THE USE OF FRACTIONAL LOWER-ORDER STATISTICS IN ACOUSTICAL ENVIRONMENTS

J. Michael Peterson, Panayiotis G. Georgiou and Chris Kyriakakis

Immersive Audio Laboratory
Integrated Media Systems Center
3740 McClintock Ave., EEB 432
Los Angeles, CA 90089-2564
johnmpet@usc.edu

ABSTRACT

Robust time delay estimation is necessary for many microphone array applications. Reverberant echoes and background noise can corrupt the estimates. Typically the signals and noise are modeled as Gaussian random processes and second moment information is used for time delay estimation. However, many signals are impulsive and should be modeled using alpha-stable distributions. Because of this impulsive nature, the second-order moments are theoretically undefined. This paper will review the important characteristics of alpha-stable random processes. Time delays will be found using lower order moments. Finally experimental evidence will show the superiority of lower order moments for TDE in real acoustical environments.

1. INTRODUCTION

While the algorithms and techniques for time delay estimation are well developed in a theoretical and laboratory setting, the implementation of existing methods lack robustness in real acoustical settings. There are many situations where accurate time delay estimates (TDE) are necessary. For example, the performance of adaptive beam-formers suffers when the actual direction of arrival (DOA) differs from the estimated DOA. Another possible application could use the time delay values to find the location of the sound source. The location could be sent to a video camera, which would then focus on the current speaker [1]. In addition the locations can be used to recreate an audio scene for remote immersi-presence or teleconferencing.

Among the challenges for acoustical TDE are the presence of reflections and reverberation, the wide-band nature of the signals, and the large amount of background noise. Typically this background noise is modeled as Guassian. With this model, the time delays are best estimated using second order statistics [2, 3]. Generalized cross-correlation (GCC) and many maximum-likelihood and eigenspace-based techniques are good examples of algorithms that use second-order statistics. While this assumption holds true for some typical noises like fan noises, in general most noise is impulsive in nature, including closing doors, dropping pens, and creaking chairs. So it is very likely that a deficiency of most existing algorithms is due to a model-mismatch.

A generalization of the Guassian distribution, called the alpha-stable distribution, can be used to model these noise signals. Gaussian distributions are alpha-stable with $\alpha = 2$. When α is less than 2, then they have heavier tails and are more impulsive. In theory, the second order moment is undefined for such signals. It has been shown that many signals encountered in acoustical environments, including speech, music, and other sounds, can be modeled as an alpha-stable distribution with $\alpha < 2$. For example, a typical segment of speech can be modeled as alpha-stable with $\alpha = 1.6$. Several algorithms based on fractional lower-order statistics has been developed for TDE [4, 5, 6, 7]. These papers have demonstrated, using simulated data, that these algorithms have greater robustness than those using second-order statistics in several cases. This paper will continue that work by showing greater robustness in less controlled situations that resemble real conditions.

First the nature of alpha-stable distributions will be discussed along with background theory. The experimental procedure will be explained. Finally the results will be shown and discussed.

2. REVIEW OF ALPHA-STABLE DISTRIBUTIONS IN TDE

Alpha-stable distributions model random variables that are considered very impulsive. In other words there are spikes in the sequence. The impulsive nature of alpha-stable distributions can be seen in fig. 1.

A symmetrical alpha-stable distribution has the follow-

0-7803-8104-1/03/$17.00 ©2003 IEEE

Authorized licensed use limited to: University of Southern California. Downloaded on May 7, 2009 at 16:51 from IEEE Xplore. Restrictions apply.

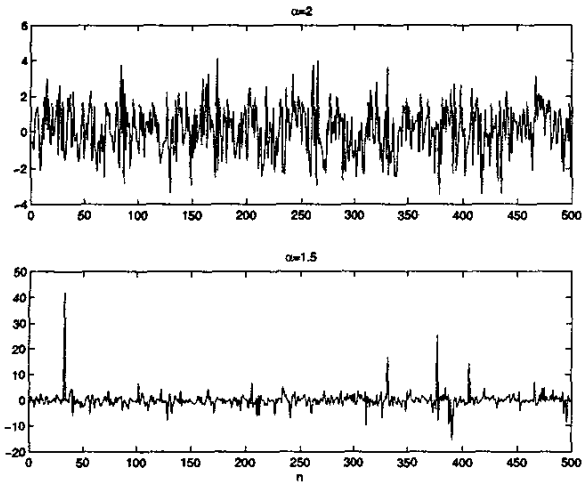

Fig. 1. Shown above are two examples of alpha-stable sequences. The top example is alpha-stable with $\alpha = 2$. It is also a Gaussian distribution. The bottom graph shows a alpha-stable sequence with $\alpha = 1.5$.

ing characteristic function.

$$\phi(t) = exp(j\lambda t - \gamma|t|^{\alpha}) \tag{1}$$

where $0 < \alpha \leq 2$ is the characteristic exponent, λ is the location parameter (the mean if $\alpha > 1$ and the median otherwise), and γ is the dispersion parameter. When $\alpha = 2$ the distribution is Guassian with $\gamma = \sigma^2/2$ and when $\alpha = 1$ the distribution is Cauchy. For most other values of α there is no closed-form solution for the probability density function.

Traditional TDE, like many array signal processing methods such as generalized cross-correlation and MUSIC, uses a second-order moment to find the time delays. However, one of the properties of alpha-stable distributions is that the n^{th}-order moment is undefined for $n > \alpha$. This potentially leads to increased errors in TDE when second order statistics are used in an impulsive noise environment. Alternative measures exist such as the fractional order covariation defined as:

$$A_{xy} = E\left[\frac{xy^*}{|x|^{1-p}|y|^{1-p}}\right] \tag{2}$$

where $0 < p < \alpha/2$. Some insight can be gained by examining the role of the parameter, p. If $p = 1$, (2) can be reduced to

$$A_{xy} = E\left[\frac{xy^*}{|x|^0|y|^0}\right] = E[xy^*] \tag{3}$$

which is the second order correlation. As p gets smaller, those values with high energy become more compressed. The compression lessens the effect of impulses when calculating time delays.

The following model is assumed for time-delay estimation.

$$x_j(t) = \sum_{i=1}^{N_s} s_i(t - \tau'_{ij}) + n_j(t) \tag{4}$$

where $x_j(t)$ is the signal received at the j^{th} microphone, $s_i(t - \tau'_{ij})$ is the i^{th} source delayed by the time delay of propagation, and $n_j(t)$ is the noise. While there are N_s sources, it is assumed that only the strongest source is the desired signal and all other sources are interfering noise sources. All signal and noise sources can be modeled as alpha-stable distributions.

Typically the operations are performed in the frequency domain, by taking the FFT of the signal $x_j(t)$. Using the fractional lower order correlation results in the following equation.

$$A_{X_iX_j}(k) = E\left[\frac{X_i(k)X_j(k)^*}{|X_i(k)|^{1-p}|X_j(k)|^{1-p}}\right] \tag{5}$$

If there is only one significant source and the noise is small

$$A_{X_iX_j}(k) \cong R(k)e^{-\sqrt{-1}\omega_k\tau_{ij}} \tag{6}$$

where $R(k)$ is the magnitude of the expectation.

One method of weighting the frequency results using cross-correlation, is to divide by the magnitude. Afterwards the inverse FFT is computed and the correct time delay is found by searching for the maximum value. This equally weights the angle found at each frequency and is known as the phase transform (PHAT) algorithm. By extension, the algorithm reviewed in this paper is known as fractional lower order statistics phase transform (FLOS-PHAT).

The basic algorithm for FLOS-PHAT can be seen in fig. 2. The fractional order covariation is computed in the frequency domain by appropriately averaging the instantaneous estimates of the covariation. In this case an IIR filter structure is used with a "memory" factor.

$$A(n,k) = (1 - \lambda)A(n-1,k) + \lambda\hat{A}(n,k) \tag{7}$$

where $\hat{A}(n,k)$ is the instantaneous estimate of the fractional lower order correlation at time, n, and frequency, k, and λ is the memory factor. Then the PHAT weighting is applied. Finally, the IFFT is computed and the time delay is determined by finding the lag of the maximum value.

3. EXPERIMENTAL SETUP

The theoretical justification for using fractional lower order statistics (FLOS-PHAT) as opposed to PHAT is fairly clear, but there have been no experiments in a real, reverberant environment to demonstrate the advantages of employing lower order statistics. This paper presents an extensive set of experiments conducted in two significantly

Authorized licensed use limited to: University of Southern California. Downloaded on May 7, 2009 at 16:51 from IEEE Xplore. Restrictions apply.

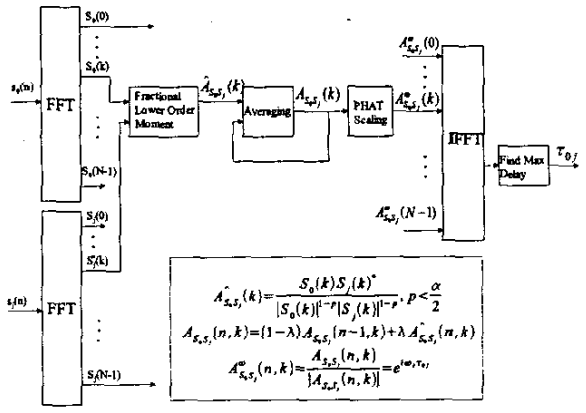

Fig. 2. A block diagram of the proposed FLOS-PHAT algorithm.

different acoustical environments to evaluate and compare FLOS-PHAT with alternative second order based methods.

In the first set of experiments, several different types of real-life noises are employed under two different scenarios. In both scenarios, loudspeakers were used to produce the sounds in order to know the precise location of the sources. A Genelec 1037B loudspeaker was used for all of the desired sources and interfering sources. In order to produce diffuse noise, a pair of Genelec 1032A loudspeakers were wired out of phase to create a dipole loudspeaker. Two of these dipole loudspeakers were used in the experimentation. The room measures 7.6 m by 6.1 m by 3.0 m and has been acoustically treated to remove strong reflections. It has a reverberation time of $T_{60} = 400ms$ at 500 Hz. The sound was measured using a circular array of eight microphones, with diameter of 25 cm, placed in the center of the room.

The first scenario investigates time delay estimation of a source in the presence of interference from different directions. The desired sound source was active throughout the entire experiment. Four different loudspeakers in four different locations were used for the interference. Each interference was a short segment of sound commonly heard in real life. The different sounds included cell phone rings, coughing, and typing. These interferers were measured to have $\alpha = 0.7$ to $\alpha = 1.6$. These interfering signals were played over alternating loudspeakers. The interference was scaled to various signal to interference ratios (SIR) over the length of the interference segment. These values ranged from 18 dB to -12 dB by intervals of 6 dB.

The second scenario investigated the performance of time delay estimation of a source in the presence of diffuse noise, which has no discernible direction of arrival to the array. Examples of the sounds used for the diffuse noise include speech, music, typing, and a dog barking. The noise sources had an alpha parameter that varied from $\alpha = 0.6$ to $\alpha = 1.7$. The desired sound sources included speech and music. In this case the two dipole speakers were used for the diffuse noise, which was active for the whole length of the experiment. The diffuse noise was scaled so that the SNR would be between 24 dB and -12 dB, with spacing of 6 dB.

In order to calculate the time delay, an FFT of length, $L = 1024$ was used. Several different values for the memory factor from 0.001 to 0.05 were used for averaging. The value of p for the fractional correlation was set to 0.2 for all the experiments. It has been determined that this is a good value that works for a wide range of scenarios [5]. This also makes sense, because the measured α has a minimum value of 0.6, which means $p < 0.3$.

The second set of experiments were conducted in a completely real setting, with people conversing at several predetermined locations. Since the locations of the speakers are known with reasonable accuracy, the time delays can be found using geometry and the speed of sound. The room is not acoustically treated in any way. Because of the nature of the walls, which are constructed using drywall, there are many strong reverberations and slap echoes present in this room. An eight-microphone array was used in this experiment, with the microphones spread throughout the room. Since this array had a larger aperture than the first experiment, a longer frame size was used, with $L = 2048$. The memory factor was kept at a constant $\lambda = .005$ and a value of $p = 0.2$ was used for the fractional order correlation just as in the first set of experiments. Using both PHAT and FLOS-PHAT, the delays were estimated in order to verify the robustness of FLOS-PHAT.

4. RESULTS

In order to compare the performance of the FLOS-PHAT algorithm with PHAT a metric should be defined. In this paper the mean square error of the time delay is used.

$$e_{TDE} = \frac{1}{K}\sum_{k=1}^{K}(\tau_k - \hat{\tau}_k)^2 \tag{8}$$

where K is the total number of estimated time delays, $\hat{\tau}_k$ is the estimate and τ_k is the true time delay. In addition to the above, another metric, based on the percent of correct estimates, was used. A time delay estimate is considered correct if $|\tau_k - \hat{\tau}_k| < th$.

The results for the interference scenario is shown in fig. 3 and 4. The error energy for FLOS-PHAT is less than that for PHAT, especially in the more adverse situations with lower SIR. In addition the percent of time delays estimated correctly show that FLOS-PHAT performs significantly better than PHAT for low SIR. In fig. 5 and 6, the results for several memory values are presented. The tables show the

Authorized licensed use limited to: University of Southern California. Downloaded on May 7, 2009 at 16:51 from IEEE Xplore. Restrictions apply.

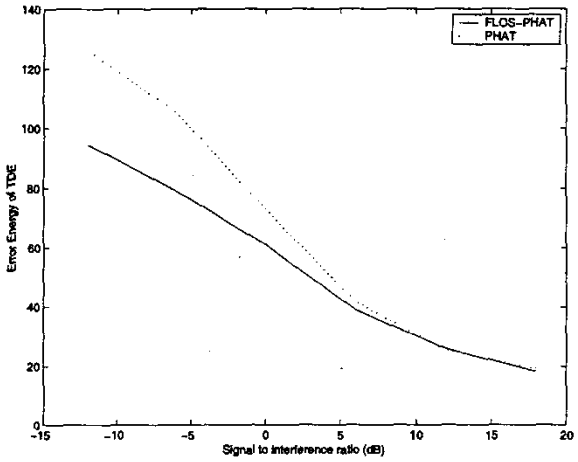

Fig. 3. A comparison of the mean square error of the PHAT algorithm vs. FLOS-PHAT

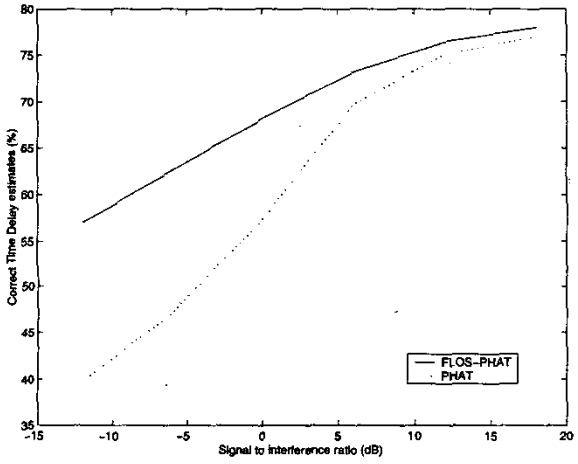

Fig. 4. A comparison of the percent of correct time delay estimates found with the PHAT algorithm vs. FLOS-PHAT. The threshold used in this graph was 2 samples.

	Percent Improvement for several values of λ				
SIR	0.001	0.005	0.01	0.02	0.05
18 dB	4.88	4.66	-18.18	-68.25	-81.43
12 dB	25.25	-1.97	-37.45	-67.55	-25.98
6 dB	26.83	6.68	-14.08	-2.62	-2.81
0 dB	23.98	16.32	19.2	20.92	11.38
-6 dB	24.37	24.71	32.30	32.54	18.75
-12 dB	21.68	25.06	35.41	35.38	22.91

Fig. 5. Percent improvement of MSE of FLOS-PHAT over PHAT using several values of λ.

	Percent Improvement for several values of λ				
SIR	0.001	0.005	0.01	0.02	0.05
18 dB	1.19	1.3	-0.26	-4.31	-2.35
12 dB	2.72	1.73	-1.44	-5.86	-1.62
6 dB	7.72	5.03	0.98	-1.29	0.79
0 dB	22.73	18.85	12.92	8.59	6.71
-6 dB	34.00	32.91	26.63	21.21	13.71
-12 dB	40.73	43.22	36.78	30.07	19.95

Fig. 6. Percent improvement of correct estimates of FLOS-PHAT over PHAT using several values of λ.

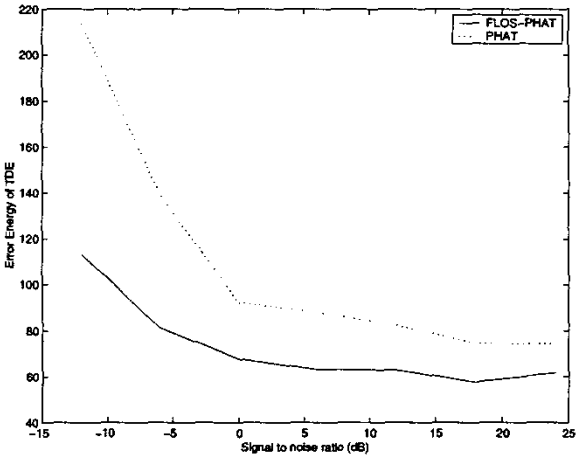

Fig. 7. TDE error results for diffuse scenario.

percent improvement of the MSE of the time delay estimates and the number of correct estimates. In adverse conditions FLOS-PHAT consistently outperforms PHAT for all values of λ. As conditions improve it appears that PHAT outperforms FLOS-PHAT for larger memory values.

The MSE TDE results for the diffuse scenario seems to show a more consistent advantage to FLOS-PHAT. However the percent of correct time delay estimates gives a somewhat conflicting story. This time there is a slight improvement in the percent of correct estimates for high SNR and basically no difference for lower SNR.

Since there is no way to know the energy of the signal or noise in the experiments using real speakers, all of the results were combined. FLOS-PHAT had an MSE of 2363, whereas PHAT had an MSE of 8719. So FLOS-PHAT performed almost four times better than PHAT. The numbers are much larger for this experiment than they were for those previous. This is due to the fact that the array is much larger. The percent of correct estimates is not used in this experiment due to the large uncertainty in what the correct time delay should be. This uncertainty doesn't effect the MSE as much.

Authorized licensed use limited to: University of Southern California. Downloaded on May 7, 2009 at 16:51 from IEEE Xplore. Restrictions apply.

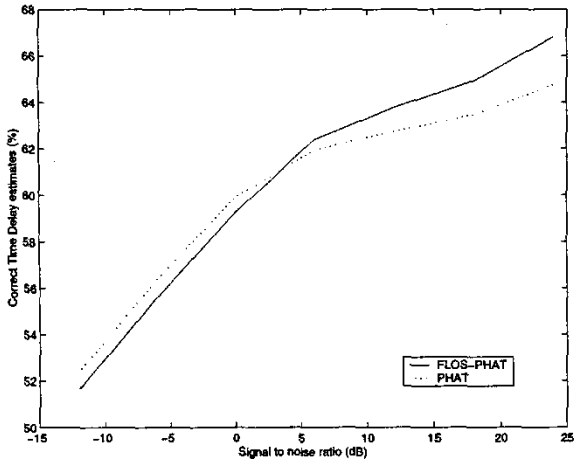

Fig. 8. Percent of correct time delay estimates for diffuse scenario.

5. CONCLUSION

Many signals encountered in acoustical TDE are impulsive and can be modeled as alpha-stable with $\alpha < 2$. Some of the properties of alpha-stable distributions have been reviewed, along with an algorithm, based on fractional order statistics, for finding TDE. The theoretical benefit of this algorithm is that the fractional order moment is defined for alpha stable random variables provided that $p < \alpha/2$. It has been shown that FLOS-PHAT does indeed perform better than PHAT, not only theoretically, but in real situations. This was especially evident in the interference scenario and the experiments in the real environment.

6. ACKNOWLEDGMENTS

This research was funded in part by by the Integrated Media Systems Center, a National Science Foundation Engineering Research Center, Cooperative Agreement No. EEC-9529152. The project was also sponsored in part by the U.S. Army. The content of the information does not necessarily reflect the position or the policy of the Government of the United States of America, and no official endorsement should be inferred.

7. REFERENCES

[1] Yiteng Huang, Jacob Benesty, and Gary W. Elko, "Passive acoustic source localization for video camera steering," in *Proceedings of ICASSP*, 2000.

[2] M. Omologo and P. Svaizer, "Use of the crosspower-spectrum phase in acoustic event location," *IEEE Transactions of Signal and Audio Processing*, vol. 5, no. 3, May 1997.

[3] Joe C. Chen, Ralph E. Hudson, and Kung Yao, "Maximum-likelihood source localization and unknown sensor location estimation for wideband signals in the near-field," *IEEE Transactions of Signal Processing*, vol. 50, no. 8, August 2002.

[4] Panayiotis G. Georgiou, Chris Kyriakakis, and Panagiotis Tsakalides, "Robust time delay estimation for sound source localization in noisy environments," in *Proceedings of WASPAA*, 1997.

[5] Panayiotis G. Georgiou, Panagiotis Tsakalides, and Chris Kyriakakis, "Alpha-stable modeling of noise and robust time-delay estimation in the presence of impulsive noise," *IEEE Transactions on Multimedia*, vol. 1, no. 3, September 1999.

[6] Xinyu Ma and Chrysostomos L. Nikias, "Joint estimation of time delay and frequency delay in impulsive noise using fractional lower order statistics," *IEEE Transactions of Signal Processing*, vol. 44, no. 11, November 1996.

[7] Panagiotis Tsakalides and Chrysostomos L. Nikias, "The robust covariation-based music (roc-music) algorithm for bearing estimation in impulsive noise environments," *IEEE Transactions of Signal Processing*, vol. 44, no. 7, July 1996.

Authorized licensed use limited to: University of Southern California. Downloaded on May 7, 2009 at 16:51 from IEEE Xplore. Restrictions apply.