic with respect to bilinguals. An example of a model of bilingual listening, BIMOLA, based on a more traditional structure will be discussed in Section 2.3.2. Probability-based models might predict that in any listening scenario a low probability is also assigned to words from an alternate lexicon to account for evidence of lexical competition (e.g., Lagrou et al., 2015). Future general speech processing models should take the challenge of not assuming monolingual and monolectal listeners. Some relevant models which have tried to undertake this task will be discussed in the following chapter.

Chapter 2 Bilingual listening

This chapter will briefly focus on existing models trying to explain how bilinguals perceive and map phonemic variation in their L2. As mentioned in Chapter 1 many models of speech perception assume that incoming speech is parsed into phoneme-like units which help its decoding. Therefore, a lot of research has been dedicated to finding out how second language phonology is developed, based on the existing first-language phonology, and where the challenges lie for achieving native-like perception and production of individual phonemes. That type of research has been generally guided by the Speech Learning Model (SLM) by Flege (1995) and the Perceptual Assimilation Model (PAM) by Best (1994) and Best and Tyler (2007). However, there are also accounts of bilingual listening that take a more global approach, like the Bilingual Model of Lexical Access (BIMOLA) by Léwy and Grosjean (2008), and the Automatic Selective Perception (ASP) model by Strange (2011), which are more relevant to the problem of general L2 speech processing, as opposed to segment discrimination. Before discussing these models, a definition of bilingualism is in order.

2.1. Defining bilingualism

When addressing the problem of how a second language is acquired, inevitably one is faced with the problem of defining what counts as a second language competence and what counts as a language. These two concepts feed into the understanding of what bilingualism is. This section does not aim to provide a universal solution to these issues but merely create a working definition that will be of use when reviewing the literature on the topic bilingual listening.

One of the seminal authors in the field, Grosjean (2010, p. 10), states that bilingualism is: "the regular use of two or more languages (or dialects), and bilinguals are those people who use two or more languages (or dialects) in their everyday lives." This usage-based view has been supported by one of the pioneers in bilingual study, Haugen (1956). However, many previous academic definition of bilingualism consider the high mastery of both languages as an important component of the definition (Bloomfield, 1933; Huston, 2002; Thiery, 1978). This approach assumes that the monolingual is the norm and that increased mastery should lead to a monolingual approximation.

The academic field of bilingualism has shifted to recognise the dynamic nature of bilingualism, where any form of competence in another language is recognised as a form of bilingualism (de Bot and Jaensch, 2015). Depending on the usage of and exposure to different languages, individuals' competencies are in a constant flux. This broad definition, however, questions the very nature of the concept. If any competence counts, then the majority of the population on the planet need to be considered not only bilingual but multilingual and multilingualism has to be regarded as the norm and monolingualism as a small facet of that spectrum (Walters, 2004). As this dissertation focuses on the interaction between the two dominant languages of the participants, for convenience they will be called bilinguals, even though it is expected that many of them would have considerable competencies in other languages. This approach seems to be common in the literature on bilingualism, even when the complexities of the term are recognised (Pavlenko, 2014).

The definition for a bilingual in this dissertation is a person who can communicate using at least two languages. The minimum functional proficiency required for the bilinguals to take part was to be able to read and understand the instructions and consent form in English and to have been raised as a Bulgarian speaker in their childhood.

2.2. Models of bilingual phonologies

2.2.1. Speech Learning Model

The Speech Learning Model (SLM) (Flege, 1995, 2005) is one of the most influential models of second language perception and production. Its aim is to explain the variation in the successes and failures to acquire L2 perception and production (Flege, 2005). One of the main innovations of SLM is that it predicts that the phonetic inventories of both L1 and L2 can change throughout the lifetime and can influence each other bidirectionally. Unlike its predecessors the model is much more concerned with the specific phonetic realisations of phonemes as opposed to reducing phonemes to the phonological features that characterise them.

It also bases a lot of its predictions on the level of similarity between phonemes in L1 and L2. However, the way proposed to establish similarity between phones is based on the ability of learners to discriminate between L1 and L2 phonemes, not by analysing the acoustic characteristics of the sounds. According to SLM the greater the perceived distance between phone A (in L1) and phone A' (in L2) the greater the likelihood that a separate phonetic category will be formed for L2 (i.e., a process of "dissimilation"). However, if they are too similar (not determined by objective acoustic similarity but by listener discrimination and ranking (Flege, 2005) they will "merge" into one category and will be produced the same way in both languages. According to the perceptual magnet model, mentioned in the previous chapter, a greater ability to discriminate between two sounds means that at least one of them is at the fringes of a prototypical category.

In general, however, SLM does not make predictions about the ability to auditorily discriminate between phonemes, because the discriminability between sounds is the assumption used to make predictions about L2 production. While systematic discriminability in L2 might be linked to more accurate and fast processing of native L2 speech productions, it is not a direct concern of SLM. The model is pertinent to the present discussion as far as it predicts that the listeners' ability to discriminate native L2 sounds improves over time and that listeners develop an L2 phonological system that is distinct from their L1 phonological system as they gain more experience with the language. SLM predicts that at an earlier stage of their L2 learning the listeners share phonetic categories with the L2. Hence, albeit indirectly, SLM leads to the prediction that L1-accented L2 speech might be easier to process especially for lower proficiency listeners, who still rely on merged categories but that might change as listeners increase their proficiency and exposure to L2.

2.2.2. Perceptual Assimilation Model

Another model, originally designed for naïve listeners with no experience of the language they are hearing, and later adapted for L2 learning is the Perceptual Assimilation Model (PAM) (Best, 1994; Best and Tyler, 2007). The model makes specific predictions about the ability to discriminate between pairs of phones within L2 than SLM. Unlike the older models it does not assume that the units of perception are abstract features and phonemes, and unlike SLM it does not assume that the acoustic characteristics of the signal are what is being perceived. PAM is based on the direct realist approach that posits that the units of perception are the "distal articulatory events that produced the speech signal" (Best and Tyler, 2007, p. 22). Higher order mental categories are based on articulatory invariants that are composed over time with exposure to L2 sounds. This concept is not incompatible with TBIM, as the representations for the articulatory gestures underlying sounds, can also be directly accessed as part of the complex multisensory models that TBIM assumes.

Unlike SLM, which only has the options for merging and dissimilation of phones between L1 and L2, PAM allows for three types of relationship between a phone from L1 and L2. They can be separate; underlyingly merged and realised identically; or underlyingly merged and realised with a phonetic difference. According to Best and Tyler (2007) if the speaker is able to perceive some difference between two sounds, they should be able to maintain at least some phonetic contrast between them. However, this prediction is quite vague because, as discussed in Chapter 1, the processing of all acoustic input is highly contextual. For example, Hawkins and Smith (2001) pointed out a glottal stop can occur at the start and in the middle of the phrase "hand it over" when pronounced with a London accent. Although both native and non-native listeners should be able to identify these two glottal stops as different, based on their function and perhaps even their fine phonetic detail, it is unclear if this means that listeners have separate representations for them.

PAM uses the principle of the Perceptual Magnet Effect (Kuhl, 1991; Kuhl and Iverson, 1995) according to which listeners can better discriminate sounds that are less prototypical of a category. The more prototypical two sounds are of the same category, the harder it is to distinguish between them. According to PAM, a foreign sound falls in one of three types: classified (anything between a good and a bad example of a native sound), unclassified (unlike any native sound), and non-assimilated (unlike a speech sound).

Using these three categories and the predictions of the perceptual magnet effect, PAM makes predictions about how easy or hard the discrimination between two non-native sounds would be in order of increasing disriminability: Single Category < Uncategorised and Uncategorised < Category Goodness < Uncategorised and Categorised < Non-Assimilated and Non-Assimilated < Two Category Assimilation. With the exception of two category assimilation, overall, the greater the similarity to a native language phoneme, the harder the discrimination between L2 phones. An issue with PAM is that it is underspecified with regard to the effects of increased proficiency. Unlike SLM, which predicts a dissociation between L1 and L2 sounds, no such prediction is made in Best and Tyler (2007). Hence it is hard to determine how generalisable these discriminability hierarchies are across listeners with different backgrounds.

Nevertheless, the discriminability hierarchies have generally found a lot of support (Best et al., 2001; Guion et al., 2000; Polka, 1991, 1992; Rohena-Madrazo, 2013) even for early learners of L2 (Calderón and Best, 1996; Pallier et al., 1997). Harnsberger (2001) falsified the predictions by demonstrating that Single Category phones were equally as discriminable as Category Goodness assimilation phones, while the latter did not differ significantly from Two Category assimilation phones. In addition, Aaltonen et al. (1997) demonstrate that discrimination performance is less variable than identification performance, which puts into question the methods used to test the theory, which overwhelmingly involve ABX discrimination tasks. Overall, while the perceived similarity between L1 and L2 phonemes might be a guide to how differences are discriminated, the evidence suggests that it is not the absolute predictor of how foreign sounds will be identified.

In addition, languages tend to differ in their phonological inventories. If the listener's L2 has more phonemes, particularly vowel phonemes, than their L1, as is the case with English and Bulgarian, then achieving two category assimilation for all L2 phones would be impossible. Bulgarians are taught at school and are meta-linguistically aware of six to eight vowels, each of which has a unique representation in the alphabet. As will be discussed in Chapter 4, in unstressed position vowel distinctions tend to be somewhat obscured. This is fewer than Standard British English, which is sometimes analysed as having about twenty vowels, including diphthongs (Wells, 1997, 1982). As a result, it would be difficult for L1 Bulgarians to find an exact correspondence of all possible English phonemes in their L1 without any two-to-one (e.g., English /ɪ/ and /i/ to Bulgarian /i/) or three-to-one correspondences (e.g., English /ɜ/, /ə/, and /ʌ/ to Bulgarian /ə/). As a result, Bulgarian-accented English speech might not systematically distinguish these contrasts. In this case an L1 accent might lead to difficulties with out-of-context phone discrimination, but it is unclear if it would still be true in a more disambiguating context such as a word or a phrase.

Similarly to SLM, PAM does not make explicit predictions about speed of processing. However, it is consistent with SLM in positing that for consecutive bilingual listeners there are correspondences between L1 and L2 phones, such that the discrimination and potentially identification of L2 phones depends on their level of similarity with L1. The actual predictions about L2 discrimination depend on the L2 sounds in question. Good discrimination performance might be expected both with phones that are categorised as two different L1 sounds and phones that are not categorised or a combination of the two. However, the model still gives advantage to two L2 phones that are consistently assimilated to two separate L1 sounds. Hence, based on the PAM predictions, it can be expected that L1-accented L2 should not on its own hinder L2 processing for bilingual listeners.

2.3. Bilingual lexical access

While most recent research on cross-linguistic phoneme discrimination has generally been in the context of one of the major theories presented above, this is not the case for research on cross-linguistic lexical access: the process of recognising a whole word as opposed to a single phone.

There is a large amount of research demonstrating that when bilinguals perceive a word in an L2 language, they activate its translation, as well as phonologically and orthographically related words in their other language (Dijkstra et al., 1999). However, even though linguistics is usually concerned with spoken as opposed to written language, most psycholinguistic research in cross-linguistic lexical access in perception has used written words or pictures as stimuli. Studies using auditory stimuli are considerably fewer (cf., Szakay et al., 2016). Therefore, some of the most influential lexical access theories, such as Bilingual Interactive Activation (BIA) model (Dijkstra and van Heuven, 1998), and the Semantic, Orthographic and Phonological Interactive Activation (SOPHIA) model (Dijkstra and van Heuven, 2002), are concerned with lexical access in reading as opposed to listening. These are less relevant for the aims of the present research therefore they will not be discussed in detail. More relevant models such as the Language Mode framework and BIMOLA will be discussed in the following sections.

One of the biggest debates in the bilingual lexical access literature is whether lexical access is language-specific or language non-specific. The language-specific approach assumes the lexicon is organised per language that can be pre-selected in perception and production. Therefore, as a word is being perceived, selection happens only within the relevant pre-selected language, even though words within the other language might be active as well (Costa et al., 1999; Potter et al., 1984). The main support for this paradigm comes from experiments which demonstrate that a cost occurs when languages are being switched (Dalrymple-Alford, 1985; Grainger and Beauvillain, 1987; Macnamara and Kushnir, 1971). This line of evidence is criticised in Section 2.3.1 as incomplete in view of the language mode paradigm.

Another source of support for a language-specific organisation of the lexicon comes from the fact that language dominance leads to asymmetric speed of lexical access between languages. Psycholinguistic research on bilingual lexical access has overwhelmingly demonstrated that priming from the participants' first language (L1) of targets in their second language (L2) gives stronger effects than priming in the opposite direction, both with translation equivalents and cognates (Basnight-Brown and Altarriba, 2007; Duñabeitia et al., 2014; Gollan et al., 1997; de Groot and Nas, 1991; Jiang, 1999; Jiang and Forster, 2001).

A counterargument to this position is that words of the same language can in theory be more easily accessed together than across languages, because words from the same language tend to be accessed and used together. This can result in a structure resembling a language-specific organisation. If L2 words are normally less active in the mind of the bilingual because of less frequent use or later age of acquisition, then priming with the first language can result in a significant boost in the L2 word activation. L1 words would normally already have higher activations than L2 words for the same reasons. Therefore, priming with L2 words cannot contribute to as big of an increase in activation in L1 (Duñabeitia et al., 2014). However, if the listener's respective language dominance changes, the asymmetry might be reversed or balanced out (as demonstrated by Duñabeitia et al., 2010). This way, the priming asymmetry can be explained without having to rely on language-specific selection.

Lastly, a language-specific selection process leads to the prediction that interlingual homographs will be accessed equally as fast as control words, as long as the task does not require the bilinguals to switch languages. This was demonstrated in early studies, such as Gerard and Scarborough (1989). Later studies, using more robust controls for word frequency, demonstrated that bilinguals slow down when recognising homographs, compared to control words. In addition, it has been demonstrated that cognates are always recognised faster (Dijkstra et al., 1998; Kerkhofs et al., 2006; Schwartz and Kroll, 2006; Titone et al., 2011). Together these studies suggest that the lexical items in the language not used for processing affect the lexical access within the target language, thus providing strong evidence for language-non-specific lexical access.

The language non-specific approach assumes that as a word is perceived, all phonologically, orthographically and semantically relevant words in both languages are activated and competing for selection (as supported by the cognate facilitation and homograph interference effect described above). This approach is adopted and supported in the Bilingual Interactive Activation BIA (Dijkstra and van Heuven, 1998), BIA+ (Dijkstra and van Heuven, 2002) and the Inhibitory control model (Green, 1986, 1998). They are complementing models, which focus on visual word recognition or speech production, which is the reason they are not discussed in more detail here.

The models mentioned so far, mainly use visual stimuli to investigate the effects of language dominance on lexical access. However, when the modality is switched to auditory processing, many new factors need to be considered. In the context of auditory processing, if it is assumed that listeners tend to associate words with specific voices and accents (Kim, 2016; Pierrehumbert, 2002), then it is possible that lexical access is also affected by factors relating to the phonetics of the word, the environment in which it is heard (Hay et al., 2017). This section will focus on the effects of task and environment of presentation as well as the socio-phonetic characteristics of the acoustic signal. An attempt will be made to combine these various sources of influence in a comprehensive model of bilingual lexical access (BIMOLA) and suggest how the model can be improved.

2.3.1. Language mode

A crucial factor for lexical access is the language mode of the bilingual (Grosjean, 1982, 2001). Grosjean proposes that the languages of a bilingual are always active to a certain extent and that the activation varies on a continuum. When a bilingual is in a monolingual mode, one of their languages is active (e.g., Language A), while the other has very low levels of activation (e.g., Language B). That would occur for example when the environment is monolingual in language A and the person is communicating with a person who only speaks language A. In a bilingual mode, both languages are highly activated. This situation would occur during code-switching when the interlocutor is also fluent in both language A and B. As proposed by Grosjean (1989), there are intermediate states between these two extremes and the mode of a bilingual can vary dynamically. This proposal has been supported by research on both lexical access and phonetic production and has important implications for experimental design (Dunn and Fox Tree, 2012; Simonet, 2014; Soares and Grosjean, 1984).

Grosjean (2001) points out that some studies fail to take language mode into account and that can skew the interpretation of the data. He uses Caramazza et al. (1973) as an example. They tested the identification of CV syllables with varying VOT by French-English bilinguals when changing the language setting. The French or English language setting (the environment) was induced by changing the language of the instructions and the initial task, and by having either a French or an English speaker present the stimuli and by inviting the participants in a French language high-school or an English language university in Canada. However, the identification patterns of the participants did not become more French- or English-like under this manipulation. Grosjean (2001) points out that as the stimuli were language-neutral, changing the language setting was not sufficient to induce a monolingual mode and the participants simply remained in bilingual mode because of the changed language setting. This was changed in Elman, Diehl and Buchwald (1977) who used real words for stimuli and observed language-specific identification patterns but only for the bilinguals who were highly fluent in both languages.

Language mode is also linked to lexical access. Cheng and Howard (2008) use the speed of lexical access in a lexical decision task to measure the participants' relative language activation in single and multi-language tasks. If participants have slower reactions in a mixed language task than a monolingual task, that is explained as the cost of activating a previously inhibited language. However, if there is no slow-down in reaction times for a mixed language task, this signals that both languages are equally activated. Mixed-language task costs have been observed by Grainger and Beauvillain (1987). They asked their French-English bilingual participants to do a lexical decision task on two types of lists. The first type contained sets of words in only one language, while the second contained words in both languages. Reaction times were significantly faster in the monolingual blocks than in the bilingual blocks. Similar results were observed by Dalrymple-Alford (1985) and Macnamara and Kushnir (1971) in a silent reading task. While these results might suggest that lexical access is language-specific, the cost of switching is not universally observed.

Studies have shown that the observed language switching costs are due to the nature of the processing task. Language non-specific orthography might be one of the factors that leads to processing costs. An additional experiment by Grainger and Beauvillain (1987) showed that the mixed-language processing cost was observed only on words that do not exhibit language-specific orthography and only when they were preceded by a word in the other language.

Breaking language expectation might be another situation that results in a processing cost. Cheng and Howard (2008) demonstrated that Mandarin Chinese - Taiwanese bilinguals showed switching costs in a mixed language task only if they expected a monolingual task. When enough information was provided about the bilingual nature of the language task, they showed no costs when switching languages. The same results are observed in a picture naming task by Mosca and de Bot (2017). Their bilingual participants had to name pictures in a language indicated by a colour. When the colour appeared together with the image to be named, reaction times were slower than when the colour preceded the image by several hundred milliseconds. In addition, Wagner (2016) reports that when inducing the "listening to the wrong language" effect, reaction times of bilingual listeners are affected only at the first unexpected word they encounter. Subsequent items from the unexpected language are processed at a normal speed.

All of these studies conclude that bilinguals can quickly adapt to the nature of the task, rather than necessarily experience automatic costs in any mixing situation, thus supporting Grosjean's situational model of dynamic language mode adaptation and showing evidence against the universal language-specific lexical access models discussed in the previous section.

2.3.2. BIMOLA: Complete model of bilingual lexical access

Overall, most of the models discussed so far focus on a narrow aspect of word processing – usually they take into account a small number of factors (e.g., language dominance and semantic relationship between words) and relate it to the activation of whole words within one language or the other. One of the very few models that takes into account bilingual word-recognition from the acoustic input all the way to the word level is BIMOLA (Léwy and Grosjean, 2008). This perspective is valuable for investigating the question of processing whole words produced with different accents.

As seen in the schematic representation of BIMOLA in Figure 1, the model supports both bottom-up and top-down paths of activation of linguistic information. On the one hand, the top-down activation is controlled by the language mode of the bilingual: the relative activation of each language boosts the activation of words (and phonemes) from the respective language. On the other hand, in the bottom-up path the important information is detected in the acoustic signal and transmitted to higher levels of processing. Distinctive features which minimally specify certain phonemes are detected in the signal, and then they activate their respective phonemes. The level of features is shared between the two languages, while phonemes and words are organised in language-specific subsets, which are grouped together in a larger set. Once phonemes are activated by the minimal distinctive features, they also activate close phonemes within the same subset. Groups of phonemes then activate word candidates. The selection of a candidate is affected also by the semantic context, word-frequency effects and language activation, which affect the top-down flow of information.

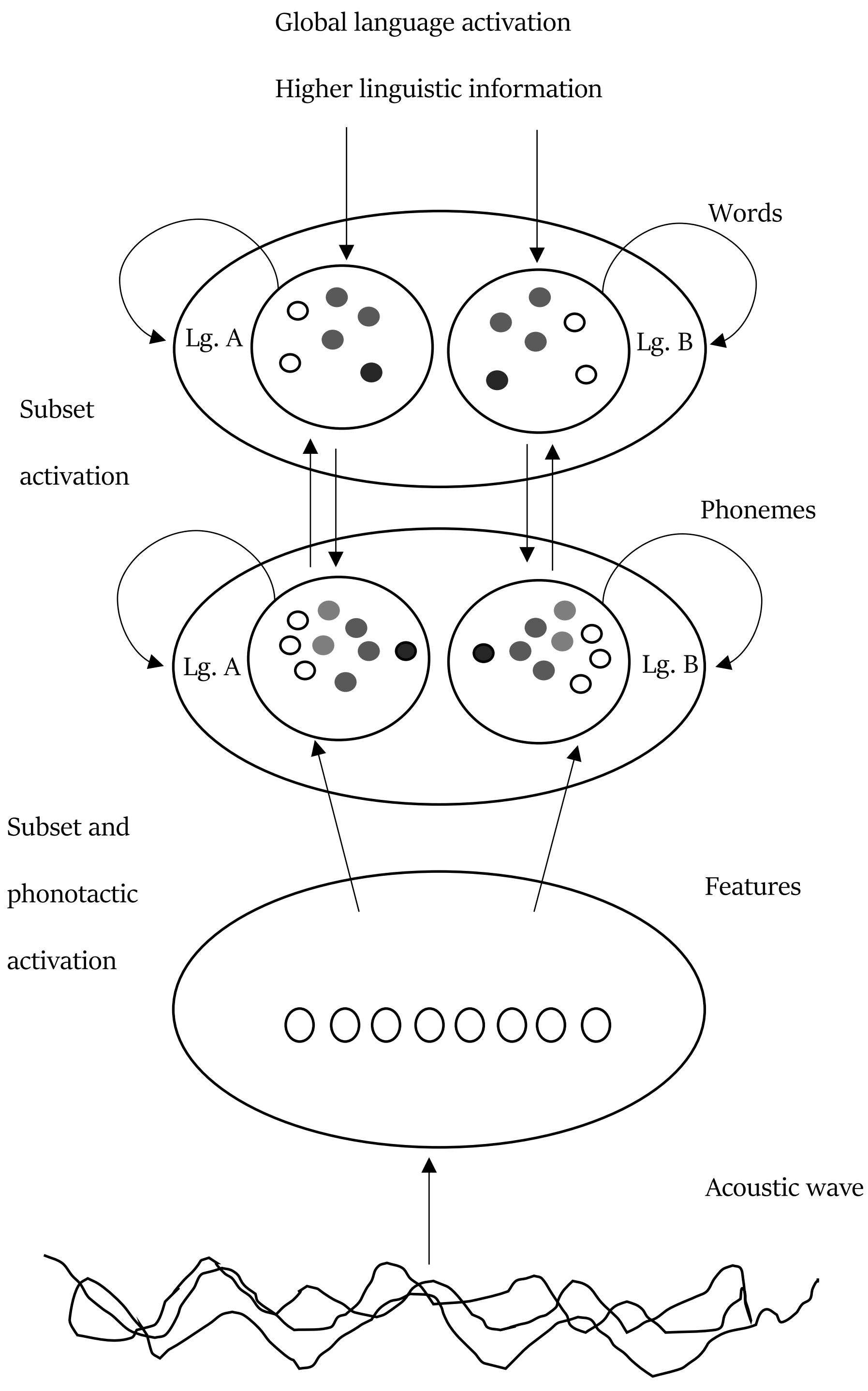


Figure 1. A visual representation of BIMOLA (reproduced from Léwy and Grosjean, 2008)

Although the model keeps the phonemes of the two languages separate schematically, in practice it recognises that some pairs of phones are more similar to each other than others based on the number of features they share (see Figure 2). This is captured through the unit similarity effect mentioned by Léwy and Grosjean (2008, p. 202): “a unit in one language, e.g. a phoneme, which shares properties with a unit in the other language, will be activated when that unit is presented.” This means that close phonemes across languages will be activated because they already share a lot of features. The features themselves are activated from the acoustic input. The bottom-up activation of phonemes in the competing language in turn activates words from that language, which stimulate other words and phonemes from that competing language in a top-down direction.

This suggests that exposure to English containing Bulgarian-like phonetic features (i.e., Bulgarian accent) should increase the activation of Bulgarian-specific phonemes and words in the Bulgarian-English bilingual listeners. If the listeners are hearing native English speech, devoid of Bulgarian phonetic characteristics, then competition from Bulgarian should be lower and lexical access in English would be faster (but see Marian et al., 2008). Hence, listening to Bulgarian-accented English in a task only requiring lexical access in English might lead to slower reaction times, due to the increased competition of the L1 Bulgarian. However, in a cross-linguistic task a Bulgarian accent might improve the recognition of Bulgarian words, if they are primed by Bulgarian-accented English instead of native English. If the listeners expect a language switching task (cf., Cheng and Howard, 2008; Mosca and de Bot, 2017) then the Bulgarian accent can only be beneficial by additionally increasing the overall activation of Bulgarian words. Similar findings are observed in Szakay et al. (2016), and will be discussed in Section 3.3.2.

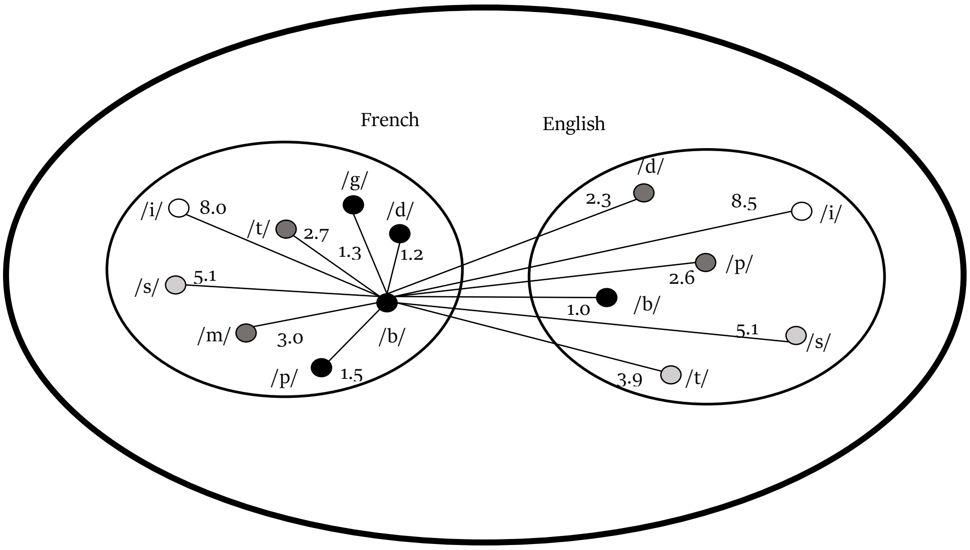


Figure 2. Visual representation of distances between phonemes in BIMOLA (reproduced from Léwy and Grosjean, 2008)

One issue with this hypothesis is that it does not specify whether L1 competition is modulated by the bilinguals' proficiency in their languages. As it stands, the model appears to predict the same reactions from all bilinguals, only accounting for different levels of bilingual mode activation, depending on the conversational situation. Drawing from the SLM, perhaps lower proficiency listeners have generally "shorter distances" between the phones of their two languages. Thus the bilingual's expected categories from L2 would actually resemble their L1 equivalents, if such are available, and processing words containing such equivalents would be faster, compared to L2-specific sounds which are not firmly established (as seen in Marian et al., 2008).

Overall, BIMOLA represents a traditional approach to bilingual speech processing, and the issues with such an approach were discussed in Section 1.1. It aims to account for the different levels of language activation in bilingual and monolingual mode of processing. This leads to the prediction that L1 would not be facilitatory in an L2 monolingual task because of increased competition of L1 words. However, an L1 accent may be helpful in a task that involves language switching because it pre-activates the other language. A drawback of BIMOLA is that it does not take into account that the listeners' proficiency may vary and change the expected inhibitory effect of an L1 accent.

2.3.3. Automatic Selective Perception (ASP)

Strange (2011) introduces a new model of both bilingual and monolingual speech perception, the Automatic Selective Perception model (ASP) that draws from the models summarised thus far. This model is particularly suitable for the purpose of the present dissertation, as it aims to account for the perception of two populations, namely functional monolinguals who have little or no experience with other languages and learners of a second language who have completed most of their formal education in L1 and have emigrated into an L2-dominant country. The latter population, Bulgarian residents in the UK, is the primary interest of this dissertation.

ASP focuses on the bottom-up auditory recognition of individual word forms stripped from any top-down effects of syntax and morphology. For that purpose, many of the studies reported by Strange (2011) test phonotactically legal nonsense words either individually or embedded in a neutral sentence. The model's focus lies in characterising the acts of perception involved to achieve different listening goals. In order to study only the acts of listening in relation to the listening goals, the model uses specific architectural components that will be briefly summarised in this section: attention (related to automatic vs. attentional processing), contrast salience, units of perception and analysis, perceptual modes, and perceptual routines.

A crucial component of the ASP model is attention. Automatic processing is contrasted with controlled processing where the former is involuntary and can occur while the listener is focused on a different cognitive task. In practice this can result in a distinct MMN brain response to stimuli that the listener is not currently focused on. For example, Japanese listeners who were engaged in a visual task and told to ignore the auditory stimuli had an MMN response when the auditory stimuli changed from [tado] to [taado] that they did not have when they were hearing

[tado] repeatedly. This was expected because vowel length is phonemically contrastive in Japanese and is therefore automatically salient to them. By comparison naïve American listeners did not react to this difference during the visual task (Hisagi et al., 2010).

Controlled processing involves controlled allocation of attention on specific parts of the input. Selective perception involves weighting phonetic cues from the input in accordance with the listeners' linguistic experience and habits. The model emphasises the importance of the selective perception routines (SPR) which a listener develops over time to increase the efficiency of their listening. SPRs involve detection and weighting of acoustic cues, which are associated with controlled processing. Nevertheless, over time they become highly automatised and efficient (Strange, 2011). For example, the American listeners in Hisagi et al. (2010) only had an increased MMN amplitude when they were instructed to pay attention to the auditory stimuli. They were capable of noticing the contrast but the lack of L1 experience means the process was not as automatic as for the Japanese listeners.

Regarding memory, the ASP model focuses on what is contained in long-term representations and the procedural knowledge employed by listeners as a response to different stimuli and tasks. The units of analysis and perception are assumed to be syllables without making claims whether listeners can or cannot delimit smaller phoneme-sized units. Unlike PAM, the basic units of perception are not composed of a direct realist perception of underlying articulatory gestures. Instead, Strange (2011, p. 460) describes them as: "structured sets of phonetically relevant acoustic parameters that specify phonological sequences." In his description, Strange (2011) posits that the representations are language-specific, stored in a syllabic frame and contain articulatory information. The model uses components of Articulatory Phonology (Browman and Goldstein, 1992) and assumes that the "articulatory synergies" are processed in a slightly more superordinate context, syllables, compared to the typical phoneme-by-phoneme processing that was reported for BIMOLA, for example.

Regarding the acts of speech processing, the model predicts that there are two general modes: phonetic (requiring controlled processing) and phonological (using automatic processing). These two types of listening, originally discussed by Werker and Logan (1985), are also used by Calabrese (2012) (cf., Section 1.1). The phonetic mode is more cognitively demanding and is likely to occur in the very early stages of L2 learning when phonological categories are not yet firmly established. In SLM that would correspond to a stage when most categories between L1 and L2 are merged. It can also be used when discriminating between individual sounds or providing judgements on the accentedness of somebody's speech. Hence it does not need to be associated only with one type of speaker, but it can also be selected, based on the task.

Phonological processing is more meaning-oriented and more likely to be employed in continuous speech and be more automatic than phonetic processing, which requires higher attentional cost. Phonological processing focuses on extracting the necessary and sufficient information from the signal to complete the listening goal. For example, Cai et al. (2017) demonstrated that the UK listeners' disambiguation of a homophone with dialect-specific meanings ("bonnet") is affected by the overall accent of the utterance, not of the individual word. However, in a different experiment Adank et al. (2010) demonstrate that listeners who tried to imitate an unfamiliar accent, thus engaging in some level of phonetic listening, improved their comprehension of that accent. Improvement was not observed when the listeners just listened attentively, repeated or just transcribed the speech. This suggests that phonetic listening affects comprehension outcomes with unfamiliar speech but also that it may vary continuously between phonetic and phonological mode. This issue is not addressed in ASP.

As suggested earlier in the discussion of TBIM and IAF, it might be useful to associate different attentional demands with different stages of the same listening task. This is where Strange's (2011) inclusion of a phonetic mode of processing might be particularly valuable. For the purposes of this dissertation, it is hereby predicted that when faced with an unfamiliar voice or an accent, or speech in noise, listeners will increase their attention, in order to maximise the use of the available phonetic information from the speech signal. This increased attention to phonetic detail will initially slow down their processing. Given continuous input from the same type of speech, this learning phase will continue until they have established short-term selective perception routines that allow them to efficiently process the input with their habitual amount of effort. If the input already matches their expectations (representations) at least partially, for instance when hearing a new speaker from a familiar accent, the listeners will adapt to the speech faster than if they have to rebuild their SPRs for a new speaker and a new accent (e.g., Kriengwatana et al., 2016). In the context of the present research interest, it is predicted that high English proficiency listeners will adapt to native English speech faster than lower proficiency listeners. However, even if high proficiency listeners adapt to native English faster than Bulgarian-accented English, because of SRPs they have developed, they should still adapt to Bulgarian-accented English at least as fast as lower proficiency listeners.

Of the models reviewed so far, ASP is the only one that provides a clear mechanism that can account for how listeners become "specialised" to a specific accent over long-term exposure to it. It is also one of the few models that make predictions about the listeners' speed of processing as opposed to their ability to achieve the end result of sound/word recognition. It is proposed here that the same concepts can be reused and tested to test short-term adaptation to unfamiliar speech.

2.4. Proficiency

One of the dissertation's aims is to investigate the effect of proficiency on the phonetic representations of Bulgarian-accented English for Bulgarian-English bilinguals. The research discussed so far shows that the listeners' relative proficiency has a direct impact on their speech processing. Therefore, this section will focus on the debate on what counts as proficiency, and how it can be measured. Lastly, it will discuss the available findings, regarding proficiency's effect on bilingual listening. In order to stay relevant to the topic of the thesis this section will only focus on aspects of proficiency that concern auditory speech processing.

2.4.1. What is proficiency?

The concept of language proficiency is generally regarded as a groupings of abilities related to language comprehension and production (Bachman and Palmer, 1996; Canale and Swain, 1980; Thurstone, 1938). A considerable number of models has been developed trying to define and create measures for the individual skills involved in language ability. One of the seminal models of language proficiency was proposed by Carroll (1961) (as cited in Verhoeven and de Jong, 1992) narrows down the areas of competence to ten items, divided in three categories: linguistic knowledge, channel control and integrated competencies.

Knowledge of structure and lexicon fall within the language knowledge category. The channel control competence consists of discrimination and production of speech sounds, reading (symbols to sounds) and writing (sounds to symbols). The integration between individual competencies of these two types leads to rate and accuracy of listening comprehension, speech and written composition. According to this conceptualisation, the aspects of proficiency that affect the ability of bilinguals to quickly and accurately process single-word utterances are vocabulary knowledge and discrimination of speech sounds (hence the importance of models for sound discrimination, like SLM (Flege, 1995) and PAM (Best and Tyler, 2007). This conceptualisation is in line with the Vocabulary model independently proposed by Bundgaard-Nielsen et al. (2011) and discussed below.

The models proposed by Carroll (1961) and others in the field (Bachman and Palmer, 1996; Canale and Swain, 1980; Lado, 1961) assume that the components of language proficiency can develop independently to a certain extent as a result of study and experience with a specific environment. However, if speech processing is a composite skill, as proposed by Carroll (1961) then it is directly dependent on both vocabulary knowledge and phoneme discrimination. Even if the two components are independent of each other and develop at different rates, speech processing problems are expected to occur when one of the two aspects is underdeveloped. Section 2.4.2 below will provide support of the necessity of rich vocabulary for speech processing. Lastly, while the field investigating general language proficiency has changed over the years, the idea that vocabulary knowledge and speed of information are core components of proficiency has been recently argued by Hulstijn (2011). The next section will investigate this relationship in more detail.

2.4.2. Vocabulary knowledge and speech processing

A number of studies demonstrate that there is an overall relationship between the listener's proficiency in a language and how they process auditory stimuli (be it connected speech or individual words). This section will present some theory-independent evidence supporting the importance of proficiency on language processing and will then try to account for the results through the use of the Vocabulary model (Bundgaard-Nielsen et al., 2011) and an alternative episodic account.

Abutalebi (2008) provides a review on the neural representation and language control of second language processing. One of the main outcomes of his review is that both representation and language control of L2 processing are affected by the person's proficiency in L2. Anatomical differences have been observed between low and high proficiency speakers in lexical decision tasks (Illes et al., 1999; Pillai et al., 2003), semantic judgement tasks (Rüschemeyer et al., 2005, 2006; Wartenburger et al., 2003); story comprehension (Perani et al., 1998). In addition, a PET study of auditory comprehension shows that bilinguals with low proficiency in L2 have different brain activation patterns for L1 and L2 listening, and no difference between L2 and unknown language listening (Perani et al., 1996). A follow-up demonstrates that high proficiency learners of L2 have the same activation patterns when listening to L2 as they do when listening to L1 (Perani et al., 1998).

These studies suggest that proficiency in a language can affect the extent to which the words in both languages of the bilingual are activated. This is also demonstrated in behavioural tasks. Blumenfield and Marian (2007) controlled for proficiency by recruiting German-native and English-native bilinguals of German and English. They presented their participants with auditory English stimuli and the participants had to select the semantically related picture from a display that also included pictures of similar sounding German distractors. Eye-movements towards the German distractors were interpreted as activation of German. Some of the target words were cognates with German words and others were English-specific. Blumenfield and Marian (2007) report that cognates activated German competitors for both groups of bilinguals. However, English-specific words activated the German translation only for the German-dominant bilinguals. Haigh and Jared (2007) report a similar cognate effect in a visual lexical decision task but only for the bilinguals who were performing it in their less dominant language. Overall, it appears that the 'other' language of a bilingual is consistently activated in a monolingual task, only if the bilingual is sufficiently proficient in it.

Vocabulary knowledge, as one aspect of language proficiency has been demonstrated to affect listening abilities. Two independent studies using the listening comprehension exam of the Cambridge certificate in proficiency in English demonstrate that vocabulary knowledge is significantly correlated with performance at the listening exam (Atas, 2018; Stæhr, 2009). Specifically, Stæhr (2009) demonstrated the importance of vocabulary breadth (i.e., vocabulary size), revealing that 98% vocabulary coverage is required for about 70% comprehension of spoken texts. Previously Bonk (2000) showed that lexical coverage of 80% is insufficient for successful listening comprehension for most participants. Teng (2016) also supported the relationship between vocabulary breadth and listening comprehension using listening tasks from an academic version of the International English Language Testing System (IELTS). However, he demonstrated that vocabulary depth (how well a specific word is integrated in the listener's mental lexicon) was a better predictor for listening performance than vocabulary breadth.

In addition to these studies Strange (2011) reports that vowel discrimination abilities in L2 are correlated neither with the length of residence in an environment where the language is spoken nor with self-reports of language usage but instead they are correlated with language proficiency. While vowel discrimination and listening to a text of connected speech might involve different listening methods (see Coleman, 2002; Strange, 2011) it is conceivable that processing single words falls between the two. Hence, it is important to consider the fact that vocabulary knowledge and a high general proficiency are important factors for both types of listening tasks.

By drawing parallels to child language acquisition, Bundgaard-Nielsen, Best and Tyler (2011) argue for the importance of vocabulary learning for phonological discrimination in L2 acquisition in adulthood specifically. Studies that investigate infant phonetic discrimination and vocabulary size report that young infants with small vocabularies learn words which phonetically match the dialect and speaker characteristics they are exposed to and are only able to develop phonological constancy after experiencing a vocabulary spurt around 19 months (Best et al., 2009; Swingley, 2003). Metsala (1999) also demonstrates that children with larger vocabulary size have better phonological awareness than children with smaller vocabularies.

On the basis of this research Bundgaard-Nielsen, Best and Tyler (2011) propose the Vocabulary model. According to it, vocabulary has a direct impact on the development of the L2 phonology. With the expansion of (presumably aurally acquired) vocabulary in the beginning stages of learning a second language, the L2 phonology is revised and attuned, particularly when the learning takes place in an environment among native speakers of that language. They predict that when the L2 listener conflates several L2 phones into a single L1 category, native L2 speech might activate competing lexical items and slow down processing (Cutler et al., 2006). With the increase of vocabulary size, it is argued that the listeners will develop consistent mappings between L1 and L2 sounds that fully exploit the L1 phonemic system and improve the listeners' chances for correct discrimination. As discussed in Section 2.2.2, achieving consistent one-to-one mappings between all L1 and L2 phones might be impossible in cases like Bulgarian as L1 and English as L2 due to the sizes of the phonemic inventories. Instead, it is more likely that with the increase of proficiency listeners develop separate language systems for language processing and production targets (cf., Flege, 1995; Léwy and Grosjean, 2008; Strange, 2011).

The Vocabulary model assumes that early stage L2 listeners use their native phonemes as models in L2 discrimination. Hence, it is consistent with the prediction that L2 listeners would benefit from processing their own native accent in L2 speech compared to a standard native accent, although Bundgaard-Nielsen et al. (2011) do not pursue the question. However, Ludwig and Mora (2017) who measured proficiency through vocabulary size support this claim. The less proficient Catalan – English and German – English bilinguals processed English words with Catalan and German accent respectively faster than native English speech. This effect was not present for the higher proficiency bilinguals.

One drawback of the Vocabulary model is that it is not clear how long these early stages of SLA last. It implicitly predicts that at some point of vocabulary acquisition the learners will have formed a fully functional phonological system in L2, which would not be affected by the acquisition of more vocabulary. Another drawback is that it does not describe the mechanism, according to which the expansion of vocabulary directly affects individual phones.

The proposal raised here, congruent with an episodic representation of language, is that by hearing a greater variety of words produced by native speakers the non-native listener can observe phones, clusters and syllables in different phonetic contexts and thus improve their phonetic models of these speech segments. It is assumed that in most cases a larger vocabulary goes hand in hand with increased access to the native speech of that language. As long as there is input by new speakers from a variety of backgrounds (that the listener marks as prestigious or sociolinguistically important), it is predicted that the perceptual phonological boundaries of L2 listeners will be continuously updated (e.g., Tamminen et al., 2015).

Bundgaard-Nielsen et al. (2011) support the vocabulary model by demonstrating that Japanese L1 - English L2 listeners with richer vocabularies in English had better discrimination of vowels than listeners with smaller vocabularies at fewer than 3 months after arrival in Australia. However, spending an additional 3 months in Australia did not affect their discrimination abilities and they were performing similarly to native Australian English listeners from a previous study (Bundgaard-Nielsen et al., 2008). The authors hypothesise that any gains in discrimination might have occurred in the first weeks after arrival before the first test, but as this was not directly recorded, the hypothesis still needs to be verified.

Overall, this section attempted to define language proficiency and highlighted the importance of vocabulary knowledge as part of that definition. After that it was argued that vocabulary knowledge is directly related to improved speed and accuracy of speech processing. The importance of a rich vocabulary for auditory comprehension is consistent both with the more traditional, direct realist approach of Bundgaard-Nielsen et al. (2011) and also with an episodic account of speech processing.

2.5. Summary of the predictions so far

The models discussed in Chapter 2 have focused on the experiences of bilinguals processing their second language. In the discussion of these models, specific predictions were made regarding the ability of bilinguals to process L2 speech with different accents. It was discussed that SLM lends itself to the hypothesis that L1-accented speech will be easier to process for L2 listeners, especially at a lower L2 proficiency, while PAM is more inconclusive, particularly for a language pair like Bulgarian as L1 and English as L2. According to BIMOLA, listening to L1-accented L2 speech might generally slow speech processing as the L1-like phonetic characteristics present within L2 speech might increase the activation of competing L1 categories. Consistent with the evidence that listeners create a model of the speaker they are hearing, an L1 accent might indicate a fellow bilingual and activate the listeners' bilingual mode and increase the number of hypothetical words they have to exclude prior to lexical decision.

ASP introduced concepts such as phonetic and phonological modes of listening, tendency for automaticity in processing and selective perception routines. ASP explicitly predicts that at the very early stages of familiarisation with an L2, listeners will rely more on phonetic listening, although the same mode could be employed by all listeners depending on their listening task. It was elaborated that on encountering an unfamiliar accent, listeners might increase their attention to detail (i.e., the phonetic mode of listening) which might slow their processing. As familiarisation to the accent happens the listeners will rely on selective perception routines (SPRs) more and pay less attention to phonetic detail. In the long run this adaptation and development of specific SPRs might lead the listeners to adapt to their second language but it is an open question if they can achieve the same efficiency and automaticity as in their L1. It is thus predicted that higher proficiency listeners residing in the UK will have a higher specialisation to prestigious native speech English typical of the UK than lower proficiency listeners. It is unclear whether that will happen at the cost of their specialisation to Bulgarian-accented speech or if both specialisations can develop in parallel over time.

The following chapter will further explore the topic of proficiency in the context of L2 speech production and speech perception as an L2 listener, from the perspective of the specific challenges and affordances presented by L2 speech.

Chapter 3 Listening to a foreign accent

3.1. Processing of foreign accents by non-native listeners

Studies of non-native listeners or non-native speech have demonstrated that the accuracy and speed at which different types of speech are processed do not necessarily correlate (Jensen and Thøgersen, 2017; Ludwig and Mora, 2017; Munro and Derwing, 1995). Sometimes speech can require longer time for processing but lead to accurate results, while others even when effort is applied the intended message is not reached. The following review will focus on speech processing by non-native listeners, and in particular when listening to L1-accented L2 speech. There are very few studies which have included this combination of listeners and stimuli and of them only one had it as a central question for the study (Ludwig and Mora, 2017).

It is important to clarify that although from a sociolinguistic perspective all speech inherently has some accent, the term "accented speech" here is used with a different connotation. For the purposes of this dissertation the phrase "accented speech" refers to speech that has an accent different from that of the implied listener or refers to non-native speech (even if the implied listener has a similar non-native accent), depending on the context.

As noted earlier, a foreign-accent disadvantage for native listeners has been of interest for phoneticians for at least half a century. The research overwhelmingly demonstrates that native listeners find foreign-accented speech less intelligible (i.e., less accurately interpreted). Lane (1963) reports that native English listeners were more inaccurate at identifying Serbian-, Japanese-, and Punjabi-accented English than native speech in a large variety of signal-to-noise ratio and filtering conditions. More recently, van Wijngaarden (2002) estimated that the strength of foreign accent (as determined by self-ratings and the ratings of native listeners) is a good predictor of the level of accuracy for native listeners. However, not all listeners struggle to process foreign-accented speech in the same way.

The term Interspeech Intelligibility Benefit (ISIB) was coined by Bent and Bradlow (2003) to capture the case when non-native speakers of a language identify non-native speech in a language as accurately as or more accurately than native speakers of the language (ISIB for listeners). Since then, the term has evolved. Hayes-Harb et al. (2008) distinguish between ISIB for listeners and ISIB for talkers. ISIB for talkers reflects that non-native speakers of a language might be equally or better able to process non-native speech in that language than native speech (Bent and Bradlow, 2003; Hayes-Harb et al., 2008; Stibbard and Lee, 2006; van Wijngaarden et al., 2002). ISIB for talkers (or ISIB-T) is the phenomenon that is directly related to the present investigation.

According to Bent and Bradlow (2003), if ISIB occurs when the non-native listeners and speakers share a native language, it is called matched ISIB and when they come from different language backgrounds it is a case of mismatched ISIB. Mismatched ISIB is even harder to observe. For example, Stibbard and Lee (2006) and Harding (2012) found no evidence for mismatched ISIB. No ISIB-T was found in Hayes-Harb et al. (2008) and Munro et al. (2006). Some mixed evidence is reported in Major et al. (2002). Chinese listeners were at a disadvantage when listening to Chinese-accented English speech compared to native English speech, but native Spanish speakers performed better when listening to Spanish-accented English. This dissertation focuses on matched ISIB by studying the processing of Bulgarian-accented English by Bulgarian-English bilinguals.

Increased experience with the L2 (or proficiency) may reduce the amount of ISIB-T benefit. Pinet et al. (2011) compared the accuracy of monolingual Standard Southern British English (SSBE) listeners, balanced French-English bilinguals, experienced French (L1) – English (L2) bilinguals and inexperienced French (L1) – English (L2) bilinguals. The listeners were exposed to English speech embedded in speech-like noise. The speech had SSBE, French, Korean or Northern Irish accent. The results show that the amount of listener experience with British English affects which of the speech variants is understood more accurately. The speakers with least experience with English benefitted the most from the French-accented stimuli, while those with the most experience in English, the balanced bilinguals, found the SSBE stimuli the most intelligible, similarly to the native English speakers.

Overall, the evidence suggests that listeners with lower proficiency in their L2 and a stronger connection to the L1 will have L1-accent intelligibility benefits. One drawback of this summary is that all studies used different criteria to establish the listeners' proficiency. Hence the present dissertation will focus on providing a validated English proficiency measure based on vocabulary size as in Ludwig and Mora (2017). In addition, the concept "benefit" has been defined in different ways. In some studies, it is taken to mean that L1-accented L2 processing is at least as efficient as native-accented L2 processing (e.g., Bent and Bradlow, 2003). In this dissertation "benefit" in speech processing is defined as systematically better outcomes compared to the outcomes from listening to a prestige native variety.

While the evidence of a matched ISIB-T is not very solid, even less attention has been dedicated to the processing of L1-accented L2 by non-native listeners. The next section will summarise the few relevant papers that are available and will then link them to existing theoretical frameworks developed for monolingual listeners. One of the main questions that needs to be addressed when discussing this topic is whether hearing L1-accented L2 by non-native listeners would improve or inhibit their comprehension. The next section summarises some of the research that directly addresses this question. Section 3.2 discusses more general evidence for one or the other position. Lastly, Section 3.3 brings up the importance of sociolinguistic bias on speech processing, as processing speech is inherently linked with sociolinguistic judgement.

3.1.1. L1-matched foreign accent is facilitatory

While some research exists on matched foreign accent intelligibility (under the ISIB Hypothesis), there are less than half a dozen studies that have investigated the processing of non-native listeners and L1-matched foreign accent. Of them, only Ludwig and Mora (2017) made it the central focus of their research and compared it to L1-mismatched and native accent processing. This subsection will focus on the studies that report some benefits from L1-accent, or L1 similarity in L2, in speech processing.

One of the most comprehensive studies on the topic compared the effects of proficiency and L1-match in foreign accented L2 processing (Ludwig and Mora, 2017). The results of this study show that both factors play a role in speech processing, while interacting in a multitude of complex ways. Five groups of age-matched listeners were recruited: low and high proficiency Catalan learners of English, low and high proficiency German learners of English and a group of native speakers of English who are unfamiliar with Catalan- and German-accented English. They all listened to Catalan- and German- English and native English productions. The stimuli were presented in isolation (the participants had to make animacy decisions), as well as in sentences (the participants had to assess whether they are true or false). In addition, the listeners had to rate both the words and the sentences according to how easy they are to comprehend.

Only low-proficiency non-native listeners had faster RTs for non-native English than for native English productions. However, this only happened when the non-native stimuli were spoken by L1-matched non-native speakers. This finding suggests that lower proficiency listeners have not developed clear representations of native speech (be it of individual phonemes or whole-word exemplars) and that with increased proficiency that obstacle to efficient processing is overcome. In addition, lower-proficiency listeners did not process faster or more accurately the non-native speech produced by non-native speakers with a different L1 to theirs compared to their own L1 accent or the native accent. Even though high-proficiency non-native listeners did not process L1-matched foreign accented speech faster than native English speech, they still outperformed native English listeners who overall struggled with foreign-accented speech. This evidence is consistent with an exemplar approach to speech processing (cf., Section 1.2). The high-proficiency listeners increase their competencies in both native English speech and L1-accented speech and thus inherently differ from monolingual listeners of their L2.

Across the board, non-matched non-native accents were processed the slowest. This is an important finding because the research in accuracy of speech processing has occasionally suggested that non-native productions mismatched for L1 are also easier for non-native listeners compared to native speech (Bent and Bradlow, 2003; Major et al., 2002). It appears that matched L1 fine-grained phonetic detail can be useful for the lexical processing of non-native listeners compared to non-matched L1 variation (contrary to the findings of Weber et al., 2014). One factor that could have affected the results of Ludwig and Mora (2016) is that the non-native participants were all tested in their native countries where they were learning

English in a formal setting taught by non-native English teachers. It is not known what differences there might be between proficiency matched non-native English listeners who live in their home country, and those who live in an Anglophone country.

One of the key studies exploring the interaction between L1 and L2 phonology and lexical access examines the ability of Russian-English bilinguals to block competing words in their L1 and L2 (Marian et al., 2008). The bilinguals had started learning English around 9-years old and at the time of the experiment their average age was 22 years and English was their dominant or preferred language. This study used a lexical decision experiment, in which the participants had to decide if the word they heard was a real word or not. The target words had a controlled number of shared phonemes between Russian and English. They had 10 words in each of the following categories: 0 phoneme overlap (only phonemes unique to each language), 2 unique:1 shared, 1 unique:2 shared, or 3 shared phonemes. A native English speaker and a native Russian speaker recorded the English and Russian words respectively.

It was found that when a word in English had a phonological overlap with a word in their native Russian this facilitated their response compared to non-overlapping words. Phonological overlap had a less straight-forward effect when processing their L1 but generally led to inhibition and inaccuracies. These results show, first, that the competitor language is activated in both scenarios and, second, that the relationships between L1 and L2 phonetic categories are asymmetrical. While this study does not directly investigate the effects of L1 accent in L2, it provides support of PAM in their claim that L2 listeners base some of their L2 phonetic categories on L1. Sounds which contain phonetic cues habitually recognised in the L1, even if rendered by a native speaker of the L2 led to processing facilitation. The L1 categories of the listeners in Marian (2008) appear to be more stable than their L2 phonetic categories, despite their immersion in L2.

One of the few other studies that investigated the perception of L2 accent on L1-matched and -mismatched non-native listeners focused specifically on the effect of initial-phoneme substitution (Hanulíková and Weber, 2012). They wanted to determine if word perception in L2 listeners will be affected when they are listening to TH-initial words, in which the target /θ/ is substituted by either the more perceptually similar /f/ or the more frequently preferred in production /s/ (for German (L1) – English (L2) bilinguals) or /t/ (for Dutch (L1) – English (L2) bilinguals). In two preliminary experiments they determined that all of their participants perceive /f/ to be the most acoustically similar substitute to /θ/. In addition, in a production task they determined that while both German and Dutch speakers produce all three substitutions, /s/ is the most common one for German speakers and /t/ is the most common for Dutch speakers.

In their main eye-tracking experiment Hanulíková and Weber (2012) tested how much the participants fixate their gaze on the target word depending on whether the auditory stimulus had the word-initial /θ/substituted by /t/, /s/ or /f/. They found that Dutch listeners fixated on the target word more when the auditory stimulus started with /t/. German listeners fixated more when the target stimulus started with /s/. A control group of native English listeners showed no preference for any of the substitutions. Also, surprisingly, the listeners did not show different behaviour depending on whether the speaker was Dutch or German. The listeners were flexible in accepting Dutch-specific variation from a German speaker and German variation from a Dutch speaker and responded favourably to their preferred substitution even when it appeared in the wrong L1 accent.

The main finding of this study is that the more frequent substitute in one's variety, not the more objectively and perceptually similar one, leads to improved processing. They also found no correlation between each individual's rate of production of a specific substitution and the eye-tracking results. Despite it being only a post-hoc correlation test, it suggests that no link can be made between individual production of a phoneme-substitution and its processing in perception. The results are only explained with the variety-specific frequency of the substitution.

Regarding the lack of effect of L1-match between listeners and speakers, the authors propose that the listeners employ an accent-general, as opposed to accent-specific word-recognition mechanism for the L2 listeners. Before the eye-tracking phase of the experiment the participants of Hanulíková and Weber (2012) had heard different non-native English accents, which might have led them to perceptually adjust to an accent-general mechanism by the time the eye-tracking was carried out (see also Weber et al., 2014). This interpretation contradicts the results of Ludwig and Mora (2017) and is at odds with finding that after all the accent-specific /θ/ substitutions did affect processing.

It is possible the speakers' L1 did not affect the listeners' processing in Hanulíková and Weber (2012) because of the initial position of the crucial phoneme and the nature of the method. Ludwig and Mora (2017) used reaction times that were recorded after the whole word was heard and they investigated the effect of an overall Catalan and overall German accent in English. Contrary to that, Hanulíková and Weber (2012) used eye-tracking, which measures duration and frequency of gaze fixations during the processing and not after the processing is competed. In addition, they included the accent-specific /θ/ substitution at the onset of the word. It has been demonstrated in L1 listening that the onsets of words carry more weight in word-processing (Marslen-Wilson and Zwitserlood, 1989; Tyler and Wessels, 1983). This is also one of the underpinnings of the Cohort model of speech processing (Marslen-Wilson and Welsh, 1978). Therefore, it is possible that *during* word-processing and *after* hearing the accent-specific initial substitution of /θ/, the phonetic nuance of the overall accent has less of a role on the rest of the processing. The mechanics of accent processing by non-native listeners require a lot more investigation.

Overall, all studies discussed in this section have demonstrated evidence of some L2 listeners processing speech with L1 phonetic characteristics more efficiently than native speech or other accents. This is consistent with the predictions of ASP and SLM for low proficiency bilinguals. The next section will present some counterevidence to these observations.

3.1.2. L1-matched foreign accent is inhibitory

The studies reviewed so far are directly challenged by several pieces of research, discussed in this subsection.

It has been reported that Dutch-English bilinguals had slower reaction times when they performed a single-word auditory lexical decision task while listening to Dutch-accented English, compared to when they listened to native English speech, even though that was not the major focus of their study (Lagrou et al., 2011). Their proficiency was recorded as a self-report on a 7-point Likert scale. One possible explanation of this effect, in line with BIMOLA, is that the Dutch accent activated competitor representations in Dutch, which interfered with the task of recognising words in English. However, if this explanation is accepted in the present case, then it is unclear why it would not affect the lower proficiency listeners in Ludwig and Mora (2017) for whom the L1 representations would be just as, if not more, salient. Given the contradictory results of Lagrou et al. (2011) and Ludwig and Mora (2017), the factors affecting native accent processing in L2 deserve more attention and systematic study.

Similarly to the Dutch listeners, Danish-English bilinguals did not process Danish-accented English faster or more accurately compared to other non-native accents of English, such as Swedish, German or Japanese (Jensen and Thøgersen, 2017). In addition, the Danish-accented speaker, who had a General American English target pronunciation was understood a lot faster than the Danish-accented speaker with an Estuary English pronunciation and the native English speaker with an Estuary accent. This suggests that the variety of the native speakers and the intended variety of the Danish L2 speakers of English can also affect speech reaction times for the Danish-English bilingual listeners.

In that study, some utterances were overtly rated as having a strong accent but were associated with high accuracy in a simple task (sentence processing). However, when two of the speakers who had similar intelligibility ratings and accuracy were used in a complex task (presenting a university lecture), the listeners had lower accuracy for the speaker who yielded slower reaction times in the simple task. This suggests that the accuracy of speech processing is dependent on the cognitive complexity of the listening task. The accuracy in a complex task is better predicted by the reaction times not the accuracy of listeners during a simple task. This suggests that speed of processing is a more nuanced measure of the difficulty with an accent than accuracy.

As mentioned earlier, one of the reasons for the difficulty of processing L1-accented L2 might be because L1-like phonetic detail activates L1 vocabulary which competes for selection during L2 processing. This effect was demonstrated in L1 processing using L2 phonetic information. Spanish-English bilinguals were presented with Spanish words whose initial stop consonants had either English-specific or Spanish-specific voice onset time (VOT) (Ju and Luce, 2004). Eye-tracking data showed that the participants fixated their gaze more often on non-target pictures whose English names had a phonological similarity with the Spanish target when the initial consonant of the Spanish auditory stimulus had English-specific VOT. This means that the listeners looked, for example, at a picture of “plyers” instead of the target picture of a beach ("playa") when the initial consonant was aspirated. Although this study attests processing of L2 accent in L1, these findings fit with of BIMOLA, as the English-specific VOT would activate more strongly competitor words in English and affect the gaze of the participants. These results are also consistent with the L2 phonological interference in L1 reported by Marian et al. (2008).

Contrary to the results of Lagrou et al. (2011) and Jensen and Thørgersen (2017), Wagner (2016) reports no systematic difference in the processing of native English and Dutch-accented English by Dutch-English bilinguals. The primary purpose of the study was to induce the effect of listening to the wrong language. Dutch-English bilingual listeners, also university students as in Lagrou et al. (2011), were led to believe that they would do a sentence verification task in one language, only to have a guest language introduced after one third of the experiment. The speakers of the guest stimuli were also either native or non-native speakers of the language (i.e., Dutch-accented English and English-accented Dutch). The results show that the listeners were slower to process the less familiar to them English-accented Dutch than native Dutch. However, there was no systematic difference in their processing times of Dutch-accented English compared to native English. The author argues that this effect was caused by the listeners' familiarity with both Dutch-accented and native English speech. In addition, the fact that they were processing whole sentences as opposed to single words may have allowed the listeners more time to adjust to any unpredicted variation. This result resembles the findings of Ludwig and Mora (2017) for their higher proficiency listeners, and is indicative of a flexible listener, in the terminology of IAF (Kleinschmidt and Jaeger, 2015a).

Another study, which shows a lack of L1-accent processing benefit compared the processing of a genuine and an arbitrary foreign accent to a standard native-like accent (Weber et al., 2014). A native Italian speaker who was also highly proficient in English recorded English stimuli with lengthened /ɪ/ (as in /tri:k/ for "trick") as genuine accent stimuli and words with short /i/ (as in */trɪt/ for "treat") as arbitrary accent stimuli. The words were used as auditory primes for written targets in a lexical decision task. The participants were native and L1-matched non-native listeners. Although not the focus of the study, the results show that the productions which were closest to native-like productions led to faster processing by Italian-English bilingual listeners than Italian-accented productions (also true for native English listeners and Dutch-English bilingual listeners).

Independently of this lack of L1 accent benefit, Weber et al. (2014) argue that bilinguals are more flexible listeners than monolinguals. Functionally monolingual native English listeners were successfully primed by the genuine Italian accent stimuli but were not primed by the "arbitrary" accented stimuli. L1-matched (Italian) and L1-unmatched (Dutch) listeners took part in the same task and they were primed by both the genuine and arbitrary foreign accent realisation. Additional tests showed that there was an overall difference between the L1 English listeners but there were no differences between the L1 Dutch and L1 Italian listeners with respect to their preference for both the arbitrary and genuine accent. Weber et al. (2014) suggest that their result is partially due to the monolingual-bilingual distinction between the listeners, claiming that bilingualism improves perceptual flexibility. They support this claim by replicating their study using multilingual L1 Dutch listeners (contrary to the monolingual L1 English listeners in the first experiment) who listened to Italian-accented Dutch. They found that the listeners were equally primed by the arbitrary and genuine accent forms. These results contradict the findings of Ludwig and Mora (2017) whose participants performed worse when listening to accents different from their own and who were also not better than monolingual English speakers at the unfamiliar foreign accents.

It is beyond the scope of this chapter to provide a complete critique of this study. The main criticism that needs to be pointed out is that short /i/ realisations, which are considered an "arbitrary" accent, are not in fact arbitrary. They can be frequently observed in Italian learner productions of English from the Speech Accent Archive (Weinberger, 2015). Both /i/ shortenings and /ɪ/ elongations are part of the Italian accent (which also has considerable geographical variation) and are anecdotally attested as stereotypical features of the accent as well, such as in the punchlines of the viral comedy video "The Italian Tourist" (ktz_33, 2007). Nevertheless, even if short /i/ is indeed a legitimate part of the variation within the Italian accent of English, it may still be rarer than the long /ɪ/ pronunciation.

In that case, the suggestion that bilingual listeners are more flexible than monolingual listeners would mirror an earlier discussion in Section 1.3.1. The Ideal Adapter Framework predicts that listeners with a greater repertoire of phonetic experience will be more "flexible" than listeners with a more limited experience. This explanation accounts for the data of the Italian-English bilinguals. Weber et al. (2014) offer no concrete explanation from a theoretical point of view as to why the Dutch-English bilinguals would be flexible with regard to Italian-accented phonetic variation in English. According to IAF, that would be possible if they have experience listening to Italian-accented English or if similar phonetic variation is also observed in Dutch-accented English.

Another evidence for a lack of L1 accent benefit comes from a study by Bien et al. (2016), a follow-up to Hanulíková et al. (2012). It focuses on the perception of the same replacement phonemes /s/ and /t/ instead of /θ/ but within non-words, using the Mismatch Negativity approach (MMN), a type of early event-related potential (ERP) of the brain. MMN effects are observed when a listener gets accustomed to a specific type of repetitive auditory input and that input is noticeably changed. For example, in the experiment in question the participants got familiarised with the non-word 'θond' repeated multiple times with different pitches and was at some point replaced with the non-word 'sond' or 'tond' produced by the same speaker. It was expected that Dutch (L1) - English (L2) listeners will have a smaller MMN effect for 'tond' than 'sond' because /t/ is a habitual substitution for /θ/ for Dutch speakers of English. Similarly, it was expected that German (L1) - English (L2) listeners will have a smaller MMN effect for 'sond' than 'tond' because /s/ is a habitual substitute for /θ/ for German speakers of English (see Hanulíková and Weber, 2012). However, the results suggest that both German and Dutch listeners had very similar MMN effects for both 'tond' and 'sond'. This result means that increased experience with a specific phonemic substitute within one's L2 English speech does not affect its representation in pre-attentive processing. This result is taken to reinterpret the evidence of Hanulíková and Weber (2012). The preference for own-accent substitutes reported by the eye tracking study has to occur at a later stage, during lexical processing, not purely as a result of phonetic similarity (cf., Cai et al., 2017). This contradicts the predictions of BIMOLA, that the perceptual similarity of phonemes is determined based on a number of shared phonological features.

Some innovative studies investigating accented-speech processing by non-native listeners provide ambiguous results, whose interpretation is hindered by the incomplete understanding of the methods. Nevertheless, they suggest novel possibilities about where a potential advantage or disadvantage may come from. Song and Iverson (2018) investigated how L1 English listeners and Korean (L1) – English (L2) bilinguals process native English speech and Korean-accented speech, using EEG. They focused on the process of entrainment, which has been previously observed in the presence of more intelligible and clear acoustic signal (Kong et al., 2015; Rimmele et al., 2015).

Song and Iverson (2018) define neural entrainment as synchronicity (phase-locking) between the firing of action potentials of neurons and the amplitude envelope of an utterance. This definition has also been used in all of the literature they refer to. Hence, it is assumed that entrainment involves adjustment to suprasegmental speech characteristics. While in previous studies, entrainment has been interpreted as a process of exploiting higher-level linguistic information to decode the acoustic signal (a top-down process), the experiment of Song and Iverson (2018), shows more entrainment when participants are listening to accented speech and also in L2 listeners compared to L1 listeners processing native speech. The listeners' proficiency was not measured but they had lived in an English-speaking country for an average of one year and had started learning English around the age of ten.

According to Song and Iverson (2018) this result poses issues because it is generally considered that L2 listeners' higher linguistic structures, such as syntax and vocabulary, are less robustly set up than those of L1 listeners. Following this assumption, it would be more likely for L1 listeners to show higher entrainment. Song and Iverson (2018) attribute the past observations of lower entrainment to poorer sound quality as in previous studies speech has been usually embedded in noise, other speech or has been vocoded (e.g., Howard and Poeppel, 2010). In Song and Iverson's (2018) experiment, the sound quality was not manipulated (although the task still required the listeners to inhibit a distractor speaker), hence entrainment can be considered an effect of the stimuli's accent and/or the listeners' experience of the given accent. However, in the absence of noise native English listeners should be able to reach the intended meaning from the acoustic input relying less on context than non-native English listeners. This interpretation of the results would be consistent with the Ideal Adapter Framework (Kleinschmidt and Jaeger, 2015a), discussed in Section 1.3.1. There it was suggested that when listeners are faced with bottom-up phonetic input that does not match their stored variability for the phonetic cues in question, they need to rely more heavily on the context (including suprasegmental cues) to reconstruct the intended meaning.

However, despite similar accuracy levels, the L1 Korean listeners showed higher entrainment when listening to Korean-accented English speech than native English speech. This suggests that the results are consistent with Lagrou et al. (2011) who demonstrated that Dutch-accented English leads to slower RTs for Dutch listeners than native English speech. Unfortunately, there is no specific information on the English proficiency of the Korean participants in Song and Iverson (2018), so it is hard to make further interpretations of this result.

3.1.3. Summary

Overall, most of the research reviewed in the previous two sections provides highly mixed evidence regarding any benefits of L1-accented L2 processing. Generally, benefits in speed and accuracy of processing are associated with lower proficiency in the L2 and some level of personal experience with the accent. Multiple theoretical explanations were brought up.

Most models reviewed in the earlier chapters allow for a dynamic updating of phonetic representations and would allow for shared representations across languages (e.g., Ideal Adapter Framework, Thousand Brains Intelligence Model, Speech Learning Model). This means that information processed in one language can activate related information in the other. According to the architecture of one of the more conservative models, namely BIMOLA, phonetic information that is not relevant to the activation of a specific phonemic feature, necessary for the recognition of a phoneme, is discarded. However, under such a rigid architecture, fine phonetic detail signalling L1 accent might not affect speech processing and activate the other language, unless it contributes to the decoding of a relevant phonological feature.

Overall, the majority of studies suggest that non-native listeners might have disadvantages when processing L1 matched and mismatched foreign accents. The evidence of benefits is scarcer but not completely inexistent (e.g., Hanulíková and Weber, 2012; Hayes-Harb et al., 2008; Ludwig and Mora, 2017; Major et al., 2002).

3.2. Broader perspective on foreign-accent processing: phonetic similarity and phonetic exposure

In order to understand the source of the potential benefits or disadvantages of processing an L1 accent that matches the listener's own, two factors need to be disentangled. The L1-match could produce processing benefits because of its phonetic similarity to the listener's own accent or because the listener has greater exposure for this particular accent from other speakers. Both factors, accent similarity and the amount of exposure to an accent, feed into the concepts of phonetic representations and predictability discussed earlier in the chapter. Even assuming that listeners' representations are constantly being updated, according to IAF (Kleinschmidt and Jaeger, 2015a) and TBIM (Hawkins et al., 2019), at any given point in time there are specificities of the acoustic input that are more predictable than others and would therefore be processed faster. After a short exposure to a new and unusual accent, the training effect can wear off and a representation, similar to the original one, remains (Bieber and Gordon-Salant, 2017). If Accent A requires more time for adaptation or requires adaptation in more contexts than Accent B then it can be assumed that in that set of times, Accent B is more closely aligned with the listener's internal representation than Accent A.

If the individual's own accent is at the core of their phonetic representations (because it resembles the accent spoken in that person's surroundings, or due to its higher social status or attractiveness) then it is expected that speech that resembles the listener's own variety would be processed faster in most contexts. However, if circumstances such as immigration or negative attitude to one's own accent render the listener's own variety a less prestigious minority of their input, then it is possible that accents dissimilar to one's own would be easier to process. This section will investigate both possibilities for accent preference and try to contextualise them using the speech processing models discussed earlier.

3.2.1. Accents different from the listener's may be inhibitory

The concept of a foreign accent (or even regional accent for that matter) is hard to narrow down to a specific phonetic pattern of divergence from a native/standard production. Some elements of a foreign accent that can be expected in single-word production can be related to stress-reassignment, achieving native targets for phonemic segments and coarticulation. For a non-native speaker any of these deviations can be caused by transfer from L1 phonology, a misinterpretation of how the L2 target is supposed to be articulated or both. In addition, people with the same L1 background can do different types of transfers and targeting different native varieties in the L2 they are learning. For example, in Jensen and Thøgersen (2017) the two Danish-English bilinguals spoke with an American and an Estuary accent, which affected both the speed and accuracy with which they were understood. On the opposite side of the process, every listener (be it native or non-native) will have their own slightly different prior expectations of what speech in the language they are listening to is supposed to sound like. Because traditionally a native or a standard variety speaker is assumed as a default it is hard to predict how a listener who diverges from that mould would interact with accents that also diverge.

3.2.1.1. Theoretical discussion

Two theoretical stances have been proposed in the investigation of perception of unfamiliar accents by native listeners: the Perceptual Distance Hypothesis (PDH) (Clarke and Garrett, 2004; Floccia et al., 2006) and the Different Processes Hypotheses (DPH) (Goslin et al., 2012). Both focus on native listeners. According to the former, the more phonetically distant an accent is from the optimal native accent for a listener, the harder it will be to process. PDH does not make a distinction between natively and non-natively produced accents. The latter hypothesis, DPH, makes a distinction between native accents (e.g., regional variation) and foreign accents (produced by non-native speakers of a language), stating that different pre-lexical perceptual processes underlie the listening of these two types of accents and that they will be associated with different processing costs.

The Perceptual Distance Hypothesis has the advantage of assuming a gradient difference between a dialect and language, which is consistent with some recent theoretical analyses of the topic (de Bot and Jaensch, 2015). However, as mentioned earlier, specifying a phonetic recipe of divergence between two varieties is a challenging task, which renders the Perceptual Distance Hypothesis vague and hard to test. In addition, as discussed in Section 3.1.1, Hanulíková and Weber (2012) demonstrated that it is not the most acoustically similar substitute of /θ/ that triggers the most gazes to the target word but it is the substitute that occurs most frequently in the listeners' variety of English. This is evidence that in the context of word processing the experience with a phonetic substitution trumps the independent judgements of phonetic similarity.

The Different Processes Hypothesis predicts that listening to regional native accents requires attuning to phonological categories that already exist in the native variety but with realisations that are predictably and systematically deviant (Goslin et al., 2012). In contrast, listening to a foreign accent requires major pre-lexical adjustments to prosody and segmental alignment that might be unpredictably influenced by the speaker's L1 and proficiency (Goslin et al., 2012). This has been supported by studies from the developmental literature (Girard et al., 2008) as well as studies demonstrating different processing times and ERP negativity for regional vs. foreign accents (Adank et al., 2009; Floccia et al., 2009; Goslin et al., 2012).

In the context of the Automatic Selective Perception (ASP) model by Strange (2011) this hypothesis might be accounted for by different types of attention required for listening to different accents. According to Strange (2011), when faced with a familiar accent a listener would rely on well-established selective perception routines that allow them to efficiently use only the minimal necessary phonetic information for successful word recognition. This type of processing is automatic and is associated with the so-called phonological listening. However, an unusual accent that requires drastic remapping of phonological categories would be associated with the more cognitively demanding phonetic listening (neurological evidence for this type of listening might be found in Perani (1996, 1998).

The main issue that can be raised with the formulation of the Different Processes Hypothesis is that the notions of foreign accent and native variety need to be specified in more detail. In the case of global languages such as English and French, for example, it is possible to have native listeners of a specific variety who are much more experienced with a foreign-accented speaker group they interact frequently with than with a geographically distant but native variety of English or French. Treating "foreign accents" and "native accents" as monolithic groups is an obvious starting point in challenging this hypothesis (cf., Porretta et al., 2016). However, when reviewed in the context of the ASP model (Strange, 2011), it appears that the Perceptual Distance Hypothesis might be applicable to non-native and native listeners alike, assuming they are engaged in very specific listening tasks.

Indeed, it has been demonstrated that monolingual listeners can have a fairly nuanced perception of different accents. In a free classification task Bent, Atagi, Akbik, and Bonifield (2016) tested how monolingual American English listeners group US regional accents, other native English accents and foreign accents. They found that foreign and native accents are perceptually distinct to the listeners. Within the native accents, some non-US accents were clustered together with the regional US ones. Within the non-native accents, there were three subgroups: first, French and German; second, Asian accents; third, a mix of European, Asian and African accents. The study demonstrates that listeners can identify different varieties of English in a nuanced way. It remains to be investigated what the relationship between this categorisation and on-line speech processing is.

A criticism that can be raised for both hypotheses is that they implicitly make a strong assumption that each listener has a single most optimal-for-processing variety and exclude the possibility that a listener is equally good at processing several varieties. Therefore, these hypotheses predict that any divergences from the most optimal variety would inhibit speech processing. This has been potentially supported by Sumner and Samuel (2009) who showed that rhotic speakers of American English, who live in areas where a non-rhotic accent is common, use non-rhotic primes less optimally than actual speakers of a non-rhotic variety.

Assuming each listener has a single most-optimal variety for processing is problematic. It is not known if the most optimal variety is more likely to match the listeners' own variety or the standard prestige variety of the region. The latter phenomenon is observed by Clopper (2014) and Szakay et al. (2016), who report overall faster processing for the standard as opposed to the own variety. In addition, numerous studies have demonstrated that native listeners can adapt to phonetic variation and accents in the duration of an experimental task (Clarke and Garrett, 2004; Kraljic et al., 2008; Lennon et al., 2016; Maye et al., 2008). With every new input the perceptual system of an individual can incur unpredictable short-term and long-term modifications (Kleinschmidt and Jaeger, 2015a).

Even if this assumption can be reconciled for native listeners with limited experience of different varieties and languages at a single time point, the problem becomes more complex when L2 listeners are taken into account. Because their L2 competencies are developing after their childhood, there is an underlying assumption that an adult L2 learner is continuously changing, and "attuning" their listening skills to the novel language they are exposed to (even if it is assumed that they plateau after some time). At an earlier point in their development of L2 phonetic categories and lexicon they might have a processing benefit from listening to L1-accented L2 compared to a native accent (Ludwig and Mora, 2017). However, as they acquire more exposure to native speech in their L2, it is possible that they reach an ambivalent point where both a native accent and an L1-matched foreign accent are processed in a similar way (Ludwig and Mora, 2017; Wagner, 2016). At a stage of L2 acquisition when the non-native listeners do not have firmly established phonemic boundaries (Broersma, 2012), they might require extra time to exclude all possible lexical competitors, regardless of the specific accent of the input. It might also be possible that as the bilinguals become highly proficient and mostly exposed to native productions of their L2 it is possible that they start processing L1-accented L2 slower than natively produced L2 (Lagrou et al., 2011). As discussed below, Porretta et al. (2016) demonstrate that experience and foreign-accent strength affect native listeners' reaction times in a non-linear fashion. It needs to be investigated in more detail to what extent a bilingual listener has a single optimal variety for processing.

While many points of criticism can be raised against the Different Processes Hypothesis and the Perceptual Distance Hypothesis, they still provide a useful starting point when discussing foreign accent processing and helpfully provide an angle of analysis when new research questions are investigated. As mentioned earlier, they emphasise the qualities of the input as the main reason for specific perceptual outcomes and downplaying the role of the listener. By contrast, a dynamic approach to L2 perception would also require the consideration of the effects of L1 knowledge on L2 processing.

3.2.1.2. Examples

According to Porretta et al. (2016) both hypotheses are partially supported because of their vagueness, although the Perceptual Distance Hypothesis was better suited to capture the gradient nature of their results. Porretta et al. (2016) investigated the effect of gradual foreign accentedness increase and the listeners' prior experience with the accent on lexical activation. They used native English listeners with varying experience of Mandarin-accented English and auditory stimuli with different strengths of Mandarin accent. They discovered that the listeners' experience with the accent improved their processing times, in line with the Perceptual Distance Hypothesis. In addition, an eye-tracking experiment demonstrated that as accentedness increased, word recognition slowed down, even though experience with the accent improved performance. The main innovation in their study is that they treated accent as a multi-dimensional continuum, as opposed to a factorial variable as done before in a study with similar findings (Witteman et al., 2013). However, Porretta et al. (2016) found an interaction between experience and accent strength, where a mild-to moderate accent was the fastest to process compared to milder and stronger variants of the accent. In that way the study contradicts the linear predictions of the Perceptual Distance Hypothesis.

Another study which focused on native listeners, like Poretta et al. (2016), operationalised the accent strength variable by acoustically modifying the stimuli instead of relying on accentedness ratings. Wingstedt and Schulman (1987) tested Swedish listeners' ability to understand artificial foreign-accent in Swedish, created by carefully controlling different acoustic characteristics of the stimuli. The authors induced three types of changes: of segmental rules, word stress, and vowel insertion in consonant clusters. The listeners' mistakes in word identification were affected by the type of acoustic modification and increased when more than one type of source of accent were combined. Identification was at its best when only one source of accent was introduced and at its worst when all three were included. Intermediate combinations led to intermediate accuracy. The main outcome of this study is that the overall accuracy of understanding an accent might be predicted depending on the acoustic characteristics of the stimuli. This evidence suggests that a foreign accent can affect word-processing at all stages of the word structure – segmental, syllabic and suprasegmental (cf., Reinisch and Weber, 2012).

Unlike the studies reviewed in this section, Ockey and French (2016) investigated the effect of accent strength in native English speech (as opposed to non-native speech) on the comprehension of non-native listeners. They investigated the listening comprehension results of 21,726 TOEFL iBT test takers from 148 countries. In this case, accent strength was measured from the point of view of native speakers of American English listening to speakers with American, British and Australian accents. The results suggest that both familiarity with an accent and accent-strength played a role for the comprehension of non-native listeners. One of the most striking results was that accents that were rated as only mildly unusual (but not requiring extra concentration) by the American listeners led to a significant decline in the test takers' results. This result changed if the test takers had a higher familiarity with the accent (cf., Adank et al., 2009; Harding, 2012; Major et al., 2002).

The studies reviewed in this section support the more intuitive hypothesis that listening to an accent different from one's own would lead to poorer listening outcomes. This can be explained by the lower familiarity and even lower prestige of the different accent. The Perceptual Distance Hypothesis assumes that increased acoustic distance from one's optimal accent would lead to poorer listening outcomes, while the Different Processes Hypothesis also takes into account the level of variability within the different accent. Both hypotheses were evaluated and criticised as vague, in light of the complex evidence of regional and L2 speech processing. The next section will explore studies that report the opposite phenomenon and thus contradict the expectations of both PDH and DPH.

3.2.2. Accents different from the listener's may be facilitatory

This section will review evidence of facilitatory effects of listening to accents different from the listener's own and the process of adapting to an unfamiliar accent.

There is some evidence suggesting that listener's own productions are not always the best predictor of their internal representations, in addition to the studies discussed in the previous section. As reviewed in Section 3.1.1 Hanulíková and Weber (2012) found that listeners process the phonetic substitution of /θ/ that was typical of their overall accent group (German or Dutch) more efficiently than other substitutions. However, they found no relationship between listeners' own initial /θ/ substitutions and their preferred substitution in perception. It appears that in this case the benefit of one's own accent was only observed on a community but not on an individual level.

In addition, Cooper et al. (2018) investigated the phonetic representations for 30-36 month old children. In an eye-tracking study they presented toddlers with their own speech, the speech of another toddler, their own mother, and another adult. They found that young children model their perceptual representations on adult speech and not on their own or on other toddlers' speech. There was no own-voice or own-mother voice benefit. It remains an open question why the toddlers' representations were closer to adult speech as opposed to toddler speech. Without knowledge of the children's experience with different voices, this preference could be explained either by a greater amount of exposure to adult speech or by an implicit importance associated with the voices of grown-ups.

The role of prestige of a certain variety can be traced in studies involving listening to regional and standard varieties. One such example is the study of Glaswegian and Standard Southern British English (SSBE) listeners exposed to Glaswegian and SSBE speech (Adank et al., 2009). The Glaswegians found both varieties equally intelligible, while the SSBE listeners struggled with Glaswegian speech. In a study reviewed earlier, Māori-English bilinguals generally processed Pākehā English (Standard New Zealand English) speech faster than Māori English (Szakay et al., 2016). The processing benefit of listening to a standard accent compared to one's own was also demonstrated by Clopper and Bradlow (2008). In a study on dialect identification and processing for native speakers of American English in different levels of noise, the main finding was that regardless of the listener's dialect, the standard General American variety was the processed most accurately.

The bilingual listeners' ability to discriminate specific sounds is related to their ability to produce them but that does not have to be reflected in their habitual realisations. Llompart and Reinisch (2018) investigated the relationship of L2 user's ability to imitate, perceive and produce difficult vowel contrasts. Specifically, they looked at the /ɛ/ - /æ/ contrast for German learners of English and discovered that the ability to imitate was strongly related to the ability to categorise the vowels. However, there was no link between the bilinguals' typical realisations of the vowels and their ability to imitate them. This study raises the particularly interesting issue of how habitual productions are not necessarily the same as the speaker's ultimate ability to produce a given sound (or a sequence or sounds or not) and discriminate it from other sounds. The existence of this divergence might be related to the different underlying mechanisms involved in imitation and spontaneous production (Hao and de Jong, 2016) but it is also dependent on the speaker's sense of identity and the speaking context. It is beyond the scope of this dissertation to discuss how individuality may be expressed in L2 speech. However, it is important to point out that the factors related to the speaker's sense of identity might be a strong reason for the divergences between the phonetic representations guiding speech processing and the representations guiding habitual speech production.

In addition, it needs to be pointed out that having some experience with a language or dialect does not guarantee that processing will improve linearly. Holliday (2016) found out that naïve listeners of Korean discriminated between the fricatives /s^ha/ and /s*a/ more accurately than novice learners but were not more accurate than advanced learners. This effect is in line with the predictions of the Perceptual Assimilation Model (Best and Tyler, 2007) and to some extent ASP (Strange, 2011). The naïve listeners would have paid more attention to the phonetic detail of the sounds, while the novice learners would have started to employ some selective perceptual routines (according to ASP) or develop category assimilations with their L1 according to PAM.

Ultimately, in order to achieve optimal accuracy and speed of processing (in particular, in a situation where top-down lexical information is limited) the input needs to resemble some of the listener's internal representations of the intended content. Factors such as experience with the variety and its high prestige can be beneficial for processing even if the listener's own variety is less considered less prestigious. The discussion so far has focused mainly on the effects of processing one variety compared to another. A more dynamic approach to the topic of other-accent processing would require an investigation into the process of developing other-accent competency over time.

3.2.3. Adaptation to other accents

Adaptation to speech can be studied as the change of reaction times and accuracy of speech processing over time. Most studies on accent adaptation involve a training state, sometimes preceded by a pre-test and compare the listening outcomes from these stages to at least one post-test. This section will summarise some of the important factors affecting adaptation to a novel accent.

Some studies such as Bradlow and Bent (2008) and Sidaras et al. (2009) found that a training phase including voices of several speakers leads to an accent-general, not just speaker-specific adaptation. Xie and Myers (2017) add further nuance to these findings. In a training phase that included the voice of only one speaker, they found evidence that listeners develop a speaker-specific but no accent-general adaptation. Native English listeners had a benefit when listening to a speaker with a novel accent who had a similar voice to the voice heard during training compared to a dissimilar voice of the same dialect. Xie and Myers (2017) interpret this finding as lack of evidence for a speaker-independent top-down adaptation to Mandarin accent. In a follow-up study Xie et al. (2018) showed that adaptation to a novel dissimilar accented speaker improves if the participants have slept after the training session (i.e., the overnight as opposed to the same-day group of listeners). These studies suggest that the process of creating a speaker-independent generalisation relies on a diversity of speakers heard during training and the consolidation processes the brain undergoes in a sleep state (also see Cairney et al., 2016; James et al., 2017).

Another piece of evidence for speaker-specific adaptation is provided by Reinisch et al. (2013). They presented half of their Dutch-English bilingual participants with Dutch-accented English stimuli or Dutch stimuli in a lexical decision task. In each language group, half of the listeners heard /s/ sounds replaced either with an ambiguous /s-f/ sound and the other half of the listeners heard an ambiguous /f/. After this exposure, the listeners heard the same Dutch speaker pronounce minimal pairs in Dutch that they had to categorise. Both language groups had adjusted their phonetic boundary of to the target sound to include the ambiguous realisation, suggesting that phonetic speaker adaptation happens regardless of the language and even within L2 speech with strong foreign accent markings. However, this recalibration was not generalised beyond the allophone. This was later replicated by Reinisch and Holt (2014) who found that attuning to a manipulated sound can happen within a global foreign accent this time for native English listeners who were unfamiliar with Dutch accented English. However, they also built on the initial study by showing that adaptation is not necessarily speaker specific. Indeed, it can generalise to a novel speaker but only if the stimuli from the speakers were sampled in a way that made them appear more perceptually similar (thus also supporting Xie and Myers, 2017).

The benefits of exposure to multiple speakers is demonstrated by Baese-Berk et al. (2013). They compared the processing of novel accents by three groups of listeners. The group which was trained using a variety of foreign accents had the highest accuracy scores compared to a group trained on only one speaker with a foreign accent or a group trained to no foreign accent. In addition, the listeners trained with multiple accents improved their scores for novel speakers whose accents were either included or not included in the training sample. This study suggests that experience with a greater phonetic variability can improve perception of novel foreign accents. This would be consistent with the IAF's predictions that "flexible listeners" might perform better than listeners who are selectively adapted to speech variation that is irrelevant for the task.

Long-term familiarity with a specific accent can also affect the level of adaptation in a short task. Smith et al. (2014) demonstrate that when listeners listened to ambiguous speech in loud noise, if the speakers shared the accent of the listener, their productions led to overall higher accuracy. However, Standard Southern British English (SSBE) listeners who were inexperienced with Glasgow speech achieved the least improvement, while Glaswegian listeners, who had long-term familiarity with both SSBE and their own accent achieved the biggest gains in accuracy in the post-test. In addition, this study also demonstrates that listeners improved more after listening to the same as opposed to a different speaker.

The listeners' long-term familiarity with the sex differences between voices compared to their long-term familiarity with the accent is investigated in Kriengwatana et al. (2016). They tested Dutch and Australian English listeners in a Go/No-go task, where they had to categorise the Dutch /ɛ/ and /ɪ/ (one of the vowels would be the Go and the other one the No-go stimulus). After they were familiarised with these stimuli within the task, the listeners were presented with four follow-ups: a block where the stimuli were pronounced by a different speaker of the same sex and accent; a different speaker with a different sex but with the same accent; a different speaker of the same sex but with a different accent; and a different speaker with a different sex and a different accent. The participants were given feedback whether their response was correct or incorrect on the familiar but not on the unfamiliar trials that introduced a variation of the speaker characteristics.

Under these conditions the listeners adapted to a new speaker of the opposite sex but were unable to adapt to a speaker with a different accent or a different accent and sex. That was true even for the Dutch listeners who had more experience with Dutch accents than Australian listeners. The authors concluded that adaptation to a novel accent requires a two-step process that starts with a pre-lexical adaptation to the novel speaker. Kriengwatana et al. (2016) posited that in order to successfully complete the second step of adaptation to an accent, the listeners need to know what lexical form was intended by the speaker. They demonstrated that with a follow-up experiment which only involved the Australian listeners. In that experiment the listeners got feedback about their correctness in the novel trials as well, allowing them to adapt even to an Accent + Sex change in the speaker.

From the point of view of a probabilistic model of speech perception this result is unsurprising. Sex differences between voices are observed more frequently on a daily basis compared to accent differences, and they tend to share similarities across languages. What is intriguing, however, is that in the no-feedback experiment there was no difference between the Dutch and the Australian listeners. This result suggests that in an experimental setting with timed responses prior experience with the accent may fail to be accessed on time. It would have been interesting to include Dutch listeners in the feedback experiment and see if their experience would have allowed them to use the feedback even more efficiently than the Australian listeners.

The evidence presented so far suggests that adaptation to a novel accent is a complex process that includes an accent-independent initial stage. One conclusion that can be drawn from this section is that successful adaptation may occur regardless of the language background of the listener and their initial level of familiarity. Given the necessary training and conditions (Tamminen et al., 2015; Xie et al., 2018) and sometimes even without them, adaptation can occur even with very little exposure to a specific accent (Clarke and Garrett, 2004) but may not last long (Bieber and Gordon-Salant, 2017).

3.3. Sociolinguistic and top-down factors

The insight that listeners tend to first adapt to a speaker before generalising that adaptation to the speakers' group identity (e.g., their accent), suggests that the idiosyncratic characteristics of the speaker might also affect the intelligibility of their speech. All human-produced, non-synthesised speech has inherent social characteristics; hence, it is important to be aware of their potential effect on speech processing. The following sections demonstrate that beliefs about the speaker's social characteristics can have a profound impact on speech processing.

As there is relatively little research on sociolinguistic factors affecting bilingual auditory perception, and they are rarely incorporated into models of speech perception in general, this section will draw upon some relevant studies on bidialectal perception. As argued by de Bot and Jaensch (2015) there is no clear distinction between bilingualism and bidialectism, hence it is assumed that the parallels drawn here will lead to useful insights for bilingual listening as well. The following discussion will centre on studies that demonstrate how the listener's beliefs about the language, variety, social class, age and race of the speaker affect their comprehension. After reviewing the evidence of top-down factors affecting speech processing, this evidence will be discussed in relation to some of the models presented earlier in the chapter.

3.3.1. Circumstantial priming and reverse stereotyping

In a highly controlled listening environment, even very subtle priming can affect listeners' beliefs about which accent they are listening to, which in turn affect how they perceive the acoustic stimuli. In Niedzielski (1999) American listeners from Detroit listened to the vowel production from a fellow Detroit speaker and had to match them to a set of synthesised vowels. One group of participants used a worksheet with the word "Canadian" on top and the other group had the word "Michigan", which was the only difference between the two groups. The "Canadian" group correctly perceived and matched the vowel raising in the stimuli, while the "Michigan" group did not and matched the stimuli to lower vowel variants. Niedzielski (1999) concludes that circumstantial sociolinguistic information affects the perception of vowels (however, see discussion on language settings in Caramazza et al., 1973, in Section 2.3.1).

In a similar design, Hay and Drager (2006) replicated the main results of Niedzielski (1999) using Australian listeners and a New Zealand speaker. However, none of their listeners were deceived by the guise and recognised the nationality of the speaker. This led Hay and Drager (2006) to argue that the misperception is due to the activation of a concept, rather than the overt categorising the speech with a specific social label. The idea of a social tag also appears in Szakay et al. (2016) and Cai et al. (2017), discussed further below. This hypothesis is supported by Hay and Drager (2010). They changed the listeners' perception of category boundaries by activating the concept of Australia and New Zealand using stuffed toys of kiwis and kangaroos even though the effects were less robust than in Hay and Drager (2006). However, a recent attempt for replication suggests the effect might be result of the task design and not the circumstantial primes (Walker et al., 2019). The studies mentioned so far tend to suggest that the listeners' meta-awareness of the speaker variety is independent of their subliminal ability to use environmental cues in the interpretation of fine acoustic detail. However, a failed replication is an important red flag, signalling that the perception tasks used in this line of research need to be more robustly tested in a variety of environments.

Directly related to these results is the research of Kang and Rubin (2009) on reverse linguistic stereotyping (RLS). It has been widely attested that listeners apply social judgements based on the speaker's accent. Speakers of certain stigmatised accents are often judged as less likeable (Campbell-Kibler, 2007) or less intelligent (Seligman et al., 1972). RLS refers to the phenomenon where listeners are affected by their beliefs about the speaker which leads to distorted perception of the speaker's language and proficiency to match the initial belief. Rubin (1992) even demonstrated that American listeners who expected to hear a non-standard accent had poorer accuracy of comprehension than other listeners who were led to believe that they were listening to a standard American pronunciation (even though both groups heard the same recording of clear native standard speech). One of the implications from this work is that just a belief about the speaker is sufficient to lower the listener's understanding, thus questioning how automatic speech processing of a familiar accent really is (cf., ASP in Strange, 2011). Hence, if the actual accent of the speaker is enough to signal their sociolinguistic background and that background happens to be stigmatised, the social belief about the speaker may have an added effect to the objective acoustic intelligibility of their speech (if objective intelligibility can be considered an independent construct).

Kang and Rubin (2009) expanded the same method and investigated if any of the listeners' characteristics predisposed them to greater RLS. Over a hundred listeners with diverse backgrounds listened to recordings of the same expert native speaker of standard North American English. Between the two recordings, which were accompanied with a picture of either an Asian or Caucasian face, they heard a filler recording produced by a speaker with a distinct East-Asian accent and lower intelligibility. Only ratings on the voice of the of standard NAE speaker were analysed. Native listeners had better accuracy than non-native listeners but both groups had the same scores on the cloze test for the Caucasian guise, which means that their understanding of the speech was affected. Previous experience of teaching English led to more "compassionate" rating of the Asian guise (Kang and Rubin, 2009, p. 453), however, experience of communicating with other non-native speakers did not have a significant effect (consistent with the review of Rubin and Lannutti, 2001). One of the main conclusions from this research is that both native and non-native listeners of a given language also engage in reverse linguistic stereotyping under the influence of circumstantial primes of race. Regular interactions with non-native speakers do not modify this effect.

Previously formed beliefs about speakers can affect the listeners' ability to remember information associated with these speakers and the listeners' judgement on their trustworthiness. Using monolingual listeners Foucart et al. (2017) also observed RLS. They demonstrated that linguistic profiling of foreign-accented speakers can have a long-term effect on information processing. Listeners were exposed to three types of speakers: with a socially superior foreign accent (FA), with a socially superior native accent (NA) and with a socially inferior NA, along with photographs and biographies of these speakers. Social status was cued only by the achievements and biographies introduced with the speaker, while the accent of the same speaker was used for both "higher" and "lower" social status profiles. After a learning phase, during which the speakers familiarised themselves with the photographs, biographies and the accents, they had to judge the truthfulness of trivia statements presented visually alongside one of the speakers' photograph. Participants tended to rate statements as 'false' more often when they were associated with the FA speaker than with the high-status NA speaker. A memory task revealed that participants remembered the information best when associated with high-status NA, followed by high-status FA, followed by low-status NA. Considering that both native accents were sampled from the same speaker, using the same accent, it appears that perception of foreign accented speech affects cognitive processes in listeners not just directly because of its acoustic properties but because of the social judgement that they trigger, as suggested earlier. This result parallels the reverse linguistic stereotyping reported by Kang and Rubin (2009) but using social class, implied by biographies and photographs instead of race. One criticism that can be raised towards Foucart et al. (2017) is that the separate effects of the biographies and photos associated with the auditory stimuli cannot be easily separated and it cannot be excluded that any observed effects are entirely due to the photographs.

In general, the studies described so far require two explanations. The first question is, what mechanism underlies the activation of a concept of a sociolinguistic identity, as in Niedzhielski (1999) and Hay and Drager (2006)? The second question is whether this process of concept activation is the same as reverse stereotyping. Of the speech processing models discussed so far, only TBIM explicitly specifies the convergence of information from all sensory experiences. According to TBIM, the environmental cues will pre-activate neurons of representations associated with those cues. In Niedzhielski (1999) that would be the words “Michigan” and “Canadian”. That pre-activation would allow the following congruent bottom-up acoustic input to reach these cells first and thus lead to an automatic adjustment of phonetic perception. The automaticity of this process is also compatible with the results of Hay and Drager (2006) who report that the listeners perception of phonetic detail was affected even though their meta linguistic beliefs of the speaker’s identity were not.

This explanation could also be applied to the process of reverse linguistic stereotyping (RLS). However, one major difference is that in addition to the automatic readjustment of phonetic boundaries, when RLS is induced the listeners also form explicit meta-linguistic beliefs about the speakers’ identity, which may play an additional and less predictable role on linguistic processing. In RLS the automatic adjustment of phonetic boundaries would be unproductive as the expected input would not match the perceived input. In Kang and Rubin (2009) the listeners perceived an East-Asian face as an indicator of a non-native or unfamiliar English accent. As a result, they might have either pre-activate their idea of a stereotypical East-Asian accent (see also Weber et al., 2014) and struggled to process the words that do not match this pre-activation, even though they happened to be in an accent they were familiar with. In addition, Kang and Rubin (2009) reported that the East-Asian condition led the listeners to rate the speaker with lower social attractiveness and Foukart et al. (2017)’s participants had lower confidence in speakers who were not the high-prestige native speaker. This negative social judgement might have caused them to automatically pay less attention to the speech and thus fail to notice that all the necessary phonetic information for optimal comprehension was present in the input. As discussed before in relation to ASP (Strange, 2011), the amount of attention that is paid to speech can potentially be an important aspect accounting for speech processing phenomena.

This discussion is crucial because if listeners form a specific negative meta-belief about the speaker (whether due to circumstantial or phonetic and lexical cues) they may also automatically lower the overall attention they pay to the speech and the phonetic information. As a result, all studies that demonstrate less accurate processing of “other” (e.g., non-local and non-native) accents may reflect not only the fact that the accent was unfamiliar to the listener but also that the listener automatically assigned less attention to the task of understanding that accent due to its lower sociolinguistic prestige.

3.3.2. Language-inherent priming

Several studies suggest that lexical access and acoustic processing in bilinguals/bidialectals can be affected by the top-down activation of a language triggered by the overall language or accent of the speaker, not just circumstantial primes. One of the ways top-down effects can arise from bottom-up evidence is when the listener creates a model of the speaker that they are hearing on the basis of their voice characteristics. Kim (2016) demonstrated that the listeners' belief of the speakers' age, cued by their overall vocal characteristics affects the speed at which the listeners achieve lexical access for some words. In an auditory lexical decision task, Korean listeners in their twenties responded to three types of words: youth slang, neutral and archaic vocabulary. The stimuli were recorded by two types of speakers: two over 60 years old and two in their early twenties. The results suggest that the vocabulary that was mismatched with the age of the speaker was processed more slowly and incorrectly by the listeners than stereotype-neutral vocabulary. This is taken as evidence supporting exemplar models of speech processing, which allows the listener to store the lexical item and the phonetic form that is most commonly observed together.

The relationship between the speaker's overall accent and the listeners' speed of processing has been investigated in-depth in the monolingual study by Cai et al. (2017). In a series of experiments, they determined that the strength of accent for an individual word (i.e., the raw phonetic characteristics of the signal) does not affect lexical access, instead the effect was achieved by the overall British or American accent in which the word (e.g., "bonnet") was embedded. They also found that a dialect-specific meaning is accessed faster when it is heard in the congruent accent (e.g., "bonnet" meaning a hat in a British accent), which is another piece of evidence supporting exemplar models of speech processing. The authors argue that listeners build up a speaker model (complete with indexical information about the speaker's dialect, age, health etc.) that can affect word processing in parallel with bottom-up information. However, they also assume that the phonetic speaker characteristics contribute to a speaker model in parallel to a process of gradual phoneme and syllable identification. They assume that the speaker characteristics affect the final stages of lexical access, as opposed to being an inherent part of subcomponent identification (also see Bien et al., 2016). This interpretation contradicts the IAF, which assumes that a speaker's phonetic characteristics are an inherent part of low-level phonetic recalibration and selective adaptation.

Cai et al. (2017)'s explanation is reminiscent of Grosjean's Language mode concept if it were applied in a cross-dialectal setting. When processing overall British-accented speech, the UK listeners are in a more monolectal mode compared to when they are listening to American-accented English. One of the earlier studies that demonstrates the effects of a language mode is by Elman et al. (1977). They found that bilingual listeners categorised the same acoustic input (syllables varying on the VOT continuum between 'pa' and 'ba') differently depending on the language that it was embedded in. A different result would be expected if the words were presented outside context, or within context that exhibits the same strength of accent as the target word. The concept of creating a speaker model and interpreting bottom-up evidence accordingly is compatible with the IAF and TBIM will be discussed again at the end of this section.

One of the very few studies that explicitly investigates sociolinguistic variation in a bilingual context is by Szakay et al. (2016). They studied the perception of bilingual participants (English L1 and Māori L2) listening to Māori English and Pākehā English (i.e., standard New Zealand English). Māori English is a variety of English, which is identifiable through systematic phonological and phonetic characteristics, and it is not Māori-accented L2 English. It is sociolinguistically related to Māori as some Māori English speakers are bilingual with Māori, and few people report native-like proficiency in Māori. Māori English was the native language of the bilingual participants. The study found that auditory Māori primes facilitated the processing of Māori English translation equivalents but not of Pākehā English words. Szakay et al. (2016) argued that the sociolinguistic link that unites Māori English and Māori is one of the main factors for this facilitation.

The English - Māori bilingual participants took part in an auditory lexical decision task with auditory priming. When the primes were in Māori, the targets were translation equivalents in Māori English or Pākehā English and vice versa. In forward priming (L1 to L2), both Pākehā and Māori English facilitated Māori words and the effect of Māori English had a much greater magnitude than Pākehā English. However, in backward priming, Māori words only facilitated Māori English but not Pākehā English. The effect was strongest for the participants who were the most immersed in Māori language.

As mentioned previously, backward priming is usually reported as more difficult (Basnight-Brown and Altarriba, 2007; de Groot and Nas, 1991; Dunabeitia et al., 2014; Gollan et al., 1997; Jiang, 1999; Jiang and Forster, 2001). Szakay et al. (2016) use that to suggest that a shared social category between Māori and Māori English is at play, resulting in a facilitative priming effect during the otherwise difficult backward priming. They propose that the words in Māori and Māori English share a sociolinguistic tag that allows faster processing of Māori English, when primed with Māori translations. Although not discussed by the authors, in neurological terms, a "tag" potentially stands for neural connections. An alternative explanation for the results that has not been discussed by Szakay el al. (2016) is that there may have been a stronger objective phonetic similarity between Māori and Māori English compared to Māori and Pākehā English. Following the predations of BIMOLA, the Māori phonetics could have activated Māori English vocabulary without needing to assume the presence of an overt social tag. The concept of a social tag or facilitation through phonetic characteristics could be explained by TBIM as a result of preactivation of relevant cortical columns as a result of past exposure to that phonetic variation in a similar situation. Perhaps the listeners were more accustomed to hearing Māori and Māori English spoken in the same environment than Māori and Pākehā English.

There is also a possibility to account for the Māori priming effect solely by using the asymmetry of dominance between the different languages. As discussed in Section 2.3, the asymmetry between forward and backward priming is usually explained by positing that the L1 vocabulary is so highly active anyway, that priming with L2 cannot contribute much more. That would suggest that Māori English vocabulary is less active for the English – Māori bilinguals than the socially dominant Pākehā variety (cf. Clopper, 2017). Hence even if they did not share a social tag, the L2 vocabulary may have facilitated the Māori English L1 vocabulary relative to the Pākehā words. To support this explanation a new experiment needs to replace the L2 Māori with a sociolinguistically unrelated L2. Nevertheless, as there is no easy way of establishing language dominance in bilinguals (which might differ depending on the type of vocabulary that is tested) these predictions are speculative.

It is hard to account for the results of Szakay et al. (2016) by assuming that the listeners were building a speaker model, as different-gendered speakers were used for the prime and the target respectively. Similarly to the finding of Hay and Dragger (2006) discussed above, it needs to recognised that not all top-down effects can fit the speaker-model explanation. As this type of priming experiment does not resemble a very natural communicative situation, it shows how the processing apparatus is able to cope with large amounts of variability and unpredictability without involving conscious decisions about the interlocutor. Szakay et al. (2016) report that overall the standard Pākehā English was processed the fastest, except when it was preceded by Māori primes. Hence it can be concluded that the bilingual listeners processed the dominant variety of their environment most optimally unless it was heard in a situation that assumed the presence of a Māori speaker. In that respect a more episodic/exemplar explanation fits the data.

The studies summarised here suggest that the surface accent of words can affect the lexical access and that listeners use the acoustic information preceding the crucial bottom-up input to form their listening predictions. The concept of shared social categories fits better with exemplar modes of perception than with more hierarchical models like BIMOLA. The latter is not concerned with different varieties of languages and their sociolinguistic status. Instead, it focuses only on the extent to which certain phonemic features overlap between two otherwise separate language systems, regardless of the sociolinguistic characteristics of the speaker. BIMOLA is not concerned with the specific mechanism of how non-linguistic information can affect lexical access. BIMOLA would predict that as far as accents are concerned, their effects on processing can be predicted on the basis of their phonetic distance between the L1 and L2 categories of the listeners, measured through a formula based on overlapping features. In that respect delays in processing would be a product of the phonetic distance of the incoming acoustic input from the idealised phonological categories that the listeners have. That is problematic in the case of Szakay et al. (2016) unless it is demonstrated that Māori shares more phonetic features with Māori English than Pākehā English.

This section has demonstrated that listeners base their processing on fast adaptations, which are related to both the environment and the phonetic characteristics of the speaker. The creation of a speaker model might be an unusual challenge for more traditional models of speech processing, such as TRACE, on which BIMOLA is based. However, it is trivial for dynamic models of processing such as IAF and TBIM. By accumulating sufficient evidence of a speaker with a particular accent, the listeners would respectively adjust the probability of hearing specific lexical items and acoustic variability from that speaker. In cases where visual input is available, TBIM allows for integration of sensory information to improve predictions (perhaps responsible for Reverse Linguistic Stereotypic effects). Hence, when seeing a child beginning to speak, a listener can automatically adjust their predictions of pitch range. Thus, even if the bottom-up evidence of the lexical item itself does not always perfectly match its most highly expected acoustic form, given the context, the correct lexical item will have sufficient pre-activation that would compensate for the bottom-up acoustic ambiguity. If no context is available, unexpected changes in accent and language lead to communication breakdown (see Kriengwatana et al., 2016; and Wagner, 2016, discussed in Section 3.2.3).

3.4. Summary

Chapter 3 focused on some key concepts relating to speech processing. In addition, it was explored whether processing an accent different from one's own would be an inherently easy or difficult task. Although there is some evidence that L1-accented speech is easier to understand for L2 listeners than standard native speech, most studies conclude the opposite. One drawback of the literature is that the proficiency of the L2 listeners was not systematically recorded.

This literature was discussed in the context of processing of accents that are similar or different to one's own. Two contradicting hypotheses have been developed for monolingual listeners: the phonetic distance and different processes hypothesis. The former states that as long as the observed accents gradually become more phonetically distant from the listener's own, then the ease of processing will gradually decline. The latter hypothesis predicts that there will be a difference between the processing of a different variety of native speech compared to foreign accented speech. In order to directly test the hypotheses, continua of dialects and accents are required as stimuli. Both hypotheses were critiqued for being vague and for assuming that a native listener has only one optimal for processing variety. A more episodic approach to speech processing would assume that listeners have multiple phonetic representations and their respective ease of processing is determined by the probability of encountering them.

Lastly, literature on adaptation to different accents was reviewed. There is evidence that before listeners can create an accent-general model of speech processing, they first form speaker models based on voice similarity, presumably dependent on a similarity of fine phonetic characteristics. The development of the accent general process may even be improved after a night of sleep. The complexity of the process of accent decoding and adaptation suggests the need to consider the sociolinguistic characteristics of the speech in relation to the listener.

This dissertation aims to explore the status of L1-accented L2 for L2 listeners by including several gradient approaches: the proficiency of the listeners, their adaptation to accents throughout the experiment and by exploring different strengths of L1 accent. As suggested in the introduction, the listeners under investigation are Bulgarian-English bilinguals living in an anglophone environment in the UK and their relationship with Bulgarian-accented English. Before outlining the main predictions based on the literature review so far, the status of English for Bulgarians, specifically Standard British English, needs to be explored. As indicated in Section 3.2.2, prestige can play an important role in speech processing and adaptation. Hence, the next chapter will focus on the issue of prestige for Bulgarian learners of English and the sociophonetic status of Bulgarian-accented English for them.

Chapter 4 Bulgarian accent

The challenge of defining 'language', 'dialect' and 'bilingualism' has been discussed in Section 2.1, reaching the conclusion that these are dynamic systems without clear boundaries that defy being restricted to a definition based on phonemic features. The same challenge is encountered when attempting to define the concept of a foreign accent that might vary in terms of intelligibility or strength. This section will address the issue of what a Bulgarian accent is, first, by exploring which English variety is the most dominant native model for Bulgarian learners of English, and second, by highlighting certain phonetic aspects of a Bulgarian accent in English in relation to the phonetics of standard Bulgarian.

4.1. UK and US influences in the teaching of English as a foreign language in Bulgaria

Variation in foreign accents can be caused by the preferred native variety of the L2 user that they choose to adopt as a model (or the one which is their teacher's model). Elliot (2018) demonstrates that the intended variety of English is an important predictor in the Slovak-English bilinguals' productions. Jensen and Thøgensen (2017) found that their two Danish-English bilinguals spoke with an American English and Estuary English dialect, and that difference affected the processing speed of other Danish-English bilingual listeners. Similarly, American English and Standard British English (SBE) are the two dominant competing standard varieties that Bulgarian learners of English might choose to adopt when starting their English education in Bulgaria.

The dominant native model when teaching English as a foreign language in Bulgaria, is particularly pertinent for the present dissertation. As stated in the introduction, the majority of Bulgarians in the UK are first generation immigrants. This means that many of them would have received some English education in Bulgaria prior to arrival, as this is the most commonly picked second language in the country nowadays (Georgieva, 2010).

Both British and American English, although admittedly these are politically loaded terms according to O'Reilly (1998), are dominant English varieties in Bulgaria and both have strong linguistic and cultural influence in Bulgaria. According to O'Reilly (1998), both varieties managed to coexist and have separate spheres of influence in the Bulgarian education system, while the Bulgarian government ensured that neither power became dominant in the first half of the 1990s.

During the Cold War it was the UK that provided English language teachers for language medium schools in Bulgaria where diplomats were trained in English (O'Reilly, 1998), though Georgieva (2010) reports that there was an overall focus on developing academic as opposed to communicative skills. In the same period, the UK also maintained connections with the Sofia

University, which is still the most prestigious higher-education institution in Bulgaria, particularly for language studies (Bulgarian Ministry of Education, 2019a; O'Reilly, 1998). After a 41-year hiatus during the Cold-war, the British council opened its doors again in 1991. However, even after the end of the Zhivkov government in 1989, Soviet-era English textbooks remained in use favouring UK culture over the US late into the 1990s. Galabova (2009) mentions that teachers from England were also employed in regular high-schools specialising in foreign languages (i.e., not just for the children of the political elite) and that they were used as consultants in the writing of textbooks. However, Galabova (2009) cites many primary sources, demonstrating the pervasive censorship and control that Bulgarian governmental officials had over their content.

During the Cold War American English had next to no influence on the English language education in Bulgaria. Black (1958), as cited by O'Reilly, claims that the US influence in Bulgaria ended in September 1942, although it appears that some US presence remained as the American College in Sofia (a prestigious private high-school) was closed in 1947. After the end of the Socialist regime, the US was quick to re-establish its presence in Bulgaria, however mostly in the sphere of math, science and technical disciplines (O'Reilly, 1998). The American University in Bulgaria was established in 1991 providing courses in Applied Economics and Computer Science. The American College in Sofia reopened its doors in 1992, admitting 100 new students. In the early years after the changes, a US fellow also joined the Graduate School of Economics in Varna (O'Reilly, 1998). Griffin (2001) mentions the prevalence of American television channels included in cable television packages, which is still the case nowadays. Anecdotally, many young Bulgarians, known to the author, claim that their first contact with English was via the undubbed American channel CartoonNetwork, which was one of the few television channels for children in the 1990s. Apart from establishing a model for second language learning, long-term exposure to English media in the native country can lead to non-trivial incidental language learning (Kuppens, 2010).

In their learning outcome requirements for high-school students, the Bulgarian Ministry of Education does not specify a native language model that needs to be followed in the teaching of phonetics (Bulgarian Ministry of Education, 2019b). Where phonetic and phonology learning outcomes are mentioned, it is only required that they approach the pronunciation of native speakers without further specifications. The learning outcomes for final year students who specialise in English in language high-schools state that students need to have in-depth understanding of history and literature of both the UK (predominantly England) and US with a balanced representation of topics from both countries. However, despite this neutrality, most English-language textbooks currently approved by the Bulgarian Ministry of Education (2019c) are either published in the UK or were written by Bulgarian authors who were likely educated during the British-centric period of English teaching in Bulgaria. As that period lasted throughout the socialist period and into the 1990s, it is expected that most Bulgarian teachers of English would have a British pronunciation model. This is also supported by the research of Elliot (2018) who reports that Socialist-era teaching practices still persist in Slovakia and British English is still considered the preferred variety among both students and teachers.

In support of this analysis, Georgieva (2010) reports a survey carried out among 120 Bulgarians living in Bulgaria or abroad regarding their language repertoire and attitudes towards English. Most participants have a positive attitude of English as a symbol of prestige and professional advancement and respond that English has found place in all spheres of their life: home, work and entertainment. When asked about their preferences for a standard variety, Georgieva interprets her results as highly indeterminate. British English is the most popular result (39), followed by a mix of British and American English (38). Only American English and only English as an International Language (EIL) were selected by 16 people respectively. Lastly, only two picked Bulgarian English. However, even they reported some native variety in addition to that response. Even though the author interprets these results as showing indeterminacy, they do not contradict the historical and political analysis of the dual presence of American and British English in Bulgaria. Overall, while the historical analysis and the survey data suggest that American pronunciation models have started to make way in Bulgaria, Standard British English is still the most common model for learning English as a foreign language in the country.

4.2. Attitudes to Bulgarian accent in English

When attempting to describe or summarise how Bulgarian-accented English is supposed to sound and what the attitudes towards it are, one faces the conclusion that there is no common stereotype of this accent. While some accents are better represented in the anglophone cultural landscape, like the Italian accent stereotypes employed in Weber et al. (2014), this is not the case for the Bulgarian accent. This section argues that the concept of Bulgarian accented speech is not very coherent among native English speakers in the UK and perhaps among Bulgarian speakers of English.

Bulgarians are more likely to be aware of Bulgarian-accented English from two main sources. Firstly, formal teaching of English is extremely widespread in the country both in the public education system for young people but also in private language schools targeted at all ages (see previous section). All hearing individuals engaged in formal learning of English from a Bulgarian instructor and/or in a classroom setting in Bulgaria will have had the experience of observing their own and other learners' speech. Secondly, all speakers of Bulgarian who have lived in Bulgaria in the last thirty years would be aware of the numerous English words which regularly enter the language through the media and advertising. The widespread use of social media in the last ten years alone has led to the introduction of a whole new specialist vocabulary, such as the verbs "лайквам", "френдвам", "събскрайбвам", "шервам" (to like, to friend, to subscribe, to share), which can be conjugated like regular Bulgarian verbs. The extent to which English words have become a part of Bulgarian can be illustrated with the fact that the suffix "-ing" (pronounced as [ink] in Bulgarian) has led to the coinage of colloquial words such as "пейкинг" [ˈpejkink] (spending a lot of time with friends on a bench, from "пейка" [pejka] – "bench").

These borrowings are readily spelled using the Bulgarian alphabet, which reinforces their being read out loud following Bulgarian phonetic processes, such as vowel reduction, final consonant devoicing etc. Although some Bulgarian scholars consider them “pollution” of the language (Shipkovenski, 2011), it is the author’s personal opinion that the use of English words in Bulgarian provides a pronunciation stepping stone for Bulgarian learners of English giving them ready phonemic equivalents (and mergers) between English and Bulgarian, as proposed in the Perceptual Assimilation Model (Best and Tyler, 2007) and thus giving confidence to learners to start using spoken English. However, as evidenced further down in this section, Bulgarian learners typically adopt a Standard North American or British model of pronunciation, which differentiates their English accents from being entirely reflective of Bulgarian phonetics.

Despite the lived experience that Bulgarians have of Bulgarian-accented English there are almost no universally known examples of that accent that can anchor it in people’s consciousness. One of the few examples would be the Sofia metro announcements in English, which were recorded by the journalist Nadya Obretenova and used since 2015 (Anonymous, n.d.). When thinking about prominent Bulgarian-accented English speakers or caricatures of Bulgarian accent in anglophone media, also very few examples come to mind. They are either known for being extremely silent and withdrawn (Viktor Krum from the *Harry Potter* film series, portrayed by Stanislav Yanevski) or are minor characters serving as a plot device (Veronika from *T2 Trainspotting*, played by Anjela Nedyalkova). Lexi Viktorova from the BBC radio 4 drama *Archers* is also a minor character and is portrayed by a London-born actor of Polish origin, Ania Sowinski. Anecdotally, most British and Bulgarian UK residents known to the author are aware of Bulgarian accents only through personal connections as opposed to the media, which would be expected to lead to heterogenous representations of Bulgarian accent across the population.

The primary reason for that is the size of the Bulgarian population – about 7 million, according to the latest census (National Statistical Institute, 2010). Compared to the world-wide population of L1 speakers of other languages such as French (600 mln) (Central Intelligence Agency, 2017; Nations Online, 2020) or Spanish (477 mln) (Instituto Cervantes, 2018), the population of Bulgarian speakers is only a fraction of their size. This means that most native English speakers are statistically less likely to come across Bulgarian speakers and for Bulgarians to participate in the cultural life of the UK and other anglophone places to compared to the speakers of the other European languages cited here. There are no well-known Bulgarian public figures with English as a second language who can serve as a representation and anchoring of the concept of Bulgarian-accented English like Marion Cottillard, Gérard Depardieu, Arnold Schwarzenegger, Salma Hayek, Antonio Banderas or Sophia Loren do for French, German, Spanish or Italian accents, respectively.

Another source of evidence for the low prominence of the concept of a Bulgarian accent can be found by searching the English language corpus linked to the Google NGram viewer. This tool has been suggested by Brysbaert et al. (2011) and Friginal et al. (2014) as a useful source of information for psycholinguistic research as it is the largest collection of books available online (Phillpott, 2019). One drawback of this tool is that it includes only books with 40 or more occurrences of the search terms and that the books available before the 20th century are scarce (Phillpott, 2019).

Two NGram searches for the period 1800-2008 with a smoothing factor of 3 are presented in Figure 3 and Figure 4. The string 'Bulgarian accent' has no instances in the British English corpus, therefore the following results are based on the general English corpus. The first search compares the prominence of the string 'Bulgarian accent' to supra categories of accents that it belongs to: 'Eastern European accent', 'Slavic accent', and 'Balkan accent'. The second compares 'Bulgarian accent' to popularly stereotyped English accents. While it is possible that the word "accent" is not used in the sense of pronunciation in cases in the corpus, the data still give an impression of the prominence of the Bulgarian identity compared to others.

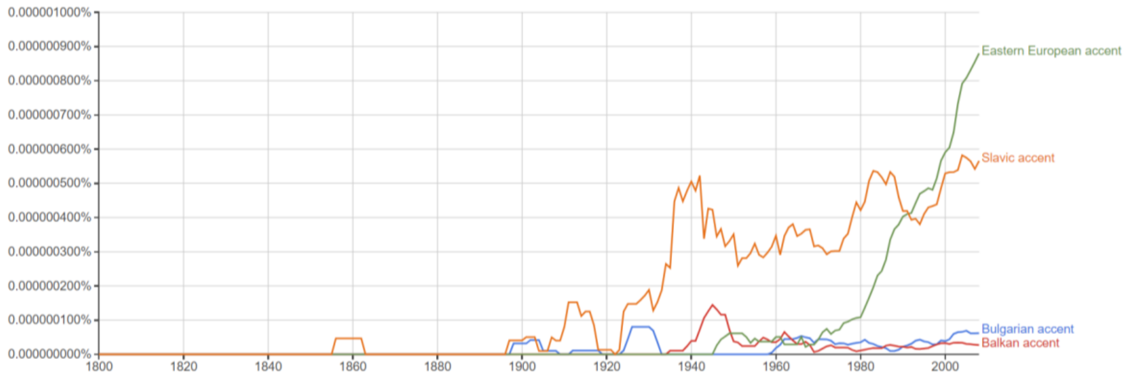


Figure 3. Google NGram view of the normalised percentage of occurrences of the phrases 'Bulgarian accent', 'Eastern European accent', 'Slavic accent', and 'Balkan accent' in published written texts between 1800-2008.

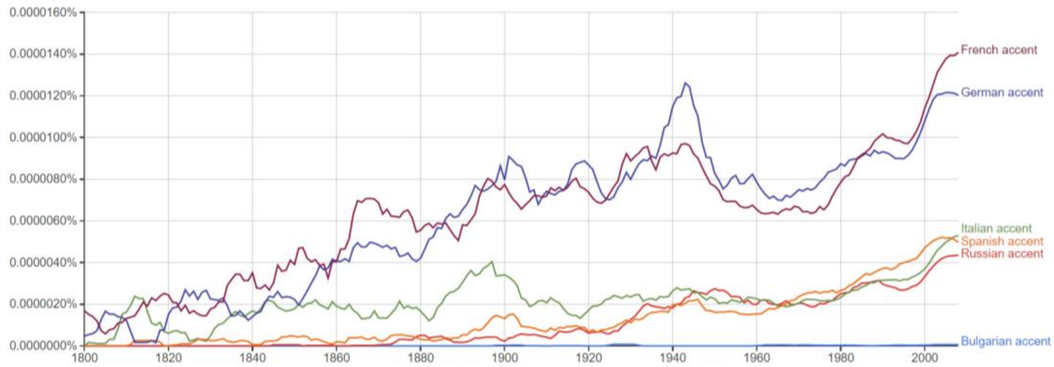


Figure 4. Google NGram view of occurrences of the normalised percentage of occurrences of the phrases 'Bulgarian accent', 'French accent', 'German accent', and 'Italian accent', 'Spanish accent', 'Russian accent' in published written texts between 1800-2008.

The phrase begins to be used consistently in the start of the 20th century only a few years after Bulgaria gains complete independence from the Ottoman empire. Even so throughout the 20th century and the beginning of 21st century the concept of a Bulgarian accent has appeared extremely rarely in English-language texts compared to other well-known accents. The political connotations of different accents are apparent in Figure 3. Slavic accents started steadily increasing in prominence in written texts after the creation of USSR in December 1922. On the other hand, the concept of the Eastern European accent has seen a sharp rise in prominence in 1980. That is likely due to the rise of the Solidarity movement in Poland at the end of 1980, which drew attention to the political crisis of the Eastern Bloc, and the death of Brezhnev in 1982, which set in motion dramatic changes within the USSR and the Eastern Bloc.

However, while Bulgaria may also be considered an Eastern European country, it is uncertain to what extent Bulgarian-accented English shares features with, for example, Romanian-accented, Russian-accented English or Polish-accented English, which, anecdotally, tend to be more commonly associated with Eastern-European accents. Bulgarian is a South Slavic language and does not have nasal vowels like Polish, which is a west Slavic language, and it does not have palatal consonants as independent phonemes like Russian does, an east Slavic language. Romanian belongs to an entirely different language family from Bulgarian. Some sources of phonetic commonality between the Eastern European accents might be the overall smaller vowel inventories than English, they tend to be voicing languages and potentially have similar articulatory setting (Honikman, 1964; Mennen et al., 2010; Schaeffler et al., 2008). In terms of attitudes, if the Bulgarian accent is perceived as Eastern European then that might evoke political associations in non-Bulgarian listeners associated with the Cold war and the recent increase of Eastern European residents in the UK, following the expansion of the EU in 2004 and 2007.

Given the low prominence of the concept of Bulgarian accent, it is not entirely surprising that very few studies have included Bulgarian-accented English in attitudes surveys. One of these studies investigated the attitudes towards different English accents among native English speakers (in this case of General American English) (Said, 2006). It found that Eastern European accents (Bulgarian-accented English clustered with Romanian-accented English) received the most positive ratings compared to the other accent groups (Arab, Latino and South East Asian). The Bulgarian speaker received an average score of 33.3 (of minimum 10 and maximum 50). The author hypothesised that this might be due to the Eastern European speakers belonging to the same language family as the raters, which would make their speech easier to process. This analysis would suggest that for the Michigan-based raters, Latino accents (also Indo-European in origin) were more stigmatised than Eastern-European accents. As this study dates to about thirteen years ago and it concerns US, not UK, listeners, it is questionable to what extent its conclusions are pertinent for the present investigation. Nevertheless, the low stigma associated with Eastern European accents (and the Bulgarian accent in particular) among native listeners of English could be also a product of the low saliency of its stereotype, as discussed earlier. The expected result of low stigma would be that Bulgarian speakers of English experience less discrimination compared to L2 speakers of other language backgrounds and would have a more positive attitude towards their own pronunciation.

This expectation is partially supported by Georgieva (2010) who investigated how L1 Bulgarian – L2 English speakers view their own speech. Her investigation demonstrates a complex dynamic between the speakers' attitudes to a model variety of English and a variant that they are comfortable adopting for their own speech. One such example is the following: "*Deep in my heart I am all for British English, although I am perfectly aware that what I speak is International English*" (Georgieva, 2010, p. 175). Georgieva (2010) posits that the purpose of British English in this context is a 'model for emulation', which is not for everyday use. It is a "'frame of reference' against which individuals' creativity, ingenuity, originality, or beauty of expression is measured." (Georgieva, 2010, p. 175). According to her, this abstract role allows the speakers' other languages to fulfil different roles and thus coexist with the emulation model.

Another finding that emerged from Georgieva (2010)'s data is that the Bulgarian bilinguals valued the content of their speech more than the form, accompanied by contradicting attitudes towards the Native Speaker model of correctness. Only 10% of the participants reported preferring to abide by grammar books and dictionaries, which Georgieva (2010) interprets as abiding by the Native Speaker model. However, without access to the rest of the questions in her survey, this result can instead be interpreted as unwillingness to abide by the rules of a standard variety in favour of a different potentially regional Native Speaker model. In the qualitative data, there were three predominant groups of respondents: those who supported a native speaker model, those who rejected it, and those who valued adapting their speech according to the situation. Georgieva (2010, p. 176) exemplified these groups with the following three quotes:

1. I do not think that I have to respect the native speakers because it is their language, but I have to respect the language rules. After all, I am not the one who has the right to choose.

2. I do not approve of bending the language to suit one's own personality; the preference for speaking some kind of an 'adjusted' and 'individualistic' English rather than the native speaker's version, in my opinion, indicates acute self-consciousness and in almost all cases is a sign of insecurity and low self-esteem or fear from losing one's identity (which in its turn again points towards insecurity. (....) Aiming to reach a native speaker level of proficiency does not threaten in any way one's identity; this is merely a matter of perfectionism.

3. For me, when I speak English it is very much like when I speak Bulgarian. It depends very much on who I speak to. I mean the person I speak to, no matter native or non-native, influences my performance so much that in every instance I may sound an entirely different person altogether.

Even though Georgieva (2010) uses these three quotes to represent three different views, they can be interpreted as two general positions: one that assumes that a language is an entity that remains independent from and uninfluenced by the users (except in detrimental ways), and another group that assumes that a language is a flexible entity that has to be adjusted according to the circumstances. It remains an open question whether speakers who represent the former attitude will display lower levels of phonetic accommodation and have a more consistent and stable realisations than the latter group and how that might reflect on their ability to process speech. In cases like the latter group it might be difficult to talk about Bulgarian accent as a single entity.

A similar complex picture of different speech models has also been reported for Slovak residents in Scotland. Elliot (2018) reports that L1 Slovak – L2 English bilinguals who reside in Scotland have a similar attitude to Standard British English, as their Scottish peers. Both groups considered Standard British English as the most prestigious but least attractive variety. Although non-native speech was rated as less prestigious than native speech, the bilinguals rated the speech of an integrated Slovak immigrant as less attractive than the speech of a non-integrated one. Despite these preferences and after many years of residence in Scotland, most of her informants surprisingly stated that Standard British English was their model accent of choice. Many of the Slovak participants of Elliot-Slosarova identified as Slovak, European or Scottish but still chose the less socially attractive to them Standard British English as an emulation model. This echoes the results for the Bulgarian participants of Georgieva (2010) for the status of Standard British English for English-speaking Bulgarians. The analysis supports the idea that this variety has a special status of prestige for expats in the UK even if it does not serve as a direct model for their speech. Hence there is motivation to expect that that Standard British English will serve as a perception model and will lead to the "prestige benefit" discussed in Section 3.2.2.

4.3. Bulgarian phonetics

As reviewed in the earlier sections of this chapter, foreign accents, especially in the early stages of language learning are usually characterised by phonetic transfers from the native language (Best and Tyler, 2007; Flege, 1995). It has been established that both British English and American English can serve as target varieties for Bulgarian L2 English speakers. Hence, it is important to find out what are the typically Bulgarian phonetic characteristics that might determine the nature of the Bulgarian accent. The remaining part of this section will provide a brief overview of Bulgarian phonetics and phonology, occasionally drawing attention to differences with English. Following that, several case studies of Bulgarian-accented English will be discussed.

As mentioned earlier, Bulgarian is a South Slavic language. It tends to be analysed as having 22 consonants and 6 vowels, listed in Table 1 below.

	Bilabial	Labio-dental	Dental/ Pre-alveolar	Alveolar	Palatal	Velar
Plosive	p b		t d			k g
Nasal	m		n			
Trill			r			
Fricative		f v	s z	ʃ ʒ		x
Affricate			ts dz	tʃ dʒ		
Approximant					j	
Lateral			l			
Vowels: i, ɛ, a, ə, ɔ, u						

Table 1. Bulgarian phonemes.

Although some of the same IPA symbols are used in English, that does not do justice to some of the fine phonetic differences that may occur between the two languages. The next section highlights the aspects of Bulgarian phonetics that might differ from English, although the same symbols are used to illustrate the sounds.

The voiceless stop consonants are non-aspirated and the voiced stop consonants are produced with prevoicing in initial position and full closure voicing in medial position (Dokovova, 2015; Rangelov, 2008). Regarding the liquids, /r/ is normally produced as a tap or trill, although some speakers never acquire it and prefer using a uvular trill or fricative instead, more rarely /l/. Clear /l/ exists in Bulgarian, and it is almost always realised in initial position when followed by front vowels, however, there might be a phonological change underway, such that lateral gliding is observed with a velarised or dark /ɫ/ (or /w/) in coda position or preceding /a, ə, ɔ, u/ (Radkova, 2009). Anecdotally, when asked about what the most obvious characteristic feature of the Bulgarian accent in English is, many Bulgarians report that it is the realisation of /l/.

According to the traditional analyses of Bulgarian (Tilkov and Boyadziev, 1981), in addition to the consonants listed above, there is also a series of palatalised consonants. However, Ignatova-Tsoneva and Baeva (2009) and Choi (1998) argue that in Contemporary Standard Bulgarian 'palatalised' consonants are best interpreted as 'plain' undergoing palatalisation when followed by the only truly palatal consonant /j/ in restricted phonemic contexts, instead of the traditional analysis with full sets of 'hard' and 'soft' consonants, which is only typical of East Slavic languages (Russian, Belarusian and Ukrainian). While this is not explicitly stated in any analyses of Bulgarian, it is likely that research produced during the Socialist regime was influenced by the need to point out similarities between Bulgarian and Russian, the latter being the politically influential language at the time. While political neutrality is assumed in most modern-day research in phonetics, it is important to consider that this has not always been the case in Bulgaria when citing resources of that era.

Regarding the Bulgarian vowel inventory, a distinction needs to be drawn between stressed and unstressed vowels. In stressed position there are six vowels, while in some unstressed positions some of the vowel distinctions are neutralised. There are contradictory analyses regarding the exact nature of unstressed vowel neutralisation and particularly relating to vowel height. According to Boyadziev and Tilkov (1997; 1981) and Ternes and Vladimirova-Buhtz (1990), unstressed vowels are neutralised into an intermediate centralised position, involving raising for the lower vowels and lowering for the higher vowels. The unstressed central vowels /a/ and /ə/ are completely neutralised in all dialects, /ɔ/ and /u/ are less so, while /ɛ/ and /i/ only neutralise in Eastern dialects. This account has been supported by Zhobov (2004) although none of these authors report any experiments from which they have obtained the formant data that they cite and is possible that some are based on perceptual judgement. For instance, Ternes and Vladimirova-Buhtz (1990) transcribe unstressed /a/ and /ə/ as [ɐ] and /ɔ/ and /u/ as [o].

In the first and so far only corpus-based acoustic study of casual Bulgarian speech, drawing from a more representative sample, Andreeva et al. (2013) find no evidence of higher vowels being lowered in unstressed position. On the contrary, they find that the lower unstressed vowels are consistently raised, while the higher unstressed vowels are either higher than or equal to their stressed counterpart (as inferred from F1) (also demonstrated in Dokovova et al., 2019). Sabev (2015)'s recent study confirms these findings for the back vowels but finds evidence of lowering for unstressed /ə/. He reports that unstressed /ɔ/, /u/ and stressed /u/ overlap in F1. In addition, Sabev (2015)'s perceptual experiment demonstrates that native Bulgarian listeners cannot reliably discriminate between unstressed central and back vowels out of context, although they do discriminate between the front unstressed vowels.

Regarding the suprasegmental aspects of Bulgarian single-word phonetics, there is only one stressed syllable per word, which is realised in terms of increased intensity and vowel duration, compared to unstressed vowels and in terms of specific vowel quality, as discussed above. In that way Bulgarian differs from English, which can distinguish between three levels of stress. According to Dimitrova (1998), Bulgarian takes an intermediate position between the syllable-timed and stress-timed categories proposed by Pike (1945), unlike English, which is widely reported to be stress-timed (Dauer, 1987). It is likely that there are differences between stressed and unstressed consonants in Bulgarian, judging on what has been reported cross-linguistically by Lisker and Abramson (1967), Klatt (1975), and Cho and McQueen (2005), but no studies on Bulgarian were found.

Overall, Bulgarian phonetics differs from standard British English varieties in terms of both its segmental inventory and realisations and its stress realisation patterns. The next section will investigate several case studies of Bulgarian accent in English.

4.4. Exploration of Bulgarian-accented English

This section will present a descriptive review of some characteristics of Bulgarian accent in English. The data for the following study are taken from the Speech Accent Archive (Weinberger, 2015). The database contains recordings of twenty L2 English speakers (nine female) whose native language is Bulgarian. Their mean age is 28.7 (*SD* = 7.8). Only two of the twenty speakers have an identifiably British model of pronunciation, probably due to the fact that most speakers are residents of the USA. On average their age of onset of English is 14.2 years (*SD* = 8.4) but as the number is skewed due to two outliers, the median is 10 years. Two of the speakers are not residents of an anglophone country. For the rest, the mean time of residence is 8.45 (*SD* = 5.4).

The speakers are recorded reading a short paragraph, called "Please, call Stella", which contains instances of all English phonemes in different environments. The website provides phonetic transcription and a phonetic analysis of the Bulgarian accent for several of the speakers (see Table 2). The phonetic characteristics of the remaining speakers were analysed auditorily by the author.

Analysis	Speaker number
Phonetic analysis provided on the website	Speakers 1-13
Summary of atypical phonological features (e.g., vowel shortening) provided on the website	Speakers 1-8
Summary of atypical phonological features (e.g., vowel shortening) made by the author based on the phonetic analysis provided on the website	Speakers 9-13
Summary of atypical phonological features (e.g., vowel shortening) made by the author based auditory judgement	Speakers 14-20

Table 2. Types of analysis used for the Bulgarian speakers from the Speech Archive.

The following phonetic generalisations emerged from the analysis. In consonants there was final obstruent devoicing (e.g., [bop] for "Bob", or [bɹɪŋk] for "bring"), interdental fricative turned into a stop/tap/labiodental (e.g., [di] for "the" or [tik] for "thick"), consonant voicing (e.g., [znæ̃k] for "snack"), /h/ to [x] (e.g., [xɝ] for "her"), no aspiration (e.g., [pi:s] for "peas"), r to trill (e.g., [frʌ̃m] for "from"), palatalisation (e.g., [xʲɜ] for "her" or [meɪbʲɪ] for "maybe"), [lˠ] or [ɫ] instead of [l] (e.g, [stɛlˠʌ] for "Stella"). For vowels there was shortening (e.g., [plis] for "please"), elongation (e.g., [fɹɔ:k] for "frog"), raising (e.g., [snek] for "snack"), lowering (e.g., [braðəɹ] for "brother"), and backing (e.g., a particularly retracted vowel [blu̠] for "blue").

This analysis recorded what percentage of the twenty speakers exhibited a certain feature at least once. The results of this analysis are summarised in Figure 5 below.

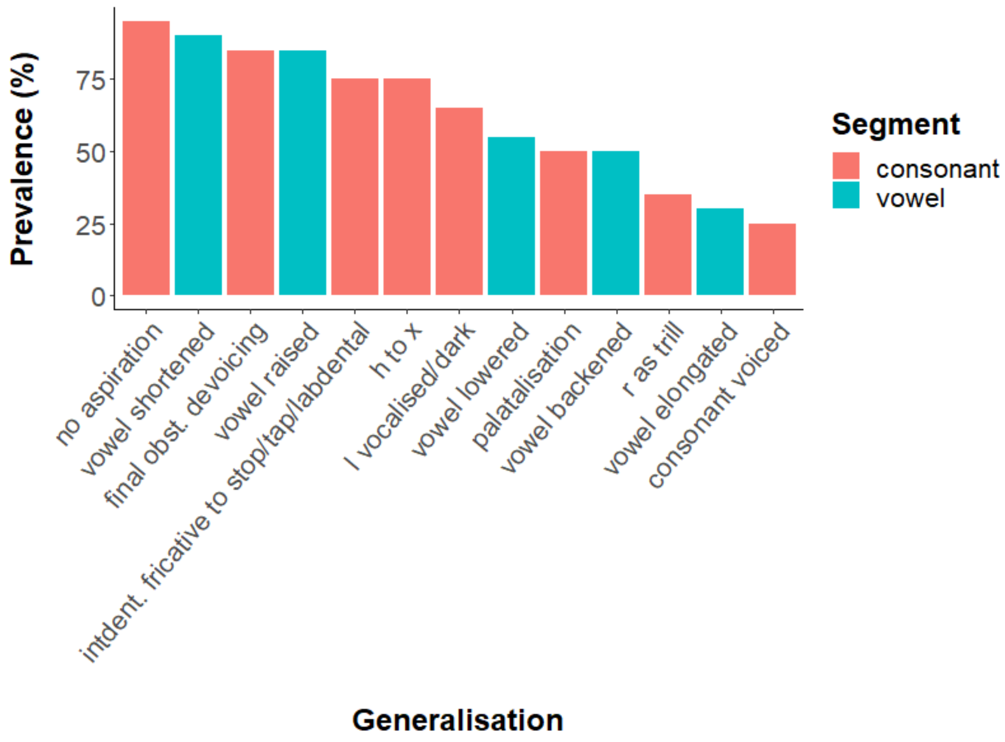


Figure 5. Prevalence of Bulgarian accent characteristics across twenty Bulgarian-English bilinguals.

The most widespread sources of deviation from the native General American or British varieties are a lack of aspiration in initial stops, shortening of vowels, final obstruent devoicing (typical of Bulgarian), vowel raising (e.g., /æ/ to [ɛ]), pronouncing interdental fricatives as a stop, tap or very rarely as a labiodental fricative, substituting /h/ with [x], often a more palatalised version [xʲ], realising a dark or vocalised /l/ in positions that require a clearer variant (e.g. in the word “Stella”). Some of the slightly less prevalent characteristics were vowel lowering (e.g., /o/ to [ɔ]), palatalisation (after some stops or /r/ followed by a vowel), vowel backing (e.g., /a/ or /æ/ to [ɑ]), realising /r/ as a trill, inappropriate elongation of a vowel or voicing consonants that should be unvoiced (primarily in word-medial position).

Impressionistically, it appears that the Speech Learning Model by Flege (1995) has some support from this data, although this analysis might be an example of confirmation bias. Nevertheless, it is worth observing that sounds that might have an obvious equivalent in terms of articulatory realisation and abstract category between the speakers’ L1 and L2 (stops, /h/ and /l/) are involved in more wide-spread substitutions compared to /r/, which requires a completely novel articulation in English. In that respect interdental fricatives are an odd case as they do not have a direct abstract equivalent in Bulgarian. However, it appears that the articulatory similarity between them and the dental stops (typical of Bulgarian), taps or labiodental fricatives has led the speakers to establish an equivalence between these categories. Flege's phone-equivalences are based on the L2-users ability to draw perceptual distinction between the sounds. However, without access to this information, these data suggest that similarity between phone-categories might also depend on articulatory similarity.

Chapter 5 Overview of experiments

As indicated throughout the chapters so far, the main aim of this dissertation is to explore the phonetic representations of Bulgarian-accented English for Bulgarian-English bilingual listeners. In this dissertation Bulgarian accent (BA) in English is defined as English speech produced by speakers of Bulgarian origin who started learning English during puberty and whose target in English is the Standard British English (SBE) variety. The prestige status of this variety is relevant because it allows the assumption that residents across the United Kingdom and Northern Ireland are familiar with it and are likely to have an advantage when processing it over any other local variety (Smith et al., 2014). In addition, as explored in the previous section, it is also of high-status for Bulgarian learners of English, suggesting that all potential participants will have at least some experience with it.

It is recognised that the realisations of Bulgarian-accented English can vary with respect to the strength of Bulgarian and Standard British English identity they convey. In order to explore the status of these phonetic representations for Bulgarian-English bilinguals, their reaction times and accuracy are measured when processing different strengths of Bulgarian-accented English speech, compared to the native SBE productions.

To this end the following two experiments are undertaken. The first experiment aims to find out if there is any link between the phonetic representations of Bulgarian-accented English and Bulgarian for Bulgarian-English bilinguals and the effect of English proficiency. The second experiment investigates the effect of proficiency on the processing of Bulgarian-accented and Standard British English, without the use of primes. A follow-up of the second experiment compares the Bulgarian-English bilinguals with the highest proficiency in English to native English listeners in their processing of Bulgarian-accented English.

The first experiment uses a lexical decision task with auditory priming. Bulgarian-accented primes and Standard British English primes preceded Bulgarian words or non-words (and vice versa). The listeners' reaction times of lexical decisions were measured, where relatively shorter reaction times between a type of prime-target pair signals that the prime facilitates the recognition of the target. The main hypothesis is that Bulgarian-accented English will facilitate the processing of Bulgarian words and that Bulgarian primes will facilitate the processing of Bulgarian-accented English words.

The second experiment uses Bulgarian-accented and Standard British English words, produced by Bulgarian-English bilinguals and native English speakers, respectively. The participants have to recognise them in a lexical decision task and their accuracy and reaction times are measured for analysis. It is predicted that the listeners with the lowest English proficiency in the sample will process Bulgarian-accented words more accurately and quickly than Standard British English words. In comparison, the listeners with the highest proficiency in English in the sample are expected to process the Bulgarian-accented words at least as efficiently as the lower proficiency listeners and they are expected to process Standard British English faster than the lower proficiency listeners.

In order to check if the bilinguals process Bulgarian accent faster and more accurately compared to native English listeners with no knowledge of Bulgarian (interspeech intelligibility benefit for listeners), the follow-up experiment presents the same task from Experiment 2 to native English listeners and then compares their performance to the listeners with highest proficiency in Experiment 2. The hypothesis is that the bilinguals will process Bulgarian-accented English faster and more accurately that native English listeners.

Overall, these experiments aim to add more depth to the Interspeech Intelligibility Benefit Hypothesis by treating proficiency as a continuous variable and by investigating reaction times and gradual within-task adaptation, in addition to overall accuracy of processing. The insights gained from this approach allow the results to illuminate multiple theories, discussed in the previous chapters. All experiments were conducted in accordance with the QMU ethics regulations and received separate ethical approvals from Prof. Janet Beck, head of the Division of Speech and Hearing Sciences.

Chapter 6 Experiment 1: Cross-linguistic auditory priming

6.1. Introduction

This study investigates whether there is a representational link between Bulgarian-accented English words and their Bulgarian translation and vice versa. This is tested using auditory priming in an auditory lexical decision task. It is expected that after hearing Bulgarian primes (e.g., "връзка" [vrəzkə], meaning "link") the listeners will have shorter reaction times to target words pronounced with Bulgarian-accented English (e.g., [liŋk] or [link]) compared to English words pronounced with a Standard British English (SBE) accent (e.g., [lɪŋk]). In addition, it is expected that if the primes are in Bulgarian-accented English (e.g., [link]) as opposed to SBE (e.g., [lɪŋk]) that will facilitate the recognition of the Bulgarian target word (e.g., [vrəzkə]).

One of the only studies using a similar design in a cross-linguistic auditory task is by Szakay et al. (2016) (for a complete discussion, see Section 3.3.2). They observed that Māori words prime Māori-English translations more effectively than translations pronounced with the Standard New Zealand English accent (see Figure 6). This priming effect was only observed in the L2-to-L1 priming condition, which is usually harder to replicate (cf., discussion in Section 2.3). The effect was also only observed for the listeners who were most immersed in Māori language. The authors speculated that the sociolinguistic link between the sociolinguistic relationship between Māori and Māori-English was the reason for this facilitation effect.

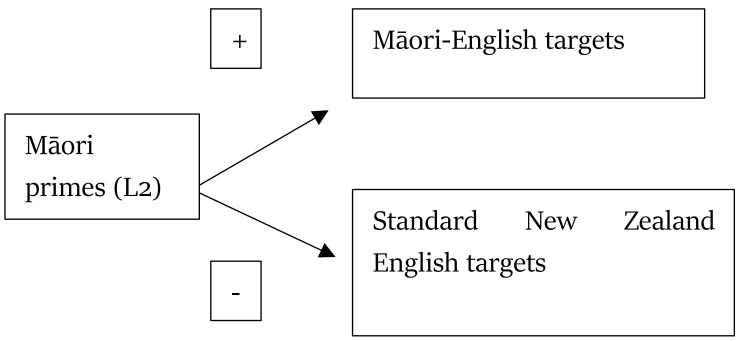


Figure 6. Pattern of sociophonetic priming in Szakay et al. (2016), where "+" indicates a prime that facilitates the recognition of the target and "-" indicates a prime which does not facilitate the recognition of the target when compared to a baseline.

As sociolinguistic indexes and their phonetic realisations are inherently linked (cf., Section 3.3), it is hypothesised here that a similar connection should be observed in a cross-linguistic task with L1 words and L1-accented L2 words. In the case of Bulgarian and Bulgarian-accented English it is possible that the effect is less strong, as Bulgarian-accented English is not an independent and socially validated variety of English like Māori-English and it is not the target realisation for the majority of Bulgarian L2 speakers of English (cf., Section 4.2). Nevertheless, as discussed in Sections 2.3.1 and 3.3.2, sometimes even very subtle phonetic cues can affect the listeners' speech processing and change their language mode.

In addition, in Section 2.3.3, a case is made that when listeners are processing an accent that they are willing to imitate, or produce themselves, they would be more likely to process it efficiently, than if they are only passively engaged with it. As lower proficiency speakers are generally more likely to speak with a Bulgarian accent than higher proficiency speakers (cf., Section 2.2.1) it is also expected that the lower proficiency listeners will process Bulgarian-accented speech faster than higher proficiency listeners (similar to what was observed with the highly immersed L2 Māori listeners in Szakay et al., 2016). As the phenomenon of cross-linguistic phonetic priming is understudied, this experiment will explore it in both directions: L1-to-L2 and L2-to-L1.

6.2. Methods

6.2.1. Overview

There were 20 Bulgarian (L1) – English (L2) bilinguals who participated in this lab-based study. The main experimental task was an auditory lexical decision task. It consists of 4 blocks in which a Bulgarian word serves as an auditory prime for a Bulgarian-accented English (BA) target or for a Standard British English target, henceforth SBE. For example, the Bulgarian word "връзка" [vrəzkə], meaning link, is used to phonetically prime the BA target [link] or the SBE target [lɪŋk]. In addition, there are 4 blocks, in which Bulgarian-accented English words or SBE words serve as phonetic primes for Bulgarian target words (e.g., [link] or [lɪŋk] priming [vrəzkə]). Lastly, there are 3 monolingual blocks, in which Bulgarian primes Bulgarian (e.g., [vrəzkə] priming [vrəzkə]), SBE primes SBE (e.g., [lɪŋk] priming [lɪŋk]) and BA English primes BA English (e.g., [link] priming [link]). The listeners have to decide if the target is a real word or not. The listeners' speed of recognition is measured via their key-press reaction times.

The participants' English proficiency was measured via the LexTale test, which is a visual lexical decision task (Lemhöfer and Broersma, 2012a). Lastly, they completed a demographic and linguistic background questionnaire.

Initially the experiment aimed to compare the size of the semantic priming effect between the BA and SBE forward and backward priming conditions. A semantic priming effect is the RT from semantically matching prime and target minus RT from semantically unmatching prime and target. However, during the implementation an insufficient number of semantically unrelated prime-target pairs in each block was included which undermines the validity of the semantic priming effect (i.e., it cannot be distinguished from a surprise effect of having fewer tokens of semantically non-matching pairs). Hence the conditions are compared overall, looking only at the effect of having a specific accent present in the prime-target pair without interpreting any semantic priming effects. The only priming effects that can be interpreted from this analysis, are phonetic (not semantic), between Bulgarian phonetics and Bulgarian-accented English phonetics.

6.2.2. Design

The auditory lexical decision task was built and presented on OpenSesame (Mathot et al., 2012). The participants were asked to decide whether the second word in a series of auditory word-pairs they heard was a real word or not. The participants' reaction times were measured and used for analysis. The lexical decision task included 14 training trials; 8 cross-linguistic blocks, where the prime was in Bulgarian and the target was in English or vice versa; and three monolingual trials, where both prime and target were in the same language and accent: Bulgarian, Standard British English or Bulgarian accented (BA) English (see Table 3). Overall, there were 88 experimental word pairs (8 per block), of which half were distractors. Table 3 also reflects that initially the English stimuli (BA and SBE) were split into more subgroups depending on whether they underwent acoustic manipulation or not. This subdivision was not used in the final analysis due to the low sample size, but more information about the procedure is available in Section 6.2.4.

Within each block of trials there were three types of relationships between the prime and target (see Table 4). First, they could be semantically related. In the cross-linguistic blocks the target was the translation equivalent of the prime (e.g., "крило" [kri'lɔ] - "wing"). In the monolingual blocks the prime and target were identical (e.g., "finish" - "finish"). There were two test and two filler pairs in each block that were semantically related. Second, the prime and target could be semantically unrelated (e.g., "skinny" - "picture"). In that case there was no connection in meaning between the prime and target. There was one test and one filler pair in each block that were semantically unrelated. Lastly, the target could be a non-word, even though the prime is a real word (e.g., [resisk]). In these cases, the non-word was derived from what the appropriate semantically related target word should have been and one or two non-initial phonemes in it were changed in order to render it a non-word in both Bulgarian and English. There was one test and one filler pair with a non-word target in each block.

Types of Blocks		
Block number	PRIME language	TARGET language
1	BG	BA + manipulation
2	BG	BA – manipulation
3	BG	RP + manipulation
4	BG	RP - manipulation
5	BA + manipulation	BG
6	BA - manipulation	BG
7	RP + manipulation	BG
8	RP - manipulation	BG
9	BA - manipulation	BA - manipulation
10	RP - manipulation	RP - manipulation
11	BG	BG

Table 3. All the possible blocks depending on the languages of the prime and target. The number of tokens is presented in parentheses.

Within a Block		
Number of tokens	PRIME	TARGET
2	word	semantically matching word (related)
1	word	semantically non-matching word (unrelated)
1	word	non-word

Table 4. Relationships between pairs of primes and targets within a block.

6.2.3. Participants

The study required the participation of bilingual adults who, according to self-reports, started actively learning their second language (L2) English, after having acquired their first language (L1) Bulgarian. There were 20 Bulgarian L2 speakers of English living in Scotland that took part in this study. In this case, the term 'learner' is used broadly even for people who acquire the language by using it outside of a formal learning environment.

The participants had lived in the UK for 6.3 years on average (SD = 1.58) and at the time of the experiment all were residents of Edinburgh. Their mean age was 25.9 (SD = 2.91). All of them had relatively high English proficiency with mean LexTale score 88.25% (SD = 11.13). All participants had completed their secondary education in Bulgaria and were raised in Bulgarian-speaking families. Bulgarian was the primary language used by them when growing up. All had completed at least one university degree and were actively using both English and Bulgarian in their everyday lives.

6.2.4. Materials

The stimuli were chosen, recorded and validated, as described further on in this section. As indicated in Section 6.2.2, some of the stimuli underwent acoustic manipulation. Its validity was tested quantitatively, as reported in Section 6.2.4.4. The validity of the manipulation was also tested using native listener judgements, described in Sections 6.2.4.5 to 6.2.4.7.

The auditory priming experiment discussed here had a total of 65 Bulgarian words used as primes (n = 18), targets (n = 13), both as a prime and a target in the monolingual block (n = 2), or distractors (n = 32) and 10 non-words derived from Bulgarian words. There was a total of 76 English words, of which 20 were used for primes (n = 10 with Bulgarian accent (BA) and n= 10 with SBE), 14 words were used as targets (n= 7 in BA and n= 7 in SBE) and 4 were both the prime and target in the monolingual block (n = 2 in BA and n = 2 in SBE English). Lastly, there were 38 distractors and 12 non-words derived from English words. All the stimuli are listed as they were presented to the participants in Appendix 1.B.

6.2.4.1. Stimuli selection

The English stimuli were selected from the British National Corpus (BNC) web (Hoffman and Evert, 2013) with minimal spoken frequency log 1.71 and maximal frequency 4.61. The Bulgarian translations of the words had comparable frequencies, between log 0.6 and 3.54, drawn from the frequency dictionaries based on the Bulgarian National Corpus (Koeva and Kolkovska, 2011) of the style "informal/fiction". This style was chosen as the closest equivalent to the spoken BNC. The former contains 5 million words and the latter 10 million.

The English words had the vowel /ɪ/ (as in KIT) in the syllable carrying the major stress. They were not cognates with their Bulgarian translation. The Bulgarian translation did not have /i/ (the most similar vowel in the Bulgarian inventory to /ɪ/) in its stressed syllable, and it did not start with the same consonant cluster as the English word. These precautions were taken in order to ensure that the cross-linguistic priming is due to the phonetic, not phonological characteristics of the words. In addition, the words were chosen, so that a substitution of the stressed vowel with /i/ would not result in a real word (e.g., /pɪŋk/-[piŋk]). Due to a difficulty of finding a sufficient number of words that satisfy all of the listed criteria, information on their neighborhood size was not controlled at the time. Future research, however, should take that information into account as neighbourhood size and spread may affect the speed of lexical access at least in monolinguals (Luce and Pisoni, 1998; Vitevitch, 2007). An additional 38 English distractor words were selected from the BNC corpus. The only criteria for them were to not have the vowel /ɪ/ in their stressed syllable and to have a single-word translation in Bulgarian.

6.2.4.2. Stimuli recording

The speaker was a 24-year old female native speaker of Bulgarian, a proficient user of English and a trained phonetician. She started learning Standard Southern British English as a target variety at 11 years of age in a formal instruction setting and at the time of the recording, she had lived in the UK for about 5 years.

Specific guidelines were followed during the recording to standardise the sound of the Bulgarian accent. All the /r/ sounds were pronounced as taps or trills. Consonants that would normally have alveolar or postalveolar place of articulation in English /n, s, z, ʃ, ʒ, t, d/ were produced with dental place of articulation. Voiceless stops /p, t, k/ were pronounced with less aspiration and voiced stops /b, d, g/ had prevoicing. Voiced final consonants were devoiced. Since Bulgarian has a smaller vowel inventory than English, the distinction between some groups of vowels was purposefully obscured. The following equivalents were aimed for in stressed positions: /ɑ, æ, a/ were pronounced more like [a]; /ɔ ɒ o/ were pronounced more like [ɔ] in stressed syllables and [o] in unstressed; /ɪ, i/ were pronounced more like [i]; /ɜ, ʌ/ were pronounced like [ə] (usually represented with the Bulgarian letter 'ъ'); /ɛ, e/ were pronounced more like [ɛ]; /ʊ, u/ were pronounced more like [u] and with a more back place of articulation than would be typical of English.

All the words were recorded in a sound-proof studio at Queen Margaret University. A Neumann U89i microphone with a sampling rate 80 kHz was connected to a Digidesign Digi 003 Mixer and Pro Tools console, from which it was digitally recorded on Pro Tools 8 on a Mac desktop. There was about 10 cm distance between the speaker's face and the microphone. All words were recorded first in non-rhotic Standard Southern British English accent. Then their translations were recorded in Bulgarian to ease the transition into Bulgarian-accented English. Lastly, to ensure that the words recorded with a Bulgarian accent sounded natural, the list was recorded twice in a row and most of the words used for the experiment came from the second reading.

6.2.4.3. Acoustic manipulation

One of the initial goals of the experiment was to create a four-step continuum of English stimuli that have increasing phonetic similarity to a Bulgarian accent. Due to a small sample size of participants, this continuum was not used as a predictor in the final analysis. It is described here for the sake of completeness.

The stressed vowel /ɪ/ of 7 BA and 7 SBE target words was acoustically manipulated, in order to increase its acoustic similarity with the equivalent Bulgarian vowel. In the acoustic manipulation of the vowels, the distance between the first and second formant was artificially widened to increase their similarity with a Bulgarian /i/ vowel. This was achieved using the programme Praat (Boersma and Weernink, 2019) and a script by Winn (2014) (see Appendix 3.A). The script allows the user to change the shape and position of Praat's formant trackings, thus changing how the segment sounds.

For this experiment, a whole vowel formant was selected and moved up or down, preserving its shape. The first formant in the stressed /ɪ/ was lowered between 150 and 250 Hz and the second formant was raised by the same magnitude. This range of change was determined based on initial trial-and-error when using the script by Winn (2014). The aim was to achieve an audible difference from the original, but the main consideration was to avoid glitches so that the stimuli sound like natural human productions. As a result, it was decided to increase the distance between F1 and F2 as little as possible, while still achieving an audible difference, as judged by the author. A change of 150-250 Hz was enough to achieve this goal. The naturalness of the manipulated stimuli was judged by the author and two phoneticians who are native English speakers.

The onsets and offsets of the vowels were determined by using a mix of cues from the waveform and the spectrogram. The main cue was the waveform, while the spectrogram was mostly used to confirm decisions based on the waveform. The zero crossing of the first upward movement of regular periodicity in the waveform was selected as onset and the zero crossing of the last upward movement of regular periodicity was selected as offset of the vowel. Where the vowels were preceded or followed by a liquid consonant /r/ or /l/ or a semi-vowel /w/ the whole cluster of consonant and vowel was selected to become manipulable (using the criteria listed above). However, the formants were actually manipulated only in the sections where they were stable and horizontal (i.e., associated with the vowel, not the consonant), and the consonant was left unchanged. In general, such words were not chosen for manipulation due to the lack of a clear boundary between /w/, /r/ or /l/ and the vowel.

An example of the manipulation process is presented in Figure 7 and Figure 8. The top figure shows the formant tracing of the [wi] portion in the word "switch". The original was pronounced with Bulgarian accent. In this case, the whole F1 was lowered, including in the glide portion. This was done to avoid sudden changes in an otherwise flat F1, thus preserving the integrity of the sound. F2 was increased only within the highlighted region, where it was horizontal. The resulting spectrogram of the whole word is presented in Figure 8.

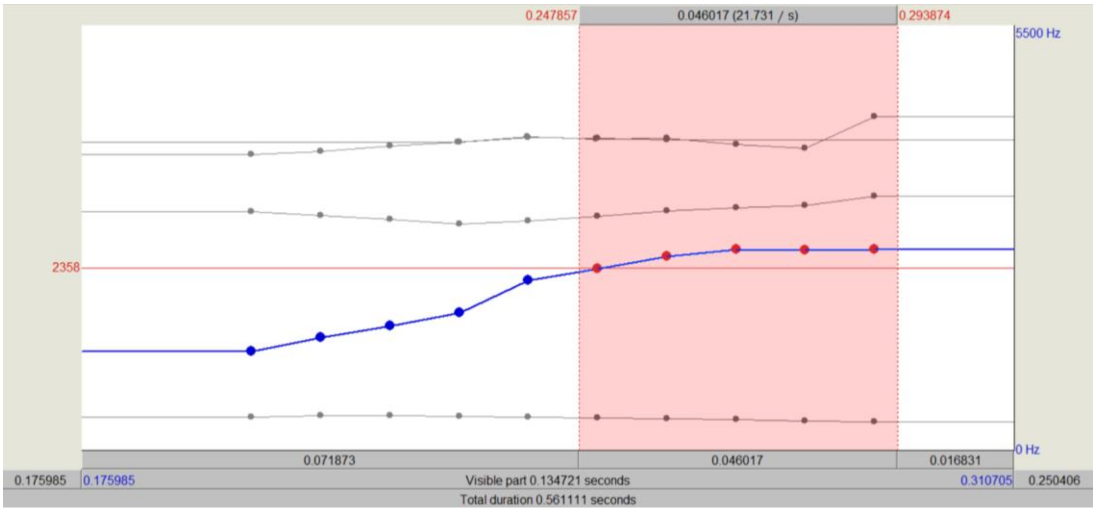


Figure 7. The formant structure of [wi] from the Bulgarian accented production of "switch" before increasing the distance between F1 (at the bottom) and F2 (highlighted in blue) in the pink region.

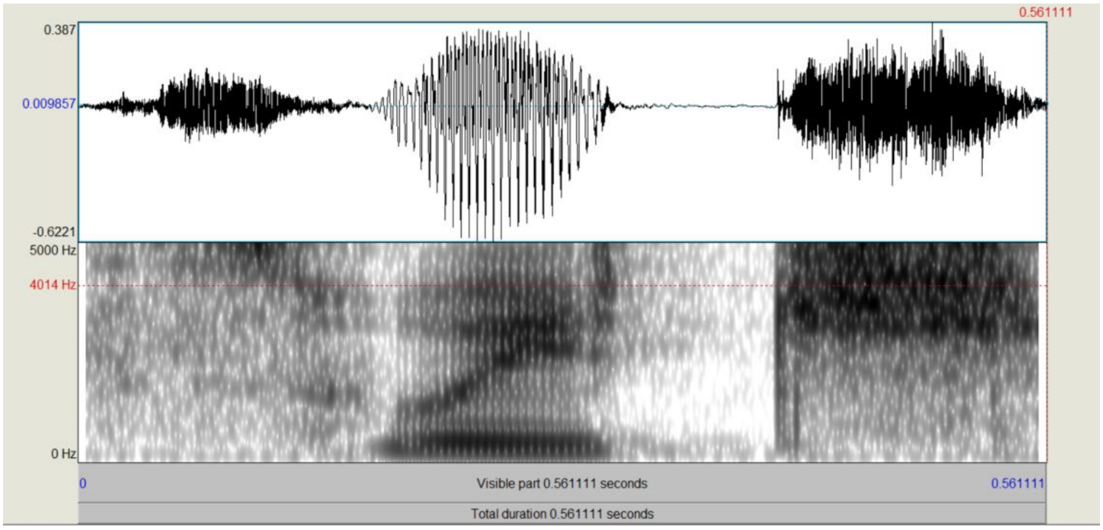


Figure 8. The spectrogram of the Bulgarian accented production of "switch" after formant manipulation of the stressed vowel.

In some cases, the process resulted in natural-sounding stimuli which, however, were judged by the author not to sound different from the original. In these cases, the process of manipulation was started again, by increasing the formant distance, while staying within the pre-defined limits of 150 – 250 Hz change per formant. Due to the inherent subjective nature of this initial auditory judgement and the manual modification of the formants, the manipulation is not perfectly replicable. However, by using the same or similar baseline stimuli, the guidelines provided here and the script in Appendix 3.A, it should be straightforward to achieve impressionistically similar albeit not identical results.

In order to avoid attracting attention to the manipulated words, 14 of the 38 distractors (7 from each accent group) were also manipulated by increasing the distance between F1 and F2. The same procedure was followed as in the manipulation of target words with /ɪ/. F1 was lowered between 150-250 Hz and the F2 was raised between 150-250Hz. As described earlier, these words did not contain stressed /ɪ/. These manipulated distractors were then included in the priming experiment.

It would have also been appropriate to present both the manipulated and unmanipulated stimuli to a group of independent listeners and ask them to rate the how natural they sound. This could have been done when English listeners for rated the stimuli for their foreign accentedness (described in Section 6.2.4.5). Some of the listeners could have rated half of the stimuli for naturalness and half of them for foreign accentedness, while other listeners could have rated the opposite groups. As it stands, the naturalness of the stimuli was evaluated by the author and two native English advisors.

6.2.4.4. Manipulation validity

To ensure that the manipulations were successful two forms of checks were carried out. Firstly, the Euclidean distances (explained below) between two sets of vowels were measured and compared. Secondly, the F2-F1 distance between the manipulated and non-manipulated vowels was directly compared.

Euclidean distance is a straight-line distance between two points in a space where each point is defined by two coordinates. When measuring Euclidean distances between vowels in this experiment, each vowel is defined by the value of F1 and F2 (measured in Bark) in the mid-point of the vowel. The formula used for the Bark transformation from the Herz values is the following: Bark = ((26.81*F)/(1960+F))-0.53 (Traunmüller, 1990). The onsets and offsets of vowels were manually marked using Praat (Boersma and Weernink, 2019) as explained earlier in the section and then the formant measurements were taken using a Praat script (Remijsen, 2013) (see Appendix 3.B).

For each accent (BA or SBE), three sets of vowels were compared. First, the mean F1 and F2 (in Bark) were calculated for all the words that were not undergoing any manipulations (Set A in Figure 9). This mean represented one point in the Euclidean plane. Then, there were the words undergoing manipulations (Set B1, B2 in Figure 9). F1 and F2 measurements (in Bark) were taken from these words before the manipulation (Set B1) and after the manipulation (Set B2). Euclidean distances were calculated between each vowel in the sets on the right-hand side of Figure 9 and the points defined by Set A. It is important to note that Sets B1 and B2 contain the same words respectively, which differ between each other only in the formants of the stressed vowel /ɪ/. The labels "b1", and "b2", are used for grouping the Euclidean distances according to stage of manipulation, 1 is pre-manipulation and 2 is post-manipulation. This measurement ensures that the B1 words (i.e., before manipulation) differed significantly from their B2 (after manipulation) counterparts compared to unmanipulated words. The un-manipulated /ɪ/ vowels were chosen for this comparison because they represent the accent (unmanipulated BA) that the manipulated stimuli had to stand out from. As the unmanipulated stimuli were also used in the experiment, this comparison verifies that the manipulated stimuli were distinct from both their original version and the other accent presented to the participants (unmanipulated

BA). Choosing a schwa as an anchoring point would have been less relevant to the target stimuli presented in the experiment.

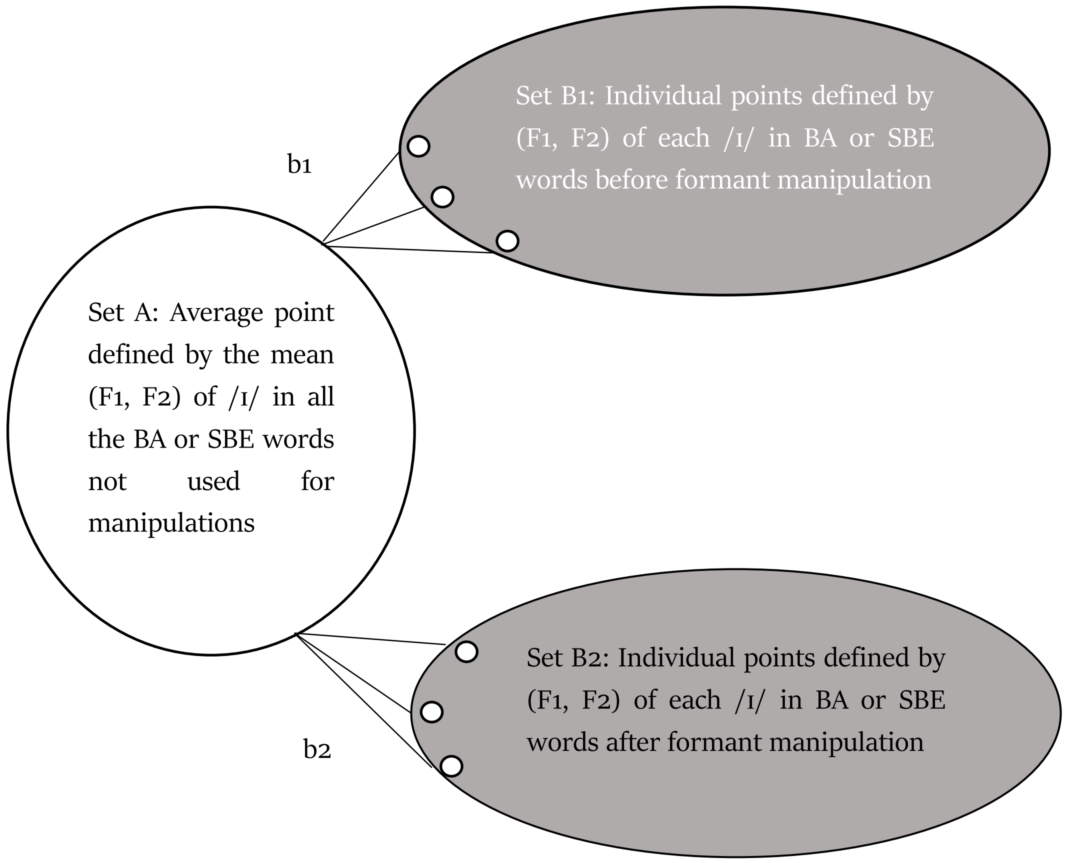


Figure 9. Three sets of Bulgarian-accented words used for calculating Euclidean distances between BA vowels.

A mixed design ANOVA met the requirement of homogeneity of variance (Levene's test F (1, 15) = 0.09, p = 0.77 pre-manipulation and F (1, 15) = 0.34, p = 0.57 post-manipulation). It had word-accent (BA or SBE) as a between-item factor and manipulation stage (pre- and post-) as a within-item factor. The dependent variable was the Euclidean distances between the words, which underwent no manipulation and the words that were manipulated before and after their manipulation (previously calculated as b1, b2, see Figure 9.) The test showed a significant main effect of manipulation stage (F (1, 15) = 85.04, p < 0.001, h^2_p = 0.85) but no significant main effect for accent (F (1, 15) = 0.08, p = 0.775, h^2_p = 0.006), and no significant interaction (F (1, 15) = 0.502, p = 0.489, h^2_p = 0.032). These results are illustrated in Figure 10. They mean that the vowels from both accents were manipulated to a similar extent and that the before and after states differ to a large enough degree to yield significant differences with large effect sizes.

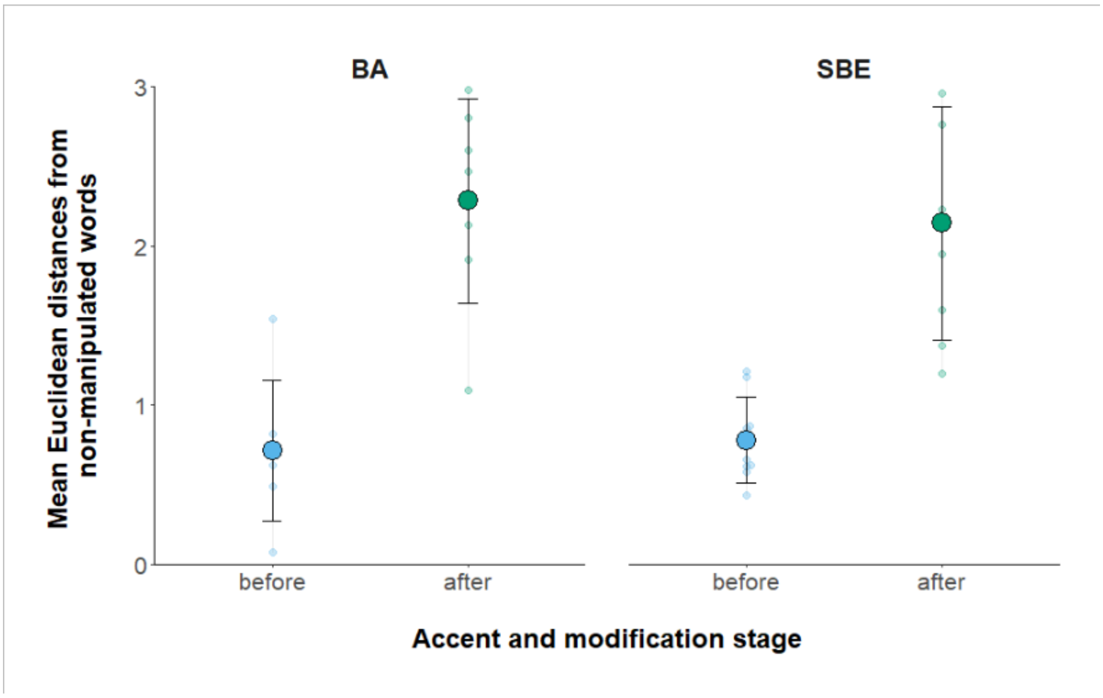


Figure 10. Mean Euclidean distances between the formants of the stressed vowel in the words with no manipulation and the targets that underwent manipulation, grouped by accent and stage of manipulation.

In addition to this test, the difference between F2 – F1 (in Bark) was compared for the tokens with and without manipulation (see Figure 11). This test is not concerned with the before and after effect of the manipulation but with the direct comparison of the formant structure of the manipulated and non-manipulated vowels that the participants heard. A two-way ANOVA with the difference of F2 – F1 (Bark) as a dependent variable shows that there are significant main effects of the Accent of the word (F (1, 39) = 36.3, $p < 0.001$, $h^2_p = 0.009$) and a significant main effect of the Manipulation status of the word (F (1, 39) = 52.4, $p < 0.001$, $h^2_p = 0.014$) but no significant interaction (F (1, 39) = 0.01, $p = 0.93$, $h^2_p < 0.001$).

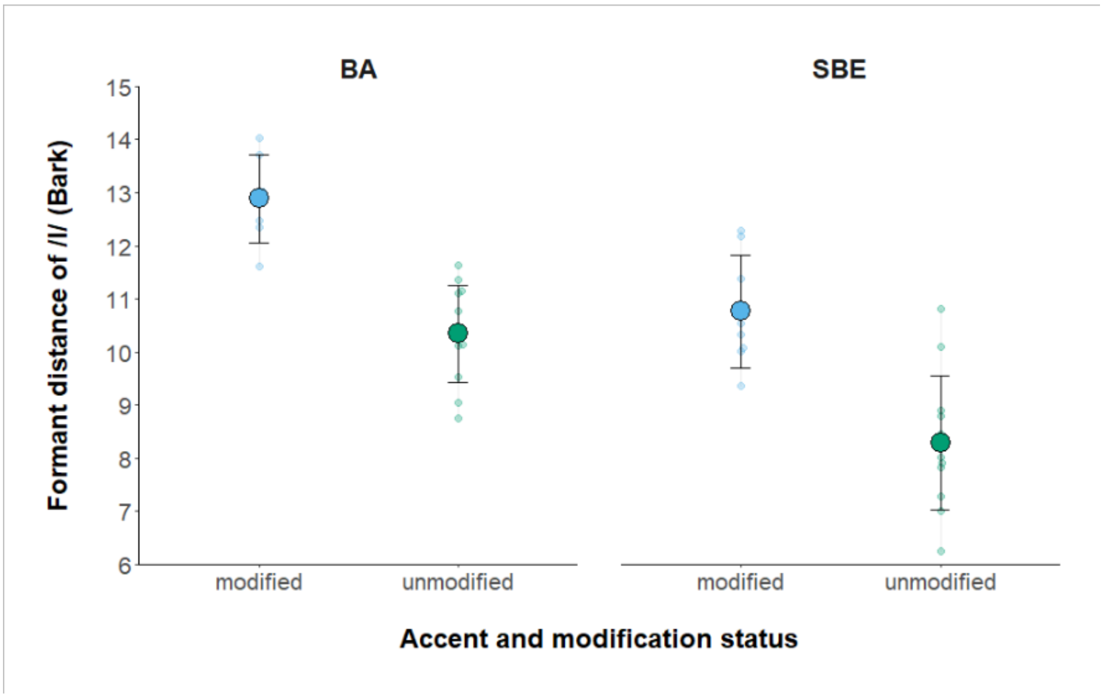


Figure 11. Formant difference between the words that underwent manipulation and the words that were presented with no manipulation grouped by Accent and by Manipulation of the word.

Overall, there is acoustic evidence that the phonetic manipulation produced a significant acoustic difference in the stimuli compared to their original state. In addition, they differed significantly from the non-manipulated stimuli.

6.2.4.5. Stimuli rating procedure

The stimuli were presented to the raters via Google forms. All the sound files were stored in a Google drive folder, with a masked name.

After reading the study information, and consenting to participate, the raters filled out a short questionnaire. It was made explicit to them that they could omit any of the answers without giving an explanation. The questionnaire asked for their age, gender, languages spoken at home when growing up, languages spoken at the time of participation (even as rarely as once a month) and how frequently they communicated with non-native speakers of English.

The raw and manipulated BA and SBE words were presented to the native English listeners who rated their level of accentedness. Initially there were total of 38 raw and manipulated BA (N = 12 raw and N = 7 manipulated) and SBE (N = 12 raw and N = 7 manipulated) target words and 38 distractors were prepared for the experiment. They were split into two wordlists that contained 38 unique tokens (both target and distractors). Unfortunately, due to a technical error one of the manipulated Standard British English words, “artificial”, was not properly included and lacks rating information. Each participant was randomly assigned to one of the two wordlists. Each of the stimuli within the wordlist was presented to the listeners twice in random order.

In the main part of the listening task, the raters saw a slide with 38 questions presented in random order. Each of the questions contained a link to a unique sound-file (19 target and 19 filler items). Each question contained the instructions: "Click on the link and listen to the recording no more than twice. Rate it based on your immediate overall judgement." Underneath each question there was a 9-point scale. Next to 1, the raters saw the label "no accent" and next to 9 there was a label "very strong accent." When the participants finished rating each sound-file, they moved on to the second page of the experiment, in which they saw the same files in a random order of presentation and had to rate them again. It was not possible for the listeners to go back to the previous page and change their answers.

6.2.4.6. Raters

The items were rated by 10 native English speakers (mean age 35.4, *SD* = 11), 3 female. The raters had only spoken English growing up and five of them were monolingual. To ensure that they had sufficient experience with Standard British English, which was the target variety of the speaker, recruitment was restricted to listeners brought up in the UK. All the raters reported that they communicated with non-native speakers of English on a regular basis (summarised in Table 5).

The raters were recruited online. Recruitment materials, including selection criteria and summary of the experiment were shared via social media. The raters had the opportunity to directly click on a link that led them to detailed information about the experiment and the experiment itself, without having to communicate with the researcher.

Rater	Gender	Age	Languages spoken in childhood	Languages spoken now (even only once or twice a month)	Frequency of talking to non-native speakers of English
1	Female	28	English	English	A few times a month
2	Male	32	English	English, Mandarin, Spanish	Daily
3	Male	39	English	English	A few times a week
4	Male	25	English	English	Daily
5	Male	50	English	English, Spanish	A few times a week
6	Male	24	English	English, French, Japanese, German	Daily
7	Male	33	English	English and Spanish	Daily
8	Male	54	English	English	A few times a week
9	Female	43	English	English, Spanish, Portuguese	Daily
10	Female	26	English	English	Daily

Table 5. The raters' questionnaire data.

6.2.4.7. Rating outcomes

Both the acoustic manipulation and the intended accent influenced the listeners' foreign accentedness ratings. The results are summarised in Table 6 and illustrated in Figure 12. A list of the words and their rating results is available in Appendix 1.A.

A linear mixed effects model was carried out. It had the predictors *Accent* (SBE or BA) and *Manipulation Status* (modified and unmodified), which were sum coded, and their interaction. The model included random intercepts for *Test Item* and random slopes between *Accent, Manipulation Status* and *Participant*. A random intercept means that means that the final model takes into account that each individual *Test Item* can systematically receive higher or lower rating regardless of its superordinate *Accent* or *Manipulation*. A random slope between the predictors and *Participant* means that the model takes into account between-participant variation with respect to these predictors. The *Manipulation Status* of a word and its *Accent* can both affect average ratings that each participant gave to a different degree.

The test was based on 365 observations from 10 participants and 37 tokens.

Predictor	Estimate	*t*-value	*p*-value
Intercept	5.13	19.26	<0.001
Accent	-1.59	-7.05	<0.001
Status	-0.70	-3.53	0.003
Accent : Status	-0.37	-1.84	0.07

Table 6. Summary of accent rating results.

The results suggest that the there was a main effect of *Accent* and of *Manipulation Status*. The Bulgarian-accented words were rated as overall stronger accented, while the words with an intended SBE accent were rated as less foreign accented. The words with manipulation were also rated as more foreign accented than the words with no manipulation. There was no significant interaction between Accent and Status.

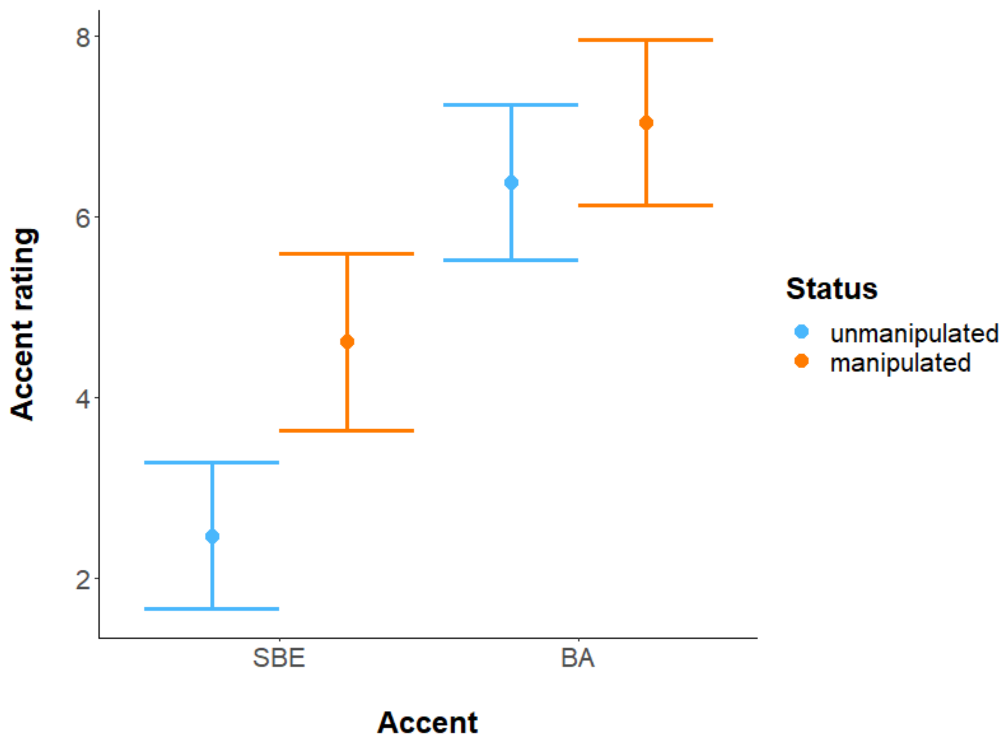


Figure 12. Native English speaker rating of auditory stimuli. Scale: 1 = no accent at all, 9 = very strong accent.

6.2.5. Procedure

In order to reach as many participants as possible, this experiment was carried out either in a sound-proof room at Queen Margaret University or in quiet rooms outside of the campus, using noise cancelling headphones Sony mdr zx770bn. After the participants were debriefed and signed a consent form, they were invited to try the practice trials of the cross-linguistic auditory lexical decision task. Upon completing the practice run, the participants had a chance to ask for clarifications, after which they were left alone in the room to finish the experiment. The participants had 2500 ms to enter a response, after which the experiment automatically proceeded to the next stimulus. The interstimulus interval was 1250 ms and between each block there was a break of ten seconds. On average the participants finished the experiment in 3 minutes, not counting the training phase.

After completing the timed auditory lexical decision task, the participants carried out LexTale (Lemhöfer and Broersma, 2012b). This is an orthographic lexical decision task in English and it has been shown to have a high reliability in determining the proficiency and vocabulary knowledge for learners of English (Lemhöfer and Broersma, 2012a). The participants saw strings of letters on the screen and had to decide whether they were real words or not. The test consists of 20 non-word and 40 word items and the final score is calculated by using the following formula (Lemhöfer and Broersma, 2012b):

((number of words correct/40*100) + (number of nonwords correct/20*100)) / 2

The test produces a proficiency score between 0 and 100%. Intermediate to high proficiency Korean and Dutch listeners achieved 67% and 76% average scores on the test, respectively.

Lastly, the participants filled out a linguistic background questionnaire (see Appendix 2.A). It was divided in five parts: usage of Bulgarian in childhood, usage of Bulgarian currently, usage of English in childhood, usage of English currently, and attitudes towards Bulgarian and Anglophone culture and language.

6.2.6. Analysis

In an attempt to simplify the structure of the experiment, several decisions were made that reduce the generalisability of the results. These decisions were made in order to keep the tasks short and to focus on testing the tools used for this reaction-time auditory experiment. The initial goal at the time was to later expand the experiment for a larger scale version of the same study, hence power was not the main concern.

Firstly, the number of test tokens per condition is low hence all the statistical tests have a low power. Also, only correct responses were included in the analysis. Overall, the accuracy levels were very high. Of the initial 460 observations for analysis, only 10 were excluded due to inaccuracy.

Second, the initial aim for this experiment was to investigate the effect of gradually increasing the strength of the Bulgarian accent of the stimuli, using both the /i/-like and /ɪ/-like BA and SBE stimuli as primes and targets. However, due to the low power, the effect of artificially increasing the Bulgarian accent was not investigated and only the effect of an overall Bulgarian vs. SBE accent was investigated pooling, together the unmanipulated and manipulated per overall accent. Unfortunately, this makes a large part of the experiment redundant.

In order to keep the experiment short and focused on the critical cross-linguistic conditions, the following conditions were not included in the experiment: Standard British English priming Bulgarian-accented English and Bulgarian-accented English priming SBE.

This meant that the analysis of the centred RTs had to be carried out in three subgroups: a. monolingual conditions, b. conditions with Bulgarian primes and English targets and c. conditions with English primes and Bulgarian targets. It was deemed that within each of these subgroups the results can be used to meaningfully assess the importance of the separate variables. As several baseline blocks were not included in the experimental design, analysing all of the data at the same time could lead to misleading results.

Previous research has shown that priming from L2 to L1 is less effective, but in Szakay et al. (2016) that was the condition in which a facilitatory link was found between Māori and Māori English. Hence, the main expectation for this experiment was that Bulgarian-accented English, not SBE primes, would lead to faster processing times for Bulgarian words.

6.3. Results

6.3.1. RT in blocks with monolingual prime and target

This section reports the results from the monolingual blocks of the experiment. It investigates the question whether in the absence of cross-linguistic priming the listeners with different proficiencies will process Bulgarian, Bulgarian accented (BA) English and Standard British English (SBE) with different overall reaction times. The results suggest that there are no significant preferences for accents across the proficiency scale.

When studying the reaction times (RTs) in the monolingual blocks, a linear mixed effects regression model was constructed with *Centred RT* as the outcome variable, and two predictors: sum coded *Condition* (Bulgarian vs. Bulgarian-accented English, and Bulgarian vs. SBE) and *Proficiency* (the LexTale score centred around its mean). The model included random slopes of *Condition* per *Participant* and random intercepts per *Target*. The results are based on 114 observations from 20 participants and 6 tokens.

None of the predictors had a significant effect on the RTs (see Table 7).

Predictor	Estimate	*t*-value	*p*-value
Intercept	-3.11	-0.08	0.94
Condition (Bulgarian vs BA)	0.24	0.01	1.00
Condition (Bulgarian vs SBE)	0.77	0.02	1.00
Proficiency	-0.31	-0.12	0.91
Condition (Bulgarian vs BA) : Proficiency	1.23	0.44	0.67
Condition (Bulgarian vs SBE) : Proficiency	-2.22	-0.84	0.41

Table 7. Summary of results from monolingual blocks.

When observing the model predictions in Figure 13, there is an intriguing tendency. On the one hand, participants with a lower Proficiency score tended to react faster to Bulgarian-accented English and Bulgarian compared to Standard British English targets. On the other hand, the participants with the highest Proficiency results reacted faster to Standard British English targets and slower to Bulgarian and Bulgarian-accented targets.

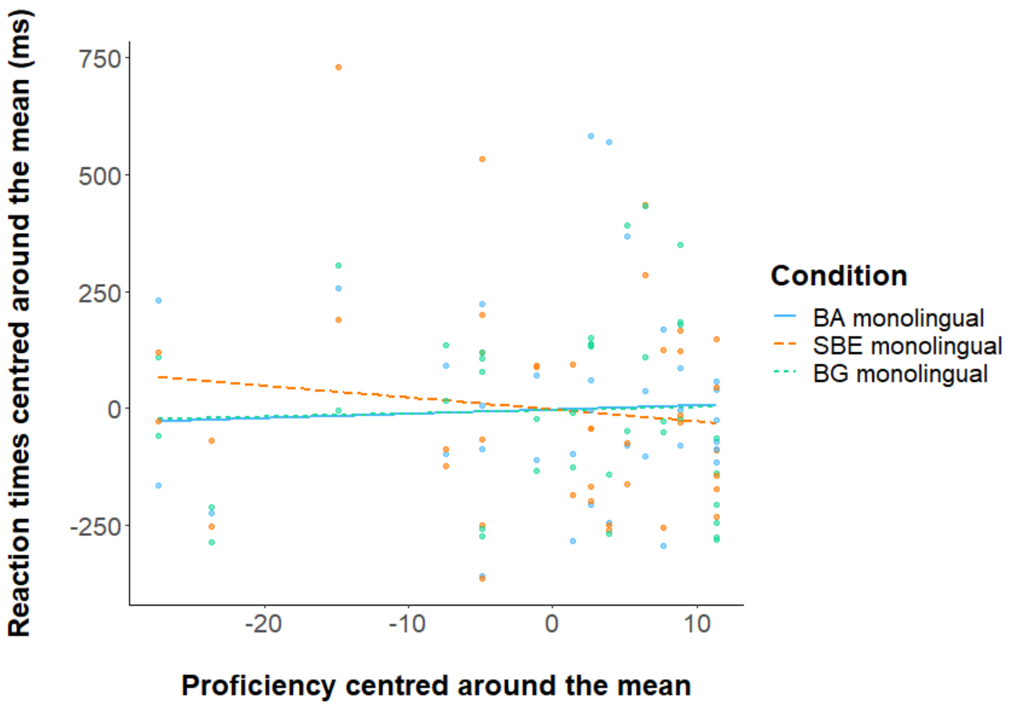


Figure 13. Reaction times to Bulgarian-accented English, Standard British English, and Bulgarian target words in monolingual blocks.

6.3.2. RT in blocks with English primes and Bulgarian targets

This section investigates the question whether BA or Standard British English serves as a better prime for Bulgarian translation equivalent targets for Bulgarian-English bilinguals with different proficiencies. The results suggest that none of the accents was significantly better than the other as a prime for Bulgarian target words.

When studying the RTs for English words priming Bulgarian targets, a new linear mixed effects model was created for the cross-linguistic blocks with Bulgarian targets. The outcome variable was *Centred RT* and it had two predictors: sum-coded *Condition* (Bulgarian- accented English vs. Standard British English) and *Proficiency* (the LexTale score centred around its mean). The model included random slopes of *Condition* per *Participant* and random intercepts per *Target*.

The results are summarised in Table 8. They are based on 159 observations from 20 participants and 8 tokens. While, again, none of the predictors were significant, an observation of Figure 14 shows a similar pattern of results to the previous section. Participants with a higher proficiency score responded to Bulgarian targets faster when they were preceded by SBE primes, as opposed to Bulgarian-accented primes. The primes had less of an effect on the target recognition for the lower proficiency listeners.

Predictor	Estimate	*t*-value	*p*-value
Intercept	-0.52	-0.02	0.98
Condition (BA vs. SBE)	40.50	1.88	0.10
Proficiency	0.07	0.04	0.97
Condition : Proficiency	1.00	0.85	0.41

Table 8. Summary of results from blocks with English primes and Bulgarian targets.

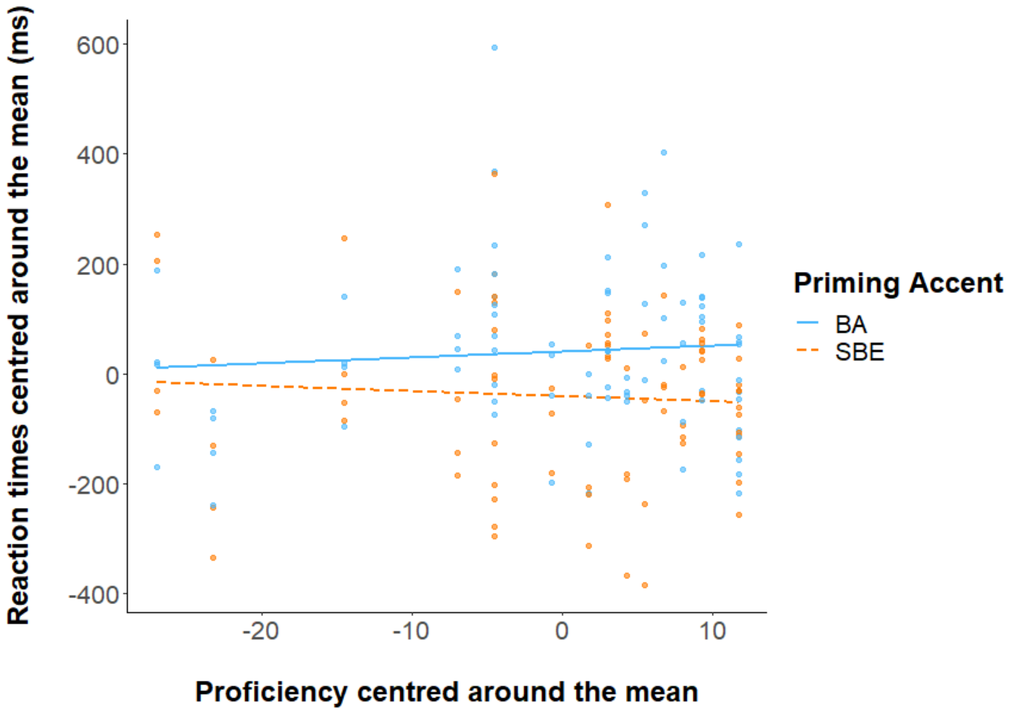


Figure 14. Reaction times to Bulgarian target words with Bulgarian-accented English or Standard British English primes.

6.3.3. RT in blocks with Bulgarian primes and English targets

This section aims to answer the question whether BA or Standard British English is processed faster when preceded by a Bulgarian prime by listeners with different proficiencies in English. The results suggest that listeners with higher English proficiency processed SBE faster than listeners with lower English proficiency, while BA English was processed faster by lower proficiency listeners compared to higher proficiency listeners. This is similar to the tendency observed in the monolingual blocks.

In order to investigate the RTs in blocks with Bulgarian as a priming language and English as a target language a similar model was constructed to the one used in the previous section. The outcome variable was *Centred RT* and it had two predictors: sum-coded *Condition* (Bulgarian-accented English vs. SBE) and *Proficiency* (the LexTale score centred around its mean). The model included random slopes of *Condition* per *Participant* and random intercepts per *Target*.

The results are summarised in Table 9. They are based on 177 observations from 20 participants and 9 tokens. Of the predictors included in this model only the interaction between *Proficiency* and *Condition* was significant (β = -7.4, t = -2.4, p = 0.03). Figure 15 suggests that as the participants' proficiency increased, their RTs to the SBE words lowered, while their RTs to Bulgarian-accented words increased. The opposite relationship is present in the lower end of the proficiency continuum.

Due to the limitations of the design and the missing baseline levels, there is no way of distinguishing between the effect of the target word's accent and the joint effect between the prime's Bulgarian language and the target's accent in English. However, as all prime-target pairs in these conditions had Bulgarian as primes and English as targets, it can be tentatively assumed that if there was any priming between Bulgarian and Bulgarian accent in English, it would be observed for L1 Bulgarian - L2 English listeners with lower proficiency in English and not for listeners with higher proficiency.

Predictor	Estimate	*t*-value	*p*-value
Intercept	-0.05	-0.001	1.00
Condition (BA vs. SBE)	-3.14	-0.09	0.93
Proficiency	-1.49	-0.60	0.56
Condition : Proficiency	3.70	2.40	0.03

Table 9. Summary of results from blocks with Bulgarian primes and English targets.

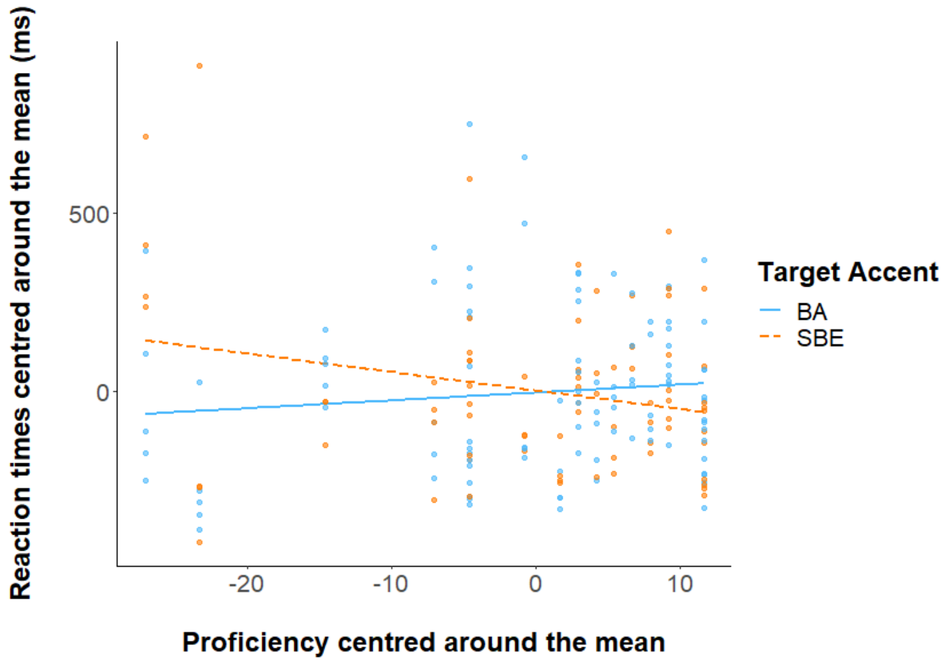


Figure 15. Reaction times to Bulgarian-accented English or Standard British English target words with Bulgarian primes.

6.4. Discussion

As outlined earlier, the methodological drawbacks of this experiment limit the theoretical inferences that can be drawn based on the data. Even though there were almost no significant results, as reported from the three models in the previous section, the interaction plots follow a coherent pattern. The following analysis is mostly based on the tendencies observed in the interaction plots and should not be considered as conclusive.

Overall, there appears to be no unambiguous link between Bulgarian accent in English and Bulgarian vocabulary either in L1 to L2 or in L2 to L1 priming. The clearest pattern observed in the data is that low proficiency English listeners respond faster to Bulgarian-accented stimuli (such as [link]), compared to Standard British English stimuli (such as [lɪŋk]) when they are preceded by the Bulgarian translation (such as [vrəzkə], "link"). The results differ from Szakay et al. (2016) in relation to the direction of language priming. Szakay et al. (2016) report a facilitatory link in the L2 to L1 condition (between Māori and Māori-English speech), and here it is observed in the L1 to L2 condition (Bulgarian priming English). That being said, there are no proper baselines for the L1 to L2 condition in this experiment and it is possible that the lower proficiency listeners' preference for BA English is not a response to the Bulgarian primes but a general preference for Bulgarian-accented speech. All the plots suggest that lower proficiency listeners tend to process BA targets or primes more efficiently than higher proficiency listeners. The opposite trend is not observed at all.

One caveat is that all the stimuli are ultimately the productions of the same L2 English speaker, even if the strength of the foreign accent and /ɹ/ realisations systematically differ between them. In one sense, the effects of the two groups of English stimuli can be interpreted as a strong and weak Bulgarian accent, instead of Bulgarian accent and SBE. Under this assumption, the facilitation results for the lower proficiency listeners can be interpreted as the effect of not just a sociolinguistic Bulgarian "tag" activation (as in Szakay et al., 2016) but rather a link between a predominantly Bulgarian phonetic cues in English with Bulgarian. Admittedly, the Bulgarian sociolinguistic activation may have increased due to the higher number of Bulgarian phonetic cues, but that also means that they have played a role in speech processing. Also, any tendency for benefit from stronger Bulgarian accent for the lower proficiency listeners and weaker accent for the higher proficiency listeners may be due to a similarity with their own accents in English. This possibility warrants further research.

The main expectation, that Bulgarian words would be more successfully primed by Bulgarian-accented stimuli than by SBE stimuli, is not supported. In fact, the opposite trend is observed, especially for the higher proficiency listeners. If SBE has a tendency to be a better prime for Bulgarian words, this suggests that SBE words allow the listeners to reach the correct lexical representation more efficiently than BA words and are perhaps closer to the listeners' core representations of English. It follows that in this case there is no sociophonetic shortcut from BA English to Bulgarian, if BA English is not the core representation of the listeners in the first place.

These results grossly pattern as some of the observations from the literature. Ludwig and Mora (2017) found that lower proficiency L2 listeners have faster processing of L1-accented L2 speech than native speech and that this benefit was not observed for higher proficiency bilinguals. In addition, Hayes-Harb et al. (2008) observed an interspeech intelligibility benefit only for low proficiency listeners when responding to low proficiency speech. Together with the observations in this study, it can be summarised that if an L1-accented L2 speech benefit is to be observed at all, it will likely happen for lower proficiency listeners who are listening to a stronger L1 accent.

The different listening outcomes related to proficiency can be linked to the predictions of some of the models of bilingual perception. SLM assumes that lower proficiency listeners maintain correspondences or mergers between L1 and L2 phones, which might become more distinct as the L2 learners gain more exposure to L2. ASP predicts that with long-term exposure to a variety the listeners will develop efficient listening routines for picking up relevant phonetic information in that variety. These models in different ways are consistent with the observation that the higher proficiency English listeners might be more specialised to a standard UK variety as opposed to a more Bulgarian-like phonetic realisation.

These models and the data observed here contradict the view of the higher proficiency listeners as more “flexible” and having accumulated high expertise in several accents at once, as seen in Ludwig and Mora (2017). The data in the monolingual trials of this experiment suggest that as listeners become more proficient in L2 they may have to trade off their most optimal accent for processing. If the proficiency results are interpreted as a cross-sectional look at L2 listening development, then it might speculated that there is be a point in their development, at which they are balanced in both accents, but balance is unlikely to persist as the listeners’ circumstances change.

6.5. Limitations

There are several limitations that prevent the study from having more conclusive results.

There were only three blocks of monolingual prime-target pairs (only Bulgarian, only Bulgarian-accented English and only Standard British English), compared to the eight cross-language blocks (see Table 3). This makes the monolingual blocks poor baselines for the cross-linguistic blocks. The listeners may have become accustomed to language changes and may have been surprised to encounter monolingual prime-target pairs, which might explain why their reaction times in the monolingual blocks are slower than in the cross-linguistic blocks (cf., Task-Set Inertia Hypothesis in Allport et al., 1994).

Lastly, within each block there are fewer tokens of the relationship which is expected to be costlier to process – the semantically unmatching one. As a result, any potential effect of semantic relationship cannot be disentangled from the effect of unequal frequency of representation. This aspect of the study prevents any conclusions regarding the expected L1-to-L2 and L2-to-L1 priming asymmetry.

6.6. Conclusions

To conclude, despite the lack of significant results and the methodological flaws, the results of this experiment tend to align with the discoveries of other studies addressing L1-accented L2 benefits for processing. Lower proficiency English listeners were generally faster when processing Bulgarian-accented English than higher proficiency English listeners. The lower proficiency listeners also tended to process Bulgarian accented English faster than SBE. These results suggest that the lower proficiency English listeners were more specialised to Bulgarian accented English perhaps due to the stage of development of their L2 phonology. This interpretation is consistent with several models of L2 phonological and listening development, such as SLM and ASP.

In order for this hypothesis to be explored more reliably, it is necessary to focus only on the relationship between listeners' proficiency in English and the accents of the speakers. However, as observed previously in Chapter 3 there are very few studies, which investigate this research question and measure the listeners' proficiency consistently and measure L1 accent processing as their primary goal. Of those, no other studies operationalise the listener proficiency as a continuous variable, which would be more representative of the population of L2 listeners. In addition, Ludwig and Mora (2017) whose study is one of the few to investigate this question tested their participants within predominantly L1 speaking environment, where the majority of their L2 input was L1-accented. This puts into question whether their low proficiency listeners' preference for L1-accented L2 was a part of their development of L2 or if it was not mostly influenced by their external input. Hence the following chapter will investigate the processing of Bulgarian-accented English by Bulgarian-English bilinguals who live in the UK and have varying proficiencies in English.

Chapter 7 Experiment 2 - Processing of Bulgarian and English accented words by Bulgarian-English bilinguals

7.1. Background

The main aim of this study is to find out whether Bulgarian-accented English (e.g., [xɛə] for "hair") or Standard British English (e.g., the standard [hɛə]) is systematically easier to process for Bulgarian-English bilinguals and if the accent specialisation changes depending on the listeners' proficiency.

As discussed in Chapter 3, previous research has demonstrated that the listeners' proficiency and experience with listening to a particular accent can improve their speed and accuracy of processing speech with that accent (Ludwig and Mora, 2017; Porretta et al., 2016; Sumner and Samuel, 2009; Witteman et al., 2013). The results from the study, reported in the previous chapter, indicate that Bulgarian-English bilinguals living in the UK tend to process Standard British English words faster than Bulgarian-accented words, especially if the listeners have high proficiency in English. On the other side of the spectrum, lower proficiency bilinguals tend to be faster with Bulgarian-accented English, particularly when it is preceded by Bulgarian primes, compared to the Standard British English targets.

The small scale of Experiment 1 is the reason for the lower variability in the participants' proficiency scores, which makes the conclusions less certain. However, the results are similar to Hayes-Harb et al. (2008) and Ludwig and Mora (2017), who also find that lower proficiency listeners have some processing speed and accuracy benefits from L1-accented L2 speech compared to higher proficiency listeners. No other studies seem to systematically investigate the effect of proficiency on the processing of L1-accented L2 speech.

One of the dominant theories on L2 phonological development, SLM (Flege, 1995) suggests that in the early stages of L2 acquisition listeners share phonological categories with L1 and as they increase their proficiency the categories might separate. That leads to the prediction that lower proficiency listeners might have faster and more accurate processing when they encounter phonetic cues in L2 that acoustically resemble their L1. On the other hand, BIMOLA (Léwy and Grosjean, 2008), which focuses on speech perception from the encounter of acoustic input to the final lexical access, predicts that phonetic characteristics consistent with L1 will increase the competition of L1 vocabulary for the listener. According to the structure of BIMOLA, L1-like phonetic information in an L2 phonetic context might slow down the processing of L2 due to increased cost of inhibition of competing L1 vocabulary (as observed in Lagrou et al., 2015). However, BIMOLA makes no predictions regarding different levels of L2 proficiency. Assuming a given group of low and high proficiency L2 listeners have similar levels of access to L1 phonetics and vocabulary, then the predictions of BIMOLA should hold for all, regardless of their specific L2 proficiency.

While proficiency may be viewed as the result of long-term adaptation to a specific variety, it is also important to investigate the listeners' ability to rapidly adapt to unfamiliar accents and speakers. It is the accumulation of such short-term adaptations that ultimately leads to generalised learning outcomes. Some of the theories reviewed in Chapter 1 and Chapter 2, such as IAF, TBIM and ASP have specific mechanisms that allow listeners to adapt to novel acoustic variation within a short time. Studies such as Clarke and Garret (2004), Sidaras et al. (2009) suggest that naïve listeners can improve their accuracy and speed of processing when exposed to an unfamiliar accent within a short time period, although Xie and Myers (2017), Xie et al. (2018) and Kriengwatana et al. (2016) point out that accent adaptation necessarily passes through a speaker-specific adaptation stage. Hearing a varied selection of voices allows listeners to more efficiently adapt to a new voice of the same accent. Hence it is expected that listeners with lower proficiency in English will adapt faster to Bulgarian accented English than SBE, while listeners with the highest proficiency in English may have accumulated experience in both accents and will adapt to them at similar rates.

7.2. Predictions

The present experiment investigates the following predictions by focusing on a group of Bulgarians (L1) – English (L2) bilinguals who have migrated from Bulgaria to the UK.

Based on this discussion and mainly the results of Ludwig and Mora (2017), it is expected that the listeners with lower proficiency will be more specialised in processing Bulgarian-accented English compared to native Standard British English. They would be expected to respond faster and more accurately to [xɛə] than [hɛə]. The higher proficiency bilinguals are expected to be similarly well-adapted to Bulgarian-accented English compared to the lower proficiency listeners and better adjusted than them at native Standard British English. This means that they are predicted to be similarly accurate and fast with both [xɛə] and [hɛə] for "hair". Hence it is predicted that there will be an interaction between the stimuli's accent and listeners' proficiency in their speed and accuracy of word recognition.

In terms of within-experiment short-term adaptation, it is expected that lower proficiency listeners will adapt faster to Bulgarian accented English than SBE, while high proficiency listeners will adapt to these two varieties at a similar rate. In addition, as stated earlier, the highest proficiency listeners are likely to be more specialised in SBE than low proficiency listeners. Hence it is expected that they will adapt to this baseline accent faster than the lower proficiency listeners.

7.3. Methods

7.3.1. Overview

In this study Bulgarian-English bilingual residents in the UK completed an auditory lexical decision task, an English proficiency test, which was a visual lexical decision task, and a questionnaire (see Appendix 2.B). The experiment was carried out online and distributed via social media. The participants heard four speakers (two native English and two Bulgarian-English bilinguals) produce 16 target words and 16 non-words each and had to decide if they are hearing a real word or not by pressing a button. The speed and accuracy of their responses were recorded for analysis.

7.3.2. Design

The experiment has a within-subject design, in which all participants heard words produced by all four speakers. In order to restrict the length of the experiment and to avoid exposing the listeners to the test words more than once (which can affect their reaction times), each participant heard only 16 words and 16 non-words per speaker, adding up to the total of 64 words and 64 non-words and resulting in 4 versions of the experiment in which each quarter of the words was produced by a different speaker. Each listener heard each of the words produced by no more than one speaker.

The words were blocked by speaker and variety in order to allow the listeners to adjust to a voice and avoid affecting the reaction times by random changes in the speaker identity. To prevent order effects, the accent blocks and the speaker blocks within them were counterbalanced across participants. Within each block the stimuli were presented in a random order. Before the main task the participants heard 10 training trials, which contained three non-words and seven real words. These were words that had been previously used as stimuli in Experiment 1. A summary of the structure is available in Figure 16.

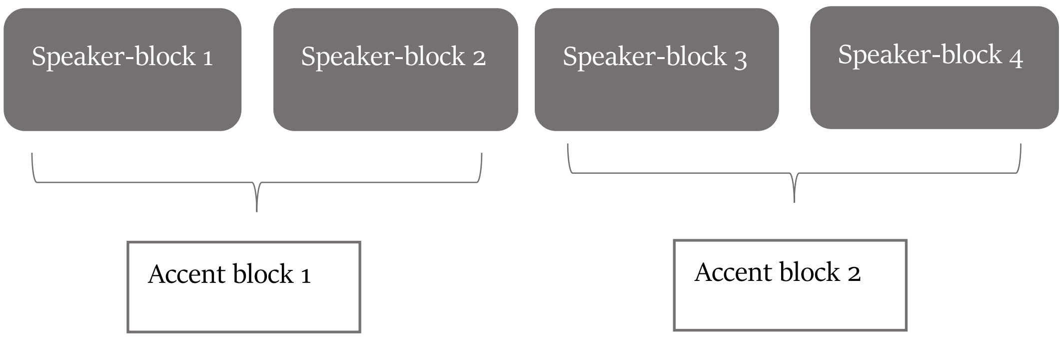


Figure 16. Structure of Experiment 2.

7.3.3. Participants

A total of 169 participants attempted the experiment. The call for participants invited people who consider themselves residents of the UK and who have been raised as Bulgarian speakers. People were invited to take part if they were comfortable reading the information about the experiment and the consent form in English, as a minimal requirement of functional ability in the language. Only the data of those who completed both the reaction time experiment and the proficiency were kept for further analysis. Sixteen participants never entered a response for the auditory task. As all "timeout" and incorrect responses were excluded from the final analysis, these participants were automatically completely filtered out. A further 6 participants were excluded from the analysis due to an error in the recording of their data. Hence the final number of participants whose data was analysed was 94. Their mean age was 30.2 (SD = 9). Of them 63 were female, 30 were male and 1 chose 'other'. Table 10 below summarises the distribution of participants across the conditions.

Accent order	Speaker order	Number of participants
SBE-BA	En1, En2, Bg2, Bg1	22
SBE-BA	En2, En1, Bg1, Bg2	18
BA-SBE	Bg2, Bg1, En1, En2	24
BA-SBE	Bg1, Bg2, En2, En1	26

Table 10. Distribution of participants across experimental conditions.

The participants were asked to estimate what percentage of their weekly time is spent talking to people whose native language is English and what percentage of that time is spent talking to native English speakers from England. England was chosen, as opposed to the UK in general, because it is more likely that residents of England sound like the native speakers of the stimuli recordings, than residents of Scotland or Northern Ireland, although it is recognised that SBE speakers also live in these two countries.

Based on these results a percentage of time spent talking to English people was calculated for each participant by multiplying these numbers and dividing them by 100 (M = 38.4%, SD = 32.7) (based on Porretta et al., 2016). For convenience this is called English exposure score. Similarly, the participants were asked to estimate what percentage of their time weekly is spent talking to non-native speakers of English and what percentage of that time is spent talking in English to Bulgarian speakers of English (M = 6.5 %, SD = 12). These two numbers were multiplied together and divided by 100. This variable is called Bulgarian exposure score.

Based on these two scores two Pearson correlations were calculated between the English exposure score and Age and another one between Bulgarian exposure score and Age. There was a positive correlation between Age and the English exposure score (t = 2.93, r (92) = 0.29, p = 0.004). There was a negative correlation between Age and the Bulgarian exposure score (t = -

2.4, r (92) = -0.24, p = 0.018). These results can be taken as an indication that the older participants were more integrated in the UK society and spoke to fewer Bulgarians in a mixed context, while the opposite was true for the younger speakers. There were no significant correlations between reported weekly exposure to Bulgarian-accented (t = -0.47, r (92) = -0.05, p = 0.64) or native English speech (t = -0.42, r (92) = -0.04, p = 0.68) and the listeners' proficiency in English.

When asked to report which is their target variety of English 64 participants responded Standard British English as taught at school (of them 1 specified northern English, 1 north western English, 3 RP); 13 North American; 9 Scottish English, 3 a mix of British and North American, 2 were not sure, 1 Bulgarian-English, 2 Irish English. These results are in line with the survey results on Bulgarians by Georgieva (2010) reviewed in Section 4.2 and demonstrate that the model of British English taught at school is the dominant standard model for L1 Bulgarian – L2 English speakers. Unfortunately, the online experimental platform used in this study did not record the IP addresses of the participants. Future work could use these data as an additional source of information about the participants' exposure to English accents.

7.3.4. Materials

7.3.4.1. Stimuli selection

The stimuli for this experiment were 64 monosyllabic English words and 64 monosyllabic non-words.

The words were chosen from the web-CELEX database (Baayen et al., 1995; Max Planck Institute for Psycholinguistics, 2001), starting with 100 monosyllabic words. This initial number was chosen based on the number in a comparable experiment by Lagrou et al. (2011), who included 88 target stimuli. It was planned that only a subset of these original stimuli would be included in the final experiment. They had a frequency over 3500 (of total corpus size 17.9 mln) and were not cognates with Bulgarian words. As the study focuses on people of varying English proficiency, only highly frequent words were chosen to minimise the overall effect of proficiency on reaction times due to lexical familiarity (and thus highlight the effects of accent familiarity).

The list was then narrowed down to the 64 best sounding tokens, according to the Foreign Accentedness ratings, in a procedure described in Section 7.3.4.3. The average phonological neighbourhood size of the final list was 21.31 (SD = 11.71) (Marian et al., 2012). Due to the high neighbourhood size it was possible that some words, when pronounced with Bulgarian accent might sound like an unintended lexeme. Such risk was identified for words containing stressed /æ/, like "had" or "land", which could be substituted with [e] and for "third" and "through" where substituting the initial /θ/ with [t] might lead to other real words. Upon auditory inspection of the actual Bulgarian accented words, it was judged that there was no risk of this type of misinterpretation as the speakers produced the difficult vowels unambiguously.

The non-words were 100 monosyllabic tokens with a comparable phoneme number to the real words. They were chosen from the ARC non-word database (Rastle et al., 2002) with a specification of only including lexically legal bigrams. After the recording, some non-words were removed, such as tokens that could be perceived as real words in Bulgarian or as a Bulgarian-accented English word (e.g., /sɪf/, which might be confused with /sɪv/ in English due to final devoicing, which is a typical feature of Bulgarian pronunciation, and it also means "grey" in Bulgarian). Hence, the non-word list was also narrowed down to 64 tokens. Their final phonological neighbourhood size was 3.93 (SD = 2.82).

7.3.4.2. Stimuli recording

The words were recorded by four female speakers. Two speakers from England who speak Standard British English were recorded in a sound-attenuated booth at Queen Margaret University Edinburgh. Although both speakers were from Yorkshire, only one of them occasionally used Yorkshire-like vowels of the BATH and STRUT lexical sets. Apart from these few instances of regional variation, her speech was considered sufficiently representative of a standard British variety by two UK-native phonetics experts, to be included in the experiment. Two female native speakers of Bulgarian were selected to record the Bulgarian-accented stimuli. They had learned the same Standard British English target variety of English in their teenage years and used it regularly in their professional lives. They were recorded in a sound-attenuated recording studio in Varna, Bulgaria. All four speakers had completed higher education degrees and were working in universities at the time of recording.

The same equipment was used for all recordings made at sampling rate 44.1 kHz. A TASCAM DR-100 recorder was placed on a desk, 20 cm away from the speaker's mouth. The speakers read the words from a list twice, while seated. They were instructed to have a two second break between reading each word. It was assumed that would minimise list-reading intonation, as exemplified in the pitch contours in Figure 17 to 20, from each of the speakers which were all taken around the same 107 second mark in their respective recordings.

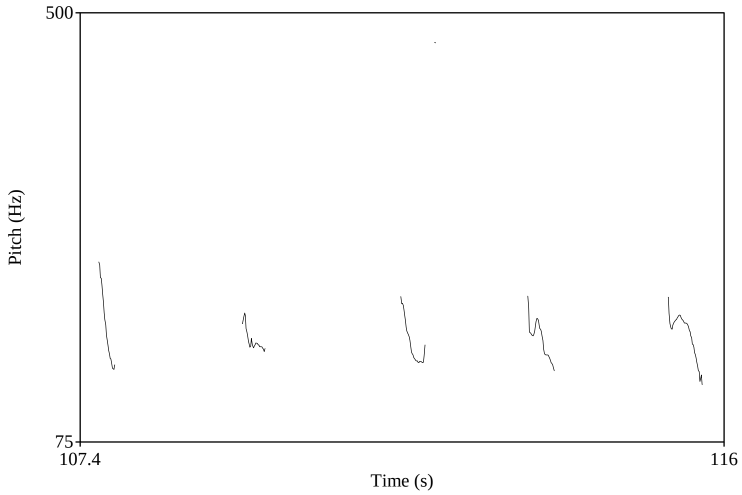


Figure 17. Pitch contours of Speaker Bg1 (L1 Bulgarian) from 107 sec to 116 sec of the recording.

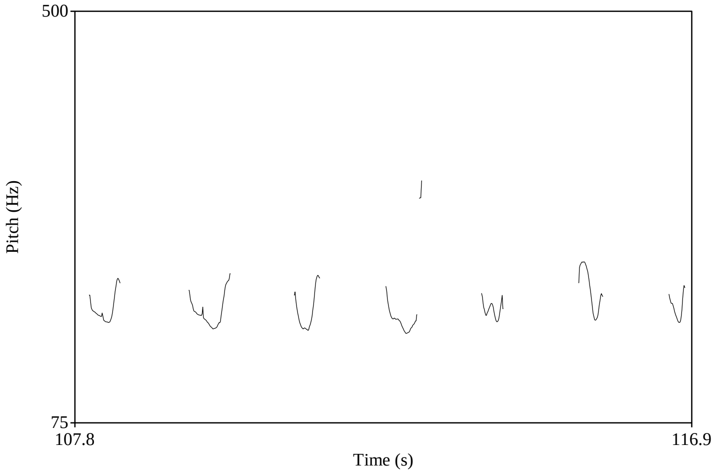


Figure 18. Pitch contours of Speaker Bg2 (L1 Bulgarian) from 108 sec to 117 sec of the recording.

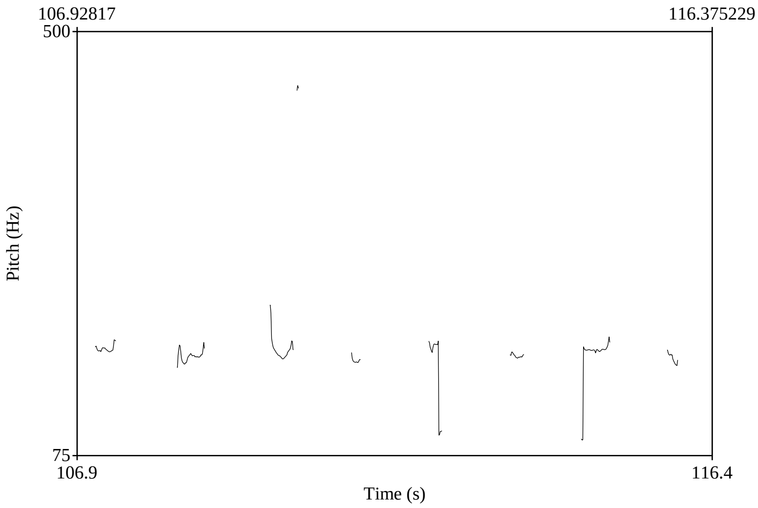


Figure 19. Pitch contours of Speaker En1 (L1 English) from 107 sec to 116 sec of the recording.

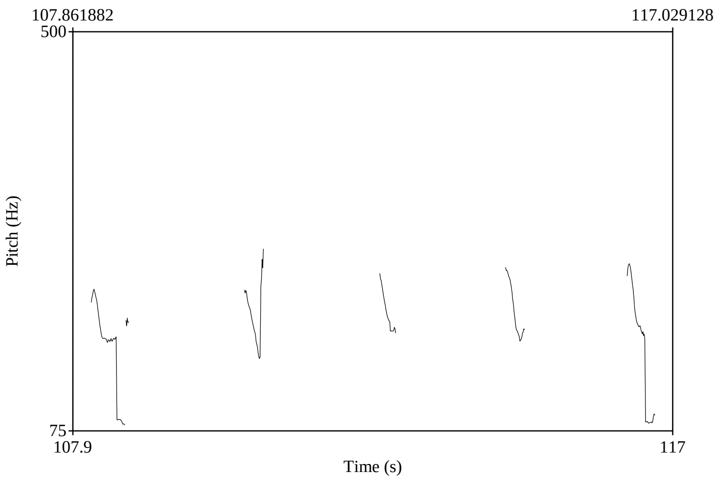


Figure 20. Pitch contours of Speaker En2 (L1 English) from 108 sec to 117 sec of the recording.

7.3.4.3. Stimuli rating procedure

In order to select only the words rated as having the strongest foreign accent, the initial 100 words were rated by native English listeners from the UK. The 100 real-word stimuli were split into four wordlists, containing 25 non-repeating words from each speaker. A rating experiment was designed on the PsyToolkit platform (Stoet, 2010, 2017). The experiment was accessible online through a web browser. In the experiment a listener could play each stimulus from one of the four wordlists, one at a time, in a random order, and rate the strength of foreign accent they perceived. The rating scale was from 0 to 8, where 0 was labelled as "no foreign accent". 0 to 8 was preferred over 1 to 9 because 0 had to indicate the absence of all foreign accent and it was assumed that 0 would represent that more intuitively than 1.

The listeners were instructed to listen to each file no more than twice. In addition, the participants could see the target word written on the screen as they were rating the pronunciation. One of the four wordlists was assigned randomly to the participants at the time of opening the link with the experiment.

After rating the words, the participants filled out a debriefing questionnaire for exploratory purposes. They were asked about general demographic information, the identity of the accents they thought they had perceived, their background in Bulgarian, their own variety of English and their frequency of interaction with non-native speakers of English and Bulgarians in particular.

After the end of the experiment the participants could enter a draw for a 25-pound voucher for online shopping. The rating experiment was advertised via Twitter and Facebook.

7.3.4.4. Raters

The participants in this experiment were over 16 years old, and they had to have grown up in the UK with English as a native language. The technical requirements for participating were access to an internet connection, a keyboard, headphones and a quiet room.

Forty-three participants (27 female) took part in the experiment (mean age = 38.02, SD = 13.36). All the participants had grown up speaking English in the UK and none of them had studied Bulgarian or had a Bulgarian background. On average the listeners spent 16.6% of their time interacting with non-native speakers of English (SD = 23.4) and 2.7% of that time (SD = 8.4) interacting with non-native speakers of English of Bulgarian origin. One of the listeners had studied Russian in their teenage years, but apart from that none of them had any current or past experiences with Slavic languages. The majority of raters reported hearing Eastern European or French accents, and only one suggested a Bulgarian accent (see Figure 21).

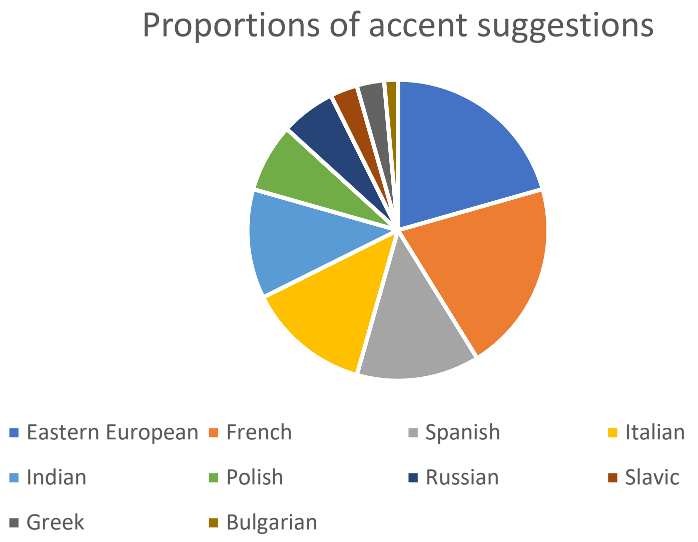


Figure 21. Proportions of non-native English accents that the raters reported hearing.

7.3.4.5. Rating outcomes

A linear mixed effects model was constructed with *Foreign Accent score* as an outcome variable. The model had one predictor, *Speaker*, in which one of the native English speakers (En1) was picked as a baseline level and the scores of the rest were compared to hers because she is the speaker with the occasional Yorkshire realisations and it was expected that some of the raters might have interpreted her less standard variation as a foreign accent. The model included random slopes for *Speaker* by *Participant* and by *Word*. The results of this model are summarised in Table 11.

Predictor	Estimate	*t*-value	*p*-value
Intercept	0.56	4.64	< 0.001
Speaker (EN 2)	-0.28	-2.45	0.02
Speaker (BG 1)	4.37	18.08	< 0.001
Speaker (BG 2)	4.21	16.79	< 0.001

Table 11. Summary of the model on Rating scores per Speaker.

There was a small difference between the two native English speakers, such that the second speaker was rated as having a little lower foreign accent. The two Bulgarian speakers were rated as similarly more foreign accented than the reference native English speaker. Figure 22 summarises their scores.

Words for each accent category (SBE vs. BA) were picked based on the Foreign accent scores (FAS) from the present experiment. For each of the 100 words two average scores were calculated: mean FAS of the native English speakers and mean FAS of the Bulgarian-English bilinguals. Then the difference between the two FAS scores was calculated for each of the 100 words. As the average difference between the En1 speaker and the two Bulgarian speakers rounds to 4 (see Table 11*),* it was decided that only words with a FAS difference of at least 4 will be selected for the listening experiment. There were 64 words with such score, and they were selected for the following experiment (see the full list in Appendix 1.C).

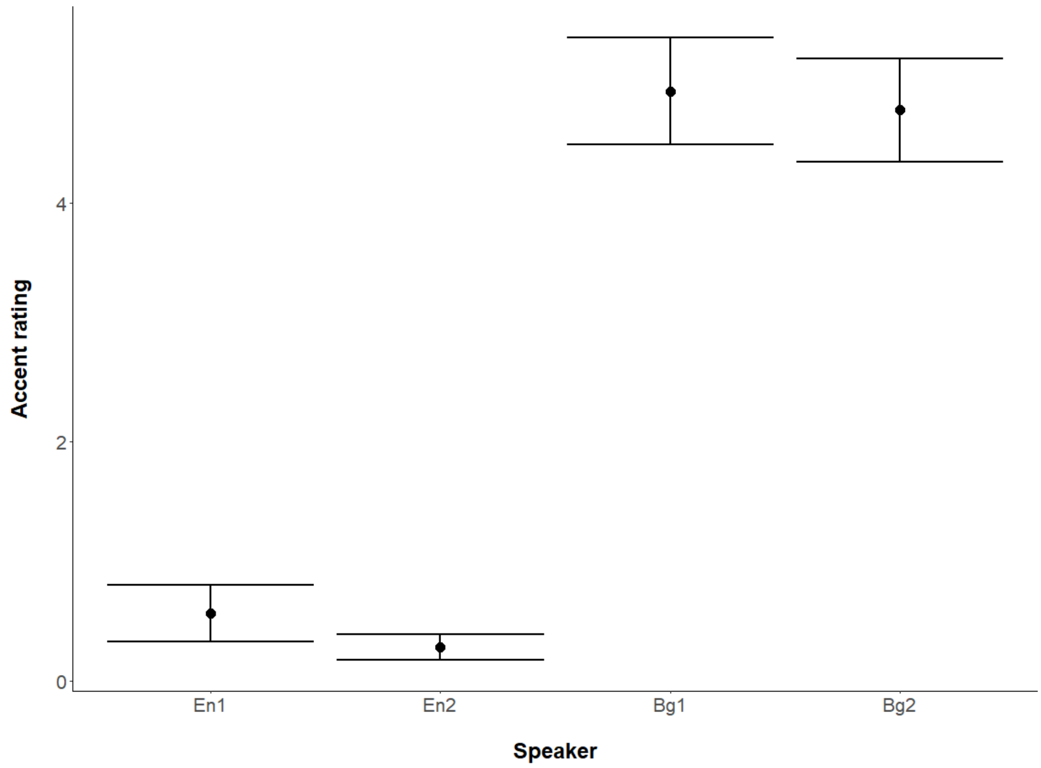


Figure 22. Model estimate and error bars of the foreign accent rating of the two native English speakers and the two L1 Bulgarian - L2 English speakers.

7.3.5. Procedure

Participants were reached online via social media, such as Twitter and Facebook as well as Queen Margaret University's internal email recruitment system. This method for data collection was chosen in order to reach as many participants as possible and improve the power of the experiment and the variability in the participants' proficiency scores. Prior pilot studies had proven that recruiting Bulgarian participants for in-person laboratory-based experiments in the area of Edinburgh and Musselburgh was highly inefficient.

Before participating the listeners were informed about data storage and their right to withdraw from the experiment at any time. After providing informed consent the participants were given written and visual instructions for the auditory lexical decision task. They proceeded at their own pace. After a countdown the training trial for the lexical decision task started automatically. The whole experiment was carried out using the online platform PsyToolkit (Stoet, 2010, 2017).

When making their lexical decisions, the participants had to respond by pressing either button 4 or 6 on the keyboard with their index finger. When waiting to hear a word and make a decision, the participants were instructed to rest their finger over button 5. These particular buttons were picked because it was anticipated that there might be differences in the layout across the participants' keyboards. Buttons 4 and 6 are consistently close to each other across the most common Bulgarian layouts as well as the English (UK) and English (United States) layouts. The correspondence to words and non-words for the two buttons were randomised across participants. After hearing each sound-file the participants had 2500 ms to enter their response, after which the following test item was automatically loaded. As soon as they entered a response, or just before the new item was loaded if they entered no response, the word "LOGGED" appeared on the screen, to signify that their response was recorded.

After the training task the participants proceeded with the main experiment, which started after a countdown. The participants heard the 128 trials of words and non-words without a break. With a maximum delay for each answer set at 2.5 seconds, the whole task was expected to take about 5 minutes and 20 seconds at most.

The auditory lexical decision task was followed by the proficiency test LexTale (Lemhöfer and Broersma, 2012a). The participants saw a word or non-word displayed on screen in capital letters. Using their mouse or touch pad, they had to click on a green button saying "YES" if they thought the item was a real word or on a red one saying "NO" if it was a non-word. Their responses were not timed.

Lastly, the participants filled out a general questionnaire collecting demographic and language background data for exploratory purposes (see Appendix 2.B). A summary of the questionnaire results is represented in the following section.

On average the whole study was completed in 18 minutes (*SD* = 14.7).

7.4. Results

7.4.1. Overall reaction times, linear analysis

One of the main questions addressed in this experiment is whether the Bulgarian accent of the stimuli will facilitate the speed of recognition of English words for Bulgarian (L1) – English (L2) bilinguals, particularly for participants with low English proficiency.

A linear mixed effects analysis was performed in order to find the effect of the listeners' English proficiency and the stimuli's accent on the overall reaction times of the listeners within the experiment. The results suggest that overall listeners were slower with Bulgarian accented stimuli than with native English stimuli and that with increased proficiency the Bulgarian accented stimuli were recognised more slowly compared to the SBE stimuli.

The linear mixed effects regression analysis had three predictors: *Proficiency* (the LexTale score centred around the mean), *Accent* (Standard British English as a baseline, and Bulgarian accent) and their interaction. The outcome variable was *Reaction times* in ms centred around the mean. The model had a random slope of *Participant* by *Accent* and a random intercept for each *Soundfile*. As each soundfile is a unique combination of a word (e.g., “link”) and one of the four speakers, including random intercepts per soundfile accounts for variation caused by both word and speaker factors. This approach was taken in all following statistical models.

Only correct real-word responses between 150 ms and 2500 ms were included. The analysis included 5378 observations based on 256 sound files and 94 participants. The reaction times were not log transformed because it was judged that the interpretation of the results would be more transparent using the raw milliseconds rather than log transformations. In addition, the RT data were normally distributed.

The detailed results are presented in Table 12. They suggest that there is no effect of Proficiency, however, there is an effect of Accent. This means that words with an SBE accent are overall recognised faster than words with a Bulgarian accent. There is a significant interaction between Proficiency and Accent. Figure 23 shows that listeners with higher English proficiency have slower reaction times when hearing the words produced with Bulgarian accent than an SBE accent. From the plot it appears that there is no systematic accent preference for the lower-proficiency listeners. Each point on the plot represents one data point.

Predictor	Estimate	*t*-value	*p*-value
Intercept	-25.34	-1.62	0.11
Proficiency	-1.31	-1.43	0.16
Accent (Bulgarian)	67.74	3.97	< 0.001
Proficiency by Accent (Bulgarian)	1.98	2.76	0.007

Table 12. Summary statistics for the linear analysis of the Proficiency and Accent effects on the Reaction times.

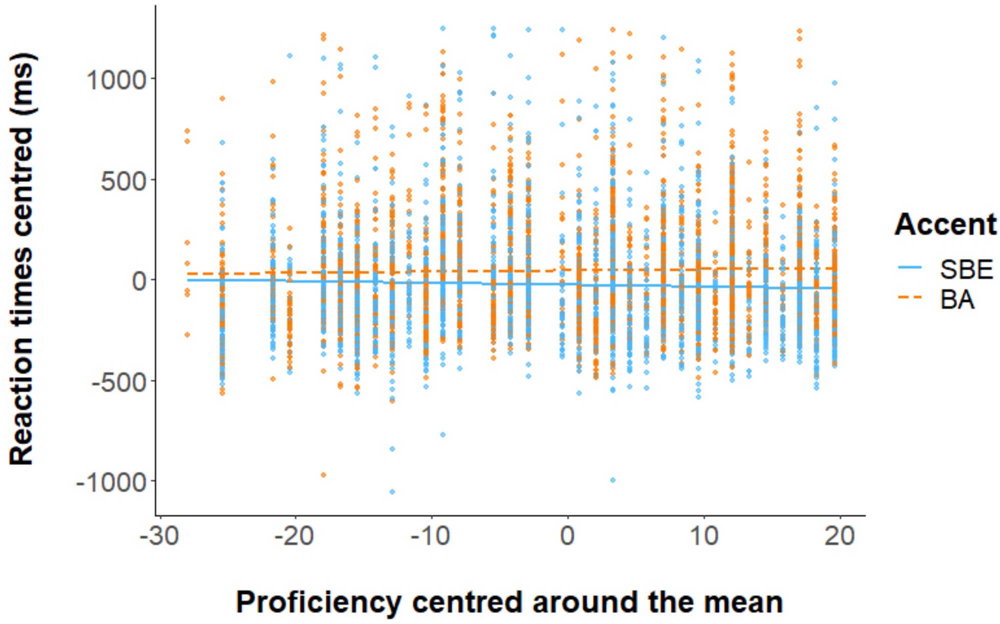


Figure 23. Model prediction for the centred reaction times to Bulgarian- and English-accented words. The x-axis shows the centred proficiency scores and the y-axis shows the centred reaction times. Each point represents one observation. The two accents are Standard British English (SBE) and Bulgarian accented English (BA).

7.4.2. Overall accuracy, linear analysis

This section aims to answer the question whether Bulgarian accented English facilitates the accuracy of word recognition of BA stimuli compared to SBE stimuli by Bulgarian (L1) – English (L2) bilinguals with different proficiencies. The accuracy results are similar to the reaction time results from the previous section. There was an overall lower accuracy rate with Bulgarian accented words than English accented words (see Table 13 and Table 14). In addition, as the listeners' English proficiency increased, their accuracy with Bulgarian-accented English declined compared to native English.

Accent	Number correct
Bulgarian accent (BA)	2645 (91%)
Standard British English (SBE)	2733 (94%)

Table 13. Overall accuracy results on real word stimuli per accent.

In order to investigate the effects of Proficiency and Accent on the accuracy of word recognition in L1 Bulgarian L2 English listeners, a binomial generalised linear mixed effect model was tested. The analysis included correct and incorrect answers to real word stimuli, that received responses within 150 ms and 2500 ms.

The model included the interaction between *Proficiency* (the LexTale scores centred around the mean) and *Accent* (SBE as baseline, compared to Bulgarian). In addition, the model had random intercepts per *Participant* and *Soundfile*. There were 5838 observations, 94 participants and 256 soundfiles considered in this model. The results of the model are summarised in Table 14. The outcome variable was coded with 0 (correct) and 1 (incorrect), hence estimates in the negative direction suggest an increased number of correct answers.

Predictor	Estimate	z-value	p-value
Intercept	-3.30	-22.58	<0.001
Proficiency	-0.006	-0.99	0.32
Accent (Bulgarian)	0.34	2.21	0.03
Proficiency : Accent (Bulgarian)	0.02	2.78	0.005

Table 14. Summary of the generalised linear mixed model on the listeners' Accuracy.

There was no overall significant effect for Proficiency. However, there was a significant effect of Accent, such that words produced in Bulgarian-accented English accent were recognised incorrectly more often than words in Standard British English. There was also a significant interaction between Proficiency and Accent, such that with increased proficiency there was increased accuracy for English-accented words.

Figure 24 shows that listeners with higher proficiency in English had higher accuracy for SBE words and lower accuracy for BA words. At the bottom end of the proficiency spectrum the accuracy of the two accents completely overlaps, similarly to the reaction times findings. There is also a small tendency for the opposite direction, a preference for Bulgarian-accented words by lower proficiency listeners, which was not observed in the reaction times analysis. The standard error area is very wide because the plotting function does not reflect the random effects structure of the model, hence these are very conservative estimates.

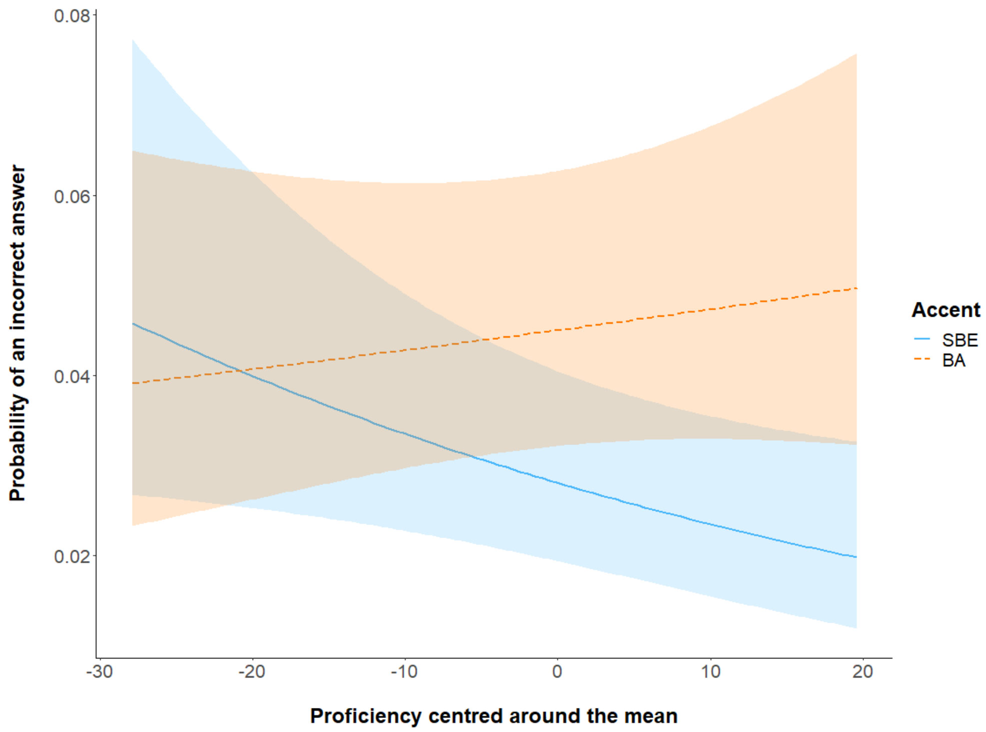


Figure 24. Modelled interaction between Proficiency and Accent on the outcome variable Accuracy. The x-axis shows the Proficiency score centred around the mean. The shaded area reflects the standard error for the main effects, not taking the random slopes into account. The two accents are Standard British English (SBE) and Bulgarian accented English (BA).

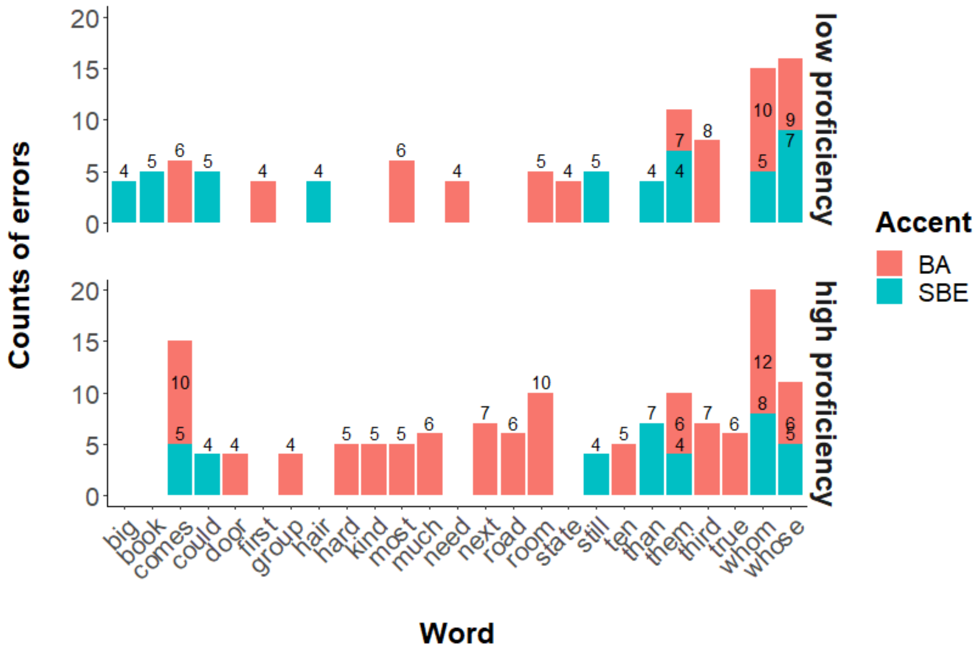


Figure 25. Exploration of the counts of errors per word of listeners with lower and higher proficiency (groups split at the median LexTale score 81.25%). L1 of the speakers is signified by the colour. The plot includes only words with four or more errors. The two accents are Standard British English (SBE) and Bulgarian accented English (BA).

The exploratory plot in Figure 25 suggests that the listeners with higher than the median English proficiency tended to produce more errors on Bulgarian-accented words (*N* = 117) than listeners with English proficiency below the median (*N* = 63) (for more detail see Appendix 4.B). By comparison, the higher proficiency group had only 33 errors with native English speech, and the lower proficiency group had 43 errors. Additional statistical exploration using groups of phones as predictors was not possible as none of the models converged. However, based on observations of Figure 25 it may be speculated that words with initial /θ ð h ɹ/ or words containing /ʉ/ led to more errors. These phones involve articulator configurations that are less familiar for L1 Bulgarian listeners. Hence, their representations of these sounds might be less well established.

However, the words starting with /θ ð h/ are also function words. Despite their high frequency in general, they are not the majority class of words among the experimental stimuli and are less imageable. It is possible that these factors, in addition to the words' phonetic realisation also affected the participants' accuracy. The exploratory plot of RT per word in Appendix 4.A also shows that the correct responses to “whom”, “whose”, “third” and “them” but not “than” were among the words with the slowest response times. This suggests that the difficulty cannot be simply explained by word class. Further research can investigate this by systematically controlling the specific phonetic makeup of the test items.

7.4.3. Reaction time adaptation within a block, curve analysis

This section investigates the effect of English proficiency and Accent on the participants' short-term reaction time adaptation to the two different accents. It was predicted that all listeners would adapt to the accents and reduce their reaction times when hearing a new speaker of an accent but at different speeds. Adaptation here means, achieving faster reaction times than the initial one and reaching at least the overall mean reaction times. It was predicted that the low proficiency listeners would adapt faster for Bulgarian accented English than Standard British English. High proficiency listeners were expected to adapt to both accents at similar rates.

The results only partially supported the predictions. The reaction times of the listeners with highest proficiency in English changed in a way that would be expected of listeners who are highly specialised to SBE. The most unexpected result was that some listeners appeared to need no adaptation time at the start of a block. They started faster than the global average and slowed down with the progression of the block, while the listeners at the opposite side of the proficiency continuum started slower than the average and increased their speed throughout the block.

In order to test the hypotheses, a generalised additive mixed models (GAMM) was created. A GAMM analysis allows for the investigation of both linear and non-linear relationships between the predictors through their inclusion as parametric (linear) and smooth (non-linear) terms in the model. The linear terms test similar hypotheses to those presented in Section 7.4.1, while the smooth terms test if the outcome variable is affected non-linearly by one or more continuous variables. A significant smooth term (also called a smooth) suggests that the outcome variable changes in a non-linear fashion along a continuous predictor. Often the main continuous predictor is *Time* or a proxy for *Time*, as is in this case with the use of *Within-block trial number*. Hence conceptually, a smooth term resembles an interaction between the predictor of interest and a continuous variable (here: *Within-block trial number*). In addition, like the mixed effects models used so far, this type of analysis also allows the use of random structures (here: random smooths) to account for the fact that multiple RT data-points came from the same participants and that multiple participants were presented with the same soundfiles. A random smooth accounts for the effect of non-linear but systematic variation from the model. This section focuses on the non-linear relationship between the continuous predictors *Proficiency,* and *Within-block trial number* and their interaction with the two *Accents* (SBE and BA).

As the research question focuses on adaptation to a new speaker from an unexpected accent, the analysis includes only responses to the first speaker within an accent-block. Of them, only correct responses between 150 ms and 2500 ms to real words were included in the analysis. The reaction times to the words were centred around their mean.

The model included a parametric term for *Accent* (SBE vs. BA) a smooth term for the token number *Within-block,* a smooth term for *Proficiency,* an interaction smooth for *Within-block* number by *Accent* with k = 10 and an interaction between *Within-block* number and *Proficiency*. The variable k (knots) is a specification of the model, which is related to the degrees of freedom for each predictor and sets the upper limit of base functions that the model can employ to represent the curving of the outcome variable (Sóskuthy, 2017). Hence k specifies how "curvy" the model can be. In a GAMM context its role is limited by an in-built smoothing parameter which automatically picks the necessary number of base functions (Sóskuthy, 2017).

In addition, the model included random smooths for *Within-block* number per *Trajectory* (where *Trajectory* is the adaptation trajectory for one participant for one accent block, allowing individual variation at each point within a block) and a random smooth for *Within-block* by *Participant* (allowing individual non-linear variation per participant as there are two trajectories per participant).

Smooth terms	Edf	*F*	*p*-value
Within-block	1.00	0.34	0.56
Proficiency	2.59	1.75	0.118
Within-block by Accent (Bulgarian)	5.59	6.67	< 0.001
Within-block by Proficiency	1.00	6.49	0.01
Within-block by Proficiency and Accent	1.00	4.39	0.04
Random smooth: Within-block per trajectory	52.51	0.14	< 0.001
Random smooth: Within-block per participant	71.55	0.57	< 0.001

Table 15. Summary statistics for the smooth and random terms of the full GAM model.

The results of the smooth terms are relevant for the research questions posed in this section and they are summarised in Table 15. There was no effect of *Proficiency* and the *Within-block* smooth term. There was a significant interaction involving *Proficiency* and *Within-block* number, which means that the RTs of people with different proficiencies did not change at the same rate as the accent block progressed. There was a significant non-linear interaction between *Within-block number* and *Accent,* which means that the listeners adapted to the two accents at different rates as the block progressed. Lastly, there was a significant triple interaction between *Proficiency, Accent* and *Within-block number,* which means that as the block progressed listeners with different levels of English proficiency adapted differently to the two accents.

In order to interpret the results, three plots are provided, demonstrating how fast the listeners responded to the two accents at the first 16 tokens within an Accent block (i.e., the first speaker of the accent block). Figure 26 splits the participants into two proficiency groups using the median LexTale score 81.25 of the group as a boundary. Listeners with a proficiency equal to or higher than the median were included in the high proficiency group and listeners with a score below that number were included in the low proficiency group. This division was created just for the purposes of exploring the data and interpreting the statistical results. The reason why the median was chosen as a boundary was to create two participant groups with similar sizes, differing in proficiency. Figure 26 shows the mean and standard deviation per group over the consecutive trial numbers.

Figure 26 shows that responses to the native English tokens received overall faster responses than the Bulgarian-accented words. At the start of the block, the higher proficiency listeners were faster than the mean when responding to English accented words and a lot slower than the mean with Bulgarian-accented words. The lower proficiency listeners tended to be slower than the mean with both accents. By the middle of the Accent block, trial 16, the responses of both listeners groups tended towards and even below the mean. Both listener groups display steep adaptation curves to Bulgarian accented words, which even out between trial 4 and 8, which is about the middle of the speaker block.

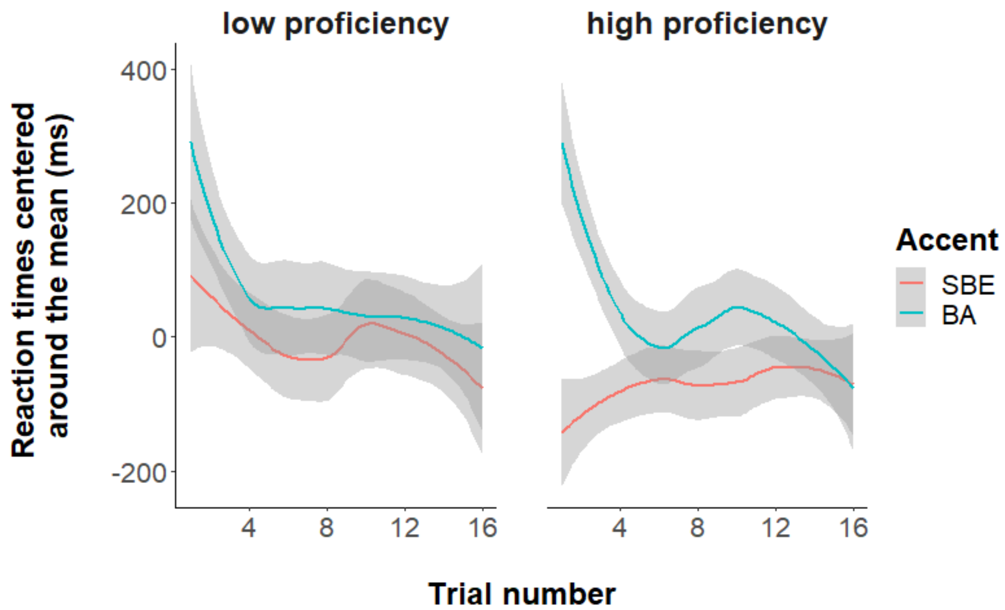


Figure 26. Centred reaction times to Bulgarian- and English-accented words. The x-axis shows the consecutive trial number within a language block and the y-axis shows the reaction times centred around the mean. The left panel shows the average response times of the low English proficiency listeners and the right panel shows the results of the high proficiency group. The two accents are Standard British English (SBE) and Bulgarian accented English (BA).

The heatmaps in Figure 27 and Figure 28 provide a more nuanced look into the triple interaction between Accent, Proficiency (y-axis) and Trial number (x-axis) with the listeners' reaction times, pictured as gradient colours. The rection times in this plot are centred around the average RT across all participants for the accent on the plot. By comparison Figure 26 uses the overall mean across all participants and accents. This means that the average RT across all participants for Bulgarian accented words in Figure 27 is 0 and is represented in green. The yellow colour represents RT above the mean (slower responses), and blue is below the mean (faster responses).

Zero on the y axis represents the average LexTale score (i.e., English *Proficiency*) across all participants. The value 20 above zero represents participants who scored 20% higher than the mean and 20 below zero represents participants who scored 20% below the average. Increase on the x-axis represents subsequent trial numbers within an accent block.

The plots suggest that both accents triggered the opposite tendencies for the two ends of the proficiency continuum. In Figure 27 the Bulgarian-accented stimuli yielded faster RT from the low proficiency listeners in the initial trials (the magnitude of the effect was around 100 ms below the mean for this accent) and slower RTs towards the 16^{th} trial of the block (again, by about 100 ms). The higher proficiency listeners started about 50-70 ms slower than the mean and ended about 20-50 ms faster than it.

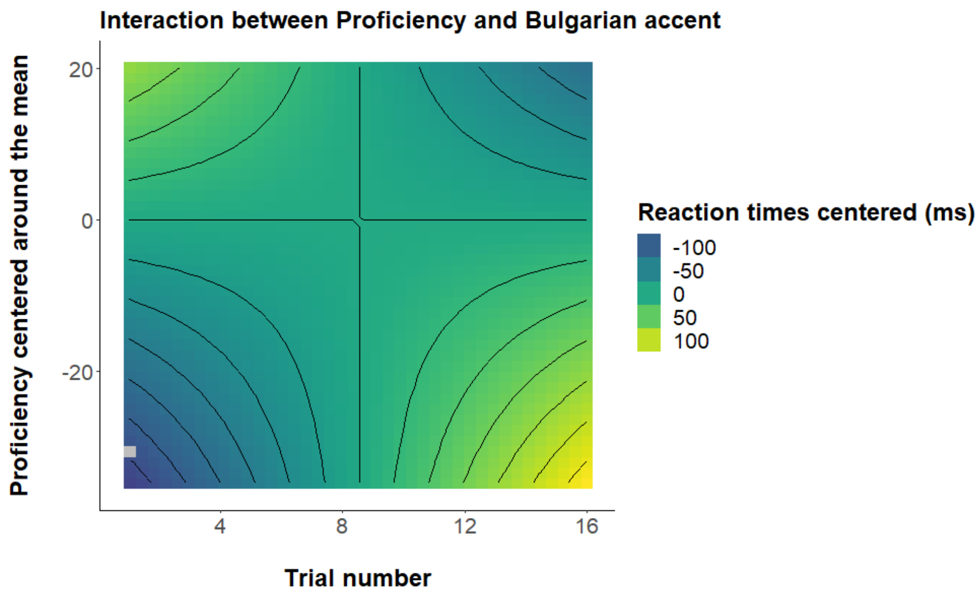


Figure 27. Heatmap of the interaction between the trial number within a language block, the listeners' proficiency centred around the mean and the reaction times (ms) centred around the mean for Bulgarian accented words. Zero on the y axis represents the average LexTale score across all participants. The value 20 above zero represents participants who scored 20% higher than the mean. Increase on the x-axis represents subsequent trial numbers within an accent block.

In the English-accented blocks, in Figure 28 the results are exactly the opposite, with similar effect magnitudes. The higher proficiency listeners were about 50 ms faster than the mean in the beginning and gradually slowed down to about 50 ms below the mean, while the lower-proficiency listeners were initially about 100 ms slower than the mean and about 100 ms faster than the mean.

Both Figure 27 and Figure 28 also illustrate the interaction between Proficiency and Trial number. The higher proficiency listeners have smaller fluctuations of their RTs (100 ms range) than the lower proficiency listeners (200 ms range).

In both heatmaps the intermediate proficiency listeners had a gradual RT pattern with smaller differences between the start and end of the block than the listeners with more extreme proficiencies.

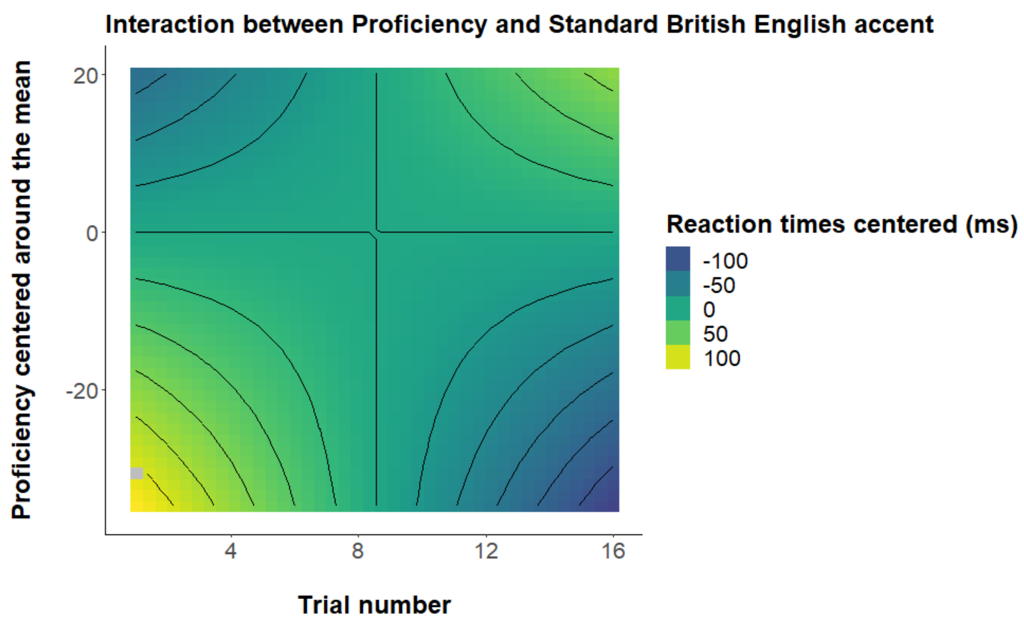


Figure 28. Heatmap of the interaction between the trial number within a language block, the listeners' proficiency centred around the mean and the reaction times (ms) centred around the mean for the native English accent.

7.4.4. Reaction time adaptation across blocks, linear analysis

This analysis aimed to determine if the listeners adapted to the two global accents independently of their adaptation to individual speakers. It was expected that having heard the same accent in the previous block would lead to faster reaction times in the current block for both BA and SBE.

For the purpose, the analysis excluded data from the first speaker-block (see Figure 29). The primary interest was in the interaction between the *Accent* of the speaker (either SBE or BA) and whether the *Preceding* block had the same or different accent. The analysis shows that both predictors had a significant effect on the listeners' reaction times, but their interaction did not.

This suggests that, overall, the listeners benefitted equally from having heard the target accent from a different speaker in both BA and SBE.

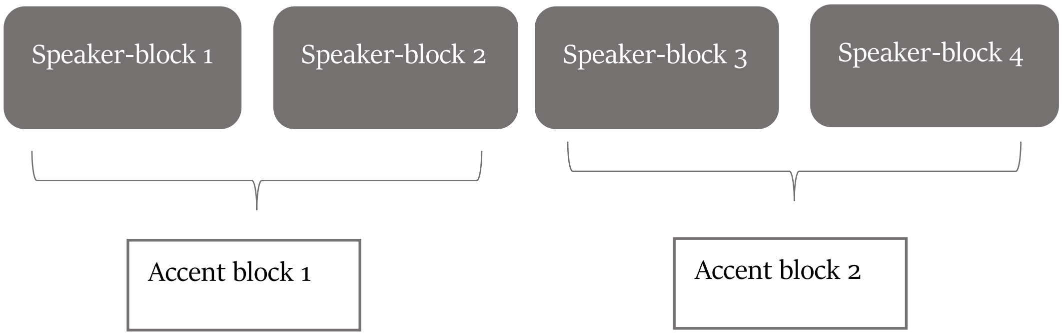


Figure 29. Structure of the auditory experiment.

A linear mixed effects regression model was run to investigate the effect of the *Current Accent* (SBE, BA) and *Preceding Accent* (different, same) on the overall reaction times of the listeners. The predictors were sum coded, which means that Table 16 below reflects the distance between the first level of each predictor from the mean of all means (the intercept), similar to main effects in ANOVA. The outcome variable was Reaction times in ms centred around the mean. The model had a random slope of *Participant* by *Accent* and *Preceding Accent* and a random intercept for each *Soundfile*. As in the previous subsection, the analysis only included correct answers, between 150 ms and 2500 ms from participants who finished the whole experiment. The analysis included 4047 observations based on 256 sound files and 93 participants.

Predictor	Estimate	*t*-value	*p*-value
Intercept	17.17	1.22	0.22
Current accent	-38.21	-3.90	<0.001
Preceding accent	15.50	3.07	0.003
Accent : Preceding accent	-5.11	-0.99	0.33

Table 16. Summary statistics for the linear analysis of the Current and Preceding accent effects on the Reaction times.

The results presented in Table 16 and illustrated in Figure 30 suggest that, as demonstrated in the previous section, overall the SBE words led to faster reaction times than Bulgarian-accented words. In addition, there was a significant increase in reaction times when a Bulgarian-accented speaker was preceded by an English-accented speaker (a different condition), compared to when they were preceded by another Bulgarian-accented speaker. The interaction was not significant, suggesting that there was a similar amount of increase in reaction times if the preceding accent did not match the current accent, regardless of whether the new accent was English or Bulgarian. This means that the listeners had a smaller cost of processing a same-accent speaker than a different-accent speaker, regardless of the actual accent. Despite this lack of significant interaction, Figure 30 shows a tendency for a larger gain for having heard the same accent before a BA block than having heard the same accent before an SBE block.

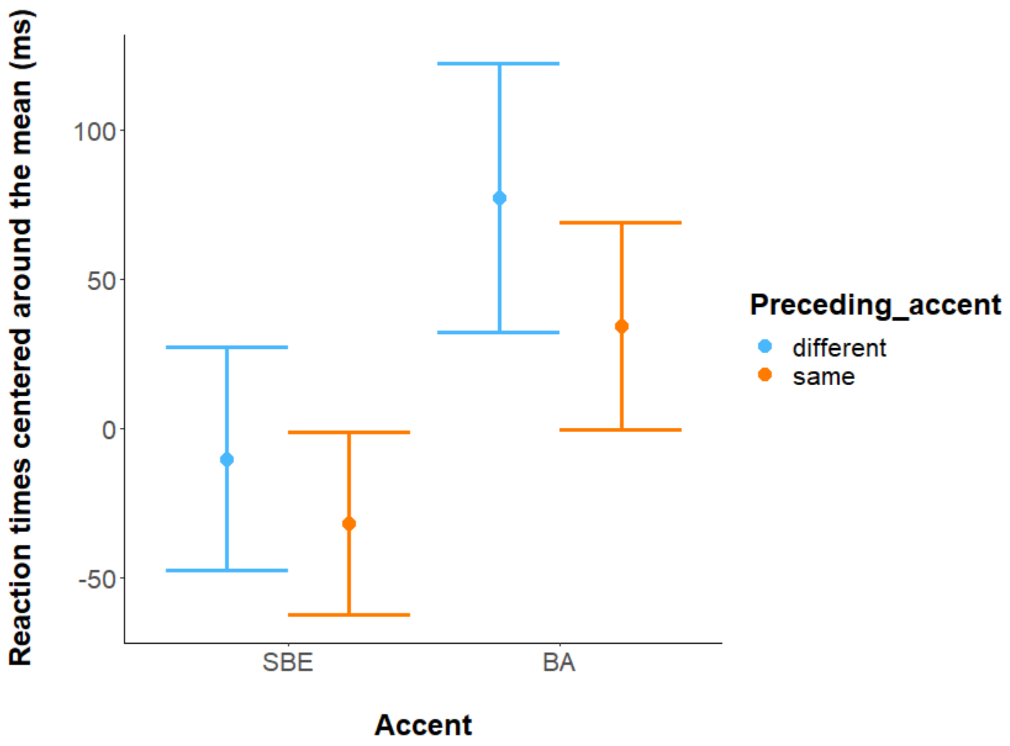


Figure 30. Model prediction for the centred reaction times to Bulgarian- and Standard British English-accented words depending on the preceding accent (estimates and standard error).

7.4.5. Accuracy adaptation across blocks, linear analysis

Similarly to the previous section, it was investigated whether listening to a speaker with the same accent as the speaker from the preceding block improved the listeners' accuracy scores. Again, it was expected that the accuracy would be higher for both BA and SBE if the listeners had just heard the same accent in the previous block. This time the analysis yielded different results, suggesting that having heard the same accent before was important only for the Bulgarian accent trials. In the native English blocks, accuracy was at ceiling level regardless of what the previous accent was.

A binomial generalised linear mixed effects model was performed using the predictors *Accent* (SBE, BA) and *Preceding Accent* (different, same) and their interaction. The outcome variable was coded with 0 for correct answers and 1 for incorrect answers. The model included random intercepts per *Soundfile* and random slopes per *Participant* by *Accent* and *Preceding Accent*. The analysis was based on 4388 observations, 256 soundfiles and 93 participants.

Predictor	Estimate	*t*-value	*p*-value
Intercept	-3.30	-20.12	< 0.001
Preceding accent	0.18	2.31	0.02
Accent (English)	-0.35	-2.71	0.007
Preceding accent : Accent (Different : English)	-0.21	-2.90	0.004

Table 17. Summary statistics for the linear analysis of the Current and Preceding accent effects on the error rates.

The results, presented in Table 17 and illustrated in Figure 31 suggest that overall the English-accented words led to lower error rate. There is a small significant main effect of Preceding accent, although Figure 31, presenting the model estimates and standard errors, suggests that the effect is probably entirely driven by the high inaccuracy probability of Bulgarian-accented words when they have been preceded by a different accent.

In addition, there was a significant increase in errors when a Bulgarian-accented speaker was preceded by an English-accented speaker (a different accent), compared to when they were preceded by another Bulgarian-accented speaker. The interaction was significant, suggesting that having heard a Bulgarian speaker prior to the target block led to better accuracy scores, while for the English-accented target block it did not matter what the preceding accent was. Figure 31 suggests that English-accented words led to ceiling accuracy results.

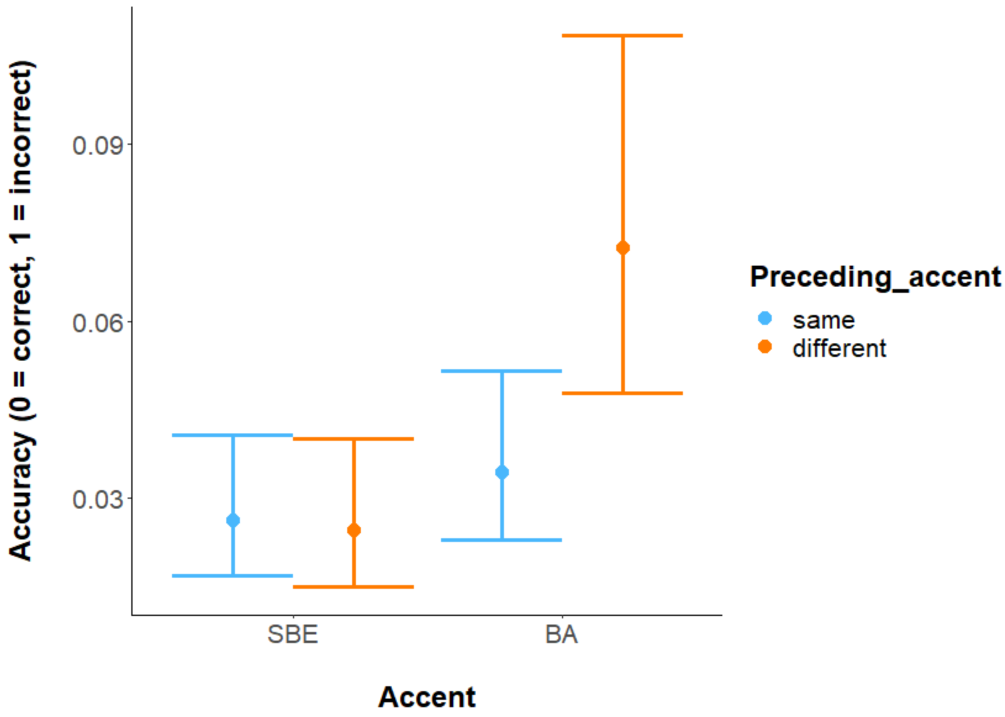


Figure 31. Effect and CI of the model predicting listener accuracy through the Accent of the block and the Preceding accent. The CI reflect the variance of the fixed effects excluding the random factors. Accuracy improves as it tends towards 0 and it deteriorates as it approaches 1. The two accents are Standard British English (SBE) and Bulgarian accented English (BA).

7.5. Discussion

This study set out to determine how the listeners' proficiency interacts with the Bulgarian-English bilinguals' overall speed and accuracy of processing Bulgarian- and English-accented speech, and also with their ability to adapt to these two varieties in the short-term. The complete in-depth discussion of the results can be found in in Chapter 9.

The results of this experiment show that the listeners processed native Standard British English speech (e.g., [hɛə] for "hair") faster than Bulgarian accented speech (e.g., [xɛə] for "hair"), although that was not observed for the listeners with the lowest English proficiency in the sample. There was a similar interaction in terms of the listeners' accuracy. Listeners with higher proficiency in English processed native English speech more accurately than lower proficiency listeners and they processed Bulgarian accented speech less accurately than lower proficiency listeners.

Firstly, these results suggest that accuracy and speed of processing were generally aligned. Secondly, unlike what has been reported previously by Ludwig and Mora (2017), Hayes-Harb et al. (2008), and Experiment 1, the lowest proficiency listeners had no clear benefit of processing

Bulgarian-accented speech over native speech, although they did appear to be faster than the listeners with the highest English proficiency. It is the high proficiency listeners that had clear accent preferences in their reaction times and accuracy. In that sense, L1-accented L2 speech leads to no benefits for word processing, when compared to native speech realisations. There is L1-accent benefit only when lower proficiency listeners are compared to higher proficiency listeners (i.e., benefit for listeners, not benefit for talkers, in the terminology of ISIB).

The results suggest that the lowest proficiency listeners have no particular accent specialisation for speed of processing and accuracy. As the participants' proficiency increases, a trade-off appears between Bulgarian-accented English and native English speech. It appears that high proficiency Bulgarian-English bilinguals living in the UK do not accumulate listening expertise in Bulgarian-accented English and Standard British English but instead, as predicted by ASP (Strange, 2011), and SLM (Flege, 1995), end up adapting to a native-like variety. BIMOLA (Léwy and Grosjean, 2008) fails to predict any possibility for accent specialisation over time.

There is also evidence of accent adaptation across speakers. The listeners had faster RTs after a block of a same-accented speaker, as opposed to after a block of a different-accented speaker. The size of the adaptation was similar for Bulgarian-accented and English-accented words. However, in terms of accuracy the adaptation was only present for the Bulgarian-accented words. The reason for that might be the ceiling levels of accuracy for the English-accented blocks. These results suggest that the listeners found the two Bulgarian speakers and the two English speakers more similar to each other respectively, as opposed to across accent groups and that they benefitted from these similarities for subsequent experimental blocks.

In addition, the listeners showed rapid within-speaker adaptation at the first few instances of hearing a new accent (cf., Clarke and Garrett, 2004). At initial trials the high proficiency listeners responded faster than the general mean for SBE words and by trial 8 out of 16 their RTs were close to the general mean. The lower proficiency listeners started slower than the general mean RTs but they also approached the mean by mid-speaker block (see Figure 28). The results were reversed for Bulgarian-accented words, showing that initially the low English proficiency listeners were faster at Bulgarian-accented words than high proficiency listeners (see Figure 27). These within-block adaptation results suggest that the higher proficiency listeners are better adapted to a native English accent (e.g., [hɛə]), while the lower proficiency listeners are better adapted to a Bulgarian accent (e.g., [xɛə]). That, however, does not prevent them from achieving a short-term adaptation in both accents.

An unexpected result is illustrated in Figure 26, which shows that the higher proficiency listeners were faster than the overall mean at the start of processing native English speech and gradually slowed down. In addition, Figure 27 and Figure 28 show that the listeners who started out the fastest finished the block slowly. Initially, it was predicted that all listeners are going to be slower at initial trials, while they adjust to the new speaker, however, these plots contradict the prediction.

One possibility for this result is that the high proficiency listeners had a strong specialisation for native Standard British English (which is consistent with their results on RTs and accuracy), which means that at the start of a new speaker block, when they were more alert and attentive, the input perfectly matched their implicit expectations and they responded faster. Similarly, for the lower proficiency listeners when first exposed to Bulgarian-accented English. As these expert listeners (high proficiency with SBE and low proficiency with BA) continued with the task, their reaction times could have slowed down due to allocating fewer attentional resources to the task they perceived as easy at first. By contrast, the listeners who initially responded slowly to a given accent, continued to reduce their RT until the end of the speaker-block as they continued adapting, retaining full engagement with the task. This line of argumentation is theoretically justified using ASP and IAF in Chapter 9.

However, there is little research on auditory accent adaptation that reports data on continuous reaction time adaptation to different accents or speakers (as opposed to accuracy results or overall block comparisons). One exception is experiment 1 in Floccia et al. (2009), which does not completely match the pattern observed in the present experiment. In Floccia et al. (2009) the expert listeners from South West England maintained RTs below the RP baseline for speech with their familiar Plymouth accent in Block 2. Towards the end of the block they had a tendency for additional RT improvement with that accent. This is unlike the present study where the expert participants slowed down at the end of the block. By contrast the Floccia et al. (2009) responses to French and Irish-accented speech did not show consistent adaptation patterns. At the end of the task they were slower than or close to the RP Baseline, respectively.

While these results are inconsistent with the observations drawn from the current study, it is important to point out that Floccia et al. (2009) included the lexical decision targets in sentences with up to two levels of embedding. It can be argued that the single, high-frequency word processing used in this experiment poses lower demands on attention than processing sentences like: “In the evening Virgil and Thomas usually complain about their tummy” or “Eric shouted very loud when he saw that we had broken the present” (Floccia et al., 2009, p. 383). Hence, it is still possible that the participants in the present experiment lost some of their engagement with the task when finding it easier at first, compared to the participants in Floccia et al. (2009) who had to pay attention throughout in order to predict the end of the sentence. Lastly, the first response to the Plymouth accent in Floccia et al. (2009) was also faster than the baseline compared to some of the later responses, which matches the observation for the expert listeners paying attention in the present experiment.

Finally, the exploration of the listeners’ errors in Figure 25 requires an explanation. This systematicity is reminiscent of the predictions of PAM (Best and Tyler, 2007) and the Vocabulary model (Bundgaard-Nielsen et al., 2011). Specifically, the Vocabulary model predicts that as listeners expand their vocabulary in a native-speaker environment they would develop systematic correspondences between L1 and L2 (like the two-category assimilation, described in PAM), which would lead to improved discrimination ability in L2. However, in the present experiment, it appears that the systematicity of correspondences is associated with processing errors when encountering Bulgarian-accented speech. For example, a lower proficiency listener would easily interpret both [bed] and [bad] as “bad” but they would also interpret [bed] as “bed”. However, a higher proficiency listener would have established a more narrow representation for the [æ] sound, where [bed] is mostly associated with “bed” and [bad] as “bad” (supported by Broersma, 2012). This is a very simplistic illustration of how the lower proficiency listeners could be described as more flexible, and the higher proficiency listeners as narrower in the terms of IAF (Kleinschmidt and Jaeger, 2015a), which is to a certain extent related with the Vocabulary model (Bundgaard-Nielsen et al., 2011). This illustration is supported by the error exploration, as well as the RT and accuracy results in this experiment.

7.6. Limitations

There are several limitations of this study, that could also be seen as opportunities for future development of the topic on L1-accented speech processing.

Firstly, as it was expect that both Bulgarian-accented English and native UK English might lead to benefits for Bulgarian-English bilinguals in the UK (particularly those of high English proficiency), it would have been pertinent to include a completely unfamiliar English accent as a baseline. The challenge of doing that was the wide target population of this experiment. In attempting to reach a high number of UK-based Bulgarian listeners, it was expected that most of them would be multilingual and living in a multicultural environment. Hence, it was difficult to pick an existing accent of English that would be similarly unfamiliar to all potential participants. As suggested by some recent research, the effects of incidental learning on speech processing should not be overlooked (Chang, 2019). A more controlled follow-up could set stricter selection criteria for the participants and add an extra level to the Accent variable.

This experiment could have also benefitted from a robust measure of listener exposure to Bulgarian-accented English. At the time when the experiment was being prepared, self-report was the most efficient option available. The informal feedback from the participants suggested that the self-reporting question was not worded very transparently (even though a Bulgarian translation was also provided). The two relevant questions were: “On a WEEKLY basis, what percentage of your time talking to people is spent speaking in English with *non-native* speakers of English? If you speak only in English and only with non-native speakers of English, it's 100%. If you never speak in English with non-native speakers of English it's 0%. If you don't fall in these extremes, estimate the most accurate intermediate answer. What percentage of those interactions in English include speakers whose native language is Bulgarian?”. This complaint precluded the measure from being used in the statistical analysis. A more controlled study could also collect a sample of spontaneous speech from the participants and assess if their foreign-accentedness scores correlate with their proficiency scores.

Lastly, one of the main ways this experiment could have been improved is by controlling more strictly the acoustic characteristics of the stimuli. As the goal here was to find the effects of the holistic Bulgarian and English accents, rather than individual correlates, this step was not taken.

However, this decision also limits the scope of the conclusions that can be drawn from the study. For example, for now it can only be speculated whether it is the presence of L1-specific cues that led to the difficulty of the high proficiency listeners, as opposed to a greater phonetic variability in the Bulgarian-accented stimuli compared to the native English realisations.

7.7. Conclusion

Overall, the results of this study support the idea that Bulgarian-English bilinguals in the UK are not a homogenous group in terms of their English comprehension abilities. The listeners' proficiency in English affected their speed of processing, accuracy and within-block adaptation. Higher-proficiency listeners appeared to be more specialised in native English speech, than Bulgarian-accented speech. At the same time, lower-proficiency listeners were faster to process Bulgarian-accented speech than the higher-proficiency listeners, although they had no specific processing benefit with it than native English words. This is the most unexpected discovery of the experiment, which contradicts the predictions laid out in Section 8.1 and the results of Ludwig and Mora (2017) and the predictions of ISIB (Bent and Bradlow, 2003). It was expected that listeners with higher proficiency in English would have overall greater experience with the different accents in English, including Bulgarian accent over the lower proficiency listeners and that they would process Bulgarian accented English as quickly and accurately as native English speech.

Theories of speech processing need to take into account the fact that as listeners gain more experience with a language in an environment, where the native prestige variety dominates, they specialise their perception tp that accent and do not necessarily accumulate balanced competencies in multiple accents of their L2. However, before challenging ISIB, it needs to be verified whether the listeners with the highest proficiency in the sample still did not have some advantage of processing Bulgarian-accented speech, at least compared to native English listeners. The next chapter presents a follow up which directly compares the processing speed and accuracy of the bilinguals with the highest proficiency in English to native English listeners who have been raised and reside in the UK.

Chapter 8 Follow-up Experiment - Processing of Bulgarian and English accented words by native English listeners

8.1. Background and predictions

The previous experiment investigated how Bulgarian-English bilinguals process Bulgarian-accented English and Standard British English words out of context. The results suggested that as the listeners increased their proficiency their reaction times and accuracy favoured the Standard British English pronunciations (e.g., [hɛə]) over the Bulgarian-accented words (e.g, [xɛə]). Increased proficiency did not lead to improved processing in both varieties, or just the Bulgarian variety. The results suggest that as the Bulgarian-English bilingual listeners increase their proficiency, they become more native-like in their processing of English. This result contradicts some of the discussion in Chapter 2. Numerous studies reveal that bilinguals' L2 skills, particularly speech processing, are dynamically influenced by L1 phonetics and lexemes (Blumenfeld and Marian, 2007; Lagrou et al., 2011; Marian et al., 2008), which is why their performance in speech processing is not comparable to that of a monolingual listener. However, on the surface this was not observed in the reaction time and accuracy results of an auditory experiment. When compared to the low proficiency bilinguals, the high proficiency listeners performed in a way that would be expected of native English listeners.

In order to explore these results further, the same experimental setup was used to collect data from native English listeners who were raised in the UK and were based in the country at the time of the experiment. Their data were compared to the data of the Bulgarian-English bilinguals who achieved the highest scores on the LexTale proficiency measure (Lemhöfer and Broersma, 2012a). Based on the research discussed in Chapter 2 and Chapter 3 (Bent and Bradlow, 2003; Hayes-Harb et al., 2008; Lee et al., 2012; Ludwig and Mora, 2017) it was predicted that even the highest proficiency Bulgarian-English bilinguals would have an advantage processing Bulgarian-accented words, in both accuracy and reaction times, and that they will adapt to the Bulgarian-accented stimuli more efficiently than the native English speakers.

8.2. Methods

The stimuli, design and procedure in this experiment were the same as the previous experiment. The questionnaire for the follow-up was slightly modified for participants who are native speakers of English raised in the UK and currently residing in their country of origin (see Appendix 2.C). For example, they were not asked about their length of residence in the UK but were asked about their previous experience with Bulgarian.

The participants were reached via social media and the Survey Circle website. Of the 24 participants who opened the experiment, 16 completed all sections and were included in the analysis. The listeners' LexTale score was also measured, similarly to the previous experiment. One participant received a score of 52%, which would correspond to the scores achieved by the lowest proficiency Bulgarian participants in the previous experiment, so their responses were also excluded, leaving 15 participants for analysis (13 female).

The average age of the participants was 30.3 (*SD* = 11). On average the participants completed the experiment in 14 minutes (*SD* = 3). The listeners' Bulgarian exposure score (details on how it was calculated are available in Section 7.3.3) was 0.07% (*SD* = 0.17) and their native English exposure score was 55.1 (*SD* = 32.4). Four participants stated that their accent is British English. The rest were split between London, SSBE, Northern English, Scottish, Welsh and East Midlands. One participant ambiguously stated that they speak Modern English and two chose not to answer. Four of the participants also spoke a language other than English.

In order to directly compare the results of the native English speakers to the high proficiency Bulgarian listeners, the 25 Bulgarian listeners with scores on LexTale over 90% were selected for further analysis (14 female). Their mean age was 28.7 years (*SD* = 9.4). On average they completed the experiment in 15 minutes (*SD* = 5.7). Their self-reported Bulgarian exposure score was 4.7 (*SD* = 8.24) and their English exposure score was 39.7 (*SD* = 35.2). Three participants had experience with a third language, apart from Bulgarian and English.

8.3. Results

8.3.1. Overall reaction times, linear analysis

This section investigates the effect of the listeners' L1 and the accent of the stimuli on the participants' overall reaction times. It was expected that the L1 Bulgarian listeners would have faster reaction times with Bulgarian accented words than L1 English listeners and slower RTs with SBE than L1 English listeners.

For the purpose a linear mixed effects model was prepared with the centred reaction times as an outcome variable. The predictors were the *Listeners' L1* (Bulgarian or English) and the *Accent* (BA or SBE) and their interaction. Both predictors were sum coded. The model also included random slopes of *L1* per *Soundfile* and *Accent* per *Participant*. The analysis included only correct answers to real word stimuli, which amounted to 2330 observations based on 256 soundfiles and 40 participants. The results are summarised in Table 18 and illustrated in Figure 32, which shows the model estimate and standard error.

Predictor	Estimate	*t*-value	*p*-value
Intercept	-67.70	50.36	<0.001
Listeners' L1	-3.62	-0.21	0.84
Accent	63.99	6.66	<0.001
Listeners' L1 by Accent	-8.15	-1.53	0.26

Table 18. Summary statistics on the RT differences between the native and non-native English listeners for Bulgarian-accented and native English stimuli.

The only significant predictor was the Accent of the stimuli L1. Overall, all listeners processed English-accented words slower than the mean compared to Bulgarian-accented words. There was no effect of the listeners' L1. The Standard Error whiskers in Figure 32 below suggests that the two groups tend to differ with respect to variance. The native English listeners demonstrate greater variability than the native Bulgarian listeners. Also, there is a small tendency for Bulgarian listeners to react faster to Bulgarian-accented words compared to native English listeners and slower to English-accented words compared to native English listeners, but these tendencies are very small.

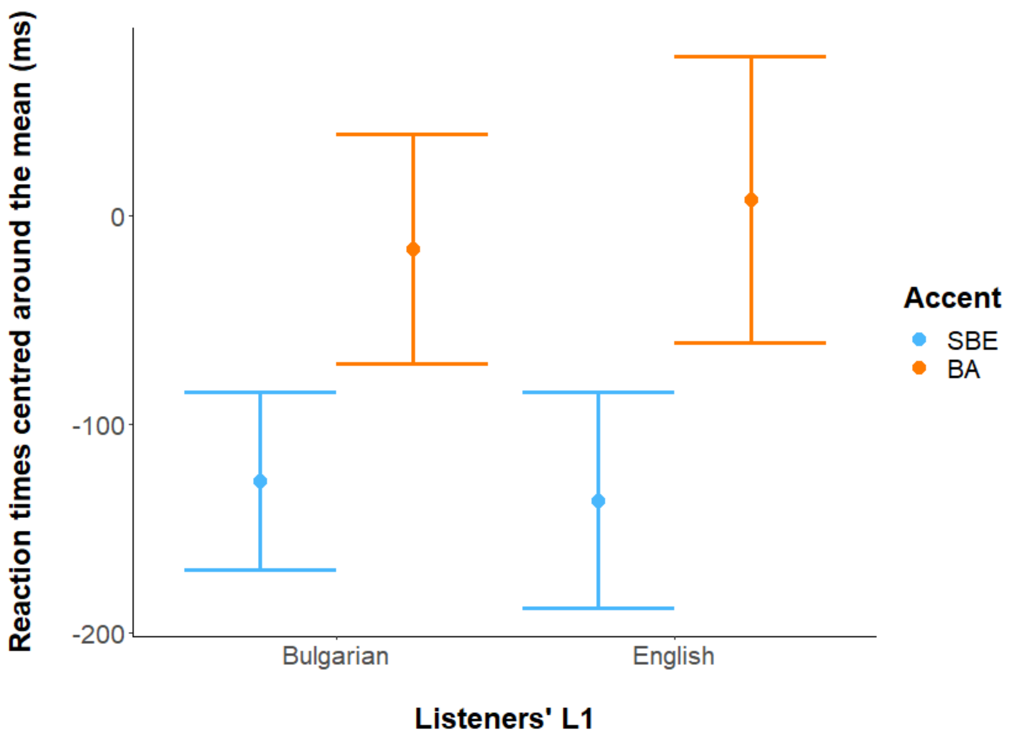


Figure 32. Model prediction for the centred reaction times to Standard British English and Bulgarian-accented words, depending on the listeners' native language (estimates and standard error).

8.3.2. Overall accuracy, linear analysis

This section investigates the effect of the listeners' L1 and the stimuli's accent on the participants' overall accuracy. It was expected that L1 Bulgarian listeners would have higher accuracy with BA items than L1 English listeners and lower accuracy at SBE items than L1 English listeners.

The summary of the accuracy results per Accent and listeners' L1 is presented in Table 19 below.

Listeners' L1	Accent	Number correct
Bulgarian	Bulgarian accent (BA)	707 (89.4%)
Bulgarian	Bulgarian accent (BA)	744 (93.5%)
English	Standard British English (SBE)	419 (89.3%)
English	Standard British English (SBE)	460 (96.4%)

Table 19. Overall accuracy results on real word stimuli per accent

A linear mixed effects model was prepared with the *Accuracy* as an outcome variable. Correct answers were coded as 0 and the incorrect answers were coded as 1. The predictors were the *Listeners' L1* (Bulgarian and English) and the *Accent* (BA and SBE) and their interaction. Both variables were sum coded. The model also included random slopes of *Listeners' L1* per Soundfile and *Accent* per Speaker. The analysis included answers to real word stimuli, which amounted to 2330 observations based on 256 soundfiles and 40 participants. The results are summarised in Table 20 bellow and illustrated in Figure 33, which shows the model estimates and standard error.

Predictor	Estimate	*z*-value	*p*-value
Intercept	-3.51	-14.2	<0.001
Listeners' L1	0.23	1.1	0.30
Accent	0.49	3.2	0.001
Listeners' L1 by Accent	-0.17	-1.49	0.14

Table 20. Summary statistics on the Accuracy differences between the native and non-native English listeners for Bulgarian-accented and native English stimuli.

The only significant predictor in the model was a main effect of the stimuli's *Accent*. As illustrated by Figure 33, there was no systematic difference between the two groups of listeners, though there was a small tendency for the Bulgarian listeners to give more incorrect answers, compared to the native English listeners regardless of the accent.

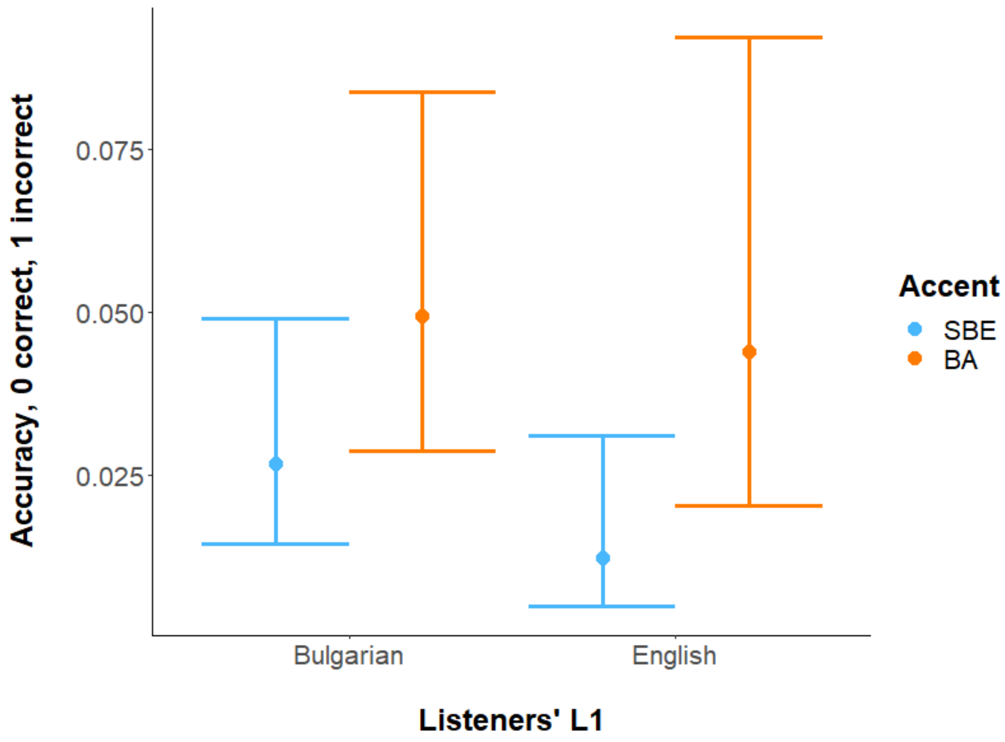


Figure 33. Model prediction for the Accuracy to Standard British English and Bulgarian-accented words, depending on the listeners' native language (estimates and standard error).

8.3.3. Reaction time adaptation within a block, curve analysis

This section investigates the hypothesis that L1 Bulgarian listeners would adapt faster to BA stimuli than L1 English listeners and that they would adapt slower to SBE stimuli than L1 English listeners.

A generalised additive mixed model (GAMM) was created, in order to investigate the effect of the listeners' native language and the accent of the stimuli on their speed of word recognition. Similarly to the GAMM in the previous chapter, the purpose of this analysis is not to focus on the process of getting used to the task but their adaptation to the accent, so only the responses from the first speaker block from each accent were analysed. Of them, only correct responses to real words were included in the analysis, between 150 ms and 2500 ms. The reaction times were centred around their mean.

Due to the specificities of the mgcv package (Pedersen et al., 2019) on R (R Core Team, 2018), the triple interaction between predictors *Accent* (SBE vs. BA) and *Listeners' L1* (English vs. Bulgarian) and the *Within Block Trial Number* could only be included in the model by combining Accent and *Listeners' L1* as one predictor with four levels (k = 10): Bulgarian accent x Bulgarian listener (baseline), Bulgarian accent x English listener, SBE accent x Bulgarian listener, SBE accent x English listener. The model also included a smooth term for the token number *Within-Block*, random smooths for *Within-Block* number per *Trajectory* and a random smooth for *Within-Block* by *Participant*.

Smooth terms	Edf	*F*	*p*-value
Within-block	4.64	4.68	< 0.001
Within-block by Accent (SBE) and L1 (Bulgarian)	3.04	4.60	0.001
Within-block by Accent (BA) and L1 (English)	1.81	0.73	0.583
Within-block by Accent (SBE) and L1 (English)	1.00	7.10	0.008
Random smooth: Within-block per Trajectory	37.75	0.34	< 0.001
Random smooth: Within-block per Participant	35.36	0.67	< 0.001

Table 21. Summary statistics for the smooth and random terms of the full GAM model.

Table 21 summarises the results for the smooth terms of the model and Figure 34 illustrates them by showing the group means and standard deviations across the block trials.

- There was a significant effect for the *Within-block* smooth, meaning that there was a non-linear change of RT with increasing token number within a block.
- Crucially, there was no significant interaction between the *Within-block* trial, Bulgarian-accented words and Bulgarian and English listeners (pale blue and dark blue lines in Figure 34). This suggests that the Bulgarian-English bilinguals and native English listeners adapted to the Bulgarian accent in a similar trajectory, although there is a clear tendency of faster adaptation for the native Bulgarian listeners.
- The baseline of Bulgarian-accented words with Bulgarian respondents (pale blue line in Figure 34) had a significantly different shape of RT adaptation from SBE-accented words with Bulgarian respondents (dark orange line) and English respondents (pale orange line). The SBE stimuli led to initial RTs which were faster than the mean and then gradually slowed down, while the Bulgarian-accented stimuli generally led to the opposite pattern.

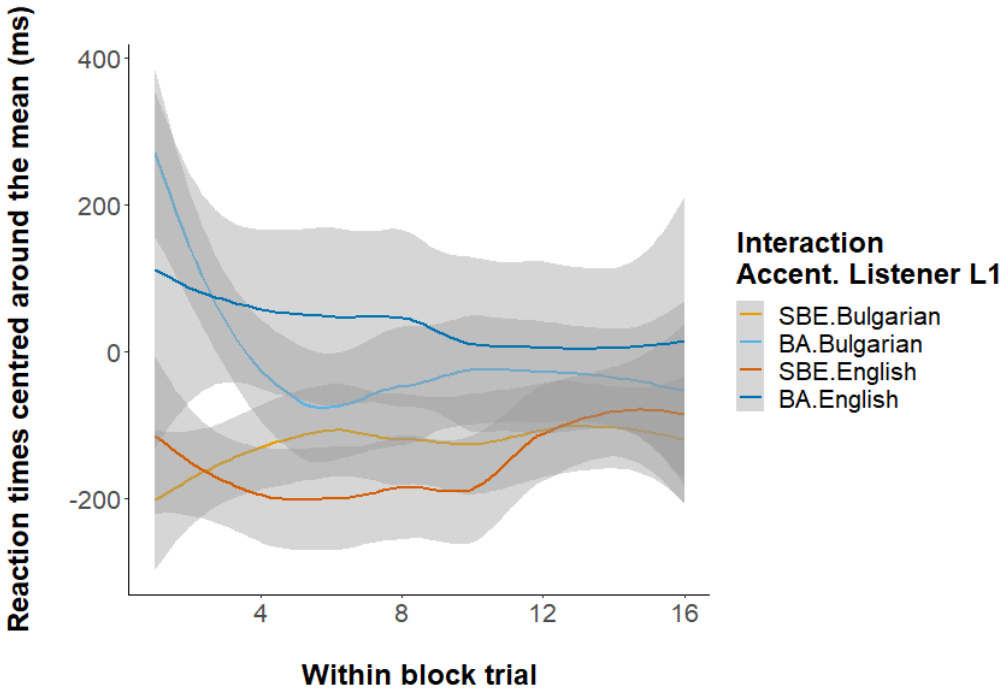


Figure 34. RT adaptation to Standard British English and Bulgarian-accented stimuli over the first half of an Accent block by native Bulgarian and native English listeners.

8.3.4. Reaction time adaptation across blocks, linear analysis

This section investigates the effect of the listeners' L1, the accent of the current speaker block and the accent of the stimuli in the preceding block on the listeners' reaction times. Based on the results in Experiment 2, it was expected that all listeners would react faster if they had just heard a block with the same accent as the target accent. Also, it was expected that the L1 Bulgarian listeners would react faster to Bulgarian accented English than L1 English speakers, based on ISIB-L. Lastly, it was expected than the L1 Bulgarian listeners would have a bigger gain of having heard BA English in the previous block than L1 English participants.

For the purpose a linear mixed effects model was prepared with RT (ms), centred around the mean, as an outcome variable. The predictors were the *Listeners' L1* (Bulgarian or English) and the *Accent* (BA or SBE), the *Preceding Accent* (same or different) and their interactions. All variables were sum coded. The model also included random slopes of *L1* per Soundfile and *Accent* per Speaker. The analysis included only correct answers to real word stimuli, which amounted to 1739 observations based on 255 soundfiles and 40 participants. The results are summarised in Table 22 bellow and illustrated in Figure 35, which shows the estimates and standard errors from the model.

Predictor	Estimate	*t*-value	*p*-value
Intercept	-57.96	-3.01	0.004
Listeners' L1	1.25	0.07	0.95
Accent	63.37	5.72	<0.001
Preceding Accent	8.53	1.14	0.26
Listeners' L1 by Accent	-12.75	-1.42	0.16
Preceding accent by Accent	1.12	0.14	0.89
Listeners' L1 by Preceding accent	17.53	2.65	0.01
Listeners' L1 by Preceding accent by Accent	0.82	0.12	0.91

Table 22. Summary statistics for the linear analysis of the Current and Preceding accent effects on the Reaction times of the native English listeners and the Bulgarian-English bilinguals.

The results suggest that there was a main effect of *Accent*, where all listeners were generally slower when reacting to Bulgarian-accented words. In addition, there was an interaction between the *Listeners' L1* and the *Preceding accent*. This result means that the L1 Bulgarian listeners were slower to recognise words if the preceding accent was different compared to when it was the same as the target accent, while the native English listeners did not show any consistent change, also illustrated in Figure 35 below.

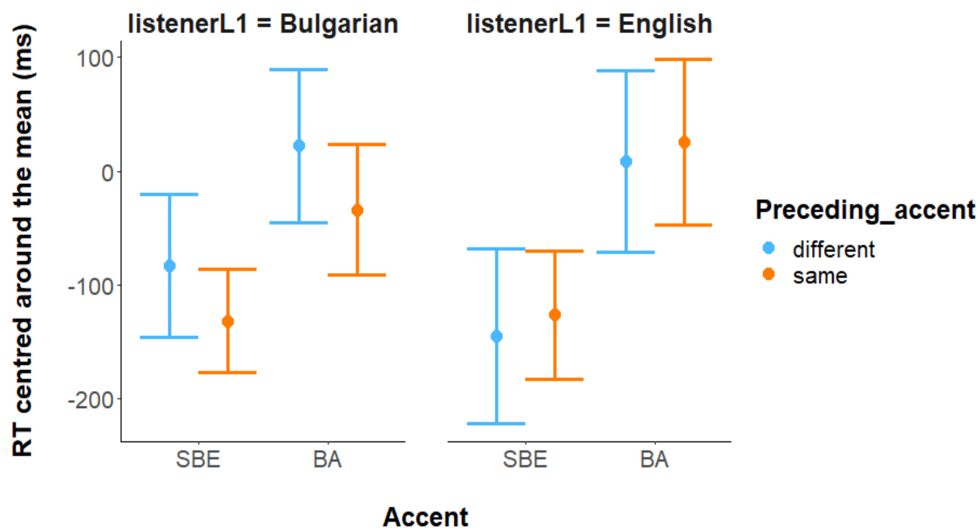


Figure 35. Model prediction for the centred reaction times to Standard British English and Bulgarian-accented accented words depending on the preceding accent and the listeners' L1 (estimates and standard error).

8.3.5. Accuracy adaptation across blocks, linear analysis

This section investigates the effect of the listeners' L1, the stimuli's accent and the accent in the preceding block on the listeners' accuracy. As in the previous section, it was expected that all listeners would have a higher accuracy if the previous speaker block had the same accent as the current speaker block. According to ISIB-L, it was expected that the L1 Bulgarian listeners would have a higher accuracy with Bulgarian accented English than L1 English speakers. Also, it was expected than the L1 Bulgarian listeners would have a bigger gain of having heard BA English in the previous block than L1 English participants.

In order to investigate the effect of the listeners' L1 and the speakers' accent and the effect of the preceding accent, a linear mixed effects model was prepared with *Accuracy* as an outcome variable. Correct answers were coded as 0 and incorrect answers were coded as 1. The predictors were the *Listeners' L1* (Bulgarian or English) and the *Accent* (BA or SBE), the *Preceding Accent* (same or different) and their interactions. All variables were sum coded. The model also included random slopes of *L1* per Soundfile and *Accent* per Speaker. The analysis included only answers to real word stimuli, which amounted to 1897 observations based on 256 soundfiles and 40 participants. The results are summarised in Table 23 bellow and illustrated in Figure 36, which shows the model's estimates and standard errors.

Predictor	Estimate	*t*-value	*p*-value
Intercept	-3.59	-12.13	<0.001
Listeners' L1	0.19	0.72	0.47
Accent	0.64	3.33	<0.001
Preceding Accent	0.11	0.73	0.47
Listeners' L1 by Accent	-0.16	-1.09	0.27
Preceding accent by Accent	0.26	1.74	0.08
Listeners' L1 by Preceding accent	0.05	0.36	0.72
Listeners' L1 by Preceding accent by Accent	-0.01	-0.10	0.92

Table 23. Summary statistics for the linear analysis of the Current and Preceding accent effects on the Accuracy of the native English listeners and the Bulgarian-English bilinguals.

The only significant predictor in the model was the words' *Accent*. Overall, there was a higher probability of an incorrect answer when the words were produced with a Bulgarian accent. As apparent from Figure 36 below and from Table 23 above, there was also a tendency for an interaction between the words' *Accent* and the *Preceding accent* of the block, where Bulgarian-accented words are recognised more accurately if they have been preceded by Bulgarian accent compared to when they were preceded by native English speech.

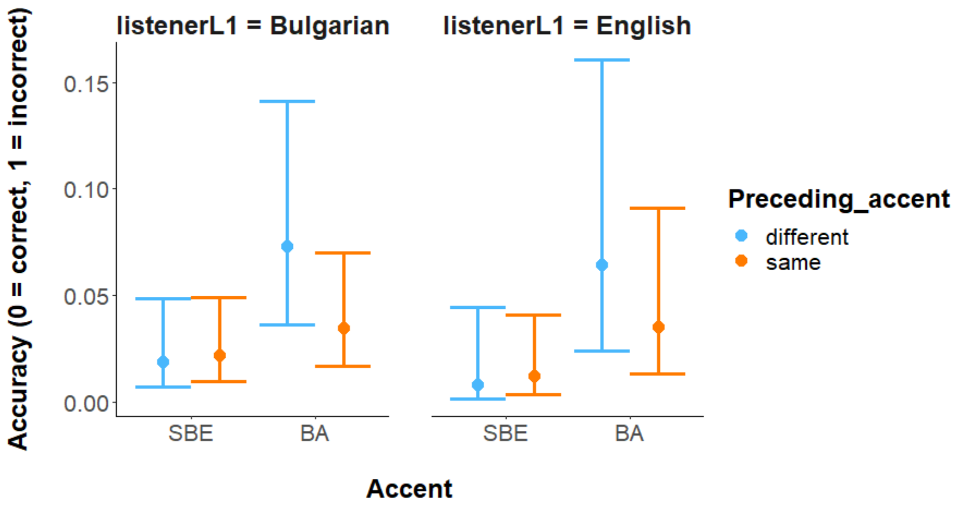


Figure 36. Model prediction for the accuracy to Standard British English and Bulgarian-accented words depending on the preceding accent and the listeners' L1 (estimates and standard error).

8.4. Discussion

This study explored whether the Bulgarian-English bilinguals who achieved the highest English scores on the LexTale test (over 90%) had any advantage in processing Bulgarian-accented words compared to native English listeners. All five types of analysis lack evidence of a specific advantage for Bulgarian-accented words, thus failing to support the ISIB hypothesis.

In the linear analyses of reaction times and accuracy, the only significant predictor was the accent of the stimuli: Bulgarian-accented words, such as [xɛə] were processed slower and more inaccurately by both groups of listeners, compared to the SBE alternative [hɛə]. The curve adaptation analysis also showed that both groups of listeners had similar patterns of adaptation to Bulgarian-accented words at the beginning of an Accent block. Nevertheless, there was a tendency for a slightly steeper adaptation curve for the Bulgarian-English bilinguals compared to the L1 English listeners.

Lastly, the effect of the accent of the preceding listening block was analysed with respect to the listeners' reaction times and accuracy. The accent of the preceding block interacted with the listeners' native language only with respect to their reaction times. The Bulgarian-English bilinguals processed words faster after hearing the same accent on a previous occasion, while the native English speakers seemed not to improve their reaction times either for the Bulgarian-accented words or the SBE ones after hearing the same accent earlier. These results suggest that the Bulgarian-English bilinguals needed less evidence to adapt their reaction times, or that they were more willing/able to adapt at all. The fact that they did this for both Bulgarian-accented and English accented words suggests that they had no special preference for the Bulgarian-accented words. No such interaction was observed in the accuracy results.

The theoretical significance of this result will be fully discussed in the general discussion. However, it is worth linking this result with the idea of flexible and narrow listeners, suggested by Kleinschmidt and Jaegger (2015a). These two terms are potentially ambiguous with respect to bilingual listeners. Bilinguals could be considered more flexible than monolinguals, as they have access to two phonetic systems (cf., Weber et al., 2014), while the native monolingual listeners of a language would tend to be more specialised listeners (although research needs to compare multi-dialectals to multi-linguals). However, a second language listener of a language would have less cumulative exposure to that particular language, compared to a native listener. Hence, in a situation where a variety of native English accents have to be processed, a native listener might have an advantage and be more flexible than an L2 English listener.

As suggested in Section 8.2, only four participants in the native English group spoke an additional language, while in the bilingual group there were three people who had a third language. Hence, without taking the results into consideration it may be assumed that in the present sample the Bulgarian-English bilinguals are the more flexible listeners compared to the native English listeners as the listening situation involves a high prestige native variety and a non-native accent. In that case, the fact that the Bulgarian-English bilinguals adapted to Bulgarian-accented speech and native English speech is consistent with the expectation of them being the flexible listeners. The native English listeners did not adapt to native English speech, likely because their reaction times were already at ceiling level (see

Figure 35). Their inability to achieve cross-speaker adaptation in Bulgarian-accented speech is consistent with the expectation that they are specialised in native Standard British English with poor representations of Bulgarian-accented English. This discussion is continued in Section 9.2.3.

8.5. Limitations

There are several limitations to this experiment. First, there were only 15 L1 English listeners whose data could be analysed. Second, as emerged from the discussion, the experiment could have benefitted from a reliable metric of exposure to English and Bulgarian-accented English specifically. A future study could compare the validity of self-reported time, length of residence, self-reported size of anglophone social circle and years spent speaking English (Age of Onset subtracted from Age) for general exposure to English. Narrowing down linguistic exposure to Bulgarian accent specifically would be an even greater challenge. Using these metrics as predictors without a sense of their validity could return misleading results.

8.6. Conclusion

This additional exploratory experiment found no evidence of Bulgarian accent advantage for high proficiency Bulgarian-English bilingual listeners compared to native English listeners. With the exception of their adaptation results there were no systematic differences between both listener groups. The main systematic finding was that Bulgarian-accented English is processed slower and more inaccurately than the native English listeners. However, the adaptation results suggest that both groups of listeners managed to adapt their reaction times to the global RT mean between both accents.

The results so far suggest that high proficiency Bulgarian-English bilinguals' processing of Bulgarian-accented English is more in line with native English speakers than low-proficiency Bulgarian-English bilinguals. There results could be indicative of the dynamic nature of language dominance, but they could also signal a tendency in L2 development, that leads to a stricter separation between the L1 and L2 fine phonetic systems and specialisation in a prestige variety of L2.

Chapter 9 General Discussion

This chapter will explore the results from the previous three chapters in the context of the theories presented in the literature review.

Across the experiments a Bulgarian accent advantage was not consistently observed. In Experiment 1 the bilinguals with low English proficiency responded faster to trials that contained Bulgarian accent (BA) compared to trials that contained Standard British English. In Experiment 2 Bulgarian-English bilinguals with the lowest English proficiency responded faster and more accurately to BA English than the listeners with high English proficiency. However, in the Follow-up experiment the latter did not have a similar advantage (nor a disadvantage) when compared to native English listeners.

These results suggest that there is little evidence for an Interspeech Intelligibility Hypothesis (ISIB) for Talkers: apart from Experiment 1 the bilinguals did not benefit from BA English compared to SBE, and some were even at a disadvantage. There is weak support for ISIB for Listeners: the low English proficiency listeners had some advantage over the high English proficiency listeners when responding to BA English, but no advantage was observed when the high-proficiency listeners were compared to native English listeners. These results suggest that the high proficiency listeners in this sample are not relying on their L1 phonetics when processing L2 speech. In fact, as predicted by some models of L2 phonetic acquisition, they perhaps rely on native-like representations in English word processing.

The majority of the following discussion will be dedicated to the results of Experiment 2, but the remaining studies will also be discussed in relation to Experiment 2, to shed further light on the conclusions.

9.1. Relevance for models of bilingual listening

Overall, it appears that the bilinguals in this study do not perform as a monolithic group. The listeners' proficiency in English is an important predictor for their reaction times and accuracy in response to different accents. This provides cross-sectional evidence for developmental changes in L2 processing, which needs to be accounted for in bilingual models of speech processing. The following discussion of these models will then be incorporated in the critique of general models of speech processing in Section 9.2.

9.1.1. BIMOLA

Based on Grosjean's (2001) concept of language modes, BIMOLA (Léwy and Grosjean, 2008) is one of the few models that addresses lexical access for bilingual listeners starting from the level of phonetics. The model explicitly states that both languages of the bilingual are in a constant state of activation flux and that both top-down and bottom-up levels of activation are possible.

Regarding the processing of Bulgarian-accented English by Bulgarian-English bilinguals, BIMOLA would predict that the features of Bulgarian accent in the input activate phonemes from both the Bulgarian and English phonemic subsets. Consistent with the studies that demonstrate language non-specific lexical activation (Blumenfeld and Marian, 2007; Ju and Luce, 2004; Lagrou et al., 2011; Marian et al., 2003, 2008), BIMOLA predicts that the activation of Bulgarian-specific phonemes can further trigger Bulgarian words and an overall increase of the activation of the Bulgarian language. In comparison, exposure to input with no Bulgarian accent would not lead to the increased activation of Bulgarian, even if according to the model, the competing language cannot be completely turned off.

Under this architecture, it was expected that Bulgarian-accented words would increase the competition between Bulgarian words and English targets compared to native English accented words and respectively slow down processing times due to increased inhibition effort. Research on cognate recognition suggests that cognates are indeed processed faster than unrelated words (Dijkstra et al., 1999; van Hell and Dijkstra, 2002) because the lexical access to the correct lexical item is phonologically supported from both languages of the bilingual. In a case of non-cognates (as in this experiment), it was expected that any interference of the L1 phonetics would be detrimental as opposed to facilitatory, as suggested by research on interlingual homophone/homograph processing (Marian et al., 2003, 2008; Spivey and Marian, 1999; Weber and Cutler, 2004).

At first sight, the results of Experiment 2 overall support the predictions of BIMOLA, as Bulgarian-accented words are recognised more slowly and incorrectly than native English utterances. This could have been the result of the Bulgarian accent activating competing Bulgarian vocabulary, thus requiring increased overall inhibition of Bulgarian. Some of the studies discussed in the literature review suggest that if the listeners have clear expectations about the speaker's accent, their processing can be affected. The presence of a Bulgarian accent can increase the salience of Bulgarian competitor words due to a top-down activation of a Bulgarian speaker model. Such top-down accent adjustments have been observed by Cai et al. (2017). According to BIMOLA, this can happen in addition to the bottom-up activation of Bulgarian words and would then require the inhibition of more candidates.

This explanation is flawed, however, as the general Bulgarian disadvantage is mediated by the listeners' English proficiency. As the model is underspecified regarding proficiency and language dominance, it is unclear why high proficiency in English would lead to slower reaction times and lower accuracy with Bulgarian-accented English compared to lower proficiency in English. If, according to BIMOLA, poorer performance is associated with greater competition from L1, that would lead to the unlikely explanation that the higher proficiency listeners experience higher competition with Bulgarian at the encounter of Bulgarian-accented words compared to lower proficiency listeners. Durlik et al. (2016), who used written words as stimuli, expected that L1 interference would be stronger for lower proficiency bilinguals but found no effect of proficiency. These results suggest that the L1-vocabulary interference explanation is not well supported.

In Experiment 1, there is very weak evidence of Bulgarian facilitating Bulgarian-accented English words, and even this observation is modulated by the listeners' proficiency. BIMOLA does predict facilitation between Bulgarian and Bulgarian-accented English thanks to the listeners' bilingual mode during the testing. However, the different proficiency results observed here and in Szakay et al. (2016) call for an investigation in the link between proficiency and language mode (also see Antoniou et al., 2010).

BIMOLA makes few predictions about short-term accent adaptations, such as the reaction times adaptation to Bulgarian-accented speech observed here. The only short-term adaptation suggested in the model is the increased bottom-up and top-down activation of a language that results from a lateral increase of activation of words within the same language. Despite an initial activation of Bulgarian due to a Bulgarian accent in Experiment 2, the continued exposure to English words (despite the Bulgarian accent) would increase the activation of the subset of English words as opposed to the Bulgarian ones. According to BIMOLA this should also lead to top-down activation of native English-like phonemes and could prevent the adaptation to a Bulgarian accent top-down (see Figure 1). Without the possibility for recalibrating the association of Bulgarian-like phonetic input with L2 phonemes, BIMOLA is restricted in accounting for short-term perceptual adaptation to an L1-accented L2.

While BIMOLA can capture the overall tendency of Bulgarian accent disadvantage it does not provide a good fit of the dynamic analysis of the data, as it cannot account for proficiency and adaptation effects. A further development of the model would require provisions for long-term (due to proficiency) and short-term (due to recent exposure) phonetic adaptations to different accents.

9.1.2. SLM and PAM

As discussed in the literature review, some of the most influential models in second language speech perception, SLM (Flege, 1995) and PAM (Best and Tyler, 2007), base their architecture on the assumption that L2 listeners employ correspondences between their native and second language phonetic and phonological systems. Although they do not make explicit predictions about word processing and L1 accent processing, in Section 2.2 it was argued that based on their existing predictions, the models would predict that in some circumstances L2 similarity to L1 on a phonetic level might be beneficial for speech processing for L2 listeners. Yet that is not observed in the results of the present study, at least compared to native speech processing.

The fact that on the whole the listeners in this study process Bulgarian-accented English more slowly than natively pronounced English suggests that most of the listeners are using dissimilated sounds in L2 (reliant on a native English model). SLM captures the interaction between proficiency and accent, as the listeners with lower proficiency appear to process Bulgarian accent more easily than listeners with higher proficiency. SLM does predict that with the increase in proficiency listeners are more likely to achieve separate phone representations for both languages, even for the so-called similar phones. However, SLM predicts that low proficiency listeners use L1 phonetics for the representations of equivalent sounds in L2. That would lead to the prediction that lower proficiency listeners should have processed Bulgarian-accented speech faster and more correctly than native English speech, which is not observed.

In this context the concept of phone can be reinterpreted as phonetic cue realisations, typical of one language or the other. Bundgaard-Nielsen et al. (2011) bring up circumstantial evidence suggesting that at their arrival in an L2-speaking country, listeners achieve very rapid perceptual adaptation, followed by a plateau. Hence it is possible that Bulgarian beginner learners of English who have just arrived in the UK would process Bulgarian accent faster and more accurately than native English speech. This possibility can be explored in targeted future studies.

PAM (Best and Tyler, 2007) would predict an advantage of Bulgarian accent only if the listeners have established systematic correspondences between L1 and L2 sounds in a two category assimilation. The Vocabulary model, proposed by the same team (Bundgaard-Nielsen et al., 2011), predicts that as the listeners gain more exposure to their L2, they stabilise the correspondences between L2 and L1 phones. As discussed earlier, achieving one-to-one correspondences would be difficult in the case of L1 Bulgarian and L2 English, hence PAM would generally not favour a Bulgarian accent advantage (although it would predict English-accent advantage in Bulgarian for L1 English – L2 Bulgarian listeners). Section 7.4.2 also provides exploratory evidence that higher proficiency listeners tend to systematically mishear Bulgarian-accented words, while lower-proficiency listeners make more sporadic errors. As discussed in Section 7.5 the systematicity is linked to the higher proficiency listeners developing narrower expectations for native English speech (hence, fewer systematic errors with that accent), while the lower proficiency listeners are more flexible.

Apart from long-term adaptation as a result of proficiency changes, SLM and PAM make no specific predictions about short-term adaptations in speech processing. In that respect, the findings of rapid adaptation, reported in this dissertation, highlight a gap in the theories, which should be considered in future updates as potential drivers of long-term change.

Overall, while SLM and PAM are very useful in accounting for the discrimination and production in individual sounds, they lead to vague predictions about word and L1-accent processing. Nevertheless, they provide a useful framework that can be used to enrich discussions on general speech processing, such as in Section 9.2.

9.1.3. ASP

One of the models perhaps best suited for discussing the L1-accent benefit, as studied in this experiment, is the Automatic Selective Perception (ASP) model by Strange (2011). The model is focused on the L2 listeners who are immersed in their L2 language environment, which was one of the main criteria used for selecting participants. The ASP model is mostly concerned with bottom-up listening without any top-down semantic and syntactic influences, which is also consistent with the type of stimulus presentation used for Experiment 2 and the follow up. The main prediction drawn from this model is that as L2 listeners gain more experience they move away from a “phonetic” style of listening, which is more costly because it requires paying attention to more phonetic information, and move towards a “phonological” listening style. That style of listening relies on selective perception routines (SPRs) and attention is paid to a small number of familiar phonetic cues.

There is some evidence for the support of this prediction. The interaction between Proficiency and Accent suggests that as the listeners’ proficiency in English increased, they are faster and more accurate at processing native Standard British English. This is consistent with a more automatic style of speech processing, specialised for native speech. The data for Bulgarian-accented processing suggest that as the listeners’ proficiency in English increases, they lose some of the SPRs that they had for this variety.

While the model explains that lower proficiency listeners process L1-accent in L2 faster and more accurately compared to the higher proficiency listeners, it remains an open question why the lower proficiency listeners do not have this benefit over SBE speech. It is possible that the low proficiency listeners in this experiment are at a stage of developing English-specific SPR but also rely on Bulgarian-like SPRs to a certain extent, perhaps in a similar situation as the higher proficiency listeners of Ludwig and Mora (2016) who used English mostly in the classroom.

As suggested earlier, the ASP model does not make direct predictions for short-term adaptations to speech. However, it is stated that the amount of attention paid to speech can vary gradiently between the phonetic and phonological modes of listening. With these assumptions the contrasting reaction times to both accents at the start of a new accent block can be accounted for. When presented with a new accent, listeners in general, not just L2 listeners, would be expected to increase their attention and their phonetic mode of listening, in order to adjust their otherwise highly automatic SPRs to that particular speaker or accent. As this type of listening is oriented towards greater phonetic nuance, it would require more processing cost, which would slow down reaction times (Strange, 2011).

In the present study, when listeners with variable proficiency in English start hearing stimuli with a different accent, the higher proficiency listeners respond faster than the global mean to native English words and lower proficiency listeners respond slower than the global mean. After 6 to 8 trials all listeners converge to the mean. This result might be interpreted as short-term SPR establishment for lower proficiency listeners but provides no explanation for the initially rapid and then slow responses of the higher proficiency listeners. If fast reaction times only signal the use of efficient SPRs, this would lead to the unlikely conclusion that the high proficiency listeners responded correctly at the start of the accent block using less attentional resources but increased their attention to detail as the block continued. This interpretation contradicts the well-established finding that people pay more attention when exposed to novelty and then get habituated (for a discussion see Mather, 2013).

This interpretation raises the question to what extent reaction times are a good measure of the ambiguous concept of listening effort (or processing cost). RT can still be used as an indicator for how much time is required for lexical access to be achieved and longer time may sometimes be associated with more effort, particularly in adverse listening conditions. However, an increased amount of attention, which would require higher processing cost may sometimes lead to faster lexical decisions, as will be discussed in Section 9.2.3. In cases like that, RT might become a poor indicator of attention, effort or processing cost. An alternative measure of attention during listening might be pupil dilation (Wainstein et al., 2017). Hence the Attention variable which is inherent of the ASP model cannot be directly attested in the experiments of this study. The discussion of the future directions Section 9.4 proposes a solution to this problem.

Overall, with some accommodations, the ASP model provides a good fit of the results observed in this dissertation. Future research can benefit from systematically studying the interaction of attention and reaction times over time when processing familiar and unfamiliar accents.

9.2. Relevance for general models of speech processing

This section discusses the implications of the results presented so far in the context of models of general speech processing, which do not specialise in bilingual listening. As stated earlier, the major challenge for a general speech processing model is to account for speech processing in listeners who have more than one language system.

Chapter 1 distinguished between two rough classes of speech perception models – those based on strict hierarchical processing of acoustic bundles of features (or inferred articulatory gestures) into increasingly complex structures (e.g., phonemes, syllables, words etc.) and exemplar models, which compare the acoustic details of the input to rich episodic memories of speech without assumptions of the perceptual building blocks. In addition, the Bayesian Ideal Adapter Framework (IAF) and the neuroscientific Thousand Brains Intelligence Model (TBIM) were discussed in the context of unfamiliar speech processing by more and less experienced listeners.

9.2.1. Traditional approaches

One of the characteristics of the traditional models is that they assume that accent-specific and speaker-specific information is discarded from the process of lexical access at an early stage of speech processing (Calabrese, 2012; Marslen-Wilson and Welsh, 1978; McClelland and Elman, 1986), and that as long as all the minimal distinctive features that specify a phoneme can be recovered, then the processing of a given word should not be inhibited. This setup by itself does not explain why word processing might be successful but delayed for different accents as in the present study and for different meanings within accents (Cai et al., 2017).

Traditional models can in principle account for problems with accuracy, or mishearing. If L2 speech, like the Bulgarian-accented speech in this experiment, contains phonological features incompatible with the target language or if it lacks important distinctive features in the acoustic input, that could lead to the decoding of an incorrect phoneme, and hence to mishearing. If the lower proficiency listeners rely on L1 distinctive features within their L2 (based on the predictions of SLM and BIMOLA), and the L1 accent contains them, this can explain why the lower proficiency listeners have a higher accuracy with Bulgarian-accented speech compared to higher proficiency listeners. However, the lower proficiency listeners in this experiment also process native English speech with similar accuracy as the Bulgarian-accented speech. This result is similar to the findings of Bent and Bradlow (2003) who interpret the lack of disadvantage as a processing benefit. From the point of view of a speech processing models based on the sequential decoding of distinctive features this result suggests that listeners with a LexTale around 60% (lower-intermediate) have unstable phonemic representations that use both L1-specific and L2-specific distinctive features. Another problem is that the listeners with intermediate to high proficiency seem to gradually specialise their representations.

While it might be possible to account for the accuracy results, it is unclear why decoding distinctive features in some accents for some listeners might sometimes take longer. For instance, it is unclear why high proficiency bilinguals and the native English listeners are slow but still generally successful at processing Bulgarian-accented English.

Calabrese's (2012) traditional model attempts to address this question using the concept of echoic memories. According to Calabrese (2012), in some situations (such as exposure to an unfamiliar language or a new word) listeners need to work to achieve parity between their existing representations and the input, as opposed to achieving it automatically. That requires analysis-by-synthesis, a several-step process of decomposing the rich phonetic input, preserved as echoic memories, into familiar phonemic features and segments, which adhere to the grammatical and phonotactic rules of the listener's language. It is unclear under what conditions a misinterpretation might occur. This process of gradual abstraction can be time and energy consuming, hence leading to slower reaction times.

Calabrese (2012) hypothesises that the exact mapping from acoustic detail onto more abstract features is achieved through the acoustic-landmark spotting proposed by Stevens (2002). After the landmarks have been spotted, the listener searches around them (presumably using the rich memory of the speech) to identify bundles of acoustic features that might help them identify a phoneme. In addition, according to Calabrese (2012, p. 373), "Listeners must be able to detect, know, and remember all types of phonetic detail about the speech of other members of the community in order to have adequate social interactions" but there is no elaboration how this knowledge is acquired and implemented on a practical level. Analysis-by-synthesis is not required for interpreting familiar words produced with a familiar accent, which may be interpreted directly from the raw input (Calabrese, 2012). The fact that the high proficiency listeners were slower with Bulgarian-accented words suggests that they were using analysis-by-synthesis when processing these stimuli.

If analysis-by-synthesis can be differentiated from phonological listening by the amount of attention used in the process, then there is a direct way to test this conclusion, namely by using pupillometry. A similar study has been carried out by Porretta and Tucker (2019), who investigated the effect of accent intelligibility and listener familiarity with the accent on their pupil size, over the course of exposure to a word. Their plots show that lesser experience and lower intelligibility tend to be associated with earlier onset of pupil dilation, between 400 ms and 600 ms after the onset of the stimulus, compared to more experience and greater intelligibility when dilation starts after 600 ms and does not reach the same magnitude. This suggests that listeners use more attentional resources overall and earlier in the listening process when it is difficult to understand the accent (be it due to its intelligibility or experience with it).

Assuming that the Bulgarian listeners have experience with Bulgarian phonetics and "phonological features", as Bulgarian is their native language, and with native English, as the dominant prestigious variety of their environment, it is not immediately obvious why the high proficiency listeners would struggle to extract the acoustic features from either variety. Calabrese (2012) emphasises the difficulty for adult L2 learners to associate the acoustics of an unfamiliar language with the features and grammatical restrictions available to them, which is only possible through the long-term preservation of echoic memories of the input.

In order for this model to usefully capture the different nature of L1-accent processing by higher and lower proficiency listeners, it needs to account for the fact that listeners might be habituated to certain combinations of acoustic markers more than others and that this habitation is a subject to change from situation to situation. If L1-accented L2 exhibits L1-like phonetics, then the access to L1 phonetics in high proficiency listeners must be more restricted compared to lower proficiency listeners. In the words of Calabrese (2012), this means that higher proficiency Bulgarian-English bilinguals cannot process Bulgarian-accented English based on a "rough phonological sketch", and they have to engage in analysis by synthesis, realigning the Bulgarian accent with their habitual mental representations of Standard British English.

In addition, Calabrese's model provides no specific account for rapid adaptation to a speaker/accent, hence it cannot be explained why the listeners, particularly the native English listeners and the higher proficiency Bulgarian-English bilinguals rapidly improve their reaction times to Bulgarian-accented stimuli, while at the same time they slow down their processing of native English speech. The analysis-by-synthesis described by Calabrese involves bottom-up and top-down loops, extracting features from the acoustic input and comparing them to mental representations until parity is achieved. As suggested in the discussion of ASP, the fact that the listeners slowed down towards the middle of the block would lead to the unlikely conclusion that as the block progressed, they started paying more phonetic attention.

Although traditional models do not focus on rapid adaptation, some mechanisms for it are available. TRACE (McClelland and Elman, 1986) for example mentions that top-down feedback is used to support learning and that when units are activated simultaneously a connection between them is learnt. This is supported by Kriengwatana et al. (2016) who demonstrate that listeners could adjust to changing the gender of the speaker without feedback, but they needed to know whether their guesses were correct, in order to adjust to an unfamiliar accent.

However, the present study provided no feedback or context at all to the listeners and they still changed improved their reaction times over time. This means that the listeners had to use their own judgement (perhaps in the feedback loops described by Calabrese, 2012) of their own accuracy to improve the efficiency of their listening and move from the time-consuming analysis-by-synthesis process to "phonological listening." This would suggest that the difference between analysis-by-synthesis and phonological listening is gradient, as opposed to categorical, as suggested by Calabrese (2012) and Strange (2011). If the relationship is gradient, this means that the model has to account for an intermediate stage of processing. In addition, it is unclear why at the same time-course, the listeners became slower at processing native English words. As it stands, the initial rapid RTs to native English speech would suggest that the listeners were engaged in "phonological listening" and as the experiment unfolded, they had to switch to the more time-consuming "analysis-by-synthesis". This contradiction will be resolved in the following section.

The traditional models of speech processing, specifically the model of Calabrese (2012), can account for difficulties in L1 accent intelligibility, including some differences caused by the listeners' proficiency. In addition, there are some attempts at resolving the issue of why the L1 accent is processed more slowly by high proficiency listeners than the Standard British English accent, but there is no explanation of a gradual transition between efficient and careful listening, as observed in this dissertation. In addition, the model emphasises the high priority of L1 phonology and grammar when learning an L2, which contradicts the observation of an overall L1-accent disadvantage in the data.

9.2.2. Exemplar approaches

The exemplar models of speech processing put a focus on the individual's experience with the phonetic nuance of a particular accent (or idiolect) to account for differences in speed of processing (Coleman, 2002; Goldinger, 1996, 1998, 2000; Hawkins and Smith, 2001; Lev-Ari, 2017).

When applying exemplar model predictions to the context of L1-accent processing in L2, it is expected that higher proficiency listeners would have had more exposure to different speakers of English than lower proficiency listeners. On a group level this leads to the expectation that higher proficiency listeners should be more flexible in their processing of different accents of English and thus process both accents either equally well or faster than the lower proficiency listeners. Alternatively, they should be expected to process the Bulgarian accent similarly to lower proficiency listeners and the native English speech faster than the lower proficiency listeners. However, this is not observed in the data. Instead, the higher proficiency listeners are more specialised in the native variety over the Bulgarian accent, while lower proficiency listeners do not distinguish between the two accents as much, in terms of reaction times and accuracy. The results in this experiment support a model that allows for a long-term trade-off between competencies, as opposed to their accumulation.

Episodic models have mechanisms that allow for adaptation in listening. The more time passes after the exposure to a certain example of speech, the more the memory of it decays (Goldinger, 1998). In addition, exemplar philosophical models of conceptual categories (Medin and Schaffer, 1978) and similar counterparts in speech processing (Pierrehumbert, 2002) suggest that as novel input is encountered, the exemplars that are most similar to the prototype have the largest weight when updating the category. It is likely that the high proficiency listeners in this experiment have changed the nature of their perceptual prototypes and although they may encounter examples of Bulgarian-accented speech as frequently as the lower proficiency listeners, that speech may have less influence on the updating of their default expectations of phonetic variation in English. A similar prestige benefit in speech processing has been reported in monolingual studies, particularly when processing speech in noise (Clopper and Bradlow, 2008; Clopper and Tamati, 2010). Clopper (2014) argues that a standard variety benefit is a result of exposure from the media and public space. She argues that frequent exposure to multiple varieties leads to wider distributions of phonetic cue representations, reminiscent of Kleinschmidt and Jaegger's (2015a) Ideal Adapter Model, discussed in the following section.

In the case of Bulgarian-English bilinguals, living in the UK, this means that increased proficiency is associated with increased exposure to and importance assigned to the native Standard British English variety at the expense of Bulgarian-accented English. The Standard British English phonetic cues become the default expectations for high proficiency listeners, at the expense of Bulgarian accent phonetic cues (e.g., distance between F1 and F2 of /ɪ/ as demonstrated in Experiment 1). The effect of the recency of exposure in the current data is demonstrated in the accent generalisation effect. The bilingual listeners were faster after hearing a speaker of the same accent than they were after hearing a speaker of a different accent. This effect was not present for the native English listeners. However, that evidence speaks to the bilingual listeners' ability to rapidly "tune in" to a specific variety, discussed more in the following section. Hence, the results of this dissertation support an exemplar model of speech processing, which is based on the existence of prototypes and prestige of different accents.

Overall, the exemplar models of speech processing can account for some of the overall effects observed in the present dissertation, but yet again, a more detailed look is required to explain some of the short-term accent-adaptation processes. The main source of unclarity comes from the interplay between several important factors influencing Bulgarian-accented English processing: access to Bulgarian-like and Standard British English phonetic variation while listening to English, ability for long-term and short-term adaptation and attention-focusing. While general exemplar models may recognise the existence of some of these influences, there are few specific predictions that can account for how they interact, in order to account for the present data. The next section will focus on how models of speech processing and adaptation can provide a specific account of the results.

9.2.3. Dynamic approaches: Ideal Adapter Framework

As discussed in the literature review, the dynamic approach to modelling speech processing is, on the whole consistent with exemplar models, as both approaches put their focus on the listeners' ever-changing experiences.

In summary, the IAF uses the Bayesian equation to account for the listeners' tracking the mean and the spread of multiple acoustic features, and that the recognition of any type of segment relies on both sources of information. In order to achieve a temporary adjustment of a phonetic representation, the listeners need to experience a critical accumulation of input that would change both or either the central tendency and the distribution of a given feature and that this can lead to expansion or narrowing of the variability of phonetic representations for that category.

In the present experiment, the focus is on the likelihood part of the Bayesian equation. The decision about the probability of hearing a given word is proportional to the prior probability of that word, given its frequency, recency of mention and semantic/pragmatic context. It also takes into account the likelihood that the observed acoustic input should be observed given the hypothesis for that word. The stimuli in Experiment 2 have a similar objective frequency and are presented out of context, which aims to keep the prior part of the Bayesian formula uniform. In addition, each listener hears each word only once, so as not to increase its predictability of a word due to recency of exposure. A listener's prior experience with a particular accent would allow them to make more precise judgements regarding the likelihood of hearing specific acoustic input given a hypothesis for a specific word or a segment. In this context of biasing towards the concept of Bulgarian-ness, only the lower proficiency listeners had a small bias for Bulgarian-accented pronunciation. That was observed in their RTs at the start of a listening block, and when compared to higher proficiency listeners.

Two important types of listeners can be distinguished based on their expectation of narrow or wide variance of phonetic cues. Flexible listeners expect greater variability in the input, and hence need more information to achieve certainty in their interpretation. Narrow listeners expect small variability in the phonetic input, which would allow them to achieve certainty faster, but only if the input matches their narrow preconception. It is hard to neatly translate the flexible and narrow listener paradigm to the present experiment, as achieved by Lev-Ari (2017), for example. Many of the non-native English listeners are multilingual and measuring all their respective proficiencies and the size of their social circle would have been beyond the scope of this dissertation. The follow-up experiment has a clearer distinction because the majority of the native English listeners are monolingual. However, the issue with the predictions about narrow and flexible listeners in Kleinschmidt and Jaeger (2015a) is that the ability to efficiently adapt to a stimulus could be associated with both groups, depending on who they are compared to. If the listener has no representations for a given type of speech, they will need the most time to adapt to it (depending on the situation this could be anyone). If the listener has broad representations (a flexible listener), they might be faster than the completely unspecialised listener. However, a flexible listener would be slower to adapt to (or process) speech they recognise, compared to a listener who is specialised for that particular type of speech.

There is some evidence to support the following flexibility hierarchies in the listening situation investigated in the present dissertation: high-proficiency L2 listeners > L1 listeners; lower proficiency listeners > high proficiency listeners. The comparison between native and non-native English listeners suggests that the former successfully used their recent exposure to Bulgarian-accented and native English to speed up their reaction times in same-accent blocks, while the native English listeners achieved ceiling results with native English speech and did not adapt to Bulgarian-accented speech across speakers. However, that pattern was not observed in their accuracy adaptations.

There is generally stronger evidence that the higher proficiency listeners are specialised to process native English at the expense of Bulgarian-accented English, compared to the lower proficiency listeners. Despite its flaws, Experiment 1 in this dissertation showed no evidence of the higher proficiency listeners processing Bulgarian-accented words more efficiently than lower proficiency listeners. In fact, the opposite tendency was observed. Experiment 2 provides more conclusive evidence to that end. The specialisation of the higher proficiency listeners is observed both in their overall accuracy and reaction times and in the curve analysis of their initial response to the two accents at the start of an accent block.

In the terminology of IAF, specialisation can be expressed as “selective adaptation”, when the mean value of a specific phonetic cue has been shifted in the representations of the listener and the variance of that cue has narrowed down around the mean. Selective adaptation is observed after prolonged exposure to input with a different mean and narrow variability than the existing phonetic representations (Vroomen et al., 2007). The adaptation of the high proficiency listeners to SBE is unsurprising, as a result of their immersion, but the fact that it happens at the expense of Bulgarian-accented English needs to be explained with selective adaptation (as opposed to recalibration, which leads to flexibility).

In the ASP model, listener specialisation is explained with SPRs, which means that listeners pick out only some of the cues in the input when they are specialised to the accent. In IAF that can be “translated” as the listeners “tracking” the mean and variance of only a few of the cues present in the speaker’s speech as opposed to many cues. However, this type of specialisation cannot account for evidence of accent preferences observed in the so-called “phonetic” (or attentive) type of listening, when listeners attend to more cues than normal (e.g., in the start of an accent block). In the present dissertation this type of specialisation is observed in the curve analysis at the start of a new accent block.

IAF, however, can account for specialisation (or selective adaptation) even when many cues are tracked in attentive listening by positing that the expected variability of the phonetic cues is narrowly distributed around their native English means. This is a form of assigning more or less weight to cues depending on how far the observed realisations are from the expected distribution. In summary, the high English proficiency of the Bulgarian listeners in the UK, is associated with more exposure to (or engagement with) the prestige Standard British English speech. As a result of this exposure or engagement, they selectively adapt to many of its phonetic cues for English speech processing. This means that they have shifted the expected means and variance of the majority of phonetic cues in English to match those observed in the speech of prestige-variety native speakers, at the expense of Bulgarian(-accented) speakers. It is an open question whether the starting point of their cue distributions is determined by the variability they have observed in Bulgarian and Bulgarian-accented English (from Bulgarian teachers of English) or just the latter (see discussion on SLM and PAM). In the process of reaching selective adaptation to variability B from a starting point of selective adaptation to variability A, the listeners must pass through a stage of recalibration, when they expand the expected variance of A to include B. This is likely the stage in which the lower proficiency listeners in this dissertation.

The data in the present experiment contradict the initial predictions because it was expected that the lower proficiency listeners would be more selectively adapted to Bulgarian accent, while high proficiency listeners would be more flexible due to their overall greater experience with English. The data show the opposite. The highest proficiency listeners appeared selectively adapted to SBE at the expense of BA, when compared to the lowest proficiency in English in the sample, who overall appear flexible with regard to Bulgarian-accented and Standard British English. Overall, the latter do not process Bulgarian-accented English faster or more accurately than native English. Hence the data for the lowest proficiency listeners in the sample are less informative, but it might also be evidential to a lack of specialisation. They may have recalibrated their phonetic categories to include the variance observed in native English and Bulgarian-accented speech. If the participants with the lowest English proficiency in the present sample were compared to beginner learners of English, the beginners would have been more specialised to Bulgarian phonetics and the current participants would have also performed as more flexible than them (similar to what was observed in Ludwig and Mora, 2017).

While this analysis may account for the long-term and overall tendencies in accent adaptation observed in the present data, IAF should also be able to explain the short-term adaptation in the reaction times of the participants across accent blocks. One problem of using IAF to account for the present data is that it is generally concerned with the outcome of speech recognition. In that respect IAF can make direct predictions about processing accuracy results or phoneme identification and discrimination. However, IAF does not directly explain why after the end of any auditory input listeners need different amounts of time to achieve lexical access. It is mentioned that listeners may need time to eliminate possible competing candidates for selection but there is no specific time course for that. In psycholinguistic literature it is usually assumed that any extra time beyond the minimal time required to engage with a basic auditory stimulus is used to achieve lexical access. The fact that lexical access occurs several hundred milliseconds after the end of the auditory input is not explicitly addressed by most of the models reviewed so far. If the roughly 150-100 ms required for pressing a button (Shinya et al., 2015) are subtracted from the observed RTs and the approximately 60-70 ms it takes for the listener to achieve a basic sensation of the last auditory stimulus at the end of the word (Efron, 1967), there are between 100 and 900 ms of processing time that need to be accounted for (based on the observations in the present experiments).

The delay in lexical access (at least in single-word out-of-context processing) suggests that after the end of auditory input listeners use their short-term memory to consolidate the input they had just received and engage in elimination of competing models for selection (suggested by IAF and TBIM and Calabrese, 2012). The role of consolidation is highlighted by studies like Xie et al. (2018) which report that a night's sleep after initial exposure might lead to shorter processing times on subsequent exposure to the same accent type (see Flores, 2010 for a related discussion in the context of syntax and attrition).

The data in the present experiment suggest that at the start of a new accent block the high proficiency and native listeners are faster than average when processing SBE speech and slower than average for Bulgarian-accented speech. If it is assumed that the initial exposure to a new accent triggers attentive listening, then the listeners should be picking up on more phonetic cues at the start than later in the block and that leads to faster RT performance. The fast processing of native English speech by high-proficiency and native listeners at the start compared to the middle of the block is inconsistent with Strange's (2011) prediction for phonetic vs. phonological listening. Strange (2011) states that phonetic listening is more costly and time consuming than phonological listening. A similar contradiction was mentioned in the discussion of Calabrese's (2012) model. However, as suggested earlier in the IAF discussion, if the high proficiency listeners are already selectively adapted to the native Standard British English accent, then perceiving a greater amount of phonetic information that correctly matches their narrow predictions should logically lead them to reach certainty faster and have faster reaction times, compared to situations in which they have tracked fewer matching phonetic features. This point will be elaborated in the following discussion of TBIM.

Hence both the concepts of selective listening routines (tracking few vs. many phonetic cues), introduced by Strange (2011), and selective adaptation (expecting small vs. big variation for the cues) are required to account for the observed data. Using only one of them is contradicted by the data. It is hereby proposed that the IAF can be extended to include different attentional modes of listening (described in Calabrese, 2012; Strange, 2011; Werker and Logan, 1985).

This conclusion is applied to the listeners in this experiment and is assumed that everyone engages in attentive listening at the start of a block due to novelty and switches to more efficient listening for the rest of the block. Listeners who are adapted to the accent are expected to perform faster than average at the start compared to listeners who are less adapted to the accent. The left side of the heatmaps in Section 7.4.3 suggests that the lowest proficiency listeners are most specialised in BA English while the highest proficiency listeners are least specialised in it, compared to listeners with average proficiency, the opposite tendency being observed for SBE. The heatmaps analyse each accent independently of the other, hence it is still consistent with the earlier interpretation that the lowest proficiency listeners do not have well-differentiated specialisation between BA and SBE. Their lack of specialisation (or state of recalibration) is what gives them an advantage with BA English over higher proficiency listeners, who first need to recalibrate from SBE before they can adapt to BA.

In summary, the Ideal Adapter Framework can account for all the data observed in this dissertation but only when it is supported by the Automatic Selective Perception model. There is evidence that the high proficiency L2 listeners, similarly to the native English listeners are selectively adapted to a prestige native English variety over Bulgarian-accented English. Overall, the listeners with the lowest English proficiency in the sample appear to have flexible representations of both Bulgarian-accented and Standard British English. Further research is required to investigate the negotiation between long-term and short-term specialisation.

9.2.4. Dynamic approaches: Thousand Brains Intelligence Hypothesis

This section will discuss how aspects of speech processing models that were already discussed and deemed useful in accounting for the observed data fit the TBIM (Hawkins et al., 2017, 2019), which models the underlying neural reality of sensory processing. The purpose is to provide a rough sketch for a non-expert in neuroscience on how some of the processes discussed so far can operate of a neural level, using the Thousand Brains Intelligence Model (Hawkins et al., 2019).

TBIM assumes that different parts of the sensory input (e.g., the abstract concept of cortex columns, each of which is connected to a different part of the retina) create complete models of the perceived category. This means that the already activated models depolarise other cells that are consistent with complete models of that category, which acts like preactivation of a representation (Hawkins et al., 2017). The final perceived object is the one model which has the highest probability among all other competing models. For example, if one is observing the image in Figure 37 and the viewer is not particularly biased by their environment, there would be two major competing models: a rabbit and a duck. If the eyes fixate on the right, a larger portion of the retina cells would be modelling ducks, while if the eyes fixate towards the left, more of the retina cells would be giving input for rabbit models, although input from the peripheral vision might still be contributing to a few instances of duck models. This is possible because each stream of sensory input is location tagged with respect to the model of the whole object. In either case the viewer concludes that they are seeing either a rabbit or a duck but never a hybrid, which is similar to categorical perception in speech processing.

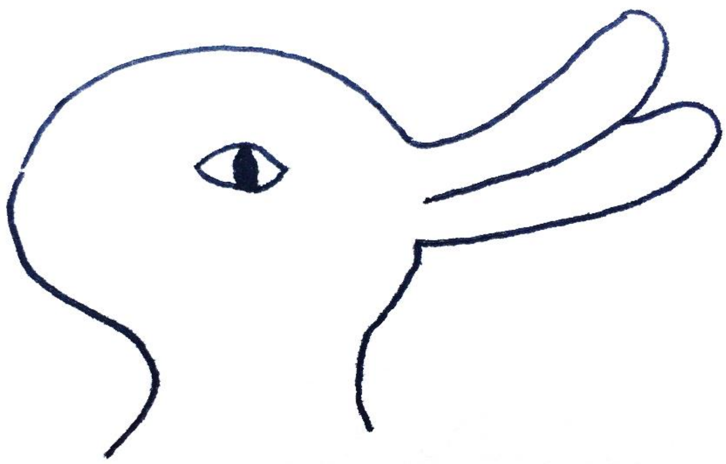


Figure 37. The rabbit-duck optical illusion, drawn by the author, inspired by Anonymous Illustrator (1892).

The perception of an image and of a word would differ in many situations. In one case the input is more likely to be static, and the sensory organs are free to move and explore and reinforce their certainty. In the other case the input is ephemeral and time conditioned, while the hearing organs have only one opportunity in time to encounter the raw input. However, if the auditory sensory input is also location-tagged, with respect to the order of the memories of what was just heard and the predictions of what is expected next (using the help of silence duration, breathing patterns, intonation and prosodic markers, voice quality and certain acoustic signals predicting each other), then at every new time point in the unfolding of speech, new competing models of what is heard are created.

One of the ways in which TBIM accounts for the processing of more input at a single time point (as in attentive or “phonetic” listening) is by involving more cortical columns. According to the canonical example of Hawkins et al. (2019), if the sensory input from one finger trying to identify a cup is processed by one cortical column, the finger might need to touch the cup in three different places before recognition occurs. That way a single cortical column would require more input over time to eliminate competitors like bottle and vase and finally recognise the object. This is similar to the prediction of Kleinschmidt and Jaeger (2015a) on speech processing by flexible listeners. However, if three fingers touch the cup at different places at the same time and three columns are involved simultaneously, then the three columns would be able to converge on a winning model in a shorter amount of time.

When applying this example to auditory processing it could be hypothesised that due to increased attention at each time point, more cortical columns are involved in the sampling of the auditory input. This could be a formal way of expressing the phonetic (or attentive) style of listening described by Strange (2011) and Calabrese (2012). However, there is a contradiction because Strange (2011) and Calabrese (2012) assume that phonetic listening is more costly in terms of effort and therefore requires more consolidation time, while TBIM predicts that involving more cortical columns will be more time efficient.

This contradiction is resolved by exploring the underlying assumptions of the two models. Strange (2011) assumes that phonetic listening involves the processing of a greater volume of information, while TBIM assumes (at least for the somatosensory system) that the same amount of information is processed in parallel instead of sequentially, by using three fingers to touch three surfaces simultaneously instead of sequentially by using only one finger. Taking these assumptions into account, there does not need to be a contradiction in the integration of these two models. However, it is still not clear to what extent attentive listening can benefit from this type of efficiency, as the production and auditory perception of a spoken word inherently unfolds in time compared to the visual or somatosensory processing of an already existing physical object, which could happen in one instant.

As the listeners are paying more attention and track more phonetic cues at each unfolding time point, that would correspond to a person using three fingers simultaneously to identify a mug instead of just one finger. Although a word unfolds in time, parts of it (perhaps clustered around acoustic landmarks, as proposed by Stevens, 2002), rich in phonetic information, are processed simultaneously and they all lead to more specific predictions, regarding the following parts of the input. An adapted perceiver might be able to recognise a cup with a single touch of one finger, but even they will reach the point of recognition faster when using three fingers simultaneously. Similarly, an experienced listener can recognise a familiar word by focusing on a smaller proportion of the phonetic input (e.g., the efficient listening of native English as in the middle of the speaker block, Section 7.4.3) but if they use more attentional resources they can recognise the same familiar accent at a faster rate (e.g., attentive listening of native English by highly proficient and native listeners as in the start of an accent block, Section 8.3.3).

As discussed in the previous section, attentive listening might be more time consuming if the input itself is not easily predictable. Hard to predict input (e.g., a foreign language or a foreign accent) might lead to attentive listening and hence more time delays, but that should not be inherent to the mode of listening itself. Attentive listening should be possible and highly efficient even with highly perceptually coherent input, as demonstrated with the rapid RTs at the start of a listening block in Section 7.4.3 and 8.3.3.

The overall slower reaction times and lower accuracy for the Bulgarian-accented words in the present experiment suggests that this variety of English introduced more unpredictability for the listeners. As discussed in the literature review, speech noise (unpredicted speech sounds) would lead to the activation of unpredicted cortical columns, which involves the activation of more neurones than if just the predicted column was activated. The additional cell activations that result from a “burst” simply mean that the process of recognition is more costly, it does not mean that the listener has recognised additional words, which need to be inhibited. This explains why the high proficiency listeners are slower with the Bulgarian accent without having to resort to the explanation discussed with BIMOLA in Section 9.1.1, that preactivation of Bulgarian lexical items slows down recognition. It is still possible that Bulgarian words are preactivated to a small extent in a fashion similar to Ju and Luce (2004) as the stimulus unfolds, but they are unlikely to affect the speech processing of high and low proficiency listeners differently (if it is assumed that they have the same proficiency in Bulgarian).

This experiment also gives an indication as to the size of phonetic unit that the listeners used in this experiment in order to adapt to the within- and across- speaker variation. The fact that no words were repeated, the words were monosyllabic and rapid adaptation of RTs occurred (see Section 7.4.3 and 8.3.3), means that the listeners used some level of perceptual compositionality smaller than a syllable to adjust their predictions for future phonetic cue identifications. One of the advantages of the TBIM is that it does not rely on specific abstract auditory units such as syllables or phonemes. TBIM accounts for this learning as a process of refining which cells are preactivated, so that they can be accessed faster when the matching input becomes available. Such units may end up being recognised in some listening situations but if the aim is to recognise a whole word then the model does not require the occurrence of intermediate steps of recognising individual meta-linguistic segments (Coleman, 2002; Hawkins and Smith, 2001). This allows the correct processing of completely or partially degraded input, such as speech in noise or hearing an unpredictable accent.

TBIM does not have a direct explanation for the fact that the lower proficiency listeners were overall a little faster and a little more accurate with Bulgarian-accented English than the higher proficiency listeners. Of the models reviewed so far, SLM (Flege, 1995) provides an explanation for that by suggesting that lower proficiency listeners have overlapping phonetic categories (e.g., expectations of phonetic realisations) between L1 and L2, compared to higher proficiency listeners, which means that Bulgarian-accented sounds are considered good examples of the respective phonetic categories. This was also the explanation offered in the previous section using the concepts of perceptual recalibration on a phonetic as opposed to phonemic level from IAF.

One way that selective adaptation and perceptual recalibration can be accounted for in TBIM is by assuming that different rates of cortical column preactivation. In selective adaptation, a smaller range of cortical columns would become preactivated. They become active only through input closer to the mean (prototype). In perceptual recalibration, a wider range of cortical columns is preactivated, increasing the chance of activation in the presence of a greater range of input. As suggested in the literature review, in order to explain speech processing, TBIM has to incorporate a complex network of specialisation and attention distribution, in order to account for grammatical and phonotactic processes. While that is beyond the scope of this dissertation, if TBIM can be used to account for long-term adaptation to a specific accent, then there is the potential for the model to be extended to other types of specialisation.

Overall, this section provided a possible neurological explanation of otherwise abstract model predictions (selective adaptation, recalibration, selective perception routines and attentive listening). It was argued that the listeners in the present experiment can rapidly update their expectations about the realisation of input smaller than a syllable, in out-of-context single word processing.

9.3. Revisiting the Perceptual Distance Hypothesis, Different Processes Hypothesis, Interspeech Intelligibility Benefit hypothesis

Using the framework of general speech processing discussed in the previous section and the results from the studies in this dissertation, this section will review the three hypotheses discussed in Section 3.2: Perceptual Distance Hypothesis (PDH), Different Processes Hypothesis (DPH), Interspeech Intelligibility Benefit hypothesis (ISIB).

9.3.1. Different Processes Hypothesis

In summary, the Perceptual Distance Hypothesis predicts that as different accents gradually diverge from the listener's native accent, the listener's efficiency in processing them to them will decrease. Contrary to that prediction, the Different Processes Hypothesis assumes that listening to native accents and foreign accents requires completely different cognitive processes because the former diverge in predictable ways while the latter require major adaptations.

Firstly, the problem of taking a native listener perspective and the problem of drawing boundaries between dialects and languages has already been addressed in Section 2.1. It is the uncertain status of Bulgarian-accented English for the Bulgarian-English bilinguals that makes the exploration of these two hypotheses so interesting. On the one hand, the Bulgarian accent is likely characterised by a phonetic similarity to the native language of the listeners. This gives it the potential to be perceptually treated like a native accent according to the Different Processes Hypothesis. On the other hand, the native Standard British English variety is consistently used as a model in English teaching in Bulgaria, is the dominant prestigious variety in the country of residence of the participants and is potentially characterised by a more predictable and stereotypically recognisable acoustic variation than Bulgarian-accented English. Hence, there is a possibility that native Standard British English speech is processed as a "native" variety by the bilinguals at the expense of Bulgarian-accented English according to DPH.

The results of this experiment offer mixed to negative support to the DP hypothesis. A case can be made for DP when considering the accuracy and the reaction results of the high proficiency bilinguals only. Both suggest that the Standard British English productions are systematically more predictable for the high-proficiency bilinguals than BA. Also, at the start of a new accent block the high proficiency listeners process SBE faster than the overall average, while they need several seconds to adjust to Bulgarian-accented speech. As discussed in the previous sections, it is possible that the higher proficiency listeners have a long term selective adaptation to SBE and they do not require increased attention to adapt to it as they do after an initial exposure to Bulgarian-accented English (which might explain the results of Goslin et al., 2012). In addition, the native listeners are not able to achieve cross-speaker adaptation for Bulgarian-accented English, while having ceiling performance with native English speech.

However, the adaptation results show that the listeners adapt to speakers of all accents within the first six to eight words. This contrasts with Floccia et al. (2009) whose native English participants' reaction times did not return to baseline levels when listening to a French accent of English. The lower proficiency Bulgarian-English listeners do not systematically differentiate between the two accents in terms of reaction times and accuracy. This suggests that as the listeners' proficiency gradually changes, so does the status of L1-accented speech. When proficiency is viewed as a continuous variable, the status of different accents appears to change gradually too. This interpretation of the data is more supportive of the Perceptual Distance Hypothesis. To shed a more definite light on the matter further studies, similar to Porretta et al. (2016), need to test the hypothesis by including stimuli with gradually varying accent strength. The stimulus preparation reported in Experiment 1 can serve as a starting point in design of such stimuli.

Overall, it appears that listening to an L1 accent within L2 is a more complex and dynamic process than can be comfortably captured by the DP hypothesis. The potential for so much variation is better captured by the IAF and TBIM, discussed in the previous section. The overall evidence suggests that the Different Processes Hypothesis is not supported in this case: mean reaction times within a speaker can be achieved for both SBE and Bulgarian-accented words from Bulgarian-English listeners of all proficiency levels. Further studies are required to determine whether hearing an unfamiliar (native or non-native) accent in L2 would follow the same pattern of adaptation.

9.3.2. Perceptual Distance Hypothesis

Of the experiments reported here, Experiment 1 uses stimuli whose foreign accent strength in English varies in a more gradient manner (as they were all produced by the same L1 Bulgarian - L2 English speaker), and whose results could therefore be used to evaluate the Perceptual Distance Hypothesis. Unfortunately, as stated in Section 6.5, the methodological flaws of Experiment 1 make its results unreliable. Nevertheless, it might be useful to consider them, with appropriate caution.

In the condition of Bulgarian primes and English targets, lower proficiency listeners respond faster to targets with a stronger Bulgarian accent than targets with a more Standard British English sound. This tendency is reversed for higher proficiency listeners. This result mirrors the finding from Experiment 2, where the two types of accents are produced by native and non-native speakers of English. It does not matter whether the Standard British English accent is actually produced by a native speaker or a high proficiency L2 speaker of English (cf., Hayes-Harb et al., 2008). As long as certain acoustic patterns are consistent enough with Standard British English or with Bulgarian phonetics (as judged by native English listeners from the UK), the higher proficiency listeners will prefer the former and will be slower with the latter. This conclusion is more in line with the Perceptual Distance Hypothesis than the Different Processes Hypothesis.

As the PDH and DPH hypotheses are usually discussed in opposition to each other, the previous section has already pointed out that PDH has stronger support in this dissertation. The gradient effect of proficiency and the gradual adaptation to a speaker's variation suggests that the listeners adjust to speech incrementally in the short- and long-term. This is consistent with the IAF prediction that listeners track the mean and variance of the phonetic cues they are exposed to. In cases that might appear as different processes, such as the native listeners' inability to adjust their RTs across speakers is more likely due to their lack of relevant representations for Bulgarian-accented speech, which they had to build up on the spot. As suggested by Xie et al. (2017) and Kriengwatana et al. (2016) accent adaptation starts with a speaker adaptation, and may not occur immediately if the speakers do not sound holistically similar. Hence, it is a temporary state, that could gradually change with more exposure, and sleep (Xie et al., 2018).

More research is needed comparing different degrees of Bulgarian-accentedness to fully evaluate the Perceptual Distance Hypothesis.

9.3.3. Interspeech Intelligibility Benefit Hypothesis

Unlike the previous two hypotheses, the evaluation of the Interspeech Intelligibility Benefit Hypothesis is more straightforward because the hypothesis was developed for non-native listeners in the first place.

No overall ISIB for talkers is observed for the participants in Experiment 2. This means that the Bulgarian-accented words are not recognised more accurately than the native English words overall. In fact, the recognition of Bulgarian-accented words tends to be more inaccurate than the recognition of native English speech. In that respect the study's findings are aligned with Hayes-Harb et al. (2008) and Munro et al. (2006) who found no overall ISIB-T. Even Bent and Bradlow (2003) define an intelligibility benefit for L1 accent processing, as a lack of a disadvantage compared to native speech processing. Hence, they also have no evidence of L1 accent processing benefit as defined in this dissertation. Hayes-Harb et al. (2008) do report that some benefit is observed over native speech processing but only for low proficiency listeners listening to low proficiency L1-accented speech. In that context, it is conceivable that if lower L2 proficiency listeners had participated in the present experiment (with a score under 53% on LexTale), they could have demonstrated a benefit of Bulgarian accent over native English accent.

ISIB-L means that non-native listeners have an advantage when processing interspeech compared to native listeners. This hypothesis has received more support than ISIB-T (Bent and Bradlow, 2003; Munro, 1998; Smith et al., 2003; van Wijngaarden, 2001). In the present experiment there is mixed evidence for ISIB-L. On the one hand, the lower proficiency listeners were more accurate in processing Bulgarian-accented English compared to higher proficiency listeners. On the other hand, on the high end of the proficiency spectrum there is no L1-accent benefit for listeners (ISIB-L) compared to native UK listeners. Experiment 2 shows that high proficiency Bulgarian-English bilinguals process the two experimental accents with comparable accuracy and speed to native English listeners, which contradicts some of the existing literature. However, without consistent measures of L2 proficiency across studies it is difficult to directly compare the results. As suggested in the previous section, if listeners with very high L2 proficiency have also undergone selective adaptation to their L2's phonetics, then it is possible that their listening performance in a lexical decision task might resemble that of native listeners and thus, have no ISIB-L. However, the results suggest that this type of selective adaptation is dependent on proficiency, hence, beginner L2 listeners are predicted to have matched ISIB-L with respect to native listeners. These results suggest that accent benefit (or disadvantage) is always relative to the participants in a communicative situation and hence there are limits to the generalisations that can be drawn from studies on ISIB, including this one.

In addition to these concerns, studies such as Munro et al. (2006) Major et al. (2002) and Jensen and Thøgersen (2017) suggest that when the bilingual processing of L1-accented L2 is compared to the processing of unfamiliar foreign L2 accents, the interpretation of the results is even more complex. These studies provide evidence that in some cases speech with a foreign accent different from the listeners' might be processed faster or more accurately than their native accent.

Overall, no ISIB-T is observed in this study and ISIB-L exists only when comparing bilinguals with different proficiencies. Considering the variety of mixed results reported on the topic, the results of this dissertation are consistent with at least some of the literature. Nevertheless, the theoretical account of the present results suggests that the variability in the literature might be in part caused by the inconsistent measures of the listeners' L2 proficiency. Considering the complex interactions between listener and speaker proficiency in L2 and their L1 background, the results of the present study should be treated only as a steppingstone into the exploration of how Bulgarian-English bilinguals process Bulgarian-accented English. Future work can reverse the design of this study and test how several more homogenous proficiency groups respond to gradient changes in Bulgarian-accented English.

9.4. Further directions

Considering the controlled setting of the experiments, there is ample scope for future research that can test the conclusions of this dissertation in novel settings. Experiment 2 has been purposefully designed to be easily replicable with other language pairs.

If the study were to be replicated, however, there are some modifications that could be implemented. Firstly, in order to exclude the possibility that the observed proficiency effect is not also driven by phonetic attrition in Bulgarian, the L2 RT and accuracy results of the participants need to be compared to their performance in their native language in a similar lexical decision task. If increase in English proficiency is associated with slower and less accurate responses to Bulgarian speech, then listeners who have little phonetic attrition in Bulgarian and high proficiency in English might potentially process Bulgarian-accented speech as efficiently as lower proficiency listeners who also do not have Bulgarian phonetic attrition. This would indicate that the selective adaptation observed in this study is not a result of less exposure to Bulgarian-accented phonetic variation in English but less exposure to Bulgarian and suggest directly that even high-proficiency L2 listeners rely on their L1 phonetics. Hence, it is possible that the selective adaptation to Standard British English is not an inherent outcome of increased L2 proficiency. In addition, it needs to be explored to what extent strong L1 phonetic attrition and high L2 proficiency co-occur among expats in the UK.

One of the main conclusions of this study is that listeners who are engaged in attentive listening (or "phonetic listening" in Strange, 2011) to a familiar variety will have faster word recognition than if they are engaged in less attentive listening (called "phonological listening" by the same author), hence demonstrating that attentive listening is not inherently slower, and that it mostly happens to be employed for speech that is not familiar to the listener. The following is a suggestion for an experiment that might help disentangle these questions.

In a similar experimental design to the one used for the main study, the listeners' pupil dilation can be used as an independent measure to detect periods of high and low attention (Wainstein et al., 2017). It is expected that when they are listening to a well-known variety and during periods of high attention, signalled by dilated pupils, the listeners will respond with faster RT and higher accuracy, compared to baseline periods of lower attention.

If high attention is induced through the introduction of a new speaker of the same well-known variety, then fast RTs and high accuracy should still be observed (inspired by experiments with infants' adaptation using rate of sucking responses of Eimas et al., 1971). The same prediction should be valid for listeners hearing a less known variety, such as a foreign accent, but only after familiarisation with the speaker and with high attention levels. During the familiarisation process to a new speaker and variety, the listeners will use more attention and they will have a decline in their response times.

There are neurological studies which demonstrate that listeners use different amounts of cognitive resources depending on the variety they are exposed to (Abutalebi, 2008; Goslin et al., 2012; Song and Iverson, 2018). The question needs to be investigated simultaneously with the listeners' RTs in order to shed more light on the interpretation of this more accessible method of investigation. If there is specific evidence on the effects of increased attention during the processing of well-known and less familiar varieties on RTs, this will need to be taken into consideration for models of general speech processing.

As proficiency appears to be an important factor in bilingual speech processing, BIMOLA and perhaps even ISIB are less suitable frameworks to pursue, as they do not account directly for listener proficiency or dynamic changes in experience. However, the results of this dissertation can be used to make more detailed predictions testing some of the assumptions of SLM and PAM. The theories can be further explored through the lens of Bulgarian-English bilinguals with different proficiency in English process English words containing specific phones, such as stressed /i/ or /ɪ/. For the sake of specificity and based on the performance in Experiment 2 low proficiency is defined as less than 55% score on LexTale and high proficiency is defined as more than 85% score on LexTale.

The following hypotheses test the SLM and PAM predictions that low proficiency listeners might rely on L1 equivalents when processing L2 phones and that as they increase in proficiency, they develop distinct representations of phones for their L2.

1. Low proficiency bilinguals will struggle to discriminate between minimal pairs with /i/ and /ɪ/ (e.g., “beet” and “bit”), provided that the vowel duration is kept constant, compared to high proficiency listeners, as the formant variability of both vowels is associated with a single vowel in Bulgarian.

2. Low proficiency listeners will process words containing stressed /i/ (e.g., “beet”) faster than the minimal pair counterpart containing /ɪ/ (e.g., “bit”), as its formant variability is closer to the formant variability of a prototypical Bulgarian stressed /i/.

3. High proficiency listeners will not have systematically faster reaction times for either /i/ (e.g., “beet”) or /ɪ/ (e.g., “bit”), provided that other characteristics of the words, such as frequency and imageability, are constant.

4. If English words containing stressed /ɪ/ (without minimal pair counterparts with /i/, such as “link”) are presented to Bulgarian-English bilinguals and the vowel /ɪ/ is replaced with [i], high proficiency listeners will have lower accuracy and slower RT (also perhaps higher entrainment, larger pupil size and a presence of P300 effect in an EEG investigation) when compared to a standard realisation of the word and when compared to low proficiency bilinguals.

IAF does account for listeners with different levels of adaptation to the source of variability (e.g., accent, or speaker), hence the results from the present experiment can lead to the following hypothesis.

5. As demonstrated by the GAMM analyses in Experiment 2, it is expected that both high and low proficiency listeners will dynamically update their representations of the vowel. At the start of a block of trials, high proficiency listeners, would process an [i] replacement in a word like “link” slower than low proficiency bilinguals, because it is expected that the former are selectively adapted to [ɪ]-realisations, and the latter to [i]-realisations of /ɪ/. However, within 8-10 tokens of the target accent the difference between the two groups of listeners will diminish. Listeners with intermediate LexTale results (55-85%) are expected to initially respond faster to [liŋk] than high proficiency listeners because as flexible listeners they do not need to recalibrate from a [lɪŋk] expectation as the high proficiency listeners. However, they will initially require more time to selectively adapt their responses to a [liŋk] production compared to the low proficiency listeners who are expected to already be selectively adapted to that.

Another important future path of research, highlighted by this dissertation, is the need for distinction between listener familiarity (high and low proficiency) and the accent of the speech (native or L2). That may be achieved, for instance, by presenting Bulgarian low-intermediate learners of French (expected to be in a “recalibration stage”) with four types of accents: standard Parisian French (native familiar), Bulgarian-accented French (non-native familiar), Canadian French (native unfamiliar), Dutch-accented learner French (non-native unfamiliar). Such experiment may reveal whether there is any inherent advantage to speech being produced by a native speaker, even when they speak a less familiar variety of the L2, compared to a non-expert, non-native speaker of a different L1 background. If both native and non-native varieties are equally unfamiliar to the listener, any advantage of native speech might be potentially attributed to a greater consistency in the realisation of important phonetic cues.

In the introduction of this dissertation it was stated that there is no complete model of speech processing, due to the complexity of the process and the large number of factors influencing it. The future directions for research, outlined in this section, suggest that the area of L1-accented L2 processing can answer fundamental questions about how the phonetic systems of multiple languages can interact during listening and to what extent selective adaptation in an inevitable side effect of increased proficiency in L2. Additionally, future research needs investigate the separate roles of the listeners' prior experience with a variety and the present effort they are putting into the listening process. These insights can be directly added to models of speech perceptions, such as TBIM and IAF. There are a large number or additional research paths that can follow from this dissertation, but this section highlights the topics which have direct implications on the validity of the conclusions drawn from this dissertation.

9.5. Practical implications

This dissertation has provided an exploration of the phonetic representations of Bulgarian-accented English for Bulgarian-English bilinguals with varying English proficiency living in the UK. It has led to the main conclusion that most of the participants find Bulgarian-accented speech disadvantageous for processing but are able to adapt to it during the experiment.

Although the experimental setting is highly controlled, using only high-frequency words, it is worthwhile to consider what might be the practical implications of these results for Bulgarian-English bilingual listeners in their everyday life as residents in the UK. Firstly, the overall accuracy rate with Bulgarian accented words is just over 90%. According to Bonk (2000) at least 80% vocabulary knowledge is required to achieve holistic auditory text comprehension. However, according to Stæhr (2009) the number is probably 98%. This dissertation demonstrates that as long as the vocabulary is appropriate for the listeners, Bulgarian accent should not hinder the overall comprehension. In addition, even the high proficiency listeners who appeared to be specialised in Standard British English, managed to adapt to Bulgarian-accented speech within and across speakers. Of course, this is a preliminary conclusion, and it needs to be validated using a larger variety of Bulgarian-accented speech.

Yet, a number of studies demonstrate that in optimal listening conditions listeners might respond in a comparable way to a several accents, but when the same types of speech are heard in noise, the speed and accuracy of processing may be seriously affected (Clopper and Bradlow, 2008; Munro, 1998). Hence, it important to also investigate the speed and accuracy of processing Bulgarian-accented speech in a more noisy, naturalistic setting.

Overall, the data in this dissertation suggest that in the interest of creating optimal listening conditions for Bulgarian-English bilinguals, living in the UK, a Standard British English accent should be favoured over Bulgarian-accented speech.

9.6. Conclusion

The main question raised in this dissertation is whether it is possible for L2 listeners of English to process their Bulgarian accent in English faster and more accurately than a Standard British English accent. The results from the experiments reported here suggest that the answer is no. However, this is not the case for all L2 listeners. All results point towards the conclusion that the Bulgarian-English bilingual listeners' proficiency in English decreases, their processing of two speakers' Bulgarian-accented English improves.

This result is consistent with some of the literature, which contains examples of both facilitatory (Hanulíková and Weber, 2012; Hayes-Harb et al., 2008; Ludwig and Mora, 2017) and inhibitory effects of processing L1 accent within L2 (Lagrou et al., 2011; Weber et al., 2014). Some of the potential explanations for the combination of facilitatory and inhibitory effects, considered in this dissertation, are related to the phonetic similarity between Bulgarian-accented English and Bulgarian (Best and Tyler, 2007; Bundgaard-Nielsen et al., 2011; Flege, 1995), and the listeners' experience with English, measured with LexTale as the their proficiency.

The results of this dissertation suggest that the similarity between a Bulgarian accent and Bulgarian phonetics is insufficient. If the listeners have access to Bulgarian phonetics while listening in English (as suggested by Blumenfeld and Marian, 2007; Marian et al., 2003), all of them should have benefitted from the Bulgarian accent. This is contradicted by the results in the present study which demonstrate consistent interactions between accent and listener proficiency, showing disadvantages of L1 accent processing for the highest proficiency listeners. In addition, listeners with higher proficiency in English would have been expected to have a greater general exposure to different varieties of English, including Bulgarian-accented English. This is not observed in the present study. The highest proficiency listeners performed worse than lower proficiency listeners and on par with native English listeners in BA processing. The evidence suggests that as the Bulgarian-English bilinguals' proficiency in English increases, they specialise in processing Standard British English at the expense of Bulgarian-accented English, by gradually fine-tuning their phonetic expectations for SBE speech. This process is consistent with the predictions of the Automatic Selective Perception model (Strange, 2011) and the Ideal Adapter Framework for phonetic recalibration and selective adaptation and the flexible and narrow listening that occurs as a result (Kleinschmidt and Jaeger, 2015a).

In addition to the evidence of long-term accent specialisation to SBE, this dissertation demonstrates that on a group level the bilingual listeners improved their reaction times and accuracy for Bulgarian-accented speech while listening to one speaker and improved their RTs across speakers. In contrast, native English participants demonstrated no RT and accuracy adaptation to either a BA or a SBE across speakers. This was explained through a ceiling effect for SBE speech and insufficient experience with BA. In addition, there was no evidence of L1-accent benefit for high-proficiency Bulgarian-English bilinguals over native English participants. Future research should continue to focus on variability within the Bulgarian-English bilingual listeners and test their selective adaptation to fine phonetic detail by tracking several phonetic characteristics of the Bulgarian-accented speech, similarly to Experiment 1.

On a theoretical level, this dissertation has pointed out the similarities across several models of speech processing, which are representative of very different schools of thought. For instance, both the traditional approach in Calabrese (2012) and the more exemplar model of Strange (2011) make use of the "phonetic" and "phonological" listening modes proposed by Werker and Logan (1985). Although throughout the dissertation the terms "attentive" and "efficient" listening appeared more fitting, on the whole it was demonstrated that the concepts are an indispensable addition to a general model of speech processing. Those two modes of listening were accounted for by the Thousand Brains Intelligence model (Hawkins et al., 2019). The latter, in turn has commonalities with the Ideal Adapter Framework (Kleinschmidt and Jaeger, 2015b). For instance, both Kleinschmidt and Jaeger (2015a) and Hawkins et al. (2019) propose that listeners evaluate several different complete models of the perceived object (or word) before they reach an optimal decision based on a multitude of cues from the environment. TBIM places a strong emphasis on this aspect of processing, and even alludes to it in its name, while IAF instead focuses on how acoustic cues are systematised in the mind (using both the central tendency and the variance). Overall, this dissertation has demonstrated support for these commonalities between the models, but only when they are used together, as opposed to independently.

In order to successfully tackle the enormous complexity of speech processing, it is hoped that future research will continue to draw inspiration across disciplines and from traditionally opposing views within disciplines. As pointed out by de Bot (2012), a person's phonetic representations of speech are like an ever-changing river; this dissertation has been an attempt to study the river both by joining the flow and observing it from the side.

References

Aaltonen, O., Eerola, O., Hellström, Å., Uusipaikka, E. and Lang, A.H. (1997), "Perceptual magnet effect in the light of behavioral and psychophysiological data", *The Journal of the Acoustical Society of America*, Vol. 101 No. 2, pp. 1090–1105.

Abbs, J.H. and Sussman, H.M. (1971), "Neurophysiological Feature Detectors and Speech Perception: A Discussion of Theoretical Implications", *Journal of Speech, Language, and Hearing Research*, Vol. 14 No. 1, pp. 23–36.

Abutalebi, J. (2008), "Neural aspects of second language representation and language control", *Acta Psychologica*, Vol. 128 No. 3, pp. 466–478.

Adank, P., Evans, B., Stuart-Smith, J. and Scotti, S. (2009), "Comprehension of familiar and unfamiliar native accents under adverse listening conditions", *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 35 No. 2, pp. 520–529.

Adank, P., Hagoort, P. and Bekkering, H. (2010), "Imitation Improves Language Comprehension", *Psychological Science*, Vol. 21 No. 12, pp. 1903–1909.

Alcántara, J.I., Weisblatt, E.J.L., Moore, B.C.J. and Bolton, P.F. (2004), "Speech-in-noise perception in high-functioning individuals with autism or Asperger's syndrome", *Journal of Child Psychology and Psychiatry*, Vol. 45 No. 6, pp. 1107–1114.

Allport, A.D., Styles, E. and Hsieh, S.A. (1994), "Shifting attentional set – Exploring the dynamic control of tasks", in Umiltà, C. and Moscovitch, M. (Eds.), *Attention and Performance Series. Conscious and Nonconscious Information Processing*, MIT Press, Cambridge, MA, pp. 421–452.

Andreeva, B., Barry, W. and Koreman, J. (2013), "The Bulgarian Stressed and Unstressed Vowel System. A Corpus Study.", in Bimbot, F., Cerisara, C., Fougeron, C., Gravier, G., Lamel, L., Pellegrino, F. and Perrier, P. (Eds.), *Proceedings of 14th Annual Conference of the International Speech Communication Association*, presented at the Interspeech, Lyon, France, available at: http://www.isca-speech.org/archive/interspeech_2013 (accessed 9 May 2018).

Anonymous. (2015). "Виж кой говори вместо метрото /Vizh koy govori vmesto metroto.", *Trud Online*, available at: https://trud.bg/article-4743531/ (accessed 5 February 2019).

Anonymous Illustrator. (1892), "Welche Thiere gleichen einander am meisten?", *Fliegende Blätter. Braun & Schneider*.

Antoniou, M., Best, C.T., Tyler, M.D. and Kroos, C. (2010), "Language context elicits native-like stop voicing in early bilinguals' productions in both L1 and L2", *Journal of Phonetics*, Vol. 38 No. 4, pp. 640–653.

Atas, U. (2018), “The Role of Receptive Vocabulary Knowledge in Advanced EFL Listening Comprehension”, *TESL-EJ*, Vol. 21 No. 4, available at: https://eric.ed.gov/?id=EJ1172563 (accessed 5 February 2019).

Baayen, R.H., Piepenbrock, R. and Gulikers, L. (1995), “The CELEX Lexical Database (CD-ROM)”, *Linguistic Data Consortium*, University of Pennsylvania, Philadelphia.

Bachman, L.F. and Palmer, A.S. (1996), *Language Testing in Practice: Designing and Developing Useful Language Tests*, Oxford University Press, Oxford.

Baese-Berk, M.M., Bradlow, A.R. and Wright, B.A. (2013), “Accent-independent adaptation to foreign accented speech”, *The Journal of the Acoustical Society of America*, Vol. 133 No. 3, pp. EL174–EL180.

Basnight-Brown, D.M. and Altarriba, J. (2007), “Differences in semantic and translation priming across languages: The role of language direction and language dominance”, *Memory & Cognition*, Vol. 35 No. 5, pp. 953–965.

Bent, T., Atagi, E., Akbik, A. and Bonifield, E. (2016), “Classification of regional dialects, international dialects, and nonnative accents”, *Journal of Phonetics*, Vol. 58, pp. 104–117.

Bent, T. and Bradlow, A.R. (2003), “The interlanguage speech intelligibility benefit”, *The Journal of the Acoustical Society of America*, Vol. 114 No. 3, pp. 1600–1610.

Best, C.T. (1994), “The emergence of native-language phonological influences in infants: a perceptual assimilation model.”, in Nausbaum, H.C. (Ed.), *The Development of Speech Perception: The Transition from Speech Sounds to Spoken Words*, MIT Press, Cambridge, MA, pp. 167–204.

Best, C.T., McRoberts, G.W. and Goodell, E. (2001), “Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener’s native phonological system”, *The Journal of the Acoustical Society of America*, Vol. 109 No. 2, pp. 775–794.

Best, C.T. and Tyler, M.D. (2007), “Nonnative and second-language speech perception”, in Bohn, O.-S. and Munro, M.J. (Eds.), *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege*, John Benjamins Publishing Company, Amsterdam, pp. 13–34.

Best, C.T., Tyler, M.D., Gooding, T.N., Orlando, C.B. and Quann, C.A. (2009), “Development of Phonological Constancy: Toddlers’ Perception of Native- and Jamaican-Accented Words”, *Psychological Science*, Vol. 20 No. 5, pp. 539–542.

Bieber, R.E. and Gordon-Salant, S. (2017), “Adaptation to novel foreign-accented speech and retention of benefit following training: Influence of aging and hearing loss”, *The Journal of the Acoustical Society of America*, Vol. 141 No. 4, p. 2800.

Bien, H., Hanulíková, A., Weber, A. and Zwitserlood, P. (2016), “A Neurophysiological Investigation of Non-native Phoneme Perception by Dutch and German Listeners”, *Frontiers in Psychology*, Vol. 7, p. 56.

Black, F.H. (1958), *The American College of Sofia: A Chapter in American-Bulgarian Relations*, The Trustees of Sofia American Schools, Boston, MA.

Blake, A. (2018), “Population of the UK by country of birth and nationality”, *Office for National Statistics*, 29 November, available at: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/internationalmigration/bulletins/ukpopulationbycountryofbirthandnationality/july2017tojune2018 (accessed 4 June 2019).

Bloomfield, L. (1933), *Language*, H. Holt, New York.

Blumenfeld, H.K. and Marian, V. (2007), “Constraints on parallel activation in bilingual spoken language processing: Examining proficiency and lexical status using eye-tracking”, *Language and Cognitive Processes*, Vol. 22 No. 5, pp. 633–660.

Boersma, P. and Weernink, D. (2019), *Praat: Doing Phonetics by Computer*, available at: http://www.praat.org/ (accessed 3 January 2019).

Bonk, W.J. (2000), “Second Language Lexical Knowledge and Listening Comprehension”, *International Journal of Listening*, Vol. 14 No. 1, pp. 14–31.

de Bot, K. (2012), “Rethinking multilingual processing”, in Amaro, J.C., Flynn, S. and Rothman, J. (Eds.), *Third Language Acquisition in Adulthood*, John Benjamins Publishing Company, available at: https://benjamins.com/catalog/sibil.46.07bot (accessed 14 August 2018).

de Bot, K. and Jaensch, C. (2015), “What is special about L3 processing?*”, *Bilingualism: Language and Cognition*, Vol. 18 No. 2, pp. 130–144.

Boyadziev, P. and Tilkov, D. (1997), *Фонетика на българския книжовен език/Fonetika na balgarskiya knizhoven ezik*, Abagar, Veliko Tarnovo.

Bradlow, A.R. and Bent, T. (2008), “Perceptual adaptation to non-native speech”, *Cognition*, Vol. 106 No. 2, pp. 707–729.

Broersma, M. (2012), “Increased lexical activation and reduced competition in second-language listening”, *Language and Cognitive Processes*, Vol. 27 No. 7–8, pp. 1205–1224.

Browman, C.P. and Goldstein, L. (1992), “Articulatory phonology: An overview”, *Phonetica*, Vol. 49 No. 3–4, pp. 155–180.

Bruderer, A.G., Danielson, D.K., Kandhadai, P. and Werker, J.F. (2015), “Sensorimotor influences on speech perception in infancy”, *Proceedings of the National Academy of Sciences*, Vol. 112 No. 44, pp. 13531–13536.

Brysbaert, M., Keuleers, E. and New, B. (2011), “Assessing the Usefulness of Google Books’ Word Frequencies for Psycholinguistic Research on Word Processing”, *Frontiers in Psychology*, Vol. 2, p. 27.

Bulgarian Ministry of Education. (2019a), "Рейтингова система на висшите училища в България: филология/Reytingova sistema na visshite uchilishta v Balgaria: filologiya.", available at: https://rsvu.mon.bg/rsvu3/#RankingPlace:YFFfUoZcUmJUTkxK (accessed 4 July 2019).

Bulgarian Ministry of Education. (2019b), "Английски език за постигане на ниво B2/Angliyski ezik za postigane na nivo B2.", *Учебни програми по чужд език. Ниво B2/Uchebni programi po chuzhd ezik. Nivo B2.*, available at: https://www.mon.bg/bg/1698 (accessed 4 July 2019).

Bulgarian Ministry of Education. (2019c), "Учебници/Uchebnitsi", available at: https://www.mon.bg/bg/10 (accessed 6 October 2019).

Bundgaard-Nielsen, R.L., Best, C.T. and Tyler, M.D. (2011), "Vocabulary size is associated with second-language vowel perception performance in adult learners", *Studies in Second Language Acquisition*, Vol. 33 No. 3, pp. 433–461.

Bundgaard-Nielsen, R.L., Best, C.T., Tyler, M.D. and Kroos, C. (2008), "Evidence of a near-merger in Western Sydney Australian English vowels", *Proceedings of the 9th Annual Conference of the International Speech Communication Association*, presented at the Interspeech, Brisbane, available at: https://researchdirect.westernsydney.edu.au/islandora/object/uws%3A7700/ (accessed 8 June 2019).

Cai, Z.G., Gilbert, R.A., Davis, M.H., Gaskell, M.G., Farrar, L., Adler, S. and Rodd, J.M. (2017), "Accent modulates access to word meaning: Evidence for a speaker-model account of spoken word recognition", *Cognitive Psychology*, Vol. 98, pp. 73–101.

Cairney, S.A., Lindsay, S., Sobczak, J.M., Paller, K.A. and Gaskell, M.G. (2016), "The Benefits of Targeted Memory Reactivation for Consolidation in Sleep are Contingent on Memory Accuracy and Direct Cue-Memory Associations", *Sleep*, Vol. 39 No. 5, pp. 1139–1150.

Calabrese, A. (2012), "Auditory representations and phonological illusions: A linguist's perspective on the neuropsychological bases of speech perception", *Journal of Neurolinguistics*, Vol. 25 No. 5, pp. 355–381.

Calderón, J. and Best, C.T. (1996), "Effects of bilingualism on the perception of nonnative consonant voicing contrasts", *Journal of the Acoustical Society of America*, Vol. 99 No. 4, p. 2602.

Campbell-Kibler, K. (2007), "Accent, (ING), and the social logic of listener perceptions", *American Speech*, Vol. 82 No. 1, pp. 32–64.

Canale, M. and Swain, M. (1980), "Theoretical bases of communicative approaches to second language teaching and testing", *Applied Linguistics*, Vol. 1, pp. 1–47.

Caramazza, A., Yeni-Komshian, G., Zurif, E.B. and Carbone, E. (1973), "The acquisition of a new phonological contrast: The case of stop consonants in French-English bilinguals", *The Journal of the Acoustical Society of America*, Vol. 54 No. 2, p. 421.

Carroll, J.B. (1961), “Fundamental considerations in testing for English language proficiency of foreign students.”, *Testing the English Proficiency of Foreign Students*, Center for Applied Linguistics, Washington DC, pp. 30–40.

Central Intelligence Agency. (2017), “The World Factbook”, available at: https://www.cia.gov/library/publications/the-world-factbook/rankorder/2119rank.html (accessed 18 January 2020).

Chang, C.B. (2019), “Language change and linguistic inquiry in a world of multicompetence: Sustained phonetic drift and its implications for behavioral linguistic research”, *Journal of Phonetics*, Vol. 74, pp. 96–113.

Cheng, Y.-L. and Howard, D. (2008), “The time cost of mixed-language processing: an investigation”, *International Journal of Bilingualism*, Vol. 12 No. 3, pp. 209–222.

Cho, T. and M. McQueen, J. (2005), “Prosodic influences on consonant production in Dutch: Effects of prosodic boundaries, phrasal accent and lexical stress”, *Journal of Phonetics*, Vol. 33 No. 2, pp. 121–157.

Choi, G.-J. (1998), “The phonological value of the feature [palatalness] in contemporary Bulgarian language”, *Proceedings of the 12th International Congress of Slavic Scholars*, presented at the International Congress of Slavic Scholars, Krakow, Poland, available at: https://liternet.bg/publish1/choi/mekost.htm (accessed 13 October 2017).

Cieśla, K., Wolak, T., Lorens, A., Heimler, B., Skarżyński, H. and Amedi, A. (2019), “Immediate improvement of speech-in-noise perception through multisensory stimulation via an auditory to tactile sensory substitution”, *Restorative Neurology and Neuroscience*, Vol. 37 No. 2, pp. 155–166.

Clarke, C.M. and Garrett, M.F. (2004), “Rapid adaptation to foreign-accented English”, *The Journal of the Acoustical Society of America*, Vol. 116 No. 6, pp. 3647–3658.

Clopper, C. (2017), “Dialect Interference in Lexical Processing: Effects of Familiarity and Social Stereotypes”, *Phonetica*, Vol. 74 No. 1, pp. 25–59.

Clopper, C.G. (2014), “Sound change in the individual: effects of exposure on cross-dialect speech processing”, *Laboratory Phonology*, Vol. 5 No. 1, pp. 69–90.

Clopper, C.G. and Bradlow, A.R. (2008), “Perception of Dialect Variation in Noise: Intelligibility and Classification”, *Language and Speech*, Vol. 51 No. 3, pp. 175–198.

Clopper, C.G. and Tamati, T.N. (2010), “Lexical recognition memory across dialects.”, *The Journal of the Acoustical Society of America*, Vol. 127, p. 1956.

Coleman, J. (2002), “Phonetic representations in the mental lexicon”, in Durand, J. and Laks, B. (Eds.), *Phonetics, Phonology and Cognition*, Oxford University Press, Oxford, pp. 96–130.

Cooper, A., Fecher, N. and Johnson, E.K. (2018), "Toddlers' comprehension of adult and child talkers: Adult targets versus vocal tract similarity", *Cognition*, Vol. 173, pp. 16–20.

Costa, A., Miozzo, M. and Caramazza, A. (1999), "Lexical selection in bilinguals: Do words in the bilingual's two lexicons compete for selection?", *Journal of Memory and Language*, Vol. 41 No. 3, pp. 365–397.

Cutler, A., Weber, A. and Otake, T. (2006), "Asymmetric mapping from phonetic to lexical representations in second-language listening", *Journal of Phonetics*, Vol. 34 No. 2, pp. 269–284.

Dalrymple-Alford, E.C. (1985), "Language switching during bilingual reading", *British Journal of Psychology*, Vol. 76 No. 1, pp. 111–122.

Dauer, R.M. (1987), "Phonetic and phonological components of language rhythm.", *Proceedings of the 11th International Congress of Phonetic Sciences*, Vol. 5, presented at the ICPhS, Tallinn, Estonia, pp. 447–450.

Dienes, Z. (2008), *Understanding Psychology as a Science: An Introduction to Scientific and Statistical Inference*, Palgrave Macmillan, NewYork.

Dijkstra, A. and van Heuven, W. (1998), "The BIA model and bilingual word recognition", in Grainger, J. and Jacobs, A.M. (Eds.), *Scientific Psychology Series. Localist Connectionist Approaches to Human Cognition*, Lawrence Erlbaum Associates, Mahwah, NJ, pp. 189–225.

Dijkstra, T., Grainger, J. and van Heuven, W.J.B. (1999), "Recognition of Cognates and Interlingual Homographs: The Neglected Role of Phonology", *Journal of Memory and Language*, Vol. 41 No. 4, pp. 496–518.

Dijkstra, T. and van Heuven, W.J.B. (2002), "The architecture of the bilingual word recognition system: From identification to decision", *Bilingualism: Language and Cognition*, Vol. 5 No. 3, pp. 175–197.

Dijkstra, T., Van Jaarsveld, H. and Brinke, S.T. (1998), "Interlingual homograph recognition: Effects of task demands and language intermixing", *Bilingualism: Language and Cognition*, Vol. 1 No. 1, pp. 51–66.

Dimitrova, S. (1998), "Bulgarian Speech Rhythm: Stress-Timed or Syllable-Timed?", *Journal of the International Phonetic Association*, Vol. 27 No. 1–2, pp. 27–33.

Dokovova, M. (2015), *Bulgarian Voicing Contrast: At Home and Abroad*, MA, University of Edinburgh, Scotland.

Dokovova, M., Sabev, M., Scobbie, J.M., Lickley, R. and Cowen, S. (2019), "Bulgarian Vowel Reduction in Unstressed Position: An Ultrasound and Acoustic Investigation", *Proceedings of the 19th International Congress of Phonetic Sciences*, presented at the ICPhS, Melbourne, Australia, available at: https://icphs2019.org/icphs2019-fullpapers/pdf/full-paper_818.pdf (accessed 30 June 2019).

Duñabeitia, J.A., Hernandez, J.A., Anton, E., Macizo, P., Estevez, A., Fuentes, L.J. and Carreiras, M. (2014), "The inhibitory advantage in bilingual children revisited", *Experimental Psychology*, Vol. 61 No. 3, pp. 234–251.

Duñabeitia, J.A., Perea, M. and Carreiras, M. (2010), "Masked translation priming effects with highly proficient simultaneous bilinguals", *Experimental Psychology*, Vol. 57 No. 2, pp. 98–107.

Dunn, A.L. and Fox Tree, J.E. (2012), "More on language mode", *International Journal of Bilingualism*, Vol. 18 No. 6, pp. 605–613.

Durlik, J., Szewczyk, J., Muszyński, M. and Wodniecka, Z. (2016), "Interference and Inhibition in Bilingual Language Comprehension: Evidence from Polish-English Interlingual Homographs", *PLoS ONE*, Vol. 11 No. 3, available at: https://doi.org/10.1371/journal.pone.0151430 (accessed 15 May 2019).

Efron, R. (1967), "The Duration of the Present", *Annals of the New York Academy of Sciences*, Vol. 138, pp. 713–729.

Eimas, P.D., Siqueland, E.R., Jusczyk, P. and Vigorito, J. (1971), "Speech perception in infants", *Science*, Vol. 171 No. 3968, pp. 303–306.

Elliot, Z. (2018), *Sociolinguistic Variation among Slovak Immigrants in Edinburgh, Scotland*, PhD, University of Edinburgh, Scotland.

Elman, J.L., Diehl, R. and Buchwald, S. (1977), "Perceptual switching in bilinguals", *The Journal of the Acoustical Society of America*, Vol. 61 No. 4, pp. 971–974.

Ernestus, M., Baayen, H. and Schreuder, R. (2002), "The Recognition of Reduced Word Forms", *Brain and Language*, Vol. 81 No. 1–3, pp. 162–173.

Fassnidge, C. and Freeman, E.D. (2018), "Sounds from seeing silent motion: Who hears them, and what looks loudest?", *Cortex*, Vol. 103, pp. 130-141, available at: https://doi.org/10.1016/j.cortex.2018.02.019 (accessed 17 October 2019).

Flege, J.E. (1995), "Second Language Speech Learning: Theory, Findings, and Problems", in Strange, W. (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, York Press, Timonium, MD, pp. 233–277.

Flege, J.E. (2005), "Origins and development of the Speech Learning Model", presented at the 1st ASA Workshop on L2 Speech Learning, Vancouver, available at: http://jimflege.com/files/Vancouver_April_2005.pdf (accessed 12 May 2019).

Floccia, C., Butler, J., Goslin, J. and Ellis, L. (2009), “Regional and Foreign Accent Processing in English: Can Listeners Adapt?”, *Journal of Psycholinguistic Research*, Vol. 38 No. 4, pp. 379–412.

Floccia, C., Goslin, J., Girard, F. and Konopczynski, G. (2006), “Does a regional accent perturb speech processing? A lexical decision study in French listeners.”, *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 32 No. 5, pp. 1276–1293.

Flores, C. (2010), “The effect of age on language attrition: Evidence from bilingual returnees”, *Bilingualism: Language and Cognition*, Vol. 13, pp. 533–546.

Foucart, A., Santamaría-García, H. and Hartsuiker, R. (2017), “First impression matters. Impact of a foreign accent on cognitive processes.”, presented at the Conference on Multilingualism, Groningen, the Netherlands, p. 33.

Fowler, C.A. (1996), “Listeners do hear sounds, not tongues”, *Journal of the Acoustical Society of America*, Vol. 99 No. 3, pp. 1730–1741.

Friginal, E., Walker, M. and Randall, J.B. (2014), “Exploring mega-corpora: Google Ngram Viewer and the Corpus of Historical American English”, *EuroAmerican Journal of Applied Linguistics and Languages*, Vol. 1 No. 1, available at: https://doi.org/10.21283/2376905X.1.4 (accessed 21 March 2020).

Fromkin, V. (1971), “The Non-Anomalous Nature of Anomalous Utterances”, *Language*, Vol. 47 No. 1, pp. 27-52.

Galabova, N. (2009), “Социалистическият английски език/Sotsialisticheskiyat angliyski ezik”, available at: https://seminar-bg.eu/87-uncategorised/42-e.html?showall=1 (accessed 21 June 2019).

Georgieva, M. (2010), “EFL: From ‘You sound like Dickens’ to International English”, in Saxena, M. and Omoniyi, T. (Eds.), *Contending with Globalization in World Englishes*, Multilingual Matters, Bristol, pp. 113–131.

Gerard, L.D. and Scarborough, D.L. (1989), “Language-specific lexical access of homographs by bilinguals.”, *Journal of Experimental Psychology: Learning, Memory, and Cognition*, Vol. 15 No. 2, pp. 305–315.

Girard, F., Floccia, C. and Goslin, J. (2008), “Perception and awareness of accents in young children”, *British Journal of Developmental Psychology*, Vol. 26 No. 3, pp. 409–433.

Golden, H.L., Nicholas, J.M., Yong, K.X.X., Downey, L.E., Schott, J.M., Mummery, C.J., Crutch, S.J., et al. (2015), “Auditory spatial processing in Alzheimer’s disease”, *Brain*, Vol. 138 No. 1, pp. 189–202.

Goldinger, S.D. (1996), "Words and voices: Episodic traces in spoken word identification and recognition memory", *Journal of Experimental Psychology: Learning, Memory, and Cognition*, Vol. 22 No. 5, pp. 1166–1183.

Goldinger, S.D. (1998), "Echoes of echoes? An episodic theory of lexical access", *Psychological Review*, Vol. 105 No. 2, pp. 251–279.

Goldinger, S.D. (2000), "The role of perceptual episodes in lexical processing", *Proceedings of SWAP (Spoken Word Access Processes)*, pp. 155–159.

Gollan, T.H., Forster, K.I. and Frost, R. (1997), "Translation priming with different scripts: Masked priming with cognates and noncognates in Hebrew-English bilinguals", *Journal of Experimental Psychology: Learning, Memory, and Cognition*, Vol. 23 No. 5, pp. 1122–1139.

Goslin, J., Duffy, H. and Floccia, C. (2012), "An ERP investigation of regional and foreign accent processing", *Brain and Language*, Vol. 122 No. 2, pp. 92–102.

Grainger, J. and Beauvillain, C. (1987), "Language blocking and lexical access in bilinguals.", *The Quarterly Journal of Experimental Psychology*, Vol. 39 No. 2, pp. 295–319.

Green, D.W. (1986), "Control, activation, and resource: A framework and a model for the control of speech in bilinguals", *Brain and Language*, Vol. 27 No. 2, pp. 210–223.

Green, D.W. (1998), "Mental control of the bilingual lexico-semantic system", *Bilingualism: Language and Cognition*, Vol. 1 No. 2, pp. 67–81.

Griffin, J. l. (2001), "Global English infiltrates Bulgaria", *English Today*, Vol. 17 No. 4, pp. 54–60.

de Groot, A.M.B. and Nas, G.L.J. (1991), "Lexical representation of cognates and noncognates in compound bilinguals", *Journal of Memory and Language*, Vol. 30 No. 1, pp. 90–123.

Grosjean, F. (1982), *Life with Two Languages: An Introduction to Bilingualism*, Harvard University Press, Cambridge, MA.

Grosjean, F. (1989), "Neurolinguists, beware! The bilingual is not two monolinguals in one person", *Brain and Language*, Vol. 36 No. 1, pp. 3–15.

Grosjean, F. (2001), "The bilingual's language modes.", in Nicol, J. (Ed.), *One Mind, Two Languages: Bilingual Language Processing*, Blackwell, Oxford, pp. 1–22.

Grosjean, F. (2010), *Bilingual: Life and Reality*, Harvard University Press, Cambridge, MA.

Guion, S.G., Flege, J.E., Akahane-Yamada, R. and Pruitt, J.C. (2000), "An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants", *Journal of the Acoustical Society of America*, Vol. 107 No. 5, pp. 2711–2724.

Haigh, C.A. and Jared, D. (2007), “The activation of phonological representations by bilinguals while reading silently: evidence from interlingual homophones”, *Journal of Experimental Psychology. Learning, Memory, and Cognition*, Vol. 33 No. 4, pp. 623–644.

Halle, M. (1985), “Speculation about the representation of words in memory”, in Fromkin, V. (Ed.), *Phonetic Linguistics*, Vol. 1, Academic Press, New York, pp. 101–114.

Hanulíková, A. and Weber, A. (2012), “Sink positive: Linguistic experience with ‘th’ substitutions influences nonnative word recognition”, *Attention, Perception, & Psychophysics*, Vol. 74 No. 3, pp. 613–629.

Hao, Y.-C. and de Jong, K. (2016), “Imitation of second language sounds in relation to L2 perception and production”, *Journal of Phonetics*, Vol. 54, pp. 151–168.

Harding, L. (2012), “Accent, listening assessment and the potential for a shared-L1 advantage: A DIF perspective”, *Language Testing*, Vol. 29 No. 2, pp. 163–180.

Harnsberger, J.D. (2001), “The perception of Malayalam nasal consonants by Marathi, Punjabi, Tamil, Oriya, Bengali, and American English listeners: A multidimensional scaling analysis”, *Journal of Phonetics*, Vol. 29 No. 3, pp. 303–327.

Haugen, E. (1956), *Bilingualism in the Americas: A Bibliography and Research Guide*, American Dialect Society.

Hawkins, J., Ahmad, S. and Cui, Y. (2017), “A Theory of How Columns in the Neocortex Enable Learning the Structure of the World”, *Frontiers in Neural Circuits*, Vol. 11, available at: https://doi.org/10.3389/fncir.2017.00081 (accessed 5 June 2019).

Hawkins, J., Lewis, M., Klukas, M., Purdy, S. and Ahmad, S. (2019), “A Framework for Intelligence and Cortical Function Based on Grid Cells in the Neocortex”, *Frontiers in Neural Circuits*, Vol. 12, available at: https://doi.org/10.3389/fncir.2018.00121 (accessed 5 June 2019).

Hawkins, S. and Smith, R.H. (2001), “Polysp: a polysystemic, phonetically-rich approach to speech understanding”, *Italian Journal of Linguistics/Rivista Di Linguistica*, Vol. 13, pp. 99–188.

Hay, J. and Drager, K. (2006), “From fush to feesh: Exemplar priming in speech perception.”, *The Linguistic Review*, Vol. 23 No. 3, pp. 351–379.

Hay, J. and Drager, K. (2010), “Stuffed toys and speech perception”, *Linguistics*, Vol. 48 No. 4, pp. 865–892.

Hay, J., Podlubny, R., Drager, K. and McAuliffe, M. (2017), “Car-talk: Location-specific speech production and perception”, *Journal of Phonetics*, Vol. 65, pp. 94–109.

Hayes-Harb, R., Smith, B.L., Bent, T. and 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*, Vol. 36 No. 4, pp. 664–679.

van Hell, J.G. and Dijkstra, T. (2002), "Foreign language knowledge can influence native language performance in exclusively native contexts", *Psychonomic Bulletin and Review*, Vol. 9 No. 4, pp. 780–789.

Hisagi, M., Shafer, V.L., Strange, W. and Sussman, E.S. (2010), "Perception of a Japanese vowel length contrast by Japanese and American English listeners: behavioral and electrophysiological measures", *Brain Research*, Vol. 1360, pp. 89–105.

Hoffman, S., Evert, S., Lehmann, H.-M. and Schneider, P. (2013), "BNCweb", available at: http://bncweb.lancs.ac.uk (accessed 10 April 2017).

Holliday, J.J. (2016), "Second Language Experience Can Hinder the Discrimination of Nonnative Phonological Contrasts", *Phonetica*, Vol. 73 No. 1, pp. 33–51.

Honikman, B. (1964), "Articulatory Settings", in Abercombie, D., Fry, D.B., MacCarthy, P.A.D., Scott, N.C. and Trim, J.L.M. (Eds.), *In Honour of Daniel Jones: Papers Contributed on the Occasion of His Eightieth Birthday 12 September 1961*, Longmans, Green, and Co, London, pp. 73–84.

Howard, M.F. and Poeppel, D. (2010), "Discrimination of Speech Stimuli Based on Neuronal Response Phase Patterns Depends on Acoustics But Not Comprehension", *Journal of Neurophysiology*, Vol. 104 No. 5, pp. 2500–2511.

Hulstijn, J.H. (2011), "Language Proficiency in Native and Nonnative Speakers: An Agenda for Research and Suggestions for Second-Language Assessment", *Language Assessment Quarterly*, Vol. 8 No. 3, pp. 229–249.

Huston, N. (2002), *Losing North: Musings on Land, Tongue and Self*, McArthur, Toronto.

Idemaru, K. and Holt, L. (2011), "Word Recognition Reflects Dimension-based Statistical Learning", *Journal of Experimental Psychology. Human Perception and Performance*, Vol. 37, pp. 1939–56.

Ignatova-Tsoneva, D. and Baeva, D. (2009), "Палатални или смекчени са българските съгласни/Palatalni ili smekcheni sa bylgarskite syglasni." *Nauchni Trudove Na Rusenskiya Universitet*, Vol. 48 No. 6.3, pp. 17–24.

Illes, J., Francis, W.S., Desmond, J.E., Gabrieli, J.D., Glover, G.H., Poldrack, R., Lee, C.J., et al. (1999), "Convergent cortical representation of semantic processing in bilinguals", *Brain and Language*, Vol. 70 No. 3, pp. 347–363.

Instituto Cervantes. (2018), *El Español: Una Lengua Viva*, available at: https://www.cervantes.es/imagenes/File/espanol_lengua_viva_2018.pdf (accessed 6 January 2020).

Jakobson, R., Fant, C.G.M. and Halle, M. (1952), *Preliminaries to Speech Analysis: The Distinctive Features and Their Correlates.*, MIT, Massachusetts: Acoustics Laboratory.

James, E., Gaskell, M.G., Weighall, A. and Henderson, L. (2017), “Consolidation of vocabulary during sleep: The rich get richer?”, *Neuroscience & Biobehavioral Reviews*, Vol. 77, pp. 1–13.

Jensen, C. and Thøgersen, J. (2017), “Foreign accent, cognitive load and intelligibility of EMI lectures”, *Nordic Journal of English Studies*, Vol. 16 No. 3, pp. 107–137.

Jepsen, M.L. (2010), *Modeling Auditory Processing and Speech Perception in Hearing-Impaired Listeners*, PhD, Technical University of Denmark, Kgs. Lyngby, Denmark.

Jiang, N. (1999), “Testing processing explanations for the asymmetry in masked cross-language priming”, *Bilingualism: Language and Cognition*, Vol. 2 No. 1, pp. 59–75.

Jiang, N. and Forster, K.I. (2001), “Cross-Language Priming Asymmetries in Lexical Decision and Episodic Recognition”, *Journal of Memory and Language*, Vol. 44 No. 1, pp. 32–51.

Johnson, K. (1997), “Speech perception without speaker normalization: an exemplar model”, in Johnson, K. and Mullenix, J. (Eds.), *Talker Variability in Speech Processing*, Academic Press, pp. 145–166.

Ju, M. and Luce, P.A. (2004), “Falling on sensitive ears: constraints on bilingual lexical activation”, *Psychological Science*, Vol. 15 No. 5, pp. 314–318.

Jusczyk, P.W. (1993), “From general to language-specific capacities: The WRAPSA model of how speech perception develops.”, *Journal of Phonetics*, Vol. 21 No. 1–2, pp. 3–28.

Kang, O. and Rubin, D., L. (2009), “Reverse Linguistic Stereotyping: Measuring the Effect of Listener Expectations on Speech Evaluation”, *Journal of Language and Social Psychology*, Vol. 28 No. 4, pp. 441–456.

Kerkhofs, R., Dijkstra, T., Chwilla, D.J. and De Bruijn, E.R. (2006), “Testing a model for bilingual semantic priming with interlingual homographs: RT and N400 effects.”, *Brain Research*, Vol. 1068 No. 1, pp. 170–183.

Kim, J. (2016), “Perceptual Associations Between Words and Speaker Age”, *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, Vol. 7 No. 1, p. 18.

Klatt Dennis H. (1975), “Voice Onset Time, Frication, and Aspiration in Word-Initial Consonant Clusters”, *Journal of Speech and Hearing Research*, Vol. 18 No. 4, pp. 686–706.

Kleinschmidt, D.F., Fine, A.B. and Jaeger, T.F. (2012), “A belief-updating model of adaptation and cue combination in syntactic comprehension”, in Miyake, N., Peebles, D. and Cooper, R.P. (Eds.), *Proceedings of the 34th Annual Conference of the Cognitive Science Society*, presented at the 34th Annual Conference of the Cognitive Science Society, Sapporo, Japan, pp. 599–604.

Kleinschmidt, D.F. and Jaeger, T.F. (2015), “Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel”, *Psychological Review*, Vol. 122 No. 2, pp. 148–203.

Koeva, S. and Kolkovska, S. (2011), “Frequency dictionaries from the Bulgarian National Corpus”, available at: http://dcl.bas.bg/bulnc/en/dostap/retchnitsi/ (accessed 21 April 2017).

Kong, Y.-Y., Somarowthu, A. and Ding, N. (2015), “Effects of Spectral Degradation on Attentional Modulation of Cortical Auditory Responses to Continuous Speech”, *Journal of the Association for Research in Otolaryngology*, Vol. 16 No. 6, pp. 783–796.

Kraljic, T., Samuel, A.G. and Brennan, S.E. (2008), “First impressions and last resorts: How listeners adjust to speaker variability”, *Psychological Science*, Vol. 19 No. 4, pp. 332–338.

Kriengwatana, B., Terry, J., Chládková, K. and Escudero, P. (2016), “Speaker and Accent Variation Are Handled Differently: Evidence in Native and Non-Native Listeners”, *PLoS ONE*, Vol. 11 No. 6, available at: https://doi.org/10.1371/journal.pone.0156870 (accessed 14 March 2018).

ktz_33. (2007), *The Italian Tourist*, available at: https://www.youtube.com/watch?v=5cPocuQUJ7I (accessed 8 April 2019).

Kuhl, P.K. (1991), “Human adults and human infants show a ‘perceptual magnet effect’ for the prototypes of speech categories, monkeys do not”, *Perception & Psychophysics*, Vol. 50 No. 2, pp. 93–107.

Kuhl, P.K. (1992), “Infants’ perception and representation of speech: development of a new theory”, in Ohala, J.J., Neary, T.M., Derwing, B.L., Hodge, M.M. and Wiebe, G.E. (Eds.), *Proceedings of the International Conference on Spoken Language Processing,* presented at the International Conference on Spoken Language Processing, Banff, Alberta, Canada, pp. 3-10.

Kuhl, P.K. and Iverson, P. (1995), “Linguistic experience and the ‘perceptual magnet effect’”, in Strange, W. (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, York Press, Timonium, Maryland, pp. 121–154.

Kuppens, A.H. (2010), “Incidental foreign language acquisition from media exposure”, *Learning, Media and Technology*, Vol. 35 No. 1, pp. 65–85.

Ladd, D.R. (2006), “‘Distinctive phones’ in surface representation”, in Goldstein, L., Whalen, D.H. and Best, C.T. (Eds.), *Laboratory Phonology 8*, De Gruyter Mouton, Berlin, Boston, available at: https://doi.org/10.1515/9783110197211.1.3 (accessed 14 June 2018).

Lado, R. (1961), *Language Testing*, Longmans, Green, and Co, London.

Lagrou, E., Hartsuiker, R.J. and Duyck, W. (2011), “Knowledge of a second language influences auditory word recognition in the native language”, *Journal of Experimental Psychology. Learning, Memory, and Cognition*, Vol. 37 No. 4, pp. 952–965.

Lagrou, E., Hartsuiker, R.J. and Duyck, W. (2015), “Do semantic sentence constraint and L2 proficiency influence language selectivity of lexical access in native language listening?”, *Journal of Experimental Psychology. Human Perception and Performance*, Vol. 41 No. 6, pp. 1524–1538.

Lane, H. (1963), “Foreign accent and speech distortion”, *Journal of the Acoustical Society of America*, Vol. 35 No. 4, pp. 451–453.

Lee, J.-K., Han, J.-I., Choi, T.-H. and Lim, I. (2012), “The Interlanguage Speech Intelligibility Benefit (ISIB) of English Prosody: The Case of Focal Prominence for Korean Learners of English and Natives”, *Phonetics and Speech Science*, Vol. 4 No. 4, pp. 53–68.

Lemhöfer, K. and Broersma, M. (2012a), “Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English”, *Behavior Research Methods*, Vol. 44 No. 2, pp. 325–343.

Lemhöfer, K. and Broersma, M. (2012b), “LexTALE”, available at: http://www.lextale.com/ (accessed 27 February 2018).

Lennon, R., Smith, R. and Stuart-Smith, J. (2016), “Ambiguous rhoticity in Glasgow: Short term exposure promotes perceptual adaptation for experienced and inexperienced listeners”, *Proceedings of the 15th Biennial Conference on Laboratory Phonology,* presented at the LabPhon, Cornell University.

Lev-Ari, S. (2017), “Talking to fewer people leads to having more malleable linguistic representations”, *PLoS ONE*, Vol. 12 No. 8, available at: https://doi.org/10.1371/journal.pone.0183593 (accessed 27 February 2018).

Léwy, N. and Grosjean, F. (2008), “The Léwy and Grosjean BIMOLA model”, in Grosjean, F. (Ed.), *Studying Bilinguals*, Oxford University Press, pp. 201–210.

Lisker, L. and Abramson, A.S. (1967), “Some Effects of Context On Voice Onset Time in English Stops”, *Language and Speech*, Vol. 10 No. 1, pp. 1–28.

Llompart, M. and Reinisch, E. (2018), “Imitation in a Second Language Relies on Phonological Categories but Does Not Reflect the Productive Usage of Difficult Sound Contrasts”, *Language and Speech*, available at: https://doi.org/10.1177/0023830918803978 (accessed 8 January 2018).

Loureiro, C. de S., Willadino Braga, L., Souza, L. do N., Filho, G.N., Queiroz, E. and Dellatolas, G. (2004), “Degree of illiteracy and phonological and metaphonological skills in unschooled adults”, *Brain and Language*, Vol. 89 No. 3, pp. 499–502.

Luce, P.A. and Pisoni, D.B. (1998), “Recognizing Spoken Words: The Neighborhood Activation Model”, *Ear & Hearing*, Vol. 19 No. 1, pp. 1–36.

Ludwig, A. and Mora, J.C. (2017), “Processing time and comprehensibility judgments in non-native listeners’ perception of L2 speech”, *Journal of Second Language Pronunciation*, Vol. 3 No. 2, pp. 167–198.

Macnamara, J. and Kushnir, S.L. (1971), “Linguistic independence of bilinguals: The input switch.”, *Journal of Verbal Learning and Verbal Behavior*, Vol. 10 No. 5, pp. 480–487.

Major, R.C., Fitzmaurice, S.F., Bunta, F. and Balasubramanian, C. (2002), “The Effects of Nonnative Accents on Listening Comprehension: Implications for ESL Assessment”, *TESOL Quarterly*, Vol. 36 No. 2, p. 173.

Marian, V., Bartolotti, J., Chabal, S. and Shook, A. (2012), “CLEARPOND: Cross-Linguistic Easy-Access Resource for Phonological and Orthographic Neighborhood Densities”, *PLOS ONE*, Vol. 7 No. 8, p. e43230, available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3423352/ (accessed 26 February 2018).

Marian, V., Blumenfeld, H.K. and Boukrina, O.V. (2008), “Sensitivity to phonological similarity within and across languages”, *Journal of Psycholinguistic Research*, Vol. 37 No. 3, pp. 141–170.

Marian, V., Spivey, M. and Hirsch, J. (2003), “Shared and separate systems in bilingual language processing: Converging evidence from eyetracking and brain imaging”, *Brain and Language*, Vol. 86 No. 1, p. 70 82.

Marslen-Wilson, W. and Welsh, A. (1978), “Processing interactions and lexical access during word recognition in continuous speech”, *Cognitive Psychology*, Vol. 10 No. 1, pp. 29–63.

Marslen-Wilson, W. and Zwitserlood, P. (1989), “Accessing spoken words: The importance of word onsets”, *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 15 No. 3, pp. 576–585.

Masapollo, M., Polka, L. and Ménard, L. (2017), “A universal bias in adult vowel perception – By ear or by eye”, *Cognition*, Vol. 166, pp. 358–370.

Mather, E. (2013), “Novelty, attention, and challenges for developmental psychology”, *Frontiers in Psychology*, Vol. 4, available at: https://doi.org/10.3389/fpsyg.2013.00491 (accessed 20 January 2020).

Mathot, S., Schreij, D. and Theeuwes, J. (2012), “OpenSesame: an open-source, graphical experiment builder for the social sciences”, *Behavior Research Methods*, Vol. 44 No. 2, pp. 314–324.

Max Planck Institute for Psycholinguistics. (2001), “WebCelex”, available at: http://celex.mpi.nl/ (accessed 26 June 2018).

Maye, J., Aslin, R.N. and Tanenhaus, M.K. (2008), “The Weckud Wetch of the Wast: Lexical Adaptation to a Novel Accent”, *Cognitive Science*, Vol. 32 No. 3, pp. 543–562.

McClelland, J.L. and Elman, J.L. (1986), “The TRACE model of speech perception”, *Cognitive Psychology*, Vol. 18 No. 1, pp. 1–86.

McGurk, H. and MacDonald, J. (1976), “Hearing lips and seeing voices”, *Nature*, Vol. 264 No. 5588, pp. 746–748.

Medin, D.L. and Schaffer, M.M. (1978), “Context theory of classification learning”, *Psychological Review*, Vol. 85 No. 3, pp. 207–238.

Mennen, I., Scobbie, J.M., de Leeuw, E., Schaeffler, S. and Schaeffler, F. (2010), "Measuring language-specific phonetic settings", *Second Language Research*, Vol. 26 No. 1, pp. 13–41.

Metsala, J.L. (1999), "Young children's phonological awareness and nonword repetition as a function of vocabulary development", *Journal of Educational Psychology*, Vol. 91 No. 1, pp. 3–19.

Mitterer, H. and McQueen, J.M. (2009), "Processing reduced word-forms in speech perception using probabilistic knowledge about speech production.", *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 35 No. 1, pp. 244–263.

Moos, A., Smith, R., Miller, S.R. and Simmons, D.R. (2014), "Cross-modal associations in synaesthesia: vowel colours in the ear of the beholder", *I-Perception*, Vol. 5, pp. 132–142.

Morais, J., Cary, L., Alegria, J. and Bertelson, P. (1979), "Does awareness of speech as a sequence of phones arise spontaneously?", *Cognition*, Vol. 7 No. 4, pp. 323–331.

Mosca, M. and de Bot, K. (2017), "Bilingual Language Switching: Production vs. Recognition", *Frontiers in Psychology*, Vol. 8, p. 934.

Mullennix, J.W., Pisoni, D. and Martin, C. (1989), "Some effects of talker variability on spoken word recognition", *The Journal of the Acoustical Society of America*, Vol. 85 No. 1, p. 365.

Munro, M. (1998), "The effects of noise on the intelligibility of foreign-accented speech", *Studies in Second Language Acquisition*, Vol. 20, pp. 139–154.

Munro, M.J. and Derwing, T.M. (1995), "Processing Time, Accent, and Comprehensibility in the Perception of Native and Foreign-Accented Speech", *Language and Speech*, Vol. 38 No. 3, pp. 289–306.

Munro, M.J., Derwing, T.M. and Morton, S.L. (2006), "The mutual intelligibility of L2 speech", *Studies in Second Language Acquisition*, Vol. 28 No. 1, pp. 111–131.

National Statistical Institute. (2010), *Revision of the Population*, available at: https://www.nsi.bg/en/content/6731/revision-population (accessed 9 March 2020).

Nations Online. (2020), "Countries by languages", available at: https://www.nationsonline.org/oneworld/countries_by_languages.htm (accessed 9 March 2020).

Niedzielski, N. (1999), "The effect of social information on the perception of sociolinguistic variables", *Journal of Language and Social Psychology*, Vol. 18 No. 1, pp. 62–85.

Nielsen, K. (2011), "Specificity and abstractness of VOT imitation", *Journal of Phonetics*, Vol. 39 No. 2, pp. 132–142.

Norris, D. (1999), "The merge model: Speech perception is bottom-up", *Journal of The Acoustical Society of America*, Vol. 106 No. 4, p. 2295.

Norris, D. (2003), “Perceptual learning in speech”, *Cognitive Psychology*, Vol. 47 No. 2, pp. 204–238.

Norris, D. and McQueen, J.M. (2008), “Shortlist B: A Bayesian model of continuous speech recognition”, *Psychological Review*, Vol. 115 No. 2, pp. 357–395.

Ockey, G.J. and French, R. (2016), “From One to Multiple Accents on a Test of L2 Listening Comprehension”, *Applied Linguistics*, Vol. 37 No. 5, pp. 693–715.

O’Reilly, L.M. (1998), “English language cultures in Bulgaria: a linguistic sibling rivalry?”, *World Englishes*, Vol. 17 No. 1, pp. 71–84.

Pallier, C., Bosch, L. and Sebastian-Galles, N. (1997), “A limit on behavioral plasticity in speech perception”, *Cognition*, Vol. 64 No. 3, pp. B9–B17.

Pavlenko, A. (2014), *The Bilingual Mind: And What It Tells Us About Language And Thought*, Cambridge University Press, Cambridge.

Pedersen, E.J., Miller, D.L., Simpson, G.L. and Ross, N. (2019), “Hierarchical generalized additive models in ecology: an introduction with mgcv”, *PeerJ*, Vol. 7, p. e6876, available at: https://peerj.com/articles/6876/ (accessed 10 June 2019).

Peperkamp, S. and Dupoux, E. (2003), “Reinterpreting loanword adaptations: the role of perception”, *Proceedings of the 15th International Congress of Phonetic Sciences*, presented at the ICPhS, Barcelona, Spain, pp. 367–370.

Perani, D., Dehaene, S., Grassi, F., Cohen, L., Cappa, S.F., Dupoux, E., Fazio, F., et al. (1996), “Brain processing of native and foreign languages”, *Neuroreport*, Vol. 7 No. 15–17, pp. 2439–2444.

Perani, D., Paulesu, E., Galles, N.S., Dupoux, E., Dehaene, S., Bettinardi, V., Cappa, S.F., et al. (1998), “The bilingual brain. Proficiency and age of acquisition of the second language”, *Brain: A Journal of Neurology*, Vol. 121 (Pt. 10), pp. 1841–1852.

Phillpott, M. (2019), “Course: Text Mining for Historians, Topic: An introduction to text mining”, *School of Advanced Study University of London*, available at: https://port.sas.ac.uk/course/view.php?id=53&section=1 (accessed 28 March 2020).

Pierrehumbert, J. (2002), “Word-specific phonetics”, in Gussenhoven, C. and Warner, N. (Eds.), *Laboratory Phonology VII*, Mouton de Gruyter, Berlin, pp. 101–140.

Pierrehumbert, J. (2016), “Phonological representation: Beyond abstract versus episodic”, *Annual Review of Linguistics*, Vol. 2, pp. 33–52.

Pike, K.L. (1945), *The Intonation of American English*, University of Michigan Press, Ann Arbor.

Pillai, J.J., Araque, J.M., Allison, J.D., Sethuraman, S., Loring, D.W., Thiruvayaru, D., Ison, C.B., et al. (2003), “Functional MRI study of semantic and phonological language processing in bilingual subjects: preliminary findings”, *NeuroImage*, Vol. 19 No. 3, pp. 565–576.

Pinet, M., Iverson, P. and Huckvale, M. (2011), "Second-language experience and speech-in-noise recognition: Effects of talker–listener accent similarity", *The Journal of the Acoustical Society of America*, Vol. 130 No. 3, pp. 1653–1662.

Polka, L. (1991), "Cross-language speech perception in adults: Phonemic, phonetic, and acoustic contributions", *Journal of the Acoustical Society of America*, Vol. 89 No. 6, pp. 2961–2977.

Polka, L. (1992), "Characterizing the influence of native language experience on adult speech perception", *Perception and Psychophysics*, Vol. 52 No. 1, pp. 37–52.

Porretta, V. and Tucker, B.V. (2019), "Eyes Wide Open: Pupillary Response to a Foreign Accent Varying in Intelligibility", *Frontiers in Communication*, Vol. 4, available at: https://doi.org/10.3389/fcomm.2019.00008 (accessed 10 January 2020).

Porretta, V., Tucker, B.V. and Järvikivi, J. (2016), "The influence of gradient foreign accentedness and listener experience on word recognition", *Journal of Phonetics*, Vol. 58, pp. 1–21.

Potter, M.C., So, K.F., Eckardt, B.V. and Feldman, L.B. (1984), "Lexical and conceptual representation in beginning and proficient bilinguals.", *Journal of Verbal Learning and Learning Behaviour*, Vol. 23 No. 1, pp. 23–38.

R Core Team. (2018), "R: A language environment for statistical computing" (3.5.2), [software], available at: https://www.r-project.org/ (accessed 9 December 2018).

Radkova, Z. (2009), *Phonological Processes in Contemporary Spoken Bulgarian*, PhD, University of Texas at Arlington, Arlington, TX, available at: https://rc.library.uta.edu/uta-ir/handle/10106/1645 (accessed 1 January 2018)

Rangelov, T. (2008), Fortis and Lenis Stops: On Voice Onset Time in Icelandic, Swedish and Bulgarian, MA, University of Iceland, Reykjavík, available at: http://skemman.is/handle/1946/16543 (accessed 27 February 2018).

Rastle, K., Harrington, J. and Coltheart, M. (2002), "358,534 nonwords: The ARC Nonword Database", *Quarterly Journal of Experimental Psychology*, Vol. 55 No. 4, pp. 1339–1362.

Reed, S.K. (1972), "Pattern recognition and categorization", *Cognitive Psychology*, Vol. 3 No. 3, pp. 382–407.

Reinisch, E. and Holt, L.L. (2014), "Lexically Guided Phonetic Retuning of Foreign-Accented Speech and Its Generalization", *Journal of Experimental Psychology. Human Perception and Performance*, Vol. 40 No. 2, pp. 539–555.

Reinisch, E. and Weber, A. (2012), "Adapting to suprasegmental lexical stress errors in foreign-accented speech", *The Journal of the Acoustical Society of America*, Vol. 132 No. 2, pp. 1165–1176.

Reinisch, E., Weber, A. and Mitterer, H. (2013), “Listeners retune phoneme categories across languages”, *Journal of Experimental Psychology. Human Perception and Performance*, Vol. 39 No. 1, pp. 75–86.

Remijsen, B. (2011), *msr&check_f1f2_batch_interv.psc. A script to batch process the measurements and visual checking of formant values in Praat,* available at: http://www.lel.ed.ac.uk/~bert/praatscripts.html (accessed 27 February 2018).

Rimmele, J.M., Golumbic, E.Z., Schröger, E. and Poeppel, D. (2015), “The effects of selective attention and speech acoustics on neural speech-tracking in a multi-talker scene”, *Cortex*, Vol. 68, pp. 144–154.

Rohena-Madrazo, M. (2013), “Perceptual Assimilation of Occluded Voiced Stops by Spanish Listeners”, *Selected Proceedings of the 15th Hispanic Linguistics Symposium*, presented at the Hispanic Linguistics Symposium, Somerville, MA, pp. 140–156.

Rubin, D., L. (1992), “Nonlanguage factors affecting undergraduates' judgments of nonnative English-speaking teaching assistants”, *Research in Higher Education*, Vol. 33 No. 4, pp. 511–531.

Rubin, D.L. and Lannutti, P. (2001), “Frameworks for assessing contact as a tool for reducing prejudice.”, in Milhouse, V.H., Asante, M.K. and Nwosu, P.O. (Eds.), *Transcultural Realities: Interdisciplinary Perspectives on Cross-Cultural Relations*, Sage, Thousand Oaks, CA, pp. 313–326.

Rüschemeyer, S.-A., Fiebach, C.J., Kempe, V. and Friederici, A.D. (2005), “Processing lexical semantic and syntactic information in first and second language: fMRI evidence from German and Russian”, *Human Brain Mapping*, Vol. 25 No. 2, pp. 266–286.

Rüschemeyer, S.-A., Zysset, S. and Friederici, A.D. (2006), “Native and non-native reading of sentences: an fMRI experiment”, *NeuroImage*, Vol. 31 No. 1, pp. 354–365.

Sabev, M. (2015), “Reduction of unstressed central and back vowels in Contemporary Standard Bulgarian”, *Proceedings of the 18th International Congress of Phonetic Sciences*, presented at the ICPhS, Glasgow, Scotland, available at: https://ora.ox.ac.uk/objects/uuid:cc997e7b-12b7-4ed9-8249-43f916a72c96/download_file?file_format=pdf&safe_filename=Reduction_Sabev.pdf&type_of_work=Conference+item (accessed 16 July 2017).

Said, S.B. (2006), *Attitudes towards Accented Speech: A Comparative Study of Native and Non-Native Speakers of American English*, PhD, Southern Illinois University, Carbondale, available at: https://www.researchgate.net/publication/301814340_Attitudes_towards_Accented_Speech_A_Comparative_Study_of_Native_and_Non-Native_Speakers_of_American_English (accessed 12 June 2017).

Schaeffler, S., Scobbie, J. & Mennen, I. (2008) "An Evaluation of Inter-Speech Postures for the Study of Language-Specific Articulatory Settings.", *Proceedings of the 8th International Seminar on*

Speech Production, presented at the ISSIP, Strasbourg, France, pp. 121-124, available at: https://eresearch.qmu.ac.uk/handle/20.500.12289/2010 (accessed 10 July 2019).

Schwartz, A.I. and Kroll, J.F. (2006), "Bilingual lexical activation in sentence context.", *Journal of Memory and Language*, Vol. 55 No. 2, pp. 197–212.

Seligman, C.R., Tucker, G.R. and Lambert, W.E. (1972), "The effects of speech style and other attributes on teachers' attitudes toward pupils.", *Language in Society*, Vol. 1 No. 1, pp. 131–142.

Semon, R. (1923), *Mnemic Psychology*, George Allen And Unwin Limited, London.

Shinya, M., Komuro, L., Kobayashi, H., Obata, H. and Nakazawa, K. (2015), "Batting and button-press reaction time in primary, junior high and high school baseball players", in Colloud, F., Domalain, M. and Monnet, T. (Eds.), *Proceedings of the 33rd International Conference of Biomechanics in Sports,* presented at the International Conference of Biomechanics in Sports, Poitiers, France, pp. 1193–1194

Shipkovenski, M. (2011), "On the inappropriate use of foreign words in Bulgarian language", *Bulgarian Speech. Journal of Linguistics and Language Culture*, No. 2, pp. 99–102.

Shockley, K., Sabadini, L. and Fowler, C.A. (2004), "Imitation in shadowing words", *Perception & Psychophysics*, Vol. 66 No. 3, pp. 422–429.

Sidaras, S.K., Alexander, J.E.D. and Nygaard, L.C. (2009), "Perceptual learning of systematic variation in Spanish-accented speech", *The Journal of the Acoustical Society of America*, Vol. 125 No. 5, pp. 3306–3316.

Simonet, M. (2014), "Phonetic consequences of dynamic cross-linguistic interference in proficient bilinguals", *Journal of Phonetics*, Vol. 43, p. 26 37.

Smith, B.L., Bent, T. and Bradlow, A.R. (2003), "Production and perception of temporal contrasts in foreign-accented English", *Proceedings of the 15th International Congress of Phonetic Sciences*, presented at the ICPhS, Barcelona, Spain, pp. 519–522, available at: http://faculty.wcas.northwestern.edu/ann-bradlow/publications/2003/ICPhS_0401-temp.pdf (accessed 27 February 2018).

Smith, R., Holmes-Elliott, S., Pettinato, M. and Knight, R.-A. (2014), "Cross-accent intelligibility of speech in noise: Long-term familiarity and short-term familiarization", *Quarterly Journal of Experimental Psychology*, Vol. 67, pp. 590–608.

Soares, C. and Grosjean, F. (1984), "Bilinguals in a monolingual and a bilingual speech mode: the effect on lexical access", *Memory & Cognition*, Vol. 12 No. 4, pp. 380–386.

Song, J. and Iverson, P. (2018), "Listening Effort During Speech Perception Enhances Auditory and Lexical Processing for Non-Native Listeners and Accents", *Cognition*, Vol. 179, pp. 163–170.

Sóskuthy, M. (2017), “Generalised additive mixed models for dynamic analysis in linguistics: a practical introduction”, available at: https://arxiv.org/abs/1703.05339v1 (accessed 19 April 2019).

Spivey, M.J. and Marian, V. (1999), “Cross Talk between Native and Second Languages: Partial Activation of an Irrelevant Lexicon”, *Psychological Science*, Vol. 10 No. 3, pp. 281–284.

Stæhr, L.S. (2009), “Vocabulary Knowledge and Advanced Listening Comprehension in English as a Foreign Language”, *Studies in Second Language Acquisition; New York*, Vol. 31 No. 4, pp. 577–607.

Stevens, K.N. (2002), “Toward a model for lexical access based on acoustic landmarks and distinctive features”, *The Journal of the Acoustical Society of America*, Vol. 111 No. 4, pp. 1872–1891.

Stibbard, R.M. and Lee, J.-I. (2006), “Evidence against the mismatched interlanguage speech intelligibility benefit hypothesis”, *The Journal of the Acoustical Society of America*, Vol. 120 No. 1, p. 433 442.

Stoet, G. (2010), “PsyToolkit - A software package for programming psychological experiments using Linux”, *Behavior Research Methods*, Vol. 42 No. 4, pp. 1096–1104.

Stoet, G. (2017), “PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments”, *Teaching of Psychology*, Vol. 44 No. 1, pp. 24–31.

Strange, W. (2011), “Automatic selective perception (ASP) of first and second language speech: A working model”, *Journal of Phonetics*, Vol. 39 No. 4, pp. 456–466.

Sumner, M. (2011), “The role of variation in the perception of accented speech”, *Cognition*, Vol. 119 No. 1, pp. 131–136.

Sumner, M. and Samuel, A.G. (2009), “The effect of experience on the perception and representation of dialect variants”, *Journal of Memory and Language*, Vol. 60, pp. 487–501.

Swingley, D. (2003), “Phonetic detail in the developing lexicon”, *Language and Speech*, Vol. 46 No. 2-3, pp. 265–294.

Szakay, A., Babel, M. and King, J. (2016), “Social categories are shared across bilinguals′ lexicons”, *Journal of Phonetics*, Vol. 59, pp. 92–109.

Tamminen, H., Peltola, M.S., Kujala, T. and Näätänen, R. (2015), “Phonetic training and non-native speech perception--New memory traces evolve in just three days as indexed by the mismatch negativity (MMN) and behavioural measures”, *International Journal of Psychophysiology*, Vol. 97 No. 1, pp. 23–29.

Teng, F. (2016), “An In-Depth Investigation into the Relationship between Vocabulary Knowledge and Academic Listening Comprehension”, *TESL-EJ*, Vol. 20 No. 2, available at: https://eric.ed.gov/?id=EJ1113907 (accessed 5 February 2019).

Ternes, E. and Vladimirova-Buhtz, T. (1990), “Bulgarian”, *Journal of the International Phonetic Association*, Vol. 20 No. 1, pp. 45–47.

Thiery, C. (1978), “True Bilingualism and Second-language Learning”, in Gerver, D. and Wallace Sinaiko, H. (Eds.), *Language Interpretation and Communication*, Springer US, Boston, MA, pp. 145–153.

Thurstone, L.L. (1938), *Primary Mental Abilities*, University of Chicago Press, Chicago.

Tilkov, D. and Boyadziev, T. (1981), *Българска фонетика/Balgarska fonetika*, Nauka i izkustvo, Sofia.

Titone, D., Libben, M., Mercier, J., Whitford, V. and Pivneva, I. (2011), “Bilingual lexical access during L1 sentence reading: The effects of L2 knowledge, semantic constraint, and L1–L2 intermixing.”, *Journal of Experimental Psychology: Learning, Memory, and Cognition*, Vol. 37 No. 6, pp. 1412–1431.

Traunmüller, H. (1990), “Analytical expressions for the tonotopic sensory scale”, *The Journal of the Acoustical Society of America*, Vol. 88 No. 1, pp. 97–100.

Tyler, L.K. and Wessels, J. (1983), “Quantifying contextual contributions to word-recognition processes”, *Perception and Psychophysics*, Vol. 34 No. 5, pp. 409–420.

Verhoeven, L. and de Jong, J.H.A.L. (1992), *The Construct of Language Proficiency*, John Benjamins, available at: (accessed 7 June 2019).

Vitevitch, M.S. (2007), “The spread of the phonological neighborhood influences spoken word recognition”, *Memory & Cognition*, Vol. 35 No. 1, pp. 166–175.

Vroomen, J., van Linden, S., de Gelder, B. and Bertelson, P. (2007), “Visual recalibration and selective adaptation in auditory and visual speech perception: Contrasting build-up courses”, *Neuropsychologia*, Vol. 45 No. 3, pp. 572–577.

Wade, T. and Möbius, B. (2010), “Detailed Phonetic Memory for Multi-Word and Part-Word Sequences”, *Journal of the Association for Laboratory Phonology*, Vol. 1 No. 2, pp. 283–294.

Wagner, M. (2016), *Listening in the Wrong Language: The Role of Language Dominance and Accent in Cross-Language Speech Misperceptions*, MA, Radboud University, Nijmegen.

Wainstein, G., Rojas-Líbano, D., Crossley, N.A., Carrasco, X., Aboitiz, F. and Ossandón, T. (2017), “Pupil Size Tracks Attentional Performance In Attention-Deficit/Hyperactivity Disorder”, *Scientific Reports*, Vol. 7 No. 1, p. 8228.

Walker, M., Szakay, A. and Cox, F. (2019), “Can kiwis and koalas as cultural primes induce perceptual bias in Australian English speaking listeners?”, *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, Vol. 10 No. 1, p. 7.

Walters, J. (2004), *Bilingualism: The Sociopragmatic-Psycholinguistic Interface*, Routledge, Mahwah, USA.

Wartenburger, I., Heekeren, H.R., Abutalebi, J., Cappa, S.F., Villringer, A. and Perani, D. (2003), "Early Setting of Grammatical Processing in the Bilingual Brain", *Neuron*, Vol. 37 No. 1, pp. 159–170.

Weber, A., Betta, A.M.D. and McQueen, J.M. (2014), "Treack or trit: Adaptation to genuine and arbitrary foreign accents by monolingual and bilingual listeners", *Journal of Phonetics*, Vol. 46, pp. 34–51.

Weber, A. and Cutler, A. (2004), "Lexical competition in non-native spoken-word recognition", *Journal of Memory and Language*, Vol. 50 No. 1, pp. 1–25.

Weinberger, S. (2015), "Speech Accent Archive", George Mason University, available at: http://accent.gmu.edu (accessed 27 February 2018).

Wells, J. (1997), "Whatever happened to Received Pronunciation?", in Medina Casado, C. and Soto Palomo, C. (Eds.), *II Jornadas de Estudios Ingleses*, Universidad de Jaén, Spain, pp. 19–28.

Wells, J.C. (1982), *Accents of English*, Cambridge University Press, Cambridge.

Werker, J.F. and Logan, J.S. (1985), "Cross-language evidence for three factors in speech perception", *Perception and Psychophysics*, Vol. 37 No. 1, pp. 35–44.

van Wijngaarden, S. (2001), "Intelligibility of native and non-native Dutch speech", *Speech Communication*, Vol. 35, pp. 103–113.

van Wijngaarden, S.J., Steeneken, H.J.M. and Houtgast, T. (2002), "Quantifying the intelligibility of speech in noise for non-native talkers", *The Journal of the Acoustical Society of America*, Vol. 112 No. 6, pp. 3004–3013.

Wingstedt, M. and Schulman, R. (1987), "Comprehension of foreign accents", in Dressler, W. (Ed.), *Proceedings of the 5th International Phonology Meeting*, Cambridge University Press, pp. 339–345.

Winn, M. (2014), *Make formant continuum. GUI-based wizard for creating realistic vowel formant continua from modified natural speech (version 30),* available at: http://www.mattwinn.com/praat.html (accessed 17 October 2019)

Witteman, M.J., Weber, A. and McQueen, J.M. (2013), "Foreign accent strength and listener familiarity with an accent codetermine speed of perceptual adaptation", *Attention, Perception, and Psychophysics*, Vol. 75 No. 3, pp. 537–556.

Xie, X., Earle, F.S. and Myers, E.B. (2018), "Sleep facilitates generalisation of accent adaptation to a new talker", *Language, Cognition and Neuroscience*, Vol. 33 No. 2, pp. 196–210.

Xie, X. and Myers, E.B. (2017), “Learning a talker or learning an accent: Acoustic similarity constrains generalization of foreign accent adaptation to new talkers”, *Journal of Memory and Language*, Vol. 97, pp. 30–46.

Xu, M., Baldauf, D., Chang, C.Q., Desimone, R. and Tan, L.H. (2017), “Distinct distributed patterns of neural activity are associated with two languages in the bilingual brain”, *Science Advances*, Vol. 3 No. 7, p. e1603309, available at: https://advances.sciencemag.org/content/3/7/e1603309 (accessed 27 February 2018).

Zhobov, V. (2004), *Звуковете в българския език/Zvukovete v balgarskija ezik*, Sema RS, Sofia.

Appendix 1 Stimuli

A. Experiment 1 Foreign accentedness ratings

Word	Accent	Status	Mean Rating
brick	BA	manipulated	8.4
delicious	BA	manipulated	6.9
familiar	BA	manipulated	6.9
kiss	BA	manipulated	8.3
pink	BA	manipulated	6.9
wish	BA	manipulated	4
admission	BA	unmanipulated	6.3
cheers	BA	unmanipulated	8.4
consider	BA	unmanipulated	7.1
district	BA	unmanipulated	8
give	BA	unmanipulated	4.1
liver	BA	unmanipulated	7.2
mince	BA	unmanipulated	5.6
sticky	BA	unmanipulated	6.7
switch	BA	unmanipulated	4.2
think	BA	unmanipulated	6.1
wing	BA	unmanipulated	6.2
vinegar	BA	unmanipulated	6.7
gypsies	BA	unmanipulated	7.9
big	BA	manipulated	5.8

kitchen	SBE	manipulated	6.4
lip	SBE	manipulated	4.6
rich	SBE	manipulated	3.7
shift	SBE	manipulated	4.6
simple	SBE	manipulated	2.7
difficult	SBE	unmanipulated	4
finish	SBE	unmanipulated	2.1
initial	SBE	unmanipulated	2.1
issue	SBE	unmanipulated	2.2
live	SBE	unmanipulated	2.2
miracle	SBE	unmanipulated	2.3
pigeon	SBE	unmanipulated	2.3
resist	SBE	unmanipulated	1.6
skinny	SBE	unmanipulated	2
spirit	SBE	unmanipulated	2
widow	SBE	unmanipulated	2.6
picture	SBE	unmanipulated	4.4

B. Experiment 1 Structure

semantic relationship	prime	target	word status	item	condition
match	наздраве /nəz’drave/	cheers	word	target	BG - BA unmanipulated
match	крило /kri’lɔ/	wing	word	target	BG - BA unmanipulated

unmatch	затвор /zə’tvɔr/	prisov	non word	target	BG - BA unmanipulated
match	шкаф /ʃkaf/	switch	word	target	BG - BA unmanipulated
match	облачен /ɔbləchen/	cloudy	word	filler	BG - BA unmanipulated
match	следвам /sledvəm/	follow	word	filler	BG - BA unmanipulated
match	неистов /ne’istof/	frantip	non word	filler	BG - BA unmanipulated
unmatch	осем /ɔsem/	recess	word	filler	BG - BA unmanipulated
match	целувка /t͡selufkə/	kiss	word	target	BG - BA manipulated
match	розов /rɔzof/	pink	word	target	BG - BA manipulated
match	условие /us’lɔvie/	conbition	non word	target	BG - BA manipulated
unmatch	девойка /de’vɔjkə/	gypsies	word	target	BG - BA manipulated
match	безгрешен /bez’greʃen/	flawless	word	filler	BG - BA manipulated
match	древен /dreven/	ancient	word	filler	BG - BA manipulated
match	маймуна /məj’munə/	montey	non word	filler	BG - BA manipulated
unmatch	нащрек /nə’ʃtrek/	dapper	word	filler	BG - BA manipulated
match	artificial	изкуствен /iz‘kustven/	word	target	SBE manipulated - BG

match	rich	богат /bo'gat/	word	target	SBE manipulated - BG
match	simple	проск /prɔsk/	non word	target	SBE manipulated - BG
unmatch	shift	сянка /sjankə/	word	target	SBE manipulated - BG
match	doll	кукла /kuklə/	word	filler	SBE manipulated - BG
match	ask	питам /pitəm/	word	filler	SBE manipulated - BG
match	taste	вкун /vkun/	non word	filler	SBE manipulated - BG
unmatch	private	мивка /mivkə/	word	filler	SBE manipulated - BG
match	initial	начален /nə'chalen/	word	target	SBE unmanipulated - BG
match	pigeon	гълъб /gələb/	word	target	SBE unmanipulated - BG
match	difficult	трумен /trumen/	non word	target	SBE unmanipulated - BG
unmatch	miracle	яйца /jəj'ts͡a/	word	target	SBE unmanipulated - BG

match	bury	погребвам /pogrebvəm/	word	filler	SBE unmanipulated - BG
match	veil	завеса /za'vesə/	word	filler	SBE unmanipulated - BG
match	chalk	тебешим /tebe'ʃim/	non word	filler	SBE unmanipulated - BG
unmatch	join	кисел /kisel/	word	filler	SBE unmanipulated - BG
match	кухня /kuxnq/	kitchen	word	target	BG - SBE manipulated
match	устна /usnə/	lip	word	target	BG - SBE manipulated
match	подарък /po'daryk/	gifp	non word	target	BG - SBE manipulated
unmatch	шал /ʃal/	big	word	target	BG - SBE manipulated
match	създател /səz'datel/	creator	word	filler	BG - SBE manipulated
match	лепило /le'pilo/	glue	word	filler	BG - SBE manipulated
match	твърд /tvərt/	solip	non word	filler	BG - SBE manipulated
unmatch	ведро /ve'drɔ/	labourer	word	filler	BG - SBE manipulated
match	вдовица /vdo'vit͡sə/	widow	word	target	BG - SBE unmanipulated
match	живея /ʒi'vejə/	live	word	target	BG - SBE unmanipulated

match	подобен /po’dɔben/	sinilar	non word	target	BG - SBE unmanipulated
unmatch	марка /markə/	spirit	word	target	BG - SBE unmanipulated
match	зъб /zəp/	tooth	word	filler	BG - SBE unmanipulated
match	корона /ko’rɔnə/	crown	word	filler	BG - SBE unmanipulated
match	цветя /t͡sve’tja/	floweps	non word	filler	BG - SBE unmanipulated
unmatch	гния /gnijə/	badge	word	filler	BG - SBE unmanipulated
match	finish	finish	word	target	SBE monolingual
match	issue	issue	word	target	SBE monolingual
match	resist	resisk	non word	target	SBE monolingual
unmatch	skinny	picture	word	target	SBE monolingual
match	worm	worm	word	filler	SBE monolingual
match	robust	robust	word	filler	SBE monolingual
match	cute	cufe	non word	filler	SBE monolingual
unmatch	delay	scale	word	filler	SBE monolingual
match	liver	liver	word	target	BA monolingual

match	district	district	word	target	BA monolingual
match	think	thinp	non word	target	BA monolingual
unmatch	admission	vinegar	word	target	BA monolingual
match	song	song	word	filler	BA monolingual
match	tumble	tumble	word	filler	BA monolingual
match	stop	stoch	non word	filler	BA monolingual
unmatch	sprout	nostalgic	word	filler	BA monolingual
match	mince	кайма /kəj'ma/	word	target	BA unmanipulated - BG
match	give	давам /davəm/	word	target	BA unmanipulated - BG
match	sticky	леткав /letkəv/	non word	target	BA unmanipulated - BG
unmatch	consider	напредък /nə'predək/	word	target	BA unmanipulated - BG
match	brush	четка /tʃetkə/	word	filler	BA unmanipulated - BG
match	thunder	гръм /grəm/	word	filler	BA unmanipulated - BG

match	island	острок /ɔstrok/	non word	filler	BA unmanipulated - BG
unmatch	versed	сода /sɔdə/	word	filler	BA unmanipulated - BG
match	brick	тухла /tuxlə/	word	target	BA manipulated - BG
match	delicious	вкусен /vkusen/	word	target	BA manipulated - BG
match	wish	жебание /ʒebanie/	non word	target	BA manipulated - BG
unmatch	familiar	поляна /po'ljanə/	word	target	BA manipulated - BG
match	condemned	осъден /o'syden/	word	filler	BA manipulated - BG
match	smell	миризма /miriz'ma/	word	filler	BA manipulated - BG
match	mine	мох /mɔx/	non word	filler	BA manipulated - BG
unmatch	vanish	магазин /məgəzin/	word	filler	BA manipulated - BG
match	пипам /pipəm/	пипам /pipəm/	word	target	BG monolingual
match	тих /tix/	тих /tix/	word	target	BG monolingual

unmatch	стая /stajə/	колела /kole'la/	word	target	BG monolingual
match	фалшив /falshif/	фалпив /falpif/	non word	target	BG monolingual
match	ръка /rə'ka/	нищо /niʃto/	word	filler	BG monolingual
match	дърво /dər'vɔ/	дърво /dər'vɔ/	word	filler	BG monolingual
match	мравки /mravki/	мравпи /mravpi/	non word	filler	BG monolingual
unmatch	свободен /svo'bɔden/	легло /leg'lɔ/	word	filler	BG monolingual

C. Experiment 2 Foreign accentedness ratings of target words

List of target words used as stimuli in the main and follow-up experiments. The Foreign Accent Score column represents the average foreign accent rating received by the word when pronounced by Bulgarian-English bilinguals minus the average rating received by the word when pronounced by a native English speaker. The ratings could vary between 0 and 8. For the full description of the rating procedure see Section 7.3.4.3.

	Word	FAS difference
1	back	5.67
2	bed	4.67
3	best	5.92
4	big	4.59
5	black	5.21
6	book	5.91
7	both	5.20
8	case	5.85
9	cent	4.44
10	child	4.40
11	comes	5.21
12	could	4.72
13	death	5.13
14	door	4.31
15	down	5.76
16	first	4.25
17	friend	4.86
18	front	6.12
19	gone	5.66

20	good	6.11
21	got	6.77
22	great	5.10
23	ground	4.81
24	group	5.99
25	had	5.98
26	hair	5.15
27	hands	4.67
28	hard	5.87
29	have	4.72
30	kind	4.89
31	land	4.91
32	last	4.98
33	least	5.01
34	most	6.36
35	much	6.44
36	must	5.90
37	need	5.31
38	next	5.91
39	night	5.99
40	place	6.13
41	point	5.91
42	real	4.62
43	rest	6.27
44	right	5.02

45	road	7.35
46	room	5.53
47	sense	4.85
48	state	5.73
49	still	4.82
50	street	5.34
51	take	4.26
52	ten	6.17
53	than	4.95
54	them	6.08
55	third	5.45
56	through	6.08
57	times	4.19
58	town	4.34
59	true	5.22
60	week	4.31
61	what	6.41
62	whom	5.26
63	whose	4.94
64	wrong	6.25

D. Experiment 2 Pseudo words

List of the 64 pseudo words from the ARC database, used as stimuli in the main and follow-up experiment.

	Pseudo word	Pronunciation
1	beiffed	bi:ft
2	blit	blɪt
3	bloarph	blɔ:f
4	cirnte	sɜ:nt
5	clese	klis
6	cluite	klut
7	cuiched	kut͡ʃd
8	daimth	daɪmθ
9	dasp	dæsp
10	derlt	dɜ:lt
11	dreage	dri:d͡ʒ
12	dweigh	dweɪ
13	dwugg	dwʌg
14	fruice	fru:s
15	gnabes	neɪbz
16	gnaufk	nɔ:fk
17	gneitch	neɪt͡ʃ
18	gnorged	nɔ:d͡ʒd
19	gnysp	nɪsp
20	hapth	hæpθ
21	helch	hɛlt͡ʃ

22	hirshed	hɜ:ʃt
23	hulve	hʌlv
24	keeppth	ki:pθ
25	keighndge	kɪnd͡ʒ
26	kiente	ki:nt
27	knoined	nɔɪnd
28	milp	mɪlp
29	moarmed	mɔ:md
30	molge	məld͡ʒ
31	moogs	mu:gz
32	mriz	mrɪz
33	mubb	mu:b
34	mysp	mɪsp
35	neult	nult
36	nirthed	nɜ:θt
37	nobed	nəʊbd
38	novvs	nɔvz
39	phroge	frɔ:d͡ʒ
40	plorch	plɔ:t͡ʃ
41	quaish	kweɪʃ
42	quaugue	kwæg
43	queape	kwi:p
44	quoch	kwɔ:t͡ʃ
45	ronck	rɔŋk
46	slaughsh	slɔ:ʃ
47	swin	swi:n

48	taved	teɪvd
49	tept	tɛpt
50	toamth	təʊmθ
51	tolth	tɔlθ
52	toolt	tu:lt
53	trirph	trɜ:f
54	twares	tweɪz
55	twaste	twæst
56	tweith	tweɪθ
57	twoart	twɔ:t
58	twoumb	twum
59	veps	vɛps
60	whegs	wɛgz
61	whoobbed	wubd
62	wighst	wɪgst
63	woocs	wuks
64	zepped	zɛpt

Appendix 2 Questionnaires

A. Questionnaire from Experiment 1

	Participant Number
PART I: GENERAL DEMOGRAPHIC INFORMATION	
Age...Gender................................Handedness…….... Current level of study or highest level achieved..…… Have you lived in other places in which languages (other than Bulgarian and English) are spoken? .... If so, where and for how long?... .... Place of birth (country):...	
PART II: LANGUAGES SPOKEN (You can communicate verbally even at a basic level)	
PART III: PARENTS' MOTHER TONGUES:	
What is your father's mother tongue? Does your father speak any other languages?	
What is your mother's mother tongue? Does your mother speak any other languages?	
If you have a child, which is their first language?	

Does your child speak any other languages? Please specify						
BULGARIAN:						
LANGUAGE HISTORY – ACQUISITION OF LANGUAGE						
1. First contact with the language:						
2. How long have you lived in Bulgaria? (years)						
3. Environment in which you used the language **in childhood**: Frequency of use (choose one):						
x/40	Always 4	Often 3	Sometimes 2	Rarely 1	Never 0	Not applicable
Family						
Mother						
Father						
Grandparents						
Siblings						
Other relatives						
Official						
Schooling						
Teachers						
Classmates						
Immediate Environment						
Friends						
Neighbours						

LANGUAGE USE

1. Do you continue to use the language? Yes☐ No☐ (If no, when did you stop using it? ___ yrs old)						
If yes, how often do you use it in each one of the following contexts (choose one)						
18/36	Always	Often	Sometimes	Rarely	Never	Not applicable
Family						
Partner						
Siblings/Nephews/Nieces						
Children						
Other relatives						
Official						
Colleagues						
Shopping						
Radio/TV						
Books/magazines						
Immediate environment						
Friends						
Neighbours						
Society						
English:						
LANGUAGE HISTORY – ACQUISITION OF LANGUAGE						
1. First contact with the language: at __ yrs of age						
2. How long have you lived in an English-speaking country? (please, specify country and time period) Britain – 6 years						
3. Environment in which you used the language **in childhood**: Frequency of use (choose one):						
6/40	Always	Often	Sometimes	Rarely	Never	Not applicable

Family						
Mother						
Father						
Grandparents						
Siblings						
Other relatives						
Official						
Schooling						
Teachers						
Classmates						
Immediate Environment						
Friends						
Neighbours						

LANGUAGE USE						
How often do you **currently** use it in each one of the following contexts (choose one)						
	Always	Often	Sometimes	Rarely	Never	Not applicable
Family						
Partner						
Siblings/Nephews/Nieces						
Children						
Other relatives						
Official						
Colleagues						
Shopping						

Radio/TV						
Books/magazines						
Immediate environment						
Friends						
Neighbours						
Society						

Attitudes:

For the following questions circle one answer, which best describes how you feel between the two extreme options 1 and 9.

How do you feel about Bulgarian people who speak English with a heavy Bulgarian accent?

I like strong Bulgarian accent in English. 1. 2. 3. 4. 5. 6. 7. 8. 9. I dislike strong Bulgarian accent in English.

Do you feel it's important for you to maintain contacts with Bulgarian people abroad?

I think it's very important. 1. 2. 3. 4. 5. 6. 7. 8. 9. I try to avoid Bulgarians.

Do you try to stay up to date with the cultural life in Bulgaria (e.g., follow the latest music, literature, pop culture, cinema releases)?

I try to stay up to date. 1. 2. 3. 4. 5. 6. 7. 8. 9. I don't put any effort into staying up to date.

Do you try to stay up to date with the political life in Bulgaria?

I try to stay up to date. 1. 2. 3. 4. 5. 6. 7. 8. 9. I don't put any effort into staying up to date.

Do you feel it's important for you to maintain contacts with native English speakers?

I think it's very important. 1. 2. 3. 4. 5. 6. 7. 8. 9. I try to avoid native English speakers.

Do you try to stay up to date with the cultural life in the UK/USA/other English-speaking countries (e.g., follow the latest music, literature, pop culture, cinema releases)?

I try to stay up to date. 1. 2. 3. 4. 5. 6. 7. 8. 9. I don't put any effort into staying up to date.

Do you try to stay up to date with the political life in the UK/USA/other English-speaking countries?

I try to stay up to date. 1. 2. 3. 4. 5. 6. 7. 8. 9. I don't put any effort into staying up to date.

B. Questionnaire from Experiment 2

What is your gender?

- Female

- Male

- Other or Prefer not to say

What is your age?

- {min=16, max=120} Slide the bar to enter your answer and continue.

What languages did you speak regularly when growing up?

-

What languages do you speak regularly now?

-

What variety of English do aim to sound like when you speak (e.g., British English as taught at school, North American, Scottish, Australian etc.)? Be as specific as you want.

-

Do you have Bulgarian friends or Bulgarian family members in the UK with whom you sometimes speak in English? If yes, how many such friends or family members do you have?

-

Are you left-handed or right-handed?

- left-handed

- right-handed

- Prefer not to say

On a WEEKLY basis, what percentage of your time talking to people is spent speaking in English with *native* speakers of English? If you speak only in English and only with native speakers of English, it's 100%. If you never speak with native speakers of English it's 0%, and/or if you never use English, it's 0%. If you don't fall in these extremes, estimate the most accurate intermediate answer.

(*Седмично* колко процента от времето си прекарвате в говорене на английски с хора, чийто роден език е английски? Ако говорите само на английски и само с хора, чийто роден език е английският, изберете 100%. Ако никога не говорите с хора, чийто роден език е английски, и/или никога не говорите на английски изберете 0%. Ако не спадате в тези крайности, преценете кой е най-точният междинен отговор.)

- {min=0, max=100, start = 50} Slide the bar to enter your answer and continue (even if it's 0).

What percentage of those interactions are with speakers with an accent from England (as opposed to American, Scottish etc.)?

(Колко процента от тези разговори е с хора, чийто акцент е от Англия (а не Америка, Шотландия и т.н.)?)

- {min=0, max=100, start = 50} Slide the bar to enter your answer and continue (even if it's 0).

On a WEEKLY basis, what percentage of your time talking to people is spent speaking in English with *non-native* speakers of English? If you speak only in English and only with non-native speakers of English, it's 100%. If you never speak in English with non-native speakers of English it's 0%. If you don't fall in these extremes, estimate the most accurate intermediate answer.

(*Седмично* колко процента от времето си прекарвате в говорене на английски с хора, чийто роден език НЕ е английски? Ако говорите само на английски и само с хора, чийто роден език НЕ е английският, изберете 100%. Ако никога не говорите на английски с хора, чийто роден език НЕ е английски, изберете 0%. Ако не спадате в тези крайности, преценете кой е най-точният междинен отговор.)

- {min=0, max=100, start = 50} Slide the bar to enter your answer and continue (even if it's 0).

What percentage of those interactions in English include speakers whose native language is Bulgarian?

(Колко процента от тези разговори на английски са с хора, чийто роден език е български?)

- {min=0, max=100, start = 50} Slide the bar to enter your answer and continue (even if it's 0).

At what age did you first move to a primarily English-speaking country?

- {min=0, max=120} Slide the bar to enter your answer and continue (even if it's 0).

At what age did you actively start studying English?

- {min=0, max=120} Slide the bar to enter your answer and continue (even if it's 0).

C. Questionnaire from Follow-up experiment

What is your gender?

- Female
- Male
- Other
- Prefer not to say

What is your age?

- {min=16, max=120} Slide the bar to enter your answer and continue.

What languages did you speak regularly when growing up?

-

What languages do you speak regularly now?

-

What variety of English do you speak?

-

Please select the place where you were raised and currently reside:

- UK
- USA

What accents do you think you heard during the experiment?

-

Do you have Bulgarian background?

- yes
- no

Have you ever studied Bulgarian

- yes

- no

Are you left-handed or right-handed?

- left-handed

- right-handed

- Prefer not to say

On a WEEKLY basis, what percentage of your time is spent speaking in English with *native* speakers of English? If you speak only in English and only with native speakers of English, it's 100%. If you never speak with native speakers of English it's 0%. If you don't fall in these extremes, estimate the most accurate intermediate answer.

- {min=0, max=100, start = 50} Slide the bar to enter your answer and continue (even if it's 0).

What percentage of those interactions are with speakers with an accent from England (as opposed to American, Scottish, Australian etc.)?

- {min=0, max=100, start = 50} Slide the bar to enter your answer and continue (even if it's 0).

On a WEEKLY basis, what percentage of your time is spent speaking with *non-native* speakers of English? If you speak only in English and only with non-native speakers of English, it's 100%. If you never speak with non-native speakers of English it's 0%. If you don't fall in these extremes, estimate the most accurate intermediate answer.

- {min=0,max=100, start = 50} Slide the bar to enter your answer and continue (even if it's 0).

What percentage of those interactions in English include speakers whose native language is Bulgarian?

- {min=0,max=100, start = 50} Slide the bar to enter your answer and continue (even if it's 0).

Which definition of a bilingual matches your understanding the most? (Select all that apply)

- Someone who uses 2 languages with high mastery regardless of the age at which they learned them.
- Someone who has learned two languages in infancy regardless of their current mastery.
- Someone who has learned two languages in infancy and uses both with high mastery.
- Someone who uses two languages regardless of their mastery or the age at which they learned them.
- None of the above

Appendix 3 Scripts

A. Formant editing script by Matthew Winn

```
# # ## ### ##### ########  #############  #####################
# Modified by Maria Dokovova on 30.04.2017 to include more
# opportunities to pause and inspect the process.
#################################################################
# GUI-based wizard for creating realistic vowel formant continua
# from modified natural speech
#
# version 30
#
# Matthew B. Winn
# April 2014
##################################
#####################
#############
########
#####
###
##
#
#
form Enter settings for Formant Continuum

 comment Formant and pitch analysis
	natural Enter_number_of_steps_in_the_formant_continuum 2
	natural Number_of_formants 5
	natural Maximum_formant_(Hz) 5500
	natural minimum_pitch_in_analysis_(Hz) 75
	natural maximum_pitch_in_analysis_(Hz) 250

 # select which formants you want to modify
	boolean modify_f1 1
	boolean modify_f2 1
	boolean modify_f3 1

 #override bandwidths?
	boolean Override_bandwidths 1
 #Enter settings for frequency blending
	real Crossover_frequency_to_restore_original_signal_(Hz) 4000
	real Width_of_filter_crossover_(Hz) 500

 comment LPC resampling
```

```
		optionmenu LPCchoice 3
		option Conventional_decimation
		option Variable_clean_decimation
		option Variable_decimation

	comment File settings
		real Enter_final_intensity_for_all_new_sounds_(dB) 73

	#Enter directory path for the new files (a folder that already exists)
		sentence Enter_parent_directory_that_already_exists C:\Users\Matt\Desktop\Speech_Continua

	#Enter basic name for the folder of output files
		sentence Basic_name_for_new_folder_for_continuum_files Continuum_from_script

	#Enter prefix for the filenames (to be suffixed by continuum step number)
		sentence Enter_prefix_for_the_filenames Step_
endform

##############

call nameAdjustFromForm

call setVariables

call select2Sounds master reference

call selectManipulationPortions

call alterDurationAndOnsetOfRefSound

call selectSegment precursor
call selectSegment postcursor

call checkDuration precursor overlapDuration
call checkDuration postcursor overlapDuration

call recordDurations

call userAdjustPitchContour "'master$'" minpitch maxpitch

call extractLPportion "'master$'_PitchAdjusted" highpassCutoff skirt
call extractHPportion "'master$'_PitchAdjusted" highpassCutoff skirt

call makeSource "'master$'_PitchAdjusted" samplerate

call checkSource

call matchIntensityContour "'master$'_PitchAdjusted_SOURCE" 'master$'_PitchAdjusted_LP_portion 1 'blank$'

call makeFilter "'master$'" numformants maxformant windowLength
	if reference_choice != 2
		call makeFilter "'references$'_Matched" numformants maxformant windowLength
```

```
    endif

call cleanUpFormantGridPoints

call processBandwidths

call makePointVowels

call checkFormantGrid 'master$'

call checkReferenceGrid

call recheckFormantGrids

call formantGridToFormant "'master$'" no
call formantGridToFormant "'reference$'_Matched" no

call makeHeaderRows

call makeContinuumOfFormantGrids

call sourceFilterResynth

call makeSF_LPs

call matchLPintensities

call blendHPportions

call retainOriginalIntensityContours

call extractMiddleParts

call concatenateparts

call concatenateContinuum

call drawFormantTracks

call drawGradientSpectra

call printContinuumInfo

call initiateSave

call majorCleanup

call presentContinuum

#
#
##
```

```
###
#####
########
#############
####################
#################################
########################################### PROCEDURES

procedure makePointVowels
  # uses the "whitened" voice source and makes a set of vowels that should be
  # easily recognizable as point vowels, and schwa.
  # if the user hears the quality of the original vowel (from the master sound),
  # then the source "whitening" did not work well.
        select FormantGrid 'master$'
        Copy... Blank
        for thisFormant from 1 to numformants
                call elideFormantGridPoints Blank 'thisFormant' 0 master_duration
                Remove bandwidth points between... 'thisFormant' 0 master_duration
                startManipulation_'master$'
        endfor

  # make the vowels
        #call makeVowel Blank 'master$_PitchAdjusted_SOURCE_adj_int Point_i 300 2300 2600 3500 70
        #call makeVowel Blank 'master$_PitchAdjusted_SOURCE_adj_int Point_a 800 1150 2400 3400 70
        #call makeVowel Blank 'master$_PitchAdjusted_SOURCE_adj_int Point_u 350 800 2300 3100 70
        #call makeVowel Blank 'master$_PitchAdjusted_SOURCE_adj_int Neutral_schwa 500 1500 2500 3500 70

        call makeVowel Blank 'master$_PitchAdjusted_SOURCE Point_i 290 2350 2600 3500 70
        call makeVowel Blank 'master$_PitchAdjusted_SOURCE Point_a 800 1150 2400 3400 70
        call makeVowel Blank 'master$_PitchAdjusted_SOURCE Point_u 350 800 2300 3100 70
        call makeVowel Blank 'master$_PitchAdjusted_SOURCE Neutral_schwa 500 1500 2500 3500 70

  # put the vowels together
        select Sound Point_i
        plus Sound Point_a
        plus Sound Point_u
        plus Sound Neutral_schwa
        Concatenate recoverably
        select Sound chain
        Rename... Point_vowels
        select TextGrid chain
        Rename... Point_vowels
        select Sound Point_vowels
        plus TextGrid Point_vowels
        View & Edit
        pause Listen to these vowels to ensure good resynthesis.
endproc


procedure makeVowel .formantGrid$ .source$ .vowelName$ .f1 .f2 .f3 .f4 .intensity
        # same F5 for all vowel inputs
        .f5 = 4500
```

```
		select FormantGrid '.formantGrid$'
		Copy... '.vowelName$'
		# populate the new FormantGrid with formant values (input arguments)
		  for .thisFormant from 1 to numformants
				select FormantGrid '.vowelName$'
				Add formant point... '.thisFormant' 0 .f'.thisFormant'
				Add bandwidth point... '.thisFormant' 0.1 (60+(10*'.thisFormant'))
		  endfor
		# filter the source
		  select Sound 'master$'_PitchAdjusted_SOURCE
				plus FormantGrid '.vowelName$'
				Filter
				Rename... '.vowelName$'
		# add high-frequency energy (from the original master sound) to the vowel
				Formula... self [col] + Sound_'master$'_PitchAdjusted_HPportion [col]
		# zero out everything around the manip portion
				do ("Set part to zero...", 0, startManipulation_'master$', "at nearest zero crossing")
				do ("Set part to zero...", endManipulation_'master$', master_duration,"at nearest zero crossing")
				Scale intensity... '.intensity'
		# cleanup
				select FormantGrid '.vowelName$'
				Remove
endproc


procedure initiateSave
		beginPause ("Save files?")
		comment ("Do you want to save the new sound files & continuum info?")

   choice ("Save", 1)
     option ("Yes, save (on a PC)")
     option ("Yes, save (on a Mac) - not enabled yet")
     option ("No, do not save")
    endPause ("Cancel", "OK", 2)

		if save = 1
				mac = 0
				call adjustForMacOrPC
				call saveEverything
		elsif save = 2
				mac = 1
				call adjustForMacOrPC
				pause Saving on a Mac is not tested yet, so this might not work...
				call saveEverything
		endif
endproc

procedure saveEverything
		call makeDirectories
		call saveInfoWindow "'parentDir$''spacer$''mainDir$'" 'outputFileName$'
		call saveSoundFiles
		call makeFileList "'parentDir$''spacer$''mainDir$''spacer$'Stimuli'spacer$'" 'listName$'
```

```
endproc

procedure saveSoundFiles
   # first, adjust numeric name of the continuum step to permit
   # easy alphabeticisation of one- and two-digit numbers
   # i.e. '2' becomes '02' so that 10 doesn't get ordered before 2.
		for thisStep from 1 to 'formantSteps'
			if formantSteps > 9
			  if thisStep < 10
				tempStep$ = "0'thisStep'"
			  else
				tempStep$ = "'thisStep'"
			  endif
			else
			  tempStep$ = "'thisStep'"
			endif
				select Sound 'basename$''thisStep'
				Rename... 'basename$''tempStep$'

		  call saveWavFile "'basename$''tempStep$'" 'parentDir$''spacer$''mainDir$''spacer$'Stimuli'spacer$'
		  call saveFormantGrid "FILTER_step_'thisStep'" 
'parentDir$''spacer$''mainDir$''spacer$'FormantGrids'spacer$'
		endfor

		call saveWavFile "'reference$'" 'parentDir$''spacer$''mainDir$''spacer$'Original_sounds
		call saveWavFile "'master$'" 'parentDir$''spacer$''mainDir$''spacer$'Original_sounds

		call saveWavFile precursor 'parentDir$''spacer$''mainDir$''spacer$'Original_sounds
		call saveWavFile postcursor 'parentDir$''spacer$''mainDir$''spacer$'Original_sounds
endproc


procedure saveFormantGrid .name$ .directory$
		select FormantGrid '.name$'
		Save as text file... '.directory$''spacer$''.name$'.FormantGrid
endproc

procedure saveWavFile .name$ .directory$
		select Sound '.name$'
		Save as WAV file... '.directory$''spacer$''.name$'.wav
endproc

procedure makeFileList .soundDir$ listName$
		Create Strings as file list... 'listName$' '.soundDir$'
		Save as raw text file... 'parentDir$''spacer$''mainDir$''spacer$''listName$'.txt
		select Strings 'listName$'
		Remove
endproc


procedure saveInfoWindow outputDirectory$ outputFileName$
		filedelete 'outputDirectory$''spacer$''outputFileName$'.txt
```

```
		fappendinfo 'outputDirectory$''spacer$''outputFileName$'.txt
endproc


procedure makeDirectories
  # makes new directories - one as the main directory and one for the stimuli
		system mkdir 'parentDir$''spacer$''mainDir$'
		system mkdir 'parentDir$''spacer$''mainDir$''spacer$'Stimuli
		system mkdir 'parentDir$''spacer$''mainDir$''spacer$'Original_sounds
		system mkdir 'parentDir$''spacer$''mainDir$''spacer$'FormantGrids
endproc


procedure drawFormantTracks
  Erase all
  Select outer viewport... 0 6 0 5

  Line width... lineWidth
  for thisStep from 1 to 'formantSteps'
    # create a color gradient between blue & red, based on the step number
		coolgradient = ('thisStep'-1)/('formantSteps'-1)
			r = coolgradient
			g = 0.0
			b = 1-coolgradient
		Colour... {'r','g','b'}

		select FormantGrid FILTER_step_'thisStep'
		Draw... (startManipulation_'master$'-0.015) (endManipulation_'master$'+0.015) 0 maxFormantDraw no
yes lines
  endfor

  # Annotate the dotted lines (currently omitted because it creates a little clutter)
    #One mark bottom... startManipulation_'master$' no yes yes vowel start
    #One mark bottom... endManipulation_'master$' no yes yes vowel end
endproc

procedure drawGradientSpectra
		## draw gradient-colored smoothed spectra from each continuum step
		# Select the area beneath the formant tracks area
		do ("Select outer viewport...", 0, 6, 5, 9.5)

		for thisStep from 1 to formantSteps
		  # create a color gradient between blue & red, based on the step number
		  coolgradient = ('thisStep'-1)/('formantSteps'-1)
			# create rgb blend (starts at blue, ends at red)
			r = coolgradient
			g = 0.0
			b = 1-coolgradient
		  Colour... {'r','g','b'}

		  select Sound 'basename$''thisStep'
			name$ = selected$("Sound")
```

```
			tempDur = Get total duration
			# extract only the manipulated portion
			do ("Extract part...", precursorDuration, (tempDur - postcursorDuration), "rectangular", 1, "no")

			To Spectrum... yes
			Cepstral smoothing... 'smoothing'
			Rename... 'name$'_part_smooth
			select Spectrum 'name$'_part
			Remove
			select Spectrum 'name$'_part_smooth
			Draw... drawHzLow drawHzHigh drawDBLow drawDBHigh yes

			# cleanup
				select Spectrum 'name$'_part_smooth
				plus Sound 'name$'_part
				Remove
		endfor
		# re-select the formant tracks drawing
			Select outer viewport... 0 6 0 5
endproc

procedure concatenateContinuum
  # put them all into a single annotated sound so that they can be viewed & heard together
		select Sound 'basename$'1
		for thisStep from 2 to 'formantSteps'
		  plus Sound 'basename$''thisStep'
		endfor
		Concatenate recoverably

		select Sound chain
		Rename... 'basename$'Continuum
		select TextGrid chain
		Rename... 'basename$'Continuum
endproc


procedure presentContinuum
		select Sound 'basename$'Continuum
		plus TextGrid 'basename$'Continuum
		do ("View & Edit")
endproc

procedure majorCleanup
  # clean up all the remaining objects in the list
		select IntensityTier LPportion
		plus Sound precursor
		plus Sound postcursor
		for n from 1 to 'formantSteps'
		  plus Sound Step_'n'_SF
		  plus Sound Step_'n'_SF_with_HPportion
		  plus Sound Step_'n'_middle
		endfor
```

```
            if reference_choice=1
                        #plus Manipulation 'reference$'
                        #plus DurationTier 'reference$'
                        plus Sound 'reference$'_Matched
                        plus Formant 'reference$'_Matched
            endif
            plus Sound 'master$'_PitchAdjusted_HPportion
            plus Sound 'master$'_PitchAdjusted
            #plus Sound 'master$'_PitchAdjusted_SOURCE
            plus Formant 'master$'
            Remove

            select Sound 'master$'_PitchAdjusted_SOURCE
            Rename... 'master$'_Voice_Source
endproc

procedure selectManipulationPortions
            # always select landmarks for the master sound
            call selectManipulationPortion 'master$'

              # select landmarks for the reference sound IF you chose a reference sound
                if reference_choice != 2
                                    call selectManipulationPortion 'reference$'
                        else
                                    reference$ = "'master$'"
                endif
            manipDuration = endManipulation_'master$' - startManipulation_'master$'
endproc

procedure checkDuration .name$ .duration
  # ensure that a segment is at least as long as that required for temporal overlap.
  select Sound '.name$'
  .tempdur = Get total duration
    if .tempdur <= (.duration/2)
            # sound must be expanded
            beginPause ("Duration")
            comment ("The duration of your '.name$' object is shorter than what is required")
            comment ("for cross-fading (per your overlapDuration.)")
            comment ("The segment must be greater than half the overlapDuration.")
            comment ("You can start over and adjust the overlapDuration in the script (bottom)")
            comment ("or choose one of the following options")

      choice ("alter_segment", 1)
        option ("Add extra silence to the '.name$' segment")
        option ("Shorten the overlap duration")

       endPause ("Cancel", "OK", 2)

            if alter_segment = 1
                        targetDuration = overlapDuration+0.0001
                        if .name$=="precursor"
                           # if it's the precursor, put the silence before the sound
```

```
                    call addOnsetSilence precursor targetDuration
                  else
                    # if it's the postcursor, put the silence after the sound.
                    call addOffsetSilence postcursor targetDuration
                  endif
        else
                  # increase the overlap duration to half the segment duration, minus a little bit
                  overlapDuration = (.tempdur*2)-0.00001
        endif
  endif
endproc

procedure recordDurations
  # establish variables to be used later in the script
        select Sound precursor
                  precursorDuration = Get total duration
        select Sound postcursor
                  postcursorDuration = Get total duration
endproc

procedure addOffsetSilence .name$ .matchToThisDuration
        select Sound '.name$'
        .numChannels = Get number of channels
        Copy... temp
        select Sound '.name$'
        Remove
        do ("Create Sound from formula...", .name$, .numChannels, 0, .matchToThisDuration, samplerate, "0")
        formula... self [col] + Sound_temp [col]
        select Sound temp
        Remove
endproc

procedure addOnsetSilence .name$ .matchToThisDuration
        select Sound '.name$'
        .tempdur = Get total duration
        .numChannels = Get number of channels
        .durSilenceToAdd = .matchToThisDuration - .tempdur
        do ("Create Sound from formula...", "bufferSilence", .numChannels, 0, .durSilenceToAdd, samplerate, "0")
        select Sound '.name$'
        Copy... temp
        select Sound bufferSilence
        plus Sound temp
        Concatenate

        select Sound bufferSilence
        plus Sound temp
        plus Sound '.name$'
        Remove

        select Sound chain
        Rename... '.name$'
endproc
```

```
procedure concatenateparts
    select Sound postcursor
    Copy... temp_postcursor
        for thisStep from 1 to 'formantSteps'
            select Sound precursor
            plus Sound Step_'thisStep'_middle
            plus Sound temp_postcursor
            if overlapDuration > 0
                # Concatenate with cross-fade
                Concatenate with overlap... overlapDuration
            else

                # If the overlap duration is 0, concatenate with no blending
                # if you accidentally set it to be a negative number,
                #   it is coerced to zero here
                overlapDuration = 0
                Concatenate
            endif
            Rename... 'basename$''thisStep'

            Scale intensity... 'finalIntensity'
        endfor
   select Sound temp_postcursor
   Remove
endproc

procedure extractMiddleParts
   # extract only the manipulation portion (user-defined region)
   # from the re-filtered sounds
   for thisStep from 1 to 'formantSteps'
        select Sound Step_'thisStep'_SF_with_HPportion
        Extract part... startManipulation_'master$' endManipulation_'master$' rectangular 1 no
        Rename... Step_'thisStep'_middle
   endfor
endproc


procedure retainOriginalIntensityContours
   # Each step in the continuum is matched to the intensity contour from the original master sound
   # This step occurs before the manipulated portion is boxed out,
   # because the user might use a different segment for the
   # leading / trailing segment.
   for thisStep from 1 to 'formantSteps'
        call matchIntensityContour "Step_'thisStep'_SF_with_HPportion" 'master$' 1 'blank$'
   endfor
endproc


procedure blendHPportions
   # add the high-frequency portion from the original master sound
   # to the low-passed re-filtered continuum steps.
```

```
  for thisStep from 1 to 'formantSteps'
        select Sound Step_'thisStep'_SF_LP
        Formula... self [col] + Sound_'master$'_PitchAdjusted_HPportion [col]
        Rename... Step_'thisStep'_SF_with_HPportion
  endfor
endproc


procedure matchLPintensities
  # Ensure that the manipulated portion of each continuum step
  # has an intensity contour that matches that of the
  # corresponding frequency region in the original master sound.
  for thisStep from 1 to 'formantSteps'
        select Sound Step_'thisStep'_SF_LP
        Scale intensity... 'original_LP_intensity'
  endfor
endproc


procedure makeSF_LPs
  # Take the re-filtered sound,
  # low-pass filter it so that only the frequency region chosen for manipulation
  # is present in the re-filtered signal.
  # Higher-frequency portions are restored from the original signal
  # in a later procedure.
  for thisStep from 1 to 'formantSteps'
        select Sound Step_'thisStep'_SF
        Filter (pass Hann band)... 0 highpassCutoff skirt
        Rename... Step_'thisStep'_SF_LP
  endfor
endproc

procedure checkSource
  # check for "whiteness" of voice source spectrum.
  # this step might be removed, as it is effectively replaced by the
  # "makeVowels" procedure.
  # However, this is the user's chance to filter the signal,
  # in case they want to deliberately change the spectral slope,
  # or remove a spurious peak.
        beginPause ("Check source")
          comment ("For advanced users (optional): ")
          comment ("Check the Source sound object for sound quality")
          comment ("It should be devoid of any perceptible vowel quality.")
          comment ("If you can clearly hear the original vowel, ")
          comment ("consider switching which sound is the master / which is the reference")
          endPause ("Cancel", "OK, I'm done", 2, 2)
endproc

procedure sourceFilterResynth
  # filter the voice source by each step in the formant continuum.
  for thisStep from 1 to 'formantSteps'
        select Sound 'master$'_PitchAdjusted_SOURCE
```

```
            plus FormantGrid FILTER_step_'thisStep'
            Filter
            Rename... Step_'thisStep'_SF
  endfor
endproc


procedure printFormants .formantObject$ .formant .step .start .end .timesteps
  # Print the formant values in the info windows
  print '.step''tab$''.formant''tab$'
  for thisTimeStep to .timesteps
            .timeStepSize = ('.end' - '.start')/('.timesteps'-1)
            .timepoint = (.timeStepSize*'thisTimeStep') + ('.start'-'.timeStepSize')

            # Get the formant value from the formant object
            select Formant '.formantObject$'
            .formantA = Get value at time... '.formant' '.timepoint' Hertz Linear

            # if the value is readable, print it
            if .formantA <> undefined
                        print '.formantA:0''tab$'
            else
              # if it's undefined, print something as a placeholder.
                        print undefined'tab$'
            endif
  endfor
  print 'newline$'
endproc


procedure makeContinuumOfFormantGrids
  for thisFormantStep to 'formantSteps'
    # create the working filter object for this step
            select Formant 'master$'
            Down to FormantGrid
            Rename... FILTER_step_'thisFormantStep'

    # alter the formantgrid
             for thisFormant from 1 to numformants
               # only alter it if it was selected to be altered
               if f'thisFormant'mod = 1
                 # First delete all the existing formant points for that formant
                        call elideFormantGridPoints "FILTER_step_'thisFormantStep'" 'thisFormant' 0
master_duration

                        # Next, insert the entire formant trajectory for that formant,
                        # based on the interpolation between endpoints at this continuum step.
                        call alterFormantGrid "FILTER_step_'thisFormantStep'" "'master$'" "'reference$'_Matched"
thisFormant formantSteps thisFormantStep startManipulation_'master$' endManipulation_'master$' timesteps
               else
                        # if you aren't altering the formant, at least print out the formant values at the appropriate
timepoints
```

```
                    call printFormants "'master$'" thisFormant thisFormantStep startManipulation_'master$'
endManipulation_'master$' timesteps
            endif
        endfor
  endfor
endproc


procedure alterFormantGrid .filter$ .master$ .reference$ .formant .numberOfSteps .thisStep .start .end .timesteps
  # Populate a formant row in a FormantGrid
  # across the specified number of timepoints.

  # First, delete all formant points in this formant row
        select FormantGrid '.filter$'
        Remove formant points between... '.formant' .start .end

  # the formant number is the first column in the table of printed values
        print '.thisStep''tab$''.formant''tab$'

  for thisTimeStep to .timesteps
        ## convert 'thisTimeStep' to an actual time value by interpolation
        .timeStepSize = ('.end' - '.start')/('.timesteps'-1)
        .timepoint = (.timeStepSize*('thisTimeStep'-1)) + '.start'

        # Get the formant value from the master sound
              select Formant '.master$'
              .formantA = Get value at time... '.formant' '.timepoint' Hertz Linear

        # Get the formant value from the reference sound
              select Formant '.references$'
              .formantB = Get value at time... '.formant' '.timepoint' Hertz Linear

        # Proceed only if the formant at this timepoint is a valid number in BOTH
        # the master & reference sound.
        if .formantA <> undefined && .formantB <> undefined
          if bark = 0
              ## interpolate the formant value (linear) for this step at this timepoint
                    .formantStep = ('.formantB' - '.formantA')/('.numberOfSteps' - 1)
                    .formantInterp =  (.formantStep*('.thisStep'-1)) + .formantA
          else
              ## interpolate the formant value (Bark) for this step at this timepoint
                    call freq2bark .formantA
                          barkA = freq2bark.out
                    call freq2bark .formantB
                          barkB = freq2bark.out
                    barkStep = (barkB-barkA)/('.numberOfSteps' - 1)
                    barkInterp = (barkStep*(.thisStep-1))+barkA
                    call bark2freq (barkInterp)
                    # the formant value is the product of interpolation using the bark scale
                    .formantInterp = bark2freq.out
          endif
```

```
        # Add that formant value to the FormantGrid
                    select FormantGrid '.filter$'
                    Add formant point... '.formant' '.timepoint' '.formantInterp'

          # Print out the formant value that you calculated,
          # rounded to the nearest whole number
                    print '.formantInterp:0''tab$'
      else
              # If the formants at this time point are not *both* readable
              # in the master & reference sound,
              # then print a placeholder
              print undefined'tab$'
              # and do not insert a point in the FormantGrid
      endif
  endfor
  print 'newline$'
endproc


procedure makeHeaderRows
  # Create a header row for the printed table for formant values
  # it will include two header rows: one with the time point, and one with the index of the time step.
      print 'tab$'T-step 'tab$'

      # print timestep index
      for t from 1 to timesteps
              print 't''tab$'
      endfor
      print 'newline$'

      print Step'tab$'Formant'tab$'
      for thisTimeStep to timesteps
              timeStepSize = manipDuration/('timesteps'-1)
              timepoint = timeStepSize*('thisTimeStep'-1)

              print 'timepoint:3''tab$'
      endfor
      print 'newline$'
endproc


procedure formantGridToFormant .formantGrid$ .removeOriginal
  # Convert FormantGrid object to Formant object
  # So that you can query values from it.
  # the second argument in the procedure is
  # if you want to remove the original FormantGrid from the list
      select FormantGrid '.formantGrid$'
      To Formant... 0.01 0.1
      if .removeOriginal = 1
              select FormantGrid '.formantGrid$'
              Remove
      endif
```

```
endproc


procedure recheckFormantGrids
		pause (Optional) You can now take a moment to re-check any changes you made to the FormantGrids.
endproc


procedure checkFormantGrid .sound$
  # check & alter the FormantGrid and shape it into the contour that you want.
  select FormantGrid '.sound$'
  View & Edit
		editor FormantGrid '.sound$'
		Set formant range... 0 maxformant
		Select... (startManipulation_'master$'-0.015) (endManipulation_'master$'+0.015)
		Zoom to selection


   # Pause window with some helpful tips
		beginPause ("Alter the formant tracks")
		comment ("Ensure that the formant tracks are smooth and continuous")
		comment ("Delete any spurious points by clicking them (or highlighting a section)")
		comment ("		and press ctrl+alt+T")
		comment ("Add a new formant point by pressing ctrl+T")
		comment ("Switch from one formant to another by pressing ctrl+(formant number)")
		comment ("		(e.g. ctrl+2 for F2)")
		comment ("   ")
		comment (" Note: if this is the master sound and you chose to *not* manipulate")
		comment ("    specific formants, then the contours you choose here will ")
		comment ("    be inherited by each continuum step.")
		comment ("   ")
		comment (" Click OK when you are finished altering the FormantGrid")
		endPause ("Cancel", "OK, I'm done", 2, 2)

		#Close
  endeditor
endproc

procedure checkReferenceGrid
  if reference_choice != 2
		call checkFormantGrid 'references$'_Matched
  else
		# If the user chose to simply make an alter-able copy of one sound,
		# make a copy of that sound's FormantGrid,
		# so that the user can work with two identical filters,
		# and simply change the elements of interest.

		  select FormantGrid 'master$'
		  Copy... 'references$'_Matched
		# now offer the user a chance to change that into the new opposite endpoint
		  call checkFormantGridCopy 'references$'_Matched
  endif
```

```
endproc

procedure checkFormantGridCopy .sound$
  select FormantGrid '.sound$'
  View & Edit
        editor FormantGrid '.sound$'
        Set formant range... 0 maxformant
        Select... (startManipulation_'master$'-0.015) (endManipulation_'master$'+0.015)
        Zoom to selection

   # pause window with helpful tips
        beginPause ("Alter the formant tracks")
        comment ("This is a copy of the FormantGrid that you just created")
        comment ("You can now design a new formant contour")
        comment ("for the opposite end of the continuum.")
        comment ("Delete any formant points by clicking them (or highlighting a section)")
        comment ("          and press ctrl+alt+T")
        comment ("Add a new formant point by pressing ctrl+T")
        comment ("Switch from one formant to another by pressing ctrl+(formant number)")
        comment ("          (e.g. ctrl+2 for F2)")
        comment ("   ")
        comment (" Click OK when you are finished altering the FormantGrid")
        comment (" Do not close the FormantGrid window.")
        endPause ("Cancel", "OK, I'm done", 2, 2)

        Close
  endeditor
endproc


procedure alterFormantGridBandwidths .sound$ .formant .endtime f1BW f2BW f3BW f4BW
  # Override bandwidth tracking by inserting static formant bandwidth values
        select FormantGrid '.sound$'
        Remove bandwidth points between... '.formant' 0 .endtime
        Add bandwidth point... .formant 0.1 f'.formant'BW
endproc


procedure cleanUpFormantGridPoints
  # Remove leading & trailing Formantgrid points
  # So that the user knows exactly when the manipulation portion
  # begins and ends.
  # This is mostly aesthetic, as the portions affected by this step
  # are not included after the middle-portion extraction step.
  for thisFormant from 1 to numformants
        call elideFormantGridPoints "'master$'" 'thisFormant' 0 startManipulation_'master$'
        call elideFormantGridPoints "'master$'" 'thisFormant' endManipulation_'master$' master_duration

        if reference_choice != 2
          call elideFormantGridPoints "'reference$'_Matched" 'thisFormant' 0 startManipulation_'master$'
          call elideFormantGridPoints "'reference$'_Matched" 'thisFormant' endManipulation_'master$'
ref_lengthened_duration
```

```
		endif
	endfor
endproc

procedure processBandwidths
	# Check the user settings for formant bandwidht adjustment.
	# if bandwidth override was chosen, execute that procedure.
	if bandwidthOverride = 1
	  for thisFormant from 1 to numformants
		call alterFormantGridBandwidths "'master$'" thisFormant master_duration f1BW f2BW f3BW f4BW
	  endfor
	endif
endproc

procedure elideFormantGridPoints .formantGrid$ .formant .start .end
		select FormantGrid '.formantGrid$'
		Remove formant points between... '.formant' .start .end
endproc


procedure makeFilter .sound$ numformants maxformant windowLength
	# Make a formant object from the sound
	# Convert it into a FormantGrid, which the user can customize.
	# The resulting FormantGrid will *only* be as good as the Formant object,
	# which is only as good as the settings used to generate it.
	# Those settings are user-specified.
		select Sound '.sound$'
		To Formant (burg)... 0 numformants maxformant windowLength 50
		select Formant '.sound$'
		Down to FormantGrid
		select Formant '.sound$'
		Remove
endproc


procedure matchIntensityContour .soundToAlter$ .referenceSound$ .removeOriginal .suffix$
	# Matches the overall intensity contour of one sound to that of another sound
	# The sounds are first aligned in the time domain.
	# Procedure results in Sound object named '.soundToAlter$''.suffix$'
		select Sound '.soundToAlter$'
		.originalIntensity = Get intensity (dB)

		To Intensity... minpitch 0 yes
		Down to IntensityTier
		select Intensity '.soundToAlter$'
		mean = Get mean... 0 0 dB

		# future versions: avoid extreme values that throw off the intensity contour when flipped
			# if you end up with a sound with undefined (/infinite) intensity
			# (you'll know because the spectrogram is all black),
			# then this is where it went wrong.
			# If that occurs, create a pause after Copy... flipped
```

```
                # and visually inspect it to find dubious extreme values
                # that affect the mean/flipping conversion.
                # Also, maybe check it *after* the flip as well, just in case.
                # that will solve 99% of your problems.
# flip the intensitycontour around its mean
        select IntensityTier '.soundToAlter$'
        Copy... flipped
        pauseScript: "check flipped intensity "
        Formula... -(self-mean)
        pauseScript: "check flipped intensity "

# multiply by the flipped contour to result in a FLAT contour
# (a poor man's Hilbert transform)
        select Sound '.soundToAlter$'
        plus IntensityTier flipped
        Multiply... no
        Rename... '.soundToAlter$'_flattened

# track the intensity contour of the reference sound
# (the one after which the altered sound is modeled)
        select Sound '.referenceSound$'
        To Intensity... minpitch 0 yes
        Down to IntensityTier

#  Clean up original sound - need to do it here before the re-naming occurs
  if .removeOriginal = 1
                select Sound '.soundToAlter$'
                Remove
        endif

# multiply the FLATTENED sound by the reference intensity contour
        select IntensityTier '.referenceSound$'
        plus Sound '.soundToAlter$'_flattened
        Multiply... no
        Rename... '.soundToAlter$''.suffix$'
        pauseScript: "check flipped intensity "
        select IntensityTier '.referenceSound$'
        plus Sound '.soundToAlter$'_flattened
        Multiply... no
        Rename... '.soundToAlter$''.suffix$'

# Match it to the overall intensity so there isn't just an overall level change.
        Scale intensity... '.originalIntensity'

# cleanup remaining objects
        select Intensity '.soundToAlter$'
        plus IntensityTier '.soundToAlter$'
        plus IntensityTier flipped
        plus Sound '.soundToAlter$'_flattened
        plus Intensity '.referenceSound$'
        plus IntensityTier '.referenceSound$'
        Remove
```

```
endproc

procedure findIdealLPCResampleRate
	# returns variable ideal_LPC_resamplerate appropriate for LPC resampling
	# Prevents odd decimation by ensuring that the downsampling factor is a natural number.
	# ensure that the sampling frequency is a whole number.
	# this procedure ONLY gets executed if you choose "clean decimation"
	# as the LPC method, which is not recommended.

	# establish variable
		nyquist = samplerate/2

	# ideally, we would like the LPC analysis range to be the same frequency range
	# used to track the formants.
		idealResample = maxformant*2

	# Goal: establish new resampling frequency that is closest to,
	# but not less than the maximum sampling frequency needed to
	# capture all of the desired formant peaks
	# (i.e. max formant freq. x2)
	# and below the Nyquist frequency

	# create a list of new potential sampling frequencies that are
	# whole-number factors of the orginal rate
		# the loop goes to 10 but you probably only need it to go to 3-5
	for divisor from 1 to 10
		# each of these new variables is a numeric value that can be used as a sampling frquency
		samplerate_simple_factors_'divisor' = samplerate / 'divisor'
	endfor

		# print out all possible sampling frequencies
		# (only implemented for debug mode)
		for i from 1 to 10
			# appendInfoLine(samplerate_simple_factors_'i')
		endfor
	# first initiate variable for the divisor of the sampling frequency
		thisDivisor = 1

	# increase the whole-number divisor until
	# the downsampling rate is lower than the ideal resampling
	# frequency defined by the formant settings
		while samplerate_simple_factors_'thisDivisor' > idealResample
		  thisDivisor = thisDivisor+1
		endwhile

	# After you've crossed that threshold,
	# you have an index of which divisor crossed the threshold.
	# Now get the *previous* index,
	# which produced a sampling frequency with a Nyquist frequency
	# that is higher than your maximum frequency for formant peaks,
	# (plus any extra bandwidth up to the new Nyquist freq.)
		idealDivisor = thisDivisor-1
```

```
	thisTempDivisor = idealDivisor
	# if the number rounded to zero does not equal the number,
	# choose a higher samplerate until it's a natural number

	if samplerate_simple_factors_'thisTempDivisor' mod 0 != samplerate_simple_factors_'thisTempDivisor'
		while samplerate_simple_factors_'thisTempDivisor' mod 0 != samplerate_simple_factors_'thisTempDivisor'
			# then select the previous one -
			# a lower divisor, effectively raising the Nyquist higher
			thisTempDivisor = thisTempDivisor -1
		endwhile
		idealDivisor = thisTempDivisor
	endif

	# establish the new rate by calling it from the list
	# previously made with the loop
		ideal_LPC_resamplerate = samplerate_simple_factors_'idealDivisor'
endproc

procedure makeSource .sound$ samplerate
	if lPCchoice = 1
		# conventional decimation
		lpcFreq = 11025
		numformants = 5

	elsif lPCchoice = 2
		# clean decimation
		cleanDecimate = 1
		call findIdealLPCResampleRate
		# get value from function
		lpcFreq = ideal_LPC_resamplerate

	elsif lPCchoice = 3
		variableDecimate = 1
			# "dirty" decimation -
			# resample based on user formant settings
			# this will yield the best tracking
			# and best spectrum "whitening"
		lpcFreq = maxformant*2
	endif

	# first, anti-alias filter
		lpc_antiAlias_LPF = lpcFreq/2
		antiAlias_filterSkirt = 100

	# resample to prepare for LPC
		select Sound '.sound$'
		Resample... lpcFreq 50
			# yields Sound '.sound$'_'lpcFreq' with weird character if it's not a whole number
			# re-name to ensure whole number in Object name
		Rename... '.sound$'_'lpcFreq:0'
```

```
            target_resid_intensity = Get intensity (dB)

      # create LPC object
            To LPC (burg)... lpcOrder 0.025 0.005 50
                  # yields LPC '.sound$'_'lpcFreq:0'

      # inverse filter the sound by the LPC to get the residual glottal source
            select Sound '.sound$'_'lpcFreq:0'
            plus LPC '.sound$'_'lpcFreq:0'
            Filter (inverse)
            Rename... '.sound$'_'lpcFreq'_reFilt

      # re-sample back up to the original sampling frequency
      # so that it can be combined with other sounds with the original sampling frequency
      # (The down-sampled object is never played as a wav file)
            Resample... samplerate 50
            Rename... '.sound$'_SOURCE
            Scale intensity... target_resid_intensity

      # cleanup
            select Sound '.sound$'_'lpcFreq:0'
            plus Sound '.sound$'_'lpcFreq'_reFilt
            plus LPC '.sound$'_'lpcFreq:0'
            Remove
endproc

procedure userAdjustPitchContour .sound$ .minpitch .maxpitch
  # lets the user adjust the pitch contour
  # "Ensure accurate pulses" means that for every pitch period,
  # there should be a corresponding
      select Sound '.sound$'
      To Manipulation... 0.01 minpitch maxpitch
      select Manipulation '.sound$'
      Edit
      editor Manipulation '.sound$'
            pause manipulate the Pitch contour to your liking. Ensure accurate pulses. Click Continue when
finished
            #Close
      endeditor
      select Manipulation '.sound$'
      Get resynthesis (overlap-add)
      Rename... '.sound$'_PitchAdjusted

      select Manipulation '.sound$'
      Remove
endproc


procedure selectSegment .newname$
  # user is prompted to select a sound object from the list
  # that object is subsequently referred to in the script
  # by the string variable given as an input argument
```

```
# simply to make the script more readable.
        beginPause ("choose the '.newname$' to the manipulated sound")
        comment ("I want to select the segment from...")

  choice ("Choice", 1)
    option ("'master$' file")
    option ("'reference$' file")
    option ("Object in list")
    option ("Silence (see below)")
       positive ("Duration_of_silence", "0.05")
  endPause ("Cancel", "OK", 2)

        if choice = 1
                .sound$ = "'master$'"

        elsif choice = 2
                .sound$ = "'reference$'"
        endif

        if choice <3
                # user chose the master or reference (see above)
                # open up the sound, mark landmarks for the segment
                  select Sound '.sound$'
                  Edit
                  editor Sound '.sound$'
                        pause Get start of the segment to become the '.newname$' THEN click Continue
                        Move cursor to nearest zero crossing
                        .start = Get cursor

                        pause Get end of the '.newname$' segment THEN click Continue
                        Move cursor to nearest zero crossing
                        .end = Get cursor

                        Select... '.start' '.end'
                        Move start of selection to nearest zero crossing
                        Move end of selection to nearest zero crossing
                        Extract selected sound (time from 0)
                        Close
                  endeditor
                  Rename... '.newname$'

                newname_file$ = selected$ ("Sound", 1)
                info'.newname$'$ = "A portion of '.sound$' was used as the '.newname$' file"
        endif

        if choice = 3
                # user chose a sound object from the objects list
                # that complete sound will be renamed as postcursor or precursor
                # and resampled to match to the master sound.
                        pause Click on your '.newname$' sound file in the list, then click Continue.
                        newname_file$ = selected$ ("Sound", 1)
                        info'.newname$'$ = "'newname_file$' was used as the '.newname$' file"
```

```
                    Resample... 'samplerate' 50
                    Rename... '.newnames$'
        endif

        if choice = 4
        silenceDuration'.newnames$' = duration_of_silence
          .tempduration = silenceDuration'.newnames$'
          call makeSilence .tempduration 'samplerate' '.newnames$'
                newname_file$ = selected$ ("Sound", 1)
                info'.newnames$'$ = "'.tempduration' of silence was used as the '.newnames$' file"
        endif

        select Sound '.newnames$'
        duration'.newnames$' = Get total duration
endproc


procedure alterDurationAndOnsetOfRefSound
  # only do this if there *is* a reference sound
  if reference_choice != 2
   # get the duration ratio of of manipulation portions of master & reference sound
        masterManipDuration = endManipulation_'master$' - startManipulation_'master$'
        referenceManipDuration = endManipulation_'reference$' - startManipulation_'reference$'
        master_dur_ratio = masterManipDuration/referenceManipDuration
        master_onset_ratio = startManipulation_'master$'/startManipulation_'reference$'

   # Create a DurationTier that adjusts the manipulation portion by that ratio
          select Sound 'reference$'
          To Manipulation... 0.01 minpitch maxpitch
                Extract duration tier
                 Add point... (startManipulation_'reference$'-0.00001) master_onset_ratio
                 Add point... startManipulation_'reference$' master_dur_ratio
                 Add point... endManipulation_'reference$' master_dur_ratio
                 Add point... (endManipulation_'reference$'+0.00001) 1

        # Use that DurationTier to modify the reference sound
        # to match timing landmarks to the master sound
                select Manipulation 'reference$'
                plus DurationTier 'reference$'
                Replace duration tier
        # Get the PSOLA modified version
                select Manipulation 'reference$'
                Get resynthesis (overlap-add)
                Rename... 'reference$'_Matched
                ref_lengthened_duration = Get total duration

        # cleanup
                select Manipulation 'reference$'
                plus DurationTier 'reference$'
                Remove
endif
endproc
```

```
procedure selectManipulationPortion .sound$
  # select landmarks for onset & offset of manipulation,
  # record the time landmarks in the info window
  # establish sound-specific variable names to record the values
        select Sound '.sound$'
        Edit
        editor Sound '.sound$'
          pause Get start of the segment to manipulate, then click Continue
                Move cursor to nearest zero crossing
                startManipulation_'.sound$' = Get cursor
                temp = Get cursor
                print '.sound$' manipulation landmark (start) = 'temp''newline$'
          pause Get end of the segment to manipulate, then click Continue
                Move cursor to nearest zero crossing
                endManipulation_'.sound$' = Get cursor
                temp = Get cursor
                print '.sound$' manipulation landmark (end) = 'temp''newline$'
          Close
        endeditor
endproc


procedure makeSilence .duration .samplerate .name$
        Create Sound from formula... '.name$' 1 0 .duration '.samplerate' 0
endproc

procedure extractLPportion .sound$ .cutoff .skirt
        select Sound '.sound$'
        Filter (pass Hann band)... 0 .cutoff .skirt
        original_LP_intensity = Get intensity (dB)
        To Intensity... minpitch 0 yes
        Down to IntensityTier
        Rename... LPportion
        select Intensity '.sound$'_band
        Remove
        select Sound '.sound$'_band
        Rename... '.sound$'_LP_portion

endproc

procedure extractHPportion .sound$ .cutoff .skirt
        ## create high-pass portion of the sound
        select Sound '.sound$'
                Filter (stop Hann band)... 0 .cutoff .skirt
                Rename... '.sound$'_HPportion
endproc


procedure select2Sounds .label1$ .label2$
        pause Select the '.label1$' sound
        '.label1$'$ = selected$ ("Sound")
```

```
            samplerate = Get sampling frequency
            numchannels = Get number of channels
            master_duration = Get total duration

    # if sound is stereo, convert to mono
    if numchannels = 2
            beginPause ("Warning about mono conversion ")
                    comment ("The sound that you selected is stereo")
                    comment ("This script cannot preserve dichotic channel differences")
                    comment ("so it is now a mono sound")
             endPause ("Cancel", "OK", 2)

            select Sound 'master$'
            Convert to mono
            Rename... 'master$'_mono
    # re-establish variables
            select Sound 'master$'_mono
            '.label1$'$ = selected$ ("Sound")
            samplerate = Get sampling frequency
            numchannels = Get number of channels
            master_duration = Get total duration
    endif

   beginPause ("Choose the second sound ")
            comment ("Choose the second (reference) sound to be used for analysis")
            comment ("Either select a new sound in the Objects list now to serve as the referent,")
            comment ("or just work with the sound you already selected, with no referent")

    choice ("reference_choice", 1)
      option ("A new object from the list")
      option ("No new sound - just modify the original master sound")
    endPause ("Cancel", "OK", 2)

            if reference_choice = 1
                    pause Select the '.label2$' sound
                      '.label2$'$ = selected$ ("Sound")
            else
                    #print Only one sound used for manipulation: 'master$' 'newline$'
            endif
endproc


procedure adjustForMacOrPC
# adjusts filepath settings for PC/Mac
            if mac = 1
                    spacer$ = "/"
            else
                    spacer$ = "\"
            endif
endproc

procedure printContinuumInfo
```

```
print 'newline$''newline$'
print Continuum script setup information'newline$'
print 'enter_number_of_steps_in_the_formant_continuum''tab$'steps in the continuum'newline$'
print 'number_of_formants''tab$'formants analysed'newline$'
print 'maximum_formant''tab$'maximum frequency for formant analysis'newline$'
print 'minimum_pitch_in_analysis''tab$'minimum pitch for analysis'newline$'
print 'maximum_pitch_in_analysis''tab$'maximum pitch for analysis'newline$'

print 'newline$'
print Which formants were modified: (1 is yes, 0 is no)'newline$'
print    F1: 'f1mod''newline$'
print    F2: 'f2mod''newline$'
print    F3: 'f3mod''newline$'
print    F4: 'f4mod''newline$'

print 'newline$'
print    Bandwidth override: 'bandwidthOverride''newline$'
        if bandwidthOverride = 1
          print      F1BW = 'f1BW''newline$'
          print      F2BW = 'f2BW''newline$'
          print      F3BW = 'f3BW''newline$'
          print      F4BW = 'f4BW''newline$'
          print 'newline$'
        endif

print 'newline$'
print 'crossover_frequency_to_restore_original_signal''tab$'crossover frequency to restore original
signal'newline$'
print 'width_of_filter_crossover''tab$'width of crossover filter'newline$'
print 'enter_final_intensity_for_all_new_sounds''tab$'final intensity for new sounds'newline$'
print 'newline$'
print parent directory:'tab$''enter_parent_directory_that_already_exists$''newline$'
print sub-directory:'tab$''tab$''basic_name_for_new_folder_for_continuum_files$''newline$'
print filename prefix:'tab$''enter_prefix_for_the_filenames$''newline$'

print 'newline$'

print 'infoprecursor$''newline$'
print 'infopostcursor$''newline$''newline$'

temp = startManipulation_'master$'
print Start of manipulation (master sound): 'temp''newline$'

temp = endManipulation_'master$'
print End of manipulation (master sound): 'temp''newline$'

if reference_choice = 2
                print Only one sound used for manipulation: 'master$' 'newline$'
else
        temp = startManipulation_'reference$'
        print Start of manipulation (reference sound): 'temp''newline$'
```

```
            temp = endManipulation_'reference$'
            print End of manipulation (reference sound): 'temp''newline$'
        endif

        print 'newline$'
        print Original sampling fequency: 'samplerate''newline$'
        print 'newline$'
        print LPC order: 'lpcOrder''newline$'
        print LPC resample rate: 'lpcFreq''newline$'
            decimationFactor = samplerate/lpcFreq
        print LPC decimation factor: 'decimationFactor''newline$'
endproc


## uses the Hz to Bark conversion from Traunmuller (1990);
        ## Hz to Bark: (26.81/(1+(1960/f))) - 0.53
        ## Bark to Hz: f = 1960 / [26.81 / (z + 0.53) - 1]

procedure bark2freq .bark
        .out = 1960 / (26.81 / ('.bark' + 0.53) - 1)
endproc

procedure freq2bark .freq
        .out = (26.81/(1+(1960/'.freq'))) - 0.53
endproc

procedure nameAdjustFromForm
  # variable name adjustment from the input settings form
  # (the form name was formatted to be easily readable,
  # and the values are assigned to variable names
  # that are more script-friendly.

        formantSteps = enter_number_of_steps_in_the_formant_continuum
        numformants = number_of_formants
        maxformant = maximum_formant
        minpitch = minimum_pitch_in_analysis
        maxpitch = maximum_pitch_in_analysis

        f1mod = modify_f1
        f2mod = modify_f2
        f3mod = modify_f3

        bandwidthOverride = override_bandwidths

        highpassCutoff = crossover_frequency_to_restore_original_signal

        skirt = width_of_filter_crossover

        finalIntensity = enter_final_intensity_for_all_new_sounds

        basename$ = enter_prefix_for_the_filenames$
```

```
		parentDir$ = enter_parent_directory_that_already_exists$

		mainDir$ = basic_name_for_new_folder_for_continuum_files$

		# sets output (Continuum info) file name using the basename from the folder
				outputFileName$ = "'mainDir$'Continuum_Info"

		# sets name for the file list using the basename from the folder
				listName$ = "'mainDir$'file_list"

		#initialize variables - omitted now (now in-line in the script)
				conventionalDecimate = 0
				cleanDecimate = 0
				variableDecimate = 0

		# activate lpc sampling choice based on option menu
		if lPCchoice = 1
				conventionalDecimate = 1
		elsif lPCchoice = 2
				cleanDecimate = 1
		elsif lPCchoice = 3
				variableDecimate = 1
		endif
endproc


########################################
########################################

procedure setVariables
		## set some variables for the script.
		## The user should adjust these as desired

		clearinfo

		f1BW = 75
		f2BW = 85
		f3BW = 90
		f4BW = 100
		f5BW = 210

		# number of timepoints to extract vowel information. Larger number means more accuracy for
trajectories.
		  # this controls how many timepoints are referenced from the formant tracks
		  # when you create the new stimuli.
		  # Values are interpolated between points
		  # NOTE: if you want steady-state vowels, it is recommended
		  # that you do not set this value to be 1, but rather
		  # Adjust the formant tracks by hand to reflect the contour that you want.
		  # Also, you probably want to maintain the onset & offset transitions
		  # out from and into the adjacent consonant sounds.
		  # for long vowels (including vowels in open syllables),
```

```
  # you probably want this number to be 25 or above
  # in order to properly track consonant transitions.
  # if you're specifically interested in consonant transitions, the more the better.
timesteps = 30

# window length for formant analysis
	windowLength = 0.04

# basic boolean variables
	yes = 1
	no = 0

# basic blank text placehodler
	blank$ = ""

# use bark interpolation instead of linear interpolation of formant frequencies
# set to 0 if you want linear interpolation
	bark = 1

# don't alter the fourth formant
  # set to 1 if you want F4 to vary
  # if you don't track F4, this value wont matter.
  # If F4 is above the frequency cutoff above which you blend in
  # energy from the original signal, this value wont matter.
	modify_f4 = 0
	modify_f5 = 0

# convert variable name from input window
	f1mod = modify_f1
	f2mod = modify_f2
	f3mod = modify_f3
	f4mod = modify_f4
	f5mod = modify_f5

# if you chose to modify F3 but are not tracking F3, alert the user
  if numformants < 3
	f3mod = 0
	pause F3 will not be manipulated because it is not being tracked (check input formant settings)
  endif

if highpassCutoff > maxformant
	highpassCutoff = maxformant
	beginPause ("Crossover filter adjusted")
	comment ("Crossover filter adjusted downward because of formant settings. ")
	comment ("Crossover frequency must be higher than the max frequency for formant analysis")
	comment ("to ensure a full spectrum.")
	endPause ("Cancel", "OK", 2, 2)
endif

# drawing formant tracks after the script is complete
	maxFormantDraw = 3000
	lineWidth = 2
```

```
# drawing spectra after the script is complete
		smoothing = 300
		drawHzLow = 0
		drawHzHigh = 6000
		drawDBLow = -15
		drawDBHigh = 45

# alter LPC parameters based on the user's input formant analysis settings
# LPC order = two poles for each formant in the freq range,
# plus two, to add two poles for sound source (spectral tilt)
		lpcOrder = (numformants*2)+2

		# if conventional LPC analysis, it is almost always the case that you want at least 10 poles
		### if you want to change it (as for chidrens' voices
		# or other difficult voices), change it here
			if lpcOrder < 10
			   if conventionalDecimate = 1
			      lpcOrder = 10
			   endif
			endif


# set overlap duration between segment onset, vowel and offset
  # HALF of this duration will be the length of the cross-fade time for each part
  # if you set this to zero, then it's a straight concatenation
  # the larger this value, the smoother the transition between the segments,
  # but you lose some info in each segment (because more of it is faded out)
		overlapDuration = 0.003

endproc
```

B. Formant measuring script by Bert Remijsen

```
#######################################################
#
# NAME: msr&check_formants_batch.psc
#
# INPUT: - soundfile (samplingfreq above 16000 kHz)
#        - TextGrid with segmentation for vowel(s)
#
#
# DESCRIPTION:
#
# This script calculates F1 and F2 at the midpoint of a
# specific segment in a TextGrid file. The procedure is
# repeated made for each occurrence of that segment in the
# TextGrid. The label needs to be specified by the user.
# F1 and F2 are calculated using 'To formant (burg)' and
# and the tracker. Both of these algorithms set parameters
# as a function of speaker sex. This parameter is controlled
# by the user. The Picture window shows the spectrogram and
# formant tracks (F1 & F2), and The rounded F1 and F2 values
# appear at the top.
#
# As an additional check, and for voice quality measurements,
# the script displays the spectrum, the Long-term average
# spectrum (Ltas) and the LPC spectrum at the bottom. The
# LPC spectrum at the bottom uses Praat's 'autocorrelation'
# algorithm.
#
# When the 'To formant' and 'Track...' procedures do not
# produce plausible formant values, the user can (1) run the
# script again with new tracking values, (2) on the basis
# of the spectrum/Ltas/LPC display at the bottom part of
# the Picture window, determine F1 and F2 by hand using
# the Spectrum/Ltas/LPC in the Object window.
#
# The script can be modified to produce measurements on voice
# quality with a check.
#
# BY:   Bert Remijsen
# DATE: 20/FEB/2013
#
#
#######################################################
form Calculate F1 & F2 for a specific segment (batch)
  comment See header of script for details.
  comment Directory of input and output files; search term; output file:
  word directory \\mull.sms.ed.ac.uk\Home\s1107659\Win7\Desktop\fast
```

```
  word searchterm *
  word outputfile output.txt
  comment The label of segments to be measured, and the tier in the TextGrid, and the speaker's sex:
  word the_label v
  integer the_tier 1
  optionmenu sex 1
  option male
  option female
  optionmenu play 1
  option yes
  option no
  comment Settings for Track... algorithm (MALE on the left; FEMALE on the right)
  positive left_F1_reference 500
  positive right_F1_reference 550
  positive left_F2_reference 1485
  positive right_F2_reference 1650
  positive left_F3_reference 2475
  positive right_F3_reference 2750
  positive left_Frequency_cost 1
  positive right_Frequency_cost 1
  positive left_Bandwidth_cost 1
  positive right_Bandwidth_cost 1
  positive left_Transition_cost 1
  positive right_Transition_cost 1
endform

clearinfo

# Recursion over files
Create Strings as file list... listfile 'directory$'\'searchterm$'.TextGrid
end = Get number of strings
for filecounter from 1 to end
 select Strings listfile
 file$ = Get string... filecounter
 Read from file... 'directory$'\'file$'
 textgridID = selected("TextGrid")
 filebare$ = file$ - ".TextGrid"
 Read from file... 'directory$'\'filebare$'.wav
 soundID = selected("Sound")
 select 'textgridID'
 plus 'soundID'
 counter = 0

 select 'textgridID'
 finishing_time = Get finishing time
 nlabels = Get number of intervals... 'the_tier'
 for label from 1 to 'nlabels'
  select 'textgridID'
  labelx$ = Get label of interval... 'the_tier' 'label'
  if (labelx$ = the_label$)
   counter = counter + 1
   n_b = Get starting point... 'the_tier' 'label'
```

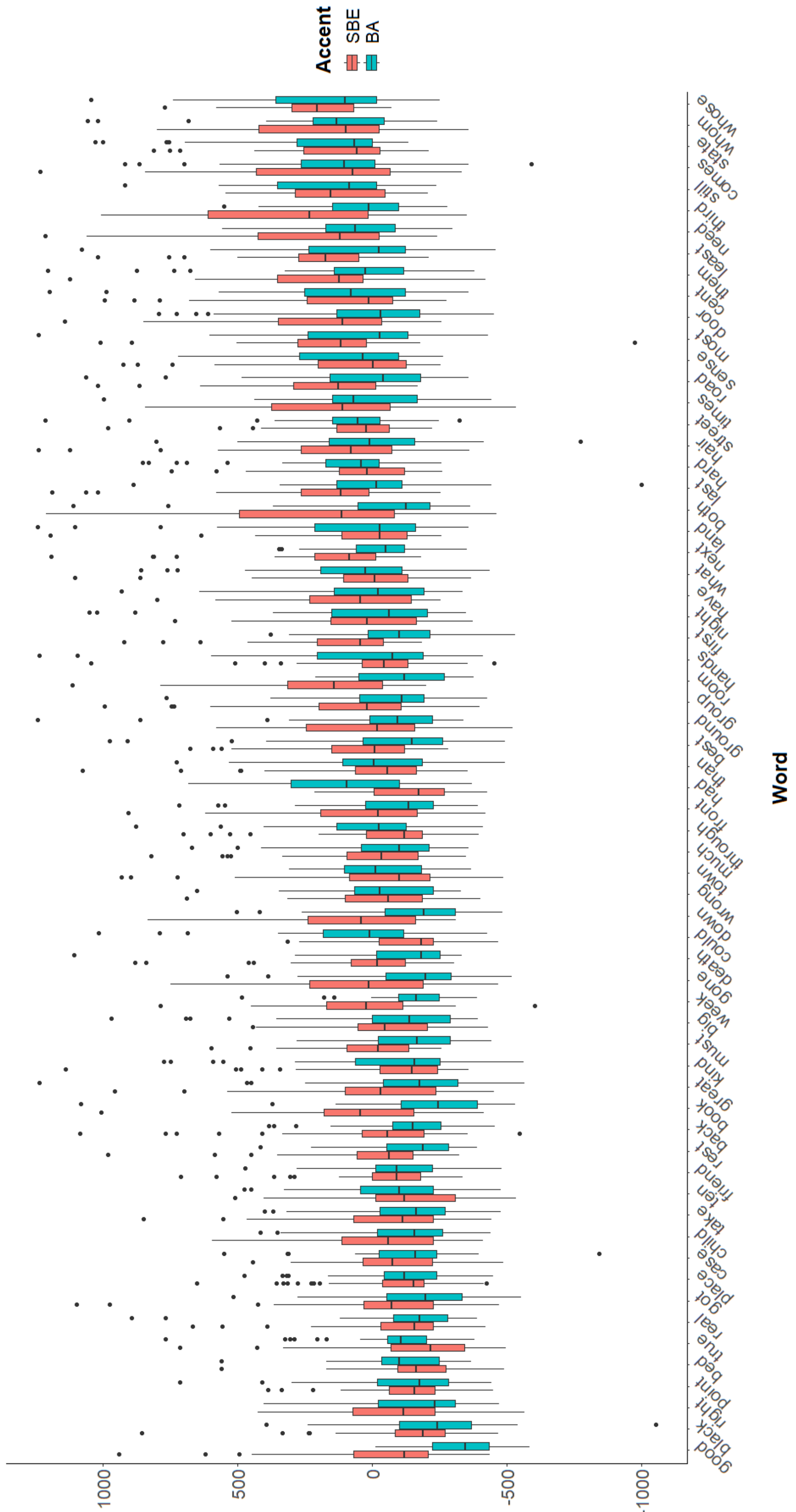
```
      n_e = Get end point... 'the_tier' 'label'
      n_d = 'n_e' - 'n_b'
      n_md = ('n_b' + 'n_e') / 2
      call vowelq 'n_b' 'n_e' 'n_md' 'filebare$'
    endif
    select 'textgridID'
    plus 'soundID'
  endfor

# Recursion over files
  select 'textgridID'
  plus 'soundID'
  Remove
endfor

procedure vowelq n_b n_e n_md filebare$
# set maximum frequency of Formant calculation algorithm on basis of sex
# sex is 1 for male (left); sex is 2 for remale (right).
  if ('sex' = 1)
    maxf = 5000
   f1ref = left_F1_reference
   f2ref = left_F2_reference
   f3ref = left_F3_reference
   f4ref = 3465
   f5ref = 4455
   freqcost = left_Frequency_cost
   bwcost = left_Bandwidth_cost
   transcost = left_Transition_cost
  endif
  if ('sex' = 2)
    maxf = 5500
   f1ref = right_F1_reference
   f2ref = right_F2_reference
   f3ref = right_F3_reference
   f4ref = 3850
   f5ref = 4950
   freqcost = right_Frequency_cost
   bwcost = right_Bandwidth_cost
   transcost = right_Transition_cost
  endif
  select 'soundID'
  Resample... 16000 50
  sound_16khz = selected("Sound")
  To Formant (burg)... 0.01 5 'maxf' 0.025 50
  Rename... 'filebare$'_beforetracking
  formant_beforetracking = selected("Formant")
  Track... 3 'f1ref' 'f2ref' 'f3ref' 'f4ref' 'f5ref' 'freqcost' 'bwcost' 'transcost'
  Rename... 'filebare$'_aftertracking
  formant_aftertracking = selected("Formant")
  Save as text file... 'directory$'\'filebare$'.Formant

# Get the f1,f2,f3 measurements.
```

```
 select 'formant_aftertracking'
 f1hzpt = Get value at time... 1 'n_md' Hertz Linear
 f2hzpt = Get value at time... 2 'n_md' Hertz Linear
 f3hzpt = Get value at time... 3 'n_md' Hertz Linear


# display the formant tracks overlaid on spectrogram.
  Erase all
  Font size... 14
  display_from = 'n_b' - 0.15
  if ('display_from' < 0)
    display_from = 0
  endif
  display_until = 'n_e' + 0.15
  if ('display_until' > 'finishing_time')
    display_until = 'finishing_time'
  endif
  select 'soundID'
  To Spectrogram... 0.005 4000 0.002 20 Gaussian
  spectrogram = selected("Spectrogram")
  Viewport... 0 7 0 3.5
  Paint... 'display_from' 'display_until' 0 3250 100 yes 50 6 0 no
  select 'formant_aftertracking'
  Yellow
  Speckle... 'display_from' 'display_until' 3250 30 no
  Marks left every... 1 500 yes yes yes
  Viewport... 0 7 0 4
  select 'textgridID'
  Black
  Draw... 'display_from' 'display_until' no yes yes
  One mark bottom... 'n_md' yes yes yes
  rf1hzpt = round('f1hzpt')
  rf2hzpt = round('f2hzpt')
  Text top... no Tracker output -- F1: 'rf1hzpt' ***** F2: 'rf2hzpt'

# display the spectrum, with Ltas and LPC
  select 'sound_16khz'
  spectrum_begin = n_md - 0.015
  spectrum_end = n_md + 0.015
  Extract part...  'spectrum_begin' 'spectrum_end' Hanning 1 no
  Rename... 'filebare$'_slice
  sound_16khz_slice = selected("Sound")
  To Spectrum (fft)
  spectrum = selected("Spectrum")
  Viewport... 0 7 4 6.5
  Draw... 0 3250 0 80 yes
  To Ltas (1-to-1)
  ltas = selected("Ltas")
  Viewport... 0 7 4 6.5
  Draw... 0 3250 0 80 no bars
  Marks bottom every... 1 500 yes yes no
  Marks bottom every... 1 250 no no yes
```

```
  select 'sound_16khz'
  To LPC (autocorrelation)... 18 0.025 0.005 50
  lpc = selected("LPC")
  To Spectrum (slice)... 'n_md' 20 0 50
  Rename... LPC_'filebare$'
  spectrum_lpc = selected("Spectrum")
  select 'lpc'
  Remove
 select 'spectrum_lpc'
  Line width... 2
  Draw... 0 3250 0 80 no
  Line width... 1
  Text top... no Spectrum [30 ms], Ltas(1-to-1) [30 ms], LPC(autocorrelation), all three overlaid

  if ('play' = 1)
    select 'soundID'
    Extract part... 'display_from' 'display_until' Hanning 1 no
    Play
   Remove
  endif

fileappend 'directory$'\'outputfile$' 'file$' 'f1hzpt:1' 'f2hzpt:1' 'newline$'

echo Settings F1ref:'f1ref' *** F2ref:'f2ref' *** F3ref:'f3ref' *** F4ref:'f4ref' *** F5ref:'f5ref' *** Frequency cost:'freqcost' *** Bandwidth cost:'bwcost' *** Transition cost:'transcost'


  select 'spectrum_lpc'
  pause ok? [occurrence 'counter' of segment 'the_label$']

  select 'spectrum_lpc'
  plus 'spectrum'
  plus 'spectrum'
  plus 'ltas'
  plus 'spectrogram'
  plus 'formant_beforetracking'
  plus 'formant_aftertracking'
  plus 'sound_16khz'
  plus 'sound_16khz_slice'
  Remove
endproc
```

Appendix 4 Plots

A. **Reaction times in Experiment 2 across correct responses to real-word test items, per Accent.**


Accent
SBE
BA
Centred RT (ms)
1000
500
0
-500
-1000
Word
good
black
right
point
bed
true
real
got
place
case
child
take
ten
friend
rest
back
book
great
kind
must
big
week
gone
death
could
down
wrong
town
much
through
front
had
than
best
ground
group
room
hands
first
night
have
what
next
land
both
last
hard
hair
street
times
road
sense
most
door
cent
them
least
need
third
still
comes
state
whom
whose

B. Accuracy in Experiment 2 across real-word test items, per Accent.