

STATIC AND DYNAMIC PHONETIC INTERACTIONS IN THE L2 AND L3 ACQUISITION OF JAPANESE VELAR VOICELESS STOPS

Mark Amengual, Lizzie Meredith, Talia Panelli

University of California, Santa Cruz
amengual@ucsc.edu

ABSTRACT

This production experiment investigates how L1 English-L2 Japanese bilinguals and L1 Spanish-L2 English-L3 Japanese trilinguals produce voiceless velar stops /k/ in each of their languages in order to ascertain if these speakers create separate phonological categories in Spanish, English, and Japanese. The results show that even though the bilingual and multilingual individuals maintain language-specific VOT patterns for each language, they produce /k/ in bilingual and trilingual sessions that display less native-like VOT values in each language in comparison to the same productions in the monolingual sessions. These findings demonstrate that increased activation of the non-target language(s) in the multilingual experimental condition creates cross-linguistic phonetic convergence in the acoustic realization of these phonemic categories, however, it does not impede these multilingual individuals from maintaining language-specific categories in each language. It is proposed that a model of multilingual processing based on the principles of episodic frameworks can explain these findings.

Keywords: VOT, L2 acquisition, L3 acquisition, language mode, speech production

1. INTRODUCTION

Many studies on bilingualism have investigated the influence of the first language (L1) on the production and perception of language-specific phonological contrasts in the second language (L2). In recent years, however, the interest in cross-linguistic influence has expanded into third language (L3) acquisition, raising issues regarding the suitability of current models of L2 phonology as a framework for the study of L3 phonological acquisition.

Despite the fact that L3 acquisition is a fairly new field of research and that phonology remains the least explored area [11], recent studies in L3 phonology have shown that L3 speakers produce L1-accented speech in their L3 [28], L2 phonological transfer into the L3 [22], a combined influence of the L1 and L2 on L3 pronunciation [32], and influence of the L3 on the L1 and L2 [7]. However, studies comparing L2 and L3 phonological acquisition are still scarce and we still do not have a clear understanding of how the

acquisition of phonological categories in a language varies as a function of being a bilingual or trilingual speaker. In addition, the effects of the language processing mechanisms at a given point in time during speech production (i.e., state of activation of each of a bilingual/trilingual's languages) also remains a largely underexplored area in multilingual individuals. This study bridges the divide between the L2 and L3 acquisition literatures with regard to the investigation of the acoustic realization of word-initial /k/ in the speech of L1 English-L2 Japanese bilinguals, and early Spanish-English simultaneous bilinguals learning Japanese as a L3.

1.1. VOT in English, Spanish, and Japanese

Voice Onset Time (VOT) refers to the relative timing of the release of the air for a stop consonant and the onset of phonation (voicing) of a following vowel. This acoustic cue is determined at a language-specific level [9,19,21] and can be classified within a continuum varying in degrees of aspiration.

Spanish /p, t, k/ have a short VOT and are always unaspirated [p, t, k]. Previous studies have reported a Spanish VOT between zero and 20 ms. [1,10]. English voiceless stops on the onset of a stressed syllable are produced with a long lag and are aspirated [p^h, t^h, k^h] except when they follow /s/ [17]. As a result, English voiceless stops have a substantial delay between the release and the onset of laryngeal vibration, resulting in a VOT from 30 ms. to 120 ms., corresponding to the aspiration interval [9]. In contrast to Spanish and English, Japanese is a language whose voiceless stops may not fit neatly into the short lag and long lag dichotomy [27]. Homma [16] and Shimizu [29] reported VOT means for /p, t, k/ between 25ms. and 65ms. Furthermore, VOT data of monolingual Japanese speakers indicate that their voiceless stops fall between the short lag and long lag boundaries suggested by Lisker & Abramson [19], with mean values of 30, 25.5, and 56.7 ms. for /p/, /t/, and /k/ [27].

1.2. Static and dynamic phonetic interactions

Phonetic interaction can occur as a result of long-term traces of one language influencing the other [30]. This is a static process as a result of cross-linguistic interactions between long-term memory representations. An extensive body of research has

provided evidence of this influence at the phonetic level, investigating language-specific VOT values [3,8,20,31].

The second type of cross-linguistic influence is more transient and occurs during short-term operations. This dynamic process is the result of interactions between representations that are simultaneously activated in working-memory. If we assume that when speakers are using two languages the representations of both languages are simultaneously activated, thus creating competition between the two languages, we can expect a deviation of the target phonetic implementation towards the non-target language during online speech processing. Grosjean [15] distinguishes these two types of cross-linguistic influence as *transfer* and *interference*, respectively. While recent work has examined the transient cross-linguistic interference of phonetic interaction in bilinguals [2,14,23,24], there are no previous studies focusing on static and dynamic phonetic interaction in both bilingual and multilingual speakers.

The present study investigates the acoustic realization of Spanish, English, and Japanese /k/ by English-Japanese bilinguals and Spanish-English-Japanese trilinguals. The experiment assumes that bilingual and trilingual mode can be induced experimentally if the stimuli come from each language and the task asked of the participants requires processing of multiple languages.

2. METHOD

2.1. Participants

Fifteen participants were recruited to participate in the production experiment. The sample comprised five L1 English-L2 Japanese bilinguals and five L1 English-L2 Spanish-L3 Japanese trilinguals. Five L1 Japanese native speakers were also recruited as a control group. Ages ranged from 19 to 22 ($M=20.8$, $SD=1.1$). They reported normal speech and hearing, and normal or corrected to normal vision.

The L1 English-L2 Japanese bilinguals ($N=5$) had been raised in a monolingual English household in the U.S., spoke English as their native language, and were not native speakers of any other language. The L1 English-L2 Spanish-L3 Japanese trilinguals ($N=5$) consisted of Spanish heritage speakers who had been raised and educated in a bilingual environment in the U.S., having extensive exposure to both Spanish and English on a daily basis. At the time of testing these learners were enrolled in Japanese 4 at UC Santa Cruz, and would be considered to be at the novice high sublevel of Japanese language study, according to the ACTFL Proficiency Guidelines. Finally, the L1 Japanese-L2 English bilingual control group ($N=5$) were native Japanese speakers, enrolled in a one-year study abroad program at UC Santa Cruz.

2.2. Materials and recording procedure

The production of the target word-initial velar voiceless stops in Spanish, English, and Japanese was elicited in a reading-aloud task. The materials consisted of 14 experimental items (3 repetitions) for each language. All target experimental words consisted of word-initial /k/ followed by a low mid vowel /a/. The target voiceless velar stop /k/ appeared in a stressed syllable in English but unstressed in Spanish and Japanese. Each target item was embedded in a carrier phrase *I say TARGETWORD for you* (English), *Digo TARGETWORD para ti* (Spanish), and *また TARGETWORD といってください*, ‘Mata TARGETWORD to itte kudasai’ (Japanese). The Spanish and English carrier phrases were presented in roman script and the Japanese carrier phrases in hiragana (ひらがな) script.

The production experiment was conducted individually in a sound-attenuated booth with participants comfortably seated in front of a computer display. Each sentence was presented on a computer screen for five second and participants were asked to read the sentences clearly and with a natural pace, speaking neither too quickly nor too slowly. The speech samples were recorded using a head-mounted microphone (Shure SM10A) and an audio interface (MOTU Ultra Lite mk3), digitized (44kHz, 16-bit quantization), and computer-edited for subsequent acoustic analysis.

The crucial manipulation in the present study is that participants were asked to participate in separate experimental sessions. In the monolingual sessions participants read words only in the target language: either Spanish, English, or Japanese. In the bilingual session, the bilingual groups read Spanish and English or English and Japanese, or Spanish and Japanese carrier phrases presented together in randomized order. Finally, the trilingual group also participated in a trilingual session in which Spanish, English, and Japanese carrier phrases were presented aleatorily. Within the bilingual and trilingual recording sessions, participants produced equal numbers of experimental words in each language, and the same items were presented in the monolingual, bilingual, and trilingual sessions.

2.3. Acoustic analysis

The VOT values of the target stops were obtained from the waveform using Praat [5]. VOTs were obtained by measuring the time interval between the stop release and the onset of voicing as discerned on the waveform as periodic (repeating) cycles. The measurement (rounded to the nearest decimal) was determined from the beginning of the burst (identified by a sharp spike where the waveform changes from quiescent to transient) to the beginning of the first

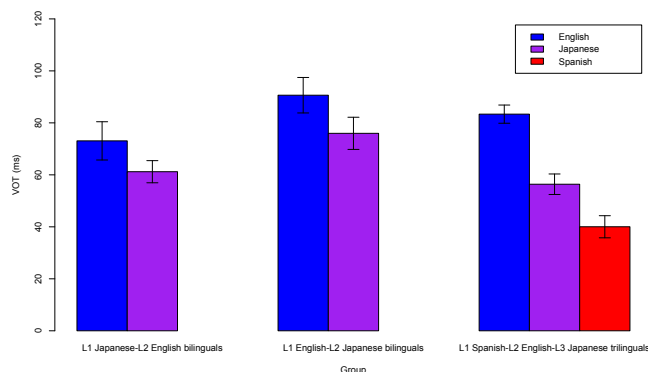
regularly repeating voicing cycle. The point in the first glottal cycle that was counted as the onset of voicing was the initial zero crossing in the waveform.

Each participant produced either 168 (bilinguals) or 504 (trilinguals) target items, for a total of 4,200 VOT tokens. 79 tokens were excluded due to mispronunciations or recording errors. As a result, the dataset comprised a total of 4,121 measurements.

3. RESULTS

In order to compare the VOT values of bilingual and trilingual participants as a function of their language and experimental session, a dataset was created including the average over subjects as a condition of language and session. As shown in Figure 1, each group maintains language-specific VOT values in their production of /k/ in each language.

Figure 1: VOT values (ms.) as a function of group and language.



The dataset was submitted to a mixed-model ANOVA, which was performed using R [26], with speaker group as between-subjects factor, language and experimental session as within-subjects factors, and subject as the random terms. The mixed-design ANOVA yielded significant main effects of *speaker group* ($F(2,12) = 11.7, p < 0.001$), *language* ($F(2,168) = 217.4, p < 0.001$) and *experimental session* ($F(6,168) = 18.4, p < 0.001$). In addition, there was a significant interaction between *speaker group* and *language* ($F(2,168) = 7.8, p < 0.001$).

The results of each fixed factor and the significant interaction between speaker group and language motivated the division of the dataset into speaker group subsets. The results of the interactions explored with pairwise comparisons using Tukey's HSD test are reported separately below for each group. In all cases, paired *t*-tests and effect sizes calculated by means of paired Cohen's *d* are reported. Figures 2, 3, and 4 present the VOT values (ms.) for each speaker group in the production of /k/ in each experimental session.

Figure 2: VOT values (ms.) of productions in monolingual and bilingual mode by the L1 English-L2 Japanese bilingual group

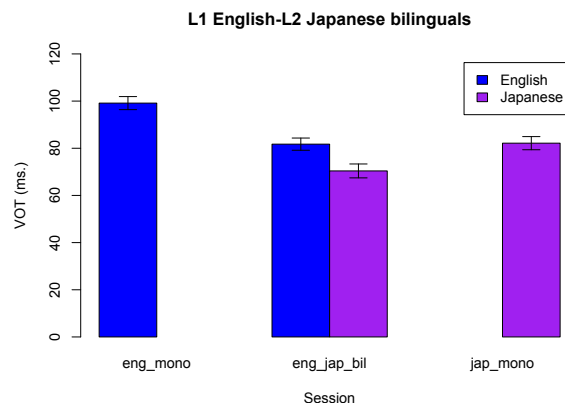


Figure 3: VOT values (ms.) of productions in monolingual and bilingual mode by the Japanese native speaker control group.

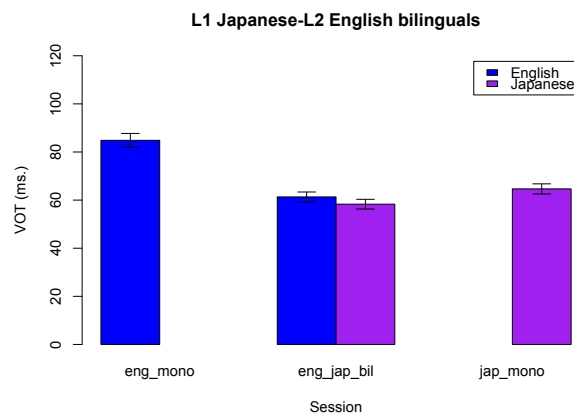
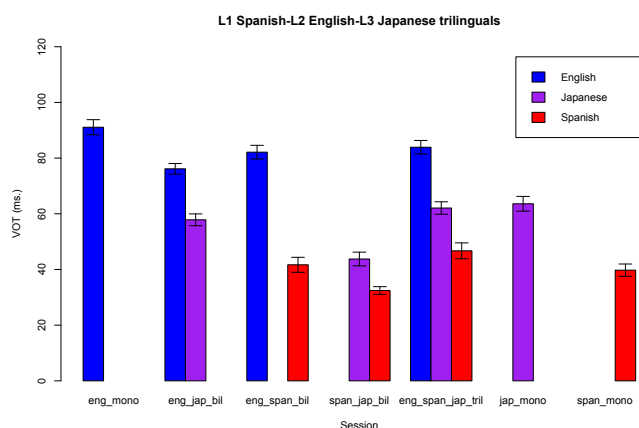


Figure 4: VOT values (ms.) of productions in monolingual, bilingual, and trilingual mode by the Spanish-English-Japanese trilingual group.



The analysis of the L1 English-L2 Japanese bilingual group revealed a significant difference between the VOT values of /k/ produced in monolingual mode and bilingual mode in English (diff. = 18, $t(9) = -3.24, p < 0.001$, Cohen's $d = 1.03$) and in Japanese (diff. = 11.2, $t(9) = -3.15, p < 0.001$, Cohen's $d = 1.009$). The pairwise comparisons investigating the effects of the experimental session in the production of English /k/ by the L1 Japanese-

L2 English bilingual group yielded a significant difference between monolingual and bilingual mode (diff. = 23.6, $t(9) = -6.57$, $p < 0.001$, Cohen's $d = 2.08$). This language mode difference was also found in their production of Japanese /k/ (diff. = 5.6, $t(9) = -3.35$, $p < 0.001$, Cohen's $d = 1.07$). Finally, a comparison of each combination of monolingual, bilingual, and trilingual experimental session was made for the VOT values of each language produced by the L1 Spanish-L2 English-L3 Japanese trilingual group. Table 1 presents the results of the pairwise comparisons on the VOT values for each language and experimental session.

Table 1: Pairwise comparisons on the VOT values of the L1 Spanish-L2 English-L3 Japanese group for each language and experimental session.

English	
English mono - Eng-Jap bil	diff. = 15, $t(9) = -5.99$, $p < 0.001$, Cohen's $d = 1.89$
English mono - Eng-Span bil	diff. = 9.2, $t(9) = 3.61$, $p < 0.01$, Cohen's $d = 1.13$
English mono - Eng-Span-Jap	diff. = 7.3, $t(9) = 5.01$, $p < 0.001$, Cohen's $d = 1.58$
Eng-Jap bil - Eng-Span bil	diff. = 5.8, $t(9) = -3.08$, $p < 0.01$, Cohen's $d = 0.98$
Eng-Jap bil - Eng-Span-Jap	diff. = 7.7, $t(9) = -4.26$, $p < 0.01$, Cohen's $d = 1.35$
Eng-Span bil - Eng-Span-Jap	diff. = 1.9, $t(9) = -1.12$, n.s
Japanese	
Japanese mono - Eng-Jap bil	diff. = 5.8, $t(9) = -2.54$, $p < 0.05$, Cohen's $d = 0.80$
Japanese mono - Span-Jap bil	diff. = 21.6, $t(9) = 3.6$, $p < 0.01$, Cohen's $d = 1.13$
Japanese mono - Eng-Span-Jap	diff. = 1.2, $t(9) = -0.78$, n.s
Eng-Jap bil - Spn-Jap bil	diff. = 15.8, $t(9) = 3.09$, $p < 0.01$, Cohen's $d = 0.97$
Eng-Jap bil - Eng-Span-Jap	diff. = 4.5, $t(9) = -3.43$, $p < 0.001$, Cohen's $d = 1.09$
Span-Jap bil - Eng-Span-Jap	diff. = 20.3, $t(9) = 3.43$, $p < 0.001$, Cohen's $d = 1.08$
Spanish	
Spanish mono - Eng-Span bil	diff. = 1.6, $t(9) = 1.4$, n.s
Spanish mono - Span-Jap bil	diff. = 7.2, $t(9) = -1.55$, n.s
Spanish mono - Eng-Span-Jap	diff. = 6.5, $t(9) = 2.09$, $p < 0.05$, Cohen's $d = 0.65$
Eng-Span bil - Span-Jap bil	diff. = 8.8, $t(9) = 1.67$, n.s
Eng-Span bil - Eng-Span-Jap	diff. = 4.9, $t(9) = -1.33$, n.s
Span-Jap bil - Eng-Span-Jap	diff. = 13.8, $t(9) = 3.01$, $p < 0.01$, Cohen's $d = 0.95$

4. DISCUSSION AND CONCLUSIONS

The present study analyzed the acoustic realization of the Spanish, English, and Japanese voiceless velar stop /k/ produced by two groups of English-Japanese bilinguals and a Spanish-English-Japanese trilingual group. The experiment also examined the effects of language activation (i.e., language mode) on the acoustic realization of /k/ in their L1, L2, and L3.

The results of the production task indicate that voiceless velar stops in Japanese, English, and Spanish differ in their VOT values. Specifically, all participant groups produce /k/ with a longer VOT in English and with a shorter VOT in Japanese. Furthermore, the trilingual group produces an even shorter VOT in Spanish, thus maintaining language-specific VOT patterns for each language.

Even though these bilingual and multilingual individuals maintain language-specific VOT patterns for each language, the acoustic analyses also reveal phonetic convergence as a result of language mode. Bilingual and trilingual individuals alike produced voiceless velar stops in bilingual and trilingual sessions that mostly displayed less native-like VOT

values in each language in comparison to the same productions in the monolingual sessions. Interestingly, not all languages of the trilingual individuals are affected in the same way: their L1 (Spanish) VOT values seem to be least affected by changes in the experimental session than English and Japanese. This is a remarkable feat if we consider that more precision may be required to maintain short lag Spanish stops in comparison to a wider span of VOT values for English and Japanese voiceless stops in bilingual and multilingual speech. The fact that larger differences were not found in their Spanish production of short-lag voiceless stops, resulting in less variability, may be explained by a gestural account as suggested by Kessinger and Blumstein [18]: lengthening a short-lag VOT requires an additional gesture (aspiration). In the same vein, Beckman et al. [4] explain that in aspirating languages lengthening a short-lag VOT would reduce or eliminate the contrast, while in prevoicing languages (such as Spanish) this gesture would simply be additional effort.

Taken together, these findings provide evidence of transient (dynamic) interlingual interference on the acoustic realization of /k/ due to the increased activation of the non-target language(s) in the bilingual and multilingual experimental conditions, even if it does not impede these multilingual individuals from maintaining language-specific categories in Spanish, English, and Japanese, thus overcoming long-term (static) traces of one language influencing the other.

A model of multilingual representation and processing based on the principles of episodic (or complementary-systems) frameworks can explain these findings [12,13]. Exemplar models of lexical representation [6,25] are able to account for the effects of the increased activation of two or three languages as a result of being in bilingual or trilingual mode. An increase in the strength of activation of words in bilingual/multilingual mode could be responsible for displacing by means of assimilation the degree of aspiration (i.e., VOT) of interconnected clouds of data, which would result in a "gestural drift". The production of words in bilingual and trilingual mode potentially draws from all language exemplars instead of restricting the possible targets to the language-specific exemplars available for each language separately in the monolingual sessions.

5. ACKNOWLEDGEMENTS

We would like to thank Sara Hatter, Emma Peyton, Daisy Soto, Yukari Iventosch, and Ruben Alfredo for their assistance with data collection and the acoustic measurements. We would also like to thank Maho Morimoto and Shigeko Okamoto for their feedback on the experimental materials.

6. REFERENCES

- [1] Abramson, A. S., Lisker, L. 1973. Voice-timing perception in Spanish word-initial stops. *Journal of Phonetics*, 1, 1–8.
- [2] Amengual, M. 2018. Asymmetrical interlingual influence in the production of Spanish and English laterals as a result of competing activation in bilingual language processing. *Journal of Phonetics*, 69, 12–28.
- [3] Antoniou, M., Best, C. T., Tyler, M. D., Kroos, C. 2010. Language context elicits native-like stop voicing in early bilinguals' productions in both L1 and L2. *Journal of Phonetics*, 38, 640–653.
- [4] Beckman, J., Helgason, P., McMurray, B., Ringen, C. 2011. Rate effects on Swedish VOT: Evidence for phonological overspecification. *Journal of Phonetics*, 39–49.
- [5] Boersma, P., Weenink, D. 2018. Praat: Doing phonetics by computer [Computer program]. Version, 6, 14.
- [6] Bybee, J. 2001. *Phonology and language use*. Cambridge: Cambridge University Press.
- [7] Cabrelli Amaro, J. 2016. Testing the Phonological Permeability Hypothesis: L3 phonological effects on L1 versus L2 systems. *International Journal of Bilingualism*, 1–35
- [8] Caramazza, A., Yeni-Komshian, G. H., Zurif, E. B., Carbone, E. 1973. The acquisition of a new phonological contrast: The case of stop consonants in French–English bilinguals. *Journal of the Acoustical Society of America*, 54, 421–428.
- [9] Cho, T., Ladefoged, P. 1999. Variation and universals in VOT: Evidence from 18 languages. *Journal of Phonetics*, 27, 207–229.
- [10] Flege, J.E., Eefting, W. 1986. Linguistic and developmental effects on the production and perception of stop consonants. *Phonetica*, 43, 155–171.
- [11] Gallardo del Puerto, F. 2007. Is L3 phonological competence affected by the learner's level of bilingualism? *International Journal of Multilingualism*, 4, 1–16.
- [12] Goldinger, S. D. 1997. Words and voices: Perception and production in an episodic lexicon. In: Johnson, K., Mullennix, J. (eds), *Talker variability in speech processing*. San Diego: Academic Press, 33–66.
- [13] Goldinger, S. D. 1998. Echoes of echoes?: An episodic theory of lexical access. *Psychological Review*, 105, 251–279.
- [14] Goldrick, M., Runnqvist, E., Costa, A. 2014. Language switching makes pronunciation less nativelike. *Psychological Science*, 25, 1031–1036.
- [15] Grosjean, F. 2011. An attempt to isolate, and then differentiate, transfer and interference. *International Journal of Bilingualism*, 16, 11–21.
- [16] Homma, Y. 1980. Voice onset time in Japanese stops. *Onsei Gakkai Kaiho*, 163, 7–9
- [17] Kent, R. D., Read, C. 1992. *The acoustic analysis of speech*. San Diego, CA: Singular Publishing Group.
- [18] Kessinger, R.H., Blumstein, S.E. 1997. Effects of speaking rate on voice-onset time in Thai, French, and English. *Journal of Phonetics*, 25, 143–168.
- [19] Lisker, L., Abramson, A. 1964. A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, 20, 384–422.
- [20] Macleod, A.A.N., Stoel-Gammon, C. 2005. Are bilinguals different? What VOT tells us about simultaneous bilinguals. *Journal of Multilingual Communication Disorders*, 3, 118–127.
- [21] Maddieson, I. 1984. *Patterns of sounds*. Cambridge: Cambridge University Press.
- [22] Marx, N. 2002. Never quite a “native speaker”: accent and identity in the L2 and the L1. *The Canadian Modern Language Review*, 59, 264–281–38.
- [23] Olson, D. J. 2013. Bilingual language switching and selection at the phonetic level: Asymmetrical transfer in VOT production. *Journal of Phonetics*, 41, 407–420.
- [24] Olson, D. J. 2016. The role of code-switching and language context in bilingual phonetic transfer. *Journal of the International Phonetic Association*, 46, 263–285.
- [25] Pierrehumbert, J. B. 2001. Exemplar dynamics: Word frequency, lenition, and contrast. In: Bybee, J., Hopper, P. (eds), *Frequency and the emergence of linguistic structure*. Philadelphia, PA: John Benjamins, 137–157.
- [26] R Core Team. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: <http://www.R-project>.
- [27] Riney, T. J., Takagi, N., Ota, K., Uchida, Y. 2007. The intermediate degree of VOT in Japanese initial voiceless stops. *Journal of Phonetics*, 35(3), 439–443.
- [28] Ringbom, H. 1987. *The role of the first language in foreign language learning*. Clevedon, MA: Multilingual Matters.
- [29] Shimizu, K. 1996. *A cross-language study of voicing contrasts of stop consonants in six Asian languages*. Tokyo: Seibido.
- [30] Simonet, M. 2014. Phonetic consequences of dynamic cross-linguistic interference in proficient bilinguals. *Journal of Phonetics*, 43, 26–37.
- [31] Sundara, M., Polka, L., Baum, S. 2006. Production of coronal stops by simultaneous bilingual adults. *Bilingualism: Language and Cognition*, 9, 97–114.
- [32] Wrembel, D. 2015. Cross-linguistic influence in second vs. third language acquisition of phonology. In: Gut, U., Fuchs, R., Wunder, E.M. (eds), *Universal or diverse paths to English phonology*. Berlin: Mouton de Gruyter, 41–70.