Influence of performance gestures on the identification of spatial sound trajectories in a concert hall

Georgios Marentakis, Joe Malloch, Nils Peters, Mark Marshall, Marcelo Wanderlay and Stephen McAdams

CIRMMT, McGill University, 527 Sherbrooke St. West, H3A 1E3 Montreal, Quebec, Canada

ABSTRACT

An experimental study was performed on the effects of the visibility of a performer's gestures on the identification of virtual sound trajectories in the concert hall. We found that when working in synchrony, the performer's gestures integrate with the audio cues to significantly increase identification performance, normalize for the effects of off-centre listening in the hall and overcome problems related to the complexity of the soundscape. In the absence of visual cues, identification performance depends on the listening seat, the sound trajectory and the complexity of the soundscape.

1. INTRODUCTION

Gesture interaction with spatialized sound is becoming increasingly popular in human-computer interaction designs. In a number of application areas such as mobile computing, presentation of background information and applications for the visually impaired, interacting with spatialized audio is known be a usable solution. We are investigating a novel application domain: the gestural control of spatialized sound for musical purposes. This work forms part of a larger project that looks into how spatial sound can be integrated into the music creation process from the composer to the performers and the audience. To this end, we evaluate the extent to which the composer's spatial intentions are communicated to the musicians and the role the performer's gestures play in this process.

The motivation for this paper stems from the lack of evaluation studies that address the aforementioned problems together in their natural environment: the concert hall. The application domain poses interdisciplinary questions that span the domains of human-computer interaction and psychoacoustics. A very common practice when composing with space is to use spatial audio trajectories. Their identification cannot be taken for granted, however, especially in realistic settings such as concert halls. This is due to the relative inefficiency of the auditory system in processing spatial information as well as the fact that most spatial audio systems are designed for the center of the listening area. The consequences for the perception of sound location and movement in off-center listening positions are not well known.

Implementation of the performance of sound trajectories is to a large extent an open design question. Many composers prefer pre-programmed spatial manipulations of sounds. This however, results in a unimodal experience for the audience because the action that triggers the spatial event is not visible. Other realizations consider using ancillary musical gestures for spatialization, in which case the spatialization action is not directly visible and the audience has to infer the mapping chosen by the composer [1]. In other cases, a third person, for example a sound engineer, performs the spatialization based on the score. However, the placement of the interface (computer and mixer) results in the actions not being visible to the audience. The possibility of using tracking devices to do gestural control of spatialization in a manner akin to direct manipulation opens new dimensions in performing spatial manipulations in music. The performer can be placed on stage together with the musicians and perform the trajectories so that their spatial semantics are mapped in the gestures. This interaction paradigm is investigated in this paper and compared to the traditional non-visible methods of performing space.

In its simplest realization this type of interaction is a tracking task. The performer has to move the sound along a certain path with velocity constraints. A number of questions arise due to the fact that the performer inadvertently receives audio feedback in the concert hall, the spatial fidelity of which depends on his position therein. If the spatial fidelity of the cues is to be maximized, the performer should be placed in the sweet spot within the audience. However, this would degrade the performance aspect, since they will not be visible by the audience. If the performer was to be placed on stage, visibility would be restored, however they would receive low fidelity audio feedback since they would be out of the loudspeaker array. A solution to this problem can be achieved by providing a binaural mix to the performer. However, no research has looked into the effects of audio feedback on target tracking.

To answer our research questions, we performed an evaluation study. From the performer's point of view, we investigated tracking a visual target along a two dimensional path by hand movements in three dimensions, with or without binaural audio feedback. From the audience point of view, we looked into the perception of the sound trajectories at different positions in the audience in the presence or the absence of visual feedback from the performer. The literature review that follows provides evidence on the importance of musician's gestures for the perception of musical performance and investigates the problems of virtual spatial audio systems in particular when deployed in concert halls.

1.1. Influence of musician's gestures

Davidson [2], examined how ratings of the expressiveness of music performance would be affected by the cross-modal information made available through the musicians' gestures. In her experiments, visual information was found to be the most important cue for distinguishing between levels of expressiveness, with performance deteriorating significantly when visual cues were not provided. Visual cues assisted the perception of subtle, expressive differences in the audio stream.

Vines et al. [3] examined the effect of cross-modal interactions on the perception of tension and phrasing. Their findings suggest that there is an emergent quality when performers are both seen and heard. They also found that hearing dominates perception of musical tension. Perceived tension increases for synergetic audio and visual cues and decreases for conflicting ones. With respect to phrasing, responses to audio and visual cues show high correlations. Some discrepancy was observed with respect to a performer's

anticipatory gestures when entering or preparing to exit a musical phrase, a fact that results in the responses to audiovisual or visual-only phrasing cues to lead responses to audio-only cues.

The interpretation of the gestures is to a large extent a social phenomenon and depends very much on the existence of a common ground between the performer and the audience. The reader is directed to Kurosawa and Davidson [4] for a classification of musical gestures. In this study, we choose to align the geometry of the performer's gestures with the spatial audio trajectory in order to provide unambiguous feedback with respect to sound movement and minimize subjective interpretation.

1.2. Perception of sound direction and movement in real and virtual environments, anechoic and enclosed spaces

A review on sound localization is provided in order to show that the perception of sound trajectories in concert halls can depart from the expectations arising from our experience with vision. This is depicted in perceptual studies, where localization accuracy is measured by either estimating absolute localization error or by measuring the Minimum Audible Angle [5-7]. Both measures are relevant for understanding spatial hearing. The first depicts the ambiguity in sound position and the second the smallest perceivable angular sound displacement from a given starting position. The reader is directed to the relevant literature for exact measurements, however it should be noted that under certain conditions, such as with lateral or elevated sounds, significant ambiguity is present with respect to the exact location of the sound event, yielding larger MAAs. Similar methodologies have been developed to study sound motion. The Minimum Audible Movement Angle (MAMA) is the angular distance a moving sound needs to traverse before its movement is perceived by a listener. MAMAs increase linearly with the velocity of a sound [6]. For quick movements, therefore, care must be taken so that the distance traveled is enough to provide the desired cue.

The perceived distance to a sound in enclosed spaces is determined based on the ratio of direct to reverberant energy as this is calculated within a short time window on the order of 6ms [8]. Information on the intensity of a sound is also used. However, it is confounded with the intensity of the sound source itself. Sensitivity to distance changes depends on the magnitude of the cues. Distance manipulations of small magnitude are therefore hard to perceive. Spectral changes, such as the attenuation of high frequencies in the air, can also affect distance perception, however they are confounded with the sound source spectrum, therefore their influence also depends on familiarity with the source. For unfamiliar sounds, common in electroacoustic music concerts, absolute distance perception is therefore further degraded.

In enclosed spaces, localization accuracy is affected to a varying extent by early reflections and late reverberation and can be severely degraded for sounds with slow onsets. The effect of reflections depends on their time of arrival and their level. Reflections arriving within a certain time window contribute to the localization of sounds, a phenomenon described as 'summing localization'. The length of the time window is about 1 ms for noise stimuli but can be longer for stimuli with slow onsets such as sine tones [9]. Reflections within the first 50-80 ms are grouped with the direct sound as long as their level is below a certain threshold, otherwise they are perceived as separate events. They do not affect localization, but influence the spatial impression, that is the auditory source width and the listener envelopment. If a listener does not have good exposure to the direct sound, sound localization can be distorted.

1.2.1. Auditory Virtual Environments

In concerts, loudspeaker arrays are used to present phantom sources in locations where no sound source or loudspeaker is otherwise present. A very common system of this kind is Vector-Based Amplitude Panning, or VBAP [10], which is essentially a method to generalize amplitude panning in arbitrary 2D and 3D loudspeaker setups. Other techniques include Ambisonics [11, 12] and Wavefield synthesis [13]. The Ambisonics technique approximates the measured or estimated sound field at a certain point using a variable number of loudspeakers. Ambisonics systems are differentiated by their order, which maps to the degree of approximation of the sound field. In wavefield synthesis systems, the sound field over an area, as opposed to the sound field at a point, is reconstructed. This is done by sampling or estimating the real or virtual sound field and then reproducing it using loudspeaker arrays. The accuracy of the method depends on the number of loudspeakers and the spacing between them. The latter introduces a limitation for the highest frequency that can be reproduced accurately, without spatial aliasing. In practical situations, it is hard for this frequency to exceed 1200 Hz. Another technique inspired by sound recording is ViMiC [14], where virtual microphones are placed in a virtual room, in which sound source propagation is simulated. The estimated microphone signals are subsequently reproduced by loudspeakers in a real room.

Localization accuracy in these systems has been little studied, and maps of localization accuracy, such as those provided for real sounds in anechoic conditions, are not largely available. Here we present relevant evaluation studies. Gröhn [15] performed a study on the localization of a moving virtual source in a virtual room and studied the effect of a distracting auditory stimulus. Participants pointed to the perceived trajectory of a moving sound, in a virtual audio system that used VBAP. The authors found that the localization error was higher compared with static sounds and also that the virtual system and room introduced higher localization errors compared to anechoic conditions. In addition, the presence of a distracting stimulus increased azimuth localization error.

Pulkki and Hirvonen [16] performed a comparative study on the localization of virtual sources in multichannel audio reproduction in an anechoic environment. The systems compared were first- and second-order Ambisonics, a spaced microphone technique and pair-wise panning in an anechoic environment. For comparisons in a 5.1 audio system, localization accuracy was better when pair-wise panning was used, followed by the spaced microphone technique and the 1st-order Ambisonics algorithm. For an eight-loudspeaker system, pair-wise panning still prevailed over 2nd-order and 1st-order Ambisonics. Similar results have been found by Guastavino et al. [17].

Bates et al. [18], performed a study on the localization accuracy of advanced spatialization techniques in small rooms and compared it with monophonic presentation. Listeners had to report which of the sixteen visible loudspeakers emitted the sound stimulus. Localization judgments were made from nine seats within the loudspeaker array. Stimuli were white noise, male speech, female speech and music. For virtual sounds, no significant variations were found for the different stimuli. The comparison ranked the algorithms in order as VBAP, Delta Stereophony and Ambisonics in terms of localization accuracy. Compared to monophonic sounds localization in the virtual systems was worse.

Based on the results of literature review we decided to use IRCAM's SPAT spatializer (Jot and Warusfel 1995) in the VBAP mode for our experiment because it provides the most reliable localization cues as well as a room model that would make the presentation more immersive. In addition, following the finding that a distracter would influence the accuracy of tracking a trajectory, we introduced distracters in the soundscape. In this study, however, we depart from the classical localization ideas and focus on the identification of sound trajectory shapes. This is more relevant to the scope of the paper, because we are interested in how people perceive sound trajectories, i.e. whether they will be able to identify them as opposed to how well the target sound can be localized. This is a higher-level cognitive process that requires listeners to integrate the localization cues into a continuous trajectory.

2. EXPERIMENT

The experiment was designed to provide insight into the following research questions:

1. How well are sound trajectories identified in different seats in a concert hall?
2. What is the effect of visibility of spatial performer's gestures?
3. Will the error introduced by a performer tracking the gestures affect the identification performance?

The experiment also aims to provide an initial investigation into the effects of audio feedback in a visual tracking task.

2.1. Experimental Design

There are four independent variables in the experiment: Trajectory (4 levels), Listening Position (9 levels), Display (3 levels) and Soundscape (2 levels).

Identification performance was measured for four different spatial audio trajectories, at nine different seats, with or without distracting sounds. Trajectories were presented through three different displays: audio only, audiovisual with the on-stage performer hearing sound spatialized out in the hall, and audiovisual with the performer hearing sound over headphones in a binaural rendering of the spatialization in the hall from the ideal listening position. Display was tested as a between-subjects variable in three independent experimental sessions. Listening Position, Soundscape and Trajectory were tested as within-subject variables with four repetitions for each unique combination of their associated levels.

2.2. Method

2.2.1. Participants

Participants were recruited from McGill University. Twenty-seven participants, 14 male and 13 female were split in the three experimental sessions. Mean age was 26 years and all participants reported having normal hearing. They were paid $10. A male performer performed the gestural control of spatialization in the two display sessions in which this was required.

Figure 1. A photo of the concert hall, illustrating the placement of the participants in the audience and the performer on stage. The performance was live (magnified in the picture).

2.2.2. Apparatus & Materials

The experiment was performed in Pollack Hall at the Schulich School of Music of McGill University [Figure 1 and Figure 3].

The trajectories were a straight line and an arc, as well as two modulated variations of them, called wobbly line and wobbly arc from here on. They were chosen to have identical start and end points to avoid identification based on absolute starting and ending positions [Figure 2]. When spatialized, the straight line would move right through the middle of the hall. The wobbly line would follow a similar trajectory, however it would swing from left to right. The arc and the wobbly arc move only in one side of the hall.

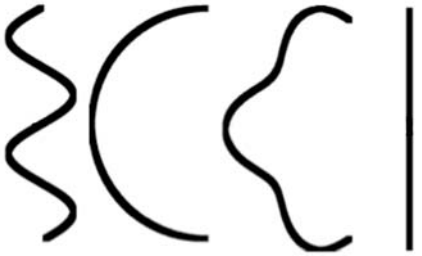

Figure 2. The four spatial sound trajectories used in the experiment as they were visualized in the graphical user interface

Sound spatialization was done using the Spat system, which comes as an add-on to MAX/MSP system (www.cycling74.com). To implement the sound system, eight Meyer UPJ-1P loudspeakers were placed as in Figure 3. Participants were seated as in Figure 3 (circled numbers). They were told to follow the movement of the target sound of a clarinet improvising and to identify the shape of the spatial trajectory. In half of the trials, the sound of the clarinet was accompanied by the sounds of percussion and cello, which were stationary and located laterally at the two sides of the hall. The clarinet, percussion and cello sequences were 71, 58 and 58 dBA, respectively, at the center of the hall. A response sheet as in Figure 4 was used to gather participants' responses. On left hand side is the trial number and on top the possible answers. Participants ticked the trajectory they perceived. Before each trial, the trial number was announced using a synthetic voice so that participants could keep track of the flow of the

experiment.

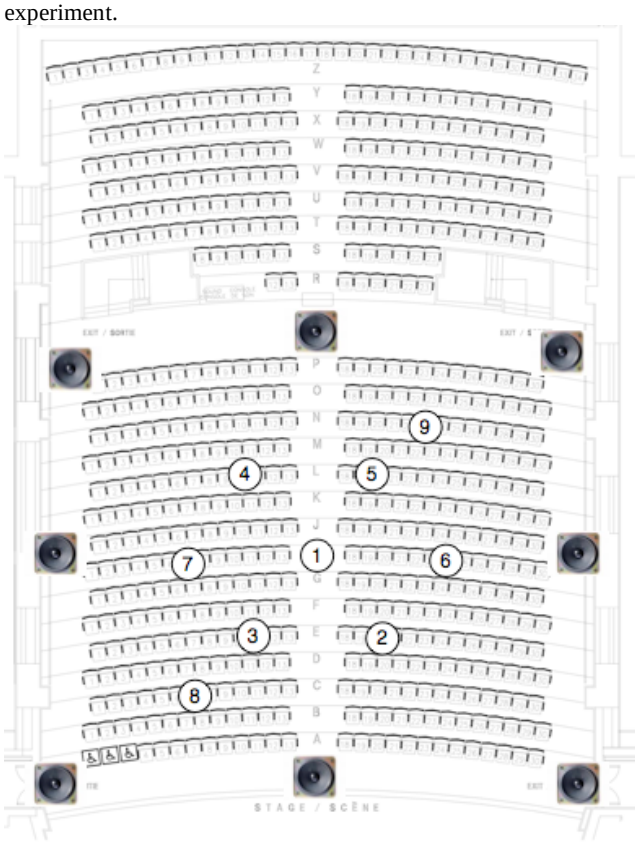

Figure 3. The seating and loudspeaker arrangement in the hall. The stage is at the bottom.

Four computers and a Polhemus Liberty magnetic position tracker connected to a LAN were used to run the experiment. The first computer controlled the experiment and provided the trial specification to the rest. The second displayed the graphical user interface that was used by the performer and generated the trajectory data. The third received the data from the trajectory control computer and rendered the spatialized sound scene in the concert hall. Finally, the Polhemus Liberty was connected to the fourth computer, which received data from the tracker and formatted it into OSC (OpenSoundControl) messages which were sent over the network to the other machines. The Polhemus tracker allowed us to track the location of the performer's hand in space with a resolution of 0.0004cm, an accuracy of 0.07cm and at a sampling rate of 240Hz.

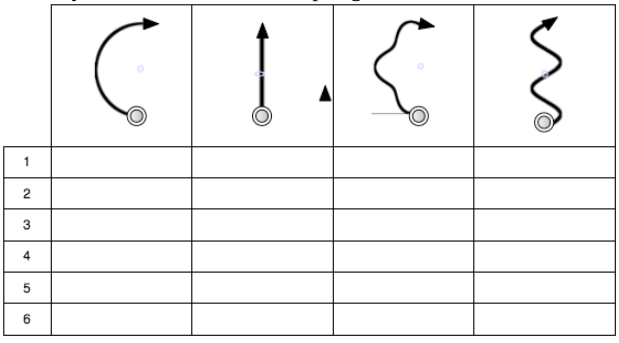

Figure 4. Response sheet used by participants.

The graphical user interface provided visual feedback indicating the desired speed, the shape of the trajectory and the current position within the trajectory. A gray circle indicated the sound position and a pink line moved with the desired speed along the trajectory. In the first group (Display Condition Auto), the gray circle was automatically aligned with the pink trace and a timer controlled its progression. There was no performer on stage. In the other two conditions, the performer controlled the position of the gray circle by way of the Polhemus tracker. His task was to track the pink trace. At the beginning of each trial, the gray circle was placed at the beginning of the trajectory and the performer had to capture it, by means of a red cursor, for 1 sec before the trajectory tracing began. Red and gray circle then moved in synchrony. While performing for the second group (Display Condition Open), the performer listened to the sound in the hall. In this sense, the quality of audio feedback was poor because the performer was outside the loudspeaker array. While performing for the third group (Display Condition Binaural), the performer listened to a binaural rendering of the sound scene as this would be experienced by a person sitting in the centre of the audience area. Data from the Polhemus tracker were normalized between -1 and 1, where these limits reflected the area that could be comfortably reached by the performer. They were scaled by a factor of 400 before being rendered into the GUI and by a factor of 3 before being rendered by the spatialization engine. SPAT assumes that the area inside the speakers is -1 to 1, values larger than these correspond to sounds further away from the loudspeaker array. The trajectories used represent virtual spatial sound movement both within and outside of the loudspeaker array.

The performer performed the gestures in the horizontal plane, so that they directly mapped to the plane spatial audio was presented using the loudspeaker array. The screen was placed diagonally between the vertical and the horizontal plane and placed lower than the arm of the performer so that his actions would still remain visible from the audience.

2.3. Procedure

Three groups of nine participants were tested in three separate experimental sessions. The first listened to the spatial audio trajectories that were automatically rendered by the computer and received no visual feedback. The second and third listened to the trajectories that were performed live by the performer and therefore received visual feedback.

In these cases, the performer was on stage using a Polhemus tracker to perform physical gestures according to the trajectory being shown in the graphical user interface. Participants were briefed and given the response sheet where the aforementioned trajectories were drawn. They were instructed to check which one they thought was performed. After each set of 24 trials, participants moved to the next seat until they had performed the experiment in all seats. Each experimental session lasted an hour with nine participants tested simultaneously in each of the nine seats in trial blocks of approximately six minutes. Prior to starting the main part of the experiment, all of the sound stimuli were presented and four training trials were performed to familiarize the participants with the task.

Figure 5. The visual interface indicating the desired trajectory (black), the current position (red circle), the part of the trajectory already completed (pink).

3. RESULTS

3.1. Identification of Sound Trajectories by the Audience

Percent correct identification was estimated from the responses of the participants and outliers were excluded based on whether they differed significantly from the median response per condition. This removed 2% of the observations. The overall identification of sound trajectories by the audience is presented in Figure 6.

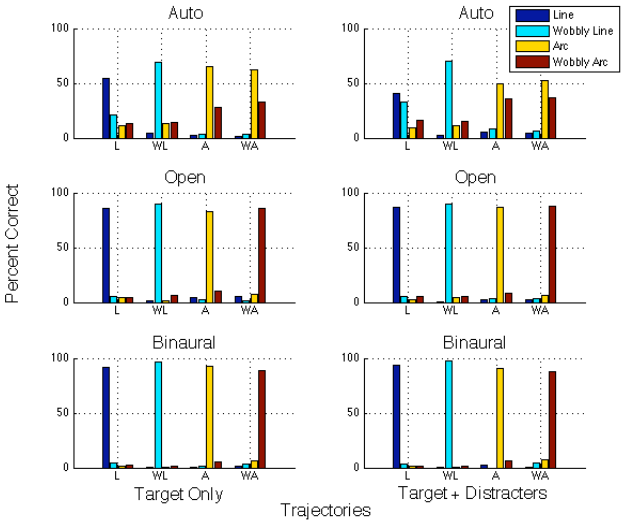

Figure 6. Percent correct averaged across listener seats. Left column for target only, Right column for target plus distracters.

Variation with respect to seat was observed only for the case where no visual feedback was given to the listeners. Figure 7 presents pooled identification performance for all seats for the automatic display. From a descriptive analysis point of view, it is evident that the presence of the performer improved the identification of sound trajectories. In the absence of the performer, the easiest trajectory to identify was the wobbly line and the worst was the wobbly arc, which was confused most often with the arc. There was a small improvement in identification performance for the third group when the performer received binaural feedback, relative to the case where he did not. The position of the listener did not affect identification performance much apart from the case of the 'line' sound trajectory. The presence of distracter sounds only degraded performance when the performer was absent.

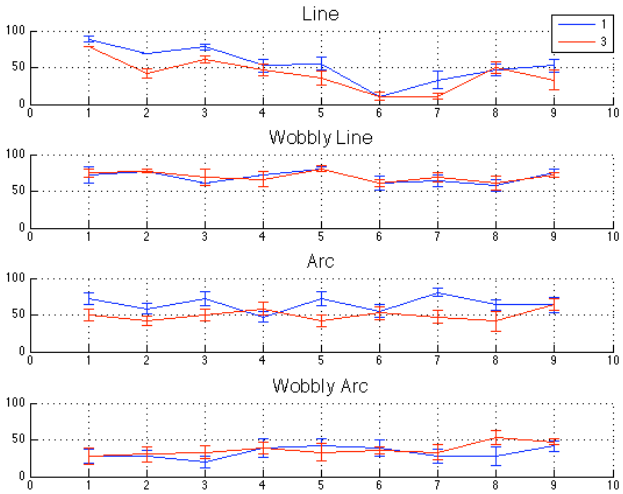

Figure 7. Variation in identification performance in the automatic display as a function of listening location, trajectory and number of sounds.

A statistical analysis (Display(3) x Location(9) x Trajectories(4) x Soundscape(2) analysis of variance) verified the observations. The was a significant main effect of Display ($F(2, 1152) = 416.611$, $p<0.001$) and trajectory ($F(3, 24) = 29.16$, $p<0.001$) on identification performance. There was a marginal effect of soundscape, ($F(1,8) = 3.8515$, $p = 0.08$)). No effect of listening seat was found. Post-hoc (TukeyHSD) pair-wise comparisons showed that identification performance was significantly different among all three interfaces. Their ranks are binaural, open, then automatic. Post-hoc tests (Tukey-Cramer HSD) for the four trajectories, showed identification performance was significantly different between all pairs except the line and arc pair. They are ranked wobbly line, arc, line, then wobbly arc. Finally, there was significant interaction between listening seat and trajectory, $F(24,192) = 2.1494$, $p<0.01$, trajectory and soundscape, $F(3, 24) = 8.1785$, $p<0.001$), trajectory and display, $F(6,1152) = 12.4829$, $p<0.001$, trajectory, soundscape and display, $F(6, 1152) = 2.1255$, $p<0.05$.

Given the significant difference between the displays, a (Listening Seat x Trajectory x Soundscape) within-subjects analysis of variance was performed for the automatic display, which showed significant main effects of listening seat, $F(8,64) = 6.0485$, $p<0.001$, soundscape $F(1,8) = 8.1554$, $p<0.5$, trajectory $F(3,24) = 72.195$, $p<0.001$ as well as all interactions between the independent variables. The interesting point here is that the absence of the performer results in listening position and soundscape effects, which are absent from the open and binaural display conditions. Other than that, post-hoc pair-wise comparisons (Tukey-HSD) showed listening seats 6 and 7 to differ significantly from all the rest (position 7 did not differ from 4) but not between them.

3.1.1. The influence of the performer

An interesting point to examine is why identification performance for the third Binaural Display group was better than that for the Open Display. To answer this question we

examined the RMS error of the performed trajectories compared to the automatic trajectories.

To achieve this, the performed trajectories were aligned in time to minimize the effect of performer lead or lag and then the RMS tracking error in pixels was calculated. Trajectories were aligned based on the lag that would maximize the cross correlation between the performer trajectory and the trajectory played by the computer. Subsequently, RMS tracking error was calculated based on these lags [Figure 8].

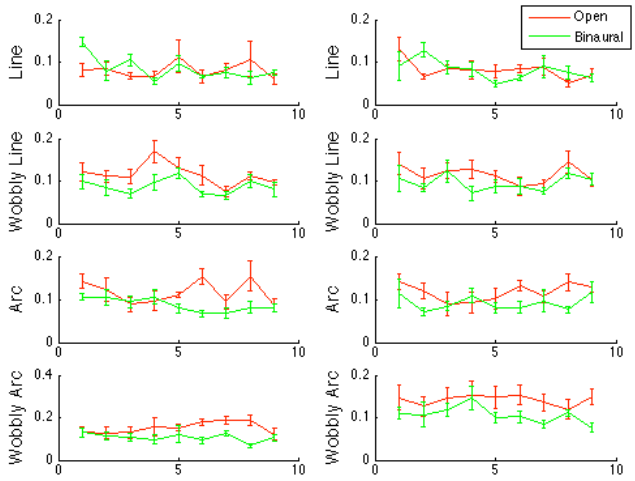

Figure 8. RMS tracking error for the conditions performed in the experiment based on the normalized data. The RMS tracking error in pixels is 400 times the indicated values and RMS error in 3D audio space in meters is 3 times this error.

Because data from only one performer were used, the lag one serial correlation was estimated for each of the repetitions within each session, which was found to be less than 0.1. This enables a factorial analysis based on an ANOVA design. A Display(2) x Trajectory(4) x Soundscape (2) x Session (9) ANOVA was performed on RMS errors. There was a significant main effect of Display, $F(1, 3) = 29.108$, $p < 0.05$, Trajectory, $F(3, 9) = 62.25$, $p<0.001$, and Session $F(8, 24) = 2.8328$, $p < 0.05$. There was significant interaction between trajectory and display, $F(3, 9) = 6.7564$, $p < 0.01$.

	Arc	Wobbly Arc	Line	Wobbly Line
Open	0.12	0.14	0.08	0.12
Binaural	0.09	0.1	0.08	0.09

Table 1. Mean RMS tracking error averaged across sessions

Tracking performance was more accurate in the third experimental session in which binaural feedback was provided to the performer. It is interesting to see that the improvement is bigger as the complexity of the trajectory is increased. The improvement is more for the Arc and Wobbly Arc and Wobbly Line as opposed to the Line trajectory, which was the most simple to perform.

RMS tracking error as a function of session averaged over trajectories is presented in Figure 9. This effect seems to be primarily due to the first session, because when this session is removed, the effect is no longer found to be significant.

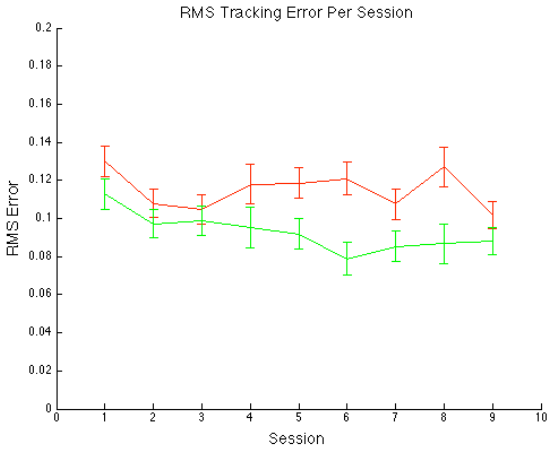

Figure 9. RMS tracking error per session averaged over trajectories.

4. DISCUSSION

The major finding of this study is that the synergy of the visual feedback provided by the performer with the spatial audio feedback results in an increase in identification performance for the sound trajectories. The effect was also beneficial in that it resulted in small identification variability as a function of the location of the participants in the concert hall and minimized the effect of a more complex soundscape.

This finding can be translated into a design guideline that states that when there is an intention to communicate spatial semantics of sound through gestures it is beneficial to align their shape to the spatial content that is to be communicated. The resulting synergy of visual and spatial audio cues, overcomes the problems associated with spatial audio perception in concert halls.

In terms of identification scores the trajectories are ranked as Wobbly Line, Arc, Line, and Wobbly Arc. The difference between Arc and Line is not significant. The differences in identification performance were less pronounced when the performer was visible, but the ranking was the same. We hypothesize that the Wobbly Line was the easiest to identify because it had the highest spatial variation. In particular we believe that the left-right swing within the concert hall, was a cue that was not present in the other trajectories and could have thus be the reason behind the high identification rates observed. The wobbly arc trajectory was the worst in terms of identification performance. The distance cues were not strong enough to sufficiently differentiate it from the arc case. The arc trajectory was identified relatively consistently across seats. However, it was confused with the wobbly arc trajectory quite often, especially when the soundscape was complex. This shows that the magnitude of spatial manipulations related to distance has to be carefully calibrated in order to be perceived by the audience. It is obvious that constancy vs. small variations in radial distance is not an easy cue to follow. Had it been that the ripple in the line trajectory was of similar magnitude as the one in the arc, it is doubtful whether the listeners would be able to distinguish line from wobbly line as well as they did in this study.

The line trajectory was well identified apart from listening locations that were too close to a loudspeaker as in 6 and 7. In addition, it is worth noting that the use of an identification task eased the process for the participants, since they could proceed

through the process of elimination using cues from a lengthy trajectory. In the more realistic situation of a concert with unknown sound trajectories and more complex soundscapes, identification for automatic presentation would probably be less accurate.

With respect to the effect of listening seat, most information can be obtained by the identification scores of the automatic display. It can be observed that identification performance was quite uniform with the exception of seats 6 and 7. The reason for this is that these seats were much closer to one of the loudspeakers in the array than the rest. As a result, the difference in the time of arrival of the sound events from the rest of the loudspeakers was not short enough to allow for summing localization to work. Therefore, the sound direction was biased towards the loudspeaker closest to the listening seat. Such a seating arrangement is not uncommon in electro-acoustic music concerts. Space constraints often result in listeners being close enough to loudspeakers for the localization of sounds to be distorted.

The three-way interaction between trajectory, soundscape and display implies that the extent of the improvement due to the visibility of the performer's gestures depends on the interaction between the simplicity of the soundscape and the ease of identification of the trajectory. In other words, the ease of identification of the trajectory and the extent that this varies as the soundscape becomes more complex, will result in varying levels of identification improvement in the presence of visual cues. For example, the improvement observed for the wobbly arc was much higher compared to the case of the wobbly line trajectory.

The interaction between listening seat and trajectory implies that the identification of a certain trajectory depends on the quality of the spatial cues at a certain position in the concert hall. For example, the arc trajectory was easy to identify in position 6, even if the localization cues were distorted due to the proximity to the loudspeaker, because the trajectory was confined in one side of the hall. This however, was not the case for the line trajectory, where the fidelity of the reproduction of the localization cues was more important.

The improvement in identification performance for the Binaural session indicates an effect of the performer's accuracy. Such a result suggests that performers need to be well trained in spatial manipulation. It is unfortunately impossible to conclude here why tracking was more accurate in the binaural display. The effect of the more reliable cues provided by the binaural rendering of the sound scene might be confounded with a practice effect. In addition, we only have data from one performer. It is therefore hard to make generalizations concerning the effect of binaural audio feedback on visual tracking. The separation between the two curves in Figure 9 points to the fact that an effect might exist. A follow-up experiment would be necessary, however, in order to understand the effects of spatial audio feedback on a visual tracking task.

The results of this study shape the specifications of a software system that provides visual and binaural feedback for the performance of 3D gesture control of spatialization. The trajectories could be loaded from the score and displayed to the performer who would then gesticulate and bring them into life in the hall. Such an implementation would increase the identification of the spatial audio manipulations by means of the visual feedback provided by the performer.

5. CONCLUSION

We presented an experiment on the effect of visibility of the performer's gestures on spatial sound trajectory identification performance. When the shape of the performer's gestures was aligned with that of the spatial audio trajectory a significant improvement in identification accuracy was observed. In the absence of this feedback, identification was generally poor even for simple trajectories and degraded depending on the listening seat.

6. REFERENCES

[1] Marshall, M., J. Malloch, and M. Wanderley. *Non-Conscious Control of Sound Spatialization*. In *International Conference on Enactive Interfaces*. 2007. Grenoble, France, p: 377-380.

[2] Davidson, J., *Visual perception of performance manner in the movements of solo musicians.* Psychology of Music 1993. **21**: p. 103-113.

[3] Vines, B., C. Krumhansl, M. Wanderely, and J. Levitin, *Cross-Modal interactions in the perception of musical performance.* Cognition 2006. **101**: p. 80-113.

[4] Kurosawa, K. and J. Davidson, *Nonverbal behaviours in popular music performance: A case study of The Corrs.* Musicae Scientiae 2005. **9**(1): p. 111-136.

[5] Blauert, J., *Spatial Hearing: Psychophysics of Human Sound Localization*. 1997.

[6] Chandler, D. and W. Grantham, *Minimum audible movement angle in the horizontal plane as a function of stimulus frequency and bandwidth, source azimuth and velocity.* The Journal of the Acoustical Society of America, 1992. **91**(3): p. 1624-1636.

[7] Saberi, K., L. Dostal, T. Sadralodabai, and D. Perrott, *Minimum Audible Angles for Horizontal, Vertical and Oblique Orientations: Lateral and Dorsal Planes.* Acustica 1991. **75**: p. 57-61.

[8] Bronkorst, A. and T. Houtgast, *Auditory Distance Perception in Rooms.* Letters to Nature 1999. **397**: p. 517-520.

[9] Hartmann, W., *Localization of Sound in Rooms.* The Journal of the Acoustical Society of America 1983. **74**(5): p. 1380-1391.

[10] Pulkki, V.P.L., *Spatial sound generation and perception using amplitude panning techniques*. 2001: Espoo, Finland.

[11] Malham, D. *Experience with large area 3-D Ambisonic sound systems*. In *Institute of Acoustics Automn Conference on Reproduced Sound*. 1992. Windermere.

[12] Malham, D., *Higher order ambisonics systems for the spatialization of sound*. In *International Computer Music Conference*. 1999. p. 484-487.

[13] Berkhout, A., D. de Vries, and P. Vogel, *Acoustic control by wavefield synthesis*. *The Journal of the Acoustical Society of America* 1993. p. 2764-2778.

[14] Braasch, J., *A loudspeaker-based 3D sound projection using Virtual Microphone Control (ViMiC).* In 118th *Convention of the Audio Engineering Society, Preprint 430*. 2005: Barcelona, Spain.

[15] Gröhn, M. *Localization of a moving sound source in a virtual room, the effect of a distracting auditory stimulus.* in *International Conference on Auditory Display*. 2002.

[16] Pulkki, V. and T. Hirvonen, *Localization of Virtual Sources in Multichannel Audio Reproduction.* IEEE Transactions on Speech and Audio Processing, 2005. **13**(1): p. 105-119.

[17] Guastavino, C., V. Larcher, G. Catusseau, and P. Boussard. *Spatial Audio Quality Evaluation: Comparing Transaural, Ambisonics and Stereo*. in *International Conference on Auditory Display*. 2006.

[18] Bates, E., G. Kearney, F. Boland, and D. Furlong. *Localization Accuracy of Advanced Spatialization Techniques in Small Concert Halls*. in *153rd meeting of the Acoustical Society Of America*. 2007.