Effects of spectral smearing on performance of the spectral ripple and spectro-temporal ripple tests

Vijaya Kumar Narne[a]

Department of Audiology, Gulf Medical University, Al Jerf, Ajman, 4178, UAE

Mridula Sharma[b,c]

[b] *Department of Linguistics, Australian Hearing Hub, 16 University Avenue, Macquarie University, NSW, 2109, Australia*

Bram Van Dun [c,d],

[d] *National Acoustic Laboratories, Level 5, Australian Hearing Hub, 16 University Ave, Macquarie University, NSW 2109, Australia*

Shalini Bansal and Latika Prabhu

Department of Audiology, All India Institute of Speech and Hearing, Manasagangothri, Mysore, 570006, India

Brian C.J. Moore

Department of Experimental Psychology, University of Cambridge, Downing street, Cambridge, CB2 3EB, UK

a) Author to whom correspondence should be addressed. Electronic mail:

vijaynarne@gmail.com

Keywords: Spectral ripple; Spectral resolution; Spectral smearing; Masking

Second revision dated October, 2016

[c] Also at The *HEARing CRC, 550 Swanston Street, Department of Audiology and Speech Pathology, The University of Melbourne, Carlton VIC 3053, Australia.*

Abstract

The main aim of this study was to use spectral smearing to evaluate the efficacy of a spectral ripple test (SRt) using stationary sounds and a recent variant with gliding ripples called the spectro-temporal ripple test (STRt) in measuring reduced spectral resolution. In Experiment 1 the highest detectable ripple density was measured using four amounts of spectral smearing (unsmeared, mild, moderate and severe). The thresholds worsened with increasing smearing and were similar for the SRt and the STRt across the three conditions with smearing. For unsmeared stimuli, thresholds were significantly higher (better) for the STRt than for the SRt. An amplitude fluctuation at the outputs of simulated (gammatone) auditory filters centered above 6,400 Hz was identified as providing a potential detection cue for the STRt stimuli. Experiment 2 used notched noise with energy below and above the passband of the SRt and STRt stimuli to reduce confounding cues in the STRt. Thresholds were almost identical for the STRt and SRt for both unsmeared and smeared stimuli, indicating that the confounding cue for the STRt was eliminated by the notched noise. Thresholds obtained with notched noise present could be predicted reasonably accurately using an excitation-pattern model.

I. INTRODUCTION

Tests of the ability to detect or discriminate spectral ripples have often been used to measure the frequency selectivity of the auditory system (also called spectral resolution) and to study spectral shape perception. Such tests employ a "carrier" that is either white noise or multiple sinusoids covering a certain frequency range. The spectrum of the carrier is modulated with a sinusoidal function on a logarithmic or a linear frequency scale (Supin *et al*., 1998; Won *et al*., 2007). The result is a broadband stimulus with regular amplitude fluctuations along the frequency axis (spectral modulations). The rate of spectral modulation, or ripple density, is specified as the number of ripples per Hz for linearly-spaced ripples and number of ripples per octave (RPO) for logarithmic spacing.

Spectral resolution and spectral shape perception have been studied using such stimuli in several tasks. Two of the most common are (1) spectral ripple detection (Litvak *et al*., 2007; Saoji *et al*., 2009; Anderson *et al*., 2012) and (2) spectral ripple discrimination (Supin *et al*., 1998; Won *et al*., 2007; Aronoff and Landsberger, 2013). In the ripple detection task, the ripple density is held constant in a given run and the spectral modulation depth is adaptively varied to estimate the smallest spectral modulation depth (in dB) that can be detected, i.e. discriminated from an unmodulated carrier. Performance is measured as a function of ripple density. In the spectral ripple discrimination task, the listener is asked to discriminate between a reference stimulus with a specific ripple density and a target stimulus that differs from the reference in some way. There are at least two variants of this task. In one, the target has the same ripple density as the reference, but the phase of the spectral modulation is changed for the target (ripple phase discrimination). The smallest detectable change in ripple phase as a function of modulation depth may be measured, or the smallest ripple depth at which a given phase change can be detected may be measured. In the other variant of the spectral ripple discrimination task, the ripple density differs for the target and

reference (ripple density discrimination). This second variant, estimating the smallest detectable difference in ripple density, was used in the present study. However, the ripple density of the reference stimulus used here was sufficiently high that the spectral ripple was essentially undetectable, so the task resembled ripple detection. The highest value of the ripple density of the target at which the task could be performed was measured using a fixed spectral modulation depth. This test is referred to hereafter as the spectral ripple test, SRt.

Measures of performance on a variety of spectral ripple tests, which are usually assumed to reflect frequency selectivity, have been found to be correlated with aspects of hearing performance (speech and music perception) for listeners with normal hearing, sensorineural hearing loss, and cochlear implants (Henry *et al*., 2005; Litvak *et al*., 2007; Won *et al*., 2007; Saoji *et al*., 2009; Jones *et al*., 2013). However, Azadpour and McKay (2012) have shown that thresholds obtained using such tests may not provide an accurate measure of spectral resolution for cochlear implant users, for whom even ripples with very low ripple densities are hard to detect and discriminate. They reported that, apart from spectral cues relating to the ripples, there were additional confounding cues, such as differences in loudness, spectral centroid (the weighted mean frequency), and changing energy at the spectral edges, that led to better thresholds than would be obtained if only cues from the spectral ripples were used.

There are several ways to reduce the use of these confounding cues, including randomizing the level of the stimuli as well as the starting phase of the spectral modulation. However, randomizing the level can be problematic for cochlear implant users, because of their very small dynamic range. Aronoff and Landsberger (2013) proposed a method for reducing the use of confounding cues using a modified version of the SRt in which the phase of the spectral modulation changed over time, called here the spectro-temporal ripple test (STRt). As a result of the temporal changes, the overall loudness evoked in each local

frequency region, the energy at the spectral edges, and the spectral centroid became unreliable cues for discriminating the target and reference stimuli. Aronoff and Landsberger (2013) evaluated the effectiveness of the STRt in assessing spectral resolution using a ripple density discrimination task. As in the present study, the ripple density of the reference stimulus was sufficiently high that the task resembled spectral ripple detection. They simulated reduced spectral resolution by vocoding the stimuli. They demonstrated that as the number of vocoder channels decreased, simulating progressively poorer frequency selectivity, the highest ripple density that could be detected also decreased, suggesting that the STRt provided a measure of spectral resolution.

Vocoder techniques have been employed in several studies to simulate the effects of reduced spectral resolution, especially for simulating the information provided by cochlear implants (Shannon *et al*., 1995; Friesen *et al*., 2001; Litvak *et al*., 2007; Aronoff and Landsberger, 2013). However, there are limitations of this approach. For example, when a noise vocoder is used, the random amplitude fluctuations in the noise carriers may affect performance and the filters used to extract the channel envelopes may limit the representation of cues relating to fundamental frequency (Stone *et al*., 2008). Also, temporal fine structure information may be reconstructed from envelope information (Gilbert and Lorenzi, 2006).

Another way to simulate reduced spectral resolution is to smear the power spectrum of the signal (Baer and Moore, 1993). This method can create excitation patterns in a normal ear that resemble those in an impaired ear with reduced spectral resolution. Any desired degree of reduced frequency selectivity can be simulated in a quantitative manner. The method gives a reasonably accurate simulation of the effect of cochlear hearing loss on frequency selectivity measured using the notched-noise method (Moore *et al*., 1992). In the present study, spectral smearing was used to simulate several degrees of reduced frequency selectivity. Ripple density discrimination with a high reference ripple density was compared

for the spectral ripple test (SRt) and the modified spectro-temporal ripple test (STRt) for each degree of smearing, to assess whether the test outcomes reflected the degree of smearing and to assess whether the tests gave similar results. Experiment 1 revealed a discrepancy between the results of the SRt and STRt for unsmeared stimuli, but not for smeared stimuli. Simulations of auditory filtering identified a potential confounding cue, not directly connected with frequency selectivity, for the STRt. The cue took the form of amplitude fluctuations in a frequency region outside the nominal frequency range of the stimulus. Experiment 2 made use of a notched noise to mask the confounding cue and to assess its influence. The results obtained in the presence of the notched noise could be predicted well using an excitation-pattern model, supporting the use of the SRt and STRt tests as measures of spectral resolution.

II. EXPERIMENT 1

A. Method

1. Participants

There were ten normal-hearing participants (2 males and 8 females) with a mean age of 20 years (range: 18 – 25, standard deviation, SD = 2.4 years). Pure-tone thresholds and speech identification scores were measured using a calibrated diagnostic audiometer (GSI-61, Grason-Stadler, Eden Prairie, MN) with Telephonics TDH-39 headphones (Huntingdon, NY). All participants had pure-tone thresholds ≤15 dB HL at octave frequencies between 0.25 and 8 kHz, as estimated using the modified Hughson and Westlake procedure (Carhart and Jerger, 1959). All had identification scores for speech in quiet greater than 90% at 40 dB SL (relative to the pure-tone average at 0.5, 1, 2 and 4 kHz) using word lists developed at the All India Institute of Speech and Hearing (AIISH), Mysore, India. All participants had normal tympanograms, with ipsilateral acoustic reflexes at normal levels in both ears, indicating normal middle ear function. None of the participants had a history of neurological and/or otological disorder.

All the testing procedures were approved by the All India Institute of Speech and

Hearing review board, and written informed consent was obtained from all participants.

2. Spectral ripple stimuli

The stimuli used in the SRt and the STRt were generated as described by Aronoff and Landsberger (2013). The stimuli were 750-ms long with 100-ms cosine-squared onset and offset ramps, and were sampled at 22,050 Hz. They were generated using a carrier composed of 201 equal-amplitude sinusoidal frequency components, spaced every 0.03rd octave from 100 to 6,400 Hz. The amplitude of each of the sinusoids was a function of the ripple density, *D* (in ripples per octave, RPO), and the ripple rate (*R*, in ripples per s) (when temporal modulation was present, i.e., for the STRt stimuli only). The STRt stimulus as a function of time, STRt(*t*), is defined by eq. 1 and 2:

$$STRt(t) = \sum_{i=1}^{i=201} \frac{P(i)}{201} \times 10^{\frac{d\left(\left|\sin\left(\pi Rt + 0.03i\pi D + \theta_2\right)\right| - 1\right)}{20}} \quad \text{(eq. 1)}$$

$$\text{where} \qquad P(i) = \sin(2\pi f_c(i)t + \theta_1 \quad \text{(eq. 2)}$$

$P(i)$ is the ith carrier component with frequency $f_c(i) = 100 \times 2^{0.03(i-1)}$ (i = 1-201), corresponding to f_C = 100-6400 Hz, θ_1 is the starting phase of each sinusoidal component in the carrier, which was chosen randomly for each component over the range 0-2π radians, θ_2 determines the phase of the spectral ripple at the onset of the stimulus, which was chosen randomly for each stimulus over the range 0-π radians, and *d* is the ripple depth in dB.[1]

To describe eq. 1: the spectral modulator was initially a sinusoid on a logarithmic frequency scale, with spectral frequency equal to 0.5*D*. This was rectified by taking the absolute value, giving a function repeating with rate *D*, with a value varying between 0 and 1. One was subtracted from this function to give a function varying between 0 and −1. This was multiplied by the desired ripple depth, *d*, in dB. The value of the result was divided by 20 and the antilog (base 10) was taken. This sequence provided the final function that determined the amplitude of each carrier component as a function of frequency. The spectral ripples were

approximately, but not exactly, sinusoidal in form on a logarithmic frequency axis. For our stimuli, *d* was 20 dB, so the peak-to-valley ratio in the spectrum was 20 dB. For the STRt stimuli, the peaks and valleys in the spectral ripple shifted along the frequency axis such that the spectral pattern repeated every 200 ms (*R* = 5 Hz). For the SRt, *R* was set to zero, so the function applied to each carrier component was simply a scaling of the overall amplitude. Example spectrograms for STRt and SRt stimuli are shown in the two left panels of Fig. 1.

The overall level of both the unsmeared and spectrally smeared stimuli (see below) was 75 dB SPL, which was reported by participants to be comfortably loud.

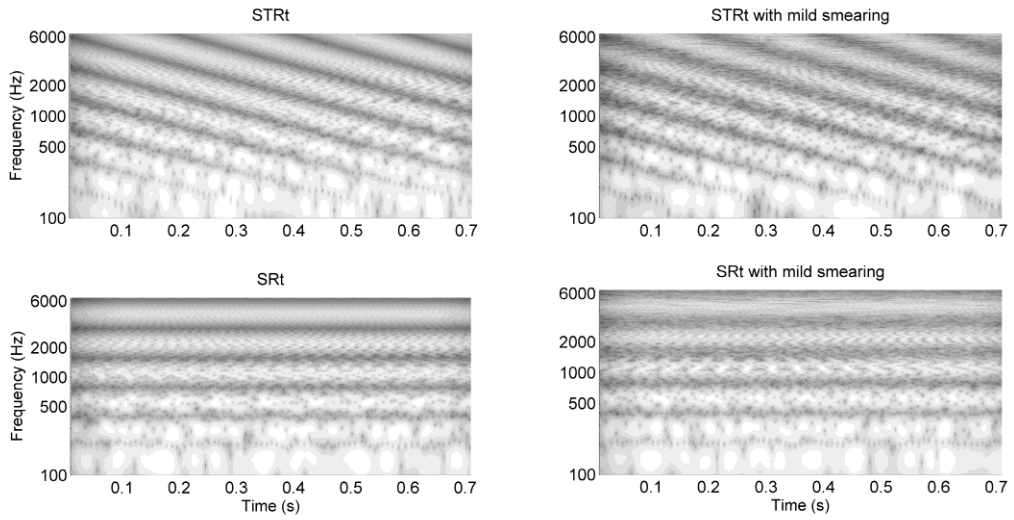

FIG. 1. *Spectrograms of unsmeared stimuli (left) and stimuli with mild spectral smearing (right) for the STRt (top) and the SRt (bottom). The ripple density was 1 RPO and the ripple rate was 5 Hz.*

3. Simulation of reduced spectral resolution

The spectral smearing used in the current study was similar to that described by Baer and Moore (1993). The smearing was carried out using an overlap-add algorithm. The spectral-ripple stimulus was segmented into frames, each 256 samples (11.6 ms) long and a Hamming window was applied to each frame. The magnitude and phase of the components in each frame were extracted using a 256-point Fast Fourier Transform (FFT). The magnitude spectrum was multiplied by a smearing matrix (described in more detail below). The

application of an inverse FFT gave the temporal waveform of the spectrally smeared frame, which was then scaled to have the same amplitude as the unsmeared signal. Finally, the time waveforms were added together using 75% overlap.

The smearing matrix was created using the following procedure. An auditory filter function was computed for each component frequency in the input spectrum. Auditory filter functions were derived using the one-parameter form of the *roex*(*p*) filter (Patterson and Nimmo-Smith, 1980). The value of *p* for the unsmeared stimuli varied with frequency according to the equation given by Glasberg and Moore (1990). A total of 128 auditory filter functions was created and stored in a 128-by-128 matrix. Widened auditory filter functions were created by dividing the value of *p* by 3 for mild smearing, 6 for moderate smearing and 9 for severe smearing. The smeared matrix was derived by multiplying the inverse matrix of a normal auditory filter bank [with values of *p* based on the equation given by Glasberg and Moore (1990)] with the matrix for the widened auditory filter bank. This compensated for the spectral smearing effects of the auditory filters of the normal-hearing participants. The right panels of Fig. 1 show spectrograms of STRt and SRt stimuli with mild spectral smearing.

4. Procedure

All participants were seated comfortably in a quiet room. Thresholds were estimated using a three-alternative forced-choice (3AFC) procedure. Two of the intervals contained a reference stimulus with a ripple density of 20 RPO (for which the ripples were assumed to be completely unresolved). The ripple density of the target was initially 0.5 RPO and it was modified with a 2-down 1-up tracking method estimating the 70.7% correct point on the psychometric function (Levitt, 1971). The step size in ripple density was initially 0.5 RPO and it was decreased to 0.1 RPO after two reversals. The arithmetic mean ripple density at the last six reversal points in a block of eight was taken as the threshold. All participants were

given two practice runs before commencing the experiment and a single threshold estimate was obtained for each condition and each participant. The ripple density at threshold was never higher than 10 and the ripple density of the target never became close to that of the reference during an adaptive track.

MATLAB 7.10.0 (Mathworks; Natick, MA, USA) software installed on a personal computer (PC) was used to control stimulus presentation. The stimuli were presented via a PC fitted with a 32 bit Realtek high definition sound card and routed through an audiometer (Madsen OB-922) to the headphones (TDH-39 with MX 41/AR cushions). The order of testing of the stimulus conditions was randomized across participants.

5. *Statistical analysis*

Modern statistical techniques were employed to minimize the potential effects of any outliers or non-normality in the data [for more details, see (Wilcox, 2012a)]. These statistical techniques employ 20% trimmed means, which are a cross between a mean and a median. Since there were 10 participants, 20% trimming meant that the thresholds for the worst-performing and best-performing participant in each condition were discarded. Trimmed means were used to estimate the means shown in the figures, for analyses of variance (ANOVAs), and for post hoc tests. Bootstrap techniques were also employed, which allow tests to be conducted using distributions based on the original data set rather than on assumed normal distributions that may not accurately reflect the data (Wilcox, 2012a). All analysis was performed using the WRS package in R 3.2.3 (Wilcox, 2012b). Thresholds across the four conditions of spectral smearing (i.e., unsmeared, mild, moderate and severe smearing) using the two tests (SRt and STRt) were compared using a within-subjects ANOVA (Wilcox, 2012b).

B. Results

The results are shown in Fig. 2. The highest ripple density of the target at which it could just be discriminated from the reference is plotted for each test and each smearing condition. There was a significant main effect of smearing condition ($F_t = 18.03$, $p < 0.0001$), increasing spectral smearing resulting in decreasing (poorer) ripple thresholds. There was a significant main effect of test type ($F_t = 156.78$, $p < 0.0001$), the STRt giving higher ripple thresholds than the SRt. There was a significant interaction ($F_t = 5.96$, $p < 0.001$) between smearing condition and test type. To investigate the interaction, multiple comparisons with Rom's method for controlling familywise error (Rom, 1990) were performed. All of the smearing conditions except moderate and severe were significantly different from one another ($p<0.001$) for both the STRt and the SRt. Pairwise comparisons were performed using Yuen's method. Thresholds were significantly higher (better) for the STRt than for the SRt only for the condition with no smearing ($p<0.001$).

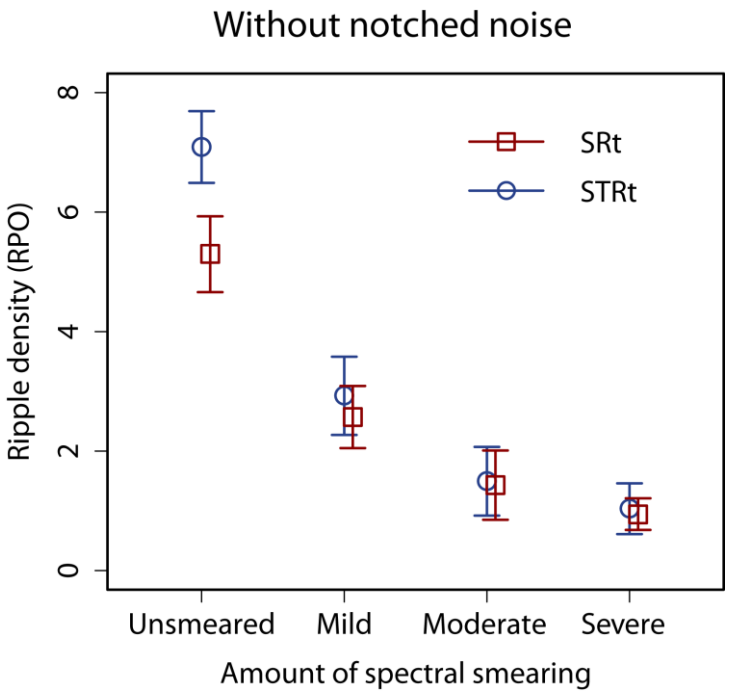

FIG. 2. *Trimmed means with error bars (95% confidence intervals estimated using 20% trimmed mean and bootstrapping method) for each spectral smearing condition, for the SRt (circles) and the STRt (squares) (color online). For visual clarity, data points are plotted slightly offset from their correct positions along the x-axis.*

C. Discussion

Increasing the amount of spectral smearing, simulating progressive broadening of the auditory filters, resulted in a significant decrease of ripple thresholds (indicating poorer spectral resolution) for both the SRt and STRt. This clearly shows that both the SRt and the STRt are sensitive to simulated changes in frequency selectivity.

The mean ripple threshold for the STRt stimuli in the condition without smearing was approximately 7.1 RPO, which is slightly less than the mean threshold of 8.2 RPO reported by Aronoff and Landsberger (2013). The difference in mean threshold across studies may be due to the use of different tracking rules. While both studies used a 3AFC procedure, Aronoff and Landsberger (2013) estimated the threshold corresponding to 50% correct using a 1-up 1-down procedure, which tracks $d' = 0.56$, whereas in the present study threshold was estimated at the 70.7% correct point using a 2-down 1-up procedure, which tracks $d' = 1.28$. The mean threshold for the SRt stimuli in the condition without smearing was 5.7 RPO which is similar to the mean threshold of 5.5 RPO found by Henry *et al*. (2005) and Peter *et al*. (2014). The higher thresholds for the STRt stimuli than for the SRt stimuli may have occurred because additional cues were available for the STRt stimuli in the condition without smearing.

Without smearing, both the STRt and SRt stimuli had sharp spectral edges at 100 and 6,400 Hz. The spectral smearing would have removed these sharp spectral edges. Hence, it seems likely that whatever extra cue was available for the STRt stimuli without smearing was connected with the sharp spectral edges. It was hypothesized that STRt stimuli with no smearing produced amplitude fluctuations at the outputs of auditory filters centered above 6,400 Hz, and that these fluctuations provided a detection cue. To assess this hypothesis, STRt and SRt stimuli with a ripple density of 7 RPO (i.e. above the threshold for the SRt test with unsmeared stimuli) were passed through an array of 4^{th} order gammatone filters, which provide a model of the auditory filters (Patterson et al., 1995). The bandwidths of the filters

were as specified by Glasberg and Moore (1990). The waveforms at the outputs of some of the simulated auditory filters are shown in part A of Fig. 3. These filters were selected because they revealed possible cues for the unsmeared STRt stimuli that were not present for the SRt stimuli. The unsmeared STRt stimuli (top) led to distinct regular amplitude fluctuations corresponding to the *R* of 5 Hz at the outputs of simulated auditory filters centred at 6,400 and 7000 Hz (see the waveform highlighted by ovals), but not for lower center frequencies. These fluctuations are more clearly visible in panel B, which shows the waveform at the output of the simulated auditory filter centered at 7000 Hz, with the envelope superimposed (white trace). The envelope shown is the Hilbert envelope after lowpass filtering using a finite-impulse-response filter with a cutoff frequency of 100 Hz. The modulation depth of the envelope was about 25%. The inset in panel B shows the spectrum of the envelope, which has a clear peak at 5 Hz, corresponding to the *R*. There was no peak at 5 Hz in the spectrum of the envelope of the reference stimulus (not shown), so the 5-Hz envelope fluctuation provided a potential detection cue for the STRt stimuli. As shown in panel C, there was no peak in the spectrum of the envelope at 5 Hz for the SRt stimulus. When smearing was introduced, the amplitude fluctuation at 5 Hz was absent for both the STRt and the SRt stimuli.

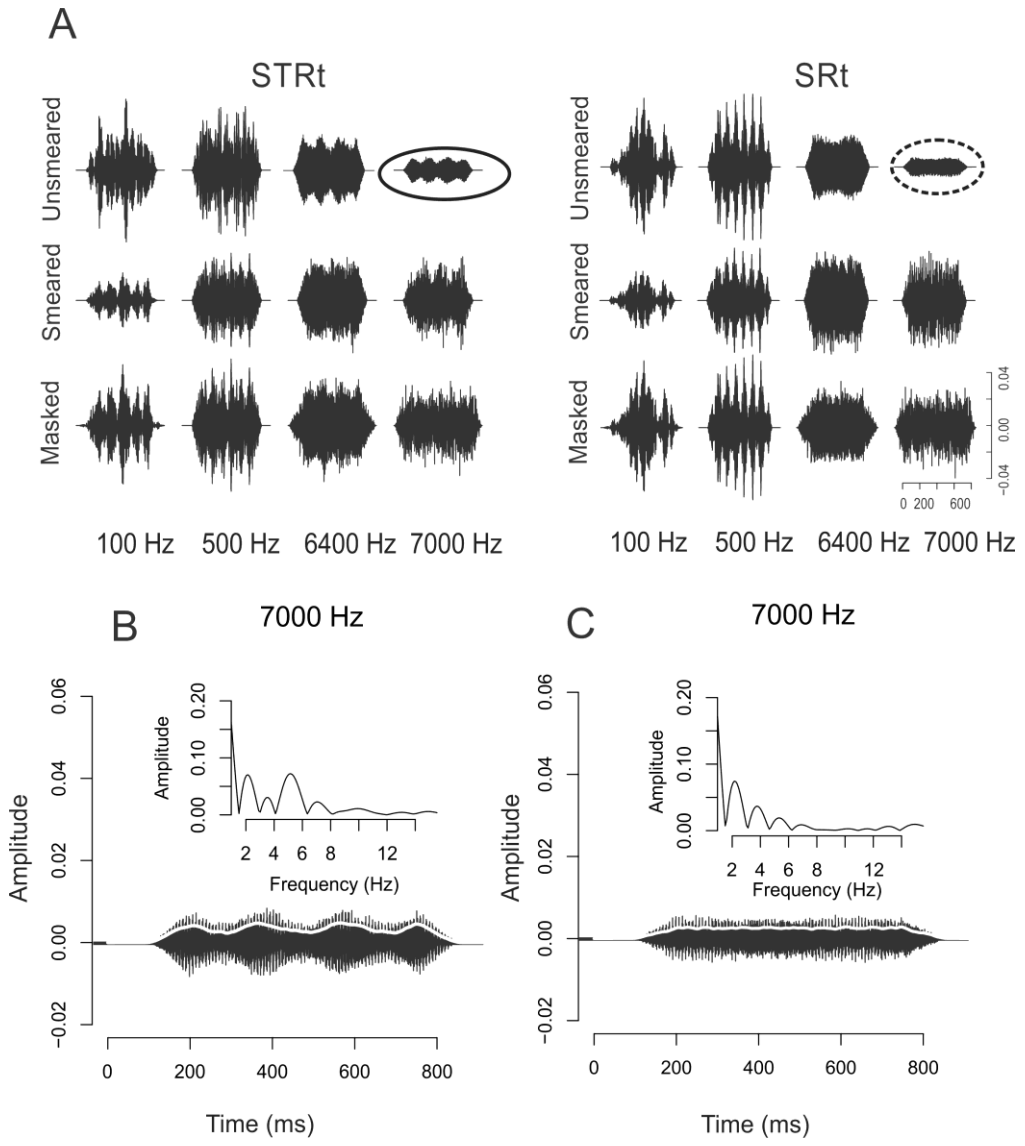

FIG. 3. *Part A shows waveforms at the outputs of simulated auditory (gammatone) filters with selected center frequencies for STRt stimuli (left) and SRt stimuli (right) with a ripple density of 7 RPO for three conditions: unsmeared ("Unsmeared"), severely smeared ("Smeared"), and unsmeared with notched-noise added ("Masked"). The filter center frequency in Hertz is indicated at the bottom. The amplitude scale is arbitrary, but the scaling was the same for all stimuli. The ovals highlights waveforms at the outputs of filters centered at 7000 Hz, showing 5-Hz amplitude modulation for the unsmeared STRt stimuli but not the unsmeared SRt stimuli. This is illustrated in more detail in panels B and C, which show the waveform and envelope (white trace) at the output of the auditory filter centered at 7000 Hz for the unsmeared STRt and SRt stimuli, respectively. The insets in panels B and C show the spectra of the envelopes, which have a peak at 5 Hz for the STRt but not for the SRt stimuli.*

For the SRt unsmeared stimuli (and to a lesser extent for the smeared stimuli), regular envelope fluctuations are apparent at the output of the gammatone filter centered at 500 Hz. These fluctuations are a consequence of beats between adjacent sinusoidal components in the carrier, whose spacing was approximately 10.5 Hz at the center frequency of 500 Hz. Such beats also occurred at the outputs of other gammatone filters (with higher rates for higher center frequencies), but they are not easily visible in the waveforms. The beats were present for both the test and reference stimuli and so presumably did not provide a cue for detecting the spectral ripple.

III. EXPERIMENT 2

To assess whether cues at the outputs of auditory filters centered above 6,400 Hz led to better performance for the unsmeared STRt stimuli than for the unsmeared SRt stimuli, a second experiment was performed in which notched noise was added to all stimuli, to make such cues unusable.

A. Method

1. Participants

There were ten normal-hearing participants (4 males and 6 females) with a mean age of 25 years (range: 20 – 35 years, SD = 5.3). Their characteristics were the same as for Experiment 1.

2. Stimulus, procedure and statistical analysis

The stimuli were the same as used in Experiment 1 except that sharply filtered (digitally generated) notched noise was added to all stimuli. The notched noise onset preceded the onset of each stimulus by 100 ms and the notched noise continued for 100 ms after the end of the stimulus. The noise was ramped on and off using a cosine-squared function with ramp duration of 100 ms. The lower and upper edge frequencies of the notch were 100 and 6,400 Hz, respectively, coinciding with the edge frequencies of the SRt and STRt stimuli. The root-mean-square level of the notched noise was 10 dB below that of the

SRt and STRt stimuli. This relative level was chosen based on the outputs of the simulated auditory filters, so that the amplitude fluctuations above 6,400 Hz would be completely masked while the notched noise had minimal effects on the outputs of auditory filters centered between 100 and 6,000 Hz. The influence of the notched noise is illustrated in Fig. 3 by the rows labelled "Masked". Procedures were identical to those for Experiment 1.

B. Results

The results of experiment 2 are shown in Fig. 4 by the open symbols. To assess whether the notched noise had any significant effect on ripple threshold across the four smearing conditions for the SRt and the STRt tests, the data for experiments 1 and 2 were compared using a mixed-model ANOVA. Amount of smearing and test condition were within-subject factors, and presence/absence of notched noise was a between-subjects factor. A significant main effect of amount of smearing was found ($F_t = 96.2$, $p < 0.0001$), thresholds decreasing with increasing spectral smearing, as expected. There was a main effect of test type ($F_t = 9.8$, $p = 0.001$), thresholds being higher for the STRt than for the SRt. There was no significant main effect of the presence/absence of notched noise ($F_t = 0.14$, $p = 0.72$). However, there was a significant interaction of test type and presence/absence of noise ($F_t = 6.1$, $p=0.013$). All other two-way and three-way interactions were not significant ($p > 0.05$).

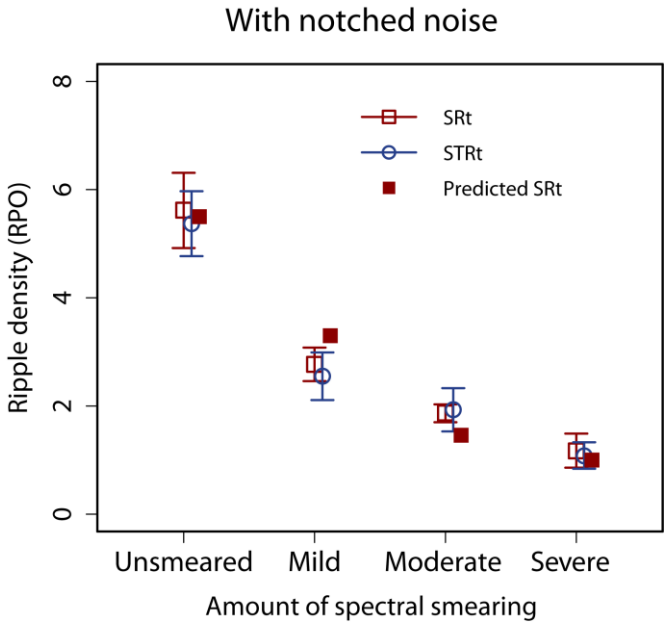

FIG. 4*. As Fig. 2 but for stimuli presented together with notched noise. Thresholds predicted using the excitation-pattern model are shown by filled symbols.*

To assess the effect of the notched noise on ripple thresholds, planned comparisons were performed using Yuen's test. The addition of notched noise led to a significantly decreased mean threshold for the STRt unsmeared stimuli (Ty = 4.80, p = 0.0013); the decrease was 1.5 RPO. The effect of the notched noise was not significant for any other amount of smearing (p>0.05).

When the notched noise was present, there was no significant difference between the thresholds for the STRt and SRt stimuli in the unsmeared condition (Ty = –0.12, p = 0.9). This confirms that the cues that led to better performance for the STRt than for the SRt stimuli in experiment 1 were removed by the notched noise used in experiment 2.

C. Discussion

The results support the hypothesis that amplitude fluctuations at the outputs of auditory filters centered above 6,400 Hz provided a cue that led to higher (better) thresholds for the STRt than for the SRt stimuli in the unsmeared condition. The addition of notched

noise eliminated these cues. The addition of notched noise did not significantly affect thresholds for any of the conditions involving spectral smearing, probably because the spectral smearing eliminated the sharp spectral edges of the unsmeared stimuli, so that the high-frequency cues were not available. The similarity of results for the SRt and STRt when notched noise was present suggests that similar cues were used to perform the two tests, for example ripples in the excitation patterns evoked by the stimuli; see below for further analyses based on excitation patterns. Although the STRt could in principle be performed by monitoring fluctuations in the output of a single auditory filter, the similarity of the results for the SRt and STRt when notched noise was present suggests that the test was not performed in this way.

It should be noted that although spectral smearing was applied to the stimuli for the spectral-ripple tests, no such smearing was applied to the notched noise. Furthermore, all the participants had normal hearing. When spectral ripple tests are used for listeners with reduced spectral resolution, the addition of notched noise might have a greater deleterious effect than found here. Thus, while the addition of notched noise has the desirable effect of eliminating confounding cues, giving a more valid measure of spectral resolution, it potentially has the undesirable effect of impairing spectral ripple discrimination by producing some masking of the components of the test and reference stimuli.

An alternative approach to reducing cues associated with sharp spectral edges is to avoid spectral edges by applying a spectral window to the test and reference stimuli. For example, Supin and co-workers (Supin *et al.*, 1994, 1998) have applied a Gaussian-shaped spectral envelope to rippled-spectrum stimuli. A potential disadvantage of this approach is that spectral ripples within the stimulus passband may produce “combination ripples” (like combination tones, especially 2f1−f2) outside the passband, and these combination ripples may be more resolvable than the ripples within the passband. For example, if peaks in the

rippled spectrum occur at 2 and 2.3 kHz (corresponding to a ripple density of 5 RPO), combination ripples would occur at 1.4 and 1.7 kHz (corresponding to a more resolvable ripple density of 3.6 RPO). These combination ripples can be masked with a lowpass noise or notched noise, but again with the potential disadvantage that the added noise has a masking effect on the ripples within the passband.

The STRt was introduced as a modification of the SRt to prevent participants using extraneous cues that might be available with the SRt stimuli. Such cues include differences in overall loudness in specific frequency regions, differences in spectral centroid, and changing energy at the spectral edges of the stimuli. However, for our results, thresholds were never poorer for the STRt than for the SRt. This suggests either that the extraneous cues were not used in the SRt or that the temporal modulation in the STRt stimuli did not prevent the use of such cues. However, the extraneous cues in the SRt might be used by cochlear implantees (Azadpour and McKay, 2012), for whom the ripple density at threshold is much lower than for normal-hearing listeners.

IV. PREDICTIONS USING AN EXCITATION-PATTERN MODEL

In this section, we assess the accuracy with which the results of experiment 2 can be explained using an excitation-pattern model. The excitation pattern of a sound is defined as the output of each auditory filter as a function of the center frequency of the filter (Moore and Glasberg, 1983). We tested the hypothesis that the threshold corresponds to a fixed peak-to-valley ratio, PVR (in dB), of the ripples in the excitation pattern, as proposed by previous researchers for similar stimuli (Green, 1988; Nechaev and Supin, 2013). Good accuracy of the predictions would support the use of the SRt and STRt as measures of spectral resolution. For a given ripple density, the spectral ripples were the same for the SRt and STRt stimuli, except that the ripples were static for the former and dynamically changing for the latter. The PVR of the ripples in the excitation pattern should therefore be the same for the two stimulus types, and so thresholds predicted on the basis of our hypothesis should be the same for the

two cases. Therefore, in what follows, excitation patterns were calculated only for the SRt stimuli.

Excitation patterns were calculated for 500 ms segments at the temporal center of the stimuli. The segments were hamming windowed prior to calculation of the Welch power spectra. The excitation patterns were calculated from the spectra using the Loudness toolbox implemented in MATLAB (Genesis, 2010), which is based on the procedure described by (Moore *et al.*, 1997). The transfer function of the TDH-39 headphones was taken into account in the calculations. For each stimulus, the excitation pattern was calculated for five different samples, and the resulting patterns were averaged. This was done to reduce small random fluctuations caused by the random phases of the frequency components in the carrier.

The first step in the modelling was to estimate the value of the PVR for each condition. This was done as follows:

1) For each amount of smearing, an excitation pattern was calculated for a SRt stimulus with ripple density corresponding to the mean measured threshold for that amount of smearing. Examples of the excitation patterns are shown in Figure 5.

2) For each calculated excitation pattern, the region in the pattern where the ripples were largest was determined and the PVR of the ripples in that region was estimated.

3) The estimates of the PVR determined in (2) were similar across unsmeared and smeared conditions, which is consistent with our hypothesis. An average PVR was estimated across conditions. This gave the estimated threshold criterion, i.e. the PVR in the excitation pattern required for threshold. The resulting value was 1.36 dB. This value is comparable to the values used in other comparable simulations (Green, 1988; Nechaev and Supin, 2013).

To further test the hypothesis, we predicted ripple thresholds as a function of amount of smearing. If the hypothesis is accurate, the predicted thresholds should be close to the obtained thresholds. To generate these predictions, the following steps were performed:

1) For each amount of smearing, excitation patterns were calculated for stimuli with ripple density varying from 0.5 to 7 RPO in 0.5-RPO steps.

2) The maximum ripple depth in each excitation pattern determined in (1) was estimated.

3) The ripple density leading to a peak-to-valley ratio in the excitation pattern of 1.36 dB was

estimated by interpolation for each amount of smearing. The ripple density estimated in this way is the predicted ripple threshold for that amount of smearing.

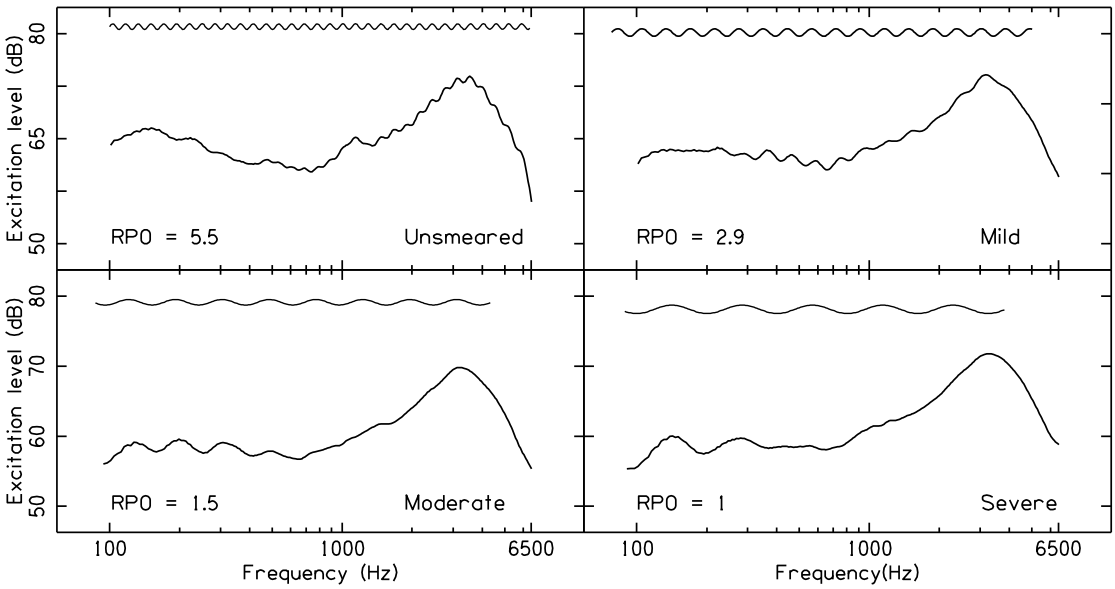

FIG. 5. *Examples of excitation patterns for stimuli with different degrees of spectral smearing. For each degree of smearing, the ripple density corresponded to the mean measured threshold. In each panel, a sine wave with frequency equal to the ripple density of the stimulus is plotted above the excitation pattern for visual guidance. The phase of the sine wave relative to the ripple in the excitation pattern is arbitrary.*

The thresholds predicted using the model are plotted in Fig. 4 as filled symbols. The predicted thresholds are similar to the experimentally obtained data for experiment 2. This supports the idea that both the STRt and the SRt assess the spectral resolution of the auditory system. However, this does not rule out the possibility that temporal cues are used in the performance of both tasks. The use of temporal cues has been proposed to underlie the detection and discrimination of stimuli with spectral ripples that are uniformly spaced on a linear frequency scale, such as iterated rippled noise (Yost *et al*. 1996). It is not known whether such cues also play a role in the detection of spectral ripples that are uniformly spaced on a logarithmic scale, as used here and in other studies (Supin *et al*., 1994, 1998).

In Fig. 5, the ripples in the excitation pattern are most prominent for the unsmeared stimuli over the frequency range from about 800 to 6000 Hz. This occurs because the relative bandwidths of the auditory filters (ERB_N/center frequency) are smallest at medium and high frequencies. However, for the stimuli with mild, moderate and severe amounts of spectral

508 smearing, the ripples are most pronounced over the ranges 300 to 2000 Hz, 100 to 500 Hz,
509 and 100 to 300 Hz, respectively. These changes with degree of smearing probably reflect side
510 effects of the way that the spectral-smearing algorithm works. The algorithm preserves the
511 phases of the components in the spectrum. As a result, when the output signal is constructed
512 from overlapping frames, the spectral structure of the original signal is partly reconstructed
513 (Baer and Moore, 1993). The effects of this reconstruction are most apparent at low and
514 medium frequencies.

515 The excitation patterns in Fig. 5 illustrate a drawback with the SRt and STRt. In
516 principle, the tests can be performed if the ripples in the excitation pattern are detectable in
517 any small frequency region. The stimuli are broadband, but there is no straightforward way of
518 assessing what frequency region is used to perform the tasks or how wide that frequency
519 region is. Furthermore, the frequency region can shift depending on the spectral resolution of
520 the participant and the way that spectral resolution varies with frequency. Researchers who
521 plan to use the SRt and STRt should bear in mind that the task does not provide a global
522 measure of frequency resolution; rather it provides a measure of the best frequency resolution
523 available to a participant for any center frequency within the passband of the stimuli.
524

525 **V. SUMMARY AND CONCLUSIONS**

526 1. As expected, simulation of one effect of cochlear hearing loss by spectral smearing
527 significantly decreased (impaired) thresholds for both the SRt and the STRt, confirming that
528 the tests are sensitive to the effects of reduced spectral resolution. Measured thresholds were
529 similar for the two tests, except when no smearing was applied. In the latter case, thresholds
530 were higher (better) for the STRt than for the SRt.

531 2. The higher threshold for the STRt in the unsmeared condition could be explained by
532 amplitude modulation cues at the outputs of auditory filters centered at the upper edge of or
533 above the frequency range of the STRt stimulus. Adding notched noise eliminated the
534 confounding cues in the unsmeared STRt stimuli and led to similar thresholds for the SRt and

the STRt. This suggests that, for normal-hearing participants, similar cues are used to perform the SRt and the STRt when confounding cues are eliminated.

3. An excitation-pattern model based on the hypothesis that the threshold in both the SRt and the STRt corresponds to a fixed peak-to-valley ratio in the excitation pattern evoked by the stimuli gave reasonably accurate predictions of the way that thresholds changed across smearing conditions for the data obtained in the presence of notched noise. This is consistent with the idea that both tasks were performed primarily using spectral cues. However, the possibility that temporal cues were used cannot be ruled out.

The experimental data and the predictions of the model indicate that both the SRt and the STRt can be employed to estimate spectral resolution. However, when STRt stimuli are used to test participants with normal hearing, notched noise should be added to eliminate cues arising in frequency regions outside the passband of the stimuli. The addition of notched noise may also be advisable when testing participants with hearing loss, although in this case the notched noise may have an effect on the detection of ripples within the passband of the stimuli. The optimal level of the notched noise remains to be determined. Researchers who plan to use the SRt and STRt should bear in mind that the task provides a measure of the best frequency resolution available to a participant for any center frequency within the passband of the stimuli.

Acknowledgements

The authors acknowledge with gratitude Prof. S. R. Savithri, Director, All India Institute of Speech and Hearing, Mysore for permitting this study to be conducted at the institute. We extend our sincere thanks to Prof. Rand Wilcox for providing guidance on statistical analysis of the data. We also extend our thanks to Dr. Justin Aronoff and Dr.Tom Bear for their technical support and stimulus generation. The authors would also like to

acknowledge the participants for their co-operation and to thank Virginia Richards and two reviewers for helpful comments on an earlier version of this paper.

Footnotes

1. Equation 1 in Aronoff and Landsberger (2013) does not include the step of taking the antilog. However, the Matlab code provided by Justin Aronoff indicates that the antilog was taken. Also, their equation 1 indicates that scaling by the ripple depth *d* (their term "*D*") was applied after 1 had been added to the absolute value of the sinusoid, whereas the Matlab code indicates that 1 was subtracted from the absolute value of the sinusoid.

References

Anderson, E. S., Oxenham, A. J., Nelson, P. B., and Nelson, D. A. (**2012**). "Assessing the role of spectral and intensity cues in spectral ripple detection and discrimination in cochlear-implant users," J. Acoust. Soc. Am., **132**, 3925–3934.

Aronoff, J. M., and Landsberger, D. M. (**2013**). "The development of a modified spectral ripple test," J. Acoust. Soc. Am., **134**, 217–222. doi:10.1121/1.4813802

Azadpour, M., and McKay, C. M. (**2012**). "A psychophysical method for measuring spatial resolution in cochlear implants," J. Assoc. Res. Otolaryngol., **13**, 145–57. doi:10.1007/s10162-011-0294-z

Baer, T., and Moore, B. C. J. (**1993**). "Effects of spectral smearing on the intelligibility of sentences in noise," J. Acoust. Soc. Am., **94**, 1229. doi:10.1121/1.408176

Carhart, R., and Jerger, J. F. (**1959**). "Preferred method for clinical determination of pure-tone thresholds," J. Speech Hear. Disord., **24**, 330–345.

Friesen, L., Shannon, R., 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.

Genesis, S. . (**2010**). *Loundness Toolbox*, http://www.genesis.fr, Available: http://www.genesis.fr. Retrieved from http://www.genesis.fr

Gilbert, G., and Lorenzi, C. (**2006**). "The ability of listeners to use recovered envelope cues from speech fine structure," J. Acoust. Soc. Am., **119**, 2438–2444. doi:10.1121/1.2173522

591 Glasberg, B. R., and Moore, B. C. . (**1990**). "Derivation of auditory filter shapes from
592 notched-noise data," Hear. Res., **47**, 103–138. doi:10.1016/0378-5955(90)90170-T
593 Green, D. . (**1988**). *Profile Analysis*, Oxford University Press, Oxford, 1-144 pages.
594 Henry, B. A., Turner, C. W., and Behrens, A. (**2005**). "Spectral peak resolution and speech
595 recognition in quiet: normal hearing, hearing impaired, and cochlear implant listeners,"
596 J. Acoust. Soc. Am., **118**, 1111–1121. doi:10.1121/1.1944567
597 Jones, G. L., Won, J. H., Drennan, W. R., and Rubinstein, J. T. (**2013**). "Relationship
598 between channel interaction and spectral-ripple discrimination in cochlear implant
599 users," J. Acoust. Soc. Am., **133**, 425–33. doi:10.1121/1.4768881
600 Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am.,
601 **49**, 467–477.
602 Litvak, L. M., Spahr, A. J., Saoji, A. A., and Fridman, G. Y. (**2007**). "Relationship between
603 perception of spectral ripple and speech recognition in cochlear implant and vocoder
604 listeners," J. Acoust. Soc. Am., **122**, 982–91. doi:10.1121/1.2749413
605 Moore, B. C. J., and Glasberg, B. R. (**1983**). "Suggested formulae for calculating auditory-
606 filter bandwidths and excitation patterns," J. Acoust. Soc. Am., **74**, 750–753.
607 Moore, B. C. J., Glasberg, B. R., and Baer, T. (**1997**). "A model for the prediction of
608 thresholds, loudness, and partial loudness," J. Audio Eng. Soc, **45**, 224–240. Retrieved
609 from http://www.aes.org/e-lib/browse.cfm?elib=10272
610 Moore, B. C. J., Glasberg, B. R., and Simpson, A. (**1992**). "Evaluation of a method of
611 simulating reduced frequency selectivity," J. Acoust. Soc. Am., **91**, 3402–3423.
612 doi:10.1121/1.402830
613 Nechaev, D. I., and Supin, A. Y. (**2013**). "Hearing sensitivity to shifts of rippled-spectrum
614 patterns," J. Acoust. Soc. Am., **134**, 2913–22. doi:10.1121/1.4820789
615 Patterson, R. D., Allerhand, M., and Giguère (**1995**). "Time-domain modeling of peripheral
616 auditory processing: a modular architecture and a software platform," J. Acoust. Soc.
617 Am., **98**, 1890–1984.
618 Patterson, R. D., and Nimmo-Smith, I. (**1980**). "Off-frequency listening and auditory-filter
619 asymmetry," J. Acoust. Soc. Am., **67**, 229–245. doi:10.1121/1.383732
620 Peter, V., Wong, K., Narne, V. K., Sharma, M., Purdy, S. C., and McMahon, C. (**2014**).
621 "Assessing spectral and temporal processing in children and adults using temporal
622 modulation transfer function (TMTF), iterated ripple noise (IRN) perception, and
623 spectral ripple discrimination (SRD)," J. Am. Acad. Audiol., **25**, 210–218.
624 doi:10.3766/jaaa.25.2.9

Rom, D. M. (**1990**). "A sequentially rejective test procedure based on a modified Bonferroni inequality," Biometrika, **77**, 663–666.

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**, 955–8. doi:10.1121/1.3179670

Shannon, R., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science (80-.)., **270**, 303–304.

Stone, M. A., Füllgrabe, C., and Moore, B. C. J. (**2008**). "Benefit of high-rate envelope cues in vocoder processing: effect of number of channels and spectral region," J. Acoust. Soc. Am., **124**, 2272–82. doi:10.1121/1.2968678

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. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7961175

Supin, A. Y., Popov, V. V, Milekhina, O. N., and Tarakanov, M. B. (**1998**). "Ripple density resolution for various rippled-noise patterns," J. Acoust. Soc. Am., **103**, 2042–2050. doi:10.1121/1.421351

Wilcox, R. (**2012**). *Modern statistics for the social and behavioral sciences A practical introduction*, CRC Press, New York, 1st ed., 1-22 pages.

Wilcox, R. (**2012**). *Introduction to robust estimation and hypothesis testing*, Academic Press, Oxford, UK, Third Edit., 2-17 pages. doi:10.1016/B978-0-12-386983-8.00005-6

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., **392**, 384–392. doi:10.1007/s10162-007-0085-8

Yost, W. A., Patterson, R. D., and Sheft, S. (**1996**). "A time domain description for the pitch strength of iterated rippled noise," J. Acoust. Soc. Am., **99**, 1066–1078.