

## **The Interaction between Spontaneous Imitation and Linguistic Knowledge**

Kuniko Y. Nielsen

[kuniko@humnet.ucla.edu](mailto:kuniko@humnet.ucla.edu)

### **ABSTRACT**

The spontaneous imitation paradigm (Goldinger, 1998), in which subjects' speech is compared before and after they are exposed to target speech, has shown that subjects shift their production in the direction of the target, indicating the use of episodic traces in speech perception as well as the close tie between speech perception and production. By using this paradigm, the current study aims to investigate the psychological reality of three levels of linguistic unit (i.e., word, phoneme, and sub-phonemic unit such as feature/gesture) through physical measurements instead of perceptual assessments. An experiment was carried out to test: 1) whether spontaneous phonetic imitation can be *generalized* across (a) new words which share the same initial phoneme, and (b) new words with a new phoneme falling in the same natural class (sharing a feature/gesture); 2) whether word-level specificity can be obtained through physical measurements of a phonetic feature; and 3) if/how the phonetic imitation interacts with linguistic representations when the change might impair linguistic (in this case, phonemic) contrast. The feature manipulated in the experiments was aspiration, or [+/- spread glottis], on the phonemes /p/ and /k/. The results revealed a significant effect of spontaneous phonetic imitation: subjects produced significantly longer VOTs after they were exposed to target speech with extended VOTs, replicating Shockley (2004) in a non-shadowing paradigm. Furthermore, this modeled feature (increased aspiration) was generalized to new instances of the target phoneme /p/ (i.e., in new words) as well as to the new segment /k/. On the other hand, the subjects did not imitate reduced VOTs, despite the fact that the (modeled) shorter VOTs occur more often than (modeled) longer VOTs in the baseline recordings. These results, taken together, indicate that 1) speakers possess sub-phonemic representations, and 2) knowledge of linguistic (or, phonemic) contrast constrains spontaneous phonetic imitation. Expected word-level specificity, tested through lexical frequency and training exposures, was not observed in this study.

### **1. Introduction**

Recent studies have shown that traces of episodic memory are retained and used in speech perception (Mullennix et al. 1989), and that both speech perception and production are more plastic than previously considered (e.g., Norris et al., 2003; Clark and Luce, 2005; Wright, 2004; Hay, 2000). The imitation paradigm, which compares subjects' speech before and after the exposure to target speech, has shown that speakers shift their productions in the direction of what they just heard. For example, Goldinger (1998) showed that subjects shifted their own F<sub>0</sub> (compared with their baseline recording) when they are asked to shadow (= immediate repetition) speech with manipulated F<sub>0</sub>. His results also revealed a word-specific advantage of imitation: stronger imitation effects were observed (through AXB perceptual assessment) among low-frequency words than high-frequency words, as well as among subjects who had a higher number of training (=target) exposures, as predicted by exemplar-based theories. A similar result was also obtained in his later work (2000), which involved non-shadowing experiments. Shockley et al. (2004) extended Goldinger's work by showing a significant Voice-Onset-Time (VOT) imitation effect (in shadowing) for voiceless stops with

artificially extended VOTs. These results demonstrate that shadowing responses, obtained in nonsocial settings, are imitative. In addition, these studies show listeners' sensitivity to variations in global phonetic dimensions such as overall pitch range, as well as sensitivity to the fine phonetic detail of a single segment such as degree of aspiration.

Although Goldinger (1998) shows evidence for word-size representations, his studies do not reveal whether sub-lexical units were also influenced by the imitation effect. That is because the post-exposure productions were elicited in the form of shadowing, and thus the listening and production lists had to be identical. The present study extends the earlier studies by using a non-shadowing task, which lets the listening (= target) and production lists *differ*; thus unheard words can be introduced into the production list. This allows us to test the generalizability of the imitation effect to sub-lexical units. Many linguistic theories (e.g., Halle, 1985) assume three levels of representations: lexical (=word), phonemic and sub-phonemic (= feature or gesture).

It is also our interest to determine the automaticity of spontaneous phonetic imitation, by comparing how two types of modeled stimuli are imitated. In addition to extended VOT (as in Shockley et al., 2004), the current study employs reduced VOT as target stimuli to see if the same degree of imitation and generalization can be observed. Unlike extended VOT, reduced VOT in voiceless stops could introduce linguistic ambiguity, namely confusion with the voiced category. If spontaneous phonetic imitation is a rather automatic process, knowledge of linguistic structure (i.e., 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 more complicated process that is filtered by abstract linguistic knowledge.

In order to address these issues, the following questions were asked:

1) Generalizability to new stimuli

- a) Will there be generalization to the same phoneme in new (unheard) words?
- b) Will there be generalization to the same feature in new phonemes?

If we observe sub-lexical generalization at the phoneme level but not the sub-phonemic (feature) level, it will provide support for phoneme-size representations. If we observe generalization at both phoneme and sub-phonemic levels *independently*, it will provide support for phoneme *and* sub-phoneme-size representations.

2) Word-specific advantage

- a) Will there be a larger imitation effect for words in the listening list?
- b) Will there be a larger imitation effect for words with lower lexical frequency?

The exemplar view predicts a stronger specificity for more recently experienced words, so we would expect a larger imitation effect for words which subjects hear in the experiment. The view also predicts a stronger specificity for low-frequency words than for high-frequency words, because the smaller the number of exemplars associated with a given word, the larger the weight of each new exemplar. These predictions were proved correct (Goldinger 1998, 2000) through AXB perceptual judgments: lower-frequency

words engender more imitation, an effect that increases with repetitions. Could a similar result be obtained through physical measurement of VOT?

- 3) Interaction between spontaneous imitation and linguistic (phonemic) contrast
  - a) If imitating one type of stimuli (i.e., reduced VOT) might impair a linguistic contrast while the other (i.e., extended VOT) does not, would speakers still imitate in the same degree?

## 2. Experiment

### 2.1. Method

#### *Participants*

Thirty-nine native speakers (20 for Group 1, 19 for Group 2) of American English with normal hearing served as subjects for this experiment. They were recruited from the UCLA undergraduate population, and included 20 females and 19 males. They received course credit for participating.

#### *Stimuli*

The production list consisted of 150 English words. Among them, 100 were words beginning with /p/ (80 target words: 40 high-frequency words and 40 low-frequency words<sup>1</sup> which were played in the study 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. The lexical frequency was determined from both Kucera & Francis (1967) and CELEX2 (Baayen et al. 1995): 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 on the 7-point Hoosier Mental Lexicon scale) (Nusbaum et al. 1984). 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 VOTs for the normally produced initial /p/ were measured (mean=72.46ms, SD=12.14). To make the 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 using PCQuirer (Scicon R&D, CA) so that the resulting VOT was *extended* by 40ms (mean=113.26 ms, SD=10.82). This splicing method was chosen, as opposed to extending the middle part of VOT as in Shockley et al. (2004), in order to maximally preserve natural formant transitions. To make the 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 the resulting tokens had VOT reduced by 40ms from the original tokens (mean=32.29ms,

---

<sup>1</sup> Each frequency group consisted of 10 monosyllabic words, 20 disyllabic words and 10 trisyllabic words.

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 word by both listeners.

### *Procedure*

The experiment used a slightly modified version of the imitation paradigm (Goldinger, 1998), in that a warm-up reading phase was added at the beginning to avoid possible hyper-articulation in the test reading due to first exposure. The stimuli were presented using Psyscope 1.2.5 (Cohen et al. 1993). Each subject was seated in front of a computer in a sound booth. Each session was divided into 4 blocks: 1) warm-up, 2) baseline, 3) listening, and 4) test. In the warm-up block, the words were presented, one at a time, on a computer screen every 2 seconds. The subjects were instructed to read the words silently without pronouncing them. 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 (80 target words and 40 filler words). There was no additional task during this block. The test block was exactly 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 (i.e., AXB testing) of the baseline versus test productions.

## **3. Results**

The 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 (group) factor in this study was:

**Listening Stimuli** (Extended VOT [Group 1] vs. Reduced VOT [Group 2])

A repeated-measures ANOVA analysis with Type of Production (within-subject) and Listening Stimuli (between-subject) revealed a clear interaction ( $F(1,75) = 23.99$ ,  $p < 0.001^*$ ) between the two factors. 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.

### **Group1**

Repeated-measures ANOVA analysis with two within-subjects factors (Type of Production and Lexical Frequency) for the 80 target words revealed significant main effects of both Type of Production ( $F(1,19)=13.13$ ,  $p < 0.01^*$ ) and Lexical Frequency

( $F(1,19)=4.79, p<0.05^*$ ). However, the interaction between the two factors was not significant ( $F(1,19)=0.08, p>0.1$ ).

Another repeated-measures ANOVA analysis with two within-subjects factors (Type of Production and Presence of Exposure [= target vs. novel stimuli]) showed a significant difference for both Type of Production ( $F(1,19)=11.99, p<0.01^*$ ), and Presence of Exposure ( $F(1,19)=13.22, p<0.01^*$ )<sup>2</sup>. However, the interaction between the two factors was not significant ( $F(1,19)=0.39, p>0.1$ ).

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. Similar to the results for items that were played in the listening block, there was a significant difference between pre- and post-exposure productions ( $F(1,19)=17.079, p<0.01^*$ ). There was also a significant difference between /p/ and /k/ ( $F(1,19)=217.845, p<0.001^*$ ) as expected due to their normally-differing VOT (e.g., Zue, 1980), while there was no interaction between the two factors ( $F(1,19)=0.82, p>0.1$ ).

Table 1 shows the medians, means, standard deviations and standard errors of VOT (ms) by stimulus types. As can be seen, the standard deviations are very large in general, due to the individual variability of VOT. On the other hand, the means of standard errors are quite small, which shows that the subjects' shifts in their production (= imitation) were rather consistent.

**Table 1: Summary of Group 1 Results**

Stimuli Type	Order of Production	Median (ms)	Mean (ms)	Std. Deviation (ms)	Std. Error of Mean (ms)
Target /P/ Low	Baseline	59.678	65.282	15.8164	3.5367
	Test	73.257	73.346	17.8417	3.9895
Target /P/ High	Baseline	59.495	64.209	15.6039	3.4891
	Test	68.937	71.867	17.4699	3.9064
Novel /P/ Low	Baseline	60.862	62.946	15.6226	3.4933
	Test	68.865	69.920	14.7472	3.2976
Novel /K/ Low	Baseline	70.615	75.143	13.9103	3.1104
	Test	81.240	80.797	13.9099	3.1103

## Group2

Repeated-measures ANOVA analysis with two within-subjects factors (Type of Production and Lexical Frequency) for the 80 target words showed significant differences for neither Type of Production ( $F(1,18)=1.14, p>0.1$ ) nor Lexical Frequency ( $F(1,18)=0.38, p>0.1$ ). The interaction between the two factors was not significant ( $F(1,18)=0.43, p>0.1$ ). Two more repeated-measures ANOVA analyses with two within-

<sup>2</sup> Note that this comparison was made between the twenty novel words (all disyllabic with initial /p/ and low lexical frequency) and the twenty disyllabic target words with low frequency.

subjects factors (Type of Production and Presence of Exposure, Type of Production and Segment) were performed, and the results revealed no significant difference in any comparison ( $F < 1$ ,  $p > 0.1$ ). That is, there was no imitation, no word-specificity, and no generalization found in Group 2. Table 2 shows the medians, means, standard deviations and standard errors of VOT (ms) by stimulus types. As can be seen, being exposed to target speech did not shift subjects' VOT in the way it did in Group 1.

**Table 2: Summary of Group 2 Results**

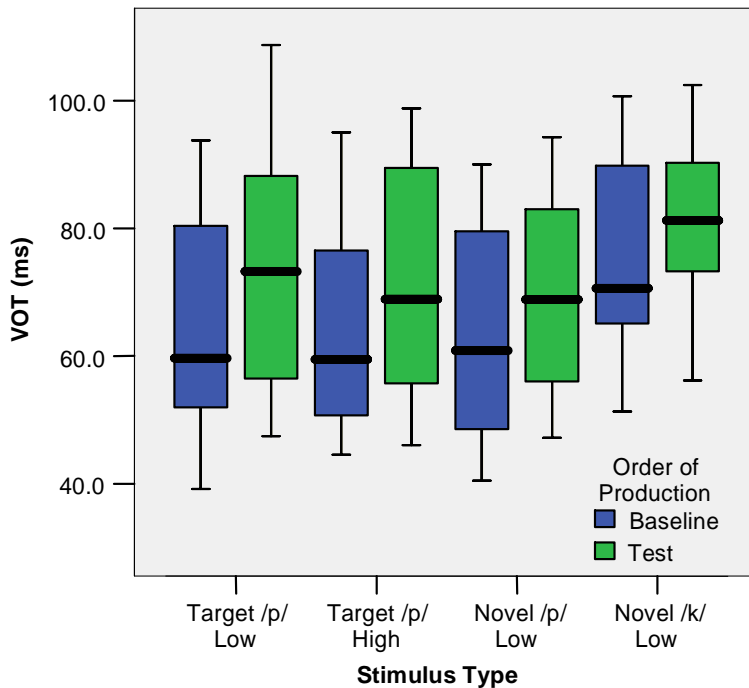
Stimuli Type	Order of Production	Median (ms)	Mean (ms)	Std. Deviation (ms)	Std. Error of Mean (ms)
Target /P/ Low	Baseline	63.500	61.610	12.5022	2.8682
	Test	63.810	64.236	14.6565	3.3624
Target /P/ High	Baseline	65.906	63.383	11.4462	2.6259
	Test	63.308	64.488	15.2054	3.4884
Novel /P/ Low	Baseline	64.512	62.755	11.9220	2.7351
	Test	63.232	63.166	14.7955	3.3943
Novel /K/ Low	Baseline	75.265	77.298	9.7106	2.2278
	Test	76.620	77.083	10.8968	2.4999

#### 4. Discussion

##### *Imitation*

Our results from Group 1 revealed a statistically significant effect of spontaneous phonetic imitation. As seen in Figure 1, test-productions show consistently longer VOTs than baseline productions, revealing that the VOT imitation effect is present even when the task involves non-shadowing elicitation-style production. This result is consistent with previous studies (e.g., Goldinger, 1998). About eight minutes after they heard the target speech, subjects appeared to have sustained the modeled speech's detailed surface phonetic information (i.e., extended aspiration).

It could be hypothesized that the imitation is due to global changes in speech style. An argument against this interpretation is provided by a post-hoc analysis of whole-word duration. 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 the change is due to more global aspects of speech, we would expect to see changes in other variables. For this reason, the whole-word duration of the low-frequency target words was measured from 8 randomly chosen subjects' data. Unlike with VOT, there was no significant difference between baseline and test ( $F < 1$ ,  $p > .1$ ) productions. Given these results, it is unlikely that global aspects of speech are solely responsible for the spontaneous phonetic imitation observed in this study.



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

### *Word-Specificity*

In order to replicate the effect of word specificity (Goldinger, 1998 & 2000) through physical measurement of a phonetic feature, the current study carefully controlled Lexical Frequency and Presence of Exposure as independent variables. Word specificity would mean that low frequency words, especially those with (more) exposure, should show a stronger imitation effect. Although significant main effects were found in both variables, expected interactions between these variables and the strength of imitation (in VOT) were not observed. Thus, the effect of word specificity, both through Lexical Frequency and Presence of Exposure, is inconclusive in this study.

There are some factors that might have contributed to this negative result, however. First, the method of current study was different from previous studies with regard to the presence of the warm-up phase. According to Goh and Pisoni (1998), a lexical frequency effect could easily disappear in such a repeated-sampling procedure. Although the modification was made in an attempt to eliminate/minimize some hyper-articulation (for low frequency items) observed in our pilot study, it did add an exemplar to each target word (with high and low frequency). Recently experienced exemplars are supposed to have a stronger echo, and thus the presence warm-up phase must have reduced the difference between high and low frequency groups. It would not be surprising if the warm-up phase contributed to the null-interaction result of imitation and lexical frequency in the current study. Second, it is also noteworthy that Goldinger

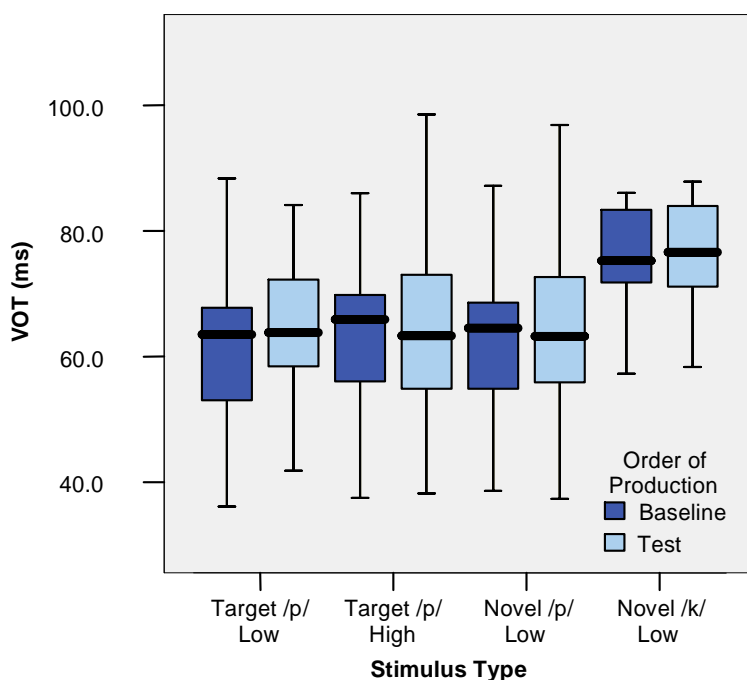
manipulated the number of training exposures from none to twelve, while the current study compared none and two. In his results (both 1998 and 2000), there was a clear interaction between the strength of imitation (i.e., correct AXB judgment) and the number of repetitions, yet the difference between zero and two exposures was rather small. Third, the current study estimated the strength of imitation from physical measurements of VOT produced by 20 subjects, as opposed to perceptual judgment by 300 subjects in Goldinger (2000). Among many features in the modeled speech signal (which were all available for the subjects to imitate), VOT is the only feature manipulated and measured in this study. It is possible that measuring one feature is just not as powerful as overall perceptual assessment (i.e., AXB judgment) to show word-specificity. And the smaller number of subjects certainly reduces the power of statistics. Taken together, the inconclusive result for word-specificity in this study cannot be taken to challenge the existence of such an effect, but it does reveal that the effect is subtle and perhaps requires strong statistical power to demonstrate it.

### *Generalizability*

Our data from Group 1 also showed that the imitation effect was generalized to novel stimuli that subjects did not hear during the listening block (see *Novel /p/* types in Figure 1). Compared with their baseline, the subjects produced significantly longer VOT in the test block even for the novel words with initial /p/. This result indicates that the locus of spontaneous phonetic “imitation” is not word-specific, and that subjects imitated something smaller than a word.

Perhaps the most important finding of the current study is that 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. Many linguistic theories assume that words are made up of discrete speech sounds (=segments) that are themselves complexes of features (Halle, 1985). Support for this notion has traditionally been provided by phonological alternations and phonotactic constraints. Recent experimental work by Goldrick (2004) suggests that speakers possess knowledge of phonological structure: in his study, subjects were able to learn phonotactic constraints at two levels of representations (segment and feature) through exposure to a set of nonwords. The current study provides additional support, through spontaneous phonetic imitation, for the sub-phonemic assumption. On the other hand, there was no interaction between the tested segment (i.e., /p/ and /k/) and Type of Production (baseline vs. target) ( $F < 1$ ,  $p > .1$ ), showing that the amount of imitation for the two segments was the same. Thus our results are inconclusive with respect to a phoneme-level representation.

Note also that these results support some sort of sub-phonemic representations, but not necessarily distinctive features per se. For example, it is entirely possible that the imitated unit is a *gesture* instead of a feature. These two theoretically contrastive views are in fact indistinguishable in the current study.

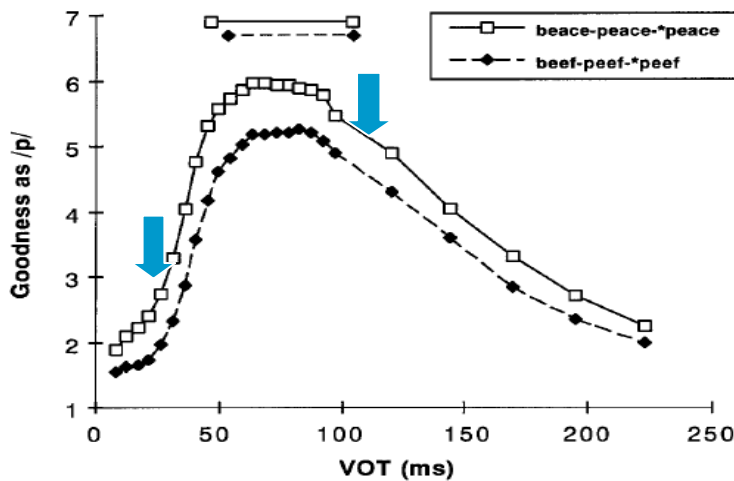


**Figure 2:** Group 2 Imitation effect (in VOT) plotted across four types of stimuli. The subjects listened to stimuli with SHORTER VOT. The horizontal (black) bar represents the median, the box represents the 25th - 75th percentile range, and the whisker represents the range of observation.

### *Asymmetry of Imitation*

Contrary to the results in Group 1, the subjects in Group 2 (who were exposed to shorter VOT) did not show any significant difference between baseline and test productions in terms of their VOT across all types of stimuli (see Figure 2). Our data from the baseline recordings of both experiments 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 heard 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 effect in Group 2, suggests that spontaneous phonetic imitation is not an automatic process as exemplar theories might predict, but rather a more complicated process modulated by other factor(s). 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. Allen and Miller's VOT Goodness-Rating study (2001) presents a similar asymmetrical result: unrealistically long VOT values (e.g., around 150ms) are considered as better exemplars of /p/ than shorter VOT values (e.g., 40ms) which occur more often in real speech. In Figure 3, two arrows that indicate the current study's listening stimuli (long and short VOT) were added to the original figure from Allen and Miller (2001). As can be seen, the VOT values played for Group 1 (113.26 ms on average) are rated much higher than the VOT values played for

Group 2 (32.29ms on average). If subjects in Group 2 felt that the listening stimuli were not very good examples of /p/, while subjects in Experiment 1 heard the listening stimuli as good examples of /p/, it is not surprising that the imitation effects in the two experiments were asymmetrical. In Allen and Miller (2001), a change in lexical status (whether the target was a word or not, such as in *beef-peef*) did not shift the entire range of best-exemplars (the horizontal lines in the Figure 3), but only the lower limit of the range, which borders on the category-boundary region, showing the subjects' sensitivities to phonemic contrast. Taking Allen & Miller's (2001) and the current study's results together, phonemic "goodness" appears to be modulated by the 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.



**Figure 3** [Figure 2 from Allen & Miller (2001), *Perception & Psychophysics*, 63 (5), p803]:  
**Group goodness ratings as a function of voice onset time (VOT) for the *beace-peace-\*peace* series and the *beef-peef-\*peef* series. Arrows were added to indicate VOT values which subjects in our experiments heard.**

Shockley et al. (2004) (as well as Fowler et al. 2003) argued that gesture theories of speech perception (Lieberman & Mattingly, 1985; Fowler, 1994) provide accounts for their results, namely the phonetic imitation of long VOT in shadowing. According to gesture theories, listeners' perception of phonemes includes motor information. When the participants are asked to shadow some speech signal, their responses are automatically guided by their perception of modeled gestures, thus predicting spontaneous imitation. Although it is not clear how long the memory of perceived gestures is sustained, our findings of phonetic imitation and its generalization at sub-phonemic level (but not phoneme level) seem *prima facie* more compatible with gesture theories than traditional acoustic theories. However, the asymmetry of phonetic imitation found in the current study reveals a more complex relationship between speech perception and production. At least in non-shadowing imitation, a perceived modeled gesture does not necessarily affect one's speech production, and other factor(s) can constrain the response.

## 5. Conclusion

In order to see if there is experimental support for the structures assumed by many linguistic theories, non-shadowing spontaneous phonetic imitation experiments were conducted that test 1) the generalizability of phonetic imitation to new instances which share (a) the same initial phoneme, or (b) the same feature; 2) the word-specific advantage predicted by exemplar view; and 3) interaction of phonetic imitation with linguistic (or, phonemic) contrast. As expected, the results revealed a significant effect of phonetic imitation in a non-shadowing paradigm: subjects produced significantly longer VOTs after they were exposed to the target speech than in their baseline productions recorded prior to the exposure. Furthermore, the results showed that the modeled feature [+ spread glottis] was generalized to new words (with initial /p/) as well as to a new segment (/k/). This result indicates that the subjects possess knowledge of sub-phonemic structure, supporting the traditional assumption in linguistics. However, when the subjects were exposed to modeled speech with reduced VOTs, there was no imitation observed, revealing the non-automaticity of phonetic imitation.

This study showed that speakers are sensitive to, and remember, sub-segmental details and phonemic contrast, which therefore must be represented in some way. The results of the current study thus call for a linguistically informed model of speech perception, which incorporates both sub-segmental and word-level representations as well as knowledge of linguistic contrast.

## 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 Pat Keating, Colin Wilson, and the audience at RIKEN BSI Forum, LSA Albuquerque, as well as Labphon 10 for their comments. Correspondence concerning this article should be addressed to Kuniko Nielsen, Department of Linguistics, UCLA (e-mail: kuniko@humnet.ucla.edu).

## REFERENCES

- 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 (5), 798-810.
- 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].

- Clark, C. M., and Luce, P.A. (2005). Perceptual adaptation to speaker characteristics: VOT boundaries in stop voicing categorization. Proceedings of ISCA Workshop on Plasticity in Speech Perception (PSP2005); London, UK.
- 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.
- Fowler, C. A. (1994). Speech perception: Direct realist theory. *Encyclopedia of language and linguistics* (Vol. 8, pp. 4199-4203). Oxford: Pergamon Press.
- Fowler, C. A., Brown, J. M., Sabadini, L., and Weihing, J. (2003). Rapid access to speech gestures in perception: Evidence from choice and simple response time tasks. *Journal of Memory and Languages*, 49, 396-413.
- Goh, W. D., & Pisoni, D. B. (1998). Effects of lexical neighbourhoods on immediate memory span for spoken words: A first report. Research on spoken language processing Progress Rep. No. 22. Bloomington, IN: Indiana University, Department of Psychology, Speech Research Laboratory.
- Goldinger, S. D. (1998). Echoes of Echoes? An Episodic Theory of Lexical Access. *Psychological Review*, 105 (2), 251-279.
- Goldinger, S. D. (2000). The role of perceptual episodes in lexical processing. In Anne Cutler, James M. McQueen and Rian Zondervan (eds.), *Proceedings of SWAP Spoken Word Access Processes*. Nijmegen: Max-Planck- Institute for Psycholinguistics, 155-159.
- Goldrick, M. (2004). Phonological features and phonotactic constraints in speech production. *Journal of Memory and Language*, 51, 586-603.
- Halle, M. (1985). Speculation about the representation of words in memory. In V. Fromkin (Ed.), *Phonetic Linguistics* (pp.101-114.) New York: Academic Press
- Hay, J. B. (2000) Causes and Consequences of Word Structure. Ph.D dissertation, Northwestern University.
- Kučera, H., & Francis, W. N. (1967). Computational analysis of present day American English. Providence, RI: Brown University Press.
- Liberman, A., & Mattingly, I. (1985). The motor theory revised. *Cognition*, 21, 1-36.
- Mullennix, J. W., Pisoni, D. B., & Martin, C. S. (1989). Some effect of talker variability on spoken word recognition. *Journal of the Acoustic Society of America*, 85, 365-378.
- Norris, D., McQueen, J. M., and Cutler, A. (2003). Perceptual learning in speech. *Cognit. Psychol.*, vol. 47, no. 2, pp. 204-238.
- 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.
- Pierrehumbert, J, B. (2002). Word-Specific Phonetics. In C. Gussenhoven and N. Warner (eds.) *Laboratory Phonology VII*, Mouton de Gruyter, Berlin. 101-140.
- Shockley, K., Sabadini, L., & Fowler, C. A. (2004). Imitation in shadowing words. *Perception & Psychophysics*, 66 (3), 422-429.

- Wright, R. (2004). Factors of lexical competition in vowel articulation. In J. Local, J., R.Ogden, and R. Temple (eds.). *Papers in Laboratory Phonology VI*. Cambridge: Cambridge University Press.
- Zue, V. (1980). *Acoustic characteristics of stop consonants: A controlled study*. Bloomington, IN: Indiana Linguistics Club.