On Cross-linguistic Natural Speech of English Initial Voiceless Stops

Lanping Zhang [1, a], Zili Chen [2, b], Hongyan Wang [3, c], Xiaobo Li [3, d]

[1] College of Inernational Studies, Shenzhen University 3688 Nanhai Ave, Nanshan, Shenzhen, Guangdong518060, P.R. China

[2] Division of Languages & Communication, College of Professional and Continuing Education, the HongKong Polytechnic University, Hong Kong

[3] College of Inernational Studies & the Europe Research Centre of Shenzhen University，3688 NanhaiAve, Nanshan, Shenzhen, Guangdong 518060, P.R. China

[a] 1377869063@qq.com, [b] 793542789@qq.com, [c] val.chen@cpce-polyu.edu.hk,

[d] xiaoboyx@aliyun.com

Abstract. Detailed phonological features in cross-linguistic study on foreign accented natural speech is important in voice identification, especially in international accented speech identification, global foreign accented identification (GFA). The voice onset time (VOT) of stop consonants are usually measured and distinguished in the speaker's language background. In this study, we take the English initial voiceless stops production of /p,t,k/ by Mandarin Chinese, American English, Dutch, and Japanese speakers. We explore the role indication of VOT in accent measurement and try to investigate the correlation between GFA and VOT. Our results suggest that non-native speakers produced Englishwith higher or lower VOT values than their equivalents in speakers' native languages to approach the target language norm in varying degrees; particularly, Chinese speakers produced the most similar VOT to the target language norm; and interestingly there is a significant correlation between the GFA and VOT.

Keywords: English initial voiceless stops; global foreign accent; voice onset time.

1. Introduction

L2 speakers unavoidably bear their native language characteristics which can be recognized apparently as the cue of foreign accents in different phonetic features. Since the Voice Onset Time (VOT) of stop consonants is always there in producing native-like voiceless stops [1,2], researchers illustrate the norms range between most L2 speakers VOT values in their native language and target language [3,4,5,6,7,8].

According to Major [2], who employed the measure of VOT values and global L2 English voiceless stop production and explored the relation between VOT and global foreign accent (GFA), there is a highly significant correlation between foreign accent and VOT. The more the accent becomes native-like, the more the VOT values approach target language. Following this research, many studies examined the relationship between foreign accent and VOT. Flege &Eefting [9] found Dutch English Foreign Language (DEFL) speakers' L2 proficiency was correlated with the alveolar voiceless stop /t/ VOT. Riney & Takagi [10] investigated the correlation between the GFA and VOT of /p/, /t/, and /k/ among Japanese English Foreign Language (JEFL) speakers. The findings are in agreement with Major [2] and Flege &Eefting [9]. These significant correlations contribute to natural speech identification by picking up only few phonemes to confirm the accented linguistic background.

In present study, we investigate the production of English initial voiceless stops: /p/, /t/ and /k/ among Mandarin Chinese, American English, Dutch, and Japanese speakers. Two experiments and analyses are conducted: In the first one we measure the VOT values of the voiceless stops /p, t, k/ in an isolated word which is inserted in the carrier phrase; and in the second one we assess the dgerees of foreign accent of the speakers.

Major [2], Flege &Eefting [9], and Riney& Takagi [10] investigated the correlation between GFA and VOT among Brazilian Portuguese, Dutch, and Japanese speakers, respectively. They found there was a positive correlation between GFA and VOT. Speakers with high GFA scores (i.e., little or no foreign accent) produced more native-like stops (close VOT values). Although these studies concluded the significant correlation between foreign accent and VOT of voiceless stops, they only compared one foreign accent/language with the target language (English stops). It is worth investigating speakers in variable languages with a wide range of VOT values of stops (e.g., short, intermediate, and long lags), comparing their L2 VOT values, and examining the effect of foreign accent on VOT. The voiceless stops (/p/, /t/ and /k/) in Dutch and Japanese are unaspirated and in a short or intermediate lag, while the voiceless stops in Mandarin and American English are aspirated and in a long lag [9,11,12,13]. A relatively large distance in voiceless stops VOT values is shown between Dutch and English, as well as Japanese and English, while a small distance exists between Mandarin and English. So, the author investigate the production of English voiceless stops by Mandarin Chinese and American English, Dutch, and Japanese speakers.

The measurement of GFA is that Native English listeners listen to the sentence materials by L2 speakers and native English speakers (control group) and then they rate each sentence ranging from "1" (strong foreign accent), to "9" (no foreign accent) [10,14]. However, in the previous studies, the sentence materials used in the rating included but not concentrated stop consonants in words. The criterion they graded the accent degrees is a subjective rating method, since many aspects, like context, could affect the judgments. As in present study we focuses on the relationship between foreign accent and VOT, the materials for GFA ratings are consistent with the materials in VOT measurements, to control the judgment for GFA on VOT differences.

This study investigates the following questions: i) what are the VOT differences and similarities between foreign accented speeches (Mandarin Chinese, Dutch, and Japanese) and native English speech? and ii) what is the relationship between GFA and VOT value among foreign accented speech?

This study provides a distinctive perspective for the features of VOT by different native language background in Global Foreign Accented measurement in the field of cross-linguistic study.

2. Experiments

2.1 Experiment on VOT in Foreign Accented Speech

2.1.1 Participants

8 speakers volunteered to participate in the experiment. Speakers were divided into 4 groups, Native Mandarin Chinese speakers, Native American English speakers, Native Dutch speakers, and Native Japanese speakers, 1 male and 1 female each. Particularly, native American English speakers were marked as the control group. According to our questionnaire before the experiment which collected the personal information of participants, non-native English speakers among the participants have learned English as a second language since they were in middle school. Half of them are advanced learners, referring to their high scores in English proficiency tests or language environments, and the others are relatively intermediate learners. Except for the native Dutch speakers, all the participants are currently living in Guangdong, China.

2.1.2 Materials

English voiceless stops /p/, /t/, /k/, are inserted in CVC syllable structure words. we chose four front vowels /i:/, /ɪ/, /e/, and /æ/ as the nucleus. The coda consonant is always voiced alveolar stop /d/. We found the corresponding orthographic forms for the CVC structure words through the Oxford English Dictionary. Most words are common words, to make sure speakers can pronounce the words. However, some CVC structure words cannot be found correspondingly (e.g.: /ki:d/, /tɪd/). Instead, we used pseudo-words (e.g.: keed /ki:d/). The words in the CVC structure were put into the carrier phrase "I say ___again". There were twelve sentences in total.

2.1.3 Procedure

Due to the distance and other requirements for subjects, the test was launched online. First, participants were required to fill in the questionnaire about their personal information. Then, participants were given two minutes to prepare. Afterward, participants orderly read the given phrases presented in the PowerPoint. Participants read the phrase twice not too fast and not too slow, being urged to ignore the meaning of the phrases. This procedure made sure that 192 tokens (12*2*8) were collected. The recordings were edited and analyzed via Praat for VOT values (in ms). Based on the definition of VOT, the interval from the deblocking burst to the onset of vibration of the subsequent vowel vocal folds, the VOT boundaries were marked manually, presenting the VOT value. We used SPSS to obtain the mean VOT value, standard deviations (SD), and graphics.

2.1.4 Results

The results are shown in Table 1. It presents the VOT values by 2 Native Mandarin Chinese speakers (NMCSs), 2 Native American English speakers (NAESs), 2 Native Dutch speakers (NDSs), and 2 Native Japanese speakers (NJSs).

Table 1. Statistic data: mean VOT values, Standard deviation, and VOT range, for English initial stops produced by Native Mandarin Chinese Speakers (NMCSs), Native American English Speakers (NAESs), Native Dutch Speakers (NDSs), and Native Japanese Speakers (NJSs).

	NMCSs VOT in ms (SD), Range	NAESs VOT in ms (SD), Range	NDSs VOT in ms (SD) Range	NJSs VOT in ms (SD) Range
/p/	90(7.5), 64-118	83(15.9), 52-106	21(6.2), 13-44	64(29.6), 23-112
/t/	99(20.5),77-118	98(28.1), 53-154	42(10.0), 20-69	68(14.5), 41-110
/k/	107(9.7), 85-175	98(21.3), 57-135	51(15.3), 23-87	86(21.3), 54-137
M	99	93	38	73

The mean VOT value for English initial voiceless stops produced by Chinese speakers is the highest at 99 ms; while the mean VOT value by Dutch speakers is the lowest at 38 ms. As for the range of VOT values, stop /p/ (SD= 29.6) produced by Japanese speakers presents more variation than others, with VOT values ranging from 23 ms to 112 ms. Additionally, the VOT values for voiceless velar stops /k/ produced by all four groups are higher than the values for voiceless bilabial stops /p/ and alveolar stops /t/. This supports the one significant relation between the place of articulation and voice onset time proposed by Cho & Ladefoged [15]. The further back the place of articulation is, the longer the VOT value is.

2.1.5 Discussion

There are significant differences among the Dutch voiceless stops VOT, Japanese voiceless stops VOT and the English voiceless stops VOT, as the voiceless stops in Dutch, Japanese, and English are in a short, intermediate, and long lag, respectively; whereas there are slight differences between English voiceless stops VOT and Mandarin Chinese voiceless stops VOT, as they are in long lags. It is worth investigating how L1 VOT patterns with slight differences or great differences to NL affect L2 VOT production.

Figure 2 compares the English voiceless stops VOT values by L2 speakers (Chinese, Dutch, and Japanese) with the VOT values of voiceless stops in their native language; as well as the VOT values of English voiceless stops by native speakers. L2 speakers produced English voiceless stops with higher or lower VOT values than their equivalents in speakers' native languages. Chinese speakers produced slightly lower values than their equivalents in Chinese (i.e., Chinese voiceless stops), reaching the most similar VOT values to the TL norm. Although Dutch and Japanese speakers produced moderately higher VOT values than their equivalents in Dutch and Japanese (i.e., Dutch voiceless stops, Japanese voiceless stops), there still were obvious differences between their English voiceless stops production and native speakers' production, especially in Dutch speakers. It shows that the phonological similarities between L1 and L2 patterns make L2 acquisition easier, whereas the more significant the phonological differences, the more difficult it is for L2 learners to acquire the TL. This supports the Contrastive Analysis Hypothesis by Lado [16]. The L1 transfer makes a difference in the L2 VOT pattern. However, the Speech Learning Model of Flege [17] claims that the phonological similarities between L1 and L2 determine the L2 phonetics learning. In the present study, the L2 voiceless stops produced by Dutch and Japanese speakers were in relatively intermediate and long lag, respectively, which shows great distance to their L1 voiceless stop patterns both in short lag. English voiceless stops could be viewed as a "new" sound in Dutch and Japanese speakers. They have to establish a new phonetic pattern for L2 sounds. So, it is difficult for Dutch and Japanese speakers to L2 sounds at the beginning. On the other hand, Chinese voiceless stops and English voiceless stops are both in long lags with slight differences. English sounds are viewed as "similar" sounds by Chinese speakers. Generally, it is difficult for L2 learners to study "similar" sounds as the L1 similar sound would interfere with the L2 sound acquisition. However, the values of L2 English VOT by Chinese speakers were close to TL norms in this study. Do these results illustrate that these Chinese speakers have acquired the English voiceless stops production or just regarded stops in TL as the same as in L1, producing TL VOT in the L1 pattern? Additionally, the deviation from the TL norm is interpreted as a foreign accent, which shows speakers' proficiency in foreign languages. Whether this deviation in VOT values can be used to determine the accent speakers? Thus, it is interesting to investigate the foreign accent of the L2 VOT pattern and the correlation between foreign accent and VOT.

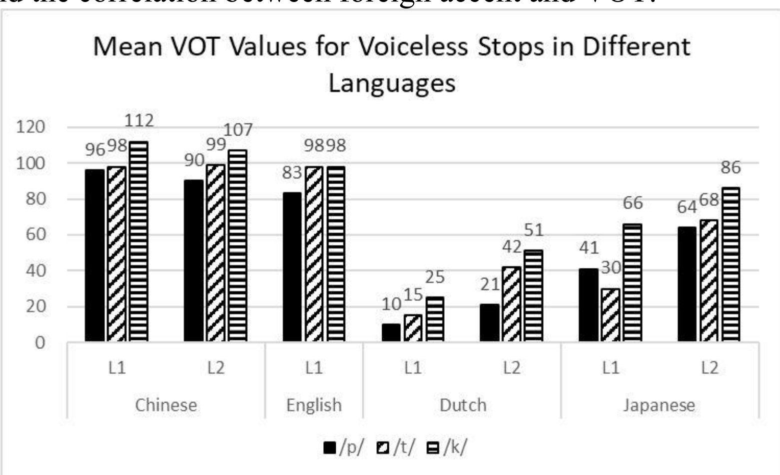

Fig. 1 Mean VOT values of Chinese voiceless stops (Chinese L1); English voiceless stops produced by Chinese speakers (Chinese L2); English voiceless stops (English L1); Dutch voiceless stops (Dutch L1); English voiceless stops produced by Dutch speakers (Dutch L2); Japanese voiceless stops (Japanese L1); English voiceless stops produced by Japanese speakers (Japanese L2).

2.2 Experiment on the GFA-VOT Correlation among Foreign Accented Speech

2.2.1 Participants

One native-English-speaking listener participated in the GFA rating test. The listener comes from the United States and now is living in Guangdong, China.

2.2.2 Materials

The materials were the recorded speech samples for eight speakers (2 Mandarin Chinese, 2 American English, 2 Dutch, and 2 Japanese) in the experiment in Chapter 3, for a total of 96 sentences. These recording sentences were edited via VSPLAY to ensure the integrity and formalness of listening materials.

2.2.3 Procedure

The listener was required to rate 96 sentences, like “I say pit again”, produced by 8 different speakers from the experiment in chapter 3 for the overall degree of foreign accent on a 9-point scale. The listener was told that he would hear native Americans and other foreign accented speeches, without knowing which sentences were produced by native Americans. The sentences were digitalized and randomized. Each sentence was presented twice in a separate block. There was a 30-second pause for the listener to rate each sentence among the blocks. The listener rated the accent by inputting a number ranging from “1” (i.e., sentence judged to have a strong foreign accent) to “9” (i.e., no foreign accent). The listener was urged to use intermediate numbers for sentences with an intermediate degree of foreign accent, make use of the whole scale, rate only on pronunciation, and ignore everything else, including any lexical or grammatical errors, recording, or background noise.

The ratings of each speaker’s accent were obtained. The VOT values were collected in the experiment in Chapter 3. we used EXCEL and SPSS to make the mean GFA scores for each speaker. The correlation between GFA and VOT was determined by calculating the Pearson CorrelationCoefficients.

2.2.4 Results

The Global Foreign Accent scores of eight speakers are shown in Table 2. The scores of eight speakers vary from the lowest (strong foreign accent) to the highest (no foreign accent). Native speakers received the highest scores at 8.7 on average, followed by Chinese speakers, Dutch speakers, and Japanese speakers at 7.2, 3.9, and 3.8 respectively. There are slight differences among the scores of English voiceless stops, /p/, /t/, and /k/, in each group.

Table 2. Mean Global Foreign Accent scores and range for Chinese speakers, native American English speakers, Dutch speakers, and Japanese speakers.

		Mean GFA Scores, Range		
	Chinese	English	Dutch	Japanese
/p/	7.0, 4-8	8.6, 8-9	4.1, 2-6	3.8, 2-6
/t/	7.3, 6-8	8.8, 8-9	3.5, 2-5	4.0, 1-6
/k/	7.4, 6-8	8.6, 8-9	4.1, 2-7	3.6, 2-6
M	7.2	8.7	3.9	3.8

The purpose of studying the GFA-VOT correlation is to investigate whether an L2 speaker’s accent influences an L2 speaker’s VOT production. Pearson correlation coefficients are illustrated in Table 4. Table 4 presents three correlation coefficients between GFA and VOT for /p/, /t/, and /k/ among six L2 speakers. The p values of all three correlations are bigger than 0.05, which means there is no significant correlation in GFA and VOT for /p/, /t/, and /k/.

However, the data of one Japanese speaker is worth noting. The GFA scores of this Japanese speaker were the lowest, whereas the VOT values were quite high and close to the NL norms. This does not conform to the general trend and previous findings (Riney and Takagi, 1999). The data points of this Japanese speaker could be viewed as outliers.

Table 3. Correlation coefficients between GFA and VOT for /p/, /t/, and /k/ for six speakers (2 Chinese speakers, 2 Dutch speakers, and 2 Japanese speakers).

	/p/	/t/	/k/
r value	0.462	0.654	0.518
p value	0.356	0.159	0.292

As the size of the samples was small, a change in the data of one sample could lead to vastly different results and conclusions. we analyzed the data without the outliers. The correlation coefficients between GFA and VOT for /p/, /t/, and /k/ for five speakers, two Chinese speakers, two Dutch speakers, and one Japanese speaker, are shown in Table 5. The r values for three stops are all over 0.9. The p values for /p/ and /t/ are less than 0.05, at 0.031 and 0.021 respectively, which indicates significant correlations at the 0.05 level (two-tailed). Also, the p value of /k/ is smaller than 0.01, at 0.001, which indicates a significant correlation at 0.01 level (two-tailed).

Thus, there is a significant correlation between GFA and VOT. The correlation is positive: The higher the GFA score, the longer the VOT value.

Table 4. Correlation coefficients between GFA and VOT for /p/, /t/, and /k/ for five speakers (2 Chinese speakers, 2 Dutch speakers, and 1 Japanese speaker).

	/p/	/t/	/k/
r value	0.911	0.932	0.993
p value	0.031	0.021	0.001

2.2.5 Discussion

As shown in Table 2, Native American English speakers received higher mean GFA scores than other non-native speakers. Several non-native speakers received the same score of 8 as native speakers, which shows the high proficiency in English of these speakers.

The correlation between global foreign accent and VOT value is significant and positive ($p<0.05$ for /p/ and /t/, $p<0.01$ for /k/). This implies that the higher the foreign accent scores the L2 speakers receive, the closer the L2 VOT values are to the NL norm. Especially, the correlation of /k/ is the most significant with the r value at 0.993.

Although there is a significant correlation between GFA and VOT, the VOT values vary among the speakers with similar GFA scores. As the analysis of the results of the correlation between GFA and VOT, one Japanese speaker (J2) who received the lowest accent scores produced higher VOT values in English voiceless stops. In Figure 3, the three data points in the upper left corner represent the GFA scores and VOT values of J2. The resting data points follow an upward trend and show a significant correlation. This indicates that there are limitations in accent assessment based on VOT measurement.

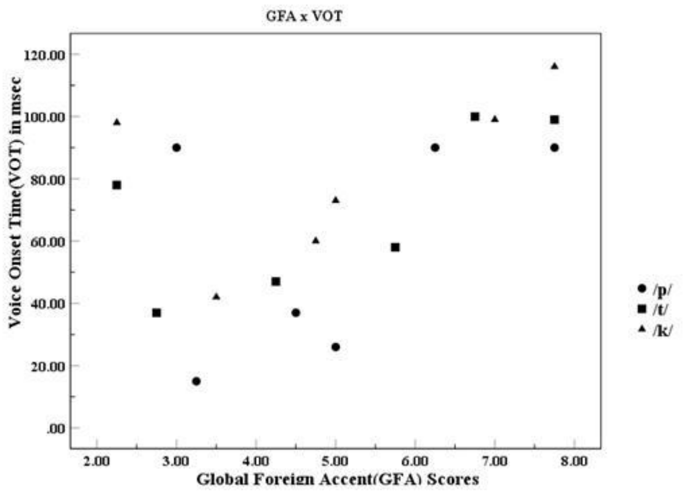

Fig. 2 GFA scores and VOT values for /p/, /t/, and /k/ by six non-native speakers.

3. Conclusion

The study demonstrated the production of English initial voiceless stops (/p/, /t/, and /k/) varied among Mandarin Chinese speakers, Native American English speakers, Dutch speakers, and Japanese speakers. L2 speakers produced English /p, t, k/ with greater or lower VOT values than their equivalents in speakers' NLs, approaching the TL norm to varied degrees. The L2 VOT by Mandarin Chinese speakers reached the closest values to the TL norm, followed by L2 VOT by Japanese speakers and Dutch speakers. These show that it is possible for L2 speakers with short-lag VOT values in their NLs to produce long-lag VOT values. Also, L2 speaker with long-lag VOT values in their NLs can produce similar long-lag VOT values. The phonological similarities between L1 and L2 patterns enhance L2 pronunciation learning, whereas significant phonological

Other factors may influence accent assessment through VOT measurement.

variances make it more difficult for L2 learners to acquire the TL.

The correlation between GFA and VOT is significant and positive. The more native-like accent L2 speakers have, the closer the VOT values are to the NL norms. Particularly, the velar voiceless stop /k/ shows the most significant correlation between global foreign accent and voice onset time. However, the VOT values of individuals are various. Foreign accent cannot be entirely judged based on VOT values. Speakers who produce native-like VOT values could be assessed as having strong accents. In the present study, one L2 speaker from Japan received low scores in GFA while producing L2 close to the NL norm, which was inconsistent with the GFA and VOT significant correlation. Even, some Chinese speakers receiving high scores in GFA produced native-like VOT with high accuracy. This supports that L2 adult speakers can acquire TL VOT with accuracy [2,10].

These findings contribute to research on how language acquisition is affected by the similarities and differences between the learner's native language and the target language. It would be an effective method for language learners to pay more attention to the similarities and differences between NL and TL. For language teaching, the difficulties of learning TL could be determined by the comparison of NL and TL. For language identification, VOT values can be a cue for language background. However, foreign accent cannot be entirely judged based on VOT values.

As for future extension, one obvious direction is to improve the diversity of materials, speakers, and listeners. Furthermore, researchers have different findings about whether language learners can produce L2 stops with equivalent or complete accuracy, e.g., [2,5,10]. In the present study, we found some L2 speakers can acquire TL VOT with high accuracy. However, this study does not further investigate how these L2 speakers process the native-like production, how common this phenomenon occurs in non-native speakers, and do these L2 speakers perceive TL sounds accurately just as they do in production. It is worth further studying these questions.

Finally, as the study focused on the production of English initial voiceless stops by Mandarin Chinese speakers, Dutch speakers, and Japanese speakers, similar studies on the GFA-VOT relation in other languages could be done. The correlation between global foreign accent and voice onset time demonstrated in this study may or may not be a universal phenomenon. The results may be different as great variation affects the VOT values and ratings of foreign accent. Therefore, research on other L2 production is significant.

Reference

[1] Caramazza, A., Yeni-Komshian, G., Zurif, E. & Carbone, E. (1973). The acquisition of a new phonological contrast: The case of stop consonants, Journal of the Acoustical Society of America, 54, 421 428.

[2] Major, R. C. (1987). English voiceless stop production by speakers of Brazilian Portuguese. Journal of Phonetics, 15, 197–202.

[3] Williams, L. (1980). Phonetic variation as a function of second-language learning. In Child phonology, vol. 2, perception (G. H.

[4] Yeni-Komshian, J. F. Kavanagh & C. A. Ferguson, editors), pp. 185- 216. New York: Academic Press.

[5] Flege, J. E. (1984). The detection of French accent by American listeners, Journal of the Acoustical Society of America, 76, 692 707.

[6] Flege, J. E. & Hillenbrand, J. (1984) Limits on phonetic accuracy in foreign language production, Journal of the Acoustical Society of America, 76(3), 708 721.

[7] Chao, K., Khattab, G., & Chen, L. (2006). Comparison of VOT Patterns in Mandarin Chinese and in English. In 4th Annual Hawaii International Conference on Arts and Humanities, pp. 840-859.

[8] Chen, L., Chao, K., & Peng, J. (2007). VOT productions of word-initial stops in Mandarin and English: A cross-language study.

[9] In: Proceedings of the 19th Conference on Computational Linguistics and Speech Processing (ROCLING).

[10] Wang, H., Xu, Y., Chen, L., & van Heuven, V. J. (2018). Native language Identification from English noise bursts by Chinese listeners. In International Conference on Intelligent and Interactive Systems and Applications (pp. 454-461). Springer, Cham. doi: 10.1007/978-3-030-02804-6_60.

[11] Flege, J. E., &Eefting, W. (1987a). Cross-language switching in stop consonant perception and production by Dutch speakers of English. Speech Communication, 6, 185–202.

[12] Riney, T.J., & Takagi, N. (1999). Global foreign accent and voice onset time among Japanese EFL speakers. Language Learning, 49, 275-302.

[13] Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. Word, 20, 384–422.

[14] Rochet, B.L., & Fei, Y. (1991). Effect of consonant and vowel context on Mandarin Chinese VOT: production and perception.

[15] Canadian Acoustics, 19, 105-106.

[16] Shimizu, K. (1990). Cross-language study of voicing contrasts of stop consonants in Asian languages. The University of Edinburgh.

[17] Munro, M.J., & Derwing, T.M. (1995). Foreign Accent, Comprehensibility, and Intelligibility in the Speech of Second Language Learners. Language Learning, 45, 73-97.

[18] Cho, T., & Ladefoged, P. (1999). Variation and universals in VOT: evidence from 18 languages. Journal of Phonetics, 27, 207229.

[19] Lado, R. (1957). Linguistics across culture. Applied linguistics for teachers. Ann Arbor: University of Michigan Press.

[20] Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. Cross-language speech perception, 233272.