Object Based Audio Reproduction using a Listener-Position Adaptive Stereo System

Marcos F. Simón Gálvez^{1*}, Dylan Menzies¹, Russell Mason²
and Filippo Maria Fazi¹

May 12, 2016

- ¹ Institute of Sound and Vibration Research, University of Southampton, Southampton, Hampshire, SO17 1BJ, United Kingdom
- ² Institute of Sound Recording, University of Surrey, Guildford, Surrey, GU2 7XH, United Kingdom

Abstract

This work introduces a listener-position adaptive stereo reproduction system that allows for the reproduction of 2D object-based audio and for a more accurate localisation when the listener is located outside the sweet spot. The adaptation is composed of two parts; a compensation system that updates the loudspeakers feeds so that the loudspeaker input signals are delivered to the listener with the same magnitude and phase independently of the listening position, as it would occur in a symmetric listening configuration, and an object-based rendering system using conventional panning algorithms. Robustness simulations show that an accurate localisation is possible when the audio objects are panned in the angle seen between the listener and the two loudspeakers. This has been further assessed by objective and subjective localisation experiments.

*E-mail correspondence should be addressed at M.F.Simon-Galvez@soton.ac.uk

1 Introduction

Even after the introduction of multichannel audio, stereo is still the most common audio reproduction set up that can be found in households. As with more complex reproduction systems involving a higher number of loudspeakers, the localisation of virtual sound images is only accurate in the so called *sweet spot*, a small area located, for a stereo system, on the plane of symmetry between both loudspeakers. If the listener, however, moves away from the sweet spot towards one of the loudspeakers, the resulting interaural level difference (ILD) and interaural time difference (ITD) differ from those found when the listener is at the same distance from all loudspeakers. This causes the perceived position of the audio image sources to shift and eventually collapse towards the closest loudspeaker, thus compromising the rendering of spatial sound and the initial intention of the sound producer [1]. Such an effect is likely to happen, as it is unlikely that the listener is always located in the sweet spot for most typical domestic listening situations.

In the past, compensation for off-sweet spot listening was performed using various techniques for virtually expanding the sweet spot size, by adjusting the directivity of the loudspeakers [1], for example, or by inverting the source input feeds [2]. Other approaches are based on adaptive methods, wherein the loudspeaker inputs are adjusted depending on the position of the listener with respect to that of the loudspeakers. An example of this approach was developed by Okuro and Kajikawa [3] wherein the loudspeaker inputs were adjusted with information gained via subjective experiments. The authors showed that such approach improved the localisation accuracy for a wide range of listening positions.

Adaptive stereo reproduction methods based on the applications of delays to both loudspeaker inputs are those presented by Merchel and Groth [4] and by Momose et al. [5]. Merchel and Groth used free-field time compensation for listener position misalignments, delaying the input of the loudspeaker to which the listener is closest to avoid the prece-

dence effect. The authors presented simulations using a binaural localisation model, and showed that such method contributed to a better localisation. The work of Momose et al. [5] used a similar approach to that of Merchel and Groth, and introduced an additional ILD target based on head-related impulse response (HRIR) functions, which allowed for better localisation for off-axis listening positions. Both these approaches, however, did not consider gain compensation to account for the distance attenuation of the loudspeaker signals.

With the introduction of new object-based audio technologies [6], it is possible to control the position of virtual sources with respect to that of the listener, for example by employing amplitude panning methods such as vector base amplitude panning (VBAP) [7, 8]. Such techniques adjust the loudspeaker input signals to give a certain perceived localisation, and are usually calculated assuming the audio signal to have a broadband nature. This causes some issues as localisation of image sources rendered by amplitude panning is not frequency independent and differs depending on the relative size of the radiated wavelength with respect to listener’s head. This was studied by Gerzon [9], who proposed different metrics to assess the localisation of amplitude panned images for different frequency bands. In the original VBAP formulation, the loudspeaker feeds are normalised following an incoherent addition assumption that does not hold at low frequencies, where the pressure waves of the active loudspeakers add coherently. Apart from the fact that this leads to incorrect loudness along the frequency band [10], it also limits the compensation for asymmetric listening. To avoid these issues, the approach presented here uses a coherent interference assumption for the loudspeaker input normalisation at low frequencies and an incoherent assumption at high frequencies based on vector base intensity panning (VBIP) [11], hence allowing the low frequency sound image positions to come for angular positions outside of the loudspeaker region.

74 This paper presents an approach to perform listening-position-adaptive reproduction and
75 continues previous work in the same area [12, 13]. Although the motivating application is
76 stereo sound reproduction, the proposed formulation can be applied to sound reproduction
77 systems involving larger numbers of loudspeakers. Differently from previous approaches
78 that have only compensated the loudspeaker signals to the listener’s position, in the system
79 proposed herein amplitude panning techniques are applied to pan different audio sources,
80 which adjust both loudspeaker signals based on listener movement. This keeps the per-
81 ceived localisation of point-source or plane-wave like audio objects static and provides the
82 listener with *motion parallax* cues, hence creating a greater 3-dimension awareness, as
83 well as perception of depth. To test the real time performance of the proposed approach,
84 this was implemented as a MAX MSP 6 application using a Microsoft Kinect sensor to
85 estimate the position of the listener with respect to the loudspeakers. The structure of
86 the paper is organised as follows: Section 2 introduces the compensation approach used
87 to correct the loudspeaker signals in amplitude and time for asymmetrical listening po-
88 sitions. Section 3 presents the formulation that is used to pan the audio objects and
89 the frequency-dependent normalisation that has been employed. The theory for source
90 localisation is then introduced, and simulations which estimate the ability of the strategy
91 to place audio objects at different positions are also shown. Section 4 introduces the
92 real-time implementation of the system, with special attention to the robustness against
93 errors in the tracking system and loudspeaker position miss-alignment, which is studied
94 by means of simulations and ITD measurements. Section 4 also includes the results of a
95 subjective study wherein the perceived localisation is assessed by listening tests.

96

97 2 Listener Position Compensation

98 The proposed method is based on the assumption that each loudspeaker behaves like an
99 ideal point-monopole source. The free-field transfer function of each loudspeaker is used

100 to control the soundfield at a given point in the space. Compensation for asymmetrical
 101 listening positions is introduced by modifying the loudspeaker input signals so that these
 102 arrive at the same time and with the intended amplitude.

103

104 2.1 Overview

Figure 1: Definition of vectors and quantities used for the adaptive stereo. The origin of the coordinates system, $C = (0, 0)$, coincides with the location of the tracking system.

105 Fig. 1 shows the reference system that is employed to find the relative position of the
 106 listener with respect to the loudspeakers. The vector notation used is as follows: a given
 107 vector $\mathbf{x} = [x_1, x_2]$ is defined by a magnitude $\|\mathbf{x}\|$ and by an unit vector $\hat{\mathbf{x}}$, e.g., $\mathbf{x} = \|\mathbf{x}\|\hat{\mathbf{x}}$.
 108 The vectors $\mathbf{r}_L$ and $\mathbf{r}_R$ define the location of the loudspeakers with respect to the origin
 109 of the coordinates system and the relative vectors $\mathbf{r}'_L$ and $\mathbf{r}'_R$ define the location of each
 110 loudspeaker relative with respect to the centre of the listener's head. These are formulated
 111 as

$$\mathbf{r}'_L(\mathbf{x}) = \mathbf{r}_L - \mathbf{x}, \quad (1)$$

112 and

$$\mathbf{r}'_R(\mathbf{x}) = \mathbf{r}_R - \mathbf{x}. \quad (2)$$

113 Using these vectors, the total field generated by both loudspeakers at a listener position

114 $\mathbf{x}$ is given by

$$p(\mathbf{x}, \omega) = \frac{\tilde{s}_L(\omega)}{4\pi\|\mathbf{r}_L(\mathbf{x})\|} e^{-jk(\|\mathbf{r}_L(\mathbf{x})\|)} + \frac{\tilde{s}_R(\omega)}{4\pi\|\mathbf{r}_R(\mathbf{x})\|} e^{-jk(\|\mathbf{r}_R(\mathbf{x})\|)}, \quad (3)$$

115 where ω is the angular frequency, $k = \omega/c_0$ the wavenumber, $c_0=344$ m/s is the speed of
 116 sound and $e^{-j\omega t}$ represents a positive time delay. The symbols $\tilde{s}_L(\omega)$ and $\tilde{s}_R(\omega)$ represent
 117 the compensated loudspeaker signals.

118

119 2.2 Compensation Algorithm

120 In order to provide an accurate localisation of virtual audio objects regardless of the lis-
 121 tener position, the first requirement is to ensure that the soundwaves propagating from
 122 the two loudspeakers arrive at the listener position at the same time instant and with an
 123 intended amplitude, as it occurs in a symmetric listening situation in which both loud-
 124 speakers are at the same distance from the listener. This is achieved by applying a gain
 125 and a delay to the loudspeaker signals to compensate for the distance attenuation and
 126 time of travel of the waves.

127

128 The gain equalisation is performed by multiplying both input signals $s_L(\omega)$ and $s(\omega)$ by
 129 $\|\mathbf{r}_L\|$ and $\|\mathbf{r}_R\|$ respectively. The delay compensation can be applied depending on the
 130 relative sign of the difference $\|\mathbf{r}_L(\mathbf{x})\| - \|\mathbf{r}_R(\mathbf{x})\|$. If the listener is closer to the left source,
 131 the compensation is applied as

$$\tilde{s}_L(\omega) = s_L(\omega)\|\mathbf{r}_L(\mathbf{x})\|e^{-jk(\|\mathbf{r}_R(\mathbf{x})\| - \|\mathbf{r}_L(\mathbf{x})\|)},$$

$$\text{and } \tilde{s}_R(\omega) = s_R(\omega)\|\mathbf{r}_R(\mathbf{x})\|, \quad (4)$$

132 being $s_L(\omega)$ and $s(\omega)$ the left and right input signals respectively. If on the other hand,
 133 the listener is closer to the source on the right hand side, the input feeds are given by

$$\begin{aligned}\tilde{s}_R(\omega) &= s_R(\omega) \|\mathbf{r}_R(\mathbf{x})\| e^{-jk(\|\mathbf{r}_L(\mathbf{x})\| - \|\mathbf{r}_R(\mathbf{x})\|)}, \\ \text{and } \tilde{s}_L(\omega) &= s_L(\omega) \|\mathbf{r}_L(\mathbf{x})\|.\end{aligned}\tag{5}$$

134 The frequency independent delays $e^{-jk(\|\mathbf{r}_R(\mathbf{x})\| - \|\mathbf{r}_L(\mathbf{x})\|)}$ and $e^{-jk(\|\mathbf{r}_L(\mathbf{x})\| - \|\mathbf{r}_R(\mathbf{x})\|)}$ are imple-
 135 mented in real time by applying a fractional delay [14] of $\delta(t - (\|\mathbf{r}_R(\mathbf{x})\| - \|\mathbf{r}_L(\mathbf{x})\|))$ and
 136 $\delta(t - (\|\mathbf{r}_L(\mathbf{x})\| - \|\mathbf{r}_R(\mathbf{x})\|))$.

137

138 3 Object Panning

139 Amplitude panning algorithms apply different gain values to the input signals of a set
 140 of loudspeakers to adjust the perceived location of a virtual source. For a given mono
 141 object with associated signal $s_0(\omega)$ that is intended to be placed in a certain position in
 142 the space given by the vector $\mathbf{p}'$, it is possible to create a set of loudspeaker signals for a
 143 stereo system that are given by

$$s_L(\omega, \mathbf{x}, \mathbf{p}') = s_0(\omega) g_L(\mathbf{x}, \mathbf{p}'),\tag{6}$$

144 and

$$s(\omega, \mathbf{x}, \mathbf{p}') = s_0(\omega) g_R(\mathbf{x}, \mathbf{p}'),\tag{7}$$

145 where $g_L(\mathbf{x}, \mathbf{p}')$ and $g_R(\mathbf{x}, \mathbf{p}')$ represent the normalised panning gains. These panning
 146 gains are adjusted in such a way that the listener localises the sound object always at the
 147 same absolute position, given by $\mathbf{p}'$, independently of the listening position $\mathbf{x}$. The whole
 148 process for adapting the panning gains to the listener position is explained below. The
 149 loudspeaker signals $s_L(\omega, \mathbf{x}, \mathbf{p}')$ and $s_R(\omega, \mathbf{x}, \mathbf{p}')$ can be further compensated to account

for the listener position and compensate for the propagation and attenuation delay of each loudspeaker, as performed in Equations 4 and 5, to give compensated loudspeaker signals $\tilde{s}_L(\omega, \mathbf{x}, \mathbf{p}')$ and $\tilde{s}_R(\omega, \mathbf{x}, \mathbf{p}')$.

3.1 Object panning strategies

The intended position of the different audio objects can be adjusted using panning techniques such as VBAP [7] and VBIP [11]. Although these techniques are normally applied to three-dimensional loudspeaker arrays, the same equations are used here for a two-loudspeaker system. For an intended position of a given audio object, as shown in Fig. 2, the normalised panning gains $g_L(\mathbf{x}, \mathbf{p}')$ and $g_R(\mathbf{x}, \mathbf{p}')$ are calculated to keep the perceived absolute position of the audio object independent of the listener position. Note that these gains are obtained differently for the low and high frequencies; below 700 Hz the gains are obtained using the VBAP formulation and above 700 Hz the gains are obtained employing the VBIP formulation.

It is noteworthy to define two terms which are often used in the next sections of the paper, “loudspeaker region” and “stereo region”, shown in Fig. 2. The loudspeaker region is the area given by the infinite strip defined between the two loudspeakers, and the stereo region is the infinite area enclosed by the two lines between the listener and the two loudspeakers.

At low frequencies, i.e., below 700 Hz, the resultant sound field at the listener position can be modelled by the sum of the particle velocity vectors of phase coherent plane-waves at that position [15]. The sound field is hence described by the pressure and the velocity, wherein the local sound field is given by a combination of the pressure and the velocity of the plane-waves incoming from the two loudspeakers. For a given listener position, a

Figure 2: Example of how the VBAP gains are adjusted for two listener position so that the object stays static relative to the listener at a desired 2D coordinate. The dashed loudspeaker symbolises the audio object.

unitary vector in the direction of the image, $\hat{\mathbf{p}}'_{(\text{LF})}(\mathbf{x})$, is created according to [7]

$$\hat{\mathbf{p}}'(\mathbf{x}) = g_{\text{L}(\text{LF})}(\mathbf{x}, \mathbf{p}')\hat{\mathbf{r}}'_{\text{L}}(\mathbf{x}) + g_{\text{R}(\text{LF})}(\mathbf{x}, \mathbf{p}')\hat{\mathbf{r}}'_{\text{R}}(\mathbf{x}), \quad (8)$$

where $g_{\text{L}(\text{LF})}(\mathbf{x}, \mathbf{p}')$ and $g_{\text{R}(\text{LF})}(\mathbf{x}, \mathbf{p}')$ are the unnormalised gain factors of the waves arriving from each loudspeaker at the listener position for low frequency and $\hat{\mathbf{r}}_{\text{L}}(\mathbf{x})$ and $\hat{\mathbf{r}}_{\text{R}}(\mathbf{x})$ are the unitary pointing vectors of each loudspeaker. In the original VBAP formulation it is assumed that the listener is placed at the same distance from both loudspeakers. In the formulation presented here, however, $\hat{\mathbf{r}}_{\text{L}}$ and $\hat{\mathbf{r}}_{\text{R}}$ account for the instantaneous listening position. Eq. 8 can be rewritten in matrix notation

$$\begin{bmatrix} p'_1(\mathbf{x}) \\ p'_2(\mathbf{x}) \end{bmatrix} = \begin{bmatrix} r'_{\text{L}1}(\mathbf{x}) & r'_{\text{R}1}(\mathbf{x}) \\ r'_{\text{L}2}(\mathbf{x}) & r'_{\text{R}2}(\mathbf{x}) \end{bmatrix} \begin{bmatrix} \check{g}_{\text{L}(\text{LF})}(\mathbf{x}, \mathbf{p}') \\ \check{g}_{\text{R}(\text{LF})}(\mathbf{x}, \mathbf{p}') \end{bmatrix}, \quad (9)$$

where the subscripts 1 and 2 indicate the orthogonal projections on the w_1 and w_2 axis.

The required gains are then obtained by [7]

$$\begin{bmatrix} \check{g}_{\text{L}(\text{LF})}(\mathbf{x}, \mathbf{p}') \\ \check{g}_{\text{R}(\text{LF})}(\mathbf{x}, \mathbf{p}') \end{bmatrix} = \begin{bmatrix} r'_{\text{L}1}(\mathbf{x}) & r'_{\text{R}1}(\mathbf{x}) \\ r'_{\text{L}2}(\mathbf{x}) & r'_{\text{R}2}(\mathbf{x}) \end{bmatrix}^{-1} \begin{bmatrix} p'_1(\mathbf{x}) \\ p'_2(\mathbf{x}) \end{bmatrix}. \quad (10)$$

At low frequency, the unnormalised gains $\check{g}_{L(LF)}(\mathbf{x}, \mathbf{p}')$ and $\check{g}_{R(LF)}(\mathbf{x}, \mathbf{p}')$ are then converted into normalised gains $g_{L(LF)}(\mathbf{x}, \mathbf{p}')$ and $g_{R(LF)}(\mathbf{x}, \mathbf{p}')$, so that the acoustic pressure at the listener position is equal to a certain value, here assumed to be unitary. This is obtained by

$$g_{L(LF)}(\mathbf{x}, \mathbf{p}') = \frac{\check{g}_{L(LF)}(\mathbf{x}, \mathbf{p}')}{\check{g}_{L(LF)}(\mathbf{x}, \mathbf{p}') + \check{g}_{R(LF)}(\mathbf{x}, \mathbf{p}')}, \quad (11)$$

and

$$g_{R(LF)}(\mathbf{x}, \mathbf{p}') = \frac{\check{g}_{R(LF)}(\mathbf{x}, \mathbf{p}')}{\check{g}_{L(LF)}(\mathbf{x}, \mathbf{p}') + \check{g}_{R(LF)}(\mathbf{x}, \mathbf{p}')}. \quad (12)$$

This normalisation requires a limitation of the loudspeaker gains when the audio objects or the listener are at certain positions, otherwise it can lead to large normalised gains which can make the system unstable, as it will be shown in Section 3.3.

At higher frequencies, i.e., above 700 Hz, when the wavelength is smaller than the head width, the sound field approximation introduced above does no longer hold. For this reason, Gerzon proposed that the localisation of a sound field was given by the sum of the energy [9]. A unitary vector on the direction of the image is created in this frequency range according to

$$\hat{\mathbf{p}}'(\mathbf{x}) = g_{L(HF)}^2(\mathbf{x})\hat{\mathbf{r}}_L(\mathbf{x}) + g_{R(HF)}^2(\mathbf{x})\hat{\mathbf{r}}_R(\mathbf{x}). \quad (13)$$

By comparison of Eq. 13 with Eq. 8 we can derive an equation for the high frequency gains equal to

$$\begin{bmatrix} \check{g}_{L(HF)}^2(\mathbf{x}) \\ \check{g}_{R(HF)}^2(\mathbf{x}) \end{bmatrix} = \begin{bmatrix} r'_{L1}(\mathbf{x}) & r'_{R1}(\mathbf{x}) \\ r'_{L2}(\mathbf{x}) & r'_{R2}(\mathbf{x}) \end{bmatrix}^{-1} \begin{bmatrix} p'_1(\mathbf{x}) \\ p'_2(\mathbf{x}) \end{bmatrix}. \quad (14)$$

The loudspeaker gains for the higher frequency region can then be normalised so that $\check{g}_{L(HF)}^2(\mathbf{x}) + \check{g}_{R(HF)}^2(\mathbf{x}) = 1$. Another preventive measure has to be taken for sources panned outside of the stereo region, as otherwise the method will give gains with a complex value. To avoid this, the feed for the loudspeaker closest to the panned object is set to 1, with the other loudspeaker feed set to 0. The normalised loudspeakers signals $g'_{L(HF)}(\mathbf{x})$ and

205 $g'_{\text{R(HF)}}(\mathbf{x})$, can then be defined by

$$g_{\text{L(HF)}}(\mathbf{x}, \mathbf{p}') = \begin{cases} \frac{\sqrt{\check{g}_{\text{L(HF)}}^2(\mathbf{x}, \mathbf{p}')}}{\sqrt{\check{g}_{\text{L(HF)}}^2(\mathbf{x}, \mathbf{p}') + \check{g}_{\text{R(HF)}}^2(\mathbf{x}, \mathbf{p}')}} & \text{otherwise} \\ 1 & \text{if } \check{g}_{\text{R(HF)}}^2(\mathbf{x}, \mathbf{p}') \leq 0 \\ 0 & \text{if } \check{g}_{\text{L(HF)}}^2(\mathbf{x}, \mathbf{p}') \leq 0 \end{cases} \quad (15)$$

206 and

$$g_{\text{R(HF)}}(\mathbf{x}, \mathbf{p}') = \begin{cases} \frac{\sqrt{\check{g}_{\text{R(HF)}}^2(\mathbf{x}, \mathbf{p}')}}{\sqrt{\check{g}_{\text{L(HF)}}^2(\mathbf{x}, \mathbf{p}') + \check{g}_{\text{R(HF)}}^2(\mathbf{x}, \mathbf{p}')}} & \text{otherwise} \\ 1 & \text{if } \check{g}_{\text{L(HF)}}^2(\mathbf{x}, \mathbf{p}') \leq 0 \\ 0 & \text{if } \check{g}_{\text{R(HF)}}^2(\mathbf{x}, \mathbf{p}') \leq 0 \end{cases} \quad (16)$$

207 A given audio object $s_0(\omega)$ can then be multiplied with the normalised low and high
 208 frequency panning gains and further compensated depending on the listening position
 209 according to Equations 4 and 5 to give the compensated and normalised loudspeaker sig-
 210 nals, $\tilde{s}_{\text{L(LF)}}(\omega, \mathbf{x}, \mathbf{p}')$, $\tilde{s}_{\text{R(LF)}}(\omega, \mathbf{x}, \mathbf{p}')$, $\tilde{s}_{\text{L(HF)}}(\omega, \mathbf{x}, \mathbf{p}')$ and $\tilde{s}_{\text{R(HF)}}(\omega, \mathbf{x}, \mathbf{p}')$.

211

212 3.2 Initial localisation simulations

213 Initial simulations of localisation can be performed using the velocity and energy local-
 214 isation vectors. At low frequency, the perceived localisation of a phantom source can
 215 be estimated by means of the Makita localisation vector, or velocity localisation vector,
 216 $\mathbf{r}_V(\mathbf{x}_0)$, which was introduced by Gerzon [9] and is based on the localisation studies of
 217 Makita [15]. The velocity localisation vector for a point field $\mathbf{x}$ is given by

$$\mathbf{r}_V(\mathbf{x}) = \frac{g_{\text{L(LF)}}(\mathbf{x}, \mathbf{p}')\hat{\mathbf{r}}'_L(\mathbf{x}) + g_{\text{R(LF)}}(\mathbf{x}, \mathbf{p}')\hat{\mathbf{r}}'_R(\mathbf{x})}{g_{\text{L(LF)}}(\mathbf{x}, \mathbf{p}') + g_{\text{R(LF)}}(\mathbf{x}, \mathbf{p}')} \quad (17)$$

218 The magnitude $r_V = |\mathbf{r}_V| = Z_0 v / p$ is the ratio of the resultant pressure and velocity mag-
 219 nitude multiplied by the characteristic acoustic impedance, $Z_0 = \rho_0 c_0$, where $\rho = 1.20$

kg/m³ is the density of the air at 20°C. Assuming that the sweet spot is not in the near-field of a given loudspeaker, it is possible to approximate the sound field radiated by a loudspeaker by a plane-wave arriving at the listener position from the direction of the loudspeaker as seen by the listener.

224

At higher frequencies, the perceived localisation is better estimated by the energy vector, which was proposed by Gerzon [9]. The energy vector for a point field $\mathbf{x}$ is given by

$$\mathbf{r}_E(\mathbf{x}) = -\frac{g_{\mathbf{L}(\text{HF})}^2(\mathbf{x}, \mathbf{p}')\hat{\mathbf{r}}_{\mathbf{L}}'(\mathbf{x}) + g_{\mathbf{R}(\text{HF})}^2(\mathbf{x}, \mathbf{p}')\hat{\mathbf{r}}_{\mathbf{R}}'(\mathbf{x})}{g_{\mathbf{L}(\text{HF})}^2(\mathbf{x}, \mathbf{p}') + g_{\mathbf{R}(\text{HF})}^2(\mathbf{x}, \mathbf{p}')}. \quad (18)$$

Some examples of how a listener localises audio objects at different positions were simulated using the formulation presented in this section, which are shown in Fig. 3. The low frequency object panning localisation has been predicted by means of the Makita velocity vector, $\mathbf{r}_V(\mathbf{x})$, as defined in Eq. 17 and is plotted in solid line. The high frequency object panning localisation has been predicted by means of the energy vector, $\mathbf{r}_E(\mathbf{x})$, as given by Eq. 18 and is plotted in dashed line. For visualisation purposes, the localisation vectors have been normalised to have unitary amplitude and emphasize the perceived localisation direction, with the origin of the arrows identifying the listening position. The simulations are performed for two loudspeakers 1.6 m apart from each other.

236

Fig. 3a shows the predicted velocity and energy vectors for an object placed at an infinite distance with a direction parallel to the axis of symmetry between both loudspeakers. The solid arrows show the low frequency localisation, and it can be seen that for listening positions outside of the stereo region (refer to Fig. 2 for the definition of stereo and loudspeaker regions), the active compensation is able to provide accurate localisation. The dashed arrows show the high frequency localisation, and it can be seen that outside the stereo region, the arrows point towards the closest loudspeaker.

244

(a) Object propagating from the infinity parallel to the loudspeakers

(b) Audio object inside the loudspeaker region and

(c) Audio object outside of the loudspeaker region

Figure 3: Simulations of perceived localisation for the frequency band below 700 Hz, using Eq. 17 with panning gains calculated according to Equations 11 and 12 (solid line); and for the frequency band above 700 Hz, which have been calculated according to Eq. 18, for panning gains obtained with Equations 15 and 16 (dashed line).

245 For point-source objects within the loudspeaker region, the localisation models for low
246 and high frequencies predict similar results for listener positions inside the stereo region,
247 as shown in Fig. 3b. For listener positions outside the stereo region in the half plane
248 where the object is placed, the high frequency localisation vectors point towards the near-
249 est loudspeaker. The localisation of point-source objects placed outside the loudspeaker
250 region follow a similar pattern, as shown in Fig. 3c. For listener positions inside the
251 stereo region, both the high frequency and the low frequency localisation are accurate
252 and offer very similar results. For listening positions outside of the stereo region, the low
253 frequency localisation points towards the object independently of the listener position,
254 however, the predicted high frequency localisation is largely biased when the listener is
255 outside the loudspeaker region, and collapses to the loudspeaker that is on the same half
256 plane as the object.

257

258 Although these simulations show that the low frequency localisation vectors predicts an
259 accurate localisation when the listener or the audio objects are outside the stereo region,
260 in practice, this is very sensitive and can not be always obtained, as it is shown in the
261 next sections.

262

263 3.3 System Robustness

264 In order to provide an accurate localisation, the proposed listener adaptive system must
265 be robust with respect to errors in listener position introduced by the tracking system
266 and/or errors in the positioning of the loudspeakers. Furthermore, the individual radi-
267 ation patterns of the loudspeakers will change between different transducer units. The
268 formulation here presented is to be setup-agnostic, and hence is necessary to investigate
269 how robust this is for different loudspeaker/tracking configurations.

270

271 Simulating errors in listener position estimation can be performed similarly to previous
 272 studies that have assessed the robustness of loudspeaker array beamforming [16]. In
 273 this case, the matrix of unitary pointing vectors of each loudspeaker is modified to give
 274 a matrix containing small inaccuracies, $\mathbf{R}_N$, to give a set of noisy, unnormalised gains
 275 $\check{g}_{LN}(\mathbf{x})$ and $\check{g}_{RN}(\mathbf{x})$

$$\mathbf{g}_N = \begin{bmatrix} \check{g}_{LN}(\mathbf{x}, \mathbf{p}') \\ \check{g}_{RN}(\mathbf{x}, \mathbf{p}') \end{bmatrix} = \mathbf{R}_N^{-1} \mathbf{g}, \quad (19)$$

276 where

$$\mathbf{R}_N = \begin{bmatrix} r'_{L1}(\mathbf{x})(1 + \lambda) & r'_{R1}(\mathbf{x})(1 + \lambda) \\ r'_{L2}(\mathbf{x})(1 + \lambda) & r'_{R2}(\mathbf{x})(1 + \lambda) \end{bmatrix}, \quad (20)$$

277 where λ is a Gaussian random noise which introduces variations with a standard deviation
 278 equal to 30 % of the nominal values of $\|\mathbf{r}_L(\mathbf{x}, \mathbf{p}')\|$ and $\|\mathbf{r}_R(\mathbf{x}, \mathbf{p}')\|$.

279

280 Another type of error is also introduced in the system due to the fact that the loudspeak-
 281 ers are not ideal point-sources with an omnidirectional radiation pattern. Instead, each
 282 loudspeaker has a radiation pattern that varies with frequency and position, and also
 283 between loudspeakers. To simulate these effects, errors can be introduced in the nominal
 284 value of the normalised loudspeaker feeds. A new set of noisy, normalised loudspeaker
 285 feeds can be defined as

$$g_{LN}(\mathbf{x}, \mathbf{p}') = \frac{\check{g}_{LN}(\mathbf{x}, \mathbf{p}')}{\check{g}_{LN}(\mathbf{x}, \mathbf{p}')(1 + \lambda) + \check{g}_{RN}(\mathbf{x}, \mathbf{p}')(1 + \lambda)}, \quad (21)$$

286 and

$$g_{RN}(\mathbf{x}) = \frac{\check{g}_{RN}(\mathbf{x}, \mathbf{p}')}{\check{g}_{LN}(\mathbf{x}, \mathbf{p}')(1 + \lambda) + \check{g}_{RN}(\mathbf{x}, \mathbf{p}')(1 + \lambda)}. \quad (22)$$

287 Hence, using Equations 19, 20, 22 and 22 to simulate tracking and directivity errors al-
 288 lows us to estimate how the localisation will be for different virtual source positions. A
 289 simulation of robustness was performed using a stereo set up in which the loudspeaker

separation is 1.6 m. The localisation was estimated for sources incoming between -90° and 90° with respect to the listening position. This analysis was repeated for three listening positions: $\mathbf{x} = (0.0 \text{ m}, \quad 1.4 \text{ m})$, corresponding to the sweet spot for a typical stereo listening configuration with a span of 30° between each loudspeaker, $\mathbf{x} = (-0.4 \text{ m}, \quad 2.0 \text{ m})$, an asymmetrical listening configuration, and $\mathbf{x} = (-1.0 \text{ m}, \quad 2.0 \text{ m})$, a listener position outside of the loudspeaker region. The simulations were performed 100 times, and the average of the results is shown here.

Figure 4: The upper graphs show the predicted localisation of a virtual source for three listener positions without mismatches (“Ideal”, solid line) and with mismatches (“Perceived”, dashed line). The dashed dotted line shows the absolute squared localisation error. The lower graphs show the norm of the panning gains against the angle of the virtual source. The dashed vertical lines show the position of both loudspeakers as seen from the listener.

The noisy localisation results are shown in Fig. 4, together with the norm of the loudspeaker gains against the target source panning angle, θ_T . The results for the three listening positions show that despite the errors introduced by the tracking system and by the individual radiation pattern of each loudspeaker, the predicted perceived localisation,

θ_P , is accurate when the audio objects are placed between the two loudspeakers. At the sweet spot, Fig. 4a, the predicted localisation is accurate until approximately $\pm 40^\circ$, where errors start to appear at those orientations for which the norm of the loudspeaker gains is large. Similarly the same phenomena also occurs for the listening position at $\mathbf{x} = (-0.4 \text{ m}, 2.0 \text{ m})$, shown in Fig. 4b. The predicted localisation of the incoming wave is accurate between the stereo region. Outside the stereo region the predicted perceived localisation is erratic, with the loudspeaker gains requiring a larger norm, as they are in phase opposition to generate a desired target velocity vector. For the listener position at $\mathbf{x} = (-1.0 \text{ m}, 2.0 \text{ m})$, shown in Fig. 4c, the results are similar, however, errors start to arise even between the loudspeaker region, especially for object positions close to the loudspeaker farthest (away) from the listener. It is possible to observe that the norm of the gains becomes unstable at about 65° , as for this listening position and object orientation the two loudspeaker gains have equal magnitude and opposite phase.

These simulations predict that, despite the errors, localisation is accurate for most listening positions inside the loudspeaker region, even when errors up to 30% in the listener position estimation and in the loudspeaker radiation patterns are present. The large norm of the loudspeaker gains requires that an additional upper bound is applied to the normalisation process, as otherwise the system may become unstable for some listening and object positions.

4 Real-Time Implementation

4.1 System Description

Using the described formulation a system was conceived to perform object-based audio rendering, which is shown in Fig. 5. The system was implemented using MAX MSP 6

327 running on a standard Windows PC. A pair of Mackie HR624 MK2 active studio moni-
 328 tors were used as loudspeakers, spaced at a distance of 1.6 m. The audio output from the
 329 computer was obtained through an RME Madiface UCX sound card. The video tracking
 330 was implemented using a Microsoft Kinect 1 using the MAX MSP patch *dp.kinect* [17].
 331 The filter bank used to perform the frequency dependent panning was implemented using
 332 a cross-over network with a cut-off frequency of 700 Hz.

333

Figure 5: Picture of the loudspeaker and listener tracking system used for the real-time implementation of the proposed listener adaptive stereo system.

334 4.2 Localisation measurements

335 In order to assess the performance of the real-time system and validate the results of the
 336 panning simulations performed in Sections 3.2, an 3.3 and initial set of measurements
 337 was carried out. These were measured in an anechoic chamber using a GRAS 455BB-3
 338 KEMAR head and torso simulator as binaural microphone. Impulse responses were ob-
 339 tained between the loudspeakers and the binaural microphone for a given set of listener
 340 positions and object panning combinations.

341

342 Prior to perform any object panning, the ability of the proposed system to compensate for
 343 off-axis listening positions was tested. The effect of compensation for gain and delay using
 344 the formulation of Section 2 on the impulse responses of each loudspeaker is shown in Fig.

6 for two asymmetric listener positions measured on the set up of Fig. 5. Position 1 corresponds to $\mathbf{x} = (-0.4 \text{ m}, 1.1 \text{ m})$ and position 2 corresponds to $\mathbf{x} = (-0.6 \text{ m}, 2.3 \text{ m})$. When no compensation is applied, a large difference of arrival time between the impulse responses can be observed in Fig. 6a and Fig. 6c. However, when the loudspeaker input signals are compensated according to Equations 4 and 5, the waves coming from both loudspeakers arrive at the listener position at the same time, as shown in Fig. 6b and Fig. 6d. Although the loudspeaker signals have also been compensated for the distance attenuation of the acoustic waves, there is a large difference in the amplitude of the compensated impulse responses for the two loudspeakers, which is compensated for Position 2, Fig. 6d, but it is not so accurately equalised for Position 1, Fig. 6b. This suggests that these differences arise due to the frequency and position dependent radiation pattern of each loudspeaker.

357

Figure 6: Measured impulse responses (IRs) of each of the loudspeaker, IR_L for left loudspeaker and IR_R for the right loudspeaker; without compensation, Figures (a) and (c), and with compensation, Figures (b) and (d).

The real-time capabilities of the proposed system to pan an audio object at different target angles was first measured objectively using the interaural time difference (ITD) function. The ITD, $\rho_{RL}(\tau)$, was obtained via the interaural cross-correlation function [18], which is calculated according to

$$\rho_{RL}(\tau) = \frac{\int_{t_1}^{t_2} h_L(t) h_R(t + \tau) dt}{\sqrt{\int_{t_1}^{t_2} h_L(t) dt \int_{t_1}^{t_2} h_R(t) dt}}, \quad (23)$$

where $h_L(t)$ and $h_R(t)$ are left and right ear impulse responses and $t_2 - t_1$ is the size of

the integration window. These responses were low-pass filtered below 700 Hz, since this frequency range is considered to be of importance for the ITDs [19].

Figure 7: ITD difference obtained from panning a virtual source between -70° and 70° . Results measured at the sweet spot, at an asymmetrical listening position, $\mathbf{x} = (-0.4 \text{ m}, 2.0 \text{ m})$, and at a listening position outside of the loudspeaker region, $\mathbf{x} = (-1.0 \text{ m}, 2.0 \text{ m})$. The black thick dashed line shown for comparison is the measured ITD from a single loudspeaker rotated around the listener and the single dots correspond to the measured ITD when the dummy head is facing the loudspeakers. The dashed horizontal lines represent an ITD of 0.05 ms, which corresponds to about 6° .

365

The measurements were performed at the same listening points as those used in the simulations of Section 3.3: namely a stereo listening configuration where the listener is in the sweet spot, an asymmetrical position, $\mathbf{x} = (-0.4 \text{ m}, 2.0 \text{ m})$, and a listening position where the listener is outside of the loudspeaker region, $\mathbf{x} = (-1.0 \text{ m}, 2.0 \text{ m})$. A virtual object consisting on white Gaussian noise was panned between -70° and 70° , in ten degree increments. For each presented angle, the binaural microphone was rotated towards the target source direction, θ_T , what, ideally, would result in 0 ITD for a plane wave propagating from such direction. The measured results are shown in Fig. 7, wherein an *ideal* localisation should give a 0 ITD for every source direction. When the binaural microphone is on the sweet spot, and each loudspeaker is at an angle of 30° from the listener position, the results show that the system is able to pan a virtual source with less than 0.1 ms of ITD difference between -40 and 60° . The measured ITD for a single loudspeaker

378 against azimuth angle is also shown for comparison, and it can be seen that an ITD of
379 0.05 ms corresponds to about 6° . The ITD difference is lowest between the loudspeaker
380 region 0 at -30° and 30° , where the binaural microphone is orientated directly towards
381 the loudspeakers. Between -60° and -40° the ITD difference grows, as in this range of
382 source directions the system is less robust to errors in the loudspeaker radiation pattern
383 and positioning mismatches, as shown in Section 3.3.

384

385 As the binaural microphone was moved towards the left of the loudspeaker region, $\mathbf{x} =$
386 $(-0.4 \text{ m}, 2.0 \text{ m})$, the ITD error becomes larger for sources panned towards negative
387 angles (i.e., towards the side where the dummy head was moved to). For this case the
388 ITD remains very close to 0 between -20° , where the closest loudspeaker is placed, and 60° .

389

390 If the binaural microphone is moved outside of the loudspeaker region, $\mathbf{x} = (-1.0 \text{ m}, 2.0 \text{ m})$,
391 the localisation is still good, i.e., the ITD is close to 0, for virtual sources incoming be-
392 tween 0° and 40° . For virtual sources panned between 40° and 60° the ITD difference
393 grows to approximately 0.1 ms at 60° . This indicates that the perceived localisation of
394 virtual sources coming from these directions may be affected by small inaccuracies in the
395 tracking system and in the loudspeaker radiation pattern, as also observed in the simula-
396 tions of Section 3.3.

397

398 The measurements show a similar behaviour to that predicted in the simulations of Sec-
399 tions 3.2 and 3.3. The listener-adaptive system is able to provide an accurate localisation
400 for objects placed inside the stereo region, however, the localisation of such objects is lost
401 if these are panned outside of the stereo region due to small inaccuracies on the system.

402

Figure 8: Schematic for the two listening tests performed; experiment 1, (a), where the localisation of panned objects was assessed for static listeners, and experiment 2, (b), where the listener is allowed to move freely in front of the loudspeaker set up. In (c) there is shown one of the listeners performing the tests.

4.3 Subjective Performance

Subjective localisation tests were carried out in an anechoic chamber to assess the perceived location of virtual audio objects at different listening positions. Two experiments were carried out: a first experiment which assessed the perceived location of objects at various static listener positions, and a second experiment to assess the 2D position of audio objects, in which the listeners were allowed to move freely to assess the ability of

the system to create motion parallax cues. The test signal used for the experiment consisted of an ensemble of 10 male and female speech excerpts in different languages. Eleven adults with normal hearing participated in the experiments. Both loudspeakers were covered with an acoustically transparent black cloth so that the perceived localisation was not biased by the loudspeaker position visual cues, as shown in Fig. 8c. The listeners were encouraged to look to the centre of the set up, given by the Microsoft Kinect, and not to rotate their head.

A diagram of the first experiment can be seen in Fig. 8a. For the first experiment three audio objects were considered: object 1, $\mathbf{x} = (0.0 \text{ m}, -0.1 \text{ m})$; object 2, $\mathbf{x} = (0.4 \text{ m}, -0.1 \text{ m})$; and object 3 $\mathbf{x} = (-1.2 \text{ m}, -0.1 \text{ m})$. Objects 1 and 2 were inside of the stereo region whilst that Object 3 was outside of the stereo region. As the reproduction set up was symmetrical, the localisation of these objects was assessed at the same three listener position as those used in the simulations in Section 3.3 and the measurements in Section 4.2: position 1, the sweet spot; position 2, the asymmetrical position, $\mathbf{x} = (-0.4 \text{ m}, 2.0 \text{ m})$; and position 3, the listening position out of the loudspeaker region $\mathbf{x} = (-1.0 \text{ m}, 2.0 \text{ m})$. The different objects were presented in a random order with the listener tracking system “ON” and “OFF”. Each object-system operation mode was rated 5 times by each listener, in a random different order every time. The listeners were asked to indicate the location of the different objects by referring to labels attached to the black cloth covering the loudspeakers, spaced by 8 cm, as shown in Fig. 8c.

It can be seen in Fig. 9a that the perceived location results for the sweet spot listening position show very similar localisation patterns for the three audio objects when the system was tracking the listener(system “ON”), and when the system was not tracking the listener (system “OFF”), i.e., a normal stereo configuration. When the listener is moved to asymmetrical listening positions, $\mathbf{x} = (-0.4 \text{ m}, 2.0 \text{ m})$, Fig. 9b; or

(a) Sweet-Spot

(b) Asymmetrical, $\mathbf{x} = (-0.4 \text{ m}, 2.0 \text{ m})$

(c) Outside of loudspeaker region, $\mathbf{x} = (-1.0 \text{ m}, 2.0 \text{ m})$

Figure 9: Box plots of the subjective tests to assess the position of audio objects. The dashed lines at $x=0$ m represent the centre of the loudspeaker set up, and this at $x=0.8$ m the position of the left loudspeaker. The intended position of each object is shown by the red circles. O1, O2 and O3 refer to the different audio objects, as shown in Fig. 8 and “ON” and “OFF” refer to the operation move of the listener adaptive system.

436 $\mathbf{x} = (-1.0 \text{ m}, 2.0 \text{ m})$, Fig. 9c, the listener tracking compensation is able to reproduce
 437 the intended location more accurately than when the position compensation is not active.
 438 In the latter operation mode, the perceived location of the objects is more widely spread,
 439 and collapses close to the loudspeaker for objects 2 and 3. The effect of the system was

Figure 10: Sum of responses for 2D objects localisation based on parallax cues. The red dots indicate the intended position of the object.

also evaluated using statistical analysis. It was found that the data were not normally distributed (mainly due to high kurtosis for some stimuli), and hence a non-parametric method was used. For each condition, the effect of the system operation was tested using a Mann-Whitney U test [20]. It was found that for the central listening position, switching the system between “ON” and “OFF” had no statistically significant effect for objects 1 & 2; for object 3 there was a relatively small statistically significant effect ($p < .01$). For the other two listening positions, switching the system between “ON” and “OFF” had a large statistically significant effect for all object positions ($p < .001$). These results are comparative with those of the simulations in Section 3.2, wherein it was predicted that audio objects can be placed outside of the stereo region at low frequency. Although panning outside the stereo region is only effective at low frequency, it worked with the speech excerpts used in the experiment.

452

The second experiment analysed how effectively the proposed system can create parallax cues to keep the position of audio objects static. This is obtained by tracking the listener

position and instantaneously adapt the panning gains. In this case, 3 objects were placed at different positions in the horizontal plane, as shown in Fig. 8b. The same listeners as in the previous experiment were allowed to move freely in front of the loudspeakers and were asked to localise the position of the audio objects, with the listener tracking system “ON” and “OFF”. The listeners were instructed to mark the position of each object in an imaginary grid spaced 0.5 m and representing the space behind the black cloth covering the loudspeakers, as shown in Fig. 8b. Each object-system operation combination was repeated 4 times in a random order. A histogram of the results can be seen in Fig. 10, depicting the sum of responses for each position on the response grid. These data were analysed using Fisher’s exact test [20] (as opposed to a chi-square test due to the low number of responses in some categories). It was found that for all three of the object positions, there was a statistically significant difference (for objects 1 & 2: $p < .001$; for object 3: $p < .05$).

From the histogram, it can be seen that the majority of the responses for each object were located close to their intended 2D position when the listener tracking system was “ON”. When the tracking system was “OFF”, however, the localisation was biased towards positions close to the loudspeakers. For the object panned in the centre of the loudspeaker region, the plots show that accurate localisation was not possible when the tracking system was not operating; in this case the responses were relatively widely spread around the intended position. This suggests that listener tracking is required to allow the perceived position to be stable when the listener moves.

5 Conclusion

This paper has presented an approach for listener position adaptive reproduction of object-based audio. The presented formulation is for a stereo system using 2 loudspeakers, but

481 the same principle can be applied in systems with a larger number of loudspeakers. The
482 compensation comprises of two parts: a first part that compensates for the listening
483 position and a second part that pans the different audio objects so that the localisation
484 is stable, using VBAP at frequencies below 700 Hz and VBIP at frequencies above 700 Hz.

485

486 Initial localisation simulations using the velocity and energy localisation vectors have pre-
487 dicted that the frequency dependent panning is able to pan virtual audio sources outside
488 of the loudspeaker region at low frequency. A robustness analysis has shown that, in
489 practice, localisation of audio objects is robust when these are panned between the stereo
490 region, and that, however, the localisation of objects outside of the stereo region is sen-
491 sitive to errors and it is affected by the accuracy of the video tracking system and the
492 homogeneity of the radiation pattern of the different loudspeakers of the system.

493

494 The proposed formulation has been implemented in a real time system. Localisation
495 predictions based on the ITD have shown that, as predicted in the simulations, the lo-
496 calisation is only robust when the virtual audio sources are panned inside of the stereo
497 region, and that the ability to pan audio objects outside of the loudspeaker region requires
498 a very accurate and calibrated system. However, a set of subjective test have shown that
499 localisation of sources outside the loudspeaker region is possible when these are close to
500 the loudspeakers. The subjective experiments have also shown that with the proposed
501 adaptive system it is possible to create parallax cues that allow for the rendering of static
502 point-source near-field objects.

503

6 Acknowledgements

The authors of the paper would like to acknowledge the support of the EPSRC Programme Grant S3A: Future Spatial Audio for an Immersive Listener Experience at Home (EP/L000539/1) and the BBC as part of the BBC Audio Research Partnership. The authors also want to thank Sandra Abu Hamdan and Andrew E. Causon for their help preparing the listening tests and to all the individuals who participated on the experiment.

References

- [1] B. B. Bauer, “Broadening the area of stereophonic perception,” *J. Audio Eng. Soc.*, vol. 8, no. 2, pp. 91–94, 1960.
- [2] Lowe et al., “Apparatus for enhancing stereo effect with central sound image maintenance circuit,” U.S. Patent 5 858 800, 1995.
- [3] S. Okuro and Y. Kajikawa, “Reproduction of varied sound image localization for real source in stereo audio system,” in *Signal Information Processing Association Annual Summit and Conference (APSIPA ASC), 2012 Asia-Pacific*, Dec 2012, pp. 1–4.
- [4] S. Merchel and S. Groth, “Adaptively adjusting the stereophonic sweet spot to the listeners position,” *J. Audio Eng. Soc.*, vol. 58, no. 10, pp. 809–817, 2010.
- [5] T. Momose, M. Otani, M. Hashimoto, M. Kayama, and K. Itoh, “Adaptive amplitude and delay control for stereophonic reproduction that is robust against listener position variations,” *J. Audio Eng. Soc.*, vol. 63, no. 1/2, pp. 90–98, 2015.
- [6] H. F. Robert Bleidt, Arne Borsum and S. M. Weiss, “Object-Based Audio: Opportunities for Improved Listening Experience and Increased Listener Involvement,” in *Annual Technical Conference Exhibition, SMPTE 2014*, pp.1-20, 2014.

- [7] V. Pulkki, “Virtual sound source positioning using vector base amplitude panning,”
J. Audio Eng. Soc, vol. 45, no. 6, pp. 456–466, 1997.
- [8] Pierre-Anthony S. Lemieux, Roger W. Dressler and Jean-Marc Jot, “Object-based audio system using vector base amplitude panning,” US Patent PCT/US2013/043.150, Nov. 24, 2015, uS Patent 9,197,979.
- [9] M. A. Gerzon, “General metatheory of auditory localisation,” in *Audio Engineering Society Convention 92*, Mar 1992.
- [10] M.-V. Laitinen, J. Vilkamo, K. Jussila, A. Politis, and V. Pulkki, “Gain normalization in amplitude panning as a function of frequency and room reverberance,” in *Audio Engineering Society Conference: 55th International Conference: Spatial Audio*, Helsinki, Finland, Aug. 2014.
- [11] Jean-marie Pernaux, Patrick Boussard and Jean-Marc Jot, “Virtual sound source positioning and mixing in 5.1 implementation on the real-time system genesis,” in *In Proc. Conf. Digital Audio Effects (DAFx-98)*, 1998, pp. 76–80.
- [12] Marcos F. Simón Gálvez, D. Menzies, F. M. Fazi, T. de Campos, and A. Hilton, “A listener position adaptive stereo system for object-based reproduction,” May 2015.
- [13] Marcos F. Simón Gálvez, D. Menzies, F. M. Fazi, T. de Campos, and A. Hilton, “Listener tracking stereo for object based audio reproduction,” in *Tecniacustica 2016 (Valencia) - European Symposium in Virtual Acoustics and Ambisonics*,, 2016.
- [14] T. Laakso, V. Valimaki, M. Karjalainen, and U. Laine, “Splitting the unit delay [FIR/all pass filters design],” *Signal Processing Magazine, IEEE*, vol. 13, no. 1, pp. 30–60, Jan 1996.
- [15] Y. Makita, “On the directional localization of sound in the stereophonic sound field,” *EBU review*, no. 73, pp. 102–108, 1962.

- 550 [16] M. F. Simón Gálvez, S. J. Elliott, and J. Cheer, “The effect of reverberation on
551 personal audio devices,” *The Journal of the Acoustical Society of America*, vol. 135,
552 no. 5, pp. 2654–2663, 2014.
- 553 [17] D. Phurrough, “dp.kinect, MAX MSP patch for Microsoft Kinect .”
- 554 [18] *BS EN ISO 3382:1-2009, Acoustics-Measurement of room acoustic parameters*, ISO
555 Std., 2009.
- 556 [19] D. M. Howard and J. A. Angus, Eds., *Acoustics and Psychoacoustics (Fourth edition)*,
557 fourth edition ed. Boston: Focal Press, 2010, pp. 488.
- 558 [20] A. Field, *Discovering Statistics Using SPSS*. SAGE Publications, 2005, pp. 821.