

Connecting Correlograms to Neurophysiology and Psychoacoustics

Malcolm Slaney

*Interval Research Corporation
1801 Page Mill Road; Building C; Palo Alto, CA 94304 USA*

1. Auditory Imaging Hypothesis

Over the years, many researchers have used auditory images to model auditory perception. A **correlogram** is an auditory image that measures the temporal regularity of an auditory stimulus. A correlogram represents an auditory stimulus as a function of three variables: cochlear position or best frequency versus short time periodicity, as a function of time. Psychoacousticians since Licklider (1951) have used the correlogram to model pitch perception and, more recently, auditory scene analysis.

We know that the auditory system encodes temporal regularity precisely. That it does is not surprising; many systems produce a periodic response to a periodic input. Yet the auditory system goes out of its way to preserve the temporal aspects of the signal. What is the brain doing with this temporal precision?

I hypothesize that the auditory system builds a representation of sound that distinguishes the temporal aspects of the sound. There are many ways to compute a correlogram. I describe three approaches and test their performance with three different auditory tasks.

In this paper, I describe the role of correlograms in explaining human and animal auditory perception. Section 2 explains what the neurophysiological data tell us about temporal processing. Section 3 describes three mechanisms for measuring the temporal regularity of a signal. Section 4 discusses the response of three correlogram models to well-known tests that are based on neurophysiology and pitch perception. Conclusions are presented in Section 5.

2 Neurophysiological Data

The neurophysiological data for correlogramlike processing are tantalizing, and incomplete. The evidence takes four forms: (1) the preservation of temporal information, (2) the existence of modulation transfer function (MTF) maps of amplitude modulation (AM) or periodicity information, (3) the evidence for delay lines in the binaural system, and (4) evidence for delayed spikes.

1. Temporal information: Many researchers (e.g., Cariani 1996) have noted that the timing of auditory neuron spikes encodes much of the information needed to model human psychoacoustic responses. But the presence of this temporal information in the spike data does not explain how the brain uses and extracts the temporal cues.
2. MTF maps: Langner (1988) and other researchers have measured the response of inferior colliculus (IC) neurons to amplitude modulation. They have found an ordered array of these neurons that maps cochlear best frequency versus modulation frequency. However, correlograms based on MTF data do not fit the psychoacoustic data, as I will discuss in Section 4.

3. Delay lines: A direct way to implement a periodicity detector is to use a delay mechanism. Carr (1990) and other investigators have mapped responses with systematic delays for binaural localization in the owl, and Suga (1990) has found long delays in the bat's echo-location system. No equivalent structure is currently known in the monaural sections of the human auditory system.
4. Delayed spikes: Depireux (1997) have reported neural data that are synchronized to the stimulus modulation, and have shown evidence of two spike sources separated by a range of fixed delays. Figure 1 shows the autocorrelation of the data from a representative site for modulation frequencies from 100Hz to 500Hz. The autocorrelation shows a strong periodicity at the fundamental period, as well as additional structure around each of the fundamental responses. Notice that, in the 100Hz case, there are two small peaks, indicated by the arrows. These peaks stay at a fixed distance from the large peaks as the modulation frequency changes. This pattern suggests that two spike sources are firing at the basic periodicity, separated by 2ms from each other.

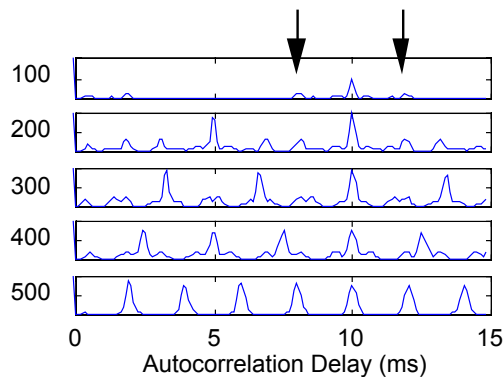


Figure 1—Autocorrelation of a spike train due to an amplitude modulated stimulus at five different modulation frequencies from 100Hz to 500Hz. The plots show a response at the fundamental periodicity, and also a second set of responses delayed by 2ms. You can see this delayed response by looking at the structure around the peak at 10ms. These data are consistent with two spike generators entrained to the basic modulation, but delayed 2ms relative to each other. (Source: Depireux, 1997)

3 Correlograms

We use a gammatone filterbank (Patterson 1995) to model the cochlea and provide the input to each correlogram. A simple half-wave rectifier serves as a detector, and the result is the predicted spike rate.

There are three different methods to summarize the temporal aspects of an auditory signal. All three methods process the output of a filterbank model of the cochlea, but they differ greatly in their perspective.

The first, the **autocorrelation correlogram** proposed by Licklider, is the simplest to describe mathematically. The second, the **auditory image model** (AIM) proposed by Patterson (1995), converts a fast moving stream of auditory nerve firings into a stable pattern. I will describe several variations of AIM. Finally, several neurophysiologists have measured the response of auditory neurons to amplitude modulation. From this data, we create an **MTF correlogram**.

No matter how we calculate the correlogram, we assemble all the channels by characteristic frequency, to create an image such as those shown in Figure 2. The correlogram shows periodicity along the horizontal axis versus characteristic frequency of each cochlea channel along the vertical axis. For any given auditory stimulus, a dark vertical line indicates that the signal is periodic with that time interval.

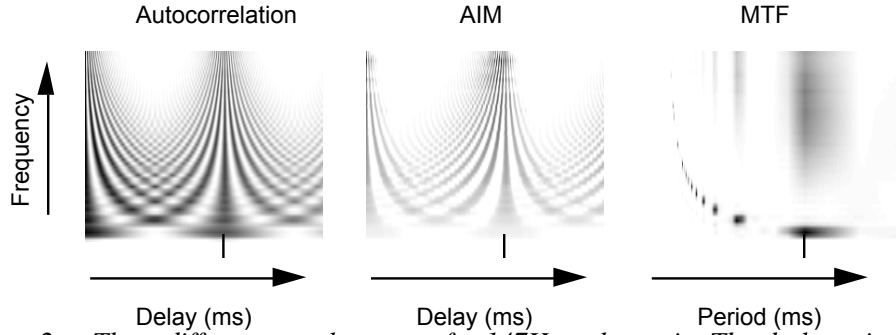


Figure 2 — Three different correlograms of a 147Hz pulse train. The dark vertical line in each picture corresponds to a common response at 6.8ms.

3.1 Autocorrelation Correlogram

Autocorrelation is the simplest method to measure the temporal regularity of a signal. The autocorrelation of a signal is defined as

$$R_{xx}(\tau) = \int_{-\infty}^{\infty} x(t)x(t+\tau)dt,$$

so $R_{xx}(\tau)$ tells us how much the signal is like itself τ seconds later.

This ideal correlogram is difficult to implement and makes sense if the signal only has finite energy and never changes. Instead, we compute the autocorrelation over short time windows. I choose a segment of the signal using a window $w(t)$:

$$x_i(s) = w(s)x(t+s).$$

Then, the local autocorrelation is written as

$$R_{xx}(t, \tau) = \int_{-\infty}^{\infty} x_i(s)x_i(s+\tau)ds.$$

This definition is convenient because we can compute it efficiently using the fast Fourier transform (FFT). Other methods—for example, using exponential weighting (Yost 1997)—produce similar results.

3.2 AIM Correlogram

Patterson’s AIM converts a pattern of auditory spikes into a stable image. The essential mechanism of AIM is a trigger, which is designed to follow the peaks of the auditory channel. Because triggering occurs at the same point in a repetitive waveform, a stable image of the waveform is preserved for periodic signals.

I consider two components of the trigger mechanism; each serves to limit the points at which the trigger can occur. The most important component of AIM is a threshold that decays over several milliseconds; the first component triggers when the auditory channel’s output exceeds this time-varying threshold. The second mechanism limits triggering to each local maxima or peaks of the output channel.

Figure 3 summarizes these mechanisms. Without these mechanisms to constrain triggering, AIM triggers at every point in the waveform, and the results look nearly identical to an autocorrelation correlogram. Adding a decay threshold (20ms in the current study) causes the largest qualitative change. Depending on the decay time of the threshold, the trigger mechanism will tend to find the largest peak of the input waveform and be set off at only this one (periodic) point. A further refinement is possible if we limit the trigger mechanism to fire only at the local maximum or peak of the waveform as Weintraub (1985) has done.

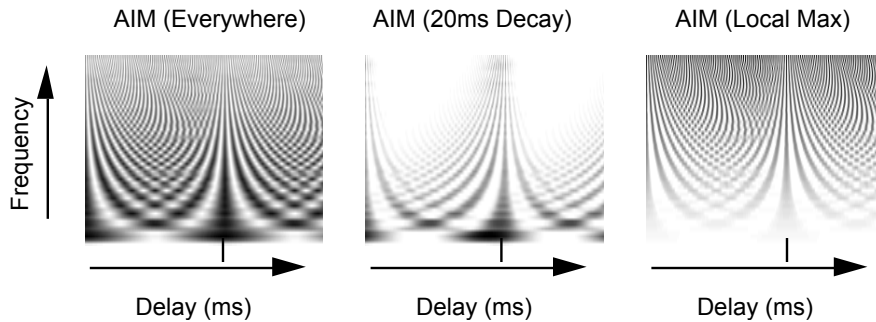


Figure 3— Examples of three simple AIM triggers. Every positive spike causes a trigger event in the left image (note similarity to autocorrelation correlogram). The middle plot shows a stable image with a threshold that decays linearly and resets after each trigger event (note asymmetry). The right plot shows the result with a trigger event at every local maximum (note sharpened peaks). Any combination of these mechanisms produces a reasonable correlogram. The AIM correlograms studied in this paper use both trigger mechanisms.

One difference between these correlograms is the way that they preserve temporal asymmetry in the original waveform. Irino (1996) reports the perceptual difference between one waveform and its time-reversed cousin. It is easy to show that the autocorrelation correlogram is symmetric and would mask any temporal asymmetry in the channel waveform. In contrast, the threshold-decay correlogram does preserve some asymmetry.

3.3 MTF Correlogram

The third correlogram model is based on MTF experiments of Langner (1988) and other scientists. One way that neurophysiologists characterize auditory neurons is by the neuron's response to amplitude-modulated tones. In many cases, neurons respond with the largest number of spikes to a narrow range of envelope frequencies.

I simulated this neurophysiological result by using a bank of modulation detectors at the output of each cochlear channel. I implemented the MTF detector bank with an FFT and rearranged the FFT bins such that the horizontal axis corresponded to the period. This axis is similar to the delay representation in the other correlogram displays.

A striking feature of the MTF correlogram is its response to the harmonics of a click train. The fundamental and the high harmonics show a strong response at the fundamental period and little response at the second through eighth harmonics (notice the blank white areas at 6.8ms delay). Of course, an auditory channel near the fundamental has a strong envelope at the fundamental period, as shown in Figure 2. The same response patterns occur for auditory channels with high CF, since several harmonics are included in a filter. A high-frequency channel's response has a strong response at the fundamental frequency, as reflected in the MTF correlogram.

The low harmonics (second through eighth) are different from the rest. Each harmonic is resolved by the cochlear filter bank, and most of the energy of the MTF correlogram is at that harmonic's frequency and the associated harmonics (due to the harmonics introduced by the half-wave rectifier detector). But there is no response at the fundamental period.

4 Correlogram Tests

This section compares the three types of correlograms based on an MTF test and two pitch tests. First, I consider the sensitivity of a cell in each correlogram to AM, and measure the cells MTF. Then, we consider the three correlograms in two types of pitch experiments. First, we see whether the correlograms model an important aspect of pitch perception—the first effect of pitch shift. Second, we study whether correlograms replicate the dominance region of pitch perception.

4.1 Correlogram MTF

The three correlograms have markedly different behavior with respect to their MTF. Because the neurophysiological data (Langner 1988) shows a narrow peak around the BMF, I plot the response of a correlogram pixel or neuron with a characteristic frequency of 5kHz and a best modulation frequency (BMF) of 200Hz. I measured the MTF using a 5kHz carrier, AM modulated at a range of frequencies.

Figure 4 shows the results of this analysis. Not surprisingly, the correlogram based on envelope detection—the MTF correlogram—has the sharpest response and the smallest off-frequency response. The auto-correlation and AIM correlograms are not as selective. Many different modulation frequencies produce a response in the neuron with a 5ms delay. There are many 5ms periodicities in a 5kHz carrier, no matter what the modulation frequency. Thus, the autocorrelation and AIM correlograms poorly model Langner’s MTF data.

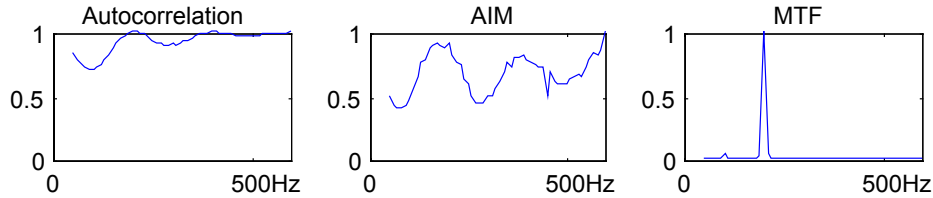


Figure 4 — Modulation transfer functions for a correlogram neuron with best frequency of 5kHz and a BMF of 200Hz. The MTF correlogram is close to the neurophysiological data (Langner 1988).

4.2 Correlogram Pitch

Correlograms have typically excelled in modeling of human pitch perception; such models often are based on the periodicity of the signal. We collapse the correlogram by integrating across frequency to form the summary correlogram. The peaks of the summary correlogram provide a good model of human pitch perception (Meddis 1991, Slaney 1990). The summary correlograms computed from each of the correlograms in Figures 2 and 3 all have the peak positioned correctly.

The results are not as uniform for more difficult pitch examples. The first effect of pitch shift is a standard test that shows how the pitch of inharmonic sounds is perceived. In this experiment, sinusoids at $f-100$, f , and $f+100$ Hz are added together and the perceived pitch is judged. Figure 5 shows the result for all three correlograms.

The fine time structure, rather than the envelope frequency, most accurately models human perception of the first effect of pitch shift. An alternate way to synthesize our example is to amplitude modulate the base frequency by a 100Hz sine wave. The MTF correlogram detects this envelope. Unfortunately, the MTF correlogram is blind to the fine temporal structure, as shown in Figure 5, where both the autocorrelation and AIM correlograms track the pitch correctly.

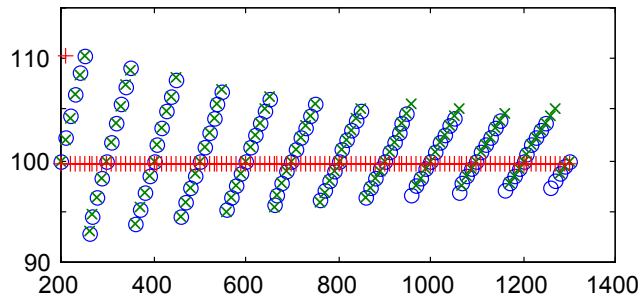


Figure 5 — Predicted pitches for three different correlograms showing the first effect of pitch as a function of center frequency, f . The \circ marks are the auto-correlation correlogram, the \times marks are the AIM correlogram prediction, and the $+$ marks are the MTF correlogram prediction. The human data matches the AIM and autocorrelation lines.

4.3 Correlogram Dominance

Many researchers have studied which parts of a sound affect pitch perception. Normally, all harmonics of a sound produce consistent evidence for a single pitch. In this experiment, we shift the low and high harmonics in different directions to test which harmonics are most important for the pitch judgment. The ratio of the summary correlogram values at the peaks due to the harmonics shifted higher versus lower is compared to 1. When the ratio is greater than 1, the high-shifted harmonics are most important; when the ratio is less than 1, the low-shifted harmonics are determining the pitch. The transition point is then plotted as a function of the fundamental frequency of the harmonic sequence. Figure 6 shows the result for two of our three correlograms.

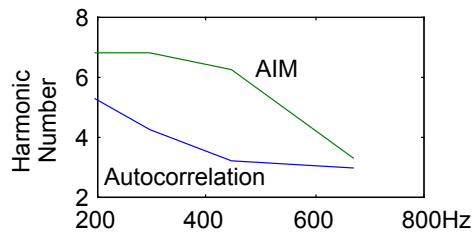


Figure 6 — A comparison of pitch dominance for autocorrelation and AIM correlograms. This plot shows the harmonic when an ambiguous sound shifts from the low to high pitch, it indicates which harmonics are most important for the pitch judgment. The psychoacoustic data are close to the autocorrelation result. The MTF correlogram can not be measured with this test.

As described by Meddis and Hewitt (1991), the autocorrelation correlogram tracks human perception. The AIM correlogram, for reasons that I do not yet understand, does not do as well in this test. No measurement of the MTF correlogram is possible due to the unimportance of the second through eighth harmonics.

5 Discussion

We discussed the connections among correlograms, to the neurophysiological data and psychoacoustics results. Analysis of the temporal aspects of a sound is important for many psychoacoustic results, yet the neurophysiological mechanisms are not understood. Auditory pathways preserve the temporal information, yet it is not clear how this information is converted into an auditory perception.

We used several kinds of temporal analysis to build three correlograms of sound. The three models match the neurophysiological and psychoacoustic data to varying degrees. Autocorrelation is the simplest to describe, and models many of the psychoacoustic data, but is neurophysiologically implausible. Several variations of Patterson's AIM produce good results and are plausible. A correlogram built with MTF neurophysiology data poorly matches the psychoacoustic pitch data.

Two types of further experiments would connect the neurophysiology to the psychoacoustics results. First, application of inharmonic sounds to cells that respond well to AM modulation would help us to rule out MTF as a model of temporal processing. Second, identifying the delay pathways and the correlators that map periodicities would help us better understand how time is used by the brain.

6 Acknowledgments

This paper was shaped by Richard F. Lyon, Roy Patterson, Ray Meddis, Steven Greenberg, William Yost, Shihab Shamma, Didier Depireux, Michele Covell, Daniel Levitin, Lyn Dupré and the participants of the Stanford Hearing Seminar.

7 References

Cariani, P.A. and Delgutte, B. (1996) Neural correlates of the pitch of complex tones. I. Pitch and pitch salience. *J. Neurophysiol.* 76, 1698-1716.

- Carr, C.E. and Konishi, M. (1990) A circuit for detection of interaural time differences in the brainstem of the owl. *J. Neurosci.* 10, 3227-3246.
- Depireux, D., Klein, D.J., Simon, J.Z., Shamma, S.A. (1997) Neural correlates of pitch in the inferior colliculus. Abstracts of the 1997 Midwinter Meeting of the ARO.
- Irino, T. and Patterson, R.D. (1996) Temporal asymmetry in the auditory system. *J. Acoust. Soc. Am.* 99, 2316-2331.
- Langner, G and Schreiner, C. (1988) Periodicity coding in the inferior colliculus of the cat: I. Neuronal mechanisms. *J. Neurophysiol.* 60, 1799-1822.
- Licklider, J.C.R. (1951) A duplex theory of pitch perception. *Experientia* 7, 128-131.
- Meddis, R. and Hewitt, M.J. (1991) Virtual pitch and phase sensitivity of a computer model of the auditory periphery: I pitch identification. *J. Acoust. Soc. Am.* 89, 2866-2882.
- Patterson, R., Allerhand, M., Giguere, C. (1995) Time-domain modeling of peripheral auditory processing: A modular architecture and a software platform. *J. Acoust. Soc. Am.* 98, 1890-94.
- Slaney, M. and Lyon, R.F. (1990) A perceptual pitch detector. *IEEE ICASSP*, Albuquerque, NM.
- Suga, N. (1990) Cortical computational maps for auditory imaging. *Neural Networks* 3, 3-21.
- Weintraub, M., (1985) A theory and computational model of auditory monaural sound separation., Ph.D. Dissertation, Stanford University.
- Yost, W. (1997) Defining the limits of pitch: Iterated rippled noise. Abstracts 1997 MWM of ARO.