AUDITORY CUES FOR GESTURAL CONTROL OF MULTI-TRACK AUDIO

Martin J. Morrell

Queen Mary University of London
Centre for Digital Music
Mile End Road, London, E1 4NS, UK
martin.morrell@eecs.qmul.ac.uk

Joshua D. Reiss

Queen Mary University of London
Centre for Digital Music
Mile End Road, London, E1 4NS, UK
josh.reiss@eecs.qmul.ac.uk

Tony Stockman

Queen Mary University of London
Interaction, Media and Communication
Bancroft Road, London, E1 4NS, UK
josh.reiss@eecs.qmul.ac.uk

ABSTRACT

This paper presents a study undertaken to evaluate user ratings on auditory feedback of sound source selection within a multi-track auditory environment where sound placement is controlled by a gesture control system. Selection confirmation is presented to the participants via changes to the audio mixture over the stereo loudspeakers or feedback over a single ear bluetooth headset. Overall five different methods are compared and results of our study are presented. A second task in the study was given to evaluate a preselection method to help find sound sources before selection, the participant altered a width control of the pre-selection that was heard in the bluetooth headset. Results indicate a specific value irrespective of genre that the pre-selection should be set to whilst the selection confirmation can be perceived to be dependant on genre and instrumentation.

1. INTRODUCTION AND RELATED WORK

A gestural control system has been developed that uses the Wii controller as an interface to control multiple monaural sound sources similar to [1, 2, 3]. An integral part of the system is correctly selecting a sound source for manipulating localisation position and gain level within an auditory mixture. The challenge is to give feedback to the user prior to altering a sound source's attributes. In this instance the feedback has to be auditory as the user may be within the auditory reproduction field and have no access to visual feedback. So this work is concerned with how best to give auditory feedback of source selection in a multi-track audio environment.

Several authors have also investigated the importance of auditory cues in related applications. In [4] it was shown that audio scene descriptors of environmental changes are often not needed for audio film, since other auditory cues of background noises provide strong cues. It was further shown that overlapping dialogue and background noises does not interfere with the listeners' perception of the scene being depicted or the spoken dialogue between characters. [5, 6] showed the importance of audio for environmental and condition changes in interactive applications. This indicated that in fields where there has been a strong historical link between audio and visual cues, especially in gaming where visuals are often given a much higher priority, audio-only versions can succeed just as well as visual cues if attention to detail is made regarding environment and object changes.

The research presented herein fits within the general context of auditory displays of human-computer interaction. However, it differs from previous work in that the system was designed to be used by audio engineers, where audio feedback is given for mixing audio material. User ratings of auditory feedback were investigated over a singular or multi-part audio system designed for gestural control. We consider two situations that may result in different solutions. If an audience is present then feedback needs to be given solely to the audio engineer, whereas when no audience are present the auditory cues can become part of the multi-track auditory mixture.

The paper is structured as follows. Section 2 describes the system for which we wish to provide the user with audio feedback. In Section 3, a pre-selection method is given in order to help the user select sound sources. Section 4 gives the proposed methods for confirmation of source selection. Details of the test system are given in Section 5. Results of evaluation are provided in Section 6, and the conclusions and directions for future work are discussed in Section 7.

2. A SYSTEM FOR GESTURAL CONTROL OF MULTI-TRACK AUDIO

For this paper the gesture control system is limited to adjustment of sound source position within a stereo reproduction field. The Wii controller is connected to OSCulater software (http://www.osculator.net/) running on a Mac computer. The data is then passed to Max/MSP as OSC data for real-time analysis and control of the software. The data used is the horizontal angular movement received from the MotionPlus attachment to the Wii controller. The OSC data received ranges from 0.0 to 1.0 representing a 180^o movement.

The controller angle θ_c is then calculated as:

$$\theta_c = (\psi_c - \psi_l)\left(\frac{90}{\psi_r - \psi_l}\right), \qquad (1)$$

ψ_c is the received controller current position value, ψ_l and ψ_r are the loudspeaker reference point values.

The user aligns the layout with the gesture control system by aiming the Wii controller at each of the two loudspeakers. The button 1 on the controller assigns the left loudspeaker and the 2 button assigns the right loudspeaker value. The stereo panning law used to move sound sources between the loudspeakers is the cosine/sine law [7]. This law ensures constant power; $\cos^2\theta + \sin^2\theta = 1$, where g_L and g_R are the gains of the left and right loudspeakers:

$$\begin{array}{c} g_L = \cos\theta \\ g_R = \sin\theta \end{array}, \qquad (2)$$

and the angular range of θ is 0^o to 90^o. The controller is limited so that if it is moved past the reference points the sound source will remain at 0^o or 90^o respective to which reference point is surpassed.

The system assumes the user to be equiangular from each loudspeaker using this mapping system although distance from the loudspeaker has no influence on mapping.

A source can have its location altered by selecting it using the ’B’ button and then moving the controller left or right between the referenced loudspeaker positions. Firstly $|\theta_n - \theta_c|$ is calculated for each sound source. Then in a second step the values are compared to one another and a priority scheme [8], is used for the case where two sound sources occupy the same position. The final step to calculate is if the closest source is within a specified localisation error θ_{le}:

$$|\theta_n - \theta_c| \leq \theta_{le}. \qquad (3)$$

After this final step the track number of the relevant sound source is passed to the system to allow movement of the sound source whilst the ’B’ button is held, $\theta_n = \theta_c$. Once the ’B’ button is released the sound source remains at the exact position is was in at the moment ’B’ was released.

The panning angle of a sound source, θ_n, always ranges from 0^o to 90^o, regardless of the panning law. However, physical layouts vary. 0 to 60 is often used for the physical angle, based on forming an equilateral triangle with endpoints at the listener and the two loudspeaker positions. Higher physical separations can result in loss of phantom image, thus ruining the psychoacoustic reasoning for which the panning laws used. Thus the physical placement of the sound source is given by $(60^o/90^o)$, $\theta_p = \frac{\theta_n}{1.5}$, as shown in Fig. 1. For the remainder of the paper, angles will be given in respect to the panning law, not the physical layout, so that they are directly transferable to setups with different user-speaker distance and stereo aperture.

3. PRE-SELECTION OF SOURCES

To assist with selection confirmation, a pre-selection method can be used to help locate a sound source rather than relying solely on the listener’s auditory localisation. An ”Audio Search Light” is proposed that plays a selection of the audio mixture via a bluetooth headset based upon the controller angle θ_c.

This pre-selection is independent of the main stereo audio mixture heard via the loudspeakers. A gain is calculate for each of the sound sources based on its distance in degrees from the position the controller is aimed at. A sound source that is to the right or clockwise on the stereo arc will be positive whilst an anti-clockwise/leftwards source will be a negative. The gain curve decreases from a multiplier of 1.0 where the controller position is equal to a sources position, as the controller moves away the gain decreases following a cosine based curve until the multiplier reaches 0. At this point the gain curve does not go into negative multiplier like a cosine would but remains at 0.0 so the source is not included in the audio mixture.

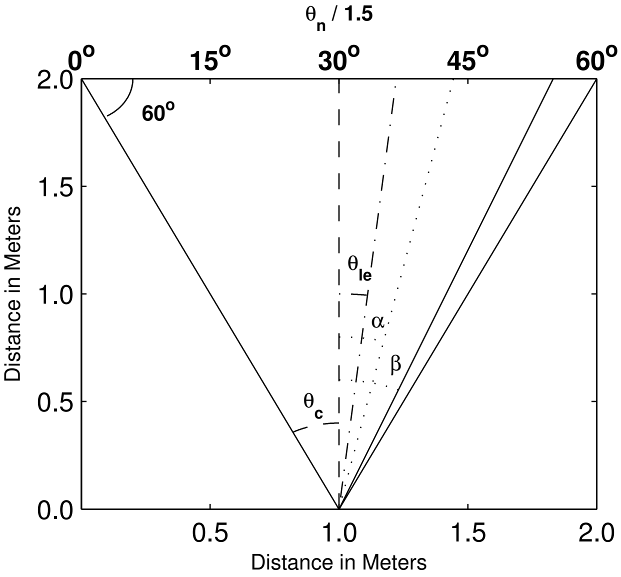


Figure 1: The physical layout of sound sources, $\theta_n/1.5$ within the 2.0m equilateral triangle with reference to the angles used in the paper; θ_c, θ_{le}, α and β.

Using [9], [10] a symmetrical gain curve can be used to achieve this. Defined by α, the -3dB point or 0.707 on the linear scale, degrees from the controller position. By altering this value sources that that are not equal to the controller position will have the gain multiplier increased or decreased in the α is increased or decreased. A factor q is calculated to shape the width of the gain curve:

$$q = \frac{1 - \sqrt{2} + \cos\alpha}{1 - \cos\alpha}, \qquad (4)$$

where the gain of a sound source in relation to the controller position (θ_c) is given as:

$$d_n = max\left\{\frac{1}{2}(1 - q + (1+q)\cos(\theta n - \theta c), 0\right\}. \qquad (5)$$

Therefore the entire ”audio search light” mixture is given as the product of each instrument track S_n and its calculated gain d_n:

$$\sum_{n=1}^{n} S_n d_n. \qquad (6)$$

The distance to the zero point, where the gain multiplier first becomes 0.0, from the controller position, β, is calculated as:

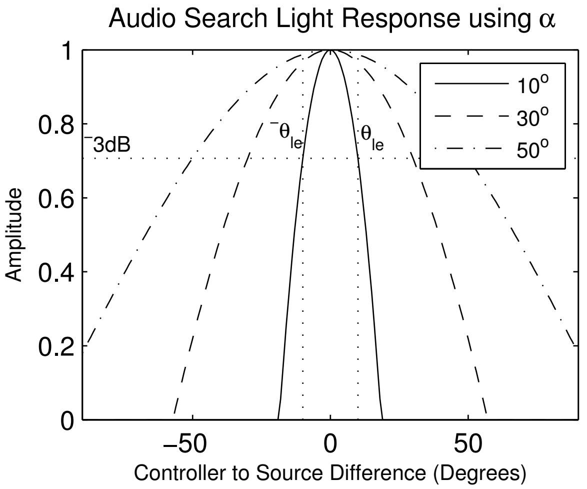


Figure 2: The "Audio Search Light" gains for α set to 10, 30 and 50 degrees. The dotted line show where θ_{le} points are vertically and the -3dB (0.707) point horizontally, this allows visual comparison between parameters used for he alteration of the audio search light.

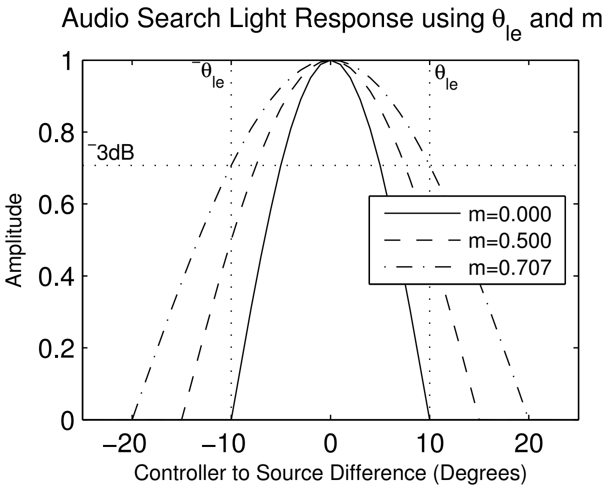


Figure 3: The "Audio Search Light" gains for $m = 0, 0.5$ and 0.707 at the localisation error points $\pm\theta_{le}$ (vertical dotted lines). In decibels the points are equal to $-\infty$, -6 and -3 dB full scale. The horizontal dotted -3dB (0.707) line is shown to compare to Fig. 2.

$$\beta = \cos^{-1}\frac{q-1}{q+1}. \tag{7}$$

Figure. 2 shows the gain curve based on widths α of 10^o, 30^o and 50^o. The horizontal dotted line denotes the -3dB points. For 10^o the auditory mixture produced is comprised mainly of the instrument directly pointed at having a β of 18.5^o, whilst the 30^0 width has a mixture using sources up to 57.1^o from the controller position. When the value is at 50^o we can calculate that in the case of a stereo panning law given in Eq. 2, where $0 \leq \theta_n \leq 90$, that every audio source has a gain > 0.0 since β is 102.7^o. Examples of $\alpha = 10^o$ 📎, $\alpha = 20^o$ 📎, $\alpha = 30^o$ 📎 and $\alpha = 50^o$ 📎 where $\theta_c = 45^o$ are embedded into this paper. However in some situations it may be prudent for the audio search light can also be defined in terms of the localisation error of the system θ_{le} (described in the following section) and an associated gain factor m. First a remapped angular value is calculated

$$x_n = \frac{(\theta_n - \theta_c)\cos^{-1} m}{\theta_{le}}, \tag{8}$$

where the whole audio mixture is made up of S_n sources each with $\cos x_n$ gain multiplier:

$$\sum_{N=1}^{N=n} S_n \cos \begin{cases} 90 & (x_n > 90) \\ x_n & (-90 \leq x_n \leq 90) \\ -90 & (x_n < -90) \end{cases} . \tag{9}$$

Figure. 2 depicts the audio search light gains for $m = 0$, 0.5 and 0.707.

The latter version of the audio search light was not used in this paper as the preferred value from participants would be in reference to a given θ_{le}. In some work it may be beneficial to test or configure the system whereby the pre-selection gain curve has a relationship with the localisation error of the system.

Figure. 4 shows an example setup where five sound sources are equally spread through the stereophonic field, whereby the leftmost and rightmost are equal to the loudspeaker positions. The gains applied to the different sound sources can be seen by the solid curve on the left hand side of the figure. The middle sound source has a gain equal to 1.0 because it's position is equal to θ_c at 0^o, the next closest two sources both have a gain just over 0.6 whilst the final two sources are both beyond $\pm\beta$ so will have a gain of 0. The user would be facing 0^o and their distance does not alter the pre-selection gain curve. The right hand side of the plot shows the angles of α, β and θ_{le} and how these angles relate to the different instrument tracks.

4. SELECTION CONFIRMATION

Various methods have been tested to attain the preferred and most reliable for delivering sound source selection confirmation to the audio engineer.

A sound source is selected by pointing the Wii controller in the direction of a sound source and pressing the B button on the bottom of the controller. There is built into the system a localisation error θ_{le}, given as the angle allowed either side of the controller angle where a source will still be selected, in this case 10^o. This feature has been incorporated to overcome errors in human auditory localisation and mapping errors.

The options for auditory feedback were:

- 📎 The name of the sound source is given as an audio scene descriptor when the source is selected e.g. Guitar, Vocal, Drums, as first documented by Frazier [11]. This can be played over the stereo loudspeaker system.
- 📎 An audio scene descriptor is given as above, but over the bluetooth headset making this option viable for audience

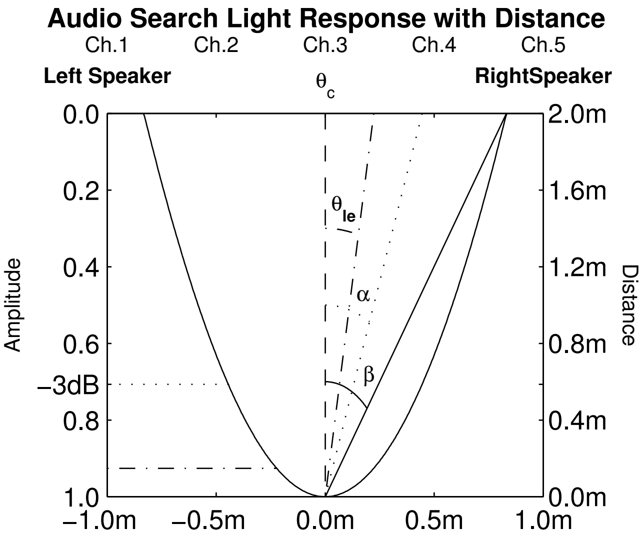


Figure 4: A setup where $\alpha = 20^o$. There are 5 equally spaced sound sources. The controller angle is pointed at 0^o and the two outer sources are at the loudspeaker positions. The (solid line) gain curve shows the gains that would be used to create the audio search light audio mixture. The horizontal dotted lines indicate represent the -3dB point with corresponding angle 20^o and the horizontal dot-dashed line represents the amplitude reduction at localisation error, $\pm\theta_{le}$.
The gain curve shows the gains that would be used to create the ”audio search light” audio mixture. The dotted horizontal lines represent the -3dB point and the dot-dash lines the amplitude for the $\pm\theta_{le}$ points.

based situations.

- The sound level of the selected source is 6dB louder than the other sound sources in the multi-track mix for a period of 3 seconds. This is achieved by means of ramping down the other tracks by 6dB over 300ms, holding for 2400ms and ramping back up for 300ms. This method is only suitable for non-audience application.

- The selected sound source is reproduced through the bluetooth headset for 3 seconds. The source is ramped up over 300ms, held for 2400ms, and ramped back down over 300ms. In order to avoid severe comb filtering the headset signal is delayed by the users distance from the speaker.

5. TEST SYSTEM DESIGN

For the testing five different music mixes of varying genres were used; latin, jazz, metal, balkan and classical. Each song contains 5 different individual tracks equally spread over the audio stereophonic field; 0^o, 22.5^o, 45^o, 67.5^o and 90^o with the order of musical mixes randomised for each participant of the test.

Prior to carrying out tests the audio tracks for each song were automatically mixed using the ITU-R BS.1770-1 standard [12] so that the average loudness was -24LUFS. Since each audio mixture includes 5 instruments their average loudness is -30.99LUFS $(-24 - 10\log_{10} 5)$). The audio scene descriptor for each instrument name was therefore average loudness matched to the instruments at -30.99LUFS to avoid biased test results.

The loudspeakers are calibrated before performing the listening tests using the AES described method [13] where the total SPL is to be 85dB A-weighted and therefore each speaker should be $82.0dB = 85 - 20\log_{10}(n), n = 2$. The calibration file used was band-limited pink noise, 500-2500Hz as found on the Blue Sky website, http://abluesky.com/ support/blue-sky-calibration-test-files/, set at a level of -24dB to match the average loudness of the audio mixes, each channel in turn was calibrated.

The bluetooth headset has a delay on reproducing the audio. Thus the audio signals of each instrument, pre-gesture control, were delayed by 280ms to align the signals. The delay was calculated using [14]. The delay to account for user distance was then subtracted from this value to align the signals in the real space.

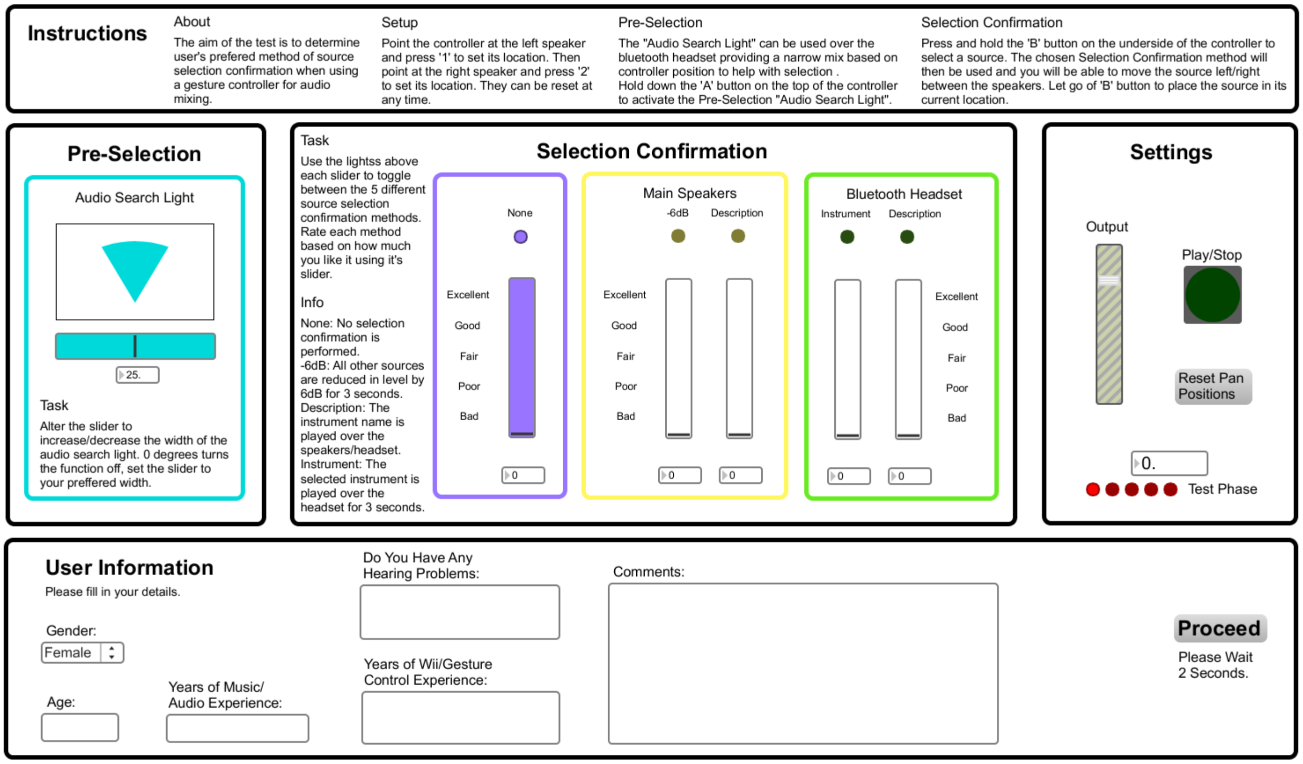
Two Genelec 8020A speakers were used for stereo output. The layout conformed to the 60^o aperture guidelines, making an equilateral triangle between the two stereo speakers and the listener, with the listener 2.0m from each speaker. The bluetooth headset used was a Jabra BT530, worn in the right ear of the user.

For each audio mixture the user was asked to set a parameter that alters α for the pre-selection ”Audio Search Light.” This ranged from $0^o - 50^o$ in steps of 0.5^o, where a setting of 0^o meant that the user preferred the ”Audio Search Light” turned off. There were five source selection confirmation methods; none, -6dB reduction for all other sources over the loudspeakers, audio descriptor over loudspeakers, instrument played over headset and audio descriptor over headset. The users were asked to rate each method 0 to 100 to whole integer steps, with general Mushra [15] type labelling; bad, poor, fair, good, excellent. However unlike a formal Mushra test, there was no reference or anchor available. Thus test subjects did not necessarily use the full range of the scale.

For each audio mixture the user is asked to set variable parameter that alters α of the pre-selection ”Audio Search Light” that ranges $0^o - 50^o$ in steps of 0.5^o. The setting of 0^o means that the ”Audio Search Lights” has no function and the result will show that the user preferred this as well as the preferred width of the audio mixture reproduced over the bluetooth headset for other users. For the five source selection confirmation methods; none, -6dB of all other sources over the loudspeakers, audio descriptor over loudspeakers, instrument played over headset and audio descriptor over headset, the users are asked to rate each method 0 to 100 to whole integer steps, with general Mushra [15] type labelling; bad, poor, fair, good, excellent.

The interface presented to the participants is shown in Fig. 5 with instructions at the top, the setting for the pre-selection given in the left part of the middle section, the ratings for the selection confirmation methods is given in the middle part of the middle section, settings given to the right of the ratings and finally user information is collected at the bottom of the interface. Extra information given to the participant is the test phase out of 5 and the angle the controller is pointing at in accordance to the pan angle $(0 \leq \theta_c \leq 90)$ rather than physical angle. The interface and gesture control system is developed using Max/MSP, http://cycling74.com/products/maxmspjitter/, a real-time audio graphical environment.

Two audio examples of the system are given. First feedback over main speakers using audio descriptors using the latin song. Secondly, sources being moved in the jazz song with -6dB



Instructions

About

The aim of the test is to determine user's prefered method of source selection confirmation when using a gesture controller for audio mixing.

Setup

Point the controller at the left speaker and press '1' to set its location. Then point at the right speaker and press '2' to set its location. They can be reset at any time.

Pre-Selection

The "Audio Search Light" can be used over the bluetooth headset providing a narrow mix based on controller position to help with selection .
Hold down the 'A' button on the top of the controller to activate the Pre-Selection "Audio Search Light".

Selection Confirmation

Press and hold the 'B' button on the underside of the controller to select a source. The chosen Selection Confirmation method will then be used and you will be able to move the source left/right between the speakers. Let go of 'B' button to place the source in its current location.

Pre-Selection

Audio Search Light

25.

Task

Alter the slider to increase/decrease the width of the audio search light. 0 degrees turns the function off, set the slider to your preffered width.

Selection Confirmation

Task

Use the lightss above each slider to toggle between the 5 different source selection confirmation methods. Rate each method based on how much you like it using it's slider.

Info

None: No selection confirmation is performed.
-6dB: All other sources are reduced in level by 6dB for 3 seconds.
Description: The instrument name is played over the speakers/headset.
Instrument: The selected instrument is played over the headset for 3 seconds.

User Information

Please fill in your details.

Gender: Female

Age:

Years of Music/ Audio Experience:

Do You Have Any Hearing Problems:

Years of Wii/Gesture Control Experience:

Comments:

Please Wait 2 Seconds.

Figure 5: The test interface presented to the participants of the study. The test interface was developed using Max/MSP real-time graphical based audio programming software.

feedback .

6. RESULTS

There were 8 male and 3 female participants, ranging in age from 23 to 38 years old. All participants had good hearing and were experienced in working with audio. Only a few had some experience of using gesture control previous to this experiment, though most were used to mixing music. Feedback towards this research was very encouraging with most enjoying the gesture based mixing system. For instance, one participant stated, ”overall I thought it was a really nice application and could definitely imagine wanting to use some kind of personalised mixing at home.” The author received feedback indicating that ”A lot of practice is needed”, though this does not necessarily mean it is unusable or not useful. Audio engineers at whom this interface was aimed may spend years honing their technical and creative skills, as do fellow researchers in related fields.

6.1. Pre Selection ”Audio Search Light”

The authors found in the tests that the average, preferred α width was 20.05^o discounting the three times it was rated as ’Off’. This was surprising to the authors since their personal preference was for a narrow beam, approximately 10^o matching the -3dB point with θ_{le}. This was also supported by one user, who stated ”Audio search light was the most useful with narrow focus”.

Figure. 6 shows a bar graph of the combined results of all the songs and the distribution of preferred values for the Audio Search Light width divided into 5^o bands, and a band for ’Off’ to show when the user disliked the function.

Song	α
Latin	18.95^o
Classical	20.85^o
Balkan	20.05^o
Jazz	21.00^o
Metal	19.40^o
Average	20.05^o

Table 1: The user settings for the ”Audio Search Light” width in terms of α in degrees.

Table. 1 shows the average values set for α for each song and the overall average. Values of 0^o were excluded as they meant the user did not like the feature. We can conclude from these results that the audio search light function is not influenced significantly by the genre of music or the instrumental make-up of the audio mixture because the averages differ only by $\pm 1.09^o$. These averages excluded the ’Off’ position which was only used 3 out of 55 times, and by only one test subject.

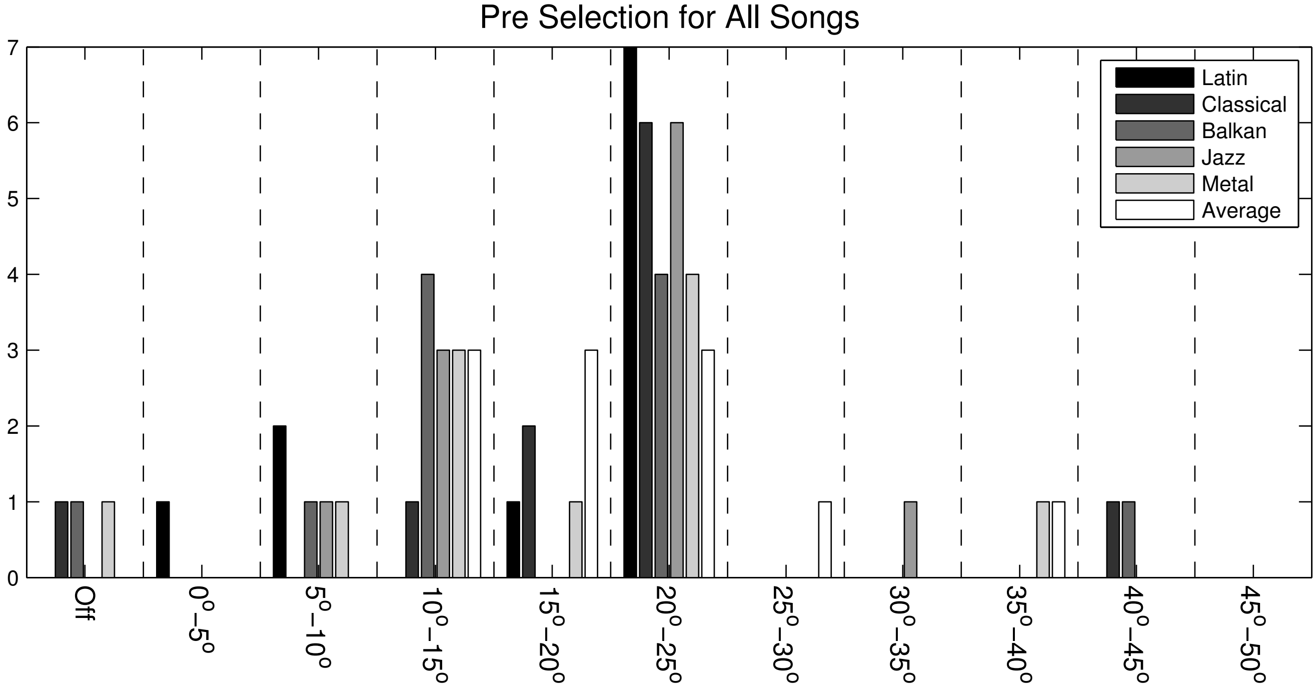


Figure 6: The ”Audio Search Light” results for all songs and the average over all songs for user preference of α.

6.2. Selection Confirmation

Figure. 7 depicts the mean value and confidence intervals using T-distribution, available at http://www.sjsu.edu/faculty/gerstman/StatPrimer/t-table.pdf, for each of the auditory feedback methods over the 5 different songs and an average of the results.

When we look at the average we can see that 'None' had the lowest mean (37.24), 15.69 points lower than the second lowest rated method. Only in the 'Classical' song is it not rated lowest. This may be explained because in that piece of music, instruments come in and out of the audio mixture. If only 1 or 2 instrument are playing simultaneously, the drop of -6dB of other instruments is less noticeable. One test subject commented, ”bits of silence in some instruments (e.g. classical) made instrument-only based confirmation methods less reliable. Generally -6dB was the least helpful to me.” . The 'Description' was also rated lower than 'None' for 'Classical' music, which indicates that the overlaid description in the main loudspeakers was intrusive to the more gentle style of music.

The two methods using the main loudspeakers for selection confirmation feedback, 'Description' and '-6dB', had the next lowest mean scores, 52.93 and 54.20 respectively. '-6dB' however has a wider confidence interval, indicating disagreement among users regarding the preference for this method. This was also shown in the comments, which included statements such as ”The -6dB seems like the most elegant idea” and ”Generally -6dB was the least helpful to me”. In all but 'Jazz' and 'Balkan' these two methods were rated lower than the methods used via the bluetooth headset. One reason 'Description' scored poorly in the 'Balkan' song, according to one test subject, was that ”The description isn't very useful in both the speakers and headset especially when I'm not familiar with the instrument names” .

Overall, the highest scoring auditory feedback methods were the two bluetooth headset methods; 'Instrument' (62.22) and 'BT Description' (64.80). The main difference for the top two results was for the 'Jazz' piece of music where the 'BT Description' was rated highest. This may be to do with the instrument makeup of the piece, which featured two different saxophones and a trumpet playing in harmony with one another during most of the piece. Even with an increased level for the participant to recognise their choice, it still was not always helpful in identifying a specific instrument. As noted by one participant, ”The genre made a fair bit of difference as different types of music had a variety of different dynamic ranges for different instruments.” This was especially true for harmonising instruments that were lower in the mix, and therefore with gain boost still not clearly audibly louder to the participant.

A reason for the 'Instrument' feedback not receiving higher ratings could be that, according to one test subject, ”ear piece causes confusion in perception of panning”. With both audio signals being time aligned, the stereo panning could be somewhat deteriorated, especially when the headset was worn in just one ear. To counteract this the 'Instrument' could be delayed by a matter of a few milliseconds, making use of the precedence effect [16]. This would result in the localisation information being correct as to the stereo position, but would also result in an increased gain.

The results showed small error bars, although the highest ranking method, BT Description, scored a mean only 27.56 points above the lowest ranking method, 'None.' This is most likely due to the lack of a hidden reference and anchor, which are used in Mushra tests to ensure that the full range of the scale is used. In this case it was not possible to provide a reference since an ideal feedback method was not known.

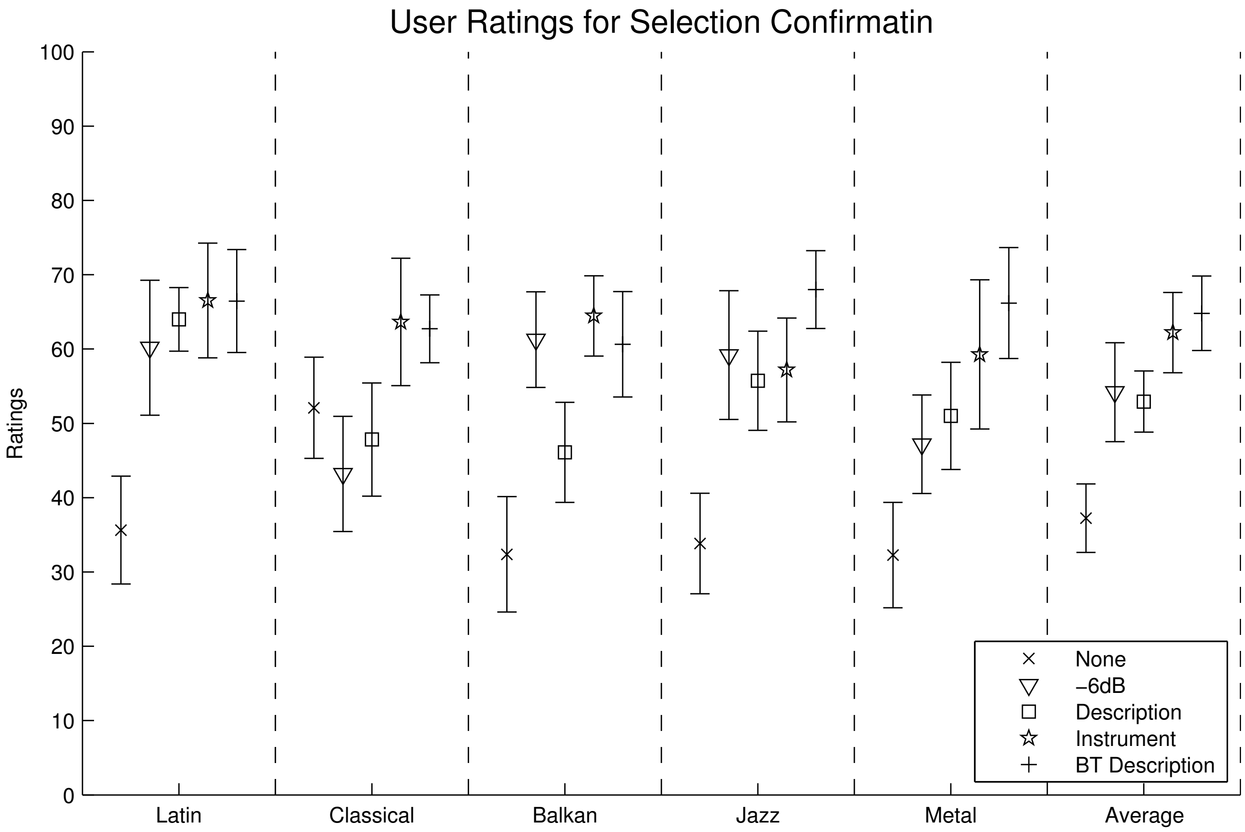


Figure 7: Error bars show 85% confidence intervals using T-distribution of the participant's rating for each selection confirmation method for each song as well as the average ratings for each method.

7. CONCLUSION

We have described evaluation of auditory feedback methods for mixing multi-track audio by use of a gesture controller. The participants' comments indicated that they liked the proposed preselection method, with only 1 participant turning it off for 3 out of 5 songs, i.e., an overall 5% 'Off' rating. When we considered the average pre-selection width, we found a preferred value of 20.05 ±1.090, indicating that the genre type had no significant impact.

However, results showed that preference for source selection confirmation method had some bias based on genre. The bluetooth headset using an audio scene description (in our case, the instrument name that had been selected), was rated highest. This was intended for when an audience is present, since it does not affect the stereo audio mixture. However, if a user wishes not to use a headset, then the '-6dB' selection confirmation was rated best, although this method was less useful for genres where instruments came in and out of the musical mix.

In this paper, the system was designed for gestural control of stereo placement of sources using a Wii controller. However the work may be expanded to multi-track surround and 3-dimensional reproduction formats, to other mixing tasks in addition to placement of sources, or other forms of gestural control. In which case the results may be considered indicative, but not definitive in their recommendations for auditory feedback.

8. ACKNOWLEDGMENT

The authors wish to thank all the volunteers from Queen Mary University of London who graciously gave up their time to take part in this study.

9. REFERENCES

[1] P. Quinn, C. Dodds, and D. Knox, “Use of novel controllers in surround sound production,” in *Audio Mostly*, Glasgow, 2009.

[2] C. Kiefer, N. Collins, and G. Fitzpatrick, “Evaluating the wiimote as a musical controller,” in *International Computer Music Conference*, Belfast, 2008.

[3] R. Selfridge and J. D. Reiss, “Interactive mixing using wii controller,” in *Audio Engineering Society Convention 130*, London, UK, 06 2011.

[4] M. J. Lopez and S. Pauletto, “The design of an audio film for the visually impaired,” in *iCAD*, M. Aramaki, R. Kronland-

Martinet, S. Ystad, and K. Jensen, Eds. Copenhagen, Denmark: Re:New – Digital Arts Forum, 18—21 May 2009. [Online]. Available: Proceedings/2009/LopezPauletto2009.pdf

[5] N. Moustakas, A. Floros, and N. Kanellopoulos, "Eidola: An interactive augmented reality audio-game prototype," in *Audio Engineering Society Convention 127*, 10 2009. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=15067

[6] N. Paterson, K. Naliuka, T. Carrigy, M. Haahr, and F. Conway, "Location-aware interactive game audio," in *Audio Engineering Society Conference: 41st International Conference: Audio for Games*, 2 2011. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=15769

[7] D. Griesinger, "Stereo and surround panning in practice," in *Audio Engineering Society Convention 112*, 4 2002. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=11308

[8] E. P. Gonzalez and J. D. Reiss, ""a real-time semi-autonomous audio panning system for music mixing," *EURASIP Journal on Advances in Signal Processing*, vol. 2010, no. Article ID 436895, p. 10, 2010.

[9] C. Faller, "A highly directive 2-capsule based microphone," in *Audio Engineering Society Convention 123*, 10 2007. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=14370

[10] C. Faller, A. Favrot, C. Langen, C. Tournery, and H. Wittek, "Digitally enhanced shotgun microphone with increased directivity," in *Audio Engineering Society Convention 129*, 11 2010. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=15609

[11] G. M. Frazier, "The autobiography of miss jane pitman: An all-audio adaptation of the teleplay for the blind and visually handicapped, film and communication," Master's thesis, San Francisco State University, 1975.

[12] ITU-R, "Bs.1770 : Algorithms to measure audio programme loudness and true-peak audio level," International Telecommunication Union, Tech. Rep., 2007-09.

[13] AES, "Multichannel surround sound systems and operations," AES Technical Council, Tech. Rep. AESTD1001.1.01-10, 2001.

[14] A. Clifford and J. Reiss, "Calculating time delays of multiple active sources in live sound," in *Audio Engineering Society Convention 129*, 11 2010. [Online]. Available: http://www.aes.org/e-lib/browse.cfm?elib=15580

[15] ITU-R, "Bs.1534 : Method for the subjective assessment of intermediate quality levels of coding systems," International Telecommunication Union, Tech. Rep., 2003. [Online]. Available: http://www.itu.int/rec/R-REC-BS.1534/en

[16] K. Wilson and T. Darrell, "Improving audio source localization by learning the precedence effect," in *IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005. Proceedings. (ICASSP '05).*, vol. 4, March 2005, pp. iv/1125 – iv/1128.