PITCH AND THE NARROWED AUTOCOINCIDENCE HISTOGRAM.

Alain de Cheveigné

ATR Interpreting Telephony Research Laboratories,
Sanpeidani, Inuidani, Seika-cho, Soraku-gun, Kyoto-fu, 619-02, Japan.
(also DRL, CNRS/Université Paris-7, 75251 Paris).

This paper is structured in three parts. In the first is outlined a model of pitch perception based on the autocoincidence histogram of auditory nerve fiber spike patterns. The second shows that the basic AC model is insufficient to account for the discrimination thresholds of pure tones, and introduces the "narrowed" autocoincidence (NAC) histogram. The third explores the NAC model in relation to complex tone pitch .

THE AUTOCOINCIDENCE HISTOGRAM PITCH MODEL

The AC model supposes that the discharge pattern for each fiber or group of fibers is processed to form an autocoincidence histogram (also called autocorrelogram or autocorrelation histogram). The AC histogram counts intervals between spikes, *consecutive or not*, and differs thus from the more familiar ISI (interspike interval) that counts intervals between *consecutive* spikes only. For a periodic stimulus, the AC histogram shows a clear peak at the period and its multiples (fig. 1).

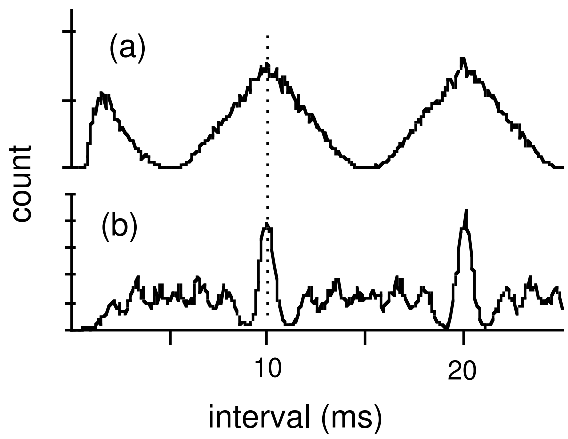

Fig. 1: AC histograms for sine tone (a) and complex tone with harmonics 3 to 6 (b). "Spike" trains are produced by a model (inhomogenous Poisson process with refractoriness and gaussian jitter). See Evans (1986) for examples of AC histograms from real nerve-fiber data.

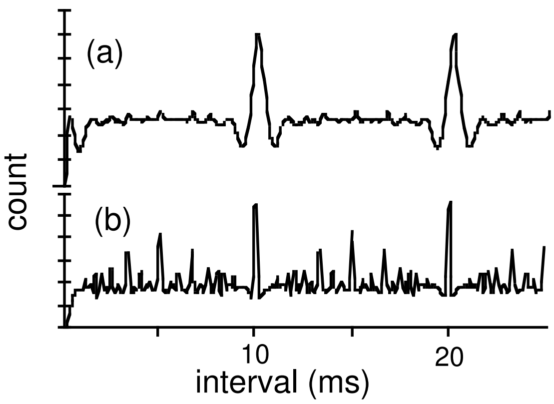

Fig. 2: NAC histograms for pure tone (a) and complex tone (b). These histograms were calculated from the same model "spike" data as in fig. 1 (a) and (b) respectively. Narrowing order is 10.

Physiologically, the AC histogram could be calculated by a neural network comprising delay-lines and coincidence summing neurons, similar to that invoked for sound localization (Yin and Chan 1988). The fibers from the cochlea would feed an array of histograms, forming a pattern of activity over two dimensions: lag and tono–topy. For a periodic stimulus, the pattern would show a ridge spanning the tonotopic dimension at a lag equal to the period (whether or not components are resolved).

For some stimuli the ridge may be unambiguous. For others (filtered noise, inharmonic stimuli, etc.), the ridge may be blunt or shallow, or its position may not be quite the same for different channels, or there may be several competing ridges at different lags. According to whether we require the pitch cue to be a single clear ridge, or allow it to be local to a tonotopic or lag region, we can choose between a precise but rigid model, or one that is more vague but that can account for imprecise, weak or ambiguous pitch, and the effects of attention. The underlying mechanism is the same.

The autocoincidence model accounts nicely for a wide range of pitch phenomena, including effects often opposed to temporal models (such as phase effects). It was originally proposed by Licklider (1959), but it has received relatively little attention since then. Moore (1982) and van Noorden (1982) proposed similar models based on the ISI histogram. The ISI histogram is however less realistic than the AC histogram, and breaks down at low frequencies (de Cheveigné 1989).

PURE TONE PITCH JNDS ~ THE NARROWED AUTOCOINCIDENCE HISTOGRAM

The AC model accounts well for many aspects of pitch. However, it is unsatisfactory in at least one respect: the peak corresponding to the period in the AC histogram for a pure tone (fig. 1a) is rather wide. How can one account for the precision with which pure tone frequency can be discriminated ($\Delta f/f \approx 0.2\%$) ?

The cue to pitch is the position of the histogram peak, which is uncertain because of the statistical nature of the histogram. It can be shown (de Cheveigné 1989) that the standard deviation σ_T of the peak position is given by:

$$\sigma_T \approx 0.12 \ R^{-1/2}(D\varepsilon)^{-1/4} \qquad (1)$$

where R is the discharge rate in spikes/s, D the stimulus duration (in s), and ε the bin width of the histogram (the coefficient was obtained by simulation). The predicted just-noticeable difference (jnd) of a pure tone as a function of duration (assuming that spikes from all fibers are pooled before the histogram calculation) is plotted in fig. 3, along with discrimination data replotted from Moore (1973). The first thing to remark is that the predicted jnd is up to an order of magnitude too large. The second is that it varies as $D^{-1/4}$, whereas Moore's data is better described by D^{-1} at low frequencies. There is thus a discrepancy between the model's predictions and psychophysical data.

The basic AC histogram is similar in shape to the autocorrelation of the rectified basilar membrane motion. Recently, Brown and Puckette (1989) proposed a method for improving the resolution of the autocorrelation function, for musical pitch extraction. The "narrowed" autocorrelation function $g_N(\tau)$ can be defined as:

$$g_N(\tau) = \sum_{k=1}^{N-1} (N-k)\, g(k\tau) \qquad (2)$$

where $g(\tau)$ is the ordinary autocorrelation, and N a narrowing factor.

It turns out that one can similarly construct a narrowed *autocoincidence histogram*, by counting, for every lag, the spike intervals that are its *multiples*, and summing up the counts using the same coefficients as in eq. 2. Physiologically, this result could be obtained with a set of summing neurons, each one connected to equally spaced outputs of the AC network mentioned above.

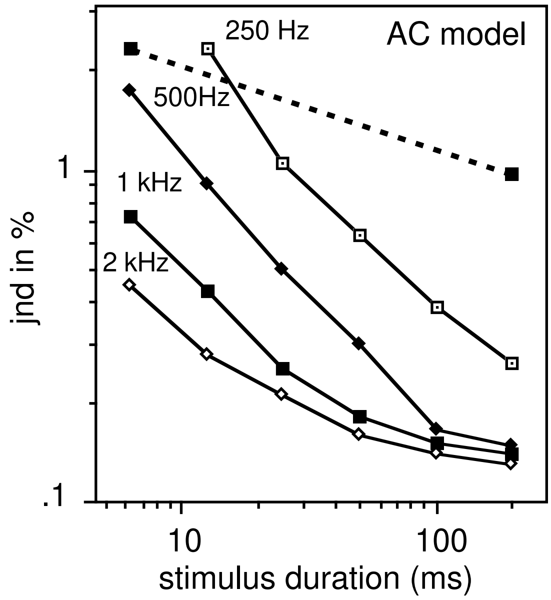

Fig. 3: Dotted-line: just-noticeable difference in pure tone frequency predicted by the AC model. Continuous lines: discrimination data replotted from Moore (1973). Data for frequencies above 2 kHz are not included.

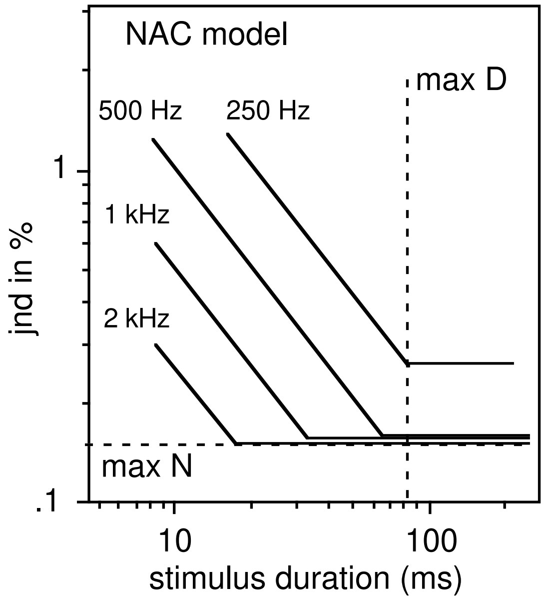

Fig. 4: Just noticeable difference in pure tone frequency predicted by the NAC model. Dotted lines represent hypothetical limits to performance: complexity (N) and delay line length (D) (shown here with arbitrary values).

Fig. 2a shows an NAC histogram in response to a pure tone, with N = 10. For large N, the width of the peak varies as N^{-1}, and so it can be made arbitrarily narrow. For time-limited stimuli, however, there is a limit to the length of intervals and therefore a limit to the allowable narrowing factor. This implies a $(\text{duration})^{-1}$ dependency of resolution, which is precisely the trend of Moore's data. Fig. 4 shows how this model predicts the jnd of pure tones as a function of duration. One can see that the slope and spacing of the curves are remarkably similar to Moore's data at low frequencies and short durations (fig. 3).

For stimuli that are not time-limited, one can expect two other factors to intervene to limit performance: maximum delay-line length (D), and maximum complexity (N) of the narrowing circuit. There is no kink in the 250 Hz curve in fig. 3, so it seems that duration is not a limiting factor in Moore's data. The dependency on duration in his data is overall flatter than predicted by the model, particularly at 2 kHz and higher (not plotted). A possible explanation is that jitter occurs in the conduction velocity of delay-line nerve fibers. This would introduce an error in the interval estimation proportional to $(\text{interval})^{1/2}$, thereby reducing the contribution of longer intervals to narrowing, especially at high frequencies.

In simulations, the narrowing mechanism loses its effectiveness above 2-3 kHz; above 5 kHz the AC histogram itself becomes flat.

THE NAC MODEL AND COMPLEX STIMULI.

The basic AC model is adequate to explain many aspects of pitch perception without the need for narrowing. Narrowing is best seen as a sort of post-processing, introduced to account for discrimination of pure tones. However it is of interest to see how the mechanism behaves in response to complex tones.

Fig. 2b shows an NAC histogram (order 10) for a complex stimulus. The histogram contains secondary peaks situated at lag values of the form n/mT, where n and m are integers (m = 1,...,N, and n arbitrary). This is understandable when one considers that the NAC is constructed by a superposition of AC histograms shrunk along the lag axis (for pure tones these peaks are not visible because the AC peaks are wider and tend to cancel each other when the histograms are added).

This peculiar pattern implies that two complex stimuli with fundamental frequencies in a simple ratio give rise to NAC histograms with many peaks in correspondence. The closest match is for the octave, followed by the fifth and other ratios, which could possibly explain the special importance of integer ratios in music and pschophysics. Such an explanation resembles in form the classical one involving a correspondence between harmonics, however here we do not require two stimuli to share common components. The only requirement is that the peaks in the AC histogram be sharp enough so that they stand out in the NAC histogram. This is sometimes the case for pure tones, when firing is localized within the stimulus cycle (Ruggero and Rich 1983).

CONCLUSION

The AC histogram and its extension, the NAC histogram, provide a basis for pitch perception that is physiologically realistic, sufficiently rich and flexible to accommodate a wide range of phenomena, and yet precise enough to be confronted with quantitative data.

ACKNOWLEDGEMENTS

The author is supported by the STP program of the Commission of the European Communities. He wishes to thank ATR for its hospitality, and the CNRS and Linguistics Department of Paris-7 University for leave of absence.

REFERENCES

Brown, J.C. and Puckette, M.S. (1989) "Calculation of a "narrowed" autocorrelation function", J. Acoust. Soc. Am. 85, 1595-1601.

de Cheveigné, A. (1986) "A pitch perception model", Proceedings IEEE ICASSP-86, 897-900.

de Cheveigné, A. (1989) "Narrowed auto–coincidence of auditory nerve spike trains as a basis for pitch perception", submitted for publication.

Evans, E.F. (1986) "Cochlear nerve fibre temporal discharge patterns, cochlear frequency selectivity and the dominant region for pitch", in "Auditory frequency analysis", B.C.J. Moore, R.D. Patterson (eds), 253-260.

Licklider, J.C.R. (1959) "Three auditory theories" in "Psychology, a study of a science", vol. 1, S. Koch (ed), McGraw-Hill, 41-144.

Moore, B.C.J. (1973), "Frequency difference limens for short-duration tones", J. Acoust. Soc. Am. 54, 610-619.

Moore, B.C.J. (1982) "An introduction to the psychology of hearing", Academic Press.

van Noorden, L. (1982) "Two channel pitch perception", in "Music, mind, and brain", M. Clynes (ed), Plenum press, 217-225.

Ruggero, M.A. and Rich, N.C. (1983) "Chinchilla auditory-nerve responses to low-frequency tones", J. Acoust. Soc. Am. 73, 2096-2018.

Yin, T.C.T., Chan, J.C.K. (1988) "Neural mechanisms underlying interaural time sensitivity to tones and noise" in "Auditory Function - Neurological bases of hearing", G.E. Edelman, W.E. Gall, W.M. Cowan (eds), Wiley, 385-430.