Dialogue Enhancement

— technology and experiments

Harald Fuchs
Fraunhofer IIS

Simon Tuff
BBC

Christofer Bustad
Swedish Radio

Finding the right balance between dialogue and ambient sound is a major challenge for sound engineers, and an increasingly common cause of audience complaints. It may therefore be desirable for audiences to influence the audio balance to suit their personal choice, their listening environment and their hearing. A new "Dialogue Enhancement" technology from Fraunhofer IIS addresses this challenge. It is backwards-compatible and adds only a moderate bitrate overhead to the transport bitstream, and is thus well suited to broadcast applications.

This article gives an overview of the underlying technology, the necessary changes required in the production chain and in the broadcast signal. An experiment at the Wimbledon Tennis Championship was used as an example of how this technology could be applied and the results of this experiment, including initial audience reactions, are shown. The article also looks at the next step in Dialogue Enhancement — a comprehensive trial at Swedish Radio in autumn 2012, using a specially-developed smartphone app.

The recent debate about loudness has drawn attention to the quality of broadcast audio. In fact, a significant number of user-complaints are related to audio. Besides the loudness differences between programmes and commercials, many complaints are also about the balance between dialogue and the ambient background "noise" [1]. For example, background music might be too loud relative to the dialogue, resulting in some audience members having difficulty understanding the presenter or actor. Similarly, in sports coverage, crowd noise may drown out the commentary.

This may especially be an issue for people with hearing impairments. In Europe, some 16% of the population have a hearing loss that calls for treatment [2]. However, only around one third of those with a medically-indicated hearing deficiency use hearing aids. For people with a mild hearing loss, this percentage is even lower – for instance below 10% in Germany and France [3]. Because hearing loss occurs gradually, this might mean that many of these individuals are unaware of their impairment. It is this latter group that would benefit most from a different balance between the dialogue level and the background level.

The same is true for fluent non-native speakers. Listening to programming in non-native languages typically requires more concentration. A higher loudness of the dialogue versus background would make listening less straining and help to improve intelligibility. Tests have showed that a 3 dB higher

SNR (approximately) is necessary to enhance the intelligibility to the level of the native language [4].

Further, dialogue intelligibility may not only be an issue for people with hearing impairments or non-native speakers, but could also be exacerbated by the listening environment or the reproduction equipment used. The listening environment has a considerable influence on the preferred setting of the mix. For instance, the audibility of a movie watched in the noisy environment of say an airport using headphones would benefit from a different balance to the one that would be optimal for viewing in the quiet environment at home with an excellent stereo or multichannel system.

These dialogue intelligibility problems are not addressed by the new loudness measurement techniques which are currently being implemented by broadcasters. To provide a solution, Fraunhofer IIS has developed a technology called “Dialogue Enhancement”, which allows the user to change the balance between ambience and dialogue according to their preferences. It was first tested by Fraunhofer IIS and the BBC during the 2011 Wimbledon Tennis Championship and will be further tested in the ongoing Swedish Radio project “Audibility in Radio” later this year.

Technology

The Dialogue Enhancement technology enables the highly bitrate-efficient transmission of individual audio sources whilst retaining compatibility with mono, stereo or 5.1 mixes. The basic idea is to describe the different components within a mixed audio signal in a way that enables the receiving device to change the balance by, for example, enhancing or attenuating the dialogue level in relation to all other signals contained in the mix.

Examples of audio sources could be a commentator’s voice, the atmosphere inside a sports stadium, or dialogue, music and effects in a feature film or TV drama.

Basic principle

The Dialogue Enhancement encoder *(Fig. 1)* analyses the input signals and produces a single mono, stereo or 5.1 mix of all those signals. In addition, the encoder generates parameters, which describe the relation of each source signal to all other sources. This description is generated in a time- and frequency-selective manner.

The mix of the input signals can be produced automatically or controlled externally by a sound engi-

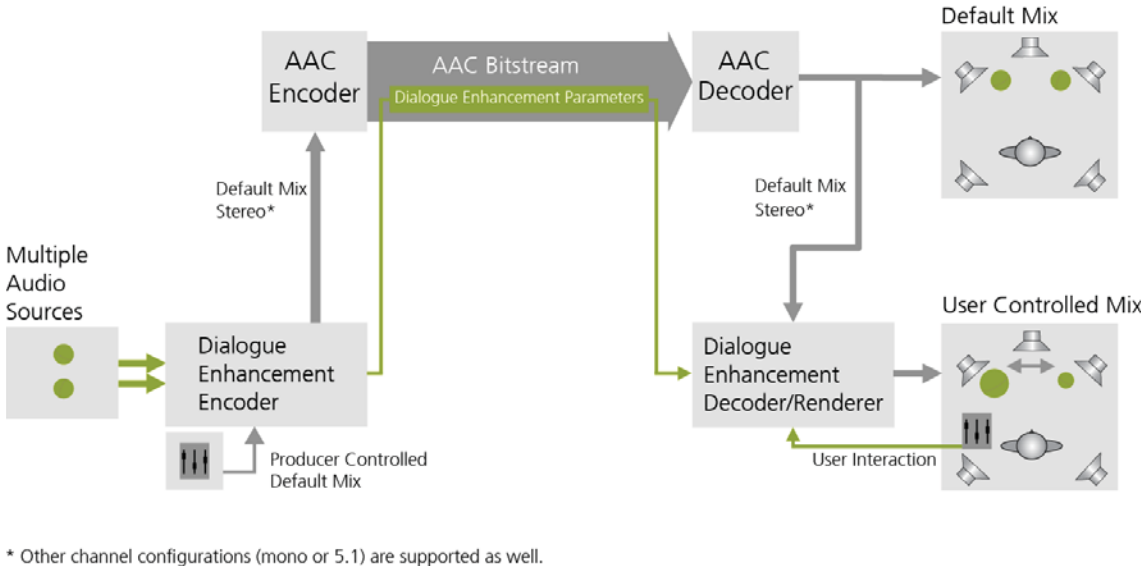

Figure 1
Block diagram of the Dialogue Enhancement system

neer. The mixed signal is encoded with an audio codec such as MPEG-4 AAC or HE-AAC. The stream of parametric side information is embedded into the encoded audio bitstream.

The transmission of the mix plus side information is substantially more bitrate efficient when compared to a separate transmission of all single sources, because each audio source representation in the parametric domain only slightly increases the overall bitrate.

On the receiving side the audio bitstream is decoded, then the Dialogue Enhancement decoder takes the decoded mix signal and uses the descriptive data from the parameter bitstream to enable access to the audio sources.

The user is then able to adjust the volume of each source individually, e.g. to improve the intelligibility of the dialogue or a sports commentary.

The technology is completely compatible with existing transmission and playback equipment. Legacy devices that are not capable of decoding the parametric side information will play back the default mix signal and ignore the side information.

Real-time encoder implementation

For testing purposes the encoder is implemented as a real-time application on a PC platform. The encoder application integrates the Dialogue Enhancement encoder and audio encoder functions as described in the previous section. It performs the following tasks:

- analyses both the source signals (e.g. the first source could be the dialogue or speech signal and the second source could be the ambience or music signal);
- creates parametric side information;
- creates the stereo downmix of both input signals;
- encodes the downmix into an MPEG-4 AAC bitstream;
- embeds the side information into the ancillary data of the AAC bitstream;
- packetizes the bitstream in an Audio Data Transport Stream (ADTS) format.

Streaming player

A software PC streaming client is used for audio playback. This player includes a Shoutcast client and an AAC decoder as well as the Dialogue Enhancement decoder and renderer, to influence the mix of both sources.

Controlling the mix is made available through an on-screen slider element in the graphical user interface. The zero position is equivalent to the default mix, while positive values correspond with enhanced commentary (or speech in general) and negative values with attenuated commentary, i.e. louder court sound (or ambience or music).

As described above, encoding and transmission is backwards compatible: every Shoutcast-compatible client (e.g. Winamp or VLC player) is able to receive the stream and decode the default downmix while ignoring the side information.

Abbreviations

AAC	Advanced Audio Coding	**LTE**	Long Term Evolution (4th generation mobile networks)
ADTS	Audio Data Transport Stream		
DTS	Digital Theatre Systems	**OB**	Outside Broadcast
HE-AAC	High Efficiency AAC	**RDS**	Radio Data System http://www.rds.org.uk/
HTTP	HyperText Transfer Protocol		

BBC Radio 5 Live – the Wimbledon experiment

Background

The All England Lawn Tennis Club's Championship is held annually at Wimbledon, in south-west London and its broadcast coverage is one the BBC's most important annual sporting commitments. 2011 saw coverage on the BBC's UK TV channels in both SD & HD with 5.1 surround sound audio as well as radio coverage on Radio 5 Live (R5L), Radio 5 Live Sports Extra (R5LSE) and BBC World Service. The Dialogue Enhancement experiment was an integral part of the R5L & R5LSE coverage and was accessed by audiences via the Radio 5 Live website. Because much of the live coverage of Wimbledon takes place during the working day, there was an expectation that this online experiment would appeal to office-working tennis fans who had access to a networked computer, but who didn't have access to normal broadcast signals. The name "NetMix" was chosen for this experiment, as it was thought to be effective shorthand for describing what the experiment offered (as well as being a title unused in Europe).

Implementation at Wimbledon

Although much of the cabling for the radio infrastructure at Wimbledon is a permanent part of the venue, all the main technology building blocks for broadcasting have to be brought on-site, installed and tested during the week before the Championship. As Radio 5 Live is primarily a medium-wave (AM) sports and news service, the radio part of this infrastructure is mono. The experiment covered the main matches from Centre Court during the second week of the Championship.

This allowed the Fraunhofer and BBC team to use the first week of play to conduct some tests and fix last-minute issues as well as to increase audience awareness of what would be available in the second week of the Championship *(more on this below)*.

Two audio feeds were required for the NetMix experiment: the first was a feed of Centre Court ambience (stereo FX) and the second was a feed of the Radio 5 Live commentary (in mono). Stereo was used for the court ambience to further enhance the listening experience over the standard mono online offering for those listeners that choose to attenuate the commentary signal. The court stereo FX feed was derived from a coincident crossed pair of microphones attached to the umpire's chair.

Because this court FX feed was only used by the experiment, it could be routed directly to the NetMix operator's position.

The commentary was taken from the main Radio 5 Live feed from the Centre Court commentary box and was routed to NetMix prior to the main Radio 5 Live commentary group fader. To create and monitor the two sources (court FX and commentary), NetMix required its own sound supervisor for a number of reasons:

1) To make sure that the signals were editorially correct and that off air microphones or other crowd noises were not put on air (the usual practice at Wimbledon is to leave box microphones live for talkback purposes).

Photo 1
A crossed pair of Neumann km100s (along with other BBC paraphernalia) mounted on the umpire's chair on Wimbledon centre court

2) To provide a feed that was Dialogue Enhancement encoded even when coverage wasn't live, so that the listener's fader always demonstrated some cross-fading effect. This meant replacing commentary feeds with sustaining announcements when the commentary box was not on-air.

3) To fade up these sustaining feeds (either R5L network or recorded announcements) and balance them with the court ambience at times when match commentary was not taking place. This arrangement was used before games had started and also when commentary was suspended for news bulletins and other contributions from the BBC's West London Studios at Television Centre.

4) To balance the two feeds (or sources) of court ambience and commentary for the default mix of the encoded output stream. Although they were not balanced in the usual way by a sound supervisor because this role was now handled by the listener, the feeds still required lining up at the start of a match, as well as adjustments to the relative levels as commentary and/or the crowd became more animated during the course of a match.

Photo 2
A view of the "NetMix" operator position showing the sub mixer, the shortcuts and instructions to leave the sustaining feed faded up when the court was off air

The additional sound supervisor position was located at the back of the Radio 5 Live control room in the basement of Wimbledon's Media Centre. This position was equipped with an audio sub-mixer, 360 Systems Shortcuts (to provide off-air announcements), an audio talkback box, plus screens showing scores, cues and court action.

The output of the NetMix operator position then fed the two sources of stereo court FX and mono commentary through to the Fraunhofer encoding position in the online operations room next door.

The encoder was configured in the following way for the experiment:

- Adaptation range: ±12 dB. The audience was able to enhance or attenuate the commentary by 12 dB compared to the downmix.
- Bitrate: 192 kbit/s was used for the encoded audio stream (AAC encoded downmix and parametric side information embedded into the AAC bitstream).
- Bitstream format: ADTS on HTTP for Shoutcast-compatible streaming.

The packetized output stream of the encoder was pushed over an HTTP connection to the StreamUK content distribution network. StreamUK was instructed to geo-lock the service to UK users only because of the constraints imposed on the BBC rights to the Championship.

It is worth noting that the encoders were shipped to the UK from Fraunhofer in Erlangen, Germany, and the whole installation was set up and tested by Fraunhofer and the BBC Outside Broadcast

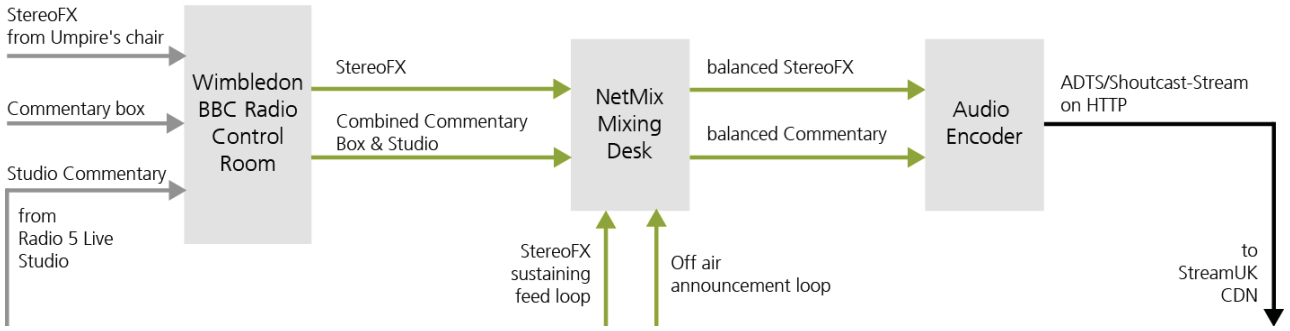

Figure 2
Signal and system diagram showing the routing of NetMix to the listener

team at the BBC's Maida Vale Studios some weeks before being moved to the Wimbledon Media Centre.

NetMix Player

A special PC software player *(Fig. 3)* was made available for download during the experiment.

The scale of the slider element between "–3" and "+3" was used for the ±12 dB adaptation range to enhance or attenuate the commentary.

Figure 3
Control slider NetMix player

This scale was also used in the survey where the audience was asked about their preferred mix, i.e. the position of the mixing slider (see Results section below). The zero position was equivalent to the default mix, while positive values correspond with enhanced commentary and negative values with attenuated commentary (i.e., louder court sound).

Implementation online

The Radio 5 Live website had a page specially designed and added to support this experiment, which was intended to achieve a number of things:

1) To explain to the audience what the experiment was trying to achieve and how it should work for them.
2) To link to the Fraunhofer player download page and the end-user licence agreement.
3) To collect audience feedback on listener reactions to the experiment. This was done via a link to a specially prepared survey and by providing a blog where listener questions were answered

Figure 4
The Radio 5 Live web page for NetMix showing its link to the Fraunhofer download site and the NetMix player

by the technical team led by Rupert Brun, the BBC's Head of Technology for Audio & Music.

4) To provide on-air promotion by the commentary and Radio 5 Live teams. A successful promotional dialogue was also conducted using the hash tag "#NetMix" on Twitter.

Results

Both the Radio 5 Live website and the NetMix player provided a button to link to the online survey (as shown above) that was prepared and analysed by BBC Marketing and Audiences in conjunction with the company eDigitalResearch. As is often the case with embedded surveys, the response was not large. About 1200 NetMix players were downloaded, and 98 listeners completed the survey. The detailed results are not published here, but some high-level findings are worth reporting:

1) Over 85% of the respondents were male.
2) Over 37% were from London and the South East of England.
3) The majority of listeners claimed they were Radio 4 or Radio 5 regulars, and the largest group (21%) said they had heard about the experiment from web content and blogs.
4) Although we were expecting a significant proportion of the audience to be at work, in fact over 75% listened from home. We suspect that the difficulties of connecting the NetMix player via manual configuration through corporate firewalls and proxies may well have significantly reduced the number of listeners who could successfully connect from work.
5) Over 72% of the listeners agreed or strongly agreed that this kind of technology would benefit radio, and 84% thought the same for TV.

The last interesting finding was that not all the listeners chose to boost the commentary over the court ambience, and in fact the survey data of preferred "fader" positions shows that the listener preferences were fairly evenly split, producing a double top result as shown in *Fig. 5*.

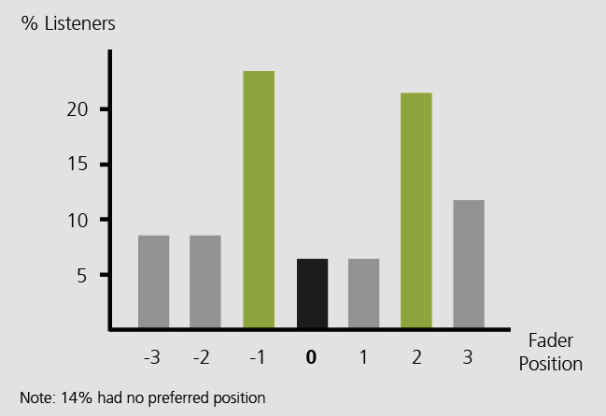

Figure 5
Percentage distribution of listener-preferred fader positions showing a double top, one on each side of the "studio balance" position (= "0" position)

This was a complex and challenging experiment requiring input from many teams including production, OBs, engineering, online, legal, sports rights and audience research. The experiment demonstrated that Dialogue Enhancement coding can provide a viable technology to address audibility issues in a way from which audiences can both understand and benefit.

Although this brought extra complexity to the production process, with thought and planning at a relatively early stage, this was manageable.

Although no firm conclusions about the future of this approach have yet been made, the BBC is keen to improve audibility for listeners in a cost-effective way and to discuss the possibilities open to the broadcasting industry.

Swedish Radio project on audibility in radio

Background

The most common source of complaints from listeners to Swedish Radio's different radio channels is that music backgrounds and sound effects reduce the speech intelligibility. This is particularly common for the channel P1. This is Swedish Radio's talk radio channel for in-depth news analysis,

current affairs and debate, and it also serves as a forum for drama and documentaries and other programmes covering the arts, the sciences and social and philosophical issues. For those in noisy environments or with a hearing impairment, having music backgrounds or sound effects mixed with the speech may cause significant difficulties in hearing what is being said. Listeners both with and without hearing impairments complain about this.

The second most common source of complaints concerns the speech-to-music balance, particularly for the channel P2 which broadcasts mainly classical and contemporary music as well as jazz and folk music. People listening in a car or in other noisy environments may want the speech to be extra loud. Those listening at home via a hi-fi system may want the speech at a lower volume.

Different technical possibilities have been tested over the years to address these issues. The Radio Data System (RDS), which is a digital data stream on a subcarrier in the FM system, includes the possibility to set a speech/music flag. Some broadcasts from Swedish Radio have included this flag, but only some RDS receivers today have the functionality to adjust the listening level based on this flag. There are also difficulties on the production side in setting this flag correctly, particularly for recorded material. The use of this flag could solve the speech-to-music balance problem when the speech is not mixed with music or sound effects. Transmitting the speech on an audio track that is separate from the music and sound effects would solve both problems, if listeners have the possibility to adjust the audio level of the speech track. Trials with 3.0 audio using Dolby Digital and DTS – with only speech in the centre channel and music and sound effects in the left and the right channels – have been conducted at Swedish Radio. Also, Swedish Television have regular broadcasts with 3.0 audio. These audio streams are, however, not automatically backward compatible with stereo reproduction and may not be very efficient (bitrate-wise) for reproduction in stereo. Backward compatibility to stereo reproduction would be very advantageous.

The possibility to adjust the speech level in a backward compatible and bitrate efficient way, even when there is a mix between the speech, music and sound effects, would be very beneficial for listeners. This would be in line with the philosophy that the listener should have the possibility to adjust and control things as much as possible and create the radio listening experience that he or she wants. The production at Swedish Radio would set the default speech-to-music balance, and listeners would then be able to adjust this balance with an offset. These possibilities can increase the speech intelligibility for those with hearing impairments, in noisy environments, and with a different mother tongue than the language being spoken. These possibilities can also give listeners control over the speech-to-music balance so that it can be adapted to different listening environments and different preferences.

Forthcoming trial

The technology described in this article will be tested at Swedish Radio in a limited trial, in which a smartphone app for the iPhone will be developed and used by listeners to receive the audio and to adjust the speech-to-music balance. This trial and the development of this app is partially funded by the Swedish Post and Telecom Authority (PTS), since this service would be beneficial for everyone and not just for a particular group of listeners. The development of a first version of the app has started and the trial is planned to take place during the autumn of this year.

The aims of the trial are to test this technology, to learn how content can be produced in an efficient way, and to find ways of developing the use of this technology beyond the adjustment of speech-to-music balance. There are however some challenges in preparing for and in conducting this trial.

Client application

The app will be developed in such a way that the first versions can adjust the speech-to-music balance and that future versions may include more advanced audio processing such as filtering and compression. The goal is that the app will be able to receive both live and on-demand audio.

Content production

Content to listen to is also needed and our goal is to offer some live shows and also some on-demand material. For live shows in a studio, feeding the live microphones to a separate output is the easier part, but playback of recorded material so that only the speech is sent to the same separate output is more challenging.

One possibility, as a start, is to send only the live speech to the separate output and to leave the recorded speech in the same output as the music and sound effects. This may however be a source of frustration for listeners. If the speech can be fed to a separate output without the presenter having to do anything extra, this would simplify the introduction and use of this technology.

Sports content with the possibility to adjust the speech-to-ambience balance, as in the Wimbledon trial, would be great to have in this trial, but this may pose some challenges in the infrastructure. Interviews, including ambience from outside of the studio, may be produced using multitrack recorders or by recording the ambience separately before or after the interview.

Shows produced in a multitrack editor can easily be exported to 3.0 audio by doing one downmix where the speech is muted and another downmix where everything but the speech is muted.

Another challenge is how to archive the audio produced for this technology.

Distribution

On-demand material can easily be sent to listeners by making the recorded audio available to listeners through progressive download over standard HTTP. Live audio needs infrastructure that can support more audio channels than two and can feed the audio to an encoder from which streams can be transferred to listeners.

Before and during the trial, suitable bitrates for different scenarios will be investigated. For this we need to be aware of the bandwidth limitations in different networks such as 2G, 3G, LTE etc.

Additional goals for the trial and beyond

During the trial we hope to find ways of developing the use of this technology beyond the adjustment of speech-to-music balance. For those with smaller hearing impairments without hearing aids, doing some suitable filtering may increase the speech intelligibility. Perhaps techniques used in hearing aids can be used.

If the speech is (approximately) separated from everything else in the app, the filtering can be applied only on the speech or differently on the speech than on the music and sounds effects. Separation of the speech from music and sound effects also allows for compression of the speech to increase intelligibility, for example in noisy environments.

It would also allow the use of “ducking” which means that the level of the music and sound effects is lowered only when someone is talking. These temporary level reductions of the music and sound effects can be done by single- or multi-band techniques, and perhaps also in combination with compression of the speech. All of these ideas however need to be tested and evaluated. Some of the ideas may be implemented experimentally in the app used by the listeners.

Beyond this trial, we hope to test the technology on more types of smartphones, tablets and computers. Perhaps adding a separate level control to each person talking could increase the listening experience. Feeding the speech to a separate speaker in a car or home cinema system for example may also increase the speech intelligibility and enhance the listening experience. Speech intelligibility could also be increased by using stretch or pause extending algorithms on the on-demand material.

NAB Technology Innovation Award

Fraunhofer IIS presented Dialogue Enhancement and the Wimbledon experiment as a technology preview at the NAB 2012 Show in Las Vegas, USA. During the show, the National Association of Broadcasters (NAB) selected Dialogue Enhancement for its 2012 NAB Technology Innovation Award [5].

The 2012 NAB Innovation Awards, open to all organizations exhibiting a non-commercialized service or product, was started in 2009 to recognize advanced technology exhibits and demonstrations of significant merit to the NAB Show.

Summary and Outlook

The Wimbledon experiment demonstrated the feasibility of the technology, and although the sample of listeners who responded was relatively small, it did appear to show a strong appreciation of the benefits provided by such a tool. The survey results indicate that the additional requirements on the production side might be outweighed by the benefits to listeners. To prove this, more tests and a detailed analysis of the additional production costs are required.

Fraunhofer IIS is currently in discussion with different broadcasters and content providers about additional experiments to investigate these topics further.

Harald Fuchs received his diploma in electrical engineering from the University of Erlangen, Germany in 1997 and joined Fraunhofer IIS in the same year. He has more than 15 years experience in video coding, audio coding and multimedia systems. From 1997 to 2002 he was a software developer for video codecs (H.263 and MPEG-4 Part-2) and multimedia streaming systems, developing standard-based software solutions for PC and embedded devices.

From 2002 onwards, Mr Fuchs has concentrated on multimedia system aspects (IP protocols, file formats, Digital Rights Management) and standardization. He has actively contributed to several standardization organizations (including MPEG, DVB, OIPF, HbbTV, ISMA and OMA), was co-author of various standard specifications and has chaired working groups in DVB and ISMA. He was author or co-author of several IEEE conference, journal and magazine papers.

Since 2009 Harald Fuchs has focused on audio technologies for broadcasting and broadband multimedia streaming and is now a Senior Business Development Manager for those application areas.

Simon Tuff has worked across most of the BBC's radio services during the last 20 years, including BBC World Service, local radio and the national radio networks. He led parts of the technology work that migrated the BBC's Radio services on to digital workflows along with the launching of five new digital stations.

As well as working on the audio for HDTV, collaborating with the BBC's R&D teams and supporting some of the Corporation's other big technology projects & strategies, Mr Tuff is currently focused on the next generation of transformations and innovations that digital production and digital broadcasting could offer.

Christofer Bustad graduated in 2006 with an M.Sc. degree in Engineering Physics from Uppsala University in Sweden with one year on exchange at Queen's University in Ontario, Canada. He did his master's thesis at Swedish Radio Technical Development in MPEG-1/2 Audio Layer II encoding, and joined Swedish Radio Method and Technical Development in 2006.

Mr Bustad has since then been working in the fields of subjective and objective audio quality, audio coding, loudness, FM processing etc.

The upcoming Swedish Radio trial is an excellent opportunity to collect more user feedback and to get additional data points on the integration into different production environments. It also offers the opportunity to combine different technologies to further improve the listener's experience.

Acknowledgements

The authors would like to thank all participants in the Wimbledon experiment at the BBC and Fraunhofer IIS for their work.

The photographs are reproduced by permission of Simon Tuff & Rupert Brun at the BBC.

The audience data was supplied by BBC Marketing and Audiences, and eDigitalResearch.

References

[1] P. Green: **Loudness Normalization Benefits Study**
Loudness Summit, London, December 2011
http://www.loudnesssummit.com/pdfs/speakers/Phil_Greene.pdf

[2] **http://www.hear-it.org/hearing-loss-in-Europe**

[3] S. Hougaard and S. Ruf: **EuroTrak I: A consumer survey about hearing aids in Germany, France and the UK**
Hearing Review, 2011;18(2):12-28.
http://www.hearingreview.com/issues/articles/2011-02_01.asp

[4] M. Florentine: **Speech perception in noise by fluent, non-native listeners**
Journal of the Acoustical Society of America, Vol. 77, Issue S1, pp. S106-S106, 1985

[5] **http://www.iis.fraunhofer.de/en/pr/presse/2012/april/NAB_Award.jsp**

This version: 15 June 2012

Published by the European Broadcasting Union, Geneva, Switzerland

ISSN: 1609-1469

Editeur Responsable: Lieven Vermaele

Editor: Mike Meyer

E-mail: **tech@ebu.ch**

The responsibility for views expressed in this article rests solely with the authors