MASTERARBEIT/ MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

"Perception of English /æ/ versus /e/ by native speakers of Bosnian/Croatian/Serbian"

verfasst von / submitted by

Olja Dobrić, BA

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Arts (MA)

Wien, 2023 / Vienna 2023

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 066 812
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium English Language and Linguistics
Betreut von / Supervisor:	Univ.-Prof. Mag. Dr. Nikolaus Ritt

Acknowledgements

Completing this research project and writing the thesis would not have been possible without the unwavering support of a number of people who directly or indirectly helped me in the process.

First and foremost, I would like to express my sincere gratitude to my supervisor, Univ.-Prof. Mag. Dr. Nikolaus Ritt, whose guidance, encouragement, and valuable insights have helped me put my previous knowledge and ideas into reality.

Furthermore, my gratitude extends to my friends for their understanding and support, and, in the context of this research project, their eagerness to provide speech samples. I would also like to thank all the participants in the present study for dedicating their time and effort to complete the experiment.

Finally, my appreciation goes out to my treasured family, especially my parents and my partner. To my father for his continuous emotional and financial support. To my mother for being the absolute best role model anyone could have, and who is solely responsible for awakening and supporting my love for and interest in languages. To my partner, for his belief in me, and his willingness to submit himself to my endless rants about linguistics, phonology, and my thesis.

Thank you all!

Hvala vam svima!

Table of Contents

1. **Introduction** 1
2. **The impact of L1 phonology on the perception of L2** 2
   - 2.1 The Perceptual Assimilation Model and patterns of assimilation 3
   - 2.2 L2 experience and the Perceptual Assimilation Model 6
3. **The vowels in English** 9
   - 3.1 Coarticulation effects in pre-nasal, pre-fortis, and pre-lenis environments 11
   - 3.2 An analysis /e/ and /æ/ in Pennsylvanian English 12
4. **The vowel system in Bosnian/Croatian/Serbian** 18
   - 4.1 Classification and description of BCS vowels 20
   - 4.2 Vowels /e/ and /a/ in different BCS dialects: A small-scale study 22
     - 4.2.1 *Methods and participants* 22
     - 4.2.2 *Results and conclusion* 24
5. **ELF versus EFL in acquisition of L2 phonology** 27
   - 5.1 Conformity to the native-speaker ideal in SLA of phonology 29
   - 5.2 A review of previous studies in the context of Bosnian/Croatian/Serbian 30
6. **The present study design** 32
   - 6.1 Research questions and hypotheses 32
   - 6.2 Experimental design 34
   - 6.3 Participants 35
   - 6.4 Task 1: Auditory perception of words in isolation/ minimal pair distinction 36
   - 6.5 Task 2: Category goodness rating of /æ/ and /e/ 42
   - 6.6 Statement rating: Exposure to English and attitudes to NS norms 45
   - 6.7 Testing procedure 48
7. **Statement ratings: Results and discussion** 49
   - 7.1 Participants' exposure to English 49
   - 7.2 Participants' attitudes to native-speaker norms 51
   - 7.3 Discussion: Exposure to L2 and attitudes to NS norms 55
8. **Auditory perception tasks: Results and discussion** 57
   - 8.1 Results of Task 1: Perception of words in isolation/minimal pair distinction 58
     - 8.1.1 *Pre-lenis environment* 60
     - 8.1.2 *Pre-nasal environment* 61
     - 8.1.3 *Pre-fortis environment* 64
   - 8.2 Results of Task 2: Similarity ratings between English vowels and BCS /e/ 65
   - 8.3 Discussion: Acquisition routes according to PAM 68
     - 8.3.1 *Category-Goodness (CG type) assimilation pattern* 70
     - 8.3.2 *Single-Category (SC type) assimilation pattern* 72

8.3.3 *Uncategorised-Categorised (UC type) assimilation pattern* 73

9. **Bringing it all together: Assimilation of /e/ and /æ/ into BCS /e/, exposure to English, and (non-)conformity to native-speaker norms** 75

10. **Conclusion** 79

11. **Bibliography** 81

12. **Appendix** 85

12.1 English abstract 85

12.2 German abstract 86

12.3 Appendix A: Complete dataset of formant frequencies and vowel duration of Pennsylvanian /æ/ and /e/ in 24 tokens used in Task 1 87

12.4 Appendix B: Complete dataset of /e/ and /a/ formants and durations in different BCS dialects 87

12.5 Appendix C: Design and visual stimuli for 24 acoustic tokens in Task 1 90

12.6 Appendix D: Formant frequencies in acoustic stimuli in Task 2 96

12.7 Appendix E: Participant information sheet 96

12.8 Appendix F: Task 1 complete dataset 101

12.9 Appendix G: Task 2 complete dataset 102

12.10 Appendix H: L2 experience statement ratings (complete dataset) 104

12.11 Appendix I: attitudes to NS norms statement ratings (complete dataset) 106

List of abbreviations

L1.....The first language

L2.....The second language

SLA.....Second language acquisition

PAM.....Perceptual Assimilation Model

SC.....Single-Category assimilation pattern

TC.....Two-Category assimilation pattern

CG.....Category-Goodness assimilation pattern

UC.....Uncategorisable-Categorisable assimilation pattern

UU.....Uncategorisable-Uncategorisable assimilation pattern

NA.....Non-Assimilable

NS.....Native speaker

NNS.....Non-native speaker

EFL.....English as a foreign language

ELF.....English as a lingua franca

LFC.....Lingua Franca Core

RP.....Received Pronunciation

GA.....General American

BCS.....Bosnian/Croatian/Serbian

F1.....the first formant

F2.....the second formant

IPA.....International Phonetic Alphabet

CGR.....Category goodness rating

SD.....Standard deviation from the mean

List of figures

Figure 1. Acoustic versus articulatory representation of vowels in American English (Ladefoged & Maddieson 1996: 286).....10

Figure 2. Mean F1 and F2 of /e/ and /æ/ in 12 tokens per vowel, taken from a native speaker from Doylestown, Pennsylvania. The x-axis shows F2 (2500-0 Hz), while F1 is represented by the y-axis (0-1200 Hz).....13

Figure 3. Mean F1 and F2 of the vowel /æ/ in 12 words, as produced by a native speaker from Doylestown, Pennsylvania. The x-axis shows F2 (2500-0 Hz), while F1 is represented by the y-axis (0-1200 Hz).....14

Figure 4. Mean F1 and F2 of the vowel /e/ in 12 words, as produced by a native speaker from Doylestown, Pennsylvania. The x-axis shows F2 (2500-0 Hz), while F1 is represented by the y-axis (0-1200 Hz).....15

Figure 5. The visual representation of the vowel space in BCS (Subotić, Sredojević & Bjelaković 2012: 45).....20

Figure 6. The four accents in BCS (Subotić, Sredojević & Bjelaković 2012: 98).....21

Figure 7. Vowel /e/ in BCS, as pronounced by speakers from different regions. The x-axis shows F2 (2500-0 Hz), while F1 is represented by the y-axis (0-1000 Hz).....24

Figure 8. Vowel /a/ in BCS, as pronounced by speakers from different regions. The x-axis shows F2 (2500-0 Hz), while F1 is represented by the y-axis (0-1000 Hz).....25

Figure 9. The Lingua Franca Core, as proposed by Jenkins (2009: 142-143). The left column shows EFL, while the right column shows EFL targets in phonology.....28

Figure 10. Mean F1 and F2 of the vowel /æ/ in 12 words, as produced by a native speaker from Doylestown, Pennsylvania. The x-axis shows F2 (2500-0 Hz), while F1 is represented by the y-axis (0-1200 Hz).....38

Figure 11. Mean F1 and F2 of the vowel /e/ in 12 words, as produced by a native speaker from Doylestown, Pennsylvania. The x-axis shows F2 (2500-0 Hz), while F1 is represented by the y-axis (0-1200 Hz).....39

Figure 12. The visual representation of the first task.....41

Figure 13. The design of Task 2: category goodness rating.....43

Figure 14. Graphic representation of vowel F1 and F2 values for 8 words used in Task 2, together with the representation of BCS vowel /e/ in different regions. The x- axis shows the F2 (2500-0 Hz), while the y-axis shows F1 (0-1200 Hz). English vowels are shown in blue, while BCS /e/ is orange.....45

Figure 15. Ratings of statements about the participants' exposure to English-language media. The x-axis shows the three statements in question (music, TV shows/films/YT videos, reading books/newspapers/articles in English). The y-axis shows the percentages of participants who

answered “strongly disagree” (orange), “agree” (yellow), “neutral” (grey), “agree” (blue), and “strongly agree” (green).....50

Figure 16. Ratings of statements about the use of English. The x-axis shows the five statements in question. The y-axis shows the percentages of participants who answered “strongly disagree” (orange), “agree” (yellow), “neutral” (grey), “agree” (blue), and “strongly agree” (green).....51

Figure 17. Mean score $\bar{x}$ (blue) and SD (orange) in five EFL statements. The y-axis shows the mean scores of the statements ($\bar{x}$=0-5). The x-axis shows the statements in question.....52

Figure 18. Mean score $\bar{x}$ (blue) and SD (orange) in seven ELF statements. The y-axis shows the mean scores of the statements ($\bar{x}$=0-5). The x-axis shows the statements in question.....53

Figure 19. Ratings of EFL-oriented statements. The x-axis shows the five statements in question. The y-axis shows the percentages of participants who answered “strongly disagree” (orange), “agree” (yellow), “neutral” (grey), “agree” (blue), and “strongly agree” (green).....54

Figure 20. Ratings of ELF-oriented statements. The x-axis shows the five statements in question. The y-axis shows the percentages of participants who answered “strongly disagree” (orange), “agree” (yellow), “neutral” (grey), “agree” (blue), and “strongly agree” (green).....55

Figure 21. Overall results of Task 1. The x-axis shows six possible variables: /æ/ correctly selected, /e/ selected in words with /æ, distractions for /æ/ selected, /e/ correctly selected, /æ/ selected in words with /e/, distractions for /e/ selected. The y-axis shows the total number of selections per variable.....60

Figure 22. Results of Task 1 for /æ/ and /e/ in pre-lenis environment. The x-axis shows the words in question, while y-axis shows the number of participants.....61

Figure 23. Results of Task 1 for /æ/ and /e/ in pre-nasal environment. The x-axis shows the words in question, while y-axis shows the number of participants.....63

Figure 24. Results of Task 1 for /æ/ and /e/ in pre-fortis environment. The x-axis shows the words in question, while y-axis shows the number of participants.....65

Figure 25. Similarity rating between English vowels and BCS /e/. The x-axis shows the words, vowels, and their environments, while y-axis shows the values from 0 to 100. Mean score ($\bar{x}$) values are represented by the blue columns, while the standard deviation (s) is represented by the orange line. The exact s and $\bar{x}$ values are given in the table under the chart.....67

Figure 26. Illustration of the CG type assimilation, based on the results of Task 1 and Task 2. The illustration shows how an acoustic stimulus is translated into the mental representation of the minimal pairs.....71

Figure 27. Illustration of the SC type assimilation, based on the results of Task 1 and Task 2. The illustration shows how an acoustic stimulus is translated into the mental representation of the minimal pairs.....73

Figure 28. Illustration of the UC type assimilation, based on the results of Task 1 and Task 2. The illustration shows how an acoustic stimulus is translated into the mental representation of the minimal pair.....75

List of tables

Table 1. Mean F1 and F2 comparisons of /e/ and /æ/ in pre-nasal position.....16

Table 2. Vowel quantity of /æ/ and /e/ in 12 minimal pairs pronounced by a native speaker from Doylestown, Pennsylvania.....17

Table 3. BCS words containing /a/ and /e/.....23

Table 4. Vowel quantity of BCS /e/ and /a/, as pronounced by native speakers from different regions.....26

Table 5. Minimal pairs/words used in Task 1, together with their phonological context.....37

Table 6. Vowel quantity of /æ/ and /e/ in 12 minimal pairs pronounced by a native speaker from Doylestown, Pennsylvania.....39

Table 7. Words used in Task 2, together with their phonological environment, IPA representation, and vowel quantity in seconds.....44

Table 8. Statements about NS norms.....47

Table 9. Mean F1 and F2 comparisons of /e/ and /æ/ in pre-nasal position.....64

Table 10. 10 individual answers to Task 2 category goodness rating (CGR).....68

1. Introduction

Second language acquisition of phonology can be quite a laborious task, particularly due to the process known as phonological transfer (Marković & Jakovljević 2016: 215). During the acquisition of a second language, "the L1 phonological system operates as a filter that accommodates linguistic input to its structure", leading to the interference with and the replacement of L2 phonemes in production and perception of the target language (Boomershine 2013: 105). Such a strong influence of native phonology on the acquisition of the non-native phonological system is rooted in the fact that, unlike L2 phonemes, L1 phonemes are not 'taught' or a product of instruction, but rather 'learned' during the cognitive development of a child (Naser 2011: 4-5). Given that speakers of a language are "tuned in to the specific phonetic categories used in their native language", they will perceive and articulate L2 sounds in terms of similarity to their own native phonology (Tyler et al. 2014: 4).

One of the most prominent models for explaining this phenomenon is the Perceptual Assimilation Model (PAM), which predicts assimilation and poor discrimination of target sounds that are similar to the L1 repertoire, and, at the same time, successful perception and the small likelihood of assimilation of the sounds that are perceived as less ideal candidates for the native categories (Best & Strange 1992; Best 1994; Best 1995; Strange et al. 2004; Best & Tyler 2007; Hallé & Best 2007; Wrembel, Marecka & Kopečková 2019). To account for different assimilation possibilities and predictions about the perception of contrasts, the model posits a number of assimilation patterns/types, such as Two-Category, Single-Category, Category-Goodness, Uncategorised-Categorised, Uncategorised-Uncategorised, and Non-Assimilable (Best 1995: 194-195). Furthermore, L2 experience and exposure, as a sociolinguistic variable, has a profound impact on assimilation patterns, as the increase in exposure to the second language facilitates better perception of target contrasts, diminishes assimilation into L1 categories, and leads to subsequent formation of distinct L2 categories (Best & Strange 1992).

In the context of Bosnian/Croatian/Serbian (BCS), the model has been applied in the study of assimilation of English vowels into BCS vowel categories, as the former greatly outnumber the latter (Marković 2009a; Marković 2009b; Marković & Jakovljević 2016; Krebs-Lazendić 2008; Krebs-Lazendić & Best 2007; Paunović 2011). Out of all English vowels, the /e/-/æ/ vocalic contrast and their assimilation into the single BCS category /e/ have been studied extensively, but the focus has mostly been on production rather than perception, and, when the

perception of English sounds is investigated, the participants in the experiments are majors of English, who are clearly not representatives of the wider population. Moreover, the influence of BCS phonology on the results is heavily stigmatised and deemed as highly detrimental.

Therefore, the purpose of this thesis is to examine the perception of /e/-/æ/ vocalic contrast by adult native speakers of BCS and to describe the perception in terms of the Perceptual Assimilation Model (PAM), i.e. to determine the exact pattern of assimilation proposed by the model. In addition to the phonetic/phonological part, the relationship between English language experience/exposure and assimilation into BCS category /e/ will be investigated, together with the interrelatedness of (non-)conformity to native-speaker norms and patterns of assimilation.

With regard to the structure of the thesis, sections 2-5 will provide a theoretical overview of the Perceptual Assimilation Model, an analysis of vowels produced by a native speaker for this study, a small-scale study of BCS vowels /e/ and /a/ in different regions, as well as the discussion of EFL and ELF in SLA of phonology, which includes conformity to native-speaker norms and a review of previous studies in the context of BCS. The methodology of the present study is outlined in section 6, and it covers the research questions and hypotheses, participant information, experimental design and testing procedure, auditory perception tasks, and statement ratings. The results of the statement rating tasks and the two perception tasks are reported and discussed in sections 7 and 8, respectively. Finally, section 9 serves as a general discussion combining the phonetic/phonological and the sociolinguistic part of the experiment.

2. The impact of L1 phonology on the perception of L2

The phonological system of the mother tongue has a considerable effect on "the perceptual mapping from the highly variable acoustic speech signal to a linguistic representation", that is, the recognition of phonemic contrasts in L2 (Holt & Lotto 2010: 1218). In this case, the influence of native phonology and phonological transfer is discernible in hindered recognition of L2 words and slower processing of the target language (Flege & Mackay 2004: 2). According to Best and Strange (1992: 305-306), perception of L2 phonemes is shaped by L1 experience, meaning that adult learners find it strenuous to detect phonemes or phones which are not "employed contrastively in the phonological system of their native language", and they perceive target sounds "in terms of their similarities (and dissimilarities) to native phonemes". In contrast to the "perceptual attunement" to the L1 repertoire in the early childhood, "there is a sort of 'tuning out' of non-native contrasts that are irrelevant in the native language" (Tyler

et al. 2014: 4). For instance, monolingual Korean and Japanese speakers find it difficult to recognise the difference between the English liquids /r/ and /l/, since these phonemes do not make up a contrast in their first language (Best & Strange 1992: 305). Those acquiring a foreign language, therefore, "perceive the L2 in terms of their L1 and cannot perceive L2 differences that are not made in their L1" (Major 2008: 75).

What needs to be noted, however, is that the learners of a second language have a tendency to rely primarily on specific "acoustic dimensions" of sounds while listening to L2 speech input, such as spectral (vowel quality, place of articulation) and temporal (vowel quantity, especially in pre-fortis and pre-lenis contexts) cues (Polka 1992: 39). Furthermore, in the absence of L2 "spectral counterparts in the native language", listeners typically resort to vowel duration as the indicator of difference, regardless of the fact that vowel quantity is not used to signal meaning contrasts in L1 (Strange et al. 2004: 1793). For example, a study by Marković (2009b: 256) found that intrinsic duration of English /e/ and /æ/ was the primary cue for native speakers of Bosnian/Serbian/Croatian in the perception of the vocalic contrast. In the tasks where the participants had to rely more on spectral features (i.e. tongue position) rather than quantity, the discrimination between the two English vowels was relatively poor, as compared to the task in which the duration of the vowels in question was over-emphasised in the speech sample and thus became the primary cue (Marković 2009b: 256). Similarly to native speakers of BCS, Japanese speakers distinguish English vowel contrasts by relying heavily on the duration of the vowels rather than their quality (Strange et al. 1998). Duration is, therefore, highly relevant in signalling vocalic distinctions for non-native listeners.

2.1 The Perceptual Assimilation Model and patterns of assimilation

One of the most prominent models for explaining the impact of L1 on perception of L2 sounds, primarily contrasts, is the Perceptual Assimilation Model or PAM, as a part of the larger Motor-Theory of speech perception (Best 1994; Best 1995; Best & Tyler 2007; Strange & Best 1992; Hallé & Best 2007). The model "emphasises native language and second language perceptual similarity as a predictor of difficulties in the discrimination of non-native contrasts", meaning that L2 sounds that are perceptually dissimilar to the sounds in one's native language will be easier to discriminate, while the sounds that are too similar and almost identical those sounds in the native inventory will be difficult to perceive (Boomershine 2013: 105). Furthermore, L2 sounds are assimilated into L1 phoneme categories that are perceived as closest to the target categories (Krebs-Lazendić 2008: 7-8). The listener perceives L2 sounds as as similar to L1 phonemes and assimilates them accordingly on the basis of "gestural similarity (or

discrepancy) between non-native and native segments", i.e. the similarity in the movements of the tongue (raising-lowering, fronting-backing), or different airflow constrictions between the tongue and other articulators, the shape and movements of the lips and the jaw, movements of the glottis and the velum, etc. (Krebs-Lazendić & Best 2007: 282). In other words, non-native sounds will be assimilated if the listener detects that the acoustic signal is highly identical to the articulatory movements that are equivalent to their L1 sounds. An example would be assimilation of the English voiced dental fricative /ð/ into the voiced plosive category /d/ in languages where the consonant is fully dental or leaning towards the dental place of articulation. Similarly, Best (1995: 193-194) posits that, if a listener's language lacks dental stops, but has velar, bilabial, and alveolar stops, they will associate non-native dental stops with alveolar stops in their own language, since the "tongue tip construction of the dental stop is straightforwardly closer in native phonological space to the alveolar place than to the others". In contrast, the likelihood of the assimilation of phones that are too distant from the native inventory is rather small (Wrembel, Marecka & Kopečková 2019: 515). As a consequence, new L2 categories will be established with greater ease for such sounds (Strange et al. 2004: 1791).

The assimilation of non-native phonemic contrasts into native categories occurs in several possible patterns, which can used to predict discrimination possibilities between sounds (Strange et al. 2001: 1692). What needs to be taken into account, however, is that each pattern of assimilation is reliant on how the listeners themselves perceive similarity/dissimilarity between the source and the target sounds, and that "listeners themselves may differ from one another, even within the same language, with respect to which phonetic properties of a non-native phone they may detect or attend to in perception" (Best & Strange 1992: 307). It is therefore plausible to assume that the same non-native contrast will be perceived and subsequently assimilated differently among those who share the same L1. The remainder of this subsection summarises the patterns of assimilation proposed by PAM.

The first pattern of assimilation, namely, the assimilation of the two members of the L2 contrast into two distinct categories in the L1 inventory, is known as Two-Category assimilation or TC type (Best 1994: 14). In this case, both phones are "perceived as acceptable exemplars of two different native phonemes" and the likelihood of category formation for the specific target sounds and their separation from L1 phonemes is small (Best & Tyler 2007: 23). Auditory discrimination between L2 sounds in TC assimilation would be most accurate, meaning that the listeners are able to hear the difference between the auditory stimuli (Tyler et al. 2014: 6).

An example of two category assimilation would be the assimilation of English dental fricative voicing contrast /ð/-/θ/ into /z/ and /s/ in German or into BCS /d/ and /t/ respectively, or the assimilation of Hindi retroflex stop and aspirated dental stop sounds into English /d/ and /ð/ (Best 1994:14).

In Single-Category assimilation (SC type), both L2 sounds are assimilated into a single L1 category and judged as either "equally acceptable" or "equally deviant" exponents of the native category (Best 1995: 195). For example, native speakers of Spanish assimilate the English vowel pair /i:/-/ɪ/ into the vowel space corresponding to their Spanish long equivalent (Boomershine 2013: 115-116). In the cases of SC type, detecting the acoustic differences between the members of the L2 contrast would be almost impossible (Levy & Strange 2008: 145). As a consequence, "minimally contrasting words would be perceived as homophones" (Best & Tyler (2007: 29). The eventual category formation could only be facilitated by the high frequency of such words (Best & Tyler 2007: 30).

Similarly to SC type, Category-Goodness (CG type) predicts assimilation of the L2 contrast into a single category in L1 (Best & Tyler 2007: 23). The two members of the contrast differ, however, in category-goodness, that is, the proximity to the "native ideal" (Best 1995: 195). While one sound is perceived as a "good instance of a native category", the other segment in the same contrast is judged as a "poor instance of that same category" (Strange et al. 2001: 1692). Discrimination of sounds in the CG type is expected to be better than in the SC Type, but is heavily dependent on the category-goodness difference between the sounds (Wrembel, Marecka & Kopečková 2019: 515). An example of CG type is can be observed in the study by Polka (1995; cited in Tyler et al. 2014: 7), who examined the perception of German /u/ and /y/ contrasts by English speakers and found that German /u/ is perceived as a better candidate for the native category than /y/. A new category "is reasonably likely to be formed eventually for the deviant L2 phone, while the L2 phone that is judged as a better exemplar would be perceived as phonologically and phonetically equivalent to the L1 category", and the likelihood of learning the new category for the phone equivalent to L1 is rather small, but not impossible (Best & Tyler 2007: 29).

According to the Uncategorisable-Uncategorisable pattern (UU type), members of a non-native contrast may also be assimilated into L1 phonological space but they cannot be associated with any of the native categories or exist "in between specific native categories", i.e. they are uncategorisable (Best 1995: 194-195). In most cases, the difference is difficult to detect, due to the absence of an association with a "clear L1 phonological contrast" (Faris, Best & Tyler:

2018: 2). Differentiating between such sounds is influenced by "the phonetic similarity of the two phones to each other and to native phonological categories" (Tyler et al. 2014: 7). For instance, those exposed to Inuktitut sounds, and whose native languages do not employ uvular sounds, may be expected to assimilate Inuktitut uvular stops into their L1 phonological space, but fail to categorise such sounds (Johnson 1978: 132).

It may also happen that, while one member of the L2 phonemic contrast is assimilated into an L1 category, the other one "falls in phonetic space, outside of native categories", which is known as UC type (Uncategorised-Categorised) (Best 1995: 195). The contrast that falls under the UC Type assimilation is "discriminated very well because it reflects a phonological distinction between an exemplar of a known phoneme versus something that is not an exemplar of that phoneme", and it is essentially a difference between a source and a target sound (Best & Tyler 2007: 23). UC assimilation is evidenced in native speakers of Australian English, who assimilate the Danish vowel /oe/ into their L1 category, but fail to do the same to /ø/, the other member of the Danish contrast (Faris, Best & Tyler 2018: 2).

The final pattern is the Non-Assimilable or the NA type, where L2 sounds are neither assimilated nor perceived as human speech-sounds at all, because they are "so deviant from the articulatory properties of native phonemes" (Best & Tyler 2007: 23). The discrimination of such sounds is expected to be excellent, as the "listener [attends] to psychoacoustic differences between them" (Polka 1992: 39). Such sounds can be either "made by vocal tract events" (choking and coughing) or even be the sounds that are not produced by the speech apparatus (the clapping of hands, finger cracking, etc.) (Best 1995: 193). The linguistic examples of NA type would include the perception of Zulu click consonants by speakers of English, since their "double-stop articulations and suction release gestures are highly deviant from any English gestural constellations"; however, the discrimination between such consonants would be excellent, even without prior exposure to stimuli (Best 1995: 196). Nevertheless, NA type is only possible with non-vowel sounds, as "vowels that are non-assimilable to speech (perceived as non-speech sounds) are unlikely to exist at all" (Tyler et al. 2014: 7).

2.2 L2 experience and the Perceptual Assimilation Model

As previously mentioned, the degree of language transfer, and the patterns of assimilation are contingent on the individual listeners with the same L1 background (Best & Strange 1992: 307). Indeed, "every learner brings to bear a unique set of learning styles, abilities, and personality traits" that have an immense impact on how they perceive and subsequently

produce sounds of a foreign language (Moyer 2014: 448). Among these various factors, two sociolinguistic factors, namely the amount of experience with and exposure to a foreign language, and the age when the acquisition began have been widely examined in the studies of the SLA of phonology (Paunović 2003: 389). Nonetheless, the question of which one of the aforementioned factors is more crucial needs to be addressed. According to Moyer (2014: 446-447), the age of learning cannot be singled out, and it is coupled with L2 experience/exposure. In other words, an earlier start equals more experience with the phonological system of the target language, which refers to

a) "the length of residence, or time spent in the target language country", i.e. the experience gained through direct immersion into L2,
b) "years spent studying the foreign language in a classroom", including the extracurricular activities such as English language courses,
c) "daily or weekly hours spent using L2", and
d) "the extent to which L2 is relied upon to perform certain functions (e.g., personal and social expression, opposed to work-related and perfunctory language use)" (Moyer 2014: 448).

With reference to b), Vilke (1991: 487) argues that the amount of experience with the target language in classroom settings has a meagre impact on the SLA of phonology, as phonological transfer occurs because pupils are not properly instructed (or not instructed at all) in learning to "identify distinctive features of the L2 phonemes" and their differences to the native repertoire. As a result, classroom learners of L2 "hear approximations of L1 phonemes and they use them in speech" (Vilke 1991: 487). Furthermore, the absence of a proper phonetic training will eventually lead to students putting more emphasis on "reading comprehension, grammar, spelling, and writing" (Naser 2011: 4-5). Apart from years of contact with the L2 in the classroom, Housen (2012: 130-131) highlights the intensity and high volume of L2 instruction on a weekly basis. What also needs to be noted is that c) and d) are interrelated, given that the speakers will use L2 in performing social functions, be it watching TV and reading in L2, or using the target language to communicate with peers (Paunović 2003: 389).

In the context of PAM, the model was initially created for naive, inexperienced listeners with little to no contact with L2, but has since been frequently used in the context of L2 learning (Bjelaković 2018: 114). The degree of assimilation is expected to be the greatest within inexperienced listeners, while "increased L2 experience may foster improved recognition of the discrepancies between L1 and L2" (Best & Strange 1992: 307). As the acquisition of the

second language progresses, the "linguistic pressure for the L2 learner to rephonologise perception of the target contrasts" will be more powerful, and the emergence of distinct categories for target sounds is guaranteed (Best & Tyler 2007: 32). In other words, as the listener's exposure to non-native phonology increases, it will facilitate recognition of the subtle phonetic details of L2 sounds, and their association with and assimilation into L1 categories will start to diminish, resulting in clear L2 categories. When looking at individual assimilation patterns, particularly Single-Category and Category-Foodness, it can be argued that increased L2 experience will eventually result in category formation (Best & Tyler 2007: 29). In CG scenarios, prolonged exposure to and experience with the target sounds will facilitate category formation for the L2 phone that is judged as a better exponent of the L1 category due to a perceived high degree of similarity to the native phoneme, whereas an increase in the exposure to high frequency words containing the L2 contrasts could eventually enable the recognition of differences between the two L2 sounds that have been assimilated on the basis of the SC pattern (Best & Tyler 2007: 29). The newly formed ability to discriminate between the two signals will result in one of the phones being perceived as a better token of the L1 category, that is, the listener will move from SC type to CG, and, finally, achieve separate category formation for both non-native sounds (Kopečková 2012: 237). Therefore, increased experience with the target language facilitates better discrimination of L2 contrasts and triggers the establishment of distinct L2 categories.

The role of L2 experience in diminished assimilation is apparent in the study by Strange and Best (1992: 331), in which Japanese listeners with more experience with English were more successful in detecting the phonetic and phonological differences between the members of the English approximant contrasts than the inexperienced Japanese listeners. In another study, Kopečková (2012: 250) examined the variation in L1-L2 similarity judgements among several groups of native speakers of Polish living in Ireland and Poland. Child and teenage Polish speakers of English demonstrated weaker similarity judgements between Irish English vowels and Polish vowels, due to their increased experience gained through English immersion programmes (Kopečková 2012: 250). Nevertheless, adult native speakers of Polish who had been living in Ireland for years prior to the experiment were more successful in detecting English vowels and judged them as less similar to Polish vowels, in contrast to adult native speakers of Polish who were living in Poland at the time (Kopečková 2012: 251). Finally, Levy and Strange (2008: 1) investigated the perception of Parisian French contrasts / u-œ/, /i-y/, /u-y/ as well as /y- œ/ by native speakers of English, and found that listeners with more experience

with L2 vowels were better at distinguishing the vowel contrasts in non-existent syllables, notwithstanding the phonetic environment of the vowels (pre-bilabial or pre-alveolar). However, the relative success of experienced participants varied from contrast to contrast, despite the years of experience with French (Levy & Strange 2008: 162). The findings from the French-English context are supported by Levy (2009: 1140), who concludes that, "with language experience, L2 learners may begin to perceive segments on a more abstract level, less affected by acoustic variation".

3. The vowels in English

Following the discussion of phonological transfer in cross-language speech perception, the Perceptual Assimilation Model and its patterns, as well as the interrelatedness of PAM and L2 experience, this section will focus on English vowels, with the emphasis on their articulatory and acoustic description, and coarticulation effects in pre-nasal, and pre-fortis/pre-lenis environments. The section will conclude with the analysis of spectral and temporal features of /e/ and /æ/, as pronounced by a native speaker from Pennsylvania.

What needs to be mentioned first is the two 'ways' of describing vowels: phonological and phonetic. In phonology, vowels are defined as units that form the nuclei of syllables (Bjelaković 2016: 13). With the exception of syllabic consonants, syllables in English cannot exist without vowels (Plag et al. 2015: 57). In contrast to the rather simple definition of vowels at the phonological level, the phonetic definition of vowels is slightly more detailed. Vowels are "produced on a pulmonic egressive airstream with central airflow", accompanied by vocal-fold vibration, meaning that the air is expelled upwards from the lungs and it escapes through the centre of the mouth, and the resulting sounds are voiced (McMahon 2001: 68). It is possible, however, for the air to escape through the nasal cavity, if the velum is lowered, resulting in nasal vowels or vowels with nasalised quality. While nasal vowels are employed contrastively in languages like French, the nasal quality of English vowels is regarded as a product of coarticulation (Krebs-Lazendić 2008: 13). Moreover, vowels are defined as "continuants", i.e. the sounds that are produced with an almost uninterrupted airflow, which makes it quite arduous for the speaker to pinpoint their exact 'place' in the mouth cavity, "since the tongue never moves close enough to the roof of the mouth [...] to make its position easy to feel" (McMahon 2001: 68). Indeed, it would be much easier for the speaker to ascertain the exact place of articulation of /b/ or /d/ than that of /e/ or /u:/, as the constriction for the former is either clearly visible or physically perceptible.

Such continuous nature of vowels makes their classification more complex. Whereas consonants are classified according to the place and manner of articulation, as well as voicing, vowels are classified in relation to the "tongue body positions", that is, its height and frontness (Ladefoged & Maddieson 1996: 282). Vowel height refers to the relative raising or lowering of the tongue, resulting in either low, mid, or high vowels, and the term frontness is used to designate the closeness of the tongue to the lips and teeth during the articulation, and the vowels are either front, central, or back (McMahon 2001: 69-71). Equally important is the shape of the lips during articulation, which, together with height and frontness of the tongue, determines the quality of vowels (Krebs-Lazendić 2008: 13). In English, however, vowel quantity or vowel length is an important cue attested at the phoneme level, and it signals meaning differences (McMahon 2001: 73). Therefore, the height of the tongue, its degree of fronting, and the shape of the lips (quality) and the length of articulation (quantity) are all distinctive features of English vowels.

Nevertheless, according to Ladefoged & Maddieson (1996: 285), the representation of vowels based solely on their articulatory features is "not as close to the usual linguistic representation [...] as is that provided by the acoustic data", as shown in the following figure:

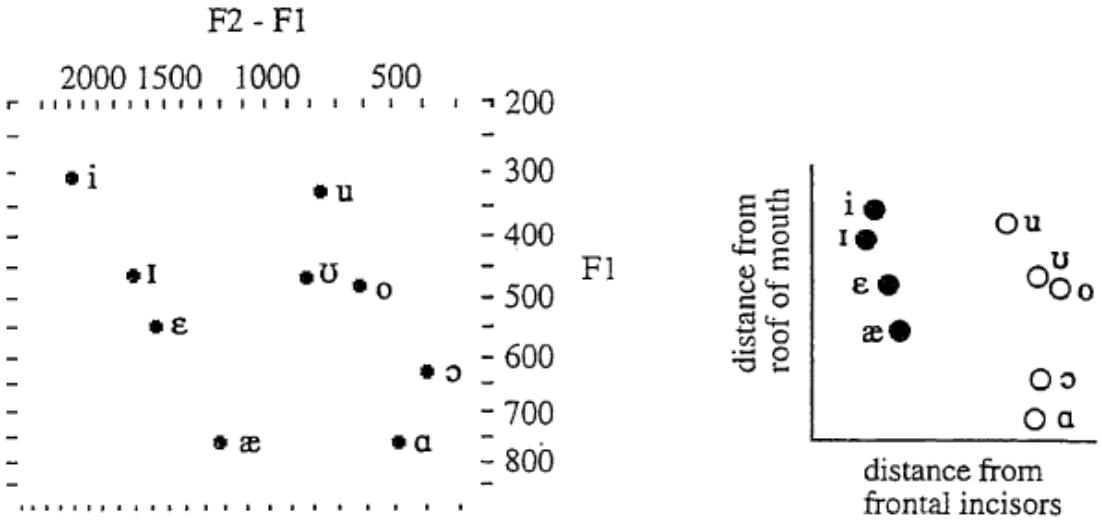


Figure 1. Acoustic versus articulatory representation of vowels in American Englis (Ladefoged & Maddieson 1996: 286).

Indeed, the more accurate representation of vowels is based on their formants, or "modes of vibration (resonance) of the column of air in the vocal tract" (Krebs-Lazendić 2008: 16). Formant frequencies, expressed in Hz, are determined by the quality of vowels, and the two most relevant formants for the analysis and classification of English vowels are F1 and F2

(Subotić, Sredojević & Bjelaković 2012: 38). What is the correlation between the first two formants and the quality of vowels? The first formant, F1, corresponds to the height of the tongue in reverse, meaning that the higher F1 value, the lower the vowel (Bjelaković 2018: 15). The value of F2 denotes the degree of tongue fronting and backing, with front vowels having a higher F2 value than the back vowels (Krebs-Lazendić 2008: 15). When the vowels are plotted on a graph, F1 corresponds to the y-axis, while F2 is presented on the x-axis, and the starting point (f=0 Hz) is shown in the top-right corner, as in Figure 1 (Subotić, Sredojević & Bjelaković 2012: 38). Thus, in contrast to the measurements of distances of the tongue in from the front of the mouth/teeth and the hard palate, plotting vowel positions relative to their formant frequencies results in a more exact two-dimensional representation of English vowels, as their height and frontness are indicated by F1 and F2, respectively.

3.1 Coarticulation effects in pre-nasal, pre-fortis, and pre-lenis environments

In many languages, vowels are seldom produced in isolation, but rather in combination with other sounds, leading to coarticulation, or "overlapping production of adjacent phonetic segments" (Krebs-Lazendić 2008: 18). During the articulation of any linguistic unit above the segment level, i.e. syllables, words, and even sentences, articulatory movements in the vocal tract occur in rather rapid sequences, and certain active articulators, particularly the tongue, may alter their position in anticipation of the upcoming sounds (Krebs-Lazendić 2008: 18). Such changes in the position of the sound modulators are reflected in the "changes in the acoustic properties" of vowels (Krebs-Lazendić 2008: 18). In other words, both quality and quantity of English vowels may be affected by the consonants that follow.

Despite the existence of numerous languages that employ oral and nasal vowels contrastively (French being one of them), meaning that there are minimal pairs containing either an oral or the nasal counterpart of the seemingly same vowel, nasalisation in English is purely a product of coarticulation, and is an allophonic feature (Ladefoged & Maddieson 1996: 298). As a rule, all English vowels are produced with a nasal quality before /m/, /n/, or /ŋ/, due to the lowering of the velum in preparation for the articulation these nasal consonants (McMahon 2001: 86). The resulting vowels are neither "purely oral" nor nasal, but the airflow escapes through both the oral and the nasal cavities (Krebs-Lazendić 2008: 19). This is particularly evident in the different qualities of the vowel /ɪ/ in words *bin* and *bid*: /ɪ/ is produced with a lowering of the velum in anticipation of the upcoming /n/ in *bin*, and the air escapes through the nose and the mouth, while the quality and the oral release remains unchanged in *bid*.

In addition, “articulatory strength” or “fortition” of the following consonants has an impact on the quantity and quality of English vowels (Ladefoged & Maddieson 1996: 95). English obstruents are divided into fortes (/p/, /t/, /k/, /f/, /θ/, /ʃ/, /tʃ/, /s/) and lenes (/b/, /d/, /g/, /v/, /ð/, /ʒ/, /dʒ/, /z/), the former requiring more articulatory strength and pressure during production than the latter (Stahlke 2003: 191). Fortis consonants are often referred to as voiceless, and lenis consonants as voiced, which can be “considered misleading”, as the complex features distinguishing the two groups cannot be solely attributed to the vibration of the vocal cords, but strength (Ladefoged & Maddieson 1996: 95). Fortition influences vowel quantity in the sense that “vowels in stressed syllables are shortened before syllable final fortis obstruents”, which is known as “fortis vowel shortening” or “pre-fortis shortening” (Stahlke 2003: 200). The vowels are shortened because of the rapid movement of the tongue in the obstruction of the airflow for the upcoming fortis consonant (Krebs-Lazendić 2008: 19). In contrast, if a vowel occurs before a lenis consonant, its expected duration will be longer, which is regarded as “pre-lenis lengthening” (Stahlke 2003: 191). A similar effect on vowel quantity can be observed in pre-sonorant (approximant and nasal) contexts, i.e. before non-obstruents (Stahlke 2003: 201). According to this rule, the short vowel /ɪ/ will have different vowel quantity in the words *bid, bin,* and *bit*: the vowel will be the longest before the lenis /d/, followed by the nasal /n/, and the shortest before the fortis consonant /t/.

3.2 An analysis /e/ and /æ/ in Pennsylvanian English

Having investigated English vowels in general, as well as the effects of coarticulation on their quality and quantity, this subsection will analyse the front vowels /e/ and /æ/ based on the acoustic data obtained from a native speaker from Doylestown, Pennsylvania (north of Philadelphia). The quality of both vowels in different environments, together with the differences in length will be taken into consideration.

What needs to be mentioned first is that the tokens were recorded in PRAAT (speech-analysis software), using the mono-sound option at a sampling frequency of 44100 Hz with the Jabra Evolve2 40 headset. The acoustic data for both vowels was taken from 24 tokens, or 12 minimal pairs containing the /e/-/æ/ contrast. The words were either monosyllables or disyllables. The vowels appeared in pre-nasal, pre-lenis, and pre-fortis environments. The following chart illustrates the plotting of the mean F1 and F2 of the two vowels, as extracted from the tokens:

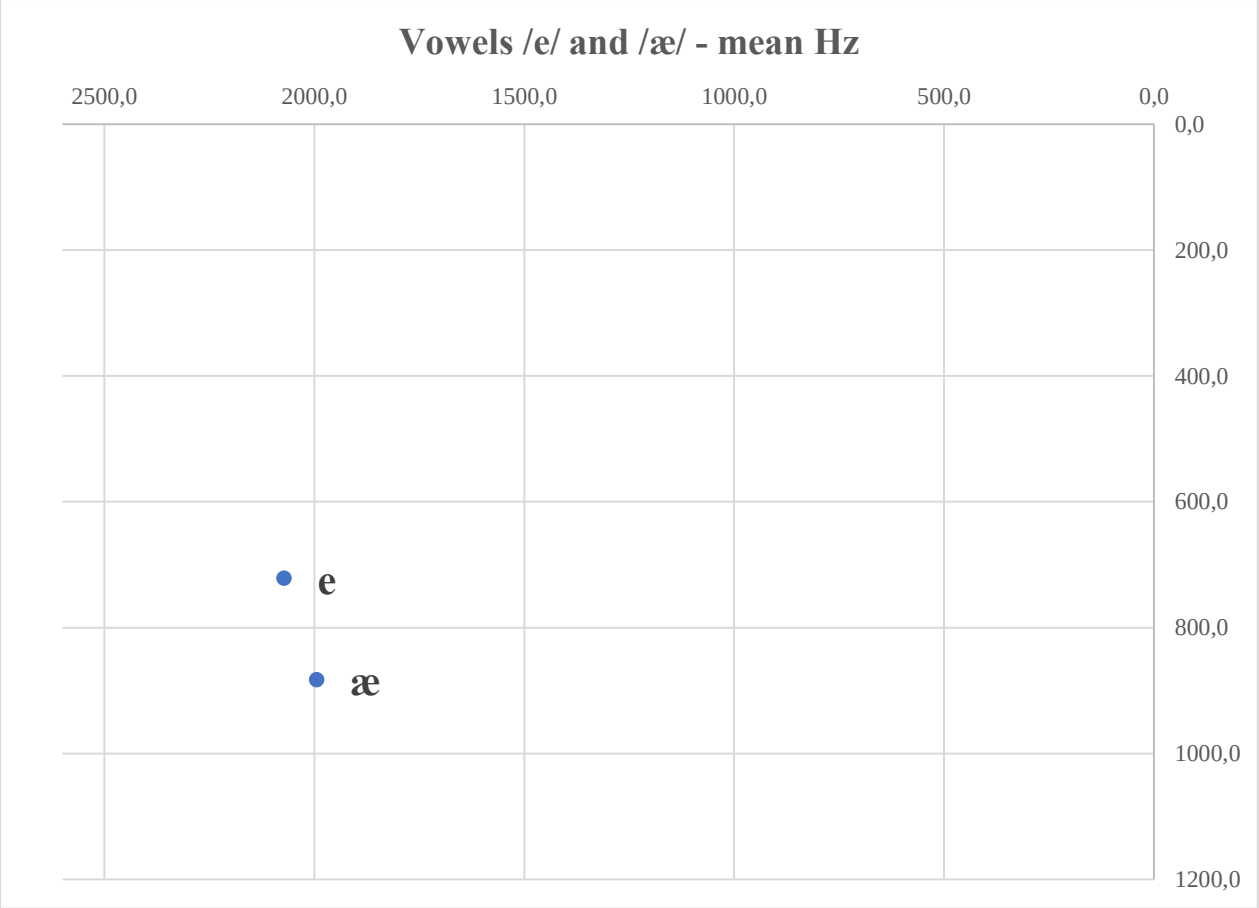


Figure 2. Mean F1 and F2 of /e/ and /æ/ in 12 tokens per vowel, taken from a native speaker from Doyelstown, Pennsylvania. The x-axis shows F2 (2500-0 Hz), while F1 is represented by the y-axis (0-1200 Hz).

As demonstrated in the above chart, /e/ is higher and more fronted than /æ/. Nevertheless, the distinction is not as pronounced as in the chart provided by Ladefoged and Maddieson (1996: 286; cf. Figure 1). Indeed, the mean formant values for /æ/ are F1=882,6 Hz and F2=1994,8 Hz, while the mean values for /e/ are F1=721,3 Hz and F2=2072,4 Hz. The F2 differences in Ladefoged & Maddieson (1996: 286) are more conspicuous than that of the native speaker in this study. Furthermore, the plotting of Pennsylvanian /e/ and /æ/ as more front than in Figure 1 seems to correspond to the findings from Pittsburgh (PA), where no tendency for the backing of these vowels was detected (Labov, Ash & Boberg 2006: 272). If individual words, however, are taken into account, the variation in vowel quality becomes apparent, for /æ/ in particular (for all F1 and F2 frequencies in words, see Appendix A):

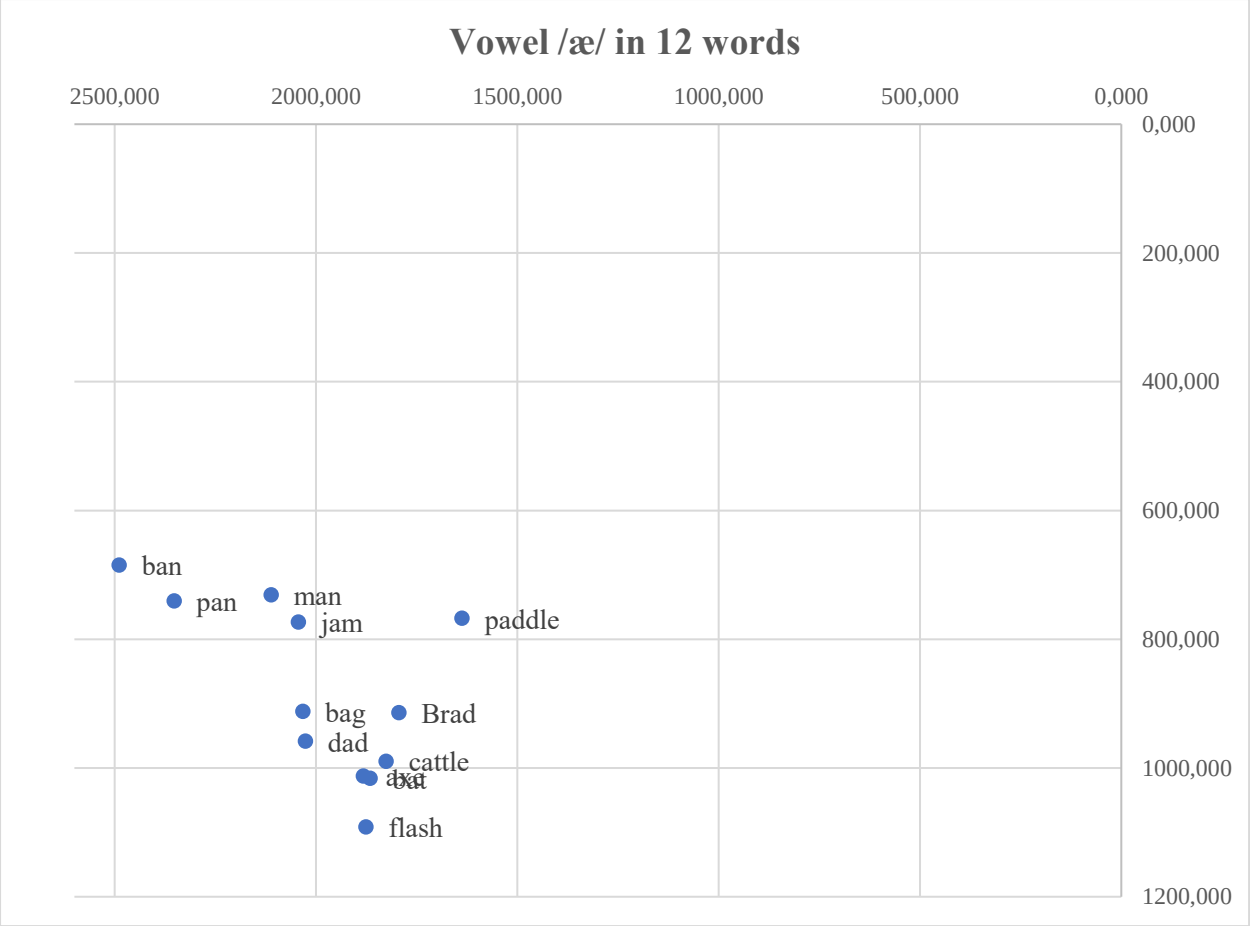


Figure 3. Mean F1 and F2 of the vowel /æ/ in 12 words, as produced by a native speaker from Doylestown, Pennsylvania. The x-axis shows F2 (2500-0 Hz), while F1 is represented by the y-axis (0-1200 Hz).

A significant degree of tongue raising, accompanied by fronting, can be observed in pre-nasal environments, meaning that, in *ban*, *man*, *pan* and *jam*, nasalisation leads to changes in vowel quality. As Labov et al. explain (2016: 276), in all North-American varieties, there is a general tendency for /æ/ to change its quality before bilabial and alveolar nasals and become a tense vowel. The vowel is therefore "articulated in a more extreme and definite way" and "more peripheral", if it occurs before /n/ and /m/ (McMahon 2001: 73). Labov et al. (2016: 277) contribute the tensing of /æ/ in Pennsylvania (and the Midlands) to the emergence of a distinct nasal system that "replaces a pattern of complex allophony", creating "the opportunity to observe the competition of the nasal system with a phonological system to full complexity". In pre-lenis, in this case pre-plosive /d/ and /g/, contexts, the quality of /æ/ is identical to that of the mean F1 and F2 values plotted in Figure 2. The vowel is relatively lower and more central (the degree of tongue fronting varies from word to word) than in pre-nasal position, with *paddle* being the only exception. The heightening and backing of /æ/ in *paddle* may be attributed to coarticulation: in anticipation for the following two consonants, the tongue is raised and backed

towards the alveolar ridge. Furthermore, the backing of the syllabic /l/ towards the velum may have had an effect on the vowel quality change. Interestingly, /æ/ in *cattle* seems to be spectrally closer to the vowel in pre-fortis positions, despite the fact that the alveolar flap [ɾ] should be voiced (Plag et al. 2015: 50). The flap is either produced with more "fortition" than /g/ and /d/, or /l/ has an impact on the vowel quality (Stahlke 2003: 191). Finally, in pre-fortis contexts, namely before /t/, /k/, and /ʃ/, the vowel has undergone a significant degree of lowering, especially in *flash*. The higher F1 of /æ/ is likely influenced by the articulatory nature of the following consonant. In /bæt/ and /æks/, the mean F1 values are almost identical, with the difference of roughly 3 Hz. The degree of tongue fronting, nevertheless, is constant in the three words containing the fortis consonants. Therefore, in the 12 tokens analysed, the quality of /æ/ is variable, depending on the consonants that follow: a significant degree of vowel heightening and fronting is observed in pre-nasal position, /æ/ is lower before lenes and its frontness varies from word to word, and, finally, the vowel is lowered in pre-fortis position, with a relatively small degree of word-to-word variation regarding tongue fronting.

In contrast to /æ/, the quality of /e/ in the 12 tokens pronounced by the native speaker does not show such a high degree of variation:

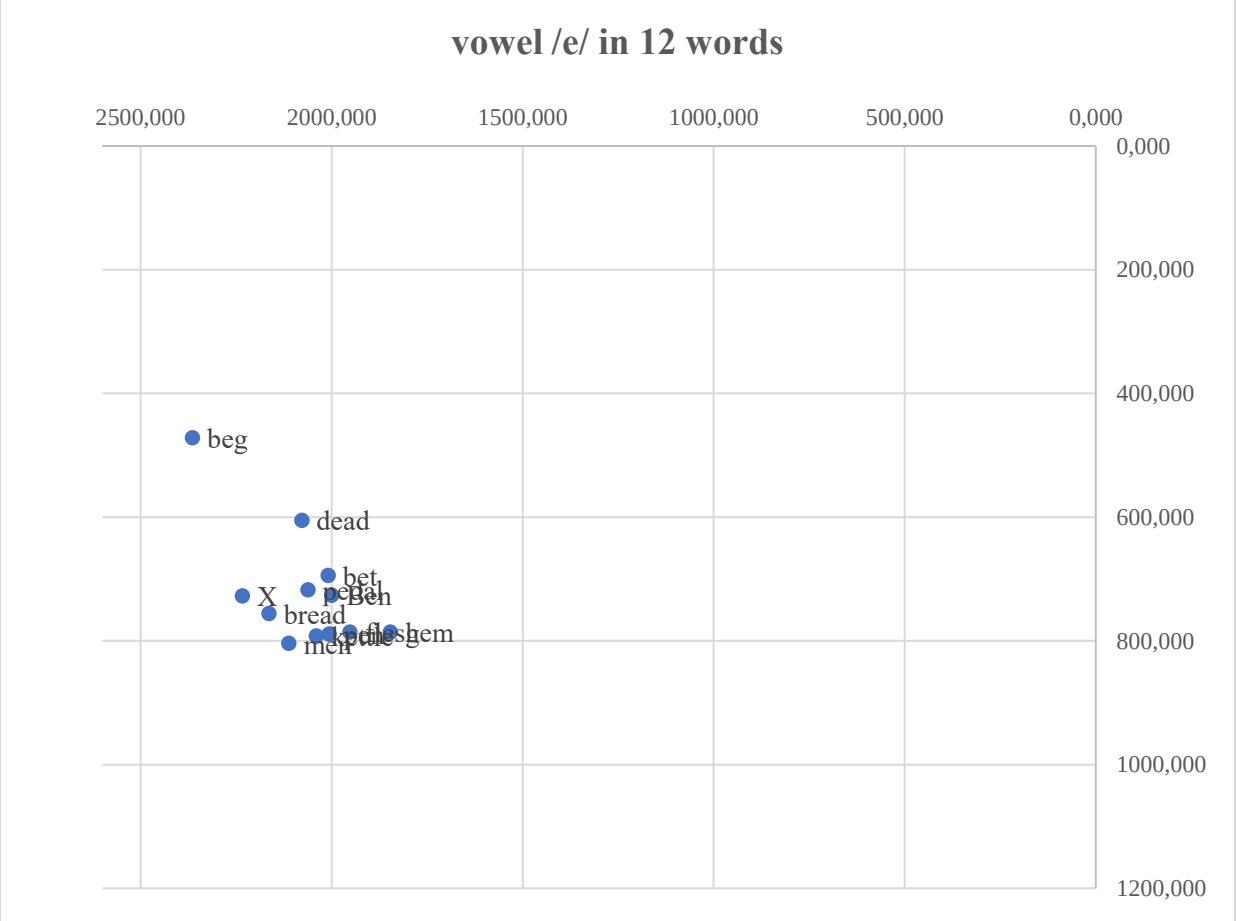


Figure 4. Mean F1 and F2 of the vowel /e/ in 12 words, as produced by a native speaker from Doylestown, Pennsylvania. The x-axis shows F2 (2500-0 Hz), while F1 is represented by the y-axis (0-1200 Hz).

The mean F1 in selection in the 10 of the 12 words (exceptions are *beg* and *dead*) seems to remain constant, ranging from approximately 605 Hz to 804 Hz, regardless of the articulatory nature of the consonants that follow. The heightening of the vowel in *dead*, however, could potentially be attributed to the phonological context, as the vowel may have been initially lowered in *pedal* and *bread* due to the fact that the vowel is either followed or preceded by two consonants. Unfortunately, more data is needed to support this. The significant degree of vowel heightening in *beg* is not caused by the pre-lenis position, but rather by vowel diphthongisation, with /e/ raised towards /ɪ/. The degree of diphthongisation is obvious in vowel quantity comparisons: the duration of /e/ varies between 0,058s and 0,140 s, but its duration is t=0,221 s in *beg*. Equally interesting is the observed slight degree of /e/ lowering in pre-nasal position, which, taken together with /æ/ raising in the same context, means the vowels in minimal pairs are perceptually close in terms of the height of the tongue:

Table 1. Mean F1 and F2 comparisons of /e/ and /æ/ in pre-nasal position.

Word	**F1 in Hz**	**F2 in Hz**	**Word**	**F1 in Hz**	**F2 in Hz**
pan	740,518	2352,179	*pen*	789,297	2006,909
ban	684,852	2488,592	*Ben*	726,176	2000,291
jam	773,406	2044,146	*gem*	786,002	1846,352
man	731,029	2111,712	*men*	804,243	2112,452

As indicated in the above table, the spectral (quality) differences between the vowels before /n/ and /m/ are rather minuscule, and it is plausible to assume that the listeners would need to pay close attention to the extremely subtle differences in either height or tongue fronting, to be able to distinguish between words. Although the quality of /e/ is remains rather unchanged throughout the tokens, a slight lowering of /e/ before nasals, combined with raising of /æ/ in the same position, could cause difficulties in discrimination between the two vowels.

Finally, the impact of the phonological environment on vowel quantity ought to be examined. As evidenced in the following table, /æ/ is generally longer than /e/, particularly in pre-lenis and pre-nasal position:

Table 2. Vowel quantity of /æ/ and /e/ in 12 minimal pairs pronounced by a native speaker from Doylestown, Pennsylvania.

Duration of /æ/		Duration of /e/	
Word	**Duration in /s**	**Word**	**Duration in /s**
pan	0,230	*pen*	0,140
bag	0,279	*beg*	0,221
bat	0,104	*bet*	0,122
ban	0,224	*Ben*	0,122
dad	0,283	*dead*	0,141
flash	0,185	*flesh*	0,134
axe	0,141	*X*	0,112
jam	0,224	*gem*	0,055
paddle	0,120	*pedal*	0,059
Brad	0,220	*bread*	0,097
man	0,217	*men*	0,123
cattle	0,120	*kettle*	0,058
Mean	0,196	**Mean**	0,115

Vowel /æ/ is significantly longer in pre-lenis and pre-nasal position (*pan*, *bag, ban, dad, jam, Brad, man*) than in pre-fortis position (*bat, flash, axe*), which corresponds to Stahlke's descriptions of pre-lenis lengthening and pre-fortis clipping (2003: 191). The only exceptions to this pattern are *paddle* (t=0,120 s) and *cattle* (t=0,120 s). It can be assumed that the fact that the vowel is followed by two consonants may have led to a significant shortening of the vowel. The duration of /e/ in 12 tokens ranges from 0,055 to 0,140 seconds, regardless of the consonants that follow. The only exception is *beg*, where the vowel is slightly diphthongised, but still shorter than /æ/ in *bag*. For example, /e/ is longer in *flesh* than in *gem* and *bread,* and its duration is the same in *bet* and *Ben*, notwithstanding the articulatory strength of the following consonants. While the short duration of /e/ in *bread, pedal,* and *cattle,* can be understood as an effect of the two consonants that follow or the increased force of the alveolar flap, it is unclear why pre-nasal environment does not lead to vowel-lengthening. In spite of such inconsistencies in the quantity dimension of /e/, durational differences between /e/ and

/æ/ are still notable, and it can be argued, that in the cases where "spectral overlap might cause perceptual confusion", such as the pre-nasal raising of /æ/ and lowering of /e/, the subtle vowel quantity distinctions can facilitate better vowel category recognition, as /e/ is generally perceived as shorter than /æ/ (Fridland, Tyler & Farrington 2014: 343). In other words, in minimal pairs where /e/ and /æ/ are perceptually identical and their F1 and F2 are relatively close, the listeners may rely on duration as the primary distinguishing cue. The exception would be the *bat-bet* contrast, as /æ/ is shorter than /e/. Given that /æ/ is considerably lower in *bat* than /e/ in *bet* ($\Delta F1 \approx 300$ Hz), the listeners are more likely to perceive the vowels on the basis of spectral cues.

4. The vowel system in Bosnian/Croatian/Serbian

Having examined English vowels, more precisely /e/ and /æ/, this section will investigate the phonology of Bosnian/Croatian/Serbian, its vowel inventory, and analyse the vowels /e/ and /a/ in different regions where the language is spoken.

What needs to be mentioned first is that BCS has three main dialects: Štokavski, Čakavski, and Kajkavski, which have been named according to their versions of the interrogative *what*, namely *šta/što, ča,* and *ka* (Brown & Alt 2004: 9). Out of the three, Štokavski has been used as a model for the standardisation of Bosnian, Croatian, and Serbian, due to the fact that it is spoken by the wider population than the other two dialects, and that it covers Bosnia, most of the territory of Croatia (inland), and almost all the territory of Serbia (Alexander 2006: 388). Čakavski can be found in Dalmatia, while Kajkavski is spoken in the area surrounding Zagreb and in the northwest, towards the border with Slovenia (Brown & Alt 2004: 9). Nevertheless, the differences between the three dialects go further than a simple interrogative form, with their "radically divergent accentual systems", the overall phonetic system, and some lexical disparities so pronounced that, when mutual intelligibility is taken into consideration, the three dialects "would qualify much more readily as different languages than standard Serbian, Croatian, and Bosnian" (Alexander 2006: 388). This also serves to show how the existence of Bosnian, Croatian, and Serbian today is a direct result of centuries of cultural, religious, and, primarily ideological (nationalistic) differences between the peoples in Bosnia, Croatia, and Serbia, and, of course, the wars that ensued following the breakup of Yugoslavia (Brown & Alt 2004: 7). Therefore, Bosnian, Croatian, and Serbian are mutually intelligible standardised varieties based on a single dialect (Alexander 2006: 401).

Another significant dissimilarity among speakers of Štokavski, and therefore the majority of BCS speakers, is dictated by the remnants of the old-Slavic phoneme "jat", and it is the "source of the single most readily perceptible pronunciation difference" (Alexander 2006: 391). The phoneme in question is an old-Slavic vowel located in the perceptual space between the vowels /e/ and /i/, which, over the centuries, developed into either /e/, /i/, or the "diphthongoids" /je/ and /ije/, and triggered the development of the dialectal subunits or pronunciation forms: ekavian, ikavian, and ijekavian (Subotić, Sredojević & Bjelaković 2012: 124). The differences between the three pronunciation forms can be observed in the BCS word for *child*, which is *dete* (ekavian), *dite* (ikavian), *dijete* (ijekavian). Both ekavian and ijekavian were taken as the bases of standard Serbian, and ijekavian is accepted in standard Croatian and Bosnian, while ikavian is consdered a non-standard pronunciation form (Alexander 2006: 391). What needs to be mentioned, however, is the fact that the distribution of the three dialects and pronunciation forms does not "correspond geographically to the major religious, cultural and political boundaries" (Alt & Brown 2006: 9). Thus,the three dialects and pronunciation forms are strictly region-bound, regardless of social, cultural, historical, religious, and even standard language influences.

Finally, before continuing with the vowel system, the orthography in BCS and its relation to phonology need to be taken into account. Today, BCS is written in two alphabets, Cyrillic and Latin, with both scripts being used officially in Serbia and Bosnia, and only the latter attested in standard Croatian (Alexander 2006: 1). Both alphabets are considered phonemic, meaning that each phoneme is always represented by a single grapheme, and one grapheme only covers a single phoneme in all contexts (Subotić, Sredojević & Bjelaković 2012: 19). In other words, there is one-to-one "correspondence between orthography and phonology", and the conversion from spelling to pronunciation and vice versa is pretty straightforward (Havelka & Rastle 2005: 150). For example, the BCS word for *hand* is /ruka/, and it is written as either <ruka> or <рука>. Furthermore, any sound alternations that occur within words as a result of coarticulation are also attested in orthography, and foreign names are usually transcribed to conform to the native script, especially in Cyrillic (Subotić, Sredojević & Bjelaković 2012: 122). There are also no silent letters, i.e. all "individual letters within each alphabet have a phonemic interpretation that is invariant over letter contexts" (Havelka & Rastle 2005: 150) While this principle of sound-letter agreement holds true for all the letters in the Cyrillic alphabet, there are three exceptions in the Latin alphabet: the phonemes /ʎ/, /ɲ/, and /dʒ/ are represented as digraphs <lj>, <nj>, and <dž> respectively (Alexander 2006: 1). Thus, BCS

generally employs two writing systems that are characterised as “spelling for the ear”, since one can rely solely on their hearing in order to learn them (Subotić, Sredojević & Bjelaković 2012: 122).

4.1 Classification and description of BCS vowels

Similarly to a number of world languages, BCS has a rather simple vowel system, with only five phonemes, which are “evenly distributed in the vowel space” (Marković & Jakovljević 2016: 218). The vowels in question are represented acoustically in the following graph, in “the form of the classic vowel triangle” (Petrović 2001: 26):

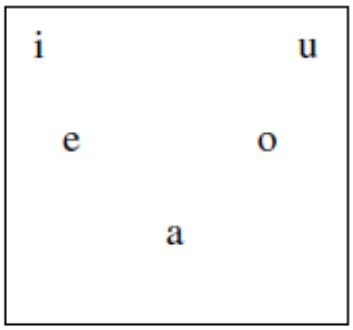


Figure 5. The visual representation of the vowel space in BCS (Subotić, Sredojević & Bjelaković 2012: 45).

As can be observed in the tongue height, /i/ and /u/ are high (closed) vowels, /e/ and /o/ are mid vowels, while /a/ is the only low (open) vowel (Subotić, Sredojević & Bjelaković 2012: 44). Interestingly, formant values of both pairs of high vowels and mid vowels are “found at approximately the same height” (Petrović 2001: 26). In terms of F2, /i/ and /e/ are front vowels, with /e/ slightly more centralised, /a/ is purely central, whereas /u/ and /o/ are back vowels, and /o/ is also slightly centralised (Marković 2009a: 3). During the articulation of front BCS vowels, the tongue is moved towards the hard palate, and during the articulation of the back vowels, it moves towards the velum, while the tongue remains in the space between the velum and the front teeth when /a/ is produced (Subotić, Sredojević & Bjelaković 2012: 44). Furthermore, all of the BCS vowels are pronounced “very cleanly and sharply”, without any reductions in their quality (Alexander 2006: 3). Nevertheless, minor changes in quality may occur as a consequence of coarticulation, meaning that, when /a/, /e/, /i/, /o/, and /u/ are followed by other sounds, they may be nasalised, diphthongised in certain dialects, or altered due to prosodic features (Subotić, Sredojević & Bjelaković 2012: 45). Therefore, in terms of

vowel quality, five phonemes can easily be distinguished in BCS, and their quality may vary, depending on the phonological environment.

The above presented vowels, however, do not seem to have any long-short distinctions, indicating that vowel quantity is not a phonemic feature in BCS. Indeed, the vowels can be pronounced as long or short, but their quantity is "the exponent of long or short tonic (falling or rising) accent" (Marković 2009a: 5). This means that "pitch accents and vowel length are merged and represent a single phonemic category", with length only attested at the level of pronunciation (Čubrović 2017: 59). The term "accent", rather than "stress" is used to signify the involvement of both pitch and quantity in the realisation of syllables, instead of just prominence (Brown & Alt 2004: 16). There are four tonic accents in BCS, with a completely free distribution, illustrated in the following chart:

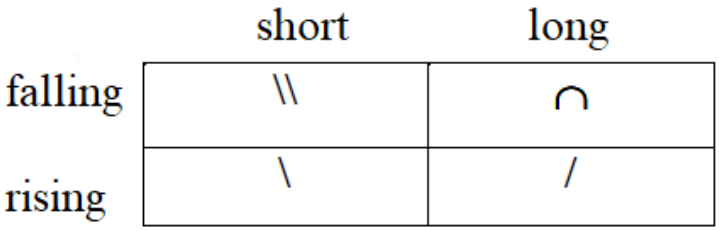


Figure 6. The four accents in BCS (Subotić, Sredojević & Bjelaković 2012: 98).

The complex polytonic system represented above is a direct consequence of the changes in Štokavian since the 15th century, with the set of falling accents inherited from old-Slavic being expanded with a set of newly developed rising accents (Subotić, Sredojević & Bjelaković 2012: 99). Prior to the change, the accents were only differentiated on the basis of quantity (Subotić, Sredojević & Bjelaković 2012: 99). Today, the falling short and long accents are reserved for monosyllabic words and the initial syllables of polysyllabic words, while the long and short rising accents cannot occur in monosyllabic words and final syllables (Simonović & Kager 2020: 309). Despite the existence of a polytonic system, native speakers can easily distinguish between the long and short accented vowels, but fail to do the same with regards to rising-falling qualities of the vowels, which is attested in all three standard varieties (Brown & Alt 2004: 17). As a consequence, "notable attempts have been made to re-interpret the prosodic system and bring it closer to the current linguistic understanding", that is, separate vowel length from pitch and postulate the existence of ten vowels in BCS, divided into long and short pairs (Čubrović 2017: 59-60). This approach, however, is highly dependent on regional variation, as, for example, speakers from Central Serbia and Belgrade differentiate between long and

short vowels, while southeastern dialects in Serbia do not rely on this particular distinction (Čubrović 2017: 59-60). Equally significant is the finding that, for certain vowels, it is quantity, not pitch, that modifies and influences the formant values of vowels /e/, /o/, /i/, and /a/ (Bjelaković 2018: 84-85). Rather than referencing the polytonic system or arguing for the expansion of the vowel space, it would make more sense to refer to such instances as allophones of particular vowels (Bjelaković 2018: 84). Thus, in standard BCS, vowel quantity is not regarded as a prosodic feature, but the differences in length are still easily recognised by speakers, and can be considered as allophonic.

4.2 Vowels /e/ and /a/ in different BCS dialects: A small-scale study

Given that the main focus of this thesis is the perception of English front vowels /e/ and /æ/ by native speakers of BCS, this subsection will examine the production of BCS vowels /e/ and /a/ in different dialects of the language. What is worth noting, however, is that BCS /e/ and /a/ are not exactly equivalents of the English front vowels, as BCS /a/ is a low central vowel, and BCS /e/ is "markedly different from the nearest English approximations", namely /e/ and /æ/ (Vaszolyi 1977: 8). Nonetheless, English /æ/ tends to "be identified with, and substituted by, either BCS /a/ or /e/", hence the common production and perception the word *bat* as /bet/ or /bat/ (Vaszolyi 1977: 9).

4.2.1 Methods and participants

In this small-scale study, the samples were collected from 7 adult native speakers of BCS, from Northwestern Bosnia (Banja Luka), Northeastern Bosnia (Tuzla), Northwestern Croatia (Zagreb), Southern Croatia/ Dalmatia (Split), Northern Serbia (Kikinda), Central Serbia (Smederevska Palanka), Central Montenegro (Podgorica).

The participants were recorded pronouncing primarily monosyllabic words using the recording option in PRAAT (speech-analysis software), at a sampling frequency of 44100 Hz. To eliminate any background noise, mono sounds were recorded, and Jabra Evolve2 40 headset was used. The instructions for pronounciation, pauses, and other relevant information were provided in writing in BCS, and the table with the actual words that needed to be articulated also included the sentence context for each word in question, in order to avoid any confusion. The participants were given time to read each sentence and familiarise themselves with the pronunciation patterns and prepare properly for the recording. In total, 13 words containing either /e/ or /a/ were recorded separately, and are presented with their English translation in the following table:

Table 3. BCS words containing /a/ and /e/.

Words containing /e/	**Translation into English**
led	*ice*
let	*flight*
ego	*ego*
eto	*there (lo)*
sve	*all, everything*
test	*test*
ten	*tan*
Words containing /a/	**Translation into English**
rad	*work*
rak	*crab/cancer*
ada	*river island*
at	*horse*
panj	*log*
zla	*evil* (3rd person feminine)

Aa can be observed in the above table, the vowels in question appeared in different phonological environments: pre-fortis, pre-lenis, pre-nasal, as well as in initial, medial, and final position. The fortis consonants used were /t/, /k/, /s/, or the /st/ cluster, and the lenis consonants were /g/ and /k/, meaning that the vowels appeared before plosives and fricatives. The nasal sounds used were /n/ and /ɲ/ (as in *panj*). Such different contexts were chosen due to their presupposed influence on both quality and quantity on the vowels, based on the already performed analysis of English vowels articulated by a native speaker from Pennsylvania (cf. section 3.2). What also needs to be mentioned, however, is the fact that a large number of highly frequent words containing the vowel /e/ were avoided due to the previously discussed ekavian/ijekavian/ikavian distinction (Alexander 2006: 391). In other words, while *dete* would have been acceptable to speakers of ekavian (primarily Serbian speakers), speakers of ijekavian would pronounce it as *dijete*, and ikavian speakers would articulate *dite*, leading to a difference in both quality in quantity, and, in the case of ikavian, a completely different vowel /i/.

Therefore, the words listed in the table were chosen as not only due to their appropriate phonological environments of /e/ and /a/, but also to avoid confusion due to ijekavian, ekavian, and ikavian pronunciation forms.

Following the data collection, F1, F2, and vowel quantity for each word and speaker were manually extracted from PRAAT and the mean formant values were compiled for each speaker. Finally, the vowels were plotted into a scatter chart, and adjusted to correspond to the vowel approximations in the mouth.

4.2.2 Results and conclusion

In terms of vowel quality, the data gathered from the pronunciation of /e/ by the native speakers in 7 BCS words is plotted in the following chart (for the complete dataset, see Appendix B):

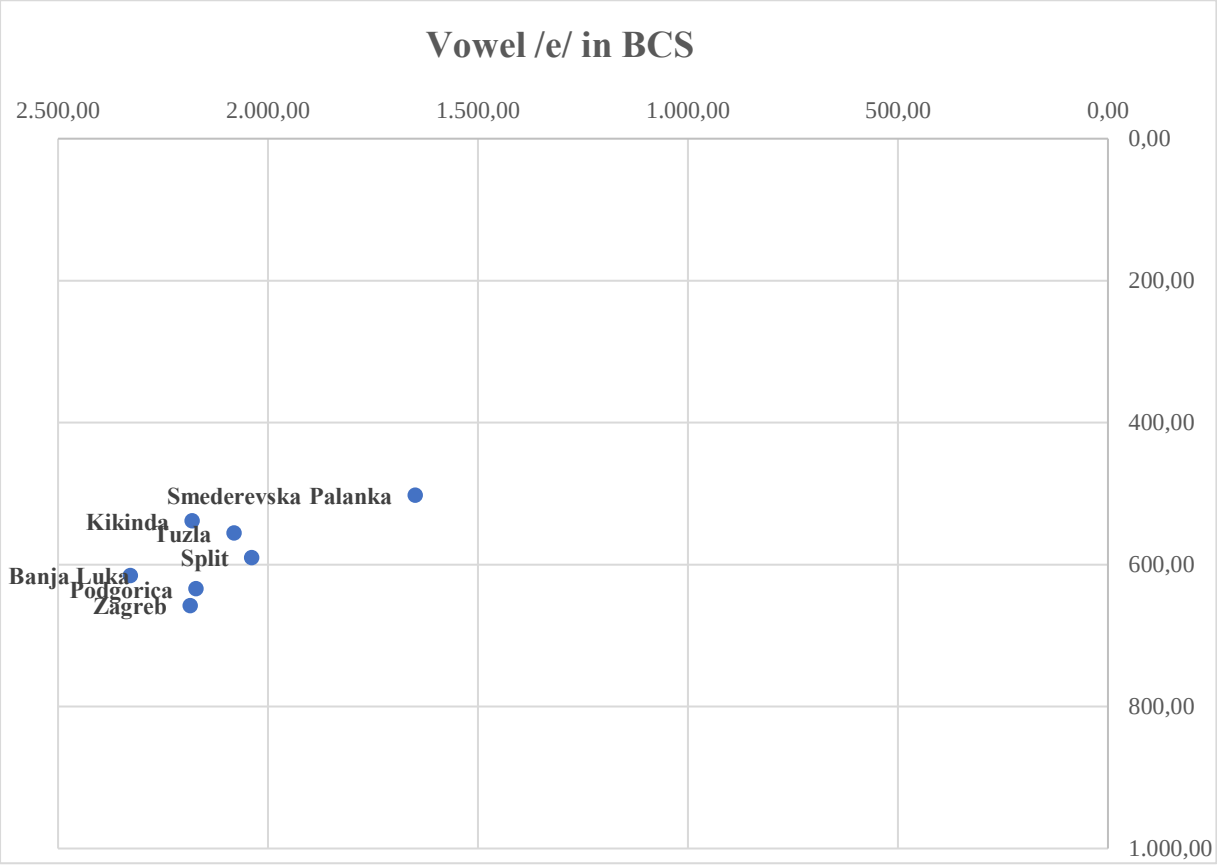


Figure 7. Vowel /e/ in BCS, as pronounced by speakers from different regions. The x-axis shows F2 (2500-0 Hz), while F1 is represented by the y-axis (0-1000 Hz).

The above chart illustrates that BCS vowel /e/ is generally a front mid vowel, which corresponds to the vowel triangle presented earlier (Subotić, Sredojević & Bjelaković 2012: 45). Nevertheless, a considerable degree of regional variation is detected, primarily in reference to tongue fronting, i.e. F2 values. The vowel is considerably fronted in Banja Luka

(Northwestern Bosnia), while the speaker from Smederevska Palanka (Central Serbia) demonstrates a significant degree of tongue backing when /e/ is pronounced, almost approaching the vowel space of BCS /a/. Other regions do not exhibit such a high degree of variation in terms of F2. With regard to F1, a significant degree of tongue raising can be observed in Smederevska Palanka, followed by Kikinda, while the vowel is lowered by the speakers in Podgorica (Central Montenegro) and Zagreb (Northwestern Croatia). Interestingly, both F1 and F2 values in Podgorica and Zagreb are almost identical, despite such physical distance between the regions where the dialects are spoken. When the values of /e/ in BCS are compared to that of its 'equivalent' in English (cf. section 3.2), only Banja Luka and Smederevska Palanka do not have the F2 values approximate to the F2 values found in Pennsylvania. Moreover, BCS /e/ is generally more raised than its English 'counterpart', with only Podgorica and Zagreb F1 values slightly approaching the F1 of the native speaker of English. In terms of the English vowel /æ/ and BCS /e/, the vowels are too spectrally distant to make any comparisons.

Furthermore, the F1 and F2 values of /a/ in 6 BCS words are plotted in the following chart (for the complete dataset, see Appendix B):

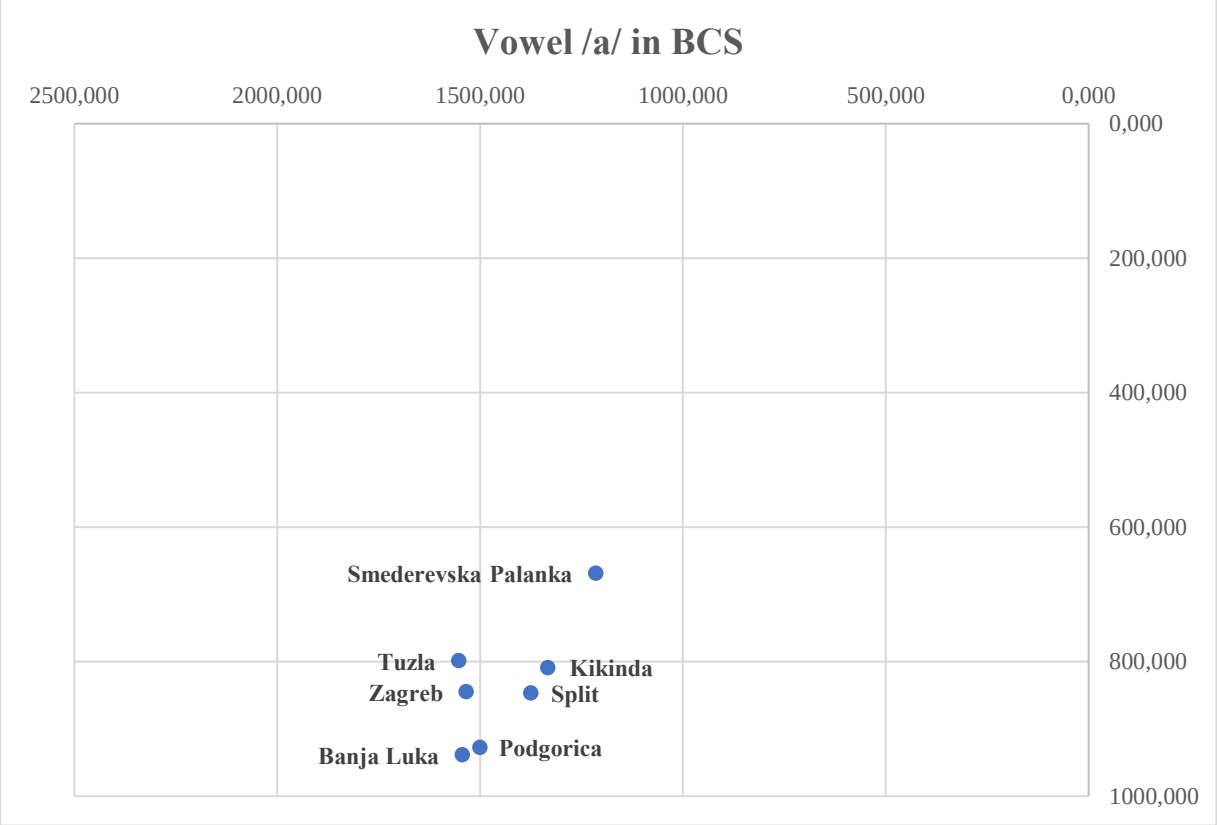


Figure 8. Vowel /a/ in BCS, as pronounced by speakers from different regions. The x-axis shows F2 (2500-0 Hz), while F1 is represented by the y-axis (0-1000 Hz).

The BCS /a/ is generally a low central vowel, which coincides with the vowel triangle taken from Subotić, Sredojević and Bjelaković (2012: 45). Regional variation is mostly observed in the speaker from Smederevska Palanka (Central Serbia) pronouncing the vowel as considerably raised and backed, in contrast to speakers from other regions. Moreover, the speakers from Split (Southern Croatia) and Kikinda (Northern Serbia) demonstrate a slight tendency towards tongue backing when producing the vowel, but their F1 remain similar to other speakers. The highest degree of regional similarity is evidenced in Banja Luka (Northwestern Bosnia) and Podgorica (Central Montenegro), where only a small difference in F2 prevents almost a complete spectral overlap. Equally interesting is the fact that the difference in F1 between the aforementioned regions is less than 10 Hz. When compared with English vowels /e/ and /æ/, Tuzla, Zagreb, Kikinda, and Split have the F1 values approximate to that of English /e/, while the lowering of the vowel in Banja Luka and Podgorica is nearing the values of /æ/. BCS vowel, however, is more centred than the English vowels.

With regards to vowel quantity, the following table was introduced to illustrate the mean duration of both BCS /e/ and /a/:

Table 4. Vowel quantity of BCS /e/ and /a/, as pronounced by native speakers from different regions.

City (Region)	Duration of /e/ in /s	Duration of /a/ in /s
Banja Luka (Northwestern Bosnia)	0,182	0,197
Tuzla (Northeastern Bosnia)	0,160	0,167
Zagreb (Northwestern Croatia)	0,163	0,347
Split (Southern Croatia/Dalmatia)	0,200	0,247
Podgorica (Central Montenegro)	0,161	0,156
Smederevska Palanka (Central Serbia)	0,143	0,152
Kikinda (Northern Serbia)	0,220	0,208
All regions	0,176	0,189

The data on vowel quantity suggests that /e/ is slightly shorter than /a/, but that the durational differences between the vowels differ from region to region. The only exceptions are the speakers from Kikinda and Podgorica, whose /e/ sounds are longer than /a/ in the words pronounced. Focusing solely on the inter-regional variation of /e/, the highest degree of regional difference is observed between the north and centre of Serbia: the vowel is pronounced

significantly longer by the speaker from Kikinda. In addition, the vowel /a/ tends to be shorter in Central Serbia and Central Montenegro, as well as Northeastern Bosnia, but it is longer in Northwestern Bosnia, Northern Serbia and both Croatian regions, in particular. With regards to the changes in quantity in pre-fortis and pre-lenis positions, /a/ undergoes clipping before fortes, but its duration is longer before lenes consonants, while /e/ exhibits none of the effects of the following consonants on its duration. For instance, /a/ is quite longer in *rad* than in *rak* (t=0,232 s vs. t=0,148 s), but /e/ in *led* is shorter than in *let* (t=0,186 s vs. t=0,211 s). Finally, in contrast to Pennsylvanian /e/ (t=0,115 s), both BCS /e/ (t=0,176 s) and /a/ (t=0,189 s) are longer. Nonetheless, English /æ/ (t=0,196 s) is longer than both BCS /e/ and /a/.

Therefore, the BCS vowel /e/, as pronounced by native speakers, can be described as a front mid vowel, whereas /a/ is a central low vowel. A relatively high degree of variation among the speakers from different regions in Bosnia, Croatia, Montenegro, and Serbia is observed in both vowels in terms of their F1 and F2 values. However, some of the regions analysed demonstrate a high degree of similarity, despite physical distance between the areas where the dialects are spoken. If the values of the BCS vowels are compared to the English /e/ and /æ/, there are minor similarities in either F1 or F2 detected between some of the BCS dialects and Pennsylvanian vowels. In terms of vowel quantity, /a/ is longer than /e/, with apparent regional differences. Both vowels have a longer duration than the English /e/, but are shorter than its pair /æ/.

5. ELF versus EFL in acquisition of L2 phonology

Obtaining a near-native accent has been continuously regarded as the pinnacle of one's proficiency in English by those adhering to the EFL principle in language teaching (Thir 2014: 36). This so-called 'nativeness principle' has shaped the idea that the impact of one's L1 is highly detrimental, and any problems with perception and articulation of target phonemes, as well as miscommunication, are seen as a consequence of L2 repertoire being 'contaminated' with source phonemes (Jenkins 2009: 141). Nevertheless, the spread of English globally has lead to non-native speakers outnumbering the native speakers, and many have started disputing the need to appropriate and adhere to NS norms, as the large number of non-native speakers of English has slowly shaped English into Englishes (Wang 2013: 256). Nowadays, the focus has slowly been shifting from the native-like correctness towards the speakers' own desires and the intelligibility principle (Thir 2016: 5). The former refers to the idea that, in using English as a Lingua Franca, speakers have the freedom to choose to appropriate the native speaker

accent or retain features of their L1, or combine L1 and L2 phonology in their speech, allowing them to "achieve equilibrium between identity and intelligibility in English pronunciation" (Thir 2014: 64). The latter stresses the need for "functional and communicative" intelligibility in communicative situations, rather than pressuring the learners to emulate the accent of native speakers (Wach 2011: 248). Based on numerous miscommunication and accommodation data samples, Jenkins (2002: 96) has created a list of targets needed for successful communication, known as the Lingua Franca Core:

	EFL targets	ELF targets
1. The consonantal inventory	• all sounds	• all sounds except /θ/, /ð/ and [ɬ]
	• RP non-rhotic /ɹ/ • GA rhotic /ɾ/	• rhotic /ɹ/ only
	• RP intervocalic [t] GA intervocalic [ɾ]	• intervocalic [t] only
2. Phonetic requirements	• rarely specified	• aspiration after /p/ /t//k/ • appropriate vowel length before fortis/lenis consonants
3. Consonant clusters	• all word positions	• word initially, word medially
4. Vowel quantity	• long-short contrast	• long-short contrast
5. Vowel quality	• close to RP or GA	• L2 (consistent) regional qualities
6. Weak forms	• essential	• unhelpful to intelligibility
7. Features of connected speech	• all	• inconsequential or unhelpful
8. Stress-timed rhythm	• important	• does not exist
9. Word stress	• critical	• unnecessary/can reduce flexibility
10. Pitch movement	• essential for indicating attitudes and grammar	• unteachable/incorrectly linked to NS attitudes/ grammar
11. Nuclear (tonic) stress	• important	• critical

Figure 9. The Lingua Franca Core, as proposed by Jenkins (2009: 142-143). The left column shows EFL, while the right column shows EFL targets in phonology.

The above figure suggests that, in terms of vowel quality, ELF speakers only need to consistently produce the vowel "features of the respective L2 regional accent", which allows them to preserve their own L1 ethnic identity while speaking English (Jenkins 2009: 143). Vowel quality substitutions from L1 are recognised as aspects of NNS regional accents only, and not analysed as errors (Jenkins 2002: 97). In other words, vowel quality in ELF is a non-essential feature, as long as pronunciation remains constant, as not to impede intelligibility of speech (Newbold 2021: 399). Together with other non-core features, vowel quality allows "L2 English speakers the same sociolinguistic rights as those enjoyed by L1 speakers", thus 'legitimising' non-native accents and endorsing the idea of Englishes (Zoghbor 2011: 287). Therefore, acquiring ELF phonology and focusing mostly of the essential features of the Lingua Franca Core validates the speakers' L1 accents, without endangering the expression of

their personal and ethnic identities, and stresses mutual intelligibility as a key factor in (international) communication, and, finally, is more attainable than a "close approximation to a NS pronunciation model" (Thir 2016: 13). More importantly, ELF phonology provides speakers with a choice, meaning that the speakers can still opt for RP or GA, but they should not be pressured (especially in education) into it (Jenkins 2009: 143).

5.1 Conformity to the native-speaker ideal in SLA of phonology

The spread of English as a lingua franca has not only resulted in researchers questioning the perceived supremacy of native-speaker norms, but has also generated interest in the attitudes to NS norms and the desire to conform to the NS ideals by the "followers of exonormative English", i.e. speakers from the Expanding Circle countries (Wang 2013: 259). Generally, the decades of dominance of EFL standards in language teaching have lead to positive attitudes to native-speaker norms and an apparent preference for native-speaker accents (Wach 2011: 250). In an international study, involving approximately 600 participants from 45 countries, Timmis (2002: 242) has discovered a general preference for native-speaker norms, and that attaining a native-speaker accent is considered "a benchmark of achievement". This idea seems to prevail regardless of the communicative context, since the participants emphasised the importance of sounding like a native speaker when communicating both with native and non-native speakers of English (Timmis 2002: 248). Timmis (2002: 248) concludes that, despite the apparent motivation to communicate in English, speakers still seem to cling to the "traditional idea of mastering a language", that is, being NS-like at all levels of language.

Similar results can be observed in a study by Wach (2011: 260), who found that native-speaker norms were viewed in a positive light by selected groups of English majors in Poland. For example, the statements defending the retainment of the native Polish accent in English speech have been rated rather negatively by both groups examined, while the statements praising those striving for a NS accent were rated favourably (Wach 2011: 256). Nevertheless, there is a significant degree of variation among the participants, as the group of students who took pronunciation classes voiced more positive opinions of EFL-oriented statements than the other group. Equally important is the fact that the educational context and career goals had an immense effect on the answers, as all the participants were English majors, whose ultimate attainment of the native-like accent is expected to reflect positively on their careers (Wach 2011: 260). Interestingly, both groups deemed native-speaker norms slightly less relevant in ELF settings (Wach 2011: 260). Therefore, English majors, depending on instruction in phonology and career goals, are generally in favour of NS norms, but they admit that sounding

like a native English speaker in international settings may not be a key factor in successful communication.

A higher degree of variation among participants was unveiled by Tengs Sannes (2013) in a survey involving high school students in Norway. On one hand, the Norwegian students who opted for a native-speaker accent did so because it was considered "real English" and more comprehensible, particularly for native speakers of English (Tengs Sannes 2013: 80). On the other hand, those who did not strive to speak English with a native-like accent either did not even think about their own accent or had negative attitudes to the idea of sounding like a native speaker of English, since it was "unnatural and unnecessary" (Tengs Sannes 2013: 81). For those students, mutual intelligibility was in the forefront of successful communication in English (Tengs Sannes 2013: 96). Apart from the participants' own desires to conform to NS norms, their judgements of different English varieties were investigated. Participants rated RP (standard British English) as "intelligent, polite, successful, rich, and nice", and GA (standard American English) was "easy to understand, cool, normal, and nice", while heavily accented English produced by a Norwegian speaker was "ugly, weird, unintelligent, unsuccessful, difficult to understand", and slightly accented English by a Norwegian speaker was generally viewed in a positive light (Tengs Sannes 2013: 92-93). Such opinions may have been influenced by the years of prolonged exposure to the near-native speech of the teachers, hence the association between standard English varieties, and intelligence (Tengs Sannes 2013: 93).

5.2 A review of previous studies in the context of Bosnian/Croatian/Serbian

Previous studies of the perception and production of English vowels by scholars from Bosnia, Croatia, and Serbia have a rather strong bias towards native-speaker norms (cf. Marković 2009a; Marković 2009b; Marković & Jakovljević 2016; Krebs-Lazendić 2008; Krebs-Lazendić & Best 2007; Paunović 2011). This subsection serves as a review of such studies and their strict EFL perspective, exemplified in the choice and characterisation of their participants, and their representation of the L1 interference as highly inimical.

What needs to be mentioned first is that the studies in question mostly research perception and production of English vowels by English majors with a higher degree of exposure to English phonology than the general population. For example, both studies by Marković (2009a; 2009b), investigated production and perception of English vowel contrasts by first year students of English in Serbia. In one of the studies (Marković 2009b: 255), the prerequisite for participating was familiarity with IPA and English phonology, as the participants were asked

to select an IPA symbol corresponding to the vowels they heard. In other words, there was a strong correlation between the knowledge of English phonology plus the IPA and the experimental design. Moreover, a study by Marković and Jakovljević (2016) also emphasised B2 or higher English language knowledge of their participants. Despite the fact that such choice of participants may be a matter of convenience, methodology, or simply due to the relevance of production and perception for those studying English at university level, it is plausible to assume that the focus group in these studies cannot possibly be a fair representative of the general population. In contrast, a study by Krebs-Lazendić and Best (2007) and a subsequent doctoral thesis of Krebs-Lazendić (2008) have both examined vowel perception and production by two groups of speakers on the basis of their English language experience and length of residence in Australia, namely early and late learners. Interestingly, both of these studies use the terms "learner" and "bilingual" interchangeably to refer to their participants, particularly those that started learning English in adulthood. Likewise, the papers by Marković (2009a; 2009b) and Marković and Jakovljević (2016) consistently characterise their participants as "learners", notwithstanding their prolonged exposure to English, especially in university settings. By being negatively portrayed as perpetual learners, participants are being denied "the same sociolinguistic rights as those enjoyed by L1 speakers" (Jenkins 2009: 143). Therefore, even the highly experienced speakers remain in the perpetual learning phase.

Finally, the deviant production and perception of vowels stems from the need to examine the degree of L1 interference, which is always disadvantageous and heavily stigmatised. In order to investigate the influence of the source language, both L1 and L2 speech samples are obtained and their quality and quantity are compared, as observed in Marković's recording of both English high vowels /ɪ/, /i:/, /u:/, /ʊ/, and BCS /i/ and /u/ (2009a: 3). In the studies of vowel perception, predictions are based on either acoustic or phonemic properties of vowels in the L1 (Marković & Jakovljević 2016). Similarly, Krebs-Lazendić and Best (2007: 282) base their hypotheses about vowel perception and production on the agreement that the "failure in non-native segmental production and perception in caused by previous linguistic experience with the first language". With reference to the effects of L1, the term 'interlanguage' is widely used to signal that the phonological repertoire is in-between BCS and English, further supporting the idea that the participants are in the state of perpetual learning (Marković 2009a; Marković 2009b; Marković & Jakovljević 2016). Although Paunović (2011) claims that their paper does not explore the degree of L1 interference in the articulation of English vowels by BCS speakers, the analysed vowel system is still described as 'interlanguage'. The influence of the source

language on either production or perception of English vowels, therefore, is typically characterised as inimical and the system developed during the learning process is referred to as 'interlanguage'.

6. The present study design

The present study investigates the perception of English quality contrast /e/ and /æ/ by native speakers of Bosnian/Croatian/Serbian. Moreover, the participants' exposure to English, as well as their attitudes to native-speaker norms are examined. This section discusses the research questions and predictions, experimental design, participants, both auditory perception tasks and statement ratings, as well as the testing procedure.

6.1 Research questions and hypotheses

Based on the theoretical concepts discussed in previous sections, the present study seeks to answer two research questions, one phonological/phonetic (main) question and one sociolinguistic (additional) question:

1. How do native speakers of Bosnian/Croatian/Serbian (BCS) perceive the English /e/ versus /æ/ contrast, since these two vowels only have a single native 'counterpart' /e/? How can the perception of this vocalic contrast be explained by Best's Perceptual Assimilation Model (PAM)?
2. Given the emphasis on L2 exposure in PAM, what is the relationship between this sociolinguistic variable and the perception of vowels? Moreover, how do attitudes to native-speaker norms in SLA of phonology relate to the perception of the English vowel contrast?

For the first question, PAM suggests three possible answers, each corresponding to one assimilation pattern, and ranked from most to least likely:

1. Category-Goodness assimilation/CG type (Best & Tyler 2007: 23): discrimination between /e/ and /æ/ is expected to be poor-to-intermediate, depending on the phonological context of the stimuli (pre-lenis, pre-nasal, pre-fortis). Both sounds are perceived as instances of BCS /e/, but English /e/ is judged as a better exemplar of the native category. On the other hand, /æ/ may be more deviant, due to its increased quantity and changes in quality (cf. section 3.2). In that sense, English /e/ and /æ/ are assimilated into BCS /e/ on a continuum, where English /e/ is closer to the quality of the native phoneme than /æ/, and therefore rated as a better candidate for the native /e/.

The vowel /æ/ would need to receive high similarity ratings with BCS /e/, but not higher than English /e/.

2. Single-Category assimilation/SC type (Best 1995: 195): a great deal of confusion between /e/ and /æ/ is observed across all phonological contexts and tokens, i.e. the minimal pairs are mixed up, and both sounds are perceived as equally good or bad examples of BCS /e/. In all the words containing /e/, the vowel is heard as /æ/, and vice versa. Moreover, both English /æ/ and /e/ would need to be rated as highly similar to BCS /e/, or both receive extremely low category goodness ratings. Different phonological contexts, however, make this prediction unlikely, given the particularly changeable quality and quantity of /æ/ before nasals, lenes, and fortes (cf. section 3.2).
3. Uncategorised-Categorised/ assimilation/UC type (Best & Tyler 2007: 23): the phonological difference between /e/ and /æ/ is clearly perceived, and the optimal discrimination between the vowels is facilitated by the fact that /e/ is assimilated into the BCS native category, while /æ/ is assimilated somewhere in the phonetic space, outside of the L1 category. In other words, distinguishing one English vowel from another is essentially discrimination between one native and one foreign sound. Nonetheless, in order to make this prediction possible, /e/ and /æ/ would need to be perceived successfully across all phonological contexts, and /æ/ would need to be perceived as a deviant example of BCS /e/, which is highly unlikely, but still hypothesised. Another problem with the UC type is the fact that the assimilation of /æ/ somewhere in the phonetic space in L1, outside of BCS /e/, would be extremely difficult to demonstrate and interpret.

Furthermore, the above predictions are dependent on the phonological context of the English phonemes. As described above (section 3.2), both /æ/ and /e/ have a variety of different phonetic realisations that are conditioned by the phonological environments in which they occur. In pre-nasal position, /æ/ is significantly raised, while /e/ is lowered, resulting closeness in terms of F1 and F2. As a consequence, a high degree of variation and confusion are expected in the answers of BCS listeners, who are likely to rely on temporal cues, because /æ/ is generally longer than /e/ before nasals (cf. Table 2). Moreover, vowel identification is expected to be better in pre-lenis than in pre-fortis environments. Given that the duration of /æ/ is reduced before fortes in the tokens analysed in section 3.2, it is plausible to assume that differentiating between /æ/ and /e/ would be more difficult in such contexts. Nevertheless, both vowels, /e/ in particular, exhibit a high degree of word-to-word variation. In contexts where realisations of

/æ/ and /e/ do not differ in terms of duration, quality will be a more reliable indicator of vocalic disparity.

Since the target group are non-majors of English who have had little to no knowledge of linguistics or phonology of English, there are two possible predictions for the second question. On one hand, those who have had more extensive exposure to English in primary and secondary education or extracurriculars (EFL English), and were/still are heavily exposed to English, be it media, work, spoken interaction, are expected to distinguish between /æ/ and /e/ on a poor-to-intermediate basis, but still fall under the CG pattern of assimilation (i.e. rate both vowels as good exponents of BCS /e/, but English /e/ is a better candidate). Such participants are also highly likely to exhibit slightly positive attitudes to NS norms, due to years of prolonged EFL instruction. On the other hand, the participants who had relatively limited contact with L2 through formal instruction (primary school, high school, or both), or learned English solely through exposure to TV, music and other media, are likely to assimilate both English vowels as equally good exponents of BCS /e/ (Single-Category assimilation), and have great difficulties with discriminating between /æ/ and /e/, especially in contexts where the acoustic differences between the vowels are 'neutralised' by the following consonants (e.g. the closeness in terms of F1 and F2 before nasals in section 3.2.). Given their moderate-to-minimal experience with EFL, it can be assumed that these participants will express highly negative attitudes to native-speaker norms. Nonetheless, a high degree of variation is anticipated in the answers pertaining to native-speaker norms. Section 9 of this thesis will be a general discussion of the phonetic and sociolinguistic parts of the experiments, and provide a number of individual examples of the interrelatedness of assimilation patterns, English language experience, and the attitudes to native-speaker norms.

6.2 Experimental design

With the purpose of answering the research questions and testing the hypotheses, an experiment was designed using QuestionPro, an online survey software. This particular software was chosen for a number of reasons. First, it is extremely user-friendly, it allows for a more interesting visual representation of questions, and ensures that the participants are able to complete the tests and questions thoroughly. Secondly, since it is online, it can be easily distributed to the target group of native speakers of BCS in their respective countries, without the need for the author to be physically present during data collection. Moreover, the tests can be completed on mobile phones. What is perhaps most important, QuestionPro enables the integration of acoustic stimuli (sounds) into questions, which is the core of this experiment.

Furthermore, it offers question types that are necessary for testing minimal pair distinctions and proper category goodness ratings. Because of this, combining different tasks/questions is feasible, rather than conducting separate experiments. Finally, the software automatically analyses data for each question/token and generates charts (4 different charts are offered), together with the computing of mean, variance, standard deviation, standard error, and confidence intervals, and therefore no additional statistical software/tool is needed for data analysis and representation, which is extremely useful when working with a larger number of answers.

To collect both phonetic/phonological and sociolinguistic data, 5 parts were created:

1. Task 1: perception of words in isolation containing /æ/ and /e/ (i.e. minimal pair distinction)
2. Task 2: category goodness rating i.e. similarity rating between English vowels and BCS /e/
3. Open questions about the participants (basic information)
4. Rating of statements about the exposure to English
5. Rating of statements pertaining to NS norms

The design of each of the parts will be discussed thoroughly in the following subsections, with the exception of open questions about the participants, as its results will be discussed immediately in the subsection that follows.

6.3 Participants

Apart from the perception tasks and statement ratings, the participants were asked to provide basic information about their country and region, age, the use of English and English accents in education, the time spent in English-speaking countries, as well as if they had taken any phonology courses. The questions have shown that the participants are adult native speakers of Bosnian/Croatian/Serbian (N=87; 18-47 years), and are well-dispersed throughout different regions in their respective countries. When their numbers are taken into consideration, the majority of participants are from Serbia (n=33), followed by Bosnia and Herzegovina (n=31), Croatia (n=19), and 2 respondents are from Montenegro.

With regards to their educational background and prior experience with English (for the complete dataset, please refer to Appendix E), the vast majority of participants (n=82) reported learning English in primary school and high school, meaning that they have had extensive exposure to EFL in the classroom. Apart from formal, mandatory education, a number of

participants attended private lessons in English. Learning English through exposure to music, films, and books was also reported, however in combination with formal instruction and private English lessons. Furthermore, the English variety/accent that was used most in education was British/RP (n=40), followed by American (n=13), and 22 participants indicated that both varieties were used. In addition, only a small number of participants has spent more than 6 months in countries where English is the *de facto* language (14 months, 4 years, 7 years), meaning that these participants may be more attuned to native English accents and may distinguish between the phonemes more successfully, as opposed to those that were not physically immersed in English varieties.

Finally, what is most significant in the present study, none of the participants have had any instruction in phonology and phonetics or have taken pronunciation classes, meaning that the perception of English vowels should not be influenced by their pre-existing familiarity with English phonology. According to Vilke (1991: 487) and Naser (2011: 4-5), in the absence of phonetic training, the chances of distinguishing between the phonemes are rather small, and may lead to students not focusing properly on developing listening and speaking skills. In other words, the participants are highly unlikely to recognise the difference between /æ/ and /e/ because they are unaware that the difference exists, and will rely on either spectral or temporal cues or both. Therefore, phonological awareness as a consequence of formal instruction is not observed, and should not have an impact on the findings. However, this point only refers to phonetic training per se, not on the possible guidelines provided by the teacher in the classroom (e.g. the teacher explaining that *man* and *men* have different vowels), which could lead to some phonological awareness.

6.4 Task 1: Auditory perception of words in isolation/ minimal pair distinction

The first phonetic/phonological task was essentially a minimal pair distinction test, where the participants heard a word in isolation containing either /æ/ or /e/ and had to select one of the four images that best corresponds to the word they heard. One of the remaining images was of the word's minimal pair (the words differed only in a single vowel, and if the participants heard *pen,* there was also an image of a *pan*), while the others were visual distractions. Since assimilation patterns outlined by Best (1995) correlate to discrimination possibilities between sounds, minimal pair distinction was introduced with the purpose of determining the exact pattern of assimilation (and if assimilation occurs at all) on the basis of this proposed link between perception and assimilation. If the discrimination between /e/ and /æ/ is poor-to-intermediate, but /æ/ stands out more in perception, then the sounds are assimilated into BCS

/e/ on the basis of the Category-Goodness type. The tendency to confuse /æ/ with /e/ and vice versa across all phonological contexts and tokens would indicate a Single-Category assimilation. The optimal perception of difference between the vowels could be facilitated by UC type assimilation, as it is essentially a distinction between an L1 and L2 sound.

For this task, 24 words/12 minimal pairs containing the /æ/-/e/ quality contrast were produced by a female native speaker from Doylestown, Pennsylvania (north of Philadelphia). All stimuli were recorded as mono sounds in PRAAT (speech-analysis software), at a sampling frequency of 44100 Hz. The pairs and their phonological environments are listed in the following table:

Table 5. Minimal pairs/words used in Task 1, together with their phonological context.

Phonological environment	**Minimal pair containing /æ/**	**Minimal pair containing /e/**
Pre-nasal	*pan*	*pen*
	jam	*gem*
	ban	*Ben*
	man	*men*
Pre-lenis	*bag*	*beg*
	paddle	*pedal*
	dad	*dead*
	Brad	*bread*
Pre-fortis	*flash*	*flesh*
	axe	*x*
	cattle	*kettle*
	bat	*bet*

As observed, there were 4 occurrences of a single vowel in three different environments to ensure more optimal results and later analysis. The vast majority of tokens are monosyllabic words. The only exceptions are disyllables *cattle, kettle, pedal,* and *paddle*. Furthermore, the vowels appeared before the nasal consonants /m/ and /n/, plosives /g/, /d/, /k/ and /t/, and the fricative /ʃ/. In terms of place of articulation i.e. point of airstream constriction, the consonants following /æ/ and /e/ were alveolars /n/, /d/, /t/ and its allophone [ɾ], velars /g/ and /k/, bilabial /m/, and palato-alveolar /ʃ/.

Given that the vowels have already been analysed in section 3.2 in a greater detail, this subsection will only summarise the features of /æ/ and /e/ and provide the vowel charts, as well as the vowel quantity table (for the complete table of F1 and F2 and duration see Appendix A):

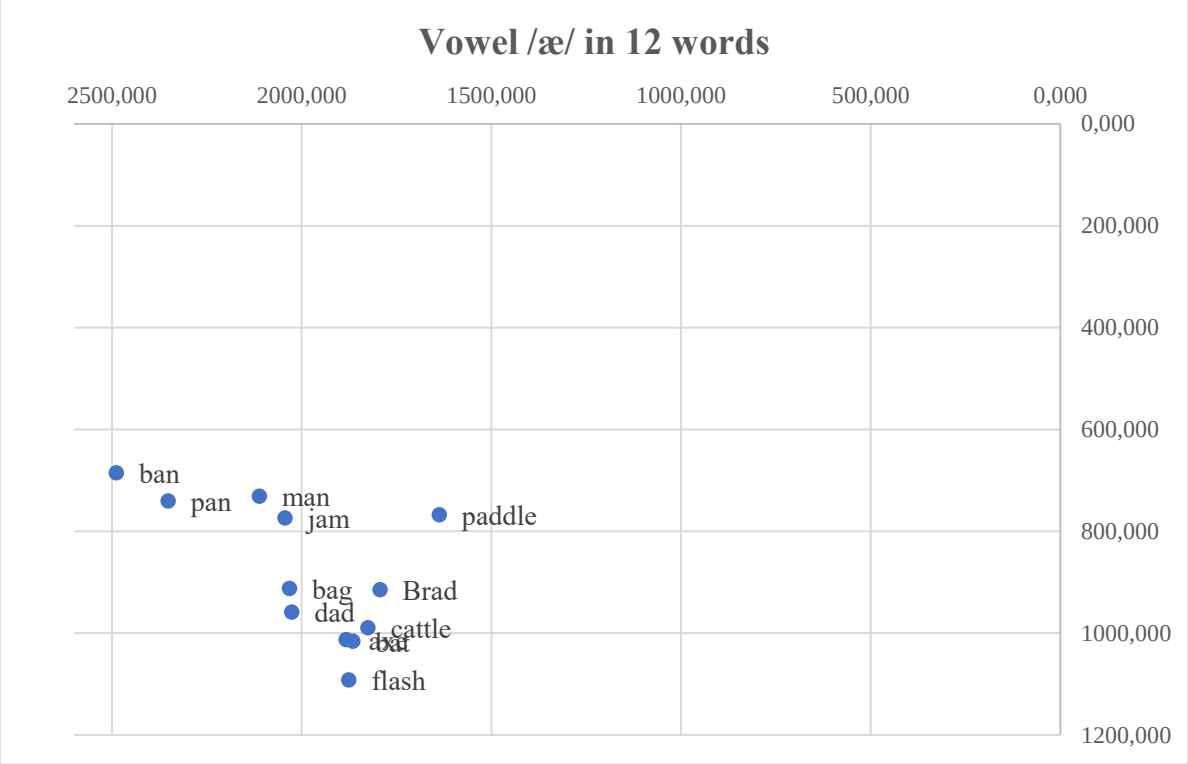


Figure 10. Mean F1 and F2 of the vowel /æ/ in 12 words, as produced by a native speaker from Doylestown, Pennsylvania. The x-axis shows F2 (2500-0 Hz), while F1 is represented by the y-axis (0-1200 Hz).

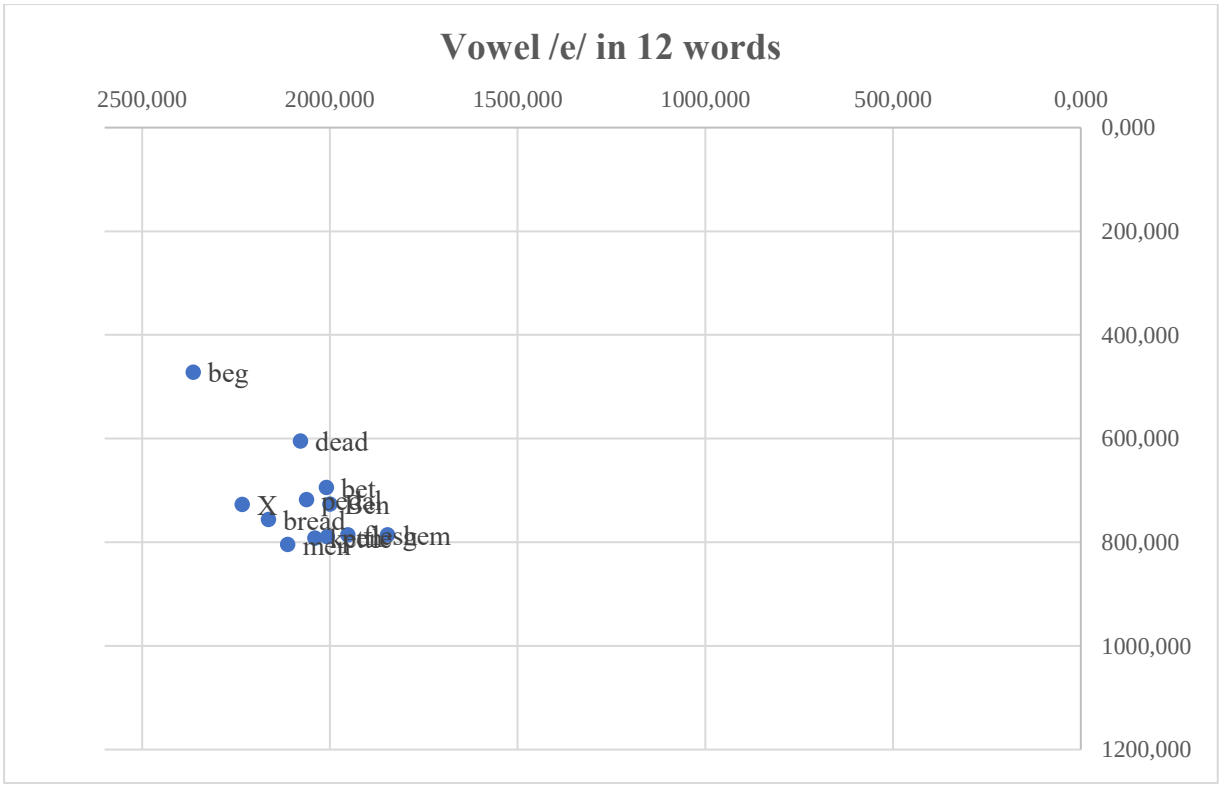


Figure 11. Mean F1 and F2 of the vowel /e/ in 12 words, as produced by a native speaker from Doylestown, Pennsylvania. The x-axis shows F2 (2500-0 Hz), while F1 is represented by the y-axis (0-1200 Hz).

Table 6. Vowel quantity of /æ/ and /e/ in 12 minimal pairs pronounced by a native speaker from Doylestown, Pennsylvania.

Duration of /æ/		**Duration of /e/**	
Words	Duration in /s	Words	Duration in /s
pan	0,230	*pen*	0,140
bag	0,279	*beg*	0,221
bat	0,104	*bet*	0,122
ban	0,224	*Ben*	0,122
dad	0,283	*dead*	0,141
flash	0,185	*flesh*	0,134
axe	0,141	*X*	0,112
jam	0,224	*gem*	0,055
paddle	0,120	*pedal*	0,059
Brad	0,220	*bread*	0,097
man	0,217	*men*	0,123
cattle	0,120	*kettle*	0,058
Mean	0,196	Mean	0,115

In the 24 tokens/12 minimal pairs articulated by a native speaker from Pennsylvania, quality and quantity of /e/ and /æ/ are impacted by the following consonants to varying degrees. Both quality and quantity of /æ/ are affected in pre-lenis, pre-fortis, and pre-nasal positions. In terms of quality, the vowel is raised and fronted before nasals, slightly raised or remains constant before lenes (the frontness varies), and lowered in pre-fortis positions. The vowel is lengthened before lenes and nasals, while shortened before fortes. The same degree of variation in vowel quality of /e/ was not detected in the 12 words examined, with the exception of the diphthongisation in *beg*, the raising in *dead*, as well as the slight lowering in pre-nasal position, causing a potential overlap with the raised /æ/ before nasals. The length differences of /e/ demonstrate an extreme degree of word-to-word variation, and the exact patterns of vowel

shortening and lengthening cannot be determined. It is only certain, with the exception of *bet* versus *bat*, that /æ/ is longer than /e/ in the acoustic data provided by the native speaker. In order to successfully distinguish between /æ/ and /e/ the listeners will either rely on spectral or temporal cues, or both, depending on the words in question. In addition, the charts and the table demonstrate why the alveolar flap [ɾ] in *cattle* and *kettle* is grouped with the fortes. Despite it being voiced, meaning that it should be listed with lenis consonants, the vowels preceding it are spectrally closer to their quality in pre-fortis position, rather than pre-lenis (Plag et al. 2015: 50). In terms of temporal cues, vowel quantity for both vowels is significantly shorter before [ɾ] than in pre-lenis position.

With regards to the task design and procedure, each word was tested on separate page. The participants heard a word, and were instructed to choose the image that best corresponds to the word they heard. The instructions were provided in English before the very first word was presented and were also translated into BCS and placed next to the audio player (marked with "?", see Figure 12). Task 1 is illustrated in the following image:

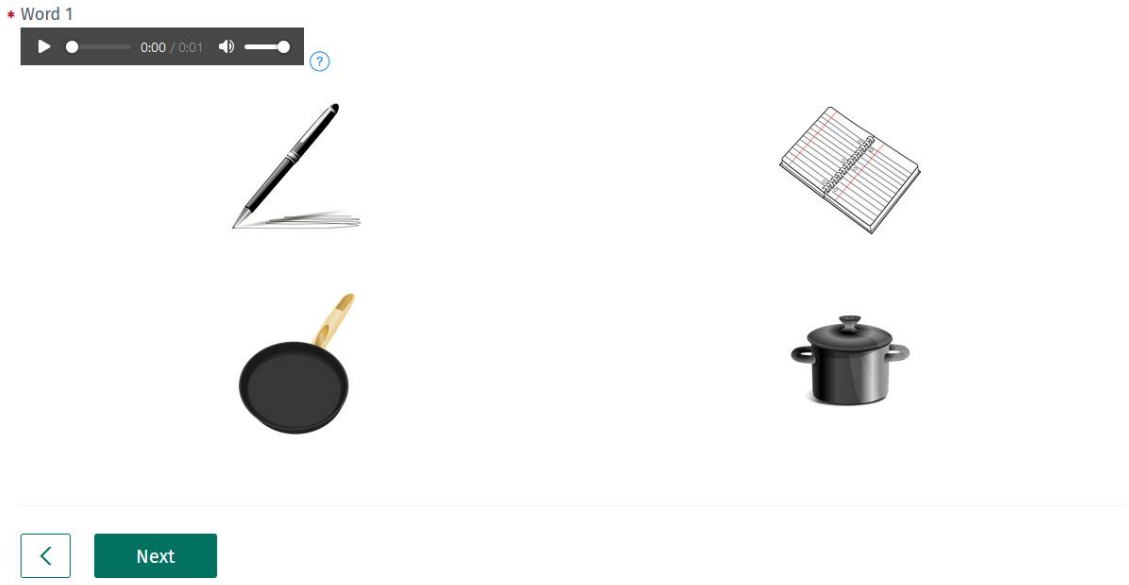


Figure 12. The visual representation of the first task.

The above example shows the first word in the sequence, which is *pan.* The participants were presented with a clipart of the object, as well as the clipart of its minimal pair *pen*. The other two images belonged to the similar semantic domain as the minimal pair, and therefore served as visual distraction stimuli, to ensure that the participants could perceptually isolate the minimal pair. In each of the trials, the solutions were presented in a randomised order, computed by the software. The words were only represented acoustically and visually, and

there was no orthography involved, as to prevent the association between the grapheme and the phoneme, and, with it, the impact of the native orthography (Havelka & Rastle 2005: 150, cf. section 4).

With reference to the visual stimuli, only simple images that best correspond to the words uttered and that show only the relevant shapes without any background were chosen (for the complete list of the clipart, see Appendix C). All images were taken from OpenClipart, a website where the artists release their rights to the public domain, and their art can be used both privately and commercially (a similar disclaimer was provided at the end of the whole experiment, together with the licence). What needs do be mentioned, however, is the fact that selecting the images that best correspond to the words containing the examined vowels proved to be a rather strenuous task, and, as a consequence, a number of words/minimal pairs that were initially picked were discarded. While for the words such as *bag*, finding the appropriate image was quite easy, doing the same for its minimal pair *beg* was more difficult, since *bag* as a noun is more concrete and straightforward, as well as significantly easier to represent visually than a verb. The possible effects of the visual stimuli will be taken into consideration while reporting the results of this task.

Given the high number of tokens in Task 1, the whole experiment was divided into two smaller ones. Only the acoustic stimuli were changed, but the visual images and the subsequent tasks remained the same. Each participant was never exposed to both members of the minimal pair (e.g. they heard either *pan* or *pan* in the first sequence and had to choose one of the images). There are several advantages to this approach:

1. It avoided phonological priming effects, i.e. if the participants heard both *pan* and *pen* and were given the same images, they would probably establish a link between the image and the word that first appears in the sequence (Radeau, Morais & Deqier 1989: 525).
2. It ensured that the participants would complete all the tasks and remain focused, because they only had to identify 12 rather than 24 words.
3. It diminished the chances of the participants developing sensitivity to the difference between /æ/ and /e/ by being exposed to more acoustic stimuli.
4. It warranted a more reliable testing of the hypothesised PAM patterns, SC, CG, or UC (Best & Tyler 2007: 23). For instance, if both [pæn] and [pen] were paired with the same visual stimulus, be it the image of a pan or a pen, this would be an attestation of the assimilation into a single native category.

Finally, the predictions for Task 1 in terms of assimilation patterns in PAM and the phonological environments in which the vowels occur need to be discussed. The discrimination between the vowels in this task in expected to be poor-to-intermediate, and correspond to the Category-Goodness pattern, particularly in pre-lenis environment. While some difference between /e/ and /æ/ can be perceived, /e/ is still assimilated into BCS /e/, and /æ/ may stand out in terms raising before lenes and increased vowel quantity. The assimilation on the basis of the Single-Category type is possible before nasals and fortes. Since /æ/ is raised before nasal consonants, and /e/ is lowered, the participants are likely to confuse one vowel with another because vowel quality is not a reliable cue. In this case, vowel duration could become a primary cue, and facilitate recognition, as /æ/ is significantly longer than /e/ in pre-nasal context. Similarly, the reduced duration of /æ/ before fortes may result in the participants mixing up the vowels, but it is also possible that quality becomes a primary indicator of difference, since /æ/ is lowered before fortis consonants. In other words, the results in pre-fortis position could potentially demonstrate SC or CG, depending on which cue the participants focus on. In order for the participants to all under the Uncategorised-Categorised type, they would need to successfully discriminate between /æ/ and /e/ in all contexts, which is highly unlikely, given the difference in acoustic realisation (i.e. quality and quantity difference) of the vowels before lenes, fortes, and nasals by the native speaker of English.

6.5 Task 2: Category goodness rating of /æ/ and /e/

The assimilation patterns proposed by Best (1995: 195), presuppose the existence of perceptual similarity between the native and target phonemes, particularly the Single-Category (SC) and Category-Goodness type (CG). In order to test this, goodness-of-fit ratings are introduced (Krebs-Lazendić & Best 2007; Krebs-Lazendić 2008). Thus, Task 2, i.e. similarity rating between the vowels heard and the native vowel was designed to investigate which of the two English phonemes is a better exemplar of the native category /e/, and, with it, determine whether the assimilation into BCS category is CG or SC type. The acoustic stimuli were provided by the same native speaker of English as in Task 1 (from Doylestown, PA; recorded using PRAAT at sampling frequency of 41100 Hz). The recording procedure was slightly changed: instead of recording words separately, the words in this task were produced in a sequence, but each word was announced prior to pronouncing it (e.g. *Word number one is* ____, *word number two is* ___, etc.), and brief pauses were included to allow the participants enough time to answer. The design of the second task is presented in the following image:

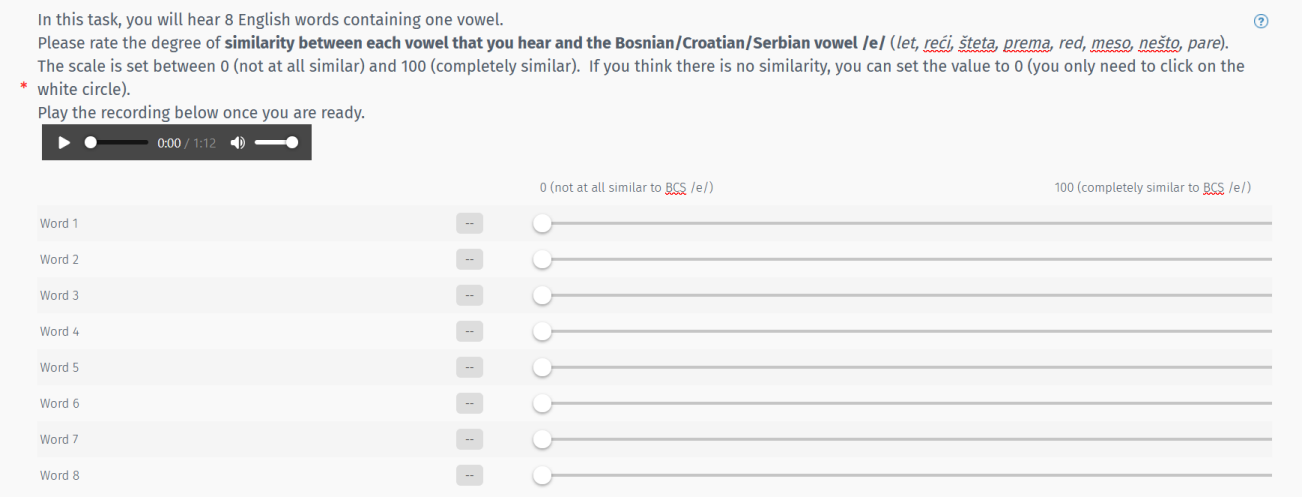


Figure 13. The design of Task 2: category goodness rating.

As observed, the participants were instructed to rate the degree of similarity between the vowels in each of the 8 words they heard and the BCS vowel /e/, on a scale from 0 ("not at all similar to BCS /e/") to 100 ("completely similar to BCS /e/"). Both examples from BCS and instructions in L1 were provided under "?". The experimental design was similar to that of Krebs-Lazendić and Best (2007: 285), who asked their participants to rate the degree of similarity between the English pairs of /æ/-/e/ and /i:/-/ɪ/ and their perceived equivalents in BCS, in order to test if the participants assimilated the vowels into the native categories on the basis of the Category-Goodness assimilation pattern. To test the category goodness, a 7-point Likert scale was used, ranging from 1 ("very little similarity") to 7 ("exactly the same vowels"). In their subsequent doctoral thesis, Krebs-Lazendić (2008: 75) used a 5-point Likert scale to assess the goodness-of-fit the same English vowels /æ/-/e/ and /i:/-/ɪ/ into BCS vowel categories. The only difference between this study and the previously mentioned studies is that a scale from 0 to 100 was used to ensure more nuanced values assigned to the degree of similarity.

With reference to the acoustic stimuli, 8 English words were recorded:

Table 7. Words used in Task 2, together with their phonological environment, IPA representation, and vowel quantity in seconds.

Environment	Spelling	IPA (AmE)	Duration in /s
Pre-fortis	*bat*	/bæt/	0,190
	bet	/bet/	0,138
	bought	/bɑːt/	0,163

	but	/bʌt/	0,127
Pre-lenis	*bad*	/bæd/	0,285
	bead	/bi:d/	0,265
	bid	/bɪd/	0,143
	bed	/bed/	0,128

All the words appeared in the pattern *b_d* or *b_t*, therefore in pre-lenis and pre-fortis position, but only before alveolar plosives (the words were not recorded in this sequence). As evidenced in the table, pre-lenis position has an effect on all the vowels apart from /e/, which is longer before /t/ than before /d/. The quantity of /æ/ is increased before the lenis plosive /d/ (t=0,285 s), but the vowel is also significantly longer than /e/ in its pre-fortis position (t=0,190 s vs. t=0,138 s in *bet* and t=0,128 s in *bed*). When compared to BCS vowels analysed earlier (cf. 4.2.2), English /e/ in both environments is shorter than BCS /e/ and /a/, while /æ/ is longer than BCS /e/ and /a/ (BCS /e/ t=0,176 s; BCS /a/ t=0,189 s).

In total, 8 English words are recorded for this task, and their vowels were manually extracted and plotted in the following chart, together with BCS /e/ in different regions (the exact F1 and F2 values of the vowels recorded for this task can be seen in Appendix D):

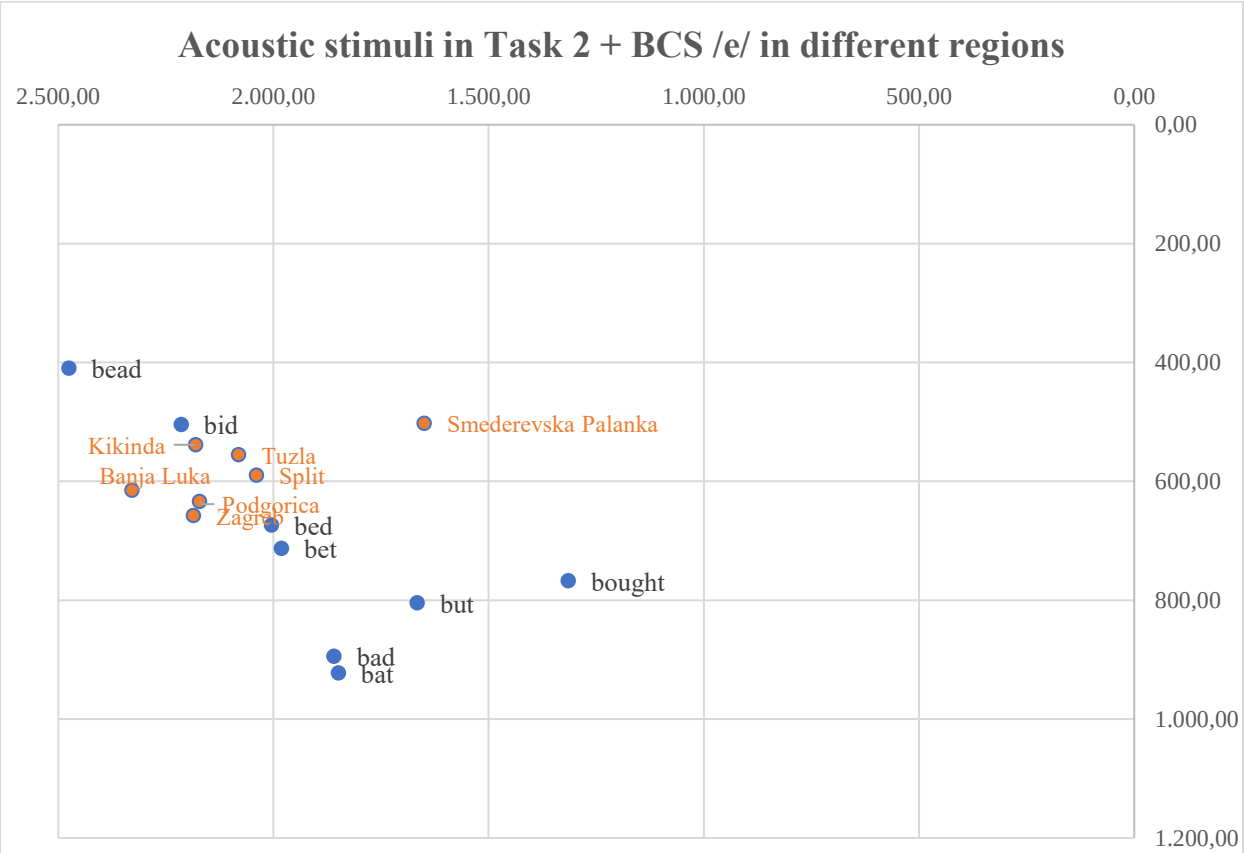


Figure 14. Graphic representation of vowel F1 and F2 values for 8 words used in Task 2, together with the representation of BCS vowel /e/ in different regions. The x- axis shows the F2 (2500-0 Hz), while the y-axis shows F1 (0-1200 Hz). English vowels are shown in blue, while BCS /e/ is orange.

The above chart demonstrates that, in comparison to pre-lenis context, both /æ/ and /e/ before fortis consonant /t/ are slightly lowered. In this task, English /e/ is lower than BCS /e/ but still close (the exception is Smederevska Palanka), which could potentially justify high similarity ratings between the English and BCS vowel. In contrast, BCS /e/ is significantly higher than /æ/, which could account for lower similarity ratings. Other vowels in this task were chosen to prevent the participants from hyper-focusing on /æ/ and /e/ only, and to determine the goodness-of-fit in a greater detail, which is why the task was created in the first place. In other words, if all the vowels presented are rated equally, then they are all perceptually close in the minds of the participants. If the ratings for /e/ are significantly higher than other vowels, but /æ/ receives high ratings that are close to that of /e/, this is a great indicator of CG type assimilation, meaning that both members of the vocalic contrast are assimilated into the L1 category /e/, but on a continuum, where /e/ is a better candidate for BCS /e/, whereas /æ/ is seen as a less optimal example of the native category. This scenario is highly likely, given how close English /e/ and BCS /e/ are in terms of F1 and F2 in the above chart. If /e/ and /æ/ receive similarly high ratings, this demonstrates that the assimilation occurs in the Single-Category pattern and that both vowels are equally good exponents of BCS /e/. As already mentioned, SC assimilation is less likely, especially in this task, since English /e/ and /æ/ are too distant in terms of F1, and English /e/ is much closer to the L1 phoneme. Finally, if English /e/ is rated highly similar to BCS /e/, whereas the ratings for the other vowels, including /æ/ are low, this could also affirm the UC type assimilation, however unlikely. Therefore, while Task 1 tested the perception of /æ/ and /e/ in individual words and helped predict the assimilation patterns on the basis of their link to vowel discrimination, Task 2 allowed me to ascertain which exact pattern of assimilation is responsible, i.e. to provide the right answer to my initial hypothesis.

6.6 Statement rating: Exposure to English and attitudes to NS norms

Following the phonetic/phonological part of the experiment, the participants were asked to rate a number of statements pertaining to their exposure to and experience with English, as well as their attitudes to native-speaker norms. Since the participants already performed minimal pair distinction/word identification, and had to rate the similarity between the English vowels and the native vowel /e/, for this part, only a simple 5-point Likert scale was used (1= “strongly disagree” to 5= “strongly agree”).

Given that "increased L2 experience may foster improved recognition of the differences between L1 and L2", meaning that the assimilation of foreign phonemes into native categories may diminish, leading to the formation of distinct L2 categories, the degree of the participants' exposure to and experience with English needed to be investigated (Best & Strange 1992: 307). With regards to the patterns of assimilation discussed in section 2.1, studies have demonstrated that a higher degree of L2 exposure facilitates phoneme discrimination, separate category formation, and lower similarity judgements for some speakers in Category-Goodness type scenarios (Strange & Best 1992; Levy & Strange 2008; Levy 2009; Kopečková 2012). As a result, the participants in this study were shown the following statements first:

1. I often speak English with my friends and family.
2. I often use English for work-related purposes.
3. The music I listen to is mostly in English.
4. During conversations, I can switch between English and my mother tongue effortlessly.
5. I use many words borrowed from English when I am speaking my language.
6. I watch films/TV shows/YT videos in English on a daily basis.
7. I read books/newspapers/blogs/articles in English on a daily basis.
8. I mostly use English to communicate with non-native speakers of English.

The above statements were used to ascertain the degree of exposure to English (3, 6, 7), as well as "the extent to which L2 is relied upon to perform certain functions", be it work-related purposes (2), or "personal and social expression" (1) (Moyer 2014: 448). Statements 4, 5, and 8, also examined the level of experience with English, but in the context of EFL and ELF. More precisely, the issue of preserving the purity of the mother tongue, as well as the use of English in mostly international, lingua franca settings were investigated (Jenkins 2009: 143).

Having completed the part about their experience with and exposure to English, the participants were asked to rate the statements in the table:

Table 8. Statements about NS norms.

Nr.	**EFL + positive attitudes to NS norms**
1.	The only way to be fully proficient in English is to sound like a native speaker from Britain or the US.
2.	When I speak with native speakers of English, I am afraid they would judge my competence if I do not sound like them.

3.	I think that the fact that BCS is my mother tongue makes it difficult to learn English, particularly in terms of listening and speaking.
4.	When I speak English with other non-native speakers, I tend to judge their language competence on the basis of how they speak.
5.	I think that teachers who sound like native speakers of English are more competent and appealing to students.
Nr.	**ELF + negative attitudes to NS norms**
1.	I do not care about my English accent and learning pronunciation as long as I am understood.
2.	There is no point in wanting to sound like someone from the UK, US, Canada, etc., when English is a global language.
3.	I think that people who are majoring in English should be the only ones concerned with sounding like native speakers.
4.	I do not like it when speakers of Bosnian/Croatian/Serbian try to sound like native speakers of English, when they do not live in English-speaking countries.
5.	I think that schools and universities in Bosnia/Croatia/Serbia should pay more attention to grammar and vocabulary, and not waste time on speaking and listening skills.
6.	I think it is unnatural for speakers of my language to insert English words in sentences and use the original pronunciation.
7.	I don't really pay attention to other people's accent because I am more interested in what they are actually saying.

The purpose of introducing the statements about EFL and ELF in the experiment was to examine how the patterns of assimilation relate to the (non-)conformity to native-speaker norms. Statements supporting the EFL principle and native-speaker norms are mostly concerned with the nativeness ideal and the "traditional idea of mastering a language", i.e. being NS-like at all levels of language, both with native and non-native speakers alike (Timmis 2002: 248). Moreover, statement 5 is concerned with English language teaching context, where Wach (2011) found that educational and career goals of English majors may have had a strong influence on the respondents rating the EFL statements favourably. The purpose of statement

3 was to examine whether L1 is regarded as a detriment to acquiring L2 speaking and listening skills, which is quite a common notion in EFL-oriented studies (cf. section 5.2).

In contrast, the topics of ELF-oriented statements and those with negative attitudes to native-speaker norms are (international) intelligibility (1, 2, 7), retainment of L1 identity (4, 6), and ELF in education (3, 5). Statements about intelligibility emphasise the fact that mutual intelligibility should be key in international English communication, which is "a more realistic and feasible goal" than approximating RP or GA (Thir 2016: 2013). Similar statements were rated positively by participants in Wach (2011) and Tengs Sannes (2013), who found that, despite a high degree of variation, having a near-NS accent is less important than being understood in international communicative contexts. Furthermore, while both 4 and 6 examine the retainment of L1 identity in English, statement 6 is more concerned with the idea of BCS being "contaminated" with English (Jenkins 2009: 141). Finally, the significance of ELF and negative attitudes to NS norms in educational settings were also investigated. Whereas statement 3 stresses the idea that acquiring a native-like accent should only be reserved for English-language teachers and those majoring in English, who, according to Wach (2011) have positive attitudes to NS norms, statement 5 emphasises the need to focus more on grammar and lexis in the classroom, which Naser (2011: 4-5) argues is the root cause of L1 transfer.

6.7 Testing procedure

Before distributing the whole experiment to participants, a trial run was conducted with 20 participants to examine the experimental design, the adequacy of acoustic and visual stimuli, as well as the functioning of the software. The only issues detected were technical in nature: a number of participants in the trial run reported that the audio files could not be played on iOS, which was resolved by converting the files from .wav into .mp4.

Following the trial run, the experiment was distributed mostly online via E-Mail or social media to the target group. A number of participants were asked complete the experiment in person, on my computer. Due to the aforementioned division into two tests (Task 1 was different, other questions were unchanged), separate links were distributed. However, I ensured to have approximately the same number of answers in both tests. Apart from the specific target group in mind for the whole experiment, there were no additional criteria for distributing the two tests, and they were later analysed as one. Finally, the online distribution and practicality of the experiment allowed for a rather rapid data collection process, which was initiated in February and lasted approximately three months, before the commencement of the analysis.

7. Statement ratings: Results and discussion

In this section, the data collected from the statement ratings pertaining to participants' English language exposure and their attitudes to native-speaker norms will be reported and discussed, particularly in comparison to previously reviewed studies in subsection 5.2.

7.1 Participants' exposure to English

As already mentioned in section 6.3, the vast majority of participants in this study reported learning English in classroom settings, be it primary school, high school, or extracurriculars (private language lessons). Only a small number mentioned music, films and other media as a source of learning, but only in combination with formal instruction. To obtain data on the current degree of exposure to English, the participants were given 8 statements (cf. section 6.6). The software used in this study (QuestionPro) automatically calculated the percentages for each point on the Likert scale ("strongly disagree", "disagree", "neutral", "agree", "strongly agree").

A high degree of self-reported exposure to English-language media is observed in the figure below (for the complete dataset, see Appendix H):

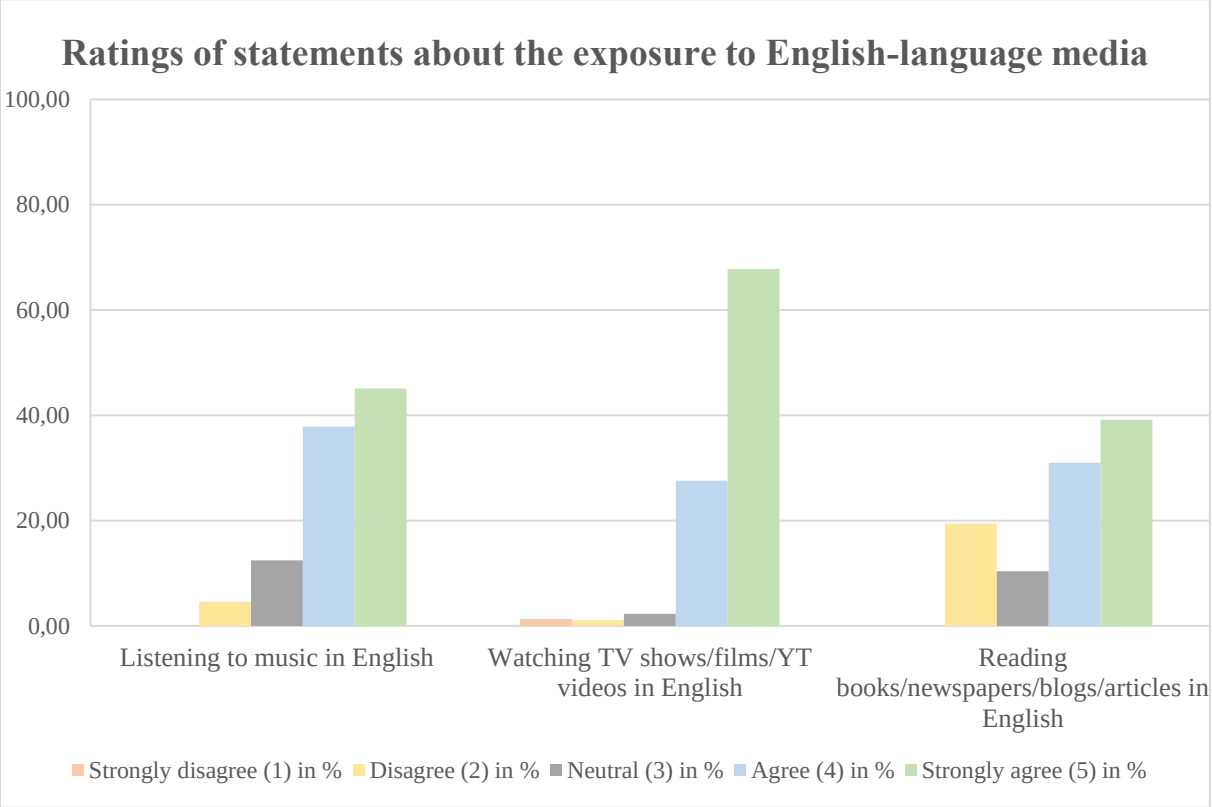


Figure 15. Ratings of statements about the participants' exposure to English-language media. The x-axis shows the three statements in question (music, TV shows/films/YT videos, reading books/newspapers/articles in English). The y-axis shows the percentages of participants who answered "strongly disagree" (orange), "agree" (yellow), "neutral" (grey), "agree" (blue), and "strongly agree" (green).

As observed, there is a general preference for spoken English, particularly the 'real-time' speech in TV shows, films, and YouTube videos. Furthermore, the smallest degree of variation in the sample is observed in the statement about the exposure to English films/TV shows/YT videos, with the vast majority of participants expressing agreement with the statement (Σ= 94,4%), followed by the statement about listening to music in English (Σ= 82,94%). The highest degree of data dispersion is evidenced in the statement about the exposure to written English. Despite the overall agreement with the statement (Σ=70,16%), 19,45% of participants voiced their disagreement. Therefore, there is a general tendency towards exposure to spoken English through music or visual media.

The remaining five statements were concerned with the actual use of English, be it the context in which English is used (work or private life), the status of the interlocutors (non-native speakers), as well as the relationship between the BCS and English. The detailed results are shown in the following chart:

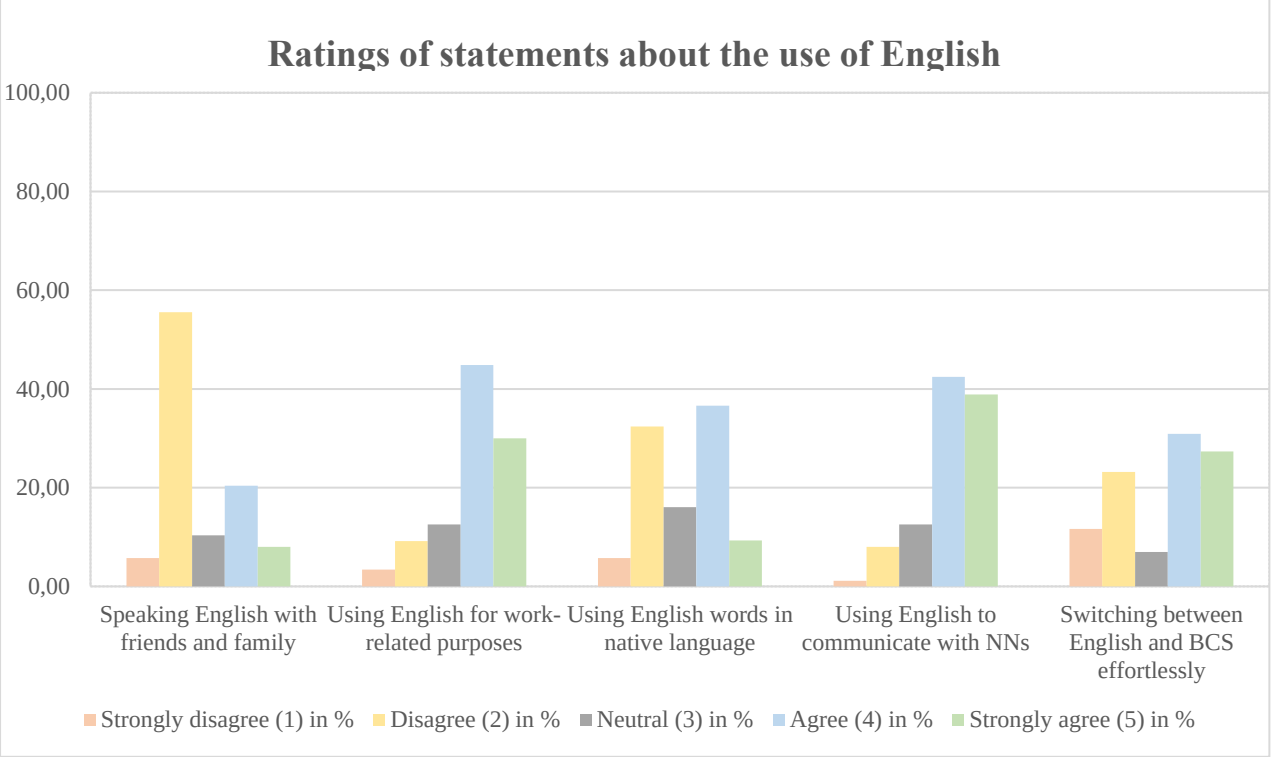


Figure 16. Ratings of statements about the use of English. The x-axis shows the five statements in question. The y-axis shows the percentages of participants who answered "strongly disagree" (orange), "agree" (yellow), "neutral" (grey), "agree" (blue), and "strongly agree" (green).

With reference to the context of English-language use, it is apparent that English is used mainly for work-related purposes, with the majority of participants agreeing with the statement (Σ=74,84%), as opposed to the use of English in the private sphere, i.e. with friends and family

(Σ_{agree}=28,42% vs. $\Sigma_{disagree}$=61,28%). When comparing with others, the statement about the use of English in private life is the one with the highest percentage of disagreement from the participants. Furthermore, the vast majority (Σ=81,34%) have reported the use of English to communicate with other non-native speakers, making them the speakers of ELF. This statement has the highest number of participants rating it with "strongly agree" (=38,88%). Finally, the statements examining the relationship between the source and the target language show a high degree of variation. Whereas the participants express general agreement with the use of many anglicisms while speaking BCS, the percentage of those disagreeing with the statement is rather significant (Σ_{agree}=45,87%, $\Sigma_{disagree}$=38,10%). Relatively similar findings are observed in relation to the statement about switching between BCS and English effortlessly during conversations (Σ_{agree}=58,18%, $\Sigma_{disagree}$=34,85%).

7.2 Participants' attitudes to native-speaker norms

The responses to the 12 statements about the attitudes to native-speaker norms were computed by QuestionPro: the software assigned the values to each point on Likert scale (1= "strongly disagree", 2= "disagree", 3= "neutral", 4= "agree", 5= "strongly agree"), and automatically calibrated the average score ($\bar{x}$) and standard deviation from the mean (s), together with the percentages. On the whole, the data suggests a general preference for ELF statements, as seen in the $\bar{x}$ and s in the following two charts (for the complete dataset, see Appendix I):

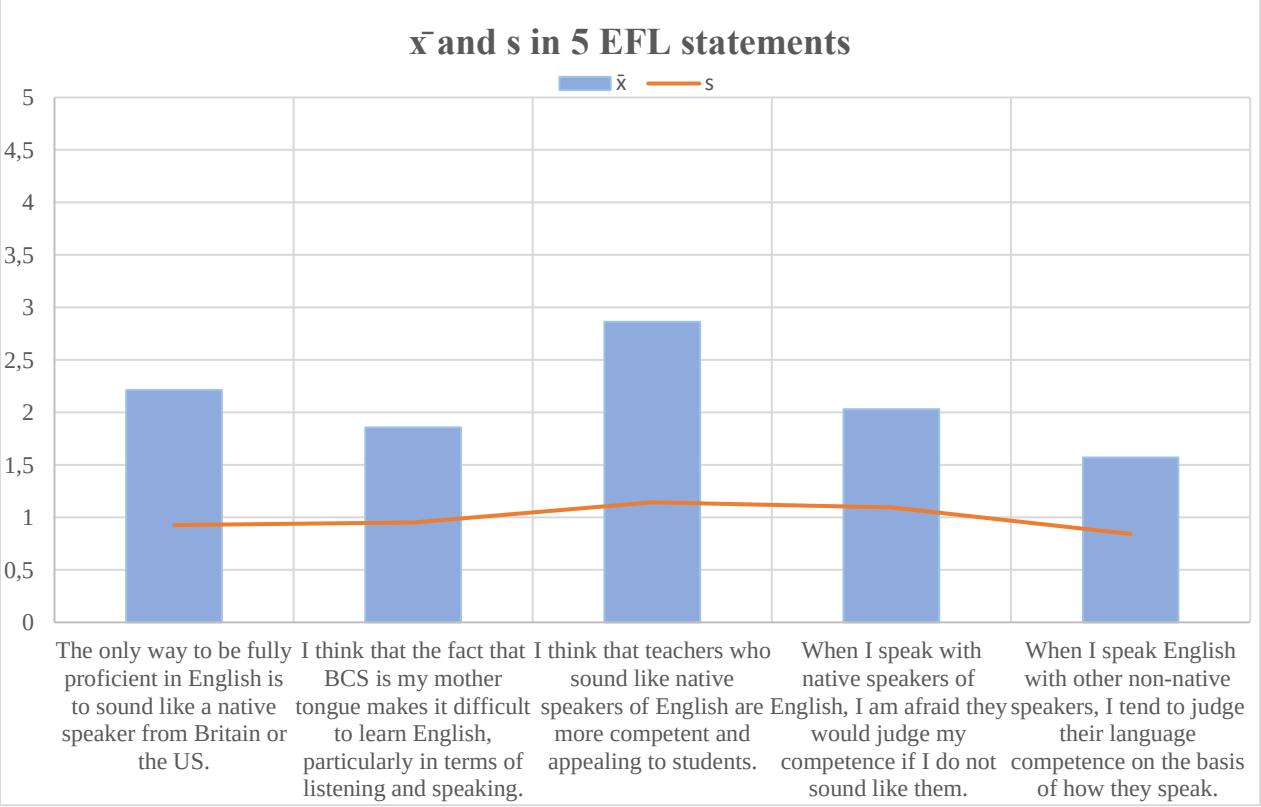


Figure 17. Mean score x̄ (blue) and SD (orange) in five EFL statements. The y-axis shows the mean scores and SD of the statements (0-5). The x-axis shows the statements in question.

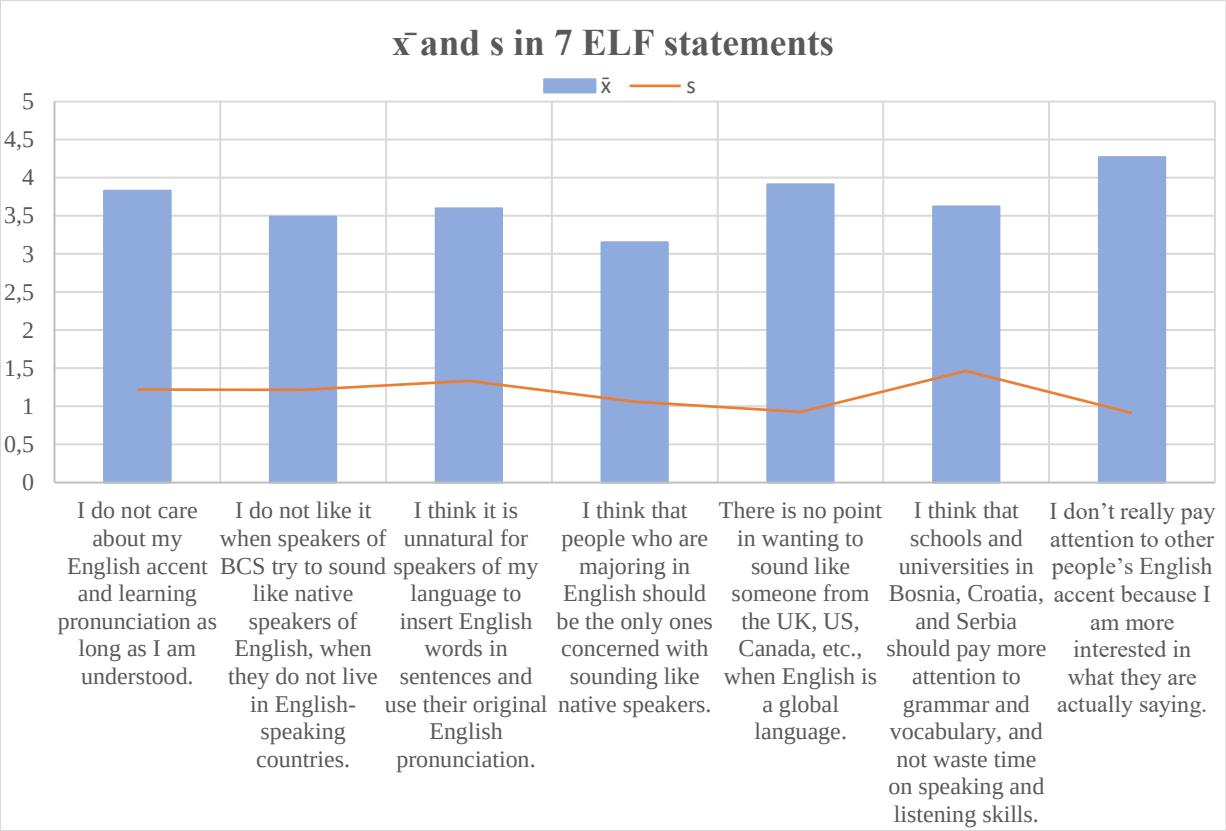


Figure 18. Mean score x̄ (blue) and SD (orange) in seven ELF statements. The y-axis shows the mean scores and SD of the statements (0-5). The x-axis shows the statements in question.

As observed in the above charts, the participants in this study expressed general agreement with ELF-oriented statements, mostly agreeing with the statements favouring their own intelligibility, as well as that of the others, and understanding the global nature of English today. Furthermore, the statement about being more interested in what the interlocutors are saying rather than their accent was the most positively rated statement (x̄=4,272). In contrast, EFL-oriented statements were rated less favourably by the participants (x̄=1,571-2,218). In this group, the lowest mean score was awarded to the statement about judging the competence of others on the basis of their accent (x̄=1,571), while the statement about the competence and appeal of teachers who sound like native speakers of English was given the highest average score (x̄=2,219). With regards to standard deviation from the mean (s) in both charts, the data appears to be relatively dispersed from the sample x̄ for each statement, meaning that there is a moderate degree of variation among the participants in their responses to both EFL and ELF statements (s=0,842-1,465). When compared with others, only the statement about focusing on grammar and lexis rather than listening and speaking skills in schools exhibits a higher

degree of variation in the answers (s=1,465) Apart from the overall disagreement, the statement concerned with the judgements of English language competence on the basis of accent demonstrates the lowest degree of variation among participants. Thus, ELF-oriented statements were rated more favourably than the ELF ones, and the degree of unity in the answers varies, but there is generally more variation in ELF statements.

Nevertheless, each statement needs to be examined in a greater detail, beginning with EFL statements:

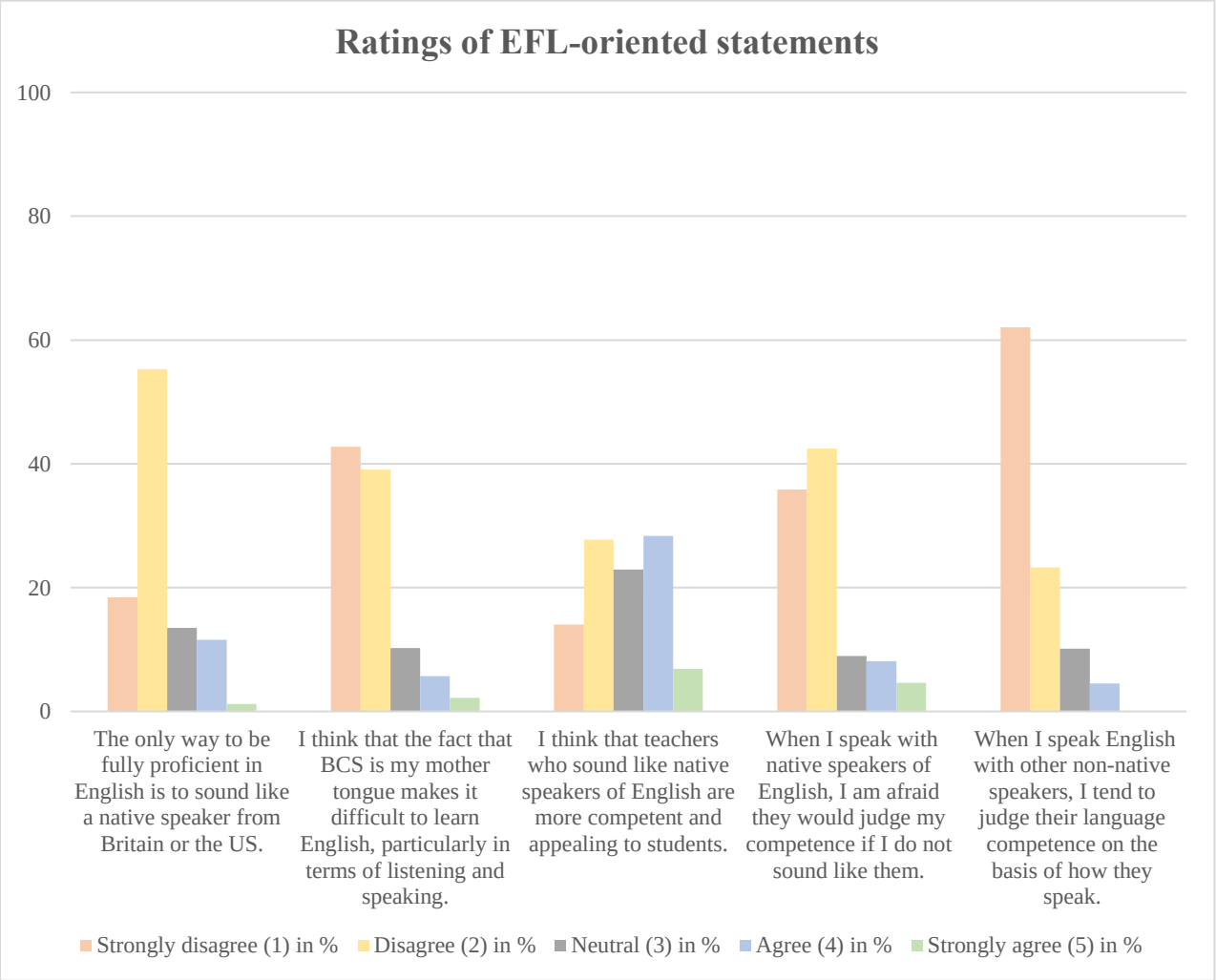


Figure 27. Ratings of EFL-oriented statements. The x-axis shows the five statements in question. The y-axis shows the percentages of participants who answered "strongly disagree" (orange), "agree" (yellow), "neutral" (grey), "agree" (blue), and "strongly agree" (green).

When considering the disagreement with the first statement (Σ=73,73%), it is apparent that emulating a native-speaker variety (RP or GA) is not considered a determining factor for mastering the English language by the participants in this study. Moreover, BCS is not regarded as a detriment to learning English, with the majority of participants rating this statement

negatively (Σ=81,53%). Interestingly, the participants seem to care less about the accent of other non-native speakers during communication ($\Sigma_{disagree}$=85,31%) but are slightly less confident in their competence when communicating with native-speakers of English. Despite the majority of participants disagreeing with this statement (Σ=78,34%), the percentage of those voicing agreement, albeit small, has to be taken into account (Σ=12,69%). Nonetheless, the highest degree of variation is detected in the responses to the statement about the competence and appeal of teachers who sound like native speakers of English. Focusing on the sum of those disagreeing (Σ=41,83%), as compared to those agreeing (Σ=35,24%), it appears that the negative ratings prevail.

Furthermore, the detailed overview of ELF statements is provided in the following chart:

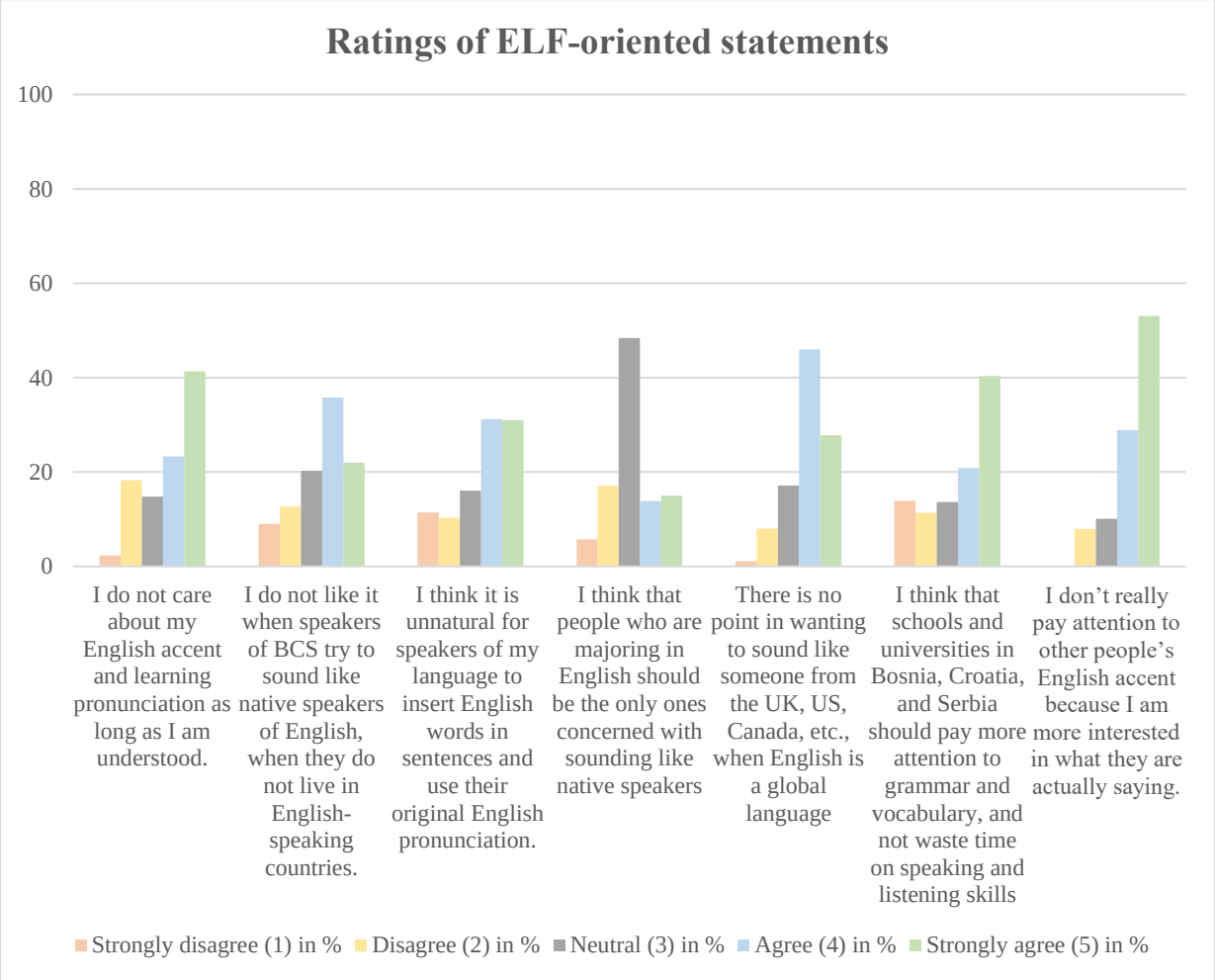


Figure 28. Ratings of ELF-oriented statements. The x-axis shows the five statements in question. The y-axis shows the percentages of participants who answered "strongly disagree" (orange), "agree" (yellow), "neutral" (grey), "agree" (blue), and "strongly agree" (green).

The overall agreement with ELF-oriented statements, but a higher degree of response dispersion, can be observed on a first glance. Statements favouring mutual intelligibility of self

and the others were rated rather positively, with the participants expressing most agreement with the statement about paying more attention to what the interlocutors are saying (Σ=81,98%). While rating this statement, none of the participants selected "strongly disagree", and only 7,94% somewhat disagreed with the statement. The participants seem to be more conflicted about their own intelligibility than that of the others, which is evident in the higher degree of variation in the answers ($\Sigma_{disagree}$=20,55% ; neutral=14,76%, Σ_{agree}=64,68). Similarly, the participants agree with the notion that English is a global language used in international communication, and therefore the need to attain RP or GA (or other NS varieties) is not that pronounced (Σ=73,74%). Only a small percentage of participants remain neutral in their stance (=17,14%), or voice disagreement (Σ=9,12%). With reference to statements about the retainment of L1 identity and preserving the 'purity' of the native language, there is an overall agreement with the negative attitude towards speakers of BCS who appropriate NS varieties of English (Σ=57,78%), as well as the insertion of English words and their original pronunciation in BCS (Σ=62,22%). Finally, the participants exhibited rather mixed responses to the prevalence of EFL and ELF in education. On one hand, more than a half (Σ=61,12%) voiced agreement with the idea that teaching listening and speaking skills in BCS schools is a waste of time, and that the efforts should be refocused on grammar and lexis, whereas a quarter of participants expressed disagreement (Σ=25,24%), and 13,65% were neutral. On the other hand, the endorsement of the idea that emulating a NS accent is reserved for those majoring in English was rated both negatively (Σ=22,74%) and positively (Σ=28,81%), but almost a half of respondents, i.e. 48,41% choose to remain neutral on this issue.

7.3 Discussion: Exposure to L2 and attitudes to NS norms

Upon closer examination of the responses from the sociolinguistic part of this study, a number of conclusions can be drawn about the status of native speakers of BCS as users of English, their reliance on English to perform certain functions, the interrelatedness of BCS and English, as well as their attitudes to native-speaker norms in communication, and, finally, the preservation of L1 identity in English.

First of all, the participants in this study seem to be highly experienced with English through years of exposure in the classroom and extracurricular settings, in which the mode of communication were the two standard varieties of English, RP and GA (Moyer 2014: 448). Notwithstanding the prolonged exposure to EFL in education, the participants identify as users of English as a lingua franca, based on the fact that the vast majority uses English to communicate with other non-native speakers. Furthermore, for this population, English is only

used in professional contexts, which Moyer characterises as "work-related and perfunctory language use", meaning that there is not much place left for English use in the personal lives of the participants. The participants are, however, still heavily exposed to spoken English through music and TV/films/YouTube, indicating that the language is used primarily for the purposes of leisure and relaxation. This aligns with an earlier study by Paunović (2003: 394), who found that the majority of Serbian high school students (N=459) seldom used English for personal purposes (outside of classroom settings), but were still heavily exposed to the language through TV and films. Finally, a high degree of data dispersion in the statements regarding the relationship between BCS and English (switching between L1 and L2 effortlessly, and the use of anglicisms in BCS), implies that the participants are rather conflicted about the interrelatedness between the source and the target language in terms of experience/exposure to L2. Therefore, the population examined in this study identify as ELF users, despite learning English in school, who utilise English for a very specific, professional purpose, but are still exposed to spoken English in their free time.

With reference to attitudes to native-speaker norms, the ratings of EFL and ELF-oriented statements generally indicate that BCS speakers in this study are against native-speaker norms, which is opposed to the findings of Timmis (2002). In the present study, mastering a native-like accent is not considered "a benchmark of achievement", and is not required to be regarded as proficient in English (Timmis 2002: 248). Furthermore, the need to attain a native-speaker accent in English is questioned, since English is regarded as a global language today, supporting the idea of "multiple Englishes" (Zoghbor 2011: 287). Given that the participants have already identified as ELF speakers, their negative attitudes to the statements favouring their own accent and that of others, together with judgements of English language competence on the basis of pronunciation, are justified. Similar attitudes were expressed by Wach's respondents (2011: 260), who doubted the relevance of native-speaker norms in lingua franca settings. Likewise, the statements favouring intelligibility of participants, as well as of their interlocutors are in line with the participants identifying as users of English as a lingua franca, confirming the results by Tengs Sannes in the Norwegian context (2013: 81). The findings in this study have elucidated that, in (international) communicative contexts, which they are a part of, native speakers of BCS are generally focused on mutual intelligibility and the contents of their own and their interlocutors' speech, rather than making assumptions about English language proficiency on the basis of accent. Furthermore, the participants exhibit a tendency towards being more reflective of their own speech when communicating with native speakers

of English, but do not seem to be interested in the accent of other non-native speakers. In other words, despite the preference for mutual intelligibility, the participants are still slightly aware of their (foreign) accent and fear judgements by native speakers. Thus, in contexts that pertain to them personally, native speakers of BCS tend to focus more on mutual intelligibility.

Finally, native speakers of BCS seem to have particular attitudes to the relationship between English and their mother tongue. Unlike Jenkins' observations about EFL (2009: 141), in this thesis, L1 is not regarded as an obstacle to mastering English, as noted in the disagreement with the statement which recognises that having BCS as the first language makes the acquisition of listening and speaking skills in English more difficult. One possible interpretation is that participants in this study consider themselves relatively experienced and proficient in English, who never attributed any potential difficulties in learning English to their own mother tongue, but rather to their individual language aptitude. The general agreement with the statement negatively portraying speakers of BCS who want to attain a native-like English accent demonstrates the desire to preserve and freely express the speakers' ethnic and linguistic identity in international communication, which creates "equilibrium between identity and intelligibility in English pronunciation" (Thir 2014: 64). Similar opinions were voiced by high school students in Norway, who judged the acquisition of native English varieties as "unnatural and unnecessary", and acknowledged the significance of mutual intelligibility (Tengs Sannes 2013: 81). Furthermore, the statement emphasising the 'unnaturalness' of inserting English words and retaining their original pronunciation while speaking BCS, is understood as stemming from the desire to preserve L1 from being 'contaminated' by external influences. Taking the tumultuous history of the region where BCS is spoken, together with the marked nationalism in the formation of linguistic identity (cf. Alexander 2006; Brown & Alt 2004), it is clear why the participants in this study exhibit adversity towards English potentially 'polluting' their mother tongue. Such an attitude may also stem from the fact that English is mostly used in professional context, thus fulfilling a particular function, hence the desire to separate 'the language of work' from the mother tongue, which is the primary mode in private communication. At the same time, however, the participants report that they use anglicisms in BCS, meaning that the realm of phonetics/phonology is where the line is drawn.

8. Auditory perception tasks: Results and discussion

This section will present the findings from the two auditory perception tasks, collected from 87 participants in the span of three months. Both overall results, and the results according to

the phonological environment of the examined vowel quality contrast in Task 1 will be reported. This section will proceed with the data on category goodness ratings (Task 2), and conclude with a discussion of the assimilation patterns according to the Perceptual Assimilation Model.

8.1 Results of Task 1: Perception of words in isolation/minimal pair distinction

Before reporting the results of Task 1, the limitations in the experimental design need to be taken into consideration, and they are mainly concerned with the use of visual stimuli. As already outlined in section 6.4, choosing the appropriate images to represent the minimal pairs tested in Task 1 was quite a laborious task, as some words were more difficult to represent visually and are less iconic than the others. Furthermore, some minimal pairs that were initially chosen had to be discarded because they were impossible to represent visually. For instance, the image for *bag* is more concrete and iconic than the image for the verb *beg*, which may trigger the preference for the noun and bias against the verb. Likewise, some members of the pairs were phonologically more similar to the words in BCS (e.g., *pedal* and BCS *pedala*), and the participants may have been more biased towards them, while exhibiting bias against proper names (*Ben, Brad*), verbs (*beg*), and adjectives (*dead*). Therefore, the participants' bias towards nouns and words that are identical to L1, as well as the bias against proper names, verbs, and adjectives have unveiled the need to approach the results with great caution, and have emphasised the importance of taking the influence of the visual stimuli on the results into account. Despite the fact that images were picked in order to avoid the impact of orthography on the results, the results of Task 1 cannot be properly analysed without examining the potential effects of the visual stimuli, particularly in the words where the bias is apparent, which will be done in the following subsections concerned with the results in pre-lenis, pre-nasal, and pre-fortis environments.

What needs to be presented first are the overall results for /æ/ and /e/ for all 87 participants. The results were calculated for six possible variables:

1. /æ/ correctly selected
2. /e/ selected in place of /æ/
3. distractions for /æ/ selected
4. /e/ correctly selected
5. /æ/ selected in place of /e/
6. distractions for /e/ selected

Since there were two vowels and 12 words containing either one of the vowels, the number of selections per word was summarised for each variable. The exact numbers per variable are presented in the following chart (the complete dataset can be found in Appendix F):

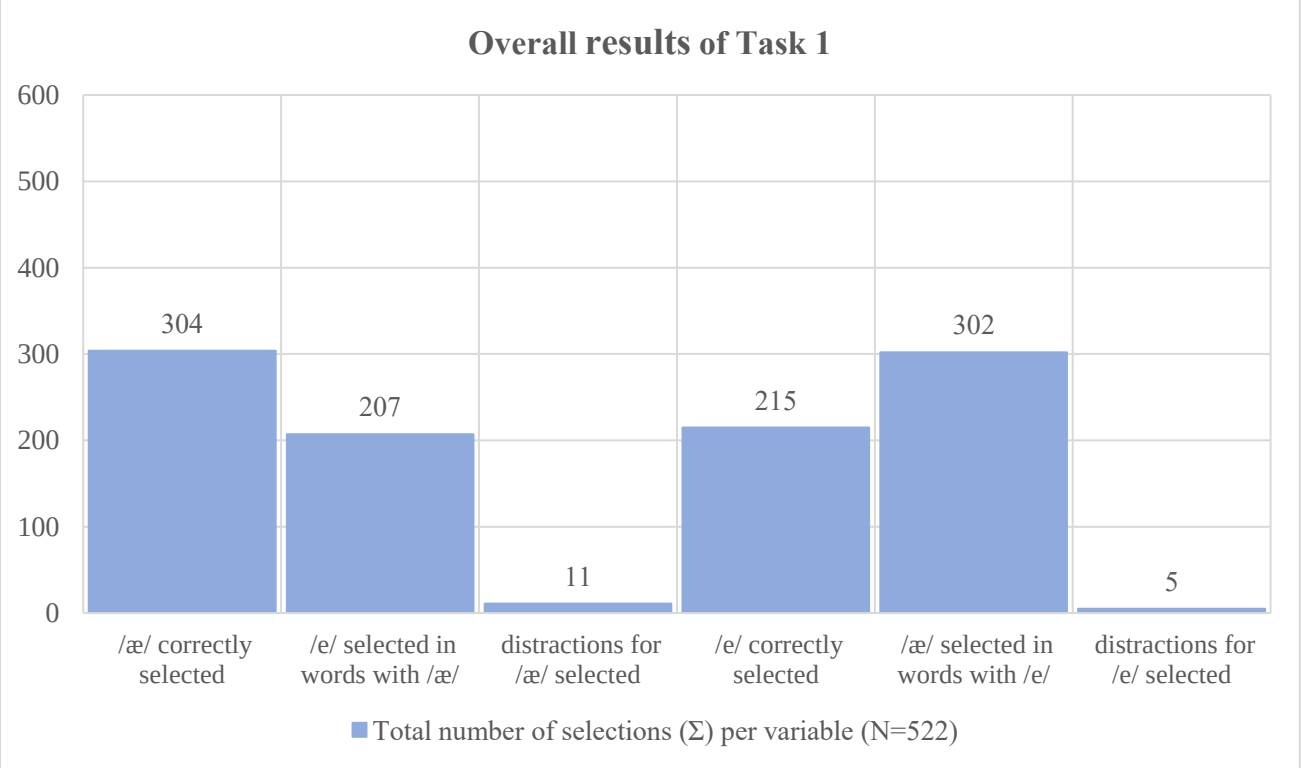


Figure 21. Overall results of Task 1. The x-axis shows six possible variables: /æ/ correctly selected, /e/ selected in words with /æ, distractions for /æ/ selected, /e/ correctly selected, /æ/ selected in words with /e/, distractions for /e/ selected. The y-axis shows the total number of selections per variable.

The chart illustrates that /æ/ is generally better identified (Σ=304), but that there is still a high degree of confusion among the participants who heard the vowel as /e/ (Σ=207), meaning that the acoustic difference between the two sounds is still heavily disregarded. The participants do not seem to identify /e/ as well as /æ/, which is evident in the overall number of selections of /æ/ in words that contain /e/ (Σ=302). Furthermore, in the tokens that contain the phoneme /e/, the participants also did not attend to the acoustic difference between the sounds, judging by the high numbers of both correct and incorrect selections (Σ=215 vs. Σ=302). Therefore, the quality and quantity differences between the realisations of the two vowel phonemes seem to be insufficient for the participants discriminating between the vowels, but /æ/ still stands out in the responses. As already discussed in section 3.2, both F1 and F2 values of /æ/, as well as its duration, tend to vary greatly in different phonological environments (pre-lenis, pre-nasal, pre-fortis), while /e/ is more stable in terms of quality and its quantity is not impacted by the following consonants. Moreover, F1 and F2 of English /e/ are much closer to BCS /e/ (cf.

section 4.2.2). Taking the phonetic features of both vowels into consideration, it is clear why /æ/ is identified more often, especially in words containing /e/: hypercorrection. In other words, English /e/ is too similar to the L1 vowel /e/, that the participants tend to be biased towards /æ/ because its varying realisation in different environments it considered as too deviant from the BCS /e/. Since the participants are aware that they are hearing words from a foreign language, they think that they should have heard a different sound when listening to words that contain /e/, because this vowel is more similar to their L1. Thus, BCS listeners cannot really trust their own perception, so they opt for a vowel that is more different from their native vowel, even in the tokens where English /e/ is apparent.

8.1.1 Pre-lenis environment

In the pre-lenis environment, /æ/ is identified more often than /e/, as shown in the chart:

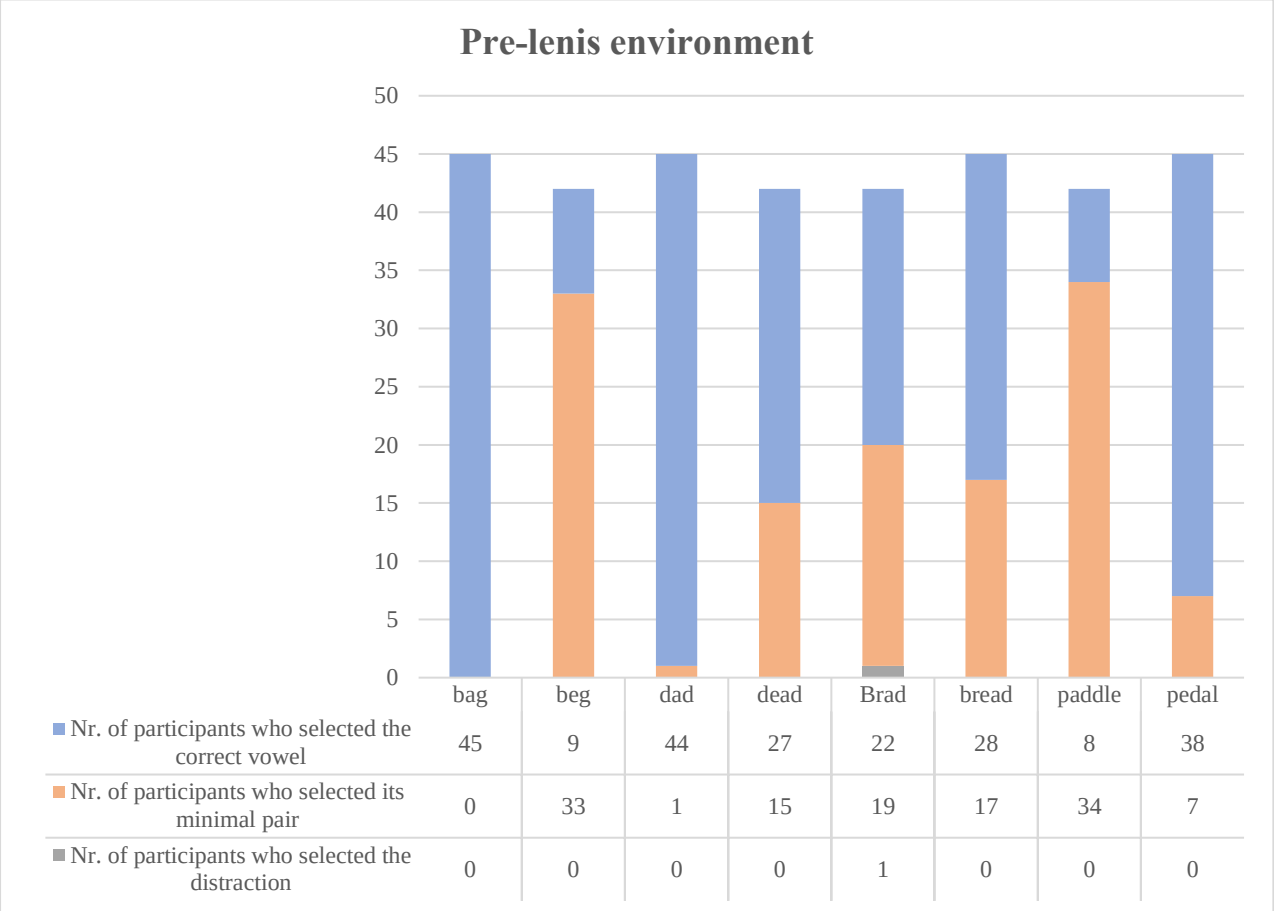


Figure 22. Results of Task 1 for /æ/ and /e/ in pre-lenis environment. The x-axis shows the words in question, while y-axis shows the number of participants.

In *bag* and *dad*, /æ/ is typically heard by the participants, with *bag* having a perfect score (n=45), followed by *dad* (n=44). Both *bag* and *dad* are nouns, and their visual representations are more concrete and iconic than the images chosen for *beg* and *dead*, highlighting the bias

towards nouns and against verbs and adjectives. There is a high degree of variation in *Brad,* and the number of correct and incorrect selections seem to be very close (n=22 vs. n=19), indicating the participants' bias against proper names, meaning that the results were influenced by the visual stimuli. In *paddle*, the vowel is heard more frequently as /e/, and, in combination with the correct identification of /e/ in *pedal*, demonstrates the bias towards the words that are phonologically similar to or attested in L1 (*pedal* and BCS *pedala*). When the phonetic properties of the vowel, analysed thoroughly in section 3.2, are taken into account, /æ/ does not generally change its height or frontness before lenes, but its quantity is significantly increased. For instance, /æ/ is the longest in *bag* (t=0,279 s) and *dad* (t=0,283 s), which, in combination with the visual stimuli, shows why the vowel stands out.

Furthermore, in words with /e/, there is a higher degree of variation among participants, as can be observed in *dead* and *bread*, which can be attributed to the bias against adjectives and proper names. Likewise, the fact that 33 participants who heard *beg* and 45 who heard *bag* selected the image of an actual bag also shows the preference for nouns in the visual stimuli. The vowel is identified in *pedal* and *paddle* due the words being perceived as acoustically identical to the BCS word *pedala*. In terms of the F1 and F2 and durational properties of /e/, such results may have been determined by the fact that phonological environment, be it pre-lenis, pre-fortis, or pre-nasal, does not have a notable impact on the quality and quantity of /e/. Moreover, the vowel is still closer to BCS /e/ in terms of quality (cf. section 3.2).

Taking minimal pairs into account, *bag-beg* and *paddle-pedal* show that only a single vowel is associated with the two members of each pair: in *bag-beg* the vowel is /æ/ (n=45 in *bag*, n=33 in *beg*), while /e/ is heard by the majority in *paddle-pedal* (n=34 in *paddle*, n=38 in *pedal*). Both words in the *dad-dead* pair seem to be associated with distinct vowels, meaning that there is almost no substitution by the participants. Nonetheless, *Brad-bread* demonstrates a high degree of variation and slight substitution. The results in the pre-lenis can be attributed to both phonetic properties of the vowels before lenes (cf. section 3.2), as well as the effects of the visual stimuli, i.e. the bias towards nouns as opposed to adjectives, verbs and proper names.

8.1.2 Pre-nasal environment

A significantly higher degree of variation and misperception in responses for /æ/ and /e/ before nasal consonants can be observed in the figure below:

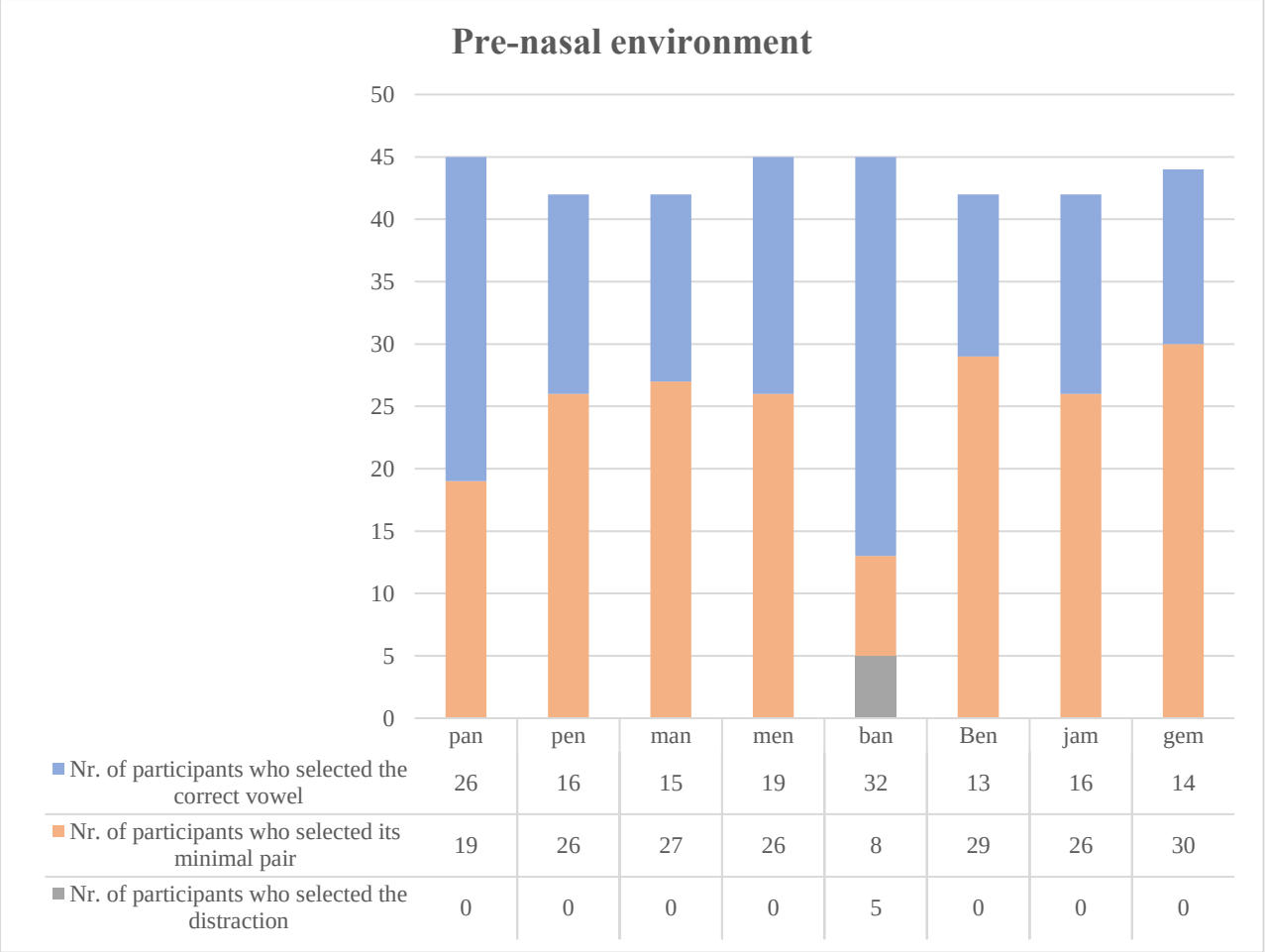


Figure 23. Results of Task 1 for /æ/ and /e/ in pre-nasal environment. The x-axis shows the words in question, while y-axis shows the number of participants.

Vowel /æ/ is heard correctly in *ban*, with only 8 participants hearing /e/ and 5 clicking on a distraction, unveiling a potential bias towards nouns, as opposed to proper names in the visual stimuli (*Ben*). It is followed by *pan*, which exhibits a high degree of variation (n=26 heard correctly, while n=19 heard /e/). In *man* and *jam*, the vowel is heard by the majority of participants as /e/ (n=27 for *man;* n=26 for *jam*). Unlike the tokens in pre-lenis position, the words used here are mostly nouns, and therefore only the bias against proper names in *ban-Ben* can be taken into consideration. In other words, the results were more likely to be influenced by the phonetic properties of the vowels. As discussed earlier, the general raising of /æ/ towards /e/ and the lowering of /e/ towards /æ/ before nasals was observed, meaning that the vowels in question are perceptually close before /n/ and /m/, as shown in the table from section 3.2:

Table 9. Mean F1 and F2 comparisons of /e/ and /æ/ in pre-nasal position.

Word	F1 in Hz	F2 in Hz	Word	F1 in Hz	F2 in Hz
pan	740,518	2352,179	*pen*	789,297	2006,909
ban	684,852	2488,592	*Ben*	726,176	2000,291
jam	773,406	2044,146	*gem*	786,002	1846,352
man	731,029	2111,712	*men*	804,243	2112,452

The potential overlap of /e/ and /æ/ in terms of quality before nasals has clearly left the participants relying on duration as the only distinguishing factor between the vowels, particularly in *pan* (t=0,230 s), in contrast to *pen* (t=140 s). Vowel quantity, however, has proved to be an insufficient cue in *jam* and *man*, and has not aided the participants in discriminating between the vowels.

In all the words containing /e/ (*pen, men, Ben,* and *gem*), the vowel was identified more often as /æ/. Despite the potential bias against proper names in *Ben*, the acoustic features of /e/ before /n/ and /m/ cannot be disregarded. Coupled with the heightening of /æ/, the subtle lowering of the vowel may account for the majority of participants hearing /e/ as /æ/. Furthermore, vowel length does not have an impact on the results (t=0,055-0,140 s). For example, /e/ in *gem* (t=0,055 s) is remarkably shorter than /æ/ in *jam* (t=0,224 s), but the word *gem* as heard as having /æ/ by the vast majority of participants (n=30). In such words, the fact that /æ/ is higher and /e/ is lower than they should normally be may have lead the participants to hypercorrecting in their perception, and identifying /æ/ more often, because the vowel stands out by deviating from its normal F1 values.

When minimal pairs are taken into consideration, *pan-pen* and *ban-Ben* demonstrate that, in both minimal pairs, /æ/ is more often identified than /e/ (n=26 in *pan,* n=26 in *pen,* n=32 in *ban,* n=29 in *Ben*). The vowel in *pan* is raised and significantly longer than the vowel in *pen*. The fact that the noun *ban*, despite being abstract, is still preferred more than the proper name *Ben* shows a bias towards nouns. In contrast, in pairs *man-men* and *jam-gem*, there is a higher degree of vowel substitution and variation among participants, which may have been the result of hypercorrection and the participants not trusting their own perception. For instance, 27 participants heard /e/ in *man*, while 26 heard /æ/ in *men*. Likewise, /e/ was heard in *jam* by 26 participants, whereas /æ/ was heard by 30 participants in *gem*. Therefore, in some minimal pairs, /æ/ is more frequently heard, while the vowels are not properly discriminated in others.

8.1.3 Pre-fortis environment

Both /æ/ and /e/ are generally not well identified before fortis consonants by BCS listeners, with only two exceptions, as shown in the following chart:

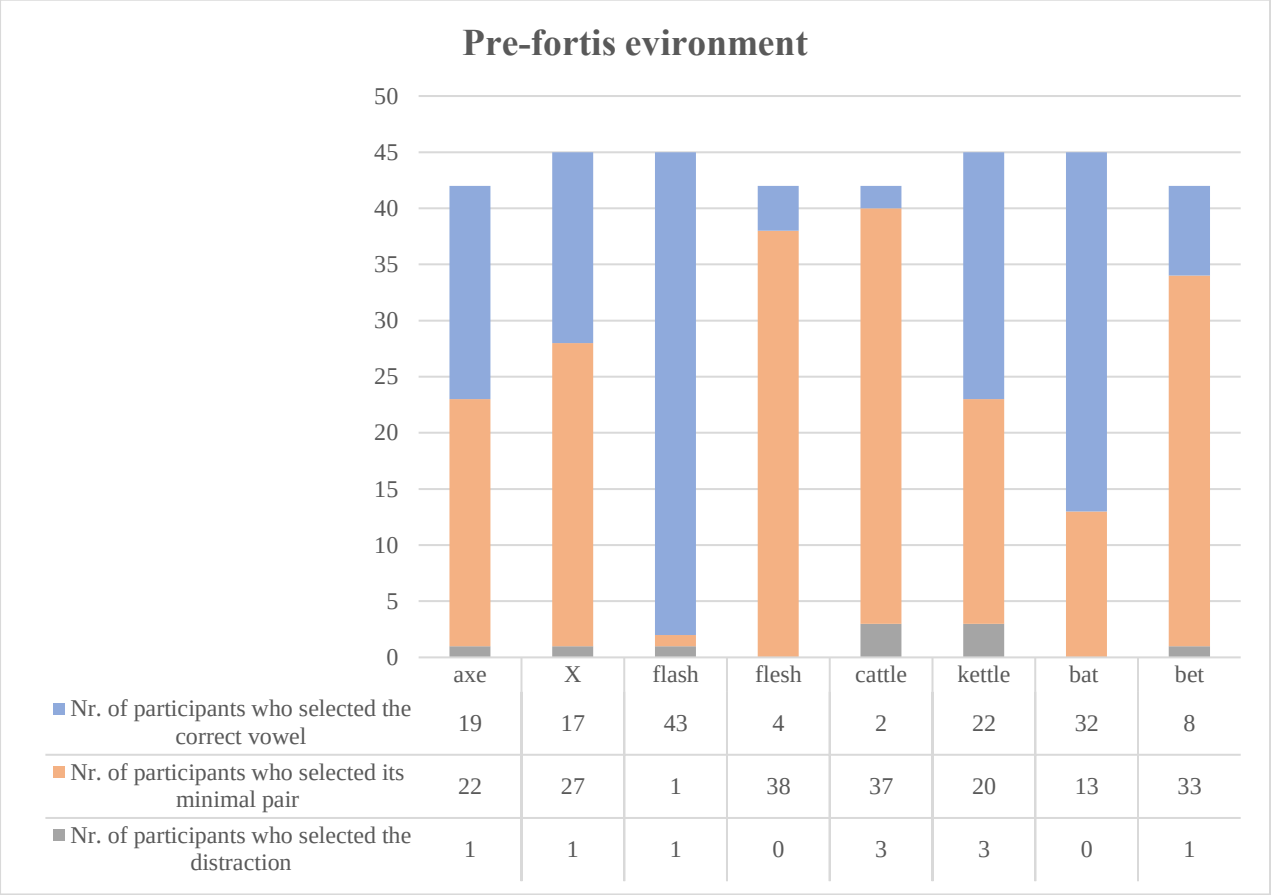


Figure 24. Results of Task 1 for /æ/ and /e/ in pre-fortis environment. The x-axis shows the words in question, while y-axis shows the number of participants.

In the words that contain /æ/ (*axe, flash, cattle, bat*), the vowel is correctly heard by most of the participants in *flash* (n=43), and *bat* (n=32). The high number of correct vowel identifications for *bat* may have been influenced by the nature of the visual stimuli, i.e. the participants' preference for nouns. The vast majority heard /e/ in *cattle* (n=37), while the highest degree of variation is detected in *axe* (n=19 for /æ/, n=22 for /e/). Since the quality and quantity of /æ/ were greatly impacted by the fortes that follow, BCS respondents seem to have relied heavily on both spectral and temporal cues (i.e. quality and quantity), as already discussed in section 3.2. In comparison to pre-lenis and pre-nasal position, the vowel is lowered in *cattle, axe* and *bat,* and significantly lowered in *flash* (F1=1091 Hz), which explains why the vowel was correctly heard by the majority in *flash*: the change in F1 has made it the most acoustically distant realisation from the BCS vowel /e/ (F1= 500-700 Hz). In terms of vowel

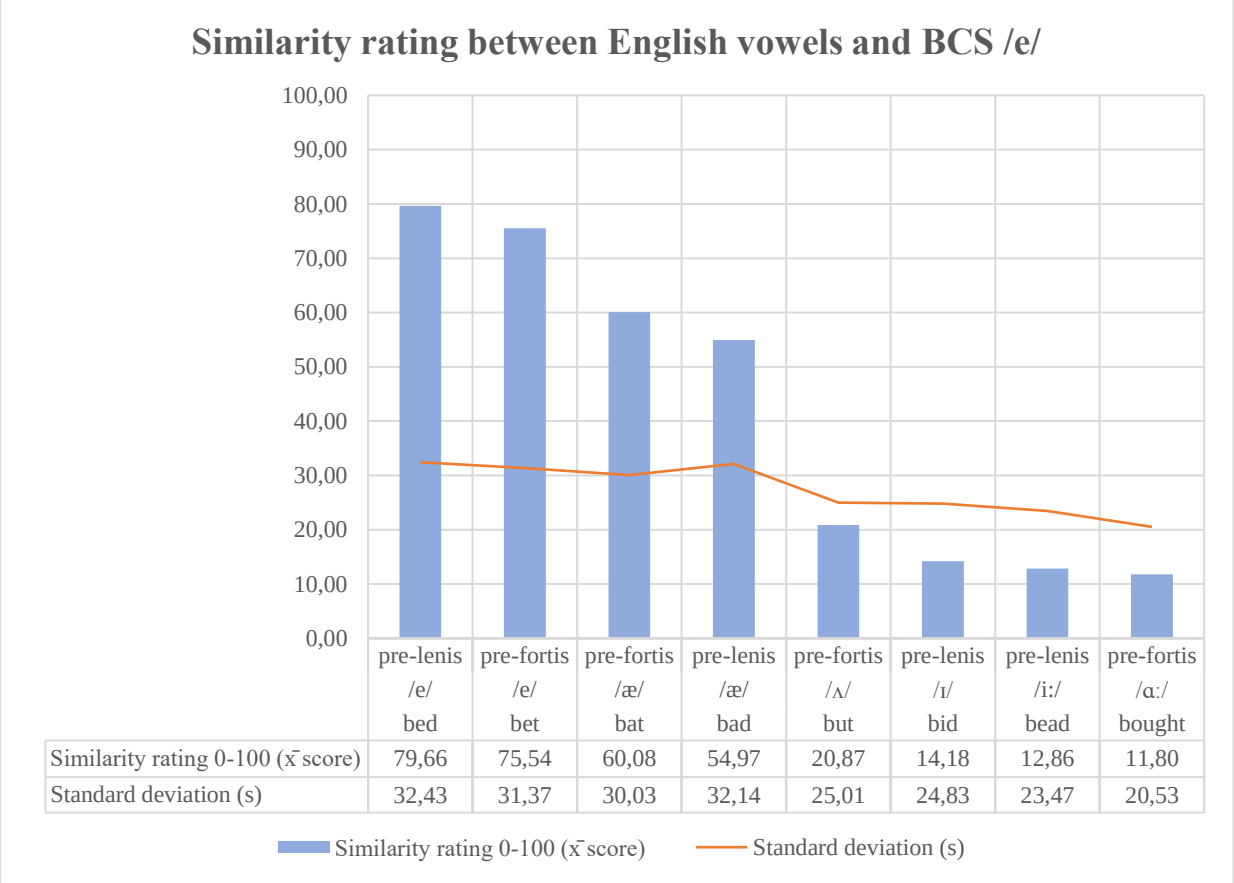
length, the reduced duration of /æ/, as the effect of the following fortes, may have diminished the participants' reliance on temporal cue as a distinguishing factor between /æ/ and /e/ and therefore influenced the results. This is particularly evident in the fact that /æ/ in *axe* is only slightly longer than its pair in *X* (t=0,141 s vs. t=0,112 s), hence the high degree of confusion and the hearing /eks/ (n=22) in place of /æks/ (n=19).

As before nasals, /e/ is frequently misheard as /æ/ in pre-fortis position (*X, flesh, kettle, bet*). The only exception is the word *kettle*, where the number of participants who heard the vowel correctly prevails (n=22), but it is closely followed by the number of those who heard /æ/ (n=20). As previously mentioned, the pre-fortis environment does hot have a notable impact on quantity of /e/. For example, /e/ in *bet* is longer than /æ/ in *bat* (t=0,122 s vs. t=0,104 s), but both tokens are heard by the majority as having /æ/. Nevertheless, in the case of *bat-bet*, the effects of the visual stimuli and the preference for nouns cannot be disregarded. In this case, the participants may have been biased towards *bat,* as it is a noun and is represented as more concrete and iconic than the abstract verb *bet*.

If minimal pairs are taken into consideration, *flash-flesh* and *bat-bet* are heard by the majority as having /æ/ (n=43 in *flash*, n=38 in *flesh*, n=32 in *bat*, n=33 in *bet*). In *flash-flesh*, /æ/ is heard more often because of its high degree of deviation from the native phoneme /e/, and the participants heard the vowel in *flesh* due to hypercorrection. Furthermore, *bat-bet* demonstrate the bias towards nouns, as opposed to verbs in the visual stimuli. The other two pairs, namely *axe*-X and *cattle-kettle* show even a higher degree of substitution of the vowels and the inability to properly discriminate between the members of the vocalic contrast. In *axe*, the majority heard /e/ (n=22), and in *X*, the majority heard /æ/ (n=27). In *cattle*, mostly /e/ was heard (n=37), while the results for *kettle* demonstrate that both vowels are almost equally good candidates (n=22 for /e/, n=20 for /æ/). Similarly to *flash-flesh*, the high number of selections of /æ/ in words with /e/ may be attributed to hypercorrection, and the participants opting for the acoustic stimulus that is heard as more distant from the L1 phoneme.

8.2 Results of Task 2: Similarity ratings between English vowels and BCS /e/

Following the data collection process, all 87 individual ratings of the 8 tokens tested were extracted from QuestionPro, and their $\bar{x}$ score and standard deviation were calculated and generated in the following chart:



	pre-lenis /e/ bed	pre-fortis /e/ bet	pre-fortis /æ/ bat	pre-lenis /æ/ bad	pre-fortis /ʌ/ but	pre-lenis /ɪ/ bid	pre-lenis /i:/ bead	pre-fortis /ɑ:/ bought
Similarity rating 0-100 (x̄ score)	79,66	75,54	60,08	54,97	20,87	14,18	12,86	11,80
Standard deviation (s)	32,43	31,37	30,03	32,14	25,01	24,83	23,47	20,53

Figure 25. Similarity rating between English vowels and BCS /e/. The x-axis shows the words, vowels, and their environments, while y-axis shows the values from 0 to 100. Mean score (x̄) values are represented by the blue columns, while the standard deviation (s) values are represented by the orange line. The exact s and x̄ values are given in the table under the chart.

As predicted, English /e/ is rated as the best exemplar of the BCS /e/ category, according to the mean x̄ scores. The highest degree of similarity is perceived in pre-lenis, i.e. pre-/d/ position (x̄=79,65), closely followed by the pre-fortis context (x̄=75,54). In other words, the slight lowering of the vowel before /t/ (cf. section 6.5), has not led to significantly different ratings of *bed* and *bet*, and the vowel can be regarded as an optimal candidate for the native category /e/ in both contexts. The quantity of English /e/ before /t/ (t=0,138 s) may have had a subtle influence on the high ratings of *bet*, given that the vowel is longer before /t/ than before /d/ (t=0,128 s). Furthermore, the high ratings of similarity between /æ/ and BCS /e/ were also expected, particularly in pre-fortis position (x̄=60,08). Despite the extremely lowered position of the vowel before /t/, making it spectrally distant from the native BCS /e/ in terms of F1, the vowel still received a high score. The vowel was also rated as relatively similar to BCS /e/ in its pre-lenis position (x̄=54,96), with a slightly bigger difference between pre-fortis and pre-lenis position (Δx̄=5,11) than /e/. In this case, it is plausible to assume that, apart from quality, the length of /æ/ before /t/ (t=0,190 s), as opposed to pre-lenis duration (t=0,285 s), has

influenced the ratings, meaning that the lower x̄ score before /d/ is explained by its increased quantity. Surprisingly, /ʌ/ in *but* was rated more similar than expected (x̄=20,87), which is a direct result of the vowel's spectral closeness to English /e/, i.e. tongue fronting during articulation. In contrast, /ɑː/ was rated as the least similar to BCS /e/ (x̄=11,80), regardless of its proximity to /ʌ/. Equally interesting is the fact that both /ɪ/ in *bid* (x̄=14,18) and /iː/ in *bead* (x̄=12,86) received a higher mean score than /ɑː/. Here, /iː/ was rated as less similar to BCS /e/ due to its significant fronting and raising, as well as duration (t=0,265 s), as opposed to /ɪ/, which was spectrally and temporally closer to English and BCS /e/.

Nevertheless, the computed standard deviation demonstrates a high degree of data dispersion in the sample, particularly with reference to English /æ/ and /e/ in both pre-lenis and pre-fortis position (s>30). The highest standard deviation was detected in bed (s=32,43) and bad (s=32,13), closely followed by bet (s=31,36) and bat (s=30,03). In other words, in spite of the words having similarity x̄>50 to the BCS /e/ and the vowel in bed being the best candidate for the native category, the respondents were not as unitary in their responses, particularly in pre-lenis environments. Moreover, a higher degree of data dispersion in relation to the mean score x̄ was expected in *but* (s=25,01), *bid* (s=24,82), *bead* (s=23,46), and *bought* (s=20,52). The lowest SD in *bought* shows that the participants were more uniform in their responses. To further illustrate variation among participants, 10 interesting individual responses were chosen and tabulated:

Table 10. 10 individual answers to Task 2 category goodness rating (CGR).

Participant ID	**Symbol**	**/bɑːt/**	**/bet/**	**/bæd/**	**/biːd/**	**/bæt/**	**/bɪd/**	**/bʌt/**	**/bed/**
85074804	A	0	55	51	48	51	49	51	0
85125947	B	0	100	0	0	100	0	0	50
97726799	C	0	81	0	0	0	0	0	83
97727055	D	0	100	0	0	0	0	0	0
98247786	E	8	70	91	5	40	12	13	4
96439502	F	1	28	1	0	39	1	3	35
96436349	G	100	95	100	0	100	0	100	100
102110289	H	0	99	68	3	71	0	13	100

102153396	I	0	100	91	0	96	0	0	100
101898906	J	31	99	77	34	91	12	19	99

Taking individual participants into account, only G rated /ɑː/ as fully similar to the native category (CGR=100). Interestingly, all the vowels apart from /ɪ/ and /iː/ were almost equally good candidates for BCS /e/ for this respondent. Furthermore, respondents B, C, and I mostly gave 0 CGR score to all vowels apart from English /e/ and /æ/, which were given high scores, albeit with some exception. For instance, /e/ before fortis /t/ was judged as a better candidate for BCS /e/ than in the lenis position by B and C. Out of these three, only B perceived similarity between /æ/ and BCS /e/ in *bat*. Participant I rated highly (CGR>90) both /æ/ and /e/, regardless of the following consonant. For participant D, only English /e/ in *bet* was judged the same as BCS /e/ (CGR=100). Similarly, participant H gave extremely low ratings to all vowels, except /e/ and /æ/, but /e/ was perceived as a better exponent of BCS /e/ (CGR=99 and CGR=100 for /e/ vs. CGR=68 and CGR=71 for /æ/). This participant also perceived some similarity between /ʌ/ in *but* and the L1 vowel, which can be justified by F1 closeness of the English vowel to both English and BCS /e/. Furthermore, participants A and E gave higher ratings to almost all the words than they gave to *bed*. Participant A rated /e/ before /t/ and /æ/ in both environments as similar to BCE /e/, while E rated /æ/ in *bad* (CGR=91) as more similar for BCS /e/ than English /e/ in *bet* (CGR=70) and *bed* (CGR=4). Finally, participant J gave relatively high ratings to all vowels, with /ɪ/ being the least similar (CGR=12) and /e/ being the most similar to BCS /e/ (CGR=99 for *bet* and *bed*, followed by CGR=91 for *bat*). In contrast, participant F rated poorly all the vowels, but gave their highest ratings to the vowels in *bet* (CGR=28), *bat* (CGR=39), and *bed* (CGR=35). Of course, the examples used cannot possibly account for the whole sample of 87 participants, but they do serve to illustrate the high degree of data dispersion and variation among the respondents in Task 2, particularly for /e/ in *bed*, whose SD was the highest (s=32,43). The complete dataset for Task 2 be found in Appendix G.

8.3 Discussion: Acquisition routes according to PAM

The results of Task 1 and Task 2 have shown that the discrimination between the English front vowels /æ/ and /e/ by BCS listeners is on a poor-to-intermediate level, meaning that both vowels are assimilated into BCS /e/ category. Why does this happen? According to the Perceptual Assimilation Model, both members of the L2 contrast are assimilated into BCS /e/ on the basis of their similarity to the native category, that is, the similarity in articulatory movements that are translated from the acoustic signal (Krebs-Lazendić & Best 2007: 282).

Indeed, the overall analysis of both English (cf. section 3.2) and BCS vowels (cf. section 4.2.2) has demonstrated that /e/ and /æ/ are partially close to the BCS vowel /e/ in terms of quality, but the perceived spectral (F1 and F2) and temporal (durational) proximity of English /e/ to the L1 category have resulted this vowel being assimilated in a higher rate than the other member of the contrast, making it the better exemplar of the native category. Furthermore, as shown in the results of Task 1, /æ/ is more frequently heard by the participants, and even the words that contain /e/ are more often identified as having /æ/, primarily due to hypercorrection. When the participants hear /e/, the vowel is closer to the native /e/, making the participants think that they have possibly heard a different sound and, as a consequence, they opt for /æ/. In other words, the participants are choosing a more 'foreign' vowel because they are aware that they are listening to words in a foreign language, meaning that they cannot trust their own perception. The results of Task 2 have shown that /æ/ is indeed more deviant from BCS /e/, since it has received high similarity ratings to the L1 vowel, but still lower than English /e/. When the analysis of quality and quantity of the vowels in section 3.2 in taken into consideration, English /e/ is more stable in terms of quality, and its quantity is not impacted by the following consonants. In contrast, /æ/ is more acoustically deviant from BCS /e/, due to its changing quality and vowel length before lenes, fortes, and nasals. This difference between the members of the English contrast is particularly evident in the results of Task 2, where /e/ in *bet* and *bet* has received significantly higher category goodness ratings than /æ/ in *bad* and *bat*, which were still highly similar to BCS /e/. However, the already reported effects of the visual stimuli need to be taken into account: the results were also influenced by the participants' preference for nouns and words that are phonologically similar to L1, and their bias against adjectives, verbs, and proper names.

With reference to the patterns of assimilation proposed by Best in section 2.2 (Best 1994; Best 1995; Best & Tyler 2007; Strange & Best 1992; Hallé & Best 2007), the findings of this study indicate that, for the vast majority of BCS respondents, the assimilation of /æ/ and /e/ happens in the Category-Goodness pattern, confirming the initial hypothesis. As stated in in section 6.2, CG type assimilation was predicted as most likely, because of the expected poor-to-intermediate discrimination between the vowels and the perception of both vowels as instances of BCS /e/, with English /e/ being a better candidate. Nonetheless, auditory perception is highly subjective, and "listeners themselves may differ from one another, even within the same language, with respect to which phonetic properties of a non-native phone they may detect or attend to in perception" (Best & Strange 1992: 307). Furthermore, as discussed in sections 3.1

and 3.2, phonetic properties of English vowels have a tendency to change depending on the environment, which also contributes to the difference among the same group of speakers of a single language. Therefore, this section will continue by discussing the PAM patterns individually.

8.3.1 Category-Goodness (CG type) assimilation pattern

The initial hypothesis in this study was that the CG type assimilation, out of all, is most likely to occur, given that /æ/ is more deviant from the L1 vowel than /e/, but both are still quite similar to the BCS category /e/. As already mentioned, Category-Goodness or CG type assimilation predicts that both members of the L2 contrast will be assimilated into a single category, but will differ in goodness-of-fit into the native category (Best 1995: 195). The sound that is in the closest phonetic proximity to the L1 sound will be perceived as "a good instance of the native category", whereas the sound that is slightly perceptually distant will be judged as "a poor instance of that *same* category" [my emphasis] (Strange et. al. 2001: 1692). Both Task 1 and Task 2 have demonstrated that, the results for the vast majority of the participants have been a direct product of CG type assimilation. To provide a more thorough account of the assimilation pattern in this study, the following illustration was created:

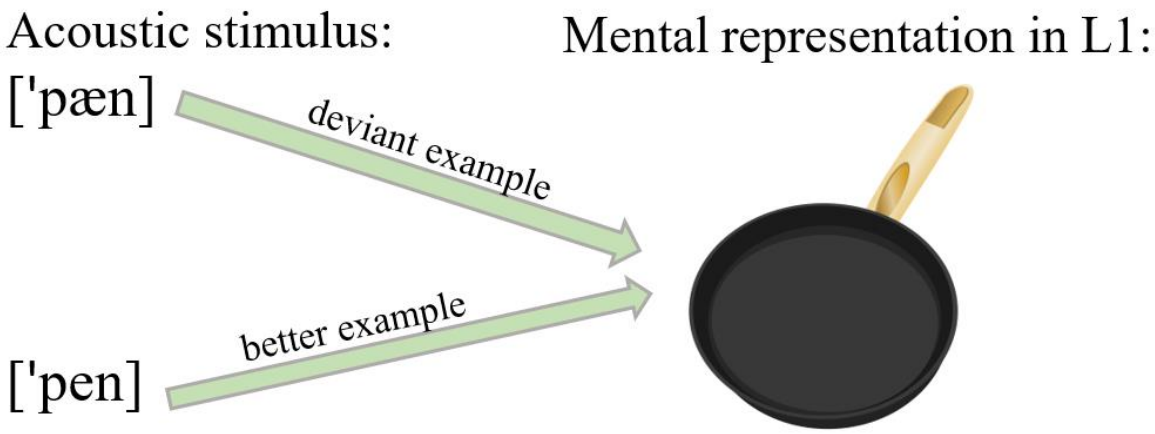


Figure 26. Illustration of the Category-Goodness type assimilation, based on the results of Task 1 and Task 2. The illustration shows how an acoustic stimulus is translated into the mental representation of the minimal pairs.

The above figure shows the perception of vowels in *pan* and *pen* in Task 1, where /æ/ was successfully selected by the majority (n=26), while the same number of participants heard the vowel in *pen* as /æ/. The *pan-pen* pair was chosen as to avoid the previously discussed effects of the visual stimuli. In this case, it is plausible to assume that both *pan* and *pen* are

phonologically represented as /pen/, but the vowel in the word is the native BCS vowel /e/. Since English /e/ is more similar to the native category and the mental representation in L1, it was not as frequently identified as /æ/, which is distant from the native /e/. The vowel /æ/ was heard more often in *pen* due to hypercorrection: the participants were aware that they were listening to words in a foreign language, and therefore opted for a more 'foreign' vowel out of the two. Even when they heard /e/, they thought they heard a different, more deviant sound, meaning that they could not really trust their own perception. This conclusion is further corroborated by previous literature, as category goodness difference between /e/ and /æ/ has had an impact on the success rate of vowel discrimination, resulting in the 'deviant' vowel being recognised more frequently (Wrembel, Marecka & Kopečková 2019: 515). Moreover, in Task 1, CG type is easiest to detect in pre-lenis environments, because /æ/ remains somewhat stable, i.e. does not drastically change its quality before lenes and is not approaching English /e/, but its duration is significantly increased, making it the primary cue for the listeners. As we move from pre-lenis to pre-nasal and pre-fortis position, /æ/ is still more often heard than /e/, albeit with more difficulty. The data in this study is aligned with the results by Marković and Jakovljević (2016: 220), who found that "the vocalic duration plays a significant role in the perception" of /æ/ versus /e/ contrast, as the highest number of misperceptions have occurred in pre-fortis position, thus eliminating vowel length as the primary distinguishing cue.

Furthermore, the results of Task 2, i.e. category-goodness rating, clearly elucidate the CG type assimilation, which could not be solely justified by minimal pair distinction. For the majority of participants in this study, /e/ is judged as the best candidate for BCS /e/, and a better exemplar of the native category than /æ/, in both pre-lenis and pre-fortis context ($\bar{x}$=79,65 for *bed*, $\bar{x}$=75,54 for *bet*). Apart from the mean score values, /e/ has the highest number of very good (CGR>75) and optimal (CGR>90) category goodness ratings, for both *bed* (n=64 for CGR>75; n=56 for CGR>90) and *bet* (n=57 for CGR>75; n=49 for CGR>90). In other words, more than a half of BCS respondents rated /e/ as highly similar to their native category. The other member of the contrast /æ/ is still assimilated into the native category but with lower goodness-of-fit ratings ($\bar{x}$=60,08 in *bat*, $\bar{x}$=54,96 in *bad*), making it a less optimal example of BCS /e/. Taking raw numbers into account, approximately a third of participants gave this vowel a high category goodness rating (n=37, CGR>75 for b*at*, n=27, CGR>75 for *bad*). Nevertheless, /æ/ is still a better exemplar of the native vowel than other English vowels tested in Task 2, such as /ɪ/ and /i:/, which are also acoustically (primarily in terms of F2) similar to BCS /e/. These findings align with Krebs-Lazendić's doctoral thesis (2007) and the subsequent paper co-authored with

Best (Krebs-Lazendić & Best 2008), which have both unveiled assimilation of Australian English /e/ and /æ/ into BCS /e/, with English /e/ being perceived as a better instance of the native category. Finally, with reference to L2 category emergence, the results for /æ/ in both tasks stipulate that the category formation "is reasonably likely to be formed" for this 'deviant' sound, whereas a new category for /e/ is unlikely to be formed and the sound would continue to "be perceived as phonologically and phonetically equivalent to the L1 category", meaning that the separation of the L2 phoneme from the source phoneme is quite unlikely (Best & Tyler 2007: 29).

8.3.2 Single-Category (SC type) assimilation pattern

In the initial hypothesis, the SC type was predicted as less likely than CG, particularly due to the different phonological contexts of the vowels. According to Best (1995: 195), SC type makes the discrimination between /e/ and /æ/ almost impossible, since both sounds are perceived as either "equally acceptable" or "equally deviant" instances of the native category. Moreover, SC type makes the detection of subtle acoustic details that help the listeners distinguish between the sounds quite strenuous (Levy & Strange 2008: 5). In other words, both /æ/ and /e/ would be heard as allophones of BCS /e/, and, as a consequence, "minimally contrasting words would be perceived as homophones" (Best & Tyler 2007: 29). If this were true, then /æ/ would get continuously substituted by /e/ in Task 1, and vice versa. Nevertheless, the Single-Category assimilation pattern was detected in the answers of a smaller number of participants, and the logic behind the assimilation is shown in the following illustration:

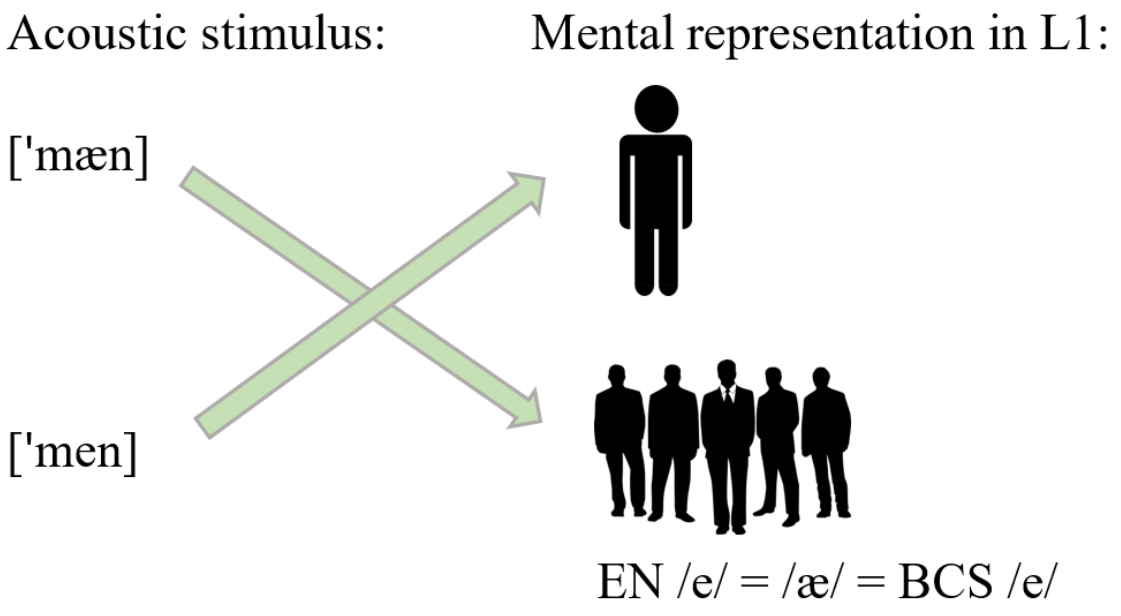


Figure 27. Illustration of the Single-Category type assimilation, based on the results of Task 1 and Task 2. The illustration shows how an acoustic stimulus is translated into the mental representation of the minimal pairs.

As observed, the SC type assimilation postulates complete and continuous substitution of the members of the same contrast, i.e. *man* is associated with the plural and *men* with the singular form on the same noun, indicating that the minimal pairs are perceived as homophones (Best & Tyler 2007: 29). Moreover, both vowels, regardless of the phonological context, are given equally high or low similarity ratings in Task 2 (CGR>50 or CGR<25 and CGR=0). When looking at individual participants, there were only two instances of complete substitution of the vowels in Task 1 and continuously high ratings for both vowels in Task 2 (ID: 102058202, 101899064). These participants heard /e/ in all the words that contain /æ/, and vice versa, proving full SC type assimilation. Also, this pattern of assimilation is the best example of the participants not being able to trust their own perception, as they consistently hear one vowel as the other. There were no participants who both switched /æ/ and /e/ in Task 1 and rated both vowels as equally bad examples of BCS /e/ in Task 2.

For the participants that fall under the SC type, the continuous substitution and equal assimilation were undoubtedly influenced by the quality and quantity of the vowels. As discussed in section 3.2 and in the results of Task 1, /æ/ is raised towards /e/ and /e/ is lowered towards /æ/ before nasals, meaning that the vowels are in approximate position to one another and the difference between them is almost impossible to hear, hence the substitution. In this case, SC type is caused by vowel quality becoming an unreliable cue in perception. Also, the participants were mostly not able to rely on duration as a distinguishing cue. Section 3.2 has also shown that, when the vowels are placed before fortes, /æ/ undergoes quantity reduction, diminishing the reliance on duration as the primary cue, and becoming similar to /e/. Interestingly, SC type is not as pronounced in pre-lenis environments, with only *Brad-bread* as an exception for six participants. This can also be attributed to the effects of the visual stimuli and the observed bias against proper names.

8.3.3 Uncategorised-Categorised (UC type) assimilation pattern

Predicted as least likely, the UC type presupposes a complete assimilation of one member of the L2 contrast, while the other sound is assimilated somewhere in the phonetic space in L1, outside of any known native category (Best 1995: 195). The discrimination between the sounds is expected to be optimal, as it is essentially a difference between a foreign and a native phoneme (Best & Tyler 2007: 23). This assimilation pattern is exemplified in the following:

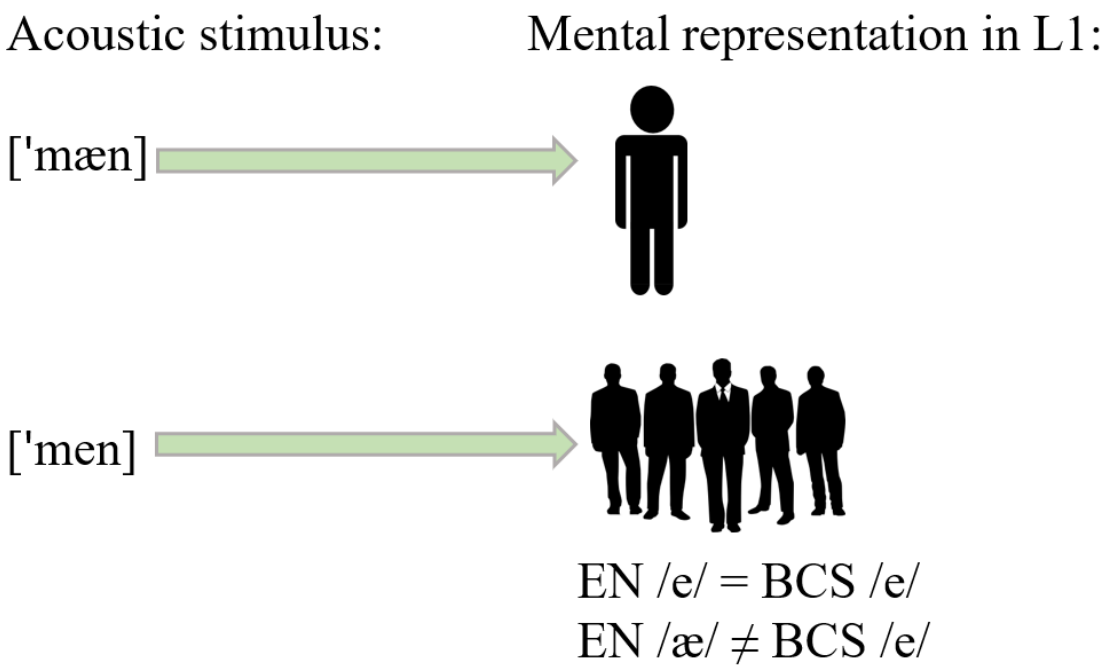


Figure 28. Illustration of the Uncategorised-Categorised type assimilation, based on the results of Task 1 and Task 2. The illustration shows how an acoustic stimulus is translated into the mental representation of the minimal pairs.

According to the UC type, both /æ/ are /e/ distinguished successfully and perceived as separate from one another in Task 1, i.e. the minimal pairs are not homophones, and /e/ is regarded as an instance of BCS /e/ (CGR>75), while /æ/ receives extremely low scores (CGR<20 or CGR<10). In this study, there are only two participants who assimilate fully according to the UC Type (ID: 85084334, 85086730). The participants in question discriminated between all the tokens in Task 1, and also rated /e/ as an excellent example of the native category, whilst /æ/ was regarded as deviant. However, /i:/ and /ɪ/ received high category goodness ratings, meaning that /e/ is not the only English sound undergoing assimilation. Another participant (ID: 97726799) falls under the scope of the UC type, but they perceived /e/ as /æ/ in a single word, meaning that they also tend to hypercorrect their perception, but still distinguish between the vowels. Therefore, only two, potentially three, BCS respondents (N=87) are encompassed by the Uncategorised-Categorised pattern, and can be regarded as the only participants who can actually trust their own perception. However, the UC type needs to be approached with extreme caution, because it is impossible to pinpoint where exactly /æ/ is assimilated, and if one can talk about assimilation at all.

9. Bringing it all together: Assimilation of /e/ and /æ/ into BCS /e/, exposure to English, and (non-)conformity to native-speaker norms

The main focus of this thesis was investigating the perception of /e/ versus /æ/ by native speakers of Bosnian/Croatian/Serbian and explaining the results using the Perceptual Assimilation Model (PAM). In addition, from a sociolinguistic perspective, this study also examined the relationship between assimilation patterns according to PAM and exposure to English, as well as the interrelatedness of (non-)conformity to native-speaker norms and assimilation into the native category /e/. Having reported and discussed the results of both auditory perception tests and statement ratings, this section will combine the phonetic/phonological and sociolinguistic part of this study, in order to provide an answer to the additional research question.

First and foremost, the results of the perception experiments have shown that, for the vast majority of participants, English front vowels /e/ and /æ/ are assimilated into the native category /e/, with English /e/ judged as a better exemplar of the BCS 'counterpart'. Furthermore, /æ/ is identified successfully more often, and is even frequently heard in the tokens containing /e/ due to hypercorrection. Since it is heard as more distant to BCS /e/ in terms of F1 and F2, the participants tend to select this vowel in their answers in Task 1. In other words, the participants cannot trust their own perception and will opt for a more 'foreign' sound even when they hear English /e/, because English /e/ is just too close to their own native vowel, and they are aware that they are listening to words in a foreign language. For instance, participants commonly associated both [pæn] and [pen] with the image of a pan, indicating that one vowel is successfully heard, while the other, as a consequence of high degree of similarity to the native phoneme and hypercorrection, is not. Therefore, based on Task 1 and Task 2, the majority of participants assimilate English vowels on the basis of the Category-Goodness assimilation pattern (Best 1994; Best 1995; Best & Tyler 2007; Strange & Best 1992; Hallé & Best 2007).

As already mentioned in section 2.3, L2 experience and exposure have a great impact on assimilation of target sounds into L1 phonological repertoire, since "increased L2 experience [fosters] improved recognition of the discrepancies between L1 and L2", resulting in rephonologisation of the assimilated sounds and the subsequent emergence of distinct categories (Best & Strange 1992: 307). Increased experience with the target language results in better discrimination possibilities between L2 sounds and allows for detection of subtle

acoustic nuances that separate one phoneme from another (Kopečková 2012: 237). Moreover, the listeners "may begin to perceive segments on a more abstract level, less affected by acoustic variation" (Levy 2009: 1140). Nevertheless, statement ratings regarding experience with English and the results of auditory perception tasks have shown that this is not entirely the case. Despite years of exposure to English in school (EFL), the relatively frequent use of English for both professional purposes and leisure (especially exposure to spoken English in media), BCS listeners have not learned to completely separate /e/ and /æ/ from one another and did not establish proper L2 categories for these English vowels, hence the assimilation into BCS /e/ on the basis of category-goodness. Moreover, in spite of the exposure to English since childhood, BCS listeners are not able to trust their own perception and still heavily separate English from their mother tongue, which is apparent in the fact that /æ/ is identified more often in both the tokens that contain /æ/ and /e/. The listeners are aware that they are listening to words in a foreign language, and even if they hear /e/ in a word, they still think they have heard a different vowel, and opt for a sound that is more 'foreign' and distant from what they have in their L1, hence the pervasiveness of /æ/, i.e. the vowel that is a more deviant exponent of the L1 /e/.

The almost full assimilation of English /e/ and misperception across the contexts has elucidated that phonological awareness/phonological training is also needed. According to Vilke (1991: 487), the only way to properly distinguish between L2 phonemes and separate them from the native repertoire is to instruct pupils to "identify distinctive features of L2 phonemes", i.e. develop phonological awareness in the classroom. It would be plausible to assume that, with the proper phonetic training, the effects of the CG type assimilation would diminish. That way, the sound that is judged as a better example of the L1 category, which in this case is /e/, would eventually be established as a separate category (Best & Tyler 2007: 29). Furthermore, phonological awareness, combined with experience, would undoubtedly facilitate more successful perception of both /e/ and /æ/ across different phonological environments, which was shown to vary primarily in pre-nasal and pre-fortis contexts in Task 1, thus decreasing the impact of acoustic and regional variation. Moreover, it can be assumed that increased exposure to L2 and phonological training would decrease the tendency for hypercorrection, making the participants more confident in their own perception abilities. In the future, it would be interesting to examine the impact of phonological awareness on the CG type assimilation of /e/ and /æ/ into BCS /e/, and to see if results similar to Levy and Strange (2008) and Levy (2009) can be obtained. Moreover, given that the participants in this study were adult speakers who learned English as a part of formal instruction, it would be worthwhile to compare the

assimilation of /e/ and /æ/ between the learners with more and less English language experience, as in Kopečková's study (2012).

Nevertheless, the relevance of phonology, accent, and successful perception of vowels is questioned. From the statements concerned with experience and native-speaker norms, it is apparent that BCS listeners mostly identify as users of English as a lingua franca, for whom English fulfills a very specific purpose (business or leisure). In spite of the prolonged exposure to native-speaker norms in classroom settings (EFL) and NS speech in media (music, films, TV shows), English is still mostly used to communicate with other non-native speakers. As a result, the overall intelligibility in communication takes precedence over accent and learning about the phonology of L2, which is particularly evident in agreement with ELF-oriented statements, and, accordingly, disagreement with EFL statements (Jenkins 2009: 143). In this study, the participants can distinguish between /e/ and /æ/ to a certain degree, but the assimilation into BCS /e/ shows that establishment of categories is not a priority, which can be attributed to the fact that vowel quality is a non-core feature in the Lingua Franca Core (Newbold 2021: 399). In the cases of CG type assimilation, particularly in the contexts where words that contain /æ/ are selected when the words with /e/ are heard, it would be interesting to investigate the impact of assimilation on intelligibility and the potential strategies that are used in order to avoid miscommunication. Moreover, assimilation of two English vowels into a single BCS category can be understood as retainment of L1 properties in English, and developing English with regional aspects of the first lanugahe, which gives BCS speakers "the same sociolinguistic rights as those enjoyed by L1 speakers" (Zoghbor 2011: 287). Furthermore, assimilation into the BCS vowel category can be interpreted as the participants' way of 'subjecting' English phonemes/sounds to their native repertoire, which may stem from the increased nationalistic approach to linguistic identity in former Yugoslav states and the desire to preserve the language(s) and thus the ethnic identity (cf. Alexander 2006; Brown & Alt 2004).

The above conclusions were drawn from the data of the majority of the participants in this study. However, the discussion of auditory tasks has already demonstrated an immense degree of variation among participants and the attestation of more assimilation patterns than just CG, due to the highly subjective nature of perception of phonemes/phones and the differences between listeners in terms of acoustic cues they "attend to in perception" (Best & Strange 1992: 307). As a consequence, detailed results for a number of participants who assimilate according

to Category-Goodness, Single-Category, as well as Uncategorised-Categorised pattern will be taken into consideration.

First of all, the participants 102152638 and 102110289 both belong to the CG type assimilation: they identified /æ/ successfully across the different phonological environments, but continuously misperceived /e/ as /æ/. In terms of experience and NS norms, both participants used English for work and leisure (music and TV) and identified as speakers of English as a lingua franca. Because of that, they heavily disagreed with EFL statements, but expressed agreement with ELF-oriented statements. More precisely, for these CG respondents, mutual intelligibility (their own and that of others) is the forefront of successful communication in English, which is understood as a global language, and, listening and speaking skills should be disregarded in English language teaching. BCS language identity should be preserved, which is evident in both participants agreeing with the statement about unnaturalness of retaining original pronunciation of English words in BCS. Only participant 102110289 agreed with the statement that the attainment of NS accent is the only key to proficiency in English, which is a clear influence of prolonged exposure to EFL in the classroom.

Similarly, the two participants who assimilated /e/ and /æ/ equally well into BCS /e/ and fully substituted the vowels in perception, i.e. the respondents who fall under the SC type (101897747 and 102058202), reported a high degree of English-language exposure (102058202 also attended an English course, apart from school), meaning that the years of experience with L2 are irrelevant. It can only be assumed that, with proper phonetic training, SC will eventually evolve into CG, and finally into distinct categories (Kopečková 2012: 237). The participants who belong to the SC type also expressed negativity towards EFL statements and support for ELF statements. There is only one exception in the answers of participant 102058202, who exhibited fear of being judged by native speakers of English, which is a direct consequence of EFL in formal education.

Furthermore, the participants who assimilate in the UC pattern, that is, those who successfully discriminate between /e/ and /æ/, but still assimilate /e/ into the native category, demonstrate a high degree of variation. For example, participant 85084334 reported a high degree of experience with English through school, private, classes, and they had English as subject at university, and has spent time in countries where English is the *de facto* language. Such exposure to EFL in education has positive attitudes to NS norms as a consequence, but the participant still thinks that sounding like a native speaker is pointless when English is a global language, regardless of their judgements of others' accents. In contrast, participant 85086730,

who also belongs to the UC type, had negative attitudes to both EFL and ELF statements, signifying a complete disregard for native-speaker norms and English as a lingua franca.

Therefore, notwithstanding the amount of exposure to English, as well as the experience with the language in education, work, and private life, CG type assimilation still prevails for the majority of BCS respondents, and highlights the need for proper phonetic training and the raising of phonological awareness in the second language. This would undoubtedly diminish assimilation, hypercorrection, and increase the listeners' confidence in their own perception. At the same time, the attitudes and preferences of the speakers need to be taken into account, specifically in the context of the international communicative settings. Finally, despite the ability to generalise from the data, there is still variation among participants who speak the same first language, due to both auditory perception and attitude expression being highly subjective (Best & Strange 1992: 307). Moreover, the effects of the visual stimuli on the results and auditory perception cannot be disregarded.

10. Conclusion

The aim of this thesis was to examine the perception of the English vocalic contrast /e/-/æ/ by native speakers of Bosnian/Croatian/Serbian. More precisely, as a part of the larger Motor-Theory of speech perception, the Perceptual Assimilation Model (PAM) was used to investigate the assimilation of the two members of the English vocalic contrast into the native category /e/. To obtain the data on perception, two auditory perception tasks were employed. The first task was concerned with perception of words in isolation, i.e. minimal pair distinction, where the participants listened to acoustic stimuli containing either /e/or /æ/ and were instructed to select one of the four visual stimuli that best corresponds the tokens they heard. The purpose of the second task was to test category-goodness and, with it, determine the exact assimilation pattern according to PAM (Best & Strange; Best 1994; Best 1995; Best & Tyler 2007; Hallé & Best 2007). The participants were exposed to tokens in isolation and were asked to determine the degree of similarity between the vowels they heard and the native phoneme /e/. In addition to the phonetic/phonological part of the thesis, the relationship between exposure/experience and assimilation into the L1 category was explored, given the emphasis on the impact of this sociolinguistic feature on assimilation of target sounds into the native phonology in literature (Best & Strange 1992; Kopečková 2012; Levy & Strange 2008; Levy 2009). Moreover, the degree of (non-)conformity to native-speaker norms in SLA of phonology was taken into

consideration as well. The primary data collection method for the sociolinguistic part of the thesis were statement ratings (5-point Likert scale).

The results of the perception tasks have unveiled assimilation of both /e/ and /æ/ into the BCS category /e/, which, for the majority of 87 participants, occurs in the Category-Goodness assimilation pattern. English vowel /e/ is judged as a better exemplar of the native category, and is commonly misperceived as /æ/, while /æ/ is identified more often, even in the words with /e/, due to hypercorrection. The vowel /æ/ is too 'foreign' and distant from the native vowel that the participants, while being aware that they are listening to a foreign language, tend to choose the more deviant vowel. In other words, the participants cannot trust their own perception, and even if they hear English /e/, they still think they have heard a different sound. Other patterns of assimilation proposed by PAM (Best 1995: 195), such as Single-Category, and Uncategorised-Categorised were also detected, but in a small number of the participants. Moreover, the ratings of statements in the sociolinguistic part have elucidated that the CG type assimilation prevails, despite the years of prolonged exposure to English, be it education, work, or leisure. Furthermore, the relevance of fully mastering the phonology of English as a global language, i.e. 'sounding like a native speaker', is also questioned, as BCS respondents largely identify as users of English as a lingua franca, for whom the forefront of successful communication in English is their own intelligibility, as well as that of the others.

In conclusion, the findings in this thesis highlight the need for a slightly different approach to teaching of English phonology in BCS classrooms. In order to properly familiarise themselves with English phonology, as not to endanger intelligibility when communicating with NSs or NNSs, pupils need to be trained to attend to acoustic properties of English phonemes on a more abstract level, and develop phonemic awareness from the early stages in education. This way, the pupils would be more confident in their own perception and not resort to hypercorrection while listening to second-language speech. At the same time, the interests and preferences of the majority of BCS speakers of English, who identify as users of ELF, need to be taken into account as well, and decisions about English language teaching cannot not be solely based on the data of the minority studying English at university level and on the norms coming from the inner-circle countries. In other words, a higher degree of phonemic awareness is direly needed, not at the expense of BCS speakers as users of ELF, but to serve as means of amelioration of mutual intelligibility, which may be hindered by hypercorrection, assimilation, and the participants not trusting their own perception. The curricula in the English-language classroom would need to be structured in a way to achieve balance between the two.

11. Bibliography

Alexander, Ronelle. 2006. *Bosnian, Croatian, Serbian, a grammar with sociolinguistic commentary*. Madison (WI): The University of Winsconsin Press.

Best, Catherine; Strange, Winifred. 1992. "Effects of phonological and phonetic factors on cross language perception of approximants". *Journal of Phonetics* 20(3), 305-330.

Best, Catherine. 1994. "The emergence of native-language phonological influences in infants: a perceptual assimilation model". *Haskins Laboratories Status Report on Speech Perception* 108, 1-30.

Best, Catherine. 1995. "A direct realist view of crosslanguage speech perception". In Strange, Winifred (ed.). *Speech perception and linguistic experience.* Timonium (MD): New York Press, 171-204.

Best, Catherine; Tyler, Michael. 2007. "Nonnative and second-language speech perception: commonalities and complementaries". In Munro, Murray; Bohn, Ocke-Schwen (eds.). *Language experience in second language speech learning: in honour of James Emil Flege.* Amsterdam: John Benjamins, 13-45.

Bjelaković, Andrej. 2018. "Vokali savremenog standardnog britanskog izgovora i njihovo usvajanje kod izvornih govornika srpskog jezika [Contemporary standard British vowels and their acquisition by native speakers of Serbian]*".* PhD thesis, University of Belgrade.

Boomershine, Amanda. 2013. "The perception of English vowels by monolingual, bilingual, and heritage speakers of Spanish and English". In Howe, Chad; Blackwell, Sarah; Lubbers Quesada, Margaret (eds.). *Selected proceedings on the 15th Hispanic Linguistics Symposium*. Somerville, MA: Cascadilla Proceedings Project, 103-118.

Brown, Wayles; Alt, Theresa. 2004. *A handbook of Bosnian, Serbian, and Croatian.* Durham (NC): The Slavic and Eurasian Language Resource Center (SEELRC).

Čubrović, Biljana. 2017. "The acoustic analysis of phonetic duration: A case study of English and Serbian English". *Linguistica* 57(1), 59-72.

Faris, Mona; Best, Catherine; Tyler, Michael. 2018. "Discrimination of uncategorised non-native vowel contrasts is modulated by perceived overlap with native phonological categories". *Journal of Phonetics* 70, 1-19.

Flege, James Emil. 1995. "Second language speech learning: Theory, findings and problems". In Strange, Winifred (ed.). *Speech perception and linguistic experience: issues in cross-language speech research*. Timonium, MD: York Press, 233–277.

Flege, James Emil; MacKay, Ian. 2004. "Perceiving vowels in a second language". *Studies in Second Language Acquisition* 26, 1-34.

Fridland, Valerie; Kendall, Tyler; Farrington, Charlie. 2014. "Durational and spectral differences in American English vowels: dialect variation within and across regions". *The Journal of the Acoustical Society of America* 136(1), 341-349.

Hallé, Pierre; Best, Catherine. 2007. "Dental-to-velar perceptual assimilation: A cross-linguistic study of the perception of dental stop+/1/ clusters". *The Journal of the Acoustical Society of America* 121(5), 2899-2914.

Havelka, Jelena; Raslte, Kathleen. 2005. "The assembly of phonology from print is serial and subject to strategic control: Evidence from Serbian". *Journal of Experimental Psychology: Learning, Memory, and Cognition* 31(1), 148-158.

Holt, Lori; Lotto, Andrew. 2010. "Speech perception as categorisation". *Attention, Perception, & Psychophysics* 72(5), 1218-1234.

Housen, Alex. 2012. "Time and amount of L2 contact inside and outside the school: Insights from European Schools". In Muñoz, Carmen (ed.). *Intensive Exposure Experiences in Second Language Learning*. Bristol: Multilingual Matters, 111-138.

Jenkins, Jennifer. 2000. *The phonology of English as an international language.* Oxford: Oxford University Press.

Jenkins, Jennifer. 2002. “A sociolinguistically based, empirically researched pronunciation syllabus for English as an international language”. *Applied Linguistics* 23(1), 83-103.
Jenkins, Jennifer. 2009. “Teaching pronunciation for English as a lingua franca: A sociopolitical perspective”. In Gnutzmann, Klaus; Intemann, Frauke (eds.). *The Globalisation of English and the English language classroom*. Tuebingen: Gunter Narr Verlag, 141-154.
Johnson, Marion. 1978. “A note on the Inuit uvular nasal”. *Études Inuit Studies* 2(1), 132-135.
Kopečková, Romana. 2012. “Differences in L2 segmental perception: The effects of age and L2 learning experience”. In Muñoz, Carmen (ed.). *Intensive Exposure Experiences in Second Language Learning*. Bristol: Multilingual Matters, 234-255.
Krebs-Lazendić Lidija; Best, Catherine. 2007. “Early and late bilinguals’ vowel perception and production: English vowel contrasts that give Serbian-English bilinguals a H(E)AD-ache”. In Rauber, S.;Watkins, M., Baptista B. (eds.). *New Sounds 2007: Proceedings of the Fifth International Symposium on the Acquisition of Second Language Speech*. Florianópolis: Federal University of Santa Catarina, 282-292.
Krebs-Lazendić Lidija. 2008. “Early vs. late Serbian-English bilinguals' responses to two Australian English vowel contrasts”. PhD thesis, University of Western Sydney.
Labov, William; Ash, Sharon; Charles, Boberg. 2006. *The atlas of North-American English: Phonetics, phonology and sound change*. Berlin: Mouton de Gruyter.
Labov, William; Fisher, Sabriya; Gylfadottír, Duna; Henderson, Anita; Sneller, Betsy. “Competing systems in Philadelphia phonology”. *Language Variation and Change* 28(3), 273-305.
Ladefoged, Peter; Maddieson, Ian. 1996. *The sounds of the world’s languages.* Oxford: Blackwell Publishers.
Levy, Erika; Strange, Winifred. 2008. “Perception of French vowels by American English adults with and without French language experience”. *Journal of Phonetics* 36(1), 141-157.
Levy, Erika. 2009. “Language experience and consonantal context effects on perceptual assimilation of French vowels by American-English learners of French”. *The Journal of the Acoustical Society of America* 125(2), 1138-1152.
Major, Roy. 2008. “Transfer in second language phonology: A review”. In Zampini, Mary; Edwards Hansen, Jete (eds.). *Phonology and second language acquisition*. Amsterdam, PH: John Benjamins Publishing Company, 63-95.
Marković, Maja. 2009a. “Different strategies in acquiring L2 vowels: The production of high English vowels /iː, ɪ, uː, ʊ/ by native speakers of Serbian”. In Čubrović, Biljana; Paunović, Tatjana (eds.). *Ta(l)king English phonetics across frontiers*. Newcastle upon Tyne: Cambridge Scholars Publishing, 3-18.
Marković, Maja. 2009b. “The perception and production of the English vowels /e/ and /æ/ by native speakers of Serbian”. In Tsangalidis, Tasos (ed.). *Selected Papers from the 18th ISTAL*. Thessaloniki: Monochromia Publishing, 253-262.
Marković, Maja; Jakovljević, Bojana. 2016. “Phonetic cue interpretation in the acquisition of a nonnative vocalic contrast”. *Annual Review of the Faculty of Philosophy Novi Sad* 41(1), 215-227.
McMahon, April. 2001. *An introduction to English phonology*. Edinburgh: Edinburgh University Press.
Moyer, Alene. 2014. “What’s age got to do with it? Accounting for individual factors in second language accent”. *Studies in Second Language Learning and Teaching* 4(3), 443-464.

Naser, Eddine. 2011. “Second language acquisition: The articulation of vowels and the importance of tools in the learning process”. In Arabski, Janusz; Wojtaszek, Adam (eds.). *The Acquisition of L2 Phonology*. Bristol: Multilingual Matters, 3-15.
Newbold, David. 2021. “English lingua franca: New parameters for the teaching (and testing) of English pronunciation?” *EL.LE* 10(3), 393-410.
Paunović, Tatjana. 2003. “On some sociolinguistic variables in the acquisition of a phonological system”. *FACTA UNIVERSITATIS- Linguistics and Literature* 2(10), 389-402.
Paunović, Tatjana. 2011. “Sounds Serbian? Acoustic properties of Serbian EFL students’ speech”. *Selected Papers from the 19th ISTAL*. Thessaloniki: Monochromia Publishing, 357-369.
Petrović, Dragoljub. 2001. “Languages in contact: Standard Serbian phonology in an urban setting”. *International Journal of the Sociology of Language* 151, 19-40.
Plag, Ingo; Arndt-Lappe, Sabine; Braun, Maria; Schramm, Mareile. 2015. *Introduction to English linguistics*. (3rd edition). Berlin: Mouton de Gruyter.
Polka, Linda. 1992. “Characterizing the influence of native language experience on adult speech perception”. *Perception & Psychophysics* 52(1), 37-52.
Polka, Linda. 1995. “Linguistic influences in adult perception of non-native vowel contrasts”. *The Journal of the Acoustical Society of America* 97,1286-1296.
Radeau, Monique; Morais, Jose; Dewier, Agnes. 1989. “Phonological priming in spoken word recognition: Task effects”. *Memory & Cognition* 17(5), 525-535.
Simonović, Marko; Kager, René. 2020. “Serbo-Croatian is developing stem-based prosody. Why so?” In Marušič, Franc; Mišmaš, Petra; Žaucer, Rok (eds.). *Advances in formal Slavic linguistics*. Berlin: Language Science Press: 305-322.
Stahlke, Herbert. 2003. “Fortis and lenis obstruents in English”. *WORD* 54(2), 191-216.
Strange, Winifred; Akahane-Yamada, Reiko; Kubo, Rieko; Trent, Sonja; Nishi, Kanae; Jenkins, James. 1998. “Perceptual assimilation of American English vowels by Japanese Listeners”. *Journal of Phonetics* 26(4), 311-344.
Strange, Winifred; Bohn, Ocke-Schwen; Trent, Sonja; Nishi, Kanae. 2004. “Acoustic and perceptual similarity of North German and American English vowels”. *The Journal of the Acoustical Society of America* 115(4), 1791-1807.
Subotić, Ljiljana; Sredojević, Dejan; Bjelaković, Isidora. 2012. *Fonetika i fonologija: ortoepska i ortografska norma standardnog srpskog jezika [Phonetics and phonology: Orthoepic and orthographic norm of the standard Serbian language]*. Novi Sad: University of Novi Sad.
Tengs Sannes, Maria. 2013. “From the native speaker norm towards English as an international language: A study of exposure and attitudes to native and non-native varieties in the teaching of English in Norway”. MA thesis, University of Bergen.
Thir, Veronika. 2014. “PPOCS1 revisited: Exploring the implications of ELF for English pronunciation teaching”. MA thesis, University of Vienna.
Thir, Veronika. 2016. “Rethinking pronunciation teaching in teacher education from an ELF perspective”. *VIEWS* 25, 1-28.
Timmis, Ivor. 2002. “Native-speaker norms and international English: A classroom view”. *ELT Journal* 56(3), 240-249.
Tyler, Michael; Best, Catherine; Faber, Alice; Levitt, Andrea. 2014. “Perceptual assimilation and discrimination of non-native vowel contrasts”. *Phonetica* 71(1), 4-21.
Vaszolyi, Eric. 1977. *Features of Serbo-Croatian for language teachers.* Mount Lawley: Mount Lawley College of Advanced Education.

Vilke, Mirjana. 1991. “A contemporary view on the compromise system of Serbo-Croatian learners of English”. In Ivir, Vladimir; Kalogjera, David (eds.). *Languages in contact and contrast.* Berlin: Mouton de Gruyter, 483-492.

Wach, Aleksandra. 2011. “Native-speaker and English as a lingua franca pronunciation norms: English majors’ views”. *Studies in Second Language Learning and Teaching* 1(2), 247-266.

Wrembel, Magdalena; Marecka, Marta; Kopečkova, Romana. 2019. “Extending perceptual assimilation model to L3 phonological acquisition”. *International Journal of Multilingualism* 16(4), 513-533.

Zoghbor, Wafa Shahada. 2011. “Teaching the pronunciation of English as a lingua franca: reducing scepticism and increasing practicality”. *International Journal of Humanities and Social Sciences* 1(8), 285-288.

12. Appendix

12.1 English abstract

The main focus of this master thesis is the impact of L1 phonological repertoire on acquisiton of phonology in the second language. Using the Perceptual Assimilation Model (PAM), as a part of the Motor-Theory of speech perception, this thesis examines the perception of the English vowel contrast /e/-/æ/ by adult native speakers of Bosnian/Croatian/Serbian (BCS), whose first language only has a single phoneme /e/. The main method for data collection are two auditory perception tasks: perception of words in isolation i.e. minimal pair distinction (listening to words in isolation and selecting the image that best corresponds to the acoustic stimuli), and category-goodness ratings (listening to words in isolation and rating the similarity between the vowels heard and the BCS phoneme /e/). In addition to the phonetic/phonological part of the thesis, the interrelatedness of L2 exposure and assimilation into L1 categories is taken into consideration, together with the relationship between assimilation and conformity to native-speaker norms in SLA. In the sociolinguistic part, statement ratings are employed as a data collection method. To analyse the data, this thesis relies on descriptive statistics, as well as comparisons with spectral and temporal features of /e/ and /æ/ obtained from the native speaker of English and with their vocalic 'counterparts' /e/ and /a/ in BCS. For the majority of respondents, the findings suggest that the English vowel contrast /e/-/æ/ is assimilated into BCS /e/ on the basis of category-goodness assimilation, with English /e/ being judged as a better exemplar of the native category, and therefore more difficult to identify. In contrast, /æ/ is assimilated into BCS /e/ as a more deviant example, due to hypercorrection and it being heard as a more deviant exponent of the L1 phoneme. The ratings of statements about English language experience indicate that CG type assimilation happens in spite of years of prolonged exposure to English, highlighting the need for proper phonetic training. At the same time, responses to statements concerned with native-speaker norms unveil the irrelevance of mastering the phonology of English, and that more focus should be placed on mutual intelligibility.

Key terms: second language acquisition, phonetics, phonology, auditory perception, Perceptual Assimilation Model, Motor-Theory of speech perception, PRAAT, vowel quality, vowel quantity, L2 experience, conformity to native-speaker norms, statement ratings, minimal pair distinction, category-goodness ratings

12.2 German abstract

Der Schwerpunkt dieser Masterarbeit liegt auf dem Einfluss des phonologischen Systems der Muttersprache auf den Erwerb der Phonologie in der zweiten Sprache. Unter Verwendung des Perceptual Assimilation Models (PAM), als Teil der Motor-Theorie der Sprachwahrnehmung, untersucht diese Arbeit die Wahrnehmung des englischen Vokalkontrasts /e/-/æ/ durch erwachsene Muttersprachler von Bosnisch/Kroatisch/Serbisch (BKS), dessen Muttersprache nur ein einziges Phonem /e/ hat. Die Hauptmethode für die Datenerfassung sind zwei auditive Wahrnehmungsaufgaben: die Wahrnehmung von englischen Wörtern in Isolation (Minimalpaarunterschiedung), und Bewertung der Ähnlichkeit zwischen den gehörten Vokalen und dem BKS /e/. Zusätzlich zum phonetischen/phonologischen Teil der Arbeit wird der Zusammenhang zwischen Erfahrung mit der zweiten Sprache und Assimilation in die Kategorien in der Muttersprache, sowie der Zusammenhang zwischen Assimilation und Konformität mit den muttersprachlichen Normen im Erwerb der zweiten Sprache berücksichtigt. Im soziolinguistischen Teil werden Statementratings als Methode der Datenerfassung verwendet. Um die Daten zu analysieren, stützt sich diese Arbeit auf deskriptive Statistik, Vergleiche mit Qualität und Quantität der Vokalen /e/ und /æ/, die vom englischen Muttersprachlerin erhalten wurden, sowie Vergleiche mit dem vokalen „Gegenstücken“ /e/ und /a/ in BKS. Für die Mehrheit der Befragten deuten die Ergebnisse darauf hin, dass der englische Vokalkontrast /e/-/æ/ auf der Grundlage der Category-Goodness-Assimilation in BKS /e/ assimiliert wird, wobei Englisch /e/ als besseres Beispiel beurteilt wird und daher schwieriger wahrzunehmen. Im Gegensatz dazu wird /æ/ als schlimmeres Beispiel in BKS /e/ assimiliert, aufgrund seiner Qualität und Quantität, der Abweichung von dem BKS Phonem, sowie Hyperkorrektion. Die Bewertungen von Aussagen über Erfahrungen mit der englischen Sprache belegen, dass eine Category-Goodness-Assimilation trotz jahrelangem Kontakt mit Englisch stattfindet, was die Notwendigkeit eines phonetischen Trainings betont. Gleichzeitig zeigen Antworten auf Aussagen, die sich mit den mutterspraclichen Normen befassen, dass es irrelevant ist, die Phonologie des Englischen zu beherrschen, und dass die Muttersprachler von BKS mehr Wert auf Interkomprehension legen.

Schlagwörter: Zweitsprachenerwerb, Phonetik, Phonologie, auditive Wahrnehmung, Perceptual Assimilation Model, Motor-Theorie der Sprachwahrnehmung, PRAAT, Vokalqualität, Vokalquantität, L2-Erfahrung, Konformität mit muttersprachlichen Normen, Aussagebewertungen, Minimalpaarunterscheidung, Kategorie-Gütebewertungen

12.3 Appendix A: Complete dataset of formant frequencies and vowel duration of Pennsylvanian /æ/ and /e/ in 24 tokens used in Task 1

vowel /æ/				vowel /e/			
words	mean F1 in selection (Hz)	mean F2 in selection (Hz)	duration in /s	words	mean F1 in selection (Hz)	mean F2 in selection (Hz)	duration in /s
pan	740,518	2352,179	0,230	*pen*	789,297	2006,909	0,140
bag	912,049	2032,838	0,279	*beg*	471,701	2364,067	0,221
bat	1015,804	1865,701	0,104	*bet*	694,146	2009,404	0,122
ban	684,852	2488,592	0,224	*Ben*	726,176	2000,291	0,122
dad	958,349	2026,607	0,283	*dead*	605,076	2078,170	0,141
flash	1091,643	1876,347	0,185	*flesh*	785,808	1952,065	0,134
axe	1012,517	1882,710	0,141	*X*	727,283	2233,314	0,112
jam	773,406	2044,146	0,224	*gem*	786,002	1846,352	0,055
paddle	767,430	1637,212	0,120	*pedal*	717,532	2061,813	0,059
Brad	914,164	1793,911	0,220	*bread*	755,913	2163,696	0,097
man	731,029	2111,712	0,217	*men*	804,243	2112,452	0,123
cattle	989,330	1825,866	0,120	*kettle*	791,996	2040,161	0,058
Mean	**882,6**	**1994,8**	**0,196**	**/e/**	**721,3**	**2072,4**	**0,115**

12.4 Appendix B: Complete dataset of /e/ and /a/ formants and durations in different BCS dialects

Vowel /e/ in words *led, let, ego, eto, sve, test, ten*

	LED			**LET**		
Speaker	**F1**	**F2**	**duration /s**	**F1**	**F2**	**duration /s**
Banja Luka	545,65	2588,26	0,16	546,12	2598,27	0,22
Tuzla	573,31	2174,68	0,14	475,13	2215,45	0,17
Zagreb	595,82	2192,17	0,19	553,23	2286,71	0,19
Podgorica	854,16	2199,67	0,15	545,28	2258,07	0,21
S. Palanka	466,38	1744,99	0,15	441,69	1674,14	0,16
Split	565,85	1912,57	0,23	557,94	2150,02	0,24
Kikinda	456,30	2214,99	0,28	376,08	2344,76	0,30
Mean	579,64	2146,76	0,19	491,56	2218,20	0,21

	EGO			**ETO**		
Speaker	**F1**	**F2**	**duration /s**	**F1**	**F2**	**duration /s**
Banja Luka	525,18	2696,35	0,21	739,90	1837,10	0,16
Tuzla	470,48	2308,63	0,19	649,16	2084,58	0,13
Zagreb	645,32	2367,98	0,19	701,33	2079,25	0,15
Podgorica	547,53	2314,84	0,18	661,07	2142,86	0,13
S. Palanka	443,67	1818,84	0,17	579,45	1550,17	0,13
Split	524,71	2212,35	0,22	586,64	2025,92	0,17
Kikinda	369,35	2398,83	0,21	742,23	1966,61	0,21

Mean	503,75	2302,55	0,20	665,68	1955,21	0,15

	SVE			**TEST**		
Speaker	**F1**	**F2**	**duration /s**	**F1**	**F2**	**duration /s**
Banja Luka	754,97	1742,88	0,18	717,91	2110,83	0,15
Tuzla	654,71	1720,84	0,15	538,39	1756,89	0,15
Zagreb	683,65	1948,16	0,15	677,06	2126,25	0,11
Podgorica	681,11	1962,11	0,14	662,58	1942,34	0,15
S. Palanka	562,73	1476,72	0,13	577,35	1500,78	0,14
Split	795,92	1876,28	0,14	573,08	1977,25	0,14
Kikinda	745,64	1967,15	0,20	698,50	2015,23	0,13
Mean	696,96	1813,45	0,16	634,98	1918,51	0,14

	TEN		
Speaker	**F1**	**F2**	**duration /s**
Banja Luka	548,19	2723,47	0,21
Tuzla	526,28	2303,14	0,20
Zagreb	748,05	2297,54	0,16
Podgorica	487,27	2381,88	0,16
S. Palanka	445,93	1779,25	0,12
Split	526,07	2120,33	0,26
Kikinda	380,98	2357,08	0,23
Mean	523,25	2280,38	0,19

Speaker	**Mean F1 /e/**	**Mean F2 /e/**
Banja Luka	615,28	2.328,17
Tuzla	555,35	2.080,60
Zagreb	657,78	2.185,44
Podgorica	634,14	2.171,68
S. Palanka	502,46	1.649,27
Split	590,03	2.039,24
Kikinda	538,44	2.180,66

Vowel /a/ in words *rad, rak, ada, at, panj, zla*

	RAD			**RAK**		
Speaker	**F1**	**F2**	**duration /s**	**F1**	**F2**	**duration /s**
Banja Luka	908,44	1493,00	0,25	964,28	1757,58	0,13
Tuzla	860,35	1684,53	0,26	673,60	1666,16	0,12

Zagreb	737,57	1567,07	0,25	867,50	1554,02	0,10
Split	848,53	1382,51	0,27	855,41	1490,84	0,19
Podgorica	936,16	1582,63	0,19	961,55	1662,22	0,13
S. Palanka	652,15	1300,79	0,16	596,16	1202,98	0,16
Kikinda	802,74	1340,67	0,23	804,88	1416,83	0,21
Mean	820,85	1478,74	0,23	817,63	1535,80	0,15

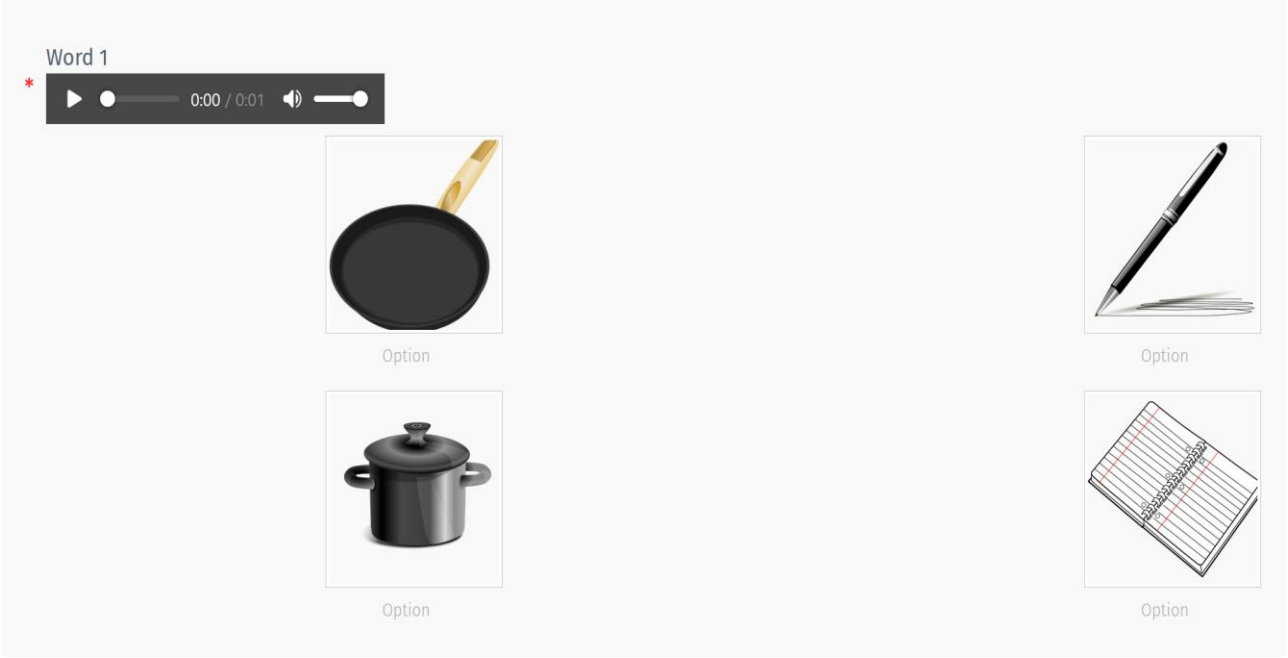
	ADA			**AT**		
Speaker	**F1**	**F2**	**duration /s**	**F1**	**F2**	**duration /s**
Banja Luka	870,08	1291,41	0,25	1030,05	1560,79	0,16
Tuzla	872,53	1545,45	0,19	835,34	1542,05	0,13
Zagreb	851,57	1548,47	0,26	888,06	1517,97	0,26
Split	854,78	1324,39	0,23	847,24	1390,03	0,29
Podgorica	913,85	1433,19	0,22	948,28	1450,80	0,12
S. Palanka	695,31	1162,99	0,20	707,41	1212,38	0,16
Kikinda	807,56	1283,41	0,27	780,81	1281,57	0,16
Mean	837,95	1369,90	0,23	862,46	1422,23	0,18

	PANJ			**ZLA**		
Speaker	**F1**	**F2**	**duration**	**F1**	**F2**	**duration**
Banja Luka	896,38	1661,99	0,24	961,91	1498,47	0,14
Tuzla	770,59	1565,68	0,20	779,09	1313,07	0,11
Zagreb	881,20	1641,57	0,23	841,14	1379,55	0,09
Split	891,60	1314,82	0,30	782,75	1345,82	0,21
Podgorica	872,62	1530,48	0,16	932,53	1345,57	0,11
S. Palanka	747,47	1300,89	0,14	611,44	1110,07	0,10
Kikinda	868,84	1451,91	0,28	787,25	1225,67	0,10
Mean	846,96	1495,34	0,22	813,73	1316,89	0,12

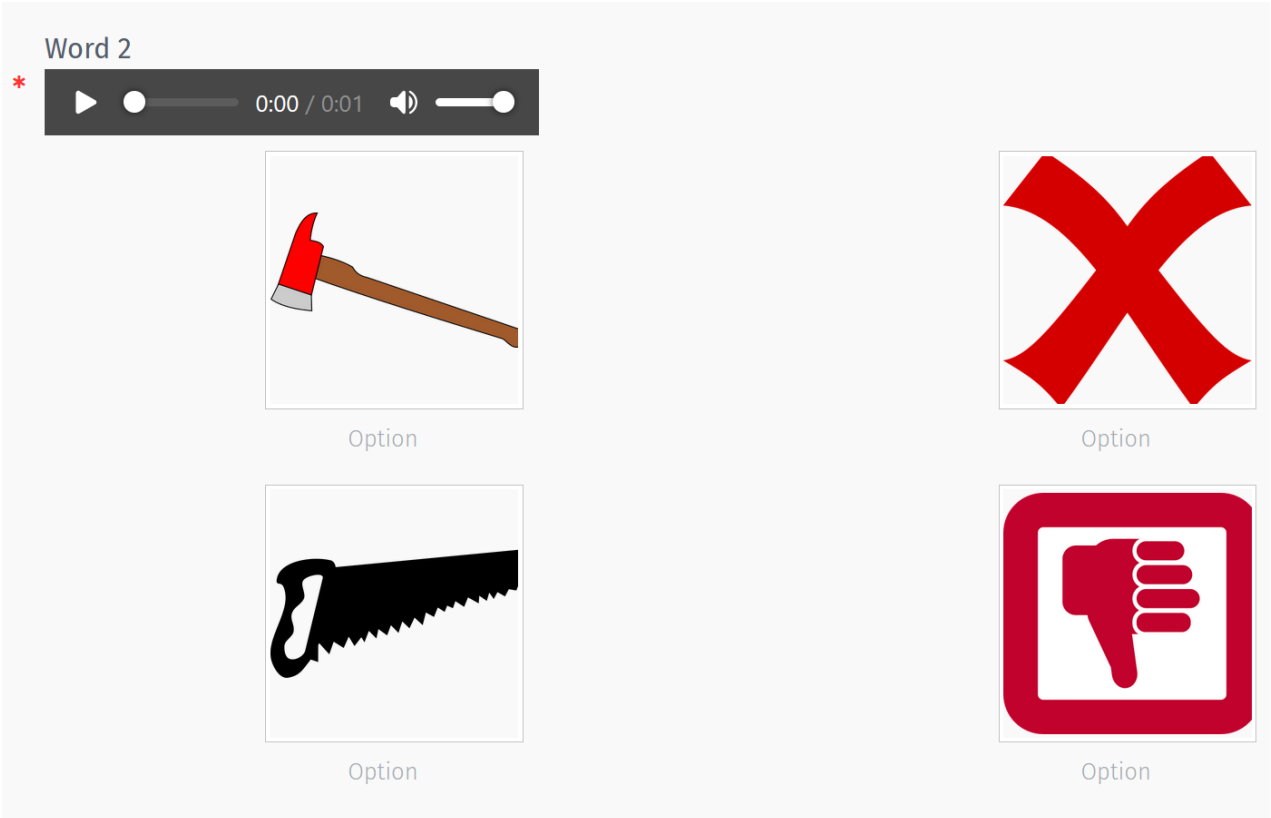
Speaker	**Mean F1 /a/**	**mean F2 /a/**
Banja Luka	938,52	1543,87
Tuzla	798,58	1552,82
Zagreb	844,51	1534,78
Split	846,72	1374,73
Podgorica	927,50	1500,81
S. Palanka	668,32	1215,02
Kikinda	808,68	1333,34

12.5 Appendix C: Design and visual stimuli for 24 acoustic tokens in Task 1

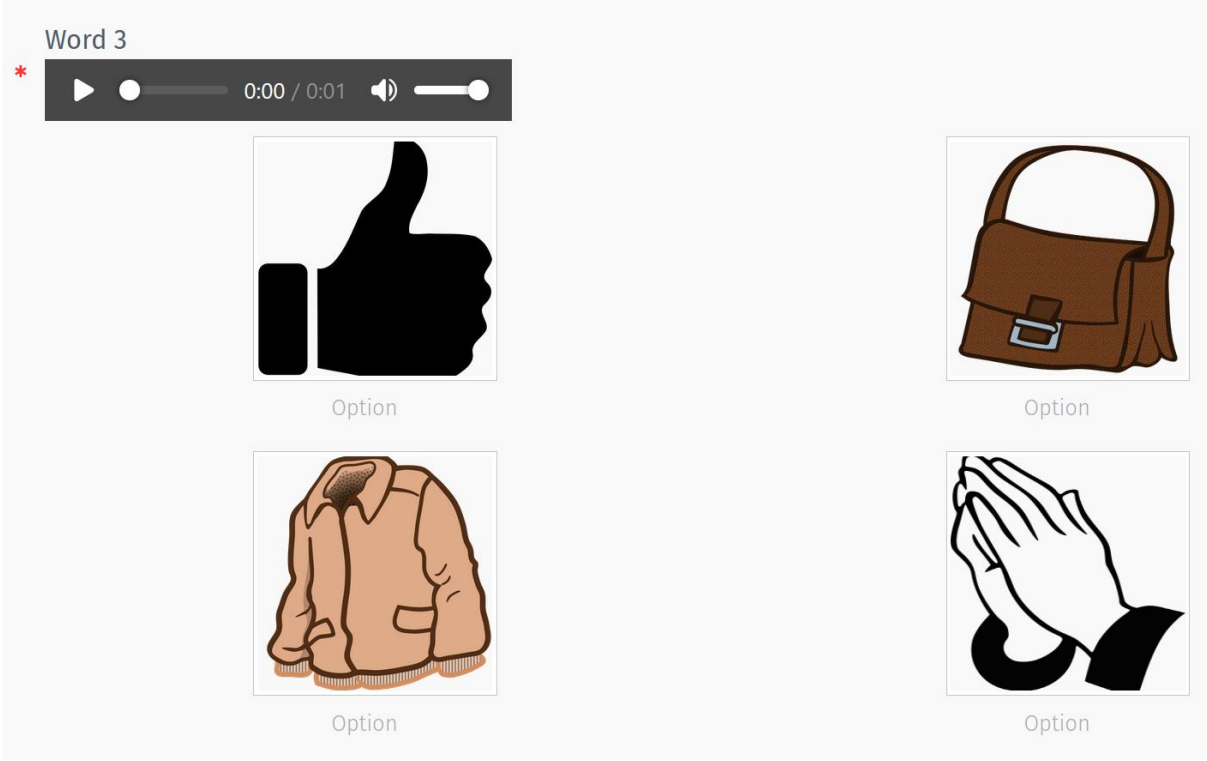
Pair 1: *pan-pen*



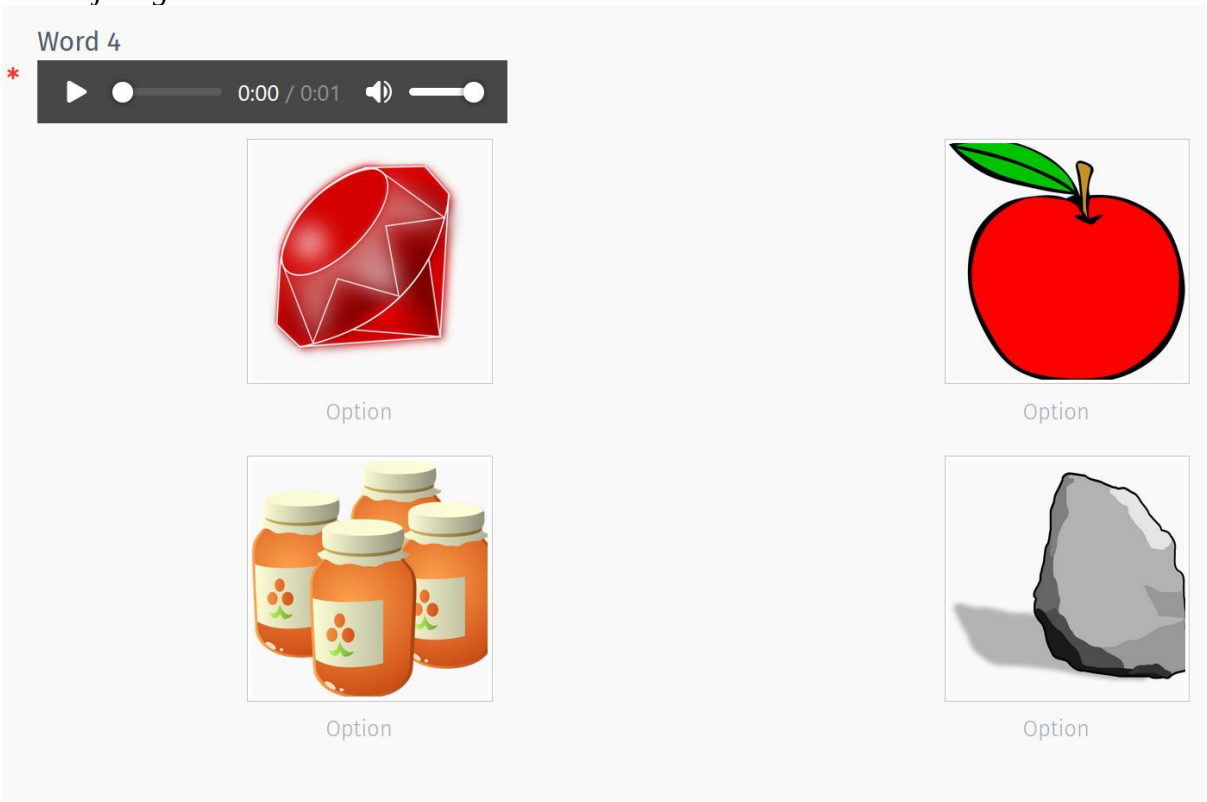
Pair 2: *axe-X*



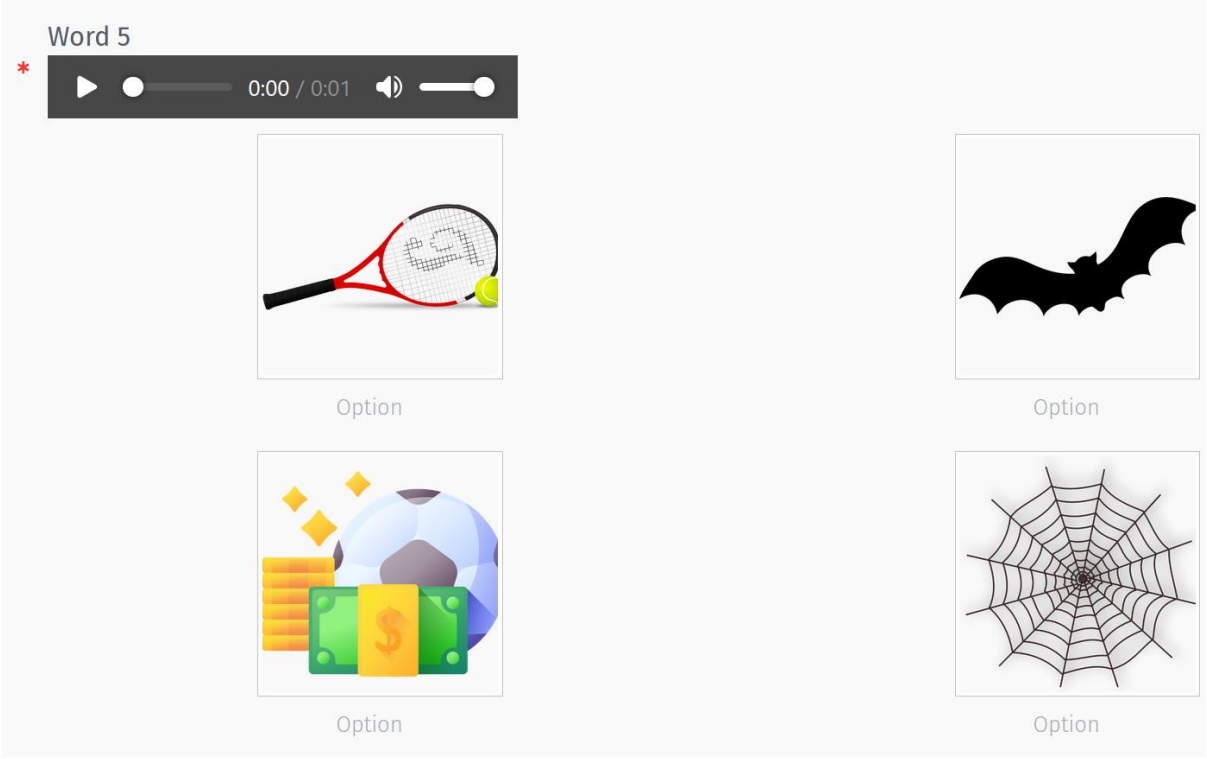
Pair 3: *bag-beg*



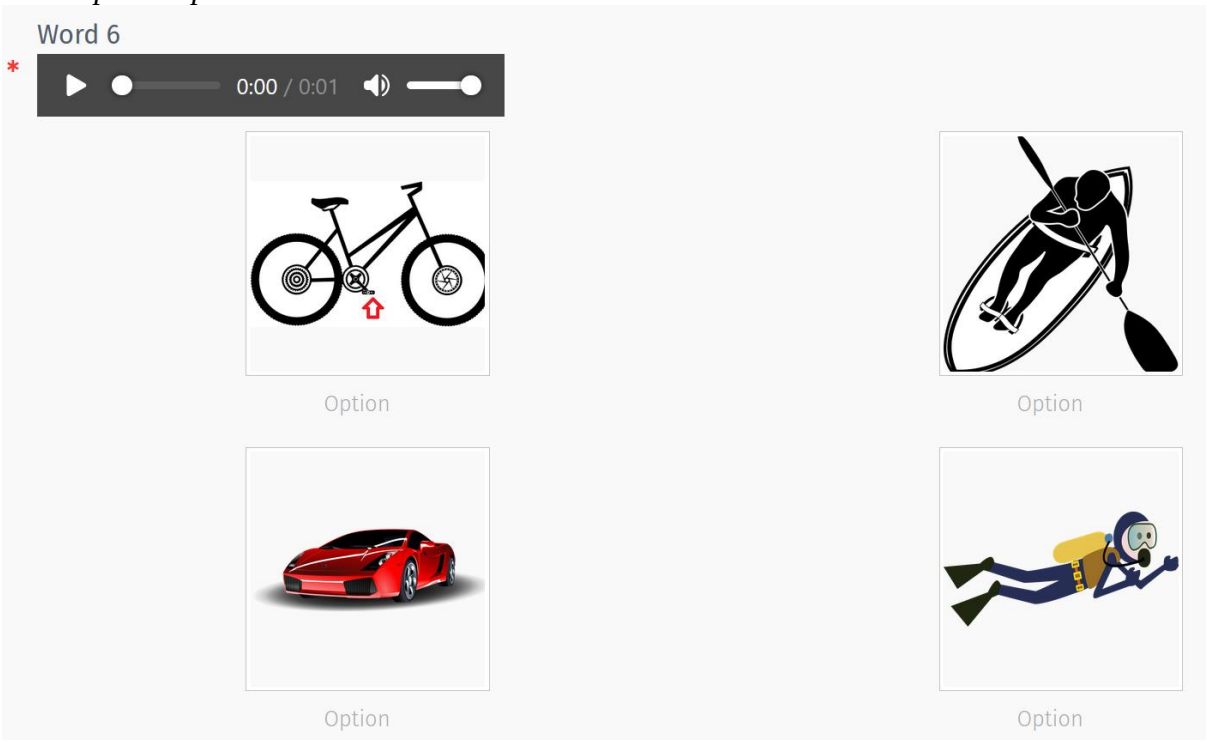
Pair 4: *jam-gem*



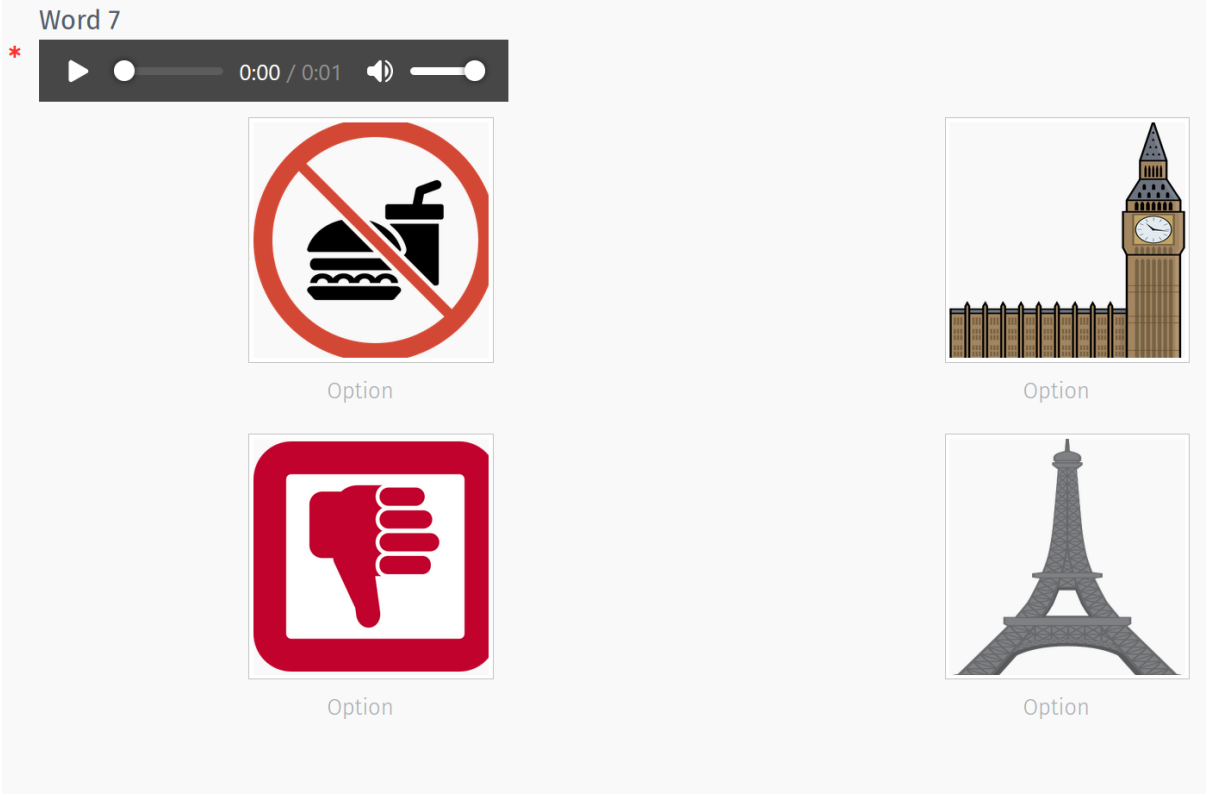
Pair 5: *bat-bet*



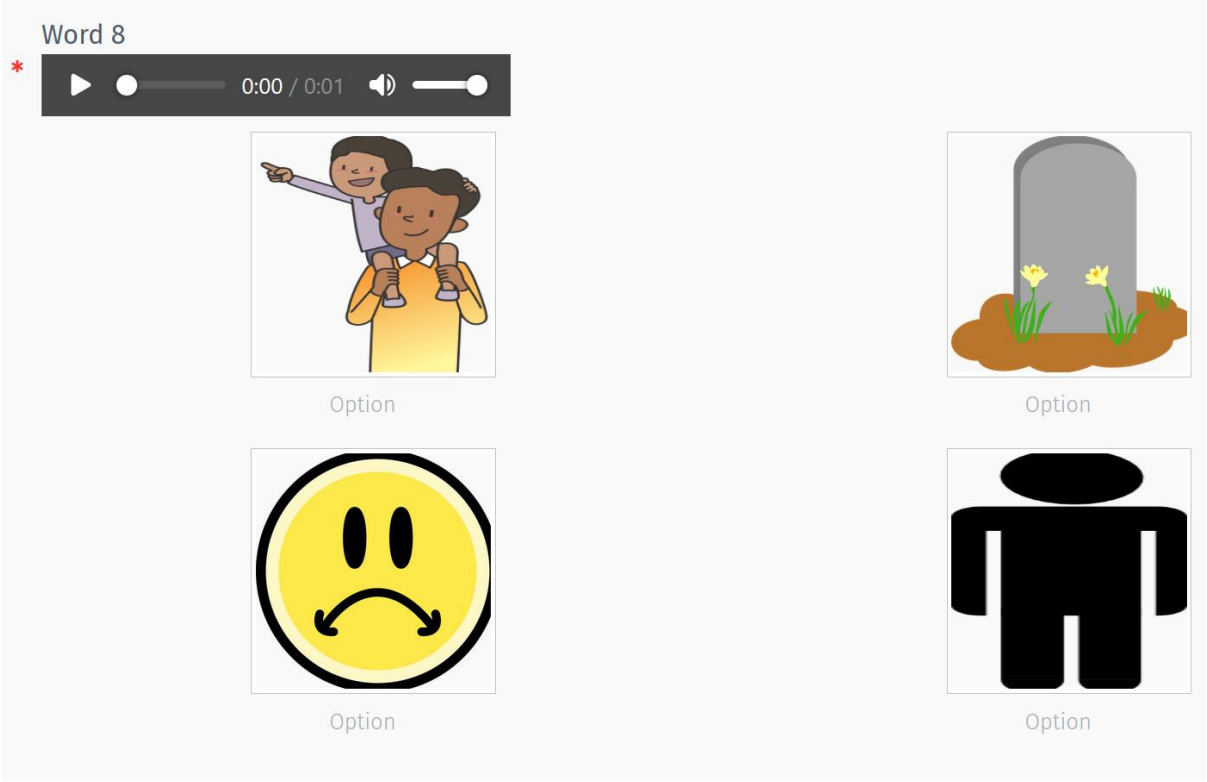
Pair 6: *paddle-pedal*



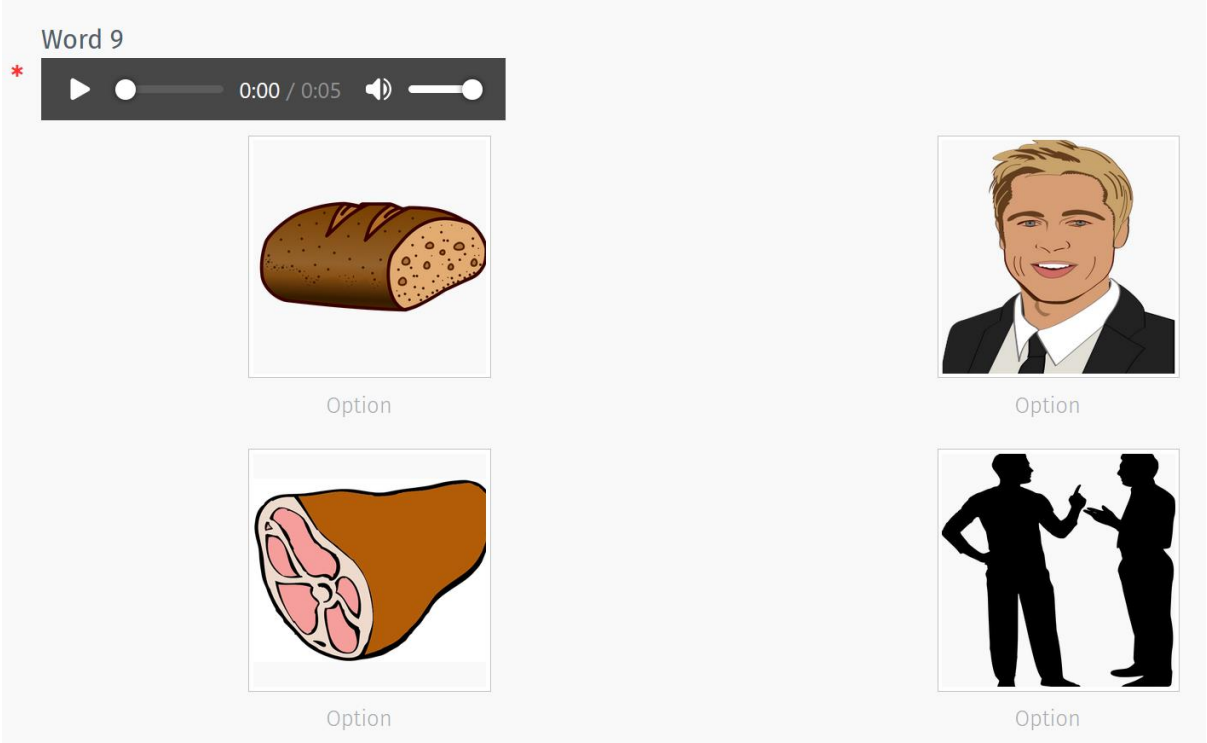
Pair 7: *ban-ben*



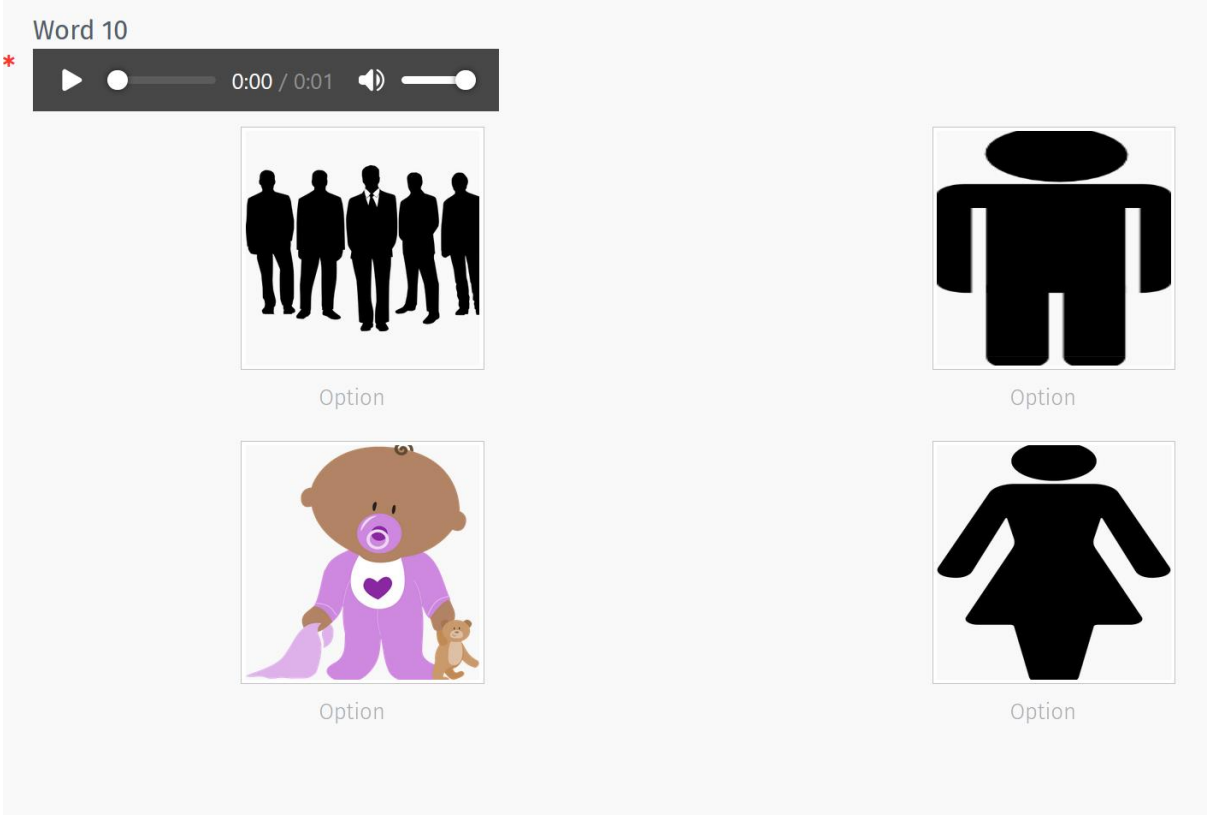
Pair 8: *dad-dead*



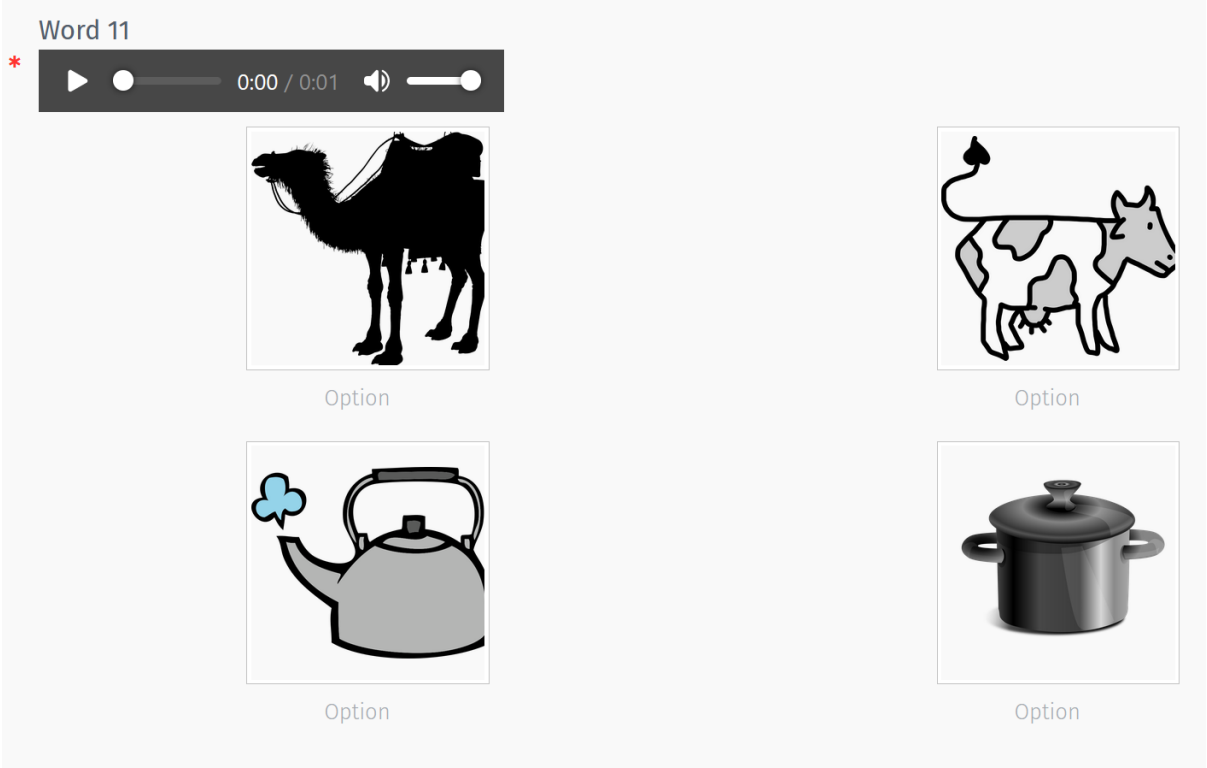
Pair 9: *Brad-bread*



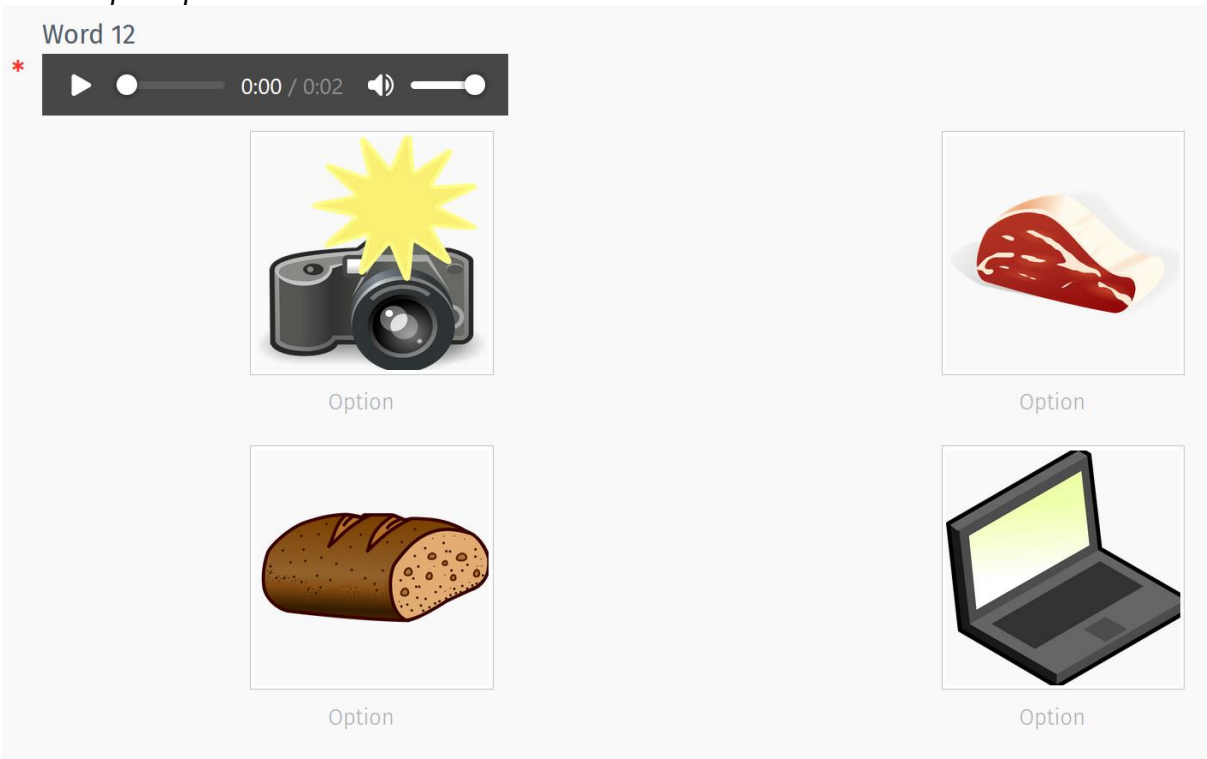
Pair 10: *man-men*



Pair 11: *cattle-kettle*



Pair 12: *flash-flesh*



12.6 Appendix D: Formant frequencies in acoustic stimuli in Task 2

Word	Vowel	Mean F1 in selection	Mean F2 in selection	Duration in /s
bought	/ɑː/	767,45	1.315,25	0,163
bet	/e/	712,74	1.981,28	0,138
bad	/æ/	894,34	1.859,48	0,285
bead	/iː/	409,75	2.475,12	0,265
bat	/æ/	922,20	1.849,27	0,19
bid	/ɪ/	504,33	2.214,39	0,143
but	/ʌ/	804,69	1.665,93	0,127
bed	/e/	673,61	2.004,47	0,128

12.7 Appendix E: Participant information sheet

ID	Age	Country	City	Where was EN learned?	English dialect in education	Time spent in EN countries	Phonetics courses
83562159	25	Bosnia, Austria	Banja Luka, Vienna	School and highschool	British	No	No
85051189	27	Serbia	Belgrade	School	American	None	None
85074804	19	Serbia	Beograd	Beograd			
85084334	26	Croatia	Zagreb	School, private classes, university	British and Scottish	2 weeks	Yes
85084893	25	Bosnia	Bosnia	School	American	0	0
85086730	47	Croatia, Turkey	Sisak + Istanbul	In school	British and American	N/a	no
85125947	26	BiH	Vienna	School		0	0
95770473	26	Serbia	Belgrade	School, uni, course	British	0	Yes
95770940	20	Bosnia	Banja Luka	School	British	no	no
97725127	24	Serbia	Smederevska Palanka	Elementary/ High School	British	0	/
97725153	24	Bosnia	Bijeljina	School	British	No	No
97725382	27	Bosnia and Herzegovina	Banja Luka			0	
97725500	22	Croatia	Split	School	UK english	16 months	.

97725708	39	Serbia Austria	Belgrade/ Graz	In school, private courses	Standard RP	Months	No
97726582	27	Serbia	Belgrade	School	British	4 months	/
97726799	26	Bosnia and Herzegovina	Banja Luka + Sarajevo	Self-taught	British English	None	None
97727055	25	serbia	belgrade	school	british	7 years	no
97727526	27			Courses	American	None	None
97729543	25	bosnia		School	American	0	
97730200							
97734100	28	Bosnia and Herzegovina	Bih Banja Luka 78000				
97736023	25	Bosnia and Herzegovina	Banja Luka	School	British English	0	0
97736902	28	Bosnia and Herzegovina + Austria	Bijeljina+Vienna	In school	AE/BE	0	0
97745530	27	Serbia + Austria	Kraljevo + Vienna	In school in Serbia	American English	No time spent	8 years in school
97754570	26	BiH/ Serbia	Banja Luka+Belgrade	At school and courses	British English, which dialect not sure	0	
97758607	25	Serbia - Austria	Novi Sad - Vienna	school, TV	Standard American	0	None
97759274	27	Serbia	Novi Sad	In middle and high school	AmE	Few months	/
97759865	30	Serbia	Novi Sad	School, private lesson	British English	0	0
97783429	33	Serbia	Belgrade	Serbia	US	0	0

98247786	28	Serbia - Serbia	Smederevska Palanka - Belgrade	School	british english	-	-
102110202	25	Serbia	Smederevska Palanka, Belgrade	in school	I don't remmember	no	no
102110289	26	bosnia	Doboj, Banja Luka	Primary school, grammar school, private lessons	british and american	no	no
102110438	19	Bosnia	Banja Luka	school	Britain	no	no
102110489	18	Bosnia	Banja Luka	school and high school	British	0	no
102110549	23	Croatia	zagreb, Wien	school	uk and us	none	no
102110595	22	Serbia	Novi Sad	school amd course	british and american	no	no
102110705	29	Serbia	S.Palanka, Belgrade	school	american and uk	0	no
102110875	26	BIH, Republika Srpska	Banja Luka	School	both	0	?
102111998	32	Serbia	Smederevska Palanka, Belgrade	School and high school	Uk and USA	no	no
102112158	24	Bosnia, Austria	Banja Luka, Wien	School and high school	standard British	no	no
102112276	19	Serbia, Germany	Belgrade, Frankfurt	school	British	no	no
102152638	29	Bosnia amd Herzegovina	Banja Luka	School	British and American	none	no
102153089	25	Serbia	Subotica, Belgrade	school and high school	UK	none	no
102153396	38	Croatia	Sisak, Zagreb	In primary school and high-school	standard British	a month in the US	no
102154131	26	B&H, Republic of Srpska	Kotor Varos, Banja Luka	school and grammar school	standard British	0	none
96435049	24	Serbia	Belgrade	Highschool	British, American	no	no
96436349	26	Montenegro - Austria	Podgorica - Vienna	School	American	-	-

96439502	37	Croatia/Serbia	Glina/Novi Sad	School		Non	
96440693	24	Serbia+ Austria	Belgrade+ Vienna	Serbia	British	Few years	
96512693	25	Bosnia	Banja Luka	Highschool	Uk/us	No	No
96513877	26	BiH, Austria		School		0	
96545045	26	Serbia + Austria	Novi Sad + Vienna	School, private english lessons, movies, music, books, internet	American mostly	A month in USA and a week in UK	No
96572179	23	Croatia	Zagreb	School	american, british	0	no
97727250	26	Bosnia and Herzegovina	Banja Luka	School	UK	6 months	
97730088	27	Montenegro, Austria	Podgorica, Vienna	Serbia, school	British	None	None
97733782	26	Bosnia and Herzegovina + Serbia	Banja Luka + Novi Sad	School	British	0	/
97734372	26	Serbia	Novi Sad	Private school	American english	6 mo	None
97738968	28	Croatia	Zagreb	School, college	No idea	0	No
97740254	28	Serbia/Austria	Smederevska Palanka/Vienna	Online + Work	British	6 months	/
97742246	26	Bosnia	Banjaluka	School	British	I haven't	
97742614	26	Bosnia and Hercegovina	Banja Luka	At school	British	Zero	
97743013	25	Serbia	Novi Sad	school			
97765260	35	Serbia Austria	Novi Sad Vienna	In Serbia (School)	British	1 Week US (Miami)	

98918534	30	Bosnia, Austria	Banja Luka, Vienna	School	British	Less than a month	
101897256	25	Croatia	Osijek	school, highschool	british	no	no
101897747	20	croatia		Rijeka	british	no	no
101898144	23	Serbia	Kragujevac	school	us and british	0	0
101898478	22	Serbia	Kraljevo	school, course	uk us	no	no
101898643	30	Bosnia	Bijeljina	primary school	american	no	no
101898906	32	Croatia	zagreb	highschool	American	no	no
101899064	27	BiH	Mostar	gymnasium	British	no	no
101899403	28	Croatia	Slavonski Brod	School + course	British and American	0	no
101899811	27	Croatia	Zagreb	Primary school and gymnasium	American	zero	none
101900038	26	Croatia	Sisak	School and highschool	British	0	no
101901061	24	Serbia	Belgrade	Highschool and private lessons	British and American	0	no
101902294	31	Serbia	Smederevska Palanka	school	British	no	no
102058080	33	Bosnia	Bihac	School and highschool	British and American	no	no
102058202	21	BIH	Zenica	Schhol, highschool, Emglish language course	RP and Standard American	0	no
102058308	28	Serbia	Subotica	School, grammar school, TV and music	Mostly British and American	2 weeks in India	No
102060417	26	Croatia	Zagreb	School	Both	no	no
102060796	20	Croatia	Split	School	American,British	nope	nope
102061059	38	Croatia	Sisak,Zagreb	School	Uk	none	none
102149804	28	Serbia	Belgrade	school	british	no	no

102150191	34	Croatia, Serbia	Glina, Novi Sad	school, high-school	standard British and standard American	0	no
102150602	25	Croatia	Zagreb, Wien	school	British	no	no
102151032	21	Bosnia, Austria	Zenica, Vienna	School, grammar school	British	0	no
102151275	27	Bosnia	Banja Luka	school and gymnasium	standard British mostly	0	no

12.8 Appendix F: Task 1 complete dataset

Word	Number of participants who selected the correct vowel	Number of participants who selected its minimal pair	Number of participants who selected the distraction	Total number of participants per word
pan	26	19	0	45
pen	16	26	0	42
man	15	27	0	42
men	19	26	0	45
ban	32	8	5	45
Ben	13	29	0	42
jam	16	26	0	42
gem	14	31	0	45
bag	45	0	0	45
beg	9	33	0	42
dad	44	1	0	45
dead	27	15	0	42
Brad	22	19	1	42
bread	28	17	0	45
paddle	8	34	0	42
pedal	38	7	0	45
axe	19	22	1	42
X	17	27	1	45
flash	43	1	1	45
flesh	4	38	0	42
cattle	2	37	3	42
kettle	22	20	3	45
bat	32	13	0	45
bet	8	33	1	42

Results by phonological context:

PRE-NASAL CONTEXT				
Word	Number of participants who selected the correct vowel	Number of participants who selected its minimal pair	Number of participants who selected the distraction	Total number of participants per word

pan	26	19	0	45
pen	16	26	0	42
man	15	27	0	42
men	19	26	0	45
ban	32	8	5	45
Ben	13	29	0	42
jam	16	26	0	42
gem	14	30	0	44
PRE-LENIS CONTEXT				
Word	Number of participants who selected the correct vowel	Number of participants who selected the incorrect vowel	Number of participants who selected the distraction	Total number of participants per word
bag	45	0	0	45
beg	9	33	0	42
dad	44	1	0	45
dead	27	15	0	42
Brad	22	19	1	42
bread	28	17	0	45
paddle	8	34	0	42
pedal	38	7	0	45
PRE-FORTIS CONTEXT				
Word	Number of participants who selected the correct vowel	Number of participants who selected the incorrect vowel	Number of participants who selected the distraction	Total number of participants per word
axe	19	22	1	42
X	17	27	1	45
flash	43	1	1	45
flesh	4	38	0	42
cattle	2	37	3	42
kettle	22	20	3	45
bat	32	13	0	45
bet	8	33	1	42

12.9 Appendix G: Task 2 complete dataset

Similarity rating between English vowels and BCS/e/								
Participant ID	**/bɑːt/**	**/bet/**	**/bæd/**	**/biːd/**	**/bæt/**	**/bɪd/**	**/bʌt/**	**/bed/**
83562159	0	75	42	0	70	0	0	100
85051189	0	100	80	0	70	15	10	100
85074804	0	55	51	48	51	49	51	0
85084334	0	99	0	50	0	100	0	100
85084893	0	100	0	0	53	0	58	100
85086730	1	100	2	94	1	100	98	100
85125947	0	100	0	0	100	0	0	50

95770473	0	41	58	0	88	0	0	100
95770940	0	71	46	0	100	0	7	100
97725127	0	50	0	50	50	80	70	50
97725153	0	73	18	0	0	0	0	89
97725382	3	43	100	5	100	3	3	100
97725500	14	34	40	9	48	22	13	28
97725708	37	2	65	21	46	67	12	28
97726582	50	54	61	65	66	57	60	73
97726799	0	81	0	0	0	0	0	83
97727055	0	100	0	0	0	0	0	0
97727526	100	28	100	11	27	22	50	42
97729543	10	19	16	20	12	13	10	11
97730200	100	69	49	51	53	54	53	58
97734100	0	0	0	0	0	0	0	1
97736023	3	95	90	10	91	13	2	100
97736902	0	20	5	0	25	0	40	40
97745530	15	98	32	100	29	12	87	37
97754570	0	100	89	0	91	0	0	99
97758607	0	10	20	10	50	0	0	100
97759274	11	9	10	79	50	78	25	79
97759865	0	2	0	0	0	0	0	0
97783429	3	79	74	4	76	7	24	53
98247786	8	70	91	5	40	12	13	4
102110202	7	99	85	0	79	0	11	100
102110289	0	99	68	3	71	0	13	100
102110438	7	99	79	15	84	3	6	98
102110489	7	98	86	7	78	0	10	99
102110549	4	98	84	7	76	6	13	99
102110595	6	99	71	2	80	8	20	100
102110705	6	95	66	10	78	7	28	99
102110875	6	100	73	0	63	0	1	98
102111998	5	98	74	2	52	0	15	100
102112158	5	89	81	0	69	0	12	99
102112276	11	85	60	9	79	3	15	100
102152638	0	93	66	0	81	0	0	100
102153089	0	96	87	0	88	0	0	100
102153396	0	100	91	0	96	0	0	100
102154131	0	99	90	0	92	0	0	99
96435049	0	71	59	0	36	0	0	87
96436349	100	95	100	0	100	0	100	100
96439502	1	28	1	0	39	1	3	35
96440693	0	39	78	0	5	30	2	98
96512693	0	92	74	0	46	57	15	100
96513877	33	55	30	14	69	33	21	65
96545045	0	95	22	0	41	0	0	78
96572179	0	97	0	0	62	0	0	98

97727250	18	75	20	80	27	100	100	84
97730088	13	35	40	4	6	6	5	100
97733782	2	100	67	5	94	22	31	99
97734372	22	49	37	53	33	27	84	46
97738968	5	100	9	2	70	0	0	95
97740254	23	31	100	5	85	26	4	21
97742246	56	49	98	86	65	63	61	61
97742614	12	74	33	14	67	9	31	80
97743013	19	41	68	33	18	21	4	89
97765260	0	0	0	0	0	0	0	0
98918534	2	9	0	0	18	0	0	0
101897256	54	100	69	0	92	0	15	100
101897747	6	100	87	0	82	0	20	100
101898144	0	98	87	0	72	0	9	99
101898478	15	99	72	0	72	0	25	98
101898643	13	94	59	12	72	0	35	99
101898906	31	99	77	34	91	12	19	99
101899064	8	99	81	17	86	15	13	99
101899403	9	94	54	2	82	16	46	100
101899811	7	99	72	0	77	0	29	100
101900038	24	99	79	0	37	9	42	100
101901061	6	93	57	0	73	13	30	99
101902294	10	92	55	0	61	14	37	100
102058080	3	98	71	8	81	0	28	100
102058202	9	99	90	0	88	0	9	100
102058308	0	95	46	0	87	0	0	100
102060417	0	87	69	0	54	0	10	93
102060796	9	99	78	5	88	5	45	100
102061059	13	99	68	0	88	0	15	96
102149804	17	97	81	0	78	0	15	99
102150191	14	91	73	13	84	1	15	99
102150602	13	85	66	13	77	9	32	99
102151032	20	98	74	20	92	6	18	100
102151275	21	99	81	12	79	8	18	99

12.10 Appendix H: L2 experience statement ratings (complete dataset)

Table with percentages, mean, and standard deviation:

Statement	Strongly disagree (1) in %	Disagree (2) in %	Neutral (3) in %	Agree (4) in %	Strongly agree (5) in %	Mean	Standard deviation
Speaking English with friends and family	5,72	55,56	10,32	20,40	8,02	2,69	1,10
Using English	3,41	9,21	12,54	44,84	30,00	3,89	1,05

for work-related purposes							
Listening to music in English	0,00	4,60	12,46	37,86	45,08	4,23	0,84
Using English words in native language	5,72	32,39	16,04	36,59	9,29	3,11	1,14
Watching TV shows, films, YT videos in English	1,19	1,11	2,30	27,62	67,78	4,60	0,71
Reading books, newspaper, blogs, articles in English	0,00	19,45	10,40	31,03	39,13	3,90	1,14
Using English to communicate with NNs	1,11	8,02	12,54	42,46	38,88	4,04	0,95
Switching between English and BCS effortlessly	11,67	23,18	6,98	30,88	27,31	3,39	1,36

Table with raw data:

Statement	**Strongly disagree**	**Disagree**	**Neutral**	**Agree**	**Strongly agree**
I often speak English with my friends and family.	5	48	9	18	7
I often use English for work-related purposes.	3	8	11	39	26
The music I listen to is mostly in English.	0	4	11	33	39
I use many words borrowed from English when I am speaking my language.	5	28	14	32	8
I watch films/TV shows/ YT videos in English on a daily basis.	1	1	2	24	59
I read books/newspapers/blogs/articles in English on a daily basis.	0	17	9	27	34

I mostly use English to communicate with non-native speakers of English.	1	7	11	37	31
During conversations, I can switch between English and my mother tongue effortlessly.	10	20	6	27	24

12.11 Appendix I: attitudes to NS norms statement ratings (complete dataset)

Table with percentages, mean, standard deviation:

Statement	**Strongly disagree (1) in %**	**Disagree (2) in %**	**Neutral (3) in %**	**Agree (4) in %**	**Strongly agree (5) in %**	**Mean**	**Standard deviation**
The only way to be fully proficient in English is to sound like a native speaker from Britain or the US.	18,415	55,315	13,49	11,59	1,19	2,218	0,9255
I think that the fact that BCS is my mother tongue makes it difficult to learn English, particularly in terms of listening and speaking.	42,78	39,05	10,235	5,715	2,22	1,8555	0,951
I think that teachers who sound like native speakers of English are more competent and appealing to students.	14,05	27,775	22,935	28,335	6,905	2,863	1,143
When I speak with native speakers of English, I am afraid they would judge my competence if I do not sound like them.	35,875	42,46	8,97	8,095	4,6	2,031	1,096
When I speak English with other non-native speakers, I tend to judge their language competence on the basis of how they speak.	62,06	23,255	10,16	4,525	0	1,5715	0,842
I do not care about my English accent and learning pronunciation as	2,3	18,255	14,76	23,33	41,35	3,8315	1,2185

long as I am understood.							
I do not like it when speakers of BCS try to sound like native speakers of English, when they do not live in English-speaking countries.	9,045	12,7	20,28	35,795	21,985	3,4895	1,215
I think it is unnatural for speakers of my language to insert English words in sentences and use their original English pronunciation.	11,425	10,315	16,035	31,19	31,03	3,6005	1,335
I think that people who are majoring in English should be the only ones concerned with sounding like native speakers	5,715	17,06	48,41	13,81	15	3,153	1,0615
There is no point in wanting to sound like someone from the UK, US, Canada, etc., when English is a global language	1,11	8,015	17,145	45,965	27,775	3,913	0,928
I think that schools and universities in Bosnia, Croatia, and Serbia should pay more attention to grammar and vocabulary, and not waste time on speaking and listening skills	13,89	11,35	13,65	20,795	40,32	3,623	1,4645
I don't really pay attention to other people's English accent because I am more interested in what they are actually saying.	0	7,935	10,08	28,885	53,095	4,2715	0,913

Table with raw data:

Statement	**Strongly disagree**	**Disagree**	**Neutral**	**Agree**	**Strongly agree**
The only way to be fully proficient in English is to sound like a native speaker from Britain or the US.	16	48	12	10	1
I think that the fact that Bosnian/Croatian/Serbian is my mother tongue makes it difficult to learn English, particularly in terms of listening and speaking.	37	34	9	5	2
I think that teachers who sound like native speakers of English are more competent and appealing to students.	12	24	20	25	6
When I speak with native speakers of English, I am afraid they would judge my competence if I do not sound like them.	30	37	8	7	5
When I speak English with other non-native speakers, I tend to judge their language competence on the basis of how they speak.	54	20	9	4	0
I do not care about my English accent and learning pronunciation as long as I am understood.	2	16	13	20	36
I do not like it when speakers of Bosnian/Croatian/Serbian try to sound like native speakers of English, when they do not live in English-speaking countries.	8	11	18	31	19
I think it is unnatural for speakers of my language to insert English words in sentences and use their original English pronunciation.	10	9	14	27	27
I think that people who are majoring in English should be the only ones concerned with sounding like native speakers.	5	15	42	12	13
There is no point in wanting to sound like someone from the UK, US, Canada, etc., when English is a global language.	1	7	15	40	24
I think that schools and universities in Bosnia, Croatia, and Serbia should pay more attention to grammar and vocabulary, and not waste time on speaking and listening skills.	12	10	12	18	35
I don’t really pay attention to other people’s English accent because I am more interested in what they are actually saying.	0	7	9	25	46