SOFTWARE OPTIMIZATION OF THE MPEG-AUDIO DECODER USING A 32-BIT MCU RISC PROCESSOR

Keun-Sup Lee, Young Cheol Park[1] and Dae Hee Youn
MCSP Lab., Dept. of Electrical and Electronic Eng., Yonsei University, Seoul, Korea
[1]Dept. of Division of Information Technology, Yonsei University, Wonju, Korea

ABSTRACT

This paper proposes optimization techniques to implement MPEG audio decoding algorithms in real-time using general-purpose 32-bit MCU RISC processor. Both MP3 and AAC LC profile decoding algorithms can be partitioned into two parts: control-intensive part and computation-intensive part. Optimization techniques in this study are focused on developing methodologies that are suitable for each part. We implemented MP3/AAC decoder using ARM-based RISC MPU and its test board. The both designed decoders with proposed optimizations could achieve decoding processes in real-time within an operating frequency of 35 MHz. And ISO 13818-4 compliance test results confirmed that the proposed decoders ensured full compliance with ISO 13818-3 audio decoder. These implementation results illustrate the efficiency of the proposed design methodology in both performance and cost.

1. INTRODUCTION

Recently, portable MPEG audio devices has been actively developed and already become as popular as portable Compact Disc (CD) or Mini-Disc (MD) devices in the consumer market. Since digital audio players do not require mechanical components needed for other devices such as CD, decoding processors consume a large portion of the battery power in use. Therefore, the power consumption of the decoding processors affects the battery life of the player.

Many portable players adopt DSP because of decoding complexity. But high power consumption of the DSP architecture brings short battery life. Moreover, complete digital audio decoding system may need an additional micro-controller for the bitstream parsing and overall system control. Therefore low power architecture suitable for portable devices is required.

As processor design technology develops, the performance of micro-controller unit (MCU) reaches old DSP level. If MPEG audio decoding algorithm can be implemented using MCU without DSP, it is possible to integrate system control functionalities with audio decoding on a single MCU processor. It would be beneficial to the overall power consumption of portable MPEG player. However, since MCU has architectural disadvantages such as simple and limited hardware resource, MCU-based implementation requires a great deal of optimization compared with DSP-based case. And different optimization techniques for DSP-based case are required.

This paper proposes optimization techniques to implement MPEG audio decoding algorithms in real-time using general-purpose 32-bit MCU RISC processor. Decoding algorithms implemented in the paper are MPEG-1/2 Layer-III (MP3) and MPEG-2 AAC (Advanced Audio Coding) Low Complexity profile[1][2]. We adopted ARM7TDMI architecture as the target 32-bit MCU RISC processor. MP3/AAC decoding software was written in C codes and compiled with the ARM C-cross compiler. But the compiled codes have much redundancy. According to implementation result, the C-compiled MP3 code with no optimization required about 60~70 MIPS to decode a 128 kbps 44.1 kHz stereo bitstream. Therefore, we performed further optimization for these codes.

The organization of this paper is as follows. In Section 2, the design methodology of MPEG-audio decoders is presented. Software optimization processes and techniques based on this proposed methodology are described in Section 3. In Section 4, we will report implementation results containing complexity and output audio accuracy. Finally, we summarize our research in Section 5.

Manuscript received June 24, 2002

0098 3063/00 $10.00 © 2002 IEEE

This material may be protected by Copyright law (Title 17 U.S. Code)

Google Exhibit No. 1063
Page 1318

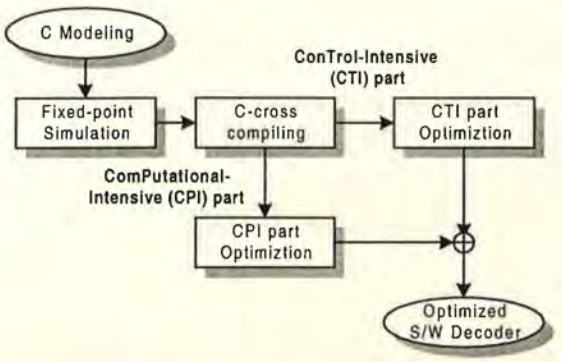


Figure 1. design methodology

2. DESIGN METHODOLOGY

The overall flow of the design methodology is illustrated in Figure 1. The flow describes implementation steps from C modeling to the optimized software decoder. In the fixed-point simulation stage, quantization errors caused by fixed-point operations are to be minimized. After the fixed-point simulation, C codes are converted into Assembly coded using a C-cross compiling tool. Since the resultant Assembly codes can contain much redundancy due to the inefficiency of the C-compiler, further optimization is required. As shown in Figure 1, there exist two distinct paths regarding algorithmic properties.

Figure 2 shows MP3 and AAC decoding flows. Both MP3 and AAC decoding algorithms consist of bit parsing, bit allocation decoding and unique synthesis filtering of their own. As shown in Figure 2, each decoding algorithm can be partitioned into two parts, which are control-intensive part and computation-intensive part. Control-intensive part consists of various operations related to control program flow. They have relatively low computational loads, but are difficult to control. On the contrary, Computation-intensive part has regular operations of high computational loads. Considering these algorithmic properties, distinct optimization technique must be performed for each part of the algorithmic. The optimized software decoder is obtained from these two distinct optimizations.

3. DESIGN OPTIMIZATION

Before the main optimization, we performed C-code level optimization to improve the performance of the code. We modified the code to be more suitable for the ARM processor and to help the compiler generate better code.

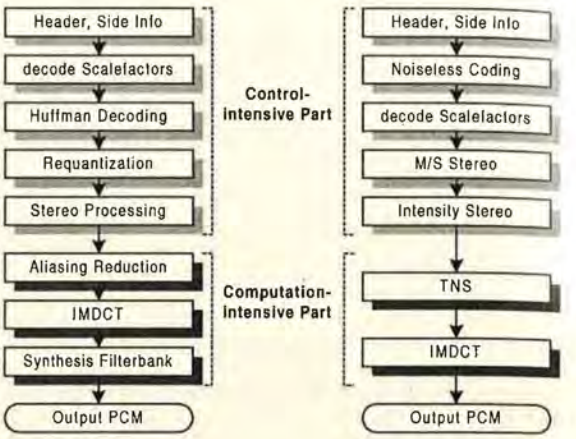


Figure 2. Partitioning MP3 and AAC decoding algorithm

3.1. C-compiler based Implementation

In general, assembly code from C-compiler has much redundancy. So, in order to get more efficient assembly code, hand assembly coding technique is required. But hand assembly coding requires so much time, and therefore, development cost increases.

Although C-compiler produces much redundancy, the redundancy from compiler varies according to processor architecture. Figure 3 shows the result of performance comparison between hand-optimized code and C-cross compiled assembly code of ARM7 Core. Target C code is Huffman decoding core routine, which belongs to control-intensive part.

According to Figure 3, two assembly codes are the same and the efficiency of C-cross compiler is very high. It is because of simple architecture and minimum resources of MCU processor that these results are obtained. Few instructions and resources due to simple architecture reduce the flexibility of assembly coding. So the optimizer of MCU compiler has more efficiency than that of DSP compiler with versatility.

Therefore, in this paper we compiled C code using C-cross compiler firstly, and then performed further optimization.

Google Exhibit No. 1063
Page 1319

Original C-code

```
while( (x = *val++) < 0 ){
  if( Mask >= 0 )    val -= x;
  Mask <<= 1;
  num++;
}
```

Hand Assembly Code

```
    B ReadX
Loop:
    CMP Rd, #0
    SUBGE pHuff, pHuff, Re, LSL #2
    MOV Rd, Rd, LSL #1
    ADD Rb, Rb, #1
ReadX:
    LDR Re, [pHuff], #4
    CMP Re, #0
    BLT Loop
```

Cross Compiled Code

```
0xf894:B 0xf8a8
0xf898:cmp r4, #0
0xf89c:subgr r12, r12, r0, lsr r3
0xf8a0:mov r4, r4, lsl #1
0xf8a4:add r14, r14, #1
0xf8a8:ldr r0, [r12], #4
0xf8ac:cmp r0, #0
0xf8b0:blt 0xf898
```

***Figure 3.** Comparison between two assembly coding methods*

3.2. Control-intensive Part Optimization

The control-intensive part of MPEG audio algorithm accompanies instructions related to the program flow control such as loop, conditional branch, and so on. Of all control-intensive routines Huffman decoding and scalefactor unpacking routines are very complex routines.

Huffman decoding consists mainly of bit unpacking and Huffman table searching routine, and these routines are iterated. Since bit unpacking and table search routines refer to bitstream buffer and ROM tables respectively, Huffman decoding routine requires to access memory frequently. These frequent memory accesses degrade the performance of overall system based on MCU, which has a weak point in memory access cycle. To minimize the number of bitstream buffer access, we use in-line hand-assembly to read maximum bits of Huffman codebook at once from bitstream buffer. This implementation reduces unpacking function call and bitstream buffer access overheads in iteration loop.

It is bit unpack routine that extracts Huffman code and scalefactor from bitstream buffer. Since bitstream buffer is finite size, to extract information from it circular addressing mode is required. But most of MCU's don't have complex addressing modes such as circular and bit-reverse addressing mode. So, to implement circular addressing with MCU additional computational loads due to if-clause are required. Whenever bit unpack routine is called, it must be checked whether bitstream read point reaches the end of bitstream buffer.

In order to remove this additional routine, we modified bitstream buffer control method as shown in Figure 4. By shifting unused bitstream to the first position at the start of each frame decoding, read point can be prevented from reaching the end of buffer. Considering the number of bit unpacking routine call during decoding process, complexity for shifting unused bitstream is very low. Furthermore, since bit unpacking routine called very often during control-intensive routines, the loads of subroutine call are somewhat heavy. So, it is advantageous to apply inline function to bit unpacking routine.

3.3. Computation-intensive Part Optimization

In MPEG audio decoding algorithm, frequency to time mapping synthesis filterbank routines come under computation-intensive part. For heavy computational loads of these routines fast algorithm commonly applies to them.

From the algorithmic point of view, most of operations are memory access, addition, and multiplication rather than branch. For example, as shown in Equation 1 one MAC operation needs one addition, one multiplication and three memory accesses that are two loads and one store.

$$y(i) = \sum_{n} x(n) \times c(n,i) \qquad (1)$$

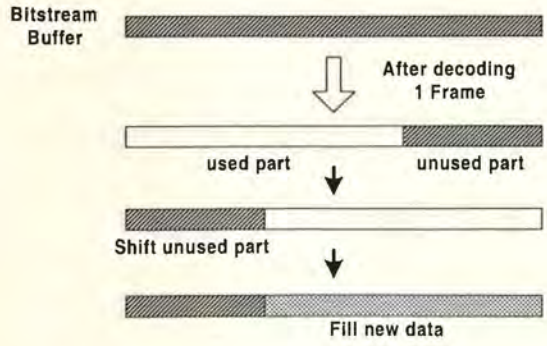


***Figure 4.** Modified bitstream buffer control method*

Google Exhibit No. 1063
Page 1320

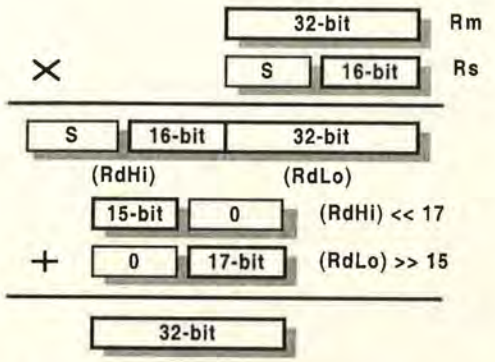


Figure 5. *32-bit x 16-bit fixed-point multiplication*

Considering that MAC is the major operation of MPEG audio synthesis filterbank, computational loads for memory access are dominant. And due to low performance of multiplier in MCU multiplication also requires many clock cycles. Therefore we concentrated optimization on reducing the computational loads of memory access and multiplication.

Many MCU processors support multiple memory access operation. Multiple memory access makes it possible to load from and store to memory within less clock cycles than sequential memory access. For example, although ARM7 processor is capable of implementing a memory load in 3 cycles and store 2 cycles, by organizing data so that memory access is via consecutive addresses it is possible to achieve effective load and store operations in about single-cycle[3]. So with this multiple access the performance of synthesis filterbank can be significantly enhanced.

As mentioned above, computation-intensive routines consist of many multiplications. Since multiplication in MCU is a heavy burden, before the optimization for this we adopted fast algorithms that have few multiplications. Contrary to DSP, many MCU processors cannot operate full-length multiplication in 1 clock cycle due to simple multiplier. For that reason, many MCU processors support partial multiplication such as 32-bit x 8-bit. But, since multiplication with short data length can result in degrading data precision, it is important to determine the length of multiplicand and multiplier. Previous work showed that 32-bit x 16-bit multiplication is enough to obtain good audio output quality in MPEG audio decoder[4]. Therefore, considering not only data precision but also complexity, we adopted 32-bit x 16-bit. 16-bit data is mainly used for coefficient table. Figure 5 shows fixed-point multiplication operation we used. These operations require 5 clock cycles. But, since MAC operation don't need to execute last two instructions for extracting high product before the accumulation is completed, the computational loads for multiplication is smaller than 5 clock cycle in many cases.

4. IMPLEMENTATION

4.1. Target RISC Processor Architecture

In order to determine target MCU RISC processor for implementing MPEG audio decoder, next two requirements must be considered; the first requirement is high operational performance of processor suitable for real-time decoding process, and the second one is high accuracy of decoder's PCM output.

For real-time implementation of audio decoder, the performance of MCU RISC processor must be able to reach at least 30 DSP MIPS[4]. Therefore, considering the architecture of general MCU, target processor is required to operate in very high clock frequency. And, since MPEG audio decoding algorithms consists of many MAC (Multiply-ACcumulation) operations, target processor must have a multiplier to reduce the computational loads.

The sound quality of decoder's PCM output is affected by the accuracy of decoding operation. Especially, since fixed-point ALU produces a lot of quantization noise, it is required to reduce these errors for the cost of increasing operational complexity. To minimize the both these efforts and additional complexity, 32-bit data path was chosen for fixed-point ALU architecture. Since long data path produces high resolution and large dynamic range, 32-bit data path has higher accuracy than 16-bit or 24-bit data path.

After considering above requirements, we adopted ARM7TDMI as target MCU RISC processor. ARM7TDMI satisfies above all requirements, and besides, now it is used for digital audio codec applications. The components of this processor are as follows[3]:

- 32-bit RISC Architecture
- 65 ~ 100 MHz clock frequency
- 31 x 32-bit registers and 6 status registers
- 32-bit x 8-bit multiplier and 32-bit ALU
- 32-bit Barrel shifter

Google Exhibit No. 1063
Page 1321

4.2. Implementation Result

In order to measure the sound quality for the output PCM of designed decoders objectively, ISO 13818-4 compliance test was performed[5]. According to Table 1 both MP3 and AAC decoder don't exceed the criteria of compliance test, and therefore, both satisfy ISO 13818-4 full compliance.

Table 2 summarizes the computational complexity of designed software decoders. For MP3 bitstream popular music signal and for AAC bitstream white noise signal are used for performance test. All two bitstream have the specification of 2 channels, 128 kbps and sampling frequency of 44.1 kHz.

In order to verify these results, we implemented MP3/AAC decoder using Samsung S3C4510B01 ARM-based RISC MPU and its test board. According to this implementation results, the both designed decoders with proposed optimizations could achieve decoding processes in real-time within a maximum operating frequency of 35 MHz. According to software libraries fully optimized by ARM Ltd., the operating frequency of MP3 decoder is 29 MHz and that of AAC decoder is 22 MHz (without TNS) and 28 MHz (with TNS)[6]. Considering that the proposed methodology is the optimization in C-cross compiling level with low development cost, the performance of designed software decoder is comparable with that of ARM's software libraries. These implementation results illustrated the efficiency of the proposed optimization techniques in both performance and cost.

Table 1. *Results of compliance test*

	RMS of Error	Maximum Error
Full Compliance	**8.8097×10^{-6}**	**6.1035×10^{-5}**
MP3	**4.7483×10^{-6}**	**2.8729×10^{-5}**
AAC	**5.6012×10^{-6}**	**3.0518×10^{-5}**

Table 2. *Maximum complexity of designed decoder*

MP3	Complexity	AAC	Complexity
Bitstream	**17 %**	Bitstream	**18 %**
Decoding	**5 %**	Decoding	**6 %**
IMDCT	**16 %**	TNS	**22 %**
Filterbank	**62 %**	IMDCT	**54 %**
Total	**33.4 MHz**	Total	**32.6 MHz**

5. CONCLUSIONS

In this paper, we proposed optimization techniques for implementing MPEG audio decoding algorithms in real-time using a general-purpose 32-bit MCU RISC processor. Target MPEG audio algorithms were MPEG-2 Layer-III (MP3) and MPEG-2 AAC Low Complexity profile. Considering algorithmic properties, both MP3 and AAC decoding algorithms could be partitioned into two parts, which are control-intensive part and computation-intensive part. Therefore, we proposed optimization techniques suitable for each part.

Test result confirmed that the proposed decoding algorithm ensured full compliance with ISO 13818-3 audio decoder. According to the implementation results, the both designed decoders with proposed methodology could achieve decoding processes in real-time within an operating frequency of 35 MHz. These implementation results illustrated the efficiency of the proposed optimization techniques in both performance and cost.

REFERENCES

[1] ISO/IEC 13818-3 "Generic Coding of Moving Pictures and Associated Audio (Part 3: MPEG-Audio)", 2nd Edition, Feb., 1997

[2] ISO/IEC JTC1/SC29/WG11 No.1650 "IS 13818-7 (MPEG-2 Advanced Audio Coding)", Apr., 1997.

[3] ARM, *ARM7TDMI Data Sheet*, 1995

[4] Keun-Sup Lee, Hyen-O Oh, Young-Cheol Park, and Dae Hee Youn, "High Quality MPEG-Audio Layer III Algorithm for a 16-BIT DSP", in Proc. IEEE ISCAS, vol. II, pp.205-208, May 6-9, 2001

[5] ISO/IEC 13818-4 "Generic coding of moving pictures and associated audio information (Part 4: Compliance testing)"

[6] ARM Ltd. Homepage (http://www.arm.com)

Google Exhibit No. 1063
Page 1322

BIOGRAPHY



Keun-Sup Lee was born in Seoul, Korea, on 1975. He received his B.S. and M.S. degrees in Electronic Engineering from Yonsei University, Seoul, Korea in 1997 and 1999 respectively. He is currently a Ph.D. candidate in Department of Electrical and Electronic Engineering at Yonsei University, Seoul, Korea. His research interests include audio signal processing, DSP & ASIC implementation, high quality audio coding, HDL chip design, and algorithm optimization.



Young Cheol Park received his B.S., M.S., and Ph.D. degrees in Electronic Engineering from Yonsei University, Seoul, Korea in 1986, 1988, and 1993 respectively. From 1993 to 1995, he was a Post Doctoral Research Scholar of the graduate program in acoustics in Pennsylvania State University, PA, USA. From 1996 to 1998, he was with Samsung Electronics Co., Ltd., Kiheung Korea, working on the development of digital hearing aids. In 1999 he joined InTime Corp. in Seoul, Korea. His research interests include digital audio, speech coding, and adaptive filters. He is currently an assistant professor in Yonsei University, Wonju, Korea.



Dae Hee Youn received his B.S. degree in Electronic Engineering from Yonsei University, Seoul, Korea, in 1977, and the M.S. and Ph.D. degrees in Electrical Engineering from Kansas State University, Manhattan, Kansas, in 1979 and 1982, respectively. From 1982 to 1985, he was an assistant professor at the University of Iowa, Iowa City, Iowa. Since 1985, he has been with the Department of Electrical and Electronic Engineering at Yonsei University, Seoul, Korea, where he is currently a professor. His research interests include adaptive digital filter and its application, speech and audio signal processing, and real-time implementation of DSP algorithms.

Google Exhibit No. 1063
Page 1323

Ex. 1040B

LIBRARY | CATALOG

1 of 1

PERIODICAL OR NEWSPAPER

IEEE transactions on consumer electronics

Full Record MARC Tags

Main title
IEEE transactions on consumer electronics.

Published/Created
New York : Institute of Electrical and Electronics Engineers, ©1975-

Request this Item | LC Find It

More Information

LCCN Permalink	https://lccn.loc.gov/75643472
Publication history	Vol. CE-21, no. 1 (Feb. 1975)-
Description	volumes : illustrations ; 28 cm
Current frequency	Quarterly, Feb. 1980-
Former frequency	Quarterly, Feb. 1975-May 1979 Five times a year, July 1979-Nov. 1979

Request this Item | LC Find It

Google Exhibit No. 1063
Page 1325

LC classification	TK6563 .I2
Continues	IEEE transactions on broadcast and television receivers 0018-9308 (DLC) 57039724 (OCoLC)1604343
Variant title	Institute of Electrical and Electronics Engineers transactions on consumer electronics
Serial key title	IEEE transactions on consumer electronics
Abbreviated title	IEEE trans. consum. electron.
Related names	IEEE Consumer Electronics Society. IEEE Broadcast, Cable, and Consumer Electronics Society. IEEE Consumer Electronics Group. IEEE International Conference on Consumer Electronics.
LC Subjects	Household electronics--Periodicals. Radio--Receivers and reception--Periodicals. Television--Receivers and reception--Periodicals. Electronic apparatus and appliances--Periodicals.
Other Subjects	Electronic apparatus and appliances. Household electronics. Radio--Receivers and reception. Television--Receivers and reception.
Form/Genre	Periodicals. Periodicals.
Browse by shelf order	TK6563
Notes	Title from cover. Some volumes have supplements. Issues for <Nov. 1982, Aug. 1985> include some papers from International Conference on Consumer Electronics. A publication of: the IEEE Consumer Electronics Group, <Feb. 1976>-May 1976; IEEE Broadcast, Cable, and Consumer Electronics Society, Aug. 1976-<Nov. 1982>; IEEE Consumer Electronics Society, <May 1984->
Additional formats	Online version: IEEE transactions on consumer electronics (Online)

Request this Item | LC Find It

Google Exhibit No. 1063
Page 1326

LCCN	75643472
CODEN	ITCEDA
Dewey class no.	621.38
National bib agency no.	PH 24/03/94
Other system no.	(OCoLC)ocm02241216
Additional Links	http://ieeexplore.ieee.org/servlet/opac?punumber=30
Type of material	Periodical or Newspaper
Content type	text
Media type	unmediated
Carrier type	volume

Item Availability

CALL NUMBER	TK6563 .I2 Set 1
Request in	Jefferson or Adams Building Reading Rooms
Status	Not Charged
Older receipts	v.22 (1976), v.24 (1978), v.27-v.61 (1981-2015), v.62:no.2-v.62:no.4 (2016:May-2016:Nov.), v.63:no.2 (2017:May). v.63:no.4-v.64:no.4 (2017:Nov.-2018:Nov.)
Supplements	v.30:no.2 (1984:May)

CALL NUMBER	TK6563 .I2 Copy 1 Unbound issues

Request this Item | LC Find It

Google Exhibit No. 1063
Page 1327

Status	Not Charged
CALL NUMBER	TK6563 .I2 Copy 2 Unbound issues
Request in	Newspaper & Current Periodical Reading Room (Madison LM133)

Request this Item LC Find It

Google Exhibit No. 1063
Page 1328

LIBRARY CATALOG

1 of 1

PERIODICAL OR NEWSPAPER

IEEE transactions on consumer electronics

Full Record MARC Tags

000		04336cas a2200877 a 4500
001		11189359
005		20240917071747.0
008		751021c19759999nyuqr p 0 a0eng
010	__	|a 75643472
016	00	|a PH |z 24/03/94
022	0_	|a 0098-3063 |l 0098-3063 |2 1
030	__	|a TCEDA
032	__	|a 256480 |b JSPS
035	__	|a (OCoLC)ocm02241216
040	__	|a DLC |b eng |c DLC |d VSD |d OCL |d NSD |d DLC |d NST |d OCL |d NSD |d AIP | d OCL |d NST |d DLC |d NST |d DLC |d OCL |d DLC |d OCL |d NST |d NSD |d NST | d HUL |d VSD |d MYG |d CUS |d MYG |d OCL |d LYU |d IEEEX |d OCLCQ |d ORE | d VSD |d _VB |d OCLCQ |d OCLCO |d OCLCF |d OCLCO |d IUL |d OCLCQ |d BRF | d OCLCO |d PUL
042	__	|a pcc |a nsdp
050	00	|a TK6563 |b J2
082	04	|a 621.38 |2 19
210	0_	|a EEE trans. consum. electron.
222	_0	|a EEE transactions on consumer electronics
245	00	|a EEE transactions on consumer electronics.
246	3_	|a nstitute of Electrical and Electronics Engineers transactions on consumer electronics
249	0_	|i ACQUIRE Title: |a IEEE transactions on consumer electronics.
260	__	|a NewYork : |b Institute of Electrical and Electronics Engineers, |c ©1975-

Google Exhibit No. 1063
Page 1329

300	__	|a volumes : |b illustrations ; |c 28 cm
310	__	|a Quarterly, |b Feb. 1980-
321	__	|a Quarterly, |b Feb. 1975-May 1979
321	__	|a Five times a year, |b July 1979-Nov. 1979
336	__	|a text |b txt |2 rdacontent
337	__	|a unmediated |b n |2 rdamedia
338	__	|a volume |b nc |2 rdacarrier
362	0_	|a Vol. CE-21, no. 1 (Feb. 1975)-
500	__	|a Title from cover.
525	__	|a Some volumes have supplements.
550	__	|a Issues for <Nov. 1982, Aug. 1985> include some papers from International Conference on Consumer Electronics.
550	__	|a A publication of: the IEEE Consumer Electronics Group, <Feb. 1976>-May 1976; IEEE Broadcast, Cable, and Consumer Electronics Society, Aug. 1976-<Nov. 1982>; IEEE Consumer Electronics Society, <May 1984->
590	__	|a SERBIB/SERLOC merged record
592	__	|a ACQN: aq 94014910
592	__	|a 5 times a year
592	__	|a FRO, title update per FAXON list dated 9-4-86
650	_0	|a Household electronics |v Periodicals.
650	_0	|a Radio |x Receivers and reception |v Periodicals.
650	_0	|a Television |x Receivers and reception |v Periodicals.
650	_0	|a Electronic apparatus and appliances |v Periodicals.
650	_6	|a Appareils électroniques domestiques |v Périodiques.
650	_6	|a Radio |x Récepteurs et réception |v Périodiques.
650	_6	|a Télévision |x Récepteurs et réception |v Périodiques.
650	_6	|a Appareils électroniques |v Périodiques.
650	_7	|a Electronic apparatus and appliances. |2 fast |0 (OCoLC)fst00906772
650	_7	|a Household electronics. |2 fast |0 (OCoLC)fst00962119
650	_7	|a Radio |x Receivers and reception. |2 fast |0 (OCoLC)fst01087124
650	_7	|a Television |x Receivers and reception. |2 fast |0 (OCoLC)fst01146597
655	_7	|a Periodicals. |2 fast |0 (OCoLC)fst01411641
655	_7	|a Periodicals. |2 lcgft |0 http://id.loc.gov/authorities/genreForms/gf2014026139 |0 (uri) http://id.loc.gov/authorities/genreForms/gf2014026139
710	2_	|a IEEE Consumer Electronics Society.
710	2_	|a IEEE Broadcast, Cable, and Consumer Electronics Society.
710	2_	|a IEEE Consumer Electronics Group.

Google Exhibit No. 1063
Page 1330

711	2_	|a IEEE International Conference on Consumer Electronics.
776	08	|i Online version: |t IEEE transactions on consumer electronics (Online) |x 1558-4127 | w (DLC) 2005212495 |w (OCoLC)60639335
776	08	|i Online version: |t IEEE transactions on consumer electronics |w (OCoLC)752430525
780	00	|t IEEE transactions on broadcast and television receivers |x 0018-9308 |w (DLC) 57039724 |w (OCoLC)1604343
850	__	|a ArU |a CaBVaS |a DLC |a FU |a ICU |a IaAS |a MH-GM |a MoU |a NN |a NcD |a NcRS | a PPi |a PU |a TxArU |a ViBlbV
859	41	|u http://ieeexplore.ieee.org/servlet/opac?punumber=30
853	20	|8 1 |a v. |b no. |u 4 |v r |i (year) |j (month) |w q |x 02
863	41	|8 1.1 |a <34> |b <1> |i <1988> |j <02>
890	__	|a IEEE transactions on consumer electronics. New York. |i 75-643472
906	__	|a 7 |b cbc |c serials |d 3 |e ncip |f 19 |g n-oclcserc
920	__	|a Keep 1
923	__	|d 20020726 |s Group Registration
963	__	|a k.m.burke@ieee.org
984	__	|a srvf
985	__	|e eserial 200406
991	__	|b c-GenColl |h TK6563 |i .I2 |w SERIALS
991	__	|b o-NLS/REF |w ACQUIRE
992	__	|b 93000437 |w ACQUIRE
992	__	|b SER |w SERLOC

Request this Item | LC Find It

Item Availability

CALL NUMBER	TK6563 .I2 Set 1
Request in	Jefferson or Adams Building Reading Rooms
Status	Not Charged
Older receipts	v.22 (1976), v.24 (1978), v.27-v.61 (1981-2015), v.62:no.2-v.62:no.4 (2016:May-2016:Nov.), v.63:no.2 (2017:May). v.63:no.4-v.64:no.4

Google Exhibit No. 1063
Page 1331

(2017:Nov.-2018:Nov.)

Supplements	v.30:no.2 (1984:May)

CALL NUMBER	TK6563 .I2 Copy 1 Unbound issues
Request in	Newspaper & Current Periodical Reading Room (Madison LM133)
Status	Not Charged

CALL NUMBER	TK6563 .I2 Copy 2 Unbound issues
Request in	Newspaper & Current Periodical Reading Room (Madison LM133)

Google Exhibit No. 1063
Page 1332

Audio Engineering Society
Convention Paper 5871
Presented at the 115th Convention
2003 October 10–13 New York, NY, USA

This convention paper has been reproduced from the author's advance manuscript, without editing, corrections, or consideration by the Review Board. The AES takes no responsibility for the contents. Additional papers may be obtained by sending request and remittance to Audio Engineering Society, 60 East 42nd Street, New York, New York 10165-2520, USA; also see www.aes.org. All rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

A closer look into MPEG-4 High Efficiency AAC

Martin Wolters[1], Kristofer Kjörling[2], Daniel Homm[1], Heiko Purnhagen[2]

[1] *Coding Technologies, Deutschherrnstr. 15–19, 90429 Nürnberg, Germany*

[2] *Coding Technologies, Döbelnsgatan 64, 11352 Stockholm, Sweden*

Correspondence should be addressed to Martin Wolters (wol@codingtechnologies.com)

ABSTRACT

MPEG Spectral Band Replication (SBR) is the newest compression technology available as part of the MPEG standards. It is combined with MPEG Advanced Audio Coding (AAC) and improves coding efficiency by more than 30%. The resulting scheme is called High-Efficiency AAC (HE AAC).

This presentation explains MPEG SBR and its integration into the existing MPEG-4 bitstream format. The SBR technology itself as well as the implications on systems based on MPEG-4 technology are described. Signaling through MPEG-4 Systems and other transport formats is introduced and typical applications and usage scenarios are listed.

1 INTRODUCTION

MPEG-4 High Efficiency AAC (HE AAC) is not a replacement for AAC, but rather a superset which extends the reach of high-quality MPEG-4 Audio to much lower bitrates. High Efficiency AAC decoders will decode both plain AAC and the enhanced AAC plus SBR. The result is a backward compatible extension of the standard. Listening tests show that MPEG-4 HE AAC offers a significant benefit over proprietary codecs and over AAC without extensions, and place it clearly as the most efficient audio codec in existence. Coding Technologies and its customers have been using this technology for a number

of years under the *aacPlus* brand name.

The technical specifications of HE AAC are now available as a final draft amendment (FDAM) [1] and official standardization is expected in September 2003. These specifications enable manufacturers to work on final implementations and integrations. This paper provides an implementation guideline and provides an overview of the technologies and system integration aspects of HE AAC.

2 INTRODUCTION INTO MPEG-4

2.1 MPEG-4: A leading multimedia framework

In the context of evolving multimedia applications, new demands for efficient and flexible representation of audiovisual content have arisen. Besides high coding efficiency required to cope with the limited bandwidth of the internet or in mobile communication, new functionality like flexible access to coded data and manipulation by the recipient is also desired. To address these requirements and develop inter-operable solutions, ISO/IEC started its MPEG-4 standardization activities "Coding of audiovisual objects" [2], [3], [4]. Part 3 of the MPEG-4 standard (Audio) [5] provides tools for coding of natural and synthetic audio objects and composition of such objects into an "audio scene." Similarly, Part 2 (Visual) [6] provides tools for coding of natural video and synthetic 2D and 3D objects, while Part 1 (Systems) [7] covers general tools for scene description, synchronization of objects, and multiplexing for transport and storage.

2.2 MPEG-4 Audio Coding

MPEG-4 is based on the notion that the audio part of the audiovisual scene presented at the receiver is composed of one or more so-called *audio objects*. Different audio compression tools (codecs) are available to enable an efficiently coded representation of the audio object(s) in a scene. Natural audio objects, such as recorded speech and music, can be coded at bitrates typically ranging from 2 kbit/s (for narrowband speech) to 64 kbit/s/ch (for CD quality music) using parametric speech coding (HVXC), CELP-based speech coding, parametric audio coding (HILN) or transform-based general audio coding (AAC, TwinVQ). The natural audio and speech coding tools support bitrate scalability, also known as embedded coding. In addition, the parametric coding tools also provide speed and pitch change functionality in the decoder. Synthetic audio objects can be represented using

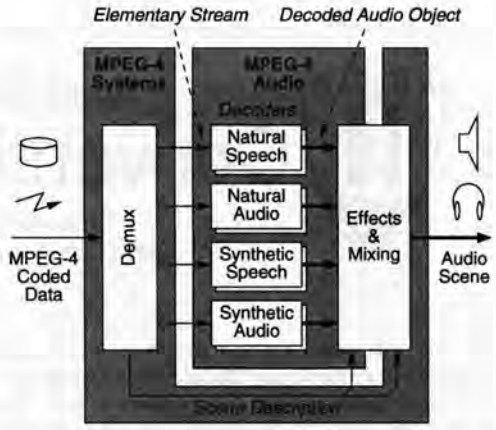


Figure 1: Block diagram of a complete MPEG-4 Audio decoder.

a Text-To-Speech Interface (TTSI) or the Structured Audio (SA) synthesis tools. Other uses of the SA tools are adding effects, like reverberation, and mixing different audio objects to compose the final "audio scene" that is presented to the listener.

Fig. 1 shows the block diagram of a complete MPEG-4 Audio decoder. It includes the decoding tools for the audio objects defined in Part 3 (Audio) of the MPEG-4 Standard [5] as well as bitstream demultiplexing and scene composition defined in the Systems part [7].

The information needed to decode an audio object at the decoder is conveyed by means of a so-called *elementary stream* (ES), as shown in Fig. 2. In case of bitrate scalable configurations, a base-layer ES and one or more enhancement-layers ES(s) are used. The initial ES_Descriptor contains an AudioSpecificConfig element, which carries the *audio object type* (AOT) and further information required to instantiate the required audio decoder. Using the ES_ID, the ES_Descriptor points to the stream of the actual compressed audio data, which is a sequence of so-called *access units* (AU), i.e., decodable bitstream frames.

A common audio object type is AAC LC, which indicates that MPEG-4 Advanced Audio Coding (AAC) in its low-complexity (LC) version is used. This is the primary general audio coder in MPEG-4, which is a compatible extension of the MPEG-2 AAC transform-based audio coder [8]. The principles of perceptual audio cod-

Google Exhibit No. 1041
Page 002

Google Exhibit No. 1063
Page 1334

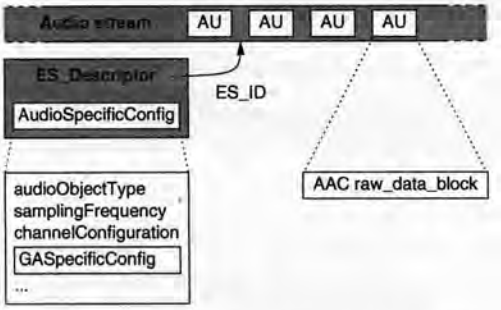


Figure 2: MPEG-4 Elementary Stream (ES) conveying an audio object.

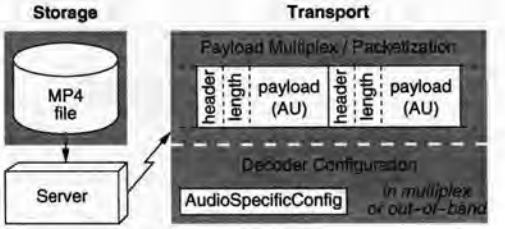


Figure 3: Storage and transmission of MPEG-4 Audio object with single elementary stream (ES).

ing and the specifics of the AAC are well documented in the literature, e.g. [9], [10], [11], [12], and hence not reviewed in this presentation.

2.3 Storage and streaming of MPEG-4

In order to allow flexible yet interoperable utilization of MPEG-4 coding technology for the largest possible range of application scenarios, MPEG-4 is designed as a transport-agnostic standard with an abstract interface to carry the compressed data, the so-called *sync layer* (SL). Basically, the SL adds mechanisms like time stamps for AUs to allow synchronization of several ESs, such as an audio and a video stream.

Depending upon the requirements of the application and the characteristics of the transmission channel, different schemes can be used to store or convey MPEG-4 content (Fig. 3). MPEG defined the MP4 file format [7], [13] for storage, which can hold all data available at the SL interface. When MPEG-4 content is carried over a transmission channel, various techniques for multiplexing, synchronization, and for establishing a connection can be appropriate. For inherently packetized channels, like those using the internet protocol (IP), no syncwords are needed and AUs can often directly be mapped to packets. A fixed framing might be available from the modulation scheme for radio channels, but random bit errors are possible. Serial channels might need some syncword mechanisms to establish basic synchronization. If the MPEG-4 content is carried in more than one ES, these streams need to be multiplexed, which can be done by the MPEG-4 FlexMux tool [7], [13].

To carry simple, audio-only MPEG-4 content with low overhead, the Part 3 (Audio) of the MPEG-4 standard defines a *Low-overhead MPEG-4 Audio Transport Multiplex* (LATM) to multiplex several MPEG-4 Audio payloads and AudioSpecificConfig() elements [5, subclause 1.7]. Based on LATM, it also defines a self-synchronized syntax of an MPEG-4 Audio transport stream which is called *Low Overhead Audio Stream* (LOAS).

A simple scheme to carry MPEG-4 over IP based on the realtime transport protocol (RTP) [14] is defined in RFC 3016 [15]. It is a subset of the general framwork defined in Part 8 (MPEG-4 over IP) of the MPEG-4 standard [16]. RFC 3016 uses LATM for audio and allows for conveying the AudioSpecificConfig either in-band using LATM or out-of-band, i.e., as a MIME parameter conveyed in the Session Description Protocol (SDP).

Other schemes to convey MPEG-4 content include MPEG-4 over MPEG-2 transport streams, as defined in an amendement to the MPEG-2 Systems standard, or proprietary schemes which provide an SL compatible interface.

MPEG-4 AAC LC coded data can also be stored using the *Audio Data Interchange Format* (ADIF) or transmitted using an *Audio Data Transport Stream* (ADTS), see Fig. 4. Both these formats are defined in the MPEG-2 AAC standard [8] but are only informative in MPEG-4, i.e., an MPEG-4 decoder is not required to support ADIF or ADTS.

2.4 Recent Developments: High-Efficiency Audio Coding becomes part of MPEG-4

Standardization of the technology submitted in response to the initial MPEG-4 Call for Proposals in July 1995 was carried out using the core experiment procedure and finalized in two steps: Version 1 in October 1998 [17]

Google Exhibit No. 1041
Page 003

Google Exhibit No. 1063
Page 1335

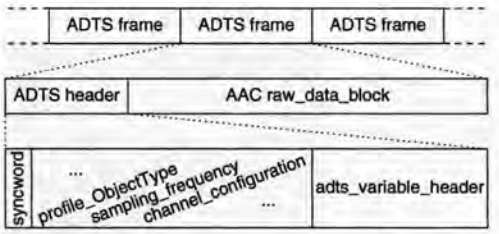


Figure 4: MPEG-2 Audio Data Transport Stream (ADTS).

and Version 2 in December 1999 [18]. In 2001, both versions and their corrigenda were merged to form a 2nd edition of the standard [5].

In July 2000, MPEG issued a Call for Evidence [19] in order to find out whether there are new developments which may provide improvements over the existing MPEG-4 standard. Based on the responses received, a Call for Proposals [20] for new tools for audio coding was issued in January 2001. It specifically asked for technology in the field of MPEG-4 compatible bandwidth extension and in the field of high quality parametric coding.

Coding Technologies submitted its Spectral Band Replication (SBR) techniques used in combination with AAC as bandwith extension tool. Based on a listening test, it was selected as inital reference model (RM0) and refined in the course of MPEG's core experiment procedure. In March 2003, the standardization of the MPEG-4 SBR tool was finalized [1].

3 AT THE HEART OF HE AAC: MPEG SBR

3.1 Fundamental Concept: High Frequency Reconstruction by Spectral Band Replication

The basic principles of SBR have been elaborated on in several papers [21], [22], [23], [24]. For the convenience of the reader a short review is given below, and some aspects not previously discussed are explained in more detail.

3.1.1 The SBR encoder

The SBR principle stipulates that the missing high frequency region of a lowpass filtered signal can be recovered based on the existing lowpass signal and a small amount of control data. The required control data is estimated in the encoder given the original wide-band signal. The HE AAC codec is a dual rate system, where the underlying AAC encoder/decoder is operated at half the sampling rate of the SBR encoder/decoder. The basic principle of the HE AAC encoder is depicted in Fig. 5.

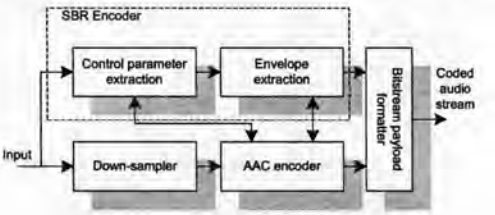


Figure 5: The HE AAC encoder.

In the SBR encoder, where the wide band signal is available, control parameters are estimated in order to ensure that the high frequency reconstruction results in a reconstructed highband that is perceptually as similar as possible to the original highband. The majority of the control data is used for a spectral envelope representation. The spectral envelope information has varying time and frequency resolution to be able to control the SBR process as good as possible, with as little bitrate overhead as possible. The other control data mainly strives to control the tonal-to-noise ratio of the highband. Fig. 6, Fig. 7, and Fig. 8, illustrate some of the characteristics of the control data.

In Fig. 6 the spectrum of the original signal at a given point in time is displayed at the top and the spectrum of the HF generated highband at the same point in time is displayed below. As is evident from the figure, the patching algorithm used to regenerate the highband was not successfull in generating all the strong tonal components of the original highband. The missing sinusoids in the regenerated highband are identified in the encoder and a parametric representation of them is incorporated in the SBR control data.

In Fig. 7 the spectra of the original and the HF generated signal is once again displayed albeit at a different point in time. Here it is evident that the tonal-to-noise ratio of the regenerated highband is not similar to that of the original. The highband of the original signal constitutes mainly of noise, whilst the highband of the regenerated signal is very tonal. This information is also incorporated into the

Google Exhibit No. 1041
Page 004

Google Exhibit No. 1063
Page 1336

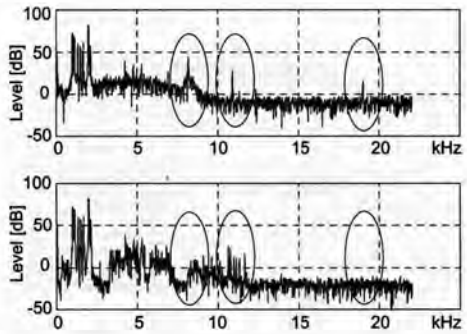


Figure 6: The spectrum of the original signal (top) and the HF generated signal without any additional signal components (bottom). The basic patching algorithm has not regenerated the strong tonal components of the original highband.

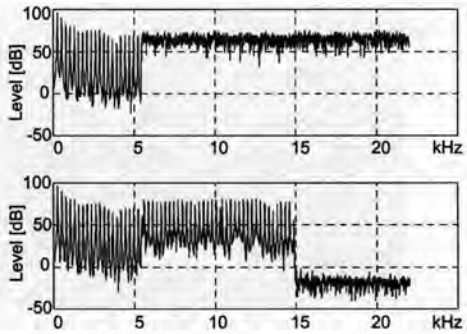


Figure 7: The spectrum of the original signal (top) and the SBR signal without any additional signal components (bottom). The basic high frequency reconstruction algorithm has not maintained the original tonal-to-noise ratio.

SBR control data so that the decoder by means of inverse filtering and noise addition can achieve a highband with a tonal-to-noise ratio similar to that of the original.

Finally, in Fig. 8, a spectrogram of the original signal is displayed with a superimposed time/frequency grid of the spectral envelope data transmitted to the decoder. Here it is evident that the time/frequency resolution of the spectral envelope varies over time, giving a higher time resolution for transient passages and a higher frequency resolution for stationary passages.

The bitrate of the control data varies depending on encoder tuning, but is in general somewhere in the region of 1-3 kbit/s per audio channel. This is far lower than the bitrate that would be required to code the highband with any conventional wave-form coding algorithm. The SBR data format is indicated in Fig. 13. A header flag indicates that an SBR header part is present. The SBR header part contains fundamental information such as SBR frequency range as well as control signals that do not require frequent changes. The sbr_data part can be subdivided into side info and raw data, where side info is defined as signals needed to decode the raw data and some decoder tuning signals. Raw data consists of Huffman coded envelope scalefactors and noise floor estimates.

3.1.2 The SBR decoder

The SBR enhanced decoder can be roughly divided into the modules depicted in Fig. 9.

All SBR processing is done in the QMF domain. Hence, the output from the underlying AAC decoder is firstly analyzed with a 32 channel QMF filterbank. Secondly, the HF generator module recreates the highband by patching QMF subbands from the existing lowband to the high band. Furthermore inverse filtering is done on a per QMF subband basis, based on the control data obtained from the bitstream. The envelope adjuster modifies the spectral envelope of the regenerated highband, and adds additional components such as noise and sinusoids, all according to the control data in the bitstream. Since all operations are done in the QMF domain the final step of the decoder is a QMF synthesis to retain a time-domain signal. Given that the QMF analysis is done on 32 QMF subbands for 1024 time-domain samples, and the high frequency reconstruction results in 64 QMF subbands upon which the synthesis is done producing 2048 time-domain samples, an up-sampling by a factor of two is obtained.

The following Fig. 10 displays the spectrum of the SBR signal at different stages in the SBR process. In the upper figure the spectrum of the signal after the high frequency reconstruction, but prior to the envelope adjustment is displayed. In the middle figure the spectrum of the output signal is displayed, and in the bottom figure the spectrum of the original signal is displayed for reference.

Google Exhibit No. 1041
Page 005

Google Exhibit No. 1063
Page 1337

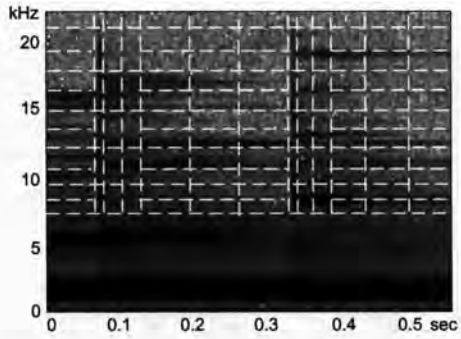


Figure 8: The spectrogram of a part of the input signal with the envelope time/frequency grid superimposed.

3.2 Low power SBR

Platforms with heavy constraints on the computational complexity such as portable devices can use Low power SBR. The main difference between High Quality SBR and Low power SBR is how the data is represented during the SBR process. High Quality SBR uses a complex representation of the subband samples and subsequently all calculations are computed with complex values. Low power SBR simply requires a real-valued representation and hence the computational complexity is heavily reduced.

The complex-valued filterbank is designed to prevent aliasing even in cases where modifications are performed [25]. However, the real-valued filterbank does not have that property. Therefore, an additional algorithm is introduced to minimize the aliasing occurring due to the real-valued filterbank. The aliasing reduction algorithm is devised to avoid introducing aliasing for strong tonal components.

The properties of the real-valued filterbank result in mirroring components (aliasing) in adjacent subbands if the adjacent subbands are given different gain-values in the envelope adjuster. This can occur if adjacent subbands are modified independently from each other. Hence, it is the objective of the aliasing reduction algorithm to identify the subbands where strong aliasing will be introduced if the subbands are modified independently. This is accomplished by observing the reflection coefficients obtained for every subband by first order linear prediction. A sign is introduced for every subband indicating whether a strong tonal component is situated in the upper (a positive sign) or lower (a negative sign) part of the subband. The aliasing reduction algorithm identifies the cases where the signs of two adjacent subbands are opposite, the sign being positive for the lower subband and negative for the higher subband. These subbands are grouped together, meaning that they will not be modified independently, and hence no aliasing will be introduced by the envelope adjustment in the real-valued filterbank for this signal. However, there are many signals where there are no strong tonal components that can be identified by the above outlined algorithm, but where aliasing is introduced nevertheless. This aliasing is audible when compared to the High Quality SBR system, but not as offensive as the aliasing that would occur for a strong tonal component if the aliasing reduction algorithm was not in place.

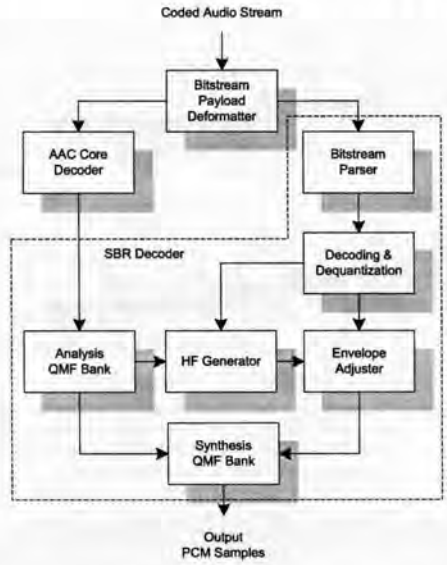


Figure 9: The HE AAC decoder.

3.3 Down-sampled SBR

Since the transform size in AAC is fixed to 1024 for long blocks and 128 for short blocks, it becomes evident that choosing the optimum sampling rate for a certain bitrate

Google Exhibit No. 1041
Page 006

Google Exhibit No. 1063
Page 1338

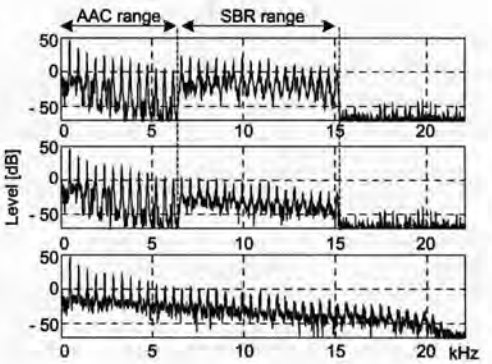


Figure 10: The spectrum of the audio signal after the HF generator (top), the spectrum of the audio signal after envelope adjustment (middle), and the spectrum of the original signal (bottom).

is crucial. In general, in the low-mid bitrate region, i.e., 20 - 32 kbit/s per audio channel, a sampling rate in the 22.05 – 32 kHz region is the best choice since the frequency resolution at the lower sampling rate improves the coding gain, compared to using a higher sampling rate. However the lower sampling rate provides a slower time response of AAC, and at higher bitrates the best trade-off between time and frequency resolution for the AAC codec may be obtained by using a higher sampling-rate. Furthermore, the low sampling rates naturally limits the highest frequency that can be covered by AAC.

The HE AAC codec is always operated in a dual rate mode. This has mainly two advantages. Firstly, as outlined above, for lower bitrates it is generally beneficial to operate the core coder at a lower sampling rate than that of the original signal. Secondly, the HE AAC decoder can always assume that if SBR data is found in the bitstream the SBR Tool should be operated at twice the sampling rate of the core coder.

There are however, scenarios where it is beneficial to have a system that, from the outside, displays the behavior of a single-rate system. If the system is operated at a higher bitrate, the core coder may be operated at the sampling rate of the original signal, e.g. 44.1 kHz. However, since the HE AAC codec is always a dual rate system, this implies that the signal has to be upsampled in the encoder prior to encoding. The result of this upsampling is that the subsequent output from the decoder will have a sampling rate twice that of the input, in this case 88.2 kHz. In order to circumvent situations in which the output signals have a sampling rate twice that of the input, a down-sampling capability is desirable on the decoder side.

The down-sampling tool is integrated as a part of the synthesis QMF bank in the decoder. Hence, the SBR decoder operates as normal with the exception of the synthesis filterbank, which only performs a 32 subbands synthesis instead of the normal 64 subbands synthesis. The 32 subbands synthesis is done on the 32 lowest (in frequency) subbands out of the 64 available subbands, and hence a down-sampling by a factor of two is obtained, or rather the inherent upsampling by a factor of two is avoided. By using the down-sampled SBR for the above outlined use-case scenario, the output sampling rate can be assured to be the same as that of the input signal even though the AAC coder is operated at the sampling frequency of the original signal and the HE AAC codec is a dual rate system.

Furthermore, by introducing the down-sampling capability, it is possible for hand-held devices to use down-sampled SBR in order to reduce cost of D/A conversion, and also reduce computational complexity by the reduced number of synthesis subbands in the QMF synthesis. In this case the decoder might even generate an output sampling rate lower than the one of the orginial signal.

3.4 Test results

During the course of standardization within MPEG several listening tests have been conducted in order to assure the quality of the new algorithm. In Fig. 11 and Fig. 12 below, listening test results are displayed for SBR enhanced AAC and MPEG-4 AAC without SBR. The tests were done at the T-Nova test-site in Berlin, using the MUSHRA test method, and the items used were the twelve items commonly used within MPEG (castanets, pitchpipe, glockenspiel, plucked strings, speech, orchestra and pop music).

From Fig. 11, where MPEG-4 AAC at 24 and 30 kbit/s mono is compared with MPEG HE AAC at 24 kbit/s, it is evident that the new HE AAC profile decoder outperforms the MPEG-4 AAC by a wide margin. The absolute performance of the HE AAC profile codec is very good at this low bitrate. From Fig. 12, where MPEG-4 AAC at 48 and 60 kbit/s stereo is compared with MPEG HE AAC

Google Exhibit No. 1041
Page 007

Google Exhibit No. 1063
Page 1339

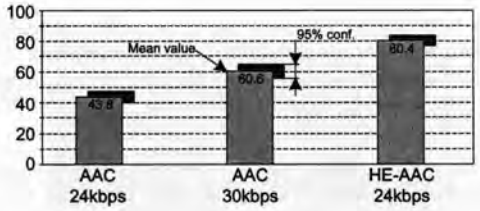


Figure 11: Listening test results for MPEG-4 AAC compared with MPEG-4 HE AAC, mono.

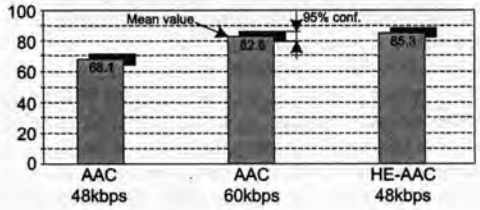


Figure 12: Listening test results for MPEG-4 AAC compared with MPEG-4 HE AAC, stereo.

at 48 kbit/s, it is evident that the new HE AAC profile decoder once again outperforms the MPEG-4 AAC.

Clearly, based on the above test results, HE AAC profile codec has overtaken the leading position in audio-compression efficiency from the previous technology leader MPEG-4 AAC.

4 INTEGRATING MPEG SBR INTO THE MPEG-4 FRAMEWORK

4.1 Embedding SBR data into AAC: Extension payload

SBR data is embedded into the AAC bitstream by means of the extension_payload() element, as shown in Fig. 13, and can therefore be combined with various AAC object type flavours (for example AAC LC, AAC Scalable, ER AAC LC). Two types of SBR extension data can be signalled through the extension_type field of the extension_payload(): EXT_SBR_DATA and EXT_SBR_DATA_CRC. The latter includes a 10 bit CRC in addition to the SBR data.

SBR can also be used in combination with a bitrate scalable system:

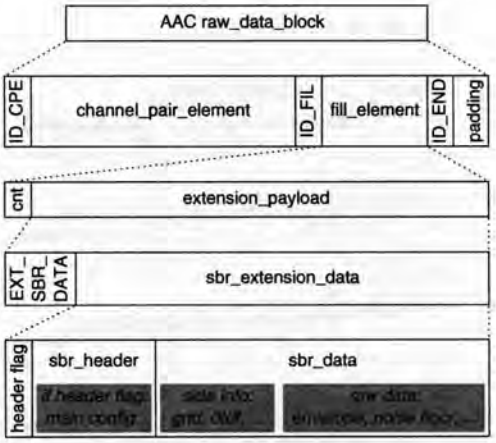


Figure 13: Conveying SBR data as extension payload of an AAC bitstream frame (this example: Stereo AAC LC).

MPEG-4 Audio is inherently scalable. If, for example, a transmission uses an error-prone channel with limited bandwidth, an audio stream consisting of a small base layer and a larger extension layer provides a robust solution. Strong error protection on the base layer (adding only little overhead to the overall bitrate) makes sure there is always a signal, even with difficult reception. The extension layer (with little error protection) and base layer together give excellent quality in normal conditions. Any errors lead only to a subtle degradation of quality but never in a total interruption of the audio stream. [26]

Scalable AAC example As an example Fig. 14 shows a possible three layer scalable system without the use of the new SBR tool, similar to the one used in the 1998 MPEG-4 Audio verification test [27]. The base layer includes a 20 kbit/s single channel AAC AOT with a sampling rate of 24 kHz, which covers an audio bandwidth of typically up to 8 kHz. The second layer implements a mono/stereo scalability and results in a stereo signal with 24 kHz sampling rate and 8 kHz bandwidth. Such a layer typically requires another 16 kbit/s. The third layer extends the bandwidth of both channels to about 12 kHz and requires approx. another 16 kbit/s. Depending on the

Google Exhibit No. 1041
Page 008

Google Exhibit No. 1063
Page 1340

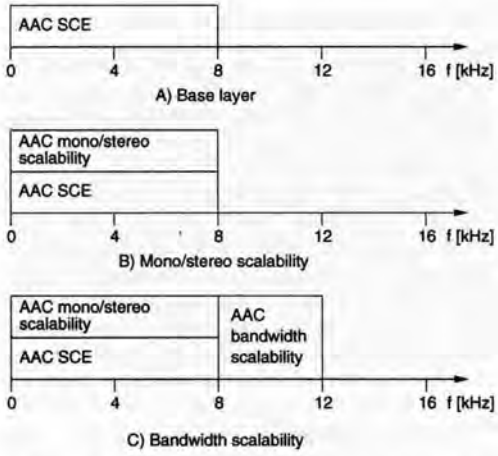


Figure 14: Example of a scalable system without the SBR tool.

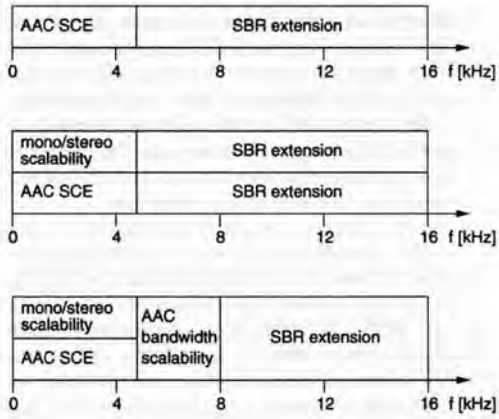


Figure 15: Example of a scalable system with the SBR tool. Contrary to a scalable system without SBR, it provides a constant bandwidth and requires lower bitrates by reducing the crossover frequency between the AAC and SBR cores.

quality of the transmission channel, a decoder might be able to decode one or more layers. Specifically when switching between the second and third layer, the change in bandwidth will be clearly noticeable and unpleasant to the listener.

Scalable HE AAC example With MPEG SBR a new tool is standardized that significantly increases the quality of such a scalable system. As indicated in Fig. 15, the SBR tool allows for maintaining a constant bandwidth of the audio signal of typically 16 kHz. Therefore the perceived audio quality in each layer is increased and the artifacts when switching between bandwidth scalable layers of the AAC core codec are reduced. Note, that the bandwidth scalability of the AAC core now becomes a crossover frequency scalability between the AAC and SBR algorithms while the overall bandwidth remains constant. HE AAC also allows for reducing the bitrate in each of the layers while maintaining the same audio quality. This is achieved by lowering the crossover frequency between the AAC and SBR algorithms. In real-world applications often a two-layer scalable configuration is used. Mono ->stereo scalability and crossover scalability are then combined into a single enhancement layer.

Embedding SBR data into a scalable system The scalable SBR data is embedded into the MPEG-4 stream in the same way as for non-scalable SBR data elements, by means of using the extension_payload(). For core coder bandwidth scalability, the SBR data is transmitted in the lowest AAC core coder layer and the SBR data covers the largest SBR frequency range used in the scalable system. In case mono/stereo scalability is chosen, the lowest stereo layer also carries an SBR data element. Note, that in the example above there is no additional SBR data being transmitted as part of the third layer. The SBR decoder will automatically adapt to the new crossover frequency used by the AAC core codec.

4.2 Methods for signaling SBR

The combination of AAC and SBR provides a significant increase of audio compression efficiency. At the same time it enables compatibility with existing AAC-only decoders. An AAC-only decoder will play only the AAC part of an HE AAC bitstream resulting in lower audio quality. There are several ways to signal the presence of SBR data:

1. **implicit signaling:** if SBR extension elements (EXT_SBR_DATA or EXT_SBR_DATA_CRC) are

Google Exhibit No. 1041
Page 009

Google Exhibit No. 1063
Page 1341

detected in the bitstream, this implicitly means that SBR data is present. This mode provides easy backward compatibility with AAC-only decoders but can introduce challenges when operating the decoder in a complex system such as an embedded device. The decoder needs to parse the payload at least partially in order to detect SBR. Only then can the output sampling rate be determined.

2. **explicit signaling:** the presence of SBR data is signaled by means of the AOT SBR in the AudioSpecificConfig(). This permits to convey configuration data specific to the SBR decoder, which includes separate specifications of the sampling rates for the SBR and AAC decoders. These specifications are also used to implicetly signal the down-sampling mode described in 3.3. If the sampling rates for the SBR and AAC decoders are identical, the down-sampled SBR tool is used. Two types of explicit signaling are available:

   (a) **hierarchical signaling:** if the first AOT is signaled as SBR, a second AOT is signaled which indicates the underlying AOT, e.g. AAC LC. This is a non backward compatible signaling method.

   (b) **backward compatible signaling:** the extensionAOT is signaled at the end of the AudioSpecificConfig(). This signaling method can only be used in systems that convey the length of the AudioSpecificConfig(). Because of this restriction, backward compatible explicit signaling can for example not be used with most LATM configurations. This mode does allow to explicitely signal the absence of SBR data as well and thus to explicitely signal the absence of implicit signaling. This circumvents the challenges that can occur when the decoder needs to check for implicit signaling.

To achieve backward compatibility with existing AAC-only decoders, a profile containing AAC (except HE AAC Profile) should be indicated on MPEG-4 Systems level (see Section 4.3) and either signaling method 1 or 2b shall be used. Method 1 can also be used in context with MPEG-2 AAC, where the MPEG-4 AudioSpecificConfig() is not available.

4.3 New Profiles & Levels

MPEG-4 provides a large and rich set of tools for the coding of audio objects. In order to allow effective implementations of the standard, subsets of the tool set have been identified that can be used for specific applications. The function of these subsets, called "Profiles," is to limit the tool set a conforming decoder must implement. For each of these Profiles, one or more Levels have been specified, thus restricting the computational complexity.

The High Efficiency AAC Profile is introduced as a superset of the AAC Profile. Besides the AOT AAC LC (which is present in the AAC Profile), it includes the AOT SBR. Levels are introduced within these Profiles in such a way, that a decoder supporting the High Efficiency AAC Profile at a given level can decode an AAC Profile stream at the same or lower level.

Level	Max. channels /object	Max. AAC sampling rate, SBR not present [kHz]	Max. AAC sampling rate, SBR present [kHz]	Max. SBR sampling rate, [kHz] (in/out)
1	NA	NA	NA	NA
2	2	48	24	24/48
3	2	48	48	48/48
4	5	48	24/48	48/48
5	5	96	48	48/96

Table 1: Levels within the HE AAC Profile

4.4 HE AAC signaling: Use Scenarios

Contrary to previous sections which described the technical specifications of HE AAC signaling from a MPEG-4 point of view, this section will look at these methods from a system integrator's point of view. On a high level, it can be distinguished between content providers/creators, decoder manufacturers, and system designers (e.g. closed systems for digital broadcasting etc.). Whereas the first group is in control of the encoding side only, and the second group is in control of the decoder only, the third group has the advantage of being in control over the complete signal chain.

Choosing a format The combination of the different signaling methods, as described in Section 4.2, with

Google Exhibit No. 1041
Page 010

Google Exhibit No. 1063
Page 1342

the different transport multiplexes as defined in Section 2.3, results in a large variety of bitstream formats for HE AAC. It should be noted that these different transport formats do not have any effect on the format of AAC raw_data_block. Thus transforming content from one into another transport format is a rather simple process. Therefore, as indicated in Fig. 3, a content provider might store data in the MPEG-4 file format (MP4FF) and then repackage it for the different distribution channels and configurations. On the other hand a decoder manufacturer needs to decide up-front which formats to support. It is conceivable that many MPEG-4 decoder manufacturers will include support for the ADIF and ADTS formats in order to be able to play back the widest possible range of bitstream formats.

4.4.1 HE AAC for content providers

As a content provider/creator, one can decide who will be able to decode the bitstreams. In general, HE AAC is backward compatible to existing AAC-only decoders. However, using decoders without the SBR tool will of course result in lower audio quality. It is possible to use signaling such that only users with new HE AAC decoders can decode the content. This is useful in order to guarantee the full audio quality for all listeners. In such a scenario, backwards compatibility to existing, older, AAC-only decoders is explicitely disabled.

Supporting only HE AAC decoders This can be achieved by indicating the new HE AAC profile. Legacy AAC decoders will not play back such audio content. On the other hand new HE AAC decoders have to use both, AAC and SBR decoder tools.

Supporting legacy MPEG-4 AAC-only decoders There are two signaling methods that will allow even legacy MPEG-4 AAC-only decoders to decode new HE AAC content. See Section 4.2 for details. In either case legacy AAC decoders will skip the SBR data, while new HE AAC decoders use that data to configure and call the SBR tool.

Supporting MPEG-2 and MPEG-4 decoders As long as implicit signaling for the SBR tool and the AAC LC AOT without the PNS tool are used, HE AAC can be formatted such that both MPEG-2 and MPEG-4 decoders can decode the content. MPEG is preparing a new release of the MPEG-2 standard to support this configuration [28]. Thus it is possible to create HE AAC content which can be decoded by MPEG-2 and MPEG-4 AAC decoders. Such HE AAC bitstreams will be decoded by AAC-only and HE AAC decoders, thus guaranteeing the largest number of compatible decoders.

4.4.2 HE AAC for decoder manufacturers

The MPEG standard defines profiles and levels to support interoperability between different encoder and decoder implementations. In general a decoder manufacturer is free to support any subset of tools as required. However, a *MPEG conformant* decoder needs to implement at least the subset of tools specified for a specific profile and level. Ultimately it will be important that the decoder can decode the desired content, which is dependent on the settings used by the content creator. An MPEG conformant encoder will also take into account the different profiles and levels so that interoperability between encoders and decoders is guaranteed. However, one cannot foresee at this point, what kind of levels might be used mostly by content creators.

Computational complexity Specifically for manufacturers of embedded decoders the computational complexity is a main concern. The MPEG standard allows for controling the computational complexity of a HE AAC decoder by means of the following three methods:

1. In situations where the listening environment is not as demanding, a specific low-power decoder can be used. See Section 3.2 for details on this tool, and Section 5.1 for detailed complexity assessments.
2. In general the SBR decoder operates at double the sampling rate of the AAC core coder. (Dual-rate system). However, in situations where a lower output sampling rate is sufficient (e.g. some portable devices) the SBR tool can run in a down-sampled mode, where the input and output sampling rate are identical. (See Section 3.3 for details.)
3. As the complexity increases with increasing levels it is important to only support the required level. The different levels differ in the maximum number of channels as well as the maximum sampling rates of the AAC core and SBR tool as indicated in Table 1.

Google Exhibit No. 1041
Page 011

Google Exhibit No. 1063
Page 1343

Following a few examples of how these options could be combined:

- *Basic configuration for cell phones:* Since the listening environment is not very demanding but computational resources might be extremely limited, a down-sampled, low-power HE AAC decoder supporting level 2 of the HE AAC profile might be suitable.
- *Basic configuration of a portable flash player:* The listening environment might be more demanding than in the case of a cell phone. A high quality HE AAC decoder supporting level 2 of the HE AAC profile might be suitable.
- *Basic configuration of a high-end hard-disk player:* Assuming the player does not support multichannel playback, a high quality HE AAC decoder supporting level 2 and 3 of the HE AAC profile might be suitable. Level 3 will allow to play back high-quality, high-bitrate files that utilize a samplerate of up to 48 kHz for the AAC core coder.
- *Basic configuration of a desktop decoder:* A high-quality HE AAC decoder supporting levels 2 to 5 of the HE AAC profile might be suitable for a state of the art desktop decoder. This will allow play back of up to 5 channels and an output sampling rate of up to 96 kHz.

4.4.3 HE AAC for digital broadcasting

The category of digital broadcasting applications includes proprietary digital radio systems such as XM Satellite Radio [29], digital radio standards such as Digital Radio Mondiale (DRM) [30], Internet-radio applications, or standards such as HDTV, DVD etc. These scenarios are similar in that they describe closed systems where MPEG *conformance* is not of major concern. However, using tools that are described in the MPEG standard and combining them into a new system is not only possible, but is a major concept behind the complete MPEG-4 standard. Although conformance and thus specific profiles and levels are unimportant, computational complexity of the decoder continues to be an area of major concern. So the same concepts as in the previous section also apply here.

A special case is digital broadcasting over noisy channels. For such a situation, a system that can adapt to the quality of service of the transmission channel is desireable. MPEG-4 introduces scalable audio codecs for such a scenario and the SBR tool supports and enhances this configuration as well. Section 4.1 includes an example for such a scalable system.

5 IMPLEMENTATIONS AND APPLICATIONS

With HE AAC, MPEG once more standardized a leading technology that is suitable for many existing and emerging markets and applications. The coding efficiency is high enough to support multichannel applications at low bitrates. A 5.1 bitstream at 112 kbit/s has been demonstrated and can support areas such as audio/video applications.

MPEG-4 HE AAC is a proven technology which has been widely deployed and is ready for use today. Reference decoder source code will soon be made available through MPEG and optimized source code for both encoders and decoders is available for license from companies such as Coding Technologies. Optimized binary implementations for Win32, Linux, MacOS X, ST Micro, ARM, Motorola, TI C64x & C55x, and Trimedia are presently available and other firmware implementations are being completed.

This section will first list the computational requirements of HE AAC decoder implementations. It will be shown, that despite the fact that the quality of HE AAC compared to plain AAC is significantly increased, the computational complexity of HE AAC is still comparable to a plain AAC decoder. In case of the already mentioned Low power decoder, the computational complexity is even identical to that of a plain AAC decoder. Thus HE AAC can indeed be implemented on low-power devices such as cell phones, enabling the delivery of full-band audio signals at a quality level that was not conceivable until today.

Section 5.2 then lists a few examples of applications and use scenarios that are well suited for HE AAC.

5.1 Computational complexity

The computational complexity of decoders is a major factor in deploying codecs in real-world applications. Specifically in mobile devices, computational complexity and memory requirements are directly related to power consumption and implementation cost. HE AAC was specifically designed to minimize these resources.

Google Exhibit No. 1041
Page 012

Google Exhibit No. 1063
Page 1344

Since HE AAC adds additional algorithms to the decoding process compared to plain AAC, one might expect a significant increase in computational complexity. However, as described in Section 3.3, the AAC core typically runs at half the output sampling rate. Thus the AAC core only requires about 50 % of the computational complexity compared to an AAC core running at the full sampling rate. The remaining resources are then being used by the additional SBR tool.

The MPEG standard indicates the computational complexity of various profiles, levels, and decoders by means of an abstract assessment based on Processor Complexity Units (PCU) and RAM Complexity Units (RCU). Although these numbers are suitable as a general, platform independent indicator of the computational complexity, a set of real-world numbers for embedded implementations is often a better approach. The numbers published in this paper are derived from implementations based on Coding Technologies Fixed-Point Firmware Reference Code (FFR) [31] [32] which is optimized in terms of memory usage and processing power. This code has been ported to various platforms so that the resulting performance numbers provide a realistic overview of the complexity.

Table 2 indicates the distribution of processing power between the AAC and the SBR decoder for a 24-bit and a 16-bit processor. All numbers given in this document relate to 44.1 kHz stereo operation. To achieve the superior audio quality of HE AAC a data width of at least 20 bits is required. Thus large parts of the SBR decoder need to be implemented using double precision arithmetic on devices with smaller data width, such as common 16-bit devices. Table 2 shows this difference in computational complexity.

Low power decoder As already described in Section 3.2, the MPEG SBR tool is available in two versions: High-quality (HQ) and Low power (LP). The difference only affects the decoding process. The encoder and the bitstream are identical and do not contain any information regarding these two versions. Which decoding process is used can be chosen depending on the application and the need to reduce power consumption and/or chip size. The high-quality version offers better audio quality at the cost of greater computational complexity. It should be used in situations where audio quality is most important. The Low power version results in a reduced audio quality but also requires less computational resources for both memory and complexity. Typical scenarios for the Low power version are mobile devices where complexity and thus power consumption is critical while the listening environment is usually not as demanding as in other applications. Table 2 shows that the computational complexity of the Low power decoder does not exceed the requirements of a plain AAC decoder. Thus better compression technology at the same computational cost is achieved.

	AAC	SBR	Total
HE AAC @24 bit	5	3	8
HE AAC @16 bit	9	4	13

Table 3: Distribution of ROM usage in KWords between AAC decoder and SBR decoder

Memory requirements Tables 3 and 4 summarize the memory requirements of the different HE AAC decoder implementations. The total data memory required can be as low as 20 KWords depending on the target platform and the chosen decoder implementation.

5.2 Applications and Use scenarios

The unique capability of HE AAC to achieve high quality at very low bitrates not only enhances existing markets but also enables new markets for digital audio. Where transmission bandwidth is constrained, the value of "High-efficiency AAC" is magnified. It not only enhances audio-only services, but also video services like digital TV. By coupling HE AAC with MPEG-4 Video, more bits can be allocated to the video signal without degrading the quality of the audio signal. This is true for mono, stereo, and multichannel applications. Combined with the new MPEG/ITU Advanced Video Coding (AVC) standard (included in MPEG-4 as part 10), even more significant gains in quality are possible.

Licensing Licensing of the SBR technology follows the simple licensing model of MPEG-4 AAC with annual caps for personal computer applications, per-unit fees for hardware devices and no additional fees for electronic music distribution. Licensing for HE AAC consists of two parts with schedules and details available from Coding Technologies [33] for the SBR object types and from Via Licensing [34] for the AAC object types.

Google Exhibit No. 1041
Page 013

Google Exhibit No. 1063
Page 1345

	AAC[a]	SBR	Total
HE AAC HQ @ 24-bit	10	15-20	25-30
HE AAC LP @ 24-bit	10	ca. 10	ca. 20
HE AAC HQ @ 16-bit	20	30-40	50-60
HE AAC LP @ 16-bit	20	ca. 20	ca. 40
plain AAC (full sampling rate) @ 24-bit	20	-	20
plain AAC (full sampling rate) @ 16-bit	40	-	40

[a] runs at half sampling rate in HE AAC

Table 2: Distribution of computational complexity in MIPS between the AAC and SBR decoder (stereo @ 44.1 kHz f_s). The complexity on 16-bit devices is largely dependent on the number of cycles needed to execute a double precision multiply-accumulate (MAC) instruction. Above estimates assume three cycles per double-precision MAC.

	input	AAC		SBR		output	Total
		static	dynamic	static	dynamic		
HE AAC HQ @24 bit	0.7	2.5	2	4	3	4	≈ 15
HE AAC LP @24 bit	0.7	2.5	2	ca. 3	ca. 1	4	≈ 12
HE AAC HQ @16 bit	0.7	4	4	7	8	4	≈ 24
HE AAC LP @16 bit	0.7	4	4	ca. 6	ca. 4	4	≈ 19

Table 4: Distribution of RAM usage in KWords between the AAC core and SBR (stereo)

Mobile streaming and download There is significant expectation for mobile multimedia services associated with the new 2.5G and 3G mobile service networks. Streaming video and other high-bandwidth services have been showcased as the ideal applications for this new infrastructure. The problem is that for much of the coverage area the peak bandwidth in these networks is generally around 144 kbit/s with individual users sustaining connections of about 40 kbit/s [35]. Delivering quality video over this type of connection is problematic and may not meet the consumer expectation. HE AAC is well suited to solve this problem by being able to provide consumer-grade download and streaming audio services within today's bandwidth. 48 kbit/s HE AAC provides CD-quality stereo programming while at 32 kbit/s, it provides excellent quality stereo programming. These bitrates combined with the growing market for subscription audio services show a strong business opportunity starting in 2003 on into 2004 and beyond.

Digital broadcast via satellite and cable *aacPlus* gained its first commercial success with XM Satellite Radio [29]. It allowed XM Radio to launch the most successfull digital radio system in the US to date. *aacPlus* was also selected by the Digital Radio Mondiale [30] consortium as part of the standard for Shortwave and AM digital radio. This heritage has brought credibility for HE AAC in the open standards world of digital satellite and cable broadcasting. As operators look to enhance their services with more channels or with high-definition, the efficiency of HE AAC gives them more options to either consolidate audio bandwidth to make room for more video or to layer more audio services like multilingual and 5.1 surround. Since Spectral Band Replication (SBR) is being added to the MPEG-2 standard as well, operators have the flexibility to use HE AAC within the MPEG-2 or MPEG-4 context as desired.

Internet distribution Video on demand and subscription content services are continuing to grow on the Internet. In these services, aggregate server bandwidth usage and last-mile bandwidth constraints make it difficult for operators and aggregators to provide high-quality, reliable services to consumers. By cutting the bandwidth requirements for audio nearly in half, MPEG-4 both reduces costs and increases reliability for Internet content services. In addition the flexible signaling of HE AAC within the MPEG-4 framework as described in Section 4.2 allows to control the user's listening experience. The content provider can for example decide if

Google Exhibit No. 1041
Page 014

Google Exhibit No. 1063
Page 1346

the new HE AAC bitstreams can be played back on plain AAC decoders as well as on HE AAC decoders. This trade off between backward compatibility and guaranteed audio quality at the decoder can be chosen for individual bitstreams.

6 CONCLUSIONS

The flexible integration of the new MPEG SBR tool into the MPEG-4 framework allows to control the backwards compatibility of HE AAC. Audio content can be packaged such that only HE AAC decoders can play back the audio streams or it can be packaged so that legacy AAC decoders can play back the new content as well, albeit at a lower audio quality. As such MPEG SBR integrates perfectly into MPEG-4 and enhances this open toolbox with an important audio coding method.

MPEG-4 also provides several transport protocols that support streaming applications as well as file based storage. The MPEG-4 framework guarantees the easy integration of HE AAC with MPEG video, thus combining state of the art video coding with state of the art audio coding. High quality multichannel and multilingual audio content at low bitrates thus becomes available for deployment in various audio/video application, enabling a whole new market for content storage and distribution.

In addition MPEG standardized for the first time two audio decoders that can both decode the same audio content: A high-quality version as well as a low power version. The availability of both decoders for HE AAC enables the standard to run on the widest possible range of processors in mobile and portable device applications. The flexibility in content creation as well as decoder implementation make HE AAC a highly effective yet generic MPEG-4 profile. This reflects the true spirit of the MPEG-4 standard: Providing a set of state of the art multimedia technologies that can be combined to support all multimedia applications.

7 REFERENCES

[1] ISO/IEC JTC1/SC29/WG1, "Text of ISO/IEC 14496-3:2001/FDAM1, Bandwidth Extension," ISO/IEC JTC1/SC29/WG11 N5570, Mar. 2003.

[2] ISO/IEC JTC1/SC29/WG11, "Overview of the MPEG-4 standard," ISO/IEC JTC1/SC29/WG11 N4030, Mar. 2001, available: `http://mpeg.telecomitalialab.com/standards/mpeg-4/mpeg-4.htm`.

[3] ISO/IEC JTC1/SC29/WG11, "Official MPEG Home Page," available: `http://mpeg.telecomitalialab.com/`.

[4] F. Pereira and T. Ebrahimi, Eds., *The MPEG-4 Book*, Prentice Hall, Englewood Cliffs, NJ, US, 2002.

[5] ISO/IEC, "Coding of audio-visual objects – Part 3: Audio (MPEG-4 Audio, 2nd edition)," ISO/IEC Int. Std. 14496-3:2001, 2001.

[6] ISO/IEC, "Coding of audio-visual objects – Part 2: Visual (MPEG-4 Visual, 2nd edition)," ISO/IEC Int. Std. 14496-2:2001, 2001.

[7] ISO/IEC, "Coding of audio-visual objects – Part 1: Systems (MPEG-4 Systems, 2nd edition)," ISO/IEC Int. Std. 14496-1:2001, 1999.

[8] ISO/IEC, "Generic coding of moving pictures and associated audio information – Part 7: Advanced Audio Coding (AAC)," ISO/IEC Int. Std. 13818-7:1997, 1997.

[9] K. Brandenburg, "Perceptual coding of high quality digital audio," in *Applications of Digital Signal Processing to Audio and Acoustics*, M. Kahrs and K. Brandenburg, Eds., chapter 2, pp. 39–83. Kluwer, Boston, 1998.

[10] T. Painter and A. Spanias, "Perceptual coding of digital audio," *Proc. IEEE*, vol. 88, no. 4, pp. 451–513, Apr. 2000.

[11] M. Bosi, K. Brandenburg, S. Quackenbush, L. Fielder, K. Akagiri, H. Fuchs, M. Dietz, J. Herre, G. Davidson, and Y. Oikawa, "ISO/IEC MPEG-2 Advanced Audio Coding," *J. Audio Eng. Soc.*, vol. 45, no. 10, pp. 789–814, Oct. 1997.

[12] J. Herre and H. Purnhagen, "General audio coding," in *The MPEG-4 Book*, F. Pereira and T. Ebrahimi, Eds., chapter 11. Prentice Hall, Englewood Cliffs, NJ, US, 2002.

[13] C. Herpel, G. Franceschini, and D. Singer, "Transporting and storing MPEG-4 content," in *The MPEG-4 Book*, F. Pereira and T. Ebrahimi, Eds., chapter 7. Prentice Hall, Englewood Cliffs, NJ, US, 2002.

Google Exhibit No. 1041
Page 015

Google Exhibit No. 1063
Page 1347

[14] H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson, "RTP: A transport protocol for real time applications," IETF RFC 1889, Jan. 1996, available: http://www.ietf.org/rfc/rfc1889.txt.

[15] Y. Kikuchi, T. Nomura, S. Fukunaga, Y. Matsui, and H. Kimata, "RTP payload format for MPEG-4 audio/visual streams," IETF RFC 3016, Nov. 2000, available: http://www.ietf.org/rfc/rfc3016.txt.

[16] ISO/IEC JTC1/SC29/WG11, "Text of ISO/IEC FDIS 14496-8, Carriage of ISO/IEC 14496 contents over IP networks," ISO/IEC JTC1/SC29/WG11 N4712, Mar. 2002.

[17] ISO/IEC, "Coding of audio-visual objects – Part 3: Audio (MPEG-4 Audio Version 1)," ISO/IEC Int. Std. 14496-3:1999, 1999.

[18] ISO/IEC, "Coding of audio-visual objects – Part 3: Audio, AMENDMENT 1: Audio extensions (MPEG-4 Audio Version 2)," ISO/IEC Int. Std. 14496-3:1999/Amd.1:2000, 2000.

[19] ISO/IEC JTC1/SC29/WG11, "Call for evidence justifying the testing of audio coding technology," ISO/IEC JTC1/SC29/WG11 N3483, July 2000, available: http://www.tnt.uni-hannover.de/project/mpeg/audio/public/w3483.pdf.

[20] ISO/IEC JTC1/SC29/WG11, "Call for proposals for new tools for audio coding," ISO/IEC JTC1/SC29/WG11 N3794, Jan. 2001, available: http://www.tnt.uni-hannover.de/project/mpeg/audio/public/w3794.pdf.

[21] M. Dietz, L. Liljeryd, K. Kjörling, and O. Kunz, "Spectral Band Replication, a novel approach in audio coding," in *112th AES Convention, Munich, May 10 - 13*, 2002, Preprint 5553.

[22] S. Meltzer, F. Henn, and R. Böhm, "SBR enhanced audio codecs for digital broadcasting such as "Digital Radio Mondiale" (DRM)," in *112th AES Convention, Munich, May 10 - 13*, 2002, Preprint 5559.

[23] T. Ziegler, A. Ehret, P. Ekstrand, and M. Lutzky, "Enhancing mp3 with SBR: Features and Capabilities of the new mp3PRO Algorithm," in *112th AES Convention, Munich, May 10 - 13*, 2002, Preprint 5560.

[24] A. Ehret, M. Dietz, and K. Kjörling, "State-of-the-Art Audio Coding for Broadcasting and Mobile Applications," in *114th AES Convention, Amsterdam, March 22 - 25*, 2003.

[25] P. Ekstrand, "Bandwidth Extension of Audio Signals by Spectral Band Replication," in *IEEE Benelux Workshop on Model based Processing and Coding of Audio (MPCA-2002), Leuven, Belgium, Nov. 15*, 2002.

[26] MPEG-4 Industry Forum, "MPEG-4 - The Media Standard," Nov. 2002, available: http://www.mpegif.org/public/documents/vault/m4-out-20027.pdf.

[27] ISO/IEC JTC1/SC29/WG11, "MPEG-4 Audio verification test results: Audio on the Internet," ISO/IEC JTC1/SC29/WG11 N2425, Oct. 1998, available: http://www.tnt.uni-hannover.de/project/mpeg/audio/public/w2425.pdf.

[28] ISO/IEC JTC1/SC29/WG11, "Text of ISO/IEC 13818-7:2003/FPDAM 1, Signalling of Bandwidth Extension," ISO/IEC JTC1/SC29/WG11 N5568, Mar. 2003.

[29] "XM Satellite Radio," WWW page, 2003, http://www.xmradio.com/.

[30] "Digital Radio Mondiale," WWW page, 2003, http://www.drm.org.

[31] Coding Technologies, "Embedded aacPlus Decoder Implementations - Product Sheet," June 2003.

[32] Coding Technologies, "Embedded HE-AAC: Implementation Guideline," June 2003.

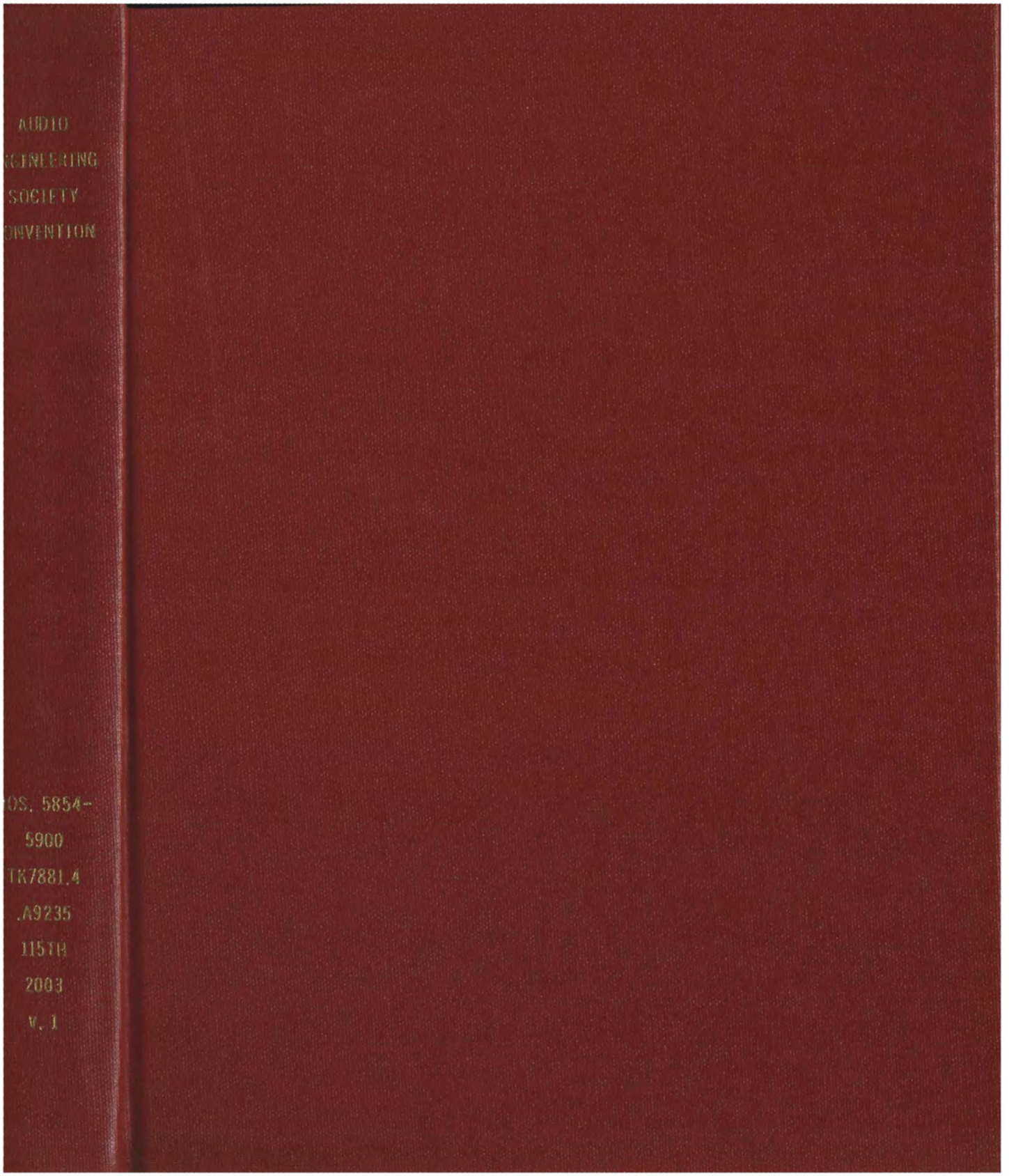
[33] "Coding Technologies," WWW page, 2003, http://www.codingtechnologies.com.

[34] "Via Licensing," WWW page, 2003, http://www.vialicensing.com.

[35] D. Frerichs, "New MPEG-4 High-efficiency AAC Audio," Apr. 2003, available: http://www.mpegif.org/public/documents/vault/m4-out-30034.zip.

Google Exhibit No. 1041
Page 016

Google Exhibit No. 1063
Page 1348



Google Exhibit No. 1041
Page 017

Google Exhibit No. 1063
Page 1349

Ex. 1041A

Audio Engineering Society
Convention Paper 5871
Presented at the 115th Convention
2003 October 10–13 New York, NY, USA

This convention paper has been reproduced from the author's advance manuscript, without editing, corrections, or consideration by the Review Board. The AES takes no responsibility for the contents. Additional papers may be obtained by sending request and remittance to Audio Engineering Society, 60 East 42nd Street, New York, New York 10165-2520, USA; also see www.aes.org. All rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

A closer look into MPEG-4 High Efficiency AAC

Martin Wolters[1], Kristofer Kjörling[2], Daniel Homm[1], Heiko Purnhagen[2]

[1] *Coding Technologies, Deutschherrnstr. 15–19, 90429 Nürnberg, Germany*

[2] *Coding Technologies, Döbelnsgatan 64, 11352 Stockholm, Sweden*

Correspondence should be addressed to Martin Wolters (wol@codingtechnologies.com)

ABSTRACT

MPEG Spectral Band Replication (SBR) is the newest compression technology available as part of the MPEG standards. It is combined with MPEG Advanced Audio Coding (AAC) and improves coding efficiency by more than 30%. The resulting scheme is called High-Efficiency AAC (HE AAC).
This presentation explains MPEG SBR and its integration into the existing MPEG-4 bitstream format. The SBR technology itself as well as the implications on systems based on MPEG-4 technology are described. Signaling through MPEG-4 Systems and other transport formats is introduced and typical applications and usage scenarios are listed.

1 INTRODUCTION

MPEG-4 High Efficiency AAC (HE AAC) is not a replacement for AAC, but rather a superset which extends the reach of high-quality MPEG-4 Audio to much lower bitrates. High Efficiency AAC decoders will decode both plain AAC and the enhanced AAC plus SBR. The result is a backward compatible extension of the standard. Listening tests show that MPEG-4 HE AAC offers a significant benefit over proprietary codecs and over AAC without extensions, and place it clearly as the most efficient audio codec in existence. Coding Technologies and its customers have been using this technology for a number

of years under the *aacPlus* brand name.

The technical specifications of HE AAC are now available as a final draft amendment (FDAM) [1] and official standardization is expected in September 2003. These specifications enable manufacturers to work on final implementations and integrations. This paper provides an implementation guideline and provides an overview of the technologies and system integration aspects of HE AAC.

2 INTRODUCTION INTO MPEG-4

2.1 MPEG-4: A leading multimedia framework

In the context of evolving multimedia applications, new demands for efficient and flexible representation of audiovisual content have arisen. Besides high coding efficiency required to cope with the limited bandwidth of the internet or in mobile communication, new functionality like flexible access to coded data and manipulation by the recipient is also desired. To address these requirements and develop inter-operable solutions, ISO/IEC started its MPEG-4 standardization activities "Coding of audiovisual objects" [2], [3], [4]. Part 3 of the MPEG-4 standard (Audio) [5] provides tools for coding of natural and synthetic audio objects and composition of such objects into an "audio scene." Similarly, Part 2 (Visual) [6] provides tools for coding of natural video and synthetic 2D and 3D objects, while Part 1 (Systems) [7] covers general tools for scene description, synchronization of objects, and multiplexing for transport and storage.

2.2 MPEG-4 Audio Coding

MPEG-4 is based on the notion that the audio part of the audiovisual scene presented at the receiver is composed of one or more so-called *audio objects*. Different audio compression tools (codecs) are available to enable an efficiently coded representation of the audio object(s) in a scene. Natural audio objects, such as recorded speech and music, can be coded at bitrates typically ranging from 2 kbit/s (for narrowband speech) to 64 kbit/s/ch (for CD quality music) using parametric speech coding (HVXC), CELP-based speech coding, parametric audio coding (HILN) or transform-based general audio coding (AAC, TwinVQ). The natural audio and speech coding tools support bitrate scalability, also known as embedded coding. In addition, the parametric coding tools also provide speed and pitch change functionality in the decoder. Synthetic audio objects can be represented using

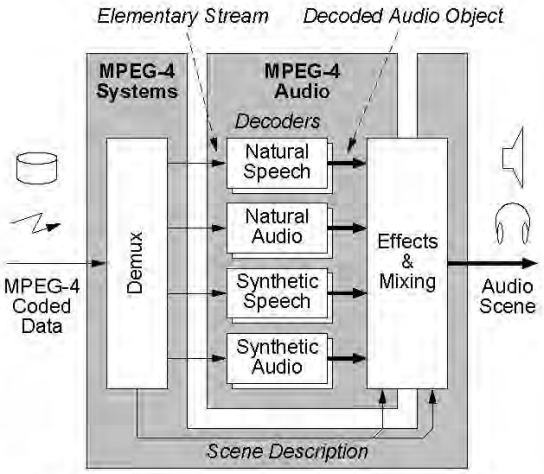


Figure 1: Block diagram of a complete MPEG-4 Audio decoder.

a Text-To-Speech Interface (TTSI) or the Structured Audio (SA) synthesis tools. Other uses of the SA tools are adding effects, like reverberation, and mixing different audio objects to compose the final "audio scene" that is presented to the listener.

Fig. 1 shows the block diagram of a complete MPEG-4 Audio decoder. It includes the decoding tools for the audio objects defined in Part 3 (Audio) of the MPEG-4 Standard [5] as well as bitstream demultiplexing and scene composition defined in the Systems part [7].

The information needed to decode an audio object at the decoder is conveyed by means of a so-called *elementary stream* (ES), as shown in Fig. 2. In case of bitrate scalable configurations, a base-layer ES and one or more enhancement-layers ES(s) are used. The initial ES_Descriptor contains an AudioSpecificConfig element, which carries the *audio object type* (AOT) and further information required to instantiate the required audio decoder. Using the ES_ID, the ES_Descriptor points to the stream of the actual compressed audio data, which is a sequence of so-called *access units* (AU), i.e., decodable bitstream frames.

A common audio object type is AAC LC, which indicates that MPEG-4 Advanced Audio Coding (AAC) in its low-complexity (LC) version is used. This is the primary general audio coder in MPEG-4, which is a compatible extension of the MPEG-2 AAC transform-based audio coder [8]. The principles of perceptual audio cod-

Google Exhibit No. 1063
Page 1352

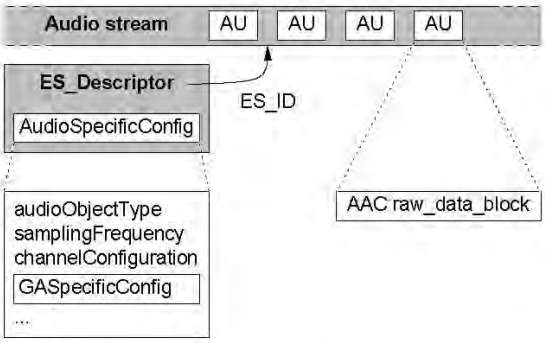


Figure 2: MPEG-4 Elementary Stream (ES) conveying an audio object.

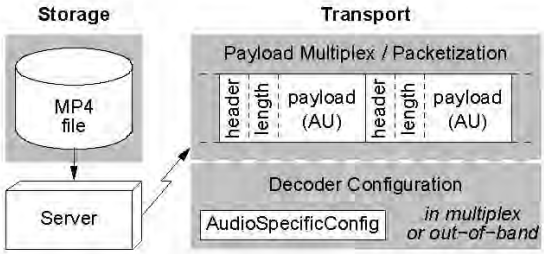


Figure 3: Storage and transmission of MPEG-4 Audio object with single elementary stream (ES).

ing and the specifics of the AAC are well documented in the literature, e.g. [9], [10], [11], [12], and hence not reviewed in this presentation.

2.3 Storage and streaming of MPEG-4

In order to allow flexible yet interoperable utilization of MPEG-4 coding technology for the largest possible range of application scenarios, MPEG-4 is designed as a transport-agnostic standard with an abstract interface to carry the compressed data, the so-called *sync layer* (SL). Basically, the SL adds mechanisms like time stamps for AUs to allow synchronization of several ESs, such as an audio and a video stream.

Depending upon the requirements of the application and the characteristics of the transmission channel, different schemes can be used to store or convey MPEG-4 content (Fig. 3). MPEG defined the MP4 file format [7], [13] for storage, which can hold all data available at the SL interface. When MPEG-4 content is carried over a transmission channel, various techniques for multiplexing, synchronization, and for establishing a connection can be appropriate. For inherently packetized channels, like those using the internet protocol (IP), no syncwords are needed and AUs can often directly be mapped to packets. A fixed framing might be available from the modulation scheme for radio channels, but random bit errors are possible. Serial channels might need some syncword mechanisms to establish basic synchronization. If the MPEG-4 content is carried in more than one ES, these streams need to be multiplexed, which can be done by the MPEG-4 FlexMux tool [7], [13].

To carry simple, audio-only MPEG-4 content with low overhead, the Part 3 (Audio) of the MPEG-4 standard defines a *Low-overhead MPEG-4 Audio Transport Multiplex* (LATM) to multiplex several MPEG-4 Audio payloads and AudioSpecificConfig() elements [5, subclause 1.7]. Based on LATM, it also defines a self-synchronized syntax of an MPEG-4 Audio transport stream which is called *Low Overhead Audio Stream* (LOAS).

A simple scheme to carry MPEG-4 over IP based on the realtime transport protocol (RTP) [14] is defined in RFC 3016 [15]. It is a subset of the general framwork defined in Part 8 (MPEG-4 over IP) of the MPEG-4 standard [16]. RFC 3016 uses LATM for audio and allows for conveying the AudioSpecificConfig either in-band using LATM or out-of-band, i.e., as a MIME parameter conveyed in the Session Description Protocol (SDP).

Other schemes to convey MPEG-4 content include MPEG-4 over MPEG-2 transport streams, as defined in an amendement to the MPEG-2 Systems standard, or proprietary schemes which provide an SL compatible interface.

MPEG-4 AAC LC coded data can also be stored using the *Audio Data Interchange Format* (ADIF) or transmitted using an *Audio Data Transport Stream* (ADTS), see Fig. 4. Both these formats are defined in the MPEG-2 AAC standard [8] but are only informative in MPEG-4, i.e., an MPEG-4 decoder is not required to support ADIF or ADTS.

2.4 Recent Developments: High-Efficiency Audio Coding becomes part of MPEG-4

Standardization of the technology submitted in response to the initial MPEG-4 Call for Proposals in July 1995 was carried out using the core experiment procedure and finalized in two steps: Version 1 in October 1998 [17]

Google Exhibit No. 1063
Page 1353

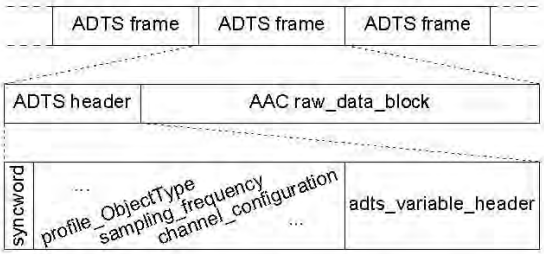


Figure 4: MPEG-2 Audio Data Transport Stream (ADTS).

and Version 2 in December 1999 [18]. In 2001, both versions and their corrigenda were merged to form a 2nd edition of the standard [5].

In July 2000, MPEG issued a Call for Evidence [19] in order to find out whether there are new developments which may provide improvements over the existing MPEG-4 standard. Based on the responses received, a Call for Proposals [20] for new tools for audio coding was issued in January 2001. It specifically asked for technology in the field of MPEG-4 compatible bandwidth extension and in the field of high quality parametric coding.

Coding Technologies submitted its Spectral Band Replication (SBR) techniques used in combination with AAC as bandwith extension tool. Based on a listening test, it was selected as inital reference model (RM0) and refined in the course of MPEG's core experiment procedure. In March 2003, the standardization of the MPEG-4 SBR tool was finalized [1].

3 AT THE HEART OF HE AAC: MPEG SBR

3.1 Fundamental Concept: High Frequency Reconstruction by Spectral Band Replication

The basic principles of SBR have been elaborated on in several papers [21], [22], [23], [24]. For the convenience of the reader a short review is given below, and some aspects not previously discussed are explained in more detail.

3.1.1 The SBR encoder

The SBR principle stipulates that the missing high frequency region of a lowpass filtered signal can be recovered based on the existing lowpass signal and a small amount of control data. The required control data is estimated in the encoder given the original wide-band signal. The HE AAC codec is a dual rate system, where the underlying AAC encoder/decoder is operated at half the sampling rate of the SBR encoder/decoder. The basic principle of the HE AAC encoder is depicted in Fig. 5.

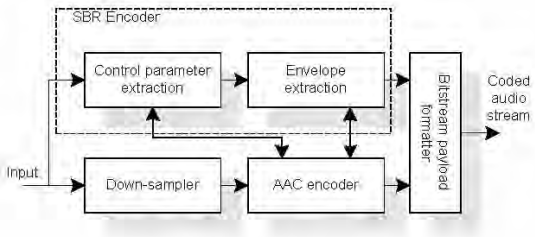


Figure 5: The HE AAC encoder.

In the SBR encoder, where the wide band signal is available, control parameters are estimated in order to ensure that the high frequency reconstruction results in a reconstructed highband that is perceptually as similar as possible to the original highband. The majority of the control data is used for a spectral envelope representation. The spectral envelope information has varying time and frequency resolution to be able to control the SBR process as good as possible, with as little bitrate overhead as possible. The other control data mainly strives to control the tonal-to-noise ratio of the highband. Fig. 6, Fig. 7, and Fig. 8, illustrate some of the characteristics of the control data.

In Fig. 6 the spectrum of the original signal at a given point in time is displayed at the top and the spectrum of the HF generated highband at the same point in time is displayed below. As is evident from the figure, the patching algorithm used to regenerate the highband was not successfull in generating all the strong tonal components of the original highband. The missing sinusoids in the regenerated highband are identified in the encoder and a parametric representation of them is incorporated in the SBR control data.

In Fig. 7 the spectra of the original and the HF generated signal is once again displayed albeit at a different point in time. Here it is evident that the tonal-to-noise ratio of the regenerated highband is not similar to that of the original. The highband of the original signal constitutes mainly of noise, whilst the highband of the regenerated signal is very tonal. This information is also incorporated into the

Google Exhibit No. 1063
Page 1354

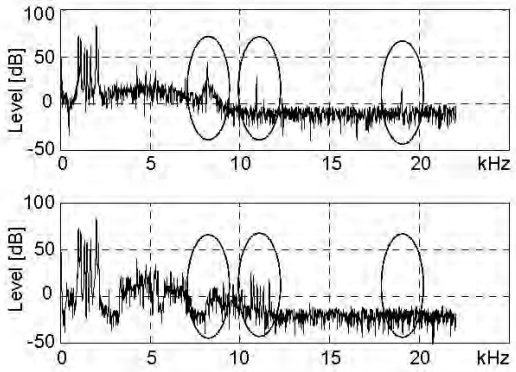


Figure 6: The spectrum of the original signal (top) and the HF generated signal without any additional signal components (bottom). The basic patching algorithm has not regenerated the strong tonal components of the original highband.

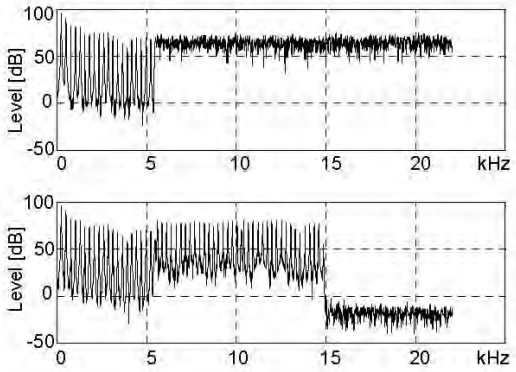


Figure 7: The spectrum of the original signal (top) and the SBR signal without any additional signal components (bottom). The basic high frequency reconstruction algorithm has not maintained the original tonal-to-noise ratio.

SBR control data so that the decoder by means of inverse filtering and noise addition can achieve a highband with a tonal-to-noise ratio similar to that of the original.

Finally, in Fig. 8, a spectrogram of the original signal is displayed with a superimposed time/frequency grid of the spectral envelope data transmitted to the decoder. Here it is evident that the time/frequency resolution of the spectral envelope varies over time, giving a higher time resolution for transient passages and a higher frequency resolution for stationary passages.

The bitrate of the control data varies depending on encoder tuning, but is in general somewhere in the region of 1-3 kbit/s per audio channel. This is far lower than the bitrate that would be required to code the highband with any conventional wave-form coding algorithm. The SBR data format is indicated in Fig. 13. A header flag indicates that an SBR header part is present. The SBR header part contains fundamental information such as SBR frequency range as well as control signals that do not require frequent changes. The sbr_data part can be subdivided into side info and raw data, where side info is defined as signals needed to decode the raw data and some decoder tuning signals. Raw data consists of Huffman coded envelope scalefactors and noise floor estimates.

3.1.2 The SBR decoder

The SBR enhanced decoder can be roughly divided into the modules depicted in Fig. 9.

All SBR processing is done in the QMF domain. Hence, the output from the underlying AAC decoder is firstly analyzed with a 32 channel QMF filterbank. Secondly, the HF generator module recreates the highband by patching QMF subbands from the existing lowband to the high band. Furthermore inverse filtering is done on a per QMF subband basis, based on the control data obtained from the bitstream. The envelope adjuster modifies the spectral envelope of the regenerated highband, and adds additional components such as noise and sinusoids, all according to the control data in the bitstream. Since all operations are done in the QMF domain the final step of the decoder is a QMF synthesis to retain a time-domain signal. Given that the QMF analysis is done on 32 QMF subbands for 1024 time-domain samples, and the high frequency reconstruction results in 64 QMF subbands upon which the synthesis is done producing 2048 time-domain samples, an up-sampling by a factor of two is obtained.

The following Fig. 10 displays the spectrum of the SBR signal at different stages in the SBR process. In the upper figure the spectrum of the signal after the high frequency reconstruction, but prior to the envelope adjustment is displayed. In the middle figure the spectrum of the output signal is displayed, and in the bottom figure the spectrum of the original signal is displayed for reference.

Google Exhibit No. 1063
Page 1355

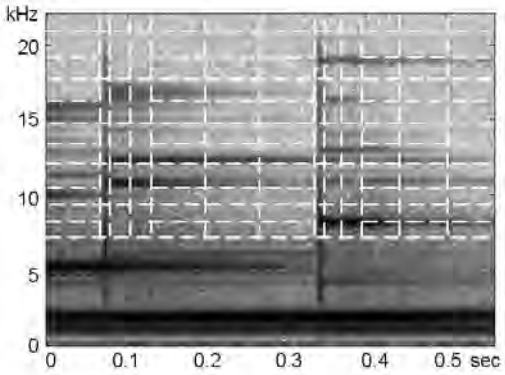


Figure 8: The spectrogram of a part of the input signal with the envelope time/frequency grid superimposed.

3.2 Low power SBR

Platforms with heavy constraints on the computational complexity such as portable devices can use Low power SBR. The main difference between High Quality SBR and Low power SBR is how the data is represented during the SBR process. High Quality SBR uses a complex representation of the subband samples and subsequently all calculations are computed with complex values. Low power SBR simply requires a real-valued representation and hence the computational complexity is heavily reduced.

The complex-valued filterbank is designed to prevent aliasing even in cases where modifications are performed [25]. However, the real-valued filterbank does not have that property. Therefore, an additional algorithm is introduced to minimize the aliasing occurring due to the real-valued filterbank. The aliasing reduction algorithm is devised to avoid introducing aliasing for strong tonal components.

The properties of the real-valued filterbank result in mirroring components (aliasing) in adjacent subbands if the adjacent subbands are given different gain-values in the envelope adjuster. This can occur if adjacent subbands are modified independently from each other. Hence, it is the objective of the aliasing reduction algorithm to identify the subbands where strong aliasing will be introduced if the subbands are modified independently. This is accomplished by observing the reflection coefficients obtained for every subband by first order linear prediction. A sign is introduced for every subband indicating whether a strong tonal component is situated in the upper (a positive sign) or lower (a negative sign) part of the subband. The aliasing reduction algorithm identifies the cases where the signs of two adjacent subbands are opposite, the sign being positive for the lower subband and negative for the higher subband. These subbands are grouped together, meaning that they will not be modified independently, and hence no aliasing will be introduced by the envelope adjustment in the real-valued filterbank for this signal. However, there are many signals where there are no strong tonal components that can be identified by the above outlined algorithm, but where aliasing is introduced nevertheless. This aliasing is audible when compared to the High Quality SBR system, but not as offensive as the aliasing that would occur for a strong tonal component if the aliasing reduction algorithm was not in place.

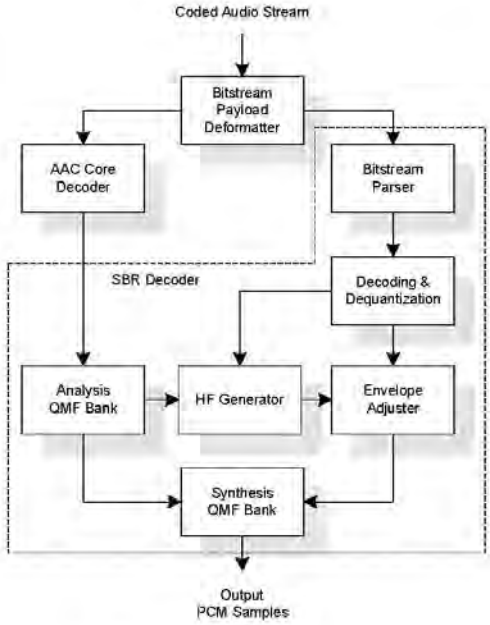


Figure 9: The HE AAC decoder.

3.3 Down-sampled SBR

Since the transform size in AAC is fixed to 1024 for long blocks and 128 for short blocks, it becomes evident that choosing the optimum sampling rate for a certain bitrate

Google Exhibit No. 1063
Page 1356

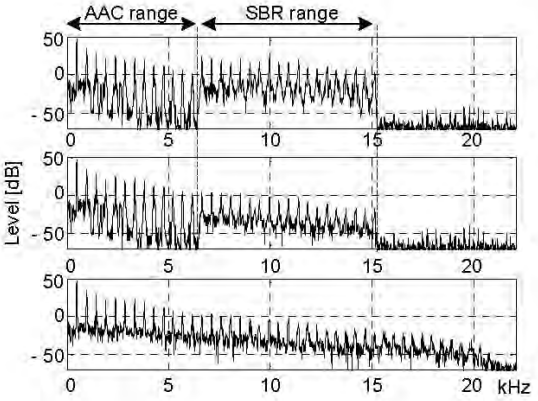


Figure 10: The spectrum of the audio signal after the HF generator (top), the spectrum of the audio signal after envelope adjustment (middle), and the spectrum of the original signal (bottom).

is crucial. In general, in the low-mid bitrate region, i.e., 20 - 32 kbit/s per audio channel, a sampling rate in the 22.05 – 32 kHz region is the best choice since the frequency resolution at the lower sampling rate improves the coding gain, compared to using a higher sampling rate. However the lower sampling rate provides a slower time response of AAC, and at higher bitrates the best trade-off between time and frequency resolution for the AAC codec may be obtained by using a higher sampling-rate. Furthermore, the low sampling rates naturally limits the highest frequency that can be covered by AAC.

The HE AAC codec is always operated in a dual rate mode. This has mainly two advantages. Firstly, as outlined above, for lower bitrates it is generally beneficial to operate the core coder at a lower sampling rate than that of the original signal. Secondly, the HE AAC decoder can always assume that if SBR data is found in the bitstream the SBR Tool should be operated at twice the sampling rate of the core coder.

There are however, scenarios where it is beneficial to have a system that, from the outside, displays the behavior of a single-rate system. If the system is operated at a higher bitrate, the core coder may be operated at the sampling rate of the original signal, e.g. 44.1 kHz. However, since the HE AAC codec is always a dual rate system, this implies that the signal has to be upsampled in the encoder prior to encoding. The result of this upsampling is that the subsequent output from the decoder will have a sampling rate twice that of the input, in this case 88.2 kHz. In order to circumvent situations in which the output signals have a sampling rate twice that of the input, a down-sampling capability is desirable on the decoder side.

The down-sampling tool is integrated as a part of the synthesis QMF bank in the decoder. Hence, the SBR decoder operates as normal with the exception of the synthesis filterbank, which only performs a 32 subbands synthesis instead of the normal 64 subbands synthesis. The 32 subbands synthesis is done on the 32 lowest (in frequency) subbands out of the 64 available subbands, and hence a down-sampling by a factor of two is obtained, or rather the inherent upsampling by a factor of two is avoided. By using the down-sampled SBR for the above outlined use-case scenario, the output sampling rate can be assured to be the same as that of the input signal even though the AAC coder is operated at the sampling frequency of the original signal and the HE AAC codec is a dual rate system.

Furthermore, by introducing the down-sampling capability, it is possible for hand-held devices to use down-sampled SBR in order to reduce cost of D/A conversion, and also reduce computational complexity by the reduced number of synthesis subbands in the QMF synthesis. In this case the decoder might even generate an output sampling rate lower than the one of the orginial signal.

3.4 Test results

During the course of standardization within MPEG several listening tests have been conducted in order to assure the quality of the new algorithm. In Fig. 11 and Fig. 12 below, listening test results are displayed for SBR enhanced AAC and MPEG-4 AAC without SBR. The tests were done at the T-Nova test-site in Berlin, using the MUSHRA test method, and the items used were the twelve items commonly used within MPEG (castanets, pitchpipe, glockenspiel, plucked strings, speech, orchestra and pop music).

From Fig. 11, where MPEG-4 AAC at 24 and 30 kbit/s mono is compared with MPEG HE AAC at 24 kbit/s, it is evident that the new HE AAC profile decoder outperforms the MPEG-4 AAC by a wide margin. The absolute performance of the HE AAC profile codec is very good at this low bitrate. From Fig. 12, where MPEG-4 AAC at 48 and 60 kbit/s stereo is compared with MPEG HE AAC

Google Exhibit No. 1063
Page 1357

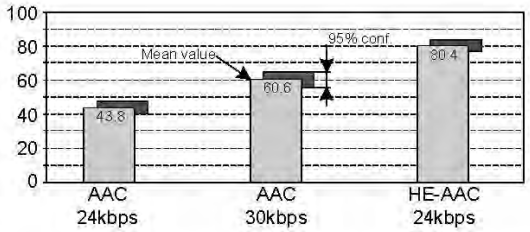


Figure 11: Listening test results for MPEG-4 AAC compared with MPEG-4 HE AAC, mono.

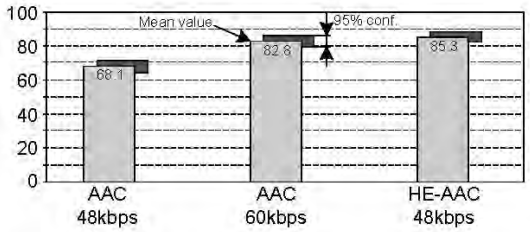


Figure 12: Listening test results for MPEG-4 AAC compared with MPEG-4 HE AAC, stereo.

at 48 kbit/s, it is evident that the new HE AAC profile decoder once again outperforms the MPEG-4 AAC.

Clearly, based on the above test results, HE AAC profile codec has overtaken the leading position in audio-compression efficiency from the previous technology leader MPEG-4 AAC.

4 INTEGRATING MPEG SBR INTO THE MPEG-4 FRAMEWORK

4.1 Embedding SBR data into AAC: Extension payload

SBR data is embedded into the AAC bitstream by means of the extension_payload() element, as shown in Fig. 13, and can therefore be combined with various AAC object type flavours (for example AAC LC, AAC Scalable, ER AAC LC). Two types of SBR extension data can be signalled through the extension_type field of the extension_payload(): EXT_SBR_DATA and EXT_SBR_DATA_CRC. The latter includes a 10 bit CRC in addition to the SBR data.

SBR can also be used in combination with a bitrate scalable system:

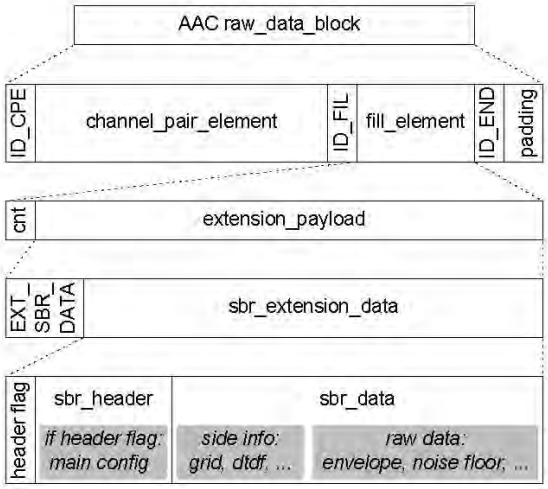


Figure 13: Conveying SBR data as extension payload of an AAC bitstream frame (this example: Stereo AAC LC).

> MPEG-4 Audio is inherently scalable. If, for example, a transmission uses an error-prone channel with limited bandwidth, an audio stream consisting of a small base layer and a larger extension layer provides a robust solution. Strong error protection on the base layer (adding only little overhead to the overall bitrate) makes sure there is always a signal, even with difficult reception. The extension layer (with little error protection) and base layer together give excellent quality in normal conditions. Any errors lead only to a subtle degradation of quality but never in a total interruption of the audio stream. [26]

Scalable AAC example As an example Fig. 14 shows a possible three layer scalable system without the use of the new SBR tool, similar to the one used in the 1998 MPEG-4 Audio verification test [27]. The base layer includes a 20 kbit/s single channel AAC AOT with a sampling rate of 24 kHz, which covers an audio bandwidth of typically up to 8 kHz. The second layer implements a mono/stereo scalability and results in a stereo signal with 24 kHz sampling rate and 8 kHz bandwidth. Such a layer typically requires another 16 kbit/s. The third layer extends the bandwidth of both channels to about 12 kHz and requires approx. another 16 kbit/s. Depending on the

Google Exhibit No. 1063
Page 1358

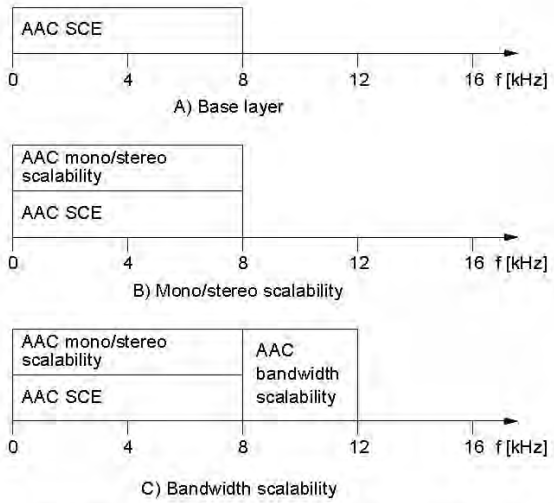


Figure 14: Example of a scalable system without the SBR tool.

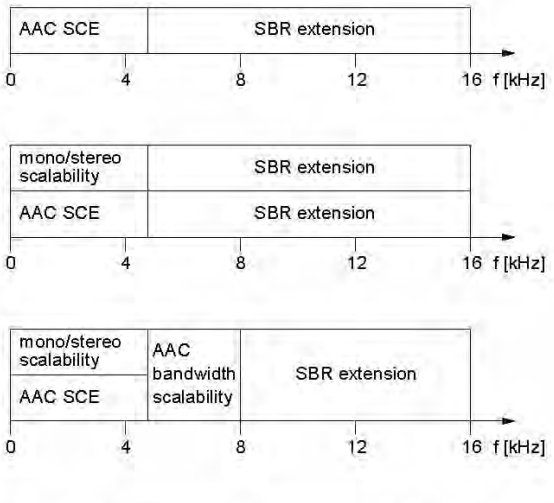


Figure 15: Example of a scalable system with the SBR tool. Contrary to a scalable system without SBR, it provides a constant bandwidth and requires lower bitrates by reducing the crossover frequency between the AAC and SBR cores.

quality of the transmission channel, a decoder might be able to decode one or more layers. Specifically when switching between the second and third layer, the change in bandwidth will be clearly noticeable and unpleasant to the listener.

Scalable HE AAC example With MPEG SBR a new tool is standardized that significantly increases the quality of such a scalable system. As indicated in Fig. 15, the SBR tool allows for maintaining a constant bandwidth of the audio signal of typically 16 kHz. Therefore the perceived audio quality in each layer is increased and the artifacts when switching between bandwidth scalable layers of the AAC core codec are reduced. Note, that the bandwidth scalability of the AAC core now becomes a crossover frequency scalability between the AAC and SBR algorithms while the overall bandwidth remains constant. HE AAC also allows for reducing the bitrate in each of the layers while maintaining the same audio quality. This is achieved by lowering the crossover frequency between the AAC and SBR algorithms. In real-world applications often a two-layer scalable configuration is used. Mono ->stereo scalability and crossover scalability are then combined into a single enhancement layer.

Embedding SBR data into a scalable system The scalable SBR data is embedded into the MPEG-4 stream in the same way as for non-scalable SBR data elements, by means of using the extension_payload(). For core coder bandwidth scalability, the SBR data is transmitted in the lowest AAC core coder layer and the SBR data covers the largest SBR frequency range used in the scalable system. In case mono/stereo scalability is chosen, the lowest stereo layer also carries an SBR data element. Note, that in the example above there is no additional SBR data being transmitted as part of the third layer. The SBR decoder will automatically adapt to the new crossover frequency used by the AAC core codec.

4.2 Methods for signaling SBR

The combination of AAC and SBR provides a significant increase of audio compression efficiency. At the same time it enables compatibility with existing AAC-only decoders. An AAC-only decoder will play only the AAC part of an HE AAC bitstream resulting in lower audio quality. There are several ways to signal the presence of SBR data:

1. **implicit signaling:** if SBR extension elements (EXT_SBR_DATA or EXT_SBR_DATA_CRC) are

Google Exhibit No. 1063
Page 1359

detected in the bitstream, this implicitly means that SBR data is present. This mode provides easy backward compatibility with AAC-only decoders but can introduce challenges when operating the decoder in a complex system such as an embedded device. The decoder needs to parse the payload at least partially in order to detect SBR. Only then can the output sampling rate be determined.

2. **explicit signaling:** the presence of SBR data is signaled by means of the AOT SBR in the AudioSpecificConfig(). This permits to convey configuration data specific to the SBR decoder, which includes separate specifications of the sampling rates for the SBR and AAC decoders. These specifications are also used to implicetly signal the down-sampling mode described in 3.3. If the sampling rates for the SBR and AAC decoders are identical, the down-sampled SBR tool is used. Two types of explicit signaling are available:

   (a) **hierarchical signaling:** if the first AOT is signaled as SBR, a second AOT is signaled which indicates the underlying AOT, e.g. AAC LC. This is a non backward compatible signaling method.

   (b) **backward compatible signaling:** the extensionAOT is signaled at the end of the AudioSpecificConfig(). This signaling method can only be used in systems that convey the length of the AudioSpecificConfig(). Because of this restriction, backward compatible explicit signaling can for example not be used with most LATM configurations. This mode does allow to explicitely signal the absence of SBR data as well and thus to explicitely signal the absence of implicit signaling. This circumvents the challenges that can occur when the decoder needs to check for implicit signaling.

To achieve backward compatibility with existing AAC-only decoders, a profile containing AAC (except HE AAC Profile) should be indicated on MPEG-4 Systems level (see Section 4.3) and either signaling method 1 or 2b shall be used. Method 1 can also be used in context with MPEG-2 AAC, where the MPEG-4 AudioSpecificConfig() is not available.

4.3 New Profiles & Levels

MPEG-4 provides a large and rich set of tools for the coding of audio objects. In order to allow effective implementations of the standard, subsets of the tool set have been identified that can be used for specific applications. The function of these subsets, called "Profiles," is to limit the tool set a conforming decoder must implement. For each of these Profiles, one or more Levels have been specified, thus restricting the computational complexity.

The High Efficiency AAC Profile is introduced as a superset of the AAC Profile. Besides the AOT AAC LC (which is present in the AAC Profile), it includes the AOT SBR. Levels are introduced within these Profiles in such a way, that a decoder supporting the High Efficiency AAC Profile at a given level can decode an AAC Profile stream at the same or lower level.

Level	Max. channels /object	Max. AAC sampling rate, SBR not present [kHz]	Max. AAC sampling rate, SBR present [kHz]	Max. SBR sampling rate, [kHz] (in/out)
1	NA	NA	NA	NA
2	2	48	24	24/48
3	2	48	48	48/48
4	5	48	24/48	48/48
5	5	96	48	48/96

Table 1: Levels within the HE AAC Profile

4.4 HE AAC signaling: Use Scenarios

Contrary to previous sections which described the technical specifications of HE AAC signaling from a MPEG-4 point of view, this section will look at these methods from a system integrator's point of view. On a high level, it can be distinguished between content providers/creators, decoder manufacturers, and system designers (e.g. closed systems for digital broadcasting etc.). Whereas the first group is in control of the encoding side only, and the second group is in control of the decoder only, the third group has the advantage of being in control over the complete signal chain.

Choosing a format The combination of the different signaling methods, as described in Section 4.2, with

Google Exhibit No. 1063
Page 1360

the different transport multiplexes as defined in Section 2.3, results in a large variety of bitstream formats for HE AAC. It should be noted that these different transport formats do not have any effect on the format of AAC raw_data_block. Thus transforming content from one into another transport format is a rather simple process. Therefore, as indicated in Fig. 3, a content provider might store data in the MPEG-4 file format (MP4FF) and then repackage it for the different distribution channels and configurations. On the other hand a decoder manufacturer needs to decide up-front which formats to support. It is conceivable that many MPEG-4 decoder manufacturers will include support for the ADIF and ADTS formats in order to be able to play back the widest possible range of bitstream formats.

4.4.1 HE AAC for content providers

As a content provider/creator, one can decide who will be able to decode the bitstreams. In general, HE AAC is backward compatible to existing AAC-only decoders. However, using decoders without the SBR tool will of course result in lower audio quality. It is possible to use signaling such that only users with new HE AAC decoders can decode the content. This is useful in order to guarantee the full audio quality for all listeners. In such a scenario, backwards compatibility to existing, older, AAC-only decoders is explicitely disabled.

Supporting only HE AAC decoders This can be achieved by indicating the new HE AAC profile. Legacy AAC decoders will not play back such audio content. On the other hand new HE AAC decoders have to use both, AAC and SBR decoder tools.

Supporting legacy MPEG-4 AAC-only decoders There are two signaling methods that will allow even legacy MPEG-4 AAC-only decoders to decode new HE AAC content. See Section 4.2 for details. In either case legacy AAC decoders will skip the SBR data, while new HE AAC decoders use that data to configure and call the SBR tool.

Supporting MPEG-2 and MPEG-4 decoders As long as implicit signaling for the SBR tool and the AAC LC AOT without the PNS tool are used, HE AAC can be formatted such that both MPEG-2 and MPEG-4 decoders can decode the content. MPEG is preparing a new release of the MPEG-2 standard to support this configuration [28]. Thus it is possible to create HE AAC content which can be decoded by MPEG-2 and MPEG-4 AAC decoders. Such HE AAC bitstreams will be decoded by AAC-only and HE AAC decoders, thus guaranteeing the largest number of compatible decoders.

4.4.2 HE AAC for decoder manufacturers

The MPEG standard defines profiles and levels to support interoperability between different encoder and decoder implementations. In general a decoder manufacturer is free to support any subset of tools as required. However, a *MPEG conformant* decoder needs to implement at least the subset of tools specified for a specific profile and level. Ultimately it will be important that the decoder can decode the desired content, which is dependent on the settings used by the content creator. An MPEG conformant encoder will also take into account the different profiles and levels so that interoperability between encoders and decoders is guaranteed. However, one cannot foresee at this point, what kind of levels might be used mostly by content creators.

Computational complexity Specifically for manufacturers of embedded decoders the computational complexity is a main concern. The MPEG standard allows for controling the computational complexity of a HE AAC decoder by means of the following three methods:

1. In situations where the listening environment is not as demanding, a specific low-power decoder can be used. See Section 3.2 for details on this tool, and Section 5.1 for detailed complexity assessments.
2. In general the SBR decoder operates at double the sampling rate of the AAC core coder. (Dual-rate system). However, in situations where a lower output sampling rate is sufficient (e.g. some portable devices) the SBR tool can run in a down-sampled mode, where the input and output sampling rate are identical. (See Section 3.3 for details.)
3. As the complexity increases with increasing levels it is important to only support the required level. The different levels differ in the maximum number of channels as well as the maximum sampling rates of the AAC core and SBR tool as indicated in Table 1.

Google Exhibit No. 1063
Page 1361

Following a few examples of how these options could be combined:

- *Basic configuration for cell phones:* Since the listening environment is not very demanding but computational resources might be extremely limited, a down-sampled, low-power HE AAC decoder supporting level 2 of the HE AAC profile might be suitable.
- *Basic configuration of a portable flash player:* The listening environment might be more demanding than in the case of a cell phone. A high quality HE AAC decoder supporting level 2 of the HE AAC profile might be suitable.
- *Basic configuration of a high-end hard-disk player:* Assuming the player does not support multichannel playback, a high quality HE AAC decoder supporting level 2 and 3 of the HE AAC profile might be suitable. Level 3 will allow to play back high-quality, high-bitrate files that utilize a samplerate of up to 48 kHz for the AAC core coder.
- *Basic configuration of a desktop decoder:* A high-quality HE AAC decoder supporting levels 2 to 5 of the HE AAC profile might be suitable for a state of the art desktop decoder. This will allow play back of up to 5 channels and an output sampling rate of up to 96 kHz.

4.4.3 HE AAC for digital broadcasting

The category of digital broadcasting applications includes proprietary digital radio systems such as XM Satellite Radio [29], digital radio standards such as Digital Radio Mondiale (DRM) [30], Internet-radio applications, or standards such as HDTV, DVD etc. These scenarios are similar in that they describe closed systems where MPEG *conformance* is not of major concern. However, using tools that are described in the MPEG standard and combining them into a new system is not only possible, but is a major concept behind the complete MPEG-4 standard. Although conformance and thus specific profiles and levels are unimportant, computational complexity of the decoder continues to be an area of major concern. So the same concepts as in the previous section also apply here.

A special case is digital broadcasting over noisy channels. For such a situation, a system that can adapt to the quality of service of the transmission channel is desireable. MPEG-4 introduces scalable audio codecs for such a scenario and the SBR tool supports and enhances this configuration as well. Section 4.1 includes an example for such a scalable system.

5 IMPLEMENTATIONS AND APPLICATIONS

With HE AAC, MPEG once more standardized a leading technology that is suitable for many existing and emerging markets and applications. The coding efficiency is high enough to support multichannel applications at low bitrates. A 5.1 bitstream at 112 kbit/s has been demonstrated and can support areas such as audio/video applications.

MPEG-4 HE AAC is a proven technology which has been widely deployed and is ready for use today. Reference decoder source code will soon be made available through MPEG and optimized source code for both encoders and decoders is available for license from companies such as Coding Technologies. Optimized binary implementations for Win32, Linux, MacOS X, ST Micro, ARM, Motorola, TI C64x & C55x, and Trimedia are presently available and other firmware implementations are being completed.

This section will first list the computational requirements of HE AAC decoder implementations. It will be shown, that despite the fact that the quality of HE AAC compared to plain AAC is significantly increased, the computational complexity of HE AAC is still comparable to a plain AAC decoder. In case of the already mentioned Low power decoder, the computational complexity is even identical to that of a plain AAC decoder. Thus HE AAC can indeed be implemented on low-power devices such as cell phones, enabling the delivery of full-band audio signals at a quality level that was not conceivable until today.

Section 5.2 then lists a few examples of applications and use scenarios that are well suited for HE AAC.

5.1 Computational complexity

The computational complexity of decoders is a major factor in deploying codecs in real-world applications. Specifically in mobile devices, computational complexity and memory requirements are directly related to power consumption and implementation cost. HE AAC was specifically designed to minimize these resources.

Google Exhibit No. 1063
Page 1362

Since HE AAC adds additional algorithms to the decoding process compared to plain AAC, one might expect a significant increase in computational complexity. However, as described in Section 3.3, the AAC core typically runs at half the output sampling rate. Thus the AAC core only requires about 50 % of the computational complexity compared to an AAC core running at the full sampling rate. The remaining resources are then being used by the additional SBR tool.

The MPEG standard indicates the computational complexity of various profiles, levels, and decoders by means of an abstract assessment based on Processor Complexity Units (PCU) and RAM Complexity Units (RCU). Although these numbers are suitable as a general, platform independent indicator of the computational complexity, a set of real-world numbers for embedded implementations is often a better approach. The numbers published in this paper are derived from implementations based on Coding Technologies Fixed-Point Firmware Reference Code (FFR) [31] [32] which is optimized in terms of memory usage and processing power. This code has been ported to various platforms so that the resulting performance numbers provide a realistic overview of the complexity.

Table 2 indicates the distribution of processing power between the AAC and the SBR decoder for a 24-bit and a 16-bit processor. All numbers given in this document relate to 44.1 kHz stereo operation. To achieve the superior audio quality of HE AAC a data width of at least 20 bits is required. Thus large parts of the SBR decoder need to be implemented using double precision arithmetic on devices with smaller data width, such as common 16-bit devices. Table 2 shows this difference in computational complexity.

Low power decoder As already described in Section 3.2, the MPEG SBR tool is available in two versions: High-quality (HQ) and Low power (LP). The difference only affects the decoding process. The encoder and the bitstream are identical and do not contain any information regarding these two versions. Which decoding process is used can be chosen depending on the application and the need to reduce power consumption and/or chip size. The high-quality version offers better audio quality at the cost of greater computational complexity. It should be used in situations where audio quality is most important. The Low power version results in a reduced audio quality but also requires less computational resources for both memory and complexity. Typical scenarios for the Low power version are mobile devices where complexity and thus power consumption is critical while the listening environment is usually not as demanding as in other applications. Table 2 shows that the computational complexity of the Low power decoder does not exceed the requirements of a plain AAC decoder. Thus better compression technology at the same computational cost is achieved.

	AAC	SBR	Total
HE AAC @24 bit	5	3	8
HE AAC @16 bit	9	4	13

Table 3: Distribution of ROM usage in KWords between AAC decoder and SBR decoder

Memory requirements Tables 3 and 4 summarize the memory requirements of the different HE AAC decoder implementations. The total data memory required can be as low as 20 KWords depending on the target platform and the chosen decoder implementation.

5.2 Applications and Use scenarios

The unique capability of HE AAC to achieve high quality at very low bitrates not only enhances existing markets but also enables new markets for digital audio. Where transmission bandwidth is constrained, the value of "High-efficiency AAC" is magnified. It not only enhances audio-only services, but also video services like digital TV. By coupling HE AAC with MPEG-4 Video, more bits can be allocated to the video signal without degrading the quality of the audio signal. This is true for mono, stereo, and multichannel applications. Combined with the new MPEG/ITU Advanced Video Coding (AVC) standard (included in MPEG-4 as part 10), even more significant gains in quality are possible.

Licensing Licensing of the SBR technology follows the simple licensing model of MPEG-4 AAC with annual caps for personal computer applications, per-unit fees for hardware devices and no additional fees for electronic music distribution. Licensing for HE AAC consists of two parts with schedules and details available from Coding Technologies [33] for the SBR object types and from Via Licensing [34] for the AAC object types.

Google Exhibit No. 1063
Page 1363

	AAC[a]	SBR	Total
HE AAC HQ @ 24-bit	10	15-20	25-30
HE AAC LP @ 24-bit	10	ca. 10	ca. 20
HE AAC HQ @ 16-bit	20	30-40	50-60
HE AAC LP @ 16-bit	20	ca. 20	ca. 40
plain AAC (full sampling rate) @ 24-bit	20	-	20
plain AAC (full sampling rate) @ 16-bit	40	-	40

[a] runs at half sampling rate in HE AAC

Table 2: Distribution of computational complexity in MIPS between the AAC and SBR decoder (stereo @ 44.1 kHz f_s). The complexity on 16-bit devices is largely dependent on the number of cycles needed to execute a double precision multiply-accumulate (MAC) instruction. Above estimates assume three cycles per double-precision MAC.

	input	AAC		SBR		output	Total
		static	dynamic	static	dynamic		
HE AAC HQ @24 bit	0.7	2.5	2	4	3	4	≈ 15
HE AAC LP @24 bit	0.7	2.5	2	ca. 3	ca. 1	4	≈ 12
HE AAC HQ @16 bit	0.7	4	4	7	8	4	≈ 24
HE AAC LP @16 bit	0.7	4	4	ca. 6	ca. 4	4	≈ 19

Table 4: Distribution of RAM usage in KWords between the AAC core and SBR (stereo)

Mobile streaming and download There is significant expectation for mobile multimedia services associated with the new 2.5G and 3G mobile service networks. Streaming video and other high-bandwidth services have been showcased as the ideal applications for this new infrastructure. The problem is that for much of the coverage area the peak bandwidth in these networks is generally around 144 kbit/s with individual users sustaining connections of about 40 kbit/s [35]. Delivering quality video over this type of connection is problematic and may not meet the consumer expectation. HE AAC is well suited to solve this problem by being able to provide consumer-grade download and streaming audio services within today's bandwidth. 48 kbit/s HE AAC provides CD-quality stereo programming while at 32 kbit/s, it provides excellent quality stereo programming. These bitrates combined with the growing market for subscription audio services show a strong business opportunity starting in 2003 on into 2004 and beyond.

Digital broadcast via satellite and cable *aacPlus* gained its first commercial success with XM Satellite Radio [29]. It allowed XM Radio to launch the most successfull digital radio system in the US to date. *aacPlus* was also selected by the Digital Radio Mondiale [30] consortium as part of the standard for Shortwave and AM digital radio. This heritage has brought credibility for HE AAC in the open standards world of digital satellite and cable broadcasting. As operators look to enhance their services with more channels or with high-definition, the efficiency of HE AAC gives them more options to either consolidate audio bandwidth to make room for more video or to layer more audio services like multilingual and 5.1 surround. Since Spectral Band Replication (SBR) is being added to the MPEG-2 standard as well, operators have the flexibility to use HE AAC within the MPEG-2 or MPEG-4 context as desired.

Internet distribution Video on demand and subscription content services are continuing to grow on the Internet. In these services, aggregate server bandwidth usage and last-mile bandwidth constraints make it difficult for operators and aggregators to provide high-quality, reliable services to consumers. By cutting the bandwidth requirements for audio nearly in half, MPEG-4 both reduces costs and increases reliability for Internet content services. In addition the flexible signaling of HE AAC within the MPEG-4 framework as described in Section 4.2 allows to control the user's listening experience. The content provider can for example decide if

Google Exhibit No. 1063
Page 1364

the new HE AAC bitstreams can be played back on plain AAC decoders as well as on HE AAC decoders. This trade off between backward compatibility and guaranteed audio quality at the decoder can be chosen for individual bitstreams.

6 CONCLUSIONS

The flexible integration of the new MPEG SBR tool into the MPEG-4 framework allows to control the backwards compatibility of HE AAC. Audio content can be packaged such that only HE AAC decoders can play back the audio streams or it can be packaged so that legacy AAC decoders can play back the new content as well, albeit at a lower audio quality. As such MPEG SBR integrates perfectly into MPEG-4 and enhances this open toolbox with an important audio coding method.

MPEG-4 also provides several transport protocols that support streaming applications as well as file based storage. The MPEG-4 framework guarantees the easy integration of HE AAC with MPEG video, thus combining state of the art video coding with state of the art audio coding. High quality multichannel and multilingual audio content at low bitrates thus becomes available for deployment in various audio/video application, enabling a whole new market for content storage and distribution.

In addition MPEG standardized for the first time two audio decoders that can both decode the same audio content: A high-quality version as well as a low power version. The availability of both decoders for HE AAC enables the standard to run on the widest possible range of processors in mobile and portable device applications. The flexibility in content creation as well as decoder implementation make HE AAC a highly effective yet generic MPEG-4 profile. This reflects the true spirit of the MPEG-4 standard: Providing a set of state of the art multimedia technologies that can be combined to support all multimedia applications.

7 REFERENCES

[1] ISO/IEC JTC1/SC29/WG1, "Text of ISO/IEC 14496-3:2001/FDAM1, Bandwidth Extension," ISO/IEC JTC1/SC29/WG11 N5570, Mar. 2003.

[2] ISO/IEC JTC1/SC29/WG11, "Overview of the MPEG-4 standard," ISO/IEC JTC1/SC29/WG11 N4030, Mar. 2001, available: `http://mpeg.telecomitalialab.com/standards/mpeg-4/mpeg-4.htm`

[3] ISO/IEC JTC1/SC29/WG11, "Official MPEG Home Page," available: `http://mpeg.telecomitalialab.com/`.

[4] F. Pereira and T. Ebrahimi, Eds., *The MPEG-4 Book*, Prentice Hall, Englewood Cliffs, NJ, US, 2002.

[5] ISO/IEC, "Coding of audio-visual objects – Part 3: Audio (MPEG-4 Audio, 2nd edition)," ISO/IEC Int. Std. 14496-3:2001, 2001.

[6] ISO/IEC, "Coding of audio-visual objects – Part 2: Visual (MPEG-4 Visual, 2nd edition)," ISO/IEC Int. Std. 14496-2:2001, 2001.

[7] ISO/IEC, "Coding of audio-visual objects – Part 1: Systems (MPEG-4 Systems, 2nd edition)," ISO/IEC Int. Std. 14496-1:2001, 1999.

[8] ISO/IEC, "Generic coding of moving pictures and associated audio information – Part 7: Advanced Audio Coding (AAC)," ISO/IEC Int. Std. 13818-7:1997, 1997.

[9] K. Brandenburg, "Perceptual coding of high quality digital audio," in *Applications of Digital Signal Processing to Audio and Acoustics*, M. Kahrs and K. Brandenburg, Eds., chapter 2, pp. 39–83. Kluwer, Boston, 1998.

[10] T. Painter and A. Spanias, "Perceptual coding of digital audio," *Proc. IEEE*, vol. 88, no. 4, pp. 451–513, Apr. 2000.

[11] M. Bosi, K. Brandenburg, S. Quackenbush, L. Fielder, K. Akagiri, H. Fuchs, M. Dietz, J. Herre, G. Davidson, and Y. Oikawa, "ISO/IEC MPEG-2 Advanced Audio Coding," *J. Audio Eng. Soc.*, vol. 45, no. 10, pp. 789–814, Oct. 1997.

[12] J. Herre and H. Purnhagen, "General audio coding," in *The MPEG-4 Book*, F. Pereira and T. Ebrahimi, Eds., chapter 11. Prentice Hall, Englewood Cliffs, NJ, US, 2002.

[13] C. Herpel, G. Franceschini, and D. Singer, "Transporting and storing MPEG-4 content," in *The MPEG-4 Book*, F. Pereira and T. Ebrahimi, Eds., chapter 7. Prentice Hall, Englewood Cliffs, NJ, US, 2002.

Google Exhibit No. 1063
Page 1365

[14] H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson, "RTP: A transport protocol for real time applications," IETF RFC 1889, Jan. 1996, available: `http://www.ietf.org/rfc/rfc1889.txt`.

[15] Y. Kikuchi, T. Nomura, S. Fukunaga, Y. Matsui, and H. Kimata, "RTP payload format for MPEG-4 audio/visual streams," IETF RFC 3016, Nov. 2000, available: `http://www.ietf.org/rfc/rfc3016.txt`.

[16] ISO/IEC JTC1/SC29/WG11, "Text of ISO/IEC FDIS 14496-8, Carriage of ISO/IEC 14496 contents over IP networks," ISO/IEC JTC1/SC29/WG11 N4712, Mar. 2002.

[17] ISO/IEC, "Coding of audio-visual objects – Part 3: Audio (MPEG-4 Audio Version 1)," ISO/IEC Int. Std. 14496-3:1999, 1999.

[18] ISO/IEC, "Coding of audio-visual objects – Part 3: Audio, AMENDMENT 1: Audio extensions (MPEG-4 Audio Version 2)," ISO/IEC Int. Std. 14496-3:1999/Amd.1:2000, 2000.

[19] ISO/IEC JTC1/SC29/WG11, "Call for evidence justifying the testing of audio coding technology," ISO/IEC JTC1/SC29/WG11 N3483, July 2000, available: `http://www.tnt.uni-hannover.de/project/mpeg/audio/public/w3483.pdf`.

[20] ISO/IEC JTC1/SC29/WG11, "Call for proposals for new tools for audio coding," ISO/IEC JTC1/SC29/WG11 N3794, Jan. 2001, available: `http://www.tnt.uni-hannover.de/project/mpeg/audio/public/w3794.pdf`.

[21] M. Dietz, L. Liljeryd, K. Kjörling, and O. Kunz, "Spectral Band Replication, a novel approach in audio coding," in *112th AES Convention, Munich, May 10 - 13*, 2002, Preprint 5553.

[22] S. Meltzer, F. Henn, and R. Böhm, "SBR enhanced audio codecs for digital broadcasting such as "Digital Radio Mondiale" (DRM)," in *112th AES Convention, Munich, May 10 - 13*, 2002, Preprint 5559.

[23] T. Ziegler, A. Ehret, P. Ekstrand, and M. Lutzky, "Enhancing mp3 with SBR: Features and Capabilities of the new mp3PRO Algorithm," in *112th AES Convention, Munich, May 10 - 13*, 2002, Preprint 5560.

[24] A. Ehret, M. Dietz, and K. Kjörling, "State-of-the-Art Audio Coding for Broadcasting and Mobile Applications," in *114th AES Convention, Amsterdam, March 22 - 25*, 2003.

[25] P. Ekstrand, "Bandwidth Extension of Audio Signals by Spectral Band Replication," in *IEEE Benelux Workshop on Model based Processing and Coding of Audio (MPCA-2002), Leuven, Belgium, Nov. 15*, 2002.

[26] MPEG-4 Industry Forum, "MPEG-4 - The Media Standard," Nov. 2002, available: `http://www.mpegif.org/public/documents/vault/m4-out-20027.pdf`.

[27] ISO/IEC JTC1/SC29/WG11, "MPEG-4 Audio verification test results: Audio on the Internet," ISO/IEC JTC1/SC29/WG11 N2425, Oct. 1998, available: `http://www.tnt.uni-hannover.de/project/mpeg/audio/public/w2425.pdf`.

[28] ISO/IEC JTC1/SC29/WG11, "Text of ISO/IEC 13818-7:2003/FPDAM 1, Signalling of Bandwidth Extension," ISO/IEC JTC1/SC29/WG11 N5568, Mar. 2003.

[29] "XM Satellite Radio," WWW page, 2003, `http://www.xmradio.com/`.

[30] "Digital Radio Mondiale," WWW page, 2003, `http://www.drm.org`.

[31] Coding Technologies, "Embedded aacPlus Decoder Implementations - Product Sheet," June 2003.

[32] Coding Technologies, "Embedded HE-AAC: Implementation Guideline," June 2003.

[33] "Coding Technologies," WWW page, 2003, `http://www.codingtechnologies.com`.

[34] "Via Licensing," WWW page, 2003, `http://www.vialicensing.com`.

[35] D. Frerichs, "New MPEG-4 High-efficiency AAC Audio," Apr. 2003, available: `http://www.mpegif.org/public/documents/vault/m4-out-30034.zip`.

Google Exhibit No. 1063
Page 1366

Ex. 1041B

CONFERENCE PROCEEDING

[Convention papers : presented at the 115th Convention : 2003, October 10-13, New York, N.Y. / Audio Engineering Society]

Audio Engineering Society. Convention (115th : 2003 : New York, N.Y.); Audio Engineering Society

2003

Available at Linda Hall Library Closed Stacks - Books (TK7881.4 .A9235 115th 2003) >

Top
Send to
Get It
Details
Virtual Browse
Links

Send to

Email | QR | Permalink | Citation

EasyBib | Export RIS | Export to Excel

Get It

Back to Locations

LOCATION ITEMS

Linda Hall Library
Available , Closed Stacks - Books ; TK7881.4 .A9235 115th 2003

Item in place (0 requests)
Loanable

Item in place (0 requests)
Loanable

Item in place (0 requests)
Loanable

Details

Title	[Convention papers : presented at the 115th Convention : 2003, October 10-13, New York, N.Y. / Audio Engineering Society]
Creator/Contributor	Audio Engineering Society. Convention (115th : 2003 : New York, N.Y.) > Audio Engineering Society >
Subjects	Electroacoustics > Acoustical engineering >
Identifier	OCLC : (OCoLC)56052226
Publisher	New York, N.Y. : Audio Engineering Society
Creation date	2003
Format	[139] papers : ill. ; 28 cm.
General notes	Caption title. Paper nos. 5854-5993 with omissions. Includes bibliographical references.

Virtual Browse

ers ented at 111th vention, ... [Papers presented at the 112th Convention of the Audio ... [Papers presented at the 113th Convention of the Audio ... [Papers presen the 114 Conver the Auc

Links

Display Source Record

Linda Hall Library MARC record for Convention papers : presented at the 115th Convention copied from https://catalog.lindahall.org/discovery/sourceRecord?vid=01LINDAHALL_INST:LHL&docId=alma993980363405961&recordOwner=01LINDAHALL_INST

leader 00927nam a2200253Ia 4500
001 993980363405961
005 20050214081800.0
008 040729s2003 nyua b 100 0 eng d
035 ##$a(OCoLC)56052226
035 ##$9547577
035 ##$a(MoKL)398036-lhalldb
035 ##$a(lhalldb)398036-lhalldb
040 ##$aLHL $cLHL
049 ##$aLHLA
090 ##$aTK7881.4 $b.A9235 115th 2003
110 2#$aAudio Engineering Society. $bConvention $n(115th : $d2003 : $cNew York, N.Y.)
245 10$a[Convention papers : $bpresented at the 115th Convention : 2003, October 10-13, New York, N.Y. / $cAudio Engineering Society].
260 ##$a[New York, N.Y. : $bAudio Engineering Society, $c2003]
300 ##$a[139] papers : $bill. ; $c28 cm.
500 ##$aCaption title.
500 ##$aPaper nos. 5854-5993 with omissions.
504 ##$aIncludes bibliographical references.
650 #0$aElectro-acoustics $vCongresses.
650 #0$aAcoustical engineering $vCongresses
710 2#$aAudio Engineering Society
948 ##$aLTI 09/01/2008

Ex. 1041C

libraries.org

A directory of libraries throughout the world

found **2** items where the item contains **Linda hall Library**. Showing item **1** of **2**

Linda Hall Library

Linda Hall Library of Science, Engineering and Technology

Address: 5109 Cherry Street
Kansas City, Missouri
64110-2498
United States

County: Jackson

Phone: 816-363-4600

Connect to: Library Web Site✔ Online Catalog

Library Hours		
Day	Open	Close
Sunday	Closed	
Monday	9:00 am	5:00 pm
Tuesday	9:00 am	5:00 pm
Wednesday	9:00 am	5:00 pm
Thursday	9:00 am	5:00 pm
Friday	9:00 am	5:00 pm
Saturday	Closed	



Linda Hall Library



map location✔

Library details: Linda Hall Library is a **Independent Research** library. This library is affiliated with **Linda Hall Library (view map)** . The collection of the library contains **860,000** volumes. The library circulates **30,000** items per year.

Other Info: The Linda Hall Library is the world's foremost independent research library devoted to science, engineering, and technology. A not-for-profit, privately funded institution, the Library is open to the public free of charge and does circulate. This library shares an ILS with the Spencer Art Reference Library (of the Nelson-Atkins Museum of Art, also in Kansas City, MO). The combined catalog was formerly called LEONARDO (named for the artist/inventor Leonardo da Vinci). LHL is a member of the Independent Research Libraries Association.

Permalink: https://librarytechnology.org/library/2294
(Use this link to refer back to this listing.)

Administration: The director of the library is **Eric Dorfman** .

Organizational structure: The library is part of a **privately-funded** organization. It operates on a **Non-profit** financial model.

Collection size	860,000	volumes
Annual Circulation	30,000	transactions

Technology Profile		
	Product Name	Year Contracted
Current Automation System	Alma	2018

Related Libraries

- Libraries located in **Kansas City** (Missouri)
- Libraries located in **Jackson** county (Missouri)
- **View map of libraries in Jackson County**
- **all Independent Research libs in Missouri**

About libraries.org

libraries.org is a free directory of libraries throughout the world.

This directory is a component of **Library Technology Guides**.

Google Exhibit No. 1063
Page 1373

Previous Automation System	**Voyager**	2010
Previous Automation System	**Horizon**	1994
Previous Automation System	**None**	
Discovery Service (w/index)	**Ex Libris CDI**	2010
Discovery Interface	**Primo**	**2010**
Digital Collection Management	**CONTENTdm**	
Web Content Management	**WordPress**	
Item ID Type	Barcode	

The library's automation system is hosted by the vendor through an **Software-as-a-Service** (**SaaS**) arrangement.

This library is responsible for the procurement of the library automation system.

Please **contact Marshall Breeding** to report corrections about information listed for this library.

Automation Survey: We are conducting a study on the automation systems used in libraries. This survey should be completed by a person familiar with your libraries use of **Alma**.

Respond to the 2024 Library Automation Perceptions Survey

Identifiers	
libraries.org ID	2294
OCLC Symbol	LHL
WorldCat Registry ID	**99283**

Record History: This listing was created on Aug 11, 1998 and was last modified on Dec 10, 2024.

Updates: Corrections or Updates? Registered members of Library Technology Guides can submit updates to library listings in libraries.org. **Registration** is free and easy. Already registered? **login**. Or, you can report corrections just by sending a message to **Marshall Breeding**.

libraries.org | Update this Entry | Library Technology Guides

Maintained by **Marshall Breeding**

terms of service
Copyright 1994-2024

Google Exhibit No. 1063
Page 1374

Ex. 1041D

Government of Canada / Gouvernement du Canada

Record number

b2015693

Search

Save this record | MARC display

Corporate author	Audio Engineering Society. Convention (115th : 2003 : New York, N.Y.)
Title	**[Convention papers : presented at the 115th convention, 2003 October 10-13, New York, New York / Audio Engineering Society]**
Publisher	[New York, N.Y. : Audio Engineering Society, 2003]

Request this item

Location	Call number & Local holdings	Status
NRC Ottawa	TK5981 A91 2003 Oct no. 5854-5920	In library use only

Google Exhibit No. 1063
Page 1376

NRC Ottawa	TK5981 A91 2003 Oct no. 5921-5993	In library use only

Description	[136] papers : ill. ; 28 cm.
Note	Title supplied by cataloguer.
Corporate name	Audio Engineering Society
Subject	Electro-acoustics -- Congresses.
	Acoustical engineering -- Congresses.

Permanent link

Save this record | MARC display

Date modified:
2024-07-11

Google Exhibit No. 1063
Page 1377

Government of Canada / Gouvernement du Canada

MARC display

Save this record | Return to record

```
LEADER 00000cam a2200000 a 4500
008    031212s2003    nyua          100 0 eng d
040    CaOON|beng|cCaOON
090    TK5981 A91 2003 Oct
110 2  Audio Engineering Society.|bConvention|n(115th :|d2003 :
       |cNew York, N.Y.)
245 10 [Convention papers :|bpresented at the 115th convention,
       2003 October 10-13, New York, New York /|cAudio
       Engineering Society]
260    [New York, N.Y. :|bAudio Engineering Society,|c2003]
300    [136] papers :|bill. ;|c28 cm.
500    Title supplied by cataloguer.
583 1  committed to retain|c20240522|d25 years|fFSLN Shared Print
       Program|5oon
590    n-bb
650  0 Electro-acoustics|vCongresses.
650  0 Acoustical engineering|vCongresses.
710 2  Audio Engineering Society
```

Location	Call number & Local holdings	Status
NRC Ottawa	TK5981 A91 2003 Oct no. 5854-5920	In library use only

Google Exhibit No. 1063
Page 1378

NRC Ottawa	TK5981 A91 2003 Oct no. 5921-5993	In library use only

Date modified:

2024-07-11

Google Exhibit No. 1063
Page 1379

FirstSearch

University Libraries

Libraries that Own Item

| Back to CUA Libraries

- This screen shows libraries that own the item you selected.

Title: [Convention papers : presented at the 115th convention, 2003 October 10-13, New York, New York Author: Audio Engineering Society Accession Number: 927388398

Libraries with Item: "[Convention papers : / pr..."(Record for Item | Get This Item)

Location	Library	Code
CA,ON	NATIONAL SCI LIBRARY/BIBLIO SCI NAT	CAI

Record for Item: "[Convention papers : / pr..."(Libraries with Item)

GET THIS ITEM

Availability: Check the catalogs in your library.

- Libraries worldwide that own item: 1

External Resources:

- Find it@CU Find or Request It at CU
- ILL@CU Request it via ILL
- Cite This Item

FIND RELATED

More Like This: Advanced options ...

Find Items About: Audio Engineering Society. (max: 14)

Title: **[Convention papers : presented at the 115th convention, 2003 October 10-13, New York, New York /**

Corp Author(s): Audio Engineering Society. Convention (115th : 2003 : New York, N.Y.); Audio Engineering Society.

Publication: [New York, N.Y.] : [Audio Engineering Society],

Year: 2003

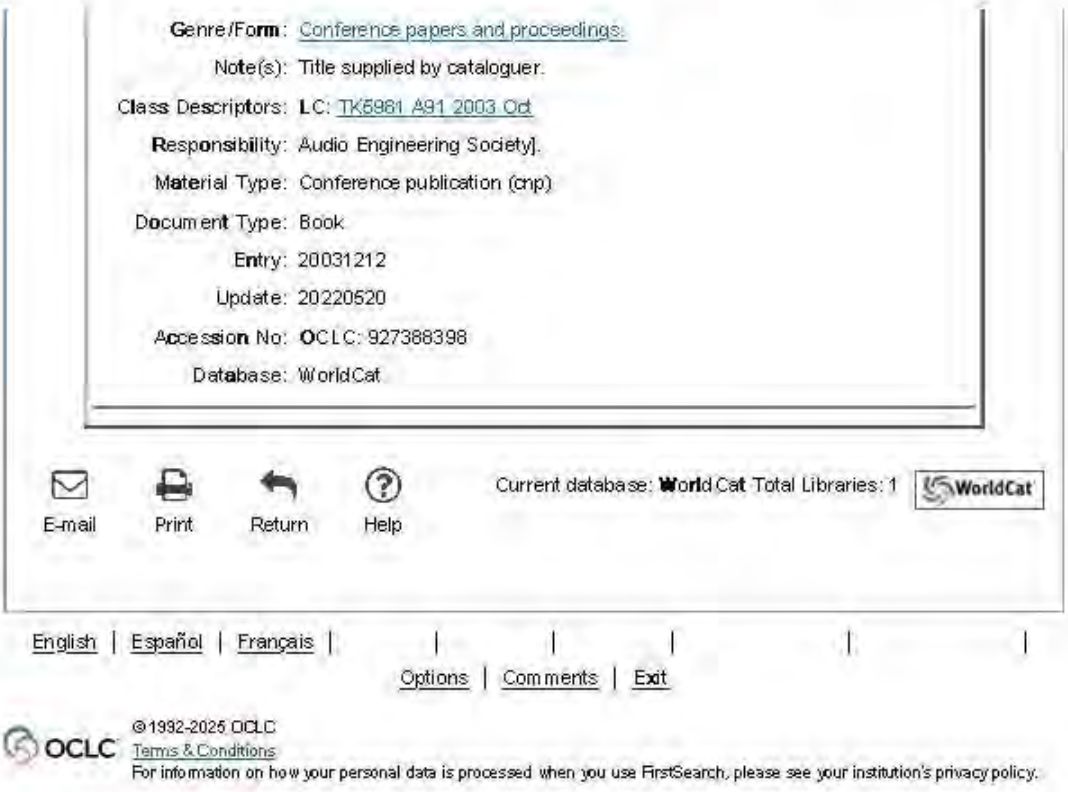
Description: [136] papers : illustrations ; 28 cm

Language: English

SUBJECT(S)

Descriptor: Electro-acoustics -- Congresses.
Acoustical engineering -- Congresses.
Électroacoustique -- Congrès.
Acoustique appliquée -- Congrès.
Acoustical engineering.
Electro-acoustics.

Google Exhibit No. 1063
Page 1380



OCLC ©1992-2025 OCLC
Terms & Conditions
For information on how your personal data is processed when you use FirstSearch, please see your institution's privacy policy.

Ex. 1041E

https://web.archive.org/web/20030831165223/http://www.aes.org/events/115/papers/SessionC.cfm
http://www.aes.org/events/115/papers/SessionC.cfm
Go
20 captures
31 Aug 2003 - 13 Apr 2024
JUL AUG OCT
2002 2003 2005

Audio Engineering Society
Conventions and Conferences

Registration
Exhibitors
Detailed Calendar (in Excel)
Calendar (in PDF)
Surround - Live: Symposium
Paper Sessions
Tutorial Seminars
Workshops
Special Events
Exhibitor Seminars



Friday, October 10 2:00 pm – 4:30 pm
Session C Low Bit-Rate Audio Coding

C-1 Scalable Perceptual and Lossless Audio Coding Based on MPEG-4 AAC—*Ralf Geiger, Gerald Schuller,* Fraunhofer IIS AEMT, Ilmenau, Germany; *Jürgen Herre , Ralph Sperschneider,* Fraunhofer IIS AEMT, Erlangen, Germany; *Thomas Sporer*, Fraunhofer IIS AEMT, Ilmenau, Germany
This paper presents a scalable lossless enhancement of MPEG-4 Advanced Audio Coding (AAC). Scalability is achieved in the frequency domain using the Integer Modified Discrete Cosine Transform (IntMDCT), which is an integer approximation of the MDCT providing perfect reconstruction. With this transform, and only minor extension of the bitstream syntax, the MPEG-4 AAC Scalable codec can be extended to a lossless operation. The system provides bit-exact reconstruction of the input signal independent of the implementation accuracy of the AAC core coder. Furthermore, scalability in sampling rate and reconstruction word length is supported.

C-2 Robust MPEG Advanced Audio Coding Over Wireless Channels—*T. H. Yeo*, National University of Singapore, Singapore; *W. C. Wong*, National University of Singapore, Singapore, Institute for Infocomm Research, Singapore; *Dong-Yan Huang*, Institute for Infocomm Research, Singapore
In this paper a concatenated system combining turbo product codes and convolutional codes with soft

The Wayback Machine - https://web.archive.org/web/20030831165223/http://www.aes.org:80/events/115/papers/SessionC...

Audio Engineering Society
Conventions and Conferences

Registration
Exhibitors
Detailed Calendar (in Excel)
Calendar (in PDF)
Surround - Live: Symposium
Paper Sessions
Tutorial Seminars
Workshops
Special Events
Exhibitor Seminars
Technical Tours
Student Program
Historical
Standards Mtgs
Hotel Information
Return to 115th



Friday, October 10 2:00 pm – 4:30 pm
Session C Low Bit-Rate Audio Coding

C-1 Scalable Perceptual and Lossless Audio Coding Based on MPEG-4 AAC—*Ralf Geiger, Gerald Schuller,* Fraunhofer IIS AEMT, Ilmenau, Germany; *Jürgen Herre , Ralph Sperschneider,* Fraunhofer IIS AEMT, Erlangen, Germany; *Thomas Sporer,* Fraunhofer IIS AEMT, Ilmenau, Germany
This paper presents a scalable lossless enhancement of MPEG-4 Advanced Audio Coding (AAC). Scalability is achieved in the frequency domain using the Integer Modified Discrete Cosine Transform (IntMDCT), which is an integer approximation of the MDCT providing perfect reconstruction. With this transform, and only minor extension of the bitstream syntax, the MPEG-4 AAC Scalable codec can be extended to a lossless operation. The system provides bit-exact reconstruction of the input signal independent of the implementation accuracy of the AAC core coder. Furthermore, scalability in sampling rate and reconstruction word length is supported.

C-2 Robust MPEG Advanced Audio Coding Over Wireless Channels—*T. H. Yeo,* National University of Singapore, Singapore; *W. C. Wong,* National University of Singapore, Singapore, Institute for Infocomm Research, Singapore; *Dong-Yan Huang,* Institute for Infocomm Research, Singapore
In this paper a concatenated system combining turbo product codes and convolutional codes with soft decision Viterbi decoding algorithm and diversity is proposed to enhance the robustness of the Unequal Error Protection (UEP) scheme for wireless transmission of MPEG-4 Advanced Audio Coding (AAC). The proposed scheme has been tested over the Gaussian and Rician fading channels. Under severe channels with random, burst, and mixed bit error rates (BER) of 6.00×10^{-2} and above, the proposed scheme provides a 90 percent improvement in residual BER performance, which is approximately 3 dB, with 19 percent increase in bandwidth over the original UEP scheme. At a high channel BER of 6.00×10^{-2}, the proposed scheme gives an error-free header frame. Compared with the concatenated convolutional codes, the proposed scheme provides 0.5 dB BER performance improvements with the same bandwidth. With diversity, the performance can be further improved by 3 dB at low SNR for proposed scheme and 2.5 dB for original UEP scheme.

C-3 MP3 in MPEG-4— *Bernhard Grill, Harald Gernhardt, Michael Härtl, Johannes Hilpert, Manfred Lutzky, Martin Weishart,* Fraunhofer IIS, Erlangen, Germany
MP3 is the most commonly used audio coding scheme worldwide. Today it is present on virtually any computer, CD, DVD-player, many car radios, and of course in the many MP3 audio player devices. This paper describes a new MPEG-4 working draft currently under development by ISO/MPEG, which will allow MP3 to be fully integrated into MPEG-4 audio-visual systems. Among other new features, such as enhanced editability, MP3 is scheduled to get a full multichannel audio capability, which can be implemented with a few additional lines of code on top of a standard stereo MP3 codec.

C-4 A Closer Look into MPEG-4 High Efficiency AAC—*Martin Wolters,* Coding Technologies, Nürnberg, Germany; *Kristofer Kjörling,* Coding Technologies, Stockholm, Sweden; *Daniel Homm,* Coding Technologies, Nürnberg, Germany; *Heiko Purnhagen,* Coding Technologies, Stockholm, Germany
MPEG Spectral Band Replication (SBR) is the newest compression technology available as part of the MPEG standards. It is combined with MPEG Advanced Audio Coding (AAC) and improves coding efficiency by more than 30 percent. The resulting scheme is called High-Efficiency AAC (HE-AAC). This paper explains MPEG-SBR and its integration into the existing MPEG-4 bitstream format. The SBR technology itself, as well as the implications on systems based on MPEG-4 technology, are described. The signaling through MPEG-4 systems and other transport formats is introduced and typical applications and usage scenarios are listed.

C-5 MPEG-4 Lossless Coding for High-Definition Audio—*Tilman Liebchen,* Technical University of Berlin, Berlin, Germany
Lossless coding will become the latest extension of the MPEG-4 audio standard. In response to a call for proposals, many companies have submitted lossless audio codecs for evaluation. The codec of the Technical University of Berlin was chosen as reference model for MPEG-4 Audio Lossless Coding, attaining working draft status in July 2003. The encoder is based on linear prediction, which enables high compression even with moderate complexity, while the corresponding decoder is straightforward. The paper describes the basic elements of the codec, points out envisaged applications, and gives an

Google Exhibit No. 1063
Page 1384

outline of the standardization process.



(C) 2003, Audio Engineering Society, Inc.



Google Exhibit No. 1063
Page 1385


https://web.archive.org/web/20031217213041/http://www.aes.org/events/115/papers/SessionC.cfm
133%
http://www.aes.org/events/115/papers/SessionC.cfm
Go
OCT
DEC
FEB
17
20 captures
31 Aug 2003 - 13 Apr 2024
2002
2003
2005
Audio Engineering Society
Conventions and Conferences
Registration
Exhibitors
Detailed Calendar (in Excel)
Calendar (in PDF)
Convention Planner
Surround - Live: Symposium
Paper Sessions
Tutorial Seminars
115th AES Convention
October 10-13, 2003
Jacob K. Javits Convention Center
New York, NY, USA
Friday, October 10 2:00 pm – 4:30 pm
Session C Low Bit-Rate Audio Coding
C-1 Scalable Perceptual and Lossless Audio Coding Based on MPEG-4 AAC—Ralf Geiger, Gerald Schuller, Fraunhofer IIS AEMT, Ilmenau, Germany; Jürgen Herre , Ralph Sperschneider, Fraunhofer IIS AEMT, Erlangen, Germany; Thomas Sporer, Fraunhofer IIS AEMT, Ilmenau, Germany
This paper presents a scalable lossless enhancement of MPEG-4 Advanced Audio Coding (AAC). Scalability is achieved in the frequency domain using the Integer Modified Discrete Cosine Transform (IntMDCT), which is an integer approximation of the MDCT providing perfect reconstruction. With this transform, and only minor extension of the bitstream syntax, the MPEG-4 AAC Scalable codec can be extended to a lossless operation. The system provides bit-exact reconstruction of the input signal independent of the implementation accuracy of the AAC core coder. Furthermore, scalability in sampling rate and reconstruction word length is supported.

The Wayback Machine - https://web.archive.org/web/20031217213041/http://www.aes.org:80/events/115/papers/SessionC...

Audio Engineering Society
Conventions and Conferences

Registration
Exhibitors
Detailed Calendar (in Excel)
Calendar (in PDF)
Convention Planner
Surround - Live: Symposium
Paper Sessions
Tutorial Seminars
Workshops
Special Events
Exhibitor Seminars Room A
Exhibitor Seminars Room B
Technical Tours
Student Program
Historical
Free Hearing Screenings
Heyser Lecture
Tech Comm Mtgs
Standards Mtgs
Press Information
Return to 115th



Friday, October 10 2:00 pm – 4:30 pm
Session C Low Bit-Rate Audio Coding

C-1 Scalable Perceptual and Lossless Audio Coding Based on MPEG-4 AAC—*Ralf Geiger, Gerald Schuller,* Fraunhofer IIS AEMT, Ilmenau, Germany; *Jürgen Herre , Ralph Sperschneider,* Fraunhofer IIS AEMT, Erlangen, Germany; *Thomas Sporer,* Fraunhofer IIS AEMT, Ilmenau, Germany
This paper presents a scalable lossless enhancement of MPEG-4 Advanced Audio Coding (AAC). Scalability is achieved in the frequency domain using the Integer Modified Discrete Cosine Transform (IntMDCT), which is an integer approximation of the MDCT providing perfect reconstruction. With this transform, and only minor extension of the bitstream syntax, the MPEG-4 AAC Scalable codec can be extended to a lossless operation. The system provides bit-exact reconstruction of the input signal independent of the implementation accuracy of the AAC core coder. Furthermore, scalability in sampling rate and reconstruction word length is supported.

C-2 Robust MPEG Advanced Audio Coding Over Wireless Channels—*T. H. Yeo*, National University of Singapore, Singapore; *W. C. Wong*, National University of Singapore, Singapore, Institute for Infocomm Research, Singapore; *Dong-Yan Huang*, Institute for Infocomm Research, Singapore
In this paper a concatenated system combining turbo product codes and convolutional codes with soft decision Viterbi decoding algorithm and diversity is proposed to enhance the robustness of the Unequal Error Protection (UEP) scheme for wireless transmission of MPEG-4 Advanced Audio Coding (AAC). The proposed scheme has been tested over the Gaussian and Rician fading channels. Under severe channels with random, burst, and mixed bit error rates (BER) of 6.00×10^{-2} and above, the proposed scheme provides a 90 percent improvement in residual BER performance, which is approximately 3 dB, with 19 percent increase in bandwidth over the original UEP scheme. At a high channel BER of 6.00×10^{-2}, the proposed scheme gives an error-free header frame. Compared with the concatenated convolutional codes, the proposed scheme provides 0.5 dB BER performance improvements with the same bandwidth. With diversity, the performance can be further improved by 3 dB at low SNR for proposed scheme and 2.5 dB for original UEP scheme.

C-3 MP3 in MPEG-4— *Bernhard Grill, Harald Gernhardt, Michael Härtl, Johannes Hilpert, Manfred Lutzky, Martin Weishart,* Fraunhofer IIS, Erlangen, Germany
MP3 is the most commonly used audio coding scheme worldwide. Today it is present on virtually any computer, CD, DVD-player, many car radios, and of course in the many MP3 audio player devices. This paper describes a new MPEG-4 working draft currently under development by ISO/MPEG, which will allow MP3 to be fully integrated into MPEG-4 audio-visual systems. Among other new features, such as enhanced editability, MP3 is scheduled to get a full multichannel audio capability, which can be implemented with a few additional lines of code on top of a standard stereo MP3 codec.

C-4 A Closer Look into MPEG-4 High Efficiency AAC—*Martin Wolters,* Coding Technologies, Nürnberg, Germany; *Kristofer Kjörling,* Coding Technologies, Stockholm, Sweden; *Daniel Homm,* Coding Technologies, Nürnberg, Germany; *Heiko Purnhagen*, Coding Technologies, Stockholm, Germany
MPEG Spectral Band Replication (SBR) is the newest compression technology available as part of the MPEG standards. It is combined with MPEG Advanced Audio Coding (AAC) and improves coding efficiency by more than 30 percent. The resulting scheme is called High-Efficiency AAC (HE-AAC). This paper explains MPEG-SBR and its integration into the existing MPEG-4 bitstream format. The SBR technology itself, as well as the implications on systems based on MPEG-4 technology, are described. The signaling through MPEG-4 systems and other transport formats is introduced and typical applications and usage scenarios are listed.

C-5 MPEG-4 Lossless Coding for High-Definition Audio—*Tilman Liebchen*, Technical University of Berlin, Berlin, Germany
Lossless coding will become the latest extension of the MPEG-4 audio standard. In response to a call for proposals, many companies have submitted lossless audio codecs for evaluation. The codec of the Technical University of Berlin was chosen as reference model for MPEG-4 Audio Lossless Coding, attaining working draft status in July 2003. The encoder is based on linear prediction, which enables high compression even with moderate complexity, while the corresponding decoder is straightforward. The paper describes the basic elements of the codec, points out envisaged applications, and gives an

Google Exhibit No. 1063
Page 1387

outline of the standardization process.

(C) 2003, Audio Engineering Society, Inc.

Google Exhibit No. 1063
Page 1388

Ex. 1041F



AUDIO ENGINEERING SOCIETY

Publications

What's New | Search | Site Map | Contact Us | Home | About AES

AES Special Publications

Convention Preprints

Single Convention Preprints are available through the AES Preprint Search and Shop facility. At the bottom of this page you can find complete sets.

	Member	*Other*
Single Convention Preprints through #3347 (91st AES Convention)	4.00	5.00
Single Convention Preprints after #3348 (92nd AES Convention)	5.00	10.00

Audio Engineering Society Store

Shopping Instructions

- To select multiple preprints, check the select box below each item you would like to add to your shopping basket and click submit at bottom of page.
- To add single item to your shopping basket, click on the +1 button +1 below item.

The Wayback Machine - https://web.archive.org/web/20040218195627/http://www.aes.org:80/publications/convention_cd...

AES Special Publications

Convention Preprints

Single Convention Preprints are available through the AES Preprint Search and Shop facility. At the bottom of this page you can find complete sets.

	Member	*Other*
Single Convention Preprints through #3347 (91st AES Convention)	4.00	5.00
Single Convention Preprints after #3348 (92nd AES Convention)	5.00	10.00

Audio Engineering Society Store

Shopping Instructions

- To select multiple preprints, check the select box below each item you would like to add to your shopping basket and click submit at bottom of page.
- To add single item to your shopping basket, click on the +1 button +1 below item.

(add to shopping cart)	Item Name	Member	Other
	Convention Preprint Sets		
☐ +1	104th Convention (CD-ROM) (includes 103rd Preprints)	$140.00	$200.00
☐ +1	105th Convention (CD-ROM)	$130.00	$190.00
☐ +1	106th Convention (CD-ROM)	$140.00	$200.00
☐ +1	107th Convention (CD-ROM)	$110.00	$160.00
☐ +1	108th Convention (CD-ROM)	$130.00	$190.00
☐ +1	109th Convention (CD-ROM)	$130.00	$190.00
☐ +1	110th Convention (CD-ROM)	$140.00	$200.00
☐ +1	111th Convention (CD-ROM)	$110.00	$160.00
☐ +1	112th Convention (CD-ROM)	$110.00	$160.00
☐ +1	113th Convention (CD-ROM)	$100.00	$150.00
☐ +1	114th Convention (CD-ROM)	$110.00	$160.00
☐ +1	115th Convention (CD-ROM)	$110.00	$160.00
	Single Convention Preprints	$5.00	$10.00

Add Checked Items to Shopping Cart

Clear Checked Items

Prices subject to change without notice.

Google Exhibit No. 1063
Page 1391

(C) 2004, Audio Engineering Society, Inc.

Google Exhibit No. 1063
Page 1392

Ex. 1041G

https://web.archive.org/web/20040615041014/http://aes.org/publications/preprints/lists/115.cfm
http://aes.org/publications/preprints/lists/115.cfm
Go
29 captures
15 Jun 2004 – 26 Jan 2021
MAY JUN NOV
2003 15 2004 2005
About th

Audio Engineering Society Preprints

AES 115th Convention

New York, New York, USA
October 10-13, 2003

AES Preprint O

Single Convention Preprints are available through the AES Preprint Search and Shop facility.

Preprints Listing

5854
Matti Hamalainen,
This paper discusses interoperable synthetic audio applications and content formats for mobile devices. It is important for these devices that technologies are compact, efficient and can be applied to many different types of applications. Some of these applications utilize network connectivity and have to be interoperable between different devices. These requirements introduce several technical challenges that are discussed in more detail, focusing on MIDI based technologies. A new interoperability solution is proposed for synthetic audio content for hybrid synthesizer architectures.
Interoperable Synthetic Audio Formats for Mobile Applications and Games

5855
Christian Grigg,

The Wayback Machine - https://web.archive.org/web/20040615041014/http://aes.org:80/publications/preprints/lists/115.cfm

Audio Engineering Society Preprints

AES 115th Convention

New York, New York, USA
October 10-13, 2003

AES Preprint Ordering

Single Convention Preprints are available through the AES Preprint Search and Shop facility.

Preprints Listing

5854
Matti Hamalainen,
This paper discusses interoperable synthetic audio applications and content formats for mobile devices. It is important for these devices that technologies are compact, efficient and can be applied to many different types of applications. Some of these applications utilize network connectivity and have to be interoperable between different devices. These requirements introduce several technical challenges that are discussed in more detail, focusing on MIDI based technologies. A new interoperability solution is proposed for synthetic audio content for hybrid synthesizer architectures.
Interoperable Synthetic Audio Formats for Mobile Applications and Games

5855
Christian Grigg,
A working group in the Interactive Audio SIG of the MIDI Manufacturers Association has produced a draft specification for a public standard file format supporting the interchange of advanced interactive audio soundtracks. It uses a cue-oriented model, is not tied to any particular authoring or playback platform, is programming languageneutral, can be used without license agreements or royalty payments. It is technically extensible in several dimensions. A model for the underlying soundtrack engine is articulated, as is a plan for an open-source software project to speed implementation on any given platform by leveraging its existing media playback APIs.
Preview: Interactive XMF - A Standardized Interchange File Format for Advanced Interactive Audio Content

5856
Rafael Kassier,Slawomir K. Zielinski,Francis Rumsey,
The effect of division of attention between the evaluation of multichannel audio quality degradations and involvement in a visual task (playing a computer game) was investigated. Time-variant impairments (drop-outs) were used to provide degradations in audio quality. It was observed that involvement in a visual task may significantly change the results obtained during the evaluation of audio impairments for some experimental conditions.
Computer Games and Multichannel Audio Quality Part 2 – Evaluation of Time-Variant Audio Degradations Under Divided and Undivided Attention

5857
Brian E. Schmidt,
As interactive audio soundtracks mature, they become more and more complex. It's not uncommon for games to support tens of thousands of lines of dialog, hundreds of music cues, dozens of ambiences and thousands of individual sound effects. The mixing of all these audio elements presents a unique challenge. Unlike linear media, 'mixing' of interactive audio happens as they game is being played, not ahead of time. Therefore, existing traditional post-productions techniques do not necessarily apply. This paper will discuss some of the unique challenges associated with mixing interactive audio content, including trying to determine what exactly is meant by 'mixing' game audio in the first place.
Interactive Mixing of Game Audio

5858

Google Exhibit No. 1063
Page 1395

Jae-Jin Jeon,Lae-Hoon Kim,Koeng-Mo Sung,
Multiple Point Equalization scheme based on least square(LS) method for wavelet-filtered signal is proposed. As the variations of the room transfer function (RTF) are different at different frequency bins depending on wavelength of frequencies, equalization with different frequency resolution is desirable. Decomposing received signal with discrete wavelet transform, we can assign different kinds of filters to each bandpass signal. Moreover, RTF measurements at various receiver positions should be utilized to make the inverse filter insensitive to source/receiver position changes. These two methods are well combined to guarantee wider sweet region in a listening room. Real measurement data are used to construct an inverse filter.
Wavelet Based Multiple Point Equalization of Room Transfer Function

5859
Kittiphong Meesawat,Dorte Hammershoi,
This study aims to investigate, whether the reverberation tails of binaural room impulse responses (BRIRs) for different locations and directions in a given room can be arbitrarily interchanged in e.g. virtual environment generation. BRIRs were measured in a quardratic lecture room and postprocessed to minimize any possible noise and concatenation artifacts. Four subsets of combinations of BRIR heads and tails were selected to test for dependence of location and head direction. A 3 AFC test has been carried out with–so far–4 listeners, and the results suggest that–for the room in question–the BRIRs may be cut around 40-60 ms and arbitrarily combined with no or little perceptual consequence.
The Time When the Reverberation Tail in a Binaural Room Impulse Response Begins

5860
Joel Preto Paulo,Carlos Rodrigues Martins,Jose L. Bento Coelho,
The measurement of the Room Impulse Response is often evaluated in the presence of non-stationary noise showing a rms value and a power spectral density that significantly varies with time. Under these conditions, the mean square value, MS, of the sequence must be minimized to improve the overall SNR. This suggests that the analysis should be performed by considering the energy of the noise in the time domain and in the frequency domain. A modified MLS measurement method working in the time and in the frequency domain for applications in the room acoustics field is presented. Experimental results obtained in real conditions are described and shown in the paper. The new approach, the Hybrid Sequences technique proved to lead to a significant increase of the SNR, when compared with the classical MLS technique.
Hybrid M Sequences for Room Impulse Response Estimation

5861
Hideo Miyazaki,Takayuki Watanabe,Shinji Kishinaga,Fukushi Kawakami,
Technology for controlling sound field by electro-acoustic means is often called Active Field Control, AFC, which is used to improve auditory impressions such as liveness, loudness, and spaciousness in auditoria. The AFC system, which has been developed at Yamaha, utilizes feedback control techniques to re-create natural reverberation based on the existing acoustics of the room. Time varying control, including EMR (Electric Microphone Rotator) and fluc-FIR (fluctuating FIR), is implemented in the AFC system to improve stability, preventing the coloration caused by a feedback loop in the system. In this paper, these technologies are summarized, together with an introduction to the recent representative venues using AFC. A system plan using core devices named AFC1, which has been developed at Yamaha and released recently in the US, is also presented.
Active Field Control (AFC) -Reverberation Enhancement System Using Acoustical Feedback Control

5862
Bradford N. Gover,James G. Ryan,Michael R. Stinson,
Spherical microphone array designs were investigated from the point of view of suitability for directional analysis of reverberant sound fields. Four array geometries (tetrahedron, cube, dodecahedron, geodesic sphere) were considered. Beamforming filters were designed using a constrained gain maximization process. The theoretical performance of each array was then predicted. A room acoustic simulator was used to help assess sufficient directionality and evaluate the suitability of each design. A 32-element geodesic sphere array was constructed and used to make directional measurements in real sound fields.
Designing a Spherical Microphone Array for the Directional Analysis of Reflections and Reverberation

5863
D. B. (Don) Jr. Keele,
To maintain constant beamwidth behavior, CBT circular-arc loudspeaker line arrays require that the individual transducer drive levels be set according to a continuous Legendre shading function. This shading gradually tapers the drive levels from maximum at the center of the array to zero at the outside edges of the array. This paper considers approximations to the Legendre shading that both discretize the levels and truncate the extent of the shading so that practical CBT arrays can be implemented. It was determined by simulation that a 3-dB stepped approximation to the shading maintained out to –12 dB did not significantly alter the excellent vertical pattern control of the CBT line array. Very encouraging experimental measurements were exhibited by a pair

Google Exhibit No. 1063
Page 1396

of passively-shaded prototype CBT arrays using miniature wide-band transducers.
Practical Implementation of Constant Beamwidth Transducer (CBT) Loudspeaker Circular-Arc Line Arrays

5864
Wolfgang Hess,Jonas Braasch,Jens Blauert,
The output of computational models of human auditory localization (e.g., [Lindemann, 1986], [Gaik, 1993]) is often given in the form of a 3-dimensional plot which takes into account the effect of binaural arrival-time and binaural level differences, the so-called binaural activity pattern. This pattern, which is a function of time, can be used for the detection, identification and separation of incoherent sound sources, the determination of their azimuths, the detection of echoes and the estimation of the amount of auditory spaciousness, among other things. This paper investigates how the binaural activity patterns of head-related room impulse responses can be used for judgements on the auditory quality of different virtual rooms. To this end it is aimed at presenting the binaural activity patterns in a visual form adequate to allow experts a general overview of the perceived room-acoustics as well as to characterize the attributes of the auditory events which correlate with lateral reflections, reverberation and the distribution of energy in time and space.
Acoustical Evaluation of Virtual Rooms by Means of Binaural Activity Patterns

5865
Brian E. Anderson,Timothy W. Leishman,
Because loudspeaker drivers are electro-mechano-acoustical transducers, their parameters may be measured from physical domains other than the electrical domain. A method has been developed by the authors to determine moving-coil loudspeaker parameters through the use of acoustical measurements. The technique utilizes a plane wave tube and the two-microphone transfer function technique to measure acoustical properties of a baffled driver under test (DUT). Quantities such as the reflection and transmission coefficients of the DUT are first measured. Driver parameters are then extracted from the measurements using curve-fitting techniques and theoretical solutions to equivalent circuits of the composite system. This paper discusses the acoustical measurement apparatus, system modeling, and a comparison of acoustically measured parameters to those measured using common electrical techniques. Parameters derived from the various methods are also compared to reference parameters to establish bias errors.
An Acoustical Measurement Method for the Derivation of Loudspeaker Parameters

5866
Karsten Nielsen,Lars Michael Fenger,
A novel audio power conversion system architecture is presented, in the attempt to provide a step forward in overall system efficiency and performance. The Active pulse modulated Transducer system ("Active Transducer") converts power directly from AC mains or from a DC power supply to the acoustic output in one simplified topological stage. New perspectives in audio system design emerge.
The Active Pulse Modulated Transducer (AT) - A Novel Audio Power Conversion System Architecture

5867
Juha Backman,
A loudspeaker arrangement, consisting of two individual drivers and delay units, having omnidirectional polar pattern at low frequencies and first-order gradient pattern at middle frequencies is presented. This system combines the low-frequency dynamic range of conventional speakers with the ability of gradient speakers to reduce roominduced coloration. The benefits and limitations of mid-frequency directivity are analyzed through simulations of room-speaker interaction for omnidirectional, cardioid, and dipole loudspeakers.
Low-frequency Polar Pattern Control for Improved In-room Response

5868
Ralf Geiger,Gerald Schuller,Juergen Herre,Ralph Sperschneider,Thomas Sporer,
This paper presents a scalable lossless enhancement of MPEG-4 Advanced Audio Coding (AAC). Scalability is achieved in the frequency domain using the Integer Modified Discrete Cosine Transform (IntMDCT), which is an integer approximation of the MDCT providing perfect reconstruction. With this transform, and only minor extension of the bitstream syntax, the MPEG-4 AAC Scalable codec can be extended to a lossless operation. The system provides bit-exact reconstruction of the input signal independent of the implementation accuracy of the AAC core coder. Furthermore, scalability in sampling rate and reconstruction word length is supported.
Scalable Perceptual and Lossless Audio Coding Based on MPEG-4 AAC

5869
Theg Hong Yeo,Wai Choon Wong,Dong Yan Huang,
In this paper, a concatenated system combining turbo product codes and convolutional codes with soft decision Viterbi decoding algorithm and diversity is proposed to enhance the robustness of Unequal Error Protection (UEP) scheme for wireless transmission of MPEG-4 Advanced Audio Coding (AAC). The proposed scheme

Google Exhibit No. 1063
Page 1397

has been tested over the Gaussian and Rician fading channels. Under severe channels with random, burst and mixed bit error rates (BER) of 6.00×10-2 and above, the proposed scheme provides a 90% improvement in residual BER performance, which is approximately 3 dB, with 19% increase in bandwidth over the original UEP scheme. At a high channel BER of 6.00×10-2, the proposed scheme gives an error-free header frame. Compared with the concatenated convolutional codes, the proposed scheme provides 0.5 dB BER performance improvements with the same bandwidth. With diversity, the performance can be further improved by 3 dB at low SNR for proposed scheme and 2.5 dB for original UEP scheme.
Robust MPEG Advanced Audio Coding over Wireless Channels

5870
Manfred lutzky,Martin Weishart,Johannes Hilpert,Bernhard Grill,Harald Gernhardt,Michael Haertl,
MP3 is worldwide the most commonly used audio coding scheme. Today it is present on virtually any computer, CD and DVD-Player, many car radios and of course in the many MP3 audio player devices. This paper describes a new MPEG-4 working draft, currently under development by ISO/MPEG, which will allow MP3 to be fully integrated into MPEG-4 Audio-Visual systems. Among other new features, like enhanced editability, MP3 is scheduled to get a full multi-channel audio capability, which can be implemented with a few additional lines of code on top of a standard stereo MP3 codec.
MP3 in MPEG-4

5871
Martin Wolters,Kristofer Kjorling,Daniel Homm,Heiko Purnhagen,
MPEG Spectral Band Replication (SBR) is the newest compression technology available as part of the MPEG standards. It is combined with MPEG Advanced Audio Coding (AAC) and improves coding efficiency by more than 30%. The resulting scheme is called High-Efficiency AAC (HE-AAC). This presentation will explain MPEG-SBR and its integration into the existing MPEG-4 bitstream format. The SBR technology itself as well as the implications on systems based on MPEG-4 technology are described. The signaling through MPEG-4 Systems and other transport formats is introduced and typical applications and usage scenarios will be listed.
A Closer Look into MPEG-4 High Efficiency AAC

5872
Tilman Liebchen,
Lossless coding will become the latest extension of the MPEG-4 audio standard. In response to a call for proposals, many companies have submitted lossless audio codecs for evaluation. The codec of the Technical University of Berlin was chosen as reference model for MPEG-4 Audio Lossless Coding, attaining working draft status in July 2003. The encoder is based on linear prediction, which enables high compression even with moderate complexity, while the corresponding decoder is straightforward. The paper describes the basic elements of the codec, points out envisaged applications, and gives an outline of the standardization process.
MPEG-4 Lossless Coding for High-Definition Audio

5873
Richard Foss,Jun-ichi Fujimori,Brad Klinkradt,Shaun Bangay,
A connection management server has been developed that enables connections to be made between mLAN-compatible audio devices, via a client web browser on any web-enabled device, such as a laptop or PDA. The connections can also be made across IEEE1394 bridges, and will allow for the transport of audio and music data between mLAN devices on the same or separate IEEE 1394 buses. Multiple users will be able to make and break connections via the server.
An mLAN Connection Management Server for Web-Based, Multi-User, Audio Device Patching

5874
Shigeru Aoki,Hirokazu Nakashima,
The Japan FM Network, 38 affiliate FM broadcasting companies all over Japan, installed the digital audio program file distribution network system to abolish conventional analog distribution. This system reduces the distribution cost and duration compared with the traditional technique of broadcasting relay where one station records another station's program off-air for later use. This paper also describes the format of digital audio file for current distribution and presents solutions to some existing problems.
The Audio File Format for Digital Distribution

5875
Masahiro Ikeda,Shinjiro Yamashita,Shinji Kishinaga,Fukushi Kawakami,
In terms of the sound systems installed in auditoriums, it is difficult to design comprehensive networking systems, because of limitations in available channels, latency and cost. However, with rapid progress of networking devices, Digital Audio Networks will become more valuable also in the field of performing arts after several years. This report will discusses required functions of network audio systems for auditoriums from the system-design viewpoint by picking a 2300-seat multipurpose hall as an example. Comparing the

Google Exhibit No. 1063
Page 1398

system against networking technologies that are suggested today, the report will also discuss the feasibility of the example and point out problems in realizing such systems.
Design method of Digital Audio Network System for Auditoriums

5876
Toshiyuki Nishiguchi,Masakazu Iwaki,Kimio Hamasaki,Akio Ando,
We conducted subjective evaluation tests to study perceptual discrimination between musical sounds with and without very high frequency components (above 21 kHz). In order to conduct strict evaluation tests, the sound reproduction system used for these tests was designed to exclude any leakage or influence of very high frequency components in the audible frequency range. Most of the sound stimuli used for the evaluation tests were newly recorded by the authors to maintain the highest quality for proper sound reproduction. The subjects were selected mainly from professional audio experts and musicians. The results showed that we can still neither confirm nor deny the possibility that some subjects could discriminate between musical sounds with and without very high frequency components.
Perceptual Discrimination between Musical Sounds with and without Very High Frequency Components

5877
Malcolm O. J. Hawksford,
Progress is reported in parametrically controlled noise shaping Sigma Delta Modulator (SDM) design. As this SDM structure (introduced by the author AES 112TH) can obtain a higher SNR than normal SDM structures, Philips Research Laboratories have questioned whether further improvement could be obtained using techniques inspired by the Trellis SDM. Simulations are used here to illustrate the performance of a parametrically controlled pseudo-Trellis SDM, which is believed to be the first disclosure of its type. The technique uses a variable state step back approach to moderate loop behaviour that is shown to achieve robust stability in the presence of aggressive noise shaping and high level signals. Comparisons are made with traditional SDM structures and LPCM systems.
Parametrically Controlled Noise Shaping in Variable State-Step-Back Pseudo-Trellis SDM

5878
Michael Page,Peter Eastty,Gary Cook,Mike Smith,Eamon Hughes,
"Super Multi-channel Audio Connection" (SuperMAC) is an enhancement of the existing "Multi-channel Audio Connection for Direct Stream Digital" (MAC-DSD) to support PCM audio formats, extending the unique benefits of the technology to a universal range of studio audio applications. The link provides full-duplex multi-channel audio connection for Direct Stream Digital (DSD) or 24-bit PCM, at sample rates up to 384kHz, plus high-quality clock signals and auxiliary data. It features error correction and deterministic latency as low as 45 microseconds, and the connection medium is a standard structured wiring cable. The specification is being submitted to the AES Standards Committee with a view to open standardisation.
A Universal Interface on Cat-5 Cable for High-resolution Multi-channel Audio Interconnection

5879
Jon D. Paulo,
High-resolution digital audio systems are susceptible to various sources of electromagnetic noise from the environment, especially crosstalk from adjacent cables. The noise can induce errors and increase jitter in the recovered clock signal. We discuss the most important noise sources and their characteristics. Next, we analyze the noise susceptibility of typical transmitter and receiver circuits. Test results are provided for a system with induced common mode noise. The paper concludes with design and application considerations.
The Effects and Reduction of Common-Mode Noise and Electromagnetic Interference in High-Resolution Digital Audio Transmission Systems

5880
Jan Abildgaard Pedersen,
This paper presents a system for adapting a loudspeaker to its position and to the acoustic properties of the listening room: the ABC room adaptation system. Adaptive Bass Control (ABC) measures the acoustic radiation resistance seen by the bass drive unit and calculates a digital filter, which is inserted in the signal path before the power amplifier. The radiation resistance is calculated from measurements of sound pressures at two different positions in front of the bass drive unit. The measured radiation resistance is compared to sound pressure measurements at different listening positions. The ABC system has been found to provide a room adaptation, which is globally valid throughout the listening room, i.e. all listening positions benefit from this system.
Adjusting a Loudspeaker to Its Acoustic Environment - The ABC System

5881
Scott G. Dorsey,
Incandescent lamps have been used for over fifty years as loudspeaker protection devices, but a large amount of misinformation about them exists. The author measures static and dynamic parameters of over thirty types

Google Exhibit No. 1063
Page 1399

of auto lamps, as well as tests some types for consistency between manufacturers and production. The results contradict a lot of the common wisdom about using lamps for protection and show serious linearity problems even at low operating levels.
Lamps for Speaker Protection

5882
Steven L. Garrett,John F. Heake,
Penn State recently instituted a First Year Seminar (FYS) requirement for every student. This paper describes a "hands-on" FYS on audio engineering that has freshmen construct and test a two-way loudspeaker system during eight 2-hour class meetings. Time and resource constraints dictated that the speaker system must be assembled using only hand tools and characterized using only an oscillator and digital multimeter. The cost of the entire system could not exceed $60/side. The speaker enclosure is made primarily from PVC plumbing parts. The four laboratory exercises that the students perform and write up are designed to introduce basic engineering concepts including graphing, least-squares fitting and error estimation, electrical impedance, resonance, transfer functions, mechanical and gas stiffness, and non-destructive parameter measurement.
Hey Kid! Wanna Build a Loudspeaker? The First One's Free

5883
Steven L. Temme,Pascal Brunet,Evan Chakroff,
During the loudspeaker manufacturing process, particles may become trapped inside the loudspeaker, resulting in a distinctive defect that is easily heard but difficult to measure. To give a clearer view of the problem, Time-Frequency maps are shown for some defective loudspeakers. Based on this analysis, a reliable testing procedure using a swept-sine stimulus, high-pass filter, and RMS-envelope analysis is presented. Further possible enhancements and applications of the method are listed.
Loose Particle Detection in Loudspeakers

5884
Elena Prokofieva,
Theoretical analysis of the problem of an elastic rectangular vibrating panel placed into a baffle in the vicinity of a porous layer has been conducted. Numerical results were obtained from a special computer program written in Matlab 6.0. It has been found that the presence of the porous layer considerably alters sound emission by the panel. The effect of the porous layer characteristics as well as the air gap width between the vibrating panel and the porous layer on the acoustic pressure and surface velocity was investigated.
Radiation of Sound by a Baffled DML-Panel Near a Porous Layer

5885
Justin Baird,David McGrath,
Conventional crossover design methods utilize traditional frequency selective networks to combine multiple transducers into a single full-bandwidth system. These traditional networks, whether they are implemented in analogue or digital form, exhibit large transition bands and suffer from phase distortion. These characteristics result in poor frequency, impulse and polar responses. A practical crossover implementation is presented that removes the detrimental effects of transition bands and phase distortion. This method implements linear phase crossovers whose transition bands approach a theoretical ideal brick wall response. Comparisons to conventional crossovers will be presented. Applications to large scale array optimization are also discussed and presented.
Practical Application of Linear Phase Crossovers with Transition Bands Approaching a Brick Wall Response for Optimal Loudspeaker Frequency, Impulse and Polar Response

5886
Markus Dodd,
A compression driver with an annular two-slot phase-plug coupled to the convex side of a hemispherical diaphragm is introduced. Magnetic and thermal domains are modelled using transient and static Finite Element Methods (FEM). Structural and acoustic domains are modelled as finite elements with boundary elements used to model free space. Structural and acoustic elements are fully coupled to both each other and the boundary elements. The application of these FEM techniques to the optimisation of compression driver performance is discussed and illustrated with results. The limitations of plane-wave tube measurements are also mentioned and illustrated with FEM and measured results.
The Development of a Forward Radiating Compression Driver by the Application of Acoustic, Magnetic and Thermal Finite Element Methods

5887
D. B. (Don) Jr. Keele,
Recently Vanderkooy et al. [1, 2] considered the effect on amplifier loading of dramatically increasing the Bl force factor of a loudspeaker driver mounted in a sealed-box enclosure. They concluded that high Bl was a decided advantage in raising the overall efficiency of the amplifier-speaker combination particularly when a

Google Exhibit No. 1063
Page 1400

class-D switching-mode amplifier was used. When the Bl factor of a driver is raised dramatically, the input impedance magnitude also rises dramatically while the impedance phase essentially approaches a purely reactive condition of ±90° over a wide bandwidth centered at resonance. This is an optimum load for a class-D amplifier, they note, which not only can supply power, but can also efficiently absorb, store, and return power to the speaker. Unfortunately, the system designed with a high-Bl driver requires significant low-frequency equalization and increased voltage swing from the amplifier as compared to systems using typical much-lower values of Bl. This paper considers the effect on the driver's efficiency of raising the driver's Bl factor through a series of Spice simulations. The nominal power transfer efficiency defined in traditional loudspeaker design methods is contrasted with true efficiency, i.e. true acoustic power output divided by true electrical power input. Increasing Bl dramatically increases the driver's true efficiency at all frequencies but radically decreases nominal power efficiency in the bass range. Traditional design methods based on nominal power transfer efficiency disguise the very-beneficial effects of dramatically raising the driver's Bl product.
Comparison of Direct-Radiator Loudspeaker System Nominal Power Efficiency vs. True Efficiency with High-Bl Drivers

5888
John J. Jr. Neumann,
MEMS (microelectromechanical systems) technology is more than a scientific curiosity. Commercial MEMS products are being produced using semiconductor manufacturing techniques. What kind of audio devices can be made using this technology? Surveillance, hearing aids, and directional microphones spring to mind. Less obvious are ultrasonics, in-ear translators and surround-sound wallpaper. The small size of MEMS devices brings up issues of physical limits and appropriate size scales for acoustic applications. MEMS microphone/ speaker design involves many of the same issues as conventional microphones/speakers, but the scale difference changes their relative importance. Over the past four years, the MEMS lab at Carnegie Mellon University has developed both microphones and speakers using CMOS-MEMS micromachining, and the technology is being commercialized by Pittsburgh startup Akustica.
MEMS (Microelectromechanical Systems) Audio Devices - Dreams and Realities

5889
Gary W. Elko,Flavio Pardo,Daniel Lopez,David Bishop,Peter Gammel,
The term MEMS is an acronym for MicroElectroMechanical Systems. During the past decade numerous novel sensor devices based on MEMS technologies have been made: accelerometers for air-bag deployment detection, MEMS mirror arrays for Digital Light Processing (DLP) projectors, and inkjet printer heads to name a few well known devices. MEMS devices have been developed for data storage, wireless communication, displays, optical switching, as well as microfluidics, aerospace and biomedical applications. The application of MEMS technology to audio has been primarily focussed on microphones. There are two major application areas that are driving the interest in MEMS microphones: hearing aids where size and integration with signal processing are important, and consumer devices where there is interest in reducing costs by integrating a complete systems solution on an integrated circuit and packaging of devices to allow standard robotic pick-and-place manufacturing. This paper describes a MEMS microphone that was built at Bell Labs which was the first all-surface machined MEMS microphone. We also describe some fundamental issues in the design of MEMS microphones.
Surface-Micromachined MEMS Microphone

5890
Earl R. Geddes,Lidia W. Lee,
Historically, distortion has been measured using specific signals that were sent through a system and quantified by the degree to which the signal is modified by the system. The human auditory system has not been taken into account in these metrics. Combining nonlinear system theory with the theory of hearing, a new paradigm for quantifying distortion is proposed.
Auditory Perception of Nonlinear Distortion - Theory

5891
Earl R. Geddes,Lidia W. Lee,
A new metric to the perception of distortion was recently proposed by Geddes and Lee (2003). Psychoacoustical data were measured, correlation and regression analysis were applied to examine the relationship and predictive value of this new metric to the subjective assessment of sound quality of nonlinear distortion. Furthermore, conventional metrics such as total harmonic distortion (THD) and intermodulation distortion (IMD) were also compared. Thirty-four listeners participated in a listening task, rating twenty-one stimuli using a 7-point scale. No significant relationships were observed when comparing the subjective ratings with TDH and IMD metrics. Significant correlation ($r=0.95$, $p<.001$) was observed between the subjective ratings and the new proposed GedLee (Gm) metric. Furthermore, robust predictive power was verified utilizing the GedLee metric. GedLee metric has demonstrated remarkable potential to quantify sound quality ratings of nonlinear distortion.
Auditory Perception of Nonlinear Distortion

Google Exhibit No. 1063
Page 1401

5892
Scott G. Norcross,Gilbert A. Soulodre,Michel C. Lavoie,
In many applications it is desirable to measure and control the subjective loudness of audio signals. However, there is a lack of data regarding loudness perception for typical program material and an ITU-R study is underway to examine this matter. In the present study, a series of formal subjective tests were conducted to evaluate the perceived loudness of a broad variety of typical program materials, including music and speech. Subjects adjusted the level of various audio materials until their perceived loudness was equal to that of a reference signal. Tests were designed to examine the relative subjective loudness for a variety of playback conditions. The just noticeable differences (JND) in perceived loudness was also examined.
The Subjective Loudness of Typical Program Material

5893
Kalle Koivuniemi,Nick Zacharov,
This paper describes the design of a Calibrated Source (CalSo) loudspeaker created for audio prototyping of a mobile phone. CalSo is used in conjunction with a PC application, which enables the use of virtual audio prototypes for the enhancement of audio content creation of mobile phones. CalSo provides means to calibrate the output of a PC with the loudspeaker itself or with a pair of headphones. CalSo was primarily developed to ensure that the auralized audio output of a virtual audio prototype from the users PC accurately reproduces the audio output of the real device.
A Calibrated Source for Virtual Audio Prototyping

5894
David Isherwood,Gaetan Lorho,Nick Zacharov,Ville-Veikko Mattila,
The generalized listener selection procedure (GLS) defines criteria for the efficient creation of a permanent listening panel. This paper describes the application of this procedure to a large group of candidate listeners (>300) from which a permanent panel of 30 listeners was to be formed. The criteria presented in the original publication are augmented to make the procedure more rapid and sensitive to acuity of specific auditory percepts. Verification of the benefits of such a procedure based on a comparison of listening test results for the final GLS panel and a panel of randomly chosen listeners is presented.
Augmentation, Application and Verification of the Generalized Listener Selection Procedure

5895
William L. Martens,Charith N. W. Giragama,Susantha Herath,Dishna R. Wanasinghe,Alam M. Sabbir,
A single, common timbre space for a small set of processed guitar sounds was derived for four groups of listeners, each group comprising respectively native speakers of English, Japanese, Bengali, a language of Bangladesh, and Sinhala, a language of Sri Lanka. Members of these four groups also made ratings on 10 bipolar adjective scales for the same set of sounds, each of the four groups using anchoring adjectives taken from their native language. Whereas the two primary dimensions underlying perception of the guitar timbres was common between the four groups, the way in which directly translated adjectives were used to describe the sounds generally differed between the groups, those differences being quantified via principal components analysis. Nonetheless, the two most closely related languages, Bengali and Sinhala (both Indo-Aryan languages), showed much more semantic similarity to each other than did the Japanese language with any of the the three other languages examined.
Relating Multilingual Semantic Scales to a Common Timbre Space - Part II

5896
Scott G. Norcross,Gilbert A. Soulodre,
There are numerous applications where it is desirable to have an objective measure of the perceived loudness of typical audio signals. For example, in broadcast applications an objective measure would allow the perceived loudness of the various program materials to be equalized. In the present study several potential objective loudness measures are examined. The objective measures are evaluated in their ability to predict the results of a database derived from a series of formal subjective tests. Possible metrics for rating the performance of the objective loudness measures are considered.
Objective Measures of Loudness

5897
Jim Brown,
It has been shown that a primary cause of VHF and UHF interference to professional condenser microphones is inadequate termination within the microphone of the shield of the microphone's output wiring, a fault commonly known as the pin 1 problem. Tests using only audio frequency test signals generally fail to expose susceptibility to radio frequency (RF) interference. Simple RF tests for pin 1 problems in microphones and other audio equipment are described that correlate well with EMI observed in the field.
Testing for Radio-Frequency Common Impedance Coupling (the "Pin 1 Problem") in Microphones and Other Audio Equipment

Google Exhibit No. 1063
Page 1402

5898
Jim Brown,
It has been shown [1, 2] that radio frequency (RF) current flowing on the shield of balanced audio wiring will be converted to a differential signal on the balanced pair by a cable-related mechanism commonly known as Shield-Current-Induced Noise. This paper investigates the susceptibility of audio input and output circuits to differential signals in the 200 kHz - 2 MHz range, with some work extending to 300 MHz. Simple laboratory test methods are described, equipment is tested, and results are presented. Laboratory data are correlated with EMI observed in the field.
A Novel Method of Testing for Susceptibility of Audio Equipment to Interference from Medium and High Frequency Radio Transmitters

5899
Paul D. Henderson,
To fully describe the sound field at a listening position in an acoustical environment, the distribution of energy arriving from varying spatial directions is required, which this research accomplishes through the use of microphone array beamforming. Using the proposed technique, a multi-microphone impulse response measurement may be used to directionally locate any direct, reflected, or scattered energy arriving at the measurement location. This data may be graphically projected onto a sphere surrounding the measurement location or onto a virtual model of the measured environment, revealing the origin of the arriving energy. Preliminary measurements conducted at a prominent concert hall are presented, illustrating the analysis capability of the technique.
Directional Room Acoustics Measurement Using Large-Scale Microphone Arrays

5900
Charles Robinson,Steve Lyman,Jeffrey C. Riedmiller,
The broadcast, satellite and cable television industries have been plagued for years by the inability of personnel to accurately interpret and thus consistently control program loudness utilizing traditional measurement devices and methods. As a result, most listeners feel compelled to make adjustments to their television volume controls (in the home). A recent survey of channel-to-channel and/or program-to-program level discrepancies and subjective listening tests confirms that current the practice is unacceptable to listeners. In this paper we describe loudness measurement techniques that improve accuracy, usability, and consistency relative to previous techniques. Accuracy in this application is determined by correlation to listener opinion, with the particular goal of minimizing annoyance resulting from level mismatch. Usability is improved by minimizing the interaction required by the user. Consistency is achieved by minimizing the amount of meter interpretation required. The keys to this method are: providing a single numeric indication of loudness for a given program or segment; and isolating and measuring the portions of the program that are primarily speech, and using speech loudness as the basis for overall program level thereby improving listener satisfaction.
Intelligent Program Loudness Measurement and Control: What Satisfies Listeners?

5901
John Vanderkooy,Robert D. Stevens,
A method is presented, for measuring acoustical properties of materials, which the authors believe to be novel and unique. The method relies on an MLS (maximum length sequence) excitation signal to measure the acoustical impedance of a specimen of material placed at the termination of a long tube. Whereas the traditional methods require that measurements be made at multiple locations within the tube, either using a multi-channel data acquisition system, or by physically moving a single microphone from one location to the next, the novel method requires only a single measurement at one location in the tube, using a single microphone. The necessity to conduct only a single measurement makes this method two to four times faster than traditional methods, depending on the desired frequency range of the measurement. Other benefits of this method include the fact that it requires only a single channel data acquisition system and single microphone, and that it has the unique ability to measure the impedance of the source end of the tube (i.e., the loudspeaker) as well as the material specimen at the termination end of the tube.
A Novel Single-Microphone Method of Measuring Acoustical Impedance in a Tube

5902
Gyorgy Wersenyi,
Listening tests were carried out for investigating the localization judgments of 40 untrained subjects through equalized headphones and with HRTF synthesis. The investigation was made on the basis of the former GUIB (Graphical User Interface for Blind Persons) project in order to determine the possibilities of a 2D virtual sound screen and headphone playback. Results will be presented about the capabilities and values of typical headphone playback errors as well as minimum, maximum and average values of discrimination skills. Special localization events like left-right and up-down symmetries, missing locations in vertical localization are also discussed. The measurement method includes a special 3-categorie-forced-choice MAA report on a screen-like virtual auditory surface in front of the listeners. Test signals were presented with different spectra and movement. Conclusions are drawn both for a GUIB application as well as for the binaural synthesis about the

Google Exhibit No. 1063
Page 1403

role of the fine structure of applied HRTFs.
Localization in a HRTF-based Minimum Audible Angle Listening Test on a 2D Sound Screen for GUIB Applications

5903
Arvindh Krishnaswamy,
We discuss various tuning possibilities for the twelve basic musical intervals used in South Indian classical (Carnatic) music. Theoretical values proposed in certain well-known tuning systems are examined. Issues related to the intonation or tuning of the notes in Carnatic music are raised and discussed as well.
On the Twelve Basic Intervals in South Indian Classical Music

5904
Sylvain Choisel,Karin Zimmer,
A new laser-pointing technique providing visual feedback is presented and compared to a more traditional method, making a mark on a line; broadband noise, speech and a musical instrument served as sound stimuli. In localizing frontal sources (±30 degrees), both "real", and amplitude-panned, in a standard listening room, the new method is shown to be more intuitive, and precise, allowing for a higher consistency of responses both within and across subjects. Furthermore, the lateral displacement of the sources is overestimated in both response techniques, this inaccuracy being significantly smaller when the laser pointer is used. In a second experiment, the influence of head orientation on pointing performance towards sounds varying in frequency content is investigated. In result, responses are not affected by moving the head towards a physical sound source, but are highly sensitive to head movements when sources are panned.
A pointing Technique with Visual Feedback for Sound Source Localization Experiments

5905
Bruno M. Fazenda,William J. Davies,Mark R. Avis,
A subjective test study was carried out in order to identify the perceptibility of changes in the Q-factor of room modes. The experimental technique concentrates on the identification of difference limen for three levels of Q-factor referring to modes in rooms used for critical listening. Trends show that changes in higher Q values are more perceptible than those for lower Q values. The results may be applied in decisions for treatment of modes in common listening and control rooms.
Difference Limen for the Q factor of Room Modes

5906
Han-gil Moon,Jungmin Park,Koeng-mo Sung,Dae-young Jang,
To provide the distance information of the sound source in 3-D audio environment, we must have some information about effective distance cues and some methods to handle them properly. Through our experiments and researches, we found that C80 and EDT change systematically with source-receiver distances. We are required to validate this result physically and psychologically. So, this paper contains physical explanation about two parameters' systematic changes and psychological tests with artificially controlled curves, which imitate different distances. With these validations, we want to show the effect of early decay time on auditory depth.
The Effects of Early Decay Time on Auditory Depth in the Virtual Audio Environment

5907
Natanya Ford,Francis Rumsey,Tim Nind,
A Graphical Assessment Language (GAL) appears to provide the listener with a medium for describing the perceived spatial attributes of a reproduced audio event. Previous language development investigations have concluded that these spatial characteristics may be represented consistently by listeners using their own graphical descriptors. However, the ease with which these individual descriptors could be misinterpreted by a researcher was highlighted in a subsequent study; a notable problem since a primary aim of the GAL is to maintain the validity of the listener's original experience. To reduce potential ambiguities in interpretation, this investigation considers the development of a common descriptive language, consolidating listener's individual descriptors into a universal set of graphical terms identified as being effective for describing the experiences of all investigation participants. The process and outcome of creating a "universal" GAL is described.
Creating a Universal Graphical Assessment Language for Describing and Evaluating Spatial Attributes of Reproduced Audio Events

5908
Brandon Cochenour,David A. Rich,
Higher-order notched networks more consistently retain a desired all-pass response of loudspeakers. However, their non-coincident drivers cause deep spectral notches to occur intermittently in the crossover region. Large phase shifts are also introduced in the loudspeaker's transfer response. To evaluate the sonic impact of the deep spectral nulls and phase shifts to the overall listening experience, we propose a real-time listening test that does not involve the design of real loudspeakers or modification of a loudspeaker's sound in a listening

Google Exhibit No. 1063
Page 1404

environment. A speaker-system simulation program has been developed using Matlab. The program processes wavfiles of music clips with a virtual loudspeaker model that covers real crossover networks, offset delays, any compensation network, and raw driver frequency response characteristics. The ABX double-blind testing methodology is applied to determine the audibility of the virtual loudspeaker model under test. This approach isolates audible effects and makes them more apparent to the listener, since other effects are eliminated that might mask the changes caused by the features under study. We expect that the software can serve as a generalized template to examine other phenomena.
A Virtual Loudspeaker Model to Enable Real-Time Listening Tests in Examine the Audibility of High-Order Crossover Networks

5909
Andrew Bright,
This paper explains how, that if the total moving mass of a loudspeaker can be known in advance, all of the remaining basic linear loudspeaker parameters can be determined using only current feedback. This is explained at a theoretical and practical level. Frequency- and time-domain algorithms for tracking the parameters are presented. Examples of the tracking performance of an adaptive algorithm operating with a real music signal on an actual loudspeaker are shown.
Tracking Changes In Linear Loudspeaker Parameters with Current Feedback

5910
Rosalfonso Bortoni,Sidnei Noceti Filho,Rui Seara,
The Thiele-Small method for speaker design considers the linear loudspeaker model driven by voltage sources and operating in a small signal environment. Subsequent studies have been made to introduce into the model some nonlinear characteristics due to the operation with large signals. This paper presents a comparative analysis of the sound pressure level and cone displacement of loudspeaker systems as an infinite baffle, a closed box, a vented box and band-pass enclosure driven by voltage and current sources, under small and large signals. The nonlinearities of the voice-coil, force factor and compliance of the loudspeaker are taken into account.
Comparative Analysis of Moving-Coil Loudspeakers Driven by Voltage and Current Sources

5911
Rosalfonso Bortoni,Homero Sette Silva,
This work presents electric models for loudspeakers installed on baffles and enclosures, as closed box, bass-reflex, 4th and 6th orders band-pass enclosures, using passive crossovers from two way to three way, whose impedance curves were derived from MATLAB® simulations. The impedance curves, it's module and phase, are presented for each one of the cited models above. The transfer functions are also presented, besides the necessary considerations to get the results from the loudspeakers specifications, dimensions and box tuning. Examples of the efforts caused in the output stages of audio power amplifiers are presented and commented.
Loudspeakers' Electric Models for Study of the Efforts in Audio Power Amplifiers

5912
Alexander Voishvillo,
The compression driver has always been and remains an essential component of sound reinforcement systems despite its unavoidable nonlinear distortion. This distortion is caused by various effects, including adiabatic compression of air in the front chamber, modulation of the chamber's air stiffness, and modulation of the chamber's viscous losses. Each of these sources of distortion is inherent in the compression driver's operation, each adversely affects the compression driver's performance in its own specific way, and each is characterized by a different nonlinear "signature." Comparative analysis of these distortion sources is undertaken. Nonlinear and parametric effects are explicitly expressed. The influence of diaphragm displacement, compression ratio, and chamber sound pressure on the generation of intermodulation and harmonic distortion is explored. Some design recommendations are given.
Nonlinear Versus Parametric Effects in Compression Drivers

5913
Wolfgang Klippel,
A new technique for measuring nonlinear distortion in transducers is presented which considers a priori information from transducer modeling. Transducers are single input-multiple output systems (SIMO) where the dominant nonlinearities can be concentrated in a single source adding nonlinear distortion to the input signal. The equivalent input distortion at this source can easily be derived from the measured sound pressure signal by performing a filtering with the inverse transfer response prior to the spectral analysis. This technique reduces the influence of the acoustical environment (room), removes redundant information and simplifies the interpretation. It is also the basis for speeding up distortion measurements, for the prediction of distortion in the sound field and for the detection of noise and other disturbances not generated by the transducer.
Measurement of Equivalent Input Distortion

Google Exhibit No. 1063
Page 1405

5914
Jonathan S. Abel,David P. Berners,
Differential equations governing the behavior of first-order peak-detecting and RMS feedback and feedforward analog compressors are presented. Based on these equations, the relationship between feedback and feedforward compressor behavior is explored, and simple, accurate digital emulations are provided. Feedback and feedforward gain reduction trajectories are shown to be equivalent by transforming the feedback gain reduction into a feedforward gain reduction having a level-dependent time constant. This time constant has the effect of slowing down the transition into and out of compression, and accounts for much of the difference in compression character between the two architectures.
On Peak-Detecting and RMS Feedback and Feedforward Compressors

5915
Stephen H. Lampen,
Digital audio requires cables made to a specific impedance, 110 ohms for twisted pairs and 75 ohms for coaxial cable. But what happens when cables are the wrong impedance or are damaged or otherwise have their impedance altered? Changes in impedance can affect the signal traveling down a cable and makes a portion of the signal reflect back to the source, called 'return loss'. This paper will show how, and when, return loss can occur, how it is measured, and how it affects digital audio systems. A return loss specification is suggested as a possible addition to the AES specifications for both equipment and cable.
Return Loss and Digital Audio

5916
Knud Bank Christensen,
An efficient implementation of parametric equalizers is a cascade of biquadratic filter sections. Traditionally, a section operates in one of 3 operating modes offering Low Shelf, Bell Shaped or High Shelf families of responses. In either mode, each section offers 3 user parameters: Boost/cut gain G, corner frequency fc and bandwidth/slope Q. This paper describes the derivation and implementation of a new 4th user parameter, Symmetry, producing a smooth transition between the 3 above mentioned operating modes while retaining the meaning of the 3 conventional user parameters. All degrees of freedom in the bi-quadratic filter blocks are utilized. Therefore an inverse mapping exists, converting arbitrary (e.g. computer generated) biquadratic coefficient sets back into meaningful EQ parameters allowing further human interpretation and adjustment.
A Generalization of the Biquadratic Parametric Equalizer

5917
Suthikshn Kumar,
This paper reviews various schemes for Smart Acoustic Volume Controllers(SVCs) for consumer electronics such as Televisions, Car Stereo systems, CD Players, AM/FM Radio, Telephones, Mobile Phones etc. There are several instances of applications of smart volume controllers in consumer electronic devices. In car stereo systems, the volume level automatically dips when the system detects the telephone ringing. The mobile with smart volume controller provides an improved speech quality even in the presence of high background noises levels. The smart volume controller for television will even out any variations in volume levels from one channel to another and program to program, thus delivering an improvement in perceived quality to the consumer. The smart volume controllers can be personalized to suit the individual's hearing requirements. This paper also reviews several soft computing techniques for smart volume control such as fuzzy logic control, neural control, neuro-fuzzy control etc.
A Review of Smart Acoustic Volume Controllers for Consumer Electronics

5918
Sin-lyul Lee,Lae-Hoon Kim,Koeng-Mo Sung,
The non-individualized head related transfer function (HRTF) is known to have a few problems, which are referred to the 'hole in the middle' phenomenon and 'front-back reversals'. To overcome these problems, a HRTF refinement technique was introduced, but unfortunately, this refinement technique causes sudden degradation in sound quality to occur and difficulty in cross-talk cancellation because of notch frequency exaggeration. In this paper, an HRTF refinement using directional weighting function has been proposed to solve these problems. This newly proposed technique weights ordinary HRTF according to its direction to amplify frontal sound intensity. As a result, spectral differences in the 'cone-of-confusion' region become more pronounced within overall audible frequencies without having to exaggerate notch frequency. Also, by using this function, cross-talk cancellation filter can be made more easily. We have verified through listening tests that the proposed technique is superior to the previous HRTF refinement in terms of both sound localization and sound quality. Therefore, the refinement of HRTF through the use of the directional weighting function can be applied in the virtual reality, 3D entertainment and auralization program to require high quality sound.
Head Related Transfer Function Refinement Using Directional Weighting Function

5919
Bruce Duewer,Heling Yi,John Melanson,

Google Exhibit No. 1063
Page 1406

A robust new architecture for multi-bit delta-sigma data converters is presented. The second order mismatch shaping function is moved inside the feedback loop of a high order modulator, replacing the need for dynamic element matching (DEM) after the modulator. The mismatch shaper makes a trade-off between ensemble quantization error and mismatch induced error. Mismatch error in the frequencies of interest is decreased, and the resulting additional quantization error is dealt with by the delta-sigma feedback loop. This approach allows good non-tonal noise shaping performance even in the face of severe element mismatch. The modulator reliably collapses to second order to maintain stability if faced with particularly high noise energy. The modulator also has integrated SACD processing.
A Multi-bit Delta-Sigma DAC with Mismatch Shaping in the Feedback Loop

5920
Robert Peruzzi,Marvin White,David A. Rich,John A. Nestor,Erik Geissenhainer,Matthew Johnston,
A low-power audio amplifier with pulse width modulated power supply rails that track the output signal is presented. Because of the tracking power supply rails, the voltage drop over the power transistors is kept as low as possible and nearly constant, so that power efficiency remains high for low as well as high output level signals. A very simple digital input pulse width modulation scheme provides four power rails to a fully differential class-AB power amplifier. The simplicity of the circuit makes it an attractive solution for low cost portable audio applications, instead of using a more complex pulse width modulated class-D audio amplifier. An efficiency increase of about 10% has been simulated over the same class-AB output stage using fixed DC rails of 3 Volts and 0 Volts, with very little sacrifice in THD. Also presented are results from a 12-Volt, single-ended hardware prototype of the system.
An Efficient Low-Power Audio Amplifier with Power Supply Rails Tracking the Output by Means of Pulse Width Modulation

5921
Ronald M. Aarts,Erik Larsen,Okke Ouweltjes,
Extending the bandwidth of an audio signal may be useful at the low or high end of the frequency spectrum, depending on the application. Also, the actual bandwidth extension algorithm may rely entirely on psychoacoustic effects or may create a physical extension of the signal spectrum. We have developed a common framework for all these problems, and from this framework derived algorithms that address diverse applications in audio signal processing for bandwidth extension. Specifically, we describe algorithms for bandwidth extension applied to enhancing reproduction of bandlimited signals (at the low or high end of the frequency spectrum), and for enhancing reproduction over small loudspeakers.
A Unified Approach to Low- and High-frequency Bandwidth Extension

5922
Alberto Bellini,
Thanks to their intrinsic high efficiency audio power switching amplifier are becoming wide spread in many applications where heating and costs are major concerns. Several topologies exist to reduce augmented distortion, and many of them rely on digital signal processors (DSPs). Nowadays the market offers several products with integrated peripherals for analog signal sampling and PWM generation. However the latter operation is still a time consuming task, often performed with a number of different approaches. The paper presents a method for an efficient computation of the PWM signal corresponding to the input audio signal, suitable to feed a switching output power stage. The presented method exploits DSP capabilities and it is oriented to DSPs that integrate PWM generators. Its peculiar characteristics is the possibility to perform N taps FIR operation on the audio signal together with the computation of a suitable PWM signal with N MAC operations per sample.
Embedded Digital Filters for PWM Generators

5923
Scott G. Norcross,Gilbert A. Soulodre,Michel C. Lavoie,
Previous work has shown that inverse filtering can degrade the subjective quality of audio signals in certain conditions. Minimum phase inversion and regularization applied separately have also been studied and can be effective in some cases, but neither technique has proven to be robust. In this paper, further methods involving various regularization methods applied to the full and minimum phase part of the impulse response (IR) are studied. Subjective tests were conducted in accordance with the MUSHRA method to evaluate the performance of the various inversion methods. The results of the subjective tests were also used to determine the effectiveness of the ITU-R PEAQ objective test model as a potential tool in the development and evaluation of inverse filtering techniques.
Further Investigations of Inverse Filtering

5924
Albertus C. den Brinker,Felip Riera-Palou,
Linear prediction (LP) has traditionally been used in speech coding. Recently, variants of LP have also shown to be appropriate for audio coding. In this paper we introduce a new prediction scheme, called pure linear

Google Exhibit No. 1063
Page 1407

prediction (PLP), which combines important features from previous approaches. We show that the modelling capability of the PLP cna be tuned in a psychoacoustically relevant way making it suitablbe for speech and audio coding. Moreover, under certain restrictions, this new scheme is directly realizable, stable and retains the whitening property of conventional linear prediction. The processing of the prediction coefficients to perfrom operations such as quantisation, interpolation and spectral broadening is also addressed. As an example of the application of the PLP, we describe its use in the context of the sinusoidal coder proposed by Philips which is being standardised in "MPEG-4 Extension 2".
Pure Linear Prediction

5925
Peter Kassakian,David A. Wessel,
A fundamental goal of sound synthesis is to reproduce, and to control, as many facets of the sound as possible. By numerically solving a carefully constructed optimization problem, we are able to design low-order filters for use with a dodecahedral loudspeaker array to synthesize low order spherical harmonics over specified frequency ranges. The method, a variant of least-squares, is general, allowing for the inclusion of side constraints, arbitrary array geometry, and incorporation of measured loudspeaker characteristics. We compare the predicted loudspeaker array performance with high-resolution measurements of the physical system.
Design of Low-Order Filters for Radiation Synthesis

5926
Sidnei Noceti Filho,Homero Sette Silva,Andre Luis Dalcastagne,
This paper describes the design of three filters used in the filtering of pink noise which are described in the new Brazilian standard proposed to replace the current NBR 10303. The NBR 10303 filter does not permit the correct test of subwoofers because its low cut-off frequency is too high and for this reason we changed its component values in order to obtain three new filters with lower low cut-off frequencies. The design of these filters was carried out through a numerical method based on a modification of the NBR 10303 filter frequency response magnitude. The final component values were specified according to commercial values.
A Numerical Method to Modify the NBR 10303 Filter Frequency Response

5927
Wolfgang Ahnert,Stefan Feistel,Waldemar Richert,Steven McManus,
The processing power today available on portable computer platforms is now so far advanced that it is no longer necessary to use dedicated digital signal processing platforms for the intensive analysis required in Time Delay Spectrometry. Moving the processing from a dedicated platform onto a standard personal computer becomes possible with the TDS technology realized in a measurement and post-processing software instead. It also opens the way for more information to be extracted from a measurement after it had been made. Care must still be taken in the choice of the data gathering system, as timing between input and output data samples is critical.
Time Delay Spectrometry Processing Using Standard Hardware Platforms

5928
James A. Angus,Tim Jackson,
The design and implementation of lossless audio signal processing using a Finite Field Transform is shown. In particular Complex Mersenne Transforms are developed. Finite field signal processing techniques are described. The effects of filter length and coefficient accuracy are also discussed. Finite field transform algorithms which would be suitable for lossless signal processing are presented. The paper concludes by presenting examples of lossless processing.
Lossless Signal Processing with Complex Mersenne Transforms

5929
Clemens Kuhn,Renato Pellegrini,Etienne Corteel,Dieter Leckschat,
The reproduction of sound using Wave Field Synthesis (WFS) provides larger possibilities in rendering sonic space compared to standard 5.1 set-ups (panorama, acoustic holography, envelopment…) [Berkhout, 1988], [Verheijen, 1997]. Different microphone set-ups have been developed for this reproduction system, multi-microphone set-ups as well as microphone arrays [Hulsebos et al., 2002] Using multi-microphone techniques, the aesthetic of the sonic image is limited with regard to the localization of sound sources, spectrum and blending, especially in so called “classical” music recordings. The sources appear rather focused and dominant, due to the position of the microphones capturing the sound in the near field. Reproduction as point sources and convolution with the room impulse responses lead to a correct room reproduction in theory, but in practice the spectrum of the instruments and the impression of spatial depth require improvement. Microphone arrays are suitable for impulse response measurements but not flexible enough for direct music recording. Therefore the authors propose an approach to miking and mixing of music recordings, combining WFS techniques with phantom sources from a main microphone. It adapts this stereophonic technique to the holographic properties of WFS. This approach has been evaluated in an interactive mixing and listening test session where a panel of sound engineers was invited to perform the mix of an orchestral recording. Several mixing tasks were specified

Google Exhibit No. 1063
Page 1408

(stable localization, blending, homogeneity, envelopment…). The results of this test permit analysis of the aesthetic advantages and also the limits of the proposed mixing approach.
An Approach to Miking and Mixing of Music Ensembles Using Wave Field Synthesis

5930
Magali Deschamps,Olivier Warusfel,Alexis Baskind,
Subjective listening tests dedicated to 5.1 multichannel were conducted using various recording and mixing configurations. Two ambience microphone arrays ("Hamasaki squares"), differing in size, were used to record the hall reverberation, in addition to direct sound microphones, providing a separation between direct and reverberant sound. Differences between the reverberation recording systems, in order to study its optimization regarding size, were evaluated using a set of spatial subjective attributes. Post-processing parameters (time delay between direct sound and reverberation, front/back distribution of reverberation) were investigated along similar attributes. Results underlined significant differences between the two "Hamasaki square" systems. The time delay parameter showed low influence on listener envelopment and apparent source width, whereas front/back distribution of reverberation showed a significant effect on these attributes.
Investigation of Interactions between Recording/Mixing Parameters and Spatial Subjective Attributes in the Frame of 5.1 Multichannel

5931
Wieslaw Woszczyk,
It has been advocated that high-resolution audio means ultra-wide frequency range and that, given the limited sensitivity of human hearing for high frequencies, little is gained from high-resolution perceptually. Not much laboratory evidence is found to counter this assertion because psychoacoustic research has restricted itself largely to studying the effects of frequency range within 20Hz-20kHz rather than outside of it. The paper reviews some of the available findings in this area and focuses on remarkable complexities of auditory signals by looking at precise distinctions auditory system has to extract when analyzing time/space attributes of auditory scenes. It is shown that high-resolution in temporal, spatial, spectral, and dynamic domains together determine the quality value of perceived music and sound, and that temporal resolution may be the most important domain perceptually.
Physical and Perceptual Considerations for High-resolution Audio

5932
Ingyu Chun,P. A. Nelson,
The performance of current virtual acoustic systems is highly sensitive to the geometry of individual ear at high frequencies. The objective of this paper is to study a virtual acoustic system which may not be sensitive to individual ear shape. The incident sound field around ear is reproduced by using a multi-channel headphone. The results of computer simulations show that the desired sound pressure at the eardrum can be successfully replicated in a virtual acoustic environment by using a multi-channel headphone.
Virtual Acoustic System with a Multichannel Headphone

5933
Malcolm O. J. Hawksford,Christoph P. A. Reller,
A preliminary study is presented that investigates processing of microphone array signals for multi-channel recording applications. Generalised, perceptually and acoustically based approaches are taken, where a spaced array of M microphones is mapped to an array of L loudspeakers. The principal objective is to establish transformation matrices that integrate both microphone and loudspeaker array geometry in order to reproduce a subjectively accurate illusion of the original soundfield. Techniques of acoustical vector synthesis and plane wave reconstruction are incorporated at low frequencies migrating to an approach based upon head-related transfer functions (HRTFs) at higher frequencies. Error surfaces based on the HRTF reconstruction error are used to assess perceptually relevant solutions. Simulation results presented in a 5-channel format are calculated for processed audio material with and without acoustical boundary reflections.
Perceptually Motivated Processing for Spatial Audio Microphone Arrays

5934
Robert E. (Robin) III Miller,
Objectives: Take the next step toward reproducing human hearing AND make better recordings in 5.1. In life, we hear sources we see – but also reflections and reverberation we don't see. Each sonic arrival is individually direction-stamped by our unique HRTF, including height, tonally colored by our pinna. Preserving 3D directionality is key to life-like hearing. A practical, scalable approach is presented (pat.pend.) - a way to "transform" 3D (full sphere) recordings for uncompromised 2D reproduction in stereo or 5.1/6.1 without any decoding. By adding a decoder and speakers, full 3D is losslessly "reconstituted" from 6-channel media. Experimental "tri-play" 6-channel "PerAmbio 3D/2D" recordings have been made and demonstrations presented (AES 114th Amsterdam 3/2003 & AES 24th Banff 6/2003) with praised results.
Scalable Tri-play Recording for Stereo, ITU 5.1/6.1 2D, and Periphonic 3D (with Height) Compatible Surround Sound Reproduction

Google Exhibit No. 1063
Page 1409

5935
Joshua D. Reiss,Mark Sandler,
In recent years there has been considerable debate over the suitability of 1-bit Sigma-Delta modulation (SDM) for high-quality applications. Much of the debate has centered on whether it is possible to properly dither such a system. It has been shown that dither with a triangular probability distribution should be applied to the quantizer input in a pulse code modulation system. This is not the case for all A/D converters. We show that the dependence of error moments on input is inherently different in sigma delta modulators, and that the effect of dither depends on whether the quantiser is one bit or multibit. These statements are proven for simple SDMs and verified by simulation.
Dither and Noise Modulation in Sigma Delta Modulators

5936
Joshua D. Reiss,Mark Sandler,Derk Reefman,Erwin Janssen,
We present a mathematical framework, based on state space modelling, for the description of limit cycles of Sigma Delta Modulators (SDMs). Using a dynamical systems approach, the authors treat sigma delta modulators as piecewise linear maps. This enables us to find all possible limit cycles that might exist in an arbitrary sigma delta modulator with predefined input. We then focus on a DC input analyse their stability and show exactly the amount of dither that is necessary to remove any given limit cycle. Using several different SDM designs, we locate and analyse the limit cycles and thus verify the results by simulation.
Stability Analysis of Limit Cycles in High Order Sigma Delta Modulators

5937
Robert C. Maher,
Table look-up (or wavetable) synthesis methods have been widely used for many years in music synthesizers. Recently wavetable music synthesis has received new interest in the context of mobile applications such as personalized ring tones and pager notification signals. The limited amount of storage and processing available in such mobile devices makes the use of compressed wavetable data desirable. In this paper several issues related to wavetable compression are described and an efficient compression/decompression method is proposed for reducing the size of wavetable data while maintaining loop continuity and timbral quality.
Compression and Decompression of Wavetable Synthesis Data

5938
Helen H. Tarn,Chris H. Dick,
As audio applications become more complex, it is increasingly difficult to support the required signal processing functions using traditional DSP microprocessors due to their limited processing capability in terms of arithmetic capacity, datapath precision and input-output (I/O) bandwidth limitations. An alternative approach based on Field Programmable Gate Array (FPGA) technology is proposed, which not only preserves the versatility and flexibility of DSP microprocessors but provides the advantages of a customizable datapath which include high degrees of computational and data parallelism, arbitrary precision arithmetic and enormous I/O capacity. This paper presents a case study in which reconfigurable logic technology is employed for implementing infinite impulse response (IIR) filter banks. The study's results show that thousands of second-order filters can be implemented on a single FPGA and demonstrate the potential of reconfigurable logic technology for audio signal processing.
Reconfigurable Logic for Audio Signal Processing

5939
Jonathan S. Abel,David P. Berners,
A method for the design of discrete-time second-order shelf filters is developed which allows the response of an analog resonant shelf filter to be approximated in the digital domain. For filters whose features approach the Nyquist limit, the proposed method provides a closer approximation to the analog response than direct application of the bilinear transform. Three types of resonant shelf filter are discussed, and design examples are presented.
Discrete-Time Shelf Filter Design for Analog Modeling

5940
Srikanth Gurrapu,Doug Roberson,
Recent advances in CMOS VLSI digital technology make it practical now to pack a lot of high performance audio processing into an ASIC. To fully reap the benefits of these semiconductor technological advances, choosing the right architecture for a given application is crucial. Key considerations involve developing an architecture, which provides high audio performance, feature rich options, flexibility to adapt to market demand and highly efficient processing for a low cost. Architectural considerations are made to allow OEM developers to easily create custom filters and functional configuration settings to accommodate specific needs without any required knowledge of DSP programming.
High-performance Configurable Fixed-point Audio Processor Design Considerations

Google Exhibit No. 1063
Page 1410

5941
Sunil Bharitkar,Philip Hilmes,Chris Kyriakakis,
Traditionally, room response equalization is performed to improve sound quality at a given listener in applications ranging from automobile, home-theater, movie theater, multimedia education in classrooms. However, room responses vary with source and listener positions. Hence, in a multiple listener environment, equalization may be performed through spatial averaging of magnitude responses at locations of interest (e.g., in movie theater equalization). However, the performance of averaging based equalization, at the listeners, may be affected when listener positions change, or due to mismatch between microphone and listener positions (i.e., displacement effects). In this paper, we present a statistical approach to map displacement effects to a magnitude response averaging equalization performance metric. The results indicate that, for the analyzed listener configurations, the zone of equalization depends on, (i) distance of microphones/listeners from a source, (ii) the listener arrangement, and (iii) the source signal spectral composition. We have also provided an experimental validation of the theoretical results, thereby indicating the usefulness of the proposed closed form expression for measuring equalization performance due to displacement effects.
Sensitivity of Multichannel Room Equalization to Listener Position

5942
Allan Devantier,Todd S. Welti,
At low frequencies the listening environment has a significant impact on the sound quality of the audio system. Standing waves within the room cause large frequency response variations at the listening location. Furthermore, the frequency response changes significantly from one listening location to another; therefore, the system cannot be effectively equalized. A novel method to reduce seat-to-seat frequency response variation is described that allows the system to be equalized over a large listening area using relatively simple signal processing.
In-Room Low Frequency Optimization

5943
Lae-Hoon Kim,Jae-Jin Jeon,Sin-lyul Lee,Koeng-Mo Sung,
For home theatre systems, it is necessary to equalize the room impulse responses. It is well known that the perfect inverse filtering over the entire audio frequency range is hard to be met due to the perturbations such as a listener's head movement. For securing larger 'sweet region' we measured at 18 points 3cm at regular intervals horizontally around the both ears position. We synthesize one representative impulse response of these 18 impulse responses in the position-weighted manner using principal component analysis method. We apply this representative impulse response as the target of inverse filtering. For inverse filtering we apply two different methods into two frequency ranges. In the low frequency range we realize a perfect inverse filter using least square method, in the high frequency range we realize linear phase finite impulse response inverse filter using 1/3 octave band frequency magnitude response smoothing. Using this inverse filter we can confirm well-equalized high sound quality in the entire sweet region from the result of experiment.
Hybrid Equalization of Room for Home Theatre System

5944
Dirk Noy,Gabriel Hauser,John Storyk,
One of the major issues in small room acoustics is low frequency response and behavior. The goal of this paper is to present how and in what magnitude different commercially available low frequency absorbing devices control low frequencies and how off-the-shelf computer based simulation programs can be used to predict the low frequency behavior of small rooms. Reproducible acoustical measurements have been taken in a completely untreated rectangular concrete room, sequentially with and without a total of eight different absorbing devices as courteously provided by eight different international manufacturers. Results are compared and conclusions are presented.
Low Frequency Absorbers - Applications and Comparisons

5945
Eddy B. Brixen,
In an increasing number of radio facilities, large office open plan environments are used for audio production. Program material is prepared, edited, and finished ready for broadcast on DAWs without using studio acoustics, studio microphones, level or loudness metering, loudspeaker monitoring and even audio engineers. In many aspects this affects the sound quality. In this paper the drawbacks are discussed and suggestions for technical solutions are presented.
Audio Production in Large Office Environments

5946
Andrew Eloff,Gary Kendall,Michael R. Honig,
Artificial reverberation continues to be a source of much research and development. Currently, there is a heavy emphasis on "auralization," or the ability to simulate physical structures' sound characteristics using computer modeling. The diffuse component of early-order reflections has been acknowledged as an important

Google Exhibit No. 1063
Page 1411

component of reverb for several decades and has been implemented since the 1970s in varying forms. The current paper details a method of simulating diffuse reflections by use of fading models commonly used in wireless communications. The method is then considered as a stereo field enhancement effect and as a component of a reverberation system. Both are implemented in MATLAB and the results are discussed.
Diffuse Field Reverberation Modeled as a Flat Fading Channel

5947
Steven L. Harris,Jack Andersen,Daniel Chieng,
Digital input class D audio amplifiers will replace traditional analog types over the next 5 years. This paper describes a digital input class D amplifier controller integrated circuit (IC) which performs many of the functions needed to build a high performance class D audio amplifier module. A powerful Digital Signal Processor (DSP) is included allowing sophisticated modulation schemes, as well as additional audio signal processing. Possible processing functions include loudspeaker load compensation, equalization (EQ), time alignment, room acoustics compensation, howl prevention and other audio signal processing tasks. A novel clocking scheme decouples the input clock from the output switching clock, creating a highly jitter-tolerant design.
Intelligent Class D Amplifier Controller Integrated Circuit as an Ingredient Technology for Multi-Channel Amplifier Modules of Greater than 50Watts/Channel

5948
Brett G. Crockett,
A method of using auditory scene analysis of audio signals in conjunction with time and pitch scaling is presented. In the method described, a multi-channel audio signal is analyzed and the location and duration of the individual audio signal components that correspond to distinct auditory scene elements are identified. The audio data is then time and/or pitch scaled in such a way that the separate audio signal components are processed individually, thereby greatly reducing audible artifacts inherent in time and pitch scaling processing.
High Quality Multi-channel Time-Scaling and Pitch-Shifting Using Auditory Scene Analysis

5949
Thomas Holm Hansen,Lars Risbo,
A novel digital-domain adaptive calibration technique is proposed, which enables mitigation of analog related errors in over-sampled data converter systems. The technique is suited for all types of over-sampled A/D and D/A converters, e.g. multi-bit, 1-bit, PWM, etc. The calibration is done by adaptive fitting of a digital error model to the physical errors, due to component mismatch, etc.
Adaptive Digital Calibration of Over-Sampled Data Converter Systems

5950
James A. Angus,
Trellis Noise-Shaping Sigma-Delta modulators look forward at k samples of the signal before deciding to output a "one" or a "zero". The Viterbi algorithm is then used to search the trellis of the exponential number of possibilities that such a procedure generates. Means of making the search more computationally efficient have been proposed. This paper describes alternative tree based algorithms that can also be used search the exponential number possibilities generated by lookahead noise-shaping S-D modulators. Tree based algorithms are simpler to implement because they do not require backtracking through an array of scores to determine the correct output value. They can also be made more efficient via the use of the "Fano" or "Stack" algorithms, which are described.
Efficient Algorithms for Look-Ahead Sigma-Delta Modulators

5951
Sang-Myeong Kim,
Authentic binaural reproduction over loudspeakers using crosstalk cancellation is considered in this paper. A systematic time domain deconvolution method is presented using both stochastic and deterministic Wiener filters. This approach is advantageous over its frequency domain counterpart in that no additional stabilization process is required since the Wiener filter is inherently causal stable. A series of reproduction tests were conducted by changing the length of the Wiener filters in an anechoic chamber using a PC based reproduction system using a soundcard. Excellent performance was achieved with the filter length of only 500; less than 1% and 2% reproduction performance errors for the monaural and binaural reproduction tests, respectively.
Authentic Reproduction of Stereo Sound - A Wiener Filter Approach

5952
Vijay Parsa,Karthikeyan Umapathy,
We have evaluated objectively the comparative performance of five noise reduction algorithms. These algorithms were based on the Short-Time Spectral Amplitude (STSA) estimation, subspace projections, wavelet packets with auditory masking, and time-frequency decompositions using matching pursuits. Speech stimuli corrupted by speech-shaped noise and multi-talker babble at 5 different Signal-to-Noise Ratios (SNRs)

Google Exhibit No. 1063
Page 1412

were used to test the performance of the noise reduction algorithms. Noise reduction performance was quantified using two different methods. In the first method, the Perceptual Evaluation of Speech Quality (PESQ) measure was computed twice - once between the original and noisy speech and the other between the original and enhanced speech. The difference between these two PESQ values indicated the performance of the noise reduction algorithm. The second method was based on the "phase reversed noise" technique where the noise reduction algorithm was tested twice, once with speech + noise and then with speech + phase reversed noise. The PESQ and SNR gain measures were then computed on the combined output. The results obtained from this study indicate that the STSA based algorithm performs better in terms of the amount of noise reduction, while the wavelet packet based algorithm performs better in terms of minimizing the speech distortion introduced by the noise reduction process.
Objective Evaluation of Noise Reduction Algorithms in Speech Applications

5953
Fabio Bozzoli,Angelo Farina,
One of the most used intelligibility's parameters is Speech Transmission Index : the techniques for determining it employs artificial speaker and listener. In many cases (i.e. big rooms or system of telecommunications) the precision of directivity of artificial mouth doesn't influence too much the result; on the contrary inside cars, but also in other cases, the shape of the whole balloon of directivity is important for determining correct and comparable values and different mouths give really different results in the same situation. Moreover there isn't a current standard that fixes the whole balloon of directivity of artificial mouth but it defines only limits for some frontal position. For this reasons we have determined, in an anechoic room, the directivities of a real speaker and some artificial mouths and finally we have compared them for underlying the need of a more precise standard in this field.
Directivity Balloons of Real and Artificial Mouth Simulators for Measurement of the Speech Transmission Index

5954
Jan Holub,Radislav Smid,Michael D. Street,
The article describes an algorithm allowing intrusive speech transmission quality measurements in networks with low bit-rate coding schemes (600 – 2400 bits/second) as used in satellite and military communications. The proposed algorithm is based on PESQ (ITU-T P.862) perceptual model, enhanced by input noise resistance capability. The new algorithm reflects also noise cancellation capabilities of modern audio coders. The correlation between Absolute Category Rating (ACR) Mean Opinion Score (MOS) listening test results and output of the developed algorithm achieves 0.89 without the knowledge of original noise-free speech sample
Intrusive Speech Transmission Quality Measurements for Low Bit-Rate Coded Signals

5955
Se-Ung Kim,Sin-Lyul Lee,Lae-Hoon Kim,Koeng-Mo Sung,
The correct level alignment of the multi-channel reproduction system is critical for the quality of the reproduction. However, the level alignment in a conventional product is controlled manually by user. And if the user is not an expert, it is not easy to align correct level of each speaker. This paper provides how to apply binaural energy summation, which is from all of available positions of speakers, to level alignment for the arbitrary multi-channel reproduction system. If it is possible to measure the distance and the angle of a speaker, and if there is an mnidirectional microphone, it is possible to align the correct level automatically applying the binaural energy summation as the position of speakers obtained before.
Automatic Level Alignment for the Arbitrary Multi-channel Reproduction System

5956
Dae-young Jang,Jeongil Seo,Kyeongok Kang,Hoe-Kyung Jung,
In this paper, we introduce a new object-based 3D audio scene representation scheme. Typically, four kinds of 3D sound source objects are defined. Point sound source, line sound source, plane sound source and volume sound source are defined 3D sound source objects that are used for representation of object-based 3D audio scene. Each 3D sound source object includes sound source samples and related 3D audio scene information. User can interact with 3D sound source object and control the object by modifying the 3D audio scene information. We implement a prototype of object-based 3D audio player and several contents for demonstration of object-based 3D audio features. This object-based 3D audio representation scheme can be used very wide range of applications, like interactive game, home shopping and broadcasting realistic ambient sounds.
Object-based 3D Audio Scene Representation

5957
Harvey D. Thornburg,Randal J. Leistikow,
We propose a flexible state-space resynthesis approach that extends the sinusoidal model into the domain of source-filter modeling. Our approach is further specialized to a class of quasi-harmonic sounds, representing a variety of acoustic sources in which multiple, closely spaced modes cluster about principal harmonics loosely

Google Exhibit No. 1063
Page 1413

following a harmonic structure. We detail a variety of sound modification possibilities: time and pitch scaling modifications, cross-synthesis, and other potentially novel possibilities.
A Flexible Resynthesis Approach for Quasi-harmonic Sounds

5958
Brahim Hamadicharef,Emmanuel Ifeachor,
This study is concerned with objective prediction of perceived audio quality for an intelligent audio system for modeling musical instruments. The study is part of a project to develop an automated tool for sound design. Objective prediction of subjective audio quality ratings by audio experts is an important part of the system. Sound quality is assessed using PEAQ (Perceptual Evaluation of Audio Quality) algorithm, and this greatly reduces the time-consuming efforts involved in listening tests. Tests carried out using a large database of pipe organ sounds, show that the method can be used to quantify the quality of synthesized sounds. This approach provides a basis for the development of a new index for benchmarking sound synthesis techniques.
Objective Prediction of Sound Synthesis Quality

5959
Geoffroy Peeters,
This paper addresses the problem of classifying large databases of musical instrument sounds. An efficient algorithm is proposed for selecting the most appropriate signal features for a given classification task. This algorithm, called IRMFSP, is based on the maximization of the ratio of the between-class inertia to the total inertia combined with a step-wise feature space orthogonalization. Several classifiers - flat gaussian, flat KNN, hierarchical gaussian, hierarchical KNN and decision tree classifiers - are compared for the task of large database classification. Especially considered is the application when our classification system is trained on a given database and used for the classification of another database possibly recorded in completely diffierent conditions. The highest recognition rates are obtained when the hierarchical gaussian and KNN classifiers are used. Organization of the instrument classes is studied through an MDS analysis derived from the acoustic features of the sounds.
Automatic Classification of Large Musical Instrument Databases Using Hierarchical Classifiers with Inertia Ratio Maximization

5960
David Lowenfels,
The bandlimited digital synthesis model of Stilson and Smith is extended with a single feed-forward comb filter. Time-varying comb filter techniques are shown to produce a variety of classic analog waveform effects, including waveform morphing, pulse-width modulation, harmonization, and frequency modulation. The techniques discussed are not guaranteed to maintain perfect bandlimiting; however, they are generally applicable to other synthesis models.
Virtual Analog Synthesis with a Time-Varying Comb Filter

5961
Oliver Hellmuth,Eric Allamance,Markus Cremer,Holger Grossmann,Juergen Herre,Thorsten Kastner,
Finalized in 2001, the MPEG-7 Audio standard provides a universal toolbox for the content-based description of audio material. While the descriptive elements defined in this standard may be used for many purposes, audio fingerprinting (i.e. automatic identification of audio content) was already among the initial set of target applications that were conceived during the design of the standard. This paper reviews the basics of MPEG-7 based audio fingerprinting and explains how the technology has been used in a number of real-world applications, including metadata search engines, database maintenance, broadcast monitoring and audio identification on embedded systems. Appropriate selection of fingerprinting parameters and performance numbers are discussed.
Using MPEG-7 Audio Fingerprinting in Real-World Applications

5962
Christian Varani,Enrico Armelloni,Angelo Farina,
Car cockpit is a critical environment for music; sound reproduction is in fact quite conditioned by reflections, echoes, engine noise and loudspeakers' set up. An important technique to improve sound comfort is a spatial equalization where both magnitude and phase of signal are controlled. This technique is performed by a stereodipole system where two closely loudspeakers are setting in front of listener and digital processing is performed real-time by a DSP board. Cross-talk cancellation is achieved using FIR filters, whose coefficients are obtained by inversion of the measured cockpit impulse response. In this paper an experimental validation of a double stereodipole system, one for driver and other for passenger, is performed by subjective evaluations inside a commercial car.
Implementation of a Double StereoDipole System on a DSP Board – Experimental Validation and Subjective Evaluation Inside a Car Cockpit

5963

Google Exhibit No. 1063
Page 1414

Kenichi Taura,Masayuki Tsuji,Masayuki Ishida,Tsuyoshi Nakada,
We have developed a digital amplifier for car use. The major problems of conventional digital amplifiers are lack of the ability to reject the ripple and noise on the car power line, and the EMI (electro-magnetic interference). We have developed a novel feedback system for digital amplifiers to solve mainly the supply voltage ripple problem. By using a prototype with a feedback system, we have confirmed that it gives enough supply voltage ripple rejection for car audio systems. We also confirmed that the feedback system could contribute to reduce the EMI by relaxing the requirements to the switching stage operation, while keeping low distortion of an audio output.
Development of a Digital Amplifier for Car Use

5964
Maja Sliskovic,Hans-Juergen Nitzpon,
As new sound and video broadcasting systems are being deployed, the need for a multistandard multiband receiver increases. The compatibility of the software radio receiver with any defined broadcasting service is guaranteed by its re-programmability, i.e. by the fact that its functionality is determined by software. A receiver for a new standard can be added by a simple software download. In this paper a software implementation of the signal path from the ADC to the audio/video output will be discussed. The possibility of software reuse in different receiver functional blocks and for different broadcasting standards will be investigated.
Software Radio Receiver for Audio and Video Broadcasting Systems

5965
Hesu Huang,Chris Kyriakakis,
Additive background noise and convolutive noise in the form of reverberation are two major types of noise in audio conferencing and hands-free telecommunication environments. To reduce this type of acoustic noise, we propose a two-step approach based on subband adaptive filtering techniques. In particular, we first reduce the additive noise using a Delayless Subband Adaptive Noise Cancellation (DSANC) technique, and then suppress the convolutive noise through Subband Adaptive Blind Deconvolution. The experiments show that our method has lower computational complexity and better performance than previously proposed methods.
Subband Adaptive Filtering for Acoustic Noise Reduction

5966
Ching-Shun Lin,Chris Kyriakakis,
In this paper, a neuro-fuzzy system is proposed to remove multifrequency noise from audio signals. There are two major elements in our method. The first comprises a fuzzy cerebellar model articulation controller (FCMAC) that is used to quantize the signals. The second one is developed based on the theory of stochastic real values (SRV) that is used to search the optimal frequencies for the overall trained system. We present a DSP implementation of the SRV algorithm and results on its performance in removing spectral noise that is buried in audio signals.
Multi-Frequency Noise Removal Based on Reinforcement Learning

5967
Holger Crysandt,
Realtime music identification became more and more interesting within the past few years. Possible fields are for example monitoring a radio station in order to create a playlist or scanning network traffic is search of copyright protected material. This paper presents a client-server application which is able to do this in real-time with the help of MPEG-7. It explains how to define the similarity between two segments of music and determents its robustness towards perceptional audio coding and filtering. It also introduces an indexing system to reduce the number of segments which have to be compared to the query.
Music Identification with MPEG-7

5968
Han-Wen Hsu,Wen-Chieh Lee,Chi-Min Liu,
Current existing digital audio signals are always restricted by sampling rates and bandwidth fit for the various storages and communication bandwidth. Take for example the widely spread mp3 tracks encoded by the standard MPEG1 layer 3. The audio bandwidth in MP3 is restricted to 16 kHz due to the protocols constraints defined. This paper presents the method to reconstruct the lost high frequency components from the band-limited signals. Both the subjective and objective measures have been conducted and shown the better quality. Especially, the important objective measurement by the perceptual evaluation of audio quality system, which is the recommendation system by ITU-R Task Group 10/4 has proven a significant quality improvement.
High Frequency Reconstruction by Linear Extrapolation

5969
Doug Perkins,Amnon Sarig,
There are many ways to store audio files but traditional methods are not conducive to file sharing, which is the

Google Exhibit No. 1063
Page 1415

goal of the modern networked facility. While still evolving, the world of digital audio storage offers the technology to quickly and safely share files between users, and even allows simultaneous users on different platforms to access audio. Learn how your facility can simplify the transition to the digital world, what products are on the leading edge, and what to look for when you are ready to make the leap.
Audio Storage and Networking in the Digital Age

5970
Nicolas Sincaglia,
In a networked audio environment, metadata not only provides a human interface but is used for the identification, organization, tracking, reporting and selling of digitized sound recordings. Establishing an open industry standard for this data will enable the entire industry to streamline its ability to make content available. The result will be a more efficient and uniform exchange of data, ultimately enabling a more versatile and profitable music industry.
The Requirement for Standards In Metadata Exchange for Networked Audio Environments

5971
Peter Mapp,
A new technique for improving the spaciousness of reproduced sound has been investigated. The technique uses a combination of conventional Pistonic loudspeakers and Distributed Mode (DML) devices. Objective measurements have been made in a range of rooms and show that 'Layered Sound' affects parameters such as the Inter Aural Cross Correlation (IACC) and Lateral Energy Fraction as well as Centre Time and Early Decay Time. A number of conditions were investigated including listening room configuration and the relative sound levels of the loudspeakers. Limited subjective testing was carried out in order to ascertain the preferred relative intensities between the conventional and DML loudspeakers. This was confirmed to be optimal with the DMLs set within the range −5dB ± 3dB relative to the conventional stereo loudspeakers. The configuration of the listening room and the type of programme material (and recording technique) were also found to be significant factors. It is shown that over a range of conditions, Layered Sound enhances the perceived spaciousness, envelopment and clarity of reproduced sound, though some changes to the original stereo image were noted.
An Investigation of Layered Sound

5972
Sandra Brixen,Frank Melchior,Thomas Roder,Stefan Wabnik,Christian Riegel,
Wave Field Synthesis (WFS) permits the reproduction of a sound fi eld, which fi lls nearly the whole reproduction room with correct localization and spatial impression. Because of its properties, the WFS technology shows enormous potential to be used for creation of audio in combination with motion pictures. For this application a special authoring system is in development, which gives audio engineers the possibility to automate the process of positioning sound sources, and takes several ergonomics as well as technical features into account. This paper presents fi rst experiences of the mixing process for content production using WFS. It also demonstrates the authoring system capabilities for WFS and shows future developments.
Authoring Systems for Wave Field Synthesis Content Production

5973
Jose Vieira,Luis Almeida,
This paper proposes an intelligent acoustic sensor able to localize sound sources in acoustic environments with strong reverberation. The proposed system is inspired on the precedence effect used by the human auditory system and uses only two acoustic sensors. It implements a modified version of the algorithm proposed by Huang that uses the precedence effect in order to achieve robust sound localization even in reverberant environments. The localization system was impletented in a C31 DSP for real time demonstration and several experiments were performed showing the validity of our solution. Finally, the paper also proposes a method to estimate on-line the decay of the reverberation, using the received sound signals, only.
A Sound Localizer Robust to Reverberation

5974
Banu Gunel,
Fidelity of spatial hearing effects generated by loudspeakers degrades outside the sweet spot. This paper proposes a method to analyze a sound scene created by stereo loudspeakers and modify it with the help of assistant headphones. Loudspeaker positions are found by analyzing B-format recordings at the listener position. This information is evaluated to generate signals for assistant headphones. Headphone signals are filtered with head related transfer functions for amplitude panning to change the positions of the loudspeakers. Then, other headphone signals are added which lead the original sounds for further loudspeaker and lag the original sounds for nearer loudspeaker creating direct-echo pairs. The system improves listening conditions outside the sweet spot and is extendable to multiple listeners.
Modification of Louspeaker Generated Posision Cues through Assistant Headphones

5975

Google Exhibit No. 1063
Page 1416

Jens Meyer,Tony Agnello,
This paper describes a beamforming spherical microphone array consisting of many acoustic pressure sensors mounted on the surface of a rigid sphere. The beamformer is based on a spherical harmonic decomposition of the sound field. We show that this design allows a simple and computationally effective, yet very flexible beamformer structure. The spherical shape of the array in combination with the beamformer allows steering the look-direction the to any angle in 3-D space. Measurements from an array with 37.5mm radius that consists of 24 senors are presented. The paper focuses on the applications of directional sound pick-up and sound field analysis and reconstruction. Other suitable applications are e.g. room acoustic measurements and 'forensic beamforming'.
Spherical Microphone Array for Spatial Sound Recording

5976
Huseyin Hacihabiboglu,
The precedence effect refers to the property of the human auditory system that enables accurate localization of sound sources where many interfering echoes of the original signal are also present. Perception of the elevation, azimuth, and distance of sound sources are affected in the presence of an echo. The multivariate Gaussian mixture model proposed in this paper combines azimuth, elevation and distance perception, and provides a general framework for modeling sound source localization under the precedence effect. The model interprets the precedence effect as a spatial property rather than a temporal one.
Modeling Sound Source Localization under the Precedence Effect Using Multivariate Gaussian Mixtures

5977
Cheng-Han Yang,Hsueh-Ming Hang,
A low complexity and high performance scheme for choosing MPEG-4 Advanced Audio Coding (AAC) parameters is proposed. One key element in producing good quality compressed audio at low rates is selecting proper coding parameter values. The MPEG committee AAC reference model does not do well on this job. A joint trellis-based optimization approach has thus been previously proposed. It leads to a near-optimal selection of parameters at the cost of extremely high computational complexity. It is, therefore, very desirable to achieve a similar coding performance (audio quality) at a much lower complexity. Simulation results indicate that the proposed cascaded trellis-based optimization scheme has a coding performance close to that of the joint trellis-based scheme, but it requires only 1/70 computations.
Cascaded Trellis-Based Optimization for MPEG-4 Advanced Audio Coding

5978
Markus Lohwasser,Marc Gayer,Manfred Lutzky,
Encoder implementations of MPEG Advanced Audio Coding and Layer-3 (mp3) on 32-bit or 16-bit fixed-point processors are challenging due to the fact that the usable word length is restricted to 32 bits if low processing power is required. This paper describes the modifications and optimizations that had to be applied to the algorithms of these audio encoders to make a true fixed-point implementation on a 32-bit or 16-bit device possible. The implementation had to be done without using floating-point emulations or even 64-bit values for the signal energies and thresholds in the psychoacoustic model of the encoder. At the same time high encoding quality and speed were required. Memory and processing power requirements on various platforms as well as results from a subjective listening test will be presented.
Implementing MPEG Advanced Audio Coding and Layer-3 Encoders on 32-bit and 16-bit Fixed-point Processors

5979
Dai Yang,Takehiro Moriya,Akio Jin,Kazunaga Ikeda,
An extended-run-length coding tool called zero compaction is proposed in this paper. The proposed coding tool is far more efficient and generates significantly better results than traditional run-length coding when a certain type of data appears in the middle stage of our lossless audio compression systems. The zero-compaction technique has been integrated into our lossless audio coding scheme. When tested on seven sets of standard audio files from the ISO, the inclusion of zero compaction improved the average compression rates of systems by more than 1% in all cases. In addition, the simplicity and good performance of zero-compaction shave up to 4% from the total encoding time.
An Extended-Run-Length Coding Tool for Audio Compression

5980
Jeongil Seo,Dae-Young Jang,Gi Yoon Park,Kyeongok Kang,
We present implementation procedure of the interactive 3D audio player using MPEG-4 standards. MPEG-4 standards allow to interact with important objects in a scene as well as to provide the high efficiency coding tools for audiovisual objects. We apply the AudioBIFS system (version 1), parts of the BIFS in the MPEG-4 Systems standards, to provide an object-based interactivity for audio objects. It also provides geometric information for spatializing an audio object into the 3D audio scene. Because the method of 3D spatialization

Google Exhibit No. 1063
Page 1417

is not normative in the MPEG-4 AudioBIFS, we adopt a novel 3D spatialization method based on HRTF processing. To prohibit excessive increase of the number of audio objects, we defined the background audio object to present the atmosphere of 3D audio scene. While the important audio objects are separated as audio objects individually, the other audio objects are merged into the background audio object. Though our interactive 3D audio player is developed for a terminal player in Digital Multimedia Broadcasting (DMB) system, it can be also applied to 3DTV, virtual reality games, interactive home shopping applications, etc.
Implementation of Interactive 3D Audio Using MPEG-4 Multimedia Standards

5981
Schuyler Quackenbush,Peter F. Driessen,
This paper investigates techniques for mitigating the effect of missing packets of MPEG-4 Advanced Audio Coding (AAC) data so as to minimize perceived audio degradation. Applications include streaming of AAC music files over the Internet and wireless packet data channels. A range of techniques are presented, but statistical interpolation in the time/frequency domain is found to be the the most effective. The novelty of the work is to use statistical interpolation techniques intended for time domain samples on the frequency domain coefficients. A means of complexity reduction is presented, after which the error mitigation is found to require on average 17\% additional computation for a channel with 5\% errors as compared to a clear channel. In an informal listening test, all subjects preferred this technique over a more simplistic technique of signal repetition, and for one signal item statistical interpolation was preferred to the original.
Error Mitigation in MPEG-4 Audio Packet Communication Systems

5982
Anibal J. S. Ferreira,Andre C. Rocha,
An advanced Audio Spectral Coder (ASC) is described that implements a new approach to the problem of efficient audio compression by combining source coding with perceptual coding techniques. This approach involves the audio segmentation into three main components: transient events in the time domain, harmonic structures of sinusoids and stationary noise in the MDCT frequency domain. It is shown that this segmentation permits the independent parametrization and coding of components according to appropriate representation models and applicable psychoacoustic rules. The coding performance of ASC is characterized and it is shown that due to its underlying structure, addiitional functionalities other than compression are also possible namely bitstream semantic scalability, access and classification.
Combined Source and Perceptual Audio Coding

5983
Albertus C. den Brinker,Andy J. Gerrits,Rob J. Sluijter,
In sinusoidal coding, frequency tracks are formed in the encoder and the amplitude and frequency information of a track is transmitted. Phase is usually not transmitted, but reconstructed at the decoder by assuming that the phase is the integral of the frequency. Such a reconstructed phase accumulates inaccuracies, leading to audible artifacts. To prevent this, a mechanism is proposed to transmit phase information of sinusoidal tracks. The unwrapped phase is quantised and transmitted. The frequencies are not transmitted, but restored from the phase information by differentiation.
Phase Transmission in Sinusoidal Audio and Speech Coding

5984
David Isherwood,Ville-Veikko Mattila,
Listening tests were performed to define the perceived loudness threshold above which a number of mobile terminal alert tones became only partially masked by environmental noise. The relative intensity of the noise and alert tone at the partial masking threshold was then used to create masker+maskee samples to be objectively measured with various loudness models. The objective estimates were then compared to the subjective results to derive recommendations as to how best to objectively estimate the partial masking thresholds.
Objective Estimates of Partial Masking Thresholds for Mobile Terminal Alert Tones

5985
Arkady Gloukhov,
In accordance with the Huygens-Fresnel principle, the radiating device is simulated as an array of point sources in the aperture. The aperture complex pressure can be calculated by means of a model of wave propagation along the waveguide based on the Huygens-Fresnel principle or by means of approximating waveguide procedure using experimental polar data in a traditional form. The method predicts dispersion pattern at any distance with any needed resolution. Phase patterns as well as dispersion vs. distance relationship are demonstrated on samples. Proposed algorithms can be used in high-resolution simulators of loudspeakers and arrays. Results also point to another approach of measurements and presentation of dispersion data.
A Method of Loudspeaker Directivity Prediction Based on Huygens-Fresnel Principle

Google Exhibit No. 1063
Page 1418

5986
Peter Mapp,
Although it is well known that equalising a sound system can significantly affect the perceived sound quality and intelligibility, surprisingly there is little or no information relating to the degree of improvement in intelligibility that can be achieved. Measurement data relating to over 30 sound systems has been reviewed and a number of factors relating to typical response anomalies identified. Large discrepancies were often noted to occur between the measured in-room sound system or loudspeaker response and published anechoic frequency response data. The underlying causes and implications relating to speech intelligibility are discussed. The results of a pilot study illustrating the improvements that appropriate equalisation can produce are presented. It is shown that under some conditions, improvements of over 20 % in speech intelligibility can be achieved. However, it is also noted that none of the current electro-acoustically based measurement metrics, including STI, % Alcons, STIPa or RaSTI correctly indicate the intelligibility improvements that system equalisation produces.
Some Effects of Equalisation on Sound System Intelligibility & Measurement

5987
Dai Yang,Takehiro Moriya,
Most lossless audio-coding algorithms are designed for PCM input sound formats. Little work has been done on the lossless compression of IEEE floating-point audio files. An efficient lossless-coding algorithm that handles IEEE floating-point format data as well as PCM format data is described in this paper. In the worst-case scenario, where the algorithm was applied to artificially generated 48-kHz sampling frequency and 32-bit floating-point sound files, an average compression ratio of 65% was still achieved for sound files with 48kHz sampling frequency. Moreover, the proposed algorithm is easily extensible to lossless/variable-lossy operation, which will provide scalability to accommodate the requirements of a wider range of applications and platforms.
Lossless Compression for Audio Data in the IEEE Floating-Point Format

5988
Anibal J. S. Ferreira,
The performance of perceptual audio coders depends on the efficiency of the quantization operation in masking the quantization noise under the audio signal. This objective is better addressed by coding separately different signal components such as sinusoids, transients and stationary noise. In this paper we use an audio coder that normalizes the MDCT spectrum by a smooth spectral envelope and by periodicities due to sinusoids. The resulting flattened MDCT coefficients are shown to exhibit a probability density function with small uncertainty allowing the design of an optimum nonuniform scalar quantizer. Its distortion-rate function is derived, is compared to that of of known quantizers, and compared to that obtained under real audio coding conditions.
Optimum Quantization of Flattened MDCT Coefficients

5989
Graeme Huon,Zeljko Velican,
The future delivery of sound must create or render a realistic sound event image. A model for 3D sound capture and render format has been proposed earlier, based on Depth Perception. This paper considers the requirements for optimised low frequency reproduction and non-bass masking effects related to this model. Theoretical modeling, practical verification in real listening environments and subjective assessment with skilled and unskilled listeners is presented and conclusions drawn. It is shown that room mode influence, low frequency spatial energy distribution and main system integration for low frequency reproduction in rooms can be effectively managed. New apparatus is described that enables the loudspeaker to be placed optimally with respect to the listener so as to minimise room mode colouration of low frequencies. Its use to quantify frequency dependent loss is defined. Nonbass spatial masking effects are also reported in the context of the Depth Perception model.
Low Frequency Optimisation and Non-bass Masking Effects for Sound Field Re-creation

5990
Ruihua Ma,
Based on recent information hiding theory, information hiding may be reviewed as a game between hider (embedder/decoder) and attacker, and optimal information-embedding and attack strategies may be developed in this context. So far, there is a lot of work to apply this theory to image watermarking. However, little research work is about applying this theory to audio watermarking. This paper aims to fill in this gap. An information-theoretic model for audio watermarking is presented. We use wavelet statistical models for audio signals and compute data-hiding capacity for compressed and uncompressed host-audio sources. The simulation shows that the within experimental results, the proposed system has near-optimal performance compared to the theoretical upper bounds.
An Information-Theoretic Model for Audio Watermarking

Google Exhibit No. 1063
Page 1419

5991
Seung-Jin Yang,Do-Hyoung Kim,Jae-Ho Chung,
We suggest the watermarking techniques which inserts additive informations into quantized integer coefficients whose values are over 15, called linbits, during Huffman coding in MP3 encoding procedure. The linbits are inserted into the bitstream with binary codes as it is. We inserted watermarks by modifying the linbits, and made an experiment evaluating audible distortion through the MOS Test. In our experiment, 20 untrained listeners were asked to rate 20 samples of about 15 seconds in which watermark is inserted at 128kbps, according to perceived quality on a scale of 1(very annoying) to 5 (imperceptible). As a result, we confirmed that we could insert the additional informations or watermarks of about 60bytes/second with sound quality of MOS 4.6 on an average.
Watermark Insertion Into MP3 Bitstream Using the Linbits Characteristics

5992
Chung-Han Yang,Jiun-In Guo,Wen-Chieh Lee,Chi-Min Liu,
The FIR-based reverberators, which convolve the input sequence with an impulse response modelling the concert hall, have better quality compared to the IIR-based approach. However, the high computation complexity of the FIR-based reverberators limits the applicability to most cost-oriented system. This paper proposes a method that uses perceptual criterion to reduce the complexity of convolution methods. Also, an objective measurement criterion is introduced to check the perceptual difference from the reduction. The result has shown that the length of impulse response can be cut off by 60% without affecting the perceptual reverberation quality. The method is well integrated into the existing FFT-based approach is have around 30% speed-up. Also, the method has a high flexibility to various computation complexities with graceful degradation to the reverberation quality.
Perceptual Convolution for Reverberation

5993
Erwin Janssen,Derk Reefman,
Recently, a new type of 1-bit Sigma Delta Modulator (SDM), called Trellis Noise-Shaping Converter, has been introduced. It offers several advantages compared to a standard SDM, including better performance in stability, signal to noise ratio (SNR) and linearity. The major drawback of the Trellis architecture is the large computational requirement. This paper refines the concept of Trellis noise-shaping, and introduces a new algorithm that offers better performance at horrendously reduced cost. On this new algorithm, a comparative performance analysis has been performed. Cost savings of multiple orders of magnitude have been achieved, while maintaining all the benefits of Trellis noise-shaping. Finally, special attention has been paid on critical implementation details.
Advances in Trellis Based SDM Structures

(C) 2003, Audio Engineering Society, Inc.

Google Exhibit No. 1063
Page 1420



Audio Engineering Society Preprints

This search engine provides querying through a database of AES Preprints for AES Conventions **1** through **115**; by Author, Title, Preprint Number, Associated Conven Associated Convention date, Journal Volume, Journal Issue and Year of Journal Publication (if applicable).

The search utility will return matching results providing all available information and includes an option to automatically generate a preprint order form for one or more articles. Order forms can be generated for articles found during one or more consecutive searches.

You may also purchase full preprint sets for the 104th through the 115th Convention on CD-ROM and in paper format from Special Publications.

If you are purchasing multiple copies of the same preprint please do not do so online. Contact Andy Veloz (below) for large volume purchases.

For more information, contact Andy Veloz at AES Headquarters: 1-212-661-8528 ext 39 or by using this e-mail form.

Single Convention Preprints through #3347 (92nd AES Convention) are $4.00 for AES Members and $5.00 for non-Members. Single Convention Preprints after #3347 AES Members and $10.00 for non-Members.

The AES Preprints database is now complete through convention #115.

The Wayback Machine - https://web.archive.org/web/20040404224624/http://www.aes.org:80/publications/preprints/searc...

Audio Engineering Society Preprints

This search engine provides querying through a database of AES Preprints for AES Conventions **1** through **115**; by Author, Title, Preprint Number, Associated Convention Number, Associated Convention date, Journal Volume, Journal Issue and Year of Journal Publication (if applicable).

The search utility will return matching results providing all available information and includes an option to automatically generate a preprint order form for one or more chosen articles. Order forms can be generated for articles found during one or more consecutive searches.

You may also purchase full preprint sets for the 104th through the 115th Convention on CD-ROM and in paper format from Special Publications.

If you are purchasing multiple copies of the same preprint please do not do so online. Contact Andy Veloz (below) for large volume purchases.

For more information, contact Andy Veloz at AES Headquarters: 1-212-661-8528 ext 39 or by using this e-mail form.

Single Convention Preprints through #3347 (92nd AES Convention) are $4.00 for AES Members and $5.00 for non-Members. Single Convention Preprints after #3347 are $5.00 for AES Members and $10.00 for non-Members.

The AES Preprints database is now complete through convention #115.

AES Preprint Search Utility

Title Search:

Author First Name:

Author Last Name:

Preprint number:

Associated Convention: *date (YY-MM):* *number:*

Journal Issue: *vol:* *issue:* *page:* *year:*

Begin AES Preprint Search

(C) 2004, Audio Engineering Society, Inc.

Google Exhibit No. 1063
Page 1422

Ex. 1041H

Ex. 1041H
Early Citations to Wolters

Source Document

Wolters, M., Kjorling, K., Homm, D., & Purnhagen, H. (2003, October). A closer look into MPEG-4 High Efficiency AAC. In *Audio Engineering Society Convention 115,* Convention Paper 5871.

Citations: 2004 (in chronological order)

Hsu, H. W. (2004, June). MP3 與 AAC 解碼器上的音訊修補方法. Audio Patch Method in Audio Decoders-- MP3 and AAC. Doctoral dissertation, National Chiao Tung University. Institutional Repository metadata record for Hsu at https://ir.lib.nycu.edu.tw/handle/11536/50035.

Tseng, C. T. (2004, June). MPEG-4 先進音訊編碼 在 DSP/FPGA 平台上的實現與最佳化. MPEG-4 AAC Implementation and Optimization on DSP/FPGA. Master's thesis, National Chiao Tung University. PDF at https://ir.lib.nycu.edu.tw/bitstream/11536/44179/1/165601.pdf

Yokotani, Y. (2004, December). *Integer Transforms for Lossy/Lossless Audio Coding*. Doctoral dissertation, University of Texas at Arlington. Available from ProQuest Dissertations & Theses Global. (305058315).



Do Not Zero-Pute: An Efficient Homespun MPEG-Audio Layer II Decoding and Optimization Strategy

Patrick De Smet
Dep. of Telecommunications and Information Processing
Ghent University, TELIN/TW07
Sint-Pietersnieuwstraat 41
B-9000 Ghent, Belgium
pds@telin.UGent.be

Filip Rooms
Dep. of Telecommunications and Information Processing
Ghent University, TELIN/TW07
Sint-Pietersnieuwstraat 41
B-9000 Ghent, Belgium
frooms@telin.UGent.be

Hiêp Quang Luong
Dep. of Telecommunications and Information Processing
Ghent University, TELIN/TW07
Sint-Pietersnieuwstraat 41
B-9000 Ghent, Belgium
hiep.luong@telin.UGent.be

Wilfried Philips
Dep. of Telecommunications and Information Processing
Ghent University, TELIN/TW07
Sint-Pietersnieuwstraat 41
B-9000 Ghent, Belgium
philips@telin.UGent.be

ABSTRACT

In this paper we point out that the general principle "do not compute what you do not need to compute" can be applied easily and successfully within a MPEG audio decoding strategy. More specifically, we will discuss the problem of eliminating costly computation cycles being wasted at processing useless zero-valued data. Hence, the title: "do not zero-pute". At first, this may all sound somewhat obvious or trivial. Indeed, this can be true in many cases, but experience gathered in various teaching related projects during several academic years has also lead us to believe the opposite. Moreover, a survey of the existing literature quickly reveals that the approach discussed below has not been investigated and documented properly. Although we will only illustrate our optimization approach by discussing the MPEG-audio layer II decoding process in detail, we hope the reader will be able to apply, extend, and implement the basic principles presented here within many other applications.

Categories and Subject Descriptors

F.2.m [**Analysis of algorithms and problem complexity**]: Miscellaneous; D.1.m [**Programming Techniques**]: Miscellaneous

General Terms

Performance, Design, Algorithms

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

MM'04, October 10-16, 2004, New York, New York, USA.

Copyright 2004 ACM 1-58113-893-8/04/0010 ...$5.00.

Keywords

MPEG audio, optimized decoder

1. INTRODUCTION

Generally speaking, multimedia standards are commonly based on some type of frame based "syntax", i.e., a repetitive and conditionally templated data structure that can be used to describe an encoded bit stream. Typically, the body of a multimedia bit stream frame contains a set of "control bits" that indicate how the remaining set of condensed information bits are to be processed further. By decoding and interpreting the control bits one typically arrives at an array of data that needs to undergo some further signal processing procedure, e.g., a subband audio reconstruction filtering procedure, or a JPEG 2D- or audio M-DCT, etc. This signal reconstruction process is most often responsible for the larger part of the computational complexity of the entire decoding process. Obviously, many excellent optimization strategies have been devised in order to arrive at various memory and/or computationally efficient implementations of several specific algorithms. Unfortunately, optimizing some of the basic algorithms typically involves designing and implementing more complex control logic. Very often this may not yield any real problem, but in some cases a straight-forward search for conceptually more simple optimizations, i.e., in our case the zero-pute paradigm, may also yield some very interesting results that can also be combined with other more complex optimizations.

Up to this point, we, and many other authors with us, have implicitly assumed that signal processing algorithms are always operating on a fixed-width array of meaningful data. In principle, this assumption would make the data structures and the many functions that may depend upon them, simple and efficient to be implemented. However, in many situations a relatively large part of the data structures may not contain any real or useful decoded data, e.g., large chunks of the array may be zero-filled (by default, or even by an inefficient encoder!). Obviously, it makes no sense to

Google Exhibit No. 1042
Page 001

Google Exhibit No. 1063
Page 1425

drag along, e.g., multiply and/or add, this zero-valued data; one should try to remove this data from the computational process as early as possible: do not zero-pute.
It is important to note that the zero-pute reduction strategy has already been experimentally implemented and used for various applications, e.g. in [1] and [2]; an important goal of this paper is to return some renewed, more complete, and more formal attention to this optimization principle. Despite the emphasis of this paper on formal results, a package containing the source code implementing all techniques and formulas discussed below is freely available from the main author. Preliminary simulation results on a 200 Mhz Pentium MMX processor have already confirmed the more formal results reported below.

Let us now illustrate how we can apply the zero-pute reduction strategy within the implementation of an optimized MPEG-audio decoder. Note that while several excellent papers have documented important optimizations for building efficient decoders, the issues discussed below seem never to have been reported on fully.
Additionally, please note that although we will focus on a layer II decoder, layer I and layer III, a.k.a. "MP3", also require the subband synthesis filtering step discussed below. The layer II standard is a crucial component of the DAB (Digital Audio Broadcasting) standard.

2. MPEG-AUDIO LAYER II DECODING

The MPEG layer II audio coding scheme operates by representing the input audio signal as bit stream encoded sets of samples of 32 equally spaced subband signals. Figure 1 illustrates the general layer II syntax layout. Each frame encodes 1152 samples. The header encodes synchronization and general parameter information (mono/stereo/dual channel, bit rate, sampling frequency, etc.). The CRC provides optional error detection. The bit allocation information (BIT ALLOC) points out which subband signals have been encoded and indicates their respective quantization accuracy. The scale factor side information (SCFSI) indicates how many scale factors (SCALEFAC) are encoded. The scale factors are used to rescale the normalized and quantized samples to their original value range. The ANC field can be used to fill up the frame size to its appropriate length, or to encode other "user defined" information (e.g. song information etc.). At the encoder side, compression is achieved by using a so-called psycho-acoustical model that enables the bit allocation routine to roughly quantize, or even remove, the signal components that are less relevant perceptually. An excellent tutorial on MPEG-audio coding can be found in [3].

Let us now consider the decoder side. For each multirate filtering step at time t, the basic ISO synthesis subband filtering algorithm starts with $M = 32$ subband samples extracted from the encoded bit stream syntax, i.e., $S_{sb}[i;t]$, $i = 0,1,\ldots,M-1$, and produces M PCM audio output samples; $S_{pcm}[j;t], j = 0,1,\ldots,M-1$. First, the intermediate IPQMF (inverse polyphase quadrature mirror filter bank) vector $V[i;t]$ is shifted over $2M$ memory positions according to: $V[i;t]=V[i-2M;t-1]$, for $i = 2PM-1, 2PM-2, \ldots, 2M$ with $P = N/M = 16$ and $N = 512$ the length of the prototype filter window. Then, the basic MPEG audio matrixing operation [6] is calculated:

$$V[i;t] = \sum_{k=0}^{M-1} S_{sb}[k;t]\, cos(\pi(2k+1)(M/2+i)/2M), \qquad (1)$$

for $i=0,1,\ldots,2M-1$.
A N-value intermediate vector $U[i;t]$ can then be built using both $U[2Mk+l;t]=V[4Mk+l;t]$, $U[2Mk+M+l;t]=V[4Mk+3M+l;t]$, for $k = 0,1,\ldots,P/2-1$, and $l = 0,1,\ldots,M-1$. Finally, the so-called windowing component computes M new PCM audio output samples, i.e., $S_{pcm}[j;t] = \sum_{i=0}^{P-1} U[j+Mi;t]\, D[j+Mi]$, $j=0,1,\ldots,M-1$, in which the values D are the N synthesis window coefficients (see Table B.3 in [6]). This algorithm requires $2M^2+MP = 2560$ multiplications and $2M(M-1)+M(P-1) = 2464$ additions. Obviously, several optimizations can be used to realize important speed-ups of Eq. (1), but this always involves more complex control logic. This will be discussed further below; see Eq. (4).

3. STATIC CUT-OFF OPTIMIZATIONS

Now, examining several layer II encoded bit streams easily reveals that any real bit stream normally does not contain meaningful samples for all 32 subbands. Let us investigate this in somewhat more detail.

While the standard was being drafted, it became clear that (at low bit rates) high frequency subbands would rarely (need to) be encoded. Hence, the standard contains a set of tables that specify a hard subband cut-off index for each possible combination of the audio sampling rate and the encoded bit rate; e.g. at 32 kbit/s encoder output for a CD-quality 44.1 kHz audio input, the subband cut-off index is 8; see Table B2.c in [6]. The cut-off index is called the subbandlimit ("`sblimit`" in [6]); no subbands above it are allowed to be encoded. We will call this behavior "static" or "`sblimit` cut-off".
A simple standard decoder will zero-fill any subband sample values above the cut-off index, and use Eq. (1) with its full subband range to reconstruct the audio signal. Hence, if we drop this zero-valued subband range and reformulate Eq. (1) as

$$V[i;t] = \sum_{k=0}^{\texttt{sblimit}-1} S_{sb}[k;t]\, cos(\pi(2k+1)(M/2+i)/2M) \qquad (2)$$

we can realize an important optimization.
Table 1 indicates the static subbandlimit cut-offs defined within the MPEG-1 standard and the worst-case computational complexity of the subband synthesis matrixing procedure based on Eq. (2). You can compare these results with the $2M^2 = 2048$ multiplications and the $2M(M-1) = 1984$ additions of Eq. (1).

4. DYNAMIC CUT-OFF OPTIMIZATIONS

Next, even if we consider these predefined cut-offs, a normal encoder will still choose to allocate most of the bits to the subbands (well) below the hard cut-off index. This is due to the perceptual properties of the human hearing system that are modeled in the psycho-acoustic component of the encoder. Hence, some subbands will not be assigned any quantization bits, and many zero-valued `BIT ALLOC` values, see Fig. 1, will be encoded. If a simple syntax decoding algorithm would again be used, all subband sample values for each of these zero-allocated subbands, would be set to zero, and the output of the standard synthesis filter would be computed using Eq. (1). Again, this would be highly inefficient. Using Eq. (2) we can improve upon this situation, but even then we can easily do better: for each decoded frame, we can determine the largest subband index that has been assigned any quantization bits and use this index as a new subband cut-off index; i.e, we can use the simple algorithm

```
/*  determine alloc_cut, i.e. locate Y!=0 in:      */
/*  xxxxxxxY0<-00...00  BIT ALLOCation array       */
for (chan_i=0; chan_i < stereo; chan_i++) {
  for (j_pds = sblimit; j_pds > 0; j_pds--)
      if (bit_alloc[chan_i][j_pds] != 0)       break;
  alloc_cut[chan_i] = ++j_pds;        /* increment:*/
} /* so we can use "< alloc_cut" testing elsewhere */
```


Google Exhibit No. 1042
Page 002

Google Exhibit No. 1063
Page 1426

HEADER	CRC	BIT ALLOC	SCFSI	SCALEFAC	QUANT. SUBBAND SAMPLES	ANC

Figure 1: MPEG-1 syntax structure (layer II).

Table 1: Upper limits for the MPEG layer II subband synthesis computational complexity using the proposed static-`sblimit` cut-off optimization approach

ISO standard	Table B.2a	Table B.2b	Table B.2b	Table B.2d
`sblimit`	27	30	8	12
max. complex. ×	1728	1920	512	768
max. complex. +	1664	1856	448	704

and an adapted formula

$$V[i;t] = \sum_{k=0}^{\texttt{alloc\_cut}-1} S_{s\hat{b}}[k;t]\, cos(\pi(2k+1)(M/2+i)/2M) \quad (3)$$

to realize another speed-up. Note that due to their simplicity neither the cut-off determination algorithm, nor the adapted formula, introduce any real additional complexity.

Obviously, this cut-off for a specific frame at a given instance in the bit stream will vary, and depend on the perceptual content of that frame. Hence, we will call this behavior "dynamic cut-off".

5. INITIAL OPTIMIZATION RESULTS

Table 2 illustrates some average decoded cut-off subband indexes, i.e., μ[`alloc_cut`], determined by decoding an example audio bit stream using an adapted version of the pseudo code given above and the ISO codec (coder-decoder) from [7]. The computation of the values $V[i;t]$ in Eq. (3) requires $2M$`alloc_cut` multiplications and $2M$(`alloc_cut` -1) additions; Table 2 also indicates the complexity computed using the average (over all frames) for the dynamic cut-off values and the corresponding static cut-off results. Again, please note that Eq. (1) required $2M^2 = 2048$ multiplications and the $2M(M-1) = 1984$ additions. Despite the conceptual and mathematical simplicity of these proposed optimizations they are most often glanced over. Instead, significantly more complex optimization strategies are often pursued. These more complicated optimization strategies will yield important improvements in terms of computational complexity, but they can do so only by also significantly increasing the time needed to study, program, verify and test their required adaptations !

6. COMBINATION WITH ADVANCED OPTIMIZATION STRATEGIES

Let us now investigate the cut-off optimizations discussed above and show that it can still be combined with more complex existing optimizations. As already indicated earlier, in real-life situations it can be inefficient to implement the matrixing procedure as formulated in all three equations given above; it is a well-known fact that Eq. (1) can be simplified based on the properties of the cosine-function; see also [4], and [5].

If, for $l = 0, 1, \ldots, M/2-1$, we let the values *SDS* (Sum/Difference Symmetry) reflect the symmetry due to the factor $(2k+1)$ in the cosine argument of Eq. (1), one can easily deduce that with the values $SDS_0[l;t] = S_{s\hat{b}}[l;t] + S_{s\hat{b}}[M-1-l;t]$ and $SDS_1[l;t] = S_{s\hat{b}}[l;t] - S_{s\hat{b}}[M-1-l;t]$, we can also use

$$\begin{aligned} V[i;t] &= \sum_{k=0}^{M/2-1} SDS_{(i \bmod 2)}[k;t]\, cos(\frac{\pi}{2M}(2k+1)(\frac{M}{2}+i)) \\ &\text{with } \{V[i;t] = -V[M-i;t],\ i = 0, 1, \ldots, M/2-1\}, \\ &\{V[M+i;t] = V[2M-i;t],\ i = 1, 2, \ldots, M/2-1\}, \\ &V[M/2;t] = 0, \quad \text{and} \\ V[3M/2;t] &= -\sum_{k=0}^{M/2-1} SDS_0[k;t]\,, \end{aligned} \quad (4)$$

to implement the subband matrixing procedure. These partially optimized matrixing formulas require only $(M-1)M/2 = 496$ multiplications and $M + (M/2-1)(M-1) + (M/2-1) = M^2/2 = 512$ additions to compute the $V[i;t]$ matrixing results. Please note that, as already indicated above, using these existing optimizations requires a considerable increase in the amount of computational control and sequencing logic (further optimization would increase this even more so).

Now, implementing the proposed zero-pute reduction and cut-off principles within Eq. (4) means that first we have to discriminate between two cases (with $M/2-1 = 15$):

1. $M \geq$ `alloc_cut` $> M/2$: in this case we can only realize an adaptation of the *SDS* computation in which we can omit some additions and subtractions; i.e. $SDS_0[m;t] = SDS_1[m;t] = S_{s\hat{b}}[m;t]$ for $m = 0, \ldots, M-$`alloc_cut` -1, and for $m =$ ($M-$`alloc_cut`$), \ldots, M/2-1$ nothing changes, i.e., we can use the original *SDS* formulas given above. Thus, the matrixing process then requires only $(M-1)M/2 = 496$ multiplications and $M - 2(M-$`alloc_cut`$) + (M/2-1)(M-1) + (M/2-1) = 448 + 2$ `alloc_cut` additions.

2. `alloc_cut` $\leq M/2$: first, note that the *SDS* values no longer need to be computed, i.e., $SDS_0[m;t] = SDS_1[m;t] = S_{s\hat{b}}[m;t]$. Next, we can adapt the formulas for $V[i;t]$ and $V[3M/2;t]$ to only iterate their index k from 0 to `alloc_cut` -1. This reduces the computational complexity to $(M-1)$`alloc_cut` $=$ 31 `alloc_cut` multiplications, and $(M-1)$(`alloc_cut` -1) + (`alloc_cut` -1) $= 32$(`alloc_cut` -1) additions.

Although these optimizations only yield a relatively small improvement when compared to the adaptation from Eq. (1) to Eq. (4), they are certainly still worth the minor increase in complexity they require, especially for the lower bit rate range.

Google Exhibit No. 1042
Page 003

Google Exhibit No. 1063
Page 1427

Table 2: Static-`sblimit` and average dynamic-μ[`alloc_cut`] subband cut-off index and computational complexity as a function of the bit rate (song: U2: "40", f_S = 44.1 kHz, 6123 frames, data shown here: channel 0 only, encoder used: lsf, psycho-acoustic model 2).

	`sblimit`	μ[`alloc_cut`]	average complex. ×		average complex. +	
			`sblimit`	`alloc_cut`	`sblimit`	`alloc_cut`
64 kbit/s	8	7,9	512	505,6	448	441,6
128 kbit/s	27	18,3	1728	1171,2	1664	1107,2
256 kbit/s	30	20,7	1920	1324,8	1856	1260,8

7. ADDITIONAL ADVANTAGES AND FUTURE WORK

Additionally, it is important to note that the cut-off optimizations discussed above always yield a scalable algorithm. If, e.g., the computational load would be too high to meet any real-time decoding deadlines, or if a system over-heating or low-battery condition would be detected, the subband cut-off parameter could be gradually adapted by forceably decrementing `alloc_cut`. The system would certainly still degrade down-to a level of unacceptable behavior, but it would do so using a near continuous and soft degradation curve that makes these algorithms an ideal solution for various situations, e.g. in portable devices.

As a final extension, one could also look for zero-allocated subbands within the subband index range [0, ..., `alloc_cut`] as determined above. Note that this approach would require another increase in complexity (due to an index-dependent computation process that can only be implemented efficiently within a more complex hardware-based solution). An interesting and probably even more popular extension can also be realized within layer III of the MPEG standard: we are currently studying how to apply the zero-pute reduction principle to the layer III aliasing, windowing and MDCT computation process.

8. CONCLUSIONS

In this paper we have discussed how using the seemingly simple "do not zero-pute" design principle, important optimizations and design alternatives can be studied or used to realize efficient MPEG-audio layer II decoders. We proposed both a static strategy that could be hard-coded into an actual decoder implementation and a dynamic or "signal adaptive approach" that can realize even better speed-up and/or power-saving results. Both strategies can be combined with a well-known and very important existing optimization strategy, and the scalability of the algorithm was pointed out. The results reported here are also useful for devising efficient layer III (MP3) subband signal synthesis schemes, but these and several other adaptations are currently being studied further. Finally, incremental to [7], a package containing the source code for implementing all techniques and formulas discussed above is freely available from the main author.

9. REFERENCES

[1] JPEG decoder source code from the Independent JPEG Group; `http://www.ijg.org/`

[2] W.-K. Paik and S.-Y. Hwang. Design of a novel synthesis filter for real-time MPEG-2 audio decoder implementation on a DSP chip. *IEEE Trans. on Consumer Electronics*, 45(4):26–28, February 1999.

[3] P. Noll. MPEG digital audio coding. *IEEE Signal Processing Magazine*, pages 59–81, September 1997.

[4] K. Konstantinides. Fast subband filtering in MPEG audio coding. *IEEE Signal Processing Letters*, 1(2):26–28, February 1994.

[5] P. D. Smet and I. Bruyland. Optimized recursive subband synthesis windowing for implementing efficient MPEG audio decoders. *IEEE Signal Processing Letters*, 10(10):303–306, October 2003.

[6] ISO/IEC 11172-3, Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to about 1.5Mbit/s, 1993.

[7] ISO dist10 source code package available in `./mpeg2/software/` at: `ftp://ftp.tnt.uni-hannover.de/pub/MPEG/audio/`

Google Exhibit No. 1042
Page 004

Google Exhibit No. 1063
Page 1428

Ex. 1042A

WorldCat®

2004 multimedia conference : MM'04

Author: Association for Computing Machinery

Print Book, English, 2004
Publisher: ACM Press, 2004

Physical Description:
1 volume

ISBN:
9781581138931, 1581138938

OCLC Number / Unique Identifier:
1063590195

Subjects:
ACM
MM
Multimedia

The British Library, On Demand
The British Library, Document Supply, Boston Spa
5,289 miles from your current location.
Boston Spa, Wetherby, West Yorkshire, LS23 7BQ, United Kingdom

Explore similar titles
We can make recommendations based on this item.
Yes, show me
Recommendations provided by a third-party AI service. Only title and author data is shared.

Google Exhibit No. 1063
Page 1430

FirstSearch

University Libraries

Libraries that Own Item

| Back to CUA Libraries

- This screen shows libraries that own the item you selected.

Title: 2004 multimedia conference : MM'04 Author: Association for Computing Machinery
Accession Number: 1063590195

Libraries with Item: "2004 multimedia conferenc..."(Record for Item | Get This Item)

Location	Library	Code
United Kingdom	BRITISH LIBR	BRI

Record for Item: "2004 multimedia conferenc..."(Libraries with Item)

GET THIS ITEM

Availability: Check the catalogs in your library.
- Libraries worldwide that own item: 1
- Library Holdings

External Resources:
- Find it@CU Find or Request It at CU
- ILL@CU Request it via ILL
- Cite This Item

FIND RELATED

More Like This: Advanced options ...

Find Items About: Association for Computing Machinery. (215)

Title: **2004 multimedia conference : MM'04.**

Corp Author(s): Association for Computing Machinery.

Publication: ACM Press,

Year: 2004

Description: 1 volume

Language: English

Standard No: ISBN: 1581138938 ((pbk.)); 9781581138931; National Library: 013134726

SUBJECT(S)

Identifier: Multimedia; MM; ACM

General Info: British Library not licensed to copy/ 0.

Material Type: Conference publication (cnp)

Google Exhibit No. 1063
Page 1431

Document Type: Book
Entry: 20050315
Update: 20241204
Accession No: OCLC: 1063590195
Database: WorldCat

E-mail Print Return Help Current database: WorldCat Total Libraries: 1 WorldCat

English | Español | Français | | | | | |
Options | Comments | Exit

©1992-2025 OCLC
Terms & Conditions
For information on how your personal data is processed when you use FirstSearch, please see your institution's privacy policy.

Google Exhibit No. 1063
Page 1432

Optimized Recursive Subband Synthesis Windowing for Implementing Efficient MPEG Audio Decoders

Patrick De Smet, *Student Member, IEEE,* and Ignace Bruyland, *Member, IEEE*

***Abstract*—In this letter, we discuss a recursive optimization approach for the subband synthesis windowing component used in all three layers of the MPEG-1 and -2 audio decoding frameworks. We explain the design of the proposed algorithm and briefly discuss how it can be combined with other novel and existing optimization strategies. The impact of these optimizations on the number of multiplications and additions is shown, and the influence of the algorithm on the amount of memory required by the adapted decoder is also discussed. We show that the proposed algorithm offers very interesting possibilities for implementing efficient VLSI designs.**

***Index Terms*—Efficient decoder, MPEG audio, recursive windowing.**

I. Introduction

In the last ten years, the ISO/IEC MPEG-1 and -2 audio encoding standards have been gaining considerable importance and popularity as efficient tools for lossy compression of high-quality digital audio [1], [2]. This letter aims at contributing other, previously undocumented possibilities for obtaining faster and/or more cost-effective MPEG audio decoders. We mainly focus on realizing more efficient VLSI designs by reducing the number of multiplications; compared to adder components, multipliers are relatively slow and often result in considerably more demanding design restrictions [3]. The proposed algorithm is based upon the optimization of redundant computations carried out in subsequent synthesis filtering operations and also extends and complements techniques discussed in [4]–[6] (encoding only), [7]–[9] (decoding only), and [10]–[12] (both).

II. MPEG Audio Decoding and Subband Synthesis

Within the MPEG-1 and -2 audio encoding standards, ISO/IEC has defined three different "layers" [1], [13]. Each layer builds upon its lower ranked siblings, and for each higher layer number, (implementation) complexity and compression efficiency typically both increase. The cosine-modulated polyphase filter bank used by all three layers within the framework maps the digital audio input into 32 equally spaced and critically decimated bandpass signals [2]. Compared to the encoder, the computational complexity of the MPEG audio decoder is significantly lower; most of the computation time is due to the synthesis filter bank [7], [9].

For each multirate filtering step at time t, the ISO synthesis subband filtering algorithm starts with $M = 32$ subband samples extracted from the encoded bitstream syntax, i.e., $S_{sb}[i;t]$, $i = 0, 1, \ldots, M-1$, and produces M pulse code modulation (PCM) audio output samples; $S_{\mathrm{pcm}}[j;t]$, $j = 0, 1, \ldots, M-1$.

First, the intermediate inverse polyphase quadrature mirror filter bank (IPQMF) vector $V[i;t]$ is shifted over $2M$ memory positions according to $V[i;t]$=$V[i-2M;\, t-1]$, for $i = 2PM-1, 2PM-2, \ldots, 2M$ with $P = N/M = 16$ and $N = 512$ the length of the prototype filter window [2]. Then, the basic MPEG audio matrixing operation [13] is calculated

$$V[i;t] = \sum_{k=0}^{M-1} S_{sb}[k;t] \cos\left(\frac{\pi(2k+1)\left(\frac{M}{2}+i\right)}{2M}\right) \quad (1)$$

for i=$0, 1, \ldots, 2M-1$. An N-value intermediate vector $U[i;t]$ is then built using $U[2Mk+l;t]$=$V[4Mk+l;t]$, and $U[2Mk+M+l;t]$=$V[4Mk+3M+l;t]$, $k = 0, 1, \ldots, P/2-1$, $l = 0, 1, \ldots, M-1$. Finally, the so-called windowing component computes M new PCM audio output samples, i.e., $S_{\mathrm{pcm}}[j;t] = \sum_{i=0}^{P-1} U[j+Mi;t]\; D[j+Mi]$, j=$0, 1, \ldots, M-1$, in which the values D are the N synthesis window coefficients (see [13, Table B.3]).

This algorithm requires $2M^2 + MP = 2560$ multiplications and $2M(M-1) + M(P-1) = 2464$ additions, and, not counting the PCM and subband sample buffers, at least $3N + M = 1568$ datawords of memory; $2PM = 2N = 1024$ datawords V, $M = 32$ cosine-coefficients (contrary to [15], using all 2048 cosine-coefficients in (1), we assume appropriate address calculators always being implemented), and $N = 512$ window coefficients D.

III. Matrixing Optimizations

Obviously, (1) can be simplified based on the properties of the cosine-function. This has been discussed in [7], [9], and [12], yielding a fast odd/even butterfly M-point inverse discrete cosine tranform decomposition (see also [14]). However, except for the fact that, for $l = 0, 1, \ldots, M/2-1$

$$\begin{aligned}
\mathrm{SDS}_0[l;t] &= S_{sb}[l;t] + S_{sb}[M-1-l;t] \\
\mathrm{SDS}_1[l;t] &= S_{sb}[l;t] - S_{sb}[M-1-l;t] \\
V[i;t] &= \sum_{k=0}^{M/2-1} \mathrm{SDS}_{(i \bmod 2)}[k;t] \cos\left(\frac{\pi(2k+1)\left(\frac{M}{2}+i\right)}{2M}\right)
\end{aligned}$$

Manuscript received December 18, 2001; revised November 15, 2002. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Antonio Ortega.

The authors are with the Department of Telecommunications and Information Processing (TW07), Ghent University, B-9000 Ghent, Belgium (e-mail: pds@telin.UGent.be; ib@telin.UGent.be).

Digital Object Identifier 10.1109/LSP.2003.817173

1070-9908/03$17.00 © 2003 IEEE

Authorized licensed use limited to: Reprints Desk. Downloaded on December 23,2024 at 19:58:44 UTC from IEEE Xplore. Restrictions apply.

Google Exhibit No. 1043
Page 001

Google Exhibit No. 1063
Page 1433

with

$$\left\{V[i;t] = -V[M-i;t], \qquad i = 0, 1, \ldots, \frac{M}{2} - 1\right\}$$
$$\left\{V[M+i;t] = V[2M-i;t], \qquad i = 1, 2, \ldots, \frac{M}{2} - 1\right\}$$
$$V[M/2;t] = 0 \quad V[3M/2;t] = -\sum_{k=0}^{M/2-1} \mathrm{SDS}_0[k;t] \tag{2}$$

we do not repeat this discussion here; we will reduce the computational complexity by pursuing other possibilities. The values for sum/difference-symmetry (SDS) reflect the symmetry due to the factor $(2k+1)$ in the cosine argument of (1).

The partially optimized matrixing formulas given above require only $(M-1)M/2 = 496$ multiplications and $M+(M/2-1)(M-1)+(M/2-1) = M^2/2 = 512$ additions to compute the $V[i;t]$ matrixing results. If needed, these equations can also reduce the required amount of memory for the entire filtering procedure to $2N+2M = 1088$ datawords; $M = 32$ values SDS_0 and SDS_1, $PM = N = 512$ remaining values V, $M = 32$ cosine-coefficients, and $N = 512$ window coefficients D.

IV. Recursive Windowing Optimizations

For VLSI design, multiplications are considered significantly more expensive than additions; the required multiplier speed can often influence the choice of the entire implementation design (see the *en*coder architecture trade-offs discussed in [3] and [4]). Hence, in this section we focus on reducing the number of multiplications in the windowing component.

Now, since

$$S_{\mathrm{pcm}}[j;t] = \sum_{i=0}^{P/2-1} \left(U[j+Mi;t]\, D[j+Mi]\right) + \left(D\left[j+Mi+\frac{MP}{2}\right]\, U\left[j+Mi+\frac{MP}{2};t\right]\right)$$

and $ab+cd = (a+c)(b+d) - ad - bc$, we have

$$S_{\mathrm{pcm}}[j;t] = \sum_{i=0}^{P/2-1} \left(U[j+Mi;t] + D\left[j+Mi+\frac{MP}{2}\right]\right) \cdot \left(D[j+Mi] + U\left[j+Mi+\frac{MP}{2};t\right]\right) - \sum_{i=0}^{P/2-1} U[j+Mi;t]\, U\left[j+Mi+\frac{MP}{2};t\right] - \sum_{i=0}^{P/2-1} D[j+Mi]\, D\left[j+Mi+\frac{MP}{2}\right] \tag{3}$$

for $j = 0, 1, \ldots, M-1$. Rewriting this sum as $S_{\mathrm{pcm}}[j;t] = A[j;t] - B[j;t] - C[j]$, the last sum $C[j]$ is an M-dimensional constant vector which can be precalculated for each j. Below, each of these terms will be discussed in more detail, but first let us consider some more general issues. The A terms will remain to be calculated by the final algorithm, but the B terms can be simplified further by splitting them into two sums for even and odd multiples of M for the index i, and by replacing U by V (see Section II). This yields

$$B[j;t] = \sum_{i=0}^{P/4-1} \big(V[j+4Mi;t]\, V[j+4Mi+MP;t] + V[j+4Mi+3M;t]\, V[j+4Mi+MP+3M;t]\big). \tag{4}$$

As already indicated, t is an external clocking signal driving the subband synthesis filter; for each filter calculation at time t, $2M$ new values $V[l;t]$ are determined according to (1) or (2). Considering the previous calculations of the subband synthesis filter, it is clear that $V[l+2Mn;t]=V[l;t-n]$, for $l = 0, 1, \ldots, (2PM-1-2Mn)$, or equivalently, $V[l;t]=V[l-2Mn;t-n]$, $l = (2Mn), \ldots, 2PM-2, 2PM-1$. Hence

$$B[j;t] = B[j;(t-2)] - d[j;(t-8)] + d[j;t]$$

with

$$d[j;t] = V[j;t]\, V[j+MP;t] + V[j+3M;t]\, V[j+3M+MP;t] \tag{5}$$

for $j = 0, 1, \ldots, M-1$. This result can be easily obtained from (4) by removing the last terms for $i = P/4-1 = 3$ and adding new terms for $i = 0$, and shifting over two clocking periods t or $4M = 128$ positions.

The equations in this section yield the proposed recursive algorithm for calculating the windowing component. To our knowledge this decoder structure has not been proposed before. The recursive computation of $B[j;t]$ allows the number of multiplications to be reduced by recycling data computed in previous filtering operations [using two separate delay lines for B and d; see (5)]. Below, we will discuss the memory requirements and the computational complexity of the proposed algorithm, but first let us discuss some additional possibilities for further optimization.

V. Additional Optimizations

A. Detailed Optimizations for the Recursive Algorithm

From the equations given above, it is clear that $V[M/2;t] = V[M/2+MP;t] = 0$; and hence, $d[M/2;t] = V[M/2+3M;t]\, V[M/2+MP+3M;t]$, using only one multiplication to compute $d[M/2;t]$. The value $d[0;t]$ can be calculated by using two multiplications and one addition. Finally, based on the properties of V [see (2)] and the definition of d given above, it can be easily shown that $d[j;t] = d[M-j;t]$, $j=1, 2, \ldots, M/2-1$. Hence, the number of multiplications and additions for calculating all $d[j;t]$ can be halved. Additionally, this also means that $B[j;t]=B[M-j;t]$, $j=1, 2, \ldots, M/2-1$, and that $B[0;t]$ and $B[M/2;t]$ are unique.

Authorized licensed use limited to: Reprints Desk. Downloaded on December 23,2024 at 19:58:44 UTC from IEEE Xplore. Restrictions apply.

Google Exhibit No. 1043
Page 002

Google Exhibit No. 1063
Page 1434

B. Windowing Symmetry

Another possibility for further optimization is given by the fact that $\{D[0]=0\}$

$$\left\{ D\left[\frac{N}{2}+k\right] = -D\left[\frac{N}{2}-k\right] \quad \text{for } k\neq 2M, 4M, 6M,\ k=1,2,\ldots,\frac{N}{2}-1 \right\}$$

and $\{D[k] = D[N-k];\ k=2M, 4M, 6M\}$, and $D[N/2]$ are unique. To our knowledge, these properties have not received any real attention in the existing literature. This seems odd, since clearly, exploiting this symmetry will immediately yield a significant advantage in the amount of memory needed to store these coefficients. Also, for $j = 1, 2, \ldots, M/2 - 1$, from (3) and with $PM = N$ (see above), we get (6) (shown at the bottom of the page). Assuming now that $i=P/2-1-l$ in each parallel computation step for a fixed $j=1,2,\ldots,M/2-1$, and exploiting the symmetry of the window coefficients, i.e., $D[j+Mi]= -D[M-j+Ml+N/2]$ and $-D[M-j+Ml]=D[j+Mi+N/2]$, half of the window D suffices to compute all values A. Also, note that the parallel computations in (6) can be used to avoid *re*loading the coefficients D (from external memory) into internal registers. Next, by considering that $D[0] = 0$ and $U[M/2+2Mn;t] = V[M/2+4Mn;t] = V[M/2;t-2n] = 0$; it is clear that

$$A[0;t] = \left(U[0;t] + D\left[\frac{N}{2}\right]\right)\ U\left[\frac{N}{2};t\right] + \sum_{i=1}^{P/2-1} \left(U[Mi;t] + (-1)^i \cdot D\left[\frac{N}{2}-Mi\right]\right)\left(D[Mi] + U\left[Mi+\frac{N}{2};t\right]\right)$$

and

$$A\left[\frac{M}{2};t\right] = A_{M/2} + \sum_{i=0}^{P/4-1} \left(U\left[\frac{M}{2}+M+2Mi;t\right] - D\left[\frac{N}{2}-\frac{M}{2}-M-2Mi\right]\right) \cdot \left(D\left[\frac{M}{2}+M+2Mi\right] + U\left[\frac{N}{2}+M+\frac{M}{2}+2Mi;t\right]\right)$$

with the constant $A_{M/2} = -\sum_{i=0}^{P/4-1} D[N/2 - M/2 - 2Mi]\ D[M/2 + 2Mi]$. Finally, for the constants C it can be shown that $C[j] = C[M-j] = \sum_{i=0}^{P/2-1} D[j+Mi]\ (-D[N/2-j-Mi])$, for $j=1,2,\ldots,M/2-1$. Additionally, $C[M/2] = \sum_{i=0}^{P/2-1} D[M/2+Mi]\ (-D[N-M/2-Mi])$, and $C[0] = (D[MP/4])^2 + 2\sum_{i=1}^{P/4-1}(-1)^i D[Mi]\ D[N/2-Mi]$. These equations can aid in a faster (re)start of a decoder when the C-values are not hard-coded into the actual application or the ROM of a device.

VI. Final Optimized Algorithm

Initially, the constants $A_{M/2}$, $C[0]$, $C[M/2]$, and $C[j]$, $j=1,2,\ldots,M/2-1$, need to be available. All other values are initialized to zero but both $B[j;t=0]$ and $B[j;t=1]$ are set to $C[j]$ for $j = 0, 1, \ldots, M/2$. This absorbs the values C into the recursive Bs, simplifying the computation of the PCM output samples $S_{\text{pcm}}[j;t]$. After extracting the M encoded subband samples $S_{sb}[j;t]$ from the bitstream, the matrixing operations are computed by using (2). Next, $d[j;t]$ is computed according to the definition given in Section IV and the symmetry issues discussed in Section V–A. This enables the computation of all $B[j;t]$, again according to Sections V–A and IV. Then, in a single loop, $A[j;t]$ and $A[M-j;t]$, and next, $A[0;t]$ and $A[M/2;t]$ can be computed using the equations formulated above. Finally, we can compute $S_{\text{pcm}}[j;t]=A[j;t]-B[j;t]$ for outputing samples.

Adapted source code for the ISO codec [15] is freely available from the authors.[1]

VII. Computational Complexity

A. Recursive Synthesis Windowing

The original windowing component as described in Section II requires MP=512 multiplications and $M(P-1)$=480 additions for calculation of the PCM samples.

In the proposed recursive algorithm, the computation of the values $A[j;t]$, requires $(M-2)P/2 = 240$ multiplications and $2(M-2)(P/2) + (M-2)(P/2-1) = 690$ additions for $j = 1, 2, \ldots, M/2-1, M/2+1, M/2+2, \ldots, M-1$; see (6). The computation of $A[0;t]$ requires $1 + (P/2-1) = 8$ multiplications and $1 + 2(P/2-1) + (P/2-1) = 22$ additions. For the calculation of $A[M/2;t]$ we need $P/4 = 4$ multiplications and $2(P/4) + P/4 = 12$ additions. This makes a total of 252 multiplications and 724 additions. The computation of all M values d requires $1 + 2 + 2(M/2-1) = 33$ multiplications and $1 + (M/2-1) = 16$ additions. Similarly, exploiting the

[1]See also http://telin.UGent.be/~pds/papers/SPL2003/.

$$A[j;t] = \sum_{i=0}^{P/2-1} \left(U[j+Mi;t] + D\left[j+Mi+\frac{N}{2}\right]\right)\left(D[j+Mi] + U\left[j+Mi+\frac{N}{2};t\right]\right)$$

$$A[M-j;t] = \sum_{l=-P/2-1}^{0} \left(U[M-j+Ml;t] + D\left[M-j+Ml+\frac{N}{2}\right]\right)\left(D[M-j+Ml] + U\left[M-j+Ml+\frac{N}{2};t\right]\right) \quad (6)$$

Authorized licensed use limited to: Reprints Desk. Downloaded on December 23,2024 at 19:58:44 UTC from IEEE Xplore. Restrictions apply.

Google Exhibit No. 1043
Page 003

Google Exhibit No. 1063
Page 1435

symmetry in B (see also Section V–A), only $2(M/2+1) = 34$ additions are needed for the computation of all values B. Finally, for the calculation of the PCM output samples, adding A and B (with C absorbed into B) requires $M = 32$ additions. Hence, 285 multiplications and 806 additions are needed to implement the optimized recursive windowing algorithm. The proposed algorithm results in an increase in the required number of additions. However, the number of multiplications has been reduced significantly; as already mentioned, this will normally put considerably less restricting demands on (the speed of) a VLSI or hardware architecture [3].

B. Full Subband Synthesis Filtering

In summary, when compared to the full basic ISO subband synthesis filter as discussed in Section II, the computational complexity is reduced from 2560 multiplications and 2464 additions (see Section II), to 781 multiplications and 1318 additions (see Section III: 496 multiplications and 512 additions; and, Section VII-A: 285 multiplications and 806 additions).

VIII. Memory Requirements

A. Recursive Synthesis Windowing

Implementing the proposed algorithm requires $8(1+M/2)=136$ datawords for $1 + M/2 = 17$ values $d[j;t]$ with a maximum time delay of eight synthesis operations, $2(1 + M/2) = 34$ datawords for $B[j;t]$ due to a buffer of 2 filtering operation delays, one dataword for $A_{M/2}$, and the $1 + M/2 = 17$ constants C. However, as indicated above, these constants C are absorbed into the recursive computation of B, and they can be easily computed when the decoder is being initialized; the constants C are not needed during "steady state" operation of the decoder. Thus, the recursive algorithm can be implemented with a total of $10(1 + M/2) + 1 = 5M + 11 = 171$ additional datawords. Fortunately, by also exploiting the symmetry in the coefficients D (see Section V–B), we can reclaim $N/2 = 256$ datawords. Hence, compared to the traditional implementations described in the existing literature, the overall balance of all the optimizations we proposed in this letter is a gain of $N/2 - (5M + 11) = 85$ datawords.

B. Full Subband Synthesis Filtering

In this letter, we discussed and proposed the memory-related optimization of the MPEG matrixing operations (see Section III), the auxiliary recursive data (see Section V–A), and the window coefficients (see Section V–B). After implementing all these optimizations, the required amount of memory can be reduced from $3N+M=1568$ datawords (see Section II) to $(2N+2M)-(N/2-5M-11)=1003$ datawords; see Sections III and VIII-A.

Note that not all symmetry properties must be implemented. This results in several flexible trade-off possibilities for exchanging address calculation logic and control complexity for memory requirements and vice versa.

IX. Conclusion

In this letter, we proposed a recursive windowing algorithm useful for implementing efficient MPEG-1 and -2 audio subband synthesis filters. The proposed windowing algorithm reduces the required number of multiplications from 512 to 285, while the number of additions changes from 480 to 806. This approach can be very profitable when a certain hardware design has some limitations on (the speed of) its multiplier(s). Additionally, we proposed and discussed several memory related requirements and optimizations. In combination with an existing and well-known cosine-symmetry exploiting optimization step, this letter simplifies the original ISO subband synthesis filtering procedure from 2560 multiplications and 2464 additions to 781 multiplications and 1318 additions, while reducing the required amount of memory from 1568 to 1003 datawords.

References

[1] P. Noll, "MPEG digital audio coding," *IEEE Signal Processing Mag.*, pp. 59–81, Sept. 1997.
[2] M. Kahrs and K. Brandenburg, Eds., *Applications of Digital Signal Processing to Audio and Acoustics*. Dordrecht, The Netherlands: Kluwer, 1998.
[3] N. Ling and C.-W. Shih, "Two polyphase filter architectures for MPEG audio," in *Proc. 40th Midwest Symp. Circuits and Systems*, vol. 2, 1997, pp. 861–864.
[4] D.-Y. Chan, J.-F. Yang, and C.-C. Fang, "Fast implementation of MPEG audio coder using recursive formula with fast discrete cosine transforms," *IEEE Trans. Speech Audio Processing*, vol. 4, pp. 144–148, Mar. 1996.
[5] M. Kumar and M. Zubair, "A high performance software implementation of MPEG audio encoder," *Proc. ICASSP*, vol. 2, pp. 1049–1052, 1996.
[6] H.-O. Oh, S.-Y. Kim, D.-H. Youn, and I.-W. Cha, "New implementation techniques of a real-time MPEG-2 audio encoding system," *Proc. ICASSP*, vol. 4, pp. 2287–2290, 1999.
[7] T.-H. Tsai, T.-H. Chen, and L.-G. Chen, "An MPEG audio decoder chip," *IEEE Trans. Consumer Electron.*, vol. 41, pp. 89–96, Feb. 1995.
[8] P. De Smet, "On two new techniques for optimized MPEG audio decoding," in *Proc. IASTED Int. Conf. SPPRA 2001*, July 3–6, 2001, pp. 79–83.
[9] W.-K. Paik and S.-Y. Hwang, "Design of a novel synthesis filter for real-time MPEG-2 audio decoder implementation on a DSP chip," *IEEE Trans. Consumer Electron.*, vol. 45, pp. 1119–1129, Nov. 1999.
[10] C. D. Murphy and L. Anandakumar, "Real-time MPEG-1 audio coding and decoding on a DSP chip," *IEEE Trans. Consumer Electron.*, vol. 43, pp. 40–47, Feb. 1997.
[11] V. Britanak and K. R. Rao, "An efficient implementation of the forward and inverse MDCT in MPEG audio coding," *IEEE Signal Processing Lett.*, vol. 8, pp. 48–51, Feb. 2001.
[12] K. Konstantinides, "Fast subband filtering in MPEG audio coding," *IEEE Signal Processing Lett.*, vol. 1, pp. 26–28, Feb. 1994.
[13] ISO/IEC, "Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s,", ISO/IEC 11 172–3, 1993.
[14] B. G. Lee, "A new algorithm to compute the discrete cosine transform," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-32, pp. 1243–1245, Dec. 1984.
[15] ISO dist10 Source Code Package [Online]. Available: ./mpeg2/software/ at ftp://ftp.tnt.uni-hannover.de/pub/MPEG/audio/.

Authorized licensed use limited to: Reprints Desk. Downloaded on December 23,2024 at 19:58:44 UTC from IEEE Xplore. Restrictions apply.

Google Exhibit No. 1043
Page 004

Google Exhibit No. 1063
Page 1436

Ex. 1043A

LIBRARY OF CONGRESS
0 011 577 688 5

IEEE SIGNAL PROCESSING LETTERS 10 2003

TK 5102
.9
.I32

IEEE

SIGNAL PROCESSING LETTERS

A PUBLICATION OF THE IEEE SIGNAL PROCESSING SOCIETY

IEEE Signal Processing Society

www.ieee.org/sp/index.html

OCTOBER 2003 VOLUME 10 NUMBER 10 ISPLEM (ISSN 1070-9908)

LETTERS

Design and Implementation of Digital Signal Processing (DSP) Systems

Efficient Computation of Discrete Polynomial Transforms . . . . . *M. F. Aburdene, J. Xie, and R. J. Kozick* 285

Image and Multidimensional Signal Processing

Image Enhancement Using a Contrast Measure in the Compressed Domain . . . . . *J. Tang, E. Peli, and S. Acton* 289

High-Performance Embedded Morphological Wavelet Coding . . . . . *F. Lazzaroni, R. Leonardi, and A. Signoroni* 293

Postprocessing of JPEG-2000 Images to Remove Compression Artifacts . . . . . *A. Nosratinia* 296

A Particle-Filter-Based Detection Scheme . . . . . *Y. Boers and H. Driessen* 300

Multimedia Signal Processing

Optimized Recursive Subband Synthesis Windowing for Implementing Efficient MPEG Audio Decoders . . . . . *P. De Smet and I. Bruyland* 303

Signal Processing Theory and Methods

An Efficient Design of a Variable Fractional Delay Filter Using a First-Order Differentiator. . . *S.-C. Pei and C.-C. Tseng* 307

Detection of Explosive Lung and Bowel Sounds by Means of Fractal Dimension . . . . . *L. J. Hadjileontiadis and I. T. Rekanos* 311

Information for Authors 315

SPL EDICS List 316

IEEE

Google Exhibit No. 1063
Page 1439

IEEE SIGNAL PROCESSING SOCIETY

The Signal Processing Society is an organization, within the framework of the IEEE, of members with principal professional interest in the technology of transmission, recording, reproduction, processing, and measurement of speech and other signals by digital electronic, electrical, acoustic, mechanical, and optical means, the components and systems to accomplish these and related aims, and the environmental, psychological, and physiological factors concerned therewith. All members of the IEEE are eligible for membership in the Society and will receive this publication upon payment of the annual Society membership fee of $24.00 plus an annual subscription fee of $20.00. For information on joining, write to the IEEE at the address below. *Member copies of Transactions/Journals are for personal use only.*

SOCIETY PUBLICATIONS

Publications Board Chair
A. NEHORAI, *VP Publications*
Univ. of Illinois at Chicago
Chicago, IL 60607-7053

Signal Processing Letters
A.-J. VAN DER VEEN, *Editor-in-Chief*
Delft University of Technology
Delft, The Netherlands

Transactions on Signal Processing
A. H. SAYED, *Editor-in-Chief*
Univ. of California
Los Angeles, CA 90095-0001

SP Magazine
K. J. RAY LIU, *Editor-in-Chief*
Univ. of Maryland
College Park, MD 20742

Transactions on Image Processing
J. KOVAČEVIĆ, *Editor-in-Chief*
Carnegie Mellon Univ.
Pittsburgh, PA 15213

Trans. on SP, IP, SAP
SPS Publications Office
IEEE Signal Processing Society
Piscataway, NJ 08855

Trans. Speech & Audio Processing
I. TRANCOSO, *Editor-in-Chief*
INESC ID
Lisbon, Portugal

Signal Processing Letters
SPS Publications Office
IEEE Signal Processing Society
Piscataway, NJ 08855

ASSOCIATE EDITORS

ALEX ACERO
Microsoft Research
Redmond, WA 98052-6399

SCOTT T. ACTON
Univ. Virginia
Charlottesville, VA 22904

JOHN M. ANDERSON
Univ. Florida
Gainesville, FL 32611

AMIR ASIF
Technical Univ.
British Columbia
Surrey, B.C. V3T-2W1, Canada

VICTOR A. N. BARROSO
IST, 1096 Lisbon
Portugal

MARCELO G. S. BRUNO
Instituto Tecnológico
de Aeronautica
Sao Jose Dos Campos SP
CEP 12228-900, Brazil

JONATHON A. CHAMBERS
Center for Digital
Signal Process. Res.
King's College London,
Strand Campus
London WC2R 2LS, U.K.

P. C. CHING
Chinese Univ. Hong Kong
Shatin, NT, Hong Kong

PAMELA COSMAN
Univ. of California, San Diego
La Jolla, CA 92093-0407

RICARDO L. DE QUEIROZ
Xerox Corp.
Webster, NY 14580

ZHI DING
Univ. California
Davis, CA 95616

STEVEN L. GAY
Bell Labs,
Murray Hill, NJ 07974

MOUNIR GHOGHO
Univ. Leeds
Leeds LS2 9JT, U.K.

PAULO GONÇALVES
INRIA Rhone-Alpes
France

MARK A. HASEGAWA-JOHNSON
Univ. Illinois
Urbana, IL 61801

CHIANG HE
Scansoft, Inc.
Waltham, MA 02451

YU-HEN HU
Univ. Wisconsin-Madison
Madison, WI 53706

YITENG HUANG
Bell Labs,
Murray Hill, NJ 07974

TON A. C. M. KALKER
Philips Research Labs Eindhoven
NL-5656 Eindhoven
The Netherlands

RICHARD C. KAVANAGH
Univ. College, Cork
Cork, Ireland

STEVEN KAY
Univ. Rhode Island
Kingston, RI 02881

W. BASTIAAN KLEIJN
KTH Royal Inst. of Technol.
Stockholm 10044, Sweden

VISA KOIVUNEN
Helsinki Univ. Technol.
FIN-02015 Hut, Finland

JANUSZ KONRAD
Boston Univ.
Boston, MA 02215

GEERT LEUS
K. U. Leuven
Heverlee, Belgium

ELIAS MASOLAKOS
Northeastern Univ.
Boston MA 02115

DIPTI P. MUKHERJEE
Indian Stat. Inst.
Calcutta 700108
India

V. H. NASCIMENTO
Univ. Sao Paulo
Sao Paulo, Brazil

DIMITRIS A. PADOS
State Univ. New York at Buffalo
Buffalo, NY 14260

A. PAPANDREOU-SUPPAPPOLA
Arizona State Univ.
Tempe, AZ 85287-7206

RAFFAELE PARISI
INFO-COM Rome 00184,
Italy

SOO-CHANG PEI
National Taiwan Univ.
Taipei, Taiwan, R.O.C. 106

SEE-MAY PHOONG
National Taiwan Univ.
Taipei, Taiwan, R.O.C. 106

MARKUS PUESCHEL
Carnegie Mellon Univ.
Pittsburgh, PA 15213

ERCHIN SERPEDIN
Texas A&M Univ.
College Station, TX 77843-3128

PETR TICHAVSKY
Inst. Inform. Theory Automat.
Czech Republic

DARREN B. WARD
Imperial College Sci.,
Technol., Med.
London SW7 2BT, U.K.

XIANG-GEN XIA
Univ. Delaware
Newark, DE 19716

XI ZHANG
Nagaoka Univ. Technology
Niigata 940-2188 Japan

THE INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS, INC.

Officers

MICHAEL S. ADLER, *President*
ARTHUR W. WINSTON, *President-Elect*
LEVENT ONURAL, *Secretary*
PEDRO A. RAY, *Treasurer*
RAYMOND D. FINDLAY, *Past President*
JAMES M. TIEN, *Vice President, Educational Activities*
MICHAEL R. LIGHTNER, *Vice President, Publication Services and Products*
W. CLEON ANDERSON, *Vice President, Regional Activities*
GERALD H. PETERSON, *President, IEEE Standards Association*
LEAH H. JAMIESON, *Vice President, Technical Activities*
JAMES V. LEONARD, *President, IEEE-USA*
JOHN R. VIG, *Director, Division IX—Signals and Applications*

Executive Staff

DANIEL J. SENESE, *Executive Director*
DONALD CURTIS, *Human Resources*
ANTHONY DURNIAK, *Publications Activities*
JUDITH GORMAN, *Standards Activities*
CECELIA JANKOWSKI, *Regional Activities*
BARBARA COBURN STOLER, *Educational Activities*
MATTHEW LOEB, *Corporate Strategy & Communications*
RICHARD D. SCHWARTZ, *Business Administration*
W. THOMAS SUTTLE, *IEEE-USA*
MARY WARD-CALLAN, *Technical Activities*

IEEE Periodicals
Transactions/Journals Department

Staff Director: FRAN ZAPPULLA
Editorial Director: DAWN MELLEY
Production Director: ROBERT SMREK
Managing Editor: MARTIN J. MORAHAN
Associate Editor: GEORGE CRISCIONE

IEEE SIGNAL PROCESSING LETTERS (ISSN 1070-9908) is published monthly by the Institute of Electrical and Electronics Engineers, Inc. Responsibility for the contents rests upon the authors and not upon the IEEE, the Society/Council, or its members. **IEEE Corporate Office:** 3 Park Avenue, 17th Floor, New York, NY 10016-5997. **IEEE Operations Center:** 445 Hoes Lane, P.O. Box 1331, Piscataway, NJ 08855-1331. **NJ Telephone:** +1 732 981 0060. **Price/Publication Information:** Individual copies: IEEE Members $10.00 (first copy only), nonmembers $20.00 per copy. (Note: Add $4.00 postage and handling charge to any order from $1.00 to $50.00, including prepaid orders.) Member and nonmember subscription prices available upon request. Available in microfiche and microfilm.

Copyright and Reprint Permissions: Abstracting is permitted with credit to the source. Libraries are permitted to photocopy for private use of patrons, provided the per-copy fee indicated in the code at the bottom of the first page is paid through the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923. For all other copying, reprint, or republication permission, write to Copyrights and Permissions Department, IEEE Publications Administration, 445 Hoes Lane, P.O. Box 1331, Piscataway, NJ 08855-1331. Copyright © 2003 by the Institute of Electrical and Electronics Engineers, Inc. All rights reserved. Periodicals Postage Paid at New York, NY and at additional mailing offices. **Postmaster:** Send address changes to IEEE SIGNAL PROCESSING LETTERS, IEEE, 445 Hoes Lane, P.O. Box 1331, Piscataway, NJ 08855-1331. GST Registration No. 125634188. Printed in U.S.A.

Efficient Computation of Discrete Polynomial Transforms

LIBRARY OF CONGRESS NOV 10 2003 COPYRIGHT OFFICE

Maurice F. Aburdene, *Fellow, IEEE*, Jin Xie, *Student Member, IEEE*, and Richard J. Kozick, *Member, IEEE*

***Abstract*—This letter presents a new recursive method for computing discrete polynomial transforms. The method is shown for forward and inverse transforms of the Hermite, binomial, and Laguerre transforms. The recursive flow diagrams require only $2N$ additions, $2(N+1)$ memory units, and $N+1$ multipliers for the $N+1$-point Hermite and binomial transforms. The recursive flow diagram for the $N+1$-point Laguerre transform requires $2N$ additions, $2(N+1)$ memory units, and $2(N+1)$ multipliers. The transform computation time for all of these transforms is $O(N)$.**

***Index Terms*—Binomial transform, discrete polynomial transform, flow diagram, Hermite transform, Laguerre transform, signal processing algorithms.**

I. Introduction

THE DISCRETE Hermite, binomial, and Laguerre transforms are used in control, signal processing, and communications [1], [2]. This letter presents a recursive method based on an improvement over the method presented in [3] for computing the forward and inverse discrete polynomial transforms. In addition, we present both the time performance and the hardware complexity analysis for the recursive method.

The transforms are based on the discrete forms of the Hermite [4], binomial [4], and Laguerre [5] polynomials, which are defined as follows:

$$\text{Hermite}: p_i(z) = \sum_{l=0}^{i} (-2)^l \binom{i}{l} \frac{z^{(l)}}{N^{(l)}} \tag{1}$$

$$\text{binomial}: p_i(z) = \binom{N}{z} \sum_{l=0}^{i} (-2)^l \binom{i}{l} \frac{z^{(l)}}{N^{(l)}} \tag{2}$$

$$\text{Laguerre}: p_i(z) = \theta^i \sum_{l=0}^{i} (-1)^l \binom{i}{l} \left(\frac{1-\theta}{\theta}\right)^l \binom{z}{l} \tag{3}$$

where $|\theta| < 1$ is a parameter of the Laguerre polynomial, $z = 0, 1, \ldots, N$, and i is the order of the polynomial with $i = 0, 1, \ldots, N$. In addition, we have

$$\binom{a}{b} = a!/(a-b)!b!$$

and $z^{(l)}$ is the backward factorial function

$$z^{(l)} = z(z-1)\cdots(z-l+1).$$

Manuscript received November 11, 2001; revised October 9, 2002. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Elias S. Manolakos.

The authors are with the Electrical Engineering Department, Bucknell University, Lewisburg, PA 17837 USA (e-mail: aburdene@bucknell.edu; kozick@bucknell.edu).

Digital Object Identifier 10.1109/LSP.2003.817177

The $(N+1)\times(N+1)$ transform matrices for these discrete polynomials are denoted by H_N, B_N, and L_N, respectively. The (i,z) element of each transform matrix is $p_i(z)$. It was shown in [3] that these matrices can be computed using the following definitions, where $i, j = 0, 1, \ldots, N$:

$$\begin{aligned}
[B]_{i,j} &= \binom{j}{i}, \text{ for } i \le j \text{ and } 0 \text{ for } i > j \\
[I^-]_{j,j} &= (-1)^j, \text{ and } 0 \text{ elsewhere} \\
[I^{\sim}]_{j,N-j} &= 1, \text{ and } 0 \text{ elsewhere} \\
[Q_\theta]_{j,j} &= \theta^j, \text{ and } 0 \text{ elsewhere.}
\end{aligned} \tag{4}$$

The Hermite, binomial, and Laguerre transform matrices can be formed by products of matrices as shown in [3]

$$H_N = [I^- Q_2 B]^T [\text{Diag}[I^{\sim}(B^T B)]]^{-1} B \tag{5}$$

$$B_N = [I^- Q_2 B]^T B \tag{6}$$

$$L_N = Q_\theta [I^- \text{Diag}[[I^- B]^T \Theta^T] Q_\theta^{-1} B]^T B \tag{7}$$

where $\Theta = [\theta^0, \theta^1, \ldots, \theta^N]$, and $\text{Diag}[X]$ is a diagonal matrix with the vector X along the diagonal.

II. Discrete Polynomial Transform Matrices

The Hermite, binomial, and Laguerre discrete transforms can be simplified by factoring the matrix B in (4) into a series of binary matrices [6]

$$B = G_1 G_2 \cdots G_N \tag{8}$$

where G_i is an upper triangular matrix with diagonal elements equal to 1, upper diagonal elements $[G_i]_{j,j+1} = 1$ for $j \ge i-1$, and 0 elsewhere, for $i = 1, 2, \ldots, N$ and $j = 0, 1, \ldots, N$. The Hermite transform matrix is given by

$$\begin{aligned}
H_N &= [I^- Q G_1 G_2 \cdots G_N]^T [\text{Diag}[I^{\sim}(B^T B)]]^{-1} \\
&\quad \cdot G_1 G_2 \cdots G_N
\end{aligned} \tag{9}$$

$$= G_N^T \cdots G_2^T G_1^T Q I^- V_N G_1 G_2 \cdots G_N. \tag{10}$$

The diagonal elements of $\text{Diag}[I^{\sim}(B^T B)]$ are the binomial coefficients $\binom{N}{k}$, for $k = 0, 1, \ldots, N$ [7], so the inverse of this matrix is simply

$$V_N = [\text{Diag}[I^{\sim}(B^T B)]]^{-1} = \begin{pmatrix} v_0 & 0 & \cdots & 0 \\ 0 & v_1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & v_N \end{pmatrix} \tag{11}$$

1070-9908/03$17.00 © 2003 IEEE

Optimized Recursive Subband Synthesis Windowing for Implementing Efficient MPEG Audio Decoders

Patrick De Smet, *Student Member, IEEE,* and Ignace Bruyland, *Member, IEEE*

***Abstract*—In this letter, we discuss a recursive optimization approach for the subband synthesis windowing component used in all three layers of the MPEG-1 and -2 audio decoding frameworks. We explain the design of the proposed algorithm and briefly discuss how it can be combined with other novel and existing optimization strategies. The impact of these optimizations on the number of multiplications and additions is shown, and the influence of the algorithm on the amount of memory required by the adapted decoder is also discussed. We show that the proposed algorithm offers very interesting possibilities for implementing efficient VLSI designs.**

***Index Terms*—Efficient decoder, MPEG audio, recursive windowing.**

I. Introduction

IN THE LAST ten years, the ISO/IEC MPEG-1 and -2 audio encoding standards have been gaining considerable importance and popularity as efficient tools for lossy compression of high-quality digital audio [1], [2]. This letter aims at contributing other, previously undocumented possibilities for obtaining faster and/or more cost-effective MPEG audio decoders. We mainly focus on realizing more efficient VLSI designs by reducing the number of multiplications; compared to adder components, multipliers are relatively slow and often result in considerably more demanding design restrictions [3]. The proposed algorithm is based upon the optimization of redundant computations carried out in subsequent synthesis filtering operations and also extends and complements techniques discussed in [4]–[6] (encoding only), [7]–[9] (decoding only), and [10]–[12] (both).

II. MPEG Audio Decoding and Subband Synthesis

Within the MPEG-1 and -2 audio encoding standards, ISO/IEC has defined three different "layers" [1], [13]. Each layer builds upon its lower ranked siblings, and for each higher layer number, (implementation) complexity and compression efficiency typically both increase. The cosine-modulated polyphase filter bank used by all three layers within the framework maps the digital audio input into 32 equally spaced and critically decimated bandpass signals [2]. Compared to the encoder, the computational complexity of the MPEG audio decoder is significantly lower; most of the computation time is due to the synthesis filter bank [7], [9].

For each multirate filtering step at time t, the ISO synthesis subband filtering algorithm starts with $M = 32$ subband samples extracted from the encoded bitstream syntax, i.e., $S_{sb}[i;t]$, $i = 0, 1, \ldots, M-1$, and produces M pulse code modulation (PCM) audio output samples; $S_{\mathrm{pcm}}[j;t]$, $j = 0, 1, \ldots, M-1$.

First, the intermediate inverse polyphase quadrature mirror filter bank (IPQMF) vector $V[i;t]$ is shifted over $2M$ memory positions according to $V[i;t]=V[i-2M;\, t-1]$, for $i = 2PM-1, 2PM-2, \ldots, 2M$ with $P = N/M = 16$ and $N = 512$ the length of the prototype filter window [2]. Then, the basic MPEG audio matrixing operation [13] is calculated

$$V[i;t] = \sum_{k=0}^{M-1} S_{sb}[k;t] \cos\left(\frac{\pi(2k+1)\left(\frac{M}{2}+i\right)}{2M}\right) \quad (1)$$

for $i=0, 1, \ldots, 2M-1$. An N-value intermediate vector $U[i;t]$ is then built using $U[2Mk+l;t]=V[4Mk+l;t]$, and $U[2Mk+M+l;t]=V[4Mk+3M+l;t]$, $k = 0, 1, \ldots, P/2-1$, $l = 0, 1, \ldots, M-1$. Finally, the so-called windowing component computes M new PCM audio output samples, i.e., $S_{\mathrm{pcm}}[j;t] = \sum_{i=0}^{P-1} U[j+Mi;t]\, D[j+Mi]$, $j=0, 1, \ldots, M-1$, in which the values D are the N synthesis window coefficients (see [13, Table B.3]).

This algorithm requires $2M^2 + MP = 2560$ multiplications and $2M(M-1) + M(P-1) = 2464$ additions, and, not counting the PCM and subband sample buffers, at least $3N + M = 1568$ datawords of memory; $2PM = 2N = 1024$ datawords V, $M = 32$ cosine-coefficients (contrary to [15], using all 2048 cosine-coefficients in (1), we assume appropriate address calculators always being implemented), and $N = 512$ window coefficients D.

III. Matrixing Optimizations

Obviously, (1) can be simplified based on the properties of the cosine-function. This has been discussed in [7], [9], and [12], yielding a fast odd/even butterfly M-point inverse discrete cosine tranform decomposition (see also [14]). However, except for the fact that, for $l = 0, 1, \ldots, M/2-1$

$$\mathrm{SDS}_0[l;t] = S_{sb}[l;t] + S_{sb}[M-1-l;t]$$
$$\mathrm{SDS}_1[l;t] = S_{sb}[l;t] - S_{sb}[M-1-l;t]$$
$$V[i;t] = \sum_{k=0}^{M/2-1} \mathrm{SDS}_{(i \bmod 2)}[k;t] \cos\left(\frac{\pi(2k+1)\left(\frac{M}{2}+i\right)}{2M}\right)$$

Manuscript received December 18, 2001; revised November 15, 2002. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Antonio Ortega.

The authors are with the Department of Telecommunications and Information Processing (TW07), Ghent University, B-9000 Ghent, Belgium (e-mail: pds@telin.UGent.be; ib@telin.UGent.be).

Digital Object Identifier 10.1109/LSP.2003.817173

1070-9908/03$17.00 © 2003 IEEE

This material may be protected by Copyright law (Title 17 U.S. Code)

Google Exhibit No. 1063
Page 1442

with

$$\left\{V[i;t] = -V[M-i;t], \quad i = 0,1,\ldots,\frac{M}{2}-1\right\}$$
$$\left\{V[M+i;t] = V[2M-i;t], \quad i = 1,2,\ldots,\frac{M}{2}-1\right\}$$
$$V[M/2;t] = 0 \quad V[3M/2;t] = -\sum_{k=0}^{M/2-1} \mathrm{SDS}_0[k;t] \qquad (2)$$

we do not repeat this discussion here; we will reduce the computational complexity by pursuing other possibilities. The values for sum/difference-symmetry (SDS) reflect the symmetry due to the factor $(2k+1)$ in the cosine argument of (1).

The partially optimized matrixing formulas given above require only $(M-1)M/2 = 496$ multiplications and $M+(M/2-1)(M-1)+(M/2-1) = M^2/2 = 512$ additions to compute the $V[i;t]$ matrixing results. If needed, these equations can also reduce the required amount of memory for the entire filtering procedure to $2N+2M = 1088$ datawords; $M = 32$ values SDS_0 and SDS_1, $PM = N = 512$ remaining values V, $M = 32$ cosine-coefficients, and $N = 512$ window coefficients D.

IV. Recursive Windowing Optimizations

For VLSI design, multiplications are considered significantly more expensive than additions; the required multiplier speed can often influence the choice of the entire implementation design (see the *en*coder architecture trade-offs discussed in [3] and [4]). Hence, in this section we focus on reducing the number of multiplications in the windowing component.

Now, since

$$S_{\mathrm{pcm}}[j;t] = \sum_{i=0}^{P/2-1} \left(U[j+Mi;t]\, D[j+Mi]\right) + \left(D\left[j+Mi+\frac{MP}{2}\right]\, U\left[j+Mi+\frac{MP}{2};t\right]\right)$$

and $ab+cd = (a+c)(b+d) - ad - bc$, we have

$$S_{\mathrm{pcm}}[j;t] = \sum_{i=0}^{P/2-1} \left(U[j+Mi;t] + D\left[j+Mi+\frac{MP}{2}\right]\right) \cdot \left(D[j+Mi] + U\left[j+Mi+\frac{MP}{2};t\right]\right) - \sum_{i=0}^{P/2-1} U[j+Mi;t]\, U\left[j+Mi+\frac{MP}{2};t\right] - \sum_{i=0}^{P/2-1} D[j+Mi]\, D\left[j+Mi+\frac{MP}{2}\right] \qquad (3)$$

for $j = 0,1,\ldots,M-1$. Rewriting this sum as $S_{\mathrm{pcm}}[j;t] = A[j;t] - B[j;t] - C[j]$, the last sum $C[j]$ is an M-dimensional constant vector which can be precalculated for each j. Below, each of these terms will be discussed in more detail, but first let us consider some more general issues. The A terms will remain to be calculated by the final algorithm, but the B terms can be simplified further by splitting them into two sums for even and odd multiples of M for the index i, and by replacing U by V (see Section II). This yields

$$B[j;t] = \sum_{i=0}^{P/4-1} \left(V[j+4Mi;t]\, V[j+4Mi+MP;t] + V[j+4Mi+3M;t]\, V[j+4Mi+MP+3M;t]\right). \qquad (4)$$

As already indicated, t is an external clocking signal driving the subband synthesis filter; for each filter calculation at time t, $2M$ new values $V[l;t]$ are determined according to (1) or (2). Considering the previous calculations of the subband synthesis filter, it is clear that $V[l+2Mn;t]=V[l;t-n]$, for $l = 0, 1, \ldots, (2PM-1-2Mn)$, or equivalently, $V[l;t]=V[l-2Mn;t-n], l = (2Mn),\ldots,2PM-2,2PM-1$. Hence

$$B[j;t] = B[j;(t-2)] - d[j;(t-8)] + d[j;t]$$

with

$$d[j;t] = V[j;t]\, V[j+MP;t] + V[j+3M;t]\, V[j+3M+MP;t] \qquad (5)$$

for $j = 0,1,\ldots,M-1$. This result can be easily obtained from (4) by removing the last terms for $i = P/4-1 = 3$ and adding new terms for $i = 0$, and shifting over two clocking periods t or $4M = 128$ positions.

The equations in this section yield the proposed recursive algorithm for calculating the windowing component. To our knowledge this decoder structure has not been proposed before. The recursive computation of $B[j;t]$ allows the number of multiplications to be reduced by recycling data computed in previous filtering operations [using two separate delay lines for B and d; see (5)]. Below, we will discuss the memory requirements and the computational complexity of the proposed algorithm, but first let us discuss some additional possibilities for further optimization.

V. Additional Optimizations

A. Detailed Optimizations for the Recursive Algorithm

From the equations given above, it is clear that $V[M/2;t] = V[M/2+MP;t] = 0$; and hence, $d[M/2;t] = V[M/2+3M;t]\, V[M/2+MP+3M;t]$, using only one multiplication to compute $d[M/2;t]$. The value $d[0;t]$ can be calculated by using two multiplications and one addition. Finally, based on the properties of V [see (2)] and the definition of d given above, it can be easily shown that $d[j;t] = d[M-j;t]$, $j=1,2,\ldots,M/2-1$. Hence, the number of multiplications and additions for calculating all $d[j;t]$ can be halved. Additionally, this also means that $B[j;t]=B[M-j;t]$, $j=1,2,\ldots,M/2-1$, and that $B[0;t]$ and $B[M/2;t]$ are unique.

Google Exhibit No. 1063
Page 1443

B. Windowing Symmetry

Another possibility for further optimization is given by the fact that $\{D[0]=0\}$

$$\left\{ D\left[\frac{N}{2}+k\right] = -D\left[\frac{N}{2}-k\right] \quad \text{for } k\neq 2M, 4M, 6M,\ k=1,2,\ldots,\frac{N}{2}-1 \right\}$$

and $\{D[k] = D[N-k];\ k=2M, 4M, 6M\}$, and $D[N/2]$ are unique. To our knowledge, these properties have not received any real attention in the existing literature. This seems odd, since clearly, exploiting this symmetry will immediately yield a significant advantage in the amount of memory needed to store these coefficients. Also, for $j = 1,2,\ldots,M/2-1$, from (3) and with $PM = N$ (see above), we get (6) (shown at the bottom of the page). Assuming now that $i=P/2-1-l$ in each parallel computation step for a fixed $j=1,2,\ldots,M/2-1$, and exploiting the symmetry of the window coefficients, i.e., $D[j+Mi]= -D[M-j+Ml+N/2]$ and $-D[M-j+Ml]=D[j+Mi+N/2]$, half of the window D suffices to compute all values A. Also, note that the parallel computations in (6) can be used to avoid reloading the coefficients D (from external memory) into internal registers. Next, by considering that $D[0] = 0$ and $U[M/2+2Mn;t] = V[M/2+4Mn;t] = V[M/2;t-2n] = 0$; it is clear that

$$A[0;t] = \left(U[0;t] + D\left[\frac{N}{2}\right]\right) U\left[\frac{N}{2};t\right] + \sum_{i=1}^{P/2-1} \left(U[Mi;t] + (-1)^i \cdot D\left[\frac{N}{2}-Mi\right]\right)\left(D[Mi] + U\left[Mi+\frac{N}{2};t\right]\right)$$

and

$$A\left[\frac{M}{2};t\right] = A_{M/2} + \sum_{i=0}^{P/4-1} \left(U\left[\frac{M}{2}+M+2Mi;t\right] - D\left[\frac{N}{2}-\frac{M}{2}-M-2Mi\right]\right) \cdot \left(D\left[\frac{M}{2}+M+2Mi\right] + U\left[\frac{N}{2}+M+\frac{M}{2}+2Mi;t\right]\right)$$

with the constant $A_{M/2} = -\sum_{i=0}^{P/4-1} D[N/2-M/2-2Mi]\ D[M/2+2Mi]$. Finally, for the constants C it can be shown that $C[j] = C[M-j] = \sum_{i=0}^{P/2-1} D[j+Mi]\ (-D[N/2-j-Mi])$, for $j=1,2,\ldots,M/2-1$. Additionally, $C[M/2] = \sum_{i=0}^{P/2-1} D[M/2+Mi]\ (-D[N-M/2-Mi])$, and $C[0] = (D[MP/4])^2 + 2\sum_{i=1}^{P/4-1}(-1)^i D[Mi]\ D[N/2-Mi]$. These equations can aid in a faster (re)start of a decoder when the C-values are not hard-coded into the actual application or the ROM of a device.

VI. Final Optimized Algorithm

Initially, the constants $A_{M/2}$, $C[0]$, $C[M/2]$, and $C[j]$, $j=1,2,\ldots,M/2-1$, need to be available. All other values are initialized to zero but both $B[j;t=0]$ and $B[j;t=1]$ are set to $C[j]$ for $j = 0,1,\ldots,M/2$. This absorbs the values C into the recursive Bs, simplifying the computation of the PCM output samples $S_{\text{pcm}}[j;t]$. After extracting the M encoded subband samples $S_{sb}[j;t]$ from the bitstream, the matrixing operations are computed by using (2). Next, $d[j;t]$ is computed according to the definition given in Section IV and the symmetry issues discussed in Section V–A. This enables the computation of all $B[j;t]$, again according to Sections V–A and IV. Then, in a single loop, $A[j;t]$ and $A[M-j;t]$, and next, $A[0;t]$ and $A[M/2;t]$ can be computed using the equations formulated above. Finally, we can compute $S_{\text{pcm}}[j;t]=A[j;t]-B[j;t]$ for outputing samples.

Adapted source code for the ISO codec [15] is freely available from the authors.[1]

VII. Computational Complexity

A. Recursive Synthesis Windowing

The original windowing component as described in Section II requires MP=512 multiplications and $M(P-1)$=480 additions for calculation of the PCM samples.

In the proposed recursive algorithm, the computation of the values $A[j;t]$, requires $(M-2)P/2 = 240$ multiplications and $2(M-2)(P/2) + (M-2)(P/2-1) = 690$ additions for $j = 1,2,\ldots,M/2-1, M/2+1, M/2+2,\ldots,M-1$; see (6). The computation of $A[0;t]$ requires $1 + (P/2-1) = 8$ multiplications and $1 + 2(P/2-1) + (P/2-1) = 22$ additions. For the calculation of $A[M/2;t]$ we need $P/4 = 4$ multiplications and $2(P/4) + P/4 = 12$ additions. This makes a total of 252 multiplications and 724 additions. The computation of all M values d requires $1+2+2(M/2-1) = 33$ multiplications and $1 + (M/2-1) = 16$ additions. Similarly, exploiting the

[1]See also http://telin.UGent.be/~pds/papers/SPL2003/.

$$A[j;t] = \sum_{i=0}^{P/2-1} \left(U[j+Mi;t] + D\left[j+Mi+\frac{N}{2}\right]\right)\left(D[j+Mi] + U\left[j+Mi+\frac{N}{2};t\right]\right)$$

$$A[M-j;t] = \sum_{l=P/2-1}^{0} \left(U[M-j+Ml;t] + D\left[M-j+Ml+\frac{N}{2}\right]\right)\left(D[M-j+Ml] + U\left[M-j+Ml+\frac{N}{2};t\right]\right) \quad (6)$$

Google Exhibit No. 1063
Page 1444

symmetry in B (see also Section V–A), only $2(M/2+1) = 34$ additions are needed for the computation of all values B. Finally, for the calculation of the PCM output samples, adding A and B (with C absorbed into B) requires $M = 32$ additions. Hence, 285 multiplications and 806 additions are needed to implement the optimized recursive windowing algorithm. The proposed algorithm results in an increase in the required number of additions. However, the number of multiplications has been reduced significantly; as already mentioned, this will normally put considerably less restricting demands on (the speed of) a VLSI or hardware architecture [3].

B. Full Subband Synthesis Filtering

In summary, when compared to the full basic ISO subband synthesis filter as discussed in Section II, the computational complexity is reduced from 2560 multiplications and 2464 additions (see Section II), to 781 multiplications and 1318 additions (see Section III: 496 multiplications and 512 additions; and, Section VII-A: 285 multiplications and 806 additions).

VIII. Memory Requirements

A. Recursive Synthesis Windowing

Implementing the proposed algorithm requires $8(1+M/2)=136$ datawords for $1 + M/2 = 17$ values $d[j;t]$ with a maximum time delay of eight synthesis operations, $2(1 + M/2) = 34$ datawords for $B[j;t]$ due to a buffer of 2 filtering operation delays, one dataword for $A_{M/2}$, and the $1 + M/2 = 17$ constants C. However, as indicated above, these constants C are absorbed into the recursive computation of B, and they can be easily computed when the decoder is being initialized; the constants C are not needed during "steady state" operation of the decoder. Thus, the recursive algorithm can be implemented with a total of $10(1 + M/2) + 1 = 5M + 11 = 171$ additional datawords. Fortunately, by also exploiting the symmetry in the coefficients D (see Section V–B), we can reclaim $N/2 = 256$ datawords. Hence, compared to the traditional implementations described in the existing literature, the overall balance of all the optimizations we proposed in this letter is a gain of $N/2 - (5M + 11) = 85$ datawords.

B. Full Subband Synthesis Filtering

In this letter, we discussed and proposed the memory-related optimization of the MPEG matrixing operations (see Section III), the auxiliary recursive data (see Section V–A), and the window coefficients (see Section V–B). After implementing all these optimizations, the required amount of memory can be reduced from $3N+M=1568$ datawords (see Section II) to $(2N+2M)-(N/2-5M-11)=1003$ datawords; see Sections III and VIII-A.

Note that not all symmetry properties must be implemented. This results in several flexible trade-off possibilities for exchanging address calculation logic and control complexity for memory requirements and vice versa.

IX. Conclusion

In this letter, we proposed a recursive windowing algorithm useful for implementing efficient MPEG-1 and -2 audio subband synthesis filters. The proposed windowing algorithm reduces the required number of multiplications from 512 to 285, while the number of additions changes from 480 to 806. This approach can be very profitable when a certain hardware design has some limitations on (the speed of) its multiplier(s). Additionally, we proposed and discussed several memory related requirements and optimizations. In combination with an existing and well-known cosine-symmetry exploiting optimization step, this letter simplifies the original ISO subband synthesis filtering procedure from 2560 multiplications and 2464 additions to 781 multiplications and 1318 additions, while reducing the required amount of memory from 1568 to 1003 datawords.

References

[1] P. Noll, "MPEG digital audio coding," *IEEE Signal Processing Mag.*, pp. 59–81, Sept. 1997.

[2] M. Kahrs and K. Brandenburg, Eds., *Applications of Digital Signal Processing to Audio and Acoustics*. Dordrecht, The Netherlands: Kluwer, 1998.

[3] N. Ling and C.-W. Shih, "Two polyphase filter architectures for MPEG audio," in *Proc. 40th Midwest Symp. Circuits and Systems*, vol. 2, 1997, pp. 861–864.

[4] D.-Y. Chan, J.-F. Yang, and C.-C. Fang, "Fast implementation of MPEG audio coder using recursive formula with fast discrete cosine transforms," *IEEE Trans. Speech Audio Processing*, vol. 4, pp. 144–148, Mar. 1996.

[5] M. Kumar and M. Zubair, "A high performance software implementation of MPEG audio encoder," *Proc. ICASSP*, vol. 2, pp. 1049–1052, 1996.

[6] H.-O. Oh, S.-Y. Kim, D.-H. Youn, and I.-W. Cha, "New implementation techniques of a real-time MPEG-2 audio encoding system," *Proc. ICASSP*, vol. 4, pp. 2287–2290, 1999.

[7] T.-H. Tsai, T.-H. Chen, and L.-G. Chen, "An MPEG audio decoder chip," *IEEE Trans. Consumer Electron.*, vol. 41, pp. 89–96, Feb. 1995.

[8] P. De Smet, "On two new techniques for optimized MPEG audio decoding," in *Proc. IASTED Int. Conf. SPPRA 2001*, July 3–6, 2001, pp. 79–83.

[9] W.-K. Paik and S.-Y. Hwang, "Design of a novel synthesis filter for real-time MPEG-2 audio decoder implementation on a DSP chip," *IEEE Trans. Consumer Electron.*, vol. 45, pp. 1119–1129, Nov. 1999.

[10] C. D. Murphy and L. Anandakumar, "Real-time MPEG-1 audio coding and decoding on a DSP chip," *IEEE Trans. Consumer Electron.*, vol. 43, pp. 40–47, Feb. 1997.

[11] V. Britanak and K. R. Rao, "An efficient implementation of the forward and inverse MDCT in MPEG audio coding," *IEEE Signal Processing Lett.*, vol. 8, pp. 48–51, Feb. 2001.

[12] K. Konstantinides, "Fast subband filtering in MPEG audio coding," *IEEE Signal Processing Lett.*, vol. 1, pp. 26–28, Feb. 1994.

[13] ISO/IEC, "Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s,", ISO/IEC 11 172–3, 1993.

[14] B. G. Lee, "A new algorithm to compute the discrete cosine transform," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-32, pp. 1243–1245, Dec. 1984.

[15] ISO dist10 Source Code Package [Online]. Available: ./mpeg2/software/ at ftp://ftp.tnt.uni-hannover.de/pub/MPEG/audio/.

Google Exhibit No. 1063
Page 1445

Ex. 1043B

LIBRARY | CATALOG

1 of 1

PERIODICAL OR NEWSPAPER

IEEE signal processing letters

Full Record MARC Tags

Main title

IEEE signal processing letters.

Published/Created

New York, NY : IEEE Signal Processing Society, c1994-

Request this Item | LC Find It

More Information

LCCN Permalink	https://lccn.loc.gov/94648235
Publication history	Vol. 1, no. 1 (Jan. 1994)-
Description	v. : ill. ; 28 cm.
Current frequency	Monthly
ISSN	1070-9908
Linking ISSN	1070-9908
LC classification	TK5102.9 .I32

Google Exhibit No. 1063
Page 1447

Variant title	Institute of Electrical and Electronics Engineers signal processing letters
Portion of title	Signal processing letters
Serial key title	IEEE signal processing letters
Abbreviated title	IEEE signal process. lett.
Related names	IEEE Signal Processing Society.
LC Subjects	Signal processing--Periodicals. Image processing--Periodicals. Speech processing systems--Periodicals.
Other Subjects	Image processing. Signal processing. Speech processing systems. Signalverarbeitung Zeitschrift Online-Ressource
Form/Genre	Periodicals. Periodicals.
Browse by shelf order	TK5102.9
Notes	Title from cover.
Reproduction no./Source	IEEE Service Center, 445 Hoes Lane, PO Box 1331, Piscataway, NJ 08855-1331
Additional formats	Online version: IEEE signal processing letters (Online) 1558-2361 (DLC) 2005212489 (OCoLC)60638457 Online version: IEEE signal processing letters (OCoLC)752415014
LCCN	94648235
Invalid LCCN	sn 93002838
CODEN	ISPLEM
Dewey class no.	621.382/2/05

Google Exhibit No. 1063
Page 1448

Other system no.	(OCoLC)ocm28525247
Invalid CODEN	IESPEJ
Additional Links	http://ieeexplore.ieee.org/servlet/opac?punumber=97
Type of material	Periodical or Newspaper
Source of description	Latest issue consulted: Vol. 5, no. 1 (Jan. 1998) (surrogate).

Item Availability

CALL NUMBER	TK5102.9 .I32 Set 1
Request in	Jefferson or Adams Building Reading Rooms
Status	c.1 v. 20, no. 1/3-4/6 2013 Jan/Mar-Apr/June In Process 04-16-2014
Older receipts	v.1 (1994); v.2:no.1-v.2:no.3 (1995:Jan.-1995:Mar.), v.2:no.5-v.9:no.5 (1995:May-2002:May) v.9:no.7-v.9:no.8 (2002:July-2002:Aug.), v.9:no.10-v.9:no.12 (2002:Oct.-2002:Dec.) v.10-v.12:no.1 (2003-2005:Jan.), v.12:no.3-v.23:no.10/12 (2005:Mar.-2016:Oct./Dec.), v.24:no.4/6-v.25:no.10/12 (2017:Apr./June-2018:Oct./Dec.)

CALL NUMBER	TK5102.9 .I32 Copy 1 Unbound issues
Request in	Newspaper & Current Periodical Reading Room (Madison LM133)
Latest receipts	v. 15, no. 2 (2008 Feb.) v. 15, no. 1 (2008 Jan.)

Google Exhibit No. 1063
Page 1449

LIBRARY CATALOG

1 of 1

PERIODICAL OR NEWSPAPER

IEEE signal processing letters

Full Record MARC Tags

000		02642cas a2200649 a 4500
001		11440048
005		20240918094749.0
008		930727c19949999nyumr p 0 a0eng c
010	__	|a 94648235 |z sn 93002838
022	0_	|a 1070-9908 |l 1070-9908 |2 1
030	__	|a SPLEM |z IESPEJ
032	__	|a 025151 |b JSPS
035	__	|a (OCoLC)ocm28525247
037	__	|b IEEE Service Center, 445 Hoes Lane, PO Box 1331, Piscataway, NJ 08855-1331
040	__	|a NSD |b eng |c NSD |d DLC |d WAU |d NYG |d CAS |d DLC |d IXA |d EYM |d OLC |d UL |d MYG |d CUS |d MYG |d SYS |d _YU |d EEEX |d NSD |d CIT |d OCLCQ |d OCLCF |d OCLCO |d OCLCA |d UL |d DEFHM |d OCLCO |d OCLCA |d GILDS
042	__	|a nsdp |a pcc
050	00	|a TK5102.9 |b .I32
082	00	|a 621.382/2/05 |2 20
210	0_	|a EEE signal process. lett.
222	_0	|a EEE signal processing letters
245	00	|a EEE signal processing letters.
246	2_	|a nstitute of Electrical and Electronics Engineers signal processing letters
246	30	|a Signal processing letters
260	__	|a NewYork, NY : |b EEE Signal Processing Society, |c c1994-
300	__	|a v. : |b ill. ; |c 28 cm.

Google Exhibit No. 1063
Page 1450

310	__	|a Monthly
362	0_	|a Vol. 1, no. 1 (Jan. 1994)-
500	__	|a Title from cover.
588	__	|a Latest issue consulted: Vol. 5, no. 1 (Jan. 1998) (surrogate).
590	__	|a SERBIB/SERLOC merged record
650	_0	|a Signal processing |v Periodicals.
650	_0	|a Image processing |v Periodicals.
650	_0	|a Speech processing systems |v Periodicals.
650	_7	|a Image processing. |2 fast |0 (OCoLC)fst00967501
650	_7	|a Signal processing. |2 fast |0 (OCoLC)fst01118281
650	_7	|a Speech processing systems. |2 fast |0 (OCoLC)fst01129243
650	_7	|a Signalverarbeitung |2 gnd
650	_7	|a Zeitschrift |2 gnd
650	_7	|a Online-Ressource |2 gnd
655	_7	|a Periodicals. |2 fast |0 (OCoLC)fst01411641
655	_7	|a Periodicals. |2 lcgft
710	2_	|a IEEE Signal Processing Society.
776	08	|i Online version: |t IEEE signal processing letters (Online) |x 1558-2361 |w (DLC) 2005212489 |w (OCoLC)60638457
776	08	|i Online version: |t IEEE signal processing letters |w (OCoLC)752415014
850	__	|a CPT |a DLC |a InU |a MiU |a N |a NjP |a TxU
859	41	|u http://ieeexplore.ieee.org/servlet/opac?punumber=97
853	00	|8 1 |a v. |b no. |u 12 |v r |i (year) |j (month) |w m
863	40	|8 1.1 |a <1>- |i <1994>- |x provisional
863	41	|8 1.2 |a <8> |b <1> |i <2001> |j <01>
890	__	|a IEEE signal processing letters. |i 94-648235
906	__	|a 7 |b cbc |c serials |d u |e ncip |f 19 |g n-oclcserc
920	__	|a Keep 1
923	__	|d 20020729 |s Group Registration
984	__	|a srvf
985	__	|e eserial 200406
991	__	|b c-GenColl |h TK5102.9 |i .I32 |w SERIALS
992	__	|b SER |w SERLOC

Google Exhibit No. 1063
Page 1451

Request this Item | LC Find It

Item Availability

CALL NUMBER	TK5102.9 .I32 Set 1
Request in	Jefferson or Adams Building Reading Rooms
Status	c.1 v. 20, no. 1/3-4/6 2013 Jan/Mar-Apr/June In Process 04-16-2014
Older receipts	v.1 (1994); v.2:no.1-v.2:no.3 (1995:Jan.-1995:Mar.), v.2:no.5-v.9:no.5 (1995:May-2002:May) v.9:no.7-v.9:no.8 (2002:July-2002:Aug.), v.9:no.10-v.9:no.12 (2002:Oct.-2002:Dec.) v.10-v.12:no.1 (2003-2005:Jan.), v.12:no.3-v.23:no.10/12 (2005:Mar.-2016:Oct./Dec.), v.24:no.4/6-v.25:no.10/12 (2017:Apr./June-2018:Oct./Dec.)

CALL NUMBER	TK5102.9 .I32 Copy 1 Unbound issues
Request in	Newspaper & Current Periodical Reading Room (Madison LM133)
Latest receipts	v. 15, no. 2 (2008 Feb.) v. 15, no. 1 (2008 Jan.)

Google Exhibit No. 1063
Page 1452

Ex. 1043C

Copyright
United States Copyright Office

Try the Copyright Public Records System (CPRS) pilot with enhanced search features and filters.

Help | Search | History | Titles | Start Over

Public Catalog

Copyright Catalog (1978 to present) at DC4
Search Request: Left Anchored Title = IEEE signal processing letters
Search Results: Displaying 7 of 32 entries

previous next

Labeled View

IEEE signal processing letters.

Type of Work: Serial
Title: IEEE signal processing letters.
Serial Publication Year: 2003
Serial Key Title: IEEE signal processing letters
ISSN: 1070-9908
Description: print material.
Frequency: Monthly.
Publication History: Vol. 1, no. 1, Jan. 1994-
Description based on: Vol. 5, no. 1, Jan. 1998.
Copyright Claimant: IEEE.
Issues Registered: v. 9, no. 12, Dec02.; Pub. 2003-01-04; Reg. 2003-04-11; TX0005707747
v. 10, no. 1, Jan03.; Pub. 2003-01-21; Reg. 2003-04-11; TX0005707747
v. 10, no. 10, Oct03.; Pub. 2003-09-18; Reg. 2004-11-08; TX0006065357
v. 10, no. 11, Nov03.; Pub. 2003-10-16; Reg. 2004-11-08; TX0006065357
v. 10, no. 12, Dec03.; Pub. 2003-11-18; Reg. 2004-11-08; TX0006065357
Names: IEEE

previous next

Save, Print and Email (Help Page)
Select Download Format: Full Record | Format for Print/Save
Enter your email address: | Email

Help Search History Titles Start Over

Contact Us | Request Copies | Get a Search Estimate | Frequently Asked Questions (FAQs) about Copyright | Copyright Office Home Page | Library of Congress Home Page

Google Exhibit No. 1063
Page 1454

INTERNATIONAL ORGANISATION FOR STANDARDISATION
ORGANISATION INTERNATIONALE DE NORMALISATION
ISO/IEC JTC1/SC29/WG11
CODING OF MOVING PICTURES AND AUDIO

ISO/IEC JTC1/SC29/WG11 **N4668**

March 2002

Source:	**WG11 (MPEG)**
Status:	**Final**
Title:	**MPEG-4 Overview - (V.21 – Jeju Version)**
Editor:	**Rob Koenen (rob.koenen@m4if.org)**

All comments, corrections, suggestions and additions to this document are welcome, and should be send to both the editor and the chairman of MPEG's Requirements Group: Fernando Pereira, fp@lx.it.pt

© MPEG 1999-2002 – unlimited reproduction permitted if not modified

Google Exhibit No. 1044
Page 001

Google Exhibit No. 1063
Page 1455

Overview of the MPEG-4 Standard

Executive Overview

MPEG-4 is an ISO/IEC standard developed by MPEG (Moving Picture Experts Group), the committee that also developed the Emmy Award winning standards known as MPEG-1 and MPEG-2. These standards made interactive video on CD-ROM, DVD and Digital Television possible. MPEG-4 is the result of another international effort involving hundreds of researchers and engineers from all over the world. MPEG-4, with formal as its ISO/IEC designation 'ISO/IEC 14496', was finalized in October 1998 and became an International Standard in the first months of 1999. The fully backward compatible extensions under the title of MPEG-4 Version 2 were frozen at the end of 1999, to acquire the formal International Standard Status early in 2000. Several extensions were added since and work on some specific work-items work is still in progress.

MPEG-4 builds on the proven success of three fields:

- Digital television;
- Interactive graphics applications (synthetic content);
- Interactive multimedia (World Wide Web, distribution of and access to content)

MPEG-4 provides the standardized technological elements enabling the integration of the production, distribution and content access paradigms of the three fields.

More information about MPEG-4 can be found at MPEG's home page (case sensitive): http://mpeg.telecomitalialab.com This web page contains links to a wealth of information about MPEG, including much about MPEG-4, many publicly available documents, several lists of 'Frequently Asked Questions' and links to other MPEG-4 web pages. The standard can be bought from ISO, send mail to sales@iso.ch. Notably, the complete software for MPEG-4 version 1 can be bought on a CD ROM, for 56 Swiss Francs. It can also be downloaded for free from ISO's website: www.iso.ch/ittf - look under publicly available standards and then for "14496-5". This software is free of copyright restrictions when used for implementing MPEG-4 compliant technology. (This does not mean that the software is fee of patents). As well, much information is available from the MPEG-4 Industry Forum, M4IF, http://www.m4if.org. See section 7, The MPEG-4 Industry Forum.

This document gives an overview of the MPEG-4 standard, explaining which pieces of technology it includes and what sort of applications are supported by this technology.

© MPEG 1999-2002 – unlimited reproduction permitted if not modified

Google Exhibit No. 1044
Page 002

Google Exhibit No. 1063
Page 1456

Table of Contents

1 Scope and features of the MPEG-4 standard 5

1.1 Coded representation of media objects 5

1.2 Composition of media objects 6

1.3 Description and synchronization of streaming data for media objects 7

1.4 Delivery of streaming data 8

1.5 Interaction with media objects 9

1.6 Management and Identification of Intellectual Property 9

2 Versions in MPEG-4 9

3 Major Functionalities in MPEG-4 10

3.1 Transport 10

3.2 DMIF 10

3.3 Systems 11

3.4 Audio 12

3.5 Visual 12

4 Extensions Underway 14

5 Profiles in MPEG-4 16

5.1 Visual Profiles 16

5.2 Audio Profiles 18

5.3 Graphics Profiles 18

5.4 Scene Graph Profiles 19

5.5 MPEG-J Profiles 20

5.6 Object Descriptor Profile 20

6 Verification Testing: checking MPEG's performance 20

6.1 Video 21

6.2 Audio 23

7 The MPEG-4 Industry Forum 25

8 Licensing of patents necessary to implement MPEG-4 26

8.1 Roles in Licensing MPEG-4 26

8.2 Licensing Situation 28

9 Deployment of MPEG-4 28

10 Detailed technical description of MPEG-4 DMIF and Systems 29

10.1 Transport of MPEG-4 30

© MPEG 1999-2002 – unlimited reproduction permitted if not modified

Google Exhibit No. 1044
Page 003

Google Exhibit No. 1063
Page 1457

10.2 DMIF 32
10.3 Demultiplexing, synchronization and description of streaming data 36
10.4 Advanced Synchronization (FlexTime) Model 39
10.5 Syntax Description 42
10.6 Binary Format for Scene description: BIFS 42
10.7 User interaction 45
10.8 Content-related IPR identification and protection 45
10.9 MPEG-4 File Format 46
10.10 MPEG-J 49
10.11 Object Content Information 50
11 Detailed technical description of MPEG-4 Visual 50
11.1 Natural Textures, Images and Video 51
11.2 Structure of the tools for representing natural video 55
11.3 The MPEG-4 Video Image Coding Scheme 56
11.4 Coding of Textures and Still Images 58
11.5 Synthetic Objects 59
12 Detailed technical description of MPEG-4 Audio 63
12.1 Natural Sound 63
12.2 Synthesized Sound 68
13 Detailed Description of the Animation Framework eXtension (AFX) 69
14 Annexes 72
A The MPEG-4 development process 72
B Organization of work in MPEG 74
C Glossary and Acronyms 75

 © MPEG 1999-2002 – unlimited reproduction permitted if not modified

Google Exhibit No. 1044
Page 004

Google Exhibit No. 1063
Page 1458

1 Scope and features of the MPEG-4 standard

The MPEG-4 standard provides a set of technologies to satisfy the needs of authors, service providers and end users alike.

- For *authors*, MPEG-4 enables the production of content that has far greater reusability, has greater flexibility than is possible today with individual technologies such as digital television, animated graphics, World Wide Web (WWW) pages and their extensions. Also, it is now possible to better manage and protect content owner rights.
- For *network service providers* MPEG-4 offers transparent information, which can be interpreted and translated into the appropriate native signaling messages of each network with the help of relevant standards bodies. The foregoing, however, excludes Quality of Service considerations, for which MPEG-4 provides a generic QoS descriptor for different MPEG-4 media. The exact translations from the QoS parameters set for each media to the network QoS are beyond the scope of MPEG-4 and are left to network providers. Signaling of the MPEG-4 media QoS descriptors end-to-end enables transport optimization in heterogeneous networks.
- For *end users*, MPEG-4 brings higher levels of interaction with content, within the limits set by the author. It also brings multimedia to new networks, including those employing relatively low bitrate, and mobile ones. An MPEG-4 applications document exists on the MPEG Home page (www.cselt.it/mpeg), which describes many end user applications, including interactive multimedia broadcast and mobile communications.

For all parties involved, MPEG seeks to avoid a multitude of proprietary, non-interworking formats and players.

MPEG-4 achieves these goals by providing standardized ways to:

1. represent units of aural, visual or audiovisual content, called "media objects". These media objects can be of natural or synthetic origin; this means they could be recorded with a camera or microphone, or generated with a computer;
2. describe the composition of these objects to create compound media objects that form audiovisual scenes;
3. multiplex and synchronize the data associated with media objects, so that they can be transported over network channels providing a QoS appropriate for the nature of the specific media objects; and
4. interact with the audiovisual scene generated at the receiver's end.

The following sections illustrate the MPEG-4 functionalities described above, using the audiovisual scene depicted in Figure 1.

1.1 Coded representation of media objects

MPEG-4 audiovisual scenes are composed of several media objects, organized in a hierarchical fashion. At the leaves of the hierarchy, we find primitive media objects, such as:

- Still images (e.g. as a fixed background);
- Video objects (e.g. a talking person - without the background;
- Audio objects (e.g. the voice associated with that person, background music);

MPEG-4 standardizes a number of such primitive media objects, capable of representing both natural and synthetic content types, which can be either 2- or 3-dimensional. In addition to the

© MPEG 1999-2002 – unlimited reproduction permitted if not modified

Google Exhibit No. 1044
Page 005

Google Exhibit No. 1063
Page 1459

media objects mentioned above and shown in Figure 1, MPEG-4 defines the coded representation of objects such as:

- Text and graphics;
- Talking synthetic heads and associated text used to synthesize the speech and animate the head; animated bodies to go with the faces;
- Synthetic sound.

A media object in its coded form consists of descriptive elements that allow handling the object in an audiovisual scene as well as of associated streaming data, if needed. It is important to note that in its coded form, each media object can be represented independent of its surroundings or background.
The coded representation of media objects is as efficient as possible while taking into account the desired functionalities. Examples of such functionalities are error robustness, easy extraction and editing of an object, or having an object available in a scaleable form.

1.2 Composition of media objects

Figure 1 explains the way in which an audiovisual scene in MPEG-4 is described as composed of individual objects. The figure contains compound media objects that group primitive media objects together. Primitive media objects correspond to leaves in the descriptive tree while compound media objects encompass entire sub-trees. As an example: the visual object corresponding to the talking person and the corresponding voice are tied together to form a new compound media object, containing both the aural and visual components of that talking person.

Such grouping allows authors to construct complex scenes, and enables consumers to manipulate meaningful (sets of) objects.

More generally, MPEG-4 provides a standardized way to describe a scene, allowing for example to:

- Place media objects anywhere in a given coordinate system;
- Apply transforms to change the geometrical or acoustical appearance of a media object;
- Group primitive media objects in order to form compound media objects;
- Apply streamed data to media objects, in order to modify their attributes (e.g. a sound, a moving texture belonging to an object; animation parameters driving a synthetic face);
- Change, interactively, the user's viewing and listening points anywhere in the scene.

The scene description builds on several concepts from the Virtual Reality Modeling language (VRML) in terms of both its structure and the functionality of object composition nodes and extends it to fully enable the aforementioned features.

© MPEG 1999-2002 – unlimited reproduction permitted if not modified

Google Exhibit No. 1044
Page 006

Google Exhibit No. 1063
Page 1460

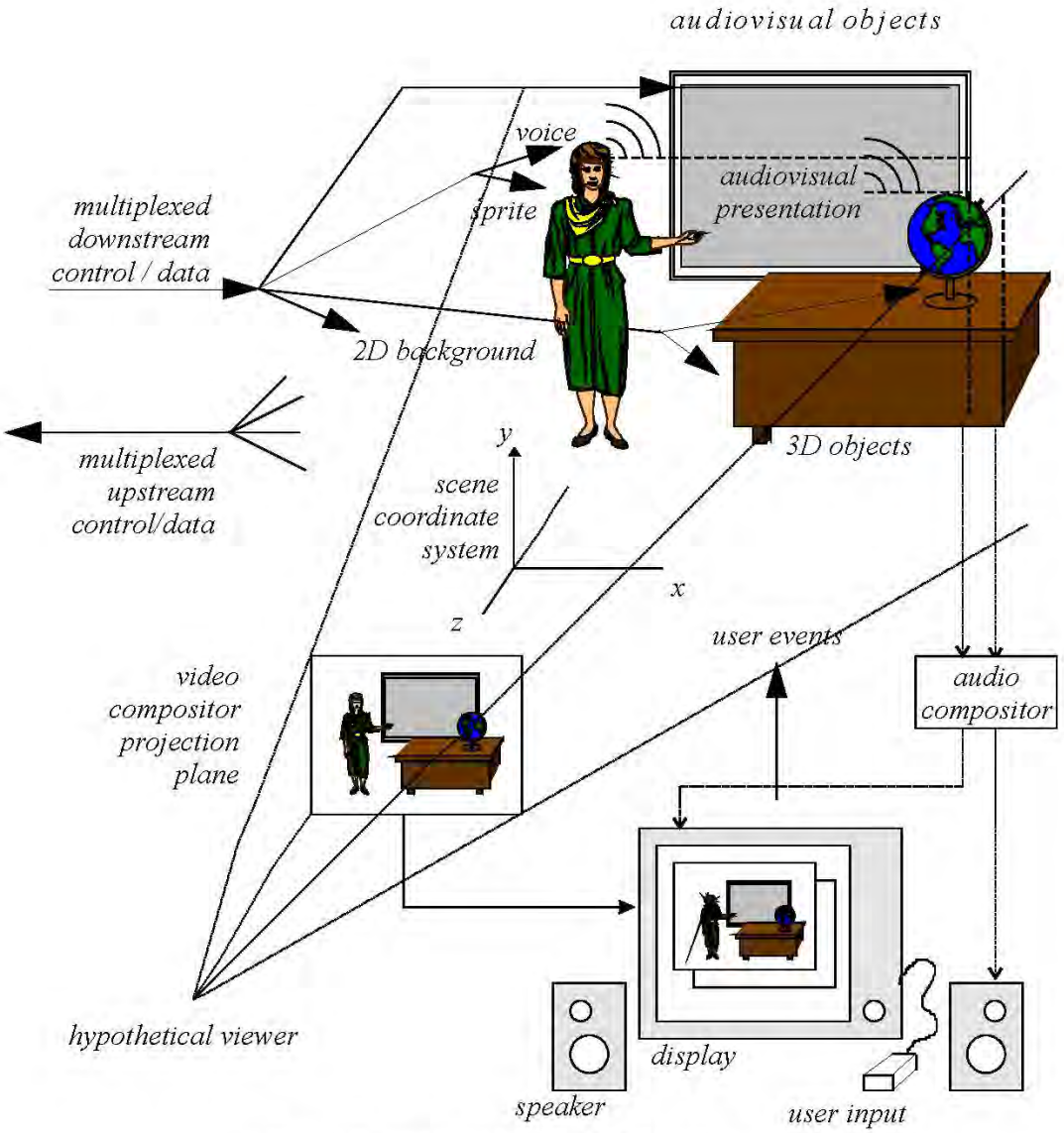


Figure 1 - an example of an MPEG-4 Scene

1.3 Description and synchronization of streaming data for media objects

Media objects may need streaming data, which is conveyed in one or more elementary streams. An object descriptor identifies all streams associated to one media object. This allows handling hierarchically encoded data as well as the association of meta-information about the content (called 'object content information') and the intellectual property rights associated with it. Each stream itself is characterized by a set of descriptors for configuration information, e.g., to determine the required decoder resources and the precision of encoded timing information. Furthermore the descriptors may carry hints to the Quality of Service (QoS) it requests for transmission (e.g., maximum bit rate, bit error rate, priority, etc.)

© MPEG 1999-2002 – unlimited reproduction permitted if not modified

Google Exhibit No. 1044
Page 007

Google Exhibit No. 1063
Page 1461

Synchronization of elementary streams is achieved through time stamping of individual access units within elementary streams. The synchronization layer manages the identification of such access units and the time stamping. Independent of the media type, this layer allows identification of the type of access unit (e.g., video or audio frames, scene description commands) in elementary streams, recovery of the media object's or scene description's time base, and it enables synchronization among them. The syntax of this layer is configurable in a large number of ways, allowing use in a broad spectrum of systems.

1.4 Delivery of streaming data

The synchronized delivery of streaming information from source to destination, exploiting different QoS as available from the network, is specified in terms of the synchronization layer and a delivery layer containing a two-layer multiplexer, as depicted in Figure 2.
The first multiplexing layer is managed according to the DMIF specification, part 6 of the MPEG-4 standard. (DMIF stands for Delivery Multimedia Integration Framework) This multiplex may be embodied by the MPEG-defined FlexMux tool, which allows grouping of Elementary Streams (ESs) with a low multiplexing overhead. Multiplexing at this layer may be used, for example, to group ES with similar QoS requirements, reduce the number of network connections or the end to end delay.

The "TransMux" (Transport Multiplexing) layer in Figure 2 models the layer that offers transport services matching the requested QoS. Only the interface to this layer is specified by MPEG-4 while the concrete mapping of the data packets and control signaling must be done in collaboration with the bodies that have jurisdiction over the respective transport protocol. Any suitable existing transport protocol stack such as (RTP)/UDP/IP, (AAL5)/ATM, or MPEG-2's Transport Stream over a suitable link layer may become a specific TransMux instance. The choice is left to the end user/service provider, and allows MPEG-4 to be used in a wide variety of operation environments.

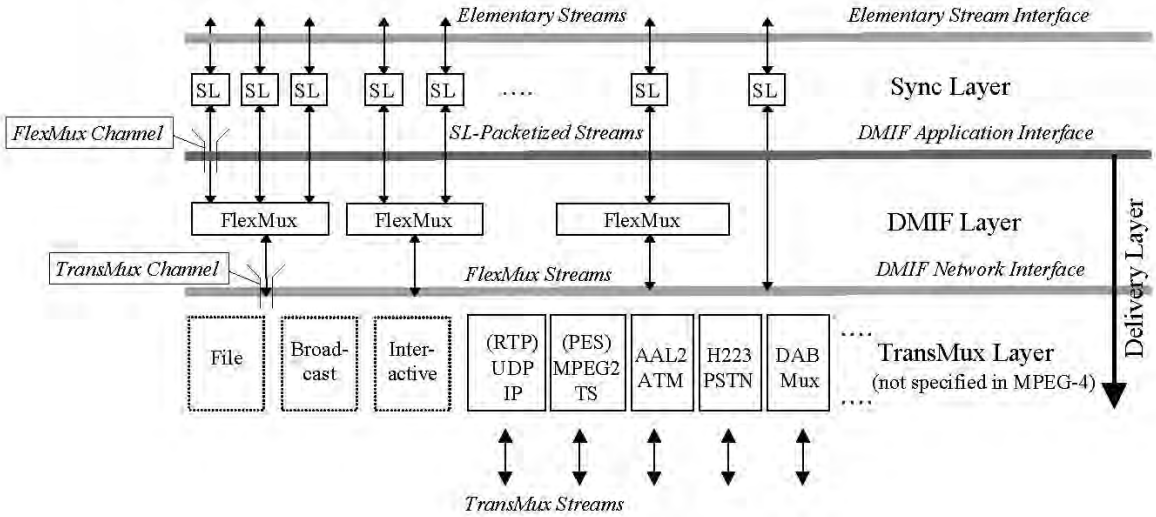


Figure 2 - The MPEG-4 System Layer Model

Use of the FlexMux multiplexing tool is optional and, as shown in Figure 2, this layer may be empty if the underlying TransMux instance provides all the required functionality. The synchronization layer, however, is always present.

© MPEG 1999-2002 – unlimited reproduction permitted if not modified

Google Exhibit No. 1044
Page 008

Google Exhibit No. 1063
Page 1462

With regard to Figure 2, it is possible to:

- Identify access units, transport timestamps and clock reference information and identify data loss.
- Optionally interleave data from different elementary streams into FlexMux streams
- Convey control information to:
- Indicate the required QoS for each elementary stream and FlexMux stream;
- Translate such QoS requirements into actual network resources;
- Associate elementary streams to media objects
- Convey the mapping of elementary streams to FlexMux and TransMux channels

Parts of the control functionalities are available only in conjunction with a transport control entity like the DMIF framework.

1.5 Interaction with media objects

In general, the user observes a scene that is composed following the design of the scene's author. Depending on the degree of freedom allowed by the author, however, the user has the possibility to interact with the scene. Operations a user may be allowed to perform include:

- Change the viewing/listening point of the scene, e.g. by navigation through a scene;
- Drag objects in the scene to a different position;
- Trigger a cascade of events by clicking on a specific object, e.g. starting or stopping a video stream;
- Select the desired language when multiple language tracks are available;

More complex kinds of behavior can also be triggered, e.g. a virtual phone rings, the user answers and a communication link is established.

1.6 Management and Identification of Intellectual Property

It is important to have the possibility to identify intellectual property in MPEG-4 media objects. Therefore, MPEG has worked with representatives of different creative industries in the definition of syntax and tools to support this. A full elaboration of the requirements for the identification of intellectual property can be found in 'Management and Protection of Intellectual Property in MPEG-4, which is publicly available from the MPEG home page.

MPEG-4 incorporates identification the intellectual property by storing unique identifiers, which are issued by international numbering systems (e.g. ISAN, ISRC, etc.1). These numbers can be applied to identify a current rights holder of a media object. Since not all content is identified by such a number, MPEG-4 Version 1 offers the possibility to identify intellectual property by a key-value pair (e.g.:»composer«/»John Smith«). Also, MPEG-4 offers a standardized interface that is integrated tightly into the Systems layer to people who want to use systems that control access to intellectual property. With this interface, proprietary control systems can be easily amalgamated with the standardized part of the decoder.

2 Versions in MPEG-4

MPEG-4 Version 1 was approved by MPEG in December 1998; version 2 was frozen in December 1999. After these two major versions, more tools were added in subsequent amendments that could be qualified as versions, even though they are harder to recognize as

1 ISAN: International Audio-Visual Number, ISRC: International Standard Recording Code

© MPEG 1999-2002 – unlimited reproduction permitted if not modified

Google Exhibit No. 1044
Page 009

Google Exhibit No. 1063
Page 1463

such. Recognizing the versions is not too important, however; it is more important to distinguish Profiles. Existing tools and profiles from any version are never replaced in subsequent versions; technology is always added to MPEG-4 in the form of new profiles. Figure 3 below depicts the relationship between the versions. Version 2 is a backward compatible extension of Version 1, and version 3 is a backward compatible extension of Version 2 – and so on. The versions of all major parts of the MPEG-4 Standard (Systems, Audio, Video, DMIF) were synchronized; after that, the different parts took their own paths.

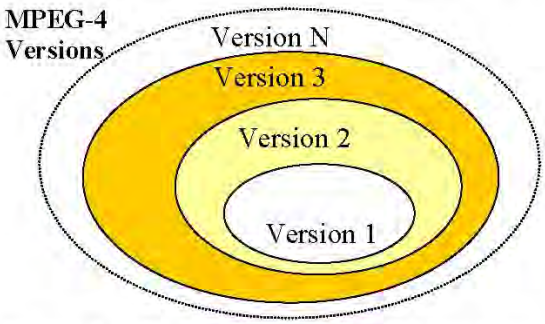


Figure 3 - relation between MPEG-4 Versions

The Systems layer of Version later versions is backward compatible with all earlier versions. In the area of Systems, Audio and Visual, new versions *add* Profiles, do not change existing ones. In fact, it is very important to note that existing systems will ***always*** remain compliant, because Profiles will ***never*** be changed in retrospect, and neither will the Systems Syntax, at least not in a backward-incompatible way.

3 Major Functionalities in MPEG-4

This section contains, in an itemized fashion, the major functionalities that the different parts of the MPEG-4 Standard offers in the finalized MPEG-4 Version 1. Description of the functionalities can be found in the following sections.

3.1 Transport

In principle, MPEG-4 does not define transport layers. In a number of cases, adaptation to a specific existing transport layer has been defined:

- Transport over MPEG-2 Transport Stream (this is an amendment to MPEG-2 Systems)
- Transport over IP (In cooperation with IETF, the Internet Engineering Task Force)

3.2 DMIF

DMIF, or Delivery Multimedia Integration Framework, is an interface between the application and the transport, that allows the MPEG-4 application developer to stop worrying about that transport. A single application can run on different transport layers when supported by the right DMIF instantiation.
MPEG-4 DMIF supports the following functionalities:

© MPEG 1999-2002 – unlimited reproduction permitted if not modified

Google Exhibit No. 1044
Page 010

Google Exhibit No. 1063
Page 1464

- A transparent MPEG-4 DMIF-application interface irrespective of whether the peer is a remote interactive peer, broadcast or local storage media.
- Control of the establishment of FlexMux channels
- Use of homogeneous networks between interactive peers: IP, ATM, mobile, PSTN, Narrowband ISDN.
- Support for mobile networks, developed together with ITU-T
- UserCommands with acknowledgment messages.
- Management of MPEG-4 Sync Layer information

3.3 Systems

As explained above, MPEG-4 defines a toolbox of advanced compression algorithms for audio and visual information. The data streams (Elementary Streams, ES) that result from the coding process can be transmitted or stored separately, and need to be composed so as to create the actual multimedia presentation at the receiver side.

The systems part of the MPEG-4 addresses the description of the relationship between the audio-visual components that constitute a scene. The relationship is described at two main levels.

- The Binary Format for Scenes (BIFS) describes the spatio-temporal arrangements of the objects in the scene. Viewers may have the possibility of interacting with the objects, e.g. by rearranging them on the scene or by changing their own point of view in a 3D virtual environment. The scene description provides a rich set of nodes for 2-D and 3-D composition operators and graphics primitives.
- At a lower level, Object Descriptors (ODs) define the relationship between the Elementary Streams pertinent to each object (e.g the audio and the video stream of a participant to a videoconference) ODs also provide additional information such as the URL needed to access the Elementary Steams, the characteristics of the decoders needed to parse them, intellectual property and others.

Other issues addressed by MPEG-4 Systems:

- A standard file format supports the exchange and authoring of MPEG-4 content
- Interactivity, including: client and server-based interaction; a general event model for triggering events or routing user actions; general event handling and routing between objects in the scene, upon user or scene triggered events.
- Java (MPEG-J) is used to be able to query to terminal and its environment support and there is also a Java application engine to code 'MPEGlets'.
- A tool for interleaving of multiple streams into a single stream, including timing information (FlexMux tool).
- A tool for storing MPEG-4 data in a file (the MPEG-4 File Format, 'MP4')
- Interfaces to various aspects of the terminal and networks, in the form of Java API's (MPEG-J)
- Transport layer independence. Mappings to relevant transport protocol stacks, like (RTP)/UDP/IP or MPEG-2 transport stream can be or are being defined jointly with the responsible standardization bodies.
- Text representation with international language support, font and font style selection, timing and synchronization.
- The initialization and continuous management of the receiving terminal's buffers.
- Timing identification, synchronization and recovery mechanisms.
- Datasets covering identification of Intellectual Property Rights relating to media objects.

© MPEG 1999-2002 – unlimited reproduction permitted if not modified

Google Exhibit No. 1044
Page 011

Google Exhibit No. 1063
Page 1465

3.4 Audio

MPEG-4 Audio facilitates a wide variety of applications which could range from intelligible speech to high quality multichannel audio, and from natural sounds to synthesized sounds. In particular, it supports the highly efficient representation of audio objects consisting of:

3.4.1 General Audio Signals

Support for coding general audio ranging from very low bitrates up to high quality is provided by transform coding techniques. With this functionality, a wide range of bitrates and bandwidths is covered. It starts at a bitrate of 6 kbit/s and a bandwidth below 4 kHz and extends to broadcast quality audio from mono up to multichannel. High quality can be achieved with low delays. Parametric Audio Coding allows sound manipulation at low speeds. Fine Granularity Scalability (or FGS, scalability resolution down to 1 kbit/s per channel)

3.4.2 Speech signals

Speech coding can be done using bitrates from 2 kbit/s up to 24 kbit/s using the speech coding tools. Lower bitrates, such as an average of 1.2 kbit/s, are also possible when variable rate coding is allowed. Low delay is possible for communications applications. When using the HVXC tools, speed and pitch can be modified under user control during playback. If the CELP tools are used, a change of the playback speed can be achieved by using and additional tool for effects processing.

3.4.3 Synthetic Audio

MPEG-4 Structured Audio is a language to describe 'instruments' (little programs that generate sound) and 'scores' (input that drives those objects). These objects are not necessarily musical instruments, they are in essence mathematical formulae, that could generate the sound of a piano, that of falling water – or something 'unheard' in nature.

3.4.4 Synthesized Speech

Scalable TTS coders bitrate range from 200 bit/s to 1.2 Kbit/s which allows a text, or a text with prosodic parameters (pitch contour, phoneme duration, and so on), as its inputs to generate intelligible synthetic speech.

3.5 Visual

The MPEG-4 Visual standard allows the hybrid coding of natural (pixel based) images and video together with synthetic (computer generated) scenes. This enables, for example, the virtual presence of videoconferencing participants. To this end, the Visual standard comprises tools and algorithms supporting the coding of natural (pixel based) still images and video sequences as well as tools to support the compression of synthetic 2-D and 3-D graphic geometry parameters (i.e. compression of wire grid parameters, synthetic text).
The subsections below give an itemized overview of functionalities that the tools and algorithms of in the MPEG-4 visual standard.

3.5.1 Formats Supported

The following formats and bitrates are be supported by MPEG-4 Visual :

- bitrates: typically between 5 kbit/s and more than 1 Gbit/s
- Formats: progressive as well as interlaced video
- Resolutions: typically from sub-QCIF to 'Studio' resolutions (4k x 4k pixels)

© MPEG 1999-2002 – unlimited reproduction permitted if not modified

Google Exhibit No. 1044
Page 012

Google Exhibit No. 1063
Page 1466

3.5.2 Compression Efficiency

- For all bit rates addressed, the algorithms are very efficient. This includes the compact coding of textures with a quality adjustable between "acceptable" for very high compression ratios up to "near lossless".
- Efficient compression of textures for texture mapping on 2-D and 3-D meshes.
- Random access of video to allow functionalities such as pause, fast forward and fast reverse of stored video.

3.5.3 Content-Based Functionalities

- *Content-based coding* of images and video allows separate decoding and reconstruction of arbitrarily shaped video objects.
- *Random access* of content in video sequences allows functionalities such as pause, fast forward and fast reverse of stored video objects.
- *Extended manipulation of content* in video sequences allows functionalities such as warping of synthetic or natural text, textures, image and video overlays on reconstructed video content. An example is the mapping of text in front of a moving video object where the text moves coherently with the object.

3.5.4 Scalability of Textures, Images and Video

- *Complexity scalability in the encoder* allows encoders of different complexity to generate valid and meaningful bitstreams for a given texture, image or video.
- *Complexity scalability in the decoder* allows a given texture, image or video bitstream to be decoded by decoders of different levels of complexity. The reconstructed quality, in general, is related to the complexity of the decoder used. This may entail that less powerful decoders decode only a part of the bitstream.
- *Spatial scalability* allows decoders to decode a subset of the total bitstream generated by the encoder to reconstruct and display textures, images and video objects at reduced spatial resolution. A maximum of 11 levels of spatial scalability are supported in so-called 'fine-granularity scalability', for video as well as textures and still images.
- *Temporal scalability* allows decoders to decode a subset of the total bitstream generated by the encoder to reconstruct and display video at reduced temporal resolution. A maximum of three levels are supported.
- *Quality scalability* allows a bitstream to be parsed into a number of bitstream layers of different bitrate such that the combination of a subset of the layers can still be decoded into a meaningful signal. The bitstream parsing can occur either during transmission or in the decoder. The reconstructed quality, in general, is related to the number of layers used for decoding and reconstruction.
- *Fine Grain Scalability* – a combination of the above in fine grain steps, up to 11 steps

3.5.5 Shape and Alpha Channel Coding

- *Shape coding* assists the description and composition of conventional images and video as well as arbitrarily shaped video objects. Applications that benefit from binary shape maps with images are content-based image representations for image databases, interactive games, surveillance, and animation. There is an efficient technique to code binary shapes. A binary alpha map defines whether or not a pixel belongs to an object. It can be 'on' or 'off'.
- *'Gray Scale' or 'alpha' Shape Coding*
  An alpha plane defines the 'transparency' of an object, which is not necessarily uniform; it can vary over the object, so that, e.g., edges are more transparent (a technique called feathering). Multilevel alpha maps are frequently used to blend different layers of image

© MPEG 1999-2002 – unlimited reproduction permitted if not modified

Google Exhibit No. 1044
Page 013

Google Exhibit No. 1063
Page 1467

sequences. Other applications that benefit from associated binary alpha maps with images are content-based image representations for image databases, interactive games, surveillance, and animation.

3.5.6 Robustness in Error Prone Environments

Error resilience allows accessing image and video over a wide range of storage and transmission media. This includes the useful operation of image and video compression algorithms in error-prone environments at low bit-rates (i.e., less than 64 Kbps). There are tools that address both the band-limited nature and error resiliency aspects of access over wireless networks.

3.5.7 Face and Body Animation

The 'Face and Body Animation' tools in the standard allow sending parameters that can define, calibrate and animate synthetic faces and bodies. These models themselves are not standardized by MPEG-4, only the parameters are, although there is a way to send, e.g., a well-defined face to a decoder.
The tools include:

- Definition and coding of face and body animation parameters (model independent):
  - Feature point positions and orientations to animate the face and body definition meshes
  - Visemes, or visual lip configurations equivalent to speech phonemes
- Definition and coding of face and body definition parameters (for model calibration):
  - 3-D feature point positions
  - 3-D head calibration meshes for animation
  - Personal characteristics
- Facial texture coding

3.5.8 Coding of 2-D Meshes with Implicit Structure

2D mesh coding includes:

- Mesh-based prediction and animated texture transfiguration
  - 2-D Delaunay or regular mesh formalism with motion tracking of animated objects
  - Motion prediction and suspended texture transmission with dynamic meshes.
- Geometry compression for motion vectors:
  - 2-D mesh compression with implicit structure & decoder reconstruction

3.5.9 Coding of 3-D Polygonal Meshes

MPEG-4 provides a suite of tools for coding 3-D polygonal meshes. Polygonal meshes are widely used as a generic representation of 3-D objects. The underlying technologies compress the connectivity, geometry, and properties such as shading normals, colors and texture coordinates of 3-D polygonal meshes.

The Animation Framework eXtension (AFX, see further down) will provide more elaborate tools for 2D and 3D synthetic objects.

4 Extensions Underway

MPEG is currently working on a number of extensions:

© MPEG 1999-2002 – unlimited reproduction permitted if not modified

Google Exhibit No. 1044
Page 014

Google Exhibit No. 1063
Page 1468

4.1.1 New Video codec: "MPEG-4 Advanced Video Coding"

Work is ongoing on MPEG-4 part 10, 'Advanced Video Coding', This codec is being developed jointly with ITU-T, in the so-called Joint Video Team (JVT). The JVT unites the standard world's video coding experts in a single group. The work currently underway is based on earlier work in ITU-T on 'H.26L'. H.26L and MPEG-4 part 10 will be the same. (H.26L will be renamed when it is done. The final name may be H.264, but that is not yet sure). MPEG-4 AVC/H.26L is slated to be ready by the end of 2002.

4.1.2 The Animation Framework eXtension, AFX

The Animation Framework extension (AFX – pronounced 'effects') provides an integrated toolbox for building attractive and powerful synthetic MPEG-4 environments. The framework defines a collection of interoperable tool categories that collaborate to produce a reusable architecture for interactive animated contents. In the context of AFX, a tool represents functionality such as a BIFS node, a synthetic stream, or an audio-visual stream.

AFX utilizes and enhances existing MPEG-4 tools, while keeping backward-compatibility, by offering:

- Higher-level descriptions of animations (e.g. inverse kinematics)
- Enhanced rendering (e.g. multi-texturing, procedural texturing)
- Compact representations (e.g. piecewise curve interpolators, subdivision surfaces)
- Low bitrate animations (e.g. using interpolator compression and dead-reckoning)
- Scalability based on terminal capabilities (e.g. parametric surfaces tessellation)
- Interactivity at user level, scene level, and client-server session level
- Compression of representations for static and dynamic tools

Compression of animated paths and animated models is required for improving the transmission and storage efficiency of representations for dynamic and static tools.

4.1.3 Audio extensions

There are two work items underway for improving audio coding efficiency even further.

a) Bandwidth extension

Bandwidth extension is a tool that gives a better quality perception over the existing audio signal, while keeping the existing signal backward compatible.
MPEG is investigating bandwidth extensions, and may standardize of one or both of:

1. General audio signals, to extend the capabilities currently provided by MPEG-4 general audio coders.
2. Speech signals, to extend the capabilities currently provided by MPEG-4 speech coders.

A single technology that addresses both of these signals is preferred. This technology shall be both forward and backward compatible with existing MPEG-4 technology. In other words, an MPEG-4 decoder can decode an enhanced stream and a new technology decoder can decode an MPEG-4 stream. There are two possible configurations for the enhanced stream: MPEG-4 AAC streams can carry the enhancement information in the DataStreamElement, while all MPEG-4 systems know the concept of elementary streams, which allow second Elementary Stream for a given audio object, containing the enhancement information.

© MPEG 1999-2002 – unlimited reproduction permitted if not modified

Google Exhibit No. 1044
Page 015

Google Exhibit No. 1063
Page 1469

b) Parametric coding

The MPEG-4 standard already provides a parametric coding scheme for coding of general audio signals for low bit-rates (HILN, "Harmonic Individual Lines and Noise"). The extension investigates parametric coding of general audio signals for the higher quality range, to extend the capabilities currently provided by HILN. Whenever possible this technology will build upon the existing MPEG-4 HILN technology.

5 Profiles in MPEG-4

MPEG-4 provides a large and rich set of tools for the coding of audio-visual objects. In order to allow effective implementations of the standard, subsets of the MPEG-4 Systems, Visual, and Audio tool sets have been identified, that can be used for specific applications. These subsets, called 'Profiles', limit the tool set a decoder has to implement. For each of these Profiles, one or more Levels have been set, restricting the computational complexity. The approach is similar to MPEG-2, where the most well known Profile/Level combination is 'Main Profile @ Main Level'. A Profile@Level combination allows:

- a codec builder to implement only the subset of the standard he needs, while maintaining interworking with other MPEG-4 devices built to the same combination, and
- checking whether MPEG-4 devices comply with the standard ('conformance testing').

Profiles exist for various types of media content (audio, visual, and graphics) and for scene descriptions. MPEG does not prescribe or advise combinations of these Profiles, but care has been taken that good matches exist between the different areas.

5.1 Visual Profiles

The visual part of the standard provides profiles for the coding of natural, synthetic, and synthetic/natural hybrid visual content. There are five profiles for natural video content:

1. The **Simple Visual Profile** provides efficient, error resilient coding of rectangular video objects, suitable for applications on mobile networks, such as PCS and IMT2000.
2. The **Simple Scalable Visual Profile** adds support for coding of temporal and spatial scalable objects to the Simple Visual Profile, It is useful for applications which provide services at more than one level of quality due to bit-rate or decoder resource limitations, such as Internet use and software decoding.
3. The **Core Visual Profile** adds support for coding of arbitrary-shaped and temporally scalable objects to the Simple Visual Profile. It is useful for applications such as those providing relatively simple content-interactivity (Internet multimedia applications).
4. The **Main Visual Profile** adds support for coding of interlaced, semi-transparent, and sprite objects to the Core Visual Profile. It is useful for interactive and entertainment-quality broadcast and DVD applications.
5. The **N-Bit Visual Profile** adds support for coding video objects having pixel-depths ranging from 4 to 12 bits to the Core Visual Profile. It is suitable for use in surveillance applications.

The profiles for synthetic and synthetic/natural hybrid visual content are:

6. The **Simple Facial Animation Visual Profile** provides a simple means to animate a face model, suitable for applications such as audio/video presentation for the hearing impaired.
7. The **Scalable Texture Visual Profile** provides spatial scalable coding of still image (texture) objects useful for applications needing multiple scalability levels, such as mapping texture onto objects in games, and high-resolution digital still cameras.

© MPEG 1999-2002 – unlimited reproduction permitted if not modified

Google Exhibit No. 1044
Page 016

Google Exhibit No. 1063
Page 1470

8. The **Basic Animated 2-D Texture Visual Profile** provides spatial scalability, SNR scalability, and mesh-based animation for still image (textures) objects and also simple face object animation.
9. The **Hybrid Visual Profile** combines the ability to decode arbitrary-shaped and temporally scalable natural video objects (as in the Core Visual Profile) with the ability to decode several synthetic and hybrid objects, including simple face and animated still image objects. It is suitable for various content-rich multimedia applications.

Version 2 adds the following Profiles for natural video:

10. The **Advanced Real-Time Simple Profile** (ARTS) provides advanced error resilient coding techniques of rectangular video objects using a back channel and improved temporal resolution stability with the low buffering delay. It is suitable for real time coding applications; such as the videophone, tele-conferencing and the remote observation.
11. The **Core Scalable Profile** adds support for coding of temporal and spatial scalable arbitrarily shaped objects to the Core Profile. The main functionality of this profile is object based SNR and spatial/temporal scalability for regions or objects of interest. It is useful for applications such as the Internet, mobile and broadcast.
12. The **Advanced Coding Efficiency (ACE) Profile** improves the coding efficiency for both rectangular and arbitrary shaped objects. It is suitable for applications such as mobile broadcast reception, the acquisition of image sequences (camcorders) and other applications where high coding efficiency is requested and small footprint is not the prime concern.

The Version 2 profiles for synthetic and synthetic/natural hybrid visual content are:

13. The **Advanced Scaleable Texture Profile** supports decoding of arbitrary-shaped texture and still images including scalable shape coding, wavelet tiling and error-resilience. It is useful for applications that require fast random access as well as multiple scalability levels and arbitrary-shaped coding of still objects. Examples are fast content-based still image browsing on the Internet, multimedia-enabled PDA's, and Internet-ready high-resolution digital still cameras.
14. The **Advanced Core Profile** combines the ability to decode arbitrary-shaped video objects (as in the Core Visual Profile) with the ability to decode arbitrary-shaped scalable still image objects (as in the Advanced Scaleable Texture Profile.) It is suitable for various content-rich multimedia applications such as interactive multimedia streaming over Internet.
15. The **Simple Face and Body Animation Profile** is a superset of the Simple Face Animation Profile, adding - obviously - body animation.

In subsequent Versions, the following Profiles were added:

16. The **Advanced Simple Profile** looks much like Simple in that it has only rectangular objects, but it has a few extra tools that make it more efficient: B-frames, ¼ pel motion compensation, extra quantization tables and global motion compensation.
17. The **Fine Granularity Scalability Profile** allows truncation of the enhancement layer bitstream at any bit position so that delivery quality can easily adapt to transmission and decoding circumstances. It can be used with *Simple* or *Advanced Simple* as a base layer.
18. The **Simple Studio Profile** is a profile with very high quality for usage in Studio editing applications. It only has I frames, but it does support arbitrary shape and in fact multiple alpha channels. Bitrates go up to almost 2 Gigabit per second.
19. The **Core Studio Profile** adds P frames to *Simple Studio*, making it more efficient but also requiring more complex implementations.

© MPEG 1999-2002 – unlimited reproduction permitted if not modified

Google Exhibit No. 1044
Page 017

Google Exhibit No. 1063
Page 1471

5.2 Audio Profiles

Four Audio Profiles have been defined in MPEG-4 V.1:

1. The **Speech Profile** provides HVXC, which is a very-low bit-rate parametric speech coder, a CELP narrowband/wideband speech coder, and a Text-To-Speech interface.
2. The **Synthesis Profile** provides score driven synthesis using SAOL and wavetables and a Text-to-Speech Interface to generate sound and speech at very low bitrates.
3. The **Scalable Profile**, a superset of the Speech Profile, is suitable for scalable coding of speech and music for networks, such as Internet and Narrow band Audio DIgital Broadcasting (NADIB). The bitrates range from 6 kbit/s and 24 kbit/s, with bandwidths between 3.5 and 9 kHz.
4. The **Main Profile** is a rich superset of all the other Profiles, containing tools for natural and synthetic Audio.

Another four Profiles were added in MPEG-4 V.2:

5. The **High Quality Audio Profile** contains the CELP speech coder and the Low Complexity AAC coder including Long Term Prediction. Scalable coding coding can be performed by the AAC Scalable object type. Optionally, the new error resilient (ER) bitstream syntax may be used.
6. The **Low Delay Audio Profile** contains the HVXC and CELP speech coders (optionally using the ER bitstream syntax), the low-delay AAC coder and the Text-to-Speech interface TTSI.
7. The **Natural Audio Profile** contains all natural audio coding tools available in MPEG-4, but not the synthetic ones.
8. The **Mobile Audio Internetworking Profile** (MAUI) contains the low-delay and scalable AAC object types including TwinVQ and BSAC. This profile is intended to extend communication applications using non-MPEG speech coding algorithms with high quality audio coding capabilities.

5.3 Graphics Profiles

Graphics Profiles define which graphical and textual elements can be used in a scene. These profiles are defined in the Systems part of the standard:

1. **Simple 2-D Graphics Profile** The Simple 2-D Graphics profile provides for only those graphics elements of the BIFS tool that are necessary to place one or more visual objects in a scene.
2. **Complete 2-D Graphics Profile** The Complete 2-D Graphics profile provides two-dimensional graphics functionalities and supports features such as arbitrary two-dimensional graphics and text, possibly in conjunction with visual objects.
3. **Complete Graphics Profile** The Complete Graphics profile provides advanced graphical elements such as elevation grids and extrusions and allows creating content with sophisticated lighting. The Complete Graphics profile enables applications such as complex virtual worlds that exhibit a high degree of realism.
4. The **3D Audio Graphics Profile** sounds like a contradictory in terms, but really isn't. This profile does not propose visual rendering, but graphics tools are provided to define the acoustical properties of the scene (geometry, acoustics absorption, diffusion, transparency of the material). This profile is used for applications that do environmental spatialization of audio signals. (See Section 12.1.7)

5.3.1 Profiles under Definition or Consideration

The following profiles were under development at the time of writing this Overview; their inclusion in the standard was highly likely, but not guaranteed.

© MPEG 1999-2002 – unlimited reproduction permitted if not modified

Google Exhibit No. 1044
Page 018

Google Exhibit No. 1063
Page 1472

5. The *Simple 2D+Text* profile looks like simple 2D, adding the BIFS nodes to display text which can be colored or transparent. Like simple 2D, this is a useful profile for low-complexity audiovisual devices.
6. The *Core 2D* Profile supports fairly simple 2D graphics and text. Meant for set tops and similar devices, it can do such things as picture-in-picture, video warping for animated advertisements, logos, and so on.
7. The *Advanced 2D* profile contains tools for advanced 2D graphics. Using it, one can implement cartoons, games, advanced graphical user interfaces, and complex, streamed graphics animations.
8. The *X3D Core* profile is the only 3D profile that is likely to be added to MPEG-4. It is compatible with Web3D's X3D core profile under development [Web3D], and it gives a rich environment for games, virtual worlds and other 3D applications.

5.4 Scene Graph Profiles

Scene Graph Profiles (or Scene Description Profiles), defined in the Systems part of the standard, allow audiovisual scenes with audio-only, 2-dimensional, 3-dimensional or mixed 2-D/3-D content.

1. The *Audio* Scene Graph Profile provides for a set of BIFS scene graph elements for usage in audio only applications. The Audio Scene Graph profile supports applications like broadcast radio.
2. The *Simple 2-D* Scene Graph Profile provides for only those BIFS scene graph elements necessary to place one or more audio-visual objects in a scene. The Simple 2-D Scene Graph profile allows presentation of audio-visual content with potential update of the complete scene but no interaction capabilities. The Simple 2-D Scene Graph profile supports applications like broadcast television.
3. The *Complete 2-D* Scene Graph Profile provides for all the 2-D scene description elements of the BIFS tool. It supports features such as 2-D transformations and alpha blending. The Complete 2-D Scene Graph profile enables 2-D applications that require extensive and customized interactivity.
4. The *Complete* Scene Graph profile provides the complete set of scene graph elements of the BIFS tool. The Complete Scene Graph profile enables applications like dynamic virtual 3-D world and games.
5. The *3D Audio* Scene Graph Profile provides the tools three-dimensional sound positioning in relation either with acoustic parameters of the scene or its perceptual attributes. The user can interact with the scene by changing the position of the sound source, by changing the room effect or moving the listening point. This Profile is intended for usage in audio-only applications.

5.4.1 Profiles under definition

At the time of writing, the following profiles were likely to be defined:

6. The *basic 2D* profile provides basic 2D composition for very simple scenes with only audio and visual elements. Only basic 2D composition and audio and video nodes interfaces are included. These nodes are required to put an audio or a video object in the scene.
7. The *Core 2D* profile has tools for creating scenes with visual and audio objects using basic 2D composition. Included are quantization tools, local animation and interaction, 2D texturing, Scene tree updates, and the inclusion of subscenes through weblinks. Also included are interactive service tools (ServerCommand, MediaControl, and MediaSensor), to be used in video-on-demand services.
8. The *Advanced 2D* profile forms a full superset of the basic 2D and core 2D profiles. It adds scripting, the PROTO tool, BIF-Anim for streamed animation, local interaction and local 2D composition as well as advanced audio.

© MPEG 1999-2002 – unlimited reproduction permitted if not modified

Google Exhibit No. 1044
Page 019

Google Exhibit No. 1063
Page 1473

9. The *Main 2D* profile adds the FlexTime model to Core 2D, as well as Layer2D and WorldInfo nodes and all input sensors. This profile was designed to be an interoperability point with tSMIL (see [SMIL]). It provides a very rich set of tools for highly interactive applications on, e.g., the World Wide Web. This name might still change.
10. The *X3D core* profile was designed to be a common interworking point with the Web3D specifications [Web3D] and the MPEG-4 standard. The same profile is will be in a Web3D specification. It includes the nodes for an implementation of 3D applications on a low-footprint engine, reckoning with the limitations of software renderers.

5.5 MPEG-J Profiles

Two MPEG-J Profiles exist: Personal and Main:

1. *Personal* - a lightweight package for personal devices.
   The personal profile addresses a range of constrained devices including mobile and portable devices. Examples of such devices are cell video phones, PDAs, personal gaming devices. This profile includes the following packages of MPEG-J APIs:
   1. Network
   2. Scene
   3. Resource

2. *Main* - includes all the MPEG-J API's.
   The Main profile addresses a range of consumer devices including entertainment devices. Examples of such devices are set top boxes, computer based multimedia systems etc. It is a superset of the Personal profile. Apart from the packages in the Personal profile, this profile includes the following packages of the MPEG-J APIs:
   4. Decoder
   5. Decoder Functionality
   6. Section Filter and Service Information

5.6 Object Descriptor Profile

The Object Descriptor Profile includes the following tools:

- Object Descriptor (OD) tool
- Sync Layer (SL) tool
- Object Content Information (OCI) tool
- Intellectual Property Management and Protection (IPMP) tool

Currently, only one profile is defined that includes all these tools. The main reason for defining this profile is not subsetting the tools, but rather defining levels for them. This applies especially to the Sync Layer tool, as MPEG-4 allows multiple time bases to exist. In the context of Levels for this Profile, restrictions can be defined, e.g. to allow only a single time base.

6 Verification Testing: checking MPEG's performance

MPEG carries out verification tests to check whether the standard delivers what it promises. The test results can be found on MPEG's home page,
http://www.cselt.it/mpeg/quality_tests.htm
The main results are described below; more verification tests are planned.

© MPEG 1999-2002 – unlimited reproduction permitted if not modified

Google Exhibit No. 1044
Page 020

Google Exhibit No. 1063
Page 1474

6.1 Video

A number of MPEG-4's capabilities have been formally evaluated using subjective tests. Coding efficiency, although not the only MPEG-4 functionality, is an important selling point of MPEG-4, and one that has been tested more thoroughly. Also error robustness has been put to rigorous tests. Furthermore, scalability tests were done and for one specific profile the temporal resolution stability was examined. Many of these tests address a specific profile.

6.1.1 Coding Efficiency Tests

a) Low and Medium Bit rates (version 1)

In this Low and Medium Bitrates Test, frame-based sequences were examined, with MPEG-1 as a reference. (MPEG-2 would be identical for the progressive sequences used, except that MPEG-1 is a bit more efficient as it uses less overhead for header information). The test uses typical test sequences for CIF and QCIF resolutions, encoded with the same rate control for both MPEG-1 and MPEG-4 to compare the coding algorithms without the impact of different rate control schemes. The test was performed for low bit rates starting at 40 kbps to medium bit rate up to 768 kbps.

The tests of the Coding Efficiency functionality show a clear superiority of MPEG-4 toward MPEG-1 at both the low and medium bit rate coding conditions whatever the criticality of the scene. The human subjects have consistently chose MPEG-4 as statistically significantly superior by one point difference for a full scale of five points.

b) Content Based Coding (version 1)

The verification tests for Content Based Coding compare the visual quality of object-based versus frame-based coding. The major objective was to ensure that object-based coding can be supported without impacting the visual quality. Test content was chosen to cover a wide variety of simulation conditions, including video segments with various types of motions and encoding complexities. Additionally, test conditions were established to cover low bit rates ranging from 256kb/s to 384kb/s, as well as high bit-rates ranging from 512kb/s to 1.15Mb/s. The results of the tests clearly demonstrated that object-based functionality is provided by MPEG-4 with no overhead or loss in terms of visual quality, when compared to frame-based coding. There is no statistically significant difference among any object-based case and the relevant frame-based ones. Hence the conclusion: MPEG-4 is able to provide content-based functionality without introducing any loss in terms of visual quality.

c) Advanced Coding Efficiency (ACE) Profile (version 2)

The formal verification tests on Advanced Coding Efficiency (ACE) Profile were performed to check whether three new Version 2 tools, as included the MPEG-4 Visual Version 2 ACE Profile (Global Motion Compensation, Quarter Pel Motion Compensation and Shape-adaptive DCT) enhance the coding efficiency compared with MPEG-4 Visual Version 1. The tests explored the performance of the ACE Profile and the MPEG-4 Visual Version 1 Main Profile in the object-based low bit rate case, the frame-based low bit rate case and the frame-based high bit rate case. The results obtained show a clear superiority of the ACE Profile compared with the Main Profile; more in detail:

- For the object based case, the quality provided by the ACE Profile at 256 kb/s is equal to the quality provided by Main Profile at 384 kb/s.
- For the frame based at low bit rate case, the quality provided by the ACE Profile at 128 kb/s and 256 kb/s is equal to the quality provided by Main Profile at 256 kb/s and 384 kb/s respectively.
- For the frame based at high bit rate case, the quality provided by the ACE Profile at 768 kb/s is equal to the quality provided by Main Profile at 1024 kb/s.

© MPEG 1999-2002 – unlimited reproduction permitted if not modified

Google Exhibit No. 1044
Page 021

Google Exhibit No. 1063
Page 1475

When interpreting these results, it must be noted that the MPEG-4 Main Profile is already more efficient than MPEG-1 and MPEG-2.

6.1.2 Error Robustness Tests

a) Simple Profile (version 1)

The performance of error resilient video in the MPEG-4 Simple Profile was evaluated in subjective tests simulating MPEG-4 video carried in a realistic multiplex and over ditto radio channels, at bitrates between 32 kbit/s and 384 kbit/s. The test used a simulation of the residual errors after channel coding at bit error rates up to 10^{-3}, and the average length of the burst errors was about 10ms. The test methodology was based on a continuous quality evaluation over a period of three minutes. In such a test, subjects constantly score the degradation they experience.

The results show that the average video quality achieved on the mobile channel is high, that the impact of errors is effectively kept local by the tools in MPEG-4 video, and that the video quality recovers quickly at the end of periods of error. These excellent results were achieved with very low overheads, less than those typically associated with the GOP structure used in MPEG-1 and MPEG-2 video.

b) Advanced Real-Time Simple (ARTS) Profile (version 2)

The performance of error resilient video in MPEG-4 ARTS Profile was checked in subjective tests similar to those mentioned in the previous section, at bitrates between 32 kbit/s and 128 kbit/s. In this case, the residual errors after channel coding was up to 10^{-3}, and the average length of the burst errors was about 10 ms (called "critical") or 1 ms (called "very critical" - this one is more critical because the same amount of errors is more spread over the bitstream than in the "critical" case).

The results show a clear superiority of the ARTS Profile over the Simple Profile for both the error cases ("critical" and "very critical"). More in detail the ARTS Profile outperforms Simple Profile in the recovery time from transmission errors. Furthermore ARTS Profile in the "critical" error condition provides results that for most of the test time are close to a complete transparency, while Simple Profile is still severely affected by errors. These excellent results were achieved with very low overheads and very fast error recovery provided the NEWPRED, and under low delay conditions.

6.1.3 Temporal Resolution Stability Test

a) Advanced Real-Time Simple (ARTS) Profile (version 2)

This test explored the performance of a video codec using the Dynamic Resolution Conversion technique that adapts the resolution to the video content and to circumstances in real-time. Active scene content was coded at 64 kb/s, 96 kb/s and 128 kb/s datarates. The results show that at 64 kbit/s, it outperforms the already effective Simple Profile operating at 96 kbit/s, and at 96 kb/s, the visual quality is equally to that of the Simple profile at 128 kbit/s. (The Simple profile already compares well to other, existing systems.)

6.1.4 Scalability Tests

a) Simple Scalable Profile (version 1)

The scalability test for the Simple Scalable Profile was designed to verify that the quality provided by Temporal Scalability tool in Simple Scalable Profile compared to the quality provided by Single Layer coding in Simple Profile, and to the quality provided by Simulcast coding in Simple Profile.

In this test, 5 sequences with 4 combinations of bitrates were used:

a) 24 kbps for base layer and 40 kbps for enhancement layer.

© MPEG 1999-2002 – unlimited reproduction permitted if not modified

Google Exhibit No. 1044
Page 022

Google Exhibit No. 1063
Page 1476

b) 32 kbps for both layers.
c) 64 kbps for the base layer and 64 kbps for the enhancement layer.
d) 128 kbps for both layers.

The formal verification tests showed that in all the given conditions, the Temporal Scalability coding in Simple Scalable Profile exhibits the same or slightly lower quality than can be achieved using Single layer coding in Simple Profile. Furthermore it is evident that the Temporal Scalability coding in Simple Scalable Profile provides better quality than the simulcast coding in Simple Profile for that condition. (Simulcast entails simultaneously broadcasting or streaming at multiple bitrates.)

b) Core Profile (version 1)

The verification test was designed to evaluate the performance of MPEG-4 video Temporal Scalability tool in the Core Profile.
The test was performed using the "Single Stimulus" method. The subjects were to evaluate how annoying the impairments of compressed sequences were, with and without use of temporal scalability. The test was conducted using a total of 45 subjects in two different laboratories and the results showed that the quality of sequences encoded using MPEG-4 temporal scalability tools are comparable to the quality of sequences encoded without temporal scalability. Furthermore it is evident that the Temporal Scalability tool in Core Profile provides better quality than the simulcast coding in Core Profile for that condition.

6.2 Audio

MPEG-4 audio technology is composed of many coding tools. Verification tests have focused on small sets of coding tools that are appropriate in one application arena, and hence can be effectively compared. Since compression is a critical capability in MPEG, the verification tests have for the most part compared coding tools operating at similar bit rates. The results of these tests will be presented progressing from higher bit rate to lower bit rates. The exception to this is the error robustness tools, whose performance will be noted at the end of this section.

The primary purpose of verification tests is to report the subjective quality of a coding tool operating at a specified bit rate. The audio tests concerned with very high subjective audio quality report this on the ITU-R BS.1116 subjective *impairment scale* while for lower bitrates the ITU-R BS.1284 *quality scale* was employed. Both are continuous 5-point scales with subjective anchors as shown here:



BS.1116 Impairment Scale

5	Excellent
4	Good
3	Fair
2	Poor
1	Bad

BS.1284 Quality Scale

Please note that it is not meaningful to compare grades across those scales.

The performance of the various MPEG-4 coding tools are summarized in the following table. To better enable the evaluation of MPEG-4 technology, several coders from MPEG-1/2 and the ITU-T were included in the tests and their evaluation has also been included in the table. In the table results from the same test are delimited by heavy lines. These results can be directly

© MPEG 1999-2002 – unlimited reproduction permitted if not modified

Google Exhibit No. 1044
Page 023

Google Exhibit No. 1063
Page 1477

compared. Results taken from different tests should not be compared, but nevertheless give an indication of the expected quality of a coding tool operating at a specific bit rate.

Coding tool	Number of channels	Total bit rate	Grading scale type	Typical subjective quality
AAC	5	320 kb/s	impairment	4.6
1995 Backward Compatible MPEG-2 Layer II	5	640 kb/s	impairment	4.6
AAC	2	128 kb/s	impairment	4.8
AAC	2	96 kb/s	impairment	4.4
MPEG-1 Layer II	2	192 kb/s	impairment	4.3
MPEG-1 Layer III	2	128 kb/s	impairment	4.1
AAC	1	24 kb/s	quality	4.2
Scalable: CELP base and AAC enhancement	1	6 kb/s base, 18 kb/s enh.	quality	3.7
Scalable: Twin VQ base and AAC enhancement	1	6 kb/s base, 18 kb/s enh.	quality	3.6
AAC	1	18 kb/s	quality	3.2
G.723	1	6.3 kb/s	quality	2.8
Wideband CELP	1	18.2 kb/s	quality	2.3
BSAC	2	96 kb/s	quality	4.4
BSAC	2	80 kb/s	quality	3.7
BSAC	2	64 kb/s	quality	3.0
AAC – LD (20 ms one-way delay)	1	64 kb/s	quality	4.4
G.722	1	64 kb/s	quality	4.2
AAC – LD (30 ms one-way delay)	1	32 kb/s	quality	3.4
Narrowband CELP	1	6 kb/s	quality	2.5
Twin VQ	1	6 kb/s	quality	1.8
HILN	1	16 kb/s	quality	2.8
HILN	1	6 kb/s	quality	1.8

Coding tools were tested under circumstances that assessed their strengths. The salient features of the MPEG-4 audio coding tools are briefly noted here.

When coding 5-channel material at 64 kb/s/channel (320 kbit/s) Advanced Audio Coding (AAC) Main Profile was judged to have "indistinguishable quality" (relative to the original) according to the EBU definition. When coding 2-channel material at 128 kbps both AAC Main Profile and AAC Low Complexity Profile were judged to have "indistinguishable quality" (relative to the original) according to the EBU definition.

The two scaleable coders, CELP base with AAC enhancement, and TwinVQ base with AAC enhancement both performed better than an AAC "multicast" operating at the enhancement layer bitrate, but not as good as an AAC coder operating at the total bitrate.

The wideband CELP coding tool showed excellent performance for speech-only signals. (The verification test result shown is for both speech and music signals.)

© MPEG 1999-2002 – unlimited reproduction permitted if not modified

Google Exhibit No. 1044
Page 024

Google Exhibit No. 1063
Page 1478

Bit Slice Arithmetic Coding (BSAC) provides a very fine step bitrate scalability. At the top of the scalability range it has no penalty relative to single-rate AAC, however at the bottom of the scale it has a slight penalty relative to single-rate AAC.

Relative to normal AAC, Low Delay AAC (AAC LD) provides equivalent subjective quality, but with very low on-way delay and only a slight increase in bit rate.

Narrowband CELP, TwinVQ and Harmonic Individual Lines and Noise (HILN) all have the ability to provide very high signal compression.

The Error Robustness (ER) tools provide equivalently good error robustness over a wide range of channel error conditions, and does so with only a modest overhead in bit rate. Verification test results suggest that the ER tools used with an audio coding system provide performance in error-prone channels that is "nearly as good" as the same coding system operating over a clear channel.

7 The MPEG-4 Industry Forum

The MPEG-4 Industry Forum is a not-for-profit organization with the following goal: *To further the adoption of the MPEG-4 Standard, by establishing MPEG-4 as an accepted and widely used standard among application developers, service providers, content creators and end users.*

The following is a non-exhaustive excerpt from M4IF's Statutes about the way of operation:

- The purpose of M4IF shall be pursued by: promoting MPEG-4, making available information on MPEG-4, making available MPEG-4 tools or giving information on where to obtain these, creating a single point for information about MPEG-4, creating industrial focus around the usage of MPEG-4
- The goals are realized through the open international collaboration of all interested parties, on reasonable terms applied uniformly and openly. M4IF will contribute the results of its activities to appropriate formal standards bodies if applicable.
- The business of M4IF shall not be conducted for the financial profits of its Members but for their mutual benefits.
- Any corporation and individual firm, partnership, governmental body or international organization supporting the purpose of M4IF may apply for Membership.
- Members are not bound to implement or use specific technology standards, or recommendations by virtue of participation in M4IF.
- There is no licensing requirement attached to M4IF membership, and M4IF will not set any licensing terms for MPEG-4 technology. (There can, however, be studies about what licensing models would suite MPEG-4's operational environments)
- The membership fee is currently US$ 3,000 per year, and US$ 300 for not-for-profit organizations.

M4IF has its homepage: http://www.m4if.org
Interested parties can visit that page for details, and they can also sign up to public mail lists (an information list and a technical list). The site also has all the information about formally joining the Forum.

© MPEG 1999-2002 – unlimited reproduction permitted if not modified

Google Exhibit No. 1044
Page 025

Google Exhibit No. 1063
Page 1479

The Forum currently has more than 100 members and some 1,500 people subscribed to its mail lists, with a very broad, worldwide representation from the following industries: Consumer Electronics, Computer, Telecommunications, Research Institutions . Also, some of the members are 'business users' of MPEG-4. Among the list of current participants, there are many large and small companies that develop or deploy MPEG-4 technology. Membership goes beyond the MPEG constituency, partly because some of the smaller companies find it hard to comply with the requirements for participation in MPEG (which vary from one country to another) and also because some companies just don't need to be involved in the standardization phase.

The activities of M4IF generally start where MPEG stops. This includes issues that MPEG cannot deal with, e.g. because of ISO rules, such as discussing licensing issues. This following is a list of M4IF's current activities:

- Promoting the standard, and serving as a single point of information on MPEG-4 technology, products and services;
- Initiating discussions leading to the potential establishment of patent pools outside of M4IF, that should grant a license to an unlimited number of applicants throughout the world under reasonable terms and conditions that are demonstrably free of any unfair competition;
  this work includes studying licensing models for downloadable software decoders, such as internet players;
- Organization of MPEG-4 exhibitions and tutorials. There have been 2 successful instances of the Workshop and Exhibition on MPEG-4; the third will be held in June 2002;
- Interoperability testing, leading to certification of MPEG-4 products.

M4IF anticipates holding some 3 physical meetings per year, with a slightly higher frequency in the current start-up phase. About a hundred people from all over the world have attended these meetings. The focus has been on initiating the patent pools, but it will now shift towards interoperability tests. See the home page for meeting details.

8 Licensing of patents necessary to implement MPEG-4

Disclaimer: the author and editor of this Section has done his best to adequately reflect the situation around licensing of MPEG-4, but he is not a lawyer and the explanation below may not be fully correct. Although he has been very careful in writing this all down, neither the author nor MPEG can be held responsible for the consequence of any errors in this text. Please consult a lawyer if you want to understand exactly how this works.

8.1 Roles in Licensing MPEG-4

Licensing MPEG-4 is an important issue, and also one that is cause for much confusion. First, on should understand the roles of the different organizations involved in getting MPEG-4 deployed. Below is a short clarification of the role of some of the main players.

ISO/IEC MPEG is the group that makes MPEG standards. MPEG does not (and cannot, under ISO rules) deal with patents and licensing, other than requiring companies whose technologies are adopted into the standard to sign a statement that they will license their patents on Reasonable and Non-Discriminatory Terms (also called RAND terms) to all parties that wish to create a standards-compliant device, (hardware or software) or create a standards-compliant

© MPEG 1999-2002 – unlimited reproduction permitted if not modified

Google Exhibit No. 1044
Page 026

Google Exhibit No. 1063
Page 1480

bitstream. 'Non-Discriminatory' means that the patent needs to be licensed to al parties that wish to implement MPEG-4 on the same terms. 'Reasonable' is not further defined anywhere.

In developing MPEG-4 part 10 / H.26L, the ITU/MPEG Joint Video Team is now attempting to establish a royalty-free baseline coder, and to this end it also asks of proposers to submit a statement that specifies whether they would want to make available any necessary patents on a royalty-free basis, if (and only if) others are prepared to license under the same terms.

The MPEG-4 Industry Forum, M4IF has in its statutes that it shall not license patents or determine licensing fees, but has nonetheless played an important role in driving the availability of licenses for the patents needed to implement MPEG-4. M4IF can discuss issues pertaining to licensing, and has acted as a catalyst, in the very literal sense, in getting patent pools going. In the years 1999 and 2000, M4IF adopted a series of resolutions recommending on ways that a joint license ('patent pool') might be established; it also mentioned names of parties/people that could play a role in this process. The process was detailed for the Systems, Visual and Audio parts of the standard, and involved an independent evaluator and a neutral administrator in all cases. Even though it will not actively pursue these, M4IF encourages alternative patent pools to be created, the more the better, and if possible even competing ones. (Competition is good, also in licensing, for the same reasons as why technology competition is good). It should be noted that no-one is forced to do business with any patent pool; one can also go straight to all the individual licensors (e.g., at least 18 in MPEG-4 Visual) as the licenses are always non-exclusive. However, doing so is cumbersome and there is a risk that negotiating 18 individual licenses turn out more costly than doing business with a one-stop joint licensing scheme.

M4If has recommended that licensors create patent joint licensing schemes for specific profiles, as profiles are the interoperability points of MPEG-4, and only when one implements a profile there is a standards-compliant implementation. Also, M4IF held a poll among its members to determine for which profile there was the most interest.

There are the 'patent pools' (joint licensing schemes) and their administrators. It is the licensors that determine who will be their licensing agent(s), and they alone. Other parties (including ISO/IEC MPEG, M4IF or the licensees) have no say in this choice. MPEG LA is an example of a licensing administrator, licensing MPEG-2 Video and MPEG-2 Systems. MPEG LA also announced licensing for MPEG-4 Visual. Dolby licenses MPEG-2 AAC, and has announced a joint licensing scheme for some of the patents needed to implement MPEG-4 AAC. Thomson licenses MP3 (MPEG-1 Layer III Audio). Patent holders determine the fees; MPEG LA, Dolby, Thomson (and there are others) collect on behalf of the patent owners and distribute the proceeds.

Patent pools usually (always?) allow new patents to enter the pool when they are found essential. Sometimes this is because patents were submitted later, sometimes because they issued only at a later date. A license pool is never (at least in the case of MPEG standards, and as far as the author knows) 'closed' after a certain date. In MPEG-2, many patents were added after the start of licensing.

It should be clear – yet cannot be stressed enough – that neither M4IF not MPEG receives even one cent of the collected royalties, nor does do they want to. M4IF has among its members licensors, licensees and entities that are neither of those. Collectively, the members have an interest in fair and reasonable licensing, because the standard will fail without it.

© MPEG 1999-2002 – unlimited reproduction permitted if not modified

Google Exhibit No. 1044
Page 027

Google Exhibit No. 1063
Page 1481

8.2 Licensing Situation

This section contains an overview of the current situation with respect to MPEG-4 licensing. See for last-minute information http://www.m4if.org/patents/index.php

8.2.1 Systems

M4IF issued a recommendation in September 2000. An independent evaluator looked at all submissions, and a number of patents form 6 companies were found essential. A press release was issued by MPEGLA on behalf of these patent holders in November 2001. Nothing was heard since; the wait is for more news. The hope is that more news will follow once licensing for Visual has been finalized.

8.2.2 Visual

The first announcement of a plan that should lead to a joint license was done on 13 April. The licensing activity was then hoped to start in October 2000. This schedule has been delayed considerably, but at least 18 holders of essential patents in Simple and Core profile were identified, and provisional licensing terms were announced on 31 January 2002. They have caused much commotion; many parties believed that these are not fit for several of MPEG-4's application areas. At the time of writing of this document (April 2002) the market was waiting for adaptations.

In the meantime, a call for essential patents for Advanced Simple, Fine Granularity Scalability and Simple Scalable was also issued. No statements about the result of this call have surfaced yet.

8.2.3 Audio

A press release announcing plans for a joint licensing scheme was announced in March 2001, seeking to encourage joint licensing schemes to be established for the Profiles Speech, Mobile Audio Internetworking and High Quality Audio. Progress was reported on 19 October 2001, and since then the group has been silent. In April 2002 Dolby announced that it would start licensing patents from 5 companies necessary to implement MPEG-4 AAC. While the terms were met with enthusiasm by the market, it is likely that more license holders exist for MPEG-4 AAC and MPEG-4 AAC does not constitute a profile in itself; rather, it is part of High Quality Audio and MAUI Profiles.

9 Deployment of MPEG-4

MPEG-4 is deployed in a number of environments. Full MPEG-4 Systems, with MPEG-4 Visual, Audio and Systems technology can be bought today from a number of providers, see http://www.m4if.org/products/ There are many providers for MPEG-4 Visual and MPEG-4 Audio codecs, and for, e.g., MPEG-4 Face Animation.

MPEG-4 Video and the MPEG-4 file format can be found in the 3GPP and 3GPP2 specifications for mobile multimedia terminals. A very widespread application of MPEG-4 Visual is the DivX coder, see www.divxnetwords.com

This implementation of MPEG-4 Visual is combined with MP3 (MPEG-1 Layer III) Audio, and it looks like the file format is headed to MP4 compliance. There is some confusion about whether this is really a compliant MPEG-4 implementation. The authors of the software have

© MPEG 1999-2002 – unlimited reproduction permitted if not modified

Google Exhibit No. 1044
Page 028

Google Exhibit No. 1063
Page 1482

joined the interoperability work in M4IF, and recently announced support for MPEG-4 Advanced Simple Profile.

Perhaps the first MPEG-4 deployment was the 'Eggy', which was sold by NTT DoCoMo. It was a small retrieval terminal for MPEG-4 visual content. NTT DoCoMo has also released the FOMA, a multimedia 3G cell phone which uses MPEG-4 Visual.

The Internet Streaming Media Alliance (ISMA), founded in by Apple, Cisco, IBM, Kasenna, Philips and Sun, has specified a fully end-to-end interoperable system for internet streaming, including MPEG-4 Visual, MPEG-4 Audio and the transport of RTP as described above.

Microsoft's Windows Media software contains an MPEG-4 encoder and decoder. This is an implementation of MPEG-4 Simple Profile, the simplest and lowest cost Visual Profile in MPEG-4. While Microsoft indicates that this implementation is fully compliant, it is 'packaged' in a Microsoft's proprietary way.

Real Networks supports MPEG-4 through a certified plug-in from Envivio. Real has announced native MPEG-4 support for an unspecified date.

Apple's Quicktime version 6, which includes MPEG-4 support, has been demonstrated at e.g. NAB 2002. Apple has stated to not release QT6 until the licensing situation is resolved to Apple's satisfaction.

MPEG members are kindly requested to send more deployment information to the editor of this document.

10 Detailed technical description of MPEG-4 DMIF and Systems

This Section contains a detailed overview of the MPEG-4 Standard. It first described the entire system, and then describes all parts of the system in subsections.

© MPEG 1999-2002 – unlimited reproduction permitted if not modified

Google Exhibit No. 1044
Page 029

Google Exhibit No. 1063
Page 1483

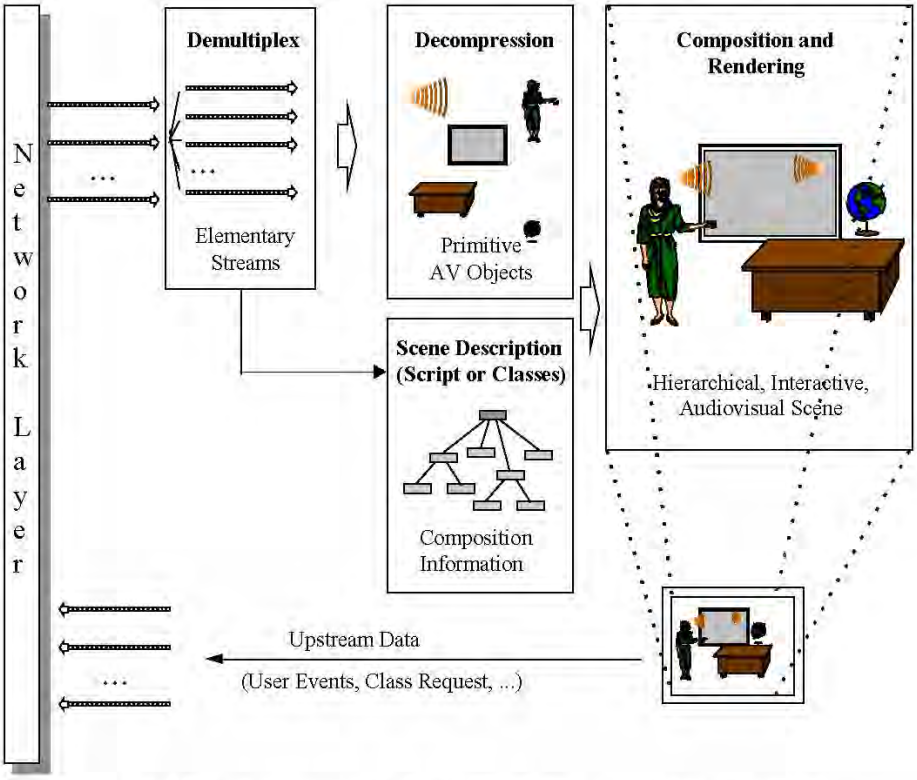


Figure 4- Major components of an MPEG-4 terminal (receiver side)

Figure 4 shows how streams coming from the network (or a storage device), as TransMux Streams, are demultiplexed into FlexMux Streams and passed to appropriate FlexMux demultiplexers that retrieve Elementary Streams. How this works is described in Section 10.3. The Elementary Streams (ESs) are parsed and passed to the appropriate decoders. Decoding recovers the data in an AV object from its encoded form and performs the necessary operations to reconstruct the original AV object ready for rendering on the appropriate device. Audio and visual objects are represented in their coded form, which is described in sections 12 and 10.11 respectively. The reconstructed AV object is made available to the composition layer for potential use during scene rendering. Decoded AVOs, along with scene description information, are used to compose the scene as described by the author. Scene description and Composition are explained in Section 10.6. The user can, to the extent allowed by the author, interact with the scene which is eventually rendered and presented. Section 10.6.1 describes this interaction. Sections 10.8 and 10.11 discuss the 'Intellectual Property Management and Protection' and Object Content Information respectively. Sections 8.9 and 8.10 describe Version 2 additions: the File Format and MPEG-J.

10.1 Transport of MPEG-4

MPEG-4 is transport-agnostic, and designed that way on purpose. This means that MPEG-4 content can be carried over many different transport layers, and move form one transport to the other. There are some cases, however, where MPEG did some work relating to the transport of MPEG-4 content.

© MPEG 1999-2002 – unlimited reproduction permitted if not modified

Google Exhibit No. 1044
Page 030

Google Exhibit No. 1063
Page 1484

10.1.1 MPEG-4 on MPEG-2

An often-asked question is 'when will MPEG-4 replace MPEG-2'? Then answer is invariably: "certainly not anytime soon, as there are tens of billions of dollars invested in MPEG-2, and MPEG is the last committee to want to make these investments useless." Rather, MPEG has taken care that MPEG-2 investments can be leveraged when deploying MPEG-4 technology. Notably, MPEG has taken care that MPEG-4 content can be transported over an MPEG-2 transport stream (MPEG-2 TS). This is done by amending the MPEG-2 Systems standard, 13818-1. This amendment is part of ISO/IEC 13818-1:2000, the 200 edition of MPEG-2 Systems.

10.1.2 MPEG-4 over IP

The specifications on the carriage of MPEG-4 contents over IP networks are developed jointly with IETF AVT working group. These include a framework and several RTP payload format specifications. The framework is standardized as both a part 8 of MPEG-4, i.e. ISO/IEC 14496-8 and informative RFC in IETF. RTP payload format specifications are only standardized as a standard track RFC in IETF.
Framework is an umbrella specification for the carriage and operation of MPEG-4 sessions over IP-based protocols, including RTP, RTSP, and HTTP, among others. It provides a framework for the carriage of MPEG-4 contents over IP networks and guidelines for designing payload format specifications for the detailed mapping of MPEG-4 content into several IP-based protocols. To assure compatibility between different RTP payload formats, framework defines a conformance point as illustrated in the Figure 1. To conform this framework all the payload formats shall provide normative mapping functions to reconstruct logical MPEG-4 SL packets. Framework also defines the standard MIME types associated with MPEG-4 contents.

Several RTP payload formats are developed under this framework including generic payload format and FlexMux payload format. Generic RTP payload format specify a homogeneous carriage of various MPEG-4 streams. It defines a simple but efficient mapping between logical MPEG-4 SL packets and RTP packets. It also supports concatenation of multiple SL packets into one RTP packets to minimize overheads. FlexMux payload format specifies a carriage of FlexMux packetized streams via RTP packets. It includes a payload formats to convey FlexMux descriptors to dynamically signal the configuration of FlexMux.

Google Exhibit No. 1044
Page 031

Google Exhibit No. 1063
Page 1485

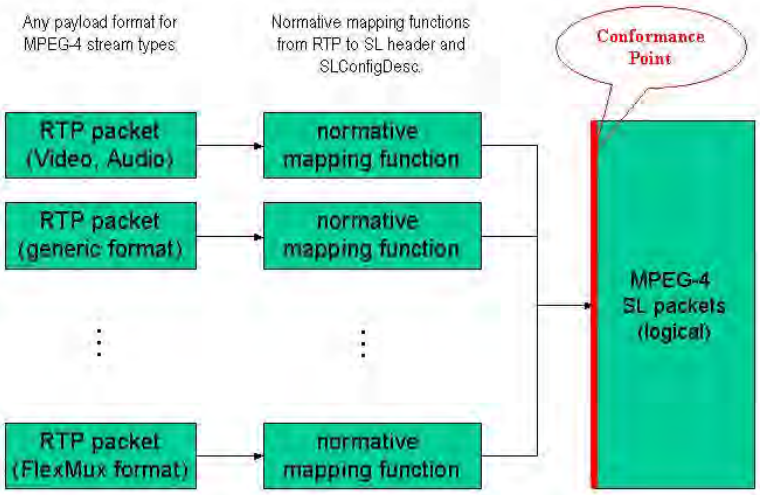


Figure 5. RTP packet to logical SL packet mapping

10.2 DMIF

DMIF (Delivery Multimedia Integration Framework) is a session protocol for the management of multimedia streaming over generic delivery technologies. In principle it is similar to FTP. The only (essential!) difference is that FTP returns data, DMIF returns pointers to where to get (streamed) data

When FTP is run, the very first action it performs is the setup of a session with the remote side. Later, files are selected and FTP sends a request to download them, the FTP peer will return the files in a separate connection.

Similarly, when DMIF is run, the very first action it performs is the setup of a session with the remote side. Later, streams are selected and DMIF sends a request to stream them, the DMIF peer will return the pointers to the connections where the streams will be streamed, and then also establishes the connection themselves.

Compared to FTP, DMIF is both a framework and a protocol. The functionality provided by DMIF is expressed by an interface called DMIF-Application Interface (DAI), and translated into protocol messages. These protocol messages may differ based on the network on which they operate.

The Quality of Service is also considered in the DMIF design, and the DAI allows the DMIF user to specify the requirements for the desired stream. It is then up to the DMIF implementation to make sure that the requirements are fulfilled. The DMIF specification provides hints on how to perform such tasks on a few network types, such as the Internet.

The DAI is also used for accessing broadcast material and local files, this means that a single, uniform interface is defined to access multimedia contents on a multitude of delivery technologies.

As a consequence, it is appropriate to state that the integration framework of DMIF covers three major technologies, interactive network technology, broadcast technology and the disk technology; this is shown in the Figure 6 below.

© MPEG 1999-2002 – unlimited reproduction permitted if not modified

Google Exhibit No. 1044
Page 032

Google Exhibit No. 1063
Page 1486