Psychology of Language and Communication 2024, Vol. 28 No. 1

sciendo

DOI: 10.58734/plc-2024-0006

Błażej Wieczorek and Arkadiusz Rojczyk
Speech Processing Laboratory, University of Silesia in Katowice, Poland

Phonetic imitation by young L2 learners: English VOTs for speakers of Polish

Phonetic imitation, understood as adjustment of one's pronunciation towards that of a model speaker, plays an important role in second language speech learning. The current study was intended to determine the degree to which young native Polish learners of English imitate native English models' speech, with gender as a potential influencing factor. Thirty-four participants shadowed words with both voiceless /p, t, k/ and voiced /b, d, g/ consonants as onsets, which differ in terms of voice onset time (VOT) in the two languages. Having compared VOT measurements in three tasks, no significant differences were found for /p, t, k/, suggesting no imitation effect regardless of gender, and an apparent increase of prevoicing for /b, d, g/ in the imitation tasks. Some of the null results may be attributed to high baseline values and to the stimuli's conflicting modalities. Noticeable variability in the data may have masked the true impact of imitative exposure.

Key words: voice onset time, consonants, phonetic imitation, second-language, voicing

Address for correspondence: Arkadiusz Rojczyk
Speech Processing Laboratory, University of Silesia in Katowice, Stefana Grota-Roweckiego 5, 41-205, Sosnowiec, Poland.
E-mail: arkadiusz.rojczyk@us.edu.pl

Phonetic Imitation in a Non-Native Language

Phonetic imitation, also known as accommodation (Babel, 2012), alignment (Trofimovich & Kennedy, 2014), or convergence (Pardo et al., 2012) is the process whereby a speaker adjusts phonetic features of their speech towards those of an interlocutor or a model speaker. Such alignment, or accommodation in communication (Giles et al., 1991), with a conversation partner is not limited to acoustics of speech production but has been found to emerge on other levels of language such as semantics (Katzir & Singh, 2013), syntax (Gries, 2005), vocabulary (Bogharti et al., 2016), or morphology (Dunstan, 2010). Two contrasting psychological approaches attempt to account for the pervasive effect of speech alignment observed both in laboratory conditions as well as in natural conversations. The first posits that imitation is fed by affective motives which rely on the speaker's intention to increase social proximity and personal affiliation with an interlocutor (Gregory & Webster, 1996; Tajfel & Turner, 1986). The second sees speech alignment as an automatic mechanism (Goldinger, 1997; Goldinger, 1998; Mitterer & Müsseler, 2013; Pickering & Garrod, 2004). Both mechanisms may also be at play in second language speech in that L2 learners align with native speaker models in order to increase language proximity and at the same time are led by automatic psychoacoustic reactions in forming new L2 sound categories.

Research on phonetic imitation in non-native speech is considered to be a window into the degree of formation of new sound categories and into how L1 speech habits may be temporarily bypassed after exposure to a native model. Previous studies have demonstrated that shadowing (direct imitation) after a model native speaker enables learners to imitate phonetic features that may be absent in their L1 and that have not yet been fully acquired in their spontaneous L2 speech. This effect has been found for phonetic-acoustic features such as voice onset time (VOT, Flege & Eefting, 1988), vowel duration (Podlipský & Šimáčková, 2015; Zając & Rojczyk, 2014), spectral properties of vowels (Jia et al., 2006; Jiang & Kennison, 2022; Llompart & Reinisch, 2018; Rojczyk, 2013), the lack of release in stop consonants (Rojczyk et al., 2013), or suprasegmentals (Hao & de Jong, 2016; Ulbrich, 2021). In a typical laboratory noninteractive imitation or shadowing tasks (Goldinger, 1998; Goldinger & Azuma, 2004; review in Pardo et al., 2017; Wagner et al., 2021), the learners are first presented with an L2 word list with orthographic representations and asked to read them using their natural speech habits (baseline task). Next, they hear and repeat the same words after a model native speaker (imitation task). The comparison of the two tasks reveals all changes in spectral and temporal properties of speech that may result from direct exposure to the native input.

Previous research on VOT in imitation has shown that this acoustic cue is relatively robust in imitative behavior. For example, Flege and Eefting (1988) tested imitation of English long-lag VOT stops (aspirated) by Spanish learners. The results showed that the Spanish imitators produced VOT values that were inbetween

Spanish short-lag VOTs and English long-lag VOTs for /p, t, k/. This was taken as evidence that imitation may lead to the formation of an intermediary category between the ones observed for L1 and L2. Rojczyk (2012) tested direct imitation (shadowing) of English voiceless stops (long-lag VOTs) by Polish learners whose native language has unaspirated voiceless stops. The results revealed that the learners were successful in imitating long-lag VOT values in English in direct uninterrupted shadowing. However, when the task with distraction ("read the number you see and then repeat") was introduced, the produced VOTs were significantly shorter than the ones produced in direct imitation, yet not as short as typically found for Polish. The current study differs from that of Rojczyk (2012) in that the imitators were young learners ranging in age from 10 to 11 years and not adult university students. Moreover, we tested the impact of the model speaker's gender on the magnitude of convergence, a variable that was not investigated in Rojczyk (2012).

Although the literature on phonetic imitation both in a native and non-native language is growing, little is known about the influence of imitators' age on the magnitude of convergence. Imitative behavior is especially important in children's development both in terms of language acquisition as well as in terms of general cognitive development. Therefore, it may be assumed that children may be more ready to imitate, however, the current results are inconclusive. The first study that provided evidence that children may perform better in speech imitation was by Nielsen (2014). In that study, children outperformed adults in converging with model productions that had artificially lengthened VOTs. However, in a more recent study, Paquette-Smith et al. (2022) compared Canadian English-speaking children and adults in imitations of native and non-native English speech along the VOT continuum and reported that they did not observe consistent differences between age groups. They concluded that both children and adults were equally successful imitators. These two studies demonstrate that the current knowledge on how age mediates the robustness of imitation is not only sparse but also contrasting, and more studies are needed to include or preclude age as a factor in imitation.

Previous studies have also looked into the effect of gender in terms of whether there are significant differences between male and female imitators (Bilous & Krauss, 1988; Namy et al., 2002; Pardo, 2006; Willemyns et al., 1997). Less attention has been paid to the gender of model speakers and the currently available results are sparse and strongly inconclusive. Babel et al. (2014) reported a trend for the male model voices to receive a stronger accommodation response. On the other hand, Wade et al. (2021) reported more VOT convergence to a female speaker in a similar population (i.e., L1 English speakers). However, the authors argue that this effect was likely due to the fact that the female speaker had longer VOTs. In the current study, we equalized VOTs in the male and female models by using acoustic manipulation. Therefore, unlike in Wade et al. (2021), any observed differences in imitation cannot be a result of nominal differences in VOTs between male and female speakers.

Voice Onset Time in English and Polish

Voice Onset Time, which is the time interval between the release of a stop occlusion and the onset of vocal fold vibration for the ensuing vowel (Lisker & Abramson, 1964), neatly categorizes differences in the voicing contrast implementation between English and Polish. English is an aspirating language in that voiceless /p, t, k/ have long-lag VOTs and phonologically voiced /b, d, g/ are characterized by short-lag VOTs. In other words, it may be said that English contrasts voiceless unaspirated /b, d, g/ with voiceless aspirated /p, t, k/ despite the fact that some prevoicing of /b, d, g/ is also observed in native English speech (Kessinger & Blumstein, 1997; Magloire & Green, 1999; Miller et al., 1986). On the other hand, Polish contrasts prevoiced (negative VOT values) /b, d, g/ with voiceless unaspirated (short-lag VOT values) /p, t, k/ (Keating, 1984; Keating et al., 1981). Accordingly, it may be said that the two languages differ in that English is an aspirating language and Polish is a truly voicing language in terms of the voicing contrast realization. Figure 1 shows the Polish word ”tak” /tak/ (yes) on the left and the English word ”tuck” /tʌk/ on the right. The voiceless stop /t/ is unaspirated (21 ms VOT) in Polish and aspirated (119 ms) in English. Figure 2 presents the Polish word ”daj” /daj/ (give Imp.) with prevoiced /d/ (-154 ms VOT) on the left and the English word ”die” /daɪ/ with the voiceless unaspirated realization of /d/ (26 ms VOT) on the left.

Previous research on the acquisition of English pronunciation by Polish learners has reported that they experience difficulties with the successful attainment of sufficiently long VOTs (aspiration) in production (Waniek-Klimczak, 2005) and find it challenging to categorize short-lag VOTs and long-lag VOTs as a categorical boundary for English /b, d, g/ and /p, t, k/ in perception (Rojczyk, 2011). Moreover, Polish learners tend to prevoice English /b, d, g/ following the articulatory patterns of their L1.

Figure 1. *The Polish word ”tak” /tak/ (yes) with unaspirated /t/ (left) and the English word ”tuck” /tʌk/ with aspirated /t/*

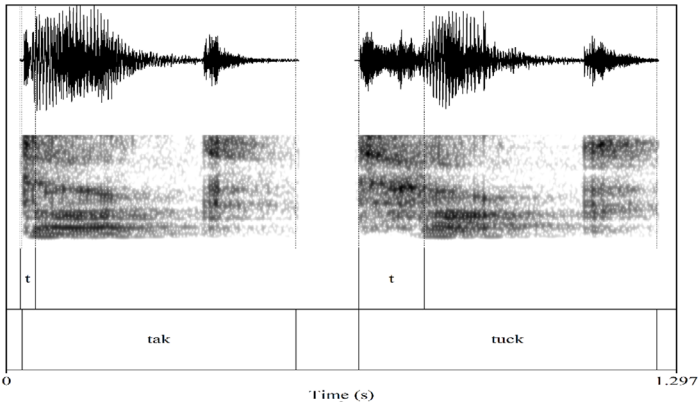
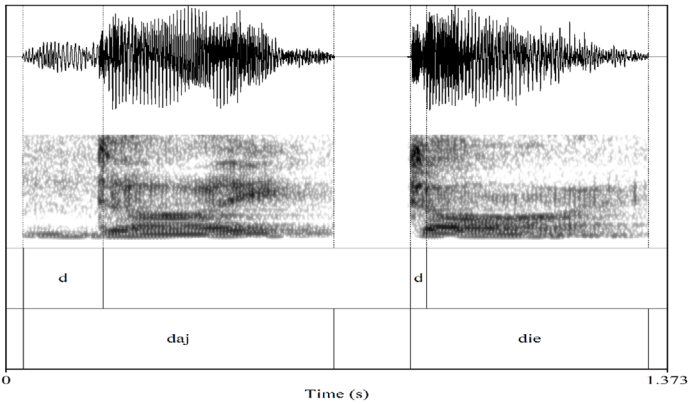


Figure 2. *The Polish word "daj" /daj/ (give Imp.) with prevoiced /d/ (left) and the English word "die" /daɪ/ with voiceless unaspirated /d/*



The Current Study

In the current study, we investigated the magnitude of phonetic imitation of the English voicing contrast by Polish young learners. We aimed to address the following questions:

1. How robust is phonetic imitation of English VOTs by Polish young learners?
2. Does the gender of the model speaker influence the degree to which young learners imitate English long VOTs?
3. Are there imitational differences between long VOTs (aspiration) for /p, t, k/ and short VOTs (no prevoicing) for /b, d, g/?

Materials

The stimuli used in the experiment included 24 English words interspersed with the same number of distractors. As can be seen in Table 1, the words were equally distributed in terms of the voicing contrast and the place of articulation of the initial plosive consonant. However, it should be noted that our main interest lay in /p, t, k/-initial words, as we believe aspiration to be more salient relative to the amount of prevoicing found in /b, d, g/-initial words, and therefore we considered it more susceptible to imitation. It was decided that the words should not be embedded in a carrier phrase to avoid potential contextual influence on VOT of /b, d, g/-initial words and to limit the exertion imposed on the participants. Another factor determining the word selection was their relative ease of pronunciation.

Four professional independent voiceover artists were hired as model speakers for the recordings to be used in the imitation tasks. They were two male and two female native speakers of Standard Southern British English. The choice to include two representatives of each gender was motivated by the need to increase the

Table 1. *The stimuli used in the experiment*

Onset type	Voiceless	Voiced
labial	part, pet, pig, pot	bath, bet, big, boy
alveolar	tell, tip, top, two	desk, dip, do, dog
velar	cake, cap, car, kid	gap, gasp, get, give

validity of the results, which were supposed to show the participants' potential tendency to imitate one of the genders more than the other, as opposed to showing their bias towards a particular speaker's overall voice quality. The model speakers were instructed to read out the words in isolation, as naturally as possible, using the falling intonation pattern, much like how the recordings for foreign language course books are made.

For the sake of limiting the effect of confounding variables potentially influencing the degree of imitation during the presentation of the stimuli, such as inconsistent loudness of the model recordings, the words' duration, and their VOT in particular, the four instances of each model word were normalized. First, the intensity of all recordings was set to 68 dB, which was the highest value possible without the risk of any of them being clipped. Second, any instances of prevoicing for /b, d, g/-initial words were cut out – such instances were found only in the case of the first male model's recordings. Third, word duration was equalized and set to word-specific average, which was particularly motivated by the second female model speaking much faster than the others. Fourth, the VOTs in the four instances of each of the /p, t, k/-initial words were set to their mean value, thus ensuring that potential differences in the degree of imitation between the genders would not be the result of the model recordings simply being different in terms of VOT. Predictably, the fourth step modified the total duration of each of the four instances, slightly altering the effect of the third step, but it was clearly a necessary yet negligible cost. Figures 3 and 4 illustrate the third and the fourth steps of the normalization process of the word "pig."

Participants

Thirty-four native Polish primary school students were asked to voluntarily participate in the study. These included 24 female and 10 male students aged 10–11 years. Being in their fifth school year, all of them had learnt English for at least four years, and most of them displayed a uniform proficiency level in English corresponding to the A2 level on the CEFR scale, as evidenced by their grades obtained through the administration of both oral and written tests. It was ascertained that none of the participants showed any signs of speech or hearing disorders.

Figure 3. *Word duration normalization of the four instances of the word "pig"*

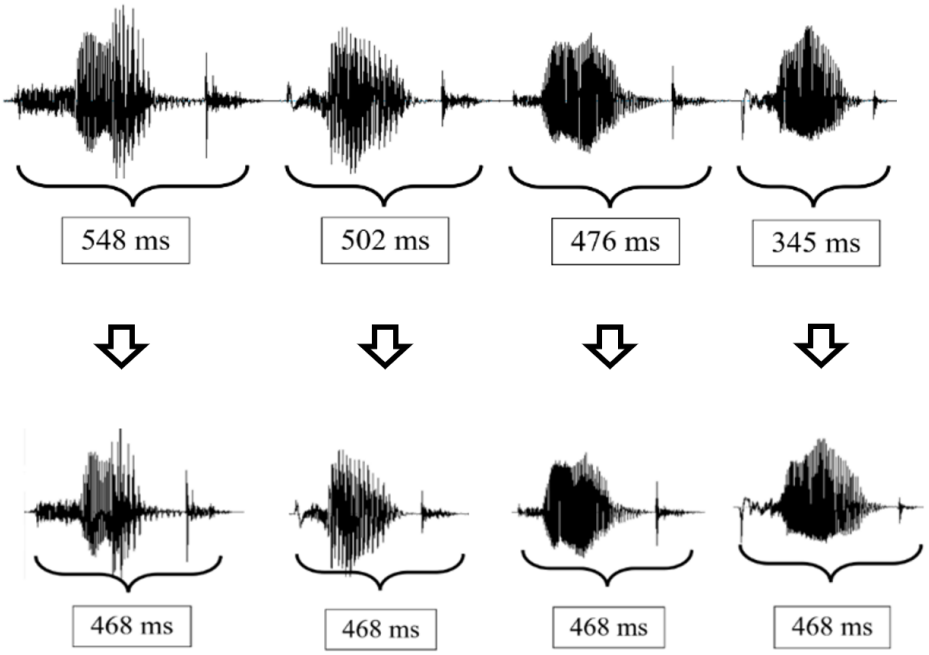
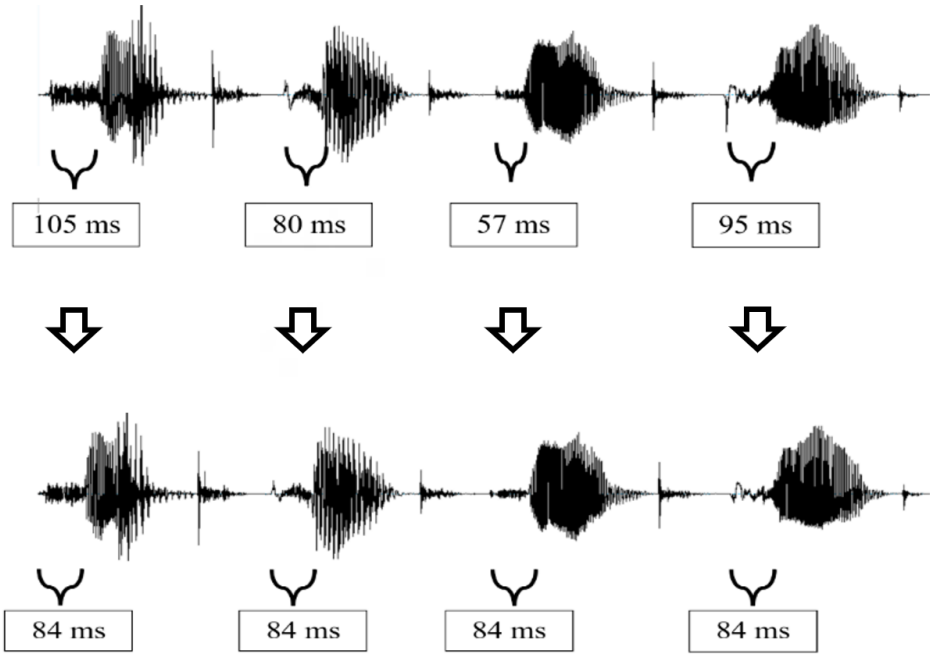


Figure 4. *VOT normalization of the four instances of the word "pig"*



Procedure, Recordings, and Measurements

The experiment took place at the school premises during the participants' classes. Each student was recorded individually in a quiet room, and each recording session took approximately 20 minutes. The participants were seated in front of a computer screen and were asked a few basic questions in English in order to activate the desired language mode, upon which they were presented with three tasks. First, they were to read orthographic representations of 48 words (including 24 distractors), presented in isolation on a computer screen in random order for each participant. This was done to establish their baseline VOT values in English. The second task was an imitation task in which half of the participants were instructed to repeat the same words approximately 500 ms after one of the two male models, and the other half – after one of the two female models. The third task was analogous to the second one, with the exception that the recordings, randomized once more, involved a model of the opposite gender. Thus, all participants imitated both genders in a counterbalanced fashion. Additionally, every other participant listened to a different representative of a given gender, which means the odd numbered participants listened to the first male model and the first female model only, while the even-numbered ones – to the other two models. The interstimulus interval was self paced for all tasks, but it did not deviate significantly across participants.

The participants' productions were captured using Samson Q2U dynamic microphone at the sampling rate of 48 kHz and 16-bit quantization. The stimuli to be imitated were presented aurally through PreSonus Eris E3.5 studio monitors. Both audio devices were connected to the 3rd Focusrite Scarlett audio interface.

Data Analysis

A total number of 2448 items (34 participants × 24 words × 3 tasks) underwent acoustic and statistical analysis. Having inspected the participants' recordings both visually and aurally, all tokens were found suitable for analysis, including a few cases which exhibited exceptionally long VOT values but still sounded natural. The VOT measurements were taken from waveform and spectrogram displays in Praat (Boersma & Weenink, 2021). Specifically, in the case of /p, t, k/-initial words, what was of interest was the time interval between the first signs of the plosive release burst and the onset of voicing of the following vowel, the latter point being evidenced by clear regularity both in the waveform and in the vowel's formant structure (see "tuck" in Figure 1). The VOTs obtained in the three tasks were averaged over the participants' productions and a repeated measures analysis of variance (ANOVA) was performed to determine if the three means were different. The main factor was within-subjects (task), with three levels (baseline, male model imitation, and female model imitation). To account for potential variability in word duration, an additional ANOVA was carried out, treating VOT as a

proportion of word duration. Furthermore, the place of articulation of the initial voiceless plosive (/p, t, k/) was included as a factor to determine its significance.

As far as /b, d, g/-initial words are concerned, the proportion of prevoicing instances was determined in the three tasks. Additionally, mean VOT values were compared across the three tasks (similarly to /p, t, k/-initial words), although, for the sake of simplicity, only cases with negative values in all three tasks were taken into account, which constituted almost 56.6% of cases (231 out of 408). The negative values were measured as time spans encompassing regular periodicity before the release. Any instances of gaps in voicing immediately preceding the release were included in the calculation of negative VOTs. Apart from absolute negative VOT values, their proportions of word duration are also reported.

Analysis and Results

Figure 5 shows mean VOT values for /p, t, k/ with their standard deviation, obtained in the three tasks: baseline word reading, male model imitation, and female model imitation. As can be seen, the two imitation tasks exhibited slight increases compared to the baseline: from 67.4 ms to 69.8 ms and from 67.4 ms to 67.8 ms, respectively. Despite a visible VOT increase in the case of the male models, analysis of variance revealed no significant effect of task, $F(2, 814) = 1.527$, $p = .218$. This suggests that not only did model speaker gender have no effect on the degree of imitation of words starting with voiceless plosives, but also that no imitation may have taken place whatsoever, as reflected in only marginal VOT increases compared to the baseline.

Figure 6 illustrates the participants' three VOT distributions obtained in the three respective tasks, along with their mean VOT values (colored vertical lines)

Figure 5. *The participants' mean VOT values of /p, t, k/ together in the three tasks*

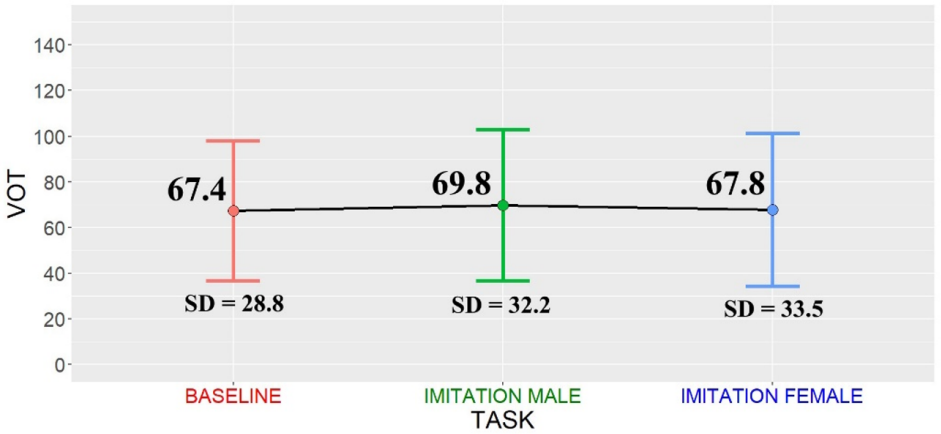
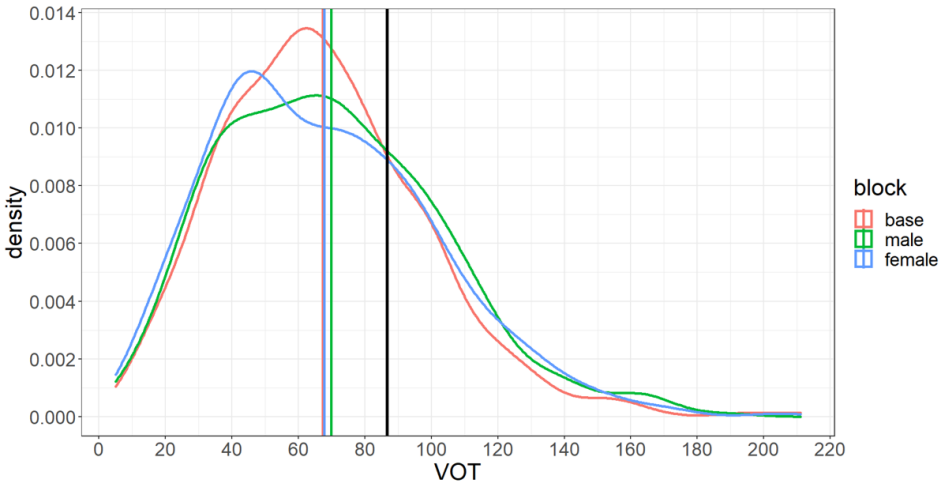


Figure 6. *Three VOT distributions (baseline, male-imitation, female-imitation), along with mean VOTs of the participants (colored) and the models (black)*

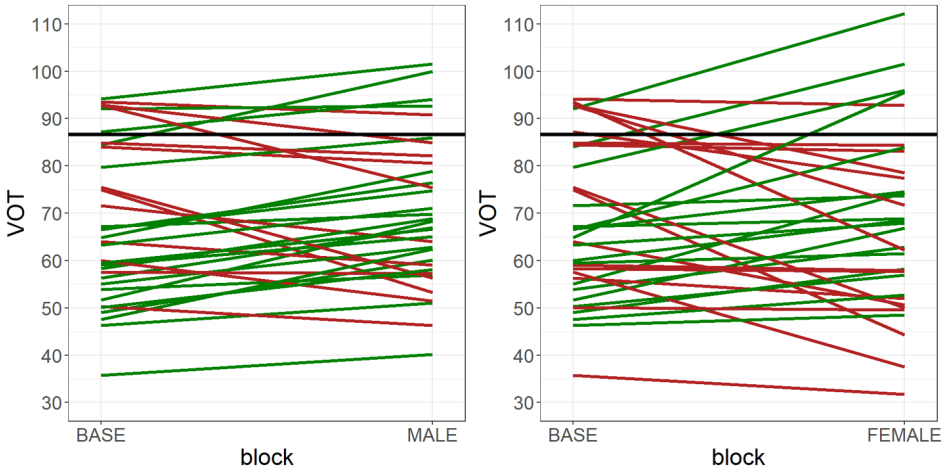


and the models' mean VOT value (black vertical line). Despite some differences in the overall shape of the three distributions, there are no discernible shifts in either direction, regardless of the imitation task. Of interest were the participants' already quite substantial VOTs obtained in the first task. It was found that in as many as 81 out of 408 cases (19.9%), the participants exceeded the models' absolute VOT values in the baseline task. Still, on the whole, their mean VOT was below the mean VOT of both male and female model speakers, equal to 86.7 ms. Interestingly, in the two imitation tasks, the number of times the participants exceeded the models' VOT values increased to 104 (25.5%) in the case of a male model and to 96 (23.5%) – in the case of a female one.

Figure 7 provides a more in-depth picture of the data by considering regression lines, representing predicted VOT changes from the baseline to the male model imitation task (left) and the female model imitation task (right) for each of the participants. The black horizontal line indicates mean model values. The first thing to notice is how variable the participants' baselines were. The intercepts roughly corresponded to each participant's actual mean baseline VOT values, ranging from 36 ms to 94 ms, with many rather uniformly distributed values in between. More importantly, it should be noted that as many as 22 participants had positive (green) slopes in the male model imitation task, and 18 had such slopes in the female model imitation task, suggesting the majority of the participants did increase their absolute VOT values to varying extents.

To confirm the apparent lack of imitation, it was deemed necessary to treat VOT as a proportion of word duration. As Table 2 indicates, the duration of /p, t, k/-initial words was found to increase significantly in the imitation tasks, $F(2,$

Figure 7. *Mean VOT changes for each participant in the male model imitation task (left) and the female model imitation task (right)*



814) = 29.12, $p < .001$, from 485.6 ms to 511.6 ms (male model imitation) and to 519.4 ms (female model imitation), despite the models' average of 482.8 ms. The increase in word duration and the apparent lack of increase of absolute VOT values seem to have negatively shifted the VOT proportions, in that the participants' baseline average of 14.2% decreased to 13.8% in the male model imitation task and to 13.2% in the female model imitation task, with the effect of task reaching significance, $F(2, 814) = 6.961$, $p = .001$. After applying the Bonferroni correction, the only significant difference was found between the baseline and the female model imitation task, $t(407) = 3.751$, $p < .001$. The decrease in ratios is surprising considering the models' average ratio of 18.2%, although, as was the case with absolute VOT values, the participants did exceed the model ratios in as many as 89 out of 408 cases (21.8%) in the baseline. This count increased in the male model imitation task to 101 (24.8%) and decreased in the female model imitation task to 87 (21.3%). The distributions of relative VOTs obtained in the three tasks resembled those presented in Figure 6.

Table 2. *Absolute mean VOT and word duration values with their ratio in the three tasks for /p, t, k/-initial words, together with p values and model values*

	Mean VOT	Mean word duration	VOT proportion
Baseline	67.4 ms	485.6 ms	14.2%
Male model imitation	69.8 ms	511.6 ms	13.8%
Female model imitation	67.8 ms	519.4 ms	13.2%
p	.218	.001 ms	0.001
Model values	86.7 ms	482.8 ms	18.2%

Figure 8 shows individual participants' slopes that represent predicted VOT proportion changes from the baseline task to one of the imitation tasks. Again, considerable variability is noticeable, with the actual mean VOT proportions ranging from approximately 8% to 20% in the baseline task. Exactly half of the participants (17) increased their VOT word duration ratios in the male model imitation task (left), and only 13 of them did so in the female model imitation task (right).

Further analysis was carried out to determine the degree of imitation depending on the place of articulation of the plosives. Figure 9 shows that while words beginning with alveolar and velar plosives showed slight increases from the baseline to imitation tasks, the mean absolute VOT differences were still found to be insufficient to reach significance, regardless of the place of articulation. Surprisingly, both imitation tasks in the case of /p/-initial words seem to have displayed a slight (insignificant) decrease from 53.1 ms to 51.8 ms and 49.2 ms, respectively, even though the models' mean VOT of these words amounted to 79.9 ms. Among the three places of articulation, the biggest, although still insignificant, increase was observed in the case of /t/-initial words. This was particularly the case in the male model imitation task, where the participants' absolute VOT values rose from 62.7 ms to 67.9 ms. This observation seems reasonable, considering the models' relatively high average of 95.0 ms for alveolar plosives. Another interesting observation is that the participants' mean VOT for /k/-initial words exceeded those of the models' (85.2 ms) in all tasks, including the baseline. When treated as proportions of word duration, only the VOTs in /p/-initial words showed apparent significance, $F(2, 270) = 3.472$, $p = .033$, across the tasks, particularly between the baseline and the female model imitation task, $t(135) = 2.945$, $p = .004$. Overall,

Figure 8. *Mean VOT proportion changes for each participant in the male model imitation task (left) and the female model imitation task (right)*

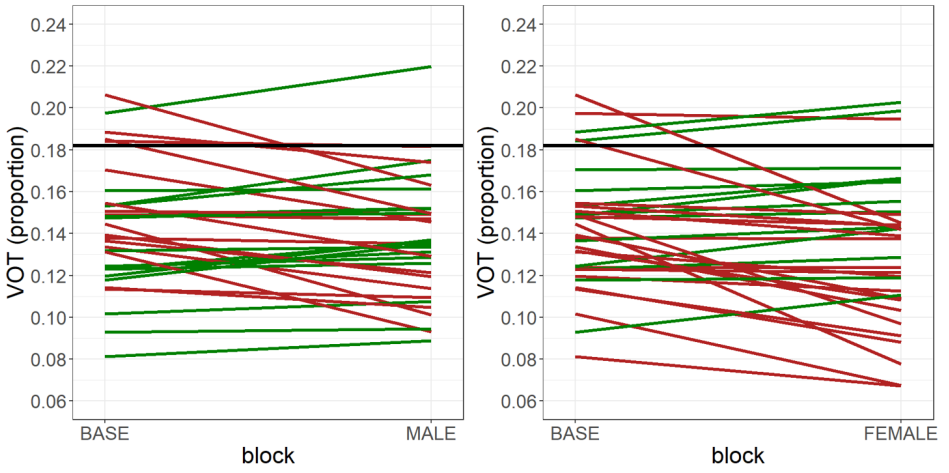
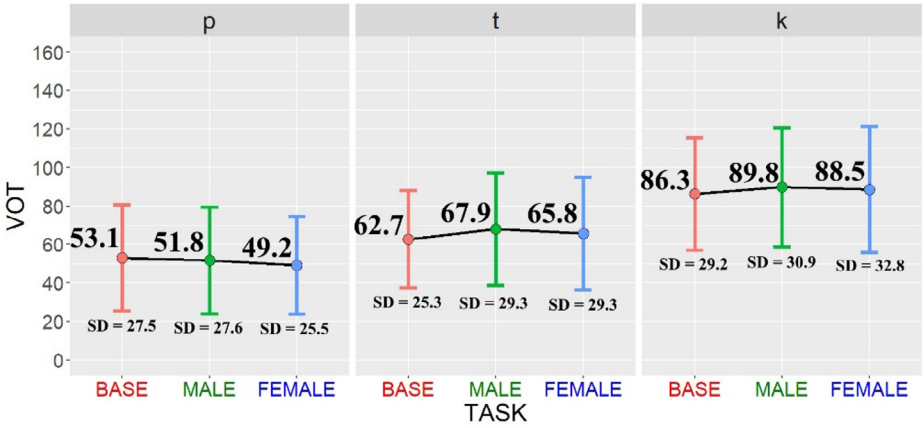


Figure 9. *The participants' mean VOT values of /p, t, k/ separately in the three tasks*



the mean VOT ratios showed consistent, small decreases in both imitation tasks for all places of articulation, namely:

- from 10.5% to 9.9% and 9.3% for /p/ (models' 16.5%),
- from 14.7% to 14.5% and 13.8% for /t/ (models' 20.6%),
- from 17.5% to 16.9% and 16.6% for /k/ (models' 17.6%).

Furthermore, it was found that the participants who exceeded the model VOTs in the baseline did so mostly in the case of velar onsets (in 40% of cases).

The analysis of /b, d, g/-initial words also yielded interesting results. Figure 10 illustrates how often the participants manifested prevoicing in the three tasks. As can be seen, out of all 408 instances of /b, d, g/-initial words (12 words × 34 participants) 319 (78.2%) exhibited negative VOT values in the first task. Despite all the model words having positive VOT values, the count was only marginally lower in the imitation tasks: 311 (76.2%) in the male model imitation task and 318 (77.9%) in the female model imitation task. The largest number of prevoicing drops was observed in the case of /b/-initial words, with 113 out of 136 cases (83%) in the baseline to 104 (76.5%) in both imitation tasks. There was an opposite tendency in the case of /g/-initial words, where the number of baseline prevoicing instances of 94 (69.1%) increased to 101 (74.2%) and 104 (76.4%) in the respective imitation tasks.

As far as the participants' VOT values of /b, d, g/ are concerned, for the sake of ascertaining statistical significance, only those cases were examined in which a given participant displayed a negative VOT for a given word in all three tasks (231 cases out of 408). Positive VOT values were not analyzed as there were only 26 of analogous cases, most of which came from only two participants. The mean negative VOT values are shown in Figure 11. Interestingly, while the substantial mean negative VOT value for the baseline task in the order of −110.3 ms was

Figure 10. *The participants' prevoicing counts in the three tasks*

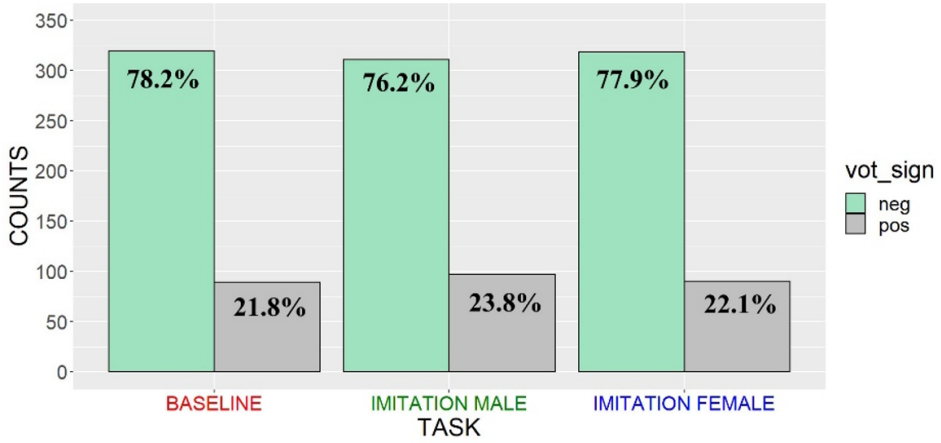
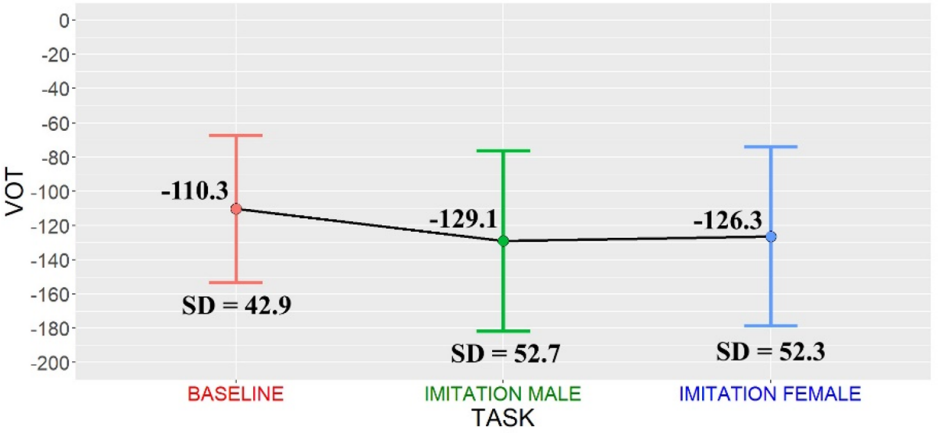


Figure 11. *The participants' mean VOT values of /b, d, g/ in the three tasks*



expected, the participants exhibited even larger absolute VOT values in the two imitation tasks. The mean values of −129.1 ms and −126.3 ms significantly differed from −110.3 ms obtained in the baseline task, $F(2, 460) = 14.07$, $p < .001$. Including all negative values (unequally distributed across the three tasks) did not seem to considerably alter the above tendency, with the three means amounting to −108.9 ms, −124.3 ms, and −121.3 ms, respectively. Initially, the above analyses seem to show that the participants may have increased their tendency to prevoice in terms of duration as a result of exposure to the model speakers whose VOTs were consistently confined within the positive range of 0-20 ms. However, it was

found that word duration increased significantly again, namely, from 562.2 ms in the baseline task to 628.0 ms and 616.5 ms in the imitation tasks, $F(2, 460) = 45.14$, $p < .001$. It was revealed that the prevoicing and word duration ratios did not change, amounting to 19.8%, 20.5%, and 20.3%, respectively across the three tasks, $F(2, 460) = 1.118$, $p = .328$. These patterns were consistent for all places of articulation, apart from the voiced velar plosives, whose VOT ratios increased quite considerably, i.e., from 16.7% to 18.4% and to 18.8%.

General Discussion

The current study was intended to determine the degree to which English VOT values are imitated by young native Polish learners of English, taking into account model speaker gender as a potential factor influencing their tendency to acquire the English voicing characteristics for voiceless and voiced plosive consonants. To that end, 34 primary school students were asked to imitate 24 monosyllables with /p, t, k, b, d, g/ as onset consonants, prerecorded by two male and two female model speakers. Considering the differences between the two languages, English exhibiting longer VOTs for /p, t, k/ and a lower tendency to prevoice /b, d, g/ than Polish, it was hypothesized that the participants' VOTs would shift towards the values found in English as a result of imitative exposure to native English speakers. Moreover, we tested the impact of the model speaker's gender on the robustness of imitation considering the fact that the results from previous studies are largely inconclusive.

The main hypothesis was not confirmed and the gender effect was not found for nominal values. The participants' VOT values for imitated /p, t, k/-initial words did not prove to be significantly longer when compared to the baseline word reading task. Consistently with Wade et al. (2021), we found no evidence for gender having an effect on the degree of imitation, apart from only marginally higher VOT increases observed in the male model imitation task. When treated as proportions of word duration, their VOTs were found to be shorter in the imitation tasks, particularly when the model was female. This was likely the result of their meager shifts in absolute VOT values lagging behind the total word durations, which increased significantly despite the models' comparable word durations to the participants' baseline. Out of the three places of articulation, the only noticeable increases in absolute values were in the case of /t/-initial words, while the opposite was found for /p/-initial words, which exhibited VOT drops both in the absolute and relative sense. This seems to correlate with the model values, which were the highest for the former, and the lowest for the latter.

Although not statistically significant, it should be noted that a more in-depth analysis of the results revealed that many of the participants did increase their VOTs in the imitation tasks to varying extents. In the male model imitation task, over half of the participants increased their absolute VOT values, and exactly half increased their VOT word duration ratios. In the female model imitation task,

positive slopes were less frequent. While approximately half of the slopes indicated increases in the case of absolute values, there was a considerable number of negative slopes for VOT proportions. On the whole, as Figures 7 and 8 illustrated, the VOT drops, represented by (often steep) red slopes, seemed to offset the VOT increases, resulting in overall insignificant results. It must also be observed that our data showed a great deal of variability in terms of baseline VOTs across different participants, which should restrict us from drawing definite conclusions. The considerable variability in the baseline may be due to the participants' age. For example, Whiteside et al. (2003), who compared VOTs of English children aged between around 6 and 12 years old, observed a pattern of increased VOT stability as a function of age. The effect was attributed to the maturation of motor speech skills of children approaching adolescence. Yu et al. (2015) compared English VOTs of young children, teenagers, and adults and revealed a similar pattern, with continually decreasing VOT variability until the age of 16. The same may be applicable to Polish children and teenagers. This would suggest that while our participants (10-11-year-olds) may have achieved a certain degree of VOT stability, our findings as well as the above studies indicate that only partial maturation had taken place in this regard.

One of the reasons why, overall, our participants did not significantly extend their VOTs in /p, t, k/ in the imitation tasks may be connected with their already long baseline VOTs. It should be noted our participants exceeded the models' both absolute and relative VOTs quite frequently in the baseline task, that is, around 20% of the time, although this number was almost consistently slightly larger in the imitation tasks. As noted by one of the reviewers, some studies, for example, Nielsen (2011), Kim and Clayards (2019), and Wade (2020), have suggested that speakers may fail to imitate, or even diverge from model speakers when their speech is already similar to the models' speech. However, we would argue that our data does not reveal this pattern, since the models' average VOTs for the majority of the stimuli were long enough to leave some room for VOT adjustments. Approximately 40% of excessive participant values came from /k/-initial words, and indeed, their mean VOT actually exceeded that of the models' mean for the velar plosive. While imitation in this case would not be expected, it would be in the case of /p/-initial words. With labial plosives, the number of excessive participant values was marginal, and the participant model distance seemed sufficiently large, that is, approximately 27 ms. Despite this distance, the participants paradoxically shortened their VOTs for /p/. What is more, when considered as word-duration proportions, model VOTs left enough room for imitation in the case of both /p/ and /t/, in the order of around 6 percentage points. It is also worth recognizing that, as the individual participants' slopes indicate, even some of those with high baseline values managed to increase VOTs further in the imitation tasks.

Still, assuming the high baseline values are responsible for the lack of observed imitation for /p, t, k/, the reason for this should not be regarded simply in terms of purportedly insufficient participant model distances. Our participants are ultimately

native speakers of Polish, a language which is characterized by short-lag VOT values for voiceless plosives, as reported earlier. Therefore, we speculate that the high initial VOT values are due to our participants already shifting their VOTs in the baseline by sole virtue of being instructed to produce words in English. This shift from their baseline L1 values to baseline L2 values may have consequently exhausted their potential to further extend the values of interest in the imitation tasks. For native English speakers, long-lag VOTs are natural and they can be considered as more flexible in shifting VOT within the long lag range than native Polish speakers. In short, our participants' high baseline values may be responsible for the lack of imitation, not simply due to a potentially reduced participant-model distance, but rather due to them having exerted their capacity to further extend their already lengthened baseline VOTs. However, this explanation does not agree with the fact that some of the participants managed to increase their VOTs further, despite high baseline VOTs.

The potentially lengthened baseline values were an unexpected element in our study, considering our Polish participants exhibited average VOTs resembling even those reported for some native English speakers. For example, in Nielsen (2013), native English children of similar age manifested comparable or even shorter mean VOT values, particularly in the case of the velar plosive (see Table 3). Moreover, the current participants seemed to have done even better than adult native Polish advanced learners of English, as suggested by the latter's overall shorter baseline VOT values found in Rojczyk (2012). However, considering that young children tend to produce longer and more variable VOT values than adults do (Yu et al., 2015), we would advise caution in drawing such conclusions. Still, relatively short Polish VOT values of Polish children of comparable age reported in Lorenc (2012) seems to suggest that the current participants may have developed the ability to shift their VOT boundary towards values typical for English /p, t, k/ categories. However, such a claim would have to be supported by measurements of Polish voiceless stops coming from our participants. Moreover, the overall impression of the experimenters is that the participants did not sound particularly native-like in the recording sessions. Furthermore, fifth grade primary school students in Poland do not typically display native-like performance in other respects of L2 proficiency. The question therefore remains as to why they would display seemingly native-like competence in this particular regard, that is, aspiration – a feature which is unlikely ever explicitly taught in schools.

Table 3. *Comparison of mean baseline VOT values for /p, t, k/ across studies*

	Current study	Nielsen (2013)	Rojczyk (2012)	Lorenc (2012)
Subjects	Polish children	English children	Polish adults	Polish children
Stimuli	English	English	English	Polish
VOT – /p/	53.1 ms	62 ms, 58 ms	56 ms	22 ms
VOT – /t/	62.7 ms	-	59 ms	29 ms
VOT – /k/	86.3 ms	75 ms	75 ms	54 ms

In the first of their experiments, examining delayed imitation among native English children and adults, Paquette-Smith et al. (2022) also observed particularly high baseline VOT values for their participants, making imitative performance difficult to observe. They enumerated three potential reasons for the lack of imitation: hyperarticulation in the baseline task, the use of high frequency words, and the employment of a delayed imitation paradigm in the sense of Nielsen (2014). In the case of hyperarticulation, comparable VOTs obtained in the baseline and the imitation task could be considered different in terms of which factor elicited them. In other words, given VOT values in the baseline could be mainly due to hyperarticulation, and seemingly similar values in an imitation task could at least party be due to imitative exposure itself. We cannot dismiss this explanation in the current study, as our participants may have indeed tried to make an increased effort to pronounce the words, especially in the first task. However, we would not rely heavily on this to account for our findings. Had hyperarticulation been responsible for high baseline values but not so much for the post-task values, we would probably have observed longer word durations in the reading (baseline) task, and this was not the case. As far as word frequency is concerned, it is known that high frequency words tend to be imitated less faithfully than low frequency ones (Goldinger, 1998), and indeed, the majority of our stimuli belong to the first category. Despite that, we would not consider these words to be high in frequency for non native English speakers in the same sense as they are to the native ones. To the latter group, the same words may easily be considered as much more frequent. Consequently, without entirely dismissing its effect, we do not believe our baseline values to have been significantly impacted by word frequency. Lastly, having attributed the lack of imitation to the delayed nature of the imitation task, Paquette Smith et al. (2022) exposed them to the immediate shadowing task. Contradictorily to our findings, they found robust imitation among both children and adults.

There is an important methodological difference between our study and similar studies dealing with VOT imitation (Nielsen, 2014; Paquette Smith et al., 2022; Shockley et al., 2004). In the quoted studies, the stimuli's VOTs were justifiably lengthened in order to create a marked participant-model difference. In our case, it was warranted to omit this step as our participants were not native English and there was already an expected participant-model difference, described earlier. The purpose of our study was to verify if young Polish learners of English imitate English words, produced as naturally as possible, with only a limited degree of manipulation required for stimulus normalization. It was also pointed out to us that the way in which the stimuli were presented, that is, through their orthographic form, may help explain the null results. Some of the studies quoted above employed a picture-naming task to elicit the participants' productions, but we did not find this method applicable in the current study. The reason for this lies, again, in that our participants were not native English speakers, and they may have had problems naming the pictures. The increased cognitive load involved in

producing these words could have made the data less reliable. Having the stimuli in their orthographic form left little doubt as to what the words actually were.

The downside of this was that the presence of spelling in the imitation tasks may have overridden the auditory dimension of the stimuli heard. Here, it should be noted that Polish is characterized by a relatively strong correspondence between its spoken and its written form (Jassem, 1981), which results in Poles' strong reliance on given letters unambiguously representing specific sounds. Polish learners of English may be suspected of transferring this reliance when speaking English, especially when the speech is read out. For example, because they are so used to the letter <i> representing the sound /i/ in their native language, they will readily associate the same letter in English words with the Polish sound /i/. Piske et al. (2002) reached similar conclusions for native speakers of Italian, whose L2 English speech was judged to contain orthography-induced L1 vowel substitutions. Similarly, when seeing the <p> letter, our participants would be expected to produce the unaspirated (Polish) variant, even if the word was English. We acknowledge that in consequence, the presence of spelling may have encouraged them to follow their native pronunciation habits to some extent, potentially limiting the influence of the auditory modality. However, it must be recognized that a similar, if not a greater effect would have been expected in the baseline task. In other words, their baseline VOT values should have been more Polish like (lower) in the presence of the only modality available, that is, the stimuli's orthographic representation.

As far as the /b, d, b/-initial words are concerned, similarly to the words starting with voiceless onsets, a clear imitation pattern was not observed. We expected the imitative effect to take the form of either a reduced number of prevoicing instances or as shorter absolute or relative values of negative VOT in /b, d, g/-initial words. Contrary to our hypothesis, negative VOTs were only marginally less frequent in the imitation tasks than in the word-reading task. Additionally, on the whole, in cases where prevoicing was observed in all three tasks, its absolute duration showed a paradoxical increase in both imitation tasks. While one may have suspected prevoicing to be too subtle a cue to allow the participants to reconstruct it from the auditory stimuli, they were not expected to show even more pronounced divergence from the model speakers in the imitation tasks. It should be recalled that the model recordings involved stimuli with invariably positive VOTs. However, as it was in the case of words with voiceless onsets, the participants were found to have produced longer words in the imitation tasks. Consequently, while their absolute negative VOT values increased, the same values, when treated as proportions of word duration, did not change significantly.

Unlike how it may have been the case with the stimuli with voiceless onsets, the baseline participant model distances for /b, d, g/-initial words were definitely not an issue. Most of the participants displayed an overwhelmingly high proportion of negative VOT values in the baseline task. It may be speculated that imitation in the case of /b, d, g/-initial words, manifested either as lower negative

VOT counts or shorter lead VOTs, did not happen because it was suppressed on account of the contrastive function of prevoicing in Polish. In Polish, the presence versus absence of prevoicing determines the voicing status of the onset plosive consonant. It seems likely that the participants did not imitate the lack of prevoicing, as it would infringe on the phonemic boundary between /p/ and /b/, /t/ and /d/ or /k/ and /g/. Mitterer and Ernestus (2008) found that native Dutch speakers, who were asked to imitate Dutch /b, d/-initial pseudowords, reproduced presence versus absence of prevoicing but not the amount of prevoicing. This was attributed to the fact that the former is phonologically relevant in Dutch while the latter is not. We think our native Polish participants did not imitate the absence of prevoicing for the same reason the Dutch speakers did, that is, to maintain the phonemic contrast applicable to their native language. The Dutch speakers were presumably exposed to the auditory stimuli only. Thus, it is not surprising they would react accordingly to the shifts across the category boundary between the prevoiced and unprevoiced items. Our native Polish participants had additional access to the stimuli's orthographic representations, which may have had a similar effect to the one described earlier. More specifically, seeing words starting with the letters <b>, <d>, or <g> just before being exposed to the auditory stimuli can be assumed to have already activated the respective phonemic categories of /b/, /d/, and /g/, along with their phonetic manifestations, including prevoicing. This prior activation may have rendered the subsequent exposure to the auditory stimuli, with their conflicting acoustic information (the lack of prevoicing), ineffective. Also, the results by Mitterer and Ernestus may also explain why the participants did not at least reduce the amount of prevoicing in the imitation tasks, the reason being that it is not phonologically relevant, in either Dutch or Polish.

However, the above interpretation does not explain why the participants exhibited significantly longer absolute lead VOTs, that is, longer prevoicing durations, in the imitation tasks. Obviously, the explanation partly lies in the participants following the models' word durations, which were longer than the former's baselines. Still, VOT ratios were also slightly higher in the imitation tasks, although not significantly. The participants appear to have been making an enhanced effort to imitate the tokens correctly, and they may have paradoxically reinforced their L1 prevoicing tendencies. This may be related to the fact that linguistic stress may lead to the lengthening of prevoicing, as was found, among others, by Simonet et al. (2014), who determined lexical stress to have such an effect on Spanish /d/. A similar observation was made by Malisz and Żygiś (2015) in the case of Polish /b, d, g/, where negative VOT values were more evident under lexical stress and sentence focus. Contrary to the current study, both of these studies also reported VOT lengthening for voiceless plosives under the same circumstances, which was hardly observed in our case. Moreover, if we assume our participants had made an increased effort for clear speech, we should have expected a similar effort made in the baseline task. The result is particularly puzzling, considering the /b, d, g/-initial words were interspersed with /p, t, k/-initial words, which did

not show signs of hyperarticulation in the imitation tasks apart from relatively modest increases in word durations. In short, we do not find a satisfactory explanation for increases in prevoicing durations other than the increases in word durations themselves.

Conculsions

The purpose of the current study was to determine the degree to which young native Polish learners of English imitate the voicing characteristics of onset stop consonants in English, taking into account the gender of the model speaker. The participants in our study did not show a clear imitation pattern for either /p, t, k/-initial or /b, d, g/-initial words. Their VOT, understood in both absolute and relative terms, did not show significant changes from the baseline to any of the imitation tasks. Although the values were negligibly higher in the case of male model speakers, overall, the model's gender did not have a significant effect on the degree of imitation. The lack of imitation in the case of stimuli with voiceless onsets may be attributable to positive VOT shifts that the participants are likely to already have made in the baseline task. These shifts potentially exhausted their potential to extend their VOTs further in the imitation tasks. In the case of words with voiced onsets, imitative performance was possibly suppressed by the presence of the stimuli's orthographic form, which was presumably responsible for the activation of the participants' native phonemes. To avoid encroaching on the voiced voiceless phoneme boundary, the participants were unaffected by the auditory modality of the stimuli (without prevoicing) and retained prevoicing for /b/, /d/, and /g/. Overall, despite the lack of significant VOT shifts in the imitation tasks, it should be noted that our data showed considerable, possibly age-induced, variability. The fact that many of our participants did increase their VOTs for voiceless onsets to varying extents suggests that our main finding should not be regarded in absolute terms.

Undoubtedly, more insight into the imitative behavior of speech exhibited by foreign language learners is needed. Future studies may consider a modified imitation paradigm, in that, for example, only auditory stimuli are presented in one of the tasks, thus isolating the influence of orthography, which may cause participants to cling to their native pronunciation habits in the imitation tasks. It may also be worth modifying the instructions given to the participants by encouraging them to aim to sound as similar to the models as possible. At the same time, it is recommended to consider enlisting participants that are balanced and accounted for in terms of age and proficiency level, which may help reveal potential patterns of variability across L2 imitators.

Conflict of Interest Disclosure

The Authors declare no conflicts of interest.

Funding

This research was supported by the National Science Centre Poland grant Phonetic imitation in a native and non-native language (UMO-2019/35/B/HS2/02767) awarded to the second author and by the funds granted under the Research Excellence Initiative of the University of Silesia in Katowice.

Research Ethics Statement

The informed consent was obtained from the school authorities.

Data Availability Statement

The authors confirm that the data supporting the findings of this study are available on the Open Science Framework: https://osf.io/587cy/?view_only=575eea0734234a65ba88be4d28fd4768

Authorship Details

Błażej Wieczorek: research concept and design, collection and/or assembly of data, data analysis and interpretation, writing the article, critical revision of the article. Arkadiusz Rojczyk: research concept and design, data analysis and interpretation, writing the article, critical revision of the article, final approval of the article.

References

Babel, M. (2012). Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics, 40*, 177–189. https://doi.org/10.1016/j.wocn.2011.09.001

Babel, M., McGuire, G., Walters, S., & Nicholls, A. (2014). Novelty and social preference in phonetic accommodation. *Laboratory Phonology, 5*(1), 123–150. https://doi.org/10.1515/lp-2014-0006

Bilous, F., & Krauss, R. (1988). Dominance and accommodation in the conversational behaviours of same- or mixed-gender dyads. *Language and Communication, 8*(3–4), 183–194. https://doi.org/10.1016/0271-5309(88)90016-X

Boersma, P., & Weenink, D. (2021). Praat: doing phonetics by computer [Computer program]. Version 6.1.57, retrieved November 2021 from http://www.

praat.org/

Bogharti, R., Hoover, J., Johnson, K. M., Garten, J., & Dehghani, M. (2016). Syntax accommodation in social media conversation. In A. Papafragou, D. Grodner, D. Mirman, & J. C. Trueswell (Eds.), Proceedings of the 38th Annual Conference of the Cognitive Science Society (pp. 1463–1468). Cognitive Science Society.

Dunstan, S. B. (2010). Identities in transition: The use of AAVE grammatical features by Hispanic adolescents in two North Carolina communities. *American Speech, 85*(2), 185–204. https://doi.org/10.1215/00031283-2010-010

Flege, J. E., & Eefting, W. (1988). Imitation of a VOT continuum by native speakers of English and Spanish: Evidence for phonetic category formation. *Journal of the Acoustical Society of America, 83*(2), 729–740. https://doi.org/10.1121/1.396115

Giles, H., Coupland, J., & Coupland, N. (1991). *Contexts of accommodation: Developments in applied sociolinguistics*. Cambridge University Press.

Goldinger, S. D. (1997). Perception and production in an episodic lexicon. In J. W. Mullennix (Ed.), *Talker variability in speech processing* (pp. 33–66). Academic Press.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review, 105*, 251–279. https://doi.org/10.1037/0033-295X.105.2.251

Goldinger, S. D., & Azuma, T. (2004). Episodic memory reflected in printed word naming. *Psychonomic Bulletin and Review, 11*(4), 716–722. https://doi.org/10.3758/BF03196625

Gregory, S. W., & Webster, S. (1996). A low frequency acoustic signal in voices of interview partners effectively predicts communication accommodation and social status perceptions. *Journal of Personality and Social Psychology, 70*(6), 1231–1240. https://doi.org/10.1037/0022-3514.70.6.1231

Gries, S. T. (2005). Syntactic priming: A corpus-based approach. *Journal of Psycholinguistic Research, 34*(4), 365–399. https://doi.org/10.1007/s10936-005-6139-3

Hao, Y. C., & de Jong, K. (2016). Imitation of second language sounds in relation to L2 perception and production. *Journal of Phonetics, 54*, 151–168. https://doi.org/10.1016/j.wocn.2015.10.003

Jassem, W. (1981). *Podręcznik wymowy angielskiej* [Handbook of English pronunciation]. Państwowe Wydawnictwo Naukowe.

Jia, G., Strange, W., Wu, Y., Collado, J., & Guan, Q. (2006). Perception and production of English vowels by Mandarin speakers: Age-related differences vary with amount of L2 exposure. *Journal of the Acoustical Society of America, 119*(2), 1118–1130. https://doi.org/10.1121/1.2151806

Jiang, F., & Kennison, S. (2022). The impact of L2 English learners' belief about and interlocutor's English proficiency on L2 phonetic accommodation. *Journal of Psycholinguistic Research, 51*, 217–234. https://doi.org/10.1007/

s10936-021-09835-7

Katzir, R., & Singh, R. (2013). A note on presupposition accommodation. *Semantics & Pragmatics, 6*, 1–16. https://doi.org/10.3765/sp.6.5

Keating, P. A. (1984). Phonetic and phonological representation of stop consonant voicing. *Language, 60*, 286–319. https://doi.org/10.2307/413642

Keating, P. A., Mikoś, M. J., & Ganong III, W. F. (1981). A cross-language study of range of voice onset time in the perception of initial stop voicing. *Journal of the Acoustical Society of America, 70*, 1261–1271. https://doi.org/10.1121/1.387139

Kessinger, R. H., & Blumstein, S. E. (1997). Effects of speaking rate on voice-onset time in Thai, French, and English. *Journal of Phonetics, 25*, 143–168. https://doi.org/10.1006/jpho.1996.0039

Kim, D., & Clayards, M. (2019). Individual differences in the link between perception and production and the mechanisms of phonetic imitation. *Language, Cognition and Neuroscience, 34*(6), 769–786. https://doi.org/10.1080/23273798.2019.1582787

Lisker, L., & Abramson, A. S. (1964). A cross language study of voicing in initial stops: Acoustic measurements. *Word, 20*, 384–422. https://doi.org/10.1080/00437956.1964.11659830

Llompart, M., & Reinisch, E. (2018) Imitation in a second language relies on phonological categories but does not reflect the productive usage of difficult sound contrasts. *Language and Speech, 62*(3), 594–622. https://doi.org/10.1177/0023830918803978

Lorenc, A. (2012). Zaburzenia dźwięczności. Analiza akustyczna i audytywna [Voicing disorders. Acoustic and auditive analysis]. *Logopedia, 41*, 71–107.

Magloire, J., & Green, K. (1999). A cross-language comparison of speaking rate effects on the production of Voice Onset Time in English and Spanish. *Phonetica, 56*, 158–185. https://doi.org/10.1159/000028449

Malisz, Z., & Żygis, M. (2015). Voicing in Polish: interactions with lexical stress and focus. *Proceedings of the 18th ICPhS in Glasgow.*

Miller, J. L., Green, K. P., & Reeves, A. (1986). Speaking rate and segments: A look at the relation between speech production and speech perception for the voicing contrast. *Phonetica, 43*, 104–115. https://doi.org/10.1159/000261764

Mitterer, H., & Ernestus, M. (2008). The link between speech perception and production is phonological and abstract: Evidence from the shadowing task. *Cognition, 109*, 168–173. https://doi.org/10.1016/j.cognition.2008.08.002

Mitterer, H., & Müssler, J. (2013). Regional accent variation in the shadowing task: Evidence for a loose perception-action coupling in speech. *Attention, Perception & Psychophysics, 75*(3), 557–575. https://doi.org/10.3758/s13414-012-0407-8

Namy, L. L., Nygaard, L. C., & Sauerteig, D. (2002). Gender differences in vocal accommodation: The role of perception. *Journal and Language and Social Psychology, 21*, 422–432. https://doi.org/10.1177/026192702237958

Nielsen, K. (2011). Specificity and abstractness of VOT imitation. *Journal of Phonetics, 39*(2), 132–142. https://doi.org/10.1016/j.wocn.2010.12.007

Nielsen, K. (2013). Investigating developmental changes in phonological representation using the imitation paradigm. *Proceedings of Meetings on Acoustics, 19*, 060086. https://doi.org/10.1121/1.4800746

Nielsen, K. (2014). Phonetic imitation by young children and its developmental changes. *Journal of Speech, Language and Hearing Research, 57*(6), 2065–2075. https://doi.org/10.1044/2014_JSLHR-S-13-0093

Pickering, M. J., Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavior and Brain Sciences, 27*(2), 169–226. https://doi.org/10.1017/S0140525X04000056

Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *Journal of the Acoustical Society of America, 119*, 2382–2393. https://doi.org/10.1121/1.2178720

Pardo, J. S., Gibbons, R., Suppes, A., & Krauss, R. M. (2012). Phonetic convergence in college roommates. *Journal of Phonetics, 40*, 190–197. https://doi.org/10.1016/j.wocn.2011.10.001

Pardo, J. S., Urmanche, A., Wilman, S., & Wiener, J. (2017). Phonetic convergence across multiple measures and model talkers. *Attention, Perception and Psychophysics, 79*(2), 637–659. https://doi.org/10.3758/s13414-016-1226-0

Paquette-Smith, M., Schertz, J., & Johnson, E. K. (2022). Comparing phonetic convergence in children and adults. *Language and Speech, 65*(1), 240–260. https://doi.org/10.1177/00238309211013864

Piske, T., Flege, J. E., MacKay, I. R. A., & Meador, D. (2002). The production of English vowels by fluent early and late Italian-English bilinguals. *Phonetica, 59*, 49–71. https://doi.org/10.1159/000056205

Podlipský, V. J., & Šimáčková, Š. (2015) Phonetic imitation is not conditioned by preservation of phonological contrast but by perceptual salience. In: *Proceedings of the 18th International Congress of Phonetic Sciences, Glasgow*.

Rojczyk, A. (2011). Perception of the English Voice Onset Time continuum by Polish learners. In J. Arabski & A. Wojtaszek (Eds.), *The acquisition of L2 phonology* (pp. 37-58). Multilingual Matters.

Rojczyk, A. (2012). *Phonetic and phonological mode in second-language speech: VOT imitation* [Paper presentation, EUROSLA 22 – The 22nd Annual Conference of the European Second Language Association, September 5-8, Poznań, Poland].

Rojczyk, A. (2013). Phonetic imitation of L2 vowels in a rapid shadowing task. In J. Levis, & K. LeVelle (Eds.), *Proceedings of the 4th Pronunciation in Second Language Learning and Teaching Conference* (pp. 66–76). Iowa State University.

Rojczyk, A., Porzuczek, A. & Bergier, M. (2013). Immediate and distracted imitation in second-language speech: Unreleased plosives in English. *Research in Language, 11*, 3–18.

Shockley, K., Sabadini, L., & Fowler, C. A. (2004). Imitation in shadowing words. *Perception & Psychophysics, 66*(3), 422–429. https://doi.org/10.3758/BF03194890

Simonet, M., Casillas, J. V., & Díaz, Y. (2014). The effects of stress/accent on VOT depend on language (English, Spanish), consonant (/d/, /t/) and linguistic experience (monolinguals, bilinguals). *Proceedings of the International Conference on Speech Prosody*, 202-206.

Tajfel, H., & Turner, J. C. (1986). The social identity theory of intergroup behaviour. In S. Worchel & W. G. Austin (Eds.), *Psychology of intergroup relations* (pp. 7–24). Blackwell.

Trofimovich, P., & Kennedy, S. (2014). Interactive alignment between bilingual interlocutors, evidence from two information-exchange tasks. *Bilingualism, Language and Cognition, 17*(4), 822–836. https://doi.org/10.1017/S1366728913000801

Ulbrich, C. (2021). Phonetic accommodation on the segmental and the suprasegmental level of speech in native–non-native collaborative tasks. *Language and Speech.* Advance online publication. https://doi.org/10.1177/00238309211050094

Wade, L. (2020). *The linguistic and the social intertwined: Linguistic convergence toward southern speech* [Doctoral Dissertation, University of Pennsylvania].

Wade, L., Lai, W., & Tamminga, M. (2021). The reliability of individual differences in VOT imitation. *Language and Speech, 64*(3), 576–593. https://doi.org/10.1177/0023830920947769

Wagner, M. A., Broersma, M., McQueen, J. M., Dhaene, S., & Lemhöfer, K. (2021). Phonetic convergence to non-native speech: Acoustic and perceptual evidence. *Journal of Phonetics, 88*, 1011076. https://doi.org/10.1016/j.wocn.2021.101076

Waniek-Klimczak, E. (2005). *Temporal parameters in second language speech: An applied linguistic phonetic approach*. Wydawnictwo Uniwersytetu Łódzkiego.

Whiteside, S. P., Dobbin, R., & Henry, L. (2003). Patterns of variability in voice onset time: a developmental study of motor speech skills in humans. *Neuroscience Letters, 347*, 29–32. https://doi.org/10.1016/S0304-3940(03)00598-6

Willemyns, M., Gallois, C., Callan, V., & Pittam, J. (1997). Accent accommodation in the job interview: Impact of interviewer accent and gender. *Journal of Language and Social Psychology, 16*(1), 3–32. https://doi.org/10.1177/0261927X970161001

Yu, V. Y., De Nil, L. F., & Pang, E. W. (2015). Effects of age, sex and syllable structure on voice onset time: Evidence from children's voiceless aspirated stops. *Language and Speech, 58*(2), 152–167. https://doi.org/10.1177/0023830914522994

Zając, M., & Rojczyk, A. (2014) Imitation of English vowel duration upon exposure to native and non-native speech. *Poznan Studies in Contemporary Linguistics, 50*(4), 495–514.