Phantom sources applied to stereo-base widening using long frequency domain adaptive filters

Ronald M. Aarts, Haiyan He, Piet C.W. Sommen

Abstract— **The problem of realising a virtual enlargement of the aperture from the real loudspeakers positions to the phantom source loudspeaker position, is discussed. The rationale of the processing is that the same pressure is obtained in both ears when the real loudspeakers are playing as in the situation when the phantom loudspeakers are playing. If the pressure at the eardrums generated by real loudspeakers is the same as that which would be generated by the phantom sources, then the subject will perceive no difference, and a virtual enlargement of the aperture is realized. Due to the small wave length at high frequencies of the audio range even small perturbations can result in large deviations of the very long acoustic impulse response. For this reason long adaptive filters are used which are implemented very efficiently in the frequency domain. In contrast to fixed filter approaches, the present method is adaptive and very efficient, while above that the phantom sources can be generated over a large frequency range.**

Keywords— **Active Noise Cancellation, Frequency Domain Adaptive Filters, Sound Reproduction, Loudspeakers**

I. Introduction

The sound received by a listener contains both directional and distance information, from which the listener can locate the positions of the sound sources. Experiences show that when the recorded stereo signals are reproduced by two separate loudspeakers the perceived sound image would be between the two loudspeakers, if the listener is in the center of the listening area. However, the distance between the loudspeakers has to be limited in order to obtain a compact arrangement for some applications like TV and monitors. To give the loudspeaker base a virtual broadening, we made use of the phantom source concept. A phantom sound source is a sound source which is perceived by the listener as a source at a different position than that of the real sound source. The basic idea behind phantom source generation is to alter the loudspeakers' input signals in such a way that the pressure at the listener's eardrums generated by the real loudspeakers is the same as which would be generated by the phantom sources. Fig. 1 shows a

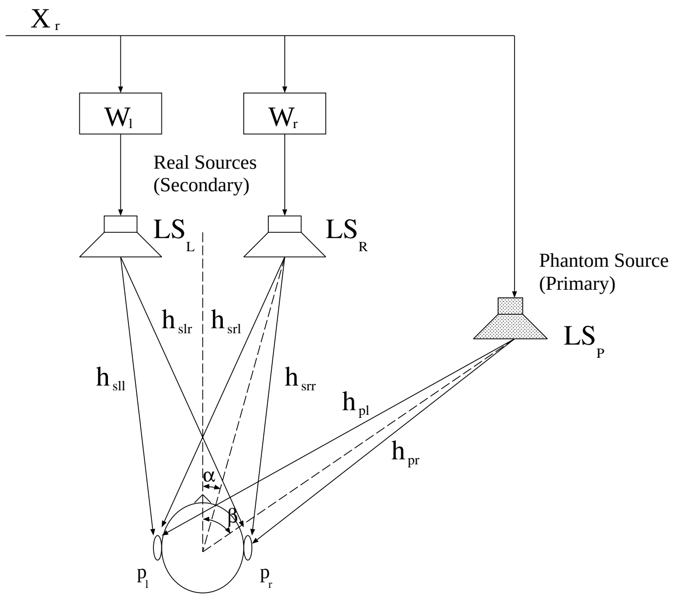

Fig. 1. Set-up for generating phantom source.

situation where a phantom sound source LS_p at angle β is generated by two real loudspeakers at angle α. W_L and W_R are filters which are used to alter the input signals to LS_L and LS_R. The design of the filters W_L and W_R is the essential part of generating a phantom source. These filters can be made fixed, however due to the small wavelength at high frequencies of the audio range, even small perturbations can result in large deviations of the very long acoustic impulse response. For this reason long adaptive filters are used which are implemented very efficiently in the frequency domain. These filters are applied as follows: Initially three loudspeakers are used, one un-filtered at the position of the phantom source, the other two filtered by adaptive filters. Microphones positioned near the subject's ears are used to control the filter's coefficients until the pressure measured is at a mini-

R. Aarts is with the Philips Research Laboratories WY81, 5656 AA Eindhoven, The Netherlands, E-mail: rmaarts@natlab.research.philips.com

P. Sommen and H.He are with the Eindhoven University of Technology, Building Eh 9.27, P.O.Box 513, 5600 MB Eindhoven, The Netherlands, E-mail: p.c.w.sommen@ele.tue.nl

mum level. The coefficients are then frozen and the third loudspeaker is removed. When the two loudspeakers are driven through the phase inverse of the function of the filter, only the illusion of a third source results. The method described above can hence be regarded as an active noise-cancellation solution.

II. Derivation of the model

The update of the adaptive weights is performed by an algorithm that is based on the Normalized Least Mean Square (NLMS) algorithm. For this reason we will first shortly describe this algorithm. In Fig. 2 we have depicted a general estimation setup in which the unknown system is given by the box '$\underline{\mathbf{w}}_u$'. Parallel to this system an adaptive filter is placed with N weights $w_{N-1}[k], \cdots, w_0[k]$. The purpose of the adaptive algorithm is that the weights converge to the unknown impulse response. The NLMS algorithm reaches this goal by minimizing the mean squared error. The up-

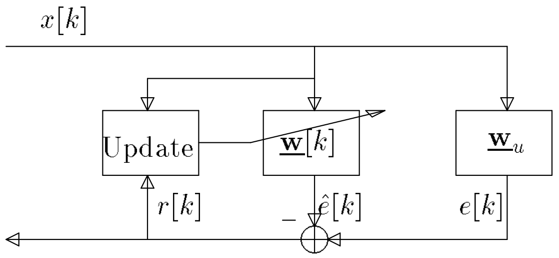

Fig. 2. NLMS update algorithm.

date is given by the following equation:

$$\underline{\mathbf{w}}[k+1] = \underline{\mathbf{w}}[k] + \frac{2\alpha}{\sigma_x^2}\underline{\mathbf{x}}[k]r[k]. \tag{1}$$

In this equation N input signal samples are contained in the vector $\underline{\mathbf{x}}[k] = (x[k-N+1], \cdots, x[k])^t$ while the N adaptive weights are in the adaptive weight vector $\underline{\mathbf{w}}[k] = (w_{N-1}[k], \cdots, w_0[k])^t$. Furthermore the adaptation process is controlled by the adaptation constant α and normalized by (an estimate of) the variance $\sigma_x^2 = E\{x^2[k]\}$.

Fig. 3-(a) shows a simple single point error cancellation case. Here $W(z)$ is the adaptive filter that is used to control the secondary source loudspeaker. In Fig. 3-(b) a model of this case is shown. Here the systems $H_p(z)$ and $H_s(z)$ are models of the paths from the loudspeaker terminals to the microphone of the primary and secondary source, respectively. The use of a simple adaptive algorithm as the NLMS algorithm for this case is complicated. The reason for this is that the difference $r[k]$ between the filter output and the desired signal $e[k]$ is no longer directly available to

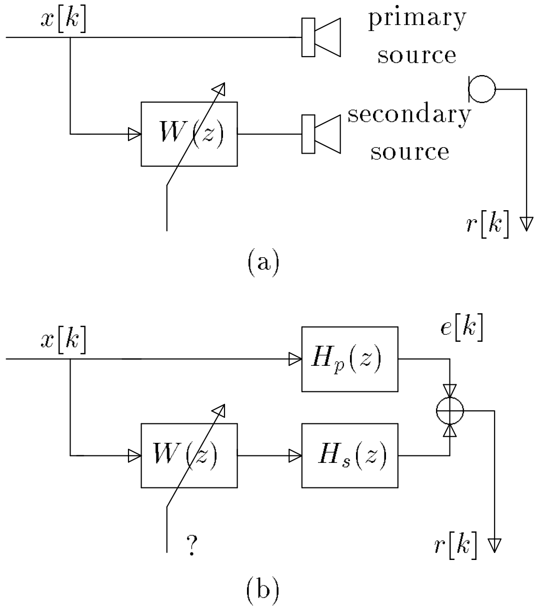

Fig. 3. Single point error cancelling.

update the coefficients. The residual signal $r[k]$ can now only be observed through the secondary source path $H_s(z)$. For the derivation of a usable model lets first look to Fig. 4-(a). Here we have a situation where the NLMS algorithm can be used. The difference with Fig. 2 is that we have preceded the adaptive filter by some unknown system with transfer function $H(z)$. In Fig. 4-(b) a system equivalent of this scheme is shown. If the adaptation of the coefficients of the adaptive filter $W(z)$ takes place on a very low time-scale, which implies slow adaptation, it can be assumed that this system is approximately linear and time-invariant. Under this condition $W(z)$ and $H(z)$ can be interchanged as showed in Fig. 4-(c). The now obtained update algorithm is the so-called filtered-x algorithm [1]. In our real-time application this approach is used and the result is as depicted in the setup of Fig. 5. In the present setup we have first made an estimate $\hat{H}_s(z)$ of the secondary system $H_s(z)$. This is done in a separate way by using the NLMS algorithm. For this measurement a white noise input signal is used while the filter $W(z)$ was not present. The unknown system $H_s(z)$ is estimated by the adaptive filter, which results after convergence in the estimate $\hat{H}_s(z)$. This (fixed) estimate is used in our system model as given in Fig. 5. Finally our goal was to be able to create an efficient realization of a phantom source over a wide frequency range. For this reason a long adaptive filter $W(z)$ is needed from which the real-time realization will be implemented by using a Block Frequency Domain Adaptive Filter (BFDAF).

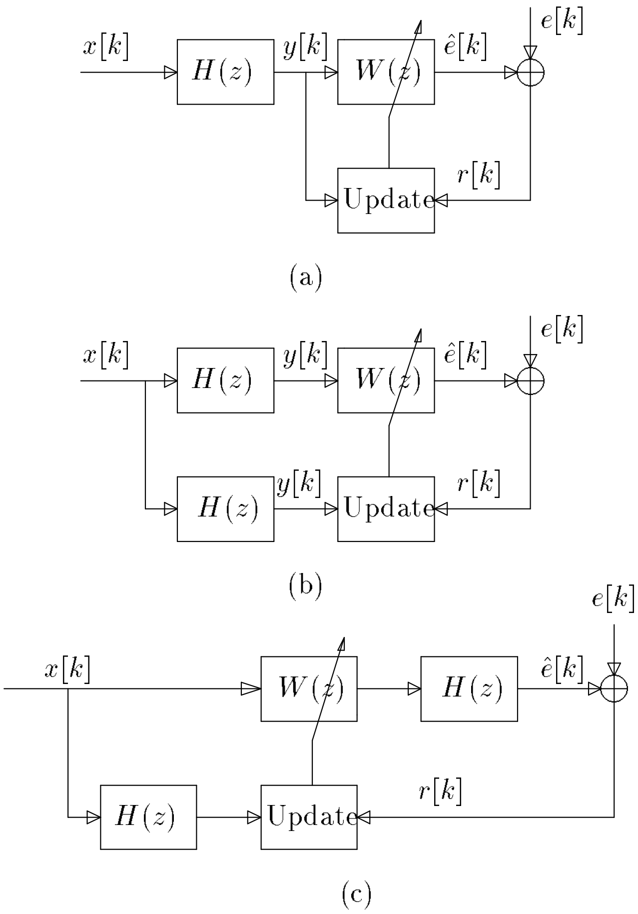

Fig. 4. Development of filtered-x algorithm.

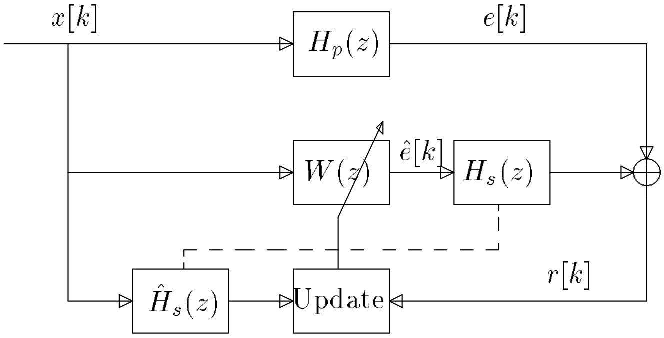

Fig. 5. System model.

III. Block Frequency Domain Adaptive Filter

In this section we will shortly derive the Block Frequency Domain Adaptive Filter (BFDAF) using the overlap and save method. In this case the update of the adaptive weight vector

$$\underline{\mathbf{w}}[kB] = (w_{N-1}[kB], \cdots, w_1[kB], w_0[kB])^t, \quad (2)$$

is performed once every B samples, with the block length $B \geq 1$. The $N \times B$ input signal matrix $\mathcal{X}[kB]$ is defined as:

$$\mathcal{X}[kB] = \begin{pmatrix} x[d] & \cdots & x[b] \\ \vdots & \vdots & \vdots \\ x[c] & \cdots & x[a] \end{pmatrix} \quad (3)$$

with $a = kB$, $b = kB - N + 1$, $c = kB - B + 1$ and $d = kB - B + 1 - N + 1$. Furthermore we define the residual signal vector $\underline{\mathbf{r}}[kB]$ as:

$$\underline{\mathbf{r}}[kB] = (r[kB - B + 1], \cdots, r[kB])^t. \quad (4)$$

First, we go from NLMS to Block NLMS (BNLMS) which results in the following update equation:

$$\underline{\mathbf{w}}[(k+1)B] = \underline{\mathbf{w}}[kB] + \frac{2\alpha}{\sigma_x^2}\mathcal{X}[kB]\underline{\mathbf{r}}[kB]. \quad (5)$$

The update of all weight coefficients is performed once every B samples and thus the output of the adaptive filter can also be expressed on block bases. Every B samples a B dimensional vector $\underline{\hat{\mathbf{e}}}[kB] = (\hat{e}[kB - B + 1], \cdots, \hat{e}[kB])^t$ is produced as follows:

$$\underline{\hat{\mathbf{e}}}[kB] = (\mathcal{X}[kB])^t\underline{\mathbf{w}}[kB]. \quad (6)$$

¿From Eqs 5 and 6 it follows that the two main operations that have to be implemented in an adaptive algorithm are: a (linear) convolution to perform the filtering of the input signal with the adaptive weights (see Eq. 6) and a (linear) correlation to calculate an estimate of the gradient that is needed for the update of the adaptive weights (see the matrix vector product $\mathcal{X}[kB]\underline{\mathbf{r}}[kB]$ in Eq. 5). For large filter lengths N these operations can be carried out very efficiently in frequency domain by using Fast Fourier Transforms (FFTs) for the transformation between time– and frequency–domain [2]. The length M of the FFT's is $M \geq N + B - 1$. An implementation of the resulting BFDAF algorithm is depicted in Fig. 6. For the convolution each step B new consecutive input signal samples $x[k]$ are collected ("S/P"). These collected samples are split into overlapping length M vectors $\underline{\mathbf{x}}[kB]$ ("overlap"). The overlap is at least $N-1$ samples. This vector is transformed to frequency domain ("F_M"). Furthermore the length N adaptive weight vector $\underline{\mathbf{w}}[kB]$ is first mirrored ("$\mathbf{J}$") and then augmented with $M-N$ zeroes ("$\underline{\mathbf{0}}$") to a vector of length M. The resulting length M vector is transformed to frequency domain $\underline{\mathbf{W}}[kB]$. The matrix $\mathbf{J}$ contains only zeroes except for the antidiagonal which contains ones. Multiplying a vector with this matrix will turn over the ordering of the elements in this vector. The

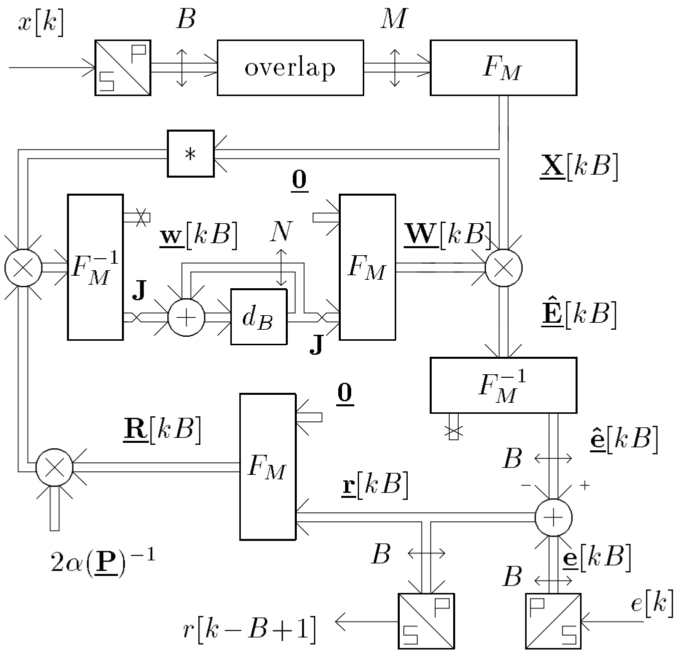

Fig. 6. Overlap save BFDAF algorithm.

(cyclic) convolution is carried out in frequency domain by a multiplication of the two vectors $\underline{\mathbf{X}}[kB]$ and $\underline{\mathbf{W}}[kB]$ as follows:

$$\underline{\hat{\mathbf{E}}}[kB] = \underline{\mathbf{X}}[kB] \otimes \underline{\mathbf{W}}[kB]. \qquad (7)$$

In this equation the symbol $\otimes$ is used to denote elementwise multiplication.

The result $\underline{\hat{\mathbf{E}}}[kB]$ is transformed back to time domain by an inverse FFT ("F_M^{-1}"). Only the last B out of M samples from this cyclic convolution represent the desired linear convolution result. Thus $M-N$ samples have to be discarded, resulting in the length B vector $\underline{\hat{\mathbf{e}}}[kB]$. The reference signal $e[k]$ is segmented in a length B vector $\underline{\mathbf{e}}[kB]$ ("S/P"). The length B residual signal vector is given by $\underline{\mathbf{r}}[kB] = \underline{\mathbf{e}}[kB] - \underline{\hat{\mathbf{e}}}[kB]$. The return to the original sampling rate is carried out by desegmenting this residual signal vector ("P/S"). This results in the delayed residual signal $r[k-B+1]$. Furthermore the gradient estimate (correlation) $\mathcal{X}[kB]\underline{\mathbf{r}}[kB]$ from the BNLMS update Eq. 5 is a correlation operation and is carried out at the left side of Fig. 6. This correlation operation is similar to the convolution operation. The difference is that the input signal needs not to be mirrored. Thus first the residual vector $\underline{\mathbf{r}}[kB]$ is augmented with $M-B$ zeroes ("$\underline{\mathbf{0}}$") to a length M vector, and this vector is transformed to frequency domain ("F_M"). The main difference between the correlation and convolution operation is reflected in frequency domain in the fact that correlation needs a complex conjugate operation ("*") of the input signal vector $\underline{\mathbf{X}}[kB]$, and the convolution does not. Finally as mentioned before the block update is performed once every B samples and this is denoted with d_B in the figure.

Notes:

• Another powerful aspect of the BFDAF algorithm is the possibility to normalize each separate frequency bin by an estimate of the power in that bin. In Fig. 6 this is done by the elementwise multiplication of the transform of the residual vector $\underline{\mathbf{R}}[kB]$ with the vector that contains the inverse power of each separate frequency bin denoted by $(\underline{\mathbf{P}})^{-1}$.

• The two mirror operations ("$\mathbf{J}$") in Fig. 6, one before and one after the updating of the adaptive weight vector, can easily be combined and left out. In order to keep the separate implementations of the convolution and correlation operations visible, this has not been done.

IV. Filtered-X in frequency domain

In this section we will describe how the model as depicted in Fig. 5 of the filtered-x algorithm can be implemented efficiently in frequency domain by using the results of the previous sections.

As mentioned before an estimate $\underline{\hat{\mathbf{h}}}_s$ of the impulse response from the secondary sound source field is available. In the present application the length of this impulse response is shorter than the length N of the adaptive filter. For this reason $\underline{\hat{\mathbf{h}}}_s$ is first augmented with zeros and then transformed to frequency domain resulting in the (fixed) length M vector $\underline{\hat{\mathbf{H}}}_s$. The filtered-x algorithm can now be implemented in frequency domain by using the setup of the BFDAF as depicted in Fig. 6. The result is given in Fig. 7. The difference with Fig. 6 is that the transformed input signal vector $\underline{\mathbf{X}}[kB]$ has first to be filtered ("filtered-x") before the correlation-part (with "*") can be carried out. The "filtered-x" is carried out in frequency domain by multiplying the input signal vector $\underline{\mathbf{X}}[kB]$ with $\underline{\hat{\mathbf{H}}}_s$. Because of the cyclic nature of the FFT this cyclic convolution result has to be transformed back to time domain ("F_M^{-1}"), from which a part has to be discarded in order to get a linear convolution result. The signal blocks have to be desegmented first in order to return to the original sampling rate ("P/S"). After this the samples have to be collected again in appropriate blocks ("S/P") and overlapped before the actual Fourier transform takes place. From this point on the scheme is the same as the BFDAF algorithm of Fig. 6. Furthermore we want to comment that, in comparison to the BFDAF algorithm, we made the

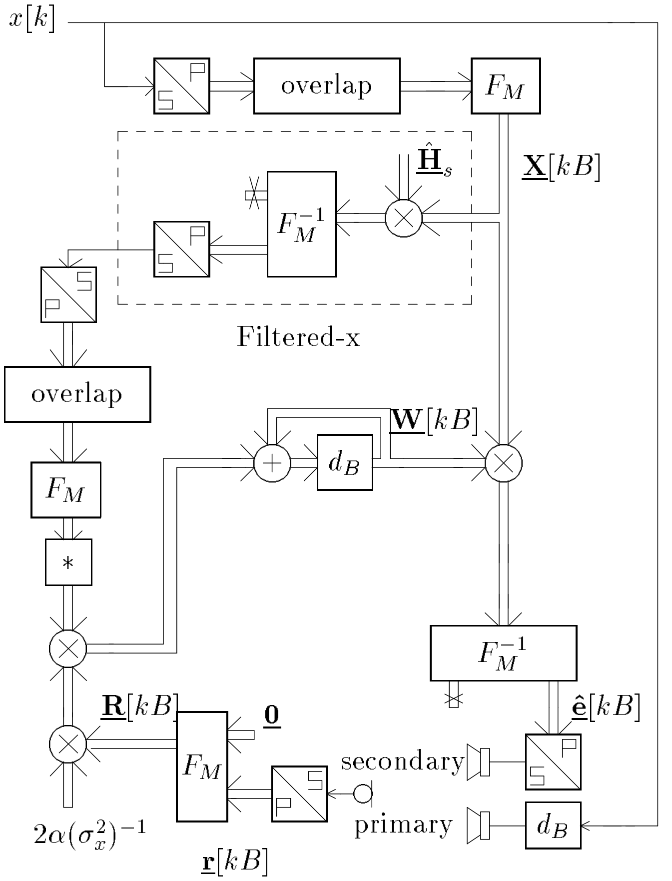

Fig. 7. Efficient implementation of filtered-x algorithm.

following simplifications:

- The normalization for each separate frequency bin is left out.
- The update of the weight vector of Fig. 6 is carried out in frequency domain. This update is now preceded by two FFT's and a window. In [3] it is shown that these blocks can be left out without distorting the convergence properties of the adaptive filter too much.

V. Phantom sources applied to stereo-base widening

In some applications the loudspeaker separation is limited in order to obtain a compact arrangement. To realize the virtual enlargement of the aperture from the real loudspeaker positions to the phantom source loudspeaker position, the real loudspeakers input signals need to be altered so that the listener can perceive the sound as coming from the phantom source loudspeaker.

The transfer function from a sound source in space to the ear is defined as the Head Related Transfer Function (HRTF). To determine the localization of a sound source, a number of cues are used. Among these cues, the interaural time, level and phase differences are playing the dominant roles in most of the positions. For distinguishing the front and back, the up and down positions, the spectral cues which are mainly contained in the high frequency range (from 4–10 kHz), have a significant effect. A human observer with two microphones at the positions of two ears participated in the experiment discussed below. The control filters have to be adapted in such a way that the signals at two positions are optimized, hence we have to use a Multiple Error (LMS) algorithm. As for the standard stereo loudspeaker setting, two real loudspeakers are used.

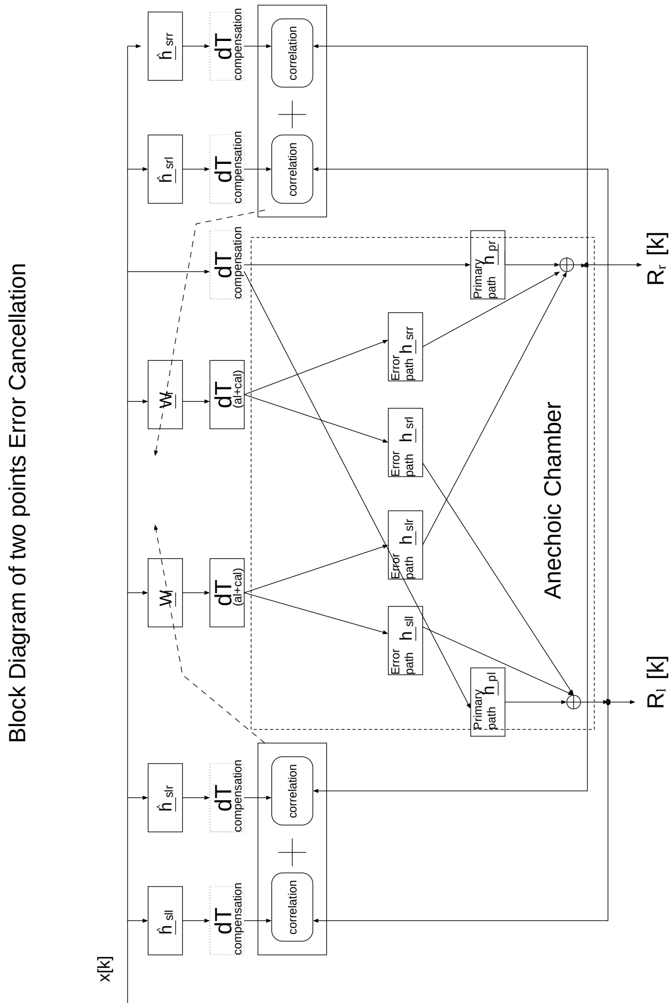

Fig. 8. The Block Diagram of two points error cancellation.

Fig. 8 shows the block diagram of two points error cancellation (cf. Fig. 1). There are four error path estimation filters ($\underline{\hat{h}}_{sll}$, $\underline{\hat{h}}_{slr}$, $\underline{\hat{h}}_{srl}$ and $\underline{\hat{h}}_{srr}$) and two control filters ($\underline{w}_l$ and $\underline{w}_r$). The filtering is performed in the frequency domain, that results in both algorithm delay and calculation delay. The delay needs to be compensated to update the coefficients correctly.

A. *Results of experiments*

The experiments of stereo-base widening are performed in an anechoic room. The results presented here are obtained under these conditions: $M = 1024$, $B = 512$ and $N = 512$. For our experiments, $\alpha = 10°$ and $\beta = 45°$ were used. Fig. 9 shows the normalized

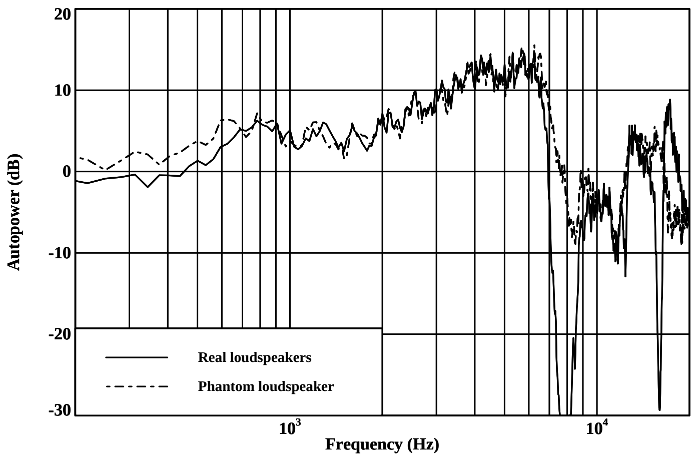

Fig. 9. The normalized HRTF (left ear) from the phantom source and the real loudspeaker.

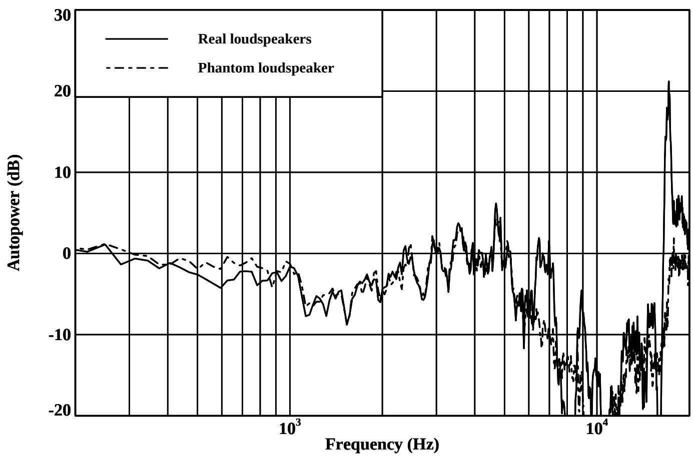

Fig. 10. The normalized HRTF (right ear) from the phantom source and the real loudspeaker.

HRTF's (for the left ear) from the phantom source and the secondary source (real) loudspeaker. Fig. 10 shows these functions for the right ear. In general, the curves of the real loudspeakers and the phantom one are very similar. Informal listening experiments showed that for all the frequency components the sound seems to come from the phantom sound source direction, when only the secondary loudspeaker is playing. ¿From additional experiments it appeared that for other angles than the above mentioned the cancellation is similar.

VI. Conclusions

A stereo-base widening system which works on a large frequency range (0–20 kHz) has been presented. The HRTFs generated by the loudspeakers located in large angles can be accurately simulated by loudspeakers located in small angles. When the listener is in the center of the listening area, the perceived sound image is coming from the phantom sound source position. For the present application, we use a two point noise cancellation.

References

[1] B. Widrow and S.D. Stearns, *Adaptive Signal Processing*, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1985.

[2] A.V. Oppenheim and R.W. Schafer, *Digital Signal Processing*, Prentice-Hall, Englewood Cliffs, N.J., 1975.

[3] P.C.W. Sommen, *Adaptive Filtering Methods*, Thesis University of Technology Eindhoven The Netherlands, June 1992, ISBN 90-9005143-0.