Pitch Perception Models - Alain de Cheveigné
Matthew Weisberg

The concept of **pitch** is an attempt to characterize the experience of the perception of the relative highness or lowness of a sound related to its frequency composition. It is is a flexible term used to model the experience of dynamic sounds and can be influenced by the lens through which you want to describe or define it. Most simply, pitch is identified by the the fundamental frequency of oscillation that best explains the patterns of oscillation observed by an observer. There are two prominent theories of how pitch is perceived in humans, known as place theory and time theory

Place theory uses spectral analysis to describe the pitch of sound. If a sound source has a fundamental wavelength L, we expect to see sympathetic vibrations in the sound at integer multiples of L. By breaking a sound down into its spectral components, we can measure the ratios between the prominent harmonics and determine the fundamental frequency which is likely to produce that harmonic series. This is done by analyzing the superimposition of periodic waves using Fourier Transform (FT) theory. In humans, this process is carried out by the tonotopic vibration of basilar membrane, as first suggested by Helmholtz. Pythagoras first discovered the concept of ratios in making a sound with his monochord, and later Mersenne started to illuminate the need for superimposition to explain harmonic content. Du Verney resolved this need with the discovery of tonotopic resonance, and Euler introduced linearity which gives us the basis of superimposition. In signal processing, we use different types of "combs" (including FT sinusoidals) to see which "comb" maps to our incoming data stream. Because frequency perception is non-linear, we use Mel-distributions for parsing speech (though this possibly disrupts pitch detection). Further applications of FT Applying an FT to a power spectrum gives auto correlation and applying an FT to a log spectrum gives cepstrum.

Time theory uses waveform data to represent pitch through periodicity as determined by the autocorrelation function. The autocorrelation function (ACF) measures a wave against itself in order to determine the period of a cycle that preserves the original source as accurately as possible. We believe that this occurs in the brain through neuron measuring a wave against a record of itself and measuring the difference between the actual and "stored" wave to calculate when it is repeating. Licklider/Meddis/Hewitt efforts put forth the idea that an ACF is done on each channel of the spectralized sound coming from the basilar membrane, and that combining the best fit ACF from all the channels into a Summed Autocorrelation function (SACF) gives a reliable measure of pitch. A related model is the cancellation model, which performs the opposite calculation of the ACF. Another related is the strobe-temporal integration (STI) model, where the signal is compared to a series of pulses (rather than itself) the determine which pulse rate best matches the periodicity of the input signal.