

## VOT of English Plosives Produced by Polish Learners of English

Michał Baran  
[michal.baran@usz.edu.pl](mailto:michal.baran@usz.edu.pl)  
Institute of Linguistics  
University of Szczecin, Poland

### ABSTRACT

This paper attempts to contribute to the description of the characteristics of the acoustics of English spoken by Polish people by focussing on the VOT of plosives produced by speakers representing three different age groups, i.e. primary school students, high school students and university students. While there have been numerous studies on the acoustics of English spoken by Polish people, there seem to be few studies that compare the speech of speakers representing different age groups and/or levels – filling this gap in knowledge is the main research goal of this study. The data collected in this experiment is compared to reference values for English and Polish plosives which allows to check whether the participants are closer to the VOT values of native English realisations or those of the corresponding plosives in Polish. The study reveals that (i) /p, t, k/ are realised with long-lag VOT and their mean values are close to the ones of native English being, at the same time, higher than the ones reported for /p, t, k/ in Polish and (ii) /b, d, g/ are realised with negative VOT characteristic of prevoicing rendering them considerably different from the ones in native English and similar to their equivalents in Polish. Moreover, the differences between the three age groups are relatively small. The university students seem to be the ones with realisations closest to the native ones as they exhibit the highest VOT results of /p, t, k/, yet they still produce long periods of prevoicing in /b, d, g/. The results of the study suggest that more emphasis should be put on pronunciation practice of /b, d, g/ as it seems to be the more difficult group of English plosives for the Polish speakers.

**Keywords:** Polish English; VOT; L2 English; laryngeal contrast; acoustics of consonants

### INTRODUCTION

Plosives are defined as sounds involving a complete closure in the oral tract (Davenport & Hannahs, 2005: 19, Bickford & Floyd, 2006: 23). It is claimed that they are the only group of consonants present in all the languages studied so far (Ladefoged & Maddieson, 1996: 49). They are also one of the first phonemes to be acquired by children during L1 acquisition (see, for instance, Crowe & McLeod, 2020). The former and especially the latter might suggest that plosives are fairly easy from the articulatory point of view. Such a claim may be supported by the fact that, in comparison to other groups of obstruents, plosives require lesser articulatory precision. It should, however, be mentioned that cross-linguistically plosives exhibit a lot of variety in (i) glottal setting, (ii) airstream mechanism, (iii) articulation events during onset and offset, (iv) length and (v) strength (Ladefoged & Maddieson, 1996: 48-49). Therefore, it becomes evident that, despite the apparent simplicity of plosives, their acquisition may still be challenging in foreign language learning.

Polish learners of English need to be aware of two key facts while attempting to learn the pronunciation of the English /p, t, k, b, d, g/. Firstly, while Polish voiceless /p, t, k/ are described as exhibiting a short-lag VOT characteristic of voiceless unaspirated plosives (Kopczyński, 1971; Keating et al., 1981; Trochymiuk, 2008; Konopska & Sawicki, 2013), their English equivalents are pronounced with a long-lag VOT characteristic of voiceless, aspirated plosives (Ladefoged & Johnson, 2010).<sup>1</sup> Secondly, Polish /b, d, g/ feature negative VOT values that identify them as prevoiced, unaspirated plosives (Kopczyński, 1971; Keating et al., 1981; Trochymiuk, 2008; Konopska & Sawicki, 2013). English /b, d, g/, on the other hand, are pronounced with a short-lag VOT characteristic of devoiced, unaspirated plosives (Ladefoged & Johnson, 2010). Therefore, Polish learners of English must achieve two goals while learning the pronunciation of the English /p, t, k, b, d, g/, i.e. they must (i) learn to produce longer VOT periods in /p, t, k/ and (ii) learn to produce /b, d, g/ with no prevoicing. These two changes are obviously not impossible to introduce, but they may require targeted practice which is usually absent in primary and high school education as, if there are any pronunciation exercises focused on consonants present in the programme, the attention is usually focused on segments absent from the sound inventory of Polish such as dental fricatives.

There have been numerous studies on the acoustics of English spoken by Polish learners discussing its various aspects (see, for instance, Rojczyk, 2010; Schwartz, 2012; Rojczyk & Porzuczek, 2012; Schwartz & Kaźmierski, 2019 and Sypiańska, 2021; Waniek-Klimczak, 2005 for plosives specifically) as well as didactic publications targeted specifically at Polish learners (see, for instance, Bałutowa, 1990; Sobkowiak, 1996). Nevertheless, due to a dynamically changing nature of the influence of English in Poland (see, for instance, Kasztalska 2014), it seems reasonable to keep researching the acoustics of English produced by its Polish learners as such research may give insight into generational changes, i.e. answer the question whether Polish people are getting better at English pronunciation due to an increasing exposure to this language or not. Additionally, studies on the acoustics of English spoken by Polish learners comparing various age groups using the same exact methodology are relatively scarce which seems to be a gap in knowledge that should be filled.

The primary focus of this study is the acoustic analysis of the English /p, t, k, b, d, g/ pronounced by Polish learners representing three different age groups, i.e. primary school students, high school students and university students. The selection of English plosives as the subject of the study is motivated by the fact that many learners may consider them identical with their Polish equivalents and, as a result, neglect practicing them. Moreover, comparing three different age groups fills a gap in scientific literature on English pronounced by Polish speakers. Looking at the realisations by participants from three different age groups allows to reveal differences between the three groups or, perhaps even more interestingly, the lack of such differences. The basic, maybe even simplistic, assumption is that the more time a learner spends on learning a given language, the better their pronunciation should get. It should, therefore, be the case that university students who have completed their primary and high school education and are in the process of doing their university degree (which includes English phonetics classes) should exhibit pronunciation that is considerably better than that of high school or primary school students. Additionally, a comparative study like the one described here may also reveal differences resulting from generational differences as the age difference between, for instance, primary school and university students is considerable. Based on all that, this study attempts to check whether (i) the VOT values of the English /p, t, k, b, d, g/ produced by Polish learners representing the three age groups under review

---

<sup>1</sup> This is the case in word-initial plosives found in stressed syllables, i.e. the position on which this study is focused.

are more similar to those of native English realisations or to those of their Polish equivalents, (ii) any of the two laryngeal series of the English plosive inventory appears more problematic to the participants of the study and (iii) there are any differences between the VOT values produced by the speakers representing the three age groups under review.

The main part of this article opens with a discussion of the theoretical background of the study, i.e. the key aspects of plosive articulation, the concept of VOT in acoustic phonetics, a brief overview of research on VOT of Polish and English plosives and a summary of a study similar to the one undertaken here. Then, the study itself is presented by discussing its methodology and results. These results are then discussed in the following section and the answers to the research questions are proposed. Finally, the paper is closed with a conclusion that, apart from summarising the key points of the paper, explores directions in which this research may be developed further.

## **THEORETICAL BACKGROUND OF THE STUDY**

### **KEY ARTICULATORY STAGES OF PLOSIVE ARTICULATION**

As aforementioned in the introduction, plosives constitute a class of consonants that exhibits a lot of variety regarding articulatory setting. The introduction to plosives presented here does not, however, attempt to discuss all of these possible articulatory variations but focuses on the aspects relevant to English and Polish plosives.

There are three key phases of articulating a plosive that have to be discussed here, i.e. hold phase, release phase and post-release phase. The hold phase starts when the articulators form a complete closure in the oral tract, and it ends when the air trapped behind the articulators is released (Davenport & Hannahs, 2005: 26). From the point of view of acoustic phonetics, the duration of the hold phase is worth attention as it may be an acoustic feature that is used in distinguishing between the homorganic plosives of different laryngeal series (see, for instance, Fulop, 1994: 56-57). Once the air trapped behind the articulators is released, the release phase begins. The key element of the release phase is the release burst which may be defined as “a small audible explosion which accompanies the release of a plosive” (Trask, 1996: 61). The release burst is another part of plosive articulation that deserves attention of phoneticians studying plosive contrasts as, similarly to the hold phase duration, it has been shown to be one of the acoustic cues marking the differences between plosives of different places of articulation and laryngeal settings (see, for instance, Chodroff & Wilson, 2014). Finally, the last phase of plosive articulation is the post-release phase. This is the period of time between the end of the release burst and the beginning of the articulation of the following sound. The post-release phase is often the one that attracts most attention in studies on plosive contrasts as it may feature aspiration which, in numerous languages, correlates with both laryngeal series and place of articulation contrasts (for a more detailed discussion of aspirating languages see, for instance, Baran 2024).

### **VOT – AN OVERVIEW**

Aspiration is frequently studied using the parameter of voice onset time (commonly abbreviated as VOT). The publication by Lisker and Abramson (1964) is universally accepted as the first major one introducing the VOT as a parameter in phonetic analysis. The idea was, however, explored earlier, especially by Adjarian in the late 19<sup>th</sup> Century and in speech synthesis research starting in 1950s (see Konopska & Sawicki, 2013: 103). VOT may be defined as the period of time between

the release of the plosive and the beginning of modal voicing (Ladefoged & Maddieson, 1996: 45). Ladefoged and Maddieson (1996: 45) distinguish three categories within which VOT values may be classified, i.e. negative VOT, short-lag VOT and long-lag VOT. Negative VOT is typical of phonetically voiced, unaspirated plosives in which voicing starts before the release of the closure, thus resulting in prevoicing. A spectrogram and an oscillogram of a word featuring a prevoiced realisation of /b/ are shown in Figure 1. below.

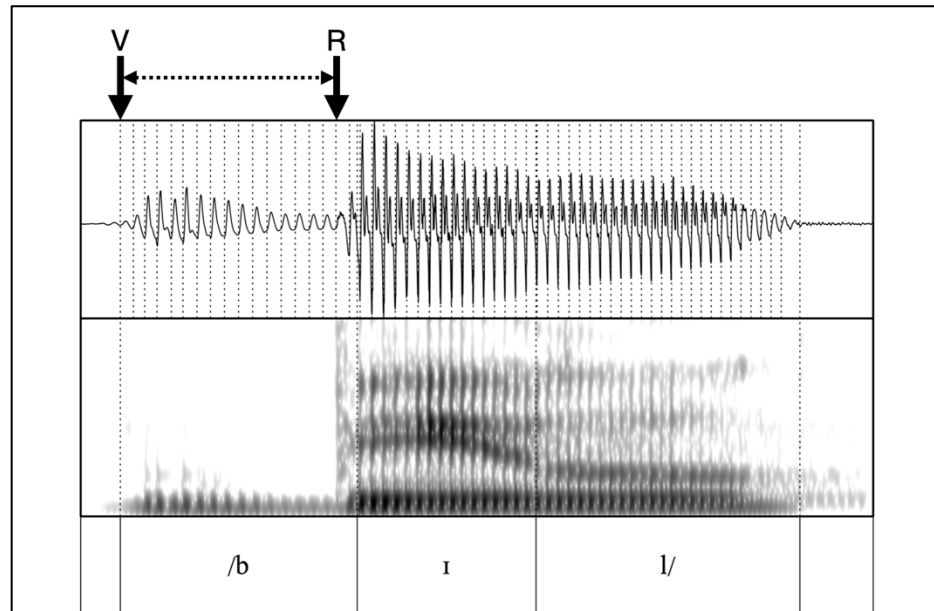


FIGURE 1. A prevoiced realisation of /b/ in the word 'bill' pronounced by S3

The arrow on the left marked with ‘V’ marks the beginning of modal voicing that can be identified both on the basis of pulses (dotted vertical lines on the oscillogram) and the presence of the voice bar (concentration of acoustic energy of low frequency). It is clearly visible that modal voicing begins before the release of /b/ marked with the arrow on the right and the letter ‘R’. As the release happens after the beginning of modal voicing rather than before, the length of this period is expressed with a negative number.

The second VOT category is referred to as a short-lag VOT. It is characteristic of phonetically voiceless, unaspirated plosives. Figure 2. below contains a spectrogram and an oscillogram of /d/ realised with a short-lag VOT.

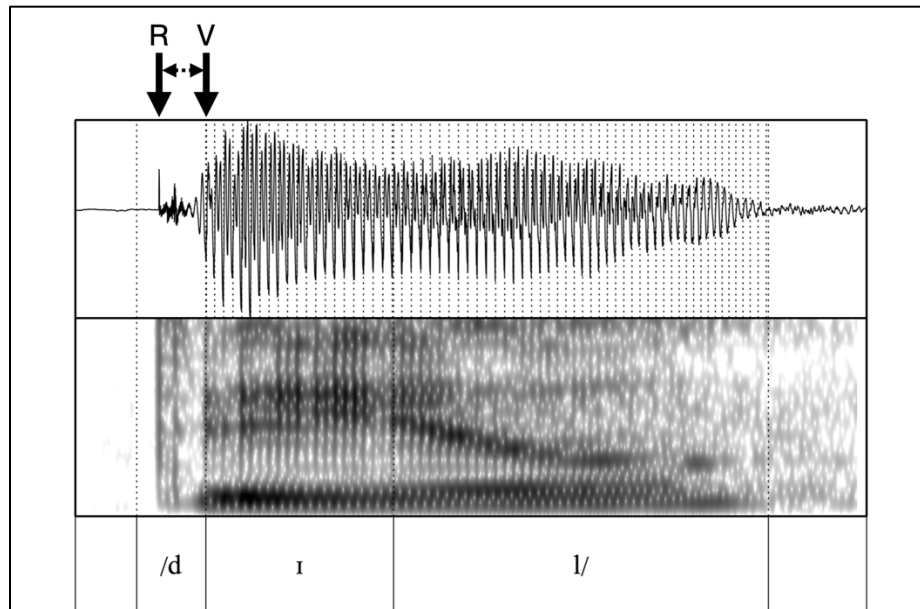


FIGURE 2. A phonetically voiceless, unaspirated realisation of /d/ in the word 'dill' pronounced by S9

In the realisation of /d/ shown in Figure 2. above, the beginning of modal voicing happens after the release burst. As a result, the length of this period is expressed with a positive number.

Finally, the third VOT category proposed by Ladefoged and Maddieson (1996) is long-lag VOT, characteristic of phonetically voiceless, aspirated plosives. Figure 3. below contains an oscillogram and a spectrogram of a long-lag VOT realisation of /k/.

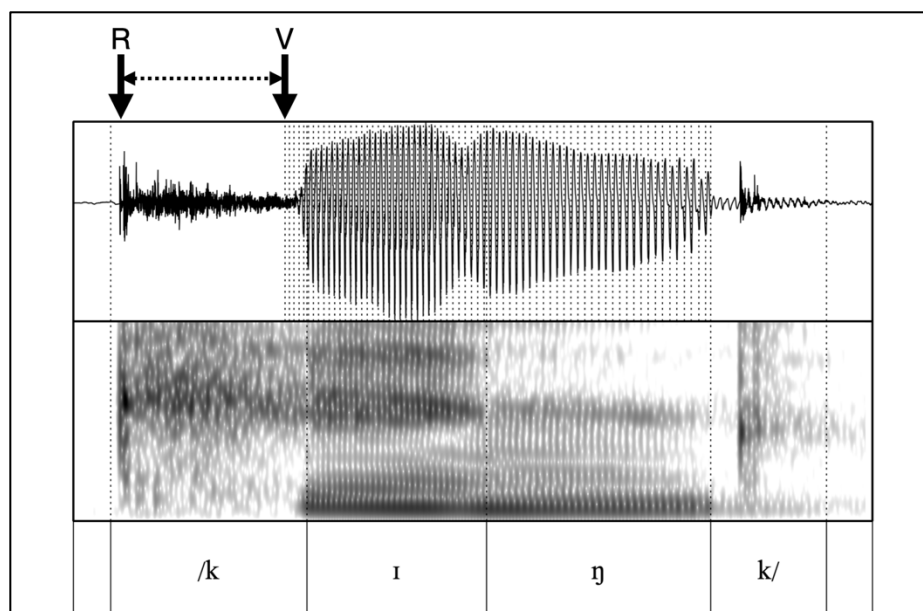


FIGURE 3. A phonetically voiceless, aspirated realisation of /k/ in the word 'kink' pronounced by S1



Similarly to Figure 2., Figure 3. also depicts a realisation in which the release precedes the beginning of modal voicing, thus resulting in a positive value of the VOT parameter. In Figure 3., however, it is visible that (i) the post-release stage is much longer than in Figure 2. and (ii) the distribution of acoustic energy in the post-release stage is also different than in Figure 2.

### VOT OF POLISH AND ENGLISH PLOSIVES

Having introduced the three most common categories of VOT values, it is now possible to discuss the VOT of English and Polish plosives answering the question of how these two languages use the VOT to mark their laryngeal contrasts.

The English language is frequently referred to as a language that features a fortis-lenis or aspirated-deaspirated laryngeal contrast, especially visible in obstruents, as opposed to a voiceless-voiced contrast.<sup>2</sup> As a result, it is said to feature two laryngeal series of plosives, i.e. fortis/aspirated /p, t, k/ and lenis/unaspirated /b, d, g/. Lisker and Abramson (1964) claim that English /p, t, k/ are realised with long-lag VOT while their lenis/unaspirated counterparts are realised with short-lag VOT. This observation is confirmed by the study that is used as a source of reference values for English plosives here, i.e. the publication by Docherty (1992). In this publication, Docherty (1992) reports mean VOT values presented in Table 1. below.

TABLE 1. Mean VOT values of English plosives (Docherty 1992)

Sound	Mean VOT Value (ms)
/p/	42
/t/	65
/k/	62
/b/	15
/d/	21
/g/	27

The values reported by Docherty<sup>3</sup> (1992) mostly go in line with the thresholds established by Lisker and Abramson (1964), who claim that (i) a short-lag VOT falls between 0 and 25 milliseconds, (ii) a long-lag VOT falls between 60 and 100 milliseconds and (iii) a negative VOT is below 0. English /p, t, k/ that are assumed to be phonetically voiceless and aspirated exhibit values that may be classified as a long-lag VOT (/p/ does have a mean value lower than 60 milliseconds, but it is still higher than the threshold for the short-lag VOT) while /b, d, g/ that are said to be phonetically devoiced and unaspirated exhibit values that may be classified as a short-lag VOT. It should be mentioned here that the data collected by Docherty included examples of prevoiced realisations of /b, d, g/ with negative VOT values, but such realisations were rare (26 out of 372 realisations) and, as a result, they were not included in the calculation of the mean VOT values.

<sup>2</sup> However, despite the fact that English is one of the best described language in the world, there is still no universally accepted description of its laryngeal system. As a result, there are studies that still use the phonological contrast voiceless-voiced to refer to English plosives (see Baran, 2024 for a more detailed discussion of this issue). This issue, however, goes beyond the scope of this study as it is not concerned with the phonological labels of the sounds under review.

<sup>3</sup> Interestingly, these values do not confirm the general assumption regarding the relationship between VOT and the place of articulation wherein velar consonants are the ones with the longest VOT periods.

While English is frequently seen as a fortis-lenis or aspirated-deaspirated language, Polish occupies the other end of the spectrum as it is usually seen as a voiceless-voiced language (see, for instance, Cyran, 2011 for an extensive discussion of this issue). Polish plosives are, therefore, described as (i) phonetically voiceless and unaspirated as in the case of /p, t, k/ and (ii) phonetically voiced and unaspirated as in the case of /b, d, g/. In light of the assumptions regarding VOT presented above, /p, t, k/ should exhibit short-lag VOT and /b, d, g/ should exhibit negative VOT. This is confirmed by the study by Konopska and Sawicki (2013) who report mean VOT values shown in Table 2. below.

TABLE 2. Mean VOT values of Polish plosives (Konopska & Sawicki, 2013)

Sound	Mean VOT Value (ms)
/p/	24
/t/	27
/k/	55
/b/	-102
/d/	-110
/g/	-82

As expected, the mean VOT values of Polish /p, t, k/ fall within the spectrum that allows to interpret it as short-lag VOT. /k/ may appear slightly problematic with a mean VOT of 55 milliseconds, but a higher value in the case of a velar sound may be explained by the general tendency that a more posterior place of articulation results in a longer VOT (Konopska & Sawicki, 2013: 105). Polish /b, d, g/ exhibit long periods of negative VOT which clearly identifies them as prevoiced.

#### A SUMMARY OF AN EXAMPLE STUDY OF VOT OF ENGLISH PLOSIVES PRONOUNCED BY POLISH SPEAKERS

The study by Sypiańska (2021) is summarised in this part of the article as it is quite similar to the one undertaken here, i.e. it is a study of VOT of English plosives pronounced by Polish learners of English (for other examples of such studies, see, for instance, Waniek-Klimczak, 2005). In her study, Sypiańska (2021) analysed recordings of L2 English spoken by Polish senior learners in order to check (i) to what extent senior learners are able to produce L2 plosives and (ii) whether the participants of the study experience L1 drift resulting from learning English. As only (i) is relevant to the present study, the summary focuses on this research question.

Sypiańska (2021) analysed the recordings of Polish learners of English (n=20, all female, average age=68.8) reading a list of English and Polish monosyllabic words (18 for English and 18 for Polish) containing /p, t, k, b, d, g/ in word-initial position, followed by a front mid vowel. The VOT was measured based on visual identification of (i) the release burst and (ii) the first regular pulse.

The mean VOT values for the English /p, t, k, b, d, g/ given by Sypiańska (2021) are presented in Table 3.

TABLE 3. Mean VOT values of English plosives produced by senior Polish learners of English (Sypiańska, 2021)

Sound	Mean VOT Value (ms)
/p/	27.2
/t/	39.5
/k/	49.4
/b/	-83.6
/d/	-100.1
/g/	-70

The values presented in Table 3. suggest that while the values for /p, t, k/ are close to the ones of native English realisations, the mean VOT values VOT of /b, d, g/ are still negative which makes them more similar to the ones in Polish. However, Sypiańska (2021) mentions that these values are indeed different for English plosives than for Polish ones produced by the same speakers, i.e. longer VOT values were recorded for the English /p, t, k/ than for the Polish equivalents and shorter negative VOT values were recorded for /b, d, g/ in English plosives than in the Polish ones. This seems to suggest that the learners produce these sounds differently when they speak Polish and English.

Sypiańska (2021) also analyses the results by comparing it to thresholds of correct realisations established on the basis of literature review. The results of this comparison are presented in Table 4. below.

TABLE 4. Percentage of realisations analysed as correct by Sypiańska (2021)

Sound	Threshold	% of correct realisations
/p/	VOT > 30 ms	30
/t/	VOT > 35 ms	27
/k/	VOT > 55 ms	39
/b, d, g/	VOT > 0 ms	22

The data presented in Table 4. above suggests that /b, d, g/ were actually more difficult to pronounce correctly for the participants of Sypiańska's (2021) study than /p, t, k/. It may, therefore, be concluded that they found it more difficult to limit the periods of prevoicing in /b, d, g/ than to extend the VOT in /p, t, k/. This is definitely an interesting finding that has to be confronted with the results of the analysis conducted here for participants representing different age groups.

## THE STUDY

### METHOD

This study is based on recordings of 30 participants representing three different age levels (primary school n=10, high school n=10 and university n=10) reading sentences from a reading list. This section presents the method by providing more details on participants, the design of the reading list and the way in which the data was gathered and analysed.



As aforementioned, the participants of this study represent three different age groups, i.e. primary school students, high school students and university students. All of the participants, regardless of the age group, are native speakers of Polish who (i) were born and raised in Poland and (ii) have never lived in an English-speaking country. Table 5 below presents the gender and age of the participants.

TABLE 5. Gender and age of the participants

Code	Age	Gender <sup>4</sup>
<b>Primary school students</b>		
P1	13	M
P2	14	M
P3	12	F
P4	12	F
P5	11	M
P6	13	M
P7	11	M
P8	14	M
P9	14	M
P10	13	M
Average age: 12.7		8 males 2 females
<b>High school students</b>		
L1	18	F
L2	18	F
L3	17	F
L4	16	F
L5	18	F
L6	17	M
L7	16	M
L8	16	F
L9	15	F
L10	16	M
Average age: 16.7		3 males 7 females
<b>University students</b>		
U1	20	F
U2	21	F
U3	21	M
U4	20	F
U5	20	F
U6	20	F
U7	21	F
U8	20	F
U9	21	F
U10	20	F
Average age: 20.4		1 male 9 females

Primary school students (codes starting with P), at the time of making the recordings, were in their 6<sup>th</sup>, 7<sup>th</sup> or 8<sup>th</sup> year of primary school education. Polish primary school students take an exam in English at the end of year 8. The exam checks whether the students can use A2/B1 English. The assumption is, therefore, that the English level of the participants representing primary school in this study should oscillate around A2.

<sup>4</sup> As the influence of gender on VOT is not studied here, gender was not taken into consideration during the selection of participants.

High school students (codes starting with L), at the time of making the recordings, where in their 1<sup>st</sup>, 2<sup>nd</sup> or 3<sup>rd</sup> year of high school education. Polish high school students take the Matura Exam at the end of high school education (in the case of high school, after year 4) which requires B1+/B2 English in the basic exam and B2+ English in the extended one. As a result, it may be assumed that the English level of the participants representing high school should oscillate around B1.

All university students (codes starting with U) recorded for the study are students of the second year of Global Communication programme at the University of Szczecin (Poland). It is a six-semester long BA programme conducted fully in English. It includes a subject called English Phonetics taught in the first and second semester of year one. In these classes students are taught basic theoretical background of English phonetics and phonology and take part in pronunciation exercises aimed at improving their pronunciation. As Global Communication students should reach C1 level of English by the time they finish the programme, it is assumed here that the English level of the participants representing university should oscillate around B2+<sup>5</sup>.

During the recording sessions, the participants were asked to read a list of sentences. The sentences on the list were created by putting a target lexeme at the beginning of a carrier phrase “X is the right word.”. The target lexemes were monosyllabic words with /p, t, k, b, d, g/ found in word-initial position followed by /ɪ/<sup>6</sup>. The sentences were repeated so that each of the target sounds had 20 repetitions by a given speaker. The participants read the list of sentences while sitting 20 centimetres away from Røde NT-USB+ microphone connected to a laptop. Sampling rate of the recordings was set to 48.000 Hz.

The recordings were analysed with Praat software (Boersma & Weenink, 2024) in order to get the VOT values of the sounds under review. VOT was measured on the basis of visual interpretation of the spectrogram and oscillogram leading to identification of the release burst and the beginning of modal voicing. The release burst was identified as the place where the acoustic activity starts after a period of silence resulting from the hold phase, while the beginning of modal voicing was identified as the place where the first pulse is visible (see Fig. 1-3). If the release burst happened before the beginning of modal voicing, the VOT value was recorded as a positive one. If, on the other hand, the release burst happened after the beginning of modal voicing, the VOT value was recorded as a negative value.

The collected data was analysed using Excel with a specific focus on mean values, standard deviation and comparing the data to thresholds of correct realisations. Results of the analysis are presented in the following section.

## RESULTS

The results regarding the VOT values of /b, d, g/ are presented first. Figures 4.-6. below present the percentage of realisations with no prevoicing for /b/, /d/ and /g/ respectively (the figures do not include the participants who had only prevoiced realisations). These realisations may be seen as correct ones, i.e. similar enough to the native English realisations.

---

<sup>5</sup> The assumptions regarding the level of English of the participants result from approximation and are not key to the study. The three groups are quite homogenous as far as age is concerned which means that the participants from a given group have had very similar exposure to English throughout their education.

<sup>6</sup> The decision that the target sounds were always followed by the same vowel was motivated by achieving the highest possible consistency of the results. Further studies should look at data gathered for plosives found before other vowels of English.

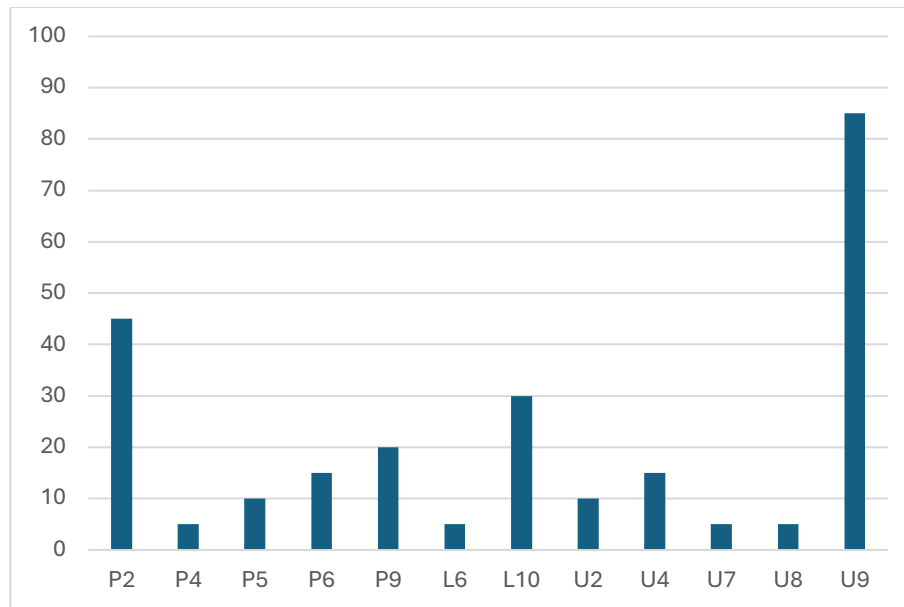


FIGURE 4. Percentage of correct realisations of /b/ by individual speakers

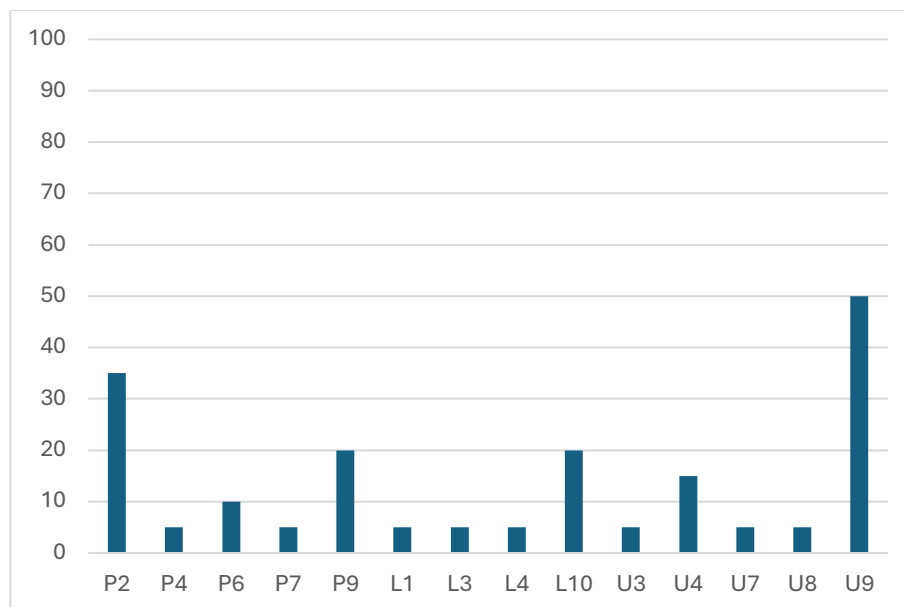


FIGURE 5. Percentage of correct realisations of /d/ by individual speakers

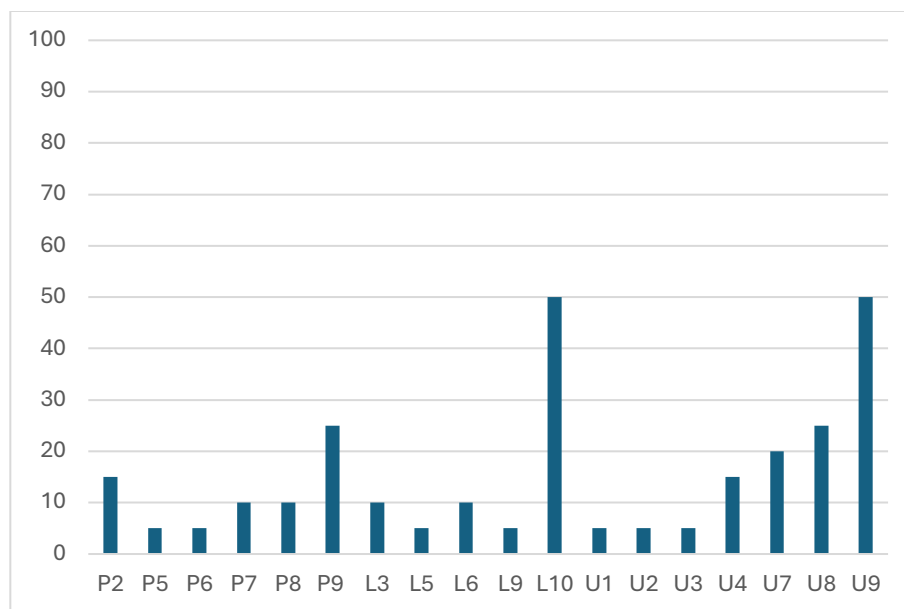


FIGURE 6. Percentage of correct realisations of /g/ by individual speakers.

In the case of /b/, correct realisations were produced by 5 primary school students, 2 high school students and 5 university students. It should, however, be noted that only 2 participants managed to produce more than 40 per cent of correct realisations of /b/ i.e. P2 with 45 per cent correct realisations and U9 with 85 per cent correct realisations. In the case of /d/, correct realisations were produced by 5 primary school students, 4 high school students and 5 university students. In this case, only U9 had more than 40 per cent of correct realisations. Finally, /g/ had correct realisations coming from as many as 6 primary school students, 5 high school students and 7 university students. L10 and U9 produced 50 per cent of correct realisations while other participants' results were lower than 30 per cent. All in all, it is visible that, as far as /b, d, g/ are concerned, some participants did not manage to produce any correct realisations. To illustrate the problem further, Figure 7. below presents the percentage of correct realisations of /b, d, g/ in all three age groups (P – primary school; L – high-school; U – university).

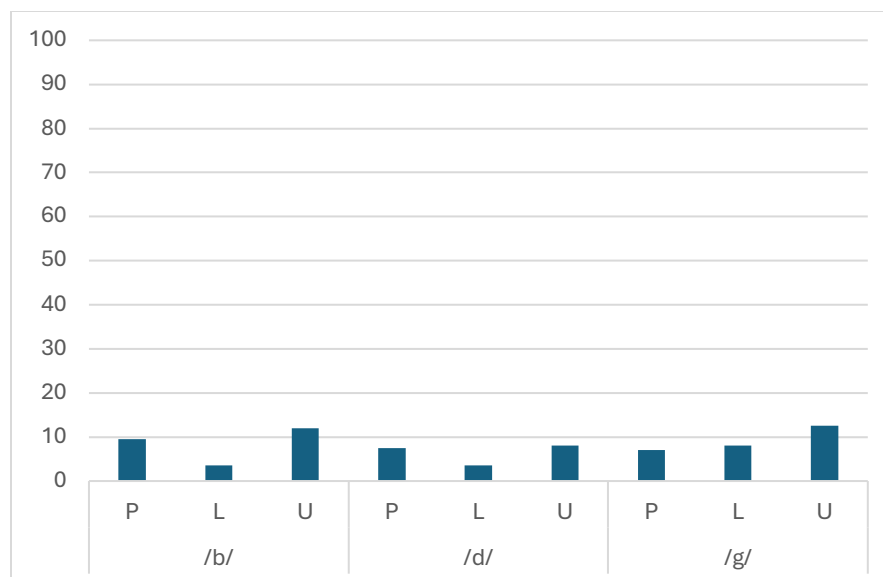


FIGURE 7. Percentage of correct realisations of /b, d, g/ in the three age groups.

Figure 7. shows clearly that correct realisations of /b, d, g/ never constitute more than 15 per cent of all realisations regardless of the group. Additionally, it should be noted that the university students have the highest percentage of correct realisations for /b/, /d/ and /g/. However, the second best result comes from the primary school students in the case of /b/ and /d/, while high school students are second only in the case of /g/.

In line with the theoretical considerations presented by Sypiańska (2021) the thresholds for correct realisations of the English /p, t, k/ are set to 30, 35 and 42 milliseconds respectively. Figure 8. below presents the percentage of correct realisations of /p, t, k/ in the three age groups under review here.

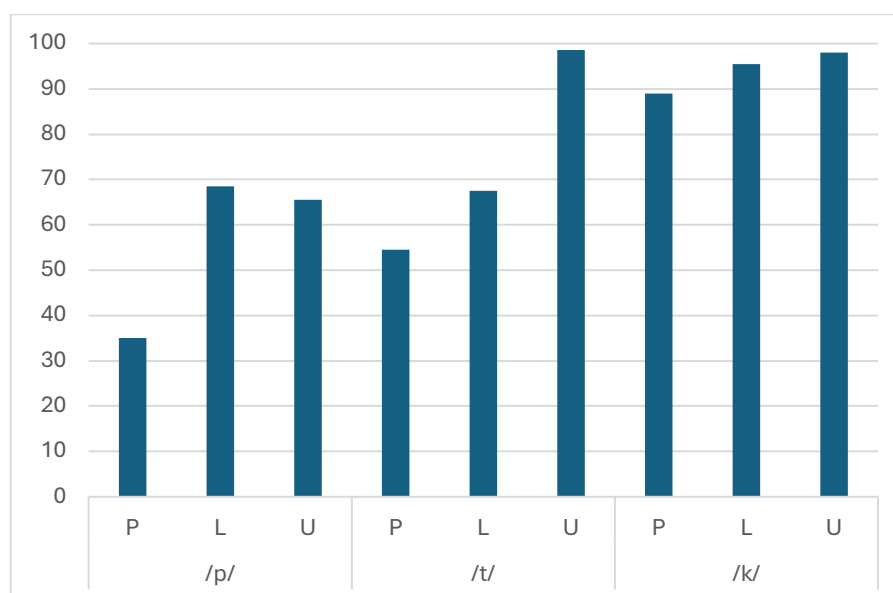


FIGURE 8. Percentage of correct realisations of /p, t, k/ in the three age groups under review

Figure 8. shows that /p, t, k/ were pronounced correctly far more frequently than /b, d, g/. While the highest results in Figure 7. is 12.5 per cent of correct realisations of /g/ pronounced by university students, in Figure 8. the lowest result is 35 per cent of correct realisations of /p/ by primary school students. Apart from /p/ produced by primary school students, all other results are higher than 50 per cent. As for the differences between the age groups, university students have the highest result in the case of /t/ and /k/, while the high school students are slightly better in the case of /p/. Primary school students have the lowest result for all three sounds under review.

Having discussed the results recorded in this study to the thresholds of correct realisations set for /p, t, k, b, d, g/, it is now possible to compare the average values for each of the three groups to reference VOT values regarding Polish and English plosives.

TABLE 6. Mean values and standard deviation of VOT values for /p, t, k, b, d, g/ pronounced by primary, high-school and university students compared to reference values for Polish and English plosives

	Polish reference (Konopska & Sawicki, 2013)	English reference (Docherty, 1992)	Primary school	High school	University
/p/	24	42	27 (SD=12.1)	38 (SD=16.6)	40 (SD=18.5)
/t/	27	65	38 (SD=16.7)	47 (SD=17.8)	75 (SD=18.9)
/k/	55	62	67 (SD=24.2)	70 (SD=19.3)	75 (SD=18.9)
/b/	-102	15	-99 (SD=62.1)	-102 (SD=51.7)	-78 (SD=54.7)
/d/	-110	21	-90 (SD=57.2)	-91 (SD=49.7)	-92 (SD=48.8)
/g/	-82	27	-80 (SD=49.5)	-77 (SD=56)	-78 (SD=54.7)

In the case of /p/, mean VOT values calculated for realisations by students from all three age groups are higher than that of the Polish /p/. However, only the realisations by high school and university students exhibit mean VOT values similar (but still lower) than the mean VOT reported by Docherty (1992) for British English. A comparison between the mean VOT values in the three age groups reveals that university students produced /p/ with the highest mean VOT value, while high-school students and primary school students rank second and third respectively.

Mean VOT values calculated for realisations of /t/ by students from the three age groups under review are all higher than those of the corresponding sound in Polish. In the case of primary and high-school students, the realisations of /t/ studied here feature VOT of a mean value much higher than that of the Polish /t/ (38ms, 47ms and 27ms respectively), but these values are still lower than the mean VOT reported for the British English /t/ by Docherty (1992). University students are the only group that produced VOT of a mean value that was actually longer than the one reported by Docherty (1992). Similarly to /p/, /t/ has highest mean VOT values in realisations by university students while high-school and secondary school students rank second and third respectively.

The realisations of /k/ by representatives of the three age groups under review exhibit VOT of mean values that are (i) higher than the reference value for the Polish /k/ and (ii) higher than the reference value for the English /k/. The three groups still rank the same as in the case of /p/ and /t/, i.e. university students exhibit the highest mean VOT value, while high-school students exhibit a mean VOT value that is lower than that of university students, but higher than that of primary school students.



The mean VOT values for /b/ pronounced by students from the three age groups are all negative, which is a reflection of the fact that a vast majority of realisations feature prevoicing. As a result, these mean VOT values are more similar to that of the Polish /b/ than to the value given by Docherty (1992) for British English. It should, however, be mentioned that the mean VOT recorded in realisations by university students is considerably higher than the Polish reference value (-78ms and -102ms respectively).

In the case of /d/, the mean VOT values are again negative for all the three age groups under review which also results from the domination of prevoiced realisations. In this particular case, however the values from the three groups are higher than the reference one for Polish but are also very similar to each other which does not allow to claim that one of the groups is less similar to the Polish reference value.

Finally, in the case of /g/, primary school students, high-school students and university school students produced /g/ with negative mean VOT values that are (i) very similar to the Polish reference data and (ii) very similar to values of other age groups.

## DISCUSSION

In order to facilitate the discussion of the results, Table 7. below summarises the observations from the previous section. To do so, it uses an arbitrary categorisation of results into one of the three categories, i.e. (i) A – a sound has been learnt successfully, (ii) B – a sound is still being learnt and (iii) C – no significant change of the sound. In the case of percentage of correct realisations column, results lower than 30 per cent are assigned category C, results between 31 per cent and 79 per cent are assigned category B and results higher than 80 per cent are assigned category A. In the comparison with reference data column, category A is assigned if the mean VOT is very close to the English reference value, category B is assigned if the mean VOT is higher than the Polish reference value, but still lower than the reference one for English and category C is assigned if the mean VOT is very close to the Polish reference value.

TABLE 7. Preliminary interpretation of results

Sound	Age group	Correct realisations (%)	Comparison with reference data
/p/	Primary school	B	B
	High-school	B	A
	University	B	A
/t/	Primary school	B	B
	High-school	B	B
	University	A	A
/k/	Primary school	A	A
	High-school	A	A
	University	A	A
/b/	Primary school	C	C
	High-school	C	C
	University	C	B
/d/	Primary school	C	B
	High-school	C	B
	University	C	B
/g/	Primary school	C	C
	High-school	C	C
	University	C	C

As far as /p, t, k/ are concerned, the results of the present study suggest that the Polish students of English are relatively successful in learning these sounds, i.e. learning to produce longer VOT periods than they do for the corresponding sounds in Polish. None of the three age groups produced realisations of /p, t, k/ that can be classified as very similar to those of the Polish /p, t, k/; they either produced realisations very close to the English reference data or ones that point to being in the transition period from the Polish VOT values to the English ones. It may, however, be claimed that the three sounds are not equally easy for the participants of the study. The results suggest that /k/ is the easiest, both in terms of the percentage of correct realisations and the comparison with reference data. It may result from the fact that VOT is generally the longer, the more retracted the place of articulation of a consonant (Konopska & Sawicki, 2013: 105), which, in turn, makes it easier for the Polish students to produce VOT of correct length in /k/. /p/ and /t/, on the other hand, are produced correctly by the university students (apart from a lower correctness score for /p/), while remaining still somewhat problematic for the primary and high-school students.

The English /b, d, g/ are definitely more problematic for the participants of the study than their voiceless counterparts. For all three sounds and in all three age groups under review, the percentage of correct realisations is very low which may be interpreted as a proof of the fact that correct realisations are on the verge of being accidental. The comparison with the reference data offers a bit more detailed insight into this matter as it demonstrated that in the case of (i) realisations of /b/ by the university students and (ii) realisations of /d/ by all three age groups the VOT values are indeed still negative, but higher than those of the corresponding sound in Polish. This is here interpreted as a sign of a transition from the Polish-like realisations towards ones that are more similar to the English ones.

As for the differences between the three age groups studied here, two main observations can be made. Firstly, the university students clearly exhibit VOT periods that are the closest to the English reference ones. In the case of /p, t, k/ the VOT periods produced by the university students are generally interpreted as very close to native realisations, both in terms of the percentage of correct realisations and the mean VOT values. In the case of /b, d, g/ the values are, as in the case of all groups in the study, far from native ones, but a look at mean VOT values reveals that those of /b/ and /d/ may be interpreted as indicative of sounds undergoing a change towards a more native-like realisations.

Secondly, primary and high-school students produced realisations that are quite similar to each other and do not reflect a longer period of studying English in the case of the high-school students. Table 7. features one notable difference between the realisations produced by the students from these two age groups, i.e. the high-school students produced realisations of /p/ with a mean VOT that is very close to the native one, while the realisations produced by primary school students feature a mean VOT that is still shorter than the reference mean VOT. Apart from that, the results for other sounds are classified in the same categories in Table 7. which suggests that the two age groups are actually quite similar in terms of their realisations of the English plosives.

## CONCLUSION

As aforementioned in the introduction, the present study was conducted to check whether (i) the VOT values of English /p, t, k, b, d, g/ produced by Polish learners of the three age groups under review are more similar to those of native English or to those of their Polish equivalents, (ii) any of the two laryngeal series of the English plosive inventory appears more problematic to the participants of the study and (iii) there are any differences between the VOT values produced by the speakers representing the three age groups under review.

As for (i) and (ii), the results presented here seem to show a visible difference between the two series of the English plosives studied here. /p, t, k/ are realised with VOT periods ranging from very similar to quite similar to the reference values given for native English realisations. They are, at the same time, considerably higher than the reference values given for Polish. These two observations allow to draw a conclusion that the participants of the study have developed a relatively good pronunciation of the English voiceless plosives. This finding should, however, be further verified by looking at the spectral features of the aspiration and by conducting a subsequent experiment that would allow for a comparison of /p, t, k/ pronounced by the same speakers in English and Polish words.

In the case of /b, d, g/, the situation is considerably different. None of the groups produced realisations that regularly featured VOT periods typical of native English realisations. Regardless of the age group, correct realisations are rare and mean VOT values are negative which is not a dominant way of producing /b, d, g/ in English. At best, the results obtained for /d/ (all three groups) and /b/ (only the university students) may suggest that the students are at the stage of developing a correct way of producing these sounds, but, for now, it is far from correct. As a result, it may be claimed that /b, d, g/ are more difficult to acquire than /p, t, k/ for the Polish learners of English whose speech was analysed here. Additionally, the difference between the results for the two series of plosives makes it difficult to answer (i) in a simple way. Instead, it has to be claimed that while /p, t, k/ produced by the participants of the study seem to be close to native realisations, /b, d, g/ still remain closer to their Polish counterparts due to prevoicing. This observation should be considered in context of potential didactic implications, i.e. it appears that both series require targeted pronunciation practice if learners are to develop natural pronunciations.

The aforementioned targeted pronunciation practice, or the lack of it, is probably the factor that explains the results of the comparison between the three age groups under review. The results of the university students are considerably better than those of primary and high-school ones, while the difference between the results of the primary and high-school students is a minor one. This may result from the fact that only the university students have been subjected to a repeated, targeted pronunciation practice. In order to support this interpretation further, or to disprove it, further research should be carried out – the results of such studies should be quite valuable from a didactic standpoint as they may contribute to making pronunciation teaching methodologies in Poland more efficient.

## REFERENCES

- Bałutowa, B. (1990). *Wymowa angielska dla wszystkich*. Warszawa: Wiedza Powszechna.
- Baran, M. G. (2024). *Welsh as a Fortis-Lenis Language*. Hamburg: Verlag Dr. Kovač.
- Bickford, A. C. & Floyd, R. (2006). *Articulatory Phonetics: Tools for Analyzing the World's Languages*. Dallas: SIL International.
- Boersma, P. & Weenink, D. (2024). *Praat: doing phonetics by computer [Computer Program]*.
- Chodroff, E. & Wilson, C. (2014). Burst Spectrum as a Cue for the Stop Voicing Contrast in American English. *Journal of the Acoustical Society of America*. 136 (5), 2762–72.
- Crowe, K. & McLeod, S. (2020). Children's English Consonant Acquisition in the United States: A Review. *American Journal of Speech-Language Pathology*. 29(4), 2155-2169.
- Cyran, E. (2011). Laryngeal Realism and Laryngeal Relativism: two voicing systems in Polish? *Studies in Polish Linguistics*. 6, 45-80.
- Davenport, M. & Hannahs, S.J. (2005). *Introducing Phonetics and Phonology*. London: Routledge.
- Docherty, G. J. (1992). *The Timing of Voicing in British English Obstruents*. Berlin: Mouton De Gruyter.
- Fulop, S. A. (1994). Acoustic Correlates of the Fortis/Lenis Contrast in Swiss German Plosives. *Calgary Papers in Linguistics*. 16, 55–64.
- Kasztalska, A. (2014). English in contemporary Poland. *World Englishes*. 33(2), 242-262.
- Keating, P., Mikoś, M., Ganong W. (1981). A cross-language study of range of voice onset time in the perception of initial stop voicing. *Journal of the Acoustic Society of America*. 70, 1261-1271.
- Konopska, L. & Sawicki, J. (2013). Studies of the VOT parameter in realizations of Polish voiceless and voiced plosive phonemes. *Logopedia*. 42, 101-122.
- Kopczyński, A. (1971). Degree of voicing in initial stops in educated Polish and American English. *Studia Anglica Posnaniensia*. 3, 75-79.
- Ladefoged, P. & Johnson, K. (2010). *A Course in Phonetics*. Belmont: Wadsworth.
- Ladefoged, P. & Maddieson, I. (1996) *The Sounds of the World's Languages*. Hoboken: Blackwell Publishers.
- Lisker, L. & Abramson, A. S. (1964). A Cross-Language Study of Voicing in Initial Stops: Acoustical Measurements. *WORD*. 20(3), 384–422.
- Rojczyk, A. (2010). *Temporal and spectral parameters in perception of the voicing contrast in English and Polish*. Katowice: Wydawnictwo Uniwersytetu Śląskiego.
- Rojczyk, A. & Porzuczek, A. (2012) Selected aspects in the acquisition of English phonology by Polish learners – segments and prosody. In D. Gabryś-Barker, (Ed.), *Readings in second language acquisition* (pp. 93-120). Katowice: Wydawnictwo Uniwersytetu Śląskiego.
- Schwartz, G. (2012). Initial glottalization and final devoicing in polish English. *Research in Language*. 10, 159-171.
- Schwartz, G. & Kaźmierski, K. (2019). Vowel dynamics in the acquisition of L2 English – an acoustic study of L1 Polish learners. *Language Acquisition*. 27(4), 227-254.
- Sobkowiak, W. (1996). *English Phonetics for Poles*. Poznań: Bene Nati.
- Sypiańska, J. (2021). Production of voice onset time (VOT) by senior Polish learners of English. *Open Linguistics*. 7(1), 316-330.
- Trask, R. L. (1996). *A Dictionary of Phonetics and Phonology*. London: Routledge.

- Trochymiuk, A. (2008). *Wymowa dzieci niesłyszących. Analiza audytywna i akustyczna*. Lublin: Wydawnictwo UMCS.
- Waniek-Klimczak, E. (2005). Predictors of Success in SLA Speech: Selected Observations. *Research in Language*. 3, 167-189.

### ABOUT THE AUTHOR

Michał Baran is an Assistant Professor at the Institute of Linguistics (University of Szczecin, Poland). His research interests include contrastive phonetics and phonetics and phonology of Celtic languages. He authored 9 scientific publications including the monograph “Welsh as a Fortis-Lenis Language” published by Verlag Dr. Kovač.