TECHNICAL PAPER

DIALOGUE ENHANCEMENT

PERSONAL AUDIO MIX FOR BROADCAST PROGRAMS

Finding the right balance between dialogue and ambient sound within a broadcast program is a major challenge for sound engineers and an increasing cause of audience complaints. Now, with Dialogue Enhancement, each viewer can individually decide what he wants to hear: Ambient sound or dialogue; the stadium atmosphere during a sports game or a commentator's up-dates. Especially for hearing-impaired viewers or non-native speakers this new opportunity will make a great difference as the technology brings an end to the well-known inconvenience of obscured dialogue when the background music's volume becomes too overwhelming.
Dialogue Enhancement from Fraunhofer IIS is the first technology that enables TV viewers and radio listeners to control the audio balance according to their personal preferences, listening environments and hearing abilities. Dialogue Enhancement allows broadcasters to meet the demand for improved speech intelligibility as part of their services for hearing impaired audiences.

Fraunhofer Institute for
Integrated Circuits IIS

Management of the institute
Prof. Dr.-Ing. Albert Heuberger
(executive)
Dr.-Ing. Bernhard Grill
Am Wolfsmantel 33
91058 Erlangen
www.iis.fraunhofer.de

Contact
Matthias Rose
Phone +49 9131 776-6175
matthias.rose@iis.fraunhofer.de

Contact USA
Fraunhofer USA, Inc.
Digital Media Technologies*
Phone +1 408 573 9900
codecs@dmt.fraunhofer.org

Contact China
china@iis.fraunhofer.de

Contact Japan
Fahim Nawabi
Phone: +81 90-4077-7609
fahim.nawabi@iis.fraunhofer.de

Contact Korea
Youngju Ju
Phone: +82 2 948 1291
youngju.ju@iis-extern.fraunhofer.de

* Fraunhofer USA Digital Media Technologies, a division of Fraunhofer USA, Inc., promotes and supports the products of Fraunhofer IIS in the U. S.

1. A GOOD AUDIO MIX IS NOT ENOUGH

Delivering a well-balanced audio mix does not guarantee an enjoyable television or radio program for a broadcast audience. There are several aspects that determine how well listeners understand the dialogue of a broadcast program, explained in the following.

1.1 Listening environment

The listening environment and the reproduction equipment have a considerable influence on how listeners perceive an audio mix. For instance, when listening with headphones, noisy background audio can mask important dialogue, and thus a different balance of dialogue versus background would be beneficial.

1.2 Foreign languages

Listening to programs in non-native languages typically requires more concentration. A higher dialogue volume in comparison to the background levels would make listening less straining and would help to improve intelligibility.[1]

1.3 Hearing abilities

Individuals with hearing impairments benefit from an increased signal-to-noise-ratio (SNR) in order to enjoy and understand broadcast programs equally as well as people with normal hearing abilities. Raising the SNR only by one dB alone achieves a great increase of the speech intelligibility.[2]

Due to different listening situations and hearing abilities, every person in the audience would benefit from their own personal mix to enhance their listening experience. The Dialogue Enhancement technology is a solution for this, allowing the audience to personally tune the audio balance according to their preferences.

---

[1]Experiments have shown that an approximately 3dB increase in signal-to-noise-ratio (SNR) enhances dialogue intelligibility to that of the listener's native language (Florentine (1). In some cases when the speech material is too complex, 3dB are not enough. According to Warzybok et al., these situations call for an increase of 5-10 dB, depending on the audience's language skills (2).

[2]For more information about the effect of an increased SNR, please see Brand and Kollmeier (4). For more information about the extension and development of hearing loss, please see Heger and Holube (5) as well as Kochkin (6).

2. INTRODUCTION: DIALOGUE ENHANCEMENT

The foundation for Dialogue Enhancement is MPEG Spatial Audio Object Coding (MPEG SAOC), a technology developed to manipulate any number of audio objects during rendering. Single instruments, dialogue or a singer's voice are just some examples of possible audio objects. The main principle is that the objects are all mixed into one downmix audio signal for transmission. The downmix signal can either be a mono, stereo or multichannel signal. Additional object-related side information is transmitted along with the downmix signal. This information enables the object manipulation in the receiver based on user interaction.

For Dialogue Enhancement, only a subset of the MPEG SAOC functionality is needed. No more than two objects (dialogue and background) are used, reducing the decoder complexity substantially. Further, Dialogue Enhancement only needs loudness level reduction or amplification for interactivity during rendering.

2.1 Basic working principle

Figure 1 gives an overview of the end-to-end signal flow from encoder input to decoder output. After analyzing the objects, the Dialogue Enhancement encoder generates a stream of parametric side information. It is possible to either mix the dialogue and background input signals within the encoder or use an external mix together with a clean dialogue or background signal. Then, any audio codec can encode the mixed signal – in this example, it is MPEG-4 AAC/HE-AAC. The side information stream is integrated into the encoded audio bit stream. As an alternative, the parameters could also be transported in a separate stream.

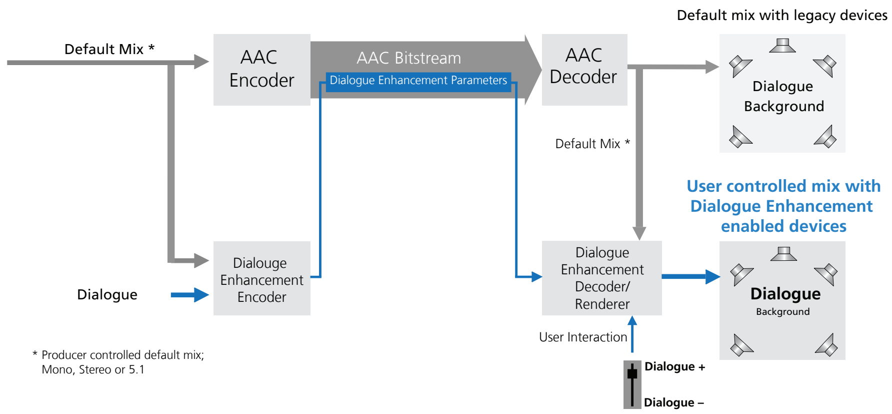

Figure 1: Dialogue Enhancement end-to-end signal flow.

2.2 Integrating Dialogue Enhancement into the signal flow

There are two ways of integrating the Dialogue Enhancement encoder into the signal flow. The most common method is shown in Figure 1, where the encoder is supplied with the externally created default mix and the dialogue signal. Alternatively, the background signal could replace the dialogue signal.

As a second option, the two audio sources dialogue and background could be fed into the Dialogue Enhancement encoder, where they are mixed into one signal that subsequently enters the AAC audio encoder.

2.3 The different modes of Dialogue Enhancement

Dialogue Enhancement offers two encoding modes for the side information:

- The basic **parametric mode** already described above: The side information stream is generated by the Dialogue Enhancement encoder and then embedded into the encoded audio bit stream.
- The **enhanced residual mode** that additionally embeds residual information for the dialogue object into the side information.

The parametric mode is typically useful for dialogue level modifications of up to 6 dB.

The residual mode allows for more attenuation or enhancement beyond 6 dB at high quality, but this method requires additional bit rate for the residual information. As a result of discussions with several broadcasters, an upper limit of 12 dB for the alterations was found to be a good compromise between bit rate and flexibility. The value of 12 dB is also used in the experiments described below.
Dialogue Enhancement can be used for any number of transmission channels, such as mono, stereo or multichannel. In the case of multichannel audio, two typical modes became apparent from discussions with broadcasters. The dialogue is either present in the center channel only or it is part of the front three channels: left, center and right.

2.4. Technical advantages

Dialogue Enhancement is backwards compatible with existing transmission equipment and receivers. Existing devices that are not capable of decoding the parametric side information will ignore it and play back the default mix.
Only one single audio track with the additional side information needs to be transmitted. This saves channel capacities compared to transmitting several pre-mixed versions or additional dialogue and background audio tracks.

3. TESTING THE TECHNOLOGY'S FEASIBILITY

The integration of Dialogue Enhancement into the production environment has been tested in a number of experiments, starting with a feasibility test in 2011 during the Wimbledon lawn tennis tournament. UK listeners of the BBC Radio 5 Live Internet stream were given the opportunity to download an audio playback client including Dialogue Enhancement as the main feature. This allowed them to control the volume of the audio component "Commentary" separately from the component "Court Atmosphere" with a maximum of 12 dB attenuation or amplification.
The subsequent feedback was consistently positive and the possibilities offered by the technology addressed the listeners' needs. A general statement was that Dialogue Enhancement would be especially useful for sports and drama programs.

Only a minority of 7% of the listeners chose to listen to the default mix, the remaining listeners chose to change the volume levels with two main groups emerging. The first one slightly turned down the commentary's volume to enjoy more of the court atmosphere. The second group significantly raised the volume of the commentary to improve intelligibility.[3]

To validate the early results from the Wimbledon experiment, a speech intelligibility experiment was conducted to verify the benefits for a hearing-impaired audience. The results of this experiment[4] indicate that the Dialogue Enhancement technology can be used as intended. An enhancement of the dialogue by 6 dB already showed a substantial improvement in intelligibility. An enhancement of 12 dB improves the speech intelligibility for hearing-impaired listeners to values comparable to normal-hearing listeners.

4. STANDARDIZATION

Dialogue Enhancement has been standardized within DVB as Advanced Clean Audio Services of the audio-video-coding toolbox.

[3]Further details on the survey results can be found in Fuchs et al. (7) and Meltzer (8).

[4]Details on the test results can be found in Fuchs et al. (9).

REFERENCES

1. Florentine, M. 1985., Speech perception in noise by fluent, non‐native listeners. J. Acoust. Soc. Am., Volume 77, Issue S1, pp. S106-S106.

2. Warzybok, A. et al., 2010. Influence of the linguistic complexity in relation to speech material on non-native speech perception in noise. DAGA 2010, Berlin.

3. Jansen, S., Luts, H., Wagener, K. C., Kollmeier, B., Del Rio, M., Dauman, R., James, C., et al., 2012. Comparison of three types of French speech-in-noise tests: a multi-center study. International Journal of Audiology, 51(3), 164–73. doi:10.3109/14992027.2011.633568

4. Brand, T. and Kollmeier, B., 2002. Efficient adaptive procedures for threshold and concurrent slope estimates for psychophysics and speech intelligibility tests. The Journal of the Acoustical Society of America, 111(6), 2801. doi:10.1121/1.1479152

5. Heger, D., and Holube, I., 2010. Wie viele Menschen sind schwerhörig? Zeitschrift für Audiologie, 49(2), pp. 61–70.

6. Kochkin, S., 2005. MarkeTrak VII: Hearing loss population tops 31 million people. Hearing Review, 12(7). http://www.betterhearing.org/pdfs/Marketrak7_Kochkin_July05.pdf

7. Fuchs, H., Tuff, S. and Bustad C., 2012. Dialogue Enhancement - technology and experiments. EBU Technology Review, June 2012. http://tech.ebu.ch/webdav/site/tech/shared/techreview/trev_2012-Q2_Dialogue-Enhancement_Fuchs.pdf

8. Meltzer, S., 2011. Dialogue Enhancement - A new approach. Nordig Sound Symposium, Sept. 2011. http://www.nrk.no/soundsymp/2011_Meltzer.html

9. Fuchs, H., Oetting, D., 2013. Advanced Clean Audio Solution: Dialogue Enhancement. IBC Conference, Sept. 2013.

ABOUT FRAUNHOFER IIS

The Audio and Media Technologies division of Fraunhofer IIS has been an authority in its field for more than 25 years, starting with the creation of mp3 and co-development of AAC formats. Today, there are more than 10 billion licensed products worldwide with Fraunhofer's media technologies, and over one billion new products added every year. Besides the global successes mp3 and AAC, the Fraunhofer technologies that improve consumers' audio experiences include Cingo® (spatial VR audio), Symphoria® (automotive 3D audio), xHE-AAC (adaptive streaming and digital radio), the 3GPP EVS VoLTE codec (crystal clear telephone calls), and the interactive and immersive MPEG-H TV Audio System.

With the test plan for the Digital Cinema Initiative and the recognized software suite easyDCP, Fraunhofer IIS significantly pushed the digitization of cinema. The most recent technological achievement for moving pictures is Realception®, a tool for light-field data processing.

Fraunhofer IIS, based in Erlangen, Germany, is one of 69 divisions of Fraunhofer-Gesellschaft, Europe's largest application-oriented research organization.

For more information, contact amm-info@iis.fraunhofer.de, or visit www.iis.fraunhofer.de/amm.