

ECHO CANCELLATION

This invention relates to echo cancellation, in particular for use in spatialised
5 audio systems such as are used in videoconference systems, that is, systems in
which two or more users at different locations participate in a conversation using
audiovisual telecommunications systems. At present most systems use a wideband
monophonic audio system with the sound reproduced using either an internal
loudspeaker fitted to the vision monitor, or an auxiliary hi-fi loudspeaker. This
10 provides good quality audio, but there is no method of co-locating the video and
sound images. For a small Visual Display Unit this is not a problem, but as the move
to more lifelike and immersive environments continues it will be necessary to have
spatialised sound.

In any two-way audio system, in which sound is travelling both to and from
15 the same user there is a problem of acoustic feedback or "echo" between the
loudspeaker of the input channel and the microphone of the output channel. As
shown in Figure 1, which shows a simple mono system, echo occurs when there is
an acoustic path between the loudspeaker (22) and the microphone (12) in Room "B"
as represented by $h_1(t)$ in Figure 1. The talker 31 in room 'A' hears an echo of his
20 own voice which passes from the microphone 11 in Room "A", the outward channel
41, the loudspeaker 22, in Room "B", the acoustic path $h_1(t)$, the microphone 12 in
Room "B", the return channel 42 and the loudspeaker 21 in Room "A". The listener
32 in room 'B' also hears an echo of the talker 31, having passed through the path
described above, and then through the acoustic paths $h_2(t)$ between the loudspeaker
25 21 and microphone 11 in Room "A" to pass a second time over the outward channel
41 to the loudspeaker 22 in Room "B". These are called 'talker echo' and 'listener
echo' respectively.

Depending on their size and time delays, echoes may be anything from
unnoticeable to a devastating impairment resulting in instability and howling.

30 The creation of echo may be avoided by the use of body-mounted devices.
Headphones can be used to isolate the incoming sound from the microphone.
Alternatively, "close" microphones (e.g. clipped to the user's clothing), can also be
used. These have a sensitivity which falls off rapidly with distance, so that the

wearer's voice is detected clearly whilst signals emanating from a loudspeaker some distance away are only detected at a low volume. However, body-mounted devices are not always practical, and are inconvenient for the user to use, and require each user at a given location to be provided for individually.

5 In situations where the creation of echo cannot be avoided, some form of echo control or cancellation is desirable, to avoid the echo signal returning to the source of the original signal. A simple form of echo control, called echo suppression, is illustrated in Figure 2. Echoes are prevented by only allowing signal transmission if the person in the room is speaking. Switches 51, 52, or variable attenuators in
10 more sophisticated systems, are controlled by analysing the send and receive signals 41, 42 and using a decision algorithm to determine when transmission is permitted. This is a very effective echo control method, but can be very intrusive. Double talk, i.e. both parties talking at the same time, is not possible.

The principle behind echo cancellation may be seen from figure 3. The echo
15 signal in the return path 42, represented by $d(t)$, is caused by the acoustic path between loudspeaker and microphone. Cancellation is achieved in a canceller 62 by creating a synthetic model $\hat{h}(t)$ of the signal path such that the echo may be removed by subtraction in a combiner 90 in the return path 42. The signal $e(t)$ is now free of
20 echoes, containing only sounds that originated in room B. The modelling of $\hat{h}(t)$ is usually achieved by adaptation techniques which drives $e(t)$ towards zero, the simplest and commonest being called the 'least mean squares' (LMS) algorithm. Its drawback is that the time taken to adapt is dependent on the signal characteristics. Other algorithms such as the 'recursive least squares' (RLS), or 'affine projection' (AP) give better performance but greatly increased processing requirements.

25 Monophonic echo cancellation is a mature technology, and widely used in telecommunications – loud-speaking telephones, teleconference systems, network echo control and data transmission are good examples. For "artificial" spatialised sound systems, in which a mono signal is reproduced in two or more loudspeakers, manipulated linearly in gain and delay to seem to originate from the required
30 direction, monophonic echo cancellation techniques can be used. However, 'real' spatial sound, in which a multi-channel signal is transmitted and the sending room

characteristics are recreated in the receiving room, presents fundamental echo cancellation problems.

For multiple channel echo cancellation the number of echo paths is the product of the number of microphones and the number of loudspeakers in room B, as shown in figure 4 for a system with two of each. Figure 4 shows the echo adaptation units 62L, 62R associated with each of two input channels 41L, 41R for a microphone 12R, whose outputs are combined in a combiner unit 91 to provide a cancellation signal $\hat{h}(t)$. Note that further adaptation units will be required for the other path 42L (not shown) associated with the other microphone 12L.

The problem may thus be seen as one of characterising an unknown system having several inputs ($g_L(t)$, $g_R(t)$) for each output $\hat{h}(t)$. The result is that adaptation algorithms can only converge to a correct solution if $x_L(t)$ and $x_R(t)$ are completely independent. This is unlikely for a stereo (or multichannel) signal in which the original sources are associated. One approach for a motionless source is to assume that one channel may be derived from the other, allowing a single channel echo canceller 62 to model both echo paths $h_L(t)$, $h_R(t)$. This is illustrated in Figure 5. Causality is maintained by a selector 60 taking the canceller input from the channel ($g_L(t)$ or $g_R(t)$) with less delay. The drawback here is that the canceller 62 is now modelling both $h(t)$ and $g(t)$. If the source moves, or a different speaker starts talking, $g(t)$ will change (the relative gains and delays of $g_L(t)$ and $g_R(t)$), requiring re-adaptation. It also assumes both channels of $g(t)$ have stable inverses.

Another approach is to try to reduce the cross correlation between channels. This must be done without degrading the audio quality or stereo image. Several methods have been tried. Adding independent noise sources to each channel has been proposed, but the noise is audible. Frequency-shifting one channel relative to the other disturbs the stereo image. Decorrelating filters have also been tried, but adequate decorrelation cannot be obtained. Adding non-linear distortion has been claimed to be inaudible due to psycho-acoustic masking effects. The use of interleaving Comb filters, in which at any given frequency only one channel contains energy is also reported to work, but performance is poor below 1kHz.

Unless the decorrelation is very good a fast adaptation algorithm such as stereo fast RLS (recursive least squares) will be required. These are computationally

very expensive, requiring the order of 28 multiplications and 28 additions for each filter coefficient. For example, a 75ms stereo fast RLS echo canceller sampling at 16kHz requires more than 10^9 multiplications per second. Since even the most modern DSP devices (e.g. Texas TMS320C6X) only just manage 2×10^8 multiplications per second.

According to the invention, there is provided a method of echo cancellation for an audio conferencing system in which a participant in the conference generates a monaural signal for transmission on an output channel to the other participants, and receives a plurality of input channels from the other participants, the input channels
10 being combined linearly to provide a spatialised audio output comprising a plurality of audio channels, characterised in that an adaptive filter is applied to the input channels to generate an echo cancellation signal for applying to the output channel.

According to another aspect, there is provided echo cancellation apparatus for an audio conferencing system having an output channel and a plurality of input
15 channels, the input channels being combined linearly to provide a spatialised audio output comprising a plurality of audio channels, characterised in that an adaptive filter is applied to the input channels for generating an echo cancellation signal for applying to the output channel.

Because a linear spatialisation technique is used, the output from each
20 loudspeaker of the spatialised output is a linear combination of the inputs. This is true for the echo component as well as for the complete signal. Since the input channels are independent, the individual echo components are uncorrelated, and each one can be identified by an adaptive filter dedicated to that input channel. The correlation problems already discussed with reference to Figure 4 are therefore not encountered
25 in this system.

A separate adaptive filter may be provided for each input channel, the outputs of the filters being combined to form the cancellation signal. These filters may be monophonic, but in an arrangement in which the input signals are individually converted to spatialised signal sets (for example stereophonic signal pairs) by
30 respective predetermined transfer functions, the adaptive filters may be multiple channel (for example stereo) filters, correlations between the spatialised signal sets

being allowed for using an adaptation process driven by data relating to each transfer function. This allows adaptation to changes in the spatialisation functions.

In a variant of this arrangement, one multiple channel adaptive filter is applied to the combined input channel set to generate an echo cancellation signal for
5 applying to the output channel, correlations between the spatialised signal set being allowed for using an adaptation process based on data relating to the predetermined transfer functions (81, 82, 83).

The embodiment to be described in detail uses a transaural stereo spatialisation process, but the invention is also applicable to panned and ambisonic
10 transfer functions, having two or more loudspeakers.

Three embodiments of an echo cancellation system according to the invention will now be described, by way of example, with reference to the Figures, and in particular Figures 6, 7 and 8.

Figure 1 is a schematic representation of the generation of acoustic
15 feedback, or "echo"

Figure 2 is a schematic representation of a simple echo suppression system

Figure 3 is a schematic representation of a simple monophonic echo cancellation system

Figure 4 is a schematic representation of a simple stereophonic echo
20 cancellation system

Figure 5 is a schematic representation of another stereophonic echo cancellation system

Figures 1 to 5 have already been discussed.

Figure 6 is a schematic representation of a first echo cancellation system
25 according to the invention

Figure 7 is a schematic representation of a second echo cancellation system according to the invention

Figure 8 is a schematic representation of a third echo cancellation system according to the invention

30 The embodiments of Figures 6, 7 and 8 apply the principles of the invention to a transaural system – that is, one in which each incoming signal 411, 412, 413 is processed by a filter bank 81, 82, 83 with filter settings appropriate to allow two

loudspeakers 22L, 22R to be used to reproduce the sound to be heard by the respective left and right ears of a listener. These settings are known as the "Head Related Transfer Function" (HRTF). Filtering 71 is used to eliminate crosstalk between left speaker and right ear, and vice versa. The principles can also be applied

5 to panned stereo systems (in which the same mono signal is transmitted from both loudspeakers, with relative intensities selected to place the virtual sound source in the correct position) or ambisonic systems (in which two or more loudspeakers are controlled to reproduce the original sound field's wavefronts) by replacing the Head Related Transfer function filter banks (HRTFs) with appropriate scaling into the

10 required number of channels.

The embodiment of Figure 6 uses mono echo cancellers. Artificial transaural processing and ambisonic processing are both well-defined linear processes. A mono signal is transmitted on the return path 42, and mono signals are also received on each input channel 411, 412, 413, and placed in the ambisonic environment using

15 appropriate predetermined Head Related Transfer Functions (HRTFs) for the filter banks 81, 82, 83 whose outputs are combined (92L, 92R) to provide the signals for the loudspeakers 22L, 22R. These filter values may be selected from a store 99 by a control system 98 according to the spatialisation required for the respective input signals 411, 412, 413. Echo control can be achieved with a mono echo canceller

20 621, 622, 623 for each input. Each canceller convolves the output signal 42 with its respective input signal 411, 412, 413, thereby identifying any correlations with that input, and generates a cancellation signal. The separate echo cancellation signals can then each be added in a combiner 91 to generate a cancellation signal which is in turn added to the return path 42 (combiner 90) to cancel that part of the echo

25 attributable to each respective input channel

In contrast to existing echo cancellers, the echo path now consists of the room response (the acoustic path between the loudspeakers 22L, 22R and the microphone 12) convolved with the HRTFs 81, 82, 83 and the crosstalk canceller 71. This approach allows established echo cancellation techniques to be used.

30 However, source locations in the ambisonic environment may change, for example to simulate the movement of one of the sound sources. Alternatively, if the number of available channels is less than the number of potential sound sources, one

channel may be used by different sound sources, requiring different spatialisation, at different times in the same conference, as discussed in the applicant's co-pending application claiming priority from European Patent Application 98302763.2. Such changes to the ambisonic environment, and hence of the Head Related Transfer

5 Functions used to define the filter banks 81, 82, 83, require re-adaptation of the echo cancellers. This disadvantage can be overcome using the system shown in figure 7.

In the embodiment of Figure 7 stereo echo cancellers 621, 622, 623 are used, driven from the outputs, and not the inputs, of the respective HRTF filter banks

10 81, 82, 83. Correlations between the pairs of channels can be allowed for using the control processor 99, which modifies the parameters of the adaptation process used in the echo cancellers 621, 622, 623 according to the values of each HRTF. This system allows source locations to be changed without readapting. However, there is some redundancy in this system as all the stereo cancellers 621, 622, 623 should

15 converge to the same values, because they are all modelling the same pair of paths 92L, 71, 22L, 12, 90 and 92R, 71, 22R, 12, 90.

With the embodiments of Figures 6 and 7, the echo canceller is modelling the crosstalk canceller 71, and one echo canceller is required for each input. These problems can be overcome using the system in Figure 8. This uses a single stereo

20 echo canceller 62 taking its input either downstream of the combiners 92L, 92R and crosstalk canceller 71. The adaptation of the stereo echo canceller will now be based either on features common to all the HRTFs used in the filter banks 81, 82, 83 or, as shown, using the controller 99 to determine input signal levels, speech content or other factors to select which of the predetermined HRTFs is the most appropriate to

25 compensate for.

If steerable microphones or arrays are used, the number of possible 'channels' becomes the number of microphone directions, multiplied by the number of loudspeaker directions. This may be met by maintaining a set of echo canceller coefficients for each loudspeaker direction in the store 99, and selecting the

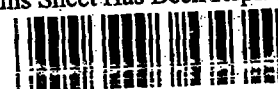
30 appropriate coefficients for the currently selected direction.



CLAIMS

1. A method of echo cancellation for an audio conferencing system in which a participant in the conference generates a monaural signal for transmission on an output channel (42) to the other participants, and receives a plurality of input channels (411, 412, 413) from the other participants, the input channels (411, 412, 413) being combined (92) linearly to provide a spatialised audio output comprising a plurality of audio channels (22L, 22R), characterised in that an adaptive filter process (621, 622, 623) is applied to the input channels (411, 412, 413) to generate a combined echo cancellation signal for applying to the output channel (42).
2. A method of echo cancellation for an audio conferencing system according to claim 1 in which a separate adaptive filter (621, 622, 623) is provided for each input channel (411, 412, 413), and the outputs of the filters are combined (91) to form the cancellation signal.
3. A method of echo cancellation for an audio conferencing system according to claim 2 in which the input signals (411, 412, 413) are individually converted to spatialised signal sets by respective filter banks having selected predetermined transfer functions (81, 82, 83), wherein the adaptive filters (621, 622, 623) are multiple channel filters, and correlations between the spatialised signal sets are allowed for by an adaptation process 61 driven by data selected according to the selected transfer functions (81, 82, 83).
4. A method of echo cancellation for an audio conferencing system according to claim 1 in which the input signals (411, 412, 413) are converted to spatialised signal sets by respective predetermined transfer functions (81, 82, 83), and combined (92) to form a single spatialised set, and wherein a multiple-channel adaptive filter (62) is applied to the combined input channel set to generate an echo cancellation signal for applying to the output channel (42), correlations between the spatialised signal set being allowed for using an adaptation process based on knowledge of the predetermined transfer functions (81, 82, 83).

This Sheet Has Been Replaced





5. A method of echo cancellation for an audio conferencing system according to any of claims 1 to 4 in which the spatialisation process is a transaural system.
- 5 6. A method of echo cancellation for an audio conferencing system according to any of claims 1 to 4 in which the spatialisation process is a panned stereo system
7. A method of echo cancellation for an audio conferencing system according to any of claims 1 to 4 in which the spatialisation process is an ambisonic system.
- 10
8. Echo cancellation apparatus for an audio conferencing system having an output channel (42) and a plurality of input channels (411, 412, 413), the input channels (411, 412, 413) being combined linearly to provide a spatialised audio output comprising a plurality of audio channels (22L, 22R), characterised in that an
15 adaptive filter (621, 622, 623) is applied to the input channels (411, 412, 413) for generating an echo cancellation signal for applying to the output channel (42).
9. Echo cancellation apparatus according to claim 8 in which a separate adaptive filter (621, 622, 623) is provided for each input channel (411, 412, 413), and the
20 outputs of the filters are connected to a combiner (91) to form the cancellation signal.
10. Echo cancellation apparatus according to claim 9 wherein the adaptive filters (621, 622, 623) are multiple channel filters for operating on spatialised signal sets
25 derived from the respective input signals (411, 412, 413) according to respective predetermined transfer functions (81, 82, 83); and an adaptation processor (99) driven by knowledge of each transfer function (81, 82, 83) is provided for controlling the adaptive filters to allow for correlations between the spatialised signal sets.





11. Echo cancellation apparatus according to claim 8 comprising a multiple channel adaptive filter (62) applied to a combined spatialised signal set derived from a plurality of spatialised signal sets derived from the respective input signals (411, 412, 413), the filter (62) being arranged to generate an echo cancellation signal for
5 applying to the output channel (42), and an adaptation processor (99) controlled in accordance with data relating to predetermined transfer functions (81, 82, 83) by which the input signals (411, 412, 413) are converted to said plurality of spatialised signal sets.
- 10 12. Apparatus according to any of claims 8 to 11, applied to a transaural spatialisation system.
13. Apparatus according to any of claims 8 to 11, applied to a panned stereo system
- 15 14. Apparatus according to any of claims 8 to 11, applied to an ambisonic system.

This Sheet Has Been Replaced

