IMPLICIT MEMORY FOR SILENT-CENTER SYLLABLES

Susan L. Hura
Department of Audiology & Speech Sciences, Purdue University

ABSTRACT

Recent work has shown that specific characteristics of spoken words are encoded in memory. Exposure to a particular token of a word improves listeners' ability to identify the word later . This study extends this paradigm to silent-center (SC) stimuli, in which vowel nuclei are attenuated to silence. Results address the claim that CV transitions in SC syllables are privileged in perception.

INTRODUCTION

A growing body of evidence suggests that information about specific voices and specific exemplars of words is encoded in memory [1]. The effects of these individual memory traces are best observed in implicit memory tasks in which listeners perform some perceptual classification of stimuli. These findings are in opposition to traditional theories of lexical representation which rest on the assumption that there is a single, abstracted entry for each item in the lexicon. Moreover, the implicit memory findings speak against the traditional belief that individual characteristics in the speech signal are a hindrance to perception, in that they will be difficult to match up with abtstract lexical or phonemic catgegories.

LEXICAL ACCESS AND IMPLICIT MEMORY

Current theories of lexical representation posit a single, abstracted phonemic representation for each item in the mental lexicon (e.g.[6]). Recognizing spoken words is thus a process of matching the variable incoming signal with stored canonical representations. Describing the process by which this match takes place has been the goal of theories of speech perception. In many cases, some invariant cue associated with each phoneme is proposed. One of the most recent claims, put forth by proponents of the theory of dynamic specification, is that dynamic vowel information contained in consonant-vowel transitions may be an invariant cue for vowel recognition [4].

Contrary to the traditional view, recent work in the field of implicit memory suggests that there may be multiple entries in the lexicon for each item. There is evidence that detailed information about specific voices and exemplars of words may be encoded in memory and used in speech perception. Goldinger [1] showed that previous exposure to particular tokens of words significantly improved listeners ability to accurately identify those words later. This advantage decreased over time, but was still observable after a delay of up to a week. These results suggest that listeners retain information about particular perceptual events and use this information in later perceptual judgments. Goldinger's study is part of the larger field of implicit memory [3, 5] which suggests that individual episodic memory traces are formed for many perceptual events.

DYNAMIC SPECIFICATION OF VOWELS

Results from Hura [2] suggest that implicit memory effects apply to silent-center (SC) versions of syllables in which vowels have been attenuated to silence. When listeners were exposured to a mixed set of unaltered Full syllables and the corresponding SC syllables, identification performance on the SC stimuli was significantly better than when listeners heard SC syllables alone. This implies that individual episodic traces of Full syllables were encoded in memory and accessed in identification of SC versions. The current study seeks to replicate this finding in a more controlled fashion.

The theory of dynamic specification characterizes vowels as gestures having intrinsic timing parameters. This theory rests on studies of silent-center stimuli, in which syllable nuclei, including static formant patterns of vowel targets, have been removed. The remaining dynamic information often enables listeners to identify vowels in SC syllables as accurately as vowels in full syllables. Consequently, dynamic portions of the speech signal are seen as privileged in perception. Moreover, hybrid SC syllables, which are composed of the initial CV transition from one talker and the final VC transition from another talker, may in some cases yield very accurate vowel perception [4]. This suggests that dynamic vowel information may be speaker-independent, thus greatly simplifying the problem of matching the incoming speech signal with stored canonical representations.

The study reported here is an initial step towards extending implicit memory paradigms to silent-center syllables. Listeners were presented with a set of unaltered CVC syllables for identification. After a variable delay, listeners were tested on a set of silent-center syllables, half of which were part of the set of Full syllables heard originally. Performance on previously heard syllables is compared with new syllables. Effect of length of delay between the two sets of stimuli is also considered.

METHOD

Stimuli

The stimuli for this study began with a set of 10 /bVt/ syllables spoken by an adult male talker. Nine of the ten syllables are real English words: /bit/, /bɪt/, /bet/, /bɛt/, /bæt/, /bɑt/, /bət/, /bot/, and /but/; the tenth syllable, /bʊt/ is a phonologically-possible nonword. Syllables were produced in isolation, and recorded on audio tape using a Sony TC-FX420R cassette recorder and an Audio-Technica ATM11 unidirectional microphone. A single version of each syllable was low-pass filtered at 8 kHz, and digitized at 20 kHz using Kay CSL equipment.

For each syllable, Full and SC stimuli were created. Full syllable stimuli consisted of whole, unmodified /bVt/ syllables, including prevoicing of the /b/ and release burst of the /t/, if present in original recordings. Syllable duration, for the purposes of generating SC stimuli, was measured from the release of initial /b/ closure to onset of final /t/ closure. Silent-center stimuli were created by attenuating to silence all but the first three and last four pitch periods of each syllable.

Each SC syllable was then embedded in white noise, with a 50 msec noise lead before onset of the syllable to guard against overshoot of masking. The amplitude of the noise used with each SC syllable was adjusted to match the energy level of the corresponding Full syllable, yielding a 0 dB S/N in each case. Due to naturally-occurring differences in the overall amplitude of vowels, SC-in-noise stimuli were not at equal dB values across the stimulus set.

In all 20 stimuli were constructed, a Full and SC-in-noise version for each of 10 /bVt/ syllables. Two listening tests were constructed for Full syllable stimuli. Each was composed of 20 repetitions each of 5 of the 10 Full syllable stimuli presented in random order. Test Full 1 contained the syllables /bit/, /bɛt/, /bɑt/, /bot/, and /bʊt/; test Full 2 contained the syllables /bɪt/, /bet/, /bæt/, /bət/, and /but/. Syllables were separated such that tense and lax versions of vowel pairs would occur on different tapes (e.g. /bit/ versus /bɪt/), and such that each tape would contain an equal number of tense and lax vowels overall. Items were presented in blocks of 10,with 3500 msec ISI and 7 sec between blocks. There were a total of 100 items on each tape, which lasted approximately 9 minutes.

A single listening test was constructed for SC stimuli. Twenty repetitions of each of the 10 stimuli occurred in random order. As in the Full syllable tests, stimuli were presented in blocks of 10, with 3500 msec ISI, and 7 sec between blocks. The SC tape lasted approximately 18 minutes. All three listening tests were recorded onto digital audio tape using a Sony DCT-690 DAT deck.

Subjects

Thirty-one undergraduate students enrolled in a speech acoustics course at Purdue University served as listeners. All were native speakers of English and none reported a history of speech or hearing problems. They received partial course credit for their participation.

Data from 28 listeners was included for analysis. The remaining listeners were excluded because they produced greater than 10% errors on the initial Full syllable test.

Procedures

Listeners were tested individually ina laboratory setting. Before the initial test, listeners read a set of printed instructions and were allowed to ask questions. Stimuli were presented via Sony MDR-V400 dynamic stereo headphones at a comfortable listening level. Listeners performed an identification task, and responded by pressing a single key on a computer keyboard which was marked with the /bVt/ word. In this initial phase, listeners heard either tape Full 1 or tape Full 2.

A delay of 5 minutes or 1 day elapsed between the intial Full syllable testing and testing on SC stimuli, with half the subjects in each delay condition. Listeners again read a set of printed instructions before the test, and also heard a short introduction tape. This tape contained 10 Full syllables spoken by an adult female talker, each embedded in wide-band noise. The purpose was to familiarize listeners with the sound of the noise. After asking any questions they had, each listener identified the stimuli on the SC tape, making their responses as for the Full syllable test.

RESULTS

Listeners' responses on both tests were scored for percent errors. Results of the Full syllable tests were used as a criterion for subject inclusion: data from any listener who made more than 10% errors identifying unmodified Full syllables was excluded.

Listeners' responses on the SC test were also scored for percent errors, which averaged 38.5% overall. Listeners were treated as two separate groups, depending on which Full syllable test they heard. Full 1 listeners averaged 39.3% errors; Full 2 listeners averaged 37.2%. Listeners were also classified according to delay between tests: listeners in the 5 Minute condition averaged 37.4% errors; listeners in the 1 Day condition averaged 39.7% errors. There were also differences in error rate across individual vowels, ranging from a high of 92% on /u/ to 1.4% for /æ/. In all subsequent analyses, results reported are collapsed across vowels.

A 2 X 2 X 2 repeated measures analysis of variance was conducted, with Full syllable tape and Delay as between subjects factors and repetion (i.e. Old versus New) as a within subjects factor. There was no significant main effect of Full syllable tape ($F(1,24) = .98$, NS), or for Delay ($F(1,24) = .77$, NS). Repetition, however, was significant ($F(1,24) = 6.73$, $p>.02$). There were significantly fewer errors on previously heard (Old) stimuli than on New stimuli (35.8% versus 41.3%, respectively).

There was only one significant interaction, that of Full syllable tape by Delay ($F(1,24) = 24.5$, $p>.01$). That is, the effects of repetition of stimuli differed for the two Full syllable tapes. Table 1 lists the average percent errors for the four Full syllable tape by repetition conditions. These data show that the repetition effect is stronger for Full 2 stimuli than for Full 1.

Table 1. Mean percent errors for Old and New stimuli for each Full syllable tape.

	Old	New
Full 1	42.3	37.3
Full 2	29.2	45.2

Although the higher order interactions do not achieve significance, trends for effects of Delay can be observed. Specifically, the difference in error rate between Old and New stimuli is greater for the 5 Minute delay than for the 1 Day delay. Table 2 lists the average percent errors for Old and New stimuli at each delay.

Table 2. Mean percent errors for Old and New stimuli for each Delay.

	Old	New
5 Minutes	33.3	41.5
1 Day	38.2	41.1

DISCUSSION

These results provide support for the claim that previous exposure to particular versions of spoken words can improve a listener's ability to correctly identify the word later . There is a strong tendency in these data for previously heard words to be more accurately recognized than words that had not been previously heard. Recall, however, that nine of the ten /bVt/ syllables tested are reasonably common words, with which all subjects were familiar. Therefore, the difference between "old" and "new" stimuli in this study is achieved solely by exposure to the words during the Full syllable test. This is strong evidence for the claim that listeners are accessing a memory trace particular to each token, rather than simply matching a single canonical representation in the lexicon.

Moreover, this study shows that when a memory trace is set up for a Full syllable, it is later accessed in identification of a SC version of the syllable. This is evidence that the portion of the syllable maintained in SC's retains some of the distinct, individual characteristics of the Full syllable. That is, the partial pattern of an SC syllable appears to be easily related to the complete Full syllable pattern for particular tokens. The question that remains, however, is whether any other partial stimulus might be as easily related to the complete memory trace. This is one question for future research in this area.

Another important future question is whether memory traces may be set up for SC syllables. That is, if a listener first hears an SC syllable, will there be an advantage in later perceptual classification of the corresponding Full syllable? Such data would allow us to make a more definitive statement about the dynamic portions of the syllable contained in silent-centers.

REFERENCES

[1] Goldinger, S. (1992), *Words and voices: Implicit and explicit memory for spoken words*. (Technical Report No. 7). Bloomington, Indiana: Indiana University, Speech Research Lab.

[2] Hura, S. (1994), Dynamic aspects of vowel perception. Doctoral dissertation: University of Texas at Austin.

[3] Jacoby, L., Marriott, M., and Collins, J. (1990), The specifics of memory and cognition. In T. Skrull and R. Wyer (Eds.), *Advances in Social Cognition, Vol. 3* , Hillsdale, NJ: Erlbaum.

[4] Jenkins, J., Strange, W., and Miranda, S. (1994), Vowel identification in mixed-speaker silent-center syllables. *Journal of the Acoustical Society of America*, vol. 95, pp. 1030-1043.

[5] Schacter, D. (1987), Implicit memory: History and current status. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 13, pp. 501-518.

[6] Studdert-Kennedy, M. (1976), Speech perception. In N. Lass (Ed.), *Contemporary Issues in Experimental Phonetics*, New York: Academic Press.