Spectrotemporal Modulation Sensitivity as a Predictor of Speech Intelligibility for Hearing-Impaired Listeners

Joshua G.W. Bernstein[*], **Golbarg Mehraei**[†], **Shihab Shamma**[‡], **Frederick J. Gallun**[§], **Sarah M. Theodoroff**[§], and **Marjorie R. Leek**[§]

[*]Audiology and Speech Center, Scientific and Clinical Studies Section, Walter Reed National Military Medical Center, Bethesda, MD

[†]Harvard-MIT Speech and Hearing Bioscience and Technology Program, Cambridge, MA

[‡]Institute for Systems Research, University of Maryland, College Park, MD

[§]VA RR&D National Center for Rehabilitative Auditory Research, Portland VA Medical Center, Portland, OR

Abstract

Background—A model that can accurately predict speech intelligibility for a given hearing-impaired (HI) listener would be an important tool for hearing-aid fitting or hearing-aid algorithm development. Existing speech-intelligibility models do not incorporate variability in suprathreshold deficits that are not well predicted by classical audiometric measures. One possible approach to the incorporation of such deficits is to base intelligibility predictions on sensitivity to simultaneously spectrally and temporally modulated signals.

Purpose—The likelihood of success of this approach was evaluated by comparing estimates of spectrotemporal modulation (STM) sensitivity to speech intelligibility and to psychoacoustic estimates of frequency selectivity and temporal fine-structure (TFS) sensitivity across a group of HI listeners.

Research Design—The minimum modulation depth required to detect STM applied to an 86 dB SPL four-octave noise carrier was measured for combinations of temporal modulation rate (4, 12, or 32 Hz) and spectral modulation density (0.5, 1, 2, or 4 cycles/octave). STM sensitivity estimates for individual HI listeners were compared to estimates of frequency selectivity (measured using the notched-noise method at 500, 1000measured using the notched-noise method at 500, 2000, and 4000 Hz), TFS processing ability (2 Hz frequency-modulation detection thresholds for 500, 10002 Hz frequency-modulation detection thresholds for 500, 2000, and 4000 Hz carriers) and sentence intelligibility in noise (at a 0 dB signal-to-noise ratio) that were measured for the same listeners in a separate study.

Study Sample—Eight normal-hearing (NH) listeners and 12 listeners with a diagnosis of bilateral sensorineural hearing loss participated.

Data Collection and Analysis—STM sensitivity was compared between NH and HI listener groups using a repeated-measures analysis of variance. A stepwise regression analysis compared

STM sensitivity for individual HI listeners to audiometric thresholds, age, and measures of frequency selectivity and TFS processing ability. A second stepwise regression analysis compared speech intelligibility to STM sensitivity and the audiogram-based Speech Intelligibility Index.

Results—STM detection thresholds were elevated for the HI listeners, but only for low rates and high densities. STM sensitivity for individual HI listeners was well predicted by a combination of estimates of frequency selectivity at 4000 Hz and TFS sensitivity at 500 Hz but was unrelated to audiometric thresholds. STM sensitivity accounted for an additional 40% of the variance in speech intelligibility beyond the 40% accounted for by the audibility-based Speech Intelligibility Index.

Conclusions—Impaired STM sensitivity likely results from a combination of a reduced ability to resolve spectral peaks and a reduced ability to use TFS information to follow spectral-peak movements. Combining STM sensitivity estimates with audiometric threshold measures for individual HI listeners provided a more accurate prediction of speech intelligibility than audiometric measures alone. These results suggest a significant likelihood of success for an STM-based model of speech intelligibility for HI listeners.

Keywords

Listeners with sensorineural hearing loss (SNHL) often have difficulty understanding speech in background noise. Audibility of the speech signal, represented clinically by the audiogram, is a reasonably good predictor of an individual's speech reception performance (Smoorenburg, 1992; Rankovic, 1998; Vestergaard, 2003). Hearing aids often succeed in improving speech-reception performance for hearing-impaired (HI) listeners by restoring much of the audibility of otherwise inaudible portions of the speech signal that are thought to be important for speech intelligibility (e.g., Keidser and Grant, 2001; Scollie et al, 2010). Nevertheless, the audiogram-based prediction of speech intelligibility is imprecise (Ching et al, 1998), with the audiogram predicting only about half of the variance in intelligibility across HI individuals (Smoorenburg, 1992). In addition to audibility limitations, HI listeners also lack certain psychoacoustic abilities that might limit the discrimination of important cues in the speech signal. For example, they nearly always show reduced frequency selectivity (e.g., Glasberg and Moore, 1986), which could limit the ability to detect spectral contrast cues. They also show increased forward masking (e.g., Oxenham and Moore, 1997; Nelson et al, 2001), such that a low-level signal is more likely to be masked by a higher-level signal immediately preceding it. This could limit, for example, the ability to detect certain temporal contrasts important for the identification of consonant-release bursts. It has also been suggested that HI listeners may have reduced ability to use temporal fine-structure (TFS) information (i.e., the cycle-by-cycle variation in the stimulus frequency as coded in auditory-nerve firing patterns), which could also negatively impact speech-reception performance (e.g., Buss et al, 2004; Lorenzi et al, 2006).

A model that can accurately predict speech intelligibility for an individual HI listener is a useful tool in the optimization of hearing-aid parameters to maximize speech intelligibility. In current clinical practice, the widely used National Acoustic Laboratories' nonlinear fitting procedure, version 1 (NAL-NL1; Byrne et al, 2001), aims to maximize speech intelligibility for a given listener's audiogram as predicted by the audibility-based Speech Intelligibility Index (SII; American National Standards Institute [ANSI], 1997). Because speech-reception performance is likely to be influenced by suprathreshold processing deficits in addition to audibility factors, some attempts have been made to include suprathreshold distortion in models of impaired speech intelligibility. Zilany and Bruce (2007) simulated the reduced frequency tuning and loss of compression that accompany hearing loss and showed a general

impact on speech intelligibility. However, this approach did not attempt to simulate different degrees of suprathreshold distortions for individual listeners. Other efforts to include suprathreshold distortion effects in models of speech intelligibility include the incorporation of a nonspecific "speech desensitization factor" (Pavlovic et al, 1986) that is directly tied to the audiogram, or a "distortion factor" to account for discrepancies between model predictions and real HI listener performance (e.g., Plomp, 1986). No current model of speech intelligibility for individual listeners includes specific suprathreshold psychoacoustic distortions associated with SNHL that are not predictable from the audiogram. A model that could accurately predict speech intelligibility for individual HI listeners based on the audiogram and specific suprathreshold distortion factors could be used in at least two ways to benefit HI patients. First, the model could be a stand-in for real patients in the development of signal-processing algorithms to compensate for suprathreshold limitations associated with hearing loss. Second, the model could be used clinically in an analogous manner to the SII to individualize the parameterization of hearing-aid settings.

Speech is often characterized by its formant peaks, spectral edges, and amplitude modulations at onsets/offsets. These important features contribute to the energy modulations seen in speech spectrograms, both over time within any given frequency channel, and along the spectral axis over a given time window. Temporal modulation is defined in terms of modulation rate (in Hz), with higher rates indicating faster energy fluctuations. Spectral modulations are defined in terms of modulation density (cycles per octave [c/o]), with higher densities indicating more densely spaced spectral peaks. It has been suggested that speech intelligibility is highly dependent on low spectral modulation densities (<8 c/o) that reflect the formant structure of speech (Elhilali et al, 2003; Liu and Eddins, 2008) and low temporal modulation rates (<16 Hz) that reflect the phonetic and syllabic rate of speech (Steeneken and Houtgast, 1980; Drullman et al, 1994; Elhilali et al, 2003). While spectral and temporal modulation sensitivities are often tested separately, they often co-occur in natural sounds. Joint spectrotemporal modulation (STM) can be thought of in terms of spectral patterns that change over time or temporal modulation patterns that vary across frequency channels.

Chi et al (1999) and Elhilali et al (2003) developed a model of speech intelligibility for normal-hearing (NH) listeners based on the STM contained within speech. A speech input signal is represented in the model by its STM profile at the output of a bank of peripheral auditory filters, that is, a description of the magnitude of STM components present in the speech signal across a range of temporal rates and spectral densities. A clean, undistorted speech signal presented to the model produces a characteristic STM profile. If the speech signal is distorted by adding noise or reverberation, the STM profile is altered from the clean version. For example, reverberation tends to smear the speech signal in the temporal domain (Steeneken and Houtgast, 1980), thereby reducing the presence of higher-rate temporal modulations in the model output. The model's intelligibility prediction is based on the similarity between the STM profile for clean speech and that for corrupted speech (i.e., the Spectrotemporal Modulation Index [STMI]). Larger differences between the clean and corrupted STM profile lead to predictions of poorer speech intelligibility for the corrupted signal. This model was successful in predicting speech intelligibility for NH listeners across a range of signal distortions, including the addition of background noise and reverberation. The model was also used to predict the effects of directional microphones on speech recognition by a group of HI listeners (Grant et al, 2008).

The success of the STM-based model in predicting speech intelligibility for NH listeners raises the possibility that such a model could be used as a tool for predicting speech intelligibility for individual HI listeners based on each listener's particular ability to detect STM. If HI listeners are impaired in their ability to detect STM, and the ability to detect

STM is correlated to speech intelligibility across listeners, this would suggest that an STM-based intelligibility model would have a significant chance of success in predicting speech intelligibility for individual HI listeners. Although STM sensitivity has been characterized for NH listeners (Chi et al, 1999), it is not known how hearing loss affects STM sensitivity.

This study examined STM sensitivity for NH and HI listeners and its relationship to other psychoacoustic measures and speech intelligibility as a step toward assessing the potential for success of a speech-prediction model based on STM sensitivity for individual HI listeners. There were three main goals of this study. The first goal was to characterize the effects of hearing loss on STM sensitivity. If STM sensitivity is negatively impacted by hearing loss, then an STM-based model of speech intelligibility would be likely to predict poorer speech-reception performance for HI listeners than NH listeners. The second goal was to determine whether variability across HI listeners in STM sensitivity would be related to deficits in psychoacoustic measures of frequency resolution, compression, or TFS processing ability (measured in the same group of subjects in a separate study; Summers et al, 2013). If so, this would point to the particular aspects of impaired peripheral processing that should be incorporated into the STM model to allow an accurate representation of STM sensitivity for individual HI listeners. The third goal was to determine whether intersubject differences in STM sensitivity predict speech intelligibility performance better than the audiogram alone. This hypothesis was based on the idea that the ability to identify speech cues relies on the ability to process modulations in time and frequency. If so, this would suggest that the successful modeling of STM sensitivity for individual HI listeners might yield more accurate predictions of speech intelligibility than those provided by audibility measures alone.

EXPERIMENT: SPECTROTEMPORAL MODULATION SENSITIVITY

Rationale

Fourier theory states that any waveform can be decomposed into a sum of individual sinusoids, each with a different frequency. The spectrum of a particular waveform describes the magnitude and phase of each of the individual frequency components that when summed together yield the waveform. In an analogous manner, any spectrogram(i.e., a description of the energy in a signal as a function of time and frequency) can be decomposed into a sum of spectrotemporal ripples, each with a different combination of temporal rate and spectral density. The STM pattern for a given spectrogram describes the magnitude and phase of each of the rate/density components that when summed together yield the spectrogram. An example of the construction of a complex spectrogram from a set of basic ripples is shown in Figure 1. Panels A, B, and C show spectrograms associated with three particular combinations of rate, density, and direction. Figure 1A shows a downward-moving ripple stimulus with a relatively slow rate (4 Hz), with few fluctuations in the horizontal time dimension at a given frequency, and a low spectral density (0.5 c/o), with few fluctuations in the vertical frequency dimension at any given time. Figure 1B shows an upward-moving ripple with a relatively fast rate (32 Hz) but the same low spectral density (0.5 c/o) as in Figure 1A. Figure 1C shows a downward-moving ripple with the same slow temporal rate (4 Hz) as Figure 1A but with a relatively high spectral density (2 c/o). When the spectrograms associated with the three STM components shown in Figures 1A–C are added together, they yield the more complex spectrogram shown in Figure 1D. The decomposition of a speech spectrogram into its constituent rate/density ripple components works in the same way, except with many more than three STM components. A reduction in sensitivity to a particular range of STM rate-density combinations would likely reduce a listener's ability to detect speech cues in the spectrogram associated with those densities and rates.

The current experiment measured NH and HI listeners' abilities to detect the presence of STM over a range of combinations of rates, densities, and directions. Chi et al (1999) measured sensitivity to combined spectral and temporal modulations for NH listeners. They showed that the STM sensitivity patterns replicated the low-pass characteristics of purely temporal (e.g., Viemeister, 1979; Yost and Moore, 1987) and spectral (e.g., Green, 1986; Bernstein and Green, 1987; Eddins and Bero, 2007) modulation-sensitivity functions observed in previous studies. No attempts have been made previously to characterize STM sensitivity for listeners with hearing loss. The effects of SNHL on spectral and temporal resolution have only been studied separately. HI listeners generally show a loss of spectral modulation sensitivity (Summers and Leek, 1994; Henry et al, 2005) consistent with reduced frequency selectivity observed in the auditory periphery (e.g., Glasberg and Moore, 1986). On the other hand, it is generally thought that HI listeners have normal temporal modulation sensitivity (Bacon and Viemeister, 1985; Bacon and Gleitman, 1992; Moore et al, 1992). Although some studies have shown poorer modulation-detection thresholds for HI than for NH listeners, it has been argued that these differences might be ascribed to either narrower effective stimulus bandwidth (Bacon and Viemeister, 1985) or to a reduced ability to spectrally resolve modulation sidebands from the carrier tone (Grant et al, 1998; Moore and Glasberg, 2001). When these factors are adequately controlled, NH and HI listeners generally demonstrate a similar ability to detect the presence of temporal modulation (Bacon and Viemeister, 1985; Moore and Glasberg, 2001). (Still, it should be pointed out that temporal modulation sensitivity might be negatively impacted by age [He et al, 2008], a factor that often covaries with hearing loss.)

Although spectral and temporal modulation sensitivity has been characterized for HI listeners, these spectral-only or temporal-only modulation-sensitivity measures do not directly reflect the characteristics found in natural sounds like speech that often have combined STM modulation. The question posed here was whether HI listeners would show a pattern of reduced sensitivity to STM stimuli that reflects sensitivity to spectral-only and temporal-only modulation. If so, we would expect STM sensitivity to be more impaired for higher than for lower spectral densities because poor frequency selectivity would reduce the ability to resolve closely spaced spectral peaks. On the other hand, we would not expect any differential effect of hearing loss on STM sensitivity as a function of temporal modulation rate, since temporal modulation sensitivity is thought to be unaffected by hearing loss.

Methods

The ability to detect the presence of STM imposed on a broadband noise was measured for eight NH and 12 HI listeners for upward- and downward-moving STM ripples with a range of rates and densities. STM detection thresholds were measured in a two-alternative forced-choice task, where one interval contained unmodulated noise and the other interval contained the STM stimulus. An adaptive procedure estimated the modulation depth required for STM detection.

Stimuli—The noise carrier consisted of 4000 equal-amplitude random-phase tones that were equally spaced along the logarithmic frequency axis spanning four octaves (354–5656 Hz). To produce the ripple stimulus, sinusoidal amplitude modulation was applied to each carrier tone. Spectral modulation was induced by adjusting the relative phases of the temporal modulation for each successive carrier tone, yielding a sinusoidal envelope at each point in time along the log frequency axis. In this manner, any combination of rate and density can be produced. When either parameter is nonzero, a set of individual ripples is created that change in frequency over time. Dynamic ripples of this type can be constructed such that the individual ripples are increasing or decreasing in (log) frequency at a rate that corresponds to the amplitude modulation rate of the individual components. For a complete

mathematical description of the ripple stimulus, see Chi et al (1999). Figure 1 shows example spectrograms of STM ripple stimuli with different combinations of rate, density, and direction.

Procedure—STM detection thresholds were measured using a two-alternative forced choice adaptive procedure. One interval contained the reference noise carrier, while the other contained the STM signal. Reference and signal stimuli were 500 msec in duration, including 20 msec raised cosine ramps. On each trial, the STM stimulus was randomly assigned to the first or second interval, each with probability 0.5. The STM modulation depth was varied in a three-down one-up adaptive procedure tracking the 79.4% correct point (Levitt, 1971). The modulation depth of the STM tracked during each run is reported in dB (i.e., 20 log m, where m is the modulation depth). The starting modulation depth for each run was 0 dB (i.e., full modulation). The modulation depth was reduced by 6 dB until the first reversal, changed by 4 dB for the next two reversals, and changed by 2 dB for the last six reversals, for a total of nine reversals per run. The threshold was determined by taking the mean of the modulation depth (in dB) of the last six reversal points. By definition, the modulation depth could not exceed 0 dB (full modulation). Therefore, if the adaptive track required a modulation depth greater than 0 dB, the next trial was presented with full modulation. If a listener was unable to detect the fully modulated signal more than five times in any one run, the run was terminated and a threshold was not collected. Visual feedback was provided after each trial to indicate the interval that contained the correct response.

The signal and the standard noise were presented at a nominal level of 80 dB SPL/octave (86 dB SPL overall) to the test ear. This level was chosen to maximize stimulus audibility for the NH and HI listeners without being uncomfortably loud. Both listener groups were presented with the same stimulus level to avoid the potentially confounding effects of stimulus-level differences on frequency selectivity (Pick, 1980; Glasberg and Moore, 2000). To reduce the effectiveness of a possible loudness cue, the overall presentation level in each interval was roved over a 5 dB range, with a random gain chosen from a uniform distribution between −2.5 and +2.5 dB. An uncorrelated noise, with the same statistics as the standard noise but 20 dB lower in level, was presented to the opposite ear to ensure that the detection was performed using the intended test ear.

For each listener, two runs were presented for each combination of density (0.5, 1, 2, or 4 c/o), rate (4, 12, or 32 Hz), and direction (upward or downward). A third threshold was collected for a given condition if the two threshold estimates differed by 3 dB or more, or if one of the two runs was terminated early due to more than five incorrect responses at full modulation depth. A fourth threshold estimate was collected if any two of the first three measurements differed by more than 6 dB. The threshold for each listener and condition was taken to be the average of the one, two, three, or four runs that did not terminate early.

Listeners—Eight NH listeners (four female, mean age: 44.5, age range: 24–60) and 12 HI listeners (one female, mean age: 75.7, age range: 70–87) took part in this study. Of the 20 listeners, 15 were tested at Walter Reed Army Medical Center, Washington, DC, and five at the National Center for Rehabilitative Auditory Research, Portland, Oregon, using an identical experimental apparatus. The mean audiogram (±1 SD) for each listener group at octave frequencies between 250 and 8000 Hz (plus 3000 and 6000 Hz) is shown in Figure 2. NH listeners had pure-tone thresholds better than or equal to 20 dBHL at all frequencies tested. On average, HI listeners had high-frequency SNHL and near-normal thresholds below 1000 Hz. The ear tested for each HI listener was determined by his or her audiogram: in general, the better ear was tested. In some cases where a HI listener had nearly equal

audiograms for both ears, the decision was determined by the ear that yielded a lower detection threshold for a 1000 Hz tone. NH listeners were tested in the ear of their choice.

Each subject completed a minimum of one hour of training. Training runs were similar to the experimental runs with the exception of an additional interval containing the noise reference. The listener was asked to identify the modulated stimulus randomly presented in interval two or three. The first interval always contained the standard noise reference. The purpose of this reference was to help the listener to better identify the modulated stimulus among the three intervals and to become familiar with the perceptual differences between the standard noise and the STM signals.

Apparatus—Sounds were generated digitally with 32-bit amplitude resolution and a sampling rate of 48,848 Hz. The digital audio signal was sent to an enhanced real-time processor (Tucker-Davis Technologies RP2.1) where it was stored in a buffer. The audio signals were then converted to analog by the processor and were passed through programmable attenuators (Tucker-Davis Technologies PA5), adjusted prior to each interval to apply the random-level rove, and a headphone buffer (Tucker-Davis Technologies HB7) before being presented to the listener through a Sennheiser HD580 headset. The listener was seated inside a double-walled sound attenuating chamber and responded by pressing virtual buttons on a touch screen representing the two stimulus intervals.

Results

Mean STM detection thresholds across the eight NH (gray symbols) and 12 HI (open symbols) listeners are shown in Figure 3 as a function of spectral modulation density (shapes) and temporal modulation rate (horizontal axis, log scale). One HI subject did not complete the 4 c/o conditions, so these conditions reflect the mean performance for the other 11 HI subjects. Although all stimuli were presented at rates of 4, 12, or 32 Hz, some data points are shifted horizontally for clarity. More negative values on the ordinate in Figure 3 indicate better performance, with the STM detectable for smaller modulation depths. The group-mean results shown in Figure 3 are discussed along with the results of a repeated-measures analysis of variance conducted with three within-subjects factors (rate, density, and direction) and one between-subjects factor (hearing loss). Data for the HI listener who did not complete all of the conditions were excluded from the analysis of variance (ANOVA). The reported degrees of freedom include a Huynh and Feldt (1976) correction wherever necessary. There were no statistically significant main effects or interactions involving direction. Therefore, the data shown in Figure 3 represent STM thresholds averaged across the upward- and downward-moving conditions.

STM sensitivity in the spectral and temporal dimensions generally demonstrated the low-pass characteristics shown previously (Chi et al, 1999), reflecting the limited spectral and temporal resolution available in the auditory system. Sensitivity generally decreased as a function of increasing density (squares to circles to triangles to diamonds) and, in most cases, also decreased as a function of increasing rate (horizontal axis). These trends were confirmed by significant main effects of rate [$F_{(1.82, 30.9)} = 22.3$, $p < 0.0005$] and density [$F_{(2.84, 48.3)} = 515$, $p < 0.0005$]. There was also a significant interaction between density and rate [$F_{(4.27, 72.6)} = 10.2$, $p < 0.0005$], indicating that STM sensitivity cannot simply be separated into independent contributions of sensitivity in the temporal and spectral dimensions.

The main question addressed by this study is whether hearing loss affects STM sensitivity and, if so, for which combinations of density, rate, and direction. It was hypothesized that the effect of hearing loss would be largest at the highest spectral densities due to the reduced frequency selectivity associated with hearing loss. This is because the relatively broadly

spaced peaks in the spectrum for low densities would be likely to be resolved by both NH and HI listeners, even with the reduced frequency selectivity associated with hearing loss. At higher densities, spectral peaks would become more closely spaced and begin to interact with group differences in frequency selectivity. It was also expected that the effect of hearing loss would be similar across the range of temporal modulation rates, based on previous results suggesting little effect of hearing loss on temporal resolution (e.g., Moore et al, 1992).

The ANOVA indicated a significant interaction between hearing loss, rate, and density [$F_{(4.27, 72.6)} = 4.05$, $p = 0.004$], supporting the observation that HI listeners showed reduced STM sensitivity, but only for certain combinations of rate and density. (Because the finding of a significant interaction between hearing loss, rate, and density reduces the meaningfulness of the lower-order interactions or main effects in the ANOVA [Keppel and Wickens, 2004], these are not reported or discussed.) To determine which conditions were responsible for the significant three-way interaction, post hoc one-tailed *t*-tests were computed on the STM data collapsed across direction. Significant differences between NH and HI listener groups were observed for three combinations of density and rate: 2 c/o and 4 Hz ($p = 0.0015$), 2 c/o and 12 Hz ($p = 0.0017$), and 4 c/o and 4 Hz ($p = 0.027$). After applying a Bonferroni correction for multiple comparisons, differences in the 2 c/o and 4 Hz ($p = 0.018$) and 2 c/o and 12 Hz ($p = 0.020$) conditions remained significant. It should be noted that the ability to observe an effect of hearing loss on STM sensitivity at 4 c/o might have been limited by floor effects, as HI STM detection thresholds were measured at or near 0 dB for some HI listeners for some 4 c/o conditions. If the modulation depth were not physically limited to a maximum of 0 dB, HI STM thresholds might have been even higher (poorer) for these conditions.

Discussion

Overall, the results indicate that hearing loss affected STM sensitivity only for conditions involving a combination of low temporal rates (4 or 12 Hz) and high spectral densities (2 or 4 c/o). Consistent with a reduction in frequency selectivity, HI listeners showed reduced sensitivity relative to normal performance for higher but not lower spectral densities. The relationship between frequency selectivity and hearing loss will be further explored below by comparing the STM sensitivity data to frequency selectivity estimates made using the notched-noise method (Rosen and Baker, 1994) as measured in the same group of HI listeners (Summers et al, 2013).However, the interaction between the effects of temporal rate and hearing loss on STM sensitivity was unexpected. HI listeners showed less than normal sensitivity for lower temporal rates but more normal performance at high rates. This is inconsistent with the hypothesis that HI listeners should show a similar reduction in STM sensitivity across temporal modulation rates, given previous results suggesting little effect of hearing loss on temporal modulation sensitivity (e.g., Bacon and Viemeister, 1985; Moore and Glasberg, 2001). Furthermore, this result is not consistent with an explanation based on reduced temporal modulation sensitivity, whereby performance would have been most affected by hearing loss for the more difficult case of high-rate modulation detection, rather than for low-rate modulation detection as was observed here.

One possible explanation for the observed reduction in STM sensitivity at low but not high rates is that listeners might have used a spectral edge cue to detect the presence of modulation. The imposition of modulation introduces spectral sidebands, which for the 32 Hz modulation rate would have lowered the low-frequency edge of the noise carrier from 353 Hz down to 332 Hz—approximately a 10% frequency shift between the reference and signal intervals. The frequency shift would have been much smaller for low-rate modulation and not as readily detectable. If detection of STM was affected by hearing loss but not the

ability to use a spectral edge cue, we would expect HI listeners to be less impaired in STM detection at higher rates where the spectral edge cue was more useful.

A second possible explanation for the observed reduction in STM sensitivity at low but not high rates is that HI listeners have a reduced ability to use TFS information (i.e., information regarding the stimulus frequency based on auditory-nerve phase locking), which might be important for the detection of low-rate but not high-rate frequency sweeps. The detection of frequency sweeps could be facilitated by two different sources of information: (1) amplitude-modulation (AM) cues generated by cochlear filters due to the movement of the spectral peaks across the filter slope and (2) changes in the spectral-peak frequencies as encoded in phase locking to individual cycles in the stimulus waveform. To investigate the contributions of these two potential cues to the detection of frequency sweeps, Moore and Sek (1996) measured frequency-modulation (FM) detection performance for NH listeners as a function of modulation rate and carrier frequency. They measured performance both with and without AM also added to both the reference and signal intervals to disrupt the possible use of AM cues to identify the FM. They found that the added AM disrupted FM detection, but mainly for high modulation rates and high absolute frequencies. This suggested that listeners were relying on TFS cues for low rates and low frequencies but AM cues at high rates or high frequencies. They argued that phase-locking information might only be useful for low-rate FM detection because the TFS decoding mechanism is too sluggish (e.g., Grantham and Wightman, 1978) to track fast changes in the tone frequency. Furthermore, TFS cues would only be useful at lower absolute frequencies due to the roll off of phase locking in the auditory nerve (Johnson, 1980).

The STM stimuli in the current study are similar to the tonal FM stimuli employed by Moore and Sek (1996) in that they involve spectral peaks with frequencies that change over time. Thus, the observation that hearing loss affected STM sensitivity for low but not high rates is consistent with an explanation based on a reduced ability to encode slow-moving spectral peak frequencies with TFS. At high rates, even NH listeners would be less likely to use TFS information to detect STM, such that a loss of TFS processing ability associated with hearing loss would have less of an effect on STM sensitivity when compared to normal performance. The possible role of TFS in STM detection will be explored below by comparing STM sensitivity estimates to thresholds for the detection of FM in the presence of added AM (Moore and Sek, 1996), measured by Summers et al (2013) in the same group of HI listeners.

PREDICTIONS OF STM SENSITIVITY AND SPEECH INTELLIGIBILITY

Rationale

The primary goal of this study was to assess the potential for success of a computational speech-intelligibility prediction model for individual HI listeners based on individual differences in peripheral processing ability and STM sensitivity. For this modeling approach to have a significant likelihood of success, two conditions need to be satisfied. First, individual differences in STM sensitivity should be predictive of individual differences in speech intelligibility. Such a predictive relationship would suggest that a model that can accurately portray individual differences in STM sensitivity would also have a good chance of predicting individual differences in speech intelligibility. Second, individual differences in STM sensitivity should be predicted by individual differences in psychoacoustic measures thought to reflect attributes of the auditory periphery. Specifically, the results of the STM experiment described above showed clues that STM sensitivity may be related to frequency selectivity and to TFS processing ability. It is therefore of interest to determine whether these observations are supported by significant correlations between STM sensitivity and psychoacoustic measures thought to reflect frequency selectivity or TFS processing. If so,

individual differences in STM sensitivity should have a reasonable chance of being successfully modeled by adjusting the bandwidths of simulated auditory filters or the strength of phase locking in the responses of simulated auditory-nerve fibers.

Three hypotheses regarding the relationship between STM sensitivity and other psychoacoustic and speech intelligibility measures were tested. First, it was hypothesized that estimates of frequency selectivity would be predictive of STM sensitivity. This is because (1) HI listeners are known to have reduced frequency selectivity (Glasberg and Moore, 1986), (2) STM sensitivity should require the ability to resolve peaks in the stimulus spectrum, and (3) the group-mean results shown above indicate a larger effect of hearing loss at higher spectral density. Second, it was hypothesized that estimates of low-rate frequency modulation (FM) sensitivity would be predictive of STM sensitivity. This was based on (1) the group-mean result whereby HI listeners showed reduced STM sensitivity for low but not for high temporal modulation rates, consistent with the idea that HI listeners have a reduced ability to use TFS information as discussed above and (2) the notion that estimates of low-rate FM sensitivity are thought to reflect the ability to use TFS information (Moore and Sek, 1996; Moore and Skrodzka, 2002). Third, it was hypothesized that STM sensitivity would be predictive of speech intelligibility, as speech signals consist of a range of rate/density combinations of STM.

Methods

To evaluate the relationship between STM sensitivity and both speech intelligibility and basic psychoacoustic capabilities thought to be important for the perception of speech in noise, psychoacoustic and speech intelligibility data were extracted from a separate study that used the same HI listeners who participated in the STM measures (Summers et al, 2013).

Frequency Selectivity—Summers et al (2013) established equivalent rectangular bandwidths (ERBs) of auditory filters measured according to methods described by Rosen and Baker (1994) in this group of HI listeners at four frequencies (500, 1000500, 2000, and 4000 Hz) and at one or two signal levels (between 70 and 85 dB SPL).

FM Sensitivity—Summers et al (2013) measured FM sensitivity in terms of the minimum modulation depth (in log frequency) needed to detect 2 Hz sinusoidal frequency modulation applied to a tone carrier of 500, 1000, 2000, or 4000 Hz. Random amplitude modulation (center modulation frequency=3 Hz, bandwidth=6 Hz)was added to both the reference and signal intervals to reduce the usefulness of amplitude-modulation cues induced by the movement of the tone carrier through an auditory filter. Data were used for a carrier level of 85 dB SPL.

Speech Intelligibility—Summers et al (2013) measured speech intelligibility for the Institute of Electrical and Electronics Engineers (IEEE) (1969) sentences presented in stationary and modulated noise at a variety of signal-to-noise ratios (SNRs). The data for speech presented in stationary noise at an SNR of 0 dB are presented here (although modulated-noise maskers and other SNRs yielded similar results). The target speech was presented to all listeners at a high level of 92 dB SPL in an attempt to overcome audibility limitations. A main question posed here is whether the STM sensitivity measure provides predictive value beyond that provided by standard audiometric measures and the SII that is based on audibility. The SII (ANSI, 1997) was computed for speech presented in quiet for each subject's audiometric configuration. The SII for speech presented in quiet was chosen instead of the SII for speech presented in noise because the SII in quiet accounted for a larger proportion of the variance in the speech scores. This might be because the calculation

of the SII for speech presented in noise is determined more by the noise statistics that are constant across listeners than by the audiogram.

Results

STM Sensitivity Prediction—The first question addressed was whether STM sensitivity could be predicted by the other psychoacoustic measures. A stepwise regression analysis was conducted to determine which of the psychoacoustic factors were predictive of STM sensitivity across the HI listeners. The stepwise regression analysis first finds the input factor that accounts for the largest proportion of the output variance. Then, the variable that accounts for the largest proportion of the residual variance is entered. This process continues until none of the remaining variables account for a significant portion of the residual variance. The inputs to the linear regression model were listener age (years), audiometric thresholds (dB HL, four frequencies), ERB estimates (Hz), and FM sensitivity (log Hz). Each of the psychoacoustic quantities was measured at four frequencies, for a total of 12 inputs. The output of the regression model was STM sensitivity for the 2 c/o, 4 Hz condition (averaged across the upward and downward conditions). Only this density/rate condition was included in the analysis because this is the condition where hearing loss had the largest impact on STM sensitivity in the group analysis. The linear regression analysis found only two factors to contribute significantly to the prediction of STM sensitivity: 500 Hz FM sensitivity and the 4000 Hz ERB. Figure 4 plots the measured STM sensitivity for individual HI listeners as a function of the STM sensitivity predicted by the regression model that includes these two inputs. Together, these two variables accounted for 67.4% of the variance in STM sensitivity across the HI listeners ($p = 0.011$), with 39.9% of the variance accounted for by the 500 Hz FM sensitivity measure and an additional 27.5% accounted for by the 4000 Hz ERB estimate. None of the audiometric thresholds provided any significant predictive power for STM sensitivity, suggesting that STM sensitivity measures an aspect of auditory perception that is not captured by traditional audiometric measures.

Speech Intelligibility Predictions—The second question addressed was whether STM sensitivity measures provided any power to predict speech intelligibility beyond that provided by audibility measures expressed in terms of the SII. Figure 5 shows that the SII alone was able to account for 39.9% of the variance in speech intelligibility ($p = 0.028$). To determine whether estimates of STM sensitivity provided additional predictive power, a stepwise regression analysis was conducted with two input variables (the SII and 4 Hz, 2 c/o STM sensitivity) and one output variable (speech intelligibility, proportion correct of the IEEE sentence keywords). Both input variables contributed significantly to the prediction of speech intelligibility. Together they accounted for 81.2% of the variance in intelligibility scores ($p = 0.0005$, Fig. 6), with STM sensitivity accounting for 61.1% and the SII accounting for an additional 20.1% of the variance. When the SII was forced to be the first variable entered in the stepwise regression (accounting for 39.9% of the variance, Fig. 5), the introduction of STM sensitivity into the regression equation accounted for an additional 41.3% of the variance (Fig. 6). Similar results were found when audiometric thresholds were included instead of the SII in the regression analysis.

Discussion

STM sensitivity was strongly related to both low-frequency, low-rate FM sensitivity and to high-frequency frequency selectivity. In the group-mean results, the HI listeners were only found to be significantly different from the NH group at higher spectral densities. The finding that the 4000 Hz ERB was a significant predictor of STM sensitivity provides further support for the interpretation that the reduced frequency selectivity associated with hearing loss negatively impacts STM sensitivity. That this predictive relationship was only observed at high frequencies is consistent with the greater audiometric hearing loss observed

at high frequencies as well as with previous results indicating that hearing loss has a greater effect on frequency selectivity at high frequencies (Hicks and Bacon, 1999).

In the group-mean analysis, hearing loss was found to significantly reduce STM sensitivity, but only for low rates. As discussed earlier, one possible interpretation of this result is that poor STM sensitivity might reflect an inability to use TFS information to follow the frequencies of the sweeping spectral peaks in the ripple stimulus. To the extent that the FM detection task measures a listener's ability to use TFS information, the finding that FM sensitivity was a significant predictor of STM sensitivity further supports the notion that the ability to use TFS information is an important factor in the detection of STM. Furthermore, the relationship between FM and STM sensitivity was observed for low carrier frequencies, consistent with the idea that TFS cues for FM or STM detection will become less salient at higher frequencies due to a roll off in phase locking at higher frequencies (Johnson, 1980).

The loss of frequency selectivity accompanying hearing loss has been well documented (Pick et al, 1977; Glasberg and Moore, 1986; Leek and Summers, 1993; Moore et al, 1999). This impairment is widely believed to reflect a loss of the narrowly tuned cochlear amplifier that also provides compression and sensitivity to low-level sounds in the NH ear (Moore et al, 1999) and is readily observed physiologically in auditory-nerve responses. It has also been suggested that HI listeners may have an impaired ability to use TFS information in speech perception (Buss et al, 2004; Lorenzi et al, 2006; Hopkins and Moore, 2007, 2010; Strelcyk and Dau, 2009), sound-source localization (Lacher-Fougère and Demany, 2005), pitch perception (Moore et al, 2006), and FM detection (Moore and Skrodzka, 2002). However, physiological evidence for reduced fidelity of phase-locking information in the impaired auditory system remains elusive. Only one study has shown a reduced auditory-nerve phase locking to TFS with SNHL (Woolf et al, 1981), while several other studies that looked at this effect have not (Harrison and Evans, 1979; Miller et al, 1997; Kale and Heinz, 2010). Kale and Heinz (2010) found enhanced temporal envelope coding in auditory-nerve fibers following noise-induced hearing loss, resulting in reduced fidelity of TFS encoding when considered relative to envelope encoding. This could be thought of as a TFS processing deficit if the impaired system were to rely on the more highly modulated temporal envelope rather than the TFS. Another possibility is that the accurate use of TFS information might require redundant coding available across many auditory nerve fibers (ANFs) to overcome encoding noise. Kujawa and Liberman (2009) found that as many as 50% of ANFs become nonfunctioning following a temporary threshold shift; a permanent threshold shift associated with SNHL might involve even more neuronal loss.

STM sensitivity was shown to be a significant predictor of speech intelligibility across the HI listeners. While the audibility-based SII was able to capture about 40% of the variance in speech scores, STM sensitivity and SII together accounted for about 80% of the variance. Together with the finding that STM sensitivity for individual HI listeners can be predicted based on psychoacoustic measurements of frequency selectivity and TFS processing ability, these results are encouraging regarding the potential for success of an STM-based approach to the prediction of speech intelligibility for HI listeners. The generic peripheral stage of the Elhilali et al (2003) model might be replaced by subject-specific peripheral processing. These results suggest that differences in the peripheral processing stage would generate different patterns of STM sensitivity for each HI listener, which would in turn generate amore accurate prediction of speech intelligibility for each listener than is available from current audibility-based models such as the SII.

The regression analyses discussed above focused on the 4 Hz, 2 c/o condition because that was the condition that yielded the largest difference in STM sensitivity between NH and HI groups. Yet in the STM-based speech intelligibility model of Elhilali et al (2003), a greater

range of rates (2–32 Hz) and densities (0.25–8 c/o) are considered, as a speech signal contains STM elements from across these ranges. Correlations were calculated across the 12 HI listeners between speech intelligibility (stationary noise, 0 dB SNR) and STM sensitivity for each of the 12 rate-density combinations (STM sensitivity thresholds were averaged across upward- and downward-moving ripple. conditions). The resulting squared Pearson correlation coefficients are shown in Table 1. Asterisks indicate correlations that were significant ($p < 0.05$) after applying a Holm-Sidak correction for multiple comparisons. In addition to the 4 Hz, 2 c/o condition, STM sensitivity was also correlated with speech intelligibility for the 4 Hz, 1 c/o and 32 Hz, 0.5 and 1 c/o conditions. These results suggest that both high- and low-rate modulations might be important for speech intelligibility. Although speech is dominated by modulation rates below 16 Hz (Steeneken and Houtgast, 1980; Drullman et al, 1994), with relatively little energy in the modulation spectrum above this rate, higher rate modulations may nonetheless convey important speech cues (Silipo et al, 1999). In any case, the predictive power of high-rate STM sensitivity was surprising, given that the HI and NH listener groups had similar performance for the high-rate conditions, consistent with previous work showing little reduction in temporal modulation sensitivity for HI listeners (Bacon and Viemeister, 1985; Bacon and Gleitman, 1992; Moore et al, 1992). It should be pointed out that there was a high degree of correlation between most of the rate-density combinations. It is possible that sensitivity to high-rate STM might only be related to speech intelligibility because of the correlation with low-rate STM sensitivity.

FM sensitivity and ERB were shown to be predictive of STM sensitivity, and STM sensitivity was in turn predictive of speech intelligibility. However, in a direct comparison between these peripheral psychoacoustic measures and speech intelligibility, Summers et al (2013) found a significant correlation between speech and FM sensitivity but not between speech and ERB. This finding corroborates other studies that have examined the relationship between TFS processing ability, frequency selectivity, and speech intelligibility measures (e.g., Buss et al, 2004; Strelcyk and Dau, 2009). It could be that frequency selectivity is an important contributor to STM sensitivity but that this aspect of STM processing is not important for speech perception. The main speech cues occurring at high frequencies, where frequency selectivity is most affected by hearing loss, involve broad spectral peaks associated with frication noise. The use of such cues might not require sharp frequency tuning. On the other hand, the detection of lower-frequency formant-peak shifts for the identification of consonant place-of-articulation is more analogous to the detection of FM and might be sensitive to the fidelity of TFS encoding.

Implications for Clinical Management

Current clinical practice for the management of an individual's hearing loss consists of an audiometric evaluation and often a measure of speech intelligibility in quiet. These measures allow the clinician to set the parameters of a hearing aid to maximize signal audibility (within the constraints of loudness and comfort) and, by extension, speech intelligibility. The current results suggest including a measure of STM sensitivity in the battery of tests available to the clinician. Because STM sensitivity, especially for low-rate, high-density stimuli, is strongly related to speech intelligibility, testing STM sensitivity could give the clinician an idea of an upper limit on the benefit that a given amplification algorithm might yield for speech perception in noise. A listener with very good STM sensitivity would be expected to benefit fully from amplification, understanding speech in noise to the extent that audibility limits will allow. A listener with poor STM sensitivity would be hampered by suprathreshold distortion and therefore expected to perform poorly, even if audibility is restored through amplification. Speech tests currently employed in the clinic can be used to address similar questions. For example, if a patient's maximum amplified speech-

recognition score in quiet (PB max) is considerably worse than expected based on the audiogram, this would suggest a processing deficit that limits speech-reception performance. The potential benefit of including a basic psychoacoustic measure in this context is that it is less likely to be influenced by other nonauditory factors that can affect speech recognition such as a patient's cognitive processing ability, educational background, country of origin (native/nonnative speaker), or age (adult/child). Furthermore, a psychoacoustic measure would target the basic acoustic elements of speech that are common to all speech materials rather than those contextual elements that vary across speech corpora. Because hearing aids are designed to enhance the acoustic properties of speech, it could be useful to isolate the acoustic aspects of speech intelligibility from these nonacoustic factors.

The long-term goal of an STM-based speech-intelligibility model for HI individuals is to provide a platform for the development of signal-processing strategies to compensate for suprathreshold distortion. The model could generate speech-intelligibility predictions for a range of signal-processing algorithms, aiding the clinician in selecting the optimal hearing-aid settings. The relationship between STM sensitivity and speech intelligibility suggests that an algorithm to enhance STM in the input signal might enhance intelligibility for an HI listener. In fact, a related algorithm operating in the spectral modulation domain has been shown to improve consonant and vowel recognition in noise for NH and HI listeners (Eddins and Liu 2009). In this context, a clinical measure of an individual's ability to detect STM could be an important component toward the individualized parameterization of a future STM enhancement algorithm, allowing the parameters of the enhancement algorithm to be tailored to the individual's psychoacoustic sensitivity.

A relationship was observed between STM sensitivity and the FM detection and frequency-selectivity tests that are thought to target the mechanisms underlying the detection of STM. Therefore, another possible avenue would be to measure these more basic psychophysical attributes instead of STM sensitivity. From a clinical perspective, it would be preferable to use the STM metric because it contains information about these underlying mechanisms within a single test, reducing the number of psychophysical attributes that would need to be measured for a given individual. Most importantly, the STM metric was more closely correlated to speech perception than were these other measures, perhaps because it is more similar to a speech signal, with a wide bandwidth and modulations that simulate those occurring naturally in speech.

The question remains as to the feasibility of conducting such an STM measurement in a clinical setting. In the current study, listeners were trained for over an hour and tested for several additional hours, a test protocol that would be prohibitively time-consuming for clinical application. However, it should be noted that while many different conditions were tested, only a single condition (the 4 Hz, 2 c/o condition) was required to account for a large proportion of the variance in speech intelligibility. Tests are currently underway to determine whether STM sensitivity for this condition could be reliably tested in a matter of minutes for clinical application. Preliminary indications are encouraging, with listeners benefiting very little from any additional training beyond a single three-minute STM threshold run.

CONCLUSIONS

HI listeners showed reduced sensitivity to STM for high spectral densities and low modulation rates. A large proportion (67%) of the variance in STM sensitivity across HI listeners was accounted for by variance in notched-noise estimates of frequency selectivity at high frequencies (4000 Hz) and FM-detection estimates of TFS processing at low frequencies (500 Hz). These results suggest that STM processing is negatively impacted by

hearing loss via a combination of reduced frequency resolution and an impaired ability to use TFS to detect spectral-peak frequency sweeps. Measures of STM sensitivity were also correlated with speech intelligibility. The inclusion of STM sensitivity data into a linear regression analysis improved the prediction of speech intelligibility for individual listeners beyond that provided by audibility-based SII. The strong relationships observed between psychoacoustic measures of peripheral function, STM sensitivity, and speech intelligibility suggest a reasonable likelihood of success for an STM-based model of speech intelligibility for HI listeners based on the model developed by Elhilali et al (2003) for NH listeners. By incorporating suprathreshold distortions caused by reduced frequency selectivity and an impaired ability to use TFS information, such an approach might yield a more accurate prediction of speech intelligibility for individual HI listeners than an existing audibility-based model such as the SII.

Acknowledgments

Abbreviations

AM	amplitude modulation
c/o	cycles per octave
ERB	equivalent rectangular bandwidth
FM	frequency modulation
HI	hearing impaired
NH	normal hearing
SII	Speech Intelligibility Index
SNHL	sensorineural hearing loss
SNR	signal-to-noise ratio
STM	spectrotemporal modulation
TFS	temporal fine structure

References

American National Standards Institute (ANSI). Methods for Calculation of the Speech Intelligibility Index, S3.5. New York: American National Standards Institute; 1997.

Bacon SP, Gleitman RM. Modulation detection in subjects with relatively flat hearing losses. J Speech Hear Res. 1992; 35(3):642–653. [PubMed: 1608256]

Bacon SP, Viemeister NF. Temporal modulation transfer functions in normal-hearing and hearing-impaired listeners. Audiology. 1985; 24(2):117–134. [PubMed: 3994589]

Bernstein LR, Green DM. The profile-analysis bandwidth. J Acoust Soc Am. 1987; 81(6):1888–1895.

Buss E, Hall JW, Grose JH. Temporal fine-structure cues to speech and pure tone modulation in observers with sensorineural hearing loss. Ear Hear. 2004; 25(3):242–250. [PubMed: 15179115]

Byrne D, Dillon H, Ching T, Katsch R, Keidser G. NAL-NL1 procedure for fitting nonlinear hearing aids: characteristics and comparisons with other procedures. J Am Acad Audiol. 2001; 12(1):37–51. [PubMed: 11214977]

Chi T, Gao Y, Guyton MC, Ru P, Shamma S. Spectro-temporal modulation transfer functions and speech intelligibility. J Acoust Soc Am. 1999; 106(5):2719–2732. [PubMed: 10573888]

Ching TY, Dillon H, Byrne D. Speech recognition of hearing-impaired listeners: predictions from audibility and the limited role of high-frequency amplification. J Acoust Soc Am. 1998; 103(2): 1128–1140. [PubMed: 9479766]

Drullman R, Festen JM, Plomp R. Effect of reducing slow temporal modulations on speech reception. J Acoust Soc Am. 1994; 95(5 Pt 1):2670–2680. [PubMed: 8207140]

Eddins DA, Bero EM. Spectral modulation detection as a function of modulation frequency, carrier bandwidth, and carrier frequency region. J Acoust Soc Am. 2007; 121(1):363–372. [PubMed: 17297791]

Eddins DA, Liu C. Consonant identification in noise with and without spectro-temporal enhancement by normal hearing and hearing impaired listeners. Assoc Res Otolaryngol Abs. 2009:892.

Elhilali M, Taishih C, Shamma SA. A spectro-temporal modulation index (STMI) for assessment of speech intelligibility. Speech Commun. 2003; 41:331–348.

Glasberg BR, Moore BCJ. Auditory filter shapes in subjects with unilateral and bilateral cochlear impairments. J Acoust Soc Am. 1986; 79(4):1020–1033. [PubMed: 3700857]

Glasberg BR, Moore BCJ. Frequency selectivity as a function of level and frequency measured with uniformly exciting notched noise. J Acoust Soc Am. 2000; 108(5 Pt 1):2318–2328. [PubMed: 11108372]

Grant KW, Elhilali M, Shamma SA, Walden BE, Surr RK, Cord MT, Summers V. An objective measure for selecting microphone modes in OMNI/DIR hearing aid circuits. Ear Hear. 2008; 29 (2):199–213. [PubMed: 18595186]

Grant KW, Summers V, Leek MR. Modulation rate detection and discrimination by normal-hearing and hearing-impaired listeners. J Acoust Soc Am. 1998; 104(2 Pt 1):1051–1060. [PubMed: 9714924]

Grantham DW, Wightman FL. Detectability of varying interaural temporal differences. J Acoust Soc Am. 1978; 63(2):511–523. [PubMed: 670548]

Green, D. Frequency and the detection of spectral shape change. In: Moore, BCJ.; Patterson, R., editors. Auditory Frequency Selectivity. Cambridge: Plenum; 1986. p. 351-359.

Harrison RV, Evans EF. Some aspects of temporal coding by single cochlear fibres from regions of cochlear hair cell degeneration in the guinea pig. Arch Otorhinolaryngol. 1979; 224 (1–2):71–78. [PubMed: 485951]

He NJ, Mills JH, Ahlstrom JB, Dubno JR. Age-related differences in the temporal modulation transfer function with pure-tone carriers. J Acoust Soc Am. 2008; 124(6):3841–3849. [PubMed: 19206810]

Henry BA, Turner CW, Behrens A. Spectral peak resolution and speech recognition in quiet: normal hearing, hearing impaired, and cochlear implant listeners. J Acoust Soc Am. 2005; 118(2):1111–1121. [PubMed: 16158665]

Hicks ML, Bacon SP. Effects of aspirin on psychophysical measures of frequency selectivity, two-tone suppression, and growth of masking. J Acoust Soc Am. 1999; 106(3 Pt 1):1436–1451. [PubMed: 10489702]

Hopkins K, Moore BCJ. Moderate cochlear hearing loss leads to a reduced ability to use temporal fine structure information. J Acoust Soc Am. 2007; 122(2):1055–1068. [PubMed: 17672653]

Hopkins K, Moore BCJ. The importance of temporal fine structure information in speech at different spectral regions for normal-hearing and hearing-impaired subjects. J Acoust Soc Am. 2010; 127(3):1595–1608. [PubMed: 20329859]

Huynh H, Feldt LS. Estimation of the Box correction for degrees of freedom from sample data in the randomized block and split-plot designs. J Educ Assoc. 1976; 1:69–82.

Institute of Electrical and Electronics Engineers (IEEE). IEEE Recommended Practice for Speech Quality Measures. New York: Institute of Electrical and Electronics Engineers; 1969.

Johnson DH. The relationship between spike rate and synchrony in responses of auditory-nerve fibers to single tones. J Acoust Soc Am. 1980; 68(4):1115–1122. [PubMed: 7419827]

Kale S, Heinz MG. Envelope coding in auditory nerve fibers following noise-induced hearing loss. J Assoc Res Otolaryngol. 2010; 11 (4):657–673. [PubMed: 20556628]

Keidser G, Grant F. Comparing loudness normalization (IHAFF) with speech intelligibility maximization (NAL-NL1) when implemented in a two-channel device. Ear Hear. 2001; 22:501–515. [PubMed: 11770672]

Keppel, G.; Wickens, TD. Design and Analysis: A Researcher's Handbook. 4. Englewood Cliffs, NJ: Prentice Hall; 2004.

Kujawa SG, Liberman MC. Adding insult to injury: cochlear nerve degeneration after "temporary" noise-induced hearing loss. J Neurosci. 2009; 29(45):14077–14085. [PubMed: 19906956]

Lacher-Fougère S, Demany L. Consequences of cochlear damage for the detection of interaural phase differences. J Acoust Soc Am. 2005; 118(4):2519–2526. [PubMed: 16266172]

Leek MR, Summers V. Auditory filter shapes of normal-hearing and hearing-impaired listeners in continuous broadband noise. J Acoust Soc Am. 1993; 94(6):3127–3137. [PubMed: 8300951]

Levitt H. Transformed up-down methods in psychoacoustics. J Acoust Soc Am. 1971; 49(2):467–477. [PubMed: 5541744]

Liu C, Eddins DA. Effects of spectral modulation filtering on vowel identification. J Acoust Soc Am. 2008; 124(3):1704–1715. [PubMed: 19045661]

Lorenzi C, Gilbert G, Carn H, Garnier S, Moore BCJ. Speech perception problems of the hearing impaired reflect inability to use temporal fine structure. Proc Natl Acad Sci U S A. 2006; 103 (49): 18866–18869. [PubMed: 17116863]

Miller RL, Schilling JR, Franck KR, Young ED. Effects of acoustic trauma on the representation of the vowel "eh" in cat auditory nerve fibers. J Acoust Soc Am. 1997; 101(6):3602–3616. [PubMed: 9193048]

Moore BCJ, Glasberg BR. Temporal modulation transfer functions obtained using sinusoidal carriers with normally hearing and hearing-impaired listeners. J Acoust Soc Am. 2001; 110(2):1067–1073. [PubMed: 11519575]

Moore BCJ, Glasberg BR, Hopkins K. Frequency discrimination of complex tones by hearing-impaired subjects: evidence for loss of ability to use temporal fine structure. Hear Res. 2006; 222 (1–2):16–27. [PubMed: 17030477]

Moore BCJ, Sek A. Detection of frequency modulation at low modulation rates: evidence for a mechanism based on phase locking. J Acoust Soc Am. 1996; 100(4 Pt 1):2320–2331. [PubMed: 8865639]

Moore BCJ, Shailer MJ, Schooneveldt GP. Temporal modulation transfer functions for band-limited noise in subjects with cochlear hearing loss. Br J Audiol. 1992; 26(4):229–237. [PubMed: 1446186]

Moore BCJ, Skrodzka E. Detection of frequency modulation by hearing-impaired listeners: effects of carrier frequency, modulation rate, and added amplitude modulation. J Acoust Soc Am. 2002; 111(1 Pt 1):327–335. [PubMed: 11833538]

Moore BCJ, Vickers DA, Plack CJ, Oxenham AJ. Interrelationship between different psychoacoustic measures assumed to be related to the cochlear active mechanism. J Acoust Soc Am. 1999; 106(5): 2761–2778. [PubMed: 10573892]

Nelson DA, Schroder AC, Wojtczak M. A new procedure for measuring peripheral compression in normal-hearing and hearing-impaired listeners. J Acoust Soc Am. 2001; 110(4):2045–2064. [PubMed: 11681384]

Oxenham, AJ.; Moore, BCJ. Modeling the effects of peripheral nonlinearity in normal and impaired hearing. In: Jesteadt, W., editor. Modeling Sensorineural Hearing Loss. Hillsdale, NJ: Erlbaum; 1997. p. 273-288.

Pavlovic CV, Studebaker GA, Sherbecoe RL. An articulation index based procedure for predicting the speech recognition performance of hearing-impaired individuals. J Acoust Soc Am. 1986; 80(1): 50–57. [PubMed: 3745665]

Pick GF. Level dependence of psychophysical frequency resolution and auditory filter shape. J Acoust Soc Am. 1980; 68(4):1085–1095. [PubMed: 7419825]

Pick, G.; Evans, EF.; Wilson, JP. Frequency resolution in patients with hearing loss of cochlear origin. In: Evans, EF.; Wilson, JP., editors. Psychophysics and Physiology of Hearing. London: Academic Press; 1977.

Plomp R. A signal-to-noise ratio model for the speech-reception threshold of the hearing impaired. J Speech Hear Res. 1986; 29(2):146–154. [PubMed: 3724108]

Rankovic CM. Factors governing speech reception benefits of adaptive linear filtering for listeners with sensorineural hearing loss. J Acoust Soc Am. 1998; 103(2):1043–1057. [PubMed: 9479758]

Rosen S, Baker RJ. Characterising auditory filter nonlinearity. Hear Res. 1994; 73(2):231–243. [PubMed: 8188552]

Scollie SD, Ching TY, Seewald RC, Dillon H, Britton L, Steiberg J, King K. Children's speech perception and loudness ratings when fitted with hearing aids using the DSL v.4.1 and the NAL-NL1 prescriptions. Int J Audiol. 2010; 49 (Suppl 1):S26–S34. [PubMed: 20109085]

Silipo, R.; Greenberg, S.; Arai, T. Temporal constraints on speech intelligibility as deduced from exceedingly sparse spectral representations. Proc of the 6th European Conference on Speech Communication and Technology (Eurospeech-99); 1999. p. 2687-2690.

Smoorenburg GF. Speech reception in quiet and in noisy conditions by individuals with noise-induced hearing loss in relation to their tone audiogram. J Acoust Soc Am. 1992; 91(1):421–437. [PubMed: 1737889]

Steeneken HJM, Houtgast T. A physical method for measuring speech-transmission quality. J Acoust Soc Am. 1980; 67(1):318–326. [PubMed: 7354199]

Strelcyk O, Dau T. Relations between frequency selectivity, temporal fine-structure processing, and speech reception in impaired hearing. J Acoust Soc Am. 2009; 125(5):3328–3345. [PubMed: 19425674]

Summers V, Leek MR. The internal representation of spectral contrast in hearing-impaired listeners. J Acoust Soc Am. 1994; 95(6):3518–3528. [PubMed: 8046143]

Summers V, Makashay MJ, Theodoroff SM, Leek MR. Suprathreshold auditory processing and speech perception in noise: hearing-impaired and normal-hearing listeners. J Am Acad Audiol. 2013; 24(4):274–292. [PubMed: 23636209]

Vestergaard MD. Dead regions in the cochlea: implications for speech recognition and applicability of articulation index theory. Int J Audiol. 2003; 42(5):249–261. [PubMed: 12916698]

Viemeister NF. Temporal modulation transfer functions based upon modulation thresholds. J Acoust Soc Am. 1979; 66(5):1364–1380. [PubMed: 500975]

Woolf NK, Ryan AF, Bone RC. Neural phase-locking properties in the absence of cochlear outer hair cells. Hear Res. 1981; 4(3–4):335–346. [PubMed: 7263519]

Yost WA, Moore MJ. Temporal changes in a complex spectral profile. J Acoust Soc Am. 1987; 81(6):1896–1905. [PubMed: 3611510]

Zilany, MSA.; Bruce, IC. Predictions of speech intelligibility with a model of the normal and impaired auditory-periphery. Proceedings of the 3rd International IEEE EMBS Conference on Neural Engineering; 2007. p. 481-485.
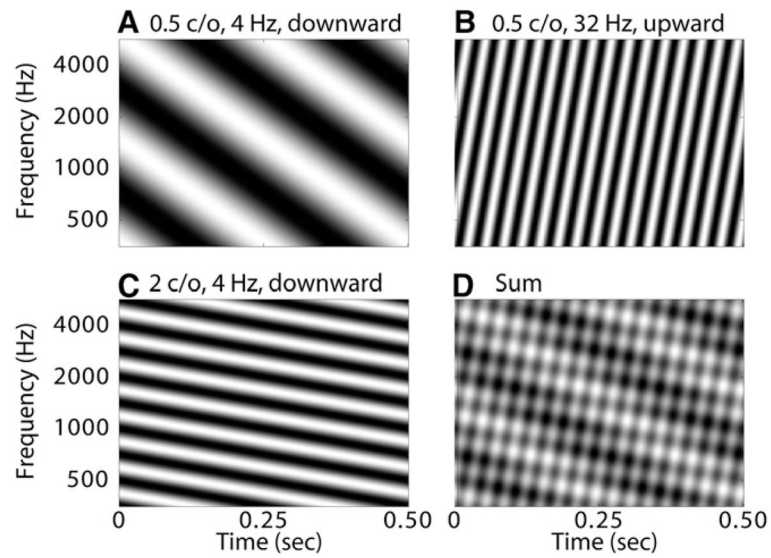
Figure 1.
A–C: Example spectrograms for STM ripple stimuli with various combinations of direction, spectral density, and temporal rate. *D*: A complex spectrogram formed by adding together the three spectrograms from panels *A–C*.
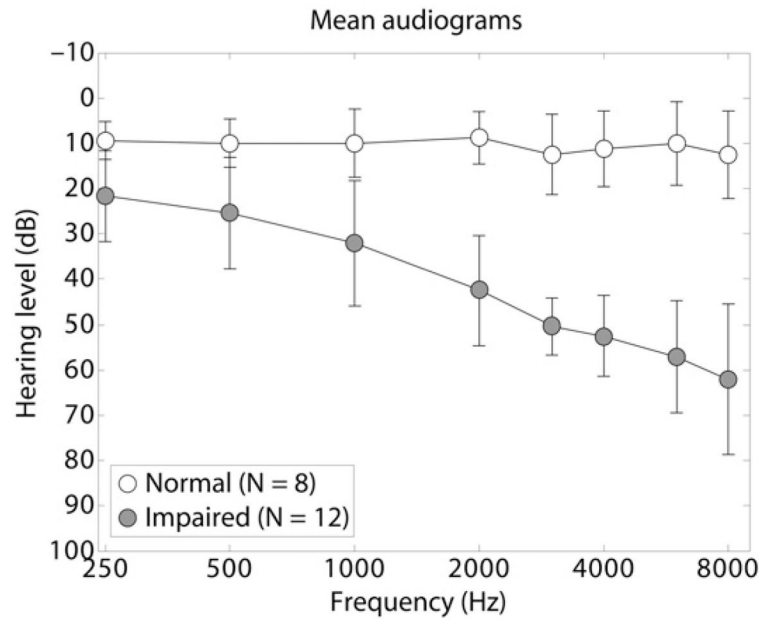
Figure 2.
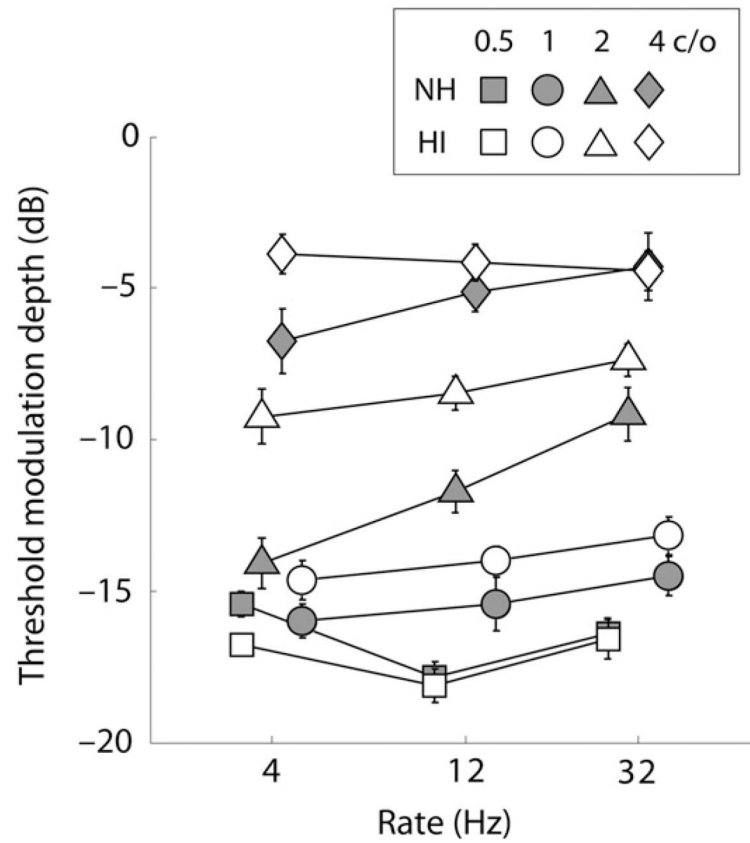Mean audiograms for the NH and HI listener groups.

Figure 3.
Group-mean STM detection thresholds averaged across upward- and downward-moving conditions. Error bars indicate ±1 SE across listeners in each group.
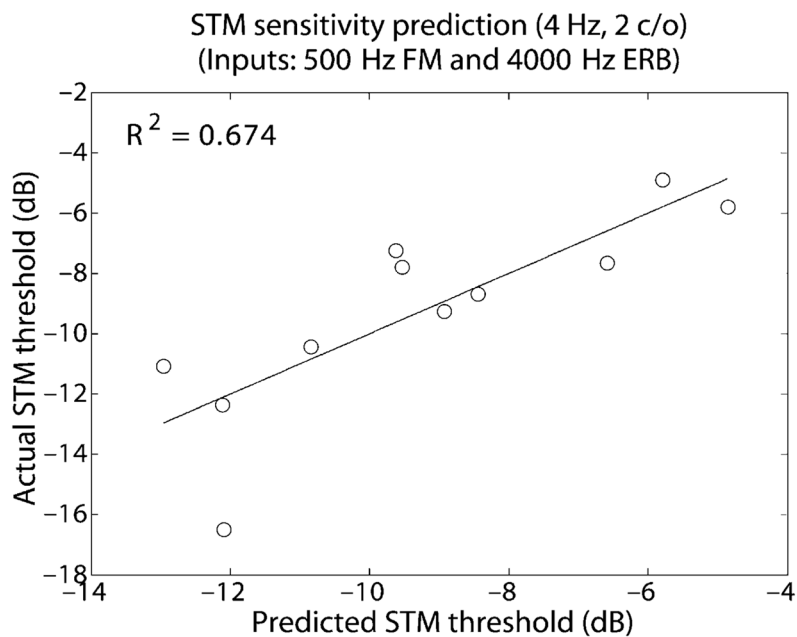
STM sensitivity prediction (4 Hz, 2 c/o)
(Inputs: 500 Hz FM and 4000 Hz ERB)

$R^2 = 0.674$

Figure 4.
Actual STM detection thresholds for individual HI listeners (for the 4 Hz, 2 c/o condition, averaged across upward- and downward-moving ripples) are plotted as a function of the threshold predicted by a linear regression model with the 500 Hz FM detection threshold and 4000 Hz ERB as inputs.
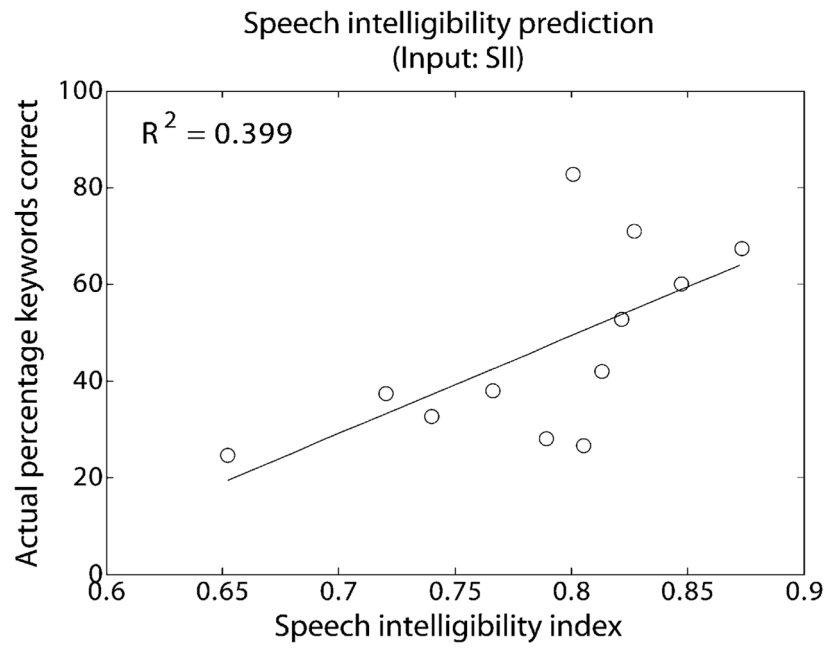
Figure 5.
Speech reception performance for sentence keywords presented in stationary noise at a 0 dB SNR is plotted as a function of the audiogram-based SII for individual HI listeners.
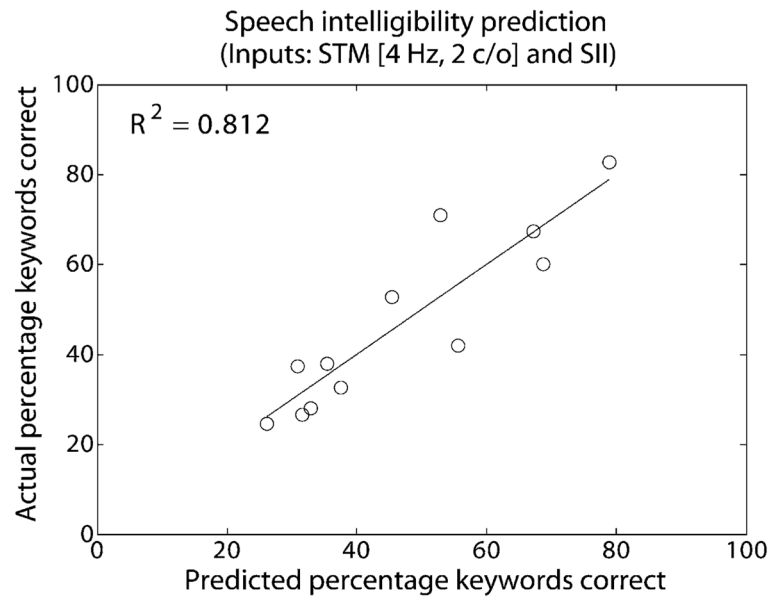
Figure 6.
Actual speech intelligibility is plotted as a function of the intelligibility predicted by a linear regression model with SII and STM sensitivity (4 Hz, 2 c/o) as inputs.

Table 1

Squared Pearson Correlation Coefficients Describing the Relationship across HI Listeners between Speech Intelligibility and STM Sensitivity for All Tested Combinations of Rate and Density

	Rate (Hz)		
Density (c/o)	4	12	32
0.5	0.48	0.05	0.56[*]
1	0.56[*]	0.36	0.71[*]
2	0.61[*]	0.22	0.03
4	0.38	0.21	0.15

[*] Significant correlations ($p<0.05$) after applying a Holm-Sidak correction for multiple comparisons.