Title: YIN, a fundamental frequency estimation for speech and music
Author: Alain de Cheveigne, Hideki Kawahara

Summary

This paper is focused on pitch detection. Based on the well-known autocorrelation method, the authors present difference function and explore several ways of bettering accuracy rates. The proposed framework first compared autocorrelation and difference function methods, then put forward cumulative mean normalized difference function to lower too high errors and absolute threshold to avoid too low errors. After that, parabolic interpolation and best local estimation are described to finer pitch estimations. Experiments among different methods prove the effectiveness of the proposed method.

Good Things about this Paper

This method is simple and efficient, with few parameters to tune and easy to implement. Also, YIN has an open-ended frequency estimation range, which outperforms the previous method without continuity problems. Lastly, this algorithm has a great scope of application. For the audio signal processing domain, it performs well on both speech signals and music signals; it can also potentially be used in the macroeconomics filed, as autocorrelation is adapted and widely used in some problems.

Major Comments

1. The proposed framework requires an absolute threshold to pick one estimation form the calculated difference function value in step 4. This process is problematic because it outputs only one estimation and discarded the others. However, these abandoned frequency candidates may contain useful information. The rigid absolute threshold picking process leads to a loss of this information which unfortunately cannot be restored. A probable improvement of this process is considering its statistical distribution and pick the most possible one instead.
2. Another drawback of this method is it's not robust to noise. And if the source audio is not monophonic, as a singer with accompanying piano sound, or if in some string instrument this might be the case that one string keeps ringing while the melody turns to a different pitch, the algorithm may not work as expected. These problems lie in every time-domain pitch estimator, so I suggest considering hybrid architectures to make use of the advantages of other domain methods. Here I recommend Noise Robust Pitch Tracking by Subband Autocorrelation Classification by Daniel P. W. Ellis.
3. Also, there are some ambiguities in the evaluation process. In section III, the author described that estimates that seemed incorrect were removed from the statistics, but they did not mention what proportion of estimates is removed and how it is to be considered removable or not. Furthermore, values differed by no more than 20% is considered as correct estimation, is this range too large? Lastly, some previous methods like fxac or fxcep have stated their working frequency range, but this is not the frequency estimated in this study. So, these methods are not suitable for compare as baselines.
4. To some extent, this article is poorly arranged with verbosities everywhere and literature review paragraphs in every section, in some they might not be expected to present, like in the method section. One interesting thing is that the author stated in the last paragraph of the Discussion part said the only thing original among these steps is step 6. However, step 6 is somehow considered useless.

Minor Comments

1. No horizontal label on figure 1, 3, 5, 6
2. No units specified on figure 1, 3, 5, 6, 9