A MODEL FOR THE DISCRIMINATION OF PURE TONE PITCH.

Alain de Cheveigné

Laboratoire de Linguistique Formelle, CNRS - Université Paris 7, France.

ABSTRACT

This paper presents a model of auditory processing that can account for the very small frequency difference limens observed psychophysically for pure tones. In a first step, an autocoincidence histogram is calculated from nerve-fiber channels synchronized to the pure tone, according to a model similar to that of Licklider [3, 4, 5]. In a second step, this histogram is "folded", resulting in a "narrowed autocoincidence histogram". The peak of this narrowed histogram is sharper than that of the autocoincidence histogram, and its width depends on stimulus duration in a way similar to frequency difference limens.

1. INTRODUCTION

Listeners can discriminate differences in the frequency of pure tones as small as 0.2% [1]. Thresholds get larger as stimuli get shorter, but discrimination remains good even when the stimuli contain only a few cycles. Moore [1] argued that the thresholds are too low to be compatible with a place mechanism of frequency discrimination based on the differences in intensity that might arise when the excitation pattern for a tone is shifted along the basilar membrane. They would be compatible, on the other hand, with a time domain mechanism. Based on this assumption, Goldstein and Srulovicz [2] proposed a theory that predicts thresholds under the hypothesis of optimum processing of interspike intervals. Goldstein and Srulovicz noted that information from as few as *nine* fibers is sufficient to account for discrimination thresholds. Since many more fibers are available for processing, performance must have other limits, perhaps due to the actual neural processing mechanism. The question arises as to whether such processing has the same behavior as optimum processing. It is therefore of interest to examine candidate processing models with respect to pure tone frequency discrimination. One such model is that of Licklider [3, 4, 5], based on the autocoincidence of nerve fiber discharges (see also [6, 7, 8, 9, 10]). If we assume this particular model, can we still predict discrimination thresholds?

In this study it is found that a) the basic autocoincidence mechanism of Licklider's model does not adequately predict performance, but b) it can be followed by a second stage of processing, described by a "narrowed autocoincidence histogram" (NAC), to form a model that predicts thresholds similar to those observed psychophysically.

2. DISCRIMINATION THRESHOLDS FOR PURE TONE PITCH

Moore [1] measured frequency difference limens for pure tones as a function of frequency and stimulus duration. His data are plotted in Fig. 1. At all frequencies, thresholds tend to be smaller for longer stimuli. Discrimination gets better as frequency increases, up to 2 kHz. For the lowest three frequencies there is a zone of durations for which threshold varies approximately as the *inverse of stimulus duration*. These frequencies are in the region for which a time-domain frequency analysis mechanism such as Licklider's is in principle applicable.

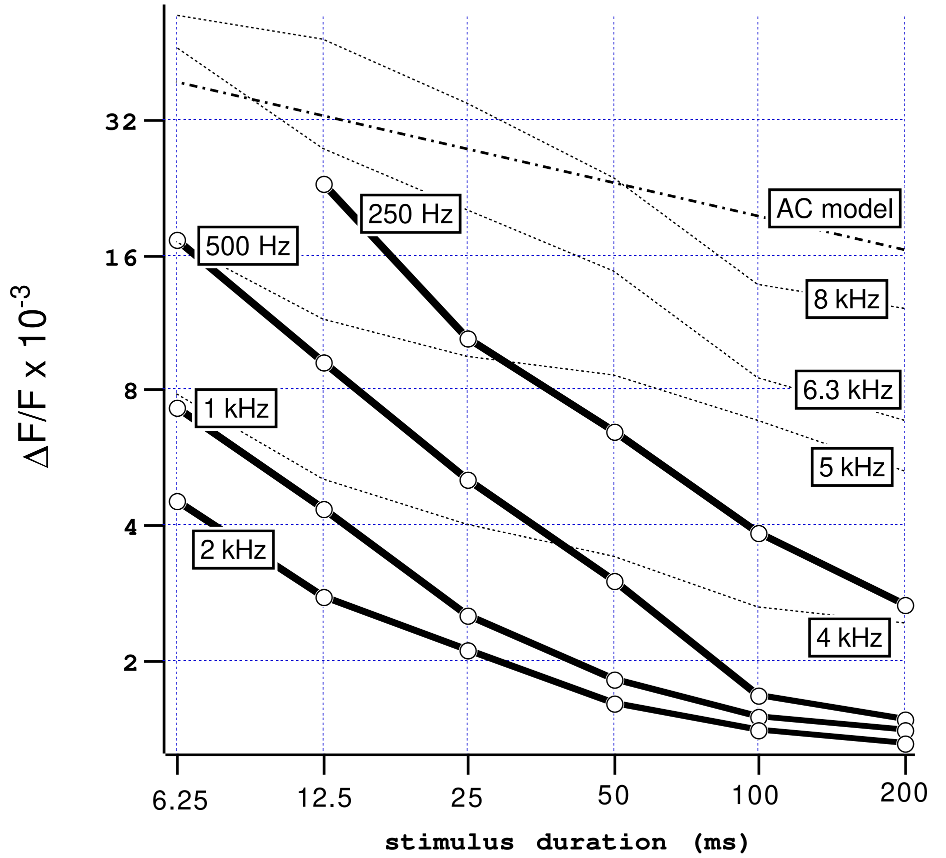

Fig. 1. Frequency difference limens ($\Delta F/F$) for pure tones as a function of stimulus duration and frequency (replotted from Moore [1]). Frequencies up to 2 kHz are plotted with continuous lines, higher frequencies with dotted lines. Straight line: difference limen predicted by basic autocoincidence model.

3. AUTOCOINCIDENCE MODEL

In Licklider's model [3, 4, 5], patterns of discharge within auditory-nerve fibers are processed in the auditory nervous system by a neural network that calculates the equivalent of an autocoincidence (or autocorrelation) histogram [11, 12]. The result is a pattern of activity over the two dimensions of *frequency* (inherited from peripheral filtering) and *lag* (provided by nerve conduction or synaptic delays). In response to a periodic stimulus, this pattern shows a ridge at a lag equal to the period, thus providing a cue to the pitch. Licklider's model was designed to explain the pitch of complex stimuli, however it works as well for pure tones. In response to a pure tone of frequency f, nerve fibers with characteristic frequencies within a band surrounding f will respond with a periodicity of 1/f. The result is an autocoincidence pattern with a ridge at 1/f. Actually, the pattern also shows ridges at period multiples; the model supposes that the position of the *first* ridge is the cue to pitch. Because synchronization deteriorates above 2-5 kHz, the model can only apply to frequencies below that limit (this excludes the upper 2 or 3 octaves of the 10 that span the audible range).

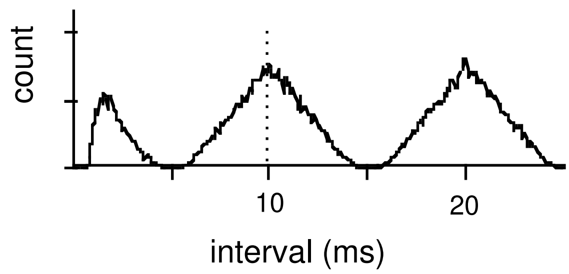

Fig. 2. Autocoincidence histogram in response to a pure tone of 100 Hz. The dotted line marks the period lag. The histogram was calculated using "spike" data produced by a model [13].

Let us define the pitch cue more precisely as the position of the *maximum* of a composite pattern obtained by taking the sum of histograms across frequency channels (alternative assumptions are possible but won't be discussed here). In response to a pure tone the histograms are all identical, so the effect of summing them is simply to reduce variability, as if a single histogram were calculated with more spikes. How precise is this cue?

As evident in Fig. 2, the histogram is "noisy", which causes the position of the maximum to be uncertain. The standard deviation of this position can be estimated [14] as a function of discharge rate R, stimulus duration D, histogram bin width ε, and number N of histograms summed together:

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

It is evident from (1) that the standard deviation varies as the inverse of the *fourth root* of stimulus duration. This dependency can be understood as follows: due to the parabolic shape of the AC histogram near its peak, the incertitude of the position of the maximum varies with the *square root* of the standard deviation of the bin "noise", itself proportional to the square root of the counts in the histogram bins. If spikes are allowed to accumulate during the entire stimulus presentation, the count within each bin is proportional to duration, hence the $D^{-1/4}$ dependency.

To get a more quantitative estimate, let us make the assumption that 1250 fibers respond each at 100 s/s, that the spike trains are pooled before histogram calculation into 10 histograms that are then summed, and that histogram resolution is 1µs (R = 12500, N= 10, $\varepsilon = 10^{-6}$). Given these assumptions, the difference limen $\Delta F/F$ (supposed equal to σ_T) varies as plotted in Figure 1. We can draw the following conclusions:

a) The dependency of $\Delta F/F$ on duration, predicted by the model as $D^{-1/4}$, does not match that observed in Moore's data at low frequencies.

b) The $\Delta F/F$ predicted by the model is almost an order of magnitude larger than the best difference limens observed.

4. NARROWED AUTOCOINCIDENCE MODEL

In the AC model, the effect of making the stimulus longer is to make more spikes available, thus reducing statistical uncertainty. Clearly this is insufficient to account for the difference limens observed and their dependency on duration. There is however a source of information that the AC model neglects: that carried by the peaks of *higher rank* of the autocoincidence histogram.

Recently, a method has been proposed for sharpening the peaks of the autocorrelation *function* (for purposes of musical pitch estimation) [15]. This method incorporates information from higher-order peaks into a compact representation called "Narrowed autocorrelation function". A similar operation can be applied to the autocoincidence histogram (AC), resulting in a "Narrowed autocoincidence histogram" (NAC):

$$NAC(\tau) = \sum_{k=1}^{N-1} (N-k)\ AC(k\tau) \quad (2)$$

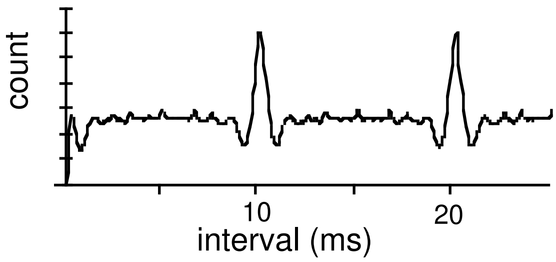

Fig. 3. Narrowed autocoincidence histogram in response to a pure tone of 100 Hz. Order of narrowing is 10.

It can be seen from Figure 3 that the period peak of the NAC is narrower than that of the AC histogram. Peak width is inversely proportional to the narrowing order N. The practical value of N is limited by the duration of the stimulus, since it is impossible to calculate an AC histogram for intervals greater than the stimulus duration. If this is the factor that limits frequency discrimination, then difference limens should vary as D^{-1}. (Whereas probabilistic factors determined the thresholds of the AC model, these factors are considered negligeable in the analysis of the NAC).

Fig. 4 displays $\Delta F/F$, under the further assumptions that the width of an AC peak before narrowing is about 10 %, and that the only effect of frequency is to vary the number of cycles within a stimulus. The effect of frequency is difficult to analyze in this model, because it affects the population of fibers that respond and their degree of synchronization, as well as the number of AC histograms peaks that can fit within a given duration.

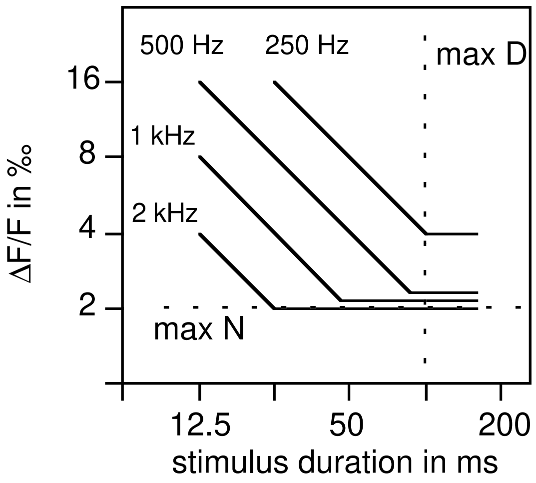

Fig. 4. Difference limens predicted by the NAC model.

The dotted lines labeled "max D" and "max N" in Fig. 4 represent additional hypothetical limits on discrimination due to two factors. The first factor limits length of AC histograms (it could be for example a limit on the allowable length of the neural delay lines assumed by Licklider's model). Making stimuli longer than this limit can bring no improvement. The second factor would limit the order of narrowing, due to the complexity of neural circuitry available for the calculation of the NAC.

The trend of the thresholds visible in Fig. 4 is similar to that of Moore's data for frequencies below 2 kHz (Fig. 1). The major differences are that the curves in Fig. 1 are somewhat shallower, and the spacing smaller than predicted by the model. There is also no evidence in Moore's data for the first of the hypothetical limits mentioned above ("max D") . Apart from these differences the agreement is quite good.

CONCLUSION

The basic autocorrelation model due to Licklider is not sufficient to account for frequency difference limens observed psychophysically. However, a modified model (the NAC model) can successfully account for these limens, and for the form of their dependency on duration. This result is of interest given the recent renewed concern for time-domain models of auditory processing.

ACKNOWLEDGEMENTS

Part of this work was carried out at ATR Interpreting Telephony Research Laboratories, under a fellowship awarded by the European Communities STP programme in Japan. The author wishes to thank ATR for its hospitality, and the CNRS for leave of absence.

REFERENCES

[1.] Moore, B. C. J. (1973), "Frequency difference limens for short-duration tones", JASA. *54*, 610-619.

[2.] Goldstein, J. L. and P. Srulovicz (1977), "Auditory-nerve spike intervals as an adequate basis for aural frequency measurement", *Psychophysics and physiology of hearing*, Evans and Wilson ed. Academic Press: London, 337-345.

[3.] Licklider, J. C. R. (1956), "Auditory frequency analysis", *Information theory*, Cherry ed. Butterworth: London, 253-268.

[4.] Licklider, J. C. R. (1959), "Three auditory theories", *Psychology, a study of a science*, Koch ed. McGraw-Hill: 41-144.

[5.] Licklider, J. C. R. (1962), "Periodicity pitch and related auditory process models", International Audiology. *1*, 11-36.

[6.] de Cheveigné, A. (1986), "A pitch perception model", *Proc. IEEE ICASSP*, 897-900.

[7.] Lyon, R. (1984), "Computational models of neural auditory processing", *IEEE ICASSP*, 36.1.(1-4).

[8.] Meddis, R. and M. Hewitt (1988), "A computational model of low pitch judgement", *Basic issues in hearing*, Duifuis, Horst and Witt ed. Academic: London, 148-153.

[9.] Moore, B. C. J. (1982), *An introduction to the psychology of hearing*, Academic Press: London.

[10.] van Noorden, L. (1982), "Two channel pitch perception", *Music, mind, and brain*, Clynes ed. Plenum Press: London, 251-269.

[11.] Ruggero, M. A. (1973), "Response to noise of auditory nerve fibers in the squirrel monkey", J. Neurophysiol. *36*, 569-587.

[12.] Evans, E. F. (1983), "Pitch and cochlear nerve fiber discharge patterns", *Auditory frequency analysis*, Moore and Patterson ed. Plenum Press: 253-264.

[13.] de Cheveigné, A. (1990), "Auditory nerve fiber spike generation model.", ATR technical report TR-I-104, 15p.

[14.] de Cheveigné, A. (1989), "The narrowed autocoincidence histogram and pure tone pitch.", submitted for publication.

[15.] Brown, J. C. and M. S. Puckette (1989), "Calculation of a "narrowed" autocorrelation function", JASA. *85*, 1595-1601.