

Generalization of phonetic imitation across place of articulation

Kuniko Y. Nielsen

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

Abstract

The imitation paradigm, in which subjects' speech is compared before and after they are exposed to target speech (=study phase), has shown that subjects shift their production in the direction of the target [1], indicating not only episodic memory in speech perception but also the close tie between speech perception and production. The purpose of the current study is to determine whether the VOT imitation effect can be observed in a non-shadowing paradigm, as well as how the effect is generalized to new stimuli. In the study phase, subjects listened to a word containing items with initial /p/ that had extended VOT. In the pre- and post-study phases, subjects produced words from lists including 1) the words in the listening list, 2) the segment /p/ in new words, and 3) the segment /k/, which did not occur during study. The results replicated the shadowed imitation study by Shockley et. al [2]: subjects produced longer VOT post-study than pre-study for all stimuli types. The results show that the imitation effect was generalized to new stimuli beginning with /p/ which the subjects did not hear during the study phase, as well as to a new segment (/k/). The effect of lexical frequency, predicted by exemplar based theories, was not found in the data.

1. Introduction

Traditional accounts of speech perception assume that linguistic representations are invariant, and that these invariant representations need to be extracted from variant speech signals [3]. According to this view, some types of variability in the speech signal (such as speech rate or F0) represent 'noise' which listeners have to filter out. However, the 'invariant' representation has yet to be discovered (lack of invariance problem: [4]), and many challenges remain on the way to mapping variant acoustic-phonetic information onto invariant abstract representations. Recently, this traditional view has been challenged by exemplar-based theories which do not assume invariant linguistic representations. Hintzman's MINERVA 2 [5] takes the idea of exemplar storage to a logical extreme, assuming that every experience creates an exemplar or memory trace. To demonstrate this view in spoken-word perception, Goldinger [1] applied the model to the single-word shadowing data, and replicated the qualitative predictions of the model: 1) when subjects' speech is compared before and after they are exposed to target speech, they shift their production in the direction of the target, 2) the effect was larger for lower frequency words, and 3) the effect was also larger for subjects who had more exposure to the target. He later replicated the same result in a non-shadowing paradigm [6], where the subjects recorded the test tokens five days after they were exposed to target speech. In addition to

this imitation effect of "voice", Shockley et al. [2] replicated the same effect by measuring one aspect of speech, namely extended aspiration (VOT), also in a single-word shadowing paradigm.

Although these results show how much of surface information affects our speech production, they provide little information in terms of how experienced speech input affects underlying linguistic representation. For example, information about the "voice" is more likely to be paralinguistic. VOT variability could carry important linguistic information; however, the shadowing task might not reveal linguistic representation in memory (even if shadowing requires deep level speech processing as Shockley et al. claim). In other words, no study has yet looked at a linguistic variable (e.g. VOT) in a long-term (non-shadowing) task. If the VOT imitation effect can be obtained through a non-shadowing paradigm, this would be stronger support for the episodic view of memory as well as the link between speech perception and production. In addition to replicating the imitation effect in a non-shadowing paradigm, it is also our interest to investigate what is imitated by testing the generalizability of the effect. In both studies [1, 2], the subjects produced and heard exactly the same word lists, and thus these results do not reveal the size of the linguistic unit(s) influenced by the effect. That is, when a subject shifts production of a particular sound in a particular word, it is uncertain whether the subject's representation of the whole word, the segment, or the feature has been influenced. In this study, to investigate whether phonetic imitation is generalized to new stimuli beginning with /p/ and /k/, the study-phase word list includes words with initial /p/ with extended VOT, while the pre- and post-study production list includes (1) the modeled words, replicating [2], (2) the modeled segment /p/ in new words, and (3) the modeled feature [+spread glottis] (= aspiration) in a new segment /k/. Lastly, lexical frequency was controlled as an independent variable to see if there is a difference in magnitude of the imitation effect between high and low frequency words as shown in [1, 6].

2. Method

The goals of our experiment were to determine 1) whether the VOT imitation effect can be observed in a *non-shadowing* paradigm, 2) how the effect is generalized to new stimuli, and 3) whether lexical frequency would induce a difference in the VOT imitation effect.

Participants. Eight native speakers of American English with normal hearing served as subjects for this experiment. They were recruited from the UCLA undergraduate population, and

included 4 females and 4 males. They received course credit for participating.

Stimuli. The production list consisted of 150 English words. Among them, 100 were words beginning with /p/ (40 high-frequency words and 40 low-frequency words which were played in the study phase, and an additional 20 low-frequency words which were not played during the study 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 words from the production list (40 high-frequency words and 40 low-frequency words beginning with /p/), and 40 filler words beginning with sonorants. The lexical frequency was determined from both Kúcera & Francis [7] and CELEX2 [8]: the threshold for low-frequency words was 5 (per million) and 300, and that for high-frequency words was 50 and 1000, respectively. The phonological neighborhood density and syllable length were controlled between the two frequency groups. All the words had equally high familiarity (6.0-7.0 on the 7-point Hoosier Mental Lexicon scale) [9]. All the target words had initial stress, and there were no onset clusters.

A phonetically trained male American English speaker recorded the 120 words in the listening list. The speaker first produced the words in the list normally, and in addition, he produced the target words (words beginning with /p/) with extra aspiration. The VOT for the normally produced initial /p/ was measured, and was spliced with the initial part of hyper-aspirated tokens using PCQuirer (Scicon R&D, CA) so that the resulting tokens have VOT extended by 40ms. The extended tokens had VOT of 113.26 ms on average (SD=10.82). This splicing method was chosen, as opposed to extending the middle part of VOT, to maximally preserve natural formant transitions.

Procedure. The experiment used a slightly modified version of the imitation paradigm [1, 2], in that a warm-up reading phase was added at the beginning to avoid possible hyper-articulation in the test reading. The stimuli were presented using Psyscope 1.2.5 [10]. Each subject was seated in front of a computer in a sound booth. Each session was divided into 4 phases: warm-up, pre-study baseline, study, and post-study test. In the warm-up phase, the words were presented, 1 at a time, on a computer screen every 2 sec. The subjects were instructed to read the words silently without pronouncing them. In the pre-study baseline phase, the subjects were instructed to “identify the word you see by speaking it into the microphone.” In the study phase, using headphones, the subjects were exposed to two repetitions of the 120 word tokens (80 target words and 40 filler words) spoken by the main speaker. There was no additional task during this block. The post-study test phase was exactly the same as the pre-study baseline phase. Across the four phases, 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 (i.e., AXB testing) of the pre- versus post-productions.

3. Results

The independent variables in this study are 1) production condition (pre- vs. post-study), 2) lexical frequency, 3) old vs.

new stimuli (presented vs. not presented in the study phase) and 4) segments (/p/ vs. /k/). A repeated-measures ANOVA with two within-subjects factors (pre vs. post, frequency: high vs. low) showed a significant difference between pre- and post-study productions ($F(1,7) = 6.488, p < .05^*$), and no significant interaction between high and low frequency words ($F(1,7) = 0.32, p > .1$). There was no significant difference between the two variables ($F(1,7) = 0.578, p > .1$). Another repeated-measures ANOVA with two within-subjects factors (pre vs. post, stimuli: old vs. new) again showed a significant difference of production ($F(1,7) = 6.857, p < .05^*$), yet no significant difference between the stimuli which were heard vs. unheard during the study phase ($F(1,7) = 3.289, p > .1$). Again, there was no significant interaction between the two variables ($F(1,7) = 0.050, p > .1$). Lastly, in order to see how the imitation effect is generalized to new stimuli, a repeated-measures ANOVA with two within-subjects factors (pre vs. post, segment: /p/ vs. /k/) were performed. Note that neither group of words was played in the study phase. Similar to the earlier tests including items that were played in the study phase, there was a significant difference between pre- and post-study productions ($F(1,7) = 6.023, p < .05^*$). As expected, there was a significant difference between /p/ and /k/ ($F(1,7) = 125.797, p < .001^*$), while there was no interaction between the order and segment ($F(1,7) = 0.032, p > .1$).

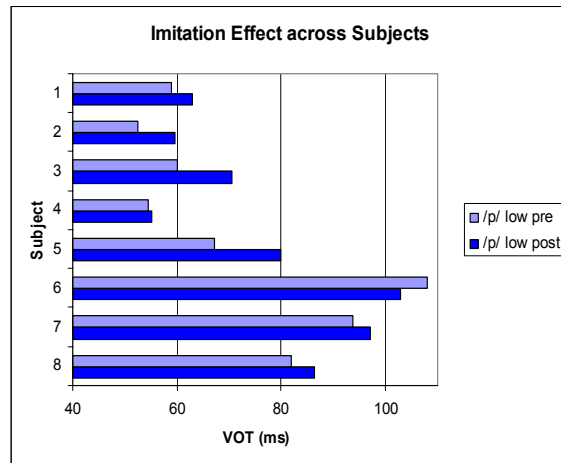


Figure 1: Imitation effect for the low-frequency tokens presented in the study phase, plotted across eight subjects (subject # 1-4 are male, 5-8 are female): seven out of eight subjects produced longer VOT in post-study phase.

4. Discussion

Our results revealed a statistically significant difference between pre- vs. post study phase tokens, revealing that the VOT imitation effect is present even when the task does not involve shadowing. As can be seen in Figure 1, seven out of eight subjects produced longer VOT in the post-study phase. This result is consistent with previous studies as well as the episodic view of speech perception: assuming that elicitation-

style production reflects underlying representation of the lexical items, about eight minutes after they heard the modeled speech, subjects sustained its detailed surface information (i.e., extended aspiration).

Goldinger [6] found that as MINERVA 2 predicts, lexical frequency affects the magnitude of the imitation effect: the higher the frequency, the weaker the effect. The current study carefully controlled lexical frequency to see whether the same result would still hold in a non-shadowing VOT paradigm. Although our results indicated a relatively weaker imitation effect for high-frequency words (see Figure 3), the difference between the two frequency groups was not significant and there was no interaction between the imitation effect and frequency. This result suggests that although the subjects were using the episodic trace, they did not use the lexicon. Given that the effect of lexical frequency, if it exists, appears to be quite small, a study with more power may be needed to detect it.

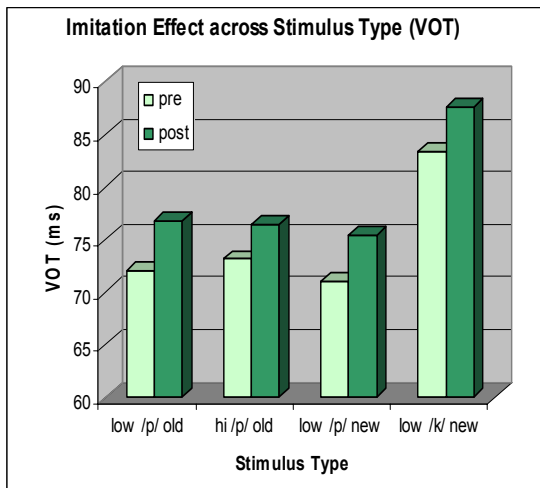


Figure 2: Imitation effect (in VOT) plotted across four types of stimuli: The imitation effect was significant for all types of stimuli, although there was no significant interaction with other factors.

Our results showed that the imitation effect was generalized to stimuli that subjects did not hear during the study-phase: subjects produced significant differences between pre- and post-study production in new words with initial /p/, and in words with initial /k/ (see Figures 2 & 3). There are at least three possible ways to interpret this result: first, if the imitation effect was triggered by exemplar, the unit of the exemplar could be smaller than words or segments, namely the “feature”. After being exposed to words containing a salient feature (extended aspiration, or [+spread glottis]), all the exemplars containing the same feature are activated and thus increase VOT regardless of initial phoneme or lexical frequency. If we had found no imitation effect for high-frequency words, it could indicate that the unit is rather like word-size and thus it would have contradicted with the generalizability of imitation effect found in this study. In other words, this generalizability of imitation effect is

consistent with another result of this study, namely the absence of lexical frequency effect.

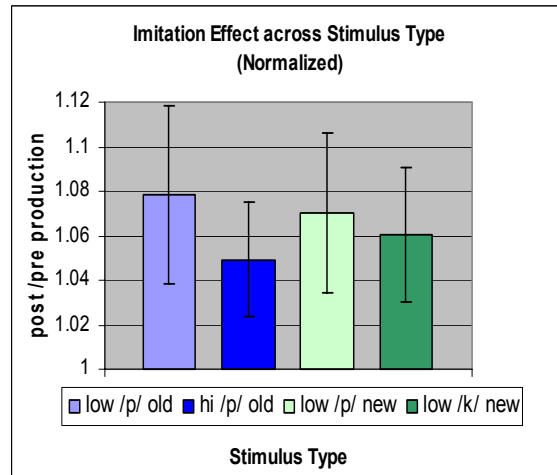


Figure 3: Imitation effect (normalized) plotted across four types of stimuli: the stimuli group consists of low-frequency words which were heard during the study phase (= low /p/ old) showed the largest effect.

Secondly, it is equally possible that the subjects learned a rule of segment (i.e., /p/) lengthening after listening to the target tokens, and applied the rule to a new segment /k/. Another way to interpret this result is that the subjects switched their “register”: after being exposed to words with strong aspiration, it is possible that they perceived the stimuli as “carefully spoken” and subconsciously adjusted their speech as well.

In order to investigate the source of the VOT imitation effect found in this study, the apparent next step is to measure other variables such as word and/or vowel durations: if the effect is due to episodic memory or rule-learning, only the manipulated variable (in this case, VOT) should be affected. On the other hand, if we observe changes in other variables, the imitation effect is more likely to be due to global aspects of speech, such as register.

In the current study, all the subjects listened to the same target stimuli, in other words, there were no between-subject factors. Their VOT was compared before and after they listened to the target speech and the pre-study production was used as baseline. However, in a strict sense, the baseline should be the post-study production of another group of subjects after they listened to non-manipulated stimuli. At this point, it is uncertain if such speakers would keep the same level of VOT between the two blocks. It is entirely possible that their first tokens are more likely to be hyper-articulated and thus their second tokens have lower VOT. If this is the case, the imitation effect may actually be larger than it appears in this study.

5. Conclusions

To investigate how experienced speech input affects linguistic representations, the current study examined the imitation effect of a linguistic variable (i.e. extended VOT) in a long-term (non-shadowing) task. The data revealed a significant imitation effect: subjects produced longer VOT in the post-study phase than in the pre-study phase. Furthermore, the results showed that the modeled feature [i.e., + spread glottis] was generalized to new words (with initial /p/) and a novel sound (/k/). Lexical frequency of the stimuli was controlled to see whether low-frequent words show larger imitation effects, as predicted by exemplar-based theories. However, there was no interaction between imitation and lexical frequency. Together, these results neither disconfirm nor support the exemplar theory, for it is possible to explain the imitation effect by non-exemplar causes, such as rule-learning or change in register. What our results indicate is that the locus of the imitation effect can be smaller than individual words or segments.

6. References

- [1] Goldinger, S. D. (1998). Echoes of Echoes? An Episodic Theory of Lexical Access. *Psychological Review*, 105 (2), 251-279.
- [2] Shockley, K., Sabadini, L., & Fowler, C. A. (2004). Imitation in shadowing words. *Perception & Psychophysics*, 66 (3), 422-429.
- [3] Halle, M. (1985). Speculation about the representation of words in memory. In V. Fromkin (Ed.), *Phonetic Linguistics* (pp.101-114.) New York: Academic Press.
- [4] Liberman, A. M., & I. G. Mattingly. (1985). The motor theory of speech perception revised. *Cognition*, 21, 1-36.
- [5] Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review*, 93, 411-428.
- [6] Goldinger, S. D. (2000). The Role of Perceptual Episodes in Lexical Processing, In *SWAP-2000*, 155-158.
- [7] Kučera, H., & Francis, W. N. (1967). *Computational analysis of presentday American English*. Providence, RI: Brown University Press.
- [8] 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].
- [9] 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.
- [10] Cohen, J., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: An interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behavior Research Methods, Instruments, & Computers*, 25, 257-271.