MULTIMODAL PERCEPTION IN CONCERT HALLS;
WHERE DO WE LOOK WHEN WE LISTEN?

A Minors Anne Minors Performance Consultants, London, UK

1 INTRODUCTION

This paper outlines a series of experiments undertaken with eye tracking equipment at Warwick University DigiLab in 2012 as part of an MA in Theatre Consulting degree. The main experiment and results are being published by Psychomusicology: Music, Mind and Brain shortly. This paper concentrates on the quest to learn more about the multimodal listening experience within concert halls, the development of a methodology to test it, and initial analysis.

Concert halls were chosen for the experiment because, as performance spaces, they are complete four-sided architectural spaces, unlike theatre and opera house forms which need to remain essentially unfinished to enable other creative artists to be able to take charge of the space, direct the audiences' viewing, and for the room to recede. In concert halls the acoustical success depends on following the laws of physics, and these have to be balanced with the architect's vision and client's desire for overall success.

When designing concert halls, the large majority of reference material is visual and images are used to judge and compare halls. The experience of being in performing arts buildings, and more importantly, the experience of hearing sound within them, is confined to a few experts and enthusiasts rather than being a common experience throughout the design team. This leads to decisions regarding form being made for visual rather than aural reasons.

This quest to find out what people look at while listening arose from being aware in concerts that, as well as looking at the performers and orchestra, some audience members close their eyes to concentrate on the sound and others focus on an object to reduce the visual 'noise' and increase the ability to hear. Multisensory perception has the potential to improve the audience and performer experience by enabling a deeper understanding of the perceptual forces at play.

2 RELATED WORK[1]

The historic relation between the acoustician (hearing sense) and the architect (visual sense) has been described as separate and often contradictory. (Forsyth 1992[2], Blesser and Salter 2007[3]).
Shoebox and surround halls have different acoustical characteristics, the principal physical difference being the intensity and density of lateral reflections, resulting in different psychophysical responses of intimacy and envelopment. This understanding and design practice has developed for half a century, with early work by Marshall (1967)[4], Keet (1968)[5], Barron (1971)[6], Schroeder et al (1974)[7], and application by Essert (1999)[8], among others. The dominance of the visual sense in modifying the actual aural information to the received aural information was demonstrated by McGurk (1976)[9]. In the late 20th century, what had been conversations between acousticians and architects were now assisted by computer software that demonstrated visually the relationships between the sound quality, the shape of the room, sound reflection patterns, and finishes (Essert 1997)[10]. While it is understood that the visual nature of the room can influence the acoustic impression, often the visual sense is excluded from acoustic experimentation as noted by Griesinger (1997)[11].

As music is a multisensory activity involving many parts of the brain and body, and it is important that it is studied as such (Hallam, Cross and Thaut, 2009)[12]. Links between music and biologically relevant, survival-related stimuli had been made. Blood and Zatorre (2001)[13] measured the areas in

the brain that were active when the participant listened to music which they liked and which induced chills or goosebumps for the participant. Relatively few studies exist of the sensory nature of architecture and of cross-modal interactions of acoustic spaces. The effect of lighting, colour and visual stimuli in the performing environment formed part of a study on intimacy in concert halls (Hyde 2002)[14].Concert halls also involve what the audience looks at while they are listening, as well as how they feel related to other members of the audience (Pallasmaa 2005)[15]. Two psychophysical studies (Valente and Braasch, 2010)[16] examined the impact of visual cues on musicians' spatial impression expectations of a performance space and the importance of congruent audiovisual presentation in the interpretation of an auditory scene (Valente, Braasch and Myrbeck 2012)[17] which concluded that in one instance, the predominance of auditory cues in the spatial analysis of the bimodal scene was key.

Chia-Jung Tsay (2012)[18] asked musicians to judge a musical performance from audio-only and visual-only clues. Neither group was able to reliably judge the winner from sound only or video and sound. However they became more reliable when judging from silent video recordings. Even when sound is consciously valued as the core domain content, the dominance of visual information is apparent. A more recent study using eye tracking equipment compared eye movements when listening to preferred music; unknown and neutral music; and no music; when viewing a picture. Schafer and Fachner (2015)[19] concluded that listening to music has a significant effect on eye movement with longer fixation times, fewer saccades and more blinks compared to silence. There was no notable difference between preferred music and neutral music. Valente and Braasch (2010)[16] found subjective impressions of spatial acoustic parameters were statistically different when the participant was presented with a uni-modal stimulus (auditory or visual) as opposed to a bi-modal stimulus (auditory and visual).

2.1 Development

An initial experiment was devised in 5 parts to compare people's responses to different concert halls:

1 Visual assessment for 30s of nine concert halls - six shoeboxes and three surround - all chosen for their good reputation, their sounds or architecture and sometimes all three. This test recorded peoples' conscious observations and impressions in words.

2 Using eye tracking equipment to view a single image of six different halls for 15s each - four shoebox (Vienna Musikverinssaal, Koerner Toronto, Birmingham Symphony Hall, Sage Gateshead) two surround halls – (Berlin Philharmonie and DR Concert hall Copenhagen). Eye tracking equipment measured the gaze points which are defined as positions in time along a saccade or eye shift. A fixation is defined as a period of more than 100 ms when the user is focusing attention in one locale.

3 A sound track was created of four different percussive sounds (xylophone, wooden box, rattle, frog croak) played simultaneously, with one sound disappearing and reappearing. Participants were asked to identify the direction from which one of the sounds disappeared and reappeared. The purpose of this change was to create some spatial sound within their visual field at a dynamic that would require the candidates to concentrate on the soundtrack, whilst looking at the picture. While the participant looked and listened, the eye tracking system logged the eye movement and this enabled a comparison to be made between what people looked at according to the eye tracking device and what a different group of people said they looked at in the commentary of the first part of the experiment.

4 The fourth part, was a control test to compare with the second part. Each candidate looked at the same single image of the six different halls for 15s each. The results were compared with the images from the second part to see if there was any pattern that could be

investigated further. Although the sample was small (only 5 easily re-assembled participants) the results were interesting enough to warrant further investigation. (see 2.2 below).

5 The fifth part of the experiment was to prepare saliency maps of the six concert hall images used and compare them with the consciously-made comments to the photographs in the first part of the experiment. Saliency maps integrate the normalized information from the individual feature maps into one global measure of conspicuity. Saliency at a given location is determined primarily by how different this location is from its surround in colour, orientation, motion, depth etc. The saliency map was designed as input to the control mechanism for covert selective attention.

2.2 Initial conclusions

Comparing the eye tracking results of the second part of the test with the people who undertook the third control test, it appeared that while listening to the sounds, their eyes were more concentrated in the area that they viewed compared to the area that they viewed when there was no sound.

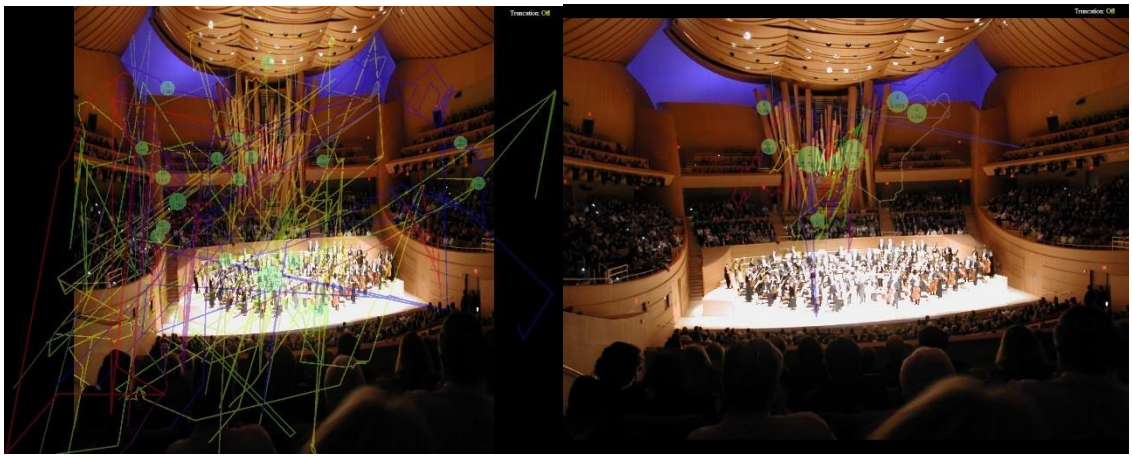

Figure 1. The same candidate looking (left) and looking and listening (right) for the same length of time at the same image of Disney Concert Hall.

The saliency images, compared with the comments that people had made, were congruent and spoke quite strongly to some people's subconscious, to the extent that they disliked some halls, as depicted by the images, because of the 'angry face' that the saliency map revealed.
Lighting was a key component in creating an impression of the hall. It had the ability to enhance or detract from the form and provide visual interest or emphasize unimportant details.

3 SECOND EXPERIMENT

Based on the observations from the first set of experiments a further experiment was devised to answer the following questions:
1. When the brain is occupied by listening, does the visual field narrow down? If so, is it in the horizontal direction only?
2. Does the visual field narrow in the same way with a surround room as a shoebox space?
3. Does some “decoration” i.e. visual interest, help the brain “listen” or cause a distraction?
4. Is there anything to be learnt from what participants look at?

3.1 Experimental Design

Four images of concert halls chosen for the experiment were taken from a similar place in relation to the stage in each concert hall. In the case of shoebox halls the view from the second balcony above the stage and some way back was chosen, and in the case of the surround halls, a similar downward view to the stage. Not all the photographs have an audience in place but all have full concert lighting on stage. The auditorium in Gateshead had a full orchestra and conductor in rehearsal; Birmingham Symphony Hall had a full orchestra conductor and extensive choir in rehearsal for Mahler's Eighth symphony; Copenhagen had a partially occupied stage during a rehearsal break with the music stands and chairs in place; Berlin had a full orchestra standing on stage at the end of the concert with audience.

3.2 Methodology

This experiment was in four parts.

A Sound Only

The first part was a control experiment to track the eyes looking at a blank, 50% grey screen (i.e. undirected viewing) while listening to 2 soundtracks 15s long, one with the characteristics of a surround concert hall, where the sound arrives at the ears from a frontal direction, and the second with a characteristic of a shoebox hall, the sound is more enveloping. Participants were asked to identify which of the two sounds appeared to be recorded 'closer' to the orchestra.

B Image Only

The second part was a control experiment to track the eyes looking at four images of different concert halls for 15 seconds each. The choice of halls was deliberate, to concentrate on scale of architectural finish and clear typology of form. All are contemporary looking halls so that there is no stylistic difference, even though the oldest hall is 50 years old. Not included were hybrid halls or less than pure shoebox or surround halls, nor any that the author has worked on. The two examples of shoebox halls were chosen for their difference from one another in their decorative treatment (visual density).

C Image and Sound

The third part of the experiment put together the same four images and a new sound track. Compared to part one, a different portion of music was used from the same recording, but the 15s sound track was manipulated to be a frontal sound (for surround halls) and an immersive sound (for shoebox halls). For the two shoebox halls, there were 2 different click patterns and for the two surround halls there were two more different click patterns, a total of 4 click patterns.

D Overall experience

The candidates were asked to rank the order of preference of the different experiences of each Hall. They were allowed to revisit the images in selecting their order of preference.

In the pilot experiment, the images were shown first in order without sound. To ensure their active listening, participants were asked to identify the direction as well as the number of clicks added to the soundtrack. The soundtrack with the appropriate acoustic was used for each Hall, i.e. the immersive sound was used for the shoebox halls and the frontal sound for the surround halls.

3.2.1 Pilot

In a pilot experiment it became apparent that there was insufficient time to write down the clicks and the different directions that they heard them in. Also a more random arrangement of the halls' images and soundtracks was introduced for the main experiment. Some early candidates counted clicks on their fingers and said the total at the end of each track. In the final experiment, it was easier for people to use a form of semaphore and point around their head in the direction that they heard the click, at the time they heard it, and for the researcher to record their actions.

3.2.2 Equipment

A laptop computer with eyetracker system faceLab 5 was linked to a second monitor with a 21" diagonal screen. Synchronisation was done by hand but did not lead to noticeable latency. The eye track co-ordinates are mapped to the pixels. Each participant was invited to sit in a comfortable position, placing their head within the combined image the infra-red camera lenses.

3.2.3 Participants and Habituation

27 participants were introduced to the equipment and asked to sign an interview release form. The form also asked participants to self assess that they had good eyesight, good hearing, if they have played a musical instrument or attended pop or classical concerts regularly. Each participant was asked to look at the screen, while the calibration equipment for the eye tracking device is aligned with the candidate's eyes. During the experiment, the researcher recorded their verbal replies and observations.

4 RESULTS

Of the original 27 participants, 20 produced usable results. The number of gaze points is very calibration dependant, while the fixations are less so. A plot of gaze points against the sound tracks was investigated for participant 1 and test 1 where the participant was looking at a 50% grey blank screen while listening to music, sound tracks S1 (frontal) and S2 (enveloping), but concluded that for this experiment it was better to concentrate on the fixation point analysis.

4.1 Method of Analysis

4.1.1 Raw data

Eye tracking data recorded the x and y co-ordinates relative to the bottom left corner of the screen for every gaze point, every 16ms, for each person; each of 4 halls; sounds 1 and 2 alone; sound and image; image alone.

4.1.2 Standard Deviation for 15s duration

To describe the extent of the eye movement, the standard deviation of the x coordinates and y coordinates over a 15s time period for each viewing was calculated. This gave the boundaries of a box that describes the extent of the eye movement. For each participant, the points were plotted on a graph in relation to a 45deg line representing a square box. For each participant, there was a summary chart of fixations by room and fixations by test.

4.1.3. Subject Summary and Room Summary Charts

The chart for each participant was put together in another file as a subject summary plot. A room summary worksheet was also prepared, with summary data for all participants for one test or one concert hall. This became a plot on a complete room summary graph showing the extent of spread among the participants by the error bars.

4.2 T test

To understand which of the large sample results were significant, and which not, a t-test was applied to the results. In this test, a value of $p < 0.05$ implies that the difference between groups on the dependent variable is statistically significant.

4.2.1 Image alone vs image w sound - all rooms

In order to establish whether the difference between subjects' eye motion for image and sound, and image alone, for all rooms, was statistically significant, a 2-sample T test was applied, with the two data series being (1) image and sound and (2) image only. The same analysis was done for the STD X and STD Y data, for the complete set of subjects.

For STD X (width of fixpoint box):

- complete set: t(df=137) = 4.09; p = .000037; statistically significant with greater than 99% confidence.

Mean of image alone results (170) is greater than mean of image and sound (128) (100:75)
For STD Y (height of fixpoint box):

- not statistically significant.

For the entire group of four halls, the visual width of the fixation focus area reduces by 25% when image and sound are presented simultaneously, compared to viewing the image alone. The visual height of the fixation focus area does not alter significantly when image and sound are presented simultaneously, compared to image only.

4.2.2. Sound alone vs image w sound - split by room type

NARROW HALLS (A,B) WITH ENVELOPING SOUND (S2):
For STD X (width of fixpoint box):

- complete set: t(df=12) = 2.05; p = 0.062; statistically significant with greater than 90% confidence. Mean of sound alone results (213) is greater than mean of image and sound (135) (100:63)

For STD Y (height of fixpoint box):

- not statistically significant.

In narrow halls, the horizontal component of the fixation focus area reduces by 27% when image and sound are presented together compared to the fixation focus area for sound alone.

WIDE HALLS (C,D) WITH FRONTAL SOUND (S1)
For STD X (width of fixpoint box):

- not statistically significant.

For STD Y (height of fixpoint box):

- complete set: t(df=23) = 1.81; p = .082; statistically significant with greater than 90% confidence. Mean of sound alone results (153) is greater than mean of image and sound (117) (100:76)

In wide halls, the vertical component of the fixation focus area reduces by 24% when image and sound are presented together compared to the fixation focus area for sound alone. So there is no great difference between hall types.

4.2.3. Sound alone vs image alone - split by room type

NARROW HALLS (Ai,Bi) VS ENVELOPING SOUND (S2)
For STD X (width of fixpoint box):

- not statistically significant.

For STD Y (height of fixpoint box):

- complete set: t(df=14) = 1.80; p = 0.093; statistically significant with greater than 90% confidence.

Mean of sound alone results (159) is greater than mean of image alone (108) (100:68)

In narrow halls, the vertical component of the fixation focus area reduces by 32% when only image is viewed compared to the fixation focus area for sound alone.

WIDE HALLS (Ci,Di) VS FRONTAL SOUND (S1) For STD X
(width of fixpoint box):
- not statistically significant.

For STD Y (height of fixpoint box):
- not statistically significant.

4.2.4. Image and sound vs image alone - Split by room type

NARROW HALLS (Ais,Bis WITH ENVELOPING S2 SOUND)
For STD X (width of fixpoint box):
- complete set: t(df=76) = 2.13; p = 0.036; statistically significant with greater than 95% confidence.

Mean of image and sound results (135) is less than mean of image alone (165) (81:100)
For STD Y (height of fixpoint box):
- complete set: t(df=94) = 1.67; p = 0.097; statistically significant with greater than 90% confidence.

Mean of image and sound results (131) is greater than mean of image alone (108) (100:82)

In narrow halls, the visual width of the fixation focus area reduces by 19% when image and sound are presented simultaneously, compared to viewing the image alone. The visual height of the fixation focus area is greater by 21% for image and sound than image alone.

WIDE HALLS (Cis,Dis WITH FRONTAL S1 SOUND)
For STD X (width of fixpoint box):
- complete set: t(df=57) = 3.19; p = 0.002; statistically significant with greater than 99% confidence.

Mean of image and sound results (120) is less than mean of image alone (168) (71:100)
For STD Y (height of fixpoint box):
- not statistically significant.

For wide halls, the visual width of the fixation focus area reduces by 29% when image and sound are presented simultaneously, compared to viewing the image alone.

4.3 Overlay of fixation boxes on room image

Of the 16 t tests, 8 had significant results. To demonstrate this information graphically and for each hall, the fixation focus area (which represents the standard deviation for all participants of that hall) for sound only, image only and image and sound, were overlaid on each concert hall image and on the saliency map of each image.

5 CONCLUSIONS

The initial conclusions drawn from the room summary graph are:
1. All ‘image only’ has much more extensive horizontal eye movement and less vertical eye movement than ‘image and sound’.
2 The visual field reduces in the horizontal direction for both surround and shoebox halls between ‘image only’ and ‘image and sound’. For sound alone, across all test types, the frontal characteristics of the sound of a surround hall engender a narrower width of visual field than the enveloping sound characteristics of a narrow hall.
3 People appeared to have their own preferences for where they tended to look- at the orchestra; above the stage, at the ceiling.

6 REFERENCES

1 Minors, A., Harvey, C., Influence of Active Listening on Eye Movements while Viewing Images of Concert Halls, Psychomusicology: Music, Mind and Brain. (2015).

2 Forsyth, M., (1992). Architect and acoustician: An historical overview. Proc.I.O.A.,14 (2), (1992).

3 Blesser, B., & Salter, L.-R., Spaces speak, are you listening? Experiencing aural architecture. Cambridge, MA: MIT Press. (2007).

4 Marshall, A.H., A note on the importance of room cross section in concert halls. Journal of Sound Vibration, 5, 100-115. (1967).

5 Keet, W. de V., The influence of early lateral reflections on spatial impression, Proc. 6th International Congress on Acoustics, Tokyo (1968).

6 Barron, M., The subjective effects of first reflections in concert halls- The need for lateral reflections, Journal of Sound and Vibration. 15, 475-494. (1971).

7 Schroeder, M., Gottlob D., & Siebrasse, K. Comparative study of European concert halls: Correlation of subjective preference with geometric and acoustic parameters, Journal of the Acoustical Society of America, 56, 1195. (1974).

8 Essert, R., Links between concert hall geometry, objective parameters, and sound quality. Paper presented at the joint meeting of the Acoustical Society of America/DAGA/Forum Acusticum, Berlin, March, 1999. (Abstract). Journal of the Acoustical Society of America, 105, 986. (1999).

9 McGurk, H. and MacDonald, J., Hearing lips and seeing voices. Nature, 264 (5588), 746-748. (1976).

10 Essert , R., Progress in concert hall design: Developing an awareness of spatial sound and how to control it. European Broadcasting Union Technical Review, 274, 31-39. (1997).

11 Griesinger, D., The psychoacoustics of apparent source width , spaciousness and envelopment in performance spacing, Acta Acustica united with Austica, 83, 721-731. (1997).

12 Hallam, S., Cross, I., and Thaut, M., The Oxford Handbook of Music Psychology. Oxford University Press. (2009).

13 Blood, A., Zatorre, R., Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. PNAS. (2001).

14 Hyde, J.R., Acoustic Intimacy in concert Halls: does visual input affect the aural exprerience? Proceedings of the Institute of Acoustics, 24, Pt 4. (2002).

15 Pallasmaa, J., The eyes of the skin: Architecture and the senses. Chichester, UK: John Wiley and Sons Ltd. (2005).

16 Valente, D. L., & Braasch, J., Subjective scaling of spatial room acoustic parameters influenced by visual environmental cues. Journal of the Acoustical Society of America. 128, 1952-1964. (2010).8

17 Valente, D.L. Braasch, J., & Myrbeck, S., Comparing perceived auditory width to the visual image of a performing ensemble in contrasting bi-modal environments. Journal of the Acoustical Society of America, 131, 205-217 (2012).

18 Tsay, C-J., Sight over sound in the judgement of music performance. Proceedings of the National Academy of Sciences, 110, 14850-14585. doi: 10.1073/pnas.1221454110. (2013).

19 Schäfer, T. and Fachner, J., Listening to music reduces eye movements. Attention, Perception & Psychophysics, 77, 551-559. (2015).