UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

JAN 22 1993

AD-A261 439

REPORT DOCUMENTATION PAGE

DTIC ELECTE MAR 02 1993 S E D

1a. REPORT SECURITY CLASSIFICATION: UNCLASSIFIED

1b. RESTRICTIVE MARKINGS:

2a. SECURITY CLASSIFICATION AUTHORITY:

2b. DECLASSIFICATION / DOWNGRADING SCHEDULE:

3. DISTRIBUTION / AVAILABILITY OF REPORT: Approved for public release; distribution Unlimited.

4. PERFORMING ORGANIZATION REPORT NUMBER(S):

5. MONITORING ORGANIZATION REPORT NUMBER(S): AEOSR-TR- 93 0042

6a. NAME OF PERFORMING ORGANIZATION: Harvard University

6b. OFFICE SYMBOL (If applicable):

7a. NAME OF MONITORING ORGANIZATION: Air Force Office of Scientific Research/NL

6c. ADDRESS (City, State and ZIP Code): Department of Psychology, 33 Kirkland Street, Cambridge, Mass. 02138

7b. ADDRESS (City, State and ZIP Code): Building 410, Bolling AFB, DC 20332-6448

8a. NAME OF FUNDING/SPONSORING ORGANIZATION: AFSOR

8b. OFFICE SYMBOL (If applicable): NL

9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER: AFOSR #89-0461

8c. ADDRESS (City, State and ZIP Code): Building 410, Bolling AFB, DC 20332-6448

10. SOURCE OF FUNDING NOS.

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT NO.
61102f	2313	A4	

11. TITLE (Include Security Classification): Perception and the Temporal Properties of Speech

12. PERSONAL AUTHOR(S): Peter C. Gordon

13a. TYPE OF REPORT: Final Report

13b. TIME COVERED: FROM 7/89 TO 7/92

14. DATE OF REPORT (Yr., Mo., Day): January 11, 1993

15. PAGE COUNT:

16. SUPPLEMENTARY NOTATION:

17. COSATI CODES

FIELD	GROUP	SUB. GR.
05	09	

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number): speech perception, prosody, context effects, phonetic segments, fricatives, stress

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

This research examines the interaction of acoustic and lexical information in the identification of words in lexically ambiguous phoneme sequences. In Experiment 1, subjects show priming for the meaning of a large word like "tulips" when presented with a sequence of combinable short words like "two lips." In Experiment 2 priming is found for the meaning of the second short word in similar sequences (e.g. "lips" in "two lips"). Finally, Experiment 3 demonstrates that listeners do not show priming for a short word like "lips" when it is pronounced as part of a larger word like "tulips." The results of these experiments show that listeners sometimes access words other than those intended by speakers, and that they may simultaneously access words associated with several alternative parses of ambiguous sequences. Furthermore, they suggest that acoustic marking of word onsets places constraints on the success of lexical access. To account for these results, we present a new model of lexical access and segmentation, the Good Start model, which gives a principled account of these properties.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT: UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS ☐

21. ABSTRACT SECURITY CLASSIFICATION: UNCLASSIFIED

22a. NAME OF RESPONSIBLE INDIVIDUAL: Dr. John Tagney

22b. TELEPHONE NUMBER (Include Area Code): (202) 767-5021

22c. OFFICE SYMBOL: NL

DD FORM 1473, 83 APR — EDITION OF 1 JAN 73 IS OBSOLETE.

93 3 1 056

SECURITY CLASSIFICATION OF THIS PAGE

15 JAN 1993

Statement of Work

Work on this project will extend previous work on the context-dependent nature of temporal cues to the identity of phonetic segments, and on the role of coarse-grained aspects of the speech signal in facilitating segment recognition. These extensions will address the following questions: Do adjacent segments exhibit mutual dependencies resulting in perceptual ambiguity that can be overcome by contextual information present in coarse-signal characteristics? Can coarse-grained aspects of the speech signal, lacking sufficient information for segment identification, convey speaking rate independently of variation in the inherent durations of the underlying segments? Do coarse-grained aspects of precursive speech contribute contextual information that is used early in the timecourse of segment recognition? Can coarse-grained aspects of the speech signal direct attention to the location of upcoming stressed syllables?

Word on the project will directly study the nature of coarse-grained aspects of the signal and their relation to processing the suprasegmental temporal aspects of speech. New techniques will be developed for creating coarse-grained representations of speech that eliminate information about segment identity but preserve prosodically-relevant aspects of the speech signal. These techniques will permit control over degree of resolution in the short-time spectrum of speech. Perceptual studies, involving direct judgments on stimuli with varying amounts of spectral resolution, will be performed to determine what the amount of spectral detail that is necessary for perceiving important temporal components of prosody.

As part of the project a computer simulation will be developed that will test the computational adequacy of the processes that are hypothesized to underlie human perception of the temporal properties of speech. This model will address three related issues: the segmentation of speech into syllables, the use of temporal relations between syllables to generate expectancies about the temporal properties of upcoming syllables, and the contextual modulation of feature analyzers for processing temporal cues to segment identity.

Status of Research

Publications

Gordon, P.C., Eberhardt, J.L., & Rueckl, J.G. (1992). Attentional modulation of the phonetic significance of acoustic cues. **Cognitive Psychology**.

Gordon, P.C., Schaeffer, C.P., & Kennison, S.M. (1991). Disambiguation of segmental dependencies by extended phonetic context. **Language and Speech, 34(2)**, 157-176.

Yaniv, I., Meyer, D.E., Gordon, P.C., Huff, C.A., & Sevald, C.A. (1990). Vowel similarity and syllable structure in motor programming of speech. **Journal of Memory and Language, 29,** 1-26.

Eberhardt, J.L., & Gordon, P.C. (1990). Effects of attention on the perception of phonologically natural forms. **Journal of the Acoustical Society of America, 88**, Suppl. 1.

Gow, D.W., & Gordon, P.C. (1990). Perceptual and acoustic measures of stress shift. **Journal of the Acoustical Society of America, 88**, Suppl. 1.

Eberhardt, J.L., & Gordon, P.C. (1989). The effects of attention on the phonetic integration of acoustic information. **Journal of the Acoustical Society of America, 86**, Suppl. 1.

Gow, D.W., & Gordon, P.C. (1989). Two paradigms for examining the role of phonological stress in sentence processing. **Journal of the Acoustical Society of America, 86**, Suppl. 1.

Manuscripts under review

Gow, D.W., & Gordon, P.C. Lexical and Prelexical Influences on Word Segmentation: Evidence from Priming.

Gow, D.W., & Gordon, P.C. Syllable stress in the processing and representation of spoken sentences.

Lexical and Prelexical Influences on Word Segmentation: Evidence from Priming

David W. Gow, Jr. Peter C. Gordon

Harvard University

Running Head: Lexical Segmentation...

Send page proofs to:

David W. Gow, Jr.
Department of Psychology
Harvard University
33 Kirkland St.
Cambridge, MA 02138

Abstract

This research examines the interaction of acoustic and lexical information in the identification of words in lexically ambiguous phoneme sequences. The cross-modal lexical priming technique is used to determine which word meanings listeners access at the offsets of such ambiguous sequences when they are presented in connected speech. In Experiment 1, subjects show priming for the meaning of a large word like "tulips" when presented with a sequence of combinable short words like "two lips". In Experiment 2 priming is found for the meaning of the second short word in similar sequences (e.g. "lips" in "two lips"). Finally, Experiment 3 demonstrates that listeners do not show priming for a short word like "lips" when it is pronounced as part of a larger word like "tulips". The results of these experiments show that listeners sometimes access words other than those intended by speakers, and that they may simultaneously access words associated with several alternative parses of ambiguous sequences. Furthermore, they suggest that acoustic marking of word onsets places constraints on the success of lexical access. To account for these results, we present a new model of lexical access and segmentation, the Good Start model which gives a principled account of these properties.

Printed English provides readers with unambiguous markers which segment sentences into their component words. Spoken language is less considerate; words usually abut one another without intervening silences, and are often woven together through coarticulation to form a seamless stream of acoustic-phonetic information. The processs of lexical access therefore includes an element of lexical segmentation. Listeners must either directly or indirectly recognize the boundaries of the words they identify. A number of schemes for isolating words have been proposed. These schemes can be broken down into two groups based on the kinds of information and processes that they emphasize. One group emphasizes the role of prelexical perceptual processes, and the other focuses on the role of lexical or contextual processes.

The prelexical approach stresses the importance of identifying acoustic cues that may mark word boundaries. The basic strategy behind this approach is to locate word boundaries without the aid of hypotheses about the identity of the actual words in a sequence; listeners only initiate lexical access after they have identified a word onset based on its local perceptual features. It follows that listeners access words sequentially when processing connected speech, and that they locate the word boundaries intended by the speaker regardless of the words he or she chooses. Work by Lehiste (1960) on English, and Garding (1967) on Swedish identified a number of acoustic correlates of word boundary including glottal stops, and/or laryngealized voicing at the onset of word-initial vowels, and increased aspiration on voiceless stops. Nakatani and Dukes (1977) manipulated these cues in forced choice listening experiments using stimuli formed by editing together portions of pairs of words like "no notion" and "known ocean". They found that these acoustic cues strongly influenced the way listeners segmented their hybrid word pairs. However, there are no such cues associated with many types of word juncture, and the phonological rules for marking word onsets by these cues tend to be optional in most dialects.

Other work has focused on continuously varying cues such as duration, amplitude and pitch that tend to vary as a function of stress, and which also tend to mark word onsets. Nakatani and Schaffer (1978) used reiterant speech (repeating a syllable, e.g., /ma/, with the prosody of a word or phrase) to produce lexically ambiguous phoneme strings with minimal spectral marking of word boundaries, but normal prosody. They found that these prosodic manipulations influence listeners' lexical segmentation judgments. Subsequent experiments using speech resynthesis to manipulate stress correlates independently showed this was primarily due to duration and rhythm cues; variations in amplitude and pitch did not affect listener segmentation performance. As in the case of spectral cues, there is evidence that duration cues are not a completely reliable. Barry (1981) found that although duration cues are present in citation form tokens, they are far less pronounced in running text.

The work of Cutler and her associates (Cutler and Carter, 1987; Cutler and Norris, 1988; Butterfield and Cutler, 1988; Cutler, 1990; Cutler and Butterfield, 1992) examines the interaction of prelexical and lexical processes in segmentation. They have focused on the distinction between strong syllables (which contain full vowels) and weak syllables (which contain reduced vowels) as a possible prelexical source of information about the location of word boundaries. Cutler and Norris (1988) showed that the patterning of strong and weak syllables influences subjects' ability to detect words embedded in nonwords. Cutler and Butterfield (1992) found that both natural, and laboratory-induced lexical segmentation errors are also affected by this patterning. Cutler and Carter (1987) examined a large database of transcriptions of natural speech to determine how reliably vowel quality marks word onsets. They found that 74% of all full vowels are found in the initial syllables of grammatical words. Their segmentation algorithm, the Metrical Segmentation Strategy (MSS), capitalizes on this distribution by positing separate lexicons for lexical and grammatical words, and initially treating all syllables with full vowels as potential lexical word onsets. The occurrence of a full vowel triggers segmentation according to the MSS, causing listeners to terminate ongoing look-ups and begin new ones starting with the syllable containing the full vowel.

The assertion that full vowels can be detected to trigger segmentation implies that vowel reduction is a discrete quality. However, Stetson (1951) notes that there is a continuous range in vowel quality between schwa and full vowels. Moreover, Fry, Abramson, Eimas and Liberman (1962) have shown that vowels are perceived along a continuum.[1] This suggests that listeners may not normally make clear distinctions between full and reduced vowels. This is reflected in English orthography which makes no distinction between full and reduced vowels.

The lexical component of the MSS is involved in verification and error correction. It states that when listeners encounter a strong syllable they access the longest word consistent with the input beginning with the previous strong syllable. This provides for strictly sequential lexical access as long as each search produces a valid word. When no candidate words are consistent with the input, the MSS calls for backtracking to access shorter words, the continuation of an ongoing search, or the reassignment of the previous syllable to the current input and the initiation of a new search. By combining these lexical and prelexical processes, the MSS provides a means for listeners to sequentially categorize speech sounds into words. Carter and Cutler (1987) tested the MSS on a paragraph-long passage, and found that it lead to the successful recognition of 82% of all words. Briscoe (1989) compared the performance of the MSS to that of segmentation strategies in which lexical access is initiated with every new phoneme or syllable a listener encounters, or at the offset of every word that is recognized, and found that it generated fewer lexical hypotheses than these strategies did across several types of phonetic transcription.

In summary, there is evidence that listeners' segmentation performance is sensitive to a variety of prelexical cues in the speech signal. However, it is unclear whether these cues are reliable enough to explain our apparent facility for accurate lexical segmentation. Spectral cues such as glottal stops, aspiration and laryngealization are generally optional, and only mark certain types of word boundaries, and rhythmic cues are diminished in connected speech. Even relatively consistent forms of marking like the occurrence of strong syllables require an elaborate processing algorithm like the Metrical Segmentation Strategy to account for good, but imperfect, lexical segmentation performance. It appears then, that a more complete understanding of lexical segmentation rests on either the discovery of more reliable acoustic indicators of word boundary location, or the identification of other sources of boundary information.

Lexical information may provide alternative routes to the discovery of word boundaries which avoid any reliance on special acoustic marking of word boundaries. Two segmentation strategies have been proposed which posit no role for acoustic marking of juncture; they are premised on the notion that lexical access leads to lexical segmentation. One approach argues that words are recognized sequentially, and that listeners locate word onsets by recognizing the words (and thus the word offsets) that immediately proceed them. Marslen-Wilson and Welsh (1978) and Cole and Jakimik (1980) have proposed lexical access models that allow listeners to identify words as soon as enough information is available to distinguish them from all other words with the same onset. These models predict that subjects may anticipate a word's offset, and thus the location of the onset of the next word, when this point occurs before its offset. However, Luce (1986) has found that words are only uniquely defined prior to their offsets roughly 40% of the time and so early recognition cannot be considered a reliable strategy. Grosjean (1985) argues that in these cases words are only identified after their offsets. He presented subjects with gated portions of sentences like "I saw the boar in the woods", and found that they were only certain that they had heard the word "boar", and not something like "board" or "born", when they had heard a portion of the next word. Grosjean also used this technique to examine the assumption that lexical access is sequential. He presented subjects with short phrases like "bun in the oven" which included nouns like "bun" which could only be recognized after their offsets, followed by prepositions, and found that subjects were more likely to recognize the noun and the preposition at the same time, or recognize the preposition before the noun, than they were to recognize the words in sequential order. This provides some evidence that the initiation of lexical search is not delayed until the previous lexical search is completed, calling into question the idea that lexical access is strictly sequential. However, Grosjean's (1985) gating procedure is an off-line task that may be subject to strategic effects. Therefore, it cannot tell us definitively if word recognition is strictly sequential in the real time interpretation of connected speech.

Another lexical segmentation strategy that depends on lexical information is to attempt to access every word which is consistent with a pattern in the input. This involves treating each new unit of speech[2] as a possible word onset. Unlike the strategies described above, this approach does not call for sequential, word by word lexical access. It hinges on listeners' ability to carry out simultaneous lexical searches associated with several different readings of segmentation ambiguities. This multiple access strategy can be found in several computer models including Klatt's (1980) SCRIBER, and McClelland and Elman's (1986) TRACE model. Klatt's proposed SCRIBER system identified the words in connected speech by matching acoustic input against sequences of templates representing portions of words and word transitions. SCRIBER simultaneously evaluated all likely sequences, and ultimately selected the best one based on an overall measure of matching between the input and the templates in a sequence.

The TRACE model (McClelland and Elman, 1986) also uses a multiple segmentation approach, and only produces lexical segmentation as a byproduct of lexical access. TRACE is an interactive activation model which can identify words in "mock speech" consisting of idealized representations of acoustic feature values at different points in time. McClelland and Elman (1986) found that TRACE was able to correctly identify, and thus appropriately segment, almost all of the words in its 211 word lexicon when they were presented in word pairs that lacked any overt boundary marking. However, TRACE does not segment accurately when word pairs have the same sequence of phonemes as longer words, showing a strong bias for recognizing long words, but not embeddable short words. This is due TRACE's mechanism for limiting lexical access. TRACE uses competition to identify a unique segmentation solution when word boundaries are ambiguous. For instance, the input representation /parti/ passes activation to the nodes which represent the words "par", "tea" and "party". In turn these lexical nodes pass inhibition to all other lexical nodes. The more activation a lexical node receives, the more inhibition it can pass to its competitors. Activation is a function of the amount of excitatory input a node gets, giving longer words an advantage over shorter words because they receive excitatory input from a larger number of input nodes.

TRACE demonstrates that, in principle, multiple access approaches to lexical segmentation can yield relatively good performance, with some important exceptions, without any reliance on acoustic boundary cues.[3] It is important to remember though that the segmentation strategies employed by TRACE or SCRIBER are motivated by design constraints rather than psychological data. As such, they must be considered largely untested as psychological models. Fortunately, the TRACE model makes strong predictions about segmentation preferences which can be evaluated using human performance data.

Goals and Research Strategy

The goal of the present work is to explore the relationship between prelexical and lexical factors in lexical segmentation. Prelexical models predict that lexical segmentation should lead to the identification of only those words associated with an accurate parse of a speaker's intended utterance. Lexical approaches like TRACE predict that segmentation will tend to reflect the lexical properties of potential words to be accessed rather than the intentions of speakers. Of course, neither approach is likely to be completely correct. It is likely that lexical segmentation processes depend on a combination of prelexical and lexical information.

In order to understand how listeners use these types of information to identify word boundaries, we must examine two fundamental issues about processing. First, we must determine whether listeners initially perceive word boundaries as speakers intend them to. Prelexical segmentation models predict that listener's should generally segment input in agreement with speakers' intentions, while lexical models predict that listener's should show a bias towards perceiving the longest potential word consistent with a phoneme string regardless of acoustic features. Second, we must determine whether lexical access is sequential or parallel. Sequential access is predicted by all of the current prelexical models of segmentation and by some lexical models, while non-sequential multiple access is predicted by models like TRACE and SCRIBER.

This work will employ the cross-modal lexical priming paradigm (CMLP) to examine the interaction of lexical and prelexical influences on lexical segmentation and thus lexical access. This paradigm has been used extensively to study the pattern of lexical access elicited by the presentation of semantically or syntactically ambiguous speech (Swinney, 1979; Onifer and Swinney, 1981; Tanenhaus, Leiman, and Seidenberg, 1979). In the CMLP paradigm, subjects generally hear a word, referred to as the priming stimulus, and then are shown a sequence of letters, the lexical decision probe, which they must classify as a word or nonword. Their task is to indicate by a button press whether the probe is a word or not. In general, subjects are able to respond to the probe more quickly when it is semantically related to the prime. By presenting subjects with primes that are somehow ambiguous and then showing them probe stimuli that are related to different interpretations of the ambiguity, researchers can determine which meanings listeners access at a given point in time. Three experiments will be presented which use a CMLP task to determine what meanings are accessed when subjects are presented with one or two words consisting of lexically ambiguous phoneme strings. Experiment 1 will directly test the prediction that listeners show a bias towards recognizing long words when lexical boundaries are ambiguous at the level of the phoneme string. Experiment 2 will pit any acoustic marking of word boundaries against the hypothesized long-word advantage to see if listeners access small words individually when they can be combined to form long words. Finally, Experiment 3 will explore listeners' tendencies to recognize embedded words.

Experiment 1

In this experiment, subjects heard sentences like "She placed her two lips on his cheek" which contained two word sequences such as "two lips" which could be combined to form a longer, semantically unrelated word, such as "tulips". The second word consisted of a single strong syllable in all cases. Several lexical models of segmentation (McClelland and Elman, 1986; Swinney, 1981) predict that these sequences should produce strong priming for the longer word formed through concatenative combination. Conversely, prelexical models predict that normal speech contains acoustic boundary cues that should lead listeners to correctly segment these sequences into two parts, and avoid accessing the concatenated long word. The Metrical Segmentation Strategy (Cutler and Carter, 1987; Cutler, 1990) predicts that concatenation should be blocked when the second word in a potentially combinable phrase has a full vowel in its first syllable.

Method

Subjects. Twenty undergraduate students participated in a single 45 minute session, and were paid between $5.00 and $5.75, based on their performance on the experimental task. All of the subjects were native speakers of American English with no reported auditory or (uncorrected) visual deficits.

Stimuli. The priming sequences were 48 combinable two-word sequences and 48 single words with identical phonemic transcriptions (e.g., . "two lips" and "tulips"). These priming sequences were embedded in the sixth and seventh syllable positions of sentences which were identical and contextually neutral prior to the primes. The single word primes were all monomorphemic words with first syllable stress. The two-syllable primes consisted of pairs of one syllable words, both words containing full vowels.

The sentences were read aloud by one of the authors (DWG) who read them relatively quickly and fluently in a manner that sounded fluent and spontaneous. The acoustic marking of one and two-word priming sequences was manipulated indirectly via the speakers' intention to pronounce either one or two words. These two types of priming stimuli provided a way of manipulating acoustic boundary marking without varying phonemic sequences. This manipulation allowed for the presence of a full range of both known and potential unknown acoustic segmentation cues. The sentences were read in a sound-attenuating chamber and were recorded on a digital audio recorder at a sampling rate of 32 kHz with a Shure SM59 microphone. The recordings were low pass filtered at 10 kHz, redigitized at 20k and equated for amplitude. The offsets of priming sequences were located using both auditory and visual inspection of digitized waveforms and spectrograms.

The lexical decision probes were single words and pronounceable nonwords. Each priming sequence was paired with a legal word that was related to the meaning of the one-syllable word. Related probe words were selected through a survey in which fifteen college students were asked to write down the first three words that they thought of after reading a single word prime. The most common responses were selected for the related probe condition. For instance, both the sequences "tulips" and "two lips" were paired with the lexical decision probe "FLOWER". Priming sequences were also paired with a lexical decision probe word that was unrelated to any of the words in the priming sequence (ex. "two lips" and "tulips" were both paired with the probe "GRAMMAR"). These unrelated probes served as related probes for other primes. Pronounceable nonword probes were constructed to match the CV pattern and syllable structure of probe words used in the task.

Lexical decision probes were presented in uppercase letters on a CRT placed roughly 30" away from subjects at eye level.

Procedure. Subjects were seated in a sound attenuating chamber facing the CRT. They wore headphones and placed two fingers from their dominant hand over two buttons on a computer mouse. Subjects initiated each trial by pressing either mouse button after being presented with a prompt. When they pressed the button two fixation lines briefly appeared on the screen to draw their attention to the position of the upcoming lexical decision probe. When the fixation lines disappeared, the subjects heard a sentence through the headphones. At some point during the sentence the lexical decision probe appeared on the screen and subjects had to determine if the probe was a word or a nonword. In experimental trials, lexical decision probes were presented immediately at the offset of priming sequences. In filler trials, probe position was varied within the sentence to prevent subjects from changing their processing strategies in anticipation of the probe. Subjects were instructed to press the left button if the probe was a word, and the right button if the probe was not a word. The sentence stopped as soon as a button was pressed. At that point, subjects were to repeat everything that they heard into the microphone. The repetition task was employed to assure that subjects attended to the sentences. After each trial, subjects were given feedback showing how quickly they responded, whether they were correct, and how many points that they earned towards a performance bonus. The number of points earned for correct responses was based on the speed of the response. Points were lost for incorrect responses.

Design. There were four experimental conditions defined by the combination of two prime conditions (one-word versus matching two-word primes) and two probe conditions (related versus unrelated to the one-word prime). An individual subject saw each of the 48 experimental prime sequences in only one of the four conditions, and across subjects each sequence appeared equally in all four conditions. In addition to the experimental trials, each subject completed 72 filler trials in which lexical decision probes were unrelated to primes.

Lexical decision probes were presented at different locations in priming sentences. The trials were broken down into five blocks of 24 sentences. The first block contained only filler trials to allow performance on the task to stabilize. Each of the last four blocks included three trials in each of the four experimental conditions. All blocks contained equal numbers of word and nonword trials in the lexical decision task.

Results

Table 1 shows the mean responses times and error rates in each of the four experimental conditions. Subjects had an overall accuracy rate on experimental trials of 93% with no significant differences in error rates between experimental conditions. There was a significant effect for relatedness, $F_1\ (1,19) = 12.7$, $p < .005$, $F_2(1,47) = 7.4$, $p < .01$, with responses to related lexical decision probes faster than unrelated ones. However, there was no significant difference between trials with single word, and concatenated word primes, $F_1(1,19) < 1$, $F_2(1,47) = 1.3$, $p > .05$. There were no significant interactions, $F_1(1,19) < 1$, $F_2(1,47) < 1$.

/insert table 1 about here/

Discussion

The results of Experiment 1 suggest that listeners access word meanings derived from unintended segmentations of strings of lexically ambiguous phonemes. This result is consistent with lexical accounts of segmentation which predict that listeners access the longest word that is consistent with a phoneme string. This lexical bias has been explained in two ways. Prather and Swinney (1977, cited in Swinney, 1981) offer the sequentialist account that lexical access is guided by a principle of minimal accretion which states that listeners activate the longest word that is consistent with input, and do not attempt to access a new word until input is inconsistent with the ongoing lexical search. For example, when hearing the word "boycott" they have shown that listeners temporarily access "boy" but do not access "cot". The TRACE model

(McClelland and Elman, 1986) offers the non-sequentialist explanation that long words are favored over short words because they receive more bottom-up activation. As they receive more activation they are in turn able to direct more inhibition to competing short words. Both accounts of the long word bias attribute it to lexical factors, and minimize the role of prelexical influences on lexical segmentation.

This result contradicts the Metrical Segmentation Strategy's prediction (Cutler and Carter, 1987; Cutler, 1990) that the full vowel in the second word of two-word sequences should trigger segmentation and prevent listeners from accessing the word formed through concatenation. According to the MSS, listeners should access only the meanings of the two short words. The priming by concatenated word meanings suggests that lexical lookup was not terminated by the full vowel in the second word of the priming sequences. However, the occurrence of syllables with full vowels may still play a role in lexical segmentation. This role may simply be masked by an overwhelming lexical bias towards perceiving long words. Alternatively, it is possible that lexical access is non-sequential, and that lexical access is initiated at all strong syllables.

Experiment 2

Experiment 1 suggested that listeners show a bias towards perceiving long words when phoneme strings are lexically ambiguous. This result almost certainly rests on a lexical mechanism. However, it is unclear whether this mechanism depends on sequential access as suggested by several models (Marslen-Wilson and Welsh, 1978; Cole and Jakimik, 1980; Swinney, 1981; Cutler and Carter, 1987), or nonsequential parallel access as suggested by models like McClelland and Elman's (1986) TRACE model. In order to determine if lexical access is sequential or not, Experiment 2 was performed to determine if the long-word bias prevents the simultaneous access of combinable short words. Subjects were presented with one word priming stimuli in contexts allowing concatenation (e.g., "lips" in "two lips") and in contexts that do not allow concatenation (e.g., "lips" in "warm lips"). If listeners consider several segmentations in parallel, then they should show both priming for long words formed by concatenation, and for the short words that make them up.

Different non-sequentialist models make different predictions about the pattern of priming these stimuli should produce. Such multiple access models attempt to access words associated with both appropriate and inappropriate parses of a speaker's intended utterance, which leads to the consideration of more lexical candidates than are ultimately necessary. Therefore, they depend on mechanisms that limit the number of words that are ultimately accessed. The TRACE model does this before lexical candidates are fully activated. Competition between lexical candidates leads to the early inhibition of words that are not fully activated. In their TRACE simulations, Frauenfelder and Peeters (1990) have demonstrated that the activation of words embedded in the second syllable of longer words actually sinks below its resting level by their offset as a result of the long word bias. Therefore, despite its non-sequentialist access strategy, TRACE predicts that only long words should be accessed when concatenation is possible, which suggests that we should see no priming for combinable short words. One can also envision a multiple access approach in which inappropriate word meanings are eliminated after they are accessed based on contextual constraints. Reddy (1976) and Cole and Jakimik (1980) have offered evidence that semantic and syntactic context can constrain lexical segmentation. Frauenfelder, Segui, and Dijkstra (1990) examined lexical effects in monitoring tasks, and found evidence for lexical facilitation, but not inhibition (competition). Without lexical inhibition, non-sequentialist lexical access would lead to the access of all words with acoustic patterns corresponding to the input.

Method

Subjects. Thirty-two subjects from the same population as the previous experiment were paid between $5.00 and $5.50 based on their performance to serve in a 30 minute experiment.

Stimuli. The priming sequences were 24 two-word sequences in which the combination of the two words formed a longer word (e.g. "two lips"), and 24 two-word sequences in which the two words could not be combined to form a longer word (e.g. "warm lips"). The second word in each combinable pair was the same as the second word in a combinable one. For example, the counterpart to the combinable words "two lips" would be the uncombinable sequence "warm lips". These priming sequences were incorporated in sentences using the same construction and recording techniques as Experiment 1. Some tokens from the first experiment were eliminated, and others were substituted in order to guarantee that the second words in these sequences had strong associates. One third of the combinable stimuli were also used in Experiment 1.

The lexical decision probes for the experimental trials were chosen on the basis of their association with the meaning of the second word in the priming sequence. For example, the sequences "warm lips" and "two lips" were paired with the related probe "KISS" as well as the unrelated word "BOAT".

Procedure and Design. Subjects were tested using the same procedure that was employed in Experiment 1. The design was also similar to that of Experiment 1. There were four experimental conditions formed by the combination of two priming stimulus types (combinable and uncombinable) and two probe types (related and unrelated). Each subject heard each priming word in only one of the four experimental conditions. There were a total of 96 trials including 24 experimental trials and 72 filler trials. These trials were broken down into four blocks of 24 trials with two trials from each of the four experimental conditions in each of the last three blocks. Trial order was different for each subject within each block.

Results

Subjects performed the task with 95 percent accuracy on experimental trials. Responses with reaction times greater than 3000 msec (less than %1 of the data) were eliminated from analysis.[4] Table 2 shows the mean response times for each of the four experimental conditions. There was a main effect for relatedness with decisions for related probes faster than decisions for unrelated probes, $F_1(1,31) = 12.5, p < .001, F_2(1,23) = 6.3, p < .02$. No main effect was observed for the combinability of the priming stimulus, $F_1(1,31) < 1, F_2(1,23) < 1$. Furthermore, there was no significant interaction between relatedness and combinability, $F_1(1,31) < 1, F_2(1,23) < 1$.

/insert table 2 about here/

Discussion

The results of Experiment 2 show that words produce priming when they are in phonemic contexts that allow concatenation, and that the amount of priming they produce is comparable to that found when they are placed in contexts which do not allow concatenation. Considered in isolation, this result suggests that listeners segment lexically ambiguous phoneme strings in a way that is consistent with speakers' intentions despite the long-word bias.

Taken together, the results of Experiments 1 and 2 suggest that listeners simultaneously access words consistent with two different segmentations. Listeners accessed the meanings of the words "tulips" and "lips" (and presumably "two") at the same time when presented with the sequence "two lips". This result carries the important implication that words are not necessarily recognized in strict sequential order as suggested by several models (Marslen-Wilson and Welsh, 1978; Cole and Jakimik, 1980; Swinney, 1981; Cutler and Carter, 1987; Cutler, 1990).

Experiment 3

The purpose of Experiment 3 is to determine the degree to which a speaker's intention, and thus pronunciation, of a phoneme string affects listeners' patterns of lexical access. The results of Experiment 1 are consistent with a long-word perceptual bias which may override prelexical segmentation cues. However, Experiment 2 showed that any such bias, if it exists, does not prevent listeners from also accessing short, potentially combinable words, that a speaker intends to pronounce. These results suggest two possibilities. The first is that listeners access every word which is consistent with a phoneme or syllable sequence in an input without regard to the speaker's intended pronunciation. The other possibility is that listeners are only able to access words which begin with sequences that speakers intend to be word onsets. While the results of Experiment 1 show that listeners may access words that speakers do not intend them to, these words do begin with sequences that the speaker intended to be word onsets. Experiment 3 examines the importance of speakers' intention to pronounce word onsets in lexical access. Specifically, it will determine if the apparent access of embeddable words in Experiment 2 was a function of speaker intention.

In the present study, listeners heard sentences containing two syllable words (e.g., "tulips") with phoneme sequences which are also consistent with two shorter words (e.g., "two lips"), and we looked for evidence of priming by the second embedded word in each pair. These sequences consist of the same phoneme strings that served as priming stimuli in the experimental trials in Experiment 2, but this time they were pronounced with the intention of saying one word rather than two.

Two studies that are similar to the current one have yielded conflicting results. Prather and Swinney (1977, cited in Swinney, 1981) found no evidence of priming by words embedded in the second syllable of longer words. They interpret this result as evidence for a long-word bias. However, Shillcock (1990) reports evidence that embedded words do produce priming, but that its magnitude is modulated by the frequency of word onsets. This suggests that the differences in results between the two experiments may be due in part to lexical factors relating to the phoneme string. By using the same phoneme sequences in the current experiment that were used in Experiment 2, we hope to provide a comparison between the two experiments that will allow us to cancel out lexical effects, and focus on the impact of prelexical factors on subjects' patterns of lexical access.

Methods

Subjects. The subjects were forty undergraduate students drawn from the same population as the subjects in Experiments 1 and 2. They were paid $5.00 to $5.50 based on their performance on the task. Subjects were tested individually in a single 30 minute session.

Stimuli, Procedure and Design. The auditory primes were twenty-four one-syllable words that appeared in each of two conditions. In the embedded condition, phoneme sequences equivalent to these words formed the second syllable of a two-syllable word. In the unembedded condition, priming words were preceded by a one-syllable word. Thus the priming word "lips" is embedded in the carrier word "tulips" and also appears unembedded in the sequence "warm lips". These priming sequences were based on the sequences used in Experiment 2. Each combinable priming sequence in Experiment 2 (e.g., "two lips") consists of the same phoneme sequence as a carrier word in this experiment (e.g., "tulips"). Furthermore, these sequences are preceded by the same words in their sentential contexts in both experiments.

The auditory priming stimuli were presented in conjunction with visually presented lexical decision probes. The lexical decision probes for experimental trials were individual words which were either related, or unrelated, to the embedded or unembedded prime word. The probes used in this experiment were identical to the ones used in Experiment 2. Subjects were tested using the same procedure and design that was employed in the previous experiments.

Results

The results of the experiment are summarized in Table 3. Responses with latencies greater than 3000 msec (less than %1 of the data) were deemed outliers and excluded from analysis as in Experiment 2. There was no main effect for embeddedness, $F_1(1,39) < 1$,

$F_2(1,23) < 1$. Similarly, there was no main effect for the relatedness of lexical probe stimuli to priming stimuli, $F_1(1,39) < 1$, $F_2(1,23) < 1$. However, there was a significant interaction between embeddedness and relatedness, $F_1(1,39) = 6.1$, $p < .025$, $F_2(1,23) = 11.5$, $p = .002$. Unembedded priming stimuli produced a 35 msec priming effect, while embedded priming stimuli produced a 22 msec reversal of the standard priming effect. Focused comparisons show that the priming effect is significant, $F_1(1,39) = 7.0$, $p < .05$, $F_2(1,23) = 8.5$, $p = .01$, while the inhibition effect is not, $F_1(1,39) < 1$, $F_2(1,23) = 3.3$, $p > .05$.

/insert table 3 about here/

Discussion

The results of Experiment 3 suggest that embedded words were not accessed, and that listeners parsed lexically ambiguous phoneme sequences as speakers intend them to. The results of Experiment 2 show that this is not attributable to any lexical bias since the same phoneme sequences produced different patterns of priming when spoken with different segmentations in mind. The contrast between these results implies that speakers mark these sequences differently when they speak. Furthermore, they show that this marking affects the likelihood of successful lexical access.

A comparison of the results of Experiments 2 and 3 suggests that acoustical factors play a larger role lexical segmentation and lexical access than is suggested by Shillcock (1990) or Prather and Swinney (1977, cited in Swinney, 1981). Prather and Swinney's minimal accretion principle predicts that listeners access only the longest potential lexical candidate during the perception of words. Experiment 3 appears to confirm this prediction. However, the results of Experiment 2 suggest that some form o. acoustic marking of word onsets can overcome a potential advantage for long words. Similarly, Shillcock argues that the size of the cohort of lexical candidates defined by a word onset determines the strength of priming for embedded word meanings. Shillcock suggests that listeners attempt to access all embedded words, but that access is blocked in some cases by inhibition from other members of the cohort. Experiment 3 was not designed to examine the effects of cohort size or lexical frequency and so we cannot address his results directly. Our results do show a suggestive trend towards the inhibition of embedded word meanings, but it is not statistically significant. Shillcock also found a trend towards inhibition in some cases. That trend was significant by subject but not by word. Given the equivocal nature of this evidence, we can neither fully reject nor accept the notion that lexical access is limited by interword competition effects. However, the priming of embeddable word meanings found in Experiment 2 suggests that if there is inhibition, it can be overcome by acoustic marking of word onsets.

The results of the current experiment also have bearing on the question of whether lexical access is initiated at all syllable boundaries, or only at boundaries which receive some special marking. Experiment 1 showed that subjects access words other than those intended by speakers, raising the possibility that listeners access the meanings of all words consistent with phoneme or syllable sequences in input. However, Experiment 3 demonstrated that listeners do not access the meanings of words which begin at syllable boundaries, but lack the special acoustic marking of word onsets. This could be taken as evidence that lexical access is only initiated at acoustically marked onsets. Alternatively, it could be argued that lexical access is initiated continuously, but that successful lexical access depends on the marking of word onsets.

General Discussion

The three experiments which we have presented indicate some limitations of existing models of lexical segmentation and lexical access, and provide the impetus for a new model built on the strengths of these earlier ones. The results of these experiments support several conclusions about the nature of lexical segmentation and lexical access. The first is that listeners simultaneously access the meanings of words associated with several parses of lexically ambiguous phoneme sequences. The second is that whether or not a word is identified in such a sequence depends on the acoustic-phonetic features of its onset. The final broad point is that acoustic marking does not play a role in the termination of a lexical search. Furthermore, this matching is not stopped by the identification of another potential word onset. In this section we will discuss the evidence for these three generalizations in the context of a new model, the Good Start model of lexical access and lexical segmentation.

The Good Start model states that lexical access is continuous, in that listeners treat each new unit of input as a potential word onset. However, because speech is a noisy and highly variable medium, listeners are most likely to successfully identify potential words when acoustic marking in the form of long phonemes and syllables, and unreduced vowels facilitates the identification of potential word onsets. Recognizable word onsets get lexical access processes off to a good start by activating potential lexical hypotheses that can supply top-down information to aid in the identification of the segments that follow them. Furthermore, our model posits that lexical search is exhaustive, and non-sequential, so simultaneous searches can access all word meanings consistent with phoneme sequences in the speech signal that begin with the acoustic clarity to allow access to get off to a good start. We will argue that this model builds on the strengths of other models of lexical access and segmentation while avoiding some of their limitations.

Our first claim of the Good Start model is that several lexical searches can occur at the same time. This is a basic assumption of the several models we have already discussed (Klatt, 1980; McClelland and Elman, 1986). This claim is supported by the finding in Experiment 1 that concatenated word meanings produce priming and the finding in Experiment 2 that the same kind of priming stimuli also produce priming for the meaning of the individual words that are concatenated.

The finding that listeners access words associated with different segmentations simultaneously argues against the notion that words are accessed in strict sequence. Sequential access would require listeners to arrive at a single interpretation of ambiguous sequences before they could go on to interpret the next word. This means that upon hearing "two lips" they would have to decide that they heard the word "two" (and not the beginning of the word "tulips" or even "toucans") before they could identify the word "lips". Our finding suggests then, that listeners cannot depend on the strategy that some have suggested (Marslen-Wilson and Welsh, 1978; Cole and Jakimik, 1980) of identifying word boundaries based on the identification of the words that lead up to them.

It is not entirely surprising that we should find evidence for simultaneous multiple access. Parallel search is a central feature of most current psychologically motivated models of lexical access (cf. Marslen-Wilson, 1987; Goldinger, Luce and Pisoni, 1989; McClelland and Elman, 1986; Grosjean and Gee, 1987). It is generally recognized that the temporal constraints of real time speech recognition demand listeners to pursue many lexical hypotheses in parallel. While this degree of functional parallelism may seem ungainly in the context of serial processing, it is a natural outgrowth of the structure of neural and neurally-inspired processing architectures. The suggestion that listeners access words associated with alternative lexical segmentations simultaneously merely adds another level of parallelism to the system.

As a multiple access model, Good Start shifts the processing burden from lexical segmentation to lexical identification. This shift has the important implication of limiting listeners' dependence on unreliable acoustic cues, and providing a potential point of contact for higher level constraints such as syntactic, semantic or pragmatic context to guide performance. Consider the extreme case of homophones which are pairs of words which cannot be distinguished by phonological information. Faced with this semantic distinction without an acoustic difference, it appears that we access all of their meanings, and eliminate the irrelevant meanings later based on contextual constraints (Swinney, 1979; Onifer and Swinney, 1981; Tanenhaus, Leiman, and Seidenberg, 1979). The strategy of multiple access provides a means of addressing the constraints imposed by the temporal and acoustical structure of the speech signal.

The Good Start model's second major claim is that lexical access depends on the physical characteristics of word onsets. We believe that the dominant way in which word "boundaries" are marked acoustically is by making word onsets more intelligible. This facilitates lexical access by giving listeners early access to lexical information which may enhance the perception of the rest of the word. The contrast in results between Experiments 2 and 3 shows that acoustic marking affects subjects' patterns of lexical access. This result is consistent with the wide range of evidence we have reviewed supporting a role for acoustic marking in lexical segmentation. In order to understand this role, we must examine the dynamics of lexical access, and the nature of this acoustic marking.

As we suggested earlier, there are two ways to interpret the role of acoustic marking of word onsets on lexical segmentation and lexical access. One interpretation is that lexical access is only initiated at acoustically marked word onsets. Another interpretation is that lexical access is continuously initiated, but that it is more successful when word onsets receive special acoustic marking. The first approach has the computational value of limiting the number of searches that are initiated, and would appear to simplify computation. However, it also carries the computational cost of requiring listeners to identify potential word boundaries as they hear them, and to backtrack if they make mistakes.

We would like to suggest instead, that lexical access is initiated continuously, but that its success depends in part on acoustic marking which makes word onsets more intelligible. By initiating lexical access continuously, or with the onset of each new unit of input, listeners could avoid the costs of backtracking to correct missed onsets. This strategy would require listeners to consider a larger number of lexical candidates in order to identify the words that make up an utterance. While this may be viewed as a considerable computational cost in the context of models which depend on serial processing (Briscoe, 1989), it may not be costly in the context of a model like TRACE (McClelland and Elman, 1986) which relies on parallel processing. Given the emerging orthodoxy that parallel processing is a central feature of human cognitive function, it is more attractive computationally to initiate lexical access continuously than it is to initiate lexical access only at likely word onsets. Furthermore, if lexical access is initiated continuously, then the effects of acoustic marking cannot be viewed as a mechanism to trigger segmentation. In order to understand the proper role of acoustic marking in lexical access and segmentation we must examine the nature of this marking.

A number of studies have examined the physical properties of word junctures. There are three logical places to look for juncture markers: at word offsets, in junctures themselves, and at word onsets. Indeed, Nakatani and Dukes (1977) have found evidence that a small set of word offsets are marked by allophonic variation of the phonemes /l/ and /r/, but such marking does not appear to exist in general. Junctures themselves are also marked in some circumstances by pauses between words. This is conceptually similar to the juncture marking used in several TRACE model simulations exploring segmentation (McClelland and Elman, 1986; Frauenfelder and Peeters, 1990). Unfortunately, pauses are generally reserved for major grammatical junctures, and are therefore of little use in marking juncture within clauses or phrases. In general, there is little evidence that acoustic marking of word offsets, or junctures themselves plays a primary role in guiding lexical segmentation performance.

It is striking that the best cues for juncture all seem to occur within the first syllable of words. Word onsets are marked by features such as aspiration and laryngealization (Nakatani and Dukes, 1977), lengthening of onset phonemes and syllables (Lehiste, 1972; Oller, 1973; Klatt, 1973; Umeda, 1975; Nakatani and Schaffer, 1978), and the occurrence of full vowels (Cutler and Carter, 1977; Cutler, 1990; McQueen and Cutler, 1992). This kind of marking is counterintuitive. If one where deliberately designing the speech signal, one would surely mark the juncture before word onsets, the way spaces mark juncture in printed test, so that listeners would not have to backtrack to recognize word onsets.

What these kinds of acoustic marking appear to do is make word onsets more intelligible. While researchers have paid little attention to issues directly concerning the intelligibility of word onsets, there is a literature examining the recognizability of stressed versus unstressed words and syllables. This is relevant because duration and vowel reduction are acoustic correlates of stress as well as word onsets. Bond and Garnes (1980) found that stressed syllables are rarely misperceived in fluent speech. Similarly, Kozhevnikov and Chistovich (1965) found that stressed syllables are detected more readily than unstressed ones in noisy environments. Lieberman (1967) also found that stressed words are more recognizable than unstressed words when they are excised from connected speech. Stressed syllables have often been described as "islands of reliability" in a noisy and highly variable speech stream. The clear implication of this work is that these syllables, with their lengthened onsets and full vowels, are more reliably identifiable based on acoustic information alone than are other syllables.[5]

The intelligibility of word onsets is critical in lexical access because word onsets provide the basis for lexical effects. Several studies have shown that lexical effects are larger at word offsets than they are at word onsets (Marslen-Wilson, 1980; Marslen-Wilson and Welsh, 1978; Gow and Gordon, under review). This can be explained in terms of the sequential nature of the speech signal. A clear word onset can define a group of potential lexical candidates. In turn, these candidates provide hypotheses about how the word is continued. These hypotheses are the source of top-down information that can be used to enhance the recognition of less intelligible segments at the ends of words. The cohort model (Marslen-Wilson and Welsh, 1978; Marslen-Wilson and Tyler, 1980; Marslen-Wilson, 1984) depends on the successful recognition of word onsets to define the cohort of lexical candidates that will supply this top-down information. According to the cohort model, lexical access cannot occur unless word onsets are accurately identified. However, Ganong (1980) has shown that under some circumstances lexical effects can affect the perception of word onsets. The TRACE model (McClelland and Elman, 1986) accounts for this effect while still explaining the tendency for lexical effects to be stronger at word offsets than it is at word onsets. In TRACE the strength of lexical activation is a function of the degree to which input is consistent with a particular lexical candidate overall. This means that lexical effects do not depend completely on the identification of word onsets. However, TRACE does produce stronger lexical effects at word offsets than at word onsets due to the sequential nature of the speech signal.

To summarize, we believe that acoustic marking of word onsets should not be considered a segmentation cue. Instead, we believe that the marking of word onsets has the effect of making them easier to perceive. Making onsets easier to perceive gets lexical access off to a *good start* by activating appropriate lexical candidates that can provide top-down support to help listeners perceive the rest of a word.

The Good Start model's third major claim is that acoustic marking does not stop lexical access. This claim comes directly from the results of Experiment 1 which showed that listeners may concatenate two separate words to access the meaning of an illusory larger word. Experiment 2 demonstrated that listeners perceive two words in this situation, and so this is not simply a matter of listeners not recognizing the juncture in these lexically ambiguous phoneme strings.

This claim is a logical outgrowth of the argument that lexical segmentation is the result of lexical access, rather than a prerequisite for it. With the hypothesized good start that acoustic onset marking affords lexical access, top-down processes can help listeners recognize segments at the end of words which reduces the necessity for special marking of word offsets. Furthermore, the GS model suggests that listeners attempt to access all word meanings that are consistent with phoneme strings in the input. As we have already shown, listeners show simultaneous activation of both short words and the longer words they can form through concatenation. We hypothesize that listeners access every potential word formed by the phoneme sequence which follows a good start simultaneously, and then suppresses irrelevant words after the fact based on contextual constraints as in the processing of homophones. In this way, contextual constraint does the work that acoustic marking of word offsets (or indirect marking of acoustic offsets via the marking of following word onsets) has been hypothesized to do in strictly prelexical accounts of lexical segmentation. This view is supported by studies by Cole and Jakimik (1980) and Reddy (1976) which show that lexical segmentation is influenced by semantic context.

In conclusion, we would like to suggest that the Good Start model provides the outline of a psychologically realistic mechanism to account for the patterns of lexical access shown by listeners when presented with lexically ambiguous phoneme strings. It makes the claim that lexical access during connected speech is non-sequential, and that alternative parses of lexically ambiguous phoneme strings are simultaneously accessed. Furthermore, it emphasizes the importance of acoustic marking of word onsets which may give listeners early access to reliable lexical information to enhance the perception of words.

References

Barry, W.J. (1981). Internal juncture and speech communication. In W.J. Barry and K.J. Kohler (Eds.), **Beitrage zur experimenteelen und angewandten phonetetik**. Kiel: AIPUK, 229-289.

Bond, Z., and Garnes, S. (1980). Misperceptions of fluent speech. In R. Cole (Ed.), **Perception and production of fluent speech**. Hillsdale, NJ; Lawrence Erlbaum, 115-132.

Briscoe, E.J. (1989). Lexical access in connected speech recognition. **Proceedings of the Twenty-Seventh Congress, Association for Computational Linguistics**, Vancouver.

Butterfield, S., and Cutler, A. (1988). Segmentation errors by human listeners: Evidence for a prosodic segmentation strategy. **Proceedings of SPEECH '88, Seventh Symposium of the Federation of Acoustic Societies of Europe, Edinburgh; Vol. 3**, 827-833.

Cole, R., & Jakimik, J. (1980). A model of speech perception. In R. Cole (Ed.), **Perception and production of fluent speech**. Hillsdale, NJ: Lawrence Erlbaum, 133-163.

Cutler, A. (1990). Exploiting prosodic probabilities in speech segmentation. In G.T.M. Altmann (Ed.), **Cognitive models of speech processing: Psycholinguistic and computational perspectives**. Cambridge, MA: MIT Press, 105-121.

Cutler, A. and Butterfield, S. (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception. **Journal of Memory and Language**, 31, 218-236.

Cutler, A., and Carter, D.M. (1987). The predominance of strong initial syllables in the English vocabulary. **Computer Speech and Language**, 2, 133-142.

Cutler, A., and Fear, B. (1991). Categoricality in acceptability judgements for strong versus weak vowels. **Proceedings of the ESCA Workshop in phonetics and phonology of speaking styles**, Barcelona.

Cutler, A. and Norris, D. (1988). The role of strong syllables in segmentation for lexical access. **Journal of Experiment Psychology: Human Perception and Performance, 14**, 113-121.

Frauenfelder, U.H., and Peeters, G. (1990). Lexical segmentation in TRACE: An exercise in simulation. In G.T.M. Altmann (Ed.), **Cognitive models of speech processing: Psycholinguistic and computational perspectives**. Cambridge, MA: MIT Press, 50-86.

Frauenfelder, U.H., Segui, J., and Dijkstra, T. (1990). Lexical effects in phonemic processing: Facilatory or inhibitory? **Journal of Experiment Psychology: Human Perception and Performance, 16(1)**, 77-91.

Fry, D.B., Abramson, A.S., Eimas, P.D., and Liberman, A.M. (1962). The identification and discrimination of synthetic vowels. **Language and Speech, 5**, 171-189.

Ganong, W.F. (1980). Phonetic categorization in auditory word perception. **Journal of Experimental Psychology: Human Perception and Performance**, 6, 110-115.

Garding, E. (1967). **Internal juncture in Swedish.** C.W.K. Gleerup: Lund.

Goldinger, S.D., Luce, P.A., and Pisoni, D.B. (1989). Priming lexical neighbors of spoken words: Effects of competition and inhibition. **Journal of Memory and Language, 28**, 501-518.

Gow, D.W., and Gordon, P.C. (Under review). Coming to terms with stress: Effects of stress location in sentence processing.

Grosjean, F. (1985). The recognition of words after their acoustic offset: Evidence and implications. **Perception and Psychophysics, 38, 299-310.**

Grosjean, F., and Gee, J.P. (1987). Prosodic structure and spoken word recognition. **Cognition, 25**, 135-155.

Klatt, D.H. (1973). Interaction between two factors that influence vowel duration. **Journal of the Acoustical Society of America, 54**, 1102-1104.

Klatt, D.H. (1980). Speech perception: A model of acoustic-phonetic analysis and lexical access. In R.A. Cole (Ed.), **Perception and production of fluent speech**. Hillsdale, NJ: Lawrence Erlbaum, 243-288.

Kozhevnikov, V., and Chistovich, L. (1965). **Speech: Articulation and perception**. US Department of Commerce Translation. IPRS 30, 543. Washington, D.C.

Lehiste, I. (1960). An acoustic-phonetic study of internal open juncture. **Phonetica Suppl. 5**.

Lehiste, I. (1972). The timing of utterances and linguistic boundaries. **Journal of the Acoustical Society of America, 51**, 2018-2024.

Lieberman, P. (1967). **Intonation, perception, and language. Research monograph no 38**. Cambridge, MA: MIT Press.

Luce, P.A. (1986). A computational analysis of uniqueness points in auditory word recognition. **Perception and Psychophysics**, 39(3), 155-158.

Marslen-Wilson, W.D. (1980). Speech understanding as a psychological process. In J.C. Simon (Ed.), **Spoken language generation and understanding**. Dordrecht: Reidel.

Marslen-Wilson, W.D. (1984). Function and process in spoken word-recognition. In H. Bouma and D. Bouwhuis (Eds.), **Attention and performance X**. Hillsdale, NJ; Lawrence Erlbaum.

Marslen-Wilson, W.D., and Tyler, L.K. (1980). The temporal structure of spoken language understanding. **Cognition, 8**, 1-71.

Marslen-Wilson, W.D., and Welsh, A. (1978). Processing interactions and lexical interaction during word recognition in continuous speech. **Cognitive Psychology**, **10**, 29-63.

McClelland, J., and Elman, J. (1986). The TRACE model of speech perception. **Cognitive Psychology, 18, 1-86.**

McQueen, J.M., and Cutler, A. (1992). Words within words: Lexical statistics and lexical access. Proceedings of the ICSLP, vol. 1, Banff, Alberta, 221-224..

Mehler, J., Dommergues, J.Y., Frauenfelder, U., and Segui, J. (1981). The syllable's role in speech segmentation. **Journal of Verbal Learning and Verbal Behavior**, **20**, 298-305.

Nakatani, L.H., and Dukes, K.D. (1977). Locus of segmental cues for word juncture. **Journal of the Acoustical Society of America**, **62**, 714-719.

Nakatani, L.H., and Schaffer, J.A. (1977). Hearing 'words' without words: Prosodic cues for word perception. **Journal of the Acoustical Society of America**, **63(1)**, 234-245.

Oller, D.K. (1973). The effect of position in utterance on speech segment duration in English. **Journal of the Acoustical Society of America, 54**, 1235-1247.

Onifer, W., and Swinney, D.A. (1981). Accessing lexical ambiguities during sentence comprehension: Effects of frequency of meaning and contextual bias. **Memory and Cognition, 9**, 225-236.

Pisoni, D.B., and Luce, D.B. (1987). Acoustic-phonetic representations in word recognition. **Cognition, 25**, 21-52.

Prather, P., and Swinney, D.A. (1977). Some effects of syntactic context upon lexical access. Presented at a meeting of the American Psychological Association, San Francisco, August 26, 1977.

Reddy, R. (1976). Speech recognition by machine: A review. **Preceedings of the IEEE. 64, 501-531,**

Shillcock, R. (1990). Lexical hypotheses in continuous speech. In G.T.M. Altmann (Ed.), **Cognitive models of speech processing: Psycholinguistic and computational perspectives**. Cambridge, MA: MIT Press, 24-49.

Stetson, R.H. (1951). **Motor phonetics**. Amsterdam: North-Holland.

Swinney, D.A. (1979). Lexical access during sentence comprehension: (Re)consideration of context effects. **Journal of Verbal Learning and Verbal Behavior, 18, 645-660.**

Swinney, D.A. (1981). Lexical processing during sentence comprehension: Effects of higher order constraints and implications for representation. In T. Meyers, J. Laver, and J. Anderson (Eds.), **The cognitive representation of speech**. North-Holland.

Tanenhaus, M., Leiman, J., and Seidenberg, M. (1979). Evidence for multiple stages in the processing of ambiguous words in syntactic contexts. **Journal of Verbal Learning and Verbal Behavior, 18**, 427-441.

Umeda, N. (1975). Vowel duration in American English. **Journal of the Acoustical Society of America, 58**, 434-445.

Author Notes

The research reported in this paper was supported by grant 80-0416 to Harvard University (Peter C. Gordon, Principle investigator) from the Air Force Office of Scientific Life Researches Directorate. We thank Julia Lee and Matei Mihalca for their assistance in testing subjects.

Direct correspondence to either author at the Department of Psychology, Harvard University, 33 Kirkland St., Cambridge, MA 02138.

Table 1

Mean reaction times (in msec) for single word and concatenated word prime conditions in Experiment 1. Accuracy rates are shown in parentheses.

Prime Condition	Related Probe (FLOWER)	Unrelated Probe (GRAMMAR)
Single Word (*tulips*)	798 (.95)	837 (.93)
Concatenated Word (*two lips*)	812 (.93)	839 (.93)

Table 2

Mean reaction times (in msec) for combinable and uncombinable prime conditions in Experiment 2. Accuracy rates are shown in parentheses.

Prime Condition	Related Probes (KISS)	Unrelated Probes (BOAT)
Combinable (*two lips*)	765 (.94)	795 (.95)
Uncombinable (*warm lips*)	782 (.96)	801 (.93)

Table 3

Mean reaction times (in msec) for embedded, and unembedded word prime conditions in Experiment 3. Accuracy rates are shown in parentheses.

Prime Condition	Related Probe (KISS)	Unrelated Probe (BOAT)
Embedded (*tulips*)	772 (.90)	751 (.90)
Unembedded (*warm lips*)	760 (.92)	796 (.94)

Footnotes

1. Cutler and Fear (1991) argue that listeners make a categorical distinction between full and reduced vowels. They found that listeners' judgements of the naturalness of word tokens produced through a cross-splicing technique were influenced by changes in vowel quality but not in stress. This demonstrates the relative importance of vowel quality. However, it does not demonstrate that vowel quality is perceived categorically since the only vowel quality distinction made was between full and reduced vowels.

2. There is a large empirical literature dedicated to the question of what sublexical units, if any, are identified. The proposed units include the phoneme (Pisoni and Luce, 1987), diphone (Klatt, 1980), and syllable (Mehler, Dommergues, Frauenfelder and Segui, 1981). The basic segmentation strategy that we are discussing could be instantiated using any of these units.

3. McClelland and Elman emphasize TRACE's ability to perform relatively accurate segmentation using a highly simplified input representation lacking any representation of potential pre-lexical boundary cues, or supralexical contextual constraints. However, they suggest that TRACE's performance would be improved by the inclusion of these factors. Both McClelland and Elman (1986) and Frauenfelder and Peeters (1990) have performed TRACE simulations using markers in input representations to guide segmentation. The use of these markers does allow TRACE to overcome the long word bias and avoid concatenation errors. It should be noted though that these markers guide segmentation by changing the alignment of the input. Lexical access in TRACE depends on a strict alignment of input over time. Each lexical node is activated by input from nodes representing phonemes in a particular temporal alignment. For instance, a node representing the word "cargo" would receive input from nodes representing the phonemes /k/, /a/, /r/, /g/, and /o/ in sequence. The insertion of a

boundary marker between the /r/ and the /g/ would shift the second syllable out of alignment, and prevent the nodes representing /g/ and /o/ from sending activation to the node representing "cargo". This would reduce the amount of activation that "cargo" received, and thus reduce the amount of inhibition it could direct towards the node representing the word "go". While this mechanism leads to accurate segmentation in the TRACE model, it is probably an unrealistic way of explaining human segmentation performance. Unlike TRACE, human listeners show flexibility in their ability to recognize words given variability in their temporal alignment. Furthermore, this kind of marking is the equivalent of placing pauses between words, and real speakers generally do not do this. It appears then that TRACE provides no realistic mechanism for utilizing acoustic information in lexical segmentation.

4. The same trimming criteria were used for all three experiments. Incorrect responses, and responses with reaction times greater than 3000 msec were excluded from all analyses. No trials were excluded from the analysis of Experiment 1 on the basis of reaction time.

5. Certain optional acoustic boundary markers like laryngealization and glottalization do not appear to make word onsets more intelligible. However, they may facilitate syllabification, which would be a vital prelexical process if syllables are identified prior to words.