Research Article

The Effects of Duration and Level on Spectral Modulation Perception

Sittiprapa Isarangura,[a] Ann C. Eddins,[a] Erol J. Ozmeral,[a] and David A. Eddins[a]

Purpose: Spectral modulation detection is an increasingly common assay of suprathreshold auditory perception and has been correlated with speech perception performance. Here, the potential effects of stimulus duration and stimulus presentation level on spectral modulation detection were investigated.
Method: Spectral modulation detection thresholds were measured as a function of modulation frequency in young, normal-hearing listeners. The standard stimulus was a bandpass noise, and signal stimuli were created by superimposing sinusoidal spectral modulation on the bandpass noise carrier. The modulation was sinusoidal on a $\log_2$ frequency axis and a $\log_{10}$ (dB) amplitude scale with a random starting phase ($0–2\pi$ radians). In 1 experiment, stimulus durations were 50, 100, 200, or 400 ms (at fixed level 81 dB SPL). In a 2nd experiment, stimuli were presented at sensation levels of 10, 20, 30, 40, and 60 dB SL (fixed at a duration of 400 ms).
Results: Spectral modulation detection thresholds were similarly low for the 400- and 200-ms durations, increased slightly for the 100-ms duration, and increased markedly for the 50-ms duration. Thresholds were lowest for 40 dB SL;

increased slightly for 20, 30, and 60 dB SL; and markedly higher for the 10–dB SL condition.
Conclusions: The increase in thresholds for the shortest durations and lowest sensational levels is consistent with previous investigations of auditory spectral profile analysis. The effects of presentation level and stimulus duration are important considerations in the context of understanding potential relationships between the perception of spectral cues and speech perception, when designing investigations and interpreting data related to spectral envelope perception, and in the context of models of auditory perception. As examples, 2 simple models based on auditory nerve output that have been used to explain spectrotemporal modulation in previous investigations produced an output inconsistent with the present results.
Plain language summary: Intensity variations across audio frequency lead to spectral shapes that are essential and sometimes signature features of various sounds in the environment, including speech. Here, we show how laboratory measures of spectral shape perception depend on presentation level and stimulus duration.

The detection of sinusoidal spectral modulation is commonly used as a general measure of auditory spectral shape perception, a fundamental auditory perceptual ability. Spectral modulation detection is analogous to the detection of sinusoidal amplitude modulation, a common index of auditory temporal processing. Measures of fundamental auditory perceptual abilities are often dependent to some extent on the duration and/or the level of the stimuli used to measure those abilities. Such abilities subserve more complex perceptual tasks such as the coding of specific acoustic features, auditory object formation,

and auditory stream segregation (e.g., Shamma, Elhilali, & Micheyl, 2011). It is important, therefore, to establish the effects of duration and level on basic auditory perception and, perhaps in the foreseeable future, to build any associated dependencies into computational models that might be used to better understand and predict more complex auditory processing.

It is known that amplitude modulation detection is robust to a wide range of presentation levels and a wide range of durations, limited only at very low sensation levels and for durations that result in low numbers of modulation cycles (Viemeister, 1979). However, it is unknown how spectral modulation depends on those two basic stimulus properties. Furthermore, knowledge of duration and/or level dependencies is important when designing and interpreting experimental tasks or clinically relevant tasks that include measures of basic auditory perception. Most investigations of spectral modulation detection have used relatively long-duration stimuli (e.g., 400–500 ms) and

[a]Department of Communication Sciences and Disorders, University of South Florida, Tampa

Correspondence to Sittiprapa Isarangura: sisarangura@mail.usf.edu

relatively high presentation levels (spectrum levels between 30 and 50 dB), though the utility of these choices will depend on the goals of the investigation. Investigations comparing the performance of listeners with normal hearing to those with hearing loss face the dilemma of whether it is better to make comparisons at equal sensation levels or an equal overall level. If high sensational levels are desirable to achieve optimum performance, then the presence of hearing loss may pose a measurement challenge because of the limited dynamic range of many listeners with hearing loss. Similarly, studies of auditory profile analysis typically have used durations of around 100 ms, limiting any direct comparisons to more recent studies of spectral modulation. Investigations that have considered the relationship between performance on speech perception tasks and performance on spectral modulation detection tasks also have measured spectral modulation superimposed on stimuli with long durations (e.g., Saoji, Litvak, Spahr, & Eddins, 2009) while the spectral modulations in speech are commonly on the order of 50–150 ms rather than 400 or 500 ms. If, for example, spectral modulation detection is inversely related to duration, then estimates of the strength of such relationships may be artificially low by virtue of the choice of stimulus duration.

Spectral envelope perception involves the encoding of patterns of intensity change across frequency and builds on the basic auditory abilities of intensity discrimination and frequency selectivity. A common measure of spectral envelope perception is spectral modulation detection, and when measured across a range of modulation frequencies, the spectral modulation transfer function (SMTF) provides a broad characterization of spectral envelope perception. Spectral modulation detection thresholds (MDTs) have been used to predict aspects of speech perception in listeners with hearing loss (e.g., Bernstein et al., 2013; Saoji et al., 2009). Van Veen and Houtgast (1985) illustrated the relation between spectral modulation and speech, noting for example that vowels can be distinguished on the basis of variations in their spectral modulation characteristics, with modulation frequencies near 2 cycles/octave being most different among vowel stimuli. Similarly, Qian and Eddins (2008) demonstrated the importance of the same modulation frequency range in variations in an elevation-related spectral shape introduced by the pinnae.

Spectral modulation detection is often measured using a noise carrier that is modulated such that the amplitude of that carrier varies sinusoidally on a logarithmic frequency axis from low to high audio frequency. Typically, the modulation phase is chosen at random to reduce the likelihood that detection is based simply on a local intensity comparison across noise bursts representing the flat-spectrum standard and the modulated signal in multi-interval listening tasks. This spectral modulation creates a series of peaks and valleys in the spectrum that, to a first approximation, are represented along cochlear space and, in theory, are represented tonotopically throughout the auditory system. As can be seen in Figure 1, increases in the modulation frequency from low (Panel A) to high (Panel B) lead to increases in the density of the corresponding spectral peaks and valleys and corresponding changes in estimated excitation patterns (Panels C and D, respectively; Moore & Glasberg, 2004). A typical SMTF is shown in Panel E. Threshold for any given modulation frequency, and thus the shape of the function, depends in part on the ability to detect a change in intensity across frequency. This ability may be limited by the frequency-resolving power of the auditory system. These basic abilities, combined with the ability of the system to compare intensity across a range of audio frequencies, ultimately determine the sensitivity to spectral modulation.

Accordingly, to anticipate the potential effects of stimulus duration and/or level on spectral modulation detection, one may consider the known effects of duration and level-on-level discrimination, frequency selectivity, and on other measures of spectral envelope perception auditory profile analysis. Because spectral modulation detection is known to be limited to some degree by the limited frequency-resolving power of the auditory system (e.g., Summers & Leek, 1994), we also consider how such limits may interact with stimulus duration and level in this context.

The Potential Effect of Duration on Spectral Modulation Detection

With respect to level discrimination, the spectral modulation detection task provides two possible cues. One is a "burst comparison" analogous to the investigation reported by Florentine (1986), and a second is an across-frequency intensity comparison analogous to the profile analysis task (e.g., Spiegel & Green, 1982). In terms of the burst comparison, if a fixed modulation phase is used, then a listener could, in theory, focus on one or more fixed frequency regions where the intensity in the signal interval is expected to increase or decrease. The data from Florentine (1986) indicate that such comparisons should result in systematic and nearly linear decreases in detection threshold with increasing duration over the range from a few milliseconds (e.g., 2 ms) to several seconds (e.g., 2 s). With random modulation starting phase, a focus on local-level differences across presentation intervals is made less reliable than with a fixed modulation phase, but it is the case that comparisons across intervals at some frequency regions might provide access to stable intensity differences. Thus, one may hypothesize that if local sequential comparisons of level across interval bursts is the cue for spectral modulation detection, then detection threshold should improve systematically and linearly over a wide range of stimulus durations from short to long based on data from Florentine (1986). In terms of simultaneous, across-frequency level comparisons, the spectral profile analysis method provides the most comprehensive body of data to date. With such methods, it is typical to have a standard stimulus made of multiple equal amplitude (typically 11–21) sinusoidal components logarithmically spaced over a wide frequency region (several octaves) and a signal stimulus with one or more components having a level increment. Because the

Figure 1. Examples of spectral modulation superimposed on a 3-octave carrier (400–3200 Hz). In Panel A, the signal stimulus has a modulation frequency of 0.5 cycles/octave (black line) at a 20-dB modulation depth (solid black line) and the standard stimulus is unmodulated (blue dashed line). Panel B shows a similar example for 2 cycles/octave. Panels C and D show estimated excitation patterns (EPs) for the standard unmodulated noise (black line) and modulated signal (blue dashed line). Panel E depicts an average spectral modulation transfer function (SMTF) for a 400-ms duration.
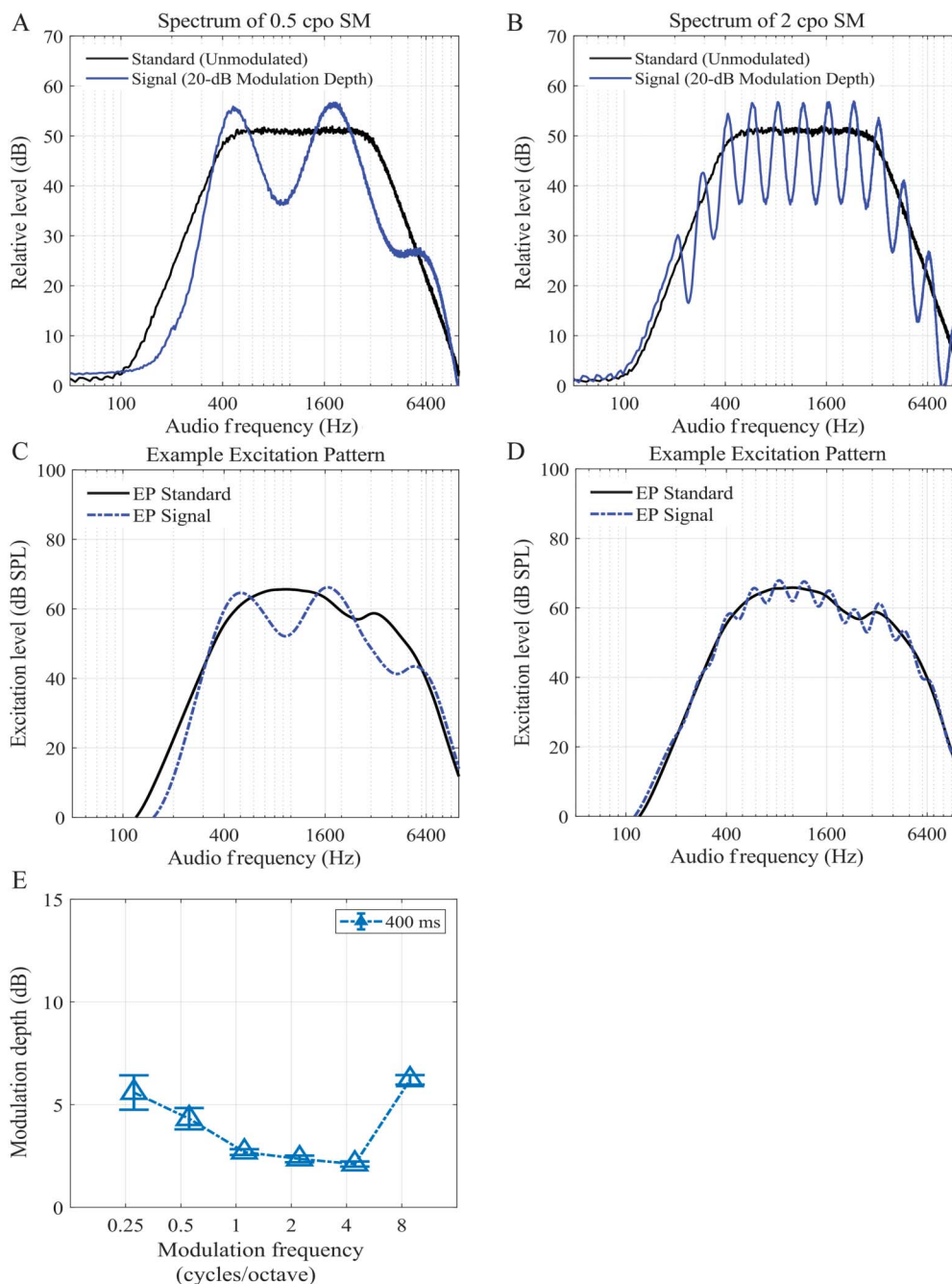
incremented component(s) is (are) fixed within a block of trials, the potential to make a local-level comparison across intervals is high. To minimize the potential to use such a cue, the overall stimulus level is randomly selected (i.e., roved) from interval to interval. In this case, Green, Mason, and Kidd (1984) and Dai and Green (1993) have shown that threshold for detecting the change in profile is dependent on duration below 100 ms, but not dependent on duration between 100 and 1000 ms. Thus, one may hypothesize that if simultaneous, across-frequency level

comparisons serve as the basis for spectral MDT, changes in duration should have the greatest impact on spectral modulation detection below about 100 ms, resulting in higher thresholds, whereas changes in duration beyond about 100 ms or so should have little impact on detection thresholds.

The Potential Effect of Level on Spectral Modulation Detection

To anticipate potential changes in spectral modulation detection with level, we can look to the same basic experimental methods. Intensity discrimination as a function of level ($\Delta I/I$ or the change in intensity divided by intensity) in a burst paradigm indicates a systematic reduction in $\Delta I/I$ from about 0.4 to about 0.1 as a function of sensation level over the range of 10–90 dB SL that is independent of frequency region (Jesteadt, Wier, & Green, 1977). On the contrary, spectral profile analysis depends little on the stimulus level (Mason, Kidd, Hanna, & Green, 1984), with only a slight decrease in threshold (~2.5 dB in units of $20\log_{10}[\Delta A/A]$ where A is amplitude and ΔA is a change in amplitude) over a wide range of levels. Note that this scale is highly expansive relative to the $\Delta I/I$ scale used by Jesteadt et al. (1977), which, when represented in units of $20\log_{10}[\Delta A/A]$, would have resulted in thresholds that spanned from about −14 to about −26 dB from 10 to 90 dB SL. We also know from the recent work of Magits et al. (2018) that spectrotemporal MDTs increase at high presentation levels ($\geq$ 75 dB SPL).

Thus, like duration, previous data lead to a testable hypothesis. If spectral modulation detection is based on sequential, across-interval level comparisons (i.e., across-interval bursts), then thresholds should be systematically dependent on level across a wide range of levels. In contrast, if spectral modulation detection is based on simultaneous, across-frequency level comparisons, then thresholds should not be strongly dependent on presentation level.

The Potential Effect of Frequency Selectivity on Spectral Modulation Detection

It is important to consider the potential impact of frequency selectivity on spectral modulation detection. As demonstrated by Summers and Leek (1994), spectral modulation detection for high modulation frequencies, where spectral peaks (and valleys) are closely spaced relative to the width of the auditory filter (e.g., Figures 1B and 1D), frequency selectivity will limit the internal spectral contrast and result in higher detection thresholds. Similarly, as shown by Eddins and Bero (2007), at low spectral modulation frequencies, frequency selectivity has progressively less impact on thresholds and changes in threshold with decreasing modulation frequency cannot be explained solely by the limits of frequency selectivity. To our knowledge, the only published investigation of the effect of duration on frequency selectivity was presented by Wright and Dai (1994). They used a notched noise method to estimate the filter

width and filter shape at 2500 Hz for short (5 ms) and long (295 ms) stimuli. The results indicated little change in filter width or shape with duration.

It is known that estimates of frequency selectivity vary with increasing level in a complex manner. In general, estimates of the width of the auditory filter increase monotonically with increasing level (e.g., Glasberg & Moore, 2000; Moore & Glasberg, 1987; Rosen & Stock, 1992). That increase in auditory filter width with level is small at very low frequencies (e.g., 125 Hz) and increases systematically with increasing frequency (Rosen & Stock, 1992). Likewise, the estimated shape of the auditory filter is level dependent, with a progressively shallower low-frequency side as stimulus level is increased. Since spectral modulation detection requires resolution of spectral peaks and valleys and limited frequency selectivity will effectively reduce the spectral contrast as the density of the modulation peaks and valleys increase, one would expect that spectral modulation detection at high modulation frequencies will be impacted by limited frequency selectivity. These results lead to the hypothesis that the stronger the dependence of spectral modulation detection on frequency selectivity, the more that spectral MDTs should increase (get worse) with increasing level. On this basis, one could put forth the hypothesis that higher spectral modulation frequencies should be impacted more with increasing level than lower spectral modulation frequencies.

At high spectral modulation frequencies, multiple modulation peaks could fall within a single auditory filter bandwidth, resulting in beating among peaks and thereby potentially providing a temporal cue to detection. If so, this temporal cue should be stronger for broader auditory filters, encompassing a greater number of spectral peaks. Because the auditory filter width increases with increasing level (Moore & Glasberg, 1987), this raises the possibility that such a temporal cue could be stronger for higher than lower presentation levels. On this basis, one might predict that threshold for 8 cycles/octave (the highest frequency tested here) might change with level to a greater extent than lower modulation frequencies as the stimulus increases from moderate to high presentation levels. Likewise, it is well established that the perception of temporal fluctuations is strongly dependent on carrier bandwidth (Eddins, 1993, 1999). Thus, if the carrier bandwidth is doubled, one might predict that any level effect at 8 cycles/octave would be even stronger than for a narrower bandwidth. Finally, if the increased bandwidth encroaches higher audio frequencies, where the relative auditory filter width, defined as the equivalent rectangle bandwidth divided by center frequency (e.g., ERB/f_c) increases (Moore & Glasberg, 1987), any temporal effects should be magnified even further. To evaluate these possibilities, we repeated the presentation level experiment for a carrier bandwidth of 3 octaves, spanning 400–3200 Hz to compare to a carrier bandwidth of 6 octaves, spanning 200–12800 Hz.

The goal of the current study is to evaluate the effects of level and duration on spectral modulation detection. The modulation detection task, in theory, could be based

on one of two fundamental processes. One is based on a simultaneous, across-frequency comparison of amplitude to encode spectral shape. The second is based on a sequential, across-interval (bursts), frequency-specific comparison of level. Randomization of modulation starting phase should discourage the use of the second process, reinforcing the use of overall spectral shape (the first process) to detect spectral modulation. On the basis of corresponding spectral profile analysis and level discrimination experiments that have manipulated stimulus duration, we evaluate several hypotheses. First, the dependence of spectral modulation detection on stimulus duration will be restricted to durations less than about 100 ms, in agreement with both types of experiments (e.g., Dai & Green, 1993; Florentine, 1986). Second, the sensitivity of the SMTF depends upon stimulus presentation level for presentation levels very near absolute detection threshold (i.e., for detecting the presence of the noise carrier). Third, based on the discussion of frequency selectivity above, we evaluate the hypothesis that, at higher presentation levels, any effect on spectral modulation detection should differentially impact higher relative to lower modulation frequencies. In addition to evaluating these hypotheses, this investigation will provide information that would be essential in the development of any models of auditory perception intended to encompass spectral or spectrotemporal modulation perception. We illustrate this by evaluating the output of a simple peripheral auditory model developed by Zilany and colleagues (Zilany, Bruce, & Carney, 2014; Zilany, Bruce, Nelson, & Carney, 2009) with two different decision statistics. Ultimately, we anticipate that mapping out the effects of level and duration on spectral modulation detection can provide theoretically and practically useful information.

Method

Participants

Participants included six young listeners (20–25 years of age) with normal audiometric hearing thresholds (≤ 20 dB HL) in the range of 250–8000 Hz. They had no history of middle ear disorders or ear surgery. Data collection was completed over 12–15 sessions lasting approximately 2 hr each, for a total of approximately 25–30 hr of testing per subject. The listeners provided written consent for study participation, and all procedures were approved by the university institutional review board. Participants were compensated for their participation time with an hourly wage.

Stimuli

All stimuli were generated in MATLAB (The Mathworks, Inc.). Stimuli were similar to those reported by Eddins and Bero (2007). The modulators were sinusoidal on a logarithmic frequency axis ($\log_2$) and a logarithmic amplitude scale (dB), such that the internal representation of the spectral modulation was, to a first approximation, sinusoidal (e.g., Figures 1C and 1D). The function representing the modulation waveform is shown by Equation (1). The spectral modulation frequencies were 0.25, 0.5, 1, 2, 4, and 8 cycles/octave. Modulators had a random starting phase (uniform distribution between 0 and 2π radians), and the modulation depth was specified as the peak-to-valley difference in dB, as shown in Figures 1A and 1B. The modulator spanned the full audio frequency spectrum and was scaled to the desired modulation depth prior to modulation of the carrier. The noise carrier stimuli had nominal bandwidths of either 3 octaves (400–3200 Hz) or 6 octaves (200–12800 Hz), with a slope outside the nominal bandwidth of −36 dB per octave. The sampling frequency was 40984 Hz.

$$x(f) = 10^{\frac{M}{20}\sin(2\pi\Omega f + \theta)}; M = \text{depth in dB}. \qquad (1)$$

Equipment

Digital stimuli were presented through a soundcard (Realtek High Definition Audio), and the analog output was amplified (Studio Linear Amplifier; SLA4) prior to routing to an insert earphone (ER-2; Etymotic Research) and presented to the left ear of the participant at experiment-specific levels as noted below. To calibrate stimulus level, earphones were coupled to a Zwislocki ear simulator (Bruel & Kjaer DB-100), fitted with a G.R.A.S. 40 AG ½″ externally polarized pressure microphone, connected to a G.R.A.S. 26 AK ½″ preamplifier, routed to a G.R.A.S. 12AA power supply, the output of which was measured with a Fluke 45 multimeter. Prior to calibrating the desired stimulus, relative level was established by coupling a sound calibrator (Bruel & Kjaer type 4230) directly to the microphone in the circuit described above.

Procedure

Participants were seated comfortably at a desk inside a double-walled, sound-attenuating chamber. Detection thresholds were estimated using a three-down, one-up adaptive staircase method estimating 79.4% correct detection (Levitt, 1971). Stimuli were presented in a three-interval, two-alternative, forced-choice presentation paradigm in which the first interval always consisted of the standard unmodulated stimulus. Responses were collected using a graphical user interface in the MATLAB environment. The graphical user interface featured three rectangular boxes from left to right that corresponded to Intervals 1, 2, and 3. During each interval, the respective box changed color. Subject responses were made by using a mouse device to click on either the Interval 2 or Interval 3 box. Feedback consisted of a red light above the interval button that was repeatedly flashed on and off over the correct intervals. A single threshold estimate was based on a block of 60 trials that included at least seven reversals. The first three reversals were always excluded from the threshold computation. The threshold for a run was based on an average of the modulation depth that occurred on the next even number of reversals, with a minimum of four reversals required to

compute a threshold. The final threshold for a condition was based on the average of three such blocks. Each participant completed both experiments, beginning with the duration experiment. Conditions within an experiment were presented in random order across participants. A given participant was provided two practice runs (~10 min) on their first duration condition prior to collection of the data reported here.

Duration Experiment

There were a total of 24 conditions (a combination of six modulation frequencies and four stimulus durations) including frequencies of 0.25, 0.5, 1, 2, 4, and 8 cycles/octave and durations of 50, 100, 200, and 400 ms. The noise carrier had a 6-octave bandwidth (200–12800 Hz), and stimuli were presented at an overall level of 81 dB SPL. For each adaptive track, the starting modulation depth was 25 dB, which was adjusted using a multiplicative step that initially was a factor of 1.587 dB (i.e., the next step down would be 15.749 dB) for the first three reversals, after which the factor was reduced to 1.122 dB. Threshold estimates for each block were based on the last even number of reversals obtained with the smaller multiplier.

Level Experiment

In this experiment, two carrier bandwidths were evaluated, a 3-ocative bandwidth (400–3200 Hz) and a 6-octave bandwidth (200–12800 Hz). To support presentation at specific sensation levels, thresholds for detecting the unmodulated bandpass noise carriers were measured first. The adaptive tracking procedure for these conditions had an initial stimulus level of 50 dB SPL that was varied adaptively using an additive step size that was 5 dB for the first three reversals and then was reduced to 2 dB for the remainder of the block of trials. Threshold estimates for each block were based on the last even number of reversals obtained with the smaller step size (after the three reversals). Spectral MDTs were measured for each modulation frequency (0.25, 0.5, 1, 2, 4, 8 cycles/octave) at stimulus levels of 10, 20, 30, 40, and 60 dB SL for a total of 60 conditions (six modulation frequencies, five levels, two bandwidths). The duration was fixed at 400 ms. The psychophysical methods for these conditions were the same as for the duration experiment. The order of conditions was randomized separately for each participant for each experiment.

Results

Effects of Duration

The effect of stimulus duration on spectral MDTs (modulation depth in dB) and modulation frequency (cycles/octave) is shown in Figure 2, with duration indicated by symbol color and type. Overall, the SMTFs demonstrate a shallow bandpass shape with peak sensitivity between 1 and 4 cycles/octave (mean threshold ranging from 4.22 to 5.38 dB across conditions for a duration of 400 ms). Visual inspection of Figure 2 indicates that the lowest

Figure 2. Spectral modulation detection versus stimulus duration. This graph shows spectral modulation detection thresholds as a function of modulation frequency (mean and standard error) with duration as the parameter indicated by symbol type.
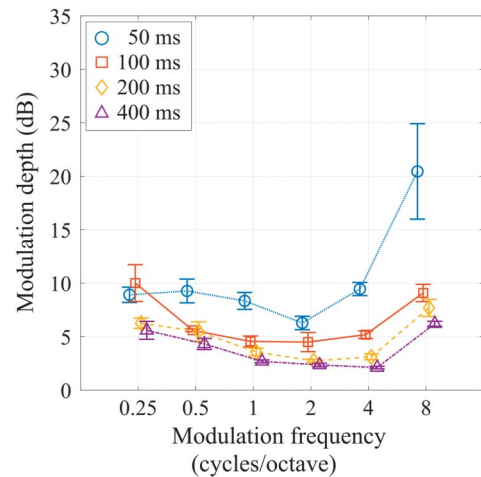
MDTs occur for the two longest durations (200 and 400 ms) and that thresholds are noticeably higher for the 50-ms condition. This effect is most pronounced for modulation frequencies greater than 0.5 cycles/octave. A two-way repeated-measures analyses of variance (ANOVAs) with a Greenhouse–Geisser correction for sphericity revealed statistically significant effects of modulation frequency, $F(1.44, 7.22) = 16.19$, $p = .003$; stimulus duration, $F(1.06, 5.28) = 27.27$, $p = .003$; and an interaction between stimulus duration and modulation frequency, $F(1.84, 9.18) = 4.70$, $p = .041$. Table 1 shows the results of separate post hoc one-way repeated-measures ANOVAs that were completed with duration as a factor for each modulation frequency. This illustrates that the significant effects were mainly driven by the duration effects at all modulation frequencies except at 0.25 cycles/octave. Visual inspection of Figure 2 at 0.25 cycles/octave shows that thresholds increased markedly for the 100-ms and 50-ms conditions.

Effects of Level

To evaluate the potential effects of presentation level, spectral modulation detection was measured at five

Table 1. F value of duration effect at each modulation frequency.

Spectral modulation frequency	Duration effect at each modulation frequency
0.25	$F(1.26, 3.00) = 3.903$, $p = .089$
0.5	$F(2.13, 10.65) = 8.423$, $p = .006$
1	$F(2.04, 10.22) = 28.51$, $p < .001$
2	$F(1.84, 9.18) = 15.66$, $p = .001$
4	$F(1.71, 8.53) = 75.13$, $p < .001$
8	$F(1.03, 5.13) = 10.54$, $p = .022$

sensation levels relative to absolute detection threshold for each individual participant. The average detection threshold for the unmodulated, 6-octave carrier was 22.5 dB SPL (SE = 2.0 dB) and ranged from 28 to 91 dB SPL. Average presentation levels ranged from 32.5 (10 dB SL) to 82.5 (60 dB SL).

Figure 3 displays the resulting SMTFs as a function of sensation level for the 6-octave carrier (200–12800 Hz), with symbol type and color denoting the different sensation-level conditions. Visual inspection reveals little difference in the resulting SMTFs for sensation levels of 20, 30, 40, and 60 dB. On the contrary, MDTs increase markedly for the 10–dB SL condition. A two-way repeated-measures ANOVA was completed with sensation level and modulation frequency for the 6-octave bandwidth. Using the Greenhouse–Geisser correction for sphericity, the results revealed significant main effects of modulation frequency, $F(2.21, 11.05)$ = 17.86, p < .001, and sensation level, $F(1.02, 5.10)$ = 8.348, p = .033. There was no significant interaction. Post hoc one-way repeated-measures ANOVAs were completed with sensation level as a factor for each modulation frequency. The results showed that the significant main effect was driven primarily by the effect of sensation level at 0.5 and 8 cycles/octave (see Table 2).

Effects of Carrier Bandwidth

To determine if the effect of level varies with carrier bandwidth, the level experiment was repeated for the 3-octave carrier (400–3200 Hz) at all previous modulation frequencies except 0.25 cycles/octave (only considering at least one and a half cycles of modulation). The average detection thresholds for the unmodulated, 3-octave carrier was 19.4 dB SPL (SE = 1.5 dB). In comparing across carrier bandwidths, a three-way repeated-measures ANOVA was computed
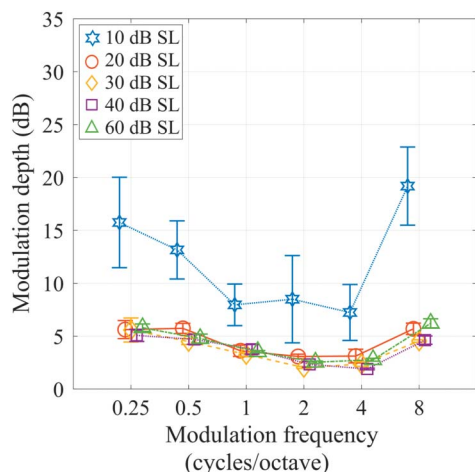
Table 2. F value of sensation-level effect at each modulation frequency.

Spectral modulation frequency	Sensation-level effect at each modulation frequency
0.25	$F(1.17, 5.87)$ = 4.84, p = .068
0.5	$F(1.16, 5.79)$ = 8.56, p = .025
1	$F(1.24, 6.18)$ = 4.73, p = .067
2	$F(1.01, 5.05)$ = 2.15, p = .203
4	$F(1.11, 5.53)$ = 3.12, p = .131
8	$F(1.07, 5.35)$ = 14.93, p = .010

with sensation level, modulation frequency, and bandwidth as within-subject factors. The results confirmed the same main effects of level and modulation frequency as previously identified, but no significant main effect of bandwidth, $F(1.00, 5.00)$ = 0.158, p = .708, or interactions, as shown in Figures 4A (SMTF comparison between 3- and 6-octave carrier bandwidths, dashed versus solid lines, respectively) and 4B (the differences between two carrier bandwidths).

In the introduction, several hypotheses regarding the potential availability of temporal cues to detection were discussed. It was noted that temporal modulation detection might be related to an interaction among adjacent spectral peaks, resulting in a temporal cue analogous to beats. This interaction should be greatest for the highest spectral modulation frequency, which has the most closely spaced spectral peaks. It also should increase with increasing auditory filter width, which in turn should increase as the presentation level increased from 40 to 60 dB SL. Contrary to this prediction, threshold at 8 cycles/octave changed little with increasing level from 40 to 60 dB SL (the largest change was from 20 to 10 dB SL). Auditory filter width would be greatest at the highest center frequency (CF) available, which would correspond to the 200- to 12800-Hz carrier bandwidth rather than the 400- to 3200-Hz carrier bandwidth. At 8 cycles/octave and the higher presentation levels, there was no difference in spectral MDT, consistent with the lack of a robust temporal cue.

Computational Auditory Model

In the introduction, it was suggested that models that are presumed to encompass fundamental auditory perception, including spectral shape perception, should be able to account for any effects of stimulus duration or level upon detection. To illustrate this concept, we considered a simple peripheral auditory model developed by Zilany and colleagues (Zilany et al., 2009, 2014) that was used previously to explain stimulus effects in spectrotemporal modulation detection (Magits et al., 2018). As a first step, we investigated whether or not this model would produce output qualitatively consistent with the effects of stimulus duration and level observed in the current behavioral data. For each duration or level condition, we submitted three stimulus types to the model: the unmodulated standard or carrier, the signal with a modulation depth equal to the average threshold shown in Figure 2 or 3, and the signal with a modulation

Figure 3. Spectral modulation detection versus presentation level (6-octave carrier bandwidth). This graph shows spectral modulation detection thresholds as a function of modulation frequency (mean and standard error) with stimulus sensation level as the parameter indicated by symbol type.

Figure 4. The effect of sensation level as a function of carrier bandwidth. Panel A displays a comparison of spectral modulation transfer functions (SMTFs) for two bandwidth conditions as a function of sensation level. The solid lines represent the results for a 6-octave carrier bandwidth (200–12800 Hz), and the dashed lines represent the results for a 3-octave carrier bandwidth (400–3200 Hz). Panel B shows the difference between the spectral modulation detection thresholds obtained with each of two bandwidth conditions (with standard error).
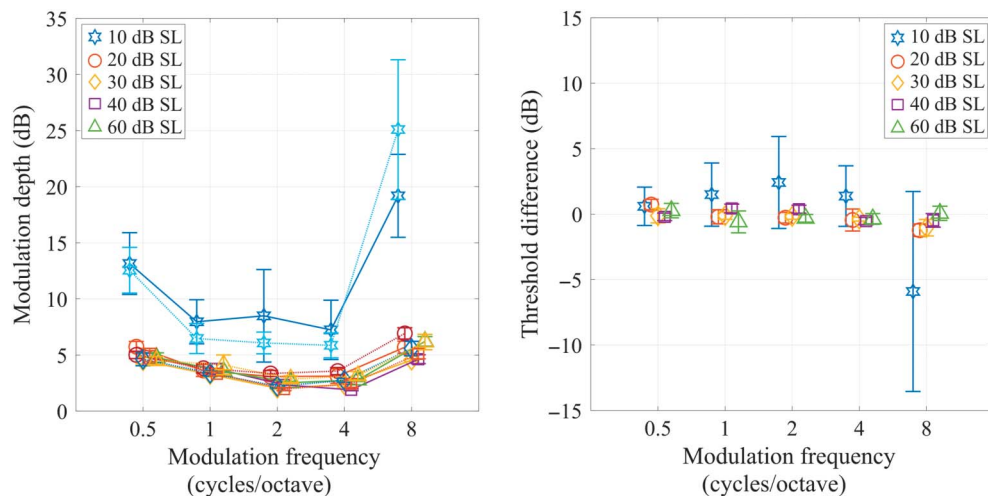
depth equal to 10 dB above the average threshold shown in Figure 2 or 3. To explore the effect of duration, the presentation level was fixed at 81 dB SPL, and duration was either 50, 100, 200, or 400 ms, as in the duration experiment above. To explore the effect of level, the stimulus duration was fixed at 400 ms, and level was either 10, 20, 30, 40, or 60 dB SL (relative to an average detection threshold of the unmodulated stimulus of 22 dB SPL).

Briefly, the peripheral model consists of an inner hair cell stage with a front end including a middle ear filter, a basilar membrane tuning, and the frequency offset of the control path filter (Zilany et al., 2009, 2014). The output is then fed into a "synapse" stage. Synapses of the auditory nerve (AN) are generated as a mean rate per CF, with 128 CFs logarithmically spaced between 125 and 15000 Hz. For each CF, the responses were simulated as the average of 50 AN fibers with different spontaneous rates: low (10), medium (10), and high (30). The resulting output across time and frequency has been termed the *early stage neurogram* (ESN), as shown in Figure 5A. The ESNs of the signal and standard were averaged across the full duration of the stimulus in each condition, as there were no obvious or expected temporal changes in the stimuli, referred to here as a *frequency profile* (see Figure 5B, signal = blue solid line and standard = black solid line). We adopted the first decision statistic from Magits et al. (2018), along with their terminology, to quantify the variability (i.e., dispersion) as shown in Equation (2), which is proportional to the stimulus conditions. The median of the interquartile frequency range at each duration or level was computed and then compared for the signal and standard versions of each stimulus, as shown in Figures 5C and 5D.

$$\text{Dispersion} = Mdn(\text{IQR}(\text{frequency profile})). \quad (2)$$

The simple assumption of any model of spectral modulation detection is that variations in the output of the model reflect variations in behavioral threshold in the modulation detection task. Thus, if the input to the model is the modulation depth that corresponds to behavioral threshold in each stimulus condition, then the output of the model should be constant across stimulus conditions (i.e., modulation frequency, stimulus duration, stimulus level) and equally different from the model output for the standard (unmodulated) stimulus for the same stimulus condition. The reference condition is the model output (dispersion) for the unmodulated standard condition (carrier alone) as shown by the asterisks for each stimulus duration (see Figure 5C) or presentation level (see Figure 5D), respectively. Clearly, the dispersion varies with both parameters, as indicated by separation among asterisks within a panel. The signal conditions reflect the model output (dispersion) corresponding to MDT (dashed lines) or modulation depths 10 dB above threshold (MDT + 10 dB; solid lines) for each condition shown as a function of modulation frequency in Figures 5C and 5D. If modulation detection is proportional to the change in dispersion (from the standard) produced by the signal modulation depth, then model output for the signal depths corresponding to MDT should be roughly equal across modulation frequency and equally different from the model output for the standard stimulus across stimulus condition (duration or level). The model data are not consistent across stimulus conditions, and the deviation is greatest for lower modulation frequencies and higher presentation levels.

As a second method of estimating sensitivity from the ESN, we subtracted the frequency profile for the standard condition from the frequency profile for a signal condition and then computed the maximum difference in the

Figure 5. Example of the auditory nerve (AN) model output for a spectral modulation frequency of 1 cycle/octave. (A): The so-called *early stage neurogram* (ESN). (B): The *frequency profile* for a modulated signal (solid blue line) and the unmodulated standard (solid black line) shown as the normalized neural activity as a function of the center frequency. (C): Dispersion plotted as a function of modulation frequency with stimulus duration and depth indicated by the solid and dashed lines. (D): Dispersion plotted as a function of modulation frequency with presentation level and depth indicated by the solid and dashed lines. (E): Dispersion difference plotted as in Panel C. (F): Dispersion difference plotted as in Panel D. See text for definitions of terminology.
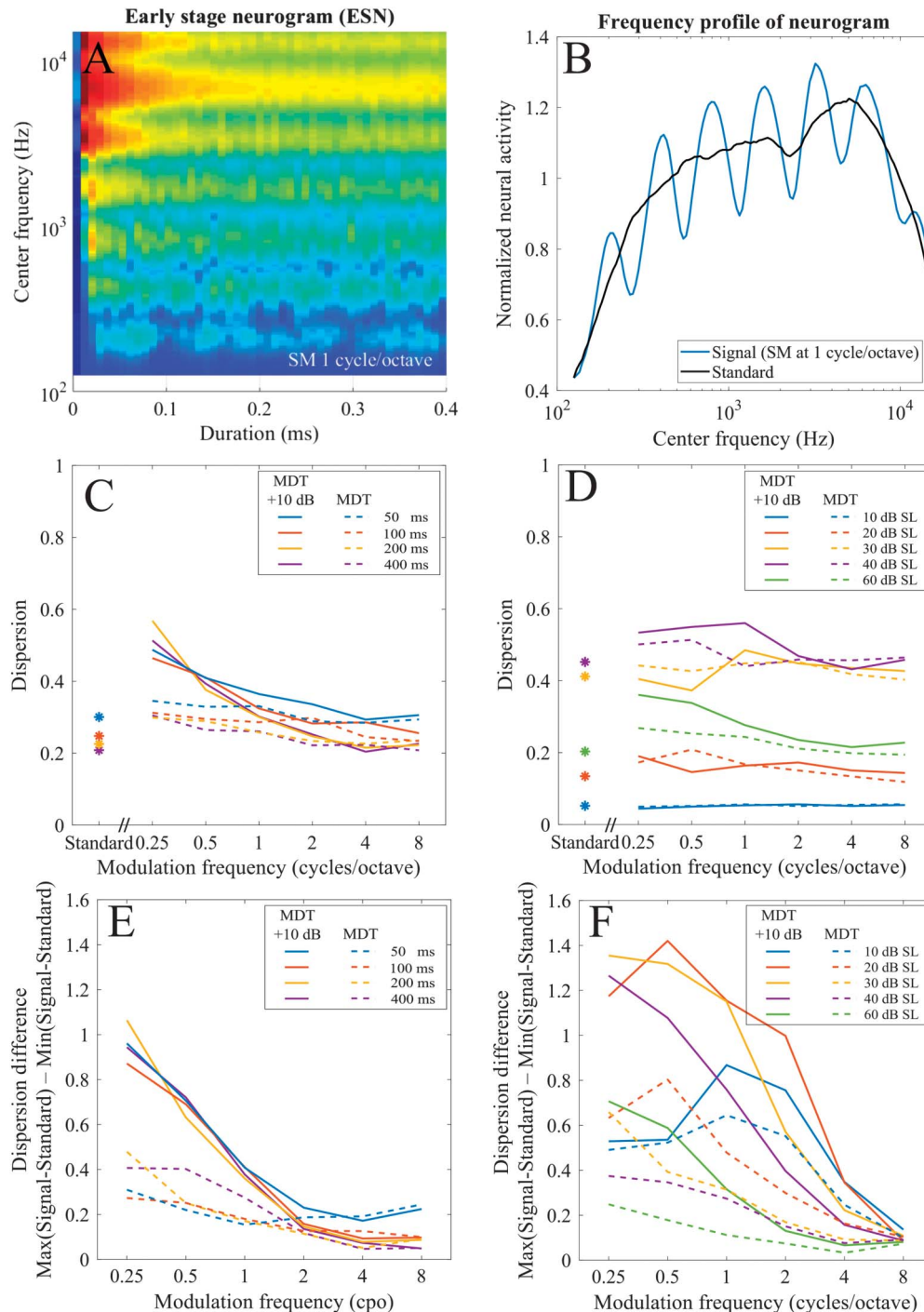
remaining function, referred to as the *dispersion difference*. Since the modulation depth for each condition at the input to the model was either equal to the average behavioral threshold or was relative to (10 dB above) behavioral threshold for that condition, the pattern of dispersion differences at the output of the model should be constant across conditions if the model accurately captures variations associated with the stimulus conditions. For the duration experiment (see Figure 5E), we first consider the results as a function of modulation frequency. The dispersion differences are smaller for the higher than the lower modulation frequencies and vary little from 4 to 8 cycles/octave. As the modulation frequency decreases, the dispersion differences increase. Thus, the model output is not constant as a function of modulation frequency, indicating that dispersion difference is not directly proportional to threshold. For the model to accurately capture changes in sensitivity to modulation with increasing duration, again one would expect the dispersion difference at threshold or 10 dB above threshold to be constant with duration. In fact, there is irregular variation with duration. For the level experiment (see Figure 5F), the dispersion differences again vary with modulation frequency, being smallest for 4 and 8 cycles/octave and increasing with lower modulation frequencies. Even more dramatic are the changes in model output with changing level. In the case of modulation frequency, stimulus duration, and stimulus presentation level, the model output is inconsistent with a constant dispersion difference at threshold and, to a first approximation, this inconsistency is inversely proportional to modulation frequency. In other words, the model is most closely related to behavioral performance for the highest modulation frequencies and increases in divergence as the modulation frequency decreases. This result is similar to the correspondence between changes in the excitation pattern and behavioral threshold as a function of modulation frequency, as demonstrated by Ozmeral, Eddins, and Eddins (2018) in the context of variations in spectral modulation threshold with hearing loss and age.

In summary, the instantiation of the simple AN model here and the associated qualitative analyses illustrate that the model output is not consistent with behavioral thresholds and their dependencies on modulation frequency, stimulus duration, and presentation level. In this regard, the present results are inconsistent with those of Magits et al. (2018), who reported output from the same model that was consistent with the detection of spectrotemporal modulation as a function of presentation level. We conclude that a more sophisticated model is needed to better capture stimulus dependencies in the spectral modulation detection task.

Discussion and Conclusions

The spectral MDTs measured in this study for long stimulus durations and moderate stimulus levels are consistent with previous measures of spectral modulation detection using similar durations and levels, both in terms of

absolute threshold value and the general bandpass shape of the SMTF (e.g., Eddins & Bero, 2007; Ozmeral et al., 2018). Thresholds were stable with decreasing stimulus duration from 400 to 200 ms. Further reduction in duration to 100 ms produced small but significant increases in MDTs, and reduction to 50 ms produced large increases in thresholds. Thus, modulation detection was robust for shorter durations until the stimulus was less than 100 ms. Such changes are consistent with previous data involving simultaneous, across-frequency level comparisons such as auditory profile analysis (e.g., Dai & Green, 1993) and are inconsistent with investigations involving sequential, across-interval level comparisons (e.g., Florentine, Buus, & Mason, 1987).

Similarly, spectral modulation detection changed little as the stimulus level was reduced from 60 to 20 dB SL. When the presentation level was 10 dB SL, however, spectral MDTs increased considerably. Thus, spectral modulation detection is quite robust to variations in presentation level between 20 and 60 dB above detection threshold for the unmodulated standard. This is promising, as it indicates that comparisons of spectral modulation detection among individuals with normal hearing and those with may be carried out at equal and modest sensation level values, such as 20 or 30 dB, or at equal suprathreshold levels, while avoiding a stimulus that is too loud for a hearing-impaired listener with hearing impairment with a markedly reduced dynamic range. Furthermore, changes with level were greatest for the highest spectral modulation frequency (8 cycles/octave), consistent with the greatest influence of level-dependent frequency selectivity on the modulation frequency with the highest spectral density. This pattern of results was similar for a 6-octave and a 3-octave bandwidth. At very low sensation levels, the shape of the internal representation of the modulator will deviate from sinusoidal, with low-amplitude portions of the spectrum being defined by audibility rather than the modulation shape, partially rectifying the modulator.

The lack of a bandwidth effect and the lack of a Bandwidth × Level interaction weakens any assertion that, at high modulation frequencies (i.e., 8 cycles/octave), a temporal cue due to beating spectral peaks within a single auditory filter could facilitate detection. Above, it was reasoned that three factors might increase the availability of any temporal cue: (a) increasing the level, thereby increasing the auditory filter bandwidth; (b) increasing the audio frequency region spanned by the carrier, thereby providing access to even broader auditory filter widths; and (c) increasing the bandwidth from 3 to 6 octaves (2800–12600 Hz), thereby increasing the sensitivity to temporal modulation. With all three combined, there remained no significant difference in MDTs, even at 8 cycles/octave.

Implementation of a simple AN model of auditory processing, the output of which has been related to spectrotemporal modulation detection, was not consistent with the present data. Comprehensive models of auditory perception should be able to account for basic stimulus parameter variables such as duration and level. It is possible that the AN model described here requires the inclusion of

subsequent stages of auditory processing to fully account for spectral modulation detection. The present results demonstrate significant interactions between spectral modulation frequencies and both level and duration, such that future tests that rely on spectral modulation sensitivity should take each factor into account. If the goal is to leverage the relationship between spectral modulation detection and speech perception in noise that exists in listeners with hearing loss, then one might choose a relatively low sensation level (e.g., 20–30 dB) to support comparisons across individuals with substantial hearing loss. Similarly, spectral modulation detection was stable for durations equal to or greater than 200 ms, increased slightly for a duration of 100 ms and increased markedly for a duration of 50 ms. These results were consistent with previous investigations of spectral envelope perception using the auditory profile analysis paradigm and should prove useful in the design and interpretation of future experiments involving spectral envelope perception.

Acknowledgments

References

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. https://doi.org/10.3766/jaaa.24.4.5

Dai, H., & Green, D. M. (1993). Discrimination of spectral shape as a function of stimulus duration. *The Journal of the Acoustical Society of America, 93*(2), 957–965.

Eddins, D. A. (1993). Amplitude modulation detection of narrow-band noise: Effects of absolute bandwidth and frequency region. *The Journal of the Acoustical Society of America, 93*(1), 470–479.

Eddins, D. A. (1999). Amplitude-modulation detection at low-and high-audio frequencies. *The Journal of the Acoustical Society of America, 105*(2), 829–837.

Eddins, D. A., & Bero, E. M. (2007). Spectral modulation detection as a function of modulation frequency, carrier bandwidth, and carrier frequency region. *The Journal of the Acoustical Society of America, 121*(1), 363–372.

Florentine, M. (1986). Level discrimination of tones as a function of duration. *The Journal of the Acoustical Society of America, 79*(3), 792–798.

Florentine, M., Buus, S. R., & Mason, C. R. (1987). Level discrimination as a function of level for tones from 0.25 to 16 kHz. *The Journal of the Acoustical Society of America, 81*(5), 1528–1541.

Glasberg, B. R., & Moore, B. C. (2000). Frequency selectivity as a function of level and frequency measured with uniformly exciting notched noise. *The Journal of the Acoustical Society of America, 108*(5), 2318–2328.

Green, D. M., Mason, C. R., & Kidd, G., Jr. (1984). Profile analysis: Critical bands and duration. *The Journal of the Acoustical Society of America, 75*(4), 1163–1167.

Jesteadt, W., Wier, C. C., & Green, D. M. (1977). Intensity discrimination as a function of frequency and sensation level. *The Journal of the Acoustical Society of America, 61*(1), 169–177.

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *The Journal of the Acoustical Society of America, 49*(2), 467–477.

Magits, S., Moncada-Torres, A., Van Deun, L., Wouters, J., van Wieringen, A., & Francart, T. (2018). The effect of presentation level on spectrotemporal modulation detection. *Hearing Research, 371,* 11–18. https://doi.org/101016/j.heares.2018.10.017

Mason, C. R., Kidd, G., Jr., Hanna, T. E., & Green, D. M. (1984). Profile analysis and level variation. *Hearing Research, 13*(3), 269–275.

Moore, B. C., & Glasberg, B. R. (1987). Formulae describing frequency selectivity as a function of frequency and level, and their use in calculating excitation patterns. *Hearing Research, 28*(2–3), 209–225.

Moore, B. C., & Glasberg, B. R. (2004). A revised model of loudness perception applied to cochlear hearing loss. *Hearing Research, 188*(1–2), 70–88.

Ozmeral, E. J., Eddins, A. C., & Eddins, D. A. (2018). How do age and hearing loss impact spectral envelope perception? *Journal of Speech, Language, and Hearing Research, 61*(9), 2376–2385.

Qian, J., & Eddins, D. A. (2008). The role of spectral modulation cues in virtual sound localization. *The Journal of the Acoustical Society of America, 123*(1), 302–314.

Rosen, S., & Stock, D. (1992). Auditory filter bandwidths as a function of level at low frequencies (125 Hz–1 kHz). *The Journal of the Acoustical Society of America, 92*(2), 773–781.

Saoji, A. A., Litvak, L., Spahr, A. J., & Eddins, D. A. (2009). Spectral modulation detection and vowel and consonant identifications in cochlear implant listeners. *The Journal of the Acoustical Society of America, 126*(3), 955–958.

Shamma, S. A., Elhilali, M., & Micheyl, C. (2011). Temporal coherence and attention in auditory scene analysis. *Trends in Neuroscience, 34*(3), 114–123.

Spiegel, M. F., & Green, D. M. (1982). Signal and masker uncertainty with noise maskers of varying duration, bandwidth, and center frequency. *The Journal of the Acoustical Society of America, 71*(5), 1204–1210.

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*(6), 3518–3528.

Van Veen, T., & Houtgast, T. (1985). Spectral sharpness and vowel dissimilarity. *The Journal of the Acoustical Society of America, 77*(2), 628–634.

Viemeister, N. F. (1979). Temporal modulation transfer functions based upon modulation thresholds. *The Journal of the Acoustical Society of America, 66*(5), 1364–1380.

Wright, B. A., & Dai, H. (1994). Detection of unexpected tones with short and long durations. *The Journal of the Acoustical Society of America, 95*(2), 931–938.

Zilany, M. S., Bruce, I. C., & Carney, L. H. (2014). Updated parameters and expanded simulation options for a model of the auditory periphery. *The Journal of the Acoustical Society of America, 135*(1), 283–286.

Zilany, M. S., Bruce, I. C., Nelson, P. C., & Carney, L. H. (2009). A phenomenological model of the synapse between the inner hair cell and auditory nerve: Long-term adaptation with power-law dynamics. *The Journal of the Acoustical Society of America, 126*(5), 2390–2412.