

**Analyzing the Contribution of Envelope Modulations to the  
Intelligibility of Reverberant Speech**

by

**Ramesh Kumar Muralimanohar**

B.E., Madras University, 2003

M.S., University of Colorado, Boulder, 2010

A thesis submitted to the  
Faculty of the Graduate School of the  
University of Colorado in partial fulfillment  
of the requirements for the degree of  
Doctor of Philosophy  
Department of Speech, Language, and Hearing Science

2018

ProQuest Number: 10792544

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



ProQuest 10792544

Published by ProQuest LLC (2018). Copyright of the Dissertation is held by the Author.

All rights reserved.

This work is protected against unauthorized copying under Title 17, United States Code  
Microform Edition © ProQuest LLC.

ProQuest LLC.  
789 East Eisenhower Parkway  
P.O. Box 1346  
Ann Arbor, MI 48106 – 1346

This thesis entitled:  
Analyzing the Contribution of Envelope Modulations to the Intelligibility of Reverberant Speech  
written by Ramesh Kumar Muralimanohar  
has been approved for the Department of Speech, Language, and Hearing Science

---

Dr. Kathryn Arehart

---

Mr. James Kates

---

Dr. Lewis O. Harvey

---

Dr. Brenda Schick

---

Dr. Neeraja Sadagopan

Date \_\_\_\_\_

The final copy of this thesis has been examined by the signatories, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.

IRB protocol #15-0533

Muralimanohar, Ramesh Kumar (Ph.D., Speech, Language, and Hearing Science)

Analyzing the Contribution of Envelope Modulations to the Intelligibility of Reverberant Speech

Thesis directed by Dr. Kathryn Arehart

We routinely hear speech that is degraded by both noise and reverberation due to the characteristics of our listening environments. These intrusions alter the spectral and temporal envelope cues that support speech perception. The effect of these perturbations is more intrusive for people with hearing impairments.

This work systematically analyzed the potential benefits of restoring the speech envelope information in reverberant speech through modulation spectral analysis, through objective predictions of intelligibility from quantifying the changes in spectrotemporal modulations due to reverberation using cepstral correlation and through listener tests of intelligibility. We examined the benefits of envelope restoration on the intelligibility of reverberant speech using 1. an ideal restoration that used the clean envelope from anechoic speech and 2. Several processing strategies that restored the envelope by expanding the reverberant envelope in multiple bands in both listeners with normal hearing and in listeners with hearing loss. Intelligibility changes in reverberation were shown to occur largely from the changes to the low-rate modulations in the speech envelope. Cepstral correlation was found to be a better descriptor of average performance with and without hearing-loss and individual performance in most of our listeners. Envelope restoration through reinstating the low-rate modulations ( $< 30$  Hz) was found to effectively restore speech intelligibility in all the reverberation conditions tested here. None of the expansion processing schemes provided significant benefits in reverberation. However, the success of the ideal envelope restoration and the usability of low-rate cues indicated the possibility of using similar strategies in signal processing aimed at improving intelligibility in hearing aids.

## **Dedication**

To my family and my mentors, who believed and supported me through this process. This work is also dedicated to the people who this research could potentially benefit in the future.

## Acknowledgements

I would first like to thank my mentors Kathy and Jim, without your mentorship and help none of this would be possible. Their dedication, work ethic, endless patience, motivation and faith in me have shaped me as a researcher and a person.

I would like to acknowledge my committee: Dr(s) Schick, Sadagopan, Harvey and Tollin, who helped shape my research questions and supported my growth. Dr(s) Ramsberger, Gilley, Sharma, Brennan, Bonino, Baiduc, Coady, Pollard and Fredrickson for the insightful discussions and support. The department as a whole for being my home and family away from home, especially Raul and Brenda D. My mentors, Curtis and Michelle, and colleagues at the NCRAR for their support and guidance.

My lab group and cohort: Kristin Sommerfeldt and Melinda Anderson who both remain huge role-models and inspire me to do more. Elizabeth McNichols and Emily Lundberg for their help and company through "that" unforgettable summer. In-ki Jin, Naomi Croghan, Cory Portnuff, Jessica Rossi-Katz for their valuable knowledge and experience. Julia Campbell, Garrett Cardon, Nick Walker, Hannah Glick, Marcia Walsh-Aziz, Katie Miller and other fellow students for their friendship, discussion and sharing in the experience of graduate school. To my students who kept me on my toes.

My research participants, for hours of their lives spent listening to sounds and for their insights that helped me understand their experience.

Most importantly, my family and friends: Dad and Mom, Karuna, Prasanna and Spandana for keeping the faith, support and patience with my eccentricities. My friends for keeping me sane.

## Contents

### Chapter

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Intelligibility Cues in the Speech Signal . . . . .	3
1.2	Room Acoustics . . . . .	7
1.3	Effects of Reverberation on the Speech Signal . . . . .	10
1.4	Effects of Reverberation on Intelligibility . . . . .	12
1.5	Reverberation and Hearing Aids . . . . .	13
1.6	Envelope Restoration for Reverberation . . . . .	15
1.7	Experimental Outline . . . . .	17
 <b>2</b>	 <b>Envelope Modulations in a Single Reflection</b>	 <b>19</b>
2.1	Introduction . . . . .	19
2.2	Methods . . . . .	21
2.2.1	Stimuli and Simulation Condition . . . . .	21
2.2.2	Analysis . . . . .	21
2.3	Results and Discussion . . . . .	23
2.3.1	Effects of Delay Time and Strength of the Reflection on Predicted Intelligibility	23
2.3.2	Exploring changes in the modulation envelope . . . . .	24
2.4	Conclusions . . . . .	28

<b>3</b>	<b>Envelope and Reverberation in Real Rooms</b>	<b>30</b>
3.1	Introduction . . . . .	30
3.1.1	Reverberation and Intelligibility of Speech . . . . .	30
3.1.2	Room Acoustics and Intelligibility . . . . .	31
3.1.3	Speech Envelope and Intelligibility in Reverberation . . . . .	33
3.1.4	Envelope Alteration and Speech Intelligibility in Reverberation . . . . .	34
3.1.5	Envelope Restoration to Restore Intelligibility . . . . .	36
3.1.6	Reversing Reverberation . . . . .	36
3.1.7	Experimental Aims . . . . .	38
3.2	Methods . . . . .	39
3.2.1	Listening Environments and reverberant stimulus generation . . . . .	40
3.2.2	Envelope Manipulation . . . . .	41
3.2.3	Envelope Expansion . . . . .	46
3.2.4	Objective Signal Analysis . . . . .	48
3.2.5	Analysis of Envelope Modulations and Predicted Intelligibility . . . . .	50
3.2.6	Listener Intelligibility Tests . . . . .	50
3.2.7	Descriptive Measures of Hearing and Cognitive Abilities . . . . .	52
3.2.8	Experimental Design . . . . .	54
3.2.9	Monaural Intelligibility Tests . . . . .	54
3.3	Results . . . . .	54
3.3.1	Effect of Reverberation and Envelope-Only Processing on Intelligibility . . . . .	56
3.3.2	Intelligibility in Reverberant Baseline Conditions . . . . .	59
3.3.3	Intelligibility in Envelope Expansion Conditions . . . . .	64
3.3.4	Cue-based Performance and Benefits . . . . .	68
3.3.5	Metric Predictions of Word Scores . . . . .	71
3.4	Discussion . . . . .	76
3.4.1	Intelligibility, envelope-related changes and cepstral correlation in reverberation	77



3.4.2	Ideal envelope restoration of reverberant speech and the role of TFS . . . . .	83
3.4.3	Expansion to improve intelligibility of reverberant speech . . . . .	85
3.5	Conclusions . . . . .	90
<b>4</b>	<b>Conclusions</b>	<b>92</b>
4.1	Summary and Conclusions . . . . .	92
4.2	Future Work and Implications for Hearing aid signal Processing . . . . .	93
	<b>References</b>	<b>96</b>

## Tables

### Table

3.1	Acoustic Environments . . . . .	40
3.2	Description of the analysis filter bank . . . . .	41
3.3	Experimental conditions used in study. Subscript ‘cln’ indicates clean speech, and ‘voc’ indicates vocoder processing. ‘rev’ indicates reverberant speech. Exponent k – level of expansion applied to the envelope. . . . .	49
3.4	Hearing Profile of Participants including age, cognitive screener score, binaural speech-in-noise measure, audiometric thresholds and measures of temporal and frequency resolution ability. Subject numbers 1XXX are from the NH group and the 2XXX are from the group with hearing loss . . . . .	53
3.5	Experimental design, listing of levels based on rationale. The first set measured baseline anechoic performance to clean and envelope vocoded speech. (This condition was used compared to other conditions to estimate relative role of envelope). The second set measured performance with reverberant speech and envelope manipulations based on reverberant speech. The final set consisted reverberant and expanded envelope conditions used to assess performance of envelope expansion. A total of 11 unique conditions were used. . . . .	55
3.6	Results of two-way mixed-model RM-ANOVA for reverberant and clean conditions.	57
3.7	Results of three-way mixed-model RM-ANOVA analyzing the baseline reverberant conditions. b : Bonferroni Corrections; c : Greenhouse-Geisser Adjustment applied .	62

3.8	Pairwise Comparisons between scores in different rooms based on processing. Average percentage correct scores and significance values are provided for each condition. All comparisons were Bonferroni-adjusted. . . . .	62
3.9	Pairwise Comparisons between processing conditions in each room based on processing. Average percentage correct scores and significance values are provided for each condition. All comparisons were Bonferroni-adjusted . . . . .	63
3.10	Results of mixed model RM-ANOVA in the envelope expansion conditions. b : Bonferroni Corrections. . . . .	65
3.11	Pairwise Comparisons in the group with hearing loss. Mean values indicate change in scores from reverberant score in that room. . . . .	66
3.12	Pairwise comparisons of the effect of expansion processing between rooms with increasing reverberation. . . . .	67
3.13	Coefficient of determination ( $R^2$ ) values for individual STI and cepstral correlation values fit to proportion of words correct. The larger values for each comparison is in bold . . . . .	73

## Figures

### Figure

1.1	Speech Envelope . . . . .	4
1.2	Effect of Noise and Reverberation on Envelope modulations . . . . .	6
1.3	Reverberation in Rooms . . . . .	8
1.4	Spectral and Temporal Effects of Reverberation on Speech Signal . . . . .	11
1.5	Expansion of Reverberant Speech . . . . .	17
2.1	(a) N2 and N3 standard audiograms. (b) Relation between STI values and cepstral correlation coefficients for the NH model. (c) Cepstral correlation as a function of delay for several reflection strengths for models of NH (solid lines) and HI (N2 audiogram - dotted lines; N3 audiogram – dashed lines). . . . .	23
2.2	Contour plot of normalized envelope cross correlations averaged for 100 IEEE sentences comparing the envelope of the reverberated to clean speech. Top and Third Rows (L-R) show the coherence of envelope modulations at every auditory frequency per modulation frequency band for increasing delay times for NH and N2 respectively with a reflection strength of 0.85. Second and Fourth Rows (L-R) show the corresponding cepstral correlations at every auditory frequency per modulation band for increasing delay times for NH and N2. The spectral ripple used in the cepstral correlation calculation is indicated in cycles/spectrum, where the speech spectrum extends from 80 to 8000 Hz based on an auditory frequency spacing. . . . .	26

3.1	Vocoded conditions a. Original envelope vocoded (Clean Voc) condition and b. Reverberant envelope vocoder (Rev Voc) condition. . . . .	44
3.2	Ideal Envelope Restoration Processing . . . . .	45
3.3	Envelope Restoration Through Expansion . . . . .	47
3.4	Auditory thresholds of Participants in group with hearing loss . . . . .	52
3.5	Performance in Clean, Clean Voc and Reverberant Baseline Conditions . . . . .	57
3.6	Performance in Reverberant Baseline Conditions . . . . .	60
3.7	Performance with Envelope Expansion Processing . . . . .	64
3.8	Effect of Envelope on the Intelligibility of Reverberant Speech . . . . .	69
3.9	Effect of Ideal Envelope Restoration . . . . .	70
3.10	Benefit from Envelope Expansion . . . . .	71
3.11	Group Level Cepstral Correlation and STI Fits . . . . .	72
3.12	Individual fits of STI and Cepstral Correlation to Word Scores for group with hearing loss . . . . .	75
3.13	Individual fits of STI and Cepstral Correlation to Word Scores for NH group . . . . .	76
3.14	Envelope Modulation Correlations in Reverberation . . . . .	82
3.15	Envelope Modulation Correlations in Ideal Envelope Restoration . . . . .	86
3.16	Envelope Modulation Correlations in $f_{lo}$ . . . . .	88
3.17	Envelope Modulation Correlations in $a_{mag}$ . . . . .	89

## Chapter 1

### Introduction

People typically communicate in indoor environments. The acoustics of these enclosed spaces often impact the fidelity of the speech signal, which, in turn, affects our comprehension of the intended message. Reverberation, caused by numerous room reflections, affects the integrity of the spectral (frequency) and temporal (envelope) cues in the speech signal. These cues are known to guide the listener's perception and the effects of these distortions are further magnified in people with hearing loss. The temporal envelope of speech, in particular, is known to support speech intelligibility not only in people with normal hearing (NH) but also in people with hearing loss (HL).

The primary purpose of this study was to understand how reverberation impairs intelligibility, particularly, its impact on the speech envelope and the extent to which envelope restoration could improve the intelligibility of reverberant speech presented monaurally over headphones. Though binaural approaches that address reverberation exist, understanding monaural speech reception should provide important insights for better individually-tuned algorithm design.

Analyses of speech intelligibility in rooms have generally focused on quantifying narrowband changes to the speech envelope and do not typically include the impact of hearing loss. An across-band spectrotemporal descriptor or metric that includes an auditory model would not only provide a better description of the envelope changes due to reverberation but would also extend these analyses to models of hearing loss. Ultimately, these methods could provide ways to evaluate signal processing in hearing aids that are designed to counter reverberation.

This dissertation reports two studies that considered the effects of reverberation on speech. The first experiment identified the changes to the speech envelope which resulted in degraded intelligibility in a simple model of reverberation that used a single reflection. A concise descriptor/metric that quantified these changes was evaluated with a range of reflection times and strengths. The evaluation also considered both normal and cochlear hearing loss models. Results showed that a) envelope fidelity at low modulation rates appears to be essential for maintaining intelligibility b) results for the models of normal hearing and of hearing loss yielded similar patterns of change across reflection time and strength and c) that the proposed metric explains a larger percentage of the variability for both models of normal and compromised hearing than the conventional narrowband measure.

The second experiment extended the evaluation of this metric to more realistic scenarios of reverberation for people with normal hearing and with hearing loss. The effectiveness of envelope restoration in reestablishing the intelligibility of reverberant speech was also examined in both these groups. Monaural intelligibility scores were obtained for speech processed through different room reverberation responses and subjected to different amounts of envelope restoration, and these listener results were compared to the metric predictions. The metric effectively explained intelligibility in both groups. Ideal envelope restoration provided a means to restore intelligibility in reverberant situations for listeners with mild-to-moderate hearing loss. Practical envelope restoration schemes based on envelope expansion were also tested. Uniform multi-band (UMB) expansion where the same amount of expansion was applied in multiple frequency bands and independent multi-band (IMB) expansion where band-specific expansion factors were used were tested. Though neither of these methods provided significant benefits over reverberation, some of the methods resulted in small improvements in intelligibility. In spite of the lack of benefit seen with these methods, the benefit seen in the ideal case establishes the case for pursuing other envelope-based methods.

The following sections of this introduction provide a background and rationale for the two experiments. We begin with a description of what envelope cues are and how they relate to speech intelligibility in different populations. We then introduce reverberation and review its effect on

both the acoustics and the intelligibility of speech. Current de-reverberation algorithms and the interaction of reverberation with other common processing algorithms used in hearing aids are also considered. Finally, we present the potential of envelope restoration as a means of restoring intelligibility in hearing aids.

## 1.1 Intelligibility Cues in the Speech Signal

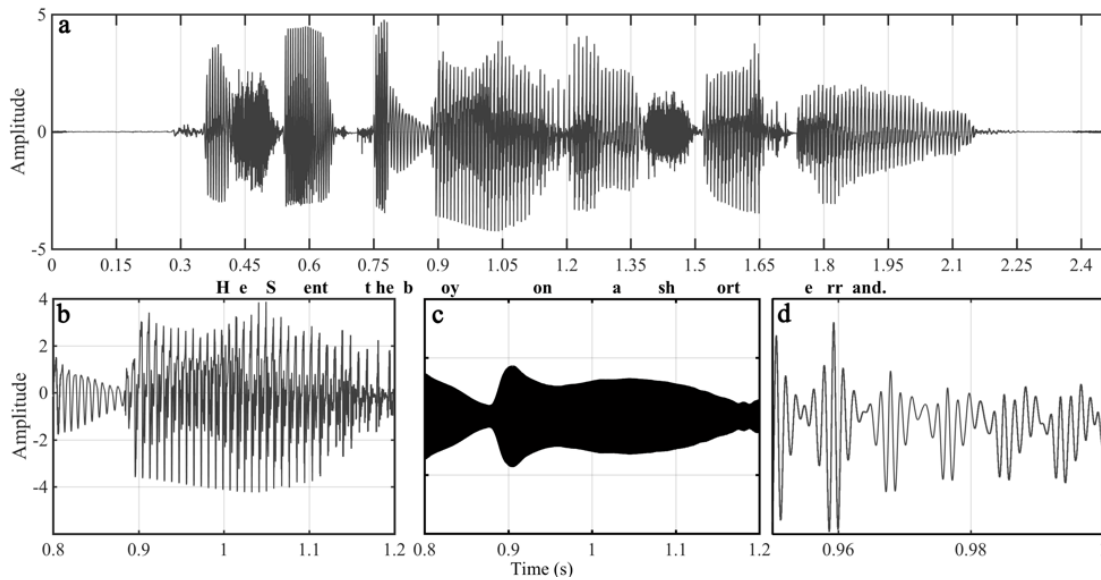
The temporal and spectral cues contained in speech carry the information required to understand the meaning behind the sounds. Some approaches to understanding the acoustics-perception relation of speech are focused on the spectral features of speech segments and how they support perception (Peterson and Barney, 1952). Other experiments have focused on temporal aspects of the speech signal. An early example of the importance of envelope cues was the artificial synthesis of speech (Dudley, 1939), in which a spectrally sparse model using only the slow changes in the speech envelope allowed for the creation of intelligible speech. Further validation of the importance of these cues has come from the success of early cochlear implants, which primarily use envelope cues to convey speech information (Hochmair-Desoyer et al., 1980).

A temporal description of speech is based on its definition as a series of modulations superimposed on carriers in multiple frequency bands. The signal in each frequency band consists of three components: 1. the slow varying (low-frequency) modulations called the temporal envelope (2 - 50 Hz) 2. periodicity cues (50 - 500 Hz) and 3. the rapid transitions that are based on the center frequency of the band called the temporal fine structure (TFS) (above 500 Hz). A sentence from the IEEE corpus (IEEE, 1969) is used to illustrate two of these components in Figure 1.1. Figure 1.1a shows the time waveform of the complete sentence. A short segment from the sentence, the word "boy", is shown in Figure 1.1b. The slow modulations of the temporal envelope (Figure 1.1c, shows the envelope for the word "boy") carry segmental information about manner, voicing and vowel identity as well as prosodic (suprasegmental) cues about tempo, rhythm, syllabicity and stress (Rosen, 1992). The periodicity cues also offer segmental information about manner and voicing and



prosodic information about stress and intonation.

Figure 1.1: a. Speech signal from IEEE Corpus, b. corresponding segment corresponding to the word “boy”. c. corresponding broadband envelope [2-50 Hz], and d. temporal fine structure extracted from about 50 ms of the same word. (TFS) [above 600 Hz].



The TFS cues (see Figure 1.1.d) are useful in identifying segmental information about manner, place, voicing and quality. Envelope, periodicity and TFS cues are all important for understanding speech. Rosen (1992) presents a comprehensive overview of the importance of these cues. Several researchers have theorized that envelope cues (also referred to as low-frequency modulations) are most important for speech perception (Smith, Delgutte and Oxenham, 2002). The importance of these cues has been reinforced by studies showing that a reduction in low frequency envelope modulations ( $<20$  Hz) results in a lowering of intelligibility for normal-hearing listeners (Houtgast and Steeneken, 1985; Drullman, Festen and Plomp, 1994; Duquesnoy and Plomp, 1980). Shannon et al. (1995) showed that good speech perception (about 80% correct) for sentences could be achieved with only four bands of temporal envelope information with envelope information lowpass filtered at 16Hz in the absence of spectral detail. Other studies have shown that envelope cues of 1-20Hz

are most useful in speech perception in both listeners with normal-hearing and with hearing loss (Kanedera et al., 1999; Plomp, 1988).

Predictive metrics of speech intelligibility have also supported the importance of slow envelope modulations for speech intelligibility (Houtgast, Steeneken and Plomp, 1980; Kates and Arehart, 2015). Kates and Arehart (2015) used a metric-based computational model of the auditory system to show that the envelope modulations (particularly in the low-to-mid auditory frequencies) were important in predicting the intelligibility of hearing aid processed speech.

In order to understand how changes to the speech envelope affect perception, it is first important to identify the modulations present in clean, anechoic speech that support intelligibility. Then, changes to these cues can be used to explain changes in intelligibility due to the external interferences like noise and reverberation. This approach was used by Houtgast and Steeneken (1973). They first obtained the speech-envelope spectrum by analyzing the temporal fluctuations of the intensity envelope of anechoic speech. This envelope extracted in octave bands was squared and low pass filtered, and then was further analyzed through 1/3 octave band modulation filters with center frequencies between 0.5 – 16 Hz. Adapted from Houtgast and Steeneken, Figure 1.2 (solid curve) shows this information. The modulation index reflects the amount of modulation, measured as the ratio of the envelope peaks to its valleys, with 1 indicating complete modulation and 0 referring to no modulation. The graph shows that the maximum modulations are centered at about 3 Hz for clean, uninterrupted speech. This pattern of modulation is found to be similar for speech produced in a variety of conditions (different talkers, rates, accents etc.).

Early attempts to characterize the relation between the physical environment and speech intelligibility examined how the modulation spectrum shown in Figure 1.2 is altered by transmission in rooms and how these changes relate to listener intelligibility. The extent to which the envelope is affected by a system (or processing) is called the modulation transfer function (MTF) and is sensitive to the effects of noise and reverberation (Houtgast and Steeneken, 1973, 1978 and 1985; Houtgast, Steeneken and Plomp, 1980). The MTF is defined as the reduction in the modulation index of the envelope of altered speech relative to clean unaltered speech for modulation rates between 0.5 – 16

Hz. The Speech Transmission Index (STI) predicts speech intelligibility based on measuring the modulation depth. The speech is divided into octave frequency bands from 125 to 8000 Hz, and the MTF is then estimated within each band over the range of modulation rates. The MTF values are then averaged across modulation rate and auditory frequency to produce the intelligibility estimate.

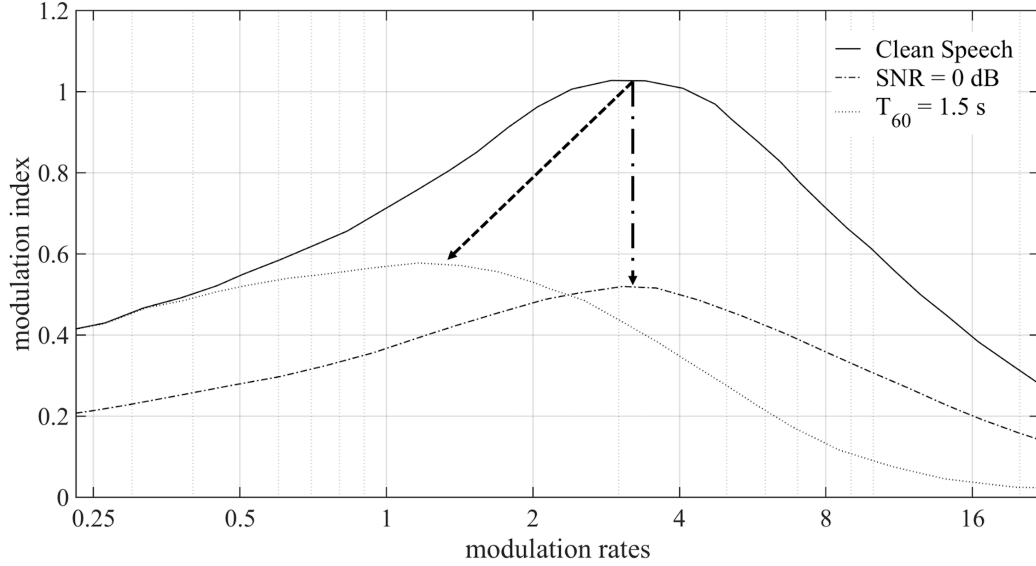


Figure 1.2: Speech envelope spectrum showing fluctuations in clean (solid curve) and the reductions in noise and reverberation (dashed lines). A modulation index of 1 indicated complete modulation and 0 no modulation (Houtgast and Steeneken, 1985). Solid curve shows clean, anechoic speech and the dashes curves show changes due to addition of noise and reverberation. Reverberation acts like a low pass filter moving the peak in modulation energies from around 3 Hz to lower frequencies whereas noise tends to lower the useful modulations at all modulation rates shown. Figure adapted from Houtgast and Steeneken, 1985)

The STI has been used to successfully relate the characteristics of listening environments and the resulting changes in the envelope to speech intelligibility. Such calculations show that reverberation causes a reduction of the low frequency modulations resulting in a decrease in STI values. We shall revisit the specifics of such a calculation in later sections when we use a short-term

version of the STI to analyze the changes in predicted intelligibility due to reverberation. Next, we will describe how listening spaces in which we communicate affect speech.

## 1.2 Room Acoustics

Listeners in typical rooms hear a superposition of the direct sound produced by the sound source (e.g. a talker) and several attenuated versions of the original sound having short delays. These short delays are followed by multiple overlapping reflections at longer delays. Called reverberation, the reflections from the walls and ceilings of the room result in the persistence of the sound energy after the original sound has ceased (see Figure 1.3b).

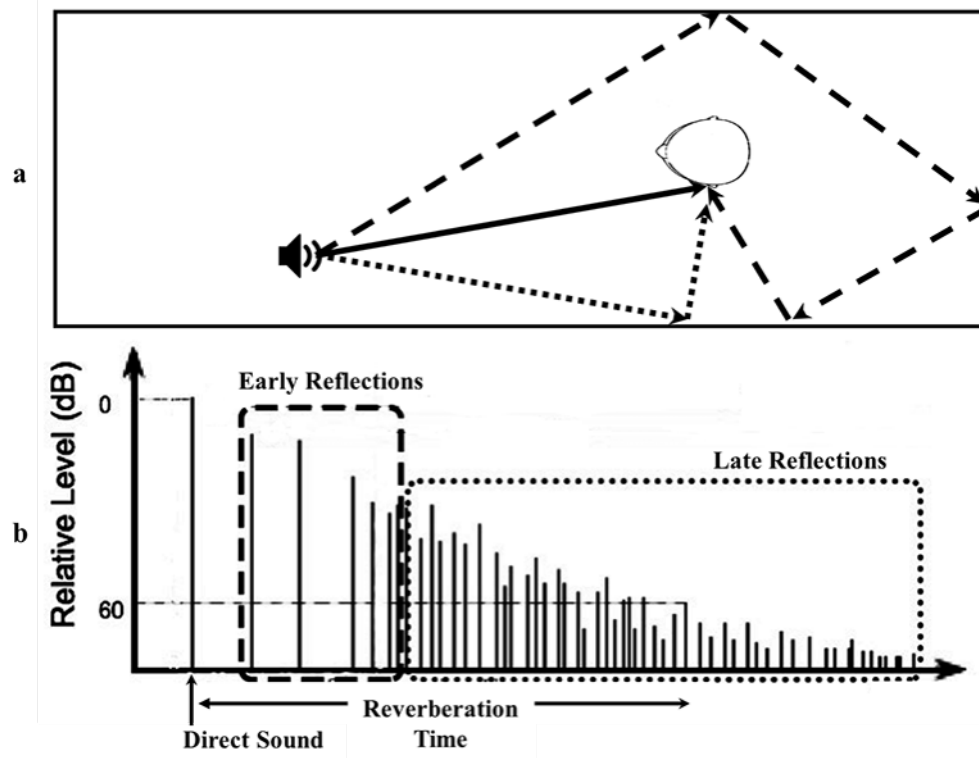


Figure 1.3: a: Sound in a room: Rectangle represents the walls and the sound from the speaker is received at the left ear of the listener. The solid line represents the direct sound to the left ear, the dotted line an earlier reflection and the dashed line a later reflection. b: Sequence of direct sound and reflections reaching the listener's ears. Figure 1.3b. adapted from Nábělek, 1994)

These reflections (indicated by the early and late reflections in Figure 1.3b right) arrive at the listener's ear quickly, i.e. within tens of milliseconds after the direct sound (first impulse indicated in Figure 1.3b). Some of these reflections blend perceptually with the direct sound. Reverberation is characterized by two distinct parts: 1. Early reflections which are the clearly-defined, stronger, well separated reflections that arrive within 50-80 ms after the direct (see Figure 1.3b), and 2. the late reflections or reverberant tail which consists of densely grouped, lower amplitude reflections that can be characterized as noise-like (see Figure 1.3b). Early and late reflections differentially affect the speech signal leading to different effects on its intelligibility.

When describing the acoustics of individual reflections, we use reflection time and reflection

coefficient. Reflection time is the time it takes for a reflection to reach the listener’s ear. The strength of each reflection is reflected in acoustic analysis through the reflection (or absorption) coefficient parameter. The reflection time depends not only on the dimensions and geometry of the room, but also on the position of the talker and listener with respect to the walls. The reflection (or absorption) coefficient usually depends on the properties of the material of the reflecting surfaces (e.g., the walls, ceilings etc.) and also varies based on the frequency of the incident sound. The strength of direct sound decreases with distance from the source but the level of reverberation remains more or less steady as a function of location. To accurately model a room, we need to consider the direct sound and all the reflections in the room. Thus, modeling the reverberation in a typical room is complicated because of the numerous reflections.

The amount of reverberation in a space can also be characterized through the reverberation time ( $T_{60}$  or  $RT_{60}$ ).  $T_{60}$  is the amount of time it takes for the sound level to fall 60 dB below the original level (shown in Figure 1.3 b) from when the source stopped producing the sound (ANSI-S1.1,1960, R 1976; Boothroyd, 2002).  $T_{60}$  describes how long the reflections “persist” in a room and might interfere with the perception of the direct sound.  $T_{60}$  can be measured individually for several frequencies or as a single number for a broadband signal to represent the acoustics of a room. Researchers often use  $T_{60}$  as a way to characterize the impact of room acoustics on the perception of speech sounds (Lochner and Burger, 1964; Nábělek, 1994; Houtgast, Steeneken and Plomp, 1980; Warzybok et al., 2013). A typical office has a  $T_{60}$  of 200 – 300 ms, while concert halls have values between 1.5 – 2.5 s. In architecture,  $T_{60}$  is used as a guideline for designing rooms for specific purposes. For example, the recommended  $T_{60}$  for classrooms for optimum speech perception is 600 ms (ANSI S12.60, 2002). Another measure used to characterize room reverberation is the Direct-to-Reverberant Ratio (DRR) which is a ratio of the energy in the direct sound to the energy in the reflections (ISO 3382). The first 2.5 ms of the sound are generally used to calculate the energy in the direct sound.

Reverberation often reduces the intelligibility of speech (Lochner and Burger, 1964; Nábělek, 1994). In general, the deleterious effects of reverberation on speech perception increase as  $T_{60}$

increases (Nábělek, 1994). This reduction in intelligibility is a direct result of the changes to the spectral and temporal cues present in the speech signal.

### 1.3 Effects of Reverberation on the Speech Signal

Reverberation results in both spectral and temporal smearing of speech. The effect of reverberation in a room with a  $T_{60}$  of 690 ms is illustrated in Figure 1.4, with clean speech in the left panels and the reverberant speech is shown in the right panels.

As shown in Figure 1.4, the multiple reflections muddle the speech signal. The two rows show the time waveform and spectrogram of clean and reverberant speech in the left and right panels respectively. The top row illustrates the effects of the superposition of multiple reflections in the received speech. The most obvious change is that the silent pauses between the words seen in the clean speech (top left panel) are filled in by the reflections (top right panel). In addition, the temporal envelope is smoothed with a reduction in the contrast between the peaks and valleys in the speech signal. This effect is also reflected in the modulation spectrum shown in Figure 1.2 (dashed and dotted lines). The dashed line shows how the envelope modulations are changed by a noise at 0 dB SNR. The dotted line shows the changes seen in reverberant speech with a  $T_{60}$  of 1.5 s. With reverberation, the modulations are not only reduced, like with the noise, but the peak modulations are shifted to the lower frequencies. Thus, the MTF associated with a reverberant space is like a low pass filter (Houtgast and Steeneken, 1985) smoothing the envelope at all acoustic frequencies. The changes in low frequency modulations are related to the loss in intelligibility associated with each interference as predicted by metrics like the STI.

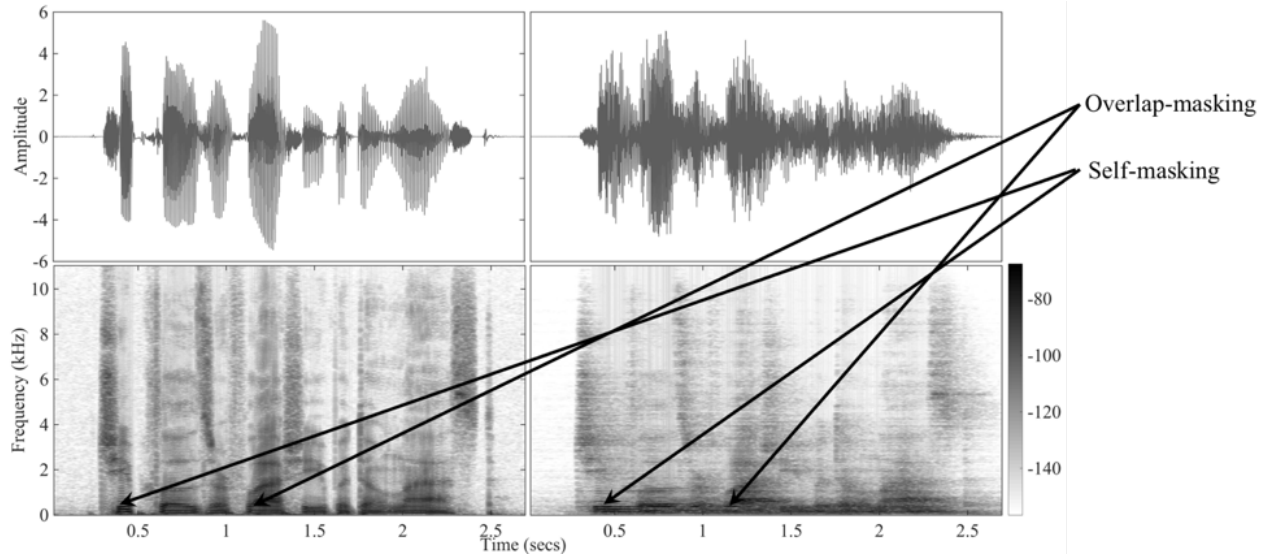


Figure 1.4: Effects of reverberation on the speech signal showing overlap and self-masking. Clean speech is shown in the left panels; the addition reverberation in a room with  $T_{60} = 690$  ms is shown in the right panels. Top panels show the waveform of speech and the bottom panels show the spectrograms of the corresponding waveforms.

The bottom panels of Figure 1.4 show the spectrotemporal changes due to reverberation. Spectral smearing of segments both by themselves and preceding elements can be observed (see bottom panels of Figure 1.4). These effects cause two main problems (Nábělek, 1994): 1. overlap masking - when energies from previous phonemes are mixed with succeeding phonemes and, 2. self-masking - when energy from the same phoneme is imposed on itself. Overlap masking, in particular, exhibits a complex pattern of interruptions based on energy overlap. The amount of energy overlap is determined by several factors. Greater  $T_{60}$ s cause greater interruptions due to longer carryover of reverberant energies. The relative intensities of current and preceding phonemes and corresponding spectral contrasts create differential patterns of change based on context. In general, consonants are more affected than vowels due to their shorter duration and lower intensities, but relative amplitudes and duration ultimately determine the effect of reverberation. The speaking rate can also factor in the effect of reverberation due to temporal salience of words and other suprasegmental changes.



However, not all the reflections in reverberation are bad for speech perception. Earlier reflections are “good” and later reflections are “bad” for speech intelligibility. That is, the early reflections ( $\leq 50$  ms after the direct sound) reinforce the direct sound, resulting in an increased signal-to-noise ratio (SNR) and are integrated with it, thereby aiding intelligibility (Rennies et al., 2014; Warzybok et al., 2013; Arweiler and Buchholz, 2011; Nábělek and Robinette 1978; Lochner and Burger, 1964). On the other hand, the late reflections ( $> 50$  ms after the direct sound) do not reinforce the direct sound but rather “smear” the speech and result in a loss of intelligibility (Nábělek, 1994) by reducing the modulation depth, causing masking and a lowering of the effective SNR (Steeneken and Houtgast, 1980; Warzybok et al., 2013).

This dichotomy makes it necessary to characterize the reverberation characteristics of spaces with more than just the  $T_{60}$  when analyzing speech intelligibility effects in those spaces (ISO 3382). For example, the clarity metric ( $C_{50}$ ) is measured as the ratio of the early to the late reverberation, with the boundary between early and late set at 50 ms. Studies have directly investigated the effects of varying acoustic spaces on speech intelligibility by using  $T_{60}$ ,  $C_{50}$  and DRR.  $T_{60}$  is the most commonly used metric, through other metrics like the  $C_{50}$  or DRR might be more applicable to certain experiments, based on their sensitivity to early reflections.

## 1.4 Effects of Reverberation on Intelligibility

Reverberation affects speech intelligibility even in quiet. Increasing reverberation, even when characterized as  $T_{60}$ , leads to lowered speech understanding scores. NH listeners are able to overcome small amounts of reverberation easily while greater amounts of reverberation impair their ability to understand speech (e.g. Lochner and Burger, 1964; Nábělek, 1994). Generally, rooms with  $T_{60}$  measures of up to 1.2 s are acceptable for speech perception for NH listeners. Among people with normal hearing, children and older adults seem to have greater problems perceiving speech in reverberant situations (e.g. Nábělek and Robinson, 1982; Finitzo-Hieber and Tilman, 1978).

Individuals with hearing loss are generally more susceptible to reverberation though the in-

telligibility scores exhibit the same patterns of change with increasing reverberation as NH listeners (Nábělek, 1994). This increased susceptibility is attributed to the reduced spectral and temporal resolution and elevated thresholds that are characteristic of cochlear hearing loss (Moore, 2007). So, even in rooms with short reverberation times ( $T_{60} = 0.4$  ms), listeners with hearing loss exhibit greater reductions in intelligibility than observed in normal-hearing listeners (Beutelmann and Brand, 2006; Gordon-Salant and Fitzgibbons, 1993; Harris and Swenson, 1990; Helfer and Wilbur, 1990; Nábělek and Pickett, 1974; Duquesnoy and Plomp, 1980). The recognition scores for reverberant speech worsen (by about 20 percentage points) as hearing loss worsens from a mild loss to a more moderate-to-severe loss and with age (Nábělek, 1982). Gordon-Salant and Fitzgibbons (1993) have shown that listeners with hearing loss experience similar decrements in reverberant speech perception even when the speech is presented at audible levels.

Background noise is also known to compound the effect of reverberation on speech intelligibility, especially at moderate signal-to-noise ratios (SNR). The effect of noise and reverberation together is more than the effect of each intrusion individually (George, Festen and Houtgast, 2008; Finitzo-Hieber and Tillman 1978; Duquesnoy and Plomp 1980; Wroblewski et al., 2012). Thus, normal everyday conditions of noise and reverberation that appear benign for normal-hearing listeners might pose great challenges for listeners with hearing loss and hearing aid users. The causes of these perceptual difficulties also seem to extend beyond just audibility and deficits due to aging.

## 1.5 Reverberation and Hearing Aids

Nearly a third of the adults with hearing loss hearing in the US use hearing aids (NIDCD, 2018). Though the use of hearing aids does improve speech intelligibility in reverberation (Ricketts and Henry, 2002), people who wear hearing aids still perform worse at speech identification in reverberation than normal-hearing listeners under the same listening conditions (Nábělek and Pickett, 1974). Listeners with hearing aids also experience high levels of displeasure in their device in the presence of reverberation (Kochkin, 2005, 2010).

De-reverberation algorithms that completely cancel reverberation in a room as the listener moves around are computationally complex (Benetsy, Sondhi and Huang, 2007). This complexity – together with the power limitations of possible implementations and the very strict requirements on allowable time delays in wearable devices - limits the availability of such processing strategies in commercial hearing aids. However, simpler algorithms that utilize the generic characteristics of reverberation have been used in hearing aids. Lebart et al. (2001) proposed a multi-channel spectral subtraction method that involved removal of the estimated reverberant tails from the speech signal. This method did not lead to significant improvements in intelligibility with reverberation in noise but was perceived as more comfortable than reverberant sounds (Fabry and Tchorz, 2005). More recently, Folkeard et al. (2017) tested a reverberation cancellation scheme based on the idea that reflected sounds are less intense than the direct sound from the source. Thus, limiting the amount of amplification provided to these sounds might help remove reverberation. This processing strategy showed modest improvements in intelligibility scores ( $\sim 7\%$  on average). However, the method combined directional microphones, noise reduction and reverberation processing, such that the specific contributions of the reverberation canceller were difficult to discern, making the evaluation of the reverberation canceller difficult.

Reverberation could potentially reduce the benefits of restored audibility from using hearing aids in difficult situations by interacting with signal processing in the hearing aid designed for other purposes. Few studies have looked at the effects of the possible interactions between hearing aid signal processing and reverberation (Reinhart et al., 2016; Giurda et al., 2017; Souza, Jenstad and Folino, 2005). Listeners with cochlear hearing loss typically lose their sensitivity to softer sounds while retaining near-normal ability for greater level sounds. This reduction in dynamic range has led to the use of non-linear (compressive) gain settings in hearing aids which might alter the normal envelope modulations seen in speech. Thus, wide dynamic range compression (WDRC) reduces envelope fidelity and as such, may exacerbate the effects of reverberation on intelligibility (Souza, Jenstad and Folino, 2005).

Current hearing aids typically have more than one microphone and these can sometimes

be combined binaurally (Kates, 2008). These microphones can be used particularly well when the target is separated from the source of interference. The signals from the microphones can be combined in ways to selectively focus on particular direction and attenuate the noise (beamforming). The benefits from such techniques diminish in the presence of reverberation and with increasing separation between sound source and receiver (Ricketts and Hornsby, 2003). Other algorithms like feedback cancellation (Kates, 2001) and digital noise reduction (Reinhart, et al., 2017) are also known to also potentially interact with reverberation and lead to reduced intelligibility of speech. Handling reverberation can thus lead to better outcomes with hearing aids. Thus, it might be worthwhile to consider whether reversal of the spectrotemporal changes caused by reverberation might restore intelligibility for listeners with hearing loss who use hearing aids.

## 1.6 Envelope Restoration for Reverberation

Information in clean speech is encoded redundantly in both its spectral and temporal patterns but reverberation degrades both types of cues. People with normal hearing have access to this information in multiple dimensions and can thus tolerate some interruptions in one or both dimensions. However, people with hearing loss, which causes a reduction in both spectral contrast and temporal acuity (Moore, 2008), have reduced access to both cues and thus experience more problems processing reverberant speech. Such listeners rely on certain cues more than others based on their hearing loss, context and task. While listeners with cochlear hearing loss sometimes show deficits on temporal processing, listeners with moderate to even severe hearing loss are often able to retain their ability to process low-rate temporal envelope information (Turner et al, 1995; Souza et al., 2015) when the sounds are sufficiently audible.

Listeners with such hearing loss might, in general, depend more on the temporal information to extract meaning from speech, especially in difficult listening situations (Souza et al., 2015; Davies-Venn and Souza, 2014). It is also true that listeners with hearing loss can use the multiple dimensions of information present in speech (Hedrick and Younger, 2001). Other evidence suggests that listeners

with hearing loss rely on temporal information when both spectral and temporal information are available (Hedrick and Younger, 2003). Souza et al. (2015) showed that listeners with hearing loss rely more on and use static spectral cues in combination with temporal cues but switched to temporal cues when the spectral information became more complex. Some of the variability in performance amongst listeners with hearing loss for speech interrupted by noise and reverberation might arise from this cue “dominance effect”. Francis, Baldwin and Nusbaum (2000) also showed that it might be possible to shift the dominant percept through auditory training. The temporal envelope therefore presents an important cue for speech intelligibility, and the envelope cues may be more important for HI listeners than for NH listeners.

We hypothesized that restoring the envelope of speech, especially the low-rate portions, could improve the intelligibility of speech in reverberant environments for such listeners. A large portion of the intelligibility deficits due to reverberation are due to a reduction in the modulation depth in multiple frequency bands (Noordhoek and Drullmann, 1997). Envelope expansion has been proposed as a solution to improving perception in situations where a reduction of envelope modulations can lead to loss of intelligibility.

An example of the effect of expansion on reverberant speech is illustrated in Figure 1.5 below. The most prominent effect seems to be the improvement in peak-to-valley ratio seen in segments around the 1.2 and 1.7s marks. The spectrogram shows slightly more apparent onsets in the expanded speech lost in the reverberant speech due to reverberant energy. Also, the fundamental frequency and first formant seem to be enhanced along with strengthening of some spectral landmarks that were smeared in the reverberant speech.

The effects of multi-band expansion, however, have not been systematically tested on reverberant speech. It remains unclear if people with hearing impairment can use these improvements in modulations to improve the understanding of reverberant speech. Simple expansion, if successful, could offer a computationally inexpensive way to restore intelligibility.

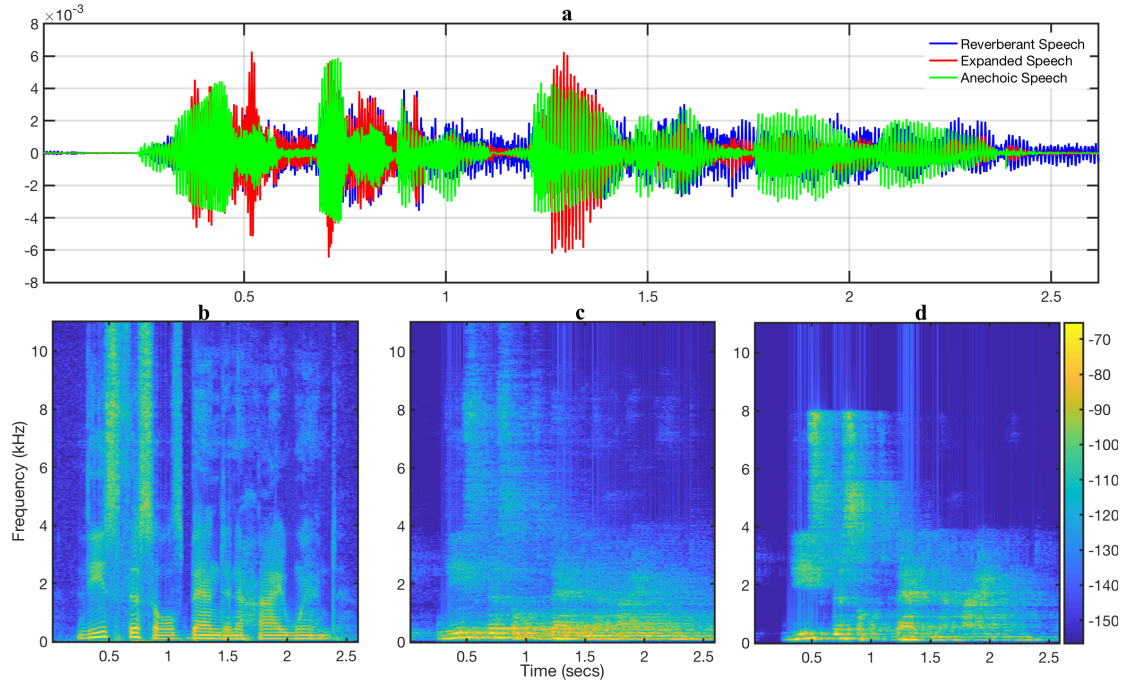


Figure 1.5: Top Row: a. Shows time waveforms of reverberant, expanded and anechoic speech overlaid on each other to show effects of expansion. Bottom Row: b. Left to Right: c. Spectrograms of clean, reverberant and c. expanded speech.

## 1.7 Experimental Outline

This work explored the role of envelope information in intelligibility of reverberant speech. The usefulness of envelope restoration to restore intelligibility of reverberant speech was evaluated. We hypothesized that envelope degradations result in a lowering in the intelligibility of reverberant speech especially for people with sensorineural hearing loss. While the MTF provides a good model that relates envelope changes to predicted intelligibility, it is limited because it ignores the spectrotemporal changes that span across several frequency bands. The model does not also account for deficits in spectral and temporal processing associated with hearing loss. People with hearing loss also vary significantly in their ability to understand speech in the presence of reverberation (Moore, 2013). Thus, we were motivated to explore the spectrotemporal changes due to reverberation across

all auditory frequencies and including a comprehensive auditory model to analyze the effect of hearing loss on reverberant speech perception. Chapter 2 presents our initial analysis of spectrotemporal causes of envelope degradations that result in intelligibility degradations in the presence of a single reflection.

Chapter 3 presents the extension of the analysis to actual reverberation in four rooms based on changes in perception in people with hearing loss and normal hearing. The idea that intelligibility can be largely predicted from changes to the envelope modulations suggests that the exact reproduction of the original envelope of speech would form the upper bound on intelligibility increments from envelope restoration and that the phase information of reverberant speech would be the only source of distortion leading to a decrement in intelligibility. Increasing envelope fidelity and retaining the low-rate cues should benefit the listener and lead to increases in the intelligibility of reverberant speech. As such, we examined the effects that several conditions with differing amounts of envelope and phase information has on speech intelligibility. Finally, we tested several envelope restoration conditions where we attempted to restore the envelope of reverberant speech using its clean envelope. The results of these analyses are also presented in this chapter. Our overall results supported the idea of pursuing envelope restoration-based solutions for improving speech communication in reverberation for use in auditory prosthetics like hearing aids.

## Chapter 2

### Envelope Modulations in a Single Reflection

#### 2.1 Introduction

Understanding speech relies on cues present in both the slowly varying temporal envelope and the rapidly changing temporal fine structure (TFS) of the speech (Rosen, 1992). Envelope information, especially at low modulation rates ( $<16$  Hz), is known to be important for speech understanding (Houtgast and Steeneken, 1985; Kates and Arehart, 2015), necessary for maintaining intelligibility in both listeners with normal hearing (NH) and listeners with hearing impairment (HI) (Drullman et al., 1994) and can provide meaning even under severe loss of spectral information (Wilson and Dorman, 2008). Evidence also exists that HI listeners can use these temporal envelope cues (Turner et al., 1995) when the stimuli are presented with sufficient audibility for that listener. Reverberation results in a lowering of intelligibility by interfering with both the envelope and TFS of speech. Given the usability of low modulation-rate envelope cues by HI listeners, restoring the envelope might allow us to restore intelligibility in the presence of reverberation for these populations. Also, HI listeners suffer greater perceptual deficits in the presence of reverberation; not only do they perform worse at a given level of reverberation but they also show problems at levels where NH listeners do not show any performance deficits (Nábělek and Nábělek, 1994; Gordon-Salant, 2006). Thus, it is important to systematically study how reverberation affects the envelope and impacts intelligibility, especially with hearing loss (e.g., elevated thresholds and decreased spectrotemporal resolution).



Reverberation in realistic listening environments (like living rooms, classrooms, auditoria, etc.) consists of multiple reflections at various amplitudes arriving with different delays, some arriving in very quick succession or overlapping with others. Early reflections (delays  $< 50$  ms) reinforce the direct sound, resulting in increased effective speech levels that aid intelligibility (Warzybok et al., 2013; Nábělek and Robinette 1978; Lochner and Burger, 1964). Later reflections (delays  $> 50$  ms), which are usually received at much lower levels compared to the direct sound and very closely grouped in time, result in a noise-like masking of the speech and reduction in intelligibility levels (Warzybok et al., 2013). Thus, reverberation in real rooms presents a very complex problem to unravel. A simpler approach to understanding the effect of reverberation is to start with a single reflection. While this approach does not reflect the full range of changes seen with a more complex reverberation, the envelope changes due to varying the delay time and the strength of the reflection can be easily understood.

Traditional approaches to speech intelligibility have focused on the within-band changes to the envelope modulation through the modulation transfer function (MTF) (Houtgast and Steeneken, 1985; Steeneken and Houtgast, 1980). Metrics like the speech transmission index (STI) form a weighted combination of the changes within each band without considering the spectral-temporal patterns across bands (Steeneken and Houtgast, 1980; Payton and Shrestha, 2013). The variations in the short-time spectrum from segment-to-segment also contain information important for speech recognition (Kates and Arehart, 2014; Zahorian and Rothenberg, 1981) which are not captured by the narrowband envelope analysis.

The current study extends these narrowband analyses (e.g., STI) by considering the changes in spectrotemporal modulations over time. The short-time spectrum is fit with a set of half-cosine basis functions. These functions represent features in a small segment of speech; for example, the second basis function describes spectral tilt. Changes in these spectral features over time can be used to identify vowels and consonants. The spectral principal components described by the basis functions explain about 97% of speech variance (Kates and Arehart, 2015; Zahorian and Rothenberg, 1981). An intelligibility metric that included a comprehensive peripheral auditory system model

(Kates and Arehart, 2014) was also used to extend these analyses to HI models.

## 2.2 Methods

The strength and delay time of a single reflection was systematically varied to investigate the spectrotemporal effects on the envelope and the resulting impact on speech intelligibility. The STI and cepstral correlation based on an auditory model were used to analyze the changes to the speech envelope due to the reflection. Changes in intelligibility were predicted based on NH and HI models. A modulation spectrum analysis was performed to analyze envelope fidelity.

### 2.2.1 Stimuli and Simulation Condition

Ten lists (100 sentences) from the IEEE Corpus (Rothhauser, 1969) were processed to simulate the effect of an added reflection. The sentences were resampled from 44 to 22 kHz for the analysis. The reflection coefficient and the length of the delay were parametrically varied in this study. The delays used were 10, 20, 30, 50, 70, and 100 ms. The strength of reflection was varied by using reflection coefficients of 0.7, 0.8, 0.85, 0.9, 0.95, and 0.99. We also analyzed the changes for two standard flat and moderately sloping [mild (N2) and moderate (N3)] audiograms (Bisgaard et al., 2010). The audiograms are presented in Fig. 1(a). Linear amplification was applied to these stimuli using the National Acoustics Lab Revised prescription (Byrne and Dillon, 1986) to analyze the HI models (model details and reference presented below).

### 2.2.2 Analysis

The amount of envelope change in the experimental conditions was measured using the correlation between the clean and reverberated envelopes in ten modulation frequency bands spanning the range from 0 to 325 Hz (see Table I in Kates and Arehart, 2015). Two metrics described below were used to calculate the predicted intelligibility from the amount of changes in the envelope modulations in each condition. Envelope regression (ER) time-domain STI method. The ER procedure

(Payton and Shrestha, 2013) was used to estimate the STI using a speech signal as the excitation. This method was chosen because it is a short-term speech-based calculation approach and it calculates the STI using envelope fidelity. The short-term ER calculation uses the band modulation metric, calculated over a rectangular window of length  $N$

$$M_i = \frac{\mu_{xi} \frac{1}{N} \sum_{k=1}^N [x_i(k)y_i(k)] - \mu_{xi}\mu_{yi}}{\mu_{yi} \frac{1}{N} \sum_{k=1}^N [x_k]^2 - \mu_{xi}^2} \quad (2.1)$$

where  $x_i(k)$  and  $y_i(k)$  are the intensity envelopes of the clean and reverberated signals and  $\mu_{xi}$  and  $\mu_{yi}$  are the respective means.

The above metric represents the correlation between the clean and reverberated envelopes and is used to compute the effective signal-to-noise ratio in each band, which is then limited and combined across bands after appropriate weighting to obtain a STI value. Payton and Shrestha (2013) do not recommend an appropriate window length  $N$ , but noted that longer windows showed greater agreements with long-term STI measures. In this study, the effects of varying  $N$  (from 0.3 s to a length corresponding to sentence length) were also analyzed. A length corresponding to 2.5 s (length of the sentences used) was used for the analysis presented here. This choice allowed the inclusion of modulation frequencies as low as 0.4 Hz in the analysis.

*Auditory model-based cepstral correlation.* Envelope fidelity was also measured using the cepstral correlation measure of Kates and Arehart (2014). This method was chosen to allow the simulation of the single reflection for impaired hearing. The cepstral correlation compares smoothed short-time spectra produced by a model of the auditory periphery in response to reverberated and clean speech. The auditory model for the HI models included elevated pure tone thresholds, a broadening of the auditory filters, a reduction both in the compressive effect of the outer hair cells and the two-tone suppression associated with hearing loss.

## 2.3 Results and Discussion

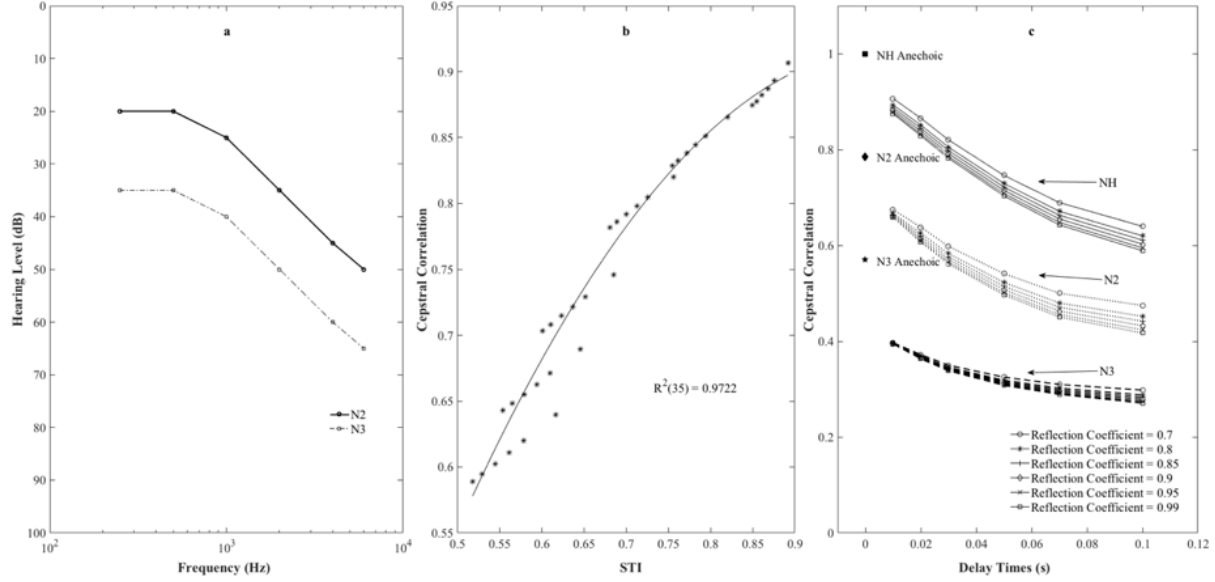


Figure 2.1: (a) N2 and N3 standard audiograms. (b) Relation between STI values and cepstral correlation coefficients for the NH model. (c) Cepstral correlation as a function of delay for several reflection strengths for models of NH (solid lines) and HI (N2 audiogram - dotted lines; N3 audiogram – dashed lines).

### 2.3.1 Effects of Delay Time and Strength of the Reflection on Predicted Intelligibility

The cepstral correlation was found to strongly correlate with the STI [ $R^2(35) = 0.9722$ ,  $p < 0.0001$ ] [see Fig. 2.1(b)]. This result is consistent with similar intelligibility models (Christiansen et al., 2010) and allows the application of the analyses to HI models through a peripheral auditory model. Figure 2.1(c) shows the cepstral correlation for increasing delay time and strength for the NH (solid lines) and HI (N2—dotted lines and N3—dashed lines) models. The cepstral correlation values for the anechoic sentences for each group are shown with the filled-in shapes (at the 0 delay). Overall, the longer the delay the greater the decrement in the correlation values. Stronger reflections led to lower predicted STI and cepstral correlation values. The cepstral correlation values follow

a trend observed by Nakajima and Ando (1991) in intelligibility values for the 0-degree azimuth presentation using a reflection that was equally strong as the direct sound, but these measures also consider the across band changes in the envelope modulations. The sentence length  $N$  we used in our analyses of STI provided the strongest correlation compared to shorter window lengths. These results support the choice of longer window lengths for calculating STI, especially when looking for small changes at the low modulation rates (1–16 Hz). It can be seen that the values are lower for the HI data and the overall patterns are similar to the NH result, which is similar to the results observed by Nábělek and Robinette (1978). Differences across conditions are reduced as a result of the interaction between the hearing loss and the characteristics of the reflection, a result which warrants future study.

### 2.3.2 Exploring changes in the modulation envelope

To analyze the changes in the speech envelope that caused the reductions in intelligibility, we analyzed the normalized cross-covariances between the clean and reverberated speech envelopes. For this purpose, we calculated the changes in envelope modulations within auditory frequency bands and the changes in the spectrotemporal modulations through the cepstral correlation coefficients across auditory bands (Kates and Arehart, 2015). To visualize these changes in the envelope modulations due to reverberation, two different sets of graphs were made to show the within- and across-band modulation spectra. Details of both these calculations are presented in Kates and Arehart (2015). For the graphs below, the normalized cross-correlations were averaged from a complete set of sentences (same as those used for intelligibility predictions above) to be more representative of changes seen in speech, rather than represent the modulations of specific stimuli. The effects of an increasing delay for reflection strength 0.85 are shown in Fig. 2.2. In these graphs, a value of 1 reflects perfect agreement (shown as black) and a value of 0 shows no agreement (shown as white) between the clean and reverberated envelopes.

The top row (Fig. 2.2, left to right) shows the correlations between the within-auditory band modulations associated with an increasing delay. For the smaller delay times (10 – 20 ms), the

reverberated speech accurately represents the low modulation rates (up to the 6 – 8Hz) in the auditory frequencies important for speech (80 - 8000Hz). The second row (Fig. 2.2) shows the changes in the spectral correlation coefficients across the auditory frequency bands for changes in the reflection delay. This measure also shows the same pattern with higher correlations for the lower modulation rates ( $< 12.5$  Hz) for the smaller delays (20–30 ms). Increasing the time of the reflection (50 through 100 ms) results in the reduction of the correlations at these modulation rates. All spectral ripple densities seem to show a higher correlation at the lower modulation rates for the shorter delay/higher predicted intelligibility conditions. These correlations are reduced with increasing delay times. The reflection strength affected the level of agreement between the reflected and clean envelopes; the greater the strength, the greater the loss of correlation between the clean and reverberated envelopes.

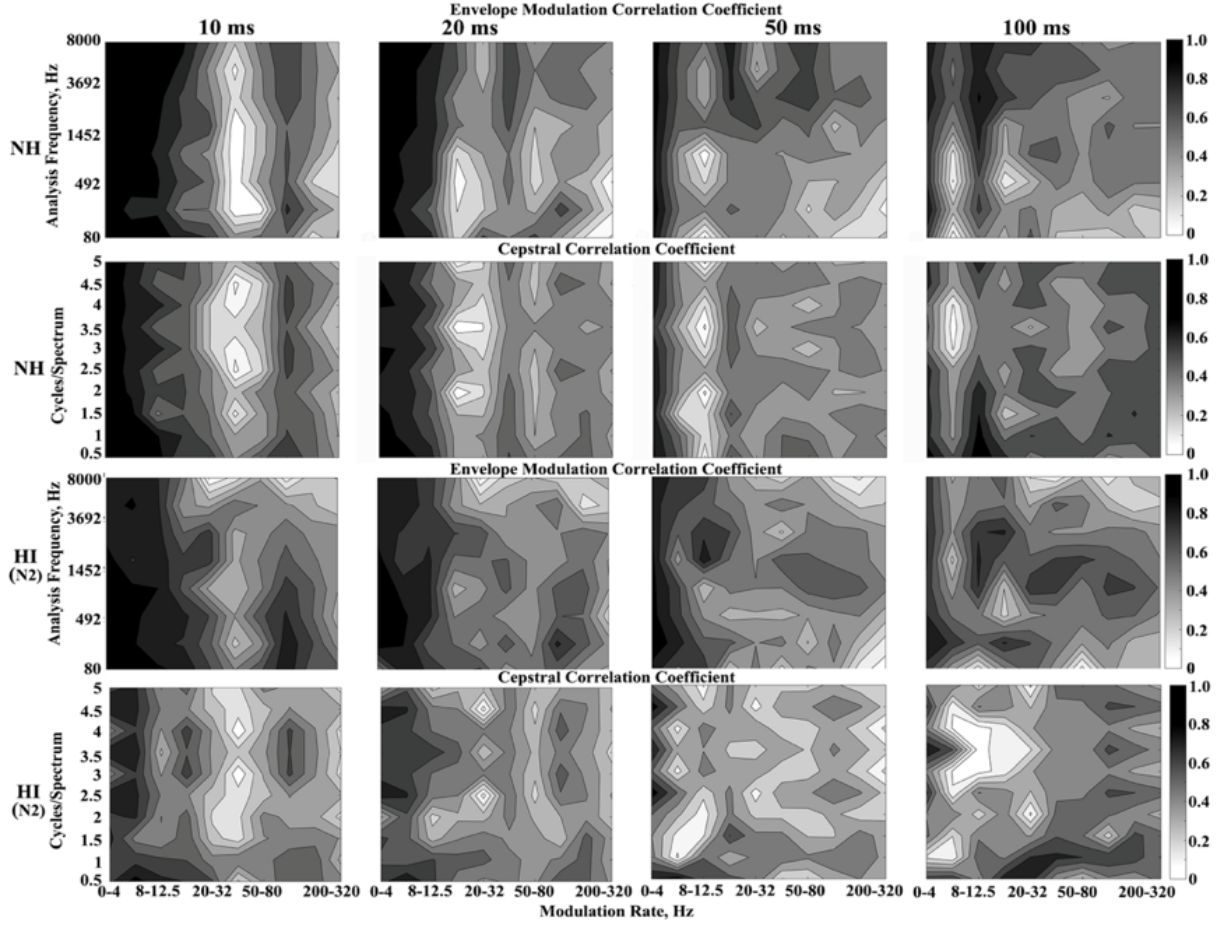


Figure 2.2: Contour plot of normalized envelope cross correlations averaged for 100 IEEE sentences comparing the envelope of the reverberated to clean speech. Top and Third Rows (L-R) show the coherence of envelope modulations at every auditory frequency per modulation frequency band for increasing delay times for NH and N2 respectively with a reflection strength of 0.85. Second and Fourth Rows (L-R) show the corresponding cepstral correlations at every auditory frequency per modulation band for increasing delay times for NH and N2. The spectral ripple used in the cepstral correlation calculation is indicated in cycles/spectrum, where the speech spectrum extends from 80 to 8000 Hz based on an auditory frequency spacing.

The results support the idea that early reflections do not impair speech intelligibility by accurately retaining the low modulation rates and auditory frequencies important for speech understanding. The importance of these low modulation rates ( $<12.5$  Hz) for maintaining speech intelligibility

has been well established in the past by several works (Kates and Arehart, 2015; Houtgast and Steeneken, 1985). The location of the high envelope agreements (and the notches with low agreement) and the level of agreement can be explained by a comb filter whose peaks change with the delay time and the strength of the reflection. Notably, for low delay times (10–30ms), this places the notches above the low modulation rates important for speech understanding. With larger delays (50, 70, and 100 ms) the notches start moving into the regions vital for intelligibility. This effect extends across all auditory frequency bands/ spectral ripple depths.

These analyses (the panels for 50 ms delay) are similar to Fig. 5 in Houtgast and Steeneken (1985), which shows the within-band MTF (for a single octave band centered at 2000 Hz) for a single reflection delayed by 50 ms. We also tested these tools using reverberation characterized as an impulse response modeled by an exponentially decayed white noise. We varied the reverberation time ( $T_{60}$ ) from 20 ms to 2 s. These simulations showed overall reduction in correlations but the same patterns of change to the modulation envelope as seen with the single reflection, with the shorter  $T_{60}$ 's showing higher correlations in the lower modulation rates and the longer reverberation times affecting the lower modulation rates. This result is consistent with that reported by Houtgast and Steeneken (1985), presented in their Fig. 5 for a reverberation time of 1.2 s. Thus, these methods which extend the analysis across all auditory frequency bands can provide a powerful analysis tool in understanding envelope changes with noise and reverberation.

The effects of reduced spectrotemporal resolution with sensorineural hearing loss (Fig. 2.2, rows 3–4) are visible when comparing the effects of the delay. The overall pattern of change is similar for both the NH and N2 audiograms. The reduced correlations in the higher auditory frequencies (above 2 kHz) can be explained by the increased thresholds at these frequencies for this audiogram. An effect similar to the NH simulations was observed in the cepstral correlation coefficient measures (Fig. 2.2, row 4) across the different auditory frequencies. The analysis clearly shows the same comb filter pattern with the location of the peaks being decided by the delay but the overall correlation values being lower for the HI condition. Also, the increased number of contours and their spread across modulation rates and auditory frequencies/ripples can be explained by the



reduced spectrotemporal resolution of the sensorineural hearing loss. The reduction in intelligibility [seen in Fig. 1(c)—dotted lines] can be attributed to the reduced audibility in the different bands combined with the reduced spectrotemporal resolution. Simulations using the N3 audiogram showed the same pattern as the N2 analysis presented above, but with a greater reduction in correlations and a larger number of contours associated with greater thresholds and reduced resolution.

## 2.4 Conclusions

This work analyzed the spectrotemporal changes in the signal envelope due to systematically varying the delay and strength of a single reflection in order to understand its effects on predicted speech intelligibility. An auditory model-based intelligibility metric was used to analyze these changes in two models of hearing loss. Both the reflection delay and strength affected this metric which was highly correlated with STI predictions in the NH model. Short time delays (10–30 ms) showed smaller losses in predicted intelligibility as predicted from the high correlations between the original and reverberated envelopes at the low modulation rates across all auditory frequencies. The improvement in predicted intelligibility can be attributed to the lack of interference from the reflection at these rates. Longer delay times showed a greater decrement in predicted intelligibility and lower correlations at these useful modulation rates across auditory frequencies. Overall, maintaining envelope fidelity across auditory frequency bands at these low modulation rates is strongly linked to intelligibility. While the overall patterns of envelope modulation effects were similar between models of NH and cochlear hearing loss, the reduced audibility and spectrotemporal resolution in the model of hearing loss caused a reduction in the access to useful envelope information. This result provides support for previous reports showing that while intelligibility is worse in HI listeners, the overall patterns of performance are similar between NH and HI listeners with a single reflection (Nakajima and Ando, 1991; Nábelek and Robinette, 1978). The cepstral correlation metric and the spectrotemporal modulation analysis method were found to effectively relate the changes in envelope modulations to patterns of intelligibility reported in the literature. Future work will investigate

the use of these tools for the analysis of more complex models of reverberation.

## Chapter 3

### Envelope and Reverberation in Real Rooms

#### 3.1 Introduction

##### 3.1.1 Reverberation and Intelligibility of Speech

We commonly communicate in rooms, stairwells, corridors, garages and other indoor environments where our ability to understand speech can be disrupted due to signal modifications caused by reverberation. Reverberation occurs, when in addition to the direct sound the listener receives from a speaker, there is a persistence of speech energy due to multiple reflections from surfaces within an enclosed space. A small amount of reverberation in quiet, though not problematic for listeners with normal hearing, might significantly reduce speech perception in quiet for listeners with cochlear hearing loss (Nábělek, 1994). Increased reverberation results in greater reductions in intelligibility (e.g. Lochner and Burger, 1964; Nábělek, 1994) in both groups of listeners.

Speech recognition scores in the same reverberant environment have been shown to decrease by as much as 20 percentage points as the hearing loss worsens from mild to moderately-severe (Nábělek, 1993). Also, the effect of background noise and reverberation together is more than the effect of each in isolation (George, Festen and Houtgast, 2008; Finitzo-Hieber and Tillman 1978; Duquesnoy and Plomp 1980; Wroblewski et al., 2012). Thus, typical everyday conditions of noise and reverberation that appear benign for normal-hearing listeners might pose greater challenges to speech understanding for listeners with hearing loss even if they are hearing aid users. Thus, it is important to mitigate the deleterious effects that reverberation has on speech intelligibility

for listeners with hearing loss. Low modulation-rate envelope cues are vital in the perception of anechoic(clean) speech (Smith et al., 2002; Shannon et al., 1995; Kates and Arehart, 2014). The alterations to envelope modulations have traditionally been used to quantify the perceptual effects of reverberation (Houtgast and Steneeken, 1985). The present study investigated the impact, for monaural listening, of modulation filtering designed to restore the envelope of reverberant speech. First, intelligibility decrements due to reverberation were linked to envelope degradations in both people with normal hearing and with mild-moderate sensorineural hearing loss through a quantification of the changes to the time-frequency modulations. Then, the extent of intelligibility benefit with ideal envelope restoration and the interaction of the different components of the envelope were systematically analyzed for both groups. Finally, envelope expansion techniques were tested to assess their effectiveness in restoring intelligibility.

### **3.1.2 Room Acoustics and Intelligibility**

The impact of reverberation on speech depends on the characteristics of the environment. The intensity and delay of the reflections, which depend on the geometry of the listening space and the energy absorption characteristics of the reflecting surfaces inside the space, result in different acoustic and perceptual changes to the original speech. The timing of the reflections, particularly, plays a significant role in the effect of reflections on perception (Bradley et al, 2003; Nakajima and Ando, 1991; Arweiler and Buchholtz, 2011; Muralimanohar et al., 2017). Reflections can be characterized as early or late, depending on the time interval between the onset of the original sound and the time of arrival of that reflection at a listener’s ear. The delay times characterized as early reflections usually range from a few milliseconds up to about 50 ms. Acoustically, these early reflections can be differentiated from the original speech both in terms of the intensity and the time of arrival at the listener’s ears. Perceptually, the early reflections are fused with the direct sound such that a listener hears a single sound source (the precedence effect; Litovsky et al., 1999), albeit with a different sound quality (Rennies et al., 2014; Warzybok et al., 2013; Arweiler and Buchholtz, 2011; Lochner and Burger, 1964). Sound reflections that arrive at the listeners’ ear more than 50

ms after the direct sound overlap more with other such reflections in time and have much smaller amplitudes. Consequently, these late reflections are perceived as a protraction of the energy decay and sound noise-like (Warzybok et al., 2013). In terms of speech intelligibility, early reflections enhance the direct sound and do not disrupt intelligibility while later reflections interfere with our ability to understand speech (Steeneken and Houtgast, 1980).

Early experiments probing speech perception in reverberation linked room acoustics to its impact on speech intelligibility (Lochner and Burger, 1964; Houtgast, Steeneken and Plomp, 1980; Nábělek, 1994). The amount of reverberation in a space can be quantified through the reverberation time ( $T_{60}$  or  $RT_{60}$ ).  $T_{60}$  is the amount of time it takes for the sound level to fall 60 dB below the original level after the source stops producing the sound (ANSI-S1.1,1960, R 1976; Boothroyd, 2004). Another measure used in the same context is the direct-to-reverberant ratio (DRR), which is a ratio of the energy in the direct sound to the energy in the reflections (ISO 3382). Other measures used to emphasize the importance of early reflections in determining speech intelligibility (ISO 3382) include the early decay time (EDT) and the clarity metric ( $C_{50}$ ). EDT is the time it takes the sound to drop by 10 dB and is extrapolated to the  $T_{60}$  value;  $C_{50}$  is an intelligibility metric computed by taking the ratio of the energy in the first 50 ms to the energy in the remainder of the room impulse response. Generally, rooms with values of  $T_{60}$  of up to 1.2 s are acceptable for speech perception for young adults with normal hearing. Though  $T_{60}$  is the most commonly used metric, other metrics like EDT,  $C_{50}$  etc. that use the early-to-late energy contrasts also provide useful information for understanding how reverberation affects intelligibility (Yang and Hodgson, 2006; Bradley et al., 2003).

While it is possible to predict the effect of the reverberation on speech intelligibility through these metrics, filters that remove the reverberation from the reverberant speech require a more detailed characterization of the listening space (Mourjopoulos and Hammond, 1983; Hirobayashi et al., 2000; Unoki et al., 2006). A complementary approach to de-reverberation involves modifying the speech signal based on an understanding (and then undoing) of the changes that reverberation causes to the perceptually salient parts of the speech signal. A crucial step in this strategy is the

identification of these relevant and usable cues that are vital to restoring intelligibility.

### 3.1.3 Speech Envelope and Intelligibility in Reverberation

Information in speech is encoded in its spectral and temporal patterns (Peterson and Barney, 1952; Shannon et al., 1995; Plomp, 1988). To understand how the encoding of these patterns supports speech perception, two methods are commonly used. The spectral approach maps acoustic patterns (amplitude-frequency characterizations) onto perceptual features (Peterson and Barney, 1952). The other approach uses the temporal patterns of the envelope to explain perception. This approach is based on the notion that speech is essentially a combination of slow information modulations imposed on carriers of different frequencies. The temporal envelope in each acoustic frequency band can be classified based on rate of modulation (Rosen, 1992). Low-rate modulations (2-50 Hz) are called the envelope. Periodicity refers to modulations in the 50 to 500 Hz rates. Temporal fine structure (TFS) refers to the rapid fluctuations of the envelope above 500 Hz. The envelope and TFS carry both segmental and supra-segmental information that are vital for speech understanding (Rosen, 1992; Shannon et al., 1995).

The speech envelope has been shown to be vital for speech understanding in listeners with normal hearing and with hearing loss (Houtgast, Steeneken and Plomp, 1980, 1985; Kates and Arehart, 2014; Taal et al., 2011; Shannon et al., 2004; Smith, Delgutte and Oxenham, 2002). Several researchers have theorized that envelope cues (also referred to as low-frequency modulations) are most important for speech perception (Smith, Delgutte and Oxenham, 2002). The importance of these cues has also been reinforced by studies showing that a reduction in low frequency envelope modulations ( $<20$  Hz) results in a lowering of intelligibility for normal-hearing listeners (Houtgast and Steeneken, 1985; Drullman, Festen and Plomp, 1994; Duquesnoy and Plomp, 1980). Shannon et al. (1995) showed that good speech perception (80% correct) for sentences could be achieved with only three bands of temporal envelope information in the absence of spectral detail. Other studies have reinforced that envelope cues of 1 to 20 Hz are most useful in speech perception in both listeners with normal-hearing and with hearing loss (Kanedera et al., 1999; Plomp, 1988).

Predictive metrics of speech intelligibility have also supported the importance of slow envelope modulations for speech intelligibility (Houtgast, Steeneken and Plomp, 1980; Kates and Arehart, 2015). Kates and Arehart (2015) used a metric-based computational model of the auditory system to show that the envelope modulations (particularly in the low-to-mid auditory frequencies) were important in predicting the intelligibility of hearing aid processed speech.

Envelope cues thus seem to provide information vital to speech intelligibility in reverberation-free “clean” speech. The addition of multiple scaled and delayed versions of the sound with the original sound results in temporal smearing of the speech signal. At the level of the utterance, reverberant energy fills in the pauses and gaps between speech features, smooths the temporal envelope, and reduces the peak-to-valley contrast and the rate of change of formants. At a segmental level, the energy from the same or previous segments affects the current portion, resulting in self and overlap masking. These changes result in the reduction of the temporal contrasts usually available to listeners to identify these features. Noordhoek and Drullman (1997) confirmed that large intelligibility degradations due to noise and reverberation could be explained by a reduction in modulation depth in multiple frequency bands. Muralimanohar, Kates and Arehart (2017) also showed that earlier reflections did not affect envelope fidelity while the later reflections resulted in a greater reduction of envelope fidelity due to larger impacts on the low modulation-rate ( $< 12$  Hz) information. Thus, envelope cues, especially low-rate cues, form an important target when trying to restore intelligibility through refurbishing distorted speech cues.

#### **3.1.4 Envelope Alteration and Speech Intelligibility in Reverberation**

Before we attempt to restore speech through restoration of low-rate envelope cues, it is important to first understand the detailed effects that reverberation has on the envelope. Changes to envelope cues have been incorporated into models that quantify the perceptual consequences of reverberation. One such model is the Speech Transmission Index (STI), which measures the impact of reverberation through a quantification of the changes in the envelope modulation depth caused by the room (Steeneken and Houtgast, 1980; Payton and Shrestha, 2013). The STI is calculated from

the modulation transfer function (MTF), which measures the peak-to-valley ratio of the modulation reproduced in each acoustic frequency band. This measure is then converted into a signal-to-noise ratio (SNR) in each band, with the computed SNRs averaged across bands to produce the final STI value. As typically implemented, the STI considers each acoustic frequency band separately, and does not consider the across-band patterns that might also influence speech understanding (Houtgast and Steeneken, 1985).

The STI does not directly allow the model to be extended to include cochlear hearing loss (Houtgast et al., 2002). The elevated thresholds and reduced spectrotemporal processing that are characteristic of cochlear hearing loss (Moore, 2007) may lead to an increased susceptibility to reverberation. While it is possible to add in the effects of the elevated thresholds into the STI model, cochlear hearing loss, termed “complex” hearing loss by the authors, is not described well by the STI model (Feston and Plomp, 2002). For people with such a hearing loss, the STI does not accurately predict the effects of intelligibility even when accounting for the increased threshold (Feston and Plomp, 2002; George et al., 2008). Thus, incorporating a model of hearing loss into a description of signal changes for reverberation analysis may allow us to more fully describe the effects of hearing loss on the perception of reverberant speech.

We recently showed that a more comprehensive description of the fidelity of spectrotemporal modulations of the speech envelope can be used to characterize the signal changes that cause a reduction of predicted intelligibility in a simple model of reverberation involving a single reflection (Muralimanohar et al., 2017). This characterization quantifies the signal change using correlations of the smoothed short-time spectra through a time-frequency modulation metric (cepstral correlation) and modulation analysis. The potential advantages of validating this cepstral correlation metric in real reverberation with listeners with hearing loss would be two-fold: 1. it combines information across the acoustic frequency range (80-8000 Hz) appropriate for hearing aids (Kates, 2014) and is thus more comprehensive, and 2. The metric includes a model of hearing loss (Kates, 2013) that has been previously validated in both listeners with normal hearing and with hearing loss in numerous conditions of noise, distortion and filtering. This addition should allow us to better



predict intelligibility in reverberant conditions for listeners with sensorineural hearing loss and help design signal processing to restore intelligibility in hearing aids.

### **3.1.5 Envelope Restoration to Restore Intelligibility**

Though not universal (Brennan et al., 2013), some individuals with hearing loss retain the ability to perceive and use envelope cues (Reed, Braida, and Zurek, 2009; Turner et al., 1995) and might even rely on temporal information when both spectral and temporal information are available (Hedrick and Younger, 2003). While age and hearing loss both affect monaural temporal perception, the ability to process envelope may be less impacted than the then ability to process TFS (Moore, 2016). In addition, listeners may be able to alter their weighting and use of different types of acoustic cues (Davies-Venn et al., 2014; Davies-Venn, Souza, Brennan, and Stecker, 2009; Souza, Jenstad, and Folino, 2005; Souza et al., 2015; Francis, Baldwin and Nusbaum, 2000). Thus, even older listeners with hearing loss might potentially gain from the restoration of envelope cues degraded by reverberation.

We therefore hypothesize that restoring the envelope to more closely match that of the original signal (increasing envelope fidelity) should result in improvements in intelligibility in reverberant conditions. However, even with envelope restoration, corrupted TFS still remains, and this source of distortion could lead to an upper bound on the potential benefits of the restored envelope. To assess the impact of preserving the envelope while corrupting the TFS, we used noise vocoder processing to create “envelope-only” versions of speech. To better understand the interaction of envelope and TFS, we tested the effects of restoring the envelope while using a noise vocoder or retaining the TFS of the reverberant speech.

### **3.1.6 Reversing Reverberation**

Studies that try to reverse the effects of reverberation attempt to restore speech through three main approaches. The first approach uses reverberation filtering to restore the envelope. The goal of the processing is to construct a model of the reverberation in the room and then use this

model to cancel the reverberation present in the signal. The effects of reverberation were reversed by recovering the clean envelope in multiple bands through envelope deconvolution using a known model of the reverberation. Mourjopoulos and Hammond (1983) used the original phase of the reverberant speech with an envelope recovered through envelope deconvolution. Hirobayashi et al. (2000) suggested a power envelope inverse filtering in subbands with an inverse filter that was derived based on an estimate of the reverberation characteristics. Unoki et al. (2006) used a blind estimator (no a priori information about the original room) to estimate the parameters used to define the inverse filter.

The second approach uses envelope modulation filtering to restore the modulation index. This method uses an estimate of the MTF reduction caused by the reverberation with different methods varying in their approach to the estimation of the filter. For example, Langhans and Strube (1982) propose a strategy in which the MTF of the reverberant signal is estimated from the room impulse response. Hirsch and Finster (1988) used an ad-hoc highpass modulation rate filter that restored the original envelope. Avendano and Hermansky (1996) used data-derived filters in each sub-band to filter the short-term power spectrum trajectories.

A third approach to restore the envelope is non-linear expansion. Noordhoek and Drullman (1997) report an 81 percent reduction of intelligibility in noise and reverb is caused by a reduction in modulation depth in multiple frequency bands. Expansion of the envelope would enhance the peaks and improve the modulation depth. Consideration of the STI also indicates that improving the modulation depth should improve intelligibility. Mourjopoulos (1982) reported an improvement in intelligibility when non-linear enhancement and expansion were performed in sub-bands. However, this model was not tested in people with hearing impairments. Lorenzi et al. (1999) suggest that expansion of the envelope in multiple bands should improve intelligibility in noise and reverberation through improving the modulation depth. This study, was, however, limited to testing the effectiveness of expansion under different SNRs and did not test intelligibility in the presence of reverberation.

It is challenging to draw general conclusions regarding the effectiveness of envelope restoration

because current studies vary in their methodologies. That is, researchers have utilized a variety of metrics for quantifying envelope fidelity, have implemented different amounts of MTF restoration, have used different signal processing strategies, have developed different ways of modeling reverberation, and sometimes have not actually measured speech intelligibility (Mourjopoulos and Hammond, 1983; Unoki et al., 2006; Hirobayashi et al., 2000; Hirsch and Finster, 1988; Langhans and Strube, 1982; Avendano and Hermansky, 1996). As such, there is a need for a controlled and consistent investigation of the benefits of envelope expansion that considers multiple processing strategies and that includes intelligibility tests in both people with normal hearing and in people with hearing loss. This study aims to fill this gap by systematically relating envelope modifications caused by expansion to measures of speech intelligibility.

We explored three types of multi-band expansion techniques in this experiment. The first method used equal expansion in all bands – uniform multi-band expansion (UMB). The effects of reverberation on the envelope are known to be frequency dependent (Houtgast and Steeneken, 1973; Sabine; 1964). So, in addition to uniform expansion, expansion that used independent amounts of expansion in each of the bands was also attempted – independent multi-band expansion (IMB). The amount of expansion in each band was decided based on the degree of mismatch between the reverberant and anechoic envelopes in that band. We specifically attempted two IMB expansion approaches; one based on matching the narrow band envelopes of a reverberant sentence to its original anechoic version and the other based on matching this envelope to an average value of expansion calculated based on an average calculated using a similar analysis of 240 sentences in that room.

### 3.1.7 Experimental Aims

The first aim of this study was to quantify the relationship between the envelope degradations of sentence-level stimuli subject to reverberation and the intelligibility of those sentence by listeners with and without hearing loss. The working hypothesis related to this aim was that reverberation affects intelligibility primarily through degradation of low-rate envelope cues. Because cepstral

correlation quantifies changes in time-frequency modulations of speech and incorporates a model of impaired hearing, we expected cepstral correlation to provide an improved characterization of the intelligibility of reverberant speech.

The second aim of this study was to determine the efficacy of ideal envelope restoration for listeners with normal hearing and with hearing loss. We hypothesized that ideal envelope restoration would improve intelligibility for both listeners with normal hearing and with hearing loss. While accurate reproduction of the envelope is important for speech intelligibility, the TFS also plays a role in the perception of speech. Thus, combining the restored envelope with the reverberant TFS should yield higher intelligibility than the envelope alone in a vocoder or with scrambled TFS.

The final aim of this study was to explore the benefit of envelope expansion to reverse the effects of reverberation. We hypothesized that both UMB and IMB envelope expansion processing schemes should restore intelligibility to the degree that they restore useful envelope modulations. The IMB methods were expected to result in higher intelligibility scores because they should provide band-appropriate expansion resulting in processing tailored to the reverberation and/or room unlike the UMB methods that apply the same changes to all speech and reverberation.

## 3.2 Methods

The following section describes the analysis tools and methods used to address the experimental questions. The two main experimental variables were listening environment (room) and type of processing. The experimental protocol involved four main phases. The first phase involved the generation of reverberant speech by filtering the unprocessed stimuli through recorded room impulse responses. Processing the reverberant speech to manipulate the envelope properties of speech formed the second phase. The third phase was the evaluation of the effectiveness of the envelope-restoration processing using objective intelligibility metrics. Finally, intelligibility was measured in both listeners with normal hearing and with hearing loss. The details of these four phases are presented in the following sections.

Table 3.1: Acoustic Environments

Label	Acoustic Environment	$T_{60}$ (s)	D.R.R (dB)	CTE (dB)
Room 1	Classroom	0.627	-3.41	4.65
Room 2	Parking Garage	1.354	-8.25	1.60
Room 3	Club Room	2.032	-8.26	0.64
Room 4	City Hall Balcony	3.002	-11.59	-4.17

### 3.2.1 Listening Environments and reverberant stimulus generation

Four listening environments were considered. The impulse response for Room 1 was measured in a classroom 393 in the Speech, Language, and Hearing Sciences (SLHS) department at University of Colorado Boulder using a custom setup that used custom MATLAB code, G.R.A.S. microphones and a USB-based National Instruments data acquisition setup (USB 4431, National Instruments). The impulse responses for rooms 2 through 4 were obtained from the Openair room impulse database (Murphy and Shelley, 2018). Listed in Table 3.1, the reverberation characteristics of these four rooms had a range of reverberation levels. Apart from reverberation time ( $T_{60}$ ) measured as an average of the values at 500 and 1000 Hz, the direct-to-reverberant energy (DRR), an early-to-late energy distribution characterization were calculated for all four spaces (CTE; ISO 3382). ISO 3382-1:2009 definitions of these parameters were used to characterize the rooms. These spaces were also chosen to represent common situations where listeners have difficulty understanding speech. In addition, the speaker and listener locations in each space were chosen to represent common listening situations. The  $T_{60}$  calculated here is based on the EDT.

Because reverberation builds up over time, the beginning of a test sentence may be affected less than later portions. To avoid this problem, each target sentence was preceded by a sentence chosen at random to provide reverberant masking at the beginning as well as during the sentence. A 200 ms gap was used between the flanker sentence and the test sentence. The test stimulus was obtained by excising the target sentence from the reverberant sentence pair. This core stimulus set was then processed for envelope restoration.

### 3.2.2 Envelope Manipulation

The envelopes from the reverberant sentences were extracted and processed to provide varying amounts of envelope restoration. The specific signal processing techniques that were implemented are described below.

#### 3.2.2.1 Envelope Extraction and Processing

Envelope manipulation required the extraction of the envelope and TFS of speech in multiple bands. The envelope was extracted using a nine-channel auditory filter bank consisting of linear phase FIR filters that spanned the frequency range from 80-8000 Hz. The lower and upper cutoff frequencies and bandwidths used in these filters are listed in Table 3.2.

Table 3.2: Description of the analysis filter bank

<b>Filter</b>	<b>Lower Cutoff (Hz)</b>	<b>Upper Cutoff (Hz)</b>	<b>Bandwidth (Hz)</b>
1	80	237	157
2	237	472	236
3	472	828	356
4	828	1259	431
5	1259	1866	607
6	1866	2720	854
7	2720	3923	1203
8	3923	5616	1693
9	5616	8000	2384

The Hilbert transform was used for both the envelope and TFS extraction. Specifically, the

envelope was extracted using the square root of the sum of squares of the real and imaginary parts of the Hilbert transform. The TFS information in each band was obtained by dividing the band-limited signal by the envelope calculated above. The envelope was then low-pass filtered using a 30 Hz cutoff frequency to provide the low-rate envelope modulation. Recursive filters (e.g. usually a 3rd to 5th - order Butterworth design) have often been used in the literature for lowpass envelope filtering (Souza, Hoover and Gallun, 2012; Apoux et al., 2004). However, the impulse responses of these filters can be negative and result in potential negative envelope values which violate the requirement that envelopes always be positive. To address this issue, a 512 point sliding window filter was designed and used to ensure that only positive envelope values were generated. This filtered envelope was then altered systematically to vary the amount of envelope restoration (described in following section). After restoration, the envelope-restored signal was then filtered a second time using the same bandpass filter as used for the analysis. This step was done to ensure that the modulation sidebands created by the envelope restoration process were removed. Finally, the level of the signal in the band was adjusted to match that of the original speech to remove any potential confounds caused by changes in the long-term spectra. Then, the signals from all the bands were summed to produce the broadband envelope-restored speech.

### 3.2.2.2 Processing Conditions

Different experimental processing conditions were used to answer our experimental aims. The first two aims of the experiment were to quantify the envelope-related changes due to reverberation that reduce intelligibility and to confirm the benefit of restoring low-rate envelope cues in restoring intelligibility in participants with and without hearing loss. To realize these two aims we used five conditions. We measured performance in clean, clean envelope-only vocoder processed and reverberant speech in four rooms. We also measured performance in each room after restoring the low-rate envelope cues ( $<30$  Hz) and using only the reverberant envelope cues processed through a noise-vocoder. These five conditions were used to characterize not only the baseline levels of monaural perception to sentence-level reverberant speech but also to probe the effects of restoring

the low-rate envelope cues to reverberant speech and the relative importance of the signal envelope and TFS. The final aim involves understanding the effect of expansion on the envelope of reverberant speech in restoring intelligibility. Six processing conditions were used to probe the effect for envelope expansion on improving intelligibility of reverberant speech. The details of these conditions are discussed in this section and are summarized in Table 3.3 at the end of this section.

The three baseline performance conditions consisted of one condition that used clean speech, another that contained only the clean envelope information and a final condition that contained reverberant speech in four rooms. In this study, the term clean refers to anechoic, noise-free speech information. The first condition (Label: Clean) was used to assess monaural speech perception through headphones using the experimental test sentences. The second condition (Label: Clean Voc – Clean Envelope Vcoded Speech) was used to assess listener perception with envelope-only information. The third set (Label: Rev) consisted of speech processed to replicate speech in the four rooms with increasing reverberation numbered one through four. The last two conditions contained different manipulations of the envelope and TFS from the reverberant speech to explore potential improvements in perception through envelope restoration. The first of these conditions included vocoded speech formed using only the envelope of reverberant speech (Label: Rev Voc – Reverberant Envelope Vcoded Speech). This condition was included to assess the decrement in perception arising from envelope-only distortions due to the reverberation. The final condition consisted of speech processed to restore low-rate envelope cues below 30 Hz (Label: Ideal Env Rest – Reverberant Speech with restored envelope).

The rest of the conditions contained envelope expanded reverberant speech (Collective Label: Exp Env – Expanded Envelope Speech) that either expanded the envelope using a pre-selected expansion factor to restore it or used the expansion factors calculated to give optimal matches between the processed and clean speech envelopes for expansion only processing. To summarize, we used clean, reverberant, vocoded and envelope-processed speech in our experiment to answer our questions. Further details of the signal processing used in the experimental manipulations of the envelope and TFS based information are provided in the following section.



### 3.2.2.3 Vcoded Processing

For narrow band signals, it is possible to extract the envelope and TFS from the speech envelope and process them separately to analyze their relative importance in perception (Swaminathan and Heinz, 2012).

The two vocoded conditions were used to determine the amount of information present in just the envelope given the information provided by the TFS in each condition. Using the reverberant envelope in a noise vocoder allowed the removal of TFS-related effects of reverberant speech. This condition (Rev Voc) provides insight into the amount of information present in the reverberant envelope. The second condition used the original envelope information in a noise vocoder (Clean Voc). This condition considered the amount of usable information present in the anechoic envelope for that person.

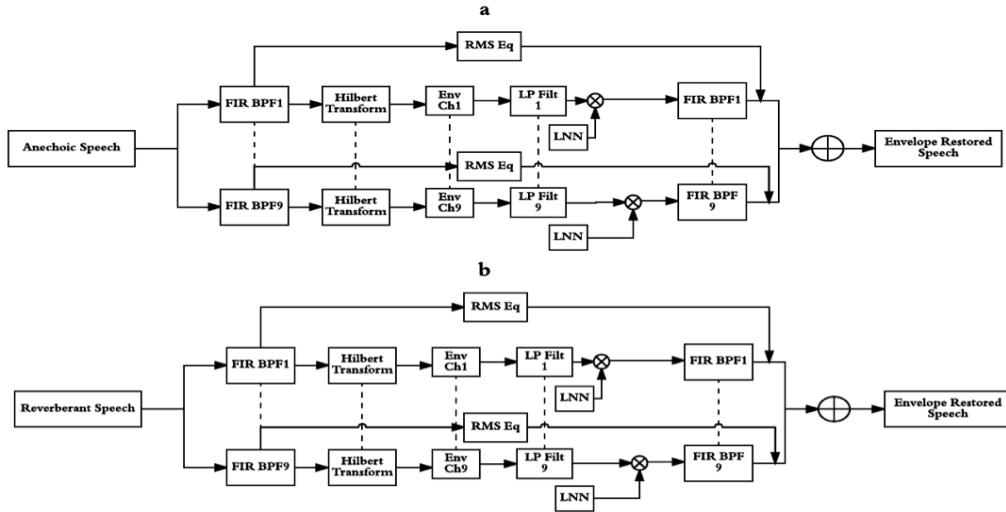


Figure 3.1: Vcoded conditions a. Original envelope vocoded (Clean Voc) condition and b. Reverberant envelope vocoder (Rev Voc) condition.

The signal processing used in the vocoder conditions is illustrated in Figure 3.1. The envelope parts of the information were extracted from clean anechoic speech (Figure 3.1a.) and reverberant speech (Figure 3.1b.) in the Rev Voc and Clean Voc conditions. For the TFS part of these signals,

low-noise noise (LNN) (Kohlrausch et al., 1997) was used in the place of traditional Gaussian-noise carriers. This choice was made to minimize the loss of intelligibility due to the random envelope fluctuations from the noise carrier (Kates, 2011). The envelopes were multiplied by the LNN carrier in each frequency band and were then level adjusted before they were added together to create the broadband restored speech.

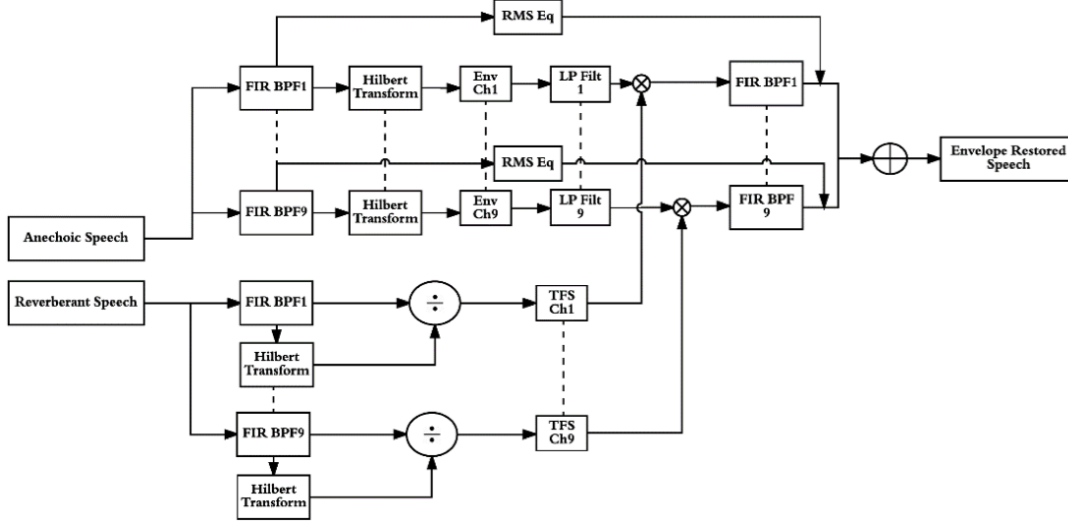


Figure 3.2: Full envelope restoration (Ideal Env Rest) condition.

### 3.2.2.4 Envelope Manipulations

In order to understand the maximum possible improvement with ideal envelope-only restoration, the clean envelope was combined with the TFS information of the reverberant speech (Label: Ideal Env Rest). The signal processing for the Ideal Env Rest condition is shown in Figure 3.2. The envelope part of the speech in this condition was provided in nine contiguous bands using the low pass filtered envelopes of anechoic speech. The TFS part of the speech was extracted from reverberant speech by dividing the reverberant speech in each band by the envelope extracted in that band using the Hilbert transform. After the second set of band pass filters, the restored-signal in each band was then level-adjusted to match the root mean square (RMS) of the original signal in that band.

The general signal processing used in the envelope expansion scheme is shown in Figure 3.3. The reverberant envelopes extracted in the nine continuous bands were low pass filtered. After low-pass filtering, these envelopes are expanded with either the same factor in all bands or a band-by-band weights specific to the stimulus (sentence and room) or the sentence average in the room. These envelopes were multiplied with the TFS of the original reverberant speech to create the signal in each band which was then filtered to limit the side bands and level adjusted to match the original anechoic RMS in that band. A final unprocessed reverberant condition (Rev) was also used as a baseline set.

### 3.2.3 Envelope Expansion

The effectiveness of both UMB and IMB envelope expansion techniques were investigated. In the UMB processing, the envelopes in the nine bands were raised to the same exponent similar to the method used in the first experiment in Apoux et al. (2004) as shown in Figure 3.3. This strategy allowed for the evaluation of restoration without any prior knowledge of the acoustics of the room (referred to as blind restoration). Two levels of expansion were tested in this method. The amount of expansion was controlled through the expansion coefficient  $k$ , illustrated in equation 1, where the output restored envelope  $env_o$ , would be obtained by raising the input reverberant envelope  $e_r$  to a constant power  $k$  referred to as the expansion factor.

$$env_o = env_r^k \quad (3.1)$$

A relatively small expansion,  $k = 1.2$  (Label:  $f_{lo}$ ), and a large expansion,  $k = 2$  (Label:  $f_{hi}$ ) were used (Lorenzi et al., 1999) in this condition.

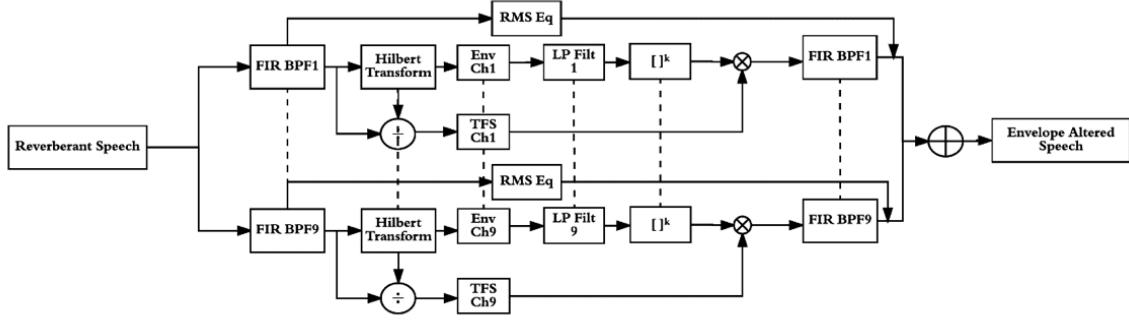


Figure 3.3: Envelope restoration through expansion

The subsequent conditions measured the possible advantage of using band and room-specific values of the expansion factor independently in each band (IMB). Different rooms have distinctive reverberation characteristics and cause consequent levels of reduction of the envelope modulations present in speech (Houtgast, Steeneken and Plomp, 1980; Houtgast and Steeneken, 1985). This effect can be seen by the decreasing values of STI for increasing  $T_{60}$  (Houtgast and Steeneken, 1985). Thus, the amount of expansion required to restore the envelope for a room might be specific to that space. The spectral information of a specific sentence may also dictate the amount of expansion required to make the sentence intelligible. Hence, the final pair of conditions attempted to restore the envelope through independent band-optimized expansion unique to that band.

The remaining methods of expansion therefore derived the value of  $k$  by minimizing the root mean squared error between the expanded reverberant and anechoic envelopes in each band obtained after the envelope rate low pass filtering step. In the first sentence-level optimized condition (Label:  $s_{mag}$ ), the optimum value of the expansion factor in a band represented the best possible restoration through expansion for that band given the anechoic and reverberant signals. It was obtained by determining the constants  $\alpha$  and  $\beta$ , that minimized the root mean square error between the clean signal  $env_o$  and the envelope restored signal  $env_p$ ,

$$env_p = (\alpha \times env_r)^\beta \quad (3.2)$$

where  $env_r$  = reverberant envelope and the RMS error in each acoustic frequency band ( $i^{th}$  band)

was calculated as,

$$\sqrt{\frac{1}{n} \sum_{i=1}^n (env_o - env_r)^2} \quad (3.3)$$

where the envelopes contained  $n$  samples.

In the second sentence-level optimized condition (Label:  $s_{\log\text{mag}}$ ), a similar minimization was performed on the root mean square difference between the clean and processed envelopes in decibels after thresholding the reverberant envelope  $env_r$  (in dB) to  $\pm 60$  dB around its root mean square level prevent the optimization from being driven by the extreme peaks or valleys.

$$env_p(\text{dB}) = (\alpha + \beta \times 20 \log_{10} env_r) \quad (3.4)$$

where  $env_r$  is reverberant envelope. The minimum RMS error criterion from Eq. 3.3 was used.

The next two levels of experimental variation ( $a_{\text{mag}}$ ,  $a_{\log\text{mag}}$ ) used expansion factors that were calculated from a subset of sentences used in the intelligibility task. The individual levels for each sentence were calculated as per the description in the sentence-level optimized conditions. The average band-level expansion factors,  $\alpha$  and  $\beta$ , were calculated by minimizing the RMS error over a set of 240 IEEE sentences (IEEE, 1969) produced by the 6 talkers. These average  $\alpha$  and  $\beta$  were determined in each frequency band for each room. The expansion was constrained to avoid issues related to recruitment or upward spread of masking (Apoux et al., 2000). An analysis of both values confirmed that the maximum amount of expansion provided was less than the expansion in the high expansion UMB condition.

### 3.2.4 Objective Signal Analysis

A measure of envelope fidelity and a graphical analysis technique were used to identify and quantify the changes to the envelope due to the reverberation and the effects of the envelope restoration. The details of these tools are presented below.

Table 3.3: Experimental conditions used in study. Subscript ‘cln’ indicates clean speech, and ‘voc’ indicates vocoder processing. ‘rev’ indicates reverberant speech. Exponent  $k$  – level of expansion applied to the envelope.

Condition			Label	Purpose
Baseline performance and envelope manipulation conditions (Clean and Reverb)	Clean Speech		Clean	Assess monaural perception of clean sentences through headphones and task performance
	$\text{Env}_{\text{cln}} \times \text{LNN}$		Clean Voc.	Maximum envelope-only information for listener
	Reverb		Rev	Unprocessed reverberant speech
	$\text{Env}_{\text{rev}} \times \text{LNN}$		Rev Voc	Usable information in reverberant envelope
	$\text{Env}_{\text{cln}} \times \text{TFS}_{\text{rev}}$		Ideal Env. Rest.	Maximum theorized improvement with ideal envelope restoration
Envelope Expansion Conditions	$\text{Env}^k \times \text{TFS}_{\text{rev}}$	UMB	Exp Env	Expanded envelope with reverberant TFS
			$k = k_{\text{flo}}$	Low fixed-expansion condition
			$k = k_{\text{fhi}}$	High fixed-expansion condition
		IMB	$k = k_{\text{smag}}$	Linear magnitude expansion sentence level
			$k = k_{\text{slogmag}}$	Log magnitude expansion sentence level
			$k = k_{\text{amag}}$	Linear magnitude expansion - 240 sentences
			$k = k_{\text{alogmag}}$	Log magnitude expansion - 240 sentences

### 3.2.5 Analysis of Envelope Modulations and Predicted Intelligibility

Cepstral correlation (Kates and Arehart, 2014) was used to measure the effects of envelope changes across auditory frequency bands. This metric compares the time-frequency envelope modulations in smoothed short-time spectra produced by a model of the auditory periphery in response to reverberant and clean signals. This metric has a high degree of correlation with the predicted intelligibility calculated using the envelope regression (ER) STI (Payton and Shrestha, 2013; Goldsworthy and Greenberg, 2004; Ludvigsen et al., 1990) for simulations of normal hearing. Models of hearing-loss also show similar patterns of change in predicted perceptual scores in models of hearing loss for a single reflection (Muralimanohar et al., 2017).

The envelope-regression implementation of the STI metric was chosen because it uses a speech signal as the excitation. The short-term ER calculation (Payton and Shrestha, 2013) used here is based on the band modulation metric in the  $i^{th}$  band ( $M_i$ ), calculated over a rectangular window of length  $N$ ,

$$M_i = \frac{\mu_{xi}}{\mu_{yi}} \cdot \frac{\frac{1}{N} \sum_{k=1}^N x_i(k)y_i(k) - \mu_{xi}\mu_{yi}}{\frac{1}{N} \sum_{k=1}^N x_k^2 - \mu_{xi}^2} \quad (3.5)$$

where  $x_i(k)$  and  $y_i(k)$  are the instantaneous intensity envelopes of the clean and reverberated signals at the  $k^{th}$  instant of time and  $\mu_{xi}$  and  $\mu_{yi}$  are the respective means. The analysis used in Muralimanohar et al. (2017) used sentence length windows to effectively characterize the changes in low modulation rate envelope cues due to reverberation (Muralimanohar, Kates and Arehart, 2017). A goal of this experiment is to validate the use of the cepstral correlation metric for assessing intelligibility in people with normal hearing and people with sensorineural hearing loss.

### 3.2.6 Listener Intelligibility Tests

#### 3.2.6.1 Participants

A total of 22 adults were recruited using University of Colorado Boulder’s Institutional Review Board (IRB)-approved procedures (Protocol# 15-0533) . Listeners were compensated for their

participation. Ten participants (range: 20-32 years; mean: 24.5 years) formed the normal hearing group (NH group) as defined by air conduction thresholds of 20 dB HL or better at octave frequencies 250 Hz through 8 kHz. Twelve participants (range: 53-77 years, mean: 65.3 years) with mild to moderate cochlear hearing loss formed the group with hearing loss (HL). Pure tone thresholds were measured in both ears and the better ear was picked for speech testing. Presented in Figure 3.5 below, participants had mild to moderately severe hearing loss; the four-frequency pure-tone threshold (4F-PTA: Average of thresholds at 0.5, 1, 2 and 4 kHz) ranged between 20-42 dB HL. Listeners lacked evidence of conductive pathology as indicated by normal tympanometry and the absence of an air bone gap exceeding 10 dB. There were no exclusions based on amplification history.

### **3.2.6.2 Processing of Stimuli**

Stimuli included 72 lists of ten sentences each from the IEEE corpus produced by six talkers (three male and three female) (IEEE, 1969). All the sentences were resampled to 22.05 kHz and processed through the impulse responses of the four rooms to create the baseline reverberant stimuli, which were then processed with the envelope restoration techniques.

### **3.2.6.3 Sound Presentation and Listener Testing**

Listeners were tested in a sound attenuating booth. The stimuli were delivered monaurally from the computer using the Tucker Davis TDT RX8 system (TDT, Alachua, FL) through headphones that have a free-field response (HD-25, Sennheiser Electronic GmbH & Co., Germany). Stimulus presentation was controlled and scored using a MATLAB<sup>TM</sup> (Mathworks, Natick, MA) based graphical interface.



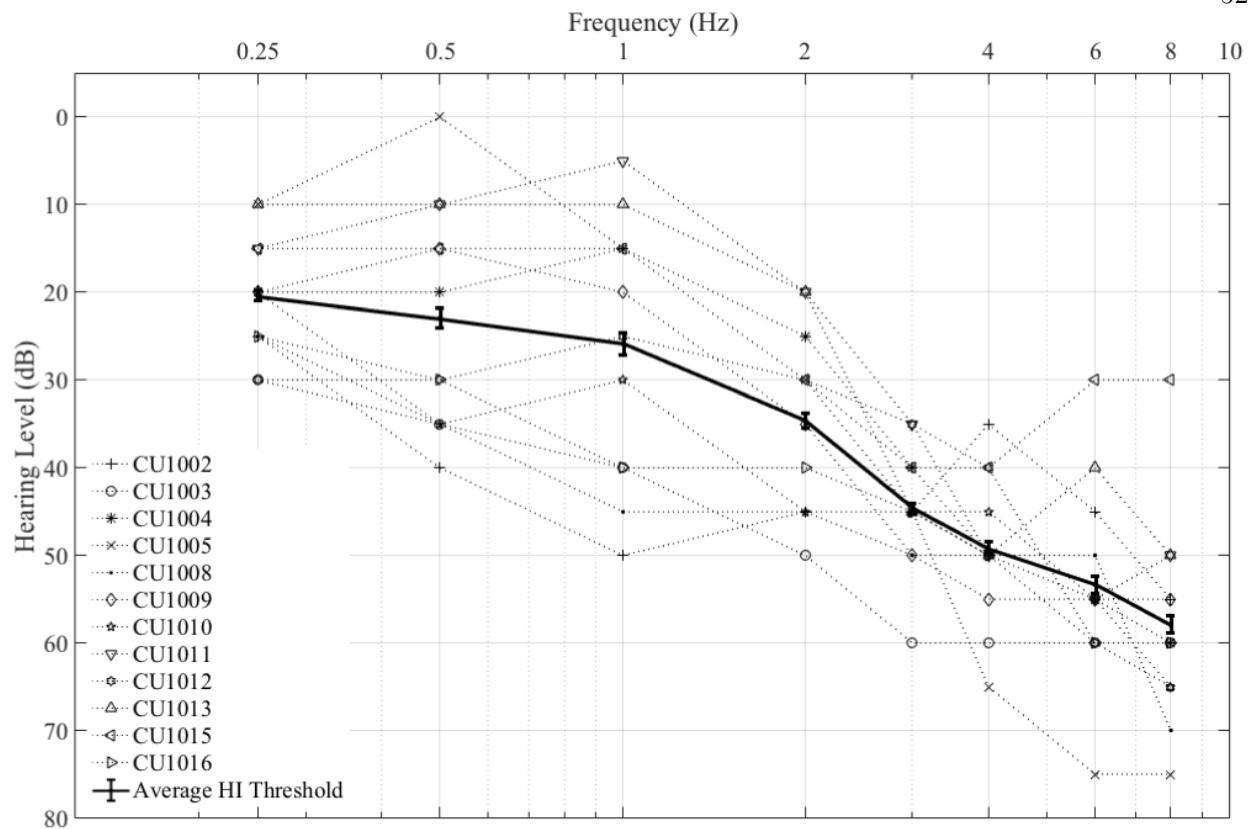


Figure 3.4: Auditory thresholds of Participants in group with hearing loss

### 3.2.7 Descriptive Measures of Hearing and Cognitive Abilities

Following the audiometric testing, all listeners passed a screening test for dementia (scores of 26 or better on the Montreal Cognitive Assessment (MoCA), Nasreddine, 2005; see Table 3.4 for individual scores.) Listeners were tested on several additional auditory tasks to characterize their supra-threshold processing. These measures included the binaural Quick Speech-in-Noise (Etymotic Research, 1993) (QuickSIN), gap detection (Brennan et al., 2013), modulation depth discrimination (Sabin et al., 2013) at 4 and 16 Hz and spectral ripple density detection (Won et al., 2007). Listener scores on these tasks are presented in Table 3.4.

Table 3.4: Hearing Profile of Participants including age, cognitive screener score, binaural speech-in-noise measure, audiometric thresholds and measures of temporal and frequency resolution ability. Subject numbers 1XXX are from the NH group and the 2XXX are from the group with hearing loss

Subject	Age	MoCA	QuickSIN	Gap Thresh.	Ripple Thresh.	Modulation Depth Discrim.	
	(Yrs)	(/30)	(dB)	(ms)	(ripples/octave)	4 Hz (dB)	16 Hz (dB)
1002	66	30	5	4.44	4.44	-14.13	-13.94
1003	77	27	3	3.99	2.73	-10.31	-12.9
1004	68	27	2	4.83	6.66	-7.48	-7.28
1005	73	27	9	6.92	5.69	-7.48	-6.65
1008	59	26	4	4.06	3.48	-10.69	-10.58
1009	57	26	1	5.12	5.22	-15.64	-9.7
1010	65	28	4	3.54	2.18	-9.9	-14.51
1011	65	29	9	7.46	3.6	-17.91	-7.17
1012	67	29	2.5	2.83	2.05	-7.28	-10.68
1013	66	29	7	6.17	4.83	-15.88	-16.89
1015	53	26	2	2.7	4.83	-18.19	-19.53
1016	68	30	5	3.27	4.16	-6.23	-6.17
2002	24	27	-1	3.39	12.71	-15.24	-11.43
2003	21	29	-1	2.64	9.66	-10.01	-11.57
2004	20	28	-1	2.5	10.8	-20.51	-13.51
2006	23	30	-1	3.25	10.71	-20.72	-22.82
2007	24	29	-1	2.96	1.27	-22.17	-15.81
2008	29	29	1.5	2.68	10.4	-12.32	-14.36
2009	32	29	-2	3.23	5.44	-12.91	-18.92
2010	23	30	2	3.1	7.93	-20.94	-17.08
2011	23	29	-1	2.52	6.97	-12.83	-11.99
2014	26	30	-3.5	3.44	8.7	-18.33	-22.37

### 3.2.8 Experimental Design

The two main experimental factors were the type of reverberant environment (room) and the type of envelope manipulation. The list of experimental conditions used to test each of the three experimental aims are listed separately based on the rationale for inclusion. As shown in Table 3.5, there were a total of 44 conditions (i.e., 4 rooms x 11 envelope manipulations) tested in this experiment.

### 3.2.9 Monaural Intelligibility Tests

The stimuli were presented to normal-hearing listeners at an average level of 70 dB SPL. To provide audibility, the 70 dB SPL stimuli were amplified for listeners with hearing loss based on the National Acoustics Laboratory-Revised (NAL-R) linear prescription (Byrne and Dillon, 1986).

For each condition, participants listened to 12 sentences (two sentences spoken by six talkers). Thus, each listener was tested on a total of 528 sentences (44 conditions x 12 sentences per condition). The participants were asked to repeat as much of the sentence that they heard. Participants received an initial practice sequence consisting of 30 IEEE sentences to provide familiarization with the task and range of stimuli they would hear in the actual experiment. This training phase was followed by the test portion where they listened to 528 sentences.

Intelligibility was scored using five keywords in every sentence. The word score (WS) in each condition was measured as the percentage of words correctly identified out of 60 possible words (12 sentences per condition, two per talker) in that condition. A sentence score (SS) was also calculated as a percentage of performance on 12 sentences in that condition. A sentence was scored as one only if all the keywords were correctly identified and zero otherwise.

## 3.3 Results

The first goal of the experiment was to characterize how envelope degradations due to reverberation affected intelligibility for listeners with normal hearing and with hearing loss. The second

Table 3.5: Experimental design, listing of levels based on rationale. The first set measured baseline anechoic performance to clean and envelope vocoded speech. (This condition was used compared to other conditions to estimate relative role of envelope). The second set measured performance with reverberant speech and envelope manipulations based on reverberant speech. The final set consisted reverberant and expanded envelope conditions used to assess performance of envelope expansion. A total of 11 unique conditions were used.

Aim	Room (4)	Processing Conditions (11)	Rationale
Link envelope changes due to reverberation to intelligibility	N/A	Clean Clean Voc	Baseline Perception (and Envelope Processing)
Determine if ideal envelope restoration improves intelligibility	Room 1 Room 2 Room 3 Room 4	Rev Ideal Env. Rest Rev Voc	Reverberant Envelope Processing Envelope Processing
Determine if envelope expansion improves intelligibility	Room 1 Room 2 Room 3 Room 4	Rev $f_{low}$ $f_{hi}$ $s_{mag}$ $s_{logmag}$ $a_{mag}$ $a_{logmag}$	Expansion Conditions

aim was to test the effectiveness of signal processing intended to restore the envelope to reverse intelligibility decrements. To establish this connection, the pattern of intelligibility scores from the various conditions are presented first, compared to each other to establish the relative impacts of different cues and are followed by results that link the envelope changes in the different conditions to the metrics.

While both word correct and sentence correct scores were obtained from testing, since they showed similar patterns only words correct scores are presented henceforth. The intelligibility scores were arcsine transformed (Studebaker, 1985) before any statistical analysis was performed. Significance was tested at the 95% level and suitably adjusted based on number of comparisons using a Bonferroni adjustment. The analyses were performed in SPSS version 24 (IBM Corp, Armonk, NY).

### **3.3.1 Effect of Reverberation and Envelope-Only Processing on Intelligibility**

Figure 3.5 shows average intelligibility scores for in each of the four reverberant rooms in addition to the scores for Clean and Clean Voc speech.

In addition to the accuracy scores in the clean and reverberant conditions, the scores in the clean envelope only condition are provided as a reference comparison to understand envelope-only perception. This comparison facilitates an examination of the effect of different levels of information (envelope and TFS) present in these conditions. A two-way mixed-model repeated-measures analysis of variance (RM-ANOVA) was completed with two groups for hearing status (normal hearing and hearing loss) and six levels of envelope information conditions (clean-anechoic, clean vocoded and the four rooms). The results of the analysis are presented in Table 3.6.

Table 3.6: Results of two-way mixed-model RM-ANOVA for reverberant and clean conditions.

Effects	df	F	p	Partial $\eta^2$	Observed Power
Envelope Information * Hearing Status	5,100	8.6	< 0.001	0.3	1
Envelope Information	5,100	263.7	< 0.001	0.93	1
Hearing Status	1,20	40.5	< 0.001	0.669	1

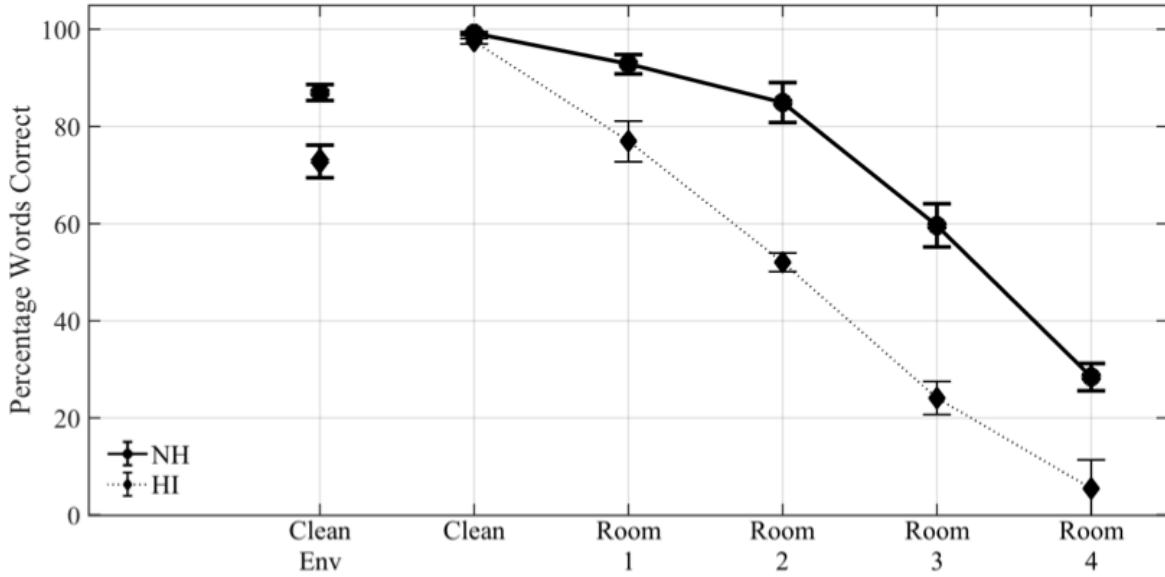


Figure 3.5: Percentage word correct scores in the Clean and 4 reverberant conditions. The performance in the Clean Vocoded condition are provided as a reference. The dashed lines indicate performance levels in the group with hearing loss while the solid lines indicate performance in the NH group. Error bars represent SEM.

The effect of varying envelope information on perception was dependent on hearing status [ $F(5,100) = 8.6$ ,  $p < 0.001$ , Partial  $\eta^2 = 0.3$ , observed power = 1]. On average across the different conditions, the scores for the hearing loss group were about 18% lower than the scores in the normal hearing group. Intelligibility scores in reduced information conditions for both groups were significantly less than the scores in the anechoic condition ( $p < 0.001$ , Bonferroni adjusted). In the Clean condition, scores in the hearing loss group were on average 2% lower than the scores from

the group with normal hearing [ $F(1,20) = 13.8$ ,  $p = 0.001$  Partial  $\eta^2 = 0.41$ , observed power = 0.942]. In the Clean Voc condition, the scores in the hearing loss group were about 14% lower than the scores in the normal hearing group [ $F(1,20) = 13.8$ ,  $p = 0.001$ , Partial  $\eta^2 = 0.409$ , observed power = 0.942]. In Room 1, the hearing loss group scores were about 16% lower than scores from the group with normal hearing [ $F(1,20) = 12.893$ ,  $p = 0.002$ , Partial  $\eta^2 = 0.392$ , observed power = 0.927]. In Room 2, the hearing loss group scores were about 33% lower than scores from the group with normal hearing [ $F(1,20) = 22.8$ ,  $p < 0.001$ , Partial  $\eta^2 = 0.533$ , observed power = 0.995]. In Room 3, the hearing loss group scores were about 36% lower than scores from the group with normal hearing [ $F(1,20) = 35.4$ ,  $p < 0.001$ , Partial  $\eta^2 = 0.639$ , observed power = 1]. In Room 4, the hearing loss group scores were about 23% lower than scores from the group with normal hearing [ $F(1,20) = 31.9$ ,  $p < 0.001$ , Partial  $\eta^2 = 0.614$ , observed power = 1].

On average across both groups, Clean Voc processing resulted in 26% lower scores than in the clean condition [ $p < 0.0005$ , Bonferroni Adjustment applied]. This difference was also significant in both groups [Hearing Loss: 25% lower,  $p < 0.001$ , Bonferroni Adjustment applied; Normal Hearing: 12% lower,  $p < 0.001$ , Bonferroni Adjustment applied].

The participants' performance in the Clean Voc condition was not significantly different from the performance in Room 1 [6% lower than Room 1,  $p = 0.730$ , Bonferroni adjustment applied] or Room 2 [11% higher than Room 2,  $p = 0.141$ , Bonferroni Adjustment applied]. For the NH group, significantly better performance was observed for Clean Voc compared to Rev in Rooms 3 and 4 [Room 3: 36% higher,  $p < 0.001$ , Bonferroni Adjustment applied; Room 4: 62% higher,  $p < 0.001$ , Bonferroni Adjustment applied]. In the group with hearing loss, the scores were better in the Env Voc condition compared to both Room 3 [46% higher,  $p < 0.001$ , Bonferroni adjustment applied] and Room 4 [70% higher,  $p < 0.001$ , Bonferroni adjustment applied].

Average scores fell by 19.8% in Room 1, 37% in Room 2, 62% in Room 3 and 88.7% in Room 4 compared to the clean scores [ $p < 0.001$ , Bonferroni adjustment applied]. Overall, the participants in the hearing loss group were more severely affected by the reverberation resulting in an average of 25% greater loss in intelligibility for the same amount of reverberation compared to the NH group

averaged across room. In the group with hearing loss, the scores fell by 21% in Room 1, 46% in Room 2, 73% in Room 3 and by 92% in Room 4 (significant,  $p < 0.001$ , Bonferroni adjustment applied). In the normal hearing group, the scores in Room 1 were lower by 6.3% (non-significant  $p = 0.051$ , after Bonferroni adjustment), in Room 2 lower by 14% lower, in Room 3 lower by 40% and in Room 4 lower by 71% (significant,  $p < 0.001$ , Bonferroni-adjusted alpha level) compared to the clean anechoic score.

### **3.3.2 Intelligibility in Reverberant Baseline Conditions**

Two additional reverberant conditions were used to assess the limits of benefit from the process of envelope restoration. The scores from these reverberant baseline conditions are illustrated in Figure 3.6. To analyze the performance in the baseline reverberant conditions, a three-way mixed ANOVA was used with the factors of room (4 rooms), processing (3 levels - Reverberant, Reverberant Vcoded and Ideal Envelope Restoration) and hearing status. The results of the analysis are presented in Table 3.7. Pairwise comparisons comparing rooms in each processing condition are provided in Table 3.8 and the effect of processing in each room are provided in Table 3.9.



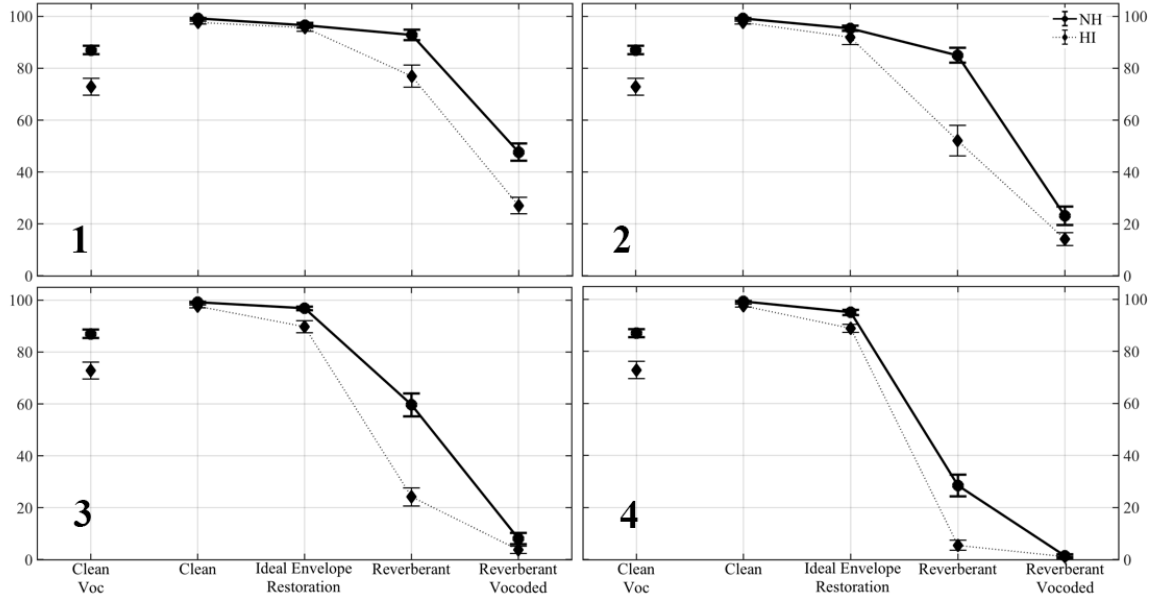


Figure 3.6: Performance in reverberant baseline conditions. The dashed lines indicate performance levels in the group with hearing loss while the solid lines indicate performance in the NH group. The room numbers are listed in the top right side of each panel. The performance in the Clean Voc condition and Clean (anechoic) speech is provided as a reference in the far left in each panel. Error bars represent SEM.

The effect of processing condition in a reverberant environment was different for the average listener with hearing loss than the average listener with normal hearing. In both the normal hearing and group with hearing loss, in each room, the effect of processing was different. The main effects of Room [ $F(3,60) = 376.5$ ,  $p < 0.001$ , Partial  $\eta^2 = 0.95$ ] and processing [ $F(2,40) = 13.62$ ,  $p < 0.001$ , Partial  $\eta^2 = 0.97$ ] were found to be significant. On average across all conditions and groups, the scores decreased with increasing reverberation with values falling from 74% in Room 1 to 34% in Room 4.

When averaged across all rooms and groups, the performance was the highest in the Ideal Restoration Condition (93%), lowest in the Reverberant Vocoded Condition (14%) and averaged about 53% in the Reverberant Condition.

The interaction between room and processing was found to be significant [ $F(6,120) = 4.149$ ,  $p$

= 0.001, Partial  $\eta^2 = 0.17$ ]. The simple two-way interactions between Room and Processing were significant (alpha level adjusted through Bonferroni adjustments) in both the normal hearing and group with hearing loss [NH:  $F(6,54)=39.62$ ,  $p < 0.001$ , Partial  $\eta^2 = 0.82$  and HI:  $F(6,66)=33.62$ ,  $p < 0.001$ , Partial  $\eta^2 = 0.75$ ]

### 3.3.2.1 Scores in the Reverberant Vcoded Condition

Scores in the Reverberant Vcoded condition for the group with hearing loss significantly decreased ( $p < 0.004$ , alpha level adjusted using Bonferroni corrections) when comparing Room 1 to 2 (14%) and Room 2 to 3 (16.5%) but due to floor effects were not significantly different between the Rooms 3 and 4 (5.6%;  $p = 0.453$ ). A similar effect was observed in the NH group where Room 1-2 and Room 2-3 decrements (23% and 19% respectively,  $p < 0.001$ , alpha level adjusted using Bonferroni corrections) were significant but the Room 3-4 difference was not (10.7%,  $p = 0.157$ ). The NH group scores reached floor levels in Room 4.

### 3.3.2.2 Scores in the Ideal Envelope Restoration Condition

For subjects in both the HL and NH groups, the pairwise comparisons between the scores in the different rooms were not statistically significant. The difference between Room 1 and 4 in the group with hearing loss was overall greater in each comparison and the difference between Rooms 1 and 4 approached significance (11%,  $p = 0.014$  based on alpha level adjusted to reflect Bonferroni corrections).

Table 3.7: Results of three-way mixed-model RM-ANOVA analyzing the baseline reverberant conditions. b : Bonferroni Corrections; c : Greenhouse-Geisser Adjustment applied

Effects	df	F	p	Partial eta sq	Est Power
Room * Proc * Hearing Status	6,120	4.149	0.001	0.17	0.972
Room * Hearing Status	3,60	0.946	0.424	0.045	0.246
Proc * Hearing Status	2,40	13.617	< 0.001	0.405	0.997
Room * Proc	6,120	68.016	< 0.001	0.773	1
Hearing Status	1,20	37.172	< 0.001	0.65	1
HL Room* Proc (Simple two-way Interactions)	6,66	33.623	< 0.001 <sup>b</sup>	0.753	1
NH Room*Proc (Simple two-way Interactions)	6,54	39.619	< 0.001 <sup>b</sup>	0.815	1
Reverb in HL	1.845,33	160.117	< 0.001 <sup>b,c</sup>	0.936	1
Reverb in NH	3,27	126.639	< 0.001 <sup>b</sup>	0.934	1
Rev Voc in HL	3,33	62.287	< 0.001 <sup>b</sup>	0.861	1
Rev Voc in NH	3,27	88.475	< 0.001 <sup>b</sup>	0.908	1
Ideal Env Rest in HL	3,33	4.283	0.012 <sup>b</sup>	0.28	0.819
Ideal Env Rest in NH	3,27	0.69	0.566 <sup>b</sup>	0.071	0.176
Room 1 - HL	2,22	139.517	< 0.001 <sup>b</sup>	0.927	1
Room 1 - NH	2,18	134.665	< 0.001 <sup>b</sup>	0.937	1
Room 2 - HL	2,22	92.96	< 0.001 <sup>b</sup>	0.894	1
Room 2 - NH	2,18	188.662	< 0.001 <sup>b</sup>	0.954	1
Room 3 - HL	2,22	189.304	< 0.001 <sup>b</sup>	0.945	1
Room 3 - NH	2,18	316.769	< 0.001 <sup>b</sup>	0.972	1
Room 4 - HL	2,22	510.942	< 0.001 <sup>b</sup>	0.979	1

Table 3.8: Pairwise Comparisons between scores in different rooms based on processing. Average percentage correct scores and significance values are provided for each condition. All comparisons were Bonferroni-adjusted.

Condition	Comparison	HL		NH	
		Mean	p	Mean	p
<b>Reverberation</b>	Room 1 – Room2	23	< 0.001	11	0.034
	Room 2 – Room 3	26	0.001	25	< 0.001
	Room 3 – Room 4	18	< 0.001	28	< 0.001
	Room 1 – Room2	14	0.004	23	< 0.001
<b>Reverberant Vcoded</b>	Room 2 – Room 3	17	< 0.001	19	< 0.001
	Room 3 – Room 4	6	0.453	11	0.157
	Room 1 – Room2	6	0.182	3	1
	Room 2 – Room 3	3	1	-2.5	1
<b>Ideal Envelope Restoration</b>	Room 3 – Room 4	2.5	1	3	1

Table 3.9: Pairwise Comparisons between processing conditions in each room based on processing. Average percentage correct scores and significance values are provided for each condition. All comparisons were Bonferroni-adjusted

Room	Comparison	HL		NH	
		Mean	p	Mean	p
<b>Room 1</b>	Rev – Ideal Env Rest	-24	< 0.001	-6	0.44
	Rev– Rev Voc	46	< 0.001	47	< 0.001
	Ideal Env Rest – Rev Voc	70	< 0.001	53	< 0.001
<b>Room 2</b>	Rev – Ideal Env Rest	-42	< 0.001	-14	0.01
	Rev – Rev Voc	36	0.001	59	< 0.001
	Ideal Env Rest – Rev Voc	78	< 0.001	73	< 0.001
<b>Room 3</b>	Rev – Ideal Env Rest	-65	< 0.001	-41	< 0.001
	Rev – Rev Voc	27	0.001	53	< 0.001
	Ideal Env Rest – Rev Voc	92	< 0.001	94	< 0.001
<b>Room 4</b>	Rev – Ideal Env Rest	-86	< 0.001	-67	< 0.001
	Rev – Rev Voc	9	0.091	35	< 0.001
	Ideal Env Rest – Rev Voc	94	< 0.001	98	< 0.001

### 3.3.3 Intelligibility in Envelope Expansion Conditions

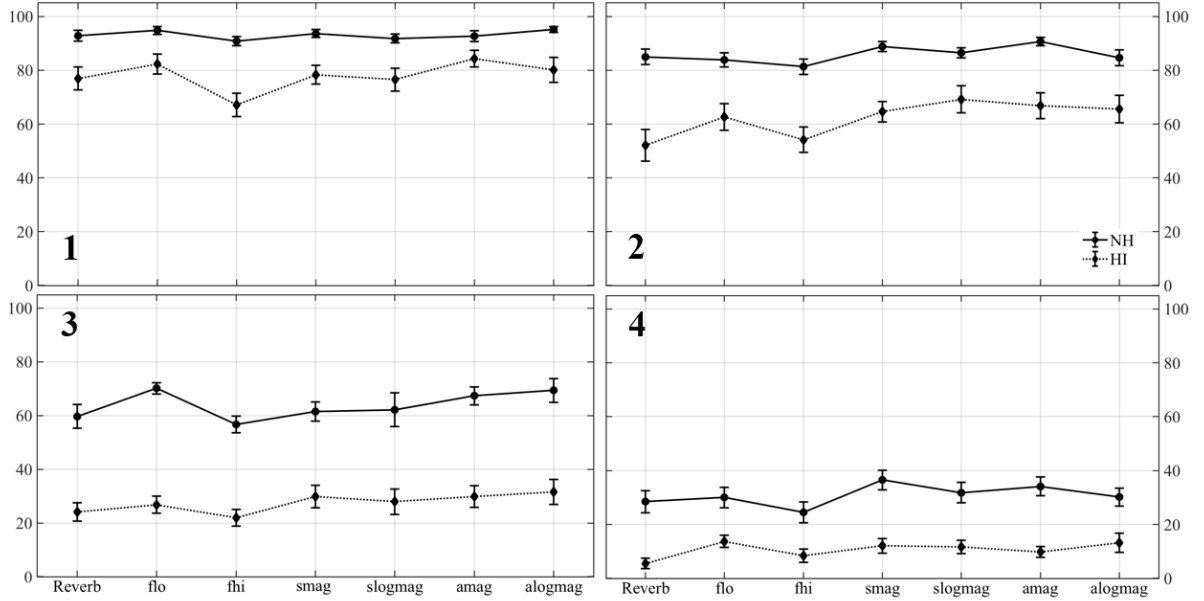


Figure 3.7: Performance in the envelope expansion conditions. The scores from the NH groups are indicated by the solid lines, while the dashed lines indicates the scores from the group with hearing loss. The room identifier is shown inside each panel. The rooms are ordered right-to-left and top-to-bottom in order of increasing reverberation. Dot-and-dashed lines represent the scores in reverberant condition for the group in that room. Error bars represent SEM.

The scores in unprocessed and expanded conditions for each room are presented in Figure 3.7. A three-way mixed model RM-ANOVA was performed to assess the effect of the three envelope expansion schemes in the two groups (normal hearing and hearing loss), four rooms and six levels of processing (two levels in each type of processing). The results of this analysis are presented in Table 3.10.

The overall effect of processing in a room was dependent on hearing status [ $F(18,360)=1.79$ ,  $p=0.026$ , Partial  $\eta^2 = 0.08$ , observed power = 0.96]. In the group with hearing loss the effect of processing was dependent on the room [ $F(18.198)=2.00$ ,  $p=0.011$  (Bonferroni adjustment applied),

Partial  $\eta^2 = 0.076$ , observed power = 0.92]. Table 3.11 provides the pairwise comparisons between the different rooms for each processing condition and Table 3.12 provides the pairwise comparisons of scores in each expansion condition to the score in that room. This effect was not significant in the NH group ( $p = 0.066$ , Bonferroni adjustment applied).

Table 3.10: Results of mixed model RM-ANOVA in the envelope expansion conditions. b : Bonferroni Corrections.

Effects	df	F	p	Partial $\eta^2$	Est Power
Room * Processing * Hearing Status	18,360	1.79	0.026	0.08	0.96
HL	1,20	38.49	< 0.001	0.66	1
HL Room * Processing (Simple two-way Interactions)	18,198	2	0.011 <sup>b</sup>	0.15	0.96
NH Room * Proc (Simple two-way Interactions)	18,162	1.37	0.152 <sup>b</sup>	0.13	0.86
Processing in Room 1 – HL	6,66	6.39	< 0.001 <sup>b</sup>	0.37	0.99
Processing in Room 2 – HL	6,66	3.95	0.002 <sup>b</sup>	0.26	0.96
Processing in Room 3 – HL	6,66	2.09	0.066 <sup>b</sup>	0.16	0.71
Processing in Room 4 – HL	6,66	3.44	0.005 <sup>b</sup>	0.24	0.4
Room – flo – HL	3,33	192.99	< 0.001 <sup>b</sup>	0.95	1
Room – fhi – HL	3,33	150.53	< 0.001 <sup>b</sup>	0.93	1
Room – smag – HL	3,33	164.56	< 0.001 <sup>b</sup>	0.94	1
Room – slogmag – HL	3,33	151.38	< 0.001 <sup>b</sup>	0.93	1
Room – amag – HL	3,33	330.16	< 0.001 <sup>b</sup>	0.968	1
Room – alogmag – HL	3,33	96.02	< 0.001 <sup>b</sup>	0.9	1

The effect of processing in each room (except in Room 3), for participants in the group with hearing loss, was significant across all processing schemes, however none of the expansion conditions resulted in scores significantly better than the score for unprocessed speech in that room (see Table 3.12). Except for the high fixed expansion ( $f_{hi}$ ) in Rooms 1 and 3, logarithmic sentence-level expansion processing in Room 1 resulted in loss of intelligibility (see Table 3.11). The increase in other conditions ranged between 1.2% ( $s_{mag}$  condition in Room 1) level to 16.3% ( $s_{logmag}$  in Room 2).

The effect of processing was dependent on the room. There was a significant decrease in scores with increasing reverberation in the scores the group with hearing loss.

Table 3.11: Pairwise Comparisons in the group with hearing loss. Mean values indicate change in scores from reverberant score in that room.

Room	Comparison	HL	
		Mean	p
<b>Room 1</b>	Rev – flo	6	1
	Rev – fhi	-10	0.38
	Rev – smag	1.2	1
	Rev – slogmag	-0.4	1
	Rev – amag	8	0.95
	Rev – alogmag	4	1
<b>Room 2</b>	Rev – flo	10	1
	Rev – fhi	2.4	1
	Rev – smag	11.4	1
	Rev – slogmag	16.3	0.43
	Rev – amag	13.9	0.89
	Rev – alogmag	12.7	1
<b>Room 3</b>	Rev – flo	3.2	1
	Rev – fhi	-2.025	1
	Rev – smag	5.6	1
	Rev – slogmag	3.4	1
	Rev – amag	5.8	1
	Rev – alogmag	7.1	1
<b>Room 4</b>	Rev – flo	12.6	0.18
	Rev – fhi	5	1
	Rev – smag	10	0.28
	Rev – slogmag	9.2	0.53
	Rev – amag	7.5	0.53
	Rev – alogmag	11.2	0.35

Table 3.12: Pairwise comparisons of the effect of expansion processing between rooms with increasing reverberation.

Room	Comparison	HL	
		Mean	p
<b>flo</b>	Room 1 – Room 2	20.0	< 0.001
	Room 2 – Room 3	32.6	< 0.001
	Room 3 – Room 4	14.4	< 0.001
<b>fhi</b>	Room 1 – Room 2	11.7	< 0.001
	Room 2 – Room 3	30.2	< 0.001
	Room 3 – Room 4	16.9	< 0.001
<b>smag</b>	Room 1 – Room 2	13.7	< 0.001
	Room 2 – Room 3	31.7	0.001
	Room 3 – Room 4	19.6	< 0.001
<b>slogmag</b>	Room 1 – Room 2	7.2	< 0.289
	Room 2 – Room 3	38.7	< 0.001
	Room 3 – Room 4	18.1	0.015
<b>amag</b>	Room 1 – Room 2	17.9	< 0.001
	Room 2 – Room 3	34	< 0.001
	Room 3 – Room 4	22.1	< 0.001
<b>alogmag</b>	Room 1 – Room 2	15.1	< 0.013
	Room 2 – Room 3	31.5	< 0.001
	Room 3 – Room 4	20.0	0.003



### 3.3.4 Cue-based Performance and Benefits

To understand the effect of change in the amount of the different cues on perception we analyzed the differences between performance in the different baseline and envelope processed conditions. We also analyzed the benefit of envelope restoration to understand its effectiveness in restoring the envelope.

#### 3.3.4.1 Intelligibility and Envelope Cues

Figure 3.8a. shows the reduction in intelligibility due to reverberation in the four rooms (compared to the scores in clean anechoic speech). Intelligibility decreased with increasing amount of reverberation. That is, for the NH group, intelligibility scores were notable worse as the room reverberation increased from Room 1 to Room 4. A similar trend was observed in the group with hearing loss, but with much greater disruption of intelligibility.

To establish the extent to which these perceptual challenges arise from envelope-only degradations, we compared the scores in the Clean Voc condition to those in the Rev Voc condition (Figure 3.8.b). The pattern of intelligibility losses is like the pattern seen in Figure 3.8.a due to reverberation; increasing reverberation results in a greater reduction of usable cues. In the first comparison (Figure 3.8.b), even in the room with least reverberation-related signal distortions, the scores fall by about 40% for participants in the NH group. The scores decreased monotonically between the last two rooms for participants with normal hearing.

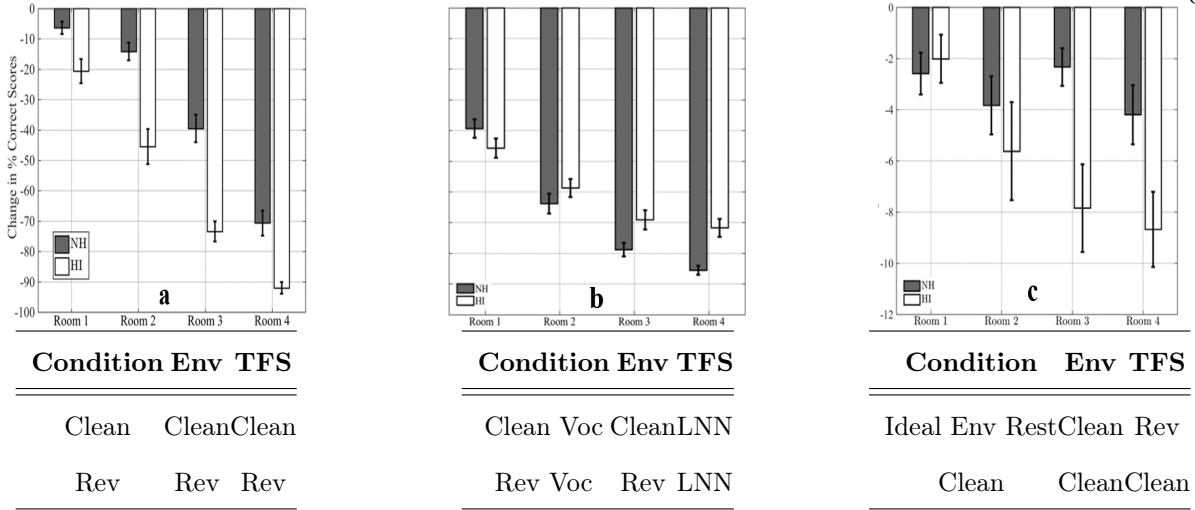


Figure 3.8: The values in each condition are calculated as the difference between scores in the conditions in Row 1 and Row 2 in each column. a: Effect of reverberation on speech intelligibility; b: Effect of envelope-only effects of reverberation; c: Relative changes in intelligibility due to reverberant TFS in the context of clean envelope. Y-axis in plot a and b are the same, while y-axis in plot c indicates a much narrower range.

We also examined the extent of interference arising from the reverberation-related degradations to the TFS by comparing the performance in the clean versus the ideal envelope restoration condition (Figure 3.8.c). The axes displaying accuracy scores in Figure 3.8.c covers a much smaller range than the other plots in the same figure. The effects of reverberant TFS was minimal compared the amount of change from envelope cues. The amount of decrement due to the TFS is very similar in both groups. The scale of the effects and the greater performance of the NH group in all the rooms in any given condition also indicate that both groups suffer similar disruptions.

### 3.3.4.2 Ideal Envelope Restoration and the role of TFS

The results of the comparisons made to illustrate the effect of ideal envelope restoration are shown in Figure 3.9. 3.9a shows the comparison between the performance with the envelope restored reverberant speech and with baseline reverberation while the graph on the right indicates the difference between the scores in the Ideal Env Rest condition to the Clean Voc condition. Ideal

envelope restoration significantly improved performance in both groups in all rooms except in Room 1 for participants in the NH group. In all other rooms the participants gained from between 10% (NH in room 2) to about 85% (HL in room 4). To understand the importance of retaining the TFS, we analyzed the differences between the Ideal Env Restored speech and Clean Env Voc conditions (Figure 3.9 b). This comparison showed that both groups gained a small but significant advantage from retaining the TFS. The level of this benefit was also pretty consistent irrespective of the level of reverberation.

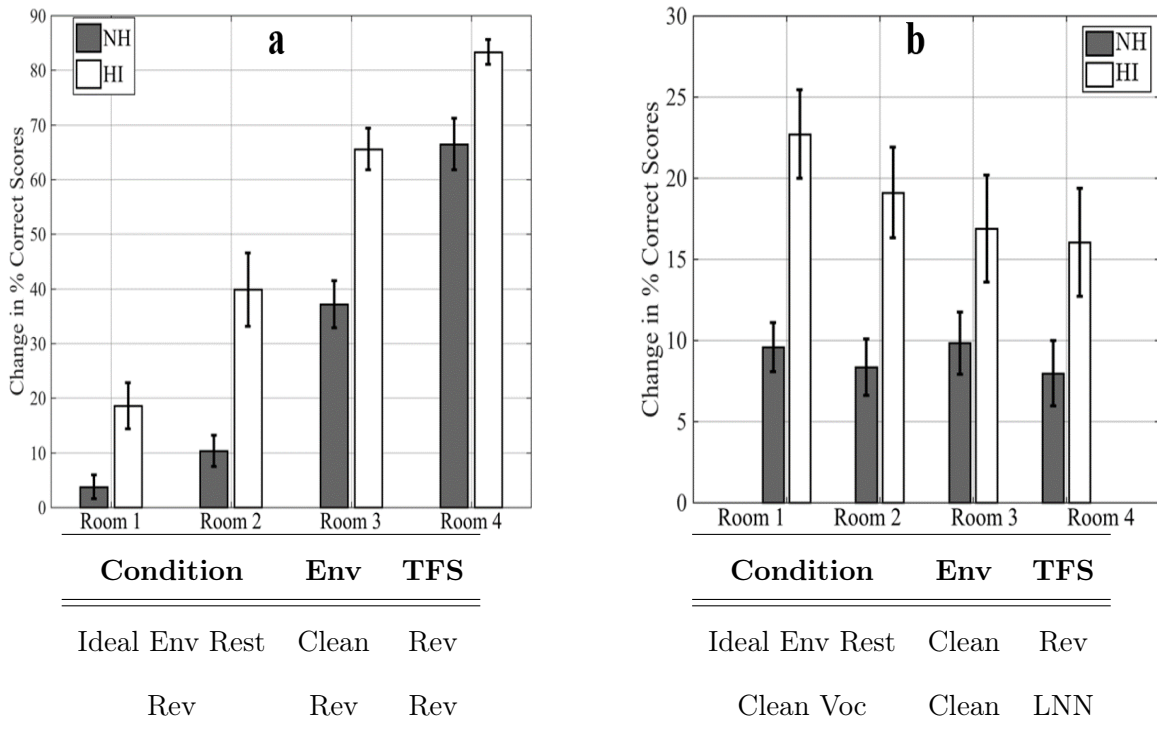


Figure 3.9: The values in each condition are calculated as the difference between scores in the conditions in Row 1 and Row 2 in each column. b: Effect of Ideal Envelope Restoration. a: Role of reverberant TFS in restored intelligibility.

### 3.3.4.3 Improvement in Intelligibility from Expansion

To analyze the benefit from expansion, we examined the difference between performance with listening to envelope expanded speech to the performance with reverberant speech in that

room. Though none of the processing resulted in significant benefit compared to the perception in reverberation, we wanted to explore the pattern of benefit. The size of the benefits in the different conditions for each room are shown in Figure 3.10. All the methods except for the high amount of UMB expansion produced some improvements in three out of four rooms. The largest gains from the processing was observed in Room 2. Also, though not significant, listeners in the group with hearing loss seemed to benefit slightly more from the processing (except with  $f_{hi}$  processing). Some of the expansion schemes produced up to 23% benefit ( $a_{mag}$ ) over reverberation for the group with hearing loss. The UMB schemes seemed to produce more benefit in most rooms.

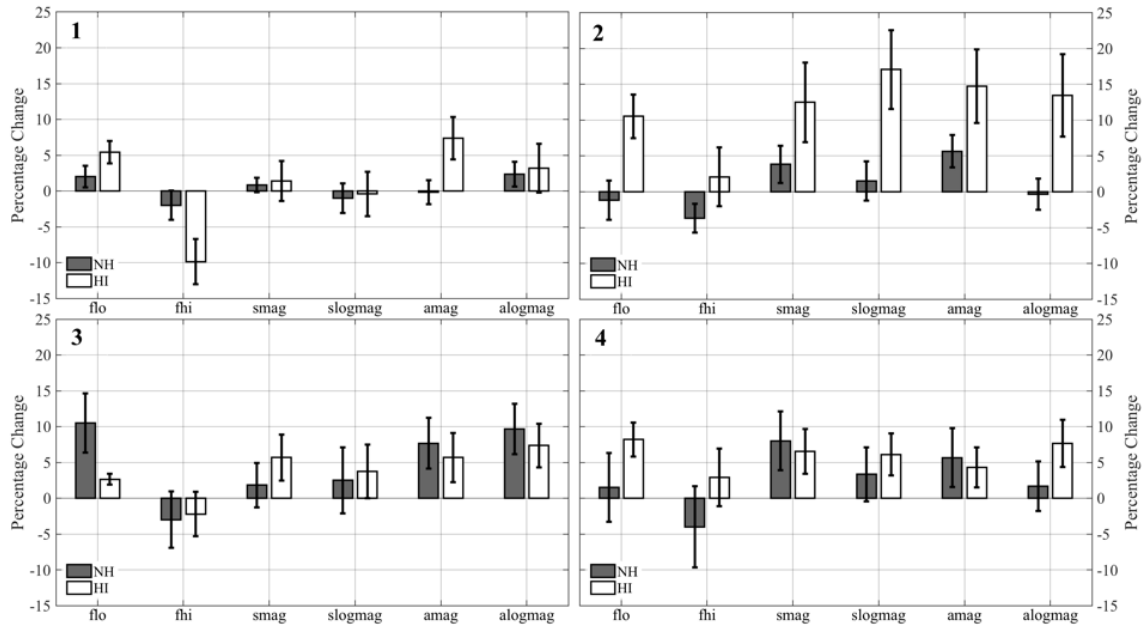


Figure 3.10: Benefit from envelope expansion. Values were calculated as a difference score of scores in a particular condition and the reverberant speech in that room. Room numbers are shown in the individual panels. Error bars are SEM.

### 3.3.5 Metric Predictions of Word Scores

Short-term STI and cepstral correlation values were calculated from the original and reverberant sounds. STI predictions for the group with hearing loss were based on the unamplified stimuli

as per normal convention (Payton and Shrestha, 2013). In contrast, similar to Kates and Arehart (2014), the current cepstral correlation predictions included stimuli that received frequency-specific amplification that were then processed through a model of the auditory periphery that considers cochlear hearing loss (Kates, 2013). In fitting the cepstral correlation predictions to intelligibility scores, an s-shaped logistic function (Kates and Arehart, 2014) of the form,

$$F(x) = \frac{a}{1 + e^{-(ax+b)}} \quad (3.6)$$

was used in fitting the predictive metric to the proportion word correct scores. These functions were fit both on a group level and an individual subject level.

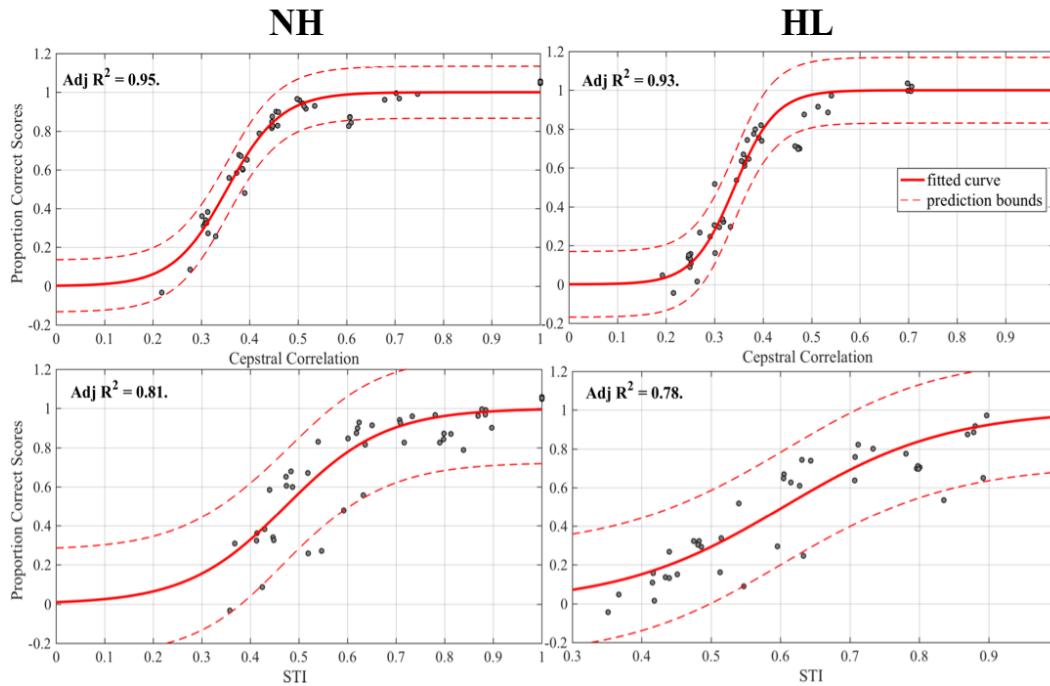


Figure 3.11: Proportion word correct scores as predicted by the cepstral correlation (top row) and STI (bottom row). The left column contains the fits for the NH data and the right column contains the fits for the HL data. The 95% confidence intervals are indicated by the bounding curves.

Traditionally, third-order (Steeneken and Houtgast, 1980) or second-order polynomials (Payton and Braida, 1999) are used in fitting STI data to behavioral intelligibility scores. The log-sig

Table 3.13: Coefficient of determination ( $R^2$ ) values for individual STI and cepstral correlation values fit to proportion of words correct. The larger values for each comparison is in bold

HI Subject	STI	Cepstral Correlation	NH Subject	STI	Cepstral Correlation
1002	0.8868	<b>0.9123</b>	2002	0.8053	<b>0.8805</b>
1003	<b>0.7960</b>	<b>0.7960</b>	2003	0.8266	<b>0.8999</b>
1004	0.8015	<b>0.8435</b>	2004	0.6792	<b>0.8956</b>
1005	<b>0.8572</b>	0.8143	2006	0.8178	<b>0.8502</b>
1008	<b>0.8396</b>	0.8320	2007	<b>0.9123</b>	0.8948
1009	0.8639	<b>0.8666</b>	2008	0.8531	<b>0.9188</b>
1010	0.9121	<b>0.9457</b>	2009	0.9028	<b>0.9258</b>
1011	0.7748	<b>0.9117</b>	2010	0.7248	<b>0.8735</b>
1012	0.8125	<b>0.9047</b>	2011	0.8529	<b>0.9208</b>
1013	0.7937	<b>0.7981</b>	2014	0.7174	<b>0.9107</b>
1015	0.7201	<b>0.8823</b>			
1016	0.7003	<b>0.8332</b>			

transformation was used here as it offers a realistic limit and near-linear change in the response regions having approximately 50 percent correct scores. The fits for the average predicted metrics and proportion word score were obtained by averaging across participants in each condition and room for each group. The results of the fitting are presented in Figure 3.11 above. The top row shows the results of fitting cepstral correlation to the data in the NH group (left) and the group with hearing loss (right) while the bottom row shows the results of fitting the STI scores for the same groups. The dots represent the individual data points, the solid line the model fit. The dotted line represents the 95% confidence bounds. The goodness-of-fit value using the Pearson correlation coefficients are displayed in each plot. High correlation coefficients were observed for the cepstral correlation fits, such that the  $R^2$  was 0.95 for the NH group and 0.93 for the group with hearing loss. The correlation coefficients for the STI fits were 0.81 for the NH group and 0.78 for the group with hearing loss.

One of main benefits offered by a metric like cepstral correlation is the ability to extend the application of the metric to models of hearing loss. To study the applicability of the model to individual participants, the average metrics were fit to the averaged proportion correct scores

in the different conditions in the group with hearing loss. The individually fit data are shown in Figure 3.12. The first three rows present the STI fits and the last three rows present the fits of the cepstral correlation data. The subject identification is present in the top-left corner in each subplot. Consistent with the average data, the fits for the individual listeners yield higher  $R^2$  values for the cepstral correlation compared to the STI. The individual data had lower residuals and better fit in the middle-of-the range metric and word score values.

The  $R^2$  for the STI and cepstral correlation fit for each subject are shown in Table 3.13. The cepstral correlation showed higher values of this parameter for ten out of the twelve HL participants and nine out of the ten NH participants. The individual fits for both the NH and HI groups are shown in Figure 3.12 and 3.13 below.

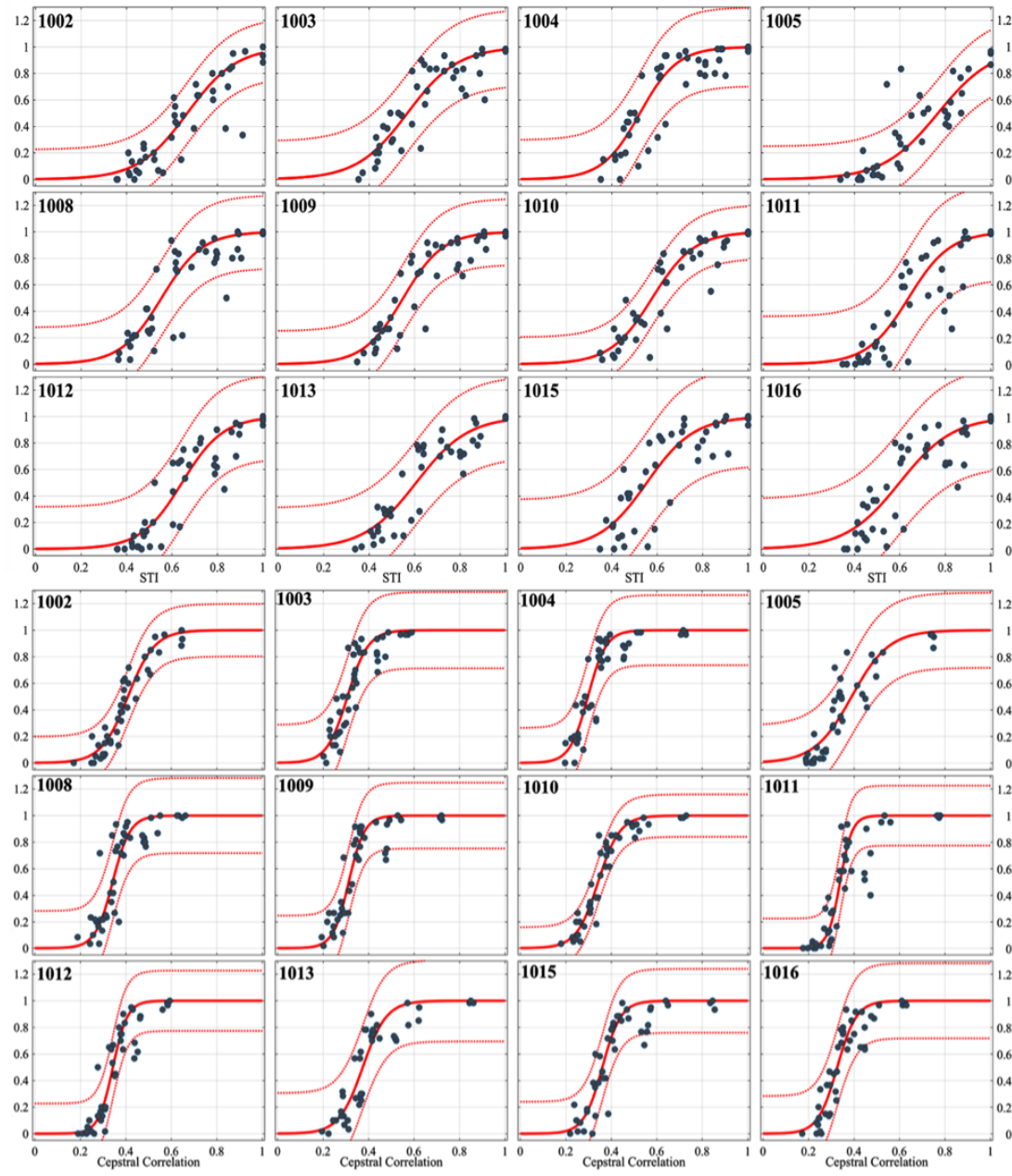


Figure 3.12: Individual best fits for calculated metrics to proportion word correct scores in group with hearing loss. Top three rows indicate the STI values fitted to the proportion word correct while the bottom three rows indicate cepstral correlation values fitted to the proportion correct scores.



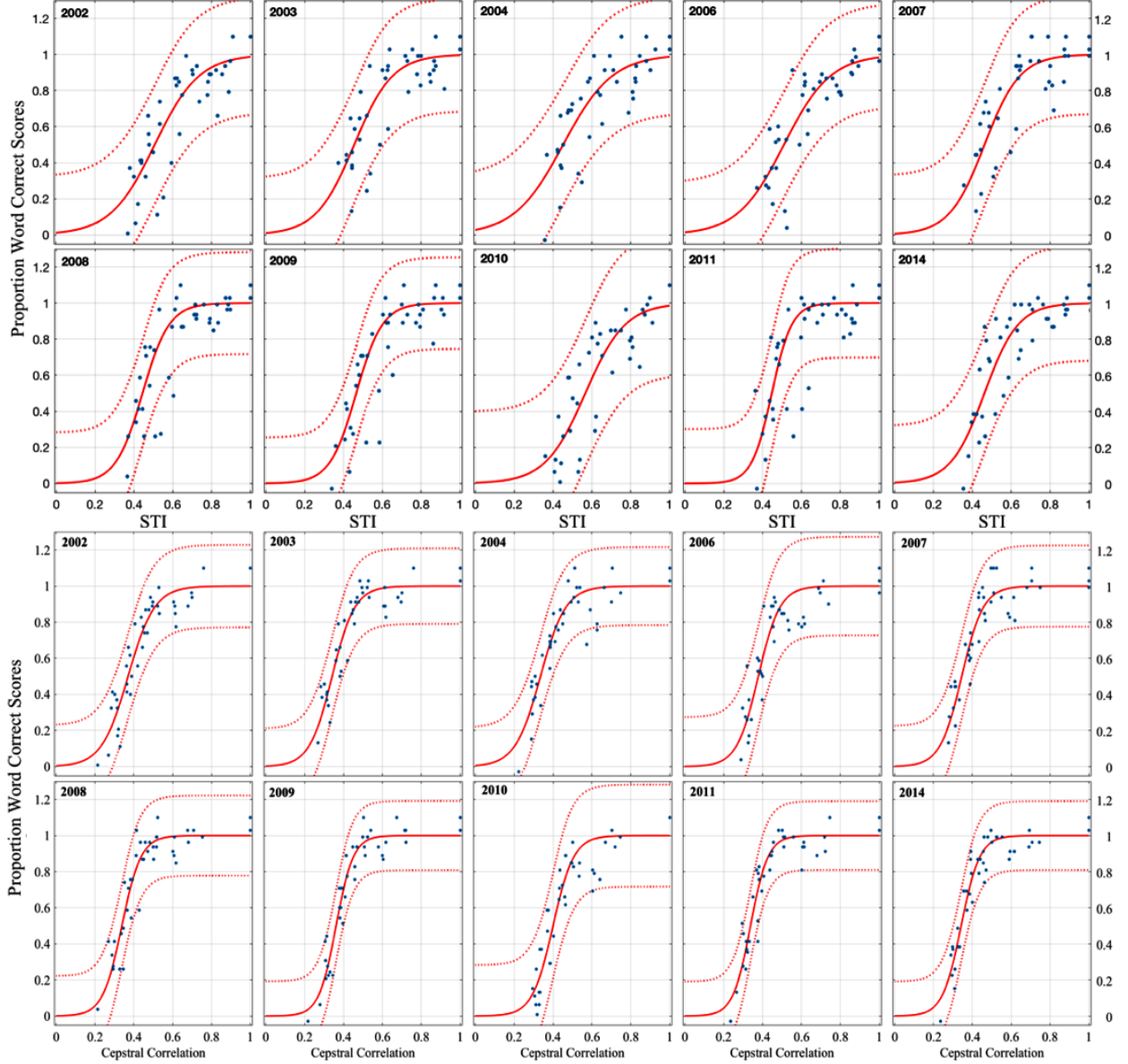


Figure 3.13: Individual best fits for calculated metrics to proportion word correct scores in NH group. Top two rows indicate the STI values fitted to the proportion word correct while the bottom two rows indicate cepstral correlation values fitted to the proportion correct scores.

### 3.4 Discussion

This experiment was designed to understand the role of envelope-based cues in the perception of monaural speech in reverberant situations and investigate the effect of envelope restoration as a

means of reversing decrements in intelligibility due to reverberation for people normal hearing and with hearing loss. This information could be used to remedy intelligibility deficits in reverberation through restoration of the envelope in reverberant environments for listeners with hearing loss. The results presented in the previous section are discussed using the framework of our hypotheses.

### **3.4.1 Intelligibility, envelope-related changes and cepstral correlation in reverberation**

#### **3.4.1.1 Intelligibility and Envelope Cues**

Our first aim was to study whether envelope-based cues were the main source of loss of intelligibility due to reverberation. Overall, both groups were affected by the reverberation (except for the NH group in Room 1) with greater deficits occurring with increased reverberation times (Figure 3.8a). The group with hearing loss were affected more by the same amount of reverberation than the NH group (Nábělek and Pickett, 1974). Even the room with the smallest reverberation time resulted in a significant loss in the ability to understand speech in the group with hearing loss. These results are consistent with previous studies that show that listeners with hearing are especially susceptible to degradations in speech intelligibility caused by reverberation (Gelfand and Hochberg, 1976; Nábělek and Robinson, 1982; Nábělek and Mason, 1981; Nábělek and Robinette, 1978).

There is a significant benefit of about 5-7% when listening with both ears irrespective of the amount of reverberation and hearing loss (Nábělek and Robinson, 1982). Thus, our results are also in line with expected intelligibility scores in binaural perception of reverberant speech (Helfer and Wilbur, 1990; Gordon-Salant and Fitzgibbons, 1999; Sato et al., 2007; George, Festen and Houtgast, 2008).

When only provided envelope cues (the baseline condition of Clean Voc), the participants in our group with hearing loss displayed lower envelope-only processing scores (Figure 3.5). This deficit in perception with cochlear hearing loss has been recorded by some studies (Souza and Boike,

2006; Turner, Chi and Flock, 1999) but is unexpected given the results of Turner et al. (1995) who showed that listeners with mild to moderate hearing loss were able to use temporal cues as well as listeners with normal hearing when accounting for audibility. Turner et al. (1999), however, also acknowledged that decrements in performance in participants with hearing loss could also occur due to age and difficulty of the listening task and due to much wider auditory filters than expected based on the thresholds. Participants in our group with hearing loss were older than participants in our NH group. The lack of significant differences in the unprocessed scores between the groups should rule out differential difficulty due to the sentences themselves, there could be a potential interaction of task complexity with the changes due to processing.

The pattern of intelligibility loss in envelope-only perception of speech were similar to the pattern seen due to reverberation (Figure 3.8.a and b); increasing reverberation results in a greater reduction of usable cues. Even with the least amount of reverberation, envelope-related distortions that affected intelligibility were significant. Increasing amounts of reverberation resulted in an increase of these envelope-related changes.

In analyzing the effect of reverberation on the TFS (Figure 3.8.c) in the clean context, we found that the reverberant TFS does not seem to impact intelligibility significantly compared to the effects of the envelope in the context of the clean envelope. Both listeners in the normal hearing and group with hearing loss are equally susceptible to changes to the changes in the TFS due to reverberation. Similar findings were observed by Srinivasan and Zahorik (2014) who showed greater reduction in intelligibility when using reverberant-envelope than reverberant-TFS cues.

#### **3.4.1.2 Cepstral correlation and intelligibility in reverberation**

The cepstral correlation calculation here models the consistency with which the time-frequency modulations of the original signal are reproduced after passing through a model of the auditory periphery (Kates and Arehart, 2014). Both on average and at the individual levels, cepstral correlation explains speech intelligibility in reverberant and envelope processed conditions better than using the STI model (Figures 3.11, 3.12 & 3.13). The cepstral correlation-based model explains more of

the variability and results in better fits of the data compared to the STI both at the group level. Cepstral correlation explains about 14-15% improvement in the  $R^2$  at the group level compared to the STI predictions. Similar results have been obtained when characterizing signal changes by Christiansen et al, (2010) who also had an auditory pre-processing model. The peripheral model used here is more comprehensive and resulted in similar or better predictions in a wider variety of hearing aid signal processing conditions (Kates and Arehart, 2014).

Increased reverberation results in greater spectral and temporal smearing of speech information in addition to filling in of the silent gaps between the speech segments (Nábělek and Nábělek, 1994). Hearing loss results in a widening of the auditory filters and loss of non-linearity in addition to the increase in thresholds (Moore, 2008). This loss of spectral and temporal acuity further decreases fidelity of the spectrotemporal modulations in the speech envelope. Thus, greater reverberation/envelope changes lead to lower metric predictions for subjects with hearing loss compared to the those with normal auditory systems. This reduction can be seen as a shift in metric values towards the lower end of the cepstral correlation axis for the group with hearing loss compared to the NH group (Figure 3.9 top row). At lower amounts of reverberation, this decrease in signal fidelity did not result in significant loss of intelligibility in Room 1 for the NH but causes significant perceptual issues for the group with hearing loss. Increasing amounts of reverberation however, resulted in larger reductions in the intelligibility among participants in both groups. Thus, cepstral correlation captures some effects of reverberation with cochlear hearing loss.

Some of the differences among the participants in the group with hearing loss arising from threshold elevation and other hearing loss-related changes (lowering of cepstral correlation values) are reflected in the individual level fits of cepstral correlation with the difference in the range of cepstral correlation values and the slope of the curves. However, the difference between the amount of variability in the data explained by cepstral correlation alone versus STI is different in different people. For four out of the twelve subjects, the STI provides a better or equally good descriptor as the cepstral correlation and the extent of additional benefits of the metric are highly variable (Table 3.13). This disparity reflects the limitations of the cepstral correlations as a within-subject level

predictor. These limitations on its usefulness could be a possible result of the cepstral correlation being a purely signal envelope related description and individual difference in the weighting of the cues or a result of other higher-level processes involved in the perception of speech (Kates and Arehart, 2014; 2015). Also, the higher predictions of intelligibility in the group with hearing loss compared to the NH group for similar values of cepstral correlation in the region above 0.5 possibly suggests a greater reliance on envelope cues by listeners with hearing loss in presence of impoverished spectral information when enough good envelope information is present but switching to use all available information with less amounts of envelope information (Souza et al., 2015). More analysis at the individual level are required to understand this effect.

### 3.4.1.3 Low-rate envelope fidelity and intelligibility

To understand the changes in the time-frequency modulations (cepstral correlation) of the envelope at the different modulation rates and acoustic frequencies at the individual level, we analyzed the effect of reverberation on the envelope with increasing reverberation (Kates and Arehart, 2015). The data from two HL subjects (1002 and 1005) are shown in Figure 3.14. The pattern of change in all four rooms is shown for both subject. These changes in each room are presented in a panel consisting of two plots. The room numbers are indicated in each panel. The top plot shows the modulation correlation between clean and reverberant speech at the different rates and different acoustic frequencies. The bottom plot shows the same information averaged across the different acoustic frequencies at each modulation rate (averaging columns in the top plot). It can be seen that envelope modulation correlations decrease with increasing reverberation, especially at the low modulation rates. This decrease is consistent with intelligibility decrements with increasing reverberation. This result is in line with the prediction made from a model with a single reflection (Muralimanohar et al., 2017). Data from two HL participants are presented. HL 1002 had overall higher-than-average thresholds at every audiometric frequency except 8 kHz while HL 1005 had normal and near-normal low frequency thresholds up to 2 kHz and sloping loss at higher frequencies with the highest thresholds above 4 kHz. The plots from both subjects show similar patterns of

decreasing envelope fidelity at the low rates with increasing reverberation. However, HL 1002 was more susceptible to the effects of reverberation than HL 1005 with lower word recognition scores in every room. This increased variability in performance with similar pure-tone thresholds is representative of speech perception for people with hearing loss in reverberant rooms (Nábělek and Robinson, 1974,1982; Nábělek and Mason, 1981; Nábělek and Robinette, 1974; Gordon-Salant and Fitzgibbons, 1999). This difference could again be possibly due to factors that cannot be wholly captured by just quantifying part of the signal changes. One possible explanation, which cannot be elicited from the current data, could occur due the source of hearing loss (inner hair cell vs outer hair cell loss) due to which the clarity of the neural representations of sounds in complex situations might be different for different listeners with similar pure-tone thresholds based on age and other factors (Gordon-Salant, 2006).

Relative to Aim 1 , we found that listeners with hearing loss exhibit greater susceptibility to reverberation. Nábělek and Robinette (1974) also concluded that though listeners with hearing loss perform worse than listeners with normal hearing for the same amount of reverberation but the pattern of changes with increasing amounts of intelligibility are similar in both groups. These participants with hearing loss also showed similar patterns of deficits when using degraded envelope-only information. Also reverberant TFS cues seem to result in only a small amount of change in perception in the context of clean envelope information for both groups.

The amount of reverberation-related changes to intelligibility is concomitant on the amount of changes to the envelope information present (Lochner and Burger, 1964; Steeneken and Houtgast, 1980; Payton and Shrestha, 2013; Muralimanohar et al., 2017) as seen through changes in cepstral correlation values and envelope fidelity plots. Lowered envelope fidelity at the low modulation rates accounts for most of the decrements in intelligibility connected with increasing reverberation (Muralimanohar et al., 2017; Houtgast and Steeneken, 1985).

Our results also confirm our sub-hypothesis that cepstral correlation is a better descriptor for modeling the effects of reverberation on speech intelligibility at group level, especially when considering the effects of cochlear hearing loss, than conventional narrow band analysis. However, this

**HL 1002**

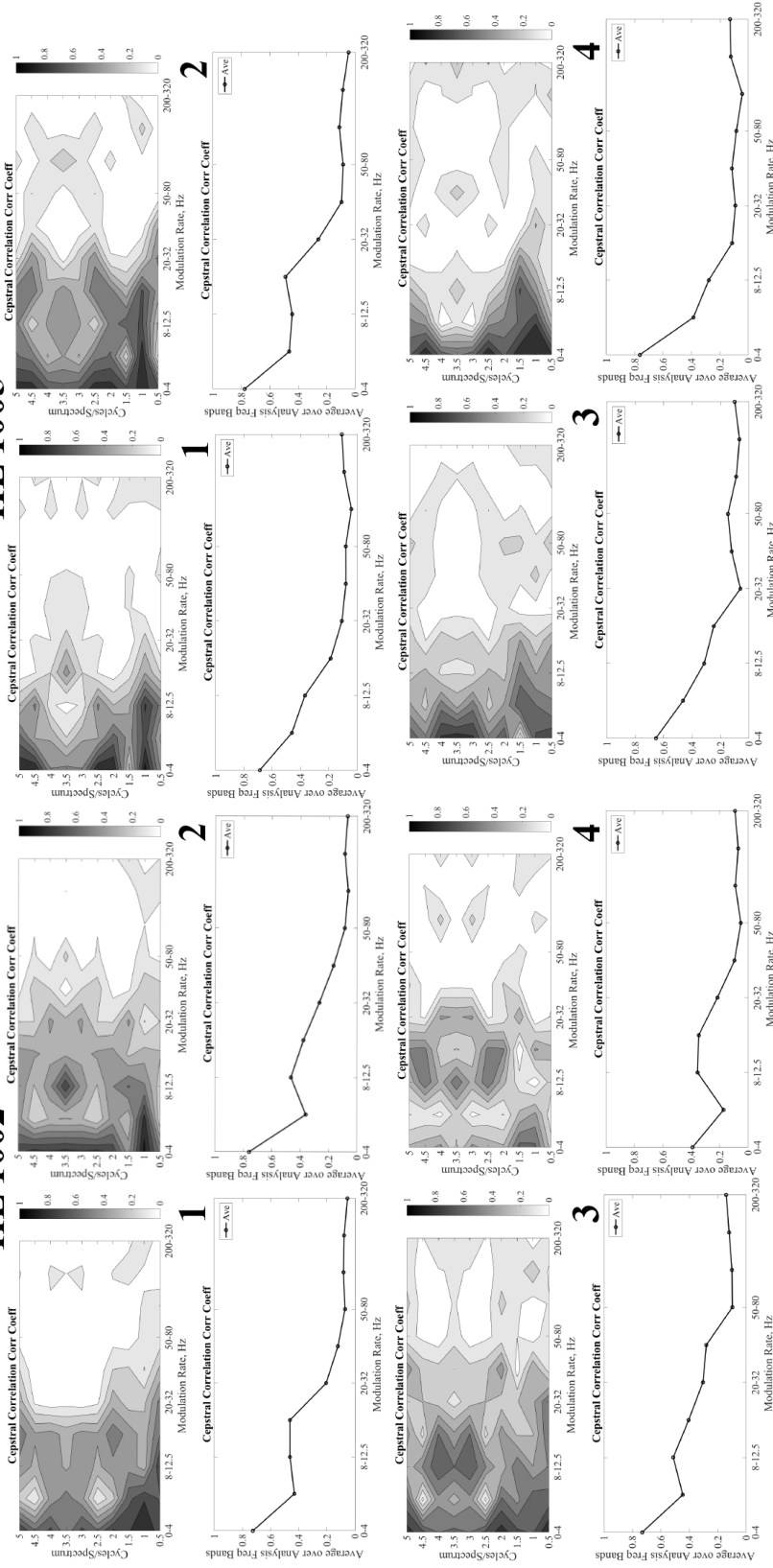


Figure 3.14: Effect of Increasing reverberation. Modulation Correlations as cepstral correlations at the different rates and acoustic frequencies are shown for two subjects (2 HL - 1002 and 1005) in the four rooms. Room numbers are indicated in each set of plots. Each room consists of two plots; the top plot indicates the envelope fidelity at the different modulation rates and acoustic frequencies and the bottom plot indicates the same correlation scores averaged across the different acoustic frequencies at each modulation rate. Within subject reductions in the envelope modulation correlations result in concomitant reductions in intelligibility.

metric does not account for all the patterns seen in behavioral data, and is less powerful when used to describe individual-level patterns in intelligibility as it does not include the effect of higher level and top-down influences on speech perception which might affect the perception of degraded speech, especially, in people with hearing loss. Thus, it is possible that the overall perception depends on the total amount of usable information present in the acoustic signal but is susceptible to other factors beyond the scope of the variables in this experiment. Listeners with normal hearing who can extract more information from the reverberant mix and have intact suprathreshold processing thus do better in reverberation. It is possible that envelope information drives perception in reverberation for listeners with and without hearing loss (Srinivasan and Zahorik, 2014; Watkins et al. 2011), especially at higher levels of reverberation and is more critical for listeners with hearing loss who might rely more on these cues (Souza et al., 2015).

### **3.4.2 Ideal envelope restoration of reverberant speech and the role of TFS**

The second experimental aim was to establish whether restoring the envelope cues below 30 Hz would improve intelligibility for both listeners with normal hearing and with hearing loss. For this aim, we hypothesized that restoring the envelope should provide benefits for listeners from both groups. We also postulated that any usable TFS information would result in an improvement in both groups of listeners.

Shown in Figure 3.9 (left), the difference between scores in the ideal envelope restoration and reverberant condition show a similar pattern to the effect of reverberation. The amount of improvement with the restored envelope is proportional to the amount of decrement due to reverberation. Both NH and HI groups seem to gain from the restored envelope. Our results showed that ideal envelope restoration improved intelligibility for participants in both the NH and group with hearing loss. Thus, restoring only the low-rate modulation cues ( $<30$  Hz) resulted in reversing the effects of reverberation in both groups. This result is consistent with the literature that shows that low-rate modulation cues are most important in the perception of speech (Kates and Arehart, 2014, 2015; Smith et al., 2002; Shannon et al., 1995).



The improvement from envelope restoration comes from combining low rate envelope cues (< 30 Hz modulation-rate cues) with the reverberant TFS. This comparison confirms the benefit from retaining the reverberant TFS is consistent irrespective of amount of reverberation. The difference between the normal hearing and group with hearing loss comes from the NH group being closer to ceiling performance with both kinds of processing. The reverberant TFS was shown to not interfere significantly with perception in a clean envelope context in the previous section. While it is expected that listeners with hearing loss will be more impacted by impaired TFS processing (Hopkins and Moore, 2008) and are not expected to benefit from TFS cues, our HL listeners did not show such effects. The source of this difference could be that in our study in comparison to the clean vocoded speech, the reverberant TFS in the Ideal Env Restoration condition provides additional enough pitch and voicing information that aids perception (through improvements in perception of vowels, voiced segments and formant transitions) even in our listeners with hearing loss. Also, it is known that with vocoder processed speech, the TFS information could potentially be combined across bands at the outputs of the auditory filters to recreate envelope information (Ghitza, 2001; Moore 2008); though the extent of the recovery would depend on the decoding of the TFS information which is imperfect in cochlear hearing loss. A simpler explanation in line with the results from the previous section could be that effects of reverberation on the TFS might be minimal and perception in reverberation (Watkins et al., 2011; Srinivasan and Zahorik, 2014) might be driven by envelope processing so in the presence of perfect envelope information, additional information that is not completely scrambled by the reverberation could continue to aid perception as long as it is accessible. Our results indicate that retaining the TFS irrespective of the level of reverberation should aid perception in attempts to restore the envelope for listeners with both normal hearing and with hearing loss.

We performed the envelope fidelity analysis similar to that of the previous section to understand how well the envelope was restored given we used scrambled TFS. Figure 3.15 shows the envelope fidelity in the ideal envelope restored conditions in the various rooms for the same two subjects from Figure 3.14. The graphs are presented in the same order as in Figure 3.14. The

envelope at rates  $< 30$  Hz is restored for all three listeners in all three rooms (compare correlations with corresponding plots in Figure 3.13 in reverberation). In terms of intelligibility benefits, HL 1002 gained from the restoration while HL 1005 did not gain in Room 1 and 2 where their scores in reverberant speech are much higher than average. HL 1002 performed poorly with all levels of reverberation and benefited significantly from the restoration in all the rooms. This evidence might also point to listeners weighting the type of cues they use in a listening situation (Souza et al, 2015).

Overall, our results confirm our Aim 2 hypothesis that restoring the envelope should improve intelligibility for participants in both groups. Also, our sub-hypothesis that retaining the TFS should improve intelligibility was also validated by the data.

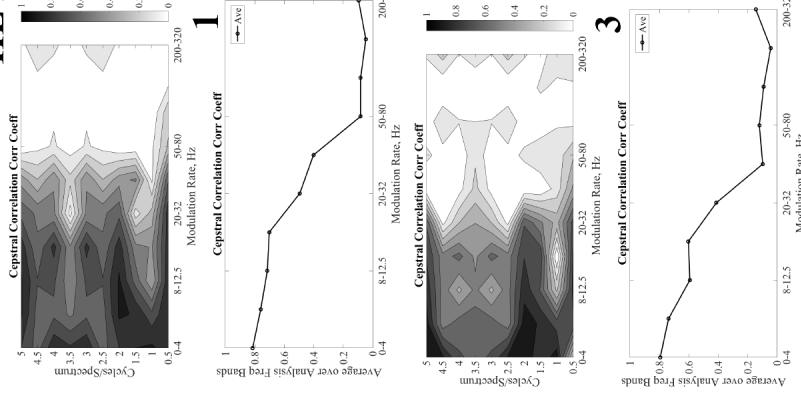
### 3.4.3 Expansion to improve intelligibility of reverberant speech

The third aim of this experiment was to ascertain the benefit of envelope expansion in restoring the intelligibility of reverberant speech. We hypothesized that the methods should produce benefits that were relative to the improvements in envelope modulations. We also posited that the IMB methods would produce greater benefits because they provided optimized-levels of restoration based on the speech stimulus and reverberation.

Our results indicated that none of the proposed methods produced significant benefits to overcome the effects of reverberant in either group. The IMB expansion seems to produce greater benefits across both groups in all rooms though these benefits were not significant at the group level. Envelope correlations at the different modulation rates and acoustic frequencies in the cepstral correlation at all the low modulation rates were not improved like the effects observed in the ideal envelope restoration condition. These results are in agreement with Apoux et al (2004) and van Buuren et al (1999). One difference between studies that showed small improvements from expansion (Lorenzi et al., 1999; Apoux et al., 2001) was that those studies expanded the envelope at rates up to 256 Hz. Thus, it remains unclear whether such expansion would benefit people with mild-moderate levels of cochlear hearing loss.

Overall the cepstral correlations showed small improvements with the expansion without a

## HL 1002



## HL 1005

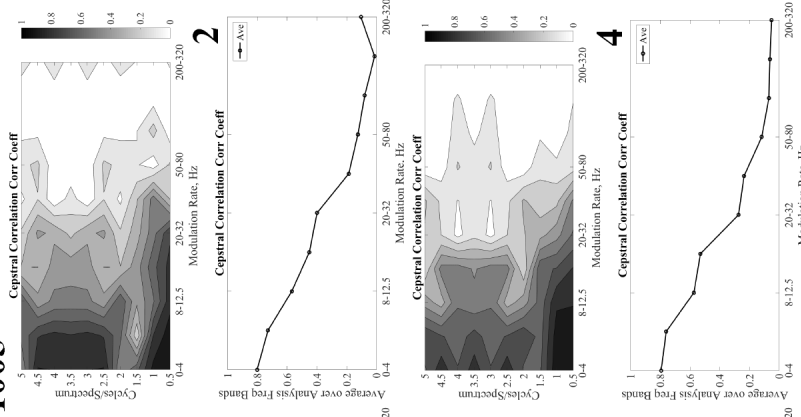
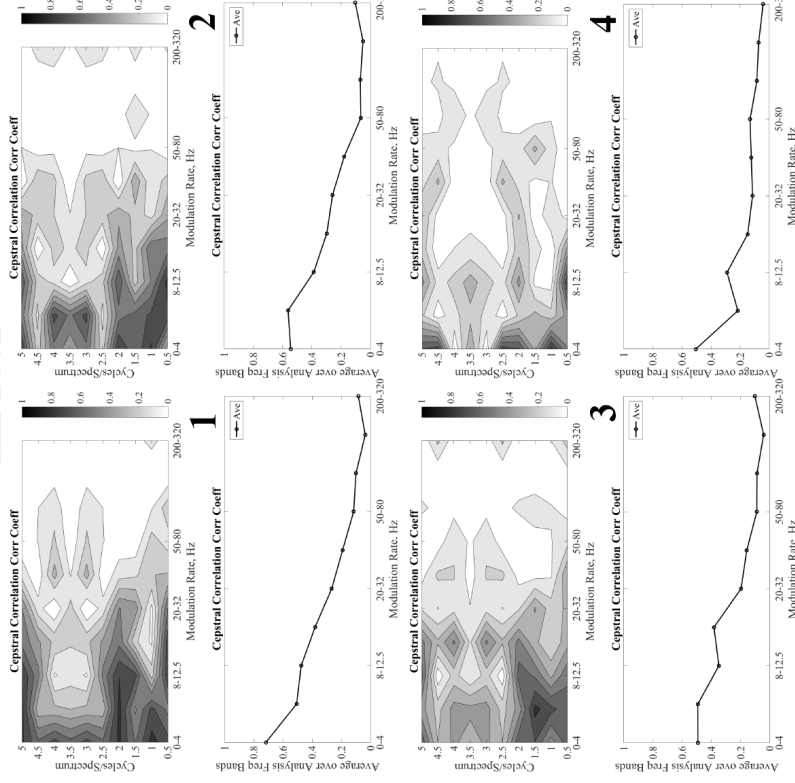


Figure 3.15: Effect of Ideal Envelope Restoration. Modulation Correlations as cepstral correlations at the different rates and acoustic frequencies are shown for two subjects (2 HL – 1002 and 1005) in the four rooms. Room numbers are indicated in each set of plots. Each room consists of two plots; the top plot indicates the envelope fidelity at the different modulation rates and acoustic frequencies and the bottom plot indicates the same correlation scores averaged across the different acoustic frequencies at each modulation rate. Within subject reductions in the envelope modulation correlations result in concomitant reductions in intelligibility.

proportional improvement in performance in all subjects (as suggested by group cepstral correlation fit in Figure 3.11). The individual performance was not completely described by the changes in the modulation correlations at the low rates. The envelope modulation correlations in the  $f_{lo}$  and  $a_{mag}$  conditions are illustrated in Figures 3.16 and 3.17 respectively. For example, subject HL 1002 who gained 26% and 18% in word recognition in those conditions due to a 3% improvement in cepstral correlation across all frequencies and rates in Room 1 and gained about 37 and 42% from a doubling in improvement of the metric in Room 2. However, subject HL 1005, did not gain from those processing schemes in any of the rooms, with very small improvements ( $< 3\%$  in the higher reverberation rooms), if any, in cepstral correlation. Also, subject HL 1005 had better low frequency hearing than subject HL 1002. HL 1002 displayed very poor scores in reverberation and gained from nearly all types of expansion processing and HL 1005's performance was at the top end of the range of the group average in reverberation (except in Room 4). From the individual fits of cepstral correlation, (row 1 of cepstral correlation fits in Figure 3.12) HL 1002 has a steeper fit compared to HL 1005. Thus, there could be a complex interaction between the person's hearing loss, susceptibility to distortions to hearing loss, level of reverberation, and possibly other factors that determine benefit from the processing. Expansion did not result in sufficient improvements in cepstral correlation for the NH subjects. One procedural reason for the lack of improvement, especially in the IMB schemes, was the limitation on the possible amounts of expansion built into the processing scheme. The amount of expansion was limited to prevent any possible recruitment or upward spread of masking and distortion from having too much expansion in any of the bands (Apoux et al., 2001). Overall, our testing indicated that envelope expansion did not sufficiently improve envelope modulations and result in significant improvements in intelligibility in either group. Our hypothesis that expansion should restore intelligibility relative to the amount of improvements to the envelope modulations provided remains unsupported. The secondary hypothesis that IMB expansion should result in greater benefit was weakly supported by the greater, non-significant improvements seen with the IMB processing.

## HL 1002



## HL 1005

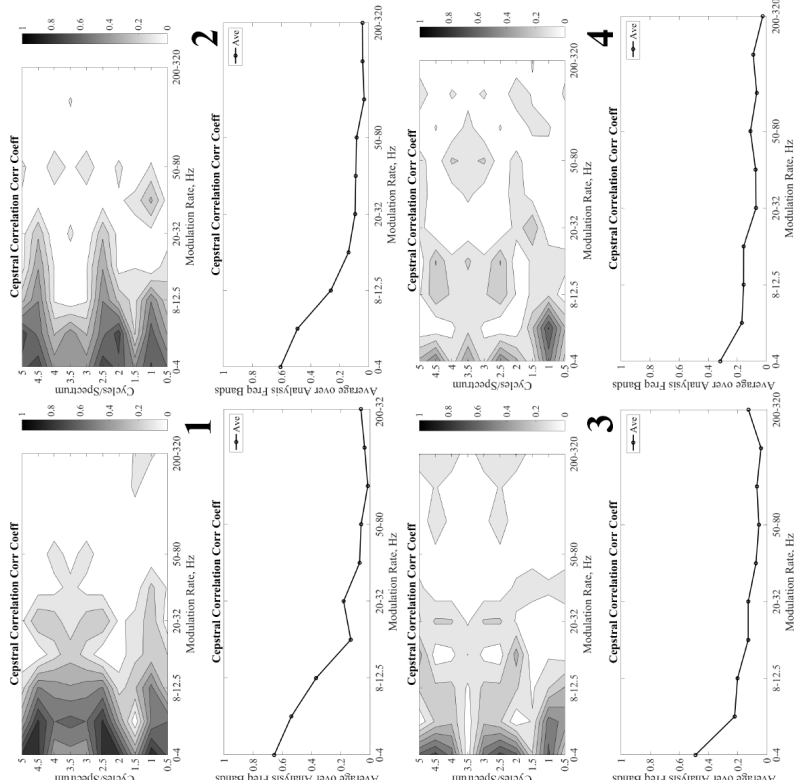
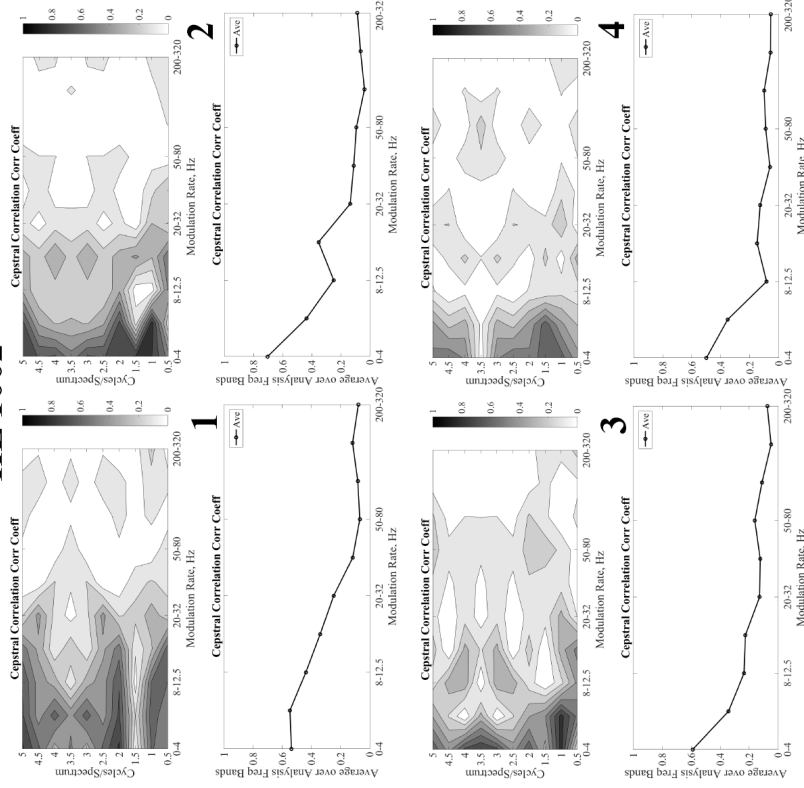


Figure 3.16: Effect of  $f_0$  processing as cepstral correlations at the different rates and acoustic frequencies are shown for two subjects (2 HL – 1002 and 1005) in the four rooms. Room numbers are indicated in each set of plots. Each room consists of two plots; the top plot indicates the envelope fidelity at the different modulation rates and acoustic frequencies and the bottom plot indicates the same correlation scores averaged across the different acoustic frequencies at each modulation rate. Within subject reductions in the envelope modulation correlations result in concomitant reductions in intelligibility.

## HL 1002



## HL 1005

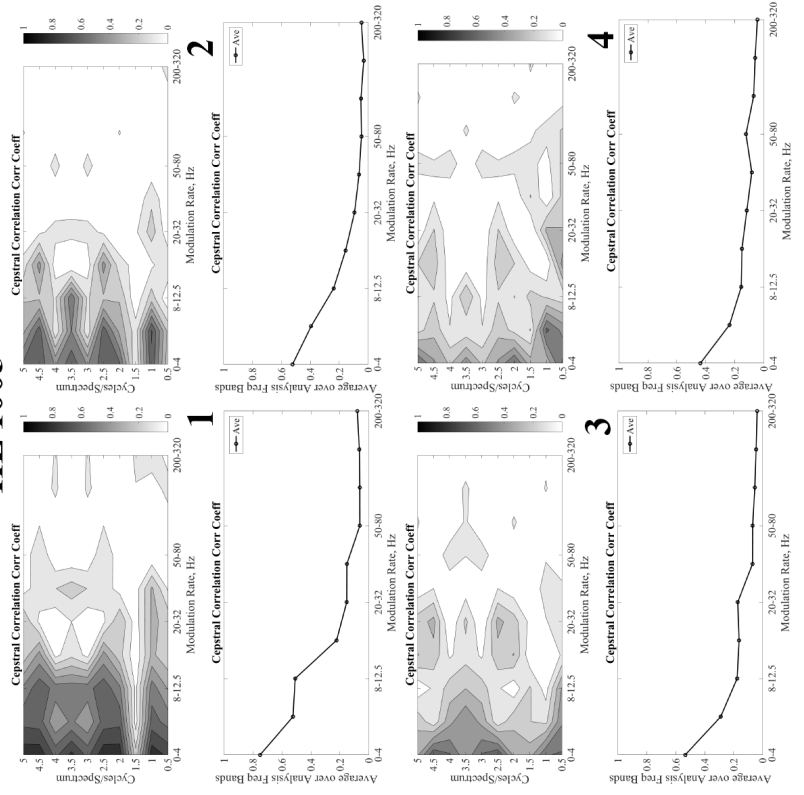


Figure 3.17: Effect of  $a_{mag}$  Modulation Correlations at the different rates and acoustic frequencies are shown for two subjects (2 HL – 1002 and 1005) in the four rooms. Room numbers are indicated in each set of plots. Each room consists of two plots; the top plot indicates the envelope fidelity at the different modulation rates and acoustic frequencies and the bottom plot indicates the same correlation scores averaged across the different acoustic frequencies at each modulation rate. Within subject reductions in the envelope modulation correlations result in concomitant reductions in intelligibility.

### 3.5 Conclusions

The effect of reverberation-related envelope degradations on the intelligibility of monaural speech-level stimuli was studied with a motivation to design envelope restoration techniques that reversed these changes. The relevant envelope modulation changes were characterized using cepstral correlation - a metric that quantified the changes in the spectrotemporal patterns in the envelope that spanned across multiple frequency bands. This metric provided a better description of the envelope-related changes to intelligibility on average in both listeners with normal hearing and with hearing loss compared to the traditional STI description. However, patterns in the individual calculations of the metric (though it described the data better for a majority of the subjects in the group with hearing loss) showed some gaps in the ability to describe intelligibility.

Cepstral correlation has been successfully used along with a description of the coherence between the clean and processed schemes to characterize the changes in speech intelligibility in a variety of noise-related and signal processing distortions including with wide-dynamic range compression (WDRC) in hearing aids which also involves changes to the envelope (Kates and Arehart, 2014). The effect of reverberation is frequency dependent (Houtgast and Steeneken, 1985). Also, the contributions of the envelope rates above 30 Hz (Zahorik et al., 2009) in reverberation and relative levels of contributions of each band of modulation rates were not considered in this experiment. Thus, it might be possible to improve this characterization by including additional information about other intelligibility-related signal changes seen in reverberation.

Use of additional descriptors of individual variability like the spectrotemporal acuity or cognitive factors could be used to improve group and individual predictions of intelligibility. For example, age (Gordon-Salant and Fitzgibbons, 1999; Fujihara, Shiraishi and Remijn, 2017) and working memory (Reinhart et al., 2018) are two known factors that are known to affect perception in reverberation and might explain some of the increased susceptibility to reverberation seen in people with hearing loss. A stepwise regression using the individual factors that described the ability to discriminate speech in noise, temporal measurements (gap and modulation depth discrimination

scores) and spectral acuity (ripple scores) and pure tone thresholds did not provide significant additional explanations of the variability of the data. Hence there is a need to further explore additional descriptions of hearing loss and subject factors.

Restoration of envelope cues, at rates below 30 Hz, was shown to improve the intelligibility of reverberant speech for both listener groups in our study. Thus, envelope restoration could potentially be used as a means to restore intelligibility in monaural perception of reverberant speech. Though envelope restoration through the envelope expansion attempted here did not offer significant benefit on average over reverberation, there was some limited benefit to some of our participants in the group with hearing loss.

Signal processing to combat reverberation in hearing aids would greatly benefit hearing aid users since listeners with cochlear hearing loss who are susceptible to the effects of reverberation. Methods that use a pre-defined set of parameters that could counter reverberation like the IMB methods that used averaged expansion factors could offer huge advantages over current more complex schemes. Thus, envelope restoration schemes could offer avenues for improvement in intelligibility in reverberant situations. These results build the case for further research into envelope-based dereverberation algorithms and integration of descriptions of subject factors into hearing aid signal.



## Chapter 4

### Conclusions

#### 4.1 Summary and Conclusions

Most people spend a considerable portion of their day in enclosed environments which could potentially interrupt the normal patterns of cues present in speech due to reverberation. Reverberation also creates much greater disruptions in the intelligibility of speech in people with hearing loss. Thus, countering reverberation to restore understanding for these listeners remains an important goal in auditory rehabilitation. The study assessed envelope-related signal and the connected intelligibility changes in reverberant speech in order to ascertain if restoring the envelope of reverberant speech would improve its intelligibility for people with mild to moderate-severe cochlear hearing loss.

Two main objectives were established to answer our experimental question. The first objective was to systematically relate the signal envelope changes due to reverberation and the consequent intelligibility through a metric that quantified relevant signal changes and included the effects of hearing loss. The second objective involved testing if envelope restoration would reverse the deleterious effects of reverberation on speech intelligibility, especially for people with hearing loss. These objectives were achieved through two experiments.

Experiment 1 was used to quantify the reverberation-related changes to the envelope in a simple model using a single reflection of varying strength and relative delay added to the direct sound. A quantification of the spectrotemporal changes to the signal envelope across frequency

bands using the cepstral correlation, which includes a model of the auditory system, that has been used before in describing the intelligibility of hearing aid processed speech (HASPI; Kates and Arehart, 2013; 2014) was proposed as an alternative. This metric was found to be highly correlated with the STI which was a good predictor of speech intelligibility in reverberation for models of normal and cochlear hearing loss. This study confirmed the link between decrements in the low-rate envelope modulations due to reverberation and predicted intelligibility in this simple model. Through analysis of the spectrotemporal changes and fidelity of modulations in reverberant speech the effects of early and late reflections were also confirmed. Late reflections were shown to result in greater interruptions of the envelope at the low-rates while earlier reflections did not interfere with envelope modulations at these rates in these simulations.

Experiment 2 was designed to address both objectives. Changes to spectrotemporal patterns in the envelope were systematically linked to the changes in intelligibility. Reverberation-related changes to the low-rate envelope changes were identified as the main drivers of intelligibility changes in reverberation. Cepstral correlation better described the group level patterns in the intelligibility of reverberant speech compared to conventional STI measures. However, cepstral correlation could not completely capture all the patterns at the level of an individual listener.

Results from experiment 2 also confirmed that restoration of the envelope at the low modulation rates restored intelligibility for people with hearing loss. Additionally, several versions of an envelope-expansion algorithm that aimed to restore the envelope did not significantly improve intelligibility in the reverberant environments tested, though some of them resulted in modest gains in intelligibility.

## 4.2 Future Work and Implications for Hearing aid signal

### Processing

Our overall results supported the idea of pursuing envelope restoration-based solutions for improving speech communication in reverberation for use in auditory prosthetics like hearing aids.

Though the specific algorithms used here failed to provide improvements, implementation of blind algorithms for de-reverberation offer a computationally cheap alternative and possibly more effective alternative to current de-reverberation strategies.

Our expansion schemes sought to improve the depth of modulation through expansion, however Apoux et al. (2004) found that using a combination of compression and expansion resulted in an improvement in intelligibility of noisy speech through improvement of the modulation depth in both NH and HL listeners. Testing such a scheme using a framework similar to our IMB method where selective compression-expansion constrained by the clean envelope offers a possible processing strategy to explore.

Also, our experiment probed monaural speech perception in reverberation and an important factor that needs to be considered before implementation in real hearing aids is the effect of binaural listening. For example, independent envelope restoration algorithms in each ear might result in potentially different gains in different ears to restore the envelope and the effects of combining such inputs across both ears needs to be determined. It may also be possible to combine reverberation processing with remote microphones. The clean signal from the remote microphone can provide the reference speech envelope used as the basis for envelope restoration at the two ears. However, the effectiveness of these cues in restoring spatial cues and the result of differences when combined binaurally need to be investigated.

In our study all of the reverberation descriptors ( $T_{60}$ , DRR, CTE) were used to create an ordinal variation in the reverberation. A closer analysis might provide clues into the contributions of early/late reflections and the reverberant tail which might potentially identify other avenues for algorithm improvements.

Thus, developing signal-processing strategies to improve speech intelligibility in reverberation remains a valid research objective. This dissertation has shown that ideal envelope restoration improves speech intelligibility for both NH and HL listeners even though the specific family of algorithms that was evaluated provided minimal benefit. Further research to develop and evaluate reverberation processing could lead to better speech intelligibility in reverberation and improved

hearing aids.

## References

- ANSI. (1960). *ANSI S12.60-2002: Acoustical performance criteria, design requirements, and guidelines for schools* (Standard).
- ANSI. (2002). *ANSI S1.1: American National Standard Acoustical Terminology* (Standard).
- Apoux, F., Crouzet, O., & Lorenzi, C. (2001). Temporal envelope expansion of speech in noise for normal-hearing and hearing-impaired listeners: effects on identification performance and response times. *Hearing Research*, 153(1-2), 123-131.
- Apoux, F., Tribut, N., Debrulle, X., & Lorenzi, C. (2004). Identification of envelope-expanded sentences in normal-hearing and hearing-impaired listeners. *Hearing Research*, 189(1-2), 13-24.
- Arai, T., Murakami, Y., Hayashi, N., Hodoshima, N., & Kurisu, K. (2007). Inverse correlation of intelligibility of speech in reverberation with the amount of overlap-masking. *Acoustical Science and Technology*, 28(6), 438-441.
- Arweiler, I., & Buchholz, J. M. (2011). The influence of spectral characteristics of early reflections on speech intelligibility. *The Journal of the Acoustical Society of America*, 130(2), 996-1005.
- Assmann, P., & Summerfield, Q. (2004). The perception of speech under adverse conditions. In *Speech processing in the auditory system* (p. 231-308). New York: Springer-Verlag.
- Avendano, C., & Hermansky, H. (1996). Study on the dereverberation of speech based on temporal envelope filtering. In *Proceedings of Fourth International Conference on Spoken Language, 1996. ICSLP 96*. (Vol. 2, pp. 889-892).
- Benesty, J., Sondhi, M. M., & Huang, Y. (2007). *Springer handbook of speech processing*. Springer.
- Best, V., Keidser, G., Buchholz, J. M., & Freeston, K. (2015). An examination of speech reception thresholds measured in a simulated reverberant cafeteria environment. *International Journal Of Audiology*, 54(10), 682-690.
- Beutelmann, R., & Brand, T. (2006). Prediction of speech intelligibility in spatial noise and reverberation for normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, 120(1), 331-342.
- Bisgaard, N., Vlaming, M. S. M. G., & Dahlquist, M. (2010). Standard audiograms for the iec 60118-15 measurement procedure. *Trends in Amplification*, 14(2), 113-120.
- Boothroyd, A. (2004). Room acoustics and speech perception. *Seminars in Hearing*, 25(02), 155-166.
- Bradley, J. S. (1986). Predictors of speech intelligibility in rooms. *The Journal of the Acoustical Society of America*, 80(3), 837-845.
- Bradley, J. S., Sato, H., & Picard, M. (2003). On the importance of early reflections for speech in rooms. *The Journal of the Acoustical Society of America*, 113(6), 3233-3244.

- Brennan, M. A., Gallun, F. J., Souza, P. E., & Stecker, G. C. (2013). Temporal resolution with a prescriptive fitting formula. *American Journal of Audiology*, 22(2), 216-225.
- Byrne, D., & Dillon, H. (1986). The national acoustic laboratories'(NAL) new procedure for selecting the gain and frequency response of a hearing aid. *Ear and Hearing*, 7(4), 257-265.
- Chi, T., Gao, Y., Guyton, M. C., Ru, P., & Shamma, S. (1999, November). Spectro-temporal modulation transfer functions and speech intelligibility. *The Journal of the Acoustical Society of America*, 106(5), 2719-2732.
- Christiansen, C., Pedersen, M. S., & Dau, T. (2010). Prediction of speech intelligibility based on an auditory preprocessing model. *Speech Communication*, 52(7-8), 678-692.
- Clarkson, P. M., & Bahgat, S. F. (1991, March). Envelope expansion methods for speech enhancement. *The Journal of the Acoustical Society of America*, 89(3), 1378-1382.
- Davies-Venn, E., & Souza, P. (2014). The role of spectral resolution, working memory, and audibility in explaining variance in susceptibility to temporal envelope distortion. *Journal of the American Academy of Audiology*, 25(6), 592-604.
- Davies-Venn, E., Souza, P., Brennan, M., & Stecker, G. C. (2009). Effects of audibility and multichannel wide dynamic range compression on consonant recognition for listeners with severe hearing loss. *Ear and hearing*, 30(5), 494-504.
- Dowell, R. C., Flynn, M. C., et al. (2003). Effects of background noise and reverberation on the aided speech perception in adults with a severe or severe-to-profound hearing impairment. *The Australian and New Zealand Journal of Audiology*, 25(2), 63-73.
- Drullman, R. (1995). Temporal envelope and fine structure cues for speech intelligibility. *The Journal of the Acoustical Society of America*, 97(1), 585-592.
- Drullman, R., Festen, J. M., & Plomp, R. (1994). Effect of reducing slow temporal modulations on speech reception. *The Journal of the Acoustical Society of America*, 95(5), 2670-2680.
- Dudley, H. (1939, October). Remaking speech. *The Journal of the Acoustical Society of America*, 11(2), 169-177.
- Duquesnoy, A. J., & Plomp, R. (1980, August). Effect of reverberation and noise on the intelligibility of sentences in cases of presbycusis. *The Journal of the Acoustical Society of America*, 68(2), 537-544.
- Ewert, S. D., Volmer, J., Dau, T., & Verhey, J. (2008). Amplitude modulation depth discrimination in hearing-impaired and normal-hearing listeners. *Journal of the Acoustical Society of America*, 123(5), 3859-3859.
- Fabry, D., & Tchorz, J. (2005). Results from a new hearing aid using "acoustic scene analysis". *The Hearing Journal*, 58(4), 30-36.
- Finitzo-Hieber, T., & Tillman, T. W. (1978). Room acoustics effects on monosyllabic word discrimination ability for normal and hearing-impaired children. *Journal of Speech, Language, and Hearing Research*, 21(3), 440-458.
- Folkeard, P., Littmann, V., Scollie, S., et al. (2017). Using a dereverberation program to improve speech intelligibility and reduce perceived listening effort in reverberation. *Canadian Audiologist*, 4(4).
- Francis, A. L., Baldwin, K., & Nusbaum, H. C. (2000). Effects of training on attention to acoustic cues. *Perception & psychophysics*, 62(8), 1668-1680.
- Fu, Q. J., & Shannon, R. V. (1999). Recognition of spectrally degraded speech in noise with nonlinear amplitude mapping. *Acoustics*, 369-372 vol.1.
- Fujihira, H., Shiraishi, K., & Remijn, G. B. (2016). Elderly listeners with low intelligibility scores under reverberation show degraded subcortical representation of reverberant speech. *Neuroscience Letters*, 1-6.

- Gelfand, S. A., & Hochberg, I. (1976). Binaural and monaural speech-discrimination under reverberation. *Audiology*, 15(1), 72-84.
- George, E. L., Festen, J. M., & Theo Goverts, S. (2012). Effects of reverberation and masker fluctuations on binaural unmasking of speech. *The Journal of the Acoustical Society of America*, 132(3), 1581-1591.
- George, E. L. J., Festen, J. M., & Houtgast, T. (2008). The combined effects of reverberation and nonstationary noise on sentence intelligibility. *The Journal of the Acoustical Society of America*, 124(2), 1269-1277.
- George, E. L. J., Goverts, S. T., Festen, J. M., & Houtgast, T. (2010). Measuring the effects of reverberation and noise on sentence intelligibility for hearing-impaired listeners. *Journal of Speech, Language, and Hearing Research*, 53(6), 1429-1439.
- Ghitza, O. (2001). On the upper cutoff frequency of the auditory critical-band envelope detectors in the context of speech perception. *The Journal of the Acoustical Society of America*, 110(3), 1628-1640.
- Giurda, R., Georganti, E., Hassager, H. G., & Dau, T. (2017). Effects of wide dynamic range compression on speech signals with respect to reverberation. *The Journal of the Acoustical Society of America*, 141(5), 3821-3821.
- Goldsworthy, R. L., & Greenberg, J. E. (2004). Analysis of speech-based speech transmission index methods with implications for nonlinear operations. *The Journal of the Acoustical Society of America*, 116(6), 3679-3689.
- Gordon-Salant, S. (2006). Speech perception and auditory temporal processing performance by older listeners: Implications for real-world communication. *Seminars in Hearing*, 27(04), 264-268.
- Gordon-Salant, S., & Fitzgibbons, P. J. (1999). Profile of auditory temporal processing in older listeners. *Journal of Speech, Language, and Hearing Research*, 42(2), 300-311.
- Harris, R. W., & Swenson, D. W. (1990). Effects of reverberation and noise on speech recognition by adults with various amounts of sensorineural hearing impairment. *International Journal Of Audiology*, 29(6), 314-321.
- Hazrati, O., & Loizou, P. C. (2012, May). The combined effects of reverberation and noise on speech intelligibility by cochlear implant listeners. *International Journal Of Audiology*, 51(6), 437-443.
- Hedrick, M., & Younger, M. S. (2001). Perceptual weighting of relative amplitude and formant transition cues in aided CV syllables. *Journal of Speech, Language, and Hearing Research*, 44(5), 964-974.
- Hedrick, M. S., & Younger, M. S. (2003). Labeling of /s/ and /f/ by listeners with normal and impaired hearing, revisited. *Journal of Speech, Language, and Hearing Research*, 46(3), 636-648.
- Helfer, K. S., & Wilber, L. A. (1990). Hearing loss, aging, and speech perception in reverberation and noise. *Journal Of Speech And Hearing Research*, 33(1), 149-155.
- Hermansky, H., & Morgan, N. (1994). RASTA Processing of Speech. *IEEE Transactions on Speech and Audio Processing*, 2(4), 578-589.
- Hermansky, H., Wan, E. A., & Avendano, C. (1995). Speech enhancement based on temporal processing. In *IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP95* (p. 405-408). IEEE.
- Hirobayashi, S., Nomura, H., Koike, T., & Tohyama, M. (2000). Speech waveform recovery from a reverberant speech signal using inverse filtering of the power envelope transfer function. *Electronics and Communications in Japan Part III-Fundamental Electronic Science*, 83(6),

77-85.

- Hirsch, H., & Finster, H. (1988). The reduction of reverberation to improve automatic speech recognition in rooms. In *7<sup>th</sup> FASE Symposium, Edinburgh, Proceedings of Speech* (p. 913-919).
- Hochmair-Desoyer, I., Hochmair, E., Fischer, R., & Burian, K. (1980). Cochlear prostheses in use: Recent speech comprehension results. *Erfahrungen mit Cochlearprothesen: Erzieltes Sprachverständnis. Archives of oto-rhino-laryngology*, 229(2), 81-98.
- Hodoshima, N., Arai, T., & Kusumoto, A. (2002). Enhancing temporal dynamics of speech to improve intelligibility in reverberant environments. *Proceedings of Forum Acusticum, Sevilla*.
- Houtgast, T., & Steeneken, H. (1985). A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria. *The Journal of the Acoustical Society of America*, 77(3), 1069-1077.
- Houtgast, T., Steeneken, H., Ahnert, W., & Braidia, L. (2002). *Past, present and future of the Speech Transmission Index*.
- Houtgast, T., Steeneken, H., & Plomp, R. (1980). Predicting speech intelligibility in rooms from the modulation transfer function. I. General room acoustics. *Acta Acustica united with Acustica*, 46(1), 60-72.
- Houtgast, T., & Steeneken, H. J. M. (1973). The modulation transfer function in room acoustics as a predictor of speech intelligibility. *The Journal of the Acoustical Society of America*, 54(2), 557-557.
- Houtgast, T., & Steeneken, H. J. M. (1978). The modulation transfer function as a link between room acoustics and speech intelligibility. *The Journal of the Acoustical Society of America*, 63(S1), S6-S6.
- Humes, L. E., & Christopherson, L. (1991). Speech identification difficulties of hearing-impaired elderly persons: the contributions of auditory processing deficits. *Journal of Speech, Language, and Hearing Research*, 34(3), 686-693.
- IEC. (2003). *Sound system equipment-part 16: Objective rating of speech intelligibility by speech transmission index* (Tech. Rep. No. IEC 60268-16).
- Kanedera, N., Arai, T., Hermansky, H., & Pavel, M. (1999). On the relative importance of various components of the modulation spectrum for automatic speech recognition. *Speech Communication*, 28(1), 43-55.
- Kates, J. (2013). An auditory model for intelligibility and quality predictions. *The Journal of the Acoustical Society of America*, 133(5), 050184,1-6.
- Kates, J. M. (2001). Room reverberation effects in hearing aid feedback cancellation. *The Journal of the Acoustical Society of America*, 109(1), 367-378.
- Kates, J. M. (2008). *Digital hearing aids*. Plural publishing.
- Kates, J. M. (2011, June). Spectro-temporal envelope changes caused by temporal fine structure modification. *The Journal of the Acoustical Society of America*, 129(6), 3981-3990.
- Kates, J. M., & Arehart, K. H. (2014). The Hearing-Aid Speech Perception Index (HASPI). *Speech Communication*, 65, 75-93.
- Kates, J. M., & Arehart, K. H. (2015). Comparing the information conveyed by envelope modulation for speech intelligibility, speech quality, and music quality. *The Journal of the Acoustical Society of America*, 138(4), 2470-2482.
- Kates, J. M., Arehart, K. H., & Souza, P. E. (2013). Integrating cognitive and peripheral factors in predicting hearing-aid processing effectiveness. *The Journal of the Acoustical Society of America*, 134(6), 4458-4469.



- Kochkin, S. (2005). MarkeTrak VII: Customer satisfaction with hearing instruments in the digital age. *The Hearing Journal*, 58(9), 30–32.
- Kochkin, S. (2010). MarkeTrak VIII: Consumer satisfaction with hearing aids is slowly increasing. *The Hearing Journal*, 63(1), 19–20.
- Kohlrausch, A., Fassel, R., van der Heijden, M., Kortekaas, R., van de Par, S., Oxenham, A. J., & Püschel, D. (1997). Detection of tones in low-noise noise: Further evidence for the role of envelope fluctuations. *Acta Acustica united with Acustica*, 83(4), 659–669.
- Langhans, T., & Strube, H. (1982). Speech enhancement by nonlinear multiband envelope filtering. In *IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP82*. (Vol. 7, pp. 156–159).
- Lebart, K., Boucher, J.-M., & Denbigh, P. (2001). A new method based on spectral subtraction for speech dereverberation. *Acta Acustica united with Acustica*, 87(3), 359–366.
- Litovsky, R. Y., Colburn, H. S., Yost, W. A., & Guzman, S. J. (1999). The precedence effect. *The Journal of the Acoustical Society of America*, 106(4 Pt 1), 1633–1654.
- Lochner, J. P. A., & Burger, J. F. (1964). The influence of reflections on auditorium acoustics. *Journal Of Sound And Vibration*, 1(4), 426–454.
- Lorenzi, C., Berthommier, F., Apoux, F., & Bacri, N. (1999). Effects of envelope expansion on speech recognition. *Hearing Research*(1-2), 131–138.
- Lorenzi, C., Gilbert, G., Carn, H., Garnier, S., & Moore, B. C. J. (2006). Speech perception problems of the hearing impaired reflect inability to use temporal fine structure. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, 103(49), 18866–18869.
- Lu, X., Unoki, M., & Akagi, M. (2008). Comparative evaluation of modulation-transfer-function-based blind restoration of sub-band power envelopes of speech as a front-end processor for automatic speech recognition systems. *Acoustical Science and Technology*, 29(6), 351–361.
- Ludvigsen, C. (1987). Prediction of speech intelligibility for normal-hearing and cochlearly hearing-impaired listeners. *The Journal of the Acoustical Society of America*, 82(4), 1162–1171.
- Ludvigsen, C., Elberling, C., Keidser, G., & Poulsen, T. (1990). Prediction of intelligibility of non-linearly processed speech. *Acta oto-laryngologica. Supplementum*, 469, 190–195.
- Moore, B. C. (2007). *Cochlear hearing loss: physiological, psychological and technical issues*. John Wiley & Sons.
- Moore, B. C. (2008). The role of temporal fine structure processing in pitch perception, masking, and speech perception for normal-hearing and hearing-impaired people. *Journal of the Association for Research in Otolaryngology*, 9(4), 399–406.
- Moore, B. C. J. (2008). The role of temporal fine structure processing in pitch perception, masking, and speech perception for normal-hearing and hearing-impaired people. *Journal of the Association for Research in Otolaryngology*, 9(4), 399–406.
- Mourjopoulos, J. (1982). *Multiband envelope techniques for the removal of reverberation and noise from speech signals*. (Tech. Rep.). Institute of Sound and Vibration Research, University of Southampton.
- Mourjopoulos, J., & Hammond, J. (1983). Modelling and enhancement of reverberant speech using an envelope convolution method. In *IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP'83*. (Vol. 8, pp. 1144–1147).
- Muralimanohar, R. K., Kates, J. M., & Arehart, K. H. (2017). Using envelope modulation to explain speech intelligibility in the presence of a single reflection. *The Journal of the Acoustical Society of America*, 141(5), EL482–EL487.
- Murphy, & Shelley. (2018, Mar). *The open acoustic impulse response library*. Retrieved from <http://www.openairlib.net/resources> ([Online; accessed 31. Mar. 2018])

- Nábělek, A., & Nábělek, I. (1994). *Room acoustics and speech perception* (4th ed.; J. Katz, Ed.). Handbook of Clinical Audiology.
- Nábělek, A. K. (1975). Intelligibility of speech with simulated single reflection for normal and hearing-impaired subjects. *The Journal of the Acoustical Society of America*, 58(S1), S114-S114.
- Nábělek, A. K. (1986). Identification of vowels in noise and in reverberation produced by a female talker. *The Journal of the Acoustical Society of America*, 80(S1), S77-S77.
- Nábělek, A. K. (1988, August). Identification of vowels in quiet, noise, and reverberation: relationships with age and hearing loss. *The Journal of the Acoustical Society of America*, 84(2), 476-484.
- Nábělek, A. K. (1993). *Communication in noisy and reverberant environments*. Needham Height, MA: Allyn and Bacon.
- Nábělek, A. K., & Dagenais, P. A. (1986). Vowel errors in noise and in reverberation by hearing-impaired listeners. *The Journal of the Acoustical Society of America*, 80(3), 741-748.
- Nábělek, A. K., & Donahue, A. M. (1984, February). Perception of consonants in reverberation by native and non-native listeners. *The Journal of the Acoustical Society of America*, 75(2), 632-634.
- Nábělek, A. K., & Letowski, T. (1984). Vowel perception in reverberation. *The Journal of the Acoustical Society of America*, 75(S1), S86-S86.
- Nábělek, A. K., & Letowski, T. R. (1988, May). Similarities of vowels in nonreverberant and reverberant fields. *The Journal of the Acoustical Society of America*, 83(5), 1891-1899.
- Nábělek, A. K., Letowski, T. R., & Tucker, F. M. (1989). Reverberant overlap- and self-masking in consonant identification. *The Journal of the Acoustical Society of America*, 86(4), 1259-1265.
- Nábělek, A. K., & Mason, D. (1981). Effect of noise and reverberation on binaural and monaural word identification by subjects with various audiograms. *Journal of Speech, Language, and Hearing Research*, 24(3), 375-383.
- Nábělek, A. K., & Pickett, J. M. (1974). Monaural and binaural speech-perception through hearing-aids under noise and reverberation with normal and hearing-impaired listeners. *Journal Of Speech And Hearing Research*, 17(4), 724-739.
- Nábělek, A. K., & Robinette, L. (1978). Influence of the precedence effect on word identification by normally hearing and hearing-impaired subjects. *The Journal of the Acoustical Society of America*, 63(1), 187-194.
- Nábělek, A. K., & Robinette, L. (1978). Reverberation as a parameter in clinical testing. *International Journal Of Audiology*, 17(3), 239-259.
- Nábělek, A. K., & Robinson, P. K. (1982). Monaural and binaural speech perception in reverberation for listeners of various ages. *The Journal of the Acoustical Society of America*, 71(5), 1242-1248.
- Nábělek, A. K., & Robinson, P. K. (1983). Consonant errors in noise and reverberation. *The Journal of the Acoustical Society of America*, 73(S1), S103-S103.
- Nakajima, T., & Ando, Y. (1991). Effects of a single reflection with varied horizontal angle and time delay on speech intelligibility. *The Journal of the Acoustical Society of America*(6), 3173-3179.
- Nasreddine, Z. S., Phillips, N. A., Bedirian, V., Charbonneau, S., Whitehead, V., Collin, I., ... Chertkow, H. (2005). The montreal cognitive assessment, MoCA: A brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society*, 53(4), 695-699.
- Neher, T., Breitsprecher, C., & Laugesen, S. (2009). The influence of hearing-aid microphone location and room reverberation on better-ear effects. In *Proceedings of the international symposium on auditory and audiological research* (Vol. 2, pp. 505-512).

- NIDCD. (2018, Apr). *Statistics and Epidemiology*. Retrieved from <https://www.nidcd.nih.gov/health/statistics> ([Online; accessed 11. Apr. 2018])
- Noordhoek, I. M., & Drullman, R. (1997). Effect of reducing temporal intensity modulations on sentence intelligibility. *The Journal of the Acoustical Society of America*, 101(1), 498-502.
- Payton, K. L., & Braida, L. D. (1999). A method to determine the speech transmission index from speech waveforms. *The Journal of the Acoustical Society of America*, 106(6), 3637-3648.
- Payton, K. L., & Shrestha, M. (2013). Comparison of a short-time speech-based intelligibility metric to the speech transmission index and intelligibility data. *The Journal of the Acoustical Society of America*, 134(5), 3818-10.
- Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *The Journal of the Acoustical Society of America*, 24(2), 175-184.
- Picou, E. M., Gordon, J., & Ricketts, T. A. (2016). The effects of noise and reverberation on listening effort in adults with normal hearing. *Ear And Hearing*, 37(1), 1-13.
- Plomp, R. (1988). The negative effect of amplitude compression in multichannel hearing aids in the light of the modulation-transfer function. *The Journal of the Acoustical Society of America*, 83(6), 2322-2327.
- Reed, C. M., Braida, L. D., & Zurek, P. M. (2009). Review of the literature on temporal resolution in listeners with cochlear hearing impairment: A critical assessment of the role of suprathreshold deficits. *Trends in amplification*, 13(1), 4-43.
- Reinhart, P., Zahorik, P., & Souza, P. (2017). The interaction between reverberation and digital noise reduction in hearing aids: Acoustic and behavioral effects. *The Journal of the Acoustical Society of America*, 141(5), 3971-3971.
- Reinhart, P. N., & Souza, P. E. (2018). Listener factors associated with individual susceptibility to reverberation. *Journal Of The American Academy Of Audiology*, 29, 73-82.
- Reinhart, P. N., Souza, P. E., Srinivasan, N. K., & Gallun, F. J. (2016). Effects of reverberation and compression on consonant identification in individuals with hearing impairment. *Ear And Hearing*, 37(2), 144-152.
- Rennies, J., Warzybok, A., Brand, T., & Kollmeier, B. (2014). Modeling the effects of a single reflection on binaural speech intelligibility. *The Journal of the Acoustical Society of America*, 135(3), 1556-1567.
- Resnick, S. B., Dubno, J. R., Hoffnung, S., & Levitt, H. (1975). Phoneme errors on a nonsense syllable test. *The Journal of the Acoustical Society of America*, 58(S1), S114-S114.
- Ricketts, T., & Henry, P. (2002). Evaluation of an adaptive, directional-microphone hearing aid: Evaluación de un auxiliar auditivo de micrófono direccional adaptable. *International Journal of Audiology*, 41(2), 100-112.
- Ricketts, T. A., & Hornsby, B. W. (2003). Distance and reverberation effects on directional benefit. *Ear and Hearing*, 24(6), 472-484.
- Rosen, S. (1992). Temporal information in speech: Acoustic, auditory and linguistic aspects. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 336(1278), 367-373.
- Rothauser, E. (1969). IEEE recommended practice for speech quality measurements. *IEEE Trans. on Audio and Electroacoustics*, 17, 225-246.
- Sabin, A. T., Gallun, F. J., & Souza, P. E. (2013). Acoustical correlates of performance on a dynamic range compression discrimination task. *The Journal of the Acoustical Society of America*, 134(3), 2136-2147.
- Sabine, W. (1964). Reverberation, The American Architect. *Collected Papers on Acoustics*, 3-68.

- Sato, H., Sato, H., & Morimoto, M. (2007). Effects of aging on word intelligibility and listening difficulty in various reverberant fields. *The Journal of the Acoustical Society of America*, 121(5), 2915-8.
- Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science (New York, N.Y.)*, 270(5234), 303-304.
- Shi, L.-F., & Doherty, K. A. (2008). Subjective and objective effects of fast and slow compression on the perception of reverberant speech in listeners with hearing loss. *Journal of Speech, Language, and Hearing Research*, 51(5), 1328-1340.
- Slama, M. C., & Delgutte, B. (2015). Neural coding of sound envelope in reverberant environments. *Journal of Neuroscience*, 35(10), 4452-4468.
- Smith, Z. M., Delgutte, B., & Oxenham, A. J. (2002). Chimaeric sounds reveal dichotomies in auditory perception. *Nature*, 416(6876), 87-90.
- Souza, P., Hoover, E., & Gallun, F. (2012). Application of the envelope difference index to spectrally sparse speech. *Journal of Speech, Language, and Hearing Research*, 55(3), 824-837.
- Souza, P. E., & Boike, K. T. (2006). Combining temporal-envelope cues across channels: effects of age and hearing loss. *Journal of Speech, Language, and Hearing Research*, 49(1), 138-149.
- Souza, P. E., Jenstad, L. M., & Folino, R. (2005). Using multichannel wide-dynamic range compression in severely hearing-impaired listeners: Effects on speech recognition and quality. *Ear And Hearing*, 26(2), 120-131.
- Souza, P. E., & Kitch, V. (2001). The contribution of amplitude envelope cues to sentence identification in young and aged listeners. *Ear And Hearing*, 22(2), 112-119.
- Souza, P. E., & Turner, C. W. (1996). Effect of single-channel compression on temporal speech information. *Journal of Speech, Language, and Hearing Research*, 39(5), 901-911.
- Souza, P. E., Wright, R. A., Blackburn, M. C., Tatman, R., & Gallun, F. J. (2015). Individual sensitivity to spectral and temporal cues in listeners with hearing impairment. *Journal of Speech, Language, and Hearing Research*, 58(2), 520-534.
- Srinivasan, N. K., & Zahorik, P. (2014). Enhancement of speech intelligibility in reverberant rooms: Role of amplitude envelope and temporal fine structure. *The Journal of the Acoustical Society of America*, 135(6), EL239-EL245.
- Steeneken, H. J. M., & Houtgast, T. (1980). A physical method for measuring speech-transmission quality. *The Journal of the Acoustical Society of America*, 67(1), 318-326.
- Studebaker, G. A. (1985). A rationalized arcsine transform. *Journal of Speech, Language, and Hearing Research*, 28(3), 455-462.
- Swaminathan, J., & Heinz, M. G. (2012). Psychophysiological analyses demonstrate the importance of neural envelope coding for speech perception in noise. *Journal of Neuroscience*, 32(5), 1747-1756.
- Taal, C. H., Hendriks, R. C., Heusdens, R., & Jensen, J. (2011). An algorithm for intelligibility prediction of time-frequency weighted noisy speech. *IEEE Transactions on Audio, Speech, and Language Processing*, 19(7), 2125-2136.
- Turner, C. W., Chi, S.-L., & Flock, S. (1999). Limiting spectral resolution in speech for listeners with sensorineural hearing loss. *Journal of Speech, Language, and Hearing Research*, 42(4), 773-784.
- Turner, C. W., Souza, P. E., & Forget, L. N. (1995). Use of temporal envelope cues in speech recognition by normal and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, 97(4), 2568-2576.
- Unoki, M., Toi, M., & Akagi, M. (2006). Refinement of an MTF-based speech dereverberation method using an optimal inverse-MTF filter. *Proceedings of SPECOM*, 323-326.

- van Buuren, R. A., Festen, J. M., & Houtgast, T. (1999). Compression and expansion of the temporal envelope: Evaluation of speech intelligibility and sound quality. *The Journal of the Acoustical Society of America*, 105(5), 2903-2913.
- Van Tasell, D. J., Soli, S. D., Kirby, V. M., & Widin, G. P. (1987). Speech waveform envelope cues for consonant recognition. *The Journal of the Acoustical Society of America*, 82(4), 1152-1161.
- Warzybok, A., Rennies, J., Brand, T., Doclo, S., & Kollmeier, B. (2013). Effects of spatial and temporal integration of a single early reflection on speech intelligibility. *The Journal of the Acoustical Society of America*, 133(1), 269-282.
- Watkins, A. J., Raimond, A. P., & Makin, S. J. (2011). Temporal-envelope constancy of speech in rooms and the perceptual weighting of frequency bands. *The Journal of the Acoustical Society of America*, 130(5), 2777-2788.
- Wilson, B. S., & Dorman, M. F. (2008). Cochlear implants: a remarkable past and a brilliant future. *Hearing research*, 242(1), 3-21.
- Won, J. H., Drennan, W. R., & Rubinstein, J. T. (2007). Spectral-ripple resolution correlates with speech reception in noise in cochlear implant users. *Journal of the Association for Research in Otolaryngology*, 8(3), 384-392.
- Wroblewski, M., Lewis, D. E., Valente, D. L., & Stelmachowicz, P. G. (2012). Effects of reverberation on speech recognition in stationary and modulated noise by school-aged children and young adults. *Ear And Hearing*, 33(6), 731-744.
- Xu, L., & Pfingst, B. E. (2008). Spectral and temporal cues for speech recognition: Implications for auditory prostheses. *Hearing Research*, 242(1-2), 132-140.
- Yang, W., & Bradley, J. S. (2009). Effects of room acoustics on the intelligibility of speech in classrooms for young children. *The Journal of the Acoustical Society of America*, 125(2), 922-933.
- Yang, W., & Hodgson, M. (2006). Auralization study of optimum reverberation times for speech intelligibility for normal and hearing-impaired listeners in classrooms with diffuse sound fields. *The Journal of the Acoustical Society of America*, 120(2), 801-807.
- Zahorian, S. A., & Rothenberg, M. (1981). Principal-components analysis for low-redundancy encoding of speech spectra. *The Journal of the Acoustical society of America*, 69(3), 832-845.
- Zahorik, P. (2009). Perceptually relevant parameters for virtual listening simulation of small room acoustics. *The Journal of the Acoustical Society of America*, 126(2), 776-791.