Object-Based Audio
Applied to Football Broadcasts
The 5 live Football Experiment
Mark Mann, Anthony Churnside, Andrew Bonney, Frank Melchior
Object-Based Audio Applied to Football Broadcasts

Mark Mann, Anthony Churnside, Andrew Bonney, Frank Melchior

Abstract

In this paper, we describe an object-based audio broadcast of a football match from Wembley Stadium, London. Listeners were given audio feeds from opposite ends of the stadium together with a commentary feed and were able to mix the balance between the three feeds using the HTML5 audio API. The experiment proved popular with the people who took part. Most listeners chose to set their preferred mix within the first minute of the broadcast and did not alter it as a result of events in the stadium. Also, preferred mixes were clearly observed.

This document was originally published in the Proceedings of the 2013 ACM international workshop on Immersive media experiences, pages 13-16. The slides used in the presentation are attached as an appendix.

Additional key words: Measurement, Performance, Design, Reliability, Experimentation, Human Factors, HTML5, audio, Javascript, broadcast, football, radio.
Object-Based Audio Applied to Football Broadcasts

Mark Mann, Anthony Churnside, Andrew Bonney, Frank Melchior

1 Introduction

The BBC has been working on a number of public facing experiments [1,2] which have explored various aspects of an object-based approach to audio broadcasting. The object-based approach can be seen as follows from responsive web design [3] in which the content responds dynamically to input and interaction from the consumer. For this method to work, broadcasters, like web designers, need to produce an set of common assets (text, pictures, sound and video) together with metadata to determine how these assets are rendered in response to the type of device asking to present them. The first experiment used proprietary technology to deliver the commentary and background audio signals separately in order to give the listener control of the balance between the two [1]. The second experiment explored the concept of perceptive media. In this experiment the content was altered based on the geolocation of the listener [2]. This experiment was based on existing web technology.

The aim of this work was to investigate a further extension of the experiences described above applied to sport events based on available web technologies, to determine the demand for such experiences and to determine how listeners used the interface and how they responded to events during the broadcast.

On Monday 27th May 2013, BBC R&D and BBC Radio 5 live provided an experimental broadcast of the English Football League Championship Play-off Final between Crystal Palace and Watford from Wembley Stadium, London. The developed system enabled the listener to choose the audio balance between the crowd and commentator and choose which set of football fans they wanted to listen to. The experiment enabled this by broadcasting three live streams over IP, with one pair of stereo microphones pointing at the Crystal Palace fans, one pair pointing at the Watford fans and a mono feed from the commentary box. The user interface employed the HTML5 web audio Javascript API [3] to control the streams thus enabling the consumer to alter the relative balance between the three streams. This paper describes the experimental system and summarizes the insight gained from data describing how people interacted with the user interface.

2 Setup of the broadcast chain

The broadcast chain can be divided into three parts, each part interdependent on the others. These were production, in this case the capture of audio at Wembley Stadium, distribution and consumption.

2.1 Production: Wembley Stadium

Four microphones were used to pick up the crowd noise; each acted as either a left or right channel for one end of the stadium. A balance needed to be struck with the position of the microphones, too close to the crowd risked being able to pick up individual voices, with the broadcasting of bad language a possible consequence; too far, and the feeling of ‘being there’ would have been removed. The microphones were positioned near the corner flags, just in front of the advertising hoardings, pointing at the crowd. A radio transmitter attached to each of the four Sennheiser shotgun 416 broadcast microphones sent the audio collected to radio receivers at the Radio 5 live commentary box as shown in Figure 1.
An analogue-digital audio converter was used to convert the audio signals fed, via a single USB, into a PC. The PC contained a multi-channel mapper and 3 purpose-built 128 kbps AAC codecs which served the streams over IP back to the BBC. Two stereo streams for the crowd noise and a third stream for the commentary (which was taken by lip microphone) were delivered in this way.

2.2 Distribution: transcoding and limiting

The audio streams from the stadium were received at a server in BBC R&D, transcoded to MP3 and Ogg Vorbis formats so that all major browsers on all major operating systems could receive at least one set of streams that HTML5 audio can play [4]. The streams were distributed by an Icecast server and the number of concurrent streams initially limited to 3000 (or 1000 listeners), but extended to 12000 (or 4000 listeners) during the course of the trial.

2.3 Consumption: the User Interface

![Figure 2. The interface used by listeners to control the sound balance they heard](image)
The user interface was designed to be as simple and intuitive as possible. The result is shown in Figure 2. Listeners were able to mix between one end of the stadium and the other with the left side 100% Crystal Palace and the right 100% Watford. This was done by dragging a microphone icon over a plan view of the stadium. Only the position of the mouse on the left-right axis was used to determine the desired balance. Listeners were also able to change the balance between the commentary and the crowd. This was done by a simple drag bar at the bottom of the interface. At the extremes, either commentary or crowd would be 8 times the amplitude of the other. Listeners were also able to stop, start and control the overall volume on the interface.

All interaction with the interface was logged. The position of the microphone was logged every 500 ms during dragging. The position of the commentary balance bar was logged when the mouse was released after dragging.

3 How listeners interacted with the broadcast

![Figure 3](image3.jpg)

*Figure 3.* A graph showing how 6 randomly selected users moved the microphone from left to right with time

5286 people tried to listen to the broadcast, but some were unable to because the number of streams were limited in the first instance or because their browser was incompatible. Logs were taken from each of the 2692 separate listeners who listened to the trial during the course of the match. A clear pattern of activity was established for each user when they first interacted with the user interface. Figure 3 shows how a random selection of listeners behaved during the first 90 seconds of interaction when choosing where to position the microphone. Some users would move the microphone to many positions during those first few seconds, but most chose a position to place the microphone and stay with it. 65% of listeners did not move their microphone 30 seconds after beginning to listen to the broadcast. This was also representative of how all users interacted with the commentary balance interface. Figure 4 shows a histogram of how all users settled on a particular commentary balance with time. The timing of each user’s activity was adjusted so that 0 was the point at which the commentary balance was first adjusted. The data is taken from 1140 listeners who joined during the first half with the total number of listeners on each point of the scale logged against adjusted time.

![Figure 4](image4.jpg)

*Figure 4.* A histogram showing how all users changed the commentary balance with time in the first half.
Note three peaks appear very quickly in Figure 4. The biggest peak was to equally balance the crowd and commentary. There were two further peaks at the extremes maximizing either the crowd or commentary feeds. Roughly twice as many people chose to increase the contribution of the crowd as compared with commentary.

The pattern of activity was repeated for people choosing the microphone position. There was no such difference in choosing which end of the stadium to listen to. Figure 5 confirms the final positions chosen by the listeners for the microphone. The preferred microphone positions were equally distributed between the centre and the two ends of the stadium.

![Figure 5](image)

**Figure 5.** A heat map showing where people chose to position their microphone and commentary balance. Blue is zero activity, red shows the most popular spot.

Having arrived at a position to listen to the experience, what listeners did next was investigated. Of particular interest was how listeners reacted to events during the match. Figure 6 shows the total activity (or number of events logged in the database) with time with a few key incidents in the match and advertisements for the trial noted on it.

At Wembley, the broadcast went smoothly, though a small adjustment in the relative crowd/commentary level needed to be made a few minutes into the broadcast. Unfortunately, the Icecast server needed to be restarted 20 minutes into the broadcast because new connections were being refused. This configuration error caused a spike in activity during the match (see Figure 6). The Ogg Vorbis commentary stream failed to restart at the beginning of the second half which also caused a spike in activity which will now be discussed in greater detail.

![Figure 6](image)

**Figure 6.** A graph comparing microphone activity, commentary activity and total activity. Key events in the stadium, together with advertisements going live and technical issues are labelled on the graph.

Figure 6 presents the activities of all listeners divided into three types of activity:
1. Change of commentary/crowd balance (blue),
2. Change of microphone position (red)
3. Cumulated total activity, which includes the sum of one and two as well as listeners joining the broadcast.

The plots in Figure 6 strongly correlate with each other. This suggests that there was no difference in the way that listeners responded to setting the microphone position and commentary balance and that whatever the cause of each peak in activity, listeners did not choose to adjust one thing over another.

To investigate the behaviour of the listeners further the number of listeners joining was separated out. Furthermore the activities of the listeners were pre-processed based on the insights of their behaviour when they first join the broadcast. Figure 3 and 4 suggested that listeners produce a burst of activity when they first join the broadcast. Figure 7 shows that people joining the broadcast was the major factor in spikes of activity.

![Figure 7](image)

**Figure 7.** A graph comparing microphone activity with the number of listeners joining the broadcast and the combined activity of all listeners but discounting each listener’s first 5 minutes of activity

Also plotted in Figure 7 is a line, which accumulated all activity of users but only 5 minutes after they had joined the broadcast. This is to eliminate the burst of activity seen by all users upon joining the broadcast. If events in the stadium prompted activity in listeners, a peak should still be visible in the line. Figure 7 shows that there is no correlation between events in the stadium and activity. It also shows that activity generally fell away during the course of the match.

One clear conclusion that can be drawn about the behaviour of listeners was that they would adjust the balance between the three audio feeds that suited them and then stuck with it for the broadcast. They did not react to events in the stadium, so the key motivation for listeners was to improve the general listening experience.

4 Feedback and Discussion

BBC Radio 5 live typically broadcasts online using a 56 kbps G.722 mono codec, so this broadcast provided a huge improvement in both sound quality and clarity which was reflected in responses to the experiment in social media. Responses fell into two categories, those hard of hearing who appreciated being able to isolate the commentary (e.g. “Brilliant idea. Now I can reduce crowd volume and actually hear the sometimes mumbled comments of the expert summarisers”) and those who liked being able to turn the crowd up and listen to the crowd (e.g. “love it, great to be able to hear the atmosphere more over the commentary”). Listeners were invited to fill in a survey with the key results from it shown in Figure 8.
Roughly three quarters of participants in the trial preferred being able to choose which end of the stadium the sound came from, roughly three quarters of participants strongly preferred being able to control the commentary/crowd mix and the experience over traditional radio coverage. Furthermore, 60% of listeners considered the interface easy to use with a roughly even split between those listeners who wanted more control and those who did not. This confirms that listeners chose not to adjust their preferred balance during a broadcast rather than this being the result of a usability problem.

This feedback, combined with demand for the trial exceeding server capacity makes a strong argument for the BBC and other broadcasters to provide an object-based service in the future. However, what this experiment did not solve was what form such a service should take. No firm conclusions can be made about what the ideal mix or mixes should be nor the degree of choice a listener would like in choosing what the mix should be or indeed if they want to choose a mix at all. A survey question on whether listeners wanted the choice was inconclusive. Also, though clear peaks appeared in both the commentary balance and microphone position, limits and a ‘centre point’ were placed on both scales which will have influenced the listeners’ selection. A further experiment with unconstrained axes and randomized start points needs to be taken in order to establish whether the peaks are real and will be the focus of future work.

5 Conclusion

This work has demonstrated how currently available technology can be used to enable new audio experiences for IP end-to-end broadcast systems. The audience feedback suggests that raising the audio quality is appreciated and the option to personalize the audio mix is preferred by three quarters of the participants of this experiment. Nevertheless the interaction patterns and importance of being able to personalise the broadcast are very different depending on which variable of the mix is changed by the listener (position in stadium or crowd commentary balance). Further work will investigate how the object-based content delivery will enable new experiences in other types of broadcasts and a redesigned unconstrained interface will be designed to better determine listeners’ preferred mixes.
6 Acknowledgements
The authors would like to thank BBC Radio 5 live, Matt Paradis and Jasmine Cox of BBC R&D and Andy Armstrong of Hexten for their help and support in realising this project.

7 References
4) Rogers, C. 2013, Web Audio API, http://www.w3.org/TR/webaudio/

8 Appendix
The presentation as delivered the 2013 ACM international workshop on Immersive media experiences, with some notes.

Slide 1

Object-Based Audio Applied to Football Broadcasts
Mark Mann, Anthony Churnside, Andrew Bonney, Frank Melchior
BBC Research & Development
R&D North Lab, 5th Floor, Dock House, MediaCity, SALFORD, Lancashire. M50 2LH.
mark.mann@bbc.co.uk
Our Role

- Define the future BBC
  - Keep the BBC at the forefront of technological and creative development and innovation.
  - Help the BBC to adapt and evolve its content and services through technology.

- Lead and shape our industry
  - Set standards
  - Make new developments widely and openly available
  - Influence and support the UK’s connected digital media economy

- Help the BBC do more for less
  - Enable financial savings
  - Environmental impact

BBC R&D Overview

- BBC Research & Development:
  - 210 staff (around 165 research engineers) working across three laboratories
    - South Lab ~ 140 staff
      (Centre House, London W12)
    - North (Media City UK) ~ 40 staff
    - Central (1 Euston Square) ~ 30 staff
  - Overall core budget of £13 – £14m per year, comprising:
    - Majority from BBC licence fee
    - Additional income from licensing of IPR and expertise
    - Collaborative funded projects (EU & UK/TSB)
Object-based audio

- Object-based approach to broadcasting closely follows responsive web design.
- The content responds dynamically to input and interaction from the consumer.
- Broadcasters produce a set of common assets (text, pictures, sound and video) together with metadata to determine how these assets are rendered in response to the type of device asking to present them.
- Netmix
  - Used proprietary technology to deliver the commentary and background audio signals separately in order to give the listener control of the balance between the two.
  - Explored the concept of perceptive media; in this experiment the content was altered based on the geolocation of the listener and was based on existing web technology.

The aim of the work described in this paper is to investigate the further extension of a new experience for sport events based on available web technologies.

---

BBC Radio 5 live broadcasts news during working hours weekdays and sport the rest of the time. Listeners can access the service through the website as well as listening on MW, DAB, Freeview and Freesat.
Project goals

- Choose location and commentary balance in stadium
- Determine public demand for service
- Develop a prototype HTML5 audio player which can multiple audio sources.
- Gather data and feedback on the player’s user interface.
- To determine limitations for outside broadcast
- To verify the requirements for scale up and infrastructure

In the stadium

Microphones were positioned at the four corners of the stadium, two point at each set of fans.
The signals were sent back to a radio receiver

Where they were grouped into a stereo mix of both sets of fans and recorded together with the commentary
For the live broadcast, this was the broadcast chain

This is a screengrab of the experiment as it appeared on the day of the broadcast. Data describing how people used the interface was grabbed from each listener. They were also invited to complete the survey.
This is a heatmap showing where listeners chose to place the microphone and commentary balance.

65% of listeners did not move their microphone 30 seconds after beginning to listen to the broadcast. This was also representative of how all users interacted with the commentary balance interface. This shows a histogram of how all users settled on a particular commentary balance with time. The timing of each user’s activity was adjusted so that 0 was the point at which the commentary balance was first adjusted. The data is taken from 1140 listeners who joined during the first half with the total number of listeners on each point of the scale logged against adjusted time.
The top graph presents the activities of all listeners divided into three types of activity:

1. Change of commentary/crowd balance (blue),
2. Change of microphone position (red)
3. Cumulated total activity, which includes the sum of one and two as well as listeners joining the broadcast.

These lines strongly correlate with each other. This suggests that there was no difference in the way that listeners responded to setting the microphone position and commentary balance and that whatever the cause of each peak in activity, listeners did not choose to adjust one thing over another.

The bottom graph compares microphone activity with the number of listeners joining the broadcast and the combined activity of all listeners but discounting each listener’s first 5 minutes of activity. The graph shows that there is no correlation between events in the stadium and activity. It also shows that activity generally fell away during the course of the match.
Audience Feedback

- “A brilliant innovation, congratulations to all involved and hope to see this more across other sports too in future.”
- “Brilliant idea. Now I can reduce crowd volume and actually hear the sometimes mumbled comments of the expert summarisers. If this is available next season I'll definitely use it. Would be really useful on TV red button as well.”
- “Think it should be used for all future matches. Much prefer to hear the sound of the crowd and this gives you an opportunity to have the best of both worlds dependant on what you want to hear. Clever stuff...”
- “Great idea, especially if you are a fan of one club or like to hear the crowd participation, along with the usual excellent BBC commentary.”
- “Please spread this service to TV Dramas and Documentaries where the music is invariably far too loud and intrusive. It would be heaven to have control of those horribly loud radio5live station idents and trailers. I find myself hitting the off button every time the jingle SHOUTS at me!”
- “Should work well with other sports too like Boxing, Tennis and F1....”
- “I am a blind person using a screenreader. The only controls which worked were those for play and pause. More thought needs to be given to making this idea work for your blind users. I did enjoy the stereo sound, which I am sure makes the match more exciting.”
- “There is a noticeable delay between the commentary feed and the reaction of the crowd. Otherwise superb experience.”
Roughly three quarters of participants in the trial preferred being able to choose which end of the stadium the sound came from, roughly three quarters of participants strongly preferred being able to control the commentary/crowd mix and the experience over traditional radio coverage.

Furthermore, 60% of listeners considered the interface easy to use with a roughly even split between those listeners who wanted more control and those who did not. This confirms that listeners chose not to adjust their preferred balance during a broadcast rather than this being the result of a usability problem.
Conclusions

• This work has demonstrated how currently available technology can be used to enable new audio experiences for IP end-to-end broadcast systems.

• Audience feedback suggests that:
  – raising the audio quality is appreciated and
  – the option to personalize the audio mix is preferred by three quarters of the participants of this experiment.

• Interaction patterns and the importance of being able to personalise the broadcast are very different depending on which variable of the mix is changed by the listener (position in stadium or crowd commentary balance).

• Further work will investigate how the object-based content delivery will enable new experiences in other types of broadcasts and a redesigned unconstrained interface will be designed to better determine listeners' preferred mixes.