Spectral Resolution and Speech Recognition in Noise by Cochlear Implant Users

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Elizabeth Susan Anderson

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Peggy B. Nelson, Ph.D., Robert S. Schlauch, Ph.D.

July, 2011

Acknowledgements

My heartfelt gratitude goes to my thesis advisers, Dr. Peggy Nelson and Dr. Andrew Oxenham, for their mentoring, support, and guidance in the process of learning how to conduct and write about research. I am so fortunate to have had this opportunity to work with them both.

Thanks to the other members of my dissertation committee: Dr. Bert Schlauch for bringing both rigor and wit to our study of the literature, and Dr. Aparna Rao for her fresh perspectives and helpful comments.

I am very grateful to Dr. David A. Nelson, in whose lab I worked many years ago as an audiology master's student, and who took me in again and supported me as a doctoral student, providing the resources and infrastructure that enabled me to do this dissertation work.

I could not have completed my doctoral program without the help of a number of people. I am ever grateful to Heather Kreft for her substantial assistance, encouragement, and camaraderie. Many thanks to Dr. Edward Carney who provided invaluable computer and signal processing support, to Dr. Christophe Micheyl for his statistical support and many helpful conversations, and to Dr. Magda Wojtczak for programming assistance.

I have appreciated the spirit of collegiality and fellowship among the doctoral students (current and former) in Shevlin Hall, including Yingjiu Nie, Sharon Miller, Melanie Gregan, Evelyn Davies-Venn, and many others who inspired and supported me through the challenges of being a doctoral student.

I have been fortunate to receive funding for my doctoral work through NIH-NIDCD grants R01DC06699-01A1 and R01-DC008306, two Bryngelsen departmental awards, a College of Liberal Arts Graduate Research Partnership Project (GRPP) award, and the Departments of Speech-Language-Hearing Sciences and Psychology.

Finally, I am grateful to all of the cochlear-implant subjects who participated in this research.

Dedication

This dissertation is dedicated to my children, Becky, Dan, and Sarah, whose unfailing encouragement and belief in me helped me to believe in myself, and to my parents, Nels and Sue, now deceased, who gave me every opportunity.

Abstract

For cochlear-implant (CI) users, the relationship between spectral resolution and speech perception in noise has remained ambiguous. An even more fundamental question has been how to measure spectral resolution in CI listeners. This dissertation describes work exploring the relationships among different measures of spectral resolution, and between each of those measures and speech recognition in quiet and in noise. Spectral ripple discrimination was found to correlate strongly with spatial tuning curves when the measures were matched in frequency region. Broadband spectral ripple discrimination correlated well with sentence recognition in quiet, but not in background noise. Spectral ripple detection correlated strongly with speech recognition in quiet, but its validity as a measure of spectral resolution was not empirically supported. Spectral ripple discrimination thresholds were compared to sentence recognition in noise, using spectrally-limited maskers that did not overlap with the entire speech spectrum. Speech reception thresholds were measured in the presence of four low- or high-frequency maskers, all bandpass-filtered from speech-shaped noise, and a broadband masker encompassing most of the speech spectrum. The findings revealed substantial between-subject variability in susceptibility to masking by each of these noises and in spectral release from masking, which cannot be explained simply in terms of energetic masking and does not appear to be strongly related to spectral resolution. Better CI users appeared to show stronger relationships between spectral resolution and speech perception than did poorer users, implying that advanced CI processing strategies designed to maximize the number of spectral channels may not benefit all CI users equally.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

CHAPTER I: INTRODUCTION

1. OVERVIEW

According to FDA figures, 219,000 people worldwide had received cochlear implants as of December, 2010. Cochlear implant (CI) technology has advanced significantly since FDA approval in the early 1980s, and today the majority of CI users achieve relatively high levels of speech understanding in quiet backgrounds (e.g. Gifford, Shallop, and Peterson, 2008), but CI users' susceptibility to interference from background noise is great (e.g. Dorman, Loizou, Fitzke, and Tu, 1998; Fu and Nogaki, 2005; Garnham, Driscoll, Ramsden, and Saeed, 2002).

Efforts to identify the underlying sources of this difficulty understanding speech in noise have motivated much research in the field of CIs. One approach has been to examine the relationships between non-speech, psychophysical measures of spectral/temporal resolution and various measures of speech recognition. Earlier work in our own laboratory has failed to demonstrate significant correlations between certain psychophysical measures and speech perception in noise by CI users (Anderson, Nelson, Kreft, Nelson, and Oxenham, 2011; Nelson, Kreft, Anderson, and Donaldson, 2011). The experiments described in this thesis attempt to elucidate the nature of the interference produced by different types of noises, and to further explore the roles of spectral and temporal resolution in speech perception in noise by CI users.

II. BACKGROUND

A. Role of spectral resolution in speech perception

Spectral (or frequency) information is critical for speech recognition; cues for vowel identity (minimally, the first two formant frequencies) and consonant place of articulation are primarily contained in the frequency domain. Even so, speech contains enough redundancy in the acoustic, phonetic, and linguistic domains that recognition of speech in quiet is possible with minimal frequency resolution (e.g. Shannon, Zeng, Kamath, Wygonski and Ekelid, 1995). In CIs, the accurate representation of spectral information can be dependent on two factors: the *number of spectral channels* into which the input signal is subdivided, determined by hardware and processing parameters of the CI system, and by the *degree of spectral smearing* of the stimulus, reflecting channel overlap due to current spread, which is primarily related to physiological factors. Studies exploring the effects of number of channels and spectral smearing on speech perception have, in general, supported the conclusion that these factors influence speech recognition performance to some degree (e.g. Shannon *et al*., 1995; Dorman, Loizou, and Rainey, 1997; Friesen, Shannon, Baskent,, and Wang, 2001).

Number of channels

On average, 4 to 7 spectral channels have been shown experimentally to be necessary for asymptotic performance by normal-hearing subjects listening to vocoded simulations in quiet, depending on the materials used.

Shannon *et al*. (1995) demonstrated that speech recognition for normal-hearing listeners can be maintained in spite of greatly reduced spectral information, primarily through the use of temporal information. They used a vocoding technique, dividing the acoustic signal into several frequency bands, extracting the amplitude envelope from each one, and using each of those envelopes to modulate a band-limited noise matched in frequency bandwidth to the analysis band. In this way, fine spectral information was removed from speech but temporal envelope information was retained. Shannon and colleagues investigated the effect of varying the number of spectral channels on vowel, consonant, and sentence identification by normal-hearing listeners. They found that speech recognition performance for all materials increased monotonically as number of channels increased, but as few as 4 time-varying bands of noise that simulated the spectral patterns of speech could produce relatively high levels of performance.

Dorman *et al*. (1997) studied the performance of eight normal-hearing subjects listening to vowels, consonants, and sentences that were processed in manner similar to Shannon *et al*. (1995) to simulate from 2 to 9 effective spectral channels. Results suggested that the number of channels necessary for good speech recognition varies, depending on the materials, with multi-talker vowels being the most difficult materials, requiring 8 channels for asymptotic performance, and sentences the easiest, requiring 5. The authors noted that extrapolating these results to the design of speech processors for cochlear implants would suggest that 8 channels might be optimal.

In background noise, however, more channels are necessary for speech recognition by normal-hearing listeners. Fu, Shannon, and Wang (1998) investigated the recognition of vowels and consonants by normal-hearing subjects as a function of signal-to-noise ratio, with number of spectral channels as a parameter. Results indicated that vowel recognition decreased monotonically for all processing conditions as the signal-to-noise ratio (SNR) decreased and as the number of effective channels decreased. The authors concluded that relatively good recognition of speech sounds by normal-hearing listeners was preserved with acoustic information limited to only 4 spectral bands in quiet listening situations, but when noise was added, more channels were needed to maintain the same level of performance. The implication of this study was that while listeners can tolerate significant spectral degradation in quiet settings, the loss of spectral information has a greater effect on speech recognition in noisy backgrounds, and the number of channels at which performance asymptotes will vary with SNR.

Dorman, Loizou, Fitzke, and Tu (1998) also investigated the effects of number of spectral channels and SNR in normal-hearing subjects listening to vocoded simulations of speech. Using sentence materials, they showed that as the SNR became worse, more channels were necessary to reach a given level of speech recognition performance. 5 channels of stimulation were sufficient to reach asymptotic performance in quiet, while 12 channels were required at +2 dB SNR, and 20 channels at -2 dB SNR.

Interestingly, while the speech recognition performance of normal-hearing listeners in noise has been shown to continue to improve with the addition of more channels up to 16 - 20, the performance of CI users in noise does not appear to improve significantly beyond about 8 (e.g. Fishman *et al.*, 1997). Friesen *et al*. (2001) looked at speech recognition in CI users as a function of number of electrodes in several levels of noise, with speech processing strategy as the parameter. Results indicated that at high (favorable) SNRs, sentence understanding was relatively good for 3 or more channels;

however, the effect of a reduction in SNR was comparable to the effect of a reduction in the number of channels. The authors concluded that with 20 electrodes/channels, better-performing CI users require a 5 to 10 dB more favorable SNR than normal-hearing simulation listeners to obtain equivalent performance. Based on the slopes of the performance-intensity (P-I) functions, a 5 dB improvement in SNR could result in a 30 - 50% improvement in sentence recognition. Interestingly, the scores of better-performing CI users improved as the number of electrodes increased up to 7, while poorer-performing CI users' performance peaked at about 4 electrodes. This trend was particularly evident with speech materials such as vowels, which require relatively more spectral information, and at poorer SNRs. These results suggest an inherent limitation to the potential benefit of increasing the number of channels in current cochlear implants.

Spectral smearing

Along with number of channels, another limiting factor in spectral resolution is spectral smearing, which causes a reduction in the spectral contrast of stimuli. The capacity to resolve and identify the frequencies of spectral peaks in speech sounds is important in speech recognition, especially that of vowels. Spectral smearing is thought to relate to the poorer ability of CI users to resolve spectral peaks, compared with normal-hearing listeners, and is most likely a function of shallower auditory filter slopes and the resulting channel overlap. In a simple auditory model, when auditory filter slopes are broad, stimulating one channel can result in a "spread" of stimulation into adjacent/nearby channels, resulting in channel interactions. In effect, individual channels become less independent. In CIs, channel interaction can occur due to uneven neural density, neural “dead regions”, irregularities in impedance characteristics of nerve fibers along the electrode array, and the specificity with which local groups of neurons respond to their characteristic frequency.

In CIs, one way in which frequency information is coded is via the spatial location of each electrode in the array, which capitalizes on the tonotopic frequency organization of neural elements along the cochlea. Several factors can affect the cochlear implant’s presumed electrode-to-place assignment of frequency information. One is the displacement of neural stimulation from the intended neural targets. Uneven auditory nerve survival may exist, where there is not a consistent density of nerve fibers along the cochlea (Zeng, Popper, and Fay, 2004). In this case, stimulation of an electrode near a “dead spot,” or area with no functioning nerve fibers, may result in stimulation of distant existing auditory fibers on either side of the desired location. Alternately, there may be irregularities in impedance characteristics of specific neural elements, causing undesired redirection of electrical current through pathways of lower impedance.

Even in the absence of such irregularities, the spatial spread of current fields generated by individual stimulated electrodes can affect the accurate representation of frequency information in the cochlear implant. In other words, stimulation of one electrode might result in the excitation of more than just one local group of nerve fibers. Current fields that are not narrowly focused may overlap with adjacent frequency channels. The degree of “spatial selectivity” and the amount of channel overlap that exists will affect the maximum number of electrodes/channels that a particular cochlear implant device can functionally use. In addition, the radial distance between the implanted electrode array and the modiolus, where the neural elements reside, is also a factor. One device-related factor governing spatial selectivity is mode of electrical stimulation, with a bipolar configuration allowing for narrower current fields than a monopolar system. The fairly recent development of tripolar configuration (Jolly, Spelman and Clopton, 1996; Mens and Berenstein, 2005) allows for current steering to create “virtual channels” between electrode sites, producing even narrower current fields than the bipolar mode.

The reduced spectral and temporal resolution resulting from any or all of these factors can have a negative effect on the potential benefit that a CI user can realize, by reducing the amount and/or accuracy of frequency and temporal envelope information that the implant system provides.

A number of studies have explored the effects of spectral smearing/channel overlap on speech perception. Fu and Nogaki (2005) manipulated independently the number of channels (4, 8, or 16) and the filter slopes (carrier filter slopes of -24 or -6 dB/oct) of the noise-band carrier in CI simulations, and found that reductions in spectral resolution through either of those means resulted in higher speech reception thresholds (SRT) for HINT sentences (i.e. poorer performance). The better CI listeners in their study performed similarly to normal-hearing simulation listeners with 8-16 channels with shallow carrier filter slopes simulating spectral smearing (-6 dB/oct. slopes), while the average CI performance was closer to that of simulation listeners with 4-band, spectrally-smeared noise-band speech.

Liu and Fu (2007) investigated the manipulation of five different spectral parameters on vowel recognition in normal-hearing simulation listeners. Their study also demonstrated a significant effect of both number of channels and spectral smearing on speech recognition.

The effect of channel interaction on the perception of speech was investigated by Verschuur (2008) with CI subjects using the Nucleus device, and normal-hearing subjects listening to vocoded speech using a comparable Advanced Combination Encoder (ACE) *n-of-m* processing strategy, in which *n* spectral maxima out of *m* total channels are selected for stimulation. Two channel/maxima conditions were used: 12-of-20 (*maximum* channel condition) and 4-of-7 (*reduced* channel condition). The bandwidth of the noise carriers was altered to simulate different degrees of channel interaction (no interaction, 1 mm spectral spread, and 3.3 mm spectral spread). Verschuur then compared performance of normal-hearing simulation listeners on a consonant identification task as a function of channel interaction condition, number of channels (maximum vs. reduced), and background noise condition (quiet or +10 SNR), to the performance of a group of CI listeners for whom number of channels/maxima and noise conditions were similarly manipulated. A 3.3 mm channel interaction model was found to match the better CI listeners' performance, but the acoustic models were poor predictors of the worse-performing CI listeners. The patterns of feature errors for both CI and simulation groups revealed only small differences between the channel number conditions, and no systematic differences in performance among the different channel interaction conditions, suggesting that channel interaction is not the limiting factor in consonant identification, but that implant processing itself plays the primary role.

Throckmorton and Collins (2002) investigated three psychophysical measures relating to channel interaction -- pitch ranking, electrode discrimination, and forward masking patterns -- to measure their relative effects on speech recognition in 11 normal-hearing subjects listening to CI simulations. They modeled three types of channel interactions, based on psychophysical data measured previously from CI subjects: non-tonotopic pitch order, indiscriminable electrodes, and variable forward-masking patterns. Also varied was the speech-processing algorithm, simulating both SPEAK and CIS strategies. Speech recognition scores (primarily vowel and consonant recognition) obtained with each channel-interaction model, for each processing algorithm, were compared to determine which expression of channel interaction produced the greatest decrement in performance. They found that the effect of each of the manipulations was dependent on the characteristics of the speech-processing algorithm (e.g. number of filters, filter bandwidths). They also examined the relative roles of spectral and temporal interactions and found that, in general, spectral interactions were more detrimental than temporal interactions, and that they were frequency-dependent, with lower-frequency spectral interactions causing greater detriment than those in higher-frequency regions.

Laneau, Moonen, and Wouters (2006) examined the success of noise-band vocoders as acoustic models for pitch perception in CI listeners, especially focusing on the effect of spectral smearing. They found that place pitch sensitivity, measured on simulated electrode discrimination and fundamental frequency discrimination tasks, drops with increased spectral smearing (i.e. decreasing the spectral filter slopes). For their group of normal-hearing simulation listeners, a space constant of 1 mm (representing 1mm of spectral spread) in the vocoders was found to match CI performance on pitch discrimination tasks. However, the results of this study may be limited by the small number of subjects; only four CI listeners and five simulation listeners were included.

Spectral resolution and speech perception in CI users

Another research approach in attempting to clarify the relationship between spectral resolution and speech perception has been to directly obtain psychophysical measures of spectral/temporal resolution in individual CI listeners, and examine their correlation with measures of speech perception in the same listeners. Given that manipulations of the speech signal simulating reductions in spectral resolution do yield reductions in speech recognition performance, as described above, it would follow that such correlations should exist in CI listeners. However, studies have reached mixed conclusions.

For instance, Boex, Kos, and Pelizzone (2003) investigated the effective interactions between electrodes, using a psychophysical forward masking paradigm (a measure of spectral resolution), and found a small but significant negative correlation between the widths of the obtained masking patterns and the recognition of medial consonants in quiet for 12 subjects.

On the other hand, Cohen, Richardson, Saunders, and Cowan (2003) reported no significant correlations between the widths of excitation patterns, obtained using electrically evoked compound action potentials (ECAP), and performance on either word or sentence recognition for seven CI users.

A study by Hughes and Stille (2008) with 18 CI users also failed to find significant correlations between speech perception, measured using CNC words in quiet and sentences in noise, and the width of forward masking patterns, measured either physiologically (using ECAP) or psychophysically (using masked thresholds).

Other measures of spectral resolution such as place-pitch sensitivity, or discriminating between loudness-balanced stimulation on adjacent electrodes, have also produced mixed results. Nelson, Van Tasell, Schroder, Soli, and Levine (1995) found moderate correlations between place-pitch sensitivity and transmitted speech information; Throckmorton and Collins (1999) also reported that electrode discrimination was strongly correlated with vowel and consonant recognition, as well as phoneme recognition for NU-6 monosyllables. However, Zwolan, Collins, and Wakefield (1997) reported no significant correlation between electrode discrimination ability and speech recognition scores, or between electrode discrimination ability and improvement in speech recognition scores when subjects were programmed with an experimental, "corrected" MAP (speech processor program), based on pitch-ranking. Henry, McKay, McDermott, and Clark (2000) reported a significant relationship between phoneme perception of CNC words and average electrode discrimination ability only under conditions of a level rove, and not for loudness-balanced stimuli, as most other studies use. One possible reason for the inconsistent pattern of results across these studies, as pointed out by Henry *et al.* (2000), is that many of the reported studies investigating the relationship between spectral resolution and speech perception may have been limited by the small numbers of subjects and sometimes by the presence of ceiling and/or floor effects. Further, the use of different speech materials by different studies (sentences vs. words vs. phonemes, in quiet or in noise) might be confounding the comparisons.

A number of investigators have used spectrally rippled noise to assess spectral resolution in both acoustic and electric hearing. Spectral ripples are peaks in the spectrum of an acoustic stimulus, spaced at regular intervals and separated by spectral valleys. Discriminating between different spectrally rippled broadband stimuli may be analogous to the perception of spectral cues in speech, such as vowel formant frequencies. A body of work with normal-hearing listeners has been reported by Supin and colleagues (e.g. Supin, Popov, Milekhina, and Tarakanov, 1994; Supin *et al.*, 1997), in which they developed and refined a technique for measuring spectral ripple resolution. Listeners were required to discriminate between a spectrally rippled noise stimulus and the same stimulus with the positions of the spectral peaks and valleys reversed. The maximum ripple rate at which the two stimuli could be discriminated was designated the ripple discrimination threshold.

Henry and Turner (2003) adapted the ripple-reversal test method proposed by Supin and colleagues, and reported a strong correlation between spectral ripple discrimination and both consonant and vowel recognition in quiet when pooling across normal-hearing, impaired-hearing and CI users. However, the correlations were weaker when only the results from CI users were examined. Won, Drennan, and Rubinstein (2007) used a similar paradigm with CI users to compare spectral ripple discrimination with closed-set spondee identification in steady noise and two-talker babble. They found high correlations between ripple discrimination and speech reception threshold (SRT) in noise, consistent with the premise that spectral resolution is particularly important for understanding speech in noise (e.g. Dorman *et al*., 1998).

Berenstein, Mens, Mulder, and Vanpoucke (2008), in a study examining virtual channel electrode configurations in Advanced Bionics CI recipients, demonstrated a small but significant correlation between spectral ripple discrimination and word recognition in quiet and in fluctuating noise.

Temporal resolution.

Another way to describe a speech signal, besides the relative amount of energy present in particular frequency regions, is in terms of the overall amplitude fluctuations of the speech signal as a function of time, or the speech signal's *temporal envelope*. The importance of temporal envelope modulations for speech recognition has been clearly demonstrated by a number of studies that have processed speech to remove the spectral information and retain only temporal information (e.g. Shannon *et al*., 1995; Rosen, 1989; Van Tasell, Soli, Kirby, and Widin, 1987) and have shown that reasonably good speech recognition is possible based only on temporal cues. The importance of temporal envelope information for speech perception is borne out by the success of the Speech Transmission Index (STI) (Steeneken and Houtgast, 1979). Extending the concept of the Articulation Index (AI), which is based on the idea that the intelligibility of speech is dependent upon the audibility of the entire spectrum of the speech signal, the STI takes into account both audibility and distortions in the time domain. With the STI, Steeneken and Houtgast reported good agreement between predicted and actual performance on speech recognition tasks, for a given level of noise.

Modulation detection, in which an amplitude-modulated target signal must be discriminated from an unmodulated standard, is a psychophysical task that is thought to relate to the perception of amplitude variations in speech. Fu (2002) reported a highly significant correlation between sinusoidal amplitude-modulation (SAM) detection thresholds and phoneme recognition in nine CI listeners. He measured SAM detection via direct stimulation, using pulse trains with a 100 Hz modulation rate presented to a single electrode in the middle of the array, at seven different presentation levels across the dynamic range. He reported that the mean SAM detection threshold, calculated over each subject's entire dynamic range, correlated strongly with both consonant and vowel identification in quiet. It is somewhat surprising that modulation detection would correlate so well with vowel recognition, since significant cues differentiating vowels are thought to be spectral in nature (e.g. formant frequency location).

Across-channel vs. within-channel interactions in CI listeners.

Certain types of channel interactions of a temporal rather than spectral nature can occur between channels that are spatially distant from one another. Chatterjee (2003) refers to two types of masking, *tonotopic* and *envelope*. Tonotopic masking is defined as the overlap of neural populations responding to the masker and the probe, at the peripheral level. Envelope masking occurs when the envelope of the noise masks the signal envelope, and is thought to arise from a potentially more central source. Envelope masking does not require tonotopic proximity, and is most likely related to the envelope interactions described in the psychoacoustic literature as ‘‘modulation masking’’ (Bacon and Grantham, 1989) or as ‘‘modulation detection interference’’ (MDI) (Yost and Sheft, 1989). Chatterjee notes that, in certain types of implant processing, steady noise can effectively become modulated noise.

Chatterjee and Oba (2004) investigated MDI and MM in nine CI listeners, as a function of the masker-signal separation and of the type of modulation in the masker. Modulation detection thresholds (50 Hz modulation rate) were measured on a target probe electrode in the presence of maskers presented on the same or each of a series of different electrodes. In this paradigm, a steady-state masker is predicted to produce primarily within-channel energetic masking, due to spectral overlap between signal and masker. If a modulated masker produces more masking than a steady-state masker on a given electrode, the difference between the two is theorized to represent the contribution of "envelope interaction" to the total masking. Chatterjee and Oba measured "envelope masking," or the difference (in dB) between masked thresholds obtained with a dynamic-envelope (modulated) masker and a steady-state masker, matched in overall energy, and they found strong evidence of MDI in their subjects, suggesting across-channel envelope interactions.

B. Effects of background noise on spectrotemporal information in speech

When speech recognition is measured with increasing amounts of steady-state background noise present, performance typically declines monotonically for normal-hearing as well as CI listeners. Noise can sometimes "swamp" critical parts of the speech signal, making them undetectable. If the energy of the noise doesn't exceed that of the speech, it can result is a reduction in the amplitude contrast of the target speech modulations, which can reduce intelligibility (e.g. Steeneken and Houtgast, 1980; Noordhoek and Drullman, 1996).

Noordhoek and Drullman (1997) investigated the effects of temporal modulation reduction (i.e. reduction in the amplitude of temporal envelope fluctuations) by comparing the effect on SRTs in normal-hearing listeners using either a "deterministic" reduction in the amplitude contrast in the acoustic speech signal, or the "stochastic" effect produced by the addition of noise to the speech signal. They found that the effect of background noise on speech recognition does not appear to be the result simply of modulation reduction; two other factors implicated are the introduction of non-relevant modulations and the corruption of the temporal fine structure of the speech signal. Dubbelboer and Houtgast (2008) further explored the notion of non-relevant modulations resulting from speech-noise interactions. They argued that the presence of spurious modulations is a primary factor undermining speech intelligibility in noise.

Cooke's *glimpsing model* of speech perception in noise (Cooke, 2006) asserts that regions of high energy in running speech are sparsely distributed, so that when background noise is present, some spectrotemporal regions of speech will be dominated by the target speech signal and others will be dominated by noise. The inherent redundancy of speech allows for identification based on limited "glimpses" of this sparsely distributed information when it is least affected by background noise – that is, in regions of favorable local SNR. When a certain threshold proportion of glimpses of clean speech information are available, speech recognition will be good.

Either the complete (energetic) masking of individual acoustic cues to speech in the spectrotemporal domain, making them inaudible, or an overall reduction in the amplitude of temporal envelope fluctuations in the speech signal could relate to Cooke's "local SNR." Importantly, not only the detection of some critical proportion of speech energy in restricted spectrotemporal regions, but also the across-channel (and across-time) integration of these bits of information, are requirements for speech perception in noise.

Broader auditory filters, or greater channel overlap, can result in a decreased local SNR within filters when noise from adjacent filters spreads, resulting in fewer effective clean glimpses of the speech signal. In addition, other factors such as upward spread of masking and modulation interference might be more likely to occur. Upward spread of masking, in which the spread of excitation is greater in the basal than the apical direction, has been demonstrated in cochlear implant listeners (Lim, Tong, and Clark, 1989; Clark, 2003); broader masking patterns, in a forward-masking paradigm, were found for more basally-located electrode pairs than in more apical pairs.

An additional factor in understanding the effects of noise on the speech signal is the potential for confusability between target and masker (e.g. Moore and Glasberg, 1988). Target-masker confusions might be considered a type of informational masking, and/or they might be related to the ability to successfully segregate auditory streams.

Background noise may have different effects on spectral and temporal cues to speech. While strong spectral cues such as formant frequencies may be relatively uncorrupted by the presence of modest levels of background noise (because they will still represent regions of "favorable local SNR"), temporal cues might be more significantly degraded, due to the decreased contrast in the temporal envelope. In support of this notion, Munson and Nelson (2005) showed that performance with a static spectral cue (*/i/-/u/* continuum) showed little effect of low-level noise (+10 SNR), while dynamically changing spectral cues (*/ra/-/la/*) and temporal cues (*say-stay*) were more poorly identified at the same SNR by CI listeners.

Noise and CIs.

Even when CI users are able to understand speech in quiet with great success, it appears that any degree of background noise can create difficulty for them. Any sound in the environment can effectively mask speech, even background music (e.g. Gfeller, 2009). In the case of maskers that are spectrally limited, the amount of within-channel energetic masking may be minimal, and interference with speech recognition may not result from a simple reduction in the proportion of clean glimpses of the signal available in regions of favorable local SNR; rather, it may be related to factors such target-masker confusions, modulation interference, an inability to integrate information across channels, or some combination of factors.

The experiments reported in this dissertation are intended to explore further the relationships among different measures relating to spectral resolution, and to investigate how differences in spectral resolution translate into speech recognition in quiet and in noise, both in wide-band and band-limited, speech-shaped background noise.

CHAPTER 2: SPECTRAL RIPPLE DISCRIMINATION

Sections II and III are reprinted with permission from
Anderson, E.S., Nelson, D.A., Kreft, H., Nelson, P.B., and Oxenham, A.J. (2011).
Comparing spatial tuning curves and spectral ripple resolution in cochlear implant users.
J. Acoust. Soc. Am. 130, 364-75.

I. BACKGROUND

Poor spectral resolution is believed to be one of the factors that limit the ability of CI users to understand speech, particularly when background noise is present. Various measures have been used to attempt to quantify spectral resolution in CI users. The spatial tuning curve (STC) is a method in which the current level needed to mask a brief, low-level signal is measured as a function of the spatial separation between the masker and the signal electrode (e.g., Nelson, Donaldson, and Kreft, 2008). This is perhaps the most direct method, but it is time-consuming. Another measure that has become more widely used involves spectral ripple discrimination, in which a spectrally rippled noise stimulus is discriminated by the listener from another spectrally rippled stimulus with the spectral positions of the peaks and valleys reversed (e.g., Henry and Turner, 2003; Henry *et al*., 2005; Won *et al*., 2007). The growing popularity of this "ripple-reversal test," has raised questions as to whether it represents a "pure" measure of spectral resolution, or whether other non-spectral cues might be available when performing this task (e.g. McKay, 2009).

The current study was undertaken to examine the relationship between spectral ripple discrimination thresholds and the STC, the more direct measure of spectral resolution.

The hypothesis was that if spectral ripple discrimination thresholds vary with spectral resolution, then they should show a strong correlation with an index of spatial tuning such as STC bandwidths. In addition, both of these non-speech measures were compared to sentence and vowel recognition, in quiet and in background noise.

II. EXPERIMENT 2.1: BROADBAND RIPPLE DISCRIMINATION

A. Subjects

Fifteen CI users (5 Clarion I, 5 Clarion II, and 5 Nucleus-22), all having had forward-masked STCs measured on at least one electrode from the middle of the array (Nelson *et al*., 2008), participated in all experiments. Table 2.1 shows individual subject characteristics.

Table 2.1. Individual subject characteristics.

Subject Code	M/F	Age (yrs)	CI use (yrs)	Etiology	Duration of deafness (yrs)	Device	Speech Processing Strategy	Stim Mode (STC)	Stim Mode (MAP)
C03	F	58.8	9.7	Familial Progressive SNHL	27	Clarion I	CIS	MP	MP
C05	M	52.5	10.2	Unknown	<1	Clarion I	CIS	MP	MP
C16	F	54.2	6.7	Progressive SNHL	13	Clarion I	MPS	MP	MP
C18	M	74.0	7.2	Otosclerosis	33	Clarion I	MPS	MP	MP
C23	F	48.1	6.4	Progressive SNHL; Mondini's	27	Clarion I	CIS	MP	MP
D02	F	58.2	6.4	Unknown	1	Clarion II	HiRes -P	MP	MP
D05	F	78.2	6.6	Unknown	3	Clarion II	HiRes -S	MP	MP
D08	F	55.9	5.0	Otosclerosis	13	Clarion II	HiRes-S	MP	MP
D10	F	53.8	5.2	Unknown	8	Clarion II	HiRes-S	MP	MP
D19	F	48.2	3.5	Unknown	7	Clarion II	HiRes-S	MP	MP
N13	M	69.9	17.5	Hereditary; Progressive SNHL	4	Nucleus 22	SPEAK	BP+2	BP+3
N14	M	63.5	13.9	Progressive SNHL	1	Nucleus 22	SPEAK	BP	BP+1
N28	M	68.8	11.8	Meningitis	<1	Nucleus 22	SPEAK	BP+1	BP+1
N32	M	40.1	10.3	Maternal Rubella	<1	Nucleus 22	SPEAK	BP	BP+1
N34	F	62.0	8.4	Mumps; Progressive SNHL	9	Nucleus 22	SPEAK	BP	BP+3

B. Stimuli

Spectrally rippled noise was generated using MATLAB software (The Mathworks, Natick, MA). Gaussian broadband (350-5600 Hz) noise was spectrally modulated, with sinusoidal variations in level (dB) on a log-frequency axis (as in Litvak, Spahr, Saoji, and Fridman, 2007), using the equation:

$$X(f) = 10^{\frac{D}{2}\sin\left(2\pi\left(\log_2\left(\frac{f}{L}\right)\right)fs+\theta\right)/20} \qquad \text{[Equation 1]}$$

where $X(f)$ is the amplitude at frequency f (in Hz), D is the spectral depth or peak-to-valley ratio (in dB), L is the low cutoff frequency of the noise pass band (350 Hz in this case), f_s is the spectral modulation frequency (in ripples per octave), and θ is the starting phase of the ripple function. Logarithmic frequency and intensity units were used as these are generally considered to be more perceptually relevant than linear units. Sinusoidal modulation was used, as this lends itself more readily to linear systems analysis and has been used in a number of studies of spectral modulation perception in normal (acoustic) hearing (e.g., Saoji and Eddins, 2007; Eddins and Bero, 2007). The peak-to-valley ratio of the stimuli was held constant at 30 dB. The stimulus duration was 400 ms, including 20-ms raised-cosine onset and offset ramps.

The stimuli were presented via a single loudspeaker (Infinity RS1000) positioned at approximately head height and about 1 m from the subject in a double-walled, sound attenuating chamber. The average sound level of the noise was set to 60 dBA when measured at the location corresponding to the subject's head. In order to reduce any possible cues related to loudness, the noise level was roved across intervals within each trial by ±3 dB. The starting phase of the spectral modulation was selected at random with uniform distribution for each trial to reduce the potential for any consistent local intensity cues that fixed-phase stimuli might create.

C. Procedure

Subjects wore their everyday speech processors at typical use settings for all experiments. A three-interval, three-alternative forced-choice (3I-3AFC) procedure was used. All three intervals in each trial contained rippled noise. In two of the intervals the spectral ripple had the same starting phase, and in the other interval the phase was reversed (180° phase shift). The interval containing the phase reversal was selected at random on each trial with equal *a priori* probability, and the listener's task was to identify the interval that sounded different. Each test run started at a ripple rate of 0.25 ripples per octave (rpo), corresponding to a single ripple across the 4-octave passband. The ripple rate was varied adaptively using a 1-up, 2-down rule, with rpo initially increasing or decreasing by a factor of 1.41. After the first two reversals the step size changed to a factor of 1.19, and after two more reversals, to 1.09. The run was terminated after ten reversals, and the geometric mean ripple rate at the last six reversal points was used to determine the threshold for ripple discrimination. If the adaptive procedure called for a ripple rate lower than 0.25 ripples per octave, the program set the ripple rate to 0.25 ripples per octave and the adaptive procedure continued. However, this "floor" was never reached during the measurement phases of the adaptive runs and no estimates of threshold included turnpoints of 0.25 ripples/octave. Each subject completed six runs, with the exception of one subject who completed only four runs (due to time and scheduling constraints). In order to minimize potential learning and inattention effects, the first threshold estimate was excluded for each subject, as were individual measurements for runs that were more than 3 standard deviations removed from the mean of the remaining measurements. In general, thresholds from the last five runs were used to compute an arithmetic mean threshold for each subject. (An alternate approach, computing the geometric mean, produced the same pattern of results.)

D. Comparison with spatial tuning curve bandwidths

As described in Nelson *et al*. (2008), a forward-masking paradigm was used to obtain spatial tuning curves. The procedure measured the masker level needed to produce a constant amount of forward masking on a specific probe electrode, for several individual masker electrodes surrounding the probe electrode. The signals were biphasic current pulses delivered in a direct-stimulation mode via a specialized CI interface (Nucleus devices) or dedicated research processor (Advanced Bionics devices). The stimulation mode was bipolar (BP) for Nucleus users and monopolar (MP) for Advanced Bionics users. For several subjects, the stimulation mode was adjusted (e.g., from BP to BP+1) to allow for sufficient levels of masking for the STC procedure, and thus was slightly different than that used in a subject's speech processor program (see Table I for details). The STCs were measured on a single electrode in the middle of the array for each of the 15 subjects. Bandwidth was defined as the width in mm of the STC, at a masker level that was 1 dB above its level at the STC tip. Figure 1 shows an example of a single spatial tuning curve (black squares) for one electrode in the middle of the array for one subject. Further details can be found in Nelson *et al*. (2008) and in Table AI in the Appendix [of Anderson *et al*., 2011]. The STC data were collected, on average, 3.5 years prior to spectral ripple data. The stimulation mode of the subjects' implants remained the same over that period.

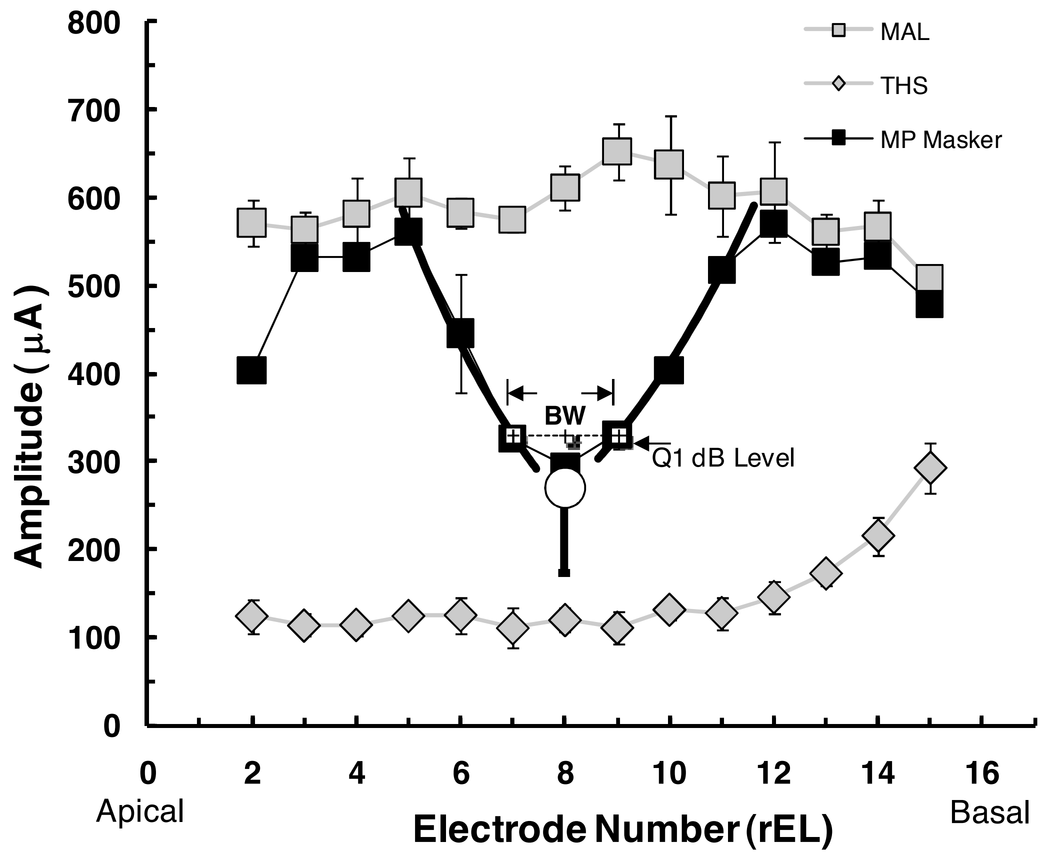


Figure 2.1. Spatial tuning curve for subject D08, measured in a previous study (Nelson *et al.*, 2008). Depicted by the black squares in this figure is the current level needed to just mask a low-level probe presented to electrode 8, as a function of masker electrode number. Stimulus amplitude (µAmps) is shown on the ordinate, with research electrode number (rEL) displayed on the abscissa. (The rEL numbering system normalizes the different numbering systems used by different implant devices; number 1 is assigned to the most apical electrode, with consecutive numbering proceeding to the basal end of the array.)

Error bars indicate standard deviations. Gray squares represent current levels at maximum acceptable loudness (MAL) for each electrode; gray diamonds indicate current levels at threshold (THS) for each electrode. The open circle at the tip of the curve indicates the probe electrode, with probe level shown by the symbol's vertical position and sensation level indicated by the height of the vertical line beneath it. Tuning curve bandwidth (BW) is defined as the width of the STC at 1 dB above the tip of the STC (Q1 dB Level).

E. Results and discussion

Spectral ripple discrimination thresholds – the highest ripple rate at which phase-reversed spectrally rippled stimuli could be discriminated – ranged from 0.41 to 4.27 rpo, with a mean of 1.68 rpo. This wide range across subjects is similar to that observed for the STC measures of bandwidth in the same subjects (Nelson *et al*., 2008). In addition, spectral ripple discrimination thresholds in these subjects are quite similar to the range of data reported by Won *et al*. (2007), despite the differences in spectral envelopes (full-wave rectified) used in that study. The left panel of Fig. 2 shows the broadband ripple discrimination thresholds as a function of the reciprocal of the BW_{STC} from Nelson *et al*. (2008) for each subject. The solid line is the least-squares fits to the data. Error bars represent one standard deviation. When all 15 subjects were included, simple regression analysis failed to reveal a significant correlation between BW_{STC} and spectral ripple discrimination thresholds (r^2=0.13, p=0.18). However, the plot indicates two obvious outliers, subjects C05 and D10, whose results did not fall within the 95% confidence intervals of the linear regression fits. When these two subjects were removed from the regression (Fig. 2, left panel), a strong and significant correlation was observed (r^2=0.79, $p<0.001$). For these remaining 13 subjects, broader spatial tuning (smaller values on the x axis) corresponded with poorer spectral ripple discrimination (fewer rpo at threshold, or smaller values on y axis).

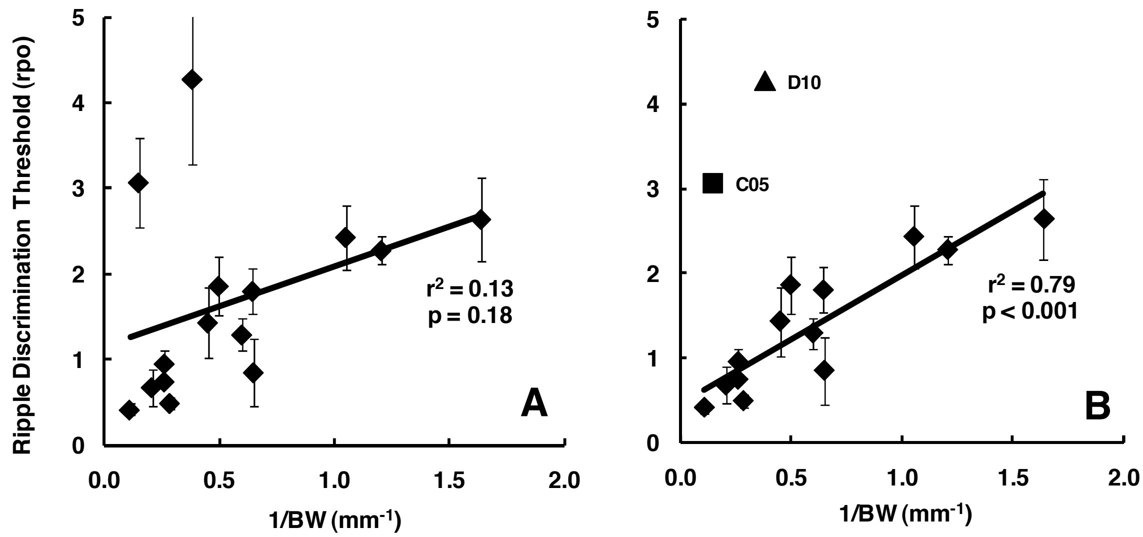


Figure 2.2 Ripple discrimination thresholds as a function of transformed STC bandwidth (from Nelson *et al.*, 2008). Panel A shows broadband ripple discrimination thresholds, in ripples per octave (rpo) with a linear regression fit to all 15 data points; Panel B shows the regression line for the least squares fit to data from 13 subjects, excluding subjects C05 and D10.

Despite the encouraging trend observed in most subjects, the overall data set does not provide compelling evidence for concluding that BW_{STC} and ripple discrimination threshold are measures of the same underlying mechanism. One possible explanation for any discrepancies between spatial tuning and spectral ripple discrimination is that the STC represents a local, focused measure of spectral resolution at a place in the cochlea corresponding to the location near the single probe electrode in the middle of the implant array, whereas the standard broadband rippled noise paradigm yields a more global measure because the stimuli encompass a frequency range that spans almost all the electrodes in the array. It may be that discrimination thresholds in the spectral ripple task were mediated by the spectral resolution associated with electrodes remote from the electrode tested in the STC task, and that the STCs for those electrodes corresponding to these areas of off-frequency listening would be narrower. This explanation is tested in Experiment 2.

Another explanation may be that good performance in the spectral ripple discrimination task does not depend solely on spectral resolution abilities. The two outliers may have been using other cues to perform the task. As mentioned in the introduction, it has been suggested that spectral edge effects may play a role. The stimuli used in Experiment 1 had sharp spectral edges, limited only by the temporal onset and offset ramps applied to the stimuli. Because of this, the level of the spectral components at the edge could vary dramatically, depending on the starting phase of the stimuli. Also, having sharp spectral edges leaves open the possibility that the change in ripple phase is detected via a shift in the centroid of the spectral envelope. To examine this possibility, we re-tested a subset of our original subjects, including the two outliers and three additional subjects representing a wide range of performance, with shallow spectral slopes on either end of the spectrum to reduce potential spectral edge effects. The re-test occurred between 15 and 20 months after the original test.

F. Experiment 1b: Effects of spectral edges on ripple discrimination

Spectrally rippled noise was generated in the same way as for Experiment 1, with either steep spectral edges ("non-windowed" stimuli) or with Hanning (raised-cosine) ramps applied to the spectral edges ("windowed" stimuli). The passband of the windowed stimuli was two octaves, with one-octave spectral ramps on either side, giving a total bandwidth of four octaves, as for the non-windowed stimuli. Figure 3 displays sample plots of spectrally windowed and non-windowed rippled noise in the left and right panels, respectively. All other stimulus parameters and the test procedure remained the same as in Experiment 1. Five of the original group of CI subjects participated. Each listener completed six runs of the adaptive procedure for windowed stimuli and six runs for non-windowed stimuli run in blocks.

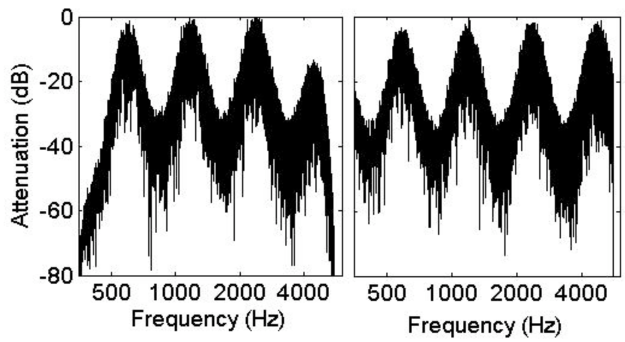


Figure 2.3. Plots of windowed (left) and non-windowed (right) stimulus spectra. Stimuli are broadband (350-5600 Hz) noise with sinusoidal spectral ripples; spectral modulation frequency is 1 ripple per octave. The windowed stimulus includes Hanning (raised-cosine) ramps applied to the spectral edges. The non-windowed stimulus has steep spectral edges**.**

The results are displayed in Fig. 4. Ripple discrimination thresholds for the windowed stimuli are plotted as a function of thresholds for non-windowed stimuli. The correspondence between thresholds for the two stimulus types is good. Using paired comparisons, no significant difference in threshold was found between the spectrally windowed and non-windowed stimuli [$t(4)=0.59$; $p=0.59$], suggesting similar performance for broadband ripple discrimination regardless of whether the noise passband has steep or shallow slopes. This in turn suggests that spectral edge effects are unlikely to have dominated performance in the main experiment, at least for this subset of listeners.

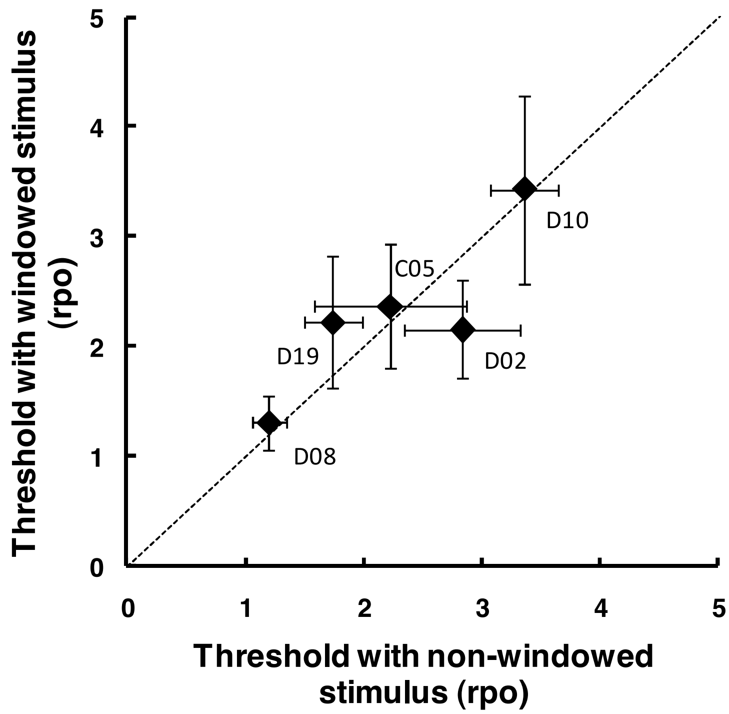


Figure 2.4. Ripple discrimination thresholds for shallow-sloped (windowed) stimuli, with half-octave Hanning ramps on each side, as a function of thresholds for steep-sloped (non-windowed) stimuli. The diagonal line represents perfect correspondence between the two types of stimuli (slope = 1).

One unexpected result of this comparison was that thresholds for broadband non-windowed stimuli for the two CI subjects who were outliers in the original data set were somewhat lower (poorer) on retest, compared to their original data. However, even when taken individually (with no post-hoc correction factor), the difference was significant ($p=0.05$) for one subject, C05, but not for the other ($p=0.06$). The three other CI subjects from Experiment 1b who were retested on broadband "non-windowed" ripple discrimination, plus one additional subject drawn from the larger group, showed no significant test-retest differences. Therefore, for this group of six subjects there was no significant test-retest difference in the ripple discrimination threshold [$t(5)=1.44$;

p=0.21]. Won *et al*. (2007) reported ripple discrimination test-retest data for 20 of their subjects (two sets of six runs completed on different days) and found no significant group differences; they concluded that the ripple reversal test appears to be a stable, repeatable measure. Our test-retest measures support their conclusions and extend them by showing that in general the thresholds can remain stable over the period of a year or more. We have no similar test-retest data for the STC measurements, and it may be that changes over the time course of these experiments (~4 years) added to the variability in comparing the measures of STC bandwidth and ripple discrimination.

III. EXPERIMENT 2.2: COMPARISON OF OCTAVE-BAND SPECTRAL RIPPLE DISCRIMINATION AND SPATIAL TUNING CURVES OBTAINED FROM THE SAME REGION OF THE COCHLEA

A. Rationale

The aim of this experiment was to explore the use of the ripple discrimination paradigm using octave-band rippled noise stimuli, matched in frequency for each subject to the probe electrodes used in obtaining their original STC measures, to test whether the correspondence between the two measures would become stronger if both measures were focused on the same cochlear location.

B. Methods

The subjects and procedure were the same as those used in Experiment 1. Spectrally rippled noise was again generated with log-spaced ripples along the frequency axis, using sinusoidal variations in level (dB) in the same way as in Experiment 1. The only difference was that the noise was band-limited to two octaves, with raised-cosine highpass and lowpass slopes applied to the amplitude spectrum of the lower and upper half-octaves, respectively, so that the unattenuated passband of the noise was one octave wide. The center frequency of the octave passband was selected for each subject individually to correspond to the center frequency of the filter from the speech processor that was assigned to the electrode for which the STC had been measured. In this way the STC and the ripple-discrimination task were as closely matched as possible in terms of assessing spatial resolution in the same region of the cochlea. As in Experiment 1, the stimuli were presented in the sound field at 60 dBA, roved by ±3 dB on each interval.

C. Results and discussion

Octave-band ripple discrimination thresholds ranged from 0.4 to 3.17 rpo, with an average of 1.59 rpo. Figure 5 shows octave-band ripple discrimination thresholds plotted against the transformed BW_{STC} for the same subjects. Comparing this plot to the ones in Figure 2, it can be seen that the data from the two subjects previously identified as outliers are now closer to the trend line fitted to data from the other 13 subjects. In fact, this experiment revealed a significant relationship between BW_{STC} and spectral ripple threshold when including the data from all subjects ($r^2 = 0.45$, $p=0.007$; solid line), although the regression excluding subjects C05 and D10 continued to be stronger ($r^2=0.73$, $p<0.001$; dotted line). Despite the differences evident in two subjects, octave-band and broadband ripple discrimination thresholds were strongly correlated across subjects overall (r^2=0.65, p<0.001).

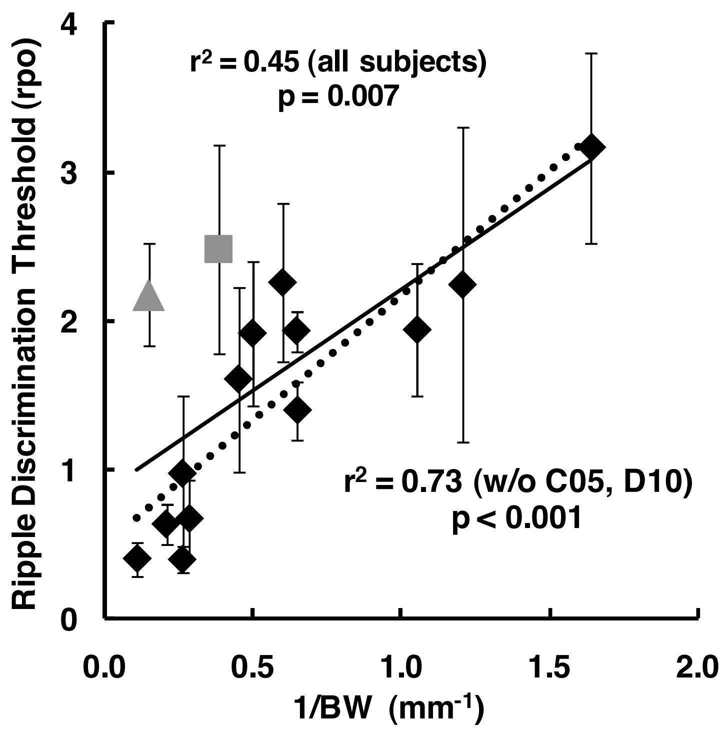


Figure 2.5. Octave-band ripple discrimination (from Experiment 2) as a function of transformed STC bandwidth. Data from all subjects are included in the solid regression line; subjects C05 and D10 are identified with different symbol shapes. Data from 13 subjects, excluding C05 and D10, are included in the dotted regression line.

One possible interpretation of the better correspondence between STCs and narrow-band ripple measures when compared with Experiment 1 is that ripple discrimination using a broadband rippled noise stimulus might not reflect local differences in spectral resolution along the electrode array for a given listener. With a broadband stimulus, a listener might be responding to acoustic information within a limited region of better spectral/spatial resolution. If ripple discrimination varies with the region of the electrode array being stimulated by the noise carrier, this might be demonstrated by using different narrow passbands of spectrally rippled noise.

IV. EXPERIMENT 2.3: FIXED OCTAVE-BAND RIPPLE DISCRIMINATION

A. Rationale

The aim of this experiment was to measure spectral ripple discrimination thresholds using four contiguous fixed octave-band noise stimuli to determine whether spectral ripple discrimination thresholds could vary across different regions of the electrode array. For each subject, broadband ripple discrimination thresholds were compared to fixed octave-band thresholds, to determine more accurately the relationship between broadband and narrowband spectral ripple discrimination measures.

B. Methods

The subjects and procedures were the same as in Experiments 1 and 2. The stimulus was one of four bands with one-octave passbands (350-700 Hz, 700-1400 Hz, 1400-2800 Hz, or 2800-5600 Hz), each with half-octave raised-cosine ramps applied to either side of the amplitude spectrum. The order in which the conditions were tested was randomized between subjects and between each repetition. For each subject, all four conditions were tested before any was repeated. Thresholds reflect the average of at least five repetitions.

C. Results

Figure 6 displays ripple discrimination thresholds for the individual listeners. The results are divided across three panels, according to implant device used by the subject. Ripple discrimination thresholds varied across listeners as well as within listeners as a function of frequency band. A repeated-measures ANOVA showed no main effect for frequency band [Greenhouse-Geisser corrected F(1.8, 24.8)=1.68, p=0.21], suggesting that there was no orderly pattern in ripple discrimination thresholds as a function of frequency band, when pooled across all subjects. Nevertheless, there were substantial variations in threshold across frequency in individual subjects. One-way ANOVAs performed on each individual subject's data showed a main effect of frequency band for 11 of the 15 subjects. If such effects were produced only by random variability in threshold measurements, a significant effect would be expected to occur by chance in only about one of the 15 subjects. Thus it seems that these variations were "real" if not consistent between subjects. Interestingly, both subjects who were outliers in Experiment 1 (C05 and D10) showed substantial variation in threshold across frequency band, as would be expected if narrowband and broadband measures of frequency resolution did not correspond well.

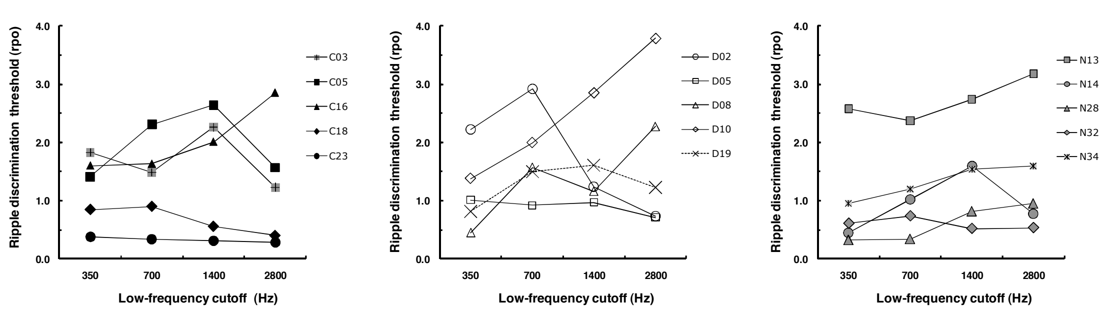


Figure 2.6. Ripple discrimination thresholds as a function of low-frequency cutoff of octave-band rippled noise stimuli, for 15 individual subjects, separated by device type. Panel A illustrates performance of Clarion I subjects, Panel B shows Clarion II subjects, and Panel C shows Nucleus 22 subjects.

Thresholds in the octave bands were compared with thresholds in the broadband conditions tested in Experiment 1. To facilitate the comparisons, summary measures of the octave-conditions were derived in three ways for each subject: 1) unweighted average of thresholds across all four octave-band conditions; 2) best threshold (i.e., greatest rpo value), and 3) worst threshold (lowest rpo value). Table 2.2 contains summary measures for each individual subject. Performance in the broadband condition (Experiment 1) was compared with performance in each of the three summary measures using paired-sample t-tests. The broadband thresholds were not significantly different than average fixed-octave band thresholds ($t=-2.00$, $p=0.07$) or the best octave-band thresholds ($t = -1.9$, $p=0.09$), but were significantly different than worst octave-band thresholds ($t=-3.71$, $p=0.003$). This result suggests that the CI users may be utilizing information from across the array when performing broadband ripple discrimination, but our method lacks the statistical power to determine whether the integration of information is “optimal” in any sense.

Table 2.2. Individual subject data for fixed octave-band spectral ripple discrimination. The first column shows the average of four octave-band discrimination thresholds (in ripples per octave) for each listener. “Best OB threshold” displays the subject’s lowest threshold, and “Worst OB threshold” displays the highest. The final column indicates ripple discrimination threshold in the broadband condition.

Subject Code	Mean OB threshold	Best OB threshold	Worst OB threshold	Broadband threshold
C03	1.70	2.26	1.23	1.29
C05	1.98	2.64	1.41	3.07
C16	2.02	2.85	1.60	2.28
C18	0.68	0.90	0.40	0.49
C23	0.33	0.38	0.28	0.41
D02	1.78	2.92	0.73	2.43
D05	0.90	1.01	0.71	0.68
D08	1.36	2.27	0.45	1.43
D10	2.57	3.79	1.38	4.27
D19	1.29	1.60	0.82	1.86
N13	2.72	3.18	2.37	2.64
N14	0.95	1.46	0.57	0.85
N28	0.61	0.95	0.33	0.74
N32	0.60	0.74	0.52	0.95
N34	1.30	1.54	0.96	1.80

D. Discussion

Ripple discrimination of octave-band noise stimuli in different frequency regions revealed significant variability within and between subjects. Ripple discrimination appears to vary across the electrode array, presumably due to many factors such as neural survival and electrode placement. Supin *et al.* (1997) measured ripple discrimination in normal-hearing subjects as a function of the center frequency (CF) of octave-band rippled noise stimuli. They found that for frequencies below 1000 Hz, ripple discrimination threshold increased with increasing CF, but above 1000 Hz, thresholds remained roughly constant with increasing CF, in line with estimates of frequency selectivity using simultaneously presented notched noise (Glasberg and Moore, 1990). Although there was substantial variation across the electrode array for many of our individual CI subjects, on average there was no systematic trend in performance with increasing CF of the octave-band carrier. The lack of a systematic effect in the pooled data is in line with expectations based on the fact that the CI processor filters have roughly constant bandwidth on a log-frequency scale, implying that the filters' spectral resolution should be independent of spectral region. Other factors, such as individual electrode placement, current spread, and neural survival patterns, may vary substantially with cochlear location in individual subjects, but such variations are unlikely to result in a systematic trend within a subject group (e.g., Hinojosa and Marion, 1983; Kawano, Seldon, Clark, Ramsden, and Raine, 1998).

V. COMPARISONS WITH MEASURES OF SPEECH PERCEPTION

A. Rationale

One benefit of the spectral ripple discrimination test is that it has been shown in some previous studies to correlate with measures of speech perception in noise (Won *et al.*, 2007). The present experiment sought to replicate and extend these earlier findings by correlating both spectral ripple discrimination thresholds and STC bandwidths with measures of speech perception in quiet and in noise.

B. Methods

The subjects tested for STCs and spectral ripple discrimination were assessed on multiple measures of speech recognition. The tests included sentence and vowel materials in quiet and in noise. Sentence recognition testing was performed using IEEE sentences (IEEE, 1969), spoken by one male and one female talker. Each sentence contained five keywords. Subjects orally repeated each sentence after presentation, and one of the experimenters, sitting in the booth with the subject, recorded the number of correct keywords. The vowel test involved Hillenbrand vowels (Hillenbrand, Getty, Clark, and Wheeler, 1995) spoken by six male talkers; this closed-set test is composed of 11 vowels in an h/V/d context. Subjects identified each vowel token by selecting the appropriate word from a list on a computer screen. Speech-shaped background noise was generated to match the long-term average spectrum of the speech materials. The speech and noise were mixed to produce the desired speech-to-noise ratio (SNR), which was verified acoustically with a sound level meter.

Speech recognition testing was performed in a sound-treated booth. All speech materials were presented at 65 dBA. Subjects used their own speech processors for speech recognition testing, with the processors set at each individual's typical use settings; programs optimized for noise reduction were not used. The speech stimuli were presented through a single speaker placed one meter in front of the subject. For sentences, one 10-sentence list spoken by a male talker and one list spoken by a female talker were presented in each condition (in quiet and at SNRs of +20, +15, +10, and +5 dB), giving a total of 100 keywords for each condition.

Vowel stimuli were presented in the sound field in the same manner, with subjects using their own speech processors at typical use settings. The 11 vowels were presented in 66-item blocks (six presentations of each vowel per block), after a 33-item practice run preceding each SNR condition. Feedback was provided during the practice runs, but not during the test trials. The average of three blocks was used to calculate percent correct scores for each condition, with the requirement that all three scores fall within a 10% performance range. Thus, each score was based on 18 presentations of each vowel token.

Speech recognition performance in quiet and in noise for each subject was compared to BW_{STC} and to broadband spectral ripple discrimination thresholds, to investigate whether these measures of nonspeech spectral resolution were predictive of speech recognition ability. Octave-band ripple discrimination was not included as an independent variable in the analyses, since speech is a broadband signal.

C. Results

Speech performance in quiet is reported as rationalized arcsine-transformed (rau) scores (Studebaker, 1985) for recognition of key words in sentence materials and for vowel identification. Since no significant differences were found for sentence recognition performance for male vs. female talkers across subjects (repeated measures ANOVA, $F(1,69)=0.68, p=0.41$), the average score for male/female talkers for each condition was used. For this group of subjects, percent correct for key words on sentence materials ranged from 5% to 99%. Percent correct on vowel recognition in quiet ranged from 54% to 96%, with a median of 87%.

Performance in noise for both sentence and vowel materials was quantified as $SNR_{50\%}$, determined as follows: performance-intensity functions (percent correct as a function of SNR) were plotted for each subject, and logistic functions were fitted to the curves, using a least-squares criterion. The fitting method was the same as that used by Qin and Oxenham (2003), with the exception that the maximum level of performance was a free parameter, and the minimum level of performance was set to ~9.1% in the vowel task to reflect chance performance in the closed-set task. The 50%-correct point was calculated from each of the fitted curves to produce the so-called speech reception threshold.

1. Broadband ripple discrimination and speech recognition

Since multiple comparisons were being made (four conditions – vowels and sentences, in quiet and in noise), a Bonferroni-corrected value of α=0.0125 was used to determine statistical significance. Broadband ripple discrimination thresholds showed a moderate, statistically significant correlation with rau scores for word recognition within sentences in quiet (r^2=0.47, p=0.005), using linear regression. Visual inspection of the data suggested a compressive relationship between percent word recognition and ripple discrimination, even after arcsine transformation. Therefore the relationship was modeled using a logarithmic function, as shown in Fig. 7A, which resulted in a somewhat stronger relationship (r^2=0.59, p=0.001). In contrast, the correlations between ripple discrimination thresholds and arcsine-transformed recognition scores for vowels in quiet failed to reach significance using either linear regression (r^2=0.19, p=0.10) or a logarithmic function (r^2=0.32, p=0.03), as shown in Fig. 7B. All 15 subjects are included in the regressions.

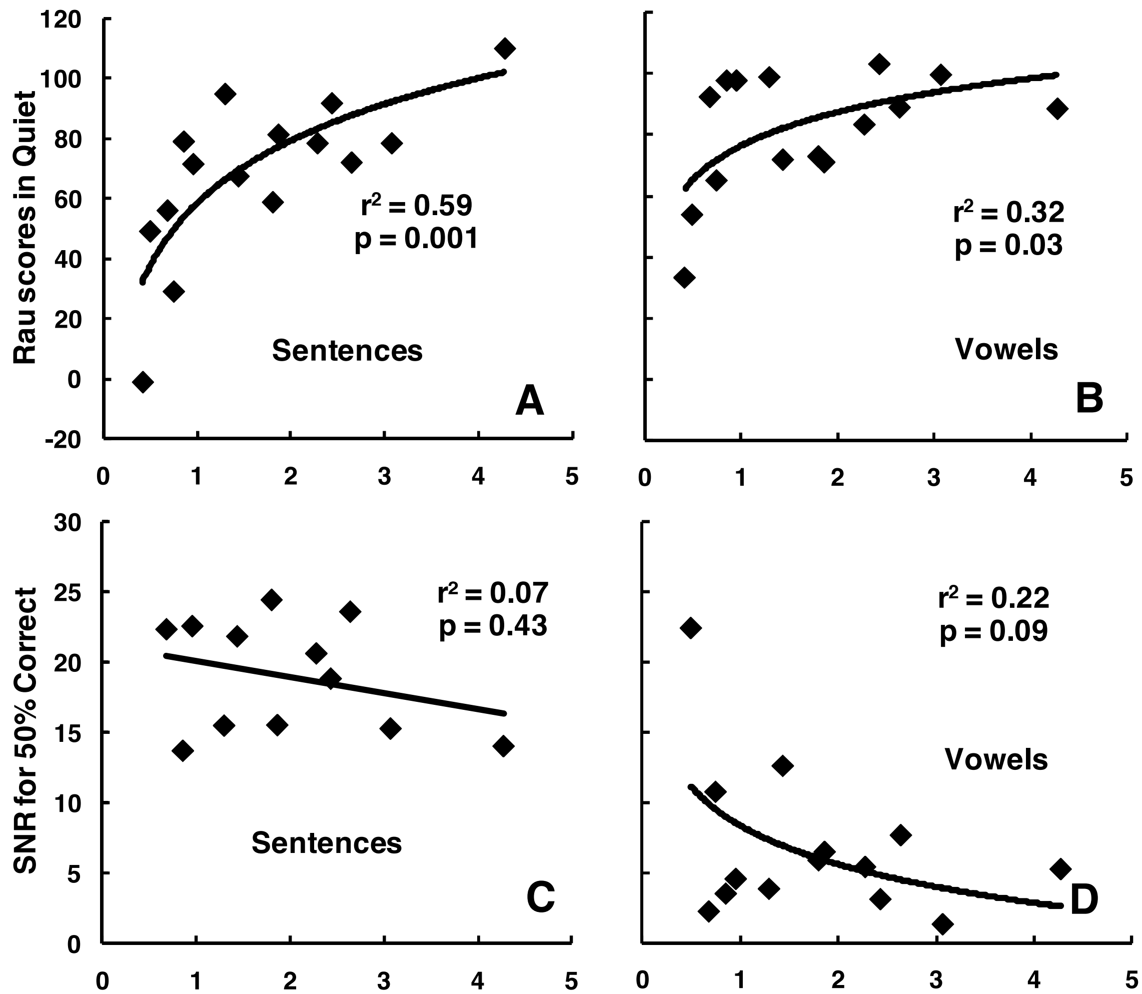


Figure 2.7. Panel A: Sentence recognition in quiet (rau scores) as a function of broadband ripple discrimination threshold. Panel B: Vowel recognition in quiet (rau scores) as a function of broadband ripple discrimination threshold. Panel C: SNR for 50% correct sentence recognition, interpolated/extrapolated from performance-intensity (P-I) functions, as a function of broadband ripple discrimination threshold. Data from 12 subjects are included. Panel D: SNR for 50% correct vowel recognition, which was interpolated/extrapolated from P-I functions for 14 subjects, as a function of broadband ripple discrimination threshold. Regression lines are all logarithmic fits, with the exception of Panel C, which shows a linear fit.

The relationships between ripple discrimination and sentence and vowel recognition in noise are shown in Figs. 7C and 7D, respectively. Plotted are the SNRs corresponding to 50% correct performance (calculated from the best-fitting performance-intensity functions, as described above) as a function of ripple discrimination threshold. Three subjects (C18, C23, N28) were excluded from sentence analysis and one (C23) from vowel analysis because their asymptotic performance fell below 50%. The SNR required for 50% word recognition in sentences for the remaining 12 subjects did not show a significant relationship with spectral ripple discrimination (r^2=0.07, p=0.43); the corresponding SNR for vowel recognition in noise for 14 subjects also failed to exhibit a significant correlation with ripple discrimination (r^2=0.22, p=0.09).

2. STC bandwidth and speech recognition

The upper two panels of Fig. 8 show sentence and vowel recognition in quiet (rau scores) as a function of $1/BW_{STC}$. A Bonferroni-corrected value of α=0.0125 was again used to determine statistical significance. The correlations between transformed BWSTC and both the log-transformed sentence and vowel (rau) scores in quiet failed to reach statistical significance (r^2=0.30, p=0.04 and r^2=0.17, p=0.13, respectively; see Fig. 8, panels A and B).

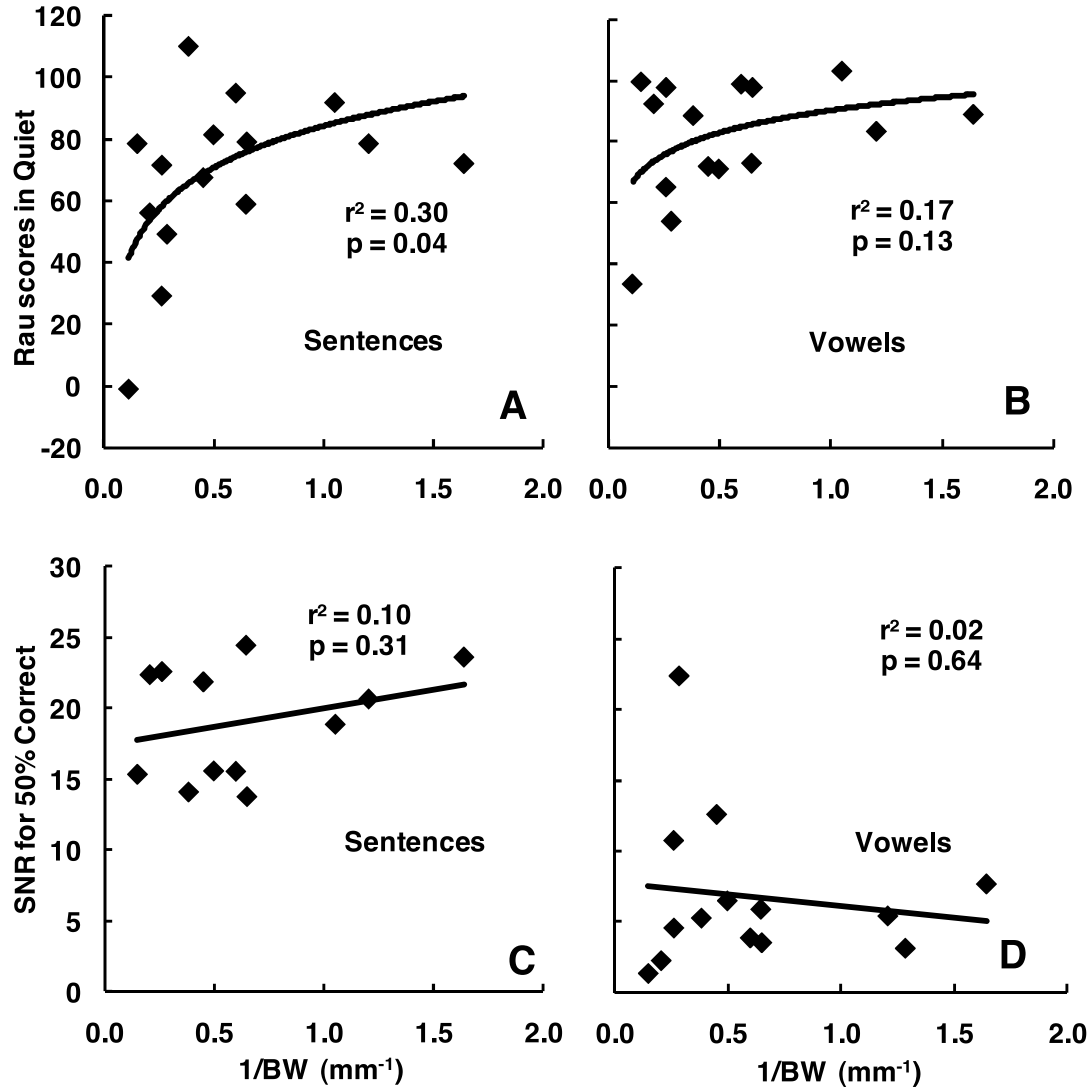


Figure 2.8. Panel A: Sentence recognition in quiet (rau scores) as a function of 1/ BW. Panel B: Vowel recognition in quiet (rau scores) as a function of 1/ BW. Regression lines are log fits. Panel C: SNR for 50% correct sentence recognition, interpolated/extrapolated from P-I functions, as a function of 1/ BW. The regression line is a linear fit. Data from 12 subjects are included. Panel D: SNR for 50% correct vowel recognition, which was interpolated/extrapolated from P-I functions for 14 subjects, as a function of 1/BW. The regression line is a linear fit.

The relationships between BW_{STC} and sentence and vowel recognition in noise are shown in the lower two panels of Figs. 8, which plot the SNRs corresponding to 50% correct performance (derived from the fitted performance-intensity functions) as a function of $1/BW_{STC}$. Again, three subjects (C18, C23, N28) were excluded from sentence analysis and one (C23) from vowel analysis because asymptotic performance fell below 50%. The BW_{STC} did not show a significant relationship with sentence recognition in noise for the remaining 12 subjects (r^2=0.10, p=0.31), nor did vowel recognition in noise for 14 subjects (r^2=0.02, p=0.64).

D. Discussion

Neither measure of spectral resolution – the spectrally local BW_{STC} measured using direct stimulation or the spectrally global spectral ripple discrimination threshold measured using the subjects' own speech processor – produced robust correlations with measures of speech perception either in quiet or in noise. The one exception was the measure of sentence recognition in quiet, which correlated significantly with spectral ripple discrimination thresholds but missed significance when correlated with BW_{STC}. The lack of strong correlations is particularly surprising for the vowel stimuli, as it is generally thought that vowel recognition relies to a large degree on the discrimination of the formant frequencies. On the other hand, earlier studies in normal-hearing listeners using noise-excited envelope-vocoded speech had shown that reasonable vowel recognition was possible with as few as four broadly tuned channels of spectral information (Shannon *et al.*, 1995).

In general, our correlations were not as strong as those reported by Won *et al.* (2007), who found robust correlations between spectral ripple discrimination thresholds and closed-set word recognition in various types of background, including noise, two-talker babble, and quiet. One potential reason is statistical power: their sample included 29 CI users, whereas our maximum sample size was 15; and Won *et al*. (2007) included no correction for multiple comparisons, whereas we adopted a more conservative approach. A more detailed comparison of the r^2 values shows that the values found by Won *et al*. (2007) ranged from 0.25 to about 0.38 (absolute *r* values between 0.5 and 0.62). Although our values were lower than this range for the speech-in-noise conditions, the ranges from the two studies do overlap, suggesting that the differences may have been due at least in part to sample size and statistical approach, rather than other differences, such as the speech material used. Thus, a conservative conclusion may be that there exists a relationship between measures of spectral resolution and speech perception, but that the variance accounted for is relatively small.

E. Conclusions

Spectral ripple discrimination thresholds for an octave-wide band of noise (but not broad-band stimuli) correlated significantly with STC bandwidth, when the two measures were matched closely in frequency by being centered on the same electrode. Differences between broadband and octave-band measures of spectral ripple discrimination support the idea that differences in resolution at different regions of the cochlea may explain, in part, the discrepancies between spectrally broad and spectrally focused measures. In general, the results suggest that spectral ripple discrimination can be considered a valid measure of spectral resolution in CI users.

Spectral ripple discrimination thresholds correlated with sentence recognition in quiet, but not in noise. These findings are consistent with other reports in the literature (e.g. Hughes and Stille, 2009; Cohen *et al*., 2003; Zwolan *et al*., 1997) suggesting that the role of spectral resolution in predicting speech perception in CI users is far from clear.

CHAPTER 3: SPECTRAL RIPPLE DETECTION

I. INTRODUCTION

The ability to discriminate spectral shapes has relevance to the auditory processing of complex acoustic signals such as speech, since some cues to the identity of speech sounds are contained in the spectral envelope. The preceding chapter described a psychophysical paradigm (the spectral ripple reversal test) that requires discrimination of one spectrally rippled noise from another one identical to the first, except the positions of spectral peaks and valleys are reversed. Another psychophysical task that involves spectral ripple resolution is one that requires the listener to distinguish a spectrally rippled noise from an unrippled (spectrally flat) noise. Unlike the spectral ripple reversal task, where the spectral ripple rate is varied and the modulation depth held constant, the spectral ripple detection paradigm typically employs a constant ripple rate while the peak-to-valley modulation depth of the spectral ripples is varied. The listener's task is to detect the presence of spectral ripples, or sinusoidal variations in amplitude on a log frequency axis, in a broadband noise signal. The spectral ripple detection threshold, or spectral modulation threshold (SMT), represents the smallest spectral contrast in a rippled noise signal that can be perceived by the listener and discriminated from an unmodulated standard noise stimulus.

When SMTs are measured individually at different spectral modulation rates, the resulting pattern of SMT as a function of modulation rate is referred to as the spectral modulation transfer function, or SMTF (e.g. Saoji, Litvak, Spahr, and Eddins, 2009). In normal-hearing listeners, sensitivity to spectral modulation has been shown to be most sensitive in the middle of the spectral modulation frequency range (approximately 1-4 rpo) and is reduced at lower and higher modulation frequencies (e.g. Saoji and Eddins, 2007).

Poor spectral resolution causes a smoothing of the spectral envelope of the internal representation of the acoustic spectrum of a complex signal such as rippled noise or speech (Horst, 1987). Reducing spectral contrasts has been shown to result in decreased speech recognition ability (e.g. Bacon and Brandt, 1982; Van Tasell *et al*., 1987). Similarly, the effective reduction of spectral contrast imposed by broader auditory filters should produce higher (poorer) thresholds for spectral modulation detection for individuals with poor spectral resolution. The reduction of spectral contrast has been shown to be greater for higher than lower spectral modulation frequencies (Eddins and Bero, 2007; Saoji and Eddins, 2007; Summers and Leek, 1994), and so should have a greater effect on the spectral ripple detection thresholds for higher ripple rates than for lower rates.

Eddins and Bero (2007) conducted a study to characterize SMTFs in normal-hearing listeners. They found that the general form of the SMTF is basically bandpass, with best modulation detection in the region between 2 and 4 rpo. Neither carrier bandwidth (1-6 octaves) nor carrier frequency region (200-12,800 Hz) influenced spectral modulation detection thresholds, with the exception of carrier bands restricted to very low audio frequencies (e.g., 200-400 Hz), where ripple detection was poorer; this exception could not be accounted for by computed excitation patterns taking into account the broader auditory filters of that frequency region.

Summers and Leek (1994) explored the relationship between frequency resolution and spectral contrast detection in normal-hearing and hearing-impaired listeners. They used a notched-noise method to define auditory filter characteristics at 1 and 3 kHz and a spectral ripple detection task for nine ripple frequencies ranging from 1 to 9 rpo. They found that while there were significant differences between groups in both spectral contrast detection and auditory filter characteristics, the calculated amount of contrast in the internal representations of the stimuli based on the auditory filter measurements was essentially equivalent for all listeners, suggesting that the reduced spectral resolution of the hearing-impaired listeners accounted for differences in spectral contrast required for detection of spectral ripples.

Under the assumption that spectral ripple detection requires spectral resolution, Litvak *et al.* (2007) used the spectral ripple detection paradigm to explore the role of spectral resolution as a possible source of variance in word recognition in CI users. They introduced varying amounts of spectral smearing to vocoder-processed rippled noise stimuli, in order to match performance of normal-hearing listeners on the spectral ripple task to that of CI listeners. Their goal was to then compare performance on measures of vowel and consonant identification by CI listeners to performance of normal-hearing simulation listeners producing the same spectral ripple detection thresholds. Twenty-five Advanced Bionics CI users were included; their speech perception and SMT data were taken from a previous report (Saoji, Litvak, Emadi, Spahr, and Greenslade, 2005). Using a Direct Connect system for direct stimulation of the electrode array, bypassing the external personal speech processors, SMTs were obtained with a cued 2-interval, 2-alternative forced choice procedure at ripple frequencies of 0.25 and 0.5 rpo. Thresholds from 3 runs were averaged for each modulation frequency, and the mean of thresholds at 0.25 and 0.5 rpo was used as the index of spectral resolution for each listener, since that combination was found to best correlate with consonant and vowel recognition in the CI listeners (Saoji *et al.*, 2005).

Vowel stimuli were presented at 60 dB SPL, and identification performance was measured in two blocks of 78 trials (13 vowels presented six times in random order) for a total of 12 presentations per vowel. Consonant identification was based on two blocks of 80 trials (16 consonants presented five times) for a total of 10 presentations per consonant.

For vocoder simulations, the temporal envelopes from 15 analysis channels were used to modulate noise bands matched in CF to the analysis channel. The filter slopes (rate of drop-off of the noise spectrum on either side of CF) varied from 5-40 dB/octave to simulate different degrees of spectral smearing (or spread of excitation in the electrically-stimulated cochlea). Ten normal-hearing subjects each listened to four different vocoder conditions with filter slopes of 5, 10, 20, and 40 dB/octave. For each of these simulation conditions, SMT at 0.25 and 0.5 rpo and vowel and consonant identification performance were measured.

Average SMTs for CI listeners ranged from 2.25 to 25 dB, similar to the range for normal-hearing simulation (NH-sim) listeners. Average SMTs for the NH-sim listeners increased from 6.3 dB for the steepest filter slopes (40 dB/octave) to 22.5 dB for the shallowest (5 dB/octave). SMT strongly correlated with vowel recognition across both groups of listeners (r^2 = 0.71), with vowel scores for NH-sim listeners decreasing as noise-band slopes broadened. The best-fit regression lines for NH-sim listeners were not statistically different from those for the CI group, implying that vocoder simulations model both spectral resolution and vowel identification performance in CI listeners. SMT also correlated with consonant recognition for both CI (r^2=0.67) and NH (r^2= 0.77) groups, but compared to vowel scores, consonant scores dropped off less with decreased spectral resolution, perhaps reflecting less dependence of consonant recognition on spectral information.

These results are somewhat counterintuitive, given that detection thresholds for higher ripple frequencies, not lower ones, should be more sensitive to reduced spectral resolution. Saoji, Litvak, Spahr and Eddins (2009) investigated this discrepancy; they extended the Litvak *et al*. (2007) study, comparing ripple detection thresholds for a greater range of spectral modulation rates (0.25, 0.5, 1.0, and 2.0 rpo) for the same 25 subjects included in the earlier study. Their supposition was that if spectral ripple detection is mediated by some single factor relating to spectral resolution, then variability in spectral ripple detection between listeners should be equivalent at all spectral ripple rates, and phoneme recognition would theoretically correlate equally well with ripple detection thresholds at all ripple frequencies. However, if spectral ripple detection involves different processes at different ripple rates, then the shapes of SMTFs might differ across subjects, and correlations with phoneme recognition should differ among spectral ripple detection thresholds at different modulation frequencies. Saoji *et al.* (2009) measured SMDTs with an adaptive, cued, 3I-2AFC procedure with feedback. Stimuli were presented at 60 dB SPL. Tracks for the four modulation frequencies were presented in random order. Thresholds were based on three runs.

Results showed that spectral ripple detection was poorer at higher modulation frequencies than lower frequencies across subjects ($p < 0.0001$), but that the shapes of SMTF varied across subjects. The strongest predictor of vowel identification was 0.25 rpo ($r^2=0.59$), and 0.5 rpo was the sole predictor for consonants. This pattern of results mirrors that of Litvak *et al.* (2007). The authors speculated that the correlation between detection of low spectral modulation frequencies and vowel and consonant identification might not be related to spectral resolution per se, but rather to differences in listeners' ability to compare widely spaced spectral maxima and minima in the spectral envelope spanning a broad frequency range.

They also conducted a small study to examine the potential for loudness cues to be used in the ripple detection task. Nine listeners from the larger pool participated in an experiment where they compared the loudness of fixed-level spectrally rippled noise stimuli to unrippled noises that varied in 1 dB steps over a range from 55-65 dB SPL, for a total of eleven comparisons. Ten repetitions for each of the four modulation frequencies were made. Results indicated a shift of 1 dB (a standard stimulus of 59 dB was matched to a spectrally rippled noise of 60 dB), but this shift represented less than half the value of the calculated difference limen, making it unlikely that loudness differences were usable cues for the ripple detection task.

The authors’ conclusions were that the different shapes of SMTFs for CI listeners reflect more than a single underlying spectral resolution capacity. The significant correlations between phoneme recognition and spectral ripple detection thresholds at low, but not high, modulation frequencies suggest that reduced spectral resolution prevents the use of fine spectral detail carried by higher spectral modulation frequencies; as a result, listeners are forced to rely on features in the broad spectral envelope (i.e. low modulation frequencies).

How does performance on the spectral ripple discrimination task (e.g. Won *et al*., 2007; Anderson *et al*., 2011) compare to SMTs in CI listeners? An experiment was undertaken to explore the relationship between the two measures. Specifically, SMTFs were measured in CI subjects and compared to their spectral ripple discrimination thresholds (Chapter 2). If the two measures are related, then the spectral modulation frequency yielding a SMT of 30 dB (the depth used in the spectral ripple discrimination task) should correlate with the spectral ripple discrimination threshold.

In addition, SMTs were compared to speech perception measures, to see if the Litvak *et al*. (2007) and Saoji *et al*. (2009) findings of a relationship between detection of broadly-spaced ripples and measures of speech recognition could be replicated.

II. EXPERIMENT 3.1: SPECTRAL RIPPLE DETECTION

A. Subjects

The same 15 CI subjects (5 Clarion I, 5 Clarion II, and 5 Nucleus-22) participated in this study as in the ones detailed in Chapter 2. See Table 2.1 for individual subject characteristics.

B. Stimuli

Stimuli were patterned after those used by Litvak, *et al*. (2007) and were generated using Matlab software (The Mathworks, Natick, MA). Sinusoidal modulations were applied to a broadband (350-5600 Hz) Gaussian noise, producing variations in level (dB) on a log-frequency axis (see Equation 1, Chapter 2).

Noises were generated at seven fixed ripple rates (0.25, 0.5, 0.75, 1.0, 1.5, 2.0, and 3.0 rpo). Modulation depth was varied adaptively for this task. The signal duration was 400 ms, which included 20-ms raised-cosine onset and offset ramps. The stimuli were presented in a double-walled, sound-attenuating booth through a single loudspeaker (Infinity RS1000) located 1 m from the subject's seated position, at approximately head height. The average sound level of the noise was 60 dBA when measured at a point corresponding to the position of the subject's head. The level was roved across intervals within each trial by ±3 dB, and the starting phase of the spectral modulation was randomized for each trial, in order to minimize potential loudness cues. A schematic of the rippled noise stimuli is shown in Figure 3.1.

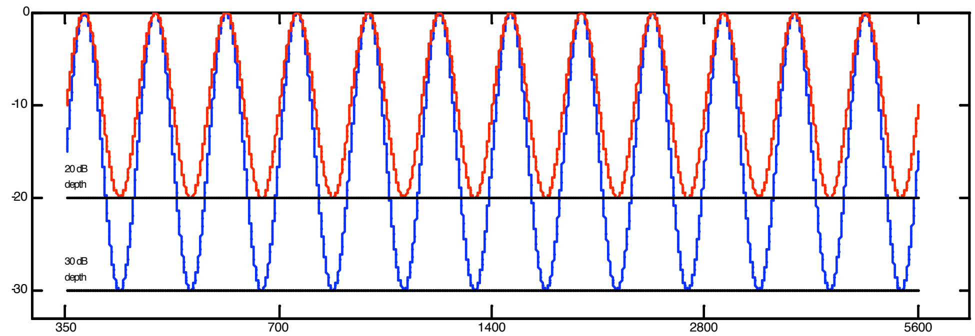


Fig. 3.1. Schematic representation of rippled noise (3 rpo, here) used for the spectral ripple detection experiment. One curve represents a modulation depth of 20 dB, and the other a modulation depth of 30 dB, with ripples per octave held constant.

C. Procedure

Subjects wore their own speech processors at typical use settings. A three-interval, three-alternative forced-choice (3I-3AFC) adaptive procedure was employed. During each trial, subjects heard two intervals of unrippled broadband noise and one interval of rippled noise, presented in random order so that each interval had the same probability of containing rippled noise on a given trial. Subjects indicated which interval they judged as sounding different (corresponding to the rippled signal) by selecting the appropriate square on a computer screen. Each test run began with a modulation depth for the rippled stimulus of 20 dB; the modulation depth was then varied adaptively in a 1-up, 2-down psychometric procedure to bracket threshold, or the smallest detectable peak-to-valley value. The initial step size was 4 dB, changing to 2 dB after the first two reversals, and to 0.5 dB following two more reversals. Termination of the run occurred following ten reversals. Spectral ripple detection threshold, or SMT, was defined for each run as the geometric mean of the PVR at the final six reversal points.

Each subject completed six runs for each of the seven ripple rates, for a total of 42 experimental runs. Thresholds from the first run for each ripple-rate condition were rejected as practice runs, and any thresholds that were more than 3 standard deviations away from the mean of the remaining measurements for that condition were excluded. Typically, thresholds from five runs for each ripple rate were used to calculate an arithmetic mean threshold for each subject.

D. Results

Spectral ripple detection thresholds varied widely across subjects and across ripple frequencies. Individual thresholds ranged from about 5 dB to 45 dB PVR. These results are in line with those reported by Saoji *et al*. (2005). In general, greater spectral modulation was required for detection as the ripple frequency increased, although some subjects showed non-monotonicities in their SMTFs, illustrated in Figure 3.2.

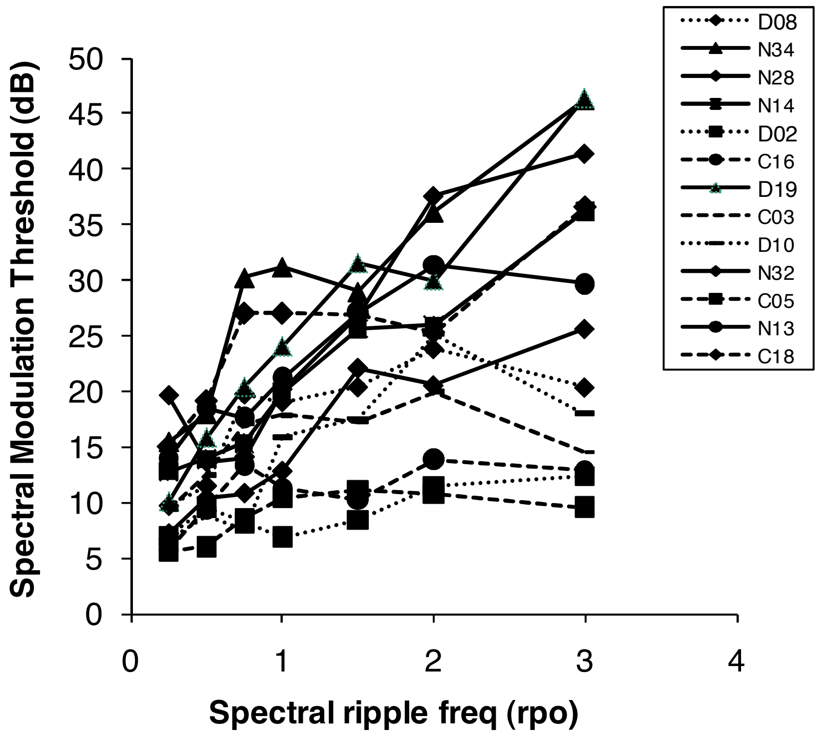


Figure 3.2. SMTFs for 15 CI subjects. SMT (in dB) for each of 7 different spectral ripple frequencies is depicted.

Spectral ripple detection and spectral ripple discrimination.

If the perceptual judgment required for spectral ripple detection is related to that required for the spectral ripple discrimination task (Chapter 2), then it might be expected that the ripple frequency at which a detection threshold of 30 dB is obtained should correspond to the spectral ripple discrimination threshold, since the modulation depth for the ripple discrimination stimuli was 30 dB. The spectral ripple detection thresholds measured did not reach 30 dB for all subjects for the highest ripple rate tested (3 rpo).

Consequently, the spectral ripple detection task was conducted at additional extended ripple frequencies (6, 9, 12, 15, 18, and 21 rpo). From this extended SMTF, the spectral ripple frequency corresponding to a ripple detection threshold of 30 dB was estimated by a linear interpolation from existing points on the SMTF. (Simple linear interpolation, rather than fitting functions to the data points, was chosen because of the irregularity and non-monotonicities present in some of the SMTFs.) See Appendix, Table A1, for the extended SMTF measures for each subject.

Figure 3.3, below, illustrates the relationship between the interpolated spectral ripple detection threshold and spectral ripple discrimination thresholds obtained from the previous experiment (Chapter 2). Linear regression analysis revealed a significant correlation between the two variables (r^2=0.41, p = 0.01), but there was not a one-to-one correspondence between the two measures. In fact, the spectral ripple detection measures were generally substantially higher than the ripple discrimination thresholds.

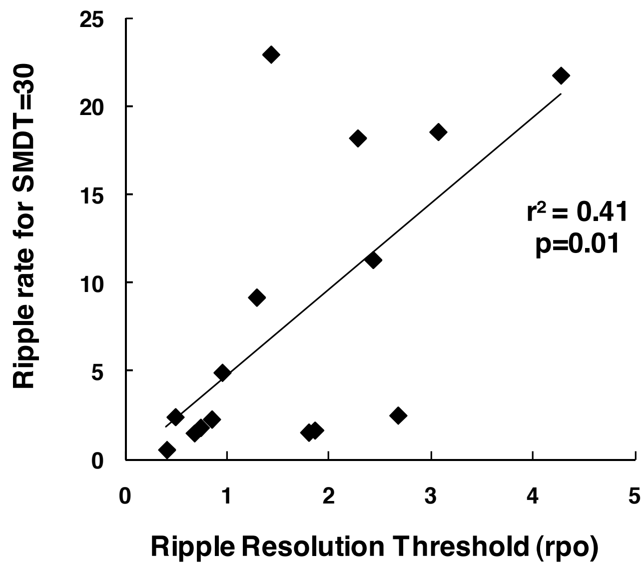


Figure 3.3. Interpolated ripple frequency for SMT=30dB, as a function of spectral ripple resolution threshold (rpo).

Spectral ripple detection and speech recognition

Spectral detection thresholds were examined in relation to IEEE sentence and Hillenbrand vowel recognition performance (see Chapter 2 for details on speech recognition testing). In addition to standard performance measures, SINFA analysis (Wang and Bilger, 1973) was completed on vowel results; rate of transmitted information (RTI) was determined for five features: F1, F2, [F1+F2], [F2-F1], and stimulus duration.

Ripple detection thresholds at several different ripple frequencies showed statistically significant correlations with speech measures. In particular, the average of ripple detection thresholds at 0.25 and 0.5 rpo showed strong relationships with sentence recognition in quiet, as shown in Figure 3.4, Panel A, and vowel recognition, Panel B. Relationships with sentences and vowels in noise (Figure 3.4, Panels C and D, respectively) showed trends but failed to reach significance.

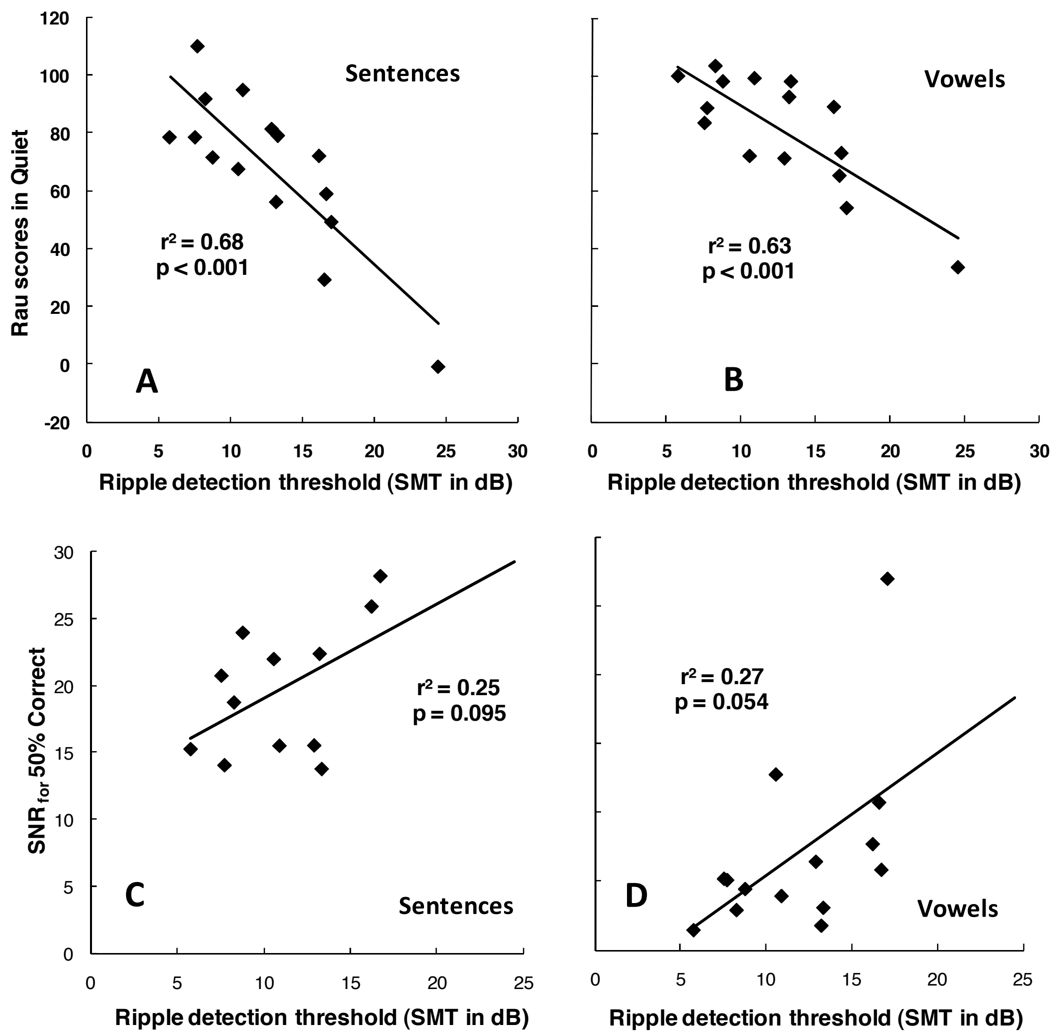


Figure 3.4. Rau scores for sentence recognition (Panel a) and vowel recognition (Panel b) as a function of average SMT for 0.25 and 0.5 rpo. Panels c and d display $SNR_{50\%}$ for sentences and vowels, respectively, as a function of the average of ripple detection thresholds at 0.25 and 0.5 rpo.

Table 3.1 displays results for all of the regression analyses for ripple detection thresholds at each ripple frequency tested, with each of the speech measures obtained. While significant correlations are seen for comparisons at a number of different ripple frequencies, the average of 0.25 and 0.5 rpo shows consistently strong correlations for sentence and vowel recognition in quiet, in agreement with other studies (Litvak *et al.*, 2007; Saoji *et al.*, 2009).

Table 3.1. Correlations and corresponding p values (with no post-hoc Bonferroni corrections for multiple comparisons) for SMT at various ripple frequencies and speech recognition measures; highlighted are r^2 values >0.60.

	ripple detection: SMT as a function of ripple frequency							
Speech Measure	*0.25 rip/oct*	*0.50*	*avg. (.25, .5)*	*0.75*	*1.00*	*1.50*	*2.00*	*3.00*
Sentence recognition, Q (rau)	**r^2=0.729, p<0.001**	r^2=0.551 p=0.002	**r^2=0.683 p<0.001**	r^2=0.512 p=0.003	r^2=0.252 p=0.067	r^2=0.3 p=0.042	r^2=0.329 p=0.032	r^2=0.344 p=0.027
$SNR_{50\%}$, Sentences	r^2=0.233, p=0.112	r^2=0.256 p=0.093	r^2=0.254 p=0.095	r^2=0.301 p=0.065	r^2=0.146 p=0.220	r^2=0.109 p=0.295	r^2=0.178 p=0.172	r^2=0.194 p=0.332
Vowel identification, Q (rau)	r^2=0.561, p=0.001	**r^2=0.601 p=0.001**	**r^2=0.627 p<0.001**	**r^2=0.631 p<0.001**	r^2=0.424 p=0.012	r^2=0.242 p=0.070	r^2=0.267 p=0.059	r^2=0.376 p=0.020
$SNR_{50\%}$, Vowels	r^2=0.218, p=0.092	r^2=0.276 p=0.054	r^2=0.276 p=0.054	r^2=0.271 p=0.056	r^2=0.424 p=0.012	r^2=0.242 p=0.074	r^2=0.267 p=0.059	r^2=0.376 p=0.020
SINFA Analysis								
RTI-Stim, Vowels	r^2=0.330, p=0.032	r^2=0.418 p=0.013	r^2=0.417 p=0.013	r^2=0.502 p=0.005	r^2=0.463 p=0.007	r^2=0.245 p=0.072	r^2=0.220 p=0.091	r^2=0.388 p=0.017
RTI-F1, Vowels	r^2=0.449, p=0.009	r^2=0.354 p=0.025	r^2=0.452 p=0.008	r^2=0.421 p=0.012	r^2=0.445 p=0.009	r^2=0.285 p=0.049	r^2=0.355 p=0.025	r^2=0.457 p=0.008
RTI-F2, Vowels	r^2=0.037, p=0.513	r^2=0.139 p=0.189	r^2=0.088 p=0.304	r^2=0.206 p=0.103	r^2=0.181 p=0.129	r^2=0.052 p=0.434	r^2=0.035 p=0.52	r^2=0.054 p=0.423

For the SINFA measures, best correlations were observed for rate of transmitted information for first formant (RTI-F1) of vowels, a measure corresponding to reception of first formant information in the vowel stimuli. Figure 3.5 displays the relationship between average SMT for 0.25 and 0.5 rpo and RTI for F1.

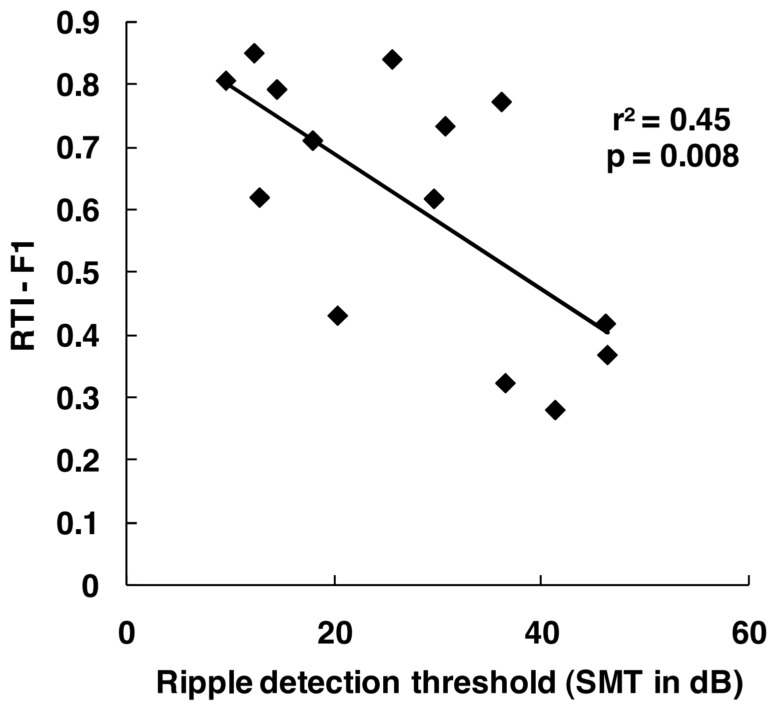


Figure 3.5. RTI-F1 for vowel identification as a function of ripple detection threshold (average of 0.25 and 0.5 rpo)

E. Discussion

Spectral ripple detection as a measure of spectral resolution.

If detection of spectral ripples at higher ripple frequencies is tapping into spectral resolution, then that ability should correlate with spectral ripple discrimination. The data, however, failed to match that prediction. Specifically, an index of ripple detection performance was selected that corresponded directly to the same ripple modulation depth that was used in the ripple reversal test. A direct, one-to-one correspondence between these results would be assumed if they measure the same phenomenon. A linear regression analysis, comparing each subject's ripple discrimination threshold (Chapter 2) to the ripple frequency yielding an SMT of 30 dB, showed that the measures were correlated but not equivalent. These results fail to support the contention of Saoji *et al.* (2009) that thresholds from the spectral ripple discrimination task represent points on the high-modulation-frequency extreme of the SMTF.

Spectral ripple detection for a modulation depth of 30 dB generally fell at a higher ripple rate than the corresponding ripple discrimination thresholds. This pattern of results does not make intuitive sense; the amount of spectral contrast between target and comparison stimuli, at any given point on the waveform, is less for the ripple detection task (which used flat noise as the comparison stimulus, producing 15 dB contrast at a spectral peak) than for the ripple-reversal discrimination task (with a phase-reversed comparison stimulus, producing 30 dB contrast at a spectral peak). Therefore, one might expect better performance on the ripple discrimination task, if both are based on perception of spectral contrast.

Potential cues for spectral modulation detection at higher ripple rates.

The extended-frequency ripple detection results raised a question of the validity of this measure. Some subjects were able to detect spectral ripples as closely spaced as 20 rpo, well beyond their spectral ripple discrimination thresholds and predicted capabilities.

Figure 3.6 demonstrates an example, where a non-monotonicity occurs in the function with a dip representing better performance (lower modulation depth at threshold) at higher ripple densities.

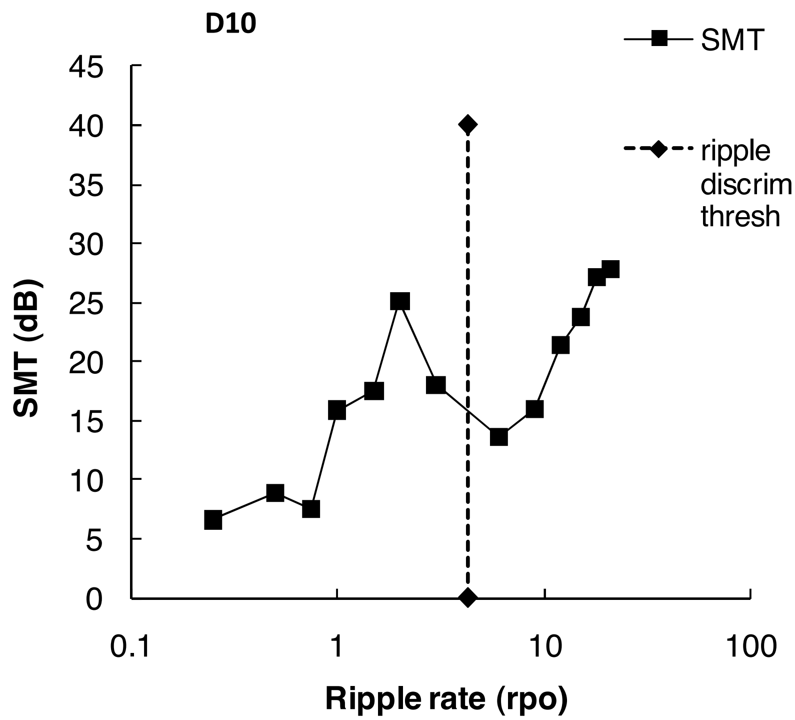


Figure 3.6. Individual SMTF (SMT as a function of spectral ripple frequency) for subject D10. The vertical line indicates the spectral ripple discrimination threshold for this listener.

It seems implausible that listeners could perform this task based solely on spectral cues. Even in normal acoustic hearing, ripple rates > 5 or 6 rpo should not be spectrally resolvable, since the equivalent rectangular bandwidths (ERBs) of the auditory filters are approximately one third- to one sixth-octave wide. To further investigate factors in performance on this task, three normal-hearing subjects were tested on an identical spectral ripple detection procedure in the sound field. Figure 3.7 shows the results. Performance was obtained well beyond the theoretical limit of 5-6 rpo.

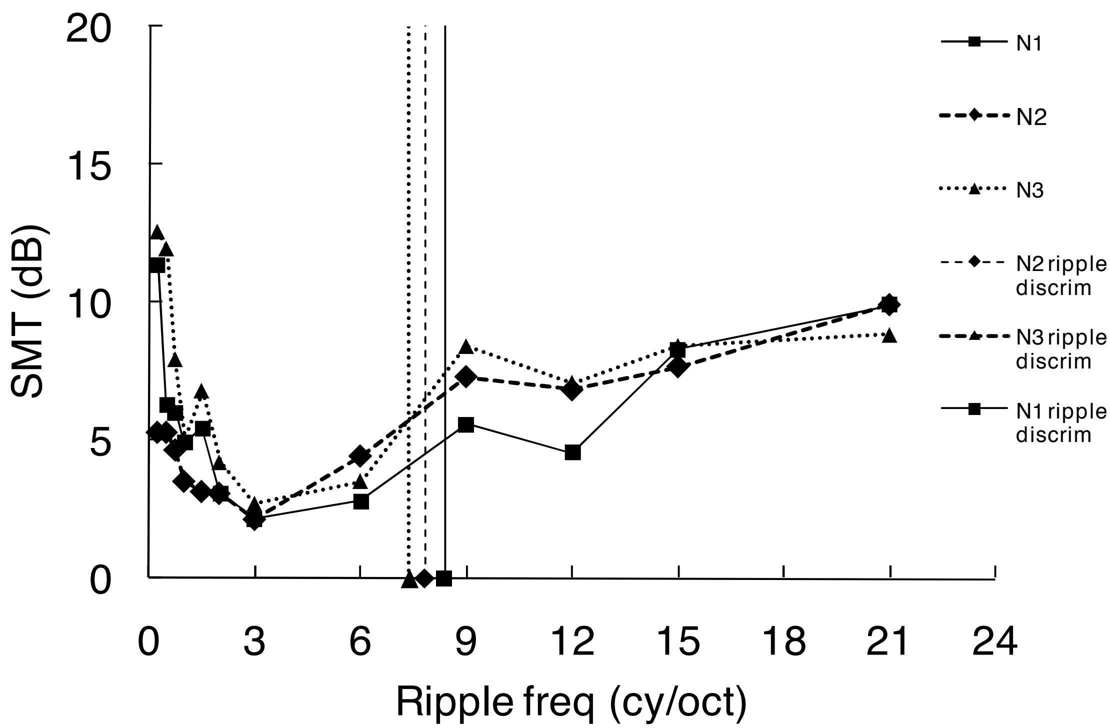


Figure 3.7. SMTFs for 3 normal-hearing subjects. SMT (in dB) for each of 12 different spectral ripple frequencies (0.25-21 rpo) is displayed. Spectral ripple discrimination thresholds are indicated by vertical lines.

This result raised the question of whether interactions between adjacent peaks in the spectrally-rippled noise might be producing temporal cues. Specifically, closely-spaced spectral peaks might produce beats or perceived “roughness” in the temporal envelope, providing a salient cue for discriminating the rippled signal from the flat-spectrum noise standard. The ability of listeners to detect this type of temporal artifact in the stimuli might correlate with their performance on a measure of temporal resolution, such as sinusoidal amplitude modulation (SAM) detection. This hypothesis was evaluated in the same group of CI subjects.

III. EXPERIMENT 3.2: SAM DETECTION

For this psychophysical procedure, the listener's task was to detect the presence of amplitude modulations, or regular variations in level as a function of time, within a broadband noise signal.

A. Subjects

The same 15 CI subjects (5 Clarion I, 5 Clarion II, and 5 Nucleus-22) participated in this study as in the previous ones. See Table 2A for individual subject characteristics.

B. Stimuli

Stimuli were generated using Matlab software (The Mathworks, Natick, MA), and differed from the stimuli used in the previous two experiments in that sinusoidal modulations were applied to the temporal envelope, rather than spectral envelope, of a broadband (350-5600 Hz) Gaussian noise, producing variations in level (dB) in the time domain.

The modulation rate was varied adaptively for this task, with modulation depth held constant at 30 dB. The signal duration was 400 ms, which included 20-ms raised-cosine onset and offset ramps. The stimuli were presented in a double-walled, sound-attenuating booth through a single loudspeaker (Infinity RS1000) located 1 m from the subject's seated position at approximately head-height. The average sound level of the noise was set to 60 dBA when measured at the location corresponding to the subject's head. The level was roved across intervals within each trial by ±3 dB, in order to reduce possible loudness cues. A random starting phase of the amplitude modulation was used for each trial.

C. Procedure

Subjects wore their own speech processors at typical use settings. A three-interval, three-alternative forced-choice (3I-3AFC) adaptive procedure was used. During each trial, subjects heard two intervals of flat broadband noise and one interval of sinusoidal amplitude-modulated (SAM) noise. The presentation order of the intervals within each trial was randomized, so that all intervals had the same probability of containing SAM noise on each trial. Subjects indicated which interval they judged as sounding different (corresponding to the SAM signal) by selecting the appropriate square on a computer screen. Each test run began with a modulation frequency for the SAM stimulus of 250 Hz, which was then varied adaptively in a 2-up, 1-down psychometric procedure to bracket threshold (the highest detectable amplitude modulation rate). The modulation rate initially increased or decreased by a factor of 2. After the first two reversals the step size changed to a factor of 1.41, and after two more reversals, to 1.19. Peak-to-valley modulation depth was held constant at 30 dB. Termination of a run occurred after ten reversals, and threshold was defined for each run as the geometric mean of the PVR at the final six reversal points.

Each subject completed six runs of the experiment. Thresholds from the first run were excluded as "practice" runs, and any thresholds that were more than 3 standard deviations away from the mean of the remaining measurements for that condition were excluded; most of the time, thresholds from five runs were used to calculate a geometric mean threshold for each subject.

D. Results and Discussion

SAM detection thresholds varied widely across subjects. Individual thresholds ranged from about 48.52 Hz to 258.69 Hz. Figure 3.8 displays individual SAM detection thresholds for all subjects.

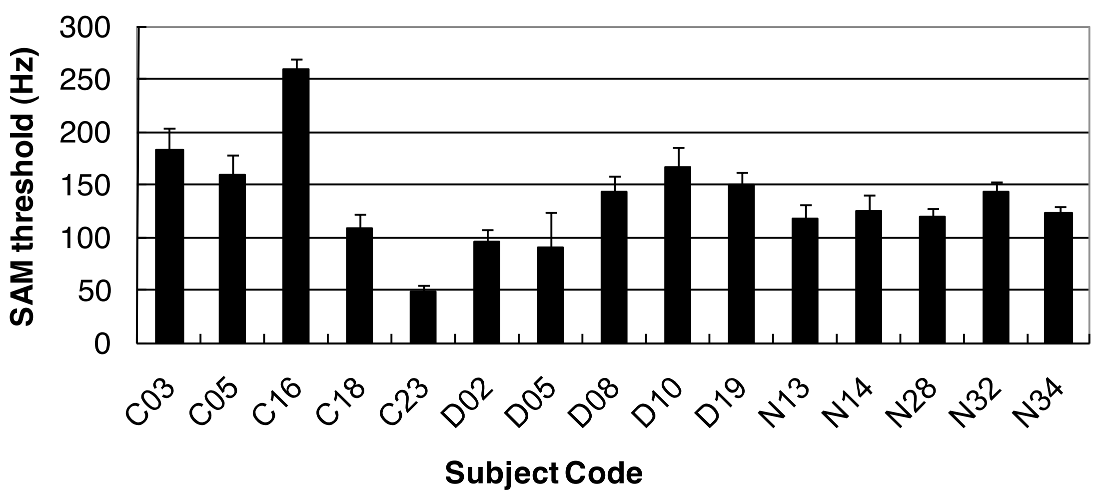


Figure 3.8. Individual SAM thresholds (in Hz). Error bars represent one sd.

The relationship between SAM detection and spectral ripple detection was examined using linear regression analysis. In order to best equate the comparison, thresholds for the highest ripple frequency detectable at 30 dB spectral modulation depth were compared to SAM detection thresholds. Figure 3.9 shows the relationships.

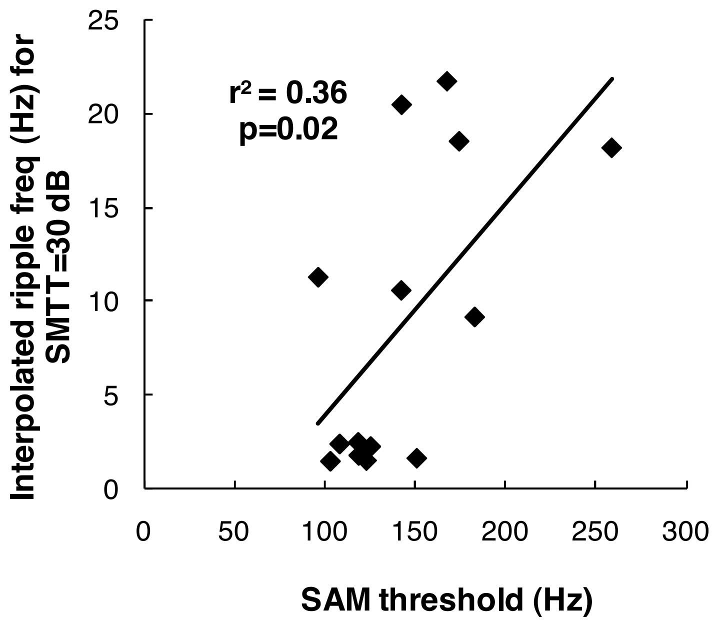


Figure 3.9. Interpolated ripple frequency corresponding to SMT of 30 dB as a function of SAM threshold.

Results indicate that the higher the detectable spectral ripple rate (at a 30 dB modulation depth), the higher the SAM detection threshold. This, in turn, suggests that the use of temporal cues at higher ripple rates in the spectral ripple detection task cannot be ruled out.

The relationship between SAM detection and spectral ripple discrimination was also examined to see if there was any evidence of the use of temporal cues in the spectral ripple reversal task. Figure 3.10 does illustrate a trend toward a correlation, but it did not reach significance.

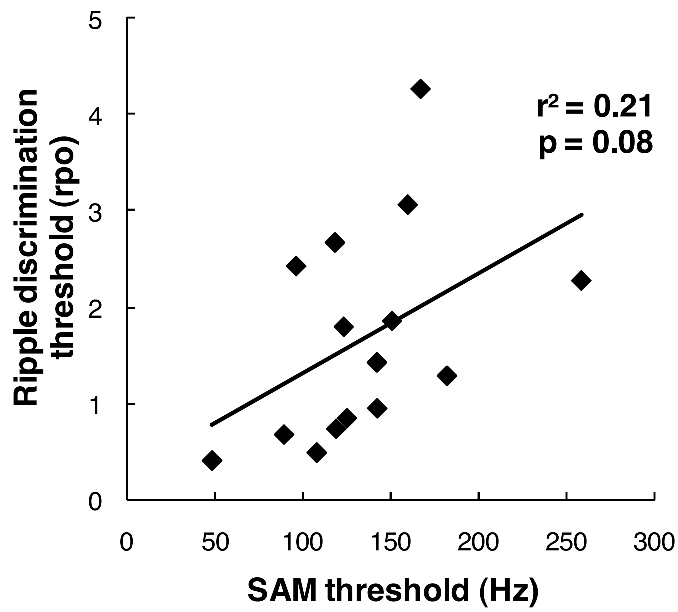


Figure 3.10. Ripple discrimination threshold as a function of SAM threshold.

Another potential non-spectral cue for performing the spectral ripple detection task relates to loudness differences between stimuli. For instance, disparities in growth of loudness across electrodes might introduce perceptible loudness differences between stimuli that could be used as cues. If listeners were using loudness cues to perform the ripple detection task, then there might be a correlation between performance on that task and on a task of intensity discrimination.

Further, the detection and discrimination of spectral ripples are fundamentally based on the perception of across-spectrum amplitude differences. This may represent a different capacity than spectral resolution. In other words, ripple detection and discrimination represent the combination of spectral resolution and (across-channel) intensity resolution, or profile analysis. This next experiment measures broadband intensity perception in order to tap into this second capacity.

IV. EXPERIMENT 3.3: INTENSITY DISCRIMINATION

The purpose of this experiment was to measure intensity difference limens, using broadband, spectrally flat noise stimuli, similar in bandwidth to stimuli from the previous experiments. Subjects had to judge which of three stimulus intervals was loudest. Performance on this task was then compared to spectral ripple detection and discrimination results.

A. Subjects

14 of the original 15 CI subjects (4 Clarion I, 5 Clarion II, and 5 Nucleus-22) participated in this study. (Data for Subject C23 was not available.) See Table 2A for individual subject characteristics.

B. Stimuli

Broadband (350-5600 Hz) Gaussian noise stimuli were generated using Matlab software. The intensity of the target stimulus was varied adaptively for this task. The signal duration was 400 ms, which included 20-ms raised-cosine onset and offset ramps. The stimuli were presented in a double-walled, sound-attenuating booth through a single loudspeaker (Infinity RS1000) located 1 m from the subject's seated position at approximately head height. The average sound level of the noise was set to 60 dBA when measured at the location corresponding to the subject's head. The starting phase of the amplitude modulation was selected at random for each trial.

C. Procedure

Subjects wore their own speech processors at typical use settings. A three-interval, three-alternative forced-choice (3I-3AFC) adaptive procedure was used. During each trial, subjects heard two intervals of the standard noise and one interval of a noise that differed in intensity from the other two. The presentation order of the intervals within each trial was randomized, so that all intervals had the same probability of containing the target on each trial. Subjects indicated which interval they judged as louder (corresponding to the target signal) by selecting the appropriate square on a computer screen. Each test run began with an increment in the target signal of 10 dB, relative to 60 dB (or an SNR of 10 dB); the SNR was then varied adaptively in a 1-up, 2-down psychometric procedure to bracket ΔL, or the smallest detectable difference in level (difference limen). The initial value of ΔL (the difference in level between the target and standard stimuli) was approximately 10 dB, changing to 5 dB after the first two reversals, and to 2.5 dB following two more reversals. Termination of the run occurred following ten reversals. Threshold was defined for each run as the geometric mean of intensity difference limen at the final six reversal points.

Each subject completed six runs of the intensity discrimination experiment. Thresholds from the first run were eliminated as "practice" runs, and any thresholds that were more than 3 standard deviations away from the mean of the remaining measurements for that condition were excluded; in general, thresholds from five runs were used to calculate an arithmetic mean threshold for each subject.

D. Results

Intensity discrimination thresholds varied widely across subjects. Individual intensity difference limens (ΔL) ranged from about 1.03 dB to 6.62 dB. Figure 3.11 displays ΔL for each subject.

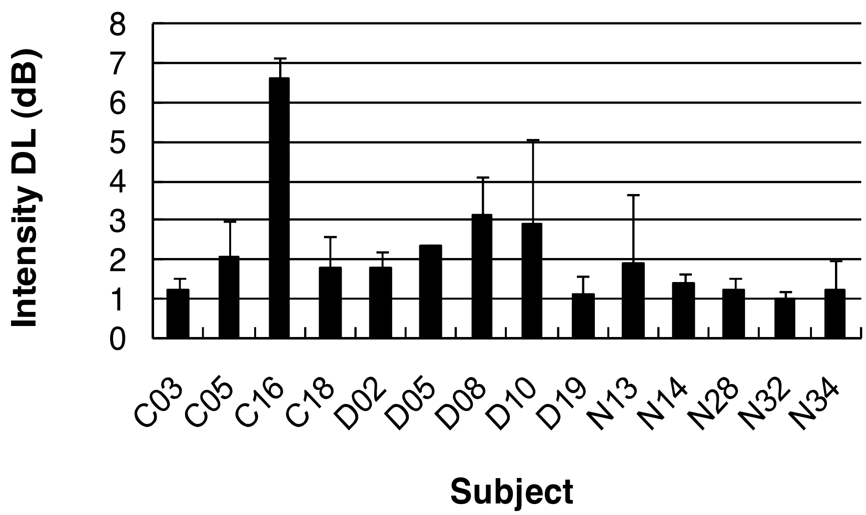


Figure 3.11. Intensity difference limens (DL) for each subject.

Intensity difference limens were compared to spectral ripple detection thresholds, using regression analysis; intensity discrimination did not correlate with spectral ripple detection at any modulation frequency (e.g. ave. SMT [0.25, 0.5 rpo] as a function of intensity discrimination: $r^2 = 0.18$, $p = 0.127$), nor with spectral ripple discrimination thresholds (all regression analyses resulted in $p>0.05$).

E. Discussion

If spectral ripple detection at very low ripple rates depends on spectral profile analysis, or the ability to discriminate differences in intensity across frequency, then we might have expected a correlation between low-rate ripple detection and intensity discrimination. The fact that none was found suggests that different mechanisms may mediate intensity discrimination over time and intensity discrimination across frequency.

V. SUMMARY

The findings from the series of experiments reported in this chapter are consistent with the idea that spectral capacities are central in determining spectral ripple thresholds, supporting the validity of the use of ripple discrimination tasks as measures of spectral resolution. However, the spectral ripple detection measures that seem to correlate best with speech perception in noise are those at very low spectral rates, which seem unlikely to test spectral resolution *per se*. We explored the relationships between low-rate ripple detection and intensity resolution, but failed to find a significant correlation between low-rate ripple detection thresholds and intensity difference limens. Therefore, it does not seem possible to explain the link between low-rate ripples and speech perception in terms of a general capacity for intensity resolution. Finally, it was determined that high-rate ripple detection may be influenced by temporal-envelope cues; however, the relationship between SAM detection and high-rate spectral ripple was not strong.

CHAPTER 4: FURTHER INVESTIGATIONS OF SPECTRAL RESOLUTION AND SPEECH PERCEPTION IN NOISE

I. BACKGROUND

While broadband spectral ripple discrimination measures showed relationships with sentence recognition in quiet (Chapter 2), they did not demonstrate similar correlations with speech measures in steady noise. This result is perhaps surprising, given that spectral resolution might be predicted to play a bigger role in speech perception in noise than in quiet (see Chapter 1). In addition, other studies (e.g. Henry *et al.*, 2005; Won *et al.*, 2007) have reported strong relationships between spectral ripple discrimination and speech recognition in noise in CI users. On the other hand, if spectral resolution can vary across the electrode array for a given individual, it might be possible that performance on the broadband spectral ripple discrimination task is selectively mediated by regions of better spectral resolution and not hurt by regions of poorer resolution, whereas speech recognition, particularly in noise, might suffer from the loss or degradation of information coded by a localized region of poorer resolution. In that case, a relationship between spectral ripple discrimination and speech recognition might be weakened, especially if the region of poorer resolution falls in a spectral region particularly important for speech recognition.

It is not well understood how the different frequency regions are weighted for intelligibility of speech for CI listeners. In acoustic hearing, with normal-hearing listeners, the most important frequency region for speech recognition has been shown to lie at approximately 1500 Hz (e.g. DePaolis and Janota, 1996), but this assumption may not hold for CI processing. Two recent reports attempted to measure frequency importance functions for normal-hearing subjects listening to vocoded consonant materials. Whitmal and DeRoy (2011) used an adaptive technique of measuring word recognition of incrementally high-pass and low-pass filtered vocoded speech materials, in order to determine a crossover cutoff frequency for performance, representing the peak in the importance function. They showed that the region around 700 Hz contained the most important information in vocoded consonant stimuli. Apoux and Healy (2011) used a modified correlational method, in which they selectively introduced spectral gaps in vocoded speech to find the frequency bands most critical for consonant recognition. They found three regions of greatest importance: approximately 500 Hz, presumably reflecting voicing information; 1800 Hz, corresponding to place of articulation; and 7000 Hz, representing plosive and fricative features. Since both of these studies focused on consonant materials, their findings may not be strictly relevant to sentence recognition, but both of these studies suggest that information in the lower frequency region of the speech spectrum (500 -700 Hz), is important for phoneme recognition in CI listeners. It follows that noise comprising a frequency region containing more speech information might be predicted to have a more adverse effect on speech understanding than a noise from a spectrally less-informative frequency range.

The next experiment to be reported examined the relationship between spectral resolution and sentence recognition in band-limited noise (i.e. noise that is bandpass-filtered in such a way that it spectrally matches only a portion of the spectrum of the speech materials.) If Cooke's (2006) glimpsing model of speech recognition in noise is correct, then limiting the spectrum of a noise masker so that it doesn't completely overlap the speech signal should result in more "clean" glimpses of the target speech, leading to better word recognition, and perhaps a better correspondence between spectral resolution and speech recognition in bandpass noise. There are two reasons for this prediction of a relationship between spectral resolution and performance in bandpass noise. First, broader auditory filters (poorer spectral resolution) should result in greater spread of masking of the band-limited noise into adjacent auditory filters, reducing the proportion of spectrotemporal regions of favorable local SNR. Second, the recognition of speech energy within the clean glimpses should be dependent to some extent upon the limits of spectral resolution within that region. Thus, it was predicted that CI listeners with better spectral resolution will be less adversely affected by a band-limited masker than will CI listeners with poorer spectral resolution.

Peters, Moore, and Baer (1998) showed that young NH listeners received approximately 15 dB of spectral masking release (better performance with a masker containing spectral gaps, versus that with a full-band masker) from four-ERB-wide (roughly 1 octave) spectral gaps in a broadband steady masker, whereas young HI listeners received only about 5 dB masking release from the same spectral glimpses. The authors presumed this difference to be related to differences in spectral resolution between the two groups.

We tested listeners' sentence recognition in quiet, broadband noise, and several band-limited noises. One advantage of this design is that each subject can act as their own control; performance in band-limited noise can be compared to performance in wide-band, spectrally-matched noise for each individual, thus controlling for individual differences in more central factors such as working memory, use of lexical context, and listening strategy. Speech recognition performance in quiet and in various band-limited noise conditions, as well as the within-subject differences in performance for each noise condition, were compared to broadband and octave-band spectral ripple discrimination thresholds to investigate whether those nonspeech measures of spectral resolution were predictive of speech recognition ability. Two different groups, CI users and normal-hearing subjects listening to vocoded simulations of the test materials, were included in this study.

II. EXPERIMENT 4.I: SENTENCE RECOGNITION IN BAND-LIMITED NOISE

A. Subjects

CI listeners.

14 of the original 15 CI subjects involved in the previous studies, plus one additional individual, participated in this experiment for a total of 15. One member of the original subject pool was excluded from participation because she could not perform near 50% correct on sentence recognition in quiet. Data from another enrolled subject was eliminated for this same reason. Performance > 50% correct in quiet was required, since the desired outcome measure was SNR for 50% correct in background noise (sometimes referred to as the speech reception threshold or SRT, which represents the lowest sound intensity level or the SNR at which 50% of words or sentences are repeated correctly). Table 4-I shows individual subject characteristics.

Table 4.1. Summary of subject characteristics. The subject identifiers C, D, and N denote Clarion I, Clarion II, and Nucleus users, respectively.

Subject Code	M/F	Age (yrs)	CI use (yrs)	Etiology	Duration of deafness (yrs)	Device	Speech Processing Strategy
C03	F	58.8	9.7	Familial Progressive SNHL	27	Clarion I	CIS
C05	M	52.5	10.2	Unknown	<1	Clarion I	CIS
C16	F	54.2	6.7	Progressive SNHL	13	Clarion I	MPS
C18	M	74.0	7.2	Otosclerosis	33	Clarion I	MPS
C23	F	48.1	6.4	Progressive SNHL; Mondini's	27	Clarion I	CIS
D02	F	58.2	6.4	Unknown	1	Clarion II	HiRes -P
D05	F	78.2	6.6	Unknown	3	Clarion II	HiRes -S
D08	F	55.9	5.0	Otosclerosis	13	Clarion II	HiRes-S
D10	F	53.8	5.2	Unknown	8	Clarion II	HiRes-S
D19	F	48.2	3.5	Unknown	7	Clarion II	HiRes-S
D24	M	60.7	3.3	Unknown, Progressive	27	Clarion II	HiRes 120
N13	M	69.9	17.5	Hereditary; Progressive SNHL	4	Nucleus 22	SPEAK
N14	M	63.5	13.9	Progressive SNHL	1	Nucleus 22	SPEAK
N28	M	68.8	11.8	Meningitis	<1	Nucleus 22	SPEAK
N32	M	40.1	10.3	Maternal Rubella	<1	Nucleus 22	SPEAK
N34	F	62.0	8.4	Mumps; Progressive SNHL	9	Nucleus 22	SPEAK

Normal-hearing simulation listeners.

10 normal-hearing participants (two males and eight females, ranging in age from 20 to 50 years of age) were recruited for this study. All of them passed pure tone audiometric screening at 15 dB HL for audiometric frequencies 250-2000 Hz, and at 20 dB HL at 4000 and 6000 Hz.

B. Stimuli

Spectral ripple discrimination

The stimuli and procedure for obtaining spectral ripple discrimination thresholds were the same as described in Chapter 2. Broadband ripple discrimination thresholds were obtained for each CI subject during one of the test sessions in which sentence recognition was measured.

Speech Materials

Sentence recognition testing was performed using the AzBio sentence corpus (Spahr and Dorman, 2004), spoken by two male and two female talkers. The AzBio sentences were created specifically for use with the CI population. They were developed using 5-channel vocoded simulations with normal-hearing listeners for validation and equivalency testing of lists. The corpus includes 15 lists of 20 sentences; each sentence contains 4-11 words. Rather than keywords, all words in each sentence are scored. The AzBio sentences are considered medium-context materials; they contain more contextual cues than do IEEE sentences, but fewer than the easier HINT materials. The AzBio sentences are produced in a natural, conversational fashion as opposed to the “clear speech” mode characterizing HINT materials (Spahr and Dorman, 2004).

Prior to testing with AzBio sentences, the HINT sentences (Nilsson, Soli, and Sullivan, 1994) were used in an adaptive procedure to determine an approximate SNR corresponding to 50% correct recognition. Each HINT list contains 2-7 keywords in high-context sentences, spoken by one male talker.

Masking noises.

Speech-shaped noise was generated to match the long-term average spectrum of the AzBio sentences. To accomplish this, one list of 20 sentences was concatenated, with silent intervals removed. From this signal, an average spectrum was calculated. A 30-sec interval of Gaussian noise was generated and filtered to match the spectrum of the speech materials. The noise was then bandpass-filtered into a 350-5600 Hz band, which was matched in overall rms amplitude to the average rms amplitude of the speech materials. This noise served as the "Full-band" masking condition, and was the only noise condition for which SNR reflected a true ratio of rms amplitude. Figure 4.1 illustrates the spectra of the speech-shaped noise (left panel) and the Full-band noise (right panel).

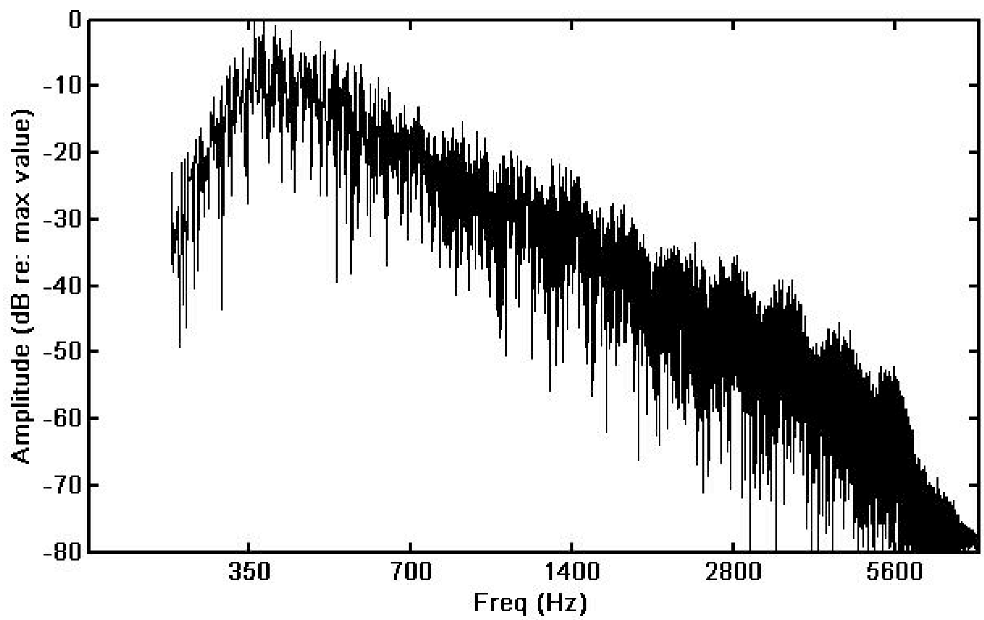


Figure 4.1. Spectrum of wide-band Gaussian noise filtered to match the long-term average spectrum of AzBio sentences, then bandpass-filtered to 350-5600 Hz for the Full-band masker condition. Log frequency is on the abscissa, and amplitude (relative to maximum value of the peak) is on the ordinate.

The full-band masking noise was further bandpass-filtered into four different sub-bands: 350-700 Hz, 350-1400 Hz, 1400-5600 Hz, and 2800-5600 Hz, producing a total of five different band-passed noises, each 30 s in duration. The spectrum level for each of these one- or two-octave bands was left unaltered from the original Full-band noise; in other words, the band-passed noises were not adjusted further after filtering to equate overall rms to the Full-band noise or the speech stimuli. Because of the roll-off in higher frequencies of the speech spectrum, the highest-frequency band, 2800-5600 Hz, contained substantially less energy than the Full-band masker. Consequently, the SNR is always referenced to the rms of the Full-band condition, and therefore refers to the SNR of the speech and noise before further octave-band filtering. Figure 4.2 illustrates the amplitude spectra for the four band-passed noises.

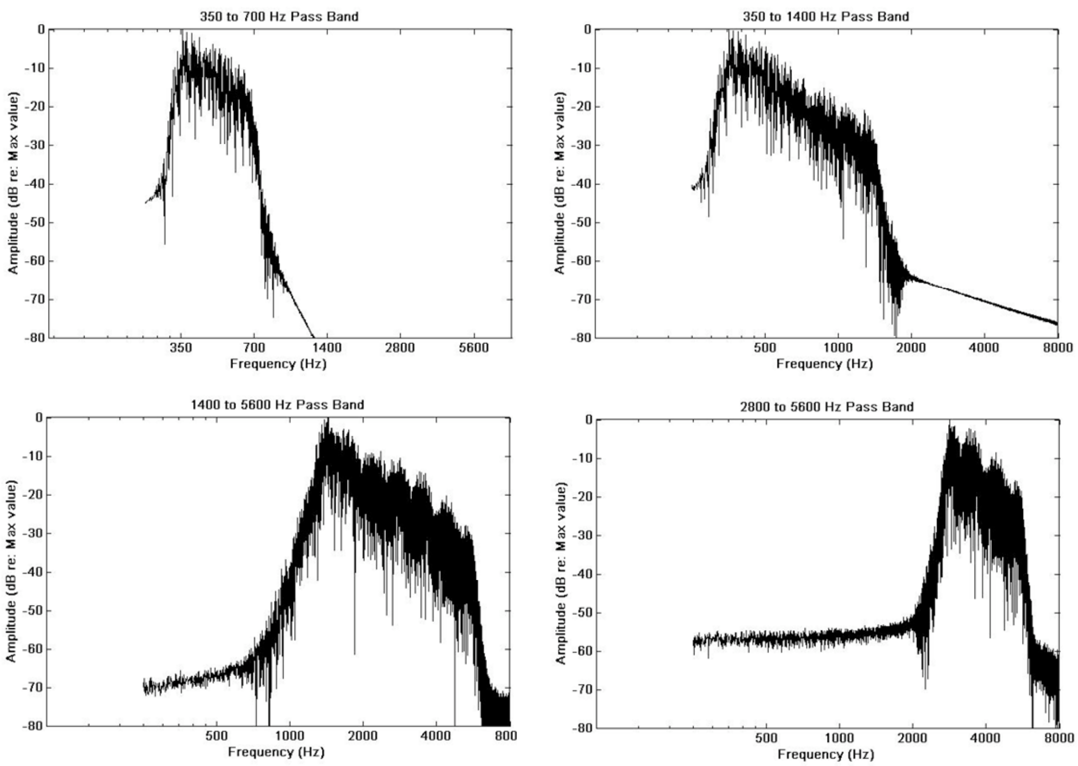


Figure 4.2. Spectra of band-passed noises, with log frequency on the abscissa and relative amplitude on the ordinate (re: maximum peak value). Left upper panel: Full-band noise filtered to 350-700 Hz (Low-1-oct). Right upper panel: full-band noise filtered to 350-1400 Hz (Low-2-oct). Left lower panel: full-band noise filtered to 1400-5600 Hz (High-2-oct). Right lower panel: full-band noise filtered to 2800-5600 Hz (High-1-oct).

Each of the sentences was individually mixed with a randomly selected segment from the 30-s samples of each of the five pre-generated noises to produce nominal SNRs ranging from +21 to -21 dB, in 3 dB steps. The speech level was held constant, and the noise level was adjusted to produce the desired SNR.

Vocoded simulations.

After the sentences were mixed with each of the different noises at each of the designated nominal SNRs, copies of the sentence-plus-noise stimuli were vocoded, or processed to simulate a typical CI processing scheme, CIS (Continuous Interleaved Sampling). A computer application called TigerCIS (Fu, 2005) was used to batch-process the stimuli. Specifically, the acoustic signal for each sentence was divided into eight frequency bands or channels, based on the Greenwood function (Greenwood, 1990); filter slopes of the analysis filters were 24 dB/oct. The temporal envelope (amplitude fluctuations as a function of time) for each of these analysis bands was extracted, and these same patterns of fluctuations were used to modulate eight carrier bands of spectrally-flat noise corresponding in frequency bandwidth to each of the analysis bands. The eight carrier bands were then combined and presented acoustically to the normal-hearing listeners.

Two different carrier filter slope conditions were used, 12 dB/octave and 24 dB/octave, to simulate two different, contrasting degrees of spectral resolution. 24 dB/octave slopes are relatively steep, and result in less spectral smearing than 12 dB/octave slopes. Therefore, 24 dB/octave vocoded speech models relatively "better" spectral resolution, and 12 dB/octave simulates "poorer" spectral resolution. Five of the subjects listened to the 24 dB/octave-processed sentences, and five listened to the 12 dB/octave-processed sentences.

Other studies using vocoder-processed speech (e.g. Bigabr, Espinoza-Varas, and Loizou, 2008; Fu and Nogaki, 2004) have demonstrated that decreasing the slopes of the vocoder carrier filters results in significant decrements in speech recognition performance. Litvak *et al*. (2007) reported that varying the carrier filter slopes for vocoder-processed rippled noise resulted in higher SMTs in normal-hearing simulation listeners.

C. Procedure

Speech recognition testing was conducted in a sound-treated booth, with sentence materials presented at a level of 65 dBA in the sound field, the same presentation level as used for the previous speech measures (Chapter 2). Most subjects required two separate test sessions of approximately two hours each to complete the testing. Subjects used their own speech processors at typical use settings; as in the previous speech recognition experiment, processor programs that were optimized for noise reduction were not used. E-Prime software was used to deliver the sentence stimuli, which were presented through a single speaker located at zero degrees azimuth, one meter from the subjects' seated position. For the adaptive HINT pretest procedure, two lists per condition were used. For the AzBio materials, one list was presented in quiet, and one list was presented at each of three SNRs for each noise condition, individually selected to produce a relatively broad range of performance. The number of words scored for each *noise x SNR* condition ranged from 133 to 154.

The subjects' task was to listen to each sentence, then repeat it verbally after presentation. They were encouraged to guess, if unsure. The experimenter sat outside the booth to control sentence/noise presentation for the HINT procedure, and in the booth with the subjects for the AzBio testing, and manually recorded the number of correct words repeated for both tests.

Target SNRs were individually estimated for each subject, for each masker condition, based on their sentence thresholds obtained using an adaptive procedure with HINT sentences. For the HINT procedure, the presentation level for the sentences was held constant and the noise level varied adaptively with a 1-up, 1-down rule, converging on 50% correct. Two randomly selected HINT lists were concatenated for each experimental condition, for a total of 20 sentences. Speech was presented at 70 dB SPL, and the first trial began with noise at an SNR that fell below the listener's speech reception threshold (SRT). A 2-dB step size was used to increase the SNR (i.e., decrease the noise level) until the listener correctly repeated all words in the original sentence. Subsequently, a new sentence was presented on each trial, with noise level adjusted according to the up-down rule. The run proceeded in this way until all 20 sentences were presented. Threshold was determined as the average of the final 6 reversals. The resulting HINT threshold for each masker condition was then used as a guide for setting an initial SNR for the AzBio experiment. The HINT procedure also served as a practice block of trials for subjects, prior to the AzBio testing.

In most cases, the HINT threshold (in dB SNR) was used as the first SNR condition for the AzBio sentences. Depending on the percent correct score for that initial AzBio condition, the next SNR condition was selected to be at a higher or lower SNR, in order to try and obtain performance levels on either side of 50% correct. Increments/decrements in SNR of 6 dB were used, unless it was predicted that this level would result in ceiling or floor performance, or would produce a score that wouldn't fall outside the confidence intervals for the preceding score; in those cases, SNRs that were 3 (or, rarely, 9) dB from the initial SNR level were selected.

The order of masker conditions was randomized for each subject, with the exception that the first condition completed was always Quiet.

D. Results

P-I functions were constructed as follows: percent correct at three SNRs was obtained, and linear, least-square regression lines fitted to the points. Logistic functions, such as those used in Experiment 2.3 (Chapter 2) were not used, because of the limited number of data points, the limited validation data available on these materials, and lack of knowledge about the actual shape of functions for these materials with these band-limited maskers.

In some cases, only two points were used: if scores at two adjacent SNR levels differed by less than the confidence intervals determined for the AzBio materials, which are based on Thornton and Raffin's (1978) binomial distribution model (Spahr, personal communication), then the more extreme of the two was eliminated. However, if two adjacent points differed by less than the confidence interval, but the two points straddled the 50% point, then both were included, so that at least one data point on either side of 50% on the function were included. Grossly non-monotonic P-I functions were excluded from analysis. From the fitted lines for each function, the SNR corresponding to 50% correct ($SNR_{50\%}$) was interpolated and used as a dependent variable in the data analysis.

Percent correct in quiet was transformed to rationalized arcsine units (rau) scores (Studebaker, 1985) for analysis. Correlations were examined between rau scores in quiet, $SNR_{50\%}$ obtained in each of the masker conditions, and spectral ripple discrimination measures across subjects. In addition, comparisons of within-subject differences between $SNR_{50\%}$ for the full-band masker condition and the band-limited masker were examined as a function of spectral ripple discrimination.

CI listeners

Sample of P-I function. Figure 4.3, below, illustrates a sample set of P-I functions, or percent correct performance as a function of nominal SNR, with masker type as the parameter. In this example, the functions are spread out along the SNR dimension in a progression of masker conditions (right-to-left) of full-band masker, low-2-octave masker, low-1-octave masker, high-2-octave masker, and high-1-octave masker.

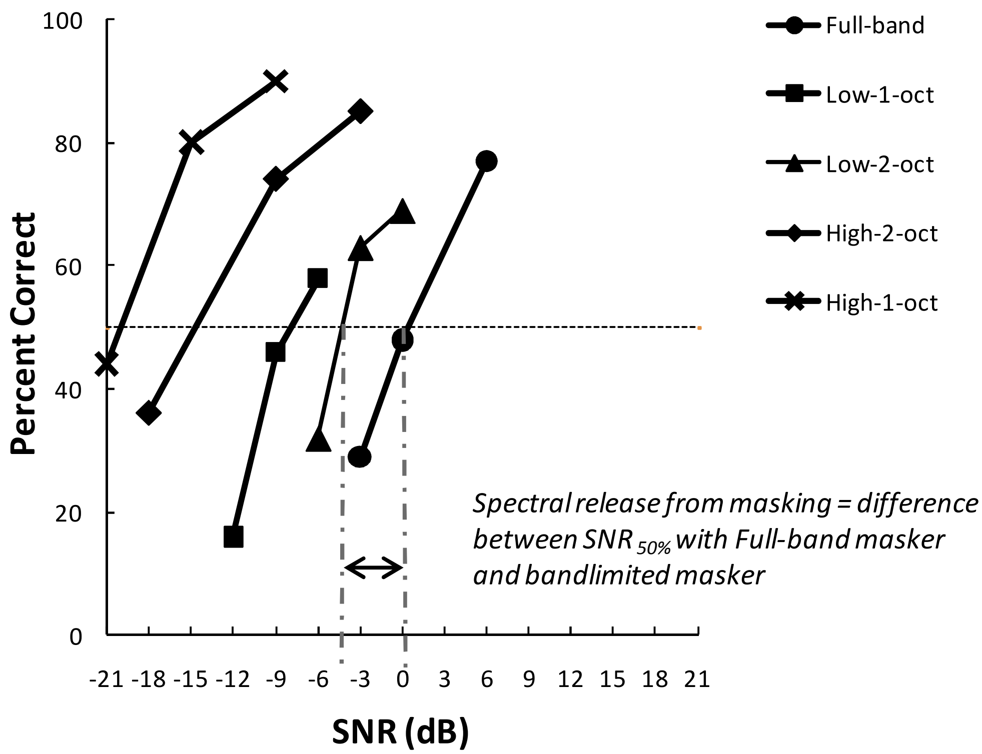


Figure 4.3. P-I functions for one subject. On the abscissa is SNR; the ordinate indicates percent correct. Masker condition is the parameter. Thus, each function is based on performance with one of the five masker conditions at 3 different SNRs. Each point represents one list of AzBio sentences. The solid arrow between the broken lines indicates the difference in $SNR_{50\%}$ for full-band and low-2-oct masker, or *spectral release from masking* for the low-2-oct condition.

This progression is intuitively consistent with masking effectiveness being a function of both the bandwidth and the actual rms of the masker, based on the region of the speech-shaped spectrum to which it is matched.

Individual P-I functions. Individual results, including SNR50% for each masker condition and corresponding spectral release from masking calculations, are listed in Table 4.2. Individual P-I functions for each subject are displayed in Figure 4.4, 4.5, and

4.6, divided into groups by implant device. As can be seen, there are substantial individual differences in relative positions of functions for the different maskers and distance between them, as well as for the slopes of the functions. In some cases, P-I functions are non-monotonic.

Table 4.2. Results for individual subjects, including performance in quiet, $SNR_{50\%}$ for each of the different masker conditions, and spectral release from masking ($SNR_{50\%}$ for Full-band condition minus $SNR_{50\%}$ in each of the band-limited masker conditions).

Subject	*% Correct*	*rau*	*$SNR_{50\%}$*					*Spectral release*			
Code	***(Q)***	***(Q)***	***Full-band***	***Low-2-oct***	***Low-1-oct***	***High-2-oct***	***High-1-oct***	***Low-1-oct***	***Low-2-oct***	***High-2-oct***	***High-1-oct***
C03	99	112.31	0.75	-4.26	-7.57	-14.68	-20.00	8.32	5.01	15.43	20.75
C05	93	97.58	0.67	-0.57	-2.78	-15.57		3.44	1.24	16.24	
C16	85	85.72	1.55	5.97	3.00	-9.00		-1.45	-4.43	10.55	
C18	50	50.00	25.89	16.41	15.79				9.48		
D02	88	89.75	6.27	4.40	1.14	-16.53		5.13	1.87	22.80	
D05			7.55	4.22	2.29	0.40	-4.87	5.26	3.33	7.15	12.42
D08	83	83.22	9.63	12.17	8.43	-1.53	-14.62	1.20	-2.54	11.16	24.26
D10	97	105.98	-3.79	-10.18	-19.10	-13.29		15.31	6.39	9.50	
D19	98	108.79	-0.13	-5.32	-10.33	-18.95		10.21	5.19	18.82	
D26	71	70.00	12.38	8.25	7.00	-10.09	-15.87	5.38	4.13	22.47	28.25
N13	63	62.14	16.20	17.17	16.64		-2.77	-0.44	-0.97	16.20	18.97
N14	65	64.06	5.22	5.81	4.50	-10.68		0.72	-0.59	15.90	
N28	64	63.10	19.00	18.08	12.33	-1.36		6.67	0.92	20.36	
N32	81	80.83	3.15	0.90	-5.29	-12.21		8.44	2.25	15.36	
N34	67	66.01	20.10	15.19	15.55	0.86	-9.00	4.55	4.91	19.24	29.10

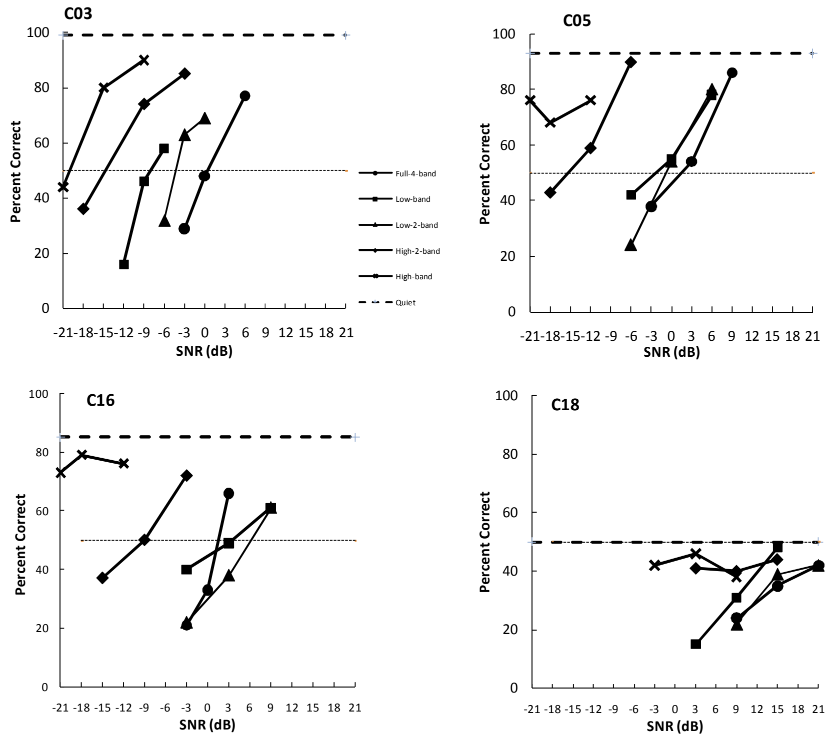


Figure 4.4. Series of P-I functions for four Clarion I subjects. SNR is represented on the abscissa; the ordinate indicates percent correct. Masker condition is the parameter. Each point represents one list of AzBio sentences. The dashed line along the top indicates performance in Quiet.

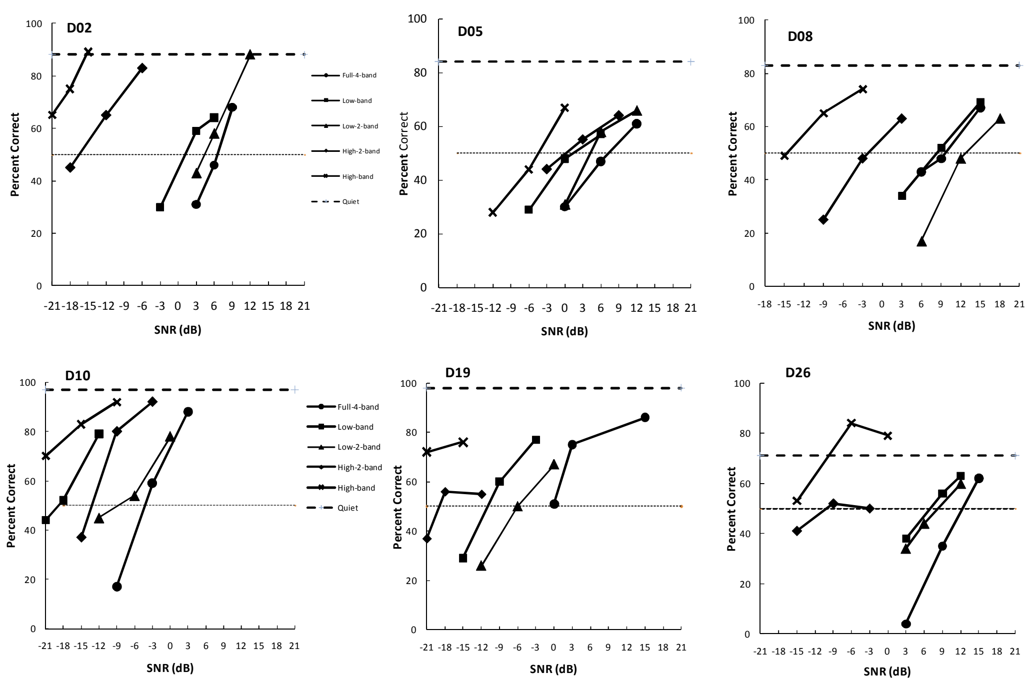


Figure 4.5. Series of P-I functions for six Clarion II subjects. SNR is represented on the abscissa; the ordinate indicates percent correct. Masker condition is the parameter. Each point represents one list of AzBio sentences. The dashed line along the top indicates performance in Quiet.

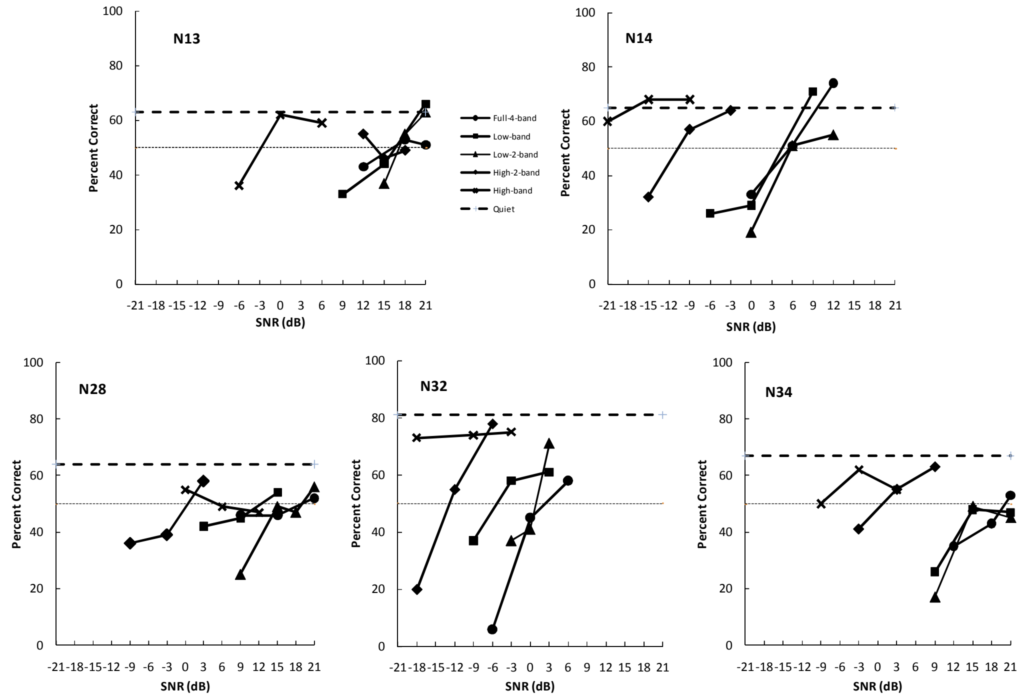


Figure 4.6. Series of P-I functions for five Nucleus-22 subjects. SNR is represented on the abscissa; the ordinate indicates percent correct. Masker condition is the parameter. Each point represents one list of AzBio sentences. The dashed line along the top indicates performance in Quiet.

Comparisons of AzBio measures with Broadband Spectral Ripple Discrimination.

Lines were fitted to the functions, as described above, defining an interpolated $SNR_{50\%}$. Regression analyses were performed on the data, comparing rau scores in quiet and $SNR_{50\%}$ in each noise condition with spectral ripple discrimination thresholds. Figure 4.7, Panel A, shows sentence recognition in quiet for 14 subjects as a function of spectral ripple discrimination. (One subject's data had to be excluded from analysis, due to a procedural error during testing.) A moderate, significant relationship was found between rau scores and broadband ripple discrimination when a logarithmic fit was used (r^2 =

0.40, p = 0.015). Panel B displays the strong relationship between performance in Full-band noise and in quiet ($r^2 = 0.75$, $p < 0.0001$).

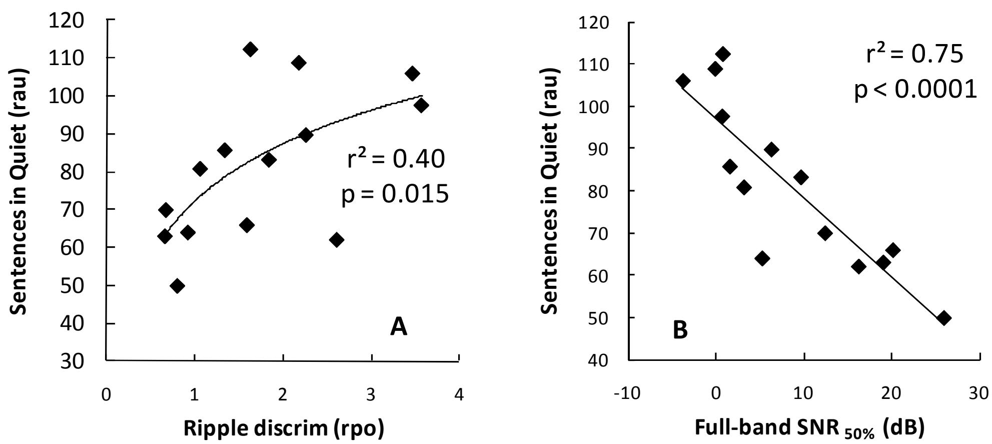


Figure 4.7. Panel A: sentence recognition in broadband noise (rau scores) as a function of spectral ripple discrimination thresholds. Panel B: sentence recognition in quiet as a function of $SNR_{50\%}$. Data from 14 subjects are included.

When sentence recognition in noise was compared to broadband ripple discrimination thresholds, the relationships failed to reach significance. Figure 4.8 shows the results. Regression analysis using a best-fit logarithmic function showed a trend between ripple discrimination and $SNR_{50\%}$ with the full-band masker (r^2=0.25, p=0.06), which weakened with each of the band-limited masker conditions (Low-2-octave masker: r^2 =0.22, p=0.08; low-1-octave masker: r^2=0.22, p=0.08; high-2-octave masker: r^2=0.26, p=0.08. All are linear fits.) Data for the high-1-octave masker are not shown, as $SNR_{50\%}$ measures for that masker condition could be defined for only a few subjects.

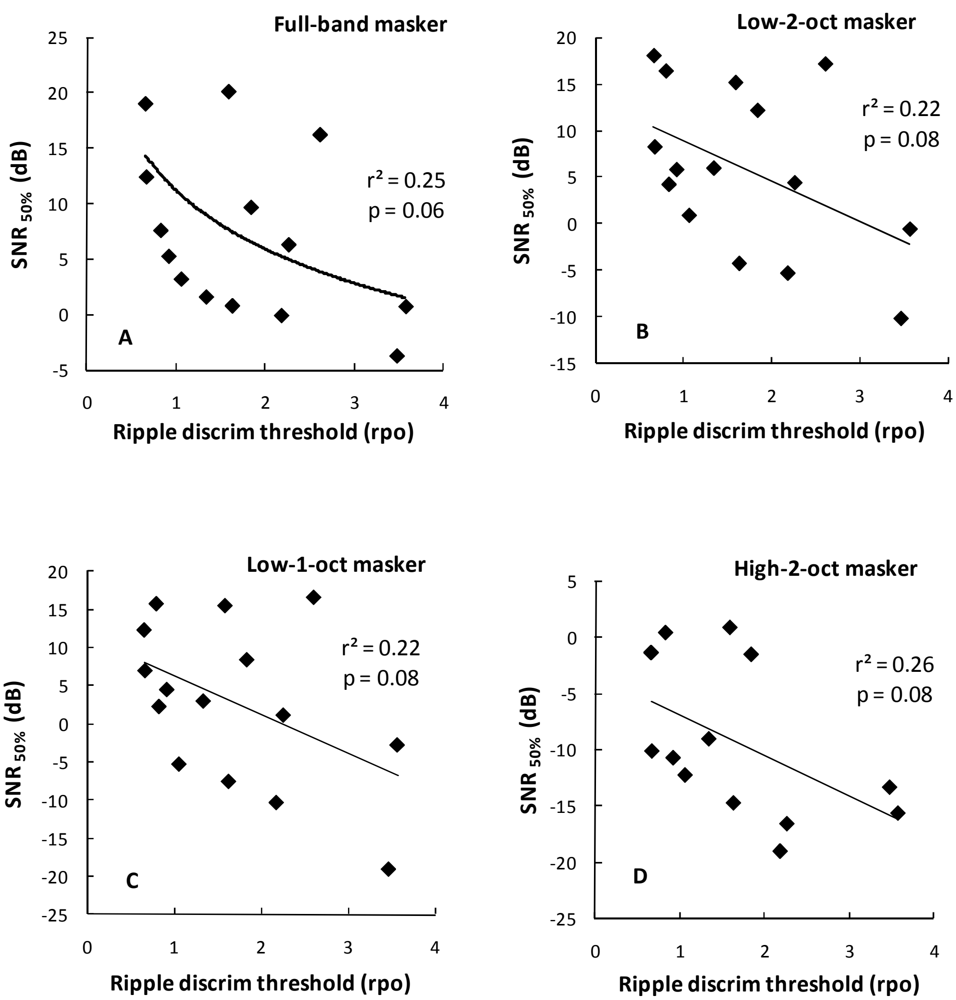


Figure 4.8. Sentence recognition in band-limited noise ($SNR_{50\%}$) as a function of broadband spectral ripple discrimination thresholds. Panel A: Full-band masker. Panel B: Low-2-oct masker. Panel C: Low-1-oct masker. Panel D: High-2-oct masker.

The relationships between spectral release from masking and broadband spectral ripple discrimination are shown in Figure 4.9. Displayed are the differences in performance between the Full-band masker condition and the low-1-octave masker condition (Panel A), the low-2-octave masker condition (Panel B), and the high-2-octave masker condition (Panel C), as a function of spectral ripple discrimination thresholds. None of the relationships are significant (all p>0.05).

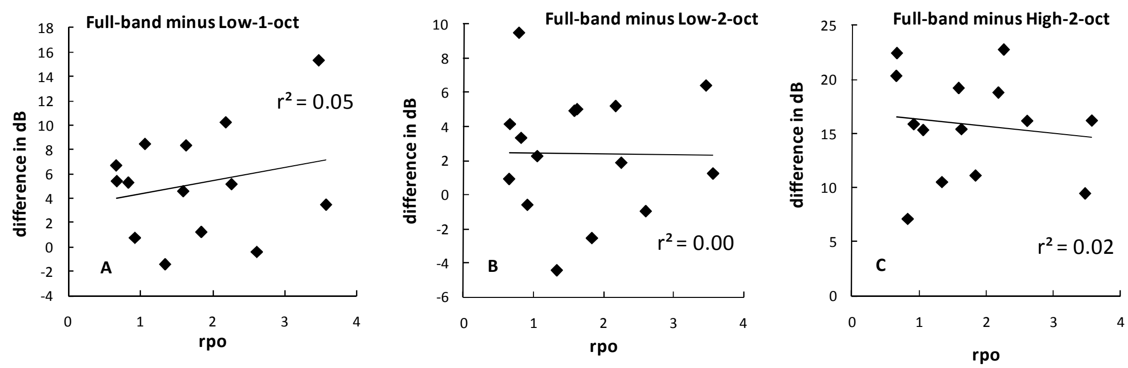


Figure 4.9. Spectral release from masking as a function of broadband spectral ripple discrimination thresholds. Panel A: Full-band $SNR_{50\%}$ minus Low-1-oct $SNR_{50\%}$. Panel B: Full-band $SNR_{50\%}$ minus Low-2-oct $SNR_{50\%}$. Panel C: Full-band $SNR_{50\%}$ minus High-2-oct $SNR_{50\%}$. The abscissa displays ripple discrimination threshold (in rpo), and the ordinate displays the difference (in dB) between the speech measures. Data from 14 subjects are included.

Repeat analyses, excluding an outlier. One of the subjects' performance produced data points that were consistently at odds with the majority of the group. Specifically, that subject demonstrated good spectral ripple resolution but performed relatively poorly on the sentence recognition tasks. The same regression analyses were repeated, as above, excluding that subject's data. Figure 4.10 displays the relationships between spectral ripple discrimination and sentence recognition in quiet and with the various maskers. The data points that are not included in the regression analyses are plotted as different symbols. Logarithmic (best-fit) regression analyses show a strong trend toward a correlation, nearly reaching significance with a Bonferroni-corrected $\alpha = 0.015$.

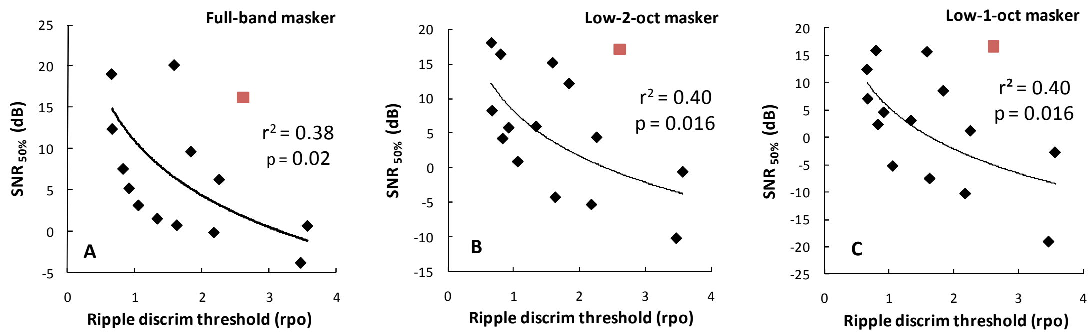


Figure 4.10. Comparisons of ripple discrimination thresholds and $SNR_{50\%}$ for Full-band (Panel A), Low-2-oct masker (Panel B), and Low-1-oct masker (Panel C), with data points for Subject N13 excluded from the regression analyses, but included on the plots as square symbols.

Comparisons of AzBio measures with Octave-Band Spectral Ripple Discrimination.

Broadband spectral ripple discrimination did not show a correlation with sentence recognition in noise, for either the full-band or the band-limited masker conditions. It is possible that a more spatially-focused measure of spectral resolution, using a signal that is more closely matched to the frequency region of the speech signal that is not overlapped by the noise masker, might bear a stronger relationship. Octave-band spectral ripple discrimination thresholds (see Chapter 2) were compared to $SNR_{50\%}$ for each of the masker conditions, in the same way as the broadband ripple thresholds, above. (Plots for each of these comparisons are displayed in Appendix, Figure A1.)

The next several figures focus on the relationships that might be predicted to correlate: i.e., ripple discrimination for low-frequency, octave-band stimuli (350-700 Hz, 700-1400 Hz) with $SNR_{50\%}$ in a high-frequency masker condition (Figure 4.11), and ripple discrimination for high-frequency, octave-band stimuli (1400-2800 Hz, 2800-5600 Hz) with $SNR_{50\%}$ in low-frequency masker conditions (Figure 4.12, panels A-D). Figure 4.11 shows $SNR_{50\%}$ for the High-2-oct masker condition as a function of spectral ripple discrimination thresholds for an octave-band rippled noise stimulus of 350-700 Hz (Panel A), and for 700-1400 Hz (Panel B). With best-fit regression analyses, each function approached but did not reach significance (p=0.09 and p=0.05, respectively) when using a Bonferroni-corrected α=0.025 for multiple comparisons.

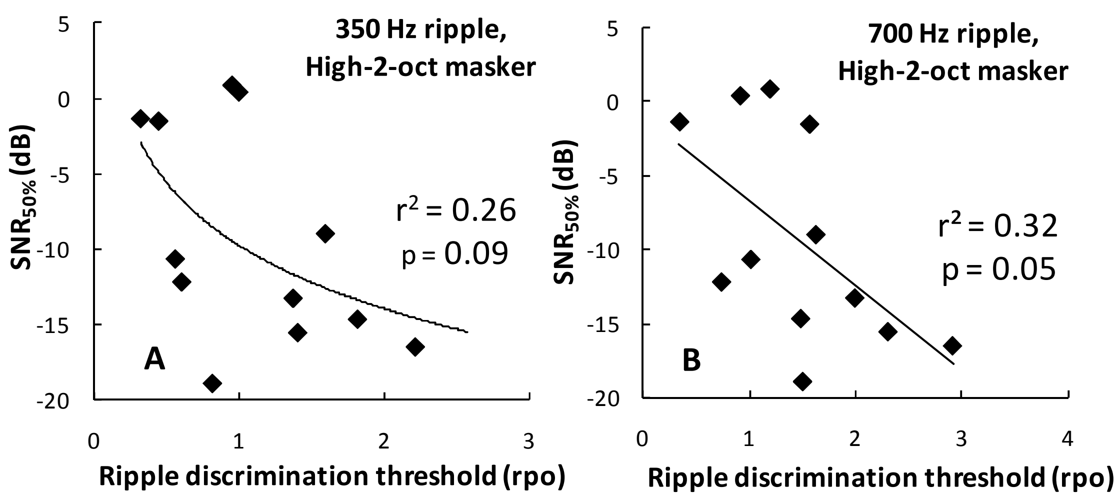


Figure 4.11. $SNR_{50\%}$ with the High-2-oct masker as a function of ripple discrimination for a 350-700 Hz (Panel A) and 700-1400 Hz (Panel B) rippled noise stimulus. Data from 12 subjects are included; one subject did not have octave-band ripple discrimination measures, and two had undefined $SNR_{50\%}$ for the high-2-oct masker conditions.

Figure 4.12 displays $SNR_{50\%}$ for sentences in low-frequency maskers (Low-2-oct, Panels A and C; Low-1-oct, Panels B and D) as a function of ripple discrimination for high-frequency rippled noises (1400-2800 Hz, Panels A and B; 2800-5600 Hz, Panels C and D). None of the best-fit regression analyses approach significance (all p>0.05).

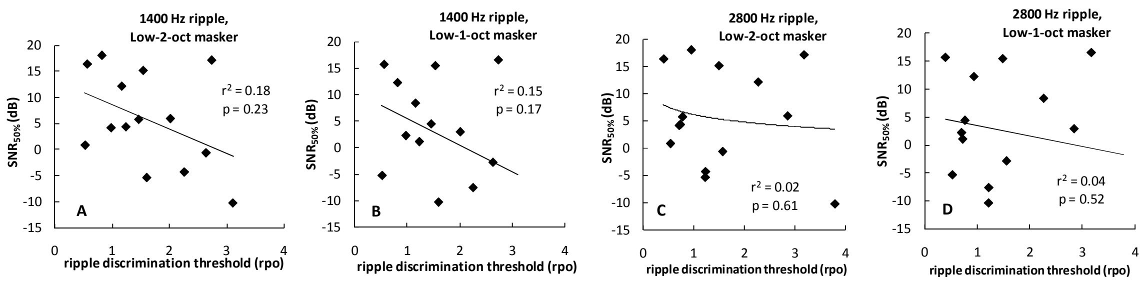


Figure 4.12. $SNR_{50\%}$ for the Low-2-oct masker (Panel A) and Low-1-oct masker (Panel B) as a function of ripple discrimination threshold for a 1400-2800 Hz rippled noise, and for the Low-2-oct masker (Panel C) and Low-1-oct masker (Panel D) as a function of ripple discrimination threshold for a 2800-5600 Hz rippled noise.

Removing data from Subject N13, as above, resulted in stronger relationships between octave-band ripple discrimination and sentence recognition in band-limited noise. Only one regression analysis is significant at a Bonferroni-corrected α=0.0125: ripple discrimination for a 1400-2800 Hz rippled noise and $SNR_{50\%}$ for the Low-2-octave condition. The relationship is shown in Figure 4.12, with the excluded point displayed as a different shape.

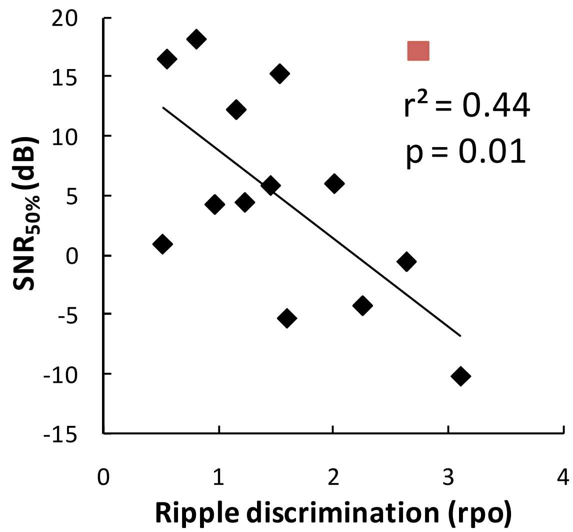


Figure 4.13. Comparisons of ripple discrimination thresholds for 1400-2800 Hz rippled noise and $SNR_{50\%}$ for the Low-2-oct masker, with data point for Subject N13 excluded from the regression line, but included on the plot as a square symbol.

Normal-hearing simulation listeners

P-I functions. P-I functions for normal-hearing simulation (NH-sim) subjects were constructed in the same way as for CI subjects: percent correct at three SNRs was obtained, and linear, least-square regression lines fitted to the points. The same rules were applied for determining which points to include in the line fits.

Figure 4.14 shows P-I functions for 5 subjects listening to 24 dB/oct simulations. Quiet performance is near 100% for all subjects. In all cases, performance with the High-1-oct masker did not degrade sufficiently at the most challenging SNR (-21 dB) to produce 50% performance. No non-monotonic functions were produced.

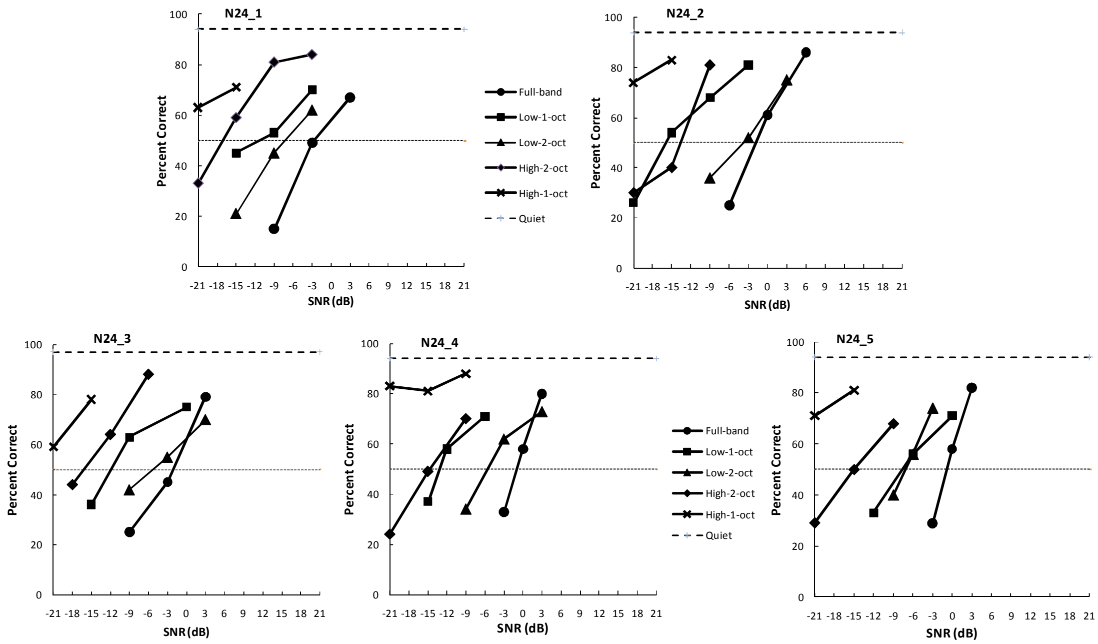


Figure 4.14. Series of P-I functions for five normal-hearing simulation listeners (24 dB/oct filter slopes). SNR is represented on the abscissa; the ordinate indicates percent correct. Masker condition is the parameter. Each point represents one list of AzBio sentences. The dashed line along the top indicates performance in Quiet.

Figure 4.15 contains the P-I functions for the 5 normal-hearing subjects listening to 12 dB/oct simulations. Quiet performance fell into a range from 60-80% for four of the listeners; one subject performed near 50% in quiet, but this was likely a spurious result since that listener's best performance in noise was near 80%, commensurate with other subjects' performance in quiet. Four of the five subjects' functions for the High-1-oct masker condition bisected the 50% correct line, suggesting greater influence of the high-frequency masker on their performance, compared to the 24 dB/oct group.

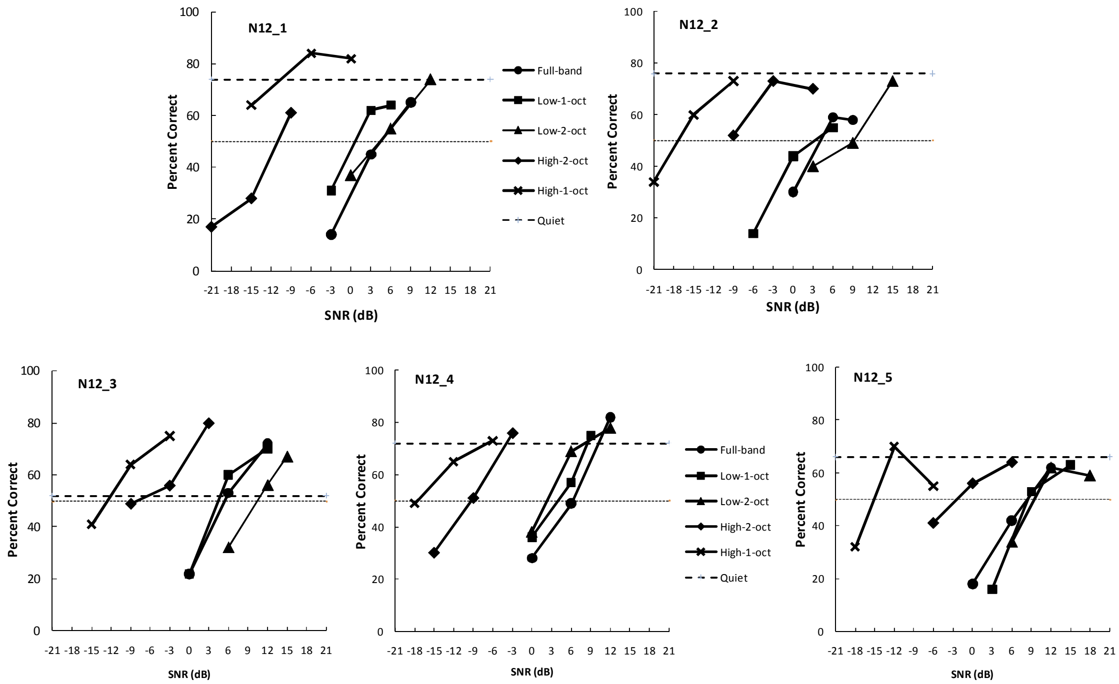


Figure 4.15. Series of P-I functions for five normal-hearing simulation listeners (12 dB/oct filter slopes).. SNR is represented on the abscissa; the ordinate indicates percent correct. Masker condition is the parameter. Each point represents one list of AzBio sentences. The dashed line along the top indicates performance in Quiet.

Group data: 24 vs. 12 dB/octave.

Analysis of the NH-sim results was performed on group data. Mean $SNR_{50\%}$ for subjects in each group (12 dB/oct and 24 dB/oct) were calculated for each masker condition. Figure 4.16 shows $SNR_{50\%}$ for each different masker condition with filter slope as the parameter. A repeated measures ANOVA was performed on the data, with filter slope (12 vs. 24 dB/oct) and masker condition as main factors. Results indicated a significant between-subjects main effect for filter slope [$F(1,8)=97.519$, $p<0.001$], a significant within-subjects main effect for masker condition ($F[1.43,11.46]=52.936$, $p<0.001$) and significant interaction between masker and filter slope ($F=8.517$, $p=0.009$).

Within-subject simple contrasts for masker condition indicated significant differences between Full-band and Low-2-oct maskers ($F=70.967$, $p<0.001$), and between Full-band and High-2-oct maskers ($F=753.687$, $p<0.001$), but not between Full-band and Low-1-oct ($F=0.084$, $p=0.780$). There was not a significant masker x filter slope interaction between Full-band and High-2-oct maskers ($F=0.057$, $p=0.817$). The interaction was significant for contrasts between Full-band and Low-1-oct ($F=9.099$, $p=0.017$), and between Full-band and Low-2-oct conditions ($F=36.144$, $p<0.001$).

In addition, paired sample T-tests were used to examine the differences in performance in the different masker conditions, for each group separately. For the 12 dB/oct group, there were no significant differences between Full-band and Low-1-oct or Low-2-oct masker conditions ($t=-1.699$, $p = 0.165$, and $t=2.27$, $p=0.086$, respectively) but the difference between Full-band and High-2-oct conditions reached significance ($t=16.348$, $p<0.001$). For the 24 dB/oct group, all differences between Full-band and band-limited conditions were significant ($t=5.88$, $p=0.004$; $t=8.52$, $p=0.001$; $t=25.595$, $p<0.001$, for Low-1-oct, Low-2-oct, and High-2-oct conditions, respectively).

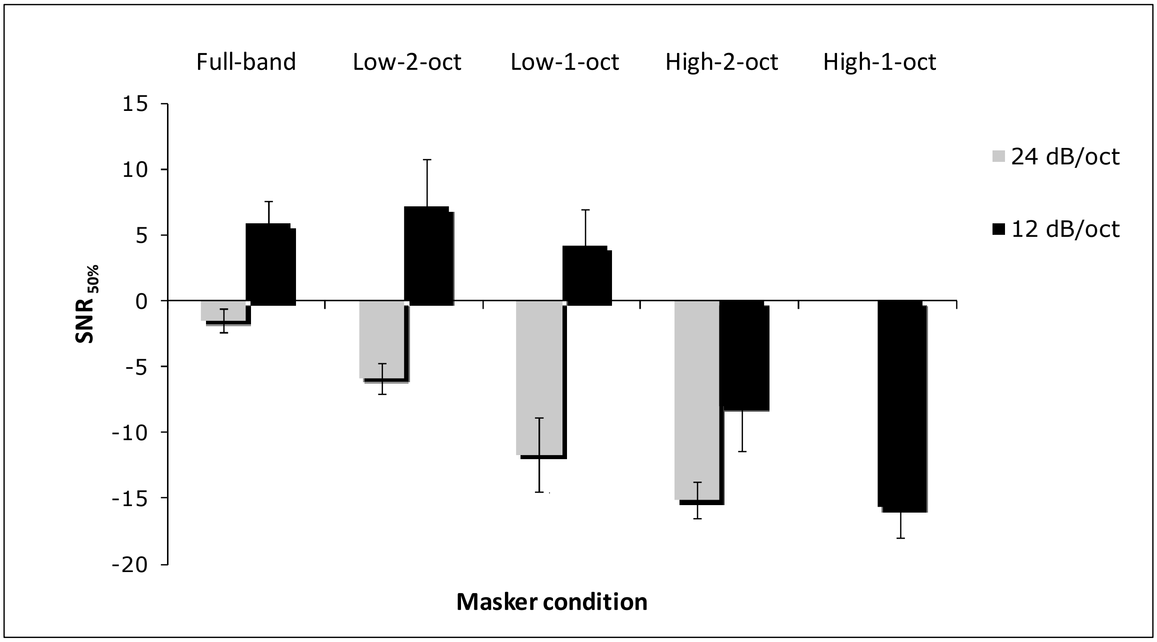


Figure 4.16. Mean $SNR_{50\%}$ for each masker condition in NH-sim subjects. Light-colored bars represent the 24 dB/oct listeners, and dark bars the 12 dB/oct group. Error bars represent one standard deviation.

CI data as "better" vs "poorer" groups

In light of the NH-sim data, CI performance was reexamined as group data, selecting the top and bottom thirds as groups representing differences in spectral resolution (analogous to the 12- versus 24 dB/oct divisions for the simulations). Specifically, the top five CI listeners, in terms of spectral ripple discrimination performance, were selected for the "better" ripple discrimination group, and the bottom five CI listeners were selected for the "poorer" ripple discrimination group. For each group, mean $SNR_{50\%}$ for each noise masker condition was calculated; results are displayed in Table 4-III and Figure 4.17.

Table 4.3 Percent correct in quiet and $SNR_{50\%}$ for different masker conditions for "better" and "poorer" subjects. Upper panel is top tertile of subjects, in terms of spectral ripple discrimination thresholds. Lower panel is bottom tertile.

Subj	% C (Q)	Full-band	Low-2-oct	Low-1-oct	High-2-oct	High-1-oct
D08	83	9.63	12.17	8.43	-1.53	-14.62
D19	98	-0.13	-5.32	-10.33	-18.95	
D02	88	6.27	4.40	1.14	-16.53	
D10	97	-3.79	-10.18	-18.75	-13.19	
C05	93	0.67	-0.57	-2.78	-15.57	
mean	*91.80*	*2.53*	*0.10*	*-4.46*	*-13.15*	
sd	*6.30*	*5.36*	*8.66*	*10.48*	*6.82*	
Subj	% C (Q)	Full-band	Low-2-oct	Low-1-oct	High-2-oct	High-1-oct
N28	64	19.00	18.68	12.33	-1.36	
D24	71	12.38	8.25	7.00	-10.09	-15.87
C18	50	25.89	16.41	15.79		
D05		7.55	5.43	2.29	0.40	-4.87
N14	65	5.22	5.81	4.50	-10.68	
mean	*62.50*	*14.01*	*10.92*	*8.38*	*-5.43*	
sd	*8.89*	*8.48*	*6.20*	*5.58*	*5.77*	

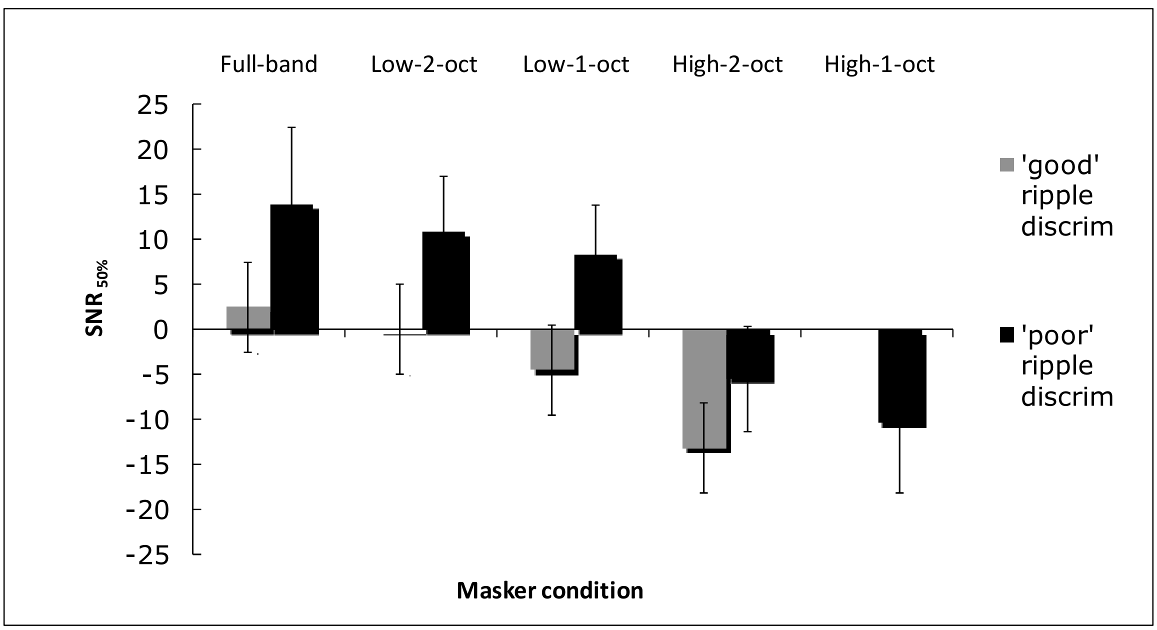


Figure 4.17. Mean $SNR_{50\%}$ for each masker condition in stratified CI subjects. Light-colored bars represent the 24 dB/oct listeners, and dark bars the 12 dB/oct group. Error bars represent one standard deviation.

While the variance is much greater for the CI group than for the NH listeners, the general pattern of results appears similar. An ANOVA was performed on the complete CI data set, with spectral resolution group (three divisions, determined by top, middle, and bottom thirds in ripple discrimination performance) and masker condition as main factors. Results indicated a significant within-subjects main effect for masker condition [Greenhouse-Geisser corrected $F(1.737,17.373)=50.733$, $p<0.001$] but the between-subjects effect of spectral resolution group was non-significant [$F(2,10)=1.893$, $p=0.201$], and there was no significant interaction between masker and spectral resolution group ($F=0.208$, $p=0.912$). Within-subject contrasts showed significant differences between Full-band and Low-1-oct ($F=18.239$, $p=0.002$) and Full-band and High-2-oct conditions ($F=108.131$, $p<0.001$), but not for Full-band and Low-2-oct ($F=4.288$, $p=0.065$). None of the contrasts for the interaction of masker and spectral resolution group were significant (all $p>0.05$).

An examination of the membership of the "better" vs. "poorer" groups suggests that membership in the former is dominated by users of the C II device. However, when an ANOVA was conducted with masker condition and CI device as factors, no significant between-subjects main effect for device was found [$F(9,2)=0.458$, $p=0.646$]. Univariate ANOVA with CI device and ripple discrimination threshold was also performed; again, no main effect for device type was seen [$F(2,12)=0.41$, $p=0.678$].

Paired sample T-tests were used to examine the differences in performance for the different masker conditions within each separate group. For the "better" spectral resolution group, there were no significant differences between $SNR_{50\%}$ for Full-band and

Low-2-oct masker conditions ($t=1.541$, $p=0.198$), but significant differences between Full-band and Low-1-oct conditions ($t=2.779$, $p = 0.05$) and between Full-band and High-2-oct conditions ($t=6.425$, $p=0.003$). For the "poorer" group, a similar pattern was found, with no difference between Full-band and Low-2-oct conditions ($t=1.831$, $p=0.141$) but significant differences between Full-band and both Low-1-oct and High-2-oct conditions ($t=3.712$, $p=0.021$; $t=4.851$, $p=0.0017$, respectively).

A final analysis was undertaken, combining CI and NH-sim data using a mixed-design, repeated measures ANOVA. The within-subjects factor was masker condition, and the dependent variable was $SNR_{50\%.}$ One between-subjects factor was spectral resolution ("good" vs. "poor"). For group assignment, CI subjects were stratified into two levels, by top and bottom halves of ordered spectral ripple discrimination performance. NH-sim subjects were divided into two groups, except based on vocoder filter slopes (with 24 dB/oct selected as "good", and 12 dB/oct as "poor"). A second between-subjects factor included was "CI status" (distinguishing CI from NH-sim). Simple contrast analysis was done, permitting pair-wise comparisons of performance in each of the masker conditions. Results revealed a significant main effect of masker condition [Greenhouse-Geisser corrected $F(1.8, 39.5)=95.37$, $p<0.001$] and significant interactions between masker condition and CI status, and between masker and spectral resolution level ($F=9.92$, $p<0.001$ and $F=6.50$, $p=0.003$, respectively). The three-way interaction was not significant. The between-subjects effect of CI status did not reach significance [$F(1,19)= 1.77$, $p=0.199$], but the main effect of spectral resolution level was significant [$F(1,19)=24.34$, $p<0.001$]; the interaction between the two was not significant.

Within-subject contrast analysis revealed significant effect of masker condition between performance in Full-band and each of the band-limited maskers ($F=7.359$, $p=0.014$; $F=44.864$, $p<0.001$; $F=92.309$, $p<0.001$ respectively). The interactions between masker condition and CI status, and between masker and spectral resolution level were significant for Full-band versus Low-1-oct and Low-2-oct contrasts, but not for Full-band versus High-2-oct. (masker x CI status: $F=10.165$, $p=0.005$; $F=10.259$, $p=0.005$; $F=1.342$, $p=0.261$; masker x spectral resolution level: $F=17.391$, $p=0.001$; $F=10.641$, $p=0.004$; $F=0.016$, $p=0.899$ respectively). The three-way contrast was significant only for the Full-band and Low-2-oct masker contrast ($F=8.385$, $p=0.009$).

In spite of the lack of a strong effect of spectral resolution level within the CI subject groups in the first ANOVA analysis, compared with the NH-sim and the combined group analyses, the pattern of results raised the possibility that spectral resolution is more closely related to speech recognition in CI listeners on the "better" end of the continuum of spectral resolution, but less so for CI listeners on the "poorer" end. Spectral release from masking (difference between $SNR_{50\%}$ in Full-band and band-limited conditions) was compared to $SNR_{50\%}$ for the Full-band masker condition for the "better" and "poorer" CI groups. Results are displayed in Figure 4.18. While the sample sizes are small, there is a clear difference in the pattern of results for the two groups: the subjects who performed better (i.e. lower $SNR_{50\%}$) in the Full-band condition showed a significant correlation between that measure and spectral release from masking in both the Low-2-oct (Panel A) and Low-1-oct (Panel B) conditions, while the subjects who performed more poorly (i.e. higher $SNR_{50\%}$) showed no systematic relationship with spectral release from masking. However, the correlation for the "better group" disappears for the High-2-oct masker condition (Panel C).

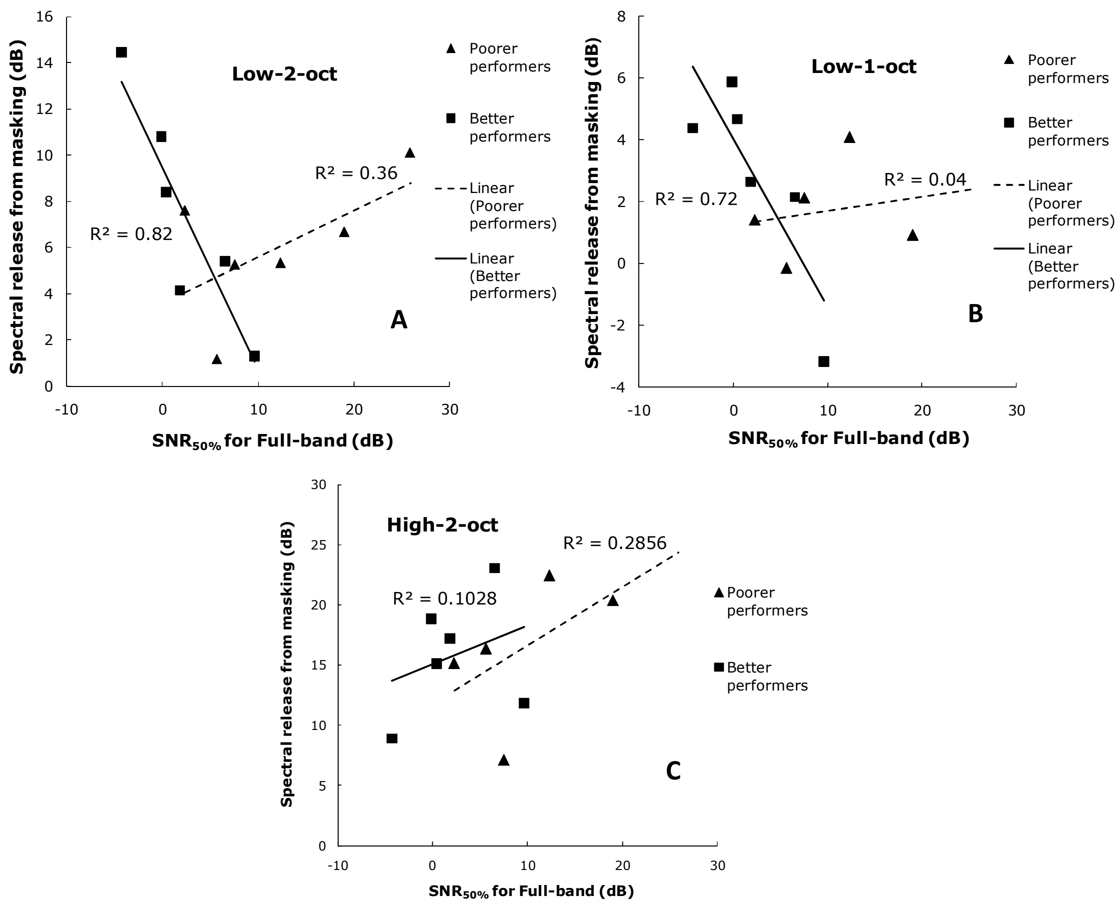


Figure 4.18. Spectral release from masking for Low-2-oct (Panel A), Low-1-oct (Panel B), and High-2-oct (Panel C) masker conditions as a function of $SNR_{50\%}$ for Full-band condition.

E. Discussion

This study attempted to investigate the relationship between spectral resolution and speech perception by examining spectral release from masking when using band-limited maskers. The results suggest that such a relationship may exist for better, but not poorer CI users.

A significant correlation was shown between spectral ripple discrimination and sentence recognition in Quiet, but the relationships weakened when speech recognition was measured in the presence of wide-band and band-limited, speech-shaped background noise. This result is fairly consistent with the data reported in Chapter 2, with similar comparisons of IEEE sentence recognition and spectral ripple discrimination.

Removing one subject's data improved the relationships. This particular subject, N13, demonstrated comparatively good ripple discrimination, but performance on speech recognition measures, particularly in noise, was relatively poor. Figure 4.19 displays Subject N13's P-I functions, most of which are shallow-sloped and/or non-monotonic. It cannot be ruled out that other (possibly central) factors contribute to this individual's performance deficits, but it is beyond the scope of this study to investigate this further.

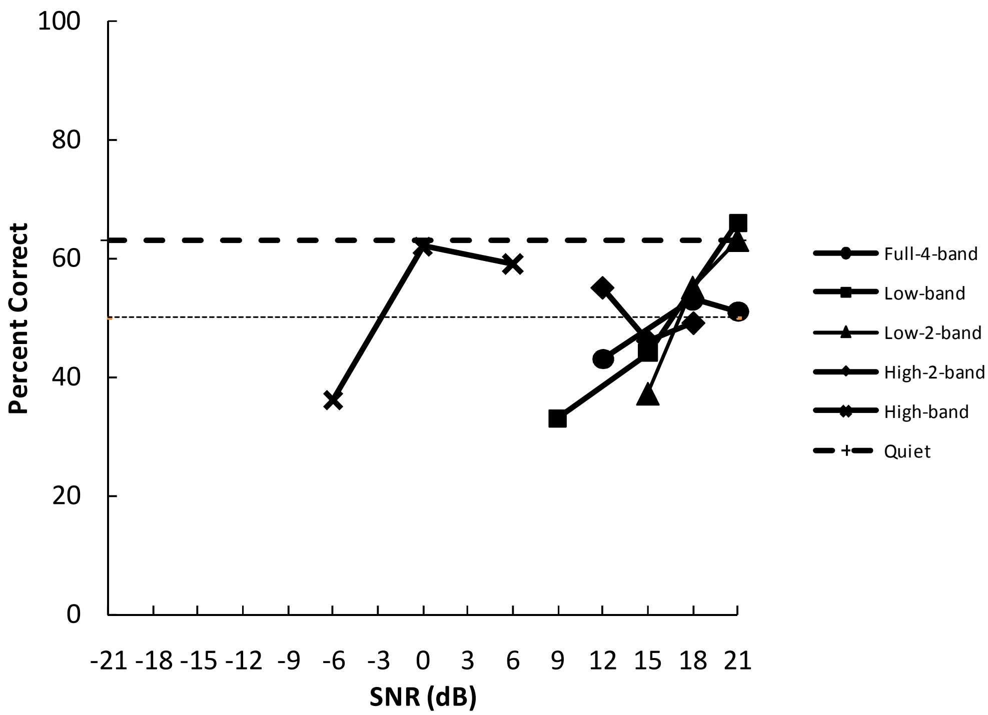


Figure 4.19. P-I functions for Subject N13.

Previous collection of spectral ripple discrimination data for octave-band stimuli allowed for systematic comparisons of localized spectral resolution with masker conditions from either similar or different spectral regions. In particular, it was predicted that spectral resolution localized to a low-frequency region might correlate with sentence recognition performance with a high-frequency band-limited masker, and spectral resolution within a high-frequency region would correlate with sentence recognition performance with a low-frequency band-limited masker. (This prediction assumes roughly equivalent speech information on either side of the cut-off frequency defining low-frequency and high-frequency band-limited maskers used in this study, or 1400 Hz. Since importance function information is not available for these materials with CI listeners, this may not be a fair assumption.) The results only partially supported this hypothesis, with one significant correlation between octave-band ripple discrimination for a rippled noise of 1400-2800 Hz and performance with the Low-2-oct masker (350-1400 Hz), for 13 subjects.

Normal-hearing simulation data revealed some interesting results, suggesting that spectral resolution is a factor not only in determining the overall level of performance in a given masker condition, but also in mediating the spectral release from masking with the two lower-frequency band-limited maskers (Low-1-oct and Low-2-oct). That is, the 24 dB/oct group showed greater spectral release from masking with the low-frequency band-limited maskers than did the 12 dB/oct group.

When NH-sim data and stratified CI data were combined in a three-way, repeated-measures ANOVA, no significant effect of group (CI vs. NH) was shown. This finding lends validity to the technique of modeling the effects of limited spectral resolution by varying the steepness of carrier filter slopes in vocoded simulations.

In addition, the pooled CI/NH-sim analysis revealed a significant interaction between spectral resolution and masker, suggestive of significant differences in spectral release from masking between "better" and "poorer" groups. This contention is further supported by the comparisons shown in Figure 4.18, which suggest that for CI users with better spectral ripple resolution, spectral release from masking covaries with the listener's "starting point" for sentence recognition in noise, the $SNR_{50\%}$ with Full-band noise.

Overall, the data from the complete set of CI subjects participating in this study failed to match the original prediction, in that measures of spectral release from masking, by themselves, did not exhibit significant relationships with spectral ripple discrimination. This finding lends support to the conclusion that spectral resolution may not, by itself, account for a significant amount of the variance in speech perception performance in noise, across a group of CI users exhibiting a wide range of performance.

CHAPTER 5: GENERAL DISCUSSION

CIs, by their nature, significantly limit the amount of spectral detail that can be transmitted to a listener. Peripheral neural representation almost certainly limits spectral representation, as well. Efforts are being made to improve the spectral resolution possible with a CI, for instance through the creation of virtual channels (e.g. Bierer, 2007). But will the provision of this additional spectral detail result in improvements in speech recognition? The series of studies described in this dissertation addressed basic questions regarding measures of spectral resolution currently in use, and their relationships to sentence and vowel recognition in quiet and in broadband and band-limited noise. In particular, some of this work allows for some comparisons between performance of better vs. poorer CI users.

Chapter 2 reviewed some experiments comparing spectral ripple discrimination with the spatial tuning curve (STC), a generally accepted index of spectral resolution, and assessed how well each of those measures correlated with speech recognition. The data revealed a strong correlation between STCs and spectral ripple discrimination when using octave-wide rippled noise stimuli, matched to the CF of the probe electrode for the STC procedure. This finding supports the conclusion that the spectral ripple discrimination task can be regarded as a valid measure of spectral resolution, but suggests that a one-point measure of ripple discrimination may not be completely informative.

When using fixed octave-band rippled noise stimuli that varied in center frequency, individual spectral ripple discrimination was shown to vary unsystematically with the region of the electrode array being stimulated, suggesting the potential for local differences in spectral resolution along the array. Importantly, this finding implies that widely-used broadband spectral ripple discrimination measures may not reveal local areas of poorer resolution.

Broadband spectral ripple discrimination correlated with speech measures in quiet, but not in noise, in contrast to other similar studies (e.g. Henry and Turner, 2003; Henry *et al.*, 2005; Won *et al.*, 2007; Won, Drennan, Kang, and Rubinstein, 2010). An interpretation of this finding might be that spectral resolution does not, by itself, account for a significant amount of the variance in speech perception performance in noise. Alternately, it is possible that an individual's performance on the broadband spectral ripple discrimination task may not be impaired by a region of relatively poorer spectral resolution, whereas their speech recognition ability might suffer from the loss or degradation of information coded by that region of poorer resolution, particularly in background noise. In this case, a relationship between broadband spectral ripple discrimination and speech recognition in noise may be weak, particularly if the region of poorer resolution falls in a spectral region particularly important for speech understanding.

Chapter 3 reported on data obtained on CI listeners for a spectral ripple detection task. Performance varied widely across subjects, with a similar range of thresholds and pattern of results across subjects (different shaped SMTFs) as reported by Saoji *et al.* (2009). Spectral ripple detection thresholds at low (0.5-1.0 rpo) ripple frequencies represent the listener's ability to perceive broadly-spaced spectral peaks, and are not thought to reflect spectral resolution. Consistent with other reports (e.g. Summers and

Leek, 1994; Litvak *et al*., 2007; Saoji *et al*., 2009), spectral ripple detection results from this study were found to correlate strongly with measures of speech perception in quiet, but not in noise. While the strongest correlations were for low (<1 rpo) ripple rates, SMTs for 1 and 3 rpo also showed trends with $SNR_{50\%}$ for vowels (correlations were not significant when Bonferroni corrections for multiple comparisons were applied to α). If detection of ripples at the upper extreme on the SMTF is influenced by spectral resolution, as suggested by Eddins and Bero (2007) and Saoji and Eddins, (2007), then demonstrating a relationship between ripple detection for higher ripple rates and speech perception might have been interpreted as implying a role for spectral resolution in speech perception in noise. The data from the current study do not support such a conclusion.

Further, when spectral ripple detection results (specifically, the ripple frequency corresponding to a spectral ripple detection threshold of 30 dB) were compared to ripple discrimination thresholds for individual listeners, the two measures correlated significantly, but were not equal to one another. This finding does not support the contention of Saoji *et al*. (2009) that spectral ripple discrimination thresholds are functionally equivalent to spectral modulation thresholds at the "higher" end of the SMTF (i.e. around 2-3 rpo). In fact, when trying to equate and compare performance between the two tasks, our data suggested that spectral ripple detection at higher modulation frequencies could involve the use of temporal cues. When spectral peaks become more closely spaced, they may interact to produce temporal artifacts (AM cues). In this group of CI subjects, detection of spectral ripples was achieved far beyond the theoretical predicted limit, raising the strong possibility that non-spectral cues were perceptible in the task. In fact, SAM detection thresholds were found to correlate with the ripple rates corresponding to ripple detection thresholds of 30 dB. It is not known at what ripple rate this spurious cue might appear; acoustic analysis of the stimuli has not been performed to verify the conjecture.

Even if it is not a “pure” measure of spectral resolution, spectral ripple detection does correlate strongly with speech recognition performance in quiet. A significant relationship between ripple detection at low ripple frequencies and sentence/vowel recognition was shown, supporting the interpretation of a reliance of CI users on broadly-spaced spectral maxima and minima in the spectral envelope, rather than spectral resolution per se, for phoneme perception (e.g. Saoji *et al*., 2009).

Some important questions regarding CI user’ speech perception in noise remain unanswered. Chapter 4 describes an experiment that attempted to address those questions by investigating the effects of band-limited noise on sentence recognition in CI users and normal-hearing simulation listeners, and whether spectral resolution is predictive of those effects. A significant relationship was found between spectral ripple discrimination and sentence recognition in quiet. Speech recognition in the different band-limited masker conditions showed a trend toward a relationship with spectral resolution. However, spectral ripple discrimination was not directly predictive of spectral release from masking (difference in performance between full-band and band-limited masker conditions), as was originally hypothesized. Results suggest that the effects of noise on speech recognition are not simply due to energetic masking, but might involve confusions of spectrotemporal "bits" of target and masker (e.g. Moore and Glasberg, 1988).

To explore this notion further, Figure 4.20 depicts the differences in performance when referenced to "nominal SNR" versus actual SNR, based on overall rms of the masker. The set of P-I functions in the left panel are from Subject C03, recopied from Figure 4.4. In the right panel, the values on the abscissa (SNR) are corrected for actual (as opposed to nominal) overall rms amplitude. The difference in rms amplitude between the High-1-oct masker and the Full-band masker is 31 dB, and for the High-2-oct masker is 20.5 dB. Differences for the two low-frequency maskers are negligible (<1 dB). Hence, the High-1-oct and High-2-oct masker P-I functions are shown shifted by 31 and 20.5 dB, respectively, while the two low-frequency maskers' P-I functions remain in the same relative position.

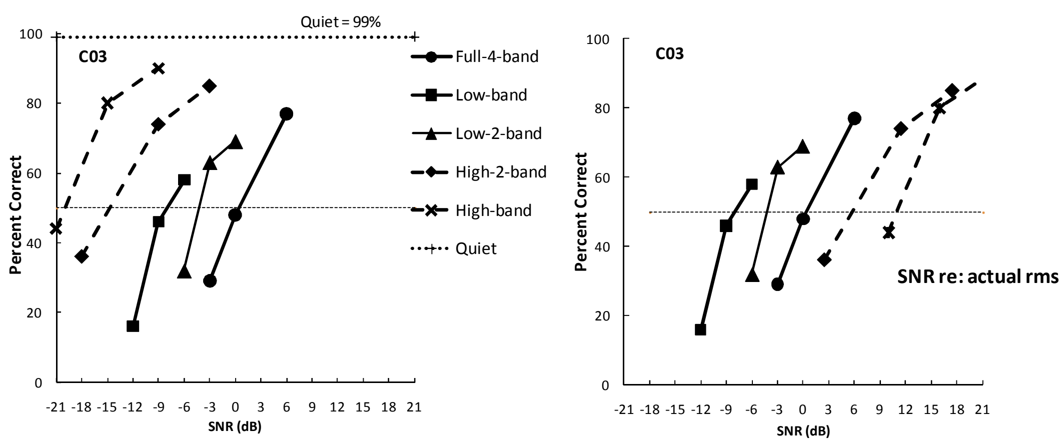


Figure 5.1 Left panel: P-I functions for Subject C03, for each of the different masker conditions. Dashed lines indicate functions for the two maskers from the high-frequency end of the spectrum. Right panel: SNR corrected for actual rms of maskers.

The shift in relative positions of P-I functions for the two high-frequency maskers is striking. If total rms was the dominating factor in determining effectiveness of a masker, then we might expect the functions to lie on top of one another. The separation between functions, as well as their relative positions, suggests that other mechanisms are involved.

It is inaccurate to say that actual rms is *not* a factor; in general, maskers with higher overall rms energy (low-frequency maskers) resulted in higher (poorer) $SNR_{50\%}$ measures. However, the spectral release from masking provided by a high-frequency band-limited masker was variable and non-uniform across subjects, suggesting that misidentification of words was not simply a function of energetic masking. In many instances, slopes of P-I functions for high-frequency masker conditions were shallower than functions for low-frequency maskers (suggestive of P-I functions in informational masking [e.g. Arbogast, Mason, and Kidd, 2002]), and non-monotonic P-I functions were obtained, suggesting the possibility of some type of non-energetic mechanism, possibly informational in nature, due to the "confusability" of components of target and masker signals (e.g. Moore and Glasberg, 1988). Difference in pattern of results for high-frequency compared to low-frequency maskers (e.g. Figure 4.18) suggests the possibility that different mechanisms for masking are at play for band-limited maskers from different regions of the spectrum.

Further, the data from the low-frequency maskers also suggest that actual rms is not the primary factor in determining masking effectiveness. For instance, the difference in rms between the Full-band and Low-1-octave maskers is very small, on the order of 1 dB. This is due to the fact that the bandwidth of the Low-1-octave band comprises the highest-amplitude part of the spectrum of the Full-band noise, and consequently contributes most to the overall rms amplitude of the Full-band noise. But in normal acoustic hearing, the bandwidth of a masker does influence masking effectiveness. For instance, a steady noise containing spectral gaps is not as detrimental to speech understanding as a noise without spectral gaps at the same overall rms amplitude (e.g. Peters, Moore, and Baer, 2006), suggesting that reducing the bandwidth of the masker provides some spectral release from masking.

In CI listeners and NH-sim listeners, masking effectiveness appears to be in part related to the region of the speech spectrum and the relative importance of the part of the acoustic signal that is "lost." Little improvement in performance was seen for either group when broadband noise was replaced by band-limited noise. In general, we might conclude from this that narrowband maskers are much more disruptive for CI and simulation listeners than they are for NH listeners.

In this study, vocoded simulation listeners showed distinct differences in spectral release from masking, related to spectral resolution; the group with steep carrier filter slopes (representing "better" spectral resolution) had significantly greater spectral release from masking than the group with shallow vocoder slopes (representing "poorer" spectral resolution). For the most part, the NH-sim data suggest that spectral resolution should be predictive of speech recognition in band-limited noise. In general, their pattern of results was broadly mirrored by the CI data when divided into groups of "better" and "poorer" CI users, but the relationships between spectral resolution and speech recognition do not hold as well for the CI group. However, there are notable outliers in nearly every data set, with some CI users performing well on speech recognition measures in quiet and noise, while performing poorly on measures of spectral resolution. It is possible that some CI users simply make better use of non-spectral cues for speech recognition.

The "better" CI group showed a strong relationship between performance in the Full-band condition and spectral release from masking (difference between the Full-band and band-limited maskers) for the low-frequency, but not high-frequency, maskers. In other words, the better their performance in the Full-band condition, the greater the spectral release from masking (for the low-frequency band-limited maskers) they showed. The "poorer" group showed no such relationship with any of the band-limited maskers.

Why might these measures of spectral resolution be limited in their predictions of speech recognition in noise? One possibility is that the speech recognition materials commonly used in CI research may not be sensitive to differences in spectral resolution. For instance, Drennan, Won, Nie, Jameyson and Rubinstein (2010) showed within-subject improvements in broadband spectral ripple discrimination with a current-steering strategy, but commensurate gains in word recognition were not found. Relationships between spectral resolution and speech recognition may be obscured by complex interactions between individual differences in spectral resolution across the array, the ambiguity surrounding the mechanism for masking with band-limited noises, and the effects of importance weighting for CI-processed speech as a function of frequency. With the limited number of subjects in this data set, it is difficult to tease apart these factors. In addition, other factors such as the possible presence of non-relevant modulations

(Noordhoek and Drullman, 1996; Dubbelboer and Houtgast, 2008) produced by speech-noise interactions might be involved.

Is sentence recognition the best outcome measure for examining the effects of spectral resolution on speech perception? Sentences are certainly the most ecologically valid speech materials. Knowledge of what predicts (or improves) sentence recognition in CI listeners could help drive improvements in speech processing algorithms. Sentence materials that allow for finer-grained phonemic error analysis might be more informative. Incorporating the use of questionnaires of self-assessed satisfaction and benefit might also be informative.

As noted previously, current clinically-efficient methods of measuring spectral resolution (e.g. spectral ripple tasks) may fail to identify local regions of poorer resolution. If such regions exist in a CI user, would it be beneficial to customize the speech processor MAP to reroute frequency information, bypassing those electrodes?

Efforts to bypass "dead regions" in the cochlea through the use of frequency-transposing or frequency-compression hearing aids have demonstrated varying degrees of success (e.g. McDermott, Dorkos, Dean, and Ching, 1999; Parent, Chmiel, and Jerger, 1997). Some potential problems with this method applied to CIs include the issues of having to compress a greater frequency range onto individual electrodes, and the inevitable resulting mismatch between spectral information and place of stimulation within the cochlea (Moore, 2001). Baskent and Shannon (2004) showed that frequency compression/expansion in CI users resulted in decrements in speech perception. However, listeners appear to be able to adapt to such shifts, given enough time and experience (e.g. Rosen, Faulkner, and Wilkinson, 1999). In fact, it can be argued that any post-lingually deafened CI user already starts with a severely frequency-shifted system, relative to normal auditory tonotopy, since the electrode array can only be inserted partway into the cochlea. The excellent performance of experienced, better CI users demonstrates that adaptation is indeed possible.

Even as performance in CI users improves with advancements in CI technology, there will likely continue to be considerable variability in performance across users, and it will still be important to know whether increasing the amount of spectral information available to an individual CI user will result in improved performance in noise. Simply increasing the number of channels may not succeed for all patients. Expanding the size of the subject pool will be important for further work in this area, especially including subjects with the newest devices. Performance in maskers containing spectral gaps would be a logical next step in the investigation of spectral release from masking. Future work could also explore the effects of band-limited noises on detection of non-speech signals, such as pure tones, in CI listeners.

In summary, the effect of noise on speech perception in CI listeners is not straightforward. Better CI users appear to show a stronger relationship between spectral resolution and speech in noise, while poorer users' limited ability to use spectral cues may restrict the potential benefit they might receive from increasing the amount of spectral information provided by CIs. Knowing which of these categories best characterizes an individual CI user might help to determine whether advanced CI processing strategies designed to maximize the number of spectral channels could be beneficial to that individual.

REFERENCES

Anderson, E.S., Nelson, D.A., Kreft, H., Nelson, P.B., and Oxenham, A.J. (**2011**). Comparing spatial tuning curves and spectral ripple resolution in cochlear implant users. J. Acoust. Soc. Am. 130, 364-75.

Apoux, F. and Healy, E.W. (**2011**). Use of random spectral gaps to estimate band-importance functions. Presented at the American Auditory Society Scientific and Technology Meeting, Scottsdale, AZ, March, 2011.

Arbogast, T.L., Mason, C.R., and Kidd, G. (**2002**). The effect of spatial separation on informational and energetic masking of speech. J. Acoust. Soc. Am. 112, 2086-98.

Bacon, S.P. and Brandt, J.F. (**1982**). Auditory processing of vowels by normal-hearing and hearing –impaired listeners. J. Sp. Hear. Res. 25, 339-47.

Bacon, S.P. and Grantham, D.W. (**1989**). Modulation masking: effects of modulation frequency, depth, and phase. J. Acoust. Soc. Am. **85**, 2575-80.

Baskent, D. and Shannon, R.V. (**2004**). Frequency-place compression and expansion in cochlear implant listeners. J. Acoust. Soc. Am. 116, 3130-40.

Berenstein, C.K., Mens, L.H., Mulder, J.J.S. and Vanpoucke, F.J. (**2008**). Current steering and current focusing in cochlear implants: comparison of monopolar, tripolar, and virtual channel electrode configurations. Ear Hear. **29**, 250-260.

Bierer, J. A. (**2007**). Threshold and channel interaction in cochlear implant users: evaluation of the tripolar electrode configuration. J. Acoust. Soc. Am., 121, 1642-1653.

Bingabr, M., Espinoza-Varas, B., and Loizou, P.C. (**2008**). Simulating the effect of spread of excitation in cochlear implants. Hear Res 241:73-79.

Boex, C., Kos, M.-I. and Pelizzone, M. (**2003**). Forward masking in different cochlear implant systems. J. Acoust. Soc. Am. **114**, 2058-2065.

Chatterjee, M. (**2003**). Modulation masking in cochlear implant listeners: envelope versus tonotopic components. J. Acoust. Soc. Am. **113**, 2042-53.

Chatterjee, M., and Oba, S.I. (**2004**). Across- and within-channel envelope interactions in cochlear implant listeners. J. Assoc. Res. Otolaryngol. **5**, 360-75.

Clark, G. *Cochlear implants: fundamentals and applications*. (**2003**) Springer-Verlag, New York.

Cohen, L. T., Richardson, L. M., Saunders, E. and Cowan, R. S. C. (**2003**). Spatial spread of neural excitation in cochlear implant recipients: comparison of improved ECAP method and psychophysical forward masking. Hear. Res. **179**, 72-87.

Cooke, M. (**2006**). A glimpsing model of speech perception in noise. J. Acoust. Soc. Am. **119**, 1562-73.

DePaolis, R.A., Janota, C.P., and Frank, T. (**1996**). Frequency importance functions for words, sentences, and continuous discourse. J. Speech Hear. Res. 39, 714-23.

Dorman, M. J., Loizou, P. C. and Rainey, D. (**1997**). Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs. J. Acoust. Soc. Am. **102**, 2403-2411.

Dorman, M. F., Loizou, P. C., Fitzke, J., and Tu, Z. (**1998**). The recognition of sentences in noise by normal-hearing listeners using simulations of cochlear-implant signal processors with 6--20 channels. J. Acoust. Soc. Am. **104**, 3583-3585.

Drennan, W.R., Won, J.H., Nie, K., Jameyson, E., and Rubinstein, J.T. (2010). Sensitivity of psychophysical measures to signal processor modifications in cochlear implant users. Hear. Res. **262**, 1-8.

Dubbelboer, F. and Houtgast, T. (**2008**). The concept of signal-to-noise ratio in the modulation domain and speech intelligibility. J. Acoust. Soc. Am. 124, 3937-45.

Eddins, D. A., and Bero, E. M. (**2007**). "Spectral modulation detection as a function of modulation frequency, carrier bandwidth, and carrier frequency region," J. Acoust. Soc. Am. **121**, 363-372.

Friesen, L., Shannon, R. V., Baskent, D. and Wang, X. (**2001**). Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants. J. Acoust. Soc. Am. **110**, 1150-1163.

Fu, Q.J. (**2002**). Temporal processing and speech recognition in cochlear implant users. Neurorep. 16, 1635-39.

Fu, Q.J. (**2005**). Noise susceptibility of cochlear implant users: the role of spectral resolution and smearing. J. Assoc. Res. Otolaryngol. 6, 19-27.

Fu, Q.J., and Nogaki , G. (**2005**). Noise susceptibility of cochlear implant users: the role of spectral resolution and smearing. J. Assoc. Res. Otolaryngol. **6**, 19-27.

Fu, Q.J., Shannon, R.V., and Wang, X. (**1998**). Effects of noise and spectral resolution on vowel and consonant recognition: acoustic and electric hearing. J. Acoust. Soc. Am. **104**, 3586-96.

Garnham, C., O'Driscoll, M., Ramsden And, R., and Saeed, S. (**2002**). Speech understanding in noise with a Med-El COMBI 40+ cochlear implant using reduced channel sets. Ear Hear. **23**, 540-52.

Gfeller, K. (**2009**, June 16). Music and cochlear implants: Not in perfect harmony. The ASHA Leader.

Gifford, R. H., Shallop, J. K., and Peterson, A. M. (**2008**). Speech recognition materials and ceiling effects: considerations for cochlear implant programs. Audiol. Neurotol. **13**(3), 193-205.

Greenwood, D.D. (**1990**). A cochlear frequency-position function for several species – 29 years later. J. Acoust. Soc. Am. 87, 2592-2605.

Henry, B. A., McKay, C. M., McDermott, H. J. and Clark, G. M. (**2000**). The relationship between speech perception and electrode discrimination in cochlear implantees. J. Acoust. Soc. Am. **108**, 1269-1280.

Henry, B. A. and Turner, C. W. (**2003**). The resolution of complex spectral patterns by cochlear implant and normal-hearing listeners. J. Acoust. Soc. Am. **113**, 2861-2873.

Henry, B. A., Turner, C. W. and Behrens, A. (**2005**). Spectral peak resolution and speech recognition in quiet: Normal hearing, hearing impaired, and cochlear implant listeners. J. Acoust. Soc. Am. **118**, 1111-1121

Hillenbrand, J., Getty, L.A., Clark, M.J. and Wheeler, K. (**1995**). Acoustic characteristics of American English vowels. J. Acoust. Soc. Am. **97**, 3099-3111.

Horst, J.W. (**1987**). Frequency discrimination of complex signals, frequency selectivity, and speech perception in hearing-impaired subjects. J. Acoust. Soc. Am. 82, 874-85.

Hughes, M. L. and Stille, L. J. (**2009**). Psychophysical and psysiological measures of electrical-field interaction in cochlear implants. J. Acoust. Soc. Am. **125**, 247-60.

IEEE (**1969**). IEEE recommended practice for speech quality measurements. Trans. Audio. Electroacoust. 17, 225-246.

Jolly C.N., Spelman, F.A., and Clopton, B.M. (**1996**). Quadrupolar stimulation for cochlear prostheses: Modeling and experimental data." IEEE Trans. Biomed. Eng. **43**, 857-65.

Kawano, A., Seldon, H.L., Clark, G.M., Ramsden, R.T. and Raine, C.H. (**1998**). "Intracochlear factors contributing to psychophysical percepts following cochlear implantation." Acta Otolaryngol. **118**, 313-326.

Laneau, J., Moonen, M., and Wouters, J. (**2006**). Factors affecting the use of noise-band vocoders as acoustic models for pitch perception in cochlear implants. J. Acoust. Soc. Am. **119**, 491-506.

Lim, H.H., Tong, Y.C., and Clark, G.M. (**1989**). Forward masking patterns produced by intracochlear electrical stimulation of one and two electrode pairs in the human cochlea. J. Acoust. Soc. Am. 86, 971-80.

Litvak, L.M., Spahr, A.J., Saoji, A.A., and Fridman, G.Y. (**2007**). Relationship between perception of spectral ripple and speech recognition in cochlear implant and vocoder listeners. J. Acoust. Soc. Am. **122**, 982-91.

Liu, C. and Fu, Q.J. (**2007**). Estimation of vowel recognition with cochlear implant simulations. IEEE Trans. Biomed. Eng. **54**, 74-81.

McDermott H.J., Dorkos V.P., Dean, M.R., Ching, T.Y.C. (**1999**). Improvements in speech perception with use of the AVR TranSonic frequency-transposing hearing aid. J. Speech Lang. Hear. Res. **42**, 1323-35.

McKay, C.M., Azadpour, M., and Akhoun, I. (**2009**). In search of frequency resolution. Presented at the Conference on Implantable Auditory Prostheses, Lake Tahoe, CA, July 2009.

Mens, L.H., and Berenstein, C. K. (**2005**). Speech perception with mono- and quadrupolar electrode configuration: A crossover study. Otol. Neurotol. **26**, 957-64.

Moore, B.C. (**2001**). Dead regions in the cochlea: Diagnosis, perceptual consequences, and implications for the fitting of hearing aids. Trends Amplif. **5**, 1-34.

Moore, B.C. and Glasberg, B.R. (**1987**). Factors affecting thresholds for sinusoidal signals in narrow-band maskers with fluctuating envelopes. J. Acoust. Soc. Am. **82**, 69-79.

Moore, B.C. and Glasberg, B.R. (**1988**). Gap detection with sinusoids and noise in normal, impaired, and electrically stimulated ears. J. Acoust. Soc. Am. **83**, 1093-1101.

Munson, B., and Nelson, P.B. (**2005**). Phonetic identification in quiet and in noise by listeners with cochlear implants. J. Acoust. Soc. Am. **118**, 2607-17.

Nelson, D. A., Donaldson, G. S., and Kreft, H. (**2008**). Forward-masked spatial tuning curves in cochlear implant users. J. Acoust. Soc. Am. **123**, 1522-1543.

Nelson, D.A., Kreft, H.A., Anderson, E.S., and Donaldson, G.S. (**2011**). Spatial tuning curves from apical, middle, and basal electrodes in cochlear implant users. J. Acoust. Soc. Am. **129**, 3916-33.

Nelson, D. A., Van Tasell, D. J., Schroder, A. C., Soli, S., and Levine, S. (**1995**). Electrode ranking of ``place pitch'' and speech recognition in electrical hearing. J. Acoust. Soc. Am. **98**, 1987-1999.

Nilsson, M., Soli, S.D., and Sullivan, J.A. (**1994**). Development of the Hearing In Noise Test for the measurement of speech reception thresholds in quiet and in noise. J. Acoust. Soc. Am. **95**, 1085-99.

Noordhoek, I.M. and Drullman, R. (**1997**). Effect of reducing temporal intensity modulations on sentence intelligibility. J. Acoust. Soc. Am. **97**, 585-92.

Parent T.C., Chmiel R., Jerger J. (**1997**). Comparison of performance with frequency transposition hearing aids and conventional hearing aids. J Am Acad Audiol **8**, 355-365.

Peters, R.W., Moore, B.C.J., and Baer, T. (**1998**). Speech reception thresholds in noise with and without spectral and temporal dips for hearing-impaired and normally hearing people. J. Acoust. Soc. Am. **103**, 577-87.

Qin, M. K., and Oxenham, A. J. (**2003**). "Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers," J. Acoust. Soc. Am. **114**, 446-454.

Rosen, S. (**1989**). Temporal information in speech and its relevance for cochlear implants. In *Cochlear implant: acquisitions and controversies* (ed. B. Fraysse and N. Cochard), pp. 3-26, Basel: Cochlear AG.

Rosen, S., Faulkner, A., Wilkinson, L (**1999**). Adaptation by normal listeners to upward spectral shifts of speech: implications for cochlear implants. J. Acoust. Soc. Am. **106**, 3629-36.

Saoji, A.A. and Eddins, D.A. (**2007**). Spectral modulation masking patterns reveal tuning to spectral envelope frequency. J. Acoust. Soc. Am. **122**, 1004-12.

Saoji, A., Litvak, L., Emadi, G., Spahr, T., and Greenslade, K. (**2005**). Spectral modulation transfer functions in cochlear implant listeners. Poster presented at the Conference on Implantable Auditory Prostheses, Asilomar, CA, July 2005.

Saoji, A.A., Litvak, L., Spahr, A.J., and Eddins, D.A. (**2009**). Spectral modulation detection and vowel and consonant identifications in cochlear implant listeners. J. Acoust. Soc. Am. **126**, 955-58.

Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J. and Ekelid, M. (**1995**). Speech recognition with primarily temporal cues. Science **270**, 303-304.

Spahr, A.J., and Dorman, M.F. (**2005**). Effects of minimum stimulation setting for the Med El Tempo+ speech processor on speech understanding. Ear Hear. 26, 2S-6S

Steeneken, H.J., and Houtgast, T. (**1979**). A physical method for measuring speech-transmission quality. J. Acoust. Soc. Am. **67**, 318-26.

Studebaker, G. (**1985**). "A ‘rationalized’ arcsine transform," J. Speech Hear. Res. **28**, 455-462.

Summers, V. and Leek, M.R. (**1994**). The internal representation of spectral contrast in hearing-impaired listeners. J. Acoust. Soc. Am. **95**, 3518-28.

Supin, A. Y., Popov, V. V., Milekhina, O. N. and Tarakanov, M. B. (**1994**). “Frequency resolving power measured by rippled noise,” Hear. Res. **78**, 31-40.

Supin, A. Y., Popov, V. V., Milekhina, O. N. and Tarakanov, M. B. (**1997**). Frequency-temporal resolution of hearing measured by rippled noise. Hear. Res. **108**, 17-27.

Thornton, A.R. and Raffin, M.J.M. (**1978**). Speech-discrimination scores modeled as a binomial variable. J. Speech Hear. Res. **21**, 507-18.

Throckmorton, C. S. and Collins, L. M. (**1999**). Investigation of the effects of temporal and spatial interactions on speech-recognition skills in cochlear-implant subjects. J. Acoust. Soc. Am. **105**, 861-873.

Throckmorton, C.S., and Collins, L.M. (**2002**). The effect of channel interactions on speech recognition in cochlear implant subjects: predictions from an acoustic model. J. Acoust. Soc. Am. **112**, 285-96.

Van Tasell, D.J., Soli, S.D., Kirby, V.M., and Widin, G.P. (**1987**). Speech waveform envelope cues for consonant recognition. J. Acoust. Soc. Amer. 82, 1152-61.

Verschuur, C. (**2009**). Modeling the effect of channel number and interaction on consonant recognition in a cochlear implant peak-picking strategy. J. Acoust. Soc. Am. **125**, 1723-36.

Wang, M.D. and Bilger, R.C. (**1973**). Consonant confusions in noise: a study of perceptual features. J. Acoust. Soc. Am. 54, 1248-66.

Whitmal and DeRoy (**2011**). Adaptive bandwidth measurements of importance functions for speech intelligibility predictions. Presented at the American Auditory Society Scientific and Technology Meeting, Scottsdale, AZ, March, 2011.

Won, J. H., Drennan, W. R. and Rubinstein, J. T. (**2007**). Spectral-ripple resolution correlates with speech reception in noise in cochlear implant users. J. Assoc. Res. Otolaryngol. **8**, 384-392.

Won, J.H., Drennan, W.R., Kang, R.S., and Rubinstein, J.T. (**2010**). Psychoacoustic abilities associated with music perception in cochlear implant users. Ear Hear. **31**, 796-805.

Yost, W.A., and Sheft, S. (**1989**). Across-critical-band processing of amplitude-modulated tones. J. Acoust. Soc. Am. **85**, 848-57.

Zeng, F.G., Popper, A. N., and Fay, R. R. (Eds.). (2004). *Cochlear implants: Auditory prostheses and electric hearing*. New York: Springer-Verlag.

Zwolan, T. A., Collins, L. M. and Wakefield, G. H. (**1997**). Electrode discrimination and speech recognition in postlingually deafened adult cochlear implant subjects, J. Acoust. Soc. Am. **102**, 3673-3685.}

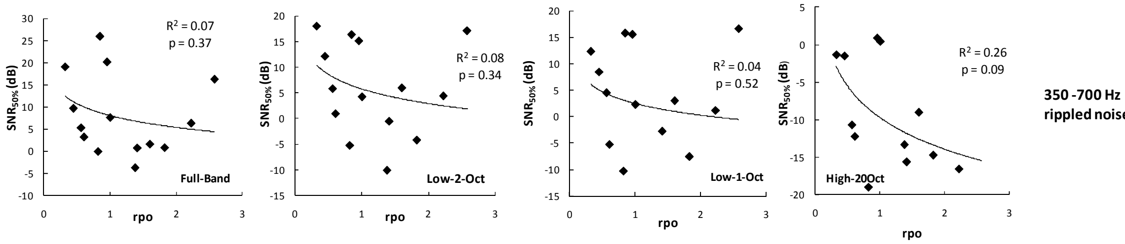
APPENDICES

Table A1

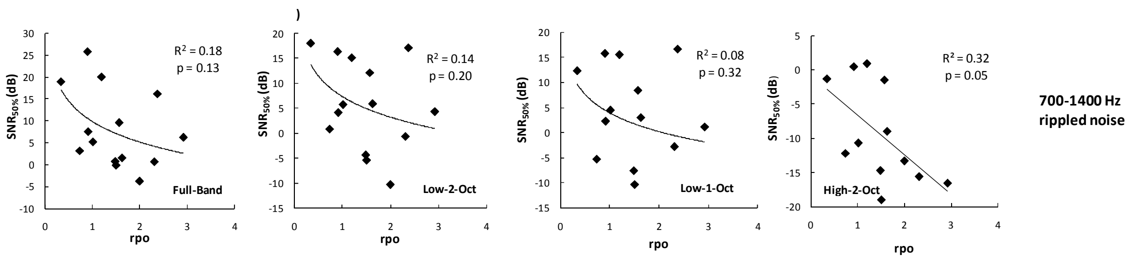
Individual spectral modulation thresholds (in dB) as a function of spectral ripple rate for extended ripple frequency range.

Subject	C03	C05	C16	C18	C23	D02	D05	D08	D10	D19	N13	N14	N28	N32	N34
	SMT (dB)														
Ripple rate	9.4	5.6	5.7	15.0	21.6	7.0	12.0	9.7	6.6	10.0	14.0	12.8	19.6	7.2	15.5
0.25	12.5	6.0	9.4	19.1	27.5	9.6	14.4	11.5	8.8	15.8	18.4	13.9	13.6	10.4	18.0
0.5	17.1	8.6	13.4	27.0	35.9	8.2	20.0	19.8	7.5	20.4	17.6	15.3	14.0	10.8	30.2
0.75	17.9	10.4	11.3	27.0		6.9	23.9	19.1	15.9	24.0	21.2	19.9	20.2	12.8	31.2
1	17.2	11.1	10.3	26.9		8.4	32.3	20.4	17.6	31.5	27.1	25.6	27.0	22.1	29.0
1.5	20.0	10.8	13.9	25.3		11.5	35.1	23.9	25.1	30.0	31.3	25.9	37.6	20.6	36.1
2	14.5	9.6	12.8	36.6		12.3	30.8	20.4	18.0	46.4	29.7	36.2	41.4	25.6	46.3
3	23.6	12.6	17.4			24.6		13.0	13.6					23.8	
6	35.6	15.1	20.8			30.1		15.8	16.0					23.8	
9	31.9	19.6	25.0			34.5		20.8	21.4					35.4	
12		24.7	28.3			33.6		23.6	23.7						
15		32.2	27.4			37.6		26.6	27.1						
18		32.3						30.7	27.8						
21															

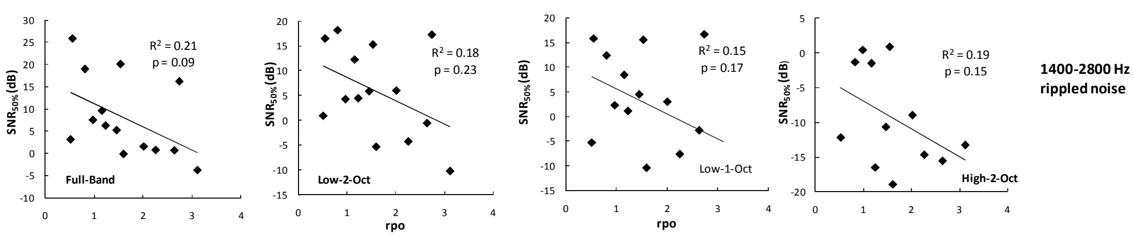
Figure A1.



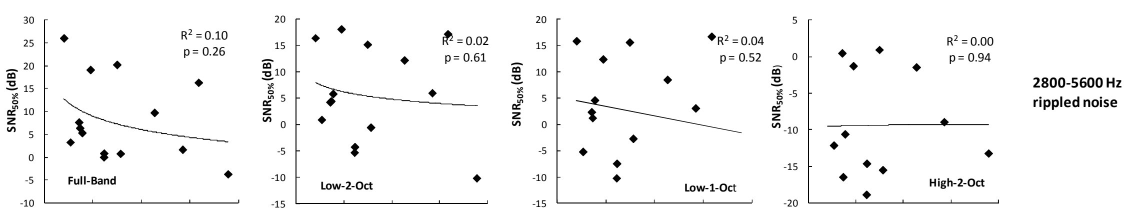
Sentence recognition in band-limited noise ($SNR_{50\%}$) as a function of octave-band spectral ripple discrimination thresholds for 350-700 Hz rippled noise stimuli. Each set of panels illustrate, from left to right, Full-band masker, Low-2-oct masker, Low-1-oct masker, and High-2-oct masker.



Sentence recognition in band-limited noise ($SNR_{50\%}$) as a function of octave-band spectral ripple discrimination thresholds for 700-1400 Hz rippled noise stimuli.



Sentence recognition in band-limited noise ($SNR_{50\%}$) as a function of octave-band spectral ripple discrimination thresholds for 1400-2800 Hz rippled noise stimuli.



Sentence recognition in band-limited noise ($SNR_{50\%}$) as a function of octave-band spectral ripple discrimination thresholds for 2800-5600 Hz rippled noise stimuli.