A model of auditory image flow II: Detection of amplitude and frequency modulation.

N.P. McAngus Todd and Duncan R. Brown

Dept. of Psychology, Univ. of Manchester, Manchester, M13 9PL, UK.

Abstract: In order to evaluate the cortical model proposed in Todd (1) we apply it to the phenomenon of modulation detection. A series of simulations were carried out and compared with psychophysical data (2). It was assumed that a discrimination task involves the comparison of the power spectrum of two image sequences, such as may occur in a typical 2A-2FC procedure. Overall the model shows the correct behaviour for AM and FM detection, but also some departures from human data. Possible explanations for these departures are discussed.

INTRODUCTION

A consistent finding in the extensive studies of modulation detection (2) is that below a certain critical modulation frequency (CMF), which is dependent on the carrier frequency, the threshold rises for AM and falls for FM with increasing modulation frequency whilst beyond the CMF AM and FM thresholds fall steeply together. It is of interest to see whether the cortical mechanism proposed in Todd (1) may give an account of AM and FM detection particularly since oriented spatio-temporal receptive fields are most sensitive to changing frequency.

METHODOLOGY

In order to test the model we used the following procedure for generating model thresholds. (a) Assume that a discrimination task involves the comparison of the power spectrum of two image sequences. (b) Select an observation strategy for comparison of power spectra. (c) Select a comparison metric $d$. (d) Linearly regress the discriminability metric as a function of the square of the modulation index. (e) Compute the slope of the regression (with no intercept) for both AM and FM. (f) Select an arbitrary threshold value and compute model thresholds.

In order to make a comparison with data of (2) we chose the following values for the stimuli. All tones had a duration of 1000 ms, carrier frequency 1 kHz and onset/offset 50 ms. Modulation frequencies chosen were 4 Hz, 16 Hz, 64 Hz and 256 Hz. Modulation depths chosen were 0.05, 0.1 and 0.2. Given the computational complexity of the model it was necessary to choose sparse spatio-temporal sampling. The following parameters were chosen. $32 \text{ Hz} \leq f \leq 8000 \text{ Hz}$, $n = 22$ or $33$; $32 \text{ Hz} \leq \phi \leq 4000 \text{ Hz}$, $n = 18$ or $28$; $2 \text{ Hz} \leq \omega_i \leq 32 \text{ Hz}$, $n = 10$; $0.05 \text{ erb}^{-1} \leq (\omega_f, \omega_f) \leq 0.2 \text{ erb}^{-1}$, $n = 3$; $\theta = 0, \pi / 4, \pi / 2, 3\pi / 4, n = 4$. We chose a "1-look" observation in which the power spectrum is frozen at the end of the stimulus and the metric $d = 10^5 (1 - R)$ (one minus product moment correlation) which gave a fit of $r^2 > 0.95$ for both AM and FM when regressed as a function of the square of the modulation index. The threshold value was arbitrarily set to $d = 1$.

RESULTS

Overall the model seems to show the correct general behaviour (see Figure 1). (a) The CMF has been correctly identified, (b) below the CMF the threshold rises for AM and falls for FM with increasing modulation frequency and (c) above the CMF AM and FM thresholds fall together. However, there are a number of important departures from the psychophysical data. (a) The ratio of AM to FM detectability is too small at low frequencies and (b) there is a systematic bias towards low modulation frequencies relative to high modulation frequencies, i.e. the model thresholds for both AM and FM at low modulation frequencies are underestimated relative to the threshold at CMF and overestimated for high frequencies.

DISCUSSION

Although the above departures from the psychophysical data seem quite serious there are several possible explanations. (a) Loudness coding. Below the CMF, the ratio of AM to FM is dependent on loudness (e.g. according to Zwicker's data (3) at 30 phons and 4 Hz the log ratio is approximately 0.9 whereas at 80 phons it is 1.2). Also relative detectability of modulation frequencies above the CMF to modulation frequencies below the CMF is highly affected by loudness. Although the model was set up so that RMS amplitude = 1 corresponds to 70 dB with the Meddis hair-cell parameters, it seems likely that a single hair-cell per channel is an inadequate coding
of intensity (4). (b) Parameter values. The slope of the model psychometric functions are sensitive to the exact parameter settings. These include the sampling rate of the image (currently 1 per erb) and the sampling rate of the image power spectrum. At a density in the temporal RFs of 1 per octave this represents a fairly severe undersampling of the temporal power spectrum. (c) Observation strategy. This could explain the departure of the model from the human data at low frequencies since the “1-look” at the end favours temporal RFs with longer time-constants. Other strategies could be tried such as a “peaks” mode where the maximum peak output of each cortical filter is stored. An alternative strategy might be to use the temporal low-pass filters to trigger the “looks” in order to form an event-based strategy (5).

In terms of comparison with other models, the model here has obvious similarities with that of Dau (6), in that modulation detection is based on monitoring the modulation spectrum across all cochlear channels. The essential difference with Dau’s model, apart from the parameter values and ranges and that it has not been applied to FM, is that the templates used here are given by the spatio-temporal power spectrum of the sequence of modulation images, rather than the time sequence of images themselves. It is interesting to note that in order to account for AM detection of pure tones, Dau found it necessary to introduce a another layer of low-pass filters after the modulation filters. This feature is also present in the model proposed here.

As suggested in (1) the fact that cortical RFs span a number of channels implies that any decision procedure based on the whole power spectrum will be subject to interference effects from channels which are remote. This has obvious implications for Modulation Detection Interference (MDI).

**CONCLUSION**

In this paper the idea that auditory image motion is coded in terms of its spatio-temporal power spectrum was tested by applying it to the phenomenon of AM and FM detection. Overall the model showed the correct trends but also some clear departures from experimental data. These departures may be accounted for by a number of factors, including intensity coding, sparseness of the sampling rates and the use of an observation strategy which biases towards low frequencies. Clearly, given these variety of factors, further experiments with the model are required.

**REFERENCES**