Vowel Imitation in Spontaneous Phonetic Imitation

Molly Babel
Dept. of Ling, University of British Columbia, Totem Fields Studio, 2613 West Mall, Vancouver, BC V6T 1Z4,
mbabel@berkeley.edu

1. INTRODUCTION

Research on phonetic imitation or phonetic convergence has established that talkers naturally accommodate during interaction (Goldinger 1998, Namy et al. 2002, Pardo 2006). Traditionally, the method of determining imitation has been an AXB judgment task where listeners judge the perceptual similarity of a participant's A and B tokens from a pre-task production and shadowed production, for example, to X – a production from a model talker in the task. At better than chance levels, listeners judge the shadowed production to be more similar to the model talker's production. This method allows listeners to make use of an array of perceptual cues for imitation. It is left unknown, however, what within the phonetic structure is being imitated. What is and what can be imitated have significant implications for the level of detail in linguistic representation. More recently, work by Shockley et al. (2004) and Nielsen (2008) has demonstrated that American English participants imitate VOT. The purpose of the current project is to examine whether spectral characteristics of the vowels are imitated in a lexical shadowing task.

2. METHOD

Fifty monosyllabic low frequency words with the vowels /i æ ɑ o u/ were used as stimuli in a lexical shadowing task. Native speakers of American English ($n = 113$) completed an auditory naming task where they shadowed productions from one of two model talkers. One talker was Black and the other was White.

Participants in the shadowing task were tested individually and were randomly assigned to one of four conditions: Black talker/No Picture, Black talker/Picture, White talker/No Picture, and White talker/Picture. The paradigm for the speech production task is a shadowing paradigm like that of Goldinger (1998). Participants were seated in a sound-attenuated room at a computer workstation. Participants wore a head-mounted AKG C520 microphone positioned about 2 inches to the side of the mouth and AKG K240 headphones. Word productions were digitally recorded to the hard drive of a PC at a 44K sampling rate. Recordings were down sampled to 22K before analysis.

2.1 Procedure

The task proceeded as follows: The first block was a Pre-task Block to establish participants' baseline productions. The words were presented randomly in 36-point font in the middle of the screen. Participants were instructed to read the words as naturally and clearly as possible. In the test blocks, the randomized word list was presented binaurally at 65 dB (SPL) over the headphones. The test blocks were comprised of three shadowing blocks where words were repeated twice per block. Participants were told that upon hearing the word, they were to repeat it as clearly and naturally as possible. In the Picture Conditions a talker photo was presented on the screen for the duration of the shadowing portion of the task. The post-test block was identical to the pre-test block where participants read the words from the screen. After the task participants in the No Picture conditions were asked to identify the race of the talker. Both talkers were identified as White by participants ($\chi(1) = 0.08$, $p = $ n.s.).

2.2 Data Analysis

First and second formants were extracted from word productions from a series of Gaussian windows spanning the middle 50% of the vowel with a 2.5 ms step size. Formant values were normalized using the Lobanov normalization routine (Lobanov 1971).

The Euclidean distance was calculated from each participant production to that of the model talker. These calculations are a measure of acoustic distance between the model talkers' productions and the participants' productions. To calculate how much a participant modified their production as a result of being exposed to the model talker the original distance for each word was subtracted from the distance for each following instance of that word. The value calculated is the difference in distance (DID). A negative DID value demonstrates that the phonetic distance between the participant and the model talker shrank. A positive value indicates an increase in phonetic distance. A value of 0 demonstrates that there was no change as the result of auditory exposure to the model talker. This DID value is used as the dependent measure in the statistical analysis.

3. RESULTS

DID values were summarized across cells and the means were used in a repeated measures analysis of variance. DID was the dependent variable; Talker Race, Picture/No Picture, and Gender were independent variables; and Vowel and Block were repeated measures. Vowel [$F(4, 396) = 56.2$, $p < 0.001$] and Block [$F(3, 297) = 60.7$, $p < 0.001$] returned as main effects. There was also a two-way Vowel x Block interaction [$F(12, 1188) = 20.2$, $p < 0.001$] and a three-way Vowel x Block x Talker Race interaction [$F(12, 1188) = 3.8$, $p < 0.001$]. There were also two-way interactions of Vowel x Gender [$F(4, 396) = 2.4$, $p < 0.05$], Talker Race x Vowel [$F(4, 396) = 8.3$, $p < 0.001$], and Picture x Gender [$F(1, 99) = 4$, $p < 0.05$]. The three-way

interaction between Picture x Vowel x Gender was also significant [$F(4, 396) = 5$, $p < 0.001$].

Figure 1 shows the effect of selective imitation for low vowels. Post-hoc tests show /◌/ and /◌/ are imitated more than /i o u/ ($p < 0.001$). In addition, /◌/ is imitated more than /◌/ ($p < 0.05$). With respect to the Block effect, post-hoc tests find that imitation is cumulative across shadowing blocks. There was more imitation in Block 5 than 4 ($p < 0.05$) and more in Block 6 than 4 ($p < 0.001$). During the shadowing task, there was more accommodation than the post-task block ($p < 0.001$). Normalized formant plots from male and female participants are shown in Figure 2. These figures indicate that the majority of vowel imitation comes from changes within the F1 dimension.
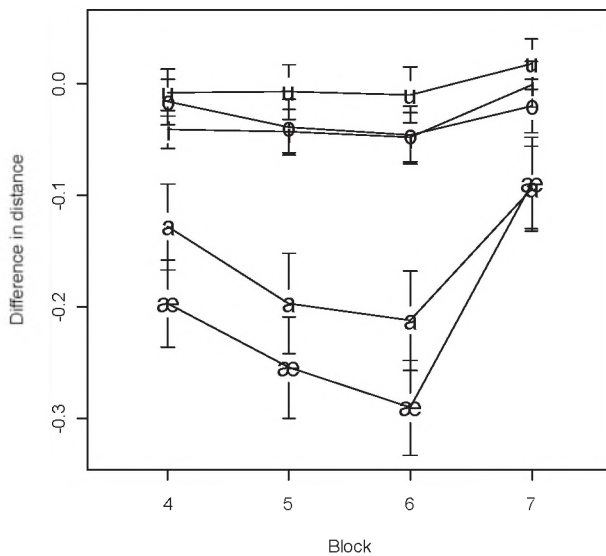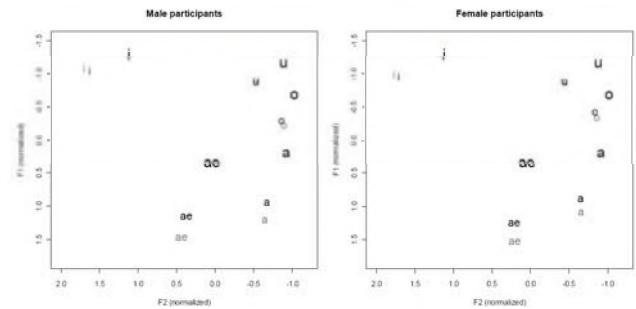

Figure 2. Formant plot displaying the direction of spontaneous phonetic imitation. The mean of the model talkers' vowels are in the slightly larger font and in black. Participants' pre-task vowel means are plotted in a smaller light gray font and their productions from the final shadowing block are in the small black print.

Figure 1. Spontaneous phonetic imitation for all participants by Vowel and Block. The DID measure on the *y*-axis indicates the amount of phonetic imitation. A negative value demonstrates phonetic imitation and a positive value demonstrates vocalic divergence. Blocks 4, 5, and 6 are Shadowing Blocks while Block 7 is the Post-task Block..

4. DISCUSSION

The results show that talkers accommodate the first and second formants of the model talker in the task, but that not all vowels are imitated to a significant degree. Only the low vowels /◌/ and /æ/ exhibit strong imitation effects, and this effect lies primarily within the F1 dimension. I argue that this is due to the increased repertoire of production variants talkers store for low vowels as a result of the difference in jaw height in accented and unaccented environments (Summers 1987, de Jong 1995).

An important aspect of this result is that it demonstrates the labile nature of linguistic segments with respect to both their perceptual encoding and their variation in production. First, listeners must perceive the detailed acoustic structure of an utterance in order to have those details influence their production. Second, in speech production, participants alter the characteristics of the output without modifying the categorical identity of the segment they produce. In sum, the exact selection of a speech production variant is determined by auditory exposure.

REFERENCES

de Jong, K. (1995). The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation. *JASA* 97:491-504.

Goldinger, S. (1998). Echoes of echoes? An episodic theory of lexical access. *Psych. Review*, 105:251–279.

Lobanov, B. (1971) Classification of Russian vowels spoken by different listeners. *JASA*, 49: 606-08

Namy, L., L. Nygaard, & D. Sauterberg. (2002). Gender differences in vocal accommodation: the role of perception. *J of Lang &SocPsych.*, 21:422–432.

Nielsen, Kuniko. (2008). *The specificity of allophonic variability and its implications for accounts of speech perception.* Doctoral Dissertation, UCLA.

Pardo, J.S. (2006). On phonetic convergence during conversational interaction. *JASA*, 119:2382–2393.

Shockley, K., L. Sabadini, & C.A. Fowler. (2004). Imitation in shadowing words. *P&P*, 66:422–429.

Summers, W.V. (1987). Effects of stress and final-consonant voicing on vowel production: Articulatory and acoustic analyses. *JASA* 82:847– 863.

ACKNOWLEDGEMENTS