# Towards The All-Digital Audio / Acoustic Chain: Challenges And Solutions

Nicolas-Alexander Tatlas, Andreas Floros, Panagiotis Hatziantoniou and John N. Mourjopoulos

AudioGroup, Wire Communications Laboratory, Electrical and Computer Engineering Department
University of Patras, 26500 Greece

Correspondence: John Mourjopoulos (mourjop@wcl.ee.upatras.gr)

## ABSTRACT

The obvious advantages of digital audio technology have up to now being manifested mainly in media storage and processing sub-components, witch are parts of a more elaborate audio / acoustic analog reproduction chain. It is envisaged that the remaining components such as cables, amplifiers and transducers will soon be also implemented in digital form, potentially leading to networked, integrated and highly optimized solutions.
The paper examines theoretical and implementation aspects related to each of the modules that could constitute such an all-digital audio / acoustic transducer, namely: (a) the digital wireless receiver, via the Bluetooth and IEEE 802.11 protocols, (b) audio decoding and format adaptation (c) DSP for acoustic compensation, based on measured results for smoothed response equalization, (d) digital amplification, and, (e) all-digital transduction.

### 1. INTRODUCTION

The common element in all current digital audio systems is a chain of largely analogue, inefficient and inflexible subsystems which are attached to digital sources before sound reaches at the listener: (a) the controller / preamplifier / amplifier, (b) the loudspeakers and (c) the connecting cables. These subsystems, for many decades have resisted standardization and integration, but this problem is now becoming even more pressing with the practical difficulties imposed on the user in order to combine and interconnect multichannel home and multimedia audio set-ups. Hence, a significant challenge for future developments should be the evolution of a universal and integrated solution to combine all these subsystems, allowing the wireless coupling of any number of digital acoustic transduction elements to digital audio sources, in networked applications.

Here, a potential scenario for integrating the above modules is discussed and analysed. According to this, as shown in Fig. 1, an integrated all-digital transducer should be allocated per audio channel allowing self-powered, adaptable and controllable wireless operation, so that stereo or multichannel signals of any of the current or future digital formats could be appropriately received, decoded, processed, amplified and acoustically reproduced. In more detail, this paper will discuss the problems and examine the merits of possible solutions for critical subsystems within this chain, when possible presenting implementations and simulation test results.

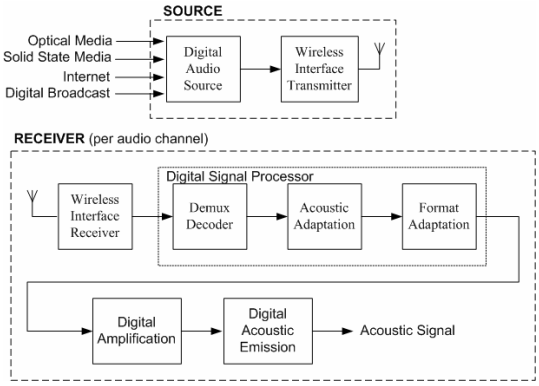



**Fig 1:** Overview of a possible form for the all-digital audio/acoustic chain

The paper initially considers the potential of current Bluetooth and IEEE802.11a/b protocols for wireless local area networks (WLANs), for multichannel audio transmission, largely based on results previously presented by the authors [11].

A subsequent section discusses general requirements for the decoding module which could allow compressed or non-compressed format audio to be converted to PCM stream in each channel receiver.

It is also likely that the response of each receiver may be sub-optimal due to physical restrictions in all-digital transduction, or even, its positioning inside the listening environment. For this, a potential equalization scheme is proposed, based on complex smoothing and equalization of each channel's combined audio/acoustic response [19]. Following decoding and processing, as shown in Fig. 1, each channel bitstream will have to be further adapted to the format employed for subsequent digital amplification and transduction. Given that PCM, Sigma-Delta modulation (SDM) and PWM appear to be the most promising candidates, these potential options are examined. With respect to these formats, issues related to the performance and complexity of the digital amplification module are examined (DAMP). The final section of the paper introduces novel methodology and preliminary results related to the potential implementation of Digital Transducer Arrays (DTAs), fed by either multibit PCM or one-bit signals (i.e. PWM or SDM). Hence, this section extends results presented previously [9], allowing useful conclusions to be drawn on the respective merits and disadvantages of each signal format for such applications.

## 2. WIRELESS NETWORKING

Wired digital networking solutions are widely accepted for connecting digital sound sources to multichannel decoders/amplifiers. The FireWire protocol has been established as a good candidate for future audio networking technology. FireWire comes as enhancement to the SPDIF protocol, providing faster bitrates, command and control, content protection and future flexibility. On the other hand, wireless protocols are widely employed for cable-free personal (WPANs) and local (WLANs) networking between workstations and other electronic devices. It is argued here that such wireless operation will greatly enhance the functionality and potential for integration for future audio devices. The most common digital networks employed, are Bluetooth [12] (WPAN), and IEEE802.11b/a [13] (WLAN).

### 2.1. Bluetooth Audio

Bluetooth is an attractive wireless specification for developing WPAN products, with a maximum range of 100m and a theoretical rate of 1Mbps. The protocol provides transmission of data and voice in point-to-point and point-to-multipoint setups. However, the bandwidth limitation together with the nature of the permitted wireless links between the connected devices present some major drawbacks. More specifically:

- Time-bounded applications, such as digital audio transmission require a constant throughput. So they must be established through Synchronous Connection-Oriented (SCO) links. The effective bandwidth in such a case is 64kbps for each link, with a maximum of 3 concurrent links.
- The establishment of SCO links creates a bandwidth overhead, which dramatically decreases the transfer capabilities of any co-existing Asynchronous Connection-Less (ACL) link.
- The packet-switched nature of the ACL links, given the limited bitrate, is not suitable for real-time applications. Especially in noisy environments, packet retransmissions are applied to ensure data integrity, which further reduce the effective bandwidth.

Given the above limitations of the Bluetooth standard, the following considerations are necessary for realizing high-quality, real-time audio reproduction through Bluetooth:

- The original audio data must be compressed prior to the transmission.[11]
- Apart from the transmitted audio data, many applications may require the transmission of control information (e.g. volume control, timing information etc), which must be multiplexed with the compressed audio data.

Below, some results are discussed, based on a Bluetooth audio implementation, reported in detail in [11]

#### *2.1.1. Bluetooth Mulitple Links*

Stereo and multichannel audio reproduction requires the concurrent transmission of discrete audio channels to two or more playback devices. Fig. 2 shows a typical example of a stereo application, where two audio channels ((L)eft and (R)ight) produced by an audio source (e.g. CD-Player, DVD-Video, etc) are compressed and individually wirelessly transmitted to the corresponding Bluetooth enabled devices through two ACL links, able to transmit both audio and control information.

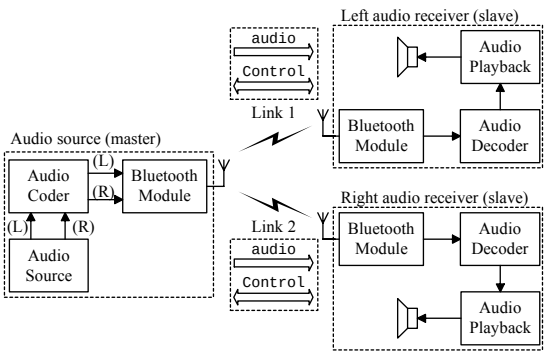


**Fig. 2:** Point-to-multipoint audio and user-controlled information transmission for stereo playback

The maximum number of supported channels in this case is equal to:

$$N_{ch} = \left\lfloor \frac{b_{max} - b_i}{b} \right\rfloor \quad (1)$$

where $b_{max}$ is the maximum available bitrate (equal to 721kbps), b is the encoding bitrate (in kbps) employed for the compression of the audio information, $b_i$ (kbps) is the mean traffic of the control data through all the established links and $\lfloor\ \rfloor$ denotes floor integer truncation.

Considering the case of concurrently transmitting audio and control data through two ACL links for stereo reproduction, a mean effective bitrate value was measured at both playback devices equal to 295kbps. However, the above application setup introduces synchronization problems between the two playback devices, as different rates of retransmissions are performed on the two ACL links.

The above problem can be overcome by pre-buffering an adequate portion of audio data, but this does not affect a possible time-mismatch of the control (user-defined) information, which is usually short and practically not noticed.

While the above application scheme is suitable as stereo reproduction scheme, the available bandwidth is inadequate for multichannel (e.g. 6-channel) reproduction.

*2.1.2. Bluetooth Broadcasting*

An alternative approach to multichannel transmission, is to employ the broadcast capability of the Bluetooth standard, which overcomes the above-mentioned channel limitation. As it is illustrated in Fig. 3, in such a case, the multichannel audio content is compressed and broadcasted, with the channel separation taking place on each of the playback devices. Hence, it is required that channel identification information should be assigned to each receiver in order to reproduce the appropriate audio channel.

In broadcast mode, the maximum number of allowable connections of a master device to slave devices equals to seven, thus a maximum of seven audio channels can be transmitted. Hence, apart from typical stereo applications, all current multichannel playback formats can be supported (e.g. AC-3 coding, DTS, etc). During tests, data integrity was not found to be assured, as no dynamic retransmission mechanism is applied in such a case. Hence, the reproduction in this case was just acceptable, due to packet losses introduced in the wireless path. Since the Bluetooth specification allows the definition of the number of retransmissions N, performed in broadcast mode (which is constant even if a packet is successfully transmitted), the above losses are partially reduced with increasing N with proportional reduction of the effective bandwidth. Typical measurements of the effective bitrate and the percentage of the lost audio information are presented in Table 1, as a function of N.

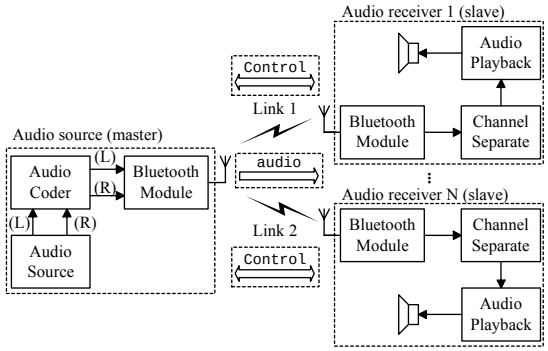



**Fig. 3:** Wireless multichannel transmission using point-to-multipoint connection setup in broadcast mode

| N | $b_e$ (kbps) | Audio data losses |
|---|---|---|
| 0 | 551 | 7%-9% |
| 1 | 325 | 2%-3% |

**Table 1:** Measured effective bitrate and audio data losses as a function of retransmissions performed in Bluetooth broadcast mode (for 1m distance).

Given that the packet losses are introduced to compressed audio data, the final audio reproduction is highly distorted. Moreover, as the losses vary in length and time occurrence between the two ACL links, they additionally introduce "phase" distortion between the two reproduced audio channels. The above observations seem to represent major restrictions for the development of multichannel audio applications, which given the bandwidth limitations of the Bluetooth protocol must be realized in broadcast mode. In practice, the above problems can be overcome by developing mechanisms in both the Bluetooth lower layer stack and the application layer, designed to perform packet loss signaling and retrieval procedures, as well as appropriate audio data restoration routines.

*2.1.3. Bluetooth Conclusions*

The conclusions drawn from the study of the Bluetooth protocol [11] can help in assessing the optimal Bluetooth-based transmission parameters for

compressed quality, real-time audio playback. Fig. 4 maps the measured effective bit rate (for 1m distance) with the audio data bitrate, indicating the maximum allowed compressed audio bitrate for all test cases considered. Ideally, uncorrupted real-time audio reproduction requires an effective bandwidth value at least equal to the compressed data bitrate. This condition is graphically represented by the Real-time threshold line. For all transmission cases examined, the maximum allowed compression bitrate is defined by the point of intersection of the mean measured effective bitrate value with the real-time threshold line. For example, from this figure it can be deduced that stereo audio and control information transmission (2 discrete ACL connections) is possible when each audio channel is encoded at a maximum of 256kbps (Stereo), while this rate is increased to 320kbps or higher when a single ACL link is considered (Mono).

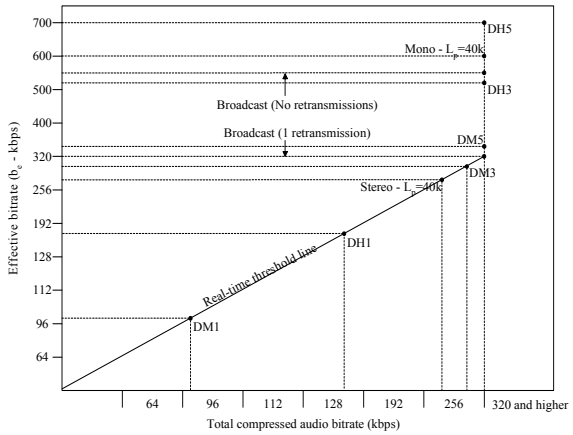



**Fig 4:** Mapping diagram of the audio data bitrate and the maximum allowed compressed audio bitrate for all test cases examined, for 1m distance (from [11])

## 2.2. IEEE802.11 Audio Networking

IEEE 802.11b and 802.11a are WLAN standards, used over large distances (up to 500m for access points) and theoretical data rates of 11Mbps for 802.11b and 54Mbps for 802.11a. The IEEE 802.11 specifications address both the Physical (PHY) and Media Access Control (MAC) layers while the others remain identical to the IEEE802.3 (Ethernet).

Two different setups for WLAN networks exist, "infrastructure" and "ad hoc". Infrastructure mode requires the use of at least one access point (AP), providing an interface to a distribution system. In this case, all network traffic goes through the AP. The "ad hoc" mode allows the network module to operate in an independent basic service set (IBSS) network configuration. With an IBSS, the devices communicate directly with each other in a peer-to-peer manner.

For audio device networking, an ad-hoc setup seems to be the most appropriate solution, because of the reduced cost and the better performance provided. The latter is in general true for a limited number of users: given that the maximum number of wireless audio devices operating in such a network should not be more than 7 (one audio transmitter and six receivers) the constraint is met.

In order to transmit audio through a IEEE802.11 ad hoc network for stereo or multichannel reproduction, the multicast/broadcast capability of the protocol must be employed. Multicasting refers to sending data to a select group whereas broadcasting refers to sending a message to everyone connected. The IEEE 802.11 multicast/broadcast protocol is based on the procedure of Carrier Sense Multiple Access with Collision Avoidance. The protocol does not offer any recovery mechanisms for the above data frames. As a result, the reliability of the multicast/broadcast service is decreased because of the increased probability of lost or corrupted frames resulting from interference or collisions.

There are two alternatives in order to solve the above mentioned issue: the first is using one of the many proposed reliable multicast protocols, where modifications are made to the MAC layer in order to add data integrity mechanisms [15]. However, such an addition could lead to synchronization problems because of the recovery mechanisms employed, as discussed with the Bluetooth protocol. Since throughput is not an issue with 802.11b or 802.11a for audio purposes, a second potential approach would be to transmit in a high definition coding that is "resistant" to errors and additionally develop an algorithm that analyzes corrupt frames (identified through faulty CRC), using undamaged components and recreating the rest through interpolation techniques. Reduction of audio quality in this case should be a matter of future investigation.

## 3. DECODING

One of the goals of the system in Fig. 1 is to implement decoders for low-latency streaming of compressed or non-compressed digital audio data delivery data through wireless networks.

Typically, the following digital audio formats may have to be incorporated on the decoder:

- Uncompressed two or multi-channel PCM audio (16 up to 24 bit resolution, 44,1 up to 192KHz sampling)
- Uncompressed DSD multichannel, at 6.2Mbit/s
- Stereo MPEG-1 Layer 3 (mp3) coding, based on the ISO-MPEG Audio coder IEC11172-3. Bit

rates supported will be in the range of 128kbit/s up to 320kbit/s as defined by the standard.

- Dolby Digital 5.1 multichannel coding based on the ATSC A/52 Digital Audio Compression Standard (AC-3). Typical rates of the AC-3 bitstream are 384kbit/sec up to 640kbit/sec.
- Multichannel MPEG-2 coding, based on the ISO/IEC 13818-3 standard. The typical bitrate for MPEG-2 bitstreams extend up to 684 kbit/s for multichannel reproduction.
- low latency subband coding formats (SBC codec), based on the specifications of the mandatory codec for the Advanced Audio Distribution Profile of Bluetooth, which is currently [14] under a standardization (voting) procedure. The bitrates supported will be in the range of 200 to 600 kbits/s.

The decoder module for the first four cases must be based on the corresponding standards. However a challenge here is to incorporate reliable extraction of high-priority data (e.g. header information) received through the wireless channel, which is essential for the decoding process.

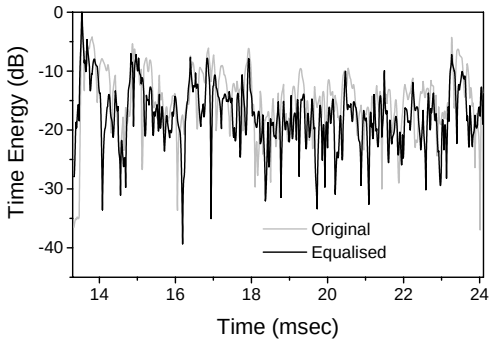
The low-latency SBC codec has the aim to provide real-time encoding and decoding of a live audio stream, for example, the sound channel of a TV program.

## 4. PROCESSING

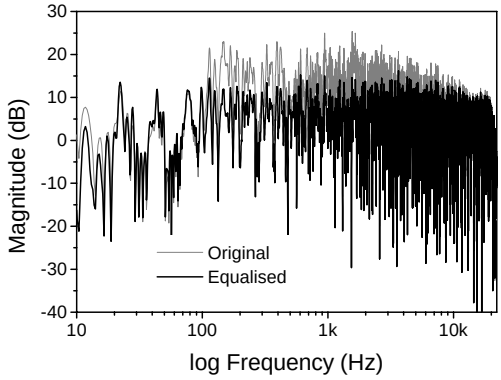
The all-digital transducer should incorporate sufficient DSP power, not only to realize simple functions such as volume control, delay, shelving filtering, but also to implement transducer response correction and adaptation to its local acoustic environment via the use of equalization methods. These can incorporate FIR filtering on PCM data by using room response inverse filters derived from responses individually measured for each of the channel receivers.

The acoustic response digital equalization method proposed here, is based on an inverted complex-smoothed acoustic room response scheme that overcomes many known problems [18-19]. The filters derived from the use of the above method can be implemented on the processor/decoder DSP within the practical limits for the FIR filters' length, taking into account possible requirements for the low-latency operation.

An example of a room response function before and after the proposed complex smoothing equalization is shown in Fig. 5. As it can be observed, in the time domain the equalised response has more power shaped in the direct and early reflection path and less power allocated in some of the reverberant components. In the frequency domain, the equalization corrects gross spectral effects due to early room reflections, without attempting to compensate for many of the original narrow-bandwidth spectral dips.




(a)




(b)

**Fig. 5:** Room Response before and after the proposed equalization method: (a) Time domain (Energy) (b) Frequency domain

The implementation of the proposed equalization scheme is shown in Fig. 6. The sound material must be pre-filtered in real-time by the equalization filter, derived from an appropriately smoothed response version, implemented within each channel receiver, and then it is reproduced via the audio chain into the room.

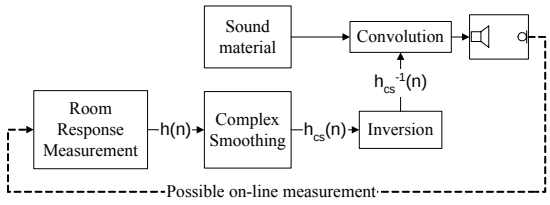
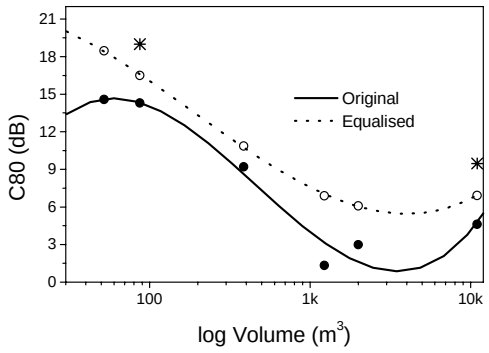


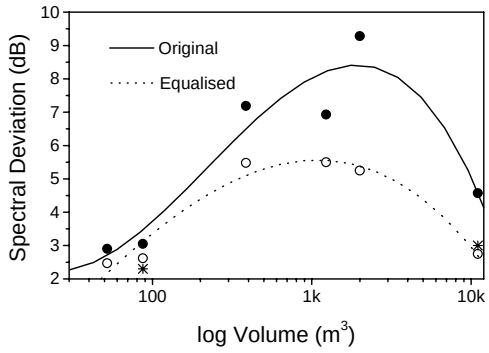
**Fig. 6:** Implementation scheme for the proposed audio/acoustic equalization method.

As was reported in [19], all tested and well-established objective acoustic criteria, were found to improve after the application of smoothed response equalization, for a number of tested spaces, ranging from a small office to a 1000-seat auditorium. Furthermore, the majority of listeners preferred the equalized sound material, in a real-time test. Specifically, the Clarity (C80) criterion, known to describe the perceived acoustic effects on music presentation, seemed to improve, irrespective of the room's volume, typically as is shown in Fig. 7. The Room Transfer Function amplitude Spectral Deviation was also found to be reduced after equalization, for all spaces, by an amount generally proportional to the degree of the room's original spectral irregularity, which was not always increasing with its volume.

Subjective tests, have also confirmed a preference for such room-corrected audio material, when was presented in real-time in a listening environment [19].




(a)




(b)

**Fig. 7:** Results for of measured acoustic space responses before and after equalization: (a) Clarity (C80) vs Volume (b) Spectral Deviation vs Volume

## 5. AMPLIFICATION

A complete digital-audio/acoustic chain requires power amplification of the audio signal in its digital form without prior analog conversion.

The switching operation of the power transistors employed form amplification of digital audio pulses yields very high power efficiency, which represents one of the major advantages of digital amplification systems. Theoretically, assuming ideal switch components, the amplification performance is 100%, which means zero heat dissipation. In practice, the efficiency is limited to 80-90%, affected by the non-zero resistance of the switches [1]

More specifically, if $P_0$ (W) is the input power fed to the amplifier, the real output power is given by:

$$P_{out} = P_0 - P_{loss} \quad (2)$$

where $P_{loss}$ (W) is the total power losses due to the non-ideal power switches (all the other possible power dissipation reasons are ignored for simplicity reasons). It is obvious that $P_{loss}$ depends on the number of times that the power switches are changing their state, which in general corresponds to the Pulse Repetition Frequency (PRF) of the control pulse stream [2]. Assuming that each state-change results to $P_{l0}$ power losses, eq. 2 becomes:

$$P_{out} = P_0\left(1 - 2\frac{P_{l0}}{P_0}PRF\right) = aP_0 \quad (3)$$

where α is the efficiency factor (α≤1). The current power switching technology implies a maximum PRF of 300 up to 500kHz in order to achieve high power efficiencies [3].

Up to now, multibit Pulse Code Modulation (PCM) was considered as the de facto signal format for the control of the power switches. Although PCM audio meets the requirement of PRF≤500kHz mentioned above, the multibit nature of such data introduces a practical restriction towards PCM-based all-digital amplifiers: assuming an N bit resolution (typically N=16 or more) and a sampling frequency equal to $f_s$ (Hz), the PCM signal values are represented by $2^N$ discrete levels and the direct digital amplification of such a signal would require $2^N$ power switch topologies operating at a frequency $f_s$. Nevertheless, as will be discussed in Section 6, lower resolution

PCM maybe employed to drive a digital transducer, which may alleviate such problems

Over the last decades, the use of 1-bit audio signals has emerged as an attractive practical alternative to multibit PCM audio. For digital amplification purposes, the 1-bit coding technology overcomes the above practical restriction of the excessive power switch number, requiring only a set of four power switches connected together in an H-bridge topology (e.g. as in an analog class D amplifier) at the expense of a much higher operating frequencies implied by the 1-bit sampling rates (typically $Rxf_s$, where R is the factor of the oversampling employed prior to 1-bit conversion). This introduces electromagnetic interference problems, increases the implementation complexity (mainly in terms of circuit design and component requirements) and decreases the amplification efficiency, given that the average PRF is increased.

Current efforts are considering two well-known 1-bit coding formats for developing all-digital amplifiers: (a) 1-bit Sigma/Delta Modulation (SDM) as employed in DSD, and (b) Pulse Width Modulation (PWM). Both approaches follow the same basic structure, with the overall amplification performance strongly depending on the modulation characteristics.

Assuming a 44.1kHz initial PCM sampling rate and given that in the case of SDM, the bitrate of the 1bit output as well as the PRF is not constant, but is derived from and hence depending on the input sample values, the SDM-based amplifiers should operate at maximum clock frequencies of 3MHz. Hence, the measured power efficiency given by eq. (3) can vary with time, but in average, is lower than for the case of direct PCM amplification by a factor equal to the specific oversampling ratio employed.

Moreover, SDM is known to suffer from slowly repeating bit-patterns (idle tones [2]), as well as high frequency noise. A number of digital signal processing algorithms have been proposed in the literature to solve these problems, such as the controlled SD bit-flipping [3] and high order noise shaping techniques.

On the other hand, PWM modulation was initially employed in analog telecommunication systems. However, previous work [4] has shown that it can be analytically described as a 1-bit digital coding technique by converting each quantised input value to a fixed level pulse with proportional time width. The PRF in this case equals to the digital input sampling frequency $f_s$, achieving high amplification efficiency according to eq. (3), while the final bit rate of the PWM bitstream is given by [4]:

$$f_p=2(2^N-1)f_s\,(\mathrm{Hz}) \quad (4)$$

where N is the digital input signal bit resolution.

However, it is well known that PWM suffers from harmonic and non-linear distortions, which decrease with the switching frequency [5]. Hence, in order to keep the overall distortion level low, increase of the PRF is required (typically by a factor of 64 [6]). For typical audio applications, this increases the final bit rate ($f_p$) in the range of GHz, rendering it inconvenient for practical implementations. For that reason, various PWM linearisation strategies have been developed [7], in the past, attempting to compensate PWM-related distortions while keeping the PRF as low as possible. Fig. 8 illustrates the effects of applying the "Jithering" technique developed by the authors, which eliminates the PWM-induced distortions without increasing the PRF [8].

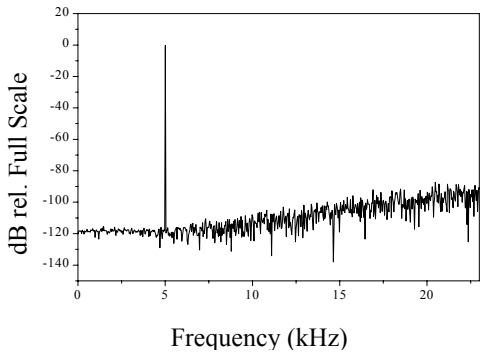



**Fig. 8.** Distortion free PWM conversion using "Jithering" distortion elimination technique

Table 2 summarizes the merits of all-digital amplification technologies with respect to signal formats. Clearly, PWM 1-bit coding represents the optimal choice for digital amplifiers, as it combines high power efficiency with low implementation complexity. High quality can be achieved using DSP preprocessing algorithms. Moreover, as SDM-based amplifiers offer lower efficiency, PWM can also be used in the case that the original audio content is SDM (e.g. as in DSD), by employing Pulse Group Modulation (PGM) [17]. As it is shown in Table 2, PGM was found to be an audibly transparent way for converting a SDM pulse stream to PWM data, able to combine the low switching frequencies of PWM and the low final bitrates of SDM.

| | PCM | SD | PWM | PGM |
|---|---|---|---|---|
| Switching rate | $f_s$ | $Rxf_s$ (max) | $f_s$ | $f_s$ |
| Bitrate | $f_s xN$ | $Rxf_s$ (max) | $2(2^N-1)f_s$ | $Rxf_s$ |
| Efficiency | High | Low | High | High |
| DSP required | No | Yes | Yes | Yes |
| Power switches required | $2^{N+1}$ | 4 | 4 | 4 |

**Table 2:** Comparison of all-digital amplifier technologies

## 6. TRANSDUCTION

A "digital loudspeaker" is a direct digital-signal to acoustic transducer, usually comprising of a digital signal processing module driving miniature elements, strategically positioned in order to reproduce the audio digital data stream. The main advantage of such loudspeaker is that the signal remains in the digital domain and is converted to analog through the element-air coupling. Other advantages of such digital arrays, relate to the flexible control of their directivity [10], currently utilized for multichannel reproduction via a single array [22]. Such directivity features are not addressed here, since the original system is envisaged to comprise of discrete, networked receivers, each reproducing an individual audio channel (Fig. 1).

Current efforts for digital loudspeakers are focusing on multi-bit signals (PCM) [9] in two different forms: Digital Transducer Arrays (DTA) and Multiple Voice Coil Digital Loudspeakers (MVCDL). Although the MVCDL appears to offer an attractive solution for many current-day practical problems, in this study the discussion will focus on DTA, as an alternative solution offering greater promise for a future integrated audio system design. Here a DTA scheme will be anaylsed, introducing a novel approach for examining its audio performance, for inputs covering all the prominent digital audio formats. Hence, the DTA will be driven by multibit PCM signal [9] and by one-bit signals such as PWM or DSD.

Generally, a DTA consists of three stages (Figs 9,10,11):

- digital signal processing (DSP)
- digital audio amplification (DAMP) (see previous section)
- digital acoustic emission (DAE)

### 6.1. Outline

*6.1.1. PCM Digital Speaker*

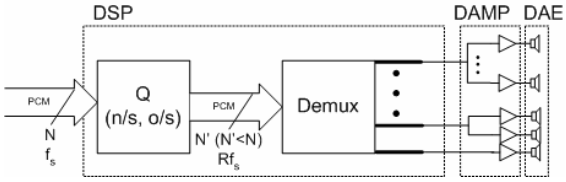



**Fig. 9:** PCM Digital Speaker

The input PCM signal $S = \begin{bmatrix} b_{1,1} & . & . & b_{1,m} \\ b_{2,1} & . & . & b_{2,m} \\ . & . & . & . \\ b_{n,1} & . & . & b_{n,m} \end{bmatrix}$, represented here as a $N \times M$ matrix, where $N$ is the number of bits per sample and $M$ is the total number of samples (in theory infinitely large). The signal is originally sampled at a frequency $f_s$, can be over-sampled $R$ times, noise-shaped and re-quantized, thus providing the signal $S' = Q[S] = \begin{bmatrix} b_{1,1} & . & . & b_{1,m'} \\ b_{2,1} & . & . & b_{2,m'} \\ . & . & . & . \\ b_{n',1} & . & . & b_{n',m'} \end{bmatrix}$, resulting to a $N' \times M'$ matrix, where $M' = R \cdot M$ and $N'$ the new bit resolution per sample (generally $N' < N$).

$N'$ different bitstreams are then created, $S'_0$ to $S'_{N'-1}$ by extracting for each one a bit from a specific position from every PCM sample. So, $S'_0 = [1 \;\; 0 \;\; . \;\; . \;\; 0] \cdot S'$, $S'_1 = [0 \;\; 1 \;\; . \;\; . \;\; 0] \cdot S'$ and $S'_{N'-1} = [0 \;\; 0 \;\; . \;\; . \;\; 1] \cdot S'$.

The bitstreams are then fed to groups of DAMPs, autonomously driving DAEs. In order to acoustically reconstruct the original digital signal, the groups consist of a variable number of elements. Each PCM bit has a different "weight" (W) where $W = 2^X$ and $X$ is the bit position (significance), i.e. $X = 0 \Rightarrow W = 1$ for LSB. So, the $S'_0$ bitstream (LSB) is fed to one transducer, the $S'_1$ to two while the $S'_{N'-1}$ (MSB) to $2^{N'-1}$ elements [9]. Thus, the total number of elements, is $2^{N'} - 1$.

*6.1.2. PWM Digital Speaker*

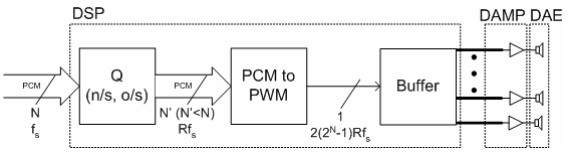



**Fig. 10:** PWM Digital Speaker

The input signal, similar to the PCM speaker case, is over-sampled R times, noise-shaped and re-quantized, thus providing the signal $S' = Q[S] = \begin{bmatrix} b_{1,1} & . & . & b_{1,m'} \\ b_{2,1} & & & b_{2,m'} \\ . & . & . & . \\ b_{n',1} & . & . & b_{n',m'} \end{bmatrix}$, $N' \times M'$ matrix, where $M' = R \cdot M$ and $N'$ the new bit resolution per sample (generally $N' < N$).

The $S'$ signal is then fed to the PCM-to-PWM modulator, where a one-bit switching bitstream is produced (see previous section). Each PCM sample is converted to a series of one bit PWM samples, thus $PWM(Q[S]) = \left[ b_{1,1} \quad . \quad . \quad b_{n",1} \quad b_{1,2} \quad . \quad . \quad b_{n",m'} \right]$, where the total number of elements is equal to $M' \cdot 2(2^{N'} - 1)$.

For ease of calculation, the above matrixes' PWM samples are "grouped", according to the "progenitor" PCM samples, creating PWM frames. Each PWM frame corresponds to one PCM sample. $S"$ is then derived, where

$$S" = G(PWM(Q[S])) = \begin{bmatrix} b_{1,1} & . & . & b_{1,m'} \\ b_{2,1} & . & . & b_{2,m'} \\ . & . & . & . \\ b_{n",1} & . & . & b_{n",m'} \end{bmatrix}, N' \times M'$$

matrix and $N" = 2(2^{N'} - 1)$. Each column of $S"$ is a PWM "frame".

The bits from each PWM sample of the signal S'' are separated, thus creating $N"$ different bitstreams, $S"_0$ to $S"_{N"-1}$ where $S"_0 = \left[ 1 \quad 0 \quad . \quad . \quad 0 \right] \cdot S"$, $S"_1 = \left[ 0 \quad 1 \quad . \quad . \quad 0 \right] \cdot S"$, and $S"_{N"-1} = \left[ 0 \quad 0 \quad . \quad . \quad 1 \right] \cdot S"$.

The difference between the PWM speaker and the PCM speaker in the DAE, relates to the fact that each PWM bit has actually the same weight (one bit signal), so each is directly fed to a unique element. Thus, the total number of elements in this case is $N" = 2(2^{N'} - 1)$, i.e. twice the number for the PCM speaker.

As with the PCM Digital Loudspeaker, where in a given time instance, taken for one PCM sample to be reproduced, a PWM bitstream frame corresponding to one PCM sample is fed to the elements. Since the PCM-to-PWM mapper output is a single bit stream, a buffering stage must be included in order to synchronize the whole PWM frame. The buffering stage is actually described by the grouping function G.

Alternative designs where fewer elements than the number of bits of one frame can be envisaged, however a trade-off exists between the number of elements and the operating frequency. Moreover, as the number of DEAs increases, the path difference between the produced signals increases resulting to increasing distortion (see following section).

### *6.1.3. Sigma - Delta Digital Speaker*

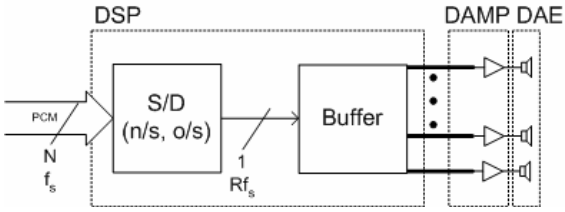



**Fig. 11:** Sigma – Delta Digital Speaker

The "Sigma-Delta" digital speaker is the second of the two possible digital speaker implementations driven by one bit digital sound. Here, typical DSD-format S-D modulation is considered.

The signal $S$, through the Sigma-Delta modulation process is over-sampled $R$ times, noise-shaped and re-quantized, thus providing the single bit stream $S - D[S] = \left[ b_{1,1} \quad . \quad . \quad b_{n',1} \quad b_{1,2} \quad . \quad . \quad b_{n',m} \right]$, represented as a single-row matrix, with $M'$ columns where $M' = R \cdot M$.

For ease of calculation, similarly to the PWM analysis, the above Sigma-Delta samples are "grouped", according to the "progenitor" PCM samples, creating Sigma-Delta frames. So, $S'$ is derived where

$$S' = G(S - D[S]) = \begin{bmatrix} b_{1,1} & . & . & b_{1,m} \\ b_{2,1} & . & . & b_{2,m} \\ . & . & . & . \\ b_{n',1} & . & . & b_{n',m} \end{bmatrix}, \qquad N' \times M$$

matrix and $N' = R$. Each column of $S'$ is a Sigma-Delta "frame".

R different bitstreams, $S'_0$ to $S'_{N'-1}$, are created where:

$S'_0 = \left[ 1 \quad 0 \quad . \quad . \quad 0 \right] \cdot S'$, $S'_1 = \left[ 0 \quad 1 \quad . \quad . \quad 0 \right] \cdot S'$ and $S'_{N'-1} = \left[ 0 \quad 0 \quad . \quad . \quad 1 \right] \cdot S'$.

The DAE part of the Sigma-Delta digital speaker is of similar concept to that of the PWM, since each Sigma-Delta bit has the same "weight". Every bitstream is directly fed to a unique DAE element. Thus, the total number of elements in this case is $N' = R$. Buffering is required in this case, in order to synchronize the Sigma-Delta frame reproduction.

## 6.2. Results

### *6.2.1. PCM DTA*

A previous work [9] has investigated the harmonic distortion produced by a digital array loudspeaker, using analytic approximation of the produced sound pressure. For the work described here, a novel arithmetic approach was applied, based on the vector addition of the pressures produced by the array elements. Using this approach, the performance of digital transducer arrays can be easily extended for non-tone audio test signals. Moreover, while the following analysis considers ideal array components (e.g. components with impulse response equal to the dirac function), it can be also extended to practical, non-ideal elements.

During the performed tests, a PCM Digital Speaker based on the topology shown in Fig. 12 was simulated for N'=6bits. In this Figure each number in the element array denotes the bit that drives the corresponding component, with 1 being the Most Significant Bit (MSB) as in [9]. Hence, the DAE module consists of 63 miniature array elements. The total array dimensions are 9d x 9d (m) where d is the distance between the miniature elements. For simulation purposes, d=10 and 20mm. The audio performance of the system was measured in terms of the resulting Total Harmonic Distortion (in dB) using 0dB-FS digital sinewave inputs of varied frequency (shown in Table 3), for two different receiver's positions: (a) On axis and (b) 45° relative to axis. For all cases the array to receiver distance was 1m.

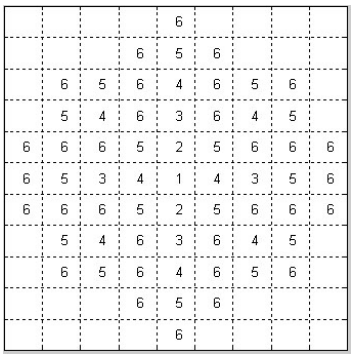



**Fig. 12.** Simulated PCM-based DAE module topology (after [9])

| **Parameter** | **Values** |
|---|---|
| Element distance (d) | 10mm, 20mm |
| N' (bits) | 6 |
| Receiver distance | 1m |
| Receiver position | On axis, $45^0$ |
| Sinewave amplitude | 0dB-FS |
| Sinewave frequency (Hz) | 125, 250, 500, 1k, 5k, 10k |
| Oversampling factor (R) | 1, 4 |
| Noise shaping | $3^{rd}$ order (with oversampling) |

**Table 3.** Simulation parameters for PCM-based DAE module

The simulations were performed for the combination of all the test parameters in Table 3. The sound pressure produced at the receiver position was derived as a function of time by adding together the vectors of the pressure produced by each array element driven by the appropriate bit value, as explained in Section 6.1. In order to retain the discrete time nature of the derived pressure, spatial quantisation was applied prior the vector addition. The final output pressure spectrum was then calculated using common FFT routines.

A typical example of the spectrum produced by the array on axis is shown in Fig. 13 for the case of 5kHz digital input and for d=10mm (without oversampling). Clearly, the sound pressure produced is a very close approximation of the digital input signal. Given the dynamic range of the output signal, it is obvious that the bit resolution of the input signal (N'=6bits) is preserved after audio reproduction. Hence, an increment of the total number of array elements would cause a corresponding improvement in the dynamic range.

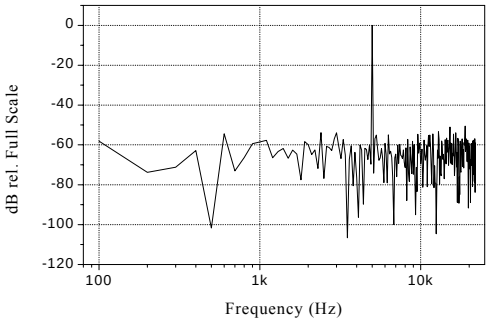



**Fig. 13.** Typical PCM DAE spectrum for 1kHz digital input without oversampling

The above measurements were verified by all test cases considered. Figures 14 and 15 show the measured THD+Noise (dB) of the output pressure spectrum as a function of the input sinewave frequency leading to the following conclusions:

(i) The measured THD+Noise increases with the element distance d, due to the longer time-delays errors introduced by the different sound propagating paths between each array element and the receiver's position. These time-delays increase with the distance d and are magnified by the spatial quantisation mechanism.

(ii) For out of axis receiver positions, the distortion values significantly increase. It was found that due to the increment of the time-delays, moving out of array axis raises harmonic and inetrmodulation distortion, especially for high input frequencies, as shown in Fig. 15.

(iii) When oversampling is applied, the benefits of increasing the input sampling rate are only evident when the element distance d is small. In addition, for large values of d, the measured distortion increases with the oversampling factor.

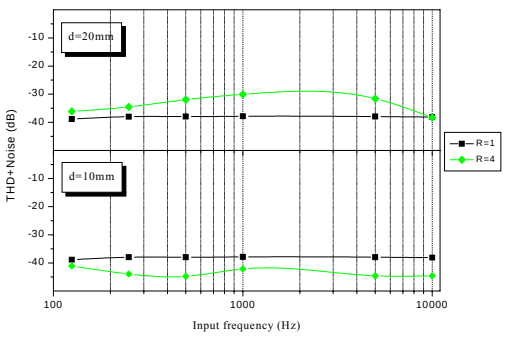



**Fig. 14**. On Axis measured THD+Noise (dB) for PCM Digital Speaker

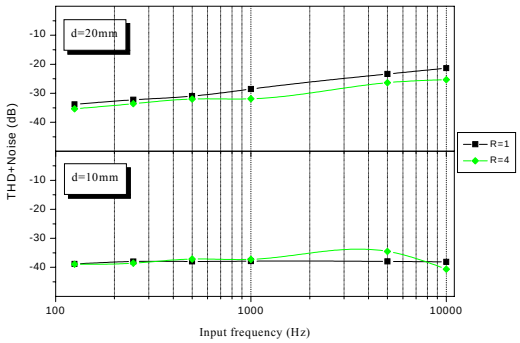



**Fig. 15**. $45^0$ measured THD+Noise (dB) for PCM Digital Speaker

### *6.2.2. PWM DTA*

In order to simulate a PWM speaker setup, the analytic PCM-to-PWM mapper [4] was used to provide the PWM bitstream without any correction of distortion via "jithering".

The DAE module consists of 510 miniature components, placed in a straight line, thus measuring a total length of 510d, where d is the diameter of each element. Even for the diameter d=10mm used, the digital speaker is quite lengthy, measuring over 5m. It is apparent that because of its size, the PWM DTA sound will be greatly distorted especially in locations who's distances from the DTA are comparable to its length.

The elements are able to reproduce either a positive or a negative pulse sound pressure. When a "0" bit is fed to the element a negative pressure is produced, while when a "1" bit is fed, a positive pressure is emitted. It is obvious that before each frame of data is fed, a "reset" phase must precede to facilitate the emission regardless of the elements' previous condition.

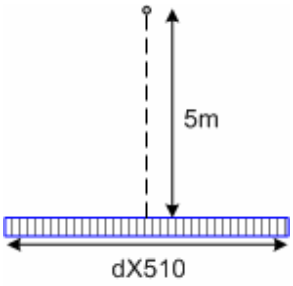



**Fig. 16:** Simulation Position for the PWM speaker

For the results presented, the pressure level was recorded in one position, as seen in Fig. 16, 5m away from the center of the DAE.
In Fig. 17 the Total Harmonic Distortion (THD) vs frequency can be seen, for the above mentioned position. Measurements were obtained for sine signals of 125, 250, 500 1000, 5000 and 10000Hz. The THD steadily increases with frequency, partly because of the 8 bit PWM representation but also due to spatial aliasing.
However, in simulations performed with oversampling (both R=2 and 4), and with no modifications made to the DAE physical setup, the speaker yields poor results, i.e. positive THD dB values for most of the cases under examination. The issue will be addressed in further studies by the authors.

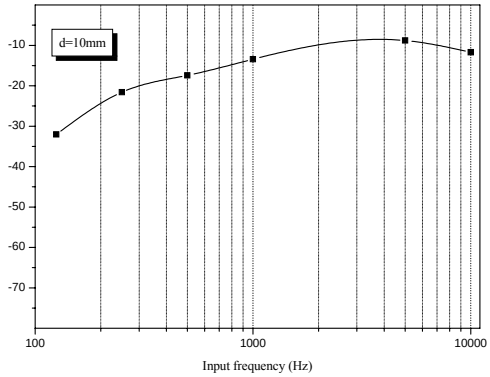



**Fig. 17:** PWM speaker THD vs frequency on axis. The simulation was performed for a distance of 5m and for d=10mm

*6.2.3. Sigma – Delta DTA*

In order to simulate the S-D digital speaker setup, the Delta-Sigma toolbox [16] was used to provide the S-D bitstream. The modulator is 5$^{th}$ order, while the oversampling ratio R varies from 32 to 128fs and $f_s$=44.1KHz. A typical output spectrum of the modulator can be seen in Fig. 18, for a sine wave input, frequency 5KHz.
The DAE module consists of 32 miniature components, placed in a straight line, thus measuring 23d, where d is the diameter of each element. For this study, the diameters used were d=10mm and d=20mm. The elements work in the same way as for the PWM speaker.

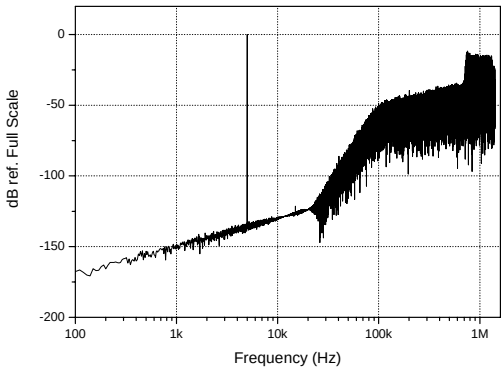



**Fig. 18:** Typical S-D bitstream spectrum

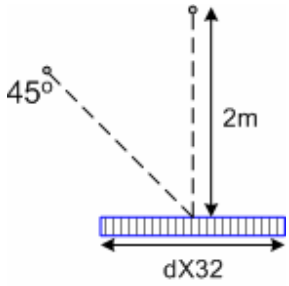



**Fig. 19:** Simulation Positions for the S-D speaker

For the results presented, the pressure level was recorded in 2 positions, as seen in Fig. 19, ranging 2m away from the center of the DAE, and for angles of 45 deg. and on axis. A typical spectrum of the pressure is shown in Fig. 20. While the input S-D bitstream has an approx. Signal to Noise Ratio (SNR) of 140dB, the sound produced from the DTA has an approx. 50dB SNR.

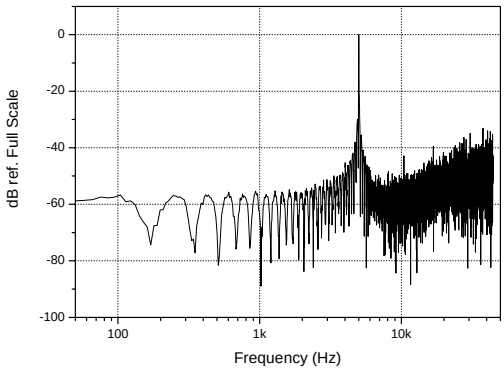



**Fig. 20:** Typical S-D DTA output spectrum

The test parameters were the same as for the PCM case, with the oversampling ratio varying between 32, 64 and 128$f_s$ The results were obtained for the

THD of the output signals. From these results it can be deducted that:

(i) THD depends heavily on the position. Regardless of the other parameters, THD values are less for on axis position than for 45 degrees. Acceptable reproduction for the frequency band under study can be achieved for the on axis position and for element diameter of 10mm.

(ii) THD generally increases with frequency. An exemption is observed for the on axis position, for 10mm elements, where the THD seems to decrease for higher frequencies.

(iii) The THD curve reaches a maximum, which depends on the diameter of the elements. The larger the element, the higher the frequency for the maximum THD value. However, in most cases, the THD is less for smaller elements.

(iv) Oversampling improves THD values for almost all frequencies and positions, for a given DTA physical setup. The average improvement is 3dB for when doubling R. This seems to be the greatest difference between the PWM DTA and the S-D DTA.

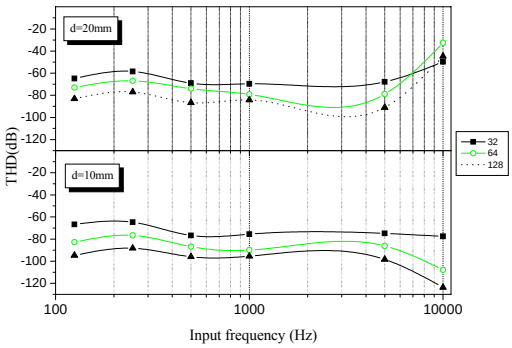



**Fig. 21:** S-D speaker THD vs frequency for different values of R and for the on axis position.

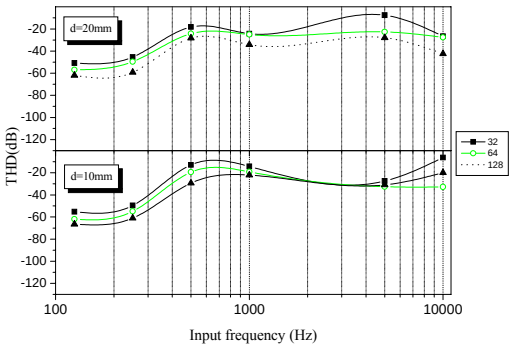



**Fig. 22:** S-D speaker THD vs frequency for different values of R and for the 45° position.

### *6.2.4. DTA Conclusions*

In the previous sections various, setups for digital loudspeakers were presented. All of the above setups share the same type of DAMP and DEA elements, arranged in different topologies and numbers. In Table 4 the total number of transducers and their maximum switching frequency is presented.

| Digital Speaker | Transducer Number | Transducer Frequency |
|---|---|---|
| PCM | $2^{N'}$-1 | $Rf_s$ |
| PWM | $2(2^{N'}$-1) | $Rf_s$ |
| SDM | R | $f_s$ |

**Table 4:** Different DTA setups (theoretical)

The number of elements and corresponding frequency for the results presented in this study are lined up in Table 5.

| Digital Speaker | Transducer Number | Transducer Frequency |
|---|---|---|
| PCM | 63 | 44-176KHz |
| PWM | 510 | 44-176KHz |
| SDM | 32 | 44-176KHz |

**Table 5:** Different DTA setups (practical)

The DSP functions (oversamping, requantization, noise shaping) for the above Digital Speakers should be considered trivial and any modern DSP processor [20]. However, for the buffering (or de-multiplexing in the case of PCM) stage, an FPGA is required in order to drive the separate DAMPs. With FPGAs supporting frequencies well above 200MHz [21], the design should be straightforward and aprox. 200 gates would be sufficient. For the latter case, a DSP core could be incorporated with the buffering stage in one ASIC, providing a single chip solution.
Practical limitations (large number of elements) deem the PWM speaker non-practical. Also, the results obtained were not in any case satisfactory for this type of speaker. On the other hand, the Sigma-Delta speaker physical characteristics as well as the simulation results indicate that an implementation is possible, though directivity issues have to be further investigated.

## 7. CONCLUSIONS

Bluetooth and 802.11a/b protocols are widely used for interconnecting computers and other portable electronic devices in a cable-free environment.

However, none of the above protocols inherently support digital audio transmission and modifications must be made to add appropriate mechanisms. The available raw bandwidth for each protocol is an indication of the maximum sound quality achievable, Bluetooth being only appropriate for compressed audio data transmission, though synchronization between receivers is dependant on a number of other parameters.

In many ways, the problem of audio decoding at each receiver seems to be rather trivial and well within the reach of current VLSI technologies, taking into account that many such products already exist in audio appliances. Taking it as prerequisite that compressed audio will have to be decoded at the receiver, the results of this study have also indicated that it may be advantageous to consider decoded streams of many alternative forms (i.e. multibit PCM, SDM or PWM), which may offer advantages for each possible application and receiver configuration.

Acoustic adaptation and compensation of each receiver (channel) is also a viable proposition, improving objective as well as subjective performance and reducing the requirements for critical placement of these integrated devices within any listening environment. Such FIR operations will clearly increase processing overheads, but it is within the reach of current DSP devices. Directivity control is also an important issue, which may improve performance and functionality, 'though it was not considered in this study.

The choice of PWM-based digital amplification appears to be advantageous when compared to PCM or SDM options, provided that appropriate preprocessing has been realized in order to alleviate potential distortion artifacts. Given that the preliminary tests described in the text indicate that PWM-based direct digital to acoustic transduction is inappropriate for high-quality audio reproduction, it appears that flexible and transparent transformation (mapping) of the datastreams between PCM, SDM and PWM should be included in such future integrated systems.

The direct digital to acoustic transduction via DTAs, clearly presents the most challenging and lesser evolved element in the complete chain. Preliminary simulation tests indicate that total array dimensions (i.e. a combination of the required number of elements and the individual element size), is a crucial parameter, the performance being better for smaller-sized DTAs. Furthermore, it was verified (see also [9]) that oversampled, low-bit PCM streams can be acoustically reconstructed with little extra loss in audio performance, using a physical DTA set-up of manageable complexity. On the other hand, SDM direct digital to acoustic transduction via DTAs, appears to achieve even better on-axis performance, requiring also smaller array dimensions and complexity.

As it is also obvious, many real-life DTA issues such as element technology, power handling and low-frequency extension, were not examined in this study and will have to be investigated. Nevertheless, apart from such practical issues, it is possible to conclude that digital audio and related technologies have now matured to a stage that most of the challenging aspects of integrating a networked all-digital audio / acoustic reproduction chain can be successfully met.

## 8. REFERENCES


[1] J. D. Martin, "Theoretical Efficiency of class-D Power Amplifiers", IEE Proc., Vol. 117, No. 6, June 1970.

[2] A. J. Magrath, M. B. Sandler, "Digital-Domain Dithering of Sigma-Delta Modulator Using Bit Flipping", *J. Audio Eng. Soc.*, Vol. 45, No. 6, pp. 467-475, June 1997.

[3] A. J. Magrath, M. B. Sandler, "Digital Power Amplification Using Sigma-Delta Modulation and Bit Flipping", *J. Audio Eng. Soc.*, Vol. 45, No. 6, pp. 476-487, June 1997.

[4] A.C. Floros, J. N. Mourjopoulos, "Analytic Derivation of Audio PWM Signals and Spectra", *J. Audio Eng. Soc.*, Vol. 46, No. 7/8, pp. 621-633, July/Aug. 1998.

[5] A.C. Floros, J. N. Mourjopoulos, "On the Nature of Digital Audio PWM Distortions", Audio Eng. Soc. 108th Convention, preprint 5123, Feb. 2000.

[6] A.C. Floros, J. N. Mourjopoulos, "A Study of the Distortions and Audibility of PCM to PWM Mapping", Audio Eng. Soc. 104th Convention, preprint 4669, May. 1998.

[7] J.M. Goldberg and M.B. Sandler, "Noise Shaping and Pulse-Width Modulation for an All-Digital Audio Power Amplifier," *J. Audio Eng. Soc.*, vol. 39, pp. 449-460, June 1991.

[8] A.C. Floros, J. N. Mourjopoulos, D. E. Tsoukalas, "Jither: The Effects of Jitter and Dither for 1-bit Audio PWM Signals", Audio Eng. Soc. 106th Convention, preprint 4656, May. 1999.

[9] S. C. Busbridge, P. A. Fryer, Y. Huang, "Digital Loudspeaker Technology: Current State and Future Developments", Audio Eng Soc. 112th Convention, May 2002

[10] M.O.J. Hawksford, "Smart directional and diffuse digital loudspeaker arrays", Audio Eng. Soc.110th Convention, May 2001

[11] A. Floros, M. Koutroubas, N.A. Tatlas, J. Mourjopoulos, "A Study of Wireless Compressed Digital-Audio Transmission', Audio Eng Soc. 112th Convention, May 2002

[12] Bluetooth SIG, Bluetooth Baseband Specification Version 1.1.

[13] IEEE 802.11 Standard, Part 11 Medium Access Control (MAC) and Physical Layer (PHY) specifications, 1999.

[14] Bluetooth SIG, Advanced Audio Distribution Voting Profile version 0.9, 2001 (Confidential).

[15] M-T Sun, L. Huang, A. Arora, T-H Lai, "Reliable MAC layer multicast in IEEE802.11 Wireless Networks"

[16] R. Schreier, "The Delta-Sigma Toolbox Version 6.0", Analogue Devices

[17] A.J. Magrath, M.B. Sandler, "Hybrid pulse width modulation/sigma-delta modulation power digital-to-analogue converter"

[18] P. Hatziantoniou, J. Mourjopoulos, "Generalised Fractional-Octave Smoothing of Audio and Acoustic Responses", *J. Audio Eng. Soc.*, vol. 48, No 4, pp. 259-280 (2000 April).

[19] P. Hatziantoniou, J. Mourjopoulos, "Results for Room Acoustics Equalization Based on Smoothed Responses", to be presented at the AES 114th Conv., Amsterdam, The Netherlands, 2003.

[20] www.ti.com, "TMS320DA610 Aureus(TM) audio processor"

[21] www.xilinx.com: "Xilinx FPGA Product Tables"

[22] B.Fox, "The First Digital Loudspeaker?", Hi-Fi News Rec. Rev., IPC Media (Nov. 1998)