Audio Engineering Society

Convention Paper 9012

Presented at the 135th Convention
2013 October 17–20 New York, NY, USA

Evaluating iBall - An intuitive interface and assistive audio mixing algorithm for live football events

Henry Bourne[1], Joshua D. Reiss[2]

[1] Queen Mary University of London - London, E1 4NS, UK
henrybourne@me.com

[2] Centre for Digital Music, Queen Mary University of London - London, E1 4NS, UK
josh.reiss@eecs.qmul.ac.uk

ABSTRACT

Mixing the on-pitch audio for a live football event is a mentally challenging task requiring the experience of a skilled operator to capture all the important audio events. iBall is an intuitive interface coupled with an assistive mixing algorithm which aids the operator in achieving a comprehensive mix. This paper presents the results of subjective and empirical evaluation of the system. Using multiple stimulus comparison, event counting, fader tracking and cross-correlation of mixes using different systems, this paper shows that lesser skilled operators can produce more reliable, more dynamic, and more consistent mixes using iBall than when mixing using the traditional fader-based approach, reducing the level of skill required to create broadcast quality mixes.

1. INTRODUCTION

Mixing the audio for a live football event is a challenging task, requiring a great deal of experience and a high level of mental focus. iBall is a system that assists the operator in their role, taking away some of the more functional work and allowing them to focus more on the creative and aesthetic decisions to create a more complete and balanced mix.

The focus of this paper is the evaluation of the iBall system. First, the method of capturing the audio for a football event is presented, and the current fader-based method for mixing the audio is outlined. This is followed by an overview of the iBall system and related work. The subsequent sections present a range of tests to compare the performance of iBall and the mixes it creates against the fader-based mixes. Finally, conclusions and possibilities or further work are discussed.

2. CAPTURING AND MIXING THE ACTION

The audio for most football events is captured and mixed using a tried and tested method [1]. Although the basic configuration and workflow is largely the same across all events, there are slight differences in a few areas.

Capturing the audio of on-pitch events involves around twelve shotgun microphones, placed at specific locations around the pitch as shown in Figure 1. The sound of the crowd and other ambient audio is captured with additional microphones located above or around the stands.

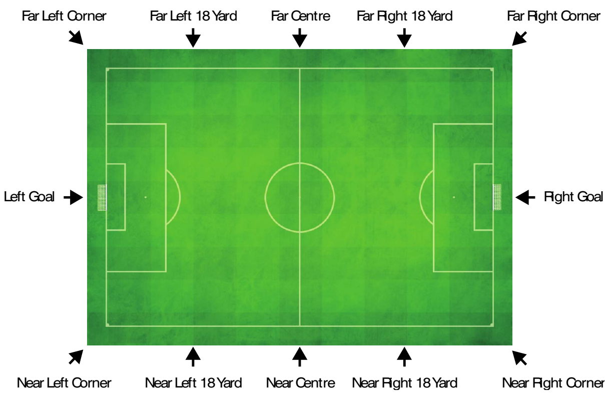

Figure 1. Microphone placement for capturing on pitch action

When the game is in play, it is the job of the operator to ensure that the audience at home can clearly hear all key components of the broadcast and recreate, as best they can, the atmosphere of the event. To bring the viewer closer to the action the operator has to ensure that all the important on-pitch events are clearly audible.

All events on the pitch could be recreated for the listener at home by simply fading up all the pitch microphones. This is problematic due to the noise of the crowd and stadium being so high, that so much of it spills onto the pitch microphones. With the combined contribution of the crowd, which is essentially wide-band noise, from the ambient microphones and all pitch microphones, the desired audio events would be masked.

During a game, the operator essentially tracks the location of the action on the pitch and fades up the relevant microphone, or microphones, to allow specific events to be heard while reducing the amount of unwanted noise. Typically just one to three microphones are faded up at any time. The operator tracks the action by monitoring a video feed from the wide angle camera mounted above the centre line; this provides the same top-down view that is used for the majority of the broadcast.

As can be seen in Figure 2, the microphones that surround the pitch are 'unfolded' onto twelve faders on the audio desk. The operators label each microphone with an arrow symbol extending from one or more lines. The lines denote the edges of the pitch and the arrows show the direction that each microphone is pointing.

Figure 2. Layout of the mixing desk

The operator must project the location of the action on the pitch, a 2D plane, onto a 1D representation of the pitch boundaries. This is a challenging mental task, especially when considering that the operator must anticipate where the next kick or impact will occur and ensure that they have faded up the correct microphone, or microphones, before that event happens. Even experienced operators who mix football games frequently still make mistakes and fade up incorrect microphones. Some very experienced operators can mix this way without looking at the faders each time they make a change, much like an accomplished piano player does not have to look at the keyboard, but some operators do need to glance down at the faders often to make sure they are in the correct location.

It is the process of simplifying the 2D to 1D mental translation that is the key focus of this paper.

3. IBALL

Based on previous work [2], which describes the algorithm and processing in detail, iBall presents a representation of the football pitch on an iPad.

The operator inputs the locations and directions of the microphones used to capture the on-pitch action, then simply touches the pitch representation to indicate where the action is occurring. This basic implementation was extended to incorporate multi-touch allowing an operator to specify multiple points of interest (POIs).

Based on these POIs, the system will select the most appropriate gain to apply to each microphone in order to best capture the on-pitch audio at those locations while reducing unwanted noise from other, non-critical microphones. The approach and algorithm is similar to the process of distance based amplitude panning [3], [4] and [5] although essentially applied in reverse.

The additional POIs were automatically faded in and out to avoid abrupt changes in the audio.

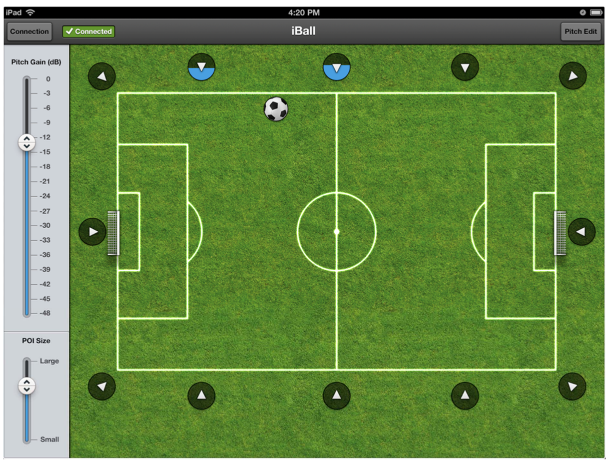

Figure 3. iBall user interface

In order to actually affect audio, the gain values calculated by iBall were converted into fader level messages to control faders on an audio console via a TCP/IP control protocol.

As proposed in [2], the system allows for the variation of the effective size of each POI. For the purposes of the tests the POI size was set to be relatively small. This means that fewer microphones are used at once to capture each POI. A large POI size would more evenly share the gain across multiple microphones.

4. OPERATIONAL TESTS

In order to subjectively evaluate the ability of iBall to create a satisfactory on-pitch mix, it was necessary to acquire representative audio source material. The following resources were captured from a football event:

- Individual audio feeds from each pitch and crowd microphone
- The full audio pitch mix as created by the audio operator
- The full audio mix of the programme, including commentary
- A video capture of the programme

For the operational tests, the individual microphone feeds were played back and sent to inputs of an audio console. The microphones were laid out on faders and labeled as they would be in a truck. The video of the game was displayed on a monitor above the console to match the location of the video monitors in a truck. Everything was configured to be as similar to an actual truck configuration as possible. Compression and EQ was applied to the crowd mix to try to match the sound of the pitch mix created by the broadcast operator.

Five subjects, who were technically competent engineers with audio mixing experience but who lacked experience in mixing live football events, were asked to create a mix of a short, three minute segment of the match using the faders on the console. Each subject was then asked to make a mix of the same segment using iBall to mix the pitch audio. Subjects were allowed one pass at the mix using each method to avoid the operator being able to remember the flow of action which would allow them to better predict how to mix the audio. The subjects were allowed to familiarise themselves with the operation of both methods prior to the recording of the mixes.

4.1. Capturing the Mix Data

For each mix, playback of the audio and video was started at a known location. The recording would end after a pre-defined period. This ensured that there was no difference in the overall length of each mix, and that the start and end positions of each mix were the same.

Mix data consisted of individual post-fader output from each source on the console, as well as a full stereo mix of the pitch and crowd. Fader level data from the audio console was captured at a rate of 10 Hz.

4.2. Operational Comments

The subjects were asked to comment on their experiences of creating each mix. While not included in this paper due to length constraints, it is quite clear from the comments that there was an overwhelming opinion that iBall allowed the subjects to concentrate more on what was happening on the pitch, and reduced the mental effort that they had to make in order to make the pitch events audible.

5. LISTENING TESTS

In order to get a subjective opinion on the performance of iBall compared to fader based mixes, a multiple-stimulus listening test process was used [6]. Participants were presented with the statement, “The mix captures all important on pitch events”, and asked to compare mixes based on this rather than comparing the various sonic or artistic qualities. Subjects were asked to rate the mix that best captured all of the important events with a score of 100. The mix that captured fewest events was to be rated as a zero. The subjects were informed that they were to create their own scaling system for rating the mixes that fell in between the highest and lowest and not given any steering in any particular direction.

The samples consisted of a 30 second segment from the five fader mixes and the five iBall mixes created during the operational tests. An anchor was provided as a source in the test and was a mix consisting of the crowd noise only, with no contribution from any pitch microphones. No ‘perfect’ reference was used, but the broadcast mix created by the professional operator was included for comparison. All mixes were accompanied with the appropriate video content.

Eight subjects participated in the listening tests. Some had a background in audio equipment design, some from a practical audio background, and others had no professional audio experience at all. Ages ranged from 23 to 65.

The audio mixes and accompanying video were looped. Each mix was assigned to a random audio channel before each listening test started. Participants were free to jump between the different mixes as required using the solo buttons on the fader strips. Subjects were presented with an equivalent number of sliders which were used to rate the mixes.

5.1. Listening Test Results

As can be seen in Figure 4, the iBall mixes consistently score higher than the fader based mixes and generally have a narrower confidence interval. Opinion of the fader based mixes had greater variation between subjects but was generally lower than the iBall mixes.

All subjects correctly identified the anchor; the crowd-only mix. iBall mix 4 was consistently rated highly, with all subjects scoring it above 80%, with half of the subjects rating it as the best mix. It was the most consistently rated of all the mixes.

Interestingly, the professional broadcast mix rating varied quite considerably between subjects and ended up with a rather low mean. This indicates that even the broadcast mix does not capture all important on-pitch events and, given the higher rating of the iBall mixes, the output from a live event could potentially be improved in terms of subjective event capture by making use of the iBall system.

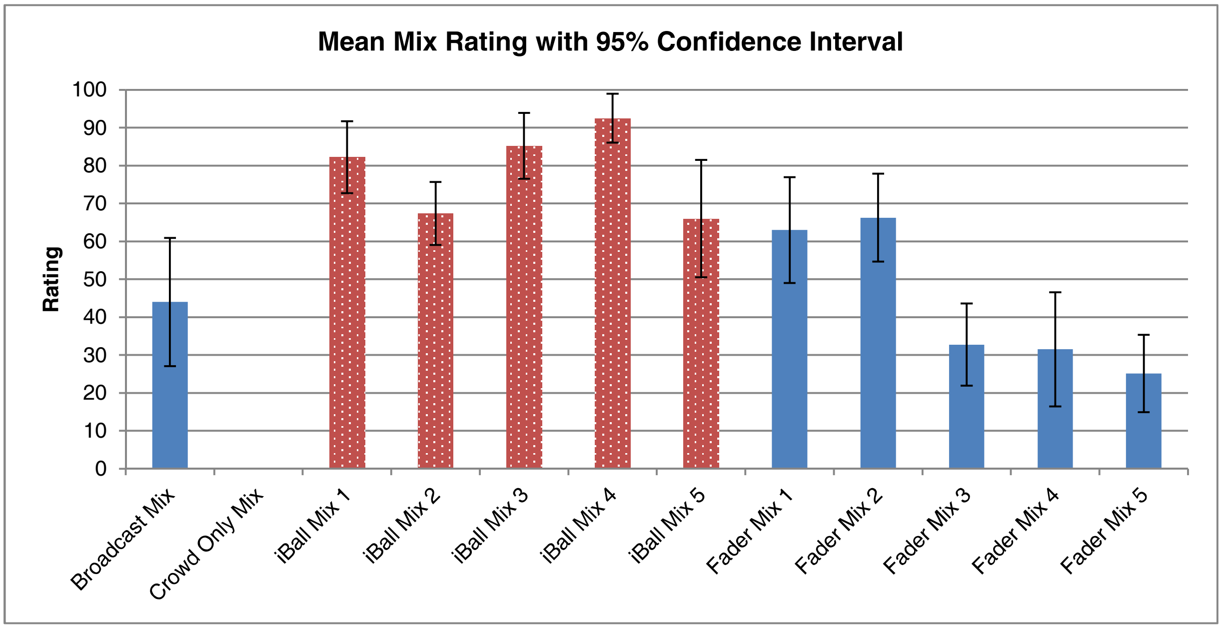

Figure 4. Mean mix rating with 95% confidence interval

6. EVENT CAPTURE

It was noticed during the listening tests and while observing the mixing processes, that iBall seemed more capable of capturing all of the on-pitch events compared to a fader based approach, where a number of key events were conspicuous in their absence. To determine this, inspired by work in [7], the number of definite, ambiguous and missed audible events were counted in each of the same mixes that were presented for the subjective listening tests.

The audible events were defined as visible events that created an expectation of accompanying audio. These were noticeable hard kicks and passes, body contact and ball bounces; there were twelve clear events in total in the examples used. Each mix was listened to three times while viewing the corresponding video. A simple scheme was used that related the numbers one, two and three to definite, ambiguous and missed events.

- Definite event: a clearly audible accompaniment to the visual event that met the listener's expectations.
- Ambiguous event: an audible accompaniment to a visual event that was either barely audible, or that didn't fully meet the expectation of the listener.
- Missed event: a visual event with no accompanying audio event. Listener expectation was not met at all.

As can be seen in Figure 5, there is a clear distinction between the iBall mixes and the fader mixes. The main statement that stands out is that across all the iBall mixes, not a single event was missed; including the more 'difficult' events such as ball bounces. The level of definite events was also consistently higher than in the fader based mixes.

There was much more variation across the fader mixes, with some performing much better than others. One fader mix actually captured more definite and ambiguous events than the professional broadcast mix, although this can not be an exact comparison due to the differences in levels and processing between the broadcast mix and the other mixes. We can also see that in three cases, the number of missed events exceeded that of the number of definite events.

The higher level of missed and ambiguous events in the broadcast mix could explain why it resulted in a low score in the listening tests.

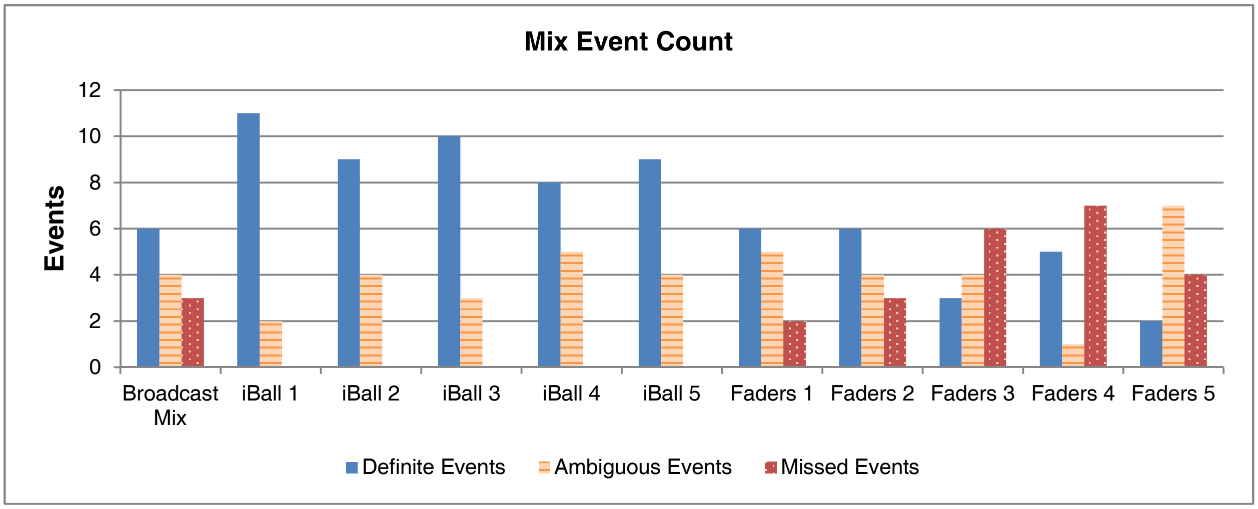

Figure 5. Event count

7. FADER ACTIVITY

Figure 6 shows the individual fader movements from the audio console over the first 30 seconds of a single iBall mix. Figure 7 shows the equivalent plot for a Fader based mix. Although plots are available for all mixes generated in this evaluation, only one plot for each mix approach is shown to due to length constraints.

It is clear from visual inspection of these plots that the iBall system promotes a more continuous adjustment of fader levels over time.

The mean fader movement over the course of each three minute mix was calculated. All iBall mixes had a mean adjustment between 45-55 dB, while the manual fader mixes all had mean adjustments in the range of 8-16 dB.

To calculate these figures, the derivative of each sample in the fader level vector was taken. This derivative vector was split into windows, each with a duration of five seconds. The absolute sum in each derivative vector window was taken, then the mean of all summed values was computed. This resulted in the average adjustment per fader per 5 second window. The mean of this value for all faders in each mix was computed. This provides a single number to explain the average adjustment made to each mix.

It is clear that iBall encourages a higher amount of adjustment to the mix. This may be due to the fact that it is easier to transfer a specific location of the ball on the pitch onto the pitch representation in iBall than it is to transfer that same location accurately onto the faders directly.

It was observed during the fader based mixes, that the subjects tended to leave up the one or two microphones that were nearest to the location of the action, even if the action was moving towards another microphone. When the action moved sufficiently further from the active microphones to become closer to a different microphone, then the old microphones would be closed and the new microphones opened.

With the iBall mixes, as the action changed position slightly the subjects would tend to respond to this change and update their touch location on the pitch representation. iBall would register this as a change in position and update the fader levels as appropriate. These more resolute changes to the fader levels could explain why listeners preferred the iBall mixes over the fader mixes, and why the number of audible pitch events is higher in the iBall mixes.

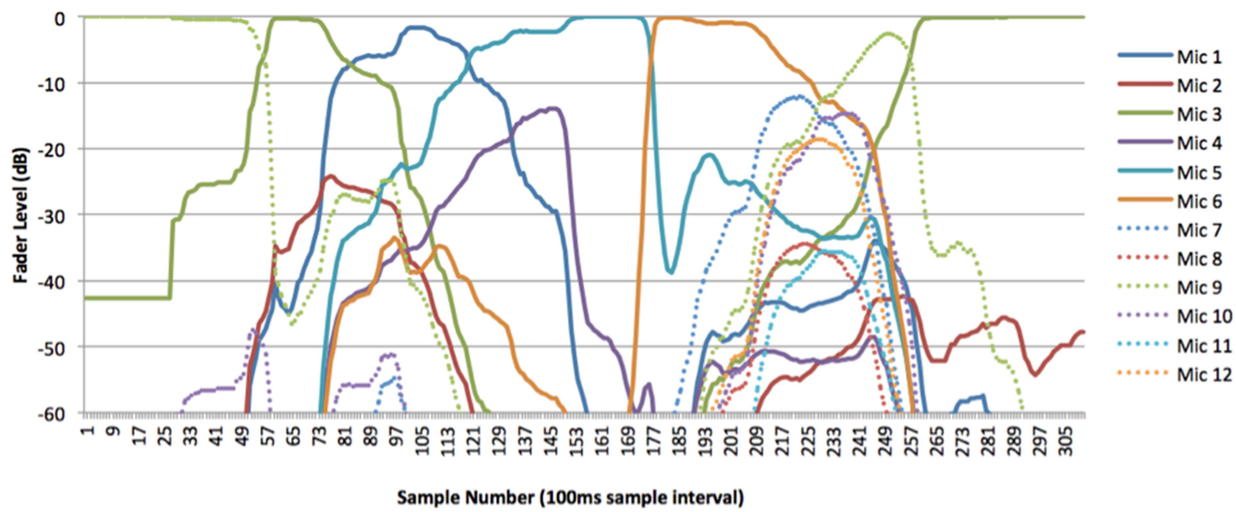

Figure 6. Example plot of fader movements from an iBall mix

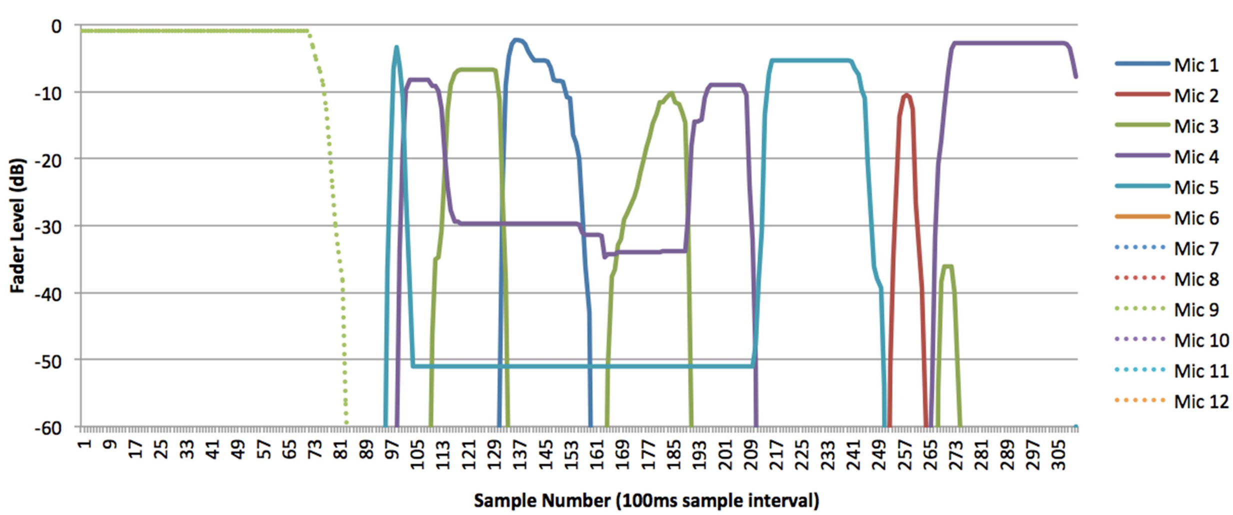

Figure 7. Example plot of fader movements from a fader-based mix

8. MIX CORRELATION

Observing the creation of the mixes, and listening to the stereo mixes afterwards, it was noticed that the iBall mixes seemed to be quite similar to one another in terms of the events they captured and their overall sound. Each fader mix seemed to be quite different to all the other mixes. This hypothesis was tested using correlation of the captured fader values.

For iBall mix 1, the level vector for fader 1 was correlated with the fader 1 level vector for all other mixes. The same calculation was performed for all other faders. Then the mean of all fader values for iBall 1 against each other mix was taken to produce a single value to show how well iBall mix 1 correlates with each of the other mixes.

This same process was performed for all other mixes, to build up a correlation matrix. Figure 8 shows this matrix in a 3D bar plot.

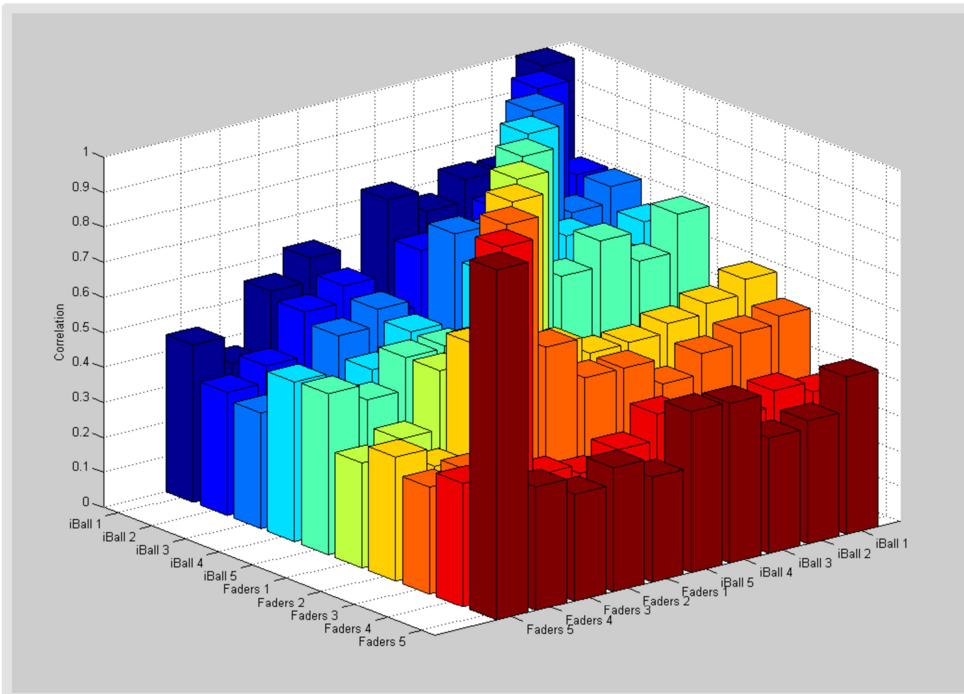

Figure 8. Mix correlation according to fader levels

This plot confirms that generally the iBall mixes are more similar to other iBall mixes, than the fader mixes are similar to other fader mixes.

This is presumed to be the case due to the way that iBall encourages operation. While it can be used in a number of ways, the most obvious and simple method is to use the location of the ball as it moves around the pitch as the POI. Every subject in the tests who followed the ball, moved the POI along a similar path, resulting in a similar output from the processing.

With the fader based mixes, the subjects were observed to take less similar approaches. Some subjects chose to open only the one or two microphones nearest to the action and open and close different microphones as the action moved around the pitch. One subject chose to open more microphones at once, sometimes five or six, while keeping them all at a lower level to try to reduce the chances of missing specific events with the compromise that an event captured may be at a lower level than if they had use fewer, more relevant microphones. As the subjects also all found it difficult to transform the two dimensional location of action on the pitch to the one dimensional fader layout, they were often slower to update the faders to reflect the ‘correct’ location. Some were faster at this than others. These differences in approaches and difficulties in location could contribute to the lower mix correlation between fader based mixes.

9. CONCLUSIONS AND FURTHER WORK

This paper presented the design and implementation of a novel interface to assist in the mixing of a live football event. Users are presented with an intuitive interface to specify the location of interest using a representation of the football pitch, rather than requiring complex 2D to 1D projection from pitch to a row of faders. This allows users to perform the task of creating a pitch audio mix with reduced mental effort and a greater capacity to concentrate on other important elements of the job.

An experienced audio engineer and broadcast quality mixing desk are still requirements for the task, and iBall does not replace either of them. It requires technical experience and good listening skill in order to create an appropriate mix structure, to apply necessary processing such as EQ and dynamics, and to attend to the myriad other tasks that are required to be performed simultaneously. The iBall system is intended as a simpler and more intuitive method of moving faders in order to better capture every important event during the football game.

Through the use of subjective listening tests and the analysis of empirical test data, it has been established that iBall has the potential to improve the ability of the audio operator to capture more of the important on-pitch events than they would do when using the faders manually. It is thought that the benefits of this system are of particular interest to lesser skilled operators, or markets who are newer to the task of producing high quality productions of football games for broadcast.

For future work, a potentially very useful development could be to involve some haptic feedback in the user interface. The operator's attention should be focused on the action on the pitch, rather than on the interface used to control the mix. With the iBall interface as it is, the operator is required to look away from the video and towards the iBall interface, although this is much less of a distraction than looking away from the video to locate the correct fader when creating a mix on faders. With further development, it may be possible for the operator to know the location of their touch on the pitch through the use of suitable haptic feedback, removing the need to look at the interface at all. For example, the device could vibrate when the touch left the boundary of the pitch, or when the operator crossed any of the painted lines on the pitch. A similar approach to tactile interaction has been presented in [8].

A more advanced improvement might be to completely remove the human element from mixing the pitch microphones by automatically tracking the location of the ball on the pitch and feeding the location into the iBall system. Commercial systems exist to perform this task, such as Fraunhofer's RedFIR [9], and research in motion tracking and object detection and identification is ongoing, such as in [10]. While this would ensure that all ball specific events were captured, by removing the human element the amount of creative input is reduced. It should also be noted that the desired audio from the pitch does not always follow the ball; there may be events occurring at the other end of the pitch that need to be captured. There may also be multiple locations of interest. Some form of human input into the process is thought necessary, which is why iBall seems a suitable compromise between reducing operator's mental involvement, while still allowing creative control.

The system as it stands provides a solid foundation for further development and testing, as proven by the positive nature of anecdotal comments and encouraging test results.

10. ACKNOWLEDGEMENTS

This work was supported by BSkyB and Telegenic. The authors wish to thank them for their assistance in capturing appropriate source material.

11. REFERENCES

[1] Bell, M. (2011) Audio for HD & 3D - UEFA Champions League Final. Calrec Audio [on- line]. Available from http://www.calrec.com/uk/DocumentsLibrary/Champions%20League.pdf [Accessed 12 May 2012].

[2] Cengarle, G., Mateos, T., Olaiz, N., and Arumí, P. (2010), A New Technology for the Assisted Mixing of Sport Events: Application to Live Football Broadcasting. In: Proceedings of the 128th Conv. Audio Eng. Soc. London, UK.

[3] Lossius, T., Baltazar, P., De la Hogue, T. (2009), DBAP - Distance Based Amplitude Panning, In:

Proceedings of 2009 International Computer Music Conference (2009)

[4] Pulkki, V. (1997), Virtual Sound Source Positioning Using Vector Base Amplitude Panning, In: J. Audio Eng. Soc., Vol. 45, No. 6, pp. 456-466, 1997 June

[5] Kostadinov, D., Reiss, J. D. and Mladenov, V. (2010). Evaluation of distance based amplitude panning for spatial audio, In: Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal (ICASSP), Dallas, March 2010

[6] International Telecommunication Union (2003), Multiple Stimuli with Hidden Reference and Anchor, ITU-R BS. 1534-1, 2003.

[7] Oldfield, G., Shirley, B. G. (2011), Automatic Mixing and Tracking of On-Pitch Football Action for Television Broadcasts. In: Proceedings of the 130th Conv. Audio Eng. Soc. London, UK.

[8] Merchel, S., Altinsoy, E. and Stamm, M. (2010), Tactile Music Instrument Recognition for Audio Mixers. In: Proceedings of the 128th Conv. Audio Eng. Soc. London, UK.

[9] Fraunhofer IIS, RedFIR, Available from http://www.iis.fraunhofer.de/en/bf/ln/referenzprojekte/redfir.html [Accessed 20th October 2012].

[10] Taj, M., and Cavallaro, A., (2009), Multi-camera track-before-detect, In: ACM/IEEE International Conference on Distribured Smart Cameras (ICDSC), Como, Italy, 30th August - 2nd September 2009.