On the use of a Haptic Feedback Device for Sound Source Control in Spatial Audio Systems

F. Melchoir, C. Pike, M. Brooks and S. Grace
On the use of a haptic feedback device for sound source control in spatial audio systems

Frank Melchior, Chris Pike, Matthew Brooks, Stuart Grace

Abstract

Next generation spatial audio systems are likely to be capable of 3D sound reproduction. Systems currently under discussion require the sound designer to position and manipulate sound sources in three dimensions. New intuitive tools, designed to meet the requirements of audio production environments, are needed to make efficient use of this new technology. This work investigates a haptic feedback controller as a user interface for spatial audio systems. The paper will give an overview of conventional tools and controllers. A prototype has been developed, based on the requirements of different tasks and reproduction methods. The implementation will be described in detail and the results of a user evaluation will be given.

This work was presented at the 134th Audio Engineering Society Convention in Rome on May 5th 2013.

Additional key words: audio, immersive, production, panning, control, HCI
White Papers are distributed freely on request.

Authorisation of the Chief Scientist or General Manager is required for publication.

©BBC 2013. Except as provided below, no part of this document may be reproduced in any material form (including photocopying or storing it in any medium by electronic means) without the prior written permission of BBC Research & Development except in accordance with the provisions of the (UK) Copyright, Designs and Patents Act 1988.

The BBC grants permission to individuals and organisations to make copies of the entire document (including this copyright notice) for their own internal use. No copies of this document may be published, distributed or made available to third parties whether by paper, electronic or other means without the BBC’s prior written permission. Where necessary, third parties should be directed to the relevant page on BBC’s website at http://www.bbc.co.uk/rd/pubs/whp for a copy of this document.
1 Introduction

The next generation of audio systems beyond 5.1 surround will be spatial audio systems capable of reproducing sound in three dimensions around the listener. Several systems have already been proposed and are in professional use [1] [2] [3]. This paper addresses the control of sound sources in spatial audio systems, focusing on the tasks of source positioning and automation including the modification of source groups and directive sources. For the production of audio in professional environments it is important to design tools to be efficient and accurate, and consider time and space constraints. Furthermore it is required that the attention of the sound designer is not averted from the sound by extensive visualisation or complicated interaction using mouse and keyboard.

The editing of 3-D position with 1-D and/or 2-D interaction devices is challenging but often applied. Conventional interaction devices found in mixing consoles, like joysticks, rotary controllers, and faders, are established and reliable. However such interfaces do not necessarily give the optimal user experience in this application.

In this work a 3-D controller with haptic feedback has been used for sound source manipulation. A prototype system has been developed based on this device, scene visualisation and open sound control of a 3-D audio rendering system. Several different interaction modes for modifying sound source position have been implemented. The device used is capable of providing haptic feedback to the user by applying a force vector with a given direction. This capability is explored in the prototype to optimise editing and manipulation tasks as well as modifying the source orientation. Furthermore the device is used to make an automation path tangible giving separate control of the trajectory and the timing of the movement. The prototype system has been evaluated by an expert panel.

2 History and state-of-the-art

The beginning of virtual sound source manipulation is very much linked to the physical constraints of electrical circuits built into mixing consoles. When these became capable of stereo mixing three way switches were first used before the pan pot was introduced. From this time a single knob was normally used to control the mix of a signal to the left and right channel in stereophonic systems. Although the design of a pan pot reflects its technical needs as a potentiometer to control a voltage, it is a very intuitive hardware controller. It has a direct mapping from its orientation to the position of the source between two speakers.

When quadraphonic sound was introduced an extension was required since now two dimensions needed to be controlled. One of the first devices [4] was the azimuth co-ordinator developed by Bernard Speaight and used for early Pink Floyd recordings and concerts. This joystick based system also has an advantage of a direct mapping of the sound position to the physical position of the device. Starting from motion picture mixing consoles the joystick has become the standard user interface device for panning sound sources in surround sound systems today. To enable more intuitive control and editing of movements, motorised joysticks were established in 2001 [5]. Recorded sound source movements can now be played back while the joystick position is modified with two motors. By adding touch sensitivity the moving joystick can be touched and the movement can be modified and controlled by the user again. Even in surround systems rotary controls are
Figure 1: The haptic feedback device used.

still in use, in addition to joysticks, to control certain dimensions in a surround mix e.g. front-back, left-right.

The concept of pan pots has been translated to the digital world using mouse control of virtual pan pots. More appropriate to the drag and drop interaction paradigm of a computer mouse, x-y plane graphical user interface elements are also used. Here the two-dimensional source position is controlled by the mouse pointer. Such methods are very intuitive to control the source position in 2D surround sound systems. Furthermore more recent hardware devices like touch pads and stylus tablets can be used to control source positioning on a x-y plane more directly [6]. Early 3D sound systems like the 50-channel sphere at the World Expo in Osaka used a spherical controller to move sound over the 50 loudspeakers in a dome. Sensors and switches arranged on a sphere were used to control the sound in a relatively direct manner [7]. Nowadays the dimensions are often decoupled. Hence the x-y plane is combined with a separate height control or azimuth and elevation is combined with a second interface element for the distance. An example of such an implementation can be found in a system for 22.2 multi-channel production with panning using a potentiometer or joystick and separate distance control [8] or in a graphical user interface for WFS based systems [9].

Alternatively a very early approach to 3D sound control was proposed and implemented by Schaeffer and Poullin 1951. They used a gestural control to position sound sources in an experimental 3D sound system. More recently Pike et al. presented a similar approach in [10] by using natural gestures to modify the source location. Such systems can be very intuitive but have drawbacks in a professional audio environment in terms of accuracy, space requirements and workflow integration.

3 Hardware

The device used in this work is the Novint Falcon [11] (see Figure 1). When set up and ready to use, the device occupies about 25 x 25 cm of space, though it requires a clear area in front so the grip can be moved in three dimensions. It has a weight of nearly 3 kg, mostly due to a weighted base that prevents it toppling over or sliding around. The device has an external mains power supply and connects to a USB port. The control knob held in the user’s hand is interchangeable. The
standard grip is a plastic sphere (about the size of a golf ball) with four small buttons on the top. The grip is mounted on three curved arms spaced at 120° angles around the circular body of the device. The driver reports its current location as a vector $p_c = (x_c, y_c, z_c)^T$ in three-dimensional Cartesian coordinates, where $x_c$ is left/right linear movement, $y_c$ is up/down linear movement, and $z_c$ is forward/backward linear movement. A transformation matrix can be applied to these coordinates, so they can be re-oriented or re-scaled into a form appropriate to the application. A similar transformation matrix can be set up for the graphics library, so that identical coordinate systems are used by the device, the software model of the environment and the screen display. The feedback forces that should be applied are specified as a vector $F_c = (x_F, y_F, z_F)^T$ in the same coordinate system used to report its position. The driver translates the requested force into an appropriate combination of forces applied via the curved arms.

### 3.1 Implementation of force feedback

The haptic feedback from the device is controlled in custom software which also provides the scene visualisation. A force is calculated based on the current position of the device and updated in a haptic rendering loop. To provide a good user experience these updates have to be calculated and applied using a rate of 1000Hz or higher. In order to illustrate the principles of force feedback rendering a simple example using only the x-component $x_F$ of the force vector $F_c$ is given here. Figure 2 illustrates it in cross section. Consider a sphere of radius $r$ around the position $p_c$ of the device controller knob. Now the movement should be constrained by a flat wall. Assume the sphere is centred at $x$-position $x_c$ along the x-axis, and the wall starts at the x-coordinate $x_w$. If all parts of the sphere are clear of the wall, no force is applied so the user can move the grip (and hence the sphere) freely. When the sphere starts to overlap with the wall, a force $F_c = (x_F, 0, 0)^T$ is applied along the x-axis, in opposite direction and proportional to the degree of overlap between the sphere and the wall hence,

$$x_F(x_c) = \begin{cases} 0 & \forall (x_c + r) < x_w \\ -k_w[(x_c + r) - x_w] & \forall (x_c + r) \geq x_w \end{cases}$$

where $k_w$ is a constant which represents the rigidity of the wall. Low values of $k_w$ produce an elastic effect where the wall feels like rubber. High values of $k_w$ give the feeling of a hard, rigid surface. The minus sign produces a force pushing away from the wall towards the origin. In a real application the distances and forces must be calculated in three dimensions. As the environment being modelled becomes more complex, care must be taken to avoid combinations of forces which lead to oscillation. Suitable terms for friction or damping may be needed in the force models.
3.2 Haptic continuity

The issue of haptic continuity between different interface modes is one that deserves consideration when designing haptic interfaces. A sudden unexpected force being applied to the controller could result in disturbing the user. Haptic continuity informed the design decisions when developing the controller modes described in Section 6. All modes, where possible, start relative to the current controller position when a momentary button activates control, rather than moving the controller using force feedback to a specific location before interaction can commence.

One problem with this approach is that a user could start a movement with the controller positioned close to a boundary of its physical motion range, for example far to the left. When the momentary button is pressed, the user would then be limited in moving the source to the left. We therefore render a representation of the controller movement envelope in the graphical user interface (GUI) when a boundary is approached. This allows a user to see that a source is already at the edge of the movement space, they can then release the momentary button and adjust the device position before re-engaging the button and moving the virtual sound source further left.

4 Prototyping environment

The prototyping environment consists of one PC with the device connected via USB and a second PC for spatial audio rendering, see Figure 3 for a block diagram. A 3D view of the scene is generated using OpenGL rendering. It is controlled either by a plugin for the digital audio workstation (DAW) or by the interaction devices. While the audio signals are fed to the spatial audio rendering by a multichannel digital audio interface the scene data (position of the sources) are transmitted between the DAW and the prototype user interface using open sound control (OSC) [12]. Based on the interaction of the user a modified dataset is sent to the rendering PC. A commercially available 3D audio processor [13] was used to render the virtual sound sources on a two layer 26 loudspeaker setup.

In order to assist the user in manipulation virtual sound sources a 3D view has been implemented. Within this view, shown in Figure 4, the virtual sound sources are represented by coloured objects. To enable orientation within the real reproduction setup the loudspeaker positions are ren-
Figure 4: Screenshot of the interaction prototype.

dered as gray boxes. The user can select a source for interaction using a button on the grip of the controller, which is then highlighted in the GUI. The user can also zoom using the arrow keys of the keyboard.

5 Requirements of different rendering algorithms

Depending on the rendering algorithm implemented in the reproduction system, different volumes can be used to position virtual sound sources. In case of a panning based system [14] virtual sound sources can only be rendered on the surface spanned by the speaker system. Source distance must be simulated by adding reverberation and/or distance-dependent damping and filtering. Sources within the volume enclosed by the speaker array cannot be rendered. Nevertheless, sources positioned within the speaker array are often visualised in a graphical user interface and used to control an additional parameter such as spreading. Spreading distributes the source signal to more than two speakers and enable smoother transition in case of diagonal movement.

For sound field synthesis systems like wave field synthesis [15], a complete plane defined by the speaker layout can be used to render the sources. Virtual sources placed at a certain distance are modelled in terms of their wave front curvature. Sources placed in front of the loudspeaker array are rendered as focused source inside the listening area. The system used in this work is based on wave field synthesis rendering but controls two layers of speakers. Therefore virtual sound sources can be rendered in the planes of the two speaker layers and faded between the two layers to adjust source height when appropriate source coordinates are received.

6 Virtual sound source control

Once a source is selected, one of the different modification modes can be used. Further settings and constraints can be applied using keyboard shortcuts. In general the main operations/source
properties investigated in this work are virtual source position and orientation. Since both required
three degrees of freedom they only can be manipulated one at a time.

6.1 Position

One of the most important tasks is the 3D positioning and automation of virtual sound sources. In this work positioning of sound sources is considered as the positioning of direct (dry) signals only. The extension to distant dependent damping or direct-to-reverberant ratio control can easily achieved by applying the 3D positioning to panning based system and add a distant model using the virtual source distance. Three different modes for position modifications have been implemented.

6.1.1 Translation mode

In this mode a change in virtual source position is made relative to a change in position of the controller. Using a scaling constant $k_s$ which can be adjusted using the keyboard, the position of the controller $p_c$ is mapped into a volume of the virtual sound source scene and controls the virtual source position $p_s$ transmitted to the rendering system,

$$p_s = p_{s0} + k_s(p_c - p_{c0}),$$

where $p_{c0}$ indicates the position of the controller at the start of the movement and $p_{s0}$ is the position of the source at the start of the movement. The interaction starts when the user grabs the controller and presses a momentary button. This allows more effective use of the interaction volume with repeated click and drag movements.

In this mode the haptic feedback capabilities are utilised to simulate a rigid boundary to constrain the interaction volume to a sphere, near the physical range limits of the device. The force feedback capabilities can also be used to constrain the movement within this volume, reducing degrees of freedom by applying a force which keeps the device on a certain $x$-, $y$- or $z$-plane. An example to constrain the movement to a user-specified $x$-coordinate $x_{fix}$ can be calculated as follows:

$$F(x_c) = \begin{pmatrix} -k_f[x_c - x_{fix}] \\ 0 \\ 0 \end{pmatrix},$$

where $k_f$ defines the strength of the force pulling the controller back to the fixed $x$-coordinate.

6.1.2 Independent control of path geometry and position on path

An alternative interaction mode is to decouple the trajectory and the timing of the virtual source. In this mode the virtual source movement is designed in two steps:

1. A path is drawn using the controller in translation mode whilst holding a momentary button on the controller knob.

2. Once the button is released the movement of the controller is constrained to positions along the drawn trajectory. The timing of the movement is then controlled by moving the controller along this path.
This approach enables detailed design of a path and then concentrate on the source timing afterwards, perhaps synchronising with activity in the audio signals or accompanying video. Compared to a solution in which a fader or mouse is used to control the position on a given trajectory, the user gets direct haptic feedback of the 3D position which enables a more precise control.

Since the device does not allow the absolute position of the control knob to be set, it must be controlled with the force update loop. The constraining feedback forces are determined by first finding the nearest point along the path from the current controller position, this is the desired position of the device. A force steering the controller towards this position is applied with strength proportional to the distance from the path. This process could potentially lead to oscillation, especially if the update rate is too low. An upper limit was placed on the applied force strength to mitigate the risk of oscillation. Another issue is discontinuities in path traversal when tight loops are present. The distance calculation was modified to include proximity along the path from the previous desired position as well as the Euclidean distance from points. Please refer to [16] for a detailed description of the implemented path following algorithm.

### 6.1.3 Velocity mode

This mode maps the movement of the controller to the velocity of the virtual source. A three degree of freedom joystick is simulated, if the user moves the controller away from the initial position, a velocity vector is derived to control the sound source. The force feedback is used to pull the controller towards the centre of the interaction space like the springs in a mechanical 2D joystick. The magnitude of the displacement vector is mapped to the speed of the source movement in the direction of that vector, using a non-linear scaling. The interaction is centred around the initial controller position when a momentary button is pressed, so the user is not required to centre the device before moving a source.

To achieve appropriate behaviour three zones are defined around the virtual centre of the joystick interaction. An inner dead-zone is used with radius \( r_i \) where no movement is applied to the source. Beyond an outer limit at radius \( r_o \) the velocity is limited to a maximum value. Source velocity \( \mathbf{v}_s \) is calculated according to the following formulae, with \( \mathbf{p}_{c0} \) the start point of interaction and \( \mathbf{p}_c \) the current position of the controller:

\[
\mathbf{v}_s = s_v k_v (\mathbf{p}_c - \mathbf{p}_{c0}),
\]

where

\[
s_v = \begin{cases} 
0 & |\mathbf{p}_c - \mathbf{p}_{c0}| < r_i \\
\frac{|\mathbf{p}_c - \mathbf{p}_{c0}| - r_i}{r_o - r_i} & r_i \leq |\mathbf{p}_c - \mathbf{p}_{c0}| \leq r_o \\
1 & |\mathbf{p}_c - \mathbf{p}_{c0}| > r_o 
\end{cases}
\]

is the non-linear scaling factor defined by the control zones, and \( k_v \) is a constant scaling factor. A force with opposite direction to the displacement vector and proportional magnitude is applied to pull the controller back to the initial position.
6.2 Orientation and source groups

Since the device is only capable of controlling three degrees of freedom, a separate mode is required to modify the orientation of a virtual sound source or the rotation of a group of virtual sound sources. To intuitively control the orientation a similar approach as presented in [17] is used. The force feedback is used to constrain the movement of the controller to the surface of a sphere of radius \( r \). Based on the position on the sphere an orientation vector \( \mathbf{o}(p_c) \) can be derived:

\[
\mathbf{o}(p_c) = p_c - p_0
\]  

(6)

The force \( F_c(p_c) \) to constrain the grip on a sphere is given by

\[
F_c(p_c) = -k (r - |\mathbf{o}(p_c)|) \mathbf{o}(p_c)
\]  

(7)

The result can be used to specify the direction of a source or the orientation of which represents the direction of the primary axis of an arbitrary directivity function. In the prototype system the orientation vector was used to modify the orientation of a group of sources.

7 Evaluation

The developed interaction methods have been evaluated in a experiment consisting of two parts. In the first part the user had to perform different predefined source movements utilising the different positioning modes as described in Section 6.1. In the second part participants were asked to judge the usability and attractiveness of the proposed interaction methods using a set of semantic differentials.

A focus group of four sound production professionals was used to evaluate the system. Two sessions were conducted per subject. First a mouse based interface including a bird’s eye view of the scene and an additional software slider for the height of the virtual sound source was used. This was to enable the participant to familiarise themselves with the setup and the spatial audio rendering. It also acted as an anchor for the judgments of the subjects. In a second session the new haptic feedback controller was used. The whole experiment took 40 minutes per subject. After a general introduction each session consisted of a practicability evaluation and a general usability survey.

7.1 Practicability

For the evaluation of practicability the participants were ask to perform a list of virtual sound source movements in a given order. The following source movements were defined:

1. Fly-by: Move the sound source on a trajectory of your choice from the front to the back of the room
2. Circle: Move the sound source in a circle around the listener at a height of your choice
3. Up-down: Move the sound source on a straight line between two speaker layers
Figure 5: Evaluation of the system in a listening room equipped with a two layer wave field synthesis based spatial audio system consisting of 26 loudspeakers.

4. Diagonal: Move the sound source on a diagonal line crossing the centre of the listening area and going from lower layer to the upper layer

5. Arbitrary: Move the sound source on a trajectory of your choice utilising the whole reproduction volume available.

These movements were chosen to use different combinations of dimensions and include typical trajectories. The sound was rendered using a 2-layer 26 loudspeaker system as depicted in Figure 5.

Subjects were asked to grade the practicability of each mode and device for each virtual source movement using a seven point scale. The results are shown in Figure 6.

7.1.1 Results

Depending of the number of dimensions required for the trajectories a clear separation between mouse interaction and the new modes using the haptic feedback controller was found. Where the translation interaction mode of the controller was graded as practicable for all movements, the mouse was only judged as practical as long as only one or two dimensions are required (fly-by, circle, up-down). When it comes to three dimensional movements the practicability was graded as poor. The path mode and the translation mode achieved an good average rating over all modes while the velocity interaction mode was graded as being not practical for circular movements. In general it can be concluded that precise movements can be achieved more easily with a mouse based interface whereas real three dimensional movement graded as being not practical using a classical mouse based interface.

7.2 Attractiveness

After source movement tasks had been performed participants were asked to position four sources consisting of dry spot microphone recordings of a string quartet using the mouse and then the haptic feedback controller. The subjects were encouraged to position the sources according to their preference as they would have done in a real mixing situation, and could use any of the available modes on the controller. After completing the task using each control method, participants were asked to judge the usability and attractiveness of the controller using the AttrakDiff semantic
Figure 6: Mean results for the practicability evaluation with a focus group of four subjects. One indicates impracticable and seven indicates practicable.

differentials [18]. This survey uses pairs of bipolar adjectives to evaluate the hedonic and pragmatic quality of interactive products.

7.2.1 Results

The mean values for each pair are shown in Figure 7. From the results it can be concluded that the proposed method is attractive and the aspect of stimulation and identity is classified as optimal. The difference to the mouse approach for interaction can clearly be recognised from the results. It can be concluded that the proposed interaction method motivates the user and stimulates him. There is still room for improvement in terms of usability.

8 Conclusions

This paper has presented the application of a three degree of freedom interaction device with haptic feedback for sound source positioning and manipulation in a 3D audio system. Four different interaction modes have been implemented and the device has been evaluated by a focus group. In comparison to traditional mouse based interaction it has been shown that the controller has advantages for virtual sound source positioning in 3D reproduction system. For the control of sources in two dimensions the mouse based interaction is considered to be more accurate. The haptic feedback capabilities of the device enable completely new interaction methods for sound source positioning like separating the timing and the trajectory of a movement in three-dimensions and the emulation of a 3D joystick. The proposed interaction design has been rated to be an attractive and motivating tool for source positioning tasks in the evaluation. Based on these promising results further investigation will study the modification of complex scenes using the device. A modification of the grip in order to make it touch sensitive would be beneficial. For the given application a modification of the device hardware to add automation playback capability for virtual source movements would be desirable.
Figure 7: Mean values of the semantic differentials within the four quality feature groups: attractiveness, hedonic quality: stimulation, hedonic quality: identity, and pragmatic quality.

9 Acknowledgement

The authors would like to thank the participants of the evaluation for their time and the fruitful input to further improve the prototype. The authors also like to thank Michael Evans for his advice on the evaluation and comments on the draft of this paper.

References


