JARO 13: 159–171 (2012)
DOI: 10.1007/s10162-011-0305-0

JARO
Journal of the Association for Research in Otolaryngology

Invited Review

Across-Channel Timing Differences as a Potential Code for the Frequency of Pure Tones

Robert P. Carlyon[1], Christopher J. Long[1,2,3], and Christophe Micheyl[4]

[1]*MRC Cognition & Brain Sciences Unit, 15 Chaucer Rd., Cambridge, CB2 7EF, UK*

[2]*Cochlear Ltd., 13059 E. Peakview Avenue, Centennial, CO 80111, USA*

[3]*Department of Speech, Language, and Hearing Sciences, University of Colorado, Boulder, CO, USA*

[4]*Department of Psychology, University of Minnesota, 75 East River Rd., Minneapolis, MN 55455, USA*

Received: 29 April 2011; Accepted: 7 November 2011; Online publication: 8 December 2011

ABSTRACT

When a pure tone or low-numbered harmonic is presented to a listener, the resulting travelling wave in the cochlea slows down at the portion of the basilar membrane (BM) tuned to the input frequency due to the filtering properties of the BM. This slowing is reflected in the phase of the response of neurons across the auditory nerve (AN) array. It has been suggested that the auditory system exploits these across-channel timing differences to encode the pitch of both pure tones and resolved harmonics in complex tones. Here, we report a quantitative analysis of previously published data on the response of guinea pig AN fibres, of a range of characteristic frequencies, to pure tones of different frequencies and levels. We conclude that although the use of across-channel timing cues provides an *a priori* attractive and plausible means of encoding pitch, many of the most obvious metrics for using that cue produce pitch estimates that are strongly influenced by the overall level and therefore are unlikely to provide a straightforward means for encoding the pitch of pure tones.

Keywords: auditory nerve, pitch, frequency, pure tone, phase transitions

Correspondence to: Robert P. Carlyon · MRC Cognition & Brain Sciences Unit · 15 Chaucer Rd., Cambridge, CB2 7EF, UK. email: bob.carlyon@mrc-cbu.cam.ac.uk

INTRODUCTION

Traditionally, theories of how the auditory system encodes pitch have focused on either the variation in firing rate across auditory nerve (AN) fibres having different characteristic frequencies (CFs; Zwicker 1970), or on the temporal pattern of firing ("phase locking") of neurons responding to the sound (Schouten 1940; Wever 1949; Siebert 1970; Cariani and Delgutte 1996). Proponents of each class of theory have pointed to potential weaknesses of the other. Rate–place representations may suffer from the facts that the firing rates of the majority of AN fibres are saturated at high stimulation levels, and that the locus of maximum firing rate may change with level (e.g. Kim et al. 1980; Chatterjee and Zwislocki 1997; Versteegh et al. 2011). As a result, rate–place profiles may become degraded and/or shifted at high levels, even though psychophysical experiments indicate good performance and strong pitch at those levels (e.g. Wier et al. 1977). Phase-locking cues are immune to these distortions, but the method for processing this temporal information, without requiring the long neural delay lines assumed by some influential "autocorrelation"-based models (Licklider 1951; Meddis and Hewitt 1991; Meddis and O'Mard 1997), remains a matter of debate (de Cheveigné and Pressnitzer 2006; Meddis and O'Mard 2006; Schnupp et al. 2010). In addition, although most temporal theorists assume that the frequency of a pure tone is encoded primarily by phase locking for frequencies up to at least 2,000 Hz (the frequency at which phase locking usually starts to drop in the mammalian species studied to

date), there is evidence that neurons in the auditory cortex phase lock only up to much lower frequencies (e.g. Eggermont 1991; Wang et al. 2008), leading to the suggestion that phase locking must be recoded into some other representation at, or below, the auditory cortex.

The pros and cons of the above arguments have been debated extensively elsewhere (Moore 2003; de Cheveigné 2005; Plack 2005). Here, we simply note that the various potential weaknesses of the two traditional approaches have led to considerable interest in an alternative class of explanation, whereby the auditory system performs an instantaneous comparison of the temporal pattern of firing in auditory neurons having different CFs (Loeb et al. 1983; Shamma 1985; Carney 1994; Heinz et al. 2001; Carney et al. 2002; Colburn et al. 2003; Oxenham et al. 2004; Loeb 2005; Moore and Carlyon 2005; Cedolin and Delgutte 2010). This idea stems from the fact that the phase of AN responses varies across CF in a manner that varies with the input frequency. It can be illustrated with reference to Figure 1, reprinted from an article by Kim et al. (1980), who recorded the responses of cat AN fibres having a wide range of CFs to a two-tone complex having component frequencies, f_1=2,100 Hz and f_2=2,700 Hz. The figure shows two sets of curves, each of which represents the phase of the AN response to one of the components. In each case, the phase is roughly constant over a range of high-frequency CFs, and there is a steep phase transition (PT) at CFs near the frequency of each tonal component (shown by the small arrows); similar results have been more recently obtained by van der Heijden and Joris (2006) using a different technique. To a first approximation, this pattern of results can be explained by the facts that a filter introduces a delay as the frequency of the input passes through its resonance, that the size of this delay is greater for narrow than for broad filters, and that the bandwidths of basilar membrane (BM) filters (as estimated from, e.g. tuning curves), when expressed in hertz, are narrower at the apex than at the base. A consequence of the variation in phase with CF is that neurons with CFs much higher than the input frequency fire at approximately the same time as each other, whereas those close to the input frequency fire out of phase. This is illustrated in Figure 2 which shows schematic "neurograms" to 1,000- and 1,200-Hz pure tones (these neurograms were not derived from the cat data obtained by Kim et al. and shown in Figure 1, but were based on fits to data obtained from another species as described in "Modelling approach").

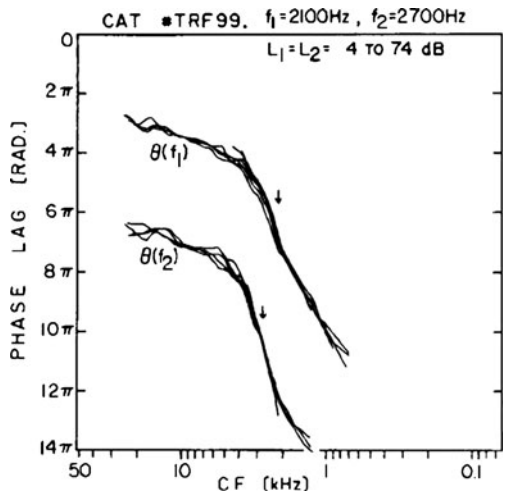

FIG. 1. From Kim et al. (1980), with permission. Phase of AN response to a 2,100+2,700-Hz complex at levels ranging from 4 to 74 dB SPL. There are two sets of curves, corresponding to each component; the curves within each set are for different sound levels and, at the scale shown here, appear to overlap. Note that, unlike all the other Figs. in this article, CF decreases from high to low along the abscissa.

A number of authors have described neural mechanisms by which the auditory system might exploit these across-channel phase differences, both for the perception of pitch, the detection of tones in noise, the coding of sound level, and for enhancing spectral representations of complex sounds (Loeb et al. 1983; Shamma 1985; Carney 1994; Heinz et al. 2001; Carney et al. 2002; Colburn et al. 2003; Loeb 2005; Cedolin and Delgutte 2010). Examples include the subtraction of the outputs of nearby channels (Shamma 1985; Cedolin and Delgutte 2010), the processing of channels whose outputs are de-correlated due to the presence of a narrowband sound (Carney et al. 2002), and the detection of coincidences between the outputs of channels having different CFs (Loeb et al. 1983; Loeb 2005). In each case, the proposed mechanisms allow for phase locking to be recoded, at an early stage of auditory processing, into some other (e.g. rate–place) representation and obviate the need for processes, such as autocorrelation, that may require long central delay lines for computation. At the same time, these across-channel timing cues are largely immune to the effects of firing rate saturation and could therefore combine the strengths of temporal and rate–place models whilst avoiding some of the drawbacks of each (Cedolin and Delgutte 2010).

Another appealing feature, noted by Shamma (1985) and by Cedolin and Delgutte (2010), is that PT cues may be level-invariant. This is illustrated in Figure 1 in which the sets of curves for each tone frequency contain individual curves for input levels of between 4 and 74 dB SPL, which substantially overlap. However, as discussed above, PTs arise from the filtering properties of the BM; these are known to be nonlinear, and so one might expect some level dependence in them too. Furthermore, it has been

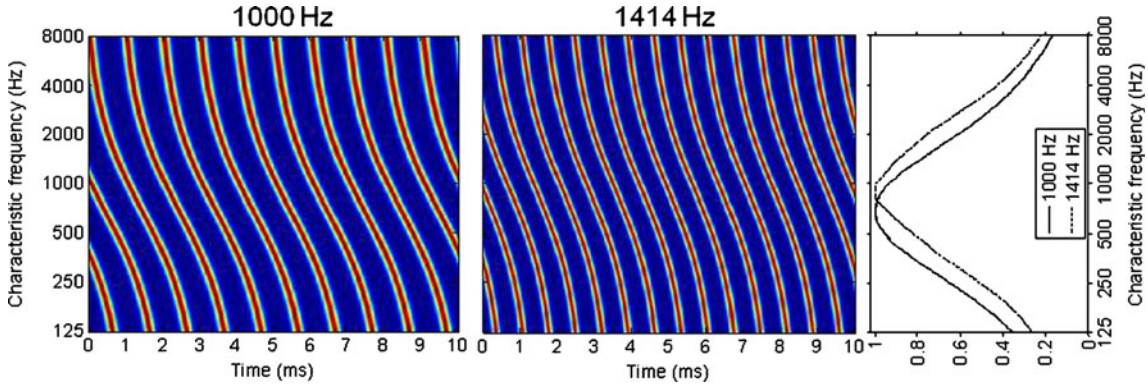

FIG. 2. Schematic "neurograms" where the firing time of neurons in each CF channel are estimated from the arc-tangent fits to the CF-vs. phase curves, described in the text, for 1,000-Hz (**A**) and 1,200-Hz (**B**) pure tones. *Bright colours* (*red, yellow*) indicate a high instantaneous firing rate; *darker colours* (*dark blue*) correspond to instantaneous firing rates close to zero. Following Colburn et al. (2003), the instantaneous firing rate was approximated as a scaled exponential function of the input waveform where the scaling factor was chosen to yield a synchronization index of 0.81. Because the purpose of this figure is to illustrate specifically the phase effects and the relative timing of neural responses to tones of different frequencies, the effects of peripheral frequency selectivity on the firing rate are not shown. The *panel on the right of the plot* shows the mean across time of the absolute value of the difference in instantaneous rate between nearby simulated auditory channels, similar to Cedolin and Delgutte (2010). The two resulting mean absolute valued difference functions (one for the 1,000-Hz pure tone and the other for the 1,200-Hz pure tone) have been normalized so that their maximum peak amplitude equals 1. Note that the peak corresponds to a higher CF for the 1,200-Hz than for the 1,000-Hz tone.

known for some time that the phase response of the BM and of individual neurons varies with the input level (Anderson et al. 1971; Rhode 1971); in fact, the slopes of the lines in Figure 1 are slightly shallower at higher levels, reflecting the broader frequency tuning at those levels (Kim, personal communication). Moreover, changes in this slope with level, arising from the operation of the cochlear amplifier and the consequent change in tuning, have been proposed as a means by which the sound level is encoded by the auditory system (Carney 1994; Heinz et al. 2001; Colburn et al. 2003), suggesting that PT cues may not provide a level-independent cue to pitch. The present study investigates this issue in an attempt to determine whether PTs can support a level-independent code for pure-tone pitch and evaluates four possible ways that phase transitions could be processed in order to provide such a level-independent code.

LEVEL DEPENDENCE OF PHASE TRANSITION CUES FOR PURE-TONE PITCH

Modelling approach

The phase transition data reported by Kim et al. (1980), and shown in Figure 1, provide an important demonstration of the presence of across-channel timing information in the AN. Here, we report the results of a quantitative treatment of some more recent data, reported by Palmer and Shackleton (2009), and which those authors generously shared with us. They measured the amplitude and phase responses of individual AN fibres of the anaesthetized guinea pig (GP) in response to tones of different frequencies and levels. Fortunately, those data were obtained for a range of fibre CFs, permitting a reanalysis in terms of the response across CFs to a tone of a given frequency and level. Our analysis differs from previous approaches that were based on *simulated* neural responses produced using models of the auditory periphery (Shamma 1985; Carney et al. 2002) or that estimated the response of neurons of different CFs to a given stimulus by recording from an *individual* neuron in response to a *range* of stimuli (Cedolin and Delgutte 2010). This allowed us to obtain direct measurements of the variation of firing phase with CF and to avoid having to make some of the simplifying assumptions associated with those alternative approaches.

Another feature of our approach is that we do not focus on trying to model the magnitude of behaviourally obtained frequency difference limens (FDLs), for two reasons, both of which apply quite generally to attempts to model perception using neural responses. First, the absolute size of predicted FDLs will be influenced by the noise present in the physiological data on which the predictions are based, by the need to combine measurements from several animals, and by assumptions concerning the number of neurons involved and the degree of statistical independence of their responses. Second, the neural data are from a species whose behavioural FDLs are larger than in humans, may vary with frequency in a different way,

and may be influenced by non-sensory factors such as attention (Heffner et al. 1971; Hienz et al. 1993). Instead, we concentrate on determining whether across-channel timing cues could provide a *level-independent* code for the frequency of a pure tone. We do so because, although the pitch of pure tones in humans does vary with sound level, this variation is small and is of the order of only a few per cent (Stevens 1935; Walliser 1969; Verschuure and van Meeteren 1975). A cue that varied substantially with the overall level would not, we assume, provide a useful means of encoding pitch, and we make the additional assumption that this will be true for the GP as well as for humans. Specifically, we compare the extent to which the predictions of each potential coding scheme vary with input level and with input frequency, f_s. Note that although "noise" in the data may obscure the effects of either f_s or level, it is unlikely to cause effects that are not significant to appear so. Note also that we are comparing the size of two effects—f_s and level—both of which are subject to the same measurement noise. Hence, the noise is unlikely to distort our estimate of the relative size of the two effects. Our analysis shows that many ways of processing PT cues—described in detail in "Behaviour of four cross-channel phase difference models"—are, in fact, strongly level-dependent. As such, those methods are unlikely, by themselves, to provide a robust level-independent code for the pitch of pure tones. A possible exception is a scheme that compares AN fibres that have different CFs and that respond out of phase to each other (cf. Carney et al. 2002); the plausibility of this scheme is discussed.

The analyses that follow use data obtained by Palmer and Shackleton (2009), and the reader is referred to that original article for experimental details. Stimuli were 50-ms sinusoids presented at a rate of 5 Hz to urethane-anaesthetized adult pigmented GPs. The phases and firing rates of the responses to these stimuli were then extracted, as a function of AN fibre CF, for a range of input levels. The data were typically obtained at 10-dB intervals; in the majority of cases, these levels corresponded (within 1 dB) to those specified in our analyses. The stimulus frequency spacing was in steps of 1.5 semitones from the CF of the considered AN fibre. Where the levels and frequencies specified in an analysis did not correspond exactly to those available in the data, phases and firing rates for a given input were estimated by linear interpolation from the values at the next highest and the next lowest levels and frequency where the measurements were obtained.

As noted above, estimates of the accuracy of any code will be affected by details specific to the method of data collection and to assumptions made in the modelling process. Hence, although plots of our model predictions are accompanied by error bars, these provide a guide to the statistical significance of the effects that we discuss, but do not indicate how small a frequency difference should be detectable by the GP.

Behaviour of four cross-channel phase difference models

Local subtraction model. Perhaps the simplest scheme for estimating the frequency (and therefore, the pitch) of pure tones, that takes advantage of the PT, is one in which an array of "second-order" neurons subtracts the response of each AN fibre from that of a fibre having a "neighbouring" CF. If the function relating phase to CF in the first-order (AN) neurons has a slope that is locally steeper over some region compared with surrounding regions, there will be a corresponding local maximum in the output of the second-order array. This concept is illustrated in Fig. 3, where part A) shows a schematic phase-vs-CF curve and where part B) shows the output of a hypothetical second-order array. It is also illustrated in Figure 2 which shows some schematic "neurograms" to 1,000- and 1,200-Hz pure tones, with the results of a local subtraction algorithm shown to the right of the plot. Note that the neurograms in Figure 2 include the effects of peripheral filtering on response phase, but not on its magnitude (firing rate). Note also that the phase-vs.-CF functions shown in Figure 3 and in all remaining plots in this article show CF increasing from low to high along the abscissa, opposite to the convention adopted by Kim et al. and shown in Figure 1.

As mentioned above, for this "local subtraction" scheme to provide a useable measure of input frequency, the phase-vs.-CF function must show a local increase in slope over some fairly narrow range of CFs, and this maximum should vary monotonically with the input frequency. To check whether this was the case, we first calculated arc-tangent fits to the functions relating phase in cycles to $\log_{10}(\mathrm{CF})$ for input levels of 50, 70 and 90 dB SPL and for f_s ranging from 250 to 2,000 Hz. The fitted functions took the form $y=A+B\cdot\arctan(x/C)$, where y is the phase of the simulated response of a fibre having CF=x and A, B, and C are free parameters. Those data and fits are shown in the top panels of Figure 4 where the different colours in each panel show data for different signal frequencies at a single level (fits to the 70 dB SPL data were also used to generate the "neurograms" in Fig. 2). The bottom panels of Figure 4 show an implementation of the "local subtraction" scheme in which the phases in each "bin" are subtracted from

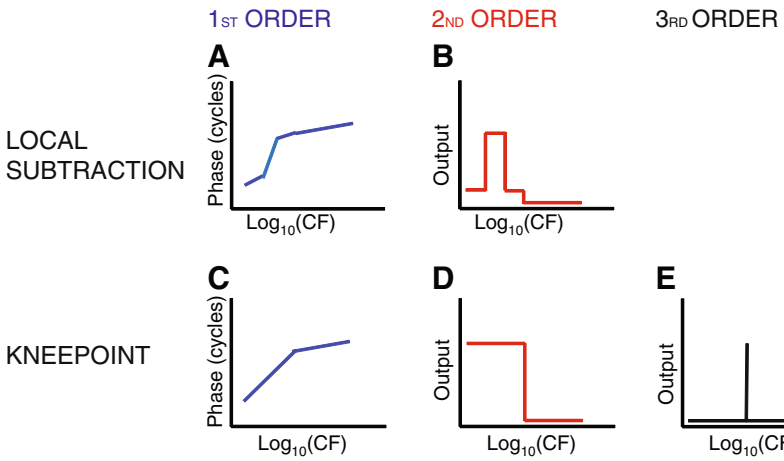

FIG. 3. The *top row* shows a schematic implementation of the local subtraction scheme. **A** Schematic representation of a phase-vs.-$\log_{10}$ CF curve consisting of a shallow portion plus a steeper portion having a region where the slope is locally maximal. **B** Output of an array of second-order neurons, whereby the instantaneous output of cells at a given CF are subtracted from that of their neighbour. This subtraction is implemented by subtracting the phases of adjacent channels. The *bottom row* shows the knee point model. **C** Phase-vs.-CF function consisting of a shallow and a steep portion. **D** Output of an array of second-order neurons, whereby the instantaneous output of cells at a given CF are subtracted from that of their neighbour. **E** Output of an array of third-order neurons operating on the different between adjacent second-order neurons shown in (**D**).

those of its neighbour; the bin widths use here correspond to units of 0.1 $\log_{10}$(CF). It can be seen that these functions do indeed show local maxima, which, although quite broad, generally vary monotonically with frequency, as illustrated for each level by the different curves in Figure 5. However, Figure 5 also shows that these "peak differences" are unfortunately not invariant with level. For example, every estimate at 90 dB SPL is higher than the corresponding value for the same input frequency at 50 or 70 dB SPL. Although the exact size of the level effect is rendered uncertain by the variability in our estimates, both it and the effect of frequency were significant when the peak difference values for three levels (50, 70, 90 dB SPL) and for ten values of f_s spaced in units of 0.1 $\log_{10} f_s$ were entered into a generalized linear model (level: $F_{1,28}$= 42.02, P<0.0001; frequency: $F_{1,28}$=42.02, P<0.0001). For this and the other metrics described in this section, we derived a simple summary of the relative effects of f_s and level by comparing its variation with f_s from 250 to 1,189 Hz, averaged over level, to its variation with level averaged across that same range of f_s. For the peak difference measure, the variation with level was 0.7 octaves, more than half the 1.24-octave effect of f_s.

Our implementation of the local subtraction scheme is a simplification for at least two reasons. First, the neural responses in any one AN fibre will not occur at exactly the same phase on every cycle, and, even on the same cycle, it is likely that different neurons with the same CF will fire at slightly different times. Second, any second-order neuron that subtracts the inputs from two adjacent channels will impose some smoothing due to variability in synaptic transmission times. This does not negate our assumption that the output of such a subtraction will on average increase monotonically with the phase difference between the two input AN channels. Furthermore, assuming that, for a given f_s, the smoothing is the same at all CFs, the output of the local subtraction algorithm will have peaks in the same locations as those shown in the bottom row of Figure 4. It does, however, have one potential implication for our evaluation of the local subtraction scheme, which is based on the subtraction of phases rather than of simulated spike trains. This is that if any component of "neural smoothing" has a time constant that is constant in milliseconds rather than in radians, its effects would be greater for higher-frequency than for lower-frequency tones. As a result, the output of the subtraction operation would have peaks that, compared with those shown in the bottom row of Figure 4, would be relatively higher for low than for high f_s.

A model based on the local subtraction idea was recently applied by Cedolin and Delgutte (2010) to the responses of cat AN fibers to harmonic complex tones. An advantage of this approach is that the across-channel timing comparisons are performed locally, and, where a complex tone consists of low-numbered harmonics, can be performed between AN fibers responding to the same harmonic (Cedolin and Delgutte 2010). This is important because it generates predictions consistent with the fact that the pitch of low-numbered harmonics is independent of their relative phase (Houtsma and Smurzynski 1990; Shackleton and Carlyon 1994) and because it would allow phase comparisons to be performed even when

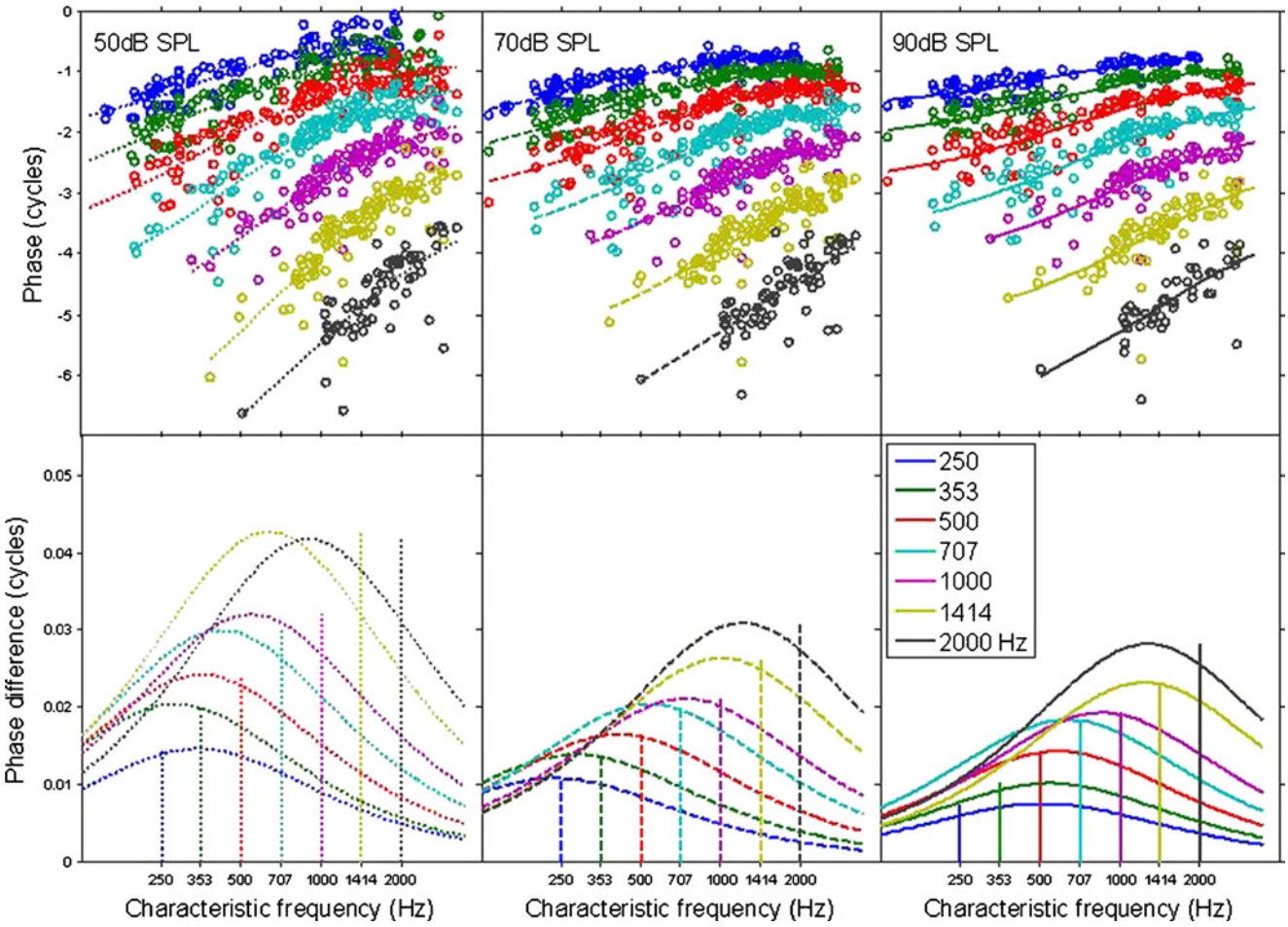

FIG. 4. *Top panels* show the phase of the AN response as a function of characteristic frequency for input frequencies ranging from 250 to 2,000 Hz. Different input frequencies are *colour-coded*; the *dotted lines* show the best-fitting arc-tangent fits. *Bottom panels* are derived from the *top panels* and show the phase difference between adjacent CF bins, where each bin is $0.1(\log_{10}(\text{CF}))$. The *vertical dashed lines* intersect the abscissa at CFs equal to the (colour-coded) input frequency. *Left*, *middle*, and *top columns* are for input levels of 50, 70 and 90 dB SPL, respectively.

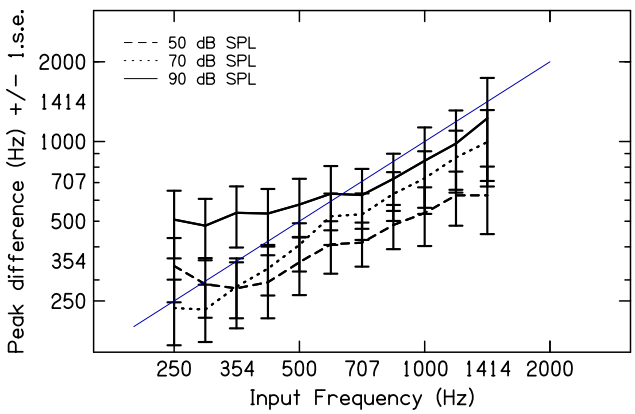

FIG. 5. Peaks in the "phase difference curves" (*bottom panels* of Fig. 4) as a function of stimulus frequency for three different levels. The peaks were estimated using the "robust regression" algorithm in Matlab (version 7.7.0, R2008b, The MathWorks, Natick, MA), which uses iteratively reweighted least squares with a bisquare weighting function. *Error bars* show 95% confidence limits.

energy from other sound sources is present elsewhere in the spectrum. We compare Cedolin and Delgutte's findings to the results of our analyses in "Comparison with other data". Although neither we, nor they, could specify exactly which neurons might perform this subtraction, Cedolin and Delgutte suggested the dorsal cochlear nucleus as one possible site.

CF-at-knee point model. A second possible code for frequency that makes use of the PT involves estimating f_s from the location of the transition between the steep and shallow portions of the functions shown in Figure 1 and in the top panels of Figure 4. This could be implemented by an array of "third-order" neurons that subtract the outputs of adjacent "second-order" neurons (Figure 3C–E). This "third-order" array would show a maximum when the second derivative of the phase-vs.-CF function is maximal. To estimate this "knee point", we refitted

the data shown in Figure 4 with "broken-stick" functions consisting of two straight lines, one of which was constrained to have a zero slope. Examples of this fit for two frequencies and levels are shown in Figure 6. When applied to the same range of f_s as shown in Figure 4, this "broken-stick" model yields almost exactly the same RMS error of fit (all frequency and level conditions combined) as the "arc-tangent" model described above: 0.2791 cycles for the broken-stick fit model vs. 0.2796 for the arc-tangent fit. The knee point varies monotonically with f_s, but, as Figure 7 shows, is not independent of level. For example, the knee point for a 250-Hz, 90-dB tone occurs at a place in the AN array having a higher CF than that for a 500-Hz, 50-dB tone. The effect of level, averaged across f_s, was 0.8 octaves, more than half the 1.33-octave effect of f_s.

Phase transition slope model. Loeb et al. (1983) proposed a model consisting of an array of detectors in the medial superior olive, with each detector receiving inputs from pairs of AN fibers that were separated along the BM by a fixed amount. They pointed out that the maximum output would occur in those detectors whose input AN fibers responded with phases that differed by one wavelength (although they also noted that optimal discrimination might be based on other phase differences). This "critical distance" depends on the slope of the function relating phase in cycles to $\log_{10}$(CF). As shown in Figures 1, 4, and 6, this slope is not constant as a function of CF but, rather, increases markedly for CFs below the knee point. Here, we focus on the slope of the steep portion as a possible cue for f_s. This slope will depend on the bandwidths of the peripheral filters through which the stimulus is passed (Heinz et al. 2001; Shera et al. 2010); a filter introduces a delay as the frequency of the input passes through its resonance, and the size of this delay is greater for narrow than for broad filters. As the slope is measured in terms of the *logarithm* of CF (which is roughly proportional to the distance along the BM), it will vary with the bandwidth of the analysing filters *relative to* their centre frequencies. This last point can be explained by considering a 1,000-Hz tone passed through a filter centred on, say 1,200 Hz and a 2,000-Hz tone passed through a filter centred on 2,400 Hz. If the bandwidths of the filters are a constant proportion of their centre frequencies, then each tone will be at the same relative position on the filter's transfer function and the phase delays (in cycles) will be equal.

As shown in Figure 8, the slopes vary as functions both of level and of f_s. The slopes are slightly shallower (less negative) at lower frequencies, presumably because, in the GP, the relative bandwidths of frequency tuning curves (e.g. as summarized by Shera

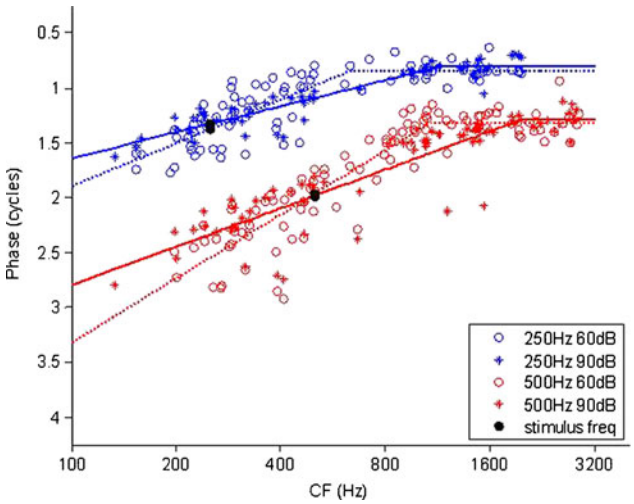

FIG. 6. Broken-stick fits to phase transition data for f_s=250 (*blue*) and 500 Hz (*red*) and for input levels of 60 dB SPL (*circles*) and 90 dB SPL (*asterisks*). The broken-stick fits are shown by *dotted* and *solid lines* for levels of 60 and 90 dB SPL, respectively.

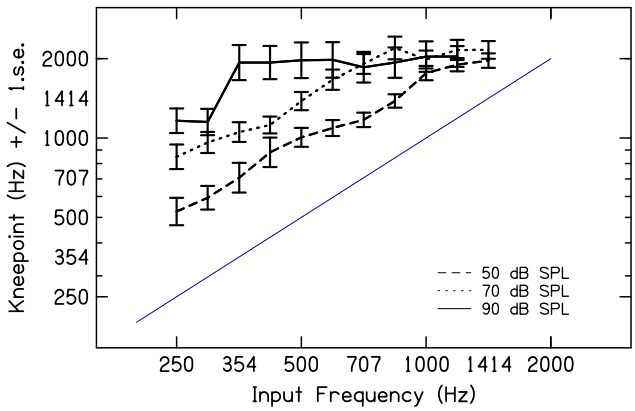

FIG. 7. The "knee point" as a function of level and input frequency. The curves shown are derived from the best-fitting model described in the text for input frequencies spaced 3 semitones apart, starting at 250 Hz, and with levels of 50, 70 and 90 dB SPL.

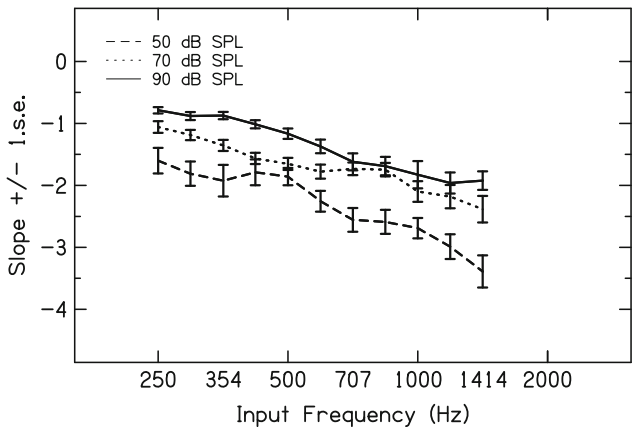

FIG. 8. Slope of the phase transition as a function of level and input frequency. The curves shown are derived from the best-fitting model described in the text for input frequencies spaced 3 semitones apart, starting at 250 Hz, and with levels of 50, 70 and 90 dB SPL.

et al. 2002) are broader at low frequencies. The slopes also become shallower (less negative) with increases in level, as would be expected with the decrease in frequency selectivity at high levels (Rhode 1971; Robles et al. 1986). Both of these effects were sufficiently reliable to produce significant main effects when the slope values for three levels (50, 70, 90 dB SPL) and for ten values of f_s spaced in units of 0.1 $\log_{10} f_s$ were entered into a generalized linear model (effect of f_s: $F_{1,28}$=28.9, P<0.00001; effect of level: $F_{1,28}$=22.1, P<0.00001). Averaged across f_s, slopes varied by a factor of 1.67 across level, approaching the factor of 2.07 for the variation with f_s. Hence, any attempt to estimate f_s based on a metric related to the slope—for example by estimating the difference along the AN array between fibers that responded in phase—would be strongly influenced by the input level.

Model involving a comparison of phase above and below the knee point. One potential decoding method is illustrated in Figure 9 which shows "broken-stick" fits to stimuli at two frequencies and levels (shown in Fig. 6), shifted so that the horizontal portions are aligned. The shift is performed because the animal has, of course, no idea about "absolute" phase and can only compare the relative phase across fibers. It can be seen that the CF at which the phase is π radians lower than that at the knee point (indicated by the black horizontal dotted lines) appears to be similar at the two levels for each frequency. Figure 10 shows this value, which we term the "π shift point", as a function of f_s for input levels of 50, 70 and 90 dB SPL. It increases monotonically with f_s and has modest error bars. A generalized linear model revealed a

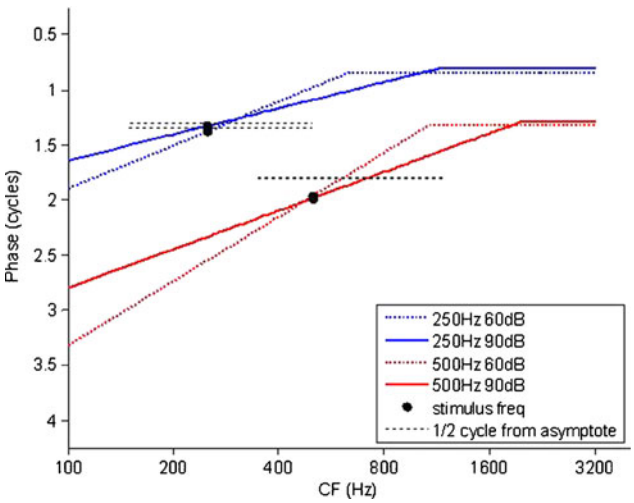

FIG. 9. "Broken-stick" fits to the phase transition curves for input frequencies of 250 and 500 Hz and levels of 60 and 90 dB SPL. The 90-dB curves have been shifted upwards so that their horizontal portion overlaps with that of the 60-dB curves. The *horizontal dashed lines* show the phases that are half a cycle lower than at the knee point.

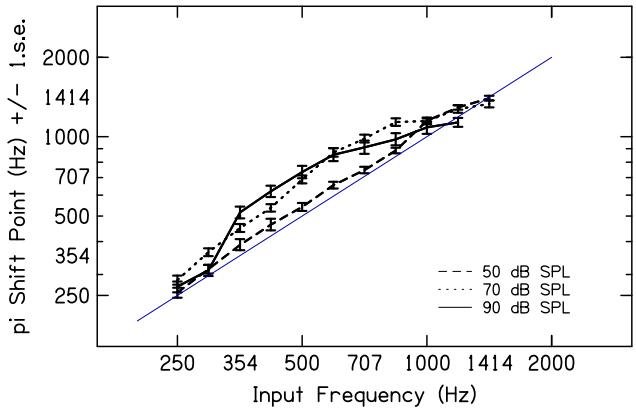

FIG. 10. Value of the "π shift point" as a function of level and input frequency. The curves shown are derived from the best-fitting model described in the text for input frequencies spaced 3 semitones apart, starting at 250 Hz, and with levels of 50, 70 and 90 dB SPL.

significant effect of f_s ($F_{1,28}$=524, P<0.0001), but not of level ($F_{1,28}$=0.23, P=0.633). Unlike the other measures described here, the effect of level, averaged across f_s, was much smaller (0.23 octaves) than the effect of f_s, which was 2.19 octaves; this, in turn, was close to the range of input frequencies (250–1,189 Hz) which was 2.24 octaves.

A matter that would still need to be resolved is how the auditory system would extract this metric. A possibly important fact is that the phase at and above the knee point is fairly constant over a wide range of CFs, and so is the phase at which most AN fibers are firing. Therefore, "all" that the auditory system has to do is to find those fibers that are firing out of phase (by π radians) with the "most common" phase. However, phase is only "fairly" constant at CFs above the knee point. The data of Kim et al. (Fig. 1) show that, if one looks over a much wider range of CFs than in the data of Palmer et al., the phase delay does decrease gradually as one moves to more basal regions. Therefore, the issue arises as to which portion of the AN array is used as the "fairly constant" reference. An alternative would be for the system to measure the CF corresponding to knee point, perhaps using a simple metric as in Figure 3C–E and to exploit this information to identify the π shift point. An important drawback is that the "π shift point" and the knee point are often far apart (more than an octave difference in CF; Fig. 9), raising the issue of how the brain would extract this cue when more than one sound is present. Furthermore, when more than one tone is present, basal fibers will respond to a mixture of both, and the auditory system would have to extract the temporal pattern corresponding to each one. Evidence from the perception of spectrally overlapping mixtures of unresolved harmonics suggests that the brain is poor at extracting two periodicities from the same set of AN fibers (Carlyon 1996a, b).

DISCUSSION

Use of firing rate profiles

The analyses described in "Level dependence of phase transition cues for pure-tone pitch" focused solely on across-channel timing information. It is, however, possible that the auditory system uses a coding scheme that combines information on both the timing and the rate of firing across channels. Although it is beyond the scope of this article to perform such a combined analysis, some preliminary considerations suggest that it may not be useful in overcoming the effects of level on the various codes considered above. Figure 11 shows the firing rate profiles, defined as firing rate as a function of CF, normalized so that the spontaneous and maximum rates for each neuron scaled to between 0 and 1, respectively. Those data were then fit by second-order polynomials. The polynomial fits, although imperfect, capture the main features of the data. Note that although the profile at 60 dB SPL has a bandpass characteristic, the one at 90 dB SPL appears largely flat over the >3 octave range of CFs from which measures were obtained. Saturation of AN rate-level functions almost certainly accounts for at least part of this effect, and it is possible that a bandpass curve could be obtained by considering only those neurons with low spontaneous firing rates, which are relatively resistant to such effects (cf. Sachs et al. 1983). However, this would invoke selective processing of low spontaneous rate fibers, the necessity of which has been a major reason for questioning the validity of purely rate–place accounts (e.g. Carney 1994).

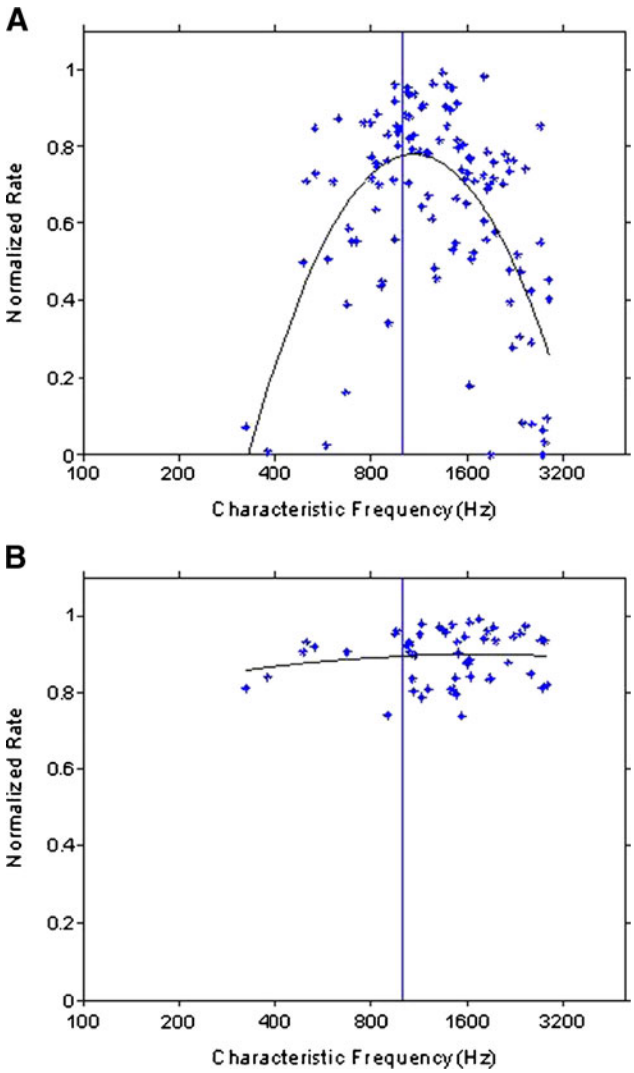

FIG. 11. Normalized firing rate as a function of CF. Input: 1,000 Hz tone with a level of 60 dB SPL (**A**) or 90 dB SPL (**B**). The *vertical blue line* indicates CF = input frequency. The *solid black lines* are second-order polynomial fits.

Comparison with other data

van der Heijden and Joris (2006) measured the amplitude and phase characteristics at different apical locations of the cat cochlea by recording the responses to inharmonic complex tones from cat AN fibers having a range of CFs. Their technique was designed to remove nonlinear contributions to the AN response arising from the transduction process (e.g. half-wave rectification and firing rate saturation). Measurements were obtained for stimuli at relatively low levels (within 35 dB of each fibre's threshold) and were not presented separately for different levels. Their data, like those presented here, reveal phase-vs.-CF functions that are quite flat for CFs much higher than the input frequency, combined with a steeply sloping portion around CF. As also shown here, they found that the slope is steeper at higher frequencies: approximately −2.1 and −3.5 cycles/$\log_{10}$(CF) at 200 and 1,000 Hz, respectively. These slopes are somewhat steeper than those shown in Figure 8, probably due to the lower level used by them and/or to species differences. A visual inspection of some cat data presented by Kim et al. (1979) reveals a slope of about −2 cycles/$\log_{10}$(CF) at L_s=45 dB SPL for f_s= 620 Hz. The data of Kim et al. 1980, plotted in Figure 1, show substantially steeper slopes of about 6 cycles/$\log_{10}$(CF), presumably reflecting the higher f_s of 2,100 Hz and the marked increase in the $Q_{10\ dB}$ values of cat AN fibers as CF is increased above about 1,000 Hz (Palmer 1995; Shera et al. 2002). In general, the results of Kim et al. and of van der Heijden and Joris agree reasonably well with those described here despite the fact that both of those earlier data sets were obtained in the cat, whereas Palmer and Shackleton's data were obtained in the GP. In addition, van der Heijden and Joris noted that they deliberately restricted their measures to fairly low

levels because increasing level could have a "drastic" effect on the phase measurements. This is consistent with evidence from recordings of individual neurons innervating the cochlear apex showing that level affects group delay in a frequency-dependent manner (Versteegh et al. 2011) and with the present evidence that PTs are strongly affected by level. Hence, although there will undoubtedly be quantitative differences between species, the general form of phase transition data is similar.

Cedolin and Delgutte (2010) measured the response of individual AN fibers to harmonic complex tones whose fundamental frequency (F_0) ranged from about 0.2 to about 0.7 times the neuron's CF. They invoked the principle of "scaling invariance" (Zweig 1976) to calculate, for each set of such measurements, the predicted amplitude and phase response of an array of neurons, with different CFs, to a complex of a given F_0. They then calculated not only the predicted rate–place profile but also a metric similar to the simple "local subtraction" method described above; the response of each "channel" was instantaneously subtracted from that of its neighbour and the absolute value of this difference then integrated over time. The variation in this summary statistic across CF, termed the "mean absolute spatial derivative (MASD)", produced peaks at CFs corresponding to the low-numbered harmonics. They found that for CFs between 1,350 and 2,800 Hz, the MASD provided a better representation of the F_0 than did a simple rate–place profile, but that the reverse was true at higher CFs, presumably due to the roll-off in phase locking at high frequencies. In addition, they estimated the best frequency (BF) of each neuron, based both on the rate–place and MASD profiles, and reported that these BFs varied less consistently with level for the MASD-based than for the rate-based measure. They also concluded that the accuracy of the F_0 representation did deteriorate slightly with increasing stimulus level, but that the significance of this deterioration was less than that observed for the rate–place model.

Cedolin and Delgutte's conclusion that their MASD model provided a relatively level-independent cue to pitch seems, at least at first sight, to be at variance with the substantial level dependence observed in our analyses of Palmer and Shackleton's GP data and by van der Heijden and Joris's observation that level can have a large effect on the phase transitions measured in the cat. One way of reconciling these two sets of findings would be if the response of the BM were more linear for complex tones than for the pure tones used by Palmer and Shackleton. In addition, the range of sound levels studied by Cedolin and Delgutte was generally lower than ours; 80% of the data they reported were obtained at levels between 25 and 65 dB SPL per component, whereas we report data for levels between 50 and 90 dB SPL. It is also worth noting that Cedolin and Delgutte focused on the finding that the PT cue was less *level-dependent than the rate–place representation*, a finding that is not inconsistent with our analysis. For example, Figure 11 shows that the rate vs. CF profile at 90 dB SPL is flat, whereas the "local subtraction" metric shown in Figure 5 does show some variation with input frequency at that level. Our focus is on whether there is an effect of level *at all* and on whether this effect is *substantially less than the effect of input frequency*.

Scaling invariance

Our analysis also sheds light on the validity of the "scaling invariance" assumption and on the implications of violations of this assumption for phase transition cues. According to the principle of scaling invariance, any PT curve in Figure 4 for a given value of f_s can be obtained by taking the curve for a frequency of $x \cdot f_s$ and shifting it horizontally by $\log_{10} x$. Figure 12 illustrates this point and shows the same general pattern for signal levels of 50, 70 and 90 dB SPL (shown in different panels). In each panel, the thin red curve shows the arc-tangent fit to the PT curve for f_s=500 Hz, and the thick red curve shows that function shifted to the right by 1.6 octaves. The thin dotted black curve shows the PT response to a pure tone 1.6 octaves higher than 500 Hz—i.e. f_s=1,516 Hz, and the thick dotted black curve shows this function shifted vertically so that it coincides with the red curve at its rightmost point, corresponding to a CF of 2 kHz, in order to facilitate visual comparison. If scaling invariance holds, then the thick red and black curves should be identical, which they are not. Rather, at a point 1.6 octaves below that at which we forced them to have the same phase, they differ by between 0.45 and 0.54 cycles. What this means is that over a frequency range (1.6 octaves) which was the same as that used by Cedolin and Delgutte to study the responses of each cat AN fibre, there is a substantial deviation from scaling invariance as measured in the GP. Although a marked deviation from scaling invariance has been observed when comparing the phase lag at the base and apex of the cat cochlea (van der Heijden and Joris 2006), we cannot be sure that the failure of scaling invariance over the relatively local range of 1.6 octaves, observed here for the GP, would necessarily occur for the cat. However, it is worth noting that the variation in the relative bandwidths of frequency tuning curves with CF is roughly similar in the two species (Shera et al. 2002).

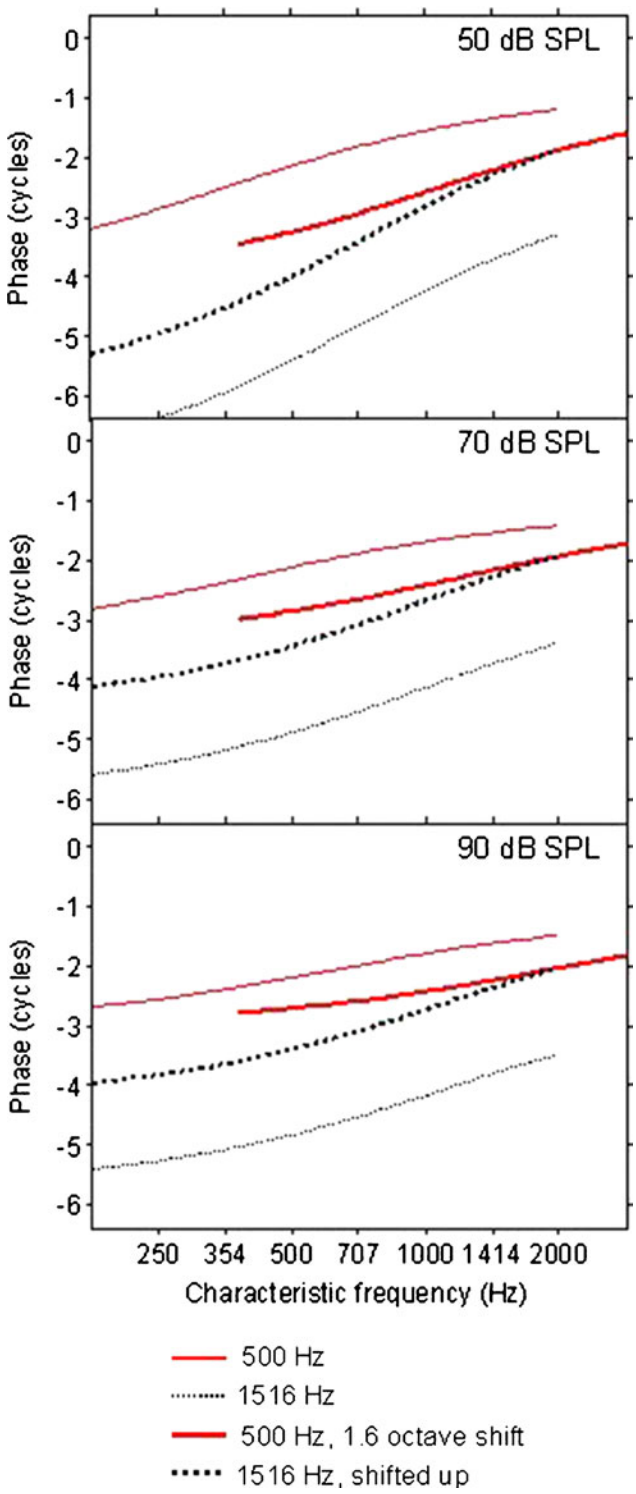

FIG. 12. Illustration of the failure of scaling invariance. Each panel shows fits to the data obtained at one level (50 dB SPL (**A**), 70 dB SPL (**B**) and 90 dB SPL (**C**)). In each panel, the *thin red line* and the *thin dotted black line* show the arc-tangent fits from Figure 4 for input frequencies of 500 and 1,516 Hz, respectively; these frequencies are 1.6 octaves apart. The *thick red solid line* shows the fit to the 500-Hz data shifted rightwards by 1.6 octaves. The *thick dotted black line* shows the fit to the 1516-Hz data shifted upwards so that it coincides with the shifted 500-Hz fit at a CF of 2,000 Hz. If scaling invariance held, these two curves should overlap completely.

Level dependence of across-channel phase cues

In the Introduction, we noted that the nonlinearity of the mechanical filtering properties of the BM made it plausible that PTs, which arise from those properties, would also be nonlinear and vary with input level. The analyses described in "Level dependence of phase transition cues for pure-tone pitch" generally support this prediction. It is possible that the phase effects we observed had not only a mechanical but also a neural basis. One potential influence of the amplitude response of the BM to a pure tone on the PT arises from a potential influence of level on the synaptic delay between inner hair cells (IHCs) and the AN. Recently, Versteegh et al. (2011) interpreted some of their data on the level dependence of AN phase responses in terms of evidence that larger input amplitudes may cause AN fibres to fire earlier, perhaps as a result of the effects of sound level on vesicle release at the IHC synapse. If this is true then, when, for a given input level, the inputs to fibres with CFs remote from f_s are reduced by BM filtering, those AN fibres may be subject to slightly different synaptic delays, thereby contributing to the variation in the AN response phase as a function of CF. Furthermore, because BM filtering itself depends on the input level, the variation in these "synaptic" effects across CF may also be level-dependent.

Summary of PT cue analysis

1. For a given f_s, the function relating AN response phase to $\log_{10}$(CF) is very shallow for CFs much higher than f_s and substantially steeper for CFs around and below f_s. This PT curve shows a local maximum, the position of which varies monotonically with f_s. However, this position also varies with level, and the average effect of a 40-dB level change on its position is more than half that produced by changing input frequency over a 2.24-octave range.
2. The "knee point"—the position on the PT curve at which there is a marked increase in slope—also varies monotonically with f_s and is also level-dependent. The average effect of a 40-dB level change is more than half that produced by changing input frequency over a 2.24-octave range.
3. The slope of the PT curve varies slightly with f_s, presumably as a result of the variation in relative tuning with CF in the GP. It becomes shallower with increasing input level, presumably reflecting the worsening of frequency selectivity with increasing level observed here. The average effect of a 40-dB level change approached that

produced by changing input frequency over a 2.24-octave range.

4. Models based on co-incidence detection are dependent on the slope of the phase-vs.-CF function. Over the steep part of that function, this slope is level-dependent and its variation with f_s is slight and depends on variations in relative bandwidth with CF—a variation that may be absent above 500 Hz in humans (Baker and Rosen 2006). Therefore, to the extent that such models assume that phase differences are compared over the steep part of the function, they do not predict a robust, level-independent code for frequency.
5. The CF at which the phase of the AN response is exactly out of phase with that at the knee point varies monotonically with f_s. Unlike the other potential codes described here, its variation with level was not significant overall. The average effect of a 40-dB level change was only about one tenth that produced by changing input frequency over a 2.24-octave range. However, this code requires the auditory system to compare channels with well-separated CFs, and it is not obvious how it could be used when more than one tonal component is present.

Overall summary and conclusions

Our analysis of Palmer and Shackleton's (2009) GP data suggests that *straightforward* methods for extracting pure-tone frequency from PTs—such as the "local subtraction" and "knee point" metrics—are likely to be too strongly affected by level to provide a robust code. Perhaps the strongest candidate code that we observed was the CF at which neurons fire out of phase with those at and basal to the knee point; this metric varied substantially and monotonically with f_s and was not substantially influenced by level. However, this code requires a comparison of AN fibres having quite different CFs, and it is less than clear how it could be implemented by the auditory system, particularly when more than one sound is present. A *caveat* that the present analyses shares with most other published studies is that the neural responses were obtained in an anaesthetized preparation.

Overall, our conclusion is therefore that although across-channel timing information provides a theoretically attractive neural code to pure-tone frequency (subjectively, pitch), the way in which it is extracted is unlikely to be straightforward. One possible conclusion is that across-channel timing cues are not used at all. At present, the most likely alternatives are either that the auditory system can estimate a quantity akin to the π-shift point, which entails comparing phase information across quite well-separated portions of the AN array, or uses one of the more "local" codes described here whilst controlling for the sometimes substantial effects of overall level, for example, by estimating total firing rate.

ACKNOWLEDGEMENTS

We are very grateful to Alan Palmer and Trevor Shackleton for providing us with their raw data and with some MATLAB scripts. Thanks are also due to Chris van den Honert, Associate Editor Laurel Carney and the anonymous reviewers for insightful comments on an earlier version of the manuscript.

OPEN ACCESS

REFERENCES

ANDERSON DJ, ROSE JE, HIND JE, BRUGGE JF (1971) Temporal position of discharges in single auditory nerve fibers within cycle of a sine-wave stimulus—frequency and intensity effects. J Acoust Soc Am 49:1131

BAKER RJ, ROSEN S (2006) Auditory filter nonlinearity across frequency using simultaneous notched-noise masking. J Acoust Soc Am 119:454–462

CARIANI PA, DELGUTTE B (1996) Neural correlates of the pitch of complex tones. I. Pitch and pitch salience. J Neurophysiol 76:1698–1716

CARLYON RP (1996A) Encoding the fundamental frequency of a complex tone in the presence of a spectrally overlapping masker. J Acoust Soc Am 99:517–524

CARLYON RP (1996B) Masker asynchrony impairs the fundamental-frequency discrimination of unresolved harmonics. J Acoust Soc Am 99:525–533

CARNEY LH (1994) Spatiotemporal encoding of sound level—models for normal encoding and recruitment of loudness. Hear Res 76:31–44

CARNEY LH, HEINZ MG, EVILSIZER ME, GILKEY RH, COLBURN HS (2002) Auditory phase opponency: a temporal model for masked detection at low frequencies. Acta Acustica United with Acustica 88:334–347

CEDOLIN L, DELGUTTE B (2010) Spatio-temporal representation of the pitch of harmonic complex tones in the auditory nerve. J Neuroscience 30:12712–12724

CHATTERJEE M, ZWISLOCKI JJ (1997) Cochlear mechanisms of frequency and intensity coding. I. The place code for pitch. Hear Res 111:65–75

COLBURN HS, CARNEY LH, HEINZ MG (2003) Quantifying the information in auditory-nerve responses for level discrimination. Jaro-J Assoc Res Otolaryngol 4:294–311

DE CHEVEIGNÉ A (2005) Pitch perception models. In: PLACK CJ, OXENHAM AJ (eds) Springer handbook of auditory research: pitch perception. Springer, New York, pp 169–233

DE CHEVEIGNÉ A, PRESSNITZER D (2006) The case of the missing delay lines: synthetic delays obtained by cross-channel phase interaction. J Acoust Soc Am 119:3908–3918

EGGERMONT JJ (1991) Rate and synchronization measures of periodicity coding in cat primary auditory cortex. Hear Res 56:153–167

Heffner R, Heffner H, Masterto B (1971) Behavioral measurements of absolute and frequency-difference thresholds in guinea pig. J Acoust Soc Am 49:1888–1895

Heinz MG, Colburn HS, Carney LH (2001) Rate and timing cues associated with the cochlear amplifier: level discrimination based on monaural cross-frequency coincidence detection. J Acoust Soc Am 110:2065–2084

Hienz RD, Sachs MB, Aleszczyk CM (1993) Frequency discrimination in noise: comparison of cat performances with auditory-nerve models. J Acoust Soc Am 17:462–469

Houtsma AJM, Smurzynski J (1990) J.F.Schouten revisited: pitch of complex tones having many high-order harmonics. J Acoust Soc Am 87:304–310

Kim DO, Siegel JH, Molnar CE (1979) Cochlear nonlinear phenomena in two-tone responses. Scand Audiol Suppl 9:63–81

Kim DO, Molnar CE, Matthews JW (1980) Cochlear mechanics: nonlinear behavior in two-tone responses as reflected in cochlear-nerve-fiber responses and in ear-canal sound pressure. J Acoust Soc Am 67:1704–1721

Licklider JCR (1951) A duplex theory of pitch perception. Experientia 7:128–133

Loeb GE (2005) Are cochlear implant patients suffering from perceptual dissonance? Ear Hear 26:435–450

Loeb GE, White MW, Merzenich MM (1983) Spatial cross-correlation. Biol Cybernetics 47:149–163

Meddis R, Hewitt M (1991) Virtual pitch and phase sensitivity studied using a computer model of the auditory periphery: pitch identification. J Acoust Soc Am 89:2866–2882

Meddis R, O'Mard L (1997) A unitary model of pitch perception. J Acoust Soc Am 102:1811–1820

Meddis R, O'Mard L (2006) Virtual pitch in a computaional physiological model. J Acoust Soc Am 120:3861–3869

Moore BCJ (2003) An introduction to the psychology of hearing, 5th edn. Academic, San Diego

Moore BCJ, Carlyon RP (2005) Perception of pitch by people with cochlear hearing loss and by cochlear implant users. In: Plack CJ, Oxenham AJ (eds) Springer handbook of auditory research: pitch perception. Springer, New York, pp 234–277

Oxenham AJ, Bernstein JGW, Penagos H (2004) Correct tonotopic representation is necessary for complex pitch perception. Proc Natl Acad Sci U S A 101:1421–1425

Palmer AR (1995) Neural signal processing. In: Moore BCJ (ed) Hearing. Academic, San Diego, pp 75–113

Palmer AR, Shackleton TM (2009) Variation in the phase of response to low-frequency pure tones in the guinea pig auditory nerve as functions of stimulus level and frequency. Jaro-J Assoc Res Otolaryngol 10:233–250

Plack CJ (2005) The sense of hearing. Lawrence Erlbaum, London

Rhode WS (1971) Observations of the vibration of the basilar membrane in squirrel monkeys using the Mössbauer technique. J Acoust Soc Am 49:1218–1231

Robles L, Ruggero MA, Rich NC (1986) Basilar membrane mechanics at the base of the chinchilla cochlea I. Input–output functions, tuning curves, and phase responses. J Acoust Soc Am 80:1364–1374

Sachs MB, Voigt HF, Young ED (1983) Auditory nerve representation of vowels in background noise. J Neurophysiol 50:27–45

Schnupp J, Nelken I, King A (2010) Auditory neuroscience: making sense of sound. MIT Press, Massachusetts

Schouten JF (1940) The residue and the mechanism of hearing. Proc Kon Akad Wetenschap 43:991–999

Shackleton TM, Carlyon RP (1994) The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination. J Acoust Soc Am 95:3529–3540

Shamma S (1985) Speech processing in the auditory system: II. Lateral inhibition and the central processing of speech evoked activity in the auditory nerve. J Acoust Soc Am 78:1622–1632

Shera CA, Guinan JJ, Oxenham AJ (2002) Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements. Proc Natl Acad Sci U S A 99:3318–3323

Shera CA, Guinan J Jr, Oxenham A (2010) Otoacoustic estimation of cochlear tuning: validation in the chinchilla. J Assoc Res Otolaryngol 11:343–365

Siebert WM (1970) Frequency discrimination in the auditory system: place or periodicity mechanisms. Proc IEEE 58:723–730

Stevens SS (1935) The relation of pitch to intensity. J Acoust Soc Am 6:150–154

van der Heijden M, Joris PX (2006) Panoramic measurements of the apex of the cochlea. J Neurosci 26:11462–11473

Verschuure J, van Meeteren AA (1975) The effect of intensity on pitch. Acustica 32:33–44

Versteegh CPC, Meenderink SWF, van der Heijden M (2011) Response characteristics in the apex of the gerbil cochlea studied through auditory nerve recordings. J Assoc Res Otolaryngol 12:301–316

Walliser K (1969) Zusammenhänge zwischen dem Schallreiz und der Periodentonhöhe. Acustica 21:319–328

Wang X, Lu T, Bendor D, Bartlett E (2008) Neural coding of temporal information in auditory thalamus and cortex. Neuroscience 157:484–494

Wever EV (1949) Theory of hearing. Wiley, New York

Wier CC, Jesteadt W, Green DM (1977) Frequency discrimination as a function of frequency and sensation level. J Acoust Soc Am 61:178–184

Zwicker E (1970) Masking and psychological excitation as consequences of the ear's frequency analysis. In: Plomp R, Smoorenburg GF (eds) Frequency analysis and periodicity detection in hearing. Sijthoff, Leiden, pp 376–394

Zweig G (1976) Basilar membrane motion. Cold Spring Harbor Symp Quant Biol 40:619–633