ISE575/CSCI575

Week 6
February 17, 2005

Mohammad Anwar Hossain
Student ID: 8813190224

YIN, a fundamental frequency estimator for speech and musica

By - Alain de Cheveigne, Hideki Kawahara

In one sentence, what the paper is about?

Discussion & Comparison of some approaches/algorithms/extensions mainly based on auto-correlation function to estimate fundamental frequency (F0) of speech or musical sound.

What is F0 estimation and Why we need it?

F0 is the lowest frequency in a periodic waveform; also called the first harmonic frequency.
Speech F0 variations contribute to prosody, and in tonal languages they help distinguish lexical categories. Several musical applications need F0 estimation, such as automatic score transcription or real-time interactive systems, but here again the imperfect reliability of available methods is an obstacle. F0 is a useful ingredient for a variety of signal processing methods, for example, F0-dependent spectral envelope estimation. Finally, a fairly recent application of F0 is as metadata for multimedia content indexing.

Discussion on Proposed Methods:

The paper is presented in a progression. The classic autocorrelation algorithm (ACF) is presented first, its error mechanisms are analyzed, and then a series of 5 improvements are introduced to reduce error rates. Then the author combines all 6 steps into one called YIN. The name YIN (from “yin” and “yang” of oriental philosophy) alludes to the interplay between autocorrelation and cancellation that it involves.

ACF:

The ACF method compares the signal to its shifted self. The ACF is the Fourier transform of the power spectrum, and can be seen as measuring the regular spacing of harmonics within that spectrum. In response to a periodic signal, the ACF shows peaks at multiples of the period. The autocorrelation method chooses the highest non-zero-lag peak by exhaustive search within a range of lags. Obviously if the lower limit is too close to zero, the algorithm may erroneously choose the zero-lag peak. Conversely, if the higher limit is large enough, it may erroneously choose a higher-order peak.

Difference function:

Where ACF is sensitive to amplitude change, the difference function is immune to this particular problem, as amplitude changes cause period-to-period dissimilarity to increase with lag in all cases. However, using difference function has the additional appeal that this function is more closely grounded in the signal model and paves the way for the next two error-reduction steps, the first of which deals with too high errors and the second with too low errors.

Cumulative mean normalized difference function:

The difference function is zero at zero lag and often nonzero at the period because of imperfect periodicity. Unless a lower limit is set on the search range, the algorithm must choose the zero-lag dip instead of the period dip and the method must fail. Even if a limit is set, a strong resonance at the first formant F1 might produce a series of secondary dips, one of which might be deeper than the period dip. A lower limit on the search range is not a satisfactory way of avoiding this problem because the ranges of F1 and F0 are known to overlap. The solution is cumulative mean normalized difference function. It reduces too high errors. A second benefit is to do away with the upper frequency limit of the search range, no longer needed to avoid the zero-lag dip. A third benefit is to normalize the function for the next error-reduction step.

Absolute Threshold:

It easily happens that one of the higher-order dips of the difference function is deeper than the period dip. If it falls within the search range, the result is a subharmonic error, sometimes called octave error. The solution the authors proposes is to set an absolute threshold and choose the smallest value of that which gives a minimum of difference deeper than that threshold. If none is found the global minimum is chosen.

Parabolic Interpolation:

The absolute threshold only works if the period is a multiple of the sampling period. If not, the estimate may be incorrect by up to half the sampling period. Worse, the larger value of difference sampled away from the dip may interfere with the process that chooses among dips, thus causing a gross error.
A solution to this problem is parabolic interpolation. Each local minimum of difference and its immediate neighbors is fit by a parabola, and the ordinate of the interpolated minimum is used in the dip-selection process.

Best Local Estimate:

Best Local Estimate is reminiscent of median smoothing or dynamic programming techniques, but differs in that it takes into account a relatively short interval and bases its choice on quality rather than mere continuity.

YIN:

The combination of above methods is YIN. It performs best of all methods over all the databases. Averaged over databases, error rates are smaller by a factor of about 3 with respect to the best competing method. Error rates depend on the tolerance level used to decide whether an estimate is correct or not. For YIN about 99% of estimates are accurate within 20%, 94% within 5%, and about 60% within 1%.

The author also tried some extensions to YIN at informal level and thus did not provide the results for them. The extensions are – Variable Amplitude, Variable F0, Additive Noise: slowly varying DC, Additive Noise: Periodic, Additive Noise: different spectrum from target, Additive Noise: same spectrum as target. These extensions did not prove to have any good impact. The key to the success of YIN was Cumulative mean normalized difference function that allows it to escape from bias parameters so that the two types of errors can be addressed independently. Parabolic interpolation gives sub-sample resolution.

Conclusion:

YIN has few parameters and they don't require fine tuning. In contrast to most other methods no upper limit need to be put on F0 search range. The method is relatively simple and may be implemented efficiently and with low latency, and may be extended in several ways to handle several forms of aperiodicity that occur in particular applications. Finally, an interesting parallel may be drawn with models of auditory processing.