

Perceptual contributions of the consonant-vowel boundary to sentence intelligibility<sup>a)</sup>

Daniel Fogerty<sup>b)</sup> and Diane Kewley-Port<sup>c)</sup>

Department of Speech and Hearing Sciences

Indiana University

Bloomington, IN 47405

Running Head: Perceptual contributions of the consonant-vowel boundary

---

- a) Portions of the data were presented at the 154<sup>th</sup> meeting of the Acoustical Society of America (J. Acoust. Soc. Am., 122, 2972) and at the 156<sup>th</sup> meeting of the Acoustical Society of America (J. Acoust. Soc. Am., 124, 2439).
- b) Corresponding author, Electronic mail: [dfogerty@indiana.edu](mailto:dfogerty@indiana.edu)
- c) Electronic mail: [kewley@indiana.edu](mailto:kewley@indiana.edu)

## ABSTRACT

Although research has focused on the perceptual contribution of consonants to spoken syllable or word intelligibility, for sentences, vowels have a distinct perceptual advantage over consonants in determining intelligibility [Kewley-Port, Burkle, and Lee, *J. Acoust. Soc. Am.*, **122**, 2365-2375. (2007)]. The current study used a noise replacement paradigm to investigate how perceptual contributions of consonants and vowels are mediated by transitional information at segmental boundaries. The speech signal preserved between replacements is defined as a glimpse window. In the first experiment, glimpse windows contained proportional amounts of transitional boundary information that was either added to consonants or deleted from vowels. Results replicated a two-to-one vowel advantage for intelligibility at the traditional consonant-vowel boundary and suggest that vowel contributions remain robust against proportional deletions of the signal. The second experiment examined the combined effect of random glimpse windows not locked to segments and the distributions of durations measured from the consonant versus vowel glimpses observed in experiment 1. Results demonstrated that, for random glimpses, the cumulative sentence duration glimpsed was an excellent predictor of performance. Comparisons across experiments confirmed that higher proportions of vowel information within glimpses yielded the highest sentence intelligibility.

PACS numbers: 43.71.Es, 43.71.Gv

## I. INTRODUCTION

All languages share a common feature: the use of both consonant and vowel sounds (Ladefoged, 2001). Intelligibility research has focused on these two fundamental groups of speech sounds, attempting to tease apart the relative perceptual contributions of each. The traditional view of these contributions has been stated clearly, as “it has become almost a commonplace statement in intelligibility testing that most of the information in speech is carried by the consonant sounds” (Owens, Talbot, and Schubert, 1968, pg. 648). More recently, this traditional view has been questioned by Cole, Yan, Mak, Fanty, and Bailey (1996) and Kewley-Port, Burkle, and Lee (2007). Both studies demonstrated an overwhelming advantage of vowels compared to consonants in sentence intelligibility testing using a segment replacement paradigm, also described as “noise replacement”. This paradigm turns successive portions of the speech signal on and off, effectively alternating the speech signal with either silence or noise. The “on” portions of speech are defined as glimpses (Cooke, 2003). Noise replacement has been used previously to examine the specific perceptual contributions of certain speech acoustics contained in consonant or vowel segments (e.g., Kewley-Port et al., 2007), as well as, the perception of interrupted speech (e.g., Miller and Licklider, 1950). While these studies were able to isolate the perceptual contributions of acoustic events present during the segment of interest, it is well known that acoustic cues for segments are distributed across consonant-vowel (C-V) boundaries (as shown for stop consonants, Liberman, Cooper, Shankweiler, and Studdert-Kennedy, 1967).

In the current study, experiment 1 sought to replicate and extend past findings by investigating how the perceptual contributions of consonants and vowels are affected by systematically varying amounts of acoustic information at the C-V boundary. Experiment 2 was

conducted as a control to isolate potential confounds of the noise replacement paradigm, in order to confirm that the perceptual pattern revealed in experiment 1 was a result of spectro-temporal information in segments, and not of particular methodological properties. In addition, experiment 2 compared the perception of randomly ordered glimpses, the preserved portions of speech between replacements, to glimpses locked to specific segments in experiment 1.

Segmentation of the highly dynamic acoustic waveform into discrete units corresponding to consonants and vowels is inherently ambiguous. Particularly problematic is the assignment of formant transitions to either the consonant or vowel segment, as these transitions provide information about both consonants and vowels (Lieberman *et al.*, 1967). Normal coarticulation means that acoustic cues for consonants and vowels largely overlap; the perceptual affects of which may be observed over multiple segments (see Kent and Minifie, 1977). Therefore, it may be that a discrete division between consonants and vowels is largely a convenience (see Ladefoged, 2001). Stevens and his colleagues (summarized by Stevens, 2002) have long advocated that prominent acoustic landmarks may be used to effectively mark different segments within the ambiguous speech signal. Therefore, identifying the C-V boundary is a convenient methodological tool for dividing the acoustic signal into mostly vocalic and mostly consonantal portions. However, the perceptual cues present within the acoustic signals associated with consonants and vowels interact with each other. Cooper, Delattre, Liberman, Borst, and Gerstman (1952) were among the first to explore the relation between perception and acoustic cues in stop consonants, such as the frequency release of the burst and the direction of the second formant transition. They described the perception of these specific cues as dependent upon their positional relationship to the neighboring vowel. Furthermore, the perception of speech sounds is highly dependent upon the acoustic-phonetic context (Miller, 1994). Therefore, the

perception of consonants and vowels and their contributions to sentence intelligibility appears to be the result of distributed and overlapping cues.

The current study uses the fundamental grouping of speech sounds into consonants and vowels to globally explore how the acoustic information contained within these segments contributes to sentence intelligibility. In addition, as the perceptual information of these segments is highly distributed and overlapping, this study investigates how segmental contributions change with the addition or deletion of transitional information between them. The current study was not designed to redefine discrete segmental boundaries, but rather to investigate the relative perceptual contributions of these two primary categories of speech sounds.

#### **A. Previous research on the relative importance of consonants and vowels**

There have been several approaches to understanding the relative contributions of consonants and vowels to speech perception. Studies that have investigated the consonant-to-vowel ratio (i.e., the difference between consonant and vowel intensity levels) have found that amplifying the intensity of the consonant relative to the vowel in CVCs enhances recognition of voiced stops (Freyman and Nerbonne, 1989; Freyman, Nerbonne, and Cote, 1991), which corroborates well with the importance of consonant acoustics for speech intelligibility. However, Sammeth, Dorman, and Stearns (1999) used a different approach that controlled for the intensity of three stop consonants. They maintained the intensity level of the consonant in CV syllables while decreasing the intensity level of the vowel by 6 or 12 dB. Bringing the consonant-to-vowel ratio to 0 dB did not improve identification performance of the consonants. However, in some conditions the consonants were effectively presented in isolation, representing

a maximum attenuation for the vowel. Reintroduction of the vowel produced a large and significant improvement in identification, suggesting that the presence of the vowel was important for the successful identification of the consonants.

Much of the work that has championed the importance of consonant segmental acoustics over that of the vowel has investigated this issue at the level of the syllable structure. Owren and Cardillo (2006) took this one step further by investigating multisyllabic words in isolation. They used a segment replacement paradigm to present consonant-only words or vowel-only words, where segments and formant transitions were replaced by silence. Listeners heard pairs of words and were asked to make two judgments: whether the two words shared the same meaning and whether they were spoken by the same talker. Results demonstrated that consonant-only words yielded higher  $d'$  scores for judging the meaning; vowel-only words had higher scores for judging the talker. Owren and Cardillo (2006) concluded that consonants may be more important for intelligibility, while vowels carry more indexical information about the talker.

Research at the sentence level has revealed a much different picture. In a preliminary report, Cole *et al.* (1996) used a segment replacement paradigm where segments were either replaced by speech-shaped noise, a harmonic complex, or silence. Words repeated correctly were scored. Their results suggested that vowel-only sentences maintained a two-to-one advantage over consonant-only sentences, regardless of the type of segmental replacement. Furthermore, this advantage remained, even after 10 ms was deleted from the onset and offset of the vowels. This two-to-one advantage of vowels was later replicated by Kewley-Port *et al.* (2007) when normal-hearing listeners were presented sentences at 70 dB SPL, as well as for elderly hearing-impaired listeners at a higher presentation level of 95 dB SPL. For young normal-hearing listeners, the 95 dB SPL level resulted in an improvement for recognition of

consonant-only sentences, but participants still maintained significantly better performance for vowel-only sentences. This advantage for vowels is even more remarkable considering that vowels actually comprise at least 10% less of the overall sentence duration than consonants (Ramus, Nespors, and Mehler, 1999).

As perceptual cues are highly distributed throughout the speech signal, it could be that certain subsegmental information plays an important role as well. Lee and Kewley-Port (2009) investigated four types of subsegmental information that presented equal proportions of consonants and vowels in each condition. Segmental onset and offset, as well as transitions and center portions were investigated. These portions of segmental information were based relative to traditional segmental boundaries. Performance based on words repeated correctly was equal across all conditions, indicating no advantage for any type of subsegmental information when equal proportions of consonant and vowel information were maintained. However, while transitional and center portions of segments may make equal perceptual contributions to overall sentence intelligibility, it remains to be seen how these two different types of acoustic signals interact.

The purpose of the current study was to investigate how transitional information at the consonant-vowel boundary as traditionally defined facilitates the perceptual contributions made by consonants and vowels to sentence intelligibility. It has become widely accepted that transitional dynamics at the consonant-vowel boundary contains important information for perception (Lieberman *et al.*, 1967; Strange, 1987; Strange, Jenkins, and Johnson, 1983), but it is less clear how this information relates to the perceptual contributions of consonants and vowels. Our investigation explores how certain types of speech information, namely consonants, vowels, and the transitions between them, interact in sentential contexts.

## **B. Glimpsing speech during noise replacement**

In studies that use noise replacement to isolate perceptual contributions, it may be that preserved perceptual cues in the signal are not the only performance factor, but also how those cues interact with a listener's ability to integrate information across preserved portions of the speech signal. The current study examined these perceptual consequences that are inherent in interrupted speech produced during noise replacement. The portions preserved between replacements are described here as glimpses, taken from the literature on interrupted speech in noise. Cooke (2003) has defined a glimpse as "an arbitrary time-frequency region which contains a reasonably undistorted view of the target signal" (p. 579). In everyday listening conditions, it may be most beneficial for a listener to focus on the glimpses of speech between background noise fluctuations, thereby attending to the preserved speech information rather than trying to extract cues from a noisy signal. The speech intelligibility index (SII) has recently been modified to include consideration of these temporal fluctuations and now provides a good account for speech recognition thresholds in interrupted speech (Rhebergen and Versfeld, 2005). The redundancy of the speech signal makes the use of such glimpses effective in speech perception (Cooke, 2006).

In glimpse studies, in order to maintain the same amount of speech information presented across all conditions there is an acoustic trade-off between interruption rate (i.e., the frequency of available glimpses) and glimpse window width (i.e., the duration of the available speech information). Miller and Licklider (1950) demonstrated that perceptual performance for monosyllabic words is affected nonmonotonically as the glimpse rate is increased and the glimpse window width is decreased with the best performance at a glimpse rate around 10 Hz.

Therefore, properties of glimpses, both in rate and size, may affect perception in noise replacement studies. This may also be related to the type of speech material used, as the perception of speech continuity is maintained at shorter glimpse durations for discourse than for isolated words (Bashford and Warren, 1987) and is associated with increased speech intelligibility when higher-order contextual cues are available (Bashford, Reiner, and Warren, 1992).

In order to investigate the segmental contributions of consonants and vowels independently of glimpse contributions, a second experiment was conducted. Experiment 2 used the same distributions of glimpse window durations measured from consonants and vowels, but randomly reordered them, thereby, creating glimpses with random portions of both consonants and vowels. This allowed for the direct comparison of: 1) any difference in performance due to the different glimpse durations measured from consonants and vowels, and 2) performance when windows were either locked to specific segmental information or when they contained random acoustic cues.

## **II. EXPERIMENT 1: SEGMENTAL CONTRIBUTIONS**

This first experiment was designed to answer some questions raised in previous studies conducted in our laboratory (Kewley-Port *et al.*, 2007; Lee and Kewley-Port, 2009). Results showed that using traditional segmental rules, vowels contributed more than consonants to the perception of sentences (Kewley-Port *et al.*, 2007), and that there appeared to be no advantage of transitional versus quasi-steady-state information when the relative proportion of consonant-to-

vowel segment durations remained equal (Lee and Kewley-Port, 2009). However, given that perceptual information is distributed across the transitional area at the C-V boundary, a major unanswered question concerns the dominance of vowel information obtained in previous studies in relation to the allocation of transitional C-V boundary information. Thus, the present investigation systematically shifted the C-V boundary to examine how transitional information contributes to the perceptual roles of consonants and vowels for sentence intelligibility.

## **A. Listeners**

Twenty young (age 19-30,  $M = 24$ ) normal hearing listeners (6 male, 14 female) were paid to participate in the study. All participants were native speakers of American English and had puretone thresholds no greater than 20 dB HL at octave intervals from 250 to 8000 Hz (ANSI, 1996).

## **B. Stimuli**

### ***1. Speech presented in sentences***

The same 42 sentences used in previous studies (Kewley-Port et al, 2007; Lee and Kewley-Port, 2009) were selected from the TIMIT database (Garfola *et al.*, 1993, [www.ldc.upenn.edu](http://www.ldc.upenn.edu)) to be used in the current study. These sentences were spoken by 21 male and 21 female talkers from the North Midland dialect region that matched the dialect region of daily exposure for the participants in this study (i.e., Indianapolis, Indiana and further north).

Sentences contained an average of 8 words (range 6-10 words) and 33 phonemes (11 vowels, 22 consonants). The C-V boundaries are specified by the TIMIT database. Phonemes were identified by traditional segmental boundaries, although TIMIT first employed CASPAR (Leung and Zue, 1984) automatic segmentation to identify the boundaries, which experienced phoneticians later confirmed (see Zue and Seneff, 1988). The TIMIT segmentation methods included rules such as: 1) finding highly salient and abrupt acoustic changes to mark phoneme boundaries for stops; and 2) dividing formant transitions in half during slow periods of change for liquids (presumably because these transitions provide information regarding both phonemes). Kewley-Port *et al.* (2007) verified the segmental boundaries provided by the TIMIT corpus and added three minor rules appropriate for identifying the vowels and consonants in sentences.

These rules were used in the present study and are noted below using the TIMIT arpabet:

(1) Stop closures (e.g., “bc|”) were combined with the stop (e.g., “b”) and treated as a single consonant.

(2) Syllable final V+[r] as in “deer” were transcribed in TIMIT as a vowel plus the consonant [r]. These V+[r] transcriptions were considered as a single rhotocized vowel in this study.

(3) The glottal stop [q] was a separate consonant in TIMIT, generally realized by silence.

When the glottal stop was transcribed as occurring between two vowels, [VqV], it was treated as part of a vowel string.

For the purposes of the noise replacement paradigm used here and elsewhere (see Kewley-Port *et al.*, 2007), consonant strings were treated as a single unit for replacement, as were vowel strings. Therefore, the experimental manipulation of C-V boundaries was only between consonant and vowel units. The acoustic boundary was shifted by a proportion of the vowel duration (VP); thereby, reassigning vowel transitions, because previous research had

demonstrated relatively low performance for consonant-only stimuli (Kewley-Port *et al.*, 2007). It was expected that manipulating by a proportion of the vowel would yield the most interpretable data by avoiding floor performance. This method allows direct investigation of how transitional information typically assigned to vowel segments influences the perceptual contributions of vowels and consonants.

Two processing strategies each created a different stimulus type. One strategy preserved the consonant information and added some proportion of the vowel transitions (C+VP), replacing the vowel centers with noise. The second strategy preserved only a proportion of the vowel center information, replacing the consonants and remaining vowel transitions at the C-V boundary with noise (V-VP). These processing strategies were manipulated by shifting the C-V boundary by six vowel proportion amounts (2 X 6 design), yielding a total of 12 conditions. Figure 1 shows a schematic of these 12 conditions. The schematic example shows the noise replacement for a single CVC; however, this method was applied to all consonant and vowel units across the entire sentence. The conditions are labeled in two ways on the figure, first according to how the C-V boundary was shifted (e.g., C+5, meaning the original consonant segment plus 5% of the vowel), and second, by an equation of the proportion of the segments preserved (e.g., C+.05V). The rest of this manuscript will use the first label type in referring to these conditions.

-----  
Insert Figure 1 here  
-----

## ***2. Noise presented in sentences***

Segments were replaced by a speech-shaped noise (SSN) generated in MATLAB that was based on a standard long-term-average speech spectrum (ANSI, 1969). The SSN was designed to be flat from 0-500 Hz and had a -9 dB/octave roll-off above 500 Hz. The noise level was presented 16 dB below the average sentence level (21dB below the vowel with loudest RMS). This level was used in a previous study for consonant noise replacement (Kewley-Port *et al.*, 2007) to avoid any potential for temporal masking of consonant segments. Every segment was replaced using a unique noise. Figure 2 displays the waveform of a single CVC, “mean,” excised from an experimental sentence under vowel-only and consonant-only noise replacement conditions with the full waveform as reference.

-----  
Insert Figure 2 here  
-----

### C. Calibration

Similar procedures to Kewley-Port *et al.* (2007) and Lee and Kewley-Port (2009) were administered to verify signal levels. First, scripts were written in MATLAB to verify that all 42 test sentences had similar average RMS levels (i.e., within +/-2 dB). Second, a MATLAB script was used to find the most intense vowel across all sentences and then iterated that vowel to produce a calibration vowel of 4 sec. As these stimuli were previously used to evaluate sentence intelligibility with listeners who had hearing impairment, all sentences and the calibration vowel were filtered by a low-pass FIR filter that was flat to 4000 Hz with a 3 dB cutoff at 4400 Hz, and a 200 dB/octave steep slope in a TDT PF1. Therefore, this allows direct comparison of this

study with previous ones (Kewley-Port *et al.*, 2007; Lee and Kewley-Port, 2009). The sound level for the calibration vowel was set to 100 dB SPL through ER-3A insert earphones in a HA-2.2-cm<sup>3</sup> coupler using a Larson Davis model 2800 sound level meter with linear weighting. Relative to the loudest calibration vowel, the mean of the distribution of the loudest vowels in the other sentences was 95 dB SPL. All sentences and the calibration vowel were passed through a programmable attenuator to present the stimuli at 70 dB SPL, being the nominal level referenced in this study. An additional low-level background noise was continuously presented during testing. The purpose of this noise was to reduce transients between speech and the replacement noise. This background noise was generated by the TDT WG2, and was also low-pass filtered at 4400 Hz. The level of this noise was presented at 50 dB less than the calibration vowel (100 dB SPL) measured using the same equipment described above.

#### **D. Procedure**

Listeners were seated in a sound attenuated booth and listened to sentences presented monaurally to the right ear via ER-3A insert earphones. Test stimuli were controlled by TDT system II hardware connected to a PC computer running a MATLAB stimulus presentation interface. Each listener was instructed regarding the task using verbal and written instruction and completed a familiarization task prior to experimental testing. The familiarization task consisted of six non-experimental sentences. The first two sentences were unprocessed. The last four processed sentences were presented in a random order of vowel-only and consonant-only conditions. The processed version was presented first followed by the full unprocessed version of the same sentence. The listener was requested to repeat aloud all of the words they heard for

both versions of the sentence.

During the experimental testing, each participant was randomly assigned to one of two condition subsets, each of which contained six of the twelve experimental conditions. Each participant heard fourteen sentences per condition (114-115 words; ~462 phonemes). The six experimental conditions included tested both processing strategies (V-VP and C+VP) for three vowel proportions. Sentences were presented in two blocks with different fixed orders of three quasi-randomly assigned conditions each, resulting in participants listening to a total of 84 sentences. Past research demonstrated little benefit of hearing the sentence presented twice, even when presented sequentially (Kewley-Port *et al.*, 2007). Here, the same sentence was never presented in the same block to participants. The possibility of a sentence repetition effect is explored in the results.

The task for each subject was to repeat each sentence aloud. Listeners were encouraged to guess, without regard to whether their responses made logical sentences. No feedback was provided. The experimenter scored responses on paper during the session and digitally recorded them for later reanalysis by an independent scorer (Inter-rater agreement on 10% of responses: 98.8%). Only words repeated exactly correct were scored (e.g., no missing suffixes). For example, a response of “pool” would be judged as incorrect if the target sentence had used the plural form “pools.” In addition, words were allowed to be repeated back in any order to be counted as correct repetitions.

## **E. Results & Discussion**

### ***1. Descriptive analysis***

In the first analysis, possible effects of learning were investigated because each sentence was presented twice to listeners, each time under a different condition. Five sentences that occurred within the first ten sentences presented in block 1 and within the last ten sentences in block 2 were selected for analysis. For each sentence, the average percent correct across each of the participants was calculated. This average score was transformed to a *z*-score for its particular segmental condition. No significant difference in performance was found,  $p = 0.98$ ; normalized scores corresponded to a performance increase of 1.6% for when the sentence was presented a second time. Furthermore, when considering *z*-scores for all of the sentences, there was no significant difference between blocks. These results indicate that listeners' performance remained consistent across trials and were unaffected by hearing a sentence presented a second time in a different condition. Thus, reported results appear reliable due to listeners having stable, repeatable performance and were not affected by procedural learning across the two blocks.

Consonants and vowels were manipulated in different directions to investigate the contributions of transitional information. This specific experimental design required a series of six paired samples *t*-tests to be conducted using Bonferroni corrections for multiple comparisons to compare performance across each of the vowel proportions tested. Overall, results indicated that the vowel-only (V only) condition ( $M = 64.7\%$  words correct) had a significant [ $t(9) = -9.7$ ,  $p < 0.001$ ], two-to-one advantage over the consonant-only (C only) condition ( $M = 30.5\%$  words correct), replicating Cole *et al.* (1996) and Kewley-Port *et al.* (2007). The other five *t*-tests between consonant (C+VP) and vowel (V-VP) conditions for each VP pair were significantly different at VP=5% [ $t(9) = -5.6$ ,  $p < .001$ ], VP=10% [ $t(9) = -5.7$ ,  $p < .001$ ], VP=30% [ $t(9) = 4.2$ ,  $p < .008$ ], and VP=45% [ $t(9) = 14.9$ ,  $p < .001$ ]. However, no significant difference was found when VP=15% [ $t(9) = -1.5$ ,  $p = 0.18$ ], indicating that consonant (C+15) and vowel (V-15)

conditions approached similar intelligibility. Figure 3 shows the data points and trends for the consonant and vowel conditions drawn with equal fit ( $R^2 = 0.98$ ) using least squares regression. A linear function approximated the C+VP conditions, while a cubic function was required to provide the same fit for the V-VP conditions. Thus, C+VP conditions increased linearly with the addition of proportional vowel information. In contrast, V-VP conditions, with a nearly constant function around 55% correct, appeared to remain robust against proportional decreases in information until 30% of the vowel was removed.

-----  
Insert Figure 3 here  
-----

These results suggest that increasing transitional information at C-V boundaries linearly predicts perceptual accuracy when consonants are presented. However, for vowels, this transitional information at the C-V boundary appears to provide little added perceptual benefit, suggesting that the C-V boundary transitions provide information redundant with the contributions from vowel centers, when measuring sentence intelligibility. Similar results were found with CVCs by Strange *et al.* (1983) who suggested that information for the vowel is distributed across the changing acoustics of the syllable. In our study, the resilience of vowel centers remained until 30% of the traditional vowel information at the boundaries was deleted. This corresponded to when intelligibility was on average 56% correct for vowels and 69% for consonants. Note that this 30% proportional C-V shift maximized performance for both consonants and vowels.

## 2. Sentence level analysis

Figure 4 shows on average, how much of the total sentence was preserved across the 12 conditions. The total sentence duration on average was 2.48 seconds. Sentence-level measurements for the original TIMIT boundaries demonstrated that consonants comprised 55% of the total sentence duration, while vowels only comprised 45%. While individual consonants have a shorter duration than vowels, more consonants ( $M = 22$ ) occurred per sentence than vowels ( $M = 11$ ). On average, 74% of consonants occurred in consonant strings, while only 12% of vowels occurred in vowel strings. When 30% of the vowel is removed in the V-30 condition performance is at 56% correct. Performance in the consonant conditions does not surpass this 56% performance level until a 30% shift in the boundary for the C+30 condition, where performance reaches 69% correct. Note that for this shift, 2/3 of the sentence is presented in this consonant condition, compared to only 1/3 presented in the corresponding vowel condition just mentioned. Clearly, the cumulative amount of the sentence presented is not an underlying factor that provides the vowel conditions an advantage. When considering perceptual performance, it is notable that vowels contribute twice as much to sentence intelligibility when the original C-V boundary is used despite preserving less of the overall sentence duration (only 45% for vowels). When considering duration of the presented speech signal, consonants must present twice as much of the total sentence duration to surpass performance of the vowel.

-----  
 Insert Figure 4 here  
 -----

### 3. *Segmental level analysis*

Another way to analyze the results is by each individual segment, regardless of whether it

occurred as part of a string of segments. This analysis corresponds to examining the effects relative to the average segment duration rather than average duration of each glimpse of the speech waveform. Figure 5 displays the percent of words correct relative to the duration of signal information present in the average speech segment calculated separately for each condition. For the V only and C only conditions, this corresponds to the average vowel or consonant duration based on the original TIMIT boundaries. On average, the C only condition provided a shorter segment duration ( $M = 62$  ms) than the V only condition ( $M = 97$  ms). However, for the rest of the conditions, a segment is defined according to the noise replacement paradigm used in this study. For example, a segment in the C+45 condition is the original consonant plus 45% of the vowel duration, with an average duration per segment of 105 ms.

-----  
Insert Figure 5 here  
-----

When the amount of speech information was varied by a proportion of the vowel, intelligibility as a function of segment duration for the consonant and vowel conditions crossed. Specifically, at the segmental level, equal perceptual contributions to sentence intelligibility occurred when the signal duration on average for consonants and vowels was equal to 83 ms. Listener performance was matched at this point for a word identification accuracy of 55% correct. While this absolute value of 83 ms is likely dependent upon factors such as speech rate, this point of perceptual equivalence occurred nearest to our 15% VP shift in both segmental conditions. As previously discussed, there was no statistical difference in the C+15 and V-15 conditions. Thus, while a 30% VP shift of the C-V boundary maximized both consonant and vowel perceptual contributions to sentence intelligibility, a 15% VP shift most successfully (out

of the conditions presented) equalized such contributions.

#### 4. *Glimpse window analysis*

The durations of glimpse windows locked to consonant and vowel segments were measured across all 42 sentences to determine if these durations composed significantly different distributions for the two segmental types (see Figure 6), thereby, providing an opportunity for different perceptual contributions related to glimpse duration. While there was a large overlap in the distributions for consonant ( $M = 113$  ms,  $SD = 67.6$  ms) and vowel ( $M = 102$  ms,  $SD = 56.6$  ms) glimpse windows, there was a significant difference between the distributions [ $t(460) = 2.642, p < 0.01$ ]. Therefore, while results for experiment 1 clearly demonstrate that glimpse windows locked to vowel information were essential for sentence intelligibility, it may be that the results were largely related to glimpse window differences rather than to segmental differences found in the acoustic content of the windows. Experiment 2 was designed to tease apart the independent contributions of segmental and glimpse properties.

-----  
 Insert Figure 6 here  
 -----

### III. EXPERIMENT 2: GLIMPSE CONTRIBUTIONS

The better performance with vowels in experiment 1 may be due to cues inherent to segment glimpse window durations rather than specific segmental acoustic cues provided in the

glimpse (as vowel windows overall were significantly shorter than consonant windows). This may have occurred because performance changes as a function of combined changes in glimpse rate and duration (Miller and Licklider, 1950). Specifically, glimpses of different durations, as noted between the different durations of consonants and vowels, may affect performance in addition to the actual acoustics presented. To investigate this issue, a second experiment was designed to control for specific segmental contributions by ensuring that acoustics from both consonants and vowels were equally likely to occur within a single glimpse. Experiment 2 examined: 1) the differential contributions to sentence intelligibility of glimpse window durations measured from consonants and vowels, and 2) the effect of randomly placed glimpse windows that are asynchronous with segmental information. In addition, these glimpse windows were varied according to a subset of six durational proportions selected from among the 12 conditions of experiment 1.

### **A. Listeners**

Ten young (age 19-25,  $M = 22$ ) normal hearing listeners (3 male, 7 female) were paid to participate in the study. These listeners were different from those who participated in experiment 1. All participants were native speakers of American English and had puretone thresholds no greater than 20 dB HL at octave intervals from 250 to 8000 Hz (ANSI, 1996).

### **B. Stimuli**

The same sentences were used as experiment 1 and differed only in processing strategy.

The same glimpse windows were used within each sentence, preserving both the number and duration of glimpses as calculated from the TIMIT database. The order of these windows was then quasi-randomized such that glimpse boundaries no longer corresponded with segmental boundaries. The quasi-randomization followed three rules. First, the windows alternated between windows measured from consonant units and those measured from vowel units, such that no two windows measured from the same segment type occurred in adjacent intervals; thereby avoiding two glimpse windows combining into a larger one. Second, for sentences that had an unequal number of windows between the two segmental types, the majority of sentences, the shortest or longest window measured from the segmental type with the most windows in that sentence was randomly selected to be presented first. This ensured that glimpse boundaries were maximally offset at the beginning of the sentence. If there was an equal number of consonant and vowel windows, this rule did not apply and the first segment window was selected at random from the opposite segmental type presented first in the sentence (i.e., if the sentence began with a consonant, a vowel window was selected at random). Third, if there were pause windows (i.e., silent intervals) present in the sentence, these windows were presented in the randomized order closest to their original location in the sentence in an attempt to preserve sentence prosody. Without this rule, the segmental window that fell within a silent period in the sentence would convey no acoustic information. For this experiment, it was important to preserve the same amount of speech acoustics as presented in experiment 1, otherwise poor performance could be related to reduced amounts of speech acoustics overall. This possibility was explicitly avoided by this rule.

As with experiment 1, two different processing strategies were used: one that used consonant glimpse window durations and one that used vowel glimpse window durations. The

durations of these windows were manipulated by three of the previous six proportions of the vowel (0%, 15%, 30%) to yield a total of six conditions. This subset of conditions was selected to explore intelligibility at glimpse durations that provided relatively stable performance for vowel conditions. The glimpse windows measured in these conditions corresponded directly to those used in experiment 1 under the identical vowel proportion condition. The only difference here was that glimpse windows were no longer time-locked to specific segmental boundaries. Calibration for this experiment used procedures identical to those in experiment 1.

### **C. Procedure**

Test procedures were identical to those of experiment 1. Each participant completed a familiarization task identical to experiment 1, but with randomized glimpse conditions. During the experimental testing, participants again completed two different blocks of 42 sentences. Each block contained three different conditions in a fixed random order. Sentences were presented once in each block under a different condition with a different randomization of glimpse windows. Participants again repeated aloud each sentence without feedback. The experimenter scored responses on paper during the session and digitally recorded them for later reanalysis by an independent scorer (Inter-rater agreement on 10% of responses: 96.5%).

### **D. Results & Discussion**

#### ***1. Glimpse window analysis***

As with experiment 1, the effect of practice resulting from procedural learning and

sentence repetition was examined to determine the reliability of participant responses. Analysis of five sentences that were presented within the first ten sentences of block 1 and the last ten sentences of block 2 were compared. No significant difference in z-scores was observed ( $p = 0.71$ ). The z-scores for these two presentations corresponded to a 3.1% increase in scores for the five sentences at the end of the second block. Thus, results for this second experiment reliably represented perceptual intelligibility under this processing strategy.

Figure 7 displays the results of experiment 2 according to how the distributions of glimpse duration were shifted by vowel proportion. All condition labels for experiment 2 are appended with an initial “g” to indicate that they only preserve the duration of speech within a glimpse and do not contain specific segmental information. Results were analyzed by a series of Bonferroni-corrected t-tests conducted between glimpses measured from consonants and vowels at each of the three vowel proportions tested. T-tests indicated significant difference between glimpse windows of gC+VP and gV-VP conditions at 15% [ $t(9) = 9.4, p < .001$ ] and 30% [ $t(9) = 21.8, p < .001$ ] VP pairs, but not at the original C-V boundary (0% VP), [ $t(9) = 1.7, p = 1.88$ ]. Thus, the two-to-one advantage obtained with segmental windows was not observed between randomly presented windows measured from consonants versus those from vowels. These results suggest that the large difference between glimpses locked to segmental information in experiment 1 was a result of perceptual cues contained in the segmental acoustics, not characteristics inherent in the distributions of glimpse windows. Thus, the initial conclusions of experiment 1 are validated: the perceptual cues contained within vowel segments do contribute more to sentence intelligibility than those within consonants segments.

-----  
 Insert Figures 7 & 8 here

-----

Figure 8 displays the six random glimpse conditions (black) plotted according to the cumulative duration of the sentence presented. As a reference, results from the corresponding conditions of experiment 1 (gray) are also plotted. A linear trend (solid line) was fit to the six glimpse conditions tested ( $R^2 = 0.93$ ). In fact, when glimpses were not locked to segmental information, the percent of the sentence presented closely predicted performance, particularly for the four longer values of the sentence presented. That is, given the percent of the sentence presented at 45, 55, 62, and 69%, the percent of the words identified was predicted by those values within 1-5%. While such close predictions did not hold for the two shorter durations (gV-15 and gV-30), overall, the predictability of word accuracy from only the duration of the sentence presented was remarkably high.

## ***2. Comparison of results with experiment 1***

An analysis was conducted to examine the difference in experimental results across the two experiments. Bonferroni-adjusted t-tests between experiment 1 and 2 for the six paired conditions (see Figure 8) demonstrated significant differences for four conditions tested, but not the two longest glimpse window conditions, C+15 [ $t(9) = 2.0, p = 0.08$ ] and C+30 [ $t(9) = 1.3, p = 0.24$ ]. Of the four conditions that were significantly different, three conditions [V-30,  $t(9) = -12.8, p < 0.001$ ; V-15,  $t(9) = -8.5, p < 0.001$ ; and V only,  $t(9) = -7.8, p < 0.001$ ] yielded higher performance in experiment 1 with segmentally locked glimpses that contained only vowel information. Performance was poorer with randomly placed glimpses that only contained partial vowel information. This benefit of vowel segmental information is clearly seen in all three comparisons (V only, V-15 and V-30), but especially at the shortest glimpse duration, V-30. At

this duration, providing glimpses of only vowel information in the segmental condition of experiment 1 provides a perceptual advantage over the random glimpses of experiment 2 by a factor of 4. The fourth significant condition, C only [ $t(9) = 12.6, p < 0.001$ ], is a special case where segmentally locked glimpses in experiment 1 contained no vowel information. For this condition, performance was significantly better with randomized glimpses, presumably because of the inclusion of partial vowel information. Therefore, our interpretation of the results for the four conditions is as follows: significant differences between segmental and random glimpses from the two experiments can be explained by increased amounts of preserved vowel information. Considering the similar performance of the two longest glimpse windows, these conditions (C+15 and C+30) contained the most segmental information distributed over both consonants and vowels. For these last two conditions, random glimpses apparently provided the same amount of perceptual information as is provided when larger amounts of vowel information were added to C only information in experiment 1. Therefore, these glimpse windows are long enough for the combined acoustics of both consonants and vowels to contribute to the perception of locked or randomized glimpse windows. In summary, the presence of more vowel information in glimpses yields higher perceptual performance. Vowel information is particularly important when less of the sentence is presented.

#### **IV. GENERAL DISCUSSION**

##### **A. Segmental contributions to sentence intelligibility**

This study investigated the perceptual contributions of vowels, consonants, and the

transitions between them, using noise replacement with locked or random glimpses. The results from this study restate the importance of vowel information to sentence intelligibility as described by Cole *et al.* (1996) and Kewley-Port *et al.* (2007). Vowels as traditionally defined carry important perceptual cues that are not found in consonants. Furthermore, even truncated portions of vowels contribute strongly to sentence intelligibility despite providing much less of the overall sentence duration than consonants. Perceptual cues are distributed across the entire vowel, such that transitions into the vowel provide little additional perceptual benefit above what is provided by the vowel center alone, when the vowel center contains 70% of the total vowel. However, these dynamic transitions do provide additional perceptual information to consonants which significantly improve sentence intelligibility. Such perceptual information in the “vowel” transitions is not found in traditionally defined consonants, as indicated by the relatively low consonant-only sentence scores. Our results correspond well with the literature that has identified C-V transitions (Liberman *et al.*, 1967) and vowel context (Cooper *et al.*, 1952) as important cues for consonant recognition, and with evidence from silent center vowels that demonstrated the importance of dynamic information for vowel recognition (Strange, 1987; Strange *et al.*, 1983).

Owren and Cardillo (2006), using a similar replacement technique with silence rather than noise, found consonants are more important to the intelligibility of isolated words. However, it may be that vowels gain greater perceptual importance from an increased amount of contextual or stylistic factors present in sentential contexts. In sentences, speech acoustics are characterized by shorter individual segments and greater coarticulation. Thus, individual consonant segments may become less distinct, requiring more transitional dynamics for recognition. This blurring of segmental boundaries results in words, originally intelligible in

sentences, becoming largely unintelligible when excised (Craig and Kim, 1990). Clearly, listeners take advantage of coarticulatory and prosodic cues in the perception of fluent sentences. In fact, some speech recognition methods explicitly define coarticulatory events between words in order to achieve higher recognition performance for continuous sentences (Giachin, Rosenberg, and Lee, 1991).

The most sonorant segments in speech, namely the vowels, are most capable of carrying prosodic information such as pitch contours or stress. Prosody spans linguistic levels (Kent and Read, 1992), providing an opportunity for multiple perceptual and cognitive levels involved in sentence recognition to interact. For example, prosody may provide information about syntactic structure through cues such as phrase-final lengthening. This particular prosodic cue, tied to the last stressed syllable in a major syntactic phrase, is perceptually available for listeners to use in recognizing the structure of speech (Read and Schreiber, 1982). Vowels may be particularly useful in conveying these cues for syntactic structure, as stress patterns temporally align with the onset of vowel energy (Tajima and Port, 2003). Indeed, Toro, Nespor, Mehler, and Bonatti (2008) demonstrated that participants were able to extract syntactic cues from vowels, but not consonants. Furthermore, they support a functional dissociation of consonants and vowels, with consonants cuing the lexicon and vowels cuing syntactic structure. It may be that the syntactic cues that vowels uniquely carry facilitate top-down processes and constrain lexical activation.

Therefore, there are many types of information that vowels may carry to provide them with an advantage over consonants during the perception of compromised sentences containing only partial information. These cues encompass acoustic, coarticulatory cues important for consonant recognition, but may also involve perceptual cues that provide higher-order mechanisms to predict information, even though many explicit acoustic cues for segment

recognition may be missing. Future studies will have to tease apart potential contributions from the complex acoustic, perceptual, and cognitive events that occur in sentences, which we as listeners navigate with relative ease in every day conversational contexts.

## **B. Glimpse contributions to sentence intelligibility**

Even though there was a significant difference between the durational distributions of consonant and vowel glimpses, no differential performance was found for perceptual intelligibility of sentences when using these glimpse windows in a randomly ordered replacement paradigm. Instead, the amount of the speech signal presented, distributed over the length of the entire sentence, predicted perceptual performance. This result is consistent with Iyer, Brungart, and Simpson (2007) who used periodic glimpses to investigate the effect of noise and speech maskers. Above an 8 Hz interruption rate, the trade-off between the glimpse window duration and number of glimpses does not have much of an effect, as long as the cumulative sentence duration glimpsed (i.e., the total duration of preserved speech information) remains constant. Also consistent with the current study, Li and Loizou (2007) found that the cumulative duration of speech glimpsed predicts performance rather than glimpse window duration. By comparing performance across several glimpse conditions that preserved different frequency regions of speech, Li and Loizou (2007) also identified the frequency region of 0-3 kHz as the most important frequency region for glimpsing. This frequency region corresponds to the F1 and F2 frequency region most important for vowel perception. Therefore, this evidence provides further support for the important contribution of vowel information to sentence intelligibility.

Experiment 2 investigated the combined perceptual effect of the distributional properties and random (i.e., aperiodic) placement of glimpse windows. In a pilot study, we separately

assessed the contributions of overall sentence duration presented while avoiding differences in the distributions of consonants versus vowels. The pilot used periodic replacement based on the duty cycle that was the average of the glimpse windows (215 ms). For the consonant glimpse condition, the on phase was 113 ms, the mean glimpse window size for consonants, while the off phase was 102 ms, the mean glimpse size for vowels. For the vowel glimpse condition, the on and off phases were reversed (e.g. the on phase was the average vowel window duration of 102 ms). These on/off phases were manipulated by a proportion of the vowel as in the experiments presented here to provide a direct comparison. The pilot for all the periodic replacement conditions yielded very similar results to those obtained in experiment 2 with aperiodic replacement, further suggesting that glimpse perception is predicted by the cumulative duration of the sentence presented. These results are consistent with Miller and Licklider (1950) who found that perception of monosyllabic words during random, aperiodic interruption was effectively the same as during periodic interruption.

Overall, there is a significant difference between the segmentally locked glimpses in experiment 1 and the random, segmentally asynchronous glimpses used in experiment 2. There is a significant advantage for glimpses that capture specific spectro-temporal information contained in vowel segments.

## **V. CONCLUSIONS**

This study used a fundamental division in speech sound categories, specifically between consonants and vowels, to investigate contributions to sentence intelligibility. The present study

replicated results from Cole *et al.* (1996) and Kewley-Port *et al.* (2007) that demonstrated a two-to-one advantage for sentences containing only vowels compared to sentences that contained only consonants. Furthermore, this study investigated how these contributions were mediated by transitional information into the vowel. This transitional information appears to be redundant with information at the vowel center, suggesting a similar result to Lee and Kewley-Port (2009) who found no difference in transitional versus quasi-steady-state information. However, increased transitional information did provide a linear benefit relative to consonant-only sentences, demonstrating that critical perceptual information is present in the transition. Finally, this study demonstrated that glimpse window distributions did not differentiate performance between consonant-only and vowel-only sentences. Instead, the cumulative amount of the sentence present and the amount of vowel information preserved served as predictors for listener performance. This study reiterates the importance of vowels to sentence intelligibility, as well as how dynamic transitional information at the C-V boundary significantly improves the perceptual contributions of consonants. Simply put, vowel information is essential for sentence intelligibility, particularly when less of the overall sentence duration is presented.

## **VI. ACKNOWLEDGEMENTS**

This work was supported by NIH-NIDCD Training Grant No. T32-DC00012 and R01-02229 to Indiana University.

## REFERENCES

- ANSI (1969). "American National Standard Methods for Calculation of the Articulation Index," ANSI S3.5-1969, American National Standards Institute, New York, NY.
- ANSI (1996). "Specifications for audiometers," ANSI S3.6-1996, American National Standards Institute, New York, NY.
- Cole, R., Yan, Y., Mak, B., Fanty, M., and Bailey, T. (1996). "The contribution of consonants versus vowels to word recognition in fluent speech", Proceedings of the ICASSP'96, 853–856.
- Cooke, M. (2003). "Glimpsing speech", J. Phonetics **31**, 579-584.
- Cooke, M. (2006). "A glimpsing model of speech perception in noise", J. Acoust. Soc. Am. **95**, 1562-1573
- Cooper, F., Delattre, P., Liberman, A., Borst, J. and Gerstman, L. (1952). "Some experiments on the perception of synthetic speech sounds", J. Acoust. Soc. Am. **24**, 597–606.
- Craig, C. H. and Kim, B. W. (1990). "Effects of time gating and word length on isolated word recognition performance", J. Speech Hear. Res. **33**, 808-815.
- Freyman, R. L. and Nerbonne, G. P. (1989). "The importance of consonant-vowel intensity ratio in the intelligibility of voiceless consonants", J. Speech Lang. Hear. Res. **32**, 524-535.
- Freyman, R. L., Nerbonne, G. P. and Cote, H. A. (1991). "Effect of consonant-vowel ratio modification on amplitude envelope cues for consonant recognition", J. Speech Lang. Hear. Res. **34**, 415-426.
- Garofolo, J., Lamel, L., Fisher, W., Fiscus, J., Pallett, D., and Dahlgren, N. (1990). "DARPA

TIMIT acoustic-phonetic continuous speech corpus CDROM”, National Institute of Standards and Technology, NTIS Order No. PB91-505065.

Giachin, E. P., Rosenberg, A., E. and Lee, C-H. (1991). “Word juncture modeling using phonological rules for HMM-based continuous speech recognition”, *Comput. Speech Lang.* **5**, 155-168.

Iyer, N., Brungart, D., and Simpson, B. (2007). “Effects of periodic masker interruption on the intelligibility of interrupted speech”, *J. Acoust. Soc. Am.* **122**, 1693-1701.

Kent, R. and Minifie, F. (1977). “Coarticulation in recent speech production models”, *J. Phonetics* **5**, 115-133.

Kent, R. D. and Read, C. (1992). *The Acoustic Analysis of Speech*. (Singular, San Diego).

Kewley-Port, D., Burkle, T. Z. and Lee, J. H. (2007). “Contribution of consonant versus vowel information to sentence intelligibility for young normal-hearing and elderly hearing-impaired listeners”, *J. Acoust. Soc. Am.* **122**, 2365-2375.

Ladefoged, P. (2001). *Vowels and Consonants: An introduction to the sounds of languages*. (Blackwell, Oxford).

Lee, J. H., and Kewley-Port, D. (2009). “Intelligibility of interrupted sentences at subsegmental levels in young normal-hearing and elderly hearing-impaired listeners”, *J. Acoust. Soc. Am.* **125**, 1153-1163.

Leung, H.C. and Zue, V.W. (1984). “A procedure for automatic alignment of phonetic transcriptions with continuous speech.” *Proc. ICASSP’84*, 2.7.1-2.7.4.

Li, N. and Loizou, P. (2007). “Factors influencing glimpsing of speech in noise”, *J. Acoust. Soc. Am.*, **122**, 1165-1172.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., and Studdert-Kennedy, M. (1967).

- “Perception of the speech code”, *Psychol. Rev.* **74**, 431–461.
- Miller, G. and Licklider, J. (1950). “The intelligibility of interrupted speech”, *J. Acoust. Soc. Am.* **22**, 167-173.
- Miller, J. L. (1994). “On the internal structure of phonetic categories: a progress report”, *Cognition* **50**, 271-285.
- Owens, E., Talbot, C. B. and Schubert, E. D. (1968). “Vowel discrimination of hearing-impaired listeners”, *J. Speech Hear. Res.* **11**, 648-655.
- Owren, M. J., and Cardillo, G. C. (2006). “The relative roles of vowels and consonants in discriminating talker identity versus word meaning”, *J. Acoust. Soc. Am.* **119**, 1727–1739.
- Ramus, F., Nespors, M. and Mehler, J. (1999). “Correlates of linguistic rhythm in the speech signal”, *Cognition* **75**, AD3-AD30.
- Read, C. and Schreiber, P. A. (1982). “Why short subjects are harder to find than long ones,” in *Language Acquisition: The state of the art*, edited by E. Wanner and L. Gleitman (Cambridge University Press, Cambridge).
- Rhebergen, K.S. and Versfeld, N.J. (2005). “A Speech Intelligibility Index-based approach to predict the speech reception threshold for sentences in fluctuating noise for normal-hearing listeners,” *J. Acoust. Soc. Am.* **117**, 2181-2192.
- Sammeth, C. A., Dorman, M. F., and Stearns, C. J. (1999). “The role of consonant-vowel intensity ratio in the recognition of voiceless stop consonants by listeners with hearing impairment”, *J. Speech Lang. Hear. Res.* **42**, 42–55.
- Stevens, K. N. (2002). “Toward a model for lexical access based on acoustic landmarks and distinctive features”, *J. Acoust. Soc. Am.* **111**, 1872-1891.

- Strange, W. (1987). "Information for vowels in formant transitions", *J. Acoust. Soc. Am.* **26**, 550-557.
- Strange, W., Jenkins, J. J., and Johnson, T. L. (1983). "Dynamic specification of coarticulated vowels", *J. Acoust. Soc. Am.* **74**, 695-705.
- Tajima, K. and Port, R. (2003). "Speech rhythm in English and Japanese". In J. Local, R. Ogden and R. Temple (Eds.) *Papers in Laboratory Phonology VI*, 317-334. (Cambridge University Press, Cambridge).
- Toro, J. M., Nespor, M., Mehler, J., and Bonatti, L. L. (2008). "Finding words and rules in a speech stream: Functional differences between vowels and consonants", *Psychol. Science* **19**, 137-144.
- Zue, V. W. and Seneff, S. (1988). "Transcription and alignment of the TIMIT database", *Proceedings of the Second Meeting on Advanced Man-Machine Interface through Spoken Language*, 11.1-11.10.

## FIGURE CAPTIONS

FIG. 1. Schematic of noise replacement conditions depicted for a single CVC within the sentence. The C only and V only (for C+0 and V-0 conditions) uses the original TIMIT C-V boundaries.

FIG. 2. Waveforms demonstrating two noise replacement conditions and the full waveform of the word “mean”. V only = vowel-only, C only = consonant-only.

FIG. 3. Results for the 12 experimental conditions in experiment 1 plotted according to the vowel proportion by which the C-V boundary was shifted. The original TIMIT C-V boundary is at 0% VP, marked by the dashed vertical line. Error bars display the standard error of the mean.

FIG. 4. The average amount of the sentence preserved across the 12 conditions with the rest of the sentence replaced by noise. The amount presented was distributed over the entire sentence according to the specific noise replacement condition. The length of the average sentence was 2.48 seconds.

FIG. 5. The average duration of an individual segment in each condition (see text for details).

FIG. 6. Glimpse window distributions measured from all segments which occurred across all 42 sentences for a) vowel glimpse windows, and b) consonant glimpse windows.

FIG. 7. Results for the 6 experimental conditions in experiment 2 plotted according to the vowel proportion by which the C-V boundary was shifted. Glimpse window durations based upon the original TIMIT C-V boundary are at 0% VP, marked by the dashed vertical line. Glimpse windows measured from the consonant (black diamonds) and vowel (grey squares) conditions of experiment 1 are coded similarly to previous figures but now with the prefix “g” to indicate random glimpses. Error bars display the standard error of the mean.

FIG. 8. The percent of words correct is displayed for the 6 conditions of experiment 2 (black) and the comparison conditions of experiment 1 (gray) as a function of the cumulative duration of the sentence preserved over the entire sentence. The black regression line is only across the six random glimpse conditions. Asterisks (\*) indicate a significant difference between experiment 1 and 2 results,  $p < 0.001$ .



Figure 2

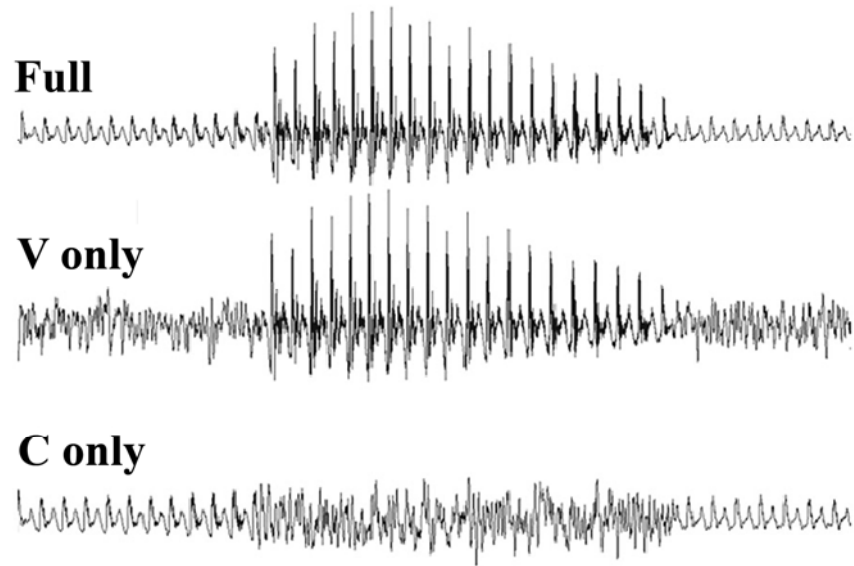


Figure 3

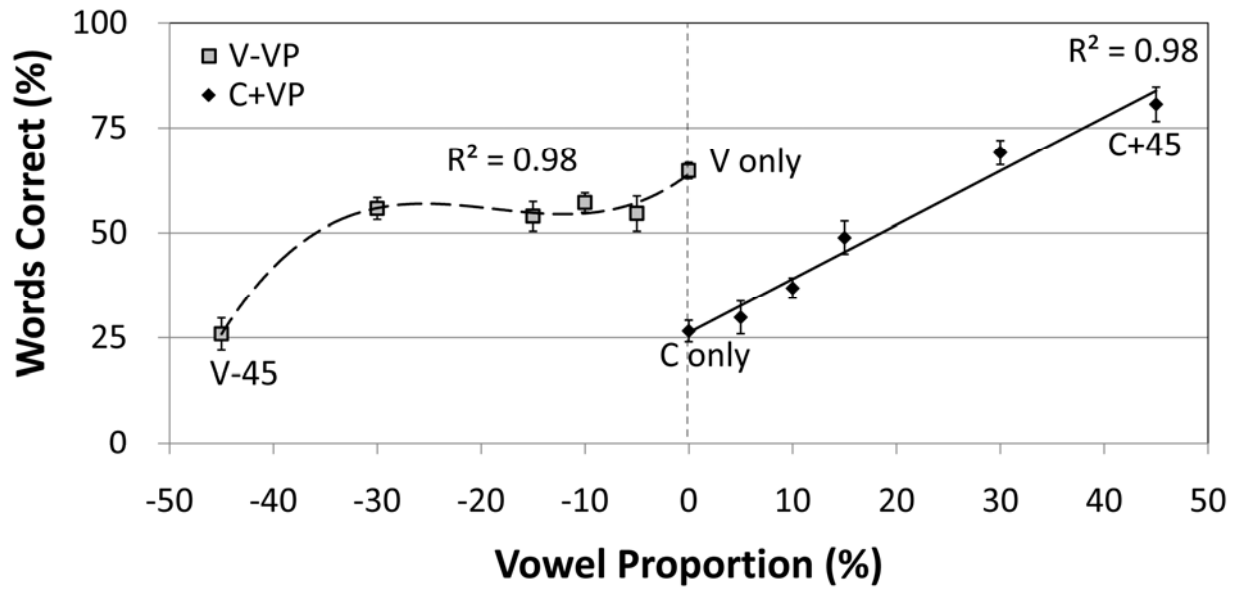


Figure 4

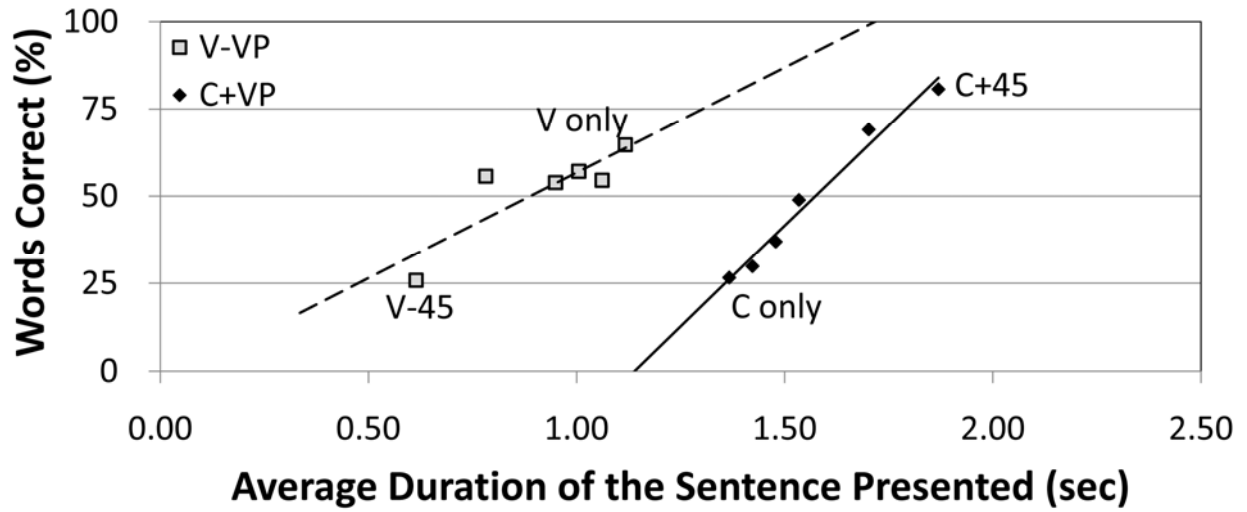


Figure 5

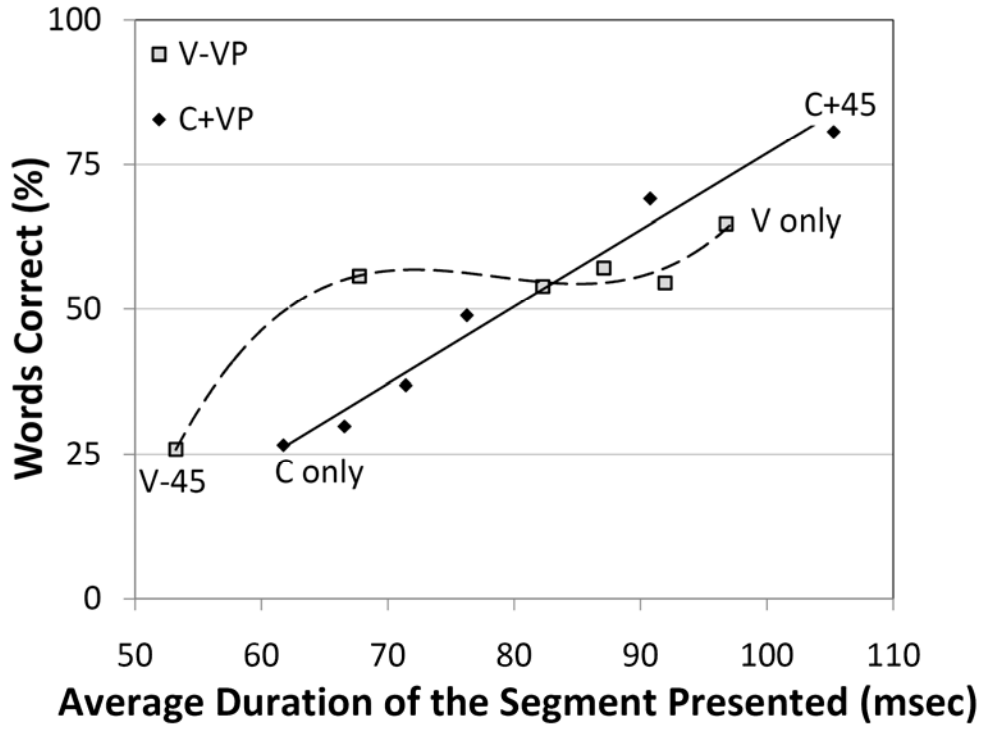


Figure 6

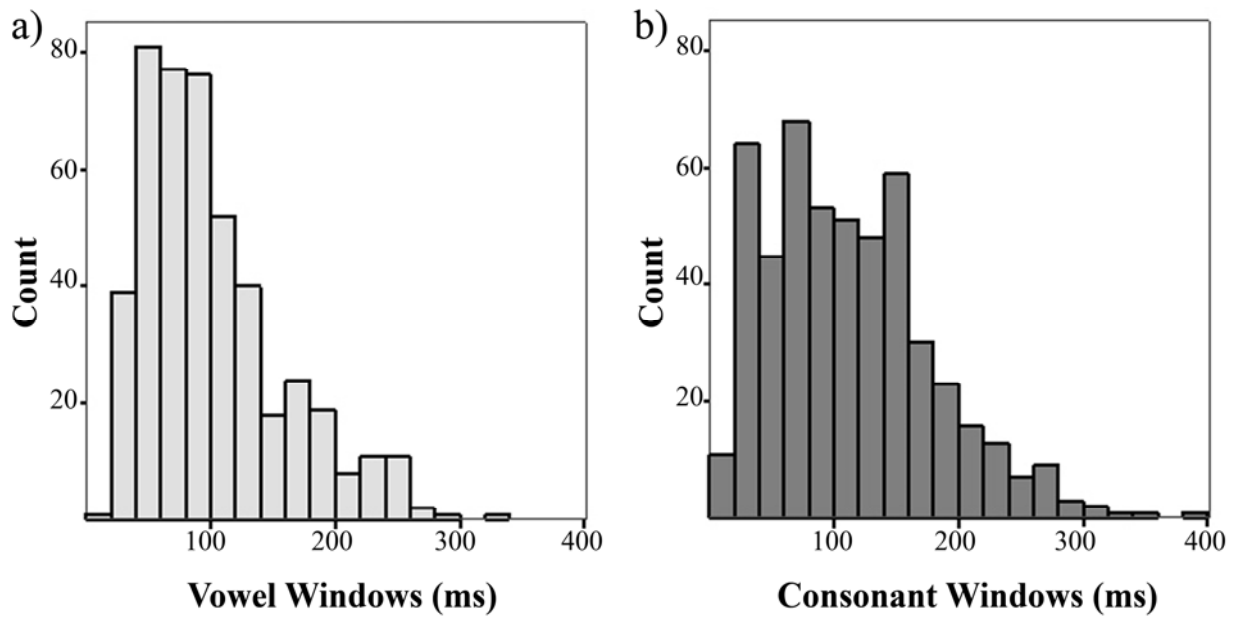


Figure 7

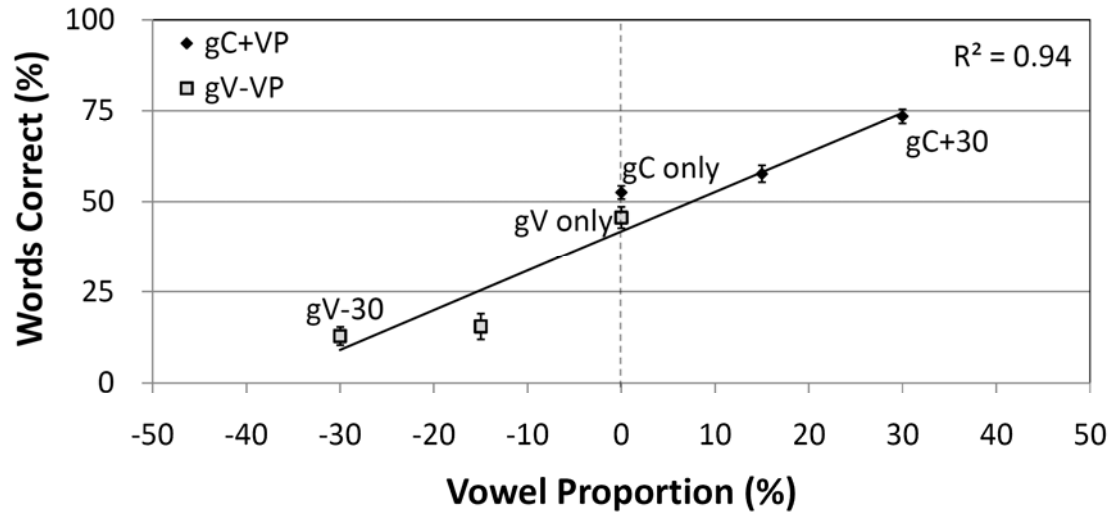


Figure 8

