Elsevier Editorial System(tm) for Hearing Research
Manuscript Draft

Manuscript Number:

Title: A novel application of a centre of mass approach to the statistical analysis of localisation performance

Article Type: Methodological paper

Keywords: auditory; spatial; directional; MANOVA

Corresponding Author: Mr Mark Edmondson-Jones,

Corresponding Author's Institution: MRC Institute of Hearing Research

First Author: Mark Edmondson-Jones, MSc

Order of Authors: Mark Edmondson-Jones, MSc; Mark Edmondson-Jones; Samuel Irving, PhD; David R Moore, PhD; Deborah A Hall, PhD

Abstract: Sound localisation is one of the key roles for listening, and measuring localisation performance is a mainstay of the hearing research laboratory. Such measurements consider both accuracy and, for incorrect trials, the size of the error. In terms of error analysis, localisation studies have frequently used general univariate techniques in conjunction with either mean signed or unsigned error measurements. This approach can make inappropriate distributional assumptions and so more suitable alternatives based on directional statistics (e.g. based on von Mises distributed data) have also been used. However these are not readily computed using most commercially available, commonly used statistical software, and are generally only defined for simple experimental designs. We describe a novel use of a 'centre of mass' approach for describing localisation data jointly in terms of accuracy and size of error. This spatial method offers powerful, yet flexible, statistical analysis using standard multivariate analysis of variance (MANOVA).

Title: "A novel application of a centre of mass approach to the statistical analysis of localisation performance"

Suggested running head: "Analysis of localisation performance"

Authors: A. Mark Edmondson-Jones(*), Samuel Irving, David R. Moore, and Deborah A. Hall.

MRC Institute of Hearing Research, University Park, Nottingham, NG7 2RD, UK.

(*) Corresponding Author: A. Mark Edmondson-Jones, MRC Institute of Hearing Research, University Park, Nottingham, NG7 2RD, UK. Telephone: +44 (0)115 922 3431. E-mail: markej@ihr.mrc.ac.uk

Abstract

Sound localisation is one of the key roles for listening, and measuring localisation performance is a mainstay of the hearing research laboratory. Such measurements consider both accuracy and, for incorrect trials, the size of the error. In terms of error analysis, localisation studies have frequently used general univariate techniques in conjunction with either mean signed or unsigned error measurements. This approach can make inappropriate distributional assumptions and so more suitable alternatives based on directional statistics (e.g. based on von Mises distributed data) have also been used. However these are not readily computed using most commercially available, commonly used statistical software, and are generally only defined for simple experimental designs. We describe a novel use of a 'centre of mass' approach for describing localisation data jointly in terms of accuracy and size of error. This spatial method offers powerful, yet flexible, statistical analysis using standard multivariate analysis of variance (MANOVA).

I. Introduction

Localising the source of auditory objects in space is one of the key roles of the auditory system, and has thus generated a substantial amount of scientific interest (Middlebrooks and Green, 1991; Moore and King, 1999; McAlpine, 2005). Spatial hearing has a dual role. In addition to localising sounds, it is important in segregating sounds from noisy backgrounds and in complex auditory scenes (Hine, Martin *et al.*, 1994; Middlebrooks and Green, 1991). Furthermore, a growing number of studies use spatial sound attributes to investigate other aspects of perception and psychophysics such as auditory plasticity and learning (Kacelnik, Nodal, *et al.*, 2006; Wright and Zhang, 2006; Keuroghlian and Knudsen, 2007). These studies increasingly favour environmentally salient presentation paradigms, via either free-field or virtual auditory space, which provide a better indication of localisation performance than the lateralisation tasks used traditionally via headphone presentation. A number of different methods have been used to quantify localisation performance, with little consensus on which might be the most appropriate summary measure of performance or the most sensitive measure for detecting change over time.

This paper presents a novel application of a 'centre of mass' (CoM) measure (Mardia and Jupp, 2000) that is both appropriate and sensitive for summarising the error characteristics of localisation performance. The CoM approach is presented in the context of alternative measures of horizontal (azimuthal) plane localisation, and direct comparisons are made using both simulated and experimental data. Examples are taken from the auditory domain, but the methods apply equally to the visual domain.

II. Typical measures of localisation performance

Accuracy and error are common measures of localisation performance. Accuracy refers simply to the proportion of correct responses made at each sound source location. While accuracy lends itself to binary logistic statistical techniques, generally authors have used accuracy merely as a summary statistic to which standard analyses such as t-tests and repeated measures ANOVA are applied (Abel and Paik, 2004; Hine, Martin *et al.*, 1994; Kacelnik, Nodal, *et al.*, 2006; Parsons, Lanyon, *et al.*, 1999). Nevertheless, accuracy can be a rather blunt tool as it provides no indication about the size of the localisation errors. For example, localisation errors both at 15° and 150° away from the sound source are treated in the same way. Usually interest relates to any systematic bias in errors or change in their dispersion, for example corresponding to experimental condition. Accuracy measures are therefore often reported alongside corresponding error measures to summarise the magnitude of the incorrect localisations. However, when the testing arena contains sound sources that span -180° to +180°, incorrect responses made to sources close to these 'boundaries' may, in some schemes, lead to large errors. As an extreme example, a $1°$ change in an observation from $\theta = 180°$ to $\theta = -179°$ would be treated as a much larger movement of $359°$ in the response. This 'boundary', or 'wrapping', effect artificially inflates the observed variance of θ. Typically mean localisation errors will be reported, with correspondingly inflated standard errors. Therefore, this kind of error statistic may not provide a particularly sensitive measure of performance, particularly if there are many 'large' errors. Previous studies have mitigated this problem by using the mean *absolute,* or *unsigned,* error, so that the error values fall between 0° and 180°. However, these studies have used either parametric tests (Parsons, Lanyon, *et al.*, 1999; Zahorik, Bangayan, *et al.*, 2006) which incorrectly assume a normal distribution of the data, or non-parametric tests (Kacelnik, Nodal, *et al.*, 2006) which offer reduced statistical power.

Stimulus-response plots (e.g. in the form of a bubble plot; Kacelnik, Nodal, *et al.*, 2006) provide a common alternative approach to presenting performance. In this type of plot, response locations (in degrees) are plotted as a function of the source locations (in degrees). Correct responses lie along the diagonal and typically the size of the data point ('bubble') reflects the frequency of that stimulus-response pairing. Stimulus-response plots provide a visually appealing way to present both accuracy and distribution of errors, but while they are a powerful tool for visualisation their distributional characteristics make them problematic as a basis for statistical analysis. Data are typically analysed by linear regression where r = 1 indicates a linear relationship between source and response and systematic changes over time are analysed by tracking the change in the slope ('*response gain'*) and intercept(*'bias'*) (Hofman, Vlaming, *et al.*, 2002; Hofman, Van Riswick *et al.*, 1998; Kacelnik, Nodal, *et al.*, 2006; Van Wanrooij and Van Opstal, 2007). As an alternative to normal linear regression, bootstrap methods have also been employed to estimate the probability of obtaining the given stimulus-response data under a null hypothesis of no effect of the experimental condition (Van Wanrooij and Van Opstal, 2005). Another potential limitation is that stimulus-response plots do not readily permit the presentation of multiple listeners or repeated testing of the same listener in the same plot, and therefore are limited to mean data, or selected single-subject plots. Moreover they are also highly sensitive to boundary effects when used as a basis for data analysis, as we illustrate later.

Stimulus-response analysis has been expanded to encompass azimuthal and elevational localisation, by observing elevation and azimuthal errors separately using a double-pole azimuth-elevation coordinate (Van Wanrooij and Van Opstal, 2005; Van Wanrooij and Van Opstal, 2007). However, this method suffers from the same variety of the issues (e.g. wrapping) described above, as it effectively performs separate analyses per coordinate.

The key features that are generally of interest when considering localisation errors are a) the directional bias; and b) the dispersion of the response. None of the techniques outlined above directly describe either of these features and furthermore are confounded by the properties of directional data. In the next section, we describe the application of directional statistical methods to the problem of analysing auditory localisation data in a 360° plane which specifically address these aspects of performance errors.

III. Circular statistics

A directional response variable, such as localisation error (θ), has peculiar properties which do not apply in general to linear measurement variables. For example, it is a feature of a circular probability density (pdf) $f(\theta)$ that $f(\theta) = f(\theta \pm 360°) = f(\theta \pm 2 \times 360°) = \cdots$. For this reason, without loss of generality, we conventionally constrain θ to a 360° range. Here we assume $-180° < \theta \leq 180°$. Such angular constraints are not readily implemented if θ is treated as a 'linear' variate with θ taking any real value. In the case of parametric analyses the underlying model will generally indicate $f(\theta) \neq f(\theta + 360°)$, with potentially both being non-zero. This is clearly in conflict with the directional paradigm and, depending on the magnitude of departure from this assumption, may result in potentially inappropriate inferences.

Analysis of summary statistics, such as the mean error within a block of trials, might be performed instead. However, this approach conceals further issues. Extreme values of θ will have a disproportionate effect on the analysis, as not only is their statistical 'leverage' large, but, as discussed earlier, a potentially small angular change in the response variable close to the (arbitrary) boundary may result in a huge change in θ.   This 'boundary' effect will serve to artificially inflate the observed sample variance of θ and thereby the standard error of $\bar{\theta}$. Alternative approaches to such problems include the adoption of non-parametric techniques, such as Mann-Whitney's U and Kolmogorov-Smirnov's Z tests. Generally these tests may be robust to distributional departures from normality but their own assumptions may be infringed; specifically they remain sensitive to the problems encountered by wrapping at the boundaries. Non-parametric techniques may also suffer reduced power and, furthermore, the implementation of non-parametric methods with complex (e.g. multi-way factorial) experimental designs may be problematic.

A family of techniques exists specifically to deal with such directional statistics (Mardia and Jupp, 2000). These techniques have not been widely used in behavioural localisation analyses, but they have been adopted in other fields. For example, the use of Rayleigh tests in vector strength analysis, in order to assess the significance of phase-locking as a departure from a null hypothesis of phase uniformity in auditory neurophysiology (Goldberg and Brown, 1969;

Buunen and Rhode, 1978). In terms of localisation comparisons (where an increased response density at certain directions might be expected), an appropriate circular distribution is potentially the von Mises distribution (Mardia and Jupp, 2000), which might be considered the circular analogue of the normal distribution. The von Mises ($M(\mu, \kappa)$) distribution is expressed in terms of a mean direction, or location, μ and a concentration κ by the pdf

$$f(\theta; \mu, \kappa) = \frac{1}{360\, I_0(\kappa)} e^{\kappa \cos(\theta - \kappa)}$$

where here we denote θ in degrees. (I_0 denotes a 0 order modified Bessel function of the first kind.) A high concentration corresponds to a distribution with little dispersion, and the special case where $\kappa = 0$ corresponds to the uniform distribution. Examples of von Mises densities for a variety of values of μ and κ are illustrated in Figure 1.

[Figure 1 about here.]

We note that, for large values of κ, $M(\mu, \kappa) \to N\left(\mu, \frac{1}{\kappa}\right)$; i.e. for high-concentration circular distributions, the normal distribution may provide a convenient approximation.

Parametric Tests

In comparing various samples of circular data that are assumed to come from distributions $M(\mu_i, \kappa_i),\ i = 1, \dots, k$, circular statistics provide separate tests for the null hypotheses H_0: $\mu_1 = \mu_2 = \cdots = \mu_k$ assuming that $\kappa_1 = \kappa_2 = \cdots = \kappa_k$ and H_0: $\kappa_1 = \kappa_2 = \cdots = \kappa_k$ assuming that $\mu_1 = \mu_2 = \cdots = \mu_k$. These one-way tests of equality of mean direction and equality of concentration respectively provide circular analogues to the normal one-way

ANOVA and test of homogeneity of variance. Unfortunately though, they are not widely available in most statistical packages. The examples provided here were implemented using the 'circular' package in R version 2.6.1 (R Development Core Team, 2007) and also using a von Mises simulation algorithm implemented in Matlab version 7.6 (Best and Fisher, 1979).

Centre of Mass

Here we propose a simple analysis that recognises the spatial nature of the data using standard Multivariate Analysis of Variance (MANOVA) methods, and effectively analyses accuracy and error size simultaneously. Each response is treated spatially, effectively as a sound source location. Planar (bivariate) Cartesian coordinates, relative to the central location of the listener, for a single error response of θ, are derived as $X = (\sin\theta, \cos\theta)$. For a sample of n observations we can then define the sample mean as

$$\overline{X} = \left(\frac{1}{n}\sum_{i=1}^{n}\sin\theta_i, \frac{1}{n}\sum_{i=1}^{n}\cos\theta_i\right).$$

Assuming a unit mass at the end-point of each unit vector (corresponding to θ_i) the sample mean represents the sample's overall centre of mass. As a summary measure the centre of mass (or CoM) avoids the influence that boundary effects exert on the mean signed and unsigned error. Having converted the angular error response into bivariate coordinates, conventional MANOVA techniques are employed to analyse the data according to the corresponding experimental design. Thereby, the benefits of such methods (use with complex experimental models, unbalanced data, etc.) can be fully and simply exploited.

IV. Comparison of Methods

a. Simulated data

To compare the performance of a variety of statistical approaches, pairs of 100 observation samples were drawn from specified von Mises distributions. Such simulations were run repeatedly in order to investigate the proportion of cases where the null hypothesis was rejected, according to each test, at a significance level of $\alpha = 0.05$. In this manner estimates of the power of each statistical test in identifying differences in the underlying samples were obtained.

In each case one sample was drawn from a reference distribution of $M(0,5)$ and the other sample was drawn from an alternative distribution of $M(\mu, \kappa)$ where $\mu \epsilon [-22.5°, 22.5°]$ and $\kappa \epsilon (0,10]$. The sampled values from each distribution were then 'discretised' into 24 bins with $15°$ separation and midpoints $\{-165°,\ -150°,\ \ldots,\ 165°, 180°\}$ in order to better represent typical experimental paradigms, in our case mimicking 24 separate loudspeaker locations separated by 15° intervals.

These data required modification to enable the investigation of stimulus-response analysis characteristics. This was achieved by applying these directional errors (as generated above) to 24 source locations. For each sample, individual sources were used four times, except for the loudspeaker locations at -90°, 0°, 90° and 180° which were each used five times. This provided an optimally balanced distribution of the 100 source locations.

Each simulated set of directional data was subjected to the following tests:

1. Centre of Mass test (using Wilks' Λ);

2. von Mises test of equality of location;
3. von Mises test of equality of concentration;
4. (Parametric) univariate ANOVA with:
    a. Signed error
    b. Unsigned error (c.f. Parsons, Lanyon, *et al.*, 1999; Zahorik, Bangayan, *et al.*, 2006);
5. (Non-parametric) Mann-Whitney U test with:
    a. Signed error
    b. Unsigned error (c.f. Kacelnik, Nodal, *et al.*, 2006); and
6. Stimulus-Response analysis based on:
    a. Parametric ANOVA F test (c.f. Hofman, Van Riswick, *et al.*, 1998; Hofman, Vlaming, *et al.*, 2002; Kacelnik, Nodal, *et al.*, 2006; Van Wanrooij and Van Opstal, 2007)
    b. Bootstrap F statistic (c.f. Van Wanrooij and Van Opstal, 2005).

The bootstrap was implemented by resampling, with replacement, 1000 times from each simulated dataset at each source location across both original distributions.

Type I Errors

For each test the Type I error rate was estimated by allowing the alternative distribution to equal the $M(0,5)$ reference distribution and observing the proportion of 10,000 simulations where the corresponding tests yielded a p-value of less than the $\alpha = 0.05$ level. If Type I errors are suitably controlled then the empirical Type I error rate $\hat{p}$ should not differ significantly from

0.05. From the observed Type I error rates given in Table 1 we observed that this was the case for all tests except those based on stimulus-response analyses which are shown to be anticonservative.

[Table 1 about here.]

The anticonservative nature of the stimulus-response ANOVA test is due to the distributional properties of the model errors, especially for source locations close to 0° or 360°, in particular due to the 'wrapping' effect. The bootstrap test suffers from relatively small numbers of trials at each source location. Rerunning this analysis with 1,000 samples from each population, we obtained an estimated type I error rate of 0.0516 (95% CI: (0.047,0.056); p=0.477). These results suggest some significant short-comings in stimulus-response analyses.

Power

In order to investigate the power characteristics of each test both the location (μ) and concentration (κ) of the alternative distribution were allowed to vary within the range specified above. 1000 repetitions for each combination (μ, κ) were performed in order to assess the observed power with regard to the underlying distributional differences at a level of $\alpha = 0.05$. The obtained power estimates are detailed in Figure 2 below. Local quadratic regression surfaces were fit for a smoothed estimation of the critical region and power differentials.

[Figure 2 about here.]

Optimum power characteristics would correspond to a plot where power equates to 0.05 (the α level) at the point where the alternative and reference distributions are equivalent (i.e. $\theta = 0$ and $\kappa = 5$) with power sharply increasing towards 1 as either location or concentration depart from this baseline.

It is apparent that most tests performed reasonably under certain conditions. Both von Mises based circular tests (Figure 2, panels 2 and 3) proved powerful under their respective constraints of homogeneity of location and concentration. The univariate ANOVA (Figure 2, panels 4a and 4b) and Mann-Whitney U (Figure 2, panels 5a and 5b) tests yielded broadly comparable results. The signed error results (Figure 2, panels 4a and 5a) were similar to those obtained from the von Mises location test (Figure 2, panel 2) at higher concentrations, indicating sensitivity to change in location but not to change in concentration when the data were typically distributed away from the boundary region. However at lower concentrations the relative performance of the univariate tests degraded. (The 'J' shape of the non-critical region is attributable to the asymmetry of extreme points being located at -165° and +180°.) All analyses of unsigned errors (Figure 2, panels 4b and 5b) show undesirable power characteristics with 'U' shaped non-critical regions. This is a feature of the destructive folding transformation whereby negative errors are mapped onto their positive counterparts. This transformation jointly affects the mean and standard deviation of the transformed variable leading to this apparent anomaly. Finally we note the poor characteristics and low levels of power obtained from the stimulus-response analyses (Figure 2, panels 6a and 6b); these are due primarily to wrapping of the response especially towards the extremes of the stimulus axis.

In contrast, the CoM MANOVA technique (Figure 2, panel 1) appeared to be sensitive to changes in location and/or concentration within the range of these comparisons, and was associated with a smaller non-critical region than any other test apart from the anticonservative stimulus-response analysis. In the presence of confidence regarding homogeneity of location or concentration, specific tests did provide a slight improvement in power within the simulations explored. The maximum power gain over the CoM test by von Mises location: +11%; von Mises concentration: +15%; signed-error ANOVA: +10%, unsigned error ANOVA: +8%; signed-error Mann-Whitney U: +10%; unsigned-error Mann-Whitney U: <+1%; stimulus-response ANOVA F test: +6%; and stimulus-response F statistic bootstrap: +2%. However, these gains were not robust to departures from these respective assumptions of homogeneity and even maximal gains would typically provide little, if any, benefit once multiple testing corrections were made to allow for separate testing of change in location and concentration. In most circumstances, unsigned error tests and stimulus-response analyses should be used with caution due to the peculiar and potentially misleading behaviour outlined above.

b. Experimental data

The data described in this section were collected as part of a study to assess the effect of a temporary unilateral conductive hearing loss on normally hearing listeners' localisation ability, as well as to ascertain whether any immediate deficits in performance could be improved by training. The listener was familiarised with the localisation task over three sessions prior to insertion of an earplug into the left ear. Five subsequent testing sessions were carried out (once daily) with the earplug in place (Plug1 – Plug5), and one final session after plug removal. The subject was a 21-year old male with normal hearing (< 20 dB HL at 0.25, 0.5, 1, 2, 3, 4, 6 and 8 kHz). The experimental apparatus was a three-metre diameter, horizontal circular configuration of 24 numbered Bose Acoustimass cube loudspeakers situated in a sound- and echo-attenuated chamber. The loudspeakers were evenly distributed at 15^{o} intervals. A BOSE Acoustimass bass unit was associated with each quadrant of six speakers to provide the low-end output (< 300 Hz). Stimuli were broadband pink noise bursts (100 ms duration with a 5 ms rise and fall time). Stimulus amplitude was roved between 50 and 70 dB SPL from trial to trial in steps randomly determined by the presenting software. A touch-screen, situated upon the listener's lap, displayed a schematic representation of the speaker ring and was used to make responses. The listener was required to indicate which of the loudspeakers produced the sound by touching the appropriate location on the screen. Feedback was given throughout training. The chosen loudspeaker location flashed green (correct) or red (incorrect). A session had 240 randomised trials. The full results of this study will be reported elsewhere. Data for one listener from the Plug1 and Plug5 sessions are presented here solely for illustration.

Results

[Figure 3 about here.]

The data are summarised in the stimulus-response plot in Figure 3. In this case it is not appropriate to assume a consistent effect across different target locations, due to an anticipated decrement in performance for sounds presented from the load speakers located on the same (left) side as the earplug. As a consequence, a simple two-sample comparison was not performed. Instead a two-way univariate ANOVA based on the signed error suggested a highly significant effect of loudspeaker location on performance ($p<0.001$) but a non-significant difference between performance on the two testing sessions ($p=0.088$). If an interpretation were placed on the results of this analysis, the conclusion would be that there was no evidence of a systematic difference between training session. However, when adopting the CoM approach (Using Wilks' Λ) we obtained strong evidence for significant effects of both loudspeaker location and testing session ($p<0.001$ and $p=0.005$ respectively). In this case the inferences based on the CoM method are stronger than those based on a simple signed-error analysis. This may be symptomatic of the general gain in power noted above. The data underlying this spatial analysis are summarised in Figure 4. The improvement over sessions is a small one, but can clearly be seen to occur for sound sources in the posterior left-hand quadrant.

[Figure 4 about here.]

To illustrate one of the problems with direct statistical analysis of stimulus-response data, a regression model was fit to these data. The model comprised session as a factor and target loudspeaker location as a covariate. Where the target and response location were both coded as angular displacements (-165° through 180°), the session effect was found to be significant ($p=0.027$). However, if both were coded as loudspeaker locations (1 through 24) then the session effect did not reach statistical significance ($p=0.339$). This type of inconsistency is clearly highly undesirable.

V. Conclusion

Examination of simulated data has illustrated difficulties with the implementation of a variety of common technique or analysing localisation data. In particular, analysis of unsigned errors can lead to misleading conclusions, while signed error and stimulus-response analyses may suffer from problematic boundary effects. Directional statistical theory provides solutions to these problems, although these can be more challenging to implement, especially with more complex experimental designs involving multiple participants, repeated testing sessions or different stimulus conditions.

The CoM approach proposed in this paper is seen to be generally powerful, providing a robust and pragmatic solution to the exploration of such data. In particular, where localising within an array of loudspeakers the underlying multivariate ANOVA methods provide a convenient method for separating out the effect of the loudspeaker location from other condition-specific effects. Post-hoc methods can be used to investigate the nature of significant effects, such as front-back location effects or interactions. Furthermore, CoM plots can be used to visualise the patterns of specific significant effects and to facilitate their interpretation. They are complementary to the stimulus-response ‘bubble’ plots shown earlier. However, where the latter is particularly suited to visualisation alone the former directly relates to a robust statistical analysis of localisation performance.

VI. References

Abel, S. M., and Paik, J. E. S., 2004. The benefit of practice for sound localization without sight, Applied Acoustics 65, 229-241.

Best, D. and Fisher, N., 1979. Efficient simulation of the von Mises distribution, Applied Statistics, 28, 152–157.

Buunen, T. J., and Rhode, W., 1978. Responses of fibers in the cat's auditory nerve to the cubic difference tone, Journal of the Acoustical Society of America 64, 772-781.

Goldberg J. M., and Brown, P.B., 1969. Response of binaural neurons of dog superior olivary complex to dichotic tonal stimuli: Some physiological mechanisms of sound localization, J Neurophysiol 32, 613-636.

Hine, J. E., Martin, R. L., and Moore, D. R., 1994. Free-field binaural unmasking in ferrets, Behav Neurosci 108, 196-205.

Hofman, P. M., Vlaming, M.S.M.G., Termeer, P.J.J., and Van Opstal, A. J., 2002. A method to induce swapped binaural hearing, Journal of Neuroscience Methods 113, 2, 167-179.

Hofman, P. M., Van Riswick, J. G., and Van Opstal, A. J., 1998. Relearning sound localization with new ears, Nat Neurosci 1, 417-421.

Kacelnik, O., Nodal, F. R., Parsons, C. H., and King, A. J., 2006. Training-Induced Plasticity of Auditory Localization in Adult Mammals, PLOS Biology 4, 0001-0012.

Keuroghlian, A.S., and Knudsen, E.I., 2007. Adaptive auditory plasticity in developing and adult animals, Progress in Neurobiology 82, 109-121.

Mardia, K. V., and Jupp, P. E., 2000. Directional Statistics, Wiley, Chichester.

McAlpine, D., 2005. Creating a sense of auditory space, J Physiol 566, 21-28.

Middlebrooks, J. C., and Green, D. M., 1991. Sound localization by human listeners, Annu Rev Psychol 42, 135-159.

Moore, D. R., and King, A. J., 1999. Auditory perception: The near and far of sound localization, Curr Biol 9, R361-363.

Parsons, C. H., Lanyon, R. G., Schnupp, J. W., and King, A. J., 1999. Effects of altering spectral cues in infancy on horizontal and vertical sound localization by adult ferrets, J Neurophysiol 82, 2294-2309.

R Development Core Team, 2007. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Van Wanrooij, M. M., and Van Opstal, A. J., 2005. Relearning sound localization with a new ear, J Neurosci 25, 5413-5424.

Van Wanrooij, M. M., and Van Opstal, A. J., 2007. Sound localization under perturbed binaural hearing, J Neurophysiol 97, 715-726.

Zahorik, P., Bangayan, P., Sundareswaran, V., Wang, K., and Tam, C., 2006. Perceptual recalibration in human sound localization: learning to remediate front-back reversals, J Acoust Soc Am 120, 343-359.

Figure Captions

Figure 1: Probability density functions (pdf) for three von Mises distributions.

Figure 2: Observed power from two sample tests using 1: Centre of Mass MANOVA; 2: von Mises location test; 3: von Mises concentration test; 4a: signed error univariate ANOVA; 4b: unsigned error univariate ANOVA; 5a: signed error Mann-Whitney U; 5b: unsigned error Mann-Whitney U; 6a: stimulus-response ANOVA F test; and 6b: stimulus-response F-statistic bootstrap. Lighter shading indicates lower power; darker shading indicates increased power. The black contour line indicates the region within which results are consistent with the null hypothesis. This region is estimated from a local quadratic regression surface.

Figure 3: Bubble plot indicating the number of responses at each location for each target location, separately for plugged sessions 1 and 5, for 100 ms stimuli. The bubble width indicates the number of trials.

Figure 4: CoM plot showing listener's performance at plugged sessions 1 and 5. Each marker indicates the 'centre of mass' of all the responses made for a particular source location. Radial movement is indicative of inaccuracy and tangential movement is indicative of error direction. Colours indicate the loud speaker and the responses associated with it.

Table 1

Test	$\hat{p}$	95% CI	p-value
Centre of Mass	0.0467	(0.043,0.051)	0.136
von Mises (mean)	0.0526	(0.048,0.057)	0.242
von Mises (concentration)	0.0524	(0.048,0.057)	0.281
Univariate ANOVA			
signed error	0.0521	(0.048,0.057)	0.347
unsigned error	0.0503	(0.046,0.055)	0.909
Mann-Whitney U			
signed error	0.0502	(0.046,0.055)	0.945
unsigned error	0.0500	(0.046,0.054)	>0.999
Stimulus-Response			
ANOVA F test	0.1059	(0.100,0.112)	<0.001
F statistic bootstrap	0.0637	(0.059,0.069)	<0.001

Table 1: Type I Error control estimated from 10,000 simulations of two 100 observation sample tests drawn from $M(0,5)$ distributions. p-values correspond to the null hypothesis $H_0: p = 0.05$.

Figure 1
Click here to download high resolution image

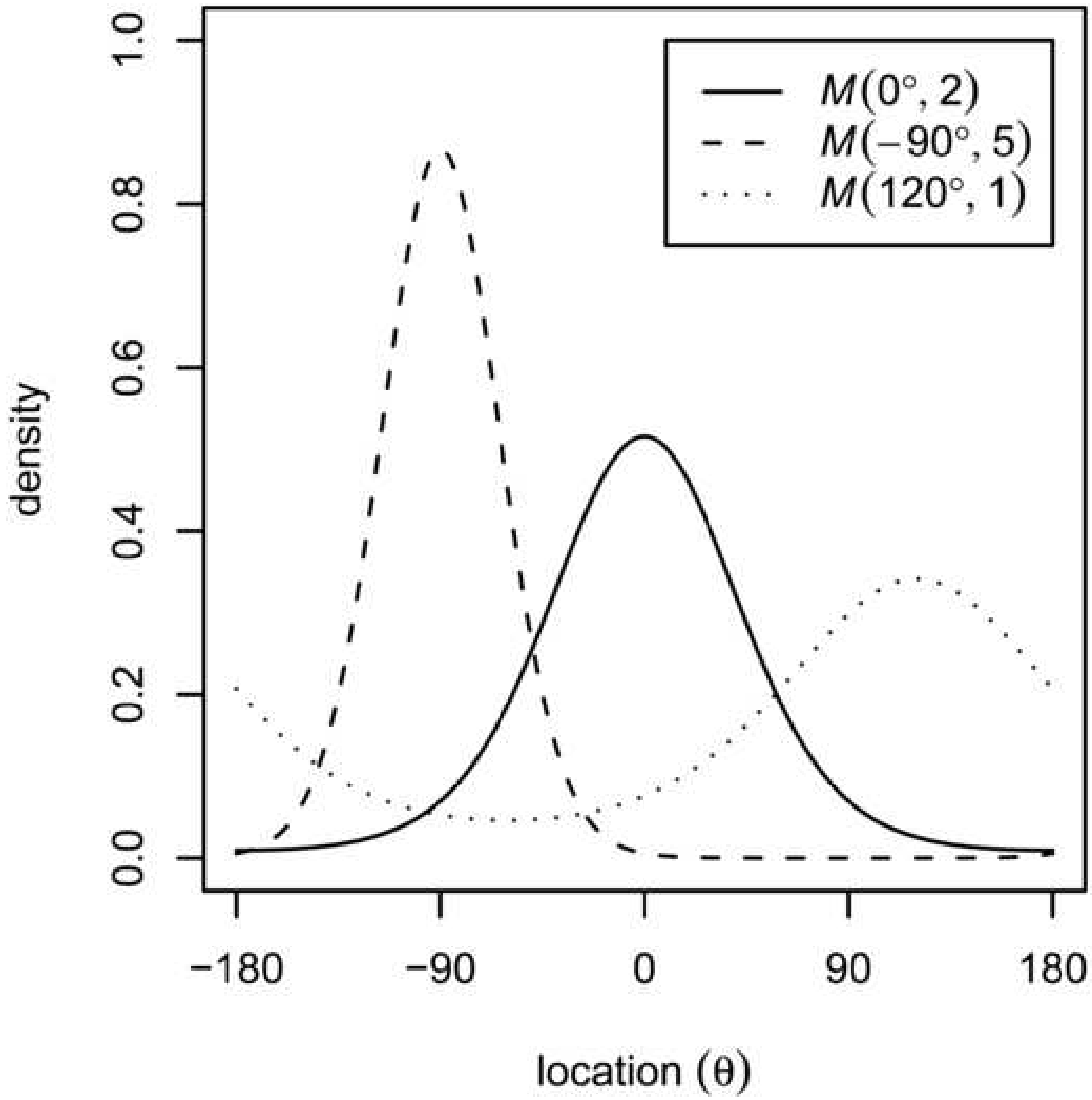


Figure 2
Click here to download high resolution image

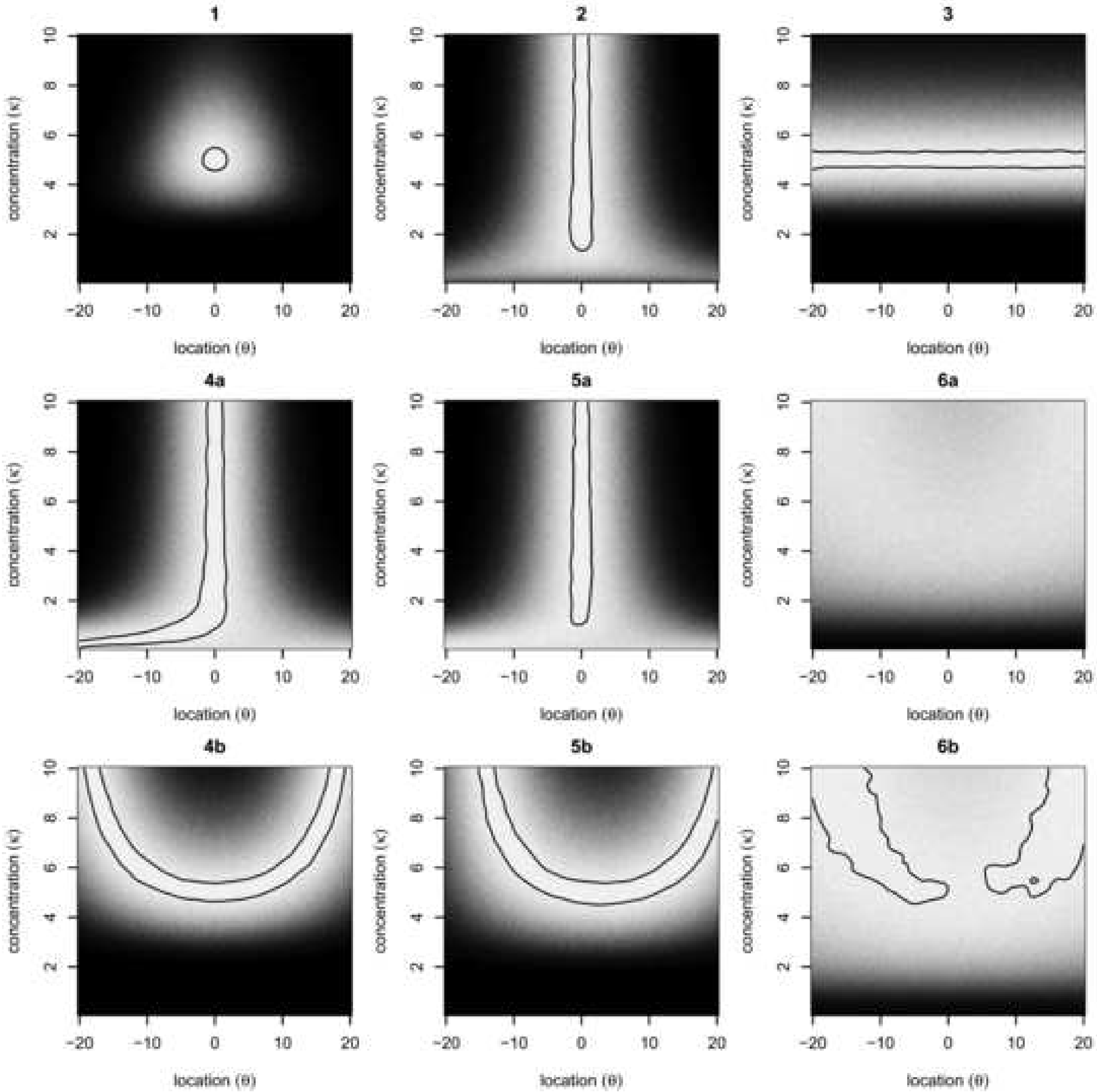


Figure 3
Click here to download high resolution image

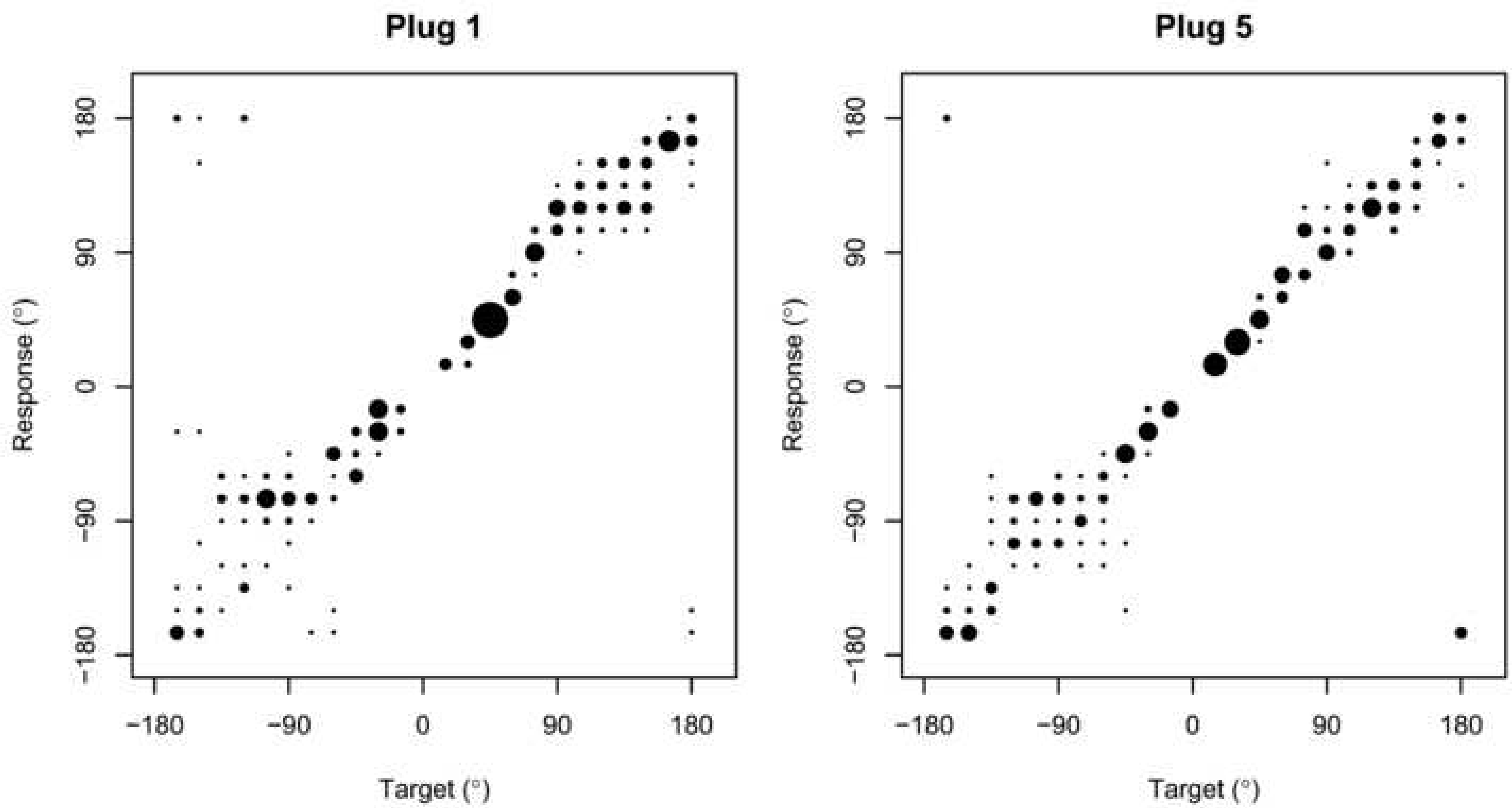
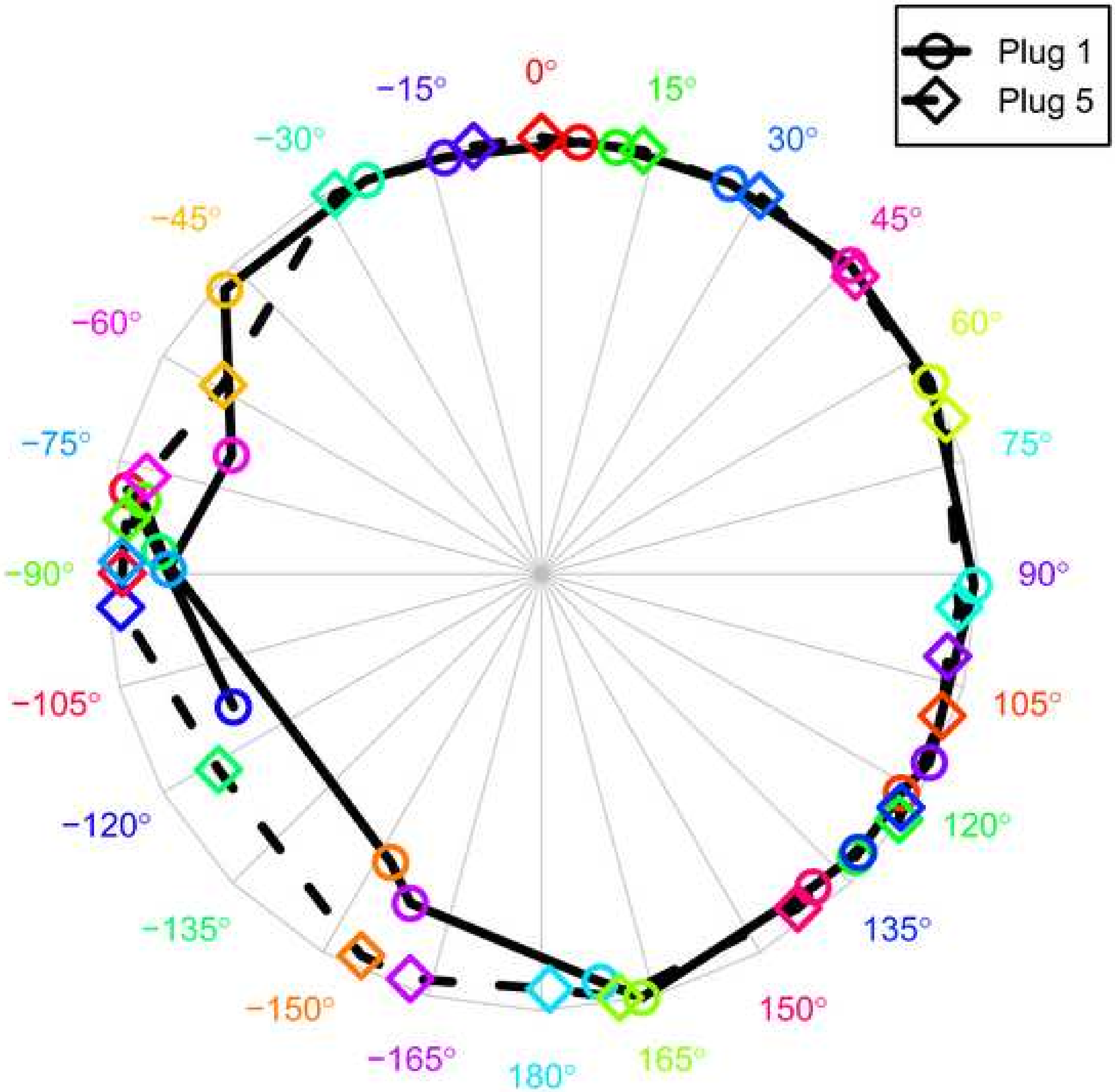


Figure 4
Click here to download high resolution image



MRC | Institute of Hearing Research

25 March 2009

Dear Professor Rubel

Methodological article: A novel application of a centre of mass approach to the statistical analysis of localisation performance

Please find enclosed the above manuscript, which I hope you will consider for publication in Hearing Research. The paper highlights some serious concerns relating to analysis methods commonly applied to localisation data collected within a 360° arena. It is common for such localisation studies to be analysed in terms of both accuracy and error magnitude. Error analyses typically apply conventional statistical techniques to signed or unsigned errors or assess the relationship between target and response locations using linear regression techniques. Both methods fail to account for the peculiar distributional properties of directional data. Here, we systematically validate a pragmatic spatial alternative (Centre of Mass approach) investigating the power characteristics of this method alongside a range of alternative techniques and applying the method to an illustrative example dataset.

Our conclusion is that this spatial method offers powerful, yet flexible, statistical analysis using standard multivariate analysis of variance (MANOVA). We hope that this statistical method will have broad application in the field of Hearing Research.

Yours sincerely

Mark Edmondson-Jones
Research Statistician
Direct line: +44 (0) 115 951 8508 + 219
E-mail: mark.edmondson-jones@ihr.mrc.ac.uk

*** Suggested Reviewers**

Suggested reviewers:

Prof. A.J. van Opstal
j.vanopstal@science.ru.nl
University of Nijmegen
Department of Medical physics and Biophysics
Phone 31 24 3614251 **Fax** 31 24 3541435

Prof Andrew King
andrew.king@dpag.ox.ac.uk
University of Oxford
Department of Physiology, Anatomy and Genetics
Phone 01865 272523 **Fax** 01865 272469

Dr Kevin Munro
kevin.munro@manchester.ac.uk
University of Manchester
School of Psychological Sciences
Phone 0161 275 3360 **Fax** 0161 275 3373