Lexical, Syntactic, and Stress-Pattern Cues for Speech Segmentation

Lisa D. Sanders
Helen J. Neville
University of Oregon
Eugene

Many sources of segmentation information are available in speech. Previous research has shown that one or another segmentation cue is used by listeners under certain circumstances. However, it has also been shown that none of the cues are absolutely reliable. Therefore, it is likely that people use a combination of segmentation cues when listening to normal speech. This study addresses the issue of how young adults use multiple segmentation cues (lexical, syntactic, and stress-pattern) in combination to break up continuous speech. Evidence that people use more than one cue at a time was found. Furthermore, the results suggest that people can use segmentation cues flexibly such that remaining cues are relied upon more heavily when other information is missing.

KEY WORDS: auditory, language, segmentation, phoneme detection

Speech comprehension requires breaking continuous streams of sounds into units that can be recognized. Most listeners solve the problem of dividing long streams of phonemes into linguistically meaningful units effortlessly, but it is unclear how this is done. The lack of knowledge about how listeners segment continuous speech is evident in the difficulties of creating automatic speech recognition software that parses speech as humans do (Bernstein & Franco, 1996; Brent, 1999).

Lexical, syntactic, and acoustic information are all available in speech and may be helpful in segmentation. However, all of the cues that have been shown to be possible sources of segmentation information are misleading under some circumstances. For example, word recognition itself provides segmentation information. Successful recognition of one word in a speech stream, which can sometimes be achieved even before the word has ended (Marslen-Wilson & Welsh, 1978) would allow a listener to predict both the rest of the word and the subsequent word boundary. It has been suggested, that along with contextual information, lexical recognition may play a primary role in segmentation (Quene, 1992). However, there are circumstances under which lexical recognition would fail as a segmentation cue, such as when short words that cannot be recognized until after their acoustic offset or words that are embedded in other words are encountered (Frauenfelder, 1985; Luce, 1986). These conditions occur frequently enough in normal speech that relying on lexical information alone would lead to inaccurate speech segmentation.

Little research has been conducted on the role of syntactic information in speech segmentation, but knowledge about phrase structure and parts of speech could be useful (Cole, Jakimik, & Cooper, 1980; Tyler &

Wessels, 1983). For example, knowing that a speaker is using an adverb would allow a listener to consider an "ly" ending as part of the current word rather than the beginning of the next. However, syntactic structure is often not obvious until well after the acoustic offsets of words and would not always provide useful segmentation information even when it was.

Many types of acoustic information have been shown to play a role in speech segmentation. For example, people can use phonotactic constraints to parse speech between phonemes that never occur in combination within a word but do occur in combination across word boundaries (Brent, 1997; Brent & Cartwright, 1996). Allophonic variation, variability in the way phonemes are pronounced, can serve as a segmentation cue when it correlates with the position of phonemes in words (Church, 1987; Umeda & Coker, 1974). In languages with clear syllable boundaries such as French, listeners have been shown to segment speech between each syllable (Cutler, Mehler, Norris, & Segui, 1986; Mehler, Dommergues, Frauenfelder, & Segui, 1981). Japanese speakers can use morae, a unit that sometimes but not always corresponds to a syllable, to segment speech (Cutler & Otake, 1994; Otake, Hatano, Cutler, & Mehler, 1993). From a corpus of spontaneous British speech, Cutler and Carter (1987) found that 90% of the words began with strong stress. After factoring in frequency, they determined that 75% of the strong stresses encountered in English speech are word initial. Native English speakers have been shown to take advantage of this typical stress pattern to segment speech (Cutler & Butterfield, 1992; Cutler & Norris, 1988; McQueen, Norris, & Cutler, 1994; Norris, McQueen, & Cutler, 1995). However, each of these cues is either sometimes misleading (stress pattern), only occasionally present (phonotactic and allophonic cues), or would result in the segmentation of speech into many more units than is necessary (syllable and mora).

Another possible segmentation cue is the transitional probabilities of syllables within and between words. Harris (1955) suggested that the probability of hearing any two syllables in succession is higher if the syllables both occur within a word (for example, "baby") than if the syllables occur as the last sound in one word and the first sound in the next (for example "-by doll"). Others have shown that adults and infants can use this type of statistical information to segment speech if they are given artificial languages made of nonsense words or environmental sounds in which the transitional probabilities of syllables within words are very high and the transitional probabilities between words is very low (Cowan, 1991; Hayes & Clark, 1970; Saffran, Aslin, & Newport, 1996; Saffran, Newport, & Aslin, 1996). However, these probabilities are not known for any non-artificial languages, so it is not clear that the information is present or useful under normal circumstances. Furthermore, word boundaries would not be marked by a dip in transitional probabilities when the last syllable of one word and the first syllable of the next could be contained within a single word (for example "bay be" and "baby"). Therefore, transitional probabilities will also be a misleading or an absent segmentation cue in some examples of speech.

Clearly, there are many sources of information people can use to segment speech, although none of them seem to be absolutely reliable. Therefore, it is likely that people use multiple cues rather than limiting themselves to a single, imperfect one. However, most segmentation studies have focused on the use of only one cue. Therefore, little is known about how the use of different sources of segmentation information may interact to produce the very accurate speech segmentation people demonstrate every day. Can people flexibly use multiple sources of information such that available cues become more important when others are absent? Is there a single cue that is relied upon most heavily such that other cues are only used when it is misleading or absent? These questions can only be addressed by measuring the effects of different segmentation cues within the same experiment.

Those studies that have considered multiple sources of segmentation information (including Norris, McQueen, & Cutler, 1995; Quene, 1992; Vroomen, van Zon, & de Gelder, 1996) have used degraded stimuli, such as one- or two-word utterances that encourage the use of some cues (usually acoustic) and discourage the use of other cues (usually contextual, lexical, and syntactic). To get an accurate measure of the relative use of different cues in normal speech processing, it is necessary to use examples of continuous speech that contain all of the usual semantic, syntactic, and acoustic information encountered in speech. The purpose of this study was to assess the use of multiple segmentation cues using full sentences presented as continuous speech. Although it was not possible to measure the effects of all the cues described here, two types of linguistic cues, lexical and syntactic, and one acoustic cue, stress pattern, were manipulated.

Another issue that must be addressed in studying speech segmentation is the type of task used. It is desirable to use a task that relates to where speech streams are segmented in a transparent way. For example, by asking people where they perceive word onsets or the lack of word onsets. The first experiment used this approach by asking subjects to detect target phonemes in sentences and to report whether the targets were word initial or word medial. This type of task has the additional advantage that all potential segmentation cues

may be made available in the stimuli by using natural speech. However, it is also a difficult task in which decisions are made well after a target is presented. This makes it impossible to determine if the information used to determine if sounds were word medial or word initial, influenced online segmentation, played a role in reparsing any missegmented sounds, affected only the metalinguistic decision, or some combination of these.

Therefore, it is also desirable to use segmentation tasks that can be performed quickly enough to be certain they are directly tapping into online segmentation. Tasks such as phoneme monitoring and syllable monitoring result in faster reaction times than phoneme localization, suggesting that these decisions can be made at an earlier point in the segmentation process than localization decisions. Converging evidence from this type of experiment would add support to the idea that information used to localize phonemes was involved in segmentation as opposed to metalinguistic decisions. However, this type of task requires the additional assumption, for which there is little evidence, that units corresponding to a segment or falling at the beginning of a segment will be detected more quickly. Furthermore, they require subjects to pay greater attention to the surface features of speech than may be typical. Experiment II of this study employs a phoneme monitoring task using the same stimuli as Experiment I so the results of the two can be compared.

Other tasks such as gating and shadowing can be used to ensure that subjects have no additional information after the point of interest and require no metalinguistic knowledge. However, they cannot be conducted without disrupting the presentation of normal continuous speech.

In an attempt to understand how continuous speech is segmented, it is necessary to use converging evidence from a variety of techniques. Because the focus of this study was to explore how multiple segmentation cues are used in normal speech processing, it was important to be certain that all possible segmentation cues were available in the stimuli. Therefore, a phoneme localization and phoneme detection task were used.

Experiment I: Phoneme Localization

Method

Participants

Sixteen monolingual English speakers participated in Experiment I (*M* age = 20.9 years, 11 women). All were right-handed university students who were paid $7 per hour.

Stimuli

A total of 900 sentences were created to vary the amount of lexical, syntactic, and useful stress-pattern information available to the listener. Lexical and syntactic information was varied by replacing all of the content words, or all of the words, in a sentence with pronounceable nonwords. Stress pattern information was varied by using words that contained targets in different positions and had strong stress on different syllables.

Thirteen single phonemes and eight phoneme combinations where chosen as targets. All of the targets were consonant sounds that occur in both word-initial and word-medial positions in English.

The following five types of words were selected for use in the experiment: (a) began with a target and had strong stress on the first syllable, (b) began with a target and had weak stress on the first syllable, (c) contained a word-medial target and had strong stress on the syllable in which the target occurred, (d) contained a word-medial target and had weak stress on the syllable in which the target occurred, and (e) did not contain a target. Examples of each type of word can be seen in Table 1, and a full list of all target-containing words can be found in Appendix A.

All of the words were two, three, or four syllables long. Word-medial targets were chosen to be the first or first two phonemes of the second syllable of a word. Syllabification in English is not always clear-cut. For example, the /v/ in “gavels” could be considered the last sound in the first syllable (gav-els) or the first sound in the second syllable (ga-vels). Both the sonority sequencing principle and the principle of maximizing onsets would predict that the targets chosen for this experiment would in fact be considered the beginning of the second syllables (e.g., ga-vels) with the exception of /s/ and /st/. An item analysis was used to determine whether or not these (or any other small group of stimuli) were producing the results reported.

Sixty words from each of the five categories (for a total of 300 words) were chosen to be matched on the targets they contained, part of speech, written and spoken word frequency, and word length. Every target phoneme

Table 1. Examples of words that contain or do not contain target phonemes.

Condition	Target sound	Word
Target present		
Strong stress, Initial position	/b/	bottles
Strong stress, Medial position	/b/	tobacco
Weak stress, Initial position	/b/	balloon
Weak stress, Medial position	/b/	timber
Target absent	/b/	afghan

and phoneme combination was equally represented in the five groups. Words that have an infrequent English stress pattern (weak stress on the first syllable, strong stress on the second) tended to be of low frequency (M = 21.38, range = 0 to 267), (Kucera & Francis, 1967), so low frequency words with typical English stress pattern (strong stress on the first syllable, weak on the second) were used as well (M = 20.73, range = 0 to 290). This selection resulted in no significant differences in written or spoken frequencies (M = 2.23, range = 0 to 55) (Brown, 1984) across the word types. Furthermore, there were no reliable differences in the number of letters in words which contained target phonemes or their non-target matches across conditions (M = 7.72, range = 4 to 11).

Sentences were then constructed around these selected words such that a target never occurred anywhere in a sentence other than in the selected word. It was important that the words that might contain targets could not be predicted before they were actually heard. Therefore, the cloze probability of each of the selected words in its sentence was measured by giving 40 naive subjects all of the words in the sentences up to the critical ones and asking them to write down the word they thought would come next. If more than 25% of the subjects chose the same word to continue a sentence (even if the word they chose was not the one to be used in the experiment), the sentence was excluded. This procedure resulted in a low cloze probability for the selected words in the sentences that were used (M = 0.032, range = 0.000 to 0.215). Additionally, target phonemes never occurred in the first three or the last three words of a sentence (M = 9th word), and there were no significant differences in target position in the sentences across word type. (See Appendix B for means and ranges describing each of these aspects of the words and sentences).

The 300 normal English sentences (semantic sentences) that were created contained extensive semantic, syntactic, and prosodic information. To make sentences that maintained syntactic and prosodic information, but had less meaning (syntactic sentences), all of the content words in these sentences were replaced with pronounceable nonwords. Morphemes such as "ed," "ing," and "ly" were kept when replacing words which contained these units with nonwords. To ensure that the resulting sentences were pronounceable, all nonwords were created by replacing every consonant with one from the same class (stop, fricative, or nasals and liquids) and every vowel with another vowel randomly. The only exceptions to this were (a) when the resulting group of sounds created another English word the procedure was repeated until a nonword was made, (b) the target phoneme or phoneme combination for a sentence was excluded from the group of sounds used in the replacement, and (c) the target phoneme or combination was not changed.

To create a group of sentences that contained normal English prosody but that had less grammatical structure than the syntactic sentences (acoustic sentences), the remaining words and morphemes were replaced in the same manner described above. A set of five sentences in each of the three sentence forms is shown in Table 2.

Table 2. Examples of semantic, syntactic, and acoustic sentences for all conditions.

Condition	Type	Sentence
Target present		
SI	Semantic	In order to recycle *b*ottles you have to separate them.
	Syntactic	In order to lefatal *b*okkers you have to thagamate them.
	Acoustic	Ah ilgen di lefatal *b*okkerth ha maz di thagamate fon.
SM	Semantic	If the only thing in it were to*b*acco it wouldn't cause so much harm.
	Syntactic	If the ilmy shord in it were do*b*atty it wouldn't gaff so much hilm.
	Acoustic	Os fa ilmy shord el ok hon do*b*atty ag hapsel gaff sha nes hilm.
WI	Semantic	The child stopped crying when a *b*alloon was given to her.
	Syntactic	The ferp trepped plawing when a *b*arreal was kaffen to her.
	Acoustic	Sa ferp trepp plawel ron i *b*arreal hof kaffem gi wem.
WM	Semantic	I saved money since lowgrade tim*b*er worked for this project.
	Syntactic	I cheft rono since miltrok del*b*er meld for this plassig.
	Acoustic	O cheft rono zalf miltrok del*b*er meld sith foch plassig.
Target absent	Semantic	Try looking under the afghan for the toy you lost.
	Syntactic	Qui medding under the ithdon for the kay you moft.
	Acoustic	Qui medden amkel fa ithdon sal cha kay wa moft.

Note. All example sentences use /b/ as the target phoneme, which is indicated by italics in the sentences. SI = strong stress, initial position; SM = strong stress, medial position; WI = weak stress, initial position; WM = weak stress, medial position.

Each sentence was digitized (22 kHz sampling rate, 16 bit) using Goldwave software on a Pentium PC by a female native English speaker at a normal speaking rate (M = 4.26 words per second). The speaker was aware of the purpose of this study and knew to record the syntactic and acoustic sentences with the same prosody as the semantic sentences they were created from. Silence at the beginning and end of all sentences was removed from the sound files, and the highest amplitude of each sound was normalized to 1. The auditory versions of the semantic, syntactic, and acoustic form of each sentence were close in total length (M difference = 49.42 ms, range = 0 to 288 ms). This confirmed that the three forms of each sentence were successfully recorded at similar speech rates. Furthermore, there were no overall differences in sentence length (M = 3,438 ms) or position of the target (M = 1,616 ms) among conditions. (See Appendix C for means and ranges of these measurements.)

Two measures of stress, length of the target phoneme with its following vowel and maximum amplitude over that range, were used to ensure that stress was consistent across the different types of sentences. Amplitude was measured on a relative scale from zero to one. The targets in the semantic sentences had an average maximum amplitude of 0.56 [range = .12 to 1] and an average length of 137 ms [range = 22 to 263 ms]. The syntactic (maximum amplitude = 0.57 [range = 0.11 to 1], length = 135 ms [range = 38 to 267 ms]) and acoustic (maximum amplitude = 0.55 [range = 0.10 to 1], length = 133 ms [range = 46 to 252 ms]) versions of the sentences did not differ in maximum amplitude or length of the targets. There were some differences in stress across the different target conditions. Target phonemes in the middle of words tended to be slightly louder [$F(1, 708) = 22.75$, $p < .01$] and slightly longer [$F(1, 708) = 53.46$, $p < .01$] than those at the beginnings of words. However, these difference were fairly small in magnitude (difference in mean maximum amplitude = 0.06, difference in mean length = 23 ms). As expected, targets that received strong stress were louder [$F(1, 708) = 722.78$, $p < .01$] and longer [$F(1, 708) = 450.86$, $p < .01$] than targets that were weakly stressed (difference in mean maximum amplitude = 0.40, difference in mean length = 66 ms).

Furthermore, pitch contours were examined across entire sentences and fundamental frequency changes over the syllables in which targets occurred were measured. For each set of three versions (semantic, syntactic, and acoustic) of the same sentence the pitch contours were judged to be equivalent. An example of the spectrograph and pitch analysis of a set of sentences can be seen in Figure 1. The three versions of each sentence had similar fundamental frequencies measured at the target onset (M difference = 27 Hz, range = 1 to 91 Hz) and at the end of the syllables in which a target occurred (M difference = 34 Hz, range = 1 to 103 Hz). In all of the sentences the frequency change between the onset of the target and offset of the syllable in which the target occurred was in the same direction for the three versions. There was no effect of sentence type or interaction of sentence type with stress or position for either of the frequency measurements or the difference between the two. Neither stress nor position had a significant effect on the first frequency measurement. However, the stress affected change in frequency [$F(1, 708) = 814.31$, $p < .01$] such that fundamental frequency increased over strongly stressed syllables (M change = +60 Hz) but tended to stay the same over weakly stressed syllables (M change = +4 Hz). Furthermore, stress and position interacted [$F(1, 708) = 114.1$, $p < .01$] such that frequency tended to increase for the weak-initial syllables (M change = +29 Hz) and decrease for the weak-medial syllables (M change = –21 Hz). Means and ranges of the changes in frequency, amplitude, and length of the syllables in which targets occurred can be found in Appendix D.

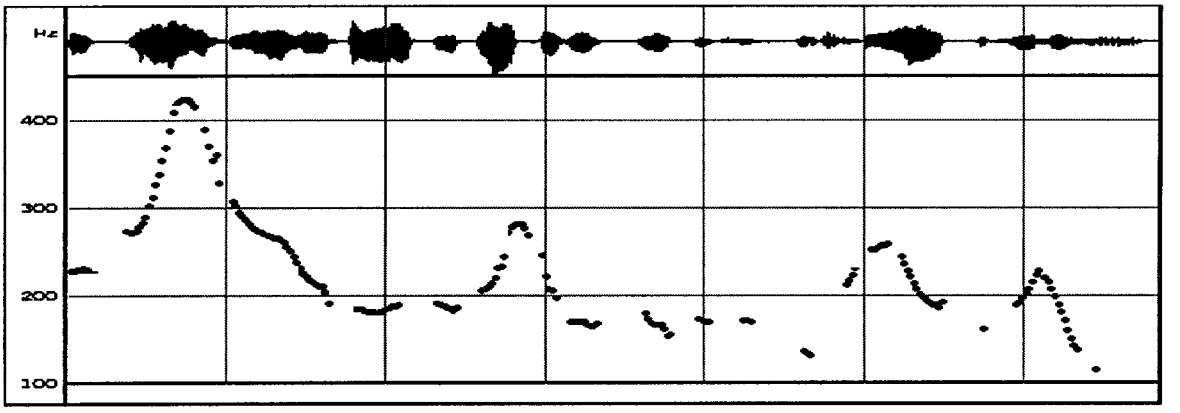
An additional test of the acoustic similarity of the semantic, syntactic, and acoustic sentences was performed by giving a group of people who did not know English a task which involved all three types of sentences. Native Spanish speakers (N = 9) who had little or no exposure to English were asked to detect target phonemes or phoneme combinations in the sentences. There was no effect of sentence type on detection accuracy [$F(2, 16) = 0.58$, $p = .571$] or on reaction times [$F(2, 16) = 0.01$, $p = .993$]. These results suggest that any differences in performance with the different types of sentences found for other groups must be dependent on experience with English, not acoustic differences in the sentences which could be detected by anyone with auditory language experience.

Examples of the target sounds pronounced in isolation were used, in part, to indicate which sound subjects were to listen for in each sentence. Each target followed by /ə/ (e.g., /b/ became /bə/, /fr/ became /frə/) was pronounced by the same speaker who recorded the sentences and was digitized in the same way as the sentences.

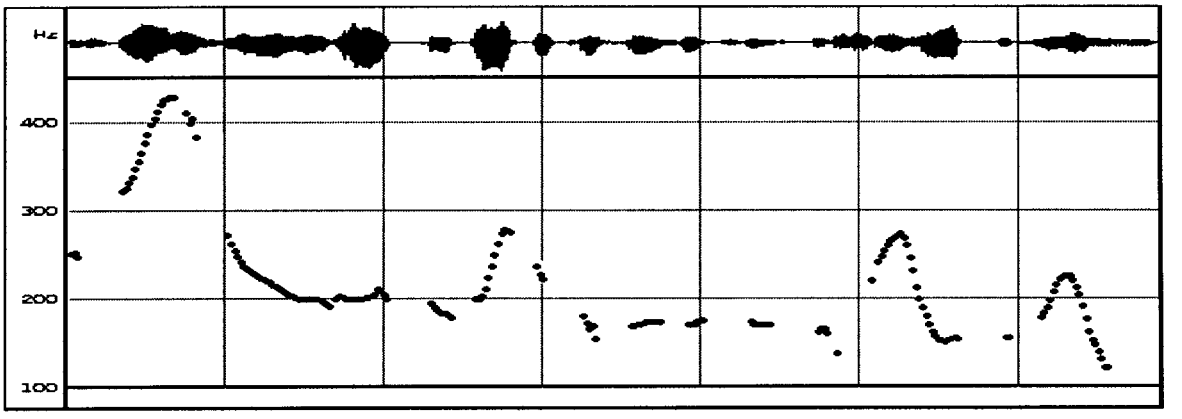
Procedure

Participants were brought in for two 1.5 hour sessions that were at least 3 days, but not more than 2 weeks apart. During each session they completed 60 practice trials and 450 test trials. They sat with headphones on in a sound-attenuated room with a computer monitor 55 inches away. All sounds were presented binaurally at approximately 60 dB above normal hearing threshold.

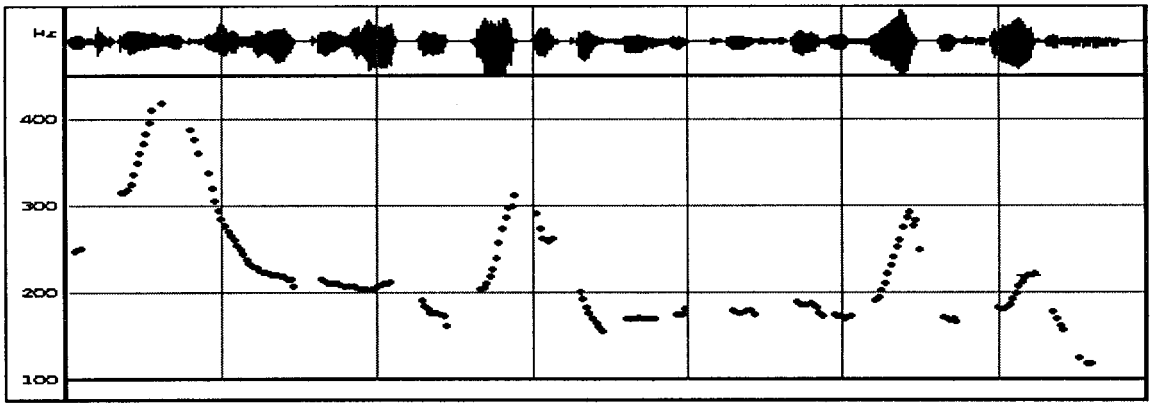
Figure 1. Spectograph and pitch analysis of the semantic, syntactic, and acoustic versions of one of the *t* test sentences. Visual inspections across entire sentences and statistical analyses of pitch and pitch change across the syllables that contained targets revealed no differences in pitch contours across sentence type.



The flowers were only there for **d**ecorative purposes, but she ate them anyway.



The sloners were only voll for **d**eblacass kelgavaz, but jee ogg them anyway.



Sha slonerf lew imro voll thal **d**eblacass kelgavaz, teg jee ogg shan ilowee.

During each trial, participants first heard the sound of the target phoneme or phoneme combination for that sentence and simultaneously saw a letter or letters representing that target appear on the screen. Subjects were instructed to listen for the target sound (not the presence of the letter or the /ə/ sound) in the sentence that followed 1,100 ms after the end of the target sound. The letter that represented the target was left on the screen for the entire trial.

For this experiment, participants had a choice of three responses. They were asked to press one button if they heard a target phoneme at the beginning of a word or nonword, were asked to press a different button if they heard the target in the middle of a word or nonword, and were asked to not respond at all if they did not hear a target. They were given examples of targets that occur at the beginning of words and nonwords, such as the /d/ in "devil," "destroy," "dossly," and "daclin," and targets that occur in the middle of words and nonwords such as the "d" in "wisdom," "tradition," "blomder," and "padell." Participants were reminded that the different buttons were to be used to indicate the two different positions a phoneme might have in a word and had nothing to do with the position of the word in the sentence. Furthermore, they were asked to press a button as soon as they heard the target sound in the sentence and were instructed not to wait for the sentence to end to make their response. Both accuracy and speed of the responses were emphasized. The next trial began 1,500 ms after the end of a sentence regardless of if, or when, the subject responded.

After 60 practice trails, participants were given feedback about their performance. They were given estimates of the percentage of trials on which they correctly detected a target, the percentage of trials on which they correctly determined where in a word a target occurred, and the average amount of time it took them to respond after a target occurred. During the test trials, participants were offered a break after every 20 sentences. Both the experimenter and the participant were required to press a button in order to continue after these breaks.

The order of presentation for the three forms (semantic, syntactic, and acoustic) of each sentence was balanced. Different forms of the same sentence were not presented with fewer than eighty trials in between. Within these constraints, sentences were randomized and presented in the same order for all subjects. Half of the participants were asked to use the left-most button to indicate a target was word initial; half were asked to use the right-most button to indicate a target was word initial.

Sounds were presented on a Pentium PC using a Data Translation (DT 2821) D-to-A converter. Sounds were low-pass filtered at 7500 Hz to prevent aliasing. Presentation and responses were controlled and recorded by C++ programs. Reaction times could be measured accurately within ±4 ms.

Results

Localization accuracy was measured by dividing the number of trials on which subjects successfully detected a target phoneme and determined whether it was word initial or word medial by the number of trials on which subjects correctly detected the target. Reaction times to perform this task were measured from the point at which the target phoneme or phonemes were presented.

A 3 (sentence type) × 2 (stress) × 2 (target position) repeated-measures ANOVA was performed. Additionally planned comparisons between sentence type (semantic and syntactic, syntactic and acoustic) and stress patterns (normal stress pattern and infrequent stress pattern) were performed. Item analyses (percentage of subjects who responded correctly for each item and mean reaction time for each item) were used to determine if the results are generalizable across words or if they were driven by a small subset of the stimuli. Concerns about violations of linearity with percentage data were met by using a natural log transformation of the proportions [score = ln(proportion correct/1 – proportion correct)]. Finally, the data were split into groups according to the type of targets subjects were listening for—consonant clusters, voiceless stops, voiced stops, fricatives, and nasals. Means and standard errors of percent correct and reaction times can be found in Appendix E. ANOVA tables and *t* tests for analysis by subject and by item can be seen in Appendix F.

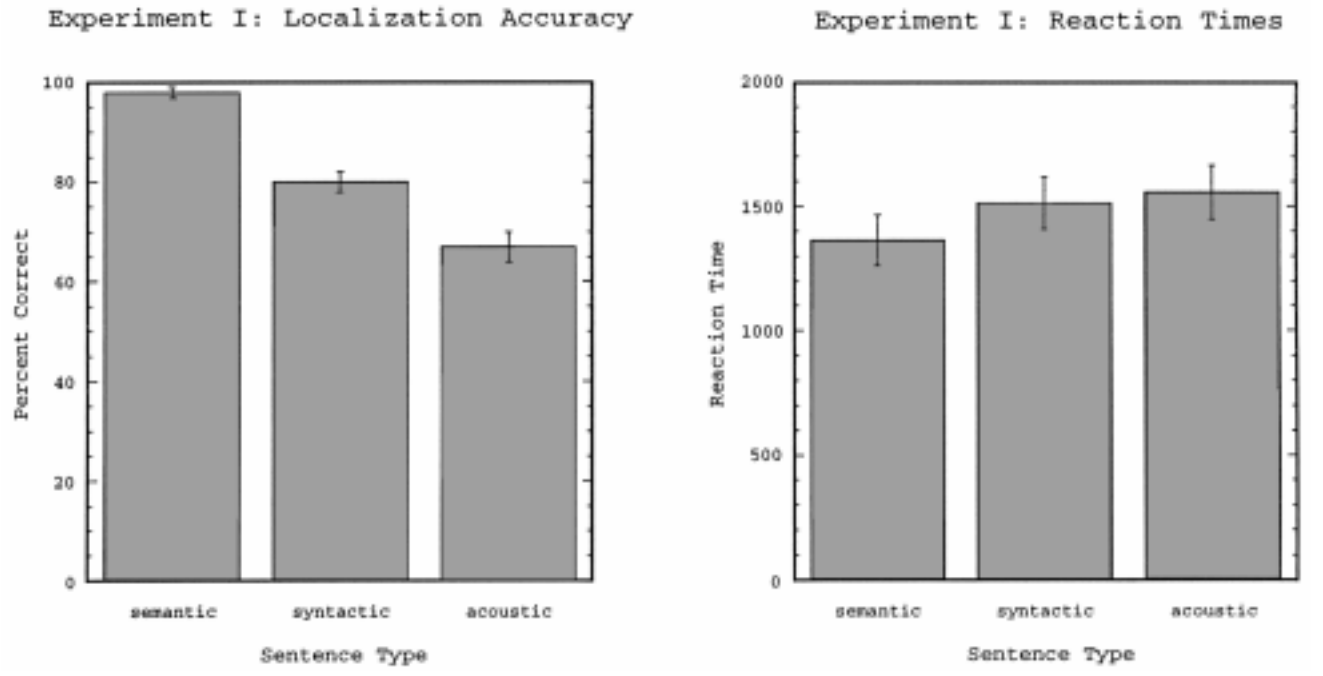
Sentence Type

Sentence type affected localization accuracy [$F(2, 30) = 385.3$, $p < .01$] such that performance was better for semantic sentences ($M = 98\%$) than syntactic sentences ($M = 80\%$), [$t(15) = 15.9$, $p < .01$] and for syntactic sentences than acoustic sentences ($M = 67\%$), [$t(15) = 13.48$, $p < .01$], as can be seen in Figure 2. Sentence type also affected reaction times [$F(2, 30) = 16.54$, $p < .01$] such that responses were faster to semantic sentences ($M = 1{,}363$ ms) than to syntactic sentences ($M = 1{,}512$ ms), [$t(15) = 4.88$, $p < .01$]. For each of the sentence types, performance across position and stress was better than chance [semantic: $t(15) = 110.7$, $p < .01$; syntactic: $t(15) = 21.8$, $p < .01$; acoustic: $t(15) = 12.49$, $p < .01$]. Each of these effects was confirmed by both the item analysis and the analysis on the natural log transformed data.

Stress Pattern

There was a stress by position interaction [$F(1, 15) = 42.49$, $p < .01$] on phoneme localization. When the

Figure 2. Experiment I, localization accuracy and reaction times by sentence type. Phoneme localization was better and faster in semantic sentences than in syntactic sentences. Performance was more accurate (but not significantly faster) for syntactic sentences than acoustic sentences.

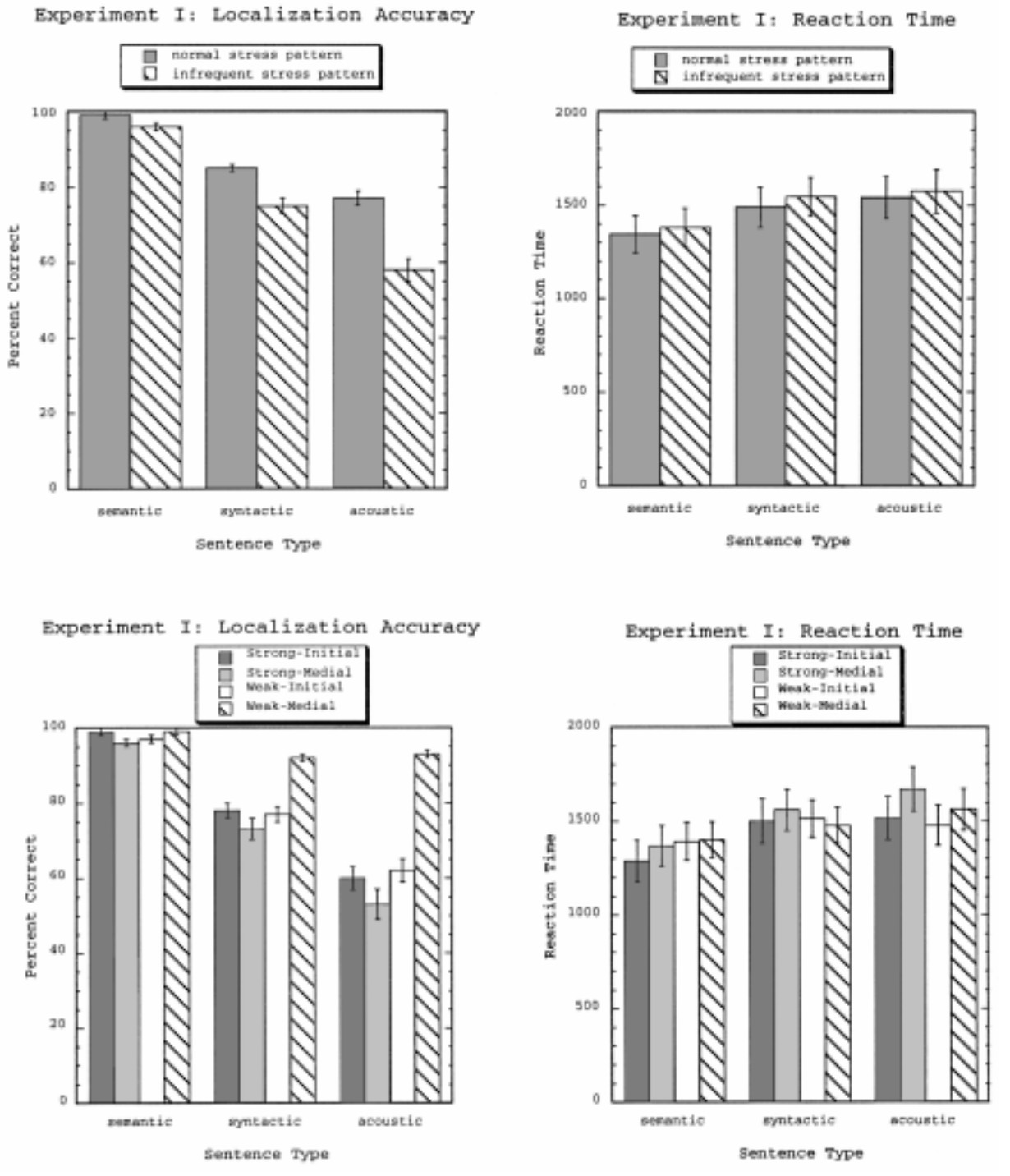


data were grouped to compare normal English stress pattern (strong-initial and weak-medial) to an infrequent English stress pattern (weak-initial and strong-medial), it was found that people were more accurate with the normal pattern [*M* normal = 87%, *M* infrequent = 76%, $F(1, 15) = 60.99$, $p < .01$]. Furthermore, stress pattern interacted with sentence type [$F(2, 30) = 17.1$, $p < .01$], such that performance was better with normal stress pattern than infrequent stress pattern for all sentence types [semantic: *M* normal = 99%, *M* infrequent = 96%, $F(1, 15) = 16.73$, $p < .01$; syntactic: *M* normal = 85%, *M* infrequent = 75%, $F(1, 15) = 59.07$, $p < .01$; acoustic: *M* normal = 77%, *M* infrequent = 58%, $F(1, 15) = 91.39$, $p < .01$], but this effect was larger when less information was available in the sentence as can be seen in Figure 3.

Additionally, the effect of stress pattern could be seen by comparing the four combinations of stress (strong and weak) and position (initial and medial) directly. For the semantic sentences, subjects were more likely to correctly identify the location of a strongly stressed phoneme when it was in the word-initial position (*M* = 99%) than when it was in the word-medial position (*M* = 96%), [$t(15) = 3.52$, $p < .01$]. Furthermore, in this same sentence type, people were better able to localize a weakly stressed phoneme when it was in the word-medial position (*M* = 99%) than when it was in the word-initial position (*M* = 97%), [$t(15) = 2.40$, $p < .01$]. For the syntactic sentences, accuracy was higher with the weakly stressed phoneme in the word-medial position (*M* = 92%) than in the word-initial position (*M* = 77%) [$t(15) = 10.17$, $p < .01$]. The means were in the predicted direction for the strongly stressed targets (*M* initial = 78%, *M* medial = 73%), but this difference was not significant. The same was true for the acoustic sentences in that the difference in word-initial and word-medial localization for the strong stress was not significant (*M* initial = 60%, *M* medial = 53%), but the difference in position for the weak stress was (*M* medial = 93%, *M* initial = 62%), [$t(15) = 10.61$, $p < .01$]. By examining these means, it is also clear that the sentence by stress by position interaction [$F(2, 30) = 24.46$, $p < .01$] is driven by the greater difference between weak-initial and weak-medial phonemes in the acoustic sentences than in the syntactic sentences and in the syntactic sentences than in the semantic sentences.

Although there was a stress by position interaction on reaction times [$F(1, 15) = 12.83$, $p < .01$], no other stress-pattern comparisons were significant. The ln transformation and item analysis of localization accuracy confirmed all of the significant results found in the ANOVA and *t* tests described above. When targets were grouped by phoneme class the following pattern of means held up across all groups: semantic > syntactic > acoustic and normal stress pattern > infrequent stress pattern.

Figure 3. Experiment I, localization accuracy and reaction times by stress pattern. Performance was better for targets in words with normal stress pattern than with infrequent stress pattern. This was true across sentence type, though the effect was larger for acoustic sentences than syntactic sentences and for syntactic sentences than semantic sentences.



Discussion

Subjects were extremely accurate (98%) at determining the position of the targets they detected in the normal English sentences. Their performance with the syntactic sentences was high (80%), but the lack of full semantic and lexical information was associated with decreased accuracy.

This suggests that word recognition played an important role in subjects' determining whether targets

were word initial or word medial. Furthermore, people were more accurate at the phoneme localization task with the syntactic sentences than with the acoustic sentences (67%). This suggests that the remaining lexical information in the syntactic sentences or the presence of more syntactic information in the form of morphemes and function words aided subjects in the phoneme localization task.

Localization accuracy was measured as the number of trials on which target position was correctly determined divided by the number of trials on which a target was successfully detected. Differences in localization accuracy on the three sentence types reflected differences in the ability of subjects to determine where in the speech stream word onsets occurred rather than differences in detectability of the targets in the different contexts. However, even when only detected targets are included it remains possible that differences in attention to, or memory for, the different types of sentences affected localization performance. For example, it is likely that people were better able to remember the normal English sentences than the nonword sentences. In a metalinguistic task like phoneme localization, better memory could allow for more accurate responses. From this experiment it is not clear whether the varying amounts of lexical and syntactic information affected online segmentation, or later reparsing and metalinguistic decisions which were likely to be influenced by many different processes including memory and attention. However, it is clear that the amount of lexical and syntactic information affected the assignment of targets to word-initial or word-medial positions, whether this effect was direct through segmentation or more indirect through memory or attention.

It is also important to recognize that even though the syntactic sentences had more grammatical information than the acoustic sentences, they also had more lexical items. It is possible that the intact function words in the syntactic sentences were recognized as lexical items and used as lexical information in the phoneme localization task. To determine if the larger amount of lexical information in the syntactic sentences was responsible for better performance, accuracy on trials that had a word immediately preceding the target was compared to accuracy on trials that had a nonword immediately preceding the target. The fact that there were no significant differences in performance for these two types of syntactic sentences suggests it was differences in grammatical structure, and not differences in the number of lexical items, that drove the better performance for the syntactic sentences than the acoustic sentences.

The fact that performance was better for words with normal English stress pattern (strong at the beginning, weak in the middle) than for words with infrequent English stress pattern (weak at the beginning, strong in the middle) for all sentence types suggests that stress pattern plays an important role in determining where word onsets occur. Although this effect was quite small for the normal English sentences (3% difference in accuracy), it suggests stress pattern has an effect even in normal continuous speech with all of the semantic, lexical, and acoustic information intact.

Furthermore, stress pattern seems to have a larger effect when other sources of information are absent as is suggested by the stress pattern by sentence type interaction. Although it is possible that the overall interaction was influenced by a ceiling effect with the semantic sentences, the interaction is still strong when only syntactic and acoustic sentences are included in the analysis. This suggests that people can flexibly use the segmentation cues available to them by relying more heavily on what is present when the number of segmentation cues are decreased. Accuracy was still high (86%) when the stress pattern gave misleading segmentation information, but only if lexical and syntactic information was present.

Strong stresses were more easily localized to word-initial positions than weak stresses were across sentence type. However, this effect was small in comparison to the difference between weak and strong stress for the word-medial targets. People were very accurate at localizing weak stresses in word-medial positions (99%, 92%, and 93% for semantic, syntactic, and acoustic sentences, respectively) and poor at localizing strong stresses to word-medial positions (96%, 73%, and 53% for semantic, syntactic, and acoustic sentences, respectively). Although this difference was not expected, one possible explanation is that people place more emphasis on one part of a stress-pattern segmentation strategy than the other. It is possible that people more strictly follow the pattern that weak stresses fall in the middle of words than the pattern that strong stresses fall at the beginnings of words. One motivation for stressing the different parts of a stress-pattern strategy unequally, may be that there is a higher cost associated with segmenting speech in places it should not be than there is in failing to segment speech where it should be. Perhaps when speech is segmented in the middle of a word based on stress pattern it is difficult to apply other cues and realize the units need to be considered together. However, when sometimes failing to segment speech where it should be, it may still be possible to consider other cues (such as lexical recognition and syntax) to make that break. Although this is a possible explanation for the pattern of data found for these experiments, more direct tests of the idea would be necessary to support or refute it.

It is also important to recognize that people were able to determine if targets were word initial or word

medial even when they had no lexical or syntactic information and when stress pattern cues were as likely to be misleading as helpful. Above chance level performance with the acoustic sentences (across the different location and stress levels) indicates that people were able to use some sources of segmentation information in addition to lexical, syntactic, and stress-pattern cues. Perhaps phonotactic and allophonic information, as well as transitional probabilities, were used in addition to the cues that were directly manipulated in this experiment.

Another important aspect of the localization task was the long response times. On average, it took people almost a second and a half to press one of the buttons after a target was presented. The long reaction times are not surprising considering the difficulty of detecting a target phoneme in over three seconds of continuous speech and determining whether that target was word initial or word medial. However, they also serve as an indication that phoneme localization was a difficult metalinguistic task that is likely to reflect many types of processing in addition to online segmentation. Therefore, an easier phoneme detection task was given using the same stimuli.

Experiment II: Phoneme Detection

Method

Participants

Seventeen monolingual English speakers participated (*M* age = 21.2 years; 13 women). All were right-handed university students who were paid $7 per hour.

Stimuli

For Experiment II, a subset of 660 of the sentences were used. Only 40 of the 60 normal English sentences that actually contained a target in each of the four possible combinations of location (word initial and word medial) and stress (strong and weak) were used. These 160 sentences were presented in each of the forms (semantic, syntactic, and acoustic) for a total of 480 sentences that contained targets. All 60 of the normal English sentences that did not contain a target and their syntactic and acoustic matches were used for a total of 180 sentences without targets.

Procedure

Participants were brought in for a single 2-hour experimental session. They completed 40 practice trials followed by 660 test trials. All presentation conditions were exactly the same as Experiment I. However, participants were simply asked to press a single button as quickly as possible if they heard the target sound in the sentence and to do nothing if they did not hear the target. For each trial, the time between the presentation of the target sound and a response were recorded.

Results

The percentage of trials on which subjects detected targets and reaction times was analyzed in the same manner as percent correct and reaction times for Experiment I. Additionally, phoneme detection rates from Experiment I were calculated by taking the percentage of trials on which subjects pressed either button (indicating the target was word initial *or* word medial) when a target was present. Phoneme detection rates for both experiments were broken down by target type. Means and standard errors for all conditions can be found in Appendix G. ANOVA tables and *t* tests can be seen in Appendix H.

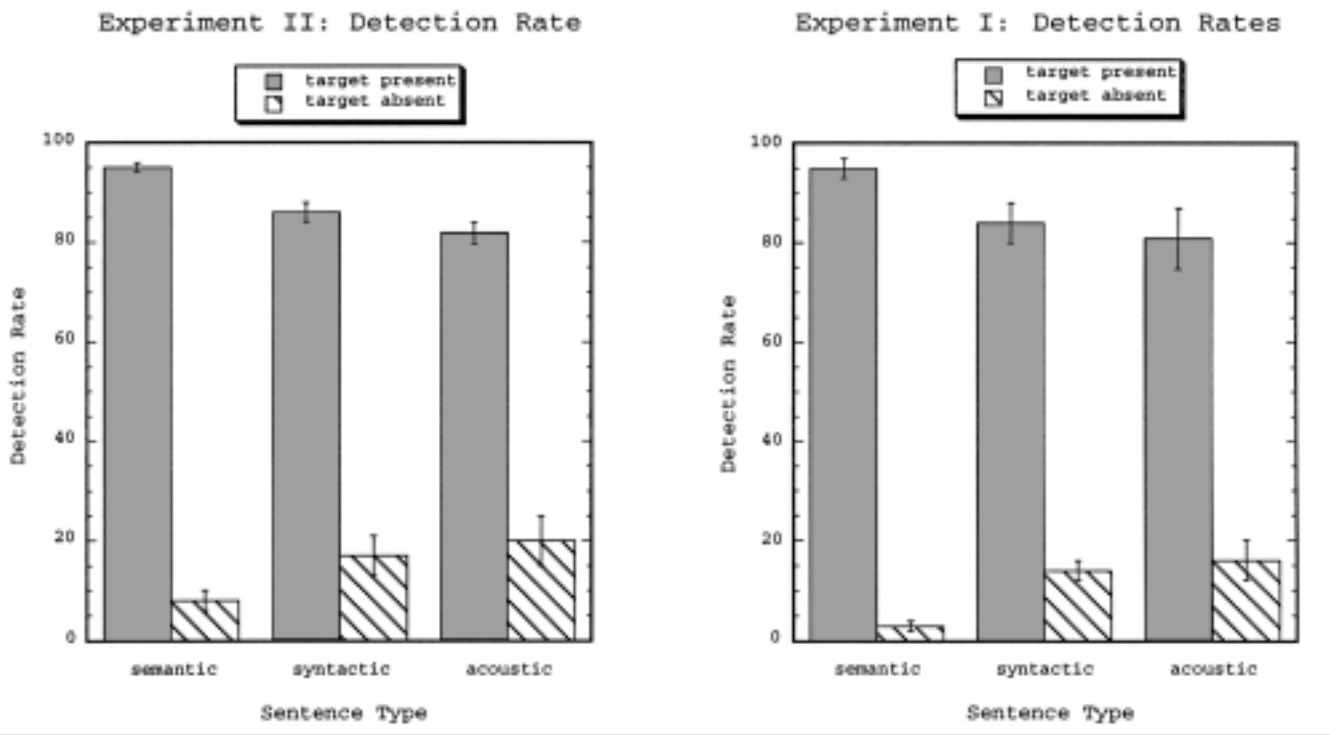
Sentence Type

People were more likely to detect targets in semantic sentences (M = 95%) than in syntactic sentences (M = 86%), [$t(16) = 5.8$, $p < .01$] and in syntactic sentences than in acoustic sentences (M = 82%), [$t(16) = 5.1$, $p < .01$] as can be seen in Figure 4. They were also more likely to make false alarms in syntactic sentences (M = 17%) than in semantic sentences (M = 8%), [$t(16) = 4.1$, $p < .01$]. Sentence type had no significant effect on reaction times (M semantic = 613 ms; M syntactic = 637 ms; M acoustic = 628 ms). The same effects were found on the natural log transformed data, using an item analysis, and for phoneme detection in Experiment I.

Stress and Position

Subjects were more likely and faster to detect target phonemes that occurred in strongly stressed syllables (M = 91%) than phonemes that occurred in weakly stressed syllables (M = 85%), [$F(1, 16) = 24.51$, $p < .01$]. Subjects were better and faster at detecting word-initial phonemes (M = 94%) than at detecting word-medial phonemes (M = 82%), [$F(1, 16) = 118.6$, $p < .01$]. Although this pattern was consistent across all three types of sentences [semantic: M initial = 97%; M medial = 93%; $F(1, 16) = 25.4$, $p < .01$; syntactic: M initial = 94%, M medial = 79%; $F(1, 16) = 75.9$, $p < .01$]; acoustic: M initial = 90%; M medial = 74%; $F(1, 16) = 86.7$, $p < .01$], a sentence type by position interaction [$F(2, 32) = 26.1$, $p < .01$] suggests that position played a larger role in detection while subjects listened to sentences

Figure 4. Phoneme detection by sentence type in Experiment II and in Experiment I. Detection rates were similar in the two experiments. Performance was better for semantic than syntactic sentences and for syntactic than acoustic sentences.



with less information (syntactic and acoustic sentences) than when they listened to normal English sentences (semantic). Although the sentence by position interaction was found for the natural log transformed data, it was not found using an item analysis.

There was also a stress by position interaction on phoneme detection [$F(1, 16) = 26.4$, $p < .01$]. Word-initial phonemes were more easily detected for both strongly stressed syllables [M initial = 95%; M medial = 86%; $F(1, 16) = 64.4$, $p < .01$] and weakly stressed syllables [M initial = 93%; M medial = 77%; $F(1, 16) = 103.7$, $p < .01$], but this position effect was larger for the weak stresses. There was no stress by position interaction on reaction time.

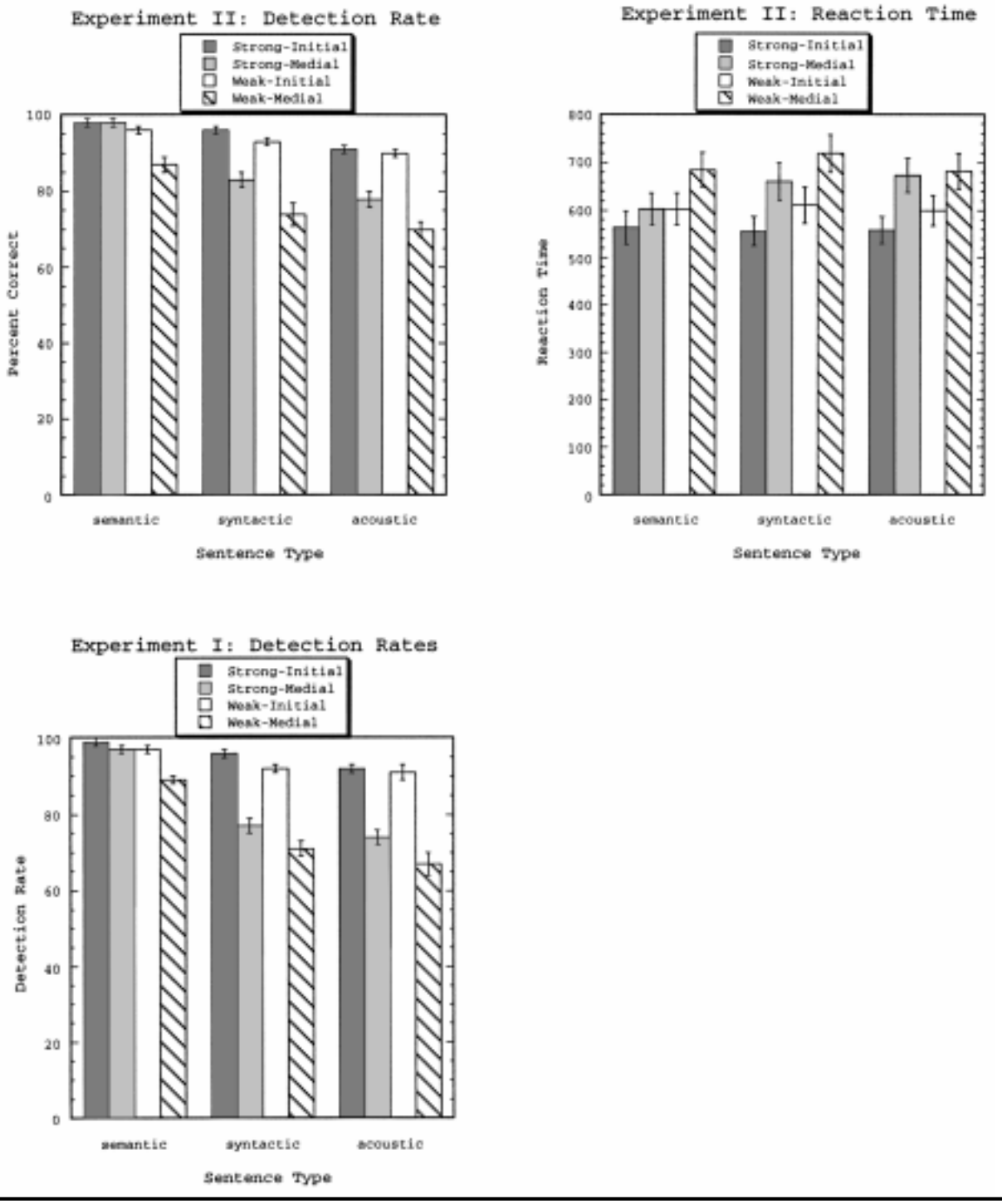
Each of these effects (stress, position, sentence by position, and stress by position) was found for both the natural log transformed data and the detection rates from Experiment I.

Discussion

Subjects were better at detecting targets in semantic sentences than in syntactic sentences and in syntactic sentences than in acoustic sentences. The finding that people were better and faster at detecting phonemes in the normal English sentences is not surprising considering that targets were presented in words in these sentences and in nonwords in the other sentence types. A reaction time advantage for words over nonwords has been found using several variants of the phoneme detection task (Cutler, Mehler, Norris, & Segui, 1987; Eimas, Hornstein, & Payton, 1990; Frauenfelder, Segui, & Dijkstra, 1990; Pitt & Samuel 1995; Rubin, Turvey, & van Gelder, 1976). Although the word advantage has not been found in some experiments (Cutler, Mehler, Norris, & Segui, 1987; Foss & Blank, 1980), it would be expected in experiments like this one that contain rich lexical and syntactic information and encourage processing at many different levels.

To understand why people were better at detecting phonemes in nonwords occurring in the syntactic sentences than at detecting phonemes in nonwords occurring in the acoustic sentences, it is important to look at context effects. Segui, Frauenfelder, and Mehler (1981) found a reaction time advantage for targets occurring after a word in comparison to targets occurring after a nonword when they presented a mixed list of isolated words and nonwords. Some of the syntactic sentences used in this experiment contained nonwords that were preceded by a function word as seen in the example of a syntactic sentence with a word-initial, weakly stressed target (Table 2). This could not happen in acoustic sentences that contained no words. Additionally, several researchers have found slower reaction times for syntactically complex sentences (Foss & Lynch, 1969; Hakes

Figure 5. Phoneme detection by stress and position in Experiment II and in Experiment I. Detection rates were higher for targets in strongly stressed syllables and for word-initial targets across sentence type.



& Foss, 1970; Hakes 1972). Although on the surface it may seem that sentences that contain function words (the syntactic sentences) are more syntactically complex than those that do not (the acoustic sentences), if participants were trying to ascertain a grammatical structure from the acoustic sentences it is likely that this process would be more taxing. The difficulty of forming a grammatical structure, rather than complexity of the syntactic information available, may have contributed to greater difficulty in detecting targets.

Previous studies have also shown phoneme detection benefits for targets in strongly stressed syllables in

comparison to targets in weakly stressed syllables (Cutler & Foss, 1977; Cutler & Darwin, 1981; Mehta & Cutler, 1988; Sheilds, McHugh, & Martin, 1974). Most of these benefits were in terms of reaction times, but this study shows stress can effect overall accuracy as well.

Fewer studies have been conducted on the effects of phoneme position on detection. Pallier et al. (1993) reported a series of five experiments in which they manipulated both target location (final phoneme of the first syllable or first phoneme of the second syllable) and the probability that a target would occur in a specific location. They found better phoneme detection for the most probable location but no affect of location independent of probability. A similar design was used with English speakers (Finney, Protopapas, & Eimas, 1996) with similar results. Because the targets in this study were equally likely to occur in word-initial and word-medial positions, the position effects cannot be attributed to probability differences. Furthermore, the previous evidence that location does not affect phoneme detection for words in isolation suggests the position effects in this study were related to processing of continuous speech or of words in context.

If improved phoneme detection for word-initial phonemes in comparison to word-medial phonemes is taken as a measure of segmentation, we would expect that this effect would be larger for the sentence types in which segmentation is easier (semantic and syntactic). However, the position effect was actually smallest for the normal English sentences. One possible explanation for this is that in normal English sentences the word-medial phonemes may be easier to predict than the word-initial phonemes. The sentences were designed to have low cloze probability for the word in which targets occur, but if subjects have already heard the first part of a word they may be able to predict the rest of it. In the case of the word-medial targets, the prediction could make phoneme detection easier. In this case, the smaller position effect for the semantic sentences could be explained as a trade-off between improved performance for the word-initial targets because of accurate segmentation and improved performance for the word-medial targets because of predictability.

Another issue is the extent to which the phoneme monitoring task did in fact tap into online segmentation. The reaction times were much faster (around 600 ms) than for the localization task. However, 600 ms is plenty of time for people to begin processing information presented after the target and does not exclude the possibility of post-lexical processing. When the same stimuli were used with people who do not speak English, they were able to do the task. This suggests that phoneme monitoring can be done without additional lexical and syntactic processing, but does not ensure that native English speakers who have access to that additional information do not use it in the task.

Conclusions

In general, this study supports the idea that multiple segmentation cues are not only available in speech, but are actually used by listeners in combination. The fact that people had more difficulty determining where word onsets fell in sentences without lexical information, even when normal stress-pattern cues were available, suggests that semantic and lexical information can be used to identify word onsets in continuous speech and cannot be completely compensated for with other cues. The fact that stress pattern affected phoneme localization performance even with normal English sentences suggests that stress is not merely a segmentation cue that gets emphasized in studies using two syllable utterances and nonsense words for input, but that it may play an important role in the normal segmentation of spoken English.

These studies also suggest that more weight may be given to remaining segmentation cues when some sources of information are absent. Stress-pattern cues became more important to phoneme localization when lexical and syntactic cues were absent. However, people did not make their localization decisions solely on the basis of stress information when listening to the acoustic form of the sentences. From this study, it is not clear whether that was caused by the presence of other segmentation cues that provided information conflicting with the stress information or by a failure of people to completely shift their segmentation strategy to take full advantage of the remaining stress pattern cues. Although people may be able to use combinations of segmentation cues flexibly, it is likely that there are some constraints on their use.

The same patterns of performance on the target localization and target detection tasks were found across the different types of phonemes used as targets. This suggests that segmentation cues such as lexical, syntactic, and stress-pattern information, are used with speech in general rather than in particular circumstances. Furthermore, finding the same pattern of results across target type and similar results using an item analysis confirms that the findings were not driven by deviant performance with a few items or subsets of items.

Phoneme localization has the advantage of directly asking subjects where word onsets occur in a speech stream. However, long reaction times and the difficult nature of this task suggest it may be measuring processes other than online segmentation. Phoneme detection has the advantage of avoiding a metalinguistic task. However, drawing conclusions about segmentation from this task requires the assumption that segment onsets

will be detected more quickly than items in the middle of a segment. Furthermore, information following the point of interest in the stimuli and post-lexical processing of the items themselves may be reflected in the responses. Both of these tasks can be used with uninterrupted continuous speech. Converging evidence from tasks like these that make all normally available segmentation cues available in the stimuli and approaches that can index pre-lexical online processing (such as event-related potential recordings) will be necessary to determine exactly which cues and combination of cues people use to initially segment speech.

Previous research has focused on determining what types of cues are available in speech that might be used in segmentation, or showing that one of these many available cues affects performance with a particular type of stimuli. However, a full understanding of speech segmentation will require the use of stimuli that make multiple segmentation cues available and that encourage people to segment speech as they normally do.

Acknowledgments

This research was presented, in part, as a poster entitled "Speech Segmentation by Bilingual Speakers" at the 1998 meeting of the Cognitive Neuroscience Society in San Francisco. The research was supported by NIH, NIDCD Grant DC000128. The authors wish to thank Dr. Douglas L. Hintzman and Dr. Susan Guion for their helpful comments in preparing this manuscript.

References

Bernstein, J., & Franco, H. (1996). Speech recognition by computer. In N. J. Lass (Ed.), *Principles of experimental phonetics* (pp. 408–434). St. Louis, MO: Mosby.

Brent, M. R. (1999). Speech segmentation and word discovery: A computational perspective. *Trends in Cognitive Sciences, 3,* 294–301.

Brent, M. R. (1997). Toward a unified model of lexical acquisition and lexical access. *Journal of Psycholinguistic Research, 26,* 363–375.

Brent M. R., & Cartwright, T. A. (1996). Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition, 61,* 93–125.

Brown, G. D. (1984). A frequency count of 190,000 words in the London-Lund Corpus of English Conversation. *Behavior Research Methods, Instruments, & Computers, 16,* 502–532.

Church, K. (1987). Phonological parsing and lexical retrieval. *Cognition, 25,* 53–69.

Cole, R. A., Jakimik, J., & Cooper, W. E. (1980). Segmenting speech into words. *Journal of the Acoustical Society of America, 67,* 1323–1332.

Cowan, N. (1991). Recurrent speech patterns as cues to the segmentation of multisyllabic sequences. *Acta Psychologica, 77,* 121–135.

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

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

Cutler, A., & Darwin, C. J. (1981). Phoneme-monitoring reaction time and preceding prosody: Effects of stop closure duration and of fundamental frequency. *Perception and Psychophysics, 29,* 217–224.

Cutler, A., & Foss, D. J. (1977). On the role of sentence stress in sentence processing. *Language and Speech, 20,* 1–10.

Cutler, A., Mehler, J., Norris, D. G., & Segui, J. (1986). The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language, 25,* 385–400.

Cutler, A., Mehler, J., Norris, D. G., & Segui, J. (1987). Phoneme identification and the lexicon. *Cognitive Psychology, 19,* 141–177.

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

Cutler, A., & Otake, T. (1994). Mora or phoneme? Further evidence for language-specific listening. *Journal of Memory and Language, 32,* 358–378.

Finney, S. A., Protopapas, A., & Eimas, P. D. (1996). Attentional allocation to syllables in American English. *Journal of Memory and Language, 35,* 893–909.

Frauenfelder, U. H. (1985). Cross-linguistic approaches to lexical segmentation. *Linguistics, 23,* 669–687

Harris, A. S. (1955). From phoneme to morpheme. *Language, 31,* 190–222.

Hayes, J. R., & Clark, H. H. (1970). Experiments on the segmentation of an artificial speech analogue. In J. R. Hayes (Ed.), *Cognition and the development of language* (pp. 221–234). New York: Wiley.

Eimas, P. D., Hornstein, S. B. M., & Payton, P. (1990). Attention and the role of dual codes in phoneme monitoring. *Journal of Memory and Language, 29,* 160–180.

Foss, D. J., & Blank, M. A. (1980). Identifying the speech codes. *Cognitive Psychology, 12,* 1–31.

Foss, D. J., & Lynch, R. H. (1969). Decision processes during sentence comprehension: Effects of surface structure on decision times. *Perception and Psychophysics, 5,* 145–148.

Frauenfelder, U. H., Segui, J., & Dijkstra, T. (1990). Lexical effects in phonemic processing: Facilitatory or inhibitory. *Journal of Experimental Psychology: Human Perception and Performance, 16,* 77–91.

Hakes, D. T. (1972). Effects of reducing complement constructions on sentence comprehension. *Journal of Verbal Learning and Verbal Behavior, 11,* 278–286.

Hakes, D. T., & Foss, D. J. (1970). Decision processes during sentence comprehension: Effects of surface structure reconsidered. *Perception and Psychophysics, 8,* 413–416.

Kucera, H., & Francis, W. N. (1967). *Computational*

analysis of present-day American English. Providence, RI: Brown University Press.

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

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

McQueen, J. M., Norris, D. G., & Cutler, A. (1994). Competition in spoken word recognition: Spotting words in other words. *Journal of Experimental Psychology: Learning, Memory and Cognition, 20,* 621–638.

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

Mehta, G., & Cutler, A. (1988). Detection of target phonemes in spontaneous and read speech. *Language and Speech, 31,* 135–156.

Norris, D. G., McQueen, J. M., & Cutler, A. (1995). Competition and segmentation in spoken word recognition. *Journal of Experimental Psychology: Learning, Memory and Cognition, 21,* 1209–1228.

Otake, T., Hatano, G., Cutler, A., & Mehler, J. (1993). Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language, 32,* 358–378.

Pallier, C., Sebastian-Galles, N., Felguera, T., Christophe, A., & Mehler, J. (1993). Attentional allocation within the syllabic structure of spoken words. *Journal of Memory and Language, 32,* 373–389.

Pitt, M. A., & Samuel, A. G. (1995). Lexical and sublexical feedback in auditory word recognition. *Cognitive Psychology, 29,* 149–188.

Quene, H. (1992). Durational cues for word segmentation in Dutch. *Journal of Phonetics, 20,* 331–350.

Rubin, P., Turvey, M. T., & van Gelder, P. (1976). Initial phonemes are detected faster in spoken words than in spoken nonwords. *Perception and Psychophysics, 19,* 394–398.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274,* 1926–1928.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language, 35,* 606–621.

Shields, J. L., McHugh, A., & Martin, J. G. (1974). Reaction time to phoneme targets as a function of rhythmic cues in continuous speech. *Journal of Experimental Psychology, 102,* 250–255.

Segui, J., Frauenfelder, U., & Mehler, J. (1981). Phoneme monitoring, syllable monitoring and lexical access. *British Journal of Psychology, 72,* 471–477.

Tyler, L. K., & Wessels, J. (1983). Quantifying contextual contributions to word-recognition processes. *Perception and Psychophysics, 34,* 409–420.

Umeda, N., & Coker, C. H. (1974). Allophonic variation in American English. *Journal of Phonetics, 2,* 1–5.

Vroomen, J., van Zon, M., & de Gelder, B. (1996). Cues to speech segmentation: Evidence from juncture misperceptions and word spotting. *Memory and Cognition, 24,* 744–755.

Received April 28, 2000

Accepted July 17, 2000

Contact author: Lisa D. Sanders, Department of Psychology, 1227 University of Oregon, Eugene, OR 97403-1227. Email: lsanders@darkwing.uoregon.edu

Appendix A (p. 1 of 2). Words that contained targets in the semantic sentences.

Target	Strong initial	Strong medial	Weak initial	Weak medial
Voiced Stops				
/b/	backgammon	abolish	*ballet*	arbitrary
/b/	*badminton*	*embellish*	balloon	embolism
/b/	*banter*	habituate	*behold*	habitat
/b/	bargain	imbedded	belated	*hibernate*
/b/	barrels	lobotomy	belief	*pebble*
/b/	binary	obey	belligerent	symbol
/b/	bottles	*tobacco*	beret	timber
/b/	butane	verbatim	bologna	trouble
/d/	daiquiri	*abduction*	deception	indicate
/d/	dangerous	condemn	*decisive*	Tuesday
/d/	daughter	*indent*	*defile*	*verdict*
/d/	decorative	pedestrians	degree	*wisdom*
/d/	*diagram*	tradition	deny	yardage
/d/	*dormitory*	vindictive	diploma	zodiac
/g/	*gadgets*	*begotten*	gazebo	bagels
/g/	gavels	forget	gazelles	haggle
/g/	ghetto	negate	gorillas	*loggers*
/g/	gophers	regurgitate	*gourmet*	staggered
Voiceless Stops				
/k/	*catalogues*	percussion	cajole	*alchemy*
/k/	*coffee*	*stockade*	*coerce*	*ticket*
/k/	company	vicariously	*cognition*	vacant
/k/	convolute	*volcano*	compete	welcome
/p/	*peanut*	apology	persuade	vampires
/p/	pesticide	*torpedo*	*potato*	*zipper*
/t/	*tabloids*	*baton*	*taboli*	*banter*
/t/	*technical*	*photography*	*tamales*	*mentoring*
/t/	*temper*	*pretentious*	*together*	*vegetables*
/t/	*towel*	*utility*	*tycoon*	*wilted*

Appendix A (p. 2 of 2). Words that contained targets in the semantic sentences.

Target	Strong initial	Strong medial	Weak initial	Weak medial
Fricatives				
/f/	figment	effect	forgotten	laughing
/f/	filter	perfect	philosophy	pilfer
/f/	phobia	perform	phonetics	waffle
/j/	juvenile	inject	judicious	injury
/s/	*satire*	*dissection*	*survival*	*rhapsody*
/th/	thimble	pathetic	theology	anthem
/v/	*victim*	*advantage*	valet	advertise
/v/	vintage	adventure	*viscosity*	savage
/v/	*vocalize*	*ovation*	vitality	*swivels*
/v/	volatile	pervasive	*vocation*	*woven*
Nasals				
/m/	magnify	amass	machines	atmosphere
/m/	mannequin	*amazed*	magnificent	caramelize
/m/	massacre	emerges	majestic	chemicals
/m/	mayonnaise	harmonics	*manipulate*	culminated
/m/	microscope	*imagine*	*minute*	similes
/m/	misanthrope	lament	mistake	terminal
/m/	*monarch*	permission	mundane	*thermostat*
/m/	*morsel*	submit	mysterious	*treatment*
/n/	*native*	*annoy*	*necessity*	annex
/n/	*natural*	inept	*negate*	*fitness*
/n/	navigate	tenacity	neglect	*tennis*
/n/	nectar	tornado	neurotic	tenor
/n/	noticeable	*vanilla*	nomadic	witness
Clusters				
/bl/	blemish	obliged	blockade	emblem
/br/	broker	sobriety	brassiere	labyrinth
/fl/	flagrantly	affluent	phlebotomy	influx
/gr/	gravity	ingredient	grammatical	fragrances
/pl/	plagiarize	complex	platonic	duplex
/pr/	promise	approach	protect	compromise
/qu/	quantity	equipment	quadratic	equal
/st/	staples	constituent	stability	*constitute*
/st/	*statue*	*mistake*	*stentorian*	institute

Note. Words in italics were used for Experiment I only.

Appendix B. Means and ranges for words in all conditions.

Condition	Cloze probability	Written frequency	Spoken frequency	Number of letters	Position in sentence
Strong stress, Initial position	0.022 (0 –0.175)	19.93 (0 –290)	2.02 (0 –44)	7.50 (5–11)	9.77 (5–17)
Strong stress, Medial position	0.016 (0 –0.215)	21.43 (0 –213)	2.08 (0 –29)	7.32 (4–11)	9.32 (5–14)
Weak stress, Initial position	0.065 (0 –0.100)	21.25 (0 –267)	2.57 (0 –34)	7.73 (4–11)	8.78 (4–15)
Weak stress, Medial position	0.026 (0 –0.200)	21.53 (0 –134)	2.33 (0 –55)	7.80 (5–11)	8.82 (5–15)
No Target	0.031 (0 –0.150)	20.71 (0 –226)	2.64 (0 –48)	7.22 (4–10)	9.24 (4–13)

Note. Cloze probability was calculated from responses of 40 participants. Written word frequencies were calculated from Kucera and Francis (1967). Spoken word frequencies were calculated from Brown (1984). There were no significant differences in any of the measures for stress, position, or stress by position interaction.

Appendix C. Means and ranges for auditory stimuli.

Condition	Speech rate (words/s)	Differences in sentence length (ms)	Differences in target position (ms)	Sentence length (ms)	Target position (ms)
Strong stress, Initial position	4.49 (3.46–7.45)	52.66 (0 –288)	77.45 (0 –293)	3,476.79 (2,416–3,822)	1,607.93 (506–2,779)
Strong stress, Medial position	4.38 (3.14–6.05)	45.92 (0 –172)	77.48 (0 –280)	3,446.95 (2,624–3,796)	1,656.42 (699–2,689)
Weak stress, Initial position	4.19 (2.94–5.41)	49.40 (0 –123)	72.19 (0 –230)	3,419.74 (2,366–3,838)	1,420.35 (293–2,598)
Weak stress, Medial position	4.02 (2.72–8.47)	45.32 (0 –135)	79.63 (0 –316)	3,468.21 (2,244–3,789)	1,778.24 (817–3,027)
No Target	4.24 (3.52–5.60)	53.81 (0 – 262)		3,382.44 (2,117–3,750)	

Note. Differences in sentence length and target position were calculated by taking the largest difference across the three versions of the same sentence and averaging these differences across the sentences in a condition. There were no significant differences in speech rate, sentence length, or target position for sentence type, stress, position, or any of the interactions among these three variables.

Appendix D. Means and ranges of amplitude, length, and fundamental frequency change.

Condition	Length (ms)		Amplitude		Frequency change (Hz)	
Strong stress, initial position:						
Semantic	163.83	(75 to 263)	0.70	(0.36 to 1.0)	+51	(–3 to +142)
Syntactic	163.43	(78 to 234)	0.78	(0.34 to 1.0)	+57	(–5 to +155)
Acoustic	156.07	(75 to 220)	0.75	(0.37 to 1.0)	+56	(–2 to +139)
Strong stress, medial position:						
Semantic	176.90	(79 to 241)	0.80	(0.34 to 1.0)	+66	(–3 to +169)
Syntactic	173.00	(90 to 267)	0.78	(0.38 to 1.0)	+65	(–2 to +171)
Acoustic	175.13	(66 to 252)	0.74	(0.30 to 1.0)	+63	(–2 to +161)
Weak stress, initial position:						
Semantic	87.90	(22 to 190)	0.31	(0.12 to 0.70)	+29	(–126 to +110)
Syntactic	83.47	(38 to 156)	0.30	(0.11 to 1.0)	+31	(–132 to +127)
Acoustic	82.45	(46 to 189)	0.32	(0.11 to 0.96)	+28	(–168 to +147)
Weak stress, medial position:						
Semantic	119.13	(32 to 225)	0.44	(0.13 to 1.0)	–20	(–110 to +3)
Syntactic	120.18	(39 to 216)	0.42	(0.11 to 1.0)	–23	(–93 to +11)
Acoustic	116.70	(46 to 119)	0.40	(0.10 to 1.0)	–21	(–102 to +5)

Note. Phoneme lengths were measured from the beginning of the target phoneme to the end of the following vowel. Maximum amplitudes are the greatest amplitude (between 0 and 1) during that span. Fundamental frequency change is the frequency at the beginning of that span subtracted from the frequency at the end of that span.

Appendix E1. Means and standard errors for Experiment I.

Condition	Percent correct (SE)		Reaction time (SE)	
Strong stress, initial position:				
Semantic	99%	(1.0)	1288 ms	(108)
Syntactic	78%	(2.0)	1499 ms	(118)
Acoustic	60%	(3.0)	1515 ms	(116)
Strong stress, medial position:				
Semantic	96%	(1.0)	1370 ms	(106)
Syntactic	73%	(3.0)	1560 ms	(110)
Acoustic	53%	(4.0)	1670 ms	(120)
Weak stress, initial position:				
Semantic	97%	(1.0)	1392 ms	(100)
Syntactic	77%	(2.0)	1513 ms	(99)
Acoustic	62%	(3.0)	1479 ms	(108)
Weak stress, medial position:				
Semantic	99%	(1.0)	1400 ms	(96)
Syntactic	92%	(1.0)	1478 ms	(96)
Acoustic	93%	(1.0)	1562 ms	(111)

Note. Localization accuracy is the percentage of trials on which subjects correctly identified the position of a target out of those trials on which the subjects correctly detected a target. Reaction times are for those trials on which subjects correctly identified the target position only. Standard Errors are shown in parenthesis.

Appendix E2. Means of localization accuracy by phoneme class in Experiment I.

Condition	Voiced stops	Voiceless stops	Fricatives	Nasals	Clusters
Strong stress, initial position:					
Semantic	99%	100%	99%	98%	99%
Syntactic	75%	81%	78%	82%	77%
Acoustic	68%	52%	53%	56%	69%
Strong stress, medial position:					
Semantic	97%	97%	95%	96%	92%
Syntactic	73%	76%	69%	78%	66%
Acoustic	66%	51%	51%	49%	39%
Weak stress, initial position:					
Semantic	97%	98%	98%	98%	92%
Syntactic	71%	76%	83%	73%	67%
Acoustic	64%	68%	59%	56%	67%
Weak stress, medial position:					
Semantic	98%	100%	99%	99%	99%
Syntactic	86%	96%	90%	97%	94%
Acoustic	96%	94%	84%	93%	97%

Note. Localization accuracy is the percentage of trials on which subjects correctly identified the position of a target out of those trials on which the subjects correctly detected a target. Voiced stops (N = 18) were /b/, /d/, and /g/; voiceless stops (N = 10) were /k/, /p/, and /t/; fricatives (N = 10) were /f/, /j/, /s/, /th/, and /v/; nasals (N = 13) were /m/ and /n/; clusters (N = 9) were /bl/, /br/, /fl/, /gr/, /pl/, /pr/, /qu/, and /st/.

Appendix F (p. 1 of 2). ANOVAs for Experiment I.

Dependent variable	Source of variance	By subject		ln transformed		By item	
		F	(*p*)	*F*	(*p*)	*F*	(*p*)
Localization accuracy:	Sentence	385.35	(.000)	303.24	(.000)	171.61	(.000)
	Stress	21.32	(.000)	23.21	(.000)	13.06	(.000)
	Position	23.01	(.000)	32.84	(.000)	61.07	(.000)
	Sentence × Stress	22.29	(.000)	21.61	(.000)	7.55	(.001)
	Sentence × Position	21.19	(.000)	12.89	(.000)	21.51	(.000)
	Stress × Position	42.49	(.000)	55.71	(.000)	59.29	(.000)
	Sent. × Stress × Pos.	24.46	(.000)	7.95	(.002)	13.15	(.000)
Reaction times:	Sentence	16.54	(.000)			27.86	(.000)
	Stress	0.24	(*ns*)			8.53	(.004)
	Position	9.61	(.007)			0.37	(*ns*)
	Sentence × Stress	12.22	(.000)			1.74	(*ns*)
	Sentence × Position	3.44	(*ns*)			1.53	(*ns*)
	Stress × Position	12.83	(.003)			5.94	(*ns*)
	Sent. × Stress × Pos.	0.13	(*ns*)			0.36	(*ns*)

Appendix F (p. 2 of 2). ANOVAs for Experiment I.

Dependent variable	Source of variance	By subject *t*	By subject (*p*)	ln transformed *t*	ln transformed (*p*)	By item *t*	By item (*p*)
Localization accuracy:	Sentence type:						
	Semantic/syntactic	15.98	(.000)	19.71	(.000)	13.11	(.000)
	Syntactic/acoustic	13.48	(.000)	12.35	(.000)	9.41	(.000)
	Stress pattern:						
	Norm/infreq:						
	Semantic	4.09	(.002)	4.14	(.001)	4.08	(.000)
	Syntactic	7.68	(.000)	6.27	(.000)	8.12	(.000)
	Acoustic	9.56	(.000)	5.86	(.000)	11.40	(.000)
	SI/SM:						
	Semantic	3.52	(.003)	3.29	(.005)	4.02	(.000)
	Syntactic	0.94	(*ns*)	0.88	(*ns*)	1.10	(*ns*)
	Acoustic	1.13	(*ns*)	1.61	(*ns*)	3.61	(.000)
	WI/WM:						
	Semantic	2.40	(.009)	2.32	(*ns*)	3.95	(.000)
	Syntactic	10.17	(.000)	7.02	(.000)	8.61	(.000)
	Acoustic	10.61	(.000)	8.36	(.000)	9.52	(.000)
Reaction times:	Sentence type:						
	Semantic/syntactic	4.88	(.000)			6.37	(.000)
	Syntactic/acoustic	1.47	(*ns*)			2.61	(*ns*)
	Norm/infreq:						
	Semantic	1.73	(*ns*)			0.97	(*ns*)
	Syntactic	1.66	(*ns*)			0.96	(*ns*)
	Acoustic	0.80	(*ns*)			0.60	(*ns*)

Appendix G1. Means and standard errors of target detection in Experiment II.

Condition	Phoneme detection (*SE*)	Reaction time (*SE*)	Phoneme detection Experiment I (SG)
Strong stress, Initial position:			
Semantic	98% (1.0)	563 ms (35)	99% (1.0)
Syntactic	96% (1.0)	556 ms (31)	96% (1.0)
Acoustic	91% (1.0)	558 ms (29)	92% (1.0)
Strong stress, Medial position:			
Semantic	98% (1.0)	603 ms (34)	97% (1.0)
Syntactic	83% (2.0)	661 ms (39)	77% (2.0)
Acoustic	78% (2.0)	673 ms (36)	74% (2.0)
Weak stress, Initial position:			
Semantic	96% (1.0)	602 ms (34)	97% (1.0)
Syntactic	93% (1.0)	611 ms (38)	92% (1.0)
Acoustic	90% (1.0)	599 ms (32)	91% (2.0)
Weak stress, Medial position:			
Semantic	87% (2.0)	685 ms (36)	89% (1.0)
Syntactic	74% (3.0)	719 ms (39)	71% (2.0)
Acoustic	70% (2.0)	682 ms (38)	67% (3.0)

Note. Phoneme detection is the percentage of trials on which subjects detected a present target. Standard errors are shown in parentheses.

Appendix G2. Means of phoneme detection by phoneme class in Experiment II.

Condition	Voiced stops	Voiceless stops	Fricatives	Nasals	Clusters
Strong stress, Initial position:					
Semantic	99 (100)	99 (100)	97 (98)	96 (97)	100 (99)
Syntactic	97 (97)	99 (98)	94 (96)	94 (95)	98 (97)
Acoustic	93 (93)	92 (92)	89 (90)	88 (90)	95 (94)
Strong stress, Medial position:					
Semantic	98 (99)	99 (99)	96 (97)	96 (92)	99 (98)
Syntactic	85 (82)	92 (88)	79 (71)	76 (69)	86 (77)
Acoustic	83 (78)	86 (81)	67 (67)	71 (69)	80 (72)
Weak stress, Initial position:					
Semantic	100 (98)	98 (97)	93 (94)	89 (97)	97 (95)
Syntactic	97 (97)	97 (95)	88 (90)	87 (87)	93 (91)
Acoustic	94 (93)	93 (93)	86 (89)	81 (85)	93 (93)
Weak stress, Medial position:					
Semantic	90 (94)	89 (96)	81 (84)	84 (79)	88 (90)
Syntactic	81 (82)	82 (74)	71 (61)	62 (63)	75 (68)
Acoustic	81 (72)	76 (73)	64 (58)	56 (60)	66 (68)

Note. Mean phoneme detection from Experiment I is shown in parenthesis. Voiced stops (N = 18) were /b/, /d/, and /g/; voiceless stops (N = 10) were /k/, /p/, and /t/; fricatives (N = 10) were /f/, /j/, /s/, /th/, and /v/; nasals (N = 13) were /m/ and /n/; clusters (N = 9) were /bl/, /br/, /fl/, /gr/, /pl/, /pr/, /qu/, and /st/.

Appendix H. ANOVAs for Experiment II.

Dependent variable	Source of variance	By subject F	(p)	ln transformed F	(p)	By item F	(p)
Phoneme detection:	Sentence	49.45	(.000)	71.77	(.000)	29.73	(.000)
	Stress	24.51	(.000)	33.17	(.000)	94.42	(.000)
	Position	118.64	(.000)	187.37	(.000)	24.43	(.000)
	Sentence × Stress	0.29	(*ns*)	15.17	(.000)	8.75	(.000)
	Sentence × Position	26.05	(.000)	10.23	(.000)	0.27	(*ns*)
	Stress × Position	26.40	(.000)	22.72	(.000)	10.22	(.000)
	Sent. × Stress × Pos.	0.46	(*ns*)	5.49	(.009)	0.14	(*ns*)
Reaction times:	Sentence	2.25	(*ns*)			0.11	(*ns*)
	Stress	22.31	(.000)			58.02	(.000)
	Position	117.36	(.000)			15.63	(.000)
	Sentence × Stress	3.57	(*ns*)			0.04	(*ns*)
	Sentence × Position	2.93	(*ns*)			4.73	(.009)
	Stress × Position	0.08	(*ns*)			1.94	(*ns*)
	Sent. × Stress × Pos.	3.12	(*ns*)			0.75	(*ns*)

Dependent variable	Source of variance	By subject t	(p)	ln transformed t	(p)	By item t	(p)
Phoneme detection:	Sentence type:						
	Semantic/syntactic	5.8	(.000)	19.71	(.000)	13.11	(.000)
	Syntactic/acoustic	5.1	(.000)	12.35	(.000)	9.41	(.000)
	Stress:						
	Semantic	25.4	(.000)	3.29	(.005)	4.02	(.000)
	Syntactic	75.9	(.000)	0.88	(*ns*)	1.10	(*ns*)
	Acoustic	86.7	(.000)	1.61	(*ns*)	3.61	(.000)