MULTIPLE-F0 ESTIMATION FOR MIREX 2015

Li Su and Yi-Hsuan Yang
Research Center for Information Technology Innovation, Academia Sinica , Taipei, Taiwan

ABSTRACT

In this submission for MIREX 2015 we propose a novel approach, referred to as combined frequency and periodicity (CFP), that detects pitches according to the agreement of a harmonic series in the frequency domain and a subharmonic series in the lag (quefrency) domain. The system takes the log-scaled spectrum as the frequency representation in the frequency domain and the generalized cepstrum as the periodicity representation in the lag domain. This approach aggregates the complementary advantages of the two feature domains and improves the robustness of the pitch detection function to the interference of the overtones of simultaneous pitches.

1. INTRODUCTION

Early in 1951, Licklider pointed out complementary advantages of frequency and periodicity in pitch perception have been discussed in the *duplex theory* of pitch perception [4]:

> "The stimulus basis for pitch is also duplex: On the one hand we have frequency, on the other hand, periodicity. That frequency and period are reciprocally related is not sufficient for throwing one away and examining only the other, for with each one is associated a method of analysis."

The idea of combining frequency and periodicity for computational single-pitch detection has started by Peeters [5] and then Emiya *et al.* [2]. Specifically, they determined the likelihood of a pitch candidate f of being the true F0 (i.e. the *pitch salience* of f) according to the multiplication of a *spectral representation* $U(f)$ and a *temporal representation* $V(\tau)$, where τ is a unit in the lag domain and can be mapped into the frequency domain via the relationship $f = 1/\tau$. As there is no simultaneous pitches for the monophonic music considered in single-pitch detection, the detection result is simply the pitch candidate with the maximal pitch salience in that given frame:

$$\hat{f}_0 = \mathrm{argmax}_f \left(U\left(f\right) V\left(1/f\right) \right) . \qquad (1)$$

The spectral representation $U(f)$, such as the magnitude spectrum, indicates the saliency of the *fundamental frequency* and its harmonics. In contrast, the temporal representation $V(\tau)$, such as the logarithm cepstrum or the autocorrelation function (ACF), indicates the saliency of the *fundamental period* (i.e. the smallest time interval the signal repeats itself) and its multiples. Mapping $V(\tau)$ into the frequency domain, $V(1/f)$ indicates the fundamental frequency and its subharmonics. Determining pitches by $U(f)$ alone is prone to octave errors [3, 6] (but is robust to sub-octave errors), while $V(1/f)$ alone is prone to sub-octave errors but not octave errors. Combining the two representations reduces the errors and leads to accuracy comparable to the famous YIN algorithm [1] for single-pitch detection [5].

To better exploit the synergy between frequency and periodicity for MPE, we propose to consider the behaviors of the full harmonic and subharmonic series in the spectral and temporal representations, and makes use of the agreement between the fundamental frequency and fundamental period for measuring pitch salience. We claim that the presence of a true pitch at frequency f_0 would imply the following three conditions:

1. A prominent harmonic series in $U\left(f\right)$.
2. A prominent subharmonic series in $V\left(1/f\right)$.
3. The fundamental frequency of the harmonic series and the fundamental period of the subharmonic series match at the same frequency f_0.

Details on how to design appropriate feature representations $U\left(f\right)$ and $V\left(1/f\right)$ which can effectively reveal pitch information and the procedure in determining concurrent pitch events are described in [7].

2. REFERENCES

[1] De Alain Cheveigné and Hideki Kawahara. YIN, a fundamental frequency estimator for speech and music. *J. Acoustical Society of America*, 111(4):1917–1930, 2002.

[2] Valentin Emiya, Bertrand David, and Roland Badeau. A parametric method for pitch estimation of piano tones. In *Proc. IEEE Int. Conf. Acoust. Speech Signal Proc.*, pages 249–252, 2007.

[3] Cheng-Te Lee, Yi-Hsuan Yang, and Homer H Chen. Multipitch estimation of piano music by exemplar-based sparse representation. *IEEE Trans. Multimedia*, 14(3):608–618, 2012.

[4] J. C. R. Licklider. A duplex theory of pitch perception. *J. Acoustical Society of America*, 23(1):147–147, 1951.

[5] Geoffroy Peeters. Music pitch representation by periodicity measures based on combined temporal and spectral representations. In *Proc. IEEE Int. Conf. Acoust. Speech Signal Proc.*, 2006.

[6] Li Su and Y.-H. Yang. Resolving octave ambiguities: A cross-dataset investigation. In *Int. Conf. Computer Music/Sound Music Computing (ICMC/SMC)*, pages 962–966, 2014.

[7] Li Su and Yi-Hsuan Yang. Combining spectral and temporal representations for multipitch estimation of polyphonic music. *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, 23(10):1600–1612, 2015.