THE EVOLUTION OF DIGITAL AUDIO TECHNOLOGY

John N. Mourjopoulos

AudioGroup, Wire Communications Lab, Electrical & Computer Eng. Dept., University of Patras, 26500 Patras, Greece

1. INTRODUCTION

From its commercial introduction, nearly 20 years ago, digital audio technology is now reaching a maturing stage, allowing sound to be an integral component of the information age revolution. Thus, via the binary representation of the audio and acoustic signals, storage, manipulation, transmission and reproduction can be achieved with extreme speed, accuracy and flexibility. In addition, the users can also benefit from improvements in system performance, as well as reductions in component size and cost. Today, two distinct divergent trends are emerging in this technology's use: one towards improved quality and a transition form stereo to multichannel delivery (e.g. with DVD-Audio and SACD) and another towards reduced audio data bitrate, utilising data compression (e.g. ISO/MPEG or propriety standards).

Here, digital audio technology will be examined from the perspective of its 3 constituent evolutionary components: (a) digital electronics and computer technology, (b) DSP theory and techniques, (c) auditory modeling. For each of these components, the significant past events are traced and future trends are predicted. Nevertheless, this technology has not only given consumers and professionals alike access to high quality audio, but has also helped to highlight many problems, some related to human auditory perception and others to signal processing methodologies. Hence, there are significant developments to be tackled by future research.

2. TECHNOLOGY OVERVIEW

2.1 Market size

Digital audio delivered from CD-DA players today represents the most successful application with approx. 800 million players, approx. 20 billion CD-DA discs sold worldwide. For 10 consecutive year, disc sales have increased, reaching 2.5 billion discs in 2000. Since 1999, the DVD-Video format allowing access to multichannel audio and digital video at home, is proving to be the fastest growing consumer electronic product in history, reaching 10 million DVD-players worldwide [1]. Furthermore, the dedicated audio high-resolution optical disk formats of DVD-Audio and SACD, having improved with respect to CD-DA audio fidelity and allowing multichannel reproduction are also emerging. It is estimated that combined sales of these formats will be 120 million discs by 2004, or 5% of the total pre-recorded disc market.

At the multimedia end of the market, reliance on limited-bandwidth digital audio delivery (e.g. through the Internet) has resulted in the widespread use of lower-quality, compressed audio formats. "On-line" music delivery will have an estimated total size of 700 million Euros by 2004. Similarly PC-audio device market (including soundcards, self-powered speakers, etc.) is an ever-expanding area with an approx. 25% annual growth rate, with total sales already covering a significant portion of the audio and hi-fi product market.

2.2 Formats and systems

The numerous digital audio devices and systems may be grouped into the following classes:

- <u>general purpose home systems</u> (stereo, multi-channel and audiovisual, e.g. CD-players, DVD-Video players, decoders, etc.)
- <u>dedicated high resolution home audio systems</u> (e.g. DVD-Audio, SACD)
- <u>portable and wireless devices</u> (e.g. Internet audio-players, Minidisc, DAB decoders, automotive systems, etc.)
- <u>desktop computer audio devices</u> (e.g. soundcards, software players, editors, processors, etc.)
- <u>professional audio systems</u> (e.g. mixing consoles, recorders, processors, codecs, measuring devices, etc.).

As with CD-DA, the great majority of these devices is employing PCM for coding the binary audio data, which was initially standarised according to the Red Book standard, for 16-bit sample resolution, 44100 Hz sampling frequency (a 22.6 μs sampling period per channel) and 2-channels (stereo), hence resulting to an audio data rate of 1.4112 Mbit/s. From this initial format, numerous other standards and formats have emerged over the years, due mainly to the following reasons:

- adaptation to computer, network and multimedia requirements (e.g. audio file formats such as .WAV, .AIFF).
- improvement of audio quality and increasing storage capacity and channel delivery (e.g. 6-channel reproduction, sampling frequency up to 192 KHz and sample resolution to 24 bit for DVD-Audio [2]).
- reduction of the overall audio bitrate for applications where either channel bandwidth or storage space are constrained (e.g. MPEG-1, MPEG-2 and propriety standards such as Dolby Digital, DTS, ATRAC, RealAudio, WMA, etc.).
- introduction of differential 1-bit encoding (as opposed to the traditional multibit PCM), termed Direct Stream Digital (DSD), for both consumer and professional applications, for the SACD format and devices [3].

3. TECHNOLOGY EVOLUTION

3.1 A brief history

The theoretical roots of digital audio technology may be traced on early information theory workers, such as Shannon [4] and Nyquist. Nevertheless, computer and electronic technologies before the sixties, did not allow the practical implementation of such systems. Early applications of digital methods into audio were for computer music synthesis and significantly, the pioneering publications by Schroeder on digital artificial reverberation [5,6]. Other significant early developments were for the implementation of PCM audio recording / playback prototypes [7-9], audio DSP [10,11], hard-disc editing [12-14] and Compact Disc prototypes later to be standarised as the CD-DA format [15,16]. The seventies may be considered as the formative period of this technology, since most theoretical principles and applications were introduced at that time. However, practical restrictions in electronic and storage technologies, resulted so that these developments largely remained at a research level. The technology evolved at fast pace during the '80s, mainly due to the emergence of real-time DSP processors [17] and optical disc storage media, so that the first products appeared in the consumer and professional audio market. Many new techniques were realised, leading to developments in perceptual audio data reduction [18,19]. In turns, these developments led in the '90s to the introduction of the various ISO/MPEG standards [20], as well as the evolutionary technologies of high-resolution digital audio in the dawn of the new century [2,3]. Table 1 gives brief list of these developments.

3.2 Evolutionary mechanisms

Assessing the previously listed developments, it is possible to trace certain dominant underlying evolutionary mechanisms, which were largely responsible and enabling for these developments to be realised.

Date	Technique / application	Name / organisation
1961	Digital artificial reverb	Schroeder & Logan
1960s	Computer music	Various, e.g. Pierce
1967	PCM prototype system	NHK
1969	PCM Audio Recorder	Nippon Columbia
1968	Binaural technology	Blauert
1971	Digital delay line	Blesser & Lee
1973	Time-delay Spectroscopy	Heyser
1975	Digital music synthesis	Chowning
1975	Audio DSP emulation	Blesser et al.
1977	Prototypes of CD & DAD	Philips,Sony
1977	Digital L/S measurement	Berman & Fincham
1978-9	PCM / U-matic recorder	Sony
1978	32-ch digital multitrack	3M
1978	Hard-disk recording	Stockham
1979	Digital mixing console	McNally
1981	CD-DA standard	industry
1983	Class D amplification	Attwood
1985	Digital mixing console	Neve
1987-9	Perceptual audio coding	Brandenburg
1991	Dolby AC-3 coding	Dolby
1993	MPEG-1 Audio standard	ISO
1994	MPEG-2 (BC) Standard	ISO
1995	DVD Video standard	industry
1997	MPEG-2 (AAC) standard	ISO
1997	SACD proposal	Sony
1999	DVD-Audio standard	industry

Table 1: Significant developments in digital audio technology

3.2.1 Digital Signal Processing

Since the mid-sixties DSP has emerged as a powerful tool for analysis, manipulation and understanding of the signal properties. Historically, audio and acoustic signals represented one of the first areas for introducing novel DSP techniques, possibly due to their relative low requirements for computer processing and to the immediate sensory (hearing) evaluation of processing results. Following this analysis, it is possible to produce a brief list of typical DSP methods and algorithms that have significantly affected audio applications (**Table 2**).

Theory / method
Signal conditioning (averaging)
Filter bank signal analysis
FFT
Homomorphic filtering (cepstrum)
Linear prediction
Adaptive filtering
Digital filter design
Time – frequency analysis & synthesis
Multirate processing
One-bit modulation & coding
Statistical signal analysis & processing
Pattern recognition
Channel error correction
Non-linear DSP methods
Perceptual audio coding

Table 2: Some of DSP methods relevant to audio technology

3.2.2 Digital electronics, computer and networking technologies

Since its earlier days, the evolution of digital audio technology has been closely linked to the digital electronics, microprocessor, computer, storage media and related technologies. Sound is a real-time signal and for the implementation of any commercially viable product, a real-time response together with local storage (typically 10 Mbyte/min) is a prerequisite, something that it is not always necessary
for audio research applications. Practically any digital audio system will realise complicated mathematical operations on digitally represented real-time audio signals, with minimum input to output latency equal to 1 sample period (22.6 μs for 1 channel 44.1 KHz audio). Many of the established and commonly-used DSP algorithms operate on such sample by sample processing basis. However, many audio DSP operations require manipulation of frequency domain data, which are derived in a block processing basis, approach which can offer speed advantages for specific applications.

3.2.2.1 Processing power

Although for the implementation of audio DSP, a variety of alternative hardware options may be adopted, ranging from the use of off-the-shelf microprocessors, field-programmable gate arrays (FPGAs) to custom integrated circuits (ASICs), a specific class of optimised microprocessors, termed “DSP processors” has proved to be the most popular approach. The architecture of these DSP processors evolved together with DSP algorithm requirements, a trend that has started in 1982, when Texas Instruments introduced the TMS32010 processor with specialised hardware to enable it to compute a multiplication in a single clock cycle [17]. Currently, improved DSP processor architectures are available, based on enhanced DSP architectures (e.g. parallel execution units and very long Instruction words – VLIW, [21] etc). Nevertheless, the emergence of desktop audio (i.e. PC-based audio applications) has led to the utilisation of the general- purpose computer CPUs, as opposed to the earlier trend of

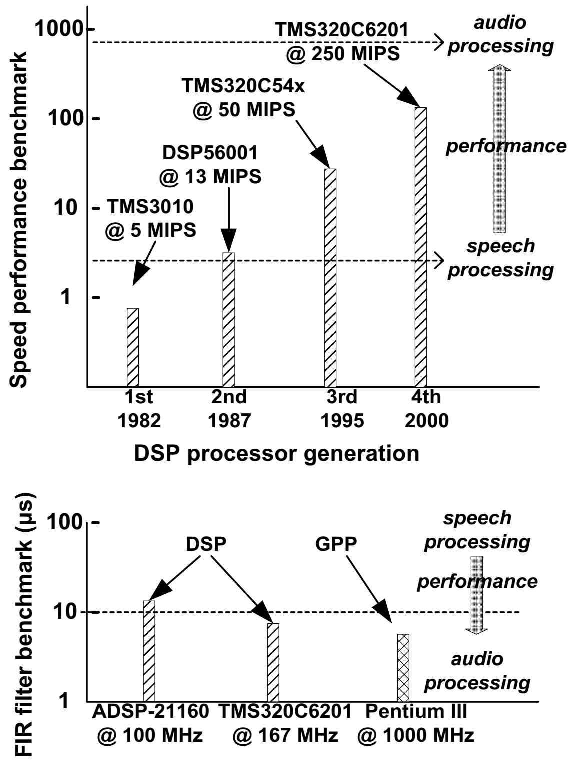

Figure 1: Real-time DSP and GPP audio performance [22]

using dedicated accelerator DSP cards attached to PCs. This trend has been supported by the ever-increasing clock frequency of such CPUs (e.g. in the region of GHz, as opposed to approx. 250 MHz for the latest DSPs) and their later adaptation to signal processing tasks. Currently such general purpose processors (GPPs) can easily challenge DSPs for audio processing speed, allowing numerous real-time audio software products to be emerge for the PC environment.

In order to evaluate processor power for a specific audio processing task, application-specific benchmarks have evolved that assess each system over a variety of relevant algorithms and tasks [22]. By application of such criteria to DSP processors of different generations, it is becoming evident that only after

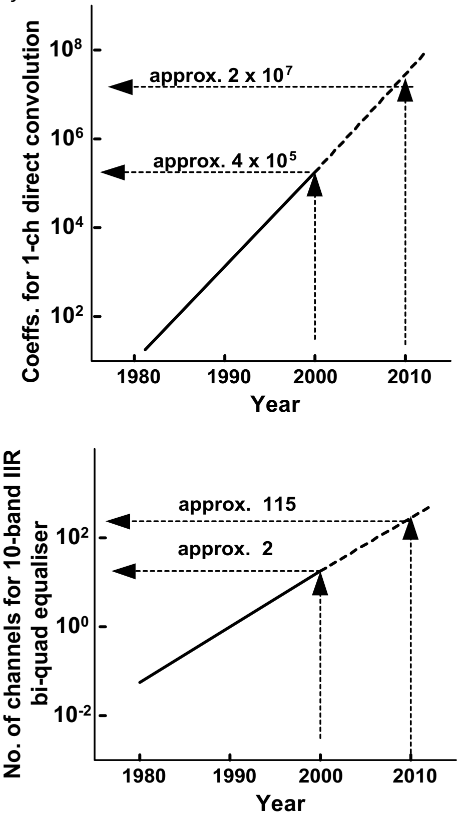

Figure 2: Real-time block and sample filtering trends [23,22]

mid-nineties processing power became sufficient to accomplish most real-time audio-related tasks (**Figure 1**). By observing similar results for current DSPs and GPPs, it is also evident that since mid-nineties GPPs became faster, so that real-time desktop audio could be efficiently implemented on software (e.g. effect processors, editors, synthesisers) without special hardware other than a soundcard. Hence, for approximately the last decade, digital audio products can be realised either as stand-alone dedicated devices, or as software attached to the common desktop computer platforms.

Let us consider typical examples for audio-related processing tasks and examine the evolutionary trends, attempting to extrapolate past and current data to predictions up to the end of the current decade. As was also stated in [23], processing power is increasing by approximately 2.25 times every 2 years, so that this trend conforms to the Moore's Law of digital electronics. **Figure 2** describes such trends for typical sample-based or block-based processing. These figures indicate that by 2010, power will suffice for all currently existing audio applications, 'though low-latency sample-based processing is just now becoming sufficient for demanding applications.

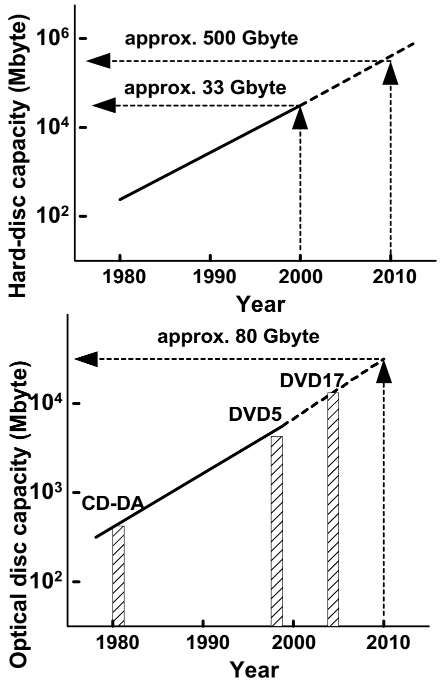

Figure 3: Hard and optical disc capacity trends

3.2.2.2 Storage capacity

Following the methodology set by the examples in the previous section, it can be predicted that by 2010, more than 500 Gbyte capacity will be readily available for most hard-disc related audio storage / retrieval tasks (approx. 500 CD-DA albums). (**Figure 3**). Considering optical disc media, future evolution will be strongly linked to standards, but at least 80 Gbyte will be available by 2010 for read-only and possibly recordable or re-writable optical disc media. In any case, the above predictions indicate that storage availability will soon exceed most requirements. The evolution of solid state memory (flash memory, EPROMs, etc) as future alternative to optical discs indicates that by 2010 available capacity will just become sufficient for storing a single CD-DA album in uncompressed form. Hence, such memories will be employed for low-quality audio applications (e.g. for portable devices), based on down-mixed or compressed audio.

3.2.2.3 Audio networking

It is becoming accepted these days that audio can be delivered to the user through the Internet, in compressed format (e.g. mp3, RealAudio, [24], etc). Although networking is dependent on many non-technological and local factors, the trends indicate (**Figure 4**) that Internet delivery is not growing at the required rate and will be only appropriate for lower-quality, compressed and / or down-mixed audio.

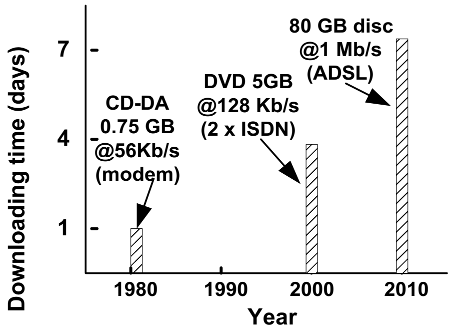

Figure 4: Internet audio delivery speed trends

Considering now the case of local delivery at home, the current wireless Blootooth platform nominally at 1Mbit/s (and up to 10m distances between devices) is not sufficient for uncompressed audio transmission [25] (**Figure 5**). Nevertheless WLAN IEEE 802.11b, may be sufficient for transmitting CD-quality and multichannel audio (from DVD-Video) and by 2010, the expected adoption of Hyperlan2 / IEEE 802.11a will allow multichannel high-resolution audio transmission, provided the cost of these will be comparable to audio wire links.

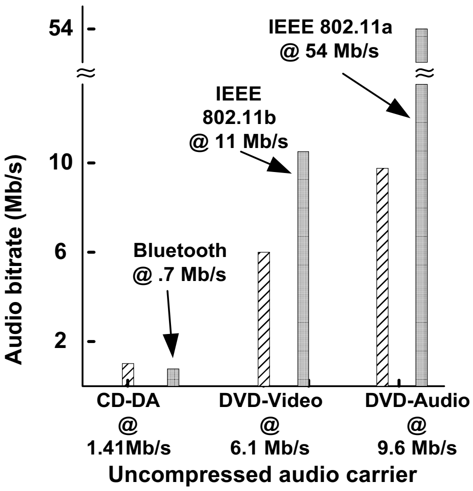

Figure 5: Audio WLAN transmission trends

3.2.3 Analysis and modeling human perception

From its introduction, audio engineering attempts to implement systems and devices conforming to the properties of human hearing and perception. Digital audio technology is supporting and expanding this approach. Typical examples of this synergy can be considered to be the increasing significance of data compression and binaural technologies. In both cases, two different peripheral perceptual mechanisms, well-defined through earlier psychoacoustic research [26,27] have been realised into practical systems and devices.
The evolution of real-time perceptual compression coders and decoders, initially on dedicated DSP processor-based hardware and more recently in software, has made possible numerous other audio applications to emerge, so that sound can be efficiently transmitted through networks and stored in minaturised devices [20].
Similarly, binaural technology is becoming widely employed in many audio applications ranging from architectural acoustics evaluation to 3-D audio rendering in Internet and multimedia presentations [28,29]. It is envisaged that many other audio applications such as source-position sensing devices, sound-quality evaluation tools, cocktail-party processors, special microphones for acoustically-adverse conditions, etc., will soon emerge from further utilisation of binaural technology [30].
However, digital audio technology is also enabling further research on many open questions concerning human auditory perception. In this sense, the previously described engineering exploitation of peripheral perceptual functions, appears to be only the beginning for future models which will utilise the largely unexplored field of cognitive psychology [28]. In order to describe these aspects of human performance, the cognitive models will have to remove parts of the known redundancy of the physical principles that govern current audio engineering practices. This will be especially beneficial to the most power-intensive applications such as sound field modeling and control.

4. EMERGING TECHNOLOGIES

Some of the emerging and promising areas in digital audio, are:

4.1 The all-digital audio chain

In the near future, audio peripherals such as power amplifiers, loudspeakers, microphones would be implemented digitally and linked via wireless means. All-digital power amplification introduces such integrated solutions form the source up to the loudspeaker, while provides high power efficiency (in the order of 90%), small size and low heat dissipation. Up to now, high-power digital amplifiers are implemented via a chain of subsystems, performing audio coding (e.g. oversampling / distortion compensation), 1-bit (e.g. PWM) conversion, class D amplification and output filtering [31,32]. Single chip solutions are available only for low-power applications.
Direct acoustic transduction of digital audio data remains a challenging research area. Some alternative solutions are emerging based on arrays of small 1-bit actuator elements, or binary multiple-voice-coil, loudspeakers.[33,34]. Nevertheless efficient, high-power, low-frequency direct digital audio transduction from small-sized radiating elements, still remains an open research field. Such technology is feasible for microphones, where research has led into the development of tiny microphones integrated within silicon integrated circuits (ICs) [35]. The possibility of further on–chip integration with analog to digital conversion (ADC) and future connection to wireless devices (see Section 3.2.2.3), will allow small sized, low cost, interactive acoustic sensing, useful for many applications.

4.2 Sound field control

Optimal acoustic reproduction, especially in reverberant enclosures, introduces many theoretical, perceptual and engineering challenges. Wavefield synthesis, i.e. accurate reconstruction of an original sound field via a large number of transducers [36], appears as alternative to current generation multichannel digital audio systems, given future availability of all-digital controllable, self-powered speakers.
Loudspeaker systems with digitally-controlled directivity may further allow accurate sound beam steering even at low frequencies, hence optimising audio reproduction into rooms or public spaces [37].
Digital filters for some time now [38,39], are employed for single-channel or multi-channel equalisation of the loudspeaker response and the room sound field, often combined with transaural processing for 3-D sound image rendering. However, such methods require adaptation to listener / source movements and their perceptual robustness cannot be easily guaranteed.
These methods will benefit from advances in acoustic monitoring elements (e.g. on-chip wireless microphone arrays), but clearly further research is required into the perception of complex sound fields by the human listeners and evolution of relevant DSP algorithms.

4.2 Portable and wireless devices

Digital audio may be soon transmitted wirelessly within devices in the home and office environments. Blootooth allows interactive compressed audio transmission to distances up to 10m [25] and the forthcoming IEEE 802.11b, 802.11a WLAN protocols will allow increased audio bandwidth for multichannel applications.
These developments, together with the reduced size and power consumption requirements for solid state storage media and low-power all-digital amplifiers, will allow small, battery-operated portable sound systems to emerge, possibly adapted into "wearable audio" form [40].

5. CONCLUSIONS

The previous sections have shown that past evolution dictates that for the next decade, highly flexible audio delivery will be possible, either via traditional optical disc carriers, or via the Internet and wireless local networks. However, technical restrictions will impose limitations on the data rate achieved by these systems, so that audio quality will be appropriately graded. DVD-Audio and SACD seem to emerge as the choice for high-fidelity multichannel audio delivery, potentially taking over from CD-DA as dedicated digital audio formats. DVD-Video and related compressed multichannel digital audio formats will clearly dominate audiovisual applications, being integral components of future multimedia, Internet and WLAN media delivery systems. Desktop audio applications will also further proliferate into consumer and professional markets, allowing flexible software-based audio storage, manipulation and reproduction, allowing flexible communication with other personal and home audio devices.

Such developments might lead to increased use of wireless portable audio devices, which will provide head-related audio devices with compressed-quality stereo, multichannel or 3-D audio utilising binaural technology. Such portable systems will wirelessly communicate with desktop computers or disc-players. Room-related audio systems will be based on loudspeakers or loudspeaker arrays. It is reasonable to predict that the reduction in size and cost of digital electronics, will allow wireless, self-powered operation and control of each loudspeaker unit, so that adaptive and equalised operation may be achieved, either for room-optimised multichannel reproduction or listener-specific transaural 3-D audio rendering. Larger scale public audio systems and installations may also benefit from such developments, especially for accurate full-bandwidth directivity control and steering.

It has been shown that for the next decade processing power and storage capacity will suffice for the above applications. Computer and communication networks as well as wireless protocols will also partially meet many of the audio-related bandwidth requirements. These developments will increase the amount of audio data stored, accessed, delivered and manipulated by the consumer and the audio professional, so that apart from logistical problems, robust and audio-transparent data security must be introduced into such systems. Nevertheless, many perception-related aspects of audio engineering will have to be resolved, if this technology can convincingly achieve many of the previously defined tasks.

6. REFERENCES

[1] Blesser B., Pilkington D., “Global Paradigm Shifts in the Audio Industry” JAES Vol.48 (9&10), 2000.

[2] Fuchigami N., et al., “DVD-Audio Specifications”, JAES vol.48(12), 2000.

[3] Nishio A., et al., “Direct Stream Digital Audio System”, AES 100th Conv. preprint 4163, 1996.

[4] Shannon C.E., “A Mathematical Theory of Communications”, Bell Sys. Tech. Jn., vol.27, 1948.

[5] Schroeder M.R., Logan B.F., “Colorless Artificial Reverberation”, JAES, vol.9, pp.192-197, 1961.

[6] Schroeder M.R., Atal B.S., “Computer simulation of sound transmission in rooms”, IEEE Int. Convention Report, Part 7, pp.150-153, 1963.

[7] Iwamura H., et al., “Pulse-Code-Modulation Recording System”, JAES, vol.21 (9), 1973.

[8] Myers J., Feinburg A., “High-Quality Professional Recording using New Digital Techniques”, JAES, vol.20(10), 1972.

[9] Sato N., “PCM Recorder”, JAES vol.21(9), 1973.

[10] Blesser B, Lee F., “An Audio Delay System Using Digital Technology”, JAES, vol.19(5), 1971.

[11] Blesser B., et al., “A Real-Time Digital Computer for Simulating Audio Systems”, JAES vol.23(9), 1975.

[12] Griffiths M., Bloom P.J., “A flexible digital sound-editing program for minicomputer system”, JAES vol. 30(3), 1982.

[13] Ingebretsen R., Stockham T.G., “Random-Access Editing of Digital Audio”, JAES, vol.32. pp.114-122, 1984.

[14] Abbot t C., “Efficient editing of digital sound on disc”, JAES, vol.32(6), 1984.

[15] Immink Schouhamer K.A., “The Compact Disc Story”, JAES vol.46(5), 1998.

[16] Carasso M.G., Peek J., Sinjou J., “The Compact Disc Digital Audio System”, Philips Tech. Rev. vol40(6), 1982.

[17] McDonough K., Mangar S., “A single-chip microprocessor architecture optimised for signal processing”, Proc. IEEE ICASSP, 1982.

[18] Brandenburg K., “High Quality Sound Coding at 2.5 bits/sample”, AES Convention preprint 2582, Paris, 1988.

[19] Johnston J., “Transform Coding of Audio Signals using Perceptual Criteria”, IEEE J. Sel. Areas Comm., vol.6(2), 1988.

[20] Noll P. “MPEG Digital Audio Coding”, IEEE Sign. Proc. Magazine, Vol.14(5), 1997.

[21] various, see “Industry DSP Technologies”, IEEE Sign. Proc. Magazine, Vol.17(2), March 2000.

[22] http://www.BDTI.com, Berkley Design Technologies Benchmark results 2000, 2001.

[23] Moorer J.A., “Audio in the New Millennium”, JAES, Vol. 48(5), 2000 May.

[24] Dietz M., et al., “Audio Compression for Network Transmission”, JAES vol 44(1/2), 1996.

[25] Floros A., Koutroubas M., Tatlas N.A., Mourjopoulos J., “A Study of Compressed Digital-audio Transmission”, to be presented, 112th AES Convention, May 2002

[26] Zwicker E., Feldtkeller R. “Das Ohr als Nachrichtenempfanger”, S.Hirzel Verlag, Stuttgart, 1967.

[27] Blauert J., “A contribution to the persistence of directional hearing in the horizontal plane”, Acustica vol.20, pp.200-206, 1968. (in German).

[28] Blauert J., “Spatial Hearing”, MIT Press , 1997.

[29] Savioja L., et al., “Creating Interactive Virtual Acoustic Environments”, JAES, vol.47(9), 1999.

[30] Blauert J., “Binaural Auditory Models”, 18th Danavox Symposium, 1999.

[31] Goldberg J.M., Sandler M.B., “Noise Shaping and PWM for All-Digital Power Amplifier”, JAES, vol.39 (6), 1991.

[32] Floros A., Tatlas N., Mourjopoulos J., “A Distortion-free PWM Coder for All-digital Audio Amplifiers”, AES 110th Conv., preprint 5361, 2001.

[33] Flanagan J.L., “Direct Digital-to-Analog Conversion of Acoustic Signals”, Bell System Tech. Jn., vol.59(9), 1980.

[34] Huang Y., Busbridge S.C., Fryer P.A. “Interactions in a Multiple-Voice-Coil Digital Loudspeaker”, JAES vol.48(6), 2000.

[35] Ouellette J., Echoes, JASA Newsletter, Vol.10 (3), 2000.

[36] Berkhout A., de Vries D., Vogel P., “Acoustic Control by Wavefield Synthesis”, JASA, vol.93, 1993.

[37] Pompei J.F., “The Use of Airborne Ultrasonics for Generating Audible Sound Beams”, JAES vol.47 (9), 1999.

[38] Clarkson P.M., Mourjopoulos J.N., Hammond J.K. "Spectral, Phase and Transient Equalization of Audio Systems", JAES, Vol.33(3), 1985.

[39] Elliot S., Nelson P. “Multi-point equalization in a room using adaptive digital filters”, J. Audio Eng. Soc., vol.37, Nov. 1989.

[40] Gough P., Eves D., “Wearable Electronics”, Philips Research Password 3, April 2000].