

The effect of temporal asymmetry on amplitude modulation detection using pure-tone carriers (L)^{a)}

Yi Shen^{b)} and Jennifer J. Lentz

Department of Speech and Hearing Sciences, Indiana University, Bloomington, Indiana 47405-7000

(Received 11 April 2011; revised 29 August 2011; accepted 1 September 2011)

The effect of temporal asymmetry on amplitude modulation detection was studied using sawtooth modulators with rising (ramped) or falling (damped) temporal envelopes within each period of modulation. For pure-tone carriers, damped modulation was more detectable than ramped modulation for a 5-kHz carrier (by a threshold difference of 3.2 dB on average) but not for a 1-kHz carrier. The threshold difference obtained at 5 kHz between the ramped and damped modulators was consistent across modulation rates (8–128 Hz). This carrier frequency dependence suggests that the effect of temporally asymmetry on modulation detection originates from envelope-based, within-channel mechanisms. © 2011 Acoustical Society of America. [DOI: 10.1121/1.3643813]

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

Pages: 2635–2638

I. INTRODUCTION

Sounds with gradual temporal onsets and sharp temporal offsets in their envelopes (“ramped sounds”) may be perceived very differently compared to their time-reversed versions (“damped sounds”) in terms of their timbre, loudness, subjective durations, and effectiveness in masking other sounds (see Phillips *et al.*, 2002, for a concise review). Because ramped and damped stimuli share identical long-term spectra, the measured perceptual differences cannot be easily explained by models based on long-term energy at the output of an auditory filterbank (e.g., Patterson, 1994a; Carlyon, 1996). The current study measures amplitude modulation (AM) detection thresholds (i.e., the lowest modulation depth at which ramped and damped stimuli can be distinguished from a pure tone) at high and low frequency regions to test two hypotheses that have been proposed to account for the perceptual effect of temporally asymmetric stimuli; one is based on across-channel, periodicity-coding mechanisms, and the other is based on within-channel, envelope-coding mechanisms.

The periodicity-based, across-channel hypothesis was proposed by Patterson (1994a,b). He modeled the perceived timbre difference between ramped and damped tones (sinusoids modulated by exponentially rising or falling envelopes, respectively) using a model of the auditory periphery, and then analyzed the temporal firing patterns of the modeled auditory nerve fibers across multiple frequency channels. He found that more frequency channels exhibited phase-locked responses to the carrier frequency for ramped compared to damped tones. He argued that the consistent periodicity across a broad frequency region might make ramped tones sound more similar to pure tones. We refer to this hypothesis as the across-channel hypothesis, because according to the proposed mechanism, peripheral frequency channels away

from the carrier frequency exhibit the greatest contrast in their representations of ramped and damped tones.

The perceived timbre difference between the ramped and damped noises, however, cannot be easily explained by the periodicity-based hypothesis because these noise carriers lack periodicity in their temporal fine structure. A few envelope-based, within-channel mechanisms have also been proposed to encapsulate data measured using ramped and damped noises (Akeroyd and Patterson, 1995; Irino and Patterson, 1996). The mechanisms are dependent upon a temporal asymmetry in the shape of the envelope. After these processes, stimuli with ramped envelopes produced relatively sustained responses while stimuli with damped envelopes produced pulsed responses. These distinct patterns of internal representations correspond well with the temporal masking patterns measured for ramped and damped stimuli. For example, Lentz and Shen (2011) found that the peaks and troughs in the temporal masking patterns for ramped noises were much less distinct than those of damped noises. These previous studies suggest the perceptual effect of temporal asymmetry can be coded via within channel, temporal envelope cues.

We can study the relative contributions of these cues by using tonal stimuli with low and high carrier frequencies. At low carrier frequencies, the bandwidth of a ramped or a damped tone exceeds the bandwidth of the auditory filter. In such situations, listeners may detect the presence of modulation based on whether they can hear the spectral sidebands introduced by the modulation (e.g., Strickland and Viemeister, 1997; Kohlrausch *et al.*, 2000). According to the across-channel hypothesis, the off-frequency components distinguish ramped from damped tones. Hence, the audibility of these side components might also differ for ramped and damped tones, causing performance differences in discriminating the ramped and damped tones from pure tones. On the other hand, because very few or no side components are within the same auditory filter with the carrier tone, ramped and damped envelopes are not well preserved after the peripheral filtering and so within-channel cues are not reliable. It is therefore not likely that the within-channel, envelope-based hypothesis

^{a)}Portions of this research were presented at the 157th meeting of the Acoustical Society of America [J. Acoust. Soc. Am. **125**, 2524 (2009)].

^{b)}Author to whom correspondence should be addressed. Electronic mail: shen.yi@uci.edu

would predict a difference in detecting ramped and damped modulations at low carrier frequencies. At high carrier frequencies, the bandwidth of a modulated stimulus is likely to be contained within a single auditory filter and so across-channel cues are not available. In this case, we would expect no difference in the AM detection threshold between ramped and damped tones according to the across-channel hypothesis. On the other hand, because the envelopes are not heavily distorted by the peripheral filtering, within-channel, envelope based mechanisms could lead to a difference in AM detection thresholds at high carrier frequencies.

Although there have been no previously reported data on AM detection using ramped and damped tones, Akeroyd and Patterson (1997) studied modulation detection using ramped and damped broadband noises. They did not observe a significant effect of temporal asymmetry in the AM detection thresholds. In the current study, we measure AM detection using ramped and damped tones to expand the findings of Akeroyd and Patterson (1997).

II. METHODS

Four young normal-hearing adults participated in the present experiment including the first author (S1), whose ages ranged from 21 to 27 years. All listeners had absolute thresholds lower than 15 dB HL in both ears and in the frequency range from 0.5 to 8 kHz. The left ear of each listener was used in the experiment. Subjects were paid for their participation, with the exception of the author. All listeners had previous experience in AM detection tasks. The experiment was conducted in 2-h sessions with no more than one session conducted on a single day for any listener. It typically took 2–3 experimental sessions for the subjects to complete the experiment, including at least one hour of training before data collection began.¹

Modulation detection thresholds were measured for amplitude-modulated pure-tone carriers and three types of modulators. In the reference conditions, sinusoidally amplitude-modulated (SAM) tones were used as the stimuli, which were identical to those typically used in previous AM detection studies (Kohlrausch *et al.*, 2000; Moore and Glasberg, 2001). The other two types of modulators were periodic sawtooth waves (generated via the MATLAB command SAWTOOTH). For these modulators, the amplitude maximum occurred at either the beginning or the end of a modulator period, producing a damped or ramped modulator, respectively.² All modulators were low-pass filtered at 300 Hz to prevent excessive spectral splatter before being applied to the carrier. This low-pass filtering only slightly

smoothed the peaks and dips of the sawtooth modulator and had no effect on the modulation depths of the resulting stimuli.³ The initial phase of the modulator was drawn from a uniform distribution ranging from 0 to 2π on each presentation.

For each modulator type, the modulation depth m required for AM to be just detectable was measured as a function of modulation rate. The stimulus carrier frequencies were 1 and 5 kHz. Five modulation rates (f_m) were used: 8, 16, 32, 64, and 128 Hz. Each stimulus presentation lasted 500 ms including 50-ms onset/offset squared-cosine ramps. The stimulus level was fixed at 70 dB SPL for the 1-kHz carrier frequency.⁴ Note that the actual presentation level at 5 kHz was lower than 70 dB SPL (by about 8 dB) due to the high-frequency roll-off in the headphones' frequency response.

All stimuli were generated digitally at a sampling rate of 22 050 Hz and presented monaurally to the listener's left ear via a sound card (Digital Audio Labs CardDeluxe), an attenuator (Tucker-Davis Technologies PA4), a headphone buffer (Tucker-Davis Technologies HB6), and a headphone (Sennheiser HDA280). The experimental sequence and tracking procedure were controlled by a customly written program running on the experimental PC. All measurements were carried out in a sound-attenuated booth.

Figure 1 shows the spectra of sample test stimuli for the three modulator conditions at a modulator rate of 32 Hz and a modulation depth of -10 dB ($m \approx 0.32$). For the SAM tone (right panel), the spectrum contains a center component corresponding to the carrier frequency f_c and two side components with the modulation rate determining the spacing between the center and side components. The magnitudes of the side components increase with increasing modulation depth. The spectra for the ramped and damped tones (left two panels) consist of a series of equally spaced side components on each side of the center component. A higher modulation rate corresponds to a wider spacing among the side components, and so the stimulus bandwidth increases with increasing modulation rate. The ramped tone is the time reverse of its damped counterpart, therefore the two stimuli have identical long-term power spectra.

At a given modulation rate, the ramped and damped tones have the same frequency span independent of the carrier frequency. In contrast, the auditory-filter bandwidth is much narrower at 1 kHz compared to 5 kHz. One equivalent rectangular bandwidth (ERB, see Glasberg and Moore, 1990) centered at the carrier frequencies of 1 and 5 kHz is indicated by the dashed and dotted lines in each panel of Fig. 1, respectively. For this modulator rate (32 Hz), most of the sound

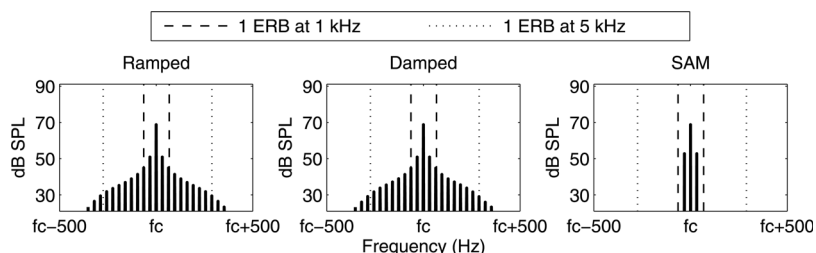


FIG. 1. Illustration of power spectra for the ramped tones (left), damped tones (middle), and SAM tones (right). f_c indicates the carrier frequency on the x -axis. For this illustration, $f_m = 32$ Hz, and the modulation depth = -10 dB ($m \approx 0.32$). Dashed and dotted lines in each panel denote the upper and lower boundary of one equivalent rectangular bandwidth centered at the carrier frequencies of 1 and 5 kHz.

power in the ramped or damped tones falls within an ERB at 5 kHz, whereas the stimulus bandwidths at 1 kHz typically exceed the auditory-filter bandwidth. Therefore, threshold differences between detecting ramped and damped AM measured at 5 kHz would support a contribution of the within-channel envelope-based cues in the effect of temporal asymmetry, whereas threshold differences observed at 1 kHz would likely reflect contributions from the across-channel, periodicity-based mechanisms.

In order to determine modulation detection thresholds, a two-interval, two-alternative forced choice task was implemented together with a three-down 1-up adaptive tracking procedure. The procedure estimated thresholds at the 79.4% point on the psychometric function (Levitt, 1971). For each trial, the listener was presented a modulated stimulus and an unmodulated stimulus in random sequence with a silent inter-stimulus interval of 400 ms. Listeners were instructed to identify the interval that contained the modulated stimulus. At the beginning of each track, the initial modulation depth was 0 dB in the form of $20 \log m$ (according to $m = 1$ or full modulation), which was also the upper limit of the modulation depth. The step size for the first two reversals was 8 dB; it was 4 dB for the following two reversals; and it was reduced to 2 dB for all remaining reversals. Each track terminated after 10 reversals; a threshold was determined by calculating the average modulation depth at the last 6 reversals. For all listeners, the 5-kHz carrier frequency condition was tested before the 1-kHz condition. Within each carrier frequency condition, the five modulation rates were tested in random sequence and were repeated four times. The threshold data reported here were based on the average of these four repetitions.

III. RESULTS AND DISCUSSION

The average modulation detection thresholds are shown in the upper panels of Fig. 2 as functions of f_m for carrier frequencies of 1 kHz (left) and 5 kHz (right). At 1 kHz, the thresholds decreased with increasing modulation rate for all three modulator types. Moreover, thresholds for the ramped and damped tones were within 3 dB of each other across all modulation rates. To better visualize the effect of temporal asymmetry, the average differences in threshold from the ramped to the damped conditions are shown in the lower left panel of Fig. 2. The threshold difference was within 1 standard error (see error bars) of 0 dB for three of the five modulation rates. The grand average of the threshold differences across listeners and modulation rates was about 1.6 dB. A within-subject analysis of variance (ANOVA) using modulator type (ramped versus damped tones) and modulation rate (five levels) as the two factors revealed neither a significant effect of modulator type [$F(1, 3) = 2.53, p = 0.21$] nor a significant interaction between the two factors [$F(4, 12) = 0.716, p = 0.60$] was found. The only significant effect observed was from modulation rate f_m [$F(4, 12) = 25.5, p < 0.001$]. The significant effect of modulation rate reflected that as modulation rate increased, the stimulus bandwidth also increased, possibly due to more separately audible spectral sidebands. Figure 2

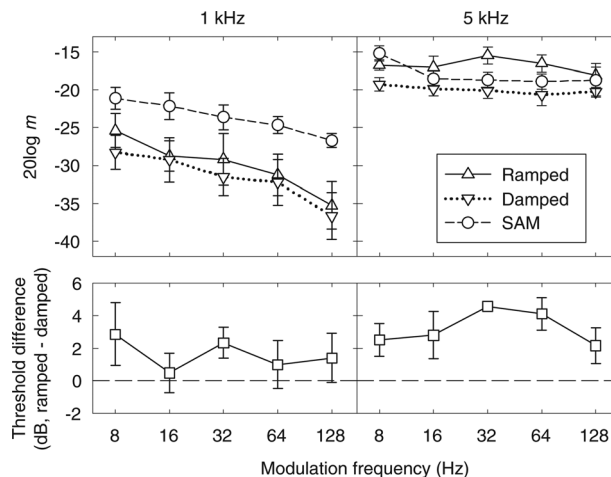


FIG. 2. Average thresholds across listeners for the ramped (upward triangles), damped (downward triangles), and SAM (circles) conditions at 1 kHz (upper left panel) and 5 kHz (upper right panel) carrier frequencies together with the average threshold differences between the ramped and damped conditions (lower panels). A positive threshold difference in the figure means better performance in the damped condition. Error bars indicate \pm one standard error of the means.

also shows that thresholds were lower for the ramped and damped tones compared to the SAM tones, which was expected if the stimulus energy in the off-frequency region dominated performance, as the ramped and damped tones had broader spectral spans than the SAM tones.

The results at the 5-kHz carrier frequency show a different pattern of thresholds compared to the 1-kHz condition. The threshold varied little with f_m . Thresholds in the ramped conditions were typically higher than those in the damped conditions, indicative of an effect (albeit small) of temporal asymmetry. The lower right panel of Fig. 2 shows the average difference in threshold between the two stimulus types. Threshold differences were at least 2 standard errors (see error bars) away from 0 dB. The grand average of the threshold differences across listeners and modulation rates was about 3.2 dB. A within-subject ANOVA using modulator type (ramped versus damped tones) and modulation rate (five levels) as the two factors revealed a significant effect of modulator type [$F(1, 3) = 204, p < 0.001$], no significant effects of f_m [$F(4, 12) = 0.982, p = 0.45$] and no significant interaction between the two factors [$F(4, 12) = 0.867, p = 0.51$]. Therefore, an effect of temporal asymmetry was found at the 5-kHz carrier frequency, and this effect did not depend on modulation rate. This result indirectly agrees with a recent study by Byrne *et al.* (2011), which suggested that the discriminability between ramped and damped stimuli depends on the contrast in the onset slopes of the modulators relative to a modulator cycle and is independent of modulation rates (for $f_m < 64$ Hz).

It is worth noting that for the SAM tones thresholds were generally higher at 5 kHz than at 1 kHz. This result has not been found in previous studies that have shown little effect of carrier frequency on AM detection below 100-Hz modulation rates (Kohlrausch *et al.*, 2000; Moore and Glasberg, 2001). Here, the thresholds obtained at 5 kHz are approximately 5 dB higher than previously reported. The

elevated thresholds may be due to practice effects, as listeners always ran the 5-kHz conditions first, or they are perhaps caused by the slightly lower presentation level at this frequency due to the headphones' frequency response (Kohlrausch *et al.*, 2000).

Inconsistent to the effect of temporal asymmetry observed at 5 kHz, Akeroyd and Patterson (1997) measured AM detection using ramped and damped broadband noises (0.1–8 kHz) and did not detect a significant threshold difference. There are several differences in experimental design between their study and the current study. First, different types of carriers are used in the two studies. Greater effect of temporal asymmetry can typically be measured using pure-tone than noise carriers (Akeroyd and Patterson, 1995; Irino and Patterson, 1996), presumably due to the probabilistic nature of broadband noise. This difference in the carrier characteristics might have made the threshold difference between the ramped and damped AM measurable using pure-tone but not noise carriers. Second, the modulators used are defined differently between the two studies. Their modulator is constructed by repeating either a rising or falling half-cycle of a sinusoidal function, whereas sawtooth waves are used in the current study. However, stimuli modulated using the half-cycle sinusoidal and sawtooth functions have very similar spectra. On each spectral component, the intensity difference is less than 1 dB. It is not likely that the slight spectral difference would account for the difference in the results.

The fact that the effect of temporal asymmetry in modulation detection was reliably observed at 5 kHz but not at 1 kHz does not support the across-channel hypothesis for the effect of temporal asymmetry. On the other hand, the processing stages that encode the within-channel, envelope cues might respond more robustly to the damped envelope patterns than the ramped patterns. This is consistent with previous studies measuring discriminability of noise carriers with temporally asymmetric AM. Akeroyd and Patterson (1995) showed that a spectral model monitoring phase-locking across multiple frequency channels could not account for the discrimination of ramped and damped noises, because the noise carriers did not produce neural firing patterns with regularity. On the other hand, within-channel, envelope-based models, such as the strobe integrator with adapting threshold by Akeroyd and Patterson (1995) or the delta-gamma process by Irino and Patterson (1996), gave reasonable predictions to the discriminability of ramped and damped noises. The AM detection results presented here are also in agreement with previous findings that damped sounds exhibit higher modulation depths in their internal envelopes than ramped sounds (e.g., Lentz and Shen, 2011).

Phillips *et al.* (2002) suggested that the perceptual effect of temporally asymmetric envelopes originates in the central auditory nervous system, which respond to temporal onsets more readily and accurately than the offsets. Whether physiological or phenomenological, all previously proposed within-channel mechanisms share a common characteristic: a nonlinear process that is sensitive not only to the power

spectrum of the envelope but also the temporal characteristics of the envelope. The current study, as well as aforementioned studies on the perceptual effect of temporal asymmetry, points to the importance of considering any asymmetry in the rise and fall in modulation in the future models of auditory temporal processing.

ACKNOWLEDGMENTS

This work was supported by the Faculty Research Support Program at Indiana University.

¹The training consisted of one repetition of all modulation rates at the 5-kHz carrier frequency (see the descriptions of experimental conditions later).

²The sawtooth modulator was implemented here as a more-common alternative to the modulator used by Akeroyd and Patterson (1997), which was made of rising or falling half-cycles of sinusoids. These two types of modulators are very similar acoustically. Sawtooth modulators were also used in a recent study investigating the discriminability between the ramped and damped noises (Byrne *et al.*, 2011).

³The rms amplitude of the modulator was always normalized to $\sqrt{2}/2$ before applying AM to the carrier.

⁴The overall stimulus level was scaled to 70 dB SPL before each presentation, so that it was not affected by the modulation depth. This is equivalent of applying a scalar to the amplitude to compensate the level change introduced by imposing AM (e.g., Viemeister, 1979).

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.

Byrne, A. J., Viemeister, N. F., and Stellmack, M. A. (2011). "Discrimination of temporally asymmetric modulation with triangular envelopes on a broadband-noise carrier," *J. Acoust. Soc. Am.* **129**, 593–596.

Carlyon, R. P. (1996). "Spread of excitation produced by maskers with damped and ramped envelopes," *J. Acoust. Soc. Am.* **99**, 3647–3655.

Glasberg, B. R., and Moore, B. C. (1990). "Derivation of auditory filter shapes from notched-noise data," *Hear. Res.* **47**, 103–138.

Irino, T., and Patterson, R. D. (1996). "Temporal asymmetry in the auditory system," *J. Acoust. Soc. Am.* **99**, 2316–2331.

Kohlrausch, A., Fassel, R., and Dau, T. (2000). "The influence of carrier level and frequency on modulation and beat-detection thresholds for sinusoidal carriers," *J. Acoust. Soc. Am.* **108**, 723–734.

Lentz, J. J., and Shen, Y. (2011). "Investigating temporal asymmetry using masking period patterns and models of peripheral auditory processing," *J. Acoust. Soc. Am.* **129**, 3194–3205.

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

Moore, B. C., and Glasberg, B. R. (2001). "Temporal modulation transfer functions obtained using sinusoidal carriers with normally hearing and hearing-impaired listeners," *J. Acoust. Soc. Am.* **110**, 1067–1073.

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.

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.

Strickland, E. A., and Viemeister, N. F. (1997). "The effects of frequency region and bandwidth on the temporal modulation transfer function," *J. Acoust. Soc. Am.* **102**, 1799–1810.

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