VOT as a phonetic cue of foreign accentedness – a case study approach of five Austrian students of English

Milena Kong-Insam*
University of Graz, Department of English Studies

Abstract

Foreign accentedness in L2 speech has been shown to be a remarkably salient feature. Numerous phonetic studies demonstrated that listeners are highly sensitive to foreign-accented speech in their interlocutors (Derwing/Munro, 2009). Yet it remains to be established which phonetic cues are dominant for perceived foreign accentedness. Studies have brought different outcomes in this regard, ranging from segmental features of consonants and vowels (e.g. Saito/Trofimovich/Isaacs, 2016; Gao/Weinberger, 2018; Labov/Ash/Boberg, 2006) over different prosodic features (e.g. Mareüil/Vieru-Dimulescu, 2006). The current study focuses on the voicing contrast in bilabial and alveolar plosives in the speech of five Austrian students of English, who are investigated in the form of a case study. It will be examined how far the measured VOT values in their L2 English speech deviate from British English VOT values. VOT was measured for the voiceless and voiced bilabial and alveolar plosives /p b t d/. The plosives chosen for analysis were assumed to be especially problematic for Austrian German speakers as voicing is realized differently in English and German. Correlations between VOT and foreign accentedness will be drawn based on the students' final grades in an English pronunciation class they were all enrolled in at the time of the study. *(received: 10 May 2021; accepted: 15 June 2021)*

*   University of Graz
    Department of English Studies
    Heinrichstrasse 36/2, 8010 Graz
    milena.kong@uni-graz.at

1. Introduction and theoretical considerations

The perception and detection of foreign accentedness in L2 speech has been a highly investigated issue in phonetic research. Over the past few decades, different studies have demonstrated that listeners can easily and quickly detect a foreign accent in their interlocutor, even in short stretches of speech (Flege, 1984), in backward speech (Munro/Derwing/Burgess, 2003), and in L2s unknown to the listeners (Major, 2007). Yet the crucial question remains of what it actually is that makes our speech sound foreign and what gives us away when we speak an L2 with a foreign accent. From a linguistic standpoint, it needs to be established which phonetic cues influence foreign-accented speech. So far, studies on foreign accentedness have yielded different results. Saito, Trofimovich and Isaacs (2016) found in their study that segmental accuracy, i.e. the accurate production of vowels and consonants, played a central part in listener ratings on foreign accentedness. Gao and Weinberger (2018) concluded that consonants, in particular, are especially critical for perceived foreign accentedness, whereas Labov, Ash and Boberg (2006) stress the importance of vowels to foreign-accented speech. Mareüil and Vieru-Dimulescu (2006) highlight the relevance of prosody in listeners' differentiation between native and non-native speech. Such diverse findings make it difficult to pin down the phonetic cues of foreign accentedness and may even suggest a combination of various factors. It should also be taken into consideration that the aforementioned studies comprise different L1 backgrounds, and that this may have also influenced the varying results (see, for instance, Bongaerts, 1999).

In an attempt to shed light on the previously outlined issue, the current study investigates five cases of Austrian L2 learners of English. Due to the inconclusive results regarding which phonetic cues influence foreign accentedness the most, various segmental and prosodic aspects in the speech of these subjects will be considered. In order to account for possible L1 influence on perceived foreign accentedness as mentioned above, the learners' L1, Austrian German, and the primary phonetic and phonological differences to their L2, British English, will be identified and examined.

This first study will focus on the subjects' production of voicing in selected consonants. As different acoustic and perceptive studies have shown, Standard Austrian German (SAG) has a neutralization of the voice contrast in bilabial and alveolar plosives in word-initial position (Wiesinger, 1996; Moosmüller/Schmid/Brandstätter, 2015). This voicing neutralization results from a combination of a lack of aspiration in voiceless and a devoicing of voiced plosives word-initially (Wiesinger, 1996). It thus makes voiced and voiceless bilabial and alveolar plosives fairly identical in initial position in SAG. In Standard Southern British English (SSBE), by contrast, there is a clear voicing distinction in bilabial and alveolar plosives, which mainly results from aspiration: voiceless plosives are strongly aspirated while voiced plosives are devoiced (Deterding/Nolan, 2007).

The present study examines whether and how far the voicing neutralization in bilabial and alveolar plosives typical of SAG is reflected in the subjects' L2 English.

Such negative L2 transfer has been shown especially during the initial stages of the L2 acquisition process (Major, 2008). The five tested subjects show different pronunciation proficiency levels and also different degrees of foreign accentedness. It will be investigated whether the transfer of the voicing contrast occurs in the subjects' speech and whether concerned students achieved a lower grade in the pronunciation class.

2. Experimental design

The present study is based on VOT measurements of bilabial and alveolar voiceless and voiced plosives in the production of five Austrian students of English. A multiple-case design was chosen for the study, which will be elaborated in the following section of the paper.

2.1. Sample description

2.1.1. The cases

The cases described in the present study include five Austrian undergraduate students of English from the University of Graz. Four of the speakers were female and one of them was male. They were between 20 and 23 years of age at the time of the study. They all had Standard Austrian German (SAG) as their L1. The subjects were all enrolled in the obligatory undergraduate English language class 'Pronunciation' at the University of Graz. The test materials were compiled during the final oral exam of the course. See Table 1 for the subjects' demographic background information and their awarded grades.

Cases	Speaker ID	Sex	Age	L1	Grade
Subject 1	S1	f	20	SAG	2
Subject 2	S2	m	21	SAG	1
Subject 3	S4	f	20	SAG	2
Subject 4	S6	f	23	SAG	4
Subject 5	S7	f	20	SAG	5

Table 1. Demographic background information and awarded grades of the subjects

The subjects' official proficiency level of L2 English can be defined as B2 according to the Common European Framework of Reference for Languages (CEFR), which is the English proficiency level that students receive upon graduating from an Austrian high school (Austrian Federal Ministry of Education, Science and Research, 2018). In spite of the subjects' supposed proficiency level of B2, their actual pronunciation proficiency level, according to their pronunciation teacher, was higher in some of them. This can be attributed to additional factors such as increased daily exposure to English through friends or relatives, extended stays abroad in English speaking countries, media influence, or simply motivational factors.

The pronunciation course was taught by a trained and experienced male English pronunciation teacher, who was a native SSBE speaker. The pronunciation teacher had been living in Graz, Austria, for eleven years at the time of the study. The intended focus of the pronunciation class, according to the course description, is on the acquisition of individual speech sounds as well as on prosodic features such as stress, rhythm, and intonation in Standard British and American English (University of Graz, 2020). In class, the teacher put the main focus primarily on segmental aspects, in particular on consonants. The underlying idea is that students should master individual speech sounds first and then venture into prosodic aspects such as intonation. In line with the actually practiced focus of the class, students' performance regarding consonant and vowel production was also the main criterion used for grading in the final oral exam.

2.1.2. Norm values

The students tested in the study had a British English accent as confirmed by their pronunciation teacher. Therefore, SSBE was chosen for the norm against which the VOT values of the students were matched. The SSBE norm values were taken from Chao and Chen (2008), who measured VOT in four speakers of standard British English. Table 2 shows their mean VOT values for the bilabial and alveolar plosives.

Plosive	/p/	/b/	/t/	/d/
VOT	62	11	73	22

Table 2. Mean VOT values (in ms) measured in four SSBE speakers (Chao/Chen, 2008)

In addition, Lisker and Abramson's (1964) categorization of VOT ranges into negative, short-lag, and long-lag VOT served as a reference frame for the study (see table 3).

Negative VOT	Short-lag VOT	Long-lag VOT
-125: -75 ms	0-25 ms	60-100 ms

Table 3. Categorization of VOT ranges (Lisker/Abramson, 1964)

2.2. Recordings and grading

The recordings took place in a sound-attenuated recording booth with a Rode NT2-A Studio Solution Set microphone of a sampling rate of 44.1 kHz and the program SpeechRecorder (Draxler/Jänsch, 2004). They were conducted during the students' final oral exam of the English pronunciation class. The students were recorded reading a selected passage from the book *Interview with the Vampire* by Anne Rice (1976). They were instructed to read the passage displayed on a screen in the sound-attenuated recording booth. The selected passage was one of the six tests

which the students had practiced in class and at home. They did not know which of the six texts would be selected for the exam.

The teacher was present in the same room during each recording. As part of his grading procedure, he took detailed notes on the students' reading performance, thereby identifying problematic aspects in terms of segmental and prosodic features. After the reading, he provided thorough oral feedback of each individual student's performance. He subsequently announced the respective grade that each student received for the class. The grades were based on the students' pronunciation proficiency while reading the selected text; their proficiency was judged based on how close their pronunciation came to SSBE in terms of, first and foremost, segmental, but also prosodic aspects.

The students were graded according to the Austrian academic grading system, comprising a range of five grades (1-5), with 1 ("very good") being the highest and 5 ("unsatisfactory") being the lowest achievable grade. In order to pass the class, students have to achieve grade 4 or higher (University of Graz, n.d.).

2.3. Acoustic analysis

In the acoustic analysis, VOT was measured for the bilabial and the alveolar plosives using the program Praat (Boersma/Weenink, 2019).

2.3.1. Data extraction

The analyzed tokens were selected on the basis of consonants assumed problematic for SAG speakers (see section 1): VOT was measured in the bilabial plosives /p/ and /b/ and in the alveolar plosives /t/ and /d/. A total number of 93 tokens were analyzed in 19 words. The tokens are listed in Table 4. The total number of tokens include two missing values (see section 2.4.2 for further explanation).

BILABIAL	voiceless	p1	p2	p3	p4	X
		pocket	painful	person	painful	X
	voiced	b1	b2	b3	b4	b5
		Boy	button	boy	boy	boy
ALVEOLAR	voiceless	t1	t2	t3	t4	t5
		Took	tell	timidly	told	two
	voiced	d1	d2	d3	d4	d5
		Date	did	don't	died	does

Total number of word tokens analyzed: 93

Table 4. Plosive labels and target words selected for VOT analysis

Previous studies have shown differences in the VOT values depending on the subsequent vowels. Docherty (1992), for instance, found longer VOT values before

high vowels and shorter ones before mid and low vowels. Neutralization of this effect by means of appropriate word token selection could not be considered in the present study because the given passages for the exams were preselected for the pronunciation class. Slight variations between the individual measured VOT values are therefore taken into consideration in the present study.

2.3.2. VOT measurements

VOT was measured according to Lisker and Abramson (1964), i.e. from the release of the plosive (indicated through a spike in the waveform) to the onset of voicing in the following vowel (indicated through the beginning of periodicity in the soundwave) (see Figure 1).
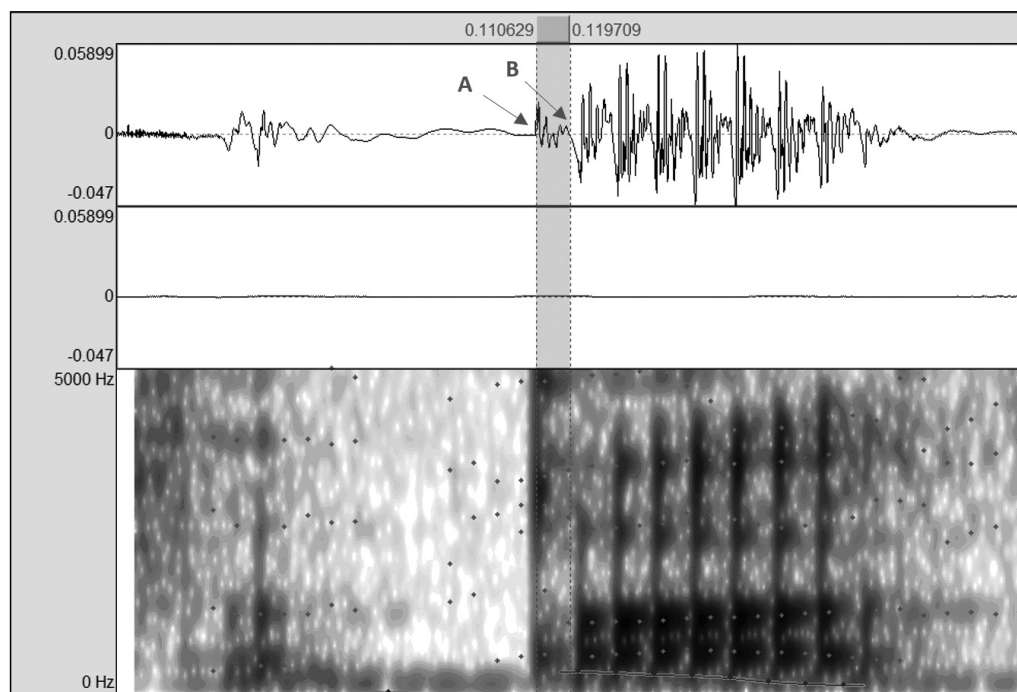

Figure 1. Example of VOT measurement in Praat, showing the oscillogram (top) and wide-band spectrogram (bottom) of the target word *button* produced by subject S2. VOT is displayed in the highlighted section and was measured from the point of release of the plosive ("A") to the onset of periodicity ("B").

In two tokens, there was no visible point of release identifiable in the respective plosive, and therefore, VOT could not be measured (see Figure 2). Auzou et al. (2000) attributed the absence of a point of release to incomplete closure in the production of a plosive, which affected up to four percent of VOTs in previous studies. The two affected tokens of the present study were thus coded as missing values and excluded.
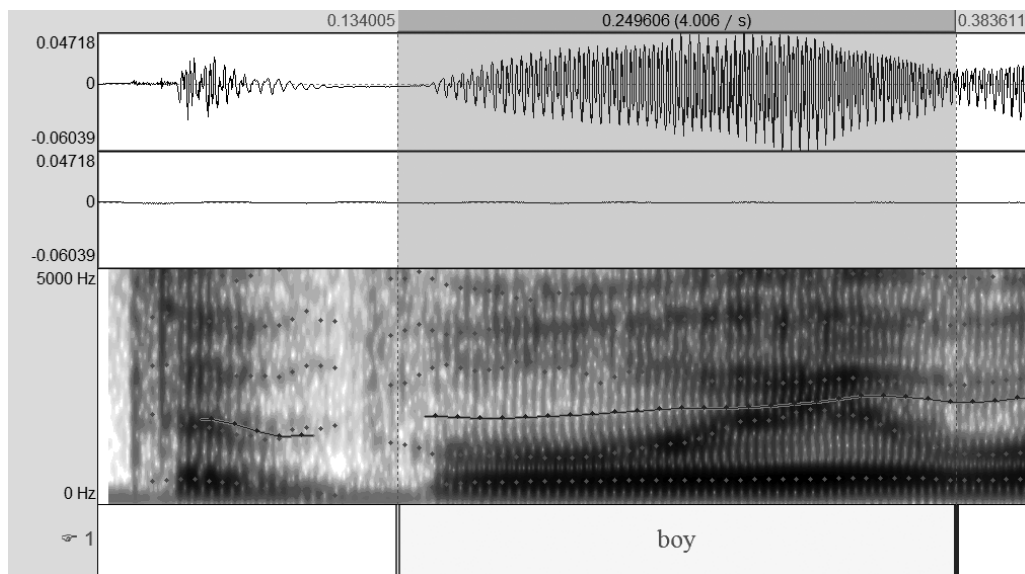
Figure 2. Representation of the oscillogram (top) and wide-band spectrogram (bottom) of the target word *boy* in Praat produced by subject S4. No point of release is visible for the plosive /b/ in the speech sample.

2.4. Research questions

The current study will focus on three major research questions, which are outlined in the following.

RQ 1: Do the individual subjects show voicing neutralization in their L2 English speech?

RQ 2: How far do the subjects' VOT values deviate from the SSBE norm values?

RQ 3: How far do the subjects' VOT values deviate from Lisker and Abramson's (1964) VOT categories?

It will further be examined whether voicing neutralization and VOT deviations influenced the students' grades, i.e. whether students with voicing neutralization and higher VOT deviations from the SSBE norm and Lisker and Abramson's (1964) categories achieved lower grades and vice versa.

3. Results and discussion

3.1. Voicing distinction

RQ 1 focused on whether the VOT values of the speakers' bilabial and alveolar plosives exhibited a clear voicing distinction. In the statistical analysis, it was found that voiceless and voiced bilabial and alveolar plosives were indeed significantly different in terms of VOT in the five speakers. A paired-sample t-test showed a

significant difference of t = 3.25, p < 0.05 between the mean VOT values of /p/ (M = 50.02, SD = 12.39) and /b/ (M = -31.61, SD = 52.07). The mean VOT values of /t/ (M = 65.52, SD = 21.09) and /d/ (M = -27.52, SD = 33.27) were also found to be significantly different with t = 4.90, p < 0.05. Overall, the speakers thus produced /p/ and /t/ with considerably longer VOTs than /b/ and /d/, respectively, and thereby clearly differentiated between voiceless and voiced plosives in their L2 pronunciation of English. The VOT ranges of the plosives are illustrated in the boxplot in Figure 3.
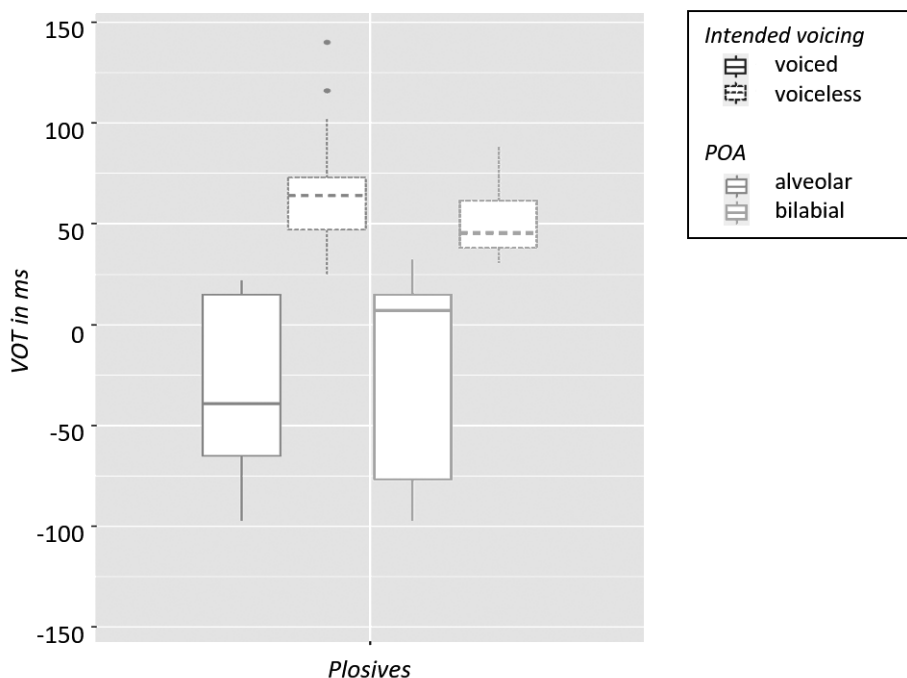

Figure 3. Boxplot comparing the VOT range of voiceless and voiced plosives in the five cases

The individual mean VOT values and the voicing distinction of each speaker were compared to the norm values and listed in Table 5. As can be seen in the table, particularly large differences between voiceless and voiced plosives were found in both bilabials and alveolars of S1 and S6. These strong differences indicate a particularly strong voicing distinction, which was much higher than that of the norm values and which mainly results from prevoicing of the voiced plosives. The voicing distinction of S2, S4, and S7 was much closer to that of the norm values, although slightly lower. Focusing on the mean values, however, it can be seen that for S7, the voicing distinction resulted from little aspiration of the voiceless plosives and prevoicing of the voiced plosives; although the voicing distinction came close to the native norm in this speaker, the actual mean values divert.

Speaker	Mean /p/	Mean /b/	Δ /p/-/b/	Mean /t/	Mean /d/	Δ /t/-/d/	Grades
S1	43.25	*-83.8*	127.05	55.4	*-47*	102.4	2
S2	54	14.8	39.2	63.8	15.8	48	1
S4	50.5	11.25	39.25	74.4	*-2.8*	77.2	2
S6	67.75	*-72.4*	140.15	95.2	*-65.6*	160.8	4
S7	34.6	*-1.5*	37.1	38.8	*-38*	76.8	5
Norm	**62**	**11**	**51**	**73**	**22**	**51**	X

Table 5. Mean VOT values and differences (Δ) between voiceless and voiced plosives in the five speakers to the norm values (indicated in ms). Negative (i.e. prevoiced) VOT values are highlighted with italics.

In three of the speakers, S2, S4, and S6, a connection could be seen between the closeness of the voicing distinction to the native norm and their awarded grades: The voicing distinction of S2 and S4, two students with the highest grades, were closest to the norm values. S6, who received a low grade, strongly exceeded the native voicing distinction due to prevoicing. S1, as well, exceeded the native voicing distinction because of prevoicing, but the speaker nevertheless achieved a high grade. S7, who received a low grade, came close to the native voicing distinction, but only because of low aspiration and prevoicing. It can thus be concluded that in S2, S4, and S6, the closeness of the voicing distinction to the native norm is indicative of their grade: the closer the voicing distinction, the higher their grade.

The statistical analysis and the comparison of the mean values of each speaker both showed that the students clearly differentiated between voiceless and voiced plosives in terms of VOT. Therefore, no evidence of neutralization which is typical of SAG was found in the speakers. This indicates that the speakers were at an already advanced learning stage where L1 influence is not as strong anymore and less negative transfer occurs (Major, 2008). Interestingly, however, as previously discussed, the voicing distinction was considerably higher than the native norm in two speakers, S1 and S6. This was especially the case because of prevoicing, as can be seen in Table 5. A possible explanation for this may be that students were made aware in class of the existence of a voicing distinction which does not exist in their native language; prevoicing may be these students' (S1 and S6) attempt at realizing the voicing distinction, which, however, resulted in a deviation from the SSBE norm.

3.2. VOT deviations

In order to investigate foreign accentedness in the speakers' L2 English speech, RQ 2 examined how far the measured VOT values of the speakers deviated from the SSBE norm values. One-sample *t*-tests and Wilcoxon signed-rank tests (when the data was not normally distributed) were conducted. The results of the statistical analysis of each speaker are listed in Table 6 and further visualized in Figure 4.

Speaker	S1	S2	S4	S6	S7
Grade	2	1	2	4	5
/p/	t = -3.53*, df = 3	t = -1.21, df = 3	t = -1.93, df = 3	t = 0.57, df = 3	t = -24.13***, df = 3
/b/	t = -14.96***, df = 4	t = 1.17, df = 4	t = 0.09, df = 3	t = -12.59***, df = 4	t = -11.78**, df = 2
/t/	t = -3.23*, df = 4	t = -1.52, df = 4	t = 0.16, df = 4	t = 1.46, df = 4	t = -8.70***, df = 4
/d/	t = -3.61*, df = 4	V = 0	t = -1.97, df = 4	V = 0*	t = -4.83**, df = 4

Table 6. Statistical results of the VOT deviations in the five speakers from the SSBE norm values
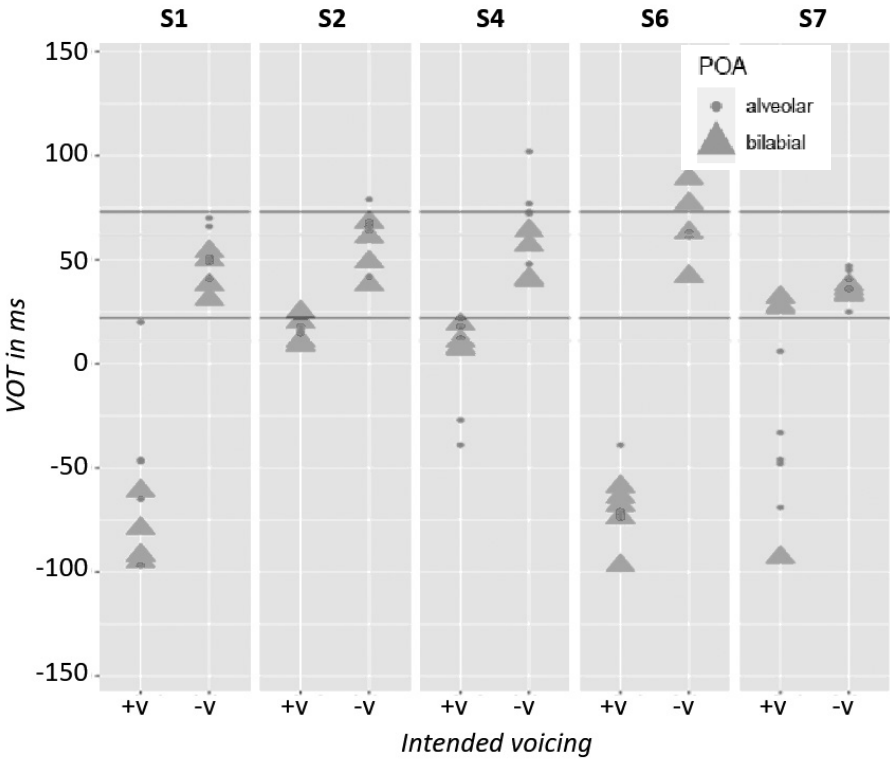

Figure 4. Visual representation of the VOT deviations in the five speakers from the SSBE norm values (v = voicing). Bilabial plosives are indicated in green, alveolar plosives in red. The lines indicate the corresponding SSBE norm values.

The statistical tests found large individual differences between the speakers in terms of deviations from the native norm, with some very close VOT values which barely deviated as well as some strongly deviating VOT values. Inter-speaker variations were also found, especially in S1 in the voiced alveolar plosives.

Strong and significant deviations were measured in S1, S6, and S7 (see Table 6). In S1, deviations were most significant in the voiced bilabial plosive /b/ (t = -14.96***, $p < 0.001$), which was prevoiced. VOT of the voiceless plosives were closer, but still significantly different to the norm values; in fact, both /p/ (t = -3.53*, $p < 0.5$) and /t/ (t = -3.23*, $p < 0.5$) had significantly lower VOT values, which means that they were produced with less aspiration. In spite of these deviations, the student received grade 2. S6 and S7, by contrast, both received low grades (grades 4 and 5, respectively). S6 showed significant differences in the voiced plosives, which were prevoiced (/b/: t = -12.59***, $p < 0.001$; /d/: V = 0*, $p < 0.5$). VOT of the voiceless plosives was considerably higher, especially in /t/, but not significantly different. This could be due to the wide range of the values, one of which was indeed very close to the norm values of the study, which may have evened out the results. S7, by contrast, who failed the class, showed significant deviations across all plosives. VOT was somewhat inconsistent in the voiced plosives, which were occasionally strongly prevoiced (/b/: t = -11.78**, $p < 0.01$; /d/: t = -4.83**, $p < 0.01$). VOT in the voiceless plosives was significantly lower than the norm values, which shows that they were not (or only slightly) aspirated (/p/: t = -24.13***, $p < 0.001$; /t/: t = -8.7***, $p < 0.001$). Interestingly, the measured values showed a voicing neutralization in the bilabials: When /b/ was not prevoiced, it was very close, almost identical to /p/, which lacked aspiration.

Two speakers, S2 and S4, came quite close to the SSBE norm values, which was reflected in their grades (grades 1 and 2, respectively). No significant deviations were found in those speakers (see Table 6). S2 was close to the SSBE norm in all VOT values, yet with slightly lower values in the voiceless plosives. S4, as well, was rather close to the SSBE norm values, with occasional minor deviations in the alveolar plosives.

RQ 2 further examined whether VOT deviations from the SSBE norm were reflected in the students' grades. For that purpose, a descriptive overview will be given of the individual grades.

Grade 1:

One student, S2, received grade 1, i.e. the highest possible grade. His VOT values were fairly close to the SSBE norm with no significant deviations (Ø 54 ms vs. 62 ms for /p/, Ø 14.8 ms vs. 11 ms for /b/, Ø 63.8 ms vs. 73 ms for /t/, and Ø 15.8 ms vs. 22 ms for /d/) (see also Table 6 and Figure 4). The grade thus clearly reflects the closeness of the VOT values to the SSBE norm.

Grade 2:

S1 and S4 received grade 2. In S1, significant deviations from the SSBE norm values were found in all plosives (see Table 6). The voiced plosives, most notably,

were highly significantly different (Ø -83.8 ms vs. 11 ms for /b/ and Ø -47 vs. 22 ms for /d/), which was due to strong prevoicing. The voiceless plosives were, in fact, closer to the norm, but still significantly lower, due to a lack of aspiration. In spite of the significant deviations across all plosives and, in particular, the high degree of prevoicing observed in S1, the student received a fairly good grade. Consequently, no such correlation between VOT accuracy and grade can be observed in S1. This shows that aspects in the student's pronunciation other than VOT influenced the high grade in this speaker.

The VOT values in S4, by contrast, came rather close to the SSBE norm values in all plosives (Ø 50.5 ms vs. 62 ms for /p/, Ø 11.25 ms vs. 11 ms for /b/, Ø 74.4 ms vs. 73 ms for /t/), with the exception of three instances of slight prevoicing in /d/ (Ø -2.8 ms vs. 22 ms for /d/). There were no significant deviations from the norm values found in S4. Except for the partial prevoicing, the closeness to the SSBE norm values is thus well reflected in the high grade of S4.

Grade 4:

S6 received grade 4. Strong prevoicing was found in the voiced plosives (Ø -72.4 ms for /b/ and Ø -65.6 ms for /d/), which deviated significantly. At the same time, the voiceless plosives were strongly aspirated and thus even slightly outscored the SSBE norm (Ø 67.75 ms vs. 62 ms for /p/ and Ø 95.2 ms vs. 73 ms for /t/), yet did not deviate significantly. S6 thus partly overachieved the VOT values for the voiceless plosives and, in fact, achieved the highest VOT values for /p/ and /t/ out of the five speakers. It seems that the student overcompensated and thus exaggerated aspiration in order to establish the voicing distinction in the tested plosives. Taking this into account, the measured VOT values thus also match the low grade of S6.

Grade 5:

S7 received grade 5 and consequently failed the class. Highly significant deviations were found in all plosives (see Table 6). VOT was inconsistent in the voiced plosives, which were occasionally prevoiced, especially in /d/ (Ø -1.5 ms for /b/ and Ø -38 ms for /d/). In addition, a lack of aspiration and, as a result, lower VOT values were found in the voiceless plosives (Ø 34.6 ms vs. 62 ms for /p/ and Ø 38.8 ms vs. 73 ms for /t/). This clearly enhanced a voicing neutralization, especially in the bilabial plosives, where the voiceless and the positive voiced VOT values were (almost) identical (see Tables 8 and 9). The VOT inaccuracies measured in the plosives of S7 as well as the observed voicing neutralization in individual VOT values due to L1 interference thus clearly reflect the student's negative grade.

Overall, it can be concluded that the grades matched the closeness of the individual VOT values measured in the tested students' speech. Those students who approached the SSBE norm indeed received higher grades than those who strongly deviated. The only exception to this observation was S1, who showed significant VOT deviations but still received grade 2. As suggested earlier, this student must have excelled in other aspects of pronunciation which influenced the grading positively.

To conclude, voicing of the plosives seemed to influence foreign accentedness in the students' speech and thus also played a central role in their final grades in the pronunciation class.

3.3. VOT categories

RQ 3 examined whether the individual VOT values measured in the speakers correspond to the VOT categories proposed by Lisker and Abramson (1964).

3.3.1. Voiceless plosives

Overall, many of the individual VOT values measured in the voiceless plosives were below the norm of 60-100 ms defined as long-lag VOT by Lisker and Abramson (1964) (see Table 3). The bilabial voiceless plosive /p/ (see Table 7), especially, was clearly lower than the SSBE norm as well as the category norm in four speakers, namely in S1, S2, S4, and S7. The lowest VOT values of /p/ were achieved by S7 with a mean of 34.6 ms. These comparatively low VOT values of /p/ show that the speakers produced the plosive with no or little aspiration. Although the values were lower, all individual VOT values exceeded 25 ms and thus did not fall into the category of short-lag VOT. Only one student, S6, came close to the SSBE norm value for /p/ with a mean of 67.75 ms and was thus within Lisker and Abramson's (1964) long-lag VOT category.

Slightly higher overall VOT values were found in the alveolar plosive /t/ (see Table 8). In two speakers, S1 and S7, the mean VOT values were below the norm. This indicates that the plosive was produced with little aspiration as well. S7 even had VOT values of 25 ms or lower, which would, accordingly, be categorized as short-lag VOT (clearly no aspiration). Interestingly, S6 overachieved the target with a noticeably higher VOT than the SSBE norm and a mean of 95.2 ms, yet was still within Lisker and Abramson's (1964) long-lag category.

The voiceless VOT values of S7, who failed the course, showed a clear lack of aspiration and were thus considerably lower than the proposed long-lag category. The other student with a low grade (grade 4), S6, by contrast, exhibited the highest individual and mean VOT values. Although they fell within the long-lag category in both plosives, the speaker exceeded the native norm and seemed to have exaggerated aspiration. S1, S2 and S4, the students with the highest grades (grades 2, 1, and 2, respectively), showed slightly lower VOT values than the proposed long-lag category in /p/ but within-average values in /t/.

3.3.2. Voiced plosives

Voiced plosives fall into the short-lag VOT category of 0-25 ms as defined by Lisker and Abramson (1964) (see Table 3). An analysis of the students' voiced plosives, however, showed a noticeable amount of prevoicing, both in /b/ and /d/ (see Tables 7 and 8). In fact, three speakers, S1, S6, and S7, showed persistent prevoicing in both plosives. S4 showed two prevoiced VOT values in /t/ but none in /p/. In those cases, negative VOT was measured instead of short-lag VOT as in the SSBE norm.

The VOT values for the voiced plosives that were not prevoiced were fairly in line with the norm values and within the range of short-lag VOT of 0-25 ms.

Two of the speakers who prevoiced, S6 and S7, received low grades (grades 4 and 5, respectively). Only S1 achieved a high grade (grade 2) in spite of prevoicing. It can be assumed that these students were trying to overcompensate in order to maintain the voicing distinction and to distinguish between voiceless and voiced plosives. S2 and S4, two students with the highest grades (grades 1 and 2, respectively), (barely) showed prevoicing.

4. Discussion

The results of the study showed that the voicing aspect of the selected plosives contributed to foreign accentedness in the speakers' L2 English speech. In fact, both closeness of the voicing distinction and individual VOT values to the SSBE norm were reflected in the students' grades, with the exception of one speaker, as will be elaborated in the following.

Voicing neutralization typical of SAG was not found in the speakers' mean VOT values (RQ 1). An examination of the individual VOT values, however, found voicing neutralization in the bilabial plosives in S7: the positive VOT values of /b/[1] and the VOT values of /p/ were almost identical due to a lack of aspiration. This voicing neutralization was not reflected in the statistical analysis of the mean values, because S7 produced a mixture of positive and negative (i.e. prevoiced) VOT values. Interestingly, S7 was the student who received a negative grade. Voicing neutralization due to a lack of aspiration may, therefore, have influenced the grade in S7. S6, by contrast, who received a low grade as well (grade 4), showed a noticeably large voicing distance. It resulted from producing unvoiced plosives with short-lag VOT (due to a lack of aspiration) and voiced plosives with negative VOT (due to prevoicing). This VOT pattern thus also deviated from the SSBE VOT pattern, although in the opposite direction compared to S7. Those students who received grades 1 and 2 (S2 and S4) showed closer proximity between voiceless and voiced plosives. The VOT pattern of these highly graded students came close to the native-like pattern exhibited in the SSBE norm (see Figure 4). The only exception was S1, whose VOT pattern resembled that of the lowly graded students but who still received a high grade. VOT in S1 was therefore not indicative of the grade, which must have been attributed to other phonetic or phonological aspects in the speaker's L2 speech. With the exception of S2, it can thus be concluded that those students who showed a more native-like voicing distinction in the bilabial and alveolar plosives generally received better grades than those who had a smaller or larger voicing distinction.

Strong deviations of the mean VOT values from the native SSBE norm were found in the students with low grades, S6 and S7. Both students (partly) exhibited prevoicing, and S7, who failed the course, additionally showed a lack of aspiration. S2 and S4, who received high grades, showed no significant deviations in their mean VOT values. S2, who received grade 1, came the closest to the SSBE VOT values. In

---

1   S7 showed both positive and negative VOT values in /b/ (see Table 7).

those four students, the deviations in VOT values from the SSBE norm were thus well reflected in their grades: The higher the VOT deviations, the lower were the grades (RQ 2). Only S1 did not match this pattern: In spite of strong deviations due to prevoicing and lack of aspiration, the student received grade 2.

Also, a connection could be seen between students' grades and deviations from Lisker and Abramson's (1964) proposed VOT categories. In the voiceless plosives, the student with grade 5 (S7) was clearly below the long-lag category due to a lack of aspiration. The student who received grade 4 (S6), however, was within the category, but showed exaggerated aspiration. The students with the highest grades (S1, S2 and S4) were somewhat below the long-lag category in /p/ but within the category in /t/. In the voiced plosive, it was found that those who received low grades (S6 and S7) showed strong prevoicing and thus fell into the negative VOT category instead of the short-lag category. The students with high grades, S2 and S4, showed little to no prevoicing. Only S1 showed prevoicing but still received a high grade.

5. Conclusion

The present study investigated the role of voicing as a determining factor of foreign accentedness in the speech of five Austrian students of English. The teacher's awarded grades in the pronunciation class served as a reference frame for foreign accentedness. Pronunciation proficiency in the students' L2 English was based on an approximation of SSBE in terms of segmental and prosodic aspects. Hence, students with little foreign accentedness would receive higher grades and vice versa. It can be concluded from the results that voicing indeed played a role in foreign accentedness as it was reflected in their grading: those students who received higher grades showed no significant deviations in VOT from the SSBE norm while students with low grades did. Only one speaker (S1) did not match this pattern as she had significant deviations (due to a lack of aspiration in the voiceless plosives combined with prevoicing of the voiced plosives) but still received grade 2. In terms of the VOT categories proposed by Lisker and Abramson's (1964), it was also found that the students with higher grades matched the categories better than the ones with lower grades, again with the exception of S1.

A voicing neutralization, which is typical of SAG, was not found in the students' speech, with the exception of the bilabial plosive in S7, where the positive VOT values of /b/ coincided with the VOT values of /p/ due to a lack of aspiration. The student received the only negative grade, which allows a tentative link between grades and voicing neutralization. The nonexistence of a voicing neutralization in the other students, by contrast, indicates that no transfer occurred in their L2 English pronunciation in this regard. These students were obviously aware of the voicing distinction in English, which suggests a growing L2 pronunciation proficiency level.

The aforementioned lack of aspiration was also found in the voiceless alveolar plosive in S7 as well as in both voiceless plosives in S1. Due to prevoicing, however, no voicing neutralization was found in these cases. It can be concluded that these

two students, as well, were apparently aware of the voicing distinction in English, yet they were not able to realize the distinction in a native-like way. The lack of aspiration can be attributed to transfer from their L1 SAG. The interlanguage stage of S1 and S7 was still more affected by transfer compared to the other students, which implies a somewhat lower pronunciation proficiency level for these students.

A noticeably large difference between voiceless and voiced plosives was found in S6, which resulted from exaggerated aspiration combined with prevoicing. Since transfer can be ruled out in this case, the speaker may be in an intermediate language proficiency state, apparently moving away from L1 patterns by exaggerating L2 patterns (i.e. aspiration and voicing distinction).

To conclude, voicing of plosives has shown to have a certain influence on the foreign accentedness of the students' L2 English. In order to investigate which other phonetic cues may play a decisive role in foreign accentedness, follow-up studies of the same speakers will focus on further segmental aspects in the speaker's speech, including voicing in /s z ʃ ʒ/, replacements in /θ ð w v/, and vowel formants and length in selected vowels. In a similar manner, it will be investigated whether and how far these factors influenced the students' awarded grades. The results will then be compared to the VOT results discussed in the present study. This will provide further insights into the prominence of voicing in plosives for foreign accentedness in comparison to other phonetic cues.

Note: This paper was presented at the Fifth Belgrade International Meeting of English Phoneticians (BIMEP 2020), 20–21 March 2020, Faculty of Philology, University of Belgrade.

References

Austrian Federal Ministry of Education, Science and Research (2018). *Modellaufgabenstellungen für die mündliche Reifeprüfung in den Lebenden Fremdsprachen.* https://www.bmbwf.gv.at/Themen/schule/schulpraxis/zentralma-tura/srdp_ahs/mrp_lfsp.html

Auzou, P., Ozsancak, C., Morris, R. J., Jan, M., Eustache, F., Hannequin, D. (2000). Voice onset time in aphasia, apraxia of speech and dysarthria: A review. *Clinical Linguistics and Phonetics 14*(2), 131–150.

Boersma, P., Weenink, D. (2019). *Praat: Doing phonetics by computer.* Version 6.1.09. http://www.praat.org

Bongaerts, T. (1999). Ultimate attainment in L2 pronunciation: The case of very advanced late L2 learners. In D. Birdsong (Ed.), *Second Language Acquisition and the critical period hypothesis* (pp. 133–160). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Chao, K., Chen, L. (2008). A cross-linguistic study of Voice Onset Time in stop consonant production. *Computational Linguistics and Chinese Language Processing 13*(2), 215–232.

Derwing, T. M., Munro, M. J. (2009). Putting accent in its place: Rethinking obstacles to communication. *Language Teaching 42*(4), 476–490.

Deterding, D., Nolan, F. (2007). Aspiration and voicing of Chinese and English plosives. In J. Trouvain and W. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences* (pp. 385–388). Saarbrücken.

Docherty, G. J. (1992). *The Timing of Voicing in English Obstruents*. Berlin/New York: Foris.

Draxler, C., Jänisch, K. (2004). SpeechRecorder: A universal platform independent multichannel audio recording software. In *Proceedings of the Fourth International Conference on Language Resources and Evaluation* (pp. 559–562).

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

Gao, Z., Weinberger, S. (2018). Which phonetic features should pronunciation instructions focus on? An evaluation on the accentedness of segmental/syllable errors in L2 speech. *Research in Language 16*(2), 135–154.

Labov, W., Ash, S., Boberg, C. (2006). *The Atlas of North American English: Phonetics, Phonology, and Sound Change*. Berlin: Mouton-de Gruyter.

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

Major, R. C. (2007). Identifying a foreign accent in an unfamiliar language. *Studies in Second Language Acquisition, 29*, 539–556.

Major, R. C. (2008). Transfer in second language phonology. In J. G. Hansen Edwards and M. L. Zampini (Eds.), *Phonology and Second Language Acquisition* (pp. 63–94). Amsterdam/Philadelphia: John Benjamins.

Mareüil, P., Vieru-Dimulescu, B. (2006). The Contribution of Prosody to the Perception of Foreign Accent. *Phonetica, 63*, 247–267.

Moosmüller, S., Schmid, C., Brandstätter, J. (2015). Standard Austrian German. *Journal of the International Phonetic Association, 45*(3), 339–348.

Munro, M. J., Derwing, T. M., Burgess, C. S. (2003). The detection of foreign accent in backwards speech. In M. J. Solé, D. Recasens and J. Romero (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 535–538). Barcelona: Futurgraphic.

Saito, K., Trofimovich, P., Isaacs, T. (2016). Second language speech production: Investigating linguistic correlates of comprehensibility and accentedness for learners at different ability levels. *Applied Psycholinguistics, 37*, 217–240.

University of Graz (2020). *ENB.01140UB 19W 2SH KS Pronunciation*. https://online.uni-graz.at/kfu_online/wbLv.wbShowLVDetail?pStpSpNr=566384&pSpracheNr=2&pMUISuche=

University of Graz. n.d. *Academic Workload and Grading Information*. https://static.uni-graz.at/fileadmin/bib/downloads/studierende/incoming/transcript/bibwww_workload_grading.pdf

Wiesinger, P. (1996). Das österreichische Deutsch als eine Varietät des Deutschen. *Die Unterrichtspraxis/Teaching German, 29*(2), 154–164.

Winke, P., Gass, S., Myford C. (2013). Raters' L2 background as a potential source of bias in rating oral performance. *Language Testing, 30*, 231–252.

Milena Kong-Insam

Sažetak

VREME NASTUPA ZVUČNOSTI KAO FONETSKO OBELEŽJE STRANOG AKCENTA – STUDIJA SLUČAJA PET AUSTRIJSKIH STUDENATA ENGLESKOG

Prisustvo stranog akcenta se pokazalo kao odlika koja se izuzetno ističe pri upotrebi L2. Brojne fonetske studije su dokazale da su slušaoci veoma osetljivi na strani naglasak kod sagovornika (Derwing/Munro, 2009). Ipak, još uvek nije utvrđeno koja su fonetska obeležja od najvećeg značaja za percepciju akcenta kao stranog. U istraživanjima koja su se bavila ovim pitanjem došlo se do različitih rezultata, od segmentalnih odlika konsonanata i vokala (npr. Saito/Trofimovich/Isaacs, 2016; Gao/Weinberger, 2018; Labov/Ash/Boberg, 2006), do različitih prozodijskih karakteristika (npr. Mareüil/Vieru-Dimulescu, 2006). Ovo istraživanje je usmereno na kontrast u zvučnosti bilabijalnih i alveolarnih ploziva u govoru pet austrijskih studenata engleskog, koji se proučava u vidu studije slučaja. Istražiće se u kojoj meri izmerene vrednosti vremena nastupa zvučnosti (VOT) u njihovom izgovoru stranog jezika odstupaju od vrednosti VOT u britanskom engleskom. VOT je meren za bezvučne i zvučne bilabijalne i alveolarne plozive /p b t d/. Pretpostavka je da su plozivi izabrani za analizu posebno problematični za govornike nemačkog iz Austrije jer se zvučnost u engleskom i nemačkom drugačije realizuju. Do zaključaka o korelaciji između VOT i stranog naglaska doći će se na osnovu završnih ocena studenata na kursu engleskog izgovora koji su oni pohađali tokom trajanja ovog istraživanja.

Ključne reči:

strani naglasak, vreme nastupa zvučnosti (VOT), fonetska obeležja, segmentne odlike, plozivi, zvučnost, izgovor engleskog jezika, engleski kao L2, austrijski nemački

Appendix

Individual and mean VOT values

Speaker		BILABIAL									
		Voiceless					voiced				
Speaker	p1	p2	p3	p4	Ø	b1	b2	b3	b4	b5	Ø
S1	38	31	54	50	43.25	-92	-61	-92	-79	-95	-83.8
S2	61	38	49	68	54	9	9	25	11	20	14.8
S4	57	41	64	40	50.5	11	7	X	19	8	11.25
S6	42	63	77	89	67.75	-68	-64	-97	-74	-59	-72.4
S7	33	36	38	34	34.6	32	X	-93	28	27	-1.5
Ø	46.2	41.8	56.4	56.2	50.02	-21.6	-27.25	-64.25	-19	-19.8	-26.33

Table 7. Individual and mean VOT values for voiceless and voiced bilabial plosives /p/ and /b/. Negative (i.e., prevoiced) VOT values in /b/ are highlighted with italics.

Speaker		ALVEOLAR										
		voiceless						voiced				
Speaker	t1	t2	t3	t4	t5	Ø	d1	d2	d3	d4	d5	Ø
S1	49	51	41	70	66	55.4	20	-47	-97	-65	-46	-47
S2	42	68	64	66	79	63.8	15	16	15	18	15	15.8
S4	73	77	48	72	102	74.4	18	-39	-27	22	12	-2.8
S6	63	116	61	96	140	95.2	-73	-71	-39	-71	-74	-65.6
S7	25	36	41	47	45	38.8	-69	-33	-48	-46	6	-38
Ø	50.4	69.6	51	70.2	86.4	65.52	-17.8	-34.8	-39.2	-28.4	-17.4	-27.52

Table 8. Individual and mean VOT values for voiceless and voiced alveolar plosives /t/ and /d/. Negative (i.e., prevoiced) VOT values in /d/ are highlighted with italics.