Discrimination of depth of sinusoidal amplitude modulation with and without roved carrier levels (L)

Mark A. Stellmack,[a] Neal F. Viemeister, and Andrew J. Byrne
Department of Psychology, University of Minnesota, Minneapolis, Minnesota 55455

(Received 16 May 2005; revised 13 September 2005; accepted 12 October 2005)

Thresholds for the discrimination of the depth of sinusoidal amplitude modulation with a broadband noise carrier were measured for three listeners in a two-alternative, forced-choice task for modulation frequencies of 8, 32, and 128 Hz. Thresholds were measured with the spectrum level of the carrier fixed at 20 dB across all trials and, separately, with the carrier spectrum level roved randomly over a 20-dB range (10–30 dB) in each interval. Mean thresholds were equal or slightly lower (but not significantly so) for the fixed conditions relative to the roved conditions, and the differences between thresholds were too small to be explained by assuming that listeners compared instantaneous intensity at corresponding phases of the modulation cycle (for example, in the troughs). Rather, it appears that listeners discriminated modulation depth by extracting an estimate of the modulation depth within each interval that was independent of the overall level. Consequently, models of envelope extraction must include normalization of the envelope fluctuations to the envelope dc. © *2006 Acoustical Society of America.* [DOI: 10.1121/1.2133576]

PACS number(s): 43.66.Mk [JHG] Pages: 37–40

I. INTRODUCTION

Amplitude modulation is an important information-bearing characteristic of sound. The ability of listeners to extract and utilize amplitude modulation has been assessed in tasks involving detection (e.g., Viemeister, 1979; Forrest and Green, 1987), discrimination of modulation depth (e.g., Ozimek and Sek, 1988; Wakefield and Viemeister, 1990; Ewert and Dau, 2004), and discrimination of envelope shape (e.g., Strickland and Viemeister, 1996), among others. Typically, the ability to perform these tasks is modeled on the basis of computation of a decision statistic that is related to the range of fluctuations in the amplitude of the stimulus envelope, such as the rms envelope power, ratio of maximum to minimum envelope amplitudes, or the ratio of the envelope maximum to envelope rms (e.g., Strickland and Viemeister, 1996). The efficiency of the auditory system in extracting these statistics is affected by its ability to encode the envelope (which varies with modulation rate) and by variability in the decision statistic either from internal or external (stimulus-based) sources of noise.

In amplitude modulation-detection tasks, one must control for the increase in power and potential concomitant loudness cue that occurs when a signal is amplitude modulated. One way in which the loudness cue can be controlled is to scale the modulated stimulus to compensate for the known increase in power (Viemeister, 1979). Another way is to impose a random rove in carrier level across trials such that the change in power associated with the presence of amplitude modulation will be so small relative to the overall rove as to be ineffective as a cue (Forrest and Green, 1987). Forrest and Green (1987) found that modulation detection thresholds for a broadband noise carrier were unaffected when they roved the carrier level over a 20-dB range (relative to thresholds with a fixed carrier level). One implication of this observation is that a successful model of modulation detection must include some normalization of the modulation envelope to its dc value in order to remove effects of a roved carrier level. The max-min statistic proposed by Forrest and Green, for example, implicitly incorporates such a normalization.

In tasks involving discrimination of the depth of amplitude modulation, the loudness cue associated with changes in power also may be a concern. In a study of modulation depth discrimination, Wakefield and Viemeister (1990) controlled for the loudness cue by compensating for the changes in total power associated with changes in the depth of sinusoidal amplitude modulation. They attempted to account for their data with a leaky-integrator model based on that of Viemeister (1979). In Wakefield and Viemeister's implementation of the model, the stimulus envelope was extracted through a process of half-wave rectification and low-pass filtering. Decisions were based on differences in the ac-coupled rms power of the extracted envelopes. The leaky-integrator model accurately predicted thresholds for standard modulation depths up to about -12.5 dB ($10 \log m^2$) but for higher standard modulation depths it predicted much lower thresholds than those measured behaviorally.

In their study of envelope discrimination, Strickland and Viemeister (1996) studied the ability of listeners to discriminate envelopes by measuring thresholds for detection of a signal modulator in the presence of a masker modulator, with both modulators applied to a single carrier. Adding the signal modulator to the masker modulator changes the effective modulation depth of the stimulus envelope. Like Wakefield and Viemeister (1990), Strickland and Viemeister (1996) controlled for loudness cues related to the signal modulator by compensating for the change in total power associated with its presence or absence. Strickland and Viemeister evaluated five different decision statistics in terms of their

---

[a] Electronic mail: stell006@umn.edu

ability to account for the data. The relevant point here is that all of the decision statistics evaluated by Strickland and Viemeister were based on representations of the envelope that were normalized to the envelope dc value and thus none predict an effect of a roved carrier level. The same is true of the model employed by Ewert and Dau (2004) to account for their modulation depth discrimination data.

While measuring modulation depth-discrimination thresholds in the context of a different study, it became apparent to the present authors that another possible cue existed for the depth-discrimination task that was not based upon estimates of the modulation depths *per se* of the standard and comparison stimuli or upon differences in overall loudness. In listening to the stimuli, it seemed that in some cases, particularly for large standard modulation depths and low modulation frequencies (below about 10 Hz), the cue for discrimination of modulation depth was related to the intensity in the troughs of the stimulus envelopes. In these cases, higher modulation depths were perceived as having gaps or brief moments of silence in the troughs of the intervals while the lower modulation depths seemed to have troughs that were "filled in." In other words, it seemed that modulation depth discrimination was based on detection of energy in the modulation troughs or discrimination of an estimate of instantaneous intensity in the troughs.

The present experiment was conducted to test whether or not a model based on discrimination of the intensities of the stimulus envelopes at corresponding modulator phases, such as at the modulator troughs, might account for modulation depth-discrimination thresholds to any extent. To our knowledge, this possibility has not been tested. The rationale behind the experiment is that if listeners were basing decisions on a comparison of instantaneous intensity across intervals, then an intensity rove imposed on the carrier across intervals should drastically interfere with modulation depth discrimination. In the present experiment, depth discrimination thresholds were measured with the spectrum level of the carrier fixed at 20 dB across intervals and, in separate conditions, with the spectrum level of the carrier randomly roved over a 20-dB range (10–30 dB). It will be seen that modulation depth discrimination thresholds were very similar in the two conditions, suggesting that listeners perform the depth discrimination task by comparing estimates of the modulation depths extracted from the standard and comparison stimuli rather than comparing estimates of instantaneous intensity across stimuli.

II. METHODS

Modulation depth-discrimination thresholds were measured in a two-interval, two-alternative forced-choice task. The stimulus within each interval was defined as

$$s(t) = A[1 + m\cos(2\pi f_m t + \phi)]n(t), \qquad (1)$$

where $n(t)$ was a broadband Gaussian noise, m was the modulation depth ranging from 0 to 1, f_m was the modulation frequency, ϕ was the starting phase of the modulator, and A was a scaling factor that determined overall stimulus level. One interval, chosen randomly with equal *a priori* probability from trial to trial, contained a stimulus carrying the standard modulation depth ($m=m_s$, fixed across trials within a block) while the stimulus in the remaining interval carried the comparison modulation depth ($m=m_c$), which was always less than the standard depth. The modulation frequency, f_m, was held constant within each block of trials at 8, 32, or 128 Hz. The starting phase of the modulator, ϕ, was randomized across intervals in order to minimize cues associated with differences in intensity at stimulus onset. Each stimulus was 1 s in duration with 500 ms of silence between the two intervals of each trial. Subjects were instructed to select the interval containing the larger (standard) modulation depth. Visual feedback indicating the "correct" interval was presented after each trial.

In one set of conditions, the unroved or fixed conditions, A in Eq. (1) was held constant across all trials such that the spectrum level of the noise carrier was fixed at 20 dB (measured at 1 kHz). In a second set of conditions, the roved conditions, the spectrum level (in dB) of every stimulus presentation was selected randomly and independently from a uniform distribution over the interval 10–30 dB. For a broadband noise carrier, detectability of amplitude modulation is independent of carrier level over a wide range of levels (Viemeister, 1979). In the fixed conditions, the rms amplitude of the signal interval was set equal to that of the nonsignal interval in order to minimize loudness cues associated with a difference in modulation depth (Viemeister, 1979). In practice, this control probably was unnecessary because the rms amplitudes of the signal and nonsignal intervals were essentially equal for these stimuli with identical carriers.

Thresholds were measured for standard modulation depths of 0, −6, and −12 dB ($10\log m_s^2$). By varying the modulation depth of the comparison, the difference between the modulation depths of the standard and comparison was varied adaptively from trial to trial in units of $10\log(m_s^2 - m_c^2)$, the intensity difference between the modulation depths of the standard and comparison expressed in dB. These units were chosen because Wakefield and Viemeister (1990) showed that psychometric functions expressed in terms of $\log d'$ vs $10\log(m_c^2 - m_s^2)$ are approximately parallel. (In the Wakefield and Viemeister experiment, the comparison modulation depth was always larger than the standard, thus the terms m_c^2 and m_s^2 are transposed in their expression.) The quantity $10\log(m_s^2 - m_c^2)$ was varied using a two-down, one-up procedure that estimated the 70.7% correct point on the psychometric function (Levitt, 1971). At the start of each block of trials, the value of $10\log(m_s^2 - m_c^2)$ was set 6 dB lower than $10\log m_s^2$. For example, when the standard modulation depth was −6 dB, $10\log(m_s^2 - m_c^2)$ was initialized to −12 dB. The initial step size in the adaptive procedure was set to 2 dB and was reduced to 0.5 dB after the first four reversals. Each block of trials was terminated after 12 reversals and the threshold was computed as the mean value of $10\log(m_s^2 - m_c^2)$ at the final eight reversals. The mean of four such threshold estimates was taken as the final estimate of threshold in each condition.

All stimuli were generated digitally in MATLAB at a sampling frequency of 44.1 kHz. For each trial, a 1-s broadband noise signal was generated in the frequency domain by drawing amplitudes from a Rayleigh distribution and starting phases from a uniform distribution for all components below 10 kHz (except for the dc component, which had an amplitude of zero). A time-domain signal was produced by applying an inverse fast Fourier transform (FFT) to the resulting spectrum. Cosine-squared ramps, 20 ms in duration, were applied to the onset and offset of each noise signal. The same noise carrier, $n(t)$, was presented in both intervals of a trial but an independent carrier was generated for each trial. The noise carrier was multiplied by the appropriate modulator to produce the stimulus for each interval.

The digital signals were generated and converted to analog signals on a PC equipped with a high-quality, 24-bit sound card (Echo Audio Gina). Stimuli were presented over Sony MDR-V6 stereo headphones to listeners seated in an IAC sound-attenuating chamber. Each block of trials was initiated by the listener. On each trial, a "ready" light flashed on the computer screen for 250 ms followed by a 100-ms pause after which a trial was presented. The intervals were marked visually by lights on the computer monitor. Listeners entered their responses on the computer keyboard at which time the correct response was indicated on the screen. Listeners were run in 2-h sessions, during which approximately 10–12 blocks of trials were run, until all stimulus conditions were completed.

All data for the conditions in which $10 \log m_s^2 = -6$ were gathered first. Within those conditions, two blocks of unroved-carrier trials were run at each modulation frequency, followed by two blocks of roved-carrier trials, then the sequence was repeated such that roved and unroved blocks of trials were interleaved. Data were gathered in the same way for standard modulation depths of −12 dB and then 0 dB. All listeners ran conditions with different standard depths on different days, alleviating concerns about the use of different standard modulation depths in successive blocks of trials (Wakefield and Viemeister, 1990; Ewert and Dau, 2004).

The three listeners consisted of the first and third authors and a female undergraduate student from the University of Minnesota who was paid to participate in the study. All listeners had pure-tone thresholds of 15 dB HL or better at octave frequencies from 250 to 8000 Hz. Listeners practiced the modulation depth-discrimination task with various modulation frequencies and standard depths until their thresholds stabilized.

III. RESULTS AND DISCUSSION

The performance of the three listeners was very similar, so only the mean data are shown. In Fig. 1, mean threshold values of $10 \log(m_s^2 - m_c^2)$ are plotted as a function of modulation frequency. The top, middle, and lower panels show data for $10 \log m_s^2 = 0$, −6, and −12 dB, respectively. Open symbols represent thresholds measured in the fixed conditions and closed symbols represent thresholds measured in the roved conditions. The error bars represent the standard error of the mean computed across the three threshold values of the three listeners.

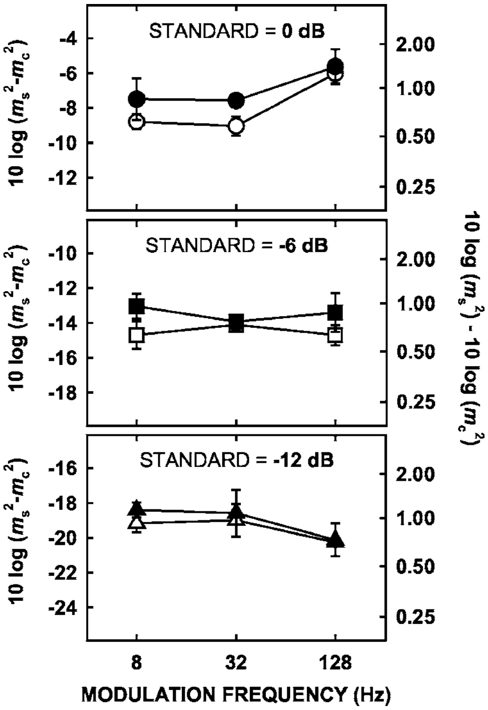

FIG. 1. Mean thresholds for three listeners in terms of the intensity difference between the modulation depths of the standard and comparison (in dB) as a function of modulation frequency. The right-hand axis scales the thresholds in terms of the modulation depth of the standard minus the modulation depth of the comparison (both expressed as $10 \log m^2$). Open symbols represent conditions in which the carrier level was fixed across trials. Closed symbols represent conditions in which the carrier level was randomly chosen from a 20-dB range of spectrum levels (between 10 and 30 dB). Error bars represent standard errors of the mean. Each panel shows data for a different standard depth (as indicated in the panels).

A three-factor, repeated measures ANOVA was performed on the thresholds expressed in terms of the dB difference between m for the standard and comparison (the right-hand axis of Fig. 1). The three factors were standard modulation depth (three levels), modulation frequency (three levels), and fixed versus roved carrier level. No main effects or interactions were statistically significant ($\alpha = 0.05$). The only visually apparent trend in Fig. 1 is that thresholds appeared to vary across standard modulation depth for $f_m = 128$ Hz, but once again this trend was not statistically significant. Most importantly in the present context, the main effect of fixed versus roved carrier level was not significant. It should be noted that an ideal observer that compares the intensities of the waveforms at corresponding points in the modulation cycle (for example, in the troughs, where the difference between instantaneous intensities is largest across intervals for modulated stimuli with fixed carrier levels) would require a change in m of approximately 8 dB in the presence of a 20-dB carrier rove in order to attain 70% correct by simply discriminating intensity at the troughs of the modulation.

The results suggest that listeners performed the modulation depth discrimination task by extracting an estimate of modulation depth that is independent of overall level. Based on the data presented here, it seems that it is appropriate and necessary for models of amplitude modulation processing, such as those noted in Sec. I, to include a normalization stage in which the modulation component of the envelope is normalized relative to its dc component, thereby eliminating effects of a rove of the carrier level.

ACKNOWLEDGMENTS

The authors thank Dr. John Grose and an anonymous reviewer for their comments. This work was supported by Research Grant No. R01 DC 00683 from the National Institute on Deafness and Communication Disorders, National Institutes of Health.

Ewert, S. D., and Dau, T. (**2004**). "External and internal limitations in amplitude-modulation processing," J. Acoust. Soc. Am. **116**, 478–490.

Forrest, T. G., and Green, D. M. (**1987**). "Detection of partially filled gaps in noise and the temporal modulation transfer function," J. Acoust. Soc. Am. **82**, 1933–1943.

Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.

Ozimek, E., and Sek, A. (**1988**). "AM difference limens for noise bands," Acustica **66**, 153–160.

Strickland, E. A., and Viemeister, N. F. (**1996**). "Cues for discrimination of envelopes," J. Acoust. Soc. Am. **99**, 3638–3646.

Viemeister, N. F. (**1979**). "Temporal modulation transfer functions based upon modulation thresholds," J. Acoust. Soc. Am. **66**, 1364–1380.

Wakefield, G. H., and Viemeister, N. F. (**1990**). "Discrimination of modulation depth of sinusoidal amplitude modulation (SAM) noise," J. Acoust. Soc. Am. **88**, 1367–1373.