IMPLICIT PHONETIC IMITATION IS CONSTRAINED BY PHONEMIC CONTRAST

Kuniko Y. Nielsen

Department of Linguistics
University of California, Los Angeles, USA
kuniko@humnet.ucla.edu

ABSTRACT

The imitation paradigm [1] has shown that subjects shift their production in the direction of the target, indicating the use of episodic traces in speech perception. By using this paradigm, two experiments were carried out to test: 1) if/how this implicit phonetic imitation interacts with linguistic representations when the change might impair linguistic contrast; 2) whether phonetic imitation can be *generalized*, and 3) whether word-level specificity can be obtained through physical measurements of a phonetic feature. The results revealed a significant effect of implicit phonetic imitation for extended VOTs, although there was no imitation observed for reduced VOTs. Furthermore, the imitated feature (extended VOT) was generalized to new instances of the target phoneme /p/ as well as to the new segment /k/. These results indicate that 1) knowledge of phonemic contrast modulates the implicit phonetic imitation, and 2) speakers possess sub-phonemic representations. Expected word specificity was not observed in the data.

1. INTRODUCTION

Recent studies have shown that traces of episodic memory are retained and used in speech perception [2], and that both speech perception and production are more plastic than previously considered (e.g., [3, 4]). The implicit imitation paradigm, in which subjects' speech is compared before and after they are exposed to target speech, has shown that speakers shift their productions in the direction of what they just heard. For example, Goldinger [1] showed that subjects shifted their own F0 when they are asked to shadow (= immediately repeat) speech with manipulated F0. His result also revealed a word-specific advantage of imitation: larger imitation effects were observed among low-frequency words than high-frequency words. This is as predicted by exemplar-based theories, because the smaller the number of exemplars associated with a given word, the larger the weight of each new exemplar. [5] extended [1] by showing a significant Voice-Onset-Time (VOT) imitation effect (in shadowing) for voiceless stops with artificially extended VOTs. These studies show listeners' sensitivity to variations in global phonetic dimensions (i.e., F0) as well as the fine phonetic detail of a single segment (i.e. VOT).

Our interest is to determine the automaticity of implicit phonetic imitation, by comparing how two types of modeled stimuli are imitated. In addition to extended VOT (as in [5]), the current study employs reduced VOT target stimuli. Unlike extended VOT, reduced VOT in voiceless stops could introduce linguistic ambiguity. If implicit phonetic imitation is an automatic process, knowledge of phonemic contrast should not constrain the effect and thus we would expect similar imitation between the two types of modeled stimuli. On the other hand, if different degrees of imitation are observed, it will suggest that the imitation is not automatic, but a more complicated process that is filtered by linguistic knowledge.

Although [1] shows evidence for word-size representations, it does not reveal whether sub-lexical units were also influenced by the imitation effect. The present study extends the earlier studies by using a *non-shadowing* task, which lets the listening and production lists differ; thus unheard words can be introduced into the production list. This allows us to test the generalization of the imitation to sub-lexical units, namely phonemic and sub-phonemic representations. Lastly, we will examine whether word-level specificity can be demonstrated.

2. METHOD

Participants: Thirty-nine native speakers of American English with normal hearing served as subjects for this experiment. They were recruited from the undergraduate population at UCLA, and received course credit for participating.

Stimuli: The production list consisted of 150 English words. 100 were words beginning with /p/ (80 target words: 40 high-frequency words and 40 low-frequency words which were played in the listening phase, and an additional 20 low-frequency words which were not played during the listening phase), and 20 were low-frequency words beginning with /k/. The remaining 30 words began with sonorants and served as fillers. The listening list consisted of 120 English words, including 80 target words from the production list (40 high-frequency words and 40 low-frequency words beginning with /p/), and 40 filler words beginning with sonorants. Lexical frequency was determined from [6, 7]. The phonological neighborhood density and syllable length were controlled between the two frequency groups. All the words had equally high familiarity (> 6.0 on the 7-point scale from [8]). All the target words had initial stress, and there were no onset clusters. A phonetically trained male American English speaker recorded the 80 target words in the production list. The speaker first produced the words in the list normally, and then with extra aspiration. The mean VOT of normally produced initial /p/ was 72.46 ms (SD=12.14). For Group 1 stimuli (with extended VOT), the normally produced tokens (including the transition between aspiration and voice onset) and the initial parts of hyper-aspirated tokens were spliced so that all the target words' VOTs were *extended* by 40ms (mean=113.26 ms, SD=10.82). For Group 2 stimuli (with reduced VOT), the most stable part of aspiration for the normally produced target words (starting with /p/) was taken out so that all the target words' VOTs were reduced by 40ms from the original tokens (mean=32.29ms, SD=12.39ms). To assure that these tokens still sounded like the target words (i.e., initial phoneme being /p/ as opposed to /b/), two native English speakers were asked to listen to the words, and record what they thought they heard. Every word was heard as /p/-initial by both listeners.

Procedure: The experiment used a modified version of the imitation paradigm [1], in that a warm-up reading phase was added at the beginning to avoid possible hyper-articulation. The stimuli were presented using Psyscope 1.2.5. Each subject was seated in front of a computer in a sound booth. Each session was divided into 4 blocks: warm-up, baseline, listening, and test. In the warm-up block, the words were presented, one at a time, on a computer screen every 2 seconds. The subjects read the words silently. In the baseline block, the subjects were instructed to "identify the word you see by speaking it into the microphone." In the listening block, using headphones, the subjects were exposed to two repetitions of the 120 spoken word tokens (with no additional task). The test block was the same as the baseline block. Across the four blocks, the words were presented in random order for each subject. The subjects' tokens were digitally recorded into a computer and VOTs were measured using both waveforms and spectrograms. Unlike in previous studies, there was no perceptual assessment (e.g. AXB testing) of the baseline versus test productions.

3. RESULTS

Within-subject factors in this study were:

Type of Production: Baseline vs. Test
Lexical Frequency: High vs. Low
Presence of Exposure: Target vs. Novel Items
Segment: /p/ vs. /k/

The between-subject factor in this study was:

Listening Stimuli: Extended VOT [Group 1] vs. Reduced VOT [Group 2]

Repeated-measures ANOVA analysis with two factors (Type of Production and Listening Stimuli) revealed a clear interaction (F(1,75) = 23.99, *p* <.001*) between them. The baseline values of the 2 groups were equivalent, while the degree of imitation was dependent on the listening stimuli. For this reason, the two groups' data were analyzed separately.

Group 1: A repeated-measures ANOVA with two within-subjects factors (Type of Production and Lexical Frequency) revealed a significant difference for both factors: F(1,19)=13.13, *p*<.01*, and F(1,19)=4.79, *p*<.05*, respectively. However, the interaction between the two was not significant (F(1,19)=0.08, *p*>.1). Another repeated-measures ANOVA analysis with two within-subjects factors (Type of Production and Presence of Exposure [= target vs. novel]) showed a significant difference for both factors (F(1,19)=11.99, *p*<.01*, and F(1,19)=13.22, *p*<.01*, respectively), while the interaction between the two factors was not significant (F(1,19)=.39, *p*>.1). Next, in order to see how the imitation effect is generalized to new stimuli, a repeated-measures ANOVA with two within-subjects factors (Type of Production and Segment) was performed. Note that neither group of words tested here was played in the listening block. There was a significant difference between pre- and post-exposure productions (F(1,19)= 17.08, *p*<.01*), and between /p/ and /k/ (F(1,19)= 217.85, *p*<.001*), as expected (e.g.[9]), while there was no interaction between the two factors (F(1,19)=.82, *p*>.1).

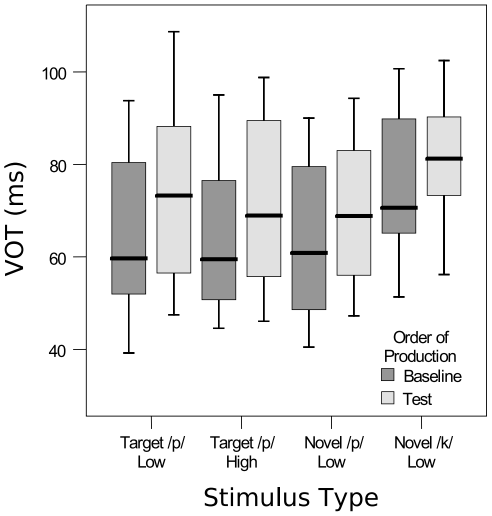

Figure 1: Group 1 imitation effects (in VOT) plotted across four types of stimuli. The subjects listened to stimuli with LONGER VOT. The horizontal bar represents the median, the box represents the 25th - 75th percentile range, and the whisker represents the range of observation.

Group 2: Repeated-measures ANOVA analyses for Group 2 found no significant differences. That is, neither imitation nor generalization was found in Group 2. As can be seen in Figure 2, being exposed to target speech did not shift subjects' VOT in the way it did in Group 1.

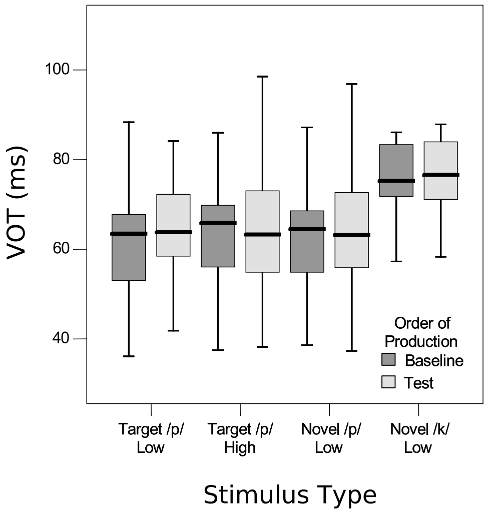

Figure 2: Group 2 imitation effects (in VOT) plotted across four types of stimuli. The subjects listened to stimuli with SHORTER VOT.

4. DISCUSSION

Our results from Group 1 revealed a significant effect of implicit phonetic imitation. As seen in Figure 1, test-productions (light bars) show consistently longer VOTs than baseline productions (dark bars), revealing that the VOT imitation effect is present in non-shadowing elicitation-style production. The data also showed that the imitation effect was generalized to novel stimuli that subjects did not hear during the listening block (see *Novel /p/* in Figure 1). This result indicates that the locus of spontaneous phonetic "imitation" is not word-specific, and that subjects imitated something smaller than a word. The imitation effect was also generalized to a new phoneme /k/, which shares the manipulated feature [+spread glottis]. This result indicates that subjects imitated a unit that is smaller than the phoneme, suggesting subjects' knowledge of sub-phonemic representation. On the other hand, there was no interaction between the tested segment and Type of Production (F<1, *p*>0.1), indicating that the amount of imitation observed for novel /p/ and /k/ was the same. Thus, this result does not provide support for phoneme-level representation.

In order to determine if the imitation is simply due to global changes in speech style, a post-hoc analysis of whole-word duration was conducted. If the effect is due to episodic memory or rule-learning, only the manipulated variable (in this case, VOT) should be affected. The whole-word duration of the low-frequency target words was measured from 8 randomly chosen Group 1 subjects' data. Unlike with VOT, there was no significant difference between baseline and test productions (F<1). Given these results, it is unlikely that global aspects of speech are solely responsible for the phonetic imitation observed in this study, again localizing the effect in a sub-lexical unit.

In contrast, the subjects in Group 2 (exposed to reduced VOT) did not show a significant difference between baseline and test productions in their VOT across all types of stimuli (see Figure 2). Data from the baseline recordings of both groups show that the distribution of VOT for /p/ ranges from around 20 ms to 130 ms, centering around 65 ms (mean 65.04 ms, median 63.71 ms). If we assume that these data represent the real-life distribution of VOT, we would expect that people have many exemplars of VOT in this range, predicting phonetic imitation to occur in both directions. The clear asymmetry of phonetic imitation found in this study, namely the absence of the imitation in Group 2, suggests that phonetic imitation is not an automatic process as exemplar theories might predict. Linguistically speaking, the difference between the two conditions is clear: imitating long VOT does not endanger the voiceless stop category, while imitating short VOT could introduce categorical ambiguity with the voiced stop. [10] presents a similar asymmetrical result: unrealistically long VOT values (e.g. around

150ms) are considered better exemplars of /p/ than shorter VOT values (e.g., 40ms), even though shorter values occur more often in real speech. That is, the VOT values played for Group 1 (113.26 ms on average) would be rated much higher than the VOT values played for Group 2 (32.29ms on average). If Group 2 subjects felt that the listening stimuli were not very good examples of /p/, while Group 1 subjects heard the listening stimuli as good examples of /p/, it is not surprising that the imitation effects in the two experiments were asymmetrical. Taking [10] and the current study's results together, phonemic "goodness" appears to be modulated by knowledge of phonemic contrast, rather than simply based on the collection of experiences. And this linguistic knowledge seems to have influenced the phonetic imitation in Group 2.

In an attempt to replicate the effect of word specificity [1] in a non-shadowing VOT paradigm, the current study manipulated lexical frequency and presence of exposure as independent variables. Although significant main effects were found on VOT for Group 1, the expected interaction between the degree of imitation (in VOT) and those variables was not observed in the data. Two factors might have contributed to this negative result, however. First, unlike [1], the current study included a silent warm-up phase. Recently experienced exemplars are expected to have a stronger echo, and thus the presence of the warm-up phase might have reduced the difference between high and low frequency groups. Second, measuring one acoustic feature might not be as powerful as overall perceptual assessment (as in [1]) in order to show word specificity. Thus the inconclusive result for word specificity in this study cannot be taken to challenge its existence, but rather reveals that the effect is at best subtle and perhaps requires strong statistical power to demonstrate it.

5. CONCLUSION

Two non-shadowing phonetic imitation experiments were conducted that test 1) interaction of phonetic imitation with linguistic contrast; 2) generalizability of phonetic imitation to new instances which share (a) the same phoneme, or (b) the same feature; and 3) word-specificity as predicted by the exemplar view. The results showed a clear asymmetry of implicit phonetic imitation between two subject groups: when the subjects were exposed to modeled speech with extended VOTs, the modeled feature [+spread glottis] was imitated, and generalized to new words (with initial /p/) as well as to a new segment (/k/). However, when the subjects were exposed to modeled speech with reduced VOTs, there was no imitation observed. These results reveal the non-automaticity and generalizability of phonetic imitation, indicating that knowledge of phonemic contrast constrained the implicit phonetic imitation, and that the subjects possess knowledge of sub-phonemic structure.

6. ACKNOWLEDGEMENT

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 Pat Keating, Colin Wilson, and the audience at LSA Albuquerque and Labphon 10 for their comments. Correspondence concerning this article should be addressed to Kuniko Nielsen, Department of Linguistics, UCLA (kuniko@humnet.ucla.edu).

7. REFERENCES

1. Goldinger, S. D. (1998). Echoes of Echoes? An Episodic Theory of Lexical Access. *Psychological Review*, 105, 251-279.
2. Mullennix, J. W., Pisoni, D. B., & Martin, C. S. (1989). Some effects of talker variability on spoken word recognition. *Journal of the Acoustic Society of America*, 85, 365-378.
3. Norris, D., McQueen, J. M., and Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, 47, 204-238.
4. Wright, R. (2004). Factors of lexical competition in vowel articulation. In J. Local, J., R.Ogden, and R. Temple (eds.). *Phonetic Interpretation: Papers in Laboratory Phonology VI* (pp75-87). Cambridge: Cambridge University Press.
5. Shockley, K., Sabadini, L., & Fowler, C. A. (2004). Imitation in shadowing words. *Perception & Psychophysics*, 66, 422-429.
6. Kučera, H., & Francis, W. N. (1967). Computational analysis of present day American English. Providence, RI: Brown University Press.
7. Baayen, R.H., Piepenbrock, R., & Gulikers, L. (1995). The CELEX Lexical Database (Release 2) [CD-ROM]. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania [distributor].
8. Nusbaum, H. C., Pisoni, D. B., & Davis, C. K. (1984). Sizing up the Hoosier mental lexicon: measuring the familiarity of 20,000 words. *Research on Speech Perception Progress Report*, No. 10. Bloomington: Indiana University, Psychology Department, Speech Research Laboratory.
9. Zue, V. (1980). Acoustic characteristics of stop consonants: A controlled study. Bloomington, IN: Indiana Linguistics Club.
10. Allen, J. S., & Miller, J. L. (2001). Contextual influences on the internal structure of phonetic categories: A distinction between lexical status and speaking rate. *Perception & Psychophysics*, 63, 798-810.