

(19)



Europäisches Patentamt

European Patent Office

Office européen des brevets



(11)

**EP 0 478 615 B2**

(12)

**NEW EUROPEAN PATENT SPECIFICATION**

(45) Date of publication and mention  
of the opposition decision:

**15.04.1998 Bulletin 1998/16**

(45) Mention of the grant of the patent:

**26.04.1995 Bulletin 1995/17**

(21) Application number: **90909155.5**

(22) Date of filing: **15.06.1990**

(51) Int Cl.<sup>6</sup>: **H04S 1/00, H04M 3/56,  
H04H 5/00**

(86) International application number:

**PCT/GB90/00928**

(87) International publication number:

**WO 90/16136 (27.12.1990 Gazette 1990/29)**

(54) **POLYPHONIC CODING**

POLYPHONISCHE KODIERUNG

CODAGE POLYPHONIQUE

(84) Designated Contracting States:

**AT BE CH DE DK ES FR GB IT LI LU NL SE**

(30) Priority: **15.06.1989 GB 8913758**

(43) Date of publication of application:

**08.04.1992 Bulletin 1992/15**

(73) Proprietor: **BRITISH TELECOMMUNICATIONS**

**public limited company**

**London EC1A 7AJ (GB)**

(72) Inventors:

- **HOLT, Christopher Ellis**  
**Woodbridge, Suffolk IP12 1QJ (GB)**
- **MUNDAY, Edward**  
**Ipswich, Suffolk IP4 3PA (GB)**
- **CHEETHAM, Barry Michael George**  
**Liverpool L19 9BQ (GB)**

(74) Representative: **Lloyd, Barry George William et al**

**BT Group Legal Services,**

**Intellectual Property Department,**

**8th Floor, Holborn Centre,**

**120 Holborn**

**London EC1N 2TE (GB)**

(56) References cited:

**EP-A- 0 273 567**

**US-A- 4 815 132**

- **IEEE International Conference on**  
**Communications, '86, 22-25 June 1986, Toronto,**  
**CA, Conference Record, volume 2 of 3, IEEE,**  
**(US), S. Minami: "A stereophonic voice coding**  
**method for teleconferencing", pages 1305-1309.**
- **Milcom '86, 1986 IEEE Military Communications**  
**Conference, 5-9 October 1986, Monterey,**  
**California, Conference Record, volume 3 of 3,**  
**IEEE, (US), A. Reichman et al.: "Applications of**  
**least-square lattice filters in adaptive data**  
**communication", pages 4751-4754.**

**EP 0 478 615 B2**

## Description

This invention relates to polyphonic coding techniques, particularly, but not exclusively, for coding speech signals.

It is well-known that polyphonic, specifically stereophonic, sound is more perceptually appealing than monophonic sound. Where several sound sources, say within a conference room, are to be transmitted to a second room, polyphonic sound allows a spatial reconstruction of the original sound field with an image of each sound source being perceived at an identifiable point corresponding to its position in the original conference room. This can eliminate confusion and misunderstandings during audio-conference discussions since each participant may be identified both by the sound of his voice and by his perceived position within the conference room.

Inevitably, polyphonic transmissions require an increase in transmission capacity as compared with monophonic transmissions. The conventional approach of transmitting two independent channels, thus doubling the required transmission capacity, imposes an unacceptably high cost penalty in many applications and is not possible in some cases because of the need to use existing channels with fixed transmission capacities.

In stereophonic (i.e. two-channel polyphonic) systems, two microphones (hereinafter referred to as left and right microphones), at different positions, are used to pickup sound generated within a room (for example by a person or persons speaking). The signals picked up by the microphones are in general different. Each microphone signal (referred to hereinafter as  $x_L(t)$  with Laplace transform  $X_L(s)$  and  $x_R(t)$  with Laplace transform  $X_R(s)$  respectively) may be considered to be the superposition of source signals processed by respective acoustic transfer functions. These transfer functions are strongly affected by the distances between the sound sources and each microphone and also by the acoustic properties of the room. Taking the case of a single source, e.g. a single person speaking at some fixed point within the room, the distances between the source and the left and right microphones give rise to different delays, and there will also be different degrees of attenuation. In most practical environments such as conference rooms, the signal reaching each microphone may have travelled via many reflected paths (e.g. from walls or ceilings) as well as directly, producing time spreading, frequency dependent colouration due to resonances and antiresonances, and perhaps discrete echos.

From the foregoing, in theory, the signal from one microphone may be formally related to that from the other by designating an interchannel transferfunction  $H$  say; i.e.  $X_L(s) = H(s) X_R(s)$  where  $s$  is complex frequency parameter. This statement is based on an assumption of linearity and time-invariance for the effect of room acoustics on a sound signal as it travels from its source to a microphone. However, in the absence of knowledge

as to the nature of  $H$ , this statement does no more than postulate a correlation between the two signals. Such a postulation seems inherently sensible, however, at least in the special case of a single sound source, and therefore one way of reducing the bit-rate needed to represent stereo signals should be to reduce the redundancy of one relative to the other (to reduce this correlation) prior to transmission and re-introduce it after reception.

In general,  $H(s)$  is not unique and can be signal- and time- dependent. However when the source signals are white and uncorrelated, i.e. when their autocorrelation functions are zero except at  $t=0$  and their cross-correlation functions are zero for all  $t$ ,  $H(s)$  will depend on factors not subject to rapid change, such as room acoustics and the positions of the microphones and sound sources, rather than the nature of the source signals which may be rapidly changing.

S. Minami, "A Stereophonic Voice Coding Method For Teleconferencing", IEEE Int. Conf. on Communications '86, 22-25 June 1986, Toronto, Canada, Vol. 2, IEEE, US, pp. 1305-1309 describes a system in which a transmitter sends one channel plus a coarse representation of a transfer function comprising a delay value and amplitude ratio which a receiver uses to estimate the second channel.

US patent no. 4815132 describes a stereophonic coding system which receives right- and left-hand channels. It transmits the right-hand channel but for the left-hand channel it uses a plural-order adaptive filter to generate filter coefficients (or a filter residual) which are transmitted instead. The receiver uses this information to control a filter which filters the right-hand channel to generate a reconstructed left-hand channel.

To realise such a system in physical form, the fundamental problems of causality and stability must be overcome. Consider for a moment a single source signal which is delayed by  $d_L$  seconds before reaching the left microphone and by  $d_R$  seconds before reaching the right microphone (although the point to be made has more general implications). If the source is near to, say, the left microphone, then  $d_L$  will be smaller than  $d_R$ . The interchannel transfer function  $H(s)$  must delay  $x_L(t)$  by the difference between the two delays,  $d_R - d_L$  to produce the right channel  $x_R(t)$ . Since  $d_R - d_L$  is positive,  $H(s)$  will be causal. If the signal source is now moved closer to the right microphone than to the left,  $d_R - d_L$  becomes negative and  $H(s)$  becomes non-causal; in other words, there is no causal relationship between the right channel and the left channel, but rather the reverse so the right channel can no longer be predicted from the left channel, since a given event occurs first in the right channel. It will therefore be realised that a simple system in which one fixed channel is always transmitted and the other is reconstructed from it is impossible to realise in a direct sense.

According to a first aspect of the invention, there is provided a polyphonic signal coding apparatus comprising:

- means for receiving a first and at least one second channel;
- means for periodically generating reconstruction data enabling the formation, from the first channel, of an estimate of the second channel, the generating means being operable to generate a plurality of filter coefficients which, if applied to a plural order predictor filter, would enable the prediction of the second channel from the first channel thus filtered;
- means for outputting data representing the said first channel and the reconstruction data;

wherein the apparatus further comprises means for filtering the first and second channel in accordance with a filter approximating the spectral inverse of the first channel to produce respective filtered channels, the first said filtered channel thereby being substantially spectrally whitened, and the generating means being connected to receive the filtered channels.

In a first embodiment, the reconstructing data are filter coefficients. In a second embodiment, the residual signal representing the difference between (for example) a difference signal and a sum signal when thus filtered is formed at the transmitter, and this is transmitted as the reconstruction data. In this embodiment, the prediction residual signal may be efficiently encoded to allow a backward adaptation technique to be used at the decoder for deriving the prediction filter coefficients. The residual is also used as an error signal which is added to the prediction filter's output at the decoder to correct for inaccuracies in the prediction of the second channel from the first. This "residual only" embodiment is also useful where the left channel, say, is predicted from the right channel (without forming sum and difference signals) - provided suitable measures are taken to ensure causality - to give high quality polyphonic reproduction. In a third embodiment, both are transmitted.

Preferably, the means for generating the filter coefficients is an adaptive filter, advantageously a lattice filter. This type of filter also gives advantages in non-sum and difference polyphonic systems.

In preferred embodiments, variable delay means are disposed in at least one of the input signal paths, and controlled to time align the two signals prior to forming the sum and difference signals so that causal prediction filters of reasonable order can be used.

This aspect of the invention has several important advantages:

- (i) The 'sum signal' is fully compatible with monophonic encoding and is unaffected by the polyphonic coding except for the introduction of an imperceptible delay. In the event of loss of stereo, monophonic back-up is thus available.
- (ii) The sum signal may be transmitted by conventional low bit-rate coding techniques (eg. LPC) without modification.
- (iii) The encoding technique for the difference sig-

nals can be varied to suit the application and the available transmission capacity between the above three embodiments. The type of residual signal and prediction coefficients can also be selected in various different ways, while still conforming to the basic encoding principle.

(iv) Overall, the apparatus encodes polyphonic signals with only a modest increase in bit-rate requirement as compared with monophonic transmission.

(v) The encoding is digital and hence the performance of the apparatus will be predictable, not subject to ageing effects or component drift and easily mass-produced.

A method of calculating approximations to  $H(s)$  when the source signals are not white (which, of course, includes all speech or music signals) is proposed in a second aspect of the invention, using the idea of a 'pre-whitening filter'.

According to a second aspect of the invention, there is provided a method of coding polyphonic input signals comprising:

- producing therefrom a sum signal representing the sum of such signals; and reconstruction data to enable the formation, from the sum signal, of a further one of the input signals;
- producing from the input signals at least one difference signal representing a difference therebetween;
- analysing said sum and difference signals and generating therefrom a plurality of coefficients which, if applied to a multi-stage predictor filter, would enable the prediction of the difference signal from the sum signal thus filtered;
- the coded output comprising the said sum signal and data enabling the reconstruction of the said difference signal therefrom;

characterised by, before said analysis, filtering the sum signal and difference signal in accordance with a filter approximating the spectral inverse of the sum signal, the sum signal thereby being substantially spectrally whitened.

Other aspects of the invention are as claimed and enclosed herewith.

The words "prediction" and "predictor" in this specification include not only prediction of future data from past data, but also estimation of present data of a channel from past and present data of another channel.

The invention will now be illustrated, by way of example only, with reference to the accompanying drawings, where it should be noted that the arrangements shown in Figures 1 to 4 do not fall within the scope of the claims as amended during opposition proceedings. In the drawings:

- Figure 1 illustrates generally a first version of en-

- coder;
- Figure 2 illustrates generally a corresponding decoder;
- Figure 3a illustrates a second version of encoder;
- Figure 3b illustrates a corresponding decoder;
- Figures 4a and 4b show respectively a corresponding encoder and decoder according to a third version;
- Figures 5a and 5b illustrate an encoder and a decoder according to an embodiment of the invention;
- Figure 6 illustrates part of an encoder according to a yet further embodiment of the invention.

The embodiments illustrated are restricted to 2 channels (stereo) for ease of presentation, but the invention may be generalised to any number of channels.

One possible way of removing the redundancy between two input signals (or predicting one from the other) would be to connect between the two channels an adaptive predictor filter whose slowly changing parameters are calculated by standard techniques (such as, for example, block cross-correlation analysis or sequential lattice adaptation). In an audioconferencing environment, the two signals will originate from sound sources within a room, and the acoustic transfer function between each source and each microphone will be characterised typically by weak poles (from room resonances) and strong zeros (due to absorption and destructive interference). An all-zero filter could therefore produce a reasonable approximation to the acoustic transfer function between a source and a microphone and such a filter could also be used to predict say the left microphone signal  $x_L(t)$  from  $x_R(t)$  when the source is close to the right microphone. However, if the source were now moved away from the right microphone and placed close to the left, the nature of the required filter would be effectively inverted even when delays are introduced to guarantee causality. The filter must now model a transfer function with weak zeros and strong poles - a difficult task for an all-zero filter. Other types of filter are not, in general, inherently stable. The net effect of this is to cause unequal degradation in the reconstructed channel when the source shifts from one microphone to the other. This further makes the simplistic prediction of one channel (say, the left) from the other (say, the right) hard to realise.

In a system according to the first aspect of the invention, better results have been obtained by forming a "sum signal"  $x_S(t) = x_L(t) + x_R(t)$  and predicting either a difference signal  $x_D(t) = x_L(t) - x_R(t)$  or simply  $x_L(t)$  or  $x_R(t)$  using an all-zero adaptive digital filter.

In practice,  $x_R(t)$  and  $x_L(t)$  (or  $x_S(t)$  and  $x_D(t)$ ) will be processed in sampled data form as the digital signals  $x_R[n]$  and  $x_L[n]$  ( or  $x_S[n]$  and  $x_D[n]$  ) and it will be more convenient to use the 'z-transform' transfer function  $H(z)$  rather than  $H(s)$

Referring to Figure 1, in its essential form the invention comprises a pair of inputs 1a, 1b for receiving a pair

of speech signals, e.g. from left and right microphones. The signals at the inputs,  $x_R(t)$  and  $x_L(t)$ , may be in digital form. It may be convenient at this point to pre-process the signals, e.g. by band limiting. Each signal is then supplied to an adder 2 and a subtractor 3, the output of the adder being the sum signal  $x_S(t) = x_R(t) + x_L(t)$ , and the output of the subtractor 3 being the difference signal  $x_D(t) = x_R(t) - x_L(t)$  i.e.  $X_D(s) = H(s) X_S(s)$ . The sum and difference signals are then supplied to filter derivation stage 4, which derives the coefficients of a multi-stage prediction filter which, when driven with the sum signal, will approximate the difference signal. The difference between the approximated difference signal and the actual difference signal, the prediction residual signal, will usually also be produced (although this is not invariably necessary). The sum signal is then encoded (preferably using LPC or sub-band coding), for transmission or storage, along with further data enabling reconstruction of the difference signal. The filter coefficients may be sent, or alternatively (as discussed further below), the residual signal may be transmitted, the difference channel being reconstituted by deriving the filter parameters at the receiver using a backwards adaptive process known in the art; or both may be transmitted.

Although it would be possible to calculate filter parameters directly (using LPC analysis techniques), one simple and effective way of providing the derivation stage 4 is to use an adaptive filter (for example, an adaptive transversal filter) receiving as input the sum channel and modelling the difference channel so as to reduce the prediction residual. Such general techniques of filter adaptation are well-known in the art.

Our initial experiments with this structure have used a transversal FIR filter with coefficient update by an algorithm for minimising the mean square value of the residual, which is simple to implement. The filter coefficients change only slowly because the room acoustic (and hence the interchannel transfer function) is relatively stable.

Referring to Figure 2, in a corresponding receiver, the sum signal  $x_S(t)$  is received together with either the filter parameters or the residual signal, or both, for the difference channel, and an adaptive filter 5 corresponding to that for which the parameters were derived at the coder receives as input the sum signal and produces as output the reconstructed difference signal when configured either with the received parameters or with parameters derived by backwards adaptation from the received residual signal. Sum and difference signals are then both fed to an adder 6 and a subtractor 7, which produce as outputs respectively the reconstructed left and right channels at output nodes 8a and 8b.

Since a high-quality sum signal is sent, the encoder is fully mono-compatible. In the event of loss of stereo information, monophonic back-up is thus available.

As discussed above, one component of the transfer functions  $H_L$  and  $H_R$  is a delay component relating to the direct distance between the signal source and each

of the microphones, and there is a corresponding delay difference  $d$ . There is thus a strong cross-correlation between one channel and the other when delayed by  $d$ .

This method, however, requires considerably processing power.

An alternative method of delay estimation found in papers on sonar research is to use an adaptive filter. The left channel input is delayed by half the filter length and the coefficients are updated using the LMS algorithm to minimise the mean-square error of the output. The transversal filter coefficients will, in theory, become the required cross-correlation coefficients. This may seem like unnecessary repetition of filter coefficient derivation were it not for the property of this delay estimator that the maximum value of the cross-correlation coefficient (at the position of the maximum filter coefficient) is obtained some time before the filter has converged. This method may be improved further because spatial information is also available from the relative amplitudes of the input channels; this could be used to apply a weighting function to the filter coefficients to speed convergence.

Referring to Figure 3a, in a preferred embodiment of the invention, the complexity and length of the filter to be calculated is therefore reduced by calculating the required value of  $d$  in a delay calculator stage 9 (preferably employing one of the above methods), and then bringing the channels into time alignment by delaying one or other by  $d$  using, for example, a pair of variable delays 10a, 10b (although one fixed and one variable delay could be used) controlled by the delay calculator 9. With the major part of the speech information in the channels time aligned, the sum and difference signals are then formed.

Referring to Figure 3b, the delay length  $d$  is preferably transmitted to the decoder, so that after reconstructing the difference channel and subsequently the left and right channels, corresponding variable length delay stages 11a, 11b in one or other of the channels can restore the interchannel delay.

In the illustrated structure, the "sum" signal is thus no longer quite the true sum of  $x_L(t) + x_R(t)$ ; because of the delay  $d$  it is  $x_L(t) + x_R(t-d)$ . It may therefore be preferred to locate the delays 10a, 10b (and, possibly, the delay calculator) downstream of the adder and subtractor 2 and 3; this gives, for practical purposes, the same benefits of reducing the necessary filter length.

In practice, the delay is generally imperceptible; typically, up to 1.6 ms. Alternatively, a fixed delay, sufficiently long to guarantee causality, may be used, thus removing the need to encode the delay parameter.

In the first embodiment of the invention, as stated above, only the filter parameters are transmitted as difference signal data. With 16 bits per coefficient, this meant that a transmission capacity of 5120 bits/sec is needed for the difference channel (plus 8 bits for the delay parameter). This is well within the capacity of a standard 64 kbit/sec transmission system used which

allocates 48 kbits/sec to the sum channel (efficiently transmitted by an existing monophonic encoding technique) and offers 16 kbits/sec for other "overhead" data. This mode of the embodiment gives a good signal to noise ratio and the stereo image is present, although it is highly dependent on the accuracy of the algorithm used to adapt the predictive filter. Inaccuracies tend to cause the stereo image to wander during the course of a conference particularly when the conversation is passed from one speaking person to another at some distance from the first.

Referring to Figure 4a, in a second embodiment of the invention, only the residual signal is transmitted as difference signal data. The sum signal is encoded (12a) using, for example, sub-band coding. It is also locally decoded (13a) to provide a signal equivalent to that at the decoder, for input to adaptive filter 4. The residual difference channel is also encoded (possibly including bandlimiting) by residual coder 12b, and a corresponding local decoder 13b provides the signal minimised to adapt filter 4. The advantage this creates is that inaccuracies in generating the parameters cause an increase in the dynamic range of the residual channel and a corresponding decrease in SNR, but with no loss in stereo image.

Referring to Figure 4b, at the decoder, the analysis filter parameters are recovered from the transmitted residual by using a backwards-adapting replica filter 5 of the adaptive filter 4 at the coder. Decoders 13c, 13d are identical to local decoders 13a, 13b and so the filter 5 receives the same inputs, and thus produces the same parameters, as that of encoder filter 4.

In a further embodiment (not shown), both filter parameters and residual signal are transmitted as side-information, overcoming many of the problems with the residual-only embodiment because the important stereo information in the first 2 kHz is preserved intact and the relative amplitude information at higher frequencies is largely retained by the filter parameters.

Both the above residual-only and hybrid (i.e. residual plus parameters) embodiments are preferably employed, as described, to predict the difference channel from the sum channel. However, it is found that the same advantages of retaining the stereo image (albeit with a decrease in SNR) are found when the input channels are left and right, rather than sum and difference, provided the problem of causality is overcome in some manner (e.g. by inserting a relatively long fixed delay in one or other path). The scope of the invention therefore encompasses this also.

The parameter-only embodiment described above preferably uses a single adaptive filter 4 to remove redundancy between the sum and difference channels. An effect discovered during testing was a curious 'whispering' effect if the coefficients were not sent at a certain rate, which was far above what should have been necessary to describe changes in the acoustic environment. This was because the adaptive filter, in addition to mod-

elling the room acoustic transfer function, was also trying to perform an LPC analysis of the speech.

This is solved in the second aspect of the invention by whitening the spectra of the input signals to the adaptive filter as shown in Figure 5, so as to reduce the rapidly-changing speech component leaving principally the room acoustic component.

In the second aspect of the invention, the adaptive filter 4 which models the acoustic transfer functions may be the same as before (for example, a lattice filter of order 10). The sum channel is passed through a whitening filter 14a (which may be lattice or a simple transversal structure).

The master whitening filter 14a receives the sum channel and adapts to derive an approximate spectral inverse filter to the sum signal (or, at least, the speech components thereof) by minimising its own output. The output of the filter 14a is therefore substantially white. The parameters derived by the master filter 14a are supplied to the slave whitening filter 14b, which is connected to receive and filter the difference signal. The output of the slave whitening filter 14b is therefore the difference signal filtered by the inverse of the sum signal, which substantially removes common signal components, reducing the correlation between the two and leaving the output of 14b as consisting primarily of the acoustic response of the room. It thus reduces the dynamic range of the residual considerably.

The effect is to whiten the sum channel and to partially whiten the difference channel without affecting the spectral differences between them as a result of room acoustics, so that the derived coefficients of adaptive filter 4 are model parameters of the room acoustics.

In one embodiment, the coefficients only are transmitted and the decoder is simply that of Figure 2 (needing no further filters). In this embodiment, of course, residual encoder 12b and decoder 13b are omitted.

An adaptive filter will generally not be long enough to filter out long-term information, such as pitch information in speech, so the sum channel will not be completely "white". However, if a long-term predictor (known in LPC coding) is additionally employed in filters 14a and 14b, then filter 4 could, in principle, be connected to filter the difference channel alone, and thus to model the inverse of the room acoustic.

Since this second aspect of the invention reduces the dynamic range of the residual, it is particularly advantageous to employ this whitening scheme with the residual-only transmission described above. In this case, prior to backwards adaptation at the decoder, it is necessary to filter the residual using the inverse of the whitening filter, or to filter the sum channel using the whitening filter. Either filter can be derived from the sum channel information which is transmitted.

Referring to Figure 5b, in residual-only transmission, an adaptive whitening filter 24a (identical to 14a at the encoder) receives the (decoded) sum channel and adapts to whiten its output. A slave filter 24b (identical

to 14b at the encoder) receives the coefficients of 24a. Using the whitened sum channel as its input, and adapting from the (decoded) residual by backwards adaptation, adaptive filter 5 regenerates a filtered signal which is added to the (decoded) residual and the sum is filtered by slave filter 24b to yield the difference channel. The sum and difference channels are then processed (6, 7 not shown) to yield the original left and right channels.

In a further embodiment (not shown), both residual and coefficients are transmitted.

Although this pre-whitening aspect of the invention has been described in relation to the preferred embodiment of the invention using sum and difference channels, it is also applicable where the two channels are 'left' and 'right' channels.

For a typical audioconferencing application, the residual will have a bandwidth of 8 kHz and must be quantised and transmitted using spare channel capacity of about 16 kbit/s. The whitened residual will be, in principle, small in mean square value, but will not be optimally whitened since the copy pre-whitening filter 14b through which the residual passes has coefficients derived to whiten the sum channel and not necessarily the difference channel. Typically, the dynamic range of the filtered signal is reduced by 12dB over the unfiltered difference channel. One approach to this residual quantisation problem is to reduce the bandwidth of the residual signal. This allows downsampling to a lower rate, with a consequential increase in bits per sample. It is well known that most of the spatial information in a stereo signal is contained within the 0-2 kHz band, and therefore reducing the residual bandwidth from 8 kHz to a value in excess of 2 kHz does not affect the perceived stereo image appreciably. Results have shown that reducing the residual bandwidth to 4 kHz (and taking the upper 4 kHz band to be identical to that of the sum channel) produces good quality stereophonic speech when the reduced bandwidth residual is subband coded using a standard technique.

Experiments with various adaptive filters for the filter 4 (and, where applicable, 12) showed that a standard transversal FIR filter was slow to converge. A faster performance can be obtained by using a lattice structure, with coefficient update using a gradient algorithm based on Burg's method, as shown in Figure 7.

The structure uses a lattice filter 14a to pre-whiten the spectrum of the primary input. The decorrelated backwards residual outputs are then used as inputs to a simple linear combiner which attempts to model the input spectrum of the secondary input. Although the modelling process is the same as with the simple transversal FIR filter, the effect of the lattice filter is to point the error vector in the direction of the optimum LMS residual solution. This speeds convergence considerably. A lattice filter of order 20 is found effective in practice.

The lattice filter structure is particularly useful as described above, but could also be used in a system in which, instead of forming sum and difference signals, a

(suitably delayed) left channel is predicted from the right channel.

Although the embodiments described show a stereophonic system, it will be appreciated that with, for example, quadrophonic systems, the invention is implemented by forming a sum signal and 3 difference signals, and predicting each from the sum signal as above.

Whilst the invention has been described as applied to a low bit-rate transmission system, e.g. for teleconferencing, it is also useful for example for digital storage of music on well known digital record carriers such as Compact Discs, by providing a formatting means for arranging the data in a format suitable for such record carriers.

Conveniently, much or all of the signal processing involved is realised in a single suitably programmed digital signal processing (dsp) chip package; two channel packages are also commercially available. Software to implement adaptive filters, LPC analysis and cross-correlations are well known.

## Claims

1. Polyphonic signal coding apparatus comprising:

- means for receiving a first ( $x_S$ ) and at least one second channel ( $x_D$ );
- means (4) for periodically generating reconstruction data enabling the formation, from the first channel, of an estimate of the second channel, the generating means (4) being operable to generate a plurality of filter coefficients which, if applied to a plural order predictor filter, would enable the prediction of the second channel from the first channel thus filtered;
- means for outputting data representing the said first channel and the reconstruction data;

characterised in that the apparatus further comprises means for filtering the first and second channel in accordance with a filter (14a, 14b) approximating the spectral inverse of the first channel to produce respective filtered channels, the first said filtered channel thereby being substantially spectrally whitened, and the generating means (4) being connected to receive the filtered channels

2. Apparatus according to claim 1, in which the generating means includes an adaptive filter (4) connected to receive the first channel and produce a predicted second channel therefrom; and means for producing a residual signal representing the difference between the said predicted second channel and the actual second channel, and in which the said reconstruction data comprises data representing the said residual signal.

3. Apparatus according to claim 1 or claim 2, wherein the reconstruction data comprises the said filter coefficients ( $h_i$ ).

4. Apparatus according to claim 2 in which the adaptive filter (4) is controlled only by the said residual signal and the said reconstruction data consists of the said residual signal.

5. Apparatus according to any one of claims 1 to 4, wherein said filtering means comprises an adaptive, master, filter (14a) arranged to filter the first channel so as to produce a whitened output, and a slave filter (14b) arranged to filter said second channel, the slave filter being configured so as to have an equivalent response to the adaptive filter of the filtering means.

6. Apparatus according to any one of claims 1 to 5, further comprising:

- input means for receiving input signals; and
- means (2, 3) for producing the said channels therefrom, the first channel being a sum channel representing the sum of such input signals and the second or further channels representing the differences therebetween.

7. Apparatus according to any one of claims 1 to 6, including variable delay means for delaying at least one of the channels, and means for controlling the differential delay applied to the channels so as to increase the correlation upstream of the generating means, the output means being arranged to output also data representing the said differential delay.

8. Apparatus according to claim 6, in which the input means includes variable delay means (10a, 10b) for delaying the least one of the input signals, and means (9) for controlling the differential delay applied to the signals so as to increase the correlation upstream of the generating means, the output means being arranged to output also data representing the said differential delay.

9. Polyphonic signal decoding apparatus comprising:

- means for receiving data representing a sum signal, and signal reconstruction data; and means operable in response to the reconstruction data to modify the sum signal so as to produce at least two output signals, the modifying means comprising:
- a configurable plural order predictor filter (5) for receiving said signal reconstruction data and modifying its coefficients in accordance therewith, the filter being connected to receive the said sum signal and reconstruct therefrom an

- output difference signal; and
- means (6) for adding the reconstructed difference signal to the sum signal, and (7) for subtracting the reconstructed difference signal from the sum signal, so as to produce at least two output signals;

characterised by an adaptive, master, filter (24a) arranged to filter the sum signal in accordance with approximately the spectral inverse of the sum signal so as to produce a whitened output, and a slave filter (24b) arranged to filter said difference signal, the slave filter being configured so as to have an equivalent response to the adaptive master filter.

10. Apparatus as claimed in claim 9, in which the difference signal reconstruction data comprises residual signal data and the apparatus includes means to add the residual signal data to the output of the filter to form the reconstructed difference signal.

11. Apparatus as claimed in claim 10, in which the predictor filter (5) is connected to receive the residual signal data and to modify its coefficients in accordance therewith.

12. A method of coding polyphonic input signals comprising:

- producing therefrom a sum signal representing the sum of such signals; and reconstruction data to enable the formation, from the sum signal, of a further one of the input signals;
- producing from the input signals at least one difference signal representing a difference therebetween;
- analysing said sum and difference signals and generating therefrom a plurality of coefficients which, if applied to a multi-stage predictor filter, would enable the prediction of the difference signal from the sum signal thus filtered;
- the coded output comprising the said sum signal and data enabling the reconstruction of the said difference signal therefrom;

characterised by, before said analysis, filtering the sum signal and difference signal in accordance with a filter approximating the spectral inverse of the sum signal, the sum signal thereby being substantially spectrally whitened.

## Patentansprüche

1. Polyphonsignalkodierungsvorrichtung, die umfaßt:

- Empfangsvorrichtung für einen ersten ( $x_S$ ) und wenigstens einen zweiten Kanal ( $x_0$ );

- Vorrichtung (4) für die periodische Generierung von Rekonstruktionsdaten, die aufgrund des ersten Kanals die Abschätzung des zweiten Kanals ermöglichen, wobei die Generatorvorrichtung (4) dazu dient, eine Vielzahl von Filterkoeffizienten zu erzeugen, welche bei Anwendung in einem Prädiktorfilter multipler Ordnung die Vorhersage des zweiten Kanals aufgrund des so gefilterten ersten Kanals ermöglichen,
- Ausgabevorrichtung für Daten, die den besagten ersten Kanal repräsentieren, und der Rekonstruktionsdaten,

dadurch **gekennzeichnet**, daß die Vorrichtung außerdem umfaßt:

- Filtervorrichtung für den ersten und zweiten Kanal in Übereinstimmung mit einem Filter (14a, 14b), das sich dem spektralen Inversen des ersten Kanals nähert, um jeweils gefilterte Kanäle zu erzeugen, wobei der erste besagte gefilterte Kanal im wesentlichen spektral weißgefärbt wird und die Generatorvorrichtung (4) so geschaltet ist, daß sie die gefilterten Kanäle empfängt.

2. Vorrichtung nach Anspruch 1, bei welcher die Generatorvorrichtung ein adaptives Filter (4) umfaßt, das so geschaltet ist, daß es den ersten Kanal empfängt und daraus einen vorhergesagten zweiten Kanal erstellt; und Vorrichtung für die Erzeugung eines Restsignals, das die Differenz zwischen dem besagten vorhergesagten zweiten Kanal und dem tatsächlichen zweiten Kanal darstellt, und bei welchem die besagten Rekonstruktionsdaten Daten umfassen, die das besagte Restsignal darstellen.

3. Vorrichtung nach Anspruch 1 oder 2, wobei die Rekonstruktionsdaten die besagten Filterkoeffizienten ( $h_i$ ) umfassen.

4. Vorrichtung nach Anspruch 2, bei welcher das adaptive Filter (4) nur durch das besagte Restsignal gesteuert wird und die besagten Rekonstruktionsdaten aus dem besagten Restsignal bestehen.

5. Vorrichtung nach einem der Ansprüche 1 bis 4, wobei die besagte Filtervorrichtung einen adaptiven Hauptfilter (14a), der dazu dient, den ersten Kanal zu filtern, um einen weißgefärbten Ausgang zu erzeugen, und einen Nebfilter (14b) umfaßt, der dazu dient, den zweiten Kanal zu filtern, wobei der Nebfilter so konfiguriert ist, daß er äquivalent auf den adaptiven Filter der Filtervorrichtung reagiert.

6. Vorrichtung nach einem der Ansprüche 1 bis 5, die außerdem umfaßt:



- Eingangsvorrichtung für den Empfang von Eingangssignalen; und
- Vorrichtung (2, 3) für die Erzeugung von deren Kanälen, wobei der erste Kanal ein Summenkanal ist, der die Summe der Eingangssignale darstellt, und der zweite oder weitere Kanäle die Differenzen dazwischen darstellen.

7. Vorrichtung nach einem der Ansprüche 1 bis 6, die variable Verzögerungsvorrichtungen für die Verzögerung wenigstens eines der Kanäle und eine Vorrichtung für die Steuerung der differentiellen Verzögerung umfaßt, die auf die Kanäle angewendet werden, um die Korrelation oberhalb der Generatorvorrichtung zu erhöhen, wobei die Ausgangsvorrichtung dazu dient, um auch Daten auszugeben, die die besagte differentielle Verzögerung darstellt.

8. Vorrichtung nach Anspruch 6, bei welcher die Eingangsvorrichtung variable Verzögerungsvorrichtungen (10a, 10b) für die Verzögerung wenigstens eines der Eingangssignale und eine Vorrichtung (9) für die Steuerung der differentiellen Verzögerung umfaßt, die auf die Signale angewendet wird, um die Korrelation oberhalb der Erzeugungsvorrichtung zu erhöhen, wobei die Ausgangsvorrichtung dazu dient, auch Daten auszugeben, die die besagte differentielle Verzögerung darstellen.

9. Polyphonsignaldenkodiervorrichtung, die umfaßt:

- Vorrichtung für den Empfang von Daten, die ein Summensignal darstellen, und Signalrekonstruktionsdaten; und Vorrichtung, die dazu dient, auf die Rekonstruktionsdaten hin das Summensignal zu modifizieren, um wenigstens zwei Ausgangssignale zu erzeugen, wobei die Modifizierungsvorrichtung umfaßt:
- ein konfigurierbares Prädiktorfilter multipler Ordnung (5) für den Empfang der besagten Signalrekonstruktionsdaten und Modifizierung der Koeffizienten in Übereinstimmung damit, wobei das Filter so geschaltet ist, daß es das besagte Summensignal empfängt und daraus das Ausgangsdifferenzsignal rekonstruiert; und
- Vorrichtung (6) für das Addieren des rekonstruierten Differenzsignals auf das Summensignal, und (7) für die Subtraktion des rekonstruierten Differenzsignals von dem Summensignal, um wenigstens zwei Ausgangssignale zu erzeugen,

gekennzeichnet durch

einen adaptiven Hauptfilter (24a) zum Filtern des Summensignals in Übereinstimmung mit in etwa dem spektral Inversen des Summensignals, um einen weißgefärbten Ausgang zu erzeugen, und ei-

nen Nebenfilter (24b) zum Filtern des Differenzsignals, wobei der Nebenfilter so konfiguriert ist, daß er eine äquivalente Antwortfunktion zum adaptiven Hauptfilter hat.

10. Vorrichtung nach Anspruch 9, bei welcher die Differenzsignalrekonstruktionsdaten Restsignalen umfassen und die Vorrichtung eine Vorrichtung zur Addition der Restsignalen zu dem Ausgang des Filters umfaßt, um das rekonstruierte Differenzsignal zu bilden.

11. Vorrichtung nach Anspruch 10, bei welcher das Prädiktorfilter (5) so geschaltet ist, daß es die Restsignalen empfängt und seine Koeffizienten in Übereinstimmung damit modifiziert.

12. Verfahren zur Kodierung polyphoner Eingangssignale, das umfaßt:

- Erzeugung eines Summensignals daraus, das die Summe solcher Signale darstellt; und Rekonstruktionsdaten, so daß die Bildung eines weiteren der Eingangssignale aus dem Summensignal ermöglicht wird,
- Erzeugung mindestens eines Differenzsignals aus den Eingangssignalen, das die Differenz dazwischen darstellt;
- Analysieren des besagten Summen- und Differenzsignals und Erzeugung einer Vielzahl von Koeffizienten daraus, welche bei Anwendung in einem mehrstufigen Prädiktorfilter die Vorhersage des Differenzsignals aufgrund des so gefilterten Summensignals ermöglichte;
- wobei der kodierte Ausgang das besagte Summensignal und die Daten umfaßt, die die Rekonstruktion des besagten Differenzsignals daraus ermöglichen,

gekennzeichnet durch

Filtern des Summensignals und Differenzsignals in Übereinstimmung mit einem Filter, das sich dem spektral Inversen des Summensignals annähert, vor dem Analysieren, wobei das Summensignal dadurch spektral im wesentlichen weißgefärbt wird.

## Revendications

1. Dispositif de codage polyphonique de signaux comprenant :

- des moyens pour recevoir un premier canal ( $x_s$ ) et au moins un second canal ( $x_D$ ),
- des moyens (4) pour produire périodiquement des données de reconstitution permettant de former, à partir du premier canal, une estimation du second canal, les moyens de production

(4) pouvant fonctionner de manière à produire un ensemble de coefficients de filtrage, qui, s'ils sont appliqués à un filtre prédictif à plusieurs ordres, permettent la prédiction du second canal à partir du premier canal ainsi filtré, et

- des moyens pour délivrer des données représentant ledit premier canal et les données de reconstitution,

caractérisé en ce que le dispositif comprend en outre des moyens pour filtrer les premier et second canaux conformément à un filtre (14a, 14b) produisant un signal qui est approximativement l'inverse spectral du premier canal, pour produire des canaux filtrés respectifs, ledit premier canal filtré étant de ce fait sensiblement blanchi du point de vue spectral, et les moyens de production (4) étant reliés de manière à recevoir les canaux filtrés.

2. Dispositif selon la revendication 1, dans lequel les moyens de production incluent un filtre adaptatif (4) relié de manière à recevoir le premier canal et à produire un second canal prédit, à partir du premier canal ; et des moyens pour produire un signal résiduel représentant la différence entre ledit second canal prédit et le second canal réel, et dans lequel lesdites données de reconstitution comprennent des données représentant ledit signal résiduel. 20
3. Dispositif selon la revendication 1 ou 2, dans lequel les données de reconstitution comprennent lesdits coefficients de filtrage ( $h_i$ ). 30
4. Dispositif selon la revendication 2, dans lequel le filtre adaptatif (4) est commandé uniquement par ledit signal résiduel, et lesdites données de reconstitution sont constituées par ledit signal résiduel. 35
5. Dispositif selon l'une quelconque des revendications 1 à 4, dans lequel lesdits moyens de filtrage comprennent un filtre maître adaptatif (14a) agencé de manière à filtrer le premier canal pour produire un signal de sortie blanchi, et un filtre esclave (14b) agencé pour filtrer ledit second canal, le filtre esclave étant configuré de manière à posséder une réponse équivalente au filtre adaptatif des moyens de filtrage. 40 45
6. Dispositif selon l'une quelconque des revendications 1 à 5, comprenant en outre : 50
  - des moyens d'entrée pour recevoir des signaux d'entrée, et
  - des moyens (2,3) pour produire lesdits canaux à partir de ces signaux, le premier canal étant un canal de somme représentant la somme de tels signaux d'entrée, et le second canal ou d'autres canaux représentant les différences

entre ces signaux.

7. Dispositif selon l'une quelconque des revendications 1 à 6, comprenant des moyens de retardement produisant un retard variable servant à retarder au moins l'un des canaux, et des moyens pour commander le retard différentiel appliqué aux canaux afin d'accroître la corrélation en amont des moyens de production, les moyens de sortie étant agencés de manière à délivrer également des données représentant ledit retard différentiel. 5 10
8. Dispositif selon la revendication 6, dans lequel les moyens d'entrée comprennent des moyens de retardement produisant un retard variable (10a, 10b) pour retarder au moins l'un des signaux d'entrée, et des moyens (9) pour commander le retard différentiel appliqué aux signaux afin d'accroître la corrélation en amont des moyens de production, les moyens de sortie étant agencés de manière à délivrer également des données représentant ledit retard différentiel. 15
9. Dispositif de décodage polyphonique de signaux comprenant des moyens pour recevoir des données représentant un signal de somme et des données de reconstitution de signaux, et des moyens aptes à fonctionner en réponse aux données de reconstitution pour modifier le signal de somme afin de produire au moins deux signaux de sortie, les moyens de modification comprenant :
  - un filtre prédictif configurable à plusieurs ordres (5) pour recevoir lesdites données de reconstitution de signaux et modifier ses coefficients en fonction de ces données, le filtre étant relié de manière à recevoir ledit signal de somme et à reconstituer, à partir de là, un signal de différence de sortie, et
  - des moyens (6) pour ajouter le signal de différence reconstitué au signal de somme, et des moyens (7) pour soustraire le signal de différence reconstitué du signal de somme, afin de produire au moins deux signaux de sortie,
 caractérisé en ce qu'il comprend un filtre maître adaptatif (24a) agencé de manière à filtrer le signal de somme conformément à un signal qui est approximativement l'inverse spectral du signal de somme pour produire un signal de sortie blanchi, et un filtre esclave (24b) agencé pour filtrer ledit signal de différence, le filtre esclave étant configuré de manière à posséder une réponse équivalente au filtre maître adaptatif. 25 30 35 40 45 50
10. Dispositif selon la revendication 9, dans lequel les données de reconstitution du signal de différence comprennent des données de signal résiduel, et le 55

dispositif comprend des moyens pour ajouter les données de signal résiduel au signal de sortie du filtre de manière à former le signal de différence reconstitué.

5

11. Dispositif selon la revendication 10, dans lequel le filtre prédictif (5) est relié de manière à recevoir les données de signal résiduel et à modifier ses coefficients en fonction de ces données.

10

12. Procédé de codage polyphonique de signaux d'entrée consistant à :

- produire, à partir de ces signaux, un signal de somme représentant la somme de tels signaux, et des données de reconstitution pour permettre la formation, à partir du signal de somme, d'un autre des signaux d'entrée, 15
- produire, à partir des signaux d'entrée, au moins un signal de différence représentant une différence entre ces signaux, 20
- analyser lesdits signaux de somme et de différence et produire, à partir de ces signaux, un ensemble de coefficients qui, s'ils sont appliqués à un filtre prédictif à plusieurs étages, permettent de prédire le signal de différence à partir du signal de somme ainsi filtré, 25
- produire un signal de sortie codé comprenant ledit signal de somme et des données permettant, à partir de là, la reconstitution dudit signal de différence, 30

caractérisé par l'étape consistant, avant ladite analyse, à filtrer le signal de somme et le signal de différence conformément à un filtre produisant un signal qui est approximativement l'inverse spectral du signal de somme, ledit signal de somme étant de ce fait sensiblement blanchi du point de vue spectral. 35

40

45

50

55

Fig. 1.

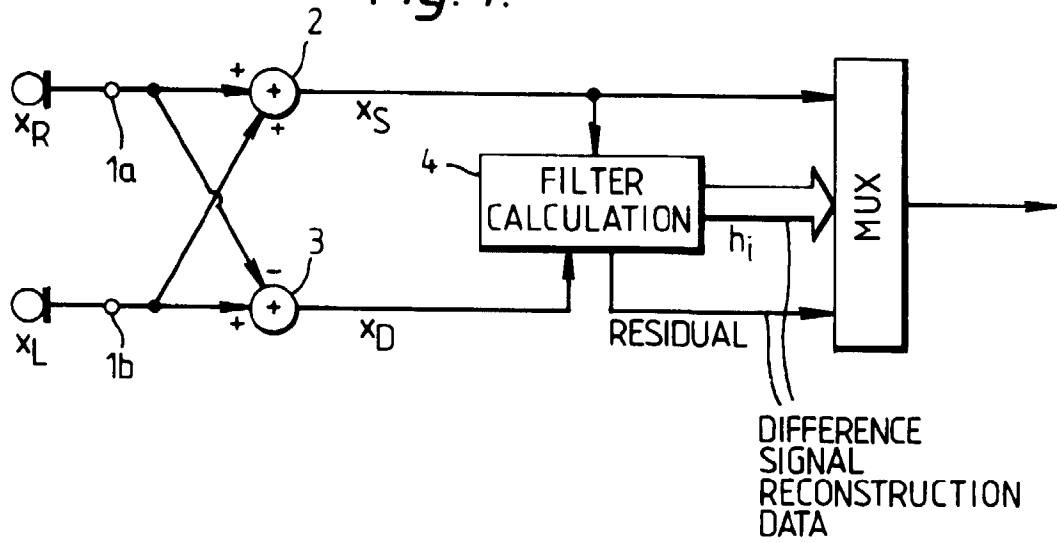


Fig. 2.

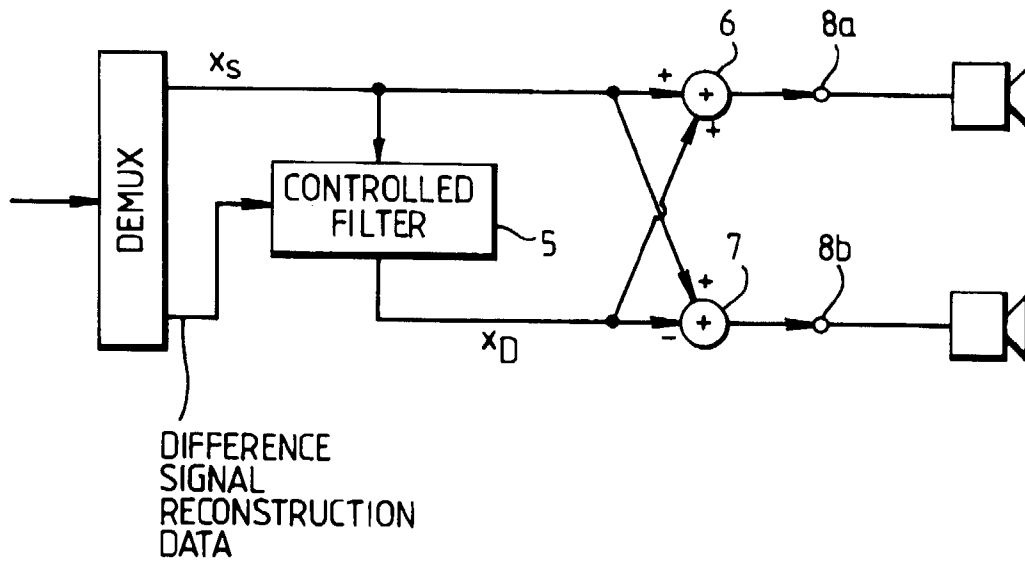


Fig. 3a

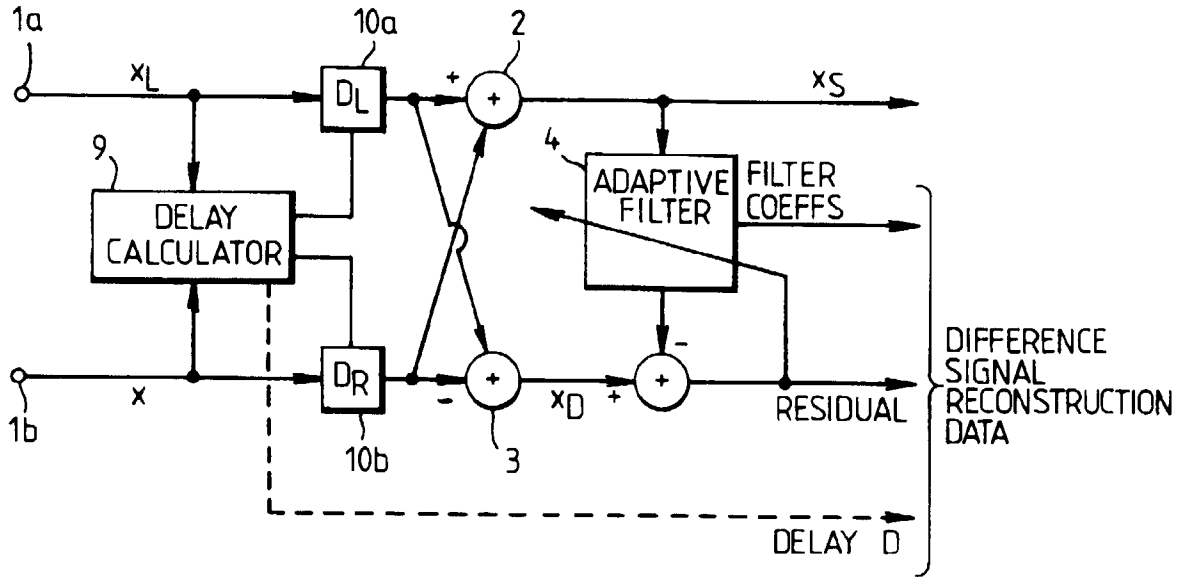


Fig. 3b

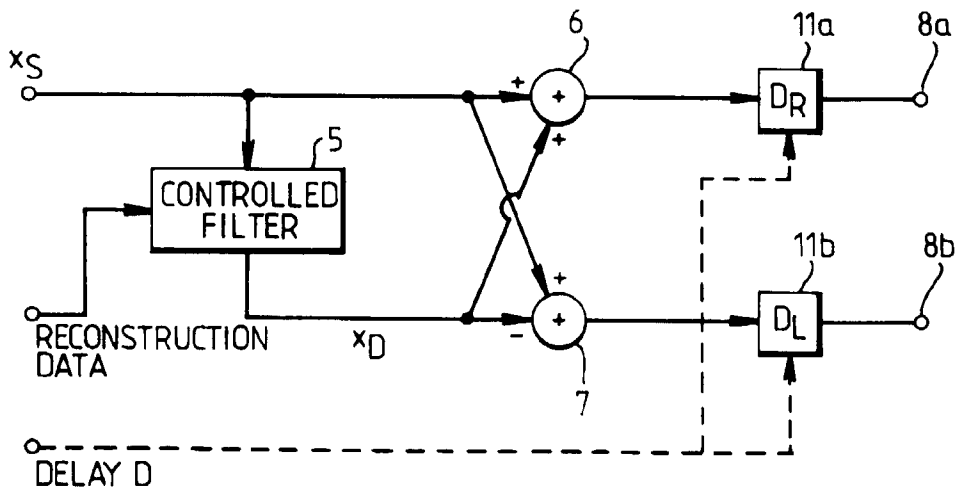


Fig. 4b

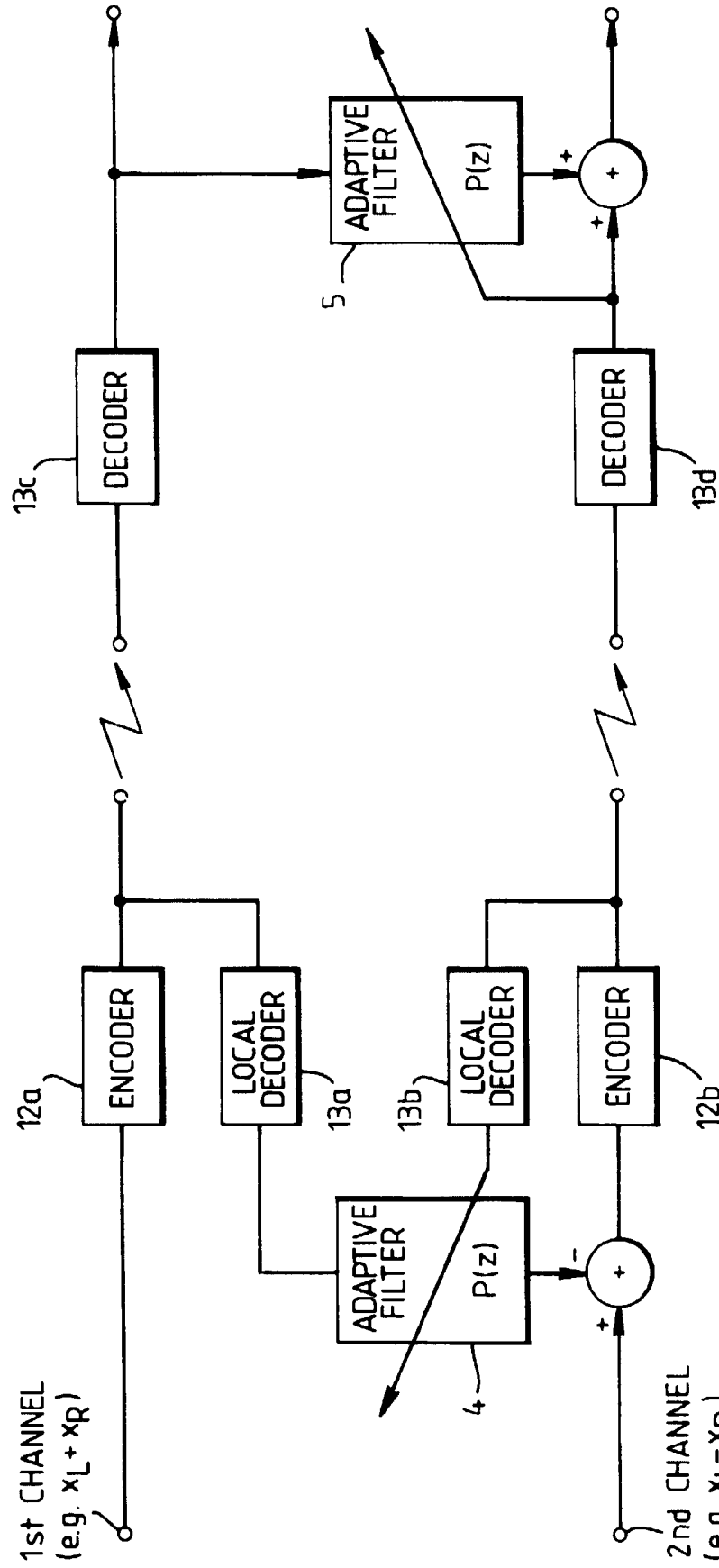


Fig. 4a

Fig. 5b

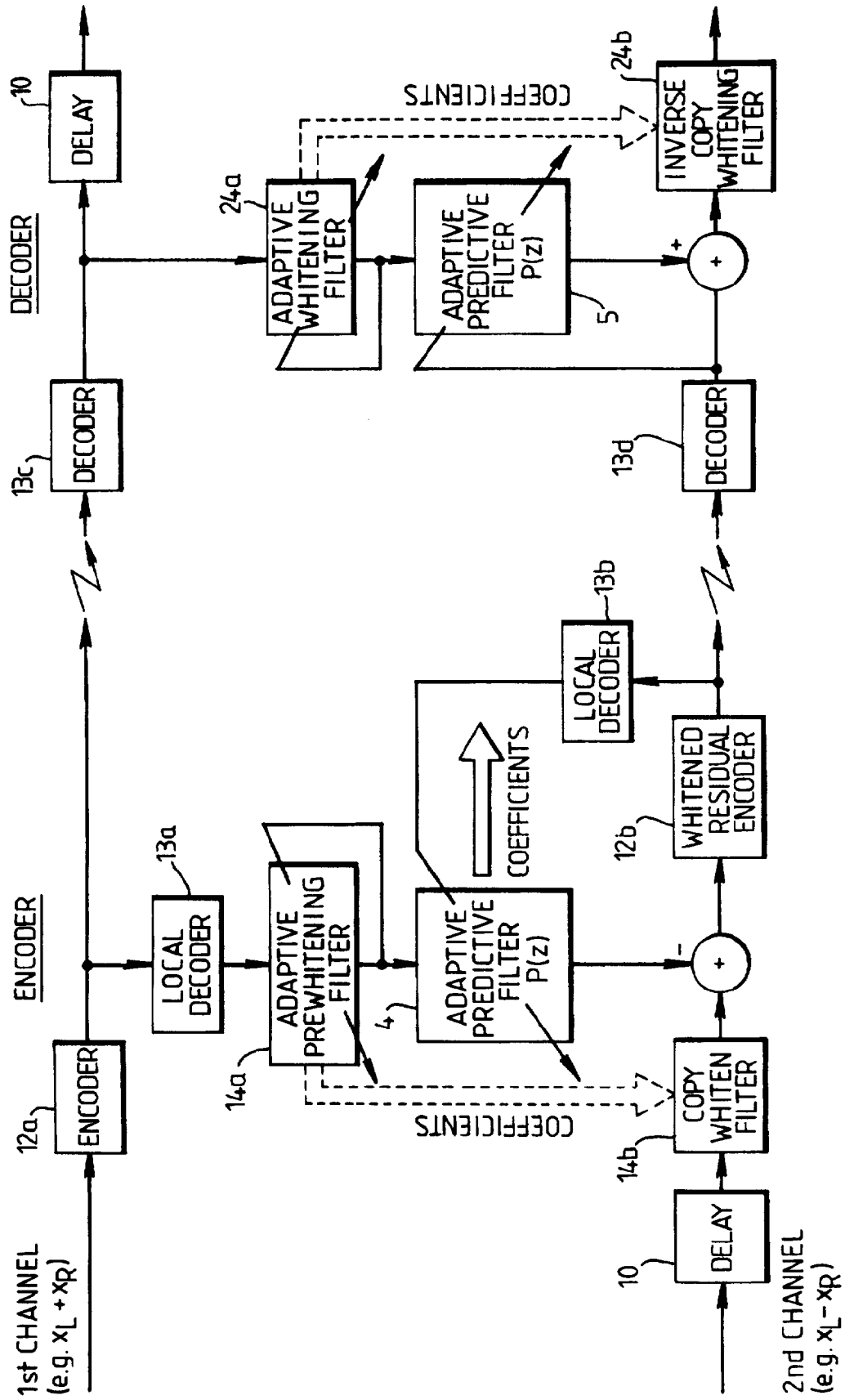


Fig. 5a

Fig. 6.

