**Open Library of Humanities**

# Matched-accent processing: Bulgarian-English bilinguals do not have a processing advantage with Bulgarian-accented English over native English speech

**Marie Dokovova,** Speech and Language Therapy, University of Strathclyde, Glasgow, Scotland, UK; Clinical Audiology Speech and Language (CASL) Research Centre, Queen Margaret University, Edinburgh, Scotland, UK, m.dokovova@gmail.com

**James M. Scobbie,** Clinical Audiology Speech and Language (CASL) Research Centre, Queen Margaret University, Edinburgh, Scotland, UK, jscobbie@qmu.ac.uk

**Robin Lickley,** Clinical Audiology Speech and Language (CASL) Research Centre, Queen Margaret University, Edinburgh, Scotland, UK, rlickley@qmu.ac.uk

The Interlanguage Speech Intelligibility Benefit (ISIB) hypothesis for Talkers suggests that there is a potential benefit when listening to one's second language when it is produced in the accent of one's first language (matched-accent processing). This study explores ISIB, considering listener proficiency. According to second language learning theories, the listener's second language proficiency determines the extent to which they rely on their first language phonetics, hence the magnitude of ISIB may be affected by listener proficiency. The accuracy and reaction times of Bulgarian-English bilinguals living in the UK were recorded in a lexical decision task. The English stimuli were produced by native English speakers and Bulgarian-English bilinguals. Listeners responded more slowly and less accurately to the matched-accent stimuli than the native English stimuli. In addition, they adapted their reaction times faster to new speakers with a native English accent compared to a Bulgarian accent. However, the listeners with the lowest English proficiency had no advantage in reaction times and accuracy for either accent. The results offer mixed support for ISIB for Talkers, and are consistent with second language learning theories according to which listeners rely less on their native language phonology when their proficiency in the second language has increased.

## 1. Background

Socially prestigious monolingual English that does not incorporate the phonetic influence of other languages is still considered the 'gold standard' against which learners' varieties are explicitly or implicitly compared, despite long-standing criticisms of such views (Cheng et al., 2021; Cook, 1999; Grosjean, 1989) and despite the fact that the majority of the population in the world is multilingual (Cenoz, 2013). This happens not just in everyday life (Roessel, Schoel, & Stahlberg, 2020), but also when prioritizing native instructors and accents when teaching and awarding certain language qualifications (Bulgarian Ministry of Education, 2019; Levis, Sonsaat, Link, & Barriuso, 2016; Seedhouse, Harris, Naeb, & Üstünel, 2014; Selvi, 2014). This gold standard is put to the test within bilingual communities, where it is often the case that people communicate with each other in their second language (L2) with an accent that reflects the phonology of their first language (L1). There is a lack of consensus in research on the relative advantage of incorporating L1 phonology in L2. Would a person listening to their L2 spoken with an accent that matches their own L1 process the incoming speech faster and more accurately than if the same L2 had been spoken by a native speaker of that language? For example, does a Bulgarian listening to Bulgarian-accented English find it easier than listening to a socially prestigious native English accent? The hypothesis that there is such a speech processing benefit for non-native listeners has come to be known as the Interlanguage Speech Intelligibility Benefit hypothesis (ISIB).

Most research has focused on bilinguals' long-term listening adaptation to L2 speech. The term 'long-term listening adaptation' refers to a listener's relative speed and accuracy of processing an accent at a single time point, and as a result of all the listener's past experiences (or lack thereof) with that accent. This has been investigated in the context of the Interlanguage Speech Intelligibility Benefit hypothesis (Bent & Bradlow, 2003) and the Perceptual Assimilation Model-L2 (Best & Tyler, 2007), among others. Relatively less attention has been paid to the moment-to-moment short-term adaptation to an L2 accent from the perspective of L2 listeners. Short-term adaptation refers to a listener's relative accuracy and speed of processing an accent, as measured over a sequence of time points, as a result of the input they are currently listening to. Short-term effects in research are important in their own right and also because it is often assumed that long-term adaptations are built on a multitude of short-term adaptations to frequently encountered input (e.g., Bundgaard-Nielsen, Best, & Tyler, 2011).

There is consensus across L2 phonological acquisition models, for example the Perceptual Assimilation Model-L2, the Speech Leaning Model, or the Automatic Selective Perception model, that adult L2 listeners draw parallels between the phonology of their L2 and their native L1 language, but that the nature of these parallels may change depending on the L2 listeners' usage of their respective languages (Best & Tyler, 2007; Flege, 1995; Flege & MacKay, 2004; Strange, 2011). The Interlanguage Speech Intelligibility Benefit hypothesis is underpinned by a presumed reliance on L1 phonetics. This study tests the Interlanguage Speech Intelligibility

Benefit hypothesis by investigating the effect of the listeners' proficiency in L2 on (a) reaction times and (b) the accuracy of identifying words, where the L2 words have been pronounced either with a matching L1 accent or, alternatively, by a native speaker of the L2.

The following two sections summarize prior (somewhat contradictory) findings about L2 listeners who share an L1 with the speaker. This type of listening situation will be referred to as 'matched-accent' processing, following from Bent and Bradlow's (2003) discussion of the matched Interlanguage Speech Intelligibility Benefit hypothesis.

## 1.1 Long-term adaptation

The Interlanguage Speech Intelligibility Benefit hypothesis (ISIB) predicts that non-native listeners would be equally or more accurate than native listeners when hearing their second language spoken by other non-native speakers (Bent & Bradlow, 2003). To illustrate, it would be expected that a non-native French listener would be equally as accurate as (or even more accurate than) a native French listener when hearing the voice of another non-native speaker of French. Hayes-Harb, Smith, Bent, and Bradlow (2008) extended the original ISIB and predicted that non-native listeners would find speech produced by non-native speakers more intelligible than monolingual speech that lacks phonetic influence from additional languages. In practice this would suggest that non-native French listeners would find non-native French speech more intelligible than native French speech. This phenomenon is called ISIB for Talkers (Hayes-Harb et al., 2008) and is investigated in this paper.

Studies that have investigated the ISIB hypothesis often operationalize intelligibility as the ability to correctly understand speech input. This ability is measured as a function of lexical accuracy. However, we want to consider a more refined view of intelligibility by also measuring speed of processing. A putative matched-accent benefit has been observed through either an accuracy or a speed benefit, but only for small samples of L2 listeners and usually but not exclusively in those with low L2 linguistic proficiency (e.g., Bent & Bradlow, 2003; Hayes-Harb et al., 2008; Ludwig & Mora, 2017; Munro, Derwing, & Morton, 2006; Pinet, Iverson, & Huckvale, 2011). However, the body of literature on the topic is small, and each study has used slightly different methods, producing variable outcomes. This leads to inconclusive support for ISIB.

One of the few studies that investigated if listeners' L2 proficiency would affect their speed of matched, native, and non-matched-accent processing is by Ludwig and Mora (2017). The participants included low- and high-proficiency Catalan learners of English, low-and high-proficiency German learners of English (all four from a foreign language acquisition setting), and a group of native speakers of English who were unfamiliar with both Catalan- and German-accented English. They all listened to Catalan- and German-accented English and native English productions. Low-proficiency L2 listeners had faster reaction times for L2 English compared

to native English productions and high proficiency listeners had no difference between these matched and native accent conditions. This is in line with ISIB for Talkers as mentioned above: Non-native listeners are expected to be more accurate with non-native accents in L2 than with native accents in L2 (Hayes-Harb et al., 2008). However, in Ludwig and Mora (2017) foreign accent reaction time advantage happened only when the L2 stimuli were matched for the listeners' L1. The listeners were, across the board, slower with non-native speech produced by non-native speakers with a different L1 to theirs (i.e., non-matched L2), compared to either a matched-accent L2 or the L2 in its native English accent. This suggests that matched-accent processing is distinct from other interlanguage phenomena.

In addition to the results of Ludwig and Mora (2017), listeners' lower experience with L2 has been associated with accuracy benefits on matched accent over native productions in L2 in Pinet et al. (2011) and Hayes-Harb et al. (2008). Pinet et al. (2011) reported that low proficiency L2 English listeners living in France more accurately repeated French-accented English than native English speech, when they heard the L2 English in different levels of background noise. This accuracy advantage for French-accented English was not, however, observed for French-English bilinguals living in the United Kingdom (UK). Hayes-Harb et al. (2008) also found that low proficiency English listeners with L1 Mandarin were more accurate at a forced choice word identification test when the stimuli were produced by low proficiency L2 English speakers with L1 Mandarin. Again, the matched-accent advantage was not observed with high proficiency L2 English listeners.

Contrary to the prediction of ISIB for Talkers, Imai, Walley, and Flege (2005) did not find that non-native listeners understand non-native accents more accurately, in a study of Spanish-English bilingual listeners. Instead, the bilinguals with high proficiency in L2 English recognized more words produced by native English speakers than those produced by Spanish-accented speakers. By comparison, the low English proficiency listeners recognized an equal number of words in both accents. The difference in accuracy with Spanish-accented stimuli was not directly tested between high- and low- proficiency listeners. However, the plots suggest that there is very little difference in their accuracy scores with Spanish-accented words.

With the exception of Ludwig and Mora (2017), the studies discussed so far have focused on listeners' accuracy. However, a number of studies focusing on reaction times in lexical decision tasks have not found evidence supporting ISIB for Talkers. Lagrou, Hartsuiker, and Duyck (2011) found that Dutch-English bilinguals tested in the Netherlands processed native English speech faster than Dutch-accented English in a lexical decision task. Weber, Betta, and McQueen (2014) report similar findings from a lexical decision task measuring Italian-English bilinguals' processing times of English with or without strong Italian accent. Unfortunately, these studies did not also investigate the effects of the listeners' proficiency.

It appears that listeners paying attention to a non-native language might process other matched-accent non-native speakers faster and more accurately than they can perceive native

speakers, but primarily when the listeners have low proficiency in their L2. In this study, the ISIB for Talkers hypothesis will be tested by measuring the reaction times and accuracy of Bulgarian L1 – English L2 listeners responding to Bulgarian-accented or native English speech, by focusing on the effect of listener L2 English proficiency. On the basis of the research discussed so far, it may be expected that Bulgarian-accented English may actually be a challenge for some Bulgarian-English bilinguals, especially those who have high English proficiency. The specifics of Bulgarian phonology may also contribute to that. The Bulgarian phonological inventory is smaller than the Standard British English one (Boyadziev & Tilkov, 1997; Ternes & Vladimirova-Buhtz, 1990). Hence many sounds in English, particularly vowels, would have two-to-one or even three-to-one correspondences with Bulgarian phonemes according to the Perceptual Assimilation Model for L2 (Best & Tyler, 2007). It is expected that this would result in greater levels of homophony in everyday speech and generate more chances for slower or incorrect processing (Edwards, Pexman, & Hudson, 2004; Rubenstein, Lewis, & Rubenstein, 1971).

As noted, at least some Bulgarian listeners could be expected to benefit from hearing Bulgarian-accented English, probably the less L2 proficient ones. Moreover, if there are communities learning L2 English, then non-native accent features could be reinforced. Bulgarian L1 – English L2 listeners are likely to have some formal training of English in Bulgaria, as English is currently the most commonly chosen foreign language option in Bulgarian schools (Georgieva, 2010). This type of exposure may have led learners to form their representations of English speech around the Bulgarian-accented model of their instructors, and their co-learners. Some listeners' prototypical phonetic categories in many-to-one correspondences may become centred around Bulgarian-like phonetic variation. For example, vowels like /ɜ/ (e.g., in "nurse"), /ʌ/ (e.g., in "nuts") and /ə/ (e.g., in "bonus") can become associated with the Bulgarian vowel /ɨ/ which often corresponds to the grapheme "ъ," given the exemplars the learner-listener is exposed to. Such listeners might be expected to experience slower processing when faced with less familiar but L2 native-like input for those categories, such as /nɜs/, /nʌts/, /ˈbəʊ.nəs/ because they would not contain the expected /ɨ/. The smaller the listeners' communicative experience with native English speech, the more likely it is that they would process Bulgarian-accented English faster and more accurately than native English speech. The topic of accent experience interacting with proficiency is further discussed in the following section.

Overall, the evidence presented in this section suggests that proficiency may play an important role for non-native listeners of a language who are exposed to matched-accented L2 speech.

## 1.2 Short-term adaptation

Listeners' accuracy and reaction times when processing different accents are not static. The research of Eger and Reinisch (2019) and Mitterer, Eger, and Reinisch (2020) suggests that frequent exposure to an accent like one's own, leads to strong long-term adjustment to and

preference for that accent. Hence, it can be argued that a multitude of short-term adaptations leads to overall long-term accuracy and reaction time advantages with specific accents (cf. Sumner & Samuel, 2009). For example, studies investigating adaptation to individual speakers have found that listeners respond faster and more accurately when they consistently hear the same speaker over time than when responding to changing voices (Choi, Hu, & Perrachione, 2018; Mullennix, Pisoni, & Martin, 1989). This suggests that when listeners encounter an unfamiliar speaker, they may need some time to adapt, slowing their processing time, even if they are listening to familiar accents with unambiguous phonemic realizations (Choi et al., 2018). However, there are no studies known to us that have investigated the effect of matched-ISIB for Talkers on token-to-token reaction time adaptation to a new speaker.

Unsurprisingly, most of the research on short-term speaker adaptation has focused on the experiences of people listening to their native language (cf. Floccia, Butler, Goslin, & Ellis, 2009; Porretta, Tucker, & Järvikivi, 2016; Witteman, Weber, & McQueen, 2013). However, it has been found that when native listeners have familiarity with a foreign accent, they adapt to the familiar accent faster: Witteman et al. (2013) investigated foreign accent adaptation by measuring cross-modal priming in response to different foreign accent strength in the first and second half of a lexical decision task. In their experiment, cross-modal priming involved presenting a prime in one modality (auditory), immediately followed by a target in another modality (written), and it was the latter to which the participant had to respond. The auditory primes were German-accented, and were either real Dutch words or non-words. The written targets were either the same as the prime or not. The participants, native Dutch speakers, had to decide if the written target was a real word or non-word, and their response times were measured. It was assumed that if the target was identical to the prime, the listeners would respond faster than if the target and prime were unrelated. The priming effect was measured as the response times of the identical pairs subtracted from the unrelated pairs. If the response times for identical pairs were the same as the response times for unrelated prime-target pairs, then there was no priming effect, and it was interpreted that the accent of the prime prevented the participants from recognizing the identical written target fast enough.

Listeners with extensive German-accent experience in Dutch had priming effects across all accent strengths, while inexperienced listeners had priming effects only when the accent was weak or medium. Exposing inexperienced listeners with strongly accented speech led to priming effects in the first half of the task, compared to a group of listeners with no training who only showed priming effects in the second half. This study suggests that both long-term experience with a strong accent and recent exposure to it can result in a priming effect in the first half of the task. However, even inexperienced listeners with no recent exposure to strongly German-accented Dutch can adapt within the duration of the experiment, achieving a priming effect in the second half of the task. This suggests that native listeners hearing an unfamiliar foreign

accent can adapt to it through the exposure they receive within one experiment (cf., Clarke & Garrett, 2004).

These results are challenged by Floccia et al. (2009). They investigated the effect of accent familiarity on listener adaptation by focusing on native accents and comparing them to a non-native accent. They found that native listeners speed up, or adapt, their reaction times within fewer response trials for familiar accents compared to unfamiliar accents in a lexical decision task. Their reaction times for non-native accents, however, did not reach the adaptation observed with familiar and unfamiliar native accents. The authors suggested that this could be caused by different processing mechanisms being involved when adapting to non-native speech. A similar detailed look at reaction time adaptation differences has not been attempted in a matched-accent and native accent L2 setting, and so this is one purpose of the present study. The following review suggests that short-term adaptation to features of matched-accent L2 speech is possible, and it highlights potential facilitating conditions based on L1 adaptation studies.

The only study known to us that investigates adaptation within matched-accent L2 processing is from Reinisch, Weber, and Mitterer (2013). They presented Dutch-English bilingual participants with either Dutch-accented English stimuli or native Dutch stimuli in a lexical decision task. In each language group, half of the listeners heard /s/ replaced with an ambiguous /s-f/ sound and the other half of the listeners heard /f/ replaced with an ambiguous /s-f/ sound. After this exposure, the listeners heard the same Dutch speaker pronounce minimal pairs in Dutch that they had to categorize. Both language groups adjusted their phonetic boundary to the target sound to include the ambiguous realization, suggesting that phonetic speaker adaptation happens even within L2 English speech with a strong Dutch accent. This study suggests that listeners can adapt to a novel phonetic realization within matched-accented L2.

One factor that can facilitate listeners' adaptation to an accent is exposure to a large variety of speakers from that accent. This is particularly relevant for the topic of matched-accent processing, as some L2 listeners might have different amounts of exposure to other L2 users who have the same non-native accent. When living in an L2 majority environment, for example, non-native listeners would likely have exposure to a greater diversity of native speakers of the L2 than matched-accent L2 speakers, or even L1 speakers. This suggests that expats in an anglophone country would be expected to have long-term adaptations to native English speakers than to matched-accent L2 English speakers and by extension they would adapt faster to new native English speakers than matched-accented speakers.

The specific experience of emigrants with phonetic adaptation was investigated by Bruggeman and Cutler (2020). Using a similar experimental paradigm to Reinisch et al. (2013), they tested the ability of Dutch emigrants in Australia to adapt either to a Dutch speaker's ambiguous /s-f/ fricatives, within their L1, or to an Australian English speaker's fricatives, within their L2. The auditory test materials included native speakers of each language. Their results suggest that the

Dutch emigrants adapted to fricative variation within the English speaker, but not within the Dutch speaker. The authors systematically excluded several alternative explanations, such as the Dutch test materials, which had been effective in a pilot experiment; the participants' proficiency in L1 Dutch, which was higher than L2 English; and their amount of lifetime exposure to Dutch, which was equivalent to the pilot participants'. It was argued that the Dutch emigrants were not able to achieve short-term adaptation to the L1 speaker, because in their daily life they used Dutch only with family members, providing little exposure to novel Dutch speakers.

The need for exposure to a diversity of voices to achieve phonetic adaptation to novel speakers has been supported by several other studies. Some studies such as Bradlow and Bent (2008) and Sidaras, Alexander, and Nygaard (2009) found that a training phase including voices of several speakers leads to an accent-general, not just speaker-specific adaptation. For example, Sidaras et al. (2009) found that native English listeners transcribed Spanish-accented speech from unfamiliar speakers more accurately after exposure to a group of Spanish-accented speakers, specifically improving their vowel identification. Such improvement was not observed for listeners who received no training, or training with native English speech. The benefits of exposure to multiple foreign-accented speakers over exposure to a single speaker are also demonstrated by Baese-Berk, Bradlow, and Wright (2013). The listeners who were trained with multiple accents improved their scores for novel speakers regardless of whether they were included in the training sample. This study suggests that experience with a greater phonetic variability can improve perception of novel foreign accents.

Another factor that can facilitate speaker adaptation is feedback about the accuracy of their perception. The participants in Kriengwatana, Terry, Chládková, and Escudero's (2016) first experiment were unable to adjust to a novel accent whether they had experience with the target Dutch and Flemish accents or not. However, in a follow-up experiment the inexperienced listeners were given feedback about their correctness, which led them to successfully adapt their accuracy when processing the accent of a new speaker with an unfamiliar Dutch accent.

In a real communicative situation, feedback on correct interpretation can be derived from contextual cues. It can be speculated than L2 listeners with higher proficiency would have access to more linguistic feedback (e.g., by being familiar with collocations and having larger vocabularies) than listeners with low L2 proficiency, who might need to rely more on metalinguistic feedback of correctness (e.g., facial expressions or body language). In addition, Eger and Reinisch (2019) show that high proficiency L2 listeners in English make more use of acoustic markers of phonemic contrast than low proficiency L2 listeners. Hence, in an everyday situation it might be expected that the former would adapt to novel speakers faster and more efficiently than the latter.

To summarize, short-term adaptation to the phonetic nuance in an L2 accent is possible for L2 listeners. Adaptation can benefit from experience with multiple novel speakers and from top-down

feedback on speech processing correctness. However, existing research is overwhelmingly based on native listeners of a language, and more information is needed on the time-course of L2 listeners' adaptation to matched-accent L2 speech. According to Floccia et al. (2009), there is a fundamental difference in how native listeners of a language adapt to native and non-native accents. They argue that the difference in adaptation patterns is caused by a higher level of phonetic unpredictability found in the non-native accents. However, for non-native listeners, particularly those with low proficiency, a matched accent would be phonetically closer to their native language, and thus it could potentially be more predictable than the less familiar native accent. Hence, a reverse effect might be observed, whereby L2 listeners could adapt faster to matched non-native accents than to native accents in their L2. If there were a matched-accent reaction time adaptation benefit, that would provide support for ISIB. If there were no matched-accent benefit, that would constitute support for Floccia et al.'s (2009) argument that L2 accents have inherently more phonetic unpredictability even to listeners who might be familiar with some of the matched-accent phonetic characteristics. However, there are no studies which have investigated this question. The goal of this study is to test ISIB in just such an accent adaptation context.

## 2. Research questions and hypotheses

There are three key elements that require further investigation within the topic of matched-accent processing. Firstly, the literature review above highlighted L2 proficiency as a potentially important factor for matched-accent benefit or lack thereof. This study will focus on the listeners' proficiency in more detail by operationalizing it as a continuous variable. Secondly, ISIB for Talkers has not been extensively investigated from the perspective of the speed of processing a matched accent. This study will investigate matched-accent processing by comparing both the response accuracy and reaction times of bilinguals in a lexical decision task. Our task is similar to the procedure used by Lagrou et al. (2011) and is chosen with the aim of focusing exclusively on the effect of the word's overall accent, without any lexical or semantic context and without a carrier sentence, which can additionally facilitate accent adaptation.

Lastly, there is no research investigating real-time reaction time adaptation to novel speakers with a matched accent compared to speakers with native accent in L2, so this will be investigated by comparing the token-to-token changes in reaction times in response to speakers with the two accents. A decreasing trend of reaction times relative to the reaction times in the initial trial within a block will be interpreted as evidence of adaptation to the speaker. These research aims will be pursued by studying the matched-accent processing of Bulgarian L1 – English L2 bilinguals who reside in the UK.

It must be recognized that the strength of the Bulgarian accent of the stimuli is likely to play a role in the results, as studies have shown that the strength of foreign accent affects how quickly

native listeners adapt their reaction times (Porretta et al., 2016; Witteman et al., 2013). Previous research has found that low proficiency listeners have both greater (Hayes-Harb et al., 2008) and lower accuracy (Eger & Reinisch, 2019) when processing the speech of low proficiency speakers compared to speakers with a higher proficiency in L2. This study keeps the strength of the matched accent constant, using the speech of high proficiency English speakers with a mild Bulgarian accent (details are presented in the following section).

First, according to ISIB for Talkers (Hayes-Harb et al., 2008) we expect the Bulgarian L1 – English L2 listeners with the lowest English proficiency to process matched-accent speech faster and more accurately than native English speech. Second, based on the results of Imai et al. (2005), Pinet et al. (2011), and Lagrou et al. (2011) we expect that the bilinguals with the highest English proficiency will process matched-accent speech more slowly and less accurately than native English speech. The listeners with intermediate scores will fall between these extremes and might not have systematic difference in their reaction times and accuracy when processing either accent (cf., Ludwig & Mora, 2017).

In addition, based on the predictions listed above, listeners with the lowest L2 proficiency are expected to speed up their reaction times earlier in response to a new speaker with a matched accent than with a native English accent. Listeners with the highest L2 proficiency are expected to speed up their reaction times earlier in response to a new speaker with a native English accent than with a matched accent. In practice this would mean that their reaction times will speed up at different rates in the initial trials of a new speaker block within the lexical decision task when responding to either Bulgarian-accented or native English speakers (cf., Floccia et al., 2009).

## 3. Methods

### 3.1 Overview

This study examines the effects of second language proficiency on bilingual listeners' speed and accuracy when perceptually processing a 'matched accent' of a second language (Bulgarian-accented English in this case) versus natively-spoken second language (native British English). The experimental tool was a timed auditory lexical decision task. Participants were categorized using an English proficiency test and further information about them was gathered via a questionnaire. The experiment was carried out online and participants were recruited via social media.

The experiment was carried out in accordance with the ethics guidelines of Queen Margaret University and the ethics application was approved by Professor Janet M. Beck, head of the Division of Speech and Hearing Sciences, on behalf of the Ethics Committee on 29 July 2018. The data and R codes used for hypothesis testing are available via the Open Science Framework (OSF) repository (Dokovova, Scobbie, & Lickley, 2021).

## 3.2 Participants

The call for participants advertised for people who considered themselves residents of the UK and who had been born and raised as Bulgarian speakers with at least a primary level of schooling in Bulgaria. We did not specifically exclude speakers who had spoken an additional language to Bulgarian in their home environment during childhood. People were invited to take part if they were comfortable reading the information about the experiment and the consent form in English, hence the recruitment process acted informally to screen for a minimal requirement of functional ability in English.

The data of 94 participants were used for analysis, out of a total of 129 participants who started the experiment. Sixteen participants never entered a response for the auditory task and nine did not complete the proficiency test. A further six participants were excluded from the analysis due to an error in the recording of their data. Two participants listed their age of arrival in the UK as less than 10 years, so they were also excluded. One participant was excluded because they had entered only one correct response. One participant had no correct responses within 2.5 seconds in the native English condition, and so they were also excluded. The participants' mean age was 30.3 ($SD = 9$). Of them 62 identified as female, 30 as male, and two chose 'other.' Their English proficiency was measured using LexTALE (Lemhöfer & Broersma, 2012), which gives a score between 0 and 100% using an accuracy formula that weighs correct and incorrect answers to an untimed written lexical decision task. The participants' English proficiency scores ranged between 45 and 100 with an average of 80 ($SD = 13.2$). Their mean age of arrival in the UK was 24 ($SD = 8$).

The following paragraphs explain how the participants' English accent exposure score and Bulgarian accent exposure score were derived, based on a method used in Porretta et al. (2016). The data of all 94 included participants were used for these calculations. One of the questions asked the participants what percentage of their time was spent talking to native English speakers on a weekly basis. The next question asked them what percentage of that time they 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, although it is recognized that Standard British English speakers also live in other countries. As a result, this is a conservative way to estimate experience with the given variety, and further research is needed to validate its reliability.

The answers to these questions were each divided by 100, then multiplied together, and then multiplied by 100. On average the listeners reported spending 62% of their time talking with native English speakers, and 58.4% of that time talking to speakers from England, meaning that on average they spent 39% ($SD = 32.5$) of their time within a week talking to native English speakers from England ($0.62 \times 0.584 \times 100$). For convenience this is referred to as English

accent exposure score. Similarly, the participants were asked to estimate what percentage of their weekly time they spent talking to non-native speakers of English (35.8%) and what percentage of that time they spent talking in English to Bulgarian speakers of English (14.4%). These two numbers were also each divided by 100, then multiplied, and that result was multiplied by 100. The result was an average of 6.5% of their weekly time spent talking to Bulgarian-accented English speakers ($SD = 12$). This variable is called Bulgarian accent exposure score.

Two Pearson correlations were calculated: one between the English accent exposure score and Age and one between the Bulgarian accent exposure score and Age. There was a positive correlation between Age and the English accent exposure score ($t = 3.3$, $r$ (90) $= 0.33$, $p = 0.001$). There was a negative correlation between Age and the Bulgarian accent exposure score ($t = -2.46$, $r$ (89) $= -0.25$, $p = 0.016$). These results can be taken as an indication that the older participants were more integrated in UK society and spoke to more native English speakers from England and to fewer Bulgarians in a mixed context, while the younger speakers spoke to more Bulgarians in English and to fewer native English speakers from England. There were no significant correlations between reported weekly exposure to Bulgarian accent ($t = -0.40$, $r$ (89) $= -0.04$, $p = 0.69$) or native English speech ($t = -0.99$, $r$ (90) $= -0.1$, $p = 0.32$) and the listeners' proficiency in English.

A limitation of this study is that the participants were not asked whether they had spoken additional languages to Bulgarian at home when growing up. However, the call for participants emphasized the criteria of early home use and schooling in Bulgarian language, which should have selected participants for whom Bulgarian was overall the dominant language in childhood, suggesting early exposure to Bulgarian phonetics. Another potential limitation is that the participants were not asked to report their current usage of Bulgarian. As demonstrated by Flege and MacKay (2004), a listeners' level of use of their L1 may affect their discrimination of L2 sounds. However, within the continuity of ISIB research, L2 proficiency has typically been the variable in focus instead of L1 use, and there is indication that L1 use is negatively correlated with L2 proficiency (Luk & Bialystok, 2013; Wilden & Porsch, 2020). To keep the focus of this study consistent with previous research on ISIB and avoid collinearity between the predictors, this study only focused on L2 proficiency.

## 3.3 Materials

The final stimuli for this experiment were 64 real monosyllabic English words and 64 monosyllabic non-words (which were phonotactically plausible for English and comparable to the real word list, see below). A full list of the real words can be found in the Appendix. Subsections 3.3 to 3.6 report how an initial list of 100 words was selected, recorded by Bulgarian-accented and native English speakers, rated by native English listeners for strength of foreign accent, then finally narrowed down to a final experimental set of materials, which

included 64 real words. It was planned from the start to only include a subset of the initial 100-word list. The initial 100 monosyllabic words were chosen from the webCELEX database (Baayen, Piepenbrock, & Gulikers, 1995; Max Planck Institute for Psycholinguistics, 2001). This number was chosen based on a comparable experiment by Lagrou et al. (2011), who included 88 target stimuli. All real words in this initial set had a frequency over 3500 (of total corpus size 17.9 million) and were not cognates with Bulgarian words. High frequency words were chosen to minimize possible interaction effects of word-frequency and listener proficiency on reaction times.

The 100 words were recorded by Bulgarian-accented and native English speakers. In order to focus on the items most representative of this accent difference, a number of steps were followed. The most important one was to gather Foreign Accentedness ratings from native English listeners in an online task (see below). First, however, it should be noted that, due to the high phonological neighbourhood size, it was possible that some words, when pronounced with a Bulgarian accent, might sound like an unintended lexeme. Such risk was identified for words containing stressed /a/, 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 high-frequency real words. Upon auditory inspection of the actual stimuli recorded by the Bulgarian-accented speakers, we judged that there was low risk of this type of misinterpretation, as the speakers produced the difficult phonemes unambiguously. If the Bulgarian-English bilingual listeners' mental representations of these phones were not sufficiently distinct from their Bulgarian equivalents, then even native-like productions could have led to lexical access to more than one real word. As these words would act as homophones in English, this could potentially have an inhibitory or facilitatory effect on their reaction times in a lexical decision task in L2 regardless of their proficiency (Broersma, 2012; Nakai, Lindsay, & Ota, 2015). However, as there was no reason to expect that any potential effects would be different in each accent, it was decided not to remove the words. The average phonological neighbourhood size of the final list of 64 words was 21.31 (*SD* = 11.71) (Marian, Bartolotti, Chabal, & Shook, 2012).

We tested the results of this decision statistically by removing the potentially problematic words from the reaction time and adaptation models and can confirm that there were no changes in the significant predictors. Removing the words from the accuracy model led to convergence problems, which were only resolved by removing the random slope per word (see Section 3.9 for an explanation of random slopes). When the model converged there were also no differences in the significances of the predictors. Therefore, the results reported below are based on responses to all 64 words selected on the basis of accentedness, as described below.

The non-words were drawn from an initial list of 100 monosyllabic tokens matched for phoneme number, as far as possible, to the real words. They were chosen from the ARC non-word

database (Rastle, Harrington, & Coltheart, 2002) with a specification of only including lexically legal bigrams. The aim was to encourage the participants to wait until the end of the auditory stimulus before deciding if the stimulus was a word or a non-word. If the stimuli had contained phonotactically illegal onsets like /tl/, the listeners would have likely disambiguated such stimuli as non-words before hearing the nucleus and coda, and done so using phonotactic rather than purely lexical knowledge. From this initial list of 100, some non-words were removed, if, for example, we considered that they might be perceived as real words in Bulgarian, or more problematically, as Bulgarian-accented English real words. For example, the non-word [sɪf] was excluded for both reasons. It might be confused with 'sieve' due to final devoicing, which is a typical feature of Bulgarian pronunciation, and it also means 'grey' in Bulgarian. Using these factors, the non-word list was also narrowed down to 64 tokens to match the size of the real word set. The final phonological neighbourhood size of the 64 non-words was 3.93 ($SD =$ 2.82).

## 3.4 Recordings

Four female speakers had been asked to produce the 100 stimulus words and 100 non-words for the experiment. They were two monolingual native speakers of Standard British English and two native Bulgarian speakers. Both Bulgarian-English bilinguals were raised as monolingual Bulgarian speakers. During their teenage years, they had learned the same Standard British English target variety of English as the monolingual speakers and used it regularly in their professional lives (scoring 87.5 and 90 in LexTALE). All four had completed university degrees and were working in universities at the time of recording.

All recordings took place in sound-attenuated recording studios (at Queen Margaret University, Edinburgh, or in Varna, Bulgaria). The same equipment was used for all recordings, made at a sampling rate of 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 twice from a list with randomized word orders, followed by randomized non-words, while seated. They were instructed to pronounce the words in a natural, everyday manner, without over-enunciation.

## 3.5 Pretest

To help select the best individual stimuli, and also to ensure a consistent group difference between the Bulgarian-accented stimuli and the native English stimuli, the initial 100 real words were rated by native English listeners from the UK. The goal was for the final experiment to include stimuli consistently judged to have a detectable foreign accent when pronounced by the Bulgarian-accented speakers and with the least strong perceived foreign accent when pronounced by the native English speakers.

Forty-three native speakers of English (27 female, mean age = 38.02, *SD* = 13.36) from the UK were recruited via advertisements on Twitter and Facebook, in return for an optional entry in a draw for a £25 voucher for online shopping. Four wordlists were created from the 100 real words. Each wordlist contained 25 non-repeating words from each of the four speakers (100 items in total). The PsyToolkit platform (Stoet, 2010, 2017) was used to present the stimuli. A rating scale allowed the participants to rate the strength of foreign accent they perceived, and the software recorded their responses.

Each rater had one of the wordlists randomly assigned to them. The words were presented in a random order over their own computer's audio system, via a web browser. The orthographic form of the target word was also presented on the screen. Participants were asked to listen to each word no more than twice. They were asked to rate the strength of the (undefined) foreign accent they perceived in each word on a scale of 0–8 (none to very strong), using the web page interface. After the rating was completed, the raters filled in a debriefing questionnaire, asking for general demographic information, what they thought the identity of the accents was, their background in Bulgarian, their own variety of English and their frequency of interaction with non-native speakers of English and Bulgarians in particular.

The questionnaire showed that all raters had grown up speaking English in the UK and none of them had studied Bulgarian or had a Bulgarian background. On average the raters 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, leading to an average 0.45 Bulgarian accent exposure score, or 0.45% of their weekly time spent talking to non-native English speakers 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. Only one suggested that they had heard a Bulgarian accent.

A linear mixed effects model was constructed to test whether the foreign accent ratings differed between the speakers with different accents. The outcome variable was Foreign Accentedness score. 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. The model included random intercepts for Rater and by-Speaker random slopes for Rater. This accounts for the fact that each rater may have had a different pattern of rating and that this pattern may have differed for each speaker. The model also had random intercepts for Word and by-Speaker random slopes for Word. This accounts for the fact that each word may have contributed to a slightly different accentedness rating and that this may have differed depending on which speaker produced it. The results of this model are summarized in **Table 1**.

| 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 1:** 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 accentedness, although the magnitude of the difference was small ($b = -0.28$), compared to the differences with the Bulgarian-accented speakers. The two Bulgarian-accented speakers were both rated as having a stronger foreign accent than the reference native English speaker, and to a similar extent. **Figure 1** summarizes the model estimates and standard errors, which do not overlap between the native English stimuli and the Bulgarian-accented stimuli. The accentedness scores for the two native speakers of English are within the typical range for foreign accentedness ratings, as observed in other studies. Porretta et al. (2016) mention that native English speakers from the Wildcat corpus were rated by other native English speakers with scores up to 4 in a similar nine-point scale, where 1 was "no foreign accent" (Van Engen et al., 2010). Hence the significant difference between the two native speakers was considered to be part of the natural variation that listeners might encounter in their everyday life even in native English.

These native English listeners were not asked to transcribe the stimuli and we did not attempt to measure their intelligibility. As pointed out in the review of ISIB research above, native English listeners might find a non-native accent less intelligible than listeners who have a matched accent. Since the aim here was to focus on non-native listeners, we decided that the intelligibility judgements of native English listeners were beyond the scope of this study and could be left to future research.

## 3.6 Final selection

With the aim of increasing the overall difference between the two accent groups, we included only a subset of items from the initial selection of 100 stimuli from webCELEX. Words for each accent category (native English versus Bulgarian accent) were picked based on the Foreign accentedness scores. For each of the 100 words, two average scores were calculated: Mean Foreign accentedness scores of the two native English speakers and mean Foreign accentedness scores of the two Bulgarian-English bilinguals. The difference between the two Foreign accentedness scores was then calculated for each of the 100 words. As the average difference between the En1 speaker and the two Bulgarian speakers rounds to four (see **Table 1**), it was decided that only

individual words with a Foreign accentedness score difference of at least four would be selected for the listening experiment. There were 64 such words exceeding the threshold, and so these were selected for the experiment (see the Appendix).
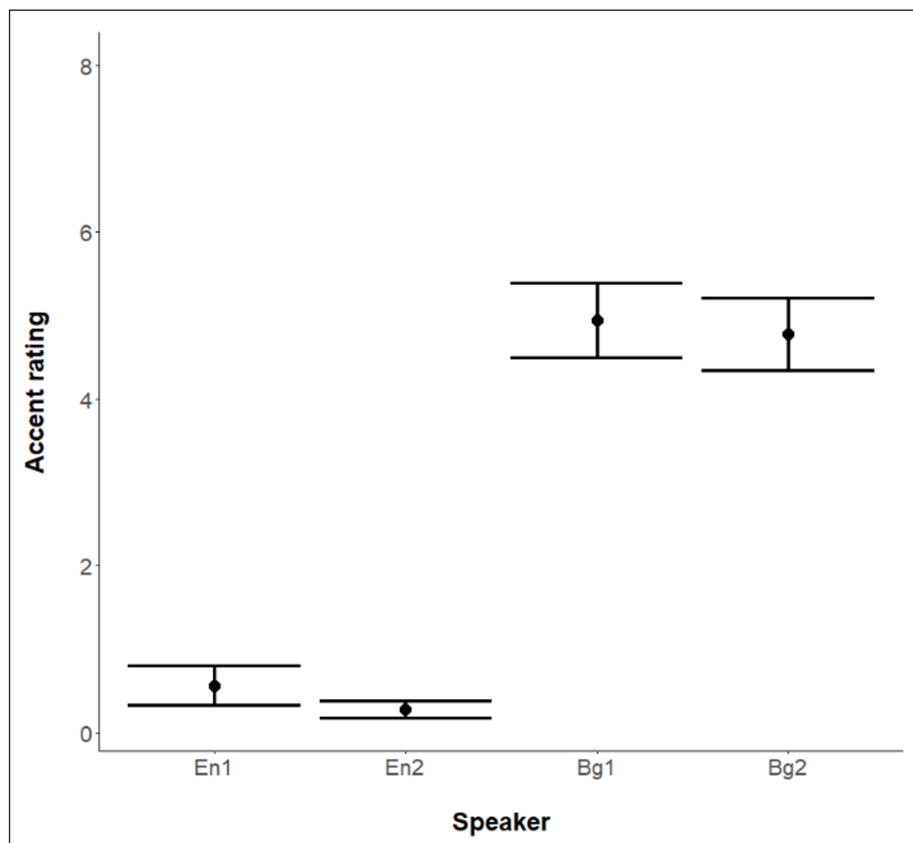


**Figure 1:** The modelled estimate and standard error of the foreign accent rating of the two native English speakers (En) and the two Bulgarian L1 – English L2 speakers (Bg).

### 3.7 Design

The main experiment was a lexical decision task with a within-subject design. As noted above, the stimuli were recorded by four speakers, two with Bulgarian-accented English and two who were native English speakers. All participants heard words and non-words produced by all four speakers. To restrict the length of the experiment and to avoid exposing the listeners to the test words more than once (which could affect their reaction times), each individual 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 four versions of the experiment in which each quarter of the stimuli was produced by a different speaker. Each listener heard each of the words and non-words only once.

The stimuli were presented in four blocks. They were blocked by speaker within accent, to allow the listeners to adjust to each voice and thus to avoid affecting the reaction times due to random changes in the speaker's 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 different random order to each listener. Before the first block the listeners heard ten training trials with three non-words and seven real words. A summary of the structure of the whole experiment is available in **Figure 2**.

The speaker for the training stimuli was the first author, a native Bulgarian speaker with 5 years of experience living in the UK at the time of the recording. Half of the training stimuli were produced with a Bulgarian accent and half with received pronunciation. The training stimuli were a subset from a previous pilot experiment, in which the Bulgarian-accented stimuli were rated as significantly more foreign accented than the received pronunciation ones by ten native English listeners.
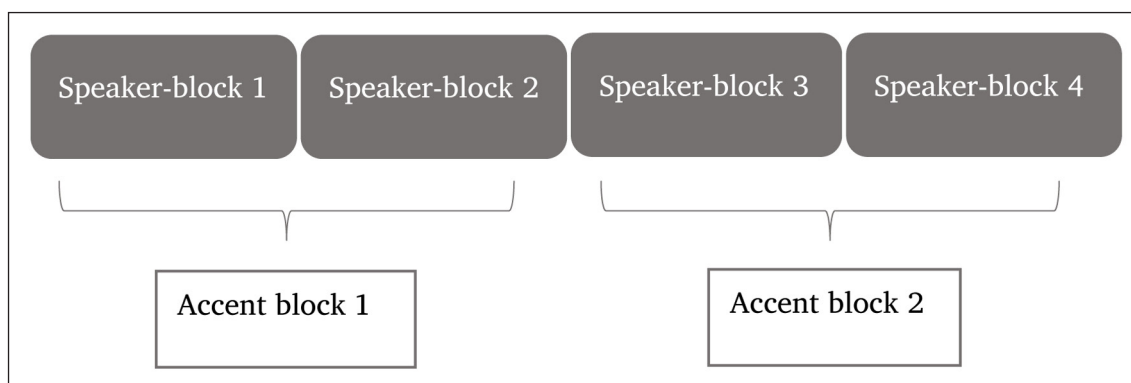


**Figure 2:** Structure of the main experiment.

## 3.8 Procedure

This subsection describes the procedure of the online experiment involving Bulgarian L1 – English L2 listeners. Participants (all living in the UK) 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 to reach as many participants as possible and increase the variability in the participants' proficiency scores. Prior pilot studies had proven that recruiting Bulgarian participants for in-person laboratory-based experiments in Edinburgh and Musselburgh was problematic. The whole experiment was carried out using the online platform PsyToolkit (Stoet, 2010, 2017) via a web-browser. The experiment was programmed to run only on computers with keyboards, so this excluded attempts to participate with tablets and mobile phones. Kim, Gabriel, and Gygax (2019) tested the validity of using PsyToolkit online over a

web browser compared to a lab-based E-Prime experiment for reaction time measurements and response choice in a psycholinguistic paradigm. They used similar sample sizes for both the online and the lab-based experiments and successfully replicated the findings from the lab experiment, demonstrating that online PsyToolkit is a valid method for both types of measurement. A recent large-scale study comparing online and lab-based experimental software also supports the validity of online experimental software (Bridges, Pitiot, MacAskill, & Peirce, 2020).

After providing informed consent the participants were given written instructions for the auditory lexical decision task. The instructions included a photo of a standard keyboard, which highlighted the keys that the participants needed to press if they wanted to select a 'word' or a 'non-word' answer. They proceeded at their own pace. After a countdown, the training trials for the lexical decision task started automatically.

The following procedure applied to the whole auditory lexical decision task. When making their lexical decisions, the participants had to respond by pressing either the '4' or the '6' key 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 the '5' key. These keys were picked because it was anticipated that there might be differences in the layout across the participants' keyboards. Keys '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 keys were randomized across participants. After hearing each auditory stimulus the participants had 2500 ms to enter their response, after which the following test item was automatically loaded. The reaction times were measured from the end of the sound file that contained the auditory stimulus. The sound files were trimmed to have no extraneous silence after the last phoneme's acoustic energy dropped away. As soon as a participant 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 (or lack of response) was recorded, and a new item was about to be played.

After the training task, in which the participants responded to ten trials, they proceeded with the main experiment, which started after a countdown. The participants heard the 128 trials of words and non-words without a break, albeit in four blocks (as described above). With a maximum delay for each answer set at 2.5 seconds, the whole task was expected to take up to five minutes. The auditory lexical decision task was followed by the proficiency test LexTALE (Lemhöfer & Broersma, 2012), which was also presented on the PsyToolkit platform. The participants saw a word or a 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. On average the whole study was completed in 18 minutes ($SD = 15.3$).

## 3.9 Analysis

The first question addressed in this experiment was whether the Bulgarian accent of the stimuli would facilitate the speed of recognition of real English words for Bulgarian L1 – English L2 bilinguals, particularly for participants with low English proficiency. Only correct real-word responses between 200 ms and 2000 ms were included, based on the similar study design, listeners, and analysis of Weber et al. (2014). The reaction times were not log transformed because they were normally distributed, as were the model residuals. An alternative log transformed analysis was performed and did not change the direction of the results, therefore the original analysis is described here. Non-words were not included in the analysis because the focus of the experiment was on the speed and accuracy of lexical access of frequent words, which would have been encountered with a variety of voices in the listeners' everyday lives. The main purpose for including non-words in the task was to create a foil, against which to elicit lexical access judgements, at speed, from the participants.

A linear mixed effects analysis was performed, to find the effect of the listeners' English proficiency and the stimuli's accent on the listeners' overall reaction times. The linear mixed effects regression analysis had three predictors: Proficiency (the LexTALE score centred around the mean 80), Accent (native English as a baseline, and Bulgarian accent) and their interaction. The outcome variable was Reaction times in ms centred around the mean 1225.3 ms. Centring was performed because it allowed for an easier interpretation of the coefficients of the model, since 0 to 200 ms were not meaningful outcomes in the dataset. The model had random intercepts of Participant and by-Accent random slopes for Participants. This means that the model accounts for the fact that each participant is likely to have a slightly different pattern of reaction times, and that this pattern could differ between different accents within a participant. The model also includes random intercepts of Word and random slopes of Word by Speaker. This means that the model accounts for the fact that each word may have elicited a different pattern of reaction times and that this pattern may have differed depending on which speaker pronounced it.

To investigate the second research question, about the effects of Proficiency and Accent on the Accuracy of word recognition, a binomial logistic mixed effect model was tested. The outcome variable included correct and incorrect answers to real word stimuli that received responses within 200 ms and 2000 ms. The model included the interaction between Proficiency (the LexTALE scores centred around the mean 80) and Accent (native English as baseline, compared to Bulgarian accent) as well as each of these predictors separately. In addition, the model had random intercepts of Participant and Speaker. This was to account for the fact that each participant could have had a slightly different pattern of accuracy. The model included by-Accent slope for Participant and a by-Speaker slope for Word. Each participant and word could have elicited a different pattern of accuracy, which could have also varied depending on the accent or speaker, respectively. The outcome variable was coded with zero (incorrect) and

one (correct), hence estimates in the positive direction would suggest an increased number of correct answers.

Lastly, this study addressed the question of whether the listeners would adapt their reaction times to a matched-accented speaker faster or slower than a native English speaker and whether their adaptation would be affected by the listeners' proficiency in English. Adaptation here is used to mean the change in reaction times over a period of exposure to stimuli. Adaptation is tested by observing the change in the listeners' reaction times with each subsequent stimulus they hear, depending on the accent of the stimuli.

As noted above, the stimuli were presented to the listeners blocked by accent and then by speaker. The change in reaction times over the subsequent stimuli was represented using curves formed of points with two coordinates: stimulus number within the speaker block on the x-axis and the reaction time on the y-axis. The curves were compared using a generalized additive mixed model (GAMM). 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 earlier, 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 it 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 described so far, this type of analysis also allows the use of random structures (here, random smooths) to account for the fact that multiple reaction time data-points came from the same participants and that multiple participants were presented with the same words. A random smooth therefore accounts for the effect of non-linear but systematic variation from the model. This model focuses on the non-linear relationship between the continuous predictors Proficiency and Within-block trial number and their interaction with the two Accents (native English and Bulgarian Accent).

Only correct responses between 200 ms and 2000 ms to real words were included in the analysis. The reaction times to the words were centred around their mean (1225.3 ms). The model included a parametric term for Accent (native English versus Bulgarian Accent), a smooth term for the token number Within-block (1 to 32 in a speaker 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).

Increasing the number of base functions ($k$) makes the model less conservative and may result in overfitting the data points. According to Sóskuthy (2017) and Wood (2021) it is recommended to run the model with different values of $k$ and to report any differences in the significances of variables. The model was tested with $k$ values from 5 to 25 (in multiples of 5) for each of the predictors. The model did not run with a $k$ value of 20 for any predictors except Proficiency, where the maximum value was tested separately and determined as $k = 27$. K values between 20 and 15 across all other predictors were tested to determine the maximum value with which the model can run. $K = 15$ was the maximum value possible for all predictors. The $p$ values remained the same within three decimal places across all iterations.

In addition, the model included random smooths for Within-block number per Trajectory where Trajectory was the adaptation trajectory for one participant for one speaker block, allowing individual variation at each trial number within a speaker block. There were also random smooths for Within-block number per Word, accounting for the fact that each word may have led to different pattern of reaction times, depending on its order within the block. There was also a random smooth for Within-block by Participant, allowing individual non-linear variation per participant at each trial number within a block.

# 4. Results

## 4.1 Overall reaction times, linear analysis

This section addresses the questions of whether the listeners' proficiency in English, and whether the speakers' accent (Bulgarian accent or native English), together or separately have an effect on the listeners' reaction times. The analysis includes 5291 observations, 64 words, and 94 participants. The detailed results are presented in **Table 2**. The listeners' English proficiency has overall no effect on their reaction times processing of native English, which is the baseline level of the Accent variable. However, there is a significant effect of Accent. This means that words with a native English accent are recognized faster than words with a Bulgarian accent for a baseline of listeners with an average proficiency (LexTALE = 80), with an estimate of 45 ms shorter processing time.

The significant interaction between Proficiency and Accent is more relevant for the research question. **Figure 3** shows that as the listeners' proficiency in English increases, the difference in reaction times between the two accents also increases, an increase differential which appears to be driven by changes in the processing times of both accents. In the lower end of the proficiency continuum there is an overlap in the confidence intervals and the smallest difference in modelled response times is observed. However, the 95% CI in **Figure 3** are based only on the fixed effects

and should be interpreted as indicative, as according to Bates, Maechler, Bolker, and Walker (2020) their estimation including the random effects is too unreliable and is therefore not included in the plotting package (Long, 2020).

| Predictor | Estimate | *t*-value | *p*-value |
|---|---|---|---|
| Intercept | 1.44 | 0.10 | 0.92 |
| Proficiency | 0.15 | 0.19 | 0.85 |
| Accent (Bulgarian) | 45.38 | 5.22 | <0.001 |
| Proficiency: Accent (Bulgarian) | 1.13 | 3.18 | 0.002 |

**Table 2:** Summary statistics for the linear analysis of the effects of Proficiency and Accent on the reaction times.
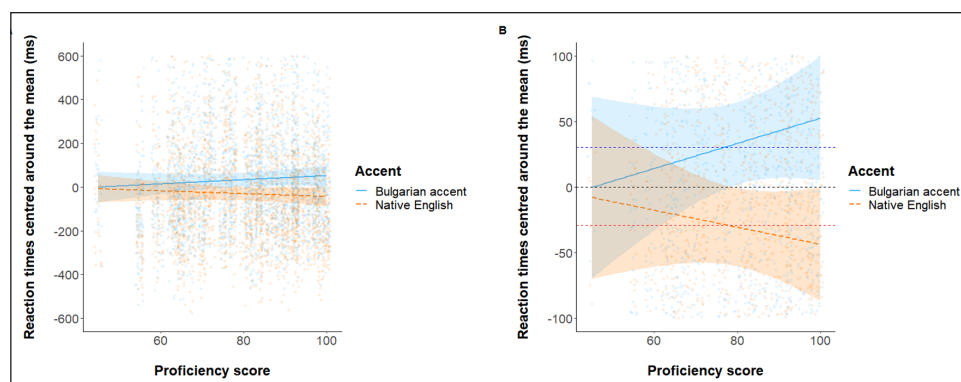


**Figure 3.A:** (left) Model prediction for the reaction times to Bulgarian- and English-accented words. The x-axis shows the proficiency scores, lowest to highest, and the y-axis shows the reaction times, centred around the overall mean 1225.3 ms. The shaded area represents the 95% confidence intervals. The raw datapoints are represented as a scatterplot. **3.B.** (right) As Figure 3.A, with y-axis showing only the range [–100, 100] ms. The blue horizontal dotted line represents mean reaction time to Bulgarian-accented words 1255.7 ms. The black horizontal dotted line represents the overall mean 1225.3 ms of both accents. The orange horizontal dotted line represents the mean reaction time to native English stimuli 1196.2 ms.

To better understand the relative effect of Proficiency on the listeners' response times for each accent, two follow-up analyses are performed. The data are separated by Accent, and the effect of Proficiency is tested on each accent subgroup separately. The first follow-up model focuses on the reaction times to Bulgarian-accented stimuli only, centred around their mean 1255.7 ms. The only predictor is the listeners' Proficiency, each score centred around the mean. The model includes random intercepts by Participant and Word and by-Speaker random slopes for Word. This means that the model accounts for the fact that different participants might have slightly

different patterns of reaction times, that different words might elicit slightly different patterns of reaction times and that words produced by different speakers can also elicit different patterns of reaction times. The model is based on 2588 observations, 64 words, and 94 participants, and the results are summarized in **Table 3**.

| Predictor | Estimate | *t*-value | *p*-value |
|-----------|----------|-----------|-----------|
| Intercept | 41.40 | 2.53 | 0.01 |
| Proficiency | 0.94 | 1.05 | 0.30 |

**Table 3:** Summary statistics for the linear analysis of the Proficiency on the Reaction times to Bulgarian-accented stimuli.

The effect of Proficiency on the reaction times to Bulgarian-accented words is not significant. The second follow-up model focuses on the reaction times to native English stimuli only, centred around the mean 1196.2 ms. The predictor is the listeners' Proficiency, centred around the mean. The model includes random intercepts by Participant and Word, and by-Speaker random slopes for Word. The model is based on 2703 observations and 94 participants. The results are summarized in **Table 4**. The effect of Proficiency on the reaction times to words with a native English accent is not significant.

| Predictor | Estimate | *t*-value | *p*-value |
|-----------|----------|-----------|-----------|
| Intercept | –27.74 | –1.78 | 0.08 |
| Proficiency | –0.59 | –0.74 | 0.46 |

**Table 4:** Summary statistics for the linear analysis of the effect of Proficiency on the Reaction times to native English stimuli.

The lack of Proficiency effect in each subset is surprising, considering the significant interaction between Accent and Proficiency in the pooled data. **Figure 3.B** helps illustrate how the result may have arisen. It illustrates the results from the pooled model, and the horizontal dotted lines represent (from top to bottom) the mean reaction times to Bulgarian accent only (blue), both accents combined (black), and native English only (orange). The blue and orange horizontal lines fall entirely within their respective confidence intervals and suggest that the effect of proficiency within each of the accents separately is possibly obscured by the great variability of reaction times across the proficiency continuum. Only when the two accents are combined is there sufficient power to detect an effect of proficiency between the two accents.

To summarize, no matched-accent benefit in reaction times is found, even for listeners towards the lower end of the proficiency continuum, providing no support for ISIB. Some of the expected

proficiency effects are observed, because there is a significant interaction between proficiency and accent. As the listeners' proficiency increases, their reaction times to Bulgarian-accented English increase relative to native English. This can be interpreted as evidence for matched-accent disadvantage. However, the two follow-up analyses reveal that when the dataset is split by accent, there is no effect of proficiency on the listeners' reaction times for either accent. This discrepancy could be the result of the high variability in reaction times combined with a small effect of proficiency, only manifesting when two accents are compared to each other.

## 4.2 Overall accuracy, linear analysis

This section addresses the question of whether Bulgarian-accented English facilitates the accuracy of word recognition compared to native English stimuli for Bulgarian L1 – English L2 bilinguals with different English proficiencies. There is an overall lower accuracy rate with Bulgarian-accented words (91%) than native English words (94%), although the accuracy of the participants is generally high. There are 5703 observations, 94 participants, and 64 words considered in the binomial logistic mixed effects model. The results are summarized in **Table 5**.

| Predictor | Estimate | *t*-value | *p*-value |
|---|---|---|---|
| Intercept | –3.46 | –18.78 | <0.001 |
| Proficiency | –0.004 | –1.73 | 0.58 |
| Accent (Bulgarian) | 0.60 | 3.14 | 0.002 |
| Proficiency : Accent (Bulgarian) | 0.02 | 2.54 | 0.01 |

**Table 5:** Summary of the overall binominal logistic mixed effects model on the listeners' Accuracy.

There is no significant effect for Proficiency for a baseline of native English stimuli. However, there is a significant effect of Accent, such that Bulgarian-accented words are recognized incorrectly more often than native English accented words for a baseline of listeners with average proficiency. Importantly, there is also a significant interaction between Proficiency and Accent, such that with increased proficiency there is decreased accuracy for Bulgarian-accented words.

**Figure 4** shows that listeners with higher proficiency in English have higher accuracy for native English words and lower accuracy for Bulgarian-accented words. At the bottom end of the proficiency continuum the accuracy of the two accents completely overlaps, similarly to the reaction times findings. As described in the previous section, no systematic advantage for Bulgarian-accented stimuli is observed and the 95% CI in **Figure 4** are only to be considered as rough estimations of the main effects, as they include only the fixed and not the random effects (Bates et al., 2020; Long, 2020).
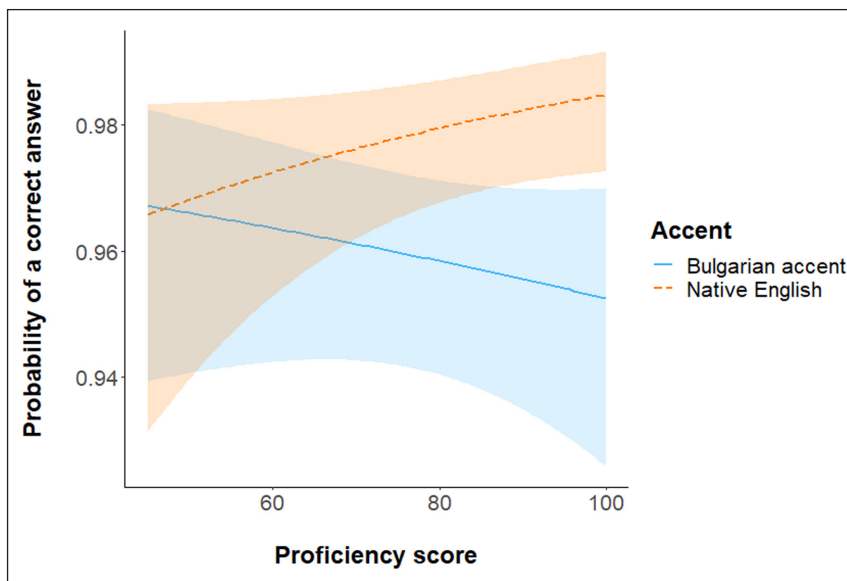
**Figure 4:** Modelled interaction between Proficiency and Accent on the outcome variable Accuracy. The x-axis shows the Proficiency score, lowest to highest. The shaded area reflects an estimation of the 95% CI for the main effects.

Two follow-up analyses are performed to fully interpret the results. The dataset is separated by Accent and the effect of Proficiency is investigated in each separate dataset, also including random intercepts of Participant and random slopes and intercepts of Word by Speaker. The model, focusing on Bulgarian-accented stimuli, is based on 2828 observations, 64 words, and 94 participants. The effect of Proficiency on the Accuracy of recognizing Bulgarian-accented words is not significant. The results are summarized in **Table 6**.

| Predictor | Estimate | $z$-value | $p$-value |
|---|---|---|---|
| Intercept | 3.15 | 17.62 | $<0.001$ |
| Proficiency | −0.007 | −1.00 | 0.32 |

**Table 6:** Summary of the generalized linear mixed model on the listeners' Accuracy for the subset of Bulgarian-accented stimuli.

The second model, focusing on the native English stimuli is based on 2875 observations, 64 words, and 94 participants. There is no significant effect of Proficiency on the Accuracy of recognizing native English words. The results are summarized in **Table 7**. The lack of proficiency effect in each accent subset echoes the results described in the previous section. Considering the wide confidence intervals in **Figure 4**, it is likely that the variability in the accuracy for each separate accent obscures any minimal effect of proficiency that is manifested when the two accents are directly compared.

| Predictor | Estimate | z-value | p-value |
|---|---|---|---|
| Intercept | 3.83 | 16.47 | <0.001 |
| Proficiency | 0.01 | 1.65 | 0.10 |

**Table 7:** Summary of the generalized linear mixed model on the listeners' Accuracy for the subset of native English stimuli.

Overall, there is no evidence of matched-accent benefit for accuracy, even for the listeners towards the lower end of the proficiency continuum, again showing no support for ISIB. There is an expected matched-accent accuracy disadvantage, relative to native accent accuracy, when the listeners' proficiency increases. Similar to the results for the reaction times, the proficiency effect is not observed when the data are split by accent. Again, this could be the result of the large variability in accuracy across listeners with different proficiency.

## 4.3 Reaction time adaptation within a block, curve analysis

This section investigates the effect of English Proficiency and stimulus Accent on the participants' short-term reaction time adaptation to new speakers with either of the two different accents. It was predicted that listeners with high English proficiency would adapt to a new speaker with a native English accent faster than to a new speaker with Bulgarian accent when they first heard the accents within the experiment. It was also predicted that low English proficiency listeners would adapt faster to a new speaker with a Bulgarian accent than to a new speaker with a native English accent. The results of the smooth terms, estimating the non-linear relationship between the predictors and the outcome variable, are summarized in **Table 8**.

| Smooth terms | Edf | F | p-value |
|---|---|---|---|
| Within-block | 1.00 | 0.42 | 0.51 |
| Proficiency | 2.86 | 2.02 | 0.10 |
| Within-block by Accent (Bulgarian) | 3.55 | 3.32 | 0.01 |
| Within-block by Proficiency | 2.95 | 1.31 | 0.22 |
| Within-block by Accent (Bulgarian) and Proficiency | 1.00 | 0.29 | 0.59 |
| Random smooth terms | Edf | F | p-value |
| Within-block per trajectory | 241.78 | 0.32 | <0.001 |
| Within-block per participant | 107.40 | 2.63 | <0.001 |
| Within-block per word | 62.37 | 4.41 | <0.001 |

**Table 8:** Summary statistics for the smooth and random terms of the full GAM model. Edf = estimated degrees of freedom.

There is no significant effect of the Within-block smooth. This means that the within-block trial number led to no systematic non-linear change in reaction times for listeners with average proficiency adapting to native English speakers. There is no significant effect of the Proficiency smooth. This means that there are no systematic non-linear differences in reaction times between listeners with different English proficiencies when responding to native English stimuli.

There is, however, a significant non-linear interaction between Within-block trial and Accent. This means that listeners change their reaction times differently for the two accents as the block progresses. A significant GAMM smooth effect suggests that there is a non-linear effect of the predictors on the outcome variable, but it does not specify the direction of change. This information is obtained by observing a plot. **Figure 5** suggests that reaction times change non-linearly over the duration of the block in response to Bulgarian-accented stimuli, and that the trend is decreasing until trials 20 to 25 when it begins to increase. This suggests that throughout the initial stage of the block the listeners adapt to the Bulgarian-accented speakers. It is less clear why towards the end of the block their reactions slow. This could be a plateau, or a temporary stage in a non-linear downward trend of reaction time adaptation, which would have manifested if the block were longer. Although the right panel of **Figure 5** suggests that the difference between the two accents is maintained throughout the whole speaker-block, the largest difference between the two accents is observed in the first five trials.

There is no significant interaction between Proficiency and Within-block trial, which means that people with different proficiencies have no systematic changes in their reaction times when responding to native English voices as the block progresses. Lastly, the most relevant interaction for the research questions is not significant. There is also no significant triple interaction between Proficiency, Accent, and Within-block number. This means that as the block progresses, listeners with different levels of English proficiency have no systematic differences in how they change their reaction times in response to the two accents.
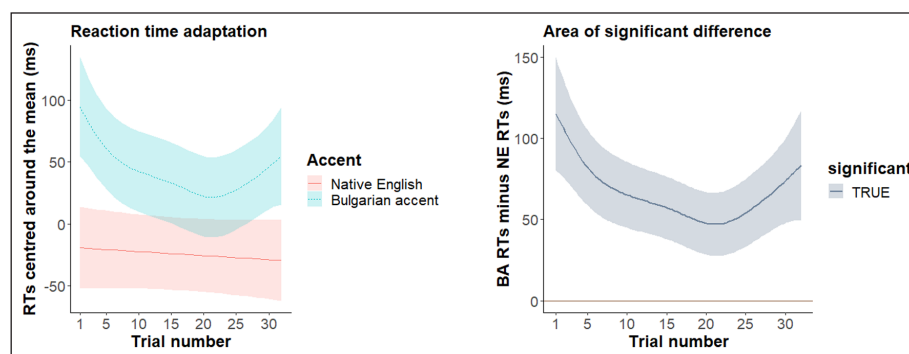


**Figure 5:** Left: y-axis shows GAM model predictions of reaction time adaptations (centred around the mean 1225.3 ms) to either of the new speakers with Native English or Bulgarian accent. Right: y-axis shows the difference between the two accent curves, with the area of significant difference highlighted in blue. Both x-axes show the trial numbers within a new speaker and a new accent block.

To summarize, Bulgarian-accented words are processed slower than native English words throughout the whole block, although the difference between the two accents gradually decreases as the block progresses. Contrary to the initial expectations, there are no significant non-linear interactions between proficiency, accent, and block trial number, suggesting that the listeners' proficiency in English does not systematically affect how they adapt to each of the accents. Similar to what was reported in the accuracy and reaction time analyses, there is no support for ISIB, and Bulgarian-accented words are slower to process than the native English words.

## 5. Discussion

The results of this study show mixed support for the Interlanguage Speech Intelligibility hypothesis for Talkers (Hayes-Harb et al., 2008). ISIB for Talkers predicts that non-native listeners of an L2 language would find non-native Talkers of the L2 more intelligible than native Talkers of the L2. Falling short of clear support, we observed effects of listener L2 proficiency on accuracy and reaction times, findings which offer support for the Perceptual Assimilation Model for L2 (Best & Tyler, 2007), a model which does however underpin ISIB for Talkers.

First, it was expected that the Bulgarian L1 – English L2 listeners towards the lowest end of the English proficiency continuum scores would process matched-accent speech faster and more accurately than native English speech. This was *not* supported by the results. Lower L2 proficiency listeners tended not to have systematic differences in reaction times and accuracy reflecting which accent of L2 they were hearing.

Second, it was expected that greater English L2 proficiency would lead to a perceptual advantage for the native-accented L2 stimuli over the matched (L1 Bulgarian-influenced) accent. This prediction was supported, as there was an overall effect of accent, suggesting that the majority of listeners processed native English stimuli faster and more accurately than Bulgarian-accented stimuli. There was also a small interaction with proficiency: The higher the English proficiency of the listeners, the slower and less accurate their responses to Bulgarian-accented stimuli were, compared to their responses to native English stimuli. However, this effect was present only when comparing the two accents directly and it was not present within each accent separately.

This latter point was a surprising finding, and may have been caused by a reduction of statistical power when the dataset was split by accent (Button et al., 2013). The wide confidence intervals of the model predictions in **Figures 3** and **4** suggest that the high variability of reaction times and accuracy for each accent separately may have obscured any small trend for a proficiency effect. The proficiency effect only becomes apparent when the confidence intervals of the two accents are compared against each other. They diverge as the listeners' L2 proficiency increases. This raises the question whether the interaction effects for reaction times and accuracy reported here are false positives, or whether they are real but small effects. To our knowledge this is the

first study to use proficiency as a continuous predictor when testing ISIB, hence more replications of this design are needed. According to Button et al. (2013), studies aiming to replicate studies with small effect sizes should collect larger datasets than the original.

Third, it was expected that there would be some listeners intermediate on the L2 English proficiency continuum who would have no systematic accent-based difference in their accuracy and reaction times. **Figures 3** and **4** suggest that this was not the case. The majority of listeners had a matched-accent disadvantage in reaction times and accuracy.

The rest of the research questions focused on token-to-token reaction time changes that might be expected when listeners encounter a new speaker, irrespective of whether the speaker has a Bulgarian or a native English accent. It was expected that there would be an interaction between listener L2 proficiency and speaker accent, such that the greater the listeners' proficiency, the greater their native English accent advantage would be. Specifically, we expected slower adaptation by proficient English L2 listeners to new speakers with a matched Bulgarian accent in English than to new speakers with a native English accent. Conversely, it was expected that the lower proficiency English L2 listeners would have a greater matched-accent benefit in reaction times and speed of adaptation compared to native accents. These predictions were not supported as there was no significant interaction between accent and proficiency. There *was* a significant effect of accent, however: Overall, the Bulgarian listeners adapted faster to native English speakers than Bulgarian-accented L2 speakers of English and maintained lower reaction times throughout the block.

Thanks to the use of proficiency as a continuous variable, the results of this study provide specific information about the mechanism by which matched-accent processing works, at least in a situation where the segmental phonetic material is its main driver. Matched-accent processing is less efficient and accurate than native accent processing in L2 for listeners with intermediate to high L2 proficiency (scoring over 45 on LexTALE), although the effects are small (on average 60 ms slower and 3% less accurate for Bulgarian-accented English). Two concrete predictions can therefore be made about where ISIB can be found, particularly in a population of emigrants in an L2-dominant country. First, ISIB would be more likely for adult learners with lower L2 proficiency (e.g., less than 45 on LexTALE). Listeners with weak L2 may be more likely to rely on similarities to their L1 phonology in processing weak L2 speech from speakers with whom they share (aspects of) their L1. Second, ISIB might be present for listeners with somewhat higher proficiency in L2 (e.g., between 45 and 80 on LexTALE), when the test materials are embedded in sentences or longer stretches of speech and are likely to contain helpful supra-segmental cues.

The current results do not support ISIB for Talkers, as it is outlined by Hayes-Harb et al. (2008), because there was no evidence of listeners processing matched-accented English faster or more accurately than native English. However, Bent and Bradlow (2003) originally described ISIB for Listeners as non-native listeners having either *no disadvantage* or having an advantage in

understanding non-native accents compared to native listeners. If ISIB for Talkers was specified in the same way as ISIB for Listeners, then a lack of systematic disadvantage with matched-accent processing would have constituted evidence for the hypothesis. When observing **Figure 4** in the present study, it may be interpreted that the listeners towards the low end of the English proficiency continuum have no systematic disadvantage relative to a native accent, which would be consistent with the no disadvantage phrasing of ISIB for Listeners.

Native listeners of a language typically have a disadvantage when processing foreign accents over native accents (e.g., Hayes-Harb et al., 2008). If data from monolingual native English listeners had also been obtained, we could have found out if the Bulgarian-accented speech induced an accuracy disadvantage for native listeners compared to the listeners at the end of the proficiency continuum, and thus test the ISIB for Listeners hypothesis. Unfortunately, that was beyond the scope of the current study.

We have, however, provided a novel approach to ISIB for Talkers by looking at the effect of the listeners' L2 proficiency on their reaction times in addition to their accuracy when processing a matched accent. The only other study with a similar design has been reported by Ludwig and Mora (2017). Their results generally differed from the results reported here. For example, their low proficiency L2 listeners had a reaction time advantage for matched L2 accent over native stimuli in L2 and their high proficiency listeners had no difference in their reaction times for the two accents. However, their results were consistent with the general trend that listeners with relatively lower proficiency in L2 would cluster around the matched-accent advantage/no difference side of the continuum, while listeners with higher proficiency cluster towards the no difference/disadvantage side of the interlanguage processing continuum. Other studies that fit this trend while investigating accuracy are Hayes-Harb et al. (2008), Imai et al. (2005), Pinet et al. (2011). However, each of these studies used different experimental tasks, measures of proficiency, and outcome variables, making direct comparison difficult. Future research can focus on replicating these studies.

As noted above, the observed trade-off between matched-accent and native accent reaction times and accuracy is consistent with the Perceptual Assimilation Model-L2 (Best & Tyler, 2007), which is one of the models of L2 phonology underpinning ISIB (Bent & Bradlow, 2003). According to the Perceptual Assimilation Model-L2, as listeners' proficiency in L2 increases during learning, their perceptual categories in L2 should become more independent of their L1 perceptual categories. Hence it might be expected that auditory stimuli that have less similarity with L1 and more similarity with native L2 accents would activate L2 representations in the listeners faster and more reliably than matched-accent L2 stimuli. Other models of L2 phonology also make similar predictions. For example the Automatic Selective Perception model (Strange, 2011) predicts that increased proficiency is linked to developing perceptual routines that are attuned to native-like phonetic characteristics, making native L2 processing more efficient than

matched-accent L2 processing would be. This is consistent with the observed results, which indicate that with increasing proficiency there is a change in how both accents are processed relative to each other. It can only be speculated that if the experiment were adapted to beginners' English vocabulary, they would rely more on Bulgarian phonological categories and therefore have a matched-accent reaction times and accuracy advantage instead of the observed native accent advantage.

As the listeners received no feedback of correctness and were only exposed to monosyllables, this suggests that they updated their representations on a sub-syllabic level during the experiment. In addition, since we used monosyllabic single words as stimuli to reduce the potential complicating effect of prosody on the listeners' responses, it may be concluded that it was specifically segmental phonetic properties that were attended to. It seems the native English stimuli were closer to the internal representations of the Bulgarian listeners.

Although not supportive of ISIB, these results are consistent with Best and Tyler (2007) who state that with additional L2 experience, L2 listeners might in some cases become better at perceiving initially-difficult L2 contrasts. Typically, difficult contrasts in L2 include cases where the contrast corresponds to a single phoneme in L1 (e.g., both the English /o/ and /ɔ/ can correspond to the Bulgarian /ɔ/). Listeners need to become aware of the phonetic nuances that distinguish the contrasting pair in L2. This is precisely the kind of phonetic nuance that is more likely to be missing in matched-accented speech stimuli (e.g., "bowl" and "ball" pronounced as homonyms by Bulgarian-accented speakers). When familiarizing to a non-native accent it is conceivable that a high-proficiency English listener would be slower to process [rɔl] as "roll" than [rol] as "roll" even if there were no competing homonym [rɔl] known to them, as observed in this study. Such sensitivity to rich phonetic information in L2 has been observed by Eger and Reinisch (2019). In their study, the listeners who were better at exploiting phonetic nuance perceptually in L2 were also better at producing such nuance. This suggests that the proficiency effect in the present study may have been found to be more robust if speech production proficiency had also been measured and incorporated in the analysis.

Our results show that the biggest difference in reaction times between the two accents was observed in the first five trials of the block. This suggests the amount of information necessary to adapt to most of the idiosyncrasies of a new speaker's voice is correspondingly small. The fact that the reaction time difference between the two accents was not completely neutralized until the end of a speaker block, however, is reminiscent of the results of Floccia et al. (2009) who report that native English listeners responding to non-native accents in English do not reach the reaction times they achieve with unfamiliar native accents. Floccia et al. (2009) propose that native listeners employ different methods of adapting to unfamiliar native and non-native accents due to their, respectively, systematic and unsystematic phonetic variability.

It is difficult to apply the same interpretation for the present results because most participants reported having more exposure to native English speakers than to matched-accented speakers. Increased exposure to native accent could reasonably be expected to lead to native accent advantage (e.g., Lagrou et al., 2011; Weber et al., 2014). Hence, it cannot be conclusively determined if the lack of adaptation to Bulgarian-accented English was caused by the lack of experience with the accent or because of its unsystematic phonetic variability. The results are also consistent with Bruggeman and Cutler (2020) who concluded that emigrants who lack exposure to novel speakers of their native language increasingly struggle to adapt to new speakers of that language. It is likely that the participants in this experiment also rarely came across new matched-accented speakers, leading to a lack of real-world experience, and a lack of adaptation within the experimental block.

Reduced exposure to native speakers might partly explain why a matched-accent benefit tends to be reported for low L2 proficiency listeners. Due to their lower experience with native accents, the L2 learners might not be able to benefit from the native accents' greater phonetic predictability and phonetic nuance to the same extent as high proficiency listeners (Eger & Reinisch, 2019), because proficiency implies exposure. In addition, low proficiency learners have greater reliance on their L1 phonology (Best & Tyler, 2007) and potentially experience a relatively higher exposure to a variety of non-native accents in formal teaching contexts (Ludwig & Mora, 2017). However, unlike the participants in Bruggeman and Cutler (2020) who were Dutch emigrants in Australia, the participants in the present experiment were geographically closer to the country with their native language and would have had more opportunities to visit and encounter new L1 speakers. This leads to a speculation that the lack of adaptation was driven by the low exposure to novel matched-accent speakers specifically, instead of low exposure to new L1 speakers.

Outside of single-word lexical decision tasks in experimental set-ups, L2 listeners are more likely to encounter matched accents in longer stretches, whether as spontaneous or read speech (e.g., conversations, media, audiobooks). Understanding ISIB for Talkers in these contexts would also require a study of matched-accent prosody and cross-linguistics influences (cf., Mennen, 2004). The presentation of experimental stimuli in sentential contexts could also affect the presence of matched-accent benefit by providing richer acoustic material for the listeners to adapt to. It is likely that there are common segmental or suprasegmental qualities of the speech, such as slower speech rate, specific boundary placement, or vowel reduction, which are characteristic of L2 speakers, regardless of L1 background (Bent & Bradlow, 2003; Bradlow, Kim, & Blasingame, 2017; Götz, 2013). The presence of similar phonetic characteristics across L2 speech from various L1 backgrounds could improve the intelligibility of L2 speech among matched and unmatched L2 listeners, perhaps even those with high L2 proficiency, compared to L1 speech aimed at other L1 listeners who are familiar with its suprasegmental patterns and phonetic shortcuts.

## 6. Conclusion

This study suggested that there was no explicit advantage for Bulgarian–English bilinguals, (specifically L1 Bulgarians living in the UK) to listen to Bulgarian-accented English (i.e., a matched-accented L2) compared to listening to native English speech. Indeed, there was a native English accent advantage. Native English speech was perceived better in terms of higher overall accuracy, lower reaction times, and more rapid short-term reaction time adaptation. The listeners' overall speed and accuracy with the two accents was affected by their English proficiency only when the two accents were compared against each other but not when the dataset was split by accent.

ISIB for Talkers, on the other hand, predicts that L2 listeners would be more effective at processing matched-accented L2 than native L2 productions (Hayes-Harb et al., 2008). However, an earlier version of ISIB accepts that finding no difference could be interpreted as evidence of a benefit (Bent & Bradlow, 2003). This earlier interpretation of ISIB received some support in the results of the participants with the lowest English proficiency in our sample: They did not appear to have systematically faster reaction times or higher accuracy scores for either Bulgarian-accented English or native English speech. Greater proficiency was associated with lower accuracy and slower processing of Bulgarian-accented English, relative to improved accuracy and faster processing of native English speech. This is consistent with the predictions of the Perceptual Assimilation Model-L2, among others, that as learners increase in L2 proficiency they might show decreased reliance on L1 phonology in L2 speech processing. The token-to-token reaction time adaptation to new speakers (with either accent) that we found can also be taken to show a native accent advantage throughout the block, with the biggest difference observed in the first five trials. This suggests that non-native listeners with relatively high L2 proficiency might experience similar difficulties even with matched non-native accents as native listeners do with foreign accents (cf., Floccia et al., 2009). This study provides specific information about matched-accent processing in a listening situation which prioritizes phonetic segmental information. As a result, future studies investigating this process can make specific predictions based on the listeners' proficiency and exposure to matched accents.

Overall, this in-depth study of the speed and accuracy of bilinguals processing a non-native language in a lexical-decision task shows how greater proficiency in their L2 is associated with a matched-accent disadvantage, but only when such processing by the listeners is directly compared to their processing of native accent stimuli in that L2.

## Abbreviations

CI – confidence intervals

edf – estimated degrees of freedom

GAMM – generalized additive mixed model

ISIB – interlanguage speech intelligibility benefit

k – knots, convergence points

L1 – first language

L2 – second language

ms – milliseconds

SD – standard deviation

## Additional file

The additional file for this article can be found as follows:

- **Appendix.** A full list of the real words. DOI: https://doi.org/10.16995/labphon.6423.s1

## Acknowledgements

## Funding Information

## Competing Interests

The authors have no competing interests to declare.

# References

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). The CELEX lexical database (CD-ROM). In *Linguistic data consortium*. University of Pennsylvania.

Baese-Berk, M. M., Bradlow, A. R., & Wright, B. A. (2013). Accent-independent adaptation to foreign accented speech. *Journal of the Acoustical Society of America, 133*(3), EL174. DOI: https://doi.org/10.1121/1.4789864

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2020). *lme4: Linear mixed-effects models using 'Eigen' and S4* (1.1-26) [Statistical package]. https://github.com/lme4/lme4/

Bent, T., & Bradlow, A. R. (2003). The interlanguage speech intelligibility benefit. *Journal of the Acoustical Society of America, 114*(3), 1600–1610. DOI: https://doi.org/10.1016/j.wocn.2008.04.002

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception. In O.-S. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 13–34). Amsterdam: John Benjamins Publishing Company. DOI: https://doi.org/10.1075/lllt.17.07bes

Boyadziev, P., & Tilkov, D. (1997). *Fonetika na balgarskiya knizhoven ezik (The phonetics of standard Bulgarian)*. Veliko Tarnovo: Abagar.

Bradlow, A. R., & Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition, 106*(2), 707–729. DOI: https://doi.org/10.1016/j.cognition.2007.04.005

Bradlow, A. R., Kim, M., & Blasingame, M. (2017). Language-independent talker-specificity in first-language and second-language speech production by bilingual talkers: L1 speaking rate predicts L2 speaking rate. *Journal of the Acoustical Society of America, 141*(2), 886. DOI: https://doi.org/10.1121/1.4976044

Bridges, D., Pitiot, A., MacAskill, M. R., & Peirce, J. W. (2020). The timing mega-study: Comparing a range of experiment generators, both lab-based and online. *PeerJ, 8*, e9414. DOI: https://doi.org/10.7717/peerj.9414

Broersma, M. (2012). Increased lexical activation and reduced competition in second-language listening. *Language and Cognitive Processes, 27*(7–8), 1205–1224. DOI: https://doi.org/10.1080/01690965.2012.660170

Bruggeman, L., & Cutler, A. (2020). No L1 privilege in talker adaptation. Bilingualism: Language and Cognition, *23*(3), 681–693. DOI: https://doi.org/10.1017/S1366728919000646

Bulgarian Ministry of Education. (2019). Angliyski ezik za postigane na nivo B2. (English language for achieving level B2). *Uchebni programi po chuzhd ezik. Nivo B2. (Study plans for foreign language teaching. Level B2)*. Retrieved from https://www.mon.bg/bg/1698

Bundgaard-Nielsen, R. L., Best, C. T., & Tyler, M. D. (2011). Vocabulary size is associated with second-language vowel perception performance in adult learners. *Studies in second language acquisition, 33*(3), 433–461. DOI: https://doi.org/10.1017/S0272263111000040

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience, 14*(5), 365–376. DOI: https://doi.org/10.1038/nrn3475

Cenoz, J. (2013). Defining Multilingualism. *Annual Review of Applied Linguistics, 33*, 3–18. DOI: https://doi.org/10.1017/S026719051300007X

Cheng, L. S. P., Burgess, D., Vernooij, N., Solís-Barroso, C., McDermott, A., & Namboodiripad, S. (2021). The problematic concept of native speaker in psycholinguistics: Replacing vague and harmful terminology with inclusive and accurate measures. *Frontiers in Psychology, 12*, 3980. DOI: https://doi.org/10.3389/fpsyg.2021.715843

Choi, J. Y., Hu, E. R., & Perrachione, T. K. (2018). Varying acoustic-phonemic ambiguity reveals that talker normalization is obligatory in speech processing. *Attention, Perception, & Psychophysics, 80*(3), 784–797. DOI: https://doi.org/10.3758/s13414-017-1395-5

Clarke, C. M., & Garrett, M. F. (2004). Rapid adaptation to foreign-accented English. *Journal of the Acoustical Society of America, 116*(6), 3647–3658. DOI: https://doi.org/10.1121/1.1815131

Cook, V. (1999). Going beyond the native speaker in language teaching. *TESOL Quarterly, 33*(2), 185–209. DOI: https://doi.org/10.2307/3587717

Dokovova, M., Scobbie, J. M., & Lickley, R. (2021). *Matched accent processing: High English proficiency may impede the processing of Bulgarian-accented English for Bulgarian-English bilinguals: dataset.* DOI: https://doi.org/10.17605/OSF.IO/2YA6G

Edwards, J., Pexman, P., & Hudson, C. (2004). Exploring the dynamics of the visual word recognition system: Homophone effects in LDT and naming. *Language and Cognitive Processes, 19*(4), 503–532. DOI: https://doi.org/10.1080/01690960344000215

Eger, N. A., & Reinisch, E. (2019). The impact of one's own voice and production skills on word recognition in a second language. *Journal of Experimental Psychology. Learning, memory, and cognition, 45*(3), 552–571. DOI: https://doi.org/10.1037/xlm0000599

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–277). Baltimore: York Press.

Flege, J. E., & MacKay, I. R. A. (2004). Perceiving vowels in a second language. *Studies in Second Language Acquisition, 26*(1), 1–34. DOI: https://doi.org/10.1017/S0272263104026117

Floccia, C., Butler, J., Goslin, J., & Ellis, L. (2009). Regional and foreign accent processing in English: Can listeners adapt? *Journal of Psycholinguistic Research, 38*(4), 379–412. DOI: https://doi.org/10.1007/s10936-008-9097-8

Georgieva, M. (2010). EFL: From 'You sound like Dickens' to international English. In M. Saxena & T. Omoniyi (Eds.), *Contending with globalization in World Englishes* (pp. 113–131). Bristol: Multilingual Matters. DOI: https://doi.org/10.21832/9781847692764-009

Götz, S. (2013). *Fluency in native and nonnative English speech.* Amsterdam: John Benjamins Publishing Company. DOI: https://doi.org/10.1075/scl.53

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

Hayes-Harb, R., Smith, B. L., Bent, T., & Bradlow, A. R. (2008). The interlanguage speech intelligibility benefit for native speakers of Mandarin: Production and perception of English

word-final voicing contrasts. *Journal of Phonetics, 36*(4), 664–679. DOI: https://doi.org/10.1016/j. wocn.2008.04.002

Imai, S., Walley, A. C., & Flege, J. E. (2005). Lexical frequency and neighbourhood density effects on the recognition of native and Spanish-accented words by native English and native Spanish speakers. *Journal of the Acoustical Society of America, 117*(2), 896–907. DOI: https://doi. org/10.1121/1.1823291

Kim, J., Gabriel, U., & Gygax, P. (2019). Testing the effectiveness of the Internet-based instrument PsyToolkit: A comparison between web-based (PsyToolkit) and lab-based (E-Prime 3.0) measurements of response choice and response time in a complex psycholinguistic task. *PLoS ONE, 14*(9), e0221802. DOI: https://doi.org/10.1371/journal. pone.0221802

Kriengwatana, B., Terry, J., Chládková, K., & Escudero, P. (2016). Speaker and accent variation are handled differently: Evidence in native and non-native listeners. *PLoS ONE, 11*(6). DOI: https://doi.org/10.1371/journal.pone.0156870

Lagrou, E., Hartsuiker, R. J., & Duyck, W. (2011). Knowledge of a second language influences auditory word recognition in the native language. *Journal of Experimental Psychology. Learning, memory, and cognition, 37*(4), 952–965. DOI: https://doi.org/10.1037/a0023217

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

Levis, J. M., Sonsaat, S., Link, S., & Barriuso, T. A. (2016). Native and nonnative teachers of L2 pronunciation: Effects on learner performance. *TESOL Quarterly, 50*(4), 894–931. DOI: https:// doi.org/10.1002/tesq.272

Long, J. (2020). *Interactions* (1.1.3) [Statistical package]. Retrieved from https://interactions. jacob-long.com

Ludwig, A., & Mora, J. C. (2017). Processing time and comprehensibility judgments in non-native listeners' perception of L2 speech. *Journal of Second Language Pronunciation, 3*(2), 167–198. DOI: https://doi.org/10.1075/jslp.3.2.01lud

Luk, G., & Bialystok, E. (2013). Bilingualism is not a categorical variable: Interaction between language proficiency and usage. *Journal of Cognitive Psychology, 25*(5), 605–621. DOI: https:// doi.org/10.1080/20445911.2013.795574

Marian, V., Bartolotti, J., Chabal, S., & Shook, A. (2012). CLEARPOND: Cross-linguistic easy-access resource for phonological and orthographic neighborhood densities. *PLoS ONE, 7*(8), e43230. DOI: https://doi.org/10.1371/journal.pone.0043230

Max Planck Institute for Psycholinguistics. (2001). *WebCelex*. Retrieved from http://celex.mpi. nl/

Mennen, I. (2004). Bi-directional interference in the intonation of Dutch speakers of Greek. *Journal of Phonetics, 32*(4), 543–563. DOI: https://doi.org/10.1016/j.wocn.2004.02.002

Mitterer, H., Eger, N. A., & Reinisch, E. (2020). My English sounds better than yours: Second-language learners perceive their own accent as better than that of their peers. *PLoS ONE, 15*(2), e0227643. DOI: https://doi.org/10.1371/journal.pone.0227643

Mullennix, J. W., Pisoni, D., & Martin, C. (1989). Some effects of talker variability on spoken word recognition. *Journal of the Acoustical Society of America, 85*(1), 365-378. DOI: https://doi.org/10.1121/1.397688

Munro, M. J., Derwing, T. M., & Morton, S. L. (2006). The mutual intelligibility of L2 speech. *Studies in Second Language Acquisition, 28*(1), 111–131. DOI: https://doi.org/10.1017/S0272263106060049

Nakai, S., Lindsay, S., & Ota, M. (2015). A prerequisite to L1 homophone effects in L2 spoken-word recognition. *Second Language Research, 31*(1), 29–52. DOI: https://doi.org/10.1177/0267658314534661

Pinet, M., Iverson, P., & Huckvale, M. (2011). Second-language experience and speech-in-noise recognition: Effects of talker–listener accent similarity. *Journal of the Acoustical Society of America, 130*(3), 1653–1662. DOI: https://doi.org/10.1121/1.3613698

Porretta, V., Tucker, B. V., & Järvikivi, J. (2016). The influence of gradient foreign accentedness and listener experience on word recognition. *Journal of Phonetics, 58*, 1–21. DOI: https://doi.org/10.1016/j.wocn.2016.05.006

Rastle, K., Harrington, J., & Coltheart, M. (2002). 358,534 nonwords: The ARC nonword database. *Quarterly Journal of Experimental Psychology, 55*(4), 1339–1362. DOI: https://doi.org/10.1080/02724980244000099

Reinisch, E., Weber, A., & Mitterer, H. (2013). Listeners retune phoneme categories across languages. *Journal of Experimental Psychology. Human perception and performance, 39*(1), 75–86. DOI: https://doi.org/10.1037/a0027979

Roessel, J., Schoel, C., & Stahlberg, D. (2020). Modern notions of accent-ism: Findings, conceptualizations, and implications for interventions and research on nonnative accents. *Journal of Language and Social Psychology, 39*(1), 87–111. DOI: https://doi.org/10.1177/0261927X19884619

Rubenstein, H., Lewis, S. S., & Rubenstein, M. A. (1971). Evidence for phonemic recoding in visual word recognition. *Journal of Verbal Learning & Verbal Behavior, 10*(6), 645–657. DOI: https://doi.org/10.1016/S0022-5371(71)80071-3

Seedhouse, P., Harris, A., Naeb, R., & Üstünel, E. (2014). *The relationship between speaking features and band descriptors: A mixed methods study* (No. 2; IELTS Research report series, pp. 1–30). British Council, Cambridge English Language Assessment, IDP, IELTS. Retrieved from https://www.ielts.org/-/media/research-reports/ielts_online_rr_2014-2.ashx

Selvi, A. F. (2014). Myths and misconceptions about nonnative English speakers in the TESOL (NNEST) movement. *TESOL Journal, 5*(3), 573–611. DOI: https://doi.org/10.1002/tesj.158

Sidaras, S. K., Alexander, J. E. D., & Nygaard, L. C. (2009). Perceptual learning of systematic variation in Spanish-accented speech. *Journal of the Acoustical Society of America, 125*(5), 3306–3316. DOI: https://doi.org/10.1121/1.3101452

Sóskuthy, M. (2017). *Generalised additive mixed models for dynamic analysis in linguistics: A practical introduction.* Working paper. Retrieved from: https://arxiv.org/abs/1703.05339v1

Stoet, G. (2010). PsyToolkit—A software package for programming psychological experiments using Linux. *Behavior Research Methods, 42*(4), 1096–1104. DOI: https://doi.org/10.3758/BRM.42.4.1096

Stoet, G. (2017). PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology, 44*(1), 24–31. DOI: https://doi.org/10.1177/0098628316677643

Strange, W. (2011). Automatic selective perception (ASP) of first and second language speech: A working model. *Journal of Phonetics, 39*(4), 456–466. DOI: https://doi.org/10.1016/j.wocn.2010.09.001

Sumner, M., & Samuel, A. G. (2009). The effect of experience on the perception and representation of dialect variants. *Journal of Memory and Language, 60*(4), 487–501. DOI: https://doi.org/10.1016/j.jml.2009.01.001

Ternes, E., & Vladimirova-Buhtz, T. (1990). Bulgarian. *Journal of the International Phonetic Association, 20*(1), 45–47. DOI: https://doi.org/10.1017/S0025100300004072

Van Engen, K. J., Baese-Berk, M., Baker, R. E., Choi, A., Kim, M., & Bradlow, A. R. (2010). The Wildcat corpus of native-and foreign-accented English: Communicative efficiency across conversational dyads with varying language alignment profiles. *Language and Speech, 53*(4), 510–540. DOI: https://doi.org/10.1177/0023830910372495

Weber, A., Betta, A. M. D., & McQueen, J. M. (2014). Treack or trit: Adaptation to genuine and arbitrary foreign accents by monolingual and bilingual listeners. *Journal of Phonetics, 46*, 34–51. DOI: https://doi.org/10.1016/j.wocn.2014.05.002

Wilden, E., & Porsch, R. (2020). Teachers' self-reported L1 and L2 use and self-assessed L2 proficiency in primary EFL education. *Studies in Second Language Learning and Teaching, 10*(3), 631–655.  DOI: https://doi.org/10.14746/ssllt.2020.10.3.9

Witteman, M. J., Weber, A., & McQueen, J. M. (2013). Foreign accent strength and listener familiarity with an accent codetermine speed of perceptual adaptation. *Attention, Perception, & Psychophysics, 75*(3), 537–556. DOI: https://doi.org/10.3758/s13414-012-0404-y

Wood, S. (2021). Package 'mgcv' (1.8-38) [Computer software]. Retrieved from https://cran.r-project.org/web/packages/mgcv/mgcv.pdf