Chapter 6.3

HARDWARE FOR AMBIENT SOUND REPRODUCTION

Ronald M. Aarts
Philips Research Eindhoven
ronald.m.aarts@philips.com

Abstract Today's and tomorrow's audio and video applications put increasing demands on sound reproduction techniques, particularly because of the advent of ambient intelligence (AmI). A good sound reproduction system is generally in conflict with the boundary conditions posed by AmI, both by size as well as by setup flexibility. Hence, improving the sound quality within these conditions is important, because the traditional means have difficulties with these constraints. Various new and old means for sound reproduction are discussed as possible candidates, including "singing display," and "BaryBass": the former uses a display as a sound generator; the latter a system, which maps the low frequency region (20–120 Hz) onto a single tone, and uses an extremely efficient transducer at that particular tone. Apart from the transducers, various other options are discussed to relax the boundary conditions of traditional sound reproduction setups required by AmI.

Keywords barybass; driver; force factor; headphones; incredible surround; loudspeaker; phantom source; sound reproduction; ultrabass

1. INTRODUCTION

Before and after the "birth" of the classical electrodynamic loudspeaker in 1925, various other concepts appeared [1–6], and some of them have left the scene, to mention a few:

- laser loudspeakers using the photo acoustic effect [7];
- loudspeaker arrays, consisting of various drivers;

Mukherjee
Hardware Technology Drivers of Ambient Intelligence

- "audio spotlight" using interfering ultrasonic sound beams [8–9];
- "flame loudspeaker" and "Ionophone," using pyroacoustic transduction;
- vibrating panels;
- (digital) sound projector (see Figure 6.3-5);
- headphones;
- neck-sets (see Figure 6.3-7);
- electromagnetic loudspeakers [1];
- piezo loudspeakers [1];
- electrostatic loudspeakers [1];
- vibrating (LC) Displays ("Singing Display," based on electrostatic forces) [6]; and
- BaryBass, a resonant loudspeaker (see Figure 6.3-10–6.3-12) [5].

Some of these systems will be discussed later, for the others, one is referred to the bibliography section. For ordinary living room applications, classic sound reproduction by ordinary loudspeakers will do for most of the time, however, for ambient audio, we might need some of the above-mentioned alternatives. The reasons can be due to size, privacy, but also whether it is to be produced locally, or it is sound for everybody, or perhaps even sound, which follows you in any room of the house.

In the following we will show some special loudspeaker systems, and then we will present various techniques to overcome several problems with traditional sound reproduction.

1.1. "Flat-Pack"

SoundpaX loudspeakers (from NXT) are "flat-pack," corrugated cardboard loudspeakers. They are very light and easy to foldaway. An example is shown in Figure 6.3-1.

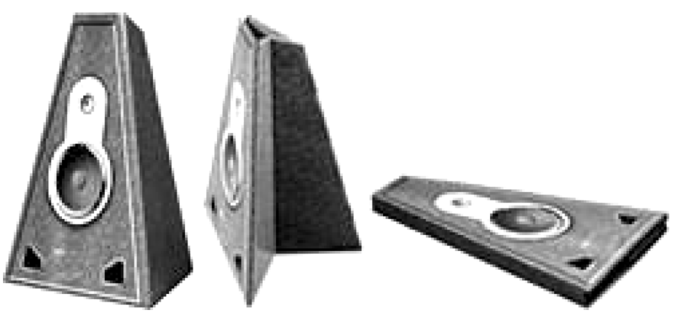

Figure 6.3-1. SoundpaX loudspeakers are "flat-pack," corrugated cardboard loudspeakers.

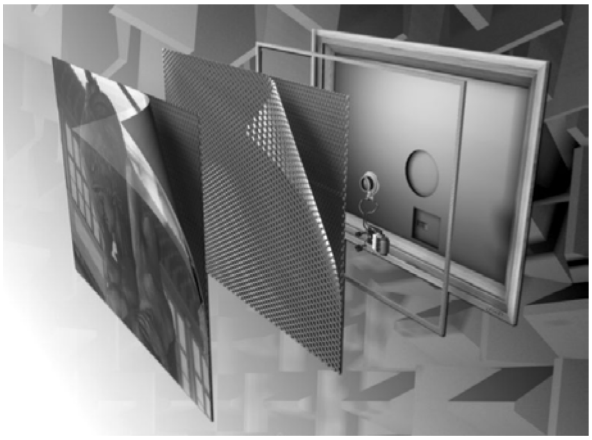

Figure 6.3-2. New generation speaker which can blend invisibly into your room, as the detachable frames allow you to insert your favorite prints.

1.2. Picture Frame

Another example is given in Figure 6.3-2; this new generation speaker by NXT's technology may be successfully applied to a wide variety of applications: multimedia, plasma TVs, home stereos, architectural acoustics, and consumer electronics. They are lightweight and flexible speakers that can reproduce high- to midrange frequencies. The panels blend invisibly into your room, as the detachable frames allow you to insert your favorite prints.

1.3. "Singing" Display

Singing Display [6] has as aim to generate sound from the display itself to save component costs and miniaturize audio-visual products.

1.3.1. Background

As displays become more pervasive, many products are becoming audio-visual. A case in point is the GSM telephone, where the display has become indispensable. Indeed, the GSM display is becoming so large to allow for games, Internet, video, etc., whilst the phone itself is becoming so small, that there is little room to accommodate the loudspeaker or microphone, which is still required for the primary GSM function (i.e., making a phone call).

1.3.2. "Singing" display

It is proposed to avoid this issue by making use of the display itself to produce (loudspeaker) or detect (microphone) the audio signals. GSM products are shown in Figure 6.3-3. In the GSM in the middle, the "singing display" has taken over the role of the loudspeaker, whilst on the right hand side, the super "singing display" has taken over the role of both loudspeaker and microphone and, optionally, the keyboard.

It is clear that this approach not only saves space and weight in the product, but also reduces the component count and could hence make the product cheaper. Whilst the idea is illustrated with a GSM embodiment, its scope of application is all possible audio-visual products due to the cost savings, particularly for portable applications (where space/weight saving becomes essential). The most common display used in portable applications is the LCD. LCDs come in many modes (TN, STN, MVA, etc.), and types (passive or active matrix), but have a common feature that they comprise a thin layer of electrooptic material sandwiched between two substrates and are driven using an electric field.

1.3.3. Layout

The proposal is to exploit the specific geometry of the LCD to induce an acoustic output. The geometry is shown schematically in Figure 6.3-4.

The LCD is driven by applying an (AC) voltage to the electrodes, which are either transparent (ITO) or reflective (Al). The observation is that under certain conditions, it is possible to use the applied voltage to cause the LCD to vibrate and create an acoustic output. The vibration is

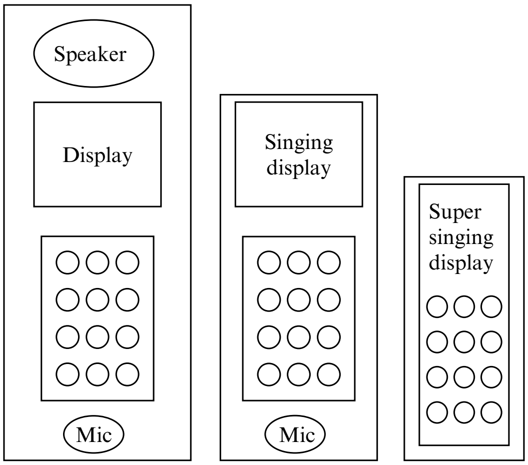

Figure 6.3-3. Traditional GSM telephone: GSM with "Singing Display" (no loudspeaker) and with super "Singing Display" (no loudspeaker, keyboard, or microphone).

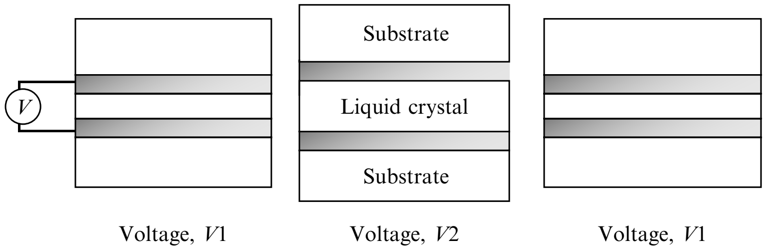

Figure 6.3-4. An LCD consists of two substrates with electrodes (grey area). The liquid crystal is situated between the substrates. Application of a variable voltage causes the cell to vibrate.

caused by electrostatic forces across the liquid crystal layer. The frequency and magnitude of the output is tunable both in frequency and amplitude, depending upon details of the applied voltages. By correct application of the applied voltages, it will be possible to use the display as a loudspeaker. The singing display concept has been proven, however, the acoustic output of the singing display is currently too low to get a sufficient sound pressure level.

1.4. Sound projector

1Limited's Digital Sound Projector is a single slim panel that connects directly to a DVD or CD player. By producing tight, focusable beams of sound, the sound projector beams the separate sound channels around the listener's room. By reflecting off walls and other surfaces in the room, these beams finally come to the listener from left and right, front and rear; see Figure 6.3-5. This single unit replaces a more conventional five loudspeaker setup for surround sound reproduction.

2. WEARABLE AUDIO MODULES

In September 2000, Philips and Levi's launched wearable electronics products. The product range is branded industrial clothing design (ICD+), and consists of four different jackets. Each of the four styles contains a simple body area network using wires integrated into the jacket design. This network allows the synchronous control of the Philips Xenium GSM mobile and Philips Rush MP3 player through the use of a unified remote control. A multidisciplinary team of textile designers, electronic engineers, and product designers have been working together on wearable electronics at Philips Research in Redhill, UK. An example is shown in Figure 6.3-6.

Figure 6.3-5. 1Limited's Digital Sound Projector is a single slim panel producing tight, focusable beams of sound, beaming the separate sound channels around the listener's room.

Another example is by Infineon, which has developed a prototype audio module for the integration in clothes. In addition to ensuring the functionality—for example, as an MP3 player—special attention was paid to a robust and textile-ready design. The components are designed so that the electronics and the interconnections between the textile structures do not interfere with a comfortable wear, allow for easy and convenient use, and allow the clothing to be washed without the need to remove the electronics. A flat keyboard is built with metallized films on an electrically conductive fabric strip. The metal films are attached with an adhesive that is commonly used in the clothing industry. A tiny sensor module is connected to the metal films and registers when the pads are "pressed." The earplug microphone set is also connected to the audio module through the fabric strip.

3. MULTICHANNEL AUDIO

The presence of digital versatile disk (DVD) and super audio CD (SACD) has made multichannel audio popular in sound systems for consumer use today. Here, a method is presented, which converts

Figure 6.3-6 An example of Philips' and Levi's wearable audio devices.

two-channel stereo to multichannel sound reproduction using a three-dimensional (3D) representation [10] (hereafter referred to as ''space mapping''). Although many have introduced multichannel sound systems with a large number of channels, we restrict ourselves to a home cinema setup, for which investigations have shown that five channels are sufficient. This setting is adopted from multichannel configuration with three loudspeakers placed in front of the listener, and the other two at the back. By use of principal component analysis, we developed an algorithm that produces a vector, which indicates the direction of both dominant signal and remaining signal. These two signals are then used as basis signals in the matrix decoding. It offers two improvements above existing multichannel techniques. Firstly, a problem associated with channel cross talk is reduced, and therefore better sound localization is achieved. The latter gives more space to the listener to enjoy the offered program rather than the restricting listening area referred to as the ''sweet spot.'' Secondly, a better sound distribution to the surround channels is achieved by using a cross-correlation technique, while maintaining energy preservation. So, it remains backward and forward compatible with ordinary stereo.

3.1. The Center Channel

We consider the three-channel approach in particular. It is known that the sound quality of stereo sound reproduction can be improved by adding an additional loudspeaker between each adjacent pair of loudspeakers. This additional center loudspeaker can be fed with the sum signal of the left and right channel. A major drawback of this approach is that cross talk with left and right channels is inevitable, and resulting in a narrowing of the stereo image. However, we derived a center channel's gain using the direction of a stereo image, which is time varying. It automatically tracks the main direction of the dominant signal.

4. POSITION INDEPENDENT STEREO

Another method to achieve correct localization for stereophonic sound reproduction in a wide listening area is to use a loudspeaker array and so-called time-intensity trading, a mechanism of the human auditory system determined via psychoacoustic experiments within a wide listening area [11]. The use of two spatially separated loudspeakers imposes restrictions on the ability of stereophony to reconstruct the correct acoustic field so that a sharp image can be perceived. Such a system can provide a well-defined image for a centrally located listener mainly at low frequencies, depending on the geometrical displacement of the speakers relative to the listener.

The basis of stereophony is the ability to create phantom sources. It is known that the brain locates a monophonic signal originated from a single source by comparing the differences in the arrival time and intensity of that signal at each ear. If the same monophonic signal is played through two loudspeakers on either side of the listener, then the sound seems to appear from midway between the two loudspeakers, since the traveling time of the signal arriving at each ear is the same. This is called a phantom (or virtual) source. We will discuss how to enlarge the region, within which the image remains reasonably. In general, it can be stated that correct localization within a wide listening area is beneficial for all applications, where a good stereophonic sound is required. The idea of achieving an enlargement of the sweet spot area in a stereophonic setup has been introduced and studied at the Philips Research Labs, Eindhoven, and the stereo sound system has been called "Position Independent" (PI). The main idea is that the directivity pattern of a loudspeaker array should have a well-defined shape so that a good stereo sound reproduction is achieved in a large listening area. Optimal digital filters are then designed and applied to individual drivers of linear loudspeaker arrays in order to obtain a directivity pattern of a specific shape. This shape has then to be adapted to the time intensity trading mechanism of the human auditory system. The goal here is to derive an optimal directivity for the PI-stereo system, which is based on parameterized time intensity trading data, and then to find, by means of an optimization process, the corresponding FIR filter coefficients that achieve this optimal directivity pattern. It has been proven that an optimal directivity pattern for a loudspeaker can be realized by using an array of drivers positioned at a specific distance from each other. In our case, a practical design to achieve PI-stereo sound reproduction is a pair of loudspeaker cabinets, each cabinet equipped with a pair of drivers for the high frequency range with a separation suitable for this frequency range and for the mid frequencies with a corresponding separation, so as to obtain the desired optimal directivity pattern.

5. ULTRA BASS

In many sound reproduction applications, it is not possible to use large loudspeakers, due to size and or cost constraints. Typical applications are ambient audio, but also portable audio, multimedia, and TV. In these applications, the devices are often of small size, and therefore the transducers are inherently small as well. Needless to say, the competitive market also dictates the highest possible audio quality of these products. However, probably the most well-known characteristic of small loudspeakers is a poor low frequency (bass) response. In practice, this means that a significant portion of the audio signal may not be reproduced (sufficiently) by the loudspeaker. For loudspeakers used in applications as mentioned above, reproduction below 100 Hz is usually negligible, while in some applications, this lower limit can easily be as high as several hundred Hertz. The bass portion of an audio signal contributes significantly to the sound "impact," and depending on the bass quality, the overall sound quality will shift up or down. Therefore, a good low-frequency reproduction is essential.

A traditional and conceptually very simple method to increase the perceived sound level in the lower part of the audible spectrum (below the loudspeaker's resonance frequency, which is usually the lower limit) is to amplify the low frequency part of the audio spectrum, by a fixed or dynamic (depending on signal amplitude and or reproduction level) amount. For very low frequencies, the mechanical limits of the loudspeaker will limit the stroke the cone can make, leading to distortion and possibly, loudspeaker overload. Thus, physically increasing the radiated sound pressure level means forcing the loudspeaker to radiate sound in a frequency range, for which it is not equipped. It may be better to prevent this completely by methods outlined below. In the process we shall discover several advantages of these methods. Now, from psychoacoustic theory, we know that a pitch perception can occur at a frequency that is not contained in the audio signal. This is possible through nonlinearities in the cochlea (difference tones), or a higher-level neural effect in the auditory system (virtual pitch). These two effects, appear to be very suitable effects for our purpose of enhancing bass perception using small loudspeakers. These effects can be utilized by some simple nonlinear (but controlled) processing, replacing very low frequencies in the audio signal by higher frequencies [12, 13]. These will still have the same perceived pitch as the original, using the psychoacoustic effects previously mentioned. Such effects also occurred in transistor radios, where undesired nonlinearities gave rise to a distorted sound. However, the method that we now propose uses nonlinearities in a controlled manner, and restricted to only the lowest frequencies, such that the effect is to our benefit. Without any information about the signal processing employed, we can immediately infer a number of advantages that such a scheme shall provide:

A higher radiated sound pressure level from a given loudspeaker, because of increased efficiency and decreased cone excursion. Furthermore, at higher frequencies the auditory system is more sensitive, which will also contribute to increased loudness;

Less power consumption, because of increased efficiency. This can be very important for portable applications; and,

Fewer disturbances in neighboring areas, because of the fact that the lowest frequencies are not physically present, while the added higher frequencies are absorbed more efficiently than the low frequencies.

6. INCREDIBLE SURROUND SOUND

It is virtually impossible to imagine sound reproduction today without stereophonic techniques, and it is to the credit of both the technology and human binaural hearing capabilities that a single pair of loudspeakers can evoke auditory perspectives so convincingly. Incredible Sound is a convincing stereo base-widening system, developed to improve the sound reproduction in applications with closely separated loudspeakers [14]. The aim of incredible surround sound is to offer a practical solution, replacing the traditional approach generally used. A filter is derived, using a simple model, where ideal loudspeakers and an acoustically transparent subject's head are assumed. This system appeared to be very practical to implement and tolerant against head movements.

7. NOMADIC RADIO: WEARABLE AUDIO MESSAGING AND AWARENESS

Nomadic Radio developed at MIT Media Laboratory is as a unified messaging system that utilizes spatialized audio, speech synthesis, and recognition on a wearable audio platform. A client–server-based messaging infrastructure is already in place, and support is added for communication and location awareness. Messages such as hourly news broadcasts, voice mail, and email are automatically downloaded to the device throughout the day. The current system operates primarily as a wearable audio-only interface, although a visual interface is used for development purposes. A combination of speech and button inputs are used to control the interface. Textual messages such as email, calendar reminders, weather forecasts, and stock reports are delivered via synthesized speech. Users can select a category, such as news or email, browse messages sequentially, and save or delete them on the server. As the system gains the capability to determine its location, a scenario is envisioned, where the listener's location context enables the system to provide relevant messages as needed. For example, as the user moves close to a particular room, she may hear a voice message left by a colleague, or more importantly she is reminded of a meeting if she is not in a desired location at a specific time.

7.1. Design of the Wearable Audio Platform

Audio output on wearables requires use of speakers worn as headphones or appropriately placed on the listener's body. Headphones are not entirely suitable in urban environments, where users need to hear other sound sources, such as traffic or in offices, where their use is considered antisocial as people communicate frequently. In these situations, speakers worn on the body could instead provide directional sound to the user (without covering the ear), yet they must be designed to be easily worn and least audible to others. The *Soundbeam Neckset* (shown in Figure 6.3-7), worn around the neck, has been modified for audio I/O from the wearable. The *Neckset* is a patented research prototype originally developed by Andre Van Schyndel at *Nortel* for use in hands-free telephony. It consists of two directional speakers, mounted on the user's shoulders, and a directional microphone placed on the chest. A button on the *Neckset*

Figure 6.3-7. MIT's Soundbeam Neckset.

will activate speech recognition, or deactivate it in noisy environments. Spatialized audio is rendered in real-time and delivered to the *Neckset*.

8. THREE-DIMENSIONAL (3D) HEADPHONES

Headphone virtualizers that are commercially available today are optimized for a head other than that of the listener. This results in large localization errors for most listeners. At Philips Research, a system is introduced that includes a calibration procedure, which can be carried out conveniently by the listener [15, 16]. This system consists of ordinary headphones, into which microphones have been mounted. The sound reproduction using headphones then gives the same listening experience to the user as if the reference multichannel loudspeaker system was being used. Besides the usual computational requirement for a headphone virtualizer, this system needs in addition two low-cost microphones.

8.1. Technology Background

The way in which sound propagates from the loudspeaker towards the ear drums of the listener depends on the loudspeaker, the room, and the physical properties of the listener (e.g., the shape of the head, ears, and torso).

The physical properties of the head and outer ears of the listener modify the sound as it travels from the source to the eardrums. The transfer functions describing this sound propagation from multiple sound sources to both ears are known as head-related transfer functions

(HRTFs). Multichannel audio can be filtered with the HRTFs of the listener prior to headphone sound reproduction. If loudspeaker reproduction is emulated using headphones, compensation for the sound reproduction characteristics of the headphones is required. In this way the multichannel loudspeaker system can be emulated very accurately. When audio is filtered with HRTFs that are measured from another person, there are large errors in the vertical and front back localization. Therefore, a system is introduced that is personalized to the listener.

The system at hand consists of headphones with integrated microphones and a digital signal processing unit (DSP), to which the headphones are connected. During the calibration, the DSP is connected to a multichannel loudspeaker setup. A noise signal is played through each of the loudspeakers and is registered by the microphones. The DSP then computes how the sounds should be processed prior to headphone reproduction, such that exactly the same sound is generated at the position of the microphones, which are very close to the ears. When the calibration is completed, the listener can manually choose between loudspeaker or headphone sound reproduction, showing the capabilities of the system.

8.2. Hexaphone

A dedicated implementation of the work has been realized which is codenamed "Hexaphone." Two examples are shown in Figures 6.3-8 and 6.3-9.

Figure 6.3-8. Prototype of headphones with integrated microphones.

Figure 6.3-9. Prototype of earphones with integrated microphones.

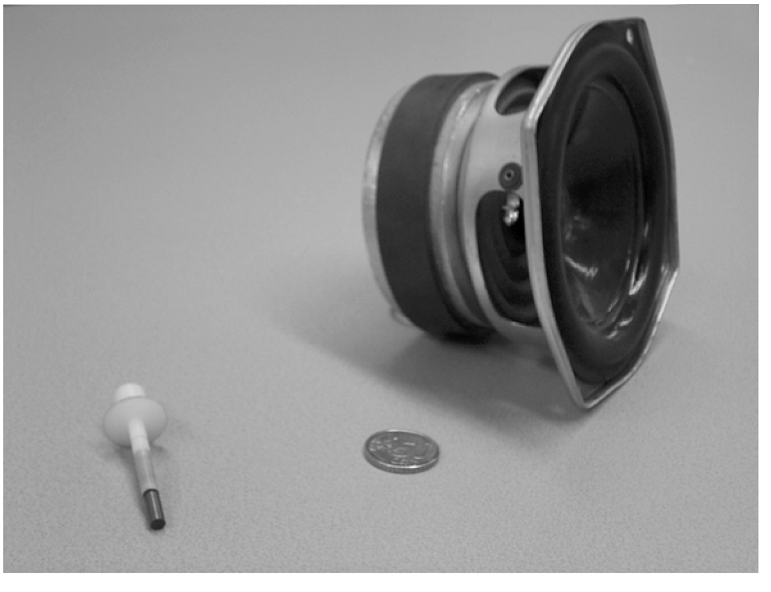

Figure 6.3-10. Left: the magnet system of the BaryBass transducer; right: a normal medium-sized bass loudspeaker. A 50 euro cents coin is shown for size comparison; the actual price is much less.

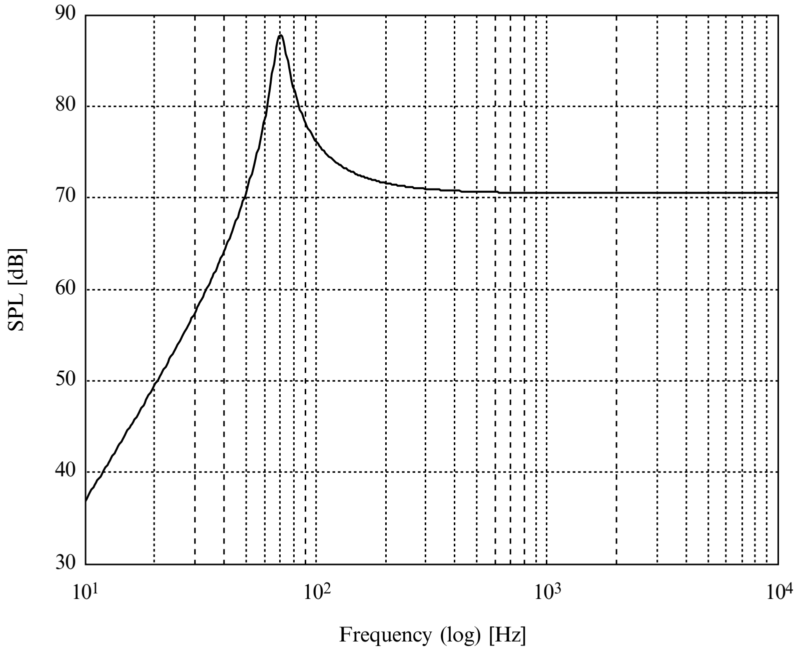

Figure 6.3-11. The response (Sound Pressure Level (SPL) [dB] versus frequency) of the BaryBass driver (log/log plot).

9. BARYBASS

Direct-radiator loudspeakers typically have a very low efficiency, since the acoustic load on the diaphragm or cone is relatively low compared to the mechanical load. On the one hand, the efficiency is inversely proportional to the moving mass, while on the other hand, it is proportional to the square of the product of the cone area and the force factor (determined by the magnet system and the voice coil). Furthermore, in order to get a sufficiently low resonance frequency, the moving mass must be high enough, and the cabinet volume—which acts as an air spring—must be large enough. However, for many consumer applications, the cone size should be small. In addition, the driving mechanism of a voice coil is quite inefficient in converting electrical energy into mechanical motion. These conflicting conditions cannot be met with a classical loudspeaker. Low frequency drivers (woofers) have a magnetic structure (see Figure 6.3-10, right side) that is rather large, so that the typical frequency response is flat enough and the efficiency is high enough. The solution consists of two steps [5]. First, we relax the requirement that the frequency response must be flat. By making

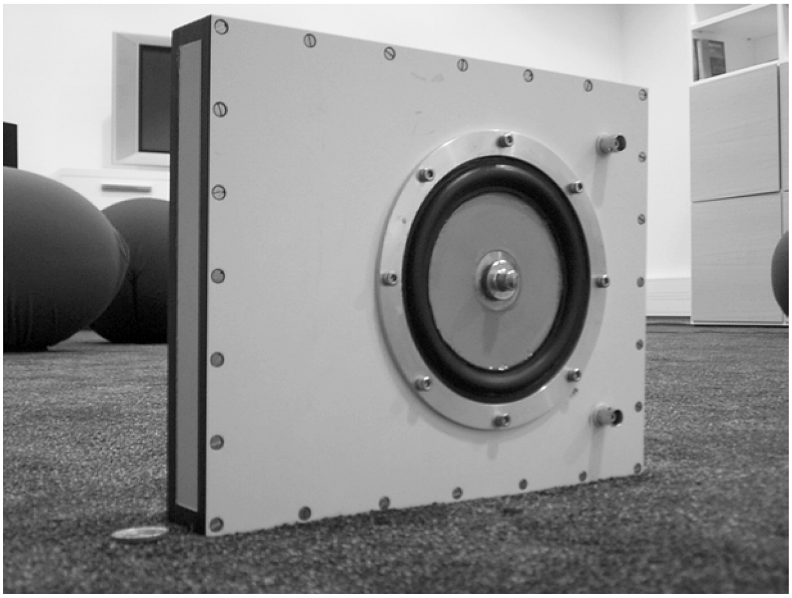

Figure 6.3-12. This new loudspeaker can be mounted in a very small volume and thin cabinets and blend invisibly into your room.

the magnet considerably smaller (see Figure 6.3-10, left side), a large peak in the sound pressure level (SPL) curve (see Figure 6.3-11) will appear. At the resonance frequency, the efficiency can be a factor 10 higher than that of a normal loudspeaker. In this case, we have at the resonance frequency of about 70 Hz—a high level of almost 90 dB @ 1 Watt input power, using only a small cabinet. Since it is operating in resonance mode only, the moving mass can be enlarged, without degrading the efficiency of the system. Due to the large peak, the normal operating range of the driver decreases considerably, however. This makes the driver not suitable for normal use. To overcome this, a second measure is applied. We map the low frequency content of the music signal, say, 20–120 Hz, to a slowly amplitude modulated tone, whose frequency equals the resonance frequency of the transducer. The modulation is chosen so that the coarse structure (the envelope) of the music signal after the mapping is the same as before the mapping. The required electronics is implemented both in the digital and in the analog domain, the latter one requiring less than a dozen transistors and a few RC components. An example of a BaryBass driver mounted in a flat enclosure with a volume of less than 1 l is given in Figure 6.3-12. The resonance frequency is about 50 Hz, which is probably, currently, the world's smallest subwoofer with such a low-resonance frequency, while the sound power efficiency is very high.

10. CONCLUSIONS

We have shown that using either special loudspeakers or signal processing, it is possible to go beyond the limits, which are usually dictated by physics. The reason that this is possible indeed is to utilize psychoacoustic phenomena, which relax these limits. This gives new opportunities for sound reproduction in general, but in particular in the ambient intelligence context.

REFERENCES

[1] Hunt, F. V., 1954, *Electroacoustics*, John Wiley.
[2] Olson, H. F., 1957, *Acoustical Engineering*, Van Nostrand.
[3] 1998, Commemorative issue: 50 years of contributions to audio engineering', *J. Audio Eng. Soc*. **46**(1/2).
[4] Vanderkooy, J., Boers, P. M. and Aarts, R. M., 2003, Direct-radiator loudspeaker systems with high Bl, *J. Audio Eng. Soc*., **51**(7/8), 625–634.
[5] Aarts, R. M., 2005, High-efficiency low-B1 loudspeakers, *J. Audio Eng. Soc*., 53(7), 579–592.
[6] Aarts, R. M. and Johnson, M. T., 2002, Sound and vision system, European Patent EP1386519, Filed 20 March 2002.
[7] Westervelt, P. J. and Larson, R. S., 1973, Laser-excited broadside array, *J. Acoust. Soc. Am*., **54**(1), 121–122.
[8] Yoneyama, M., et al., 1983, The audio spotlight: An application of nonlinear interaction of sound waves to a new type of loudspeaker design, *J. Acoust. Soc. Am*., **73**(5), 1532–1536.
[9] Pompei, F. J., 1999, The use of airborne ultrasonics for generating audible sound beams, *J. Audio Eng. Soc*. **47**(9), 726–731.
[10] Irwan, R. and Aarts, R. M., 2002, Two-to-five channel sound processing, *J. Audio Eng. Soc*., **50**(11), 914–926.
[11] Rodenas, J., Aarts, R. M. and Janssen, A. J. E. M., 2003, Derivation of an optimal directivity pattern for sweet spot widening in stereo sound reproduction, *J. Acoust. Soc. Am*., **113**(1), 267–278.
[12] Larsen, E. and Aarts, R. M., 2004, Audio bandwidth extension, *Application of Psychoacoustics, Signal Processing and Loudspeaker Design*, John Wiley.
[13] Larsen, E. and Aarts, R. M., 2002, Reproducing low-pitched signals through small loudspeakers, *J. Audio Eng. Soc*., **50**(3), 147–164.
[14] Aarts, R. M., 2000, Phantom sources applied to stereo-base widening, *J. Audio Eng. Soc*., **48**(3), 181–189.
[15] Aarts, R. M., 2004, Headphones with integrated microphones, US patent 6,829,361, December 2004.
[16] Schobben, D. and Aarts, R. M., 2005, Personalized multi-channel headphone sound reproduction based on active noise cancellation. *Acta Acoustica*, 91(3), 440–450.