Robust Maximum Likelihood Source Localization: The Case for Sub-Gaussian versus Gaussian

Panayiotis G. Georgiou and Chris Kyriakakis

***Abstract*—In this paper, we investigate an alternative to the Gaussian density for modeling signals encountered in audio environments. The observation that sound signals are impulsive in nature, combined with the reverberation effects commonly encountered in audio, motivates the use of the sub-Gaussian density. The new sub-Gaussian statistical model and the separable solution of its maximum likelihood estimator are presented. These are used in an array scenario to demonstrate with both simulations and two different microphone arrays the achievable performance gains. The simulations exhibit the robustness of the sub-Gaussian-based method while the real world experiments reveal a significant performance gain, supporting the claim that the sub-Gaussian model is better suited for sound signals.**

***Index Terms*—Alpha stable, maximum likelihood (ML), microphone arrays, sound source localization, sub-Gaussian.**

I. Introduction

INTEREST is increasing in both the academic community industry in using microphone arrays to extract more accurate representations of the signal, remove reverberation, or improve speech recognition, or in video-camera steering applications. Wang and Chu have presented [1] a commercial product based on the idea of time delay estimation for steering a camera in the speaker's direction. More recently, Yamada *et al.* [2] have visited the problem of joined microphone array acquisition and speech recognition, and have identified as one of the issues concerning the distant speech recognition the "localization and tracking algorithm." Additionally, in order to facilitate localization from time difference of arrival (TDOAs) estimates between sensor pairs several methods are being examined [3, and references therein]. Methods that exploit visual in addition to audio information are also of renewed interest and are reviewed by Strobel *et al.* in [4].

In this paper we also approach the problem of localization using microphone arrays but from a fundamentally different angle. We discount constraints of computational complexity and proceed to give a better model on which less computationally complex algorithms can be based in the future. We revisit the usual assumptions about the sound signal's characteristics and present a novel model for signals encountered in audio environments. Motivated by the observation that noise in a room environment is mostly due to reverberation rather than independent sources, we derive a model of dependent source and noise. In addition, based on the demonstration of the impulsiveness of sound in previous work from these and other authors [5], [6], we decide to use the α-stable class of distributions, and more specifically their sub-Gaussian subset for our model.

Our work will follow a brief introduction of α-stable theory in Section II. In Section III, we will introduce the proposed model, its density, ML estimator and separable solution. The performance of the ML estimator based on the new sub-Gaussian model will be assessed on real data in Section IV.

Manuscript received February 10, 2004; revised March 28, 2005. This work was supported in part by the Integrated Media Systems Center, a National Science Foundation Engineering Research Center, under Cooperative Agreement EEC-9529152, and 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. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Shoji Makino.

The authors are with the Integrated Media Systems Center, University of Southern California, Los Angeles, CA 90089-2564 USA (e-mail: georgiou@sipi.usc.edu).

Digital Object Identifier 10.1109/TSA.2005.860846

II. Background: Alpha-Stable Distributions

The Gaussian distribution has traditionally been the most widely accepted distribution and used, as a rule, as a realistic model for various kinds of noise. In recent years however, there has been a tremendous interest in the class of α-stable distributions, which are a generalization of the Gaussian distribution, but are able to model a wider range of phenomena and can be of a more impulsive nature. In fact, 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.

In 1991, Cambanis, Samorodnitsky, and Taqqu [7] gave a review of α-stable processes from a statistical point of view. Several other statisticians have provided valuable work in the theory of α-stable distributions. Cambanis, Weron, Zolotarev, Miller *et al.* have done extensive work on the properties of α-stable distributions, in the field of linear filtering problems, and in the domain of spectral representation. A textbook of comprehensive coverage of the α-stable theory was written by Samorodnitsky and Taqqu in 1994 [8]. In 1993, Nikias and Shao [9] 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 [10].

Alpha-stable distributions have been used to model diverse phenomena such as radar clutter [11], random fluctuations of gravitational fields, economic market indices, data file sizes on the Web, network traffic [12], sound signals [5], [6], etc.

A. Theory

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* and second it satisfies the *Generalized Central Limit Theorem* [8], [10], [13].

Since there is no closed form expression for the probability density function of α-stable distributions, they are usually represented by their characteristic function [10] and four parameters.

- α is the *characteristic exponent* satisfying $0 < \alpha \leq 2$. The characteristic exponent controls the thickness (also referred to as heaviness) of the tails of the density function. The tails are heavier, and thus the noise is more impulsive for low values of α, while for a larger α the distribution has a less impulsive behavior.
- 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 $\alpha = 2$ (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

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

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

A closed form expression also exists for the case of the Lévy distribution (also referred to as a Pareto type 5), which has parameters $\beta = 1$ and $\alpha = 0.5$, and therefore it is completely skewed to the positive axis

$$f(x) = \begin{cases} \frac{\gamma}{\sqrt{2\pi}} (x-\lambda)^{-(3/2)} \exp\left\{-\frac{\gamma^2}{2(x-\lambda)}\right\}, & \text{if } x > \lambda \\ 0, & \text{if } x \leq \lambda. \end{cases} \tag{3}$$

The only other closed form expression for a stable distribution is the case obtained by symmetric reflection of the Lévy, i.e., with $\alpha = 0.5$ and $\beta = -1$. The density is given by $f_{\alpha=0.5,\beta=-1}(x) = f_{\text{Levy}}(-x)$.

The properties of the α-stable distributions combined with their impulsive characteristics have encouraged 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 [5].

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

$$\begin{aligned} \alpha < 2, \quad & \mathbf{E}|X_\alpha|^p \text{ not defined } \forall\, p \geq \alpha \\ \alpha < 2, \quad & \mathbf{E}|X_\alpha|^p < \infty \quad \forall\, 0 \leq p < \alpha \\ \text{Gaussian}: \alpha = 2, \quad & \mathbf{E}|X_\alpha|^p < \infty \quad \forall\, p \geq 0. \end{aligned} \tag{4}$$

References [8]–[10], [12] treat the α-stable theory further.

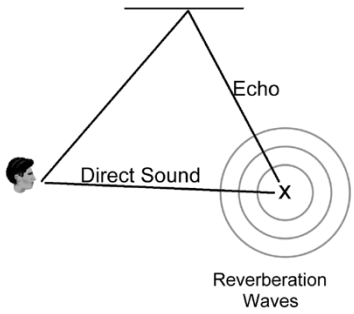

Fig. 1. Several dependent signals: the direct component, the directional echo, and the nondirectional reverberation elements.

B. Sub-Gaussian Random Variables

A Sub-Gaussian random vector $\underline{\boldsymbol{X}}$ can be defined as a random vector with characteristic function of the form

$$\varphi(\underline{\boldsymbol{u}}) = \exp\left(-\frac{1}{2}\left[\underline{\boldsymbol{u}}^T \underline{\underline{\mathbf{R}}} \underline{\boldsymbol{u}}\right]^{\alpha/2}\right) \tag{5}$$

where $\underline{\underline{\mathbf{R}}}$ is a positive-definite matrix, and $0 < \alpha \leq 2$.

Sub-Gaussian processes are variance mixtures of Gaussian processes [14]. Specifically, $\underline{\boldsymbol{X}}(t)$ is sub-Gaussian with parameter α (will be denoted by α-SG $(\underline{\underline{\mathbf{R}}})$) if there exists $S(t)$, a positive $\alpha/2$-stable process with dispersion $\cos(\pi\alpha/4)^2$, and $\underline{\boldsymbol{Y}}(t)$, a multivariate Gaussian process independent of $S(t)$, and

$$\underline{\boldsymbol{X}}(t) = S(t)^{1/2} \underline{\boldsymbol{Y}}(t). \tag{6}$$

A property of sub-Gaussian signals is that irrespective of the correlation structure of $\underline{\boldsymbol{Y}}(t)$, the components of $\underline{\boldsymbol{X}}(t)$ cannot be independent [8].

It is appropriate here to consider the relation of this model to the real world acoustic signals. Consider a sound propagated in a reverberant environment and received by a sensor. The received signal is a scaled copy of the original signal added with a significantly large number of highly dependent to the original signal reverberation components. These reverberation components can be viewed as nondirectional arrivals to the sensor, while the presence of echos can be viewed as directional and dependent arrivals to the sensor as in Fig. 1.

III. New Model

In [5] we demonstrated the impulsive nature of sound signals and provided a modified version of the common PHAT method that takes this nature into consideration. This impulsive behavior is also demonstrated in Fig. 2 with two speech signals, a trumpet and a cello. The four sounds are played from multiple locations in the room and recorded by a microphone array. All recordings were done with a 48 kHz sampling frequency. Subplot 1 shows one of the original dry signals—the cello—while the bottom subplot shows the signal from one of the microphones—containing the reverberant recordings from all four signals. In the middle subplot, the characteristic exponent α estimates using a 5 s sliding window are shown for five signals: dry cello $(\times)$, dry trumpet (line), two dry speech signals (line) and the combined reverberant signal $(\circ)$. The signals are aligned over the three

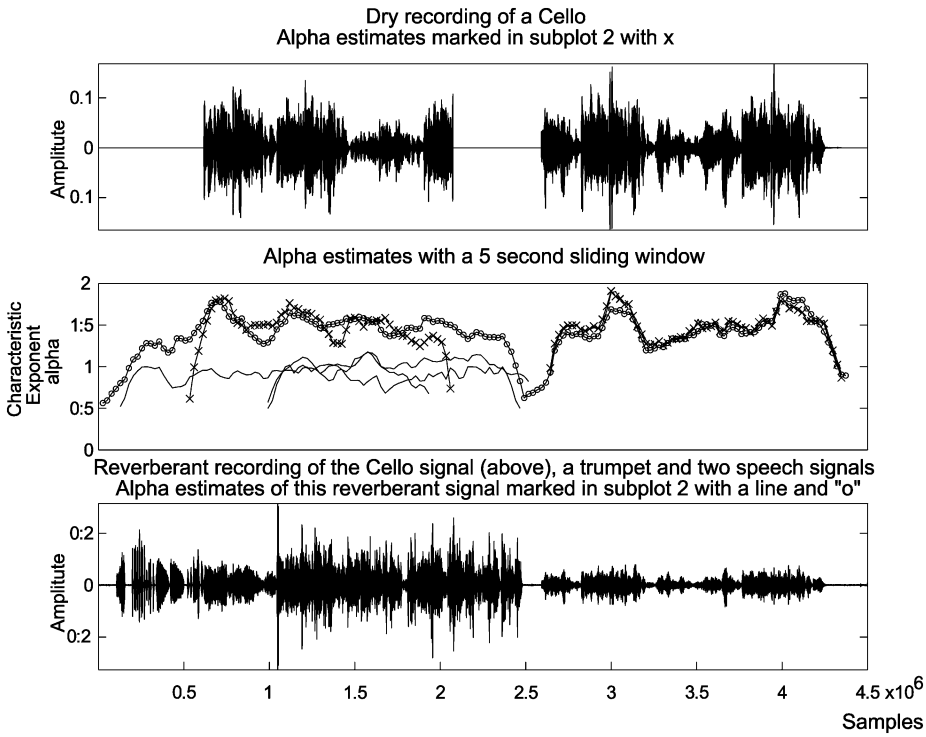

Fig. 2. Dry recordings of a cello, a trumpet and two speech signals are played through loudspeakers and recorded by a microphone array. Top signal is the dry cello and bottom is the reverberant addition of all the signals. Both are sampled with 48 kHz and the horizontal axis represents samples. The *characteristic exponent* α estimates using a 5 s sliding window are shown in the middle plot using $\times$ for the cello, $\circ$ for the reverberant addition, while the other signals' (trumpet and two speech signals) α estimates are marked with a line. It can be observed that the least impulsive signal is the cello. These signals are the ones used further on for the 20-microphone array.

plots, and we can see that the cello lowers the impulsiveness of the resulting reverberant signal.

We continue in this chapter our work on localization by focusing on the development of methods relating to the estimation of the parameters of a system where we assume multiple sources ($k = 1 \dots \kappa$) received by an arbitrary number of sensors ($r = 1 \dots \rho$, where ρ is greater than the number of sources). Additionally, we are aiming to provide a more accurate statistical description of the signals encountered in acoustical environments.

This problem, which we visit from a completely theoretical perspective in [15] and here in the microphone array signal processing framework, can have significant applications in a variety of fields. The scenario of impulsive and multiplicative noise is encountered for instance in communications owing to the presence of local scatterers in the vicinity of the mobile, or due to wavefronts that propagate through random inhomogeneous media. Gershman *et al.* [16] have, for example, presented a method that assumes a random phase perturbation along all source-sensor paths. Their method has led to a non-Gaussian model, and did not result in a ML estimator. Besson *et al.* [17] suggested a similar localization algorithm for a source, which appears as a scatter of sources. Similarly, Stoica *et al.* [18] have presented a Gaussian-based ML method in the presence of multiplicative noise, but constraining the amplitude of the noise to be 1. We expect that the model we propose in this chapter is well suited for such cases, even though experiments will be performed for audio signals only.

The transmitted signals for the development of the localization algorithm are assumed to be stochastic, and as such, the parameters of interest will be their statistics and *Directions-of-Arrival* (DOAs). Despite the wide variety of optimization criteria, the optimal detector is characterized by a single result: the maximum (ML) likelihood ratio test, which was also one of the first methods to be applied in the area of array signal processing [19].

The maximum likelihood technique applied to the source localization problem usually makes two different assumptions for the signal waveforms, resulting in two different ML methods. According to the *Stochastic ML* (SML), the signals are usually modeled as Gaussian random processes motivated by the Central Limit Theorem, and result in closed form mathematical expressions. On the other hand, in the *Deterministic ML* (DML) the signals are considered to be unknown but deterministic. In this case, estimates of the signals as well as the DOAs are desired, while in the former case, the only parameters to be estimated are the statistics and DOAs. In this paper, we deal exclusively with *Stochastic ML* estimation, and we will deviate from the usual Gaussian assumption to work with the alternative impulsive model.

A. Motivation for a Sub-Gaussian Model

The demonstration of the impulsiveness of sound signals motivates our work in improving the signal model. Additionally, one of the most important sources of noise in any acoustical environment is the reverberation (while similar effects such as multipath can be observed in other environments). As we are interested in a more accurate model for acoustical signals, we attempt to model both these effects.

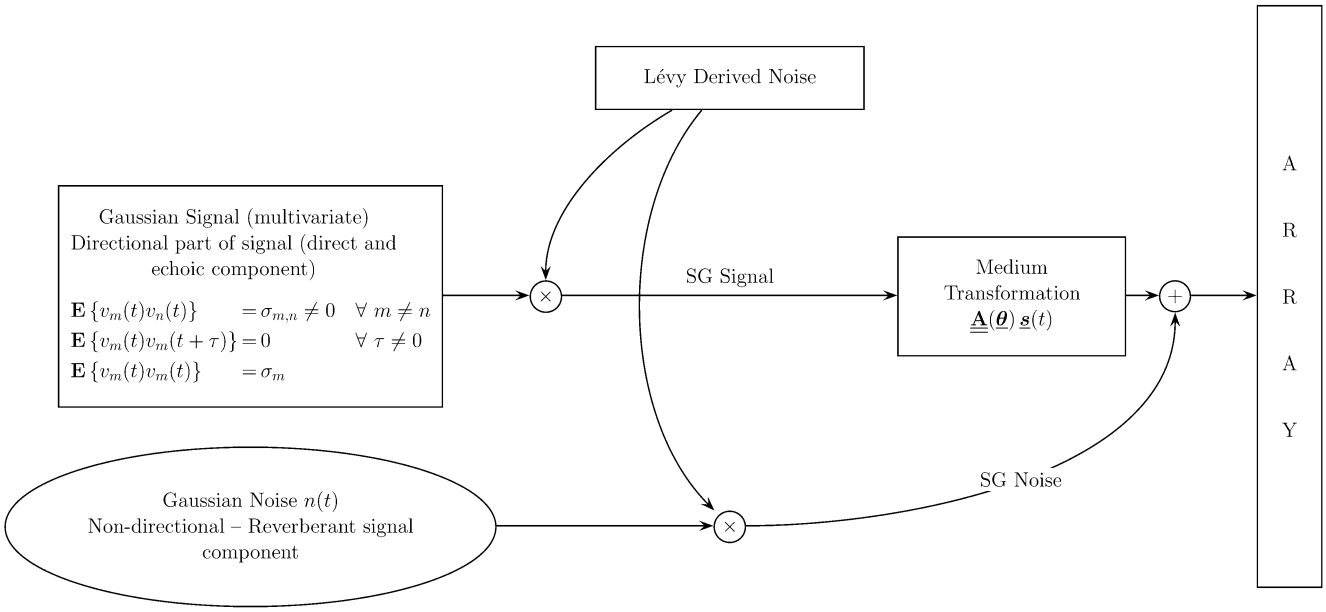

Fig. 3. Overview of the proposed signal model.

The sub-Gaussian processes are attractive in this respect for two main reasons. First, sub-Gaussian processes are impulsive, and hence are able to account for the impulsiveness of the signals. Secondly, the components of a multivariate sub-Gaussian process can not be independent. We expect this process to be a good model for reverberant noise, which is highly related to the signal itself. The noise, which consists mostly of unwanted reverberant signals, can be considered as jointly sub-Gaussian with the signal as would signals produced from the same Lévy process.

We begin with a theoretical analysis of the SML estimator of a Gaussian signal in the presence of Gaussian noise. This analysis is given as a precursor to the derivation of the sub-Gaussian density and the SML estimation of a signal modeled as a sub-Gaussian random process.

B. Framework

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

$$a_{r,k} = e^{-i\omega\tau_{r,k}}, \quad r = 1 \ldots \rho \text{ and } k = 1 \ldots \kappa \tag{7}$$

where $\tau_{r,k}$ is the delay of the signal (of source k) received at sensor r relative to the first sensor.

We assume the sources to be in the far-field and hence, $\tau_{r,k} = \tau_r(\theta_k)$. Also, we denote the vector of the medium transformations for source k by $\underline{\boldsymbol{a}}_k = [a_{1,k} \quad a_{2,k} \ldots a_{\rho,k}]^T$.

Therefore, the array's input vector is

$$\underline{\boldsymbol{x}}(f) = \underline{\underline{\mathbf{A}}} \cdot \underline{\boldsymbol{s}}(f) + \underline{\boldsymbol{n}}(f) \tag{8}$$

where

$$\underline{\underline{\mathbf{A}}} = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,\kappa} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,\kappa} \\ \vdots & \vdots & \ddots & \\ a_{\rho,1} & a_{\rho,2} & & a_{\rho,\kappa} \end{bmatrix} \text{ and } \underline{\boldsymbol{s}}(f) = \begin{bmatrix} s_1(f) \\ s_2(f) \\ \vdots \\ s_\kappa(f) \end{bmatrix}.$$

C. Gaussian Signals

The most commonly used maximum likelihood DOA estimator is the Gaussian ML derived either under the assumptions of a deterministic or a stochastic signal. We present in this section the *Stochastic ML* (SML) DOA estimator for a Gaussian signal in additive white Gaussian noise as background material for the SML DOA estimator to be presented in Section III-D, which is based on sub-Gaussian signals.

Assuming the signals to be jointly stationary Gaussian stochastic processes with covariance matrix $\underline{\underline{\mathbf{\Sigma}}} = \mathbf{E}\left[\underline{\boldsymbol{s}}(t)\underline{\boldsymbol{s}}^\dagger(t)\right] = \mathbf{E}\left[\underline{\boldsymbol{s}}(f)\underline{\boldsymbol{s}}^\dagger(f)\right]$, and the noise to be uncorrelated white noise of variance σ^2, we can express the covariance matrix of the received signal as

$$\begin{aligned} \underline{\underline{\mathbf{R}}} &= \mathbf{E}\left[\underline{\boldsymbol{x}}(f)\underline{\boldsymbol{x}}^\dagger(f)\right] \\ &= \underline{\underline{\mathbf{A}\mathbf{\Sigma}\mathbf{A}}}^\dagger + \sigma^2\underline{\underline{\mathbf{I}}}. \end{aligned} \tag{9}$$

From the assumption that the snapshots are independent and identically distributed, the density function of the complete data set of size M is

$$f(\underline{\underline{\mathbf{X}}}) = \prod_{f=f_1}^{f_M} \frac{1}{\pi^\rho \left|\underline{\underline{\mathbf{R}}}\right|} \exp\left(-\underline{\boldsymbol{x}}^\dagger(f)\underline{\underline{\mathbf{R}}}^{-1}\underline{\boldsymbol{x}}(f)\right) \tag{10}$$

where

$$\underline{\underline{\mathbf{X}}} = \underline{\boldsymbol{x}}(f_1), \underline{\boldsymbol{x}}(f_2), \ldots, \underline{\boldsymbol{x}}(f_M). \tag{11}$$

In order to solve the SML problem, we need to estimate $\hat{\sigma}^2$, $\underline{\underline{\hat{\mathbf{\Sigma}}}}$, and $\underline{\hat{\boldsymbol{\theta}}}$ by maximizing (10) with respect to these parameters.

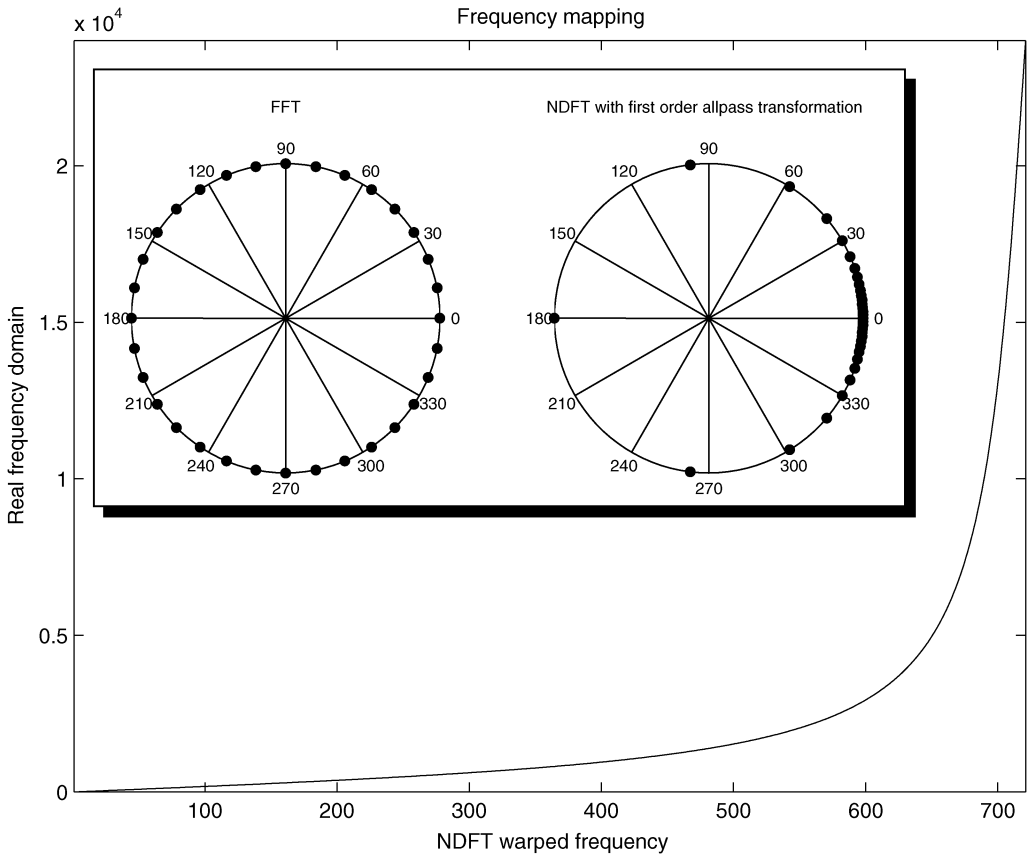

Fig. 4. Actual transformation used with an inset of a sample frequency transformation of a 1st order with fewer taps.

The problem is investigated in [20] and [21], and numerical methods are developed for the minimization of the ML function.

The DOA ML function as a function of the transmitted signal statistics is

$$\hat{\underline{\boldsymbol{\theta}}}_{\mathrm{ML}} = \arg\min_{\underline{\boldsymbol{\theta}}} \log_e \left| \underline{\underline{\mathbf{A}}} \hat{\underline{\underline{\boldsymbol{\Sigma}}}} \underline{\underline{\mathbf{A}}}^{\dagger} + \hat{\sigma}^2 \underline{\underline{\mathbf{I}}} \right|. \tag{12}$$

We should reiterate here that $\underline{\underline{\mathbf{A}}}$ above is a function of $\underline{\boldsymbol{\theta}}$, although the dependence has been dropped for notational convenience. Note also that the above requires an estimate for the statistics $\underline{\underline{\boldsymbol{\Sigma}}}$, which can be shown to be

$$\hat{\underline{\underline{\boldsymbol{\Sigma}}}}(\underline{\boldsymbol{\theta}}) = \underline{\underline{\mathbf{A}}}^{-} \left[\hat{\underline{\underline{\mathbf{R}}}} - \hat{\sigma}^2(\underline{\boldsymbol{\theta}}) \underline{\underline{\mathbf{I}}} \right] \underline{\underline{\mathbf{A}}}^{-\dagger} \tag{13}$$

where $\underline{\underline{\mathbf{A}}}^{-}$ is the pseudo-inverse of $\underline{\underline{\mathbf{A}}}$ and the statistics $\hat{\underline{\underline{\mathbf{R}}}}$ are a function of the DOA $\underline{\boldsymbol{\theta}}$. Clearly, these two optimization functions require iteration between the estimation of the statistics and the DOA.

Numerical methods have to be employed to solve this optimization problem [22].

D. Sub-Gaussian Signals

The use of a sub-Gaussian signal of equal impulsiveness provides an alternative to Cauchy distributed signals [23], and allows for dependent sources to be modeled. For this purpose, we consider a distribution of impulsiveness $\alpha = 0.5$, which is completely skewed to the positive axis, together with a multivariate Gaussian density. There is one distribution with a closed form expression, the Lévy distribution, which satisfies exactly these properties. Fig. 3 gives a top level description of the problem and signals.

- A multivariate Gaussian signal is corrupted by multiplicative Lévy noise to the half power, *i.e.*, $s_k(f) = u_k(f)^{1/2} \cdot \underline{\boldsymbol{v}}_k(f) = w_k(f) \cdot \underline{\boldsymbol{v}}_k(f)$.
- The resulting signal s_k is transformed through a set of delays $\underline{\boldsymbol{x}}(f) = \underline{\underline{\mathbf{A}}} \cdot \underline{\boldsymbol{s}}(f)$ to the receiving end of the array.
- White noise $\underline{\boldsymbol{n}}$ can corrupt the signal before it is received by the array.

The Gaussian density at a single frequency $\underline{\boldsymbol{v}}$ is

$$f(\underline{\boldsymbol{v}}) = \frac{1}{\pi^{\rho} \left| \underline{\underline{\mathbf{R}}} \right|} \exp\left(-\underline{\boldsymbol{v}}(f)^{\dagger} \underline{\underline{\mathbf{R}}}^{-1} \underline{\boldsymbol{v}}(f) \right) \tag{14}$$

and the Lévy distribution [8], [24] with $\gamma = 1/\sqrt{2}$ is given by

$$f(u) = \begin{cases} \frac{u^{-(3/2)} e^{-(1/4u)}}{2\sqrt{\pi}}, & \text{if } u > 0 \\ 0, & \text{if } u < 0. \end{cases} \tag{15}$$

From these it can be shown, as in [15], that

$$f(\underline{\boldsymbol{s}}) = \frac{1}{2\pi^{(2\kappa+1)/2} \left| \underline{\underline{\boldsymbol{\Sigma}}} \right|} \frac{\Gamma\left(\frac{2\kappa+1}{2}\right)}{\left[\frac{1}{4} + \underline{\boldsymbol{s}}^{\dagger} \underline{\underline{\boldsymbol{\Sigma}}}^{-1} \underline{\boldsymbol{s}} \right]^{(2\kappa+1/2)}} \tag{16}$$

where

$$\Gamma(z) = \int_0^{\infty} t^{z-1} e^{-t} dt. \tag{17}$$

TABLE I
DISTRIBUTIONS OF INTEREST

Lévy	$f(u) = \begin{cases} \frac{u^{-\frac{3}{2}} e^{-\frac{1}{4u}}}{2\sqrt{\pi}} & \text{if } u > 0 \\ 0 & \text{if } u < 0 \end{cases}$
Gaussian	$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}$
1-D 1-Sub-Gaussian	$f(x) = \frac{1}{2\sqrt{2}\pi\sigma} \cdot \left[\frac{x^2}{2\sigma^2} + \frac{1}{4}\right]^{-1}$
ρ-D Gaussian	$f(\underline{\underline{X}}) = \frac{1}{\pi^\rho \lvert\underline{\underline{\Sigma}}\rvert} \exp\left(-\underline{x}^\dagger \underline{\underline{\Sigma}}^{-1} \underline{x}\right)$
ρ-D 1-Sub-Gaussian	$f(\underline{X}) = \frac{\Gamma\left(\frac{2\rho+1}{2}\right)\left[\frac{1}{4} + \underline{x}^\dagger \underline{\underline{\Sigma}}^{-1} \underline{x}\right]^{-\frac{2\rho+1}{2}}}{2\pi^{(2\rho+1)/2}\lvert\underline{\underline{\Sigma}}\rvert}$

Note that if the Gaussian random variable was one-dimensional and real, then

$$f(\underline{s}) = \frac{\Gamma(1)}{2\pi\sqrt{2}\sigma} \cdot \left[\frac{1}{4} + \frac{s^2}{2\sigma^2}\right]^{-1} = \frac{1}{2\sqrt{2}\pi\sigma} \cdot \left[\frac{1}{4} + \frac{s^2}{2\sigma^2}\right]^{-1}. \quad (18)$$

As expected, at $\sigma = \sqrt{2}$ the sub-Gaussian random process is equal in distribution to the Cauchy of $\gamma = 1$. Distributions of interest to this work are listed in Table I. Note that although this expression specifically addresses the $\alpha = 1$ sub-Gaussian case, algorithms developed under these assumptions are expected to be robust for all signals of lower impulsiveness, i.e., $\alpha \geq 1$.

Now the received signal $\underline{\boldsymbol{x}} = [x_1 \ldots x_\rho]^T$ is of the form

$$\underline{\boldsymbol{x}}_r = y^{1/2} \cdot \underline{\boldsymbol{z}}_r \quad (19)$$

where again, as in the transmitted signal case, the received signal is sub-Gaussian and introducing jointly sub-Gaussian noise (the noise is a sub-Gaussian process produced by the same Lévy sequence) as shown on Fig. 3, the received signal statistics can be given by

$$\underline{\underline{\mathbf{R}}} = \underline{\underline{\mathbf{A}}}\underline{\underline{\boldsymbol{\Sigma}}}_v \underline{\underline{\mathbf{A}}}^\dagger + \sigma_n^2 \underline{\underline{\mathbf{I}}}_\rho. \quad (20)$$

Therefore, the maximum likelihood estimator over all available frequencies is

$$[\hat{\underline{\underline{\boldsymbol{\Sigma}}}}, \hat{\underline{\boldsymbol{\theta}}}, \sigma_\eta] = \arg\min_{\underline{\underline{\boldsymbol{\Sigma}}}, \underline{\boldsymbol{\theta}}} \sum_{f=f_1}^{f_M} \left\{ \log_e \left|\underline{\underline{\mathbf{R}}}\right| + \left(\frac{2\rho+1}{2}\right) \log_e \left[\underline{\boldsymbol{x}}^\dagger \underline{\underline{\mathbf{R}}}^{-1} \underline{\boldsymbol{x}} + \frac{1}{4}\right] \right\}. \quad (21)$$

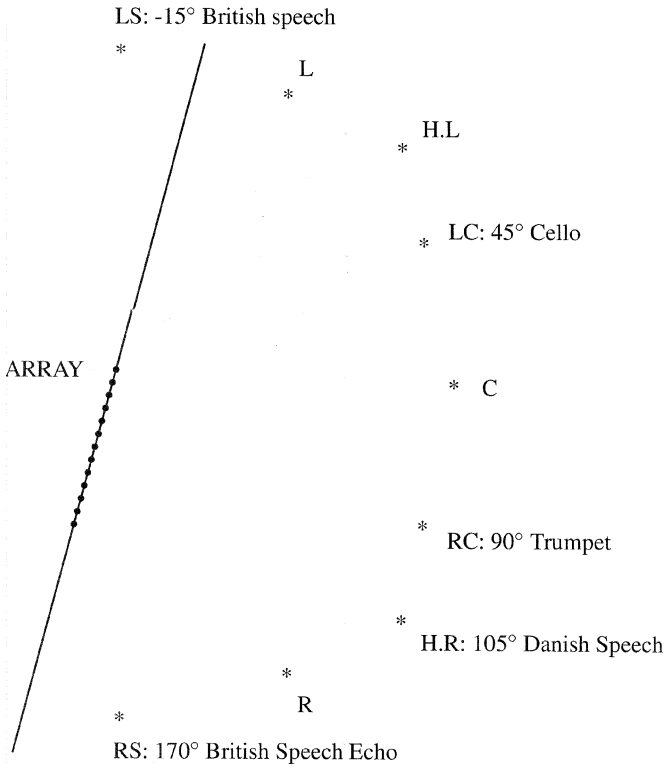

Fig. 5. Experiment setup for the 20-microphone array. Angles shown are relative to the center of the array arrangement. Only the cello and trumpet sources could be reliably localized due to the inaccuracies in sensor placement with the 20-microphone array.

E. Separable Solution

1) Estimating the Statistics: Following similar procedure as in [21] where the ML function is minimized with respect to the signal statistics and assuming known DOAs, we derived in [15] an alternative minimization function that reduces the search space

$$\underline{\underline{\boldsymbol{\Sigma}}}_{\text{ML}} = \frac{1}{M} \sum_{f=f_1}^{f_M} \left[\underline{\underline{\mathbf{A}}}^- \left(\frac{(\rho + 0.5)\, \underline{\boldsymbol{x}}\underline{\boldsymbol{x}}^\dagger}{\text{Tr}\left[\underline{\underline{\mathbf{R}}}^{-1} \underline{\boldsymbol{x}}\underline{\boldsymbol{x}}^\dagger\right] + \frac{1}{4}} - \sigma_n^2 \right) \underline{\underline{\mathbf{A}}}^{-\dagger} \right] \quad (22)$$

where $\underline{\underline{\mathbf{A}}}^- = \left(\underline{\underline{\mathbf{A}}}^\dagger \underline{\underline{\mathbf{A}}}\right)^{-1} \underline{\underline{\mathbf{A}}}^\dagger$ and

$$\underline{\underline{\mathbf{R}}}^{-1} = \frac{1}{\sigma_n^2} \left\{ \underline{\underline{\mathbf{I}}} - \underline{\underline{\mathbf{A}}} \left(\underline{\underline{\boldsymbol{\Sigma}}} \underline{\underline{\mathbf{A}}}^\dagger \underline{\underline{\mathbf{A}}} + \sigma_n^2 \underline{\underline{\mathbf{I}}} \right)^{-1} \underline{\underline{\boldsymbol{\Sigma}}} \underline{\underline{\mathbf{A}}}^\dagger \right\}. \quad (23)$$

We can observe from (22) and (23) that the unknown $\underline{\underline{\boldsymbol{\Sigma}}}_{\text{ML}}$ appears on both sides of the equation. However, the above equation can be easily and rapidly solved using the numerical iteration method.

In our experience, the iteration algorithm converges in very few cycles, and thus random initial conditions are used in the above equation. However, it is possible to estimate the underlying Gaussian statistics of a sub-Gaussian signal using fractional lower-order statistics as demonstrated in [15].

2) DOA Estimation: In the sound source localization application, the information of interest is the DOA, $\underline{\boldsymbol{\theta}}$. The statistics estimation above assumes that the DOA vector is known, and

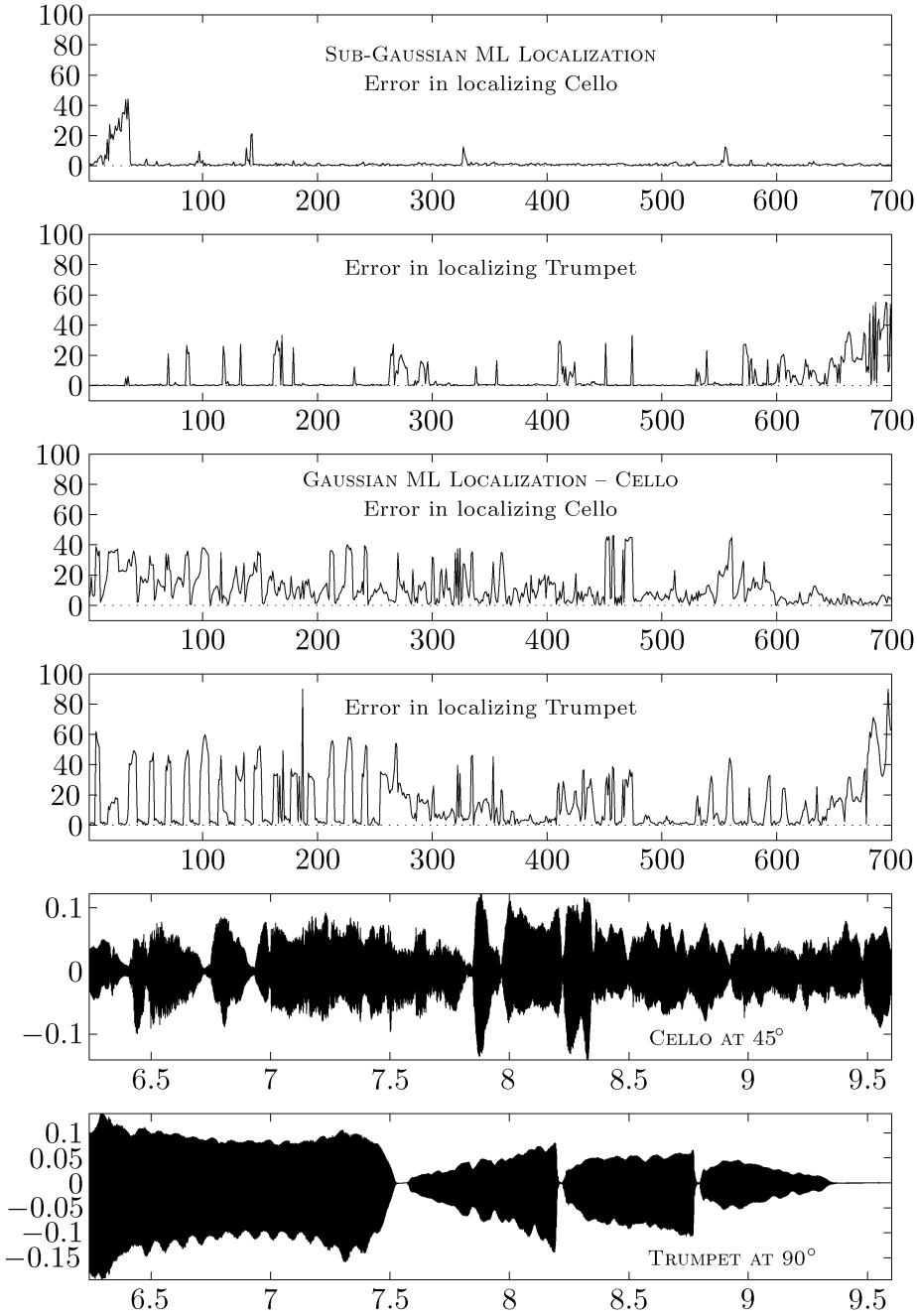

Fig. 6. Audio array ML DOA comparison for the localization of two sources at 45° and at 90° using the sub-Gaussian and Gaussian ML-based methods. Subplots 1–4 show the absolute error in the DOA measured in degrees where the horizontal axis represents time and the unit of time is measured in frames. Subplots 1&2 show the error resulting from the sub-Gaussian method and subplots 3&4 for the Gaussian methods. As can be observed the error in the Gaussian case is a significantly greater than the sub-Gaussian one. The two original sound signals are shown at the bottom two subplots—amplitude versus time in seconds—and we can see the correlation of the error rising when the amplitude of the trumpet dies off at the end. Note that, although not visible on the graph, there is actually a low amplitude signal until the very end of the figure.

here we approach the localization part of the problem. Using a pseudo-ML approach, we can express the modified ML function as

$$\hat{\underline{\boldsymbol{\theta}}} = \arg\min_{\hat{\underline{\boldsymbol{\theta}}}} \sum_{f=f_1}^{f_M} \left\{ \log_e \left|\underline{\underline{\mathbf{R}}}\right| + \left(\frac{2\rho+1}{2}\right) \log_e \left[\underline{\boldsymbol{x}}^{\dagger} \underline{\underline{\mathbf{R}}}^{-1} \underline{\boldsymbol{x}} + \frac{1}{4}\right] \right\}. \quad (24)$$

But since the statistics $\left|\underline{\underline{\mathbf{R}}}\right|$ are not a function of θ due to the $|\cdot|$ operator, the ML function can be reduced to

$$\hat{\underline{\boldsymbol{\theta}}} = \arg\min_{\hat{\underline{\boldsymbol{\theta}}}} \sum_{f=f_1}^{f_M} \left\{ \log_e \left[\underline{\boldsymbol{x}}^{\dagger} \underline{\underline{\mathbf{R}}}^{-1} \underline{\boldsymbol{x}} + \frac{1}{4}\right] \right\} \quad (25)$$

where $\underline{\underline{\mathbf{R}}}$ can be substituted with any valid statistics (identity matrix for instance). A search algorithm such as the one described in [25] can be used to find the solution of the above

equation. Note that iterating between (25) and (22) can achieve even higher accuracy estimates.

IV. Sub-Gaussian and Gaussian ML Localization Comparisons on Real Data

A. "Synthetic" Arrays

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 48 kHz, and two microphones were shifted forming 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. The room in which these sounds were collected is acoustically treated to a reverberation time of 0.6 s and there are no echo producing surfaces. Note that for all the results below the system had no memory from one frame to the next, making the DOA estimation even more challenging.[1]

The ML function for the following cases was evaluated over all frequencies by re-calculating the transformation matrix $\underline{\underline{\mathbf{A}}}$ for all possible $(\boldsymbol{\theta}, f)$ combinations, which is a computationally expensive process. For the localization part, a *Non-linear DFT* (NDFT) [26]–[28] was used in order to keep the resulting frequency domain bands narrow. Specifically, we employed the method described by Mitra *et al.* in [26], with a first order allpass filter and a 30 ms window (1440 samples). The resulting frequency mapping is shown on Fig. 4, with an inset of a more visual representation of the first-order mapping with fewer taps. Note that de Haan *et al.* [29] have evaluated the performance of NDFT's of various parameters in sound reconstruction.

1) 20-Microphone Array: In the 20-microphone array case, the aperture was 38 cm and the intersensor spacing was 2 cm, while four (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 where all five channels were active was very large. The array setup is shown on Fig. 5.

Results of localization demonstrate that the sub-Gaussian-based ML method performs significantly better than its Gaussian counterpart. Fig. 6 shows 7 s of the signal where only the cello and trumpet are being played. Each frame of the segment corresponds to a sliding window of 30 ms, and the sources were placed at 45° and at 90°. As can be observed, the sub-Gaussian ML method works significantly better. Table II 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.

TABLE II
Errors for the Gaussian-Based ML Method are More Than Double Those of the Sub-Gaussian-Based ML

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

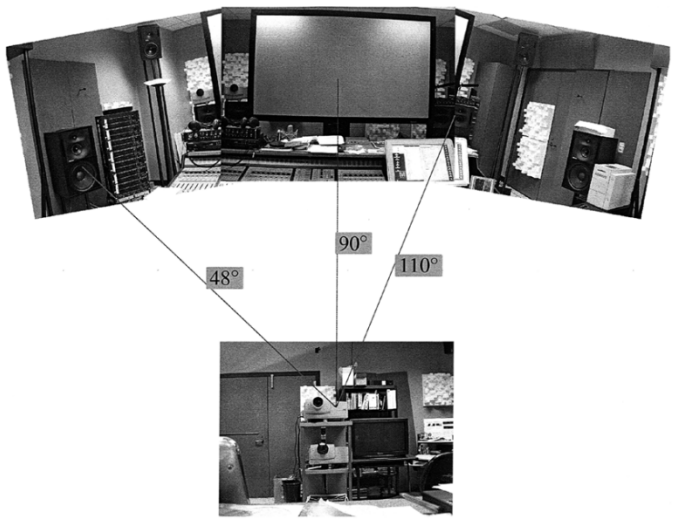

Fig. 7. Arrangement of 41-microphone array.

2) 41-Microphone Array: In the 41-microphone array, the recording conditions are similar to the previous case. However, the inter-sensor spacing is 1 cm, 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. 7 shows the positions of the sources, the array, and the flat screen,[2] which as we expect causes a strong sound reflection.

From Fig. 8 we note that, in addition to a superior performance of the sub-Gaussian-based ML, the errors of the sub-Gaussian tend to be more reasonable. In other words, the sub-Gaussian algorithm incorrectly localizes sources mostly in the range 50°–90°, which we believe corresponds to the reflections off the console, while the Gaussian-based ML is severely influenced by the noise impulsiveness and locates sources more indiscriminately. 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 III.

B. Real Arrays

In order to increase the realism of our experiments collection, we employed our 16 microphone array by rearranging the microphones in a linear pattern. The room used in this experiments

[1]This is in contrast to the PHAT-based methods such as in [5], which we also employ in our lab, where a memory for the statistics is used from one frame to the next. Clearly, in a real system that would be advantageous, but in our case we are interested in the comparison of the Gaussian- and sub-Gaussian-based ML methods.

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

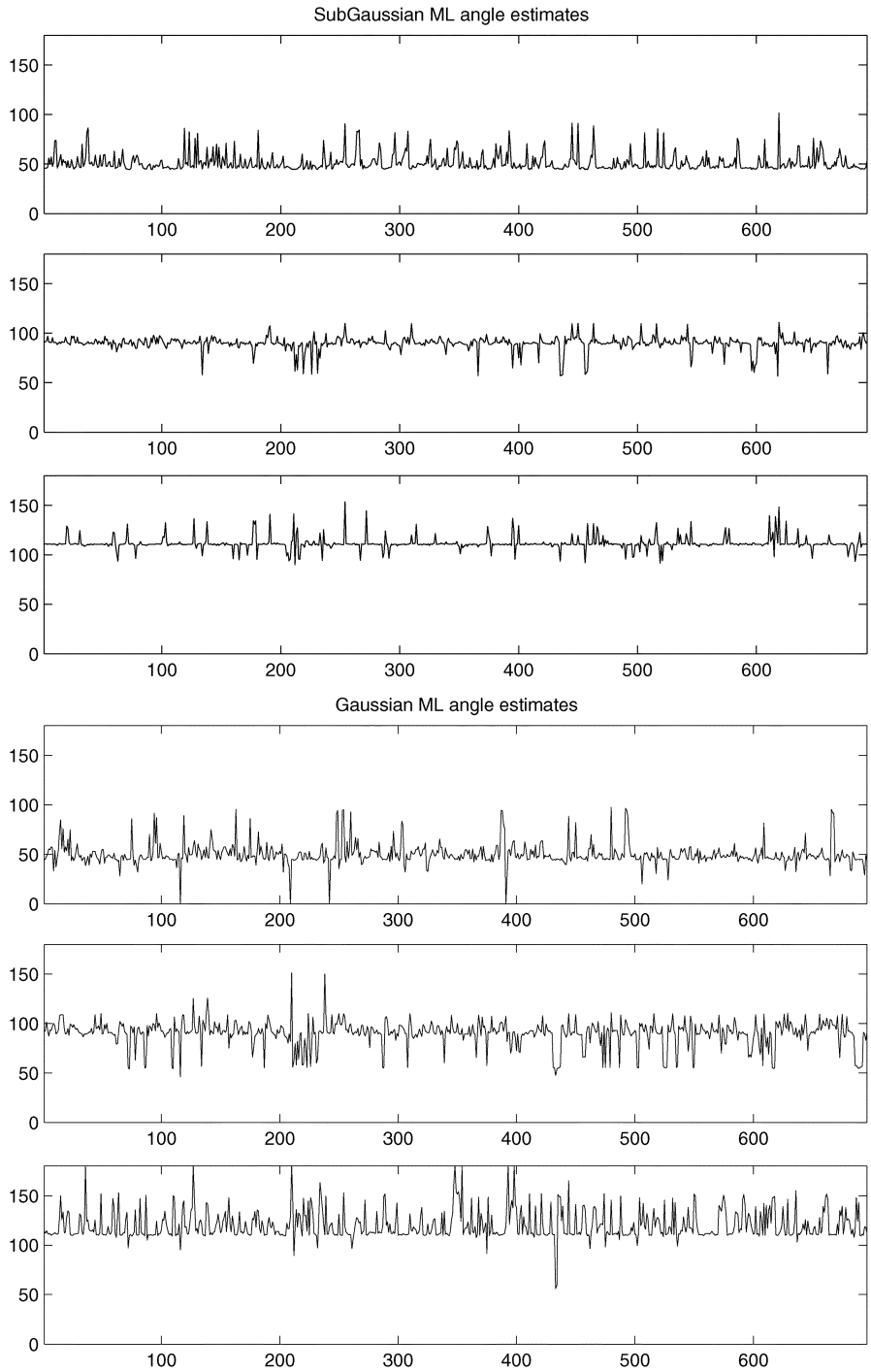

Fig. 8. Angle estimates for Gaussian- and sub-Gaussian-based ML methods for the 41-microphone setup.

is more reverberant (0.85 s) and more echoic than the one previously used. More specifically, the room is acoustically untreated and has 3 flat wall sides that are very reflective, and the opposite wall to the array is covered with a cloth to reduce echoes. The 16 microphones are placed in a linear fashion with 1 cm inter-sensor spacing, and the recordings are made at 48 kHz. Additionally, the experiments were performed using both FFT and NDFT. The signals in this case were both male speech and placed at 126° and 65° relevant to the array. Since the minimum block size is 2^{10} samples and the signal is speech, the narrow-band assumption holds well in the region of interest for speech which is above 300 Hz and below 8 kHz, as demonstrated on Fig. 9. The results between linear and nonlinear frequency transformations were statistically identical, and the average of the two methods is shown on Table IV.

The DOA estimation results are again much more accurate with the sub-Gaussian-based ML method as opposed to the Gaussian ML. Table IV shows the root mean square error in degrees for the cases of 1 cm intersensor spacing and a variable number of samples per frame. As in the previous case, the

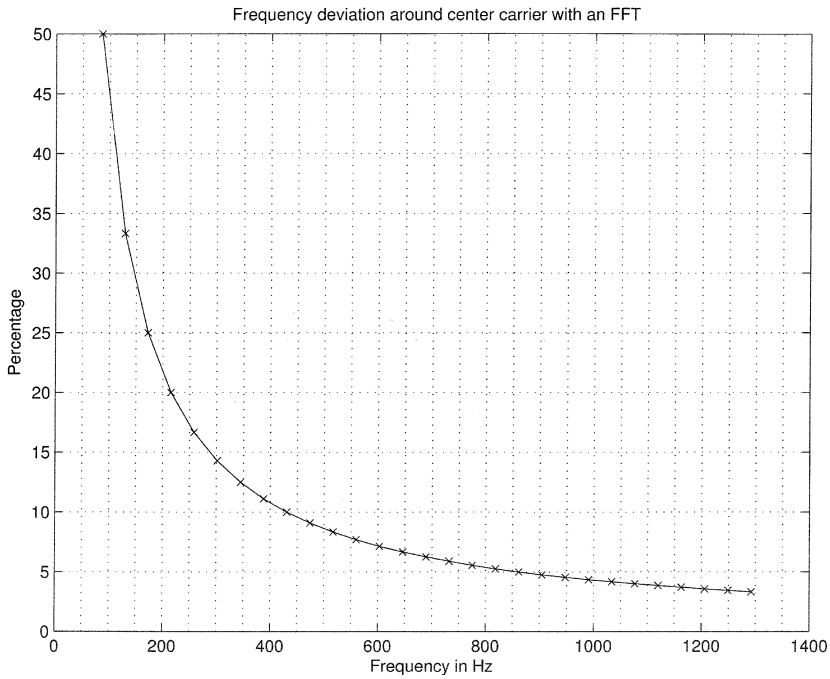

Fig. 9. Deviation around the "carrier frequency" (center frequency of the FFT) is well below 10% for the 300 Hz-8 KHz range, the region of interest for speech signals.

TABLE III
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

	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 IV
REAL ARRAY DEMONSTRATES THE IMPROVED LOCALIZATION USING SUB-GAUSSIAN RATHER THAN GAUSSIAN-BASED ML

Number of samples	R.M.S. Error Sub-Gaussian	R.M.S. Error Gaussian
2^{10}	5.9	13.8
2^{12}	4	7.9

statistics collected were only from the length of the frame and *not* averaged using any memory measure.

V. CONCLUSION

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. Our present research was motivated by existing work demonstrating the impulsiveness of sound, and by the observation that reverberation is highly dependent on the original source.

The ML solution of this model was given under a sensor array scenario, and its separable solution was derived. The separable solution assumes known statistics to localize the directions-of-arrival and known directions-of-arrival to find the statistics of the underlying processes. Although the statistics estimator could not be derived as a closed form expression, the resulting form allows for a fast iterative solution. The directions-of-arrival estimator still requires a search, but of a much smaller space.

Simulations have demonstrated the robustness of the sub-Gaussian-based ML, and encourage us to further develop methods employing the sub-Gaussian, rather than the Gaussian, model. Additionally, the performance loss of the sub-Gaussian-based ML in the case that signals are Gaussian is insignificant, which reinforces our robustness claim.

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, and in a more reverberant room (16 microphones). 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.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers and the editor for their constructive comments.

REFERENCES

[1] H. Wang and P. Chu, "Voice source localization for automatic camera pointing system in videoconferencing," in *Proc. IEEE Workshop Appl. Signal Processing Audio Acoustics*, New Paltz, NY, 1997.

[2] T. Yamada, S. Nakamura, and K. Shikano, "Distant-talking speech recognition based on a 3-d viterbi search using a microphone array," *IEEE Trans. Speech Audio Processing*, vol. 10, no. 2, pp. 48–56, 2002.

[3] Y. Huang, J. Benesty, G. W. Elko, and R. M. Mersereau, "Real-time passive source localization: A practical linear-correction least-squares approach," *IEEE Trans. Speech Audio Processing*, vol. 9, no. 8, pp. 943–956, Nov. 2001.

[4] N. Strobel, S. Spors, and R. Rabenstein, "Joint audio-video object localization and tracking," *IEEE Signal Processing Mag.*, vol. 18, no. 1, pp. 22–31, 2001.
[5] 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 Trans. Multimedia*, vol. 1, pp. 291–301, Sep. 1999.
[6] P. Kidmose, "Alpha-stable distributions in signal processing of audio signals," in *Proc. 41st Conf. Simulation and Modeling, SIMS2000*, 2000, pp. 87–94.
[7] S. Cambanis, G. Samorodnitsky, and M. S. Taqqu, Eds., "Stable processes and related topics," in *Progress in Probability*. Boston, MA: Birkhäuser, 1991.
[8] G. Samorodnitsky and M. S. Taqqu, *Stable Non-Gaussian Random Processes: Stochastic Models With Infinite Variance*. New York: Chapman & Hall, 1994.
[9] M. Shao and C. L. Nikias, "Signal processing with fractional lower order moments: Stable processes and their applications," *Proc. IEEE*, vol. 81, no. 7, pp. 986–1010, Jul. 1993.
[10] C. L. Nikias and M. Shao, *Signal Processing With Alpha-Stable Distributions and Applications*. New York: Wiley, 1995.
[11] P. Tsakalides, R. Raspanti, and C. L. Nikias, "Angle/doppler estimation in heavy-tailed clutter backgrounds," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 35, pp. 419–436, Apr. 1999.
[12] R. Adler, R. E. Feldman, and M. S. Taqqu, Eds., *A Practical Guide to Heavy Tails: Statistical Techniques and Applications*. Boston, MA: Birkhäuser, 1998.
[13] H. Stark and J. W. Woods, *Probability, Random Processes and Estimation Theory for Engineers*, 2nd ed. Eaglewood Cliffs, NJ: Prentice-Hall, 1994.
[14] S. Cambanis and G. Miller, "Linear problems in pth order and stable processes," *SIAM J. Appl. Math.*, vol. 41, no. 1, pp. 43–69, Aug. 1981.
[15] P. G. Georgiou and C. Kyriakakis, "Maximum likelihood parameter estimation under impulsive conditions. a Sub-gaussian signal approach," *Eur. J. Signal Process.*, submitted for publication.
[16] A. B. Gershman, C. F. Mecklenbrauker, and J. F. Bohme, "Matrix fitting approach to direction of arrival estimation with imperfect spatial coherence of wavefronts," *IEEE Trans. Signal Processing*, vol. 45, pp. 1894–1899, 1997.
[17] O. Besson, F. Vincent, P. Stoica, and A. B. Gershman, "Approximate maximum likelihood estimators for array processing in multiplicative noise environments," *IEEE Trans. Signal Processing*, vol. 48, pp. 2506–2518, Sep. 2000.
[18] P. Stoica, O. Besson, and A. B. Gershman, "Direction-of-arrival estimation of an amplitude-distorted wavefront," *IEEE Trans. Signal Processing*, vol. 49, pp. 269–276, Feb. 2001.
[19] D. H. Johnson and D. E. Dudgeon, *Array Signal Processing: Concepts and Techniques*. Englewood Cliffs, NJ: Prentice-Hall, 1993.
[20] J. F. Bohme, "Separated estimation of wave parameters and spectral parameters by maximum likelihood," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, vol. 4, 1986, pp. 2819–22.
[21] A. G. Jaffer, "Maximum likelihood direction finding of stochastic sources: A separable solution," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, vol. 5, 1988, pp. 2893–2896.
[22] D. Starer and A. Nehorai, "Newton algorithms for conditional and unconditional maximum likelihood estimation of the parameters of exponential signals in noise," *IEEE Trans. Signal Processing*, vol. 40, pp. 1528–1534, Jun. 1992.
[23] 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.
[24] V. M. Zolotarev, *One-Dimensional Stable Distributions*. Providence, RI: Amer. Math. Soc., 1986.
[25] W. Huyer and A. Neumaier, "Global optimization by multilevel coordinate search," *J. Global Optim.*, vol. 14, pp. 331–355, 1999.
[26] A. Makur and S. K. Mitra, "Warped discrete-fourier transform: Theory and applications," *IEEE Trans. Circuits Syst. I*, vol. 48, no. 9, pp. 1086–1093, Sep. 2001.
[27] S. Bagchi and S. K. Mitra, *Nonuniform Discrete Fourier Transform and its Signal Processing Applications*. Norwell, MA: Kluwer, 1999.
[28] A. Oppenheim and D. Johnson, "Computation of spectra with unequal resolution using the fast fourier transform," *Proc. IEEE*, vol. 59, pp. 299–301, 1971.
[29] J. M. de Haan, N. Grbic, I. Claesson, and S. Nordholm, "Design and evaluation of nonuniform dft filter banks in subband microphone arrays," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, 2002, pp. 1173–1176.

Panayiotis Georgiou received the B.S. and M.Eng. degrees from Pembroke College, Cambridge University, Cambridge, U.K., in 1996, and the M.Sc. and Ph.D. degrees in 1998 and 2002, respectively, from the University of Southern California (USC), Los Angeles.

Since 2002, he has been with the Speech Analysis and Interpretation Lab at USC, and the Integrated Media Systems Center, where he is currently a Research Assistant Professor. His research interests lie in the fields of microphone array signal processing, speech to speech translation and multimodal signal processing.

Chris Kyriakakis received the B.S. degree in electrical engineering from the California Institute of Technology, Pasadena, in 1985 and the M.S. and Ph.D. degrees in electrical engineering from the University of Southern California (USC), Los Angeles, in 1987 and 1993, respectively.

He is an Associate Professor in the Electrical Engineering Department at the USC's Viterbi School of Engineering and the Deputy Director of the Integrated Media Systems Center (IMSC), a National Science Foundation Engineering Research Center at USC. He is the head of the IMSC Immersive Audio Laboratory at USC. His research interests include high fidelity multichannel audio coding, immersive audio signal processing, microphone arrays for robust sound localization, multichannel audio streaming over high bandwidth networks, virtual microphones for multichannel audio synthesis, and multiple listener room equalization.