Influence of Active Listening on Eye Movements while Viewing Images of Concert Halls

Anne Minors and Carlo Harvey

University of Warwick

Author note

Anne Minors, Anne Minors Performance Consultants; Carlo Harvey, WMG Digital, University of Warwick

Anne Minors thanks Michael Holden and Dr Margaret Shewring, for setting up the Master's degree at Warwick University in Theatre Consulting; Prof. Alan Charmers whose lectures sparked thoughts of using DigiLab facilities for this experiment; and Bob Essert for preparing the binaural sound tracks and assisting with the data analysis.

Anne Minors read Leo Beranek's book, *Music, Acoustics and Architecture* at the age of 16, while playing cello in the Merseyside Youth Orchestra and being introduced to architecture. This inspired her to study and practice architecture, be intrigued by the power of music and music-making and ultimately become a theatre consultant. In 1996, she founded Anne Minors Performance Consultants Ltd, whose mission is to build effective rooms for rapport. Anne holds a B.A.(Hons) and Diploma in Architecture (Dist.) (1979) from Sheffield University and an M.A. in Theatre Consulting (Dist.) from Warwick University (2013).

Dr. Carlo Harvey is a research fellow on the collaborative EPSRC project, PSi, with Jaguar Land Rover. Carlo holds a B.Sc. (Hons) degree in Computer Science from the University of Bristol (2007) and a PhD in Engineering from the International Digital Laboratory, University of Warwick (2012). Coordinator of the visualisation seminar series at Warwick University, he publishes in the fields of auditory perception, olfaction, multi-sensory perceptual impact on vision and cultural heritage.

Corresponding author: Anne Minors, Anne Minors Performance Consultants, G2 Tay House, 23 Enterprise Way, London SW18 1EJ
, United Kingdom. E-mail: anne.minors@ampcstudio.com

Abstract

Concert hall design is at a crossroads between its origins of unamplified orchestral music and singing, and the forces of popular music, which depend mostly on amplified sound and multimedia accompaniment. Concurrently there has been a revolution in the way that the buildings are initiated and designed. Computer modelling techniques enable architects to conceive iconic building exteriors and acoustic engineers and performance consultants to shape building interiors to provide a rich sound experience. Despite the fact that a concert is a multisensory experience, little is known about how the visual aspects of the building interiors may impact the acoustical experience. This paper examines how the process of active listening influences where people look within a photograph of a concert hall interior. Using eye tracking equipment and images of concert halls while actively listening to sound tracks, the extent of the eye movement was recorded when looking only; listening only and looking and listening. This revealed significantly smaller standard deviations of eye movement width in the horizontal plane when looking at images of concert halls while listening to music as compared to looking at the images alone.

Keywords: multimodal perception, visual perception, aural perception, concert hall, theatre, opera, eye tracking, active listening.

Influence of Active Listening on Eye Movements while Viewing Images of Concert Halls

Concert halls are complete 'four-sided' architectural spaces. In concert halls the interplay between the musicians' playing and the effect of the hall on their sound is primary. Concert hall forms are expensive to build, and while their acoustical success depends on following the natural laws of physics, this has to be balanced with the clients' aspiration for iconic architecture to attract audiences and the architects' desire for 'being cutting edge', that is, different. Added to the design of concert halls is the increasing requirement for multimedia technology to assist some performances and be absent at others. The visual impact that the technology has on the space can be significant. Another less tangible factor, how people relate to the performer and each other in the space and how that unity feels, is also a key to a hall's success.

The architect, acoustician, and performance consultants form a triumvirate of disciplines, and together design the hall. The challenge for design teams now is to produce relevant and appropriate buildings for a fast-evolving artistic and audience landscape, and accelerating technological advances. Subtle design decisions on the shape of these buildings can have far-reaching consequences for the artistic, acoustic and operational use of the building.

While so many concert hall spaces can be viewed and visually judged from photographs in magazines, what those halls sound and feel like, and for whom, can best be experienced by being in the space and listening to live music. In the way that buildings are designed now, great time and energy goes into form-making, leaving less time for details and finishes. By revealing what people look at in concert halls, it may help architects to understand on what to focus their efforts for maximum effect.

There is an assumption that members of the audience look at the musicians on stage for their focus, however this may not be the case all of the time. Regular concert goers, particularly musicians, often comment that they close their eyes to concentrate on the music. The first author's observations of her own behavior in concerts before this experiment was that to really concentrate on the sound, it

was often necessary to look away from the brightly lit stage and focus on a static object to facilitate more attentive listening.

A longer quest beyond the scope of this paper has been investigating the following questions: Are successful concert halls successful because they get the balance right between the aural and the visual? Is there any consensus as to a successful space as an overall experience? Is there anything to do with rhythm or scale of detail that the eye picks up on and which is significant to our reading of, and relationship to, the space? What can we learn from the eye tracking information that may make us able to concentrate our design efforts on parts of the concert hall that matter? Does the eye tracking reveal anything about what people notice in a performance space?

> Can the visual help people to be 'present' to the musical performance, as described by conductor Susanna Mälkki (Dammonn, 2015)
>
> …it's the listening that's the thing. That's where the beauty is coming from. It's actually quite a precious moment when lots of people gather together in a big room to listen to something. Isn't it worth defending the practice of being fully present, just for a short time, in the music? That's how it is for the performers. Being present: in every single note, and every decision and feeling behind it, there's a whole universe in there.

How do we design attractive concert halls that promote, support and do not detract from good listening? Up to now, concert hall design discussion has been focussed on producing good acoustics and what is needed from the architecture to achieve this. Even halls built with high budgets and great care can fail to excite or sound exceptional. Clearly to be present in the universe of music communication involves the interaction between modes of perception. We can, by studying multi-modal perception and analysing existing concert halls, give some direction for improving the audience and performer experience and enable a deeper understanding of the perceptual forces at play between the aural and the visual. This paper seeks to begin the examination of visual attention while listening in concert halls.

Related Work

The historic relation between the acoustician representing the hearing sense and the architect representing the visual sense has been described as separate and often contradictory (Blesser & Salter 2007; Forsyth 1992). In the late twentieth century, the importance of room proportion and geometry on acoustical quality was established. Shoebox (typically narrow) and surround halls (typically wide) 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. The understanding of this and design practice have developed for half a century, with early work by Barron (1971), Keet (1968), Marshall (1967), and Schroeder, Gottlob, and Siebrasse, (1974), and application by Essert (1999), among others.

In an audio visual environment, although speech is generally considered as a purely auditory process, the visual influence on auditory perception cannot be neglected. McGurk and MacDonald (1976) reported that pronunciation of *ba* is perceived as *da* when accompanied by the lip movement of *ga* (the McGurk effect). 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). 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).

As music is such a multisensory activity, involving many parts of the brain and body, it is important that it is studied as such (Hallam, Cross & Thaut, 2009). There are relatively few studies of the sensory nature of architecture and of cross-modal interactions of acoustic spaces. There has been a realisation that concert halls are also about what the audience member looks at while listening as well as the extent to which the audience member feels related to other members of the audience (Pallasmaa, 2005). A psychophysical study of Valente and Braasch (2010) showed that musicians' spatial expectations of a performance space were affected by visual cues such as changing the sound source position in a multi-purpose space as well as the makeup of the sound stimuli. A subsequent

study of the importance of congruent audio visual presentation in the interpretation of an auditory scene concluded that in one instance, the predominance of auditory cues in the spatial analysis of the bimodal scene was key (Valente, Braasch, & Myrbeck 2011).

In a study by Tsay (2012), professional and amateur musicians were invited to judge the winner of a live musical performance competition 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.

Using eye tracking equipment to compare eye movements when viewing a picture or video while listening to preferred, neutral / unknown, or no music, Schafer and Fachner (2015) reported that listening to music led to reduced eye movements compared to silence. There were no differences resulting from the preferred versus the neutral or less familiar music. Valente and Braasch (2011) 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). The present study investigates the extent of eye movement in the context of attentive listening under bi-modal (visual images of concert halls and orchestral music) and uni-modal (only concert hall images or only orchestral music) conditions.

Method

An experiment was devised to measure eye tracking across still images of concert halls in the context of active listening to music that could plausibly be associated with each hall. The soundtracks were designed to engender active listening while the eye tracking equipment measured what was happening in the visual field in order to compare eye activity when the brain is occupied with looking and listening simultaneously (bi-modal condition) with uni-modal looking or uni-modal listening. While the soundtracks for the different concert hall images were given the gross spatial characteristics of those concert hall typologies, this experiment was specifically not trying to simulate or auralise the sound of the different concert halls in the photographs.

Clicks were introduced into an orchestral soundtrack to engender active listening rather than measure any correlation between the images of the concert halls and the sounds that could be heard from that position. The choice of music was not intended to sway the listener emotionally - it would not give rise to goose bumps or chills.

In addition, two control conditions were needed to capture eye tracking with a sound track only and with visual images only. In comparing the results from these three conditions, some answers might be found for the following research questions:

1. Does the visual field, (the area of eye movements in the horizontal and vertical directions) differ in the presence versus absence of music?
2. Does the visual field change in the same way with a surround room as a shoebox space?
3. Is there anything to be learnt from what participants look at?

Eye tracking equipment exploited software that records not only the movement of the eyes (saccades) and their order, but the length of time the eyes fixate on one point (gaze fixation) and the order of fixations. The experiment was conducted in three parts.

The first part was a control condition to track the eyes looking at a blank, 50% grey screen (i.e., undirected viewing) while listening to two soundtracks, 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, where the sound arrives from the sides and is more enveloping. In order to invoke some attentive listening, and demonstrate their ability to perceive the difference between the sound tracks, participants were asked to identify which of the two sounds appeared to be recorded 'closer' to the orchestra.

The second part was a control condition to track the eyes looking at four images of different concert halls for 15 seconds each in silence. The choice of halls was deliberate, to concentrate on the scale of architectural finish and clear typology of form. All are contemporary looking halls so that there is not a great stylistic difference, even though the oldest hall is 50 years old. Hybrid halls were excluded for clarity. Two examples each of shoebox and surround halls were chosen.

A- Hall One at the Sage Gateshead is a shoebox form and has an all-over architectural treatment in wood which is consistent over all the wall surfaces. The ceiling panels are also repetitive but have some differences of scale within each panel. Gateshead represents a simple visual experience within the form and typology of a shoebox. The space evokes a tall, uni-rhythmic visual and material experience.

B- Birmingham Symphony Hall is also a shoebox and has a more varied visual experience, using colour, scale of feature and different materials to add richness to the basic shoebox form. The space evokes a tall, multi-rhythmic visual and material experience. Both Gateshead and Birmingham have excellent acoustic reputations.

C- Of the two surround Halls, Copenhagen's Danish Radio Symphony Hall represents the overall surface treatment of the enclosing walls. The Copenhagen hall has an all-over decoration in wood which becomes more curvaceous and free form higher in the room. Its materiality is the same throughout the space, although some areas have perforations in the wooden panels. The backlighting from below creates another scale of visual detail. The space evokes a wide, uni-rhythmic visual and material experience.

D- Berlin Philharmonie is a visually dynamic space, with several levels of detail, particularly in the treatment over the stage where there are small reflectors, suspended light fittings and cut outs in the ceiling. This Hall represents the surround form equivalent of Birmingham, in terms of its visual richness and difference in scale. It has the added visual dimensions of light-coloured angled surfaces, which tend to be visually arresting. The space evokes a wide, multi-rhythmic visual and material experience.

Acoustically, Copenhagen and Berlin behave in a similar way, being both wide surround rooms, where the sound is received mainly from the front. The acoustic reputation of Berlin is very high and the orchestra has a very strong following, such that the hall does not need to put on any other type of concert.

The third part of the experiment combined 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 sound track was manipulated to be a frontal sound S1 (for surround halls) and an immersive sound S2 (for shoebox halls). To ensure their active listening too, 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 S2 was used for the shoebox halls and the frontal sound S1 for the surround halls).

The four images of concert halls chosen for the experiment, shown in Figure 1, 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. In three of the halls the ceiling plane was visible; in the case of Copenhagen the ceiling was absent. Not all the photographs have an audience in place but all have full concert lighting on stage. The stage occupies a similar area of the photograph in each case. 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; 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.

Binaural Sound Files - Choice and Construction of the Sounds

Two music excerpts were chosen from a single live orchestral concert recording made in the rear stalls of a 6000 seat hall with late envelopment and a large sense of space. The recording was made approximately 28m from the orchestra and the hall has a mid-frequency reverberation time of 2.5s compared to the concert halls in the images whose mid frequency reverberation times varied between 1.85s and 2.1s. The music chosen was Charles Ives' 'Variations on America'. The music was chosen as a well known song whose theme is short, about 15 s long, and encompassed a complete statement of the theme to give some sense of musical completeness.

A simplified binaural recording technique was used to give the listener a sense of presence in the space. Two high quality omni microphones were mounted on eyeglass frames approximately 1cm forward of the pinna of the person recording. The original recording was made at a comparable distance from the stage as the photographs were taken. This position was close enough to the orchestra to imply a broad soundstage, yet far enough away to hear the room, giving a good basis to create contrast between frontal and enveloping acoustics. The original auditory source covered nearly the front hemisphere of the apparent sound field as heard on headphones.

Music 1 was used in the sound alone experiment (experiment 1), and Music 2 was used in the music-with-image experiment (experiment 3). As they are two different parts of the same performance, they are acoustically similar, but musically not identical, so that having heard the music in the first phase of the experiment, the participants would still need to pay attention to the music in the second phase.

Manipulation of Music Excerpts to generate "enveloping" and "frontal" Sound Field Samples

In a narrow concert hall, the lateral component of the sound energy reflected back and forth between side walls is stronger, and the perceptual effect is that the sound field is more enveloping. In a wide hall, the lateral component of the sound energy reflected back and forth between side walls is weaker, and the perceptual effect is that the sound field is more frontal (cf., Barron, 1971; Essert, 1999; Keet, 1968; Marshall, 1967; Schroeder et al., 1974).

Excerpts 1 and 2 from the original binaural recording were used as the 'enveloping sound field' matched to the narrow halls. This sound pattern was highly lateral and enveloping in its sonic character. This enveloping sound field is characteristic of narrow, tall 'shoebox' halls. The same excerpts 1 and 2 were amplitude panned with digital audio editing software to narrow the auditory width of the sound to a more 'frontal', less lateral and enveloping sound field, in line with the gross acoustical character of the wide halls. The degree of binaural similarity can be characterised by the broadband RMS cross-correlation between left and right channels, which for the frontal sound was 2.4

dB greater than that for the enveloping sound. There was no change in total loudness or reverberation time.

Use of a single performance at the root of both frontal and enveloping samples allowed the listener to be aware of the primary difference between narrow and wide halls – the degree of spatial envelopment. This difference was included only for general congruence with the images. No attempt was made to manipulate frequency balance, sectional balance, loudness, clarity or reverberance, even though they are important and unique characteristics of individual halls and orchestras. The similarity of the musical content across all rooms was deemed more important than the precise acoustical differences between tracks. This combination of binaurally recorded sound and spatial compression reproduced a realistic sense of spatial relationships between musicians in the orchestra and overall width of sound, but did not include sound reflection paths specific to each hall. It would have been preferable to record the same orchestra and conductor performing the same piece of music in the four different halls, but that was not available for this experiment, and would be unlikely to happen as these concert halls are not on the same concert touring circuits.

Addition of Clicks to the Audio Track

Finally, for the music-with image condition, a series of equal-amplitude clicks was mixed with the Music 2 excerpts, panned to various locations in the spatial sound field in order to give the participants a specific listening task – a 'non-musical' task that would not rely on familiarity with the particular musical excerpt, genre of music or music listening expertise.

The clicks were spread from position 1= far left to 7 = far right, with position 4 in the centre, mixed in with the 15 second music excerpt. Between three and five clicks occurred in each sample, the locations randomly placed (but the same for each participant).The participants were asked to identify the order and location of the clicks. There were two sound files per visual image, the same sound files being repeated for the two shoebox halls and the two surround halls. For the two shoebox halls, there were two different click patterns and for the two surround halls there were two more different click patterns, a total of four click patterns.

Apparatus

A laptop computer with FaceLab 5 eyetracker software was linked to a second monitor with a 21" diagonal screen with a pixel count of 1280 x 1024 and was synchronised to the controlling PC. The eye track co-ordinates are mapped to the pixels so that the lower left corner is (0,0) and the upper right corner is (1280, 1024).

Two infra-red cameras at table height track one eye each. Each participant was invited to sit in a comfortable position, placing their head within the combined image formed by the infra-red camera lenses, as shown in Figure 2. The software enables switching between pupil and iris calibration. This is necessary because although pupil calibration is much more accurate, a pupil cannot be picked out by an IR filtered image from the camera when the participant has a dark colored iris. In these situations, an iris calibration was used. Sound was played through high quality stereo headphones, set at a comfortable loudness level and maintained constant for all listeners and trials.

Participants

There were 28 participants who were asked to confirm that they had good eyesight and good hearing. Eleven (5 female) of the participants were under the age of 29 years, 6 (3 female) were between 30 & 39 years of age, and 11 (7 female) were over 40 years old.

Instructions and procedure

Each participant was asked to look at the screen, while the calibration equipment for the eye tracking device was aligned with the candidate's eyes. Once the eye tracking device was calibrated, the participant was given headphones. Each participant was introduced to the conditions with the following instructions.

There will be three tests: In the first test, you will see a blank grey screen and hear the same passage of music played twice. You are to say whether track one or track two sounds as though it has been recorded closer to the orchestra. You are to look at the screen throughout the 15 seconds of music and to shout out 'one' or 'two', to identify the track. Immediately afterwards for the second test, you will be shown four images of four different concert halls, and you are to look at the screen for the

full duration of each image and refocus your eyes in the blank screen in between each image. You are not required to say anything about what you saw or looked at. In the third test, you will be shown the same four images twice, each in a random order, but with two different passages of sound. In this instance you will listen for clicks within the soundtrack and identify using your two hands to point at where around your head you hear the sounds.

After the above instructions were presented, a summary of the actions required of the participants was repeated at the end as: "first test - say which track is closer to the orchestra; second test - look at the images; third test- count the clicks."

Results

Of the 28 participants, 20 produced usable eye tracking results. Some results were lost due to blind spots when participants moved forwards or backwards in the chair and the cameras could no longer build a face model and thus interpolate eye vectors. Other people turned to the author to ask a question and their noses partially blocked their eyes so the tracking fell off or was impaired.

Apparent Proximity to Orchestra

20 participants judged sound track 2 to be closer to the orchestra; 5 participants judged sound track 1 to be closer and 3 participants could not distinguish between the sound tracks. "Immersion" in sound is a way to describe the source broadening and envelopment effects of room geometry. "Appear to be closer" was a way to convey a meaningful question to the lay-person and untrained listener.

Eye Tracking

The eye tracking system produces records of fixation points and gaze points. Gaze points are recorded every 16 milliseconds (ms) with their X and Y coordinates relative to the bottom left-hand corner of the screen. Fixation points recorded over a 15 s period with their start and end times, and their duration.

For each participant, there were 14 x 15 s collections of data, namely 2 sound only tracks, and for each of the 4 halls, 1 image only track, 1 sound only track, and 1 image and sound track. For each collection of data there were circa. 800 plus gaze points. Fixation points varied from 0 to 14 per 15 s

trial, with an average range between 4.8 and 8.7 fixations per trial. Fixations captured by software varied between 0.5 s to 2 s. Eye movement tracks with locations of the gaze points and fixations were overlaid on each photographic image, as shown in Figure 3. Overlay plots for each participant, each hall and each sound/ image combination were arranged per subject for visual inspection, as shown for Participant 3 in Figure 4.

_____

Figures 3 and 4 here

_____

Fixation Boxes. Using each of the fixation lists, the average and standard deviation of the x and y coordinates of fixation points were calculated, giving the boundaries of a box that characterised the extent of the eye movement, Average ± STD, as shown in Figure 5.

___

Figure 5 here

___

The individual fixation boxes were reduced to their height and width; the height and width were then plotted as points on an xy graph, x being width and y being height, to allow comparison of responses for each room and test for each participant as shown in Figure 6. These allowed comparisons of eye movement envelopes up-down and left-right separately, and in relation to each other.

Figure 6 here

Room summary data for all participants were assembled for 1 test or 1 concert hall to give an average of the collective SD x and y. The extent of the spread among participants is indicated by the

error bars. Figure 7 shows the spread of values across all participants of the SD of the co-ordinates of their fixation points.

From the room summary graph, it is possible to see that:

1. Image only (Ai1-Di1) involves more extensive horizontal eye movement than 'image and sound'. (Circles more to the right than the triangles.)
2. Eye tracks for 'sound only' (S1, S2) are more vertical than either 'image only' or 'image and sound'.
3. Eye tracks for immersive 'sound S2 only', has more horizontal movement compared to all 'image and sound' and 'image only' and also the frontal 'sound S1only'.

Figure 7 here

—

Statistical Analysis

Test Type. Overall the fixation box width for 'image and sound' is the smallest for all halls taken as a group. The mean width of the fixation box for image and sound (128 pixels) is less than the mean width for image alone (167 pixels). For 'sound only', the fixation box S2, (the wider sound field) however, was the greatest of all fixation boxes, 54% wider for sound S2 (213 pixels, associated with the narrow halls images) than for sound S1 (138 pixels, associated with the wide hall images).
To determine whether these differences between means were significant as a function of the test type, (Image alone, Image and Sound, Sound alone), for each participant, the standard deviation for fixation box width and box height was entered into a within-subjects ANOVA having one factor of test type (Image alone, Image and Sound, Sound alone), and another factor of spatial dimension (x/y).

It was first necessary to test for sphericity independently for SD in width and height. In neither case was sphericity violated, thus it was unnecessary to correct for degrees of freedom. Across test type, the mean SD scores in width (x) were statistically significantly different ($F(2, 38) = 10.228$, $p < 0.001$). Mean scores for STD in the x dimension for Image, Image and Sound and Sound alone test types respectively were 164.8, 106.5 and 111.2 with standard deviation 48.094, 46.643 and 63.401 respectively. It is clear that when sound is a condition the horizontal deviation in gaze points is

reduced. The mean SD scores for height (y) were not statistically significantly different ($F(2, 38) = 1.225$, $p = 0.305$). Mean scores for STD in the y dimension for Image, Image and Sound and Sound alone test types respectively were 111.4, 108.15 and 130.05 with standard deviation 48.032, 40.719 and 51.615 respectively.

Hall Type. For the test of Hall Type, it was deemed appropriate to consider the conditions of Image alone and Image with Sound only. Omission of the Sound alone condition was judged to be prudent because the test investigated the effect across Hall Types and the Sound alone condition had no Hall Type factor present. SD in x and y were used as factors across the conditions and across the Hall Type (shoebox vs. surround) for each participant.

Across Halls, the mean SD scores in width (x) ($F(1, 19) = 0.012$, $p = 0.913$), and height (y) ($F(1, 19) = 1.970$, $p = 0.177$), for the Image alone condition (Shoebox vs Surround) were not statistically significantly different. For the Image with Sound condition, across hall type, the mean SD scores in width (x) ($F(1, 19) = 0.377$, $p = 0.547$), and height (y) ($F(1, 19) = 0.534$, $p = 0.474$), were also not statistically significantly different.

Summary of ANOVAs These tests show that there is a significant difference in the SD of fixation points in the x direction, but not the y direction when comparing the effects of the 3 conditions of visual image only, image plus sound, or sound alone. Specifically, there is greater extent of movement in the x direction for the visual only condition. Given the image of the type of hall yields no significant difference across the SD in x or y, it can be said that the observed effect persists invariant of the type of hall.

Saliency Maps

To demonstrate some of this information graphically and for each hall, the fixation focus area (which represents the standard deviation for all participants for that hall) for sound only, image only and sound and image were overlaid on the concert hall image.

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 color, orientation, motion, depth etc. Saliency maps of the type developed by Itti and Koch (1998) were designed as input to the control mechanism for covert selective attention. A saliency map of each hall was produced and the same fixation focus areas were overlaid for each concert hall to indicate what visual highlights the eyes were drawn to, on average.

In Figure 8 an overlay of the fixation box patterns on the saliency diagrams of the different halls demonstrates that the average fixation pattern is not directly looking at highlighted features but to one side of them. The image of the whole screen was 1280 x 1024 pixels, an area of 1,310,000 pixels, and yet the largest SD box is 34,000 pixels in comparison.

Discussion

The research questions set up before the experiment were answered:

1. Is the visual field that is the area of eye movements in the horizontal and vertical directions, different when the brain is occupied by listening, than when it is not?

The results indicate that the visual field is reduced in the horizontal direction during listening to the presented music while looking at images of concert halls.

2. Does the visual field change in the same way with a surround room as a shoebox space?

   Yes the visual field reduces in the horizontal direction for both surround and shoebox spaces 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. Is there anything to be learnt from what participants look at?

People appeared to have individual preferences for areas at which they tended to look: at the orchestra, above the stage, at the ceiling. A remarkable number of people did not look at the performers on the stage, but upward to the ceiling, the lights and to details that caught their eye. For the two tall, narrow halls, the participants tended to look above the stage or behind the stage, whereas for the surround halls they looked at the front edge of the stage when looking at the image alone and

lifted their gaze when listening and looking. However there was not statistical evidence to support this claim.

Whilst active listening was encouraged for the sound alone and image with sound conditions, the task of the participants was different in the sound only and in the sound with image conditions. Whilst this difference was small and designed to mitigate any learning effect in the procedure, top down task related bias may have been introduced. Future work can explore more directly the links between sound alone and image and sound.

Whilst the results from this study correlate quite strongly with the work presented by Schafer and Fachner (2014), it should also be mentioned that this work may be general or indeed scene dependant. Further study would be required to determine this. It would be premature to suggest, however, that the results of the present study can be applied to the design of concert halls, although future research should investigate the implications of having the scale and complexity of features concentrated on the centre of the room.

In further experiments, fixed images could be replaced by video recordings of performances in specific halls. The sound could be recorded binaurally in each hall to further analyze the impact of the type of room. Ideally the same piece of music would be heard with the same conductor and orchestra in several different performance halls, as might be arranged through the work of Tapio Lokki and colleagues with the development of the virtual orchestra methodology as found elsewhere in this special issue. Indeed, a database of Head-Related Transfer Functions could be provided to participants so they can choose the best or most appropriate function for spatial localisation for them.

In conclusion, the introduction of a listening task to the viewing of concert hall images results in a reduction of the width of eye movement. When there is an image to look at and the brain is occupied by listening, then the horizontal eye movement is more focused than when there is an image to look at without background music. Although an auditorium may be beautiful to look at in every direction, the visual attention of the audience may be more focussed in the horizontal plane once the music starts.

References

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

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

Dammann, G. (2015, February 6). Interview: Finnish conductor Susanna Mälkki. *Financial Times*. http://www.ft.com/cms/s/0/fa1dbdf4-aba8-11e4-b05a-00144feab7de.html

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

Essert, R. (1999). 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.

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

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

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

Harvey, C. (2011). *Modality Based Perception for Selective Rendering*. Unpublished doctoral dissertation. Warwick University.

Hulusic, V., Harvey, C., Debattista, K., Tsingos, N., Walker, S., Howard, D., Chalmers, A., Acoustic Rendering and Auditory-Visual Cross-Modal Perception and Interaction. *Computer Graphics Forum, 31*, 102-131.

Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 20*, 1254-1259.

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

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

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

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

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

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

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

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

Valente, D.L. Braasch, J., & Myrbeck, S. (2012). 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).

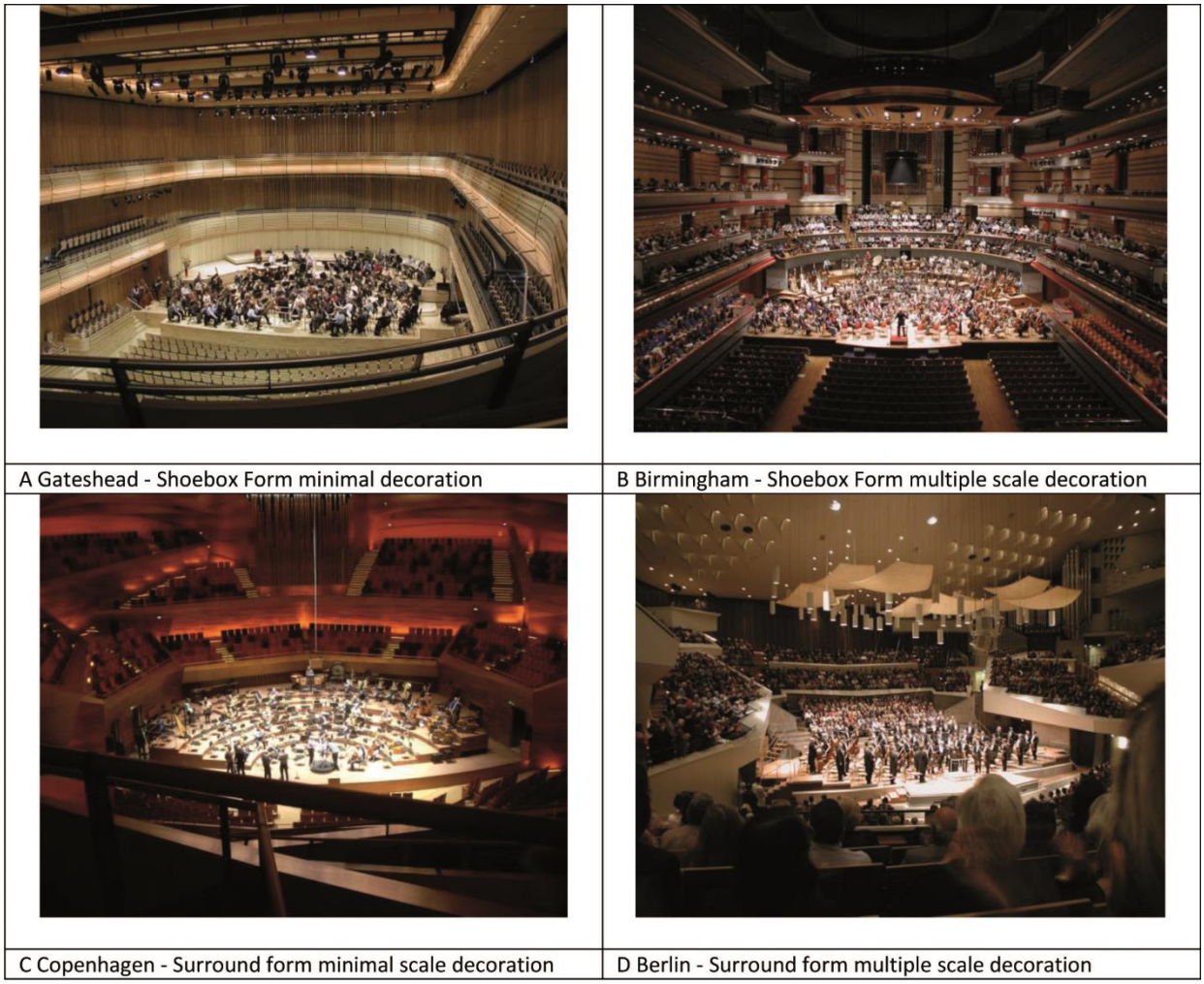

Figure 1. Images of the four halls.

Figure 2. Equipment setup: display, eye tracker and headphones.

Figure 3. Image of concert hall with calibrated fixation points. Typical green fixation points (circles) overlaid onto image on screen. The size of the green circles indicates the length of the fixation point. The lines in between the fixation points are the saccades of gaze points recorded in different colours.

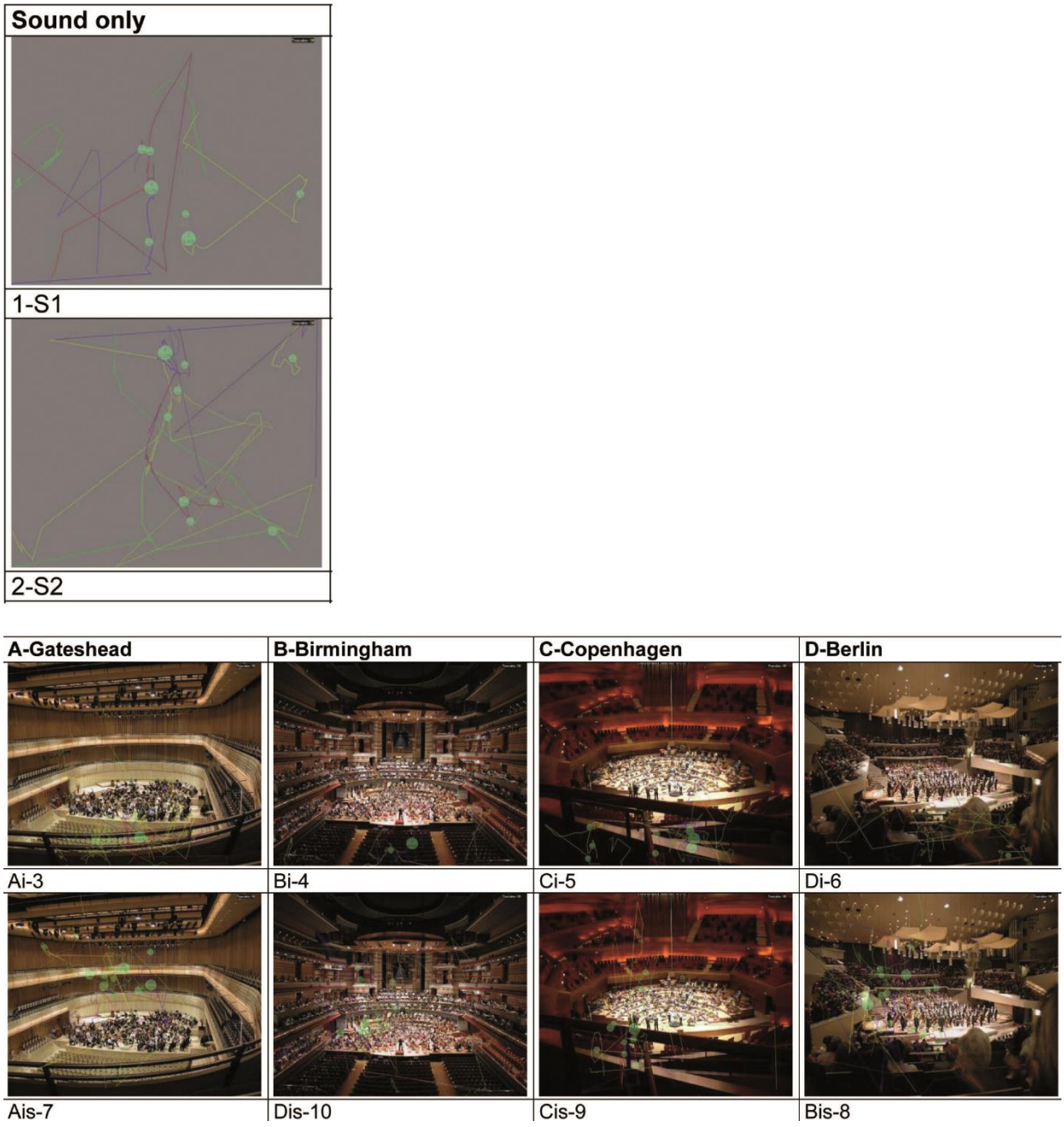

Figure 4. Eye tracks for Participant 3 (raw data on which analyses were based).

Top panel) Eye tracks on grey background for sound only. S1 for the first sound and S2 for the second sound.

Bottom panel) Tracks for image only 'i', prefaced with the concert halls A, B, C or D.

Tracks for sound and image 'is', prefaced with the concert halls A, B,C or D.

Figure 5. Example fixation boxes for Berlin. Di (green) larger box is for image only. Dis (red) smaller, upper box is for image and sound.

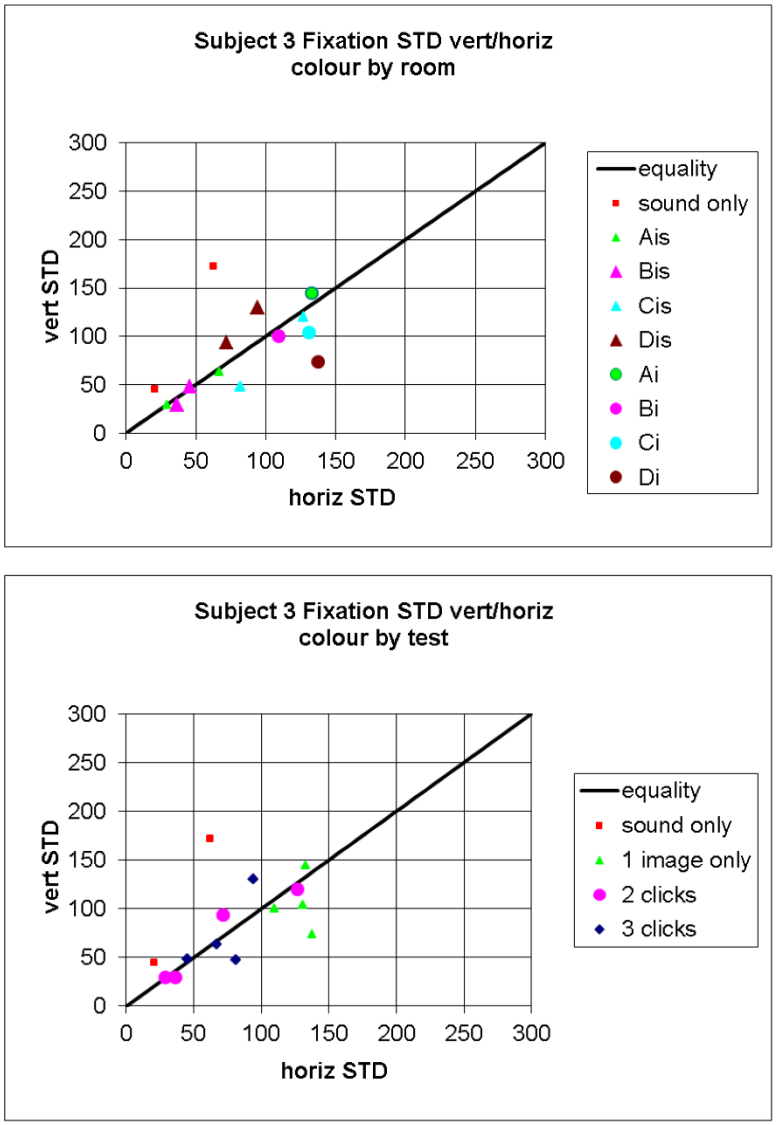

Figure 6. Typical subject summary plot by room and test.

Typical fixation standard deviation for one participant with data sorted by room. The graph shows the extents of the standard deviation in the x and y directions so that a point higher on the graph is a taller box and farther to the right represents a wider box. The units upon each axis are units of pixels.

Results for image only are titled 'i' and prefaced with the concert halls A,B,C, D.

Results for sound + image are titled 'is' and prefaced with the concert halls A,B,C, D.

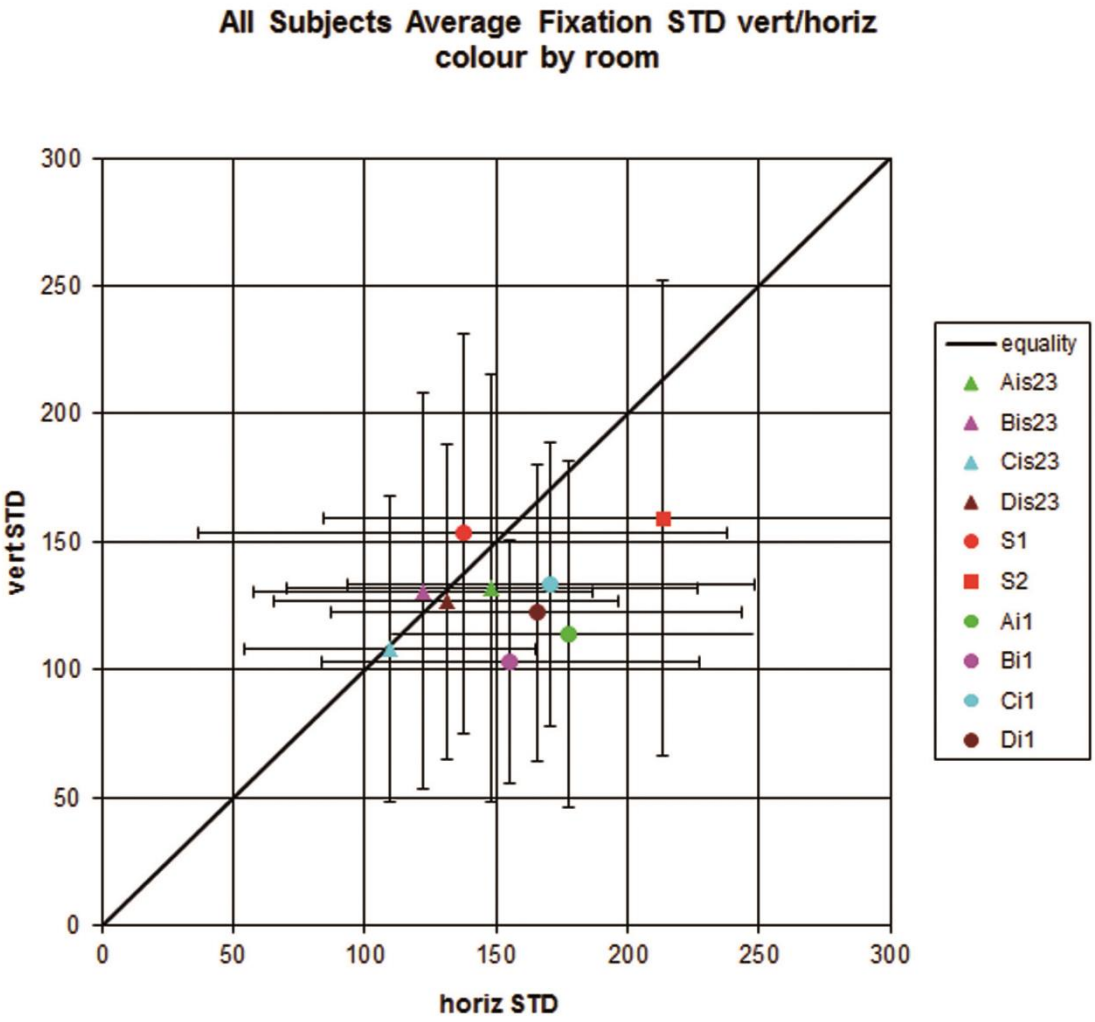

Figure 7. Average fixation box size (in pixels) for all participants for each test, showing the spread of values across all participants of the SD of the co-ordinates of their fixation points. S1 and S2 are sound tracks 1 and 2 SD; Ais23-Dis23 are image and sound for halls A-D for images 2 and 3 of each hall; Ai1-Di1 are image only for halls A-D. Points below the diagonal indicate fixation focus areas that are wider than they are tall. Points above the diagonal indicate fixation focus areas that are taller than they are wide. A square box would lie on the equality line.

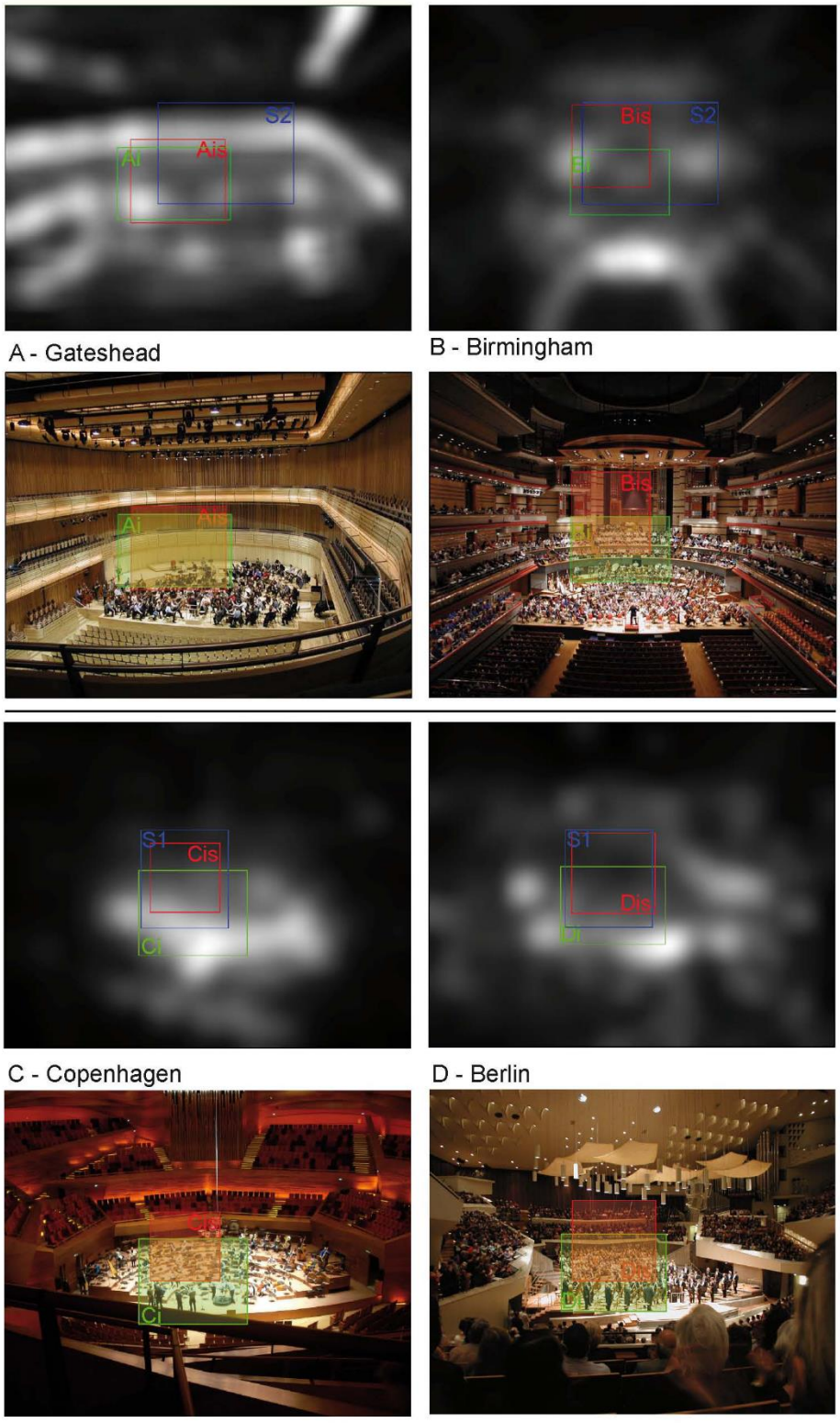

Figure 8. Overlay of fixation boxes on saliency map (above) and room image.

Tests: “i” – image only; “is” – image and sound. “S1, S2” – sound only, for comparison.