

Internal Use Only

TR-I-0139

Auditory nerve fiber spike generation model

Alain de Cheveigne

(ATR自動翻訳電話研究所)
ATR Interpreting Telephony Labs.

1990.2

ATR 自動翻訳電話研究所
ATR Interpreting Telephony Research Laboratories

©ATR 自動翻訳電話研究所
©ATR Interpreting Telephony Research Laboratories

Auditory nerve fiber spike generation model.

Contents:

Introduction.....	1
The spike generation model.....	2
• Spike generation process.....	2
• Justification of the model's basic assumptions:	2
• Parameters	2
Computer implementation.....	4
• Spike generation.....	4
• Histogram calculation.....	5
Applications.	6
Limitations of the model.	6
1) Structure.	6
2) Spike generation as distinct from adaptation.	7
3) Linearity.	7
4) Half-wave rectification.....	7
5) Refractory effects.....	7
6) Jitter.	8
Conclusion.....	8
Acknowledgements.....	8
References	8
Appendix: histogram examples.....	11

Auditory nerve fiber spike generation model.

Introduction.

Speech recognition systems are far from reproducing the capabilities of human listeners. Recognition rates remain low even in the best recording conditions, and deteriorate even further when the conditions approach those of everyday use: noise, reverberation, interfering voices, etc. In contrast, people can hear and understand speech in very adverse conditions.

To a certain degree, we can parse the complex spectrum of superimposed speech and noise into separate "streams", and concentrate on the speech while eliminating the noise. We can also "reconstruct" portions of speech that have been obliterated. Such capabilities likely involve a combination of *peripheral* processing (under central control via efferent nerve pathways), and *central* processing of auditory nerve fiber patterns.

Physiological experiments (recordings of discharge patterns within the auditory nerve) provide a picture of the data that central processing operate on, and are a starting point for models of central processing. Despite the large amount of published data available, questions may arise for which this data is insufficient.

A model of auditory nerve fiber spike generation is useful within this context, because it allows us to:

- a) experiment with models of higher-level auditory processing,
- b) extrapolate from existing experimental data to new stimulus conditions, without carrying out physiological experiments,
- c) clarify the relation between the shape of the driving function that controls nerve fiber discharge, and the shape of histograms as they can be measured from single units of the auditory nerve,
- d) assess the degree to which experimental data has been affected by the choice of histogram format used to report it (different histograms are not equivalent).

The model described here can be seen as one element of a more comprehensive model of auditory processing:

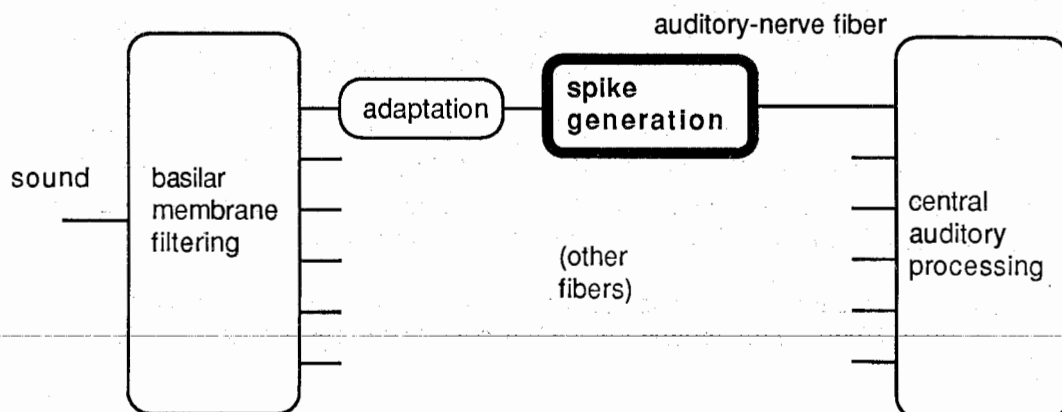


Fig. 1: Spike generation model within a model of auditory processing.

The validity of this model is discussed further on.

The model.

1) Spike generation process

This model is designed to produce spike trains as similar as possible to those recorded from first-order auditory-nerve fibers. It comprises two steps: a point process, followed by a gaussian jitter.

The point process is an inhomogenous Poisson process with refractory effects. The probability of firing between t and $t+dt$ is:

$$r(t) dt = \lambda s(t) \rho(t-t') dt$$

where λ is maximum process rate, $s(t)$ is a driving function normalized to $[0,1]$ and proportional in a first approximation to the half-wave rectified basilar membrane motion at the point of innervation of the fiber being modeled. The function $\rho(t-t')$ is a recovery function that depends only on the time since the previous spike.

Apart from the choice of parameters, the model is identical to the digital simulation model of Schroeder et al. Johnson and Swami (1983) use a similar model to account for the pattern of discharge in physiological data, and Siebert (1970), and Sruлович and Goldstein (1983) use a similar model to account for psychophysical discrimination performance.

The "spike" train produced by this process is submitted to a "jitter" process: the position of each spike is shifted by an amount that follows a zero-mean gaussian distribution.

2) Justification of the model's basic assumptions:

The approximate proportionality of auditory nerve fiber discharge probability to the half-wave rectified, linearly filtered stimulus (basilar membrane motion) has been reported by many authors (Rose et al. 1969, Brugge et al. 1969, de Boer and de Jongh 1978, Evans 1986, Carney and Yin 1988).

For constant driving function (spontaneous or high-frequency driven activity), discharge probability obeys a Poisson process with absolute and refractory effects. This was verified by Kiang (1965), Gray (1967) and Gaumond et al. (1982).

Discharge probability is independent of discharge history before the previous spike as verified by Rodieck et al. (1962), Gray (1962) and Kiang (1965). Refractory effects are represented by a multiplicative factor (ρ) that depends only on the time elapsed since the previous discharge. This was verified by Gaumond et al. (1982, 1983).

For constant input, the model describe the activity of a single fiber as being independent from that of other fibers. This independence has been tested by Johnson and Kiang (1976).

3) Parameters

The maximum process rate λ depends on stimulus level, and controls, with other factors, the average discharge rate. In the literature, the instantaneous discharge rate as seen from peaks in PSTs range as high as 2000 spikes/s (Delgutte and Kiang, 1984b), but the average rate is always lower because of a combination of refractoriness, adaptation, and driving function shape factor. Typical average discharge rates at saturation for sustained sounds are 150-200 spikes/s, with some fibers reaching 300 spikes/s (Evans, 1978; Liberman, 1978; Ruggero, 1973). We avoid using values of λ that produce rates above these maxima.

We examined various sources of data to determine the shape of the refractory function. Unfortunately there is little data offering high resolution at short intervals, where effects are largest. The absolute refractory period in the literature ranges from 0.5 ms (Rose et al., 1967) to 1.3 ms (Javel, 1980). We settled on a value of 0.8 ms. The shape of the recovery function is similar to Schroeder and Hall (1974), with a time constant of 2 ms:

$$\begin{aligned} \rho(t) &= 0 && \text{for } t < 0.8 \text{ ms,} \\ \rho(t) &= 1.14 (1 - \exp(-(t - 0.8)/2)) && \text{for } 0.8 \text{ ms} < t < 5 \text{ ms,} \\ \rho(t) &= 1 && \text{for } t > 5 \text{ ms.} \end{aligned}$$

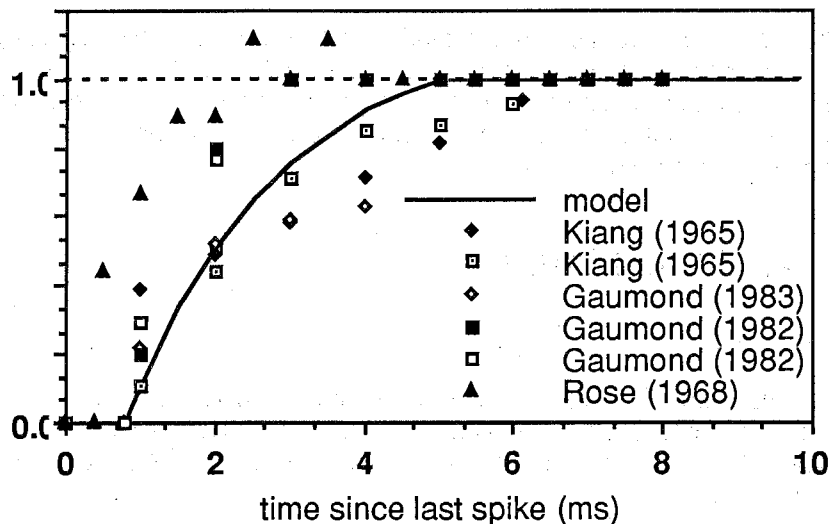


Fig. 2: Recovery function used in the model, along with estimates of recovery function values taken from the literature.

The gaussian jitter was chosen with a standard deviation of 55 μ s, to simulate the fall-off of synchrony between 2 and 6 KHz described by Johnson (1980). It is similar to the 40 μ s value proposed by Goldstein and Sruлович (1977).

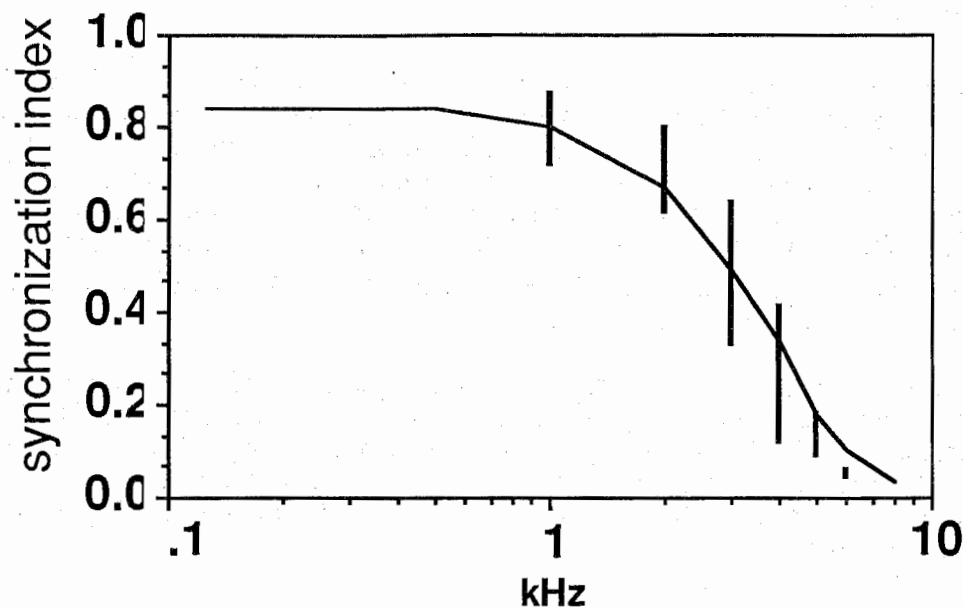


Fig. 3: Synchronization index measured from period histograms obtained with the model (continuous line). Vertical lines indicate the range of values measured by Johnson (1980).

Jitter appears to account for synchrony drop-off quite well, but the real cause may be elsewhere, such as in the decline in receptor potential due to hair cell membrane capacity (Russel and Sellick, 1980).

When using the model to generate data for models of auditory processing, it is good to keep in mind that part of the jitter in recorded data may be due to the recording process itself, so the data available to central processes might be somewhat cleaner. However, Johnson (1978) showed that noise introduced in recording cannot be the sole source of synchrony drop-off at high frequencies.

Computer implementation.

1) Spike generation

The principle of simulation is simple. Time is divided into discrete steps of length dt . At each step, a random number in $[0,1]$ is drawn and compared to $r(t)dt$. If the random number falls below that value, a spike is "fired" and the *interval* since the previous spike is output as data.

Performed exactly as described, the computation would be expensive, especially at low rates and small sampling step. Actually, we can remark that the occurrences of trials smaller than the constant λdt (maximum probability) follow a normal Poisson law. We can simulate the intervals between these occurrences separately by drawing intervals following a Poisson law. The actual firings are selected among these occurrences by comparing a random number to the value $s(t)p(t-t')$. The gain in efficiency of random number use is in a factor $1/(2\lambda dt)$.

The gaussian distribution of the jitter was simulated by summing 12 random bytes with uniform distribution.

Programs were written in C on a Vax 8800 running a version of UNIX 4.3 BSD. The random numbers were produced by the `random()` subroutine of the UNIX C library (period greater than 2^{69}). Before each series of calls to `random`, the generator was initialized with a call to `srandom()` with an argument taken from the system clock.

"Spike" interval data is stored on disk. The first value corresponds to the time elapsed between a time reference (relative to the driving function) and the first spike, rather than an interspike interval.

2) Histogram calculation.

The output of the model can be used to calculate histograms, for comparison with published physiological data. Histograms are calculated as defined in the literature.

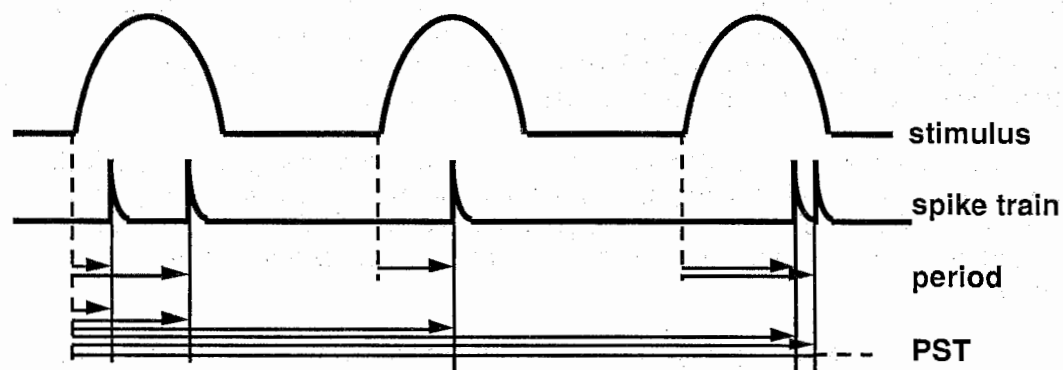


Fig. 4: Intervals accumulated in PST and period histograms.

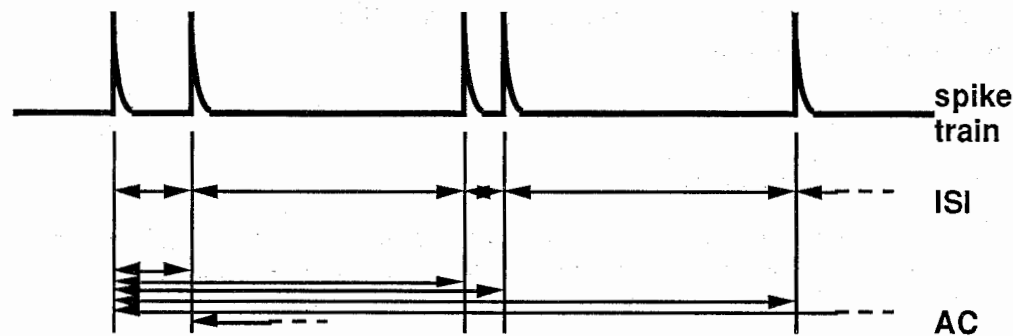


Fig. 5: Intervals accumulated in ISI and AC histograms.

ISI (interspike) histograms are calculated by simply accumulating intervals in bins according to length.

AC (autocoincidence) histograms are a little more difficult to calculate: a population of higher-order intervals must be maintained at each step. This population is implemented as a linked list. Each time a new interval is processed, it is added to the list, and the oldest interval is removed (unless the total cumulated interval would fall shorter than the histogram size). All cumulated intervals containing the new interval are calculated, and the bins corresponding to their length incremented.

Period or PST (peri-stimulus time) are calculated by maintaining a cumulative sum of all intervals. Bins corresponding to this value modulo the "stimulus" period are incremented.

Examples of histograms produced by the model are given in the appendix.

Applications.

Description of applications is outside the scope of this report. This is an outline a few themes:

- The model allows a better understanding of the relationship between the different histogram formats used to report data in the literature. Different histogram formats are not equivalent, and the choice of a given format can affect interpretation of the data. A striking example can be found in the controversy that opposed Whifield (1978,1979,1980) to Evans (1978,1983,1986) and Moore (1977,1980) on the interpretation of period histograms in response to inharmonic complexes.
- The model also allows us to understand how the transduction process can modify the effect of previous stages of processing. A possible example is the distortion of period histogram peaks for low frequencies and high discharge rates (Johnson and Swami 1983). Another is the fast initial portion of discharge rate adaptation (Smith 1977, Delgutte 1980).
- The model can serve as a "signal generator" to experiment with time-domain models of central processing, such as those of Licklider (1956, 1959,1962), Sruлович and Goldstein (1983), Moore (1986), etc. (de Cheveigné 1985a,b, 1986, 1989, de Cheveigné and Kuwabara 1989). Details of the spike generation process might have an effect on hearing performance, as suggested by Ohgushi (1978, 1983). The model allows us to examine these effects.

Limitations of the model.

This model is an approximation of a process that is difficult of access. Its behavior can differ from that of the real process in ways both known and unknown, and it is important to be aware of such limitations.

1) Structure.

This model of spike generation takes part in a model of auditory processing that is *layered*, *hierarchical*, and *"bottom-up"*.

"Layered" means for example that the spike production step is distinct from the adaptation step that precedes it, itself distinct from basilar membrane filtering.

"Hierarchical" means that each layer communicates only with the layers that immediately precede it and follow it. "Bottom-up" means that data is transferred from layer to layer in one direction only: from the periphery to the center.

Anatomical and functional descriptions appear layered, but it must be realized that "layeredness" is a constraint of the *description* process: nonlayered structures make little sense to us. Actual auditory structures might have a much finer granularity.

Systems built by humans have to be modular and hierarchical, with each layer being sufficiently simple for comprehension. It is tempting to assume the same constraints in auditory processing, but such an assumption is probably not justified.

Anatomical and physiological studies of the auditory system show a complex pattern of *afferent* innervation involving both contralateral and ipsilateral centers. This contradicts the hierarchical assumption, at least at higher levels. They also show a complex pattern of *efferent* innervation that reaches from the cortex all the way down to the outer hair cells of the basilar membrane. Since these cells are assumed to have a motor activity, it can be concluded that the purely "bottom-up" structure assumption is invalid as peripherally as the cochlear mechanical filtering stage, and as centrally as the cortex.

The auditory nerve fiber discharge patterns that our model attempts to reproduce are therefore not simply the input of central processing, but represent rather a signal present at one point of a complex, looped structure.

A further difficulty is that the model attempts to reproduce discharge patterns that can be measured in physiological experiments. These experiments are usually carried out on animals under anesthesia, which is known to modify the response of neurons to sound (Kuwada et al. 1989). Patterns measured in these conditions may not be typical of patterns that occur in actual hearing when the animal is awake (Kiang 1965).

2) Spike generation as distinct from adaptation.

When stimulus level varies, discharge rate varies over a much smaller range. The shape of discharge probability remains approximately the same, so this adaptation mechanism is usually modeled as an AGC-like process preceding spike generation. However AGC mechanisms cannot adequately explain the time course of adaptation (Smith 1977, 1979), and it is possible that spike generation and adaptation are related in a more intimate way.

3) Linearity.

The model assumes that discharge probability is proportional to the half-wave rectified version of the linearly filtered stimulus. This appears valid in a wide range of cases, but there are other cases where it is clearly incorrect (see for example the responses to low frequency tones in Ruggero and Rich 1983). It is possible that the apparent "linearity" is a by-product of the interaction of several non-linear processes (such as the combination of envelope compression and expanding non-linearity used by Seneff [1985]), and that the noisy nature of period histograms prevents such non-linearity from appearing clearly.

4) Half-wave rectification.

Other authors use a softer half-wave rectification function than the one we use here. Given the variability of data, it does not seem worthwhile to model the exact shape of half-wave rectification, especially as the "softening" visible in PST histograms may partly be the consequence of jitter.

5) Refractory effects.

Details of refractory effects are far from clear. Depending on the source, the absolute refractory period takes on a wide range of values (0.5 to 1.3 ms), as does the shape of

the function during the relative recovery period. Published data usually has very poor resolution, and it is difficult to assign the variability to any given cause.

It is possible that refractory effects are due to, not only the depressed discharge probability after a spike, but also to a reduced *conduction velocity*. Differences in absolute refractory period could be then explained in terms of differences in distance between the generation point and the measurement point. The mode shifts of conditional PST histograms for low interspike intervals in Gaumond et al. (1982) are consistent with this idea.

6) Jitter.

Gaussian jitter superimposed on the basic Poisson process reproduces the fall-off in synchrony quite well, but other causes are possible (such as low-pass filtering of hair-cell membrane potentials). The frequency limit of synchronization in fibers (represented here by a 55 μ s s.d. jitter) may be under- or over-estimated. On one hand, synchrony appears to deteriorate as one moves from the periphery to the central auditory system, meaning that higher centers may not have access to the synchrony evident peripherally. On the other, the measurement process may introduce noise (Johnson 1978), and anesthesia may prevent the peripheral transducer from operating at its best working point (Crane 1982). The upper limit of synchrony can be quite high in some animals (11 kHz in the owl, Konishi et al. 1988). In the human, ITDs (interaural time differences) can be discriminated with a precision of the order of the microsecond, suggesting that timing imprecision can be reduced or averaged out.

The limits just described do not prevent the model from being a useful modeling tool, or from producing data very similar to that obtained from nerve fiber recordings.

Conclusion.

The spike generator described in this report is a useful tool for auditory modeling. It can be used to better understand how the histogram data from physiological experiments reflect early stages of peripheral processing (filtering, transduction and adaptation). It can be used as a "signal generator" for modeling of central auditory processing. Finally, it allows us to model how auditory performance might be affected by details of the spike generation process.

Acknowledgements

Part of this work was carried out while the author was supported by an STP fellowship awarded by the Commission of European Communities. The author wishes to thank ATR for its kind hospitality, and both the Centre National de la Recherche Scientifique (CNRS, France) and Professor Culioli, Linguistics Department, Université Paris 7, for leave of absence.

References

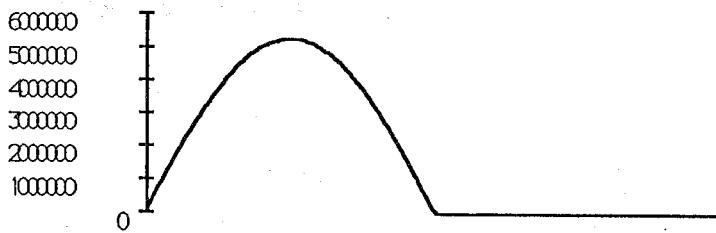
- de Boer, E., and de Jongh, H.R., (1978), "On cochlear encoding: potentialities and limitations of the reverse-correlation technique", JASA 63, 115-135.
- Carney, L.H., and Yin, T.C.T. (1988). "Temporal coding of resonances by low-frequency auditory nerve fibers: single-fiber responses and a population model", J. Neurophysiol. 60, 1653-1677.

- de Cheveigné, A. (1985), "Phase sensitivity and pitch shifts: the behavior of three temporal pitch models", *Proceedings of the Autumn ASJ meeting*, 277-278.
- de Cheveigné, A. (1985), "A nerve fiber discharge model or the study of pitch", *Trans. Committee on Hearing Res. of the ASJ*, H-85-37, 279-286.
- de Cheveigné, A. (1986), "A pitch perception model", *Proc. ICASSP-86*, 897-900.
- de Cheveigné, A., and Kuwabara, H. (1989), "Narrowed autocoincidence of nerve spike patterns and pure tone pitch", *Proceedings of the Autumn ASJ meeting*, 383-384.
- de Cheveigné, A. (1989), "Pitch and the narrowed autocoincidence histogram", *Proc. ICMPC*, 67-70.
- Crane, H.D. (1982), "IHC-TM connect-disconnect and efferent control. V", *JASA* 72, 93-101.
- Delgutte, B. (1980), "Representation of speech-like sounds in the discharge patterns of auditory-nerve fibers", *J. Acoust. Soc. Am.* 75, 843-857.
- Evans, E.F. (1978), "Place and time coding of frequency in the auditory system: some physiological pros and cons", *Audiology* 17, 369-420.
- Evans, E.F. (1983), "Pitch and cochlear nerve fiber discharge patterns" in *Hearing-Physiological bases and psychophysics*, edited by R. Klinke, R. Hartmann (Springer-Verlag, Berlin), 140-146.
- Evans, E.F. (1986), "Cochlear nerve fibre temporal discharge patterns, cochlear frequency selectivity and the dominant region for pitch", in *Auditory frequency selectivity*, edited by B.C.J. Moore, R.D. Patterson (Plenum Press), 253-264.
- Gaumond, R.P., Molnar, C.E. and Kim, D.O. (1982), "Stimulus and Recovery dependence of cat cochlear nerve fiber spike discharge probability", *J. Neurophysiol.* 48, 856-873.
- Gaumond, R.P., Kim, D.O. and Molnar, C.E. (1983), "Response of cochlear nerve fibers to brief acoustic stimuli: role of discharge history effects", *J. Acoust. Soc. Am.* 74, 1392-1398.
- Goldstein, J.L. and Srulovicz, P. (1977), "Auditory-nerve spike intervals as an adequate basis for aural frequency measurement", in *Psychophysics and physiology of hearing*, edited by E.F. Evans, J.P. Wilson (Academic Press, London), 337-347.
- Gray, P.R. (1967), "Conditional probability analyses of the spike activity of single neurons" *Biophysical J.* 7, 759-777.
- Greenberg, S. (1986), "Possible role of low and medium spontaneous rate cochlear nerve fibers in the encoding of waveform periodicity", in *Auditory frequency selectivity*, edited by B.C.J. Moore, R.D. Patterson (Plenum press), 241-251.
- Javel, E. (1980), "Coding of AM tones in the chinchilla auditory nerve: implications for the pitch of complex tones", *J. Acoust. Soc. Am.* 68, 133-146.
- Johnson, D.H. (1978), "The relationship of post-stimulus time and interval histograms to the timing characteristics of spike trains", *Biophysical Journal* 22, 413-430.
- Johnson, D.H. (1980), "The relationship between rate and synchrony in responses of auditory-nerve fibers to single tones", *J. Acoust. Soc. Am.* 68, 1115-1122.
- Johnson, D.H., and Swami, A. (1983), "The transmission of signals by auditory-nerve fiber discharge patterns", *J. Acoust. Soc. Am.* 74, 493-501.
- Kiang, N.Y.-S. (1965), "Discharge patterns of single fibers in the cat's auditory nerve", *MIT Research monograph* 35, pp154.
- Konishi, M., Takahashi, T.T., Wagner, H., Sullivan, W.E., Carr, C.E. (1988), "Neurophysiological and anatomical substrates of sound localization in the owl" in *Auditory function - neurobiological bases of hearing*, edited by G.E. Edelman, W.E. Gall, W.M. Cowan (Wiley, New York), 721-745.
- Kuwada, S., Batra, R., Stanford, T.R. (1989) "Monaural and binaural response properties of neurons in the inferior colliculus of the rabbit: effects of sodium pentobarbital", *J. Neurophysiol.* 61, 269-282.
- Licklider, J.C.R. (1956), "Auditory frequency analysis" in *Information theory*, edited by C. Cherry (Butterworth, London), 253-268.
- Licklider, J.C.R. (1959), "Three auditory theories" in *Psychology, a study of a science*, edited by S. Koch (McGraw-Hill), vol. I, 41-144.
- Licklider, J.C.R. (1962), "Periodicity pitch and related auditory process models", *International Audiology* 1, 11-36.
- Lieberman, M.C. (1978), "Auditory-nerve response from cats raised in a low-noise chamber", *J. Acoust. Soc. Am.* 63, 442-455.
- Lyon, R.F. (1984), "Computational models of neural auditory processing", *IEEE ICASSP*, 36.1.(1-4).
- Moore, B.C.J. (1980), "Neural interspike intervals and pitch", *Audiology* 19, 363-365.
- Moore, B.C.J. (1982), *An introduction to the psychology of hearing* (Academic Press, London).
- van Noorden, L. (1982), "Two channel pitch perception", in *Music, mind, and brain*, edited by M. Clynes (Plenum Press, New York), 251-269.
- Ohgushi, K. (1978), "On the role of spatial and temporal cues in the perception of the pitch of complex tones", *J. Acoust. Soc. Am.* 64, pp. 764-761.

- Ohgushi, K. (1983), "The origin of tonality and a possible explanation of the octave enlargement phenomenon", *JASA* 73, 1694-1700.
- Rodieck, R.W., Kiang, N.Y.-S., Gerstein, G.L. (1962). "Some quantitative methods for the study of spontaneous activity of single neurons", *Biophysical Journal* 2, 351-368.
- Rose, J.E., Brugge, J.F., Anderson, D.J., and Hind, J.E. (1969). "Some neural correlates of combination tones", *J. Neurophysiol.* 32, 402-423.
- Ruggero, M.A. and Rich, N.C. (1983). "Chinchilla auditory-nerve responses to low-frequency tones", *J. Acoust. Soc. Am.* 73, 2096-2108.
- Russel, I.J., and Sellick, P.M. (1980). "The responses of hair cells to low-frequency tones and their relationship to the extracellular receptor potentials and sound pressure level in the guinea pig cochlea", in *Neuronal mechanisms in hearing*, edited by J. Syka, L. Aitkin (Plenum Press, New York), 3-15.
- Schroeder, M.R., and Hall, J.L. (1974). "Model for mechanical to neural transduction in the auditory receptor", *J. Acoust. Soc. Am.* 55, 1055-1060.
- Siebert, W.M. (1970). "Frequency discrimination in the auditory system: place or periodicity mechanisms", *Proc. IEEE* 58, 723-730.
- Smith, R.L. (1977) "Short-term adaptation in single auditory-nerve fibers: some postsimulatory effects", *J. Neurophysiol.* 40, 1098-1112.
- Smith, R.L. (1979). "Adaptation, saturation, and physiological masking in single auditory-nerve fibers", *JASA* 65, 166-178.
- Srulovicz, P., and Goldstein, J.L. (1983). "A central spectrum model: a synthesis of auditory-nerve timing and place cues in monaural communication of frequency spectrum", *J. Acoust. Soc. Am.* 73, 1266-1276.
- Whitfield, I.C. (1978). "The neural code" in *Handbook of perception, vol. IV*, edited by E.C. Carterette and M.P. Friedman (Academic Press, New York), 163-183.
- Whitfield, I.C. (1979). "Periodicity, pulse interval and pitch", *Audiology* 18, 507-512.
- Whitfield, I.C. (1980). "Theory and experiment in so-called pulse-interval pitch", *Audiology* 20, 86-88.

Appendix: histogram examples.

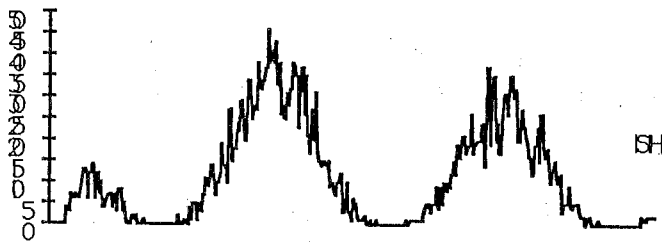
response for 100Hz sine BM motion



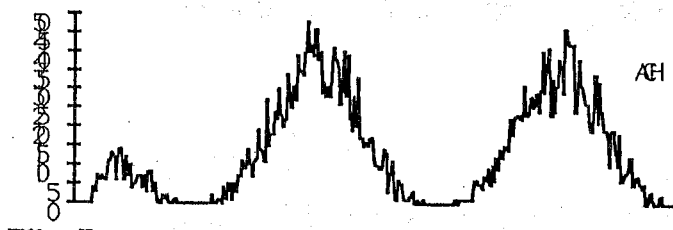
driving
function



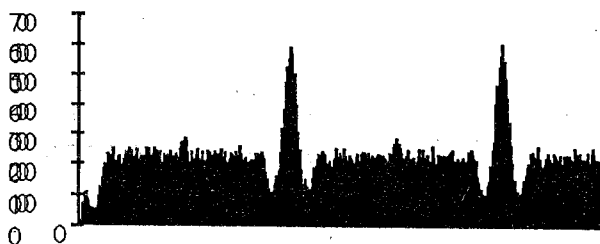
Period H



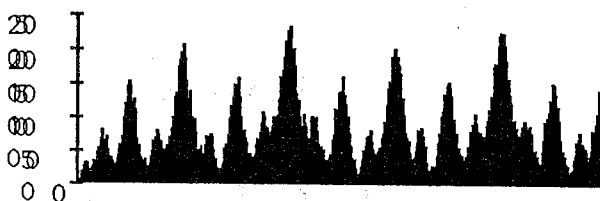
ISI H



Auto-
coincidence H

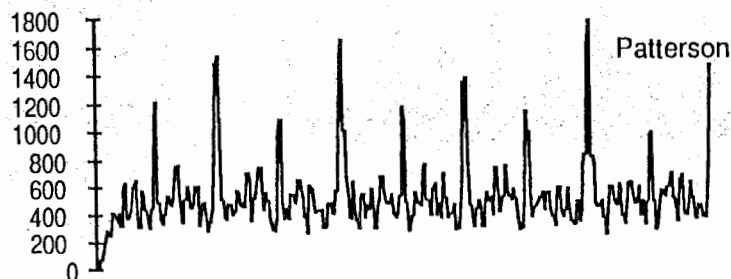
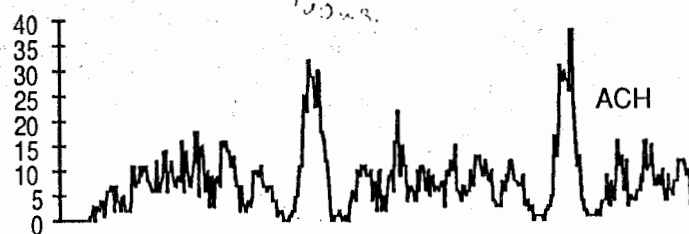
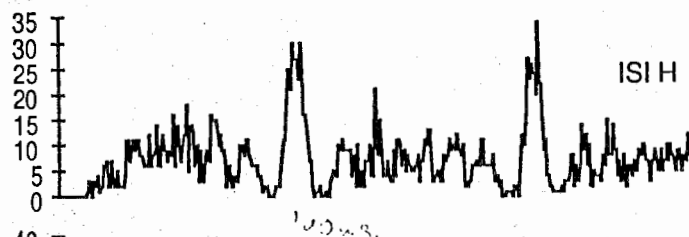


Narrowed
Auto-
coincidence H
(order 10)



Histogram
similar to
Patterson's
"pulse ribbon"
model.

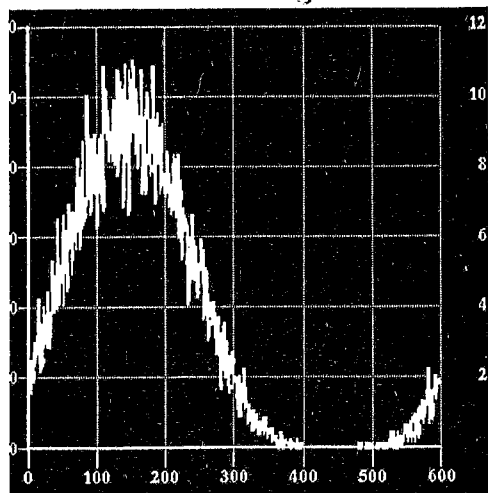
response for 100 Hz complex BM motion
(harmonics 3,4,5,6 in cosine / sine phase)



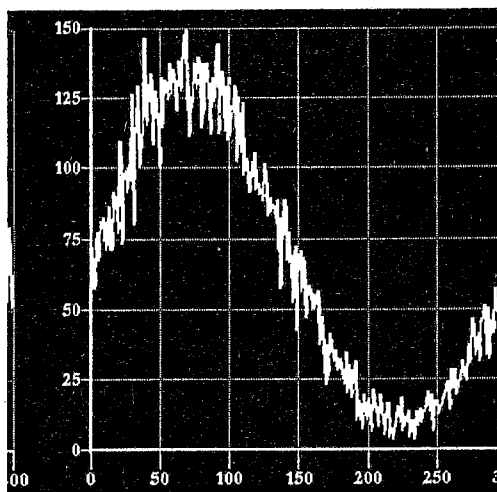
order 10

Effect of $33\mu\text{s}$ jitter as
a function of frequency

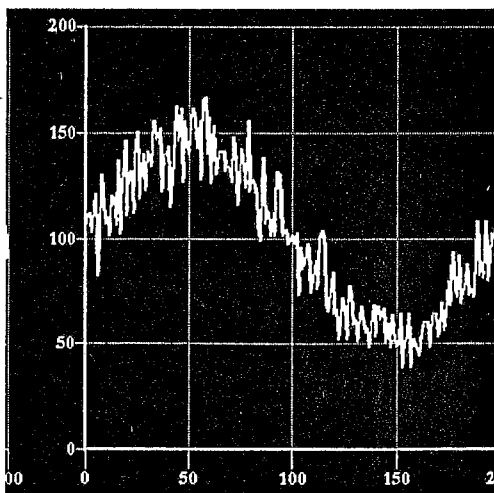
Period Histograms



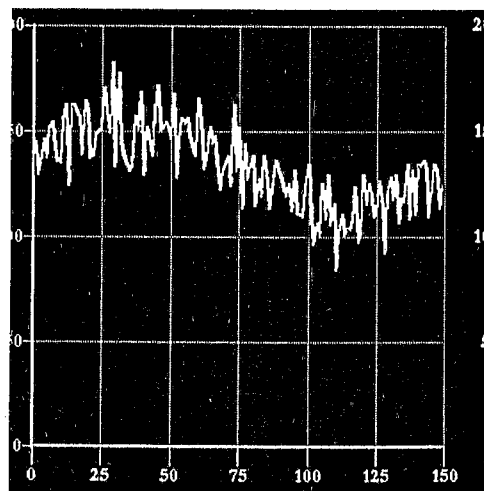
1 KHz



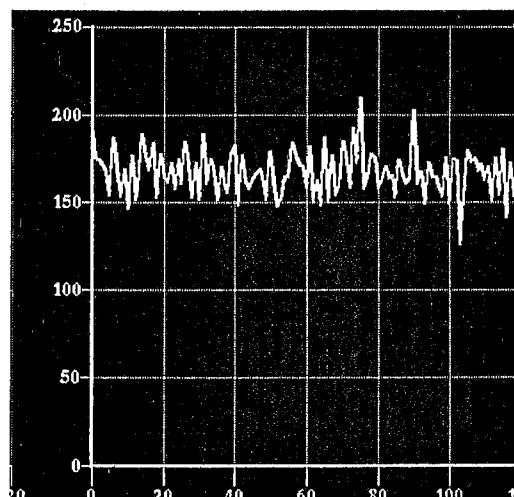
2 KHz



3 KHz



4 KHz



8 KHz