An Observational Study of the Precedence Effect

Hüseyin Hacıhabiboğlu*, Fionn Murtagh†
Sonic Arts Research Centre (SARC), School of Computer Science, Queen's University Belfast, Belfast, BT7 1NN, United Kingdom

Summary
This paper proposes a framework model for the analysis of sound source localisation under the precedence effect using Gaussian mixtures. The model quantifies the effect of the lagging sound over the angular position and the width of the perceived auditory event, and incorporates some well-known aspects of the precedence effect in a single framework. Two sound source localisation experiments were carried out in order to show the utility of the proposed model. In the first experiment, the localisation accuracy for single sound sources was measured. In the second experiment, the subjects localised spectrally coherent lead-lag stimulus pairs with different angular separations. The responses were analysed using the proposed model. It is shown that the model can be used for assessing the fusion, localisation dominance, and the lag discrimination suppression aspects of the precedence effect as well as the associated widening of the auditory event.

PACS no. 43.66.Qp, 43.66.Pn

1. Introduction

One of the most important perceptual phenomena in complex acoustical environments is the precedence effect [1, 2, 3]. The physiological, neural, or cognitive mechanisms involved in the occurrence of the precedence effect suppress the directional information conveyed in most of the reflections. However, this suppression does not void all the information conveyed in the reflections. Rather, the apparent width and distance of the sound source, and the diffuseness of the auditory event strongly depend on the presence of reflections. Although the directional information conveyed in the reflections is suppressed, some reflections can still be discriminated as being incident from the right or the left of the actual sound source. This allows the auditory system to build an internal representation of the acoustical environment and the sound source.

The precedence effect was discovered around the same time by two different groups (in 1949 by Wallach, Newman, and Rosenzweig at Harvard University, USA [4], and in 1951 by Haas at the University of Göttingen, Germany [5]). In the classical precedence effect experiment, two loudspeakers separated by 2θ on the horizontal plane (left loudspeaker at $-\theta$ and the right loudspeaker at θ azimuth) are positioned at a distance R from the listener. One of these loudspeakers is stimulated with a broadband click and the other one with the delayed replica of it (see Figure 1).

At short inter-onset delays (Δt) between the direct sound and the reflection, a single fused auditory event is perceived [2, 6]. This phenomenon is known as *fusion*. As Δt is increased beyond the *temporal echo threshold* (τ_{high}), the simulated reflection becomes audible as a separate auditory event (i.e. as an echo). For $\Delta t = 0$, the fused auditory image appears right at the middle of the two sound sources. As the delay is gradually increased, the perceived auditory image shifts toward the leading sound. At a lower threshold, τ_{low}, the fused image does not shift significantly anymore and the perceived location of the fused auditory event depends more on the location of the leading sound than the location of the lagging sound [3, 7, 8, 9, 10]. This phenomenon is known as the *localisation dominance*. This property of the precedence effect is important in the sense that it allows near-accurate detection of the actual direction of a sound source in the presence of interfering reflections within the given time limits.

A change in the location of the lagging sound source is more difficult to perceive than a change in the location of the leading sound [10, 11, 12]. This phenomenon is known as *(lag) discrimination suppression* which prevents the perception of the direction of the lagging sound for a given range of Δt values. Discrimination suppression is also an important property of the human auditory system which reduces the redundant directional information conveyed in the reflections. It has been shown that the minimum audible angles (MAAs) for the lagging sound sources are higher than either MAAs for single sound source condition or the leading sound [13, 14, 15, 16].

Received 17 November 2004, Revised 29 July 2005, accepted 8 November 2005.

* Centre for Communication Systems Research (CCSR), University of Surrey, Guildford, GU2 7XH, United Kingdom. h.hacihabiboglu@ieee.org

† Department of Computer Science, Royal Holloway, University of London, Egham, Surrey TW20 0EX, United Kingdom. fmurtagh@acm.org

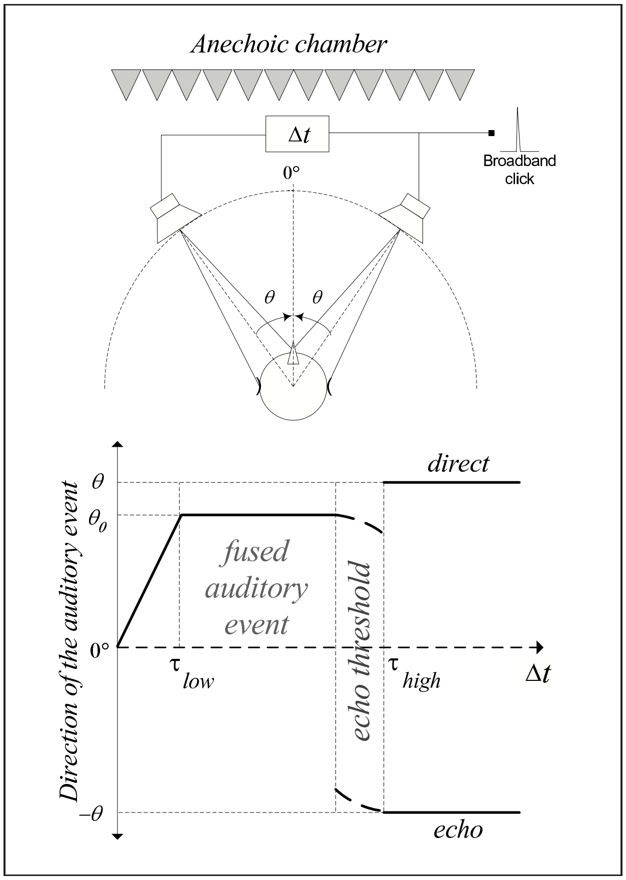

Figure 1. Classical precedence effect experiment (after Blauert, 1997). (a) The test setup, and (b) the direction of the auditory event with respect to the lead/lag delay.

Comprehensive reviews of previous work and extensive lists of previously measured average echo thresholds, and lag discrimination suppression thresholds, can be found in Zurek [17], Blauert [1] and Litovsky *et al.* [2].

Further, the presence of lateral reflections increases the perceived width and diffuseness of the auditory event in comparison with the single sound source localisation case [18]. Even the presence of a single reflection changes the perception of the auditory event substantially, adding spaciousness to the overall perception of the acoustical environment.

This paper proposes a framework model which can be utilised in the analysis of sound source localisation under the precedence effect conditions. The proposed model accommodates means of modelling the fusion, localisation dominance, lag discrimination suppression properties of the precedence effect and the associated widening of the auditory event. The applicability of the model is displayed by applying the model on data obtained in subjective localisation experiments.

The organisation of the paper is as follows: The previous models of the precedence effect are briefly reviewed, and the motivation of the paper is laid out in section 2. A parametric framework model of sound source localisation under the precedence effect is presented in section 3. Two subjective localisation experiments are reported in sections 4 and 5. The results of the experiments are interpreted in section 6 for displaying the application of the proposed model. The relation of the present study to previous studies is given in section 7. The findings are discussed in section 8 and the conclusions are drawn in section 9.

2. Background and motivation

2.1. Background

It is well-known from observational studies that the precedence effect has both temporal and localisational aspects. There have been different attempts at quantifying the relative influence of the reflections over the perceived auditory event. However, there is no general consensus on how to assess the relative prominence of a given early reflection with reference to its direction, arrival time, and level. The following discussion gives a brief account of the previous approaches.

2.1.1. Acoustical plausibility hypothesis

When a click-pair is repeated a number of times from two spatially separated loudspeakers, the temporal echo threshold (τ_{high}) gradually increases (i.e. the lagging sound is not heard as an echo even at longer delays). If the spatial positions of the leading and the lagging click are exchanged at the end of the click train instantly, the precedence effect breaks down and the sound sources are perceived as separate auditory events [19]. After a few repetitions of the new condition, the precedence effect builds up again and the location of the new leading source dominates. This phenomenon is called the Clifton effect [1].

Other than the Clifton effect, a sudden change in the frequency spectrum of the lagging sound [20], a sudden increase or decrease of the delay between the leading and the lagging sounds [21], and suddenly switching from a single-source stimulus to one with a simulated reflection [6], may also result in the breakdown of the precedence effect. In the context of an experiment investigating the influences that affect the buildup and the breakdown of the precedence effect, Clifton *et al.* [22] suggest that, a significant change in the azimuthal direction of a reflection can cause the breakdown of the built-up precedence effect. Further, it was shown that the precedence effect builds up directionally [23], and that it is unlikely for the built-up precedence effect to break down if the change in the azimuthal location of the lagging sound is small.

The acoustical plausibility hypothesis suggests that the precedence effect is a cognitive process rather than one taking place in the peripheral or central auditory systems.

2.1.2. Models of central inhibition

Central auditory system is thought to be the place in the auditory path where binaural coincidence detection (and thus the extraction of the ITD from binaural signals) takes place [24]. Therefore, the localisational aspects of the precedence effect point to the possibility that a central inhibition mechanism, which enables the extraction of the actual location of the sound source, may be present. A number of computational models of the precedence effect explore the possibility of including this inhibition mechanism.

Two models deserve special attention as they have inspired more recent models. 1) Zurek's precedence effect model is one of the earlier models employing an inhibition mechanism which depends on onset detection [17], 2) Lindemann's precedence effect model [25, 26] is an extension to the Jeffress' running cross-correlation model. Zurek's model employs an onset detection mechanism to facilitate the inhibition of the directional information conveyed in the reflections when an onset is detected. Lindemann's model includes a central inhibition mechanism, which suppresses forthcoming reflections when a local peak at the running interaural cross-correlation is detected. Coincidence detection is a process involving interaural differences. However, precedence effect is also known to occur for sound sources on the median plane where no significant interaural differences are present [27, 28]. This casts doubt on whether the precedence effect is purely a process involving binaural interaction.

2.1.3. Models of weighting

The observer-weighting approach [29] when used in a localisation task quantifies the *relative perceptual weight* of the lagging sound on the overall perceived auditory event, but does not model any localisation-related aspect of the precedence effect. It is reasonable to use this approach to assess the relative contribution of the lagging sound for different lead/lag delays [30], onset dominance [31], and binaural adaptation [32]. However, the effect of the lagging sound on the perceived mean location and width of the auditory event cannot be assessed.

A localisation based approach to modelling the salience of the precedence effect is built upon the observation that mean perceived azimuth of an auditory event is affected in the presence of a lagging sound [1, 2]. The perceived mean azimuth of a fused auditory event can be attributed to a single effective interaural time delay (τ_e) which is a weighed and noise-corrupted average of interaural time delays (ITD) of the leading and the lagging sounds (ITD_{lead}, ITD_{lag}) [7]:

$$\tau_e = c_S \cdot \mathrm{ITD}_{lead} + (1 - c_S) \cdot \mathrm{ITD}_{lag} + \mathbf{N}(0, \sigma), \qquad (1)$$

where $0 \leq c_S \leq 1$ and $\mathbf{N}(0, \sigma)$ is a zero-mean, Gaussian-distributed random variable with a standard deviation σ which is independent from presentation to presentation. The model quantifies the relative prominence of the leading ITD over the lagging ITD, by using a single precedence level measure, c_S, that denotes the level of precedence. Therefore, the precedence effect is characterized by the shift of the mean azimuth.

Litovsky and Macmillan [13] indicate that the Gaussian term in the given model represents trial-to-trial variability in localisation and not the spatial extent within a trial. Hence, the increase in the apparent width of the auditory event cannot be modelled.

Models of weighting are useful for the experimental investigation of the perceptual weight attributed to a reflection. However, the fundamental difficulty is that most of the temporal and localisational aspects of the precedence effect cannot be easily assessed.

2.2. Motivation

The hypothesis investigated in this study is as follows:

A sound source is conceptualised as an information source encoding audible positional information in terms of the interaural differences. However, additive perceptual noise corrupts this information at the receiver (i.e. listener) side. This additive perceptual noise is assumed to be Gaussian with a zero mean ($\mu = 0$) and with a standard deviation (σ) related to the *localisation blur* [1]. Single sound source localisation experiments are usually in agreement with this assumption in the sense that observational data in free field sound source localisation experiments are normally distributed.

For the classical precedence effect experiment with two such information sources, the sound localisation responses are expected to be sampled from a linear mixture of two Gaussian distributions. The means of the mixture components are determined by the processes involved in the occurrence of the precedence effect and the standard deviations are equal to the standard deviation of the perceptual noise for single sound source condition.

It must be pointed out that the intention of this study is not to investigate the psychophysical mechanism governing the precedence effect itself or to suggest a physiologically feasible explanation to the precedence effect, but to assess the *observational* aspects. That is to say, the internal representation of the perceived auditory event is under investigation, and not the psychological, physiological, neurological or cognitive processes resulting in such an internal representation. The approach taken in this paper rather aims to develop a framework model which embodies some well-known properties of the precedence effect.

3. A framework model of sound source localisation under the precedence effect

3.1. Gaussian distributions for modelling sound source localisation

A sound signal played using a single sound source in free-field conditions is the only contributor to the perceived auditory event if its level at the listener position is higher than the absolute threshold of hearing. In such a scenario, it may be observed from localisation experiments that the sound localisation can be modelled simply as a single Gaussian distribution. The single Gaussian model has two parameters: the mean value (μ) which is related to the perceived mean azimuth of the sound source, and the standard deviation (σ) which is related to the additive perceptual noise (or response variation). This variation is related to the MAA of the sound source and is a measure of subjective localisation acuity.

When there are two sound sources, it is reasonable to assume that the observational data is a linear combination of two Gaussian distributions (i.e. a Gaussian mixture). A two-component Gaussian mixture has five parameters. These are the mean values of the mixture components (μ_1 and μ_2), the variances of the mixture components (σ_1 and σ_2), and the mixture proportion (p). The total number of physical parameters that relate to the sound sources are the actual azimuth angles of the leading and the lagging sound sources (θ_{lead} and θ_{lag}), the individual levels of the sound sources (L_{lead} and L_{lag}), the temporal separation between the leading and the lagging sound sources (Δt), spectral content, and the spectral coherence of the leading and the lagging sounds. The following two sections explain how these physical parameters can be related to the model parameters.

3.2. Mixture of two Gaussian distributions

A two-component univariate Gaussian mixture is defined as:

$$\Gamma(\mu_1, \mu_2, \sigma_1, \sigma_2, p) = p \cdot \mathbf{N}(\mu_1, \sigma_1) + (1-p) \cdot \mathbf{N}(\mu_2, \sigma_2), \tag{2}$$

where μ_1, μ_2, σ_1, σ_2, and p are the means, standard deviations and the mixture proportion of the model respectively. Further specifics about a Gaussian mixture are the equivalent mean ($\overline{\mu}_{mix}$) and the equivalent variance (σ^2_{mix}) which can be expressed as:

$$\overline{\mu}_{mix} = \sum_{i=1}^{2} p_i \mu_i, \tag{3}$$

$$\sigma^2_{mix} = \sum_{i=1}^{2} p_i(\sigma_i^2 + \mu_i^2) - \overline{\mu}^2_{mix}. \tag{4}$$

The following inequalities hold for the equivalent mean and variance:

$$\begin{aligned} \mu_1 &\le \overline{\mu}_{mix} \le \mu_2, \\ \sigma^2_{mix} &\ge \max(\sigma_1^2, \sigma_2^2). \end{aligned} \tag{5}$$

A Gaussian mixture of two components can have at least one (*unimodal*) and at most two modes (*bimodal*) [33]. The modes of a Gaussian mixture cannot be obtained using a closed form equation and require an iterative algorithm. However, a sufficient condition for determining whether a Gaussian mixture is unimodal is defined in [34] as:

$$\mid \mu_1 - \mu_2 \mid \le 2\min(\sigma_1, \sigma_2). \tag{6}$$

This suggests that for $\mu_1 = \mu_2$, the mixture is unimodal regardless of the other parameters. If $\sigma_1 = \sigma_2 = \sigma$, a stricter condition which depends also on the mixture proportion, p, is:

$$\left|\mu_1 - \mu_2\right| \le 2\sigma\sqrt{1 + \left|\ln(p) - \ln(1-p)\right|/2}. \tag{7}$$

Figure 2 shows the dependence of the modality of a Gaussian mixture distribution on the means of its components.

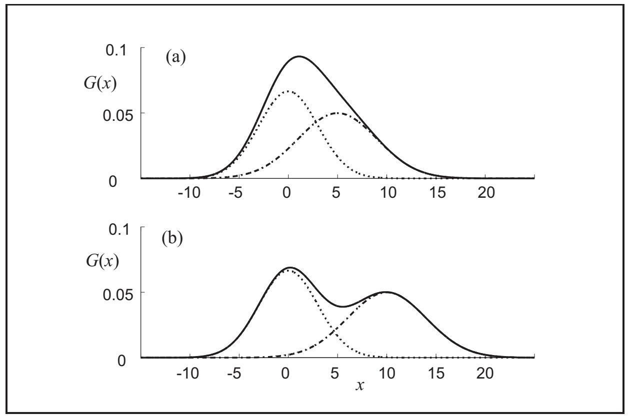

Figure 2. Unimodality *vs.* bimodality of a two-component Gaussian mixture. (a) Unimodal mixture with $\mu_1 = 0$, $\sigma_1 = 3$ (dotted) and $\mu_2 = 5$, $\sigma_2 = 4$ (dash-dot), and (b) bimodal mixture with $\mu_1 = 0$, $\sigma_1 = 3$ (dotted), and $\mu_2 = 10$, $\sigma_2 = 4$ (dash-dot) of the variate x.

The mixture proportion, mean and variance of the first component, and the variance of the second component are held fixed ($p = 0.5$, $\mu_1 = 0$, $\sigma_1 = 3$, $\sigma_2 = 4$) while the mean of the second component (μ_2) is increased. For the case $\mu_2 = 5$, this results in a unimodal distribution, and for the case $\mu_2 = 10$, in a bimodal distribution with reference to equation (6).

Given a set of observations drawn from a Gaussian mixture, the parameters of the mixture distribution can be obtained by a two-step algorithm known as the expectation maximization (EM) algorithm. The EM algorithm works by recursively calculating the expectation of the log-likelihood of data given the parameters of a Gaussian mixture model, and adjusting the mixture parameters (i.e. component means and variances, and mixture proportions) accordingly to increase the value of this expectation.

A two-component Gaussian mixture always provides an equivalent or better fit to any finite dataset than a single Gaussian (see for example [35]). Further, a three-component Gaussian mixture obtained in the same way using the EM algorithm would also provide a fit equivalent to or better than a two-component mixture. It must be noted that the number of components in the mixture model that we propose structurally has to be two and only two if we aim to separate the effect of the lag on the perceived auditory event(s) for a single lagging sound.

3.3. Modelling the precedence effect using a Gaussian mixture model

The precedence effect depends in many ways on the time delay (Δt) between the leading and the lagging sound sources, levels (L_{lead} and L_{lag}), the level difference (ΔL), the spectral content, the spectral coherence, and the azimuthal directions (θ_{lead}, θ_{lag}) of the leading and the lagging sound sources. Therefore, any proposed model of the precedence effect has to take these factors into account.

In the absence of other sound sources in the free-field, a sound source is perceived at a mean location that coincides with the actual azimuth direction of the sound source. The following three assumptions are made for the precedence effect conditions for which the leading and the lagging sounds are spectrally coherent:

- Each sound source has a probability of being perceived which is related to its relative presentation level with respect to other sound sources. This probability determines the contribution of the sound source to the perceived auditory event. It is also assumed that the lagging sound signal has a level which is smaller than or equal to the level of the leading sound signal. This assumption is consistent with the fact that, in nature, a reflection always has a level smaller than that of the direct sound.
- The variance of the perceptual noise is similar for each sound source. In the presence of a leading sound signal and at least one lagging sound signal, this noise is additive and is related to the apparent width of the perceived auditory event.
- The precedence effect is a simple process which modifies the perceived directions of the individual auditory events and not the perceptual noise associated with them. The nature of this modification depends on the time delay and the spatial separation between the leading and the lagging sound sources.

3.3.1. Temporal aspects, fusion and localisation dominance

In the classical precedence effect experiment, two sound sources at different azimuth directions are used. The perceived auditory event can be modelled as a mixture of two Gaussian distributions:

$$\begin{aligned}&\Gamma_{PE}(\mu_{lead}, \mu_{lag}, \sigma_{lead}, \sigma_{lag}, p) = \\ &p \cdot \mathbf{N}(\mu_{lead}, \sigma_{lead}) + (1-p) \cdot \mathbf{N}(\mu_{lag}, \sigma_{lag}),\end{aligned} \tag{8}$$

where $\mathbf{N}(\mu, \sigma)$ is the normal distribution, μ_{lead} is the mean perceived azimuth of the leading sound source, μ_{lag} is the mean perceived azimuth of the lagging sound source, σ_{lead} and σ_{lag} are the standard deviations of the additive perceptual noise attributed to the leading and the lagging sound sources respectively; $(1-p)$ and p are the mixture proportions which represent the perceptual weight given to each sound source in the formation of the composite auditory event. As suggested by the first assumption, $p \geq 0.5$ (i.e. the leading sound has a higher presentation level than the lagging sound).

The mean values of the mixture components can be expressed as weighted averages of the actual azimuth directions of the sound sources:

$$\begin{aligned}\mu_{lead} &= \frac{1}{2}\Big[\big(1 + c_{lead}(\Delta t, \Delta_\theta)\big) \cdot \theta_{lead} \\ &\quad + \big(1 - c_{lead}(\Delta t, \Delta_\theta)\big) \cdot \theta_{lag}\Big],\end{aligned} \tag{9}$$

$$\begin{aligned}\mu_{lag} &= \frac{1}{2}\Big[\big(1 + c_{lag}(\Delta t, \Delta_\theta)\big) \cdot \theta_{lead} \\ &\quad + \big(1 - c_{lag}(\Delta t, \Delta_\theta)\big) \cdot \theta_{lag}\Big],\end{aligned} \tag{10}$$

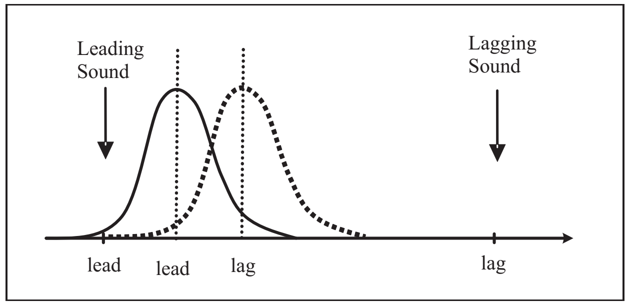

Figure 3. Two Gaussian mixture components on the perceptual azimuth axis. The actual azimuthal positions of the sound sources are denoted with arrows.

where $c_{lead}(\Delta t, \Delta_\theta)$ and $c_{lag}(\Delta t, \Delta_\theta)$, the precedence levels, are functions of both the time delay (Δt) and the azimuth separation ($\Delta_\theta = \theta_{lead} - \theta_{lag}$) between the leading and the lagging sound sources. Given that μ_{lead} and μ_{lag} are near, the auditory event would be perceived as a single fused auditory event. It must be noted that, under the precedence effect conditions with no level difference between the sound sources, the larger the values of c_{lead} and c_{lag}, the closer the fused auditory event will be to the leading sound (see Figure 3).

The following piecewise continuous functions can be used for $c_{lead}(\Delta t, \Delta_\theta)$ and $c_{lag}(\Delta t, \Delta_\theta)$:

$$c_{lead}(\Delta t, \Delta_\Theta) = \begin{cases} \Delta t \cdot \hat{\mathbf{c}}_{\mathbf{lead}}(\Delta_\theta)/\tau_{low} \\ \qquad \text{if } \Delta t < \tau_{low}, \\ \hat{\mathbf{c}}_{\mathbf{lead}}(\Delta_\theta) \\ \qquad \text{if } \tau_{low} < \Delta t < \tau_{high}, \\ 1 \qquad \text{if } \Delta t > \tau_{high}. \end{cases} \tag{11}$$

$$c_{lag}(\Delta t, \Delta_\Theta) = \begin{cases} \Delta t \cdot \hat{\mathbf{c}}_{\mathbf{lag}}(\Delta_\theta)/\tau_{low} \\ \qquad \text{if } \Delta t < \tau_{low}, \\ \hat{\mathbf{c}}_{\mathbf{lag}}(\Delta_\theta) \\ \qquad \text{if } \tau_{low} < \Delta t < \tau_{high}, \\ -1 \qquad \text{if } \Delta t > \tau_{high}. \end{cases} \tag{12}$$

where $\hat{\mathbf{c}}_{\mathbf{lead}}(\Delta_\theta)$ and $\hat{\mathbf{c}}_{\mathbf{lag}}(\Delta_\theta)$ are functions of the azimuth separation (Δ_θ) between the sound sources, τ_{low} is the lower temporal threshold and τ_{high} is the higher temporal threshold (see Figure 4). For any $\Delta t = t_0 > 0$ and $\Delta_\theta = \Theta \neq 0$,

$$\begin{aligned}&0 < \hat{\mathbf{c}}_{\mathbf{lag}}(\Theta) < \hat{\mathbf{c}}_{\mathbf{lead}}(\Theta) < 1, \\ &c_{lag}(\Delta t, \Delta_\theta) < c_{lead}(\Delta t, \Delta_\theta).\end{aligned} \tag{13}$$

This relation suggests that when $\Delta t < \tau_{high}$ (i.e. when the lagging sound is not perceived as a separate auditory event), such that the mean azimuths of the mixture components related to the leading and the lagging sound sources are between the actual azimuth angles of the leading and the lagging sound sources. Further, any change in the actual azimuth of the lagging sound source is less likely to change the mean azimuth of the fused auditory event (i.e. the equivalent mean of the Gaussian mixture). Therefore, the localisation dominance aspect of the precedence effect can be modelled using the equations (9) through (12). It

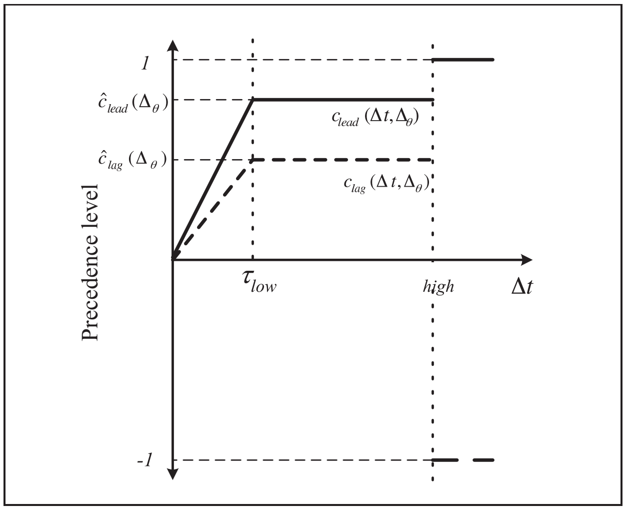

Figure 4. The precedence level functions $c_{lead}(\Delta t, \Delta_\theta)$, and $c_{lag}(\Delta t, \Delta_\theta)$.

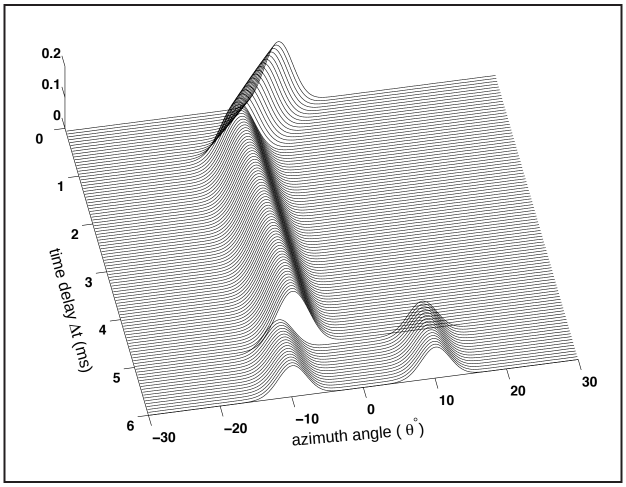

Figure 5. The Gaussian mixture model for the perception of the direction of the auditory event at different time delays for $\theta_{lead} = -10°$ and $\theta_{lag} = 10°$. The precedence levels c_{lead} and c_{lag} are as defined in the text.

should be noted that the precedence level functions are motivated by the outcome of the classical precedence effect experiment as shown in Figure 1, i.e. τ_{low} corresponds to the limit of summing localisation and τ_{high} corresponds to the echo threshold.

Figure 5 shows the model output for $0 < \Delta t < 6$ms with $\theta_{lead} = -10°$ and $\theta_{lag} = 10°$ for the condition where there is no level difference between the leading and the lagging sound sources and the well-known lower and higher temporal thresholds for broadband click pairs (i.e. $\tau_{low} = 1$ms and $\tau_{high} = 5$ms). The mixture proportion, p, is selected to be 0.5 to reflect leading and lagging sounds presented at the same level. $\mathbf{\hat{c}_{lead}}$ and $\mathbf{\hat{c}_{lag}}$ were selected as 0.9 and 0.7 for demonstration purposes. When there is no time delay ($\Delta t = 0$), there is a single fused auditory event perceived at the mid-line. If the time delay is increased such that $0 < \Delta t \leq 1$ms, the mean perceived azimuth of the auditory event gradually shifts toward the direction of the leading sound. When 1ms$< \Delta t < 5$ms, the mean perceived azimuth is stable near the actual azimuth of the leading sound source. If the time delay is increased beyond this point, the lagging sound becomes audible as a separate auditory event (i.e. as an echo). Thus, the temporal properties of the precedence effect can be modelled by defining precedence levels as two time-dependent functions.

3.3.2. Apparent width

It may be observed from Figure 5 that the Gaussian mixture model also has the capacity to model the widening of the auditory event in comparison with the single sound source case. When there's a single fused auditory event, the standard deviation of the mixture distribution is greater than the standard deviation of a single auditory event. This can be quantified in the following way:

Substituting equations (9) and (10) in equations (3) and (4) it can be shown that:

$$\overline{\mu}_{mix} = \left[\frac{1}{2}\Sigma_c\Delta_\theta + \frac{1}{2}\Sigma_\theta\right], \tag{14}$$

$$\sigma^2_{mix} = \frac{1}{4}p\,(1-p)\,\Delta_c^2\Delta_\theta^2 + \left(p\sigma^2_{lead} + (1-p)\sigma^2_{lag}\right), \tag{15}$$

where

$$\Delta_\theta = \theta_{lead} - \theta_{lag}, \tag{16}$$

$$\Sigma_c = pc_{lead} + (1-p)\,c_{lag}, \tag{17}$$

$$\Sigma_\theta = \theta_{lead} + \theta_{lag}, \tag{18}$$

$$\Delta_c = c_{lead} - c_{lag}. \tag{19}$$

These equations suggest that if the presentation level of the leading sound source is greater than the lagging sound source (i.e. $p > 0.5$), the equivalent mean of the mixture, $\overline{\mu}_{mix}$ is nearer to μ_{lag}. The equivalent variance which is related to the width of the auditory event is maximum for $p = (1-p) = 0.5$ (i.e. presentation levels of the leading and lagging sources are equal). If the level of the leading sound increases (i.e. $p : 0.5 \to 1$), the width of the fused auditory event decreases.

Together with the assumption that $\sigma_{lead} \approx \sigma_{lag} = \sigma$, the widening of the auditory event can be characterized by a single parameter, $A_p = \sqrt{\sigma^2_{mix} - \sigma^2}$ defined as:

$$A_p = \frac{1}{2}\left(\sqrt{(1-p)p}\right)\left|\Delta_c\Delta_\theta\right|. \tag{20}$$

Therefore, the widening of the auditory event can be modelled as a function of Δ_c, and the shift of the fused auditory event (from the midpoint) as a function of Σ_c. The widening of the auditory event is maximum for $p = (1-p) = 0.5$. The widening is less pronounced if the leading source has a level greater than the lagging one (i.e. $p > 0.5$).

For the classical precedence effect paradigm for which the leading and the lagging sounds are exact copies of each other presented at the same level, the mixture proportions for both components are considered to be the same (i.e.

$p = (1-p) = 0.5$). Then, the equation above for the increase in the apparent width becomes

$$A_p = \left|\Delta_c \Delta_\theta\right|/4. \tag{21}$$

3.3.3. Lag Discrimination Suppression

The lag discrimination suppression aspect of the precedence effect can be modelled using the modality of the Gaussian mixture. If the mixture is unimodal, this refers to the situation that the information whether the lagging sound source is to the right or to the left of the leading sound source is not available to the listener. If the mixture is bimodal, it means that the listener is informed whether the lagging sound source is to the right or to the left of the leading sound source.

If we assume again that $\sigma_{lead} \approx \sigma_{lag} = \sigma$, it is possible to define a mixture modality function, F_{mod}, by substituting equations (9) and (10) in equation (7):

$$\begin{aligned} F_{mod} &= \left|0.5\Delta_c\Delta_\theta\right| \qquad (22)\\ &= 2A_p \leq 2\sigma\sqrt{1+\frac{1}{2}\ln\left(\frac{p}{1-p}\right)}. \end{aligned}$$

The left side attains its minimum for the same presentation level (i.e. $p = (1-p) = 0.5$). If $p > 0.5$ the right hand side is greater than 2σ, and thus it is harder for the lagging sound to be discriminated if the presentation level of the leading sound is greater. If the lead and lag sources are presented at an equal level the sufficient condition for the mixture to be unimodal becomes:

$$F_{mod} = \left|0.5\Delta_c\Delta_\theta\right| = 2A_p \leq 2\sigma. \tag{23}$$

This condition, when satisfied, results in a unimodal distribution. However, it does not mean that the the mixture is definitely bimodal if the condition is not satisfied. Therefore, it may be stated that, if the increase in the apparent width is greater than the smallest of the standard deviations of lead and lag components (i.e. $A_p > \sigma$), the distribution is no longer guaranteed to be unimodal. Therefore, the model gives information on whether the lagging sound contributes to the spatial properties of the auditory event significantly or not. In other words, the model can give information on the extent to which the directional information conveyed in the lagging sound is suppressed. It should however be noted once again that this condition does only provide a lower bound for the lag discrimination suppression MAA, and not the lag discrimination suppression MAA per se.

Two experiments that demonstrate the application of the proposed model are reported in the following sections. The first experiment focuses on single source localisation that provides a measure of the response bias. The second experiment investigates sound source localisation for precedence effect stimuli pairs.

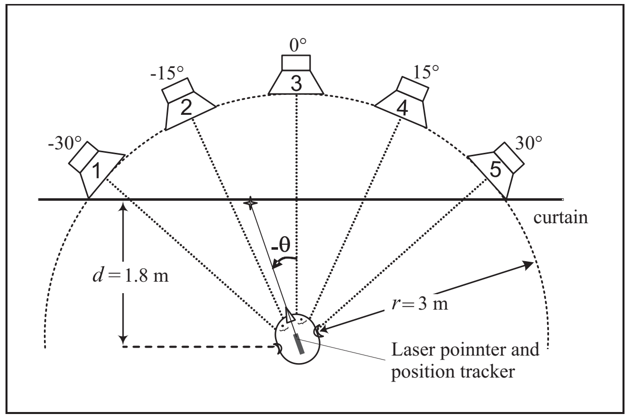

Figure 6. The test setup. The positions of the loudspeakers are denoted by numbers.

4. Experiment 1: Single source localisation

The motivation of this experiment is to characterize the source localisation acuity of the subjects and the response and/or measurement bias for sound sources at different azimuth angles for the hybrid pointing task used in the second experiment.

4.1. Method

4.1.1. Subjects

Five adults (three males including the first author ‘HH’, two females; aged 20 to 36) with normal hearing participated in this experiment. The subjects were informed about the tasks that they were required to complete. Although basic information was provided about sound source localisation and the precedence effect, the subjects except the author were not aware of the purpose of the experiments and the hypotheses being tested.

4.1.2. Apparatus and stimuli

The sound localisation tests were carried out in an acoustically isolated and attenuated $7.9\,\mathrm{m} \cdot 5.5\,\mathrm{m} \cdot 3\,\mathrm{m}$ critical listening room (ITU-R BS 1116) with a reverberation time (T_{60}) of $319\,\mathrm{ms}$. Five two-way loudspeakers (Genelec 1030A) with a flat frequency response between $55\,\mathrm{Hz}$ to $18\,\mathrm{kHz}$ ($\pm 2.5\,\mathrm{dB}$) were placed on a circular arc of radius $3\,\mathrm{m}$ at a height of $1.2\,\mathrm{m}$. The loudspeakers were positioned at $-30°$, $-15°$, $0°$, $15°$, and $30°$ azimuth (see Figure 6). The subject was seated at the centre of this circular arc on a wooden armchair. The acoustic axes of loudspeakers were pointed towards the subject. A black curtain was placed between the loudspeakers and the subjects to prevent any visual cues about the actual locations of the loudspeakers. (see Figure 7)

The experiment was carried out in complete darkness to prevent subjects from using any visual landmarks as location cues. The subject wore a headstrap with an attached laser pointer pointing to the subject’s forward direction providing the only visual feedback. The receiver part of an electromagnetic position tracker (Polhemus Isotrak II) was also attached to the headstrap in alignment with the laser pointer. The subjects were asked to remove all metallic objects on them before the test to prevent measurement errors with the electromagnetic position tracker. The gathered data consisted of the change in the angular position of the head that is known to be a good measure of auditory localisation [36].

An 8-channel audio interface (MOTU 828 Mk-I) was used to activate the loudspeakers. A mixing desk (MACKIE 1642-VLZ Pro) was used to attenuate signals between the audio interface and the loudspeakers.

The stimuli were sampled at 48 kHz with 16-bit quantization. They consisted of randomly generated 6 ms windowed white noise bursts with 2 ms rise-fall times and were regenerated for every repetition. Therefore, the stimulus spectrum changed randomly for each new presentation. The average level of the stimuli measured at the listener position was 59–61 dBA. The background noise level was less than 30 dBA. The generation and playback of the stimuli as well as data acquisition from the position tracker was controlled by using a special script running under MATLAB.

4.1.3. Procedure

On each trial, the stimulus was presented using one of the loudspeakers. The subject's task was to turn her/his head to face the perceived location of the sound source. The subject registered her/his response by using a response box attached to the left arm of the seat while s/he was facing the perceived location of the sound source. The stimulus could be repeated as many times until the subject felt confident with the response s/he gave. The subjects were encouraged to listen to the stimulus a second time after the first presentation after turning their heads towards the direction of the perceived auditory event. A forced delay of 500 ms was inserted between consecutive repetitions to prevent binaural adaptation. The order of loudspeakers to be activated was fully randomized. No feedback was provided to the subjects about the correct location of the loudspeakers. Presentation at each azimuth position was repeated 50 times over the course of the test for each subject. Therefore, each subject gave 250 responses in total in experiment 1. Each subject was trained for 30 minutes before the experiment for getting acquainted with the applied procedure. The experiment took about an hour for each subject to complete. Breaks were given during the experiment with regular intervals to prevent fatigue.

4.2. Results

The hybrid pointing task required the subjects not only to point to but also to face the perceived auditory event by turning their heads. All the subjects listened to the stimuli more than once as encouraged to do so. At each trial the subjects listened to the stimulus, turned their heads in the direction of the auditory event, and listened to the stimulus again. Therefore, before finalizing their responses, the direction of the active loudspeaker was already around $0°$ azimuth with respect to the subject's head. However, as the subjects were told not to reorient their bodies but only

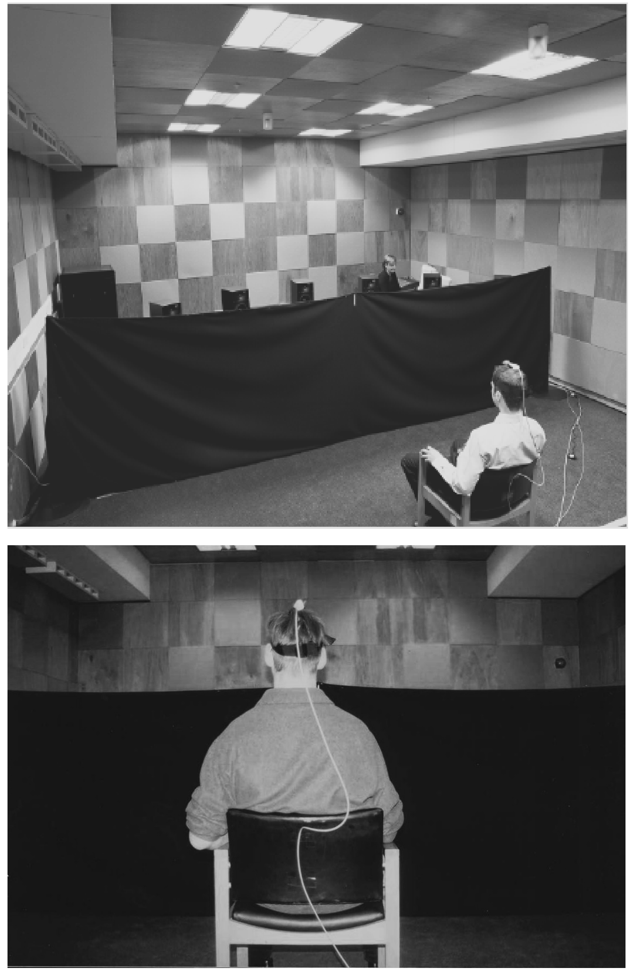

Figure 7. top: The test setup, and bottom: the subject's view during the tests.

to turn their heads, the reflections originating from their torsos differed in their direction with respect to the head for each stimulus direction in comparison with the situation where the head is aligned with the median plane. Therefore, the results of this experiment do not characterize sound source localisation accuracy for sounds at different azimuths with respect to the listener's median plane. Rather, the experiment measures the extent to which the localisation accuracy is affected due to the reflections from the torso and the measurement bias depending on the angular position of the head. The results therefore indicate the response and/or measurement bias due to the specific data collection technique that need to be eliminated.

Apart from the misalignment of the head from the median plane, the laser pointing method is also known to cause a perceptual bias i.e. subjects consistently tend to underestimate the actual angle, as a result of audiovisual interaction [37, 38]. The results in this experiment are comparable to previous findings in sound localisation tests with other laser pointing techniques in terms of the bias. A one-way independent analysis of variance (ANOVA) shows that the mean response bias depends on the presentation angle ($F = 59.37, df = 5, p \ll 0.01$). Correlation coefficients of the lines regressed to the mean values of re-

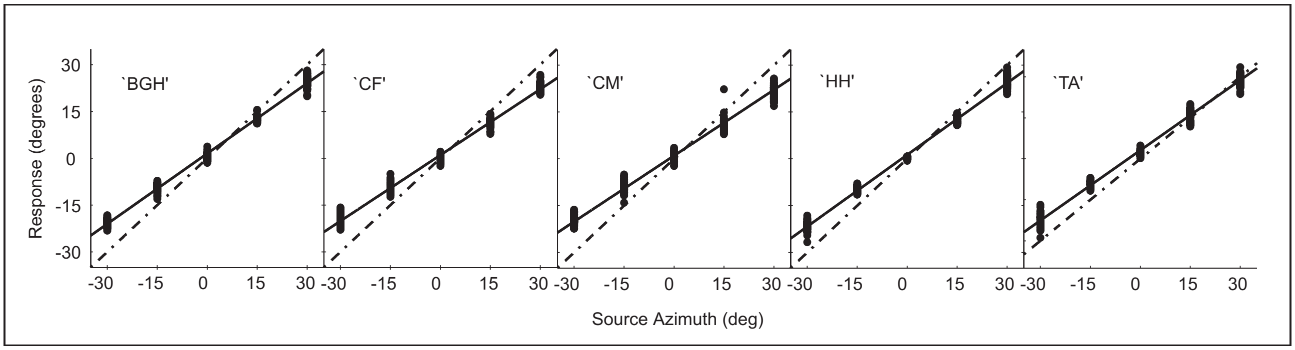

Figure 8. The responses and the regression lines fit to the means of the responses for the subjects (‘BGH’, ‘CF’, ‘CM’, ‘HH’, ‘TA’). The solid lines represent the regression lines. The dash-dot lines represent ideal responses (i.e. response=source azimuth).

sponses for each subject ($r^2 > 0.99$ for all of the cases) show that the linear model is a valid representation of the bias. The slope of the regression lines averaged over the subjects is 0.752 ($std = 0.039$) and the y-intercept (i.e. the actual response bias at $0°$ azimuth) is 1.55 ($std = 0.915$). Figure 8 shows the bias and the linear regression model for all subjects. The solid lines denote the linear regression model of bias. The linear model of the measurement bias obtained for single-source situation in this experiment is used in the analysis of the results of experiment 2 for eliminating the effect of the perceptual bias related to the measurement method.

The standard deviations of responses for different stimulus conditions reflect sound source localisation accuracy for the given condition. All the standard deviations were within the interval $0.3° < \sigma < 2.7°$ (see Figure 9). These values are comparable to the results of previous studies [39, 40, 41]. The standard deviations are higher for $|\theta| > 0$ (i.e. sound sources at azimuth angles other than $0°$). The average standard deviation was $1.23°$.

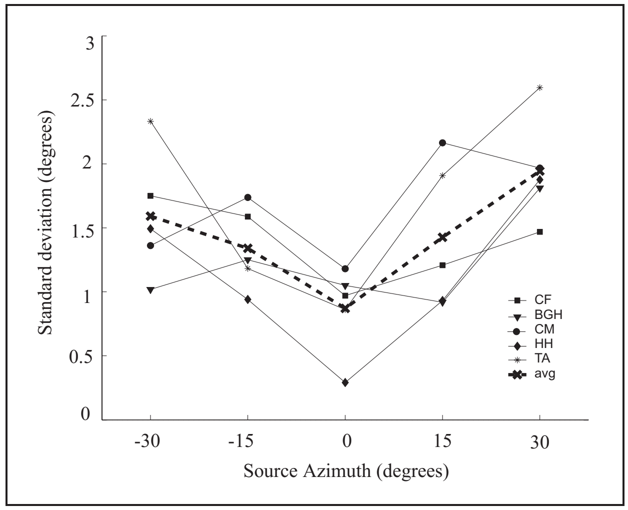

Figure 9. Standard deviations for all the subjects for the single source conditions and the standard deviations averaged across the subjects.

5. Experiment 2: Precedence effect

The motivation of the second experiment is to obtain basic localisation data under the precedence effect conditions. The observational data obtained in this experiment is used for displaying the application of the proposed model to obtain model parameters for typical precedence effect conditions.

5.1. Method

5.1.1. Subjects and apparatus

The subjects who participated in the first experiment also participated in the second experiment. This experiment also took place in the acoustically treated listening room previously described. The apparatus was the same as that used in Experiment 1.

5.1.2. Stimuli and procedure

The stimuli consisted of windowed white noise burst pairs of 6ms length with 2 ms rise-fall times (see Figure 10). The inter-onset delay was 4 ms and was kept constant throughout the test. The lag delay of 4 ms is shown by previous studies to result in localisation dominance for broadband stimulus pairs under similar listening conditions as utilised in this experiment. Each lagging noise burst was a spectrally coherent copy of its respective leading burst. However, the stimuli were generated randomly during the experiment, and the spectral characteristics of consecutive lead-lag pairs differed for every new presentation and/or repetition.

Each one of the loudspeakers was used for presenting the leading sound (i.e. loudspeakers 1, 2, 3, 4, and 5 in Figure 6). The lagging sound was presented only from one of the loudspeakers at $-30°$, $0°$ or $30°$ azimuth (i.e. loudspeakers 1, 3, and 5 in Figure 6).

The major strength of the proposed model in comparison with previous models is that it can account for the increase in the perceived width of the auditory event. A lagging signal incident from the same direction as the leading signal (such as ceiling and floor reflections) increases the localisation acuity in rooms [42] or in other words is rendered less audible [43]. Therefore we chose not to include same direction lead/lag pairs to reduce the test duration. Also note that, as t he lead and the lag signals were spectrally coherent, presenting the stimuli pair from the same

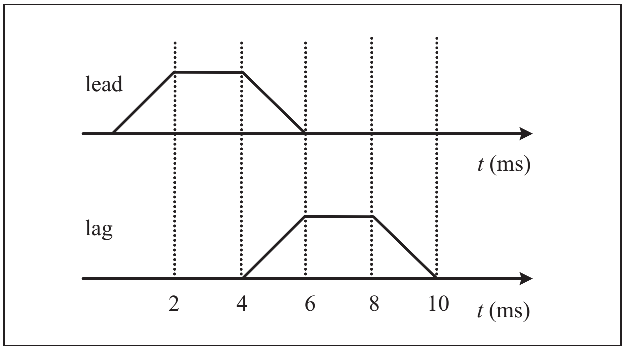

Figure 10. The temporal order of the stimuli used in the second experiment. Note that the stimulus pair is presented from two azimuthally separated loudspeakers.

Table I. The loudspeaker pairs used for stimulus presentation.

Stimulus	θ_{lead}	θ_{lag}	Δ_θ	LS pair (lead/lag)
1	$-30°$	$0°$	$-30°$	1/3
2	$-30°$	$30°$	$-60°$	1/5
3	$-15°$	$-30°$	$15°$	2/1
4	$-15°$	$0°$	$-15°$	2/3
5	$-15°$	$30°$	$-45°$	2/5
6	$0°$	$-30°$	$30°$	3/1
7	$0°$	$30°$	$-30°$	3/5
8	$15°$	$-30°$	$45°$	4/1
9	$15°$	$0°$	$15°$	4/3
10	$15°$	$30°$	$-15°$	4/5
11	$30°$	$-30°$	$60°$	5/1
12	$30°$	$0°$	$30°$	5/3

loudspeaker would in effect mean that a comb filtered version of the same signal is presented from a single direction. Such a stimulus does not convey any specific binaural features pertaining exclusively to the lagging signal that can be resolved by head movement. Therefore, that the inclusion of such conditions would introduce further complexities was another concern. The loudspeaker pairs that are used for pre senting the lead-lag stimulus pairs are listed in Table I. The order of presentation was fully randomized as in the previous experiment.

The task that the subjects had to carry out was similar to the first experiment. The subjects were instructed to turn their heads to face the more prominent auditory event if they happen to perceive two auditory events. Each lead-lag pair was presented 50 times. Thus, each subject gave 600 responses in total. The experiment took about three to four hours to complete for each subject. Regular breaks were given to prevent fatigue. At the end of each session, percussion music was played simultaneously from all loudspeakers to mark the end of the session.

5.2. Results

All the subjects verbally reported that the stimuli were "more difficult to localise, wider, and more spacious than the previous experiment" reflecting the well-known properties of the precedence effect conditions. This was also observed as an increase in the standard deviations of the responses in comparison with the single sound source conditions.

Before the analysis, the perceptual bias was corrected using the linear bias model obtained from the results of the previous experiment. This way, the relation associating the actual and the perceived locations of the sound sources can be inferred.

Figure 11 shows the localisation results for each subject and stimulus pair. Except for the subject 'TA' whose responses had a higher spread in general, the responses do not reflect huge inter-subject variability in terms of their standard deviations. However, the mean perceived locations differ slightly from subject to subject. This may be as a result of the differences in sensorimotor performance between individual subjects for the experimental task involving head movement.

An interesting feature of the obtained data is the asymmetry of the responses depending on the relative position of the leading sound source with respect to the lagging sound source. If the leading sound source is to the right of the lagging sound source the mean perceived direction is nearer to the direction of the leading sound source. If the leading sound source is to the left, the mean perceived position is pulled more towards the lagging sound source. The difference between the means of absolute shift observed in the mean perceived direction for right-leading and left-leading conditions was significant at $\alpha = 0.05$ level ($t = 10.29$, $df = 58$, $p \ll 0.001$). In other words, the localisation dominance was stronger for the leading sound sources originating from the right of the lagging sound source. A similar spatial asymmetry has also been reported by other authors [44, 45].

The responses for the lead-lag stimulus pairs were analysed by fitting two-component Gaussian mixtures. The EM algorithm [46, 47] with 100 iterations was used. The starting points of the algorithm were initialized with 10 small-EM runs. The algorithm was implemented as a part of the MIXMOD package [48] running under MATLAB.

The mixture proportions were selected to be equal ($p = 0.5$) as the leading and the lagging sound signals were spectrally coherent copies presented at the same level to the listener. The rationale of this choice was made clear earlier. The means and variances of the mixture components obtained using the EM analysis together with the Gaussian assumption are given in Figure 12.

Both hypotheses (i.e. Gaussian or Gaussian mixture) were tested with a Kolmogorov-Smirnov (KS) test. It was possible to reject neither. This suggests that the responses could have been interpreted to be single Gaussians as well. However, as suggested previously, what distinguishes the present model from that of Shinn-Cunningham *et al.* [7] which models localisation data as a single Gaussian is its ability to account for the increase in the apparent width observed with the precedence effect as well as to quantify the effect of the lagging sound source efficiently. Therefore, using a more complex model with a larger number of parameters can be justified.

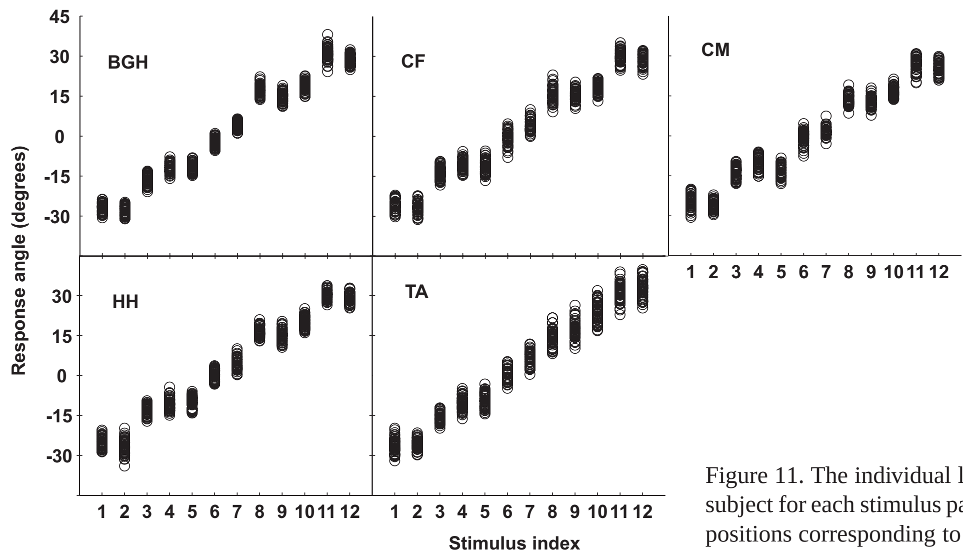

Figure 11. The individual localisation responses of each subject for each stimulus pair. See Table I for the lead/lag positions corresponding to each stimulus pair index.

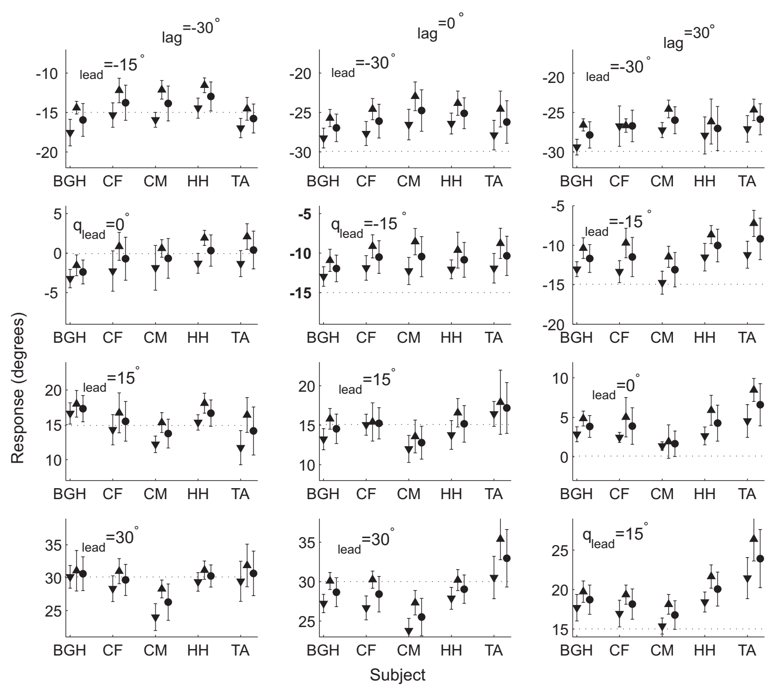

Figure 12. The mean values and standard deviations of distributions related to the two hypotheses of perceptual noise. The filled circles (●) represent the mean value for the single Gaussian hypothesis. The upward triangle (▲) and the downward triangle (▼) represent the mean values for the mixture components obtained using the EM analysis. The dotted line denotes the direction of the leading source.

6. Analysis and interpretation of the responses with the mixture model

It is virtually impossible in practice to design a sound source localisation experiment which would span the whole space of parameters that may influence the precedence effect. Therefore, the model parameters obtained from the responses given in the second experiment are valid only for the lead/lag delay (Δt) of 4ms with coherent, broadband lead/lag pairs presented at the same level. However, they provide information about a range of azimuth separations for the given time delay.

The Gaussian mixture model parameters (i.e. the component means and variances) obtained using the EM algorithm in the second experiment can be used to compute the precedence levels (c_{lead} and c_{lag}) and the related parameters/functions (Σ_c, Δ_c, A_p, and F_{mod}) for the tested stimulus conditions.

The precedence levels, c_{lead} and c_{lag} are calculated as:

$$\begin{aligned} c_{lead} &= \frac{2\mu_{lead} - (\theta_{lead} + \theta_{lag})}{(\theta_{lead} - \theta_{lag})}, \\ c_{lag} &= \frac{2\mu_{lag} - (\theta_{lead} + \theta_{lag})}{(\theta_{lead} - \theta_{lag})}. \end{aligned} \tag{24}$$

The mean value of the component nearer to the actual azimuth of the leading sound is indicated as μ_{lead} and the mean value of the other component is indicated as μ_{lag}.

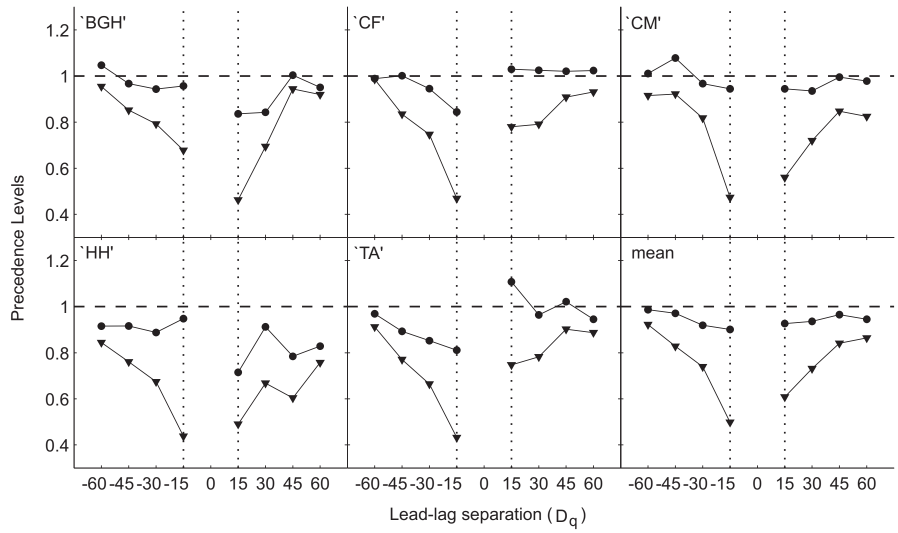

Figure 13. The precedence levels calculated from the mixture means for all subjects. Filled circles (●) represent c_{lead}, and filled triangles (▼) represent c_{lag}. The last panel shows the precedence levels averaged over subjects. Note that the number of values used for obtaining the average for each lead/lag separation is not the same.

The results show that the precedence levels for the leading sound are more stable and relatively higher than the precedence levels for the lagging sound. Furthermore, the precedence levels for the lagging sounds are lower for smaller lead/lag separations (see Figure 13). This means that, for smaller azimuth separations, influence of the lagging sound source on the perceived azimuth of the combined auditory event is greater.

Another interesting finding is that although the mean values of c_{lead} are mostly smaller than 1, responses to a number of stimuli pair yielded mean c_{lead} values larger than 1 for some of the subjects. This may be due to a negative shift of the mixture component related to the leading sound or an unintended response bias.

The precedence level difference was calculated using the lead and lag precedence levels. It is possible to obtain the A_p values directly from the calculated precedence level values. A_p represents the widening of the auditory event under the precedence effect in comparison with the apparent width of the auditory event for the single sound source condition. The calculated A_p values represent the predicted widening of the auditory event for the current experimental setting. The comparison of these with the increase in the response standard deviations reveals that the value predicted by the model is in agreement with the increase in the perceptual variation. The broken line in Figure 14 represents the increase in the response standard deviations (i.e. the difference of standard deviations observed in the first and the second experiments) averaged over the subjects. The bold line represents the calculated A_p values.

The Δ_c data was exponential in appearance. Therefore, two exponential functions $e^{k\Delta_\theta + l}$ were regressed (i.e. linear regression was applied after a Box-Cox transform) on the Δ_c data for the regions where $\Delta_\theta < 0$ and $\Delta_\theta > 0$. For $\Delta_\theta < 0$, $k = 0.0343$ and $l = -0.5722$ ($r^2 > 0.96$). For $\Delta_\theta > 0$ $k = -0.0305$ and $l = -0.6892$ ($r^2 > 0.99$). This suggests a good fit to the obtained Δ_c data. It must be noted that the asymmetry of these exponential functions was a direct result of the observed left/right asymmetry in the responses.

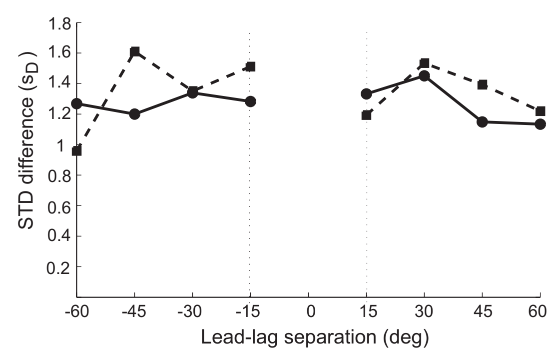

Figure 14. The increase in the apparent width of the auditory event. The broken line shows the actual increase (i.e. arithmetic difference of the standard deviations of responses to single source and lead-lad conditions). The bold line shows the predicted increase A_p.

The precedence level difference function, Δ_c, is a function of Δ_θ. This function was used for calculating the F_{mod} function for the Δt value used in the experiment (i.e. $\Delta t = 4$ms). As explained in equation (23), $F_{mod} < 2\sigma$, is a sufficient condition for the mixture to be unimodal for $p = 0.5$. Therefore, we can suggest that for lead/lag separations where $F_{mod} > 2\sigma$, the lagging sound has the potential to change the modality of the responses. The lag discrimination MAA for a lead/lag pair, is expected to lie in the Δ_θ interval on which F_{mod} is greater than or equal to twice the standard deviaion.

Figure 15(a) shows the F_{mod} function calculated using the exponential model of Δ_c. The function is not perfectly symmetric with respect to $\Delta_\theta = 0°$ as the value of the exponent, k, is different for $\Delta_\theta > 0$ and $\Delta_\theta < 0$. This asymmetry arises due to the mentioned left-right asymme-

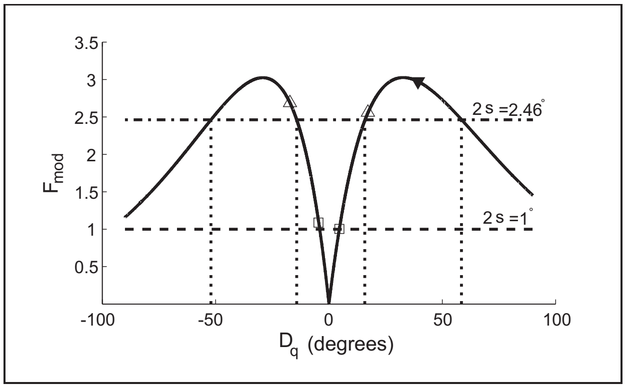

Figure 15. The modality function, F_{mod}, calculated with the exponential model of Δ_c. Triangular markers show lag discrimination MAA data from Litovsky and Shinn-Cunningham [13], and square markers represent lag discrimination MAA data from Perrott and Pacheco [15].

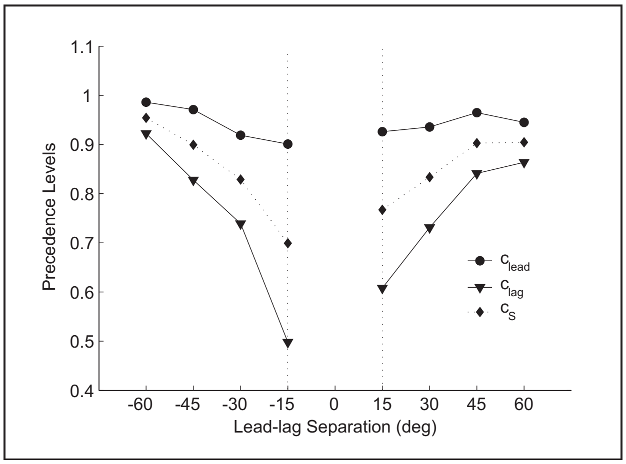

Figure 16. The lead and lag precedence levels (c_{lead} and c_{lag}) of the present study and the single precedence level measure, c_S.

try in the obtained responses. The upper horizontal broken line shows $2\sigma = 2.46°$ (i.e. twice the standard deviation for a single sound source). When the modality function is greater than this value, the response distribution is potentially bimodal. The lead-lag separation corresponding to the limit determines the bounds within which the lag discrimination MAA is expected to lie. The lower horizontal broken line shows $2\sigma = 1°$ that corresponds to localisation acuity at free-field conditions. The vertical dotted lines show the boundaries for the mixture to be unimodal. The triangle and square markers represent lag discrimination MAA values obtained from Litovsky and Shinn-Cunningham [13] and Perrott and Pacheco [15] respectively. These will be discussed shortly in the next section.

7. Relation to Previous Studies

The concept of precedence level functions in the proposed model was inspired by an earlier study by Shinn-Cunningham *et al.* [7] in that the perceptual weights assigned to each sound source (presented over headphones) was assessed using the localisation data obtained using an acoustic pointer. The stimuli used in that study was random white noise samples. Two time delays were used; $\Delta t = 1$ ms for strong precedence and $\Delta t = 10$ ms for weak precedence. The presentation levels were 80 dB and 110 dB.

The data obtained in the present experiments are not directly comparable to the results reported by Shinn-Cunningham *et al.* as the time delay (1 ms and 10 ms *vs.* 4 ms), presentation medium (headphone vs. loudspeaker), and the presentation level were not the same. In the mentioned study, a single interaural cue (ITD) was investigated. However, some general comments can be made about the similarity of the results.

It may be suggested that the time delay used in the present experiment resulted in a precedence effect that was neither as strong as the 1 ms delay nor as weak as the 10 ms delay used in the Shinn-Cunningham *et al.* study. The precedence level measure, c_S, reported by Shinn-Cunningham *et al.* can be related to the precedence levels in the present study as:

$$c_S = \frac{1}{2}\left(c_{lead} + c_{lag}\right). \qquad (25)$$

The results presented in that study show that the precedence level is generally around 1 for the strong precedence condition, and smaller than 1 for the weak precedence condition. Figure 16 presents the lead and lag precedence levels (c_{lead} and c_{lag}) and the single precedence measure c_S calculated from the results of the second experiment. It may be stated that the precedence level is stronger for larger azimuth separations for the stimulus and presentation conditions in the present experiments. This is in agreement with Shinn-Cunningham *et al.*'s general remark that the precedence level values increase with the absolute difference between the ITDs of the leading and the lagging sounds.

In another study, Litovsky and Shinn-Cunningham investigated fusion, localisation dominance and discrimination suppression properties of the precedence effect for 1 ms long Gaussian white noise stimuli with abrupt onsets and offsets presented at three different ITD values (-400μs, 0, 400μs) and six different inter-click intervals (1, 2, 3, 5, 10, and 15 ms) [8]. It is again not possible to make a direct comparison with the localisation experiment presented in this paper. However, the data presented in that study allows the lead and lag precedence levels to be calculated by using the responses to lead and lag discrimination tasks. Figure 17 shows the calculated precedence levels. Note that the precedence levels for the $\tau_{lead} = \tau_{lag} = \pm 400\mu s$ are calculated by using only τ_{lead} for c_{lead} and τ_{lag} for c_{lag} such that:

$$\begin{aligned} c_{lead} &= \frac{2\mu_{lead}}{\tau_{lead}} - 1, \\ c_{lag} &= 1 - \frac{2\mu_{lag} - \tau_{lead}}{\tau_{lag}}. \end{aligned} \qquad (26)$$

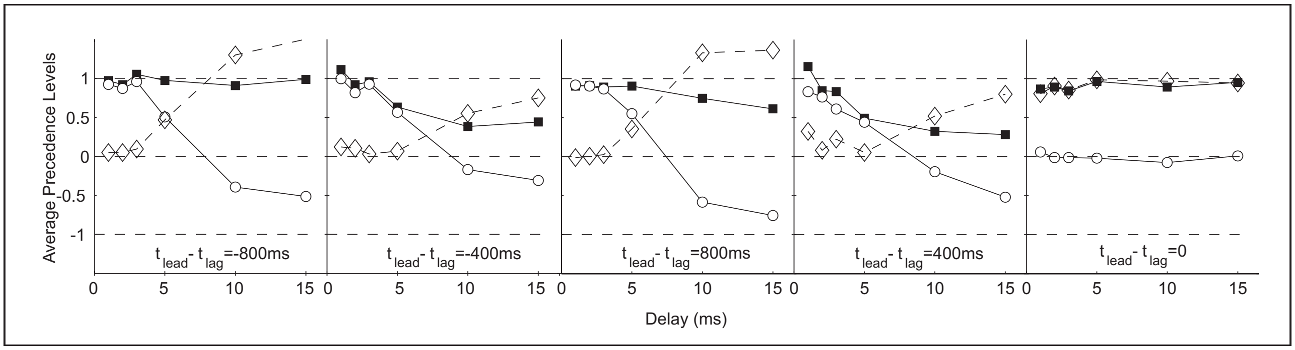

Figure 17. The lead and lag precedence levels (c_{lead} and c_{lag}) averaged across subjects calculated from the data given in [8] . The filled squares represent c_{lead}, the circles represent c_{lag}, and the diamonds represent Δ_c.

The data, although not perfectly in alignment with our initial assumption of lead and lag precedence levels shows similarity to the ideal precedence level functions presented in Figure 4. Although the shape of the modality functions resembled the modality function that we have obtained, it was not possible to predict the reported ITD jnd for lag discrimination suppression from the modality functions obtained from these data. This may be due to the fact that despite broadband stimuli was used in that study, ITD is a useful cue only for frequencies below about $1.5\,$kHz [1].

The previous studies we have mentioned above investigated the ITD of the lead and lag with the consideration that the precedence effect is more of a temporal weighting process. However, as is well-known [49], localisation as opposed to lateralisation is a more complex task involving not only interaural differences, but also other cues such as the spectral shaping applied onto the sound wave by the listener's pinnæ, head and torso, as well as the reflections present in the listening environment. In that sense, the results from these studies do not correspond to a localisation task in a natural binaural listening scenario.

Litovsky and Macmillan [13] used a similar apparatus and stimuli to the experiments presented in this paper. Pink noise bursts, low-pass filtered at 8500Hz as stimuli were presented at a presentation level of $50-52$dBA over loudspeakers in a sound deadened room. The time delay and the rise-fall times of the temporal envelope used in that study is the same as the ones used in the present study. The authors investigated whether the azimuth of presentation for the leading or the lagging sound source affects the lag discrimination MAA. Two experimental tasks were employed for evaluating the effect of a conditioner (i.e. standard) source presented from the leading source. The task in which the conditioner source was not employed (i.e. nostandard task) is comparable to the task employed in the second experiment of this paper. The major difference between that study and the present one is the experimental paradigm. Litovsky and Macmillan used a discrimination task with an adaptive procedure, while the present study employed a pointing task. However, the reported mean lag discrimination MAA values may be compared with the bound(s) of lag discrimination MAA given in equation (23). Litovsky and Macmillan report $17.17°$ and $39.28°$ as the lag discrimination suppression MAA for the leading sound source at $0°$ and $50°$ azimuth respectively. These azimuth separations satisfy the $F_{mod} > 2\sigma$ condition for $2\sigma \approx 2.46°$ (see the triangular markers in Figure 15). Here, it is assumed that the response variability of a single sound source at $0°$ and $50°$ were similar to the variability of responses obtained in the second experiment of the present study.

Perrott and Pacheco [15] report lag discrimination MAA data measured under free-field conditions for a wide range of inter-stimulus delays. This study was a lag discrimination experiment utilizing loudspeakers and using an adaptive procedure to obtain the threshold values for the MAA. The employed stimuli were band-pass filtered ($500-8000$Hz) pink noise presented at 50dBA with various lead/lag delays presented from loudspeakers. The reported lag discrimination MAA is around $4.5°$ for a lead/lag delay of 4ms. Given that humans can localise sound sources in free-field with $1°$ accuracy [39], we suggest that the MAA reported by Perrott and Pacheco also satisfies the condition $F_{mod} > 2\sigma$ if $2\sigma \approx 1°$ (see the square markers in Figure 15).

8. Discussion

The Gaussian mixture framework proposed in this paper is suitable for modelling the temporal and localisational properties of the precedence effect in the utilised audiovisual pointing task. To the best of the authors' knowledge, the presented model is unique in the sense that it allows the quantitative assessment of the precedence effect with respect to its effects on the perceived width of the auditory event. Although this effect has been well-known, its context in previous models was not fully explained.

The employment of the hybrid laser pointing task that required the subjects not only to point at but also to face the perceived auditory event allowed the direct employment of natural sound source localisation strategies by the subjects.

The proposed model is more complex in comparison with previous observer weighting models of the precedence effect with regard to the number of variables. The parsimony of the model can be questioned. However, given that the number of physical parameters that can influence certain properties of the precedence effect is plentiful, the model is a reasonable framework which can be used to assess the fusion, localisation dominance, and lag discrimination suppression aspects together with the widening of the auditory event. Furthermore, the precedence effect is modelled as a very simple core process consisting of weighted addition. In that sense, the model can be considered as a parsimonious model, as there is a good justification for the number of parameters.

Although the model does not have a physiological basis, it is not in contradiction with the physiological models of the precedence effect based on contralateral inhibition. We hypothesize without further consideration that, the perceptual noise and the temporal window [50] of the auditory system addressed in recent binaural processing models [51] may account for the data obtained in this study. However, our model has less in common with the more recent research [52] suggesting that the precedence effect may be linked with the peripheral processes in the auditory system.

The present model is shown to be plausible for the classical precedence effect scenarios (at least for the experimental setting used). The dynamic aspects of the precedence effect such as the binaural adaptation, buildup of the precedence effect, and the Clifton effect can also be integrated to the model by defining the mixture proportion, p, as a function of stimulus repetition. It should however be noted that, the major limitations of the model are its inability to account for localisation data when both the leading and the lagging sources are on the median plane, and when the leading and lagging sounds are not spectrally coherent.

Other than this, the piecewise continuous precedence level functions are oversimple generalizations and have to be replaced with more realistic functions. However, the model, as it stands, can accommodate the basic well-known temporal and localisational properties of the precedence effect, and is a suitable tool for analysing the data obtained in observational studies of the precedence effect.

The reported sound localisation experiments are limited in the sense that only a single type of stimulus (white noise bursts) and a single lead/lag delay (4ms) was used, the lead and lag were spectrally coherent, and were presented at the same level. For a formal verification of the usefulness of the proposed model, further experiments need to be carried out for different types of stimuli, presented at different relative levels, at different azimuth separations, and more importantly with different time delays.

It must be remarked that it is possible to generalize the proposed model for elevations other than 0° by using a mixture of bivariate Gaussians for modelling the precedence effect. However, it is much harder to display the applicability and obtain the parameters of such a model as it would require a substantially larger set of responses to be sampled because of the well-known *curse of dimensionality* [53]. The same accuracy as attained for the univariate model proposed in this paper would require $50^2 = 2500$ responses per stimulus condition tested for a bivariate model which would limit the practicality of such an experiment greatly.

9. Conclusions

A Gaussian mixtures approach for modelling sound source localisation under the precedence effect conditions was presented. The precedence effect was modelled as a simple core process that in effect modifies only the perceived locations of the sound sources and not the perceptual noise attributed to each. Certain properties of the precedence effect can be captured using the proposed model:

1. The temporal properties of the precedence effect are modelled by defining lead and lag precedence levels for the leading and the lagging sound sources as functions of the time delay between the leading and the lagging sounds. These precedence levels define the mean values of the mixture components.
2. The effect of the lagging sound source is quantified by using separate weights for leading and lagging sound sources.
3. The widening of the auditory event is a direct result of the properties of Gaussian mixtures such that the variance of a Gaussian mixture is larger than individual component variances.
4. Minimum bounds of the angular interval, within which the lag discrimination MAA is expected to lie, can be predicted by utilizing a modality function.

Two subjective localisation tests were presented which exemplified the applicability of the model to data obtained in a precedence effect scenario. The obtained data shows a particular left/right asymmetry that is in agreement with previous studies of the precedence effect. The model suggests that for a smaller angular separation between leading and lagging source, the lagging source has a greater influence on the perception of the auditory event.

Acknowledgment

This work was supported by a SPUR (Support Programme for University Research) grant. The authors are grateful to Dr Banu Günel Hacıhabiboğlu and Mr Chris Corrigan for the valuable discussions and the technical help they provided during this study. The authors also thank the associate editor, Prof Birger Kollmeier, and the three anonymous reviewers for their feedback and insightful comments on a previous version of this paper.

References

[1] J. Blauert: Spatial hearing: The psychophysics of human sound localization. MIT Press, Cambridge, MA, USA, 1997.

[2] R. Y. Litovsky, H. S. Colburn, W. A. Yost, S. J. Guzman: The precedence effect. Journal of the Acoustical Society of America **106** (1999) 1633–1654.

[3] P. M. Zurek: The precedence effect and its possible role in the avoidance of interaural ambiguities. Journal of the Acoustical Society of America **67** (1980) 952–964.

[4] H. Wallach, E. B. Newman, M. R. Rosenzweig: The precedence effect in sound localization. American Journal of Psychology **62** (1949) 315–336.

[5] H. Haas: Über den einfluss des einfachechos auf die horsamkeit von sprache. Dissertation. University of Göttingen, Göttingen, Germany, 1949.

[6] R. L. Freyman, R. K. Clifton, R. Y. Litovsky: Dynamic processes in the precedence effect. Journal of the Acoustical Society of America **90** (1991) 874–884.

[7] B. G. Shinn-Cunningham, P. M. Zurek, N. I. Durlach: Adjustment and discrimination measurements of the precedence effect. Journal of the Acoustical Society of America **93** (1993) 2923–2932.

[8] R. Y. Litovsky, B. G. Shinn-Cunningham: Investigation of the relationship among three common measures of precedence: Fusion, localization dominance, and discrimination suppression. Journal of the Acoustical Society of America **109** (2001) 346–358.

[9] P. L. Divenyi: Binaural suppression of nonechoes. Journal of the Acoustical Society of America **91** (1992) 1078–1084.

[10] W. A. Yost, D. R. Soderquist: The precedence effect: Revisited. Journal of the Acoustical Society of America **76** (1984) 1377–1383.

[11] X. Yang, D. W. Grantham: Echo suppression and discrimination suppression aspects of the precedence effect. Perception and Psychophysics **59** (1997) 1108–1117.

[12] K. Saberi, D. R. Perrott: Lateralization thresholds obtained under conditions in which the precedence effect is assumed to operate. Journal of the Acoustical Society of America **87** (1990) 1732–1737.

[13] R. Y. Litovsky, N. A. Macmillan: Sound localization precision under conditions of the precedence effect: Effects of azimuth and standard stimuli. Journal of the Acoustical Society of America **96** (1994) 752–758.

[14] R. Y. Litovsky: Developmental changes in the precedence effect: Estimates of minimum audible angle. Journal of the Acoustical Society of America **102** (1997) 1739–1745.

[15] D. R. Perrott, S. Pacheco: Minimum audible angle thresholds for broadband noise as a function of the delay between the onset of the lead and lag signals. Journal of the Acoustical Society of America **85** (1989) 2669–2672.

[16] D. R. Perrott, K. Marlborough, P. Merrill: Minimum audible angle thresholds obtained under conditions in which the precedence effect is assumed to operate. Journal of the Acoustical Society of America **85** (1989) 282–288.

[17] P. M. Zurek: The precedence effect. Directional Hearing, New York, USA, 1987, W. A. Yost, G. Gourevitch (eds.), Springer-Verlag, 85–105.

[18] B. Rakerd, W. M. Hartmann: Localization of sound in rooms, II: The effects of a single reflecting surface. Journal of the Acoustical Society of America **78** (1985) 524–533.

[19] R. K. Clifton: Breakdown of echo suppression in the precedence effect. Journal of the Acoustical Society of America **82** (1987) 1834–1835.

[20] D. D. McCall, R. L. Freyman, R. K. Clifton: Sudden changes in spectrum of an echo cause a breakdown of the precedence effect. Perception and Psychophysics **60** (1998) 593–601.

[21] R. K. Clifton, R. L. Freyman, R. Y. Litovsky, D. McCall: Listeners' expectations about echoes can raise or lower echo threshold. Journal of the Acoustical Society of America **95** (1994) 1525–1533.

[22] R. K. Clifton, R. L. Freyman, J. Meo: What the prececedence effect tells us about room acoustics. Perception and Psychophysics **64** (2002) 180–188.

[23] T. Djelani, J. Blauert: Modeling direction-specific build-up of the precedence effect. Proc. of Forum Acusticum Sevilla (CD-ROM), Sevilla, Spain, Madrid, Spain, 2002, EAA.

[24] P. X. Joris, P. H. Smith, T. C. T. Yin: Coincidence detection in the auditory system: 50 years after Jeffress. Neuron **21** (1998) 1235–1238.

[25] W. Lindemann: Extension of a binaural cross-correlation model by contralateral inhibition. I. Simulation of lateralization for stationary signals. Journal of the Acoustical Society of America **80** (1986) 1608–1622.

[26] W. Lindemann: Extension of a binaural cross-correlation model by contralateral inhibition. II. The law of the first wavefront. Journal of the Acoustical Society of America **80** (1986) 1623–1630.

[27] R. Y. Litovsky, B. Rakerd, T. C. T. Yin, W. M. Hartmann: Psychophysical and physiological evidence for a precedence effect in the median sagittal plane. Journal of Neurophysiology **77** (1997) 2223–2226.

[28] R. M. Dizon, R. Y. Litovsky: Localization dominance in the median-sagittal plane: Effect of stimulus duration. Journal of the Acoustical Society of America **115** (2004) 3142–3155.

[29] B. G. Berg: Analysis of weights in multiple observation tasks. Journal of the Acoustical Society of America **86** (1989) 1743–1746.

[30] M. A. Stellmack, R. H. Dye, Jr., S. J. Guzman: Observer weighting of interaural delays in source and echo clicks. Journal of the Acoustical Society of America **105** (1999) 377–387.

[31] G. C. Stecker, E. R. Hafter: Temporal weighting in sound localization. Journal of the Acoustical Society of America **112** (2002) 1046–1057.

[32] K. Saberi: Observer weighting of interaural delays in filtered impulses. Perception and Psychophysics **58** (1996) 1037–1046.

[33] E. Acuña: On the modality of the finite mixtures of normal densities. Proc. of 52^{nd} Session of the International Statistical Institute, Voorburg, the Netherlands, 1999, International Statistical Institute.

[34] J. Behboodian: On the modes of a mixture of two normal distributions. Technometrics **12** (1970) 131–139.

[35] D. M. Titterington, A. F. M. Smith, U. E. Makov: Statistical analysis of finite mixture distributions. John Wiley and Sons, Chichester, UK, 1985, (Series in Probability and Mathematical Statistics).

[36] D. R. Perrott, H. Ambarsoom, J. Tucker: Changes in head position as a measure of auditory localization performance: Auditory psychomotor coordination under monaural and binaural listening conditions. Journal of the Acoustical Society of America **82** (1987) 1637–1645.

[37] J. Lewald, W. H. Ehrenstein: Auditory-visual spatial integration: A new psychophysical approach using laser pointing to acoustic targets. Journal of the Acoustical Society of America **104** (1998) 1586–1597.

[38] S. Choisel, K. Zimmer: A pointing technique with visual feedback for sound source localization experiments. *Presented at the* 115^{th} *Convention of the Audio Engineering Society*, preprint 5904, New York, NY, USA, October 2003.

[39] A. W. Mills: On the minimum audible angle. Journal of the Acoustical Society of America **30** (1958) 237–246.

[40] D. W. Grantham, B. W. Y. Hornsby, E. A. Erpenbeck: Auditory spatial resolution in horizontal, vertical, and diagonal

planes. Journal of the Acoustical Society of America **114** (2003) 1009–1022.

[41] W. M. Hartmann, B. Rakerd: On the minimum audible angle—A decision theory approach. Journal of the Acoustical Society of America **85** (1989) 2031–2041.

[42] W. M. Hartmann: Localization of sound in rooms. Journal of the Acoustical Society of America **74** (1983) 1380–1391.

[43] S. E. Olive, F. E. Toole: The detection of reflections in typical rooms. Journal of the Audio Engineering Society **37** (1989) 539–553.

[44] D. W. Grantham: Left–right asymmetry in the buildup of echo suppression in normal hearing adults. Journal of the Acoustical Society of America **99** (1996) 1118–1123.

[45] K. Saberi, J. V. Antonio, A. Petrosyan: A population study of the precedence effect. Hearing Res. **191** (2004) 1–13.

[46] A. P. Dempster, N. M. Laird, D. B. Rubin: Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society (B) **39** (1977) 1–38.

[47] R. A. Redner, H. F. Walker: Mixture densities, maximum likelihood and the EM algorithm. SIAM Review **26** (1984) 195–239.

[48] C. Biernacki, G. Celeux, G. Govaert: Choosing starting values for the EM algorithm for getting the highest likelihood in multivariate Gaussian mixture models. Computational Statistics & Data Analysis **41** (2003) 561–575.

[49] G. Plenge: On the differences between localization and lateralization. Journal of the Acoustical Society of America **56** (1974) 944–951.

[50] B. C. J. Moore, B. R. Glasberg, C. J. Plack, A. K. Biswas: The shape of ear's temporal window. Journal of the Acoustical Society of America **83** (1988) 1102–1116.

[51] J. Breebaart, S. van de Par, A. Kohlrausch: Binaural processing model based on contralateral inhibition. I. model structure. Journal of the Acoustical Society of America **110** (2001) 1074–1088.

[52] K. Hartung, C. Trahiotis: Peripheral auditory processing and investigations of the "precedence effect" which utilize successive transient stimuli. Journal of the Acoustical Society of America **110** (2001) 1505–1513.

[53] R. E. Bellman: Adaptive control processes. Second ed. Princeton University Press, Princeton, NJ, USA, 1961.