Applications of DSP for sound reproduction improvement

Ronald M. Aarts[1]

[1] *Philips Research Laboratories, Prof. Holstlaan 4, 5656 AA Eindhoven, The Netherlands*

Correspondence should be addressed to Ronald M. Aarts (`Ronald.M.Aarts@philips.com`)

ABSTRACT

Today and tomorrow's audio and video, portable audio and multi-media applications put increasing demands on sound reproduction techniques. On one hand there is a need for reductions in both cost and size, on the other hand we wish to enhance the experience of the user beyond today's possibilities. A good sound reproduction system is in general in conflict with the boundary conditions for consumer products both by size as well as by price requirements. A possible way to ease these conflicts is to enhance the reproduction and perception of sound for listeners by exploiting the combination of psycho-acoustics, loudspeaker configurations and digital signal processing. Various examples will be given, such as increasing the perceived bass response of loudspeakers, and increasing the number of loudspeaker channels (converting stereo to multichannel sound). When multichannel reproduction through loudspeakers is not a viable option, the same percept can be simulated over headphones. This method, using active noise control principles, is discussed as well.

PART 1: TWO-TO-FIVE CHANNEL SOUND PROCESSING [1]

INTRODUCTION-PART 1

Since the introduction of digital versatile disk (DVD) and super audio CD (SACD), a revival of multichannel audio has appeared in sound systems for consumer use today. It is, however, desirable to maintain the compatibility with the existing two-channel stereo recordings and/or broadcasting. Therefore the conversion of two-channel stereo to multichannel format has been studied extensively over the decades, and a considerable number of publications exists [3–8]. Among these, Gerzon and Barton's is particularly notable in which many schemes have been proposed (see [7] and references therein).

Although many authors have introduced multichannel sound systems with a large number of channels, we restrict ourselves to a home cinema setup for which it has been shown that five channels is sufficient for creating ambience effects [9]. Hence in this part of the paper paper we focus on signal format conversion from two-channel stereo to five-channel (two-to-five) sound processing algorithm.

The desired setup is shown in Fig. 1, in which the channels are labelled L (left), C (center), R (right), S_L (left surround), S_R (right surround) according to convention. This setting is adopted from the ITU multichannel configuration [10], with three loudspeakers placed in front of the listener, and the other two at the back. The front channels are used to provide a high degree of directional accuracy over a wide listening area for front stage sounds, particularly dialogues, and the rear channels produce diffuse surround sounds providing ambience and environment effects. An additional loudspeaker (subwoofer) may be used to augment bass reproduction, which is often called 5.1 system with .1 referring to the low

[1] Based on papers [1, 2] presenting much more detail.

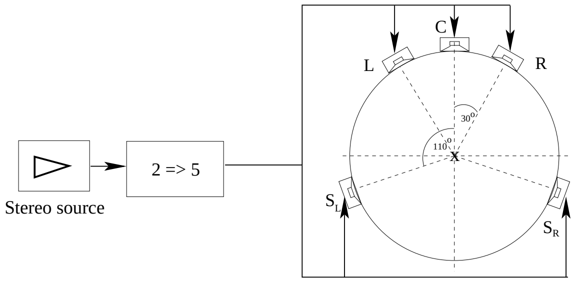

Fig. 1: ITU reference configuration [10]. The reference listening position (sweet spot) is indicated by $\mathbf{x}$. Left and right channels are placed at angles $\pm 30°$ from C, and the two surround channels are placed at angles $\pm 110°$ from C.

frequency enhancement (LFE) channel. In this paper, however, we do not use a subwoofer, since it can easily be extended when necessary without affecting the algorithm.

The algorithm presented in this paper offers two improvements above existing two-to-five channel sound systems. First a problem associated with channel crosstalk is reduced, and therefore sound localisation is better. Listening tests have confirmed that good sound localisation without the need to listen at the sweet spot gives more space to the listener to enjoy the program offered rather than restricting the listener to the sweet spot.

Second a better sound distribution to the surround channels is achieved by using a cross correlation technique. Surround channels are crucial in creating the ambience effects, which is one of the main goals of multichannel audio. At the same time, the energy preservation criterion is an important constraint that has been used to design multichannel matrices [8]. The main reason for this is to maintain backward and forward stereo compatibility. Furthermore, the preservation criterion ensures that all signals present in the two-channel transmitted signals are produced at a correct power level, so that the balance between the different signal sounds in the recording is not disturbed.

THE CENTER LOUDSPEAKER

We consider the three-channel approach first. 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. For example, as proposed by Klipsch [6], an additional center loudspeaker C can be fed with the sum signal $\sqrt{2}(x_L + x_R)/2$, where x_L and x_R represent signals from left and right, respectively. The $\sqrt{2}$ factor was introduced to preserve the total energy from the three loudspeakers, assuming incoherent addition for left, center, and right sounds recorded by two widely spaced microphones. A major drawback of this approach is that crosstalk with left and right channels is inevitable, and therefore it will narrow the stereo image considerably.

We propose an algorithm to derive the center channel without these drawbacks using Principal Component Analysis (PCA) [11, 12] which produces two vectors indicating the direction of both dominant signal y and remaining signal q as shown in Fig. 2 by dashed lines. Note that these two directions are perpendicular to each other, creating a new coordinate system. These two signals are then used as basis signals in the matrix decoding, a point that is different from other existing two-to-five sound systems.

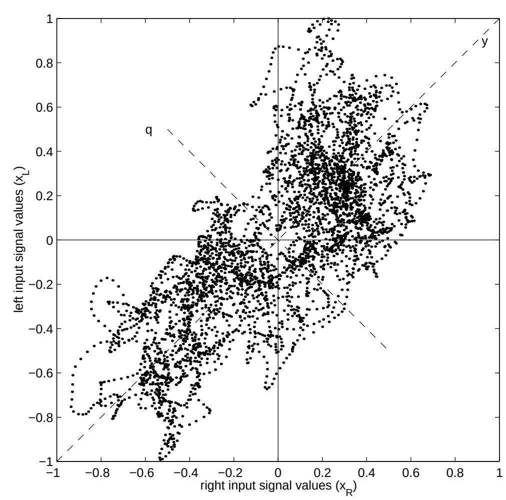

Fig. 2: A Lissajous plot of a stereo signal recorded from the fragment "The great pretender" by Freddy Mercury. Dashed lines represent new coordinate system based on both the dominant signal y and remaining signal q forming a direction of a stereo image α.

To derive the center channel's gain using the direction of a stereo image, we process the audio signal coming from a CD (sampling frequency $F_s = 44.1$ kHz) on a sample basis. Each sample of a stereo pair at a time index k can be expressed as

$$\mathbf{x}(k) = [x_L(k) \;\; x_R(k)]^T, \qquad \text{integer } k. \tag{1}$$

Let us now define $y(k)$ to be a linear combination of the input signals:

$$y(k) = \mathbf{w}^T(k)\mathbf{x}(k), \tag{2}$$

where

$$\mathbf{w}(k) = [w_L(k) \; w_R(k)]^T \tag{3}$$

is a weight vector corresponding to the left and the right channel, respectively.

In order to find the optimum weighting vectors, we maximise the energy of Eq. (2) with respect to $\mathbf{w}$, that is,

$$\frac{\partial E[y(k)^2]}{\partial \mathbf{w}} = 0, \tag{4}$$

where E denotes the expected value. Using a method presented by Haykin [12], and see [1] for more details we get

$$\begin{aligned} w_L(k) &= w_L(k-1) + \mu y(k-1)[x_L(k-1) \\ &- w_L(k-1)y(k-1)], \\ w_R(k) &= w_R(k-1) + \mu y(k-1)[x_R(k-1) \\ &- w_R(k-1)y(k-1)]. \end{aligned} \tag{5}$$

The direction of a stereo image in terms of angle in radians can easily be computed as

$$\alpha(k) = \arctan\left(\frac{w_L(k)}{w_R(k)}\right). \tag{6}$$

Figure 3 shows the values of α when it is calculated for a CD stereo music recording. Recalling Fig. 2 when left channel corresponding to $\alpha = \pi/2$, and right channel to $\alpha = 0$, we can see that α fluctuates around $\pi/4$ creating a phantom source almost equidistant between left and right channels.

Figure 4 shows the same responds of the angle α, but now measured from a DVD movie fragment where abrupt changes from one channel to the other are present. We intentionally take a shorter fragment in order to demonstrate that the algorithm is still able to detect the abrupt changes in localisations within a short period of time.

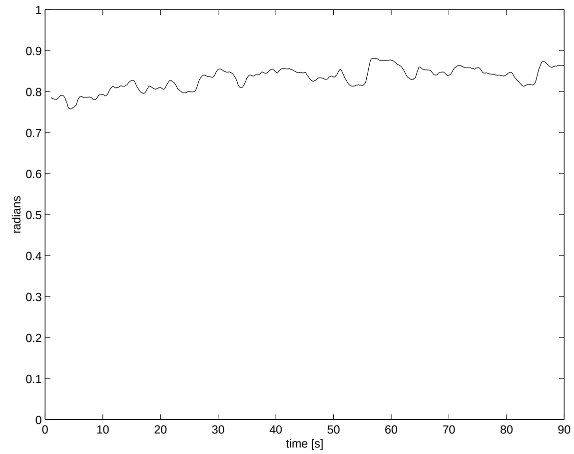

Fig. 3: Typical example of fluctuation of α, computed from a CD stereo music fragment with a stable phantom source.

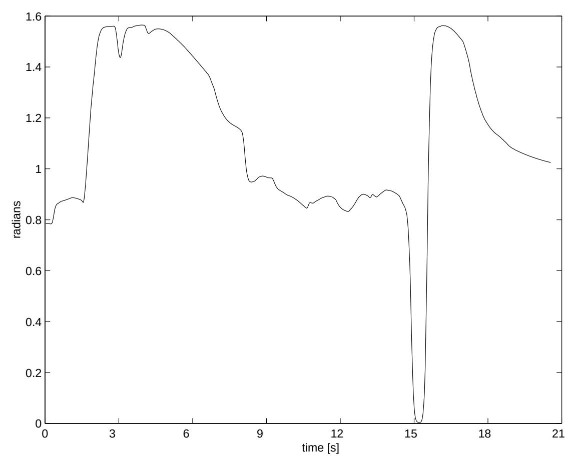

Fig. 4: Fluctuation of the direction, α, computed from a DVD fragment containing sounds of a passing car with a high speed from one channel to the other. Total duration of fragment about 20 seconds.

Now we can represent a pair of stereo signals using a vector given by Eq. (3). This is a vector of unit length having the right channel gain in the horizontal axis, and the left channel gain in the vertical axis, as shown in Fig. 5a. To map this stereo vector onto

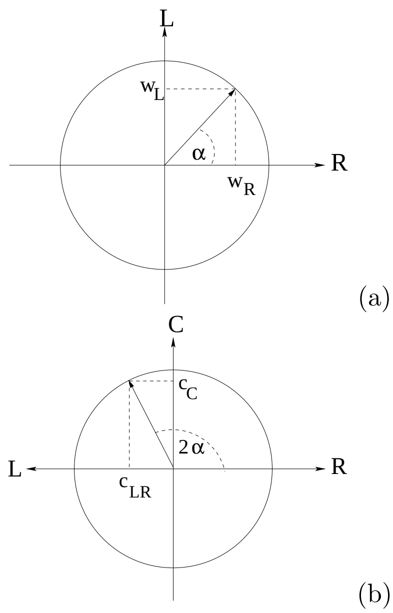

Fig. 5: (a) Direction vector plots of stereo signals. (b) Corresponding three-channel representation by doubling the angle α.

a three-channel vector, we double the angle α producing a new mapping as depicted in Fig. 5b. We can then find the projections of the vector onto the LR-axis, and C-axis using sine and cosine rules as:

$$\begin{aligned} c_{LR} &= w_R^2 - w_L^2 \\ c_C &= 2 w_L w_R. \end{aligned} \tag{7}$$

It should be pointed out that the transformation illustrated in Fig. 5 works only for nonnegative 2α. This is because for negative 2α, a multiplication by a factor two results the vector to be in a lower quadrant, and therefore no gain can be derived for the center channel. To overcome this problem, extra information should be used which is described in the next section.

THE SURROUND LOUDSPEAKERS

The surround channels are generally used to create ambience effects for music, while for applications in the film industry the surround channels are used for sound effects. A common technique for ambience reconstruction is the use of delayed front channel information for the surround channels. Dolby Pro Logic for instance, has delayed the surround sounds so as to arrive at the listeners' ears at least 10 ms later than the front sounds [8].

Environmental and ambience effects can be computed by considering left and right channel variation ($x_L - x_R$) in the original signals. This variation is usually referred to as the *anti-phase* components, the amount of which can be represented by the remaining signal q (Fig. 2). However, it can be expected that when the amount of the dominant signal equals or almost equals to that of the remaining signal, an ambiguity appears since there is no way to determine the direction vector uniquely. In this situation the distribution in Fig. 2 is no longer an ellipse but has a circle-like form ($|y| \approx |q|$), as illustrated in Fig. 6, causing α to be not well defined.

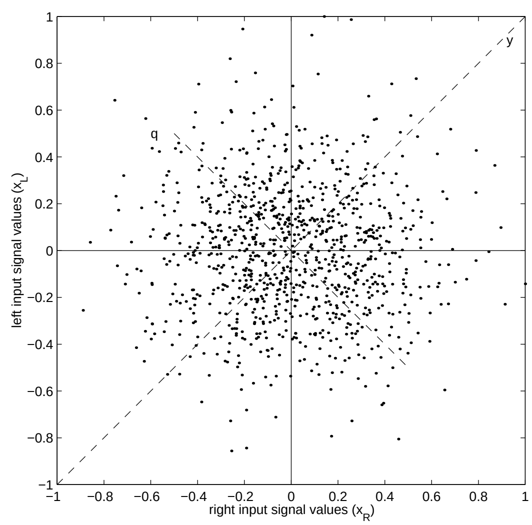

Fig. 6: A Lissajous plot of the first 23-second stereo signal recorded from the fragment "Holiday" by Madonna, where the amount of a dominant signal is almost equal to that of a remaining signal, forming a circle-like distribution.

Obviously extra information is necessary when dealing with this sort of ambiguity. In this paper we propose to use a known technique to measure the amount of anti-phase components, namely the correlation coefficient which is given in any text book on statistics as

$$\rho = \frac{\sum(\mathbf{x}_L - \bar{\mathbf{x}}_L)(\mathbf{x}_R - \bar{\mathbf{x}}_R)}{\sqrt{\sum(\mathbf{x}_L - \bar{\mathbf{x}}_L)^2 \sum(\mathbf{x}_R - \bar{\mathbf{x}}_R)^2}} \quad (8)$$

where $\bar{\mathbf{x}}_L$, and $\bar{\mathbf{x}}_R$ are the mean values of $\mathbf{x}_L$ and $\mathbf{x}_R$, respectively.

In [13] it has been shown that Eq. (8) can be computed recursively by using only a few arithmetic operations.

Three-dimensional mapping

To avoid ambiguity when the amount of the dominant signal approaches that of the remaining signal, the use of both the direction of the stereo image and the correlation coefficient is necessary. The latter is included in the mapping (Fig. 5) by, for example, placing the surround channels in the vertical plane, as shown in Fig. 7.

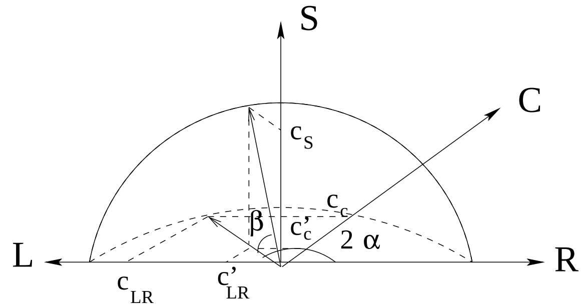

Fig. 7: Three-dimensional mapping showing front (horizontal plane) and surround channels (vertical plane). Parameter β determines the level of surround information with respect to the front channel sounds.

The angle β can be defined to represent the actual surround information by means of the adaptive correlation coefficient, for example, by using the following expression

$$\beta(k) = \arcsin(1 - \rho_0(k)), \quad (9)$$

and hence

$$0 \leq \beta(k) \leq \frac{\pi}{2}. \quad (10)$$

Thus as the amount of remaining signal increases (input signals become weakly correlated), the angle β also increases, which reduces the total distribution to the front channels. On the other hand, when the input signals are strongly correlated (quasi mono), β approaches zero, producing a larger contribution to the front channels. This principle satisfies the energy preservation criterion, which is discussed in more detail in the following section.

Since the direction vector on the horizontal plane is now lifted with an angle β, recalculation for the projections is necessary. Using straightforward trigonometry and keeping in mind that the vector is of unit length we obtain:

$$\begin{aligned} c'_{LR} &= c_{LR}\ \cos\beta \\ c'_C &= c_C\ \cos\beta \\ c_S &= \sin\beta. \end{aligned} \quad (11)$$

Matrixing

The system as described so far reproduces four channel signals as L, C, R and S from two input signals. Therefore, we have a 4×2 reproduction matrix.

We now discuss the objective requirement on the energy preservation as emphasised in the previous section. A matrix preserves energy if and only if its columns are of unit length, and the columns are pairwise orthogonal. Since the product of any two orthogonal matrices is also orthogonal, back and forward compatibility between stereo and multichannel can also be achieved.

Following the above energy criterion, we design the matrix as follows:

$$\begin{bmatrix} u_L(k) \\ u_R(k) \\ u_C(k) \\ u_S(k) \end{bmatrix} = \begin{bmatrix} c_L(k) & g\ w_L(k) \\ c_R(k) & g\ w_R(k) \\ c_C(k) & 0 \\ 0 & c_S(k) \end{bmatrix} \begin{bmatrix} y(k) \\ q(k) \end{bmatrix}. \quad (12)$$

The components of the left-hand side of Eq. 12 denote the signals for Left, Right, and Center loudspeakers, and u_S the mono surround signal. The basis signals are obtained by rotating the coordinate system of x_L and x_R:

$$\begin{aligned} y(k) &= w_L(k)x_L(k) + w_R(k)x_R(k) \\ q(k) &= w_R(k)x_L(k) - w_L(k)x_R(k), \end{aligned} \quad (13)$$

and

$$c_L = \begin{cases} -c_{LR} & c_{LR} < 0 \\ 0 & \text{otherwise} \end{cases} \quad (14)$$

$$c_R = \begin{cases} c_{LR} & c_{LR} \geq 0 \\ 0 & \text{otherwise,} \end{cases}$$

and g is a gain to control the energy preservation.

Since c_L and c_R can only produce one value depending on the condition in Eq. (14), the length of the first column of the matrix given in Eq. (12) is equal to $c_{LR}^2 + c_C^2$, which is unity. The second column of Eq. (12) contains mainly a projection of the vector onto the horizontal plane (see Fig. 7). The length of this column is equal to: $g(w_L^2 + w_R^2) + c_S^2 = 1$. The two columns are thus of unit length and pairwise orthogonal if $g = \cos^2\beta$, and therefore the matrix preserves the total energy.

Finally, the Lauridsen [14] decorrelator is used to obtain stereo surround because of its simplicity. This decorrelator can be viewed as two FIR comb filters (h_L and h_R) with two taps each for surround left and surround right. A time delay of δ (=440 samples) $\approx$ 10 ms is used between the taps, which is determined experimentally.

The choice of the time delay δ is a subtle compromise between the amount of widening and the sound diffuseness. The greater δ is, the more diffuse the sounds will be, and at some point it will lead to confusion.

Note that there are other decorrelator filters available, for example complementary comb filters in which the "teeth" are equally distributed on a logarithmic frequency scale.

CONCLUSIONS-PART 1

A method to convert two-channel stereo to multichannel sound has been presented. A three-dimensional representation has been used to produce each channel's gain, which is time varying. PCA is proven to be a powerful tool to detect the direction of a stereo image, which is then used to derive the center channel's gain. Furthermore, a robust tracking algorithm for computing the cross correlation between left and right channel has been used to improve the sound quality of the surround channels.

PART 2: BANDWIDTH EXTENSION OF BAND-LIMITED SIGNALS IN PARTICULAR FOR REPRODUCING LOW PITCHED SIGNALS THROUGH SMALL LOUDSPEAKERS [2]

INTRODUCTION-PART 2

In many sound reproduction applications, it is not possible to use large loudspeakers, due to size and/or cost constraints. Typical applications are portable audio, multimedia, TV and public address systems, to name just a few. Hence the devices are often small in size, and therefore the transducers are inherently small as well. Needless to say, the competitive market just mentioned 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 such applications reproduction below 100 Hz is usually negligible, whereas 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.

VIRTUAL PITCH

Pitch is a subjective, psychophysical quantity. According to the American Standards Association pitch is "that attribute of an auditory sensation in terms of which sounds may be ordered on a musical scale". For a pure tone, where the fundamental frequency corresponds to the frequency of the tone, the pitch is unambiguous and— if we neglect the influence of sound level on pitch— one can identify pitch with the frequency of the pure tone. For a complex tone, consisting of more than one frequency, the situation is more complicated. Pitch should then be measured by psychophysical experiments. A pitch that is produced by a set of frequency components, see Fig. 8-b, rather than by a single sinusoid, is called a *residue*. In Fig. 8-b the fundamental frequency is missing, yet will still be perceived as a residue pitch, which in this case is also called *virtual pitch*. The psychoacoustic phenomenon responsible for this effect is the 'missing fundamental' effect. There is long history of investigations into pitch perception, also

[2] Based on papers [15–22] presenting much more detail.

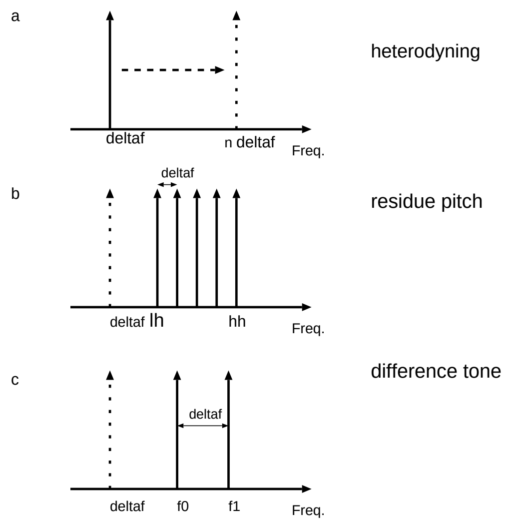

Fig. 8: Possible options for psychoacoustic bass enhancement. The dotted frequency component denotes the perceived pitch (but is not necessarily acoustically radiated). (a) Frequency doubling. (b) Residue pitch. (c) Difference tone.

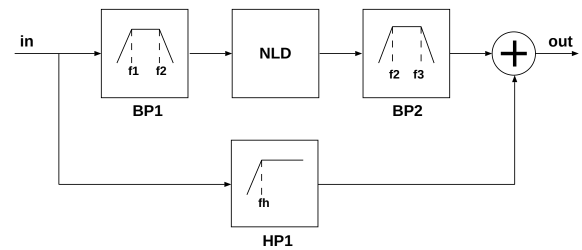

Fig. 9: Signal processing for psychoacoustic bass enhancement.The input signal is summed and filtered to obtain the bass portion. Then harmonics are created by the non-linear device (NLD) and added to left and right output signals. In the direct path a high-pass filter is implemented.

regarding virtual pitch. Famous are the experiments of Seebeck in 1843, and the controversy of him with Ohm; see Plomp [23] for a historical review. There is a vast amount of literature on this topic, just a few interesting references are [24–28]. As the frequency of a pure tone decreases to very low values, say under 100 Hz, the pitch becomes more difficult to determine. This is also true for the missing fundamental effect, and because the proposed algorithm is aimed at this very low frequency range, we need psychoacoustic data regarding the perception of virtual pitch for this range. Unfortunately, only sparse data is available. The work of Ritsma [29, 30] investigates the existence region of the tonal residue, for frequencies above 200 Hz.

PROCESSING SCHEME

Fig. 9 presents the general processing scheme that we propose for psychoacoustic bass enhancement. As the system is 'merely' based on a psychoacoustic model of pitch perception, and uses loudspeaker characteristics in a very general sense (it is only assumed that reproducing lower frequencies is less efficient than reproducing higher frequencies), the method can be employed for any kind and/or size of loudspeaker.

NON-LINEAR DEVICE - NLD

The non-linear device, or harmonics generator, 'shifts' signal components in a low frequency range to a higher frequency range. The pitch of the input signal is preserved, because the components in the higher frequency range are harmonics of the original components. The preservation of the original low pitch is due to the virtual pitch (or difference tone effect) of the harmonics signal. Because this element is a non-linear device, any single output component depends on all input components. Moreover, at the output, frequency components will be generated, which are not present at the input. This is a desired effect, since this is how the harmonics are obtained. However, it also leads to sum and difference components, which are not desired, for they are not harmonically related to the input signals.

CONCLUSIONS-PART 2

In this part of the paper we have proposed a psychoacoustically based signal processing system to enhance the perceived bass response of a loudspeaker below its cut-off frequency. The main concept of this system is to replace very low frequency components by their harmonics, through controlled non-linear processing. The resulting harmonics yield the same (virtual) pitch as the original signal, due to the missing fundamental effect. Beneficial characteristics of the system include:

- (Very) low radiation of energy below loudspeaker cut-off frequency.
- Less headroom required compared to a traditional (linear) bass boost, for a comparable bass enhancement effect.
- Computationally very efficient, simple circuit in case of analog design.
- Power efficient.
- Tuneable to any kind and size of loudspeaker.

Some drawbacks are:

- Intermodulation distortion can lead to audible artefacts. Careful tuning of filters will be beneficial.
- The added harmonics interfere with frequency components in the original audio, which in some cases may alter the timbre of the signal to some extent.

PART 3: 3D HEADPHONES BASED ON ACTIVE NOISE CANCELLATION [3]

Headphone virtualizers are systems that aim at giving the user the illusion that the sound is coming from loudspeakers rather then from the headphones themselves [35–38]. Systems that are commercially available today are not optimized for the individual listener. This results in large localization errors for most listeners [39]. The system at hand is personalized in that it requires a calibration procedure, which can be carried out conveniently by the listener. This system consists of conventional headphones into which miniature microphones have been mounted. The sound reproduction using headphones gives the same listening experience to the user as the reference (multichannel) loudspeaker system. This is achieved by taking all contributions into account: the room impulse responses, the loudspeaker characteristics, the headphone characteristics and the properties of the listener's head and torso. Besides the usual computational requirement for a headphone virtualizer, this system needs in addition two low-cost microphones and two analog to digital converters to convert the microphone signals.

[3] Based on papers [31–34] presenting much more detail.

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, torso). If loudspeaker reproduction is emulated using headphones, these sound charateristics have to be taken into account and compensation for the sound reproduction characteristics of the headphones is required.

The physical properties of the head and outer ears of the listener modify the sound as it travels from the source to the ear-drums. 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 and the inverses of the headphone to ear transfer functions prior to headphone sound reproduction. In this way the multichannel loudspeaker system can be emulated very accurately [40]. Note that only one loudspeaker driver is required at each side of the head in order to make multichannel virtual sound. Sounds add in a linear way in the air so that headphone signals of the virtual left loudspeaker can be added to those of the virtual right loudspeaker to obtain virtual stereo for example.

When audio is filtered with HRTFs that are measured from another person, there are large errors in the vertical and front/back localization [41]. Therefore the sound reproduction system should be personalized.

Configuration
Figure 10 shows a five channel loudspeaker setup. A person who is wearing headphones resides inside the sweet spot, that is, the region in which the sound reproduction of the loudspeaker setup is optimal. The headphones are equipped with integrated microphones [42] and are connected to a digital signal processing unit (DSP). During the calibration, the DSP is connected to the multichannel loudspeaker setup. A noise signal is played through each of the loudspeakers consecutively and is picked up 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. The algorithm that is

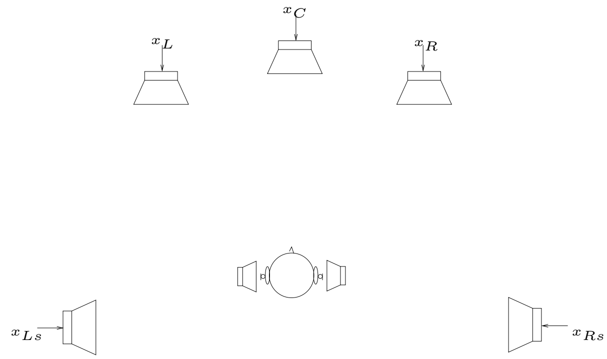

Fig. 10: A 5 channel sound reproduction setup with headphones with integrated microphones.

used is described in the next section. When the calibration is completed, the listener can manually choose between loudspeaker or headphone sound reproduction, showing the capabilities of the system. A variant is that the calibration is carried out using only one loudspeaker. The subject needs to change his/her orientation after each measurement such that this loudspeaker corresponds to the left front, right front, left rear, right rear and center loudspeaker position.

ACTIVE NOISE CANCELLATION

The algorithm that is used during the calibration is essentially an active noise cancellation algorithm. An introduction to active noise cancellation is given next, a noise cancellation primer can be found in [43]. Its application to headphone listening will be explained below.

In sound reproduction systems, sound signals can be filtered prior to reproduction by loudspeakers to ideally obtain perfect sound reproduction at a finite number of positions in space. The filters can be found by first placing microphones at the relevant positions and using the difference between the ideal sound and the reproduced sounds at these positions as errors signals for an adaptive algorithm. In classic adaptive-filter theory these error signals are obtained by comparing the desired signals with the adaptive filter outputs. In active noise cancellation the error signals are obtained by comparing the desired signals with adaptive filter outputs that are filtered by acoustic transfer functions. The classical adaptive filter is depicted in Fig. 11 (*top*) where it is used to equalize the acoustic transfer function from the loudspeaker to the microphone $H(z)$. Here, the update uses the input signal of the adaptive filter $W(z)$ and the difference between the reference signal $d[n]$ which resembles $x[n]$ and the adaptive-filter output. The reference signal $d[n]$ can be a delayed version of $x[n]$ for example, so that $W(z)$ can converge to a stable solution with $W(z)H(z)$ equal to this delay.

Instead of equalizing a signal $x[n]$ that is filtered by an acoustic transfer function $H(z)$, sound reproduction systems need to filter this signal prior to playback as depicted in Fig. 11 (*bottom*). Both systems are equivalent if the adaptive filter $W(z)$ is constant. In practical applications it suffices to demand that the adaptive filter is slowly varying. The latter system is termed the filtered-x algorithm in [44] which indicates that a filtered version of the signal $x[n]$ is used in the update. This filter $\tilde{H}(z)$, which corresponds to the acoustic transfer function $H(z)$, is not exactly known in practical situations. An estimate of it can be used however, and the filtered-x algorithm is known to be robust to estimation errors herein. In [45] it was found that superior performance is obtained by replacing $\tilde{H}(z)$ by an all-pass filter with the same phase. Application areas include noise cancellation in ducts and phantom sound source generation [43].

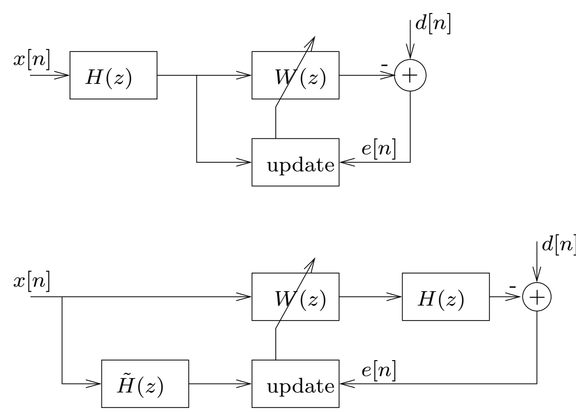

Fig. 11: Conventional adaptive filter (*top*) and filtered-x equivalence (*bottom*).

Phantom sound source generation using noise

cancellation
In Fig. 12 the filtered-x algorithm is applied in an HRTF measurement system. The system is shown for only one side of the headphones. The processing for the other side is identical and works independently. A noise signal is fed to the loudspeaker. This noise signal is also filtered before it is fed to the headphones. The filtering is done in such a way that the microphone signal is minimized. In this way the adaptive filter $h_{L,L}$ will become approximately equal to the transfer function of the loudspeaker to the microphone $g_{L,L}$ times the inverse of the transfer function of the headphone to the microphone gx_L

$$h_{L,L} = -g_{L,L}(gx_L)^{-1}. \quad (15)$$

A noise suppression of about 20 dB can be achieved at the position of the microphone. This means that the sound of the headphones is almost identical to that of the loudspeaker. When the loudspeaker is switched off the listener will have the illusion to listen to the loudspeaker[4], while he is listening via headphones.

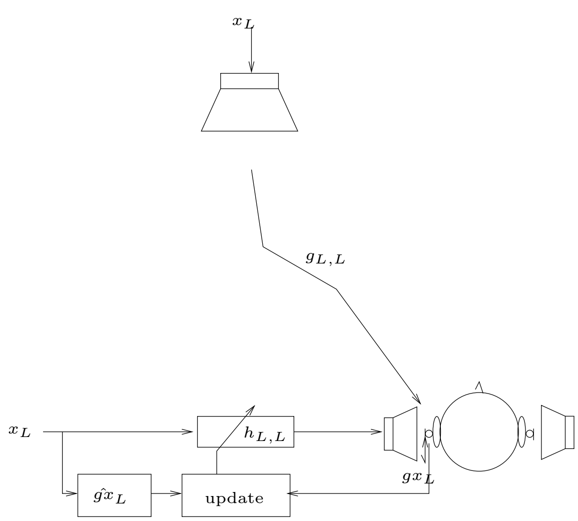

Fig. 12: HRTF measurement using noise cancellation.

CONCLUSIONS-PART 3

The performance of a system that delivers multichannel sound using headphones is analyzed. The system is calibrated using active noise cancellation techniques.

[4]Note that theoretically the headphones signals should be reversed in sign first. However, perceptually this has only a small impact.

ACKNOWLEDGMENTS

The author would like to thank his colleagues Roy Irwan, Erik Larsen and Daniël Schobben for their parts in the work presented in this paper.

REFERENCES

[1] R. Irwan and R.M. Aarts. Two-to-five channel sound processing. *J. Audio Eng. Soc.*, 50(11):914–926, November 2002.

[2] R. Irwan and Ronald M. Aarts. A method to convert stereo to multi-channel sound. *The proceedings of the AES 19th International Conference, Schloss Elmau (Kleis), Germany, June 21–24*, pages 139–143, 2001.

[3] J. Eargle, editor. *Stereophonic techniques: An anthology of reprinted articles on stereophonic techniques*. Audio Eng. Soc., 1986. ISBN 0-937803-08-1.

[4] J.M. Eargle. Multichannel stereo matrix systems: An overview. *J. Audio Eng. Soc.*, 19(7):552–559, July/August 1971.

[5] P. Scheiber. Four channels and compatibility. *J. Audio Eng. Soc.*, 19:267–279, 1971.

[6] P.W. Klipsch. Stereophonic sound with two tracks, three channels by means of a phantom circuit (2PH3). *J. Audio Eng. Soc.*, 6(2):118–123, April 1958.

[7] M. Gerzon and G. J. Barton. Ambisonic Decoders for HDTV. In *preprint 3345 AES Convention Vienna*, March 1992.

[8] R. Dressler. Dolby Pro Logic Surround decoder principles of operation. Paper on Dolby's WEB pages, 1996.

[9] G. Theile. HDTV sound systems: how many channels? In *Proceedings of the 10th AES Conference Images of Audio*, pages 147–162, September 1991.

[10] ITU-R Recommendation BS.1116, Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems, International Telecommunication Union, Geneva, Switzerland, (1994).

[11] Te-Won Lee. *Independent Component Analysis: Theory and Applications.* Kluwer Academic publisher, 1998.

[12] S. Haykin. *Neural Networks.* Prentice-Hall, N.J., 1999. Second Edition.

[13] Ronald M. Aarts, R. Irwan, and A.J.E.M. Janssen. Efficient tracking of the cross-correlation coefficient. *IEEE Trans. on Speech and Audio Proc.*, 10(6):391–402, September 2002.

[14] H. Lauridsen. Nogle forsøg med forskellige former for rumakustisk gengivelse og forslag til og forsøg med et stereofonisk system med utraditionel informationfordeling mellem kanaler. *Ingeniøren*, 47:906–910, November 1954.

[15] R.M. Aarts and S.P. Straetemans. Ultra Bass, 1996. First Filing Form PHN 15.798 EP-P, European patent: EP19970917381 19970505.

[16] R.M. Aarts and S.P. Straetemans. Circuit, audio system and method for processing signals, and a harmonics generator, Aug. 2000. US patent 6,111,960.

[17] E. Larsen and R.M. Aarts. Perceiving low pitch through small loudspeakers. *Preprint 5151 (P-2) 108th AES Convention Paris, Feb. 19–22 2000.*, 2000.

[18] G.F.M. De Poortere, C.M. Polisset, and R.M. Aarts. Ultra Bass, Oct 17. 2000. US patent 6,134,330.

[19] R.M. Aarts, J. Smeets, and P.C.W. Sommen. Bandwidth extension of narrow-band speech, Proceedings of 2nd IEEE Benelux Signal Processing Symposium (SPS-2000), Hilvarenbeek, March 23–24, 2000.

[20] R.M. Aarts. Audio system, September 2002. US patent US 6,456,718.

[21] E. Larsen and R.M. Aarts. Reproducing low-pitched signals through small loudspeakers. *J. Audio Eng. Soc.*, 50(3):147–164, March 2002.

[22] R.M. Aarts, E. Larsen, and D. Schobben. Improving perceived bass and reconstruction of high frequencies for band limited signals. *Proceedings first IEEE Benelux workshop on MPCA-20 Leuven, Belgium, Nov. 15, 2002*, pages 59–71, 2002.

[23] R. Plomp. *Experiments on tone perception.* PhD thesis, University Utrecht, 1966.

[24] F.A. Bilsen and R.J. Ritsma. Some parameters influencing the perceptibility of pitch. *J. Acoust. Soc. Am.*, 47(2 (Part 2)):469–475, February 1970.

[25] E. de Boer. *On the "residue" in hearing.* PhD thesis, University of Amsterdam, 1956.

[26] A.J.M. Houtsma and J.L. Goldstein. The central origin of the pitch of complex tones: Evidence from musical interval recognition. *J. Acoust. Soc. Am.*, 51(2 (Part 2)):520–529, 1972.

[27] J.F. Schouten. The perception of pitch. *Philips Technical Review*, 5(10):286, October 1940.

[28] J.F. Schouten, R.J. Ritsma, and B. Lopes Cardozo. Pitch of the residue. *J. Acoust. Soc. Am.*, 34(8 Part-2):1418–1424, September 1962.

[29] R.J. Ritsma. Existence region of the tonal residue I. *J. Acoust. Soc. Am.*, 34(9):1224–1229, September 1962.

[30] R.J. Ritsma. Existence region of the tonal residue II. *J. Acoust. Soc. Am.*, 35(8):1241–1245, August 1963.

[31] D. Schobben and R.M. Aarts. 3D headphones based on active noise cancellation. *Convention Paper 5713 Presented at the AES 113th Convention 2002 Oct. 5–8, Los Angeles, CA, USA*, 2002.

[32] R.M. Aarts, A.W.M. Mathijssen, P.C.W. Sommen, and J. Garas. Efficient Block Frequency Domain Filtered-x applied to Phantom Sound Source Generation. In *preprint 4650 of the 104th AES Convention Amsterdam*, May 1998.

[33] P.C.W. Sommen, R.M. Aarts, A.W.M. Mathijssen, and J. Garas. Efficient realizations of phantom sources. *ProRISC/IEEE workshop on Circuits, Systems and Signal Processing, Mierlo, the Netherlands, November 27–28*, 1997.

[34] H. He. Real-time generation of phantom sound sources in large frequency range. Master's thesis, Eindhoven University of Technology, July 1996. Master thesis with R.M. Aarts and P. Sommen.

[35] J. Blauert. *Spatial hearing: The Psychophysics of Human Sound Localization*. The MIT Press, 1983.

[36] H. Møller. Fundamentals of binaural technology. *Applied Acoustics*, 36(3–4):171–218, 1992. Special issue on auditory environment and telepresence.

[37] F.L. Wightman and D. Kistler. Headphone simulation of free-field listening. I: Stimulus synthesis. *J. Acoust. Soc. Am.*, 85(2):858–867, February 1989.

[38] F.L. Wightman and D. Kistler. Headphone simulation of free-field listening. II: Psychophysical validation. *J. Acoust. Soc. Am.*, 85(2):868–878, February 1989.

[39] H. Møller et al. Do we need individual recordings? *J. Audio Eng. Soc.*, 44(6):451–469, June 1996.

[40] A.W. Bronkhorst. Localization of real and virtual sound sources. *J. Acoust. Soc. Am.*, 98(5):2542–2553, November 1995.

[41] M. Morimoto and Y. Ando. On the simulation of sound localization. *J. Acoust. Soc. Japan (E)*, 1(3):167–174, 1980.

[42] R.M. Aarts. Headphones with integrated microphones, 2000. European patent EP1201101.

[43] S.D. Snyder. *Active noise control primer*. Springer-Verlag, 2000.

[44] P.A. Nelson and S.J. Elliott. *Active control of sound*. Academic Press, 1992.

[45] J. Garas and P.C.W. Sommen. The all-pass filtered-x algorithm. In *Proc. EUSIPCO, Rhodos Greece, pp. 101-104*, September 1998.