

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. 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) showed that in single-word shadowing 1) subjects **shifted their productions** in the direction of the auditory target, 2) the effect was larger for **lower frequency** words. Shockley et al. (2004) replicated this effect with **extended aspiration (VOT)**.

That is, speakers implicitly imitate aspects of speech they are shadowing. However, because in a shadowing task speakers repeat all and only what they hear, it is impossible to determine whether they are adapting to the heard speech generally, or specifically to the particular words

The current study aims to examine how specific this phonetic imitation really is, by using a non-shadowing paradigm. Subjects are exposed to target speech, but they then produce not only the heard words, but also a wider set of words. Because the words share only some properties with the target speech, we can tease apart word-level imitation from a more **general imitation of sub-lexical units**.

Questions

1. Can phonetic imitation effect generalize to sub-lexical representations?

For example, /p/ and /k/ share the same acoustic cue “Long VOT”. Listeners’ knowledge of linguistic structure allows (and predicts) sub-segmental generalization from heard /p/ to unheard /p/ and /k/.

2. Is there a word-specific advantage in phonetic imitation?

An episodic view predicts a stronger specificity for 1) more recently experienced words, and 2) low-frequency words (Low frequency = fewer exemplars → weight of one exemplar is relatively larger).

3. Would the imitation effect be observed when a shift in production would impair a linguistic contrast?

Very short VOT of /p/ introduces linguistic ambiguity (confusion with voiced /b/), while there is no such danger for very long VOT. An extreme episodic view predicts the same imitation whether or not it endangers a contrast.

1. Generalizability of the effect to new linguistic units (phoneme, feature)
2. Effect of word specificity in VOT imitation effect (target vs. novel items, high vs. low freq)
3. Effect of linguistic contrast (key stimuli: Long VOT vs. Short VOT)

Manipulated Variable: VOT on /p/
(+40 ms / -40 ms)

Methods

Participants:

- 39 native speakers of American English (19M & 20F): 20 in Group1 (long VOT) and 19 in Group 2(short VOT)

Stimuli

- **Listening list** (for study-phase) **80 target words with initial /p/ - manipulated 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 the listening list)
 - (2) 20 new words, also with initial /p/
 - (3) 20 new words with initial /k/, which like /p/ is [+spread glottis]

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:

Lengthened by 40ms (for Group 1)

Spliced with the initial part of hyper-aspirated tokens to preserve natural formant transitions: The resulting tokens had average VOT of **113.26 ms** (SD=10.82ms)

Shortened by 40ms (for Group 2)

The most stable part of aspiration was taken out: The resulting tokens had average VOT of **32.29ms** (SD=12.39ms)

Procedure

The experiment used a slightly modified version of the imitation paradigm from Goldinger. The participants first read the list silently, to help avoid possible hyper-articulation in the main experiment.

1. **Warm-up:** Subjects read the production list *silently*
2. **Baseline:** Subjects produced (read) the production list aloud
3. **Listening:** Subjects heard the listening list (no other task)
4. **Test:** Same as the Baseline Phase

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

Results & Discussion

Within group factors:

- **Type of Production (Baseline vs. Test)**
- **Lexical Frequency (High vs. Low)**
- **Word Specificity (Target vs. Novel Items)**
- **Imitated Unit (/p/ vs. /k/)**

Between group factor:

- **Listening Stimuli (Long vs. Short VOT)**

Group1 (Lengthened VOT)

Imitation effect (Baseline vs. Test)
Significant (longer VOT in the Test phase)
($F(1,19)= 11.037, p < .005^*$)

Word-Specificity (target vs. novel)

Significant (stronger imitation effect for Target)
($F(1,19)= 13.221, p < .005^*$)

Lexical Frequency (high vs. low)

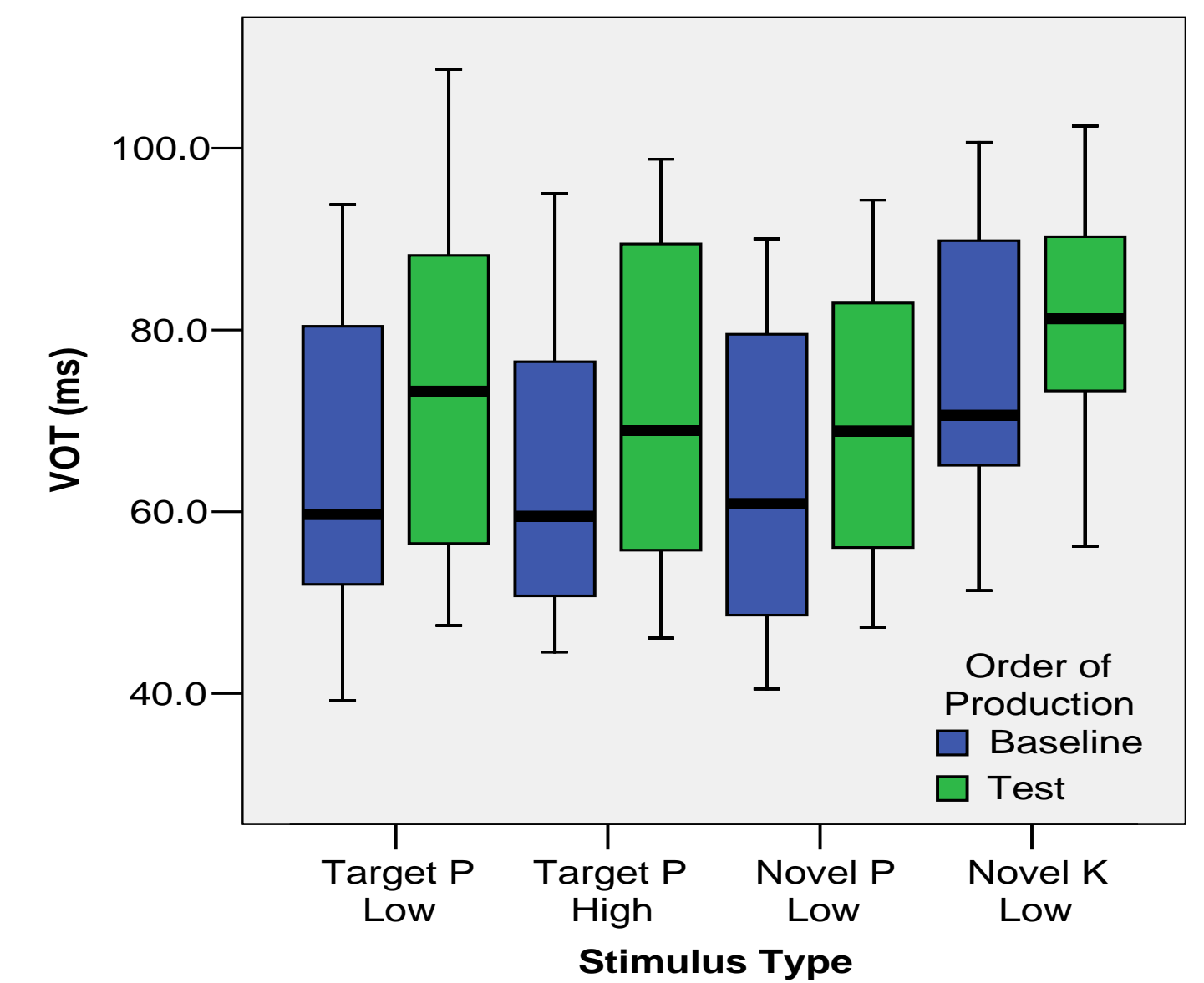
Not significant (numerically stronger imitation effect for low freq) ($F(1,19)= 3.935, p = .062$)

Imitation of lengthened VOT was **generalized** to novel items with **/p/ as well as /k/** (with **NO interaction** ($F(1,19)<1$))

Group 2: (Shortened VOT)

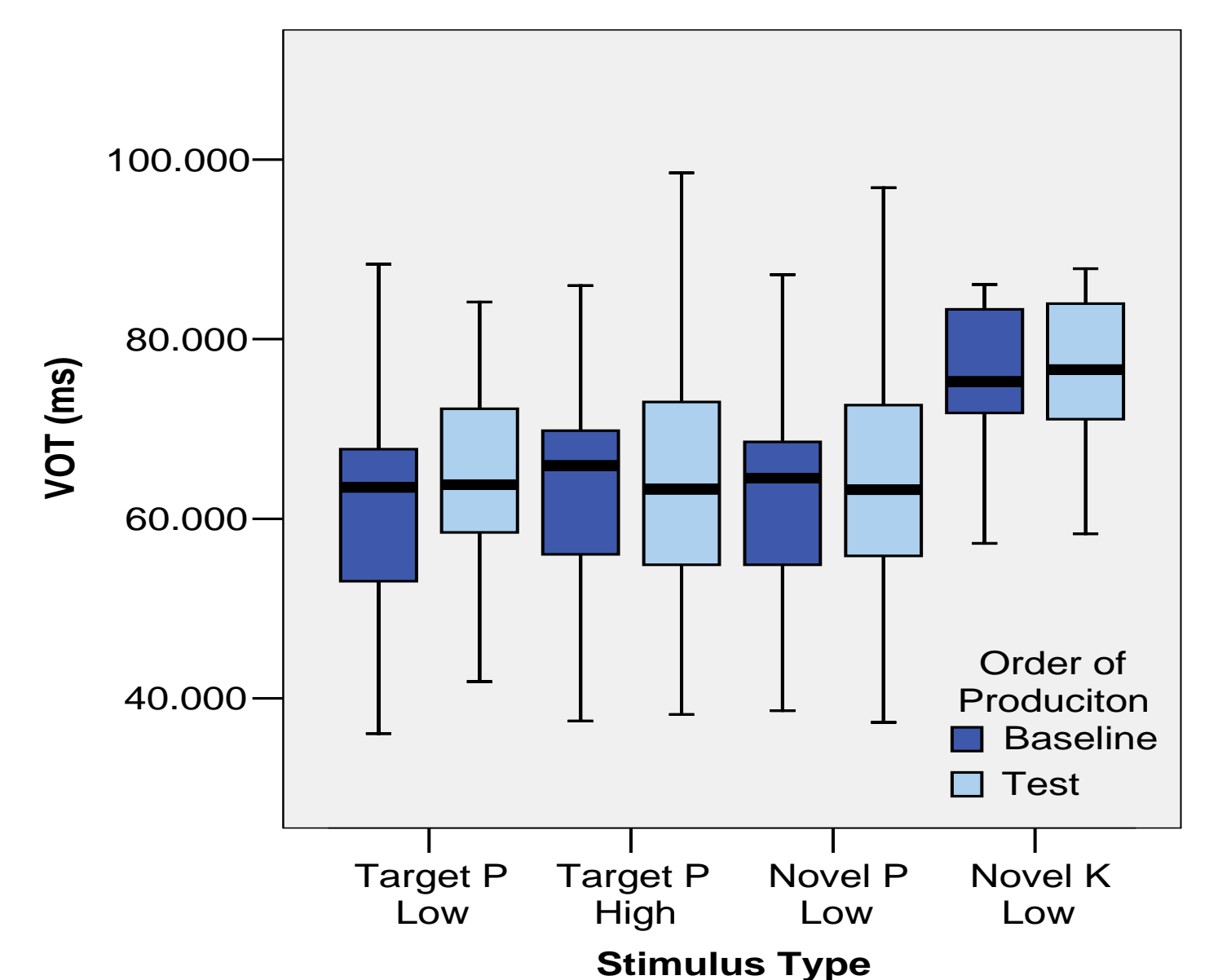
No significant difference between Baseline and Test production across all types of stimuli = **No imitation effect** ($F(1,18)= 2.201, p > .1$)
(No word-specificity, no generalization)

Clear interaction between **type of production x listening stimuli** ($F(1,75)= 23.990, p < .001^*$) = **Group Difference**



| 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 |

Figure 1: Imitation effect (in VOT) plotted across four types of stimuli (Group1: Listening Stimuli = Long VOT) (box = 1 SD, dark line = median, low/high = lexical frequency)



| 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 |

Figure 2: Imitation effect (in VOT) plotted across four types of stimuli (Group2: Listening Stimuli=Short VOT)

➤ 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**

➤ Actually heard items showed stronger imitation effect than unheard items, supporting a prediction by the exemplar view. However, the expected frequency effect on imitation was not observed.

➤ Subjects imitated long VOT, but did not imitate short VOT

Knowledge of linguistic contrasts modulates the phonetic imitation

Similar asymmetrical results in VOT Goodness Rating (Allen & Miller, 2001)

Conclusions

This study showed:

- Some predictions of the extreme exemplar view were confirmed, namely, **phonetic imitation** and **word specificity**
- But also, speakers are sensitive to linguistic structure:

Sub-phonemic Level and Phonemic Contrast

The results call for a Linguistically informed exemplar model of speech perception, which incorporates both **sub-segmental and word-level representation** as well as knowledge of **linguistic contrast**

- Compatible with models of spoken word recognition with sub-phonemic units
- Parallel to proposed exemplar models of speech production (e.g. Pierrehumbert 2002)