KU LEUVEN

STADIUS
Center for Dynamical Systems, Signal Processing and Data Analytics

Citation/Reference	Enzo De Sena, Mike Brookes, Patrick A. Naylor, and Toon van Waterschoot **Localisation experiments with reporting by head orientation: statistical framework and case study** *J. Audio Eng. Soc.*, vol. 65, no. 12, pp. 982-996, Dec. 2017.
Archived version	Author manuscript: the content is identical to the content of the submitted paper, but without the final typesetting by the publisher
Published version	https://doi.org/10.17743/jaes.2017.0038
Journal homepage	http://www.aes.org/journal/
Author contact	toon.vanwaterschoot@esat.kuleuven.be *+ 32 (0)16 321927*
IR	ftp://ftp.esat.kuleuven.be/pub/SISTA/vanwaterschoot/abstracts/16-146.html

(article begins on next page)

OPEN ACCESS Freely available online

CC BY

Journal of the Audio Engineering Society
Vol. 65, No. 12, December 2017 (© 2017)
DOI: https://doi.org/10.17743/jaes.2017.0038

Localization Experiments with Reporting by Head Orientation: Statistical Framework and Case Study*

ENZO DE SENA[1], **MIKE BROOKES,**[2] *AES Associate Member*, **PATRICK A. NAYLOR**[2], **AND**
(e.desena@surrey.ac.uk)
TOON VAN WATERSCHOOT,[3] *AES Associate Member*

[1] *University of Surrey, Institute of Sound Recording, Guildford, GU2 7XH, UK*
[2] *Imperial College London, Electrical and Electronic Engineering Department, Communications and Signal Processing Group, Exhibition Road London, SW7 2AZ, UK*
[3] *KU Leuven, Dept. of Electrical Engineering (ESAT–STADIUS/ETC), Andreas Vesaliusstraat 13, 3000 Leuven, Belgium*

This paper is concerned with sound localization experiments in which subjects report the position of an active sound source by turning toward it. A statistical framework for the analysis of the data from this type of experiment is presented together with a case study from a large-scale listening experiment. The statistical framework is based on a model that is robust to the presence of front/back confusions and random errors. Closed-form natural estimators are derived, and one-sample and two-sample statistical tests are presented. The framework is used to analyze the data of an auralized experiment undertaken by nearly nine hundred subjects. Results show that responses had a rightward bias and that speech was harder to localize than percussion sounds, which are results consistent with the literature. Results also show that it was harder to localize sound in a simulated room with high ceiling, despite having a higher direct-to-reverberant ratio than other simulated rooms.

0 INTRODUCTION

The phenomena governing human sound localization have been the subject of intense study since the turn of the twentieth century [1]. A large variety of characteristics have been studied, ranging from the just-noticeable-differences in localization accuracy, adaptation, and learning effects, to the influence of the source's spectral content and room reflections [1–3]. Recent experiments also studied the contribution of high frequency content in the presence of a noise masker [4], the degradation of localization accuracy with outer ears occlusions [5] and bilateral hearing aids [6], and the localization of multiple coherent sound sources [7].

Subjects are typically asked to indicate the direction of the perceived sound source by (a) reporting the closest loudspeaker, fixed acoustic pointer or label [7, 4, 2]; (b) steering a movable pointer [8]; (c) reporting the direction on a *graphical user interface* (GUI) or on paper [6]; or (d) turning their face toward the perceived sound source after the stimulus has been presented [9, 5]. This paper is concerned with experiments where subjects report the position of the perceived sound source by turning toward it *while* the stimulus is being presented. This methodology makes it possible to study the dynamics of how subjects rotate themselves to find a sound source, to study the mechanisms that enable them to resolve front/back confusions, and to study the reported direction of the perceived sound source. This paper focuses on the latter of the three.

Metrics of interest for the the perceived sound source include the mean direction and concentration of responses and how many subjects experience a front/back confusion or make a random error. Since the subjects turn toward an active sound source and give their answer once they believe the sound source is in front of them, the methodology considered in this paper is limited to the study of localization in frontal directions. This restriction allows subjects to fine tune their initial decisions and is particularly useful in cases where the stimuli are hard to localize (e.g., in echolocation tasks or when the auditory system is interfered with) and in experiments involving untrained subjects. The task of

*This research work was carried out in the frame of (a) the FP7-PEOPLE Marie Curie Initial Training Network "Dereverberation and Reverberation of Audio, Music, and Speech (DREAMS)," funded by the European Commission under Grant Agreement no. 316969; (b) KU Leuven Impulsfonds IMP/14/037; (c) KU Leuven Internal Funds VES/16/032 and was supported by (d) a Postdoctoral Fellowship (F+/14/045) of the KU Leuven Research Fund and by (e) EPSRC Grant EP/M026698/1. The scientific responsibility is assumed by the authors.

turning toward a sound source is, in fact, easy to understand and is a natural and intuitive reaction to sound.

The main contribution of this paper is a statistical framework designed to analyze the data obtained with this experimental methodology. The proposed statistical framework is robust to the presence of front/back confusions and random errors. The framework is then used to analyze the data of a large-scale auralized experiment. The objective of this experiment was to study localization performance in the horizontal plane in an informal setting and with little training, which are conditions of interest because they are similar to those typically encountered in consumer applications of binaural audio. An earlier version of the experiment description with partial results was presented at the 60th AES International Conference [10].

This paper is organized as follows. Sec. 1 outlines the experimental context considered here. Sec. 2 reviews concepts of circular statistics that form the basis of the proposed statistical framework, which is presented in Sec. 3. Sec. 4 describes in detail the design of the large-scale auralized experiment and presents an analysis of the data based on the proposed statistical framework. Sec. 5 concludes the paper.

1 EXPERIMENTAL CONTEXT

The experimental context considered in this paper has the following characteristics. The subject is presented with a sound stimulus and is asked to indicate the direction of the perceived sound source by turning themselves toward it. The sound stimulus stays active throughout the test, including while the subject is turning to identify the source. The sound stimulus may consist of a single sound source in free field or more complex acoustical situations, e.g., a sound source in a reverberant room or multiple coherent sound sources.

The task of the subject is to rotate their head or body until the sound source is perceived to be in front of them. Once confident about the direction of the perceived sound source, the subject confirms the choice. The perceived sound source stays in a fixed position in space.

The experiment could be carried out in an actual physical setting, e.g., with a loudspeaker in a reverberant room. Alternatively, the desired physical setting can be simulated and the resulting binaural stimulus played back through headphones. In this case, the binaural stimulus has to be smoothly updated in real-time as the subject turns, so as to mimic the change that the subject would experience in an actual physical setting with an external stationary sound source.

In order to isolate sound perception as the only factor influencing the decision, no visual cue about the position of the sound source is available. Furthermore, the initial look direction of the subject with respect to the sound source is random and uniformly distributed.

Fig. 1 shows the apparatus used in the large-scale experiment described in detail later in Sec. 4. In this experiment subjects wore headphones and stood on a rotating platform. They could freely turn themselves by applying force on a stationary wheel in the center of the platform. A gyroscope fixed to the platform measures the platform rotation, and this information is used to update the binaural stimulus in real time. Here, the subject is trying to localize a *stationary* sound source in the *stationary* virtual room by rotating themselves on the platform.



Fig. 1. Apparatus used in the large-scale auralized experiment.

Another example of the methodology described in this section is the echolocation experiment of the type considered by Pelegrin-Garcia et al. in [11] and subsequent works by the same authors. In this class of experiments, subjects wear head-tracked headphones and a lavalier microphone. Self-generated oral sounds are picked up by the microphone and are processed by a real-time audio processor that simulates the presence of a stationary virtual wall somewhere around the subject. Subjects are asked to turn toward the virtual wall. Here, the perceived sound source sought by the subjects is the acoustic echo of their own voice.

User responses can be divided into three classes. The first class consists of responses in which the subject correctly identified the sound source within a certain angular tolerance. The second class consists of responses where the subject experienced a front/back confusion. In this case the responses are concentrated around the opposite direction. This is due to the fact that when the subject turns toward the perceived sound source, the cone of confusion [1] collapses onto the median sagittal plane. The third class consists of erroneous responses; these include cases where, for instance, the subject could not identify the sound source, did not understand the task, or ended the task early.

2 ELEMENTS OF CIRCULAR STATISTICS

The data analysis of localization experiments typically involves aperiodic statistical moments, e.g., mean, variance and mean squared errors, and statistical tests that assume normally distributed data, e.g., t-test and ANOVA [3]. While the normal distribution is an acceptable approximation in some cases, angular data is periodic in nature, thus circular statistical moments and circular distributions should be used instead. This section briefly reviews elements of circular statistics. Thorough treatments of this topic can be found in Mardia and Jupp [12] and in Fisher [13].

Let $f_\Theta(\vartheta)$ be the *probability density function* (PDF) of the continous circular random variable Θ, with $f_\Theta(\vartheta) \geq 0$, $f_\Theta(\vartheta + 2\pi) = f(\vartheta)$ and $\int_0^{2\pi} f_\Theta(\vartheta)d\vartheta = 1$. The l-th trigonometric moment of Θ is defined as

$$\gamma_l' = \mathbb{E}[e^{il\Theta}] = \int_0^{2\pi} f_\Theta(\vartheta)e^{il\vartheta}d\vartheta, \tag{1}$$

which can be written in polar coordinates as $\gamma_l' = \rho_l' e^{i\mu_l'}$, with $i = \sqrt{-1}$. The parameter ρ_1' is denoted as *mean resultant length*, and μ_1' as the *mean direction*. Due to the importance of these two statistics, ρ_1' and μ_1' are usually written simply as ρ and μ, respectively. In the context of this paper μ indicates the direction of the perceived sound source. The cosine and sine moments are defined as the real and imaginary parts of γ_l': $\alpha_l' = \mathbb{E}[\cos(l\Theta)]$ and $\beta_l' = \mathbb{E}[\sin(l\Theta)]$.

The l-th *central* trigonometric moment of Θ is defined as the l-th trigonometric moment of the random variable $\Theta - \mu$ and are denoted here by γ_l:

$$\gamma_l = \mathbb{E}[e^{il(\Theta-\mu)}] = \int_0^{2\pi} f_\Theta(\vartheta)e^{il(\vartheta-\mu)}d\vartheta. \tag{2}$$

The corresponding central cosine and sine moments are $\alpha_l = \mathbb{E}[\cos(l(\Theta - \mu))]$ and $\beta_l = \mathbb{E}[\sin(l(\Theta - \mu))]$, respectively.

The central trigonometric moment can be expressed as a function of the (non-central) trigonometric moment as

$$\gamma_l = \mathbb{E}[e^{il\Theta}]e^{-il\mu} = \gamma_l' e^{-il\mu} = \rho_l' e^{i\mu_l'}e^{-il\mu}. \tag{3}$$

Therefore $\gamma_l = \rho_l e^{i\mu_l}$ with $\mu_l = \mu_l' - l\mu$ and $\rho_l = \rho_l'$.

Consider now N sample observations of Θ, denoted in the following as $\theta = [\theta_1, \ldots, \theta_N]^T$. In the context of this paper the sample observations θ are the angles reported by the subjects, and N is the number of experiments for a certain condition. The sample equivalents of α_l' and β_l' are given by

$$a_l' = \frac{1}{N}\sum_{n=1}^{N}\cos(l\theta_n) \text{ and } b_l' = \frac{1}{N}\sum_{n=1}^{N}\sin(l\theta_n). \tag{4}$$

From the sample moments a_l' and b_l', one can derive the sample equivalents of ρ and μ as

$$\overline{R} = \sqrt{{a_1'}^2 + {b_1'}^2}, \tag{5}$$

$$\overline{\theta} = \begin{cases} \tan^{-1}\left(b_1'/a_1'\right) & a_1' \geq 0 \\ \tan^{-1}\left(b_1'/a_1'\right) + \pi & a_1' < 0 \end{cases}. \tag{6}$$

The *von Mises* (vM) distribution is among the most extensively studied circular distributions. The PDF of the vM distribution is given by

$$f_\Theta(\vartheta; \mu, \kappa) = \frac{e^{\kappa\cos(\vartheta-\mu)}}{2\pi I_0(\kappa)}, \tag{7}$$

where $I_0(\kappa)$ is the modified Bessel function of the first kind of order zero. The parameter κ is the *concentration parameter*. For $\kappa = 0$, the vM distribution degenerates to a uniform distribution. On the other hand, for large κ the vM distribution tends to a normal distribution with variance $1/\kappa$. Closed-form *maximum likelihood* (ML) estimators of the parameters of the vM distribution are available in the literature, together with one-sample and two-sample statistical tests.

The vM distribution is well suited to model the angular dispersion around the perceived angle in cases where the subject correctly identified the sound source. However, as will be shown later in this paper, in the presence of front/back confusions and random errors the vM distribution and the associated statistical tests fail. In order to model front/back confusions, a suitable distribution is the so-called *3-parameter von Mises mixture* (vMM3), which is a mixture of two von Mises distributions having the same concentration parameter κ but mean directions that are π apart. This distribution has a PDF given by

$$f_\Theta(\vartheta; \mu, k, p) = \frac{pe^{\kappa\cos(\vartheta-\mu)} + (1-p)e^{-\kappa\cos(\vartheta-\mu)}}{2\pi I_0(\kappa)}, \tag{8}$$

where $p \in [0, 1]$ is the convex combination parameter. The shape of the vM and vMM3 distributions can be seen, for example, in Fig. 9. Closed-form natural estimators (i.e., method of moments-based) exist for the vMM3 distribution [12]. One-sample tests using numerical ML optimization were studied by Grimshaw et al. [14].

3 VON MISES AND UNIFORM MIXTURE (vMUM) MODEL

As will be shown later in this paper, the vMM3 model and the associated one-sample and two-sample statistical tests perform poorly in the presence of uniformly-distributed random errors. This motivates the *von Mises and uniform mixture* (vMUM) statistical model, which is presented in this section.

3.1 Model Definition

Since the initial look direction of the subject is drawn from a uniform distribution, it is reasonable to model the erroneous decisions as uniformly distributed. Consider then the following statistical model:

$$f_\Theta(\vartheta; \mu, \kappa, p_1, p_2, p_3) = \frac{p_1 e^{\kappa\cos(\vartheta-\mu)} + p_2 e^{-\kappa\cos(\vartheta-\mu)}}{2\pi I_0(\kappa)} + \frac{p_3}{2\pi} \tag{9}$$

with $p_1, p_2, p_3 \in [0, 1]$ and $p_1 + p_2 + p_3 = 1$. This model will be referred to as vMUM in the following. Here, the values p_1, p_2, p_3 can be seen as simple parameters of the model. A different interpretation of these values is to consider them as the *probability mass function* (PMF) of an unobserved latent variable describing whether the subject experienced a frontal image, a front/back confusion or made a random error. With this interpretation, the terms $\frac{e^{\kappa\cos(\vartheta-\mu)}}{2\pi I_0(\kappa)}$, $\frac{e^{-\kappa\cos(\vartheta-\mu)}}{2\pi I_0(\kappa)}$ and $\frac{1}{2\pi}$ take the meaning of the PDFs of the *incomplete data* while $f_\Theta(\vartheta; \mu, \kappa, p_1, p_2, p_3)$ takes the meaning of PDF of the *complete data*.

The central moments of Θ can be written as

$$\alpha_l = \mathbb{E}[\cos(l(\Theta-\mu))] = \frac{I_l(\kappa)}{I_0(\kappa)}\left(p_1+(-1)^l p_2\right)+p_3\delta_l, \quad (10)$$

$$\beta_l = \mathbb{E}[\sin(l(\Theta-\mu))] = 0, \quad (11)$$

where δ_l is the Kronecker delta function. Appendix A.1 provides a proof of this result.

3.2 Parameter Estimation

3.2.1 Method of Moments Estimator (vMUM-MME)

Similarly to the derivation of the *method of moments estimator* (MME) for the vMM3 distribution [12], consider the random variable associated with the double-wrapped angle, i.e., $\Phi = 2\Theta$. The PDF of a double-wrapped variable can be written as [12] $f_\Phi(\varphi) = \frac{1}{2}f_\Theta(\varphi/2)+\frac{1}{2}f_\Theta(\varphi/2+\pi)$, and thus, with simple trigonometric and algebraic manipulations:

$$f_\Phi(\varphi) = \frac{p_1+p_2}{2\pi I_0(\kappa)}\cosh(\kappa\cos(\varphi/2-\mu))+\frac{1-(p_1+p_2)}{2\pi},$$

where the dependency on the parameters is omitted for clarity. The advantage of considering the random variable Φ instead of the original random variable Θ is that the parameters p_1 and p_2 do not appear separately but only as p_1+p_2. This enables all the parameters to be estimated one at a time, as explained in the following.

The central moments of Φ can be calculated as

$$\alpha_l^w = \mathbb{E}[\cos(l(\Phi-2\mu))] = p_w\frac{I_{2l}(\kappa)}{I_0(\kappa)}+(1-p_w)\delta_{2l}$$

$$\beta_l^w = 0,$$

where $p_w = p_1 + p_2$. Appendix A.2 provides a proof of this result. Since $\beta_l^w = 0$, then $\gamma_l^w = \alpha_l^w + i\beta_l^w = \alpha_l^w$. Using Eq. (3), the l-th trigonometric moment can therefore be written as

$$\gamma_l^{'w} = \gamma_l^w e^{i2l\mu} = \left(p_w\frac{I_{2l}(\kappa)}{I_0(\kappa)}+(1-p_w)\delta_{2l}\right)e^{i2l\mu}. \quad (12)$$

Since $p_w\frac{I_{2l}(\kappa)}{I_0(\kappa)}+(1-p_w)\delta_{2l}\in\mathbb{R}$, then $\angle\gamma_l^{'w} = 2l\mu$. Applying the method of moments to the phase of the first trigonometric moment, $\gamma_1^{'w}$, gives $\overline{\phi} = 2\hat{\mu}$, where $\overline{\phi}$ is the mean sample direction of Φ, and thus

$$\hat{\mu} = \frac{\overline{\phi}}{2}. \quad (13)$$

The first and second moments of Φ are given by

$$\alpha_1^w = p_w\frac{I_2(\kappa)}{I_0(\kappa)} \text{ and } \alpha_2^w = p_w\frac{I_4(\kappa)}{I_0(\kappa)}, \quad (14)$$

respectively. Assuming that $\frac{I_2(\kappa)}{I_0(\kappa)}\neq 0$, or, in other words, that $\kappa\neq 0$, one can isolate p_w from α_1^w and replace it in the expression of α_2^w, which gives

$$\alpha_2^w = \alpha_1^w\frac{I_4(\kappa)}{I_2(\kappa)}. \quad (15)$$

By replacing the moments with their sampled equivalents, an estimate of the concentration parameter κ can be taken as the solution of

$$\frac{1}{N}\sum_{n=1}^{N}\cos(2(\phi_n-2\hat{\mu})) = \frac{1}{N}\sum_{n=1}^{N}\cos(\phi_n-2\hat{\mu})\frac{I_4(\hat{\kappa})}{I_2(\hat{\kappa})}. \quad (16)$$

The value of $\hat{\kappa}$ is found using non-linear optimization and is called MME of κ. Notice that a similar step is necessary to obtain the parameter κ of the vMM3 model.

The convex parameter p_w can now be estimated using the expression of the first central moment:

$$\frac{1}{N}\sum_{n=1}^{N}\cos(\phi_n-2\hat{\mu}) = \hat{p}_w\frac{I_2(\hat{\kappa})}{I_0(\hat{\kappa})}. \quad (17)$$

Notice that isolating $\hat{p}_w$ from the above equation may produce a solution outside the closed interval [0, 1]. This is the same problem encountered in the estimation of the vMM3 distribution parameters [12], and, more in general, in method of moments estimates. When this happens, one approach is to find the value of $p_w\in[0,1]$ that best satisfies Eq. (17), e.g., in the least square sense. Another, simpler approach is to associate the value $\hat{p}_w = 0$ to all negative estimates and the value $\hat{p}_w = 1$ to all estimates larger than one. This approach is the one used in the simulations presented in this paper for both the vMM3 and the vMUM estimates.

It only remains to estimate one of the two parameters p_1 and p_2, with the second being determined via the expression $p_w = p_1 + p_2$. Using Eq. (10), the first central moment of the unwrapped random variable Θ can be written as

$$\alpha_1 = (p_1-p_2)\frac{I_1(\kappa)}{I_0(\kappa)} = (2p_1-p_w)\frac{I_1(\kappa)}{I_0(\kappa)}. \quad (18)$$

By applying again the method of moments, the parameter p_1 can be estimated as the solution of

$$\frac{1}{N}\sum_{n=1}^{N}\cos(\theta_n-\hat{\mu}) = (2\hat{p}_1-\hat{p}_w)\frac{I_2(\hat{\kappa})}{I_0(\hat{\kappa})}. \quad (19)$$

In summary, an estimate of the model parameters can be obtained as follows:

$$\begin{cases}\hat{\mu} = \overline{\phi}/2\\ \hat{\kappa}: \frac{1}{N}\sum_{n=1}^{N}\cos(2(\phi_n-2\hat{\mu})) = \frac{1}{N}\sum_{n=1}^{N}\cos(\phi_n-2\hat{\mu})\frac{I_4(\hat{\kappa})}{I_2(\hat{\kappa})}\\ \hat{p}_w\in[0,1]: \frac{1}{N}\sum_{n=1}^{N}\cos(\phi_n-2\hat{\mu}) = \hat{p}_w\frac{I_2(\hat{\kappa})}{I_0(\hat{\kappa})}\\ \hat{p}_1\in[0,1]: \frac{1}{N}\sum_{n=1}^{N}\cos(\theta_n-\hat{\mu}) = (2\hat{p}_1-\hat{p}_w)\frac{I_2(\hat{\kappa})}{I_0(\hat{\kappa})}\\ \hat{p}_2\in[0,1]: \hat{p}_w = \hat{p}_1+\hat{p}_2\end{cases}$$

which will be referred to as the *vMUM method of moments estimator* (vMUM-MME) below.

In practice, this procedure yields large values of $\hat{\kappa}$ when the ratio of the sample moments a_2^w/a_1^w is close to unity. In order to alleviate this issue, after obtaining a first estimate of κ and p_w, one can refine the estimate of κ by solving Eq. (17) for $\hat{\kappa}$ (instead of for $\hat{p}_w$ as done in the procedure

described above). It was found empirically that small further improvements can be obtained by iteratively solving Eq. (17) for $\hat{\kappa}$ and then for $\hat{p}_w$. In the results presented in this paper, two such additional iterations are carried out.

3.2.2 vMUM Maximum Likelihood Estimator (vMUM-MLE)

The *maximum likelihood estimator* (MLE) of the vMUM model parameters is obtained as

$$\{\hat{\mu}, \hat{\kappa}, \hat{p}_1, \hat{p}_2, \hat{p}_3\} = arg\,max_{\mu,\kappa,p_1,p_2,p_3} f_\Theta(\theta; \mu, \kappa, p_1, p_2, p_3), \quad (20)$$

subject to $p_1, p_2, p_3 \in [0, 1]$ and $p_1 + p_2 + p_3 = 1$. Here, the PDF $f_\Theta(\theta; \mu, \kappa, p_1, p_2, p_3)$ is seen as a function of the parameters for a fixed set of sample observations, θ, and represents the *likelihood function*. Since a closed-form solution of this problem is not known, it is necessary to resort to nonlinear optimization. As noted in Sec. 3.3, it is important to initialize the algorithm carefully so as to avoid convergence to a local maximum. Unless stated otherwise, the initialization used in the simulations of this paper will be the vMUM-MME estimate.

3.3 Performance Analysis of Parameter Estimation

The performance of the proposed estimator is assessed via Monte Carlo simulations. A total of 10,000 sets of N samples are generated using the vMUM statistical model. The random samples are generated using the algorithm proposed by Best and Fisher [15]. For each set, the parameters are drawn from uniform distributions with $\mu \sim \mathcal{U}(0, 2\pi)$, $\kappa \sim \mathcal{U}(0, 100)$, $p_2 \sim \mathcal{U}(0, 0.3)$ and $p_3 \sim \mathcal{U}(0, 0.3)$, where the symbol $\sim$ stands for "distributed as." The parameter p_1 is then obtained as $p_1 = 1 - p_2 - p_3$.

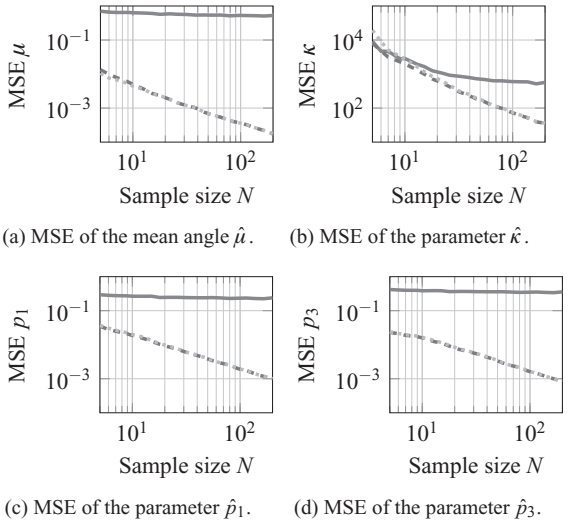
The performance is compared to the standard non-Bayesian MME of the vMM3 parameters [12], which is termed here *vMM3 method of moments estimator* (vMM3-MME).

The simulations are run using Matlab R2015b with the default random seeding. The trust-region-dogleg algorithm (Matlab command *fsolve*) is used to find κ in Eq. (16) and in a similar step required by the vMM3-MME [12]. The *sequential quadratic programming* (SQP) algorithm (Matlab command *fmincon*) is used to solve the constrained optimization problem involved in the ML estimations.

The *mean squared error* (MSE) is used here as metric to assess the estimators performance. MSE values above the 95-percentile are considered as outliers and are removed from the data. Unless stated otherwise, the sample size is $N = 20$.

3.3.1 Performance of vMUM-MLE for Different Starting Points

In order to assess the sensitivity of the vMUM-MLE estimate to its initialization, Monte Carlo simulations with different starting points were run. The first case is the vMUM-MME estimate. The second case is a random starting point. Here, μ and κ are drawn from uniform distributions with $\mu \sim \mathcal{U}(0, 2\pi)$ and $\kappa \sim \mathcal{U}(0, 100)$. Notice that this choice of κ gives the random estimate a slight advantage because the range corresponds to the actual range used to generate the data. The parameters p_1, p_2, and p_3 are all drawn from a uniform distribution between 0 and 1 and subsequently normalized such that $p_1 + p_2 + p_3 = 1$. The third case is obtained from a grid-search using 10 points across each of the five dimensions in the parameter space (it is assumed that $\kappa \in [0, 100]$, again giving it a slight advantage), which together amount to a total of 10^4 points.



(a) MSE of the mean angle $\hat{\mu}$. (b) MSE of the parameter $\hat{\kappa}$.

(c) MSE of the parameter $\hat{p}_1$. (d) MSE of the parameter $\hat{p}_3$.

Fig. 2. Mean squared error of the vMUM-MLE estimator for various initializations: random starting point (——), vMUM-MME starting point (- - -) and full-search starting point (.....). Results obtained from Monte Carlo simulations.

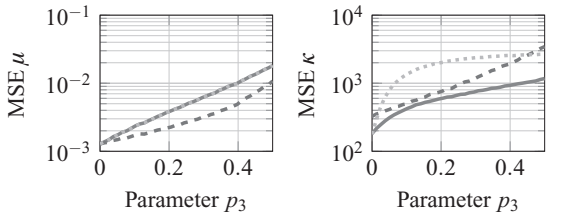
Fig. 2 shows that the random starting point results in a very poor performance, indicating that the likelihood function has multiple local maxima. The full-search starting point and the vMUM-MME starting point perform equally well. Given that the latter also requires significantly fewer computations, the vMUM-MME estimate is shown to be the most suitable starting point for the vMUM-MLE optimization problem and is used throughout the rest of this paper.

3.3.2 Performance as a Function of p_3

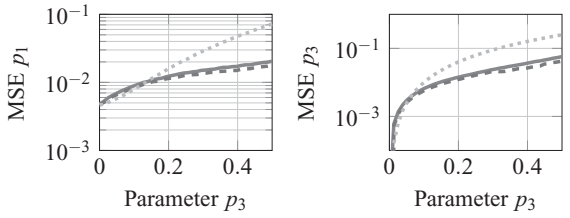
Fig. 3 shows the mean squared error of the different estimators as a function of p_3. It can be observed that all the estimators perform equally well for values of p_3 close to zero. However, as p_3 increases, the vMUM-based estimators significantly outperform the vMM3-based one. Hence, the vMUM-based estimators can be seen to be more robust to the presence of uniformly-distributed errors.

3.3.3 Performance as a Function of N

Fig. 4 shows the mean squared error of the different estimators as a function of the sample size N. It can be observed that the vMUM-based estimators outperform the vMM3-based one. The vMUM-MLE and vMUM-MME estimators perform equally well, except for the estimation

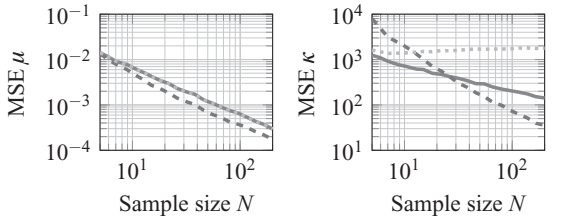


(a) MSE of the mean angle $\hat{\mu}$. (b) MSE of the parameter $\hat{\kappa}$.

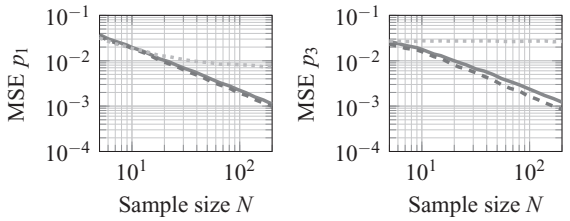


(c) MSE of the parameter $\hat{p}_1$. (d) MSE of the parameter $\hat{p}_3$.

Fig. 3. Mean squared error of the vMUM-MME estimator (——), vMUM-MLE estimator (– – –) and vMM3-MME estimator (.....) as a function of the parameter p_3. Results obtained from Monte Carlo simulations with sample size $N = 20$.



(a) MSE of the mean angle $\hat{\mu}$. (b) MSE of the parameter $\hat{\kappa}$.



(c) MSE of the parameter $\hat{p}_1$. (d) MSE of the parameter $\hat{p}_3$.

Fig. 4. Mean squared error of the vMUM-MME estimator (——), vMUM-MLE estimator (– – –) and vMM3-MME estimator (.....) as a function of the sample size N. Results obtained from Monte Carlo simulations.

of the concentration parameter κ, where the vMUM-MME outperforms vMUM-MLE for sample sizes smaller than around $N = 25$. This may seem surprising due to the fact that the MLE has a higher likelihood function than the MME. This however does not imply a lower MSE, and, in fact, some ML estimators are known to have poorer MSE than *method of moments* (MM) estimators in cases with small sample sizes.

3.4 Single-Sample Test of the Mean Direction

Consider the case where one wishes to test whether the data is drawn from a distribution with a given mean angle μ_0. For instance, a hypothesis tested later in the paper is whether the data has a zero directional mean, i.e., $\mu_0 = 0$. Assume that the a priori probability on whether this is true or not is unknown, as is the case in typical listening experiments. This amounts to the following non-Bayesian hypothesis test:

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases}, \tag{21}$$

with *null hypothesis* H_0 and *alternative hypothesis* H_1. Assume also that the model parameters are unknown. Hence, this is a non-Bayesian test of composite hypotheses [16]. Notice that whenever front/back confusions are present, the statistical model is bi-modal and π-symmetric, and therefore testing for $\mu = \mu_0$ is the same as testing for $\mu = \mu_0 + \pi$.

In this paper $P_d = Pr(\hat{H}_1; H_1)$ denotes the probability of rejecting the null hypothesis given that the alternative hypothesis is true, and $P_{fa} = Pr(\hat{H}_1; H_0)$ denotes the probability of rejecting the null hypothesis given that the null hypothesis is true.

A typical approach in this context is to seek the best P_d for a given P_{fa}. Toward this end, a commonly used test is the *generalized likelihood ratio test* (GLRT), which can be shown to have certain optimality properties [16]:

$$R(\theta) = \frac{\max\limits_{\hat{\mu},\hat{\kappa},\hat{p}_1,\hat{p}_2,\hat{p}_3} f_\Theta(\theta; \hat{\mu}, \hat{\kappa}, \hat{p}_1, \hat{p}_2, \hat{p}_3)}{\max\limits_{\tilde{\kappa},\tilde{p}_1,\tilde{p}_2,\tilde{p}_3} f_\Theta(\theta; \mu_0, \tilde{\kappa}, \tilde{p}_1, \tilde{p}_2, \tilde{p}_3)} \underset{\hat{H}_0}{\overset{\hat{H}_1}{\gtrless}} \lambda, \tag{22}$$

where $\hat{\mu}$, $\hat{\kappa}$, $\hat{p}_1$, $\hat{p}_2$, $\hat{p}_3$ are the ML estimates under the hypothesis H_1 and are termed *unrestricted MLE*, $\tilde{\kappa}$, $\tilde{p}_1$, $\tilde{p}_2$, $\tilde{p}_3$ are the ML estimates under the hypothesis H_0 and are termed *restricted MLE*, and λ is a desired threshold.

The restricted MLE is calculated as in Sec. 3.2.2 but with the constraint $\mu = \mu_0$. The unrestricted MLE is also calculated as in Sec. 3.2.2, using the restricted MLE as starting point, which guarantees that $R(\theta)$ is larger than one. Using the restricted MLE as starting point, however, sometimes causes the optimization algorithm to become stuck in a local maximum near $\mu = \mu_0$. To overcome this issue, the unrestricted MLE is calculated again using the vMUM-MLE as starting point. Between the two so-obtained unrestricted MLEs, the one with higher likelihood value is chosen.

Regarding the choice of the threshold ξ, consider the transformation $L(\theta) = 2\ln R(\theta)$ of the likelihood ratio:

$$L(\theta) = 2\ln \frac{\max\limits_{\hat{\mu},\hat{\kappa},\hat{p}_1,\hat{p}_2,\hat{p}_3} f_\Theta(\theta; \hat{\mu}, \hat{\kappa}, \hat{p}_1, \hat{p}_2, \hat{p}_3)}{\max\limits_{\tilde{\kappa},\tilde{p}_1,\tilde{p}_2,\tilde{p}_3} f_\Theta(\theta; \mu_0, \tilde{\kappa}, \tilde{p}_1, \tilde{p}_2, \tilde{p}_3)} \underset{\hat{H}_0}{\overset{\hat{H}_1}{\gtrless}} \xi, \tag{23}$$

where $\xi = 2\ln\lambda$. The above ratio is typically termed *log-likelihood ratio* (LLR). Consider the PDFs of the random variable $L(\Theta)$ for observations obtained under the two different hypotheses, $f_L(l; H_1)$ and $f_L(l; H_0)$. The probabilities P_d and P_{fa} can be written as $P_d = \int_\xi^\infty f_L(l; H_1)dl$ and $P_{fa} = \int_\xi^\infty f_L(l; H_0)dl$, respectively. The threshold ξ is chosen so as to obtain a P_{fa} that is equal to a desired value, typically $P_{fa} = 0.05$. This requires knowledge of $f_L(l; H_0)$. A powerful result in detection theory [16] states that for $N \to \infty$ the log-likelihood ratio under H_0 is chi-squared

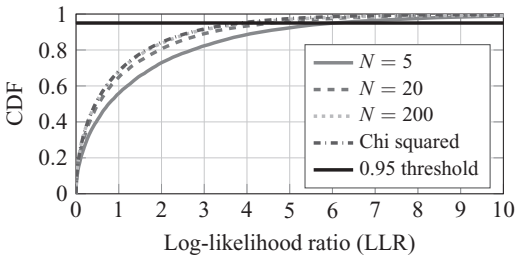


Fig. 5. Cumulative density function (CDF) of the log-likelihood ratio of the single-sample statistical test for different values of N in comparison to the χ^2_1 asymptotic approximation. Results obtained from Monte Carlo simulations.

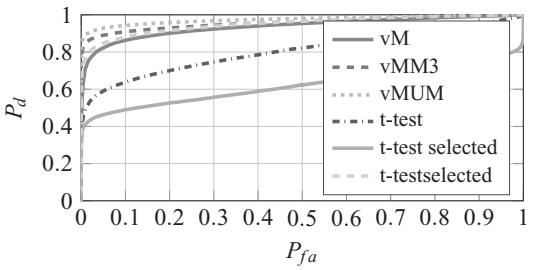


Fig. 6. Comparison of the ROC curves for the single-sample statistical test. These curves represent the available trade-offs between P_d and P_{fa} obtained by varying the value of the threshold ξ. ROC curves with points close to ($P_d = 1$, $P_{fa} = 0$) indicate a test with high performance. Results obtained from Monte Carlo simulations with sample size $N = 20$.

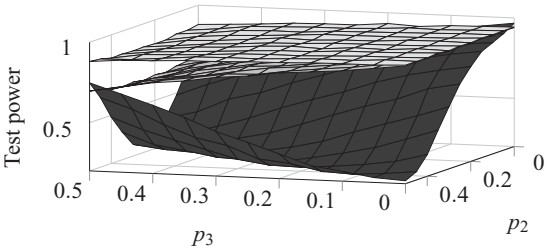


Fig. 7. Test power as a function of p_2 and p_3 for the single-sample statistical test of the mean direction. The top (grey) surface denotes vMUM, the middle (dark grey) surface denotes vMM3, and the bottom (black) surface denotes vM. Results obtained from Monte Carlo simulations with sample size $N = 20$.

distributed, i.e., $f_L(l; H_0) \sim \chi^2_r$, with r the number of degrees of freedom. For the test Eq. (21), $r = 1$.

In order to assess whether χ^2_1 is a reasonable approximation for finite N, Fig. 5 shows the PDFs of $f_L(l; H_0)$ generated using Monte Carlo simulations (10000 tests) with the same setup of Sec. 3.3. It may be observed that the χ^2_1 asymptotic approximation is already reasonably accurate at $N = 20$. Thus, if a $P_{fa} = 0.05$ is sought, the threshold can be simply chosen as $\xi = f_{\chi^2_1}(1 - 0.05) = 3.8415$. With smaller sample sets, however, it is advisable to choose the threshold in some other way, e.g., using Monte Carlo simulations directly. For instance, if one uses the threshold $\xi = 3.8415$ for $N = 5$, then the actual P_{fa} is not $P_{fa} = 0.05$ but rather $P_{fa} = 0.13$. If one wishes to achieve $P_{fa} = 0.05$ with $N = 5$, then the threshold $\xi \approx 5.5$ should be used instead.

3.5 Performance of Single-Sample Test of the Mean Direction

This section analyzes the performance of various single-sample tests using Monte Carlo simulations. The setup of the simulations is the same used in Sec. 3.3 and the number of samples is $N = 20$.

The proposed vMUM test is compared to a vMM3 test, a vM test, and the standard t-test. The vMM3 test is that described by Grimshaw et al. [14]. Here, the algorithm's starting point (which was not specified in [14]), is taken as the vMM3-MME. The vM test is the standard likelihood ratio test with unknown concentration parameter (see [12], page 122). Notice that all these tests assume the χ^2_1 asymptotic approximation. The threshold ξ is chosen such that $P_{fa} = 0.05$ in all cases.

The comparison also includes the standard t-test [16]. The t-test assumes that the data is normally distributed and, more importantly, aperiodic. The interval chosen to represent the angles is thus crucial. Indeed, if one uses the angles representation in the $[0, 2\pi)$ interval, then the sample mean, $\sum_{n=1}^{N} \theta_n$, will not be zero even when $\mu = 0$. In this section the interval is taken symmetrically around μ_0. Notice, however, that this device does not solve the issue of treating a periodic random variable as aperiodic. A typical countermeasure used in the literature is to select errors close to the hypothesized direction, which are referred to as genuine errors [9]. In this section genuine errors are taken as those within $\pm\pi/2$ of the hypothesized angle μ_0, and the associated test is referred to as *selected* t-*test*. Without loss of generality, all simulations use $\mu_0 = 0$.

3.5.1 Receiver Operating Characteristics (ROC)

Fig. 6 shows the ROC curves for different tests. Here, the vMUM test has a higher P_d than all other tests for all P_{fa}. In particular, for $P_{fa} = 0.05$, the test power is $P_d = 0.93$ for vMUM and $P_d = 0.89$ for vMM3 when averaging across all the Monte Carlo simulations. In some cases the difference in performance is much larger, especially when the mean angle μ is close to μ_0. For instance, in simulations where $|\mu - \mu_0| = \pi/50$ (i.e., 3.6°), one has $P_d = 0.40$ for vMUM and $P_d = 0.17$ for vMM3, while for $|\mu - \mu_0| = \pi/15$ (i.e., 12°), one has $P_d = 0.90$ for vMUM and $P_d = 0.64$ for vMM3. The standard vM test has a relatively poor performance ($P_d = 0.82$ for $P_{fa} = 0.05$). For small $|\mu - \mu_0|$ the performance is particularly poor, with $P_d = 0.11$ for $|\mu - \mu_0| = \pi/15$ (i.e., 12°). The t-test has a poor performance. The selected t-test is, on average, even worse. However, if one only considers cases where $|\mu - \mu_0| < \pi/4$, then the test performs on a par with the vM test. In other words, if it can be reasonably assumed that the mean direction is close to the hypothesized direction, then the selected t-test performs on a par with the vM test.

3.5.2 Performance as a Function of p_2 and p_3

Fig. 7 shows the performance of the vM, vMM3, and vMUM tests as a function of p_2 and p_3. All tests have

essentially the same power when $p_2 = 0$ and $p_3 = 0$. As p_2 increases, the power of the vM test drops considerably while both the vMM3 and vMUM tests perform well. As p_3 increases, the power of the vMM3 test starts dropping while the power of the vMUM test remains about the same. In conclusion, the vMUM test performs on a par or better than the other tests and should be preferred whenever front/back confusions and/or random errors may be present in the data.

3.6 Two-Sample Tests

Given two data samples, θ_X of length N_X and θ_Y of length N_Y, consider the following tests:

$$\begin{cases} H_0 : \mu_X = \mu_Y = \mu \\ H_1 : \mu_X \neq \mu_Y \end{cases} \text{or} \quad \begin{cases} H_0 : \kappa_X = \kappa_Y = \kappa \\ H_1 : \kappa_X \neq \kappa_Y \end{cases}. \tag{24}$$

Here, the a priori probabilities of H_0 and H_1 are unknown. The distribution parameters are also unknown and are not necessarily the same for the two distributions.

The GLRT associated to these tests is

$$2 \ln \frac{\max\limits_{\hat{\nu}_X, \hat{\nu}_Y, \hat{a}_X, \hat{a}_Y} f_X(\theta_X; \hat{\nu}_X, \hat{a}_X) f_Y(\theta_Y; \hat{\nu}_Y, \hat{a}_Y)}{\max\limits_{\tilde{\nu}, \tilde{a}_X, \tilde{a}_Y} f_X(\theta_Y; \tilde{\nu}, \tilde{a}_X) f_Y(\theta_Y; \tilde{\nu}, \tilde{a}_Y)} \underset{\hat{H}_0}{\overset{\hat{H}_1}{\gtrless}} \xi, \tag{25}$$

where ν_r is the parameter being tested (either μ or κ), with $r \in \{X, Y\}$, while a_r contains the remaining parameters (e.g., when testing the mean direction, $\nu_r = \mu_r$ and $a_r = [\kappa_r, p_{1r}, p_{2r}, p_{3r}]$).

The restricted MLE at the denominator is found numerically. As a starting point, the value of ν is taken as the vMUM-MLE estimate of the combined data set $\theta = \left[\theta_X^T, \theta_Y^T\right]^T$, while the vectors containing the remaining parameters, $\tilde{a}_r$, are taken as the restricted vMUM-MLE estimates of the two data samples separately. The maximization of the unrestricted MLE at the numerator is separable. Thus the optimal parameters $(\hat{\nu}_r, \hat{a}_r)$ are the unrestricted vMUM-MLE estimates of each distribution, which can be calculated using the method in Sec. 3.2.2. Similarly to the single-sample test, two starting points are used—the parameters of the restricted MLE and the vMUM-MLE estimates. The one with the higher likelihood value is then chosen.

3.7 Performance of Two-Sample Tests

For space reasons, only the performance of the two-sample test of mean directions is analyzed here. The proposed vMUM test is compared to the two-sample vM test with the same concentration parameter[1], which is, to the best of the authors' knowledge, the only comparable test available in the literature. For this reason, the data was drawn from two distributions with the same concentration parameter, $\kappa \sim \mathcal{U}(0, 100)$, and probabilities $p_2 \sim \mathcal{U}(0, 0.3)$, $p_3 \sim \mathcal{U}(0, 0.3)$, and $p_1 = 1 - p_2 - p_3$. In order to run a fair comparison, the vMUM test was amended so as to take into account that $a_X = a_Y$, which is a straightforward modification. The mean directions were set as $\mu_X = 0$ and $\mu_Y \sim \mathcal{U}(0, \pi)$, without loss of generality. The sample size was set to $N = 20$. The results of this comparison show that for $P_{fa} = 0.05$, the power of the vM test is $P_d = 0.76$, while the power of the vMUM test is $P_d = 0.90$.

[1]The two-sample vM test of mean direction can be found in Mardia and Jupp [12], page 132. Notice that Eq. (7.3.17) appears to contain an error. The LRT used in this paper is $2N(\hat{\kappa}_{12}(R_X + R_Y)/2 - \hat{\kappa}R - \log I_0(\hat{\kappa}_{12}) + \log I_0(\hat{\kappa}))$.

4 CASE STUDY

This section describes a large-scale auralized experiment that was held in London during the 2015 Summer Science Exhibition of the Royal Society. The objective of the experiment was to study the localization performance of a large number of untrained subjects in an informal setting and with little training. Furthermore, as explained in detail in the following, the study aimed at testing several hypothesis: (a) whether two different *head related transfer function* (HRTF) datasets led to a significant difference in localization performance, (b) whether two different sound samples led to a significant difference in localization performance, and (c) whether the localization performance was different for a free-field simulation and reverberant rooms with different ceiling heights. The latter objective was inspired by a study by Hartmann [2] where it was shown that a room with high ceiling resulted in poorer localization, despite having a higher *direct-to-reverberant ratio* (DRR) than rooms with lower ceiling. Hartmann used large rooms with long reverberation times. The objective here was to independently confirm this hypothesis in an auralized experiment and with typically-sized rooms.

4.1 Participants and Data Monitoring

A total of 893 subjects participated in the experiment. No personal information of individual subjects was collected. Their age varied greatly and no gender bias was observed.

The data associated with 40 of the subjects was removed because they declared that (a) they were deaf in one ear, or (b) they did not understand the task, or (c) they made a mistake in using the interface, or (d) because the subject was only playing with the apparatus rather than executing the experiment. An additional 27 tests were excluded because the equipment was incorrectly initialized. Another 22 tests were removed because the response was given in less than one second, which indicated that the subject touched the interface without engaging in the test or by mistake. In case a subject performed the task under identical conditions more than once, the additional data points were also excluded. The above data selection reduced the number of subjects from 893 to 844, and the total number of tests reduced from 1979 to 1655.

4.2 Apparatus and Procedure

The apparatus consisted of a circular rotating platform with a diameter of about 70 cm and a height of 20 cm. The subjects stood on the rotating platform and they could freely turn themselves by applying force on a stationary wheel. The wheel was positioned in the center of the platform at a height of about 94 cm. Subjects wore a pair of Bang &

Olufsen BeoPlay H6 headphones connected to an iPad Air that was mounted on a pole in front of them at eye level.

The iPad's motion sensor was used to measure the rotation of the platform. In order to assess the accuracy of the motion sensor, the iPad was turned 10 times around itself and back, which gave an average deviation of about 2° from the expected 180°. Leaving the iPad stationary on a stable horizontal surface gave an average maximum drift of 0.67° over 10 repetitions. Please notice that the iPad's motion sensor was used as the reference for both the audio rendering and for recording the subjects' angular response, and therefore a drift of the motion sensor will have the same effect on both.

The subjects controlled the experiment using a simple custom-made GUI displayed on the iPad Air (see [10] for screenshots of the GUI). They could choose between two conditions—an anechoic condition and a reverberant condition the details of which are described in the next section[2]. They could run the conditions in any desired order and any number of times.

The subjects were asked to remain still and to keep looking at the iPad which displayed a real-time animation of the rotation of the platform to hold their attention. Once the audio started, their task was to rotate the platform until the sound source appeared to be in front of them. Subjects were told that they could take as long as needed to make a decision. The audio sample was looped until they recorded their decision. Subjects recorded their decision by tapping on a button stating "Touch here when the sound source appears to be in front of you" on the GUI. The GUI would then show their performance in terms of angular error.

It is worth observing here that the apparatus and measurement system is of such a specific nature that one should be careful in making conclusions in absolute terms. The data analysis will thus focus on the relative differences between conditions, for which the impact of the apparatus and measurement system can be factored out.

4.3 Sound Stimuli

Two anechoic sound samples from the "Music for Archimedes" CD were used. One was track 4, a female speech sample, and the other was track 26, a sample of an African percussion instrument. The two samples were cut at 28 seconds and 25 seconds, respectively, in order to reduce memory spooling. The levels were manually adjusted until the two sound samples had the same perceived loudness, which resulted in the speech sample being reduced by 3 dB with respect to the original level.

Two HRTF datasets were used—the MIT KEMAR mannequin [17] and subject number 58 of the CIPIC database [18], both having a 5° horizontal resolution. For every subject one of the two HRTF datasets and one of the two anechoic samples was chosen at random.

The anechoic signal was convolved with HRTFs using time-domain filters running in real-time on the iPad. When the platform (and thus the iPad) rotated, the coefficients of the filters were updated accordingly in real-time. In other words, the iPad was used as a head tracker in this experiment. As mentioned earlier, subjects were asked to remain as still as possible and to keep looking at the iPad, but their head and body were not physically restrained. In updating the HRTF filters, no interpolation was used. No audible artifacts were reported, owing to the slow rotation speed allowed by the rotating platform.

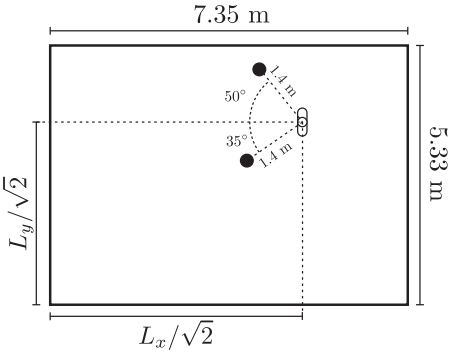


Fig. 8. Setup of the reverberant simulation. The black circles denote the position of the sound sources (only one source playing at any one time).

The subjects could choose between two conditions—an anechoic condition, where the sound source was placed in free field at the same height as the listener and at a distance of 1.4 m, and a reverberant condition, where the room acoustic response was simulated in real-time using a *scattering delay network* (SDN) [19]. SDN was chosen to generate the reverberation because of its ability to reproduce faithfully the important physical (e.g., early reflections, reverberation time) and perceptual features (e.g., normalized echo density) while running in real-time.

Table 1 lists the three room setups that were used and in each case shows the value of the DRR and the *reverberation time* (RT60). The dimensions of the "typical room" are ITU-R-compliant (BS.1116-1) and are identical to those of the "high reverberation" case. The "high reverberation" and "high ceiling" cases have the same T_{60}. The listener and sound sources were placed in the room as depicted in Fig. 8. The setup was chosen so as to be simple to describe, while at the same time avoiding the occurrence of sweeping echoes, which have been shown to occur in rectangular rooms with regular setups [20]. For each test, one of the three room setups and one of the two sound source positions was selected at random.

The frequency response of the Bang & Olufsen BeoPlay H6 headphones was equalized via monophonic minimum-phase inverse filters provided by the manufacturer.

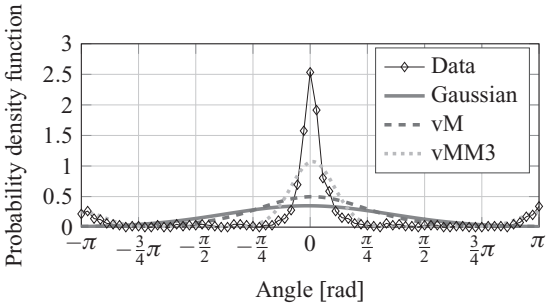
4.4 Model Comparison

Fig. 9 shows the empirical PDF of the localization errors under the anechoic condition ($N = 751$) and the result of fitting various statistical models to the data. Fig. 9a includes the Gaussian model and the vM model, both with ML estimate of the parameters, and the vMM3 model with

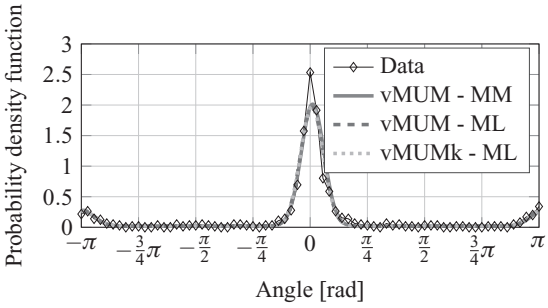
[2]A third condition with a sound source and a single echo was also included, but results were not included in this paper for space reasons.

Table 1. Characteristics of the room simulation in the reverberant condition.

Condition	Room width	Room length	Room height	Wall absorp.	RT60	DRR
Typical room	7.35 m	5.33 m	2.5 m	0.36	0.30 s	1.0 dB
High reverb.	7.35 m	5.33 m	2.5 m	0.30	0.45 s	0.2 dB
High ceiling	7.35 m	5.33 m	8.0 m	0.36	0.45 s	4.5 dB



(a) Data fitting with Gaussian, vM and vMM3 models.



(b) Data fitting with vMUM model.

Fig. 9. Empirical PDF of the anechoic condition and fitting of the statistical models discussed in this paper. The sound source is positioned at 0°.

MM estimate of the parameters. It is clear that the Gaussian model fits the data poorly. The Pearson correlation between the empirical PDF and the Gaussian PDF, used here as a measure of goodness-of-fit, is 0.42. The vM model has an improved fit to the data but does not account for the front/back confusions, and the model's concentration is insufficient for frontal sound sources (Pearson correlation 0.55). The vMM3 model achieves a better fit in respect of the front/back confusions (Pearson correlation 0.85). However, also in this case, the model fails to represent the higher concentration of the data, which is due to the fact that it is trying to fit the uniformly distributed errors by means of the front and back vM marginal distributions. By including the error marginal distribution explicitly, the vMUM model is able to provide a better fit to the data than the other models investigated, as shown in Fig. 9b. The Pearson correlation for both the vMUM-MLE and vMUM-MME models is 0.97.

In the vMUM model, the concentration parameters for the frontal and front/back confusions vM marginal distributions are identical. However, it could be hypothesized that front/back confusions are associated with a higher uncertainty and thus a lower concentration. For this reason, an additional model allowing for different concentration parameters was also investigated. This model was termed vMUMk and is shown in Fig. 9. As may be observed, the vMUMk and vMUM models fit the data equally well, with vMUMk also having a Pearson correlation of 0.97. Therefore, owing to its simplicity and the availability of closed-form MM estimators, the vMUM model remains the preferred model here.

4.5 Results

4.5.1 Angular Error and Concentration Parameter

Table 2 presents a summary of the statistics for the angular error, i.e., the difference between the response angle and true angle. The probability of identifying the frontal source is $p_1 = 0.76$, while the probability of experiencing a front/back confusion is $p_2 = 0.09$. The probability of making a uniformly distributed decision, or, in other words, a mistake, is $p_3 = 0.15$. From a frequentist point of view, 15% of subjects made a mistake. Notice that this includes cases where the subject made a mistake that was close to the correct direction (or to the opposite direction) by chance. These probabilities are similar across individual conditions.

In the anechoic case, the mean directional error is +1.4°. Using the single-sample test proposed in Sec. 3.4, the hypothesis $H_0 : \mu = 0$ can be rejected at the 0.05 significance level. The p-value, i.e. $p = \int_{L(\theta)}^{+\infty} f_L(l; H_0)dl$, is $p < 0.001$, which indicates a strongly significant result.

This rightward bias has been observed before in the literature [21] and implies some kind of physiological asymmetry in the auditory system itself. While the present experiment supports the findings in the literature, it cannot be ruled out that the bias observed here is due to systematic experimental errors such as asymmetries in the headphones or in the HRTF datasets.

Considering all the conditions together, the KEMAR and CIPIC datasets have a mean directional error of +2.3° and +1.1°, respectively. The two-sample test proposed in Sec. 3.6 reveals that the difference is statistically significant ($p = 0.02$). Furthermore, there is a borderline significant trend in comparing the concentration parameters ($p = 0.10$), with the KEMAR having larger concentration than CIPIC. This difference is strongly statistically significant if one considers the room simulation with the typical room setup alone ($p<0.001$).

The data indicates that speech yields a stronger rightward bias than the percussion instrument sound. Across all conditions there is only a borderline significant trend ($p = 0.11$). The difference is not significant in the anechoic case ($p = 0.99$) but is significant in the reverberant

Table 2. Statistics of the angular error. Positive values of μ indicate a rightward bias. The first column denotes the number of samples, *N*. The following 5 columns denote the vMUM-MLE estimates. The following 4 groups of 3 columns analyze differences in mean and concentration parameter between the KEMAR and CIPIC HRTF datsets, and mean and concentration parameter between speech and percussions audio samples. The p-values are calculated using the proposed two-sample vMUM tests. Values in boldface indicate statistical significance at the 0.05 significance level.

Condition	N	μ	κ	p_1	p_2	p_3	μ KEM.	μ CIP.	p	κ KEM.	κ CIP.	p	μ spe.	μ per.	p	κ spe.	κ per.	p
All combined	1655	1.7°	42.5	0.76	0.09	0.15	2.3°	1.1°	**0.02**	46.8	39.4	0.10	2.1°	1.3°	0.11	41.6	43.6	0.65
Anechoic	751	1.4°	40.8	0.76	0.10	0.14	1.7°	1.2°	0.51	44.3	37.9	0.31	1.4°	1.4°	0.99	43.1	38.9	0.52
Reverberant	904	1.9°	43.9	0.76	0.08	0.16	2.8°	1.1°	**0.01**	49.8	40.4	0.14	2.6°	1.2°	**0.04**	40.9	47.2	0.31
Rev. typical	312	1.9°	40.2	0.78	0.08	0.14	2.5°	1.3°	0.33	66.4	32.4	**0.00**	2.8°	0.9°	0.10	41.5	40.2	0.88
Rev. high reverb.	296	3.1°	46.2	0.74	0.07	0.18	3.6°	2.3°	0.25	41.6	55.5	0.27	4.4°	2.1°	0.06	31.1	59.7	**0.02**
Rev. high ceiling	296	0.8°	50.1	0.77	0.08	0.15	2.0°	−0.6°	**0.02**	49.5	53.9	0.72	0.9°	0.7°	0.87	54.5	46.0	0.48

conditions ($p = 0.04$). To the best of the authors' knowledge, this bias has not been observed before.

In the high reverberation condition, there is a statistically significant difference between speech and percussion sounds ($p = 0.02$), with percussion sounds having a larger concentration than speech. This suggests that, in the presence of reverberation, percussion sounds are easier to localize in comparison to the speech signal, which is consistent with results in the literature [1].

Finally, results of this experiment not included in Table 2 show that a larger number of subjects experience a front/back confusion in cases where the sound source starts behind the subject. Indeed, for tests where the initial look direction is less than 90° away from the source, the vMUM-MLE parameters are $p_1 = 0.79, p_2 = 0.06, p_3 = 0.15$, while for tests where the initial look direction is more than 90° away from the source, the vMUM-MLE parameters are $p_1 = 0.74, p_2 = 0.11, p_3 = 0.15$, which shows that the percentage of front/back confusions has nearly doubled. The percentage of front/back confusions increases even more as the initial look direction approaches 180°, with $p_2 = 0.25$ for initial look directions larger than 160° from the source.

The results presented have not been corrected for multiple comparisons.

4.5.2 Time to Complete Experiment

Subjects could choose whether to run the anechoic or the reverberant conditions first. Unsurprisingly, there was a clear learning effect, with the first condition taking much longer than the second. Those who started with the anechoic condition took on average 30.4 s for the anechoic condition and 23.4 s for the reverberant condition. Conversely, those who started with the reverberant condition took on average 22.3 s for the anechoic condition and 31.7 s for the reverberant condition. In order to compare the two conditions fairly, the dataset was pruned until it contained an equal number of subjects who started with each one[3].

The results are reported in Table 3 and show that subjects take longer in the high ceiling condition than in all the other conditions. Two-sample *t*-tests with right tail reveal that the result is statistically significant in all cases.

This result may seem surprising. Indeed, the high ceiling condition has the highest DRR of all three (see Table 1) and has the same reverberation time as the high reverberation condition. In contrast, the high reverberation condition took about the same time as the typical room condition (25.9 s and 25.5 s, respectively). As mentioned earlier this phenomenon has been observed before in the work of Hartmann [2]. Hartmann hypothesized that it is easier for the auditory system to localize the sound source if the ceiling reflection arrives before the onset of the precedence effect. The rationale is that there is an additional reflection that is temporally fused with the line of sight component and is azimuthally co-directional with the sound source. While Hartmann supported his conclusions based on experiments in a large room with long reverberation times the results presented here suggest that the same phenomenon also arises in small, ITU-R-compliant rooms.

Finally, results of this experiment not included in Table 3 reveal that the speech sample took 30.5 s on average, which is significantly longer than the percussion sample, 22.9 s ($p<0.001$). This difference could be due to subjects actively paying attention to what was being said in the speech sample, or to the fact that percussive signals with sharp onsets are easier to localize [1].

5 CONCLUSIONS

This paper has presented a circular statistical model for sound localization experiments that is robust to the presence of front/back confusions and random errors. Closed-form MM and ML estimators were presented, with the MM estimators outperforming the ML estimators for small sample sizes. Single-sample and two-sample tests were proposed and were shown to have a performance on a par with or

[3]All subjects who started with the reverberant condition were kept (because there were fewer of them). Then, of the subjects starting with the anechoic condition, only those who carried out the experiment at a similar time to the ones starting with the reverberant condition were kept. This particular choice was made because it allows to factor out possible time dependances of the experiment (e.g., at peak times a queue would sometimes form that could put subjects under pressure to complete the test faster).

Table 3. Statistics of the time to complete the test. The dataset was pruned to factor out learning effects (see text). The first two columns denote the number of samples, *N* and the mean of the time to complete the test, respectively. The last four columns present the p-value of the interactions across conditions using a two-sample *t*-test with right tail. Values in boldface indicate statistical significance at the 0.05 significance level.

Condition	N	Mean	Anech.	Reverb.typical	High. reverb.	High ceiling
All combined	776	26.9 s	–	–	–	–
Anechoic	388	26.4 s	–	0.33	0.41	0.98
Reverberant	388	27.4 s	0.24	–	–	–
Rev. typical	140	25.6 s	0.67	–	0.57	0.97
Rev. high reverb.	122	25.9 s	0.59	0.43	–	0.96
Rev. high ceiling	126	30.9 s	**0.02**	**0.03**	**0.04**	–

better than prior art tests. The performance of the proposed data modelling was substantially better in the presence of front/back confusions and/or random errors. These errors are common if the experiment is auralized (especially when using non-individualized HRTFs), if the subjects are untrained, if the auditory system is intentionally interfered with, or if the sound stimuli are particularly hard to localize, e.g., in echolocation tasks.

The statistical framework was used to analyze the data of a large-scale auralized experiment. Using the proposed statistical framework, it was shown that (a) a rightward bias is present in the subjects' localization responses, (b) the speech sample is found to have a larger rightward bias than the percussion sample under reverberant conditions, (c) the KEMAR HRTF dataset results in a larger bias than the CIPIC HRTF dataset, (d) under high reverberation, responses given with the percussions sample had a higher concentration than with the speech sample. Furthermore, an analysis of the time it took to complete the experiment showed that (a) the speech sample took significantly longer to localize than the percussions sample, and (b) it took significantly longer to localize sound in a simulated room with high ceiling.

Matlab routines implementing the proposed statistical estimators and statistical tests, together with the data of the listening experiment, are available online[4].

6 ACKNOWLEDGMENTS

The authors would like to thank Christine Evers, Sina Hafezi, Hamza Javed, Alastair Moore, Richard Stanton, Costas Yiallourides, Ray Thompson and all the Fellows of the DREAMS ITN for helping carry out the experiment, and Neofytos Kaplanis and Alastair Moore for their help and support.

7 REFERENCES

[1] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization* (Cambridge, MA: MIT Press, 1997).

[2] W. M. Hartmann, "Localization of Sound in Rooms," *J. Acoust. Soc. Amer.*, vol. 74, no. 5, pp. 1380–1391 (1983).

[3] S. Bech and N. Zacharov, *Perceptual Audio Evaluation: Theory, Method and Application* (Hoboken, NJ: John Wiley & Sons, 2006).

[4] D. S. Brungart and B. D. Simpson, "Effects of Bandwidth on Auditory Localization with a Noise Masker," *J. Acoust. Soc. Amer.*, vol. 126, no. 6, pp. 3199–3208 (2009).

[5] S. Carlile, K. Balachandar, and H. Kelly, "Accommodating to New Ears: The Effects of Sensory and Sensory-Motor Feedback," *J. Acoust. Soc. Amer.*, vol. 135, no. 4, pp. 2002–2011 (20140.

[6] M. F. Mueller, A. Kegel, S. M. Schimmel, N. Dillier, and M. Hofbauer, "Localization of Virtual Sound Sources with Bilateral Hearing Aids in Realistic Acoustical Scenes," *J. Acoust. Soc. Amer.*, vol. 131, no. 6, pp. 4732–4742 (2012).

[7] E. De Sena, H. Hacıhabiboğlu, and Z. Cvetković, "Analysis and Design of Multichannel Systems for Perceptual Sound Field Reconstruction," *IEEE Trans. Audio, Speech, Language Proc.*, vol. 21, pp. 1653–1665 (2013 Aug.).

[8] E. H. Langendijk and A. W. Bronkhorst, "Contribution of Spectral Cues to Human Sound Localization," *J. Acoust. Soc. Amer.*, vol. 112, no. 4, pp. 1583–1596 (2002).

[9] S. Carlile, P. Leong, and S. Hyams, "The Nature and Distribution of Errors in Sound Localization by Human Listeners," *Hearing Research*, vol. 114, no. 1, pp. 179–196 (1997).

[10] E. De Sena, N. Kaplanis, P. A. Naylor, and T. van Waterschoot, "Large-Scale Auralized Sound Localisation Experiment," presented at the *AES 60th International Conference: DREAMS (Dereverberation and Reverberation of Audio, Music, and Speech)* (2016 Jan.), paper 2-5.

[11] D. Pelegrín-García, M. Rychtáriková, C. Glorieux, and B. F. Katz, "Interactive Auralization of Self-Generated Oral Sounds in Virtual Acoustic Environments for Research in Human Echolocation," *Proc. of Forum Acusticum* (Kracow, Poland) (2014).

[12] K. V. Mardia and P. E. Jupp, *Directional Statistics* (Hoboken, NJ: John Wiley & Sons, 2009).

[13] N. I. Fisher, *Statistical Analysis of Circular Data* (Cambridge, UK: Cambridge University Press, 1995).

[14] S. D. Grimshaw, D. G. Whiting, and T. H. Morris, "Likelihood Ratio Tests for a Mixture of Two von Mises Distributions," *Biometrics*, vol. 57, no. 1, pp. 260–265 (2001).

[4] http://desena.org/dirstat

[15] D. Best and N. I. Fisher, "Efficient Simulation of the von Mises Distribution," *Applied Statistics*, pp. 152–157 (1979).

[16] S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory* (New York, NY: Pearson Education, 1998).

[17] B. Gardner and K. Martin, "HRFT Measurements of a KEMAR Dummy-Head Microphone,"Tech. Rep., MIT Media Lab Perceptual Computing - Technical Report 280 (May 1994).

[18] V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano, "The CIPIC HRTF Database," *IEEE Workshop Appl. of Signal Proc. to Audio and Acoustics* (New Paltz, NY, USA), pp. 99–102 (2001).

[19] E. De Sena, H. Hacıhabiboğlu, Z. Cvetković, and J. Smith, "Efficient Synthesis of Room Acoustics via Scattering Delay Networks," *IEEE/ACM Trans. Audio, Speech, Language Proc.*, vol. 23, pp. 1478–1492 (2015 Sep.).

[20] E. De Sena, N. Antonello, M. Moonen, and T. van Waterschoot, "On the Modeling of Rectangular Geometries in Room Acoustic Simulations," *IEEE/ACM Trans. Audio, Speech, Language Proc.*, vol. 23, pp. 774–786 (2015 Apr.).

[21] A. Dufour, P. Touzalin, and V. Candas, "Rightward Shift of the Auditory Subjective Straight Ahead in Right- and Left-Handed Subjects," *Neuropsychologia*, vol. 45, no. 2, pp. 447–453 (2007).

[22] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions: with Formulas, Graphs, and Mathematical Tables* (North Chelmsford, MA: Courier Corporation, 1964).

A.1 PROOF OF CENTRAL TRIGONOMETRIC MOMENTS OF Θ

This appendix derives the expression for the central trigonometric moments of Θ. Starting with the central cosine moments:

$$\begin{aligned}\alpha_l &= \frac{p_1}{2\pi I_0(\kappa)}\int_0^{2\pi}\cos(l(\vartheta-\mu))e^{\kappa\cos(\vartheta-\mu)}d\vartheta \\ &\quad+\frac{p_2}{2\pi I_0(\kappa)}\int_0^{2\pi}\cos(l(\vartheta-\mu))e^{-\kappa\cos(\vartheta-\mu)}d\vartheta \\ &\quad+\frac{p_3}{2\pi}\int_0^{2\pi}\cos(l(\vartheta-\mu))d\vartheta,\end{aligned}$$

The above integrals can be simplified by substituting $\vartheta' = \vartheta - \mu$. Furthermore, since both $\cos(l\vartheta)$ and $e^{\kappa\cos(\vartheta')}$ are 2π-periodic, then the integral over a period interval is always the same:

$$\begin{aligned}\alpha_l &= \frac{p_1}{2\pi I_0(\kappa)}\int_0^{2\pi}\cos(l\vartheta')e^{\kappa\cos\vartheta'}d\vartheta \\ &\quad+\frac{p_2}{2\pi I_0(\kappa)}\int_0^{2\pi}\cos(l\vartheta')e^{-\kappa\cos\vartheta'}d\vartheta' \\ &\quad+\frac{p_3}{2\pi}\int_0^{2\pi}\cos(l\vartheta')d\vartheta'\end{aligned}$$

The third integral above can be written as $\int_0^{2\pi}\cos(l\vartheta')d\vartheta' = 2\pi\delta_l$. From Eq. (9.6.19) in Abramowitz and Stegun [22] $I_n(z) = \frac{1}{\pi}\int_0^{\pi}\cos(n\vartheta)e^{z\cos\vartheta}d\vartheta$. Here, since the integrand is even and 2π-periodic, then $\int_0^{2\pi}\cos(n\vartheta)e^{z\cos\vartheta}d\vartheta = 2\pi I_n(z)$. Therefore

$$\alpha_l = \frac{p_1}{I_0(\kappa)}I_n(\kappa)+\frac{p_2}{I_0(\kappa)}I_n(-\kappa)+p_3\delta_l$$

Eq. (9.6.30) in Abramowitz and Stegun [22] states $I_\nu(ze^{im\pi}) = e - im\nu\pi I_\nu(z)$ with m integer, which implies for $m = -1$ and ν integer that $I_n(-z) = (-1)^n I_n(z)$. Thus the cosine central trigonometric moments are given by

$$\alpha_l = \frac{I_n(\kappa)}{I_0(\kappa)}(p_1+(-1)^n p_2)+p_3\delta_l.$$

Regarding the sine central trigonometric moments, using similar considerations used for the cosine moments:

$$\begin{aligned}\beta_l &= \mathbb{E}[\sin(l(\Theta-\mu))] \\ &= \frac{p_1}{2\pi I_0(\kappa)}\int_0^{2\pi}\sin(l\vartheta')e^{\kappa\cos\vartheta'}d\vartheta \\ &\quad+\frac{p_2}{2\pi I_0(\kappa)}\int_0^{2\pi}\sin(l\vartheta')e^{-\kappa\cos\vartheta'}d\vartheta' \\ &\quad+\frac{p_3}{2\pi}\int_0^{2\pi}\sin(l\vartheta')d\vartheta'\end{aligned}$$

Since all the integrands are odd functions and the integration interval is across a full period, then $\beta_l = 0$.

A.2 PROOF OF CENTRAL TRIGONOMETRIC MOMENTS OF Φ

This appendix derives the expression for the central trigonometric moments of Φ. Starting with the central cosine moments:

$$\begin{aligned}\alpha_l^w &= \frac{p_w}{4\pi I_0(\kappa)}\int_0^{2\pi}\cos(l(\varphi-2\mu)) \\ &\quad\times\left(e^{\kappa\cos(\frac{\varphi}{2}-\mu)}+e^{-\kappa(\frac{\varphi}{2}-\mu)}\right)d\varphi \\ &\quad+\frac{1-p_w}{2\pi}\int_0^{2\pi}\cos(l(\varphi-2\mu))d\varphi,\end{aligned}$$

The third integral can be simplified by substituting $\varphi' = \varphi - 2\mu$ and noticing that $\cos(l\varphi')$ is 2π-periodic, thus $\int_0^{2\pi}\cos(l(\varphi-2\mu))d\varphi = \int_0^{2\pi}\cos(l\vartheta')d\vartheta' = 2\pi\delta_l$. The first and second integral, on the other hand, can be simplified by substituting $\varphi' = \frac{\varphi}{2} - \mu$:

$$\begin{aligned}\alpha_l^w &= \frac{p_w}{4\pi I_0(\kappa)}\int_{-\mu}^{\pi-\mu}\cos(2l\varphi')) \\ &\quad\times\left(e^{\kappa\cos\varphi'}+e^{-\kappa\cos\varphi'}\right)2d\varphi'+(1-p_w)\delta_l.\end{aligned}$$

The function $\cos(2l\varphi'))$ is π-periodic. Furthermore, $e^{\kappa\cos\varphi'}+e^{-\kappa\cos\varphi'} = g(\varphi')$ is also π-periodic. Indeed, $g(\varphi' + k\pi)$ for k even ($k = 2m$) is given by $g(\varphi'+2m\pi) = e^{\kappa\cos\varphi'}+e^{-\kappa\cos\varphi'} = g(\varphi')$. The same holds for k odd ($k = 2m+1$): $g(\varphi'+(2m+1)\pi) = e^{-\kappa\cos\varphi'}+e^{\kappa\cos\varphi'} =$

$g(\varphi')$. Therefore

$$\begin{aligned}\alpha_l^w &= \frac{p_w}{2\pi I_0(\kappa)}\int_0^{\pi}\cos(2l\varphi'))\left(e^{\kappa\cos\varphi'}+e^{-\kappa\cos\varphi'}\right)d\varphi' \\ &\quad +(1-p_w)\delta_l \\ &= \frac{p_w}{2\pi I_0(\kappa)}\left[\int_0^{\pi}\cos(2l\varphi'))e^{\kappa\cos\varphi'}d\varphi'\right. \\ &\quad \left.+\int_0^{\pi}\cos(2l\varphi')e^{-\kappa\cos\varphi'}d\varphi'\right]+(1-p_w)\delta_l \\ &= \frac{p_w}{2\pi I_0(\kappa)}\left[\pi I_{2l}(\kappa)+\pi I_{2l}(-\kappa)\right]+(1-p_w)\delta_l.\end{aligned}$$

where the last step used Eq. (9.6.19) in Abramowitz and Stegun [22]. Since $I_n(-z) = (-1)^n I_n(z)$ (see Appendix A.1), then $I_{2l}(-z) = (-1)^{2l} I_{2l}(z) = I_{2l}(z)$ and thus

$$\alpha_l^w = p_w\frac{I_{2l}(\kappa)}{I_0(\kappa)}+(1-p_w)\delta_l.$$

Regarding the sine central trigonometric moments, using similar observations considered in Appendix A.1 for Θ one obtains $\beta_l^w = 0$.

THE AUTHORS


Enzo De Sena


Patrick Naylor


Mike Brookes


Toon van Waterschoot

Enzo De Sena is a Lecturer in the Institute of Sound Recording at the University of Surrey, UK. He received his B.Sc. degree in 2007 and his M.Sc. degree (cum laude) in 2009, both from the Università degli Studi di Napoli "Federico II", Italy, in telecommunication engineering. In 2013 he received his Ph.D. degree in electronic engineering from King's College London, UK, where he also served as a Teaching Fellow from 2012 to 2013. Between 2013 and 2016 he was a postdoctoral Research Fellow at the Katholieke Universiteit Leuven, Belgium. He held visiting positions at Stanford University, California (2013), Aalborg University, Denmark (2014/2015), and Imperial College London, UK (2016/2017). He is a former Marie Curie fellow, and his current research interests include room acoustics modelling, multichannel audio systems, microphone beamforming, and binaural modelling.

•

Patrick Naylor is a member of academic staff in the Department of Electrical and Electronic Engineering at Imperial College London. He received the BEng degree in electronic and electrical engineering from the University of Sheffield, UK, and the PhD. degree from Imperial College London, UK. His research interests are in the areas of speech, audio and acoustic signal processing. He has worked in particular on adaptive signal processing for speech dereverberation, blind multichannel system identification and equalization, acoustic echo control, speech quality estimation and classification, single and multichannel speech enhancement, and speech production modelling with particular focus on the analysis of the voice source signal. In addition to his academic research, he enjoys several fruitful links with industry in the UK, USA, and in Europe. He is the past-Chair of the IEEE Signal Processing Society Technical Committee on Audio and Acoustic Signal Processing and a director of the European Association for Signal Processing (EURASIP). He has served as an associate editor of *IEEE Signal Processing Letters* and is currently a senior area editor of *IEEE Transactions on Audio Speech and Language Processing*.

•

Mike Brookes is a Reader (Associate Professor) in signal processing in the Department of Electrical and Electronic Engineering at Imperial College London. After graduating in mathematics from Cambridge University in 1972, he worked at the Massachusetts Institute of Technology and, briefly, the University of Hawaii before returning to the UK and joining Imperial College in 1977. Within the area of speech processing, he has concentrated on the modelling and analysis of speech signals, the extraction of features for speech and speaker recognition, and on the enhancement of poor quality speech signals. He is the primary author of the VOICEBOX speech processing toolbox for MATLAB. Between 2007 and 2012 he was the Director of the Home Office sponsored Centre for Law Enforcement Audio Research (CLEAR), which investigated techniques for processing heavily corrupted speech signals. He is currently principal investigator of the E-LOBES project that seeks to develop environment-aware enhancement algorithms for binaural hearing aids.

•

Toon van Waterschoot received M.Sc (2001) and Ph.D (2009) degrees in electrical engineering, both from KU Leuven, Belgium, where he is currently a tenure-track Assistant Professor. He has previously also held teaching and research positions at Delft University of Technology in The Netherlands and the University of Lugano in Switzerland. His research interests are in signal processing, machine learning, and numerical optimization, applied to acoustic signal enhancement, acoustic modeling, audio analysis, and audio reproduction. He has been serving as an Associate Editor for the *Journal of the Audio Engineering Society* and for the *EURASIP Journal on Audio, Music, and Speech Processing*, and as a Guest Editor for Elsevier *Signal Processing*. He is a Director of the European Association for Signal Processing (EURASIP), a Member of the IEEE Audio and Acoustic Signal Processing Technical Committee, and a Founding Member of the EAA Technical Committee in Audio Signal Processing. He was the General Chair of the 60th AES International Conference in Leuven, Belgium (2016).