phonetic imitation of Japanese vowel devoicing

Kuniko Nielsen
Linguistics Department, Oakland University, USA
nielsen@oakland.edu

Abstract

Recent studies have shown that talkers implicitly imitate/accommodate the phonetic properties of recently heard speech [1, 2]. However, it has also been shown that this phonetic imitation effect is not an automatic process [3, 4]; in [3], the artificially lengthened VOT on /p/ was imitated in a non-shadowing task, while shortened VOT (which could jeopardize phonemic contrast) was not imitated, suggesting that phonetic imitation is sensitive to existing phonemic contrast. This paper explores whether phonological factors unrelated to contrast preservation also affect imitation of phonetic details, specifically, Japanese vowel devoicing. The results revealed significant imitation of Japanese devoicing, indicating that perceived fine phonetic details are imitable and can subsequently affect speech production even in phonologically constrained environments.

Index Terms: Imitation, accommodation, Japanese vowel devoicing

1. Introduction

Recent studies have shown that talkers implicitly imitate/accommodate the phonetic properties of speech presented in the form of experimental stimuli [1, 5] and of the speech produced by their interlocutors [2], indicating that perceived fine phonetic details are retained in listeners' memories, and subsequently affect their speech production. At the same time, it has also been shown that phonetic imitation is not an automatic process [3, 4]. In [3], the artificially shortened VOT on /p/ (which could jeopardize phonemic contrast with /b/) was not imitated, while lengthened VOT was imitated, suggesting that knowledge of phonemic contrast modulates the phonetic imitation.

The current study explores the extent to which phonological factors unrelated to contrast preservation also affect imitation of fine phonetic details. Many phonological factors are unrelated to contrast preservation and are weakly realized in grammar. If such factors constrain imitation of phonetic details, it will suggest that phonetic imitation occurs only in the absence of phonological factors. In this case, the episodicity previously observed in phonetic imitation would have few implications in terms of phonological representations. On the other hand, if imitation is observed despite the presence of phonological factors, it will provide a stronger argument for the episodic nature of phonological representations. We address this question by examining the imitation patterns of Japanese vowel devoicing.

In Tokyo Japanese, short high vowels tend to be devoiced or deleted when they occur between two voiceless consonants [6]. This process has traditionally been considered a phonological assimilation of the feature [+voice] [6]. An alternative account was proposed in [7], which considers Japanese vowel devoicing a gradient phonetic process involving overlap of glottal gestures. The precise phonetic implementation of Japanese vowel devoicing is not homogeneous, ranging from shortened (voiced) vowel to voiceless vowel, to pure deletion [8]. Previous studies have shown that devoicing can be optional, with a number of both phonetic and phonological factors affecting the rate of devoicing. Consonantal context, especially manner of articulation, is known to affect the devoicing rate [9, 10]. Vowel quality is also claimed to condition devoicing [11]. Other phonological factors known to influence the likelihood of devoicing include syllable structure and position, morphological/word boundary, moraic position, and presence of pitch accent [12, 6, 13, 14, respectively]. Japanese vowel devoicing is suitable for our purpose because: 1) unlike shortened VOT, its imitation does not endanger phonemic contrast as shown in [15]; 2) unlike lengthened VOT, its imitation in certain environments would violate well-established phonetic and phonological constraints just described, 3) similar to VOT, Japanese devoicing is a gradient process and is quantifiable [7]. Further, it does not have an overt association with a stylistic register such as the association of long VOT with “careful” speech: for example, [16] showed that professional teachers of hearing-impaired children reduced their vowel devoicing in order to improve their listeners’ comprehension.

2. Method

An experiment was carried out to test: 1) whether phonetic imitation of Japanese devoicing can be observed, 2) whether the imitation can be generalized across words to which subjects were not exposed [3], and 3) whether the word-level specificity effect of imitation can be observed, as in [1]. The feature manipulated in the experiments was [+ spread glottis] on the phoneme /u/.

2.1. Participants

Twenty four native monolingual speakers of Tokyo Japanese (age 18–50, mean 28.25, 12 females) with normal hearing served as subjects for this experiment. None of the participants had lived outside of the greater Tokyo area (Tokyo, Chiba, Kanagawa, and Saitama), and none had daily exposure to foreign languages. They were recruited from the TDFS (the Tokyo University of Foreign Studies) population, and were paid for their participation.

2.2. Stimuli Selection and Construction

The production list consisted of 110 Japanese words: 1) 30 Target words which were played in the listening phase (20 words containing devoiceable /u/ and 10 words containing devoiceable /e/ and /o/), 2) 60 Novel words which were NOT played in the listening phase (20 words containing devoiceable /u/, 20 words containing devoiceable /i/, and 20 words containing devoiceable /i/ and /u/, and 3) 20 Filler words containing no devoicing environment. The Novel (= unheard) words are added to test the generalizability of imitation, as in
2.3. Procedure

voiced vowels were 138 ms and 114 ms, respectively. The consonants preceding complete devoicing and (devoiceable) eliminate all the voiced portions (i.e., /fukutsu/ 'persistence', were potentially devoiceable and thus it was not possible to low vowel /a/ and the two exceptional cases where all vowels with no voiced portion of any devoiceable vowel, excluding (i.e., vowel devoicing was mostly realized as a deletion). As a although they were extremely rare in the speaker's production tokens had no partial devoicing. Voiceless vowels were left in, devoicing environments were spliced out so that the resulting environments and were indeed devoiced. To make the Target /o/ words) were never deleted in devoicing On the other hand, mid vowels (in 5 Target /e/ words and 5 /zokushutsu/ [zoku 5 Target /u/ words, the speaker produced only one [u /u/ after listening to the target speech with extreme vowel devoicing. Voiceless vowels were extremely rare in the data and most cases of complete devoicing were indeed realized as complete deletion. When voiceless vowels occurred, they were treated as a part of the preceding consonant. In order to examine the results phonetically as well as phonologically, the degree of devoicing was quantified in two ways: (1) A gradient measure by measuring the duration of devoiceable (yet voiced) vowels, and (2) a binary measure by calculating the fraction of devoiceable vowels that were completely deleted or devoiced.

Statistical by-subjects analysis of the data was performed based on repeated-measures ANOVA. The dependent variables in the gradient and categorical analyses were (voiced) vowel duration and rate of complete vowel devoicing and deletion, respectively. The independent variables (all within-subject factors) were 1) Production Type (Baseline vs. Post-Exposure), 2) Vowel (/u/, /i/, /e/, /o/), 3) Lexical Frequency (High vs. Low), and 4) Presence of Exposure (Target vs. Novel). As described above, the target speech did not contain any voiced vowels in devoicing environments (except for two cases described earlier), and thus the targets of imitation with regard to the duration of devoiceable vowels (gradient) and the rate of complete devoicing (categorical) were 0 ms and 100%, respectively.

3. Results

3.1. Phonetic Imitation

Gradient Analysis: The extreme degree of devoicing in the target speech was imitated: A significant main effect of Production Type [F(1,22)=45.72, p<0.01*] ( = imitation) was found in the gradient analysis. As expected, the mean duration of devoiceable vowels decreased significantly from 53 ms (baseline production) to 44 ms (post-exposure production). (The average whole-word duration also decreased from 549 ms to 532 ms, while average consonant duration increased from 103 ms to 106 ms.) The effect of Vowel was also significant [F(1,22)=3.76, p=0.065*], while the interaction between the two variables did not reach significance [F(1,22)=2.95, p>0.05]. A large individual variability was also observed in the magnitude of imitation. Figure 1 illustrates the speaker variability of vowel duration and degree of imitation. As seen, although most speakers decreased the duration of devoiceable /u/ after listening to the target speech with extreme vowel
deletion/devoicing, the degree of change (or imitation) varied greatly across participants.

**Categorical Analysis:** The extreme degree of devoicing was categorically imitated as well: there was a significant effect of Production Type \[F(1,22)=23.98, p<0.01^*\] and Vowel \[F(1,22)=51.83, p<0.01^*\]. The rate of complete devoicing increased in post-exposure production by 5%. The interaction between the two variables was significant \[F(1,22)=5.67, p<0.05^*\]. This is mainly due to the low degree of imitation among mid-vowels, which was expected. In addition, a clear difference between /u/ and /i/ was observed. Although this result is in agreement with [11], the current experiment was not designed to compare the baseline difference between the high vowels, and this point is not discussed further.

**3.2. Generalizability**

The generalizability of phonetic imitation was examined at two levels: phoneme level (Target /u/ vs. Novel /u/) and feature level (Novel /u/ vs. Novel /i/).

**3.2.1. Phoneme-level Generalization**

Phoneme-level Generalization (Target /u/ vs. Novel /u/ comparison) was examined by two repeated-measures ANOVA analyses: gradient and categorical. The independent variables for both analyses were Production Type and Presence of Exposure. According to the gradient analysis, the effect of Production Type was significant \[F(1,22)=24.37, p<0.01^*\], revealing that phonetic imitation occurred in both Target /u/ and Novel /u/. The effect of Presence of Exposure was not significant \[F<1, p>0.1\], while the interaction between the two was significant \[F(1,22)=4.78, p<0.05^*\], indicating that imitation was greater for Target /u/. The categorical analysis revealed significant main effects of Production Type \[F(1,22)=6.04, p<0.05^*\] and Presence of Exposure \[F(1,22)=67.81, p<0.01^*\]. The interaction between the two factors was not significant \[F<1, p>0.1\].

**3.2.2. Subphonemic Generalization**

Generalization at the subphonemic level was examined by comparing the imitation patterns of Novel /u/ and Novel /i/ (the independent variables: Production Type, Vowel (Novel /u/ and Novel /i/)). Phonetic imitation of extreme devoicing was generalized at a subphonemic level: The results of the gradient analysis revealed significant main effects of both Production Type \[F(1,22)=47.30, p<0.01^*\] and Vowel \[F(1,22)=7.31, p<0.01^*\]. The interaction between the two variables was also significant \[F(1,22)=9.91, p<0.05^*\]. Note, however, that the observed interaction between the degree of imitation and vowel quality (i.e., Novel /u/ and Novel /i/) does not show phoneme-level imitation (i.e., stronger imitation for /u/), because the degree of imitation was actually larger for /i/. The results of the categorical analysis also revealed generalization of phonetic imitation at a subphonemic level: The effects of Production Type \[F(1,22)=22.51, p<0.01^*\] and Vowel \[F(1,22)=388.69, p<0.01^*\] were both significant, while the interaction between the two variables was not \[F(1,22)=1.123, p>0.1\]. Figure 2 illustrates the result of the categorical analysis for Novel /u/ vs. Novel /i/: As seen, the rate of voicing increased in both types of stimuli.

**3.3. Word Specificity**

Lexical Frequency and Presence of Exposure were controlled in order to examine the effect of word-level specificity on the degree of imitation. Word specificity would mean that low frequency words as well as words with more exposure show a stronger imitation effect, as shown in [1] and [3].

**3.3.1. Lexical Frequency**

The effect of Lexical Frequency on the degree of imitation was examined in gradient and categorical analyses. The independent variables were Production Type and Lexical Frequency. The gradient analysis found significant main effects of both Lexical Frequency \[F(1,22)=4.61, p<0.05^*\] and Production Type \[F(1,22)=45.87, p<0.01^*\], while the interaction between the two was not significant \[F<1, p>0.1\]. The categorical analysis also revealed significant main effects of Lexical Frequency \[F(1,22)=24.89, p<0.01^*\] and Production Type \[F(1,22)=5.13, p<0.05^*\], while there was no interaction between the two factors \[F<1, p>0.1\]. In sum, there was no evidence that lexical frequency influences the degree of imitation. The observed main effect of lexical frequency was not expected, although it is in agreement with [9].

**3.3.2. Presence of Exposure**

Next, the effect of Presence of Exposure (comparison between Target and Novel words) on the degree of imitation was examined. The independent variables for both analyses were
production type and presence of exposure, while the dependent variables for the gradient and categorical analyses were the duration and rate of complete devoicing of /u/, respectively. As described in 3.2.1 (phoneme-level generalization), the effect of presence of exposure was not significant, while that of production type was significant, and the interaction between the two was significant, revealing the word-specific pattern of phonetic imitation. Also as described in 3.2.1 (phoneme-level generalization), there were significant main effects of presence of exposure [F(1,22)=67.8, p<0.01*] and production type [F(1,22)=6.04, p<0.05*] in the categorical analysis, while there was no interaction between the two factors [F=1, p>0.1]. In sum, word specificity was found only in the gradient analysis which examined presence of exposure. When the vowel durations were compared, the data revealed a greater degree of imitation for target words than novel words. On the other hand, the degree of imitation did not differ between low and high frequency words. As for the rate of complete devoicing, neither exposure to target speech nor lexical frequency showed evidence for word specificity.

4. Discussion

Our results revealed that the extreme degree of Japanese vowel devoicing/deletion in the target speech was imitated by the participants, and that the imitation was generalized to new instances of the modeled vowel /u/ as well as to a new vowel /i/ in both categorical and gradient analyses. In post-exposure production, the average rate of complete devoicing increased from 62% (baseline) to 67% (post-exposure), and the average duration of vowels in devoicing environments decreased from 53 ms (baseline) to 44 ms (post-exposure). A word-specific pattern of imitation was observed only in a gradient analysis comparing target and novel items, while there was no effect of lexical frequency (on the degree of imitation) in the data. More importantly, the data showed that phonetic imitation occurs in environments where various phonetic/phonological factors control the output, and that imitation affects the categorical allophonic realization (i.e., presence/absence of a vowel) in addition to the gradient phonetic realization (i.e., shorter vowel duration). These results suggest that details of perceived allophonic variability indeed have consequences for phonological representations.

As for the mechanism of Japanese vowel devoicing, our data suggest that it is a gradient phonetic process: A phonological voice assimilation process would predict the presence of visible devoiced vowels, which was absent in most cases in our data. Further, our data showed many cases of partial devoicing (acoustic vowel shortening) as well as longer consonant durations preceding complete devoicing, neither of which would be predicted by an assimilation account. In contrast, these patterns are in agreement with the gradient phonetic account [7].

5. Conclusions

The current study showed that the extreme degree of Japanese vowel devoicing (realized as vowel deletion) of the modeled vowel /u/ was imitated in environments where various phonetic/phonological factors unrelated to contrast preservation control the output. Further, the change was generalized to novel items with /i/, suggesting that the change in speech signal was coded at a subphonemic level. On the other hand, the degree of imitation was shown to be at least in part word-specific. These results suggest that perceived allophonic information is retained after lexical access, which could have consequences for phonological representations. As for the mechanism of Japanese vowel devoicing, our results were more compatible with the gradient phonetic account than the phonological feature assimilation account.

6. Acknowledgements

The research was supported by NSF Dissertation Improvement Grant BCS-0547578 (PI: Patricia Keating) and a UCLA Dissertation Year Fellowship. The author would like to thank Professor Peri Bhaskararao and Dr. Yuke Abe at TUF's for their generous help with data collection.

7. References