A Pitch Detector Based on a Generalized Correlation Function

Jian-Wu Xu, *Student Member, IEEE*, and Jose C. Principe, *Fellow, IEEE*

***Abstract*—This paper proposes a novel pitch determination algorithm (PDA) based on the newly introduced concept of a generalized correlation function called *correntropy*. Correntropy is a positive definite kernel function which implicitly transforms the original signal into a high-dimensional reproducing kernel Hilbert space (RKHS) in a nonlinear way, 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 computes the summary of correntropy functions from the outputs of an equivalent rectangular bandwidth (ERB) filter bank. We present simulations on pitch determination for a single vowel, double vowels, and a benchmark database test. Simulations show that correntropy exhibits much better resolution than conventional autocorrelation in pitch determination and outperforms other PDAs in the benchmark database test.**

***Index Terms*—Correntropy, pitch determination, reproducing kernel Hilbert space (RKHS).**

I. Introduction

PITCH, or the fundamental frequency $F0$, is an important parameter in speech signal analysis. 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, and others. Numerous pitch determination algorithms (PDAs) have been proposed in the literature [2]. In general, they can be categorized into three classes: time-domain, frequency-domain, and time–frequency domain algorithms.

Time-domain PDAs operate directly on the signal temporal structure. These include but are not limited to zero-crossing rate, peak and valley positions, and autocorrelation. The autocorrelation model appears to be one of the most popular PDAs for its simplicity, explanatory power, and physiological plausibility. For a given signal x_n with N samples, the autocorrelation function $R(\tau)$ is defined as

$$R(\tau) = \frac{1}{N}\sum_{n=0}^{N-1} x_n x_{n+\tau} \tag{1}$$

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

$$R(\tau, t) = \frac{1}{N}\sum_{n=0}^{N-1} x_n x_{n+\tau} w_t$$

where w_t is an arbitrary causal window that confines the autocorrelation function into a neighborhood of the current time. Other similar models can be obtained by replacing the multiplication by subtraction (or excitatory by inhibitory neural interaction) in the autocorrelation function such as the average magnitude difference function (AMDF) [3]. Cheveigné proposed the squared difference function (SDF) in [4] as

$$\mathrm{SDF}(\tau) = \frac{1}{N}\sum_{n=0}^{N-1} (x_n - x_{n+\tau})^2. \tag{2}$$

The weighted autocorrelation uses an autocorrelation function weighted by the inverse of an AMDF to extract pitch from noisy speech [5]. All these PDAs based on the autocorrelation function suffer from at least one unsatisfactory fact: the peak corresponding to the period for a pure tone is rather wide [6], [7]. This imposes a greater challenge for multiple $F0$ estimation since mutual overlap between voices weakens their pitch cues, and cues further compete with cues of other voices. The low resolution in pitch estimation results from the fundamental time–frequency uncertainty principle [8]. To overcome this drawback, Brown *et al.* presented a "narrowed" autocorrelation function to improve the resolution of the autocorrelation function for musical pitch extraction [9]. The "narrowed" autocorrelation function includes terms corresponding to delays at 2τ, 3τ, etc., in addition to the usual term with delay τ as

$$S_L(\tau) = \frac{1}{N}\sum_{n=0}^{N-1} (x_n + x_{n+\tau} + x_{n+2\tau} + \cdots + x_{n+L\tau})^2. \tag{3}$$

However, it requires an increase in the length of the signal and less precision in time. It also requires the *a priori* selection of the number of delay terms L.

Frequency-domain PDAs estimate pitch by using the harmonic structure in the short-time spectrum. Frequency-domain methodologies include component frequency ratios, filter-based methods, cepstrum analysis, and multiresolution methods. Pitch determination algorithms such as harmonic sieve [10], harmonic product spectrum [11], subharmonic summation [12], and subharmonic-to-harmonic ratio [13] fall into this category. Most frequency-domain pitch determination methods apply pattern matching [14]. Others use nonlinear or filtering preprocessing to generate or improve interpartial spacing and fundamental component cues. The frequency-domain PDAs have the advantage of efficient implementation with fast Fourier transform and theoretical strength of Fourier analysis. However, one weakness is

Manuscript received September 06, 2007; revised May 22, 2008. This work was supported in part by the National Science Foundation under Grant ECS-0601271 and by a Graduate Alumni Fellowship from the University of Florida. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. George Tzanetakis.

The authors are with Computational NeuroEngineering Laboratory, Electrical and Computer Engineering Department, University of Florida, Gainesville, FL 32611 USA (e-mail: jianwu@cnel.ufl.edu; principe@cnel.ufl.edu).

Digital Object Identifier 10.1109/TASL.2008.2002039

that they rely on the shape and size of the analysis window. Selection and adjustment of analysis windows remain a problem in estimation.

The time–frequency approach splits the signal over a filterbank, applies time-domain methods to each channel waveform, and the results are aggregated over channels. The summary, or "pooled," autocorrelation functions across all channels provides pitch information of the signal. Licklider first presented this idea as a pitch perception model [15]. Later, Lyon and Slaney further developed the methodology and called it *correlogram* [16], [17]. The correlogram is often the first-stage processor in a computational auditory scene analysis (CASA) system [18]. It has also been incorporated into a neural oscillator to segregate double vowels and multipitch tracking [7], [19]. The strength of correlogram in pitch estimation is that different frequency channels corresponding to different signal sources of different pitches can be separated, which makes it useful in multipitch estimation [7], [20]. Also individual channel weighting can be adapted to compensate for amplitude mismatches between spectral regions [21].

On the other hand, autocorrelation and power spectrum-based pitch determination algorithms only characterizes second-order statistics. In many applications where non-Gaussianities and nonlinearities are present, these second-order statistical methodologies might fail to provide all the information about the signals under study. Higher order statistics have been used in pitch determination. Moreno *et al.* applied higher order statistics to extract pitch from noisy speech [22], but only diagonal third-order cumulants were used for simplicity and computational efficiency which is given by

$$c(k) = \frac{1}{N}\sum_{n=0}^{N-1} x_n x_n x_{n+k}, \quad k = 0, \ldots, N-1$$

and pitch is found by applying autocorrelation function to the cumulants $c(k)$

$$G(\tau) = \frac{1}{2N}\sum_{k=-(N-1)}^{N-1} c(k)c(k+\tau). \tag{4}$$

In this paper, we propose a new pitch determination algorithm based on a generalized autocorrelation function which is called *correntropy* [1]. Correntropy uses a positive definite kernel function to nonlinearly project the original signal into a high dimensional reproducing kernel Hilbert space (RKHS), and calculates "a generalized correlation function" in that space. This implicit nonlinear transformation preserves the periodic signal characteristics because periodicity information is not disturbed by most instantaneous nonlinear transformations even if the bandwidth, amplitude, and phase information might change [17]. Correntropy is a function of two arguments with similar properties as the conventional correlation function, but contains a weighted combination of higher order statistical information of the input through the kernel function. Preliminary results have shown that it produces very peaky estimations of similarity and much narrower peaks corresponding to the pitch period than the conventional correlation function. Moreover, correntropy offers much smaller time–frequency bandwidth than the correlation function. Therefore, it improves the accuracy in determining the pitch period. Correntropy has also been successfully applied to various signal processing and machine learning problems such as blind equalization [1], minimum average correlation energy filter [23], principal component analysis [24], and others. The proposed PDA method is applied after the acoustical signal is processed by an equivalent rectangular bandwidth filter bank in the time domain. The equivalent rectangular bandwidth (ERB) filter bank acts as a cochlear model to transform a one-dimensional acoustical signal into a two-dimensional map of neural firing rate as a function of time and place [17]. The correntropy function for each channel is calculated and the summation across all the channels provides the pitch information. Therefore, correntropy is simple and readily integrated in the mainstream of PDA methods, only at a slightly higher computational cost. As a novel PDA, correntropy is able to offer much better resolution than the conventional autocorrelation function in pitch estimation. Moreover, our pitch determination algorithm can segregate double vowels without applying any complex model such as a neural oscillator [7].

This paper is organized as follows. In Section II, we briefly introduce the concept of correntropy and its relevant properties. The proposed PDA is presented in Section III. We applied our method in determining pitches for one-vowel and double-vowels cases, and a benchmark database in Section IV. Some specific issues are addressed in Section V, and we conclude our work in Section VI.

II. Correntropy Function

Definitions: 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 generalized correlation function, called the *correntropy function*, is defined as [1]

$$V(t,s) = E\left[\kappa(x_t, x_s)\right] \tag{5}$$

and the generalized covariance function, called the *centered correntropy function* is defined as

$$U(t,s) = E_{x_t x_s}\left[\kappa(x_t, x_s)\right] - E_{x_t}E_{x_s}\left[\kappa(x_t, x_s)\right] \tag{6}$$

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. Notice that the correntropy is the joint expectation of $\kappa(x_t, x_s)$, while the centered correntropy is the difference between the joint expectation and product of marginal expectations of $\kappa(x_t, x_s)$.

In functional analysis, a symmetric positive definite kernel is a special type of bivariate function. In the literature, the sigmoidal, Gaussian, polynomial and spline kernels are among the mostly used symmetric positive definite kernel functions in the area of machine learning, function approximation, density estimation, support vector machine, and others [25]. The widely used Gaussian kernel function is given by

$$\kappa(x_t, x_s) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(x_t - x_s)^2}{2\sigma^2}\right\} \tag{7}$$

where σ is the variance, called the kernel width parameter (or kernel size). We will apply the Gaussian kernel throughout the paper without loss of generality. Applying the Taylor series expansion to the Gaussian kernel, we can rewrite the correntropy function as

$$V(t,s) = \frac{1}{\sqrt{2\pi}\sigma}\sum_{k=0}^{\infty}\frac{(-1)^k}{(2\sigma^2)^k k!}E\left[(x_t - x_s)^{2k}\right] \tag{8}$$

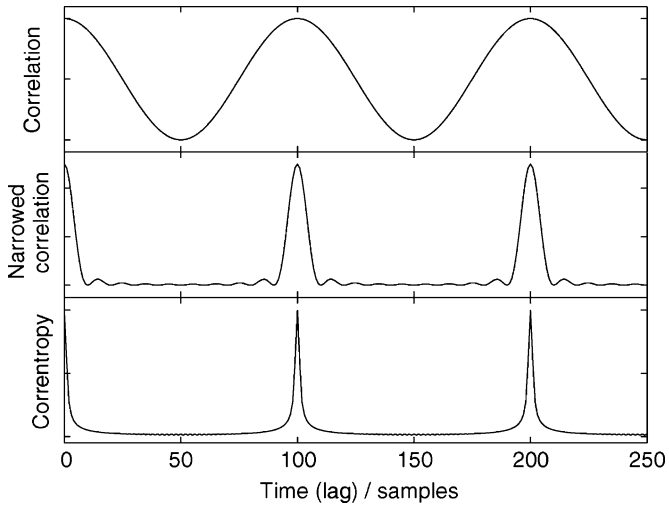


Fig. 1. Autocorrelation, narrowed autocorrelation with $L = 10$, and correntropy functions of a sinusoid signal.

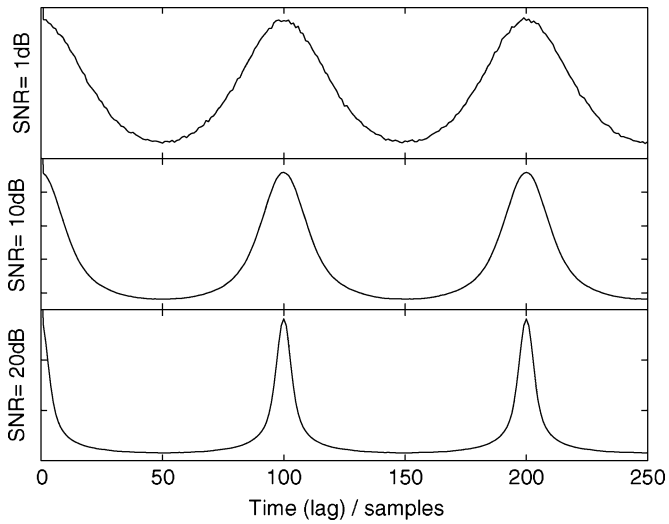


Fig. 2. Correntropy functions of a sinusoid signal with white noise at different SNR levels.

which contains all the even-order moments of the random variable $(x_t - x_s)$. Obviously different kernel functions would yield different expansions, but all the kernel functions mentioned above are nonlinear and therefore include higher order statistical information about the input random process. Therefore, the correntropy function partially characterizes higher order statistics of random processes with a compact bivariate kernel function. Note that the emphasis given to the higher order moments versus the second moment is controlled by the kernel size, which is a free parameter introduced by the method and must be selected by the user from the data.

Fig. 1 shows the application of correntropy to a sinusoidal signal of period 1 s with 100-Hz sampling frequency. The kernel size for this example is selected by the Silverman's rule (15) as 0.04. As can be observed, the information about periodicity given by correntropy is much peakier and decays faster than the one given by correlation and even the narrowed correlation. This result can be understood from first principles due to the effect of the nonlinear (Gaussian) kernel. Indeed, since the sample difference appears in the argument of the exponent, larger differences are exponentially attenuated. This means that only when the two versions of the sine wave align perfectly their correntropy is large. Minor time delays between t and s will produce an exponential decay in the similarity producing the very peaky appearance of correntropy across the lags. Of course, noise in the signal will also create values further away from zero even when the two signals are aligned, so it tends to decrease the peak as shown in Fig. 2 for three different signal-to-noise ratio (SNR) levels of white noise of 20, 10, and 1 dB, respectively. Note that the high noise case makes correntropy approach correlation. In Fig. 3, we present the Fourier transform of each function. The ordinary autocorrelation function only exhibits one harmonic and the narrowed autocorrelation produces ten harmonics which is equal to the number of terms L used in (3). The correntropy function places even more energy at higher harmonics in frequency due to the embedded nonlinearity which is controlled by the kernel size. The narrowness of correntropy function in the time domain anticipates the rich harmonics present in the frequency domain. For this reason, the lag domain seems the most interesting domain to apply correntropy.

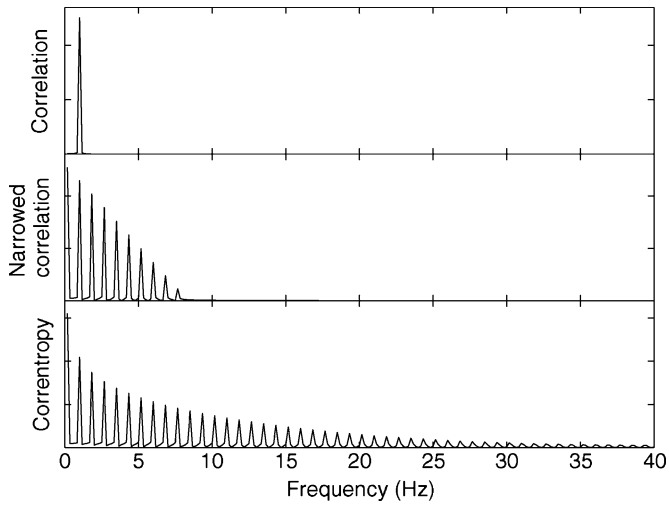


Fig. 3. Fourier transform of autocorrelation, narrowed autocorrelation with $L = 10$, and correntropy functions of a sinusoid signal.

It should also be noticed that there is a connection between the correntropy function (5) and the square difference function (2). The correntropy function also uses inhibitory neural interaction model instead of excitatory one with a Gaussian kernel function (7), but it nonlinearly transforms the subtraction of the signals by the exponential function. From another perspective, the correntropy function includes the scaled square different function as an individual term for $k = 1$ in the summation of (8). However, it contains more information with other higher order moment terms.

In the context of pitch determination, the correntropy function might as well estimate the pitch information of the signal similar to the autocorrelation function. However, compared to the autocorrelation function model, our pitch determination algorithm based on the correntropy function offers much better resolution and enhances the capacity of estimating multiple pitches. Since correntropy creates many different harmonics of each resonance present in the original time series due to the nonlinearity of the kernel function, it may also be useful for perceptual pitch determination. Finally, the structure of the correntropy definition may translate in a biologically plausible way some of the known sensitivity features of neurons as a function of synchrony in their

excitation. Its argument is sensitive to differences in time instances as correlation, but instead of being linear across differences, it gives more emphasis to values that are closer together in time. Neurons are also known to be very sensitive to time differences, and their highly nonlinear response also favors synchrony.

For completeness, we present below some of the most important properties of the correntropy function.

Property 1: The correntropy function is positive definite.

Given a positive definite kernel function $\kappa(\cdot,\cdot)$, for any positive integer n, any points $t_1,\ldots,t_n$ in $\mathbb{T}$, and any not all zero real numbers $\alpha_1,\ldots,\alpha_n$, by definition we have

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j\kappa\left(x_{t_i},x_{t_j}\right)>0.$$

Certainly, the expectation of any positive definite function is always positive definite. Thus, we have

$$E\left[\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j\kappa\left(x_{t_i},x_{t_j}\right)\right]>0.$$

This equals to

$$\sum_{i,j=1}^{n}\alpha_i\alpha_j E\left[\kappa\left(x_{t_i},x_{t_j}\right)\right]=\sum_{i,j=1}^{n}\alpha_i\alpha_j V(t_i,t_j)>0. \quad (9)$$

Property 2: $V(t,s)$ is symmetric: $V(t,s)=V(s,t)$.

This is the direct consequence of the symmetric kernel function used in the definition of the correntropy function.

Property 3: $V(t,t)>0$.

Since $\kappa(x_t,x_t)>0$ by the positive definiteness of kernel function, accordingly, $V(t,t)>0$.

Property 4: $|V(t,s)|<\sqrt{V(t,t)V(s,s)}$.

Let $n=2$ in (9), the expression reduces to

$$\alpha_1^2V(t,t)+\alpha_2^2V(s,s)>2\alpha_1\alpha_2\left|V(t,s)\right|. \quad (10)$$

We can substitute

$$\alpha_1^2=\frac{V(s,s)}{2\sqrt{V(t,t)V(s,s)}} \quad \text{and} \quad \alpha_2^2=\frac{V(t,t)}{2\sqrt{V(t,t)V(s,s)}}$$

into (10) to obtain the property above.

These properties are very similar to those of conventional correlation function. From the geometrical perspective, correntropy can be viewed as an inner product in the RKHS induced by the correntropy function; therefore, it preserves the geometrical interpretation of "correlation."

In order to obtain a univariate correntropy function, we see from (8) that all the even-order moments should be time shift invariant. This is a much stronger condition than the wide sense stationarity required by the conventional correlation function. More specifically, the sufficient condition to have

$$V(t+\tau,t)=V(\tau)=\frac{1}{N}\sum_{n=0}^{N-1}\kappa(x_n,x_{n+\tau}) \quad (11)$$

for all τ in the index set $\mathbb{T}$, is that the random process is strictly stationary on the even moments when the Gaussian kernel is used in the correntropy function.

Similar to the conventional (power) spectral density function defined for a wide-sense stationary random processes, we can also define a correntropy spectral density function for a strict stationary random processes.

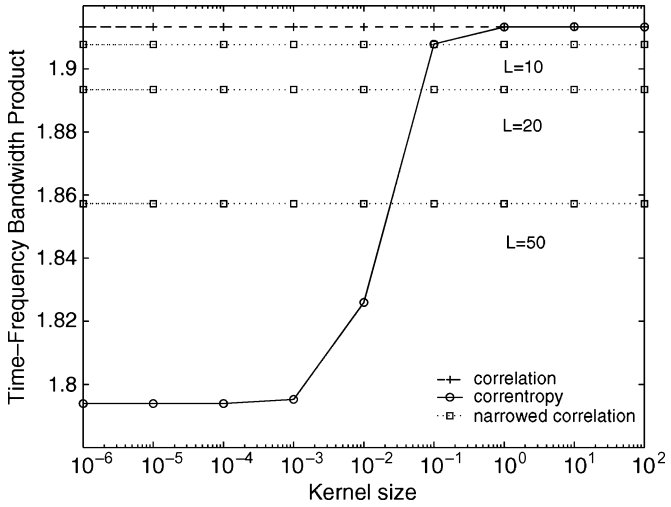


Fig. 4. Time–frequency bandwidth of autocorrelation, narrowed autocorrelation with $L=10,20,50$ and correntropy functions of a sinusoid signal.

Definition: Given a strict stationary random process $\{x_t,t\in\mathbb{T}\}$ with univariate centered correntropy function $U(\tau)$, the *correntropy spectral density function* is defined by

$$P(\omega)=\int_{-\infty}^{\infty}U(\tau)e^{-j\omega\tau}d\tau.$$

It is nothing but the Fourier transform of the univariate centered correntropy function. Therefore, we can also define the time–frequency bandwidth for the correntropy function.

Definition: Let $E=\int|U(\tau)|^2d\tau=(1/2\pi)\int|P(\omega)|^2d\omega$, and define the time bandwidth d and frequency bandwidth D as

$$d^2=\frac{1}{E}\int\tau^2\left|U(\tau)\right|^2d\tau, \quad \text{and}$$
$$D^2=\frac{1}{2\pi E}\int\omega^2\left|P(\omega)\right|^2d\omega$$

where d^2 and D^2 are the normalized variances of the magnitude-squared of the univariate centered correntropy function in, respectively, the time and frequency domain. Then the time–frequency bandwidth product for the correntropy function is defined as $B_{t\omega}=d\cdot D$.

For a finite duration signal, we estimate the time–frequency bandwidth product for the correntropy function by estimating the integrals in the definition. Because pitch determination is about pitch period localization in time or fundamental frequency localization in frequency, the smaller the time–frequency bandwidth product for the employed function is, the more accurate the pitch determination. In Fig. 4, we plot the time–frequency bandwidth product for the correlation, narrowed correlation, and correntropy functions for the same sinusoidal signal with respect to the kernel size. As expected, the narrowed correlation function generates smaller time–frequency bandwidth product than the correlation function. The figure clearly demonstrates that correntropy function achieves much smaller time–frequency bandwidth product than the correlation and narrowed correlation functions by choosing a suitable kernel size. The kernel sizes chosen in the experiments

are all fallen into the such range that the correntropy time–frequency bandwidth product is smaller than that of correlation and narrowed correlation functions. As the kernel size becomes larger, the time–frequency bandwidth product for the correntropy function approaches to that of the correlation function. This fact can be inferred from (11) because for large kernel sizes correntropy approaches correlation. However, we can no longer associate the concept of "power" to the correntropy spectral density. We will discuss the effect of kernel size on the correntropy function in Section V.

Since the kernel function appears in the definition of the correntropy, an analysis from the kernel point of view provides a different perspective. Any symmetric positive definite kernel function can induce a reproducing kernel Hilbert space whose bases can be expanded from the eigen-decomposition of the kernel function. Mercer's theorem [26] provides the theoretical foundation of eigen-decomposition of the kernel function, which states

$$\kappa(x_t, x_s) = \sum_{k=0}^{M} \lambda_k \varphi_k(x_t)\varphi_k(x_s) = \langle \Phi(x_t), \Phi(x_s) \rangle$$
$$\Phi : x_t \mapsto \sqrt{\lambda_k}\varphi_k(x_t), \quad k = 1, 2, \ldots M$$

where λ_k are the eigenvalues, φ_k are the corresponding eigenfunctions, M is the dimensionality of the reproducing kernel Hilbert space induced by the kernel function, and $\langle , \rangle$ denotes the inner product between two vectors. The nonlinear mapping Φ transforms the original 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

$$V(t, s) = E\left[\langle \Phi(x_t), \Phi(x_s) \rangle\right].$$

The correntropy function for any random process can be viewed as a "standard" correlation function for transformed random process. Kernel-based learning algorithms employ the nonlinear mapping Φ to treat *nonlinear* algorithms in a *linear* way if the problems can be expressed in terms of inner product [27]. This suggests that we can deal with nonlinear systems efficiently and elegantly in a linear fashion when applying the correntropy function. In fact, all the previous properties of the correntropy function can be derived in a kernel framework. For example, property 2 can be shown that $V(t, t) = E[\|\Phi(x_t)\|^2]$ which means that $V(t, t)$ is nothing but the expectation of the norm square of transformed random process. Property 3 is the generalized *Cauchy–Schwarz inequality* in the reproducing kernel Hilbert space

$$E\left[\langle \Phi(x_t), \Phi(x_s) \rangle\right] < \sqrt{E\left[\|\Phi(x_t)\|^2\right] E\left[\|\Phi(x_s)\|^2\right]}.$$

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

$$\begin{aligned} V(t, s) &= E\left[\langle \Phi(x_t), \Phi(x_s) \rangle\right] = \langle E\left[\Phi(x_t)\right], E\left[\Phi(x_s)\right] \rangle \\ &= \left\langle \frac{1}{N}\sum_{i=1}^{N} \Phi(x_i), \frac{1}{N}\sum_{j=1}^{N} \Phi(x_j) \right\rangle \\ &= \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} \kappa(x_i, x_j). \end{aligned} \tag{12}$$

We have used the sample mean to estimate the statistical expectation and the independence property in the computation above, where $\{x_1, x_2, \ldots, x_N\}$ is one realization of the random process. The quantity (12) is called *information potential* and corresponds to the argument of the logarithm of the quadratic Renyi's entropy when a Parzen window estimator is used [28]. In fact, for a strict stationary random process, the univariate correntropy function (11) asymptotically approaches the information potential. This clearly demonstrates the relationship between the correntropy function and quadratic Renyi's entropy. Hence, we coined the name *correntropy* to show that $V(t, s)$ includes both time structure and higher order statistical description of the random processes [1].

III. 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 [29]. The impulse response of a gammatone filter is defined as

$$q(t) = t^{n-1} e^{-2\pi a t \cos(2\pi f_0 t + \psi)}$$

where n is the filter order with center frequency at f_0 Hz, ψ is phase, and a is the bandwidth parameter. 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 and 4000 Hz [30]. This creates a cochleagram, which is a function of time lag along the horizontal axis and cochlear place, or frequency, along the vertical axis. The cochlear separates a sound into broad frequency channels while still containing the time structure of the original sound. It has served as a peripheral preprocess in the CASA model [18], and used extensively in pitch determination [7], [31].

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

$$V_i(\tau) = \frac{1}{N}\sum_{n=0}^{N-1} \kappa\left(x_n^i, x_{n+\tau}^i\right) \tag{13}$$

where i stands for channel number and z_n is the cochlear output. The kernel bandwidth is determined using Silverman's rule [32]. The time lag τ is chosen long enough to include the lowest expected pitch. Generally it is set at least 10 ms throughout the paper. In this way, a picture is formed by grey coding the correntropy values (white high, black low) with horizontal axis as correntropy lags and vertical axis as cochlear frequency. We name it *correntropy-gram*, which literally means "pictures of correntropy." If a signal is periodic, strong vertical

lines at certain correntropy lags appear in the correntropy-gram indicating times when a large number of cochlear channels display synchronous oscillations in correntropy. The horizontal bands signify different amounts of energy across frequency regions. The correntropy-gram is similar to the correlogram in structure but different in content since it does not display power, but "instantaneous" information potential. In order to reduce the dynamic range for display in the correntropy-gram, the correntropy function should be normalized such that the zero lag value is one as given by the following formula:

$$C_i(\tau)=\frac{\frac{1}{N}\sum_{n=0}^{N-1}\kappa\left(x_n^i,x_{n+\tau}^i\right)-\frac{1}{N^2}\sum_{n,m=0}^{N-1}\kappa\left(x_n^i,x_{m+\tau}^i\right)}{\sqrt{V(0)-\sum_{n,m=0}^{N-1}\frac{\kappa(x_n^i,x_m^i)}{N^2}}\sqrt{V(0)-\sum_{n,m=0}^{N-1}\frac{\kappa\left(x_{n+\tau}^i,x_{m+\tau}^i\right)}{N^2}}} \tag{14}$$

where $V(0)$ is the value of correntropy when lag $\tau = 0$. The numerator is called the *centered correntropy* which takes out the mean value of the transformed signal in the RKHS. $C(\tau)$ is also called *correntropy coefficient* that has been applied to detect nonlinear dependence among multichannel biomedical signals [33].

In order to emphasize pitch related structure in the correntropy-gram, the correntropy functions are summed up across all the channels to form a "pooled" or "summary" *correntropy-gram*

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

The summary correntropy-gram measures how likely the pitch would be perceived at a certain time lag. The pitch frequency can be obtained by inverting the time delay lag. In our experiment, the summary of correntropy functions is first normalized by subtracting the mean and dividing by the maximum absolute value. The position of pitch can be picked by various peak-picking algorithms to identify local maximum above the predefined threshold. Here we calculate the first derivative and mark the position when the value changes from positive to negative as a local maximum.

Compared to the conventional correlogram model [7], [31], [34], 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. Furthermore, since the correlogram estimates the likelihood that a pitch exists at a certain time delay, the summary correlogram may generate other "erroneous" peaks besides the one corresponding to the pitch period [17]. while the summary correntropy-gram suppresses values that are dissimilar at all other time delays by the exponential decay of the Gaussian function and only peaks at the one corresponding to the pitch period. For mixtures of concurrent sound sources with different fundamental frequencies, the summary correlogram usually fails to detect multiple pitches without further nonlinear postprocessing. However, the summary correntropy-gram is able to show peaks at different periods of each source. These characteristics of the proposed method suggest a superiority of the correntropy function over the autocorrelation function in pitch determination.

Moreover, the computational complexity of our method, whether the correntropy function (13) or the correntropy coefficient (14), remains similar to the correlogram. Although there are double summations in the correntropy coefficient, the computational complexity can be reduced to $O(N \log N)$ using the fast-Gauss transform [35]. However, the "narrowed" autocorrelation function increases computational complexity by including more delay terms.

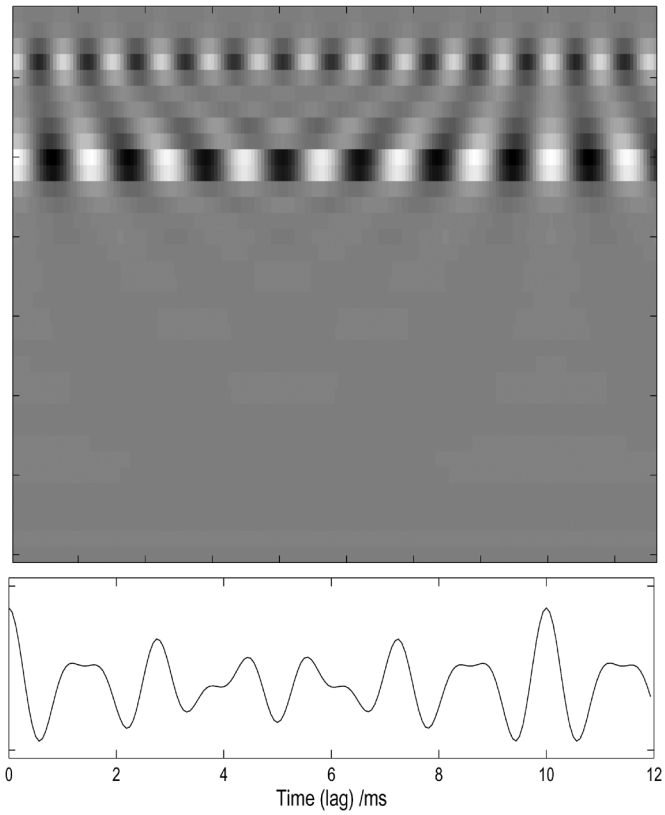


Fig. 5. Correlogram (top) and the summary (bottom) for vowel /a/ with 100-Hz fundamental frequency.

IV. Experiments

In this section, we present three experiments to validate our method. In the first two simulations, we compare our method with the conventional autocorrelation function [31], the third-order cumulants function [22], and the arrowed autocorrelation function [9] in determining pitches for a single speaker and two combined speakers uttering different vowels. The synthetic vowels are produced by Slaney's Auditory Toolbox [36]. For a fair comparison, we did not apply any postprocessing on the correlogram as was used in [31]. The conventional autocorrelation function (1), autocorrelation of third-order cumulants functions (4), narrowed autocorrelation functions (3), and correntropy functions (13) are presented after the same cochlear model. In the third experiment, the proposed method is tested using Bagshaw's database which is a benchmark for testing PDAs [37].

A. Single Pitch Determination

Figs. 5– 8 present the pitch determination results for a single synthetic vowel /a/ with 100-Hz fundamental frequency. The upper plots are the images of correlation functions, autocorrelations of third-order cumulants, narrowed autocorrelations, and correntropy functions after the same cochlear model, respectively. The bottom figures are the summaries of those four images. The kernel size σ in the Gaussian kernel (7) has been

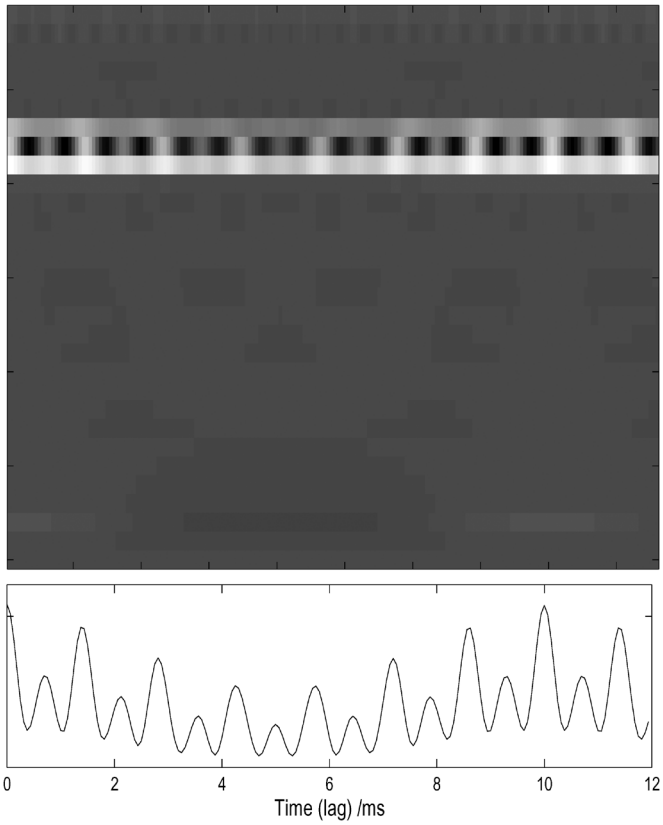


Fig. 6. Autocorrelation functions of third-order cumulants (top) and the summary (bottom) for the vowel /a/ with 100-Hz fundamental frequency.

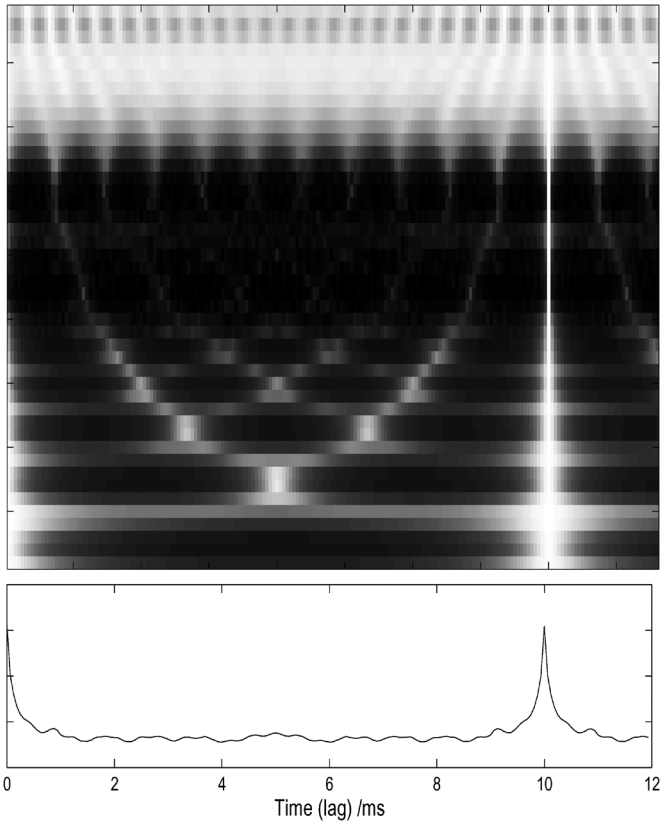


Fig. 8. Correntropy-gram (top) and the summary (bottom) for vowel /a/ with 100-Hz fundamental frequency.

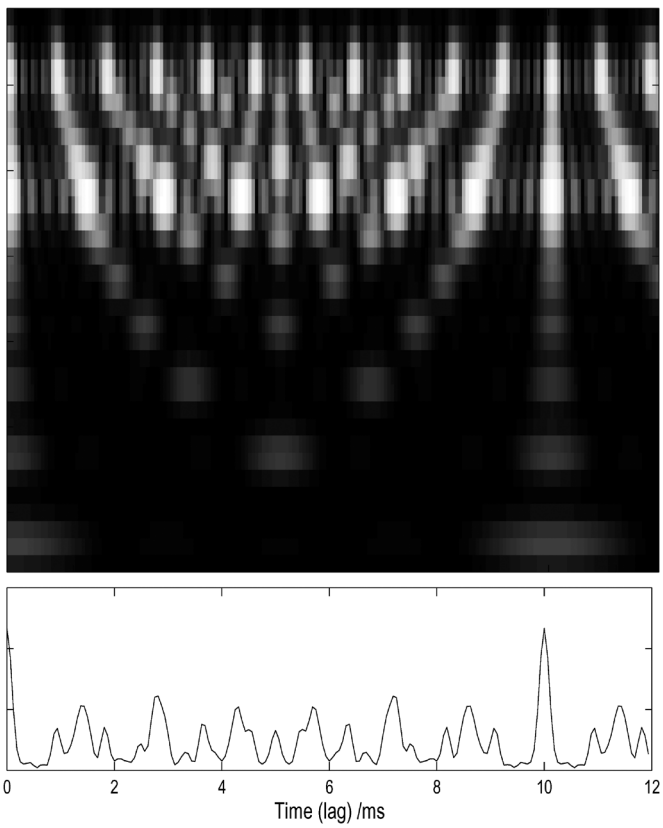


Fig. 7. Narrowed autocorrelation functions (top) and the summary for vowel /a/ with 100-Hz fundamental frequency.

chosen to be 0.01 (we will discuss further kernel size selection in Section V) and $L = 10$ in the narrowed autocorrelation function (3). The conventional autocorrelation, third-order cumulants and narrowed autocorrelation are all able to produce peaks at 10 ms corresponding to the pitch of the vowel. However, they also generate other erroneous peaks which might confuse pitch determination. On the contrary, the summary of correntropy-gram provides only one single and narrow peak at 10 ms which is the pitch period of the vowel sound, and the peak is much narrower than those obtained from other methods. The correntropy-gram clearly shows a single narrow stripe across all the frequency channels which concentrates most of the energy, including the low-frequency channels where all the other methods are weak.

The fine structure of hyperbolic contours can also be clearly seen in the correntropy-gram. Particularly, the second harmonic shows two peaks during the time interval when the fundamental frequency exhibits one. The contour results from the gamma shape cochleagram. The autocorrelation of the third-order cumulants fail to present such structures. All other three methods are able to produce hyperbolic contours. However, the correntropy-gram shows the finest structure due to its ability to generate all the harmonics of signal because of its nonlinear structure. The image of the narrowed autocorrelation functions is able to show some hyperbolic contours, the white vertical stripe at the fundamental frequency is much wider than that of the correntropy-gram and there are other large values in the high-frequency channels. These result in the wide peak at 10 ms and other erroneous peaks in the summary of the narrowed autocorrelation functions. Our proposed method clearly outperforms the conventional autocorrelation function,

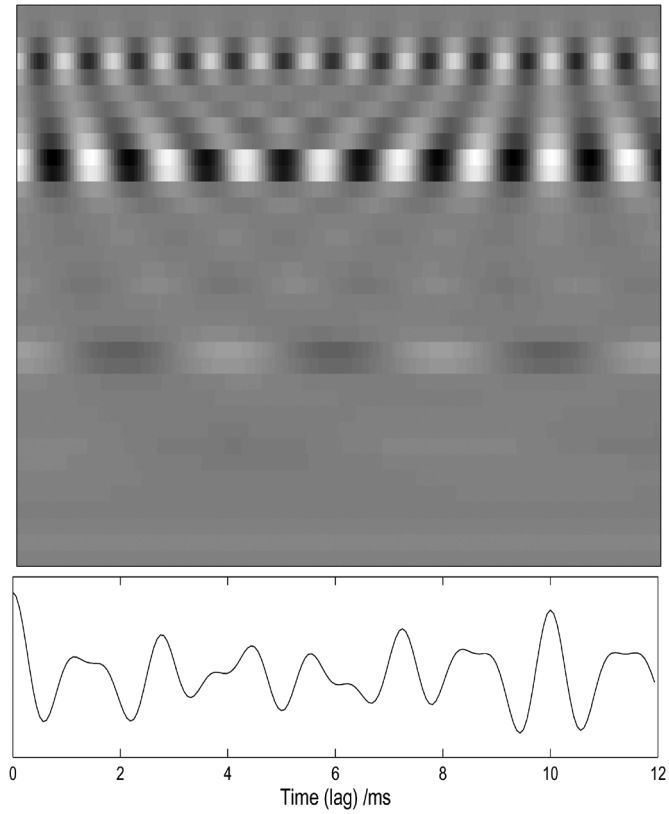


Fig. 9. Correlogram (top) and the summary (bottom) for a mixture of vowels /a/ and /u/ with fundamental frequencies at 100 and 126 Hz, respectively.

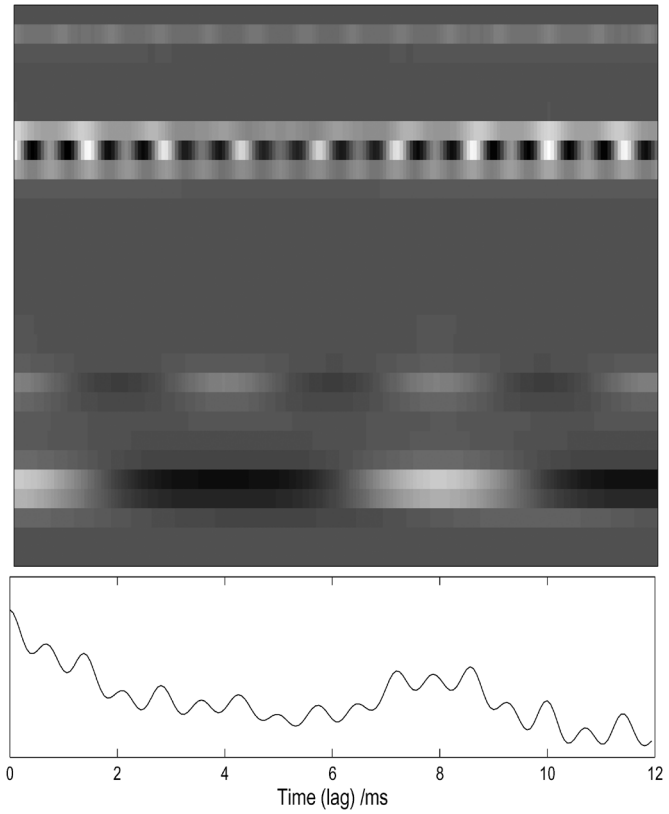


Fig. 10. Autocorrelations of third-order cumulants (top) and the summary (bottom) for a mixture of vowels /a/ and /u/ with fundamental frequencies at 100 and 126 Hz, respectively.

third-order cumulants method, and narrowed autocorrelation function in single pitch determination case.

B. Double Pitch Determination

In this example, we consider pitch determination for a mixture of two concurrent synthetic vowels with /a/ $(F0 = 100\,\text{Hz})$ and /u/ $(F0 = 126\ \text{Hz})$ which are separated by four semitones. The difference in power at the pitch between /u/ and /a/ complicates further the segregation. We compare the same four methods of the previous experiment to demonstrate that the correntropy function is able to determine two pitches presented in the mixture of two vowels.

Figs. 9– 12 present the simulation results. The correlogram method result of Fig. 9 only shows one peak corresponding to the pitch of the vowel /a/ while no indication of the other vowel /u/ at time of 7.9 ms is provided. The summary of correlogram resembles that of single vowel case in Fig. 5. The third-order cumulants method in Fig. 10 fails to detect two pitches in the mixture signal. Although there are two small peaks at 10 and 7.9 ms which correspond to the two pitch periods, respectively, their amplitudes are not large enough to be reliably detected. In Fig. 11, the summary of narrowed autocorrelation functions with $L = 15$ is able to produce only one peak at 10 ms corresponding to the pitch period of vowel /a/, but there is no peak at 7.9 ms. There are white streaks in the low-frequency channels in the narrowed autocorrelation functions which are the indications of the second vowel /u/, where the one at 7.9 ms corresponds to the pitch period of vowel /u/. However, the amplitude is too small compared with that of vowel /a/ and the information is lost in the summary plot. A complex neural network oscillator has been used to separate the channels dominated by different voices, and the summaries of individual channels are able to produce peaks corresponding to different vowels [7].

On the other hand, our method is able to exhibit two reasonable peaks from the mixture of two vowels. The kernel size σ is set to 0.07 in this experiment. The correntropy-gram in Fig. 12 shows a white narrow stripe across high-frequency channels at 10 ms corresponding to the pitch period of the vowel /a/. These channels have center frequencies close to the three formant frequencies of vowel /a/ ($F1 = 730$ Hz, $F2 = 1090$ Hz, $F3 = 2440$ Hz). The hyperbolic structure can still be seen in the high-frequency channels, but the lower frequency channels have been altered by the presence of the vowel /u/. The three high-energy white bars appear along the frequency channels centered at 300 Hz which is the first formant of vowel /u/. The second white streak is located at 7.9 ms and matches the pitch period of vowel /u/. The positions of white streaks match to those of in Fig. 11 for the narrowed correlation function. In the correntropy-gram summary, the first peak at 10 ms corresponds to the pitch period of vowel /a/. It is as narrow as the one in the single-vowel case in Fig. 8. The second peak appears at 8.2 ms which is only 4 Hz off the true pitch frequency (126 Hz). It is much less than the 20% gross error pitch determination evaluation criterion [38] or 10-Hz gross error [5]. The second peak is also much wider than the one at 10 ms. The amplitude for the peak at 8.2 ms is also smaller than that of peak at 10 ms since the energy ratio is 5.2 times higher for vowel /a/. The pitch shift and peak broadening phenomenon is due to the fact that vowel /a/ dominates the mixture signal and it generates spurious peaks which blur that of vowel /u/. However, it is remarkable that our method, with the proper kernel size, is able to detect two pitches while

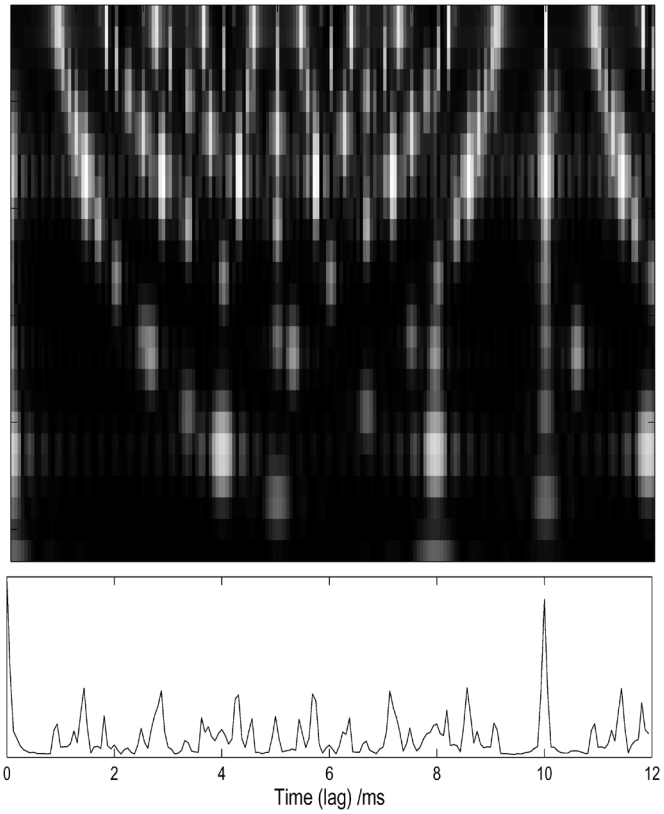


Fig. 11. Narrowed autocorrelations (top) and the summary (bottom) for a mixture of vowels /a/ and /u/ with fundamental frequencies at 100 and 126 Hz, respectively.

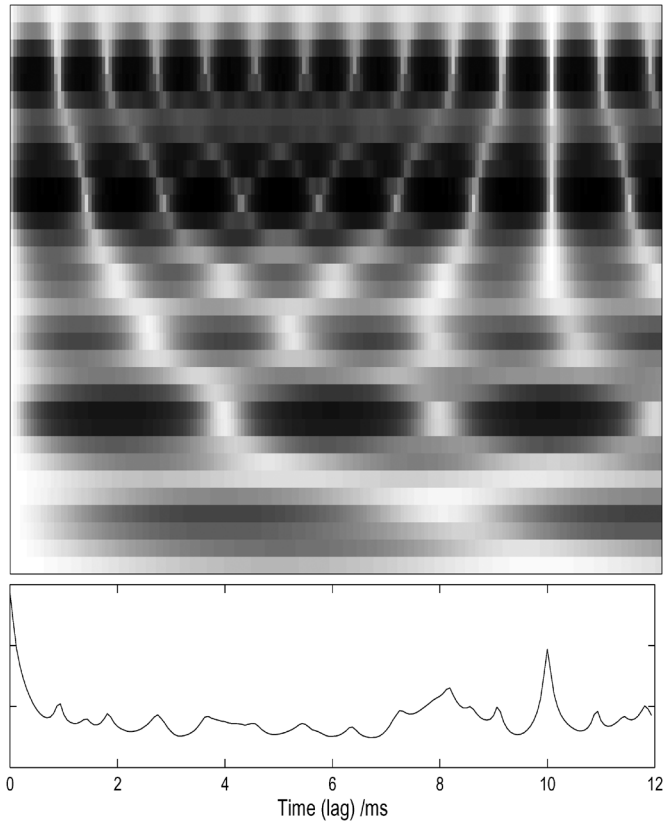


Fig. 12. Correntropy-gram (top) and the summary (bottom) for a mixture of vowels /a/ and /u/ with fundamental frequencies at 100 and 126 Hz, respectively.

all other algorithms fail in this experiment. The visual assessment clearly demonstrates the features of the proposed correntropy-based methodology and is an indicator for superior performance over the conventional correlogram, third-order cumulants, and narrowed correlation approaches for multipitch determination.

C. Double Vowel Segregation

To further investigate the performance of the proposed PDA, we generate a set of three vowels: /a/, /u/, and /i/ using Slaney's Auditory Toolbox. Each vowel is synthesized at five pitches corresponding to differences of 0.25, 0.5, 1, 2, and 4 semitones from 100 Hz, and the duration is 1 s each. For every mixture of double vowels, one is always with the fundamental frequency at 100 Hz, and the other constituent can be any vowel at any pitch value. In total, we have 45 mixtures of different combinations of vowels with different pitch values $(3 \text{ vowels} \times 3 \text{ vowels} \times 5 \text{ pitches})$. The detection functions from each of the four methods discussed above have been normalized to 0 and 1. A threshold is varied between 0 and 1 to decide the peak locations across a range of lags from 7.5 to 11 ms that covers the range of the pitches. If the difference between the detected pitch and reference is within a certain tolerance, the right pitch is detected. Since the minimum distance in this experiment is 0.25 semitone from 100 Hz, which is 1.45 Hz, we select the tolerance to be 0.725 Hz. Fig. 13 plots the receiver operating characteristic (ROC) curves for the four pitch determination algorithms based on correntropy function, autocorrelation, narrowed autocorrelation, and autocorrelation of third-order cumulants. It clearly shows that our method outperforms the other three in double-vowel pitch detection. However, none is able to get 100% detection. Notice that the ROC curve for correntropy function has an abrupt increase at zero probability of false alarm, corresponding to 45% of correct detection. This is due to the fact that correntropy function is able to suppress other erroneous peaks which are away from pitch positions and concentrate energy around the fundamental frequencies. The performance of autocorrelation and third-order cumulants are below 50% detection rate, irrespective of the number of

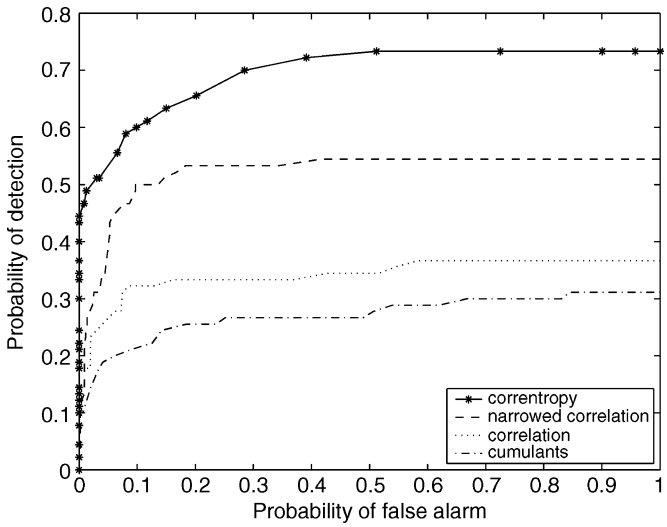


Fig. 13. ROC curves for the four PDAs based on correntropy, narrowed autocorrelation $(L = 15)$, autocorrelation, and autocorrelation of third-order cumulants in double vowels segregation experiment.

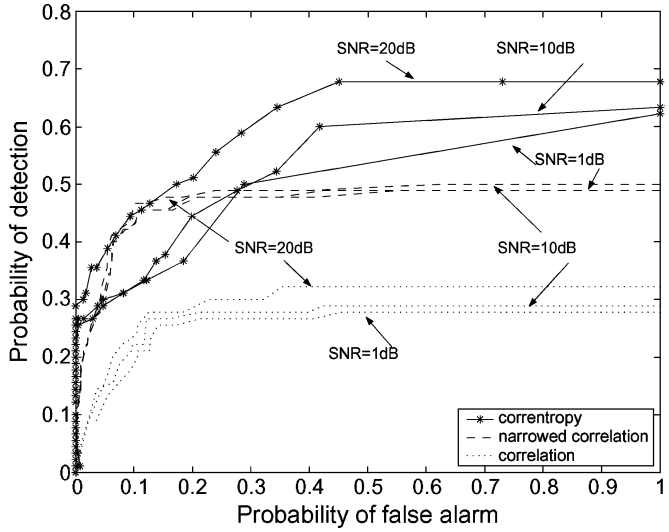


Fig. 14. ROC curves for the three PDAs based on correntropy, narrowed autocorrelation ($L = 15$), and autocorrelation in double vowels with additive white noise segregation experiment.

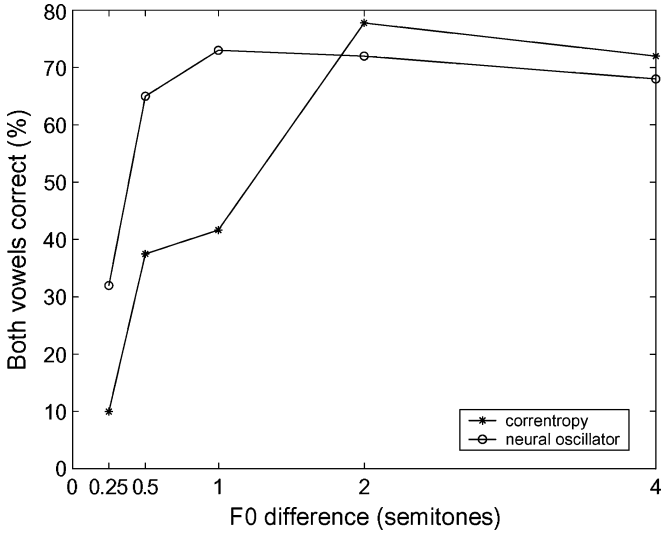


Fig. 15. Percentage performance of correctly determining pitches for both vowels for proposed PDA based on correntropy function and a network of neural oscillators model.

false alarms generated, which means that most often the second largest peak is an harmonic of the highest pitch. This is not surprising since both functions fail to present two peaks for most mixtures in this experiment.

Fig. 14 plots the ROC curves for the PDAs based on the correntropy and correlation functions for the double vowel segregation experiment where the vowels are corrupted by a white Gaussian noise of different SNR levels. The results show that the proposed PDA is fairly robust against white noise and perform better than the PDA based on the narrowed correlation and correlation function in three different SNR level of white noises.

We also present the vowel identification performance to examine the discriminating ability of correntropy function at different semitones for the mixture of double vowels. In the experiment, the threshold is chosen such that the first two peaks are detected. We compare our results with a CASA model with a network of neural oscillators [7] in Fig. 15. The CASA model outperforms our method at 0.25, 0.5, and 0.5 semitones of $F0$ differences since it uses a sophisticated network of neural oscillators to assign different channels from ERB filter bank outputs to different vowels. Our method is just based on the simple summarized correntropy-gram. The closer the fundamental frequencies of the two vowel become, the harder is for correntropy to produce two distinct peaks corresponding to different pitches. However, our method obtains comparable results to CASA model at 2 and 4 semitones of $F0$ differences. It suggests that our simple model is able to produce similar results for double vowel segregation of 2 and 4 semitones of $F0$ differences compared to the sophisticated CASA model. This certainly shows our technique is very attractive in compromising between simplicity and performance.

D. Benchmark Database Test

We test our pitch determination algorithm in Bagshaw's database [37]. It contains 7298 male and 16 948 female speech samples. The groundtruth pitch is estimated at reference points based on laryngograph data. These estimates are assumed to be equal to the perceived pitch. The signal is segmented into 38.4-ms duration centered at the reference points in order to make the comparisons between different PDAs fair. The sampling frequency is 20 kHz. The kernel size is selected according to Silverman's rule (15) for different segments. We use (14) to calculate the normalized correntropy function to yield unit at zero lag. Since the pitch range for male speaker is 50–250 Hz and 120–400 Hz for female speaker, the PDA searches local maxima from 2.5 to 20 ms in the summary correntropy function. We set the threshold to be 0.3 by trial and error so that every local maximum which exceeds 0.3 will be detected as a pitch candidate.

TABLE I
GROSS ERROR PERCENTAGE OF PDAS EVALUATION

	Male		Female		
PDA	High (%)	Low (%)	High (%)	Low (%)	Weighted Mean (%)
HPS	5.34	28.2	0.46	1.61	11.54
SRPD	0.62	2.01	0.39	5.56	4.95
CPD	4.09	0.64	0.61	3.97	4.63
FBPT	1.27	0.64	0.60	3.35	3.48
IPTA	1.40	0.83	0.53	3.12	3.22
PP	0.22	1.74	0.26	3.20	3.01
SHR	1.29	0.78	0.75	1.69	2.33
YIN					2.2
SHAPE	0.95	0.48	1.14	0.47	1.55
eSRPD	0.90	0.56	0.43	0.23	0.92
Correntropy	0.71	0.42	0.35	0.18	0.71

Table I summarizes the performance of various PDAs which are taken from [38]–[40]. The performance criterion is the relative number of gross errors. A gross error occurs when the estimated fundamental frequency is more than 20% off the true pitch value. This relatively loose criterion has been used in many studies, particularly in this benchmark dataset. The percent gross errors by gender and by lower or higher pitch estimates with respect to the reference are given in Table I. The weighted gross error is calculated by taking into account the number of pitch samples for each gender. It clearly shows that

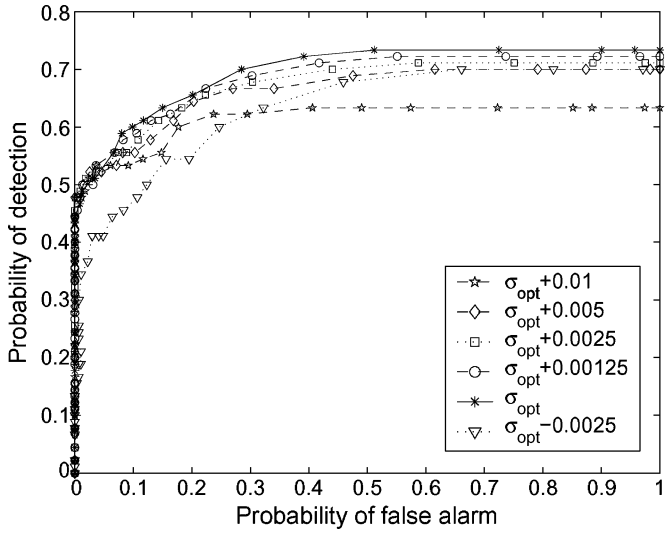


Fig. 16. ROC curves for PDA based on correntropy function with different kernel sizes in double vowels segregation experiment.

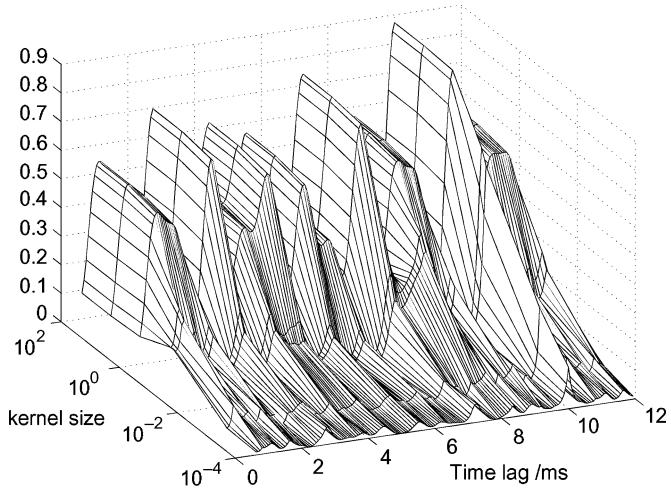


Fig. 18. Summary of correntropy functions with different kernel sizes for a mixture of vowels /a/ and /u/ with fundamental frequencies at 100 and 126 Hz, respectively.

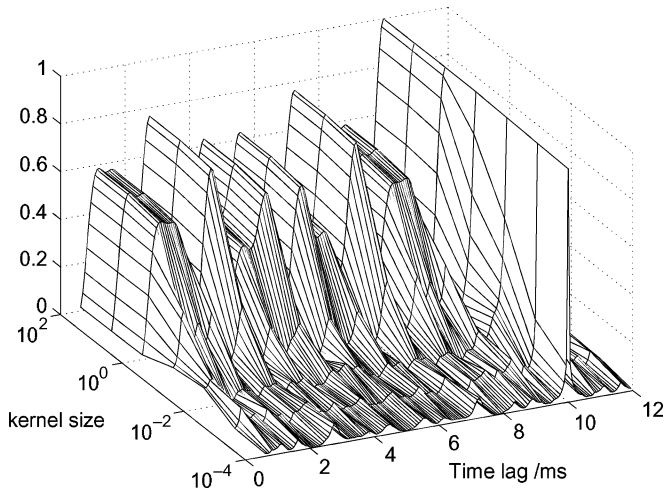


Fig. 17. Summary of correntropy functions with different kernel sizes for a single vowel /a/ with fundamental frequency at 100 Hz.

for this particular database correntropy-based PDA outperforms others.

V. Discussions

Correntropy introduces a free parameter that needs to be set by the user from the data. Actually, the kernel size plays an important role in the performance of our method since it determines the scale at which the similarity is going to be measured. It has been shown that kernel size controls the metric of the transformed signal in the RKHS [1]. If the kernel size is set too large, the correntropy function approaches the conventional 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 discrimination ability. If the problem was supervised, cross validation [27] could be easily applied to help set the kernel size as done in classification. However, in pitch detection, we do not have the luxury to use such methodology. One practical way to select the kernel size is given by the Silverman's rule of density estimation [32]

$$\sigma = 0.9AN^{-1/5} \tag{15}$$

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. Since in this case the signals are scalar, this technique is simple and provides a reasonable initial value. However, it does not always provide the best performance and therefore the kernel size remains an open problem. Perhaps the kernel size should be considered as a scale parameter as in the wavelet transform and *a priori* knowledge about the signals of interest or multi scale analysis should be used to find the proper scale.

To illustrate the effect of different kernel sizes, we simulate the summary of correntropy functions for the same experiments setup in Section IV with different kernel sizes in Figs. 17 and 18. The Silverman's rule for this data is $\sigma = 10^{-2}$ for the single vowel /a/ case and $\sigma = 0.07$ for a mixture of /a/ and /u/m, respectively. It can be seen that if the kernel size is large, σ chosen from 10^0 to 10^2 here, the summaries of correntropy functions approach those of correlation functions shown in Figs. 5 and 9. As the kernel size approaches the one given by Silverman's rule, 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 /u/. If the kernel size is too small, σ set from 10^{-4} to 10^{-3} for the mixture of two vowels, the summary of correntropy functions loses its ability to represent the two pitches. For different speech mixtures the kernel size very likely will be different. Finally, Fig. 16 shows the ROC curves for the double vowel segregation with different kernel sizes used in the correntropy function, where σ_{opt} denotes the kernel size selected through the Silverman's rule for each mixture and others are different deviations away from the σ_{opt}. As can be observed, there is a large range of values of kernel size around the Silverman's rule of thumb that the proposed PDA

performs similarly, which shows that the sensitivity to this parameter is not high.

VI. Conclusion

A novel pitch determination algorithm is proposed based on the correntropy function. Its foremost characteristic is simplicity and the immediate integration in conventional PDA and CASA methods. 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, third-order cumulants method, and narrowed correlation function in single and double pitches determination. This suggests that correntropy can discriminate better pitch when two different speakers speak in the same microphone. This is essential in the computational auditory scene analysis. Moreover, a benchmark database test for various PDAs shows that the correntropy PDA outperforms a collection of alternative algorithms tested in the same dataset without further processing besides normalization. Although these results are preliminary with synthetic sounds and much further work is needed to evaluate the methods, this technique seems promising for CASA, and warrants further analysis with real speech in different SNR environments. The automatic selection of the kernel size or of a multiple kernel size analysis needs to be further investigated to automate the pitch determination algorithm. This remains the main theoretical challenge of the approach. The future work also includes incorporating correntropy-gram channel selection to enhance the discriminating ability of proposed method in multiple pitches determination.

References

[1] I. Santamaria, P. Pokharel, and J. C. Principe, "Generalized correlation function: Definition, properties, and application to blind equalization," *IEEE Trans. Signal Process.*, vol. 54, no. 6, pp. 2187–2197, Jun. 2006.

[2] W. J. Hess, *Pitch Determination of Speech Signals*. New York: Springer, 1993.

[3] M. J. Ross, H. L. Shaffer, A. Cohen, R. Freudberg, and H. J. Manley, "Average magnitude difference function pitch extractor," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-22, no. 5, pp. 353–362, Oct. 1974.

[4] A. de Cheveigné, "Cancellation model of pitch perception," *J. Acoust. Soc. Amer.*, vol. 103, no. 3, pp. 1261–1271, Mar. 1998.

[5] T. Shimarura and H. Kobayashi, "Weighted autocorrelation for pitch extraction of noisy speech," *IEEE Trans. Speech Audio Process.*, vol. 9, no. 7, pp. 727–730, Oct. 2001.

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

[7] G. J. Brown and D. Wang, "Modelling the perceptual segregation of double vowels with a network of neural oscillators," *Neural Netw.*, vol. 10, no. 9, pp. 1547–1558, 1997.

[8] L. Cohen, *Time–Frequency Analysis*. Englewood Cliffs, NJ: Prentice-Hall, 1995.

[9] J. C. Brown and M. S. Puckette, "Calculation of A "Narrowed" autocorrelation function," *J. Acoust. Soc. Amer.*, vol. 85, pp. 1595–1601, 1989.

[10] H. Duifhuis, L. Willems, and R. Sluyter, "Measurement of pitch in speech: An implementation of Goldstein's theory of pitch perception," *J. Acoust. Soc. Amer.*, vol. 71, pp. 1568–1580, 1982.

[11] M. R. Schroeder, "Period histogram and product spectrum: New methods for fundamental frequency measurement," *J. Acoust. Soc. Amer.*, vol. 43, pp. 829–834, 1968.

[12] D. J. Hermes, "Measurement of pitch by subharmonic summation," *J. Acoust. Soc. Amer.*, vol. 83, no. 1, pp. 257–264, 1988.

[13] X. Sun, "A pitch determination alogirithm based on subharmonic-to-harmonic ratio," in *Proc. 6th Int. Conf. Spoken Lang. Process.*, Beijing, China, 2000, vol. 4, pp. 676–679.

[14] A. de Cheveigné, "Pitch perception models," in *Pitch—Nerual Coding and Perception*, C. Plack, A. Oxenham, R. Fay, and A. Popper, Eds. New York: Springer-Verlag, 2005.

[15] J. C. R. Licklider, "A duplex theory of pitch perception," *Experientia*, vol. 7, pp. 128–134, 1951.

[16] R. Lyon, "Computational models of neural auditory processing," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, San Diego, CA, 1984, pp. 41–44.

[17] 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. New York: Wiley, 1993, pp. 95–116.

[18] D. Wang and G. J. Brown, *Computational Auditory Scene Analysis—Principles, Algorithms, and Applications*. New York: Wiley, 2006.

[19] M. Wu, D. Wang, and G. J. Brown, "A multipich tracking algorithm for noisy speech," *IEEE Trans. Speech Audio Process.*, vol. 11, no. 3, pp. 229–241, May 2003.

[20] R. Meddis and M. Hewitt, "Modeling the identification of concurrent vowels with different fundamental frequencies," *J. Acoust. Soc. Amer.*, vol. 91, pp. 233–245, 1992.

[21] A. de Cheveigné, "Multiple f0 estimation," in *Computational Auditory Scene Analysis—Principles, Algorithms, and Applications*, D. Wang and G. J. Brown, Eds. New York: Wiley, 2006, pp. 45–79.

[22] A. Moreno and J. Fonollosa, "Pitch determination of noisy speech using higher order statistics," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, San Francisco, CA, Mar. 1992, pp. 133–136.

[23] K.-H. Jeong and J. C. Principe, "The correntropy MACE filter for image recognition," in *Proc. Int. Workshop Mach. Learn. Signal Process. (MLSP)*, Maynooth, U.K., 2006, pp. 9–14.

[24] J.-W. Xu, P. P. Pokharel, A. R. C. Paiva, and J. C. Principe, "Nonlinear component analysis based on correntropy," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Vancouver, BC, Canada, Jul. 2006, pp. 3517–3521.

[25] M. G. Genton, "Class of Kernels for machine learning: A statistics perspective," *J. Mach. Learn. Res.*, vol. 2, pp. 299–312, 2001.

[26] J. Mercer, "Functions of positive and negative type, and their connection with the theory of integral equations," *Philos. Trans. R. Soc. London*, vol. 209, pp. 415–446, 1909.

[27] B. Schölkopf and A. Smola, *Learning With Kernels*. Cambridge, MA: MIT Press, 2002.

[28] J. C. Principe, D. Xu, and J. W. Fisher, "Information theoretic learning," in *Unsupervised Adaptive Filtering*, S. Haykin, Ed. New York: Wiley, 2000, pp. 265–319.

[29] R. D. Patterson, J. Holdsworth, I. Nimmo-Smith, and P. Rice, "SVOS Final Report, Part B: Implementing A Gammatone Filterbank," Appl. Psychol. Unit Rep. 2341, 1988.

[30] B. R. Glasberg and B. C. Moore, "Derivation of auditory filter shapes from notched-noised data," *Hearing Res.*, vol. 47, pp. 103–138, 1990.

[31] M. Slaney and R. F. Lyon, "A perceptual pitch detector," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Albuquerque, NM, 1990, pp. 357–360.

[32] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*. New York: Chapman & Hall, 1986.

[33] J.-W. Xu, H. Bakardjian, A. Cichocki, and J. C. Principe, "A new nonlinear similarity measure for multichannel signals," *Neural Netw.*, vol. 21, pp. 222–231, 2008.

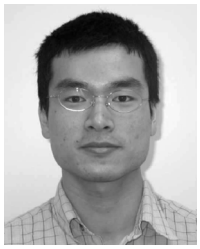
[34] R. Meddis and M. Hewitt, "Virtual pitch and phase sensitivity of a computer model of the auditory periphery: I. Pitch identification," *J. Acoust. Soc. Amer.*, vol. 89, pp. 2866–2882, 1991.

[35] S. Han, S. Rao, and J. C. Principe, "Estimating the information potential with the fast gauss transform," in *Proc. Int. Conf. Independent Compon. Anal. Blind Source Separation (ICA)*, Charleston, SC, 2006, pp. 82–89, ser. LNCS 3889.

[36] M. Slaney, "Malcolm Slaney's Auditory Toolbox to Implement Auditory Models and Generate Synthetic Vowels." [Online]. Available: http://www.slaney.org/malcolm/pubs.html

[37] P. Bagshaw, "Paul Bagshaw's Database for Evaluating Pitch Determination Algorithms." [Online]. Available: http://www.cstr.ed.ac.uk/research/projects/fda

[38] P. Bagashaw, S. Hiler, and M. Jack, "Enhanced pitch tracking and the processing of F0 contours for computer and intonation teaching," in *Proc. Eur. Conf. Speech Commun.*, 1993, pp. 1003–1006.
[39] A. Camacho and J. Harris, "A pitch estimation algorithm based on the smooth harmonic average peak-to-valley envelope," in *Proc. Int. Symp. Circuits Syst.*, New Orleans, LA, May 2007, pp. 3940–3943.
[40] A. de Cheveigné and H. Kawahara, "YIN, a fundamental frequency estimator for speech and music," *J. Acoust. Soc. Amer.*, vol. 111, no. 4, pp. 1917–1930, Apr. 2002.



Jian-Wu Xu (S'03) received the B.S. degree in electrical engineering from Zhejiang University, Hangzhou, China, in 2002 and the Ph.D. degree in electrical and computer engineering from the University of Florida, Gainesville, in 2007.

Since 2002, he has been with Computational NeuroEngineering Laboratory, the University of Florida, under the supervision of Dr. Principe. His current research interests include information theoretic learning, adaptive signal processing, control, and machine learning.

Dr. Xu is a member of Tau Beta Pi and Eta Kappa Nu.

Jose C. Principe (F'00) received the B.S. degree in electrical engineering from the University of Porto, Porto, Portugal, the M.S. and Ph.D. degrees in electrical engineering from the University of Florida, Gainesville, and the Laurea Honoris Causa degree from the Universita Mediterranea, Reggio Calabria, Italy.

He has been a Distinguished Professor of Electrical and Biomedical Engineering at the University of Florida since 2002. He joined the University of Florida in 1987, after an eight-year appointment as Professor at the University of Aveiro, in Portugal. His interests lie in nonlinear non-Gaussian optimal signal processing and modeling and in biomedical engineering. He created in 1991 the Computational NeuroEngineering Laboratory to synergistically focus the research in biological information processing models. He was supervisory committee chair of 50 Ph.D. and 61 M.S. students, and he is author of more than 400 refereed publications (three books, four edited books, 14 book chapters, 116 journal papers and 276 conference proceedings).

Dr. Principe received the Gabor Award from the International Neural Network Society in 2006 and the Career Achievement Award from the IEEE Biomedical Engineering Society in 2007. He is a Fellow of the AIMBE, past President of the International Neural Network Society, and Past Editor-in-Chief of the IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING, as well as a former member of the Advisory Science Board of the FDA. He holds five patents and has submitted seven more.