INTERSPEECH 2020
October 25–29, 2020, Shanghai, China

Exploring the use of an artificial accent of English to assess phonetic learning in monolingual and bilingual speakers

Laura Spinu[1], *Jiwon Hwang*[2], *Nadya Pincus*[3], *Mariana Vasilita*[1]

[1]City University of New York
[2]Stony Brook University
[3]University of Delaware

lspinu@kbcc.cuny.edu, jiwon.hwang@stonybrook.edu, npincus@udel.edu, m.vasilita@gmail.com

Abstract

We designed a production experiment to explore the relatively controversial phenomenon of the bilingual advantage. Our focus is on an understudied aspect of bilingual cognition, specifically phonetic learning. We presented 36 participants (17 monolinguals and 19 early bilinguals) living in New York City with an artificially constructed accent of English, differing in four ways from Standard American English. More precisely, the novel accent included a vocalic change (diphthongization of the open-mid front unrounded vowel), consonantal change (tapping of intervocalic liquids), syllable structure change (epenthesis in voiceless s-clusters) and suprasegmental change (a novel intonation pattern in tag questions). After recording their baseline accents, the participants first completed a training task, in which they listened to and then directly imitated sentences heard in the novel accent, and then a testing task, in which they were asked to read the baseline sentences in the accent they had just learned in the absence of any audio prompts. In this paper, we present acoustic results with diphthongization and tag question intonation. Our findings replicate the previously observed bilingual advantage in phonetic learning across the board and extend it to novel learning circumstances.

Index Terms: second dialect learning, phonetic learning, artificial accent, bilingualism, bilingual advantage

1. Introduction

Experimental studies show a number of cognitive consequences of bilingualism, sometimes referred to as a 'bilingual advantage' [1, 2, 3], although the term has recently been losing popularity, earning the nickname of "Loch Ness monster" [4] due to conflicting findings reported in the literature of the past decade [5, 6]. The reported lack of replicability of some of these studies is thought to arise from the difficulty of quantifying the bilingual experience [7] and from the fact that some of the posited advantages of bilingualism are thought to be most evident in childhood and old age, but 'muted' in adulthood [8]. Emerging areas of research where a consistent bilingual advantage has been identified with young adults include studies on phonetic learning. Recent findings suggest enhanced phonetic and phonological learning ability in bilinguals compared to monolinguals [9, 10]. A study in which subjects learned vocabularies that differentiated words using foreign phonetic contrasts [11] reports that bilinguals possess an advantage over monolinguals in terms of phonetic learning, which is however modulated by the universal difficulty of the specific phonetic contrast to be learned and by the phonetic similarity between the target language and the learners' native language. A study on the auditory processing of a synthesized syllable (i.e. [da]) revealed stronger subcortical encoding of sound in bilinguals [12]. Other experimental results with non-native contrasts [10] suggest enhanced speech perception abilities in bilinguals and multilinguals compared to monolinguals, even though the three groups do not differ in their ability to discriminate a non-native contrast before any training is received. Because studies such as these have only addressed limited aspects of second language phonetic and phonological learning, we are faced with the follow-up question of whether the results might differ when participants are presented with a new accent as a whole, a scenario which more closely mimics real-life situations.

A number of longitudinal studies have taken this into account, but focused on the acquisition of limited sets of sounds (see [13] on rhotics). Few studies investigated the acquisition of a novel accent as a whole. These reported a bilingual advantage in terms of segmental learning, both for vowels [14] and consonants [9], and a slight monolingual advantage in the learning of suprasegmental properties of the novel accent [15]. In [9], the performance of 17 monolinguals and 25 bilinguals from Canada was compared in a production experiment with two tasks: imitation and spontaneous reproduction of a novel foreign accent, specifically Sussex English, spoken in Southeast England. The focus was on a sound already existing in the subjects' production (i.e. the glottal stop), but differently mapped to surface representations in the novel accent to which they were exposed (i.e. as the only possible allophonic realization of coronal stops in word-final position). Bilinguals showed more effective learning, expressed as a significant increase in glottal stop production post-training, even though the two groups' performance during the imitation task was very similar. However, an investigation of a different aspect of the same accent in the same speakers [15], revealed a slight monolingual advantage in the acquisition of a new intonation pattern in declarative sentences (i.e. the phenomenon known as uptalk).

While adding to the body of work on how phonetic learning is affected by language experience, such studies present certain drawbacks intrinsic to the experimental design. Methodological issues arise from the inability to control for and measure all of the potential differences between participants' native dialect and that of a different natural dialect. While the results obtained to date remain valid, it must be noted that the participants were not asked to focus on any particular cue, and it is a distinct possibility that while some subjects attended to certain cues, a number of different cues were more salient to other subjects. For instance, recent work with Korean-accented English has shown that while segmental and prosodic information have different weights in the perception of foreign-accented speech [16], segmental information contributes substantially more to the perception of foreign accentedness. Depending on the par-

http://dx.doi.org/10.21437/Interspeech.2020-2783

ticipants' native language, different cues may have been consciously perceived as more salient and deliberately pursued in accent imitation and regeneration.

These findings thus underscore the importance of using model accents that only differ in limited ways from baseline accents, so that *all* of the existing differences can be measured and used to assess learning in various groups of speakers. One potential solution is to employ an artificial accent that sounds as naturalistic as possible, using a limited number of differences that reflect phonetic and phonological behavior attested in other languages or dialects. It is precisely this approach that we pursue in the experiment reported here.

2. Experiment

2.1. Hypotheses

Our predictions are primarily based on previous work suggesting the existence of a bilingual advantage in phonetic learning for segmental aspects of speech [9, 11, 10] and a potential monolingual advantage for intonational aspects [15].

Hypothesis 1 Bilinguals will outperform monolinguals in the learning of segmental properties of the novel accent.

Hypothesis 2 Monolinguals will outperform bilinguals in the learning of intonational properties of the novel accent.

2.2. Stimuli

An artificial accent of English, henceforth referred to as Model Speech, was created such that it differed in four distinct ways from standard North American English:

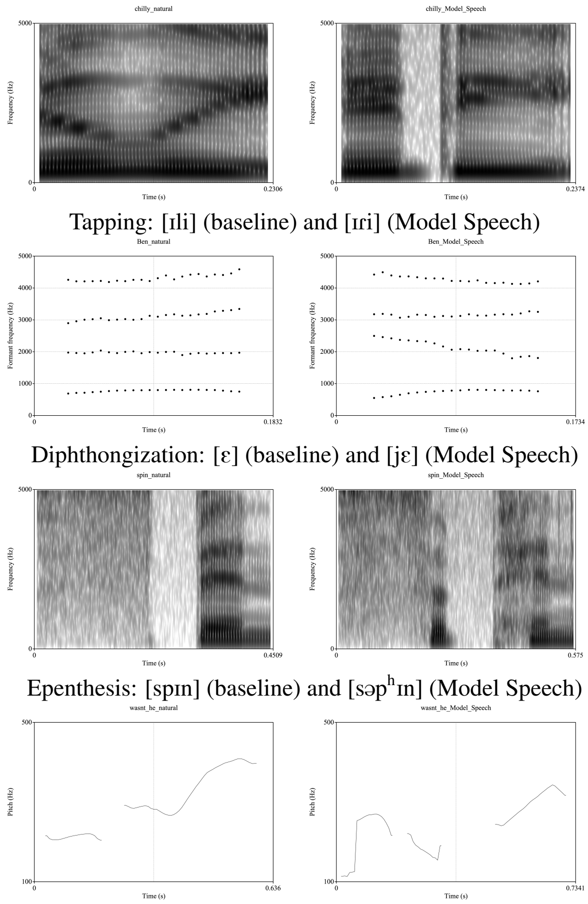
1. Tapping: intervocalic /l/ → [ɾ] e.g. 'color' →[kʌɾɚ]

2. Diphthongization: the vowel /ɛ/ → [jɛ] after an onset consonant, e.g. 'bed' → [bjɛd]

3. Vowel epenthesis: voiceless clusters of the form sC̥ → səC̥ e.g. 'spy' → [səp^haj]

4. Intonation change: tag questions were realized with a novel Mid-Low-High (MLH) pattern. Tag questions (e.g. *isn't it?*) are typically produced with either rising or falling intonation in standard American English.

The stimuli consisted of short sentences containing either one single feature e.g. *You make a good spy*, where *spy* was realized as [səp^haj] (epenthesis), two features combined e.g. *She put a spell on him*, where [spɛl] was realized as[səp^hjɛl] (epenthesis and diphthongization), or all four of them (e.g. *You set the speed alone, didn't you?* where the vowel in the word *set* was diphthongized, epenthesis occured in the word *speed*, tapping affected the [l] in *alone*, and the tag question *didn't you?* was realized with a MLH contour). The features were distributed as follows: 20 tapped /l/, 20 diphthongized vowels, 20 epenthesized vowels and 10 tag questions. The reason less tag questions were included compared to the other novel features was that they were found impressionistically to be highly salient and their presence in higher numbers was deemed to have a distracting effect on the listeners. The total list of stimuli comprised 40 sentences (of which 20 contained single features, 15 contained combinations of two features, and 5 contained all four features). A female trained phonetician recorded the full list of stimuli using the Model Speech and also in her natural Northeastern US accent (for comparison). The consistent presence of all features in the artificial accent was verified acoustically (see Figure 1).



Tapping: [ɪli] (baseline) and [ɪɾi] (Model Speech)

Diphthongization: [ɛ] (baseline) and [jɛ] (Model Speech)

Epenthesis: [spɪn] (baseline) and [səp^hɪn] (Model Speech)

Tag question: LMH (baseline) and MLH (Model Speech)

Figure 1: *Examples of the 4 features in baseline (left) and Model Speech (right). The VCV sequence for tapping was extracted from the word 'chilly'. The tracks of formants 1-4 are obtained from the vowel in 'Ben'. The spectrograms for epenthesis are obtained from the word 'spinning'. Pitch tracks for the sequence 'wasn't he?' illustrate intonation change.*

2.3. Participants

The participants were 36 undergraduate students, of which 17 were monolingual (mean age = 23.9, 6 male, 11 female) and 19 early bilingual (mean age = 22.5, 3 male, 16 female). Early bilinguals had had consistent exposure to two languages starting before the age of 3 years, and currently displayed native or near-native proficiency in both languages. Bilinguals' other languages included Arabic, Cantonese, Hebrew, Russian, Spanish, Urdu, Thai, and (Haitian / Jamaican / St. Lucian) Creole. Both age of acquisition and proficiency level were self-reported.

2.4. Procedure

The experiment was run using PsychoPy [17]. The participants recorded 40 baseline sentences, and then received training by first listening to 40 different sentences spoken in the novel accent. The sentences were accompanied by orthographic transcription. After listening to the full set once, each sentence was played once more and the participants were asked to imitate it immediately after hearing it. In the testing phase, they were asked to read the baseline sentences again, this time aiming to reproduce the novel accent in the absence of any audio prompts.

2.5. Data processing and analysis

Data processing consisted of (a) categorical judgments provided by a trained phonetician, scoring the absence or presence of each target feature with a 0 or 1, resulting in a mean accent score for each subject, and (b) measurements of continuous parameters such as duration, pitch and formant values, using the Praat software for acoustic analysis [18]. Before extracting acoustic measurements, the portions of interest were manually aligned. For diphthongization, F1 and F2 values were obtained from the first 50% of each vowel's duration, extracted at 5% intervals for a total of 11 steps. For tag questions, pitch values were similarly obtained throughout the duration of the question, from 20 distinct time steps. Durations were normalized for both vowels and tag questions.

The statistical analyses we conducted included a univariate ANOVA for the categorical judgment scores, and additional ANOVAs that compared various aspects of the two groups' performance across the different blocks, detailed in the following sections. For the analysis of diphthongization, the dependent variables were duration and mean formant frequencies. For tag questions, we first converted the F0 values to Z scores. We then designed an algorithm in Python [19] to compute the F0 change from each point in the time series to the points to its right and we identified three measures as follows: (1) *DropML*, referring to the largest drop (which in a tag question produced with the novel pattern would correspond to the drop from Mid to Low), (2) *RiseMH*, i.e. the largest mid to high change between the first identified peak and the last peak, and (3) *RiseLH*, referring to the largest low to high change, computed as the difference between the lowest drop and the last peak. Drops and rises were computed as percentage changes.

2.6. Results

2.6.1. Categorical scoring: All four features

Figure 2 shows the average scores for monolinguals and bilinguals for all four novel features in the three different conditions (i.e. Baseline, Training, Testing). Bilinguals outperformed monolinguals across the board, in both the Training (imitation) and Testing conditions, but the differences in Training were more pronounced with tapping and tag questions. In Testing, monolinguals performed best with tag questions, followed by epenthesis, and they did not show any learning of the tapping feature. A univariate ANOVA with *Score* as the dependent variable and *Group (monolingual / bilingual)*, *Block (baseline / training / testing)*, *Feature (dipthongization / tapping / epenthesis / tag question)* and *Number of features per sentence (1/2/4)* as independent variables revealed significant main effects of all independent variables (Group: F(1, 7544)=194.2, p $<$ .001, Block: F(2, 7544)=1084.9, p $<$ .001, Feature: F(3, 7544)=22.7, p $<$ .001, and # features/sentence F(2, 7544)=64.6, p $<$ .001), and also of the interactions between Group $\times$ Block, Group $\times$ Feature, Block $\times$ Feature, Block $\times$ Number of features per sentence, and Feature $\times$ # features/sentence. Post hoc tests using the Bonferroni correction revealed that each block differed significantly from the other two, and tapping differed significantly from all other features.

2.6.2. Acoustic measurements: Diphthongization

While the average vowel durations were higher in training and testing for both groups (Monolinguals: baseline 100.2 ms, training 131.7 ms, testing 126 ms; Bilinguals: baseline 92.4 ms, training 138.4 ms, testing 129.9 ms), no significant differences were found between the two groups. The formant values, however, revealed greater separation of F1 and F2 for the bilingual group in both Training and Testing (Figure 3). The change across blocks is more pronounced for the second formant, but the differences between groups were statistically significant for both formants. Specifically, we computed the F2-F1 difference, averaged over all 11 time points, for each participant and conducted a univariate ANOVA to determine whether this difference was significantly affected by Group and Block. This was indeed the case (Group: F(1, 2294)=26.8, p<.001; Block: F(3, 2294)=332.8, p<.001). The interaction between Group $\times$ Block was also significant, F(2, 2294)=10.5, p<.001. Significant differences between baseline and both training and testing were found for both groups, but the F2-F1 difference was found to be significantly larger for bilinguals, F(1, 2298)=19, p<.001.

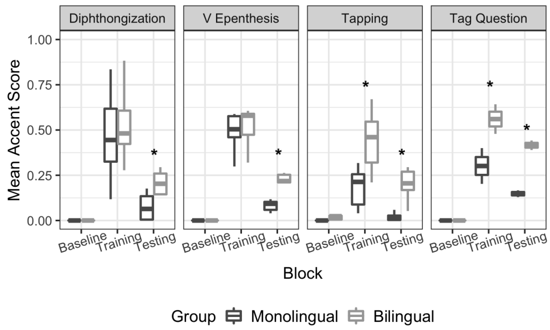


Figure 2: *Mean accent scores for each novel accent feature obtained by monolinguals and bilinguals in Baseline, Training and Testing. Significant differences are marked with an asterisk.*

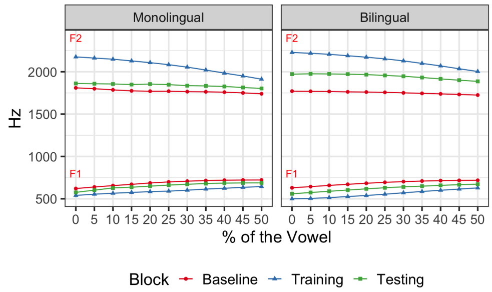


Figure 3: *Formant values, traced throughout the first 50% of the normalized vowel duration for the two groups of speakers.*

2.6.3. Acoustic measurements: Tag Questions

Figure 4 displays the pitch contours for tag questions in the model speaker (both for her baseline accent and the artificially constructed accent) and the two groups of participants in the three different blocks. The model speaker and the two groups exhibit the same baseline tag question pattern, specifically a rising one (implemented as a low to high progression). Both groups altered their tag question intonation in training and testing, though the bilinguals appear to have approximated the new pattern to a slightly greater extent. The high terminal was exhibited by both groups in both training and testing, but the dip from mid to low is more apparent with the bilinguals as a group. The ANOVA results showed a significant effect of Group on DropML, F(1, 1242)=3.8, p=.05, and of Block on RiseMH, F(2, 1242)=8.2, p<.001, and RiseLH, F(2, 1242)=12.5, p<.001. There was no significant interaction of Group $\times$ Block.

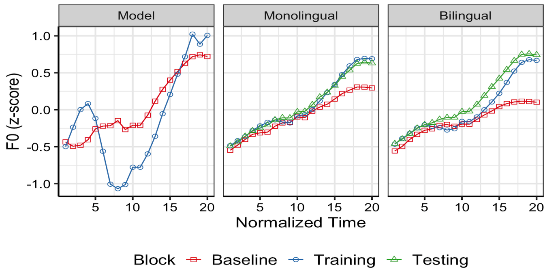


Figure 4: *Pitch tracks for the tag questions of the model speaker and the two groups. NOTE: Model Speech labeled as Training.*

To eliminate the sources of variability due to collapsing both successful and unsuccessful productions, Figure 5 displays the pitch contours for only those tag questions that had received a score of 1 on the subjective accent score. This reveals that even though as a group monolinguals did not outperform the bilinguals in the initial learning of this feature, those who did learn the feature were able to reproduce it more accurately compared to the bilinguals who learned it. This can be assessed visually as their pitch tracks more closely resemble those of the Model Speech. At the same time, presumably due to the reduced numbers of successful productions, this pattern does not reach statistical significance. The findings thus make it difficult to evaluate Hypothesis 2, as on the one hand the bilinguals showed a significant drop from Mid to Low, but on the other hand the monolinguals who did learn the novel pattern exhibited a pitch contour that is closer to that of the Model Speaker.

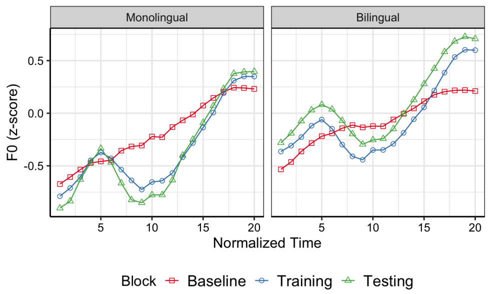


Figure 5: *Pitch tracks for the tag questions that received a subjective rating of 1 (i.e. MLH pattern deemed present).*

3. Discussion

Our findings support Hypothesis 1, as bilinguals exhibited a clear advantage in terms of subjective categorical ratings as far as the novel segmental properties of the Model Speech were concerned. This advantage was further confirmed with acoustic data for the process of diphthongization and novel tag question intonation. Hypothesis 2 received only partial and inconsistent support. The categorical ratings showed that bilinguals as a group are more successful learners of the intonational pattern compared to monolinguals. A closer look at the participants' productions, however, revealed that the monolinguals who did learn the novel pattern were closer to the Model Speech compared to bilinguals. This underscores the importance of conducting both quantitative and qualitative analyses in phonetic learning studies, as the two may not always converge.

Overall, our findings successfully replicate previous work reporting a bilingual advantage in phonetic learning. Compared to the more extensively studied executive function advantages, for which conflicting results were reported in the literature [7, 4], the advantage found with phonetic learning appears robust, despite differences in participants' bilingual background. Since different aspects of phonetic learning were tested here (i.e. vocalic, consonantal, syllabic and intonational processes), we can tentatively conclude that the specific languages spoken by the participants did not greatly affect the outcome, and the observed advantage correlates with the state of being bilingual. It must be noted however that the structures we tested did not involve any sounds that were entirely foreign to our participants. In addition, subtler differences may arise in future analyses of our data, in line with previous claims that phonetic learning is modulated by the degree of similarity between the phonologies existent in the bilinguals' repertoire and the universal difficulty of the phonetic features learnt [13].

The differences in phonetic learning between monolinguals and bilinguals may be tentatively explained by at least two (non mutually exclusive) approaches. On the one hand, the different performance with (shorter) segmental and (longer and more salient) suprasegmental features could be due to differences in auditory sensory memory [20, 9]. On the other hand, these differences might also be ascribed to conscious versus more automatic learning strategies, as tag questions, on which the monolinguals performed better compared to the other features, were arguably the most salient of all, due to having the longest duration and sentence-final position, possibly creating a recency effect [21]. Future work should be aimed at identifying some of the neural mechanisms underlying the observed phonetic learning behavior of monolinguals and bilinguals, specifically whether novel accent learning tasks reveal differences in terms of conscious awareness and regulation of accent features (typically associated with the insula, [22]) and whether monolinguals exhibit more effortful phonetic processing (as possibly reflected by greater neural recruitment in the articulatory-auditory network, [23]). The observed differences in phonetic learning may ultimately be due to the two groups' recruiting different cognitive resources to achieve learning, with more conscious and effortful processing in the case of monolinguals.

4. Conclusion

This study documents the learning of an artificially constructed, naturalistically produced accent of English by monolinguals and bilinguals. Our findings indicate potential differences in the learning strategies of speakers with different linguistic backgrounds. Learning following short initial exposure and training was more effective in bilinguals compared to monolinguals, a finding that applied to all four linguistic aspects tested. At the same time, even though fewer monolinguals successfully learned the novel intonation pattern, their production more accurately reproduced this pattern when they did learn it, compared to bilinguals. Considering also that monolinguals did not learn the tapping pattern, which had the shortest duration, another potential explanation arises in terms of auditory sensory memory. Higher auditory sensory memory span in bilinguals may feed more effectively into working memory, where rehearsal is thought to occur, consolidating transmission to long-term memory (and consequently learning).

5. Acknowledgements

We gratefully acknowledge the support of PSC-CUNY Grant # 61667-00 49 awarded to the first author.

6. References

[1] J. Bartolotti, and V. Marian. “Bilinguals’ existing languages benefit vocabulary learning in a third language”. *Language learning*, 67(1), 2017, pp. 110–140.

[2] D. Thomas-Sunesson, K. Hakuta, and E. Bialystok. “Degree of bilingualism modifies executive control in Hispanic children in the USA”. *International journal of bilingual education and bilingualism*, 21(2), 2018, pp. 197–206.

[3] M. A. Warmington, S. Kandru-Pothineni, and G. J. Hitch.“Novel-word learning, executive control and working memory: A bilingual advantage”. *Bilingualism: Language and Cognition*, 22(4), 2019, pp. 763–782.

[4] A. Marzecová. “Bilingual advantages in executive control–A Loch Ness Monster case or an instance of neural plasticity”. *Cortex*, 73, 2015, pp. 364–366.

[5] A. Dick, N. L. Garcia, S. M. Pruden, W. K. Thompson, S. W. Hawes, M. T. Sutherland, ... and R. Gonzalez, “No bilingual advantage for executive function: Evidence from a large sample of children in the Adolescent Brain and Cognitive Development (ABCD) Study”. *PsyArXiv*. 2018, July, 8.

[6] K. R. Paap, H. A. Myuz, R. T. Anders, M. F. Bockelman, R. Mikulinsky, and O. M. Sawi, “No compelling evidence for a bilingual advantage in switching or that frequent language switching reduces switch cost”. *Journal of Cognitive Psychology*, 29(2), 2017, pp. 89–112.

[7] E. Bialystok, “Bilingualism and executive function: What’s the connection? In D. Miller, F. Bayram, J. Rothman, and L. Serratrice (Eds.), *Bilingual cognition and language: “The state of the science across its subfields” Philadelphia, PA*: John Benjamins. 2018, pp. 283–306.

[8] E. Bialystok, F. I. M. Craik, and G. Luk, G. “Bilingualism: Consequences for Mind and Brain”. *Trends in Cognitive Sciences*, 16(4), 2012, pp. 240–250.

[9] L. Spinu, J. Hwang, and R. Lohmann, “Is there a bilingual advantage in phonetic and phonological acquisition? The initial learning of word-final coronal stop realization in a novel accent of English”. *International Journal of Bilingualism*, 22(3), 2018, pp. 350–370.

[10] M. C. Tremblay, and L. Sabourin, L. “Comparing behavioral discrimination and learning abilities in monolinguals, bilinguals and multilinguals”. *The Journal of the Acoustical Society of America*, 132(5), 2012, pp. 3465–3474.

[11] M. Antoniou, E. Liang, M. Ettlinger, and P. C. M. Wong, “The bilingual advantage in phonetic learning”. *Bilingualism: Language and Cognition*, 18(4), 2015, pp. 683–695.

[12] J. Krizman, V. Marian, A. Shook, E. Skoe, and N. Kraus, “Subcortical encoding of sound is enhanced in bilinguals and relates to executive function advantages”. *In Proceedings of the National Academy of Sciences 109*. 2012, pp. 787–7881.

[13] R. Kopeckova, “The bilingual advantage in L3 learning: a developmental study of rhotic sounds”, *International Journal of Multilingualism*, 13:4, 2016, pp. 410–425, DOI: 10.1080/14790718.2016.1217605

[14] Y. Kondratenko, and L. Spinu, “Being "better" with accents: evidence from bilinguals”. *Proceedings of the Forth-eighth Annual fteeting of the Chicago Linguistic Society*, eds. A. Beltrama, T. Chatzikonstantinou, J. L. Lee, M. Pham, and D. Rak, 2014, pp. 387–400.

[15] L. Spinu, and Y. Rafat, “The initial learning of intonation patterns in a novel dialect of English by monolinguals and bilinguals: An unexpected monolingual advantage”. *Proceedings of the International Symposium on Monolingual and Bilingual Speech 2019*, E. Babatsouli (ed.), 2019, pp. 122–129. ISBN: 978-618-82351-3-7

[16] J. A. Sereno, L. Lammers, and A. Jongman, “The relative contribution of segments and intonation to the perception of foreign-accented speech”. *Applied Psycholinguistics*, 37(2), 2016, pp. 303–322.

[17] J. W. Peirce, J. R. Gray, S. Simpson, M. R. MacAskill, R. Höchenberger, H. Sogo, E. Kastman, J. Lindeløv, “PsychoPy2: experiments in behavior made easy”. *Behavior Research Methods*. 2019, 10.3758/s13428-018-01193-y

[18] P. Boersma, and D. Weenink, “Praat: doing phonetics by computer [Computer program]”. Version 6.1.14, retrieved 2 May 2019 from http://www.praat.org/, 2019.

[19] Van Rossum, G., & Drake Jr, F. L. “Python reference manual”. Centrum voor Wiskunde en Informatica Amsterdam. 1995.

[20] A. Calabrese. “Auditory representations and phonological illusions: A linguist’s perspective on the neuropsychological bases of speech perception”. *J of Neurolinguistics* 25, 2012, pp. 355–381.

[21] G. Vallar, and C. Papagno. “Phonological short-term store and the nature of the recency effect: Evidence from neuropsychology”. *Brain and Cognition*, 5(4), 1986, pp. 428–442.

[22] L. Ghazi-Saidi,T. Dash, and A. I. Ansaldo, “How native-like can you possibly get: fMRI evidence for processing accent”. *Frontiers in human neuroscience*, 9, 2015, pp. 587.

[23] N. Golestani. “Neuroimaging of phonetic perception in bilinguals.” *Bilingualism: Language and Cognition*, 19(4), 2016, pp. 674–682.