Multi-pitch Determination Algorithm Based on Mixture Laplacian Distribution

Xueliang Zhang and Wenju Liu

National Laboratory of Pattern Recognition (NLPR)
Institute of Automation, Chinese Academy of Sciences Beijing, China, 100080
{xlzhang, lwj}@nlpr.ia.ac.cn

ABSTRACT

In this paper, a multi-pitch determination algorithm based on mixture Laplacian distribution (MLD) is proposed. MLD replaces the autocorrelation function (ACF) of correlogram which shows the possibility of the lag being pitch period. The peaks of summary MLDs indicate the multiple pitch periods. Compared with summary correlogram, summary MLDs has better resolution and less fake peaks which do not correspond to the pitch period. The proposed algorithm is evaluated on a database of speech utterances mixed with various types of interference. The comparisons show that our algorithm has better performances.

***Index Terms*—**multi-pitch determination algorithm, computational auditory scene analysis, mixture Laplacian distribution, correlogram.

1. INTRODUCTION

In 1951, Licklider [1] introduced a method of auto-correlation analysis in his duplex theory of pitch perception, the essence of which is that our auditory system employs both frequency analysis and autocorrelation analysis for sensation of pitch. Frequency analysis is performed by the cochlea via an array of bandpass filters, and autocorrelation analysis is performed on the activity of auditory nerve fibers, resulting in a two-dimensional pattern: characteristic frequency and time lag. These banks of ACFs are also called correlgoram. The pitch is then extracted from nerve firing patterns by finding a time lag with maximal peaks in the autocorrelation functions. Meddis and Hewitt [2] took a further step to propose a summary autocorrelation function (SACF) or called summary correlogram and pointed out that the highest point of the SACF indicates the perceived pitch. Meddis and Hewitt argued that many phenomena about pitch perception could be explained with their model including the missing fundamental, ambiguous pitch, the pitch of interrupted noise, inharmonic components, and the dominant region of pitch. However, using SACF for multi-pitch detection suffers from unsatisfactory facts. One is that the peak of ACF is rather wide [3]. It leads to that the peak corresponding to a pitch period is likely submerged by others.

To overcome the drawbacks, we proposed an algorithm based on the summary MLD. The MLD is a Laplacian mixture distribution which shows the possibility of the lag being the pitch period. The mean of Laplacian distribution equals to the peak position of corresponding ACF. We know that the Laplacian distribution has shaper peak around its mean than ACF. Hence, the resolution problem is solved to some extent. The reason for the generation of "fake" peak is that the ACF has peaks on its all integer multiples of the period. In fact, if we only reserve the peak on pitch period for each ACF, the problem is solved quite well. In MLD, it corresponds to that the mixture according to other simultaneous frequency components.

2. ALGORITHM DESCRIPTION

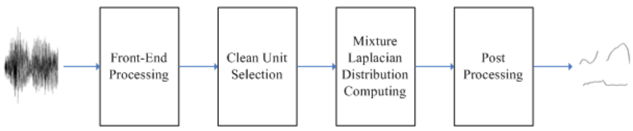

Fig.1. Diagram of proposed multi-pitch detection algorithm

2.1 Front-End processing

At first, the typical processing is performed. Input signal is decomposed by 128-channel gammatone filterbank whose center frequencies are quasi-logarithmically spaced from 80 Hz to 5 kHz and bandwidths are set according to equivalent rectangle bandwidth (ERB). Then the outputs of filterbank are transited into neural firing rate by hair cell model. Meanwhile, the envelope of filter response is extracted by performing Hilbert transform and then the squared Hilbert envelope is filtered by a filtered with passband [50Hz, 550Hz].

The time-frequency (T-F) units are formed in each channel with 20 ms window and 10 ms overlap between consecutive frames. Within each T-F units, the following features are extracted, normalized ACF, normalized envelope ACF, cross channel correlation. To remove the multiple peaks, the normalized envelope ACF is further processed into enhanced envelope ACF by method in [6]. For the different characteristic, these T-F units are classified into two categories: resolved and unresolved units. Resolved unit is defined as the unit dominated by single harmonic and unresolved one is dominated by multiple harmonics. Our previous study [4] shows that the envelope fluctuation is relative small in resolved unit and large in unresolved unit. And the same feature carrier-to-envelope energy ratio is employed for unit classification.

2.2 Clean unit selection

The T-F unit is possibly dominated by several sound sources. In this case, no peak of ACF corresponds to pitch periods. These units are deleterious to pitch perception and should be excluded. In [5], unit selection is based on the shape of ACF. Here, a different mechanism is introduced. The T-F units are first marked as candidates by their features and then these selected units are merged into segments by the time and channel continuity. At last, the units in large segment (composed of more than 20 units) are selected as clean.

2.3 Mixture Laplacian distribution

The computation of MLD in T-F unit U_{cm} at channel c and frame m is described as a by following equations.

$$\begin{aligned} p_{c,m}(\tau|\Theta) &= \sum_{n=1}^{N_p} \lambda_{c,m}(n) \times L\left(\tau \left| \mu_{c,m}(n), \sigma_{c,m}(n)\right.\right) \\ &= \sum_{n} \lambda_{c,m}(n) \times \frac{1}{2\sigma_{c,m}(n)} \exp\left(-\frac{|\tau - \mu_{c.m}(n)|}{\sigma_{c,m}(n)}\right) \end{aligned} \quad (1)$$

where, $\tau \in [0.5ms, 13.3ms]$ for the pitch range is from 75 Hz to 2 kHz; N_p is the number of peaks of ACF; $\mu_{c,m}(n)$, $\sigma_{c,m}(n)$ and $\lambda_{c,m}(n)$ stand for the mean, standard variance and mixture coefficient of n *th* Laplacian function respectively; $\Theta = <\mu_{c,m}(1)\ldots\mu_{c,m}(N_p); \sigma_{c,m}(1)\ldots\sigma_{c,m}(N_p); \lambda_{c,m}(1)\ldots\lambda_{c,m}(N_p)>$ is the parameter set.

Fig.2. shows the MLDs (after different iterative times) of a test signal with three complex sounds with fundamental frequencies 210 Hz, 390 Hz and 550 Hz. We can see that the "fake" peaks of summary MLD are gradually vanished with iteration increasing. Fig.6. shows the comparison between the conventional periodogram based on summary ACF (a) and periodogram based on summary MLD (b). From Fig.6.(a), we can see that the peaks of different utterance overlap seriously because of low resolution and there are numerous "fake" peaks. In (b), the problems are solved.

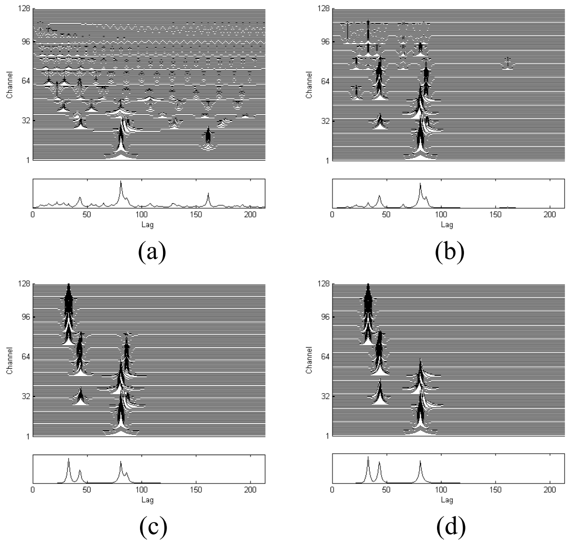

Fig.2. Results of MLDs on simulation. The upper panels are MLDs and lower panels are summary MLD. The test signal is the mixture of three complex sounds with fundamental frequencies 210 Hz, 390 Hz and 550 Hz. (a) initial MLDs; (b) after 1 iteration; (c) after 5 iterations; (d) more than 30 iterations

2.4 Post processing

The pitch estimation is based on summary of MLDs. The period candidates are sorted sort in descending order according to the heights of the peaks. The distant between estimated pitches is larger than 6%. And the peak height of estimated pitch is larger than 0.001.

5. REFERENCES

[1] J. C. R. Licklider, A duplex theory of pitch perception. *Experientia* 7(4), 128–134. 1951.

[2] R. Meddis and M. J. Hewitt. "Virtual pitch and phase sensitivity of a computer model of the auditory periphery. I: pitch identification," *J. Acoust. Soc. Amer.*, no. 6, pp. 2866–2882. 1991

[3] A. de Cheveigné, "Pitch and the narrowed autocoindidence histogram," in *Proc. Int. Conf. Music Perception and Cognition*, Kyoto, Japan, pp. 67–70, 1989.

[4] X. L Zhang, W. J. Liu, P. Li and B. Xu, "Monaural voiced speech segregation based on elaborate harmonic grouping strategy," in *Proc. IEEE ICASSP* 2009, pp. 4661-4664 Taiwan.

[5] M. Y. Wu, D. L. Wang, and G. J. Brown, "A Multipitch Tracking Algorithm for Noisy Speech," *IEEE Trans. Speech And Audio Process.*, vol. 11, No. 3. pp. 229-241 2003.

[6] T. Tolonen and M. Karjalainen, "A computationally efficient multipitch analysis model," *IEEE Trans. S.A.P.*, vol. 8, pp.708–716, 2000.