Published in final edited form as:

J Phon. 2015 September ; 52: 89–104. doi:10.1016/j.wocn.2015.06.005.

Indexical properties influence time-varying amplitude and fundamental frequency contributions of vowels to sentence intelligibility

Daniel Fogerty

Department of Communication Sciences and Disorders, University of South Carolina, 1224 Sumter Street, Suite 300, Columbia, SC 29208, fogerty@sc.edu, 1-803-777-6050

Abstract

The present study investigated how non-linguistic, indexical information about talker identity interacts with contributions to sentence intelligibility by the time-varying amplitude (temporal envelope) and fundamental frequency (F_0). Young normal-hearing adults listened to sentences that preserved the original consonants but replaced the vowels with a single vowel production. This replacement vowel selectively preserved amplitude or F_0 cues of the original vowel, but replaced cues to phonetic identity. Original vowel duration was always preserved. Three experiments investigated indexical contributions by replacing vowels with productions from the same or different talker, or by acoustically morphing the original vowel. These stimulus conditions investigated how vowel suprasegmental and indexical properties interact and contribute to intelligibility independently from phonetic information. Results demonstrated that indexical properties influence the relative contribution of suprasegmental properties to sentence intelligibility. F_0 variations are particularly important in the presence of conflicting indexical information. Temporal envelope modulations significantly improve sentence intelligibility, but are enhanced when either indexical or F_0 cues are available. These findings suggest that F_0 and other indexical cues may facilitate perceptually grouping suprasegmental properties of vowels with the remainder of the sentence. Temporal envelope modulations of vowels may contribute to intelligibility once they are successfully integrated with the preserved signal.

Keywords

indexical cues; suprasegmentals; vowels; sentences; fundamental frequency; temporal envelope; intelligibility

1. Introduction

Speech is a spectrally and temporally varying signal composed of many different acoustic properties. The functional role of these acoustic properties varies widely, from providing cues for phonetic identity (Kent & Reed, 1992), to providing suprasegmental cues useful for defining linguistic structure (e.g., Cutler, 1994), to indicating non-linguistic information about the sex and identity of the talker (e.g., Bachorowski & Owren, 1999). These multiple

layers of information are conveyed simultaneously within the same acoustic signal. Early theories of speech perception posited independent and representationally distinct acoustic and perceptual sources of information for the linguistic and non-linguistic properties of speech (e.g., Jackson & Morton, 1984; McClelland & Elman, 1986). However, others have proposed that these two properties are integrally related within the same acoustic signal and are involved in the process of speech perception (e.g., Goldinger, 1998; Jusczyk, 1993; Pisoni, 1993). A comprehensive model of speech perception must incorporate how listeners integrate linguistic and non-linguistic information coded in the acoustic speech signal to support successful spoken communication (Nygaard, 2005). The main goal of the current set of experiments is to elucidate the interaction of these different sources of information within the acoustic structure of vowels for sentence materials.

1.1 Linguistic and non-linguistic sources of speech information

There is now significant literature on how linguistic and non-linguistic acoustic features provide information that is useful for speech perception and intelligibility (e.g., Bradlow, Torretta, & Pisoni, 1996; Nygaard & Pisoni, 1998; Pisoni, 1993). Furthermore, acoustic features associated with the identity of the talker (e.g., sex, age, dialect, vocal tract size), termed indexical cues (Abercrombie, 1967), are encoded along with acoustic cues essential for determining lexical meaning (Borrie, McAuliffe, Liss, O'Beirne, & Anderson., 2013; Goldinger, 1998; Mullennix, Pisoni, & Martin, 1989). Thus, talker variation that is captured by indexical information in the signal is processed as part of the normal speech perception process (Pisoni, 1993) where talker identity is retained in the linguistic representation (Johnson, 2005). However, little is known about the interdependency of indexical and lexical levels of speech information for the intelligibility of sentences. There is a particular gap in understanding how acoustic properties jointly encode information about the speech source (i.e., the talker) and the message.

Studying the interdependency of these levels of information is somewhat problematic given that the acoustic constituents of lexical and phonetic information simultaneously encode indexical information about the speaker. For example, it has long been known that formant frequencies of F1 and F2 are essential for coding the phonetic identity of vowels (Helmholtz, 1885), but they also capture the large variability among individual talkers (Peterson & Barney, 1952). Thus, while the acoustic-phonetic properties of speech are essential for the perception of phonetic identity and later word recognition processes, systematic variation in these acoustic properties is informative about the speaker. Furthermore, individual (i.e., indexical) variation in these acoustic-phonetic details may influence the phonetic categorization and lexical processing, as predicted by exemplar-based theories (e.g., Goldinger, 1998).

There is also interplay between levels of lexical information. Phonetic cues to sound categories interact with prosodic cues that organize the sentence into structures such as words, intonational phrases, and prominence relations (Beckman, 1996). However, this phonetics-prosody interface (see Cho, 2011) is bi-directional; fine-grained phonetic detail informs and is interpreted by higher-level prosodic structure (e.g., Beckman, 1996; Byrd & Choi, 2010). This multi-level information is acoustically realized by duration, fundamental

frequency (F_0), and amplitude cues (Culter, 2005). Significant recent focus has been on the time-varying dynamics of this acoustic information (see Mücke, Grice, & Cho, 2013, as well as Port, 2003). Therefore, the current study is concerned with the contribution of time-varying dynamic F_0 and amplitude envelope (here discussed in terms of the temporal envelope) to sentence intelligibility. Durational changes across the sentence were explicitly controlled.

1.2. Purpose of the current study

The purpose of the current study was to specifically investigate the interaction between lexical (phonetic and prosodic) and indexical levels of processing for sentence intelligibility. These levels of the linguistic message were investigated by using a vowel replacement paradigm. Indexical, suprasegmental, and phonetic sources of information are all highly salient within vowels, and this replacement paradigm afforded an experimental separation of indexical and suprasegmental cues from cues to phonetic identity. The experimental procedure systematically replaced vowels with vowels from a different speaker (Experiment 1), the same speaker from a different utterance (Experiment 2), or acoustically morphed the original vowels to remove time-varying formant cues to phonetic identity (Experiment 3). All conditions removed information about the changing vowel identity across the utterance by using a single vowel for all replacements within the sentence.

Vowel replacement studies have a long history in the literature due to their utility in identifying how listeners perceptually use the preserved acoustic features (e.g., Strange, Jenkins, & Johnson, 1983; Jenkins, Strange, & Edman, 1983; Fox, Wall, & Gokcen, 1992; Kewley-Port, Burkle, & Lee, 2007; Donaldson, Rogers, Cardenas, Russel, & Hanna, 2013). The contribution of suprasegmental properties of F_0 and amplitude modulations (e.g., related to stress, duration, and speech rate) were investigated here by systematically restoring these cues through digital signal processing of the replacement vowels. In this way, the contributions of individual acoustic constituents of the sentence prosody could be investigated independent from phonetic contributions of the segment identity. The current experiment hypothesized that the preservation of indexical information from using a speaker's own productions for replacement (Experiments 2 and 3) would be critical for the ability of listeners to capitalize on restored suprasegmental cues. When indexical information is replaced (Experiment 1), it was expected that listeners would experience greater difficulty using restored suprasegmental cues to facilitate sentence intelligibility. Indexical continuity of the sentence (including during the replaced vowel) is important for perceptually assigning suprasegmental cues during the replacement vowel to the target utterance. When indexical cues are discontinuous, the replacement vowel may be interpreted as an external, interrupting signal. This study also investigates how prosodic structure facilitates intelligibility independently of phonetic information (i.e., vowel identity). This later contribution was specifically investigated in Experiment 1 by examining word recognition as a function of the prosodic structure.

1.3 Acoustic cues during vowels to sentence intelligibility

Vowels were chosen as the phonetic target in the current set of experiments because of the convergence of segmental, suprasegmental, and indexical information (Ladefoged & Disner,

2012). For example, vowel formant frequencies provide information about phonetic identity (Fry, Abramson, Eimas, & Liberman, 1962) for both vowels and consonants (Cooper, Delattre, Liberman, Borst, & Gerstman, 1952), changes in the relative amplitude of vowels across the sentence provide suprasegmental cues to speech rhythm (Port, 2003), and acoustic correlates of vocal tract length during vowels are important for talker identity (Bachorowski & Owren, 1999).

1.3.1 Contributions of the temporal amplitude envelope—The temporal envelope provides information about both consonants and vowels (e.g., Shannon *et al.*, 1995; Xu *et al.*, 2005) and may code important cues for phoneme identification (e.g., Van Tasell *et al.*, 1987; Rosen, 1992; Shannon *et al.*, 1995). However, it is unclear how these cues contribute to overall sentence intelligibility. Investigations of sentence materials have suggested that the temporal envelope present during vowel segments is a significant contributor to intelligibility (Fogerty, 2013, 2014; Fogerty & Humes, 2012). Indeed, temporal envelope-based measures of speech intelligibility (i.e., changes in relative rms levels across the sentence) better predict intelligibility for vowel-dominated sentences compared to consonant-dominated sentences (Chen & Loizou, 2012). It is clear that the temporal envelope operates at short (i.e., phonetic) and long (i.e., prosodic) time scales (Greenberg, Carvey, Hitchcock, & Chang, 2003) which feature prominently in some neural models of speech perception (e.g., Poeppel, 2003; Gihtza & Greenberg, 2009). Furthermore, vowels feature prominently in several temporal accounts to prosodic structure (e.g., Greenberg *et al.*, 2003; Port, 2003) which may set the framework for interpreting other phonetic segments within the syllable (Greenberg *et al.*, 2003). However, it is not clear how this information may (or may not) interact with indexical properties of the speech signal. The current study extends this investigation to determine how indexical properties of the talker impact how listeners use temporal envelope modulations present during vowel segments.

1.3.2 Contributions of fundamental frequency— F_0 variations also play an important role for sentence intelligibility (e.g., Laures & Weismer, 1999). Traditionally, F_0 has played a prominent role in defining prosodic structure (e.g., Fry 1958; Liberman & Pierrehumbert, 1984), although its relative contribution compared to amplitude cues is debated (e.g., Greenberg, 1999). However, F_0 may also play a significant role in perceptual grouping; enabling a talker's message to be followed during interrupted speech (e.g., Darwin, Brungart, & Simpson, 2003). Importantly, F_0 can exert considerable influence on the perceptual grouping of formant frequencies (Darwin, 1981) and therefore influence the perceived phonetic identity (Gardner, Gaskill & Darwin, 1989).

The current study explores the potential role of F_0 cues by selectively distorting or restoring F_0 -modulation of replaced vowels. As F_0 is used in perceptually grouping formant information, one possible outcome is that preserving the original F_0 -modulation will decrease performance, causing listeners to be more likely to be confused by the contradictory phonetic information of the replacement vowel. In contrast, preserving F_0 -modulation may enhance performance by contributing to a number of other processes known to be important for sentence intelligibility. For example, F_0 -modulations support perceptual continuity (Darwin *et al.*, 2003), indexical information (Johnson, 2005), and

contribute to the prosodic structure of the sentence (Lehiste, 1970). This study tested these alternate hypotheses.

1.4 The role of indexical properties in the current study

Indexical information may, in part, determine the relative importance of time-varying F_0 -modulation and E-modulation properties. For example, indexical cues may facilitate the integration of acoustic features belonging to the attended message and influence how a listener processes available speech cues. Indeed, expectations of talker identity can influence speech perception (Magnuson & Nusbaum, 2007). Therefore, indexical properties may facilitate binding the multidimensional acoustic properties of speech. In the presence of conflicting indexical cues, linguistic processing of concurrent prosodic or phonetic cues could be compromised. The hypothesis of this study is that replacement vowels that preserve indexical information (i.e., are spoken by the same talker) maximize listeners' ability to use concurrent F_0 and amplitude envelope cues, as compared to replacement vowels spoken by a different talker.

Previous investigations have demonstrated that listeners are able to use temporal modulation cues of the target sentence that are restored during interrupted noise maskers (Bashford, Warren & Brown, 1996; Shinn-Cunningham & Wang, 2008), particularly if those modulations are present during vowels (Fogerty, 2014). Shinn-Cunningham and Wang (2008) proposed that listeners were able to use these modulations by perceptually grouping these cues with the target sentence. If this perceptual grouping does take place, then it is likely that acoustic similarity to the target sentence (i.e., preservation of the talker's indexical properties) would influence the success by which restored acoustic properties facilitate speech intelligibility.

2. Overview of the experiments

2.1 Definition of experimental conditions across experiments

This set of experiments used a vowel-replacement paradigm where all of the vowels in the sentence were excised and replaced with a single vowel. The source of this replacement vowel was varied across a series of three experiments to preserve indexical properties of the target talker to different degrees (see table 1). Experiment 1 used a replacement vowel spoken by a different talker that was matched in gender. Experiments 2 and 3 used a replacement vowel spoken by the target talker saying a different sentence (Exp. 2) or the same sentence resynthesized with flattened formant frequencies (Exp. 3). These experiments replaced vowel indexical information to different degrees.

The experiments also investigated the contribution of vowel F_0 -modulation and E-modulation cues. These manipulations could potentially influence concurrent indexical cues. However, as these manipulations were consistent across experiments, differences between experiments are a reflection of indexical quality of the replacement vowel. Table 2 summarizes how acoustic properties were modified across the experiments. Vowel duration was always preserved via time compressing or expanding the replacement vowel to match the duration of the original vowel. In addition, the mean F_0 of the original vowel was also preserved in all but one condition tested in Experiment 2. The remaining three acoustic

measures highlighted in Table 2 were varied according to specific questions regarding the importance of the intrinsic vowel F_0 contour or to the overall sentence-level temporal envelope present during vowels. The latter is captured by both the mean amplitude and intrinsic amplitude modulation of the vowel, and were therefore varied concurrently in these experiments.

2.2 Confirmation of differences in indexical properties across experiments and conditions

Crucial to this set of studies is the relative preservation of indexical information that was first confirmed through a preliminary study completed by four young adult listeners with normal hearing (i.e., audiometric thresholds at octave intervals from 250–8000 Hz below 20 dB HL). All replacement vowels used in the experiments described here were excised from experimental stimuli after processing and paired with the original vowel. Vowel pairs were separated by 500 ms of silence and the original vowel was always presented first for comparison. Participants listened to each vowel pair and identified if the two vowels could have been spoken by the same talker, even though one of the vowels was distorted. Testing was completed in a sound attenuating booth targeting an average presentation level of 70 dB SPL. However, individual vowel levels varied based on their level in the experimental sentences. All experimental vowels from the three experiments reported here were presented once to the listeners, for a total of 2146 vowel pair judgments. Vowels from all experiments and experimental conditions were fully randomized. Testing was self-paced and listeners were encouraged to take frequent breaks as needed.

Results are plotted in Figure 1 according to the proportion of responses where listeners identified the two vowels as belonging to the same talker. The Flat F_0 condition (see table 2) specifically avoided introducing acoustic properties specific to the original vowel in the sentence and is therefore most representative of the similarity in intrinsic indexical properties across the Experiments. Paired-samples t-tests demonstrated significantly more indexical properties preserved in Experiment 3 compared to Experiment 1, $t(3) = 4.96$, $p = .016$, with Experiment 2 falling in the middle. This rank ordering across the experiments confirms that the stimulus processing procedures maintained the intended relative increase in indexical information across experiments.

Regarding the three main signal processing conditions, from the data using vowels from different talkers (Exp. 1), there is also a clear trend regarding increasing similarity to the original talker when acoustic properties of the original vowel are restored during the replaced interval. Therefore, indexical properties related to talker identity do appear to be coded by both dynamic F_0 and E-modulation amplitude cues that are intrinsic to the vowel segment. Interestingly, this trend was not maintained when the replacement vowels already were spoken by the original talker of the sentence (Exp. 2 and 3), presumably due to existing preservation of these indexical cues in the replacement vowel.

3. Experiment 1: Vowel replacement from a different talker spoken in a different context

The purpose of Experiment 1 was to investigate the relative contribution of F_0 -modulation and E-modulation when vowel indexical cues of the talker were replaced with vowels from a different talker.

3.1 Listeners

Eighteen young adult listeners ($M = 21$ yrs; 19–22 yrs.) participated in the study. All listeners passed a pure-tone hearing screen (i.e., audiometric thresholds at or below 20 dB HL at octave frequencies between 250 Hz–8000 Hz), were not familiar with any of the sentence materials, and were native speakers of American English. Two participant groups were tested on the same set of conditions that examined F_0 and E-modulation with a single replacement vowel. The two groups differed only in the phonetic identity of the replacement vowel, either /ə/ (N=8) or /æ/ (N=10).

3.2 Stimuli

Forty-two sentences (21 male and 21 female talkers) were selected from the TIMIT database (Garofolo *et al.*, 1993) to be used in the study. Sentences contained 6–10 words with an average of four phonemes/word. Each sentence contained an average of 33 phonemes (11 vowels and 22 consonants). Sentences were divided into three lists of 14 sentences for the three experimental conditions, each list containing 114–115 words with an average total of 163 vowel segments. The three sentence lists were balanced for number of words and talker gender. This sentence set has been used in previous studies to examine the intelligibility of sentences with noise-replaced segments (e.g., Fogerty & Kewley-Port, 2009) and has the advantage of providing time markings specified by experienced phoneticians for all phonetic boundaries in the sentence. The TIMIT phonetic classification of American English vowels was used with the inclusion of vowel+/r/ as a single rhotacized vowel and intervocalic glottal stops, which were also included as vocalic segments (consistent with previous segmental investigations; Kewley-Port *et al.*, 2007). Sentence intelligibility for the full TIMIT sentences presented in quiet to young normal-hearing adults is 99% correct (Kewley-Port *et al.*, 2007), indicating a high level of intelligibility for the original recordings.

In order to isolate suprasegmental properties from phonetic acoustic cues, all vowels in the sentence were replaced with a single vowel. One group was tested using a neutral /ə/ vowel and a second group using the /æ/ vowel. The two different vowels (/ə/ and /æ/), presented to two different listener groups, were selected to examine the interaction between the phonetic identity of the replacement vowel and the lexical selection of the words in the sentence. This vowel difference allowed for the investigation of how listeners may have used phonetic information of the replacement vowel to fill-in missing phonetic cues in the target sentences. Such factors were investigated by examining listener responses to new words created by the replacement (e.g., “lack” for “like”) as well as the accuracy for replacement items that happened to preserve the original vowel identity (i.e., /ə/ replaced /ə/ in the original word).

A male and a female production of each vowel was selected from the Hillenbrand vowel database (Hillenbrand *et al.*, 1995) in order to preserve acoustic cues to talker gender in the replacement vowel. The use of these vowels during replacement specifically replaced talker-specific indexical cues (i.e., talker identity, dialect, emotional state, etc.) during the replacement interval. Replacement vowels were spoken in the “hVd” context and selected such that they had 100% identification in the Hillenbrand database and had formants near the median values for the talkers recorded in the database. The two representative vowels from the male and female talkers were used as the replacement vowel in the sentences for male and female talkers, respectively (i.e., the replacement vowel matched the gender of the talker for each sentence).

Use of these replacement vowels specifically removed cues provided by natural formant transitions into the vowel as well as information regarding the phonemic identity of the missing vowel. Other suprasegmental cues carried by the replaced vowels were selectively restored according to the signal processing conditions assigned to the listener group.

3.3 Signal processing

Replacement vowels were first excised from the identified “hVd” Hillenbrand recordings to remove the consonantal context. Vowel intervals were identified via dual displays of the waveform and spectrogram. Vowel onsets and offsets were identified at the start or end of voicing with clear energy in the first and second formants. A 10-ms spline onset/offset ramp was applied to the replacement vowel to facilitate continuity during replacement. It should be noted that this manipulation and the subsequent signal processing conditions applied to the replacement vowel could have impacted the overall intelligibility of the vowel production.

Next, sentences were processed to replace vowels using the segmental boundaries identified by the TIMIT database and adjusted within 1-ms to the nearest local minima (i.e., zero-crossing). Processing for all replacement vowels consisted of time compressing or expanding the vowel using STRAIGHT (Kawahara, Masuda-Katsuse, & de Cheveigné, 1999), a speech analysis and synthesis program, to preserve the duration of each replaced vowel in the sentence. STRAIGHT was also used to provide a high-fidelity representation of source and filter characteristics of the speech sample. F_0 coding of the speech source was extracted using instantaneous-frequency based measures that produce continuous F_0 contours. Spectral estimates of the filter specifically eliminated interferences of source periodicity to ensure high reproductive quality and clear source-filter separation without residual cues (see Kawahara *et al.*, 1999). Next, each replacement vowel underwent specific signal processing according to one of three conditions described by the procedures below. Figure 2 displays of the amplitude (top) and F_0 (middle) contours, as well as spectrographic displays (bottom) of an example sentence following signal processing.

1. *Flat F_0* : STRAIGHT was used to replace the F_0 contour of the replacement vowel with a flat F_0 centered at the mean F_0 of the original vowel being replaced.

2. *F₀-modulated*: The replacement vowel F_0 contour was replaced using STRAIGHT with the F_0 contour of the original vowel. The temporal envelope was left unmodified, and was identical to the Flat F_0 condition.
3. *E-modulated*: The wideband temporal envelope of the original vowel was extracted through half-wave rectification and low-pass filtering at 64-Hz using a 6th order Butterworth filter, which specifically removed temporal periodicity cues related to F_0 . This envelope was then used to modulate the amplitude of the replacement vowel and then scaled to the level of the original vowel being replaced. It is important to note that this procedure used the replacement vowel as the carrier signal that was then modulated by the wideband envelope of the original vowel. To avoid concurrent contributions from F_0 , the replacement vowel was also manipulated to have a flat F_0 contour at the mean of the original vowel as in the Flat F_0 condition.

Following condition-specific processing, all replacement vowels were scaled to the average root-mean-square level of the target sentence. This scaling specifically avoided providing slow-rate E-modulation cues. Furthermore, using a constant level during the vowel intervals resulted in a more direct comparison to previous studies with noise replacement (Fogerty, 2014). The one exception to this vowel scaling procedure was for the E-modulated vowels that preserved the original level of each replaced vowel as these level differences are an amplitude cue provided by the temporal envelope. This maintained slow-rate E-modulation cues for this condition. A 4-ms cosine ramp was applied at the onset and offset of the replacement vowel to avoid introduction of any transient artifacts. The original TIMIT and Hillenbrand speech materials were originally recorded at a sampling rate of 16000 Hz and a bit rate of 16 bits. These sampling characteristics were maintained during signal processing and presentation.

3.4 Procedure

Listeners were seated in a sound-attenuating booth and listened to the stimuli monaurally via the right headphone (Sennheiser HD 280 PRO). The level of the speech (prior to replacement) was calibrated to be presented at 70 dB SPL using a Larson-Davis 800B sound level meter with linear weighting. Sentences from all three signal processing conditions were presented in a random order to listeners. Listeners were provided with a short demonstration of sentences processed according to these signal processing conditions to gain familiarity with the task prior to testing. Listeners were instructed to repeat each sentence and were encouraged to guess any real words they think they may have heard. Sentence responses were audio recorded and analyzed offline by trained raters who demonstrated accurate transcriptions on a training response set. 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.

3.5. Results

3.5.1 Investigation of experimental conditions—All word percent-correct scores, here and in subsequent experiments, were transformed to rationalized arcsine units to

stabilize the error variance prior to analysis (RAU; Studebaker, 1985). Figure 3 displays the mean performance for the two listener groups. A mixed-model Analysis of Variance (ANOVA) was conducted with the three stimulus conditions as the repeated-measure and the vowel group (/ə/ versus /æ/ replacement) as the between-subjects variable. Results demonstrated a significant main effect for the stimulus condition, $F(2,32) = 13.9, p < .001$, with no main effect of vowel group (i.e., (/ə/ versus /æ/ replacement), $F(1,16) = 1.5, p > .05$, or interaction, $F(2,32) = 2.5, p > .05$. As the phonetic identity of the vowel used during replacement did not significantly influence performance (effect of vowel group with $p > 0.2$), these groups were pooled for subsequent analyses.¹

A moderate improvement for adding F_0 modulation ($M = 24$ RAU, $SD = 13$ RAU) to the replacement vowel was observed in contrast to the flat F_0 condition ($M = 16$ RAU, $SD = 10$ RAU), $t(17) = 4.5, p < 0.001$, Cohen's $d = 0.65$. Adding envelope modulations ($M = 27$ RAU, $SD = 8$ RAU) also improved performance compared to the flat F_0 condition, $t(17) = 6.2, p < 0.001, d = 1.0$. No significant difference was obtained between F_0 and E-modulation conditions.²

3.5.2 Investigation of prosodic structure—The prosodic structure of the sentence provides high-level information that impacts the processing of lower level properties of the sentence, such as phonetic categorization (e.g., Byrd & Choi; 2010; Kim & Cho, 2013) that ultimately influences the intelligibility of the sentence (see Cho, 2011) such as in cases of lexical ambiguity (Cho *et al.*, 2007; Salverda *et al.*, 2003). Such ambiguity in the phonetic and prosodic structure of the sentence was introduced in the current study by replacing phonetic segments. To investigate this potential effect, word recognition scores were calculated across the prosodic structure of the sentence. This analysis explicitly investigated how F_0 -modulation and E-modulation cues to prosodic structure may have impacted sentence intelligibility scores.

Two raters independently identified and then reached consensus on intonational phrase boundaries and the primary (nuclear) prominence within each experimental sentence. Word recognition scores were then calculated for the pooled group of eighteen listeners for prominent words within the sentence as well as for words that occurred at the start, middle, or end of intonational phrases. There were 22–24 prominent words and 21–24 intonational phrases per experimental condition. These scores are displayed in Figure 4. All analyses were conducted on the RAU transformed scores. As a significant main effect of the acoustic condition was observed in the previous analysis, the results for each acoustic condition were analyzed separately.

¹Potential confusions related to the replaced phonetic identity were also analyzed. Participant productions of nonwords did not occur and alternative real word productions (e.g., “luck” for “like”) were rare (2% of responses for either replacement vowel) and distributed across the different signal processing conditions. Therefore, listeners do not appear to have integrated the phonetic identity of the replacement vowel with the preserved speech segments. Performance was primarily determined by the acoustic processing of the replacement vowel.

²Talker gender was also analyzed, with female speakers resulting in a 20.8 RAU improvement over male speakers, possibly due to a greater number of pitch periods per vowel (Assmann, 1999). However, talker gender did not interact with the acoustic processing conditions, indicating similar contributions of F_0 and E-modulation cues for both male and female speakers.

Prominence: Paired-samples t-tests investigated the effect of word prominence for each of the three acoustic conditions. Results demonstrated that prominent words were recognized significantly more successfully than non-prominent words only for the F_0 -modulation condition, $t(17) = 2.8$, $p < .05$. This suggests that higher word recognition scores for prominent words may be dependent on the availability of time-varying F_0 information within the vowel, and does not significantly rely on amplitude cues.

Intonational phrase: Each acoustic condition was investigated independently using a repeated-measures ANOVA to examine the effect of word position within the intonational phrase. No significant effect of word position was observed for Flat F_0 and F_0 -modulation conditions ($p > .05$), although there was a trend toward better performance for the first word of the intonational phrase. A main effect for word position was observed for the E-modulation condition, $F(2,34) = 15.6$, $p < .001$. Subsequent paired-sampled t-tests in the E-modulation condition demonstrated better word recognition for the first, $t(17)=3.7$, $p < .01$, and final words, $t(17) = 7.7$, $p < .001$, in the intonational phrase compared to words in the middle of the phrase. No significant difference was obtained between the first and final words ($p > .05$).

Of particular interest is whether the available acoustic properties influence how word recognition accuracy varies as function of the position of the word in the intonational phrase. In order to control for differences in the overall level of accuracy determined by the acoustic condition, difference scores were calculated using accuracy for words in the middle of the phrase as the baseline. A 3 (acoustic condition) $\times$ 2 (initial or final difference score) repeated measures ANOVA demonstrated a significant main effect of the acoustic condition, $F(2,34) = 12.0$, $p < .001$, with no significant main effect of position or interaction ($p > .05$). Post-hoc paired-samples t-tests with Bonferroni correction for multiple comparisons demonstrated that differences between the acoustic conditions were confined to the final word. E-modulation resulted in better final word recognition than either F_0 -modulation, $t(17) = 6.4$, $p < .001$, or Flat F_0 conditions, $t(17) = 3.3$, $p = .004$. Overall, these results suggest that when accurate time-varying amplitude information is provided, listeners perform better at identifying words at the end of intonational phrases when durational cues are maintained.

3.6 Discussion of experiment 1 replacing indexical cues

Results demonstrate that when indexical cues about the target talker are not available within the replaced vowel segments, listeners do make use of vowel-intrinsic F_0 modulation or E-modulation cues to facilitate intelligibility of the sentence. Recall that the F_0 -modulation condition only altered modulations within the vowel, but preserved mean vowel F_0 differences. Thus, listeners may be particularly sensitive to the restoration of fine-grained F_0 variations when other indexical cues are replaced.

The effect of the phonetic identity of the replacement vowel on phonetic selection of the original segment was specifically investigated through the use of two different replacement vowels, /ə/ and /æ/. For these sentences, /ə/ and /æ/ accounted for 6.0% and 5.8% of the total duration of vowels across the sentence, respectively, with a slightly greater occurrence of /ə/ compared to /æ/ (6.5% versus 3.9%). However, no significant main effect or

interaction with the signal processing condition was observed between these two vowels. Furthermore, error analyses demonstrated that listeners made very little-to-no use of phonetic cues that were provided by the replacement vowel, either to facilitate word recognition when it preserved the phonetic identity of the original vowel, or to select alternative lexical candidates that were formed by the phonetic identity of the replacement vowel. Overall, listeners appear to have processed these vowel-replaced sentences independently of the phonetic content of the replacement vowels. Instead, listeners relied on the preserved acoustic phonetics of the consonant segments and any prosodic information restored during signal processing. Evidence for this was supported by investigating word recognition as a function of the prosodic structure of the sentence. Results suggest different contributions of time-varying F_0 and amplitude cues, with F_0 contributing more to the intelligibility of prominent words, and amplitude contributing to final words in intonational phrases.

4. Experiment 2: Vowel replacement from the same talker spoken in a different context

4.1 Experiment 2a: Vowel level normalized to sentence level

The purpose of Experiment 2a was to investigate how listeners are able to make use of F_0 -modulation and E-modulation of the replaced vowels when speaker-specific indexical information is provided.

4.1.1 Listeners—Nine young, normal hearing listeners ($M = 24$ yrs; 20–24 yrs) participated in the study. All listeners passed a pure-tone hearing screen at octave frequencies (below 20 dB HL from 250–8000 Hz), were not familiar with any of the sentence materials, and were native speakers of American English.

4.1.2 Signal processing and procedures—Indexical talker-specific information across the sentence was preserved by using a replacement vowel spoken by the same talker in a different sentence context. The vowel from the “hVd” context in the sentence “She had your dark suit in greasy wash water all year” was excised from each talker’s recording using the phonetic boundaries identified in TIMIT. This effectively created a catalogue of 42 talker-specific vowels: one replacement vowel for each talker excised from the same sentence context. Variability in talker productions resulted in phonetic transcriptions of the excised vowel as either /æ/ or /ə/, dependent on the talker. As in Experiment 1, all of the vowels in the target sentence were replaced by a single vowel spoken by the same talker as that of the target sentence (see Figure 5). Participants completed testing for Flat F_0 , F_0 -modulated, and E-modulated conditions using identical signal processing procedures to Experiment 1, with the only exception being the talker-specific replacement vowel that was used.

As with Experiment 1, all replacement vowels were scaled to the average root-mean-square level of the target sentence with the exception of E-modulated vowels that preserved the original level of each replaced vowel. A 4-ms cosine ramp was applied at the onset and

offset of the replacement vowel to avoid introduction of any transient artifacts. Calibration and testing procedures were identical to Experiment 1.

4.1.3 Results and discussion—Results for Experiment 2a are plotted in Figure 6. When a vowel produced by the target talker was used for replacement, F_0 modulation ($M = 14$ RAU, $SD = 16$ RAU) for this group of listeners actually reduced overall performance compared to the flat F_0 vowel ($M = 24$ RAU, $SD = 11$ RAU), $t(8) = 4.9$, $p > .01$, $d = 0.7$. However, E-modulation ($M = 33$ RAU, $SD = 17$ RAU) resulted in significantly better performance compared to F_0 -modulated, $t(8) = 11.3$, $p < 0.01$, $d = 1.2$, and Flat F_0 conditions, $t(8) = 3.4$, $p < .01$, $d = 0.6$.

Comparison of the same-talker (Exp. 2a) and different talker (Exp. 1) vowel replacement groups was conducted using a mixed-model ANOVA with the acoustic condition as a repeated measure and the experimental group as a between subjects factor. As expected, a main effect of the acoustic condition was found, $F(2,50) = 24.9$, $p < .001$, consistent with the earlier results. No significant main effect of group was observed. However, of particular interest is the interaction, which was significant, $F(2,50) = 17.3$, $p < .001$. This interaction suggests that the presence of indexical properties in the replacement vowel specifically alters how listeners use time-varying F_0 and amplitude cues.

When combined with the earlier results, these findings suggest that the presence of vowel indexical cues (1) facilitate sentence intelligibility in general, as better performance was obtained with indexical cues in the flat F_0 condition, (2) appear to provide no additional benefit to information already coded in the natural F_0 contour (i.e., vowel-intrinsic F_0 modulations may serve a similar function as indexical information), and (3) facilitate the use of E-modulation cues for sentence intelligibility, as performance in this condition was enhanced in the presence of indexical information. Finally, note that performance for the F_0 -modulated condition in Experiment 2a is actually poorer than in Experiment 1. In this condition, indexical cues were derived from a different sentence, while F_0 modulations were restored from the target sentence. The poor performance relative to replacing talker-specific indexical cues may reflect the processing of two cues that are potentially meaningful (i.e., originating from the target talker) and serve similar functions for intelligibility, but represent conflicting information from different sentence contexts. Alternatively, there may have been individual differences among the listeners in Experiment 2a that resulted in poorer overall performance. Some indication of this was provided with comparison to Experiment 2b.

4.2. Experiment 2b: Replacement vowel scaled to the original vowel level

Experiment 1 and 2a both presented replacement vowels at a normalized intensity level equal to the average level of the sentence. This effectively distorts the slow varying amplitude changes across the sentence. The E-modulation condition restored both slow-varying amplitude changes across the sentence and faster amplitude fluctuations intrinsic to the vowel segment. In order to parse out the contributions of these slow-varying and faster-varying amplitude modulations, a second experiment was run that maintained overall changes in segmental level across the sentence for all signal processing conditions. Thus, slow-varying changes were preserved and any benefit from the E-modulation condition, in

relation to other conditions, was a reflection of intrinsic amplitude modulations. In addition, to gain a better understanding of the interaction of E-modulated and F_0 -modulated cues, additional signal processing conditions were added for the current experiment.

4.2.1 Listeners—Sixteen young, normal hearing listeners ($M = 23$ yrs; 20–29 yrs) who met the eligibility criteria outlined in the earlier experiments participated in the study. Participants were evenly divided into two groups of eight listeners and were tested on conditions that primarily explored either F_0 -modulation or E-modulation.

4.2.2 Signal processing and procedures—The talker-specific replacement vowels used in Experiment 2a were again used here. As in Experiment 1, all of the vowels in the target sentence were replaced by a single vowel spoken by the same talker as that of the target sentence. Experimental conditions were divided among two listener groups. The *F₀-mod group* was tested on vowel-scaled versions (i.e., the original vowel intensity level was maintained) of the Flat F_0 and F_0 -modulated conditions described earlier, as well as a new “*Replaced*” condition that preserved the acoustic properties of the replacement vowel (i.e., the replacement vowel maintained its own F_0 and envelope properties; no acoustic restoration was provided). The *E-mod group* was tested on vowel-scaled versions of the F_0 -modulated condition (to provide comparison with the other listener group), E-modulated condition, and a new “*F₀ & E modulated*” condition that preserved the original vowel F_0 contour and was modulated by the original vowel wideband temporal envelope (i.e., underwent processing for F_0 -modulation and E-modulation to provide the greatest amount of restoration).

New to this Experiment, all replacement vowels were scaled to the root-mean-square level of the original vowel that was replaced. As before, a 4-ms cosine ramp was applied at the onset and offset of the replacement vowel to avoid introduction of any transient artifacts. Calibration and testing procedures were identical to Experiment 1.

4.2.3 Results and discussion—Results for the two listener groups of Experiment 2b are plotted in Figure 6. A between group comparison was first conducted for the scaled F_0 -modulation condition which was completed by both groups. No significant difference between groups was obtained, indicating that performance was consistent between the two groups (i.e., effects related to differences in conditions are robust, and not reflective of inherent group differences).

Next, within group comparisons were conducted to examine the contribution of the different acoustic properties. When a vowel produced by the target talker was used for replacement, F_0 modulation ($M = 26$ RAU, $SD = 9$ RAU) did not significantly improve performance over the unmodified replacement vowel ($M = 24$ RAU, $SD = 9$ RAU) or the flat F_0 vowel ($M = 29$ RAU, $SD = 6$ RAU). However, for talker-specific replacement vowels, E-modulation did have a large effect. E-modulation ($M = 39$ RAU, $SD = 8$ RAU) resulted in significantly better performance compared to F_0 -modulated vowels ($M = 26$ RAU, $SD = 7$ RAU), $t(7) = 3.9$, $p < 0.01$, $d = 1.6$. The F_0 & E-modulation condition ($M = 39$ RAU, $SD = 9$ RAU) did not result in any additional performance improvements compared to E-modulation alone, $p > .05$.

Finally, the results of Experiment 2b were compared to Experiment 2a that scaled the replacement vowel to the average level of the sentence, not to the original vowel level. Overall, performance in Experiment 2b was better, even for the E-mod condition that used the same stimulus files during testing. The difference between experiments is largely explained by the 5 percentage-point difference in E-modulated scores. This observation, combined with the larger variability in Experiment 2a due largely to a couple poor performing listeners, suggests that the lower performance in Experiment 2a could be attributed to individual variability in performance. However, even with this lower performance, independent samples t-tests between paired conditions in Experiment 2a and 2b demonstrated no significant effects of vowel level ($p > .05$). Therefore, E-modulation results observed in the current set of experiments may be largely due to the presence of faster-rate vowel-intrinsic amplitude modulations. In contrast, the slow-varying amplitude changes defined by overall vowel level appeared to have little influence on the overall results, contrary to previous results using noise replacement (Fogerty, 2014).

4.3 Discussion of experiment 2 preserving talker-specific indexical cues

This experiment preserved indexical cues of the target talker by replacing the vowels in the sentence by a vowel from the talker's own production in a different context. It was observed that listeners performed similarly in the F_0 -modulation condition regardless of whether indexical information was provided, but performed significantly better with indexical cues when F_0 -modulation was removed. These results suggest that F_0 -modulation may serve a partially redundant function with other indexical cues present during the vowel. Indeed, F_0 , along with vocal tract length, is a primary cue to talker sex (Bachorowski & Owren, 1999). However, having information about talker identity already encoded in the replacement vowel may reduce the importance of using F_0 modulation cues. In Experiment 2b, the presence of F_0 modulation along with E-modulated cues (i.e., F_0 & E modulated condition) also did not result in any improvements in intelligibility over E-modulated cues alone. Statistical classification of vowel segments suggests that acoustic characteristics related to the vocal tract, as opposed to the source (i.e., F_0), are more important for providing information about talker identity (Bachorowski & Owren, 1999). Preserving indexical cues by using talker-specific replacement vowels may enable listeners to perceptually group other acoustic features present during the vowel, such as E-modulation, to enhance intelligibility of the sentence.

5. Experiment 3: Vowel morphing based on the same talker and original production

Experiment 3 used a morphing procedure to reduce dynamic formant frequency cues provided in the vowel spectrum. These inherent dynamic cues of vowels are particularly important for providing context-dependent acoustic-phonetic cues (e.g., Cooper *et al.*, 1952) as well as for vowel identity (Hillenbrand, 2013). However, in the previous experiments these context-dependent cues contradicted the actual phonetic context of the replacement vowel. By removing these dynamic cues, listeners in Experiment 3 were required to use amplitude envelope and F_0 contour information during replacement intervals that now provided static "context-free" spectral information related to talker identity. Bachorowski

and Owren (1999) previously determined that talkers may be identified based on static measures at the midpoint of a single vowel, which was confirmed perceptually for these stimuli (see Figure 1).

5.1 Listeners

Ten young, normal hearing listeners ($M = 21$ yrs; 20–22 yrs) participated in the study. All listeners passed a pure-tone hearing screen at octave frequencies, were not familiar with any of the sentence materials, and were native speakers of American English. One listener was excluded from analysis due to previous experience with some of the experimental sentences and processing that resulted in outlier performance.

5.2 Signal Processing and Procedures

Signal processing of the sentences was completed in two phases: (1) morphing the vowel segments to reduce cues to phonetic identity and (2) restoring F_0 -modulation or E-modulation cues according to the signal processing procedures implemented in the previous experiments. Vowel morphing was completed using STRAIGHT. The same talker-specific vowels used in Experiment 2 were identified as template vowels for the morphing procedure. STRAIGHT was used to extract the filter characteristics at the middle time point of the template vowel, creating a spectral template (0–8000 Hz). The spectral template specifically removed distortions of signal periodicity in the spectral representation (i.e., filter characteristics were separated from the glottal source characteristics) using a pitch-adaptive spectral smoothing (Kawahara *et al.*, 1999). The filter characteristics of each vowel in the target sentence were then independently extracted and vowels were resynthesized using the spectral template. Importantly, this spectral template was used for the duration of the vowel, resulting in steady-state formant values. The effect of this signal processing was a replacement vowel with static spectral cues that were independent from the experimentally manipulated instantaneous F_0 trajectory. Resynthesized vowels were then processed according to one of the three signal processing strategies implemented in the previous experiments: (1) *Flat F_0* , (2) *F_0 -modulated*, or (3) *E-modulated*. Recall that both (1) and (3) used F_0 values that were flattened to the mean of each individual vowel in the sentence (i.e., each vowel had a different F_0 value determined by the mean of its original production). Figure 7 compares spectrograms from the original and vowel-morphed versions of an example sentence. Testing procedures were identical to the previous experiments.

5.3. Results

Multiple acoustic dimensions were manipulated in these experiments. However, these acoustic manipulations did not result in equal acoustic change. To facilitate comparison between conditions, an acoustic similarity index (similar to that implemented by Fogerty, 2013) was calculated between the original and processed sentence versions. First, the correlation between the original and modified amplitude envelope was calculated for each experimental sentence. The amplitude envelope of each version of the sentence was extracted via half-wave rectification and low-pass filtering at 64 Hz using a 4th-order butterworth filter. The Pearson correlation coefficient was then calculated between the two sentence versions. Second, correlation coefficients were also calculated between the F_0

contours extracted via STRAIGHT for the original and modified versions of the sentence. These coefficients were then transformed to z-scores using condition-specific distributions to provide a *relative* comparison between F_0 and amplitude envelope similarity. Standardized F_0 and amplitude envelope coefficients were then averaged across all sentences within an experimental condition to determine an average measure of acoustic similarity that equally weighted F_0 and envelope measures. As expected, acoustic similarity measures demonstrated that E-modulation and F_0 -modulation procedures both resulted in greater acoustic similarity to the original sentence, with E-modulation resulting in the greatest similarity due to additional amplitude scaling of the vowel level. These acoustic measurements are summarized in Figure 8, plotting the results for the three experiments.

Results for Experiment 3 were consistent with those of Experiment 2. Adding F_0 -modulation ($M = 11$ RAU, $SD = 12$ RAU) to the replacement vowel did not significantly improve performance above that obtained in the flat F_0 condition ($M = 8$ RAU, $SD = 9$ RAU), $t(8) = 1.2$, $p > 0.05$. However, adding envelope modulations ($M = 23$ RAU, $SD = 11$ RAU) significantly improved performance compared to the flat F_0 condition, $t(8) = 3.4$, $p < 0.01$, $d = 1.5$, and compared to the F_0 -modulation condition, $t(8) = 3.4$, $p < 0.01$, $d = 1.0$.

Examining the performance trends for the three experiments across the measure of acoustic similarity demonstrates a marked difference between Experiment 1 that used replacement vowels from a different talker, and Experiments 2 and 3 that used replacement vowels from the same talker. With replacement vowels from a different talker, F_0 -modulation resulted in significant improvement. In contrast, F_0 -modulation did not facilitate intelligibility when the replacement vowel was spoken by the same talker, even though the signal processing resulted in greater acoustic similarity to the original vowel. Finally, the introduction of E-modulation cues aided intelligibility, particularly when indexical information was available (Exp. 2 & 3 vs. Exp. 1).

5.4 Discussion of Experiment 3

These results suggest that the morphing procedure which retained the indexical properties of the original vowel segment resulted in similar findings as in Experiment 2 when a vowel from the same talker was used to replace the vowels in the sentence. These results again suggest that F_0 -cues are most useful in the presence of conflicting indexical information. The poorer performance in Experiment 3 is somewhat surprising given that the original vowel was preserved, albeit degraded. One possible reason for this discrepancy could be due to the systematic removal of inherent spectral change within the vowel. Vowel inherent spectral change plays a particularly important role in vowel recognition (see Hillenbrand, 2013 for an overview). In addition, reduction of formant movement within and across vowels in this study effectively removed information about the relative size of the vowel space, which can be an important predictor of speech intelligibility (Turner, Tjaden, & Weismer, 1995). Furthermore, patterns of vowel inherent spectral change can be used to describe differences in dialect and context (see Nearey, 2013). However, it is important to note that while spectral change was removed, other reliable cues for talker identification were preserved, particularly characteristics of the vocal tract transfer function, such as vocal tract length, relative formant amplitudes, and midpoint formant frequencies of a single

vowel (Bachorowski & Owren, 1999) as well as characteristics of the glottal source. These static cues were reliable for talker discrimination, as demonstrated in the initial talker discrimination perceptual measures summarized in Figure 1. However, these static acoustic correlates of talker identity likely underspecify the speech source in sentential contexts. The unnatural and unvarying spectral properties of the morphed vowels may have resulted in a difficulty perceptually assigning acoustic features of the replacement vowel to the same sound source as the remainder of the sentence. Instead, the dynamic acoustic specification of vowels may be essential for assigning interrupted speech fragments to the same sound source (i.e., talker). Indeed, dynamic formant transitions suppress auditory source segregation for rapid vowel sequences (Dorman, Cutting, & Raphael, 1975). This is particularly notable given that the inherent spectral change provided in Experiment 2 was from a different speech context. Thus, the potentially “incorrect” formant change from a different phonetic context may have been informative about the sound source and therefore useful for perception, unlike the underspecified static cues provided in Experiment 3 that may be more prone to source segregation.

6. General Discussion

6.1 Interaction of indexical information with the use of vowel fundamental frequency and temporal envelope cues

Overall, results demonstrate that talker-specific indexical properties influence how different acoustic properties are used by the listener. The presence of talker-specific indexical cues appears to assist perceptual grouping of the replacement vowel E-modulation with the target speech. In addition, no significant improvement over E-modulation alone was obtained when F_0 and E-modulated cues were both available, possibly suggesting that F_0 -modulation may serve a redundant, or supplementary function with other indexical cues, at least in terms of the assignment of E-modulation. This may be because acoustic properties related to the vocal tract filtering are more important for talker identification compared to acoustic properties related to the speech source (Bachorowski & Owren, 1999). This is further supported by the better performance that was observed for same talker compared to different talker vowels when F_0 and envelope cues were removed (i.e., Flat F_0 condition). Here, listeners performed better with indexical cues available and did not obtain any substantial benefit of restoring F_0 -modulation when indexical cues were present (only when they were replaced).

The results of these vowel replacement conditions were further supported by Experiment 3. In this experiment, spectral cues to vowel identity were removed by resynthesizing each vowel in the sentence with steady-state formants determined by the talker’s production. This morphing procedure preserved indexical information, such as vocal tract size, while removing phonetic cues. Results for this experiment paralleled the vowel replacement findings in Experiment 2 that also preserved indexical cues of the talker. Listeners benefited from E-modulation cues, but not F_0 -modulation cues.

In summary, introducing the talker’s own F_0 or indexical properties during the replacement vowel facilitates integration of the vowel with the remaining sentence. In contrast, E-modulation is more directly related to speech intelligibility and is best interpreted when the

replacement intervals are perceptually grouped with the speech signal via indexical cues (e.g., talker source and/or filter characteristics).

6.2 Interface with prosodic structure

Prosody influences lexical access, as phonetic coding is sensitive to prosodic structure (Keating & Shattuck-Hufnagel, 2002). Cho et al. (2007) have suggested that a “prosody analyzer” works in parallel with the acoustic-phonetic analysis of the signal. The current study suggests that such an analyzer may operate across the high-dimensional acoustic-prosodic space (e.g., Lehiste, 1970), where acoustic cues have different functional roles that support word recognition as a function of prosodic structure.

Prominent words typically have higher pitch and are associated with greater duration and intensity (Sluijter & van Heuven, 1996). Mean pitch and duration were preserved across signal processing conditions in this study. However, pitch movement on the lexically stressed syllable is also often associated with prominence as well (see Terken & Hermes, 2000), which may have resulted in better performance for prominent words when the F_0 trajectory was restored. Prominent words can also be viewed as strong points within the metrical structure of the sentence (see Calhoun, 2010). The prominent words identified in the current study correspond with the primary or “nuclear” accent of the sentence (Calhoun, 2010). The reduction of temporal amplitude cues through normalization of vowel level may have reduced cues about the relative prominence of the sentence. However, this processing would have maintained the syllabic rhythm of the sentence, while other levels of the rhythmic hierarchy (Liberman & Prince, 1977) associated with slower stress patterning or faster phonetic cues would have been distorted (see Leong, Stone, Turner, & Goswami, 2014). It may be that restoring temporal amplitude cues did not improve the intelligibility of prominent words in this study due to the preservation of syllabic rate across signal processing conditions.

Regarding analysis of the intonational phrase, E-modulation cues were particularly important for determining final word accuracy. On the surface this contrasts with previous findings. While pitch, duration, and amplitude play important roles in defining phrasal boundaries, pitch and duration have been suggested as the principal cues for perception (Streeter, 1978). However, the results of this study suggest that amplitude information does operate *within* phrases to affect the relative intelligibility of words across the phrase (i.e., at a more fine-grained level of processing compared to phrase boundary perception).

Overall, the prosodic analysis of the current study suggests that the time-varying F_0 within vowels facilitates word recognition at points of prominence, and temporal amplitude information operates across the intonational phrase. Combined with other results from this set of experiments, F_0 also appears to be used as a grouping cue along with other indexical information that is available and both appear important for integrating time-varying information together. Thus, F_0 operates as both a prosodic and indexical cue and amplitude modulations appear to mainly be involved in facilitating intelligibility. However, these conclusions are based on temporal changes across vocalic intervals of speech. Other interactions could be observed when considering other sonorant or obstruent portions of the signal.

6.3 Contribution of vowels to sentence intelligibility

These results also help to confirm the performance levels obtained previously for vowel-replaced sentences that preserve primarily consonant cues. Previous investigations with noise-replacement of vowels have resulted in performance levels around 30% (e.g., Fogerty & Kewley-Port, 2009). This performance level was closely approximated in this study with maximal performance only reaching 38% for E-modulated talker-specific replacement vowels. This is far exceeded by conditions that preserve the natural vowels and segmentally replace the consonants (e.g., 65% in Fogerty & Kewley-Port, 2009). Thus, the use of consonant cues does not improve with use of vowel replacement compared to noise replacement and is consistent with previous results using a harmonic complex during segmental replacement (Cole *et al.*, 1996). Furthermore, results are consistent with the interpretation that vowels are important carriers of indexical cues about the talker (Owren & Cardillo, 2006), which appears to significantly impact how listeners use suprasegmental cues, particularly amplitude modulation, for sentence intelligibility.

The results of this study add to the theoretical understanding of how vowels contribute to sentence intelligibility. The physiological production of vowels ensures that they provide acoustic information about the talker's speech source and filter. Indexical information regarding the size and shape of the vocal tract, characterizations of vocal fold vibration, gender, and emotion are all acoustically encoded by vowels. This is in addition to phonetic information determined by the vocal tract configuration and suprasegmental information provided by the periodicity, timing, and energy of vocal fold vibration. The current set of experiments attempted to tease apart the multiple levels of information (linguistic and non-linguistic) coded by vowel acoustics to determine how they uniquely contribute to sentence intelligibility. The results of this study demonstrate the primary importance of the vowel temporal envelope for sentence intelligibility. This is in line with a large literature on temporal envelope contributions (e.g., Drullman *et al.*, 1995; Shannon *et al.*, 1995; Smith *et al.*, 2002; Rosen, 1992) and with previous investigations demonstrating sentence-level contributions of the temporal envelope specific to vowels (Fogerty & Humes, 2012; Fogerty, 2013) and that occur at suprasegmental levels, independent of phonetic identity (Fogerty, 2014). However, the current study adds an additional layer of complexity to this finding, suggesting that there is a significant interaction between intelligibility cues coded by the suprasegmental structure of the temporal envelope and the non-linguistic, indexical cues that are inherent in vowel acoustics and are inseparable from the linguistic processing of the vowel (Mullennix *et al.*, 1989; Goldinger, 1998; Borrie *et al.*, 2013).

7. Summary and conclusions

This set of experiments investigated temporal amplitude envelope and fundamental frequency modulation cues using a vowel replacement paradigm. Preserving the temporal envelope improved speech intelligibility of vowel-replaced sentences, but only when the replacement vowel preserved indexical information about the target talker (Exp. 2 and 3). When the vowels were replaced by a vowel from a different talker, natural variations of the F_0 contour were important for speech intelligibility (Exp. 1). Thus, talker-specific indexical cues related to the speech source and filter, present during vowels, appear to be important in

determining how listeners make use of other suprasegmental cues. This result further supports conclusions about the interplay between indexical and lexical information that are jointly coded in the same acoustic signal (e.g., Goldinger, 1998; Jusczyk, 1993; Mullennix *et al.*, 1989; Nygaard, 2005; Pisoni, 1993); non-linguistic indexical properties interact with linguistic levels of speech perception to facilitate word recognition. This study also adds to the discussion of the interface between phonetics and prosody (e.g., Beckman, 1996); results indicate that F_0 and amplitude modulation cues may have different prosodic functions that interact with word recognition. Importantly, both linguistic (i.e., prosodic) and non-linguistic (i.e. indexical) high-level cues facilitate recognition of the sentence, even in the presence of contradictory acoustic-phonetic information. Finally, time-varying aspects of these cues that result in acoustic change within vowel segments appear to make significant contributions to intelligibility, lending further support to views of the importance of temporal dynamics in speech perception (e.g., Greenberg *et al.*, 2003; Mücke *et al.*, 2014; Port, 2003). Overall, it appears that preservation of the vowel temporal envelope is essential for maximal sentence intelligibility, especially when indexical cues are present to facilitate perceptual grouping with the target sentence.

Acknowledgements

The author would like to thank Adrienne Low, Catherine Ashley, Carolyn Turbeville, Carson Holstien, Julia Porter, and Jenine Entwistle who assisted with data collection and scoring. Previous discussions with Charles Watson were also important for motivating the current study and particularly in designing the vowel morphing experiment. This work was supported, in part, by a grant from the National Institutes of Health, NIDCD R03-DC012506.

References

- Abercrombie, D. Elements of General Phonetics. Edinburgh: Edinburgh University; 1967.
- Assmann, PF. Proceedings of the 14th International Congress of Phonetic Sciences. San Francisco, CA: 1999. Fundamental frequency and the intelligibility of competing voices; p. 179-182.
- Bachorowski JA, Owren MJ. Acoustic correlates of talker sex and individual talker identity are present in a short vowel segment produced in running speech. *Journal of the Acoustical Society of America*. 1999; 106:1054–1063. [PubMed: 10462810]
- Bashford JA Jr, Warren RM, Brown CA. Use of speech-modulated noise adds strong “bottom-up” cues for phonemic restoration. *Perception & Psychophysics*. 1996; 58:342–350. [PubMed: 8935895]
- Beckman ME. The parsing of prosody. *Language and Cognitive Processes*. 1996; 11:17–67.
- Borrie SA, McAuliffe MJ, Liss JM, O’Beirne GA, Anderson TJ. The role of linguistic and indexical information in improved recognition of dysarthric speech. *Journal of the Acoustical Society of America*. 2013; 133:474–482. [PubMed: 23297919]
- Bradlow AR, Bent T. The clear speech effect for non-native listeners. *Journal of the Acoustical Society of America*. 2002; 112:272–284. [PubMed: 12141353]
- Bradlow AR, Torretta GM, Pisoni DB. Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. *Speech Communication*. 1996; 20:255–272. [PubMed: 21461127]
- Byrd, D.; Choi, S. At the juncture of prosody, phonology, and phonetics – The interaction of phrasal and syllable structure in shaping the timing of consonant gestures. In: Fougeron, C.; Kühnert, B.; D’Imperio, M.; Vallée, N., editors. *Laboratory Phonology*. Vol. 10. Berlin, New York: Mouton de Gruyter; 2010. p. 31-59.
- Calhoun S. How does informativeness affect prosodic prominence? *Language and Cognitive Processes*. 2010; 25:1099–1140.

- Chen F, Loizou PC. Contribution of cochlea-scaled entropy versus consonant-vowel boundaries to prediction of speech intelligibility in noise. *Journal of the Acoustical Society of America*. 2012; 131:4104–4113. [PubMed: 22559382]
- Cho, T. The Continuum Companion to Phonology. In: Kula, NC.; Botma, B.; Nasukawa, K., editors. *Laboratory phonology*. London/New York: Continuum; 2011. p. 343-368.
- Cho T, McQueen JM, Cox EA. Prosodically driven phonetic detail in speech processing: The case of domain-initial strengthening in English. *Journal of Phonetics*. 2007; 35:210–243.
- Cole R, Yan Y, Mak B, Fanty M, Bailey T. The contribution of consonants versus vowels to word recognition in fluent speech. *Proceedings of the ICASSP'96*. 1996:853–856.
- Cooper FS, Delattre PC, Liberman AM, Borst JM, Gerstman LJ. Some experiments on the perception of synthetic speech sounds. *The Journal of the Acoustical Society of America*. 1952; 24:597–606.
- Cutler A. The perception of rhythm in language. *Cognition*. 1994; 50:79–81. [PubMed: 8039376]
- Darwin CJ. Perceptual grouping of speech components differing in fundamental frequency and onset-time. *The Quarterly Journal of Experimental Psychology*. 1981; 33:185–207.
- Darwin CJ, Brungart DS, Simpson BD. Effects of fundamental frequency and vocal-tract length changes on attention to one of two simultaneous talkers. *Journal of the Acoustical Society of America*. 2003; 114:2913–2922. [PubMed: 14650025]
- Donaldson GS, Rogers CL, Cardenas ES, Russell BA, Hanna NH. Vowel identification by cochlear implant users: Contributions of static and dynamic spectral cues. *Journal of the Acoustical Society of America*. 2013; 134:3021–3028. [PubMed: 24116437]
- Dorman MF, Cutting JE, Raphael LJ. Perception of temporal order in vowel sequences with and without formant transitions. *Journal of Experimental Psychology: Human Perception and Performance*. 1975; 1:121.
- Drullman R. Temporal envelope and fine structure cues for speech intelligibility. *Journal of the Acoustical Society of America*. 1995; 97:585–592. [PubMed: 7860835]
- Ferguson SH. Talker differences in clear and conversational speech: Vowel intelligibility for normal-hearing listeners. *Journal of the Acoustical Society of America*. 2004; 116:2365–2373. [PubMed: 15532667]
- Fogerty D. Acoustic predictors of intelligibility for segmentally interrupted speech: Temporal envelope, voicing, and duration. *Journal of Speech, Language, and Hearing Research*. 2013; 56:1402–1408.
- Fogerty D. Importance of envelope modulations during consonants and vowels in segmentally-interrupted speech. *Journal of the Acoustical Society of America*. 2014; 135:1568–1576. [PubMed: 24606291]
- Fogerty D, Humes LE. The role of vowel and consonant fundamental frequency, envelope, and temporal fine structure cues to the intelligibility of words and sentences. *Journal of the Acoustical Society of America*. 2012; 131:1490–1501. [PubMed: 22352519]
- Fogerty D, Kewley-Port D. Perceptual contributions of the consonant-vowel boundary to sentence intelligibility. *Journal of the Acoustical Society of America*. 2009; 126:847–857. [PubMed: 19640049]
- Fox RA, Wall LG, Gokcen J. Age-related differences in processing dynamic information to identify vowel quality. *Journal of Speech, Language, and Hearing Research*. 1992; 35:892–902.
- Fry DB. Experiments in the perception of stress. *Language and speech*. 1958; 1:126–152.
- Fry DB, Abramson AS, Eimas PD, Liberman AM. The identification and discrimination of synthetic vowels. *Language and Speech*. 1962; 5:171–189.
- Gardner RB, Gaskill SA, Darwin CJ. Perceptual grouping of formants with static and dynamic differences in fundamental frequency. *Journal of the Acoustical Society of America*. 1989; 85:1329–1337.
- Garofolo J, Lamel L, Fisher W, Fiscus J, Pallett D, Dahlgren N. DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus CD-ROM, National Institute of Standards and Technology. 1993 NTIS Order No. PB91-505065.
- Ghitza O, Greenberg S. On the possible role of brain rhythms in speech perception: intelligibility of time-compressed speech with periodic and aperiodic insertions of silence. *Phonetica*. 2009; 66:113–126. [PubMed: 19390234]

- Goldinger SD. Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 1996; 22:1166–1183.
- Goldinger SD. Echoes of echoes? An episodic theory of lexical access. *Psychological Review*. 1998; 105:251–279. [PubMed: 9577239]
- Greenberg S. Speaking in shorthand-A syllable-centric perspective for understanding pronunciation variation. *Speech Communication*. 1999; 29:159–176.
- Helmholtz, HLF. On the sensations of tone. 2nd Edition. Ellis, AJ., editor. New York: Longman; 1885.
- Hillenbrand, JM. Vowel inherent spectral change. Berlin Heidelberg: Springer; 2013. Static and dynamic approaches to vowel perception; p. 9-30.
- Hillenbrand J, Getty LA, Clark MJ, Wheeler K. Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America*. 1995; 97:3099–3111. [PubMed: 7759650]
- Iyer N, Brugart DS, Simpson BD. Effects of periodic masker interruption on the intelligibility of interrupted speech. *Journal of the Acoustical Society of America*. 2007; 122:1693–1701. [PubMed: 17927429]
- Jenkins JJ, Strange W, Edman TR. Identification of vowels in ‘vowelless’ syllables. *Perception and Psychophysics*. 1983; 34:441–450. [PubMed: 6657448]
- Johnson, K. Speaker normalization in speech perception. In: Pisoni, D.; Remez, RE., editors. *The handbook of speech perception*. Malden, MA: Blackwell Publishing; 2005.
- Kawahara H, Masuda-Kastuse I, Cheveigné A. Restructuring speech representations using a pitch adaptive time-frequency smoothing and an instantaneous- frequency-based F0 extraction: Possible role of a repetitive structure in sounds. *Speech Communication*. 1999; 27:187–207.
- Keating P, Shattuck-Hufnagel S. A prosodic view of word form encoding for speech production. *UCLA Working Papers in Phonetics*. 2002; 101:112–156.
- Kent, RD.; Read, C. *The Acoustic Analysis of Speech*. San Diego: Singular; 1992.
- Kewley-Port D, Burkle TZ, Lee JH. Contribution of consonant versus vowel information to sentence intelligibility for young normal-hearing and elderly hearing-impaired listeners. *Journal of the Acoustical Society of America*. 2007; 122:2365–2375. [PubMed: 17902871]
- Kim S, Cho T. Prosodic boundary information modulates phonetic categorization. *Journal of the Acoustical Society of America*. 2013; 134:EL19–EL25. [PubMed: 23862901]
- Ladefoged, P.; Disner, SF. *Vowels and Consonants*. 3rd Ed.. Malden, MA: Wiley-Blackwell; 2012.
- Laures J, Weismer G. The effect of flattened F0 on intelligibility at the sentence- level. *Journal of Speech, Language, and Hearing Research*. 1999; 42:1148–1156.
- Lehiste, I. *Suprasegmentals*. Cambridge, MA: MIT Press; 1970.
- Leong V, Stone MA, Turner RE, Goswami U. A role for amplitude modulation phase relationships in speech rhythm perception. *The Journal of the Acoustical Society of America*. 2014; 136:366–381. [PubMed: 24993221]
- Lieberman M, Pierrehumbert J. Intonational invariance under changes in pitch range and length. *Language sound structure*. 1984; 157:233.
- Lieberman M, Prince A. On stress and linguistic rhythm. *Linguistic inquiry*. 1977:249–336.
- Magnuson JS, Nusbaum HC. Acoustic differences, listener expectations, and the perceptual accommodation of talker variability. *Journal of Experimental Psychology: Human Perception and Performance*. 2007; 33:391. [PubMed: 17469975]
- Mullennix JW, Pisoni DB, Martin CS. Somme effects of talker variability on spoken word recognition. *Journal of the Acoustical Society of America*. 1989; 85:365–378. [PubMed: 2921419]
- Nearey, TM. Vowel inherent spectral change. Berlin Heidelberg: Springer; 2013. Vowel inherent spectral change in the vowels of North American English; p. 49-85.
- Nearey TM, Assmann PF. Modeling the role of inherent spectral change in vowel identification. *Journal of the Acoustical Society of America*. 1986; 80:1297–1308.
- Nygaard, L. Linguistic and non-linguistic properties of speech. In: Pisoni, D.; Remez, RE., editors. *The handbook of speech perception*. Malden, MA: Blackwell Publishing; 2005.
- Nygaard LC, Pisoni DB. Talker-specific learning in speech perception. *Perception & Psychophysics*. 1998; 60:355–376. [PubMed: 9599989]

- Owren MJ, Cardillo GC. The relative roles of vowels and consonants in discriminating talker versus word meaning. *Journal of the Acoustical Society of America*. 2006; 119:1727–1739. [PubMed: 16583915]
- Peterson GE, Barney HL. Control methods used in a study of the vowels. *Journal of the Acoustical Society of America*. 1952; 24:175–184.
- Pisoni DB. Long-term memory in speech perception: Some new findings on talker variability, speaking rate and perceptual learning. *Speech Communication*. 1993; 13:109–125. [PubMed: 21461185]
- Poeppel D. The analysis of speech in different temporal integration windows: cerebral lateralization as ‘asymmetric sampling in time’. *Speech communication*. 2003; 41:245–255.
- Port RF. Meter and speech. *Journal of Phonetics*. 2003; 31:599–611.
- Rosen S. Temporal information in speech: Acoustic, auditory, and linguistic aspects. *Philosophical Transactions of the Royal Society B*. 1992; 336:367–373.
- Salverda AP, Dahan D, McQueen JM. The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition*. 2003; 90:51–89. [PubMed: 14597270]
- Shannon RV, Zeng F-G, Kamath V, Wygonski J, Ekelid M. Speech recognition with primarily temporal cues. *Science*. 1995; 270:303–304. [PubMed: 7569981]
- Shinn-Cunningham BG, Wang D. Influences of auditory object formation on phonemic restoration. *Journal of the Acoustical Society of America*. 2008; 123:295–301. [PubMed: 18177159]
- Sluijter AM, Van Heuven VJ. Spectral balance as an acoustic correlate of linguistic stress. *Journal of the Acoustical society of America*. 1996; 100:2471–2485. [PubMed: 8865652]
- Smith ZM, Delgutte B, Oxenham AJ. Chimaeric sounds reveal dichotomies in auditory perception. *Nature*. 2002; 416:87–90. [PubMed: 11882898]
- Stevens KN. Toward a model for lexical access based on acoustic landmarks and distinctive features. *Journal of the Acoustical Society of America*. 2002; 111:1872–1891. [PubMed: 12002871]
- Strange W, Jenkins JJ, Johnson TL. Dynamic specification of coarticulated vowels. *Journal of the Acoustical Society of America*. 1983; 74:695–705. [PubMed: 6630725]
- Streeter LA. Acoustic determinants of phrase boundary perception. *Journal of the Acoustical Society of America*. 1978; 64(6):1582–1592. [PubMed: 739094]
- Studebaker G. A rationalized arcsine transform. *Journal of Speech, Language, and Hearing Research*. 1985; 28:455–462.
- Sumby WH, Pollack I. Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America*. 1954; 26:212–215.
- Terken, J.; Hermes, D. The perception of prosodic prominence. In: Horne, M., editor. *Prosody: Theory and experiment*. Studies presented to Gösta Bruce. Dordrecht: Kluwer; 2000. p. 89-127.
- Turner GS, Tjaden K, Weismer G. The influence of speaking rate on vowel space and speech intelligibility for individuals with amyotrophic lateral sclerosis. *Journal of Speech, Language, and Hearing Research*. 1995; 38:1001–1013.
- Van Tasell DJ, Soli SD, Kirby VM, Widin GP. Speech waveform envelope cues for consonant recognition. *Journal of the Acoustical Society of America*. 1987; 82:1152–1161. [PubMed: 3680774]
- Wang X, Humes LE. Factors influencing recognition of interrupted speech. *Journal of the Acoustical Society of America*. 2010; 128:2100–2111. [PubMed: 20968381]
- Xu L, Thompson CS, Pfingst BE. Relative contributions of spectral and temporal cues for phoneme recognition. *Journal of the Acoustical Society of America*. 2005; 117:3255–3267. [PubMed: 15957791]

Figure 1.

Talker discrimination based on original and replaced vowel pairs excised from the experimental stimuli used in the three experiments. Replacement vowels were spoken by different talkers, the same talkers, or spectrally morphed presentations of the original vowels.

Figure 2.

The amplitude contour (top), F_0 contour (middle), and spectrograms (bottom) are displayed for the example sentence "Ralph prepared red snapper with fresh lemon sauce for dinner" using /æ/ as the replacement vowel. As modified contours were the same for two of the experimental conditions, only the one from the Flat F_0 experimental condition is displayed. The grey bars on the top of the amplitude and pitch contour plots display the durations of each vowel in the sentence.

Figure 3.

Accuracy for the two listener groups for vowel-replaced sentences that preserve the identity of the consonants. All vowels were replaced by a single vowel (either /ə/ or /æ/) spoken by a standard talker and were modified to preserve the original vowel duration. Three different signal processing strategies, displayed left to right, were used to present (1) vowels with a flat F_0 at the mean of the original vowels, (2) F_0 -modulated vowels that preserved the natural F_0 contour of the original vowels, (3) E-modulated vowels that preserved the temporal envelope of the original vowels. Error-bars display the standard error of the mean.

a) Prominence Analysis **b) Intonational Phrase Analysis** **Figure 4.**

Word recognition scores across acoustic conditions by the prosodic structure of the sentence for a) prominent words and b) word position within the intonational phrase.

Figure 5.

Example spectrograms are displayed to illustrate the effect of vowel replacement using a vowel produced by the same talker spoken in a different sentence context. The original (top) and vowel-replaced (bottom) versions are of the sentence displayed in Figure 2 “Ralph prepared red snapper with fresh lemon sauce for dinner.” The vowel-replaced sentence was processed according to the E-modulated condition.

Figure 6.

Accuracy for the three listener groups in Experiments 2a and 2b that heard sentences with vowels replaced by a single vowel spoken by the same talker in a different context. For comparison, results are also displayed from Experiment 1 which used a replacement vowel from a different talker (white bars). Five different signal processing strategies, displayed left to right, were used to present (1) unmodified replacement vowels, (2) replacement vowels with a flat F_0 at the mean of the original vowels, (3) F_0 -modulated replacement vowels that preserved the natural F_0 contour of the original vowels, (4) E-modulated replacement vowels modulated by the temporal envelope of the original vowels, or (5) replacement vowels that preserved both the F_0 and E modulation of the original vowels. Note that Experiment 2b scaled vowels to the original vowel level, while the other experiments scaled vowels to the mean sentence level for conditions 1–3. Error-bars display the standard error of the mean.

Figure 7.

Example spectrograms are displayed to illustrate the effect of morphing the original vowel spectrum to contain steady-state formants from a template spectrum. The original (top) and vowel-morphed (bottom) versions are of the sentence displayed in Figures 2 and 5: “Ralph prepared red snapper with fresh lemon sauce for dinner.” The displayed vowel-morphed version was processed according to the E-modulated condition.

Figure 8.

Results are plotted for Experiment 3 (solid line) that heard sentences in which the vowels were acoustically morphed to match a single spectral template. Three experimental conditions were tested (Flat F_0 = white symbol; F_0 -mod = grey symbol; E-mod = black symbol). Results from Experiments 1 (which replaced indexical cues; dotted) and Experiment 2a (which provided indexical cues about talker identity; dashed) are plotted for comparison. Results are plotted according to a standardized measure of acoustic similarity to the original production, with larger values indicating greater preservation of the original acoustic signal. Error bars display the standard error of the mean.

Table 1

Origin of the replacement vowel used across the experiments.

Experiment	Talker	Source	Vowel
1	Hillenbrand speaker; gender matched	“hVd” context	/ə/ or /æ/
2	Same as original	Standard carrier sentence spoken by all talkers	Excised from the word “had” in the carrier phrase
3	Same as original	Same recording as original	Original vowel with formants morphed to a spectral template derived from the middle time point of the replacement vowel used in Exp. 2.

Table 2

Summary of vowel acoustic modifications across the three experiments.

Condition	Exp.	Mean F ₀	F ₀ contour	Mean Amplitude (rms)	Amplitude contour (envelope)	Duration
Replaced _{scale}	2b	Replaced	Replaced	-	Replaced	-
Flat- F ₀	1,2a,3	-	Flat	Sentence	Replaced	-
Flat- F _{0scale}	2b	-	Flat	-	Replaced	-
F ₀ -mod	1,2a,3	-	-	Sentence	Replaced	-
F ₀ -mod _{scale}	2b	-	-	-	Replaced	-
E-mod	1,2a,2b,3	-	Flat	-	-	-
F ₀ &E-mod	2b	-	-	-	-	-

The dash (–) indicates values that were preserved from the original vowel. Replaced = the acoustic property of the replacement vowel was preserved, and therefore constant across all replacement intervals within the sentence; Flat = the acoustic property was presented at a constant value throughout the duration of the replacement vowel; Sentence = the acoustic property was matched to the mean of the sentence.