Alpha-Stable Modeling of Noise and Robust Time-Delay Estimation in the Presence of Impulsive Noise

Panayiotis G. Georgiou, *Student Member, IEEE,* Panagiotis Tsakalides, *Member, IEEE,* and Chris Kyriakakis, *Member, IEEE*

***Abstract*— A new representation of audio noise signals is proposed, based on symmetric α-stable (SαS) distributions in order to better model the outliers that exist in real signals. This representation addresses a shortcoming of the Gaussian model, namely, the fact that it is not well suited for describing signals with impulsive behavior. The α-stable and Gaussian methods are used to model measured noise signals. It is demonstrated that the α-stable distribution, which has heavier tails than the Gaussian distribution, gives a much better approximation to real-world audio signals.**

The significance of these results is shown by considering the time delay estimation (TDE) problem for source localization in teleimmersion applications. In order to achieve robust sound source localization, a novel time delay estimation approach is proposed. It is based on fractional lower order statistics (FLOS), which mitigate the effects of heavy-tailed noise. An improvement in TDE performance is demonstrated using FLOS that is up to a factor of four better than what can be achieved with second-order statistics.

***Index Terms*— Microphone arrays, symmetric alpha-stable distributions, time delay estimation, wideband array signal processing.**

I. Introduction

THE proliferation of integrated media technologies combined with the steady increase of available computing power and high-bandwidth networking infrastructure are beginning to lay the ground for teleimmersion applications. The ultimate goal of teleimmersion is to create the illusion of proximity among multiple participants that are geographically apart. In order to achieve this illusion, it is necessary to capture and recreate all the necessary aural and visual cues that human participants rely on.

In this paper we examine issues that relate to the robustness of methods for localization of sound sources in teleimmersion applications using microphone arrays. In particular, we are interested in identifying the location of a person speaking in the presence of background room noise. One possible application would be the capability to automatically steer a camera in the direction of the speaker during a session with multiple participants at each site. The presence of noise such as door slams, chair creaks, computer hard drives, and objects dropping (Fig. 1) causes such camera steering systems to turn at every instance.

Several distributions exist that can be good candidates for modeling audio noise signals such as the ones mentioned above. The most common in the literature, and especially in speech and audio signal processing, is the Gaussian distribution. The use of the Gaussian distribution is frequently motivated by the physics of the problem, and in most cases it ensures an analytical solution. This has led to the development of numerous algorithms based on second-order statistics. In recent years, further research into signal modeling has led to the realization that many natural phenomena can be better represented by distributions of a more impulsive nature. One type of distribution that exhibits heavier tails than the Gaussian is the class of α-stable distributions. In 1993, Nikias and Shao [1] gave an introductory review of α-stable distributions from a statistical signal processing viewpoint that was followed by a book from the same authors in 1995 [2]. Alpha-stable distributions have been used to model diverse phenomena such as random fluctuations of gravitational fields, economic market indexes [3], and radar clutter [4]. More recently, with the evolution of the World Wide Web, additional areas of application of the α-stable distributions have become apparent. For example, Crovella *et al*. have used the stable law to model data file sizes on the Web [3], while Willinger *et al*. used it to model network traffic [3].

The first half of this paper deals with the heavy tailed nature of certain types of audio noise signals in typical reverberant rooms. Sections II and III give an overview of the class of α-stable distributions and α-stable parameter estimation, while Section IV proceeds by fitting a Gaussian model to recorded audio data. The same data is subsequently used to obtain the best α-stable fit and compare the relative accuracy of the two models.

Next in this paper, we address the problem of sound source localization. Source location information can be used for tracking a moving sound source, steering a camera in teleconferencing applications to follow the speaker, selectively acquiring sound from a specific direction, improving hearing aid devices, as well as reducing noise and reverberation.

Manuscript received October 13, 1998; revised June 7, 1999. This research was supported in part by the Integrated Media Systems Center, a National Science Foundation Engineering Research Center with additional support from the Annenberg Center for Communication at the University of Southern California and the California Trade and Commerce Agency. The associate editor coordinating the review of this paper and approving it for publication was Dr. M. Iwadare.

P. G. Georgiou and C. Kyriakakis are with the Integrated Media Systems Center, University of Southern California, Los Angeles, CA 90089-2564 USA (e-mail: georgiou@sipi.usc.edu; ckyriak@imsc.usc.edu).

P. Tsakalides was with the Integrated Media Systems Center, University of Southern California, Los Angeles, CA 90089-2564 USA. He is now with the Department of Electrical and Computer Engineering, University of Patras, Patras, Greece (e-mail: tsakalid@ee.upatras.gr).

Publisher Item Identifier S 1520-9210(99)06729-2.

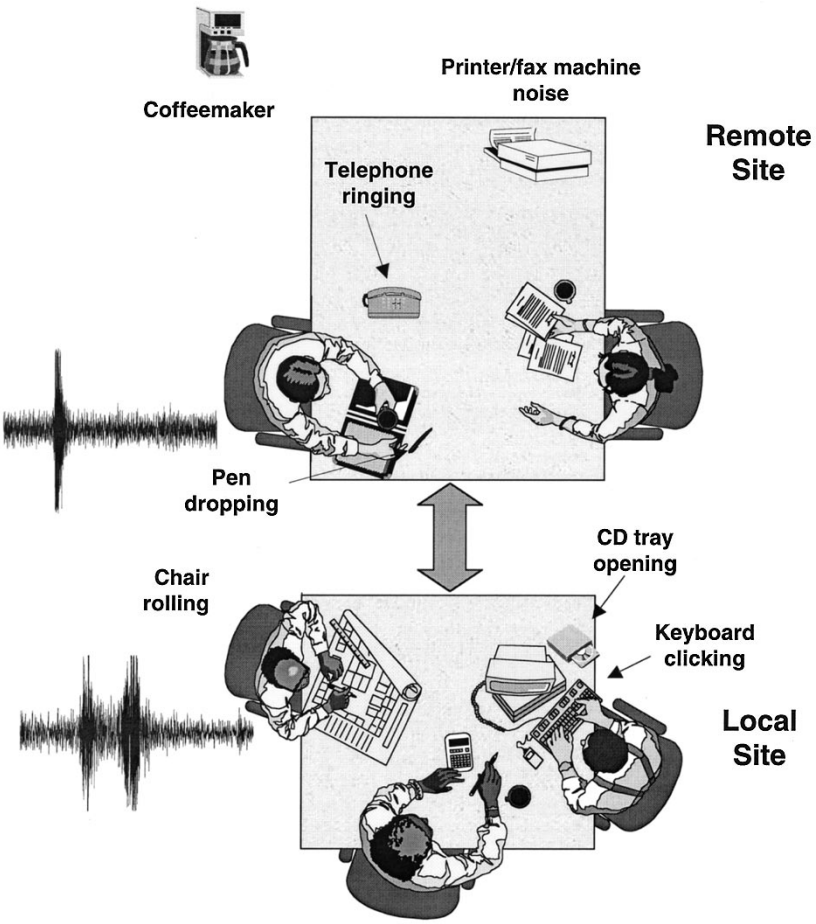

Fig. 1. Typical teleconference environment. Several instances of impulsive noise were recorded including a chair rolling, a pen dropping, and a keyboard clicking.

Intersensor *time delay estimation* (TDE) is a method commonly used to estimate source location using bearing information. TDE algorithms are well-suited to applications that involve a single wideband sound source. This is due to their low computational cost compared to high-resolution spectral estimation techniques and their reduced sensitivity to noise compared to the steered-beamformer locators.

Informative introductory tutorials on existing TDE methods are given in a paper by Knapp and Carter [5] and in a guest editorial by Carter [6]. A simple approach to estimate the bearing information for localizing a sound source in an enclosed space was proposed by Brandstein [7] in his dissertation. The majority of TDE methods proposed so far, especially for use in audio applications, assume a Gaussian noise signal and thus use second- (or higher) order statistics in order to locate the source. A drawback of second or higher order methods is that they become suboptimal when the signal deviates from the Gaussian assumption. Section V gives an overview of the TDE problem, its mathematical formulation, as well as the phase transform (PHAT) [5], [8] and the proposed fractional lower order statistics-PHAT (FLOS-PHAT) TDE methods. Finally, Section VI presents simulations on the TDE problem using both methods. We show that when the Gaussian noise assumption fails—and instead the α-stable distribution is a better approximation for the noise—then the FLOS-PHAT algorithm achieves better detection performance than the PHAT.

II. Alpha-Stable Distributions

The α-stable distribution, which can model phenomena of an impulsive nature, is a generalization of the Gaussian distribution and is appealing because of two main reasons.

- First, it satisfies the *stability property*, which states that if X, X_1, and X_2 are α-stable independent random variables of the same distribution, then there exist μ_1 and μ_2 satisfying

$$\nu_1 X_1 + \nu_2 X_2 \stackrel{d}{=} \mu_1 X + \mu_2 \tag{1}$$

  where ν_1, ν_2, μ_1, and μ_2 are constants and $\stackrel{d}{=}$ denotes equality in distribution.
- Second, it satisfies the *generalized central limit theorem* [2], [9], [10] stating: X is α-stable if and only if X is the limit in distribution of the sum

$$S_n = \frac{X_1 + X_2 + \cdots X_n}{a_n} - b_n \tag{2}$$

  where $X_1, X_2, \cdots$, are i.i.d. r.v.'s and $n \to \infty$. Parameter b_n is real and a_n is real and positive.

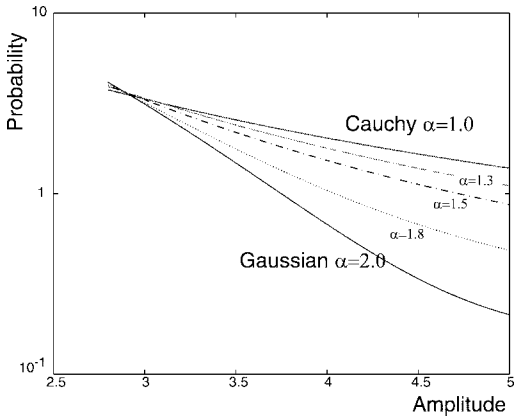

Fig. 2. The tails of the probability density function of a symmetric α-stable distribution for different values of α. The case of $\alpha = 2$ being the less impulsive case of Gaussian noise and $\alpha = 1$ the more impulsive Cauchy case. In all the above, the dispersion was kept constant at $\gamma = 1$.

There is no closed-form expression for the probability density function of α-stable distributions, but the characteristic function $\varphi(t)$ is given by

$$\varphi(t) = \exp(j\lambda t - \gamma|t|^{\alpha}[1 + j\beta\,\text{sign}(t)\omega(t, \alpha)]) \tag{3}$$

where

$$\omega(t, \alpha) = \begin{cases} \tan\dfrac{\alpha\pi}{2}, & \text{if } \alpha \neq 1 \\ \dfrac{2}{\pi}\log|t|, & \text{if } \alpha = 1 \end{cases} \tag{4}$$

$$\text{sign}(t) = \begin{cases} 1, & \text{if } t > 0 \\ 0, & \text{if } t = 0 \\ -1, & \text{if } t < 0 \end{cases} \tag{5}$$

and

- α is the *characteristic exponent* satisfying $0 < \alpha \leq 2$. The characteristic exponent controls the heaviness of the tails of the density function. The tails are heavier, and thus the noise more impulsive, for low values of α while for a larger α the distribution has a less impulsive behavior (Figs. 2 and 3).
- λ is the *location parameter* ($-\infty < \lambda < \infty$). It corresponds to the mean for $1 < \alpha \leq 2$ and the median for $0 < \alpha \leq 1$.
- γ is the *dispersion* parameter ($\gamma > 0$), which determines the spread of the density around its location parameter. The dispersion behaves in a similar way to the variance of the Gaussian density, and it is, in fact, equal to half the variance when $\alpha = 2$, the Gaussian case.
- β is the index of symmetry ($-1 \leq \beta \leq 1$). When $\beta = 0$, the distribution is symmetric around the location parameter.

The case of $\alpha = 2$, $\beta = 0$ corresponds to the Gaussian distribution, while $\alpha = 1$, $\beta = 0$ corresponds to the Cauchy distribution. The density functions in these two cases are given by

$$f_{\alpha=2}(\gamma, \lambda; x) = \frac{1}{\sqrt{4\pi\gamma}}\exp\left\{-\frac{(x-\lambda)^2}{4\gamma}\right\} \tag{6}$$

$$f_{\alpha=1}(\gamma, \lambda; x) = \frac{\gamma}{\pi[\gamma^2 + (x-\lambda)^2]}. \tag{7}$$

The impulsiveness of the α-stable distribution can clearly be seen in Fig. 3(a)–(c). However, when we take a closer look at Fig. 3(d)–(f), the time series resulting from the three different distributions does not seem very different. This encourages the use of α-stable distributions in situations where the noise has been traditionally modeled as Gaussian, but where sudden "spikes" might occur. For example, in an enclosed room, sounds produced by pages turning, pens clicking, or objects falling can give rise to the impulsiveness in the noise.

The class of α-stable distributions does not possess finite second (or higher) moments. In fact, α-stable distributions with $\alpha \neq 2$ have finite moments only for order p lower than α

$$\begin{aligned} & \alpha < 2, \quad \mathbb{E}|X_\alpha|^p \to \infty \quad && \forall\, p \geq \alpha \\ & \alpha < 2, \quad \mathbb{E}|X_\alpha|^p < \infty \quad && \forall\, 0 \leq p < \alpha \\ \text{Gaussian:} \quad & \alpha = 2, \quad \mathbb{E}|X_\alpha|^p < \infty \quad && \forall\, p \geq 0. \end{aligned} \tag{8}$$

References [1]–[3] and [10] treat the α-stable theory further. For the purposes of this paper, we will deal with the class of *symmetric α-stable* (SαS) distributions ($\beta = 0$) with finite mean, i.e., $1 < \alpha \leq 2$.

III. Parameter Estimation for SαS Distributions

The possibility that heavy-tailed noise behavior may adequately be described by the stable law gives rise to the need for fast, simple, and efficient estimators of the alpha-stable parameters (especially, the characteristic exponent, α) from real data. Several such estimators, compromising optimality for the sake of computational efficiency, have been proposed in the past. Among them, maximum likelihood methods developed by DuMouchel [11] and by Brorsen and Yang [12] are asymptotically efficient, but difficult to compute. Paulson *et al*. estimate the stable parameters by fitting the Fourier transform of the data to the characteristic function [13], a computationally intensive procedure. Hence, a number of suboptimal but simple methods have been devised. Zolotarev estimates the stable parameters by the method of moments, but requires that the location parameter is known in advance [14]. Brockwell and Brown estimate α with high efficiency in the special case of $\alpha < 1$. McCullogh [15] generalized the Famma and Roll approach [16] to provide consistent estimators of α in the range [0.6, 2].

For audio signals, we will be dealing with symmetric distributions and thus we can assume that the distribution will be of the SαS class. The two methods used in this paper to estimate the α and γ parameters are summarized below.

A. Positive-Order and Negative-Order (Sinc) Function Estimator

The *fractional lower order moment* (FLOM) of an SαS variable with zero location parameter can be shown [2] to be equal to

$$\mathbb{E}(|X|^p) = C_1(p, \alpha)\gamma^{p/\alpha} \tag{9}$$

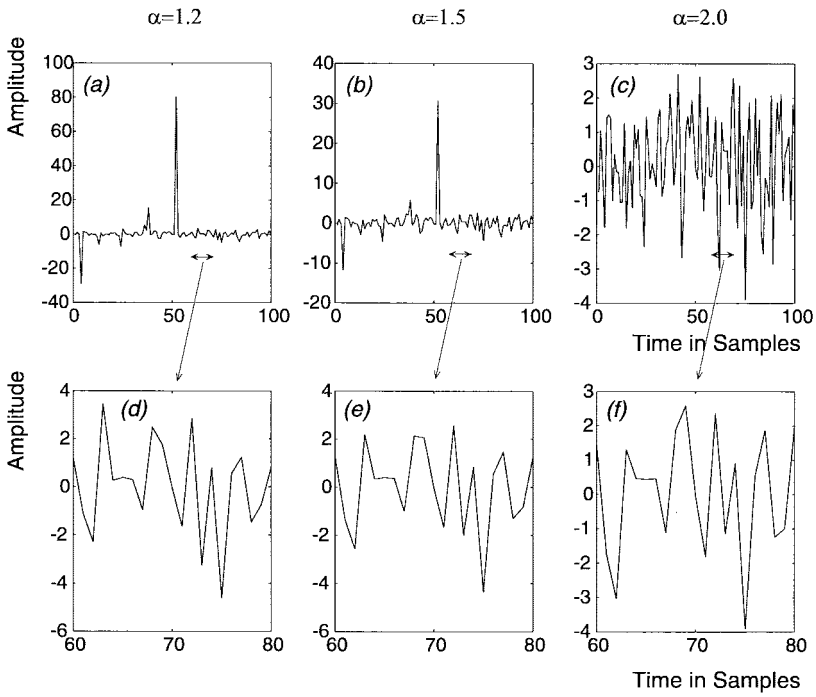

Fig. 3. Sample time series of SαS random variables. The characteristic exponents are $\alpha = 1.2$, $\alpha = 1.5$, and $\alpha = 2.0$ (Gaussian). The second row of figures shows an enlargement of parts of the above row and demonstrates the similarities between the distributions.

where $-1 < p < \alpha$ and

$$C_1(p, \alpha) = \frac{2^{p+1}\Gamma\left(\frac{p+1}{2}\right)\Gamma\left(-\frac{p}{\alpha}\right)}{\alpha\sqrt{\pi}\,\Gamma\left(-\frac{p}{2}\right)}. \tag{10}$$

The above expression for the pth-order moment of X gives rise to

$$\mathbb{E}(|X|^p)\mathbb{E}(|X|^{-p}) = \frac{2\tan(p\pi/2)}{\alpha\sin(p\pi/\alpha)} \tag{11}$$

where $0 < p < \min(\alpha, 1)$ and therefore

$$\operatorname{sinc}\left(\frac{p\pi}{\alpha}\right) = \frac{2\tan(p\pi/2)}{p\pi\mathbb{E}(|X|^p)\mathbb{E}(|X|^{-p})}. \tag{12}$$

From (12), the value of α can be estimated and γ can then be found using

$$\gamma = \left(\frac{\mathbb{E}(|X|^p)}{C_1(p, \alpha)}\right)^{\alpha/p}. \tag{13}$$

B. Logarithm of an SαS Process

Defining $Y = \log|X|$, it can be shown that the mean of Y is given by

$$\mathbb{E}(Y) = C_e\left(\frac{1}{\alpha} - 1\right) + \frac{1}{\alpha}\log(\gamma) \tag{14}$$

where $C_e = 0.57721566\cdots$ is the Euler constant. The variance of Y is given by

$$\operatorname{Var}(Y) = \mathbb{E}\{[Y - \mathbb{E}(Y)]^2\} = \frac{\pi^2}{6}\left(\frac{1}{\alpha^2} + \frac{1}{2}\right). \tag{15}$$

The estimation process involves solving (15) for α and substituting back into (14) to find γ.

IV. Alpha-Stable Modeling of Audio Signals

To demonstrate the α-stable nature of audio noise signals, several measurements were taken in a typical teleconferencing room with dimensions 8.5 m (L) × 7.0 m (W) × 3.5 m (H). The reverberation time was measured using the THX R2 analyzer and was found to be 0.5 s from 125 Hz to 4 kHz. We used an AKG omnidirectional microphone whose signal was fed to a Rane preamplifier and then to a Pentium II PC. The microphone was at a distance of 2 m from the noise sources during the recording.

The recorded sound signals were modeled by means of the SαS family of distributions. The Gaussian density was also used to model the same data, so as to compare the Gaussian and SαS fitting. The two methods described in the previous section—namely, the *positive-order and negative-order* (or sinc method) and the $\log|\text{S}\alpha\text{S}|$ method—were used to estimate the SαS distribution parameters from the data.

Figs. 4 and 5 show the *amplitude probability density* (APD) corresponding to the following.

- The real data. The sum of all the data whose amplitude exceeds the horizontal axis value gives the APD graph. The time series of the real signals is plotted above the corresponding APD's.
- A Gaussian distribution with the same variance and mean as the data.
- An SαS distribution whose parameters α and γ were estimated from the data under study.

Both the sinc method and the $\log|\text{S}\alpha\text{S}|$ method consistently gave the same estimates.

The failure of the Gaussian density to give a good approximation for the tails in the data is apparent in the APD graphs.

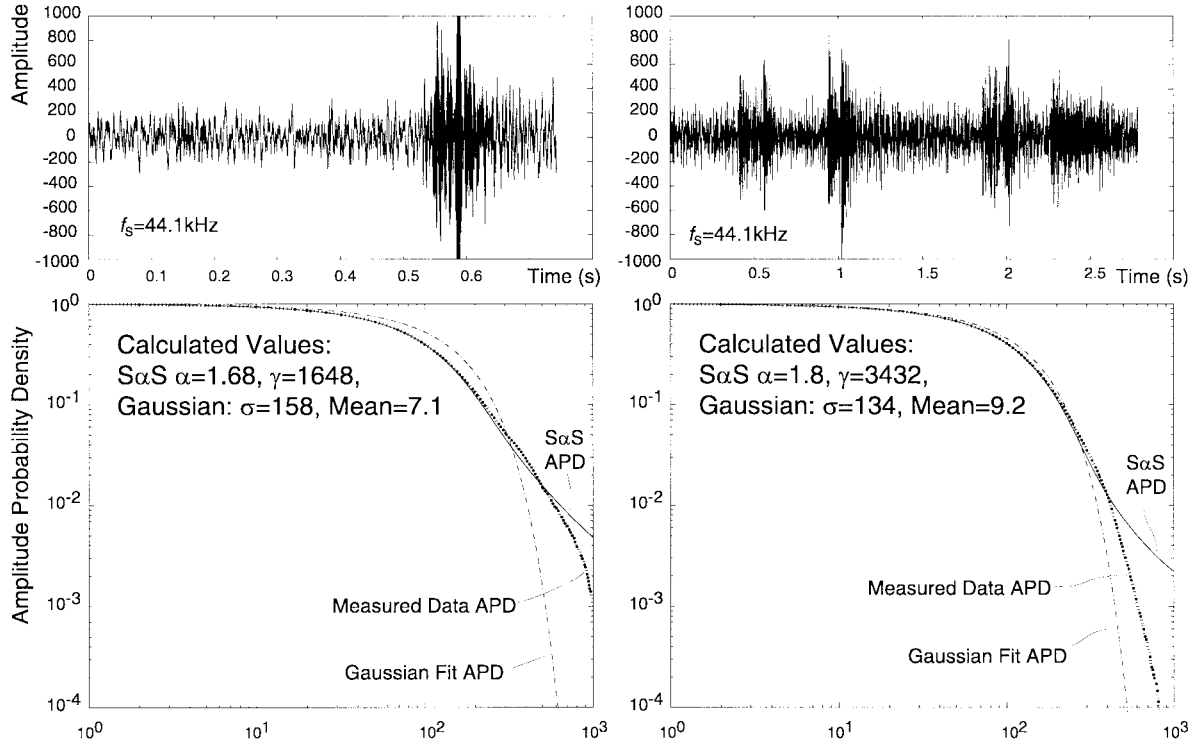

Fig. 4. Time sequences of recorded signals and their associated APD's. The upper two plots show two time sequences of signals recorded in a typical teleconferencing room. The two lower plots are the amplitude probability densities calculated from the recorded signal, as well as the APD's of the best fit SαS and Gaussian distributions. The left signal contains noise produced due to a small roll of a chair, while the right signal contains noise produced from footsteps. It is clear that the SαS gives a much better fit to the data than the Gaussian.

On the other hand, the α-stable density follows the tail of the data much more closely.

The curves shown in Fig. 6 demonstrate the α-stable behavior of the sound signals and are extracts from a much larger sequence of α estimates. The two sequences displayed here are from recordings made in the room described previously with three people present. The estimation of α was performed for each sequence using both the sinc method and the $\log|S\alpha S|$ method for comparison. As can be seen from the two pairs of curves, both methods gave approximately the same estimates. In one recording (denoted as Noise 1), the noise was due to the air conditioning and a computer. In another recording (denoted as Noise 2), the noise present was caused by a slight movement of a chair, a pen drop, and the opening of a CD case. As expected, the α parameter of the measurements changes with time, in agreement with the time-changing statistics of the acoustic environment. The two sound recordings used to generate Fig. 6 gave an average $\alpha = 1.57$ and $\alpha = 1.53$, respectively, when a length of 22.7 s (10^6 samples) was considered. In addition to these recordings, we performed measurements in several other environments such as a small office and a living room. For most of these recordings, the estimated characteristic exponent stabilized in the region between $\alpha = 1.5$ and $\alpha = 1.6$, which is well below the Gaussian model ($\alpha = 2.0$). In some cases, the noise exhibited highly impulsive behavior as shown in Fig. 5 for a recording of pages flipping.

It is important to note that if we treat speech as a random signal and try to estimate α over a large time interval, the resulting value for several speech time series is in the range of $\alpha = 1.5$ to $\alpha = 1.6$. However, when we estimate α using smaller time intervals and then average, the speech signal has a higher value of α. This can be explained by the fact that the speech signal is not random and thus tends to stay at a steady power longer. This appears as a less impulsive behavior on the small scale, but more impulsive on a longer time scale.

V. Application to TDE

Numerous applications can be envisioned in which microphone array steering is desired. For example, in teleconferencing and telepresence systems it is often required to automatically redirect a video camera so that the person speaking is in the field-of-view [7]. This is achievable by bearing estimation from a single or multiple microphone arrays. Mahieux *et al.* [17] present a microphone array for multimedia workstations, which is desirable to provide spatially selective speech acquisition as well as reduce noise and echo.

The localization of a source in audio applications has an added complexity not commonly found in other array processing fields such as radar, which arises from the wideband nature of the signal. Additionally, the statistics are not known *a priori* and they vary with time. For these reasons, the most widely used sound source localization methods are the least computationally intensive algorithms, which are based on TDE. In this section we introduce a new method for TDE based on FLOS of the received signals. We also examine the behavior of the PHAT [5], [8] algorithm, which uses second-order statistics, under stable noise.

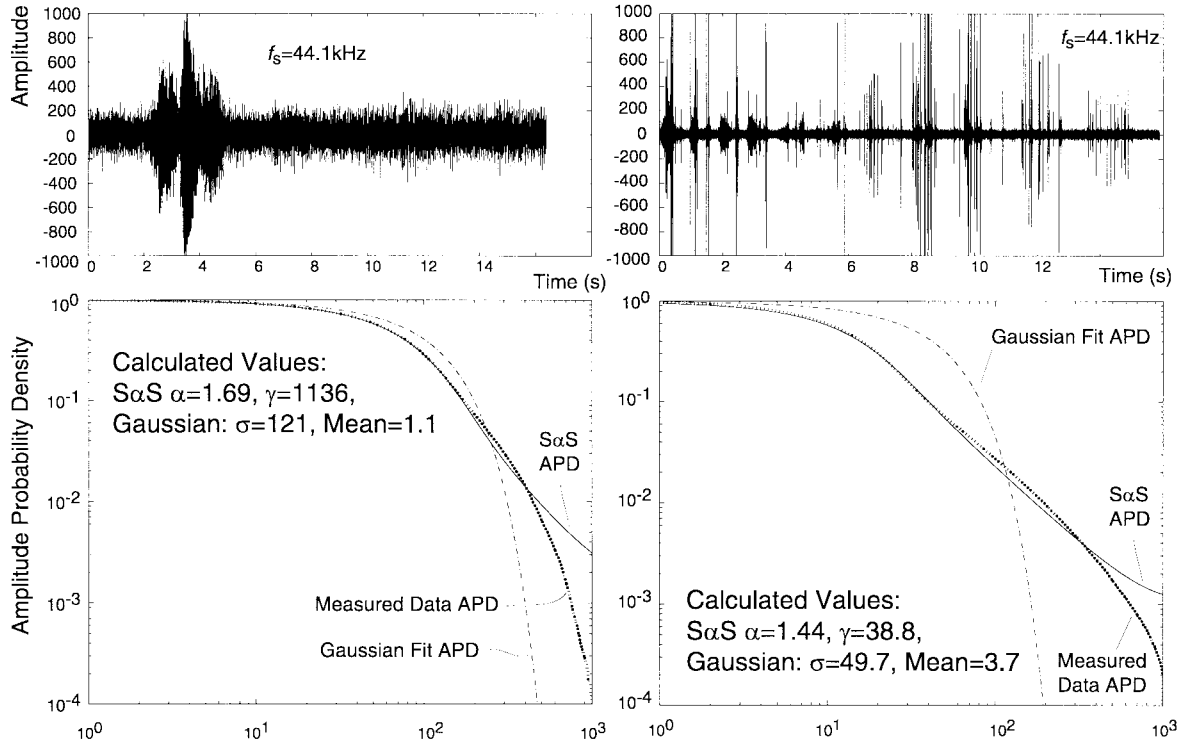

Fig. 5. Time sequences of recorded signals and their associated APD's. The upper two plots show two time sequences of signals, the first recorded in a typical teleconferencing room, and the second in a quiet office. The two lower plots are the amplitude probability densities calculated from the recorded data, as well as the APD's of the best fit SαS and Gaussian distributions. The sudden impulses in the right signal are due to the noise of pages turning.

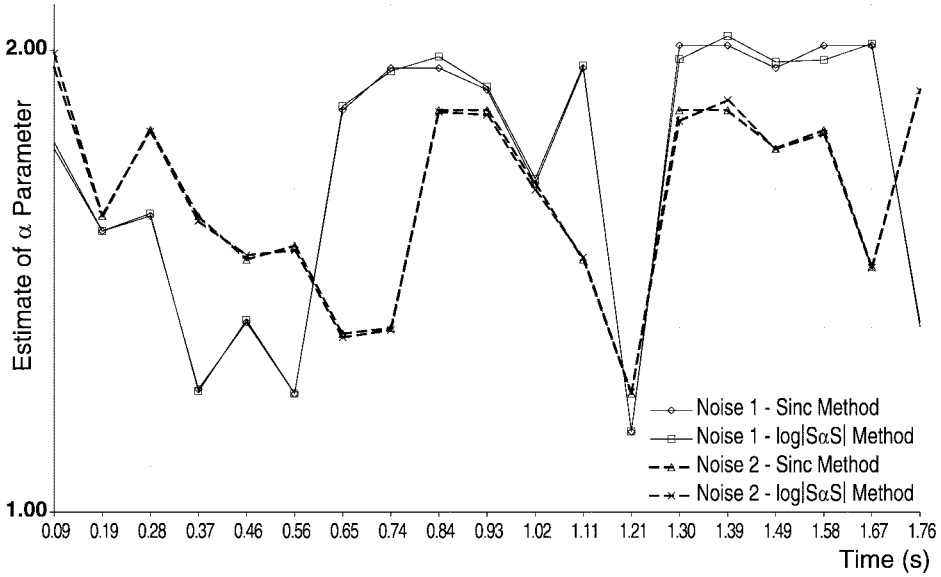

Fig. 6. Estimated α values for the noise measured in a room with characteristics of a typical teleconference room. Two noise recordings are shown. The impulsiveness of each was calculated using two different methods with resolution of 0.08 s and the signals were sampled at 44.1 kHz. It can be seen that the value of α changes with time and in some cases approaches the Cauchy noise case ($\alpha = 1$), while in other occasions follows the Gaussian ($\alpha = 2$) model. The ability to estimate the value of α in real time can offer a significant advantage in implementing near optimal algorithms.

A. Mathematical Formulation

Consider a two-element microphone array receiving signals $r_1(t)$ and $r_2(t)$

$$\begin{aligned} r_1(t) &= x(t) + n_1(t) \\ r_2(t) &= x(t-\tau) + n_2(t) \end{aligned} \tag{16}$$

in which the noise components $n_1(t)$ and $n_2(t)$ are assumed to be zero mean, uncorrelated with each other and the desired speech signal $x(t)$, i.e.,

$$\begin{aligned} &\mathbb{E}[n_1(t_1)n_1^*(t_2)] = \mathbb{E}[n_2(t_1)n_2^*(t_2)] = \sigma_n^2\delta(t_1 - t_2), \\ &\mathbb{E}[x(t_1)n_1^*(t_2)] = \mathbb{E}[x(t_1)n_2^*(t_2)] = 0, \text{ and} \\ &\mathbb{E}[n_1(t_1)n_2^*(t_2)] = 0 \qquad \forall t_1, t_2. \end{aligned} \tag{17}$$

Although most real-world noise signals are correlated, the assumptions above are commonly used in the literature when analytical solutions are desired.

The goal is to estimate the delay τ from measurements of r_1 and r_2, in order to be able to localize the sound source $x(t)$. Transforming the measurements in the frequency domain, we have that

$$\begin{aligned} R_1(k) &= X(k) + N_1(k) \\ R_2(k) &= X(k) \cdot e^{-j\omega_k \tau} + N_2(k). \end{aligned} \tag{18}$$

The second-order cross-correlation function in the frequency domain can then be found from (18). According to our assumptions above, the signal-noise cross terms as well as the noise cross term are zero

$$\begin{aligned} C_{R_1R_2}(k) &= \mathbb{E}\{R_1(k) \cdot R_2(k)^*\} \\ &= \mathbb{E}\left\{|X(k)|^2 e^{j\omega_k\tau}\right\} + \underbrace{\mathbb{E}\{N_1(k)N_2^*(k)\}}_{0} \\ &\quad + \underbrace{\mathbb{E}\{X(k)N_2^*(k)\}}_{0} + \underbrace{\mathbb{E}\left\{X^*(k)N_1(k)e^{j\omega_k\tau}\right\}}_{0}. \end{aligned} \tag{19}$$

Phase Transform Method: A fast method to use for the estimation of the delay between two signals is the PHAT method [5], [8]. According to PHAT, the signal cross spectrum $C_{R_1R_2}(k)$ is smoothed by a window inversely proportional to the magnitude cross spectrum, i.e.,

$$W(k) = \frac{1}{|C_{R_1R_2}(k)|} \tag{20}$$

which will, in turn, give a weighted cross correlation function

$$C^w_{R_1R_2}(k) = \frac{C_{R_1R_2}(k)}{|C_{R_1R_2}(k)|} = e^{j\omega_k\tau}. \tag{21}$$

The inverse Fourier transform will result in a sharp peak in the time domain corresponding to the delay τ. Although this method was expected to be quite sensitive to noise, we found as demonstrated by the simulations that it performed well even for low SNR's.

However, when the process deviates from the ideal Gaussian assumption, and is better characterized by the α-stable class of distributions, performance degrades as will be demonstrated in the simulations. To achieve better performance, we propose a new TDE method based on FLOS, which is robust to heavy noise environments.

B. TDE in Heavy-Tailed Noise—FLOS-PHAT

The *covariation* of two signals x and y is defined as

$$[X, Y]_\alpha \triangleq \int_S xy^{\alpha-1}\mu(d\mathbf{s}) = \frac{\mathbb{E}\left(XY^{<p-1>}\right)}{\mathbb{E}(|Y|^p)}\gamma_y \tag{22}$$

where S is the unit circle, $\mu(.)$ is the spectral measure of the SαS random vector (X, Y), γ_y is the dispersion parameter of signal Y, p satisfies $1 \le p < \alpha$, and $y^{<k>} = |y|^{k-1}y^*$ is the signed-power nonlinearity.

The covariation of complex jointly SαS random variables is not generally symmetric and has the following properties.

P1) If X_1, X_2, and Y are jointly SαS, then

$$[aX_1 + bX_2, Y]_\alpha = a[X_1, Y]_\alpha + b[X_2, Y]_\alpha \tag{23}$$

for any complex constants a and b.

P2) If Y_1 and Y_2 are independent and Y_1, Y_2, and X are jointly SαS, then

$$[aX, bY_1 + cY_2]_\alpha = ab^{<\alpha-1>}[X, Y_1]_\alpha + ac^{<\alpha-1>}[X, Y_2]_\alpha \tag{24}$$

for any complex constants a, b, and c.

P3) If X and Y are independent SαS, then

$$[X, Y]_\alpha = 0. \tag{25}$$

Using covariation properties (23)–(25) and assuming that both the noise and signal have the same distribution, we can now form the covariation of the frequency domain measurement

$$\begin{aligned} D_{R_1R_2}(k) &= [R_1(k), R_2(k)]_\alpha \\ &= \left[X(k) + N_1(k), X(k)e^{-j\omega_k\tau} + N_2(k)\right]_\alpha \\ &= \left[X(k), X(k)e^{-j\omega_k\tau} + N_2(k)\right]_\alpha \\ &\quad + \left[N_1(k), X(k)e^{-j\omega_k\tau} + N_2(k)\right]_\alpha \\ &= [X(k), X(k)]_\alpha \left(e^{-j\omega_k\tau}\right)^{<\alpha-1>} \\ &\quad + \underbrace{[X(k), N_2(k)]_\alpha}_{0} \\ &\quad + \underbrace{[N_1(k), X(k)]_\alpha}_{0} \left(e^{-j\omega_k\tau}\right)^{<\alpha-1>} \\ &\quad + \underbrace{[N_1(k), N_2(k)]_\alpha}_{0} \\ &= [X(k), X(k)]_\alpha \left|e^{-j\omega_k\tau}\right|^{\alpha-2} \left(e^{-j\omega_k\tau}\right)^* \\ &= \gamma_X e^{j\omega_k\tau} \end{aligned} \tag{26}$$

where γ_X is the signal dispersion, which is a real and positive number. From (22) we can see that the denominator is a positive number and thus we can again define a smoothed covariation measure

$$D^w_{R_1R_2} = \frac{D_{R_1R_2}}{|D_{R_1R_2}|} = e^{j\omega_k\tau}. \tag{27}$$

As in the PHAT transform case, the peak in the time domain resulting from the inverse Fourier transform of $D_{R_1R_2}$ will correspond to the delay τ.

In the above expression, the signed-power nonlinearity has been applied to only one of the two signals. A more robust

measure that applies the signed-power nonlinearity to both terms has been proposed by Nikias and Ma [18] to be the *fractional-order correlation function* defined as

$$A_{xy} = \mathbb{E}\{x^{<p>} y^{<q>}\} \tag{28}$$

i.e.,

$$A_{R_1R_2}(k) = \mathbb{E}\{R_1^*(k)^{<p>} \cdot R_2(k)^{<q>}\}. \tag{29}$$

We define the FLOS-PHAT method as

$$A^w_{R_1R_2} = \frac{A_{R_1R_2}}{|A_{R_1R_2}|} = e^{j\omega_k \tau}, \qquad p = q < \frac{\alpha}{2} \tag{30}$$

whose inverse Fourier transform will again result in a sharp peak in the time-domain, corresponding to τ. The moment has to remain of order $p + q < \alpha$ to be defined and thus under a reasonable assumption that both signals are of the same nature, we can choose $p = q < \alpha/2$. Note that when $p = q = 1$, the method reduces to the PHAT method as described above.

VI. Simulation Experiments

In this section, we test the performance of the above algorithms for TDE by adding simulated noise to a real signal obtained in the room described in Section IV. The noise was generated artificially in order to allow us to control the signal-to-noise ratio. In our experiments, the statistics of the real audio recordings vary and must be estimated in real time. Therefore, the TDE algorithm must be fast and able to adapt to new data and statistics. The simple method suggested by this paper is based on the use of blocks of data. All the data used in the experiments described below was obtained using various speech and music signals sampled at typical audio sampling frequencies of 22.05 kHz and 44.1 kHz; thus a block of data of about 1000 samples will introduce a maximum delay of 0.1 s in the time delay estimate. Using overlapping blocks can decrease the delay even more. The algorithm used to obtain the results below can be summarized as follows (see Fig. 7).

- A block of 1024 samples is obtained (using a rectangular window function) from each microphone and their FFT is evaluated.
- The instantaneous second- ($C^i_{R_1R_2}$) and lower-order ($A^i_{R_1R_2}$) statistics (in the frequency domain) are estimated from the data block.
- A weighted-average statistic is obtained. For example, in the case of PHAT

$$[C_{R_1R_2}]_t = (1-\rho)[C_{R_1R_2}]_{t-1} + \rho\left[C^i_{R_1R_2}\right]_t. \tag{31}$$

  The value of ρ, the *adaptation factor,* (where $0 \le \rho \le 1$) determines the tradeoff between speed of adaptation of the algorithm on new statistics (ρ near 1) versus the memory of the algorithm (ρ small).
- The PHAT or FLOS-PHAT algorithm is applied using the appropriate weighted-average statistic evaluated in (31).
- Repeat.

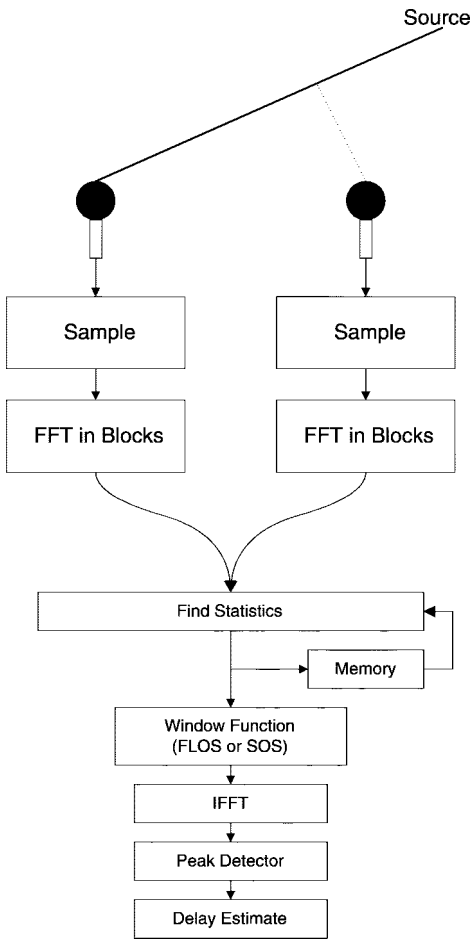

Fig. 7. Block diagram of the proposed TDE algorithm.

An important point here is to define the SNR measure used in this paper. Since power is not defined for α-stable distributions, the conventional definition of SNR cannot be used. Two alternative definitions of SNR are used in literature [19]. In this paper we use the *generalized*-SNR, defined as the ratio of the signal average power to the dispersion of the noise total in the finite interval of interest

$$\text{GSNR} = 10 \log_{10}\left(\frac{1}{\gamma M}\sum_{t=1}^{M} |s(t)|^2\right). \tag{32}$$

We used four different GSNR values of 0, 6, 12, and 25 dB. The comparative values of the GSNR and *effective*-SNR—defined as the average signal power over the average noise power in the finite interval of interest, for the specific data used in Fig. 8—are given in Table I. The FLOS-PHAT method employed constants $p = q = 0.2$ in the expression for the FLOS function [cf. (29)].

The results obtained were based on a set of Monte Carlo runs. Each run starts with an arbitrary vector of statistics and so the algorithm has to adapt to the statistics of the signal. The algorithm converges very fast in about five to ten blocks of data (depending on the GSNR) and then stabilizes until an outlier appears in the noise. After the algorithm reaches steady state, data is gathered to form a "hit/miss" performance curve.

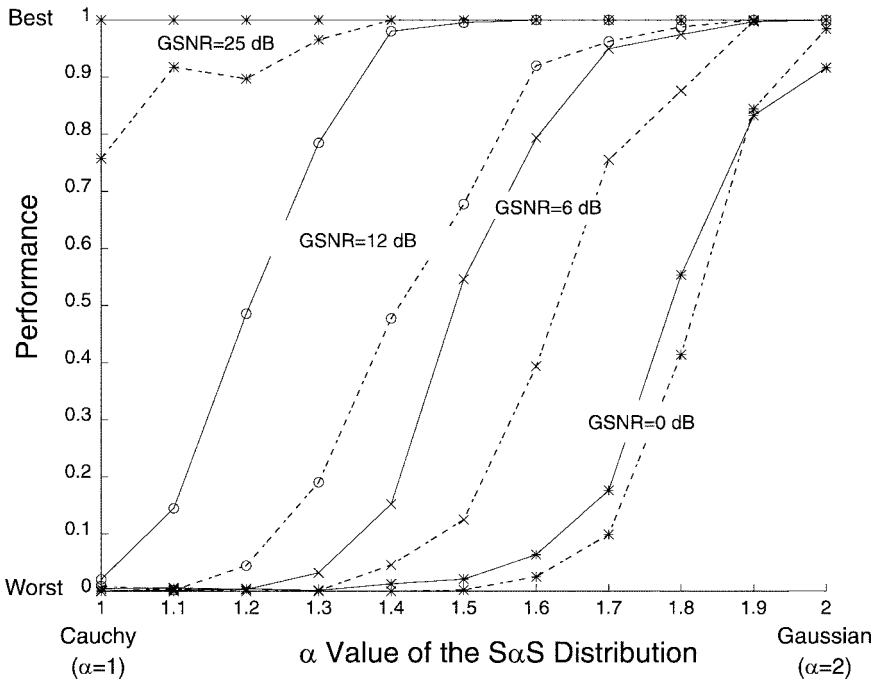

Fig. 8. Comparative performance of the PHAT and FLOS-PHAT methods. Dashed line: PHAT, solid line: FLOS-PHAT. Note, for example, the improvement at GSNR $=$ 12 dB and $\alpha = 1.3$ which is approximately a factor of four. Performance is defined here as the number of correct over the total number of time delay estimates.

TABLE I
CORRESPONDENCE OF GSNR TO EFFECTIVE-SNR. WE USED FOUR DIFFERENT GSNR VALUES OF 0, 6, 12, AND 25 dB TO PRODUCE THE SIMULATIONS OF FIG. 8. THE ASSOCIATED VALUES OF THE GSNR AND EFFECTIVE-SNR—DEFINED AS THE AVERAGE SIGNAL POWER OVER THE AVERAGE NOISE POWER IN THE FINITE INTERVAL OF INTEREST, FOR THE SPECIFIC DATA USED IN FIG. 8—ARE GIVEN BELOW

α	1.0	1.2	1.4	1.6	1.8	2.0
GSNR	Effective-SNR					
0	-52.50	-35.80	-24.79	-15.43	-8.18	-2.94
6	-41.44	-25.45	-15.87	-8.40	-1.9	3.06
12	-28.97	-16.59	-7.43	0.13	4.69	9.06
25	-2.01	4.15	11.36	16.29	19.35	22.06

In total, 4000 values for each point were considered to obtain the curves in Fig. 8.

Fig. 8 shows that in impulsive noise conditions, the FLOS-PHAT method greatly outperforms the PHAT method, sometimes by as much as a factor of four (e.g., at GSNR of 12 dB and $\alpha = 1.3$). As expected, the PHAT outperforms FLOS-PHAT only when the additive noise is Gaussian ($\alpha = 2$). The robust behavior of the introduced FLOS-PHAT method is also apparent when looking at the transient TDE response when outliers occur in the data as seen in Fig. 9. The performance of the PHAT method, based on the time-averaged estimate of $C_{R_1R_2}$, is greatly influenced by impulsive noise. This is due to the fact that an impulsive noise component in the PHAT algorithm remains unaffected while in the FLOS-PHAT it is raised to a fractional power, an operation that limits the effect of the outlier. The FLOS-PHAT method, under severe noise conditions, can also produce a wrong time delay estimate. However, due to the use of fractional lower order statistics, the significance of the outliers is diminished and thus subsequent estimates are less influenced. It should be noted that the transient response shown in Fig. 9 was produced with SαS of $\alpha = 1.2$, which gives an increased number of outliers. This is not a measure of performance, but an indication of the reaction of the two algorithms to outliers in the noise.

In the above simulations we tested FLOS-PHAT with $p = q = 0.2$ and showed that we can achieve better estimates than the ones obtained by the original PHAT method that in fact uses $p = q = 1$.

An interesting question is to determine how well this method performs for different values of p, q or rather what are the appropriate values of p, q to use for each value of α. Several simulations were performed in an attempt to find what this value is. In all instances we assumed that $p = q$ and performed the simulation for values of $p = 0$ to $p = 1$ (corresponding to the PHAT method). Although it is clear that the performance is much better at about $p = 0.15$ when the noise is quite impulsive ($\alpha = 1$ to 1.7), (cf. Fig. 10) it is difficult to draw a clear conclusion as to where the best performance lies for higher values of α. However, we do expect that the value of $p = q = 1$ will be the optimal for $\alpha = 2$. Finding the optimum p, q becomes even more difficult when the data used is from a real speech signal since there is some impulsive behavior in the speech itself, as we showed, which is nonstationary and which influences that value. The simulations of Fig. 10 show how the performance (vertical axis) changes versus p (horizontal axis) for different α.

As we can see, the optimum value of p and q initially rises slowly from about $p = 0.15$ and then faster when $\alpha > 1.8$.

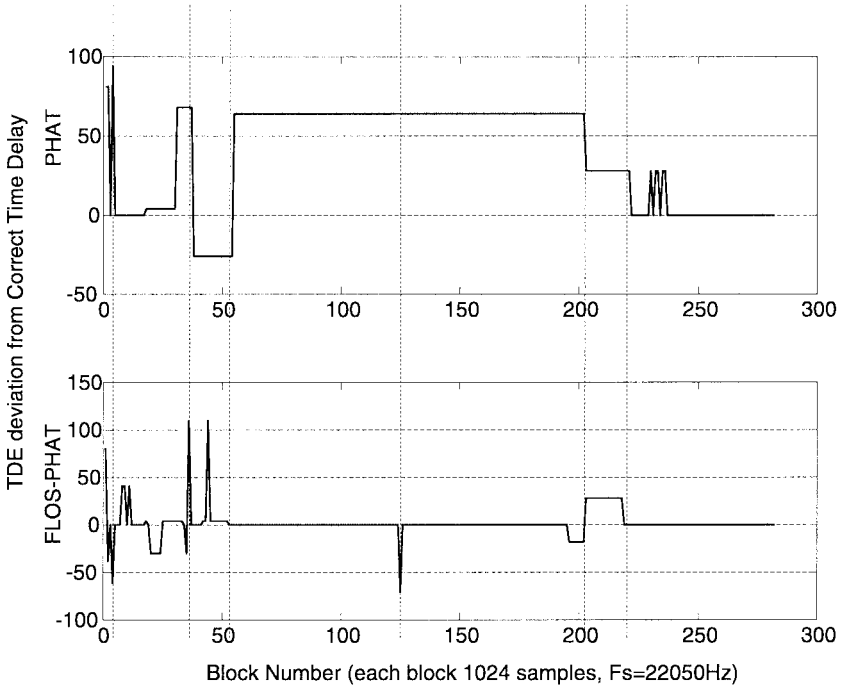

Fig. 9. Transient performance of the PHAT (top) and FLOS-PHAT (bottom) methods. Plots show the offset error from the correct TDE versus block number. It can be seen that the FLOS-PHAT is more robust than the PHAT when the noise is impulsive. In the case of the PHAT algorithm, an impulsive element in the noise creates a very large error in the statistics which takes a long time to revert, while in the case of the FLOS-PHAT the statistics are influenced much less and the algorithm returns to its correct state much faster.

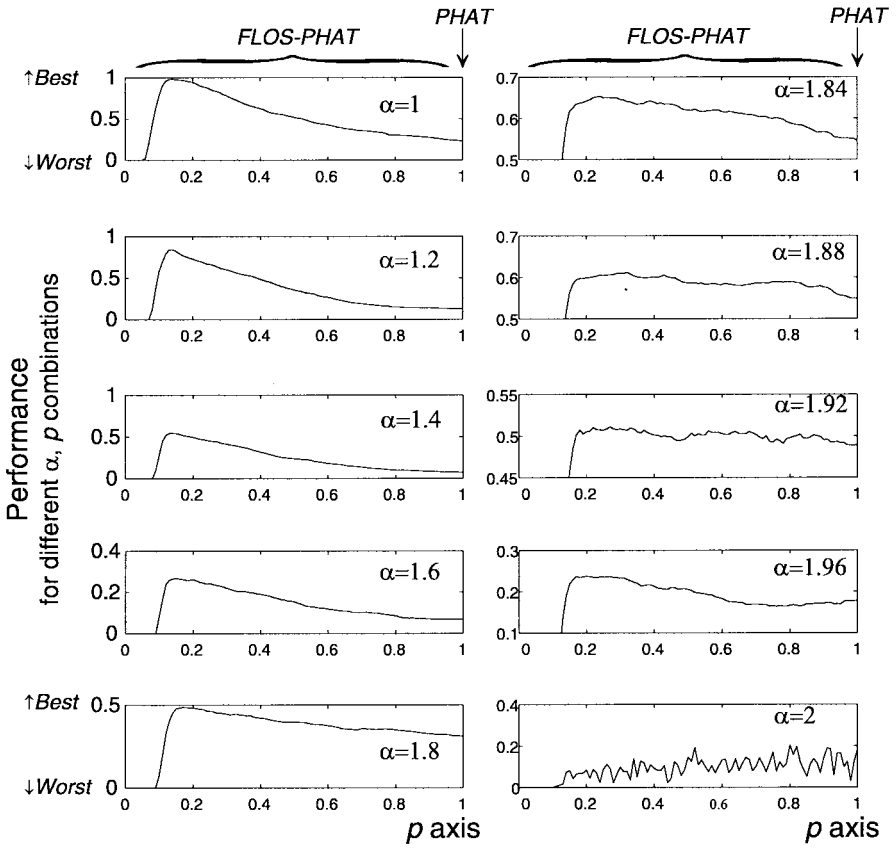

Fig. 10. TDE performance as a function of p. Values range from $\alpha = 1$ (top) to $\alpha = 2$ (bottom) with resolutions $\alpha = 0.2$ and $\alpha = 0.02$. A performance of one means that the TDE was correct at all instances, while zero means we had no correct TDE. For impulsive noises with $\alpha < 1.8$ it is clear that the performance gain by using FLOS-PHAT is significant. p is the fractional-power nonlinearity parameter of the FLOS-PHAT algorithm.

At values closer to $\alpha = 2$, the impulsiveness of the speech signal starts to become significant, and thus we cannot see a clear peak at $p = 1$, although we see the trend of an improved performance for an increasing p. As a conclusion, we could say that since the performance improvement is expected to be around 50% between 12 and 16 dB GSNR, it seems a logical choice to keep our algorithm working at a low value of p. Even if we were to operate at the $\alpha = 2$ the performance degradation would be negligible with a low value of p compared to the gain achieved at low α. In situations where the noise characteristic exponent and the GSNR where known stationary, an optimum value of p, q could be estimated. This is never the case with audio signals, and therefore the value of $p = 0.15$ to $p = 0.25$ is a reasonable choice for TDE.

VII. Conclusions

This paper has presented a better model for noise encountered in typical reverberant rooms. The model presented uses the symmetric α-stable class of distributions, of which the Gaussian is a special case, and shows that noise in the room tends to have a value of $\alpha \simeq 1.6$, which is lower than the usually assumed $\alpha = 2$ Gaussian model.

Based on this observation, we have presented a new method for adaptively steering microphone arrays in the presence of SαS noise. Our method, based on fractional lower order statistics of the measurements, was tested to be better than the second-order-based PHAT algorithm, while at the same time adding little computational expense. It is a simple algorithm that exhibits robust performance even for small values of α and can be applied to the "speaker tracking" problem.

The FLOS-PHAT time delay estimator that we introduced is a class of methods parameterized by p and q. When $p = q = 1$, the conventional second-order PHAT algorithm is obtained as a special case. By choosing the parameters p and q according to the statistics of the underlying acoustical environment, we can operate robustly close to a near optimal point. The comparison in this paper between the PHAT and the FLOS-PHAT is a demonstration of the advantages that the use of α-stable distributions and fractional lower order statistics can offer to audio applications.

Acknowledgment

The authors are very grateful to Prof. C. L. Nikias, who initiated the study on the subject, and whose knowledge and support are always available.

References

[1] M. Shao and C. L. Nikias, "Signal processing with fractional lower order moments: Stable processes and their applications," *Proc. IEEE,* vol. 81, pp. 986–1010, July 1993.

[2] C. L. Nikias and M. Shao, *Signal Processing with Alpha-Stable Distributions and Applications*. New York: Wiley, 1995.

[3] R. Adler, R. E. Feldman, and M. S. Taqqu, Eds., *A Practical Guide to Heavy Tails: Statistical Techniques and Applications*. Boston, MA: Birkhäuser, 1998.

[4] P. Tsakalides, R. Raspanti, and C. L. Nikias, "Joint target angle and Doppler estimation in stable impulsive interference," *IEEE Trans. Aerosp. Electron. Syst.,* vol. 32, Apr. 1999.

[5] C. H. Knapp and G. C. Carter, "The generalized correlation method for estimation of time delay," *IEEE Trans. Acoust., Speech, Signal Processing,* vol. ASSP-24, pp. 320–327, Aug. 1976.

[6] G. C. Carter, "Guest editorial—Time delay estimation," *IEEE Trans. Signal Processing,* vol. ASSP-29, p. 461, June 1981.

[7] M. S. Brandstein, "A framework for speech source localization using sensor arrays," Ph.D. dissertation, Brown Univ., Providence, RI, May 1995.

[8] P. A. Petropulu and C. L. Nikias, *Higher Order Spectral Analysis: A Nonlinear Signal Processing Framework*. Englewood Cliffs, NJ: Prentice Hall Signal Processing Series, 1993.

[9] H. Stark and J. W. Woods, *Probability, Random Processes and Estimation Theory for Engineers,* 2nd ed. Englewood Cliffs, NJ: Prentice Hall, 1994.

[10] G. Samorodnitsky and M. S. Taqqu, *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*. New York/London: Chapman & Hall, 1994.

[11] W. H. DuMouchel, "Stable distributions in statistical inference," Ph.D. dissertation, Dept. of Statistics, Yale University, New Haven, CT, 1971.

[12] B. W. Brorsen and S. R. Yang, "Maximum likelihood estimates of symmetric stable distribution parameters," *Commun. Statist.-Simul.,* vol. 19, pp. 1459–1464, 1990.

[13] A. S. Paulson, E. W. Holcomb, and R. A. Leitch, "The estimation of the parameters of the stable laws," *Biometrika,* vol. 62, pp. 163–170, 1975.

[14] V. M. Zolotarev, "Statistical estimates of the parameters of stable laws," *Math. Stat.: Banach Center Pub.,* vol. 6, pp. 359–376, 1980.

[15] J. H. McCulloch, "Simple consistent estimators of stable distribution parameters," *Commun. Statist.-Simul.,* vol. 15, pp. 1109–1136, 1986.

[16] E. F. Fama and R. Roll, "Some properties of symmetric stable distributions," *J. Amer. Statist. Assoc.,* vol. 63, pp. 817–836, 1968.

[17] Y. Mahieux, G. Le Tourneur, and A. Saliou, "A microphone array for multimedia workstations," *J. Audio Eng. Soc.,* vol. 44, no. 5, pp. 365–372, May 1996.

[18] X. Ma and C. L. Nikias, "Joint estimation of time delay and frequency delay in impulsive noise," *IEEE Trans. Signal Processing,* vol. 44, pp. 2669–2687, Nov. 1996.

[19] P. Tsakalides and C. L. Nikias, "Maximum likelihood localization of sources in noise modeled as a stable process," *IEEE Trans. Signal Processing,* vol. 43, pp. 2700–2713, Nov. 1995.

Panayiotis G. Georgiou (S'98) received the B.A. and M.Eng. degrees in electrical and information sciences from Cambridge University, Cambridge, U.K., in 1996. He completed the M.S. degree at the University of Southern California, Los Angeles, and is working toward the Ph.D. degree.

His research interests include time delay estimation techniques, broadband array signal processing, lower order statistics, and spatial sound rendering.

Panagiotis Tsakalides (S'93–M'95) received the Ph.D. degree in electrical engineering in 1995 from the University of Southern California (USC), Los Angeles.

He was a Research Assistant Professor at USC from 1996 to 1998. He is currently with the Department of Electrical Engineering at the University of Patras, Patras, Greece. His research interests include statistical signal processing with emphasis on space-time adaptive processing for wireless communications, sonar, and radar applications.

Chris Kyriakakis (M'96) received the Ph.D. degree in electrical engineering from the University of Southern California (USC), Los Angeles, in 1993.

He is an Assistant Professor of electrical engineering as well as an investigator in the Integrated Media Systems Center (IMSC), an NSF Engineering Research Center at USC. His research is in the areas of immersive audio signal processing, adaptive beamforming, and immersive telepresence.