A NOVEL PITCH DETERMINATION ALGORITHM BASED ON GENERALIZED CORRELATION FUNCTION

Jian-Wu Xu, Jose C. Principe

Computational NeuroEngineering Laboratory
Department of Electrical and Computer Engineering, University of Florida
Gainesville, FL 32611 U.S.A.

ABSTRACT

In this paper, we propose a novel pitch determination algorithm (PDA) based on the newly introduced concept of the generalized correlation function called *correntropy* [1]. Correntropy contains a positive definite kernel function which nonlinearly transforms the original signal into a high dimensional reproducing kernel Hilbert space (RKHS), and calculates very efficiently the generalized correlation in that RKHS. By incorporating the kernel function, correntropy is able to utilize higher order statistics to enhance the resolution of pitch estimation. The proposed PDA calculates the summary of correntropy functions from the outputs of an equivalent rectangular bandwidth (ERB) filter bank. We present simulations on pitch determination for single vowel and double vowels cases. Simulations show that correntropy exhibits much better resolution than the conventional autocorrelation method for pitch determination.

1. INTRODUCTION

Pitch, or fundamental frequency, is an important parameter of speech signals. Accurate determination of pitch plays a vital role in acoustical signal processing and has a wide range of applications in related areas such as coding, synthesis, speech recognition, etc. Numerous pitch determination algorithms have been proposed in the literature [2]. In general, they can be categorized into three different classes: frequency-domain, time-domain, and time-frequency domain algorithms.

Frequency-domain PDAs estimate pitch by using the harmonic structure in the short-time spectrum. This approach includes component frequency ratios, filter-based method, cepstrum analysis and multi-resolution method. Pitch determination algorithms such as harmonic sieve, harmonic product spectrum, and sub-harmonic summation [2] also fall into this category.

Time-domain PDAs operate on the temporal structure of signals directly. These include, but are not limited to

zero-crossing rate, peak and valley positions, autocorrelation. The autocorrelation model appears to be one of the most popular PDAs for its simplicity, explanatory power, and physiological plausibility [3]. For a given signal $x(n)$ with N samples, the autocorrelation function $R(\tau)$ is defined as

$$R(\tau) = \frac{1}{N} \sum_{n=0}^{N-1} x(n)x(n+\tau), \quad (1)$$

where τ is the delay parameter. For dynamical signals with changing periodicities, a short-time causal window can be included to compute the periodicities of the signal within the window.

Other similar models can be obtained by replacing the multiplication by subtraction (or excitatory by inhibitory neural interaction) in the autocorrelation function. For instance, the average magnitude difference function (AMDF) [4] is defined as

$$\text{AMDF}(\tau) = \frac{1}{N} \sum_{n=0}^{N-1} |x(n) - x(n+\tau)|.$$

Cheveigné proposed the squared difference function (SDF) in [3] as

$$\text{SDF}(\tau) = \frac{1}{N} \sum_{n=0}^{N-1} (x(n) - x(n+\tau))^2.$$

All these PDAs based on autocorrelation function suffer from at least one unsatisfactory fact: the peak corresponding to the period for a pure tone is rather wide [5]. This low resolution in pitch estimation results from the fundamental time-frequency uncertainty principle [6]. To overcome this drawback, Brown *et al.* presented the “narrowed” autocorrelation function to improve the resolution of the autocorrelation function for musical pitch extraction [7]. The “narrowed” autocorrelation function includes terms corresponding to delays at 2τ , 3τ , etc, in addition to the usual term with delay τ in (1). However it requires an increase in the length of the signal and less precision in time. How to select the number of delay terms also remains a problem.

On the other hand, correlation and power spectrum based pitch determination algorithms only provide second-order statistics. In many applications where non-gaussianities and nonlinearities are present, all these second-order statistical methods fail to provide all the information of the processes. Higher-order statistics have been used in pitch determination. Moreno *et. al.* applied third order cumulants to extract pitch from noisy speech [8].

In this paper, we propose a new pitch determination algorithm based on a generalized autocorrelation function, introduced in [1] called *correntropy*. Correntropy uses a positive definite kernel function to nonlinearly project the original signal into a high dimensional reproducing kernel Hilbert space. This implicit nonlinear transformation preserves the periodic characteristics of signal because periodicity information of an auditory signal is not disturbed by most nonlinear transformations even the bandwidth, amplitude and phase information change [9]. Correntropy has similar properties as the conventional correlation function but contains a compact form of higher-order statistical information of signal with the presence of a kernel function. The proposed PDA method is applied after the acoustical signal is processed by an equivalent rectangular bandwidth(ERB) filter bank in the time domain. The ERB filter bank acts as a cochlear model to transform a one dimensional acoustic signal into a two dimensional map of neural firing rate as a function of time and place [9]. The correntropy function for each channel is calculated and the summation across channels provides the pitch information.

The paper is organized as follows. In Sec. 2, we briefly introduce the concept of correntropy and its relevant properties. The proposed PDA is presented in Sec. 3. We applied our method in determining pitches for one vowel and double vowels cases in Sec. 4. Some specific issues are addressed in Sec. 5, and we conclude our work in Sec. 6.

2. CORRENTROPY FUNCTION

Conventional autocorrelation function only captures second-order statistics of random processes. This way of partial characterization provides practical measurements of random processes and also suits well linear operations on random processes. But when the problem deals with nonlinear systems or non-Gaussian signals, the second order statistics are not sufficient. A new generalized correlation function has been proposed to represent higher order statistics of a random process [1]. This generalized correlation function can also be viewed as a “correlation” function for the transformed random process in a high dimensional reproducing kernel Hilbert space via nonlinear mapping imposed by the kernel. The *linear* operations on the transformed random process in the RKHS correspond to *nonlinear* operations on the original random process. Therefore, this generalized

correlation function also quantifies the nonlinearity in the random processes.

Definition: Given a random process $\{x(t) : t \in \mathbb{T}\}$ with t typically denoting time and $\mathbb{T}$ being an index set of interest, the *correntropy function* is defined by [1]

$$V(t, s) = E[\kappa(x(t), x(s))] \quad (2)$$

for each t and s in $\mathbb{T}$, where E denotes the statistical expectation operator and $\kappa(\cdot, \cdot)$ is a symmetric positive definite kernel function. The widely used kernel function is the Gaussian kernel which is given by

$$\kappa(x(t), x(s)) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ -\frac{(x(t) - x(s))^2}{2\sigma^2} \right\}. \quad (3)$$

Applying Taylor series expansion for the Gaussian kernel, we can rewrite the correntropy function as

$$V(t, s) = \frac{1}{\sqrt{2\pi}\sigma} \sum_{k=0}^{\infty} \frac{(-1)^k}{(2\sigma^2)^k k!} E[(x(t) - x(s))^{2k}] \quad (4)$$

which contains all the even-order moments of the random variable $(x(t) - x(s))$. Therefore the correntropy function partially characterizes higher-order statistics of random processes.

The correntropy function satisfies the following properties:

$$V(t, s) = V(s, t), \quad (5)$$

$$V(t, t) > 0, \quad (6)$$

$$|V(t, s)| < \sqrt{V(t, t)V(s, s)}. \quad (7)$$

These properties are very similar to those of conventional correlation function. Moreover, the correntropy function is also positive definite. This can be easily seen from the definition (2) since it includes a positive definite kernel function. The positive definiteness is fundamental to correntropy because all the properties above can be deduced from this one and it also shows that correntropy function is indeed a correlation function for a Gaussian process in the transformed domain.

In order to obtain a univariate correntropy function $V(t + \tau, t) = V(\tau, 0)$, we see from (4) that all the even-order moments should be time shift invariant. This is a much stronger condition than the wide sense stationarity required by conventional correlation function.

Since the correntropy function uses kernel functions, an analysis from the kernel point of view would provide a different perspective. Any symmetric positive definite kernel function can be expanded as eigen-decomposition, which is an important special case of *Mercer's theorem* [10],

$$\kappa(x(t), x(s)) = \sum_{k=0}^M \lambda_k \varphi_k(x(t)) \varphi_k(x(s)),$$

where λ_k are the eigenvalues, φ_k are the corresponding eigenfunctions, and M is the dimensionality of the reproducing kernel Hilbert space (RKHS). In theory, the dimensionality M can be infinitely large. It follows that the kernel eigen-decomposition can be rewritten as an inner product between two vectors,

$$\begin{aligned}\kappa(x(t), x(s)) &= \langle \Phi(x(t)), \Phi(x(s)) \rangle \\ \Phi : x(t) &\mapsto \sqrt{\lambda_k} \varphi_k(x(t)), \quad k = 1, 2, \dots, M.\end{aligned}$$

The nonlinear mapping Φ transforms the random process from the *input space* into a high dimensional RKHS, called *feature space*, induced by the kernel function. Each vector $\Phi(x(t))$ in the RKHS consists of all the eigenvalues and corresponding eigenfunctions evaluated at different instants. Accordingly, we can rewrite the correntropy function as

$$V(t, s) = E[\langle \Phi(x(t)), \Phi(x(s)) \rangle].$$

The correntropy function for any random process can be viewed as a “standard” correlation function for the transformed random process. The higher-order statistics for any random process in the input space turns out to be “second-order” statistics in the feature space. In fact, all the previous properties of correntropy function can be derived in a kernel framework. For example, property (6) can be shown that $V(t, t) = E[\|\Phi(x(t))\|^2]$ which means that $V(t, t)$ is nothing but the expectation of norm square of transformed random process. Property (7) is the generalized *Cauchy-Schwarz inequality* in the reproducing kernel Hilbert space,

$$E[\langle \Phi(x(t)), \Phi(x(s)) \rangle] < \sqrt{E[\|\Phi(x(t))\|^2] E[\|\Phi(x(s))\|^2]}.$$

The correntropy function also has an intriguing connection to the information theory [1]. This is the direct consequence of its higher-order statistics characterization of random processes. For a pair-wise independent random process $\{x(t) : t \in \mathbb{T}\}$, the correntropy function can be computed as

$$\begin{aligned}V(t, s) &= \langle E[\Phi(x(t))], E[\Phi(x(s))] \rangle \\ &= \left\langle \frac{1}{N} \sum_{i=1}^N \Phi(x(i)), \frac{1}{N} \sum_{j=1}^N \Phi(x(j)) \right\rangle \\ &= \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \kappa(x(i), x(j)).\end{aligned}\quad (8)$$

We use sample mean to estimate the statistical expectation and the independence property in the computation above. The quantity (8) is called *information potential* when Parzen window estimator is used to estimate quadratic Renyi’s entropy [11, 12]. Hence we coined the name *correntropy* to show the similarity with correlation but to indicate that the mean over the lags is the argument of Renyi’s quadratic entropy [1].

3. PDA BASED ON CORRENTROPY

Our pitch determination algorithm first uses cochlear filtering to peripherally process the speech signal. This is achieved by a bank of 64 gammatone filters which are distributed in frequency according to their bandwidths. The bandwidth increases quasi-logarithmically with respect to the center frequency. The center frequencies of each filter are equally spaced on the equivalent rectangular bandwidth scale between 80 Hz and 4000 Hz. The cochlear model has been extensively used in pitch determination [9, 13].

The periodicity analysis is done by computing the correntropy function of the output of each cochlear frequency channel,

$$V_i(\tau) = \frac{1}{N} \sum_{n=0}^{N-1} \kappa(x_i(n), x_i(n + \tau)), \quad (9)$$

where i stands for channel number. If a signal is periodic, the correntropy function (9) for each channel will exhibit a peak in the time lag domain at the corresponding period. The position of the peak corresponds to the perceived pitch period. In order to emphasize this pitch related structure, the correntropy functions are summed up across all the channels to form a “pooled” or “summary” correntropy function,

$$W(\tau) = \sum_i V_i(\tau).$$

For mixtures of concurrent sound sources with different fundamental frequencies, the summary correntropy function will show peaks at the periods of each sources.

Compared to the conventional correlogram model [9, 13, 14], our pitch detector is able to locate the same period information as the correlogram, but has much narrower peaks. Hence the proposed method enhances the resolution of pitch determination. On the other hand, the computational complexity of our method remains the same level as correlogram, while the “narrowed” autocorrelation function in [7] increases computational complexity by including more delay terms.

4. EXPERIMENT

In this section, we compare our method with the conventional autocorrelation function [9, 13] in determining pitches for single speaker and two combined speakers uttering different vowels. For a fair comparison, we did not apply any post-processing to the correlogram. The correntropy functions (9) and conventional autocorrelation functions (1) are presented after the same cochlear model.

Fig. 1 and Fig. 2 present the results for a single vowel /a/ with fundamental frequency at 100Hz. The upper plots are the image of correlation functions (called correlogram [9])

Fig. 1. Image of the correlogram (top) for the vowel /a/ with fundamental frequency at 100 Hz. The summary correlogram (bottom) shows a peak at the autocorrelation lag at time 10ms corresponding to the period of the vowel.

and correntropy functions after the same cochlear model respectively. The bottom figures are the summary of those two. The kernel size σ in the Gaussian kernel (3) has been chosen to be 0.01 (we will discuss more details on this in Sec. 5). The image of correntropy functions clearly shows a single narrow stripe across all the frequency channels which concentrates most of the energy. The pooled correntropy function provides a single large narrow peak at 10ms corresponding to the pitch of the vowel /a/. While the conventional correlogram yields a much wider spread of energy across the frequency channels. Though the pooled correlation functions also shows a peak at 10ms, it is much wider than that of pooled correntropy functions and there are other peaks which might confuse the outcome. Our proposed method clearly outperforms the conventional correlogram for the single vowel case.

In Fig. 3 and Fig.4, we compare the same methods in a mixture of two vowels consisting of /a/ and /i/ with fundamental frequencies at 100Hz and 150Hz respectively. The correlogram method only shows one peak corresponding to the pitch of the vowel /a/ while no indication of the other vowel /i/ at time of 6.7ms is provided. The summary of correlogram resembles that of single vowel case in Fig. 1. On

Fig. 2. Image of the correntropy functions (top) for the vowel /a/ with fundamental frequency at 100 Hz. The summary of correntropy (bottom) shows a large narrow peak at time 10ms corresponding to the period of the vowel.

the contrary, our method shows two peaks in the summary of correntropy functions in Fig. 4. The peak corresponding to the pitch of the vowel /i/ at 6.7ms is not as obvious as the large and narrow one of the vowel /a/ at 10ms since the energy ratio between vowels /a/ and /i/ is 6.5 times higher for /a/. However the image of correntropy functions clearly shows a white stripe at 6.7ms indicating the presence of vowel /i/. The kernel size σ is set to 0.05 in this experiment. This simulation demonstrates the superiority of our method over the conventional correlogram approach for combination of vowels with slight different pitches.

5. DISCUSSIONS

In the previous section, we chose a specific kernel size σ in the Gaussian kernel (3). Actually, the kernel size plays an important role in the performance of our method. It has been shown that kernel size controls the spread of the transformed signal in the RKHS [1]. If the kernel size is set too large, the correntropy function approaches to the correlation function and fails to detect any nonlinearity and higher order statistics intrinsic to the data; on the other hand, if the kernel size is too small, the correntropy function loses its

Fig. 3. Image of the correlogram (top) for a mixture of vowels /a/ and /i/ with fundamental frequencies at 100 Hz and 150Hz respectively. The summary correlogram (bottom) only shows one large peak at the autocorrelation lag at time 10ms corresponding to the period of the vowel /a/ while no indication of vowel /i/ is provided.

discrimination ability [15]. One way to select an optimal kernel size is given by the Silverman’s rule [16]

$$\sigma = 0.9AN^{-1/5},$$

where A is the smaller value between standard deviation of data samples and data interquartile range scaled by 1.34, and N is the number of data samples. To illustrate the effect of different kernel sizes, we simulate the summary of correntropy functions for the same experiments setup in Sec. 4 with different kernel sizes in Fig. 5 and Fig. 6. It can be seen that if the kernel size is large, $\sigma = 1$ here, the summary of correntropy functions approaches that of correlation functions shown in Fig. 1 and Fig. 3. As kernel size approaches the optimal kernel size, the summary of correntropy functions starts to present a large and narrow peak corresponding to the pitch of vowel /a/ and show the other vowel /i/. If the kernel size is too small, $\sigma = 0.001$ here, the summary of correntropy functions loses its ability to present two vowels. This is shown in the bottom plot of Fig. 6.

We also point out one connection of our method to the cancellation model [3]. Notice that the argument of the

Fig. 4. Image of the correntropy functions (top) for a mixture of vowels /a/ and /i/ with fundamental frequencies at 100 Hz and 150Hz respectively. The summary of correntropy (bottom) shows a large peak at time 10ms corresponding to the period of the vowel /a/ and a small peak at time 6.7ms.

Gaussian kernel function (3) also uses subtraction instead of multiplication. But the kernel function nonlinearly transforms the subtraction by an exponential function that brings the nonlinearity and higher order statistics.

6. CONCLUSION

A novel pitch determination algorithm is proposed based on the correntropy function. The pitch estimator computes the correntropy functions for each channel of an ERB filter bank, and adds across all the channels. simulations on single and double vowel cases show that the proposed method exhibits much better resolution than the conventional correlation function in pitch determination. This suggests that correntropy can discriminate better pitch when two different speakers speak in the same microphone. This is essential in computational auditory scene analysis (CASA). Since correntropy creates many different harmonics of each resonance present in the original time series due to the nonlinearity, it may also be useful for perceptual pitch determination. The future work will compare our algorithm to others on

Fig. 5. The summary of correntropy functions with different kernel sizes for single vowel /a/ with fundamental frequency at 100Hz.

Fig. 6. The summary of correntropy functions with different kernel sizes for a mixture of vowels /a/ and /i/ with fundamental frequencies at 100Hz and 150Hz respectively.

benchmark datasets and look into the noisy speech case.

7. REFERENCES

- [1] I. Santamaria, P. Pokharel, and J. C. Principe, "Generalized correlation function: Definition, properties, and application to blind equalization," vol. 54, no. 6, pp. 2187–2197, 2006.
- [2] W. J. Hess, *Pitch Determination of Speech Signals*. New York: Springer, 1993.
- [3] A. de Cheveigné, "Cancellation model of pitch perception," vol. 103, no. 3, pp. 1261–1271, Mar. 1998.
- [4] M. J. Ross, H. L. Shaffer, A. Cohen, R. Freudberg, and H. J. Manley, "Average magnitude difference function pitch extractor," vol. ASSP-22, no. 5, pp. 353–362, Oct. 1974.
- [5] A. de Cheveigné, "Pitch and the narrowed autocoincidence histogram," in *Proc. Intl. Conf. of Music Perception and Cognition*, Kyoto, Japan, 1989, pp. 67–70.
- [6] L. Cohen, *Time-Frequency Analysis*. New Jersey: Prentice Hall, 1995.
- [7] J. C. Brown and M. S. Puckette, "Calculation of a "narrowed" autocorrelation function," vol. 85, pp. 1595–1601, 1989.
- [8] A. Moreno and J. Fonollosa, "Pitch determination of noisy speech using higher order statistics," San Francisco, USA, Mar. 1992, pp. 133–136.
- [9] M. Slaney and R. F. Lyon, "On the importance of time - a temporal representation of sound," in *Visual Representations of Speech Signals*, M. Cooke, S. Beet, and M. Crawford, Eds. John Wiley & Sons, 1993, pp. 95–116.
- [10] J. Mercer, "Functions of positive and negative type, and their connection with the theory of integral equations," vol. 209, pp. 415–446, 1909.
- [11] J. C. Principe, D. Xu, and J. W. Fisher, "Information theoretic learning," in *Unsupervised Adaptive Filtering*, S. Haykin, Ed. John Wiley & Sons, 2000, pp. 265–319.
- [12] A. Rényi, "On measures of entropy and information," in *Selected paper of A. Rényi*. Budapest, Hungary: Akademiai Kiado, 1976, vol. 2, pp. 565–580.
- [13] G. J. Brown and D. Wang, "Modelling the perceptual segregation of double vowels with a network of neural oscillators," vol. 10, no. 9, pp. 1547–1558, 1997.
- [14] R. Meddis and M. Hewitt, "Virtual pitch and phase sensitivity of a computer model of the auditory periphery: I. pitch identification," vol. 89, pp. 2866–2882, 1991.
- [15] J. Xu, "Correntropy: Definition, properties, and application in principal component analysis," *Ph.D. proposal, University of Florida*, June 2006.
- [16] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*. New York: Chapman and Hall, 1986.