Spectral aliasing in an acoustic spectral ripple discrimination task (L)

Jesse M. Resnick,^{a)} David L. Horn,^{a),b)} Anisha R. Noble,^{a)} and Jay T. Rubinstein^{a)}

Department of Otolaryngology—Head and Neck Surgery, University of Washington, Box 357923, Seattle, Washington 98195-7923, USA

ABSTRACT:

Spectral ripple discrimination tasks are commonly used to probe spectral resolution in cochlear implant (CI), normal-hearing (NH), and hearing-impaired individuals. In addition, these tasks have also been used to examine spectral resolution development in NH and CI children. In this work, stimulus sine-wave carrier density was identified as a critical variable in an example spectral ripple-based task, the Spectro-Temporally Modulated Ripple (SMR) Test, and it was demonstrated that previous uses of it in NH listeners sometimes used values insufficient to represent relevant ripple densities. Insufficient carrier densities produced spectral under-sampling that both eliminated ripple cues at high ripple densities and introduced unintended structured interference between the carriers and intended ripples at particular ripple densities. It was found that this effect produced non-monotonic psychometric functions for NH listeners that would cause systematic underestimation of thresholds with adaptive techniques. Studies of spectral ripple detection in CI users probe a density regime below where this source of aliasing occurs, as CI signal processing limits dense ripple representation. While these analyses and experiments focused on the SMR Test, any task in which discrete pure-tone carriers spanning frequency space are modulated to approximate a desired pattern must be designed with the consideration of the described spectral aliasing effect. © 2020 Acoustical Society of America.

<https://doi.org/10.1121/10.0000608>

(Received 7 May 2019; revised 24 December 2019; accepted 30 December 2019; published online 11 February 2020)

[Editor: Joshua G. Bernstein]

Pages: 1054–1058

I. INTRODUCTION

Spectral ripples are a family of acoustic stimuli in which the power of sinusoidal carriers spanning the frequency domain are modulated as a function of frequency; they are often leveraged in attempts to probe spectral resolution, particularly in hearing-impaired subjects. The spectral-temporally modulated ripple (SMR) test elaborates on this strategy by asking listeners to discriminate between signals consisting of sinusoidal carriers modulated in both frequency and time domains (Aronoff and Landsberger, 2013). Typically, the time-domain modulation frequency is held at 5 Hz, and subjects are asked to make judgments between stimuli varying in “ripple density,” the log-periodic modulation in the frequency domain. As such, this task is designed as a method for probing listeners’ spectral resolution in the context of temporal modulation. Performance in this task has been widely shown to correlate well with speech measures in cochlear implant (CI) listeners (DiNino and Arenberg, 2018; Holden *et al.*, 2016; Lawler *et al.*, 2017; Zhou, 2017); however, one group found this correlation disappeared when they rigorously excluded temporal cues (Anderson *et al.*, 2012).

Previous work has found that adults with normal acoustic hearing may have spectral ripple density discrimination

up to ~10–11 ripples per octave (RPO) with a comparison stimulus of 20 RPO and when 33.33 carrier sine waves per octave are used to generate the stimuli (DiNino and Arenberg, 2018; Kirby and Brown, 2015; Landsberger *et al.*, 2018). We sought to adapt this task to test infants and children to explore the development of spectral resolution for both normal-hearing (NH) and CI listeners. We created a version of the task with longer stimuli and larger RPO step sizes to facilitate observer-based testing of young subjects. While piloting this task in adult listeners, we observed an interesting non-monotonicity in which some subjects distinguished 16 RPO from a 20 RPO reference but then failed when the spectral density decreased to 13.75 RPO. This non-monotonicity prompted us to explore the stimulus generation strategy in greater depth. We found that interaction between the carrier and ripple densities produced “spectral aliasing” when the ripple density was within a few hertz of half the carrier density, akin to the Nyquist frequency limit introduced by discrete time series data. With the 33.33 carriers/octave density utilized previously, this led to both the disruption in the representation of the intended spectral ripples near the Nyquist limit and a prominent beating artifact near 16.666 RPO that remained significant at both 10- and 20-dB ripple depths. In this Letter, we show that these stimulus generation artifacts led to the estimation of aberrantly poor ripple detection thresholds with adaptive threshold-finding approaches. By utilizing a higher carrier density, NH listeners’ discrimination thresholds improved by about 5

^{a)}Also at: Virginia Merrill Bloedel Hearing Research Center, University of Washington, Seattle, WA, USA, ORCID: 0000-0003-2108-4074.

^{b)}Electronic mail: david.horn@seattlechildrens.org

RPO with the same 20 RPO reference stimulus. These data demonstrate the importance of ensuring that the carrier density is sufficiently high to avoid the introduction of unintended artifacts by the discrete-spectral-carrier stimulus generation approach taken within the spectro-temporally modulated ripple task (SMRT).

II. METHODS

A. Participants

Five NH adults (mean age = 31 years old) were recruited from a University of Washington communication studies subject pool. Inclusion criteria were the following: native American English speaker, no ear otological disease or hearing loss, and normal distortion product otoacoustic emission screening on day of the testing. Compensation was \$15 per hour for participation. Testing was completed in a single 2-h test visit. This study was approved by the Seattle Children's Hospital Institutional Review Board.

B. Stimuli

SMR stimuli were generated as described by Aronoff and Landsberger (2013) with alterations to enable modification of carrier density and stimulus length. Briefly, phase-randomized sinusoidal carriers were produced with frequencies uniformly spaced within the logarithmic frequency interval 100 to 6500 Hz at the selected carrier densities. Each carrier's amplitudes were spectro-temporally modulated with a time-variant full wave rectified sinusoid.

Onset and offset ramps, 100 ms each, were applied to the modulated stimuli.

Figure 1 shows the spectrograms of examples of the SMR stimuli used. Figures 1(A) and 1(B) show SMR stimuli produced using 33 carriers per octave with 4 and 16 RPO, respectively. Figures 1(C) and 1(D) present similar stimuli produced with 100 carriers per octave. Both the 33 and 100 carrier-per-octave stimuli produce the expected spectro-temporal sweeps at 4 RPO [Figs. 1(A) and 1(C)]. Notice, however, that the 33 carrier density spectrograms show periodic, unintended absence of signal due the discrete sampling of the ripples [Fig. 1(A)]. While this effect is not observable for the 100 carrier-per-octave-based signals' spectra [Fig. 1(C)], this is merely due to spectral-resolution limits of the discrete fast Fourier transform used. Most importantly, while the 100 carrier-per-octave, 16 RPO signal [Fig. 1(D)] maintains the expected spectro-temporal sweeps, the 33 carrier-per-octave version [Fig. 1(B)] exhibits a more checkerboard-like pattern with the desired ripples being obscured by unintended spectro-temporal sweeps with lower RPO, higher temporal frequency, and opposite spectro-temporal direction than the intended ripples.

The source of this unintended lower-frequency signal within the stimuli can be best illustrated by considering the discrete sampling of the frequency spectra modulations at a single time point (Fig. 2). The top and bottom panels illustrate the intended and discretely approximated modulations for 33 and 100 carrier-per-octave stimuli, respectively. Figures 2(A) and 2(C) illustrate the modulation approach for relatively low (4 RPO) ripple densities while Figs. 2(B) and 2(D) illustrate

FIG. 1. (Color online) Example spectrograms of SMR stimuli with 33.33 (A),(B) and 100 (C),(D) carrier densities at 4 (A),(C) and 16 (B),(D) RPO. Grayscale mapping is according to the relative level in dB. Note that these stimuli are shown without onset and offset ramps. A discrete fast Fourier transform frequency resolution of 50 Hz was used.

FIG. 2. (Color online) Time frozen spectrograms comparing ideal (blue lines) and discretely sampled (red markers) modulation spectra of SMR stimuli for 33.33 (A & B) and 100 (C & D) carrier densities at 4 (A & C) and 16 (B & D) RPO. For each modulation spectra, a single octave representing the frequency range from 100 to 200 Hz is illustrated.

that of relatively high (16 RPO) ripple densities. For clarity, only a single octave representing the frequency range from 100 to 200 Hz is shown; however, the presented dynamics are consistent across the frequency spectrum. For the 4ripple densitiesRPO stimuli, both the 33 [Fig. 1(A)] and 100 [Fig. 1(C)] carrier-per-octave discrete modulation approaches sample at, or near, both the peaks and troughs of the idealized ripple pattern. Conversely, while the 100 carrier-per-octave sampling resolution largely captures the full range of the desired rippling on every cycle of the 16ripple densitiesRPO stimulus [Fig. 1(D)], the 33 carrier-per-octave-based stimuli do not [Fig. 1(B)]. The relatively low spectral sampling density relative to that of the ripples leads to aliasing in the modulation intensity that introduces an additional lower density modulation.

C. Psychophysical task

The effect of aliasing [Fig. 2(B)] on SMR discrimination was tested using a single-interval, forced choice observer-based psychoacoustic procedure (OPP), previously used by Horn and colleagues with infants and adults (Horn *et al.*, 2017b, 2017a). Testing was conducted in a single-walled sound booth with the listener facing a loudspeaker 1.6 m away at 0° azimuth. Listeners were instructed to raise their hand when they heard the sound that activated a video clip in the room. Stimuli were presented in soundfield at 70 dB SPL with 1 s inter-stimulus interval. An observer sat outside the test booth and observed the participant's behavior through a glass window. Participants were presented with repeating background “standard” stimuli (20 RPO with spectral phase chosen at random) and were blind to trial initiation. The observer initiated trials intermittently when the participant was quiet, prepared, and facing forward. Two types of trials were utilized: “change” trials in which one of the “non-standard signals” (RPO < 20) was presented and “catch” trials in which the standard was repeated. Trial type was determined randomly. To the listener, “catch” trials were indistinguishable from the background standard stimuli. Trial presentation began the full stimulus period after

initiation and was followed by a 4-s response window. During this window, the observer marked whether the participant raised their hand with a keyboard button press.

For each condition, the task began with an initial training phase, designed to produce association between “change” trials and the reinforcer, by activating the reinforcer after each change trial regardless of the observer classification. During training, ripple density was fixed at 4 RPO and trial type was varied pseudo-randomly with six change and two no-change trials in each block of eight trials. Training ended when the observer reached a criterion of >80% hit rate and >80% correct rejection rate over five consecutive change and no-change trials, respectively. All subjects passed the training phase within one block of eight trials for all conditions. Training was followed by a testing phase which differed in three key respects. First, the ripple density for change trials varied pseudo-randomly within each condition with eight total presentations for each RPO, two each with phases of 0, $\pi/2$, π , and $3\pi/2$. Second, change and no-change trials were presented pseudo-randomly with a frequency of 4:1. Finally, the reinforcer was activated only on “change” trials on which the observer indicated a response.

Each participant completed a block for four different conditions, consisting of stimulus durations of either 0.5 or 1 s at carrier densities of both 33.333 and 100 carriers per octave. Condition order was randomized between subjects. For each condition, “standard” stimuli were 20 RPO and were produced using the same strategy as for the “non-standard” stimuli. For the 33.333 carrier/octave conditions, SMRs at ripple densities of 5.657, 8, 9.514, 11.314, 13.454, 16, and 19.027 RPO (corresponding to 2 raised to the power of 2.5, 3, 3.25, 3.5, 3.75, 4, and 4.25) were presented. For the 100 carrier/octave conditions, ripple densities 16.708, 17.448, and 18.221 RPO (corresponding to 2 raised to the power of 4.0625, 4.125, and 4.1875) were also included to enable closer approximation of thresholds.

III. RESULTS

Figure 3 presents group-mean percent-correct data for all subjects from this experiment. With the 100 carrier-per-octave

FIG. 3. (Color online) Group-mean proportion correct plotted against ripple density for the SMR discrimination task. Mean data for the 33.33 and 100 carriers-per-octave-based stimuli are plotted in red and blue, respectively. Long stimuli (1 s) are plotted as circular marks, and solid lines and short stimuli (0.5 s) are plotted as triangles and dashed lines. Error bars represent the standard error of the mean (SEM). All ripple stimuli were presented at a mean level of 75 dBA with a ripple depth of 20 dB. The black dashed line represents chance performance level.

stimuli (blue symbols), the group-mean and individual (data not shown) psychometric curves exhibited a typical shape declining monotonically from saturation to chance. In contrast, the 33.33 carrier-per-octave stimuli (red symbols) yielded strongly non-monotonic psychometric curves that showed a reduction in performance from saturation to chance between 8 and 13.454 RPO but a rebound in performance at 16 RPO. The effects of carrier density, stimulus length, and ripple density on percent-correct performance were examined using a linear mixed-effects model with a compound symmetry covariance matrix. This analysis was chosen over a repeated measures analysis of variance (ANOVA) due to the fact that subjects were not tested at all ripple depths in the 33 carrier-per-octave condition. Significant main effects for all three variables were observed with better mean percent-correct performance scores associated with larger carrier density [$F(1,148) = 63.378$, $p < 0.0001$], longer stimulus length [$F(1,148) = 8.893$, $p = 0.003$], and smaller ripple density [$F(10,148) = 63.229$, $p < 0.0001$]. A significant interaction between ripple density and carrier density was observed [$F(7,148) = 17.411$, $p < 0.0001$]. Paired comparisons between the two carrier densities at each ripple density revealed no significant differences at 4–8 RPO, 16 RPO, or 19.03 RPO (all p -values > 0.234), whereas a higher carrier density was associated with a higher percent-correct at 9.51–13.45 RPO (all p -values < 0.011). All other interactions failed to reach significance (all p -values > 0.375). With the Bonferroni-Holm correction for eight paired comparisons, the effect of carrier density remained significant only at 11.314 and 13.454 RPO. The first 70% crossing, a typical threshold in pediatric psychophysics experiments, was approximately 11 RPO for the low carrier density signals with both stimulus durations, while for the high carrier density signals, they were 16.5 and 17.5 for the short and long stimuli, respectively.

IV. DISCUSSION

Figures 1 and 2 present evidence that the use of an inadequate density of sine-wave carriers can lead to spectral

aliasing for SMR stimuli. The use of high carrier densities to produce spectrally rippled stimuli has been previously leveraged to produce static spectral ripple stimuli (Horn *et al.*, 2017b; Jones *et al.*, 2013); however, these studies did not explicitly describe the spectral aliasing effect they sought to avoid. For SMR stimuli, this aliasing effect can both interfere with intended ripple cues via under-sampling and introduce unintended low-frequency spectro-temporal beats through interaction between the ripples and sampling density. The under-sampling effect becomes progressively more prominent as the desired ripple density approaches half the carrier density, akin to the Nyquist frequency in time series data, with ripples above this cutoff being completely eliminated in the generation process. The beating cues are present with non-monotonic intensity across ripple densities and depend on how structured the pattern of interference between the carrier-density and ripples is. The net effect of these two artifacts is that listeners' ability to discriminate targets from the standard, which is also impacted by the artifacts, is generally reduced due to the under-sampling but with non-monotonicities near integer multiples of the Nyquist limit where beating cues may arise. While we specifically illustrate this aliasing occurring for SMRs, any other stimuli generated via band-limited approximation of spectral functions, including static spectral ripples, may be impacted by similar effects.

The results of our psychophysical experiment suggest that, with a sufficiently high carrier density, NH adults are able to discriminate 16.5 or even 17.5 RPO SMR stimuli from a 20 RPO standard. In contrast, the predominant carrier density used previously, 33.33 sine wave per octave, yields non-monotonic psychometric functions that suggest listeners begin losing access to desired spectral ripple cues above approximately 11 RPO due to under-sampling in stimulus generation but are again able to perform discrimination at ripple densities near 16 RPO, presumably because they are able to detect the aliasing-induced beating. This finding demonstrates the importance of ensuring a sufficiently high carrier density is selected to accurately measure SMR detection.

In psychophysics, an important distinction is drawn between the processes of discrimination (i.e., distinguishing between two quantitatively different stimuli) and detection (i.e., detecting a qualitative change between a signal and a "standard"). While the SMR test was initially conceived as a detection task, with the 20 RPO thought to be perceived as noise, listeners' ability to detect ripple densities relatively near the 20 RPO standard with the elimination of the under-sampling artifacts indicates that this strategy is probing ripple discrimination rather than ripple detection. Going forward, if a discrimination task is to be used, it would make the task more phonetically relevant to pick an RPO standard mimicking the 0.5–3 octave spectral distance between formants and report minimal discriminable ripple differences. Multiple groups have demonstrated that CI users' ability to discriminate between ripple densities in this regime correlate well with some speech outcomes (notably formant categorization), providing precedence for this approach (Henry

et al., 2005; Saoji *et al.*, 2009; Winn *et al.*, 2016). Alternatively, if absolute SMR detection capabilities are of interest, transitioning to unmodulated standard stimuli and leveraging a sufficiently large number of carriers to avoid aliasing at the highest tested ripple densities would enable interrogation of the physiological limit of ripple detection.

A significant caveat to the approach of attempting to characterize the upper limit of spectral ripple density detection is that there exists controversy over the physiological mechanism for representing high density ripples. There exists the potential for adjacent ripple peaks to introduce temporal beating when the peaks fall within an auditory filter that may provide a nonspectral cue that listeners use to distinguish dense spectra-ripples from unmodulated noise (Anderson *et al.*, 2012). Modeling the representation of SMRT stimuli upon the basilar membrane with different carrier and ripple densities to explore whether such nonspectral cues arise is a fruitful exercise but beyond the scope of this Letter. Nevertheless, the fact remains that listeners can hear and discriminate these ripple envelopes and these results are generally correlated with speech perception (DiNino and Arenberg, 2018; Holden *et al.*, 2016; Lawler *et al.*, 2017), so such tests have relevance for probing abilities critical for speech perception, particularly in the hearing-impaired, pre-lingual population.

Other groups have explored the minimum detectable ripple depth at lower ripple densities, particularly in hearing-impaired listeners, and found these measures correlate well with speech-perception measures (Bernstein *et al.*, 2013; Bernstein *et al.*, 2016; Davies-Venn *et al.*, 2015; Mehraei *et al.*, 2014; Miller *et al.*, 2018). Such approaches should not require the high carrier densities necessary to probe ripple density thresholds and, therefore, should be robust to the issue identified here. Similarly, experiments using SMR in CI users are limited to SMR densities below the limit where stimulus generation introduces aliasing, suggesting that findings in this population, notably the consistent correlation with speech outcomes, were likely robust to this effect (DiNino and Arenberg, 2018; Holden *et al.*, 2016; Lawler *et al.*, 2017). That said, CI digital signal processing requires filtering signals into discrete bands for representation on individual electrodes, a process that under-samples dense SMR stimuli and introduces another source of spectral aliasing (Winn and O'Brien, 2019).

ACKNOWLEDGMENTS

Thank you to Mariette S. Broncheau for her assistance with subject recruitment and scheduling.

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**, EL217–EL222.
- Bernstein, J. G. W., Danielsson, H., Hällgren, M., Stenfelt, S., Rönnberg, J., and Lunner, T. (2016). "Spectrotemporal modulation sensitivity as a predictor of speech-reception performance in noise with hearing aids," *Trends Hear.* **20**, 233121651667038.
- Bernstein, J. G. W., Mehraei, G., Shamma, S., Gallun, F. J., Theodoroff, S. M., and Leek, M. R. (2013). "Spectrotemporal modulation sensitivity as a predictor of speech intelligibility for hearing-impaired listeners," *J. Am. Acad. Audiol.* **24**, 293–306.
- Davies-Venn, E., Nelson, P., and Souza, P. (2015). "Comparing auditory filter bandwidths, spectral ripple modulation detection, spectral ripple discrimination, and speech recognition: Normal and impaired hearing," *J. Acoust. Soc. Am.* **138**, 492–503.
- DiNino, M., and Arenberg, J. G. (2018). "Age-related performance on vowel identification and the spectral-temporally modulated ripple test in children with normal hearing and with cochlear implants," *Trends Hear.* **22**, 233121651877095.
- Henry, B. A., Turner, C. W., and Behrens, A. (2005). "Spectral peak resolution and speech recognition in quiet: Normal hearing, hearing impaired, and cochlear implant listeners," *J. Acoust. Soc. Am.* **118**, 1111–1121.
- Holden, L. K., Firszt, J. B., Reeder, R. M., Uchanski, R. M., Dwyer, N. Y., and Holden, T. A. (2016). "Factors affecting outcomes in cochlear implant recipients implanted with a perimodiolar electrode array located in scala tympani," *Otol. Neurotol.* **37**, 1662–1668.
- Horn, D. L., Dudley, D. J., Dedhia, K., Nie, K., Drennan, W. R., Won, J. H., Rubinstein, J. T., and Werner, L. A. (2017a). "Effects of age and hearing mechanism on spectral resolution in normal hearing and cochlear-implanted listeners," *J. Acoust. Soc. Am.* **141**, 613–623.
- Horn, D. L., Won, J. H., Rubinstein, J. T., and Werner, L. A. (2017b). "Spectral ripple discrimination in normal-hearing infants," *Ear Hear.* **38**, 212–222.
- Jones, G. L., Won, J. H., Drennan, W. R., and Rubinstein, J. T. (2013). "Relationship between channel interaction and spectral-ripple discrimination in cochlear implant users," *J. Acoust. Soc. Am.* **133**, 425–433.
- Kirby, B. J., and Brown, C. J. (2015). "Effects of nonlinear frequency compression on ACC amplitude and listener performance," *Ear Hear.* **36**, e261–e270.
- Landsberger, D. M., Padilla, M., Martinez, A. S., and Eisenberg, L. S. (2018). "Spectral-temporal modulated ripple discrimination by children with cochlear implants," *Ear Hear.* **39**, 60–68.
- Lawler, M., Yu, J., and 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.
- Mehraei, G., Gallun, F. J., Leek, M. R., and Bernstein, J. G. W. (2014). "Spectrotemporal modulation sensitivity for hearing-impaired listeners: Dependence on carrier center frequency and the relationship to speech intelligibility," *J. Acoust. Soc. Am.* **136**, 301–316.
- Miller, C. W., Bernstein, J. G. W., Zhang, X., Wu, Y.-H., Bentler, R. A., and Tremblay, K. (2018). "The effects of static and moving spectral ripple sensitivity on unaided and aided speech perception in noise," *J. Speech, Lang. Hear. Res.* **61**, 3113–3126.
- 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**, 955–958.
- Winn, M. B., and O'Brien, G. (2019). "Flaws in spectral ripple stimuli for listeners with cochlear implants," *PsyArXiv*, preprint 31 May 2019, available at <https://psyarxiv.com/cbwgh/> (Last viewed 1/30/2020).
- Winn, M. B., Won, J. H., and Moon, I. J. I. (2016). "Assessment of spectral and temporal resolution in cochlear implant users: Speech and psychoacoustic approach," *Ear Hear.* **37**, e377–e390.
- 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–EL248.