A NEW MODEL OF SOURCE DEPENDENT NOISE FOR ROBUST ARRAY SIGNAL PROCESSING

Panayiotis G. Georgiou and Chris Kyriakakis

Immersive Audio Laboratory – Integrated Media Systems Center
University of Southern California
Los Angeles, CA 90089-2564
georgiou@sipi.usc.edu ckyriak@imsc.usc.edu

ABSTRACT

In this work we investigate an alternative model for signals encountered in acoustic environments to the traditional Gaussian process. The sound signals in this case are assumed to be sub-Gaussian of impulsive nature. The noise encountered in these environments predominantly stems from reverberation or multipath effects, which makes it significantly dependent on the source. Hence, the noise is also modeled as jointly sub-Gaussian.

The Lévy alpha-stable distribution, of characteristic exponent 0.5 and index of symmetry 1, is used together with a multivariate Gaussian density to derive the sub-Gaussian process. Based on this density, the stochastic *Maximum Likelihood* (ML) estimator is formulated. A separable solution of the estimator is given.

Subsequently, simulations demonstrating the performance gains relative to the Gaussian-based ML estimator are provided, as well as a comparison of the two methods on localization of real sounds gathered with a 20- and 41-microphone arrays.

1. INTRODUCTION

In many array signal processing applications one encounters an environment where sources and noise are dependent. An example of such a case is found in audio environments where microphone arrays often suffer from severe reverberation, rather than other independent noise or interference sources.

Undoubtedly, the Gaussian density has traditionally been the most widely accepted distribution, especially in microphone array applications, and used, as a rule, as a realistic model for various kinds of noise. The introduction of a more general noise structure, such as spatially correlated Gaussian noise or non-uniform spatially white Gaussian noise has also been of investigated [1, and ref. therein].

In recent years however, it has been shown in numerous disciplines that the class of α-stable distributions, which are a generalization of the Gaussian distribution and can be of a more impulsive nature, are often a more suitable model for a wider range of phenomena [2]. The Gaussian is the least impulsive α-stable distribution, while other widely known distributions of the α-stable class are the Cauchy and the Lévy. Further information on α-stable distributions can be found in [3, 4, 5].

This research was funded in part 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, and no official endorsement should be inferred.

These authors have presented in previous work [6] the appropriateness of the α-stable distributions for modeling noises encountered in audio environments and presented a computationally simple time delay estimation method for localization of speech sources. Reverberation constitutes an additional factor in acoustical environments that significantly influences the performance of localization methods. The noise introduced by reverberation effects is far from being independent white noise, but is highly dependent on the desired source. Such dependence is well modeled by the class of sub-Gaussian distributions, a special case of α-stable processes.

In this paper we express the density of a sub-Gaussian distribution, the separable solution of its Maximum Likelihood ML estimator and we provide results to support the superiority of the new model.

2. SUB-GAUSSIAN RANDOM VARIABLES

Sub-Gaussian distributions are a special case of α-stable random processes [7]. A sub-Gaussian random vector $\boldsymbol{X}$ can be defined as a random vector with a characteristic function of the form

$$\varphi(\boldsymbol{u}) = \exp\left(-\frac{1}{2}\left[\boldsymbol{u}^T \mathrm{R}\,\boldsymbol{u}\right]^{\alpha/2}\right) \tag{1}$$

where R is a positive-definite matrix, and the characteristic exponent satisfies $1 < \alpha \leq 2$. Sub-Gaussian processes are variance mixtures of Gaussian processes.

If $\boldsymbol{X}(t)$ is sub-Gaussian with parameter α and S is a positive stable process with characteristic exponent $\alpha/2$ (*i.e., S* is $\frac{\alpha}{2}$-stable random variable completely skewed to the right) and $\boldsymbol{Y}(t)$ is a multivariate Gaussian process independent of S, then:

$$\boldsymbol{X}(t) = S^{1/2}\boldsymbol{Y}(t) \tag{2}$$

Clearly from the above, irrespective of the correlation structure of $\boldsymbol{Y}(t)$, the components of $\boldsymbol{X}(t)$ can not be independent.

3. SIGNAL MODEL FOR ML ESTIMATION

In order to account for impulsiveness in signals, Tsakalides *et al* [8] have used the Cauchy distribution as a signal model. An alternative model that will allow for the high dependence of the noise on the sources is the sub-Gaussian process of equal impulsiveness. We use a distribution of impulsiveness $\alpha = 0.5$, which is completely skewed to the positive axis together with a multivariate Gaussian density. The Lévy distribution (also referred to as a Paretto type 5 distribution with an index of symmetry $\beta = 1$ and characteristic exponent $\alpha = 0.5$) satisfies these requirements and

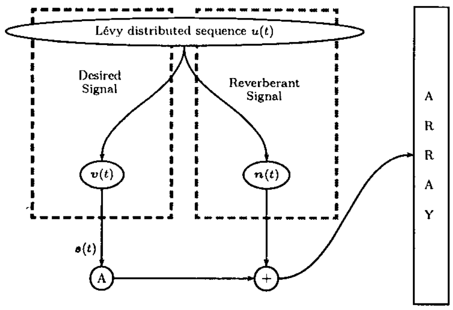

Fig. 1. A multivariate Gaussian signal, corrupted by multiplicative Lévy noise, is transformed through a set of delays to the receiving end of the array. Similarly, the additive noise is generated by the same Lévy sequence. This may be a good model for a reverberation noise, which is highly dependent on the signal of interest.

has a closed form expression. Using a multivariate Gaussian density $f(\boldsymbol{v})$ and a univariate Lévy density $f(u)$:

$$f(u) = \begin{cases} \frac{u^{-\frac{3}{2}}\,\mathrm{e}^{-\frac{1}{4u}}}{2\sqrt{\pi}} & \text{if } u > 0 \\ 0 & \text{if } u < 0 \end{cases} \tag{3}$$

We can derive the sub-Gaussian process of characteristic exponent $\alpha = 1$, which is given by:

$$f(\boldsymbol{s}) = \frac{1}{2\sqrt{\pi}\,\pi^{\kappa}\,|\Sigma_{\mathrm{v}}|} \cdot \left[{}^{1}\!/\!{}_{4} + \boldsymbol{s}^{\dagger}(t)\Sigma_{\mathrm{v}}{}^{-1}\boldsymbol{s}(t)\right]^{-1} \tag{4}$$

Derivation details are omitted due to space constraints but can be found in [9] and [10].

4. MAXIMUM LIKELIHOOD ESTIMATION

The transmitted signals in this case are assumed to be stochastic, and as such, the parameters of interest will be their statistics and *Directions-of-Arrival* (DOA's). In this work we investigate exclusively the *Stochastic ML* estimation where the signals are assumed to be random rather than deterministic.

Fig. 1 gives a top level description of the problem. We assume a scenario under which there are κ sources received by an array of ρ sensors. The transfer function each signal undergoes while traveling to the array can be modeled as an attenuation and a delay. The attenuation will be considered the same at all sensors under the assumption that the sources are in the far-field of the array. These transfer functions are

$$a_{r,k} = \mathrm{e}^{-\mathrm{i}\omega\tau_{r,k}}, \quad r = 1\ldots\rho \quad \text{and} \quad k = 1\ldots\kappa \tag{5}$$

where $\tau_{r,k}$ is the delay of the signal (of source k) received at sensor r relative to the first sensor. With the same assumption, $\tau_{r,k} = \tau_r(\theta_k)$, and it is also clear that if we are considering a linear array, $\tau_{r,k} = (r-1)\cdot\tau_1(\theta_k)$. We denote the vector of the medium transformations for source k by $\boldsymbol{a}_k = [a_{1,k}\ a_{2,k}\ \ldots\ a_{\rho,k}\ ]^T$, and A the matrix comprised of these vectors. Therefore, the array's input vector is

$$\boldsymbol{x}(f) = \mathrm{A}\cdot\boldsymbol{s}(f) + \boldsymbol{n}(f) \tag{6}$$

Assuming that the signal is of the form given by (4) and described on Fig. 1 at the transmitted end, then at the receiver, $\boldsymbol{x} = [x_1 \ldots x_\rho]^{\mathrm{T}}$ is of the same form:

$$\boldsymbol{x}_r(t) = y(t)^{1/2}\cdot\boldsymbol{z}_r(t) \tag{7}$$

where it is easy to show that the received signal is sub-Gaussian with underlying Gaussian statistics

$$\mathrm{R} = \mathrm{A}\,\Sigma_v\,\mathrm{A}^{\dagger} + \sigma_n^2\mathrm{I}_\rho \tag{8}$$

Therefore, the $\log_{\mathrm{e}}$ maximum likelihood estimator is

$$[\hat{\Sigma},\hat{\boldsymbol{\theta}}] = \arg\min_{\hat{\Sigma},\hat{\boldsymbol{\theta}}} \sum_{f=f_1}^{f_M} \Big\{ \log_{\mathrm{e}}|\mathrm{R}| + \log_{\mathrm{e}}\left[\boldsymbol{x}^{\dagger}(f)\mathrm{R}^{-1}\boldsymbol{x}(f) + {}^{1}\!/\!{}_{4}\right] \Big\} \tag{9}$$

5. ML – A SEPARABLE SOLUTION

The separable solution of the ML function can be obtained in a similar fashion to the separable solution of the Gaussian ML [11].

5.1. Estimating the Statistics

We proceed in this case to reach an alternative minimization function to reduce the search space. In order to do so, we minimize the ML function with respect to the statistics, assuming known DOA's:

$$\Sigma_{\mathrm{ML}} = \sum_{t=t_1}^{t_M} \left[\mathrm{A}^{-} \left(\frac{\boldsymbol{x}\boldsymbol{x}^{\dagger}}{\mathrm{Tr}\left[\mathrm{R}^{-1}\boldsymbol{x}\boldsymbol{x}^{\dagger}\right] + {}^{1}\!/\!{}_{4}} - \sigma_n^2 \right) \mathrm{A}^{-\dagger} \right] \tag{10}$$

where $\mathrm{A}^{-} = \left(\mathrm{A}^{\dagger}\mathrm{A}\right)^{-1}\mathrm{A}^{\dagger}$ and R^{-1} as defined in (11). The solution of (10) can be easily found using the numerical iteration method and:

$$\mathrm{R}^{-1} = \frac{1}{\sigma_n^2}\left\{\mathrm{I} - \mathrm{A}\left(\Sigma\,\mathrm{A}^{\dagger}\,\mathrm{A} + \sigma_n^2\mathrm{I}\right)^{-1}\Sigma\,\mathrm{A}^{\dagger}\right\} \tag{11}$$

An estimate of R can be used as an initial guess and can be found from the data using a covariation measure. However, simulations have shown this step to be unnecessary since the algorithm converges very fast from an initialization of $\mathrm{R} = \mathrm{I}$.

5.2. DOA Estimation

The above assumes that the DOA vector is known, and we approach here the localization part of the problem. Using pseudo-ML, we can express the modified ML function irrespective of the statistics R as

$$\hat{\boldsymbol{\theta}} = \arg\min_{\hat{\boldsymbol{\theta}}} \sum_{f=f_1}^{f_M} \left\{\log_{\mathrm{e}}\left[\boldsymbol{x}^{\dagger}(f)\mathrm{R}^{-1}\boldsymbol{x}(f) + {}^{1}\!/\!{}_{4}\right]\right\} \tag{12}$$

where R can be substituted with any valid statistics (identity matrix, for instance).

A search algorithm or even a simple alternating line search can be used to find the solution of the above equation.

6. SIMULATIONS - DOA ESTIMATION

Several sets of simulations have been performed to test the validity of the algorithm. Throughout the DOA estimation tests, $\Sigma = I$ is assumed to hold, although the test matrix had a random correlation structure, but always with diagonal elements of dispersion equal to 1. The impulsiveness was kept constant ($\alpha = 1$ for cases 1 & 4, and $\alpha = 2$ for 2 & 3 as described below).

The Generalized Signal-to-Noise Ratio is defined as:

$$\text{GSNR} = 10 \log_{10} \left(\frac{\gamma_s}{\gamma_n} \right) = -10 \log_{10} (\gamma_n) \qquad (13)$$

Fig. 2 shows the mean squared error and the probability of localization for the conditions described in Fig. 1. Four cases were simulated, and in each one the noise follows the same assumptions as the signal:

1. Exactly as per the derivation assumptions (Fig. 2a): received signal is sub-Gaussian, created from a Multivariate Gaussian and a univariate Lévy. Received signal impulsiveness is $\alpha = 1$ (impulsiveness – dependence)

2. The signal is a Multivariate Gaussian (Fig. 2b), and is created from a Multivariate Gaussian ($\boldsymbol{v}$) and a univariate Gaussian (w). Received signal impulsiveness is $\alpha = 2$ (no impulsiveness – dependence)

3. The signal is a Multivariate Gaussian (Fig. 2c) and it undergoes *no* energy fluctuation ($w = 1$, $\boldsymbol{v} = \boldsymbol{s}$). This conforms to the assumptions of the well known Gaussian based ML. Clearly, the received signal impulsiveness is $\alpha = 2$ (no impulsiveness – no dependence)

4. Finally, the received signal is sub-Gaussian(Fig. 2d), created from a Multivariate Gaussian ($\boldsymbol{v}$) and a *Multivariate* Lévy ($\boldsymbol{w}$). In this case, the signals can be viewed as simply Cauchy. Received signal impulsiveness is $\alpha = 1$ (impulsiveness – no dependence)

As expected, the sub-Gaussian ML method performs better when the derivation assumptions hold (Fig. 2a). Likewise, when the signal is a multivariate Gaussian, the Gaussian ML algorithm performs better (Fig. 2c).

In the cases that neither assumption holds however, we can see how more robust the sub-Gaussian ML method is. When the signal follows (2) (Fig. 2b), the sub-Gaussian ML performs slightly better than the Gaussian ML. The real benefit of the proposed ML method can however be observed when the signals are impulsive due to random multiplicative noise, independent from one source to the next (Fig. 2d).

7. COMPARISONS WITH REAL DATA

In order to test the localization algorithm with some real data, we constructed two synthetic microphone arrays: using the 10.2 channel system and ProTools we played back several (dry) signals (Trumpet, Cello, a female voice in English, and a female voice in Danish). These audio channels were played together in various combinations through the loudspeakers at 48kHz, and 2 microphones were shifted to form a linear array. The synchronized playback–recording feature of ProTools, confirmed by the addition of chirp synchronization signals at the start of the recording, ensured that the array was accurately created.

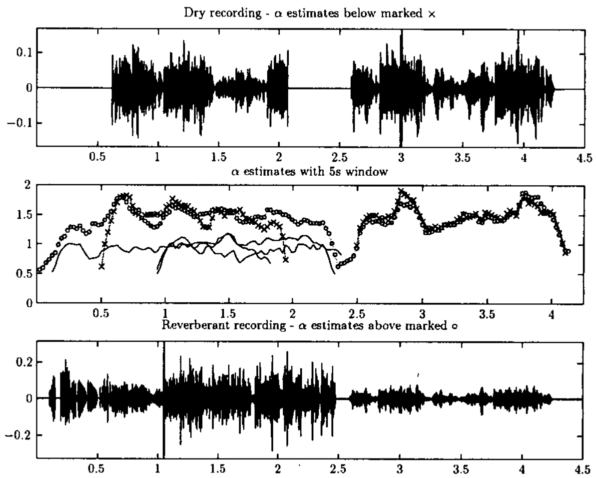

Fig. 3. Dry signal on top (Cello) and reverberant combination at the bottom. Middle figure shows estimates of α values, specifically the × of the top signal and the o of the bottom signal while the solid lines correspond to the other original/dry signals.

The ML function for the following cases was evaluated over all frequencies by re-calculating the transformation matrix A for all possible $(\boldsymbol{\theta}, f)$ combinations, which is a computationally expensive process. For the localization part, a *Non-linear FFT* (NFFT) [12, 13] was used in order to keep the resulting frequency domain signals. Specifically, we employed the method described by Mitra *et al* in [12] with a first order all-pass filter and a 30ms window (1440 samples).

7.1. 20-Microphone Array

In the 20-microphone array case, the aperture was 38cm and the intersensor spacing was 2cm, while 4 (originally dry) signals (Trumpet, Cello, a female voice in English, and a female voice in Danish) and an artificial echo of the cello were used. These 5 channels were played together in various combinations, although the results shown here are based on localization of the sources when two signals were active (the Cello and Trumpet at 45° and 90° respectively). This array was not very accurately spaced and the error rate from the part were all 5 channels were active was very large.

In order to demonstrate the impulsiveness of the signals, Fig. 3 displays one of the dry signals on top, and its α parameter estimates in the second subplot with the × line. The α estimates of the signal received by one of the array microphones is also marked in the middle subplot with the o line, and the time domain of the same signal is shown in the third subplot below. Additionally represented in the center subplot are the estimates of the α parameter of the other dry recordings that constitute the array input.

Results of localization demonstrate that the sub-Gaussian based ML method performs significantly better than its Gaussian counterpart. Fig. 4 shows 7s of the signal where only the cello and trumpet are being played.[1] Each frame of the segment corresponds to a sliding window of 30ms, and the sources were placed at 45° and at 90°. As can be observed, the sub-Gaussian ML method works significantly

[1] These correspond to 13s to 20s or 6.24×10^5 to 9.6×10^5 samples of Figs. 3.

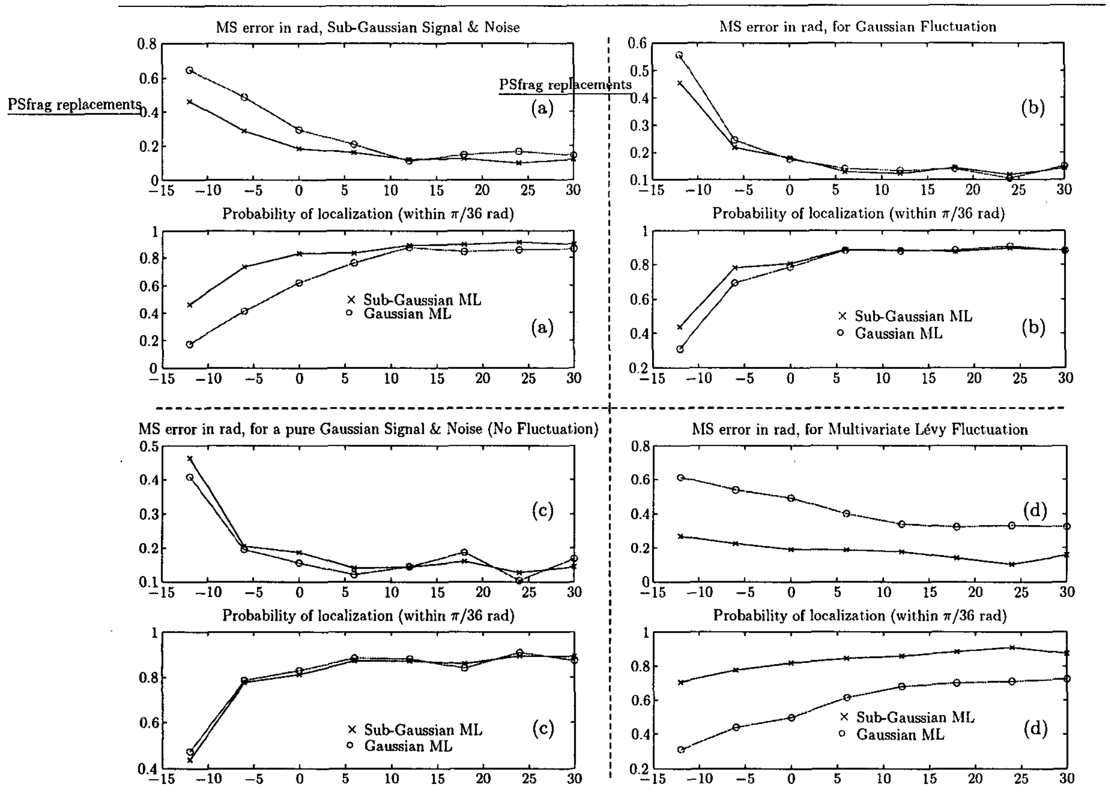

Fig. 2. Simulations demonstrate the obtained benefit in localization by using the Stochastic ML method based on the Lévy Sub-Gaussian processes versus the Gaussian ML method for the conditions described in the text. Robustness of the Sub-Gaussian method is apparent especially in case (d).

	Gaussian	Sub-Gaussian
45° angle RMS error	16	5
90° angle RMS error	22.5	10
Overall RMS error	19.5	8

Table 1. Errors for the Gaussian based ML method are more than double those of the sub-Gaussian based ML.

better. Table 1 shows the RMS error for this localization experiment, and reveals that the performance of the Gaussian based ML is significantly worse than that of the sub-Gaussian based ML.

7.2. 41-Microphone Array

In the 41-microphone array the recording conditions are similar to the previous case. However, the inter-sensor spacing is 1cm, the array is much more accurately spaced than the previous one, and the sources are the two speech signals used in the previous section placed at 48° and 110°. In addition, the arrangement is such that a strong echo is created at 90°. Fig. ?? shows the positions of the sources, the array, and the flat screen,[2] which as we expect causes a strong sound reflection.

	Gaussian	Sub-Gaussian
48° angle RMS error	11.1	9.3
90° angle RMS error	13.1	6.9
110° angle RMS error	17.7	6.6
Overall RMS error	24.6	13.3

Table 2. Errors for the Gaussian based ML method are much higher than those of the sub-Gaussian based ML, but compare better under these conditions of the larger array than in the case of the 20-microphone array.

The performance is again better when using the sub-Gaussian based ML, and especially for the 90° and 110° directions. Nevertheless, the performance difference decreases as the array size grows, a similar conclusion with the performance difference gap narrowing at increasing SNR's in the simulations. The RMS error of localization for the two methods is shown on Table 2.

[2]The screen is made from a synthetic material that is highly reflective.

8. CONCLUSIONS

We have presented in this work a model designed to account for signals that are dependent and impulsive in nature. Such signals are often encountered in many disciplines including audio. The proposed model can be used in several different scenarios: Gaussian sources undergoing the same energy fluctuations, dependent and impulsive sources, or even independent and impulsive sources. Possible applications include localizing multipath signals, which can be highly dependent and impulsive, or sound source localization in reverberant environments using a microphone array.

Based on this model we formulated an array signal processing problem, and gave the separable solution of its ML estimator. Simulations have demonstrated the robustness of the sub-Gaussian based ML versus the Gaussian based one. Additionally, real world measurements were conducted with two large arrays (20 and 41 microphones) in our audio lab, a room with the acoustics of a typical living room. These experiments have also supported the advantages of the new model. The sub-Gaussian based ML exhibits an improvement in localization up to a factor of 3 in the RMS error versus the Gaussian ML.

References

[1] M. Pesavento and A. B. Gershman, "Maximum-likelihood direction-of-arrival estimation in the presence of unknown nonuniform noise," *IEEE Trans. Signal Processing*, vol. 49, no. 7, pp. 1310–1324, July 2001.

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

[3] Stamatis Cambanis, Gennady Samorodnitsky, and Murad S. Taqqu, Eds., *Stable Processes and Related Topics*, Progress in probability. Birkhäuser, Boston, 1991, Selection of papers from the Workshop on Stable Processes and Related Topics held at Cornell University.

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

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

[6] P. G. Georgiou, P. Tsakalides, and C. 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, pp. 291–301, September 1999.

[7] S. Cambanis and G. Miller, "Linear problems in p^{th} order and stable processes," *SIAM Journal on Applied Mathematics*, vol. 41, no. 1, pp. 43–69, August 1981.

[8] 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.

[9] P. G. Georgiou and C. Kyriakakis, "Maximum likelihood parameter estimation under impulsive conditions. a Sub-Gaussian signal approach," Submitted: IEEE Trans. Signal Processing.

[10] Panayiotis Georgiou, *Robust signal processing techniques for source localization and multisource spatial sound rendering for immersive environments*, Ph.D. dissertation, University of Southern California, 2002.

[11] A. G. Jaffer, "Maximum likelihood direction finding of stochastic sources: a separable solution," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing.*, 1988, vol. 5, pp. 2893–2896.

[12] A. Makur and S. K. Mitra, "Warped discrete-fourier transform: theory and applications," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 48, no. 9, pp. 1086–1093, Sep. 2001.

[13] A. Oppenheim and D. Johnson, "Computation of spectra with unequal resolution using the fast fourier transform," in *Proc. IEEE*, 1971, vol. 59, pp. 299–301.

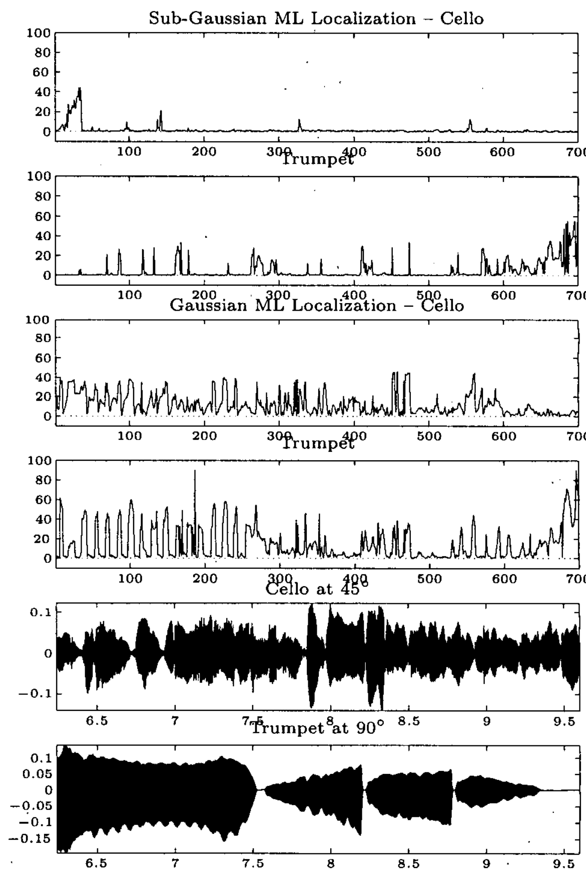

Fig. 4. Subplots 1-4 show DOA error versus time (1/100 s). The sub-Gaussian based ML (top 2 subplots) performs significantly better than the Gaussian ML (3rd and 4th subplots). Source signals were at 45° and at 90° and are plotted at the bottom two graphs. We can see the correlation of the error rising when the amplitude of the trumpet dies off at the end.