ROCZNIKI HUMANISTYCZNE
Tom LXIX, zeszyt 6 – 2021
ZESZYT SPECJALNY
DOI: https://doi.org/10.18290/rh21696s-2

EUGENIUSZ CYRAN
MARTYNA MICHALSKA

TARGET ATTAINMENT IN THE PRODUCTION OF ENGLISH VOWELS BY POLISH STUDENTS: THE INFLUENCE OF LANGUAGE INSTRUCTION IN THE TIME OF COVID-19, IMITATION, AND MUSICALITY

1. DEFINING THE PROBLEMS

The object of the study is deliberately limited to five English vowel categories: [ɪ] (HID), [æ] (HAD), [ɜ:] (HERD), [ʊ] (HOOD) and [ɒ] (HOD). The quality of these vowels is often wrongly assumed by Polish learners of English to be fairly similar to their first language (L1) categories, that is, [ɨ] (HYDrant) 'hydrant', [a] (HADes) 'Hades', [ɛ] (sCHEDa) 'heritage', [u] (CHUDy) 'slim', and [ɔ] (sCHODy) 'stairs', respectively. In fact, each of the English (L2) categories is phonetically different from the perceptual assimilations in L1. The English vowel [ɪ] is often replaced by Polish learners with the native [i] or [ɨ]. Thus, the word *bit* is pronounced as either *[b'it], which seems to be a case of spelling-induced assimilation, or *[bɨt], which is freed from the influence of spelling, but is still a case of assimilation to a Polish category. Neither rendition is correct, as the vowel [ɪ] lies somewhere

Prof. dr hab. EUGENIUSZ CYRAN – pracownik w Katedrze Językoznawstwa Porównawczego Angielsko-Polskiego Instytutu Językoznawstwa Katolickiego Uniwersytetu Lubelskiego Jana Pawła II; adres do korespondencji: Katolicki Uniwersytet Lubelski Jana Pawła II, Al. Racławickie 14, 20-950 Lublin; e-mail: eugeniusz.cyran@kul.pl; ORCID: https://orcid.org/0000-0002-3266-6961.

Mgr MARTYNA MICHALSKA – studentka w Szkole Doktorskiej Katolickiego Uniwersytetu Lubelskiego Jana Pawła II; adres do korespondencji: Katolicki Uniwersytet Lubelski Jana Pawła II, Al. Racławickie 14, 20-950 Lublin; e-mail: martyna.michalska@kul.pl; ORCID: https://orcid.org/0000-0002-6533-5819.

between the two Polish vowels and is significantly different from them. Likewise, the English vowel [æ] in *have* and *dad* is often replaced with [ɛ], as in *[hɛf], and [a], as in *[dat], respectively, while in English it is, again, rather a mixture of the two vowels. It is not mid front like [ɛ], or low central like [a], but rather front mid-low (Sobkowiak 142), or even front low (Cruttenden 84). The vowel [ɜ:] is long and central, which makes it difficult to attain for two reasons: there are no length distinctions or central vowels in Polish. This vowel is often replaced with Polish [ɛ]. The remaining two English vowels, that is, [ʊ] and [ɒ] have considerably less lip rounding than the Polish targets of assimilation. English [ɒ] has a relatively low and open articulation, while Polish [o] is more rounded and mid. Sobkowiak (p. 149) describes English [ɒ] as a category which is intermediate between Polish [o] and [a]. In this sense, it is parallel to English [æ] which lies between Polish [ɛ] and [a] (Gonet et al.). Additionally, English [ʊ] is centralized, that is, both lower in articulation, and more central than Polish [u]. It is predicted to be quite hard for Polish learners because it cannot be positioned easily as an intermediate category with respect to two Polish vowels (Sobkowiak 151), like in the case of English [ɪ] (between Polish [i] and [ɨ]), [æ] (between Polish [ɛ] and [a], or [ɒ] (between Polish [o] and [a]). In this paper, we follow the opinions which assume that the production of the English targets will be problematic for Polish learners for reasons described above. We also predict that some vowel categories will be more difficult to attain than others. For example, it is often argued that phonetic training may more readily lead to the creation of new L2 categories precisely where distinctions from phonetically adjacent L1 vowels are relatively greater (Bohn and Flege, Rojczyk). For the same reasons, the selected vowels should also be easier for goodness judgement by raters (Flege et al.). Our aim is to establish how the above-mentioned predictions play out in two experimental conditions – reading and imitation – and to what extent the results correlate with the independently measured musicality of the participants.

It is a commonly held view that in the process of learning a second language pronunciation learners need to first shift away from the initial stage of being biased perceivers of L2 categories through their native L1 ‘sieve’ (Best, Trubetzkoy, Kuhl and Iverson, Iverson et al.). This can be achieved through phonetic training and lead to establishing new L2 categories, allowing the learners to produce as well as comprehend them when using the second language (e.g. Balas, Flege). The process of attainment of near-native pronunciation is a complex matter as it involves developing perceptual awareness of phonetic

distinctions that are often distorted by native prototypes and establishing new targets in production. Thus, not only perception-based factors are at play but also articulatory ones, which may take the form of bad habits. For example, in order to attain the correct English phonetic targets [ð, θ], as in *this, bathe, thing* and *both*, Polish learners have to first sever the earlier established wrong connection between these English sounds and the phonetic segments they used as adaptation replacements from L1, that is, [z, d, v] for the voiced fricative [ð], and [s, t, f] for the voiceless [θ] (e.g. Rojczyk and Porzuczek).

Equally challenging is the problem of assessing phonetic progress in second language acquisition, which is nonetheless necessary if we want to monitor and stimulate the shift from the unwelcome L1 assimilation to the desired L2 category creation. In fact, the problem does not only concern individual sounds, but also their mutual influence and phenomena at higher prosodic levels of speech, including stress, liaison phenomena, rhythm, intonation, general fluency, etc. As for the individual sounds that a learner produces, an overall assessment is needed which must take into account the fact that students may attain some new phonetic categories, for example, *th*-sounds, but fail to create others, for example, vowels. For these reasons, measuring phonetic attainment is a very complicated matter. Szpyra-Kozłowska et al., for example, evaluate the general assessment schemes, focusing on the distinction between holistic and atomistic methods. The former are based on imprecise impressionistic judgements, which is a handy and quick method used in, for example, Cambridge English Examinations. The latter (atomistic ones) involve detailed marking schemes established in order to evaluate individual diagnostic aspects such as the correctness of vowel and consonant segments, selected phonological or phonetic processes, or prosodic elements such as stress, rhythm, and intonation (e.g. Stasiak).

Unlike with some aspects of English phonetics, such as the *th*-sounds, whose attainment is fairly easy to assess visually, auditorily or by means of acoustic analysis, measuring progress with respect to vowels is much more challenging. For example, acoustic measurements of the formants F1 and F2, especially if plotted in diagrams against the respective L2 targets coming from native English productions, may allow for a visual evaluation of some sort, whereby it is possible to estimate the phonetic distance between the pronounced vowel and its intended target, as well as its relation to the closest Polish category (adaptation replacement). It is also possible to measure the distance mathematically, given the values of F1 and F2 by calculating the Euclidean distance (Deterding, Rojczyk), or apply more complicated models identifying

perceptual target zones within an auditory-perceptual space (Miller). However, due to the nonlinear nature of the relationship between acoustics and perception, the judgement as to whether an L2 target has been attained, or how a produced vowel is different from either L1 or L2 is not a simple task. For this reason, gauging the degree of perceived cross-language phonetic distance in an objective way is still a matter of research (e.g. Flege 264). In this paper, we rely on scalar assessment by trained Polish phoneticians who are also experienced teachers of English pronunciation at academic level. They are aware of both L2 and L1 categories. Therefore, the advantage that trained Polish phoneticians offer is that they may be able to assess whether, and to what degree, a given sound is close to an L2 target, or whether it is an L1 replacement. Although their judgements are to some extent subjective, it is possible to achieve a level of objectivity of the results by means of statistics. More importantly, the judgements reflect a cognitive operation of categorical perception considering a range of permissible phonetic variation, something that acoustic analysis may document but not evaluate. Possibly, a combination of acoustic and perceptual criteria could help overcome the limitations of these methods used separately. This, however, is an issue for further research.

A number of factors have been claimed to influence attainment of native-like pronunciation in a second language, such as the age at which one begins learning a foreign language (Birdsong, Long), motivation, personality, memory, or differences between the phonemic systems of L1 and L2 (Larsen-Freeman and Long). Given the relation between perception and production and assuming that the degree of assimilation of L2 targets to the native L1 sounds may change as a result of language instruction, practice, as well as depending on the talent for imitation on the part of the students (Best), the interesting question is what underlies this linguistic talent. One hypothesis, which has been explored in recent years, is that, among others, there may be a strong link between second language acquisition and musical aptitude or training. This paper looks at the relation between inborn musical talent, rather than musical training, and the talent for foreign language learning by looking at the phonetic attainment of some notoriously difficult English vowels by Polish learners at the undergraduate university level in relation to musicality of the learners. Given the pilot nature of this study, we decided to include musicality as an additional factor affecting vowel target attainment in order to enrich the set of research options for future full-scale studies.

Several analogies between musicality and language have been noted in the literature (e.g. Chobert and Besson, Jackendoff). The relationship has

also become of interest in the context of second language acquisition, especially the pronunciation skills (e.g. Fonseca-Mora et al., Dolman and Spring, Molivanov et al.). One can think of a number of areas in foreign language pronunciation learning that could be correlated with musicality. One relation, because it directly refers to tune, may be assumed to exist between musicality and intonation (Schön et al.). Another area of interest, this time based on rhythm, can involve the relation between musicality and stress (e.g. Gralińska-Brawata and Rybińska). The relation studied in this paper, that is, between musicality and vowel target attainment in L2, in a sense also looks at the melodic aspects of speech. Acoustically, vowels are described as different patterns of arrangement of energy bands, called formants, at particular frequencies. Therefore, attainment of vocalic categories of L2 and its relation to musicality is a valid area of research.

The question is how musicality is understood and more importantly, how it can be measured. Approaches to gauging musicality may vary. Some studies simply rely on declared special interest in music or playing instruments (e.g. Peltola), while others attempt to set up tests relying on perception and production of various aspects of music, that is, rhythm and tune (Gralińska-Brawata and Rybińska). In this paper, we use an established test of musical aptitude devised by Gordon (*Advanced measures*). The test has already been used in a number of studies related to musicality and second language acquisition (e.g. Pei et al.).

2. RESEARCH QUESTIONS

A number of research questions are posed in this study. The first of them concerns the effects of phonetic training:

RQ1. What is the level of attainment of selected English vowels after intensive online language instruction and phonetic drilling? We hypothesize that most students should obtain a score above 50% percent in the measuring system applied. The benchmark takes into account the assumption that learners obtaining this score have substantially moved away from L1 categories but may not have established stable and correct new L2 ones. In this respect, any results above 50% will be viewed as moving in the direction of correct L2 categories. This study evaluates the expected attainment rather than progress from point A to point B. Therefore, it is not longitudinal in nature. It should be added that the syllabus of pronunciation teaching administered among the

students in this study is constructed in such a way that all the English vocalic categories are covered in the first semester of their study. Additionally, various segmental and supra-segmental aspects of English phonetics are also introduced. Thus, evaluation of their vowel production is realistic.

The second research question refers to predictions concerning the results obtained in the reading and the imitation experiments:

RQ2. Is there a difference in the production of read and imitated stimuli? We hypothesize that the results in the imitation experiment should be significantly better than in the list reading experiment. This is because a heard stimulus provides a correct auditory target on which the subject could base the production, while in the reading task the subject must rely on his or her own internalized target, developed during the pronunciation course.

The third research question concerns the relationship between vowel production outcomes and independently established scores for musicality:

RQ3. To what extent the goodness rates in vowel production in the two experimental contexts are related to the musical talent of the participants? We hypothesize that musicality of the participants should correlate with the higher results in both experimental settings. We assume that musicality is an important factor in the long-term establishment of L2 categories, as well as in imitation.

3. METHOD AND PROCEDURE

The participants in this study were students of English philology at The John Paul II Catholic University of Lublin. There were six male and six female students who completed a semester of intensive phonetic training covering, among other things, the theoretical and practical aspects of all the vocalic categories of English. The participants were asked to install *Praat*, free software for acoustic analysis (Boersma), which enabled them to produce recordings in the WAV format, at 44,1kHz sampling rate.[1] The recordings were made on the basis of a PowerPoint presentation with a list of English words containing the 5 English vowels in the context H_D, that is, HID, HAD, HERD, HODD and HOOD (e.g. Deterding). Each word was provided in spelling and IPA phonetic transcription in both experimental contexts. The students were familiar with phonetic transcription. In Experiment I, the participants were asked to

[1] An acoustic analysis of the recordings is planned for a separate study.

read the list of words by looking at the spelling and transcription and using falling intonation as in statements. This stage was assumed to reflect the level of attainment of the vowel targets which is a result of training. In other words, the learners were to produce vowels on the basis of the auditory form that they had managed to create in the process of learning, and which is related to their long-term memory. In Experiment II, the participants were provided with the same visual representation of the words and a recording by a native speaker of corresponding gender. Each word was imitated separately, whereby the participants could rely on their perception and short-term memory before they produced the words.

Finally, the participants were asked to do the online Gordon's Advanced Measures of Music Audiation test (AMMA), which yields separate scores for rhythm, tone and produces a total score on a percentage scale 0–100%, which we took as the final score for further comparisons.

The Gordon's Advanced Measures of Music Audiation test was designed for measuring musical aptitude among college students. It is an online test which consists of thirty pairs of brief musical statements followed by musical answers. The participant in this test hears a pair of musical pieces in each question and assesses whether they are identical, or whether they differ in terms of rhythm or tone. The procedure lasts about 15 minutes. The score is then broken down into the categories of rhythm and tonal, and a total score is calculated on a percentage scale. The scores of the AMMA test appear to be fairly good predictors of the musical achievement, with reliability coefficients ranging from .83 to .88 (Gordon, *Predictive validity*, Schleuter).

The recordings were edited out into separate files for separate words to be rated by experienced university level English pronunciation teachers. The assessment was based on a PowerPoint presentation sent to them, with embedded recordings for individual speakers and an embedded online form for anonymized evaluation. The six raters (4 men and 2 women) did not know whether they were evaluating imitation or reading. The recordings of the two stages were randomised.

A 5-point Likert scale was used for the purpose of goodness rating. The instruction to the raters defined 1 on the scale as "L1 replacement" (a total assimilation), or "complete mispronunciation". 5 was defined as "(near-)native L2 category". The intermediate marks constituted a range between the two extremes, where it was assumed that 2 was understood as slightly removed

from L1, 4 as slightly removed from the L2 vowel, and 3 as a case between a good L2 category and bad L1 replacement.

A mean rating value from the six raters was calculated for each vowel production in the two experimental conditions, as shown in table 1. With respect to the first research question (RQ1), the overall assessment of target attainment, the total scores of individual participants from the reading experiment (Experiment I) were transformed into a percentage scale with two benchmarks yielding the following distinctions: < 50% = unsatisfactory, 50–70% = sufficient, and > 70% = very good (table 2). The imitation results (Experiment II) were not used for the evaluation of vowel attainment because they were assumed to reflect the outcome of phonetic training to a lesser degree. However, the influence of phonetic training on perception of L2 categories and their imitation deserves a separate study.

With regard to research questions 2 and 3, non-parametric statistic tests were performed due to the small sample of data, that is, Wilcoxon test for comparing between two experimental conditions (RQ2), and Spearman correlation test for linear relations between musicality and rating scores (RQ3). Additionally, a post-hoc Mann–Whitney test for comparing independent samples – a group comparison between men and women – was carried out as a result of preliminary analysis of the results in tables 1 and 2.

4. RESULTS

The following table presents the rating data for each participant in the reading experiment (Experiment I) and the imitation experiment (Experiment II), as well as the results of the musicality test (AMMA). It also shows the total means for each participant in the two experiments, including the difference between the two experiments, and the total mean values for each vowel in Experiment I and II, including standard deviation.

Table 1. Mean rating scores for all participants and vowels in Experiment I and II, with total means and AMMA results in percent.
MS = male speaker, FS = female speaker, μ = total means for each vowel,
SD = standard deviation, ID = identification of the participants.

ID	Experiment I: mean rating of reading					Experiment II: mean rating of imitation					Total means for each participant			Difference	AMMA %
	HAD	HERD	HID	HOD	HOOD	HAD	HERD	HID	HOD	HOOD	Experiment I		Experiment II		
	[æ]	[ɜː]	[ɪ]	[ɒ]	[ʊ]	[æ]	[ɜː]	[ɪ]	[ɒ]	[ʊ]					
MS1	3.50	2.80	3.30	3.50	2.70	4.50	4.00	4.50	3.20	3.70	3.20	<	4.00	+0.80	68.00
MS2	3.20	3.80	3.50	1.80	2.20	3.80	3.30	3.70	3.00	2.30	2.90	<	3.20	+0.30	44.00
MS3	4.30	4.00	3.50	2.20	1.70	4.70	4.80	4.80	4.80	4.80	3.10	<	4.80	+1.80	29.00
MS4	1.50	2.70	2.80	1.30	1.80	3.70	3.30	3.80	3.30	1.70	2.00	<	3.20	+1.20	35.00
MS5	2.70	3.20	3.50	1.80	1.30	3.00	3.20	3.70	1.70	1.80	2.50	<	2.70	+0.20	20.00
MS6	2.80	3.20	2.70	3.20	2.50	4.30	4.00	4.00	3.80	2.80	2.90	<	3.80	+0.90	44.00
FS1	4.70	4.80	3.80	4.30	4.20	4.50	4.50	3.80	4.70	3.70	4.40	>	4.20	-0.20	70.00
FS2	2.50	4.20	3.20	4.20	1.80	2.80	4.20	4.00	4.20	3.30	3.20	<	3.70	+0.50	26.00
FS3	3.20	4.50	4.50	4.20	2.30	4.20	4.50	4.30	3.80	2.70	3.70	<	3.90	+0.20	87.00
FS4	4.50	4.70	4.50	4.20	4.50	4.80	4.30	4.50	4.50	3.30	4.50	>	4.30	-0.20	96.00
FS5	3.80	2.70	1.70	3.00	2.00	3.50	2.20	1.20	2.20	1.70	2.60	>	2.10	-0.50	38.00
FS6	3.80	4.00	2.00	4.00	2.50	4.50	4.50	2.50	4.00	3.20	3.30	<	3.70	+0.40	56.00
μ	**3.38**	**3.72**	**3.25**	**3.14**	**2.46**	**4.03**	**3.90**	**3.73**	**3.60**	**2.92**	**3.19**	**<**	**3.63**		
SD	**0.92**	**0.77**	**0.86**	**1.11**	**0.97**	**0.66**	**0.75**	**0.99**	**0.96**	**0.94**	**0.73**		**0.74**		

Based on this data, one can make some provisional observations. For example, the table with total means seems to indicate that generally the marks in the imitation experiment were higher than those in the reading experiment (3.19 < 3.63). Three out of the twelve participants, all female, obtained slightly worse marks in the imitation. However, the drop was rather small (between 0.20 and 0.50), and two out of the three students exhibiting a drop between the two experimental contexts, that is, FS1 and FS4 were the two highest scoring participants in the reading experiment. They also scored high in the musicality test. This suggests that there is a correlation between target attainment and musicality, but the slightly lower results in the imitation experiment may show a ceiling effect. The average results obtained by the two students in the imitation experiment, that is, 4.20 and 4.30, as compared to 4.40 and 4.50 in the reading experiment, respectively, clearly show that their productions were close to 'near-native' in both cases. The participant with the highest drop (FS5) was one of the lowest scoring students in both experiments and in the musicality test.

On the other hand, some of the participants were judged better in the imitation experiment by a mean difference of 1.80 points on the scale. All his productions in the imitation experiment were judged as almost 'near-native'. Interestingly, the participant in question, MS3, achieved one of the two lowest scores in the musicality test. This may suggest that the level of musicality correlates better with the long-term establishment of vocalic categories, and that it correlates less with imitation.

It is also possible to make some preliminary observations concerning the rating scores for individual vowels. In both experiments, the lowest mean score was given to the vowel [ʊ] as in HOOD and [ɒ] as in HOD, while the highest rated two vowels were [æ], as in HAD, and [ɜː], as in HERD. The fact that [ʊ] is potentially a very hard phonetic category to establish by Polish learners was already mentioned in Sobkowiak (*English phonetics*).

Finally, one should note that the standard deviation values in the total means of the two experiments do not change conspicuously. This suggests that despite the observed improvement in the rating scores between Experiment I and II, the raters maintained a similar degree of uniformity in their judgements.

4.1 Research question 1

The first research question concerned the level of attainment of the five selected vowel targets after intensive online language instruction and phonetic

drilling. To answer this question, scores of individual participants in the reading experiment (Experiment I) were transformed into percentage and ranked with respect to two reference points: 50% and 70% (table 2).

Table 2. Ranked percentage scale of the vowel target attainment in Experiment I, with benchmarks at 50% and 70%.

Rank	ID	Total score (max. 150)	%	
1.	FS4	134	89%	
2.	FS1	131	87%	**> 70%**
3.	FS3	112	75%	
4.	FS6	98	65%	
5.	MS1	95	63%	**70%**
6.	FS2	95	63%	⇧
7.	MS3	94	63%	
8.	MS2	87	58%	**50%**
9.	MS6	86	57%	
10.	FS5	79	53%	
11.	MS5	75	50%	
12.	MS4	61	41%	**< 50%**

It can be observed that only one participant received a total score below 50%. Thus, our assumption concerning the general level of attainment seems to have been correct. Additionally, it is noteworthy that three participants received scores above 70%, and four best scores were achieved by female participants. This observation prompted us to verify the statistical significance of the difference between scores obtained by two independent groups of participants, male and female, and whether it holds for both experimental contexts.

For this purpose, post-hoc Mann–Whitney U tests were carried out comparing the scores for each vowel between the two groups, and the total means obtained by the participants in the two groups in the reading and imitation experiments separately.

In the reading experiment, the female participants received significantly higher marks than the men in two cases (table 3): with respect to the total mean scores, $U = 4.500$, $p = 0.037$, with large effect size ($r = –0.601$), and with respect to the vowel [ɒ] in HOD, $U = 2.000$, $p = 0.013$, also with large

effect size (r = –0.717). The other scores were not significantly different with respect to gender.

Table 3. The results of the Mann-Whitney U tests comparing male and female scores obtained for vowels in the two experimental conditions.

	Experiment I: Reading				**Experiment II: Imitation**			
	p	*U*	*Z*	*r*	*p*	*U*	*Z*	*r*
HAD	0.200	9.500	–1.281	–0.370	0.936	17.000	–0.080	–0.023
HERD	0.066	6.000	–1.841	–0.532	0.298	11.000	–1.041	–0.300
HID	0.810	16.000	–0.240	–0.069	0.631	14.500	–0.480	–0.139
HOD	**0.013**	**2.000**	**–2.482**	**–0.717**	0.262	10.500	–1.121	–0.324
HOOD	0.230	10.000	–1.201	–0.347	0.810	16.000	–0.240	–0.069
Total means	**0.037**	**4.500**	**–2.082**	**–0.601**	0.810	16.000	–0.240	–0.069

The scores for female and male participants did not differ significantly in the imitation experiment, which suggests that imitation had a levelling effect between the two groups. These results shed additional light on our second research question concerning the difference between rating scores obtained in the two experimental conditions.

4.2 Research question 2

The second research question concerned the difference between the rating scores of vowel productions in the two experimental conditions. We hypothesised that imitation should produce better ratings because it is based on a correct auditory impression provided by a native speaker, which just needs to be echoed in production, while in the reading experiment the participants had to rely on the auditory impression they had established in the process of language learning.

In order to determine if there is a significant change in the results between read and imitated productions of vowels, the Wilcoxon signed-rank test with two dependent samples was used. The mean scores for individual vowels for all participants in the reading experiment were compared with those in the imitation experiment. Additionally, the total means of all participants in the two conditions were also compared.

The Wilcoxon signed-rank test revealed a statistically significant positive change with respect to the comparisons of the total means, as well as for two vowels, HAD and HID. Total rating means in the imitation (Experiment II) were significantly higher than in the reading condition (Experiment I), $T = 13.50$, $Z = 2.000$, $p = 0.045$, with a medium effect size ($r = 0.408$). The vowel [æ] in HAD was rated significantly higher in the imitation condition (Experiment II) than in the reading (Experiment I), $T = 4.50$, $Z = 2.706$, $p = 0.007$, with a large effect size (r = 0.552). The vowel [ɪ] in HID was rated significantly higher in the imitation condition (Experiment II) than in the reading one (Experiment I), $T = 6.50$, $Z = 2.141$, $p = 0.032$, with a medium effect size (r = 0.437). The other vowels, although their rating increased in the imitation condition (table 1), do not show any significant effect statistically, as shown in table 4.

Table 4. The results of the Wilcoxon signed-rank test comparing the productions of vowels and the total means in two experimental conditions: reading vs. imitation.

	T	*Z*	*p*	*r*
HAD	**4.50**	**2.706**	**0.007**	**0.552**
HERD	11.00	1.362	0.173	0.278
HID	**6.50**	**2.141**	**0.032**	**0.437**
HOD	15.00	1.274	0.203	0.260
HOOD	22.50	1.294	0.195	0.264
Total means	**13.50**	**2.000**	**0.045**	**0.408**

4.3 Research question 3

The third research question concerned the linear relationship between the goodness rates for each vowel in a given experimental context and the musicality of the participants. The same test was calculated for the total mean rates in the two experimental contexts in relation to musicality. We hypothesised that higher scores in the musicality test should correlate positively with higher grades for particular vowels and higher total means.

To answer the research question, the Spearman correlation coefficient between the two analysed variables was calculated separately for reading and imitation experiments, including the total means. The results of the Spearman's Rho tests show that in Experiment I, there is a significant positive linear relationship between musicality and vowel ratings for some vowels, as well as between

musicality and the total means. Specifically, there is a moderate positive relationship for HOD, r_s (df 10) = 0.627, p = 0.029; a very high positive correlation for HOOD, r_s (df 10) = 0.909, $p < 0.001$; and a high correlation for total means, r_s (df 10) = 0.766, p = 0.004. In the case of Experiment II (imitation), a significant positive correlation between musicality and vowel ratings was only found for the vowel [æ] in HAD, r_s (df 10) = 0.649, p = 0.022. No significant correlation was found between the ratings of other vowels and musicality as shown in table 5.

Table 5. The results of the Spearman rho tests for correlation between rating scores for individual vowels in the two experimental conditions and musicality of the participants.

	Experiment I: reading		Experiment II: imitation	
	r_s	p	r_s	p
HAD	0.575	0.051	**0.649**	**0.022**
HERD	0.484	0.111	0.411	0.185
HID	0.438	0.154	0.258	0.419
HOD	**0.627**	**0.029**	0.289	0.361
HOOD	**0.909**	**0.001**	0.282	0.375
Total means	**0.766**	**0.004**	0.522	0.082

5. DISCUSSION

A cursory look at the results in tables 1 and 2 allowed us to make a few preliminary observations concerning the best and worst scoring vowel categories in the two experiments, as well as concerning some trends, which needed to be tested statistically. Despite the changes in the overall values between the two experiments, it is important to note that the standard deviation between the mean scores remained quite uniform, that is, between 0.73–0.74. This suggests that the raters were reliable.

As for the 50% benchmark, assumed for the purpose of assessing the vowel target attainment in this study, it was to some extent chosen arbitrarily, taking into consideration the short time of training, which consisted of one semester, and the general pandemic conditions pertaining to online studying. But most importantly, we assumed that scores above 50% would be satisfactory because they indicate that on the whole the participants have managed to

move away from the native L1 categories as assimilation replacements, and possibly constructed new categories, which have not yet reached the correct L2 targets. In other words, the 50% benchmark follows from the nature of the assessment scale used and the nature of the evaluation task performed by the raters. The usual benchmark of 60% would not represent much difference in terms of goodness of the new L2 categories. It would, however, influence the general evaluation of the results, when compared, for example, with the academic scale of assessment. It is important to bear in mind that this paper does not attempt to mark students' progress, but to assess the degree of attainment. Thus, with 50%, we can say that most participants, except one, have attained satisfactory results. On the other hand, a 60% benchmark would have produced a cut-off point somewhere in the middle of the ranking list, and the results would not look so good. For this reason, we refrain from claiming that the results of phonetic training are particularly positive given the conditions of studying. However, we maintain that the results are quite satisfactory. There are both negative and positive aspects of online learning and teaching of pronunciation. For example, some aspects of teaching pronunciation online resembled laboratory situation such as the use of headphones, break-out rooms, and individual feedback. However, this advantage was often diminished by common problems with internet connection. On the whole, with respect to the first research question it may be stated that the results are not disappointing from the pedagogical perspective.

A few surprising outcomes of this study are connected with the differences between the scores obtained by female and male participants. Some of these differences reached statistical significance and concerned mainly the scores in Experiment I. To some extent, these differences may be related to high scores obtained by female participants in the musicality test, as demonstrated by the correlation tests. Another unexpected result was the levelling effect between the gender groups, which was observed and statistically supported in the imitation experiment. This effect may have a few causes. First, the relatively high scores of the female participants in Experiment I left little space for improvement in Experiment II. This ceiling effect may be due to the fact that the imitation task was too easy, allowing the male participants to come near the female scoring levels, or that the 5-point scale was too narrow and should be extended to a 7, or 9-point scale. What is important, though, is the observation, which requires further study, that imitation appears to yield better rating scores regardless of the musicality of the participants, or gender.

While a closer correlation between musicality and vowel attainment is found in the reading experiment in which long-term memory is at play.

The overall improvement of rating scores in the imitation experiment, which in the case of HID, HAD, and total means also turned out to be significant statistically, supports our hypothesis that when provided with an exemplar auditory impression, and asked to imitate it, the participants would undershoot the targets less than when they have to rely on their own auditory impressions developed during the phonetic instruction and drilling. These results also support the established relation between perception of phonetic distinctions and their reproduction (Balas, Best, Flege).

Finally, the correlation results between musicality and target attainment in the two experimental conditions seem to be quite intriguing and require further research. The strongest correlation between the scores and musicality can be found in Experiment I, when the participants rely on their long term memory, rather than in the imitation experiment, in which the production is based upon a heard example. The important observation that needs to be made here is that the scores that highly correlate with musicality concern the vowels which are on average rated lowest, that is, HOD and HOOD (table 1). Thus, musicality does have an effect in the case of more difficult categories, and it is important in the process of L2 category creation and target attainment. However, the results of correlation in Experiment II are to some extent unexpected (cf. Pastuszek-Lipińska). Here it is the highest scoring vowel HAD that is highly correlated with musicality. It is interesting to note that this is also the vowel target that was improved most between Experiments I and II, even though it was already rated high in Experiment I. Thus the relation between this vowel production and musicality is yet to be understood. It may be the case that the overall levelling effect that imitation has produced an accidental correlation with musicality.

Additionally, one should perhaps conclude that musicality has less to do with reproduction in the imitation context than with long-term establishment of L2 categories, which is the opposite of what we expected. The perception and production in the imitation context seems to be independent of musicality, unlike in the reading context. This, however, does not mean that lower musicality limits the phonetic achievement. The imitation data strongly suggest that perhaps more emphasis should be put on various aspects of imitation in the process of teaching pronunciation, if only because imitation appears to level out the differences between highly musical and non-musical participants. However, as seen in table 2, one should bear in mind that phonetic training

still leads to a broad range of attainment results, which suggests that a search for additional factors influencing phonetic progress needs to be continued.

These conclusions require further studies, which should, first of all, be designed as more statistically robust, allowing for more complex statistical analyses, and involve longitudinal studies, recording participants before and after phonetic training, with perhaps a longer interval between the two experimental contexts (cf. e.g. Nowacka). Quite a separate line of inquiry, concerning methods of measuring of phonetic attainment, could be to attempt to integrate the perceptual assessment of category goodness, like the one used in this study, with some model of acoustic analysis of vowel perception and goodness (e.g. Miller), or phonetic distance (e.g. Deterding, Rojczyk).

Acknowledgements

We would like to thank Marcin Wojtasiński for his assistance with the statistics, as well as Piotr Steinbrich and Sławomir Zdziebko for their comments on earlier drafts of this paper.

WORKS CITED

Balas, Anna. *Non-native Vowel Perception. The Interplay of Categories and Features.* Wydawnictwo Naukowe UAM, 2018.

Best, Catherine T. "A Direct Realist View of Cross-language Speech Perception." *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, edited by Winifred Strange, York Press, 1995, pp. 171-204.

Birdsong, David. "Age and Second Language Acquisition and Processing: A Selective Overview." *Language Learning,* vol. 56, no. 1, 2006, pp. 9-49.

Boersma, Paul. "Praat, a system for doing phonetics by computer." *Glot International*, vol.10, no. 9/10, 2001, pp. 341-345.

Bohn, Ocke-Schwen, and James Emil Flege. "The Production of New and Similar Vowels by Adult German Learners of English." *Studies in Second Language Acquisition*, vol. 14, 1992, pp. 131-158.

Chobert, Julie, and Mireille Besson. "Musical Expertise and Second Language Learning." *Brain Sciences*, vol. 3, no. 2, 2013, pp. 923-940.

Cruttenden, Alan. *Gimson's Pronunciation of English*. 8th ed. Routledge, 2014.

Deterding, David. "The Formants of Monophthong Vowels in Standard Southern British English Pronunciation." *Journal of the International Phonetic Association*, vol. 27, 1997, pp. 47-55.

Dolman, Matthew, and Ryan Spring. "To What Extent Does Musical Aptitude Influence Foreign Language Pronunciation Skills? A Multi-factorial Analysis of Japanese Learners of English." *World Journal of English Language*, vol. 4, no. 4, 2014, pp. 1-11. DOI: 10.5430/wjel.v4n4p1.

Flege, James Emil. "Second Language Speech Learning: Theory, Problems, Findings." *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, edited by Winifred Strange, Timonium, 1995, pp. 233-277.

Flege, James Emil, Murray J. Munro, and Robert Allen Fox. "Auditory and Categorical Effects on Cross-Language Vowel Perception." *Journal of the Acoustical Society of America*, vol. 95, 1994, pp. 3623-3641. DOI: 10.1121/1.409931.

Fonseca-Mora, Carmen, Carmen Toscano-Fuentes, and Kathleen Wermke. "Melodies that Help: The Relation Between Language Aptitude and Musical Intelligence." *Anglistik. International Journal of English Studies*, vol. 22, no. 1, 2011, pp. 101-118.

Gonet, Wiktor, Jolanta Szpyra-Kozłowska and Radosław Święciński. "Clashes with Ashes." *Issues in Accents of English 2: Variability and Norm*, edited by Ewa Waniek- Klimczak, Cambridge Scholars Publishing, 2010, 213-228.

Gordon, Edwin. *Advanced Measures of Music Audiation*. GIA Publications, 1989.

Gordon, Edwin. *Predictive Validity Study of AMMA: A One-Year Longitudinal Predictive Validity Study of The Advanced Measures of Music and Audiation.* GIA Publications, 1990.

Gralińska-Brawata, Anna, and Paulina Rybińska. "The Relationship Between the Production of Word Stress and Musical Abilities in Polish Learners of English." *Research in Language*, vol. 15, no. 3, 2017, pp. 265-283.

Iverson, Paul, Patricia Kuhl, Reiko Akahane-Yamada, Eugen Disch, Yoh'ich Tohkura, Andreas Kettermann, and Claudia Siebert. "A Perceptual Interference Account of Acquisition Difficulties for Non-Native Phonemes." *Cognition*, vol. 87, 2003, pp. B47-B57.

Jackendoff, Ray. "Parallels and Non-parallels Between Language and Music." *Music Perception*, vol. 26, no. 3, 2009, pp. 195-204. DOI: 10.1525/MP.2009.26.3.195.

Kuhl, Patricia, and Paul Iverson. "Mapping the Perceptual Magnet Effect for Speech Using Signal Detection Theory and Multidimentional Scaling." *The Journal of the Acoustical Society of America*, vol. 97, no. 1, 1995, pp. 553-562.

Larsen-Freeman, Diane, and Michael Long. *An Introduction to Second Language Acquisition Research*. Longman, 1992.

Long, Michael. "Second Language Acquisition as a Function of Age: Research Findings and Methodological Issues." *Progression and Regression in Language*, edited by K. Hyltenstam and A. Viberg, Cambridge UP, 1993.

Miller, James. "Auditory-perceptual Interpretation of the Vowel." *The Journal of Acoustic Society of America*, vol. 85, no. 5, 1989, pp. 2114-2134.

Molivanov, Riia, Päivi Pietilä, Mari Tervaniemi, and Paulo Esquef. "Foreign Language Pronunciation Skills and Musical Aptitude: A Study of Finnish Adults with Higher Education." *Learning and Individual Differences*, vol. 20, no. 1, 2010, pp. 56-60.

Nowacka, Marta. *Phonetic Attainment in Polish College and University Students of English. A Study in the Productive and Receptive Pronunciation Skills*. Unpublished PhD dissertation, Maria Curie-Skłodowska University, Lublin, 2008.

Pastuszek-Lipińska, Barbara. "Musicians Outperform Nonmusicians in Speech Imitation." *Computer Music Modelling and Retrieval. Sense of Sounds.* Springer, 2008, pp. 56-73. Lecture Notes in Computer Science, vol. 4969. DOI: 10.1007/978-3-540-85035-9_4.

Pei, Zhengwei, Yidi Wu, Xiaocui Xiang, and Huimin Qian. "The Effects of Musical Aptitude and Musical Training on Phonological Production in Foreign Languages." *English Language Teaching*, vol. 9, no. 6, 2016, pp. 19-29.

Peltola, Terhi. "Factors Affecting the Categorizing and Rating of the Goodness of L2 Vowels." *Proceedings of the XXVII Phonetics Symposium*, 2012, pp. 48-51.

Rojczyk, Arkadiusz. "Forming New Vowel Categories in Second Language Speech: The Case of Polish Learners' Production of English /e/ and /ɪ/." *Research in Language*, vol. 8, 2010, pp. 85-97.

Rojczyk, Arkadiusz, and Andrzej Porzuczek. "Selected Aspects in the Acquisition of English Phonology by Polish Learners – Segments and Prosody." *Readings in Second Language Acquisition*, edited by Danuta Gabryś-Barker, Wydawnictwo Uniwersytetu Śląskiego, 2012, pp. 93-120.

Schleuter, Stanley. "The Relationship of AMMA Scores to Sightsinging, Dictation, and SAT Scores of University Music Majors." *Contributions to Music Education*, vol. 22, 1993, pp. 57-63.

Schön, Daniele, Cyrille Magne, and Mireille Besson. "The Music of Speech: Music Training Facilitates Pitch Processing in Both Music and Language." *Psychophysiology*, vol. 41, 2004, pp. 341-349.

Sobkowiak, Włodzimierz. *English Phonetics for Poles: A Resource Book for Learners and Teachers*. Wydawnictwo Poznańskie, 2001.

Stasiak, Sławomir. *The Acquisition of English Pronunciation in a Polish Secondary School Setting*. Unpublished PhD dissertation, Maria Curie-Skłodowska University, Lublin, 2021.

Szpyra-Kozłowska, Jolanta, Justyna Frankiewicz, Marta Nowacka, and Lidia Stadnicka. "Assessing assessment methods on the reliability of pronunciation tests in EFL." *PTLC2005*, citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.546.5252&rep=rep1&type=pdf.

Trubetzkoy, Nikolaj. *Principles of Phonology.* 1939. University of California Press, 1969.

TARGET ATTAINMENT IN THE PRODUCTION OF ENGLISH VOWELS BY POLISH STUDENTS: THE INFLUENCE OF LANGUAGE INSTRUCTION IN THE TIME OF COVID-19, IMITATION, AND MUSICALITY

S u m m a r y

This pilot experimental study pursues a number of goals. Firstly, it aims to assess the phonetic attainment of selected English vowels among twelve Polish students of English philology after one term of intensive online pronunciation training in pandemic conditions. Secondly, it looks at potential differences between production outcomes in two experimental contexts, that is, reading and imitation. Finally, it seeks to determine if there is any correlation between musicality and target attainment with a view to identifying a broader scope for potential future research questions. For this purpose, recorded samples of read and imitated English words containing vowels in a uniform context /h_d/ were assessed by six raters using a 5-point Likert scale. The results, including those of an online musicality test, were analysed and subjected to statistical testing. The majority of total scores exceed the assumed acceptability benchmark of 50%. The study yielded a number of unexpected results. Firstly, female participants performed significantly better than male ones in the reading experiment, but not in imitation. Secondly, a stronger correlation was found between the reading results and musicality than between imitation results and musicality.

Keywords: pronunciation assessment; English vowels; imitation; musicality.

OSIĄGANIE DOCELOWEJ ARTYKULACJI ANGIELSKICH SAMOGŁOSEK PRZEZ POLSKICH UCZNIÓW: WPŁYW NAUCZANIA WYMOWY W CZASACH COVID-19, IMITIACJI I MUZYKALNOŚCI

S t r e s z c z e n i e

Niniejsze pilotażowe badanie eksperymentalne jest próbą oceny osiągnięcia celów fonetycznych odnośnie do wybranych samogłosek angielskich wśród dwunastu polskich studentów filologii angielskiej po jednym semestrze intensywnego szkolenia internetowego z wymowy w warunkach pandemii. Analizie (w tym statystycznej) poddane są różnice w wynikach ocen w dwóch kontekstach eksperymentalnych, tj. czytania i naśladowania. Dodatkowo, badana jest relacja między muzykalnością a osiągnięciami docelowymi w wymowie. W tym celu nagrane próbki czytanych i imitowanych angielskich słów zawierających samogłoski w jednolitym kontekście /h_d/ zostały ocenione za pomocą 5-stopniowej skali Likerta przez sześciu wykładowców z doświadczeniem w uczeniu wymowy angielskiej na poziomie akademickim. W większości przypadków, ogólna ocena wymowy samogłosek angielskich przekracza założony poziom akceptowalności wynoszący 50%. Badanie przyniosło szereg nieoczekiwanych wyników. Po pierwsze, kobiety osiągnęły znacznie lepsze wyniki niż mężczyźni w eksperymencie polegającym na czytaniu słów, ale nie w eksperymencie opartym na imitacji. Po drugie, stwierdzono silniejszą korelację między wynikami czytania a muzykalnością niż między wynikami imitacji a muzykalnością.

Słowa kluczowe: ocena wymowy; samogłoski angielskie; imitacja; muzykalność.