Viemeister, N.F., Stellmack, M.A., and Byrne, A.J. (2005).
The role of temporal structure in envelope processing.
In Pressnitzer, D., de Cheveigne, A., McAdams, S., and Collet, L. (Eds.)
Auditory Signal Processing: Physiology, Psychoacoustics, and Models.
Springer-Verlag, New York, 221-229.

The role of temporal structure in envelope processing

Neal F. Viemeister, Mark A. Stellmack, and Andrew J. Byrne

Department of Psychology, University of Minnesota, {nfv, stell006, byrn0050}@umn.edu

1 Introduction

The ability to extract information from relatively slow amplitude changes, the envelope, appears to be a crucial aspect of auditory communication. The general question we address is how can this ability best be understood and described. One approach, the "modulation filterbank" (MF) model (Dau, Kollmeier, and Kohlrausch 1997; Ewert and Dau 2000) postulates that there are filters within the auditory system that are selectively tuned to envelope frequency. This, essentially, is a translation of the well-documented notion of auditory spectral filtering into the domain of modulation frequency. The general concept of modulation filters is based on data from experiments on modulation masking that indicate "tuning" for modulation frequency (see Ewert, Verhey, and Dau 2002). It is not surprising, therefore, that the general MF model can account for these data. The model, however, has considerable predictive power and can account for envelope effects, notably the differences seen in detection of sinusoidal amplitude modulation (SAM) for different types of carriers (Dau, Verhey, and Kohlrausch 1999; Kohlrausch, Fassel, and Dau 2000).

Unfortunately, direct experimental evaluation of the critical aspect of the MF models, modulation filters, has proved elusive and inconclusive. Several experiments that attempt such evaluation have used complex modulators or maskers (Lorenzi, Berthommier, and Demany 1999; Moore, Sek, and Glasberg 1999; Sheft and Yost 2001). There recently have been examinations of phase effects that are not easily explained by the MF models, at least in their current forms (Strickland and Viemeister 1996; Moore and Sek 2000; Sek and Moore 2003). Part of the problem, one common to most models of envelope processing, is uncertainty as to that aspect of the "internal" envelope (the decision variable) that is important in detection and, more generally, in envelope-based perception.

This paper does not attempt a direct evaluation of modulation filters or, more generally, of approaches based on the modulation spectrum. Rather, it is an initial examination of detection in stimulus situations for which the local temporal structure of the envelope is manipulated systematically. The general question is: To what extent does local temporal structure determine envelope perception? The emphasis is on the time-domain representation of the envelope rather than on the modulation frequency-domain representation.

2 Methods

In all conditions, a three-interval, three-alternative forced-choice task was used. The stimulus in each interval, $x(t)$, was defined as follows:

$$x(t) = \{[1 + m(t)] + [1 + m_s \cos(\omega_s t)]\}\, n(t) \qquad (1)$$

where $m(t)$ is the masker modulator and m_s and ω_s are the modulation index and modulation frequency of the signal modulator. The carrier, $n(t)$, was a broadband Gaussian noise low-pass filtered at 10 kHz. The noise carrier was generated randomly from trial to trial but the same sample of noise was used in all three intervals of each trial. The starting phase of the signal modulator was always zero. Each interval was 1000 ms in duration with 500 ms of silence between the intervals. Correct-answer feedback was provided after each response.

The modulation index of the signal modulator (m_s) in the signal interval was varied adaptively in a 3-down-1-up procedure that estimated the 79.4 percent correct point on the psychometric function (Levitt 1971). In the non-signal intervals of each trial, m_s was set to zero. The initial step size for m_s was set to 2 dB (in units of $20 \log m_s$) and was reduced to 1 dB after four reversals. A block of trials was terminated after a total of 12 reversals, and threshold was estimated as the mean modulation depth of the final eight reversals. Four blocks of trials were run in each condition and the four threshold estimates were averaged to obtain the final threshold estimate for that condition. A block was terminated and no threshold was estimated for that block if the addition of the signal would have produced overmodulation.

Modulation-detection thresholds were measured separately for the signal modulator in the presence of several different masker modulators. In the "No Masker" condition, $m(t)$ was set to zero. In the "Sine Masker" condition: $m(t) = \cos(\omega_m t - \pi/2)$. In the "Pulsed Masker" conditions the masker modulator was a periodic rectangular pulse wave with duty factor (ratio of pulse width to signal period) ranging from 0.15 to 0.50. The fundamental frequency was equal to that of the signal (4, 8, 24, 48, and 64 Hz). The pulsed maskers were generated with the restriction that the amplitude and phase of the fundamental were always equal to those of the sine masker. Because of the manner in which $m(t)$ was generated, the DC of the pulse train with a duty factor of 0.25 was -2.8 dB and the DC for a duty factor of 0.15 was -8.1 dB relative to those of the other maskers. In a subset of conditions the overall DC of the masker was adjusted to be constant across duty factors. In all conditions, the waveform for the signal interval was scaled such that its rms was equal to that of the non-signal intervals.

In the Pulsed Masker conditions, modulation detection thresholds were measured when either the complete signal modulator was present or the signal modulator was present only in the troughs of the masker modulator. Examples of the waveforms used in these conditions are shown in Fig. 1.

Stimuli were generated and presented via Matlab (Math Works) on a PC equipped with a high-quality, 24-bit sound card (Echo Audio Gina) at a sampling rate of 44.1 kHz. Stimuli were presented monaurally via Sony MDR-V6 stereo headphones to listeners seated in a sound-attenuating chamber. The listeners were three undergraduate students (one male and two female) from the University of Minnesota who were paid to participate in the study. All listeners had normal hearing according to lab audiometric standards. Listeners were allowed to practice in a variety of modulation masking conditions until their thresholds stabilized (about 2 weeks).

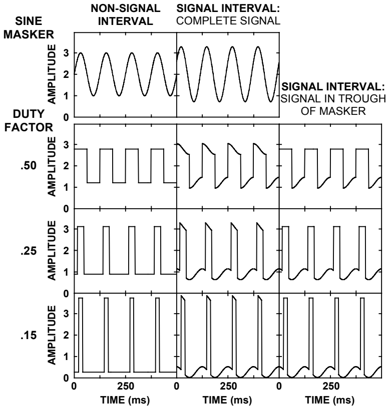

Fig. 1. Envelopes in the 8-Hz conditions. The amplitude of the fundamental but not the DC is constant across the conditions shown.

3 Results

Figure 2 shows data for two of the five modulation frequencies examined. The important trends are similar across modulation frequency and because the thresholds are similar across listeners, averaged thresholds are shown. Two reference conditions are shown by the symbols on the left-hand side of each panel. Also shown are thresholds for a condition in which the DC component of the masker for duty factors of 0.15 and 0.25 (inverted triangles) was adjusted to be equal to that in the Sine Masker and 0.50-duty factor conditions. The thresholds in the No Masker and Sine Masker conditions are in good agreement with those reported in the literature (e.g. Viemeister 1979; Wakefield and Viemeister 1990; Ozimek and Sek 1988).

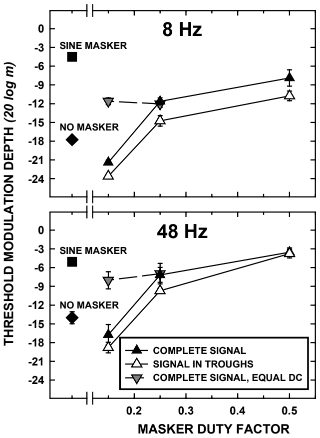

Fig. 2. Data from the 8- and 48-Hz conditions. The masker modulator, when present, was pulsed (triangles) or sinusoidal (squares). The fundamentals of all maskers were equal in amplitude.

One important aspect of the data shown in Fig. 2 is the general decrease in modulation threshold with decreasing duty factor. Similar decreases were observed for the other modulation frequencies. Another noteworthy result is that the thresholds for “signal in troughs” (open symbols) follow the same general trend as those for the complete signal but usually are slightly lower than those for the complete signal. A possible explanation for the lower thresholds is that truncating the signal introduces modulation energy in the envelope spectrum at harmonics of the signal frequency and that somehow this “splatter” can be used to aid detection of the signal. At this point we view these slight differences (the largest between the black and white symbols shown in Fig. 2 is 3.1 dB) as a secondary concern. The more important point is that the thresholds for the troughs-only condition are comparable to those for the complete signal.

The data for the “equal DC” condition make it clear that the large decrease in thresholds for a duty factor of 0.15 in the other conditions is due in part to the fact that the DC was allowed to vary across duty factor, although the change in DC had no effect for a duty factor of 0.25. Despite the obvious effects of varying the DC, the important comparison in Fig. 2 is between the thresholds in the Pulsed Masker conditions (triangles) and those for the Sine Masker (squares). For these conditions, addition of the signal causes an increment in the fundamental of the masker, which is equal across all of these conditions. The substantial changes in threshold across these conditions suggest that such a frequency-domain approach is not a useful way to interpret the data.

An alternative way to interpret the data shown in Fig. 2 is to consider that signal detectability in the Pulsed Masker conditions may be determined by the *local* modulation depth in the troughs of the masker. This notion is illustrated in Fig. 1 where it is clear that because the DC in the trough decreases with decreasing duty factor, the modulation depth of the portion of the signal in the trough is increased. The modulation depth in the trough of the masker can be computed as

$$m_t = (A_{max} - A_{\mathrm{min}})/(A_{max} + A_{min}) \qquad (2)$$

where A_{max} and A_{min} are the maximum and minimum amplitudes of the signal plus masker in the trough of the masker. For the fixed amplitude signal used in Fig. 1, 20 log m_t is -13.7, -11.0, and -0.6 dB for duty cycles of 0.50, 0.25, and 0.15 respectively.

This notion is pursued in Fig. 3. Figure 3 shows the pulsed-masker data from Fig. 2 expressed as modulation depths in the troughs, 20 log m_t, as defined above. The important point illustrated in Fig. 3 is that local modulation depth, specifically m_t, appears to eliminate the strong dependence on duty factor shown in Fig. 2. This clearly is true for the unequal DC conditions (solid symbols).

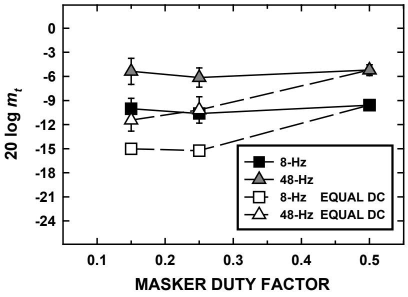

Fig. 3. A portion of the data from Fig. 2 expressed as the modulation depth in the troughs of the pulsed masker.

There are other aspects of the data that are notable. For all duty factors, thresholds are higher for 48 Hz than for 8 Hz and the increases in thresholds are in agreement with those for simple detection (the No Masker condition). However, the thresholds for the Sine Masker shown in Fig. 2 are the same for 8 and 48 Hz. This condition is modulation depth discrimination and, because it is frequency independent, appears to be a different process from that involved in the Pulsed Masker conditions. An explanation is that the Pulsed Masker conditions involve detection of the signal in the troughs of the masker rather than discrimination of a change in the amplitude of the fundamental of the envelope spectrum. This explanation is consistent with the notion that local modulation depth determines signal threshold.

It is not clear why the thresholds in Fig. 3 for the equal DC condition are lower than those for the other conditions. In the equal DC condition the overall level is higher but this should not affect the detectability of modulation (Viemeister 1979). Despite these uncertainties, Fig. 3 illustrates the main point of this paper, namely that local modulation depth may determine detectability with complex maskers.

In an experiment related to the present experiment a wider range of masker duty factors was examined using maskers for which the amplitudes of both the DC and fundamental were held constant across duty factor. To avoid overmodulation, a potential problem using the adaptive procedure, performance (percent correct) was measured for fixed signal depths (m_s). An important result is that there were large differences in performance for the complementary pairs of duty factors examined, specifically (0.15, 0.85) and (0.25, 0.75). For such pairs the amplitude spectra of the envelopes are identical. Thus, this result further underscores the importance of phase and temporal structure.

4 Discussion

The general question addressed in this paper is the nature of the decision variable used in envelope processing. This is a fundamental question and is one that has received attention over the past few decades (for a review see Strickland and Viemeister 1996). It appears that the decision variables that have been proposed can, at best, describe the data only for a limited subset of conditions, namely detection of SAM. None appears to be able to account for the present data.

Our qualitative, black-box account is essentially an extension of the notion of “multiple looks” (Viemeister and Wakefield 1991) in which information from brief samples or looks at some central representation is selectively and intelligently combined to make decisions about the stimulus. In the present situation this central representation is the result of relatively early stages of envelope processing. An example of such processing is lowpass filtering preceded by a nonlinearity such as half-wave rectification. Another is modulation filtering possibly including some type of recombination of the filter outputs. At this point in model development the details of this processing seem less important than in describing how the output is processed to arrive at a decision variable.

A decision variable that appears promising is one based on local modulation depth at the signal frequency. This is closely related to m_t in Eq 2 which provided a good, but incomplete, account of the general trends of the data from this experiment. Thus, after initial envelope processing the running modulation depth of the waveform at the signal frequency is computed and those temporal regions in which the local depth is largest are given the most weight. To be more concrete, one such statistic is based on

$$D(t) = [e'(t) - e'(t + T/2)]/[e'(t) + e'(t + T/2)] \qquad (3)$$

where $e'(t)$ is the “internal envelope” after initial processing and T is the period of the signal. A simple decision variable would be the maximum value of $D(t)$. There are, of course, other ways to compute local modulation depth including local rms and running cross correlation. The key element, however, appears to be that the statistic is computed over some relatively short time window and is allowed to run over the course of the modulation. This permits extraction of temporal structure and distinguishes this type of decision statistic from those that have been previously examined.

The duration of the window probably should be related to the period of the signal rather than fixed. This conjecture is based on our observations in experiments similar to those reported here but using maskers and signals that differ in their periods. A window related to the signal captures the notion that listeners have some knowledge of the signal (and masker) periodicity and offers the appealing possibility that the frequency selectivity shown in envelope processing, namely modulation rate discrimination and tuning in modulation masking, can be explained on a purely temporal basis.

5 Summary and conclusion

The results from this experiment strongly suggest a potentially important role for local temporal structure in envelope processing. The data indicate that the detectability of SAM in the presence of pulsatile AM maskers is approximately determined by the local modulation depth in the trough of the masker (Fig. 3). The data may not be inconsistent with spectrally-based models, however, especially if the decision variable somehow incorporates phase information. In our opinion, a more promising approach emphasizes a time-domain analysis. We have outlined the rudiments of a multiple-looks type of model based on local modulation depth that captures, at least qualitatively, the general characteristics of the data. Such a model may provide the basis for a purely temporally-based account of this fundamentally important aspect of auditory perception.

6 Acknowledgments

This work was supported by Research Grant No. R01 DC 00683 from the National Institute on Deafness and Communication Disorders, National Institutes of Health.

References

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.

Dau, T., Kollmeier, B. and Kohlrausch, A. (1997) Modeling auditory processing of amplitude modulation. I. Detection and masking with narrow-band carriers. J. Acoust. Soc. Am. 102, 2892-2905.

Ewert, S.D. and Dau, T. (2000) Characterizing frequency selectivity for envelope fluctuations. J. Acoust. Soc. Am. 108, 1181-1196.

Ewert, S.D., Verhey, J.L. and Dau, T. (2002) Spectro-temporal processing in the envelope-frequency domain. J. Acoust. Soc. Am. 112, 2921-2931.

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.

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

Lorenzi, C. Berthommier, F. and Demany, L. (1999) Discrimination of amplitude-modulation phase spectrum. J. Acoust. Soc. Am. 105, 2987-2990.

Moore, B.C.J., Sek, A. and Glasberg, B.R. (1999) Modulation masking produced by beating modulators. J. Acoust. Soc. Am. 106, 908-918.

Moore, B.C.J. and Sek, A. (2000) Effects of relative phase and frequency spacing on the detection of three-component amplitude modulation. J. Acoust. Soc. Am. 108, 2337-2344.

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

Sek, A. and Moore, B.C.J. (2003) Testing the concept of a modulation filter bank: The audibility of component modulation and detection of phase change in three-component modulators. J. Acoust. Soc. Am. (in press)

Sheft, S. and Yost, W.A. (2001) AM detection with interrupted modulation. In: D.J. Breebart, A.J.M. Houtsma, A. Kohlrausch, V.F. Prijs and R. Schoonhoven (Eds.), *Physiological and Psychophysical Bases of Auditory Function.* Shaker, Maastricht, pp. 290-297.

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.

Viemeister, N.F. and Wakefield, G.H. (1991) Temporal integration and multiple looks. J. Acoust. Soc. Am. 90, 858-865.

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.