A new parametric Stereo and Multi Channel Extension for MPEG-4 Enhanced Low Delay AAC (AAC-ELD)

María Luis Valero[1], Andreas Hölzer[3], Markus Schnell [1], Johannes Hilpert[1], Manfred Lutzky [1], Jonas Engdegård[2], Heiko Purnhagen[2], Per Ekstrand[2] and Kristofer Kjörling[2]

[1] *Fraunhofer Institute for Integrated Circuits IIS, Am Wolfsmantel 33, 91058 Erlangen, Germany*

[2] *Dolby Sweden AB, Gävlegatan 12A, 11330, Stockholm, Sweden*

[3] *DSP Solutions GmbH & Co. KG, Maximilianstraße 16, 93047 Regensburg, Germany*

Correspondence should be addressed to Manfred Lutzky (`Manfred.Lutzky@iis.fraunhofer.de`)

ABSTRACT

ISO/MPEG standardizes two communication codecs with low delay: AAC-LD is a well established low delay codec for high quality communication applications such as video conferencing, tele-presence and Voice over IP. Its successor AAC-ELD offers enhanced bit rate efficiency being an ideal solution for broadcast audio gateway codecs.

Many existing and upcoming communication applications benefit from the transmission of stereo or multi channel signals at low bit rates. With Low Delay MPEG Surround, ISO has recently standardized a low delay parametric extension for AAC-LD and AAC-ELD. It is based on MPEG Surround technology with specific adaption for low delay operation. This extension comes along with a significantly improved coding efficiency for transmission of stereo and multi channel signals.

1 INTRODUCTION

In 1999, ISO/MPEG defined the audio codec Low Delay AAC (AAC-LD) [1], intended for bi-directional (low latency) communication. The design of the codec is based on the AAC format, but uses a reduced frame length resulting in an algorithmic delay of 20 ms rather than the 55 ms of AAC [2] for packet based transmission. The AAC-LD technology provides excellent sound quality for typical audio content at bit rates between 48 and 64 kb/s per channel. In order to further increase the bit rate efficiency, a low delay derivative of the Spectral Band Replication (SBR) [3] tool was adopted by AAC-LD. This combination is known as Enhanced Low Delay AAC (AAC-ELD)[4] which was finalized at MPEG in 2008 [5]. AAC-ELD allows the coding of "super-wideband" signals (i.e. full audio bandwidth) at bit rates down to 24 kb/s while the delay is still low - approximately 32 ms. During the standardization process the main focus was mostly on mono signals. However, MPEG offers some interesting technologies for the coding of multi channel content, e.g. MPEG Surround, which has recently been adapted for low delay codecs. Such tools allow codecs to provide good quality for stereo or multi channel content at much lower bitrates as it is possible with discrete coding of each channel.

2 MPEG SURROUND AND MPEG SPATIAL AUDIO OBJECT CODING (SAOC)

Finalized in 2007, MPEG Surround (ISO/IEC 23003-1 [6]) is the ISO standard for the parametric multi channel extension of monophonic or stereophonic audio codecs. MPEG Surround is based on the technical principles of Spatial Audio Coding (SAC). The building blocks are shown in Figure 1.

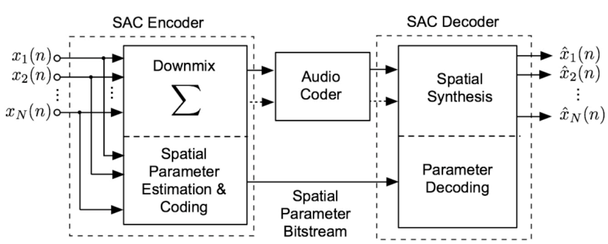


Figure 1: Principle of Spatial Audio Coding

Figure 1: SAC block diagram [7]

A spatial audio encoder acts as a preprocessor to conventional perceptual audio coder. It derives perceptually relevant parameters (like level differences and correlation between channels) from the multi channel (e.g. 5.1 channels) input signal $x_1 \ldots x_N$. The signal analysis in MPEG Surround is done based on a complex valued frequency decomposition of the input signal by means of a Quadrature Mirror Filter Bank followed by an oddly-modulated Nyquist filter bank (hybrid QMF). At the same time, a mono or stereophonic down mix of the input signal is generated by the spatial audio encoder. A compact parametric description of the spatial properties of the multi channel input signal is conveyed next to the down mix. On the receiver side, the down mix signal is expanded to the multi channel output $\hat{x_1} \ldots \hat{x_N}$ by applying the spatial parameters in the spatial audio decoder.

During the standardization of MPEG Surround, the main focus was compression efficiency, flexibility and scalability of the technology. A processing delay low enough to allow for an efficient combination with communication audio codecs was not on the requirements list at that time.

MPEG Spatial Audio Object Coding (SAOC) has been a follow-up standardization process in ISO, converting the principles of Spatial Audio Coding from a 'channel-oriented' approach towards an 'audio objects' approach. The technology was recently finalized in early 2010. SAOC supports interactive manipulation of levels and spatial rendering positions of independent audio objects at the receiver side. SAOC re-uses an MPEG Surround decoder when rendering an arbitrary number of audio objects to a multi channel loudspeaker setup. In order to support interactive teleconferencing application scenarios, a low delay operation mode of SAOC processing as well as a low delay multi channel renderer (LD MPEG Surround) were specified in ISO/IEC 23003-2 [8].

3 LOW DELAY MPEG SURROUND

3.1 Introduction

The low delay variant of MPEG Surround is directly derived from the original MPEG Surround specification but optimized towards a low algorithmic delay. In order to motivate the applied changes, all possible delay sources are identified in the following section.

3.2 Delay Analysis of MPEG Surround

MPEG Surround comprises an algorithmic delay exceeding the requirements for a combination with communication codecs. Figure 2 shows a detailed signal flow of an MPEG Surround extended AAC-ELD core coder. An analysis of the delay sources reveals the following contributions:

level/intensity differences and measures of correlation/coherence between the audio channels and can be represented in an extremely compact way. At the same time, a monophonic or two-channel downmix signal of the sound material is created and transmitted to the decoder together with the spatial cue information. The downmix can be conveyed to the receiver using known audio coders for monophonic or stereophonic signals. On the decoding side, the transmitted downmix signal is expanded into a high quality multi-channel output based on the spatial parameters.

Due to the reduced number of audio channels to be transmitted (e.g. just one channel for a monophonic downmix signal), the Spatial Audio Coding approach provides an extremely efficient representation of multi-channel audio signals. Furthermore, it is backward compatible on the level of the downmix signal: A receiver device without a spatial audio decoder will simply present the downmix signal.

Conceptually, this approach can be seen as an enhancement of several known techniques, such as an advanced method for joint stereo coding of multi-channel signals [4], a generalization of Parametric Stereo [5] [6] to multi-channel application, and an extension of the Binaural Cue Coding (BCC) scheme [7] [8] towards using more than one transmitted downmix channel [9]. From a different viewing angle, the Spatial Audio Coding approach may also be considered an extension of well-known matrix surround schemes (Dolby Surround/Prologic, Logic 7, Circle Surround etc.) [10] [11] by transmission of dedicated (spatial cue) side information to guide the multi-channel reconstruction process and thus achieve improved subjective audio quality [1].

Due to the combination of bitrate-efficiency and backward compatibility, SAC technology can be used to enhance a large number of existing mono or stereo services from stereophonic (or monophonic) to multi-channel transmission in a compatible fashion. To this aim, the existing audio transmission channel carries the downmix signal, and the spatial parameter information is conveyed in a side chain (e.g. the ancillary data portion of an audio bitstream). In this way, multi-channel capability can be achieved for existing audio

- Framing delay

  Since block based processing is applied the input samples corresponding to one frame have to be collected.

- Residual coding delay compensation

  MPEG Surround supports a tool called *residual coding* to allow a smooth transition from parametric to waveform preserving coding [7]. For a good performance this tool requires a look ahead of two frames for the analysis.

- Look ahead for the analysis window

  To determine the spatial parameterization of the input signal in the MPEG surround encoder one additional frame of look ahead for succeeding time samples is needed.

The contributions of framing, look ahead and residual coding delay are directly proportional to the frame length which is determined by the employed mono coder. Having AAC-ELD as core coder, the number of samples per frame is set to 512 samples (without SBR) or 1024 samples (dual-rate SBR). Hence, these three sources introduce a delay of four times the frame length.

- Hybrid filter banks (QMF, Nyquist)

  The time-frequency transformations are performed by means of the hybrid QMF. The QMF has a combined analysis/synthesis delay of 577 samples, while the delay introduced by the hybrid decomposition/reconstruction amounts to 384 samples. Note that the filter bank contributes to this delay twice, once on the encoder- and once on the decoder-side.

- Core coder

  As the core coder is fed with a down mix that is generated by the MPEG Surround encoder, the AAC-ELD codec delay adds up to the total system delay. The delay of AAC-ELD without SBR amounts to 256 samples, with dual rate SBR it amounts to 596 samples.

- Low Power delay compensation

  The MPEG Surround decoder supports a *Low Power* operation mode which was designed to minimize computational complexity [8]. The required processing adds a delay of 320 samples.

- Synchronization buffers (D1, D2)

The synchronization buffers D1 and D2 in Figure 2 are used to ensure the alignment of both bit stream and down mix at the decoder side. Their size is variable, with the restriction that the size of the bit stream buffer is $D2 = N \cdot frame\ length$. As it can be deduced from Figure 2:

$$D2 = N \cdot frame\ length \stackrel{!}{=} D_{QMF^{-1}} + D1 + D_{CoreCoder} + D_{QMF} + D_{LowPower} + D_{Nyquist},$$

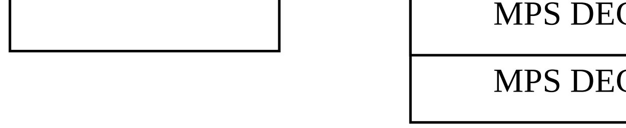
where the minimization of N results in N=4 and D1=511 samples (without SBR) and N=2 and D1=171 samples (dual rate SBR).

At a sampling rate of 48 kHz the overall delay of the combination of MPEG Surround with AAC-ELD as core coder amounts to 105.4 ms without using SBR and to 148 ms for a system with SBR enabled. ITU-G.114 [9] recommends an end-to-end delay for telecommunication applications below 150 ms. As pre- and postprocessing like gain- and echo-control, networking, etc. introduce additional delay, the given recommendation is hard to comply to.

3.3 Structural differences between MPEG Surround and LD MPEG Surround

In order to obtain a system with minimum delay, the following changes were applied to MPEG Surround which are also outlined in Figure 3.

- Down mix processed in time domain
- Residual coding not allowed
- Reduced look ahead of half a frame
- QMF replaced by Low Delay QMF
- No Nyquist decomposition
- Low Power operation not supported



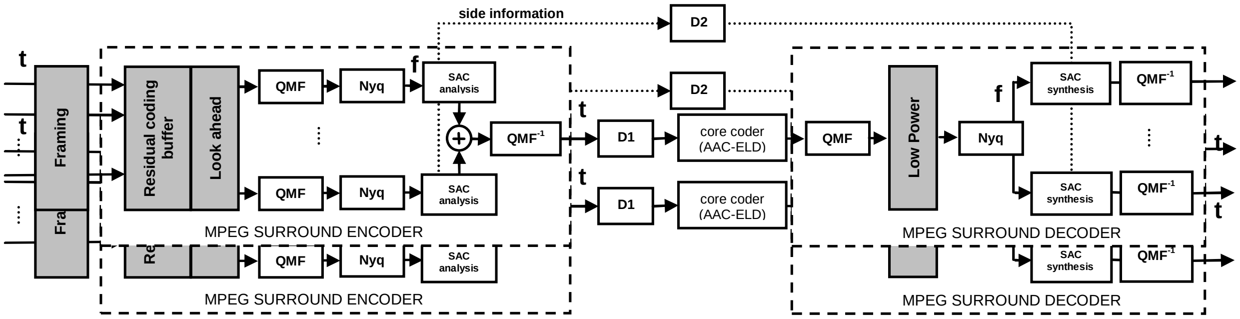


Figure 2: MPEG Surround delay sources

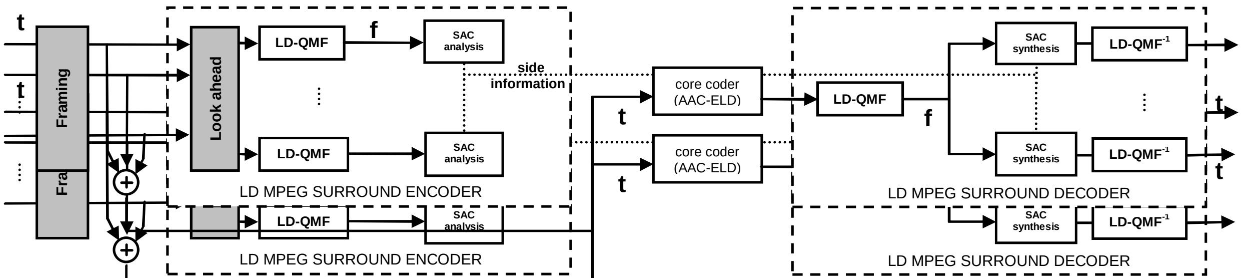


Figure 3: LD MPEG Surround delay sources

As the down mix is processed in the time domain, both LD MPEG Surround and the core coder can run in parallel which means that their delays do no longer add up. Basically, a down mix processing in the frequency domain is still possible but comes along with an increased system delay due to the synthesis filter bank and the need of synchronization buffers.

Furthermore, the structure of the filter banks has been optimized with respect to the delay by removing the Nyquist decomposition and replacing the QMF filter bank by a Low Delay QMF (see 3.5). This delay optimized filter bank has a combined analysis/synthesis delay of 256 samples which corresponds to 5.3 ms at 48 kHz sampling frequency.

Moreover, the length of the look ahead is reduced to half a frame length which coincides with the LD-MDCT delay of AAC-ELD meaning that the look ahead delay can be completely concealed by the core coder delay.

Finally, the total delay for LD MPEG Surround running with a AAC-ELD core coder without SBR amounts to 21.3 ms, adding only 5.3 ms on top of the AAC-ELD delay. For AAC-ELD with SBR, the overall delay amounts to 39 ms. A further delay reduction is achieved by interfacing AAC-ELD's SBR and LD MPEG Surround in the frequency domain (see also Section 4.2). In this case the overall delay amounts to 37.7 ms.

A summary of the delay numbers above is presented in Table 1.

In addition to those delay optimizations the following changes were applied to LD MPEG Surround:

- As a consequence of the omission of the Nyquist decomposition, the number of processing bands is reduced from 28 to 23 [8].
- The decorrelator filters are adapted to the low delay operation [8].
- The typical frame length of 512 samples results in a relatively high update rate of the side information. To decrease the impact on the resulting bit rate the encoder was adapted to support a subsampling of the spatial parameters.

Table 1: Delay Summary (for a sampling frequency of 48 kHz)

AAC-ELD	without SBR	with dual rate SBR
frame length	512 samples	1024 samples
AAC-ELD core coder delay	16 ms	33.7 ms
AAC-ELD + MPEG Surround	105.4 ms	148 ms
AAC-ELD + LD MPEG Surround	21.3 ms	39 ms (1) 37.7 ms (2)

(1) time domain interface, (2) frequency domain interface between AAC-ELD and LD MPEG Surround

3.4 2-1-2 Channel Configuration

One fundamental building block of the MPEG Surround technology is the One-To-Two box (OTT box) which takes one channel as input and produces two channels as output [6]. By the concatenation of these boxes channel configurations are realised which support up to 32 output channels. In addition to these channel configurations the low delay flavor of MPEG Surround offers a dedicated parametric stereo extension mode which is based on a single OTT box and is referred to as the 2-1-2 configuration. To sum up, Low Delay MPEG Surround covers both, a parametric stereo as well as a multichannel extension of a communication core coder with clearly increased bitrate efficiency compared to discrete coding of the channels.

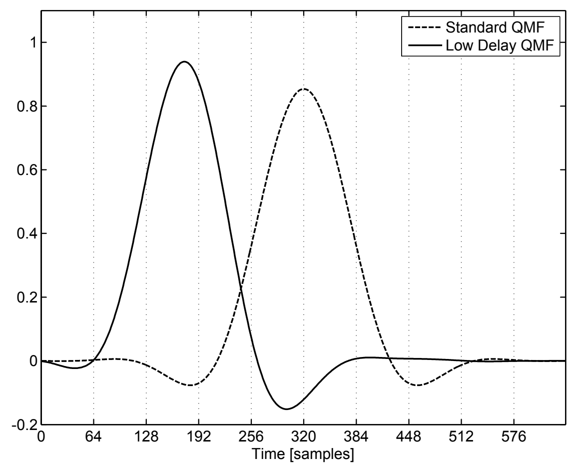


Figure 4: Low Delay QMF bank prototype

3.5 Low Delay QMF

In order to reduce the delay induced by the standard MPEG Surround QMF bank, a low delay filter bank employing complex exponential modulation of an asymmetric prototype filter was introduced to LD MPEG Surround.

The low pass prototype filter has, likewise to the standard QMF, 640 coefficients. However, as seen from Figure 4, where the prototype filter coefficients are shown together with the coefficients of the standard QMF, the filter is asymmetric and allows for a much shorter system delay. The prototype filter coefficients are optimized to supress alias artifacts emanating from independent processing of neighbouring subband channels. This is of vast importance in MPEG Surround due to the frequency dependent upmix matrices that are applied in the decoder.

The total delay of the low delay filter bank is 319 samples, and excluding the framing delay of 63 samples, this amounts to 256 samples. Actually, the framing delay does not contribute to the overall delay in a system that is inherently frame-based, such as the MPEG Surround decoder. Hence, the delay corresponds to 5.3 ms at 48 kHz sampling frequency. The filter bank has a linear reconstruction error (amplitude and phase errors) of -72 dB, e.g. the phase deviation from linear phase is smaller than $\pm 0.02°$, and the aliasing suppression is 76 dB in absence of modifications to the subband samples. For comparision, the standard QMF bank has a total delay of 640 samples, which corresponds to 12 ms at 48 kHz sampling frequency, a linear reconstruction error of -60 dB (amplitude only error) and 75 dB of aliasing suppression.

Figure 5 shows the amplitude responses of the low delay prototype filter and the standard QMF prototype filter. As seen from the figure, the low delay filter has a slightly lower stop band attenuation compared to the standard QMF filter. Potentially, this means somewhat higher susceptibility to alias artifacts when neighboring QMF bands are processed independently. However, this deficiency comes with the advantage of the much smaller delay. Thus, the low delay filter bank reaches a good compromise between delay minimization and overall subjective sound quality.

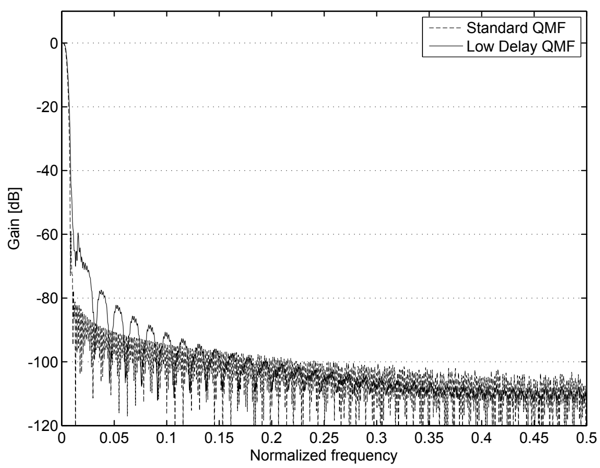


Figure 5: Low Delay QMF bank frequency response

4 INTEGRATION OF LOW DELAY MPEG SURROUND

4.1 MPEG frame work

The MPEG frame work ensures that LD MPEG Surround can be used as an extension of MPEG's low delay codecs in a fully backwards compatible manner. This extension is always signaled explicitly, either inside the Audio Specific Config (ASC) for AAC-LD or as an AAC-ELD extension inside the ELD Specific Config. For LD MPEG Surround, the Audio Object Type (AOT) 44 has been reserved. Due to the explicit nature of signaling, a suitable negotiation of the codec extensions with respect to the device properties is possible. The transport of the spatial parameter information is implemented as extension payloads where a new type has been assigned.

4.2 Interfacing AAC-ELD and LD MPEG Surround

SBR for AAC-ELD is standardized in combination with a Complex Low Delay Filter Bank (CLDFB) as QMF substitute. In order to enable the frequency domain interface between SBR and LD MPEG Surround (see Section 3.3), the time frequency transformations inside SBR are applied using the LD QMF and consequently both modules operate in the same frequency domain. This means that whenever LD MPEG Surround is signaled in the ASC, the SBR module should operate in the LD QMF domain. The time alignment of the bit streams has to be adapted according to the LD QMF. To ensure backwards compatibility of mono devices - i.e. no LD MPEG Surround support - the signal has to be delayed by 48 samples before the SBR decoder. Hereby the difference of the reconstruction delay of the two filter banks CLDFB and LD QMF is compensated.

5 PERFORMANCE EVALUATION

5.1 Overview

The subjective performance of the LD MPEG Surround extended AAC-ELD technology has been assessed for the following configurations.

- Stereo operation comparing discrete stereo AAC-ELD against LD MPEG Surround 2-1-2 with ELD mono core.
- Multi channel playback comparing discrete 5.1 AAC-ELD against LD MPEG Surround 5-1-5 with ELD mono core.

The test items were chosen to resemble realistic application scenarios. The two listening tests are described in the following subsections.

5.2 Test Methodology

The subjective listening tests were conducted in an acoustically isolated listening room that is designed to permit high-quality listening. The signals were presented via

- headphones for the LD MPEG Surround 2-1-2 listening test.
- loudspeakers for the LD MPEG Surround 5-1-5 listening test.

The test methodology followed the standard procedures used in the MPEG Surround verification tests, based on the "Multiple Stimulus with Hidden Reference and Anchors" (MUSHRA) method for the subjective assessment of intermediate quality audio [10]. In accordance with the MUSHRA methodology, the listeners were instructed to compare all test conditions against the reference. All subjects can be considered as experienced listeners. The test conditions were randomized automatically for each test item and for each listener. The subjective responses were recorded by a computer-based MUSHRA program on a scale ranging from 0 to 100. Instantaneous switching between the items under test was allowed.

Two MUSHRA tests were conducted to assess the perceptual performance of LD MPEG Surround for stereo and 5.1 playback compared to discrete AAC-ELD and the reference signal. As a hidden anchor a 3.5 kHz low pass filtered reference condition was used. For the listening test the MPEG Surround payload was embedded in the AAC-ELD bit stream to ensure an overall constant bit rate for the multiplexed bit stream. MPEG Surround was interfaced with the AAC-ELD core coder using the spectral domain interface. The sampling frequency for all systems under test was 48 kHz.

5.3 LD MPEG Surround 2-1-2 configuration

On the one hand, a LD MPEG Surround 2-1-2 extended ELD mono core coder was compared to discrete stereo AAC-ELD. The discrete stereo signals were coded using AAC-ELD with SBR at bit rates of 48 kb/s and 32 kb/s. For LD MPEG Surround 2-1-2, the mono down mixes were coded using AAC-ELD with SBR at 32 kb/s including the embedded MPEG Surround payload.

Along with four speech items, four music items were included in the listening test. The *rockyou* item has proven to be very critical during the development of LD MPEG Surround, the other three music items were previously used in MPEG stereo testing.

Table 2: Systems under test for the LD MPEG Surround 2-1-2 listening test

System name	System description
hidden_ref	Hidden reference signal
ELD_discr_48	Discrete stereo @ 48 kb/s
ELD_discr_32	Discrete stereo @ 32 kb/s
ELD_param_32	LD MPS 2-1-2 @ 32 kb/s
3.5kHz	hidden anchor (3.5 kHz)

Table 3: Item description and spatial payload for the LD MPEG Surround 2-1-2 listening test

Item name	Description	Bit rate [kb/s]
Fluch_der_karibik	Speech	2.90
reunion	Speech	2.51
rockyou	Music	3.01
sr000_4ch	Speech	2.88
tabou	Speech	2.63
te09	Music	2.84
te16	Music	3.13
te44	Music	3.09
all items		ø 2.87

Table 2 contains a description of the systems under test. Table 3 shows the list of audio items as well as the bit rates of the corresponding spatial side information. The MUSHRA scores for 17 listeners are given in Figure 6.

5.4 LD MPEG Surround 5-1-5 configuration

On the other hand, a LD MPEG Surround 5-1-5 extended ELD mono core coder was compared to discrete 5.1 AAC-ELD. The discrete coding was done using AAC-ELD with SBR at bit rates of 160 kb/s and 128 kb/s. For LD MPEG Surround 5-1-5, the mono down mixes were coded using AAC-ELD with SBR at 64 kb/s including the embedded MPEG Surround payload.

The set of items used for this listening test includes four items used during the MPEG standardization process. In addition, four speech items were included.

Table 4 contains the description of the systems under test. Table 5 contains the list of audio items as well as the bit rates of the corresponding spatial side

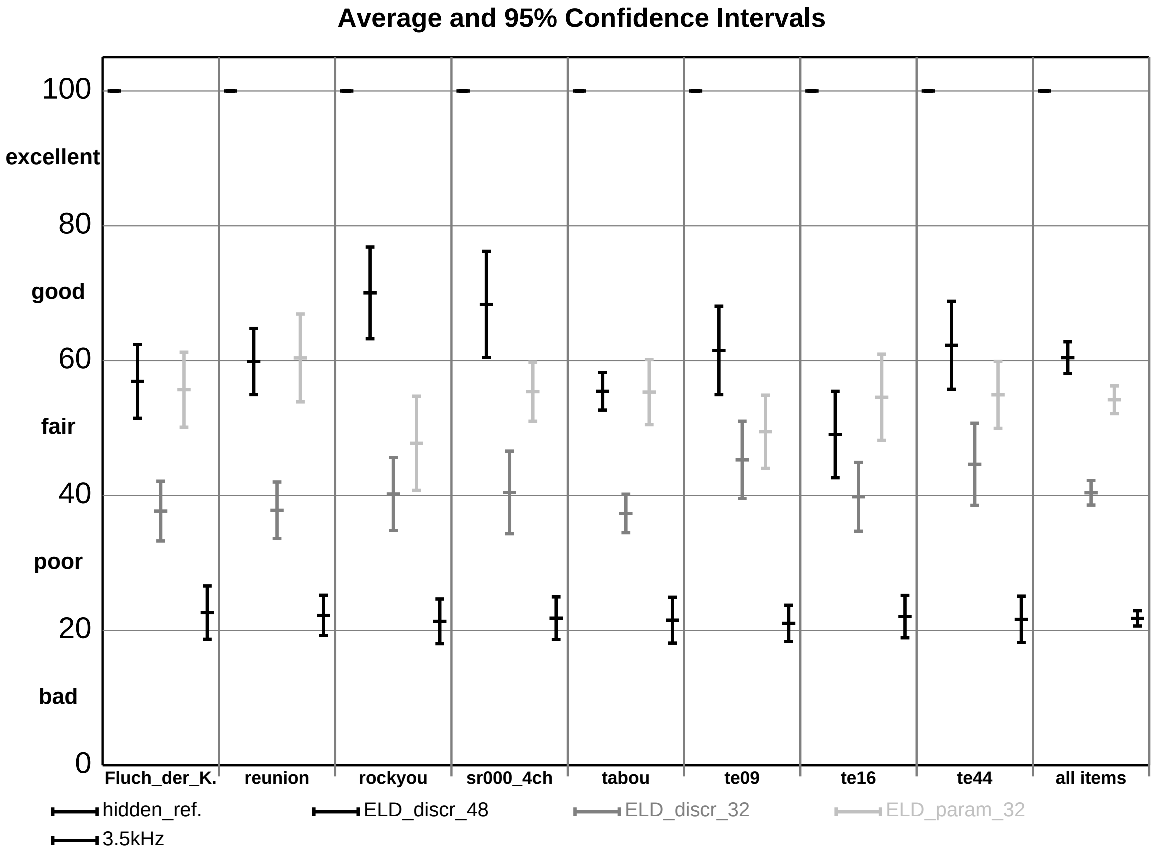


Figure 6: LD MPEG Surround 2-1-2 MUSHRA listening test results of 17 subjects

Table 4: Systems under test for the LD MPEG Surround 5-1-5 listening test

System name	System description
hidden_ref	Hidden reference signal
ELD_discr_160	discrete 5.1 @ 160 kb/s
ELD_discr_128	discrete 5.1 @ 128 kb/s
ELD_param_64	LD MPS 5-1-5 @ 64 kb/s
3.5kHz	hidden anchor (3.5 kHz)

information. The MUSHRA scores for 12 listeners are given in Figure 7.

5.5 Discussion of listening test results

A graphical representation of the listening tests results can be found in Figure 6 and Figure 7. These plots depict the average MUSHRA grading per item

Table 5: Item description and spatial payload for the LD MPEG Surround 5-1-5 listening test

Item name	Description	Bit rate [kb/s]
Stomp	Music	12.43
cckma	Speech	12.63
chostakovitch	Music	11.74
cocktailparty	Speech	13.74
indie	Film ambience	9.37
pops	Pop Music	12.03
reunion	Speech	11.53
telco_sim	Speech	9.13
all items		ø 11.29

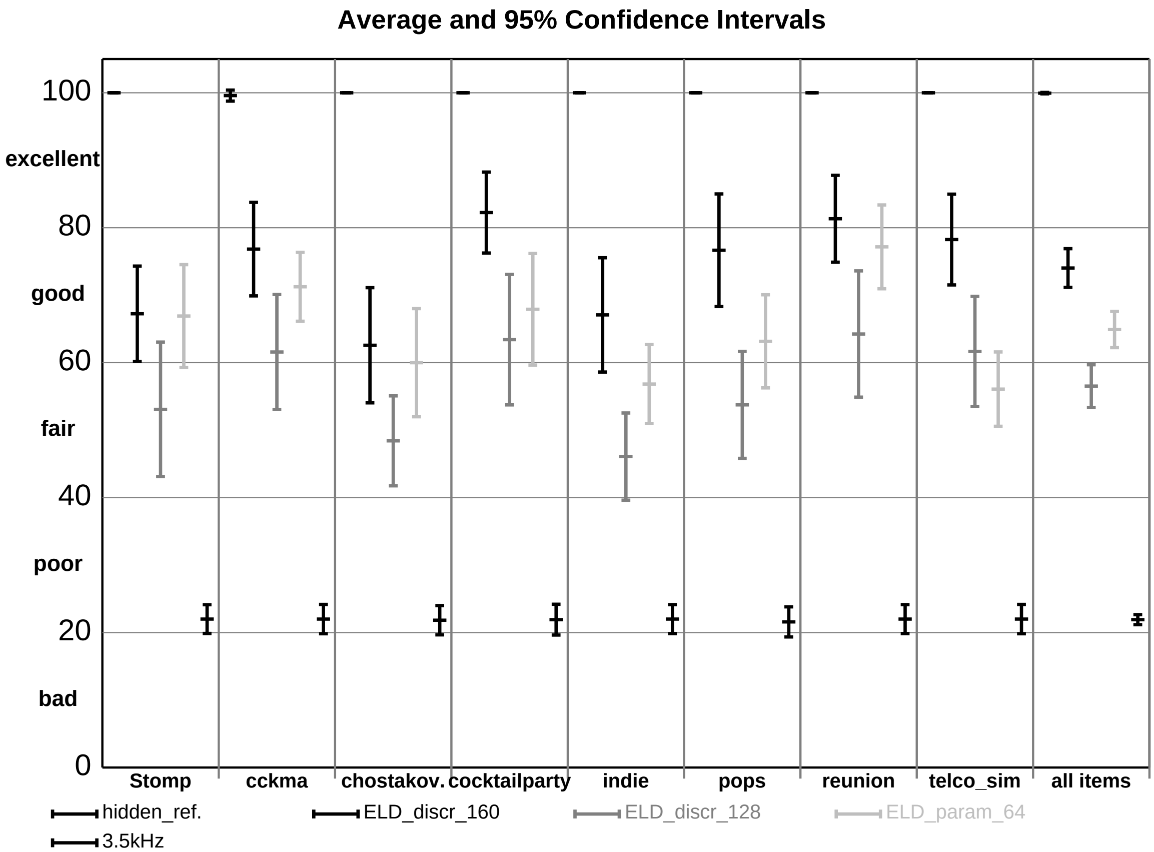


Figure 7: LD MPEG Surround 5-1-5 MUSHRA listening test results of 12 subjects

over all listeners and the statistical mean value over all evaluated items together with the associated 95% confidence intervals.

The listening test results in Figure 6 show that parametrically extended AAC-ELD at 32 kb/s achieves comparable MUSHRA results for five out of eight items (Fluch_der_Karibik, reunion, tabou, te16 and te44) compared to discrete stereo AAC-ELD at 48 kb/s.

Parametric versus discrete coding at 32 kb/s reveals a significantly better perceptible audio quality of the LD MPEG Surround 2-1-2 configuration over all items. For no test item the achieved MUSHRA score is worse, five out of the eight test items show a significant improvement.

Figure 7 compares discrete 5.1 AAC-ELD at 128 kb/s and 160 kb/s with parametrically extended AAC-ELD at 64 kb/s. Seven out of eight items show overlapping confidence intervals between the parametrically extended version at 64 kb/s and the discrete version at 160 kb/s. Nevertheless, a significant difference over all items is observed.

In contrast, compared to a discrete version of AAC-ELD at 128 kb/s the LD MPEG Surround 5-1-5 configuration is significantly better over all items while using only half the bit rate.

6 APPLICATIONS

6.1 Overview

The MPEG low delay audio codecs are well established in communication products such as telepresence systems, tele- and video conferencing systems, Voice over IP (VoIP) applications and broadcast gateways. Up to now, most of these applications only provide mono audio connections between the participants. In recent years telepresence systems have revolutionized the user experience by two main differentiators to Video-conferencing. The most prominent one is the introduction of several HD video screens which allow the reproduction of communication partner in natural size. The second but not less important one is a multichannel audio connection which enables a spatially correct positioning of the remote audio scene.

The benefit of spatial audio, the telepresence user easily gets aware of, is described in a very interesting experiment [11]. The experiment aims at comparing mono tele conferencing scenarios to setups where the participants are distributed in the spatial panorama. The conclusion is that spatial audio clearly enhances memory, focal assurance and perceived comprehension. The theory behind this result is that the tasks for speaker identification and semantic meaning can be distributed among distinct parts of the human working memory. This shows that spatial conferencing does not only create a more natural sound enviroment but also improves the efficiency of the communication itself.

Other applications which benefit from spatial audio apart from tele- and videoconferencing are telepresence at home, social communication in the living room, communication for gaming and next generation of broadcast gateways.

6.2 Broadcast Gateways

Broadcast gateways are widely used for transmitting speech and audio content from a remote location to a broadcasting studio. The currently used fixed transmission networks are either announced to be shut down (ISDN) or suffer from pure reliability (DSL) due to packet loss and jitter. The reliability issue becomes even worse as soon as wireless channels like UMTS are used. A bearer of hope to solve all theses issues are upcoming LTE networks due to the integrated quality of service, nominal high date rates up to 100+ Mb/s as well as the low delay of below 10 ms on the air link. Due to the flat “all IP” network structure, LTE speech services will become more cost efficient compared to 2G and 3G networks and replace them in the long term.

Low delay MPEG surround in conjunction with AAC-ELD enables low bit rate audio coding of stereo signals and is a promising combination for usage in UMTS and upcoming LTE networks. On the one hand, stereo capability is a key feature of this application and on the other hand, the actually available data rates can be significantly lower than the theoretical ones. This has been already experienced with UMTS where HSDPA data rates up to 7.2 Mb/s are expected but under poor network conditions only a small percentage of the data rate is experienced.

To prove this concept Fraunhofer has successfully setup a stereo conference prototype of an AAC-ELD based Communication Engine over a Fraunhofer owned multi-cell LTE-A testbed which operates in the center of Berlin.

6.3 Telepresence at Home

Applications like Telepresence at home or Room to Room communication are upcoming key applications for HD enabled devices due to the integration of IP network interfaces on modern TVs, STBs and data flatrates of triple play enabled DSL services. Users of TV and HiFi equipment are accustomed to full audio bandwidth of stereo or 5.1 audio. Adapted to those communication applications, an unique natural communication will feel like being in the same room. This trend has been recognised by a consortium of industry and scientific partners which are cooperating in the TA2 project (http://www.ta2-project.eu/Pages/contacts.html). A main goal of this project is to develop and evaluate technologies to enhance social communication between people in living room like environments. Here, natural spatial communication is one of the key features. Another one is automated orchestration of audio and video which aims at capturing and focusing on the most relevant interactions between the participants, e.g. laughing people.

7 CONCLUSIONS

Low Delay MPEG Surround enables parametric stereo and multi channel coding for MPEG-4 AAC-LD and ELD. The average data rate for the parametric stereo extension is only 3 kb/s while for multi channel, LD MPEG Surround requires 12 kb/s. At a total bit rate of 32 kb/s for stereo content, the combination of AAC-ELD and LD MPEG Surround shows a level of quality close to a discrete coding mode of AAC-ELD at 48 kb/s. For the coding of multi channel content at 64 kb/s, AAC-ELD with LD MPEG Surround shows a significant better quality at half the bit rate compared to the discrete coding mode of AAC-ELD. This proofs that LD MPEG Surround as a parametric spatial coding tool improves the performance of low delay, low bit rate and high quality communication codecs.

8 REFERENCES

[1] E. Allamanche, R. Geiger, J. Herre, T. Sporer, "MPEG-4 Low Delay Audio Coding Based on the AAC Codec", *106th AES Convention*, Munich, Germany, preprint 4929, May 1999

[2] M. Lutzky, G. Schuller, M. Gayer, U. Krämer, S. Wabnik, "A guideline to audio codec delay", *116th AES Convention*, Berlin, Germany, preprint 6062, May 2004

[3] M. Dietz, L. Liljeryd, K. Kjörling, O. Kunz, "Spectral Band Replication, a Novel Approach in Audio Coding", *112th AES Convention*, Munich, Germany, preprint 5553, Apr. 2002

[4] M. Schnell, M. Schmidt, M. Jander, T. Albert, R. Geiger, V. Ruoppila, P. Ekstrand, M. Lutzky, B. Grill, "MPEG-4 Enhanced Low Delay AAC - a new standard for high quality communication", *125th AES Convention*, San Francisco, CA, USA, preprint 7503, Oct. 2008

[5] ISO/IEC 14496-3:2008, "Coding of Audio-Visual Objects, Part 3: Audio - Amendment 9", 2008

[6] ISO/IEC 23003-1:2007, "MPEG-D (MPEG audio technologies), Part 1: MPEG Surround", 2007

[7] J. Herre, K. Kjörling, J. Breebaart, C. Faller, S. Disch, H. Purnhagen, J. Koppens, J. Hilpert, J. Rödén, W. Oomen, K. Linzmeier, K. S. Chong, "MPEG Surround -The ISO/MPEG Standard for Efficient and Compatible Multi-Channel Audio Coding", *AES 122th Convention*, Vienna, Austria, May 2007

[8] ISO/IEC FDIS 23003-2, "MPEG-D (MPEG audio technologies), Part 2: Spatial Audio Object Coding (SAOC)", 2010

[9] International Telecommunication Union, "One-way transmission time", ITU-T Recommendation G.114, 2003

[10] International Telecommunication Union, "Method for the subjective assessment of intermediate sound quality (MUSHRA)", ITU-R, Recommendation BS. 1534-1, Geneva, Switzerland, 2001

[11] J. J. Baldis, "Effects of spatial audio on memory, comprehension, and preference during desktop conferences", *CHI '01: Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM, New York, NY, USA, pp. 166–173, ISBN 1-58113-327-8, 2001