

## Research Article

# The Effects of Static and Moving Spectral Ripple Sensitivity on Unaided and Aided Speech Perception in Noise

Christi W. Miller,<sup>a</sup> Joshua G. W. Bernstein,<sup>b</sup> Xuyang Zhang,<sup>c</sup>  
Yu-Hsiang Wu,<sup>c</sup> Ruth A. Bentler,<sup>c</sup> and Kelly Tremblay<sup>a</sup>

**Purpose:** This study evaluated whether certain spectral ripple conditions were more informative than others in predicting ecologically relevant unaided and aided speech outcomes.

**Method:** A quasi-experimental study design was used to evaluate 67 older adult hearing aid users with bilateral, symmetrical hearing loss. Speech perception in noise was tested under conditions of unaided and aided, auditory-only and auditory-visual, and 2 types of noise. Predictors included age, audiometric thresholds, audibility, hearing aid compression, and modulation depth detection thresholds for moving (4-Hz) or static (0-Hz) 2-cycle/octave spectral ripples applied to carriers of broadband noise or 2000-Hz low- or high-pass filtered noise.

**Results:** A principal component analysis of the modulation detection data found that broadband and low-pass static and moving ripple detection thresholds loaded onto the first factor whereas high-pass static and moving ripple detection thresholds loaded onto a second factor. A linear mixed model revealed that audibility and the first factor (reflecting

broadband and low-pass static and moving ripples) were significantly associated with speech perception performance. Similar results were found for unaided and aided speech scores. The interactions between speech conditions were not significant, suggesting that the relationship between ripples and speech perception was consistent regardless of visual cues or noise condition. High-pass ripple sensitivity was not correlated with speech understanding.

**Conclusions:** The results suggest that, for hearing aid users, poor speech understanding in noise and sensitivity to both static and slow-moving ripples may reflect deficits in the same underlying auditory processing mechanism. Significant factor loadings involving ripple stimuli with low-frequency content may suggest an impaired ability to use temporal fine structure information in the stimulus waveform. Support is provided for the use of spectral ripple testing to predict speech perception outcomes in clinical settings.

Poor understanding of speech in background noise is a persistent complaint for individuals with hearing loss. Reduced audibility partially explains the deficits in speech-in-noise outcomes, but even among those with similar audiograms or audibility indices, variability in performance occurs (Humes, 2007; Smoorenburg, 1992). Secondary to the well-understood effects of audibility,

other factors suggested as responsible for variability in speech perception in noise include age (e.g., Dubno, Dirks, & Morgan, 1984), cognitive function (e.g., Akeroyd, 2008; Dryden, Allen, Henshaw, & Heinrich, 2017; Lunner, 2003; Zekveld, Rudner, Johnsrude, & Rönnerberg, 2013), and temporal and/or spectral resolution (e.g., Baer & Moore, 1993, 1994; Helfer & Vargo, 2009; Lorenzi, Debrulle, Garnier, Fleuriot, & Moore, 2009). Although cognitive factors are important for understanding performance in noisy environments, this study focused on characterizing auditory distortion associated with sensorineural hearing loss not captured by audibility metrics. Accurately predicting an individual's speech understanding in noise could assist the clinician in setting patient expectations regarding the potential efficacy of hearing aid (HA) treatment options.

Sensorineural hearing loss is associated with a loss of the active cochlear mechanism, causing broadly tuned

<sup>a</sup>Department of Speech and Hearing Sciences, University of Washington, Seattle

<sup>b</sup>National Military Audiology and Speech Pathology Center, Walter Reed National Military Medical Center, Bethesda, MD

<sup>c</sup>Department of Communication Sciences and Disorders, University of Iowa, Iowa City

Correspondence to Christi W. Miller: christim@uw.edu

Editor-in-Chief: Frederick (Erick) Gallun

Editor: Daniel Fogerty

Received October 2, 2017

Revision received June 6, 2018

Accepted August 4, 2018

[https://doi.org/10.1044/2018\\_JSLHR-H-17-0373](https://doi.org/10.1044/2018_JSLHR-H-17-0373)

**Disclosure:** The authors have declared that no competing interests existed at the time of publication.

auditory filters (e.g., Glasberg & Moore, 1986) and an impaired spectral representation of speech signals (i.e., an inability to resolve the frequency components of a complex sound; Leek, Dorman, & Summerfield, 1987; Summers & Leek, 1994). Broadly tuned filters also result in a decreased signal-to-noise ratio at the output of each auditory filter, increasing the masking effects of background noise (Baer & Moore, 1993, 1994). Although spectral resolution has been traditionally examined using a spectral masking paradigm to characterize the auditory filter at a specific cochlear location (e.g., Glasberg & Moore, 1986), another approach is to measure the ability to detect the presence of spectral peaks and valleys in a broadband noise stimulus (e.g., Davies-Venn, Nelson, & Souza, 2015; Summers & Leek, 1994). This approach requires the listener to perform an across-channel analysis to detect the presence of a spectral ripple. Spectral ripple sensitivity—typically characterized in terms of the modulation depth required to detect a spectral ripple with a given density (cycles/octave) or the maximum density at which a given modulation depth can be detected—is a significant predictor of speech understanding in noise for listeners with hearing impairment with acoustic hearing, explaining 23%–63% of the variance across an assortment of speech tasks (Davies-Venn et al., 2015).

Sensorineural hearing loss also causes a decline in certain temporal processing abilities (e.g., Hall & Fernandes, 1983; Irwin & Purdy, 1982; Kidd, Mason, & Feith, 1984). It is mainly the perception of the fine temporal aspects of speech that are thought to be related to speech-in-noise perception (e.g., Buss, Hall, & Grose, 2004; Hopkins & Moore, 2007; Neher, Lunner, Hopkins, & Moore, 2012). Referred to as temporal fine structure (TFS), the auditory nerve fibers encode these fast fluctuations (on the order of hundreds to thousands of hertz) in the signal by phase locking (Johnson, 1980). However, it is possible that listeners with sensorineural hearing loss have jitter in the precise timing of auditory nerve firings (e.g., Miller, Schilling, Franck, & Young, 1997), potentially disrupting the encoded sound information within a channel or changes in relative phase that could impact decoding of TFS across channels (e.g., Loeb, White, & Merzenich, 1983; Ruggero, 1994). TFS carries information for pitch perception (Moore, 2003), separation of competing talkers (Brokx & Noteboom, 1982), and cueing when to listen for a target talker in modulated background noise (Hopkins, Moore, & Stone, 2008). These cues contribute critical information for understanding speech in noise. Listeners with normal hearing are able to use TFS cues with minimal envelope cues to understand speech in noise (Lorenzi et al., 2009; Lorenzi, Gilbert, Carn, Garnier, & Moore, 2006; Lorenzi & Moore, 2008), whereas listeners with hearing impairment experience difficulty in doing so (Hopkins & Moore, 2007; Lorenzi et al., 2006).

Although historically spectral and temporal processing abilities have each been studied independently, the ability to detect the presence of joint spectrotemporal modulation (STM) has recently received much attention as a significant predictor of speech in noise (Bernstein et al., 2013, 2016; Mehraei, Gallun, Leek, & Bernstein, 2014;

Won et al., 2015). Speech contains energy fluctuations over both time and frequency, which co-occur in natural sounds, and joint STM detection is measured using moving spectral ripples that are characterized by spectral modulation density (cycles/octave) and temporal modulation rate (Hz). A speech spectrogram can be broken down into integral STM components with varying parameters of modulation density and rate (Chi, Gao, Guyton, Ru, & Shamma, 1999; Chi, Ru, & Shamma, 2005; Elhilali, Taishih, & Shamma, 2003). STM detection has been associated with both TFS processing and frequency selectivity abilities (Bernstein et al., 2013; Mehraei et al., 2014). Despite compelling evidence that STM sensitivity can predict speech-in-noise perception, there are several remaining questions.

The first purpose of this study was to determine if the temporal (moving) component of the ripple adds any predictive power to speech perception in noise over the static component of the ripple in individuals who use acoustic hearing. Impairments or enhancements in perception may exist when spectral and temporal domains are tested in combination, and although there are some data to support these assumptions (Zheng, Escabi, & Litovsky, 2017), a direct comparison between sensitivity to static and moving ripples has not been made in the same subjects with acoustic hearing. Comparing the sensitivity between static and moving ripples may facilitate a better understanding of the underlying mechanisms that are essential for speech perception in noise.

If sensitivity to static and moving ripples are correlated and both are good predictors of speech-in-noise perception, then these tasks may be measuring the same underlying function(s). In particular, the ability to detect moving ripples (i.e., STM) might rely on the ability to make use of TFS information to detect the presence of moving spectral peaks. Based on a similar argument for frequency modulation detection (Moore & Sek, 1994, 1996; Moore & Skrodzka, 2002), Bernstein and colleagues argued that the pattern of conditions whereby listeners with hearing impairment demonstrate poor STM sensitivity is consistent with the idea that poor STM sensitivity reflects impaired TFS processing (Bernstein et al., 2013, 2016; Mehraei et al., 2014). Mehraei et al. (2014) found that listeners with hearing impairment mainly show deficits for low carrier frequencies, where TFS cues are available, and for slow temporal modulation rates, where a sluggish TFS mechanism can make use of this information. At the same time, listeners with hearing impairment show relatively little deficit for high carrier frequencies, where even listeners with normal hearing do not have access to phase-locking information, and for high modulation rates, where even listeners with normal hearing are unlikely to have access to TFS cues due to the sluggish TFS processing mechanism. In contrast to the TFS argument, it has often been argued that static spectral ripple detection reflects a listener's spectral resolution ability (e.g., Anderson, Oxenham, Nelson, & Nelson, 2012; Davies-Venn et al., 2015; Summers & Leek, 1994). However, it is possible that listeners require TFS information to detect static ripples. In fact, a static ripple

can be thought of as an extreme case of a slow-moving STM stimulus. To examine whether the detection of static and moving ripples and their relationship to speech perception reflect the same underlying mechanism, modulation detection thresholds were measured for both static (0 Hz) and moving (4 Hz) STM stimuli. Furthermore, performance was evaluated in three spectral conditions—low-pass (below 2 kHz), high-pass (above 2 kHz), and broadband—to examine whether ripple detection performance and its relationship to speech understanding in noise depended on the availability of phase-locking information.

The second purpose of this study was to extend previous findings to more real-world aided conditions to gauge how well the results could be generalized, using “ecologically relevant” outcomes. Results from previous studies demonstrated that STM detection thresholds explain between 13% (Bernstein et al., 2016) and 40% (Bernstein et al., 2013; Mehraei et al., 2014) of unique variance in speech-in-noise performance, over the variability explained by the audiogram. The significant relationship between STM sensitivity and speech in noise has been established in people with hearing loss without the use of HAs, but at high listening levels (e.g., Bernstein et al., 2013; Mehraei et al., 2014), and was recently replicated under aided conditions, using well-fit HAs to achieve optimal audibility (Bernstein et al., 2016). Researchers have suggested that an STM clinical test battery is warranted based on these results. These studies used speech stimuli consisting of auditory-only (AO) cues spoken by a single-target talker, yet real-world environments often consist of interactions with multiple people throughout a conversation and frequently with the addition of visual cues. In addition, multichannel compression processing, which is the most common type of processing used in current HAs, is known to distort spectral (Bor, Souza, & Wright, 2008; Yunds & Buckles, 1995) and temporal cues (Jenstad & Souza, 2005, 2007; Souza & Turner, 1998). If the processing operates with fast time constants, distortion may occur and impair speech outcomes. Although most HAs use some form of compression processing, manufacturers make different decisions regarding the speed of processing (i.e., attack and release time) and channel implementation (e.g., frequency allocation of channels). These manufacturer-dependent decisions could alter the amount of temporal and spectral distortion imposed by HA processing. An important next step is to evaluate whether this significant relationship is maintained under ecologically relevant listening conditions, which would provide support for implementation of an STM clinical testing protocol. The current study evaluated speech perception for individuals already fit with HAs from a variety of clinics in an ecologically relevant paradigm involving speech materials produced by a range of talkers and the presence of visual speech cues.

### **Talker Variability**

Throughout a speech task using a single-target talker, listeners learn indexical cues, which refer to voice-specific

features (e.g., variations in length and resonance of vocal tract) that are influenced by age, gender, and regional dialect, among other traits (Broadbent, 1952; Mullennix, Pisoni, & Martin, 1989). One may assume that spectrotemporal cues support perception of these voice-specific features (e.g., pitch perception; Cabrera, Tsao, Gnansia, Bertocini, & Lorenzi, 2014). Availability of indexical cues lead to better speech perception than when indexical cues are not available (Mullennix et al., 1989; Nygaard & Pisoni, 1998; Nygaard, Sommers, & Pisoni, 1994). Furthermore, self-reported hearing abilities correlate better with performance on understanding multiple-target talkers (i.e., the target talker varies from sentence to sentence) than with single-target talker conditions (Kirk, Pisoni, & Miyamoto, 1997). If significant effects of STM sensitivity are observed under more typical speech conditions, implementation of a clinical STM test battery would be supported. On the other hand, it is also possible that weaker effects would be observed under multiple-target talker conditions, because indexical cue learning is not as prevalent.

### **Visual Cues**

Visual cues improve speech perception, likely due to enhancing auditory detection from comodulation of mouth movements and the acoustic envelope (i.e., auditory–visual [AV] integration; Grant & Seitz, 2000). The addition of visual cues is exceptionally helpful when the background noise is also speech (Helfer & Freyman, 2005), and when audibility is reduced or when the listener cannot take advantage of the audibility provided (Bernstein & Grant, 2009). In the case of hearing loss, acoustic coding of the temporal envelope may be impaired due to HA signal processing, the damaged auditory periphery, or a disruption in the neural coding of the signal; creating a mismatch between auditory and visual cues and AV integration may decline (Bernstein & Grant, 2009; Grant & Seitz, 2000). Visual cues also inform the listener of phonetic information, even when the signal is not audible; hence, STM sensitivity may have a weaker association with AV speech perception because perception is not merely reliant on auditory cues.

### **Method**

A quasi-experimental approach was used to evaluate associations between STM sensitivity and outcomes of speech perception in noise. The benefit of an observational approach is that a large group of typical HA users was assessed, with HAs obtained in a range of clinical settings and who vary in their degree of success with amplification, making results highly generalizable.

### **Participants**

Participants were recruited from a larger study on HA outcomes. Adult HA users were recruited from participation pool databases across two sites: the University of Washington (UW) and the University of Iowa (UI). Individuals were

identified across both databases, who were between the ages of 21 and 79 and with a history of bilateral, symmetrical, mild to moderately severe sensorineural hearing loss and HA use. Recruitment also occurred through word of mouth and advertising. Adults who responded to our initial contact attempt underwent further screening prior to being enrolled in the study. The inclusion criteria were as follows: fluent English speaker (self-reported); bilateral HA user (self-reported, at least 8 hr/week over the last 6 months); American National Standards Institute (ANSI, 2003) high-frequency average HA gain of 5 dB or greater; a total score exceeding 21 out of 30 on the Montréal Cognitive Assessment (Nasreddine et al., 2005) to minimize the probability of dementia; no self-reported comorbid health conditions (e.g., conductive hearing loss, active otologic disorders, complex medical history involving the head, neck, ears, or eyes) that would interfere with the study procedures; and bilateral, symmetrical, mild to moderately severe sensorineural hearing loss. Bilateral, symmetrical, mild to moderately severe sensorineural hearing loss was defined as thresholds no poorer than 70 dB HL from 250 through 4000 Hz, air–bone gaps no greater than 15 dB with a one-frequency exception, and all interaural air-conduction thresholds within 15 dB at each frequency with a one-frequency exception. To determine the amount of gain the HAs were providing, each aid was placed in an Audioscan Verifit test chamber, and the automatic gain control ANSI test battery was performed at user settings. The high-frequency average gain was determined to be the amount of gain reported for a 50-dB input averaged at 1000, 1600, and 2500 Hz. The number of participants in the larger study totaled 154 at the time of recruitment. Phone calls and emails were made to participants from the larger study at both sites inviting them to return for further testing.

Sixty-seven participants (38 women, 29 men) provided informed consent to participate in the study approved by the Human Subjects Review Committee at UW and UI. A power analysis was conducted using G\*Power v3.1.9.2 (Faul, Erdfelder, Buchner, & Lang, 2009; Faul, Erdfelder, Lang, & Buchner, 2007) for a multiple regression analysis. Effect sizes were initially estimated from published data (Bernstein et al., 2013; Mehraei et al., 2014) and later in the study using the results of the first 37 subjects. A sample size of 60 was suggested to reach a power of 80%. Participants were encouraged to take breaks as often as they needed throughout testing and were paid \$15/hr for their time.

### ***Stimuli and Procedure***

Experimental testing took place in a sound-attenuated booth. All stimuli were presented using a combination of Windows Media Player, MATLAB, and PsychToolBox (MATLAB and Statistics Toolbox, 2013) and a custom platform for presenting stimuli. Testing took place over two to three sessions, depending on the preference of the participant. Following screening and audiometric measures, speech perception was tested, with STM sensitivity measured in the last session.

### **Hearing Aids**

The HA processing was quantified using probe microphone measures on the Audioscan Verifit. Real-ear responses were recorded with the user's own HAs and a 65 dB SPL standard speech passage. Two metrics were derived from the real-ear response to the 65 dB SPL speech passage. Audibility was recorded as the Speech Intelligibility Index (SII) calculated by the Verifit (ANSI S3.5; American National Standards Institute, 1997), and the better-ear value was used in subsequent analysis. The amount of compression processing was quantified as the difference (dB) in speech output levels (dB SPL) between the 99th and 30th percentiles, termed the peak-to-valley ratio (PVR). Greater amounts of compression show smaller PVR values (e.g., stronger compression ratios, faster time constants; Henning & Bentler, 2008). PVR values were averaged between ears for subsequent analysis.

### **Speech Perception**

Speech perception in noise was assessed using the Multimodal Lexical Sentence Test for adults (MLST-A; Kirk et al., 2012) without HAs and with HAs worn in their typical volume and program settings for noisy environments. The MLST-A is a clinically available measure of speech recognition (GN Otometrics; LIPread) using multiple-target talkers, multiple presentation formats (audio, audio–video, video), and optional background noise. The test was composed of 320 sentences, which were presented in 30 lists of 12 sentences each. Some sentences were repeated. Each sentence was seven to nine words in length and had three key words that were used for scoring purposes. Five women and five men serve as the talkers, representing diverse racial, ethnic, and geographical backgrounds. Individual lists are equated for intelligibility, regardless of the presentation format (unpublished data; K. Kirk, personal communication, November 5, 2015). Sentences were presented through a loudspeaker (Tannoy Di5t) at 0° azimuth at 65 dB SPL with the participant seated 1 m from the speaker. Only the AO and AV presentation formats were used; the visual-only condition was not tested. A 15-in. monitor mounted directly below the 0° azimuth speaker presented the face of the person speaking each sentence during AV conditions.

Two types of background noise were used during speech recognition testing, as it was hypothesized that ripple sensitivity might be more predictive of speech perception in modulated backgrounds than unmodulated backgrounds. Steady-state noise (SSN) was created from white noise, which was shaped to the long-term average spectrum of the speech stimuli in one-third octave bands. A four-talker babble (4TB) was made from the International Speech Test Signal (Holube, Fredelake, Vlaming, & Kollmeier, 2010). Babble composed of a language that is understandable to the listener leads to greater degrees of masking than with a babble composed of a foreign language that is nonintelligible to the listener (Van Engen & Bradlow, 2007). Our primary interest was in the effects from multiple-target talkers; therefore, we wanted to minimize linguistic effects also occurring from the masker. To create the 4TB noise

condition, the International Speech Test Signal was temporally offset and wrapped to create four uncorrelated samples. The noise was presented via four loudspeakers positioned at 0°, 90°, 180°, and 270° azimuth and 1 m from the participant at output levels to create an overall +8 dB signal-to-noise ratio (i.e., 57 dB SPL) at the location of the participant's head. The decision to present speech at 65 dB SPL and noise at a +8 dB signal-to-noise ratio was based on speech and noise levels representative of those experienced in the real world (Pearsons, Bennett, & Fidell, 1977; Smeds, Wolters, & Rung, 2015; Wu et al., 2018). In total, eight MLST-A conditions were tested in a randomized order across participants and blocked in aided and unaided conditions (unaided tested first). Noise was presented throughout each test block, and the subject was tasked to repeat any words they could understand immediately following the sentence. The repeated keywords had to be identical to the target keywords to be counted correct (e.g., plurals were not counted as correct). Two lists were presented and averaged for each condition, which meant that the average of correct keywords over 24 sentences was computed.

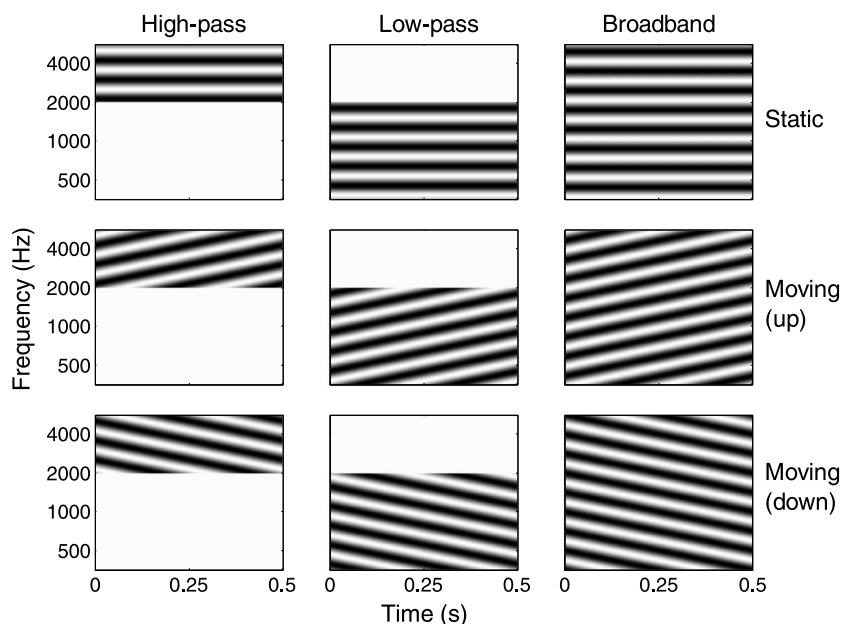
### Ripple Sensitivity

Ripple stimuli were created generally as described by Bernstein et al. (2013) and Mehraei et al. (2014). Figure 1 illustrates each spectral ripple condition at full modulation depth. Spectral and spectrotemporal modulation was applied to three different noise carriers: broadband (BB), low-pass (LP), and high-pass (HP). Each carrier consisted of equal-amplitude random-phase tones (1,000 per octave) equally

spaced along the logarithmic frequency axis. The BB carrier spanned four octaves (354–5656 Hz). The LP and HP carriers spanned 2.5 (354–2000 Hz) and 1.5 octaves (2000–5656 Hz), respectively. These bandwidths were chosen to replicate Bernstein et al. (2016) and to maintain thresholds below 0 dB for as many listeners as possible. Static spectral ripples (0-Hz modulation rate) were applied by varying the amplitudes of the carrier tones across the spectrum at a rate of 2 cycles/octave. Moving spectral ripples (4 Hz) were applied by applying temporal modulation to each of the carrier tones, while shifting the phase of the modulation applied to each tone to generate a 2-cycle/octave spectral ripple density. Furthermore, the moving ripple conditions were tested under both upward and downward directions. The 2-cycle/octave, 4-Hz stimulus was chosen because this rate–density combination was most sensitive to performance differences between listeners with normal hearing and those with hearing impairment (Bernstein et al., 2013).

A MATLAB code generated and delivered the stimuli diotically to the listener. In each interval, the root-mean-square level was chosen from a uniform distribution between –2.5 and 2.5 dB of the 85 dB SPL nominal level (i.e., 82.5–87.5 dB SPL). Each participant wore headphones (unaided) and was seated in front of a touch screen monitor. The task was to select the button that corresponded to the sound with modulations in a two-alternative forced-choice task. For the moving ripple conditions, there were two intervals: one containing the modulated stimulus and the other containing an unmodulated noise reference. For the static ripple conditions, there were three intervals: one containing the

**Figure 1.** Example spectrograms for the nine spectrotemporal modulation conditions tested in the experiment at 100% modulation depth (white area shows minimum magnitude; dark areas show maximum magnitude). Static stimuli (0 Hz) are shown in the top row, whereas upward- and downward-moving stimuli (4 Hz) are shown in the bottom two rows. Columns represent the different noise-carrier bandwidths tested.



modulated stimulus and the other two intervals containing the unmodulated reference. In this case, the first interval was always an unmodulated reference, and the modulated stimulus could only be in the second or third interval. The modulation depth varied in a three-down, one-up adaptive procedure tracking the 79.4% correct point (Levitt, 1971). Listeners completed two training blocks to familiarize themselves with the procedure and two test blocks per condition. For the training blocks, the temporal rate was set at either 0 or 4 Hz, and the spectral density was set to 1 cycle/octave. The modulation depth was defined in decibels as  $20 \log_{10} m$ , where  $m$  is the linear depth (e.g., 0 dB = full modulation). The stimuli started at full modulation and changed by 6 dB until the first reversal point, then changed by 4 dB for the next two reversals, and then by 2 dB for the final six reversals. The ripple detection threshold was estimated to be the mean modulation depth for the last six reversal values. The ripple metric used in the analysis was the average between the two test blocks. Test conditions were presented in a random order.

The maximum modulation depth was 0 dB. If a greater depth was required on any given trial, the modulation depth was maintained at 0 dB for that trial. Following more than three incorrect responses at full modulation depth, it was assumed the listener could not achieve the target performance level of 79.4%, and the test reverted to a method of fixed stimuli where the listener completed an additional 40 trials. For these runs, percent correct scores were transformed to equivalent detection thresholds (Bernstein et al., 2016). For each listener and condition, the trial-by-trial data were plotted as psychometric functions ( $d'$  vs. modulation depth), and the slope was calculated. The slopes were very similar for the six BB and LP conditions ( $0.21d'$  points per dB) and for the three HP conditions ( $0.24d'$  points per dB). A percent correct score was transformed into an equivalent modulation depth threshold by taking the difference between the measured  $d'$  value and the  $d'$  value associated with the tracked score of 79.4% and dividing by the mean slope. Percent correct scores below the tracked value of 79.4% were transformed into equivalent modulation depth values greater than 0 dB.

## Results

### Participants

Table 1 reports participants' age, audiometric, and HA details. Pure-tone hearing thresholds were averaged into a low-frequency average (LF-PTA; 250, 500, 1000 Hz) and a high-frequency average (HF-PTA; 2000, 4000, 6000 Hz). Results from Bernstein et al. (2016) and Humes (2007) showed that the HF-PTA significantly explained variance in speech perception scores whereas the LF-PTA did not.

### Speech Recognition

Percent correct absolute scores were transformed to rationalized arcsine units (rau; Studebaker, 1985) to stabilize the variance. Values of mean (and standard deviation) performance on the MLST-A for all conditions, within and across test site, are shown in Table 2. Because there were large differences in mean speech recognition performance between the listeners tested at the two sites, test site was included as a factor in the statistical analyses of both the speech recognition and the STM data. A mixed, repeated-measures analysis of variance was performed on unaided data with noise type (SSN, 4TB) and presentation format (AO, AV) as within-subject factors and test site (UW, UI) as a between-subjects factor. The main and interaction effects of noise type and visual cues were all significant: noise type,  $F(1, 65) = 28.57, p < .001, \eta_p^2 = .31$ , visual cues,  $F(1, 65) = 293.89, p < .001, \eta_p^2 = .82$ , and the interaction of noise and visual cues,  $F(1, 65) = 8.26, p = .005, \eta_p^2 = .11$ . The main effect of test site was also significant,  $F(1, 65) = 9.969, p = .002, \eta_p^2 = .13$ ; however, the interaction effects with test site were not significant for noise ( $p = .538$ ), visual cues ( $p = .066$ ), or Noise  $\times$  Visual Cues ( $p = .757$ ).

The same analysis was repeated for the aided data. The main effects of condition were all significant: noise type,  $F(1, 65) = 11.85, p = .001, \eta_p^2 = .15$ , and visual cues,  $F(1, 65) = 121.51, p < .001, \eta_p^2 = .65$ ; but the interaction of Noise  $\times$  Visual Cues was not significant ( $p = .050$ ). The main effect of test site was also significant,  $F(1, 65) = 31.13$ ,

**Table 1.** Demographics for the 67 participants.

Variable	Minimum	Maximum	<i>M</i>	<i>SD</i>
Age (years)	29.0	79.0	66.9	10.7
Air-conduction threshold (dB HL)	0.0	60.0	26.4	13.4
250	0.0	70.0	31.0	14.9
500	12.5	75.0	39.4	14.4
1,000	17.5	72.5	48.3	11.5
2,000	30.0	80.0	54.9	10.3
4,000	15.0	102.5	56.6	14.9
6,000	29.0	79.0	66.9	10.7
Low-frequency PTA (0.25, 0.5, 1.0 kHz; dB HL; better ear)	6.7	66.7	34.1	12.9
High-frequency PTA (2.0, 4.0, 6.0 kHz; dB HL; better ear)	35.0	76.7	55.6	9.9
Aided SII (65 dB SPL input; better ear)	32.0	90.0	62.4	12.9
Peak-to-valley ratio (65 dB SPL input; binaural average)	20.2	29.3	23.3	1.9

Note. PTA = pure-tone average; SII = Speech Intelligibility Index.

**Table 2.** Speech perception in noise (rationalized arcsine units) on the Multimodal Lexical Sentence Test across eight conditions and two test sites.

Condition	Speech-shaped noise		Four-talker babble	
	Auditory-only	Auditory-visual	Auditory-only	Auditory-visual
University of Washington ( $n = 40$ )				
Unaided, $M$ ( $SD$ )	41.32 (28.12)	69.86 (28.98)	34.70 (28.69)	67.18 (27.66)
Aided, $M$ ( $SD$ )	58.34 (23.93)	79.40 (24.53)	50.30 (22.17)	73.69 (25.48)
University of Iowa ( $n = 27$ )				
Unaided, $M$ ( $SD$ )	66.13 (28.83)	88.19 (21.90)	57.81 (32.52)	84.77 (23.96)
Aided, $M$ ( $SD$ )	85.85 (12.98)	97.09 (10.20)	77.32 (15.78)	96.84 (11.42)
Average ( $n = 67$ )				
Unaided, $M$ ( $SD$ )	51.32 (30.74)	77.24 (27.70)	44.01 (32.14)	74.27 (27.45)
Aided, $M$ ( $SD$ )	69.43 (24.28)	86.53 (21.75)	61.19 (23.81)	83.02 (23.79)

$p < .0001$ ,  $\eta_p^2 = .32$ ; however, the interaction effects with test site were not significant for noise ( $p = .450$ ), visual cues ( $p = .050$ ), or Noise  $\times$  Visual Cues ( $p = .263$ ).

Regarding the significant main effect of test site, calibration procedures, MATLAB code, equipment setup, protocol, and loudspeakers were identical between the two sites, and repeated equipment and procedural checks were made throughout the experiment. We further ruled out hearing loss in the examiners, participants' Montréal Cognitive Assessment scores and hearing threshold levels, and calibration of the sound-level meters as sources of site differences. To evaluate if deviations in instructions between examiners were responsible, four subjects returned to the lab for repeat testing (6 months to 1 year after initial testing), and instructions were read verbatim. Test results between initial and repeat testing were highly correlated for all conditions (averaged across MLST-A conditions, mean difference =  $-3.3$  rau,  $r = .99$ ), suggesting deviations in instructions were unlikely to be the cause of site differences. Without normative data for the MLST-A on this population, it was undeterminable which site contained outliers. Furthermore, when we examined the highest performers from UI and the lowest performers from UW, there were no demographic variables (e.g., extreme age range) or other explanations (e.g., low noise tolerance) for their performance. To investigate whether these differences in mean speech recognition performance affected the outcome of the study, the analyses comparing the STM and speech scores were conducted both with and without test site included as a predictor variable.

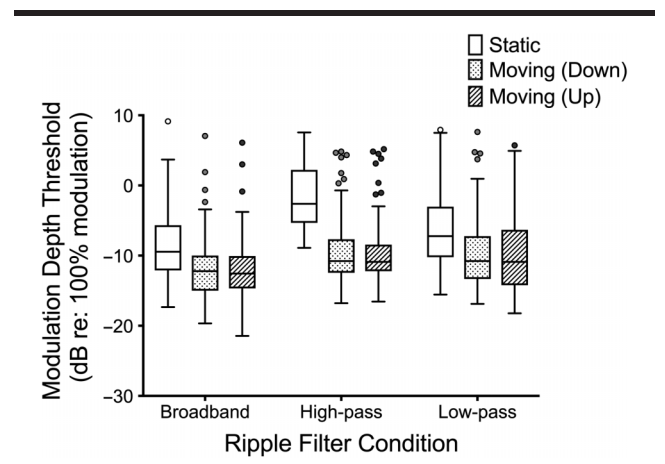
### Spectral Ripple Sensitivity

The modulation depth thresholds for the nine conditions tested are shown in Figure 2. A mixed, repeated-measures analysis of variance was used to test for differences between filter type (BB, HP, LP) and direction of the moving component (static, down, up), with test site (UW, UI) as the between-subjects variable. No significant main effects or interactions involving test site were found ( $p > .05$ ). The within-subject tests for main and interaction effects were all significant, which means that modulation depth thresholds differed depending on the task condition. There were

significant main effects of filter condition,  $F(1.42, 92.34) = 26.20$ ,  $p < .0001$ ,  $\eta_p^2 = .29$ , and direction,  $F(1.34, 86.94) = 233.47$ ,  $p < .0001$ ,  $\eta_p^2 = .78$ , and a significant interaction between filter and direction,  $F(3.11, 202.09) = 42.48$ ,  $p < .0001$ ,  $\eta_p^2 = .40$ . The interaction was further analyzed through a series of paired comparisons of direction within each filter condition, with Bonferroni corrections applied for (nine) multiple comparisons. Performance in the static condition was found to be significantly poorer ( $p < .05$ ) than in either of the moving ripple conditions for all three filter conditions, but there were no significant differences found between the upward- and downward-moving ripple conditions. Therefore, the results of the upward and downward thresholds were averaged for the remaining analysis. Paired samples  $t$  tests also showed significant differences between thresholds for each filter condition within static and moving stimuli ( $p < .0001$ ), with the exception of the comparison between HP and LP moving thresholds ( $p = .638$ ).

Although detection thresholds were statistically different between the six STM conditions (BB, HP, LP; static and

**Figure 2.** Modulation depth detection thresholds for all spectrotemporal modulation conditions. Boxes represent 25th to 75th percentile with the middle bar representing the 50th percentile. Potential Tukey outliers are noted as circles.



**Table 3.** Pearson's correlations and associated *p* values between ripple conditions.

Condition	Static ripples			Moving ripples		
	Broadband	High-pass	Low-pass	Broadband	High-pass	Low-pass
Static ripples						
Broadband		.414**	.851**	.855**	.388**	.817**
High-pass			.235	.464**	.715**	.253*
Low-pass				.664**	.284*	.883**
Moving ripples						
Broadband					.541**	.794**
High-pass						.423**
Low-pass						

\**p* < .05. \*\**p* < .001.

moving), a simple correlation analysis showed they were all correlated to each other ( $r = .23-.88, p < .05$ ; see Table 3). To better understand the collinearity or clustering between STM conditions and to answer the first question of whether detection of moving and static ripples reflects similar underlying functions, a principal components analysis was performed on the six STM conditions. Initial eigenvalues indicated that only the first two components had values of > 1.0. These first two components explained 87% of the total variance in STM thresholds, indicating that the remaining four components did not explain a significant amount of variance in STM thresholds (see Table 4). Primary factor loadings revealed that static and moving conditions for both the BB and LP stimuli loaded onto Factor 1 and accounted for 66% of the total variance. Both static and moving conditions of the HP stimuli loaded onto Factor 2 and accounted for 21% of the total variance. Secondary factor loadings were all below .3, suggesting that two distinct factors were underlying detection thresholds. For example, as shown in Table 4, the BB static variable had a high factor loading of .91 on Factor 1 and a low factor loading of .27 on Factor 2, and vice versa, the HP static variable loaded high on to Factor 2 (.92) and low onto Factor 1 (.14). Furthermore, the communalities for all STM variables entered into the principal component analysis were quite high (greater than .83), indicating that the two-factor solution is adequate to account for the variances in the six conditions (Gorsuch, 2013). See Humes (2003) for a tutorial on

the use of principal components factor analysis in HA outcome studies.

### Associations Between Predictors and Outcomes

A mixed linear model was used to separately analyze the unaided and aided speech perception data. Fixed effects considered in the unaided model were noise type (SSN, 4TB), visual cues (AO, AV), test site (UW, UI), age, LF-PTA, HF-PTA, Factor 1 (mainly reflecting BB and LP sensitivity), and Factor 2 (mainly reflecting HP sensitivity). A random intercept for participants was also included in the model. For the aided model, the predictors were the same as in the unaided model and also included two additional predictors to characterize the effects of HA processing: audibility (SII) and compression effects (PVR).

The unaided model showed significant main effects of visual cues, noise type, LF-PTA, HF-PTA, Factor 1 (mainly reflecting BB and LP), and test site ( $p < .01$ ; see Table 5 for details). The aided model revealed significant main effects of visual cues, noise type, SII, Factor 1, and site ( $p < .05$ ; see Table 6 for details). For both the aided and unaided analysis, all interactions were also examined, and none were significant ( $p > .05$ ) and therefore were not included in the final model. When the analyses were repeated without test site included as a variable, the results remained stable with very similar estimates and *p* values. The effects of Factor 1 on unaided and aided

**Table 4.** Rotated (varimax) factor weighting from a principal components analysis on depth detection thresholds across spectrotemporal modulation conditions.

Condition	Factor 1: BB and LP (66%)	Factor 2: HP (21%)	Final communality estimate
BB static	.91*	.27	.90
HP static	.14	.92*	.86
LP static	.94*	.06	.89
BB moving	.80*	.43	.83
HP moving	.23	.89*	.84
LP moving	.93*	.17	.90

Note. The percentage of variance accounted for by each of the two principal factors is indicated in parentheses. The asterisks (\*) indicate factor loadings greater than .30. BB = broadband; HP = high-pass; LP = low-pass.



**Table 5.** Effect of factors in the final linear mixed model analysis (with compound symmetric covariance matrix) for unaided outcomes.

Effect	Estimate	SE	df	t	p value
Intercept	103.05	21.0046	60	4.91	< .0001
Visual cues (AO or AV)	28.0899	1.1703	66	24.00	< .0001
Noise type (SSN or 4TB)	-5.1416	1.1703	66	-4.39	< .0001
Age	0.4021	0.2072	60	1.94	.0570
LF-PTA	-0.7042	0.2076	60	-3.39	.0012
HF-PTA	-0.7347	0.2786	60	-2.64	.0106
Factor 1 (BB & LP)	-10.6062	2.4488	60	-4.33	< .0001
Factor 2 (HP)	-0.3454	2.6868	60	-0.13	.8981
Site (UW or UI)	-24.9093	4.3984	60	-5.66	< .0001

*Note.* AO = auditory-only; AV = auditory-visual; SSN = steady-state noise; 4TB = four-talker babble; LF = low frequency; HF = high frequency; PTA = pure-tone average; BB = broadband; LP = low-pass; HP = high-pass; UW = University of Washington; UI = University of Iowa.

speech perception scores are displayed in Figures 3 and 4, respectively.

## Discussion

The goal of the current study was to determine if static and moving spectral ripple thresholds were predictive of speech perception outcomes with and without HAs, using ecologically relevant speech perception stimuli. The results of a principal components analysis suggest that thresholds for the four STM conditions involving static or moving ripples and BB or LP filters were highly correlated to one another and, therefore, likely reflect a similar mechanism of impairment. Likewise, thresholds for the two STM conditions involving static and moving ripples and an HP filter were highly correlated to one another and, therefore, likely reflect a different underlying mechanism of impairment. While controlling for audibility and age, the BB and LP factor accounted for significant variance in speech outcomes for both unaided and aided models, whereas the HP factor was not significant. These results remained stable regardless

of the type of noise or the addition of visual cues, suggesting robust effects of spectral ripple modulation sensitivity.

The first question answered with this study was whether the moving aspect of the ripple stimulus added unique information about listeners' speech perception abilities over that provided by static ripple sensitivity. It was hypothesized that the temporal component of the ripple would better reflect speech, which is both spectrally and temporally dynamic. However, the current results did not support this hypothesis, with thresholds for both types of ripple stimuli loading onto the same factor. These results imply that performance in detecting spectral ripples reflects a similar underlying mechanism or mechanisms for both static and moving ripples. Although moving spectral ripples are closer to the dynamic nature of speech than static spectral ripples, these stimuli do not capture the continually changing spectral and temporal modulation in a speech spectrogram, and it is possible that different results would have been observed with different stimulus characteristics.

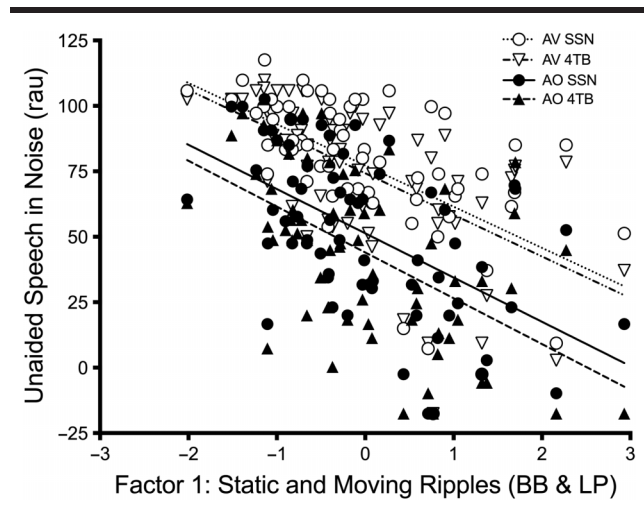
As discussed in the introduction, a reduced ability to detect STM has been attributed both to reduced frequency

**Table 6.** Effect of factors in the final linear mixed model analysis (with compound symmetric covariance matrix) for aided outcomes.

Effect	Estimate	SE	df	t	p value
Intercept	84.9600	32.7276	53	2.60	.0122
Visual cues (AO or AV)	19.8734	1.6537	61	12.02	< .0001
Noise type (SSN or ISTS)	-5.6818	1.6537	61	-3.44	.0011
Age	-0.3094	0.1674	53	-1.85	.0701
LF-PTA	0.02993	0.1847	53	0.16	.8718
HF-PTA	-0.3077	0.2405	53	-1.28	.2064
SII	0.3501	0.1696	53	2.06	.0439
PVR	0.5252	1.0967	53	0.48	.6340
Factor 1 (BB & LP)	-6.6223	1.9851	53	-3.34	.0016
Factor 2 (HP)	0.6581	2.2176	53	0.30	.7678
Site (UW or UI)	-25.7804	4.0527	53	-6.36	< .0001

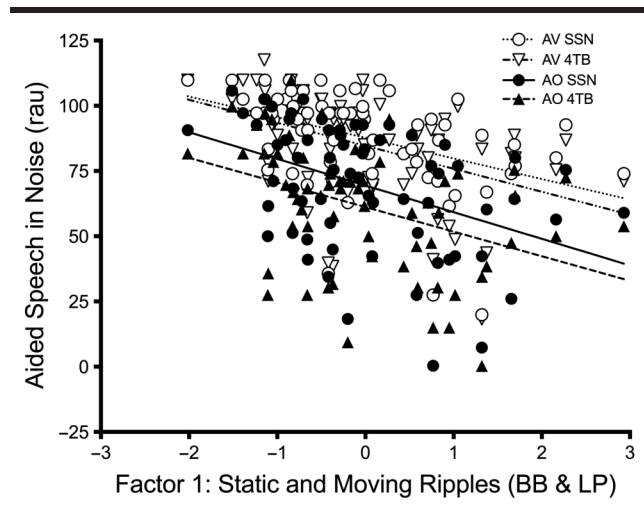
*Note.* AO = auditory-only; AV = auditory-visual; SSN = steady-state noise; ISTS = International Speech Test Signal; LF = low frequency; HF = high frequency; PTA = pure-tone average; SII = Speech Intelligibility Index; PVR = peak-to-valley ratio; BB = broadband; LP = low-pass; HP = high-pass; UW = University of Washington; UI = University of Iowa.

**Figure 3.** Factor scores for Factor 1 (broadband [BB] and low-pass [LP]) are plotted against unaided speech perception scores. Individual data and linear regression lines are shown for the auditory-only (AO; filled symbols, bold lines) and auditory–visual (AV; open symbols, dotted lines) speech perception conditions in backgrounds of steady-state noise (SSN; circles) and four-talker babble (4TB; triangles).



selectivity (at high frequencies) and to a reduced ability to make use of TFS information in the stimulus waveform (at low frequencies; Bernstein et al., 2013; Mehraei et al., 2014). Bernstein et al. (2013) demonstrated significant associations between moving ripple sensitivity and frequency selectivity at 4000 Hz, whereas Davies-Venn et al. (2015) showed similar support for static ripple performance using

**Figure 4.** Factor scores for Factor 1 (broadband [BB] and low-pass [LP]) are plotted against aided speech perception scores. Individual data and linear regression lines are shown for the auditory-only (AO; filled symbols, bold lines) and auditory–visual (AV; open symbols, dotted lines) speech perception conditions in backgrounds of steady-state noise (SSN; circles) and four-talker babble (4TB; triangles).



auditory filter bandwidths. However, the pattern of results observed in the current study seem to point more strongly to an explanation based on TFS than an explanation based on frequency selectivity. There was no significant relationship between speech perception in noise and ripple detection for the HP conditions, where impaired frequency selectivity is more likely to occur (Glasberg & Moore, 1986). In contrast, the stimuli for the LP and BB conditions both include low-frequency content, where phase-locking information tends to be strongest (Johnson, 1980), and it was this component that showed significant effects on speech perception in noise. Furthermore, Mehraei et al. (2014) found no relationship between STM sensitivity and frequency selectivity for stimuli below 2000 Hz. The current results are therefore consistent with others in the literature, suggesting an important role for TFS processing in facilitating speech understanding in noise or in the presence of interfering talkers (e.g., Bernstein et al., 2013, 2016; Lorenzi et al., 2006; Strelcyk & Dau, 2009). One caveat to this argument is that the STM stimuli in the HP condition were presented at lower audibility levels than the LP or BB conditions due to more severe high-frequency hearing loss; therefore, different results may have occurred if testing was performed at equated audibility levels.

Although the psychophysical literature highlights the importance of TFS processing, the mechanism by which the TFS deficit could arise remains unclear. There are many physiological studies in humans showing that hearing loss alters the neural representation of temporal and spectral cues; however, human studies use far-field recordings that involve multiple neural generators, which makes it difficult to identify the source of impaired processing. Animal studies provide a greater opportunity to isolate underlying mechanisms; however, they have not reported phase-locking deficits for individual auditory nerve fiber responses in animals with hearing impairment (e.g., Kale & Heinz, 2010), except as a result of more noise passing through the auditory nerve fiber response due to broader tuning (Henry & Heinz, 2012). It has been argued that the apparent lack of ability to use TFS information associated with hearing loss could be attributable to decreased conversion of TFS to temporal envelope information (Moore & Sek, 1994) due to broader filtering, although models of TFS and envelope processing suggest that normal hearing TFS processing ability cannot be explained by this mechanism (Ewert, Paraouty, & Lorenzi, 2018). Overall, the mechanisms underlying TFS processing and spectral resolution are not easily separable, and we cannot rule out the possible role of individual differences in frequency selectivity below 2000 Hz in the current study. In any case, the strong correlations observed here suggest that static and slow-moving spectral ripple discrimination performance may reflect the same basic underlying mechanism.

The second purpose of this study was to extend previous research examining the relationship between STM sensitivity and speech understanding in noise to more ecologically valid situations including speech produced by different target talkers, the availability of visual cues, and different noise types. The results showed that the association between

STM sensitivity and speech perception generalized to these other speech tasks, regardless of whether the listener was aided or not. Furthermore, the interactions between the LP and BB components were not significant for noise type or visual cue, suggesting that the importance of STM sensitivity did not change under different conditions. Collectively, these findings propose that the variability in performance explained by the distorted auditory system are robust across many types of speech communication environments. There are several potential advantages of using STM stimuli to predict speech perception instead of speech itself: (a) STM detection presumably relies most heavily on peripheral abilities, with less reliability on cognitive and linguistic processing, and (b) native English-speaking abilities are not required, so a wider population could be reached.

The lack of age effect on speech perception in noise in the current study was somewhat surprising given previous reports of older adults performing more poorly in noise than younger adults (e.g., Dubno, 2015; Humes, 2015). However, the age range was not evenly distributed. The mean age was 67 (see Table 1), and only four HA users were recruited who fit our inclusion criteria, who were under the age of 50 years. The  $p$  values for age were .05 (unaided) and .07 (aided), suggesting an age effect may have been observed had a larger number of younger adults been included in the sample. These results may not be representative of the general population, specifically to individuals with hearing loss who do not own HAs. In addition, we did not measure vision abilities, and although all participants read and signed a consent form, it is possible that some individuals were unable to take full advantage of the visual cues provided in the AV conditions.

It is also important to point out that the listeners in the study wore their HAs for the speech testing, but not for the STM test. This was done to specifically investigate the contribution of psychoacoustic factors related to the health of the auditory system, rather than factors related to individual differences in HA signal processing. Although two types of HA effects were controlled in the statistical analysis—audibility (i.e., the SII) and compression (i.e., the PVR)—other possible sources of distortion were not controlled (e.g., frequency lowering, noise reduction). It is possible that different effects would be observed with more sophisticated methods for modeling compression (e.g., Hearing Aid Speech Perception Index; Kates, Arehart, Anderson, Muralimanohar, & Harvey, 2018). Theoretically, HA distortion could interact with a listener's sensitivity to TFS cues, in that individuals with better TFS sensitivity have been shown to have better outcomes with fast compression processing (Moore, 2008; Stone, Füllgrabe, & Moore, 2009; Stone & Moore, 2007). The benefit in audibility that fast compression provides may partially compensate for any distortion in the temporal envelope, and listeners with adequate TFS processing will be able to take advantage of the improved audibility. Conversely, those with poor access to TFS cues may do better with less distortion (i.e., slow compression processing). For this reason, it could be of interest to examine the relationship between

HA outcomes and ripple detection thresholds with stimuli presented through the HA. This could help determine whether spectral ripple stimuli can characterize individual variability in the distortion imparted by HA processing and whether this affects HA outcomes.

## Conclusions

This study demonstrated that the relationship between ripple sensitivity and speech understanding in noise is generalizable to audiovisual speech conditions, different types of background noise, variance in target talkers, and aided or unaided listening. It was also shown that sensitivity to static and moving ripples appear to reflect similar underlying functions, regardless of the frequency content of the stimuli. However, only ripple stimuli consisting of low-frequency content showed a significant relationship to speech perception in noise. This pattern of results is consistent with the idea that, for listeners with hearing impairment, ripple detection performance for both static and slow-moving ripples and the associated speech understanding in noise are related to an ability to make use of TFS information in the stimulus waveform. Unexplained variance might be attributable to HA distortion factors, a topic for future study.

## Acknowledgments

We thank our funding sources for making this work possible: the American Speech-Language-Hearing Foundation (C. M.), NIH NIDCD R21 DC016380-01 (C. M.), NIH NIDCD R01 DC012769-04 (K. T. and R. B.), NIH NIDCD P30 DC004661 (Edwin Rubel), and NIH NCATS U54 TR001356 (Gary Rosenthal). We would like to thank the community practitioners for advertising our study and the participants for their time. We also thank Elizabeth Stangl, Claire Jordan, and Gina Hone for collecting data and Elisabeth Went for data analysis assistance. The identification of specific products or scientific instrumentation does not constitute endorsement or implied endorsement on the part of the authors, Department of Defense, or any component agency. The views expressed in this presentation are those of the authors and do not reflect the official policy of the Department of Army/Navy/Air Force, Department of Defense, or U.S. Government.

## References

- Akeroyd, M. A. (2008). Are individual differences in speech reception related to individual differences in cognitive ability? A survey of twenty experimental studies with normal and hearing-impaired adults. *International Journal of Audiology, 47*, S53–S71.
- American National Standards Institute. (1997). *Methods for calculation of the speech intelligibility index* (ANSI S3.5). New York, NY: Author.
- American National Standards Institute. (2003). *American national standard specification of hearing aid characteristics* (ANSI S3.22-2003). New York, NY: Author.
- Anderson, E. S., Oxenham, A. J., Nelson, P. B., & Nelson, D. A. (2012). Assessing the role of spectral and intensity cues in spectral ripple detection and discrimination in cochlear-implant users. *The Journal of the Acoustical Society of America, 132*(6), 3925–3934.

- Baer, T., & Moore, B. C. J. (1993). Effects of spectral smearing on the intelligibility of sentences in noise. *The Journal of the Acoustical Society of America*, 94(3), 1229–1241.
- Baer, T., & Moore, B. C. J. (1994). Effects of spectral smearing on the intelligibility of sentences in the presence of interfering speech. *The Journal of the Acoustical Society of America*, 95(4), 2277–2280.
- Bernstein, J. G. W., Danielsson, H., Hällgren, M., Stenfelt, S., Rönnerberg, J., & Lunner, T. (2016). Spectrotemporal modulation sensitivity as a predictor of speech-reception performance in noise with hearing aids. *Trends in Hearing*, 20, 1–17.
- Bernstein, J. G. W., & Grant, K. W. (2009). Auditory and auditory-visual intelligibility of speech in fluctuating maskers for normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, 125(5), 3358–3372.
- Bernstein, J. G. W., Mehraei, G., Shamma, S., Gallun, F. J., Theodoroff, S. M., & Leek, M. R. (2013). Spectrotemporal modulation sensitivity as a predictor of speech intelligibility for hearing-impaired listeners. *Journal of the American Academy of Audiology*, 24(4), 293–306.
- Bor, S., Souza, P., & Wright, R. (2008). Multichannel compression: Effects of reduced spectral contrast on vowel identification. *Journal of Speech, Language, and Hearing Research*, 51, 1315–1327.
- Broadbent, D. E. (1952). Listening to one of two synchronous messages. *Journal of Experimental Psychology*, 44, 51–55.
- Brox, J. P. L., & Noteboom, S. G. (1982). Intonation and the perceptual separation of simultaneous voices. *Journal of Phonetics*, 10, 23–36.
- Buss, E., Hall, J. W., & Grose, J. H. (2004). Temporal fine-structure cues to speech and pure tone modulation in observers with sensorineural hearing loss. *Ear and Hearing*, 25(3), 242–250.
- Cabrera, L., Tsao, F. M., Gnansia, D., Bertoncini, J., & Lorenzi, C. (2014). The role of spectrotemporal fine structure cues in lexical-tone discrimination for French and Mandarin listeners. *The Journal of the Acoustical Society of America*, 136(2), 877–882.
- Chi, T., Gao, Y., Guyton, M. C., Ru, P., & Shamma, S. (1999). Spectrotemporal modulation transfer functions and speech intelligibility. *The Journal of the Acoustical Society of America*, 106(5), 2719–2732.
- Chi, T., Ru, P., & Shamma, S. A. (2005). Multiresolution spectrotemporal analysis of complex sounds. *The Journal of the Acoustical Society of America*, 118(2), 887–906.
- Davies-Venn, E., Nelson, P., & Souza, P. (2015). Comparing auditory filter bandwidths, spectral ripple modulation detection, spectral ripple discrimination, and speech recognition: Normal and impaired hearing. *The Journal of the Acoustical Society of America*, 138(1), 492–503.
- Dryden, A., Allen, H. A., Henshaw, H., & Heinrich, A. (2017). The association between cognitive performance and speech-in-noise perception for adult listeners: A systematic literature review and meta-analysis. *Trends in Hearing*, 21, 1–21.
- Dubno, J. R. (2015). Speech recognition across the life span: Longitudinal changes from middle-age to older adults. *American Journal of Audiology*, 24(2), 84–87.
- Dubno, J. R., Dirks, D. D., & Morgan, D. E. (1984). Effects of age and mild hearing loss on speech recognition in noise. *The Journal of the Acoustical Society of America*, 76(1), 87–96.
- Elhilali, M., Taishih, C., & Shamma, S. A. (2003). A spectrotemporal modulation index (STMI) for assessment of speech intelligibility. *Speech Communication*, 41, 331–348.
- Ewert, S. D., Paraouty, N., & Lorenzi, C. (2018). A two-path model of auditory modulation detection using temporal fine structure and envelope cues. *European Journal of Neuroscience*, 2018. Epub ahead of print. <https://doi.org/10.1111/ejn.13846>
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G\*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41, 1149–1160.
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G\*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175–191.
- Glasberg, B. R., & Moore, B. C. J. (1986). Auditory filter shapes in subjects with unilateral and bilateral cochlear impairments. *The Journal of the Acoustical Society of America*, 79, 1020–1033.
- Gorsuch, R. L. (2013). *Factor analysis* (6th ed.). New York, NY: Taylor and Francis.
- Grant, K. W., & Seitz, P. F. (2000). The use of visible speech cues for improving auditory detection of spoken sentences. *The Journal of the Acoustical Society of America*, 108, 1197–1208.
- Hall, J. W., & Fernandes, M. A. (1983). Temporal integration, frequency resolution, and off-frequency listening in normal-hearing and cochlear-impaired listeners. *The Journal of the Acoustical Society of America*, 74(4), 1172–1177.
- Helfer, K. S., & Freyman, R. L. (2005). The role of visual speech cues in reducing energetic and informational masking. *The Journal of the Acoustical Society of America*, 117(2), 842–849.
- Helfer, K. S., & Vargo, M. (2009). Speech recognition and temporal processing in middle-aged women. *Journal of the American Academy of Audiology*, 20(4), 264–271.
- Henning, R. L., & Bentler, R. A. (2008). The effects of hearing aid compression parameters on the short-term dynamic range of continuous speech. *Journal of Speech, Language, and Hearing Research*, 51(2), 471–484.
- Henry, K. S., & Heinz, M. G. (2012). Diminished temporal coding with sensorineural hearing loss emerges in background noise. *Nature Neuroscience*, 15(10), 1362–1364.
- Holube, I., Fredelake, S., Vlaming, M., & Kollmeier, B. (2010). Development and analysis of an International Speech Test Signal (ISTS). *International Journal of Audiology*, 49, 891–903.
- Hopkins, K., & Moore, B. C. J. (2007). Moderate cochlear hearing loss leads to a reduced ability to use temporal fine structure information. *The Journal of the Acoustical Society of America*, 122, 1055–1068.
- Hopkins, K., Moore, B. C. J., & Stone, M. A. (2008). Effects of moderate cochlear hearing loss on the ability to benefit from temporal fine structure information in speech. *The Journal of the Acoustical Society of America*, 123, 1140–1153.
- Humes, L. E. (2003). Modeling and predicting hearing aid outcome. *Trends in Amplification*, 7(2), 41–75.
- Humes, L. E. (2007). The contributions of audibility and cognitive factors to the benefit provided by amplified speech to older adults. *Journal of the American Academy of Audiology*, 18(7), 590–603.
- Humes, L. E. (2015). Age-related changes in cognitive and sensory processing: Focus on middle-aged adults. *American Journal of Audiology*, 24(2), 94–97.
- Irwin, R. J., & Purdy, S. C. (1982). The minimum detectable duration of auditory signals for normal and hearing impaired listeners. *The Journal of the Acoustical Society of America*, 71(4), 967–974.
- Jenstad, L. M., & Souza, P. E. (2005). Quantifying the effect of compression hearing aid release time on speech acoustics and intelligibility. *Journal of Speech, Language, and Hearing Research*, 48(3), 651–667.
- Jenstad, L. M., & Souza, P. E. (2007). Temporal envelope changes of compression and speech rate: Combined effects on

- recognition for older adults. *Journal of Speech, Language, and Hearing Research*, 50, 1123–1138.
- Johnson, D. H.** (1980). The relationship between spike rate and synchrony in responses of auditory-nerve fibers to single tones. *The Journal of the Acoustical Society of America*, 68, 1115–1122.
- Kale, S., & Heinz, M. G.** (2010). Envelope coding in auditory nerve fibers following noise-induced hearing loss. *Journal of the Association for Research in Otolaryngology*, 11(4), 657–673.
- Kates, J. M., Arehart, K. H., Anderson, M. C., Muralimanohar, R. K., & Harvey, L. O.** (2018). Using objective metrics to measure hearing aid performance. *Ear and Hearing*. Epub ahead of print. <https://doi.org/10.1097/AUD.0000000000000574>
- Kidd, G., Mason, C. R., & Feith, L. L.** (1984). Temporal integration of forward masking in listeners having sensorineural hearing loss. *The Journal of the Acoustical Society of America*, 75(3), 937–944.
- Kirk, K. I., Pisoni, D. B., & Miyamoto, R. C.** (1997). Effects of stimulus variability on speech perception in listeners with hearing impairment. *Journal of Speech, Language, and Hearing Research*, 40, 1395–1405.
- Kirk, K. I., Prusick, L., French, B., Gotch, C., Eisenberg, L. S., & Young, N.** (2012). Assessing spoken word recognition in children who are deaf or hard of hearing: A translational approach. *Journal of the American Academy of Audiology*, 23, 464–475.
- Leek, M. R., Dorman, M. F., & Summerfield, Q.** (1987). Minimum spectral contrast for vowel identification by normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, 81, 148–154.
- Levitt, H.** (1971). Transformed up-down methods in psychoaoustics. *The Journal of the Acoustical Society of America*, 49(2), 467–477.
- Loeb, G. E., White, M. W., & Merzenich, M. M.** (1983). Spatial cross correlation: A proposed mechanism for acoustic pitch perception. *Biological Cybernetics*, 47, 149–163.
- Lorenzi, C., Debruille, L., Garnier, S., Fleuriot, P., & Moore, B. C. J.** (2009). Abnormal processing of temporal fine structure in speech for frequencies where absolute thresholds are normal. *The Journal of the Acoustical Society of America*, 125(1), 27–30.
- Lorenzi, C., Gilbert, G., Carn, C., Garnier, S., & Moore, B. C. J.** (2006). Speech perception problems of the hearing impaired reflect inability to use temporal fine structure. *Proceedings of the National Academy of Sciences of the United States of America*, 103(49), 18866–18869.
- Lorenzi, C., & Moore, B. C. J.** (2008). Role of temporal envelope and fine structure cues in speech perception: A review. In T. Dau, T. J. M. Buchholz, J. M. Harte, & T. U. Christiansen (Eds.), *Auditory signal processing in hearing-impaired listeners* (pp. 263–272). Lautrupbjerg, Denmark: Danavox Jubilee Foundation.
- Lunner, T.** (2003). Cognitive function in relation to hearing aid use. *International Journal of Audiology*, 42, S49–S58.
- MATLAB and Statistics Toolbox** [Computer software]. (2013). Release 2013b. Natick, MA: The MathWorks, Inc.
- Mehraei, G., Gallun, F. J., Leek, M. R., & Bernstein, J. G. W.** (2014). Spectrotemporal modulation sensitivity for hearing-impaired listeners: Dependence on carrier center frequency and the relationship to speech intelligibility. *The Journal of the Acoustical Society of America*, 136(1), 301–316.
- Miller, R. L., Schilling, J. R., Franck, K. R., & Young, E. D.** (1997). Effects of acoustic trauma on the representation of the vowel /*ɛ*/ in cat auditory nerve fibers. *The Journal of the Acoustical Society of America*, 101, 3602–3616.
- Moore, B. C. J.** (2003). Speech processing for the hearing-impaired: Successes, failures, and implications for speech mechanisms. *Speech Communication*, 41, 81–91.
- Moore, B. C. J.** (2008). The choice of compression speed in hearing aids: Theoretical and practical considerations, and the role of individual differences. *Trends in Amplification*, 12, 102–112.
- Moore, B. C. J., & Sek, A.** (1994). Effects of carrier frequency and background noise on the detection of mixed modulation. *The Journal of the Acoustical Society of America*, 96(2), 741–751.
- Moore, B. C. J., & Sek, A.** (1996). Detection of frequency modulation at low modulation rates: Evidence for a mechanism based on phase locking. *The Journal of the Acoustical Society of America*, 100, 2320–2331.
- Moore, B. C. J., & Skrodzka, E.** (2002). Detection of frequency modulation by hearing-impaired listeners: Effects of carrier frequency, modulation rate, and added amplitude modulation. *The Journal of the Acoustical Society of America*, 111, 327–335.
- Mullennix, J. W., Pisoni, D. B., & Martin, C. S.** (1989). Some effects of talker variability on spoken word recognition. *The Journal of the Acoustical Society of America*, 85, 365–378.
- Nasreddine, Z. S., Phillips, N. A., Bedirian, V., Charbonneau, S., Whitehead, V., Collin, I., ... Chertkow, H.** (2005). The Montréal Cognitive Assessment (MoCA): A brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society*, 53, 695–699.
- Neher, T., Lunner, T., Hopkins, K., & Moore, B. C. J.** (2012). Binaural temporal fine structure sensitivity, cognitive function, and spatial speech recognition of hearing-impaired listeners (L). *The Journal of the Acoustical Society of America*, 131(4), 2561–2564.
- Nygaard, L. C., & Pisoni, D. B.** (1998). Talker-specific learning in speech perception. *Perception & Psychophysics*, 60, 355–376.
- Nygaard, L. C., Sommers, M. S., & Pisoni, D. B.** (1994). Speech perception as a talker-contingent process. *Psychological Science*, 5, 42–46.
- Pearsons, K. S., Bennett, R. L., & Fidell, S.** (1977). *Speech levels in various noise environments* (Report No. 600/1-77-025). Washington, DC: U.S. Environmental Protection Agency.
- Ruggero, M. A.** (1994) Cochlear delays and traveling waves: Comments on “Experimental look at cochlear mechanics.” *Audiology*, 33, 131–142.
- Smeds, K., Wolters, F., & Rung, M.** (2015). Estimation of signal-to-noise ratios in realistic sound scenarios. *Journal of the American Academy of Audiology*, 26, 183–196.
- Smooenburg, G. F.** (1992). Speech reception in quiet and in noisy conditions by individuals with noise-induced hearing loss in relation to their tone audiogram. *The Journal of the Acoustical Society of America*, 91(1), 421–437.
- Souza, P. E., & Turner, C. W.** (1998). Multichannel compression, temporal cues, and audibility. *Journal of Speech, Language, and Hearing Research*, 41(2), 315–326.
- Stone, M. A., Füllgrabe, C., & Moore, B. C. J.** (2009). High-rate envelope information in many channels provides resistance to reduction of speech intelligibility produced by multi-channel fast-acting compression. *The Journal of the Acoustical Society of America*, 126(5), 2155–2158.
- Stone, M. A., & Moore, B. C. J.** (2007). Quantifying the effects of fast-acting compression on the envelope of speech. *The Journal of the Acoustical Society of America*, 121(3), 1654–1664.
- Strelcyk, O., & Dau, T.** (2009). Relations between frequency selectivity, temporal fine-structure processing, and speech reception in impaired hearing. *The Journal of the Acoustical Society of America*, 125(5), 3328–3345.
- Studebaker, G. A.** (1985). A rationalized arcsine transform. *Journal of Speech and Hearing Research*, 28(3), 455–462.

- 
- Summers, V., & Leek, M. R.** (1994). The internal representation of spectral contrast in hearing-impaired listeners. *The Journal of the Acoustical Society of America*, *95*, 3518–3528.
- Van Engen, K. J., & Bradlow, A. R.** (2007). Sentence recognition in native- and foreign-language multi-talker background noise. *The Journal of the Acoustical Society of America*, *121*, 519–526.
- Won, J. H., Moon, I. J., Jin, S., Park, H., Woo, J., Cho, Y-S., . . . Hong, S. H.** (2015). Spectrotemporal modulation detection and speech perception by cochlear implant users. *PLoS One*, *10*(10), e0140920.
- Wu, Y. H., Stangl, E., Chipara, O., Hasan, S. S., Welhaven, A., & Oleson, J.** (2018). Characteristics of real-world signal-to-noise ratios and speech listening situations of older adults with mild-to-moderate hearing loss. *Ear and Hearing*, *39*(2), 293–304.
- Yunds, E. W., & Buckles, K. M.** (1995). Multichannel compression hearing aids: Effect of number of channels on speech discrimination in noise. *The Journal of the Acoustical Society of America*, *97*(2), 1206–1223.
- Zekveld, A. A., Rudner, M., Johnsrude, I. S., & Rönnberg, J.** (2013). The effects of working memory capacity and semantic cues on the intelligibility of speech in noise. *The Journal of the Acoustical Society of America*, *134*(3), 2225–2234.
- Zheng, Y., Escabi, M., & Litovsky, R. Y.** (2017). Spectrotemporal cues enhance modulation sensitivity in cochlear implant users. *Hearing Research*, *351*, 45–54.