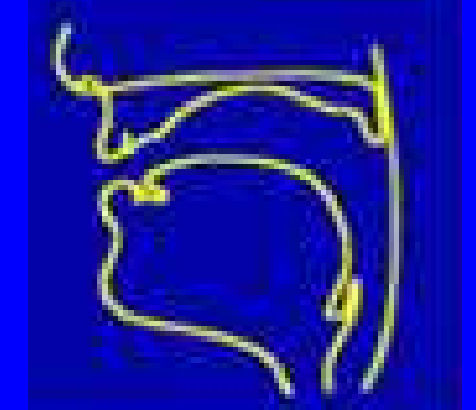


Generalization of phonetic imitation across place of articulation

Kuniko Y. Nielsen, Department of Linguistics, University of California, Los Angeles

UCLA



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. However, the 'invariant' representation has yet to be discovered, and many challenges remain in 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. In support of this view, Goldinger (1998) examined single-word **shadowing** data, comparing subjects' normal productions of words with shadowed productions of the same words. He found that 1) they **shifted their productions** in the direction of the auditory target, 2) the effect was larger for **lower frequency** words. Goldinger later (2000) replicated this result in a **non-shadowing** paradigm, in which the subjects recorded the test tokens five days after they were exposed to target speech.

Shockley et al. (2004) replicated this effect measuring by one aspect of speech, namely **extended aspiration (VOT)**, also in a single-word **shadowing paradigm**.

Goal

To develop an experiment which will test the effects of speech input on **linguistic representation**. For this purpose, the following criteria are important:

- **Variable:** Voice, F0 vs. **VOT** → More likely to carry **linguistic information**
- **Paradigm:** Shadowing vs. **Non-shadowing** → More likely to reveal underlying linguistic representation
- **No one has yet looked at a linguistic variable (e.g. VOT) in a long-term (non-shadowing) task**
- **Locus of imitation:** Word vs. **segment ?** → Words are composed of segments, Heard words vs. heard segment
- **Locus of imitation:** Segment vs. **feature ?** → Segments are composed of features Subsegmental primitive = features; voiceless stops /p t k/ share the feature value [+spread glottis] (form a **natural class**)
- **Is the imitation of linguistic information (e.g., VOT) generalized below the word? How far?**

- Exemplar theories predict a larger specificity effect for low-frequency words than high-frequency words (low frequency = fewer exemplars → weight of one exemplar is relatively bigger)
- **Is there an effect of lexical frequency in non-shadowing VOT imitation?**

The experiment

- 1) **VOT imitation effect in non-shadowing paradigm**
- 2) **generalizability of the effect to new stimuli**
- 3) **effect of lexical frequency in VOT imitation effect**

Key stimuli: Exaggerated VOT(40ms) **Are they imitated?**

Methods

Participants:

- 8 native speakers of American English (4M & 4F) with normal hearing and reading
- Received course credit for participating

Stimuli

- **Listening list** (for study-phase) **80 target words with initial /p/ - exaggerated VOT**, 40 filler words with initial sonorants
- **Production list** (for baseline and test phase)
 - 120 target words**
 - (1) the 80 modeled words (the targets in listening list)
 - (2) 20 new words, also with initial /p/
 - (3) 20 new words with initial /k/, which like /p/ is [+spread glottis]
 - 30 filler words with initial sonorants

Lexical frequency:

- **40 of the target words had high frequency, and 40 had low.** (Kučera & Francis (1967) Hi>50, Low<5: CELEX2 (Baayen, Piepenbrock and Gulikers, 1996) Hi>1000, Low<300)
- All the new words had low frequency

Phonological neighborhood density & number of syllables: controlled between frequency groups (Neighborhood density obtained from Sommers 2004)

Familiarity: 6.0-7.0 on the 7-point Hoosier Mental Lexicon scale (Nusbaum et al., 1984)

All the target words had initial stress, no onset clusters

- A phonetically trained male American English speaker recorded the 120 words in the listening list
- The speaker produced: 1) All the words normally, and 2) The target words *with extra aspiration*
- The VOT for the normally produced initial /p/ was extended by 40ms by splicing (PCquirer: Scicon R&D, CA)
- Splicing was with the initial part of hyper-aspirated tokens to maximally preserve natural formant transitions
- The extended tokens had average VOT of 113.26 ms (SD=10.82ms), which was 1.56 times longer than original (Shockley et al: 2.03 times longer)

Procedure

The experiment used a slightly modified version of the imitation paradigm from Goldinger & Shockley et. al. The participants first read the list silently, to help avoid possible hyper-articulation in the main experiment. The stimuli were presented using Psyscope 1.2.5 (Cohen, et al., 1993)

1. **Warm-up Phase:** Subjects read the production list silently
2. **Baseline (Pre-study) Phase:** Subjects read the production list aloud
3. **Study Phase:** Subjects listened to the listening list (no other task)
4. **Test (Post-study) Phase:** Same as the Baseline Phase

The subjects' tokens were digitally recorded and VOTs were measured using both waveforms and spectrograms

Results & Discussion

Independent Variables:

- Pre vs. Post-study** (= Imitation Effect)
- Lexical Frequency** (high vs. Low)
- Heard vs. Unheard Items**
- Linguistic Unit of Generalization** (segment:/p/ vs. feature:/k/)

- **Significantly higher VOT was found in the post-study phase than in the pre-study phase, across all types of stimuli = imitation effect :** (F(1,7)= 6.488, $p < .05^*$)
- No statistically significant interaction between the imitation effect and other variables was found

Repeated-measures ANOVA with two within-subjects factors

- **lexical freq:** high and low frequency (F<1, $p > .1$), Interaction (F<1, $p > .1$)
- **heard vs. unheard:** heard vs. unheard (F<1, $p > .1$), interaction (F<1, $p > .1$)
- **pre vs. post x segments:** /p/ and /k/ (F(1,7)=125.797, $p < .001^*$), **interaction (F<1, $p > .1$)**

Results & Discussion

7 (4 male, 3 female) out of 8 subjects produced longer VOT in the post-study (=test) phase

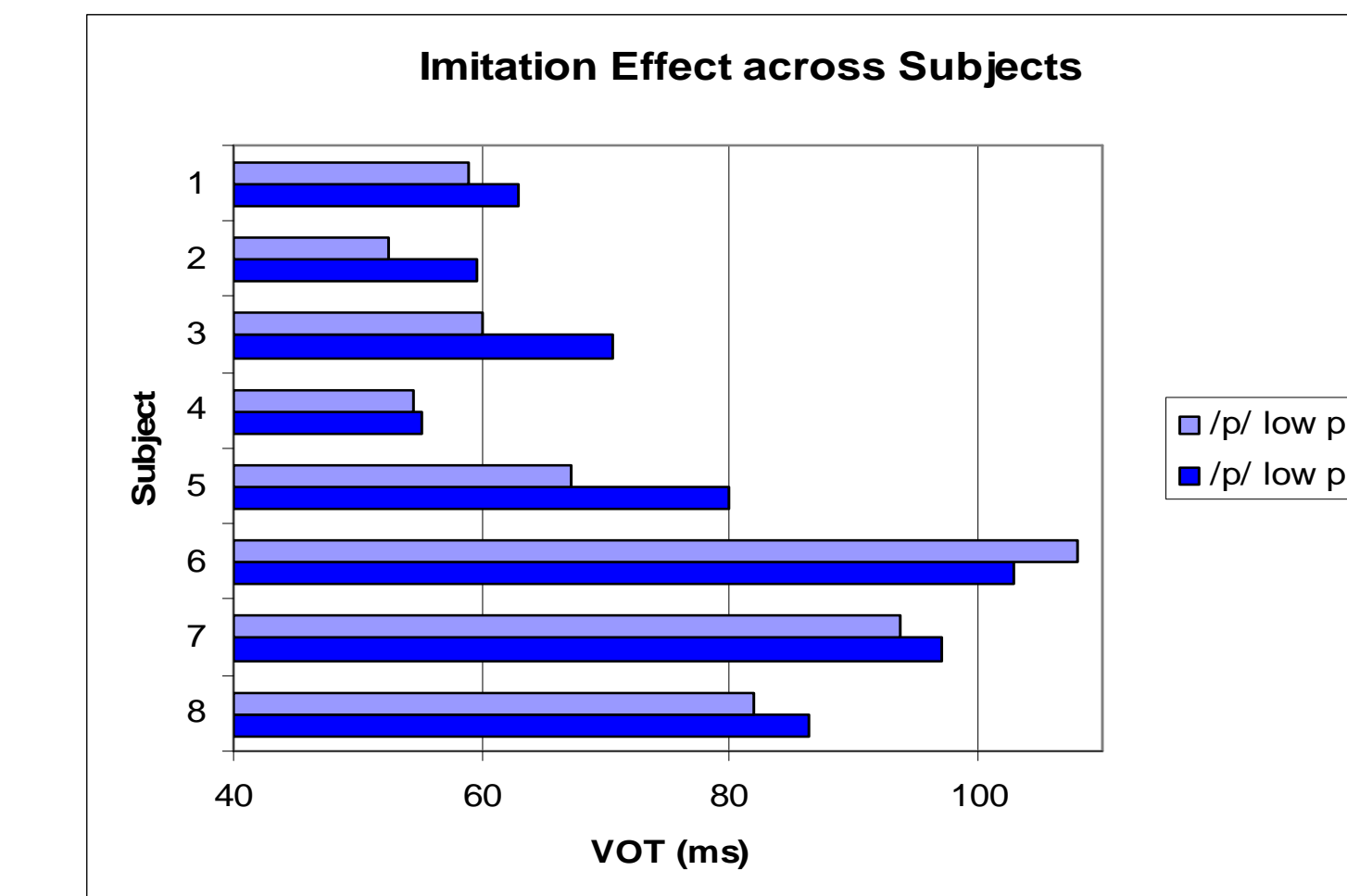


Figure 1: Imitation effect (VOT) for the low-frequency tokens presented in the study phase, plotted across eight subjects (subject # 1-4 are male, 5-8 are female).

The duration of the entire word did not change significantly: (F<1, $p > .1$)

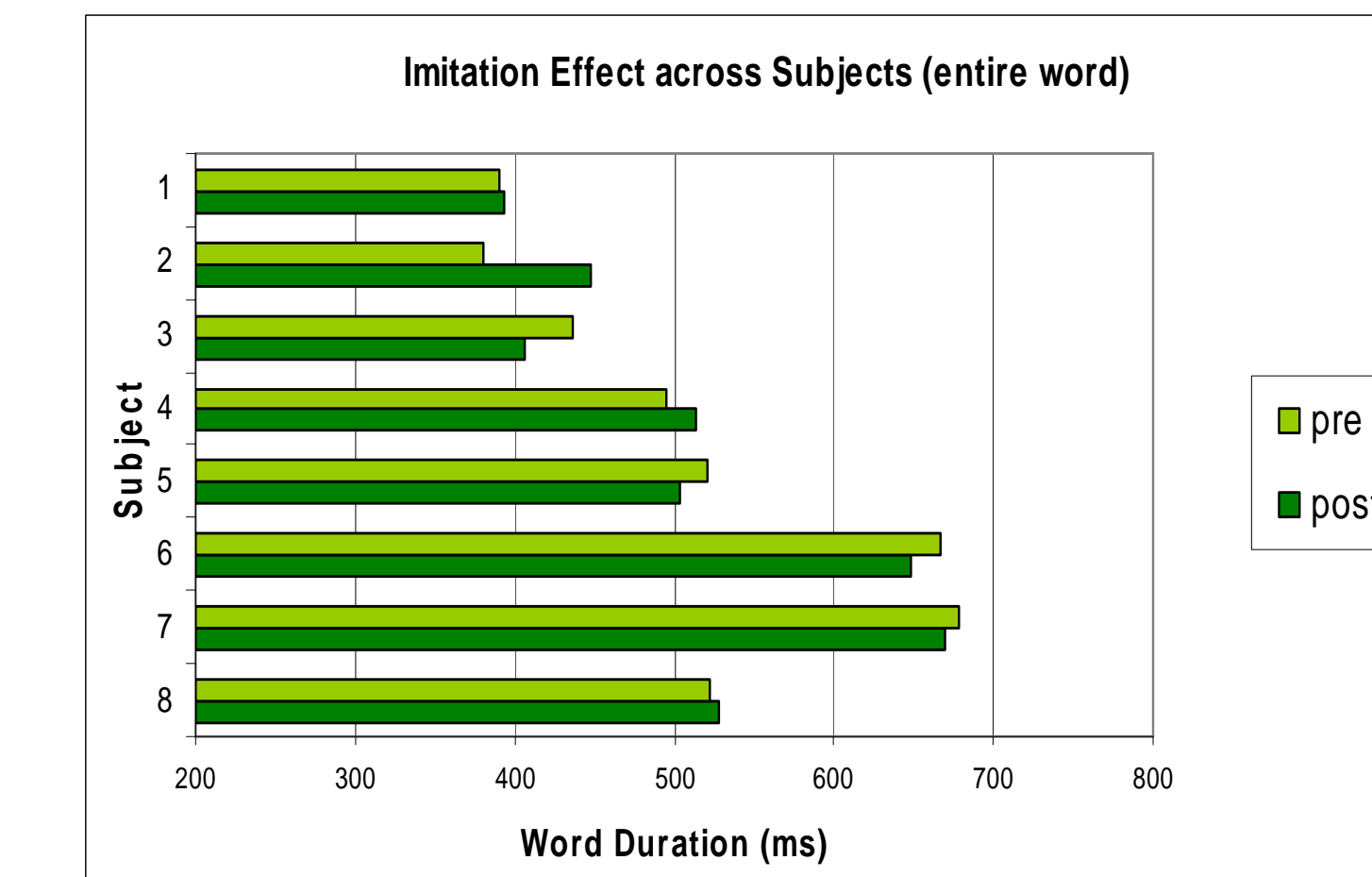


Figure 3: Whole-word durations for the low-frequency tokens presented in the study phase, plotted across eight subjects.

If the imitation effect is truly due to episodic memory, only the manipulated variable (in this case, VOT) should be affected

Results & Discussion

VOT imitation was found for the modeled words, new words with /p/, and new words with /k/; but no frequency effect

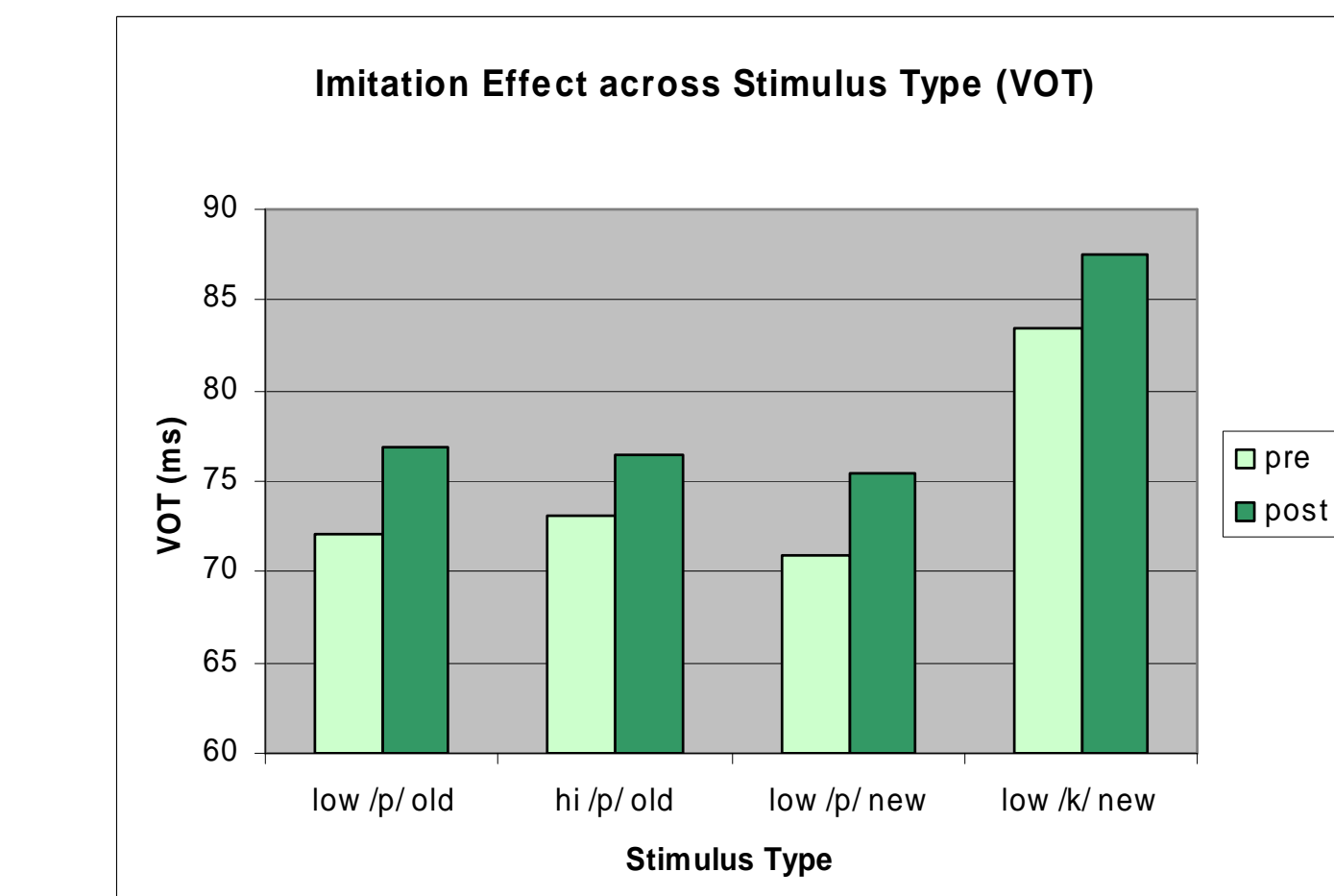


Figure 2: Imitation effect (in VOT) plotted across four types of stimuli: The imitation effect was significant for all types of stimuli, although there was no significant interaction with other factors.

		low /p/ old	hi /p/ old	low /p/ new	low /k/ new
Mean	pre	72.08	73.14	70.98	83.43
(SD)	pre	(20.31)	(19.87)	(19.63)	(19.37)
	post	76.84	76.38	75.37	87.52
	post	(17.66)	(19.26)	(18.18)	(14.96)

• **The imitation effect was generalized to:**
- **New words** which share the initial **phoneme /p/**
- **New segment /k/** which shares a feature [+spread glottis] (and: [-continuant, -sonorant, -voice]) = natural class

• **Exemplar theories predict the imitation effect to be stronger for low-frequency words**
- **Our results did not support the prediction**

Conclusions

- **The imitation effect for VOT was found in a single-word non-shadowing paradigm**
- **The effect was generalized to new words and new phonemes**
- **There was no frequency effect**

None of our results actually require an exemplar model. If the effect is triggered by exemplars, they must include featural information.

References

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.

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].

Goldinger, S. D. (1998). Echoes of Echoes? An Episodic Theory of Lexical Access. *Psychological Review*, 105 (2), 251-279.

Shockley, K., Sabadini, L., & Fowler, C. A. (2004). Imitation in shadowing words. *Perception & Psychophysics*, 66 (3), 422-429.

Goldinger, S. D. (2000). The Role of Perceptual Episodes in Lexical Processing. In *SWAP-2000*, 155-158.

Kučera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.

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.