Discrimination of temporally asymmetric modulation with triangular envelopes on a broadband-noise carrier (L)

Andrew J. Byrne,^{a)} Neal F. Viemeister, and Mark A. Stellmack
Department of Psychology, University of Minnesota, Minneapolis, Minnesota 55455

(Received 20 July 2010; revised 18 November 2010; accepted 26 November 2010)

Highly detectable, time-reversed triangular amplitude modulation, with linear increases and decreases in amplitude, was used in an adaptive task to measure just-noticeable differences for changes in the direction of envelope temporal asymmetry for different modulation depths ($m = 1.0$ and 0.5) and rates (8, 16, and 32 Hz). Thresholds were analyzed using three different measures of the modulator's shape based on (1) the change in the position of the peak within a cycle, (2) the change in the slope of the modulator's increasing amplitude portion, and (3) the change in slope measured in units of amplitude per unit cycle rather than amplitude per unit time. The amplitude per unit cycle measure resulted in the best fit to all the data, and predicted additional data that were gathered with roved modulation frequency. The results suggest that a time normalization process may be involved in the perception and discrimination of envelope shape. © 2011 Acoustical Society of America. [DOI: 10.1121/1.3531838]

PACS number(s): 43.66.Mk [TD]

Pages: 593–596

I. INTRODUCTION

The discrimination of the shape of a waveform's envelope has been studied with regards to its role in the discrimination of the depth of amplitude modulation (AM; e.g., Wakefield and Viemeister, 1990) and the detection of one modulator in the presence of others (modulation masking; Strickland and Viemeister, 1996; Viemeister *et al.*, 2005). Data from these studies suggested that, in some cases, discrimination of stimulus envelopes must be based upon a decision statistic that stems from time-domain analysis of the envelope rather than analysis of the envelope amplitude spectrum.

A particular class of stimuli, those with ramped and damped envelopes, provides a strong test of spectrally based models of modulation processing. When ramped and damped stimuli are generated such that they are merely time-reversed versions of one another (i.e., mirror images in the time domain), they have identical envelope amplitude spectra. If these stimuli are discriminable, it indicates that the listener is utilizing a decision statistic that is based on something other than the long-term amplitude spectrum of the modulation. Models that rely only on the amplitude spectrum, such as a simplified modulation filterbank model (Dau *et al.*, 1999; Ewert and Dau, 2000), cannot account for the discrimination of stimuli whose envelopes are time-reversed.

Ramped sinusoids and damped sinusoids were studied by Patterson (1994a,b), and it was shown that ramped and damped stimuli can be easily discriminated. Discrimination is presumed to be based on the different percepts of the two types of stimuli; a ramped stimulus is considered to have a more continuous "carrier-like" quality, while a damped stimulus has a greater "percussive" quality (Irino and Patterson, 1996). Other studies have indicated that damped stimuli are perceived to be quieter (Stecker and Hafter, 2000) and shorter in duration (Schlauch *et al.*, 2001) than their time-reversed

counterparts and that the ramped portion of the stimulus is weighted more heavily than the offset and likely dominates discrimination of asymmetric stimuli (Phillips *et al.*, 2002). In order to account for the discrimination of temporal asymmetry, different versions of an auditory-image model (AIM) have been developed (e.g., Patterson and Irino, 1998). These models use strobed temporal integration to capture envelope asymmetries and can predict discrimination in many cases. [For a more in-depth review of both the psychophysical and physiological literature involving temporally asymmetric waveforms, see to Phillips *et al.* (2002).]

Rather than the more transient envelopes of some of the previous studies, Akeroyd and Patterson (1995, 1997) measured detection and discrimination with different types of sawtooth-like periodic modulators with varying rates. In addition, a noise carrier was used to disrupt the earlier versions of the AIM which was based on the regularity of sinusoidal carriers (Patterson, 1994a,b). The discrimination results of Akeroyd and Patterson (1997) showed that upward-sloping cosinusoidal ramps are less discriminable from true sinusoidal AM than downward sloping cosinusoidal damps for a given modulation depth. They were not able to fully account for this trend using various modeling strategies, including the implementation of an AIM, so they suggested that the data might be explained by an unspecified decision model that they called an "envelope-shape recognizer."

The present study examines the possibility that the slope of a modulator's ramping portion could account for discrimination of its shape (in this case, the direction of its temporal asymmetry). Rather than measuring the minimum depth at which the waveforms can be discriminated, as done by Akeroyd and Patterson (1997), the present research uses highly detectable modulation to measure just-noticeable differences (JNDs) for changes in the triangular shape the modulators in an adaptive task. This method permits the examination of subtle differences in slope and shape, while also examining other possible decision statistics based on the time-domain properties of the envelope.

^{a)}Author to whom correspondence should be addressed. Electronic mail: byrn0050@umn.edu

FIG. 1. A schematic representation of the modulation envelopes for two values of P (columns) and two modulation depths (rows). For clarity, only two cycles of 16-Hz modulation are shown.

II. METHOD

The present experiment examined listeners' ability to discriminate the envelopes of amplitude-modulated noise. Each cycle of modulation had a linearly increasing and decreasing amplitude, such that both the envelope and its time-reversed version could be characterized by a simple slope metric. The modulation waveforms were characterized in terms of the proportion of each cycle (P) that was rising in amplitude (Fig. 1). In this scheme, an envelope for which $P = 0.5$ is a true, symmetrical triangle wave, while $P = 0.0$ and 1.0 are damped and ramped saw-tooth waves, respectively. If each cycle of modulation is considered to start at the minimum of a cycle, when $P < 0.5$, the peak of each cycle is shifted toward the beginning of the cycle and for $P > 0.5$ the peak is shifted toward the end. The envelope-magnitude spectra of stimuli with $P = x$ and $P = (1 - x)$ are equal, as they are merely time-reversed versions of one another.

Two modulation depths were used in separate conditions, $m = 1.0$ and 0.5 , with m defined as $(\max - \min) / (\max + \min)$, where $\max$ and $\min$ were the maximum and minimum amplitudes of the envelope, respectively (Fig. 1). Modulation frequencies (f_m) of 8, 16, and 32 Hz were used, in which the f_m was fixed for an entire block of trials. The modulation rates and depths were chosen to produce not only highly detectable envelopes but also highly discriminable envelopes at the extreme (sawtooth) peak positions based on Akeroyd and Patterson (1997); therefore, JNDs for more subtle changes in P could be measured. Using these variables, the slope of the modulator's ramp can be defined as $(m * f_m / P)$ in units of amplitude/time, with the slope being directly related to the modulation's depth and frequency and inversely related to the peak position.

The carrier consisted of broadband Gaussian noise with a frequency range of 1 Hz to 10 kHz and an average spectrum level of 30 dB. A new noise sample was generated for each trial, but the same sample was used for all three intervals within a trial. All stimulus intervals were 500 ms in duration with 25-ms raised-cosine on-and-off ramps and 250 ms of silence between intervals.

Data were gathered using an adaptive procedure in which the value of P in the signal interval (P_{SIG}) was changed from

trial to trial based on previous performance but was always less than $P = 0.5$, and the non-signal modulation's P (P_{NOS}) was set equal to $1 - P_{\text{SIG}}$. The starting phase of the modulation was randomized for each interval of every trial, and the root-mean-square (*rms*) amplitudes of the stimuli in all three intervals were equated. A cued, two-interval forced-choice (2-IFC) task was used with the first (cue) interval always carrying the non-signal modulation, and the second and third (listening) intervals having one signal and one non-signal modulation presented each trial. The listening interval containing the signal modulation was determined randomly with equal *a priori* probability. The listener's task was to choose the odd interval, i.e., the one with a modulator that was different from that of the cue and the remaining listening interval.

A three-down, one-up tracking procedure using geometric step sizes was used to estimate the 79.4% correct point of the listener's underlying psychometric function (Levitt, 1971). A block was terminated after 12 reversals and the geometric mean of the final eight reversals was defined as threshold. For the first four reversals, a step-size multiplier/divider of 1.25 on P_{SIG} was used, which was then reduced to 1.12 for the final eight reversals. The decision to use a geometric stepping procedure was based on the preliminary measurements of psychometric functions and the observation that they were orderly on logarithmic scales [d' as a function of $(P_{\text{NOS}} - P_{\text{SIG}})$, or ΔP]. The initial probe level was set above the expected threshold for easier discrimination at the start of each block of trials. In the event that the value of P_{SIG} was adapted below zero, the block was terminated and discarded; this only occurred during the initial training of the listeners.

Each block was initiated by the listener and each trial began with a "ready" light which flashed for 250 ms, followed by a 250-ms pause. Listening intervals were marked by lights, and the listener entered responses on a keyboard. The light marking the correct interval was then illuminated for feedback. Listeners typically ran in 2-h sessions and the order of conditions run was different for each listener. A condition was complete after a total of four blocks and the four threshold estimates from those blocks were averaged to yield the final threshold estimate for that condition.

Four subjects participated in the experiment (two males and two females). Two were authors, while the others were undergraduate students from the University of Minnesota who were paid to participate. All listeners had normal hearing based on pure-tone thresholds of 15 dB hearing level (HL) or better at octave frequencies from 250 to 8000 Hz. The subjects' performance was asymptotic before data collection began. Listeners were run individually in a sound-attenuating chamber (Industrial Acoustics Company, Bronx, NY). The stimuli were generated digitally with MATLAB on a personal computer (PC) equipped with a 24-bit sound card (GINA 3G, Echo Digital Audio, Santa Barbara, CA) and were presented over stereo headphones (MDR-V6, Sony Corporation, Tokyo, Japan) to only the listener's left ear.

III. RESULTS

The expectation that the decision variable for discrimination was the slope of the modulator's ramp was tested relative

FIG. 2. Average thresholds for four listeners using three different dependent variable metrics as a function of modulation frequency. Filled circles: $m = 1.0$; open squares: $m = 0.5$. On the left-side panels, the right-most discrete points are for the roved- f_m conditions (8–32 Hz). The dependent variable for each panel is described in the text. The error bars represent the standard error of the mean.

to other metrics describing aspects of the modulator's shape across the various experimental conditions. Figure 2 shows geometric mean thresholds for the four listeners plotted in three different ways as a function of f_m . In the upper left panel, thresholds are plotted in terms of the change in P (the change in proportion of cycle that was ramped) that was varied adaptively in the experiment. Thresholds expressed in this way show a difference between the modulation depths (filled circles vs open squares). Thresholds are reasonably similar across f_m , although the best performance was for the 16-Hz modulator.

The change in the slope of the ramp across intervals was expressed by the difference in normalized amplitude units per unit time, or $\Delta(m^*f_m/P)$, and is plotted in the upper-right panel of Fig. 2. This measure produced good agreement between the data for a given f_m , but there is an obvious discrepancy across f_m , suggesting that slope expressed in this way is not the decision statistic used by listeners for this type of envelope discrimination.

Given the trends of the previous two metrics, an additional slope-like metric was formulated in an attempt to align the data across the various conditions. This new decision statistic, $\Delta(m/P)$ or the change in slope in units of normalized amplitude per unit cycle, is shown in the bottom-left panel of Fig. 2 and describes the change in the ramp's slope relative to a cycle of modulation rather than time (i.e., normalized time for a cycle of modulation and independent of f_m). It can be seen that $\Delta(m/P)$ accounts for the data remarkably well in the variety of conditions represented.

The fact that the decision metric that best accounts for the data does not include f_m suggests that discrimination should be possible in the presence of roved f_m . In order to test this prediction, the discrimination task was performed again with roved f_m at each of the two modulation depths (fixed within a block), with f_m chosen randomly and independently for each interval of each trial from a rectangular distribution over the range of 8 to 32 Hz. The results of the roved- f_m condition are plotted as discrete points in the left-side panels of Fig. 2 (the top-right panel's units of slope are dependent on

the f_m and therefore vary with the roved f_m). The listeners' thresholds were slightly elevated by the uncertain modulation rate across intervals within a trial, but thresholds could be obtained nevertheless, and the thresholds for both modulation depths are similar on the $\Delta(m/P)$ scale.

Statistical tests were performed on the observed trends and are consistent with the qualitative observations noted above. A repeated measures analysis of variance (ANOVA) was performed for each of the three dependent variables with the factors of f_m (8 vs 16 vs 32 Hz vs roved f_m) and modulation depth ($m = 1.0$ vs 0.5). The dependent variable for the first 4×2 ANOVA was ΔP (Fig. 2, top-left panel); there were not significant differences between modulation frequencies, $F(3,9) = 2.96$, $p = 0.09$, but there was a significant difference for modulation depth, $F(1,3) = 121.48$, $p = 0.002$, and there was not a significant interaction, $F(3,9) = 1.31$, $p = 0.33$. For the dependent variable of $\Delta(m^*f_m/P)$ (Fig. 2, top-right panel), a 3×2 ANOVA was performed (without the inclusion of the roved- f_m data): f_m , significant, $F(2,6) = 19.65$, $p = 0.002$; modulation depth, not significant, $F(1,3) = 0.63$, $p = 0.49$; interaction, not significant, $F(2,6) = 0.90$, $p = 0.45$. Finally, for the dependent variable of $\Delta(m/P)$ (Fig. 2, bottom-left panel), there were no significant differences shown by a 4×2 ANOVA: f_m , $F(3,9) = 0.42$, $p = 0.75$; modulation depth, $F(1,3) = 0.99$, $p = 0.39$; interaction, $F(3,9) = 1.01$, $p = 0.43$.

IV. DISCUSSION

Over the two-octave range of f_m used, a constant change in the modulator's slope was not a good criterion for discrimination, while the $\Delta(m/P)$ metric does provide a good account of the data across modulation depths and frequencies. [However, preliminary examination with rates of 64 Hz and above has indicated that the $\Delta(m/P)$ criterion value does not remain constant, consistent with what would be predicted from the discrimination results of Akeroyd and Patterson (1997) at higher frequencies.] $\Delta(m/P)$ itself may not be immediately intuitive. It is the change in the slope of the

modulator relative to the proportion of a cycle over which it occurs, rather than the duration over which it occurs. The metric is simply the change in the slope of the ramp with f_m (or unit time) factored out.

Since the $\Delta(m/P)$ metric does not change with f_m , those units may quantify the cue, or the perceptual envelope shape, which the listeners used to perform the task. The fact that performance in the roved conditions was not significantly degraded from the fixed- f_m conditions indicates that the listeners were computing a statistic that was independent of modulation rate. From the listeners' subjective interpretation of the roved- f_m conditions, the task was performed based on similar percepts regardless of the randomized rate, with their decisions based on the relative "percussiveness" of the stimuli.

The $\Delta(m/P)$ metric may represent a decision statistic of the hypothesized envelope-shape recognizer and may account for some of the data of Akeroyd and Patterson (1997). The value of (m/P) as a function of P is a reciprocal function; therefore, the slope is much steeper when $P < 0.5$ than for $P > 0.5$. If the decision statistic for discrimination is the change in (m/P) , this function predicts that a modulator with $P = 0.5$ is less discriminable from one with $P = 1.0$ than from one with $P = 0.0$. This trend is consistent with the general differences in discriminability observed by Akeroyd and Patterson (1997) when using sinusoidal ramps, damps, and true sinusoidal AM (similar to P values of 1.0, 0.0, and 0.5, respectively).

Since an AIM-based modeling attempt failed to predict the results of Akeroyd and Patterson (1997), it is less likely that such a model would be able to account for the results of the present study even when f_m was fixed. A modulation filterbank model (e.g., Dau *et al.*, 1997a,b) may predict the fixed- f_m discrimination, if the model preserved phase information beyond 32 Hz. The ability to perform discrimination while roving f_m is an issue that would need to be addressed by any modeling attempt and a filterbank model with some type of time normalization included may do so.

The neural mechanisms underlying computation of the putative decision statistic [$\Delta(m/P)$] are highly uncertain, since it would involve some sort of time normalization. Presumably this requires higher-order perceptual processes for which no neural basis is evident. However, time normalization has been proposed in the speech perception literature (e.g., rate normalization, Miller and Liberman, 1979; dynamic time warping, Qi, 1992).

In summary, the present study reinforces the notion that temporally based models of envelope processing can account for discrimination of certain types of modulation, especially highly detectable, temporally asymmetric modulation at low rates (below 40 Hz). Furthermore, certain metrics related to the global shape of the stimulus envelope and independent of modulation frequency, such as $\Delta(m/P)$, may account for

discrimination of some types of AM and thus may begin to capture the physical correlates of perceptual envelope shape and its influence on timbre.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Torsten Dau and two anonymous reviewers who provided helpful comments and suggestions for improving this manuscript. This work was supported by Research Grant No. R01 DC 00683 from the National Institute on Deafness and Communication Disorders, National Institutes of Health.

- Akeroyd, M. A., and Patterson, R. D. (1995). "Discrimination of wideband noises modulated by a temporally asymmetric function," *J. Acoust. Soc. Am.* **98**, 2466–2474.
- Akeroyd, M. A., and Patterson, R. D. (1997). "A comparison of detection and discrimination of temporal asymmetry in amplitude modulation," *J. Acoust. Soc. Am.* **101**, 430–439.
- Dau, T., Kollmeier, B., and Kohlrausch, A. (1997a). "Modeling auditory processing of amplitude modulation. I. Detection and masking with narrow-band carriers," *J. Acoust. Soc. Am.* **102**, 2892–2905.
- Dau, T., Kollmeier, B., and Kohlrausch, A. (1997b). "Modeling auditory processing of amplitude modulation. II. Spectral and temporal integration," *J. Acoust. Soc. Am.* **102**, 2906–2919.
- Dau, T., Verhey, J., and Kohlrausch, A. (1999). "Intrinsic envelope fluctuations and modulation-detection thresholds for narrow-band noise carriers," *J. Acoust. Soc. Am.* **106**, 2752–2760.
- Ewert, S. D., and Dau, T. (2000). "Characterizing frequency selectivity for envelope fluctuations," *J. Acoust. Soc. Am.* **108**, 1181–1196.
- Irino, T., and Patterson, R. D. (1996). "Temporal asymmetry in the auditory system," *J. Acoust. Soc. Am.* **99**, 2316–2331.
- Levitt, H. (1971). "Transformed up-down methods in psychoacoustics," *J. Acoust. Soc. Am.* **49**, 467–477.
- Miller, J. L., and Liberman, A. M. (1979). "Some effects of later-occurring information on the perception of stop-consonant and semivowel," *Percept. Psychophys.* **25**, 457–465.
- Patterson, R. D. (1994a). "The sound of a sinusoid: Spectral models," *J. Acoust. Soc. Am.* **96**, 1409–1418.
- Patterson, R. D. (1994b). "The sound of a sinusoid: Time-interval models," *J. Acoust. Soc. Am.* **96**, 1419–1428.
- Patterson, R. D., and Irino, T. (1998). "Modeling temporal asymmetry in the auditory system," *J. Acoust. Soc. Am.* **104**, 2967–2979.
- Phillips D. P., Hall S. E., and Boehnke S. E. (2002). "Central auditory onset responses, and temporal asymmetries in auditory perception," *Hear. Res.* **167**, 192–205.
- Qi, Y. (1992). "Time normalization in voice analysis," *J. Acoust. Soc. Am.* **92**, 2569–2576.
- Schlauch, R. S., Ries, D. T., and DiGiovanni, J. J. (2001). "Duration discrimination and subjective duration for ramped and damped sounds," *J. Acoust. Soc. Am.* **109**, 2880–2887.
- Stecker, G. C., and Hafter, E. R. (2000). "An effect of temporal asymmetry on loudness," *J. Acoust. Soc. Am.* **107**, 3358–3368.
- Strickland, E. A., and Viemeister, N. F. (1996). "Cues for discrimination of envelopes," *J. Acoust. Soc. Am.* **99**, 3638–3646.
- Viemeister, N. F., Stellmack, M. A., and Byrne, A. J. (2005). "The role of temporal structure in envelope processing," in *Auditory Signal Processing: Physiology, Psychoacoustics, and Models*, edited by D. Pressnitzer, A. de Cheveigne, S. McAdams, and L. Collet (Springer-Verlag, New York), pp. 221–229.
- 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.