REPRESENTATIONAL SPECIFICITY OF LEXICAL FORM IN THE PRODUCTION AND PERCEPTION OF SPOKEN WORDS

Paul A. Luce, Jan Charles-Luce, and Conor McLennan
University at Buffalo, Buffalo, New York, USA

ABSTRACT

We examined the specificity and abstractness of form-based lexical representations of spoken words. Using a repetition-priming paradigm, we attempted to determine if flapped intervocalic alveolar stops in American English are mapped onto underlying representations of /t/ and /d/ or if flaps have an independent representational status. Subjects shadowed spoken words in two blocks of trials. Stimuli in the first block served as primes and those in the second block as targets. Primes and targets consisted of flapped and carefully articulated bisyllabic words. We measured shadowing times to target words in the second block as a function of prime type. We also examined subjects' productions of the targets. In particular, we measured vowel duration as a correlate of voicing. The shadowing time results provide evidence for underlying perceptual representations. The analysis of subjects' productions indicate that underlying representations subserving speech production can be primed to accentuate the surface realization of underlying /t/ and /d/.

1. INTRODUCTION

Recent psychological theories of lexical representation have proposed that mental representations of lexical form are exemplar-based, preserving much of the surface information present in the speech signal [5,6]. According to these models, lexical representations contain specific details of particular spoken words, such as allophonic variation, talker identity, and speaking rate.

Exemplar models of the mental lexicon contrast with traditional psychological and linguistic models that postulate the existence of abstract, underlying forms [11]. According to these models, surface details of spoken words are discarded during perceptual processing and a match is attempted between the resulting recoded information and an underlying lexical representation stored in memory. Thus, traditional models of spoken word perception have assumed that the underlying representation is recovered from its surface phonetic realization. Similarly, traditional models of speech production [3,4] have assumed that underlying representations of lexical form are realized as surface forms via phonetic implementation and phonological rules. In short, traditional models of production and perception assume abstract underlying representations that do not include specific details.

We evaluated exemplar and traditional models of lexical representation by examining flapping in American English. Under a traditional analysis, flaps are thought to be neutralized surface realizations of the underlying phonemic forms /t/ and /d/. Thus, it is hypothesized that *only* forms containing fully specified /t/ and /d/ are represented in memory. However, according to an exemplar-based view of lexical representation, flapped tokens of spoken words should have full representational status. Indeed, given that listeners of American English most frequently encounter flapped segments intervocalically, exemplar-based models would ascribe representational priority to the flapped representation because of its higher frequency of occurrence.

We investigated the representation of flaps in American English using the repetition priming paradigm, a standard experimental technique used to examine the nature of representations in memory. In long-term repetition priming experiments, stimuli are presented in two blocks. The first block consists of *primes* and the second block consists of *targets*. The dependent variable of interest is a change in the subject's response to the target caused by prior presentation of a prime. Facilitation of target processing by a related prime indicates that the prime and target share a common representation in memory. More specifically, presentation of a prime activates its representation. As a results subsequent stimuli that also activate this representation are made easier to process. The phenomenon of facilitative repetition priming can be used to determine if two nominally different stimuli activate the same mental representation.

In the current experiment, we presented both carefully articulated and flapped versions of bisyllabic words (e.g., "rater," "raider") as primes and targets. Subjects shadowed, or repeated, the targets as quickly as possible. Shadowing times served as indices of speed of processing. We attempted to determine if flapped primes facilitate processing of carefully articulated targets, and vice versa. Failure to observe equivalent levels of priming between flapped and carefully articulated stimuli would indicate that the two do *not* share a common underlying representation. This result would be consistent with exemplar-based models of lexical representation. However, equal facilitation of target processing by flapped and careful primes would indicate the existence of a shared, and presumably abstract, representation. This result would be consistent with more traditional models.

2. EXPERIMENT

2.1. Method.

2.1.1. Subjects. Twenty-four members of the University at Buffalo community participated in the experiment. All subjects were native American English speakers and reported no history of speech or hearing disorders.

CONDITION	BLOCK 1: PRIMES		BLOCK 2: TARGETS
1	Flap (ɹeɪɾɚ)	➡	Flap (ɹeɪɾɚ)
2	Flap (ɹeɪɾɚ)	➡	Careful /t/ (ɹeɪtɚ)
3	Flap (ɹeɪɾɚ)	➡	Careful /d/ (ɹeɪdɚ)
4	Careful /t/ and /d/ (ɹeɪtɚ/ɹeɪdɚ)	➡	Flap (ɹeɪɾɚ)
5	Careful /t/ (ɹeɪtɚ)	➡	Careful /t/ (ɹeɪtɚ)
6	Careful /d/ (ɹeɪdɚ)	➡	Careful /d/ (ɹeɪdɚ)
7	None (control)	➡	Flap (ɹeɪɾɚ)
8	None (control)	➡	Careful /t/ (ɹeɪtɚ)
9	None (control)	➡	Careful /d/ (ɹeɪdɚ)

Table 1. Stimulus conditions.

2.1.2. Materials. Twelve pairs of bisyllabic English words containing intervocalic /t/ or /d/ were selected. Examples of word pairs are *rater-raider* and *utter-udder*.

The 12 pairs of words were read twice by a phonetically sophisticated male speaker, once carefully articulated and once with the intervocalic /t/ or /d/ flapped. The stimuli were recorded in a sound-attenuated room, low-pass filtered at 10 kHz, and digitized at a sampling rate of 20 kHz using a 16-bit analog-to-digital converter. All words were edited into individual files and stored on computer disk.

The /t/ and /d/ members of the stimulus pairs were matched on overall log frequency of occurrence [7]. The average log frequencies for /t/ and /d/ words were .59 and .31, respectively. This difference was not significant (t (22) = 1.23, $p > .20$).

The stimuli were screened by ten listeners who indicated whether each word contained a /t/ or /d/. For the carefully articulated stimuli, at least nine out of ten listeners correctly identified the intended voiced or voiceless segment. For the flapped stimuli, no more than six out of ten subjects labeled any of the individual flapped items as containing a /t/ or /d/. On the average, fifty percent of listeners identified the flaps as /t/ and fifty percent as /d/. Thus, the flapped segments were perceived to be ambiguous.

In addition to the 12 stimulus pairs, 24 bisyllabic filler words were also recorded and digitized. These items contained intervocalic velar and bilabial stops.

The stimuli were presented in two successive blocks consisting of 24 stimuli each. Block 1 consisted of *primes* and Block 2 consisted of *targets*. Orthogonal combination of three prime types (flapped, careful, and none) and three target types (flapped, careful /t/, and careful /d/) resulted in nine conditions, which are illustrated in Table 1. For example, in Condition 1, a flapped prime in Block 1 was followed by the same flapped item in Block 2. In Condition 2, a flapped prime in Block 1 was followed by a carefully articulated version of the same word in Block 2, and so on. The remaining stimuli in Blocks 1 and 2 were fillers.

No version of a given word was presented more than once in a given block. Across all 24 subjects, each word appeared in each condition.

2.1.3. Procedure. Subjects performed a single-word shadowing task in both blocks of trials. At the beginning of the experiment, subjects were *not* instructed that they would be presented with two blocks of trials. Subjects were not informed about the number of blocks to avoid inducing any undesired strategic processing of stimuli in Block 1. The subjects were tested individually. A Macintosh Centris 650 controlled stimulus presentation and response collection. The stimuli were presented over headphones at a comfortable listening level. Subjects responded by repeating the word that they heard as quickly as possible into a microphone attached to their headphones. Shadowing times were recorded from the beginning of the stimulus to the onset of the subject's verbal response.

Ten practice trials were presented to familiarize the subjects with the task. After completing Block 1, subjects performed math problems for five minutes before being presented with Block 2. A typical trial proceeded as follows: A cue light at the top of a response box situated in front of the subject illuminated for 500 ms, indicating that a stimulus was about to be presented. A stimulus was then presented 250 ms after the offset of the cue light. Once the subject responded, an intertrial interval of 500 msec was initiated. Each subject received a different random order of stimuli in each block.

Subjects' shadowing responses were recorded on digital minidisk for accuracy analysis and measurement.

2.2. Results.

2.2.1. Shadowing times. A 3 (prime: flap, careful, none) X 3 (target: flap, careful /t/, careful /d/) repeated measures analysis of variance was performed on the shadowing times for target stimuli in Block 2. We obtained significant effects of target (F (2,46) = 19.28, $p < 0.05$) and prime (F (2,46) = 4.09, p (F<1.0). The main effect of target was due to faster shadowing times for flapped compared to carefully articulated stimuli. More important, the main effect of prime revealed that subjects' shadowing times to targets in the second block varied as a function of the prime.

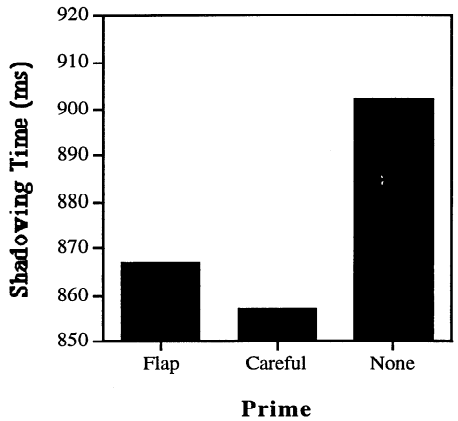

Figure 1. Shadowing times for targets in Block 2 as a function of prime in Block 1. Results are collapsed across target type.

As shown in Figure 1, targets primed by flapped and careful stimuli were shadowed more quickly than targets in the control condition (i.e., targets preceded by no prime). Planned contrasts revealed significant differences between the flapped and no prime conditions (F (1,46) = 4.46, $p < .05$) and between the careful and no prime conditions (F (1,46) = 7.44, $p < .05$). However, the 10 ms difference between the flap and careful prime conditions was not significant ($F<1.0$).

Flapped and carefully articulated primes produced significant facilitative effects on shadowing times to targets in Block 2. The magnitude of the priming effect was not statistically different across the three types of targets (flapped, careful /t/, careful /d/). Furthermore, flapped primes facilitated target shadowing as much as carefully articulated primes. In other words, flapped primes were just as effective as carefully articulated primes. These results are consistent with a model that posits shared underlying representations for flapped and carefully articulated stimuli.

2.2.2. Production measurements. In addition to examining shadowing times, we measured vowel duration preceding the intervocalic stop in the target words produced by the subjects during the shadowing task. Preceding vowel duration is a well-known correlate of medial stop voicing, with longer durations associated with voiced consonants [see, for example, 8]. We were interested in determining the degree to which priming a representation in memory affects subjects' subsequent articulation of similar repeated items. Again, differential effects of priming may be indicative of the nature of the representations subserving production of flapped and carefully articulated items.

Vowel duration was measured for target stimuli for a subset of the 24 subjects. The results of this analysis are shown in Figure 2. Results are plotted as a function of prime and target type. The top panel of Figure 2 shows mean vowel durations for flapped, careful /t/, and careful /d/ targets preceded by flapped primes. The middle and bottom panels show the mean results for each target type in the careful and no prime conditions, respectively.

No significant differences in vowel duration for flaps, /t/s, and /d/s were observed for the flap and no prime conditions (both Fs < 1.0; top and bottom panels in Figure 2). In these two prime conditions, subjects' productions failed to preserve a voice contrast across the three target types, at least as indicated by differences in vowel duration. However, subjects produced significant differences in vowel duration for the careful prime condition (F (2,16) = 3.40, $p < .05$). Subjects produced shorter preceding vowel duration in the /t/ target words and significantly longer preceding vowel duration in the /d/ and /r/ target words.

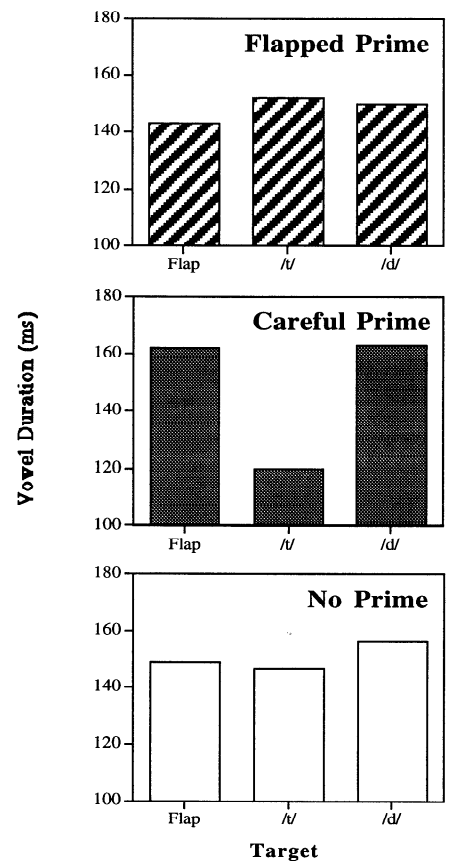

Figure 2. Target vowel duration as function of prime type.

3. DISCUSSION

Our investigation of lexical representations of spoken words containing intervocalic flaps revealed two notable findings: First, evidence from shadowing times demonstrated that flapped and carefully articulated words are equally effective perceptual primes for flapped and carefully articulated targets. That is, both types of primes produced equivalent levels of facilitation for both types of targets. Second, evidence from measurements of subjects' productions indicated that carefully articulated primes served to enhance the voice contrast (as indexed by vowel duration) for carefully articulated /t/ and /d/ words.

The shadowing time results suggest that representations shared by both flapped and carefully articulated words are contacted during the perceptual process. These shared

representations are presumably responsible for the equivalent levels of facilitation observed in this experiment (see Figure 1). If effects of priming were restricted to veridical, surface representations of primes and targets -- as predicted by a strict exemplar model -- we would have expected markedly different patterns of priming between flapped and carefully articulated words. In particular, facilitative priming between *identical* items (i.e., flap to flap and careful to careful) should have been greater than facilitative priming between *different* versions of the stimuli (i.e., flap to careful and vice versa). Instead, we observed equivalent degrees of facilitative priming for flapped and careful primes and targets. It appears that flapped stimuli activate representations containing both /t/ and /d/, which in turn serve to facilitate subsequent processing of flapped and carefully articulated stimuli.

However, the present shadowing results do *not* constitute evidence against exemplar-based representations of form. There is considerable evidence that specific, instance-based representations of spoken words are stored in memory [2,5,9] and that representational specificity has implications for processing [10]. Instead, our results suggest that the type of allophonic variation examined in the present study requires activation of underlying representations in addition to veridical surface forms [see also 9]. We propose that because the listener must, under normal listening conditions, map a potentially ambiguous flapped word onto one of two lemmas, phonological forms corresponding to clearly delineated voicing categories (i.e., /t/ and /d/) are contacted during the recognition process. Thus, our results suggest that -- at least in the recognition of flapped items in American English -- recovery processes do indeed map the surface manifestation of the flap onto representations in which underlying abstract voicing categories are well specified.

The vowel duration results support a similar -- but nonetheless somewhat divergent --view regarding the nature of the representations supporting production of spoken words. We found that, in all but one condition, vowel duration failed to preserve a voice contrast. When carefully articulated /t/ or /d/ target words were preceded by *carefully articulated* primes, subjects produced vowel duration differences that preserved the /t/-/d/ contrast. However, when carefully articulated /t/ and /d/ words were preceded by *flapped* or *no* primes, speakers did not preserve the contrast. To account for these findings, we hypothesize that the voice contrast is most frequently neutralized for intervocalic alveolar stops in American English. However, the effect of the careful prime is to activate more strongly those production codes responsible for implementing phonemic /t/ and /d/ (as opposed to /ɾ/) [1]. Thus, as indexed by preceding vowel duration, the voice contrast was preserved only in subjects' repetitions of the *primed*, carefully articulated stimuli. The effect of the careful prime on the careful /t/ or /d/ target was to minimize the tendency of the speaker to neutralize the /t/-/d/ contrast.

The priming of a flapped target by carefully articulated /t/ and /d/ words was apparently not sufficient to overcome the stronger and more natural tendency to flap. Inspection of the vowel duration data revealed that there was no appreciable difference between target flaps primed by careful /t/s and those primed by careful /d/s. It is also possible that the flap itself may have an independent representational status in the production system and be unaffected by priming from a carefully articulated stimulus. Thus, unlike in the recognition system, in which flaps are mapped onto representations with clearly bifurcated voicing categories, the production system may assign equal and independent representational status to /ɾ/, /t/, and /d/.

Our results suggest that models of word recognition and production must incorporate aspects of both abstract and exemplar-based representations. Whereas production models may need to assign full representational status to the flap as one of a triumvirate of alternatives for invervocalic alveolar stops, recognition models appear to require some sort of recovery process to map flapped patterns onto representations with underlyingly well-specified voicing categories.

ACKNOWLEDGEMENTS

This research was supported (in part) by research grant number 1 R01 DC 0265801 from the National Institute on Deafness and Other Communication Disorders, National Institutes of Health. We thank Stacey Zelkind for her assistance with the acoustic measurements and Dan Gaygen for his help running subjects.

REFERENCES

[1] Charles-Luce, J. (1997). Cognitive factors involved in preserving a phonemic contrast. *Language and Speech*, 40, 229-248.

[2] Church, B. A., & Schacter, D. L. 1994. Perceptual specificity of auditory priming: Implicit memory for voice intonation and fundamental frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 43-57.

[3] Dell, G. 1986. A spreading activation theory of retrieval in sentence production. *Psychological Review*, 93, 283-321.

[4] Dinnsen, D. A. & Charles-Luce, J. 1985. Phonological neutralization,phonetic implementation and individual differences. *Journal of Phonetics*, 12,49-60.

[5] Goldinger, S. D. 1996. Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 22, 1166-1183.

[6] Goldinger, S. D. 1998. Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, 105, 251-279.

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

[8] Luce, P. A., & Charles-Luce, J. 1985. Contextual effects on vowel duration, closure duration, and the consonant/vowel ratio in speech production. *Journal of the Acoustical Society of America, 78,* 1949-1957.

[9] Luce, P. A., & Lyons, E. A. 1998. Specificity of memory representations for spoken words. *Memory and Cognition, 26,* 708-715.

[10] Mullennix, J. W., Pisoni, D. B., & Martin, C. S. 1989. Some effects of talker variability on spoken word recognition. *Journal of the Acoustical Society of America*, 85, 365-378.

[11] Pisoni, D. B., & Luce, P. A. 1987. Acoustic-phonetic representations in word recognition. *Cognition*, 25, 21-52.