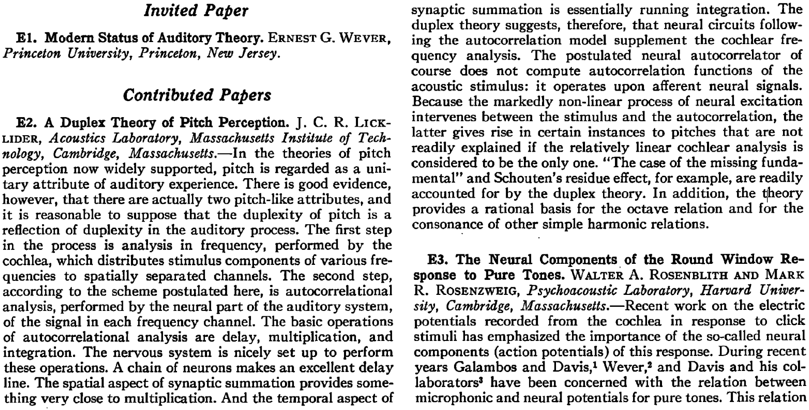
Lecture 6: Duplex Theory of Pitch Perception

Mark Hasegawa-Johnson

ECE 537: Speech Processing Fundamentals

Outline

Licklider’s Proposal



Invited Paper

E1. Modern Status of Auditory Theory. Ernest G. Wever, *Princeton University, Princeton, New Jersey.*

Contributed Papers

E2. A Duplex Theory of Pitch Perception. J. C. R. Licklider, *Acoustics Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts.*—In the theories of pitch perception now widely supported, pitch is regarded as a unitary attribute of auditory experience. There is good evidence, however, that there are actually two pitch-like attributes, and it is reasonable to suppose that the duplexity of pitch is a reflection of duplexity in the auditory process. The first step in the process is analysis in frequency, performed by the cochlea, which distributes stimulus components of various frequencies to spatially separated channels. The second step, according to the scheme postulated here, is autocorrelational analysis, performed by the neural part of the auditory system, of the signal in each frequency channel. The basic operations of autocorrelational analysis are delay, multiplication, and integration. The nervous system is nicely set up to perform these operations. A chain of neurons makes an excellent delay line. The spatial aspect of synaptic summation provides something very close to multiplication. And the temporal aspect of synaptic summation is essentially running integration. The duplex theory suggests, therefore, that neural circuits following the autocorrelation model supplement the cochlear frequency analysis. The postulated neural autocorrelator of course does not compute autocorrelation functions of the acoustic stimulus: it operates upon afferent neural signals. Because the markedly non-linear process of neural excitation intervenes between the stimulus and the autocorrelation, the latter gives rise in certain instances to pitches that are not readily explained if the relatively linear cochlear analysis is considered to be the only one. “The case of the missing fundamental” and Schouten’s residue effect, for example, are readily accounted for by the duplex theory. In addition, the theory provides a rational basis for the octave relation and for the consonance of other simple harmonic relations.

E3. The Neural Components of the Round Window Response to Pure Tones. Walter A. Rosenblith and Mark R. Rosenzweig, *Psychoacoustic Laboratory, Harvard University, Cambridge, Massachusetts.*—Recent work on the electric potentials recorded from the cochlea in response to click stimuli has emphasized the importance of the so-called neural components (action potentials) of this response. During recent years Galambos and Davis,[1] Wever,[2] and Davis and his collaborators[3] have been concerned with the relation between microphonic and neural potentials for pure tones. This relation

The Duplex Theory of Pitch Perception

- Cochlear analysis
  - Bandpass filtering

$$v_k[n] = h_k[n] * s[n]$$

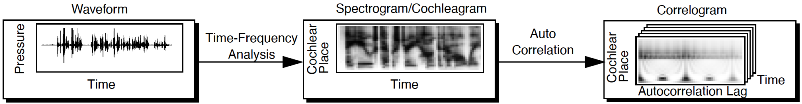
  - Nonlinear amplitude compression

$$a_k[n] = G(v_k[n])$$

- Autocorelation in each band
  - Chain of neurons = delay line
  - Synaptic summation = multiply two signals, and compute a running integral

$$\phi_k(\tau) = \sum_n a_k[n] a_k[n-\tau]$$

Slaney's Implementation



Copyright Malcolm Slaney, *Auditory Toolbox*, 1989

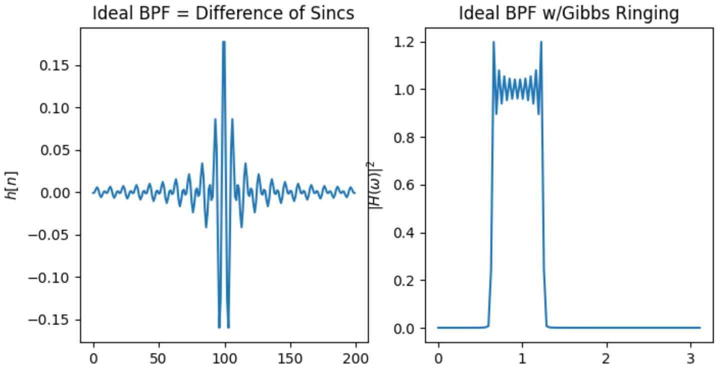
Outline

1. Duplex Theory of Pitch Perception
2. **Cochlear Filtering**
3. Cochlear Nonlinearity
4. Correlogram
5. Conclusions

Ideal Bandpass Filters?

Until now, we've been assuming ideal bandpass filters:

$$h_k[n] = \frac{\omega_{k+1}}{\pi}\text{sinc}\left(\omega_{k+1}n\right) - \frac{\omega_k}{\pi}\text{sinc}\left(\omega_k n\right)$$



Gammatone Filters

Roy D. Patterson ("Auditory filter shapes derived with noise stimuli," 1976) showed that the shape of the auditory filter, as a function of Ω (radians/second), is

$$|H_k(\Omega)|^2 = \left(\frac{\Omega_k^2 + B_k^2}{(\Omega - \Omega_k)^2 + B_k^2}\right)^2$$

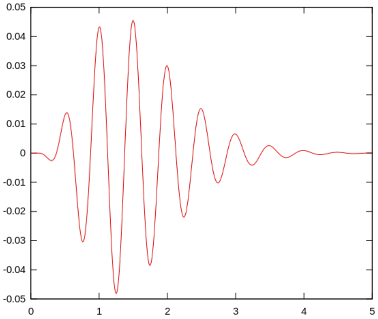
If you make reasonable assumptions about phase, and take the inverse continuous-time Fourier transform, you get

$$h_k(t) = te^{-B_k t} \sin\left(\Omega_k t\right) u(t)$$

This is called a "gammatone filter" because the $te^{-B_k t} u(t)$ part is like the gamma function in mathematics.

Gammatone Filters

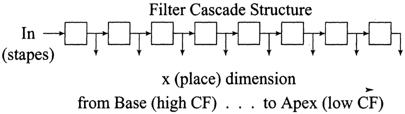
$$h_k(t) = te^{-B_k t} \sin\left(\Omega_k t\right) u(t)$$



Public domain, `https://commons.wikimedia.org/wiki/File:Sample_gammatone.svg`

Sequence of Lowpass Filters

Richard Lyon modeled the cochlea as a sequence of lowpass filters:



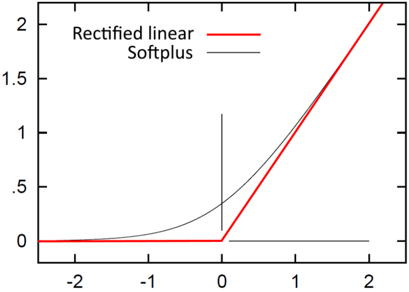
Copyright 1998, Richard Lyon, Filter Cascades as Analogs of the Cochlea

Outline

1. Duplex Theory of Pitch Perception
2. Cochlear Filtering
3. Cochlear Nonlinearity
4. Correlogram
5. Conclusions

Half-Wave Rectification

The inner hair cells perform a sort of soft half-wave rectification (HWR). This HWR may be modeled as a ReLU (red curve), although it is closer to the form of a softplus (gray curve).

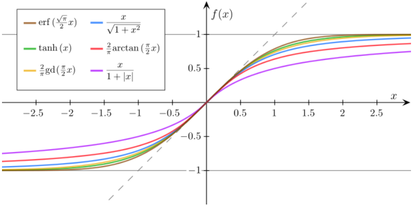


CC-SA 4.0, `https://commons.wikimedia.org/wiki/File:Rectified_linear_and_Softplus_activation_functions.png`

Half-Wave Rectification

In addition to the HWR, there should also be some kind of compressive nonlinearity, so that the neural activation, $a_k[n]$, is proportional to perceived nonlinearity. For example, Slaney proposes that the basilar membrane velocity $v_k[n]$ is compressed as

$$a_k[n] = \frac{1}{2}\left(1 + \tanh\left(v_k[n] + a\right)\right)$$



CC-SA 3.0, `https://commons.wikimedia.org/wiki/File:Gjl-t(x).svg`

Compressive Nonlinearity

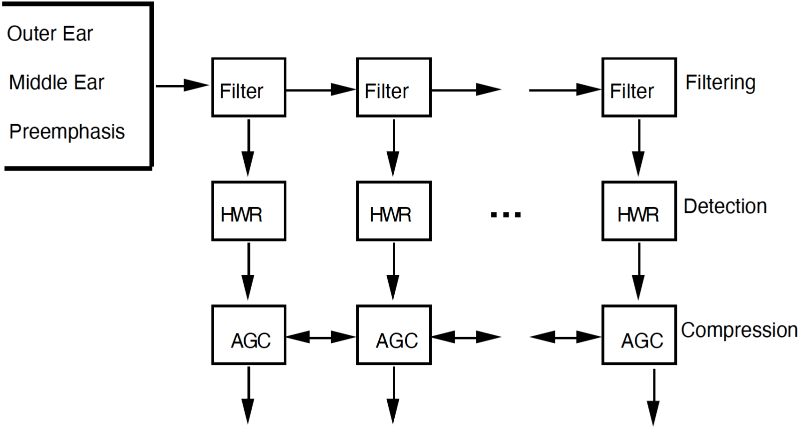
Alternatively, we could use Fletcher's $G()$ function directly, like this:

$$a_k[n] = G(L_k)u\left(v_k[n]\right),$$

where $u()$ is the unit step function, and $G(L_k)$ is computed from the loudness level $L_k(f_k, v_k[n])$ of the tone $v_k[n]$ at frequency f_k.

HWR + AGC

Slaney models the compressive nonlinearity using an automatic gain control (AGC):

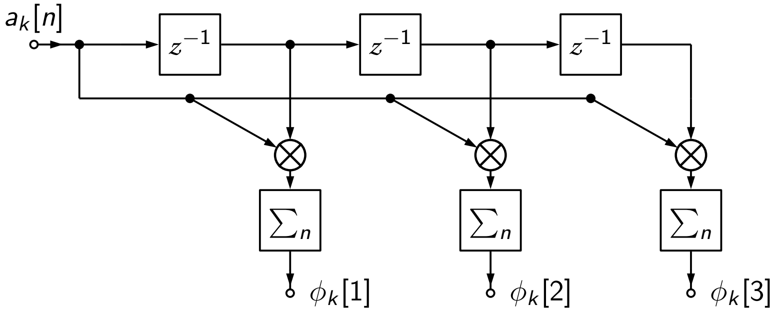


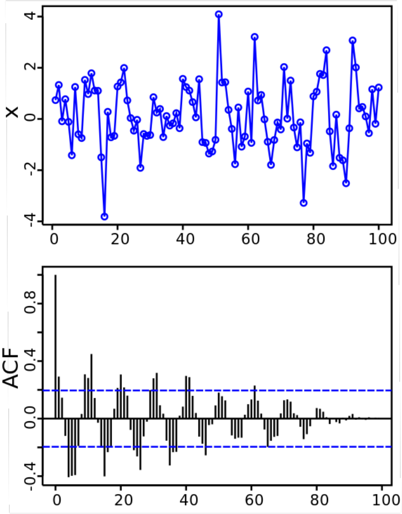
Copyright Apple, Inc., 1988, "Lyon's Cochlear Model," Malcolm Slaney

Outline

Autocorrelation as a Neural Circuit

$$\phi_k(\tau) = \sum_n a_k[n] a_k[n-\tau]$$





CC-SA 2.5, `https://commons.wikimedia.org/wiki/File:Acf_new.svg`

Autocorrelation is basically a phase normalizer. For example, it finds the periodic component hidden in this noisy signal, and aligns that periodic component so its peaks are at $\tau = 0$, and at $\tau =$ multiples of the period.

Correlogram: Autocorrelation as Function of Lag and Frequency

→Visit the CCRMA Correlogram Museum←

Outline

1. Duplex Theory of Pitch Perception
2. Cochlear Filtering
3. Cochlear Nonlinearity
4. Correlogram
5. Conclusions

Conclusions

- Licklider's "Duplex theory" suggested that the ear hears pitch by performing an autocorrelation separately, on each of the neural channels coming from the inner ear.
- Slaney seems to have invented the term "correlogram" as the name for the visual representation of this perceptual construct.
- Note: the neural circuitry necessary to perform this task has never been found in any anatomical experiment. Nevertheless, this model is a good description of human perceptual behaviors.