

位相効果とピッチ・シフト
 —3つの時間ピッチ・モデルの性質—*

ド シュベニエ, アラン (京都大学工学部)

Introduction

Pitch perception theories are of two types: "place" theories based on spectral cues, and "temporal" theories based on time cues. The debate between the two is old, but active. Complex tones with different phase relationships have in most cases the same low pitch, and this argues against temporal models, because both envelope and fine-time cues are sensitive to phase. This paper examines three temporal models, asking two questions:

- a) Do the models give pitch estimates that are phase sensitive?
- b) Is their behaviour phase insensitive in all respects?

I The models

All three models are of a "shift-and-compare" type: the signal is compared to a delayed version of itself, and the first match gives the period.

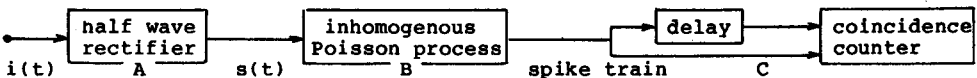
1) Auto-correlation is familiar. Short-term ACF can be defined as:

$$acf(L) = 1/T \int_0^T s(t)s(t-L)dt$$

2) AMDF (average magnitude difference function) can be similarly defined:

$$amdf(L) = 1/T \int_0^T |s(t) - s(t-L)| dt$$

3) The third model is physiologically more realistic:



A: Input $i(t)$ represents motion at a point along the basilar membrane (displacement or velocity). Only positive excursions are effective.

B: Instantaneous nerve-fiber discharge probability is modeled as:

$$p(t) = s(t) \cdot r(t-t_0),$$

where t_0 is the time of occurrence of a previous spike. The recovery function $r(t)$, representing refractoriness, is zero until 0.7ms and then rises linearly from 0.5 to 1.0 at $t = 5.0$ ms (Johnson and Swami 1983).

C: An auto-coincidence histogram is calculated by summing the occurrences of inter-spike intervals, regardless of whether the spikes are consecutive or not (the more common ISI histogram only sums intervals between consecutive spikes). Ruggero (1973) and Evans (1983) published AC histograms for nerve-discharge data, Licklider (1959) suggested auto-coincidence as a pitch perception mechanism, and Lyon (1984) uses it in his computational model.

In a complete model, different points along the basilar membrane would feed an array of such processors and the output pattern would be scanned for cues to pitch (for example a peak spanning all channels). Here, we consider individual channels.

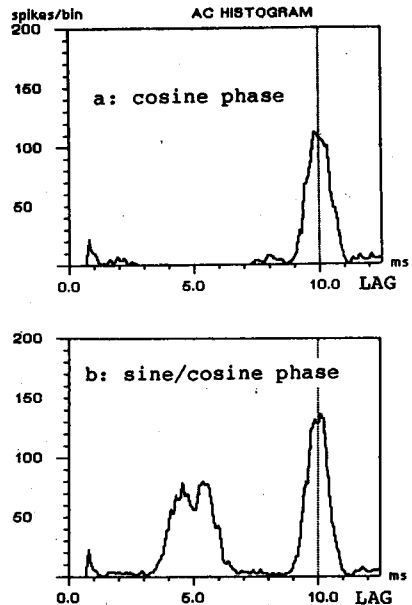


Fig. 1: AC histogram for a 6-harmonic complex

* "Phase sensitivity and pitch shifts: the behaviour of three temporal pitch models" Alain de Cheveigne (Kyoto University and LIMSI, Paris-Sud-Orsay).

II Is the low pitch estimate phase sensitive?

Figure 1 shows the output of model 3 driven by a sum of harmonics 1 to 6, in cosine phase (a) or alternate sine-cosine phase (b). The major maximum in the AC histogram remains clearly at a lag equal to the period (10ms). The same is of course true of the ACF, and of the AMDF. This behaviour (a priori evident) was verified for a range of complex signals typical of psycho-acoustical stimuli. It is thus incorrect to say that all temporal pitch models are phase sensitive.

III Are the models "phase-deaf"?

The ear is not quite phase-deaf. Beats of mistuned consonances are one example of phase sensitivity (Plomp 1967). If we present the ear with a sum of two sine waves, at $n.f_0$ and $m.f_0 + df$ (with m, n integers, $m > n$), under certain conditions a beating sensation is heard, at a rate of $n.df$ beats per second. This is equivalent to slow variations in the relative phases of a $m.f_0 + n.f_0$ complex. The amplitude ratio for best beats suggests that they occur when both partials have a similar level at the place (along the basilar membrane) of the higher partial.

For a 1:3 complex, Plomp (1967) reported in addition a strange phenomenon: the pitch of the higher partial seemed to beat.

In figure 2, the ACF of the rectified signal (similar to model 3 for low frequencies) has been plotted for a 100Hz + 300Hz complex, for various phases. It can be seen that the minor peak near 3.3ms moves sideways in addition to vertically. If this peak were the cue to the pitch of the higher partial, this shift with phase would account for the pitch beat. Figure 3 shows this cue (in terms of frequency ratio to the higher partial) as a function of phase, for various amplitude ratios. The peak disappears below -4dB, so the maximum excursion is about 1.3. For a higher partial of 600Hz, this would lead to a pitch between 600 and 780Hz. Plomp (1967) obtained matches between 600 and 750Hz, so the agreement is good.

We can go one step further by considering the patterns of vibration of the partials on the basilar membrane. If the 600Hz pattern is sharp enough, at a certain level ratio, beats would originate from two regions: a) the peak of the 600Hz pattern, b) the tail of the 600Hz pattern. In the first region, where $a_3 > a_1$, the pitch would remain at about 610Hz whatever the phase. In the second (broader) region where $a_3 < a_1$, a weaker pitch would arise, oscillating widely with phase. This would lead to a pitch ambiguity similar to that seen in figure 5 of Plomp (1967).

In conclusion, all three models examined provide phase-insensitive lower pitch estimates. In addition the third model displays, in at least one situation, a phase sensitivity quite similar to one observed in hearing.

Johnson, D.H. (1983), JASA 74, 493-501.

Ruggero, M.A. (1973) J. Neurophysiol. 36, 569-587.

Licklider, J.C.R. (1959) "Three auditory theories" in "Psychology, the study of a science", S. Koch (ed), McGraw-Hill, NY, 41-144.

Evans, E.F. (1983) in "Hearing-Physiological bases and psycho-physics" R.

Klink, H. Hartman (ed) Springer-Verlag, 140-146.

Plomp, R. (1967) "Beats of mistuned consonances" JASA 42, 462-474.

Fig. 2: ACF of rectified stimulus for 100Hz + 300Hz:

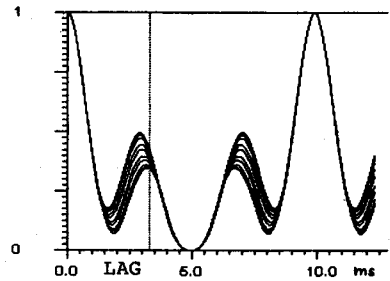


Fig. 3: Lag at minor peak (inverse ratio to 3rd harmonic period):

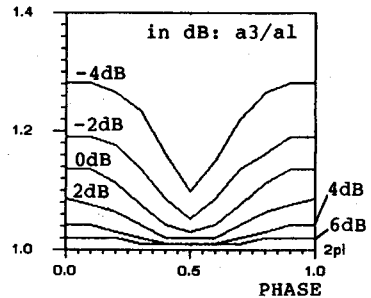


Fig. 4: ACF at minor peak (ratio to ACF at zero lag):

