THE ROLE OF TEMPORAL FINE STRUCTURE IN SOUND QUALITY PERCEPTION

by

MELINDA CHURCH ANDERSON

B.A., University of Florida, 1999

M.S., Vanderbilt University, 2002

A dissertation submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirement for the degree of

Doctor of Philosophy

Department of Speech, Language, and Hearing Sciences

2010

This dissertation entitled:

The Role of Temporal Fine Structure in Sound Quality Perception

written by Melinda Church Anderson

has been approved for the Department of Speech, Language, and Hearing Science

_____

Kathryn H. Arehart, PhD: committee chair

_____

James M. Kates: committee member

Date: _____

The final copy of this dissertation 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 #: 0310.7

Anderson, Melinda Church (Ph.D., Speech, Language, and Hearing Sciences)

The Role of Temporal Fine Structure in Sound Quality Perception

Dissertation directed by Associate Professor Kathryn H. Arehart

Speech perception depends on access to spectral and temporal acoustic cues. Temporal cues include slowly-varying amplitude changes (temporal envelope) and higher-rate amplitude changes (temporal fine structure, TFS). This study sought to quantify the effects of alterations to temporal structure on the perception of speech quality by parametrically varying the amount of TFS available in specific frequency regions.

The three research aims were to: 1) establish the role of TFS in quality perception, 2) determine if the role of TFS in quality perception differs for listeners with normal hearing and listeners with hearing loss, and 3) quantify the relationship between intelligibility scores and quality ratings. Quality ratings were obtained using an 11-point scale for three signal processing types (two types of vocoding noise and total band removal), and with different amounts of background noise (none, 18, and 12 dB signal-to-noise ratios (SNRs)) for a range of frequency regions.

TFS removal above 1500 Hz had a small, but measurable, effect on quality ratings for speech in quiet (i.e. a 2.2-point drop on an 11-point scale). For speech in noise, TFS removal had a smaller effect (at most a 1.2-point drop). TFS modifications also influenced the temporal envelope. Analyses using the Hearing Aid Speech Quality Index (HASQI) (Kates & Arehart, 2010) showed that temporal envelope modifications provide a partial, though incomplete,

description of sound quality degradation. Thus, TFS is important to consider in models of quality perception of speech.

Intelligibility performance was correlated with quality ratings, with larger correlations evident for poorer intelligibility. However, a significant relationship between intelligibility and quality was documented even when intelligibility remained above 95%.

The results of this study have both scientific and clinical implications. The findings provide insight into the mechanisms that affect sound quality perception, including the role of TFS. Additionally, this knowledge may be applied to future signal processing design, helping to maximize both speech intelligibility and sound quality in new hearing aids.

Dedicated in loving memory to my father, who always pushed me to excel,

and in grateful honor of my mother, who never doubted that I could.

Acknowledgements:

I would like to extend my gratitude to my doctoral committee: Gail Ramsberger, Phillip Gilley, Lew Harvey, Pamela Souza, and Christopher Long. I am fortunate to have such strong leaders guiding me.

My mentors Kathy Arehart and Jim Kates deserve a special nod of acknowledgement. They have spent countless hours steering me, teaching me, and occasionally forcing me to find my feet as a researcher. I hope I've made them proud.

My listeners, who contributed their time to participate in this project. I hope they had as much fun as I did. It is with a sad heart that I say a special thank-you to listener I10, who has been a participant in studies in this lab for as long as I have been a student here. He was a wonderful person, and his passing is much grieved.

My colleagues at University of Colorado Hospital have been a valuable source of reality these past several months. They have shown me that there is life after graduate school.

My research lab partners: Naomi Croghan and Ramesh Kumar Muralimanohar. I wouldn't have made it through this without you.

Jessica Rossi-Katz and Kristin Uhler, two of my favorite women, have helped me keep my eye on the prize. We've all made it!

My sister, Katie Church. Life is better with you in Colorado. It will be better still now that we'll have time to play.

And most of all, my family. Ryan, I know at times it seemed this journey might never end. We did it. This belongs to you, too. Hazel, you are the light in my day and the stars in my night. Grizz, you have been my constant companion over much of this work. Thank you for staying put until it got finished. I can't wait to meet you.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

Many of the 31.5 million Americans with hearing loss are candidates for hearing aids (Kochkin, 2005a). While recent clinical trials document the benefit of hearing aids (e.g. Larson, et al., 2000), only 20-40% of individuals who are candidates actually own them (Kochkin, 2005a; Dubno et al., 2008). Approximately 65-75% of those who wear hearing aids are satisfied with their instruments (Kochkin, 2005a; Dubno, et al., 2008). Numerous factors contribute to the lack of satisfaction, some related to speech intelligibility and sound quality (Kochkin 2005a, b). Research has shown that alterations to the signal, through noise, nonlinear and linear signal processing, can affect both speech intelligibility and speech quality (e.g., Moore & Tan, 2004; Davies-Venn et al., 2007; Arehart et al., 2007; Anderson et al., 2009; Arehart et al., 2010). These types of processing can take the form of modifications to the spectral domain and the temporal domain. Temporal information, as used in this dissertation, refers both to temporal envelope (slowly-varying amplitude changes over time) and temporal fine structure (higher-rate amplitude changes over time with rates close to the center frequency of the band or auditory filter). While maintaining high levels of speech intelligibility is important for user satisfaction with hearing aids, it is possible to have high levels of subjective intelligibility with poor sound quality (e.g. Preminger & Van Tasell, 1995a). The purpose of this study is to quantify the effects of alterations to the temporal structure of speech on the perception of its quality. The research focus is to determine how removal of temporal fine structure in specific frequency regions affects sound quality in situations where speech intelligibility remains at high levels.

Temporal envelope has been shown to be a salient cue in sound quality perception. Greater amounts of temporal envelope degradation lead to greater reductions in predicted and subjective quality ratings using models of sound quality (e.g., Huber & Kollmeier, 2006; Stone

& Moore, 2007; Arehart et al., 2007; Anderson et al., 2009; Kates & Arehart, 2010). Many signal manipulations used in hearing aids take the form of nonlinear digital signal processing (e.g. dynamic range compression and spectral subtraction). These signal manipulations create alterations to the temporal envelope of speech, and may have unintended perceptual consequences. For example, as the amount of dynamic range compression increases, the effects on the temporal envelope increase. These changes in temporal envelope are associated with decreased sound quality ratings (e.g. van Buuren et al., 1999). However, little literature exists regarding the effects of temporal fine structure modifications on sound quality perception.

In contrast to the limited number of studies which address the role of temporal fine structure in sound quality perception, a significant number of recent studies have examined how temporal fine structure is used in speech intelligibility. For a single talker in quiet, speech with no temporal fine structure is highly intelligible for listeners with normal hearing and for listeners with mild to moderate hearing loss due to cochlear damage (e.g., Shannon et al., 1995; Başkent, 2006). However, when listening to speech in the presence of competition, temporal fine structure plays a more important role (e.g., Qin & Oxenham, 2003; Lorenzi et al., 2006; Başkent 2006; Hopkins et al., 2008; Hopkins & Moore, 2009). When speech is presented with a competing sound, temporal envelope only cues are insufficient for high speech intelligibility for both listeners with normal hearing and listeners with hearing loss. Listeners with normal hearing achieve better speech understanding in noise from inclusion of temporal fine structure only up to about 5000 Hz, while listeners with hearing loss benefit from inclusion of temporal fine structure up to about 1500 Hz (Hopkins et al., 2008). This decreased ability to use temporal fine structure in listeners with hearing loss has been attributed, in part, to broader auditory filters associated with cochlear hearing loss (Hopkins et al., 2008).

This study has three research aims. The first aim is to quantify the role of temporal fine structure in sound quality perception. The second aim is to determine if the role of temporal fine structure differs for listeners with normal hearing and listeners with hearing loss. The third aim is to objectively quantify the relationship between intelligibility and quality. As in studies examining the role of temporal fine structure in speech intelligibility (e.g. Hopkins et al., 2008), this study parametrically varies the amount of temporal fine structure in different frequency regions. Given the increased importance for temporal fine structure for speech intelligibility when speech is in the presence of competition, quality ratings for speech are obtained in quiet, and at 18 and 12 dB signal-to-noise ratios (SNRs). Some previous work has shown that quality ratings are largely determined by subjective intelligibility when subjective intelligibility is poor (Preminger & Van Tasell, 1995a). This study provides objective data examining the relationship between intelligibility and quality for a range of speech conditions.

The purpose of this study is to quantify the effects of alterations to the temporal structure of speech on the perception of its quality. The research focus is to determine how removal of temporal fine structure in specific frequency regions affects sound quality in situations where speech intelligibility remains at high levels. Based on data from the speech intelligibility literature, we know that the type of vocoding noise, the amount and type of background noise, and the frequency region of the vocoded speech all affect listener perception. Because of differential ability to use temporal fine structure in listeners with normal hearing and listeners with hearing loss, quality perception based on temporal fine structure removal may also differ between listeners with normal hearing and listeners with hearing loss.

Studies which examine the role of temporal fine structure in intelligibility provide a framework for studying factors that may be important in the study of quality perception for both

listeners with normal hearing and listeners with sensorineural hearing loss. In this project, several factors identified to be important in intelligibility are manipulated in order to increase our understanding of the role temporal fine structure plays in quality perception. Specifically, this study uses three types of signal processing, including two types of vocoding noise (a noise-envelope-intact and a noise-envelope-removed) and removal of high-frequency bands of the signal. The vocoding noise is designed to remove the temporal fine structure and keep the temporal envelope intact. We also manipulate the frequency region of the altered signal through changes to the band cutoff, such that as the band cutoff decreases, the amount of the signal that is vocoded or removed is increased. A final factor is the amount of background noise. Because background noise increases the importance of temporal fine structure to speech intelligibility, three levels of background noise are included in this study.

The results of this study have both scientific and clinical implications. First, the results provide insight into the mechanisms underlying sound quality perception. In addition to previously existing documentation regarding the importance of the temporal envelope to sound quality perception, the results provide objective data supporting the role of temporal fine structure in quality perception. Furthermore, the findings provide insight into how best to improve signal processing design for hearing aid applications. Digital signal processing in hearing aids has the goal of maximizing speech intelligibility and sound quality. Integrating these results with the existing speech intelligibility literature may allow for improved hearing aid signal processing design.

In addition to providing knowledge regarding the relative importance of the temporal envelope and temporal fine structure in sound quality perception, we also provide evidence to support objective modeling of sound quality. As described above, many quality metrics

currently employ measures of temporal envelope degradation. The results from this study may allow the addition of a component which also examines temporal fine structure degradation, enhancing the effectiveness of objective models of quality perception. Given that a significant amount of patient satisfaction with hearing aids is related to the sound quality of the device, facilitating an increase in our understanding of sound quality perception may enhance our ability to improve current signal processing strategies.

CHAPTER 2: BACKGROUND AND SIGNIFICANCE

The purpose of this study is to quantify the effects of alterations to the temporal structure of speech on the perception of its quality. The research focus is to determine how removal of temporal fine structure in specific frequency regions affects sound quality in situations where speech intelligibility remains at high levels. This chapter provides a framework for understanding how alterations to the temporal fine structure of a speech sample might affect a listener's perception of the quality of that speech sample.

Sound Quality Perception

In order to effectively study the perception of sound quality, it is important to understand how a listener makes decisions about sound quality. A typical hearing aid wearer might report that a particular signal processing strategy sounds "shrill" or "mumbled" or "loud". The number of adjectives used by a listener suggests that more than one factor affects sound quality perception.

Several researchers have published reports regarding the multi-dimensional nature of sound quality perception (e.g. Gabrielsson & Sjogran, 1979; Gabrielsson et al., 1988; Gabrielsson et al., 1990; Preminger & Van Tasell, 1995a, b; Neuman et al., 1998; Arehart et al., 2007). Gabrielsson and colleagues have shown that listeners use multiple factors to judge the sound quality of an acoustic sample. The number of specific dimensions depends on the nature of the signal, the sound source of the reproduction, and the hearing status of the listener. For example, in Gabrielsson and Sjogren (1979), subjects were given a list of 200 adjectives and asked to rate the relevance of each adjective to multiple sound reproductions using three sources: loudspeakers, headphones, and hearing aids. The 200 adjectives were consolidated into eight

distinct dimensions, with the number of dimensions per sound source varying between two and five. The number of dimensions was dependent upon the type of processing conducted with each reproduction system. Overall, the eight dimensions were separated into facets representing clearness/distinctness, sharpness/hardness-softness, brightness-darkness, fullness-thinness, feeling of space, nearness, disturbing sounds, and loudness**.** Although multiple dimensions may be required to fully describe a particular type of processing, accurate measurement of sound quality perception is possible with a single dimension (e.g., van Buuren et al., 1999; Moore & Tan, 2004; Arehart et al., 2010, in press).

Another factor which may affect sound quality perception is intelligibility. Preminger and Van Tassel (1995a) found that when subjective intelligibility was allowed to vary from poor to good, individual quality dimensions of clarity, pleasantness, listening effort, loudness, and total impression were not significantly different from subjective intelligibility or from each other. In other words, a high degree of correlation exists between each dimension such that when subjective intelligibility was varied, there was a consistent variation between subjective intelligibility and each of the other dimensions. In contrast, when subjective impressions of intelligibility remained high, significant differences in quality ratings among the individual dimensions were found. These results indicate that when subjective intelligibility is poor, listeners do not identify multiple unique dimensions of quality, and subjective impressions of intelligibility drive quality ratings. In contrast, when listeners rate intelligibility as high, they are able to perceive and accurately rate variations in sound quality perception.

For a speech signal with high intelligibility, listeners may still have decreased sound quality, particularly for speech in background noise and/or under different types of hearing aid signal processing (e.g. Preminger & Van Tasell, 1995a; Gabrielsson et al., 1990; Anderson et al.

2009). Given this relationship between speech intelligibility and speech quality, it may be possible to develop a framework for determining which types of signal manipulations will affect speech quality. If intelligibility has been degraded, sound quality may be adversely affected. However, in those instances where signal manipulations do not harm intelligibility, degradations to sound quality may still occur. One factor that may alter the quality of a signal without harming intelligibility is manipulation of the temporal fine structure.

Temporal Structure

An acoustic signal takes place in two domains: the frequency (spectral) domain and the time (temporal) domain. When a signal is described in terms of the temporal domain, an attempt is made to describe the overall temporal structure of the signal. The temporal structure of an acoustic speech signal is commonly divided into two parts: the slowly varying temporal envelope which is superimposed upon the quickly varying temporal fine structure. The temporal envelope, also known as the amplitude envelope, provides information about the intensity and duration of a signal (Rosen, 1992). Temporal fine structure includes information about voicing and manner of articulation for speech sounds, intonation and stress (Rosen, 1992). In this dissertation, the temporal envelope is defined as frequency information equal to or below 300 Hz, with temporal fine structure including all frequency content above 300 Hz. Because the temporal envelope range is extended to 300 Hz, the temporal envelope will contain information related to the fundamental frequency (F0) and lowest harmonic(s) of most talkers.

Temporal Structure in Models of Sound Quality

The relative importance of the temporal structure to sound quality perception has not been explicitly studied. However, objective models of quality perception, which include estimates of temporal envelope degradation, are reasonably accurate at predicting subjective ratings of sound quality for listeners with normal hearing and listeners with hearing loss.

Current models of sound quality have been shown to be accurate in predicting speech quality ratings for a range of noise, nonlinear, and linear processing conditions. Several of these models of quality are envelope based (e.g. Hansen & Kollmeier, 1997; Huber & Kollmeier, 2006; Stone & Moore, 2007; Kates & Arehart, 2010). Typically, these models function by calculating the difference in envelopes between a clean reference signal and degraded output signal, and correlating the difference to subjective measures of sound quality. Though each model uses a different mathematical basis, the overall implication is that changes to the temporal envelope are correlated with changes in sound quality perception. For example, the PEMO-Q developed by Huber and Kollmeier (2006) functions by cross-correlating the envelope modulations of the input and output across auditory frequency bands. The Cepstrum Correlational model (Cep Corr) developed by Kates and Arehart (2007) fits mel cepstrum coefficients to the short-time spectra (produced by sampling the envelope across frequency) and then cross-correlates the cepstrum coefficients. Both the PEMO-Q and Cep Corr models ignore the F0 of most speakers, as well as any harmonics. The Hearing Aid Speech Quality Index (HASQI) (Kates & Arehart, 2010) is comprised of a noise and nonlinear term and a linear filtering term. The noise and nonlinear portion of the quality model measures the changes in the envelope time-frequency modulation. The evolution of the spectral shape over time for the processed signal is then compared to that for the clean reference signal. Differences indicate a loss in quality. The linear portion of HASQI measures changes in the long-term internal signal

spectrum. The difference between the spectrum of the processed and clear signals is computed as a function of frequency, and the standard deviation of the difference is one input to the linear quality index. The HASQI model multiplies the noise and nonlinear term by the linear filtering term to produce the final index value.

The use of these models in fitting the same subjective quality judgments supports the idea that utilizing the difference in signal envelope between reference and output is a valid technique (Kates & Arehart, 2010). The modeling studies generally show correlation coefficients above r = 0.8 between actual quality ratings and predicted quality ratings for both judgments by listeners with normal hearing and by listeners with sensorineural hearing loss. Although the use of envelope differences has proven to be accurate, these models do not take into account how temporal fine structure may affect sound quality perception. A methodical examination of temporal fine structure variations will provide an additional feature (temporal fine structure) that may be implemented in objective models of subjective sound quality perception.

In addition to providing evidence to support objective modeling of sound quality, knowledge regarding the relative importance of the temporal envelope and temporal fine structure in sound quality importance may aid in the development of signal processing strategies for hearing aids. Many signal processing strategies have an effect on the temporal structure or audibility of the signal. Given that a significant amount of patient satisfaction with hearing aids is related to the sound quality of the device, facilitating an increase in our understanding of sound quality perception may enhance our ability to improve current signal processing strategies.

Effects of Hearing Aid Signal Processing on Temporal Structure and Sound Quality

Many factors of hearing aid design may affect sound quality. These factors include both linear processing (e.g., Gabrielsson & Sjögren, 1979; Gabrielsson et al., 1990; Gabrielsson et al., 1991; Preminger & Van Tasell, 1995 a, b; Moore & Tan, 2003; Ricketts et al., 2008) and noise and nonlinear processing (e.g., Lawson & Chial, 1982; Crain, 1992; Kates & Kozma-Spytek, 1994; Kozma-Spytek et al., 1996; Stelmachowicz et al,. 1999; Versfeld et al., 1999; Tan & Moore, 2003; Arehart et al., 2007; Davies-Venn et al., 2007; Tan & Moore, 2008). State-of-the-art commercial hearing aids involve complex nonlinear signal processing designs, such as dynamic range compression and noise reduction (Chung, 2004; Kates, 2008). An additional linear design, increased spectral bandwidth beyond the traditional 4-6 kHz, is also beginning to be implemented in commercial devices.

Dynamic range compression is designed to restore the audibility to a range of sounds by providing more amplification for low-level portions of the signal than high-level portions, and to normalize loudness for hearing-impaired listeners (e.g., Souza, 2002; Chung, 2004; Kates, 2008). A physical consequence of dynamic range compression is a reduction of the peak-to-valley ratio of the temporal envelope, thereby smoothing the temporal envelope (e.g. Stone & Moore, 1992, 2007; Jenstad & Souza, 2005). These physical changes to the temporal envelope have been shown to have negative perceptual consequences for some listeners, such as decreased sound quality (e.g. Neuman et al., 1995; Neuman et al., 1998, Lunner et al., 1998; van Buuren et al., 1999; Souza et al., 2005; Moore & Tan, 2008; Ricketts et al., 2008; Anderson et al., 2009). Neuman et al. (1998) found that ratings of clarity, pleasantness, and overall impression decreased as compression ratio increased, while ratings of background noise increased with increasing compression ratio. Anderson et al. (2009) found the sound quality preference was less for

compressed speech in noise when compared to unprocessed speech in noise. However, there appears to be a perceptual boundary where listeners do not report decreased sound quality for compression systems (e.g. Neuman et al., 1994; Shi & Doherty, 2008). For example, Neuman et al., (1994) found that in conditions in which there was background noise, listeners preferred unprocessed speech and a compression ratio of 1.5:1 to higher compression ratios. When the amount of background noise was decreased, listeners did not differentiate between unprocessed speech and compression ratios up to 2:1. Shi and Doherty (2008) found judgments of clarity were similar between compressed and unprocessed speech.

In contrast to the purpose of compression, spectral subtraction has the goal of decreasing the amplitude of low-level portions of the signal that are presumably noise in an attempt to provide a cleaner speech signal. Spectral subtraction involves estimation of the power spectrum of the assumed noise, followed by the subtraction of the estimated noise spectral magnitude from the noisy speech. This process is implemented in multiple channels and will adaptively reduce the gain in each channel in response to the noise estimate (Kates, 2008). Spectral subtraction alters the temporal envelope by increasing the peak-to-valley ratio. Research on sound quality perception for spectral subtraction is mixed (e.g. Arehart et al., 2003; Ricketts & Hornsby 2005; Anderson et al., 2009; Arehart et al., 2010). Arehart et al. (2010) found there was no difference in quality ratings between speech in noise and speech in noise processed with spectral subtraction. In contrast, Anderson et al. (2009) found that listeners preferred speech in noise processed with spectral subtraction compared to speech in noise without spectral subtraction. Although the spectral subtraction routine was the same in both studies, the speech in Anderson et al. (2009) was noisier, perhaps leading to some of the differences.

One additional design receiving interest in the literature is increasing the available spectral bandwidth in hearing aids. Current hearing aids typically provide amplification through 6 kHz. Depending on the degree of hearing loss, many high frequency speech sounds may not be audible. In an attempt to improve audibility of high frequency speech cues, high-frequency amplification may be provided through a hearing aid. However, even with appropriate amplification though the typical available bandwidth, some speech sounds will still not be audible to a hearing-aid user. This lack of audible bandwidth may be detrimental to some listeners, especially for some speech sounds (e.g. Stelmachowicz, et al., 2002; Stelmachowicz et al., 2004; Pittman et al., 2005; Stelmachowicz et al., 2007; Stelmachowicz et al., 2008). However, increased bandwidth frequently results in increased high-frequency gain, which in turn may cause audible feedback (a high-pitched whistling sound). Audible feedback results from an amplified signal leaking back into the hearing aid and getting re-amplified (Dillon, 2001). One of the most common ways to remove audible feedback is to reduce the amount of amplification in the frequency region where the feedback is occurring, limiting the amount of amplification that can be provided. If the signal at the output of the hearing aid is highly correlated with the signal at the input, the feedback cancellation system will often attempt to cancel the input signal rather than model the feedback path. This situation arises, for example, when the input is a single tone or musical note. The inaccurate modeling of the feedback path can then lead to system instability (e.g. “whistling”). Removing the temporal fine structure from the hearing-aid output at high frequencies also removes the correlation between the output and input, greatly reducing the error in modeling the feedback and thus allowing higher gain.

Expanded bandwidth also has conflicting results regarding benefit to speech quality (e.g., Gabrielsson et al., 1990; Moore & Tan, 2004; Ricketts et al., 2008; Arehart et al., 2010).

Listeners with normal hearing consistently rate signals with increased bandwidth higher than signals with decreased bandwidth. Ricketts et al. (2008) found that increasing the low-pass filter (LPF) cutoff from 5.5 kHz to 9 kHz improved speech quality for listeners with normal hearing. Arehart et al. (2010) also reported that speech quality for listeners with normal hearing improved as the LPF cutoff frequency increased from 2 kHz to 7 kHz. In contrast, speech quality rating are affected less by changes in bandwidth in listeners with hearing loss. In Ricketts et al. (2008), only a subset of listeners with hearing loss judged sound quality to be better for increased bandwidth, while in Arehart et al. (2010) average quality ratings for increased bandwidth did not change for listeners with hearing loss. Taken together, these results indicate that additional high-frequency information may improve sound quality perception for listeners with normal hearing and for some, albeit not all, listeners with hearing loss.

Although not all listeners benefit from increased sound quality for each of these signal processing designs, some listeners do achieve improved sound quality perception. However, it is unclear how much of the temporal signal is actually required to achieve this benefit. It is clear that both compression and spectral subtraction affect the temporal envelope. Research from the compression and noise reduction literature indicates that listeners are able to withstand a limited amount of degradation to the temporal envelope before quality perception is affected. These quality reductions can be accurately predicted by objective models of sound quality perception. Additionally, while increasing the spectral bandwidth may improve sound quality, it does not necessarily indicate that the full temporal content must be available to achieve this benefit. However, there is little to no information available regarding the effect of temporal fine structure alterations to sound quality perception. What is available is a significant amount of literature regarding how temporal fine structure affects speech intelligibility. It may be possible to derive

some insight into the role temporal fine structure plays in sound quality perception from the speech intelligibility literature. Determining the limits of perceptual tolerance to temporal fine structure manipulation may be of benefit to hearing aid signal processing development, and provide complementary knowledge to our understanding of temporal envelope effects on sound quality perception.

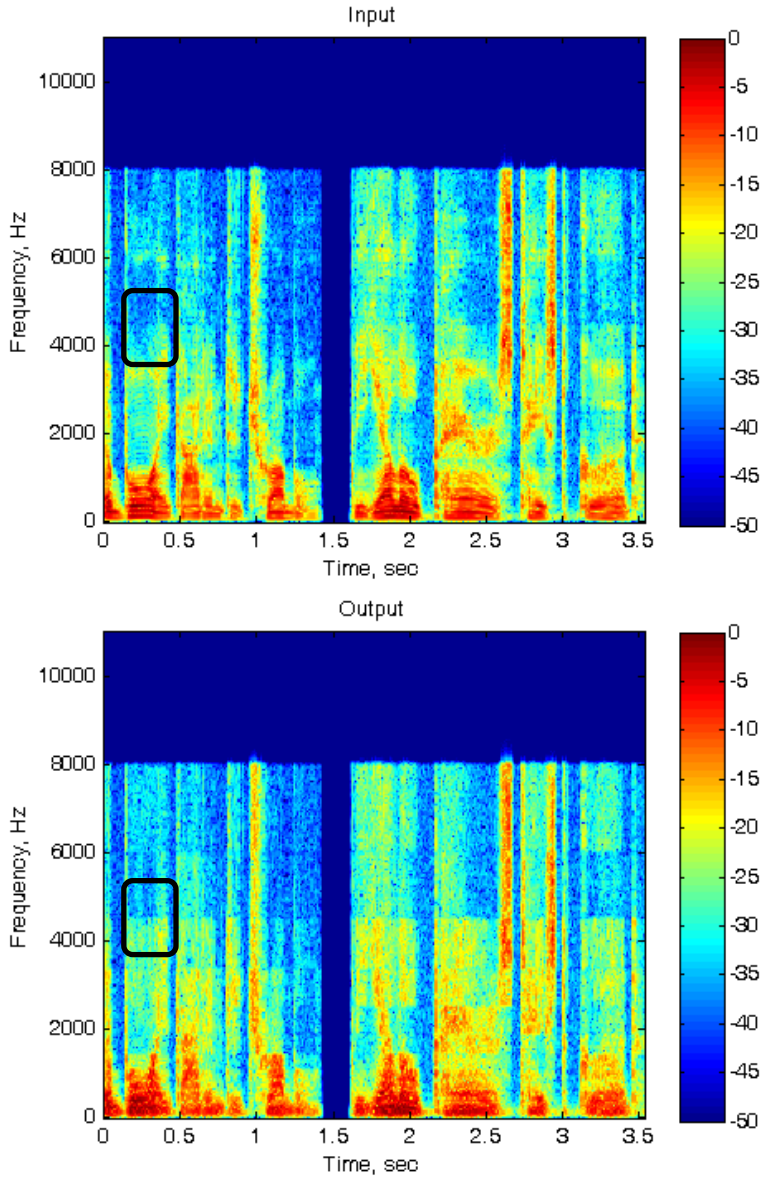
Vocoding

In order to study the separate effects of temporal envelope and temporal fine structure on speech perception, signals are often created which separate the two cues through a process called vocoding (e.g. Dudley, 1939; Shannon, et al., 1995). This technique has been used to study the importance of temporal envelope and temporal fine structure cues to pitch and speech perception. When a signal is vocoded, it is filtered into a specified number of bands, and the envelope of each band is used to modulate a carrier signal (either noise or sine waves). As the number of bands increases more of the temporal envelope structure remains intact. Each band is then re-filtered (using the same filter bank) to remove any out-of-band modulation effects and the bands are recombined. The recombined signal includes the temporal envelope and limited portions of the temporal fine structure (dependent on specific envelope filter cutoff frequencies). The carrier signal used to replace the temporal fine structure can be either a pure tone at the center frequency of the band (sine vocoder) or a Gaussian noise (in a noise vocoder). In sine-vocoded speech, a sine wave at the center frequency of the band is multiplied by the extracted envelope for that band. In noise-vocoded speech, a Gaussian noise is multiplied by the extracted envelope for each band, thereby randomizing the phase of the periodicity and temporal fine structure contained in the band. The Gaussian noise traditionally used in noise vocoders has intrinsic random amplitude fluctuations over time, meaning that at any given point in time, the noise has

its own noise envelope. It has been suggested that this intrinsic noise envelope may have a detrimental impact on speech understanding when combined with the temporal envelope of the speech. Removal of the noise envelope from the Gaussian noise is possible, and both noise-envelope-intact vocoding noise and noise-envelope-removed vocoding noise have been described in the literature (Pumplin, 1995; Kohlrauch et al., 1997; Whitmal et al., 2007; Kates, submitted).

Regardless of the type of carrier used in the vocoding process, a signal processing confound exists (Kates, submitted). Although vocoding is designed to remove temporal fine structure cues, vocoding also affects the temporal envelope because the temporal envelope in each band is forced to have the same amplitude for all of the frequencies within the band, thus removing the spectral ripple that would normally exist across the band. As the number of bands decreases, the amount of degradation to the temporal envelope increases. The spectrum of the vocoded output shows a stepped pattern in the temporal envelope of the final signal. Figure 1 provides a visual representation of this stepped pattern using spectrograms. The set of spectrograms shows the input signal in the top panel (2 sentences spoken by a male talker) and the version reproduced using a 16-channel noise-envelope-intact vocoder (bottom panel). The spectrogram of the noise-vocoded speech has edges in the time and frequency domains because the envelope structure that varies within a frequency band in the input signal is replaced by constant values across the band in the modulated noise signal (the output). For instance, at 0.25 s, there is a clear difference between the input and the output at 4500 Hz. In the input, there is a marked variation across frequency that is missing in the output. Therefore, while it is accurate to say that in a vocoded signal the temporal fine structure has been removed, it is not accurate to say that the complete temporal envelope structure is intact. Even with this limitation, vocoding

Figure 1. A set of spectrograms showing the input signal in the top panel (2 sentences spoken by a male talker) and the version reproduced using the 16-channel vocoding noise (bottom panel). The vocoded noise spectrogram has more obvious edges in the time and frequency domains and the envelope structure that varies within a frequency band in the input signal is replaced by constant values across the band in the vocoded signal.



is still a valuable signal processing tool. Vocoding provides a consistent method of temporal fine structure removal, allowing for experimentation on the effects of temporal fine structure on speech perception.

Temporal Structure in Speech Intelligibility

The effects temporal fine structure removal (through the process of vocoding) on speech intelligibility has been shown to vary based on a number of factors. These factors include, among others, the number of bands, the type of vocoding carrier signal, the frequency region of the temporal fine structure, and the presence or absence of background noise. Additional factors, such as the hearing status or age of the listener, may also play a significant role in the importance of temporal fine structure to speech intelligibility, although the mechanisms responsible for differences in perception may vary between groups. Given the relationship between speech intelligibility and sound quality (e.g. Preminger & van Tasell, 1995a), understanding the effects of temporal fine structure removal on speech intelligibility for these disparate groups of listeners may provide insight for understanding the effects of temporal fine structure removal on speech quality perception.

In one of the earliest studies on temporal fine structure removal, Shannon et al. (1995) tested normal-hearing listeners with speech in quiet that was subjected to noise-envelope-intact vocoding with 1, 2, 3, or 4 bands. As the number of bands was increased from 1 to 4, speech understanding improved for vowels, consonants and sentences. With four bands, listeners showed speech intelligibility that exceeded 80%. This result shows that robust speech intelligibility can be achieved even with the limited cues available in the vocoded speech. Shannon et al. (1995) also investigated the importance of temporal envelope cues by measuring speech intelligibility for vocoded signals in which the envelope-modulation low-pass filter (EM

LPF) cutoff vas varied from 16 Hz to 500 Hz. Listeners showed poorer performance for the intelligibility of vowels and sentences for 16 Hz EM LPF when compared to higher EM LPF cutoff values. Consonant recognition was not affected by changes in the EM LPF. The effects of the EM LPF cutoff indicate that temporal envelope modulation above 16 Hz is important for speech that contains periodic voiced information (vowel sounds), but not for aperiodic voiced sounds (consonants). Limited temporal envelope cues (low EM LPF cutoff) and limited spectral resolution (small number of bands) are all that are needed to convey sufficient information for good speech understanding in quiet for listeners with normal hearing.

Souza and Boike (2006) also varied the number of bands available to listeners using restricted temporal fine structure information, from 1 to 8 bands, as well as an unprocessed (full temporal fine structure) condition using vowel-consonant-vowel (VCV) stimuli. There was not a significant difference in speech understanding for the 1-band or 2-band conditions between the listeners with normal hearing and listeners with hearing loss and performance for both groups improved with increasing amounts of temporal fine structure information. However, listeners with hearing loss did not show as much improvement in speech understanding when additional temporal fine structure was available, compared to the listeners with normal hearing.

In an expansion of the work of Shannon et al. (1995), Başkent (2006) studied the effects of the number of vocoded bands for both listeners with normal hearing and listeners with hearing loss for speech in quiet and speech noise. Specifically, a noise-envelope-intact vocoder was used and the number of spectral bands varied from 2 to 32 bands. In this study, normal hearing listeners had speech understanding scores >60% for vowel and consonant recognition in the 4 band condition in quiet, and reached asymptote (~85% correct) in the 8 band condition. Speech understanding in listeners with sensorineural hearing loss was consistently poorer than that of

normal-hearing listeners in all conditions (with asymptote reached at ~80% correct), though the trends in performance were similar in the two groups. Listeners with hearing loss achieved maximum benefit with the same number of bands (8) to listeners with normal hearing in quiet and low noise conditions (10 dB Signal to Noise Ratio [SNR]), with no further improvement in consonant and vowel recognition when the number of bands was increased. In contrast, in conditions with higher levels of background noise (0 and –5 dB SNR), listeners with hearing loss were not as able as listeners with normal hearing to make use of increased information available when the number of spectral bands was increased. For listeners with hearing loss, maximum benefit was achieved with 8 bands of information for both vowel and consonant recognition in all conditions (with ~50% correct asymptote in the 0 dB SNR conditions), while listeners with normal hearing required 12 bands to reach asymptote for vowel recognition and 10 bands to reach asymptote for consonant recognition (with asymptote reached at ~65% correct). At an SNR of -5 dB, listeners with normal hearing did not reach asymptote until 16 bands. As this study shows, listeners with normal hearing and with sensorineural hearing loss need only limited temporal information to achieve asymptote levels speech understanding when speech is presented in quiet. However, when speech is presented in background noise more information is required to achieve best performance, and even when asymptote is reached, intelligibility scores may not be comparable to scores achieved in quiet.

Examining differences in listeners with normal hearing and listeners with hearing loss, Hopkins and Moore (2007) found that listeners with hearing loss were not able to discriminate differences in complex tones that had the same temporal envelope but different temporal fine structure. The consequences of this disability may come in the form of decreased speech understanding in noise. Hopkins et al. (2008) explored the benefit listeners with normal hearing

and listeners with sensorineural hearing loss receive from inclusion of temporal fine structure in a 32-band speech sample in the presence of a competing babble. Temporal fine structure was removed using a noise-envelope-intact vocoder. In contrast to listeners with normal hearing, listeners with hearing loss received only 4.9 dB of benefit in speech reception threshold (SRT) scores when the stimuli went from no temporal fine structure to full temporal fine structure. Listeners with normal hearing received 15.8 dB of benefit. Listeners with hearing loss showed less benefit than normal-hearing listeners from the addition of temporal fine structure information, especially at mid- and high- frequencies. It is worth noting that there was no additional benefit to increasing the availability of temporal fine structure above 4102 Hz for listeners with normal hearing, while listeners with hearing loss received no benefit from additional temporal fine structure over 1605 Hz. In addition to differences in performance based on the amount of temporal fine structure available, between-group differences increased when the number of spectral bands was increased. Normal-hearing listeners improved when a sine-wave vocoded signal was increased from 16 bands to 32 bands, while listeners with hearing loss did not. These results indicate that listeners with hearing loss are less able to make use of added temporal fine structure information and less able to make use of added spectral cues (thorough additional spectral bands).

The type of vocoding noise also influences intelligibility of speech. For example, Whitmal et al. (2007) reported differences in intelligibility for a vocoding noise which has an intact noise-envelope and a vocoding noise which has had the noise-envelope removed. Listeners performed more poorly on measures of speech intelligibility for vocoded speech processed with a noise-envelope-intact vocoding noise, compared to a noise-envelope-removed vocoding noise. Similarly, Souza and Rosen (2009) speculated that the differences seen

between tone and noise vocoders (with an intact noise envelope) may be due to differences in how the noise envelope affects perception of the speech envelope. If a listener is using the speech envelope to identify individual sounds, then it is possible that altering the speech envelope will affect intelligibility. The greater the increase in speech envelope distortion, the greater the likelihood of decreased speech intelligibility.

Taken together, these results show that several factors influence speech understanding for listeners with normal hearing and listeners with hearing loss. These factors include the number of bands used to process the speech, the frequency region of the temporal fine structure removal, the amount of background noise, and the type of vocoding noise used in the processing. The above results show that for a single talker in quiet, speech with no temporal fine structure is highly intelligible for listeners with normal hearing and for listeners with mild to moderate hearing loss due to cochlear damage (e.g., Shannon et al., 1995; Başkent, 2006). However, when listening to speech in the presence of competition, temporal fine structure plays a more important role (e.g., Qin & Oxenham, 2003; Lorenzi et al., 2006; Başkent 2006; Hopkins et al., 2008; Hopkins & Moore, 2009). When speech is presented in the presence of a competing sound, temporal-envelope-only cues are insufficient for high speech intelligibility for both listeners with normal hearing and listeners with hearing loss. The amount of intelligibility degradation depends on both the type of target speech sample (e.g. individual sounds, words, or sentences) and the amount of background noise. Listeners with normal hearing achieve better speech understanding in noise from inclusion of temporal fine structure only up to about 5000 Hz, while listeners with hearing loss benefit from inclusion of temporal fine structure up to about 1500 Hz for target sentences in milti-talker babble (Hopkins et al., 2008).

As the above research shows, access to and the ability to use, temporal fine structure differs between listeners with normal hearing and listeners with cochlear hearing loss. However, hearing loss may not be the only factor in this decreased ability. Research also shows that older listeners, both with and without clinically significant hearing loss, perform more poorly on speech intelligibility tasks compared to younger listeners with normal hearing. These differences have been attributed, in part, to differences in temporal resolution abilities (Gordan-Salant & Fitzgibbons, 1993, 1999; Pichora-Fuller, 2003).

When the temporal fine structure is removed through vocoding, removal of temporal fine structure cues do not have the same impact on these groups, with the evidence showing that removal of temporal fine structure leads to sharper decreases in speech understanding for younger listeners with normal hearing (e.g., Hopkins et al., 2008; Arehart et al., 2010). For example, Arehart et al. (2010) found that younger listeners were better able than older listeners to use temporal fine structure to discriminate two vowels that were concurrently presented. This differential effect between groups suggests that older and hearing-impaired listeners are less able to make use of temporal fine structure cues.

When specifically considering the differences between younger listeners with normal hearing, older listeners, and listeners with hearing loss, two main factors stand out. Younger listeners with normal hearing show more benefit than both older listeners and listeners with hearing loss when more spectral information is provided by a greater number of bands. Second, the ability to use additional temporal fine structure is consistently different between listeners with normal hearing and listeners with hearing loss. When additional temporal fine structure cues are provided, there is benefit to listeners with normal hearing, while older listeners and listeners with hearing loss are comparatively not able to access these additional cues. While it is

known that aging also has a negative impact on the ability to use temporal fine structure, it may be that the mechanisms responsible for this decrease differ between older listeners and listeners with hearing loss. For this reason, it is important to consider each group individually.

Relationship between frequency selectivity and temporal structure

While the role of temporal fine structure in speech perception is not clearly understood, its assumed utility in speech perception depends on the underlying assumption that accurate neural encoding of resolved and unresolved harmonics is important. Understanding the encoding of temporal fine structure in the auditory pathway may provide some insight into how much temporal fine structure is actually available to a listener for speech perception. There are two primary ways that temporal fine structure is encoded in the auditory system: place coding and temporal coding. Research suggests that these differences in encoding abilities (and therefore available temporal fine structure) may partially explain the differences between listeners with normal hearing and listeners with hearing loss regarding the ability to utilize temporal fine structure (Moore, 2008). These encoding differences may result from physiologic differences in the auditory pathway, such as in the sharpness of the traveling wave, broadness of auditory filters, and the phase-locking abilities of the impaired auditory system.

Encoding of temporal fine structure begins in the cochlea, which functions as a frequency analyzer (Plomp, 1964). Frequency analysis, or frequency selectivity, is the ability of the cochlea to separate an incoming complex signal into its component frequencies. Two physiologic factors enhance the ability of the cochlea to perform this task. The first is the tonotopic organization of the basilar membrane which provides a "place method" for frequency analysis (Plomp, 1964). The second is the compressive nature of a healthy cochlea, which sharpens the

traveling wave of an incoming signal and intensifies low-level sounds (e.g. Oxenham & Bacon, 2003).

Each location along the basilar membrane responds best to a particular frequency, but will respond to other frequencies if the traveling wave is big enough, meaning that each location can be thought of as a bandpass filter. The compressive nature of a healthy cochlea sharpens the peaks of low intensity traveling waves. This effect is not seen in a damaged cochlea, resulting in wider traveling waves (Glasberg & Moore, 1990). Narrow filters will pass single resolved harmonics of a complex tone. Wider auditory filters will pass multiple unresolved harmonics, which interact and form a complex waveform with a period corresponding to that of the fundamental frequency, leading to the recovery of envelope cues (Ghitza, 2001). This process is true for both a healthy and an impaired cochlea. The transitional frequency region between resolved and unresolved harmonics is expected to occur in lower frequency regions in a damaged cochlea with abnormally broad filters (e.g., Bernstein & Oxenham, 2006 a, b; Hopkins et al., 2008). Therefore, listeners with cochlear hearing loss may have less access to resolved harmonics because of the increased width of the auditory filter when compared to a listener with an intact cochlea and normal hearing.

Speech intelligibility research indicates that speech information extracted from low-frequency resolved harmonics is more beneficial than temporal fine structure extracted from high-frequency unresolved harmonics (e.g. Hopkins et al., 2008; Hopkins & Moore, 2010). However, even high frequency temporal fine structure is at least somewhat beneficial for speech intelligibility (Hopkins & Moore 2010), by providing information about frication of consonants (Rosen, 1992) and vowel formant information from higher frequencies (Young & Sachs, 1979).

Listeners with greater access to resolved harmonics may have improved speech understanding compared to listeners with limited access to resolved harmonics.

In addition to wider auditory filters, temporal encoding of the signal may also be affected by cochlear hearing loss. Just as in the cochlea, the auditory nerve is tonotopically organized. Peripheral neural representation of the signal includes phase-locking to individual resolved harmonics, encoded by the timing of the nerve impulse, as well as to the envelope of unresolved harmonics, which are coded by changes in the firing rate over time (Rose, 1967). Neural phase locking has been shown to decrease in high frequency regions in mammals such as cats, especially above 4- 5 kHz (e.g. Joris & Yin, 1992). Because of the wider auditory filters in a damaged cochlea, there is less phase-locking to individual harmonics and increased phase-locking to the envelope of unresolved harmonics. Additionally, some evidence suggests that phase locking may also be decreased in the presence of cochlear hearing loss (Woolf et al., 1981), leading disrupted neural encoding of both resolved and unresolved harmonics, regardless of auditory filter width. However, the contention that phase locking is disrupted by cochlear hearing loss remains controversial (Harrison & Evans, 1979).

These physiologic results are supported by recent work modeling neural encoding of temporal fine structure (Heinz & Sawaminathan, 2008, 2009; Sawaminathan, 2010). Heinz and Sawaminathan (2008) showed that models of auditory nerve firing patterns in listeners with sensorineural hearing loss have poorer across-fiber correlation, due to the wider auditory filters. Decreased correlation may result in degraded spatiotemporal neural response patterns for acoustic temporal envelope and temporal fine structure. The modeling results of Sawaminathan (2010) showed that temporal fine structure in normal hearing listeners is a relatively weakly encoded signal compared to the temporal envelope. However, even the weakly encoded

temporal fine structure was correlated with improved performance on speech in noise intelligibility tasks for listeners with normal hearing (Sawaminathan, 2010). These results are consistent with speech intelligibility research showing that as temporal fine structure is added to a speech signal, speech intelligibility improves, with reduced benefit for listeners with sensorineural hearing loss (e.g., Hopkins et al., 2008; Hopkins & Moore, 2010).

The physiologic literature suggests that listeners with hearing loss are limited by their inability to utilize and discriminate temporal fine structure. Reduced frequency selectivity and/or reduced phase locking may provide a partial explanation for some of the differences seen between listeners with normal hearing and listeners with hearing loss. Differences in the physiology of the auditory system may not allow listeners with hearing loss the access to the temporal fine structure cues that are available to listeners with normal hearing.

Summary

The purpose of this study is to quantify the effects of alterations to the temporal structure of speech on the perception of its quality. The research focus is to determine how removal of temporal fine structure in specific frequency regions affects sound quality in situations where speech intelligibility remains at high levels. Based on data from the speech intelligibility literature, we know that the type of vocoding carrier signal, the amount and type of background noise, and the frequency region of the vocoded speech all affect listener perception. Because of the differential ability to use temporal fine structure in listeners with normal hearing and listeners with hearing loss, quality perception based on temporal fine structure removal may also differ between listeners with normal hearing and listeners with hearing loss.

The literature on the intelligibility of speech with limited-to-no temporal fine structure described above provides a framework for factors which may be important in studying the impact of temporal fine structure on quality perception for both listeners with normal hearing and listeners with sensorineural hearing loss. In this project, several factors identified to be important in intelligibility are manipulated in order to increase our understanding of the role temporal fine structure plays in quality perception. Specifically, this study uses three types of signal processing, including two types of vocoding noise (a noise-envelope-intact and a noise-envelope-removed) and removal of high-frequency bands of the signal. We also manipulate the frequency region of the altered signal through changes to the band cutoff, such that as the band cutoff decreases, the amount of the signal that is vocoded or removed is increased. A final factor is the amount of background noise. Because background noise increases the importance of temporal fine structure to speech intelligibility, three levels of background noise are included in this study.

The results of this study will have both scientific and clinical implications. First, we will increase our understanding of the mechanisms underlying sound quality perception. Second, the results will provide insight into how best to improve signal processing design for hearing aid applications. Digital signal processing in hearing aids has the goal of maximizing speech intelligibility and sound quality. Determining the limits of perceptual tolerance for manipulation to changes to temporal fine structure may benefit signal processing development.

Statement of Purpose

The purpose of this dissertation is to quantify the effects of temporal fine structure removal on speech quality ratings for different stimuli in both a group of listeners with normal hearing and a group of listeners with sensorineural hearing loss. To this end, this study has three specific aims.

Specific Aim 1

The first aim is to establish the relationship between temporal fine structure and speech quality ratings by parametrically varying the amount of temporal fine structure available in the signal of both a male and a female talker. Because background noise can affect the importance of temporal fine structure, the stimuli will be presented in both quiet and background noise.

Research Question 1

Do speech quality ratings vary with removal of the temporal fine structure contained in different frequency regions of the acoustic signal for quiet or noisy speech?

Prediction 1

Removal of temporal fine structure is predicted to affect quality perception differently based on frequency region. It is also predicted that the impact of temporal fine structure removal may be greater in noisy speech compared to quiet speech.

Specific Aim 2

The second aim is to quantify the effect of hearing status on speech quality ratings as the amount of temporal fine structure is varied by frequency region for speech in quiet and speech in noise.

Research Question 2

Does the role of temporal fine structure in speech quality ratings depend on the hearing status of a listener?

Prediction 2

Hearing status is expected to differentially affect speech quality ratings, such that listeners with normal hearing and listeners with hearing loss will show variations in the relative importance temporal fine structure by frequency region. It is expected that listeners with normal hearing will show greater variation is sound quality ratings when temporal fine structure is removed, given their increased sensitivity to temporal fine structure.

Specific Aim 3

The third aim is to establish the relationship between quality ratings and intelligibility scores ratings by objectively measuring both quality and intelligibility for the same conditions in a group of listeners with normal hearing and a group of listeners with hearing loss.

Research Question 3

Are speech quality ratings correlated with intelligibility scores for the same conditions?

Prediction 3

Quality ratings are predicted to vary with intelligibility when intelligibility is poor. Quality ratings are predicted to vary independently of intelligibility scores when intelligibility is high.

CHAPTER 3: METHODS

Research Design

The purpose of this experiment is three-fold. The first goal is to establish the relationship between temporal fine structure and quality perception. The second goal is to determine if this relationship differs between listeners with normal hearing and listeners with sensorineural hearing loss. The third goal is to explore the relationship between speech intelligibility and sound quality ratings. It is known that temporal fine structure plays a role in pitch perception, as well as assisting in speech understanding in complex environments. However, the temporal envelope appears to contribute more to speech understanding than temporal fine structure (e.g. Hopkins et al., 2008). What is unclear is how much these temporal fine structure cues contribute to a listener's perception of speech quality (Specific Aim 1), and whether the contribution is affected by hearing status (Specific Aim 2). Although this study focuses on sound quality perception, the overall experimental design is similar to Hopkins et al. (2008). Specifically, we divide the signal into 32 bands and vocode the signal in individual bands (in groups of two), beginning with the highest two bands first and working progressively downward in frequency. Because Hopkins et al. (2008) showed that listeners with hearing loss are able to utilize temporal fine structure information below 1419 Hz for speech intelligibility, speech vocoding is limited to the frequency region above 1500 Hz.

Specific test conditions are discussed below (see section titled *Test Conditions*). Two types of vocoding noise are used to remove the temporal fine structure from the speech signal (see section titled *Stimulus Generation).* The first vocoding noise uses traditional vocoding techniques, and includes both the speech envelope and the noise envelope (FL noise). The

second vocoding noise type replaces the noise envelope with the speech envelope (SM noise). Control conditions using band removal (REM) are also included. In these conditions, the signal is low pass filtered at the band edge, so the entire portion above the band cutoff is removed. In the case that there is no change to quality ratings due to signal vocoding, the inclusion of control conditions allows us to determine if removal of the entire signal is detrimental. Quality ratings are expected to decrease more rapidly for band removal than for vocoding. However, a lack of difference in quality ratings between an intact signal and a band-removed signal may indicate either an inability to hear the difference (lack of audibility) or it may be that although portions removed through the REM conditions are audible, band removal does not detract from quality perception.

Subjects

This study includes a total of twenty listeners: ten with normal hearing and ten with hearing loss. Listeners underwent an audiometric evaluation at their initial visit. Listeners in the normal-hearing group (NH) were required to have air conduction thresholds 20 dB HL or better at octave frequencies from 250 to 8000 Hz (ANSI, 2004), with all other audiometric tests within normal limits. The NH group included nine women and one man with a mean age of 44.3 (range: 20- 64; σ = 16.2). Listeners in the hearing-impaired group (HI) were required to have at least a mild sensorineural hearing loss that would be compatible with a hearing aid fitting, with other audiometric test results consistent with a sensorineural hearing loss. Specifically, the air-bone gap was less than or equal to 10 dB and acoustic reflexes were consistent with the degree of hearing loss. The HI group included three women and seven men with a mean age of 67.4 (range: 47-81; σ =11.3). The better hearing ear was used for testing, with the default of the right ear for the NH group in cases where both ears met the criteria. All listeners were consented after

audiometric testing, but before any experimental procedure. All participants were recruited from the Boulder/Denver metro area and were native speakers of American English. For the experimental listening tasks, subjects were tested monaurally and individually in a double-walled sound-treated booth. Table 1 shows the age, sex, and air conduction thresholds for the test ear for all subjects. Table 1 also shows results from the Mini Mental State Exam (MMSE) (Flostein et al., 1975) and the Quick Speech in Noise Test (QuickSIN) (Killion et al., 2004). Table 2 shows the test results for the TFS1 test (Moore & Sek, 2009) (calculated as a d' value) for F0s of 200, 300, and 400 Hz. These tests are described in more detail in the following sections.

As shown in Figure 2, there is noted variability in the audiometric thresholds of the individual listeners in the HI group. Listeners of variable hearing acuity were purposefully recruited in order to explore the contribution of hearing loss on the role of temporal fine structure on sound quality perception. Although there is a positive relationship between age and amount of hearing loss, this relationship is not statistically significant (Pearson correlation coefficient: $r = 0.1006$, $p = 0.782$).

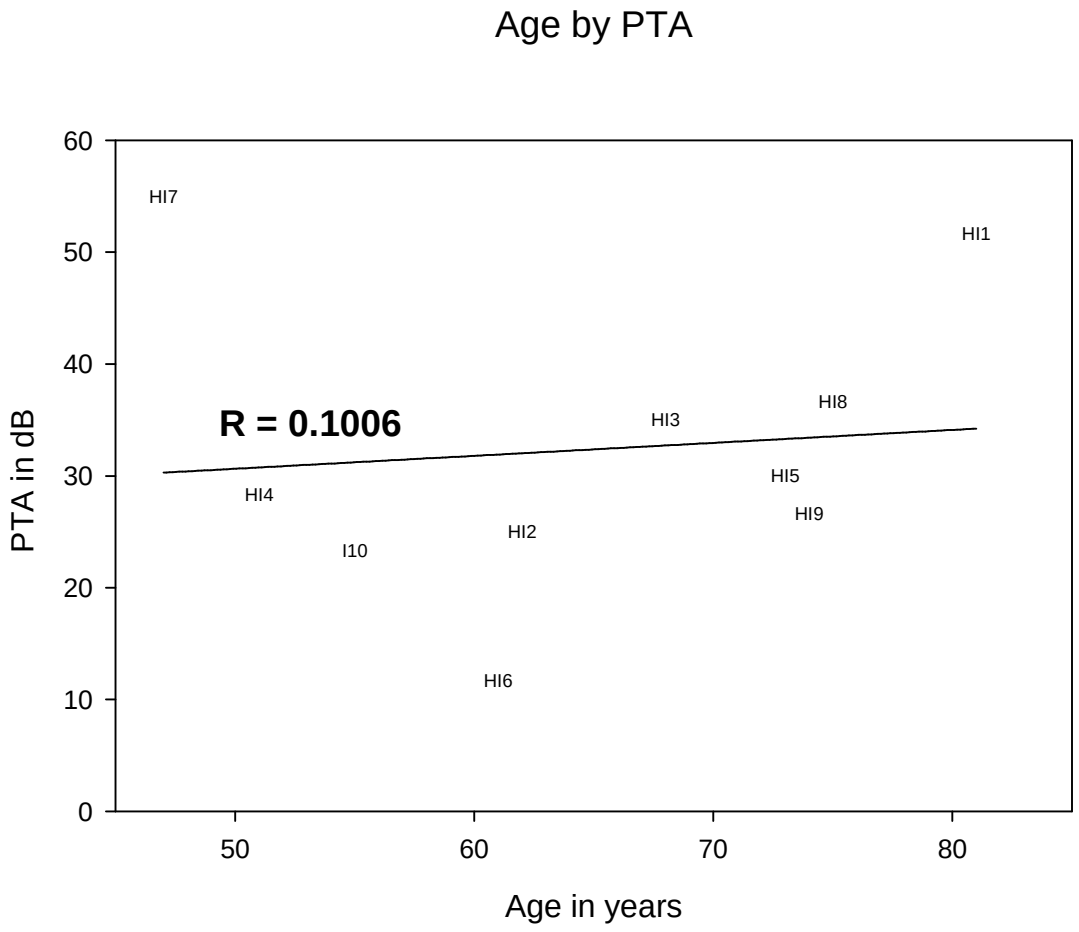
Table 1. Individual subject information. HA user = hearing aid user, TE = test ear, PTA = pure tone average (average of 0.5, 1, and 2 kHz), HFPTA = high frequency pure tone average (average of 1, 2, and 4 kHz).

Sub ID	sex	age	HA user	TE	QuickSIN TE	MMSE	PTA	HFPTA	250	500	1000	2000	3000	4000	6000	8000
HI1	F	81	Y	R	6.5	27	52	55	40	45	55	55	45	55	50	70
HI2	M	62	Y	L	2	30	25	38	15	15	20	40	55	55	40	15
HI3	M	68	N	L	3	29	35	47	35	35	35	35	60	70	60	65
HI4	M	51	N	R	1	30	28	33	20	25	25	35	40	40	45	40
HI5	M	73	Y	R	8.5	28	30	43	35	25	15	50	55	65	55	65
HI6	M	61	N	R	3	30	12	22	5	10	15	10	20	40	30	35
HI7	F	47	Y	L	DNT	30	55	68	20	30	60	75	65	70	65	55
HI8	M	75	N	R	8.5	30	37	48	25	30	30	50	60	65	65	65
HI9	F	74	N	L	0.5	28	27	30	20	25	20	35	35	35	35	40
I10	M	55	Y	R	1.5	30	23	35	25	15	25	30	30	50	45	50
NH1	F	51	N	L	2	30	10	7	5	15	5	10	5	5	5	5
NH2	F	56	N	R	3.5	30	5	5	5	5	5	5	0	5	0	0
NH3	F	57	N	R	0.5	30	12	10	20	20	5	10	10	15	10	10
NH4	F	60	N	R	2	30	8	10	5	5	10	10	10	10	10	10
NH5	F	64	N	R	-0.5	30	18	18	20	20	15	20	20	20	20	10
NH6	M	39	N	R	2.5	29	12	12	15	10	15	10	10	10	5	0
NH7	F	21	N	R	0.5	30	13	13	5	10	10	20	15	10	0	0
NH8	F	46	N	R	1	30	5	5	5	5	5	5	5	5	15	20
NH9	F	20	N	R	1.5	27	13	10	20	15	15	10	10	5	10	0
N10	F	29	N	R	0.5	29	5	7	0	5	10	0	0	10	10	15

Table 2. TFS1 results for individual subjects. Scores are reported in d'. Scores $\geq$ 0.78 indicate the listener scored above chance on the task. CNT = could not test (listener was not able to complete task due to playout intensity limitations).

Subject	200 Hz d'	300 Hz d'	400 Hz d'
NH1	3.93	1.83	3.89
NH2	5.53	1.43	6.73
NH3	1.59	1.50	2.05
NH4	0.44	1.27	1.71
NH5	0.46	0.22	0.04
NH6	2.60	1.44	0.13
NH7	4.97	0.40	4.32
NH8	4.80	2.81	6.09
NH9	10.92	6.20	10.40
N10	3.09	2.39	6.93
HI1	0.10	CNT	CNT
HI2	0.05	0.34	0.04
HI3	2.65	CNT	CNT
HI4	0.00	0.39	0.38
HI5	CNT	CNT	CNT
HI6	3.25	3.76	2.94
HI7	CNT	CNT	CNT
HI8	-0.14	CNT	CNT
HI9	-0.20	0.01	0.10
I10	0.90	2.30	CNT

Figure 2. Age (horizontal axis) and pure tone average (PTA; vertical axis) . Although there is a positive relationship between age and amount of hearing loss, this relationship is not statistically significant (p = 0.782).



Stimulus Generation

The primary goal of this study is to examine the role of temporal fine structure in sound quality perception. The temporal fine structure is separated from the temporal envelope using standard accepted vocoding techniques in the literature. Three general signal processing methods are used to manipulate the signals. The first two methods involve noise-excited vocoders. The third method involves removal (filtering) of specific high-frequency portions of the signal.

The two vocoding noise types (fluctuating and smooth) are based on speech vocoding techniques. These two types of vocoding noise allow for examination of the impact an intact noise envelope has on quality perception. An intact noise envelope introduces unrelated modulations to the speech envelope, and may causing interference with perception of the speech envelope. In the fluctuating (FL) vocoding noise, the noise envelope remains intact and is combined with the speech envelope (Figure 3a). In the smooth (SM) vocoding noise, the noise envelope is removed before being combined with the speech envelope (Figure 3b). The amount of vocoding in a speech sample is determined by the band cutoff (BC). Bands above the BC are vocoded, while bands below, and including, the BC maintain their original fine structure. Specific signal processing techniques are discussed below, including total signal removal (REM) and the addition of background babble.

Figures 3a and 3b depict the order of processing for both the FL and SM vocoding procedures. The speech sample is first combined with the background babble at the appropriate SNR. The speech sample is then passed through a bank of 32 band-pass, linear-phase finite impulse response (FIR) filters. The band edges and center frequencies of the 32 bands are listed in Table 3 and are based on a standard equivalent rectangular bandwidth (ERB) (Slaney, 1993). In vocoded bands, the signal envelope is generated via the Hilbert transform. The signal in each

band is then divided by the envelope to give the fine structure. In instances where the envelope hits zero (which potentially would return an output with no meaning), MATLAB provides a division output of zero. A linear-phase envelope-modulation low-pass filter (300 Hz) is used to filter the speech envelope in each vocoded band. The speech is then reconstructed with the original fine structure and filtered envelopes. The Gaussian noise used for the noise vocoding is passed through the same linear-phase FIR filer bank as the speech. One of two things is then done to the noise: 1) the noise is multiplied by the speech envelope (fluctuating; FL; Fig. 3a), or 2) the noise envelope is removed by dividing the noise signal in the frequency band by its envelope after which it is multiplied by the speech envelope (smooth; SM; Fig. 3b). Both the speech and noise are then passed through the same filters as in the first filtering stage to remove any out-of-band modulation products. The RMS level of the noise and speech in each frequency band is equalized separately to the RMS level of the original input speech in the corresponding band. The band removed conditions (REM) are created using the same 32 band-pass linear phase FIR filer. The amplitude of the signal in the bands above the BC is then set to zero.

Figure 3a. Stimulus generation block diagram: The additive noise is modulated by the filtered speech envelope (fluctuating noise; FL). The background babble is added to the speech at the appropriate SNR before processing begins. The speech line refers to the speech + background babble. The noise line refers to the vocoding noise.

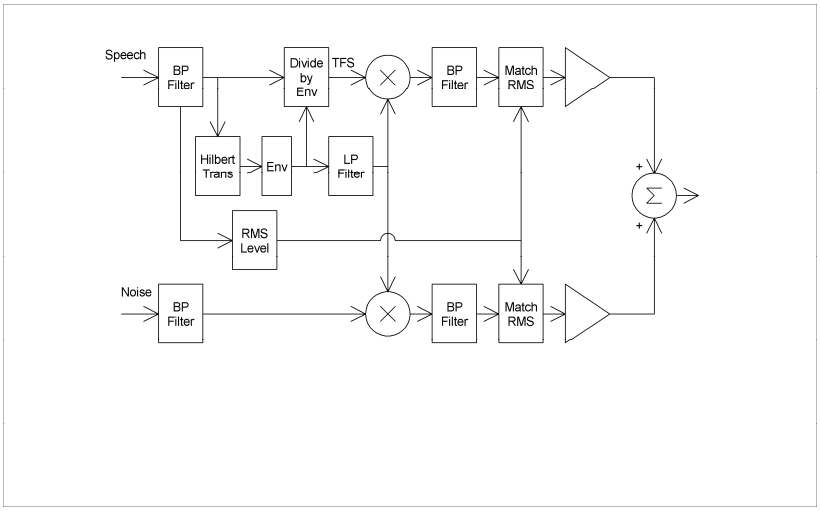


Figure 3b. Stimulus generation block diagram: The noise envelope is replaced by the filtered speech envelope (smooth noise; SM) and then added to the speech. The background babble is added to the speech at the appropriate SNR before processing begins. The speech line refers to the speech + background babble. The noise line refers to the vocoding noise.

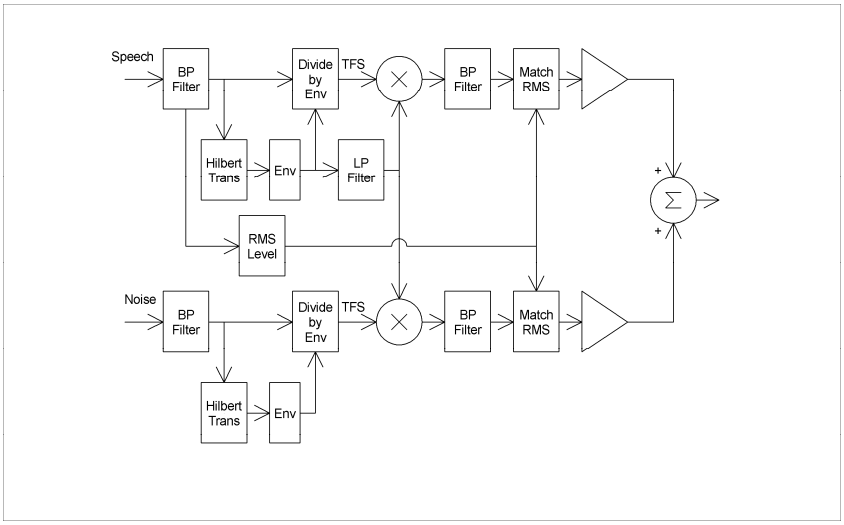


Table 3. Cutoff frequencies for band edges. CF = center frequency of band.

Band	CF	Lower Edge	Upper Edge
1	118	100	137
2	157	138	179
3	201	180	225
4	250	226	277
5	304	278	334
6	365	335	398
7	432	399	469
8	507	470	548
9	591	549	636
10	684	637	734
11	787	735	843
12	902	844	965
13	1030	966	1100
14	1173	1101	1251
15	1332	1252	1418
16	1509	1419	1605
17	1706	1606	1813
18	1926	1814	2045
19	2170	2046	2303
20	2442	2304	2590
21	2745	2591	2909
22	3082	2910	3265
23	3458	3266	3661
24	3876	3662	4102
25	4341	4103	4594
26	4860	4595	5140
27	5437	5141	5749
28	6079	5750	6427
29	6795	6428	7182
30	7591	7183	8022
31	8478	8023	8958
32	9465	8960	10000

Stimuli

The test materials for sound quality and speech intelligibility are the sentences from the IEEE corpus (Rosenthal, 1969). All stimuli are digitized at 44.1 kHz and are down-sampled to 22.05 kHz to reduce computation time. The vocoding is implemented with a customized MATLAB routine (see *Stimulus Generation*). The two sentences in the quality portion were "A saw is a tool used for making boards." and "Take the winding path to the lake." These two sentences cover a broad range of sounds typical in American English.

The IEEE sentences are spoken by a male talker and by a female talker. These same two talkers are utilized in both the quality and the intelligibility potions of the experiment. The difference between talkers in terms of the long-term average speech spectrum for the quality sentences is shown by Figure 4. Speech from the male talker contains more energy in the low-frequencies (below 200 Hz). Speech from the female talker contains more energy in the mid-frequencies (500-3000 Hz). The average fundamental frequency (F0) for the male talker is 114 Hz with a range of 75 – 161 Hz. The average F0 for the female talker is 237 Hz with a range of 143 – 383 Hz. Formant regions for the male and female talker also differ, such that the male talker's formant regions were lower in frequency (Figure 5a,b).

A multi-talker babble is the background noise of choice. It is a common type of background noise encountered by listeners in everyday life. The multi-talker babble background noise is the six-talker babble background from the Connected Speech Test (CST) (Cox et al., 1988). The duration of the babble is matched to the duration of each sentence, and pauses were inserted in the babble to duplicate the pauses between the sentences. The sentences and babble are gated on and off using 5-msec raised cosine windows. The overall RMS of the speech samples is kept a constant level, so in conditions where background noise is added there is a slight decrease in target speech level.

Figure 4. Long term average speech spectrum of male and female talker. The male talker has more energy in the low-frequencies (below 200 Hz). The female talker has more energy in the mid-frequencies (500-3000 Hz).

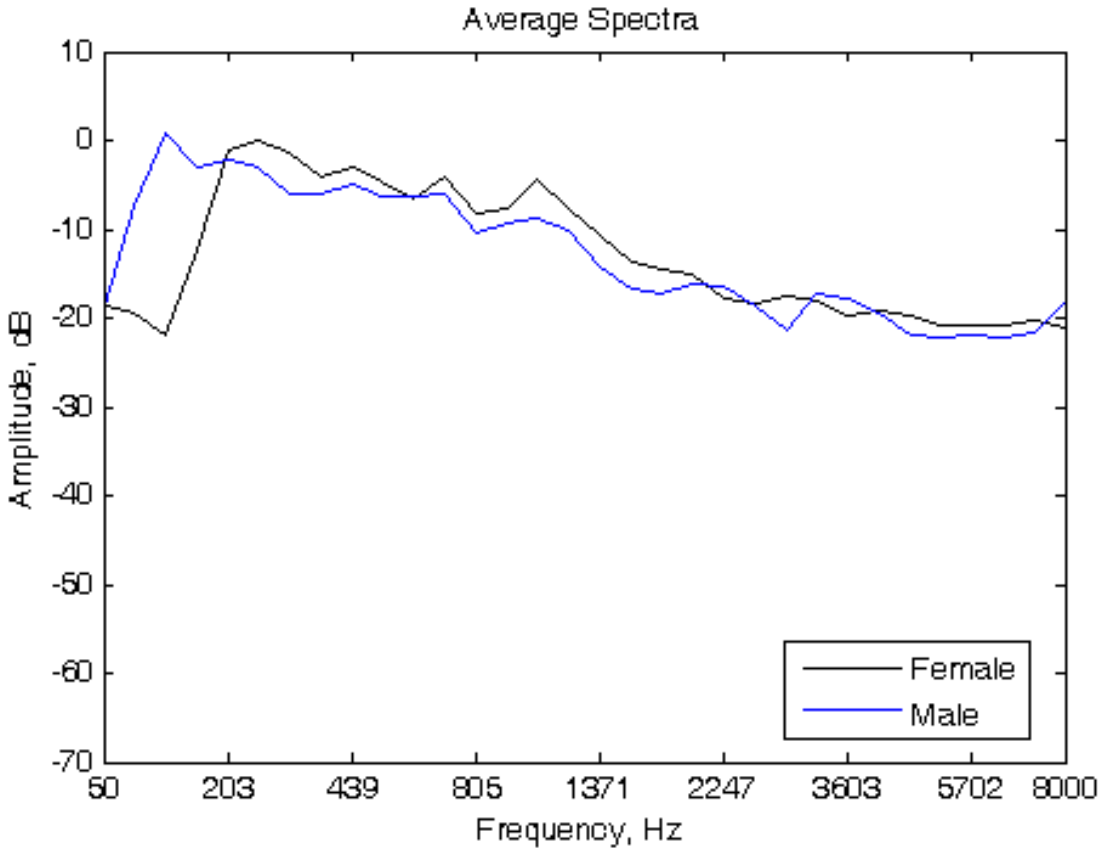
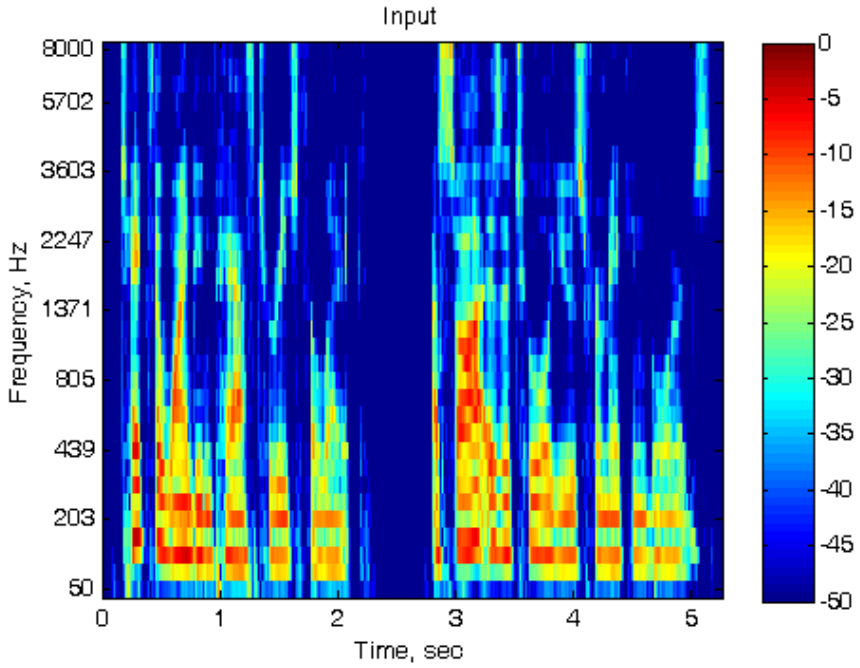
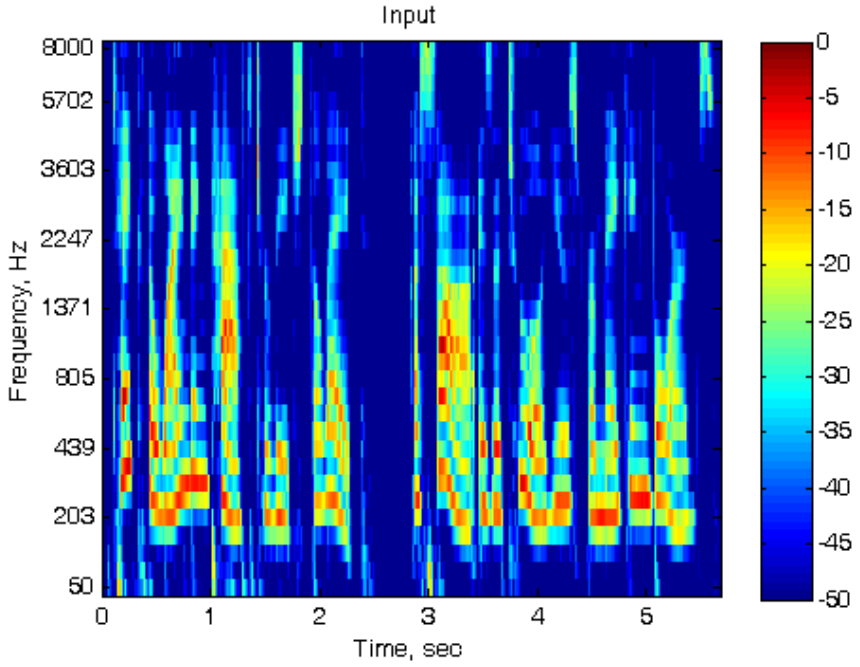


Figure 5 a, b. Output of gammatone filter bank for male and female talker. A) The male talker has substantial energy under 500 Hz. B) The female talker has primary energy between 200-1300 Hz.

A.



B.



Speech is played out at an average level of 65 dB SPL for listeners in the normal hearing group. Individualized frequency-specific linear amplification using NAL-R (Byrne & Dillon, 1996) is applied to each stimulus to compensate for elevated thresholds for listeners in the group with hearing loss.

Test Conditions

This goal of this study is to establish the role of temporal fine structure in sound quality perception. The aim is to determine the impact that type of signal processing (type of vocoding noise and band removal), amount of background noise, frequency region, and hearing status have on the role of temporal fine structure to quality perception. To this end, this study includes four sets of stimuli, for a total of 75 test conditions. In all sets of stimuli, original temporal fine structure information below (and including) band 16 is kept intact. The upper edge frequency for band 16 is 1605 Hz. Table 4 provides an overview of the 75 test condition.

Set 1

The first set of conditions is created by replacing bands of speech with FL vocoded noise (see Figure 3a). The noise is passed through an envelope-modulation low-pass filter at 300 Hz to retain the same temporal envelope as the original speech while removing temporal fine structure cues. The original speech is divided in 32 bands. In groups of two, beginning with the highest two frequency bands, two bands of FL noise replace two bands of original speech until the highest 16 bands are vocoded with FL noise. These 8 conditions are presented in quiet and at a 12 and 18 dB SNR, for a total of 24 conditions.

Set 2

The second set of stimuli is similar to the first, but uses SM vocoded noise (see Figure 3b). Again, the noise is passed through an envelope-modulation low-pass filter at 300 Hz to retain the same temporal envelope as the original speech while removing temporal fine structure cues. The original speech is divided in 32 bands. In groups of two, beginning with the highest two frequency bands, two bands of SM noise replace two bands of original speech until the highest 16 bands are vocoded with SM noise. These 8 conditions are presented in quiet and at a 12 and 18 dB SNR, for a total of 24 conditions.

Set 3

The third set of stimuli functions as the control set, and removes entire bands of speech information. In groups of two, beginning with the highest two frequency bands, two bands of speech information are systematically removed two bands at a time. In the most limited condition speech will only be available in the lowest 16 bands. These 8 conditions are presented in quiet and at a 12 and 18 dB SNR, for a total of 24 conditions.

Set 4

An additional 3 fully intact conditions are also presented. The original speech is divided in 32 bands and presented in quiet, and at 12 and 18 dB SNR.

Table 4. Experimental test conditions. I = intact temporal fine structure, FL =fluctuating vocoding noise, SM = smooth vocoding noise. Each of the 25 conditions was presented in levels of background noise (quiet, 18 db SNR, and 12 db SNR) for a total of 75 total conditions.

Cond #	**16**	**18**	**20**	**22**	**24**	**26**	**28**	**30**	**32**
1	I	I	I	I	I	I	I	I	I
2	I	I	I	I	I	I	I	I	
3	I	I	I	I	I	I	I		
4	I	I	I	I	I	I			
5	I	I	I	I	I				
6	I	I	I	I					
7	I	I	I						
8	I	I							
9	I								
10, 18	I	I	I	I	I	I	I	I	FL, SM
11, 19	I	I	I	I	I	I	I	FL, SM	FL, SM
12, 20	I	I	I	I	I	I	FL, SM	FL, SM	FL, SM
13, 21	I	I	I	I	I	FL, SM	FL, SM	FL, SM	FL, SM
14, 22	I	I	I	I	FL, SM	FL, SM	FL, SM	FL, SM	FL, SM
15, 23	I	I	I	FL, SM	FL, SM	FL, SM	FL, SM	FL, SM	FL, SM
16, 24	I	I	FL, SM	FL, SM	FL, SM	FL, SM	FL, SM	FL, SM	FL, SM
17, 25	I	FL, SM	FL, SM	FL, SM	FL, SM	FL, SM	FL, SM	FL, SM	FL, SM

Test Procedures

This study includes two main experimental tasks, a sound quality rating task and a speech intelligibility task.

Quality:

In the sound quality rating task, listeners judge the sound quality of the 75 conditions (described above, see *Test Conditions*). The 11-point rating scale ranges from 0 (min) – 10 (max) in 0.1 increments, and is modeled after the overall impression scale in Gabrielsson et al. (1988). Figure 6 shows an example of the visual screen produced by the custom MATLAB routine that is used by the listener to rate the sound quality. A practice set is rated on the scale and includes a sample of 27 of the 75 test conditions. This practice set is intended to familiarize the listener to the quality rating tasks and range of test conditions. Each of the 75 conditions are presented a total of 4 times for each talker, with 2 presentations for one talker and then two presentations for the second talker before repeating the cycle. The order of presentations is randomized, such that half of the listeners hear the male talker first and half hear the female talker first. Listeners are responsible for the pace of presentation and are given breaks after 50 trials, or more frequently as needed. Test instructions for the sound quality task are included in the Appendix.

Intelligibility:

For each of the 75 conditions, listeners are presented with five different sentences for each of the two talkers, for a total of 750 sentences in the intelligibility experiment (see *Test Conditions*). Each sentence contains five keywords, for a total of 3750 keywords. Conditions are presented in random order, with half of the listeners presented with the male talker first, and

half presented with the female talker first. Given that the IEEE corpus has only 720 sentences, a small number of sentences are repeated. All sentences within a talker (male or female) are unique. Listener responses are scored by the examiner for number of keywords correctly identified, and verbal responses are recorded for offline scoring as needed. Completion of the intelligibility task allows us to quantify intelligibility and familiarize listeners to the test conditions. Test instructions for the speech intelligibility task are included in the Appendix.

Cognitive and Temporal Resolution Screenings (MMSE and TFS1 test)

In addition to measures of speech intelligibility and quality, listeners also participate in two additional tasks. The first, the Mini Mental Status Exam (MMSE) (results listed in Table 1), is designed to provide a screening test for general cognitive status (Folstein et al., 1975). All listeners have a passing MMSE score of 27 or higher.

The second task, the TFS1 test (Moore & Sek, 2009a), is designed to determine sensitivity to temporal fine structure changes between harmonic (H) and inharmonic (I) complex tones (results listed in Table 2). In this task, listeners are to discriminate between a consistent series containing all H tones from a fluctuating series, containing both H and I tone. In the consistent condition, four tones are presented in an H H H H order. In the fluctuating condition, four tones are presented in an H I H I order. The results of the task are computed as the smallest ΔF detectable by a listener. Following two correct responses in a row, the value of ΔF is decreased, while following one incorrect response it is increased. The procedure continues until eight changes in direction have occurred. Maximum ΔF allowed is 0.5 ΔF. Temporal fine structure thresholds above this point are extrapolated based on the percentage correct (out of 40) at 0.5 ΔF. Normal temporal fine structure sensitivity has been shown to be a ΔF of 0.3 or better. For participants in this project, scores are converted to d' values in order to accurately compare

abilities between listeners. There was a level limitation with playout of the TFS1 stimuli, such that listeners with thresholds above 55dB HL for the harmonic frequencies to be tested (2200, 3300, and 4400 Hz) are not able to participate in the task.

Stimuli Presentation and Playout

Listeners are first tested using the TFS1 test and the MMSE. Listeners then participate in the intelligibility task, with a total of 5 repetitions of each condition per talker. The intelligibility task serves two purposes. First, it allows for documentation of intelligibility for the processing conditions chosen for this study. Second, it provides listeners with familiarization to the talkers and the various processing conditions they encounter in the quality portion of the experiment. The final stage of the experiment is the quality ratings portion. Listeners judge the overall speech quality of the processing conditions a total of 4 times for each talker.

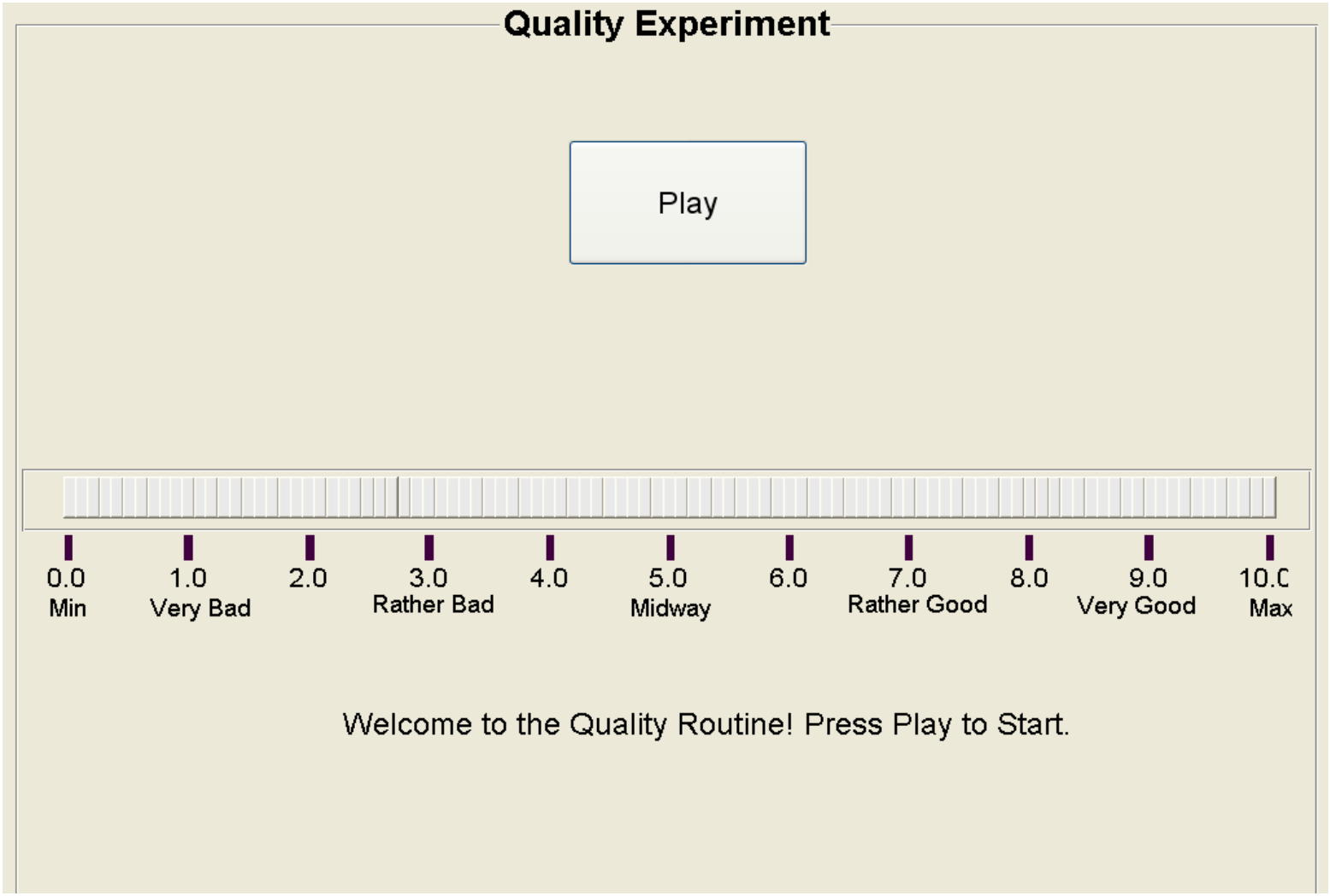
For listener presentation, the digitally-stored intelligibility and quality speech stimuli are processed through a digital-to-analog converter (TDT RX8), an attenuator (TDT PA5) and a headphone buffer amplifier (TDT HB7). Finally, the stimuli are presented monaurally to the listeners' test ear through a Sennheiser HD 580 earphone.

The TFS1 test is presented to the listener on a laptop inside a double-walled sound booth. The stimuli are processed through an external sound card (E-MU 440) and are presented to the listener's test ear through Sennheiser HD-25 headphones.

Table 5: Experimental tasks broken down by hour of participation

Hour	Task	Details
1	Audiometric Evaluation	typanometry, acoustic reflexes, air and bone conduction pure tones, Speech Reception Threshold testing (SRT), and word recognition testing (using NU-6 recorded word lists), QuickSIN
2	MMSE, TFS1	MMSE: Cognitive screening task; TFS1: sensitivity to changes in temporal fine structure at F0s of 200, 300, and 400 Hz
3-5	Speech Intelligibility	2 talkers * 5 presentations * 75 conditions= 750 total sentences (3750 keywords)
6-8	Sound Quality	2 talkers * 4 presentations * 75 conditions = 600 total quality ratings

Figure 6. Screen capture of custom MATLAB interface used by each subject to rate the quality of the stimuli.



Audibility

An excitation pattern model (Hopkins et al., 2008) is employed in order to quantify the audibility of the target signal after NAL-R gain has been applied for each subject with hearing loss. Mean excitation levels between 100 and 10,000 Hz are calculated for each subject. This model incorporates a middle ear transfer function. Default values for the proportion of inner and outer hair cell damage are assumed based on audiometric thresholds in dB HL. The model gives estimates of the excitation level at threshold as a function of frequency for each subject. This threshold level was compared to the level of the stimulus for both the male and female talker. For both the talkers, it is found that the target signal was audible through at least 5,500 Hz. Figure 7 shows the threshold level compared to the stimulus level for the male talker for each HI listener, as well for the level for an NH listener, while Figure 8 shows the same for the female talker. The entire signal was audible for all listeners in the NH group and for 5/10 listeners in the HI group. For the 5 listeners who were unable to hear the full signal, it was ensured that the signal was audible through at least 5500 Hz (band 27). Specifically, listeners with limited audibility were HI9 (audibility through band 27), HI2 (band 28), HI3 (band 29), HI4 (band 31), and HI5 (band 31).

Figure 7. Excitation levels of the male target speech (dotted lines) and excitation levels at threshold (solid line) for individual hearing-impaired subjects. Excitation levels for normal hearing subjects are shown for comparison in the bottom right panel.

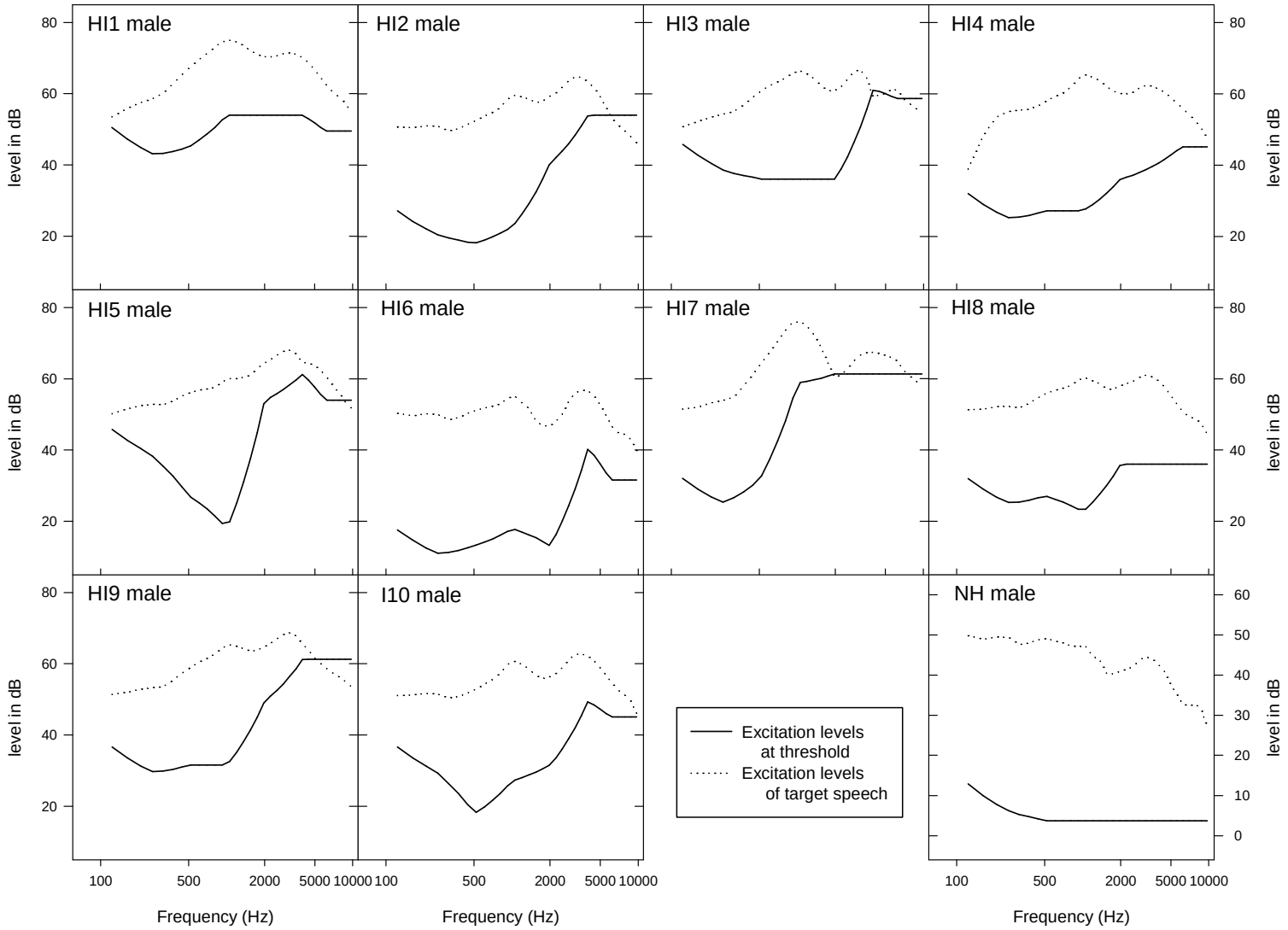
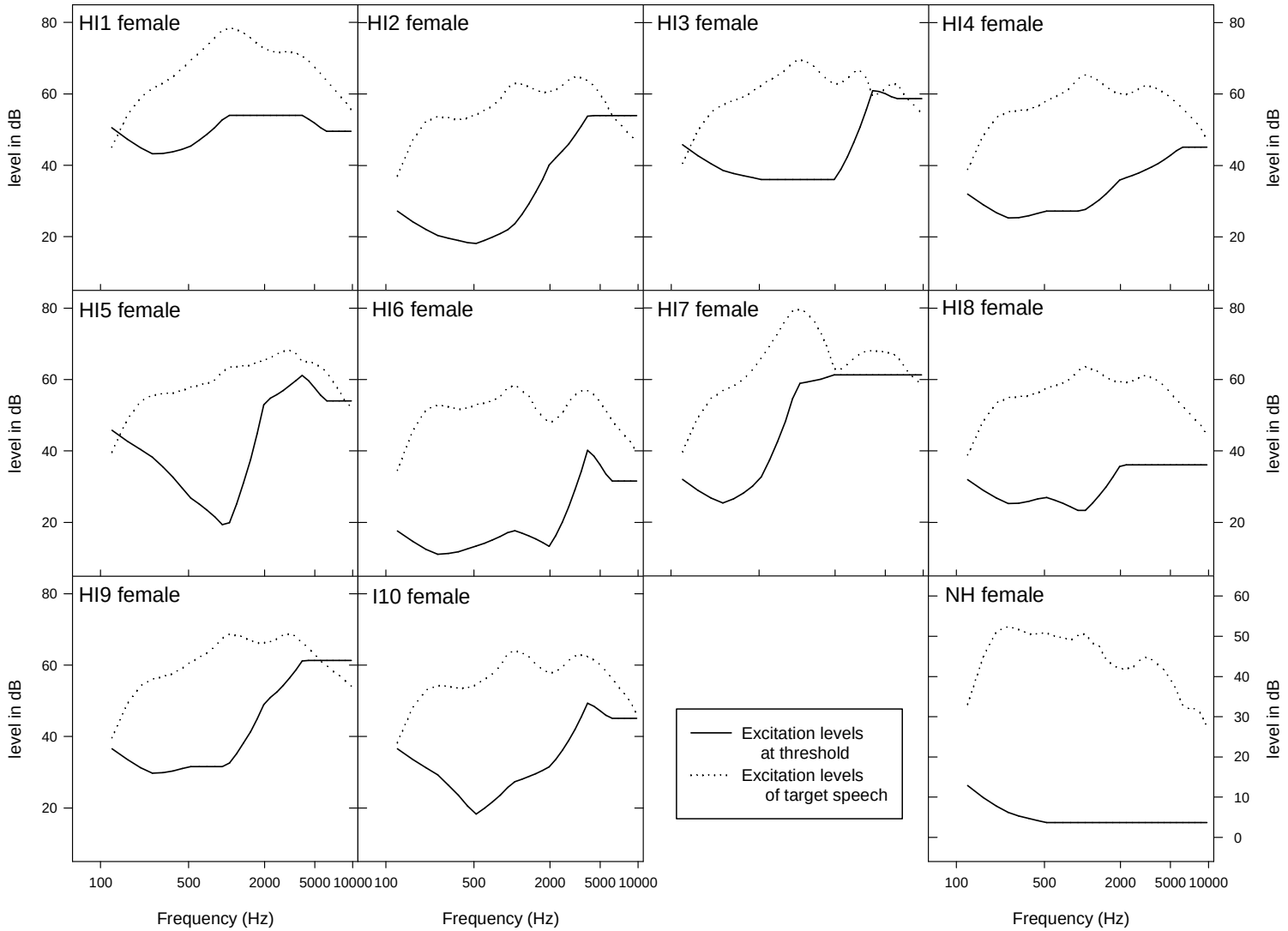


Figure 8. Excitation levels of the female target speech (dotted lines) and excitation levels at threshold (solid line) for individual hearing-impaired subjects. Excitation levels for normal hearing subjects are shown for comparison in the bottom right panel.



Chapter 4: Results

Results of the study include a data set of quality ratings and a dataset of intelligibility scores for 20 listeners (10 with normal hearing and 10 with hearing loss). The quality ratings are calculated from four presentations for two talkers in 75 conditions for a total of 600 quality ratings per listener (not including practice trials). Quality ratings are reported, pictured, and analyzed in terms of the raw score given on the 0-10 point scale. The intelligibility scores are calculated from five presentations for two talkers in 75 conditions for a total of 750 sentences per listener (3750 keywords). Intelligibility scores are reported and illustrated in terms of percent correct of number of keywords, although all statistics were performed on arcsine transformed data (Studebaker, 1985) in an effort to stabilize the error variance.

The quality data set is first analyzed to determine reliability and consistency of the ratings. The quality data set is then analyzed to explore differences between a) the male and female talker, b) types of signal processing (FL, SM, and REM), c) effects of SNR (quiet, 18 dB and 12 dB SNR), d) band cutoff (vocoding or removal of the highest 16 bands in 8 consecutive 2-band steps moving from highest to lowest), and e) effects of hearing status. This analysis is used to examine the data in context of Specific Aims 1 and 2 by considering the relationship between quality perception and temporal fine structure. The final portion of the results section examines Specific Aim 3 and details the relationship between the quality ratings and intelligibility scores for both groups of listeners.

Unless otherwise indicated, all statistics were completed using SPSS version 18 using a mixed-model repeated measures analysis of variance (RM ANOVA). In situations where the

assumption of sphericity was violated, the Greenhouse-Geisser correction factor was used to determine significance.

Quality Dataset

Figures 9 and 10 show the average quality ratings given by the NH group (Figure 9) and the HI group (Figure 10). Each of the twelve panels contain the average scores with standard error bars for all three signal processes (FL, SM, and REM) as a function of band cutoff (vocoding or removal of the highest 16 bands in 8 consecutive 2-band steps moving from highest to lowest, with the full intact signal as the right-most point in each panel). Each panel shows one SNR (quiet, 18 dB SNR, or 12 dB SNR) for one talker (male or female). The left panels show data from quality ratings for quiet speech, the middle panels show 18 dB SNR and the right panels show 12 dB SNR.

The raw data have been examined based on a number of factors: talker sex, signal processing, SNR, and band cutoff. For each of these factors, the ranges of quality ratings, as well as the general movement in quality rating changes, are similar between listeners with normal hearing and listeners with hearing loss. Quality ratings for the male and female talker are similar in range; however, the overall trend is for the female talker to be rated more poorly than the male talker for the same condition. When examined by signal processing, the trend is for vocoded conditions to be rated more highly than band removal. Within the two types of vocoded noise, SM noise is consistently rated more highly than FL noise. Quality ratings based on SNR show a large differentiation, with quality ratings decreasing as the amount of background noise increases, irrespective of signal process. Finally, as band cutoff decreases (more of the signal is

manipulated) quality ratings decrease. This effect is most noticeable in quiet, with the effects of band cutoff decreasing as the SNR decreases.

Specifically, the male talker ratings for the NH group range from 2.33 to 8.92 and for the female talker range from 2.22 to 9.0. The average quality ratings for the HI group for the male talker range from 2.63 to 8.65 and for the female talker range from 2.76 to 8.63. Quality ratings for the NH group for the FL vocoding noise range from 4.63 to 9.0, ratings for SM range from 5.38 to 8.87, and for REM range from 2.22 to 8.91. Quality ratings for the HI group for the FL vocoding noise range from 4.81 to 8.65, ratings for SM range from 5.02 to 8.64, and for REM range from 2.76 to 8.62. Quality ratings for quiet speech range from 2.87 to 8.97 for the NH group and from 3.87 to 8.65 for the HI group. Quality ratings for 18 dB SNR range from 2.45 to 6.34 for the NH group and from 2.99 to 6.42 for the HI group. In the conditions with the most background noise, 12 dB SNR, quality ratings range from 2.22 to 5.95 for the NH group and from 2.76 to 5.77 for the HI group. For band cutoff quality ratings ranged from 2.22 (in the 16 band cutoff condition) to 9.0 (in the intact 32 band condition) for the NH group and from 2.63 to 8.65 for the HI group. While band cutoff has 9 levels, from 16 BC to 32 BC, the statistical analysis contains just 8 levels (16 BC to 30 BC), as the intact 32 BC was not collected for all signal processes. The intact condition served as its own set of conditions, as there was no signal processing factor associated with the intact signal.

Figure 9. The average quality ratings for listeners in the NH group are shown here. Each panel contains the three signal processing types for one talker (male or female) and one SNR (quiet, 18, or 12 dB SNR). The FL vocoding noise is represented by open triangles, the SM vocoding noise by closed circles, and the REM conditions by closed squares. The top row shows the results for the male talker, the bottom row for the female talker. Speech in quiet is displayed in the left panels, 18 dB SNR in the middle panels, and 12 dB SNR in the right panels. The error bars represent the standard error.

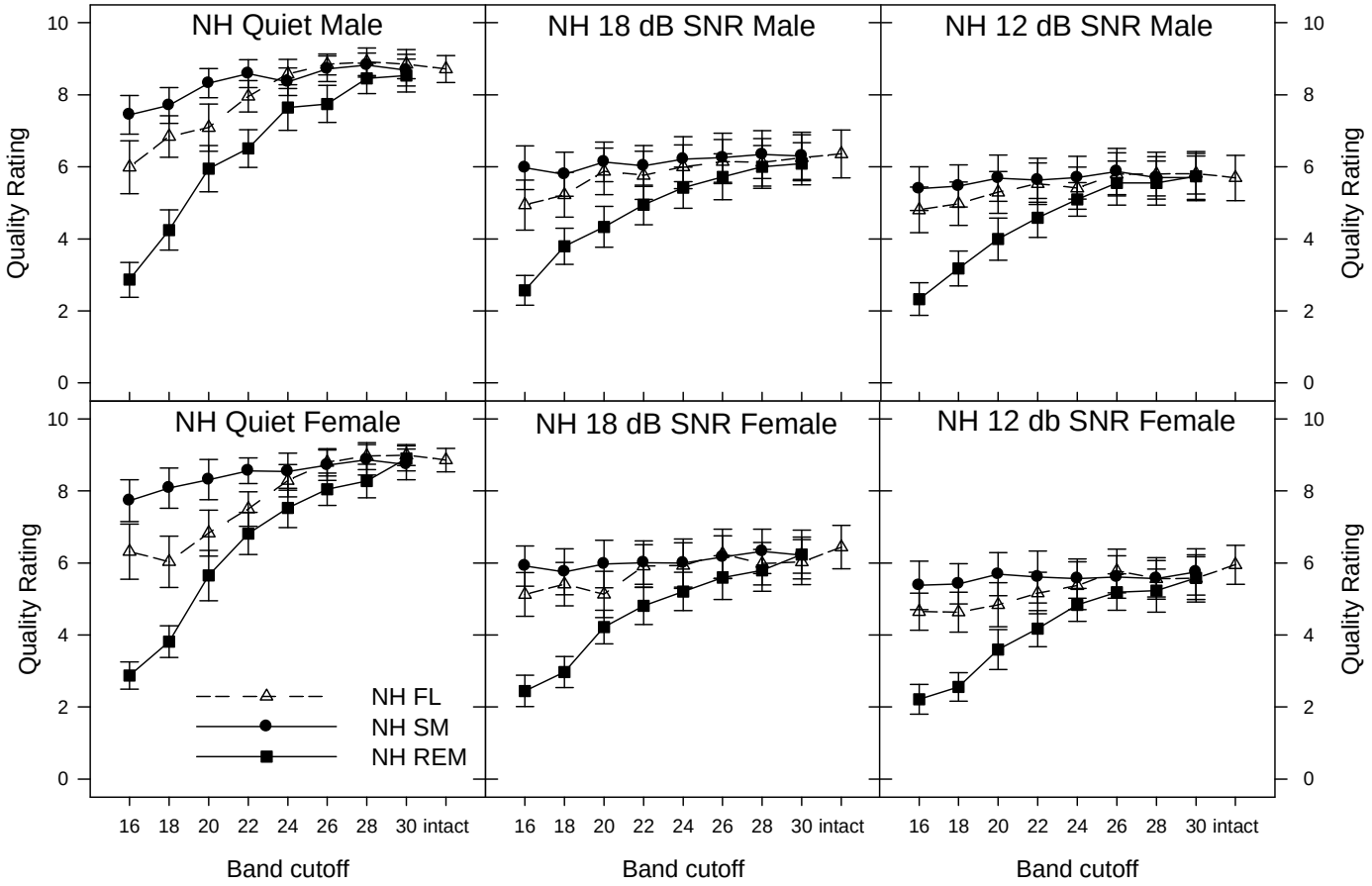
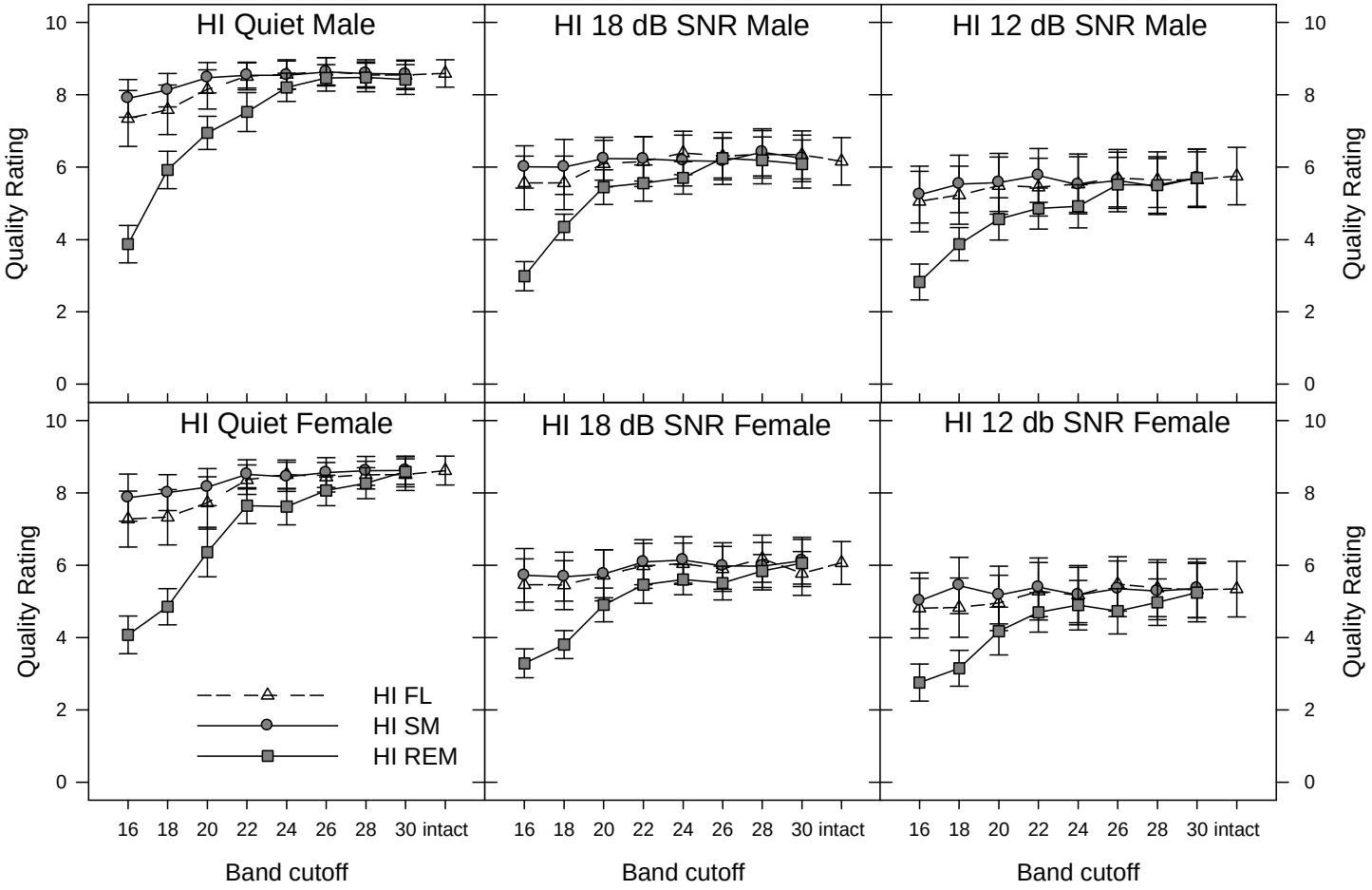


Figure 10. The average quality ratings for listeners in HI group are shown here. Each panel contains the three signal processing types for one talker (male or female) and one SNR (quiet, 18, or 12 dB SNR). The FL vocoding noise is represented by open triangles, the SM vocoding noise by closed circles, and the REM conditions by closed squares. The top row shows the results for the male talker, the bottom row for the female talker. Speech in quiet is displayed in the left panels, 18 dB SNR in the middle panels, and 12 dB SNR in the right panels. The error bars represent the standard error.



Within visit and across visit reliability

In this section, the results are discussed in terms of listeners' consistency within a visit and across visits, differences between the male talker and the female talker, and the effects of temporal fine structure removal, SNR, and band cutoff.

Listeners provided quality ratings for each processing condition for each talker a total of four times, with the first two presentations rated within the first session, and the second two presentations rated within a session during a second, separate visit. Bivariate correlations provide a means to quantify how consistent listeners were within and across visits for the same stimulus (same condition and some talker). Figure 11 shows four separate scatter plots for trial 1 plotted against trial 2 for the NH listeners for visit 1 and visit 2 with both the male and female talkers, while Figure 12 shows the same for the HI listeners. Each data point represents a single listener's rating on trial 1 of the given visit (horizontal axis) against trial 2 (vertical axis) of the same visit. Both groups of listeners demonstrate consistent ratings within a visit. The Pearson correlation coefficients are 0.89 and 0.92 ($p< 0.001$) for the HI listeners and 0.93 and 0.96 ($p<0.001$) for the HI listeners.

Figure 13 shows an additional four scatter plots for the correlation between visit 1 and visit 2 for the male talker and the female talker for each group. The female talker is depicted in the top panels, the male talker in the bottom panels. Each data point represents a single listener's rating for trial 1 of visit 1 plotted against trial 1 of visit 2 and trial 2 of visit 1 plotted against trial 2 of visit 2. The Pearson correlation coefficient is between 0.84 and 0.89 ($p < 0.001$) for the four comparisons. When the two quality ratings from visit 1 are averaged and two quality ratings from visit 2 are averaged the Pearson correlation coefficient between visit 1 and visit 2 rises to

range from 0.86 - 0.92 ($p<0.001$). The strength of the correlations for both groups of listeners suggests that the rating scale used in this study was a reliable instrument in quantifying the effects of stimulus processing on quality perception. As such, the mean score of all 4 trials for each condition form the basis of the remaining statistical analyses.

Figure 11. Scatter plots of the within-session ratings for the NH group for the male talker (top panels) and the female talker (bottom panels). Each data point represents a single subject's rating of a specific stimulus for trial 1 (horizontal axis) and trial 2 (vertical axis) within a visit (visit 1 in left panels and visit 2 in right panels). The dashed line would be perfect match (0 to 0 and 10 to 10). The solid line represents actual regression line. All correlations are significant at $p < 0.01$.

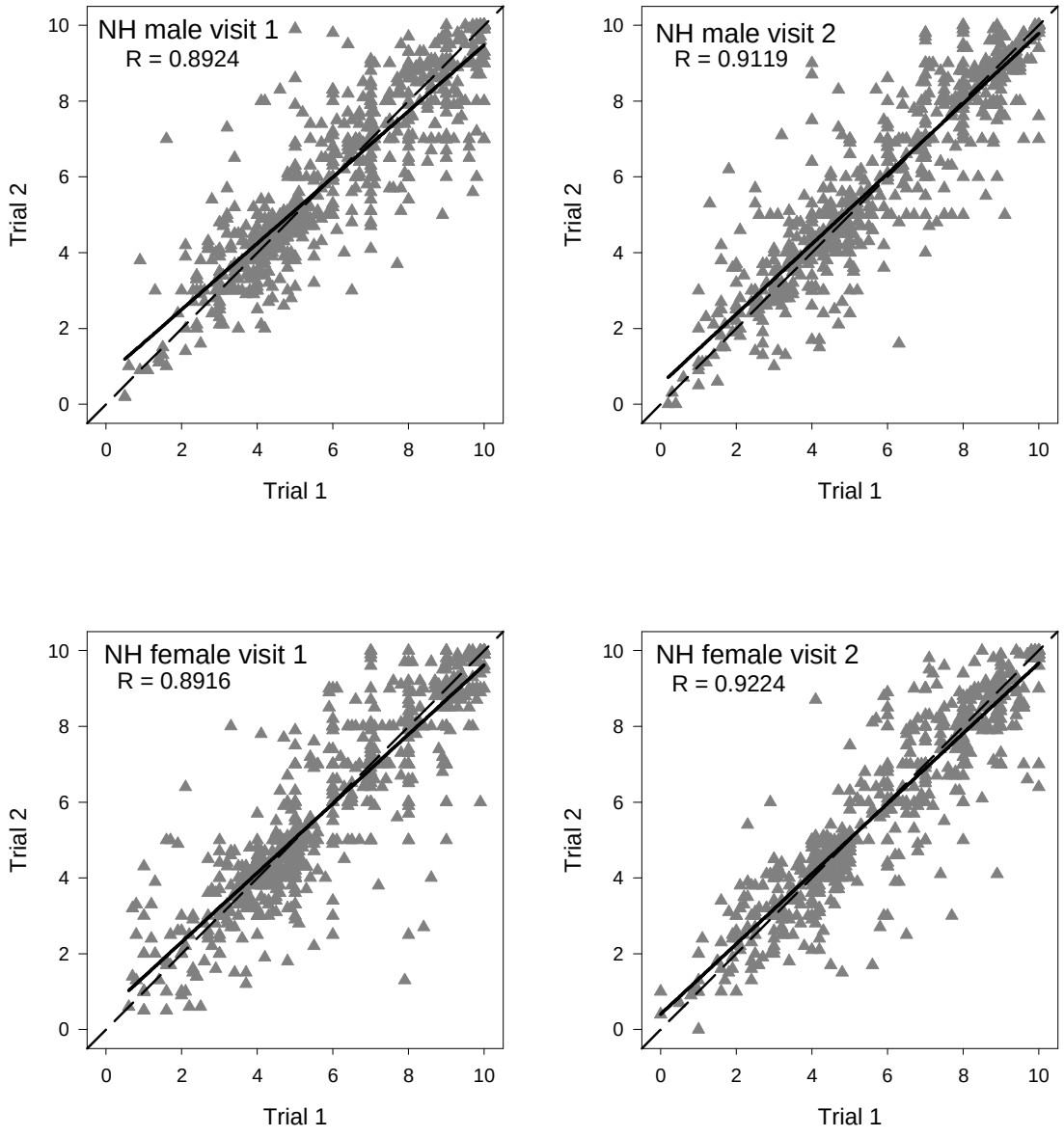


Figure 12. Scatter plots of the within-session ratings for the HI group for the male talker (top panels) and the female talker (bottom panels). Each data point represents a single subject's rating of a specific stimulus for trial 1 (horizontal axis) and trial 2 (vertical axis) within a visit (visit 1 in left panels and visit 2 in right panels). The dashed line would be perfect match (0 to 0 and 10 to 10). The solid line represents actual regression line. All correlations are significant at $p < 0.01$.

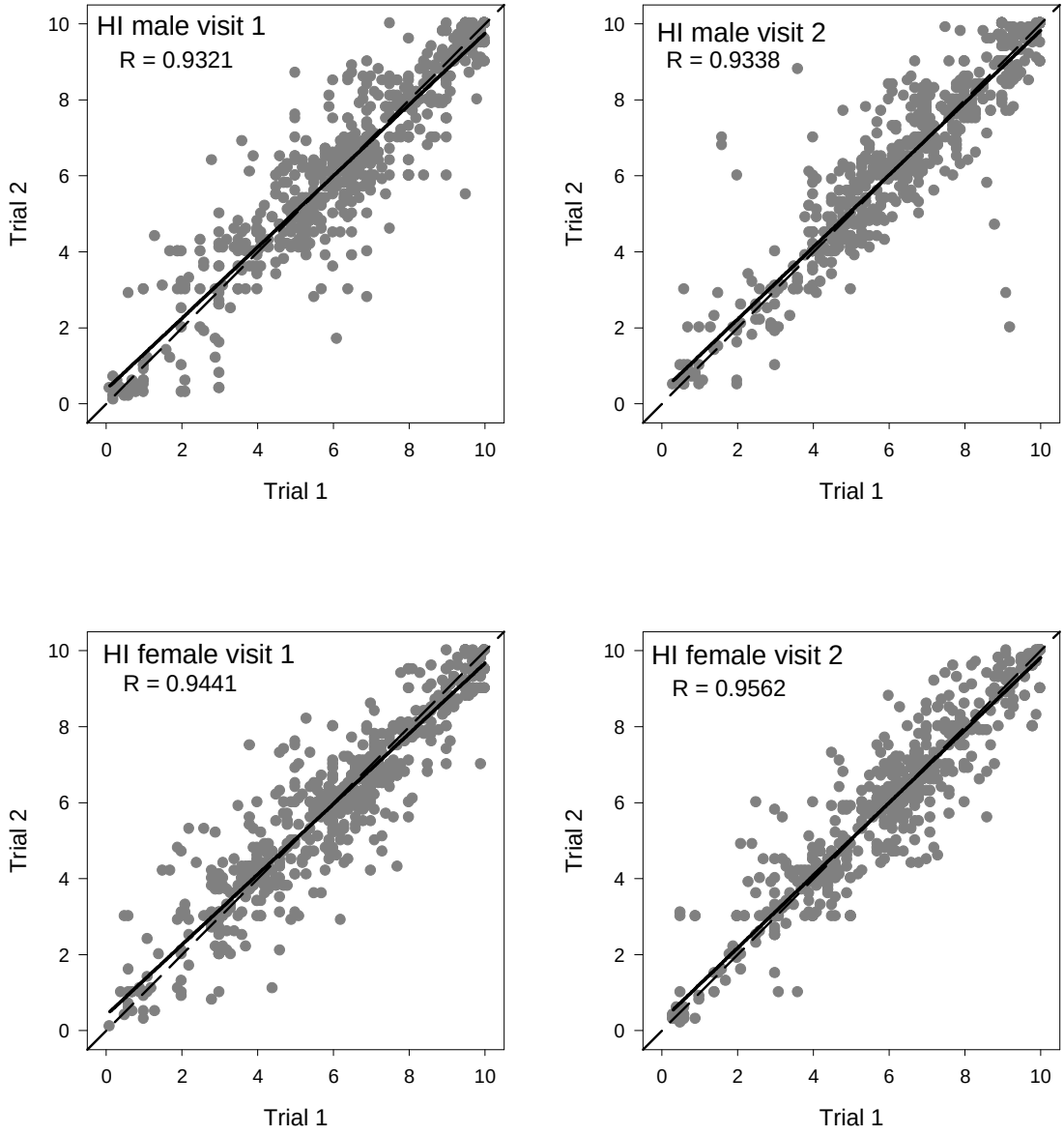
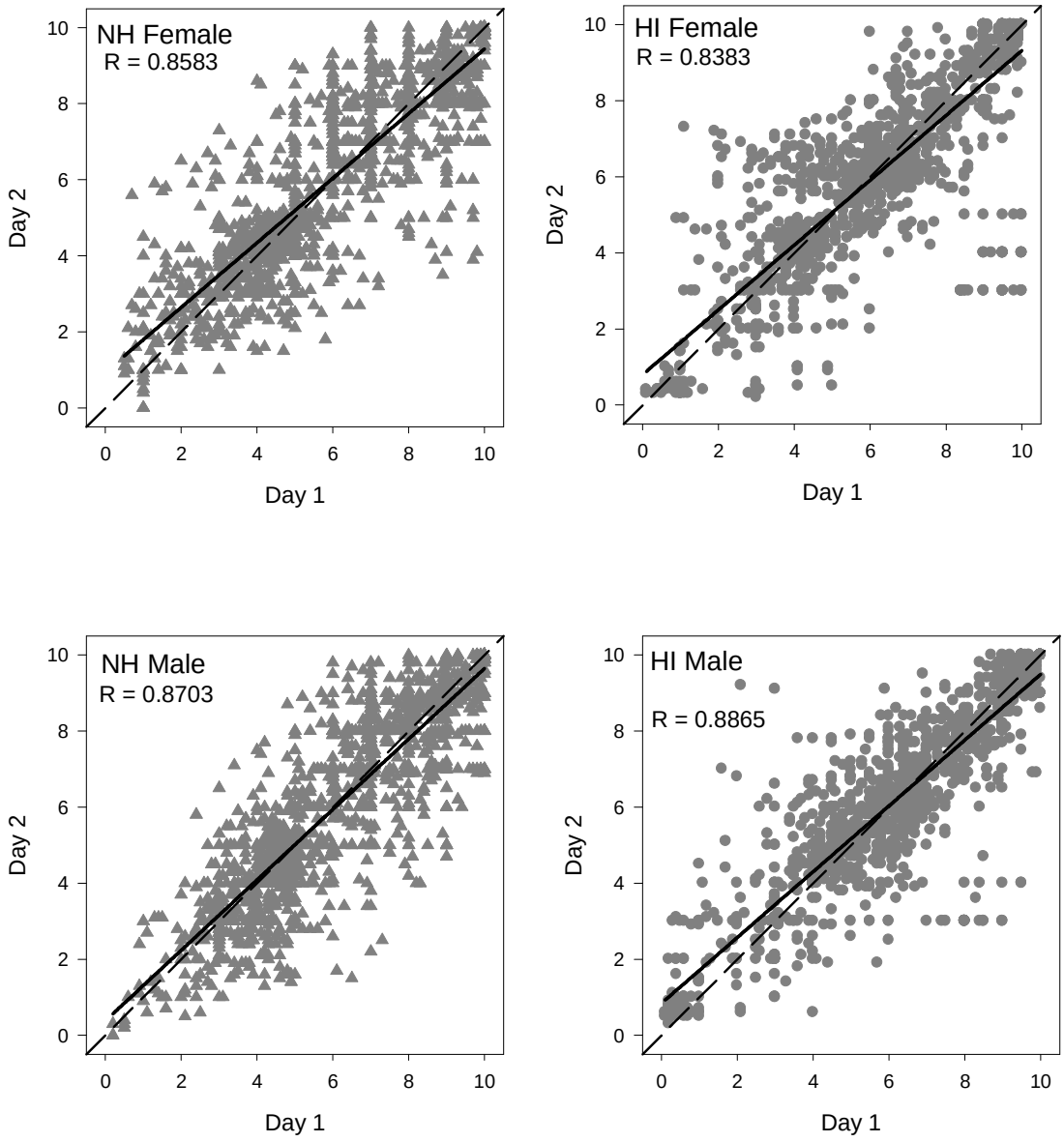


Figure 13. Scatter plots of the across-session ratings for the female talker (top panels) and the male talker (bottom panels). The NH group is represented in the left panels, the HI group in the right panels. There are two data points for each listener in each panel. The first data point represents trial 1 of visit 1 (horizontal axis) and trial 1 of visit 2 (vertical axis). The second data point plots trial 2 of visit 1 (horizontal axis) and trial 2 of visit 2 (vertical axis). The dashed line would be perfect match (0 to 0 and 10 to 10). Solid line represents actual regression line. All correlations are significant at $p < 0.01$. Similar results were found when the within-visit ratings were averaged.



Omnibus Statics: Quality Ratings

An omnibus RM ANOVA provides an overview of the entire quality dataset and includes the within-subject factors of talker sex (2 levels: male and female), signal processing (3 levels: FL, SM, and REM), SNR (3 levels: quiet, 18 dB and 12 dB SNR), and band cutoff (8 levels: 16 BC to 30 BC). There is a single between-subject factor of group. Table 6 presents the results of this statistical analysis. All four within-subject main effects are significant, indicating that all main effects are significantly related to quality ratings: talker sex [$F(1,18) = 10.2$; $p = 0.005$; partial $\eta^2 = 0.362$], signal processing [$F(2,36) = 32.6$; $p < 0.001$; partial $\eta^2 = 0.644$], SNR [$F(2,36) = 38.8$; $p < 0.001$; partial $\eta^2 = 0.683$], and band cutoff [$F(7,126) = 81.8$; $p < 0.001$; partial $\eta^2 = 0.82$]. The between-subject factor of group is not significant [$F(1,18) = 0.067$; $p = 0.798$; partial $\eta^2 = 0.004$], indicating that there was no difference in quality ratings between the groups.

In addition to the main effects, there are also several significant interactions, five two-way interactions and one three-way interaction. The first significant interaction is talker sex * signal processing [$F(2,36) = 4.2$; $p = 0.036$; partial $\eta^2 = 0.189$], which indicates that the effect of signal processing on quality ratings is dependent on the sex of the talker. The second significant interaction is SNR * signal processing [$F(4, 72) = 10.7$; $p < 0.001$; partial $\eta^2 = 0.372$], indicating that the effect of signal process on quality ratings are dependent on SNR. The third through fifth significant interactions are based on band cutoff: talker sex * band cutoff [$F(7, 126) = 4.3$; $p = 0.007$; partial $\eta^2 = 0.192$], SNR * band cutoff [$F(14,252) = 22.2$; $p < 0.001$; partial $\eta^2 = 0.552$], and signal processing * band cutoff [$F(14,252) = 28.6$; $p < 0.001$; partial $\eta^2 = 0.614$], which indicates the effect of band cutoff on quality ratings are sensitive to changes in talker sex, SNR,

and signal process. There is one significant three-term interaction, SNR * signal process * band cutoff [$F(28, 504) = 5.1$; $p < 0.001$; partial $\eta^2 = 0.221$], which indicates that the effects of signal process on band cutoff are dependent on SNR.

Figures 9 and 10, along with the raw data, show the significant differences for the main within-subject factors revealed by the statistical analysis. The male talker is rated significantly higher than the female talker. The difference in type of signal processing is especially visible in the figures, with REM conditions showing the lowest overall ratings and SM conditions showing the highest quality ratings. The amount of background noise also has a significant impact on quality ratings, with ratings highest in the quiet conditions and decreasing significantly as the amount of background noise increases. And finally, band cutoff is a significant factor, as quality ratings decrease as more information is removed from the signal from 30 BC to 16 BC, through vocoding (FL and SM) and total removal (REM).

The significant interactions can also be interpreted in the context of the data in Figures 9 and 10. Consider, by way of example, the top three panels in Figure 9, which show the average quality ratings for the NH group for the male talker at 3 SNRs. The significant signal processing * SNR interaction indicates that the differences in quality ratings between signal processing types decreases as the SNR decreases, which can be seen in the top three panels of Figure 9. We can also easily see the SNR * band cutoff interaction in those same three panels. As SNR decreases, the effects of band cutoff are reduced. In the quiet condition, there is a large effect of band cutoff, with reduced effects of band cutoff as the amount of background noise increases to 12 dB SNR.

Table 6. Statistical results for the omnibus RM ANOVA with within-subject variables of talker-sex, signal processing, SNR, and band cutoff (BC). The between group variable is hearing status (group). The dependent variable is quality rating. Significant effects are highlighted in gray.

Variable	df	*F*	*p*	Partial η2	Observed Power
talker sex	1, 18	10.194	0.005*	0.362	0.855
signal process	2, 36	32.559	<0.001*	0.644	1.000
SNR	2, 36	38.849	<0.001*	0.683	1.000
BC	7, 126	81.827	<0.001*	0.820	1.000
group	1, 18	.067	0.798	0.004	0.057
talker sex * group	1, 18	1.501	0.236	0.077	0.213
signal process * group	2, 36	1.055	0.359	0.055	0.220
SNR * group	2, 36	.130	0.732	0.007	0.064
BC * group	7, 126	3.268	0.053	0.154	0.564
talker sex * SNR	2, 36	2.945	0.083	0.141	0.458
talker sex * BC	7, 126	4.283	0.007*	0.192	0.860
SNR * BC	14, 252	22.211	<0.001*	0.552	1.000
talker sex * signal process	2, 36	4.201	0.036*	0.189	0.609
SNR * signal process	4, 72	10.653	<0.001*	0.372	0.984
signal process * BC	14, 252	28.626	<0.001*	0.614	1.000
talker sex * signal process * group	2, 36	.436	0.595	0.024	0.106
talker sex * SNR * group	2, 36	.033	0.933	0.002	0.054
SNR * signal process * group	4, 72	.402	0.675	0.022	0.110
talker sex * SNR * signal process	4, 72	1.134	0.347	0.059	0.339
talker sex * BC * group	7, 126	.557	0.789	0.030	0.233
SNR * BC * group	14, 252	1.934	0.161	0.097	0.366
talker sex * SNR * BC	14, 252	1.505	0.177	0.077	0.590
signal process * BC * group	14, 252	.666	0.535	0.036	0.160
talker sex * signal process * BC	14, 252	1.853	0.113	0.093	0.597
SNR * signal process * BC	28, 504	5.106	<0.001*	0.221	0.994
talker sex * SNR * signal process * BC	28, 504	1.357	0.222	0.070	0.597
SNR * signal process * BC * group	28, 504	1.203	0.309	0.063	0.474
talker sex * signal process * BC * group	14, 252	.759	0.578	0.040	0.257
talker sex * SNR * BC * group	14, 252	.433	0.869	0.023	0.179
talker sex * SNR * signal process * group	4, 72	1.708	0.173	0.087	0.434
talker sex * SNR * signal process * BC * group	28, 504	1.172	0.321	0.061	0.522

Results based on Specific Aims 1 and 2

The overall omnibus RM ANOVA reveal several interesting findings. First, quality ratings differ between the male talker and the female talker, with the female talker rated consistently more poorly than the male talker. In addition, there is a significant effect for type of signal processing. However, because three types of signal processing are included in the omnibus analysis (FL, SM, and REM), it is difficult to directly determine the separate effects of the FL and SM vocoding noise. Given that the goal of the study is to examine the effects of temporal fine structure removal, a second series of analyses are completed. In this second analysis the male talker and the female talker datasets are considered separately. In addition, the factor signal process is reduced to two levels in order to explicitly examine the effects of the two types of vocoding noise, and is referred to as noise type.

This second set of analyses allows for more direct examination of the research questions for Specific Aims 1 and 2. Specifically, these analyses answer four questions. First, does the type of vocoding noise used to remove temporal fine structure affect quality ratings? Second, does the presence and amount of background noise affect the importance of temporal fine structure to quality ratings? Third, does frequency region (defined by band cutoff) affect the importance of temporal fine structure to quality ratings? And finally, is there an effect of hearing status on the effects of temporal fine structure removal based on signal process, SNR, and band cutoff? The final question will be answered in a separate section related to Specific Aim 2.

Results of Specific Aim 1

Specific Aim 1: Establish the relationship between temporal fine structure and sound quality perception.

Table 7 displays the results for the statistical analysis for the dataset including the male talker with two levels of noise type (FL and SM vocoding noise). In answer to question 1, does the type of vocoding noise affect quality ratings? The answer is yes, with SM noise rated significantly higher than FL noise for the male talker [$F(1,18) = 19.4$; $p > 0.001$, partial $\eta^2 = 0.518$]. In answer to the second question: Does the presence and amount of noise influence quality ratings for speech which has been vocoded? The answer is yes, as the main effect of SNR is significant [$F(2, 36) = 36.3$; $p > 0.001$, partial $\eta^2 = 0.669$]. Quality ratings for vocoded speech decrease as SNR decreases. In answer to the third question, is there an effect of frequency region of quality ratings for vocoded speech? Again, the answer is yes [$F(7, 126) = 23.7$; $p > 0.001$, partial $\eta^2 = 0.569$]. Quality ratings for vocoded speech decrease as the band cutoff decreases (more of the speech signal is vocoded), such that vocoded speech with lower band cutoffs are rated more poorly.

Interestingly, the interaction between noise type and SNR is not significant [$F(2,36) = 1.9$; $p = 0.187$, partial $\eta^2 = 0.093$], indicating that the amount of background noise does not differentially affect the two types of vocoding noise. There is, however, a significant interaction between noise type and band cutoff [$F(7,126) = 11.0$; $p > 0.001$, partial $\eta^2 = 0.379$], such that quality ratings for the FL vocoding noise show greater decreases as band cutoff decreases when compared to SM vocoding noise. The interaction between SNR and band cutoff is significant [$F(14, 252) = 11.0$; $p > 0.001$, partial $\eta^2 = 0.379$], such that as SNR decreases, the importance of

band cutoff to quality ratings also decreases. This finding indicates that importance of temporal fine structure to quality ratings decreases as the amount of background noise increases.

Results for the female talker show similar, though not identical, trends. Table 8 displays the results of the statistical analysis for the dataset using the female talker with two levels of noise type (FL and SM vocoding noise). As with the male talker, the effects of noise type, SNR and band cutoff are all significant factors in quality ratings. There were also some similarities to the male talker for specific interaction terms. There was a significant interaction between noise type and band cutoff , such that quality ratings for the FL vocoding noise show greater decreases as band cutoff decreases when compared to SM vocoding noise. SNR * band cutoff was also significant, such that as SNR decreased, the importance of band cutoff to quality ratings also decreased. This finding indicates that importance of temporal fine structure to quality ratings decreases as the amount of background noise increases.

In contrast to the finding for the male talker, the interaction between noise type and SNR was significant for the female talker [$F(2,36) = 6.2$; $p = 0.005$, partial $\eta^2 = 0.256$], indicating that the amount of background noise differentially affects the two types of vocoding noise, such that the difference in quality ratings between the noise type is reduced as the amount of background noise increases. A three-term interaction related to Specific Aim 1 was found to be significant, noise type * SNR * band cutoff [$F(14, 252) = 3.5$; $p = 0.005$, partial $\eta^2 = 0.162$]. This finding indicates that the effects of band cutoff on noise type are dependent on SNR, such that there is a greater effect of band cutoff on noise type for vocoded speech in quiet and this effect is reduced as the amount of background noise increases.

In summary, the differences between the male talker and female talker are limited, but significant. There is an overall difference in average quality ratings for the two talkers, with the female talker rated more poorly than the male talker. However, the main within-subject factors show the same trends. Given that similarity, there are some differences in the interactions between the talkers, dealing specifically with noise type * SNR. In the female talker, decreases in SNR have a significant differential effect on quality ratings for vocoding noise type, which is not seen for the male talker.

Results of Specific Aim 2

The second specific aim seeks to understand the effects of hearing status on quality ratings for speech with limited temporal fine structure. Similar to the analysis which included all three signal processing types and both talkers, the between-subject factor of group is not significant for the male talker [$F(1, 18) = 0.03$; $p = 0.885$, partial $\eta^2 = 0.002$] or for the female talker [$F(1, 18) = 0.001$; $p = 0.983$, partial $\eta^2 < 0.001$]. This lack of significance indicates the groups provided similar overall quality ratings.

For the male talker, the only significant interaction involving group is the three-term interaction for SNR * band cutoff * group [$F(14, 252) = 2.7$; $p > 0.048$, partial $\eta^2 = 0.129$], indicating that the effects of band cutoff on quality ratings for SNR depend on group, such that quality ratings from listeners in the group with hearing loss are less affected by band cutoff for each SNR for the male talker.

For the female talker, however, there is a significant two-term interaction, noise type * group [$F(1,18) = 6.1$; $p = 0.024$, partial $\eta^2 = 0.254$], revealing that the quality ratings for noise

type were dependent on group, such that listeners with normal hearing are more sensitive to noise type and showed larger differences in noise type ratings by rating the FL noise more poorly. There is also a significant three-term interaction involving group for noise type * band cutoff * group [$F(7, 126) = 3.9$; $p = 0.025$, partial $\eta^2 = 0.177$]. This indicates that the effects of band cutoff on quality ratings for noise type depend on group, such that quality ratings from listeners in the group with hearing loss are less affected by band cutoff for each noise type.

Table 7. Statistical results for the male talker RM ANOVA with within-subject variables of noise type (FL vs. SM vocoding noise), SNR, and band cutoff (BC). The between group variable is hearing status (group). The dependent variable is quality rating. Significant effects are highlighted in gray.

Variable	df	*F*	*p*	Partial η2	Observed Power
noise type	1, 18	19.4	<0.001*	0.518	0.986
SNR	2, 36	36.3	<0.001*	0.669	1.000
band cutoff	7, 126	23.7	<0.001*	0.569	1.000
group	1, 18	0.13	0.72	0.01	0.06
noise type * group	1, 18	4.2	0.054	0.190	0.495
SNR * group	2, 36	0.076	0.798	0.004	0.058
band cutoff * group	7, 126	2.2	0.148	0.107	0.340
noise type * SNR	2, 36	1.9	0.187	0.093	0.283
noise type * band cutoff	7, 126	11.0	<0.001*	0.379	1.000
SNR * band cutoff	14, 252	11.0	<0.001*	0.379	1.000
noise type * band cutoff * group	7, 126	1.6	0.185	0.083	0.445
SNR * band cutoff * group	14, 252	2.7	0.048*	0.129	0.667
noise type * SNR * group	2, 36	0.366	0.602	0.020	0.093
noise type * SNR * band cutoff	14, 252	2.1	0.080	0.102	0.656
noise type * SNR * band cutoff * group	14, 252	1.4	0.251	0.070	0.455

Table 8. Statistical results for the male talker RM ANOVA with within-subject variables of noise type (FL vs. SM vocoding noise), SNR, and band cutoff (BC). The between group variable is hearing status (group). The dependent variable is quality rating. Significant effects are highlighted in gray.

Variable	df	*F*	*p*	partial η^2	Observed Power
noise type	1, 18	27.6	< 0.001*	0.605	0.999
SNR	2, 36	41.1	< 0.001*	0.695	1.000
band cutoff	7, 126	22.2	< 0.001*	0.552	1.000
group	1, 18	< 0.001	0.983	< 0.001	0.050
noise type * group	1, 18	6.124	0.024*	0.254	0.649
SNR * group	2, 36	0.1	0.760	0.006	0.061
band cutoff * group	7, 126	2.1	0.147	0.104	0.366
noise type * SNR	2, 36	6.2	0.005*	0.256	0.865
noise type * band cutoff	7, 126	13.3	< 0.001*	0.426	0.998
SNR * band cutoff	14, 256	7.8	< 0.001*	0.303	0.976
noise type * band cutoff * group	7, 126	3.9	0.025*	0.177	0.699
SNR * band cutoff * group	14, 256	1.4	0.245	0.074	0.339
noise type * SNR * group	2, 36	1.2	0.306	0.063	0.221
noise type * SNR * band cutoff	14, 256	3.5	< 0.001*	0.162	0.910
noise type * SNR * band cutoff * group	14, 256	2.1	0.071	0.104	0.683

Results for Specific Aim 3

Specific Aim 3: Establish the relationship between intelligibility and quality ratings.

Intelligibility scores for the NH group (Figure 14) and the HI group (Figure 15) are presented in 12 panels. Each panel contains the average scores for all three signal processing types (FL, SM, and REM) and band cutoffs (vocoding or removal of the highest 16 bands in 8 consecutive 2-band steps). Each panel contains one SNR (quiet, 18 dB SNR, or 12 dB SNR) and one talker (male or female). Overall, intelligibility decreases as the amount of signal manipulation increases. However, for both groups of listeners, the intelligibility for even the most vocoded conditions remains above 90%. Intelligibility for the REM conditions is more variable, with intelligibility dropping as low as 44% for the most limited signal (16 BC) in the greatest amount of background noise (12 dB SNR).

Specifically, the average intelligibility scores range from 54% to 100% for the NH group and from 44% to 100% for the HI group. When examined by type of signal processing, the NH group intelligibility scores for the FL vocoding noise range from 97% to 100%, for SM vocoding noise range from 97% to 100% and for REM range from 54% to 99%. The signal processing intelligibility scores for the HI group for the FL vocoding noise range from 90% to 99%, for SM range from 90% to 100%, and for REM range from 44% to 98%. Intelligibility scores for quiet for all conditions range from 73% to 100 % for the NH group and from 70% to 100% for the HI group. Intelligibility scores for 18 dB SNR range from 73% to 100% for the NH group and from 64% to 98% for the HI group. For the conditions in 12 dB SNR, intelligibility scores range from 54% to 100% for the NH group and from 44% to 98% for the HI group. The final factor,

band cutoff, has intelligibility scores from 54% (in the 16 band cutoff condition) to 100% (in the intact 32 band condition) for the NH group and 44% to 100% for the HI group.

Given the potential influence of intelligibility on sound quality, one goal of this study is to use listening conditions in which speech intelligibility was high. Consistent with this intention, the great majority of the test conditions have intelligibility greater or equal to 90% for both groups. For the NH group, 71 /75 conditions for the male talker and 70 /75 conditions for the female talker have intelligibility scores greater or equal to 90%. Out of 150 total conditions, 140 yield intelligibility scores over 95%. For the HI group, 67/75 male conditions and 66/75 female conditions yields intelligibility scores above 90%. Out of 150 conditions, 106 yielded intelligibility scores over 95%. For both groups, all conditions with intelligibility scores below 90% are from REM conditions.

The purpose of specific aim 3 is to quantify the relationship between quality ratings and intelligibility scores. Figure 16 shows four plots of quality ratings as a function of intelligibility scores, divided by listener group and talker. Each point in a panel represents the intersection of quality rating and intelligibility score for a given group (either NH or HI) for a given talker (either male or female). The quality rating is plotted along the horizontal axis, with the intelligibility score along the vertical axis. For example, the top left panel plots the responses of the NH group for the male talker. Quality ratings vary from 4 to 9 for both vocoded noise types, while intelligibility stays high (above 90%). In contrast, more variation is seen for both quality ratings and intelligibility scores for the REM signal process.

For the NH group, for quality ratings above 4 there is little to no variation in intelligibility, with intelligibility above 90%. Similarly, in the HI group, for quality scores above

5 there is little to no variation in intelligibility scores. These figures indicate that listeners were able to make distinct quality judgments for stimuli that were similar in intelligibility.

The relationship between quality and intelligibility is examined using twelve bivariate correlations (Table 9). Each comparison is based on one group (NH or HI) and one talker (male or female) for each signal process (FL, SM, or REM). In the band removed conditions, there is a high degree of positive correlation between intelligibility and sound quality for both groups and both talkers with magnitudes ranging from 0.64 to 0.78 ($p < 0.001$), with decreases in intelligibility linked to decreases in sound quality ratings. Average intelligibility scores for the REM signal processing type range from 44% to 100%, while average quality ratings range from 2.2 to 8.9.

Unexpectedly, however, even when the range of intelligibility scores is small, and intelligibility is high, for three of the eight vocoded conditions there is still a significant relationship between intelligibility and quality ratings. The HI group shows a significant positive correlation between intelligibility scores for both the male talker and the female talker for the FL noise type, again with large magnitude of 0.64 to 0.68 ($p = 0.001$), such that as intelligibility decreases, so too do sound quality ratings. The intelligibility scores for the FL noise type for the HI group conditions ranges from 90% to 99%, while the quality ratings ranges from 4.8 to 8.7. The NH group shows a significant positive relationship between intelligibility scores and quality ratings for the female talker for the SM noise type, although the magnitude is smaller at 0.44 ($p = 0.031$). An even smaller range of intelligibility is evidenced for this condition, with scores ranging for 97% to 100%, while quality ratings range from 5.4 to 8.9, again with decreases in intelligibility related to decreases in sound quality ratings.

The results indicate that there is a clear relationship between intelligibility scores and quality ratings for all conditions with poor intelligibility, as well as in some situations where intelligibility is high. As intelligibility increases, so too does quality perception.

Figure 14. The average intelligibility scores for listeners in the NH group are shown here. Each panel contains the three signal processing types for one talker (male or female) and one SNR (quiet, 18, or 12 dB SNR). The FL vocoding noise is represented by open triangles, the SM vocoding noise by closed circles, and the REM conditions by closed squares. The top row shows the results for the male talker, the bottom row for the female talker. Speech in quiet is displayed in the left panels, 18 dB SNR in the middle panels, and 12 dB SNR in the right panels. The error bars represent the standard error.

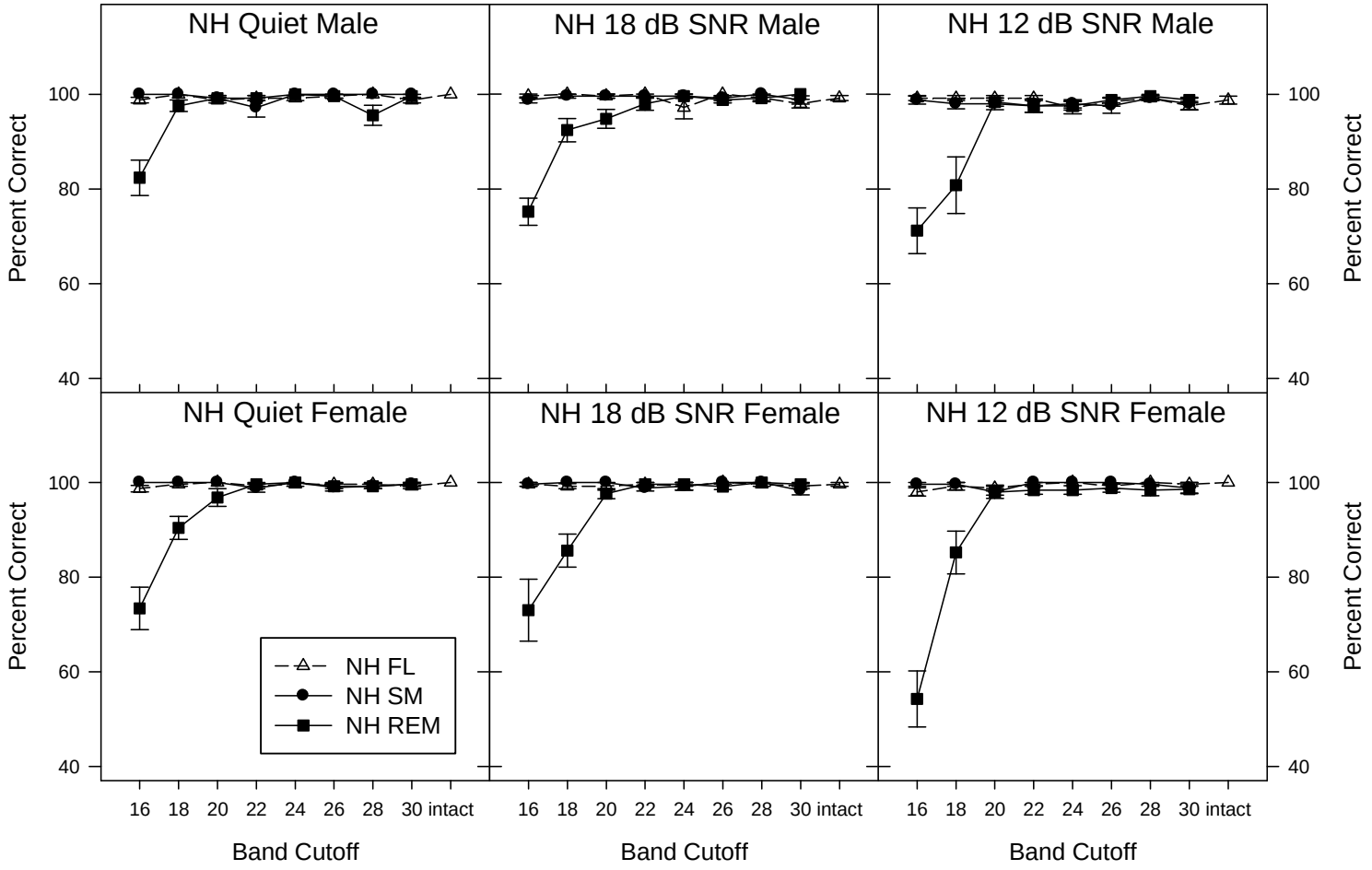


Figure 15. The average intelligibility scores for listeners in the HI group are shown here. Each panel contains the three signal processing types for one talker (male or female) and one SNR (quiet, 18, or 12 dB SNR). The FL vocoding noise is represented by open triangles, the SM vocoding noise by closed circles, and the REM conditions by closed squares. The top row shows the results for the male talker, the bottom row for the female talker. Speech in quiet is displayed in the left panels, 18 dB SNR in the middle panels, and 12 dB SNR in the right panels. The error bars represent the standard error.

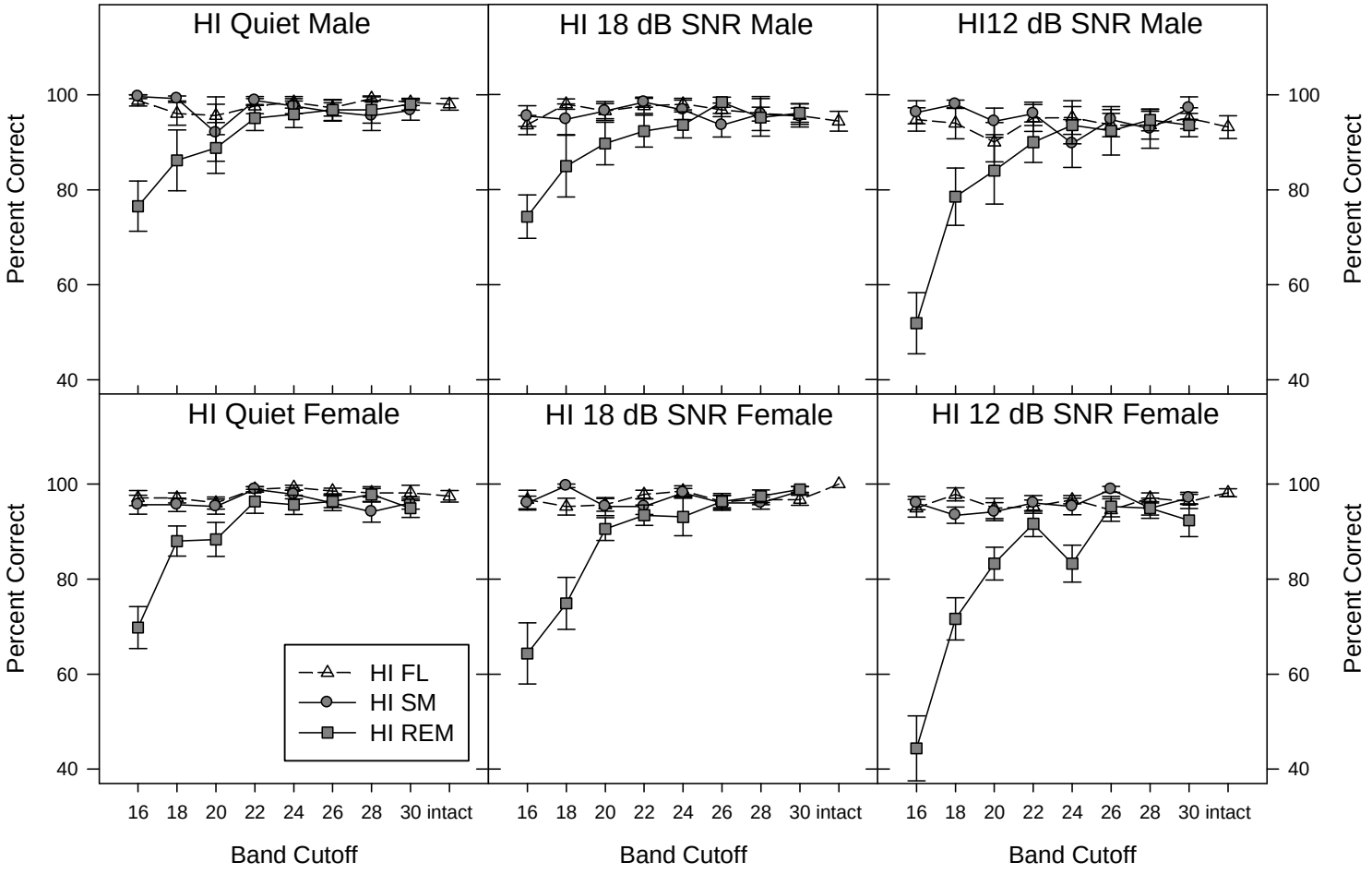


Figure 16. Average results of intelligibility scores plotted as a function of quality ratings for the male talker (top row) and the female talker (bottom row). Scores from the NH group are presented in the left panels, the HI group in the right panels. FL conditions are represented by filled triangles, SM by open circles, and REM by filled squares.

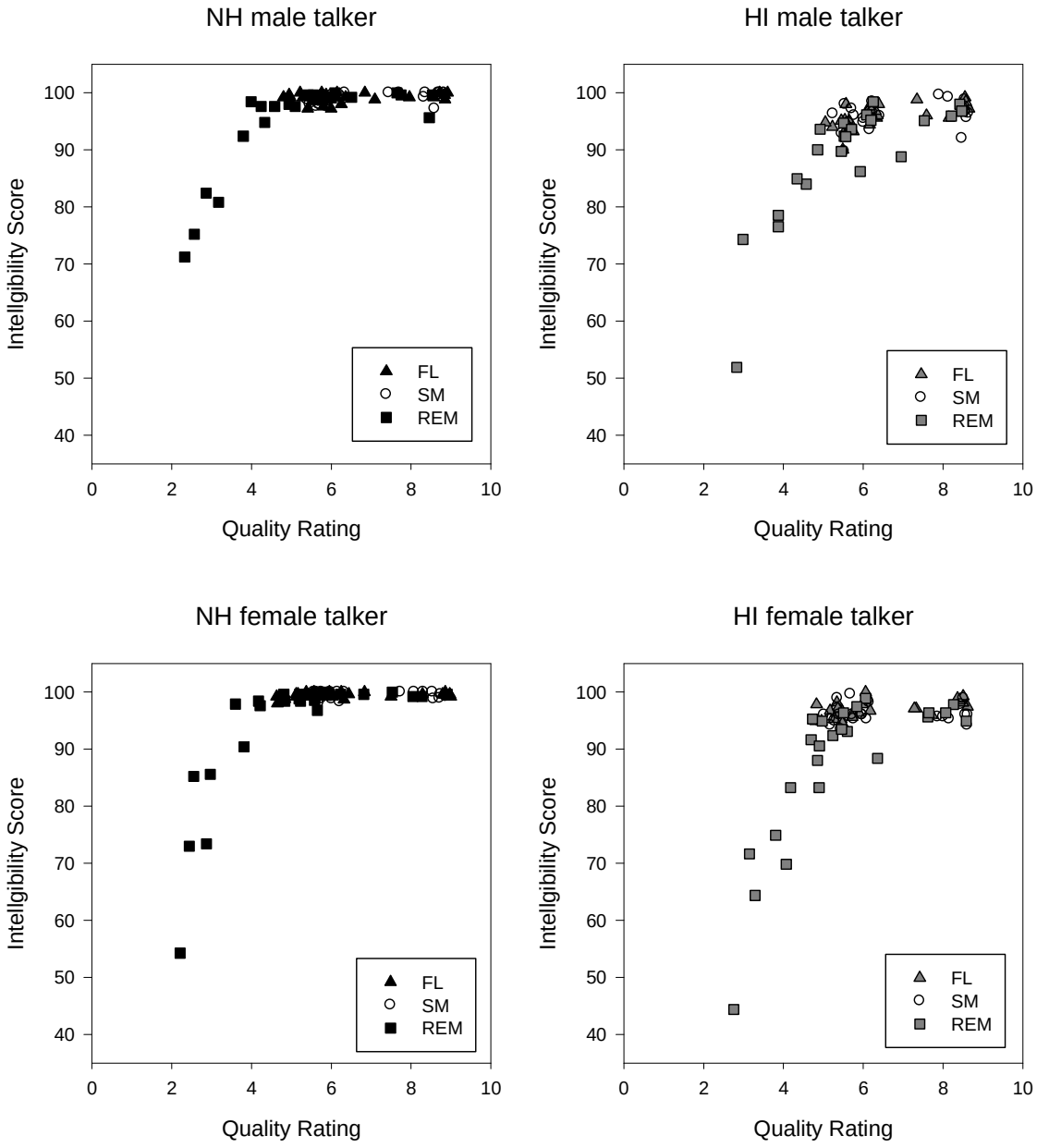


Table 9. Bivariate correlations between intelligibility scores and quality ratings. Significant correlations are highlighted in gray. The specific p-value is indicated within the parentheses.

	NH male	NH female	HI male	HI female
FL noise	0.161 (0.452)	0.23 (0.279)	0.64 (0.001*)	0.676 (<0.001*)
SM noise	0.442 (0.031*)	0.005 (0.98)	0.335 (0.109)	0.042 (0.845)
REM	0.722 (<0.001*)	0.684 (<0.001*)	0.764 (<0.001*)	0.732 (<0.001*)

C**hapter 5: Discussion**

The main goal of this study is to examine the relationship between temporal fine structure and sound quality perception for listeners with normal hearing and listeners with hearing loss. In this chapter, the empirical results are interpreted in the context of this goal. As shown in Figures 9 and 10, processing the speech using vocoding techniques, as well as removal of specific bands, has a measurable effect on sound quality perception. Accordingly, the first aim of this study establishes the nature of the relationship between these signal processing techniques and sound quality perception for speech with different amounts of background noise and different amounts of vocoding.

Outcomes related to specific aim 1

Specific Aim 1: Establish the relationship between temporal fine structure and sound quality perception.

Similar to the methodology of Hopkins et al. (2008), speech is presented to listeners for a range of band cutoff frequencies that determine the amount of high frequency temporal fine structure removal or band removal. In addition to measures of intelligibility, listeners were asked to rate the sound quality of the stimuli using an 11-point scale. Three factors are manipulated in the processing of the speech: 1. type of signal processing (type of vocoding noise, as well as total signal removal), 2. amount of background noise, and 3. band cutoff. Each of these factors will be discussed in terms of their effect on sound quality perception.

Type of Signal Processing

The first manipulated factor is the type of signal processing used on the speech sample, either a vocoding procedure or removal of portions of the signal. In order to study the effects of temporal fine structure removal, this study employs two types of vocoding noise, a fluctuating (FL) noise which has an intact noise envelope, and a smooth (SM) noise which has the noise envelope removed through an additional processing step. As a control condition, a third signal processing type (no signal: REM) is included to measure the effects of signal removal of specific high-frequency bands on sound quality perception. These three signal processing types are intended to provide data representing the effects of temporal fine structure removal and the effects of overall signal bandwidth on quality perception.

In this study, it was found that vocoding the signal had a negative impact on sound quality ratings from both listeners with normal hearing and listeners with hearing loss. That is, removal of temporal fine structure results in decreased quality ratings, indicating that the removal of temporal fine structure is detrimental to sound quality perception. In addition, the type of vocoding noise influences the nature of this impact. Speech with an intact noise envelope (FL) is rated significantly lower than speech without a noise envelope (SM). This finding supports the idea that listeners are sensitive to the presence of the noise envelope in combination with the speech envelope. An intact noise envelope has been shown to be detrimental to speech intelligibility (e.g. Whitmal et al., 2007), and the results from this study indicate that it is also detrimental to sound quality. The intact noise envelope leads to extraneous modulations in the speech envelope which are negatively impact perception.

Historically, the literature supports the assumption that the vocoding process removes temporal fine structure. There are also associated envelope degradations resulting from the vocoding process (Kates, submitted). That is, the speech envelope is affected in two ways. First, the speech envelope is degraded when the signal is divided into bands because the envelope is forced to have the same amplitude within the band. Second, the noise has its own envelope, which is imposed on the speech envelope during the vocoding process. The use of the FL noise leads to greater envelope degradations (resulting from influence of the noise envelope) compared to the SM vocoding noise. Based on the significant, but limited, effects of vocoding on quality ratings, the findings from this study suggest that temporal fine structure removal has no more than a small impact on sound quality perception for both listeners with normal hearing and listeners with hearing loss. Additionally, increased degradation to the speech envelope from the FL vocoding noise envelope increases the amount of quality degradation.

The results from this study also show that removal of the signal in the high-frequencies has a significant detrimental impact on sound quality perception for both listeners with normal hearing and listeners with hearing loss. Quality ratings for listeners with normal hearing and listeners with hearing loss are negatively affected when the speech is low-pass filtered, as in the REM condition. Listeners gave the lowest quality ratings to speech in this condition, compared to vocoding of the signal.

Consideration of a listener's sensitivity to an intact noise envelope may be an important factor when determining the most appropriate type of vocoding noise to use in an experimental design. Signal bandwidth also bears examination when determining an appropriate experimental design.

Amount of background noise

The second factor manipulated in the processing of the speech is the amount of babble noise added to the signal. There are three levels of noise presented to the listeners: speech in quiet, and speech at 18 dB SNR and speech at 12 dB SNR. Based on pilot data, it was discovered that the intelligibility of the speech signal was affected when more noise was added to the signal (poorer SNRs). A great deal of speech quality research shows that the addition of noise is detrimental to sound quality perception (e.g., Arehart et al., 2007; Anderson et al., 2009, Arehart et al., 2010). Additionally, studies of intelligibility using vocoded speech show that as background noise increases, the importance of temporal fine structure to speech understanding also increases (e.g., Qin & Oxenham, 2003; Lorenzi et al., 2006; Başkent, 2006; Hopkins et al., 2008; Hopkins & Moore, 2009).

The results of the study indicate that adding background babble to the speech significantly decreases quality ratings. Unexpectedly, however, the addition of background to the signal reduces the importance of vocoding to the quality ratings. That is, as background babble was added, the effects of decreasing the band cutoff (so an increased amount of the signal was vocoded) are reduced. Listeners are not as sensitive to increased amounts of vocoding when there is background babble added to the signal. These findings are counter to the predicted findings, which are based on the speech intelligibility literature. When listening to speech in the presence of competition, temporal fine structure plays a more important role in speech intelligibility (e.g., Qin & Oxenham, 2003; Lorenzi et al., 2006; Başkent 2006; Hopkins et al., 2008; Hopkins & Moore, 2009). It was expected that as the amount of background noise increased, the importance of temporal fine structure to quality perception would also increase.

However, the addition of noise appears to dominate sound quality perception. Even just a small amount of background noise, where intelligibility remains above 90%, reduces the impact of temporal fine structure removal on sound quality ratings.

Several possible reasons exist for this finding. For example, these results may be due to the increased effects of temporal envelope modification by the background noise. The addition of background noise, even low-level background noise, may mask the more subtle effects of vocoding. In quiet, a listener may be better able to perceive the reduction in temporal fine structure, and its associated temporal envelope effects, from the vocoding process. The addition of a competitor, background babble in this study, has proportionally greater effects on the temporal envelope of the signal, as well as affecting temporal fine structure in the lower frequency regions not impacted by the vocoding process. The addition of the background babble has the effect of smoothing the overall temporal envelope by introducing increased energy into low-level valleys in the target speech, thereby reducing the overall peak-to-valley ratio of the envelope. The background babble may also act as a partial masker of the temporal fine structure of target speech in frequency regions where the temporal fine structure is intact. These larger effects from the addition of background babble may overwhelm the comparatively smaller effects of vocoding process, making vocoding a much more insignificant factor in quality perception.

Frequency region

The third factor manipulated is the band cutoff. As the band cutoff is decreased the amount of vocoding in the signal increased (more of the signal was vocoded). Temporal fine structure is removed from the signal beginning in the high-frequencies because of its limited

utility in speech understanding (e.g., Hopkins et al., 2008). The goal of this specific manipulation is to understand how decreasing the presence of temporal fine structure, without harming intelligibility, affects sound quality perception. By moving in two-band steps, it was possible to examine the effects of temporal fine structure removal in small steps, while maintaining a reasonable number of test conditions.

Overall, increasing the number of bands that are vocoded decreases quality ratings. Of note, however, is the fact that there was not a significant change in quality perception until the signal was altered above 4594 Hz (26 BC). This indicates that the listeners are not sensitive to high-frequency vocoding of the speech signal. The maximum change in quality ratings seen from the intact condition (32 BC) to a cutoff of 4594 Hz (26 BC) was for the female talker in quiet for the HI group, where quality ratings decreased from 8.62 to 8.43, a total decreased of only 0.19 points.

This finding is consistent with the physiological literature regarding the utility of temporal fine structure in high-frequency regions. Based on physiologic data it has been suggested that listeners may not be as sensitive to temporal fine structure above about 5000 Hz (e.g. Joris & Yin, 1992). This decreased sensitivity appears to be the case in this study, as quality ratings did not improve when temporal fine structure was added to the signal above 5000 Hz. If a listener is not as sensitive to temporal fine structure, it may be expected that removal of said fine structure would not impact the perception of speech quality. The present results indicate that our HI group is sensitive to temporal fine structure up to 4594 Hz with regards to quality perception, in contrast with their limited abilities to utilize that region of temporal fine structure above 1500 Hz for speech intelligibility (e.g. Hopkins et al., 2008).

The third type of signal processing, signal removal of specific high-frequencies is also significantly affected by band cutoff. Overall, as with vocoding, increasing the amount of the signal removed decreases quality ratings. This effect is more pronounced for the band removal conditions, where quality ratings decrease by much greater amounts for the same band cutoff level. For example, in the quiet REM condition, quality ratings decreased from a high of 8.8 for the intact signal to 2.9 for the most limited condition. Similar to the vocoding conditions, there is no change in quality ratings for total signal removal above 4594 Hz. The largest decrease in quality ratings from the full intact signal to the 26 BC condition was less than 0.9 points.

Table 10 shows the point at which removal of temporal fine structure becomes significantly detrimental to speech quality perception for speech in quiet for the normal-hearing listeners. As the table shows, removal of high-frequency temporal fine structure above 5000 Hz does not affect speech quality ratings. In addition, speech from the male talker is more resistant to quality degradation compared to the female talker. As expected from the results discussed above, the SM vocoding noise for both talkers is also more resistant to quality degradation. These findings are consistent with the speech intelligibility literature, showing that when lower-frequency temporal fine structure is available there is little benefit to the addition of higher-frequency temporal fine structure (e.g., Hopkins et al., 2008; Hopkins & Moore, 2010). The results presented here are also consistent with the literature related to the physiology of the auditory system, which indicates that listeners are not as sensitive to even the presence of temporal fine structure above 5000 Hz (Heinz & Swaminathan, 2008, 2009; Moore & Sek, 2009b), and that temporal fine structure is itself a weakly coded signal (Swaminathan, 2010).

Table 10. Critical frequency for significantly decreased sound quality perception for the NH group for speech in quiet.

	Male Talker	Female Talker
SM vocoding noise	2590 Hz	4102 Hz
FL vocoding noise	3265 Hz	5140 Hz
REM	6427 Hz	8022 Hz

However, it is interesting to note that quality ratings do not decrease until such a point where there would be expected good access to temporal fine structure. These points of degradation may be related to the acoustics of the specific talkers chosen for this study.

The differences between the male talker and the female talker can be examined in light of their respective acoustic differences. The male talker has a consistently lower fundamental frequency, along with consistently lower formant frequencies (see Figure 5). The removal of temporal fine structure affects the upper formant regions for the female talker at a higher frequency compared to affected formant regions for the male talker. Consequently, sound quality degradation begins at a higher frequency for the female talker. Degradation in sound quality ratings for removal of temporal fine structure begins at 5140 Hz for the female talker and at 3265 for the male taker (see Table 10). For both talkers, these are the frequency regions which correspond to the third formant for some speech sounds. These results are consistent with literature that shows these formants are important for speech understanding (Leek et al., 1987) and speech quality (e.g., Simpson et al., 1990; Baer et al., 1993). Another point for speculation concerns resolved harmonics. F0 for the female talker is about an octave higher than F0 for the male talker. Thus the ability of the auditory system to resolve the harmonics will extend higher in frequency for the female talker given the greater separation between the harmonics.

The results of the limited effect of band-cutoff in the high frequencies fit within the mixed context of the literature. In this study, increasing the low-pass-filter cutoff limited amounts improved quality ratings for both listeners with normal hearing and listeners with hearing loss. Historically, studies show that for listeners with normal hearing, there tends to be a consistent improvement in quality as the bandwidth of a signal increases (e.g., Arehart et al., in

press; Arehart et al., 2010; Ricketts et al., 2008; Moore & Tan, 2003; Gabrielsson et al., 1990). Arehart et al. (2010) reported that increasing the low-pass filter cutoff frequency from 2 kHz to 7 kHz improved speech quality for listeners with normal hearing, while Ricketts et al. (2008) found that increasing the low-pass filter cutoff from 5.5 kHz to 9 kHz improved speech and music quality for listeners with normal hearing. In contrast, quality ratings for increased bandwidth did not improve for listeners with hearing loss in Arehart et al. (2010), while only some listeners with hearing loss judged sound quality to better in for increased bandwidth in Ricketts et al. (2008). In this study, both groups of listeners showed improved quality ratings for extending the bandwidth up to 5 kHz, with no further improvement in quality ratings beyond this point.

Several possible explanations exist for these findings. As discussed above, one possible factor may be the limited utility of temporal fine structure above 5000 Hz. Listeners, even those with normal hearing, are not as sensitive to temporal fine structure above 5000 Hz due to limitations in phase locking (Joris & Yin, 1992). Additionally, recent research modeling of auditory nerve responses to noisy speech has shown that there is limited representation of TFS in the neural coding of a noisy speech sample (Swaminathan, 2010). Given this limited representation, removal of TFS information would be expected to have limited effects on sound quality perception. However, further experiments which examine the role of TFS in quiet are needed to more clearly define possible physiologic sources for the results found in this study.

A second explanation may lie in the high-frequency audibility of the signal. Although these findings suggest that audibility may be at least partly responsible for these findings, the results of the excitation pattern analysis would seem to rule out audibility as a factor in the limited improvement in quality ratings beyond 5000 Hz. For listeners in the NH group and five

of the 10 in the HI group, the speech information above 5500 Hz is audible. Eight of the ten listeners in the HI group have audibility through 7100 Hz. Given that there are no significant between group differences, it seems unlikely that high-frequency audibility is responsible for this lack of change in quality ratings. As a follow-up measure to audibility, the omnibus statistical analysis was re-run with just the group of 15 listeners who had complete audibility of the full 10,000 Hz signal, and no difference in statistical significance for the main effects was found.

A third reason may be the acoustics of the specific speech sample used. Both the male talker and the female talker had speech energy up to 10,000 Hz (see Figure 4 and Figure 5). However, the speech information above 5000 Hz was limited. It may be that given this particular speech sample, the decrease or removal in speech information above 5000 Hz simply is not enough to alter quality perception. The use of a different speech sample, one with more high frequency (> 5000 Hz) content, may reveal differences in quality ratings for the listeners. Additionally, although the quality ratings are not significantly different between male and female talkers, there is a difference in the band-cutoff where quality perception is degraded, as discussed above. It is possible that this difference is due to the difference in formant locations for the two talkers. Although the linguistic content of the talkers was the same, there is a difference in upper formant locations between the talkers, with the female talker having upper formants in higher-frequency regions. Removal of important formant regions would happen at higher frequency regions for the female talker, possibly leading to the faster decline in quality perception.

Outcomes related to specific aim 2

Specific Aim 2: Is there an effect of hearing status on the effects of temporal fine structure removal on quality perception. Does the effect of hearing status differ based on type of vocoding noise, amount of background noise, or frequency region?

The results show that no overall significant differences in quality perception exist between the two groups. That is, listeners with normal hearing and listeners with hearing loss gave similar quality ratings to the conditions presented.

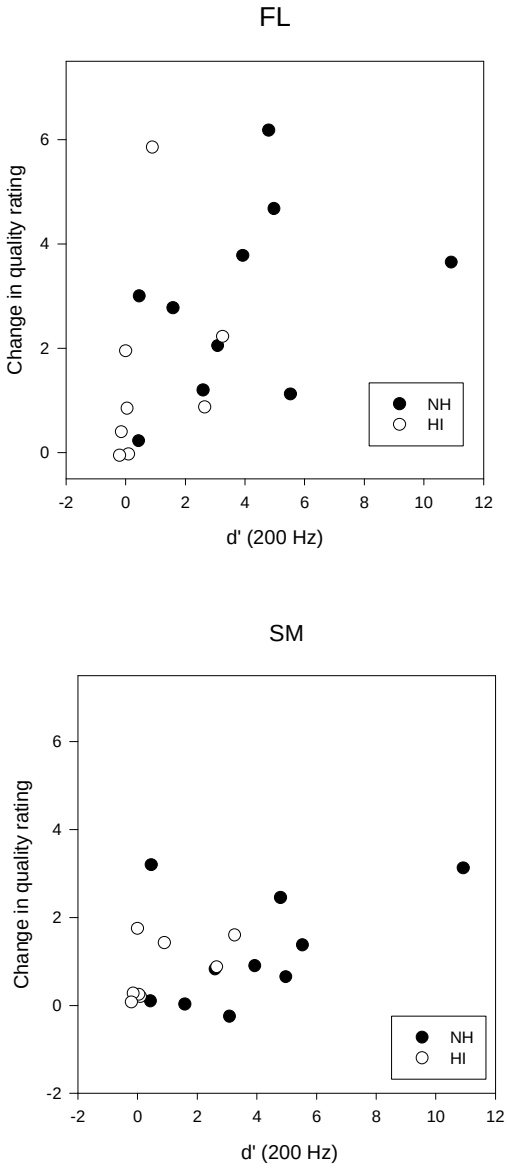
These findings are unexpected given the difference in the abilities between listeners with normal hearing and listeners with hearing loss to utilize temporal fine structure in speech intelligibility. As previous research has shown (e.g., Qin & Oxenham, 2003; Lorenzi et al., 2006; Başkent 2006; Hopkins et al., 2008; Hopkins & Moore, 2009; Hopkins & Moore, 2010), listeners with normal hearing are able to make good use of temporal fine structure even when in the presence of background noise. In fact, temporal fine structure importance increases when in the presence of background noise for listeners with normal hearing. In contrast, intelligibility does not improve as much for listeners with hearing loss when temporal fine structure is added to a speech signal. Quality ratings of listeners with normal hearing were predicted to be more adversely affected by temporal fine structure removal when compared to listeners with hearing loss. Surprisingly, no significant main effect in quality ratings between the groups was found. However, some specific significant interaction terms (e.g. noise type * group for the female talker) show that the groups may have a difference in quality perception thresholds for different aspects of the vocoding procedure.

There is a difference in the abilities of listeners with normal hearing and listeners with hearing loss to detect changes to temporal fine structure using the TFS1 test. Hopkins and Moore

(2010) showed a significant relationship between scores on the TFS1 test and the ability to benefit from temporal fine structure in speech understanding. The better the score on the TFS1 test, the greater the intelligibility benefit from the addition of temporal fine structure to a speech signal. Figure 17 presents a scatter plot of TFS1 scores (horizontal axis) and the amount of change in quality ratings between the least vocoded condition and the most vocoded condition (vertical axis) for the FL noise (top panel) and the SM vocoding noise (bottom panel). Bivariate correlations were not significant for the NH group (FL vocoding noise: $r = 0.388$, $p = 0.268$; SM vocoding noise: $r = 0.466$, $p = 0.197$) or the HI group (FL vocoding noise: $r = 0.261$, $p = 0.532$; SM vocoding noise: $r = 0.519$, $p = 0.187$), and reveal that there is not a relationship between scores of the TFS1 test and change in quality perception from a vocoded speech sample to an intact speech sample.

One possible explanation for the surprising lack of an overall group difference is that modifications to the temporal envelope may be a determining factor in the listeners' quality ratings. Research has shown that modification to the temporal envelope is a strong predictor of quality ratings for both listeners with normal hearing and with hearing loss. It may be possible that while temporal fine structure removal plays a small role in quality perception, the effects of temporal fine structure removal are overshadowed by the effects of temporal envelope modification on quality perception. These effects will be explored in upcoming sections.

Figure 17. Relationship between TFS1 scores for an F0 of 200 Hz (horizontal axis) and the amount of change in quality ratings between the least vocoded condition and the most vocoded condition for the male talker (vertical axis) for the FL noise (top panel) and the SM vocoding noise (bottom panel). Bivariate correlations were not significant for the NH group (FL vocoding noise: r = 0.388, p = 0.268; SM vocoding noise: r = 0.466, p = 0.197) or the HI group (FL vocoding noise: r = 0.261, p = 0.532; SM vocoding noise: r = 0.519, p = 0.187)



Outcomes related to specific aim 3

Specific Aim 3: Establish the relationship between intelligibility and quality ratings.

The results related to specific aim 3 indicate that, as expected, a significant link exists between intelligibility and quality ratings when intelligibility is allowed to vary. For example, in the REM conditions, as intelligibility increased from 44% to 100%, the corresponding average quality rating ranged increased from 2.2 to 8.9. This result is consistent with the results of Preminger and van Tassel (1995a) who found that predicted intelligibility was highly correlated with quality ratings. Using a subjective measure of intelligibility, Preminger and van Tassel reported corresponding increases in quality perception for increases in intelligibility. By directly measuring both intelligibility and quality for the same conditions, this study is able to provide objective data regarding the relationship between intelligibility and quality which supports previous qualitative findings. Results such as these indicate that listeners are likely to use speech intelligibility as a major factor in speech quality ratings. Specifically, when intelligibility is poor, quality ratings will also be poor.

Interestingly, findings from this study indicate that in some situations where intelligibility remains high there is still a correlation with quality ratings. One possible explanation for this correlation may be a difference in the threshold of temporal structure manipulation between speech intelligibility and sound quality perception. Listeners are able to withstand more temporal structure manipulation before speech intelligibility is affected. The threshold for sound quality perception may be lower, leading to larger reductions in sound quality for small reductions in speech intelligibility. For example, the addition of background babble has an impact on the temporal structure of speech. In this study, the addition of background babble at 12 dB SNR has

a very limited effect on speech intelligibility scores. Intelligibility scores for the intact condition decrease from an average of 99% in quiet to an average of 97% in 12 dB SNR. However, it has a larger effect on quality ratings. Average quality ratings for the same condition decreased from a high of 8.7 down to 5.5.

Use of Quality Models

To date, no published studies have explicitly considered the role of temporal fine structure on sound quality perception. However, several studies have indirectly considered the role of the temporal envelope on quality perception. Studies of modeling have shown that quantification of the change in the temporal envelope from a clean signal to a modified signal can be used to accurately predict the quality ratings given by listeners with normal hearing and listeners with hearing loss. While these models provide accurate estimations of sound quality perception, they do not account for the entire picture, as they do not account for all of the variation in quality perception.

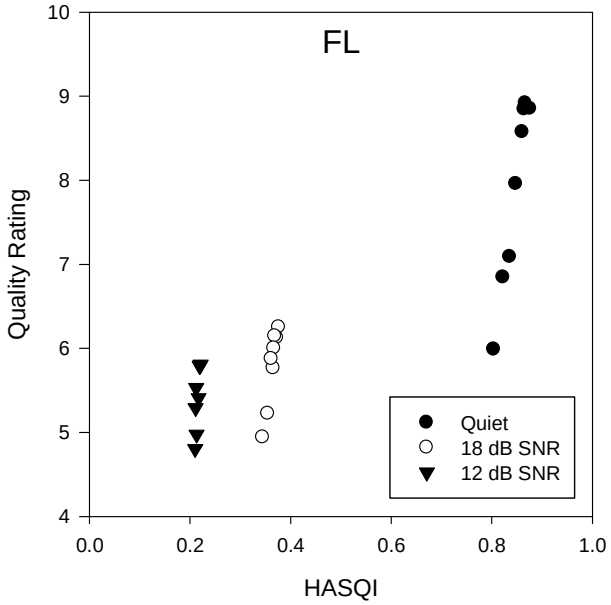
One such model of sound quality is HASQI (Kates & Arehart, 2010), which uses differences in the time-frequency modulations between an unprocessed and a processed speech sample. HASQI had been successfully used to model speech sound quality ratings (Kates & Arehart 2010). HASQI has also been used as a tool to calculate the amount of modification to the temporal envelope of a signal that has undergone temporal fine structure modification through various vocoding processes (Kates, submitted). Given the signal processing confounds that exist with the vocoding procedure for temporal fine structure removal (see Chapter 2: Vocoding), this method provides an objective measure of the amount of temporal envelope modification. Table 10 provides HASQI values for all vocoded conditions for the male talker for

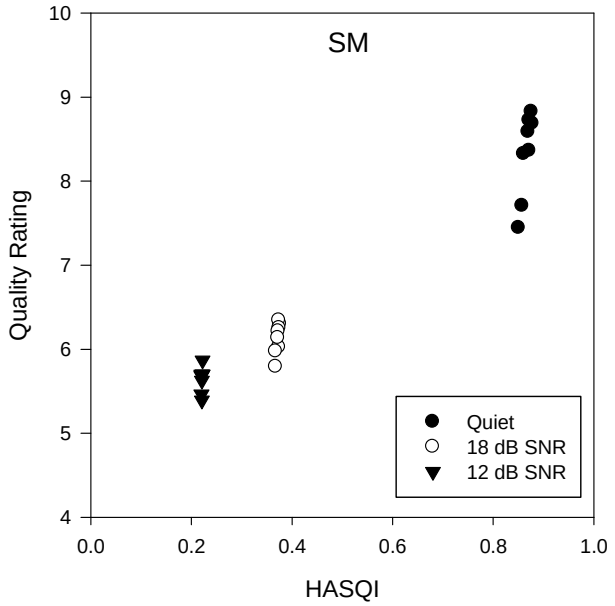
the NH group. Bivariate correlations are calculated between quality ratings for vocoded conditions and the corresponding HASQI values and are included in Table 11. Figure 18 shows scatter plots of the HASQI values as a function of quality ratings for the FL (top panel) and SM vocoding noise for the male talker (bottom panel).

The correlations between HASQI and the quality ratings indicate that envelope modifications are predictive of sound quality ratings for the vocoded signals. The results from this study provide evidence to support using envelope-based modeling metrics. It was suggested above that adding in background noise has a large effect on the envelope of a signal by decreasing the overall peak-to-valley ratio, with the noise filling in the low intensity valleys of the target speech. HASQI provides an objective measure of the amount of this envelope modification. A comparison of the FL-noise 16 BC vocoded speech sample in quiet has a HASQI rating of 0.808, while the same condition at 12 dB SNR has a HASQI rating of 0.201. Compare this to the HASQI rating for FL-noise 30 BC in quiet of 0.876 and at 12 dB SNR of 0.221. The vocoding process (moving from a 30 BC to a 16 BC), with full temporal fine structure removal and slight temporal envelope modifications, are responsible for only a small drop in HASQI values (<0.1). This is consistent with the small change in quality ratings from the 30 BC to the 16 BC conditions in quiet ($\leq$2.2 points) and especially in noise ($\leq$1.3 points). However, the large drop in HASQI due to the addition of background noise (>0.6) is associated with a larger change in temporal envelope structure and larger change in quality ratings (>3 points).

Even with the slight change in HASQI values due to the vocoding process, there is still a measureable effect on quality ratings. Based on the findings of this study, it can be concluded

Figure 18. HASQI values (horizontal axis) and quality ratings (vertical axis). Small changes in HASQI within an SNR are correlated with a small, but measurable, change in quality ratings for increases in the amount of vocoding for both FL and SM vocoding noises.





that full temporal fine structure removal from above 1500 Hz has a small, but measureable, impact on quality ratings for both listeners with normal hearing and listeners with hearing loss. In addition, the data show that even mild to moderate alterations to the temporal envelope have a significant impact on quality perception Interestingly, as modification to the temporal envelope increases (through the addition of background noise), the importance of temporal fine structure to quality ratings decreases for both listeners with normal hearing and listeners with hearing loss.

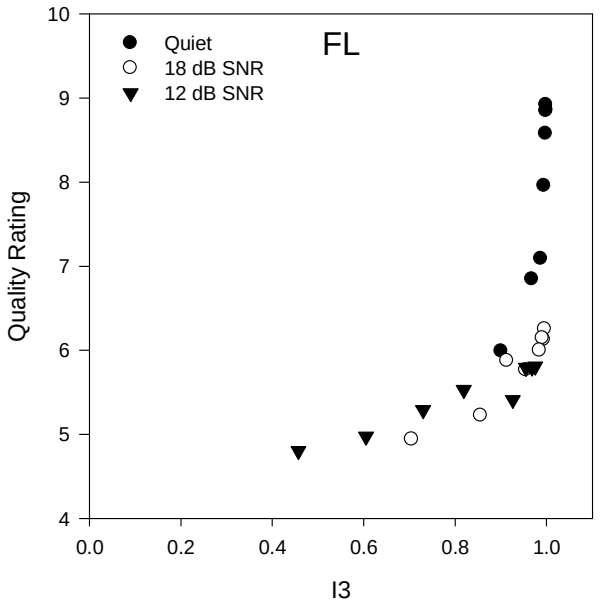
In order to explore this confound between the effect of temporal fine structure and temporal envelope on quality ratings, a metric more directly exploring temporal fine structure changes was also examined for correlations to quality ratings. An additional intelligibility metric, the I3 metric (Kates & Arehart, 2005) is based on the coherence (normalized cross-correlation) between the output and reference signals. Using the Speech Intelligibility Index (SII) 21-band analysis, the coherence is computed for the low-, mid-, and high-level signals. The SII for each of the three signal levels is computed based on the coherence values. The three levels are then combined to predict the intelligibility. The largest weight, for both NH and HI listeners, is for the mid-level SII. For example, if I3 is high (near 1.0), then the cross-correlation between the reference and processed signals must be high at all signal levels. That is, both the signal envelope and temporal fine structure have to match up. I3 is reduced by changes in the envelope modulation, but temporal fine structure changes have a much stronger effect. Figure 19 shows the scatter plots for I3 values plotted as a function of quality ratings, arranged in the same manner as Figure 18. Table 11 provides I3 values for all vocoded conditions for the male talker for stimuli for the NH group. As with HASQI, bivariate correlations were calculated between the

I3 score and quality ratings. A significant relationship exists between I3 and the quality ratings, supporting the conclusion that changes to the temporal fine structure affect quality ratings.

As can be seen the significantly high correlations between quality ratings and the HASQI and I3 metrics, envelope and temporal fine structure changes are highly correlated with quality ratings when processed using the vocoding processing technique. Therefore, it can be said that while both temporal fine structure and temporal envelope play a role in quality perception, the role played by temporal fine structure is measurably smaller. Also, as temporal envelope degradations increase, the influence of temporal fine structure on quality ratings decreases. Given the signal processing confounds that exist with the vocoding process, it is not possible to fully separate from the effects of temporal fine structure from temporal envelope modifications. However, the use of HASQI and I3, as described above, indicate that both aspects of temporal structure do play a measurable role in quality perception.

Several conclusions regarding quality perception emerge from this study. First, listeners are sensitive to removal of temporal fine structure between 1500 and 5000 Hz. Second, the addition of background noise reduces the relevance of temporal fine structure removal across all band cutoff frequencies. The HASQI correlations, as well as the I3 correlations, indicate that both envelope and temporal fine structure changes are related to changes in quality ratings.

Figure 19. I3 values (horizontal axis) and quality ratings (vertical axis). Changes in I3 within an SNR are correlated with a small, but measurable, change in quality ratings for increases in the amount of vocoding for both FL and SM vocoding noises.



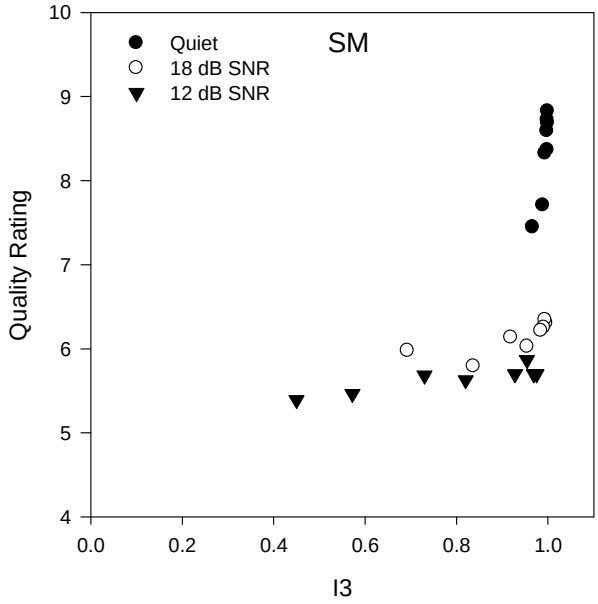


Table 11. HASQI and I3 values were correlated with the quality ratings of the of the NH group for each condition. All correlations were significant at the 0.05 level except for SM vocoding noise at 12 dB SNR.

	HASQI : FL	HASQI/ QR corr: FL	HASQI: SM	QR corr: SM	I3: FL	I3/ QR corr: FL	I3: SM	QR corr: SM
overall		0.868 (< 0.001)		0.854 (<0.001)		0.611 (0.002)		0.541 (0.006)
quiet		0.986 (<0.001)		0.929 (0.001)		0.852 (0.007)		0.891 (0.003)
30 BC	0.876		0.878		1		1	
28 BC	0.867		0.876		0.999		1	
26 BC	0.865		0.872		0.999		0.999	
24 BC	0.861		0.872		0.998		0.999	
22 BC	0.848		0.87		0.995		0.998	
20 BC	0.836		0.861		0.988		0.994	
18 BC	0.823		0.858		0.968		0.989	
16 BC	0.804		0.851		0.901		0.967	
18 dB SNR		0.963 (<0.001)		0.839 (0.009)		0.947 (<0.001)		0.73 (0.04)
30 BC	0.377		0.376		0.996		0.996	
28 BC	0.373		0.374		0.994		0.994	
26 BC	0.369		0.375		0.991		0.991	
24 BC	0.367		0.373		0.985		0.985	
22 BC	0.366		0.374		0.955		0.955	
20 BC	0.362		0.372		0.913		0.919	
18 BC	0.355		0.368		0.856		0.837	
16 BC	0.345		0.368		0.705		0.693	
12 dB SNR		0.833 (0.01)		0.429 (0.288)		0.957 (<0.001)		0.889 (0.003)
30 BC	0.221		0.222		0.975		0.975	
28 BC	0.22		0.22		0.968		0.969	
26 BC	0.219		0.222		0.955		0.954	
24 BC	0.217		0.222		0.926		0.928	
22 BC	0.212		0.221		0.819		0.82	
20 BC	0.211		0.22		0.73		0.73	
18 BC	0.213		0.22		0.605		0.572	
16 BC	0.21		0.221		0.457		0.45	

Conclusions

The goal of the study is to establish the role of temporal fine structure in sound quality perception. Considered in the context of the specific aims, we have established the role of temporal fine structure in quality ratings, quantified the difference between listeners with normal hearing and listeners with hearing loss, and objectively measured the relationship between speech intelligibility and quality perception.

Specific aim 1 seeks to establish the relationship between temporal fine structure and quality perception. The outcomes resulting from specific aim 1 show that removal of temporal fine structure information above 1500 Hz has a small, but measurable, effect on quality ratings. This finding is true for speech in quiet and speech in the presence of a babble background at an 18 and 12 dB SNR. Using an 11-point quality rating scale, for speech in quiet, temporal fine structure removal resulted in changes that range from 9.5 to 7.3, a 2.2-point drop in quality ratings. For speech in the presence of background noise, temporal fine structure removal has a smaller effect, with the biggest change dropping from 6.9 to 5.7, a 1.2-point drop in quality ratings. The overall results indicate that the addition of the multi-talker babble has a greater influence on quality ratings when compared to the influence of temporal fine structure removal. However, it is important to acknowledge that when the temporal fine structure is removed from the signal, there is a coexisting degradation in the temporal envelope. Given this confound, it is difficult to isolate the effects of temporal fine structure removal only on quality ratings. HASQI provides an objective measure of envelope degradation. FL vocoding noise shows more envelope degradation than SM vocoding noise, as would be expected. However, even in the presence of 12 dB SNR, when HASQI changes from least vocoded to most vocoded are less than

0.1, there is still a small drop in quality ratings for increasing vocoding of about 1 point. Therefore, while it can be said that temporal fine structure removal is responsible for no more than a small drop in quality ratings, it remains that it is a factor in these rating decreases.

Specific aim 2 seeks to quantify the effect of hearing status on the role temporal fine structure plays in quality perception. The outcomes resulting from specific aim 2 indicate that there is no overall significant difference in quality ratings between the NH group and the HI group for temporal fine structure removal, although there are some differences in quality perception between groups as evidenced by significant interaction terms (e.g. noise type * group for the female talker). One possible explanation for the lack of main effect of group is that modifications to the temporal envelope may more heavily influence quality ratings, limiting the ability to detect between group differences based on temporal fine structure removal.

Specific aim 3 seeks to objectively quantify the relationship between intelligibility performance and quality ratings for the same processing. The outcomes resulting from specific aim 3 indicate that there is an objective, measurable relationship between quality perception and intelligibility for listeners with normal hearing and listeners with hearing loss. This supports previous research documenting the relationship between subjective intelligibility and quality perception (Preminger & van Tasell, 1995a). This relationship is strongest when objective intelligibility is poor, but remains even for slight decreases in intelligibility performance. Quality ratings are associated with intelligibility performance, indicating that listeners may utilize similar mechanisms in decoding speech for intelligibility performance and quality perception. However, the tolerance threshold appears to be lower for temporal fine structure

removal for quality perception, as there is a greater reduction in quality ratings when compared to the reduction in intelligibility performance for the same processing.

Future Directions

The findings of this study have implications for both objective models of quality perception and the future design of signal processing for hearing aids. These results show that quality ratings from both listeners with normal hearing and with hearing loss are not sensitive to removal of temporal fine structure at frequencies above 5000 Hz for any of the background noise and vocoding noise conditions. However, quality ratings are more significantly affected when temporal fine structure is removed between 1500 and 5000 Hz. The experimental design of this study limits the extent to which we are able to define specific cutoff frequencies above which quality perception is not harmed. Trends exist, however, indicating that the specific type of vocoding noise and specific amount of background noise affect the amount of temporal fine structure removal that is acceptable. Future work should seek to isolate these effects and explore the possibility of defining cutoff frequencies above which temporal fine structure can be removed using various techniques in different environments for listeners with normal hearing and listeners with hearing loss.

Currently HASQI (Kates & Arehart, 2010) considers only the envelope portion of the speech temporal structure. The addition of a temporal fine structure factor to the modeling equation may increase its overall accuracy in predicting quality perception. The implementation of temporal fine structure variable in the modeling equation may help to identify the threshold for temporal fine structure removal. This threshold may differ for listeners with normal hearing and listeners with hearing loss. Although our work does not show a main effect of group

differences, some specific significant interaction terms (e.g. noise type * group for the female talker) show that the groups may have a difference in quality perception thresholds for different aspects of the vocoding procedure.

Future work may also take the form of addressing the effects of temporal fine structure "holes" on quality perception. While this project focused on the high-frequency region only, it may be that there are differences for quality perception if the high-frequency region is left intact, at least partially, and greater amounts of low and mid frequency temporal fine structure are removed. Additionally, given that recent work shows there may be availability high frequency temporal fine structure above 5000 Hz (e.g. Heinz & Swaminathan, 2008, 2009; Moore & Sek, 2009b), it would be useful to determine the perceptual utility of this temporal fine structure on sound quality perception. While in most hearing aid applications the high-frequency region is the most likely to be modified through signal processing, increasing our understanding of the mechanisms of sound quality perception and its relationship to temporal fine structure may allow for anticipated and as yet unanticipated advances in signal processing.

These results may also contribute to advances in hearing aid signal processing design by providing evidence that the threshold for quality perception regarding manipulation to temporal fine structure is quite high. Based on this study, we have objective data indicating that removal of temporal fine structure above 5000 Hz is not detrimental to quality perception, regardless of the type of vocoding noise or amount of background noise.

Several signal processing strategies in use in today's hearing aids function by modifying the temporal structure of the signal. Processes such as dynamic range compression and noise reduction function primarily on the temporal envelope of the signal. However, other processes

such as feedback cancellation and increased bandwidth will also include a more significant impact on temporal fine structure. For example, we know that removing the temporal fine structure at high frequency regions may help to increase the stability of the hearing aid (Ma, 2010). This increased stability will allow for increased gain in these frequency regions without feedback. Future work establishing a processing frequency-cutoff would aid in the design of such systems, by allowing an objective determination of when it would be perceptually safe to remove the temporal fine structure of the original speech while gaining the benefit of increased stability via feedback cancellation and increased spectral bandwidth.

References

American National Standards Institute (ANSI) (2004). Specifications for audiometers *ANSI S3.6*. New York.

Anderson, M. C., Arehart, K. H., & Kates, J. M. (2009). The Acoustic and Perceptual Effects of Series and Parallel Processing. *Eurasip Journal on Advances in Signal Processing*. doi: 619805, 10.1155/2009/619805

Arehart, K., Kates, J., & Anderson, M. (in press) The effects of noise, nonlinear and linear processing on music quality. *International Journal of Audiology*.

Arehart, K., Kates, J., & Anderson, M. (2010) The effects of noise, nonlinear and linear processing on speech quality. *Ear and Hearing, 31,* 420-436.

Arehart, K. H., Kates, J. M., Anderson, M. C., & Harvey, L. O. (2007). Effects of noise and distortion on speech quality judgments in normal-hearing and hearing-impaired listeners. *Journal of the Acoustical Society of America, 122(2),* 1150-1164.

Arehart, K., Souza, P.,Muralimanohar, R., and Miller, C. (in press). Effects of age on concurrent vowel perception in acoustic and simulated electro-acoustic hearing. *Journal of Speech, Language, and hearing Reseach,*

Baer, T., Moore, B., and Gatehouse, S. (1993). Spectral contrast for speech in noise for listeners with sensorineural hearing impairment: effects of intelligibility, quality, and response times. *Journal of Rehabilitation Research, 30,* 49-72.

Baskent, D. (2006). Speech recognition in normal hearing and sensorineural hearing loss as a function of the number of spectral channels. *Journal of the Acoustical Society of America, 120*(5), 2908-2925. doi: 10.1121/1.2354017

Bernstein, J. and Oxenham, A. (2006a) The relationship between frequency selectivity and pitch discrimination: Effects of stimulus level. *Journal of the Acoustical Society of America, 120,* 3916-3928.

Bernstein, J. and Oxenham, A. (2006b) The relationship between frequency selectivity and pitch discrimination: Sensorineural hearing loss. *Journal of the Acoustical Society of America, 120, 3929-3945.*Byrne, D. and Dillon, H. (1996) The national acoustical Laboratories' (NAL) new procedure for selecting the gain and frequency response of a hearing aid. *Ear and Hearing, 7,* 257-265.

Crain, T. R. (1992). *The effect of Peak clipping on the speech recognition threshold.* University of Minnesota, Minneapolis.

Davies-Venn, E., Souza, P., & Fabry, D. (2007). Speech and music quality ratings for linear and nonlinear hearing aid circuitry. *Journal of the American Academy of Audiology, 18*(8), 688-699.

Dillon, H. *(2001). Hearing Aids*. New York: Thieme.

Drullman, R. (1995). Temporal envelope and fine-structure cues for speech-intelligibility. *Journal of the Acoustical Society of America, 97*(1), 585-592.

Dubno, J., Matthews, L., Fu-Shing, L., Ahlstrom, J., & Horwitz, A. (2008). *Predictors of hearing-aid ownership and success by older adults*. Paper presented at the International Hearing Aid Research Conference, Lake Tahoe, CA.

Dudley, H. (1939). Remaking speech. *Journal of the Acoustical Society of America, 11*(2), 169-177.

Gabrielsson, A., Hagerman, B., Bechkristensen, T., & Lundberg, G. (1990). Perceived sound quality of reproductions with different frequency responses and sound levels. *Journal of the Acoustical Society of America, 88*(3), 1359-1366.

Gabrielsson, A., Schenkman, B. N., & Hagerman, B. (1988). The effects of different frequency responses on sound quality judgments and speech-intelligibility. *Journal of Speech and Hearing Research, 31*(2), 166-177.

Gabrielsson, A., & Sjogren, H. (1979). Perceived sound quality of sound-reproducing systems. *Journal of the Acoustical Society of America, 65*(4), 1019-1033.

Ghitza, O. (2001). On the upper cutoff frequency of the auditory critical-band envelope detectors in the context of speech perception. *Journal of the Acoustical Society of America,110,* 1628-1640.Glasberg, B. R., & Moore, B. C. J. (1986). Auditory filter shapes in subjects with unilateral and bilateral cochlear impairments. *Journal of the Acoustical Society of America, 79*(4), 1020-1033.

Glasberg, B. R., & Moore, B. C. J. (1990). Derivation of auditory filter shapes from notched-noise data. *Hearing Research, 47*(1-2), 103-138.

Gordon-Salant, S. , and Fitzgibbons, P. (1993). Temporal factors and speech recognition performance in young and elderly listeners. *Journal of Speech and Hearing Research, 36,* 1276-1285.

Gordon-Salant, S. and Fitzgibbons, P. (1999). Profiles of auditory temporal processing in older listeners. *Journal of Speech, Language, and Hearing Research, 42,* 300-311.,

Hansen, M. & Kollmeier, B. (1997). Using a quantitative psychoacoustical signal representation for objective speech quality measures. *IEEE*.

Harrison, R and Evans, E. (1979). Cochlear fiber response in guinea pigs with well defined cochlear lesions. *Scandinavian Audiology, 9*, 89-92.Heinz, M. and Swaminathan, J. (2008). Neural cross-correlation metrics to quantify envelope and fine-structre coding in auditory-nerve responses. *Journal of the Acoustical Society of America. 123*, 3056.

Heinz, M. and Swaminathan, J.(2009) Quantifying envelope and fine-structure coding in auditory nerves responses to chimeric speech. *Journal of the Association for Research in Otolaryngology 10*, 407-423.

Hopkins, K., & Moore, B. C. J. (2007). Moderate cochlear hearing loss leads to a reduced ability to use temporal fine structure information. *Journal of the Acoustical Society of America, 122*(2), 1055-1068. doi: 10.1121/1.2749457

Hopkins, K., & Moore, B. C. J. (2009). The contribution of temporal fine structure to the intelligibility of speech in steady and modulated noise. *Journal of the Acoustical Society of America, 125*(1), 442-446. doi: 10.1121/1.3037233

Hopkins, K., & Moore, B. C. J. (2010). The importance of temporal fine structure information in speech at different spectral regions for normal-hearing and hearing-impaired subjects. *Journal of the Acoustical Society of America, 127(3),* 1595-1608.

Hopkins, K., Moore, B. C. J., & Stone, M. A. (2008). Effects of moderate cochlear hearing loss on the ability to benefit from temporal fine structure information in speech. *Journal of the Acoustical Society of America, 123*(2), 1140-1153. doi: 10.1121/1.2824018

Huber, R., & Kollmeier, B. (2006). PEMO-Q--A new method for objective audio quality assessment using a model of auditory perception. *IEEE, 14*(6), 1902-1911.

Jenstad, L. M., & Souza, P. E. (2005). Quantifying the effect of compression hearing aid release time on speech acoustics and intelligibility. *Journal of Speech Language and Hearing Research, 48*(3), 651-667.

Joris, P. X., & Yin, T. C. T. (1992). Responses to amplitude-modulated tones in the auditory-nerve of the cat. *Journal of the Acoustical Society of America, 91*(1), 215-232.

Kates, J. (2008). *Digital Hearing Aids*. San Diego, CA: Plural Publishing.

Kates, J. (submitted for publication). Envelope changes caused by fine structure modification. *Journal of the Acoustical Society of America.*

Kates, J., & Arehart, K. (2010). The Hearing Aid Speech Quality Index (HASQI). *Journal of Audio Engineering Society, 58(5),* 363-381.

Kates, J. M., & Arehart, K. H. (2007). *A time-frequency modulation model of speech quality*. Paper presented at the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New Paltz, NY.

Kates, J. M., & Kozma-Spytek, L. (1994). Quality ratings for frequency-shaped peak-clipped speech. *Journal of the Acoustical Society of America, 95*(6), 3586-3594.

Killion, MC, Niquette, PA, Gudmundsen, GI, Revit, LJ, and Banerjee, S (2004). Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners. *Journal of the Acoustical Society of America, 116(4),* 2395-2405.

Kochkin, S. (2005a). Customer satisfaction with hearing instruments in the digital age. *The Hearing Journal/MarkeTrak VII, 58*(9), 30-37.

Kochkin, S. (2005b). Hearing loss population tops 31 million. *Hearing Review, 12*, 16-29.

Kohlrausch, A., Fassal, R., van der Heijden, M., Kortekaas, R., van de Par, S., Oxenham, A., and Puschel, D. (1997). Detection of tones in low-noise noise: Further evidence for the role of envelope fluctuations. *Acta Acoustica, 83,* 659-669.

KozmaSpytek, L., Kates, J. M., & Revoile, S. G. (1996). Quality ratings for frequency-shaped peak clipped speech: Results for listeners with hearing loss. *Journal of Speech and Hearing Research, 39*(6), 1115-1123.

Larson, V. D., Williams, D. W., Henderson, W. G., Luethke, L. E., Beck, L. B., Noffsinger, D., et al. (2000). Efficacy of 3 commonly used hearing aid circuits - A crossover trial. *JAMA-Journal of the American Medical Association, 284*(14), 1806-1813.

Lawson, G. D., & Chial, M. R. (1982). Magnitude estimation of degraded speech quality by normal-hearing and impaired-hearing listeners. *Journal of the Acoustical Society of America, 72*(6), 1781-1787.

Leek, M., Dorman, M., and Summerfield, A. (1987). Minimum spectral contrast for vowel identification by normal hearing and hearing impaired listeners. *Journal of the Acoustical Society of America, 81,* 148-154.

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.

Lunner, T., Hellgren, J., Arlinger, S., & Elberling, C. (1998). Non-linear signal processing in digital hearing aids. *Scandinavian Audiology, 27*, 40-49.

Ma, G. (2010). *Feedback suppression in digital hearing aids.* Unpublished doctoral dissertation, Technical University of Denmark, Lyngby, Denmark.

Moore, B. C. J., & Sek, A. (2009a). Development of a fast method for determining sensitivity to temporal fine structure. *International Journal of Audiology, 48*(4), 161-171. doi: 10.1080/14992020802475235

Moore, B.C.J. and Sek, A. (2009b). Sensitivity of the human auditory system to temporal fine structure at high frequencies. *Journal of the Acoustical Society of America, 125,* 3186-3193.

Moore, B. C. J., & Tan, C. T. (2008). Perceived naturalness of spectrally distorted speech and music. *Journal of the Acoustical Society of America, 114*(1), 408-419.

Moore, B. C. J., & Tan, C. T. (2004). Development and validation of a method for predicting the perceived naturalness of sounds subjected to spectral distortion. *Journal of the Audio Engineering Society, 52*(9), 900-914.

Neuman, A. C., Bakke, M. H., Hellman, S., & Levitt, H. (1994). Effect of compression ratio in a slow-acting compression hearing-aid - paired-comparison judgments of quality. *Journal of the Acoustical Society of America, 96*(3), 1471-1478.

Neuman, A. C., Bakke, M. H., Mackersie, C., Hellman, S., & Levitt, H. (1995). Effect of release time in compression hearing aids: Paired-comparison judgments of quality. *Journal of the Acoustical Society of America, 98*(6), 3182-3187.

Neuman, A. C., Bakke, M. H., Mackersie, C., Hellman, S., & Levitt, H. (1998). The effect of compression ratio and release time on the categorical rating of sound quality. *Journal of the Acoustical Society of America, 103*(5), 2273-2281.

Nilsson, M., Soli, S. D., & Sullivan, J. A. (1994). Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise. *Journal of the Acoustical Society of America, 95*(2), 1085-1099.

Oxenham, A. and Bacon, S. (2003). Cochlear Compression: perceptual measures and implications for normal and impaired hearing. *Ear & Hearing, 24,* 352-366.

Pittman, A. L., Lewis, D. E., Hoover, B. M., & Stelmachowicz, P. G. (2005). Rapid word-learning in normal-hearing and hearing-impaired children: Effects of age, receptive vocabulary, and high-frequency amplification. *Ear and Hearing, 26*(6), 619-629.

Plomp, R. (1964). The ear as a frequency analyzer. *Journal of the acoustical Society of America,36,* 1628-1636.Pichora-Fuller, M.K. (2003). Processing speech and timing in aging adults: psychoacoustics, speech perception, and comprehension. *International Journal of Audiology, 42,* S59-S67

Preminger, J. E., & Van Tasell, D. J. (1995a). Quantifying the relation between speech quality and speech-intelligibility. *Journal of Speech and Hearing Research, 38*(3), 714-725.

Preminger, J. E., & Van Tasell, D. J. (1995b). Measurement of speech quality as a tool to optimize the fitting of a hearing-aid. *Journal of Speech and Hearing Research, 38*(3), 726-736.

Pumplin, J. (1985). Low-noise noise, *Journal of the Acoustical Society of America, 78,* 100-104.

Qin, M. K., & Oxenham, A. J. (2003). Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers. *Journal of the Acoustical Society of America, 114*(1), 446-454.

Ricketts, T., & Hornsby, B. (2005). Sound quality measures for speech in noise through a commercial hearing aid implementing "Digital Noise Reduction". *Journal of the American Academy of Audiology, 16,* 270-277.

Ricketts, T. A., Dittberner, A. B., & Johson, E. E. (2008). High-frequency amplification and sound quality in listeners with normal through moderate hearing loss. *Journal of Speech Language and Hearing Research, 51*(1), 160-172. doi: 10.1044/1092-4388(2008/012)

Rose, J., Brugge, J., Anderson,D., and Hind, J. (1967). Phase locked response to to low-frequency tones in in single auditory nerve fibers of the squirrel monkey. *Journal of Neurophysiology, 30,* 769-793.

Rosen, S. (1992). Temporal information in speech - acoustic, auditory and linguistic aspects. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences, 336*(1278), 367-373.

Rosenthal, S. (1969). IEEE: Recommended Practices for Speech Quality Measurements. *IEEE Transactions on Audio and Electroacoustics, 17,* 227-246.

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science, 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. doi: 10.1044/1092-4388(2008/07-0196)

Simpson, A., Moore, B., and Glasberg, B. (1990). Spectral enhancement to improve the intelligibility of speech in noise for hearing impaired listeners. *Acta otolaryngology, S469,* 101-107.Slaney, M. (1993). An efficient implementation of the Patterson-Holdsworth auditory filter bank. Apple Computer Technical Report #35. Cupertino, CA: Apple Computer Library.

Souza, P. E. (2002). Effects of compression on speech acoustics, intelligibility, and sound quality. *Trends in Amplification, 6*(4), 131-165.

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. doi: 10.1044/1092-4388(2006/011)

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.

Stelmachowicz, P. G., Lewis, D. E., Choi, S., & Hoover, B. (2007). Effect of stimulus bandwidth on auditory skills in normal-hearing and hearing-impaired children. *Ear and Hearing, 28*(4), 483-494.

Stelmachowicz, P. G., Lewis, D. E., Hoover, B., & Keefe, D. H. (1999). Subjective effects of peak clipping and compression limiting in normal and hearing-impaired children and adults. *Journal of the Acoustical Society of America, 105*(1), 412-422.

Stelmachowicz, P. G., Nishi, K., Choi, S., Lewis, D. E., Hoover, B. M., Dierking, D., et al. (2008). Effects of Stimulus Bandwidth on the Imitation of English Fricatives by Normal-Hearing Children. *Journal of Speech Language and Hearing Research, 51*(5), 1369-1380. doi: 10.1044/1092-4388(2008/07-0115)

Stelmachowicz, P. G., Pittman, A. L., Hoover, B. M., & Lewis, D. E. (2002). Aided perception of vertical bar s vertical bar and vertical bar z vertical bar by hearing-impaired children. *Ear and Hearing, 23*(4), 316-324. doi: 10.1097/01.aud.0000027406.51909.06

Stelmachowicz, P. G., Pittman, A. L., Hoover, B. M., Lewis, D. E., & Moeller, M. P. (2004). The importance of high-frequency audibility in the speech and language development of children with hearing loss. [Proceedings Paper]. *Archives of Otolaryngology-Head & Neck Surgery, 130*(5), 556-562.

Stone, M. A., & Moore, B. C. J. (1992). Spectral feature enhancement for people with sensorineural hearing impairment - effects on speech-intelligibility and quality. *Journal of Rehabilitation Research and Development, 29*(2), 39-56.

Stone, M. A., & Moore, B. C. J. (2007). Quantifying the effects of fast-acting compression on the envelope of speech. *Journal of the Acoustical Society of America, 121*(3), 1654-1664.

Swaminathan, J. (2010) The Role of Envelope and Temporal Fine Structure in the Perception of Noise Degraded Speech. unpublished doctoral dissertation, Purdue University, West Lafayette, IN.

Tan, C. T., & Moore, B. C. J. (2003). The effect of nonlinear distortion on the perceived quality of music and speech signals. *Journal of the Audio Engineering Society, 51*(11), 1012-1031.

van Buuren, R. A., Festen, J. M., & Houtgast, T. (1999). Compression and expansion of the temporal envelope: Evaluation of speech intelligibility and sound quality. *Journal of the Acoustical Society of America, 105*(5), 2903-2913.

Versfeld, N. J., Festen, J. M., & Houtgast, T. (1999). Preference judgments of artificial processed and hearing-aid transduced speech. *Journal of the Acoustical Society of America, 106*(3), 1566-1578.

Whitmal, N., Poissant, S., Freyman, R., and Helfer, K. (2007). Speech intelligibility in cochlear implant simulations: Effects of carrier type, interfering noise, and subject experience. *Journal of the Acoustical Society of America, 122(4),* 2376-2388.

Woolf, N. K., Ryan, A. F., & Bone, R. C. (1981). Neural phase-locking properties in the absence of cochlear outer hair-cells. *Hearing Research, 4*(3-4), 335-346.

Young, E. and Sachs, M. (1979). Representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of auditory nerve fibers. *Journal of the Acoustical Society of America, 66,* 1381-1403.

Appendix: Subject Instructions

In addition to verbal instructions, these instructions are also printed and left with the listener for reference during the experimental session.

TFS1 Test Instructions

In this task, each trial will contain two intervals in which sounds will play. Each interval will consist of a background noise with 4 tones. In one interval the pitch of the tones will be consistent. In the other, the pitch of the tones will fluctuate. Your task it to decide if interval 1 or interval 2 contains the fluctuating tones, by using the mouse to click on the corresponding box.

If your choice is correct, the box will flash green briefly. If your response is wrong, the box will flash red briefly. After a short delay the next trial will start. Please always make a response even if you think you are guessing.

Instructions provided to the listeners are reproduced below. In addition to verbal instructing these instructions are also printed and left with the listener for reference during the experimental session.

Intelligibility Instructions

In this experiment you will be listening to sentences that have been digitally processed. Some of the sentences may be difficult understand, or sound "fuzzy". Your task is to repeat as much of the sentence as you understood, even if only one or two words.

To begin sentence playout please click on the button marked PLAY. When the speaker has finished talking, please repeat back as much of the sentence as you understood. If you did not understand any of the words please say "I understood nothing". To begin the next speech sample click PLAY. You will be given a break at the end of each block of trials. If you would like a break before the end of the block, do not click PLAY.

Instructions provided to the listeners are reproduced below. In addition to verbal instructing these instructions are also printed and left with the listener for reference during the experimental session. The instructions were adapted from Gabrielsson et al. (1998) and Davies-Venn et al. (2007).

Quality Instructions

Your task today is to judge the sound quality of the programs you listened to in the previous sessions. You shall now try to describe how they sound by means of an overall impression scale. It is graded from 10 (maximum) to 0 (minimum). You decide yourself on the accuracy that you consider necessary.

As you can see it is also possible to use decimals. The integers 9, 7, 5, 3, and 1 are defined on the scale. 10 means maximum (highest possible) sound quality, 9 means very good, 7 rather good, 5 midway, 3 rather bad, 1 very bad, and 0 minimum (lowest possible) sound quality.

To begin each trial, click on the button marked PLAY. After the sample has ended, mark your rating on the slider bar using the mouse. Click CONFIRM to indicate that you have made a final decision. Click PLAY again to begin the next trial. If you would like a break before the end of the block, do not click PLAY.