Int. Adv. Otol. 2011; 7:(3) 391-397

ORIGINAL STUDY

Envelope Expansion of Sine Wave Vocoded Speech

Pitchai Muthu Arivudai Nambi, Jayashree Sunil Bhat, Haralakatta Shivananjappa Somashekara

Department of audiology and speech language pathology, Kasturba Medical College (Manipal University) (PMAN, JSB, HSS)

Objective: Present study investigated the effect of envelope expansion on sine wave vocoded speech. Speech recognition score for sine wave vocoded stimuli was assessed in quiet as well as in presence of background noise.

Materials and Methods: The test stimuli consisted of four lists, each containing ten HINT (Hearing in noise test) sentences spoken by a female speaker, subsequently each of these stimulus sentences were divided into 8 frequency bands and an envelope was extracted from each frequency bands at 400Hz cut off. Envelope was expanded using overlap and add algorithm (OLA) at the modulation frequencies from 1 to 30 Hz. Ten adult individuals participated in the perceptual task.

Results: Speech recognition scores in the presence of background noise was significantly poor when compared to speech recognition in quiet Paired't' test revealed no significant main effect of envelope expansion on speech recognition in quiet as well as in noise.

Conclusion: Envelope cues alone are sufficient for good speech recognition in quiet but not in the presence of noise. Envelope expansion scheme does not affect the sentence perception containing only envelope cues in quiet as well as in noise.

Submitted : 28 January 2011 *Accepted : 26 April 2011*

Introduction

Cochlear implants are an accepted and effective treatment for restoring hearing sensation to people with severe-to profound hearing loss. Contemporary cochlear implants consist of a microphone, a sound processor, a transmitter, a receiver, and an electrode array that is positioned inside the cochlea. The sound processor is responsible for decomposing the input audio signal into different frequency bands and delivering information about each frequency band to the appropriate electrode in a base-to-apex tonotopic pattern. The bandwidths of the frequency bands are approximately equal to the critical bands, where low-frequency bands have higher frequency resolution than high-frequency bands. The actual stimulation to each electrode consists of nonoverlapping biphasic charge-balanced pulses that are modulated by the low pass-filtered output of each analysis filter. However, the improvement in speech perception necessitates a highly sophisticated way of processing the speech and coding to the auditory nerve fibers.

Human speech is highly redundant with spectral and temporal cues. Importance of these cues for speech recognition has been the research interest in the recent decades. Smith et al. studied the relative importance of temporal envelope and fine structure cues for speech recognition. They reported that envelope cues are important for speech perception whereas fine structure cues are important for pitch perception and localization [1]. Most of the speech coding strategies of today's cochlear implants mainly rely on the temporal envelope cues for the speech recognition. The

Corresponding address:
Haralakatta Shivananjappa Somashekara
Department of audiology and speech language pathology
Kasturba Medical College (Manipal University) Mangalore – 575001
Phone: +919886373606 • Fax: 0824 2428379
E-mail: som.shekar@manipal.edu

temporal envelope cues from 3-4 bands are sufficient for the speech recognition in quiet [2]. However, recent studies have indicated that the envelope cues alone are not sufficient for the robust speech recognition in noise [3-6]. The noise causes degradations to these amplitude modulations which in turn affects the speech perception [7]. The presence of steady background noise fills up the valleys of speech thus reducing the modulation depth [8]. This envelope smearing of speech adversely affects the speech intelligibility [9,10] .

Clarkson and Bahgat reported that, envelope expansion can compensate for the deleterious effect of noise, as it increases the depth of amplitude modulation present in the speech [11]. Envelope expansion resulted in small improvements at 0 dB signal-to-noise (S/N) ratio, but no improvement was found at -5dB and -15dB S/N ratio [11]. Enhancing the modulation depth by 15 dB improved the speech perception even in the individuals with auditory neuropathy [12]. The temporal fine structure was not fully excluded in the above studies. The envelope enhancement scheme needs to be evaluated on a speech material containing only envelope cues to understand the role of envelope alone on speech recognition and also for its application in cochlear implants.

Fu and Shannon synthesized the envelope speech using noise wave vocoders and enhanced the modulation depth using power law transformations. Results of their study revealed a significant reduction in vowel and consonant recognition scores in quiet condition. Further, the study indicated that modulation depth enhancement distorts the speech signal in quiet [13]. Similarly Lorenzi and co-workers also reported that expanding the envelope decreases the consonant identification scores in quiet when the spectral information is restricted below 2500 Hz and scores are unaffected when full range of spectrum is applied [14]. In the presence of background steady state noise, consonant identification scores improved when presented at 0 dB SNR [14], as well as at -6 dB, 0 dB and +6 dB SNR [15]. The current study aims to extend the above studies to sentence recognition in background noise using sine wave vocoders.

Materials and Methods

Four lists each containing 10 English sentences taken from HINT [16] served as stimuli or the current study. Familiarity of the sentence lists were ascertained on individuals exposed to English language at least for ten years. In each list three key words were identified, thus having a total of 30 key words in each list. Only the key words were considered for the scoring. The speech stimuli were spoken by a female speaker with Indian English accent recorded digitally on a data acquisition system at 44.1 kHz sampling frequency and using a 16-bit A/D converter in a sound treated room.

Signal processing:

Speech signals with only envelope cues were synthesized using following steps. Signals were first processed through a pre-emphasis filter at1200 Hz cutoff, with a slope of 3dB/octave and then band passed into 8 frequency bands using sixth-order Butterworth filters. The envelope of the signal was extracted by full-wave rectification, and low-pass filtering (second-order butterworth) with a 400 Hz cutoff frequency. Sinusoids were generated with amplitudes equal to the root mean-square (RMS) energy of the envelopes and frequencies equal to the center frequencies of the band pass filters. Finally outputs of each filter were summed to produce the synthesized speech. To produce the speech in noise condition, white noise was added at 0 dB SNR by taking RMS value into consideration and the steps mentioned above were repeated.

Envelope enhancement was performed using PRAAT software version 4.1.21. The software incorporates overlap and add (OLA) procedure for speech processing. The speech signal was passed through stop band (300Hz and 8000Hz) (hamming) frequency domain filter after FFT. The spectral values at frequencies between 400 and 7900Hz were set to zero. The spectral values from 200 to 400Hz and from 7900 to 8100Hz were multiplied by a raised sine, so as to give a smooth transition without ringing in the time domain. Finally, a backward Fourier transformation was done to obtain the stop band signal. The processing was carried out on the spectrum in the 300Hz to 8000Hz range. The spectrum was divided

into different critical bands using Hamming window. Each frequency band is one Bark wide (based on bark scale), with 100 Hz overlap. Each critical band was converted to a band pass filtered sound by means of the backward Fourier transform. The band pass filtered sound was subjected to FFT and Envelope (intensity modulations) is obtained from the absolute value of FFT. The envelope within each band was then passed through a Frequency domain filter using hamming window to extract frequency range of interest. The filtered band was converted into power. Based on the power and the maximum enhancement, a factor required for envelope expansion was calculated in dB scale and the filtered sound was multiplied by this factor. The manipulated band pass signals were added and a backward Fourier transformation was done to obtain the manipulated signal. The manipulated signal was finally added to the stop band signal to get the envelope-enhanced signal. Modulation depth was enhanced by 15 dB at 1 to 30 Hz envelope bandwidth. This modulation frequency was selected based on the findings of previous research that, normal hearing individuals require envelope modulation of 20 Hz or greater for correct identification of syllables and the performance of individuals with cochlear hearing loss improve with enhanced modulations for larger bandwidths [17].

Subjects:

Ten adult males with the age ranging from 18 to 24 years participated in the current study. The hearing thresholds of the participants were ≤ 15 dB at octave frequencies from 250 Hz to 8000 Hz. All participants were exposed to English language for past ten years.

Procedure:

The experiment was performed on a PC equipped with a Creative Labs SoundBlaster 16 soundcard. The subjects listened to the sentences via Senheiser stereo headphones at a comfortable level set by the subjects, subsequently listened to the stimuli unilaterally and ear selection was randomized across the subjects. Written responses were obtained from the subjects on open set task. The responses were scored using the 'loose method' in which a response was recorded as correct if the root of it matches the root of the presented word [18].

Results

The current study investigated the speech recognition scores of sine wave vocoded speech with and without envelope expansion. Speech recognition ability was assessed in both quiet and noise condition using sentence recognition task. Henceforth in this article, speech recognition task in quiet without envelope expansion will be referred as QUIET; speech recognition task in quiet with envelope expansion as QUIETENH; speech in noise without envelope expansion as WN; and speech in noise with envelope expansion as WNENH. Speech recognition task was carried out using the standardized sentence material. The sentence recognition scores were calculated by counting the correctly identified key words within the sentences. Further, the scores were converted into percentages, for all the four conditions (quiet with and without envelope enhancement, white noise with and without envelope enhancement). Mean speech recognition scores obtained for QUIET, QUIETENH, WN and WNENH were 86.63%, 88.96%, 15.31% and 18.30% respectively. Mean and standard deviation for each condition is depicted in Figure 1.

The main effect of noise on speech recognition scores was evaluated using parametric paired 't' test. In view of Shapiro-Wilk tests for normality which indicated that the difference in scores between QUIET and WN was normally distributed ($W=0.94$, $p>0.01$), parametric test was chosen to investigate the main effect. Paired 't' test revealed that, the mean difference between speech recognition scores in quiet and in noise was statistically significant ($t9=34.8$, $p < 0.01$). Speech recognition scores in noise were significantly poorer than speech recognition scores in quiet. Shapiro-Wilk's tests for normality indicated that the difference in scores between QUIET & QUIETENH ($W=0.93$, $p>0.01$) and WN & WNENH ($W=0.82$, $p>0.01$) was normally distributed. Since, difference in scores were normally distributed, the main effect of envelope expansion was also investigated using parametric test. Paired sample't' test revealed that there was no significant difference between QUIET & QUIETENH ($t9= -1.3$, $p>0.01$) as well as WN & WNENH ($t9= -2.3$, $p>0.01$).

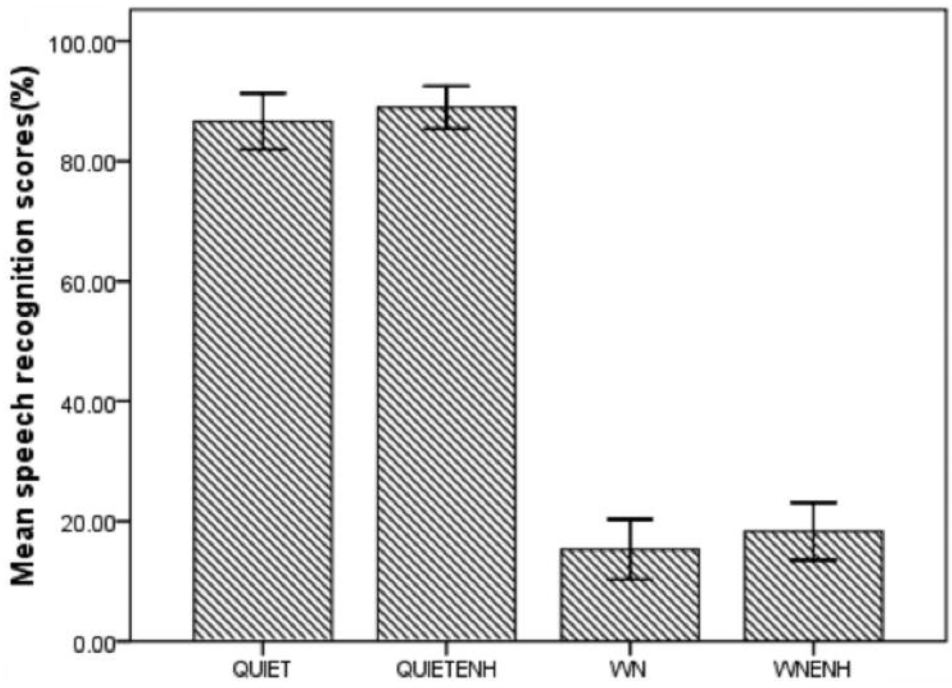

Figure 1. Mean speech recognition scores in quiet with and without envelope enhancement, and speech in noise with and without envelope enhancement (QUIET, QUIETENH, WN and WNENH). Error bar indicates ±1 standard deviation.

Discussion

In this study it was observed that the participant's performance yielded good scores only with envelope cues in QUIET, in line with the findings of Shannon et al [2]. However, in the presence of steady background noise the speech recognition scores were degraded when compared to QUIET, which can be observed from the figure 1. When the speech and noise were together, they produce a new mixed envelope which was different from target speech [6], which can be observed from Figure 2. This mixed envelope is used to modulate the carrier sine waves, because of which the listener may not be able to segregate the speech and noise into two perceptual streams [19].

The envelope enhancement did not have any significant effect on speech recognition scores in quiet, which was proven statistically too. Earlier study by Fu and Shannon showed that envelope expansion decreases the speech recognition scores in quiet [13]. This difference in the results between the current study and the earlier studies could be attributed to the difference in the type of speech material used. Earlier studies focused only on the consonant identification ability, with an observation that the consonant identification scores were degraded when the envelope was expanded. In the current study also, amplitude of consonantal portions was decreased when the envelope was expanded (Figure 3). Even though the acoustic analysis revealed decreased consonant amplitude, the sentence recognition scores were unaffected. This can be explained by glimpsing model [20], which states that auditory glimpsing involves taking brief "Snap shot" from an ongoing temporal sequence. It is the process by which distinct regions of the signal, separated in time, are linked together when intermediate regions are masked or deleted. Even though intermediate regions have been affected by the envelope expansion, unaffected portion of the sentences would have permitted the listener to "glimpse" the acoustic structure of the target speech. The results can also be attributed to the extrinsic redundancies of the speech, where sentences have more redundancies than syllables or word. Lorenzi and co-workers have reported that expanding the envelope does not affect the consonant identification scores when full range of spectrum is applied. The current study also supports their findings [14].

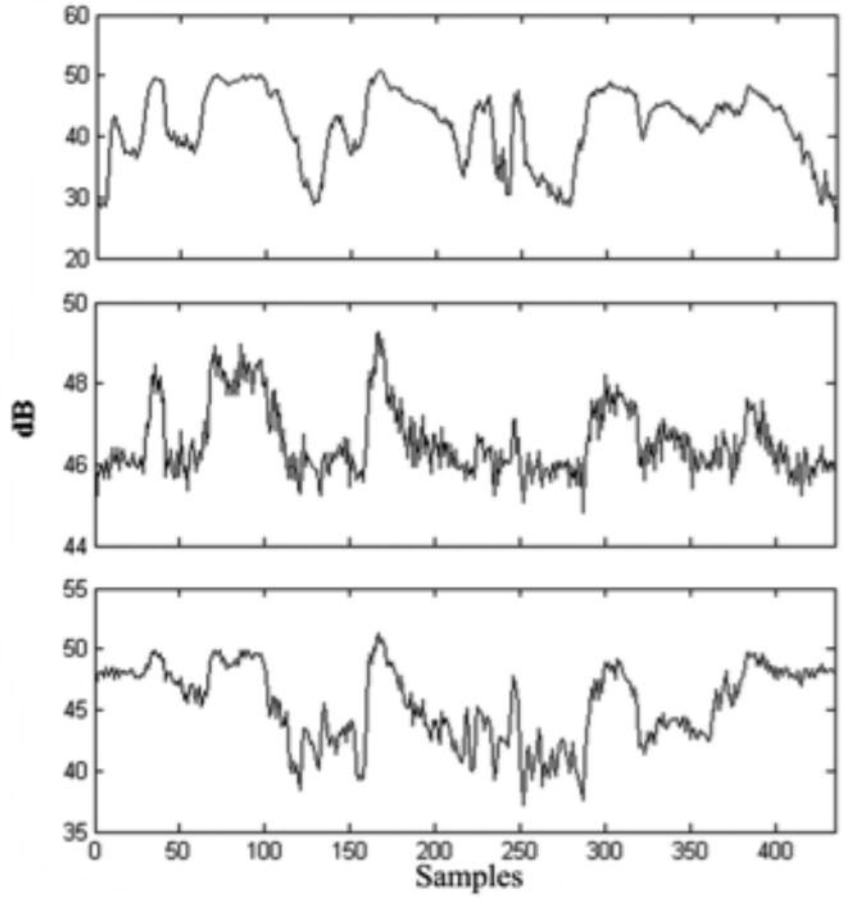

Figure 2. Energy plots for the sentence "the ball bounced very high". Top panel represents quiet speech without envelope expansion. Middle panel represents the noisy speech without envelope expansion. Bottom panel represents noisy speech with envelope expansion.

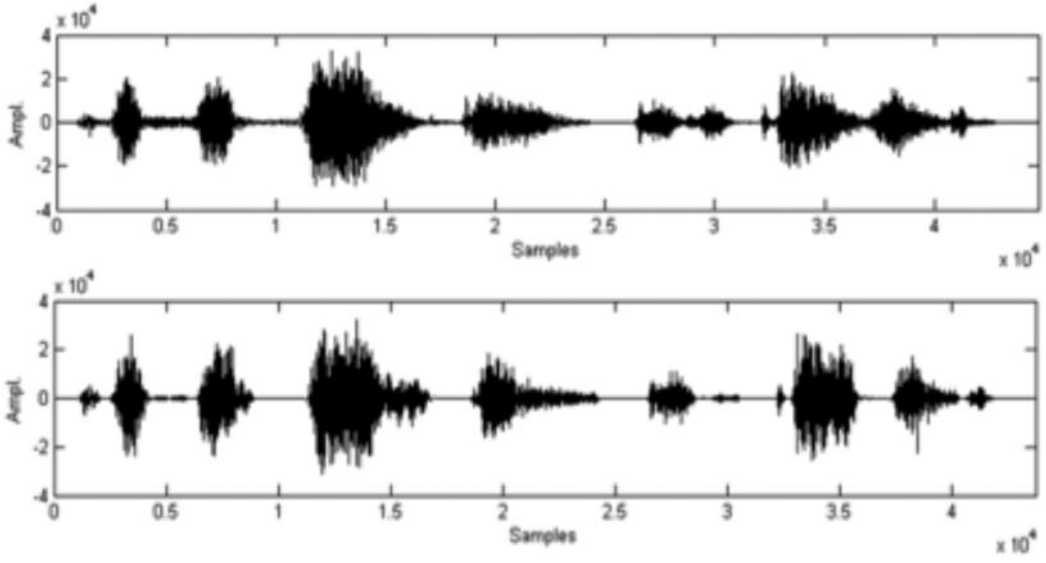

Figure 3. Wave form of the "The foot ball game is over". Top panel represents the quiet speech without envelope expansion. Bottom panel represents quiet speech with envelope expansion.

When the speech and noise are combined together they produce a new mixed envelope. This mixed envelope (speech + noise) not only differs in terms of modulation depth from the target speech, but also alters the modulation spectrum [6]. The effect of white noise and envelope expansion on speech is depicted using modulation spectrum (Figure 4).

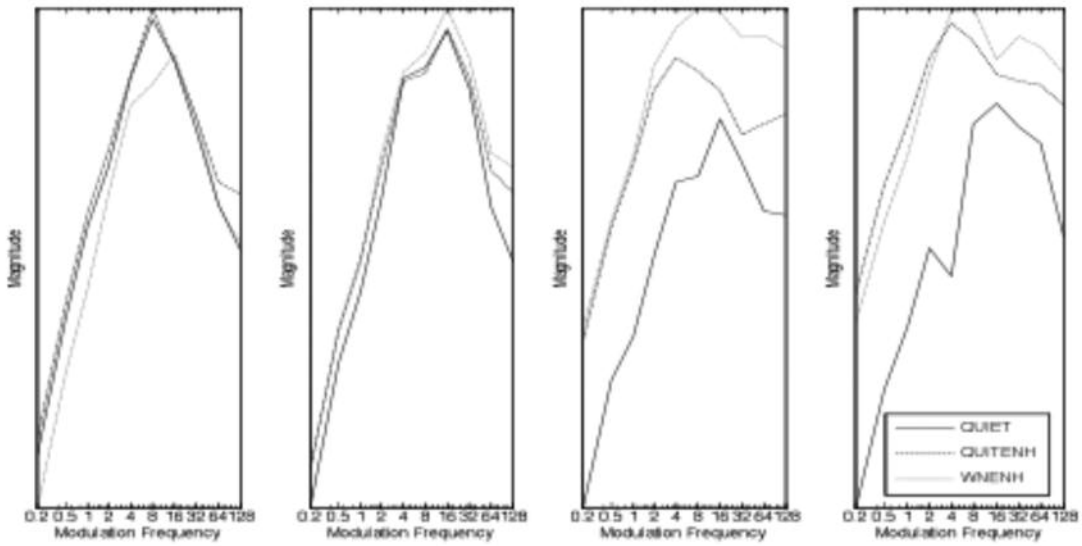

Figure 4. Modulation spectrum of the sentence "the ball bounced very high". Each panel represents four spectral bands of the same sentence (QUIET, WN and WNENH)..

It can be observed from the above figure that the white noise alters the modulation spectrum of the speech signal at least in high frequency spectral bands. When the envelope is expanded for this altered envelope, it may not result in improved scores. Probably envelope expansion could have been useful if only modulation depth is being reduced in the presence of background noise. The modulation spectrum of the speech in noise with envelope expansion is not similar to the modulation spectrum of the speech in quiet. This implies that envelope expansion does not restore the cues which are altered by the noise. It was also observed that some of the modulation dips were completely filled up or nearly flattened (Figure 4), posing difficulty to the algorithm to identify those dips and expand it. These apparent reasons determine the poor performance of the participants under envelope expansion scheme.

Conclusion

The current study evaluated the effect of envelope expansion when primarily temporal cues are used. The envelope expansion was evaluated in quiet and in the presence of background white noise. 40 HINT sentences were subjected to envelope extraction at 400 Hz and the envelope was expanded by 15dB at 1-30Hz bandwidth. It is concluded that the envelope cues alone are sufficient for good speech recognition in quiet but not in the presence of noise. Envelope expansion scheme does not affect the sentence perception containing only envelope cues in quiet as well as in noise.

References

1. Smith ZM, Delgutte B, Oxenham AJ. Chimaeric sounds reveal dichotomies in auditory perception. Nature. 2002; 416:87–90.

2. Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M. Speech recognition with primarily temporal cues. Science. 1995; 270:303-304.

3. Fu QJ, Shannon RV. (1999) Phoneme recognition by cochlear implant users as a function of signal-to-noise ratio and nonlinear amplitude mapping. Journal of the Acoustical Society of America.1999; 106:L18–L23.

4. Zeng FG, Galvin JJ. Amplitude mapping and phoneme recognition in cochlear implant listeners. Ear and Hearing. 1999 Feb; 20 (1):60-74.

5. Stickney G, Zeng FG, Litovsky R, and Assmann P. Cochlear implant speech recognition with speech masker. The Journal of the Acoustical Society of America. 2004; 116:1081–1091.

6. Nie K, Stickney G, Zeng FG. Encoding Frequency Modulation to Improve Cochlear Implant Performance in Noise. IEEE transactions on biomedical engineering. 2005; 52(1):64-73.

7. Duquesnoy A, Plomp R. Effect of reverberation and noise on the intelligibility of sentences in cases of presbycusis. Journal of the Acoustic Society of America. 1980; 64:537-544.

8. Assmann PF, Summerfield Q. The perception of speech under adverse acoustic conditions. In Speech Processing in the Auditory System. Springer Handbook of Auditory Research. 2004; 14.

9. Drullman R, Festen JM, Plomp R. (1994) Effect of reducing slow temporal modulations on speech reception. Journal of Acoustic Society of America. 1994; 95: 2670-2680.

10. Hou Z, Pavlovic CV. Effects of temporal smearing on temporal resolution, frequency selectivity, and speech intelligibility. Journal of Acoustic Society of America. 1994; 96: 1325–1340.

11. Clarkson PM, Bahgat SF. Envelope expansion methods for speech enhancement. Journal of Acoustic Society America. 1991; 89: 1378–1382.

12. Narne VK, Vanaja CS. Effect of envelope enhancement on speech perception in individuals with auditory neuropathy. Ear Hearing. 2008; 29:45-53.

13. Fu QJ, Shannon RV 1998. Effects of amplitude nonlinearity on phoneme recognition by cochlear implant users and normal-hearing listeners. Journal of Acoustic Society America. 1998; 104: 2570-2577.

14. Lorenzi C, Berthommier F, Apoux F, and Bacri N. Effects of envelope expansion on speech recognition. Hearing Research. 1999; 136:131–138.

15. Apoux F, Crouzet O, and Lorenzi C. Consonant recognition in noise with temporal cues: II. Effects of envelope enhancement on response times. 2000; 139th Meeting of the Acoustical Society of America. Atlanta, USA.

16. Nilsson M, Soli SD, Sullivan JA. Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise. Journal of the Acoustical Society of America. 1994; 95:1085–1099.

17. Apoux F, Tribut N, Debruille X, and Lorenzi C. Identification of temporally expanded sentences in normal-hearing and hearing-impaired listeners. Hearing Research. 2004; 189: 13-24.

18. Rosen S. Temporal information in speech: Acoustic, auditory and linguistic aspects. Philosophical Transactions of the Royal Society of London, Series B.1992; 336: 367-373.

19. Bregman, AS. What is auditory scene analysis? (In special issue on auditory scene analysis). Journal of the Acoustical Society of Japan. 1994; 50 (12):1007-1010.

20. Cooke, MP. A glimpsing model of speech perception in noise. Journal of Acoustic Society of America. 2006; 119:1562–1573.