Importance of envelope modulations during consonants and vowels in segmentally interrupted sentences^{a)}

Daniel Fogerty^{b)}

Department of Communication Sciences and Disorders, University of South Carolina, Columbia, South Carolina 29208

(Received 24 June 2013; revised 26 November 2013; accepted 15 January 2014)

The present study investigated the importance of overall segment amplitude and intrinsic segment amplitude modulation of consonants and vowels to sentence intelligibility. Sentences were processed according to three conditions that replaced consonant or vowel segments with noise matched to the long-term average speech spectrum. Segments were replaced with (1) low-level noise that distorted the overall sentence envelope, (2) segment-level noise that restored the overall syllabic amplitude modulation of the sentence, and (3) segment-modulated noise that further restored faster temporal envelope modulations during the vowel. Results from the first experiment demonstrated an incremental benefit with increasing resolution of the vowel temporal envelope. However, amplitude modulations of replaced consonant segments had a comparatively minimal effect on overall sentence intelligibility scores. A second experiment selectively noise-masked preserved vowel segments in order to equate overall performance of consonant-replaced sentences to that of the vowel-replaced sentences. Results demonstrated no significant effect of restoring consonant modulations during the interrupting noise when existing vowel cues were degraded. A third experiment demonstrated greater perceived sentence continuity with the preservation or addition of vowel envelope modulations. Overall, results support previous investigations demonstrating the importance of vowel envelope modulations to the intelligibility of interrupted sentences.

© 2014 Acoustical Society of America. [<http://dx.doi.org/10.1121/1.4863652>]

PACS number(s): 43.71.Gv, 43.71.Es, 43.72.Dv [ICB]

Pages: 1568–1576

I. INTRODUCTION

Speech interrupted by noise has been used as a simplified case for listening to speech in the presence of a fluctuating masker. Performance between these two conditions is highly correlated (Jin and Nelson, 2010). However, in the case of interruption, deletion of the interrupted speech segments ensures that speech cues during the interruption interval do not contribute to perception. Instead, acoustic cues from the remaining speech segments must be combined with linguistic and contextual knowledge for optimum speech intelligibility. Therefore, interrupted speech conditions can help identify the acoustic cues that are important for speech perception, particularly if such interruptions are locked to specific acoustic cues, such as speech segments.

Such interruption studies have a long history in the speech perception literature, including prominent silent-center syllable studies that were designed to index the contribution of formant transitions into vowels (e.g., Strange *et al.*, 1983). In recent years this paradigm has been used to investigate the relative perceptual contributions of different segmental and acoustic cues in words (e.g., Owren and Cardillo, 2006; Fogerty and Humes, 2010) and sentences (e.g., Kewley-Port *et al.*, 2007; Fogerty and Kewley-Port, 2009; Stilp and Kluender, 2010; Chen *et al.*, 2013). Results suggest

that the context can greatly influence vowel contributions leading to better intelligibility in sentences (Fogerty and Humes, 2010; Fogerty *et al.*, 2012). This additional contribution in sentences appears to be related to the vowel temporal envelope (Fogerty and Humes, 2012; Fogerty, 2013). The current study investigated envelope modulations provided during consonant and vowel segments in an attempt to identify how segmental modulation contributes to sentence intelligibility.

A. Importance of preserving the envelope during temporal interruption

Modulating the noise masker by the missing speech improves intelligibility (Bashford *et al.*, 1996) through the assignment of the masker modulation to the speech stream (Shinn-Cunningham and Wang, 2008). Masker modulations are used as informative speech cues even though they occur during the masker and are not required to potentially mask the speech signal. This can be viewed as potentially separate from other perceptual restorations where the masker is required to potentially energetically mask the missing speech information (Bashford *et al.*, 1992).

Additional studies of interrupted speech have highlighted the importance of preserving the amplitude envelope as it appears that interruption and fluctuating maskers can directly interfere with the use of envelope cues (e.g., Nelson *et al.*, 2003; Nelson and Jin, 2004; Gilbert *et al.*, 2007). Under such conditions listeners appear to shift perceptual weighting to other acoustic properties or frequency regions that contain temporal and spectral cues that remain relatively

^{a)}Portions of this data were presented at the 164th Meeting of the Acoustical Society of America [J. Acoust. Soc. Am. 132 (1967)], Kansas City, MO.

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

intact during the interruption (Fogerty, 2011). Thus, maintaining temporal envelope cues during interruption, even if cues regarding phonetic identity are lost, may be important for maximizing speech intelligibility during interruption.

Recently, Fogerty (2013) completed an acoustic analysis of sentences interrupted to replace consonants or vowels with noise. Duration, temporal envelope, and voicing cues were measured. Results demonstrated that preservation of the temporal envelope best predicted sentence intelligibility during segmental interruption, with vowels best preserving the overall envelope of the sentence. In contrast, intelligibility of interrupted words was best predicted by the preserved duration. For sentences, it appears that slow varying amplitude changes across the entire sentence contribute essential information for sentence-level speech intelligibility.

B. Importance of consonant and vowel envelope cues

In order to understand the intended message, listeners must integrate information over time, which may be assisted by the acoustic restoration of certain speech cues during the interrupted interval (Shinn-Cunningham and Wang, 2008). The present study investigated the importance of overall segment amplitude and intrinsic segment amplitude modulation to sentence intelligibility by restoring these cues during the interrupted (i.e., noise-replaced) segment. Dynamic changes in the overall segment level help to convey sentence-level temporal envelope modulation, while intrinsic segment modulation conveys faster envelope cues within the phoneme. Evidence suggests that restoring these envelope modulations may be important for accessing cues provided by both consonants and vowels (e.g., Shannon *et al.*, 1995; Xu *et al.*, 2005).

Regarding consonants, many of these phonemes are characterized by rapid, transient changes in amplitude, such as result from noise bursts (Stevens, 2002). Thus, faster rate phoneme-level modulations within consonants may be particularly important for conveying consonant cues necessary for intelligibility. However, it is important to note that while these consonant-level temporal modulations may code important cues for phoneme identification (e.g., Van Tasell *et al.*, 1987; Rosen, 1992; Drullman *et al.*, 1994; Shannon *et al.*, 1995), it is unclear how these cues contribute to overall sentence intelligibility above what is provided by vowel cues.

Envelope modulations also provide important cues present during vowels. Recently, by selectively manipulating the acoustics of consonants or vowels, Fogerty and Humes (2012) demonstrated that the temporal envelope of vowels is essential for providing cues to sentence intelligibility, but not word intelligibility. In addition, the acoustic analysis of segmentally interrupted sentences supports the hypothesis that sentence intelligibility during interruption can be predicted based upon the preservation of the temporal envelope, which performs better than durational metrics for sentence materials (Fogerty, 2013). In addition, the analysis revealed better preservation of the temporal envelope during vowel segments. This is supported by the finding that temporal envelope-based measures of speech intelligibility predict better intelligibility for vowel-dominated segments compared to consonant-dominated segments (Chen and Loizou,

2012). Based upon these previous results, it was hypothesized that restoring the overall vowel level during interruption would facilitate sentence intelligibility.

Using the intelligibility metric of Chen and Loizou (2012), dynamic changes in the temporal envelope that are characteristic of consonant-vowel boundaries and that preserved relatively similar proportions of consonants and vowels resulted in the highest speech intelligibility predictions. Thus, it appears that the greatest contribution of the temporal envelope to speech intelligibility may occur during periods of dynamic change. As consonants are typically of weaker intensity than vowels, consonant-vowel boundaries are characterized by relatively rapid changes in instantaneous amplitude. During segmental noise replacement, this rapid change can be completely obliterated, particularly when a low-level noise is used (see, for example, Kewley-Port *et al.*, 2007; Fogerty and Kewley-Port, 2009). The current study investigates the rapid amplitude change at this boundary by using a steady-state replacement noise presented at the level of the replaced segment. In such a condition, cues to vowel identity as well as amplitude modulations during the vowel remain completely obliterated, but boundary changes in the amplitude envelope are relatively preserved.

The purpose of the current study was to examine contribution of envelope modulations across the sentence (i.e., segment level) and within the interrupted segment (i.e., segment modulation). In addition, overall performance level could also be a factor mediating the degree of improvement between conditions replacing vowels or consonants. Therefore, a second experiment was conducted to more closely match performance between the conditions through additional noise masking of the preserved segments. Finally, the contribution of perceived continuity of the sentence was investigated to index perceptual organization and segregation of speech and noise sources. Listeners could have more difficulty using masker modulations to facilitate sentence recognition when the informative noise source is perceptually segregated from the target speech (e.g., Shinn-Cunningham and Wang, 2008).

II. EXPERIMENT 1: MODULATION OF NOISE-REPLACED CONSONANTS AND VOWELS

A. Listeners

Two groups of young normal-hearing listeners were recruited to participate in the experiment. The first group of listeners completed the vowel replacement conditions ($N=8$, $M=21$ yrs, female = 8) while the second group of listeners completed the consonant replacement conditions ($N=9$, $M=23$ yrs, female = 8). All listeners were native speakers of English and passed a pure-tone audiometry screening.

B. Stimuli and design

Forty-two sentences were selected from the TIMIT database (Garfalo *et al.*, 1990) to be used in the study. These sentences were spoken by 21 male and 21 female talkers. This sentence set has been used in previous studies of segmentally

FIG. 1. Waveforms and spectrograms for the different stimulus processing conditions. (A) The full, unprocessed sentence, “We always thought we would die with our boots on,” was processed in the following six segmentally interrupted conditions. (B) Vowel-replacement conditions preserve the consonant and replace the vowel with one of three types of SSN. Likewise, (C) consonant-replacement conditions preserve the vowel and replace the consonant with one of the same three types of SSN.

interrupted sentence intelligibility (e.g., [Kewley-Port et al., 2007](#); [Fogerty and Kewley-Port, 2009](#)). The consonant-vowel boundaries are specified by the TIMIT database. These boundaries are typical of traditional segmental boundaries that were marked by automatic segmentation and confirmed by experienced phoneticians.

Sentences were processed to segmentally interrupt speech using the segmental boundaries identified by the TIMIT database and adjusted within 1-ms to the nearest local minima (i.e., zero-crossing). Fourteen sentences were presented in each of three noise replacement conditions. Each listener group had the same types of noise replacement,

but replacement preserved either the vowels or the consonants in the sentence as defined by the listener group.

Speech was interrupted to replace vowel or consonant segments with noise matching the long-term average spectrum for the unprocessed sentences. The same speech-shaped noise (SSN) was then processed further to create three different noise types (see Fig. 1). These types were:

- (1) Low-level noise that distorted the overall sentence envelope,
- (2) segment-level noise that restored the overall syllabic amplitude modulation of the sentence, and
- (3) segment-modulated noise that further restored faster temporal envelope modulations intrinsic to each phoneme.

The low-level noise was presented at a constant level across the sentence at 16 dB below the average sentence level. This level was below the average level of both the vowels (−21 dB re: Vowel) and the consonants (−11 dB re: Consonant). This low-level noise resulted in distorting the overall amplitude envelope of the sentence.

The segment-level noise presented each interrupting noise at the level of the replaced segment. Thus, the amplitude of each noise interval across the sentence was dynamically varied. This method served to restore the gross temporal envelope of the sentence. However, each individual noise interval had a steady envelope. Thus, amplitude modulations intrinsic to the replaced segment were distorted.

The segment-modulated noise resulted in the highest preservation of the amplitude envelope by modulating each noise interval by the envelope of the replaced segment. Thus, this method restored both sentence-level slow-varying modulations as well as faster modulation within each replaced phoneme. The amplitude envelope of the segment was extracted by half-wave rectification followed by low-pass filtering at 64 Hz using a sixth order Butterworth filter. This envelope was then used to modulate the SSN used to replace the segment and scaled to the root-mean-square level of that segment.

An acoustic analysis of the stimuli, similar to the procedures defined by [Fogerty \(2013\)](#) for calculation of the Envelope Correlation Index, was conducted to demonstrate the effect of these different processing strategies on the amplitude envelope of the sentence. The envelopes of the final processed stimulus and the original sentence were extracted according to the procedures above using half-wave rectification and low-pass filtering the envelope at 64 Hz. The correlation between these extracted envelopes was calculated and the resulting coefficient was arcsine transformed to ensure equal variance across the conditions. Average values across 42 processed sentences are displayed in Fig. 2. This analysis demonstrates the high preservation of the envelope for consonant-replaced sentences when the vowel is preserved. However, for vowel-replaced conditions the envelope is degraded, particularly when a constant low-level noise is used for replacement.

A series of paired-samples *t*-tests were conducted to assess the difference between the conditions. Consonant-replaced conditions all had higher indices of envelope

FIG. 2. The preserved amplitude envelope, an arcsine transformation of the correlated envelopes of the stimulus and original sentences, is displayed for the three different noise types under consonant and vowel replacement conditions. Error bars indicate the standard error of the mean calculated across the 42 sentences analyzed.

preservation compared to vowel-replaced conditions [low: $t(41) = 23.0$, $p < 0.001$; segment: $t(41) = 3.5$, $p < 0.001$; modulated: $t(41) = 12.8$, $p < 0.001$]. In addition, for both consonant- and vowel-replaced sentences, greater preservation of the envelope was obtained for segment-level versus low-level stimuli [vowel-replaced: $t(41) = 6.8$, $p < 0.001$; consonant-replaced: $t(41) = 42.4$, $p < 0.001$], and modulated versus segment-level conditions [vowel-replaced: $t(41) = 8.6$, $p < 0.001$; consonant-replaced: $t(41) = 9.8$, $p < 0.001$]. These measures confirm that each condition significantly impacted the preservation of the temporal envelope across the sentence.

C. Procedure

Listeners were seated in a sound-attenuating booth and listened to the stimuli monaurally via the right headphone (Sennheiser HD 280PRO). The level of the speech (prior to replacement) was calibrated to be presented at 70 dB sound pressure level using a Larson-Davis 800B sound level meter with linear weighting. Listeners were instructed to repeat each sentence and were encouraged to guess. Sentence responses were audio recorded and analyzed offline by trained raters. All words in the sentence were scored. A strict scoring procedure was used in that listeners were required to repeat the word exactly to be scored as correct (e.g., no missing or extra suffixes), but words could be repeated in any order.

D. Results and discussion

Results were first examined using an analysis of variance with noise-type as a repeated measures variable and segment replacement group as a between-subjects factor. Results demonstrated a significant main effect of noise condition [$F(2,28) = 31.2$, $p < 0.001$] and segment replaced [$F(1,14) = 100.2$, $p < 0.001$], as well as a significant interaction [$F(2,28) = 11.6$, $p < 0.001$]. Therefore, further analysis

was conducted separately for the two segment replacement groups.

1. Vowel-replaced

Results for the vowel replacement conditions are displayed in Fig. 3(a). All pairwise contrasts between the three noise types were significantly different after using Bonferroni-correction. The low-level noise [$M = 27\%$, standard deviation (SD) = 6%] resulted in poorer performance than the segment-level noise ($M = 42\%$, SD = 9%), $t(7) = 4.3$, $p < 0.01$. In addition, the performance with the segment-level noise was incrementally poorer than performance with the segment-modulated noise ($M = 52\%$, SD = 10%), $t(7) = 3.2$, $p < 0.05$. Thus, performance improved with greater resolution of the temporal envelope conveyed during vowels.

2. Consonant-replaced

Results for the consonant replacement conditions are displayed in Fig. 3(b). In these conditions the vowels were preserved and the consonants were replaced with one of three types of SSN. In agreement with previous results (e.g., Kewley-Port *et al.*, 2007; Fogerty and Kewley-Port, 2009), better overall performance was observed for these consonant replacement conditions (i.e., with preserved vowels) than for the former vowel replacement conditions that only preserved the consonants. Collapsed across all noise modulations, this resulted in a 34 percentage-point advantage for preserving vowel segments in the consonant-replaced conditions. However, in contrast to the vowel replacement conditions, only incremental performance was observed with increasing resolution of amplitude modulations carried by the consonants. Following Bonferroni-correction for multiple comparisons, the only significant difference was between the low-level ($M = 70.5$, SD = 9.1) and envelope-modulated ($M = 76.5$, SD = 8.8) noise conditions, $t(7) = 4.6$, $p < 0.01$. This represents only a 6 percentage-point benefit in adding consonant amplitude modulations compared to the 26 percentage-point improvement in adding vowel amplitude modulations. Segment-level noise was not significantly different from either condition (Bonferroni-adjusted $p > 0.05$). Even though overall performance was greatest in this condition, two listeners actually performed slightly worse (2 to 4 percentage-points) compared to low-level noise replacement, which resulted in the non-significant result compared to envelope-modulation in which all listeners received benefit. These results suggest that faster rate phoneme-level modulations within each consonant contribute to intelligibility, possibly in addition to overall consonant level that was used inconsistently by the listeners in this study.

As indicated previously, there was a significant increase in performance when provided vowel envelope cues. On average, intelligibility scores doubled as a result of introducing modulations of the vowel envelope to the interrupting noise for an average increase of 24 percentage-points. In contrast, only minimal performance gains were observed in the case of modulating the interrupting noise by the consonant envelope. Intelligibility scores for the consonant group only improved by 9%, an average increase of 6 percentage-points.

(a) Vowel-replaced sentences

(b) Consonant-replaced sentences

FIG. 3. Performance for the three noise types during (a) vowel replacement when only the consonants were preserved and (b) consonant replacement when only the vowels were preserved during segmental interruption of the sentence. Consonant and vowel segments were replaced with one of three types of noise: Low-level (16 dB SNR), segment-level, and segment-modulated that incrementally restored amplitude modulation properties of the replaced segments. Error bars indicate the standard error of the mean.

Therefore, vowels may provide the predominant amplitude modulation cues necessary for sentence-level interpretation. This observation of the behavioral results is consistent with the acoustic analysis on the preserved temporal envelope conducted here and by [Fogerty \(2013\)](#).

In addition, it is important to note that restoring both the sentence-level modulation provided by the overall vowel intensity and the phoneme-level modulation present during individual vowels is informative. Thus, vowels appear to carry temporal modulation cues at different time scales, slow across the entire sentence, and rapid changes within the segment. In contrast, while some consonants are characterized by rapid amplitude changes, such as stops, little perceptual benefit was derived even from within-segment modulation. However, given the high heterogeneity between consonants (e.g., [Singh and Allen, 2012](#)), a significant effect of noise modulation may be observed for individual consonants, particularly as distortions in consonant amplitude have differential effects depending on the consonant ([Freyman et al., 1991](#)).

A number of factors could be related to the lack of significance obtained for these preliminary results, including higher overall performance levels limiting the range for performance improvements, and less relevance of temporal envelopes obtained from consonants. Experiment 2 specifically addresses possible ceiling effects limiting the range for improvement in consonant-replaced conditions.

III. EXPERIMENT 2: NOISE MASKER AND PERFORMANCE LEVEL

Experiment 1 demonstrated better performance for sentences that preserved the acoustic cues present during the vowel segments (i.e., consonant-replaced conditions). Furthermore, while partially restoring envelope modulations across the sentence during vowels resulted in systematically better performance with increased envelope restoration, restoring envelope cues present during the noise-replaced consonants did not lead to systematic improvements in intelligibility. Therefore, preliminary conclusions, consistent with [Fogerty and Humes \(2012\)](#), suggest that envelope modulation of vowel segments is the primary contributor to systematically improve intelligibility for interrupted sentences, with consonant modulations having a more limited contribution in sentence contexts. However, the performance level was much higher for the consonant-replaced sentences that preserved vowel segments. Therefore, it may be that an effect of envelope restoration was not observed in this condition due to the better overall performance. To examine this issue in more detail, Experiment 2 was designed to test the consonant-replaced conditions within the same performance range as the vowel-replaced conditions. This was done by degrading the preserved vowel segments by adding a continuous SSN. This noise effectively degraded spectral and temporal modulation cues selectively for the vowel segments. Importantly, in addition to possible ceiling effects, cues conveyed by the consonants may take on added importance when spectral-temporal cues from the vowels are degraded.

A. Methods

A new group of six young, normal hearing adult listeners completed Experiment 2 ($M = 23$ yrs, female = 5). Stimuli and methods were identical to the consonant-replaced condition tested in Experiment 1 with one exception. Stimuli were presented in the context of a steady-state SSN at 3 dB signal-to-noise ratio (SNR). The continuous SSN matched the long-term average speech spectrum of the concatenated, unprocessed sentences. The effect of adding this noise was to selectively reduce the spectral and temporal detail contained within preserved vowel segments. This masking noise was only presented during the preserved vowel segments and was gated off during the noise-replaced consonant segments. The purpose of this background noise was to reduce overall performance to a similar performance range obtained in the vowel-replaced condition of Experiment 1 by degrading the vowel segment while preserving the envelope cues introduced during the replaced consonant segments.

B. Results and discussion

Average performance across all consonant-replaced conditions was 39%. Compared to the average performance of 40% correct for vowel-replaced conditions in Experiment 1, the vowel distortion employed in Experiment 2 was successful in approximating equal overall performance levels between the two replacement conditions. Figure 4 compares the results for these conditions of the two experiments. As can be seen from this figure, while modulating the interrupting noise by the vowel significantly impacts performance, little effect is observed for consonant modulations that are introduced in the interrupting noise. Pairwise comparisons between the three different noise conditions in Experiment 2 demonstrated no significant differences in performance. Low-level ($M = 39.7$, $SD = 5.9$), segment-level ($M = 40.3$, $SD = 7.9$), and envelope-modulated ($M = 37.4$, $SD = 13.0$) resulted in a highly similar performance. Thus, the results were largely consistent with Experiment 1, except consonant envelope modulations did not facilitate sentence intelligibility when spectrotemporal cues from neighboring vowels are also degraded.

IV. EXPERIMENT 3: PERCEPTUAL CONTINUITY OF SEGMENTALLY INTERRUPTED SENTENCES

In the present case of consonant envelope modulation of interrupting noise, it may be that the system actually works the way [Shinn-Cunningham and Wang \(2008\)](#) state that it should—by segregating speech and noise sources. This would lead to listeners not assigning and using the consonant modulations to facilitate speech recognition. This segregation could, in part, be due to the preserved periodicity of vowel segments. While speech is naturally composed of alternating periodic and aperiodic segments from vowels and consonants, it may be that the discrete and artificial replacement of aperiodic segments created more discontinuous percepts between such segments, facilitating segregation of the

FIG. 4. Sentence intelligibility scores are compared for the vowel-replacement conditions in Experiment 1 (light gray) and the new consonant-replacement conditions from Experiment 2 (dark gray). Overall performance across the two experiments was closely approximated through noise masking of the remaining vowel in the consonant-replaced conditions (Experiment 2). Error bars indicate the standard error of the mean.

noise source. In contrast, in the vowel replacement conditions, the preserved speech segments (i.e., the consonants) were already predominantly aperiodic sources. Thus, few alternating periodic and aperiodic segments occurred, resulting in a more continuous percept of the speech, similar to aperiodic whispered speech. This continuous aperiodic quality may assist listeners in assigning modulations of the noise to the speech, facilitating speech recognition while at the same time failing to segregate sound sources.

This third experiment was designed to determine if perceptual continuity of the sentence could explain how listeners assigned (or did not assign) segmental envelope modulations present during the interrupting noise to the speech source in order to facilitate intelligibility.

A. Methods

Six young normal-hearing listeners (five female) participated in the third experiment. All listeners had previously participated in one of the earlier experiments. All listeners completed a continuity judgment task where they listened to processed sentences and responded to whether the speech sounded (1) continuous or (2) interrupted.

Sentences were processed according to six conditions. Four of the conditions used stimuli that were presented in Experiment 1: Consonant or vowel replacement with either low-level noise or segment-modulated noise. Two additional conditions were also tested that presented continuous speech, but selectively presented noise simultaneously during consonant or vowel segments. The level of the noise was presented at 0dB SNR relative to the average sentence level. The segmentally interrupted masking noise was created to match the average speech spectrum of the target sentence and modulated by a different sentence recorded from the same talker. Thus, the interrupted masking noise contained speech envelope modulations independent of the target sentence and was selectively presented to partially mask either consonant or vowel segments of the target speech sentence. Ten sentences were selected from the TIMIT database to be presented in each of these 6 conditions for a total of 60 experimental sentences.

Testing procedures were identical as in the previous experiments with the exception that listeners responded by typing “C” or “I” on the response interface, corresponding to whether they perceived the sentence to be continuous or interrupted, respectively.

B. Results and discussion

The results for the six experimental conditions are displayed in Fig. 5. Sentences that were continuous with modulated noise turned on during consonant or vowel segments (white bars) were largely perceived as continuous ($M = 92\%$ and 92% of sentences judged as continuous for consonant and vowel masking, respectively). For the segmentally interrupted sentences (gray scale), a series of planned-comparison paired t -tests were conducted to determine the difference between the perceptual continuity of vowel and consonant conditions.

The addition of vowel modulations to the existing consonant resulted in significantly more sentences judged as

FIG. 5. The proportion of “continuous” responses is displayed for the sentence continuity judgment task in Experiment 3. Vowel- and consonant-replaced sentences were interrupted by a low-level noise, a segment-modulated noise, or presented continuously with a speech-modulated masker selectively added to the consonant or vowel segments of the target speech at 0 dB SNR. Error bars indicate the standard error of the mean.

continuous ($M = 55\%$) compared to low-level noise replaced sentences ($M = 15\%$), $t(5) = 6.9$, $p < 0.001$; whereas the addition of consonant modulations to preserved vowel segment resulted in a trend toward greater perceived continuity ($M = 65\%$ with consonant modulations compared to $M = 47\%$ for low-level consonant replacement), but did not reach significance, possibly due to the small sample size, $t(5) = 2.5$, $p = 0.06$. Overall, sentences that preserved vowel envelope modulations (i.e., consonants with vowel-modulated noise and vowels with noise-replaced consonants) resulted in the greatest perceived continuity, with no significant difference observed between these conditions.

The difference between low-level and segment-modulated noise replacement was investigated. A greater improvement in sentence continuity was observed for adding vowel modulations to preserved consonants (an increase of 40 percentage-points) compared to adding consonant modulations (18 percentage-points), $t(5) = 3.1$, $p < 0.05$.

Overall, these results suggest that envelope modulations significantly contribute to the perceived continuity of sentences. This was particularly observed when restoring the vowel envelope in vowel-replaced sentences and is consistent with what would be predicted based on the acoustic analysis of envelope preservation (see Fig. 2). This is also consistent with the improvement in sentence intelligibility for vowel-replaced sentences observed as a result of envelope restoration. Overall, envelope modulations of vowel segments are essential for sentence intelligibility and perceived continuity of interrupted sentences. This may be due to the low-to-mid frequency and high intensity nature of vowels that are specifically capable of carrying the syllabic rhythm and stress of the sentence (Port, 2003).

V. GENERAL DISCUSSION

Previous investigations have demonstrated that modulating an interrupting noise by the temporal envelope of the target speech significantly improves speech intelligibility

(Bashford *et al.*, 1996; Shinn-Cunningham and Wang, 2008). The present study expanded upon this discussion by examining the contribution of envelope modulations from consonant and vowel segments. Previous investigations of segmentally interrupted speech have suggested the importance of vowel envelope cues to sentence intelligibility (Fogerty and Humes, 2010; Fogerty, 2013). The present study supports these previous observations, as overall sentence intelligibility doubles when the temporal envelopes of missing vowels are used to modulate the interrupting noise intervals. A benefit by a factor of 1.5 is even observed for restoring sentence-level amplitude modulations provided by changes in the overall amplitude of individual vowels across the entire sentence. In contrast, only minor to no improvement was observed for modulating interrupting noise segments by the envelope of the missing consonants. Thus, it appears that vowel envelopes contribute essential cues for the intelligibility of interrupted sentences.

This is interesting due to the similarity of interrupting noise sources compared to consonant acoustics, many of which are characterized by aperiodic noise sources (Stevens, 2002). Shinn-Cunningham and Wang’s (2008) account of this effect was that listeners made a perceptual misassignment of noise-source modulations to that of the target speech. This would suggest that noise maskers, qualitatively more similar to consonant segments, would be more likely to result in errors in perceptual segregation. Indeed, consonant-replaced sentences were judged more continuous and resulted in better overall intelligibility. However, such an effect should also predict better improvements in intelligibility with the addition of noise modulation by consonant and not vowel envelopes. However, this was not observed. Even after degrading vowel spectral and temporal modulation cues through noise masking (Experiment 2), restoration of consonant envelopes during the interrupting noise impacted performance to a minimal degree. Instead, greater intelligibility improvements were observed when vowel modulations were added to preserved consonants. This was also associated with a large increase in the proportion of continuous judgments (i.e., the difference between low-level and modulated conditions during vowel-replaced sentences) and an increase in an acoustic index of envelope preservation.

These results can be summarized by stating that vowel-modulated noise restores the sentence envelope and perceived continuity of sentences preserving consonants, and results in improved sentence intelligibility. In contrast, sentences that remove the consonants maintain a relatively preserved sentence envelope and perceived continuity. Therefore, the addition of consonant-modulated noise results in significant but modest gains in sentence envelope preservation as well as improvements in perceived continuity that approach significance for this small sample. In this condition, perhaps due to only modest improvements, little to no benefit in intelligibility was observed, even after noise masking of the preserved vowel segments (Experiment 2).

The results are further surprising given previous work that has demonstrated significant contributions of the temporal envelope to consonant identification in nonsense syllables (e.g., Van Tasell *et al.*, 1987), particularly for voicing and

manner class distinctions (see also [Rosen, 1992](#)). Considering these previous results, although with very different methods and stimuli, it would be reasonable to expect improvements in intelligibility with restoration of consonant envelopes. However, consonant cues, such as manner and voicing, are restricted to phonemic identification, while the measures of continuity, intelligibility, and envelope preservation assessed here are more global measures of the sentence where context plays a significant role. It may be that these cues are redundant with information about neighboring consonants already preserved in the neighboring formant transitions of the preserved vowels (e.g., [Cooper et al., 1952](#); [Strange et al., 1983](#)). However, still no contributions to intelligibility were observed for consonant envelope modulations in Experiment 2 where preserved vowels were selectively noise-masked. These conditions successfully degraded vowel cues, as indicated by the large drop in average performance from 75% to 39% with the additional noise. In such conditions, it would be expected that additional, even redundant, phonemic cues for individual consonant identity would facilitate sentence recognition.

The results from these three experiments lead to the possible conclusion of fundamental differences between the information carried by modulations intrinsic to consonant and vowel segments. Consonants, on average, contain higher frequency components compared to vowels. These high-frequency bands are characterized by having fast-rate envelope modulations, compared to the mid- and low-frequency bands ([Greenberg et al., 1998](#)). This acoustic difference in modulation rate likely underlies some of the differences in the perceptual contribution of such segments. The fast-rate modulations of consonants may be constrained to phonemic-level interpretations. In contrast, slow-rate modulations intrinsic to vowels appear capable of indexing slower amplitude changes across the entire sentence. These sentence-level amplitude modulations are less likely to convey cues specific to phonemic identity.

Nevertheless, they are essential to maximizing speech recognition during long speech utterances by providing suprasegmental cues about speech rate, stress, and rhythm. In isolated word and nonword contexts where phoneme identity is the primary determiner of performance, such slow rate changes from vowels have a greatly reduced role in speech intelligibility ([Fogerty and Humes, 2010](#); [Fogerty et al., 2012](#)), whereas consonant contributions are essential ([Owren and Cardillo, 2006](#)).

Given the numerous differences between consonants and vowels, there are a multitude of plausible explanations that could potentially explain performance favoring vowel segments in sentences. Several previous investigations have examined factors that could be related differences between consonants and vowels on the basis of (1) acoustics ([Fogerty and Humes, 2012](#); [Stilp and Kluender, 2010](#); [Chen and Loizou, 2012](#); [Fogerty, 2013](#)), (2) context ([Owren and Cardillo, 2006](#); [Fogerty and Humes, 2010](#); [Fogerty et al., 2012](#)), (3) distributional probabilities ([Stilp, 2011](#)), (4) language ([Chen et al., 2013](#)), (5) durational differences ([Fogerty and Kewley-Port, 2009](#); [Fogerty, 2013](#)), among others. Durational differences are of particular importance as

these differences could potentially impact both the independent contributions of the segments themselves as well as the properties of the interruption paradigm. This could be problematic if perceptual factors related to the interruption properties confound interpretations of segmental importance. This was specifically investigated by [Fogerty and Kewley-Port \(2009\)](#) that examined properties of the interruption paradigm independent of the actual segments preserved. Their results demonstrated that interruption rate and duty cycle (or glimpse duration) differences between consonant and vowel conditions did not result in differences in performance. Rather, performance was dictated by the type of segment preserved. This was further supported by an acoustic analysis that demonstrated performance differences between consonants and vowels are best explained by the preservation of the sentence temporal envelope, not the duration of speech preserved ([Fogerty, 2013](#)). In addition, [Fogerty and Humes \(2012\)](#) demonstrated the importance of the temporal envelope for providing sentence-level vowel cues by selectively degrading acoustic properties across the entire sentence. Combined with these previous investigations, the current study empirically supports the importance of vowel temporal envelope cues, demonstrating improved intelligibility when these cues are re-introduced during interrupted intervals.

VI. SUMMARY AND CONCLUSIONS

The current experiment investigated modulation cues for sentence intelligibility carried by consonants and vowels at two different time scales: Slow modulation across the entire sentence and fast modulation that occurs within segments. The contribution of these modulations was investigated in segmentally interrupted speech using noise maskers. The following results were observed.

- (1) When the consonants were preserved, modulations at the level of the sentence (segment-level noise) and the phoneme (segment-modulated noise) present during vowel portions progressively improved performance with greater preservation of the temporal envelope.
- (2) In contrast, less benefit was observed by adding information regarding the overall level and modulation of consonants to existing vowels.
- (3) When preserved vowel cues were degraded through additional simultaneous noise masking, modulation of the replaced consonant still had limited contributions to intelligibility, particularly compared to the overwhelming importance of vowel modulations.
- (4) While consonant-replaced sentences were judged to be relatively continuous, listeners received minimal intelligibility benefit from the introduction of consonant modulations. In contrast, the presence of vowel modulation was an important predictor of perceived sentence continuity and was associated with better sentence intelligibility.

Combined, these results suggest that the broadband temporal modulations contained within vowels significantly contribute to intelligibility of interrupted sentences. Performance was poorest when vowel segments were maximally degraded

by replacement with a low-level noise. This type of noise maximally degraded the preserved temporal envelope (see acoustic analysis of Fig. 2) and was also associated with the greatest disruption of sentence continuity (Fig. 4). Restoring amplitude properties of vowels by modifying the level or temporal envelope of the replacement noise improves sentence intelligibility, possibly through the restoration of global, suprasegmental cues to sentence intelligibility. In contrast, modulation of consonant segments appear to contribute minimal additional cues to sentence intelligibility above and beyond what might already be conveyed during preserved vowel portions, even after noise-masking vowels. Overall, changes in vowel level and vowel modulation are important sources of information for the intelligibility of interrupted sentences.

ACKNOWLEDGMENT

This work was supported, in part, by a grant from the National Institutes of Health, NIDCD R03-DC012506.

- 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.
- Bashford, J. A., Jr., Warren, R. M., and Brown, C. A. (1996). "Use of speech-modulated noise adds strong 'bottom-up' cues for phonemic restoration," *Percept. Psychophys.* **58**, 342–350.
- Chen, F., and Loizou, P. C. (2012). "Contributions of cochlea-scaled entropy and consonant-vowel boundaries to prediction of speech intelligibility in noise," *J. Acoustic. Soc. Am.* **131**, 4104–4113.
- Chen, F., Wong, L. L., and Wong, E. Y. (2013). "Assessing the perceptual contributions of vowels and consonants to Mandarin sentence intelligibility," *J. Acoust. Soc. Am.* **134**, EL178–EL184.
- 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.
- Drullman, R., Festen, J. M., and Plomp, R. (1994). "Effect of temporal envelope smearing on speech reception," *J. Acoust. Soc. Am.* **95**, 1053–1064.
- Fogerty, D. (2011). "Perceptual weighting of the envelope and fine structure across frequency for sentence intelligibility: Effect of interruption at the syllabic-rate and periodic-rate of speech," *J. Acoust. Soc. Am.* **130**, 489–500.
- Fogerty, D. (2013). "Acoustic predictors of intelligibility for segmentally interrupted speech: Temporal envelope, voicing, and duration," *J. Speech Lang. Hear. Res.* **56**, 1402–1408.
- Fogerty, D., and Humes, L. E. (2010). "Perceptual contributions to monosyllabic word intelligibility: Segmental, lexical, and noise replacement factors," *J. Acoust. Soc. Am.* **128**, 3114–3125.
- Fogerty, D., and Humes, L. E. (2012). "The role of vowel and consonant fundamental frequency, envelope, and temporal fine structure cues to the intelligibility of words and sentences," *J. Acoust. Soc. Am.* **131**, 1490–1501.
- Fogerty, D., and Kewley-Port, D. (2009). "Perceptual contributions of the consonant-vowel boundary to sentence intelligibility," *J. Acoust. Soc. Am.* **126**, 847–857.
- Fogerty, D., Kewley-Port, D., and Humes, L. E. (2012). "The relative importance of consonant and vowel segments to the recognition of words and sentences: Effects of age and hearing loss," *J. Acoust. Soc. Am.* **132**, 1667–1678 (2012).
- Freyman, R. L., Nerbonne, G. P., and Cote, H. A. (1991). "Effect of the consonant-vowel ratio modification on amplitude envelope cues for consonant recognition," *J. Speech 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 CD-ROM, National Institute of Standards and Technology, NTIS Order No. PB91-505065.
- Gilbert, G., Bergeras, I., Voillery, D., and Lorenzi, C. (2007). "Effects of periodic interruptions on the intelligibility of speech based on temporal fine-structure or envelope cues," *J. Acoust. Soc. Am.* **122**, 1336–1339.
- Greenberg, S., Arai, T., and Silipo, R. (1998). "Speech intelligibility derived from exceedingly sparse spectral information," in *Proceedings of the International Conference on Spoken Language Processing*, Sydney, Australia, pp. 2803–2806.
- Jin, S.-H., and Nelson, P. B. (2010). "Interrupted speech perception: The effects of hearing sensitivity and frequency resolution," *J. Acoust. Soc. Am.* **128**, 881–889.
- 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.
- Nelson, P. B., and Jin, S.-H. (2004). "Factors affecting speech understanding in gated interference: Cochlear implant users and normal-hearing listeners," *J. Acoust. Soc. Am.* **115**, 2286–2294.
- Nelson, P. B., Jin, S.-H., Carney, A. E., and Nelson, D. A. (2003). "Understanding speech in modulated interference: Cochlear implant users and normal-hearing listeners," *J. Acoust. Soc. Am.* **113**, 961–968.
- 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.
- Port, R. F. (2003). "Meter and speech," *J. Phonetics* **31**, 599–611.
- Rosen, S. (1992). "Temporal information in speech: Acoustic, auditory, and linguistic aspects," *Philos. Trans. R. Soc. B* **336**, 367–373.
- Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., and Ekelid, M. (1995). "Speech recognition with primarily temporal cues," *Science* **270**, 303–304.
- Shinn-Cunningham, B. G., and Wang, D. (2008). "Influences of auditory object formation on phonemic restoration," *J. Acoust. Soc. Am.* **123**, 295–301.
- Singh, R., and Allen, J. B. (2012). "The influence of stop consonants' perceptual features on the Articulation Index model," *J. Acoust. Soc. Am.* **131**, 3051–3068.
- Stevens, K. N. (2002). "Toward a model for lexical access based on acoustic landmarks and distinctive features," *J. Acoust. Soc. Am.* **111**, 1872–1891.
- Stilp, C. E. (2011). "The redundancy of phonemes in sentential context," *J. Acoust. Soc. Am.* **130**, EL323–EL328.
- Stilp, C. E., and Kluender, K. R. (2010). "Cochlea-scaled entropy, not consonants, vowels, or time, best predicts speech intelligibility," *Proc. Natl. Acad. Sci. U. S. A.* **107**, 12387–12392.
- Strange, W., Jenkins, J. J., and Johnson, T. L. (1983). "Dynamic specification of co-articulated vowels," *J. Acoust. Soc. Am.* **74**, 695–705.
- Van Tasell, D. J., Soli, S. D., Kirby, V. M., and Widin, G. P. (1987). "Speech waveform envelope cues for consonant recognition," *J. Acoust. Soc. Am.* **82**, 1152–1161.
- Xu, L., Thompson, C. S., and Pfingst, B. E. (2005). "Relative contributions of spectral and temporal cues for phoneme recognition," *J. Acoust. Soc. Am.* **117**, 3255–3267.