

PHONETIC IMITATION OF VOWEL DURATION IN L2 SPEECH

MAGDALENA ZAJĄC

University of Łódź

zajac1234@gmail.com

Abstract

This paper reports the results of a pilot study concerned with phonetic imitation in the speech of Polish learners of English. The purpose of the study was to investigate whether native speakers of Polish imitate the length of English vowels and to determine whether the extent of phonetic imitation may be influenced by the model talker being a native or a non-native speaker of English. The participants were asked to perform an auditory naming task in which they identified objects and actions presented on a set of photos twice, with and without the imitation task. The imitation task was further sub-divided depending on the model talker being a native or non-native speaker of English (a native Southern British English speaker and a native Polish speaker fluent in English). As the aim was to investigate the variability in durational characteristics of English vowels, the series of front vowels /æ e ɪ i:/ were analysed in the shortening and lengthening b_t vs. b_d contexts. The results of the study show that the participants imitated the length of the investigated vowels as a result of exposure to the two model talkers. The data suggest that the degree of imitation was mediated both by linguistic and social factors and that the direction of convergence might have been affected by the participants' attitude toward L2 pronunciation.

Keywords: phonetic imitation, phonetic convergence, phonetic accommodation, L2 pronunciation, vowel length in L2 pronunciation, pre-fortis clipping in L2 pronunciation, Polish learners of English, social factors affecting phonetic imitation, linguistic factors affecting phonetic imitation.

1. Phonetic imitation

Phonetic imitation, sometimes termed phonetic convergence or phonetic accommodation (e.g. Babel, 2009; Pardo, 2010; Kim et al., 2011), can be defined as the process in which a talker takes on acoustic characteristics of another individual as a result of exposure to his or her speech (Babel, 2011). Many instances of this phenomenon have been reported to take place in “cooperative, socially rich, dyadic interactions” (Babel, 2011: 178). For example, Gregory and Webster (1996, in Babel, 2009) examined F0 convergence in the interviews from the Larry King Live television programme and observed that the show's host accommodated more toward guests with higher social status. Bilous and Krauss (1988, in Pardo, 2010) examined convergence in spontaneous conversations between same-gender and mixed-gender dyads. One of the observations they made was that both

male and female participants converged in average utterance length and frequency of pauses. More recently, Pardo et al. (2010) asked pairs of participants to give each other instructions and cooperate in order to complete a map task. The degree of phonetic imitation was calculated by using perceptual similarity judgments, measures of articulation rate and measures of vowel spectra. It was found that some of the subjects imitated the speech of their conversational partners and that a talker's gender and his or her role in the interaction affected the degree of phonetic convergence.

Phonetic imitation has also been found to take place in "socially minimal situations where talkers are simply producing single words" (Babel, 2011: 178). In a study by Babel (2009), the subjects read and then repeated a series of words containing different English vowels after two model talkers. The results showed that phonetic imitation did take place and that factors such as implicit racial biases and attractiveness ratings influenced the degree of convergence. In a similar study, Babel (2010) investigated whether New Zealand English speakers imitated the speech of an Australian talker. The participants performed an auditory naming task, the stimuli were single-word productions from the Australian model talker. It was found that the participants imitated the model talker and that "[s]ocial biases about how a participant feels about a speaker predicted the extent of accommodation" (Babel, 2010: 437). In Nielsen's (2011) study, the subjects listened to a model talker producing a series of words with extended VOT values and were then asked to read the words. The results indicated that after exposure to the model talker's speech, the participants produced significantly longer VOTs.

As referred to above, phonetic imitation may be affected by various social factors, such as gender, model talker's perceived attractiveness or the subject's implicit attitude towards race. As stated by Babel (2009: 23) "a talker's social knowledge and desires mediate the strength and nature of convergence in language". Nonetheless, it has been observed that phonetic imitation can also be conditioned by linguistic factors. For instance, Babel (2009) observed that participants in her study imitated /æ/ and /a/ to a greater extent than other investigated vowels. Analogously, the results of the study on NZE speakers (Babel, 2010) revealed that not all analysed vowels were imitated to the same extent. Nielsen (2011), on the other hand, discovered that lexical frequency had an effect on the degree of VOT imitation and that productions with reduced VOT were not imitated.

2. Phonetic imitation in non-native speech

An interesting issue related to phonetic imitation is whether or not it occurs in non-native speech. Kim et al. (2011) investigated phonetic convergence in conversations between subjects who had either the same or different regional dialects, and between native and non-native speakers of English. The degree of imitation was measured by asking an independent group of listeners to judge the similarity of utterance samples taken from one participant to the utterance samples taken from his or her conversational partner. The results of the study demonstrated that "sharing the same language and dialect was the only condition amongst the three language distance conditions where phonetic convergence was likely to occur" (Kim et al., 2011: 139). The authors attributed the apparent lack of phonetic convergence on the part of the non-native talkers who interacted with the native speakers to the fact that "the extra demands of second

language production may have interfered with any alignment process” (Kim et al, 2011: 143).

Conversely, in an earlier study, Beebe (1981, in Zuengler, 1991) explored the pronunciation of Chinese-Thai children and adults, who were interviewed in Thai by an ethnically Chinese and an ethnically Thai interlocutor. Beebe (ibid.) analysed the subjects’ L2 pronunciation of six Thai vowels and found that they realised five of these vowels significantly more Thai-like when talking to the Thai interviewer. Most recently, Rojczyk (2012a) demonstrated that native Polish speakers of English imitated the realisation of /æ/ when required to immediately repeat a series of words after a native English model talker. In addition, Rojczyk (2012b) found Polish learners of English imitated English VOT values under similar conditions.

3. The Current Study

Given the relative scarcity of studies on phonetic imitation in non-native speech as well as their varied results, the issue seems an interesting and important subject for research. The main purpose of the study reported here was to investigate whether native speakers of Polish imitate the length of selected English vowels. The second goal, related to the issue of phonetic convergence being mediated by various social factors, was to explore whether the extent of imitation is influenced by the model talker being a native or a non-native speaker of English. Finally, the study aimed to determine if any potential imitation was selective from a linguistic perspective, i.e. to check whether the imitation of vowel length by Polish learners of English is affected by linguistic factors.

3.1 Variables

The dependent variables under investigation were the durational characteristics of English /æ e i: ɪ/, which were analysed in the shortening and lengthening b_t and b_d contexts. Such variables were chosen as pre-fortis clipping, a feature characteristic of English pronunciation, may cause difficulties for Polish learners. As explained by Waniek-Klimczak (1998: 397):

Vowel duration is used in English at the phonological level as an inherent feature of individual vowels, enhancing the articulatory differences for individual vowel phonemes, and at the phonetic implementation level as a cue for voicing of the following obstruent; consequently, English can be claimed to be vowel-length sensitive, as compared with languages like Polish, which are vowel-length insensitive due to the lack of phonological use of inherent vowel duration or the use of vowel lengthening / shortening cue for consonant voicing.

In other words, vowel duration in shortening and lengthening contexts was selected as a variable in order to explore whether articulatory habits typical of the participants’ L1 would prevent them from imitating this L2 feature.

3.2 Participants

Twenty native speakers of Polish, twelve females and 8 males, took part in the study. The subjects were all first-year students of English Studies, recruited from the University of Lodz in Poland. Two model talkers were included in the study, one of them a native speaker of Southern British English, the other a native speaker of Polish, fluent in English but speaking with a relatively heavy foreign-accent. Both model talkers were male and in their mid-twenties.

3.3 Stimulus

The following eight monosyllabic words were selected as stimuli: *bad, bat, bed, bet, bead, beat, bid, bit*. The model talkers were recorded while reading the analysed words in carrier sentences (*I'm saying ____ again*). This was done to prevent the model talkers from using a special intonation pattern associated with reading word lists (Ladefoged, 2003). The stimuli were extracted from the recordings and presented as isolated words in the imitation (shadowing) task.

Vowel durations in the model talkers' productions were calculated, making up a total of 16 measurements. The obtained data is presented in Table 1. The abbreviations NM and NNM denote the native model talker and the non-native model talker respectively; b_d and b_t represent the voiced and voiceless contexts. As expected, the native model talker used noticeably longer vowels in the voiced context in each of the analysed pairs of words. The vowels in the non-native model's productions, on the other hand, were shorter in the voiced environments in two instances, /i:/ and /ɪ/.

vowel	NM		NNM	
	b_d	b_t	b_d	b_t
æ	140	98	145	128
e	127	77	138	94
i:	167	145	114	118
ɪ	103	81	81	105

Table 1. Vowel durations in the model talkers productions.

3.4 Procedure

The experiment consisted of three tasks: a written matching exercise, an auditory naming task, and a shadowing task. First, each participant was given a sheet of paper with the eight analysed words and a set of black-and-white photos (see Appendix A), and then asked to match the words with the photos that represented them. This exercise was designed to ascertain that the subjects knew all the words and their meaning before being asked to produce them.

Next, the eight photos from the matching exercise were presented sequentially on a computer monitor in five-second intervals. The subjects were instructed to name what they see in each photo using only the words from the written matching exercise. This was done to establish the participants' baseline productions of the analysed words. The photos were presented in a random order, which was the same for all participants.

In the shadowing task, sixteen photos were presented sequentially on a computer monitor in five-second intervals. Each photo was accompanied by either the native or the non-native model's voice pronouncing the word that was represented in the photo, i.e. each of the eight photos from the two previous task was shown twice, once with the native model's voice and once with the non-native model's voice. The photos were shown in a random order, which was the same for all subjects. As in the previous task, the participants were instructed to identify what they see, the difference being that this time they were required to listen to another person pronouncing the words before producing them themselves. It is important to note that the subjects were never explicitly instructed to imitate what they heard.

The words spoken by the model talkers in the shadowing task were presented together with the photos to make certain that the participants knew to which word they were listening at a particular moment. This was essential for the study because the non-native speaker's realization of /æ/ and /e/ closely resembled Polish /ɛ/, while his /ɪ/ and /i:/ appear to have been realized as Polish /i/. Without being able to see the photos, this would render it exceedingly difficult for the participants to determine whether the Polish model talker was saying *bad* or *bed*, *bead* or *bid*, *bat* or *bet*, etc, which, in turn, would make it impossible to correctly interpret the results of the study.

3.5 Data analysis

Vowel durations in baseline and shadowed productions were measured for each participant, giving a total of 24 vowel length measurements for each subject (8 words x 3 productions = 24 measurements). The obtained data was analysed in two stages. The first stage consisted of comparing vowel durations in the three productions (baseline, shadowing after the native model, shadowing after the non-native model). Its purpose was to examine whether the participants imitated vowel length in individual words. The second stage involved examining vowel durations in the shortening and lengthening contexts and comparing them across the three productions. This was done to establish whether exposure to the model talkers' pronunciation made the subjects modify the degree of pre-fortis clipping. The data obtained for the model talkers (see section 3.3.) was also included in the analysis as it was integral to the process of interpreting the results.

3.6 Results

The following two tables show mean vowel durations in the subjects' baseline productions contrasted with vowel durations in the model talkers' productions (NM and NNM stand for native model talker and non-native model talker respectively). The mean

durations in the baseline productions are, on the whole, noticeably longer than the durations in the model talkers' productions. The only exception is the word *beat*, in which the mean vowel duration in the baseline is very close to that of the native model. It is also worth mentioning that, as opposed to the native model talker and the participants, the non-native speaker's vowel durations are in two cases shorter in the *voiced* context (*bead* vs. *beat*, *bid* vs. *bit*).

word	vowel	baseline	NM	NNM
<i>bad</i>	æ	202	140	145
<i>bed</i>	e	194	127	138
<i>bead</i>	i:	205	167	114
<i>bid</i>	ɪ	140	103	81

Table 2. Mean vowel durations in baseline b_d productions contrasted with vowel durations in model talkers' b_d productions (in milliseconds).

word	vowel	baseline	NM	NNM
<i>bat</i>	æ	162	98	128
<i>bet</i>	e	143	77	94
<i>beat</i>	i:	148	145	118
<i>bit</i>	ɪ	138	81	105

Table 3. Mean vowel durations in baseline b_t productions contrasted with vowel durations in model talkers' b_t productions (in milliseconds).

Tables 4 and 5 show the subjects' mean vowel durations before and after hearing the native model's speech (NM stands for native model talker). The significance of the differences between the mean values was calculated by carrying out one-tailed paired-samples t-tests. The probability levels for non-chance difference between the mean values are tabulated in the last column of each table. The data in the two tables shows that the general tendency among the participants was to decrease vowel length after exposure to the native talker's pronunciation. In addition, the differences between the means are statistically significant in nearly all of the investigated words. These findings imply that the subjects systematically imitated the durational characteristics of most of the native model's vowels, except for the ones in *bid* and *beat*.

word	vowel	baseline N=20	shadowing NM N=20	p
<i>bad</i>	æ	202 (46)	160 (31)	0.000**
<i>bed</i>	e	194 (44)	160 (40)	0.001**
<i>bead</i>	i:	205 (45)	184 (33)	0.008**

word	vowel	baseline N=20	shadowing NM N=20	p
<i>bid</i>	ɪ	140 (32)	131 (29)	0.108

Table 4. Mean vowel durations in the b_d context (in milliseconds; SD given in brackets).

word	vowel	baseline N=20	shadowing NM N=20	p
<i>bat</i>	æ	162 (38)	143 (25)	0.011*
<i>bet</i>	e	143 (25)	111 (26)	0.000**
<i>beat</i>	i:	148 (36)	141 (28)	0.207
<i>bit</i>	ɪ	138 (42)	106 (21)	0.002**

Table 5. Mean vowel durations in the b_t context (in milliseconds; SD given in brackets).

Mean vowel durations before and after exposure to the non-native model's speech are presented in Tables 6 and 7 (NNM stands for non-native model talker). The results of one-tailed paired-samples t-tests are tabulated in the last column of each table and show the probability levels for non-chance difference between the mean values. As can be seen, mean vowel durations in the shadowing task are generally shorter than those in baseline productions and almost all of the differences in means are statistically significant. These results indicate that the participants converged toward the non-native model talker by decreasing vowel length. The only irregularity in the data is the lack of systematic imitation in the case of *beat*.

word	vowel	baseline N=20	shadowing NNM N=20	p
<i>bad</i>	æ	202 (46)	170 (33)	0.001**
<i>bed</i>	e	194 (44)	164 (29)	0.000**
<i>bead</i>	i:	205 (45)	162 (34)	0.000**
<i>bid</i>	ɪ	140 (32)	125 (22)	0.039*

Table 6. Mean vowel durations in the b_d context (in milliseconds; SD given in brackets).

word	vowel	baseline N=20	shadowing NNM N=20	p
<i>bat</i>	æ	162 (38)	136 (23)	0.001**
<i>bet</i>	e	143 (25)	119 (19)	0.000**
<i>beat</i>	i:	148 (36)	132 (30)	0.066
<i>bit</i>	ɪ	138 (42)	111 (26)	0.003**

Table 7. Mean vowel durations in the b_t context (in milliseconds; SD given in brackets).

Table 8 shows mean vowel durations in the shortening and lengthening contexts under three different conditions: in the baseline productions, after hearing the native model's pronunciation (shadowing NM) and after hearing the non-native model's pronunciation (shadowing NNM). One-tailed paired-samples t-tests were conducted to calculate the significance of the differences between the mean values. The probability levels for non-chance difference between the values are presented in the last column of each table. The results reveal that the subjects generally tended to shorten the investigated vowels in the voiceless context. Interestingly, although the mean length of /ɪ/ in the baseline productions is almost the same in both the shortening and the lengthening environment, in the shadowed productions the same vowel is significantly shorter in the b_t context. This may signify that the participants converged toward the native speaker by increasing the amount of pre-fortis clipping. However, the same is not true in the case of the non-native model talker, as the results show that he produced a longer /ɪ/ in the voiceless context. Also, after listening to the non-native speaker's pronunciation, the subjects used a significantly shorter /i:/ in the b_t context despite the fact that the model talker shortened /i:/ in the b_d environment. These observations suggest that, in terms of pre-fortis clipping, the subjects *converged* toward the native speaker and *diverged* from the non-native speaker.

vowel	baseline			shadowing NM			shadowing NNM		
	b_d N=20	b_t N=20	p	b_d N=20	b_t N=20	p	b_d N=20	b_t N=20	p
æ	202 (46)	162 (38)	0.000**	160 (31)	143 (25)	0.001**	170 (33)	136 (23)	0.000**
e	194 (44)	143 (25)	0.000**	160 (40)	111 (26)	0.000**	164 (29)	119 (19)	0.000**
i:	205 (45)	148 (36)	0.000**	184 (33)	141 (28)	0.000**	162 (34)	132 (30)	0.000**
ɪ	140 (32)	138 (42)	0.423	131 (29)	106 (21)	0.000**	125 (22)	111 (26)	0.031*

Table 8. Participants' mean vowel durations under three conditions (in milliseconds; SD given in brackets).

4. Discussion

One of the findings of the study was that in baseline productions the participants used considerably longer vowels than both model talkers. It seems that the large discrepancy between vowel durations stems from the fact that the model talkers were recorded while reading the analysed words in frame sentences (see section 3.3.), which might have caused them to produce the stimuli with a faster speech tempo. The participants produced the words in isolation and, as a consequence, might have used a slower tempo. This way another independent variable was unintentionally introduced in the study. One way of resolving this problem in follow-up studies would be to make the elicitation procedure the same for the model talkers and the participants, i.e. let the model talkers

familiarise themselves with the words selected as stimuli and then ask them to use the words to identify what they see in a set of photos.

As far as individual words are concerned, the results indicate that the participants imitated both model talkers by significantly decreasing vowel length in the shadowing task. Nonetheless, some inconsistencies were also found in the data. Firstly, the subjects did not significantly shorten the vowels in *bid* and *beat* after listening to the native speaker. The apparent lack of systematic imitation of the native model's *beat* could be attributed to the fact that mean vowel length in the baseline production of this word was already very close to that of the native model (see Table 2), thus providing the subjects with no context for imitation. The lack of regularity in the case of *bid*, on the other hand, is particularly intriguing. Unfortunately, no convincing explanation for this finding was found. Another interesting observation was that the subjects did not significantly decrease vowel duration in *beat* after hearing the non-native speaker. Similarly as with *bid*, it proved difficult to explain the lack of systematic convergence toward the non-native model. Nonetheless, the fact that the participants imitated vowel length in some of the investigated words to a greater extent than in others suggests that imitation might have been linguistically-selective. If so, the obtained data supports the observations made by Babel (2009, 2010) and Nielsen (2011) (see section 1).

As regards examining vowel durations across shortening and lengthening contexts, the results imply that exposure to the model talkers' pronunciation caused the subjects to modify the degree of pre-fortis clipping in their pronunciation. The participants appear to have converged toward the native speaker by increasing the vowel length difference between *bit* and *bid* in the shadowing task; they also seem to have diverged from the non-native speaker by maintaining a longer /i:/ in the b_d environment and increasing the vowel length difference between *bid* and *bit*. This indicates that the subjects might have realised that one of the model talkers spoke with a foreign accent and that the other was a native speaker of English. If so, it is probable that they diverged from the NN model in order to distance themselves from other foreign-accented speakers, whereas their convergence toward the N model was the result of a desire to sound more native-like. Hence, it seems that the direction of phonetic imitation might have been influenced by the participants' attitude toward L2 pronunciation. These findings seem to endorse the claim that phonetic imitation may be affected by social factors (Babel, 2009). In this case, it appears that imitation was to some extent mediated by the model talker's status as a native/non-native speaker of English and the subjects' desire to sound native-like. Notice also that social aspects seem to have a bearing on the degree of imitation even if the experiment takes place in "socially minimal situations where talkers are simply producing single words" (Babel, 2011: 178).

5. Conclusions

The results of the study indicate that Polish learners of English are able to imitate durational characteristics of English vowels as a result of exposure to the speech of different model talkers. The obtained data suggest that, just as in the case of native speakers, phonetic imitation in L2 speech may be selective from both a linguistic and a social perspective. It was also found that the direction of convergence may be influenced by the participants' attitude toward L2 pronunciation. Finally, the results of the study

show that phonetic imitation in non-native speech can take place in socially minimal situations.

References

- Babel, M. 2009. *Phonetic and Social Selectivity in Speech Accommodation*. Unpublished PhD dissertation.
- Babel, M. 2010. Dialect divergence and convergence in New Zealand English. *Language in Society* 39, 437–456. DOI: 10.1017/S0047404510000400
- Babel, M. 2011. Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics* 40, 177–189. DOI: 10.1016/j.wocn.2011.09.001
- Beebe, L. 1981. Social and Situational Factors Affecting the Communicative Strategy of Dialect Code-Switching. *International Journal of the Sociology of Language* 32, 139-149.
- Goldinger, S. 1998. Echoes of Echoes? An Episodic Theory of Lexical Access. *Psychological Review* 105, 251- 279.
- Honorof, D., Weihing, J. and Fowler, C.A. 2011. Articulatory events are imitated under rapid shadowing. *Journal of Phonetics* 39, 18–38. DOI: 10.1016/j.wocn.2010.10.007
- Kim, M., Horton, W. S. and Bradlow, A. R. 2011. Phonetic convergence in spontaneous conversations as a function of interlocutor language distance. *Laboratory Phonology* 2, 125-156. DOI: 10.1515/LABPHON.2011.004
- Ladefoged, P. 2003. *Phonetic Data Analysis. An Introduction to Fieldwork and Instrumental Techniques*. Oxford: Blackwell Publishing.
- Nielsen, K. 2011. Specificity and abstractness of VOT imitation. *Journal of Phonetics* 39, 132-142. DOI: 10.1016/j.wocn.2010.12.007
- Pardo J. 2010. Expressing oneself in conversational interaction. In Morsella, E. (ed) *Expressing oneself/expressing one's self: Communication, cognition, language, and identity*. London: Psychology Press, 183–196. DOI: 10.3758/APP.72.8.2254
- Pardo, J., Cajori Jay, I. and Krauss, R. M. 2010. Conversational role influences speech imitation. *Attention, Perception, & Psychophysics* 72 (8), 2254-2264.
- Rojczyk, A. 2012a. *Spontaneous phonetic imitation of L2 vowels in a rapid shadowing task*. Poster presented at PSLLT 2012 - Pronunciation in Second Language Learning and Teaching Conference, Vancouver, Canada, 24-25 August.
- Rojczyk, A. 2012b. *Phonetic and phonological mode in second-language speech: VOT imitation*. Paper presented at EuroSLA22 – 22nd Annual Conference of the European Second Language Association, Poznań, Poland, 5-8 September.
- Waniek-Klimczak, E. 1998. Sources of variability in L2 phonology: vowel duration in 'Polish English'. In Catherine Gruber, C., Higgins, D., Olson, K. and Wycsocki, T. (eds) *CLS 34/2: The Panels*. Chicago Linguistic Society, 387-401.
- Zuengler, J. 1991. Accommodation in Native-Non-native Interactions: Going Beyond the “What” to the “Why” in Second-Language Research. In Coupland, J., Coupland N. and Giles H. (eds) *Contexts of Accommodation: Developments in Applied Sociolinguistics*. Cambridge: Cambridge University Press, 223-244.

Appendix A

The written matching exercise

Dopasuj poniższe angielskie słowa do przedstawiających je obrazków:

bead bad bat beat bet bid bit bed



a).....



b).....



c).....



d).....



e).....



f).....



g).....



h).....