Linguistic Portfolios

Volume 8 Article 7

2019

An Acoustic Phonetic Account of VOT in Russian-Accented English

Mikhail Zaikovskii
St Cloud State University, mzaikovskii@stcloudstate.edu

Ettien Koffi
St. Cloud State University, enkoffi@stcloudstate.edu

Linguistic Portfolios–ISSN 2472-5102 –Volume 8, 2019 | 74

AN ACOUSTIC PHONETIC ACCOUNT OF VOT IN RUSSIAN-ACCENTED ENGLISH

MIKHAIL ZAIKOVSKII AND ETTIEN KOFFI[1]

ABSTRACT

Russian is known as a true voice language with pre-voicing of voiced stops and no aspiration. It has been noted in many linguistic studies that the native language plays an important role in second language acquisition. There is a study which shows that people continue using their L1 processing strategies of linguistic information to communicate in their second language (Culter and Norris, 1988; Koda, 1997). In this research, we are interested in seeing whether or not Russian speakers prevoice voiced stops when speaking English.

1.0 Introduction

Maddieson and Ladefoged claims that stop segments are unique in that they are found in all languages (as cited in Koffi, 2016, p. 146). Linguists use various terms such as stop and plosive to name these sounds. Each label highlights a specific feature. "Stop" indicates that the air coming from lungs is stopped, whereas "plosive" indicates that a released burst is heard when a sound is produced. The purpose of this paper is to analyze the production of English stops by native Russian speakers. In particular, we are interested in the measurements of Voice Onset Time (VOT). We will measure the VOTs of [p, t, k, b, d, g] produced by Russian speakers and contrast them with existing measurements for these segments in General American English (GAE).

2.0 Stop consonants

There are six segments which are considered to be stops, namely [p, b, t, d, k, g]. According to Koffi (2016), they are classified as this group of sounds because air which is coming from the lungs is "stopped, blocked, or obstructed in one fashion or another when they are produced" (p. 146). Linguists divide stop sounds into three groups, based on their place of articulation and voicing features. These types are listed in Table 1.

Type	Voiced	Voiceless
Alveolar	d	t
Velar	g	k
Bilabial	b	p

Table 1: Stop consonants in English

Vocal folds vibrate when voiced segments are produced but voiceless stops are pronounced without any vibration. As can be seen in Figure 1, there are three phases that are involved in the production of stops. They are the "Close", "Stop", and "Release" phases as illustrated by Figure 1:

[1] **Authorship responsibilities:** This paper was originally written in the second author's acoustic phonetics class. He encouraged the first author and Sofia Logvineko to investigate stops in Russian-accented English for their final course project because both are native speakers of Russian. He encouraged both students to develop their term paper further for publication. Sofia graduated and moved on. The first author presented a preliminary version of this paper at the 10th annual Pronunciation Second Language Learning and Teaching (PSLLT) conference at Iowa State University in September 2018. He thereafter submitted a preliminary version to be considered for publication. The second author has thoroughly rewritten several aspects of it, and re-analyzed some aspects of the data for this publication. To the extent that the original measurements are accurate, the second author assumes full responsibility for any interpretive errors.

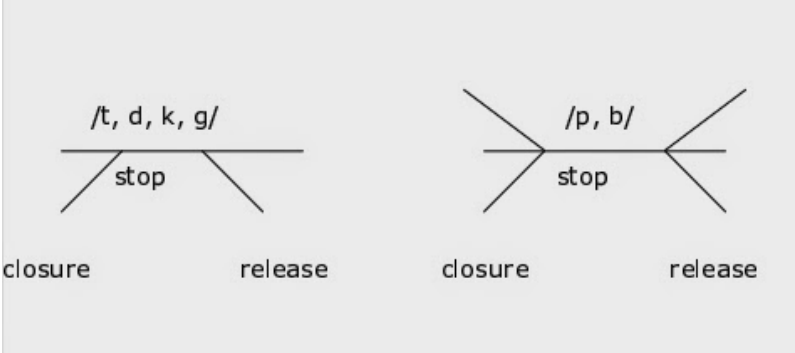

Figure 1: Articulation of stop segments (Linares, 2014)

There are also measurable points at which hearers can distinguish among stop segments. These points are called "category boundaries." A duration of ≥25 ms is necessary to distinguish between [p] and [b]; a duration of between 34 and 39 ms helps to differentiate [t] from [d], while a duration of ≥ 42 ms discriminates between [k] from [g] (as cited in Koffi, 2016, p. 156). Additionally, the percentage of voicing known as the 40/60 threshold is important role for determining whether a stop is fully voiced, devoiced, or voiceless. According to Gradoville (as cited in Koffi, 2016, p. 158), if 40% or more of a segment is voiced, then the whole segment is perceived as voiced by hearers. However, if less than 40% of a segment is voiced, the whole segment is heard as devoiced or voiceless. If less than 10% is voiced, the segment is voiceless.

3.0 Voice Onset Time

In general, voicing is important for discriminating among stops. Additionally, phoneticians have singled out Voice Onset Time (VOT) as a very robust correlate. Lisker and Abramson (1964) define VOT as "[The] time span between the burst of a stop consonant and onset of voicing." In other words, as stated by Koffi (2016), VOT equals the amount of time "that elapses between when two articulators come into contact and when they part away from each other" (p. 150). According to Kong et al, (2012, pp. 726-727), there are three types of VOT in world languages, as depicted in Figure 2.

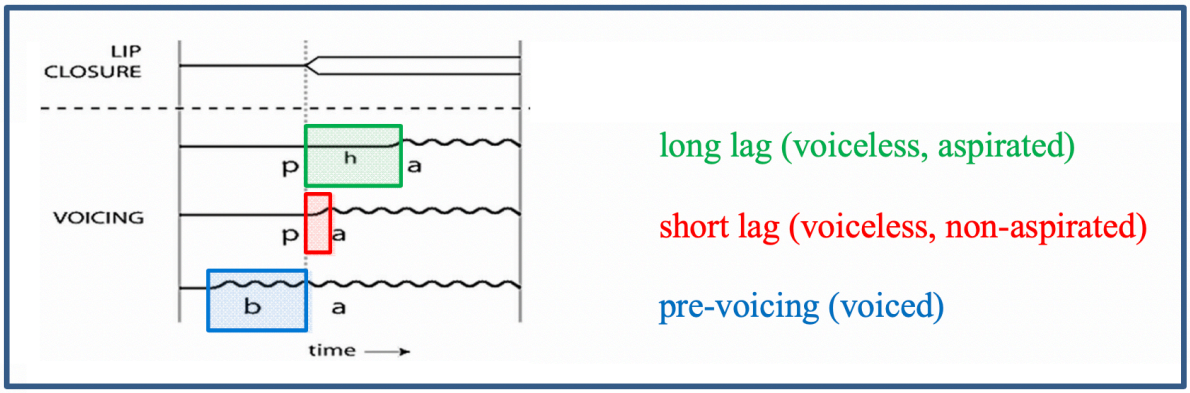

Figure 2: Three types of VOT (Color)

Lisker and Abramson (1964) observed three types of VOT in different languages. However, they noticed that the time of the beginning of voicing varies from one language to another. In order to measure VOT, first of all, they determined release of the closure is time 0. In one of the types of languages, they found voicing began before the stop closure. They concluded that in this case, VOT is negative since voicing started before the release of a stop. Stop segments with **negative VOT** are also called prevoiced stops. In some other cases, the length of VOT is measured and it turns out to be a small positive number. These stops are said

to have a **short-lag VOT**, or they are called **unaspirated stops**. Finally, in the third type of voicing, it was found that a rather long interval of time occurs from the moment of closure until the release. These are **long-lag VOTs** or **aspirated stops**. Dittmers et al. (2017, p. 4) illustrate these categories of VOT in Figure 3. They also list some languages that belong to these categories:

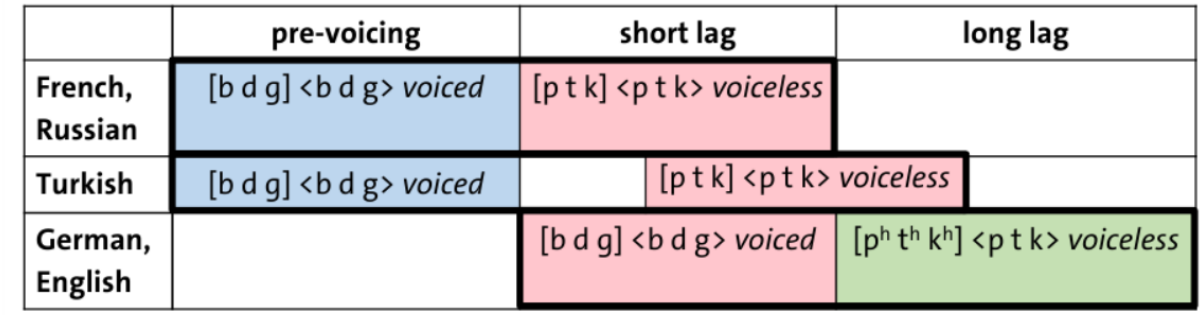

Figure 3: Voicing in different languages (Color)

As can be seen from Figure 3, French, Russian, and Turkish have pre-voicing or negative VOT but English and German do not. Yet, they have long-lag VOTs for voiceless stops, while French and Russian do not.

4.0 VOT of Russian Speakers

Ringen and Kulikov (2012), while studying VOT of Russian consonants, encountered a problem with phonetic terminology. Instead of using the terms "voiced" and "voiceless," the authors switched to "lenis" and "fortis," accordingly. They made such decision because oftentimes it is difficult or even impossible to find out the exact phonetic nature of a stop under investigation. For instance, sometimes "voiced" English segments do not have prevoicing in word-initial position, whereas Russian counterparts do. Moreover, "voiceless" stops in Russian are not aspirated, whereas in German they are (p. 270).

In their study, Ringen and Kulikov (2012, p. 270) found out that "Russian has a two-way laryngeal contrast, and it is classified as a true-voice language." French, Hungarian, and Spanish are also classified as true-voice languages. As a rule, voiced stops in these languages are prevoiced, which means that negative VOT can be observed. In addition, voiceless stops are not accompanied by aspiration. These observations were confirmed by Ringen and Kulikov (2012, p. 278), where they analyzed VOT of fourteen monolingual speakers of Russian. The findings of this study are reported in Table 2.

	Bilabial	Dental/Alveolar	Velar
Fortis (Tensed)	18	20	38
Lenis (Lax)	-70	-75	-78

Table 2: The findings of Ringen and Kulikov (2012)

They found that over 97% of the initial lenis stops were prevoiced by their Russian-speaking participants. These results are similar to the findings in Hungarian in which Gósy & Ringen (2009) found that initial voiced stops in Hungarian were prevoiced 100% of the time.

5.0 VOT in L2 Acquisition

Some studies suggest that if speakers are bilingual, their voicing might be changed due to the influence of the second language. For instance, Nagy and Kochetov (2013) conducted a

cross-linguistic study of VOT. They measured it in three generations of bilinguals with Italian, Ukrainian, or Russian Heritage Languages. Their findings indicate that VOT of Russian speakers move from short-lag towards long-lag. Sancier & Fowler (1997) found that the length of residence in the United States was influential in the VOT of the speakers of Brazilian Portuguese after an extended stay in the country. They produced longer VOTs. With this in mind, we now turn to the VOT of Russian-accented English. We seek to provide an answer to the following questions:

1) Since Russian is a true voice language, do Russian speakers of English transfer the voicing patterns of their L1 into their pronunciation of stop segments in English?
2) Or do they acquire the VOT characteristics of English when learning English?

The answers to these questions are provided by our acoustic phonetic analyses of the recordings of ten native speakers of Russian obtained from http://accent.gmu.edu/. There are 10 participants (five males and five females). They range in age from 20 to 68 years old. At the time of recording, they had been living in the USA for various amounts of time. Their mean length of residence is 9.1 years. Additional details about the participants are found in Table 3.

ID of participants	Age	AOG[2]	LOR	Other languages
Speaker 1M	37	36	1	None
Speaker 2M	66	12	27	None
Speaker 3M	54	13	15	French, German
Speaker 4M	33	18	6	None
Speaker 5M	23	13	1	German
Speaker 6F	68	38	9	Albanian
Speaker 7F	25	15	0.3	Hebrew
Speaker 8F	33	19	8.5	German
Speaker 9F	20	5	0.5	French, Polish, German, Japanese, Welsh
Speaker 10F	23	17	5	None
Mean	18.6	9.1	7.33	

Table 3: Participants' Profile

Two acronyms used in the the table require an explanation. "AOG" stands for Age of Acquisition, while "LOR" means Length of Residency. All these pieces of information and many more are found at the Speech Accent Archieve website.

5.1 Methodology and Materials

Ten target words were selected for the study. They all occur in the George Mason University's Speech Accent Archieve text. The elicitation paragraph reads as follows:

> *Please **call** Stella. Ask her to bring these things with her from the store: Six spoons of fresh snow **peas**, five thick slabs of blue cheese, and maybe a snack for her brother Bob. We also need a small plastic snake and a big **toy** frog for the kids. She can scoop these things into three red **bags**, and we will **go** meet her Wednes**day** at the train station.*

The lexical items whose VOTs were analyzed are the following:

[2] "AOG" stands for Age of Onset, and "LOR" means Length of Residency. Both are measured in years. The suffix "F" stands for "Female" while M is for "Men."

1) Voiceless stops: <peas, call, toy>
2) Voiced stops: <bag, Wednesday, go>

The stop segment of each word was annotated as shown in Figure 4.

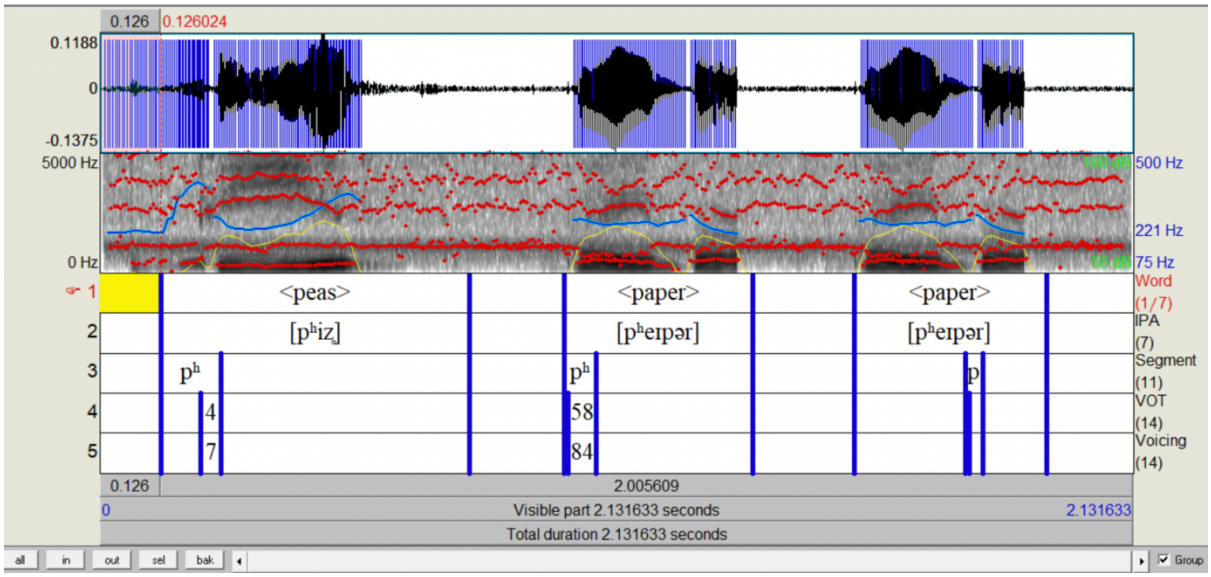

Figure 4: Annotation Procedure (Color)

All in all, the participants produced 60 tokens (6 stops x 10 participants). The tokens include 30 voiceless and 30 voiced stops. The segmentation, annotations, and measurements were done manually using Praat (Boersma & Weenink 2018).

6.0 Results

Table 4 displays the average VOT measurements of voiceless stops produced by the Russian speakers. Their measurements are compared and contrasted with the VOTs produced by GAE speakers in running speech.

ID of participants	VOT in ms		
	peas	**t**oys	**c**all
Speaker 1M	19	17	28
Speaker 2M	22	22	25
Speaker 3M	70	112	53
Speaker 4M	43	59	89
Speaker 5M	24	44	42
Speaker 6F	15	31	20
Speaker 7F	9	65	19
Speaker 8F	19	47	43
Speaker 9F	20	81	40
Speaker 10F	51	81	53
Mean	**29**	**56**	**41**

Table 4: VOTs of Voiceless Stops

We can see from Table 4 that Russian speakers' VOT for [p] and [k] are perceptually identical with the one produced by GAE speakers according to Abramson and Lisker (1964). We notice, however, that the VOT of [t] in Russian-accented English is longer than that of GAE speakers by as much as 17 ms. The bar graphs in Figure 5 show this very clearly.

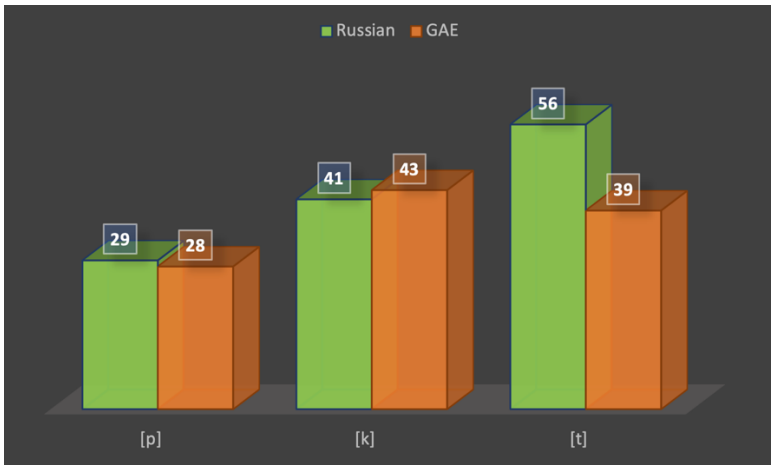

Figure 5: VOT of voiceless stops (Color)

The measurements of voiced consonants are reported in Table 5:

ID of participants	VOT (ms)		
	bags	Wednes**d**ay	**g**o
Speaker 1M	-58	12	12
Speaker 2M	-73	-82	-82
Speaker 3M	17	-104	-104
Speaker 4M	-76	-42	-42
Speaker 5M	-75	-24	-24
Speaker 6F	-93	-82	-82
Speaker 7F	-77	-48	-48
Speaker 8F	-148	12	12
Speaker 9F	-34	19	19
Speaker 10F	-76	-61	-61
Mean	**-69**	**-40**	**-40**

Table 5: VOTs of Voiced Stops

As we can see from Table 5, Russian speakers' VOT for all the voiced segments differ significantly from those of GAE. This is also reflected in the bar graphs in Figure 6:

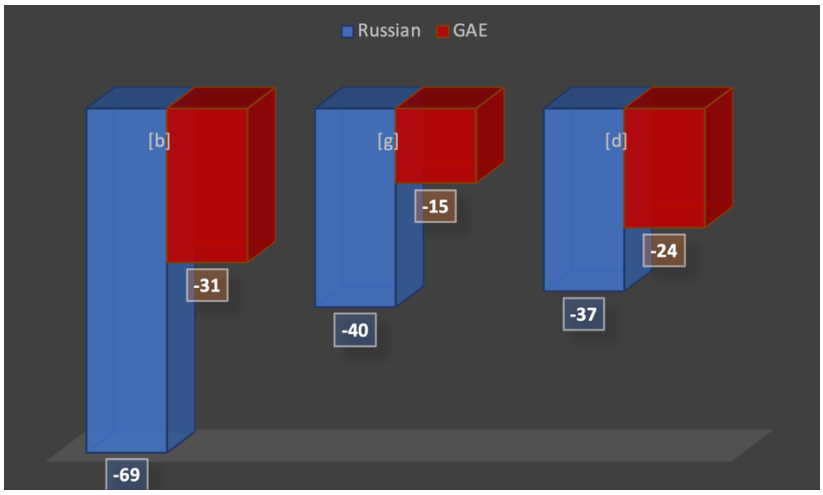

Figure 6: VOT of voiced stops (Color)

7.0 Discussions

Our results confirm the findings of Ringen and Kulikov (2012) regarding negative VOT. We see that for voiced stops, Russian speakers produced considerably longer VOTs than their GAE counterparts. We also came across something that we did not anticipate with regard to the VOT of voiceless stops. We thought that Russian speakers of English would transfer the short-lag VOT from their native language into English; but this is not what we see. Instead, we see that their VOT of [p] and [k] are as long as those of the speakers of GAE. We also see that the VOT of their [t] is longer than the one in English. These results appear to confirm the view that L2 speakers can acquire the VOTs of their L2, as evidenced by the findings of studies mentioned in 5.0.

This finding has implications that go far beyond language acquisition. It can be used for speaker identification or verification. Voiceless stops cannot be relied on to differentiate Russian L2 speakers of English from native speakers of American English. However, the length of negative VOT in voiced stops and its pervasines among Russian L2 speakers of English are helpful discriminatory acoustic cues. In Lisker and Abramson (164, p. 395) one in four Americans prevoiced [b, d, g]. In other words, prevoicing voiced stops is not as widespread among native speakers of American English as it is among Russian L2 speakers of English, or broadly speaking, among Slavic L2 speakers of English. For this, see Koffi and Abat (2014) for a discussion of negative VOT in Montenegrin-accented English. In fact, prevoicing of voiced stops is so uncommon among native speakers of American English that Wolf (1972, p. 2051) sees it as an "efficient acoustic parameter" for speaker recognition.

8.0 Summary

The results of this study are preliminary, but they show that the participants' production of stop consonants is fully intelligible. Furthermore, they reveal that these Russians speakers of English produce voiceless [p] and [k] similarly to GAE speakers. Their [t]s are more strongly aspirated than the ones produced by GAE speakers. Their voiced stops indicate clearly that they transfer the negative VOTs of their native Russian into English. They produce 90% of their voiced consonants with pre-voicing. The only three cases where voiced consonant was produced with a positive VOT is [g] in <go> as produced by Females 3 and 4, and Male 1. The results presented in this paper are preliminary because no attempt was made to correlate level of proficiency with VOT. Also, we did not attempt to correlate length of residency with VOT. Future studies may consider such correlations. It would also be useful to increase the number of participants and tokens. Even so, the results discussed here point in the same

direction as others that non-native speakers, in this case Russian, can acquire the VOT characteristics of voiceless stops in English. The particular contribution of this study is that the Russian speakers in this study overwhelmingly transfer the pre-voicing features of the native language into English.

ABOUT THE AUTHOR

Mikhail Zaikovskii is an M.A. TESL student and EAP (English for Academic Purposes) and ESL (English as a Second Language) instructor at Saint Cloud State University (SCSU). He holds an M.A. in Finance from Russian State Social University (Moscow, Russia) and is about to complete his M.A. in TESL/Applied Linguistics. He has been awarded the Interdisciplinary Enrichment Fellowship to continue his research on language and sexuality while pursuing a Ph.D. in Applied Linguistics at Arizona State University, Tempe, AZ. He can be reached at mzaikovskii@asu.edu.

Ettien Koffi, Ph.D. in linguistics from Indiana University, teaches linguistics at Saint Cloud State University, MN. Author of many peer-reviewed articles on various topics in linguistics and of four books: *Language and Society in Biblical Times* (1996), *Paradigm Shift in Language Planning and Policy: Game Theoretic Solutions* (2012), *Applied English Syntax* (2010, 2015), and the *New Testament in Anyi Morofu* (2017), a task which took over 25 years. Specializing in acoustic phonetics, dialect variation, and emergent orthographies, his current research centers on speech acoustics of L2 English (within the Speech Intelligibility Framework), Central Minnesota English, and Anyi. He can be reached at enkoffi@stcloudstate.edu.

References

Boersma, Paul & Weenink, David (2018). *Praat: Doing phonetics by computer* [Computer program]. Version 6.0.39. Retrieved from http://www.praat.org/

Cutler, A., & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human perception and performance, 14*(1), 113-121.

Dittmers, T., Gabriel, C., Krause, M., & Topal, S. (2017). *Positive transfer from the heritage language? The case of VOT in German/Turkish and German/Russian learners of L3 French, Russian, and English*. Universität Hamburg

Gabriel, C., & Thiele, S. (2017). Learning and teaching of foreign language pronunciation in multilingual settings. A questionnaire study with teachers of English, French, Italian and Spanish. *Migration, Mehrsprachigkeit und Inklusion. Strategien für den schulischen Unterricht und die Hochschullehre, Stuttgart, ibidem*, 79-104.

Gósy, M., & Ringen, C. (2009). Everything you always wanted to know about voicing in Hungarian. Paper presented at the *International Conference on the Structure of Hungarian* (Vol. 9), Debrecen, Hungary

Koda, K., 1997. Orthographic knowledge in L2 lexical processing. In: Coady, J., Huckin, T. (Eds.), *Second language vocabulary acquisition*. (35-52). United Kingdom: Cambridge University Press.

Koffi, E. (2016). *Relevant Acoustic Phonetics of L2 English: Focus on Intelligibility*. Course Manuscript: St. Cloud State University.

Koffi, E., & Abat, M. (2014). Negative VOT in three Montenegrin-accented English idiolects. *Linguistic Portfolios* 4:91-97

Kong, E. J., Beckman, M. E., & Edwards, J. (2012). Voice onset time is necessary but not always sufficient to describe acquisition of voiced stops: The cases of Greek and Japanese. *Journal of phonetics, 40* (6), 725-744.

Linares, L. (2014). Allophonic variants for English Consonants (Part 1). Retrieved from: http://easypeasyenglishlearning.blogspot.com/2014/12/allophonic-variants-for-english_24.html

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word, 20* (3), 384-422.

Nagy, N., Kochetov, A. (2013). VOT across the generations: A cross-linguistic study of contact-induced change. In Peter Siemund, Ingrid Gogolin, Monika Schulz & Julia Davydova (Eds.), *Multilingualism and language diversity in urban areas*, (19–38). Amsterdam: John Benjamins.

Ringen, C., & Kulikov, V. (2012). Voicing in Russian stops: Cross-linguistic implications. *Journal of Slavic Linguistics, 20* (2), 269-286.

Sancier, M. L., & Fowler, C. A. (1997). Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of phonetics, 25* (4), 421-436.

Wolf, J. J. (1972). Efficient acoustic parameters for speaker recognition. *Journal of the Acoustical Society of America*, 51 (6), 2044-2056.