Edward Stein
EE675 (ISE575)
February 17, 2005

A Perceptual Pitch Detector
By Malcolm Slaney and Richard F. Lyon

In the paper Slaney and Lyon propose a process of pitch detection that accounts for human perception in tonal interpretation instead of straight frequency content as is the case in most models. They provide a good start, but seem to fall short with an overly simplified model with little or no error checking.

The idea was originally proposed by Licklider in the early 50's, but at the time lack of computational power and the heavy use of autocorrelation prevented actual implementation of an algorithm. Following Licklider's methods they have broken the pitch detection process down to three key elements. A model of the inner ear or Cochlea, a temporal correlogram to mimic the assumed processing of pitch information and a pitch detector are used to arrive at a single "best" pitch to represent the input signal. Recently there have been publications providing evidence that the human perception may not involve a stage equivalent to autocorrelation. Even so, it is a fair approximation that provides reasonably similar results.

The Cochlear model is a combination of three stages. The first is a series of 2^{nd} order filters that emulate the basilar membrane and frequency detection. The transition from basilar sensation to hair cells and nerve endings is modeled by a rectifier and automatic gain adjustment stages. The result although not exactly part by part the same as the working of the ear provides an approximate black box representation for the transition from acoustic energy to electrical brain activity.

The information is represented internally as a function representing the correlogram of the signal. This function considers frequency vs. time to verify periodicity or perceived periodicity of the signal. Time lag bins with strong energy over all of the channels becomes an indication a possible pitch period. The function is summed and integrated to wait these peaks. A final narrow autocorrelation function generates weighting of all the subharmonics for the possible pitch peaks. A final selection is made to approximate what the "best" pitch to represent the input signal. In this system no validation is made to decide if the pitch calculated is an actual pitch or somehow related to the zero crossing rate of noise for example.

However, because of this process other perceptual phenomena in pitch detection are caught. For example, pitch detection of a residue pitch which has no energy in the fundamental can be detected by the sub-harmonics. Also non-harmonic tones can be detected in periodic signals such as quantization noise. There are still pitch effects that are not accounted for by this model. Such as more than one pitch at a time, making this system geared more towards voice streams. It also lacks sufficient brain modeling, which can make final decisions or other filtering and averaging operations.

The results section had surprisingly vague demonstrations with out true verification of valid results. The testing was all tone on relatively simple monophonic signals below 300Hz. There were interesting results for things like the Shepard tones which revealed the systems ability to match human interpretation of volume curves and their effect on pitch. But there was no verification against known pitches which would need to be established to prove greater and equal to previous pitch detection systems. In addition the lack of polyphonic tone detection means the system can never completely match that of the human perception.