Perceptual contributions to monosyllabic word intelligibility: Segmental, lexical, and noise replacement factors^{a)}

Daniel Fogerty^{b)} and Larry E. Humes

Department of Speech and Hearing Sciences, Indiana University, Bloomington, Indiana 47405

(Received 4 June 2010; revised 17 August 2010; accepted 31 August 2010)

This study investigated the relative contributions of consonants and vowels to the perceptual intelligibility of monosyllabic consonant-vowel-consonant (CVC) words. A noise replacement paradigm presented CVCs with only consonants or only vowels preserved. Results demonstrated no difference between overall word accuracy in these conditions; however, different error patterns were observed. A significant effect of lexical difficulty was demonstrated for both types of replacement, whereas the noise level used during replacement did not influence results. The contribution of consonant and vowel transitional information present at the consonant-vowel boundary was also explored. The proportion of speech presented, regardless of the segmental condition, overwhelmingly predicted performance. Comparisons were made with previous segment replacement results using sentences [Fogerty, and Kewley-Port (2009). *J. Acoust. Soc. Am.* **126**, 847–857]. Results demonstrated that consonants contribute to intelligibility equally in both isolated CVC words and sentences. However, vowel contributions were mediated by context, with greater contributions to intelligibility in sentence contexts. Therefore, it appears that vowels in sentences carry unique speech cues that greatly facilitate intelligibility which are not informative and/or present during isolated word contexts. Consonants appear to provide speech cues that are equally available and informative during sentence and isolated word presentations.

© 2010 Acoustical Society of America. [DOI: 10.1121/1.3493439]

PACS number(s): 43.71.Es, 43.71.Gv [AA]

Pages: 3114–3125

I. INTRODUCTION

Segmental contributions to the perceptual intelligibility of speech have recently received increased attention in the literature. Whereas it is clear that consonant and vowel acoustics are highly overlapping and distributed, an investigation of the predominantly consonantal or vocalic intervals of speech can inform us as to how listeners use the basic acoustic cues present during these segments to understand speech. Rather than investigating how segments contribute to vowel or consonant identification, Experiment 1 pursues how the dynamic acoustic cues present during these segments contribute generally to word intelligibility. Furthermore, in Experiment 2 this study directly manipulates the amount of transitional acoustics associated with the consonant or vowel acoustics to determine how segmental contributions are mediated by these boundary acoustics.

The traditional view of segmental contributions was summarized 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” (e.g., Owens *et al.*, 1968, p. 648). However, several recent studies have demonstrated that vowels are the most important segment for the perceptual intelligibility of sentences (Cole *et al.*, 1996; Fogerty and Kewley-Port, 2009; Kewley-Port *et al.*, 2007). These studies examined sentences using a

noise replacement paradigm to present consonant-only sentences (which had all of the vowels deleted and replaced by noise) or vowel-only sentences (the consonants having been removed and replaced by noise).

It is important to note that such a segment replacement method does not remove all acoustic information involved in the identification of that segment. That is, removing all consonant segments does not remove all acoustic information available to identify consonants. For example, it has long been known that the perception of stop consonants is influenced by the positional relationship of their acoustic cues to the neighboring vowel (Cooper *et al.*, 1952) and that transitional information between segments provides information about both vowels and consonants (Liberman *et al.*, 1967). However, the segment replacement paradigm does vary different types of acoustic information that are generally associated with one of the two segmental categories. For example, periodicity and a clear formant structure are generally associated with vowels whereas transient acoustic events related to vocal tract constriction are more often associated with consonants (Stevens, 2002) and have been used as cues for segmentation (e.g., Li and Loizou, 2008). Stevens has long argued for acoustic landmarks identifying segmental units, which the segmental replacement method explicitly removes (see Stevens, 2002). Semivowels such as /l/ or /w/ have either been investigated as a separate category (see Cole *et al.*, 1996) or included in the consonant category based on linguistic descriptions (see Kewley-Port *et al.*, 2007; Owren and Cardillo, 2006). Vowels and consonants are often cited as the most fundamental distinction between

^{a)} Portions of the data were presented at the 2009 ASHA Convention, New Orleans, LA.

^{b)} Author to whom correspondence should be addressed. Electronic mail: dfogerty@indiana.edu

speech sounds (Ladefoged, 2001; Stevens, 2002). By removing the most predominant consonant or vowel acoustic cues, it is possible to determine the general role of these acoustic cues to overall sentence or word intelligibility. In an additional phonemic error analysis, this study also explores contributions of the segmental acoustics to individual phoneme identification. Furthermore, Experiment 2 explicitly investigated the contribution of transitional cues at the consonant-vowel (C-V) boundary.

In a preliminary report, Cole *et al.* (1996) used a segment replacement paradigm to present consonant-only and vowel-only sentences. They replaced segments with speech-shaped noise, a harmonic complex, or silence. Across all types of segmental replacement, their results suggested that vowel-only sentences maintained a two-to-one advantage in word identification over consonant-only sentences. Furthermore, this advantage remained, even after 10 ms was deleted from the onset and offset of the vowels.

This word-identification advantage for vowel acoustic information was later replicated for normal-hearing and elderly hearing-impaired listeners (Kewley-Port *et al.*, 2007). In addition, the perceptual contributions of vowels in sentences appear to remain robust, even after 30% of the transitional vowel information is removed (Fogerty and Kewley-Port, 2009). Thus, the greater contribution of vowels in sentence contexts is a strong result, especially given that vowels in totality actually comprise at least 10% less of the overall sentence duration than consonants (Fogerty and Kewley-Port, 2009; Ramus *et al.*, 2000).

A very different picture was demonstrated using multisyllabic words. Owren and Cardillo (2006) used a similar segment replacement paradigm with the exception that they used silence for replacement. Owren and Cardillo (2006) explicitly removed all formant transitions from both consonant and vowel conditions. 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, but vowel-only words had higher scores for judging the talker. Thus, consonants appear most important for the lexical access of isolated words necessary for semantic discriminations.

While these studies are very different in the stimulus processing and methodology, they suggest that the contributions of consonants and vowels may be dependent upon the linguistic or acoustic context. Namely, vowels carry acoustic cues that are more informative for constraining lexical access in sentences and not isolated words. The current study directly addressed this question by replicating the methods of Fogerty and Kewley-Port (2009), to determine how segmental acoustics contribute to the open-set recognition of isolated words. In addition, the current study addressed a number of other questions related to the differences with the Owren and Cardillo (2006) study. These were the importance of consonants for lexical access and the influence of the noise level used during replacement.

Several researchers have suggested that consonant acoustics are most important for lexical access, while vowel

acoustic cues are important for perceiving indexical properties of the talker (e.g., Owren and Cardillo, 2006; Toro *et al.*, 2008a). In reading tasks, consonants, but not vowels, are involved in lexical priming (New *et al.*, 2008) and produce greater influences on event-related potentials during lexical decisions (Carreiras *et al.*, 2009). In addition, changing the vowel in pseudowords increases brain activation in prosodic-related areas while changing the consonant produces changes in areas associated with go-responses, which is presumed to be due to insufficient lexico-semantic processing (Carreiras and Price, 2007). These reading tasks are fundamentally different from spoken language processing, which includes the added complexity of supra-linguistic acoustic cues. However, they do point to dissociated contributions of consonants and vowels. The large advantage demonstrated for vowel acoustic cues in spoken sentences brings into question the predominant role of consonants in lexical processing during spoken word recognition.

Are the acoustic cues of vowels involved in the lexical access of individual words? Or, in contrast, do vowels only provide higher level structure to the global perception of sentences for constraining lexical selection rather than directly accessing it? To investigate this question, the lexical difficulty of words was varied in the current study using the Neighborhood Activation Model (Luce and Pisoni, 1998). This model states that the difficulty of lexical access is determined by a word's frequency of occurrence in the language and the number of phonemic neighbors that it has (determined by adding, deleting, or changing one phoneme). Words that occur frequently in the language and have few neighbors (i.e., lexical competitors) are recognized easier, as indexed by speed and accuracy, than words that are less frequent and have many neighbors. If consonant acoustics are more involved in lexical access, then consonant-only presentations of words should be mediated more by differences in the lexical difficulty of items.

Recall that Owren and Cardillo (2006) used silence instead of noise during the replacement interval. Bashford *et al.* (1992) investigated the effect of noise level, including silence, during interruption studies. They found that the level of replacement does not influence the perception of isolated words. However, as Owren and Cardillo (2006) posit, it is possible that the noise level may have a differential effect on vowel as compared to consonant replacement. Therefore, this study systematically varied the level of the replacement noise. A specific result of interest was if high levels of noise interfered more with consonant-only presentations due to masking and facilitated vowel-only presentations due to phonemic restoration.

Finally, as stated in the opening paragraph, vowel and consonant acoustics do not occur as isolated events, but overlap in time. Therefore, any segmentation strategy to isolate individual vowels will ultimately contain information about the identity of neighboring consonants and vice versa. This study investigated this issue through two complementary means. First, phoneme scoring allows for the measurement of segment identification accuracy when it was removed by the replacement paradigm. Therefore, this measurement can be used as an index of how much information isolated vow-

els or consonants contain about their missing neighbors. Second, Experiment 2 was conducted to directly manipulate the amount of transitional acoustics assigned to either the presented consonant or vowel acoustics. Fogerty and Kewley-Port (2009) found, in sentences, that transitional information into the vowel facilitates intelligibility of consonant-only sentences, but does not provide additional benefit for vowel-only sentences above what is already contained in the vowel center. The current study investigated contributions of the transitional information into both the consonant and the vowel to overall word intelligibility.

This study presented listeners with monosyllabic consonant-vowel-consonant (CVC) words. These CVC words preserve the natural durational differences between consonants and vowels across the utterance, which facilitates comparison of this study with our previous work with sentences (Fogerty and Kewley-Port, 2009). In the sentences used by Fogerty and Kewley-Port, consonants constituted 55% of the utterance, while vowels comprised 45%. Close agreement was made for the CVC words used here, with 54% and 46% of the utterance presented by consonants and vowels respectively. This agrees with measurements of English in general (Ramus *et al.*, 2000). Consonants also occur twice as often in sentences, which the CVC syllable structure preserves. While previous visual studies have balanced the number of vowels and consonants (see New *et al.*, 2008), the current auditory study used stimuli that matches the acoustics of consonants and vowels as they occur in natural spoken language. The use of monosyllabic words also minimizes additional supra-linguistic cues that modify acoustic cues between syllables. This being said, it is recognized that the stimuli used here represent a limited sampling of speech.

In summary, the literature appears to suggest different roles for consonants and vowels. Furthermore, the relative importance of these roles appears to be modulated by context. In order to resolve a few seemingly contradictory results, the current study investigated five main phenomena.

- (1) Owren and Cardillo used very different methods from the open-set responses required by Kewley-Port *et al.* (2007). Thus, an investigation of the contributions of segmental acoustics to monosyllabic words using open-set responses will help to further explain how the contributions of these acoustic cues are dependent upon the linguistic or acoustic context.
- (2) The examination of lexically easy and lexically hard words further helps to investigate which segmental acoustics mediate lexical access.
- (3) Low-level perceptual processes of filling-in or phonemic restoration that may differentially contribute to perceptual performance when consonants versus vowels are replaced by noise was investigated by varying the level of the replacement noise.
- (4) The contribution of segments to the identification of missing phonemes was investigated through phoneme scoring. Thus, this scoring method explicitly investigates for CVCs the amount of acoustic information contained

within the central vowel for the identification of the missing neighboring consonants, or within the consonant segments to identify the missing vowel.

- (5) Finally, as there is no discrete boundary between consonant and vowel acoustics due to the normal dynamics of speech production, the amount of transitional information at the C-V boundary was systematically varied in Experiment 2.

II. EXPERIMENT 1: CONTRIBUTION OF SEGMENTAL ACOUSTICS

Experiment 1 investigated the contribution of consonant and vowel acoustics to monosyllabic word intelligibility using a traditional definition of segment boundaries (see methods). Also investigated was how the contributions of these segmental acoustics are modulated by lexical difficulty of the words and by the level of the replacement noise used. Finally, phoneme scoring indexed the amount of information about segment identity that remained following its replacement.

A. Listeners

Twenty-seven normal-hearing young listeners (18–28 y, mean=22 y) were paid to participate in the experiment. All participants were native speakers of American English and had pure-tone thresholds no greater than 20 dB HL at octave intervals from 250 to 8000 Hz (ANSI, 2004). Listeners were randomly assigned to one of three different noise-replacement groups which differed only in the level of noise used during replacement.

B. Stimuli

Stimuli used in the current study were 148 monosyllabic CVC words previously recorded by Takayanagi *et al.* (2002). These words had been previously selected using an online lexical database (the 20,000-word Hoosier Mental Lexicon; Nusbaum *et al.*, 1984) derived from Webster's 1967 *Pocket Dictionary*. Word lists were spoken by a single male talker and were varied according to two levels of lexical difficulty based upon the neighborhood activation model (Luce and Pisoni, 1998), defined as lexically easy words (low neighborhood density, high frequency of usage) and lexically hard words (high neighborhood density, low frequency of usage). All words were normalized in RMS amplitude and presented at a level of 70 dB SPL. The average word duration was 437 ms (consonants: 237 ms; vowel: 200 ms). Word lists included the following consonants: /b, d, g, p, t, k, tʃ, dʒ, ʃ, ʒ, θ, f, s, v, z, h, m, n, ŋ, l, r, w, j/ and vowels: /i, ɪ, e, æ, a, ʌ, ɔ, o, u, ʊ, ə, aɪ, aʊ, ɔɪ/. Word lists, including measurements of frequency, density, and phoneme occurrence, have been published by Dirks *et al.* (2001).

1. Segmentation

Segmental boundaries between consonants and vowels were independently identified by two phoneticians and the midpoint between phonetician-marked boundaries was selected for use in the study. Any boundaries that were greater

than 10 ms from a phonetician-marked boundary were settled in consultation with a third expert in speech acoustics. All boundary markers attempted to explicitly model boundary locations following the rules (see Seneff and Zue, 1988) used for the TIMIT database (Garofolo *et al.*, 1990, www.ldc.upenn.edu) which had been used in previous noise-replacement studies using sentences (see Kewley-Port *et al.*, 2007; Fogerty and Kewley-Port, 2009). These boundaries were identified in PRATT (Boersma and Weenick, 2006, www.pratt.org) using dual displays of the waveform and spectrogram. Segmentation used landmarks of acoustic change 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 (as these transitions provide information regarding both phonemes). The full stop closure and burst was assigned to the consonant. Finally, a vowel followed by /r/ was treated as a single rhoticized vowel as done previously. Boundaries were adjusted within 1-ms to the nearest local amplitude minima (i.e., zero-crossing) to minimize introduction of transients.

2. Noise replacement

A background speech-shaped noise (SSN) and that used during replacement were based on the long-term average speech spectrum of a concatenation of all experimental test words. The low level speech-shaped background noise was continuously presented 20 dB below the speech level to mask any residual transients and began and ended 100 ms before and after the stimulus, respectively. The replacement interval was substituted either with silence, SSN at -6 dB SNR, or SSN at $+6$ dB SNR according to the specific experimental group to which the participant was assigned.

Half of the 148 words were presented with noise replacing the consonants and half with noise replacing the vowels. The 74 words in each segmental condition were further subdivided into 37 lexically easy and 37 lexically hard words. All participants heard the same words with one of two different segmental assignments (e.g., half of the participants heard “boss” in the consonant-only condition and “feed” in the vowel-only condition, while the remaining participants heard the reverse segmental assignment). Consonant-only words preserved all consonants while replacing the vowel with silence or noise depending upon the experimental group. Vowel-only words preserved the primary vowel acoustics and replaced consonant segments.

C. Design and procedure

This study used a 2 (segmental replacement) $\times$ 2 (lexical difficulty) $\times$ 3 (noise replacement level) mixed-model design with noise replacement level as a between-subject measure and the other two variables as repeated-measures variables. All stimuli were presented using Tucker-Davis Technologies System III hardware and passed through ER-3A insert earphones. Calibration was conducted using the speech-shaped noise designed from these stimuli and equated for RMS amplitude. The system was calibrated to present this speech-shaped noise at a sound level of 70 dB SPL by using a 2-cm³

FIG. 1. Word scores in RAUs according to the segmental replacement condition and lexical difficulty of the words. Error bars plot the standard error of the mean.

coupler and a Larson Davis model 2800 sound level meter with linear weighting. All participants were tested alone in a sound attenuating booth. Participants were instructed to listen to each word and type what they thought they heard on a PC running a MATLAB open-set response interface. All words were presented to the participants in a random order. No word was presented more than once to a given listener and no feedback was provided.

D. Scoring

Responses were automatically corrected for common typing and spelling errors and were phonetically transcribed from the CELEX database (Baayen *et al.*, 1995) using custom-made software and were also visually inspected. Automatic correction of responses occurred only for phonetic misspellings ($<1\%$ of responses), and were made to enable phonetic transcription using the database. Analysis of all corrections demonstrated that no scoring measures were influenced by these automatic corrections. Responses were scored using both word and phoneme scoring. All word and phoneme percent correct scores were transformed to rationalized arcsine units to stabilize the error variance prior to analysis (RAU, Studebaker, 1985).

E. Results

1. Word scoring

The between-subject factor of noise replacement level was first investigated using a mixed-model ANOVA. No main effect of group and no two-way group interactions were observed ($p > 0.05$). Therefore, groups were collapsed for all subsequent analyses. Figure 1 depicts the group data for segment types with lexical difficulty as the parameter. The subsequent repeated-measures ANOVA demonstrated a significant main effect of lexical difficulty [$F(1,26)=150.8$, $p < 0.001$] and no main effect for segmental type ($p > 0.05$). Thus, no overall difference in consonant-only and vowel-only words was observed. The interaction between these two factors was significant [$F(1,26)=16.5$, $p < 0.001$]. Post-hoc *t*-tests using Bonferroni correction demonstrated significant

FIG. 2. Phoneme scores are plotted in RAU according to the type of segmental replacement. Identification performance of each phoneme type (i.e., consonant or vowel) is displayed for each segmental replacement condition. Error bars plot the standard error of the mean.

lexical difficulty effects for both consonant-only [$t(26)=9.7$, $p<0.001$] and vowel-only [$t(26)=6.9$, $p<0.001$] words, with better performance for lexically easy words. Again, no significant difference was observed between segmental types for either easy or hard words ($p>0.01$). However, the difference between easy and hard words was twice as large for consonant-only words as compared to vowel-only words which was significant statistically [$t(26)=-4.0$, $p<0.001$].

2. Phoneme scoring

Results in Fig. 2 are plotted according to phoneme scoring. Significant main effects were found for the type of segment replacement [$F(1,26)=16.1$, $p<0.001$] and lexical difficulty [$F(1,26)=32.2$, $p<0.001$], but not the type of phoneme scored ($p>0.05$). However, as expected, post-hoc analysis demonstrated that participants made significantly more consonant errors than vowel errors in vowel-only words [$t(26)=-21.0$, $p<0.001$] and significantly more vowel errors in consonant-only words [$t(26)=19.0$, $p<0.001$]. That is, there were more errors for the segment type replaced by noise than the segment that was preserved in each word. However, there was still preserved identification of the missing phoneme. Missing vowels were identified with 34% accuracy from the preserved consonants, while missing consonants were identified with 41% accuracy from persevered vowels. Missing consonants were identified significantly better than missing vowels ($p<0.001$). In addition, identification of missing consonants from vowels was not influenced by the lexical difficulty of the word ($p>0.05$), whereas identification of missing vowels from the remaining consonants was ($p<0.001$). Consistent with Woods *et al.* (2010), the initial consonant was identified better than the final consonant in C-only words (60% vs 53% respectively, $p<0.001$). However, in V-only words the final consonant was identified best (37% for initial C vs 44% for final C, $p<0.001$).

As with word scoring, phoneme scoring showed a greater effect of lexical difficulty for consonant-only words vs vowel-only words when the vowel and final consonant

were scored ($p<0.01$). The effect of lexical difficulty for consonant-only and vowel-only stimuli did not differ significantly ($p>0.05$), however, when the initial consonant phoneme was scored. Thus, while performance between the segmental conditions resulted in similar overall performance, there was a qualitative difference in performance that was dependent upon the type of segmental errors made.

F. Discussion

The replacement interval was filled with silence or noise above or below the speech level for the different listener groups. However, the noise level used during replacement did not influence word identification or differentially influence consonant or vowel contributions. There was also no significant difference between C-only and V-only words, indicating similar contributions to intelligibility. However, the type of information conveyed by these segmental conditions appears to be functionally different. First, performance with C-only words is modulated more by the lexical difficulty of the items. This suggests that the contributions of consonants may be more influenced than vowels by the difficulty of lexical access, possibly indicating a greater lexical role. Second, phoneme scoring, not surprisingly, indicated that consonants and vowels conveyed fundamentally different information for the identification of phonemes in isolated words. However, information about missing neighboring phonemes is conveyed by both vowels and consonants to similar degrees (34% vs 41% respectively), either due to acoustic information retained in segmental transitions or top-down cognitive processes. This result is consistent with the silent-center syllable data of Strange *et al.* (1983) who found relatively preserved identification of the vowel when only transitional information at the beginning and end of the syllable was present, similar to the C-only condition here. Interestingly, in V-only contexts, the identification of the missing consonants is not influenced by the lexical difficulty of the word. Vowels appear to convey equal information for identifying consonants in both lexically easy and hard words, suggesting that identification of the missing phoneme may be conveyed acoustically, and is a bottom-up perceptual process. However, the identification of the missing vowel in C-only words is mediated by the lexical difficulty of the item. Vowels are identified better when missing in easy, more predictable words. This possibly suggests the involvement of top-down “filling-in” processes for the missing phoneme for C-only words. Combined, these results provide further support for consonants being involved more directly in lexical access. Experiment 2 investigates the contribution of segmental transitional information by successively adding or deleting this transitional information from the vowel or consonant.

III. EXPERIMENT 2: CONTRIBUTION OF TRANSITIONAL ACOUSTICS

Segmentation of the highly dynamic acoustics of speech into discrete consonants and vowels is inherently problematic. This is due in part because formant transitions provide information about the identity of both consonants and vowels

(Liberman *et al.*, 1967) and coarticulation results in overlapping acoustics which influence perception over multiple segments (see Kent and Minifie, 1977). Therefore, segmentation of discrete segmental units may be more a matter of convenience (Ladefoged, 2001). However, examining the contributions of mostly vowel or mostly consonant acoustic cues can help identify how listeners use general acoustic categories for understanding speech. This second experiment directly manipulated the amount of vowel and consonant information presented to investigate how these distributed acoustic cues at segmental boundaries contribute to the perception of isolated monosyllabic words. This experiment follows methods used by Fogerty and Kewley-Port (2009) to study transitional contributions in sentences.

A. Listeners

Twenty young normal-hearing listeners (18–26 y, mean = 20 y) participated in Experiment 2. No listeners completed both experiments. All participants were native speakers of American English and had pure-tone thresholds no greater than 20 dB HL at octave intervals from 250 to 8000 Hz (ANSI, 2004). Listeners were randomly assigned to one of four listening groups, five listeners per group. Each listener group listened to stimuli that contained different C-V transitional information.

B. Stimuli and design

The same monosyllabic words were used as in Experiment 1. Lexically easy and lexically hard words were again presented. Stimuli were processed according to the four listening groups for whom the segmental boundary was shifted by either a proportion of the vowel (VP) or a proportion of the consonant (CP). These boundaries were shifted by either 15% or 30% of the segment. Therefore, the four groups were 15%CP, 30%CP, 15%VP, and 30%VP. Within each group, listeners heard the acoustics preserved from either the consonants or the vowel. Thus, there were a total of eight conditions that are schematized in Fig. 3. Each condition is defined by the proportion of segmental acoustics presented. For example, 0.30C+V indicates that 30% of the consonant acoustics plus the entire vowel was presented to the listener. Note that this 30% is representative of the consonant acoustics presented over the entire syllable. The C-V boundary was moved into the initial and final consonant by 15%, resulting in a 30% gain in consonant information across the entire syllable. Similarly, 0.70 V indicates that 70% of the vowel was presented, with 15% removed at the initial boundary and another 15% deleted from the end of the vowel.

As there was no effect of noise level found in Experiment 1, silent replacement was used for all conditions. A constant speech-shaped background noise was again presented simultaneously with all stimuli at –20 dB SNR. This background noise began and ended 100 ms from the stimulus. The same procedures as Experiment 1 were used in replacing segmental acoustics.

FIG. 3. Schematic of the experimental conditions displaying the mean duration for each segment. The proportion of segmental information presented is displayed on the left. Each experimental listening group is listed on the right. Conditions in the shaded box display the original consonant-vowel boundary and were presented during Experiment 1. Solid bars = presented speech; Stippled bars = replaced segment.

C. Procedure

The same testing and scoring procedures were used as in Experiment 1. All listeners were seated in a sound-attenuating booth and typed their responses to the word stimuli presented monaurally via ER-3A insert earphones at 70 dB SPL. Listeners completed a familiarization task identical to Experiment 1 that matched the stimulus processing of their listener group. No feedback was provided to their open-set responses either during familiarization or testing.

D. Results and discussion

1. Word scoring

Figure 4 displays the mean performance in RAU for each of the experimental conditions. Experiment 1 results were obtained using the original C-V boundary. Only data from the 9 listeners of Experiment 1 who were in the silent

FIG. 4. Mean performance for each of the experimental groups. Exp. 1 group is data from the silent replacement group of Experiment 1 using the original C-V boundary. Each bar is coded by the segment preserved and the lexical difficulty of the word (e.g., Vowel Easy=word intelligibility for predominantly vowel acoustics preserved in easy words). Error bars plot the standard error of the mean.

replacement group are included in subsequent analyses, as stimuli for this condition underwent identical processing as those used in Experiment 2. Recall that the CP groups heard stimuli that reassigned transitional acoustics of the consonant to the vowel resulting in reduced consonant durations and increased vowel durations. In contrast, VP groups heard vowels with reduced durations and consonants with increased durations due to reassignment of vowel transitions. From Fig. 4 it is clear that listener performance between groups reflect these changes in the available segment duration. Means for lexically easy words were also higher than lexically difficult words across all conditions.

A mixed model ANOVA was conducted with segment replacement and lexical difficulty as within-subject factors and experimental group as a between-subject factor. Results demonstrated significant main effects for lexical difficulty [$F(1,24)=205.0$, $p<0.001$] and segment replacement [$F(1,24)=17.9$, $p<0.001$]. No main effect of group was obtained ($p>0.05$) for pooled conditions. However, a significant interaction of segment and group was obtained [$F(4,24)=100.4$, $p<0.001$] indicating that groups performed differently for the different segment replacement conditions as expected. A significant interaction between lexical difficulty and segment replacement was obtained [$F(4,24)=31.4$, $p<0.001$] as well as a three-way interaction with group [$F(4,24)=4.9$, $p=0.005$].

Next, repeated-measures ANOVAs were conducted for each group individually. A significant main effect for segmental replacement was obtained for all groups ($p<0.01$) with the exception of the original segmental boundary as obtained in Experiment 1. Overall, adding vowel or consonant acoustics to a segment significantly improved performance, while removing segmental information significantly deteriorated performance. A significant main effect for lexical difficulty was also obtained for all groups ($p<0.05$). The difference between performance for lexically easy and lexically hard words was larger when consonantal acoustics were presented. However, this was only significant for Exp1, VP%15, and VP%30 experimental groups ($p<0.01$). Thus, when consonantal acoustics were reduced in CP%15 and CP%30 conditions, both consonant and vowel replacements mediated the differences between lexical difficulty similarly.

As would be predicted by the Neighborhood Activation Model, listeners' errors were typically high-frequency words. Word frequency of each response error was determined from the CELEX database. The word-frequency difference between the target word and the response word was then calculated. Results demonstrated that across all conditions, word errors had significantly higher word frequency than the target for lexically hard words than for lexically easy words ($p<0.05$). However, no significant difference for this bias was observed for consonant-errors versus vowel-errors for any of the conditions tested, indicating that this effect was not mediated by segmental contributions.

Finally, the influence of transitional acoustics was also investigated to compare intelligibility differences from the original C-V boundary. Adjusting 15% of the vowel transitional information improved intelligibility for consonant-only words, and reduced intelligibility for vowel-only words

a) Vowel Identification

b) Consonant Identification

FIG. 5. Phoneme scores for the four experimental conditions of Experiment 2 and the silent replacement condition from Experiment 1 are plotted in RAU for (a) identification of the vowel, or (b) identification of the consonants. Phoneme scores are displayed according to the predominant segmental type presented (vowel or consonant) and the lexical difficulty of the word (easy or hard). Error bars plot the standard error of the mean. Asterisks (*) indicate significant differences, $p<0.01$.

($p<0.001$). Adjusting 15% of the consonant acoustics yielded similar results, by improving the intelligibility of vowel-only words ($p<0.001$). However, 30% of consonant transitional acoustics needed to be removed to significantly decrease the intelligibility of consonant-only words ($p<0.001$). Note that Fig. 4 displays a clear linear trend between the intelligibility of consonant-only words and the amount of consonant acoustics preserved.

2. Phoneme scoring

Figure 5 displays the results of the phoneme scoring for the four experimental groups of Experiment 2 and the silent replacement from Experiment 1 as comparison. The results displayed are divided into identification scores for the central vowel (top) and the initial and final consonants (bottom). A mixed model ANOVA comparing identification of vowels and consonants in RAU for the two segmental replacement conditions across the five experimental groups demonstrated significant main effects for segment replacement [$F(1,24)=8.1$, $p<0.001$], phoneme identification [$F(1,24)=18.8$, $p<0.001$], and lexical difficulty [$F(1,24)=95.4$, $p<0.001$] but not for group ($p>0.05$). However, significant two-way interactions were observed with group for segment replacement, phoneme identification, and lexical difficulty ($p<0.01$).

All groups demonstrated the same trend as observed in Experiment 1 for replacement at the original C-V boundary. That is, vowels were identified best when predominantly vowel acoustic cues were presented [$t(29)=-2.2$, $p<0.05$], and consonants were identified best when predominantly consonant acoustic cues were presented [$t(29)=3.1$, $p<0.01$]. This trend was noted even when 30% of the transitional acoustics of the neighboring segment were added. However, all groups demonstrated preserved identification of the missing phoneme. For example, vowels were still identified with 22% accuracy when only consonants were presented with 30% of the transitional acoustics removed. In this context, the consonants themselves were only identified with 38% accuracy. Across all conditions tested, there was no significant difference between identifying consonants from vowels, or vowels from consonants ($p>0.05$). As with Experiment 1, the initial consonant was identified better than the final consonant when predominantly consonant acoustics were presented ($p<0.001$), while the final consonant was identified better than the initial consonant when predominantly vowel acoustics were presented ($p<0.001$).

When listeners have minimal speech information, they might be expected to respond with lexically easy words a higher proportion of the time and therefore guess lexically easy words more often. Therefore, we might expect to see the greatest impact of the lexical difficulty of word items when minimal speech information is available. However, listeners did not significantly vary the word frequency of their responses across the different proportions of speech information presented ($p>0.05$). In addition, the phoneme scoring results presented here demonstrate that the lexical difficulty of the words influences the identification accuracy when maximal vowel information is presented with the consonants (i.e., presentation of the consonants at 30%VP). Note that lexical difficulty influenced the results most when maximal consonantal speech information was available. Significant differences in phoneme identification between easy and hard words are marked by asterisks in Fig. 5.

The types of listener errors were also investigated in a low-level feature-based analysis. Errors of manner, place, and voicing made in the identification of consonants systematically decreased as more speech information was presented, and likewise for high-low and front-back distinctions for vowel identification. In general, this progression was true regardless of whether consonants or vowels were replaced. The most notable finding from this analysis resulted from a comparison of identification of the initial and final consonants when the vowel was presented. A significantly higher proportion of errors in manner, place, and voicing occurred for initial consonant identification than for final consonant identification when the vowel was presented ($p<0.001$). This indicates that the vowel was more informative about all aspects of the final consonant than for the initial consonant. Indeed, all error types occurred on less than 25% of error trials with the exception of manner (51%) and place (40%) measurements for the initial consonant when the vowel was presented, indicating how uninformative the vowel is of the preceding consonant for these CVC words. Also notable was that the proportion of errors per error type for identifying the

final consonant was not any different for when the consonant was available compared to when the vowel was available. The same was true for vowel identification.

Between-group contrasts demonstrated that shifting the C-V boundary significantly impacted phoneme identification. Specifically, adding 15% of transitional consonant information onto the vowel significantly improved the identification of missing consonants from the vowel [$t(12)=-3.6$, $p<0.01$]. Adding 30% of the consonant onto the vowel also improved vowel identification [$t(12)=-4.1$, $p<0.01$], presumably due to overall improvements in word identification. Removing this same 30% from the consonant also significantly decreased consonant identification when consonant acoustic cues were presented [$t(12)=3.6$, $p<0.01$]. Rather noteworthy is the finding that in this condition where 30% of transitional information from the consonant is removed, the identification of the missing vowel was not significantly different from vowel identification when the entire consonant was present ($p>0.05$). This is consistent with [Parker and Diehl \(1984\)](#) who found that vowel identification remained above chance for closed-set vowel identification when 90% of middle of the CVC syllable was replaced by silence.

As for adjusting the transitional information of the vowel, adding 15% of the vowel to the consonant significantly improved the identification of both the vowel [$t(12)=-4.9$, $p<0.01$] and consonants [$t(12)=-4.3$, $p<0.01$] when predominantly consonant cues were present. Removing this 15% from the vowel significantly decreased identification of consonants from the remaining vowel acoustic cues [$t(12)=3.7$, $p<0.01$]. This was expected, as formant transitions into the vowel provide significant cues for consonant identification (e.g., [Liberman et al., 1967](#)). Removing 30% of the vowel also significantly decreased identification of the vowel [$t(12)=3.9$, $p<0.01$]. However, vowels were still identified with 36% correct identification in this condition.

3. Total speech available

By varying the duration of segmental acoustics presented, each condition presented a different amount of speech information that was available for word recognition. Figure 6 displays the duration of speech information presented for each experimental condition. Conditions that presented predominantly vowel or predominantly consonant acoustics are coded by squares and diamonds, respectively. Lexically easy and hard words were pooled for this analysis. There was a high Pearson-r correlation between performance and the duration of speech information presented, with 93% of the variance accounted for across all conditions by this one measure.

Several of the consonant and vowel conditions contained approximately equal durations of available speech acoustics. Paired t -tests were conducted between consonant and vowel conditions at each of these durations. No significant differences were obtained between consonant and vowel conditions that were matched in total speech duration ($p>0.01$). If differences between the two segmental conditions were obtained, then two different trend lines would be observed. However, all conditions fall along the same trend which is best described by the duration of the acoustic speech signal

FIG. 6. Each condition is plotted according to the duration of speech presented across the entire word. The linear trend is pooled for both consonant and vowel conditions. Squares=predominantly vowel acoustic cues; Diamonds=predominantly consonant acoustic cues. Conditions at the original C-V boundary from Experiment 1 are highlighted.

presented. Given these results, the segmental type preserved does not appear to influence performance above what is already accounted for by the proportion of available speech alone.

IV. GENERAL DISCUSSION

Results from Experiment 1 demonstrated that the level of noise replacement did not have an effect for either vowel or consonant replacement, despite the noise level being effectively presented about 16 dB above the average consonant intensity level in one case and nonexistent in another case. Bashford *et al.* (1992) also found no effect of noise level during the interruption of isolated words.

Results from Experiments 1 and 2 confirm that there is no vowel advantage in CVC word contexts. Indeed, no difference in speech intelligibility was obtained for consonant and vowel acoustics at the original C-V boundary. In addition, the amount of consonant and vowel acoustics was varied by adjusting the C-V boundary to include more or less transitional information. When the segment duration was accounted for, little difference was obtained between predominantly vowel and predominantly consonant segments. Instead, these two segmental conditions fell along the same linear trend, with performance highly predicted by the duration of speech presented (Fig. 6). The lack of a vowel advantage for words is consistent with Owren and Cardillo (2006). However, Owren and Cardillo (2006) found that listeners were more sensitive at detecting the meaning of consonant-only words than vowel-only words. This may be due to the simpler task requirements of a 2-alternative forced-choice detection task, rather than the open-set response required here, or the multisyllabic nature of Owren and Cardillo's stimuli. However, while listeners may be better able to detect word meaning from consonants than from vowels, this translates into no difference in open-set recognition performance between consonants and vowels across a range of segment durations.

FIG. 7. Word intelligibility (filled symbols) for CVCs in the current study and sentences (open symbols) in Fogerty and Kewley-Port (2009) are plotted according to the proportion of the utterance presented for that condition. Squares=predominantly vowel acoustic cues; Diamonds=predominantly consonant acoustic cues. Symbols that are circled indicate conditions that are at the original C-V boundary.

Furthermore, the overall performance of consonant and vowel conditions was not differentially influenced by the lexical difficulty of words. However, the difference between easy and hard words was at least twice as large for the consonant-only condition when the entire consonant was presented, suggesting that consonants may be more susceptible to the difficulty of lexical access. On the other hand, no difference was obtained when transitional acoustics were removed from the consonants and added to the vowel. For example, even when the vowel condition for the 30%CP group was 105 ms longer than the consonant condition, no significant difference between these segmental conditions was obtained. Thus, the difficulty of lexical access appears mediated by consonant acoustic cues and not the duration of speech information presented. This appears to corroborate interpretations that suggest consonants are more important for lexical identity than vowels (Owren and Cardillo, 2006; Toro *et al.*, 2008a).

A. Comparison to sentence contexts

This study was specifically interested in the conflicting results for the relative segmental contributions made by vowels and consonants in isolated words versus sentential contexts. In order to investigate this issue, comparison was made to a previous study conducted in our laboratory (see Fig. 7) using noise replacement in sentences (Fogerty and Kewley-Port, 2009). As in the current study with CVCs, Fogerty and Kewley-Port removed consonant or vowel segments from 42 TIMIT sentences (Garofolo *et al.*, 1990) and replaced them with low-level noise (-16 dB SNR). The sentences were then presented to listeners in the context of a continuous low-level background noise similar to the one used here. Listeners responded by repeating aloud each sentence they heard. Comparison of results at the original consonant-vowel boundary (circled symbols in Fig. 7) demonstrated that performance was significantly poorer for vowel-only words than vowel-only sentences [$t(9)=18.6$, $p<0.001$]. Sentence presentations resulted in improvements in intelligibility by a factor of 2. In particular, for word recognition performance

in a word context, vowel-only words had a mean score of 24%. Performance from [Fogerty and Kewley-Port \(2009\)](#) on the comparable condition in a sentence context was 65%. Comparison of consonant-only items across contexts demonstrated no significant differences ($p=0.31$). In particular, the mean word-recognition score for the word context was 25% whereas the mean word-recognition score in the sentence context was 27% for consonant-only stimuli. Thus, it appears that the linguistic and acoustic contextual setting of sentences benefits the contributions of vowels to speech intelligibility, but does not influence the contributions of consonants. This provides further support that vowels may carry acoustic information that is specific to sentences, such as prosody or coarticulation that may constrain lexical access. Furthermore, while the contributions of vowels appear mediated by context, consonant contributions appear to be independent of contextual factors. The lack of improvement for consonant-only stimuli in sentences is further highlighted by the fact that intelligibility is generally better in sentence contexts than in isolated words ([Miller et al., 1951](#)). These results suggest that the improvements that have been noted with sentence stimuli may be due in large part to the additional contribution of information carried by vowels in sentence contexts.

As dynamic variations in speech acoustics occur at the C-V boundary, Experiment 2 also examined how segmental contributions to intelligibility are modulated by transitional acoustics between consonants and vowels. The study by [Fogerty and Kewley-Port \(2009\)](#) also modified the segment durations presented. By calculating the proportion of the utterance presented, performance between CVC words and sentences may be compared across these conditions. Figure 7 displays both the data for Experiment 2 using CVC words (displayed as filled symbols) and for [Fogerty and Kewley-Port \(2009\)](#) who used sentences (displayed as open symbols). Sentences that contained predominantly consonant acoustics fall along the same trend line as obtained for CVC words with either segmental type. Indeed, the proportion of the utterance presented accounts for 92% of the variance across these three segmental conditions combined (i.e., CVC vowels, CVC consonants, and Sentence consonants). However, sentence vowels do not fall along this trend. While a similar slope is obtained, vowel acoustics in sentences afford much better intelligibility although less of the utterance is presented. Therefore, the speech information provided by vowels is influenced by the context, while consonants provide the same benefit to intelligibility regardless of the type of speech utterance.

B. Segmental transitions for CVC identification

[Fogerty and Kewley-Port \(2009\)](#) noted that acoustic transitions into vowels differentially influenced the intelligibility of sentences presented with predominantly consonant or vowel acoustics. Specifically, vowel transitions linearly improved the intelligibility of consonant-only sentences, while these same transitions, when removed, did not decrease the intelligibility of vowel-only sentences until over 30% of the vowel was removed. These results were inter-

preted as suggesting that vowel transitions provided additional speech information for consonant-only sentences, but were redundant with information contained in the vowel center.

This was not the case for the word context examined here. Vowel transitions both facilitated the intelligibility of consonant-only words and decreased the intelligibility of vowel-only words when only 15% of the vowel was manipulated. Transitions into the consonant also had a similar effect of influencing word intelligibility. Thus, vowel and consonant acoustics influence intelligibility linearly in word contexts. For word contexts, it is the duration of speech information that is important, rather than specific segmental information.

C. The roles of consonant and vowel acoustic cues

A number of studies have observed the importance of consonants in lexical processing (see [Nespor et al., 2003](#)), detection of synonymous meaning ([Owren and Cardillo, 2006](#)), and word segmentation ([Mehler et al., 2006](#)). Consonants also appear to constrain lexical selection more so than vowels across a number of languages ([Cutler et al., 2000](#)). The current study adds to this literature by demonstrating that the contribution of consonants is impacted by differences in the lexical difficulty of items to a greater extent (i.e., a factor of two) than that of vowels. In addition, even when transitional consonant acoustics are added on to the vowel so that vowel conditions are 1.5 times longer than consonant conditions, and therefore present more of the word, the difference between easy and hard words in the consonant and vowel conditions is no different.

Note that the lexical difficulty of the words is based on phonetic similarity. Therefore, for CVCs, there are twice as many opportunities for consonant errors to produce similar sounding lexical neighbors. As such, the lexical difficulty of CVC words may be determined, to a greater extent, by the substitution and deletion of consonants rather than vowels. Conceivably, this could have caused the larger difference in consonant errors between easy and hard words compared to vowel errors, because more lexical neighbors resulting from consonant changes were available. However, this distributional pattern of twice as many consonants as vowels was also measured for sentences ([Fogerty and Kewley-Port, 2009](#)), and therefore is not unique to CVCs. The frequency of occurrence of consonants relative to vowels may indeed be a factor in what underlies the importance of consonants for lexical access. Other factors may include the greater number of consonants than vowels in the language system as well as the greater distinctiveness of consonants, as vowels tend to lose distinctiveness and harmonize within words ([Nespor et al., 2003](#)).

While the evidence appears to suggest that consonants are more important for lexical access and processing for CVCs, it is ever more troublesome that vowels provide a two-to-one benefit over consonants for the intelligibility of sentences ([Fogerty and Kewley-Port, 2009](#); [Kewley-Port et al., 2007](#)). The current study demonstrated that this vowel advantage is limited to sentence contexts. However, only a

single monosyllabic structure was investigated here. Multi-syllabic words or vowel-initial words might yield different results. Recall, though, that [Owren and Cardillo \(2006\)](#) found no vowel advantage in their detection task using multisyllabic words that included initial consonant or vowel segments.

It is well known that sentence contexts greatly enhance the accuracy of word recognition relative to isolated words, even at the same signal-to-noise ratios (e.g., [Boothroyd and Nitttrouer, 1988](#); [Miller et al., 1951](#)). However, [Woods et al. \(2010\)](#) recently reported that sentence comprehension appears to rely more heavily upon vowel cues and semantic constraints for a patient with SNHL, as much larger elevations of consonant-identification thresholds were found than for sentence reception thresholds. Indeed, prior research suggests that vowels mark syntactic constituency ([Nespor et al., 2003](#)) and generalize over structural patterns in artificial languages, even when reduced in duration or when compared to sonorant consonants ([Toro et al., 2008b](#)). The importance of vowels in providing syntactic structural cues is consistent with a sentence-specific advantage for intelligibility that was described here. The prosodic contour has been found to be important for providing syntactic cues and constraining lexical access ([Wingfield et al., 1989](#)). Indeed, prosody appears to suppress possible word-like sequences that span prosodic units ([Shukla et al., 2007](#)). In addition, variations in fundamental frequency significantly benefit the intelligibility of sentences ([Laures and Weismer, 1999](#)). Therefore, fundamental frequency information is one likely acoustic cue that underlies the contribution of vowels in sentence contexts. Vowels appear to provide acoustic cues to the higher-order global structure of speech, while consonants provide acoustic cues to lower-order local units important in the lexical access and disambiguation of neighboring phonemes.

V. SUMMARY AND CONCLUSIONS

Overall, the results of this study demonstrate that a vowel advantage for speech intelligibility is specific to sentence contexts, while consonant contributions to intelligibility appear to be independent of context. Furthermore, the replacement of consonants and vowels with noise resulted in fundamentally different errors between the two types of replacements even though overall performance between conditions was the same. Nevertheless, both segments provided information regarding the missing phonemes, illustrating the distributed nature of speech cues.

Distributed acoustics were investigated in Experiment 2 by specifically manipulating the amount of transitional acoustics presented in either the consonant or vowel conditions. The results demonstrated that the proportion of the utterance presented overwhelmingly predicted performance, regardless of whether the word contained predominantly vowel or consonant acoustics. Furthermore, adding or deleting transitional information from consonants or vowels influenced intelligibility similarly. Comparison to the sentence results of [Fogerty and Kewley-Port \(2009\)](#) demonstrated that

consonants contribute to intelligibility similarly regardless of context, while vowels provide a much larger contribution to intelligibility when presented in sentences.

The results of this study are summarized as follows.

- (1) Consonant and vowel acoustics contribute equally to the intelligibility of isolated monosyllabic words.
- (2) Consonant contributions are impacted more by differences in lexical difficulty of the items.
- (3) The level of the noise replacement does not impact performance, nor are consonants versus vowels differentially influenced by the presence of noise during replacement.
- (4) Segments provide significant acoustic information to identify, although to a reduced degree, missing neighboring phonemes.
- (5) Transitional information linearly influences the contributions of vowels and consonants to word intelligibility.
- (6) Regardless of the amount of transitional acoustics provided, sentence contexts do not influence the contribution of consonant acoustics to intelligibility, but greatly enhance the contribution of vowels.

For isolated words, no difference in the contribution of consonants and vowels was observed for overall open-set word recognition. This was true even when the contribution of transitional acoustic cues at the C-V boundary were investigated. For isolated monosyllabic words, the proportion of speech information conveyed, not the segmental type, was most predictive of performance. The results of this study, combined with previous investigations, strongly suggest that vowels are extremely important for the intelligibility of meaningful sentences. Vowels, in sentences, uniquely transmit specific speech information that is not conveyed by the consonants and that strongly facilitates intelligibility.

ACKNOWLEDGMENTS

The authors would like to thank Diane Kewley-Port for her comments regarding this study and Robert Felty for his assistance in marking segmental boundaries and transcription computer code. This work was supported, in part, by NIA R01 AG008293.

- ANSI (2004). ANSI S3.6-2004, American National Standard Specification for Audiometers (American National Standards Institute, New York).
- Baayen, R. H., Piepenbrock, R., and Gulikers, L. (1995). "The CELEX lexical database," Linguistic Data Consortium, University of Pennsylvania, Philadelphia.
- Bashford, J. A., Jr., Riener, K. R., and Warren, R. M. (1992). "Increasing the intelligibility of speech through multiple phonemic restorations," *Percept. Psychophys.* **51**, 211–217.
- Boersma, P., and Weenink, D. (2006). "Praat: Doing phonetics by computer," (Version 4.2) (computer program): <http://www.praat.org/> (Last accessed March 17, 2008).
- Boothroyd, A., and Nitttrouer, S. (1988). "Mathematical treatment of context effects in phoneme and word recognition," *J. Acoust. Soc. Am.* **84**, 101–114.
- Carreiras, M., Gillon-Dowens, M., Vergara, M., and Perea, M. (2009). "Are vowels and consonants processed differently? Event-related potential evidence with a delayed letter paradigm," *J. Cogn. Neurosci.* **21**, 275–288.
- Carreiras, M., and Price, C. (2007). "Brain activation for consonants and vowels," *Cereb. Cortex* **18**, 1727–1735.
- Cole, R., Yan, Y., Mak, B., Fanty, M., and Bailey, T. (1996). "The contri-

- bution of consonants versus vowels to word recognition in fluent speech," in Proceedings of ICASSP '96, pp. 853–856.
- 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.
- Cutler, A., Sebastián-Gallés, N., Soler-Vilageliu, O., and van Ooijen, B. (2000). "Constraints of vowels and consonants on lexical selection: Cross-linguistic comparisons," *Mem. Cognit.* **28**, 746–755.
- Dirks, D. D., Takayanagi, S., and Moshfegh, A. (2001). "Effects of lexical factors on word recognition among normal-hearing and hearing-impaired listeners," *J. Am. Acad. Audiol.* **12**, 233–244.
- Fogerty, D., and Kewley-Port, D. (2009). "Perceptual contributions of the consonant-vowel boundary to sentence intelligibility," *J. Acoust. Soc. Am.* **126**, 847–857.
- 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.
- Kent, R., and Minifie, F. (1977). "Coarticulation in recent speech production models," *J. Phonetics* **5**, 115–133.
- 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), pp. 1–191.
- Laures, J., and Weismer, G. (1999). "The effect of flattened F0 on intelligibility at the sentence-level," *J. Speech Lang. Hear. Res.* **42**, 1148–1156.
- Li, N., and Loizou, P. C. (2008). "The contribution of obstruent consonants and acoustic landmarks to speech recognition in noise," *J. Acoust. Soc. Am.* **124**, 3947–3958.
- Liberman, A. M., Cooper, F. S., Shankweiler, D. P., and Studdert-Kennedy, M. (1967). "Perception of the speech code," *Psychol. Rev.* **74**, 431–461.
- Luce, P. A., and Pisoni, D. B. (1998). "Recognizing spoken words: The neighborhood activation model," *Ear Hear.* **19**, 1–36.
- Mehler, J., Peña, M., Nespor, M., and Bonatti, L. L. (2006). "The 'soul' of language does not use statistics: Reflections on vowels and consonants," *Cortex* **42**, 846–854.
- Miller, G., Heise, G., and Lichten, W. (1951). "The intelligibility of speech as a function of the context of the test material," *J. Exp. Psychol.* **41**, 329–335.
- Nespor, M., Peña, M., and Mehler, J. (2003). "On the different roles of vowels and consonants in speech processing and language acquisition," *Lingue e Linguaggio* **2**, 201–227.
- New, B., Araújo, V., and Nazzi, T. (2008). "Differential processing of vowels and consonants in lexical access through reading," *Psychol. Sci.* **19**, 1223–1227.
- Nusbaum, H. C., Pisoni, D. B., and Davis, C. K. (1984). "Sizing up the Hoosier mental lexicon: Measuring the familiarity of 20,000 words," Research on Speech Perception Progress Report No. 10, Speech Research Laboratory, Psychology Department, Indiana University, Bloomington, IN.
- 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.
- Parker, E. M., and Diehl, R. L. (1984). "Identifying vowels in CVC syllables: Effects of inserting silence and noise," *Percept. Psychophys.* **36**, 369–380.
- Ramus, F., Nespor, M., and Mehler, J. (2000). "Correlates of linguistic rhythm in the speech signal," *Cognition* **75**, AD3–AD30.
- Seneff, S., and Zue, V. W. (1988). "Transcription and alignment of the TIMIT database," in Proceedings of the Second Meeting on Advanced Man-Machine Interface Through Spoken Language, pp. 11.1–11.10.
- Shukla, M., Nespor, M., and Mehler, J. (2007). "An interaction between prosody and statistics in the segmentation of fluent speech," *Cogn. Psychol.* **54**, 1–32.
- 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., Jenkins, J. J., and Johnson, T. L. (1983). "Dynamic specification of coarticulated vowels," *J. Acoust. Soc. Am.* **74**, 695–705.
- Studebaker, G. (1985). "A rationalized arcsine transform," *J. Speech Hear. Res.* **28**, 455–462.
- Takayanagi, S., Dirks, D., and Moshfegh, A. (2002). "Lexical and talker effects on word recognition among native and non-native listeners with normal and impaired hearing," *J. Am. Acad. Audiol.* **16**, 494–504.
- Toro, J. M., Nespor, M., Mehler, J., and Bonatti, L. L. (2008a). "Finding words and rules in a speech stream: Functional differences between vowels and consonants," *Psychol. Sci.* **19**, 137–144.
- Toro, J. M., Shukla, M., Nespor, M., and Endress, A. D. (2008b). "The quest for generalizations over consonants: Asymmetries between consonants and vowels are not the by-product of acoustic differences," *Percept. Psychophys.* **70**, 1515–1525.
- Wingfield, A., Lahar, C. J., and Stine, E. A. L. (1989). "Age and decision strategies in running memory for speech: Effects of prosody and linguistic structure," *J. Gerontol.* **44**, 106–113.
- Woods, D. L., Yund, E. W., and Herron, T. J. (2010). "Measuring consonant identification in nonsense syllables, words, and sentences," *J. Rehabil. Res. Dev.* **47**, 243–260.