ALPHA-STABLE ROBUST MODELING OF BACKGROUND NOISE FOR ENHANCED SOUND SOURCE LOCALIZATION

P. G. Georgiou, P. Tsakalides, and C. Kyriakakis

Integrated Media Systems Center
University of Southern California
3740 McClintock Ave., EEB 432
Los Angeles, CA 90089, USA.
email: georgiou@sipi.usc.edu

ABSTRACT

In this paper we address the problem of sound source localization in the presence of impulsive noise for application in immersive telepresence and teleconferencing. Traditional Gaussian modeling of noise signals fails when the signals exhibit impulsive behavior. A new model is used, namely the Symmetric α-Stable ($S\alpha S$), which can better account for the outliers that exist in real-world signals. Real data is used to compare the performance of both the Gaussian and the α-stable models. We demonstrate that the α-stable model gives a much better approximation to the noise signal than the Gaussian model.

Furthermore, we study the problem of Time Delay Estimation (TDE) and we demonstrate the shortcomings of TDE techniques based on second-order statistics when the noise is of $S\alpha S$ nature. We propose an alternative to second-order based methods, based on Fractional Lower-Order Statistics, and demonstrate the achieved improvement via simulation experiments.

1. INTRODUCTION

Although several distributions exist that are good candidates for signal modeling, it is common in the literature to use the Gaussian distribution model. The majority of *Time Delay Estimation* (TDE) [1, 2] methods proposed so far for audio applications assume a Gaussian noise signal and use second or higher-order statistics to locate the source. A drawback of this assumption is that should the signal deviate from the Gaussian model, the method becomes suboptimal and in many cases unusable.

In this paper we first assume a Gaussian noise signal and compare it with real measured data in a real-world teleconferencing environment. We then proceed to describe the class of α-stable distributions, and show that this class of distributions gives a more accurate model of the measured audio signals.

The α-stable model is then applied to the problem of time delay estimation using both a traditional second-order statistics method (PHAT - Phase Transform Method [1]) as well as a *Fractional Lower-Order Statistics* method (FLOS-PHAT). 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 gives better detection than the PHAT.

2. ALPHA-STABLE DISTRIBUTIONS

The α-stable distribution is more impulsive than the Gaussian, and is appealing because it satisfies the *Stability Property*, as well as the *Generalized Central Limit Theorem* [3, 4] .

Although there is no closed form solution for the probability density function of α-stable distributions, the characteristic function is given by

$$\varphi(t) = \exp\left(j\lambda t - \gamma|t|^{\alpha}\left[1 + j\beta \mathrm{sign}(t)\omega(t,\alpha)\right]\right) \qquad (1)$$

in which α is the *characteristic exponent* satisfying $0 < \alpha \leq 2$. The characteristic exponent controls the heaviness of the tails of the density function. For low values of α the tails are heavier and thus the noise is more impulsive, while for a larger α the distribution exhibits less impulsive behavior. The *location parameter* is denoted by λ and corresponds to the mean for $1 < \alpha \leq 2$ and the median for $0 < \alpha \leq 1$. The *dispersion* parameter (γ) behaves similarly to the variance, and is in fact equal to one-half the variance in the Gaussian case ($\alpha = 2$). Finally, the parameter β is the index of symmetry.

In this paper, we will deal with the class of *Symmetric α-Stable* ($S\alpha S$) distributions ($\beta = 0$) with finite mean, *i.e.* $1 < \alpha \leq 2$. It should be noted that the class of α-stable distributions, does not possess finite second (or higher) moment statistics. In fact, α-stable distributions with $\alpha \neq 2$ have finite moments only for order $p < \alpha$.

3. ALPHA-STABLE MODELING OF SOUND

Several methods have been proposed for the estimation of the parameters of the α-stable distribution [3, 5]. For audio signals we can assume that the distribution will be of the $S\alpha S$ class. In this paper we use the *Positive-Order and Negative-Order (Sinc) Function* and the *Logarithm of $S\alpha S$ process* estimation methods.

To demonstrate the heavy-tailed nature of sound, several measurements were taken in a typical noisy environment. We then estimated the α and γ parameters assuming that the distributions of the measured signals were of the $S\alpha S$ type (which includes the Gaussian case). As expected the mean was zero for large data samples.

In Fig. 1 we show the *Probability Density Function* (PDF) and *Amplitude Probability Density* (APD) of the real measured data and compare it to: (i) a Gaussian distribution with the same variance as the entire data set; (ii) a Gaussian distribution with the same variance as the bulk of the data. The variance for this case is found after the tails above 30% of the maximum amplitude are cut off; and (iii) a $S\alpha S$ matching the calculated α and γ. A histogram is plotted in the PDF graph, while the sum of all data values whose

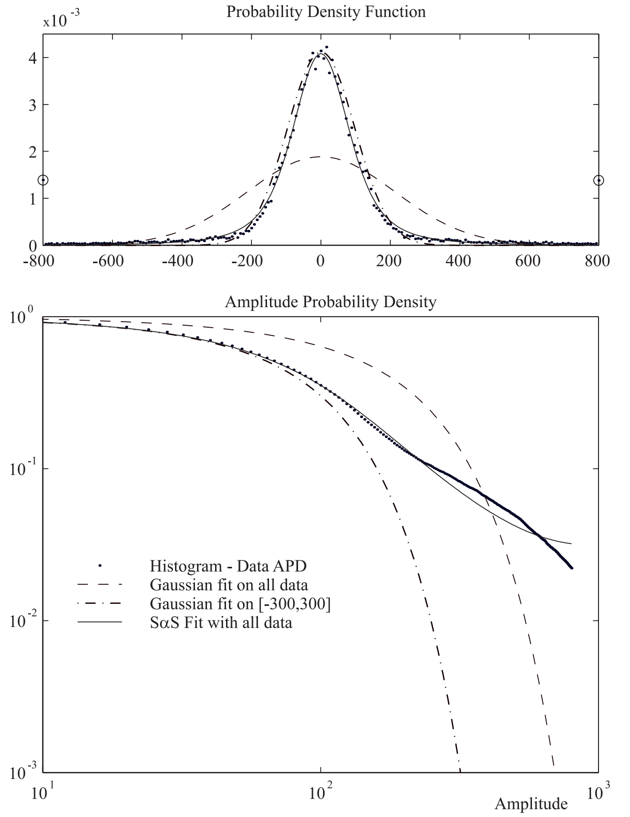

Figure 1: Comparison of the Gaussian, $S\alpha S$, and real data PDF and APD's.

amplitude exceeds the horizontal axis value is used for the APD graph.

Although it may seem under casual observation that the Gaussian with a variance calculated from the bulk of the data might be a good fit, a careful examination of the PDF curves shows that the Gaussian approaches zero probability much faster than the real data histogram. This can be seen more clearly in the APD graphs. It should also be noted that there is a large amount of data with amplitude greater than 80% that appear at ±800 on the PDF plot of Fig. 1 (data were normalized between -1000 and 1000).

The two graphs shown in Fig. 2 demonstrate the α-stable behavior of sound and are extracts from a much larger sequence of α estimates. Both sequences displayed here are from a recording made in a relatively quiet room with three people engaged in normal activities. In the first signal, the noise is mainly from the air flow through the air-conditioning vent in the ceiling and from the spinning hard drive in the computer. In the second signal there was noise from slightly moving a chair, dropping a pen, and opening a CD case. The α parameter of the measurements, as expected, changes with time. This is in agreement with α-stable simulations in which the $S\alpha S$ noise behaves in a relatively Gaussian-like fashion for a large stretch of time, but occasionally presents outliers of much higher amplitude. In fact most of the recordings we made, stabilized in the region of $\alpha = 1.5$ to $\alpha = 1.6$ depending on the noise environment. The time scale considered exceeded one million samples and the sampling rate was 44.1 kHz. The two noise signals described above gave an $\alpha = 1.57$ and $\alpha = 1.53$ respectively for a 22.7 second recording (1 million samples). Other recordings that contained additional background noise (from another computer) showed very similar behavior, with a slightly lower overall value of α.

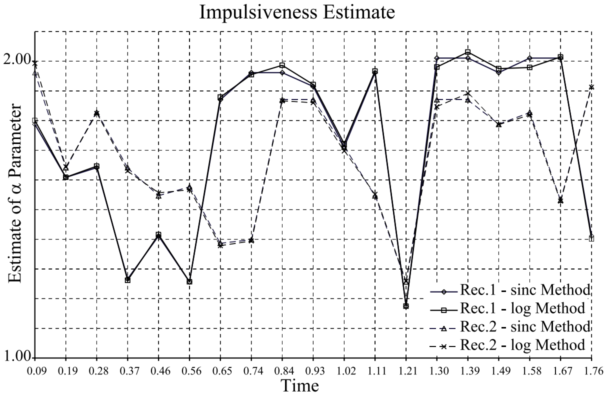

Figure 2: Estimated α values for noise recordings in a typical teleconference environment.

4. APPLICATION TO TIME DELAY ESTIMATION

Numerous applications can be envisioned in which microphone array steering is desired. For example, in teleconferencing and telepresence systems it is often required to redirect a video camera so that the person speaking is in the field-of-view. In multi-participant environments it is desirable to provide spatially-selective speech acquisition as well as noise and echo cancellation.

The localization of a source for audio applications adds complexity not commonly found in other signal processing applications, which arises from the wideband nature of the signal. Additionally, the statistics are not known *a-priori* and they may vary with time.

Inter-sensor *Time Delay Estimation* (TDE) is a method commonly used ([6] and ref. therein) to estimate the position of the source using bearing information. The majority of TDE methods proposed so far in audio applications use second or higher-order statistics of the measurements to locate the signal of interest. A drawback of these methods is that in impulsive noise or severe interference environments, which as we showed above are best described by the α-stable family of distributions, second or higher-order statistics are not theoretically defined.

In this section we introduce a new method for TDE based on *Fractional Lower-Order Statistics* (FLOS-PHAT) of the received signals. We also examine the behavior of the *Phase Transform* (PHAT) algorithm, which uses second-order statistics, under stable noise. We show that when the Gaussian noise assumption fails, then the FLOS-PHAT algorithm gives better detection than the PHAT.

4.1. Mathematical Formulation

Consider a two-element microphone array receiving

$$r_1(t) = x(t) + n_1(t) \text{ and } r_2(t) = x(t-\tau) + n_2(t) \quad (2)$$

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

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 $x(t)$. We are interested in localizing wideband signals, hence we transform the measurements into the frequency domain

$$\begin{aligned} R_1(k) &= [X(k) + N_1(k)] \\ R_2(k) &= [X(k) \cdot e^{-j\omega_k \tau} + N_2(k)] \end{aligned} \tag{3}$$

The second-order cross-correlation function, in the frequency domain can then be found (4). Note that the signal-noise and noise-noise cross terms are zero according to our assumptions above.

$$\begin{aligned} & C_{R_1 R_2}(k) = E\Big\{R_1(k) \cdot R_2(k)^*\Big\} \\ &= \quad E\Big\{|X(k)|^2 e^{j\omega_k \tau}\Big\} + \underbrace{E\Big\{N_1(k) N_2^*(k)\Big\}}_{0} \\ &+ \quad \underbrace{E\Big\{X(k) N_2^*(k)\Big\}}_{0} + \underbrace{E\Big\{X^*(k) N_1(k) e^{j\omega_k \tau}\Big\}}_{0} \end{aligned} \tag{4}$$

4.1.1. Phase Transform Method

A fast method to use for the estimation of the delay between two signals is the *Phase Transform* method [1]. In PHAT the signal cross spectrum $C_{R_1 R_2}(k)$ is smoothed by a window inversely proportional to the magnitude cross spectrum.

$$W(k) = \frac{1}{|C_{R_1 R_2}(k)|} \tag{5}$$

This in turn gives a weighted cross correlation function

$$C^w_{R_1 R_2}(k) = \frac{C_{R_1 R_2}(k)}{|C_{R_1 R_2}(k)|} = e^{j\omega_k \tau} \tag{6}$$

The inverse Fourier transform generates a sharp peak in the time domain corresponding to value of the delay τ. Although this method was expected to be quite sensitive to noise, our simulations showed 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 significantly as we demonstrate in Section 5 below.

4.2. TDE in Heavy-Tailed Noise

4.2.1. FLOS-PHAT

The *covariation* of two signals, x and y is

$$[X, Y]_\alpha = \int_S x y^{\alpha - 1} \mu(d\mathbf{s}) = \frac{E(XY^{<p-1>})}{E(|Y|^p)} \gamma_y \tag{7}$$

in which S is the unit circle, $\mu(.)$ is the spectral measure of the $S\alpha S$ random vector (X,Y), $1 \leq p < \alpha$ and $y^{<k>} = |y|^{k-1} y^*$.

For α-stable distributions, the frequency domain representation of a signal does not converge as $T \to \infty$, but does exist for finite T (*i.e.* after smoothing by a window) [7]. Thus we can express the received signals in the frequency domain as shown in eq. (3). Using the properties of $S\alpha S$ distributions [8], and assuming that both the noise and signal have the same distribution, we can now form the covariation

$$\begin{aligned} & D_{R_1 R_2}(k) = \Big[R_1, R_2\Big]_\alpha \\ &= \Big[X(k), X(k) e^{-j\omega_k \tau} + N_2(k)\Big]_\alpha \\ &\quad + \Big[N_1(k), X(k) e^{-j\omega_k \tau} + N_2(k)\Big]_\alpha \\ &= \Big[X(k), X(k)\Big]_\alpha \Big(e^{-j\omega_k \tau}\Big)^{<\alpha - 1>} + \Big[X(k), N_2(k)\Big]_\alpha \\ &= \Big[X(k), X(k)\Big]_\alpha \Big|e^{-j\omega_k \tau}\Big|^{\alpha - 2} \Big(e^{-j\omega_k \tau}\Big)^* = B e^{j\omega_k \tau} \end{aligned} \tag{8}$$

in which B is a real and positive number. We can again define a smoothed covariation measure

$$D^w_{R_1 R_2} = \frac{D_{R_1 R_2}}{|D_{R_1 R_2}|} = e^{j\omega_k \tau} \tag{9}$$

As in the PHAT transform case, the peak in the time domain, resulting from the inverse Fourier transform of $D_{R_1 R_2}$, will correspond to the delay τ.

It has been shown [9] that an even better measure is the *Fractional Order Correlation Function* defined as:

$$A_{R_1 R_2}(k) = E\left\{R_1^*(k)^{<a>} \cdot R_2(k)^{<b>}\right\} \tag{10}$$

Based on this we define [10] the FLOS-PHAT method

$$A^w_{R_1 R_2} = \frac{A_{R_1 R_2}}{|A_{R_1 R_2}|} = e^{j\omega_k \tau} + \varepsilon_k, \qquad a = b < \frac{\alpha}{2} \tag{11}$$

whose inverse Fourier transform will again result in a sharp peak in the time-domain, corresponding to τ.

5. SIMULATION RESULTS

To test the performance of the above algorithms we must make use of estimation techniques due to the absence of second and fractional lower-order statistics. In the case under consideration, the statistics of the problem are not known and they vary with time. The algorithm therefore must be fast and able to adapt to new data and statistics. The simple method suggested in this paper is based on the use of blocks of data and can be summarized as follows: a block of 1024 samples is obtained from each microphone and their FFT is evaluated. The instantaneous second and lower-order statistics (in the frequency domain) are found and a weighted-average statistic is obtained *i.e.*

$$\Big[C_{R_1 R_2}\Big]_t = (1 - \rho) \Big[C_{R_1 R_2}\Big]_{t-1} + \rho \Big[C_{R_1 R_2}\Big]_t \tag{12}$$

in which ρ is the *adaptation factor* satisfying $0 \leq \rho \leq 1$. The PHAT or FLOS-PHAT algorithm is applied using the appropriate weighted-average statistic and the process is repeated.

An important point here is to define an SNR measure. Because power is not defined for α-stable distributions, the conventional definition of SNR can not be used. Two alternative definitions of

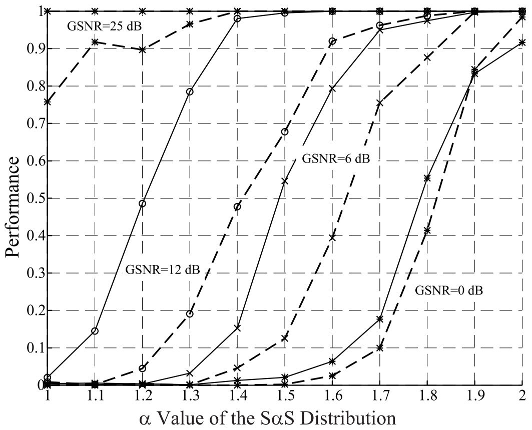

Figure 3: Comparative Performance of the PHAT and FLOS-PHAT methods with $\rho = 0.0125$ and $a = b = 0.2$. Dashed line: PHAT, Solid line: FLOS-PHAT

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	-1.95	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

Table 1: Correspondence of GSNR and average *Effective*-SNR for the specific noise present in the measurements.

SNR have been proposed [11]. 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

$$\mathrm{GSNR} = 10\log_{10}\left(\frac{1}{\gamma M}\sum_{t=1}^{M}|s(t)|^2\right) \tag{13}$$

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. The results obtained were based on a set of Monte-Carlo runs. Each run starts with a "wrong delay" vector of statistics and so the algorithm has to adapt to the statistics of the signal. After the algorithm reaches steady state, data is gathered to form a "hit/miss" performance curve. In total, 4000 values for each point were considered to obtain the curves in Fig. 3.

The tests were all done with a constant $a = b = 0.2$ value and for GSNR's 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 – are summarized in Table 1.

Our results indicate that in impulsive noise conditions, the FLOS-PHAT method greatly outperforms the PHAT method for Time Delay Estimation, sometimes by as much as 50% except for the Gaussian ($\alpha = 2$) case in which the PHAT performs better.

6. CONCLUSIONS

In this paper we presented a method for modeling the noise encountered in audio environments based on the symmetric α-stable class of distributions. Our results show that noise signals in a typical office have an α in the range of 1.5 to 1.6, which deviates from the Gaussian case ($\alpha = 2.0$) that is typically assumed.

Based on these findings that are supported by our measurements in a real-world environment, we have presented a new method for adaptively steering microphone arrays in the presence of such $S\alpha S$ noise. Our method, based on fractional lower-order statistics of the measurements, performed better than the second-order based PHAT algorithm, while at the same time adding little computational expense. It is a simple algorithm that gives very good performance even for small values of α, and can be applied to the speaker tracking problem for real-time applications.

The enhanced performance of the FLOS-PHAT over the PHAT method demonstrated in this paper shows the advantages of using α-stable distributions in audio applications. Further research directions include modeling the correlation structure of heavy-tailed noise using sub-Gaussian processes as well as other impulsive multidimensional distributions and studying algorithms based on negative order moments.

7. REFERENCES

[1] C. H. Knapp and G. C. Carter, "The generalized correlation method for estimation of time delay," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. ASSP-24, no. 4, pp. 320–327, August 1976.

[2] G. C. Carter, "Guest editorial - time delay estimation," *IEEE Transactions on Signal Processing*, vol. ASSP-29, no. 3, pp. 461, June 1981.

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

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

[5] X. Ma and C. L. Nikias, "Parameter estimation and blind channel identification for impulsive signal environments," *IEEE Transactions on Signal Processing*, vol. 43, no. 12, pp. 2884–2987, December 1995.

[6] M. S. Brandstein, *A Framework for Speech Source Localization Using Sensor Arrays*, Ph.D. thesis, Brown University, May 1995.

[7] E. Masry and S. Cambanis, "Spectral density estimation for stationary stable processes," *Stochastic Processes and Their Applications*, vol. 18, pp. 1–31, 1984.

[8] P. Tsakalides and C. L. Nikias, "The robust covariation based music (roc-music) algorithm for bearing in impulsive noise environments," *IEEE Transactions on Signal Processing*, vol. 44, no. 7, pp. 1623–1633, July 1996.

[9] X. Ma and C. L. Nikias, "Joint estimation of time delay and frequency delay in impulsive noise," *IEEE Transactions on Signal Processing*, vol. 44, no. 11, pp. 2669–2687, November 1996.

[10] P. G. Georgiou, C. Kyriakakis, and P. Tsakalides, "Robust time delay estimation for sound source localization in noisy environments," *IEEE Proceedings*, WASPAA 1997.

[11] P. Tsakalides and C. L. Nikias, "Maximum likelihood localization of sources in noise modeled as a stable process," *IEEE Transactions on Signal Processing*, vol. 43, no. 11, pp. 2700–2713, November 1995.