Research White Paper

WHP 185

September 2010

Enabling and Enriching Broadcast Services by Combining IP and Broadcast Delivery

Mike Armstrong, James Barrett & Michael Evans

BRITISH BROADCASTING CORPORATION

BBC Research White Paper WHP 185

Enabling and Enriching Broadcast Services by Combining IP and Broadcast Delivery

Mike Armstrong, James Barrett & Michael Evans

Abstract

This paper explores the opportunities and challenges of supplementing television broadcast channels with additional content using IP delivery. It looks at the way in which the additional material could be resynchronised with the broadcast content and what level of accuracy is required by different types of material. It focuses on the use case of an alternative soundtrack to provide improved intelligibility for viewers who have difficulty understanding speech when presented with background sound. It then goes on to describe our demonstration system and discusses the opportunities for further research into hybrid delivery and the way it could enable a richer broadcasting landscape.

This document was originally published as an IBC 2010 Technical Paper

Additional key words: audio, alternative audio, clean audio, hard of hearing.

White Papers are distributed freely on request.

Authorisation of the Head of Broadcast/FM Research is required for publication.

ENABLING AND ENRICHING BROADCAST SERVICES BY COMBINING IP AND BROADCAST DELIVERY.

Mike Armstrong, James Barrett & Michael Evans

BBC R&D, UK

ABSTRACT

This paper explores the opportunities and challenges of supplementing television broadcast channels with additional content using IP delivery. It looks at the way in which the additional material could be resynchronised with the broadcast content and what level of accuracy is required by different types of material. It focuses on the use case of an alternative soundtrack to provide improved intelligibility for viewers who have difficulty understanding speech when presented with background sound. It then goes on to describe our demonstration system and discusses the opportunities for further research into hybrid delivery and the way it could enable a richer broadcasting landscape.

INTRODUCTION

At IBC 2009, NHK STRL and BBC R&D jointly demonstrated the delivery of alternative subtitles using IP and synchronised with broadcast DVB content. The work was also presented at ICCE 2010 the following January, Matsumura et al (1). This year BBC R&D have begun to explore the possibilities for other synchronised content delivered over IP and the different challenges there are in achieving synchronisation for different types of content, both within a single IP/DVB hybrid receiver and between two separate devices, such as a TV receiver and a laptop or other mobile device.

An increasingly ageing and diverse population means that broadcasters will be serving an audience with multiple and various accessibility needs. The current model of providing a TV or radio service with all its components in the broadcast DVB multiplex is unsustainable if accessibility enhancements need to become more diverse. DTT multiplexes are already full in the UK and adding extra components to the existing channels would only further degrade the quality of the main services.

In this paper we focus on delivering an alternative audio soundtrack using IP delivery. This application requires synchronisation to the same standard as normal TV sound, with no perceptible lip-sync issues. An audio stream must maintain continuous synchronisation unlike intermittent components, such as subtitles, which can be individually aligned.

Our application is the delivery of an alternative audio track with an alternative mix of speech, music and effects to assist the intelligibility of the speech content. As the audience for such an audio stream may also rely on lip-reading, the maintenance of lip-sync is critical. Other audio services, such as an alternative language or a director's commentary do not have a lip-sync requirement as the speakers are not in vision.

In order to deliver additional content to people with accessibility needs without impacting on the main services, we stream the content using IP delivery and then resynchronise the IP and broadcast components in the home. This paper will look at the spectrum of synchronisation requirements and the needs of different types of component and the methods that might be employed to achieve that degree of accuracy. We are presenting work which is very much still in progress and sharing ideas rather than results.

HYBRID DELIVERY

Broadcasting is a very efficient method of delivering content to a mass audience. It scales well, requiring no greater cost or effort to serve a large audience rather than a small one. Streamed IP delivery, on the other hand, is very cost effective for small audiences, as the infrastructure can be relatively inexpensive when compared with broadcasting. However, costs can rise unsustainably when serving large audiences with AV content. So, combining broadcast delivery of the main programme with IP delivery of personalisation and enhancements for relatively small groups could take advantage of the strengths of the two delivery paths. Furthermore, the delivery of personalisation or enhancements to an individual via a portable device is more practicable using IP delivery. This is because indoor reception of DTT on a small portable device is usually less reliable than domestic wifi reception.

Resynchronisation Scenarios

The end-to-end latency of IP delivery is variable and this will cause problems when trying to resynchronise IP-delivered data with content delivered through a constant latency broadcast chain.

In the case where the broadcast and IP components are recombined in a single receiver, there is the opportunity to use DVB clocks and time stamps to align the IP-delivered component with a high level of accuracy. Provided that the IP-delivered components were accurately time stamped at source and that there is sufficient buffering in the receiver, the two streams can be recombined and displayed with the same accuracy as the components in the broadcast stream. The main disadvantage is the considerable time delay in switching to an IP enhanced service whilst the buffers are filled. The length of buffering required for acceptable performance will depend on the characteristics of the local internet connection and is likely to have to be implemented with some form of adaption to cope with changing network conditions.

Where the DVB content is displayed on a conventional receiver, but the enhancements are delivered to a separate mobile device, the problems immediately become more complex. If the main receiver is a standard DVB receiver then it will lock locally to the incoming DVB signal and will display the content after a delay caused by the MPEG decoder buffer and any video frame reordering. The mobile device meanwhile will have to cope with the variability of the IP delivery path and will have no access to the main receiver's timing. Also, the delay in the broadcast path is likely to be much smaller than the delay encountered in the IP delivery path and the IP delivery delay will be variable, even to the extent of some packets turning up out of order. In the case of pre-recorded programmes the IP-delivered content could be coded and streamed in advance of the DVB content, but the problem of resynchronisation still remains a difficult one

Resynchronisation could be achieved between a DVB receiver and a portable device if the two communicate directly over the local network. In this case the DVB receiver could receive the IP and DVB content and resynchronise them by buffering them, before passing the additional content to the portable device with a suitable time offset. This offset is needed to take account of the decode and presentation time in the portable device in a similar manner to the way a time delay is manually adjusted in an external 5.1 decoder.

Resynchronisation Requirements

Depending on the nature of the additional data being provided, the required accuracy of resynchronisation can vary widely. Content that is sent as discrete units, such as subtitles, can be individually synchronised. By contrast, continuous streams, such as audio, require continuous synchronisation to be maintained and would be disruptive to the viewer if the synchronisation were adjusted during a programme.

The least challenging type of content to synchronise is supporting information such as text and pictures sent to supplement a programme. This kind of content need only coincide with a particular segment in a programme and so the viewer is likely to tolerate a variation in the display time of a few seconds.

More exacting is an audio channel offering live simultaneous translation, where there is noticeable latency in the process of translation, so synchronisation whilst would probably not be acceptable if it was out by more than a second.

Audio description for visually impaired users, whilst not requiring exact lip-sync, does require alignment with the gaps in the main audio so needs to be more or less frame aligned. Similarly a director's commentary track without the main audio does not require lip-sync.

Tighter synchronisation is required for lip-sync. The BBC requires its programmes to be delivered with a soundtrack synchronised to within 10ms (2), and recent work suggests that viewers are most sensitive to loss of sync with higher resolution video, Mason & Salmon (3). Even tighter synchronisation is necessary if audio phase alignment is required between, say a stereo audio and a separate surround component.

Table 1 shows the range of synchronisation challenges and some candidate mechanisms for achieving the required level of accuracy. For example, where the signals are all encoded together and share a common clock reference, such as DVB PCR/PTS, then it is possible to achieve the same level of sync as if they had been delivered together, albeit with increased buffering requirements. Harder still is the case where the additional content is decoded in a separate device, such as a laptop, mobile phone or media player, as might be the case if only one member of a family group wishes to hear audio description. In this case the second device needs some way of timing itself against the main receiver – a significant challenge given the diverse nature of many handheld and portable devices.

Required sync accuracy	**10s – 1s**	**1s – 0.1s**	**0.1s – 10ms**	**10ms – 1ms**	**1ms – 0.1ms**	**0.1ms - 10µs**
Otherwise known as...			Frame Sync	Lip Sync		Audio Phase Alignment
Type of content	Fact sheets Photographs Web links	Simultaneous translation Director's commentary	Pre-recorded subtitles Audio Description	Main sound		Difference or additional channels for surround sound
Timing methods	UTC DVB TDT		DVB PCR & PTS			Audio specific

Table 1 – the spectrum of synchronisation challenges

INTELLIGIBILITY AND BROADCAST SOUND

The UK Office for National Statistics projects that over the next 20 years the number of people aged over 60 will nearly double (4). Whilst many people of this age and older can still have what is regarded as normal hearing, their ability to understand speech is significantly degraded when the speech has been distorted in some way or mixed with background noise, Bergman (5). This degraded ability to hear individual words causes the person to rely more and more on contextual cues to understand what is being said. The reliance on context means that the listener uses more of their working memory and has to cope with an increased cognitive load. As a result older listeners are less able than young adults to remember what they have heard when speech is presented against background noise, Pichora-Fuller (6). Therefore, over the next couple of decades, a growing proportion of the UK's population will have difficulty with the intelligibility of speech in television programmes, particularly where it is presented against a background of music and/or noise.

There is a popular perception that the level of background in TV programmes has become more of a problem in recent years, for example Burrell (7). However, the broadcasters' research work into this issue can be traced back at least 20 years, through a study published as a Research Department Report in 1991, Mathers (8). This particular study was largely inconclusive in its results, but the author tentatively suggested that a 6dB lowering of background sound levels might be both acceptable to most people and useful for people with hearing loss. Subsequently the ITC Clean Audio Project looked at intelligibility, but this work was predicated on the use of 5.1 sound and the use of the centre channel as a dedicated speech channel, Shirley et al (9). The tests lowered the level of the left and right channels, relative to a speech only centre channel, and did not make any use of the rear channels. It showed that lowering the level of the left and right channels by 6dB gave a statistically significant improvement in reported dialogue clarity for hearing impaired listeners and a greater benefit in providing centre channel only. This is similar to the results from Mathers. Generally both pieces of work support the benefit of lower background sound levels for hard of hearing viewers but give no guidance on relative music and background loudness. More recent work by NHK has investigated the measurement of the relative loudness of speech and background, Komori et al (10). They have built prototype hardware to measure the relative loudness and provide an indication as to when the balance may cause problems for elder listeners.

Other research investigating the effectiveness of processing television sound in the receiver to improve intelligibility has failed to show any significant improvement in clarity or intelligibility over the original sound when tested on the target audience, Carmichael (11), Uhle et al (12). This is likely to be as a result of the processing introducing distortions into the speech which affect older and hard of hearing subjects, but do not affect younger adults with normal hearing.

The BBC reissued its guidance to programme makers on the needs of viewers with hearing loss in 2007 (13), and continually monitors the public reaction to background sound in its programmes. Research is currently ongoing to improve the BBC's understanding of the range and nature of the issues impacting on intelligibility in its programmes. Cost-effective, intelligible programmes should be built around a primary soundtrack which is acceptable to the widest possible audience. The provision of one or more additional soundtracks would then enable intelligibility for people with particular needs.

DEMONSTRATION SYSTEM

The system we have implemented illustrates the possibility of the remote provision of an alternative soundtrack using IP delivery, see Figure 1. This approach does not require the re-engineering of the main DVB broadcast coding and multiplex infrastructure, but it does come with some caveats about achieving the level of synchronisation accuracy required for good lip-sync.

The main AV services are coded and multiplexed as normal in a conventional DVB output chain and transmitted via DTT or DSAT. A feed of the DVB stream, either off-air or cabled, is then used to recover the programme clock reference (PCR) from the DVB stream. This clock is then used as the timing reference for a separate audio playout system. This reference clock, combined with a time offset to account for coding delays, is used to produce presentation time stamps (PTS) to add to the alternative audio before it is streamed using real-time transport protocol (RTP) over the IP delivery network. The accuracy of the time offset and stability of the coding delays will have an impact on the timing accuracy that can be achieved.

The receiver then obtains the broadcast service off-air and the alternative audio via an internet connection. The receiver buffers both the streamed audio and the DVB service to overcome the variability of the IP delivery path. It then allows selection between either the main audio or the alternative streamed audio before presenting it to the viewer.

This system could also handle other types of alternative content such as subtitles, interactive content or even alternative video content.

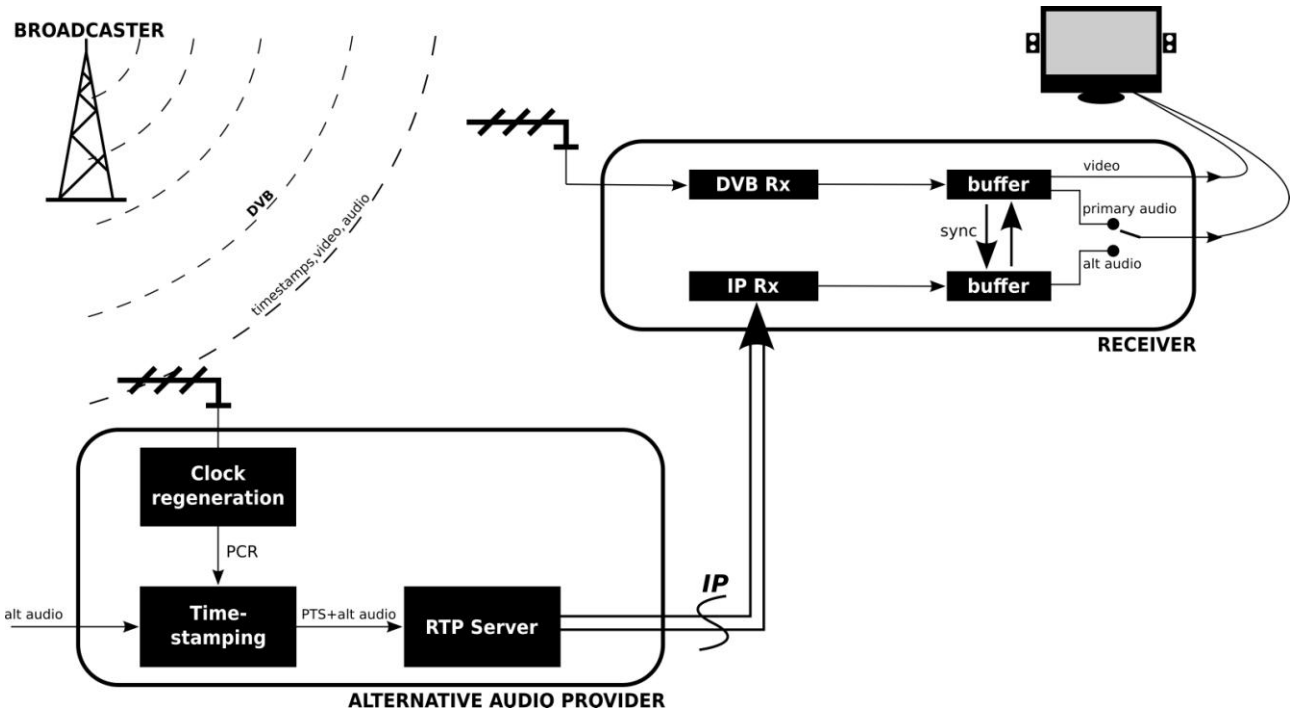

Figure 1 Block diagram of hybrid DVB/IP delivery system of alternative audio via IP

PRELIMINARY RESULTS - SYSTEM PERFORMANCE

Our demonstration system's receiver substitutes the alternative audio transport stream packets for the main audio before the programme is passed to the decoder. With this system, since the decoder's pre-existing ability to synchronise audio is employed for the actual playback, there is no chance of audio synchronisation drift. This means that the results are binary: either the audio is inserted correctly in the right time window, in which case the audio is synchronised correctly, or the audio is inserted late (or early) in which case no audio is decoded. It is possible under some circumstances for some audio data to be inserted at the correct time, and for other packets to be inserted too late or dropped entirely. In this case the audio track will be in-sync one moment, then disappear entirely leaving silence, before it returns again.

To regulate the insertion of the alternative audio track the demonstration software uses a number of different parameters which can be adjusted to alter performance:-

The timing window used for the insertion of audio packets into the stream is modifiable. The insertion window governs how much earlier than the expected PCR timestamp an audio packet can be placed in the stream (the system never inserts a packet late). With experimentation it was determined that a window with a size of 0.1s was the largest usable size for this value. Any larger and the audio packets would often be discarded by the decoder (VLC in this case) before they were due to be decoded; whilst a smaller value would often result in packets being dropped as they could not be inserted in time.

The IP stream buffer exists to store auxiliary packets which have arrived earlier than is required for insertion to take place. Currently it is set to the same size as the primary buffer, but is larger than necessary for audio.

The DVB transport stream buffer in the receiver must always be long enough to contain the complete length of the overall system delay and contain enough extra packets for the code to efficiently process in one go. It was found experimentally that a DVB buffer length around 0.2s longer than the overall system delay was sufficient for these purposes. The overall system delay itself exists in the set-up to allow for three factors:

a) The delay between the receipt of a timecoded packet by the alternative audio streamer (from the DVB transmission) and the transmission, via IP multicast, of a packet by that machine with the same timecode. This delay can, in fact, be negative when working with pre-prepared content files. In our demonstration system it is positive but is generally reasonably short, less than one 1s in all cases. In some potential use-cases this delay might be quite large.

b) The delay between the transmission of an auxiliary audio packet and its reception by the receiver. This is primarily caused by network latency, and hence is variable. In our demonstration it is always very small, but could be significantly longer.

c) The processing delay caused by the insertion of the alternative audio packets in the transport stream. This is generally very small.

Taking these factors into account our choice of two seconds overall delay in the system works well in the laboratory. In practice it is possible that a significantly larger delay could be needed. This would occur if the content from the auxiliary source is not pre-prepared and so needs to be transformed in some way before being transmitted. In these cases some method of adapting the overall delay to different scenarios will be required.

DISCUSSION

The demonstration system above clearly has many more aspects that need exploring, from buffering strategies to deal with IP network performance, to recovery from signal loss of either stream and the acceptability of the latency involved in changing channel. There is also the need to carry out user testing on some of these issues, and that will require the addition of a suitable user interface and a library of test material.

The practicality of the origination of alternative audio components needs to be explored. It might be possible, for example, to provide a simultaneous translation of a TV programme using IP delivery directly to the home from a location remote from the broadcaster. This could open up the provision of such services to third parties.

The system could also be tested streaming a variety of other forms of content such as a second video to be overlaid on the main video. This could be used to demonstrate a closed signing feed or show audience or commentator reaction shots.

Before the IP delivery of a soundtrack with increased intelligibility could become practicable and affordable, the origination of this material needs to be cost effective. Unlike the main access services, subtitles and audio description, which add an extra component to the programme, the production of a high-intelligibility audio track would require the remixing of the programme sound, thus requiring an additional production process. For example, for a music-free version of an episode of “The Nature of Britain”, the production team added narration to fill the gaps left by the lack of music, effectively creating a new version of the programme. Clearly, tools to assist in the generation of alternative soundtracks in a semi-automated manner would be an advantage. More radically, tools which enabled the delivery of separate audio “objects” along with mix-down parameters, could enable appropriate mixes to be derived at the receiver tailored to the individual’s needs, in the same way that style sheets can be used to tailor the presentation of appropriately coded web pages. Such “object orientated’ presentation of sound could provide further possibilities for customisation and interactive presentation of programmes.

The example we have used is based on the resynchronisation of the two signals in the receiver. If the signals can be delivered to separate devices, then it opens up the possibility of streaming different signals to individuals watching the same programme. Anecdotally some families watching the same TV are already using laptops or mobile phones at the same time. So it is quite reasonable to envisage using the laptop or phone to deliver associated content to individual viewers and for the benefit of a wider range of viewers. For example, if only one member of the family wants to watch a TV programme with audio description the AD could be delivered to that person via a mobile phone or audio player. The synchronisation issues are more complex for a separate device, and there are no existing conventions for the way the two devices should interact. This is a rich area for research, both from the technical challenge of synchronising presentation on multiple devices and from the way in which the devices should interact with the users.

Finally there are methods which might be used to recover from lost IP packets. As the system currently exists no attempt is made to recover lost packets at all, any lost packets will result in audio drop-outs, and potentially an undesirable experience. The most common means of dealing with lost packets on an IP system is to make use of TCP to ensure that the system will request the retransmission of any packets which were lost in transmission. However, with multicast transmission this technique is not practicable. This application would seem to be an opportunity to make use of Forward Error Correction codes. The processing time required to do this is likely to be small enough not to affect the system’s performance and it is anticipated that such a system should enable the recovery from even quite significant packet loss.

CONCLUSION

We have discussed the issues faced when attempting to resynchronise television service components sent via diverse paths, and the spectrum of different temporal requirements for a range of different types of content. We have highlighted the particular need for improved intelligibility in soundtracks for a growing part of the television audience, and used the example of providing an alternative soundtrack for this audience using IP delivery. Hybrid IP/broadcast delivery opens up many different possibilities for broadcasters and third party service providers to enrich the viewing experience and enable more diverse provision of future access services.

REFERENCES

1. Matsumura, K., Evans, M.J., Shishikui, Y. and McParland, A., 2010. Personalization of a Broadcast Program using Synchronized Internet Content. International Conference on Consumer Electronics. January, 2010.
2. BBC Technical Standards for Network Television Delivery, Version 01.13. November 2009. Available from http://www.bbc.co.uk/guidelines/dq/pdf/tv/tv_standards_london.pdf
3. Mason, A. and Salmon, R., 2008. Factors affecting perception of audio-video synchronisation in television. 125th Convention of the Audio Engineering Society paper 7518. October, 2008.
4. Office for National Statistics, 2009. National populations projections, 2008-based. 21 October 2009. Available from http://www.statistics.gov.uk/pdfdir/pproj1009.pdf
5. Bergman, M., 1971. Hearing and Aging: Implications of Recent Research Findings. International Journal of Audiology. 1971, Vol 10, No. 3, Pages 164-171.
6. Pichora-Fuller, M.K., Schneider, B.A. and Daneman, M., 1995. How young and old adults listen to and remember speech in noise, J. Acoust. Soc Am. 97, 593-607, 1995.
7. Burrell, I., 2009. Great drama – but can you hear a single word they are saying? The Independent. 1st June 2009. Available from http://www.independent.co.uk/arts-entertainment/tv/features/great-drama-ndash-but-can-you-hear-a-single-word-they-are-saying-1693835.html
8. Mathers, C. D., 1991. A Study of Sound Balances for the Hard of Hearing. BBC Research Department Report 1991/3. 1991. Available from http://downloads.bbc.co.uk/rd/pubs/reports/1991-03.pdf
9. Shirley, B., and Kendrick, P., 2004. ITC Clean Audio Project. 116th Convention of the Audio Engineering Society. May 2004.
10. Komori, T., Takagi, T., Kurozumi, K. And Murakawa, K., 2008. An Investigation of Audio Balance for Elderly Listeners using Loudness as the Main Parameter. 125th Convention of the Audio Engineering Society. October 2008.
11. Carmichael, A.R., 2004. Evaluating digital “on-line” background Noise suppression: Clarifying television dialogue for older, hard-of-hearing viewers, Neuropsychological Rehabilitation: An International Journal, 1464-0694, Volume 14, Issue 1, 2004, Pages 241 – 249.
12. Uhle, C., Hellmuth, O., and Weigel, J., 2008. “Speech enhancement of movie sound”, 125th Convention of the Audio Engineering Society, paper 7628, October 2008.
13. BBC Editorial Policy Guidance Note, 2007. Television Viewers with Hearing loss. Available from http://www.bbc.co.uk/guidelines/editorialguidelines/advice/viewerswithhearingloss/