Published in final edited form as:

J Speech Lang Hear Res. 2013 October ; 56(5): 1402–1408. doi:10.1044/1092-4388(2013/12-0203).

Acoustic predictors of intelligibility for segmentally interrupted speech: Temporal envelope, voicing, and duration

Daniel Fogerty

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

Daniel Fogerty: fogerty@sc.edu

Abstract

Purpose—Temporal interruption limits the perception of speech to isolated temporal glimpses. An analysis was conducted to determine the acoustic parameter that best predicts speech recognition from temporal fragments that preserve different types of speech information, namely consonants and vowels.

Method—Young normal-hearing listeners previously completed word and sentence recognition tasks that required them to repeat word and sentence material that was temporally interrupted. Interruptions were designed to replace various portions of consonants or vowels with low-level noise. Acoustic analysis of preserved consonant and vowel segments was conducted to investigate the role of the preserved temporal envelope, voicing, and speech duration in predicting performance.

Results—Results demonstrated that the temporal envelope, predominantly from vowels, is most important for sentence recognition and largely predicts results across consonant and vowel conditions. In contrast, for isolated words the proportion of speech preserved was the best predictor of performance regardless of whether glimpses were from consonants or vowels.

Conclusions—These findings suggest consideration of the vowel temporal envelope in speech transmission and amplification technologies for improving the intelligibility of temporally interrupted sentences.

Identifying important acoustic properties for speech understanding is important for the successful design of speech transmission technologies. The goal of these technologies is to preserve or enhance, such as in the case of assistive listening devices, essential speech acoustic cues to achieve maximal speech understanding. However, identification of these cues is complicated by the overlapping, time-varying, multidimensional nature of speech. Furthermore, it is likely that the acoustic and linguistic context, as well as the auditory environment influence the physical signal such that the cues that are most important vary according to the listening situation. This is further complicated by listener experience and abilities, which may interact with the perceptual use of different acoustic cues for determining speech recognition performance.

One theory for how listeners recognize speech in noise proposes that listeners “glimpse” speech at moments of relatively favorable signal-to-noise ratios (SNR; Cooke, 2003). Listeners base their speech recognition judgments primarily on the available glimpses. A

method used to investigate how listeners use temporal glimpses of speech has been to interrupt the speech signal by periodically replacing temporal portions with noise (see Miller & Licklider, 1950). In contrast to a concurrent fluctuating noise that could possibly allow for the availability of energetically masked speech cues during the noise interval; this replacement method ensures that speech cues are limited to the preserved speech intervals between noise replacements.

A number of factors influence performance for periodically interrupted speech, including interruption frequency (e.g., Nelson & Jin, 2004), duration of individual glimpses (Li & Loizou, 2007), and proportion of the total speech duration preserved (PTD; e.g., Miller & Licklider, 1950). However, these factors are often confounded with each other, as increasing the frequency of interruption can decrease the glimpse duration. To disentangle these findings, Wang and Humes (2010) systematically varied each of these parameters within the same range of values used in the present investigation. They demonstrated for CVC words that the PTD is the primary predictor of performance, with other factors playing a secondary role. This finding was later confirmed by Kidd and Humes (2012) for sentences from the Revised Speech Perception in Noise test (R-SPIN; Bilger, Nuetzel, Rabinowitz, & Rzeczkowski, 1984). Therefore, while isolated words and sentences may involve different sources of information, such as contextual cues in sentences, PTD still remains the primary periodic interruption parameter that best predicts performance.

However, it is likely that not all acoustic speech events contribute equally to intelligibility. For example, Fogerty and Kewley-Port (2009) limited temporal glimpses to either predominant vowel or predominant consonant speech events in sentences. They found that across different proportions of speech, glimpses locked to vowel events contributed more to intelligibility than those locked to consonant events. Thus, Fogerty and Kewley-Port documented an exception to the PTD “rule” identified by Wang and Humes (2010). Glimpsing vowel acoustic cues may be essential for maximizing speech intelligibility. However, for isolated word contexts, Fogerty and Humes (2010) and Owren and Cardillo (2006) found no such advantage for vowel segments, suggesting that the vowel acoustic property responsible for greater intelligibility could be related to slow-varying temporal cues distributed across longer utterances. In word contexts, the PTD “rule” remains true.

Recently, Fogerty and Humes (2012) investigated three different acoustic properties that could underlie the superiority of glimpsing acoustic cues conveyed during vowel segments in sentences. These properties were the fundamental frequency (F_0) contour, the temporal envelope, and the temporal fine structure. These acoustic properties have been described as important for the intelligibility of temporally interrupted speech. F_0 is important for auditory streaming, source segregation, and integration across interruptions (Bregman *et al.*, 1990; Darwin *et al.*, 2003). The temporal envelope is important for speech intelligibility (Shannon *et al.*, 1995), timing (Port, 2003), and word prediction (Waibel, 1987). Finally, the temporal fine structure is believed to be most useful in glimpsing preserved speech fragments between interfering or interrupting maskers (Lorenzi *et al.*, 2006). In their study, Fogerty and Humes (2012) presented listeners with sentences and words limiting cues from one of these acoustic properties. Consistent with a number of studies (e.g., Laures & Wiesmer, 1999; Watson & Schlauch, 2008), they found that flattening or removing F_0 cues did reduce performance.

However, limiting F_0 cues did not affect the differential contribution between vowel and consonant acoustic cues. That is, sentence-level vowels without F_0 cues still resulted in superior speech intelligibility. This occurred even though F_0 is believed to contribute significantly to interrupted speech intelligibility (e.g., Ba Kent & Chatterjee, 2010) and that vowels contain the primary voicing for speech. In contrast, temporal envelope cues present during vowels resulted in a performance advantage over cues during consonants in sentences. This effect did not occur in words, mirroring the context-dependent advantage of natural vowels. Temporal fine structure cues were also limited to the vowel, but this occurred in both word and sentence contexts, and therefore, is not investigated here as an acoustic explanation of differences between consonants and vowels. Thus, Fogerty and Humes (2012) interpreted these behavioral results as suggesting that slow-varying amplitude cues of the temporal envelope are essential for sentence intelligibility, and that vowels are best able to preserve the continuity of this temporal signal across the sentence.

The current investigation represents an acoustic analysis of different speech properties that may contribute to speech intelligibility. Namely, these are duration as measured by PTD, the preservation of the wide-band temporal envelope, and preservation of F_0 . Therefore, the current analysis was designed to compare: (1) the duration of speech preserved (PTD), (2) the envelope preserved during segmental interruption for natural, unprocessed speech, and (3) F_0 cues to voicing that clearly are more present during vowels but were not behaviorally implicated in explaining the different contributions of consonants and vowels in sentences (Fogerty & Humes, 2012). Importantly, this analysis seeks to define an acoustic parameter that can explain performance collapsed across both consonant and vowel interruption conditions without a priori labeling of the segmental condition. That is, which acoustic parameter best predicts performance across different types of segmental interruption?

Recently, Fogerty, Kewley-Port, and Humes (2012) collected intelligibility data for young and older listeners on sentence and word materials when listening to consonant or vowel glimpses. Importantly, they controlled and equally varied the total proportion of speech presented in consonant and vowel conditions across four proportions. Thus, the PTD as a predictor for consonant and vowel performance can be compared against the other acoustic measures investigated here. The current investigation represents a summary of previous consonant and vowel findings for the recognition of words and sentences by young normal-hearing listeners. Furthermore, it extends previous investigations by identifying how the preservation of different acoustic properties in the speech signal might explain findings across consonant and vowel glimpse conditions and refines the PTD “rule” to account for contribution differences between consonants and vowels. That is to say, this study investigates whether a single acoustic parameter can predict performance regardless of the type of segment preserved.

Acoustic predictions of performance

An acoustic analysis was conducted to directly test the predictions of PTD, preserved temporal envelope, and preserved F_0 . Results from Fogerty and Humes (2012) with acoustically-limited speech stimuli suggest that the preservation of the temporal envelope is the best predictor of sentence intelligibility across consonant and vowel glimpses. In

contrast, F_0 preservation should not be as effective of a predictor, even though it is biased toward favoring vowel contributions, as vowels are the primary voiced unit of speech.

Fogerty, Kewley-Port, and Humes (2012) investigated vowel and consonant contributions across the same four PTD values by systematically shifting the vowel-consonant boundary either into the vowel or the consonant. Thus, PTD values were held constant across vowel and consonant conditions so that a measure of the preserved temporal envelope could be calculated to predict differences between the two segmental conditions. Only data for young, normal-hearing listeners are included here. Fogerty et al. (2012) tested twenty-four listeners, separated into four different groups of six listeners (i.e., one group for each boundary condition tested). These listeners completed tasks for both words and sentences.

Methods

Speech materials analyzed here were the same as those tested previously (Fogerty et al., 2012). Forty-two sentences spoken by 21 male and 21 female talkers were selected from the TIMIT database (Garofolo *et al.*, 1990). Consonant-vowel-consonant word materials (N=148) were selected from recordings of a single male talker (Takayanagi *et al.*, 2002). Phoneticians previously labeled the segmental boundaries for these materials. Speech materials were analyzed at four average PTD values (within .02) evaluated with the predominant vowel and consonant segments (PTD = 0.38, 0.46, 0.54, 0.62). PTD was calculated by summing the duration across all preserved speech segments and dividing by the total duration of the speech sample. Average PTD values were used across the entire word or sentence list. Measurements were also made at the additional PTD values tested by Fogerty and Kewley-Port (2009) and Fogerty and Humes (2010) for additional comparison of the robustness of these measures as predictors across studies. These additional PTD values were not included in the statistical analysis as these studies did not equate PTD across vowel and consonant conditions.

Previous experiments consisted of a word or sentence repetition task. Listeners were seated in a sound attenuating booth and listened to segmentally interrupted speech materials presented monaurally to the right ear via an ER-3A insert earphone. Speech materials were calibrated at 70 dB SPL. Segmental interruption was conducted by deleting the consonant or vowel segment according to the PTD condition and replacing it with a low-level noise spectrally-matched to the long-term average spectrum of the speech corpus. Segment boundaries (labeled in the corpus by phoneticians) were adjusted to the nearest amplitude minima (i.e., zero-crossing). The low-level noise was presented 16 dB below the average speech level. These procedures resulted in segmental interruption with minimal transients. In addition, the low-level noise aided continuity and ecological validity of the interruption while avoiding phonemic restoration effects that require the noise to be at a level that could potentially mask the speech. The focus of the analysis was to compare listener performance, previously obtained on these tasks, across different metrics of speech preservation.

Envelope preservation analysis—Of primary interest here was whether a measure of the preserved temporal envelope, after segmentation, could account for the observed differences in vowel and consonant contributions in sentences at the same PTD. The amount

of envelope preserved was indexed by correlating the overall speech envelope of the full speech sample with the segmentally-replaced speech sample at the various PTD values tested here. The speech envelope was extracted by down-sampling the full speech stimulus to 1000 Hz, followed by full-wave rectification and low-pass filtering at 64-Hz using a 6th-order Butterworth filter. Correlations between the full sentence envelope and segmented envelope are termed here as the Envelope Correlation Index (ECI). A higher ECI, that is, a higher correlation between the original and segmented speech envelope, indicates greater preservation of the speech envelope and should be viewed as relative to ECI values from other conditions. This ECI analysis used a noise replacement level at a signal-to-noise ratio (SNR) of 0 dB. Using a 0 dB SNR noise, that is, replacing the missing segment with the average speech level of the sentence or of the word list, effectively removes the noise level as a factor in calculating the ECI by preserving the relative direction of amplitude differences from vowels and consonants. Similarly, the ECI provides an index of time-varying amplitude information, specific to the preserved vowels or consonants. The ECI is modeled after similar measures of temporal envelope agreement by Fortune, Woodruff, and Preves (1994) who calculated envelope differences and Gallun and Souza (2008) who examined the spectral distribution of correlated envelope information.

F0 preservation analysis—For comparison, a correlational measure of agreement between the segmented F_0 and full F_0 contour was calculated in a similar way as the ECI and will be termed here as the Fundamental Frequency Correlation Index (F_0CI). STRAIGHT (Kawahara *et al.*, 1999), a high-fidelity speech analysis and synthesis tool, was used to extract the full F_0 contour for the segmented and full speech sample. A correlation was calculated between these extracted contours. This correlational measure was in high agreement ($r > 0.99$) with the proportion of total voiced duration preserved calculated by the duration of voicing in the segmented version divided by the duration of total voicing in the speech sample. This high agreement was true for both word and sentence materials. This agreement occurred because the experimental conditions did not alter the kinematic F_0 trajectory between the full and segmented versions. Therefore, in this case, F_0CI can be viewed as a measure of the preserved voiced segments.

Results

Analysis here is conducted only over the data from Fogerty et al. (2012) that controlled for PTD between consonant and vowel glimpses. Table 1 provides the mean results for the acoustic measures analyzed here for all speech materials. Displayed in Figure 1 is performance as a function of PTD, ECI, and F_0CI for (a) sentences and (b) words. As the range of values was different among the different measures, results are plotted according to standardized z-scores to facilitate comparison. Shaded symbols display the performance data from Fogerty et al. (2012) for young normal-hearing listeners who were tested at equal PTDs for vowels and consonants. As a reference for including additional listeners and PTD conditions, the open symbols display sentence data from Fogerty and Kewley-Port (2009) and word data from Fogerty and Humes (2010).

Analyses were conducted separately for sentences and words to determine the extent to which PTD, ECI, and F_0CI account for the variance in performance across vowel and

consonant conditions. It is important to note that in order to capture any perceptual difference between consonants and vowels, the acoustic metric must differentially weight these segments based upon general acoustic factors. However, even if a metric differentially weights these segments based upon the different acoustic properties inherent to these segmental types, this does not mean that the metric will be successful at predicting variations in performance across conditions. The statistical analyses described here specifically investigate whether PTD, ECI, and F_0 CI are able to capture performance across all segmental conditions without a priori knowledge of the type of segment being replaced. Therefore, this is an investigation of intrinsic acoustic properties across both consonants and vowels that are predictive of performance across all segmentally interrupted materials.

Sentences

1. Independence of Acoustic Measures: Correlations between the acoustic measures were first investigated to determine any shared variance in predicting performance. ECI and PTD were not significantly correlated ($r = 0.44$, $p > 0.05$). However, while indexing separate phenomena, these measures are not mutually exclusive as increasing PTD would also result in greater preservation of the envelope. In contrast, ECI and F_0 CI were highly correlated ($r = 0.97$, $p < 0.05$). Partial correlations of ECI with performance, controlling for F_0 CI were not significant. The same was true for F_0 CI with performance, controlling for ECI. This indicated that ECI and F_0 CI were not independent measures. However, Fogerty and Humes (2012) behaviorally confirmed no effect of the presence of voicing or of F_0 variation on relative segmental contributions in sentences, but a significant effect of the temporal envelope. Therefore, the F_0 CI was dropped from additional analysis as it provided a redundant measure with ECI for these sentence materials.

2. Linear Regression Analysis: Stepwise linear regression with ECI and PTD as predictors of listener performance across the segmental conditions demonstrated that for sentences, ECI is the primary predictor [$F(1,7) = 71.1$, $r^2 = 0.92$, $p < 0.001$]. Indeed, the ECI captures 92% of the variance across vowels and consonants, compared to 30% by PTD and 85% for F_0 CI. As observed in Figure 1a for sentences, the ECI largely captures performance across all conditions, with vowel conditions clearly preserving more of the overall temporal envelope (as indicated by the greater ECI values for most vowel conditions).

3. Analysis of Residual Error: Central to this analysis is whether these acoustic measures account for performance without prior knowledge of whether consonants or vowels were preserved in the replacement condition. If a single trend across the collapsed consonant and vowel conditions is as accurate at predicting performance as two independent trends for consonants and vowels, then it could be considered that the acoustic measure does not require a priori labeling of the segmental category. It is of particular note that such an analysis is biased toward predicting better performance for two separate trends due to the additional fitting parameters. Thus, a single trend indicates a more parsimonious explanation of participant performance.

To directly investigate this possibility, separate correlations were calculated for vowels and consonants across PTD and ECI and the residual errors were compared. By examining

Figure 1a, it is not surprising that comparison of the residual errors between using a single trend across PTD, or two separate trends for consonants and vowels, demonstrated a significantly better fit for two versus one trend [$t(7) = 5.25$, $p < 0.001$]. In contrast, a similar analysis for ECI demonstrated no significant difference [$t(7) = 0.02$, $p > 0.05$]. Thus, compared to PTD, ECI was better able to capture performance across consonants and vowels, indicating a more parsimonious account of participant performance.

4. Summary: The ECI analysis did much better at accounting for performance than PTD, suggesting that vowels preserve more of the amplitude envelope of sentences. Furthermore, the amount of the envelope preserved, regardless of the segment presented, largely predicts overall performance. Thus, ECI is able to predict performance without a priori labeling of vowel and consonant conditions. Given that ECI was highly correlated with the preservation of voiced speech segments (i.e., F_0CI), it is conceivable that time-varying amplitude envelope cues during voiced speech provide the most information to sentence intelligibility. However, initial acoustic analysis of a combined measure of ECI values during preserved voiced segments did not facilitate prediction above that provided by envelope cues alone. This may be because the envelope already provides cues for voicing (Rosen, 1992). Thus, the ECI appears to capture one important component necessary for the intelligibility of segmentally interrupted speech.

Isolated Words

1. Independence of Acoustic Measures: Correlations between PTD, ECI, and F_0CI were first investigated to determine any shared variance for explaining performance on the isolated words. In contrast to the sentence materials, no significant correlations were obtained between these measures. Therefore, all three measures were investigated as possible predictors of performance.

2. Linear Regression Analysis: Stepwise linear regression across segmental conditions with PTD, ECI, and F_0CI as predictor variables, demonstrated that the PTD is the primary predictor of performance [$F(2,7) = 88.6$, $r^2 = 0.94$, $p < 0.0001$]. This is further supported by the fact that only PTD was significantly correlated with performance across collapsed vowel and consonant conditions, accounting for 94% of the variance. Furthermore, recall that no differences for consonant and vowel words were observed by Fogerty and Humes (2012) when words were processed to preserve predominant envelope cues, or when the words were unprocessed (Fogerty and Humes, 2010; Fogerty et al., 2012). Contrasting with these behavioral results, ECI values for the unprocessed consonant ($M = 0.80$) and vowel ($M = 0.57$) words were significantly different [$t(148) = 10.5$, $p < 0.001$]. This further suggests that durational measures (e.g., PTD) and not acoustic measures of voicing or the amplitude envelope are predictive of performance for isolated words. While ECI values account for at least part of the performance differences observed between consonants and vowels in sentences, the envelope does not appear to account for the similarity in consonant and vowel performance for isolated CVCs.

3. Analysis of Residual Error: An analysis was again conducted, this time for word data, on the accuracy of predicting performance across vowel and consonant conditions versus

predicting these measures with prior knowledge of the condition label. The analysis of residual errors for one versus two trends was investigated, this time for PTD as it was the primary predictor of performance. No significant difference was obtained [$t(2) = 0.7$, $p > 0.05$], indicating that PTD was able to accurately predict performance when collapsed across segmental conditions.

4. Summary: The acoustic analysis demonstrates that the proportion of vowel envelope cues and proportion of voicing cues do not differentially assist in isolated word recognition processes. It appears that temporal envelope cues provide higher-level structural cues that assist speech perception at a more global level that is not informative in isolated word contexts. This observance of reduced envelope contributions may be due to the shorter duration of words ($M = 437$ ms) versus sentences ($M = 2480$ ms), as some slow-rate modulations may require longer speech durations to unfold. This may be supported by the finding that vowels, characterized by slow-varying amplitude changes, best preserve the sentence envelope. In addition, these slow modulations may also be linked with specific linguistic information present only in the sentence context that facilitates word identification (Waibel, 1987). Furthermore, while this study directly controlled the proportion of transitions between phonemes, it may be that greater coarticulation that occurs in natural sentences also influenced the information carrying capacity of the envelope.

This analysis suggests that PTD is still an important predictor of performance for consonants and vowels, particularly in isolated word contexts. These results support the conclusions of Fogerty and Humes (2012) in that the envelope does account for segmental differences in sentences, but is not as predictive in isolated words. However, other factors may also be involved in determining performance, that are not captured by the ECI.

General Discussion

The analysis conducted here investigated the contribution of duration, temporal envelope, and fundamental frequency to the contribution of speech glimpses from predominant consonant or vowel segments. Measures of the preserved duration (PTD), temporal envelope (ECI), and fundamental frequency (F_0 CI) were calculated to predict performance across segmental conditions. The high correlation between ECI and F_0 CI is interesting given that the presence of voicing or of natural F_0 variation did not explain the benefit of vowel acoustic cues in sentences (Fogerty and Humes, 2012). However, accounting for the combined contribution of ECI and F_0 CI indicated no improvements in prediction over ECI alone, possibly because the envelope already provides voicing cues (Rosen, 1992). Therefore, while the presence of these two cues is highly correlated, it appears that the temporal envelope is most predictive of performance. It may be that vowels are not the most informative units per se, but rather, it is the temporal amplitude envelope that contributes most to speech intelligibility. As it happens, vowels appear to best preserve sentence-level temporal envelope modulations during segmentally interrupted speech, possibly due to their relatively high intensity compared to many consonants. Thus, vowels capture the inherent rhythm of speech (Port, 2003) that appears crucial for sentence intelligibility.

As predicted by Fogerty and Humes (2012) who examined envelope contributions behaviorally through acoustic manipulations of the speech stimulus, the quantitative measurement of the preserved envelope (ECI) predicted differences in vowels and consonants at the same PTDs tested, reflecting better performance for the vowel conditions. ECI predicted performance across vowel and consonant conditions in sentences. In contrast, PTD, not the ECI or F_0 CI, predicted performance better in word contexts. Interestingly, while envelope and F_0 measures were highly correlated in sentences, these measures were not significantly correlated for isolated words. Furthermore, consonants better preserved the isolated word envelope and vowels better preserved the voicing F_0 CI cue. Thus, while ECI and F_0 CI resulted in similar conclusions in sentences, their predictions diverged for words, and neither accurately captured performance across segmental conditions. On the other hand, PTD was clearly associated with performance. For isolated words, performance was similar between vowel and consonant contributions at the same PTD, but not the same ECI or F_0 CI.

This set of acoustic analyses demonstrate that the speech envelope carried by consonants and vowels is sensitive to word or sentence context, although it is only predictive of performance in sentences. Therefore, in line with previous behavioral examinations, contributions of the preserved envelope are limited to the sentential context and favor the contribution of vowels. This may be why similar analyses of the envelope weighted toward the consonants only capture a small amount of the perceptual variance (Hoover, Souza, and Gallun, 2012). The results of the present analysis combined with the previous behavioral results strongly suggest that the temporal envelope during voiced speech segments is the acoustic property responsible for the large perceptual advantage observed for sentence-level vowel cues. Thus, we may say that it is not vowels, but supra-linguistic cues related to the amplitude variations and preserved most during vowel segments that provide a supra-additive contribution to sentence intelligibility, above what may be provided by the identity of individual phonemes.

The results from this analysis suggest that further investigations of temporally interrupted speech need to consider distortion of the temporal envelope, particularly during vowels, in predicting performance. The results suggest that in order to obtain maximal speech intelligibility for natural speech in temporally interrupted contexts, speech transmission and amplification technologies need to preserve temporal envelope cues. Furthermore, it is the vowel segments that best preserve these slow-varying sentence-level temporal cues. This analysis highlights the importance of preserving temporal envelope modulation during vowels.

Acknowledgments

The author would like to thank Larry Humes for his comments regarding this work. This work was supported, in part, by the National Institute on Deafness and Other Communication Disorders grant R03-DC012506.

References

- Ba kent D, Chatterjee M. Recognition of temporally interrupted and spectrally degraded sentences with additional unprocessed low-frequency speech. *Hearing Research*. 2010; 270:127–133. [PubMed: 20817081]
- Bilger RC, Nuetzel JM, Rabinowitz WM, Rzeczkowski C. Standardization of a test of speech perception in noise. *Journal of Speech and Hearing Research*. 1984; 27:32–48. [PubMed: 6717005]
- Bregman AS, Liao C, Levitan R. Auditory grouping based on fundamental frequency and formant peak frequency. *Canadian Journal of Psychology*. 1990; 44:400–413. [PubMed: 2224643]
- Cooke M. Glimpsing speech. *Journal of Phonetics*. 2003; 31:579–584.
- 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]
- Fogerty D, Humes LE. Perceptual contributions to monosyllabic word intelligibility: Segmental, lexical, and noise replacement factors. *Journal of the Acoustical Society of America*. 2010; 128:3114–3125. [PubMed: 21110607]
- 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]
- Fogerty D, Kewley-Port D, Humes LE. The relative importance of consonant and vowel segments to the recognition of words and sentences: Effects of age and hearing loss. *Journal of the Acoustical Society of America*. 2012 in press.
- Fortune TW, Woodruff BD, Preves DA. A new technique for quantifying amplitude envelope contrasts. *Ear and Hearing*. 1994; 15:93–99. [PubMed: 8194684]
- Gallun FJ, Souza P. Exploring the role of the modulation spectrum in speech recognition. *Ear and Hearing*. 2008; 29:800–813. [PubMed: 18596640]
- Garofolo, J.; Lamel, L.; Fisher, W.; Fiscus, J.; Pallett, D.; Dahlgren, N. NTIS Order No. PB91-505065. 1990. DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus CD-ROM, National Institute of Standards and Technology.
- Hoover EC, Souza PE, Gallun FJ. The consonant-weighted Envelope Difference Index (cEDI): A proposed technique for quantifying envelope distortion. *Journal of Speech, Language, and Hearing Research*. 2012 Epub ahead of print, March 12, 2012. 10.1044/1092-4388(2012/11-0255)
- Kawahara H, Masuda-Kastuse I, Cheveigne 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.
- Kidd GR, Humes LE. Effects of age and hearing loss on the recognition of interrupted words in isolation and in sentences. *Journal of the Acoustical Society of America*. 2012; 131:1434–1448. [PubMed: 22352515]
- 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.
- Li N, Loizou PC. Factors influencing glimpsing of speech in noise. *Journal of the Acoustical Society of America*. 2007; 122:1165–1172. [PubMed: 17672662]
- Lorenzi C, Gilbert G, Carn H, Garnier S, Moore BCJ. Speech perception problems of the hearing impaired reflect inability to use temporal fine structure. *Proceedings of the National Academy of Sciences*. 2006; 103:18866–18869.
- Miller G, Licklider J. The intelligibility of interrupted speech. *Journal of the Acoustical Society of America*. 1950; 22:167–173.
- Nelson PB, Jin S-H. Factors affecting speech understanding in gated interference: Cochlear implant users and normal-hearing listeners. *Journal of the Acoustical Society of America*. 2004; 115:2286–2294. [PubMed: 15139640]

- Owren MJ, Cardillo GC. The relative roles of vowels and consonants in discriminating talker identity versus word meaning. *Journal of the Acoustical Society of America*. 2006; 119:1727–1739. [PubMed: 16583915]
- 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.
- Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M. Speech recognition with primarily temporal cues. *Science*. 1995; 270:303–304. [PubMed: 7569981]
- Takayanagi S, Dirks D, Moshfegh A. Lexical and talker effects on word recognition among native and non-native listeners with normal and impaired hearing. *Journal of the American Academy of Audiology*. 2002; 16:494–504.
- Waibel A. Prosodic knowledge sources for word hypothesization in a continuous speech recognition system. *Acoustics Speech, and Signal Processing, IEEE International Conference on ICASSP*. 1987; 87:856–859.
- Wang X, Humes LE. Factors influencing recognition of interrupted speech. *Journal of the Acoustical Society of America*. 2010; 128:2100–2111. [PubMed: 20968381]
- Watson PJ, Schlauch RS. The effect of fundamental frequency on the intelligibility of speech with flattened intonation contours. *American Journal of Speech Language Pathology*. 2008; 17:348–355. [PubMed: 18840699]

a) Sentences

b) Words

Figure 1.

Results of unprocessed, natural speech during vowel and consonant replacement containing varied amounts of transitional cues into the vowel or consonant are displayed for (a) sentences and (b) words. For both contexts, results are plotted according to the proportion of the total duration (PTD), the preserved temporal envelope (ECI), and the preserved fundamental frequency (F₀CI). Independent variable scores were converted to standardized z-scores to facilitate comparison. Filled symbols are results from Fogerty et al. (2012). Unfilled symbols are from (a) Fogerty and Kewley-Port (2009) for sentences and (b) Fogerty and Humes (2010) for words.

Table 1

Mean values across all speech materials for the proportion of the total duration (PTD), the preserved temporal envelope (ECI), and the preserved fundamental frequency (F_0CI).

	Sentences			Words		
	PTD	ECI	F_0CI	PTD	ECI	F_0CI
Vowels Preserved	0.38	0.51	0.55	0.39	0.51	0.74
	0.45	0.53	0.63	0.46	0.57	0.79
	0.53	0.55	0.69	0.54	0.61	0.83
	0.62	0.58	0.75	0.62	0.67	0.86
Consonants Preserved	0.39	0.38	0.25	0.38	0.61	0.07
	0.47	0.42	0.30	0.46	0.71	0.11
	0.55	0.46	0.35	0.54	0.80	0.16
	0.62	0.48	0.40	0.61	0.84	0.22