

# Distortion of Spectral Ripples Through Cochlear Implants Has Major Implications for Interpreting Performance Scores

Matthew B. Winn<sup>1</sup> and Gabrielle O'Brien<sup>2</sup>

The spectral ripple discrimination task is a psychophysical measure that has been found to correlate with speech recognition in listeners with cochlear implants (CIs). However, at ripple densities above a critical value (around 2 RPO, but device-specific), the sparse spectral sampling of CI processors results in stimulus distortions resulting in aliasing and unintended changes in modulation depth. As a result, spectral ripple thresholds above a certain number are not ordered monotonically along the RPO dimension and thus cannot be considered better or worse spectral resolution than each other, thus undermining correlation measurements. These stimulus distortions are not remediated by changing stimulus phase, indicating these issues cannot be solved by spectrotemporally modulated stimuli. Speech generally has very low-density spectral modulations, leading to questions about the mechanism of correlation between high ripple thresholds and speech recognition. Existing data showing correlations between ripple discrimination and speech recognition include many observations above the aliasing limit. These scores should be treated with caution, and experimenters could benefit by prospectively considering the limitations of the spectral ripple test.

**Key words:** Cochlear implants, Spectral ripple, Spectral resolution.

(Ear & Hearing XXX:XX:00–00)



This article has received an OSF badge for Open Materials.

## INTRODUCTION

A major focus in cochlear implant (CI) research is the evaluation of spectral resolution, which underlies a listener's ability to perceptually distinguish differences in frequency patterns (or the spectral envelope) in sounds. Errors in word and phoneme recognition by CI listeners are driven heavily by impaired differentiation of speech sounds that rely on spectral contrasts (e.g., consonant place of articulation, c.f. Munson et al. 2003), suggesting spectral resolution is a chief limiting factor of the device.

The spectral ripple test is a popular psychophysical method aimed at quantifying spectral resolution in individual CI listeners. Although there are various testing paradigms, ripple stimuli are typically presented in a discrimination task, where listeners distinguish between two sounds whose spectral envelopes are sinusoidally modulated at equal modulation depths, but with inverted phases (i.e., the spectral peaks are alternated with spectral valleys). The number of spectral peaks within a set frequency range—and hence the spectral *density*—is increased

until the listener can no longer discriminate between the phase inversions. So long as the frequency difference between spectral peaks exceeds the bandwidth of the narrowest auditory filter, the two stimuli should be discriminable by the listener, without time-consuming conventional tuning-curve methods. In the case of a CI listener, the auditory filters at the mechanical level of the cochlea are bypassed, but the frequency selectivity of the spiral ganglion activation is the target of the evaluation.

One reason for the spectral ripple test's popularity is that thresholds can be obtained relatively quickly, and have been demonstrated to correlate with speech perception outcomes, including word, consonant, and vowel recognition (Henry et al. 2005), and speech recognition in noise (Won et al. 2007; Lawler et al. 2017), although not all studies observe robust correlations with speech perception (Anderson et al. 2011). The ripple test appears to be sensitive to physical and psychophysiological properties of the implant, including differences between processing strategies (Drennan et al. 2010; Zhou 2017), and indices of electrical current spread or channel interaction (Jones et al. 2013; Won et al. 2014; Scheperle & Abbas 2015). Enthusiasm for the spectral ripple test is therefore understandable from clinical and basic science perspectives.

## Concerns About the Spectral Ripple Test

Previous studies have raised methodological concerns about the spectral ripple test in CIs. Azadpour and McKay (2012) suggest that listeners could attend to differences in loudness, spectral centroid, and changes to the spectral edges rather than differences in broadband spectral density. Aronoff and Landsberger (2013) proposed a temporally dynamic spectral ripple test (SMRT) that addressed some of these issues by introducing drifting phase of the spectral modulation which results in fixed-rate amplitude modulations. It has been observed that perception of dynamic spectral ripples could be driven by amplitude modulations rather than spectral resolution (Lawler et al. 2017). This issue is particularly relevant at spectral edges, where the valleys of the amplitude envelope would be less likely to be filled by activation from two neighboring electrodes. Modulations at the edges lack interfering spread of activation from at least one neighboring electrode, and therefore should contain deeper more perceptible modulations that do not require the same spectral resolution that would be required in the middle of the array. To address this problem, Archer-Boyd et al. (2018) proposed a further modification of ripple stimuli (called STRIPES) that neutralizes spectral edge cues by probing for perception of the *direction* of spectral drift in stimuli where temporal modulations have equal modulation rates at the edges.

Despite the improvements in ripple testing methods offered by the SMRT and STRIPES, several potentially serious concerns with spectral ripple stimuli in CI listeners remain unaddressed,

<sup>1</sup>Department of Speech-Language-Hearing Sciences, University of Minnesota, Minnesota, USA; and <sup>2</sup>School of Information, University of Michigan, Ann Arbor, Michigan, USA.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and text of this article on the journal's Web site ([www.ear-hearing.com](http://www.ear-hearing.com)).

which are the subject of the current perspective. The central issue described in this paper is that the spectral density transmissible by a CI is necessarily limited by the number of frequency channels and the bandwidth of each channel. Stimuli that exceed this limitation are transformed in a way that complicates comparison of individual ripple scores and challenges correlations between ripple scores and other test results. At the very least, the expression of a spectral ripple score as representing spectral resolution is a problematic position.

The limitation of spectral sampling in a CI is akin to the Shannon–Nyquist sampling theorem, but in the spectral domain rather than the more-familiar temporal domain. For a rippled spectrum of  $N$  ripples per octave, there must be at least  $N \times 2$  frequency channels for each octave to encode the spectral envelope (just as there must be a time sampling rate of  $f \times 2$  to encode frequency  $f$ ). When the spectrum is undersampled, the signal will be aliased, meaning novel spectral components will be introduced which were not in the original signal. The novel frequencies should in principle differ from the maximum possible encodable frequency by the same amount as the intended frequency, but in the opposite direction. However, calculating the maximum sampling density of a CI is not straightforward, because frequency sampling in a CI is not equally spaced on a logarithmic (octave) scale, and because the spacing between electrodes within the cochlea is not guaranteed to subtend equal octave spacing across characteristic frequencies. As a result, the transformation of the spectral ripple stimulus in a CI is not uniform, and is not structured in a principled way; it is therefore more generally a distortion rather than pure aliasing.

The importance of frequency sampling for spectral ripples has been acknowledged both implicitly and explicitly in previous studies. The requirement for a large number of spectral components is built into the stimulus—Won et al. (2007) created ripples using 200 frequency components between 100 and 5000 Hz and the SMRT uses 202 pure tones between 100 and 6400 Hz (Aronoff & Landsberger 2013). However, this need for dense frequency sampling is commonly overlooked when sending the stimuli through CI processors. The CI processor has basic constraints on the range of possible stimuli that can be presented, as it is an immutable part of the stimulus delivery chain. For those interested in mapping perceptual results to spectral resolution specifically, results obtained from stimuli that extend beyond the device's capacity should be treated with caution and possibly treated as qualitatively different than stimuli that were more faithfully transmitted.

The problem of spectral undersampling/aliasing has been described by Anderson et al. (2011), Gifford et al. (2018), and O'Neill et al. (2019). Anderson et al. explained that full reconstruction of each ripple in the stimulus spectrum is not necessary for discrimination of ripple stimuli; any spectral difference could theoretically be used, so long as it is perceptible. This description is both true and problematic; the stimulus could have numerous properties other than the intended spectral density, so successful auditory discrimination cannot be directly linked to any particular property such ripple-per-octave (RPO) density. Furthermore, if only one part of the spectrum is used to discriminate ripple stimuli, the value of the test as a probe of broadband spectral resolution is diminished. Anderson et al. (2012) speculated that when listeners encounter spectral ripples that are more dense than the estimated capability of the CI processor, that they might be switching to a different perceptual regime altogether, highlighting how it could be misleading to describe CI ripple stimuli as differing by a singular underlying factor.

The specific focus on the electrode activation pattern can seem like a paralyzing and unreasonable demand when generalizing beyond spectral ripple stimuli. Other signals such as speech and environmental sounds are more conveniently described by their acoustic structure rather than their transformed electrical representation. However, spectral ripple stimuli are treated as controlled psychoacoustic stimuli with experimenter-defined parameters that are used to make correlations and judgments about the resolution of auditory system. There is an expectation that discriminating spectral ripple stimuli does not simply imply that some stimulus difference was perceptible (e.g., differentiating a “b” and a “d”), but rather that a specific attribute of the stimulus was perceptible (“the listener can resolve four ripples per octave”). This article examines whether that expectation can be satisfied. If some spectral ripple stimuli are transformed in a disorderly fashion, then corresponding scores cannot be linked to the experimenter-controlled stimulus parameters, and therefore not ordered or compared along a single dimension. In the Method section, we describe the distortions to spectral ripple stimuli that highlight the need to be extra cautious when handling stimuli that exceed the maximum spectral density that can be supported by a CI.

## MATERIALS AND METHODS

To investigate the problem of spectral ripple distortion, a series of spectral ripple inputs were synthesized in the frequency domain, which had a standard modulation depth of 30 dB, and which varied in incremental steps of spectral density expressed as ripples per octave. The spectra were entirely synthetic, with 8192 frequency samples that were logarithmically spaced between  $2^5$  and  $2^{15}$ ; these frequency ranges extend past the limits of CI frequency analysis but include padding that results in extra precision in measuring spectral density (*a la* padding in a fast Fourier transform). At each frequency sample, the spectrum power was defined using the following formula:

$$(1 - (\text{mod\_depth} / 2)) + \left( \sin \left( \left( \left( \left( \log_2(\text{frequency}) \right) - k \right) \times 2 \times \pi \times \text{RPO} \right) + \right) \right) \times (0.5 \times \text{mod\_depth})$$

where modulation depth is expressed in dB (we used 30 dB as a default), RPO is ripples per octave, and  $k$  is an optional constant phase shift of  $\log_2(10)$  that resulted in perfect alignment of spectral peaks with octave frequencies displayed on the chart, purely to visually confirm intended spectral density of integer levels of RPO.

For simplicity, the spectra were analyzed through a bank of idealized (i.e., infinitely steep “brick wall”) bandpass filters matched to the default channel-frequency allocations in the fitting software for the devices made by the Cochlear Corporation (Sydney, Australia). In a real device, there would be additional processing steps following the initial spectral analysis that add complexity to the output, such as pre-emphasis, dynamic range optimization, and various noise-reduction algorithms. Understanding of all of these processing steps would require electrodiagrams. We chose a simplified and idealized approach to spectral analysis here so that it was focused on the analysis

of frequency bands in a way that was not bound to the particular details of any manufacturer's processing approach, and which is agnostic to the definition of the frequency domain in an electrodiagram.

The output of the idealized filters was further simulated to model the pattern of activation in the cochlea, with interactions between electrodes using 2 dB/mm roll-off inspired by the average value of studies by Nelson et al. (2008, estimating 1.2 dB/mm) and Bingabr et al. (2008, estimating 2.8 dB/mm). This simulation does not model the exponential loudness function that would result from electric hearing, which would depend heavily on the device mapping and individual's dynamic range. Analyses for other devices are available in Supplemental Digital Content 1, <http://links.lww.com/EANDH/A968>. The mapping of frequencies to cochlear space was estimated using a model of the basilar membrane (Greenwood 1990) for simplicity. A CI would directly stimulate the spiral ganglion, which has an appreciable difference in the frequency map (Stakhovskaya et al. 2007; Landsberger et al. 2015), but the current analysis concerned mainly with the interfrequency intervals (i.e., spectral modulation) rather than absolute frequencies.

Figure 1 illustrates some examples of acoustic spectral ripples in the acoustic domain (row A), the corresponding discretized electrode activations (row B), cochlear simulations that impose an idealized electrode-frequency tonotopic match (row C), or simulations of the mapping of physical electrode placement along the basilar membrane (row D). Row E illustrates the output of a processor that stimulates the eight electrodes with the highest output, in a crude simulation of the advanced combination encoder peak-picking strategy used in Cochlear devices. At a low spectral density (e.g., 0.5 RPO), the processor output can theoretically represent the input spectral density with some fidelity (row B), although overlap between electrode

activations can smooth over much of the spectral valleys (rows C and D). At higher densities (e.g., above 2 RPO), the processor is unable to represent the spectral density even in an idealized form—exactly what would be expected considering the limited number of independent channels in the device. In each simulation of cochlear activation, the spectral modulation depth does not reflect the nonlinear transformation that would map acoustic levels to corresponding current levels within an individual's dynamic range of electrical stimulation.

Figure 2 shows spectral density that would result after spectral ripples are analyzed either in idealized form or through the various CI simulations described above. Spectral density was calculated using a log-spectral modulation frequency analysis, similar to a typical spectral analysis (e.g., Fast Fourier Transform), but using the log-sampled spectral domain instead of the time domain (code available in Supplemental Digital Content 2, <http://links.lww.com/EANDH/A997>).

Unsurprisingly, spectral densities that exceed the maximum density supported by the simulated CI processor result in aliased spectra. The critical value of input density is difficult to calculate in a precise manner because the spectrum is not uniformly sampled; hence it does not have a single spectral sampling rate. One could estimate the maximum spectral density supported by the narrowest channel or pair of channels, the width of the widest channel, some average across the array, or any other weighted combination of channel widths. We estimate that the highest spectral density before problematic distortion appears to be roughly 2 RPO for the Cochlear device. Even though channel number 9 in the cochlear speech processor has a narrow 0.144 octave bandwidth (theoretically supporting 3.46 ripples per octave), the principle of the ripple stimulus is to express spectral density across channels; the device does appear capable of representing any spectral density higher than

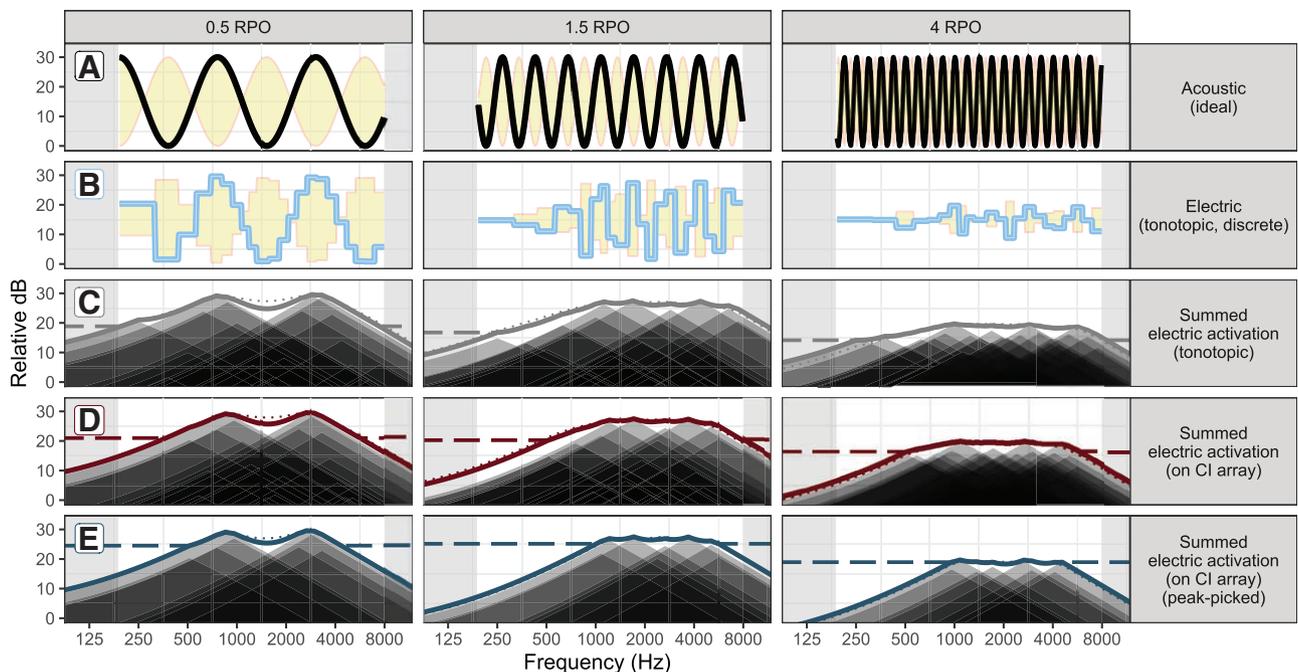


Fig. 1. Spectra of idealized acoustic spectral ripples (A), analyzed through a synthetic CI processor (B), with corresponding simulated cochlear activation patterns at ideal cochlear positions (C), or aligned with estimated electrode array placement (D). The bottom row (E) represents the activation from row D but with only top eight highest-energy channels kept. Rows A and B show inverted-phase ripples, with the difference in spectral power shaded. Straight dashed lines in rows C, D, and E reflect the lower of the two endpoint values corresponding to the extreme apical or basal electrodes.

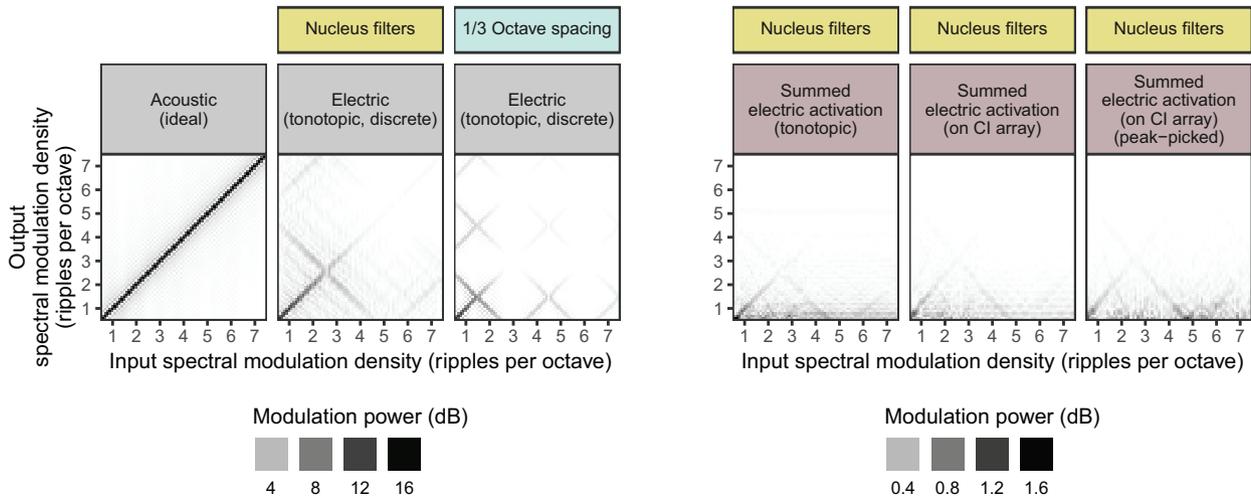


Fig. 2. Input-output functions showing the spectral density (RPO) of the output of various simulations resulting from a change in RPO of the input stimulus. The three panels on the left illustrate idealized spectra, and the panels on the right show analyses of spectra that are simulated from summed electrical activation using a constant spread of 2 dB/mm, but not rescaled for exponential growth in loudness perception. Any deviation from the straight diagonal line on the far-left panel represents distortion of the output spectrum. RPO indicates ripple-per-octave.

2 RPO on a broadband scale. Even when rotating the phase of the spectral modulation, the limit does not change in any appreciable manner, implying that temporally dynamic ripple stimuli (which use drifting phase) cannot overcome the aliasing/distortion problem.

Because of nonmonotonic changes in both output spectral density and modulation depth, spectral ripples above the critical limit are represented *differently* rather than being ordered along a single-dimension commensurate with the density of the input. In practical terms, this means that a threshold score of 5 RPO does not necessarily ensure that a listener's spectral resolution is better than that of a person who has a threshold of 3 RPO. It also means that the differences between stimuli with 3, 4, and 5 RPO are not equivalent.

To be clear, there are true spectral differences between stimuli more dense than the aliasing limit, but those differences are not related to the density of the input stimulus in any principled or ordered way. Past a certain density, ripples are not distorted in a linear well-behaved fashion. Instead, the density has disorderly peaks and valleys, including a cluster of high-density energy in low spectral densities when the RPO is set to a value just below 5 RPO. The RPO dimension is not a clear proxy for the stimulus features modulated by the experimenter in the acoustic domain. In some parameter ranges, it may be neither proportional nor monotonically related to the parameter varied by the experimenter. Linear correlations involving ripple scores demand more than discrimination of *any* unspecified spectral difference; they convey the assumption that the difference between 1 RPO and 2 RPO is the same as the difference between 3 RPO and 4 RPO, but this assumption is not met. The discrimination of a 5-RPO stimulus is better than the inability to discriminate a 5-RPO stimulus, but it does not mean that the listener's spectral resolution is better than someone who discriminated only 4 RPO, because the number does not correspond to any particular spectral dimension of the stimulus. Standard correlational metrics using RPO threshold scores are therefore highly questionable when they include data points beyond the point of aliasing,

because at least one dimension of the correlation uses a scale that is not monotonic.

As spectral ripple density increases, the spectral modulation depth of the output stimulus decreases, because spectral peaks and valleys will average within the same filter, neutralizing cross-channel differences. At a glance, this seems desirable, since any test of spectral resolution should be more difficult or impossible when spectral components are less resolvable. However, it is not possible to interpret the results specifically as the highest density at which the listener's auditory system fails to resolve the inputs, or whether we are finding the density at which the *device* is simply failing to deliver the intended stimulus. At stake is whether the experimenter has control over the relevant stimulus parameters such as ripple density and depth when stimuli are sent through a CI processor. Alternatively, the issue could be considered a matter of communicating the stimulus attribute that is thought to govern the perceptual responses; it is understandably more attractive to express the results as spectral resolution since that is a crucial issue in CI research, but the problem is that it is unknown whether performance is driven by resolution in the spectral or amplitude domains, since they are conflated.

The effect of spectral density on stimulus modulation depth is nonmonotonic, meaning there is not a correction factor that scales neatly with the input. Some channels show maximum-depth power changes at 3 RPO that are weaker than those at 4.5 RPO. Anderson et al. (2012) found that there was not an interpretable relationship between spectral ripple discrimination threshold and the ripple density at which a 30-dB modulation was detectable (with detection densities exceeding 15 RPO in multiple participants). Since modulation clearly has a substantial effect on ripple perception, and since modulation depth is partly outside the experimenter's control, this factor presents a significant challenge to the goal of interpreting results as measuring spectral resolution.

If a CI processor had perfectly even-spaced sampling of the frequency spectrum, then the spectrum distortions described

here would not be fully alleviated, but would become orderly. A simulation of such a filter design was implemented using 1/3-octave filters with 15 center frequencies spanning between 250 and 6349.4 Hz. Figure 3 shows a series of increasingly dense ripple stimuli and their corresponding filter activation strengths. As expected, the 3-RPO stimulus has equal strength in each 1/3-octave filter, and ripples with density greater than 3 RPO yield a weak and nonmonotonic pattern of filter activation.

### Interpreting Results From Previous Literature

The technical issues raised here lead to the question of how to interpret previous studies that found correlations between spectral ripple discrimination and speech recognition in CI listeners. Jones et al. (2013) argued that even if there are contaminating factors affecting spectral ripples through CIs, which the history of positive correlations with speech perception justifies the use of the ripple test. We acknowledge that not every tester necessarily needs to know *why* a person succeeds in the ripple test, as long as it gives useful information. However, for researchers investigating mechanistic hypotheses about the source and implications of individual differences in spectral resolution in cochlear implant outcomes, and for those researchers interested in spectral resolution specifically, the issues outlined in previous sections in this article pose a considerable challenge. At the very least, it appears that the characterization of better-performing listeners on the spectral ripple test as having “better spectral resolution” is up for debate.

In the published literature, spectral ripple thresholds above the aliasing limit are common occurrences. We reviewed 34 articles (Supplemental Digital Content 3, <http://links.lww.com/EANDH/A970>) that tested discrimination or either static ripples (22 studies) or spectrotemporally modulated ripples (i.e., SMRT; 13 studies, including one that also used static ripples), and which illustrated individual data. Figure 4 shows the trend revealed by this literature review. Among thresholds for discriminating static ripples, 40% thresholds were above 2 RPO, and 30% were above 2.5 RPO. Static ripple thresholds above 4 and 5 RPO, represented 12% and 8% of the data, respectively. For SMRT, these numbers were substantially higher, with 73% of scores above 2, and 62% of scores above 2.5 RPO. SMRT thresholds above 4 and 5 RPO, represented 38% and 19% of the data respectively. Given these degrees of prevalence, ripple thresholds that exceed the actual capacity of CI devices clearly have a substantial impact on the conclusions of these studies. It is not possible to know the spectral resolution of

listeners who achieve these scores, and whether they are ordered in any meaningful way. In some data sets, correlations with speech perception tasks appear to be heavily influenced or leveraged by outliers with unusually high ripple thresholds (e.g., data published by Won et al. 2007 and 2010). Therefore, the technical issues highlighted here cannot be dismissed as inconsequential.

Dynamic ripple tests such as the SMRT consistently show higher (better) thresholds than static ripple tests. This observation is the opposite of what one might predict if the extra features of the SMRT successfully eliminated spurious cues that are unrelated to spectral resolution, such as spectral centroid or amplitude cues at edge frequencies that might lead to misleadingly good performance in static ripple tests (articulated by Aronoff & Landsberger 2013). After those cues are neutralized, one would expect poorer thresholds, since performance would no longer be artificially inflated by perception of those cues. Instead, RPO thresholds for SMRT stimuli are consistently better than those for static ripple stimuli. This is likely because spectrotemporally modulated stimuli can potentially be discriminated by amplitude modulations, as postulated by Lawler et al. (2017), who point out that these modulations will be affected by spectral resolution (as poorer resolution would fill in the amplitude valleys, rendering discrimination more difficult). Consistent with that notion, Zhou et al. (2020) found that scores for SMRT had no correlation with static ripple scores ( $r^2 = 0.0009$ ), but did have a positive correlation with modulation detection ( $r^2 = 0.46$ ), suggesting that the increased success for SMRT does not reflect a more targeted probe of spectral resolution, but rather a pattern of listeners exploiting temporal cues. Thus, even though the SMRT removes edge-frequency intensity differences as a cue, it introduces edge-frequency modulations as a cue that listeners can use to feign spectral resolution. Amplitude modulations are a useful auditory property to perceive, but for the experimenter who is interested in probing for spectral resolution in particular, they present a confound. The data published by Zhou et al. (2020) suggest that the amplitude modulation cues in SMRT stimuli account for a significant contribution to the performance score. This means that the SMRT results should not be interpreted as a pure measure of spectral resolution, since the experimenter is unable to determine whether spectral or temporal cues were used to perform the discrimination task. Expressing the threshold solely in terms of spectral ripples per octave is therefore potentially misleading, given Zhou et al. (2020) findings. Regardless of the perceptual cue used to successfully discriminate dynamic ripples,

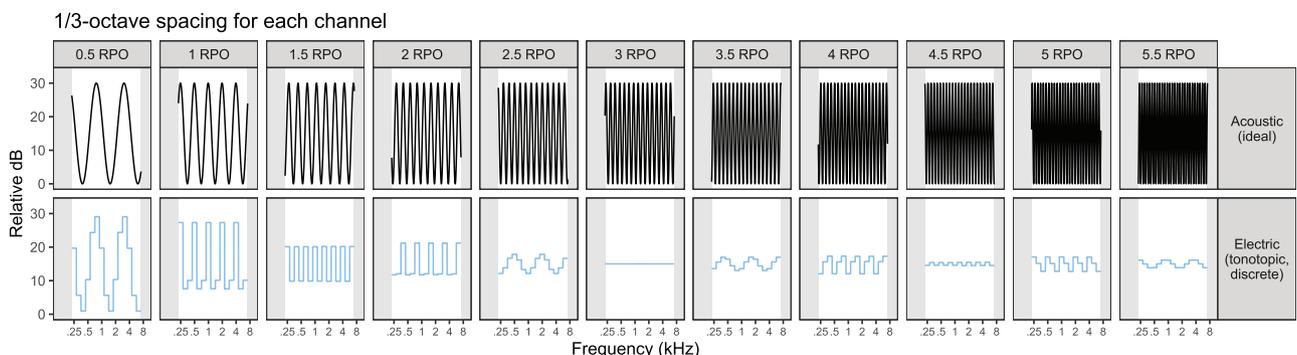


Fig. 3. Same as Figure 1A and B, except the spectrum is filtered into bands that subtend exactly 1/3 octave. The simulated pattern of electrical activity shows an orderly but not monotonic output based on changes in the acoustic RPO density. RPO indicates ripple-per-octave.

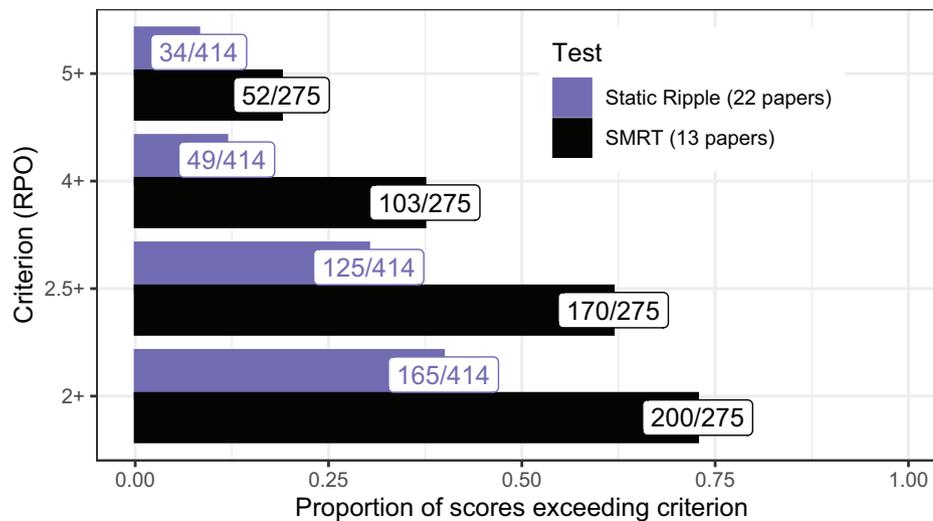


Fig. 4. Proportion of individual spectral ripple scores exceeding various criteria, based on results of 34 studies with published individual data or data shared with the current authors by the original experimenters. Raw numbers reflect the number of individual observations that exceed each criterion indicated on the y axis. SMRT refers to scores obtained with the spectrotemporally modulated ripple test (Aronoff & Landsberger 2013) and static ripples refer to discrimination of phase-inverted ripples with static spectra.

the modulation depth/nonmonotonicity problem remains unaddressed even in these dynamic stimuli because there is no phase that would support the experimenter's intended spectral modulation depth or spectral density beyond the aliasing limit.

In the process of reinterpreting previous results, we considered what conclusions would change as a result of avoiding spectral ripple stimulus aliasing. When experimenters have limited their stimuli to spectral densities below 2, reported correlations with speech perception are *stronger* than when such a limit is not imposed (Henry et al. 2005; Litvak et al. 2007; Saoji et al. 2009), motivating a new retrospective analysis. We reanalyzed data from several published studies while excluding scores above a critical value. There is no singular aliasing limit, even when limiting analysis to one CI processor (which we did not do), since the frequency sampling along the array is not uniform. But it appears from the simulations presented here that the aliasing limit is somewhere close to 2 RPO. Different studies have different levels of granularity for constructing ripple stimuli and also for estimating thresholds. We initially aimed for 2.5 RPO as a generous upper limit of possible spectral density transmitted through the CI processor. However, finding that some studies used RPO values with variable precision, 2.56 RPO was arbitrarily chosen to be inclusive of many scores reported in studies by Won et al. and Anderson et al. (which include that specific number), as well as being in the middle range of the thresholds reported by Jeon et al. (2015) and Henry et al. (2005) after averaging over psychometric tracking reversals above and below the estimated thresholds.

Correlation of ripple thresholds with word recognition reported by Winn et al. (2016) grew stronger when high ripple scores were omitted ( $r^2$  grew from 0.45 to 0.59). Similar trimmed correlations between ripple scores and speech formant perception also grew stronger, with  $r^2$  increasing from 0.33 to 0.50. Trimmed correlations for vowel recognition (originally measured by Anderson et al. 2011) were only marginally improved, with  $r^2$  changing from 0.196 to 0.211. Interestingly, for Anderson et al.'s results for sentence recognition, which depends on more central mechanisms, omitting high ripple

scores weakened the correlation from 0.41 to 0.31, possibly suggesting that the higher ripple scores might reflect higher-level auditory skills that are relevant for sentence perception rather than lower-level peripheral auditory encoding. But importantly, the nature of the perceptual cues used in these different tasks is complex and has not been shown to be directly or monotonically linked. Illustrations of these revised patterns are available in Supplemental Digital Content 1, <http://links.lww.com/EANDH/A968>.

Lower-RPO scores are not only more easily interpretable based on the constraints of CI processing, they are also more reliable. Data illustrated in studies by Won et al. (2007), Anderson et al. (2011), Jung et al. (2012), Drennan et al. (2016), and Winn et al. (2016) show that higher RPO scores have much wider variability when averaging across multiple test runs, suggesting that a single high RPO score is not always replicable within an individual. Using data from the studies by Anderson et al. (2011) and Winn et al. (2016), Figure 5 shows thresholds for individual test runs for participants discriminating spectral ripples. The average RPO thresholds above 2.56 are substantially more variable than those below this limit, with many individual test-session values exceeding the "mean" performance by over 1 full RPO. This pattern of instability is not observed for any participant whose average RPO score was below the aliasing limit. This pattern suggests that scores above the aliasing limit are unreliable even within an individual, and possibly contaminated by spurious adaptive tracking patterns. For example, a listener might make a successful (but random) guess on a stimulus near the aliasing/distortion limit and then subsequently be presented with an aliased stimulus with spurious low-rate spectral densities that are discriminable, raising the score above 2.56. If that fortunate random guess was not made in a separate test session, then the listener is not given the opportunity to exploit aliased ripples, and is relegated to lower-RPO stimuli. It is possible that a combination of ascending tracking, descending tracking, and method of constant stimuli could clarify how often this problem occurs.

The complications that we point out in this article do not directly lead to an understanding of what drives perception of

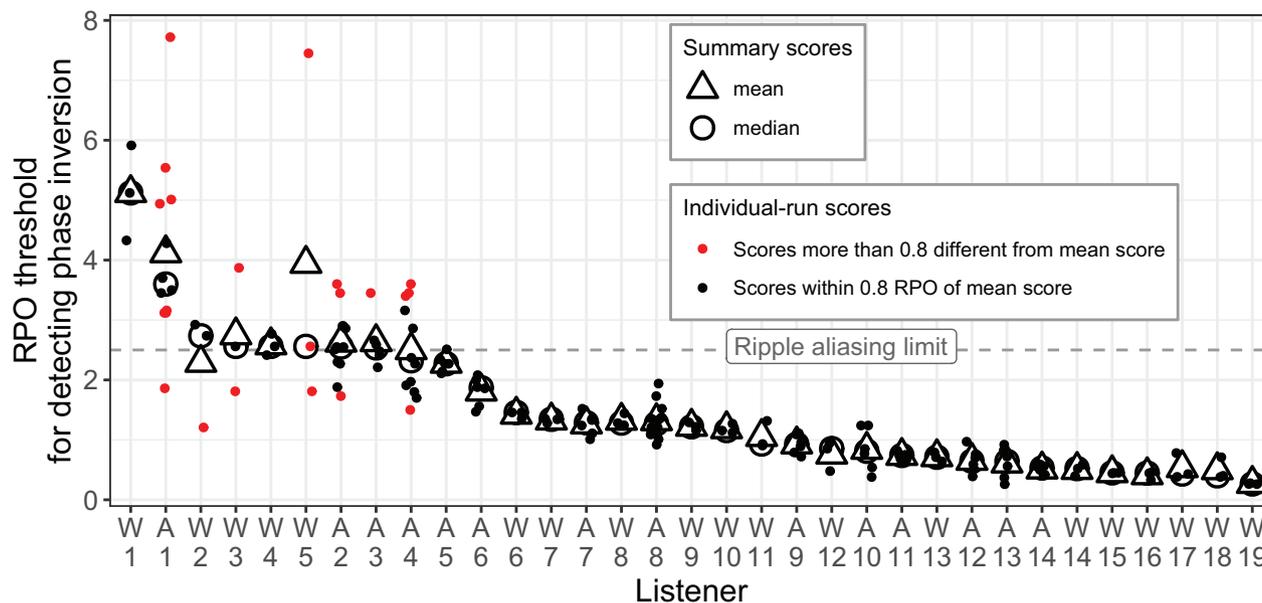


Fig. 5. Ripple-per-octave (RPO) phase inversion thresholds obtained from 33 CI listeners in previous studies by Winn et al. (2016; participants identified with “W”) and by Anderson et al. (2011; identified with “A”). Participants are ordered by median RPO threshold (open circle), with mean score displayed with an open triangle. Single test-run scores that deviated from the individual’s mean score by 0.8 RPO or more are colored red.

spectral ripples through a CI. In fact the analyses implies that one might not be able to isolate any perceptual cue at all, since multiple things are changing in unplanned and nonlinear ways. Although we are agnostic as to what cues drive performance in previous studies, the analysis presented here suggests that those cues cannot be described using the spectral dimension of RPO, particularly when the RPO is above the aliasing limit. The auditory dimensions that change with increasing input RPO, whatever they may be, appear to reliably correlate with speech recognition scores but also appear to be elusive in definition.

Previous studies that used spectral ripples to evaluate cochlear mechanisms in listeners with acoustic hearing (Narne et al. 2016; Nechaev et al. 2019; Supin et al. 2019) are not constrained by the issues raised in this article, as the problems stem mainly from the frequency sampling in a CI processor. However, those studies that focused on cochlear mechanisms shed further insight into extra cues that could be used by CI listeners to discriminate spectral ripples, so long as the transduction of spectral cues into temporal cues shares common mechanisms.

### Relating Spectral Ripple Perception to Speech Perception

Although speech involves both temporal and spectral modulations, there is virtually no theoretical framework connecting spectral ripple stimuli to speech sounds. Thresholds considered to represent better performance in spectral ripple tests are not reflective of spectral densities observed in speech sounds, and the modulation rates used in dynamic ripple stimuli do not appear to reflect corresponding modulations in speech. One possibility is that listeners who excel at discriminating the low-density, nonuniform spectral patterns produced by aliased ripple stimuli are likely to also excel at distinguishing between the low-density, nonuniform spectral patterns that characterize vowels, which generally contain less than one peak per octave (Liu & Eddins 2008; also see Supplemental Digital Content 4, <http://links.lww.com/EANDH/A971>). However, when ripple stimuli

are aliased, the nonlinear transformations are not well described mathematically, and their similarity to speech does not increase commensurate with their original prealiased spectral density. Aliased spectral ripples could therefore demand generally good listening skills to discriminate, but those skills cannot be easily mapped to specific acoustic properties of speech.

The lack of correspondence between the spectra of experimental ripple stimuli and speech sounds is not an inherent weakness of any test of spectral resolution. However, it does complicate the interpretation of why perception of spectral ripples should be an attractive proxy for speech perception abilities. The observed pattern of aliasing combined with reports of cognitive factors (Kirby et al. 2019) and the change in ripple performance with increasing exposure to test stimuli (Drennan et al. 2016; de Jong et al. 2018) imply that the performance in these various studies can be affected by the same factors that one intends to avoid by choosing ripple stimuli instead of speech (as expressed in the titles of articles by Gifford et al. 2014 “...A non-language based measure of performance outcomes” and Drennan et al. 2016 “Nonlinguistic outcome measures in adult cochlear implant users...”). If an experimenter’s goal is to design a study that is free from the influence of learning and cognitive processing, the spectral ripple discrimination does not necessarily accomplish that goal.

### Other CI Processors

Critical limits of spectral density for the Advanced Bionics and Med-El devices have not been described in this article, for simplicity. It would be reasonable to suspect that these processors, having fewer electrodes, would permit even less spectral density than the implant arrays currently manufactured by the Cochlear Corporation. However, they might offer other advantages unrelated to the analysis discussed here, and the ability to transmit spectral ripples is not the primary goal of CI processors in any case. When considering only the number of electrodes and the spacing between electrodes, the critical density for spectral aliasing in

the Advanced Bionics device (without current steering) is about 2 RPO, and about 1.1 RPO for the Med-El device. For the Cochlear device, this number is highly dependent on whether peak-picking is activated; the value is likely between 2 and 2.5. The threshold of modulation depth saturation (i.e., smallest ripple whose spectral modulation period fits into the narrowest filter) is 3.46 RPO for the Cochlear device, 2.02 RPO for the Advanced Bionics device and 1.26 RPO for the Med-El device. All of these numbers are further qualified by the amount of electrode interaction within the device and the spread of neural activation within an individual ear.

### What Should We Do Instead?

The full development and validation of a new sensitive and robust test of spectral resolution is beyond the scope of this article, but it is worth considering what should be done, since the measurement of spectral resolution remains critically important. First of all, we recommend not grouping data that include spectral ripple thresholds both above and below the limit of aliasing/distortion. Scores above the limit might or might not indicate better resolution, and do not indicate that spectra with specific density can be discriminated. One useful step would be to limit stimulus presentation to only spectral densities that can be supported by the device, just as one would do for other domains of stimulus control. This would possibly result in lack of differentiation among many listeners who achieve scores at the upper limit. Another potential solution would be to abandon adaptive tracking, since it might result in a listener being “lost” in the parameter region where stimuli are different but not controlled (see earlier discussion of unreliability of high ripple scores in studies that used adaptive tracking).

Some behavioral tests hold promise to stand in for the conventional task of detecting phase inversion of increasingly dense acoustic ripples sent through the CI processor. The STRIPES test (Archer-Boyd et al. 2018) described earlier has been validated as a tool that is sensitive to experimenter-controlled systematic manipulations in analysis filter width (i.e., “spectral smearing”), and explicitly addresses some but not all of the issues raised with regard to spectral ripples in CIs. Low-density spectral ripples could be used to measure limits in the perception of spectral modulation depth (i.e., spectral modulation detection; Litvak et al. 2007; Zhang et al. 2013; Gifford et al. 2018), which requires a listener to discriminate a flat-spectrum sound from a sound that has spectral peaks and valleys (as opposed to discriminating when the peaks and valleys have changed positions in the spectrum). Studies of spectral modulation detection arguably measure within-channel intensity discrimination rather than spectral resolution per se (Anderson et al. 2012), which might explain why they have not been adopted more broadly. An alternative approach to detection of ripple phase inversion could be to restrict the spectral density to a degree that can be faithfully transmitted, and then adaptively find the minimum *phase change* that is detectable by a listener.

### CONCLUSIONS

There are major complications in the transmission of spectral ripples through CIs that render results questionable, or at least difficult to interpret, when the stimulus is above a critical ripple density. Spectral aliasing and neutralization of modulation depth are nonmonotonic distortions, meaning that above a certain spectral density (RPO value), spectral ripples of increasing

density are different but not ordered in any systematic way in the spectral domain. Different spectral ripple densities above the aliasing limit therefore cannot be clearly interpreted as indicating a specific degree of spectral resolution. This severely complicates the interpretation of thresholds above the limit, and undermines any linear correlations involving this metric. The relationship between spectral ripple scores and speech recognition in CI users is reliable, but is suggested here to not reflect an underlying mechanism of spectral resolution.

For the Cochlear device, the critical limit for spectral aliasing is around 2 RPO, and the theoretical upper limit for complete saturation of the narrowest filter is 3.46 RPO. The current analysis was mainly directly at static ripple stimuli, although there is no ripple phase at which these limits would not apply, so temporal modulation (e.g., phase changes used in SMRT stimuli) would not solve either of the problems raised here, if the experimenter is specifically trying to quantify spectral resolution. Additionally, the spectral modulations at these high densities fail to represent the modulations found in actual speech sounds, and correlations with speech recognition improve when thresholds above the critical value are excluded, suggesting that the mechanistic link to speech perception is tenuous. Experience arguably plays a role in spectral ripple discrimination, violating assumptions about its robustness to perceptual learning. Various alternative testing approaches exist, some of which have already been used in the published literature.

We conclude that the practice of identifying the threshold of spectral ripple density be limited because (1) it lacks the well-behaved mathematical properties a psychophysical experimenter requires and (2) high spectral densities are a poor representation of ecologically relevant speech cues, precluding an explanatory mechanistic account of performance correlations. We recommend that alternative testing strategies be designed with the specific consideration of real speech acoustics in their development.

### ACKNOWLEDGMENTS

We are grateful Andrew Oxenham, David Landsberger, Justin Aronoff, and Erin O'Neill for thoughtful and critical discussions of this project. We are also grateful to Liz Anderson for providing data collected in her 2011 study and to Erin O'Neill for providing data from her 2019 study. Kate Teece contributed to the gathering of information of this manuscript.

Funding was received grants from the NIH NIDCD R03 DC01439 and R01 DC017114 (M.B.W.) and from NIH Auditory Neuroscience Training Grant 2T32DC005361-16 (G.B.).

Portions of this work were presented at the Conference on Implantable Auditory Prostheses, Lake Tahoe, CA, July 2017, and the Acoustical Society of America, Louisville, KY, May 2019, and in a preprint at <https://psyarxiv.com/cbwgh/>.

The authors have no conflicts of interest to disclose.

Address for correspondence: Matthew B. Winn University of Minnesota Department of Speech-Language-Hearing Sciences 164 Pillsbury Drive SE, Minneapolis, MN 55455, USA. Email: [mwinn@umn.edu](mailto:mwinn@umn.edu)

### OPEN PRACTICES

This manuscript qualifies for an Open Materials Badge. The data have been made publically available at [https://github.com/ListenLab/Spectral\\_ripple/blob/master/SDC\\_2\\_Recreate\\_Spectral\\_Ripple\\_plots.R](https://github.com/ListenLab/Spectral_ripple/blob/master/SDC_2_Recreate_Spectral_Ripple_plots.R). More information about the Open Practices Badges can be found at <https://journals.lww.com/ear-hearing/pages/default.aspx>.

Received August 20, 2020; accepted September 25, 2021

## REFERENCES

- Anderson, E. S., Nelson, D. A., Kreft, H., Nelson, P. B., Oxenham, A. J. (2011). Comparing spatial tuning curves, spectral ripple resolution, and speech perception in cochlear implant users. *J Acoust Soc Am*, *130*, 364–375.
- Anderson, E., Oxenham, A., Nelson, P., Nelson, D. (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.
- Archer-Boyd, A. W., Southwell, R. V., Deeks, J. M., Turner, R. E., Carlyon, R. P. (2018). Development and validation of a spectro-temporal processing test for cochlear-implant listeners. *J Acoust Soc Am*, *144*, 2983.
- Aronoff, J. M., & Landsberger, D. M. (2013). The development of a modified spectral ripple test. *J Acoust Soc Am*, *134*, EL217–EL222.
- Azadpour, M., & McKay, C. M. (2012). A psychophysical method for measuring spatial resolution in cochlear implants. *J Assoc Res Otolaryngol*, *13*, 145–157.
- Bingabr, M., Espinoza-Varas, B., Loizou, P. C. (2008). Simulating the effect of spread of excitation in cochlear implants. *Hear Res*, *241*, 73–79.
- de Jong, M. A. M., Briaire, J. J., Frijns, J. H. M. (2018). Learning effects in psychophysical tests of spectral and temporal resolution. *Ear Hear*, *39*, 475–481.
- Drennan, W. R., Won, J. H., Nie, K., Jameyson, E., Rubinstein, J. T. (2010). Sensitivity of psychophysical measures to signal processor modifications in cochlear implant users. *Hear Res*, *262*, 1–8.
- Drennan, W. R., Won, J. H., Timme, A. O., Rubinstein, J. T. (2016). Nonlinguistic outcome measures in adult cochlear implant users over the first year of implantation. *Ear Hear*, *37*, 354–364.
- Gifford, R. H., Hedley-Williams, A., Spahr, A. J. (2014). Clinical assessment of spectral modulation detection for adult cochlear implant recipients: A non-language based measure of performance outcomes. *Int J Audiol*, *53*, 159–164.
- Gifford, R. H., Noble, J. H., Camarata, S. M., Sunderhaus, L. W., Dwyer, R. T., Dawant, B. M., Dietrich, M. S., Labadie, R. F. (2018). The relationship between spectral modulation detection and speech recognition: Adult versus pediatric cochlear implant recipients. *Trends Hear*, *22*, 2331216518771176.
- 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., Turner, C. W., 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.
- Jeon, E. K., Turner, C., Karsten, S., Henry, B., Gantz, B. (2015). Cochlear implant users' spectral ripple resolution. *J Acoust Soc Am*, *138*, 2350–2358.
- Jones, G. L., Won, J. H., Drennan, W. R., Rubinstein, J. T. (2013). Relationship between channel interaction and spectral-ripple discrimination in cochlear implant users. *J Acoust Soc Am*, *133*, 425–433.
- Jung, K. H., Won, J. H., Drennan, W. R., Jameyson, E., Miyasaki, G., Norton, S. J., Rubinstein, J. T. (2012). Psychoacoustic performance and music and speech perception in prelingually deafened children with cochlear implants. *Audiol Neurootol*, *17*, 189–197.
- Kirby, B. J., Spratford, M., Klein, K. E., McCreery, R. W. (2019). Cognitive abilities contribute to spectro-temporal discrimination in children who are hard of hearing. *Ear Hear*, *40*, 645–650.
- Landsberger, D. M., Svrakic, M., Roland, J. T. Jr, Svirsky, M. (2015). The relationship between insertion angles, default frequency allocations, and spiral ganglion place pitch in cochlear implants. *Ear Hear*, *36*, e207–e213.
- Lawler, M., Yu, J., Aronoff, J. M. (2017). Comparison of the spectral-temporally modulated ripple test with the arizona biomedical institute sentence test in cochlear implant users. *Ear Hear*, *38*, 760–766.
- Litvak, L. M., Spahr, A. J., Saoji, A. A., Fridman, G. Y. (2007). Relationship between perception of spectral ripple and speech recognition in cochlear implant and vocoder listeners. *J Acoust Soc Am*, *122*, 982–991.
- Liu, C., & Eddins, D. A. (2008). Effects of spectral modulation filtering on vowel identification. *J Acoust Soc Am*, *124*, 1704–1715.
- Munson, B., Donaldson, G., Allen, S., Collison, E., Nelson, D. (2003). Patterns of phoneme misperceptions by individual with cochlear implants. *J Acoust Soc Am*, *113*, 925–935.
- Narne, V. K., Sharma, M., Van Dun, B., Bansal, S., Prabhu, L., Moore, B. C. (2016). Effects of spectral smearing on performance of the spectral ripple and spectro-temporal ripple tests. *J Acoust Soc Am*, *140*, 4298.
- Nechaev, D. I., Milekhina, O. N., Supin, A. Y. (2019). Estimates of ripple-density resolution based on the discrimination from rippled and nonrippled reference signals. *Trends Hear*, *23*, 2331216518824435.
- Nelson, D. A., Donaldson, G. S., Kreft, H. (2008). Forward-masked spatial tuning curves in cochlear implant users. *J Acoust Soc Am*, *123*, 1522–1543.
- O'Neill, E. R., Kreft, H. A., Oxenham, A. J. (2019). Speech perception with spectrally non-overlapping maskers as measure of spectral resolution in cochlear implant users. *J Assoc Res Otolaryngol*, *20*, 151–167.
- Saoji, A. A., Litvak, L., Spahr, A. J., Eddins, D. A. (2009). Spectral modulation detection and vowel and consonant identifications in cochlear implant listeners. *J Acoust Soc Am*, *126*, 955–958.
- Scheperle, R. A., & Abbas, P. J. (2015). Relationships among peripheral and central electrophysiological measures of spatial and spectral selectivity and speech perception in cochlear implant users. *Ear Hear*, *36*, 441–453.
- Stakhovskaya, O., Sridhar, D., Bonham, B. H., Leake, P. A. (2007). Frequency map for the human cochlear spiral ganglion: Implications for cochlear implants. *J Assoc Res Otolaryngol*, *8*, 220–233.
- Supin, A., Nechaev, D., Milekhina, O., Sysueva, E. (2019). Discrimination of ripple depth in rippled spectra: Contributions of spectral and temporal mechanisms. *Proc Meet Acoustics*, *39*, 050001, 1–9.
- Winn, M. B., Won, J. H., Moon, I. J. (2016). Assessment of spectral and temporal resolution in cochlear implant users using psychoacoustic discrimination and speech cue categorization. *Ear Hear*, *37*, e377–e390.
- Won, J. H., Drennan, W. R., 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., Rubinstein, J. T. (2010). Psychoacoustic abilities associated with music perception in cochlear implant users. *Ear Hear*, *31*, 796–805.
- Won, J. H., Humphrey, E. L., Yeager, K. R., Martinez, A. A., Robinson, C. H., Mills, K. E., Johnstone, P. M., Moon, I. J., Woo, J. (2014). Relationship among the physiologic channel interactions, spectral-ripple discrimination, and vowel identification in cochlear implant users. *J Acoust Soc Am*, *136*, 2714–2725.
- Zhang, T., Spahr, A. J., Dorman, M. F., Saoji, A. (2013). Relationship between auditory function of nonimplanted ears and bimodal benefit. *Ear Hear*, *34*, 133–141.
- Zhou, N. (2017). Deactivating stimulation sites based on low-rate thresholds improves spectral ripple and speech reception thresholds in cochlear implant users. *J Acoust Soc Am*, *141*, EL243.
- Zhou, N., Dixon, S., Zhu, Z., Dong, L., Weiner, M. (2020). Spectrotemporal modulation sensitivity in cochlear-implant and normal-hearing listeners: Is the performance driven by temporal or spectral modulation sensitivity? *Trends Hear*, *24*, 1–11.