# SAOC FOR GAMING – THE UPCOMING MPEG STANDARD ON PARAMETRIC OBJECT BASED AUDIO CODING

**LEONID TERENTIEV[1], CORNELIA FALCH[1], OLIVER HELLMUTH[1], JOHANNES HILPERT[1], WERNER OOMEN[2], JONAS ENGDEGÅRD[3], HARALD MUNDT[4]**

[1] *Fraunhofer Institute for Integrated Circuits IIS, Erlangen, Germany*
[2] *Philips Applied Technologies, Eindhoven, The Netherlands*
[3] *Dolby Sweden AB, Stockholm, Sweden*
[4] *Dolby Germany GmbH, Nürnberg, Germany*

Correspondence should be addressed to: leonid.terentiev@iis.fraunhofer.de

Following the current trend of employing parametric enhancement tools for increased coding and spatial rendering efficiency, MPEG audio pursues a standardization activity on Spatial Audio Object Coding (SAOC) technology. The SAOC system extends the MPEG Surround standard by exploiting its coding efficiency. Moreover, the SAOC technology introduces user-controllable rendering functionality together with flexible choice of various playback configurations. These aspects are of potential interest for a large range of gaming applications that will benefit from the efficient coding and interactive rendering. Although SAOC targets many different application scenarios, besides describing the basic SAOC architecture and the manifold of enhancement tools, this paper will focus on the relevance for gaming applications.

## INTRODUCTION

According to a series of reports presented at the Game Developers Conference by the IGDA Online Games SIG the popularity of computer games continues to grow and a major expansion of the global gaming market is forecast for the computer game industry worldwide [1]. This growth can be attributed to a proper representation of the virtual worlds offered to players. One of the most important but sometimes hidden aspects when creating an attractive atmosphere in the game is the realistic rendering of the audio scene. User interaction in virtual worlds requires consistent time-variant positioning of multiple sound sources representing the players and other objects in the game. In addition, this interaction involves various sound effects reflecting the influence of the virtual world surroundings. Such an audiovisual scene may be acoustically represented by a number of individual audio objects plus a scene description stored or conveyed as multiple discrete audio tracks with corresponding descriptive side information as shown in Figure 1. Each coded audio bitstream (BS) features its own decoding stage. The subsequent rendering block exploits the scene description and furthermore provides the user-control interface. In practice, this approach can become undesirable due to large bandwidth requirements and computational complexity of the array of decoding processes, both denoting critical design measures that grow approximately proportional with the number of audio objects.

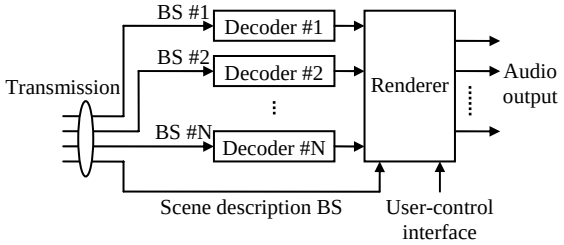



Figure 1: Audio scene representation with multiple discrete audio tracks

Spatial Audio Object Coding (SAOC) is a novel interactive audio format describing an entire

transmission chain comprising an object encoder and a corresponding decoder plus rendering unit. As illustrated in Figure 2, the SAOC system is designed to transmit several audio objects mixed into one or two downmix channels. A parametric description of all audio objects is stored in a dedicated SAOC bitstream. Although the parametric object related SAOC data grow linearly with the number of objects, in a typical scenario the bitrate consumption of the coded object data is negligible compared to that of a coded audio (downmix) channel. Therefore, the SAOC scheme is capable of transmitting multi-track content at bitrates only slightly exceeding those for mono or stereo signals.

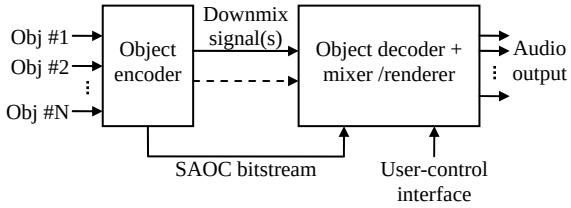


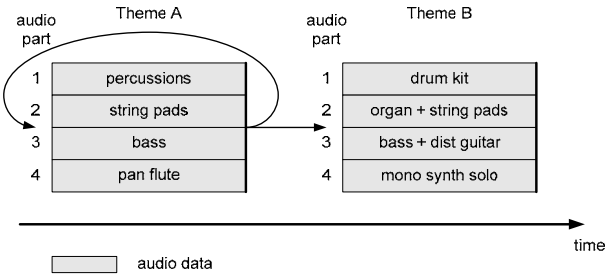

Figure 2: SAOC conceptual overview

## 1 SAOC AS GAME AUDIO ENGINE

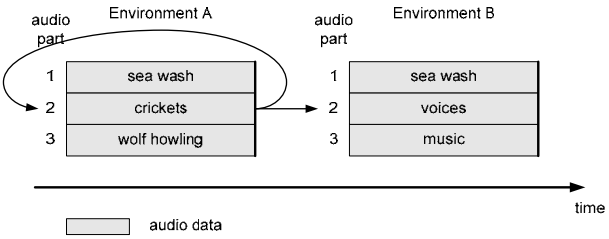
Since SAOC handles both the audio object decoding and the mixing, it can act as a complete or part of a game audio engine. In the following sections, typical game audio functions that particularly fit the SAOC tool are outlined. In order to demonstrate the advantages provided by the SAOC technology potentially relevant gaming application scenarios are also presented.

### 1.1 Non-linear Music and Soundscapes

Non-linear music, also known as interactive music, has been around almost since the creation of the first computer game. Though, in the early days it was mostly limited to simple sequencing of a few themes, e.g. one "combat theme" and one "main theme". There are several ways of implementing interactive music for games. One of them is the use of symbolic music representation, today typically MIDI, allowing very flexible mixing and sequencing options even including advanced automatic composition rules for transitions. Another approach, definitely being more deployed today, is to use dynamically mixed pre-authored audio parts. A number of real audio tracks typically require more memory than the corresponding MIDI track, and often but not always, the computational complexity of playback is higher. However this might depend on choice of audio codec and MIDI synthesizer as well as possible hardware acceleration. As a consequence, MIDI is used mainly for game clients that are severely constraint by memory / bitrate or complexity today.




*(a) Audio parts with music*




*(b) Audio parts with environmental sounds*

Figure 3: Interactive music with audio parts

Figure 3 (a) illustrates a traditional implementation of interactive music with four audio layers and two music themes. Thus being a somewhat simplified model, it demonstrates the possibility for non-linear performance with respect to mix (vertical dynamics) and sequencing (horizontal dynamics). Hence, after playing *theme A* it is possible to either loop *theme A* again or move to *theme B*. The same concept can be applied to environmental sounds or "soundscapes" as exemplified in Figure 3 (b).

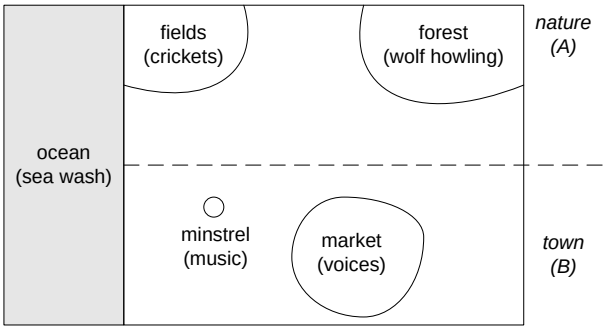



Figure 4: Game world example

In Figure 4, a 2D game world example is illustrated, where the corresponding soundscapes of the two areas, "nature" (A) and "town" (B), are represented by the multi-track loops as shown in e.g. Figure 3 (b). Similarly as in the interactive music application, the soundscapes can be dynamically changed and adapted to the game-play. The mixing and sequencing, typically based on a geometric distance model is here clearly demonstrated as the player moves between the two

scenes, (A) and (B), and changes proximity to the different objects. As can been observed, the sea wash part is common for both (A) and (B), which in this case facilitates a seamless transition between the two audio scenes. An alternative to dynamic mixing based on geometrics is to apply a mix that just randomly varies in time. This technique is sometimes used to make a short sound loop appear less repetitive.

The corresponding SAOC implementation of the non-linear music and soundscape concept previously outlined is illustrated in Figure 5. Whereas the sequencing functionality works exactly the same way as before, the implementation of the mixing is radically different. According to the principles of SAOC, all the audio parts within "environment" or "theme" (A) or (B) are downmixed into one mono or stereo track, here labeled "mix A" and "mix B". Along with the audio track the SAOC data is available, enabling re-mixing of the audio parts without fully decomposing them. It should be noted that the re-mixing also includes spatial re-rendering in case the output format is stereo or even higher channel modes.

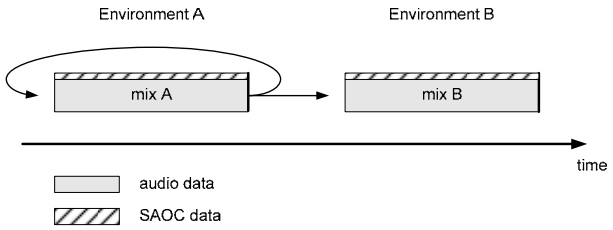



Figure 5: virtual audio parts using SAOC

An obvious benefit of the SAOC implementation is the memory saving as a potentially large number of audio tracks are reduced to one mono or stereo track. This is possible because the data-rate of the SAOC data is far lower than the data-rate of the audio data. Memory consumption also much depends on which audio codec is used for the audio parts or downmix. However, since the SAOC system is agnostic to the choice of downmix codec, it will always result in data reduction. Another benefit of the SAOC implementation is a strong potential playback complexity saving due to the reduced number of audio codec channels running in parallel. This is a rather typical scenario as long audio tracks in games usually are compressed with an audio codec. Furthermore, the complexity of a large number of decoders easily exceeds the extra overhead required by the SAOC tool. If the audio structure of a game application allows, music parts and soundscape parts could even be downmixed together to fully exploit the advantages of SAOC. However, excessive modifications of the mix often drop object separation and general audio quality. SAOC is therefore preferably deployed where full audio object separation is not a necessity. For instance, if a music-off / SFX-off functionality is desired, music and soundscapes should probably be derived from separate downmixes. Nevertheless, even for the case of common downmix, the subjective quality of separated signals can be increased by means of adding waveform coded components.

### 1.2 Online Voice Communication (voice chat)

In Massively Multiplayer Online gaming (MMO) and other social-oriented virtual worlds, live voice communication or "voice chat" between the players has become a popular phenomenon. As the application itself on a basic level is very similar to a teleconference application, it comes with two additional challenges. Firstly, to achieve a realistic spatial rendering of the voices according to game coordinates. This is typically applicable for first-person shooters where the player and his character both share the same correct orientation in the game world. Secondly, since indeed several thousands of players could share the same MMO game world, voice mixing and other server functions face practical problems due to large scales. There are several solutions available, both external generic voice chat tools such as various instant messaging systems, and voice chat systems with possible embedding in games such as Xbox LIVE and Dolby Axon. Common for all systems is the server's responsibility to control voice mixing and distribution to the clients. A corresponding SAOC implementation also relies on this server functionality. However, multiple voice channels and SAOC bitstreams can be downmixed separately, hence avoiding an additional SAOC decoding and re-encoding. For large scale cases the server still needs to exclude players, leaving only a smaller number of relevant voice channels in the SAOC bitstream and downmix. Still, many voices are distributed to each client for independent spatial rendering given a low bitrate.

### 1.3 Application Scenarios

Mobile games can be characterized by the presence of strong limitations on the available memory capacity and computational complexity of the game application. These are typical requirements for many gaming platforms, including Flash (e.g. Adobe Flash Lite™), Java-based games and mobile gaming in general. On such platforms, audio rendering capabilities are often kept to a simplistic level in order to limit the size of stored audio data and decoding complexity. Therefore, it is a desired objective to develop and implement a powerful interactive multi-channel audio handling technology as demonstrated in example 1 below.

Similarly, since in online games a large amount of audio data could be continuously transmitted or exchanged

there is a potential problem due to restrictions on the capacity of transmitting data channels. Nowadays, the vast majority of mobile phones are capable to access Internet, thereby enhancing the possibility for online mobile games supporting streaming audio and a rich set of user interactivity. The development of advanced web-based technologies such as Flash and Java allow for complex web browser games as well. Therefore, as demonstrated in example 2 below, an efficient technical solution of the functional audio data exchange mechanism suitable for the online gaming is demanded by the game developing industry.

**Example 1: SAOC Application for Mobile Games**

The SAOC technology application can be demonstrated using the following example of a game simulating farm activities.


Figure 6: Farm simulation game (© by Alawar Entertainment)

Figure 6 shows a game scene containing several animated figures representing farm animals and a farmer. According to the game scenario, all animals can move around and occasionally they utter their characteristic sounds. The farmer character that represents the player can also continuously change its position. In addition, some ambient sounds reflecting the environment of this virtual farm are present and background music is played. For the described game situation the appropriate rendering of all individual audio sources is beneficial for making the game atmosphere realistic and attractive. For the considered example, the SAOC technology allows implementation of the game sound handling using a single coded downmix signal. This signal contains all audio objects, which appear at any particular location in the game scene. At the decoder side, the SAOC data is either transcoded into an MPEG Surround (MPS) compatible representation for multi-channel playback or decoded directly for mono, stereo or binaural (3D headphone virtualizer) output, which would typically be the case for mobile gaming. This process is controlled by the dynamic rendering describing the level mapping from the audio objects to the playback channels. The corresponding mapping function implemented in the game core unit can include (among other parameters) coordinates of all present audio objects (farm animals) and position of the character representing the listener (farmer).

**Example 2: SAOC Application for Online Games**

The SAOC technology can be successfully applied to online games. The following example demonstrates one possible application. Let us consider the online game scenario, which allows voice chat functionality during game playing. In this example it is required that the players' voices are interactively rendered according to their positions in the virtual environment of the game scene (see Figure 7).


Figure 7: Office simulation game (© by Alawar Entertainment)

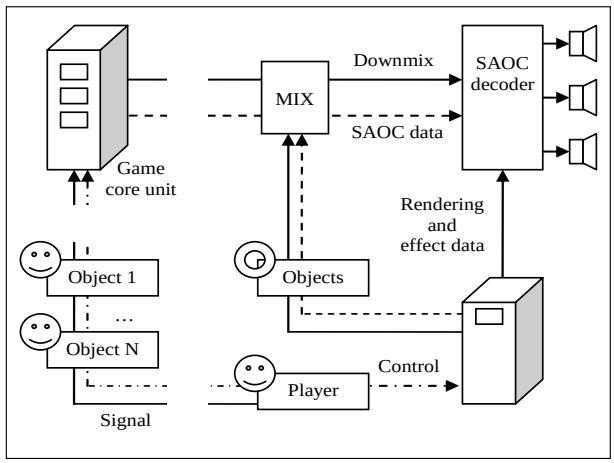



Figure 8: SAOC application for online games

All online players continuously transmit audio streams containing their coded voices and other game-relevant control data to the remote game server. This game server produces the corresponding data stream for each individual active game user. Next to the game control data, this stream contains the audio signal related part, consisting of the coded downmix and corresponding SAOC data. The local game client can add and mix additional audio objects, which are typically stored locally (e.g. background music, sounds of different objects present in the game scene, etc.) to the incoming downmix signal. This can be realized using the Multipoint Control Unit (MCU) functionality, which is able to modify the initial SAOC data without de/re-encoding. The local game client transforms the received game control information and commands of the local game user (e.g. coordinates of player's positions) into rendering settings and data describing audio effect, which should be applied to the resulting mix. The data, modified downmix signal and corresponding SAOC parameters are provided to the SAOC decoder, which produces the final audio output for the desired playback configuration (see Figure 8).

### 1.4 Benefits for SAOC Technology

The main advantages of the SAOC technology for the considered game examples are moderate bitrate consumption for the transfer of multiple interactively-controllable audio objects and low decoder/rendering complexity. The total bitrate of the audio related part is generally defined by the downmix core coder bitrate and is almost independent of the number of controllable audio objects.

The SAOC decoder complexity depends mainly on the type of playback configuration rather than on the quantity of coded input objects (see Figure 9).

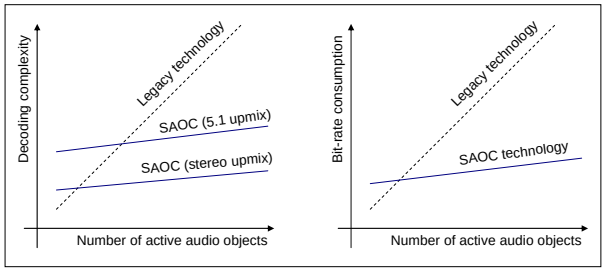



Figure 9: Benefits for SAOC relative to technology using multiple discrete audio track processing

## 2 TECHNICAL DESCRIPTION OF SAOC

### 2.1 Background and Concept

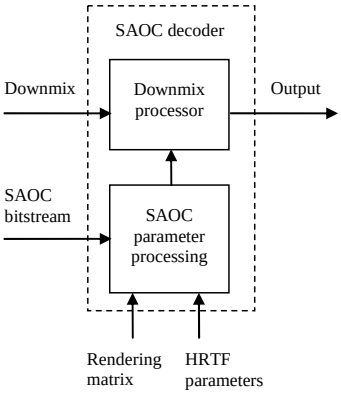
Spatial audio coding technology, such as the MPEG Surround (MPS) standard [2], has introduced a new paradigm of highly efficient multi-channel audio processing. For complex audio content, this technology provides extremely high compression rates and computationally efficient rendering. Building upon this technology, in January 2007 MPEG issued a Call for Proposals for SAOC system. After evaluation of the responses the present SAOC approach has been chosen as Reference Model (RM) [3]. The resulting SAOC work item provides a wealth of flexible user-controllable rendering tools based on transmission of conventional downmix extended with parametric audio object side information. At the receiver side, the object parameters are combined with the user-controllable rendering matrix, describing the (level) mapping from the audio objects to the playback channels. Depending on the intended output channel configuration, the receiver module may either act as a decoder that directly generates a mono, stereo or binaural (reproduced over headphones) output signal, or it may incorporate a transcoder module and re-use an MPS decoder as rendering engine yielding a 5.1 multichannel output signal. The task of transcoding consists of transforming the SAOC bitstream and rendering information associated with each audio object to a standard compliant MPS bitstream. The system also facilitates the integration of effects, parametric Head Related Transfer Function (HRTF) processing for binaural playback and an enhanced mode for improved object separation (e.g. Karaoke-type applications).

### 2.2 SAOC Architecture

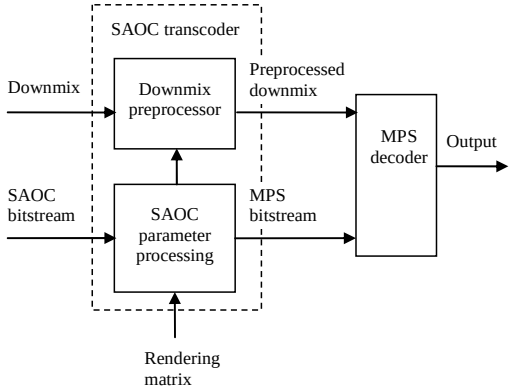
The SAOC system processes different types of audio objects (mono, stereo or multi-channel) into a specified (mono or stereo) downmix signal and the SAOC data. A hybrid Quadrature Mirror Filter (QMF) bank is used for enabling frequency selective processing of all parameters. The filterbank structure and frequency grid defining the parameter bands on which the parameters operates on, are re-used from previous MPEG technologies such as Parametric Stereo [4][5] and MPEG Surround [2]. Each data frame of the SAOC bitstream contains one or more sets of parameters for each parameter band, where every set corresponds to a certain block of samples in time. The concept of SAOC is to extract perceptually relevant cues like Object Level Differences (OLD) and Inter-Object cross Coherences (IOC) from the input audio objects. The downmix information is retained by Downmix Gains (DMG) and for a stereo downmix additional Downmix Channel Level Differences (DCLD). These parameters are quantized and entropy coded yielding the SAOC data being transmitted in the ancillary data portion of the downmix bitstream. The overall size of the SAOC bitstream depends on the number and type of input objects. It approximately constitutes of 3 kbit/s per

object using high parameter resolution, but could also be significantly lowered for certain applications.

Figure 10 shows block diagrams of the SAOC processing architecture. Plot (a) illustrates the decoder mode. The SAOC parameter processing engine decodes the SAOC bitstream and has an interface for additional input of time-variant rendering information and HRTF parameters. The downmix processing module directly provides the output signal in a mono, stereo or binaural configuration by applying these user-specified parameters and transmitted SAOC data to the corresponding downmix signal. In plot (b) the transcoding mode structure is depicted. Its architecture consists of an SAOC parameter processing engine and a downmix preprocessing module followed by an MPS decoder. Again, the SAOC parameter processing engine prepares parameters for the downmix preprocessor and provides a standard compliant bitstream to an MPS decoder. This transcoding functionality is what enables SAOC to perform a mix in 5.1 format.




*(a) SAOC decoder processing mode*




*(b) SAOC transcoder processing mode*

Figure 10: Architecture of SAOC system

For application scenarios requiring high level of amplification or attenuation of individual audio objects, an enhanced downmix preprocessor mode can be activated at the SAOC decoder/transcoder side.

Figure 11 shows the architecture of the enhanced processor comprising an object separation block and a rendering unit. The object separation process is done by using a "One-To-N"/"Two-To-N" (OTN/TTN) unit, which extracts available audio object groups from downmix channels. Using the enhanced processor, it is convenient to classify all input audio objects into a static background object (BGO) and flexibly rendered foreground object (FGO) groups (e.g. background music and controllable game objects). These groups can be efficiently recovered from the common downmix signal by the OTN/TTN unit, which supports partially waveform coded objects [2]. In the subsequent stage they are appropriately rendered by the successive rendering unit.

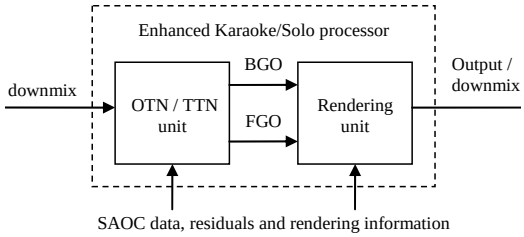



Figure 11: Architecture of the enhanced processor

Some applications request the possibility to combine a number of distributed clients. In general, this task is assigned to MCU. The SAOC concept incorporates an MCU functionality ensuring flexible merging of several audio objects on a parameter level, without the need for additional de/re-encoding of the downmix signal. As illustrated in Figure 12 the MCU combines the input SAOC bitstreams into one common SAOC bitsteam in a way that the parameters representing all audio objects from the input bitstreams are included in the resulting output bitstream.

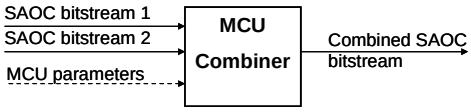



Figure 12: Outline of the SAOC MCU combiner

The binaural decoding mode facilitates the use of headphones as playback device. The SAOC scheme provides a parametric HRTF database interface and is computationally more efficient than the conventional approach. In SAOC the binaural rendering for all audio objects is realized by processing the single mono downmix signal with two HRTFs (two pairs of HRTFs for a stereo downmix signal), which are derived from a combination of rendering information, SAOC and

HRTF-database parameters. Full freedom in 3D positioning (including positioning outside the horizontal plane) is possible by mapping the objects to any combination of virtual loudspeakers represented by the HRTFs. Consequently, the incorporation of head tracking can be implemented by dynamically updating the render matrix according to head movements. The parametric approach ensures a compact representation of HRTFs yielding minimum storage requirements for mobile devices.

In addition, the flexibility of the SAOC technology is extended by an effects interface providing means for insert- and send-effects applied either in series or parallel to the signal processing chain. The SAOC effects interface provides means to incorporate room acoustic simulation in a very flexible manner and to reflect the influence of the virtual world surroundings.

## CONCLUSIONS

This paper introduces the SAOC concept as a novel interactive audio technology for games. SAOC offers joint transmission/storage and manipulation of multiple individual audio objects in an exceptionally efficient way compared to conventional game audio engines. Besides describing the basic SAOC architecture, this paper primarily focuses on the relevance for online and mobile gaming. Some typical application examples for the SAOC technology are presented and potential benefits of low bitrate consumption and computational complexity are discussed. Finally, the benefits of the SAOC technology is put in a context of today's rapid growing world of mobile and online games.

## REFERENCES


[1] "2005 Mobile Games White Paper", Presented at the Game Developers Conference by the IGDA Online Games SIG, 2005. http://www.igda.org/online/IGDA_Mobile_Whitepaper_2005.pdf

[2] "MPEG audio technologies – Part 1: MPEG Surround", ISO/IEC Int. Std. 23003-1:2007, 2007.

[3] J. Engdegård, B. Resch, C. Falch, O. Hellmuth, J. Hilpert, A. Hölzer, L. Terentiev, J. Breebaart, J. Koppens, E. Schuijers, W. Oomen, "Spatial Audio Object Coding (SAOC) - The Upcoming MPEG Standard on Parametric Object Based Audio Coding" 124th AES conv., Amsterdam, Netherlands, May 2008, pp. 7377.

[4] "Coding of audio-visual objects Part3: Audio, AMENDMENT 2: Parametric coding for high quality audio", ISO/IEC Int. Std. 14496-3:2001/Amd.2:2004, 2004.

[5] E. Schuijers, J. Breebaart, H. Purnhagen, and J. Engdegåard, "Low complexity parametric stereo coding", in AES 116th Convention, Berlin, Germany, May 2004.