Comments

Comment by Kollmeier:

Your stimulus conditions with low modulation rates seem to explore the lower limit of where an envelope power spectrum model is valid. Similar to the audio frequency domain, special care has to be taken in any modulation filterbank model for the lowest frequency band. Dau *et al.* therefore used a lowpass filter for this band and hence preserve the phase of the modulation.

Also, an increase in level becomes asymmetrical to a decrease in level if the modulation depth is increased – this clearly limits the validity of your null hypothesis, that an identical modulation spectrum and hence identical thresholds should result for the conditions with the respective inverted duty cycle (such as, e.g., the pair .25 and .75). The extreme extrapolation of your hypothesis would even assume that a periodic delta-pulse modulator should be perceptually identical to a constant envelope interrupted with small gaps – this hypothesis does not make much sense. Hence, your data may not be as unexpected as you state it.

Reply:

As indicated in the chapter and in the talk, the issue was *not* spectral vs. temporal models of envelope processing. The issue was whether temporally-based models, such a those based on local temporal structure, can provide insights into complex envelope processing. The data we presented suggest that such models are perhaps more conducive than spectral models for understanding the masking produced by pulse-train maskers. This, however, does not mean that spectral models, such as those incorporating a modulation filterbank, can not account for the data. As we indicated, spectrally-based models can, in principle, account for the data if phase information is preserved.

The second part of the comment refers to the possible role of compression. Clearly, in the experiment in which the maskers had complementary duty factors (e.g., 0.15 and 0.85) and therefore the same amplitude spectra, compression of the input will affect these spectra and make them unequal. This raises the interesting possibility that with compression a spectral model could account for the data even if phase information were not preserved. This should be examined. However, in the first experiment it was observed that for a given duty factor the thresholds were similar for the complete signal and the trough-only signal (Fig. 2). This suggests that the detectability was determined primarily by the local modulation depth in the troughs. As shown in Fig. 3 the detection thresholds, expressed as this local modulation depth (via eq. 2), are approximately constant. Compression would change the “internal” modulation depth but since the thresholds are nearly constant it appears that compression is not an important factor in this experiment. That is, regardless of the compression exponent, as typically defined, the modulation depth after compression will still be constant over the various duty factors assuming, of course, that the exponent does not depend on duty factor.

Comment by Dau:

A modulation filter-bank analysis that discards temporal (phase) information can certainly not account for the data of the present study. However, the modulation filter-bank model in *Dau et al.* (1997) preserves phase information at low modulation frequencies and analyses the temporal internal representation of the stimuli using a matched filtering process based on cross correlation. Such a combined spectro-temporal analysis would be worth testing explicitly. The analysis might be successful, at least at the low-frequency (signal) modulation rates.

More generally, an important key aspect is how absolute (monaural) phase is processed in the system. If phase turns out to be fully preserved throughout the *entire* modulation frequency range and no "information loss" is introduced at some stage of processing, then a modulation filter-bank analysis would, of course, be completely equivalent to a purely temporal analysis. The data of the present study suggest that phase information might be available at higher rates (48 Hz) but this needs further examination. Again, a quantitative test and comparison with model predictions need to be done.

Finally, I think that a purely temporal process (as suggested in the paper) can not account for modulation masking as found in the studies by Bacon and Grantham (1989), Houtgast (1989) and Ewert and Dau (2000). Such a process, even if combined with a "sub-optimal" detector will most likely not predict sufficiently broad tuning. What would be your strategy to account for the tuning data?

Reply:

We are in agreement about the potential importance of preserving phase information in spectral models. See the response to Kollmeier. I certainly would like to see a proof that a purely temporal process can not account for modulation tuning and, more generally for modulation frequency selectivity. It would save me considerable time…