Comments

Comment by Bernstein:

To what extent might any short-term spectral consequences (e.g., splatter) serve as a cue for listeners in the physical phase-reversal conditions? Such cues would, of course, not be present in the continuity illusion conditions. Could that explain listeners' poor sensitivity to phase-reversals in the latter conditions?

Reply:

The long-term spectral cues would have been at least as large for a modulation rate of 20 Hz (where subjects are close to chance) as at 2.5 and 5 Hz, where they perform very well. In addition, we have performed the experiment with a carrier frequency of 6 kHz and the same relative FM depth. At 6 kHz the ERB of an auditory filter (according to Glasberg and Moore) is 11% of the center frequency, which is narrower in relative terms than the value of 13% at 1 kHz. Despite this, performance was considerably worse at the higher carrier.

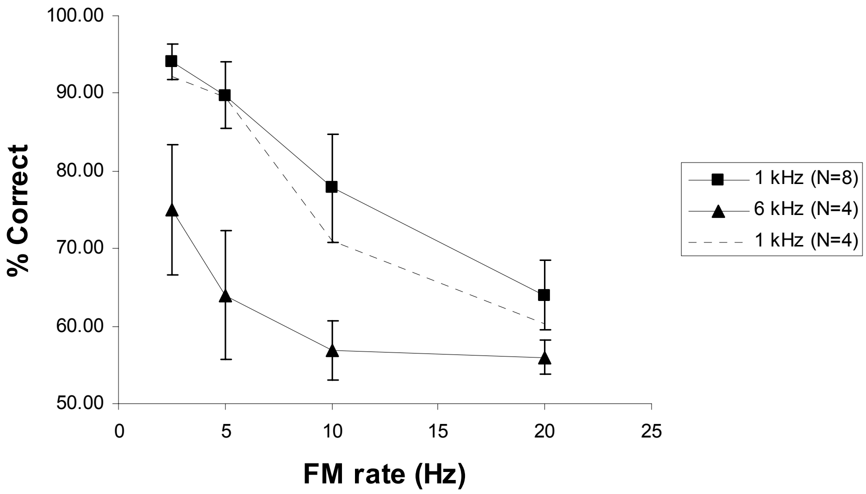

Fig. A1. Detection of a phase discontinuity in the FM as a function of rate and carrier frequency.

Comment by Furukawa:

It is possible that the auditory system filled the noise portion with illusory modulations with slightly different rates for the normal- and the reversed-phase conditions. Such modulations might be one cycle of 5-Hz FM for the normal-phase condition, and a 1.5 cycle of 7.5-Hz FM for the reversed-phase condition. The subject was not able to tell the difference between the conditions, maybe because the S/N for the illusory FM portion was too low to detect the difference between the 5-Hz and the 7.5-Hz FMs, like in a simultaneous masking experiment. To explore that possibility, you might test whether the listener could distinguish between 5-Hz and 7.5-Hz of *actual* modulations, in the presence of a noise masker. Also, you might try a rate-matching experiment for the normal and the reversed-phase condition. If the above notion were true, you would expect the best-matched rate to be around 7.5 Hz.

Reply:

It would be hard to equate the S/N in the test you propose to that in our "illusory" condition, in which the signal was absent during the noise. But, as a start, we *could* determine, in the absence of noise, whether listeners could tell whether the central 200-ms of a 5-Hz FM tone was replaced with 1.5 cycles of a 7.5-Hz modulator.

Comment by Marvit:

Based my understanding from the methods described in the paper and (especially) the example stimuli, I am surprised that the subjects could not use one additional cue in Experiments 3 and following to substantially improve their performance—the starting and ending phase of the modulation in each interval. That is, the first and final 50 msec of the 1 s standard interval always had an FM going in the same direction, while the 1 s signal had the first and last 50 ms going in opposite directions. The exception, of course, was in condition 3d where duration of the segments varied ; however, only the 100 ms elongation of one segment in an interval would render that cue inoperable and then only if in conjunction with a 0 or 200 ms elongation of the other segment. Informally, Didier Depireux and I have observed that we can determine the velocity direction of a ripple by attending to the initial part of the sound, even when we could not otherwise identify the direction from the main part. In the talk, I noticed these cues when example stimuli were played. I wonder, therefore, whether your subjects had feedback during the experiments that might have given them opportunity to develop this strategy or whether the results might have been rather different if they had instructions to attend to the relative direction of "pitch change" at the extremes of the intervals. Of course, this cue would be rendered moot had the stimuli varied randomly by 50 ms increments—preserving the zero crossings at the edges of the segments, but thereby randomizing the direction of the edges of the intervals.

Reply:

In experiment 3, the duration randomisation was applied to all conditions (as stated in the text). The increase was applied independently to the first and second burst *within each interval*, and could be applied either to the start or end of that burst. As you point out, this would have removed any *consistent* cue relating to the starts or ends of the standard and signal stimuli. In experiment 4 this duration randomisation was removed. However, the randomisation of the starting FM phase from interval to interval was still performed. So, subjects would still have to have compared the "glide" at the start and end of each stimulus. Usually our subjects are pretty crafty, but the chance performance in the noise condition of experiment 4 suggests that they were unable to use this trick. This may have been due to the long duration of the sounds. Another possible interpretation is that they can perform phase comparisons between sounds (as in the gap condition) but not within sounds (as in the steady condition). In fact, their ability to do the task in the gap condition suggests that they could encode the starting phase of one sound and compare it to that of another.

Comment by Demany and Pressnitzer

You argue that your data "point to a lack of explicit encoding of FM phase by the auditory system", for periodic FM with a rate of a few Hertz. This suggestion is at odds with observations reported by Demany and McAnally (1994) concerning the perception of periodic FM functions with a complex waveform. They found that when the FM waveform consists of the sum of a sinusoid and its third harmonic, a repeating melodic motif is heard that typically consists of the successive peaks (local maxima) of the FM waveform. The melodic pattern heard depends on the relative phase of the FM components. This perceptual phenomenon seems to imply an explicit encoding of FM phase.

Demany, L., and McAnally, K.I. (1994) The perception of frequency peaks and troughs in wide frequency modulations. J. Acoust. Soc. Am. 96, 706-715.

Reply:

I think that what your comment highlights is that I (and others) should be more explicit when talking about the "explicit coding of FM". What I do not mean to imply is that FM phase can never affect perception. For example, one can easily tell the difference between a phase reversal that occurs at a positive zero crossing (e.g. left hand part of Fig. 1h) and one that occurs at a negative zero crossing. In terms of the "smoothed integrator" explanation, one produces a dip in the averaged instantaneous frequency and the other produces a peak. What I am arguing, however, is that at some level of representation the existence (and probably the depth and rate) of FM are preserved, but that the phase information is not. A (probably over-simplistic) analogy is with fine-structure phase: when you hear a 1 kHz tone you encode its frequency, loudness, and duration, but are not aware of its starting phase. However, its phase can still affect perception when that tone is added to (say) two other tones of nearby frequencies.

Having said this, what is striking about the Demany and McAnally finding is that it shows an asymmetry not accounted for by the sliding integrator that we show in Fig. 3. Incidentally, the model also does not account for the fact that portions of the stimulus where instantaneous frequency is changing quickly have a reduced impact on the pitch of the sound (Gockel *et al.* 2001). This latter finding is roughly consistent with the first derivative of the instantaneous frequency *vs* time function being calculated within the integration window, and an appropriate weight applied. To account for the Demany and McAnally finding, the model would (arbitrarily) have to do something like weight portions according to the second derivative of that function. So there are clearly phenomena that this particular model, in its simplest form, cannot account for, but I do not think that this is inconsistent with explicit FM phase information being discarded at some fairly early stage of auditory processing.

Gockel, H., Moore, B. C. J. and Carlyon, R. P. (2001) Influence of rate of change of frequency on the overall pitch of frequency modulated tones, J. Acoust. Soc. Am. 109, 701-712.