(12)   **EUROPEAN PATENT APPLICATION**

(84) Designated Contracting States:
     **AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU
     MC NL PT SE**
     Designated Extension States:
     AL LT LV MK RO SI

(30) Priority: **30.12.1998 US 223485**

(71) Applicant: **SIEMENS CORPORATE RESEARCH,
     INC.
     Princeton, New Jersey 08540 (US)**

(72) Inventors:
     • **Rosca, Justinian
       Monmouth Junction, NJ 08852 (US)**

     • **Darken, Christian
       Riverton, NJ 08077 (US)**
     • **Petsche, Thomas
       Neschanic Station, NJ 08853 (US)**
     • **Holube, Inga
       91058 Erlangen (DE)**

(74) Representative: **O'Connell, David Christopher
     Haseltine Lake & Co.,
     Imperial House,
     15-19 Kingsway
     London WC2B 6UD (GB)**

(54)   **Blind source separation for hearing aids**

(57)   An electronic filtering device for performing real-time unmixing of a signal desired to be recovered by a user of the device, where the desired signal emanates from one of a plurality of independent signal sources. Two microphones positioned along a common axis develop first and second electrical input signals in response to reception by the microphones of acoustic signals from the plurality of independent signal sources. The common axis of the microphones is controllable in real time by the user to align the common axis so it points in the direction of the source of the desired signal. An adaptive unmixing signal processor responsive to the input signals develops output signals wherein the desired signal is separate from the mixture signal. A preprocessor may be provided to subject the input signals to one or both of a time delay processing and a decorrelation processing before their application to the unmixing signal processor, to enhance recovery of the desired signal. A selected output of the unmixing signal processor can be applied as an input to a speaker for reproduction, or can be further processed for signal enhancement by an additional processor before reproduction.

EP 1 017 253 A2

Description

BACKGROUND OF THE INVENTION

1. Field of the Invention

[0001] The present invention generally relates to electronic filtering for enhancing a desired signal component of a mixed signal, and more specifically to a method and apparatus for real-time unmixing (separation or deconvolving) of a desired signal from a mixture of independent signals, particularly useful, for example, in a hearing aid.

2. Description of the Prior Art

[0002] When one is listening to someone or something, "noise" or undesired signals that interfere with the voice or desired signal, are ubiquitous. People with hearing impairment are especially vulnerable to noise. Background conversations, interference from digital devices (mobile telephones), car, or other specific environment noises, can make it very difficult for a hearing impaired person to understand a desired speech signal. A reduction in the noise level of a signal, coupled with an automatic focus on a desired signal component, can significantly improve the performance of an electronic voice processor, such as one used in an advanced hearing aid.

[0003] In recent years, hearing aids using digital signal processing have been introduced. They contain one or more microphones, analog to digital converters, digital signal processors, and speakers. Usually the digital signal processors divide the incoming signals into several frequency regions using filter banks. Within each of those regions, signal gain and dynamic compression parameters can be individually adjusted in accordance with the requirement for a particular user of the hearing aid, in an attempt to improve intelligibility. Additionally, digital signal processing algorithms for feedback reduction and noise reduction are available, however they have major limitations. For example, some of the disadvantages of the currently available algorithms for noise reduction are the limited improvement they obtain when speech and background noise are in the same frequency region, due to their inability to distinguish between speech and background noise.

[0004] One relatively new digital signal processing approach currently finding use for noise reduction in areas such as speech recognition, data communication and sensor signal processing, involves a technique known generally as Independent Component Analysis (ICA), and in more specific applications as Blind Source Separation (BSS). This technique searches an input signal having multiple components, for a signal transformation which will minimize the statistical dependence between its components. Accordingly, BSS is a signal separation technique capable of delivering dramatic improvements in signal to noise ratio for mixtures of independent signals, such as multiple voices or mixtures of voice and noise signals.

[0005] It is an object of the present invention to provide an electronic filtering technique incorporating BSS processing which can operate in real time to enhance reception of a desired signal, such as the voice of a nearby person, and furthermore, if desired, can be incorporated in a hearing aid.

SUMMARY OF THE INVENTION

[0006] An electronic filtering device for performing real-time unmixing of a signal desired to be recovered by a user of the device, where the desired signal emanates from one of a plurality of independent signal sources. Two microphones positioned along a common axis develop first and second electrical input signals in response to reception by the microphones of acoustic signals from the plurality of independent signal sources. The spatial position of the common axis of the microphones is controllable in real time by the user to align the common axis so it points in the direction of the source of the desired signal, thereby imparting an inherent directionality to the input signals. An adaptive unmixing signal processor responsive to the input signals develops output signals wherein the desired signal is separated from the mixture signal. In one preferred embodiment of the invention a preprocessor is provided to enhance the inherent directionality of the input signals by establishing a relative time delay therebetween. Furthermore, the preprocessor may subject the enhanced input signals to a decorrelation processing before their application to the unmixing signal processor. A selected output of the unmixing signal processor can be applied as an input to a speaker for reproduction, or can be further processed for signal enhancement by an additional processor before reproduction.

BRIEF DESCRIPTION OF THE DRAWINGS

[0007]

Figure 1 illustrates in block diagram form an electronic filtering device constructed in accordance with the principles of the present invention;

Figure 2 illustrates in block diagram form the preprocessing stage of the electronic filtering device shown in Figure 1;

Figure 3 illustrates in block diagram form the technique of Blind Source Separation as used in the electronic filtering device of the invention; and

Figure 4 illustrates in block diagram form an exemplary embodiment of a Blind Source Separator useful in the electronic filtering device of the invention.

DETAILED DESCRIPTION OF THE INVENTION

[0008] Figure 1 illustrates in block diagram form an application of the invention for use in hearing aids. A hearing aid 10 includes two microphones 12 and 14 for developing two input signals 1 and 2, respectively. In accordance with one aspect of the invention, the microphones are mounted in the hearing aid such that a common axis of their positioning always extends substantially in the direction in which the wearer of the hearing aid looks when being attentive to a signal source such as a voice. This microphone positioning imparts an inherent directionality to input signals 1 and 2. Since each microphone develops electrical signals representative of the acoustic waves received thereby from sound sources within it's operating range, each input signal may comprise a mixture of unknown signals from an unknown number of signal sources. Input signals 1 and 2 are processed in three main stages. At a first stage 16, the input signals are preprocessed for enhancing the inherent directionality already imparted thereto by their positioning. At a second stage 18, the resulting signals are subjected to an unmixing processing (sometimes referred to as separation processing), which is designed to produce estimates of the original unknown signals picked-up by microphones 12 and 14. At a third stage 20, the outputs of the unmixing processing are preferably postprocessed to produce the desired signal 22, which can then be applied to a speaker 24 of the hearing aid 10 for reproduction and presentation to a user.

[0009] As illustrated in Fig. 2, preprocessing stage 16 begins with normalization of the raw input signals. Automatic Gain Control is used to normalize input signals 1 and 2 to a [-1,+1] range. The inputs 1 and 2 are now given in by a vector $\mathbf{x} = (x_1(t), x_2(t))$.

[0010] In accordance with one aspect of the invention, in order to adapt a blind source separation (BSS) technique for use in a device as small as a hearing aid, and to have it operate in real-time, preprocessing stage 16 also provides at least the first, and preferably both of the following additional processing:

- Enhancement of signal source directionality inherent in the input signals, resulting from a directional arrangement of microphones 12 and 14 with respect to a source of interest. In the hearing aid exemplary embodiment, the directionality of the source of interest is presumed to be in the direction that the user is looking. Accordingly, the microphones are positioned on the hearing aid along an axis that is in the direction that the user would be looking, and the direction of the source of interest is presumend to be at zero degrees with respect to such axis. The direction of a second source can be estimated in the preprocessing stage (delay box in 16) resulting in an adaptive delay (δ). The delay is a positive or negative fractional delay, such that the most powerful component of the inputs other than

the one approximately aligned with the microphone axis arrives synchronously at the two microphones. For example this would be zero if the second source were perpendicular to the microphone axis. For this enhancement, the normalized input signals $x = (x_1(t), x_2(t))$ are modified as follows:

$$x_1(t) = x_1(t)$$

$$x_2(t) = x_2(t-\delta)$$

- Decorrelation of the input signals. In the exemplary embodiment decorrelation is carried out by a diagonalization of the correlation matrix. More specifically, let $C = \text{Covariance}(\mathbf{x}^T)$, where $\mathbf{x}^T$ is a transpose of $\mathbf{x}$. If significant correlation exists between the two input signals (x_1, x_2), a decorrelation over a time window D means transformation of the signals in two steps: (1) centering around the mean over the data in the window D; and (2) Affine transformation of the resulting data points in order to diagonalize the covariance matrix of the resulting signals. Assuming that x is centered around its mean, we use the following transformation:

$$x = 2(\sqrt{C})^{-1} x$$

[0011] In the illustrated embodiment, the window D comprised 16,000 samples.

[0012] The above described preprocessing facilitates the subsequent BSS processing to arrive at a solution in a shorter time than if the preprocessing was not provided, and furthermore, increases the probability that the BSS processing will arrive at a valid solution instead of a local minimum.

[0013] Figure 3 illustrates the principles of the operation of a BSS algorithm upon which the unmixing or separation of the desired component from the input signals is based. The technique is called Blind Source Separation because it makes few assumptions about the type of signals present in the mixture. As well known by those of ordinary skill in this technology, BSS processing is intended to recover the set of n unknown source signals from a set of their mixtures, assuming that the n source signals are independent. More specifically, as shown in Figure 3, if s is a vector of n sources, and $\mathbf{x}$ is a vector of m observations of those sources (i.e., the raw input signals from the m microphones), the goal of a BSS processor is to discover the m by n mixing matrix A:

$x = As$, where x is the preprocessed signals shown in Figure 2 (i.e., x").

or equivalently, and as is done in the present invention, to find an unmixing or separating matrix W such that

$z = Wx = \hat{s} \approx s$ where $\mathbf{z}$ is the vector of the independent estimates of component signals s and z is an

estimate of the source signals.

[0014] As previously noted, the sources $s=(s_1, s_2)$ and the environment-dependent mixing matrix A are unknown. The BSS processor (which as well known, may be implemented using a neural network) only sees the inputs $x=(x_1, x_2)$ coming from two microphones in order to determine estimates $z=(z_1, z_2)$ of the independent component signals s. In this case, the inputs x are actually the preprocessed signals $x"$, previously described.

[0015] Figure 4 illustrates a block diagram of the main components of a BSS processor 400. BSS processor 400 comprises: an unmixing component 402 for recording and updating the state of the unmixing process defined by parameters W and v; a nonlinear component 404 for generating statistics used in the adaptation process; and an adaptation component 406 for computing changes in the values of the unmixing parameters, ΔW and Δv.

[0016] As will now be described in greater detail, the BSS processor 400 continuously adapts two state variables: the 2 by 2 unmixing matrix W, and the 2 by 1 bias vector b. The unmixing component 402 buffers the most recent N samples input to BSS processor 400. It computes the output z corresponding to the most recent input sample x by using the current values of the parameters W. These parameters are initialized with small random values at the beginning of the process (while v=0):

$$z = Wx$$

[0017] The nonlinear component 404 transforms the output of the system using an *invertible* mapping. The objective of component 404 is to avoid processing very large numeric values of the outputs, which may be infinities from a computational point of view. This objective is carried out by processing statistically equivalent quantities, obtained after running the outputs z through the invertible mapping. An example of a nonlinear transformation used in component 404 is the sigmoidal nonlinearity y, defined below, taking as arguments z translated with v over the input buffer.

$$y = \frac{1}{1 + \exp(-z - v)}$$

[0018] The adaptation component 406 determines changes in the unmixing parameters W and v: i.e., ΔW and Δv. The objective is to maximize the mutual information that the outputs y contain about the inputs x, as well known to those skilled in this technology, and as described, for example by A.J. Bell and T.J. Sejnowski in their article entitled "An information-maximization approach to blind separation and blind deconvolution" published in Neural Computation, 7:1129--1159, 1995, and as also described in Bell's US patent 5,706,402. This objective reduces to a condition on the joint entropy

$H=H(y_1, y_2)$ of the outputs y:

$$\frac{\partial H(y_1, y_2)}{\partial w} = 0$$

$$\frac{\partial H(y_1, y_2)}{\partial v} = 0$$

[0019] The resulting adaptations rules are modified to perform a "natural gradient" step known by those skilled in this technology, such as described by S. Amari in his publication entitled "Minimum mutual information blind separation, published in Neural Computation, 1996.

[0020] We obtain the following update rules:

$$\Delta W = \eta(W + (1-2y) \cdot u)$$

$$\Delta v = \eta(1 - 2y)$$

[0021] A typical value for the learning rate η is 0.005.

[0022] Referring again to Figure 1, following unmixer 18 is the postprocessing step 20, wherein a determination is made of which output estimate of unmixer 18 is more likely to represent voice rather than noise, as well as a normalization of the power of the outputs by scaling them to the level of the input powers. The output signal section can be based on multiple criteria using, for example, voice specific feature extraction and analysis, and/or dominant speaker detection, which can also be accomplished using feature extraction and analysis.

[0023] As previously noted, in the illustrated embodiment of the present invention, the BSS processing is applied for use in hearing aids. The inputs to the system are given by two microphones which, with the present invention, can be situated very close to one another. In terms of the notation in the BSS processor shown in Figures 3 and 4, the system has two inputs and two ouputs ($n=m=2$).

[0024] Particularly for the case of hearing aids, the present invention addresses the following problems:

- It works with real world mixtures of signals in anechoic environments. The challenge is that a hearing aid using BSS would incorporate two microphones which, given the physical limitation imposed by in the ear hearing aids, may be less than 11 mm apart.
- It can cope with more signals than the number of microphones. Until now, this was thought to be impossible since the existing theory behind BSS guarantees that a solution exists only when $n>m$.
- It works under non-stationary mixing conditions in order to follow moving sources and adapt to changing listening environments.
- It works in real time so that the user is not subjected

to disconcerting delays in the signals and so that the hearing aid can adapt as necessary.

[0025] Thus, there has been shown and described a novel method and apparatus for real-time unmixing of a desired signal from a mixture of independent signals. Many changes, modifications, variations and other uses and applications of the subject invention will, however, become apparent to those skilled in the art after considering this specification and its accompanying drawings, which disclose a preferred embodiment thereof. For example, although pre- and post- BSS processors 16 and 18 are described, as noted herein, they are not strictly necessary in the broadest application of the present invention. Additionally, the various components of BSS processor 400 can be biased with *a priori* knowledge about the input signals to facilitate its operation, for example, knowledge about the distribution of the amplitude values of the source signals or even that one input signal represents speech. Furthermore, signal processing for enhancing source signal directionality can be incorporated into preprocessor 16. Even furthermore, the teaching of the present invention can be extremely useful for interference cancellation, separation of one voice from a mixture of many voices ("cocktail party" problem), and for preprocessing sound mixtures for noise reduction in order to allow further processing of a desired sound signal. **x**. All such changes, modifications, variations and other uses and applications which do not depart from the teachings herein are deemed to be covered by this patent, which is limited only by the claims which follow as interpreted in light of the foregoing description.

Claims

1. An electronic filtering device for performing real-time unmixing of a signal desired to be recovered by a user of the device, where the desired signal emanates from one of a plurality of independent signal sources, comprising:

   two microphones positioned along a common axis for developing first and second electrical input signals in response to reception by the microphones of acoustic signals from the plurality of independent signal sources, wherein the spatial position of the common axis of the microphones is controllable in real time by the user to align the common axis so that it substantially continuously points in the direction of the source of the desired signal; and
   an adaptive unmixing signal processor responsive to said input signals for developing output signals wherein the desired signal is separate from the mixture signal.

2. The apparatus of claim 1, wherein the common axis is positioned on the user in a manner so as to point in the direction of the source.

3. The apparatus of claim 2, wherein said microphones are mounted in a common housing that is intended to be co-located with the ear of the user.

4. The apparatus of claim 1, further including a preprocessor for modifying the input signals before they are applied to the unmixing signal processor.

5. The apparatus of claim 4, wherein the preprocessor introduces a relative delay between components of the input signals.

6. The apparatus of claim 4, wherein the preprocessor subjects the input signals to a decorrelation processing.

7. The apparatus of claim 1, further including a post-processor responsive to the output signals of the unmixing signal processor for selecting the desired signal for application to a signal reproduction device.

8. The apparatus of claim 1, wherein the unmixing signal processor comprises a blind source signal separator.

9. The apparatus of claim 8, wherein the blind source signal separator comprises a neural network for performing an unsupervised learning process that operates to maximize the joint output entropy of the output signals.

10. A method for performing real-time unmixing of a signal desired to be recovered by a user, where the desired signal emanates from one of a plurality of independent signal sources, the method comprising the following steps:

    positioning two microphones along a common axis, for developing first and second electrical input signals in response to reception by the microphones of acoustic signals from the plurality of independent signal sources, said positioning being such that the common axis of the microphones is controllable in real time by the user to align the common axis so that it substantially continuously points in the direction of the source of the desired signal; and
    subjecting said input signals to an adaptive unmixing signal processing for developing output signals wherein the desired signal is separated from the mixture signal.

11. The method of claim 10, wherein said positioning

locates the common axis proximate the user in a manner so that it points in the direction that the user is looking.

12. The method of claim 11, wherein said positioning locates the common axis on a common housing that is intended to be co-located with the ear of the user.

13. The method of claim 10, further including a preprocessing step for modifying the input signals before they are subjected to the unmixing signal processing.

14. The method of claim 10, wherein the preprocessor step introduces a relative delay between the input signals.

15. The method of claim 14, wherein the preprocessing step subjects the relatively delayed input signals to decorrelation processing.

16. The method of claim 15, wherein the decorrelation processing step is carried out by a diagonalization of a correlation matrix formed using the relatively delayed input signals.

17. The method of claim 10, further including a post-processing step responsive to the output signals of the unmixing signal processing step for selecting the desired signal for application to a signal reproduction device.

18. The method of claim 10, wherein the unmixing signal processing step comprises blind source signal separation processing.

19. The method of claim 18, wherein the blind source signal separation processing comprises an unsupervised learning process that operates to maximize the joint output entropy of the output signals.

20. A method for performing real-time unmixing of a signal desired to be recovered by a user, where the desired signal emanates from one of a plurality of independent signal sources, the method comprising the following steps:

> positioning two microphones along a common axis, for developing first and second electrical input signals in response to reception by the microphones of acoustic signals from the plurality of independent signal sources;
> preprocessing the first and second electrical input signals so as to enhance signal source directionality inherent therein due to the positioning of the microphones; and
> subjecting said directionality enhanced input signals to an adaptive unmixing signal process-

ing for developing output signals wherein the desired signal is separated from the mixture signal.

21. The method of claim 20, wherein said positioning is such that the common axis of the microphones is controllable in real time by the user to align the common axis so that it substantially continuously points in the direction of the source of the desired signal.

22. The method of claim 20, wherein said preprocessing comprises introducing a relative delay between the input signals so as to further enhance their directionality.

23. The method of claim 22, wherein said preprocessing also includes a decorrelation processing of the relatively delayed input signals.
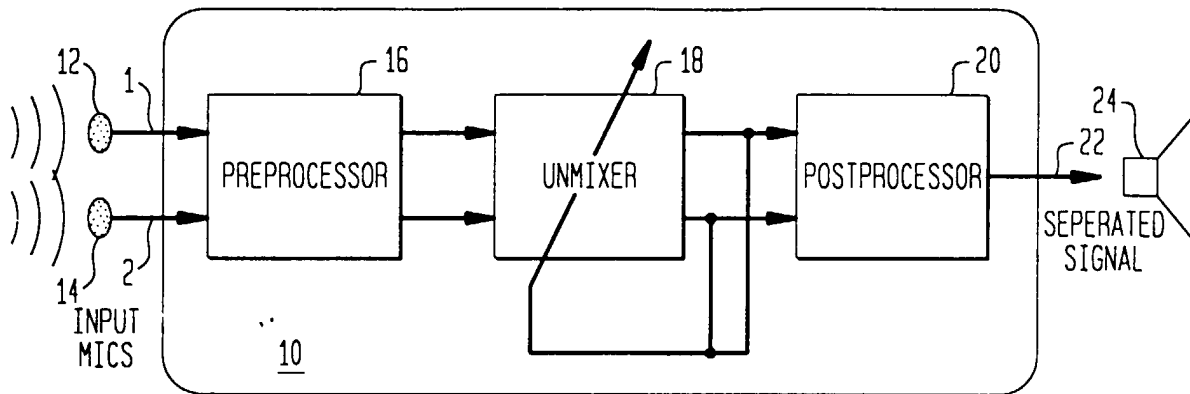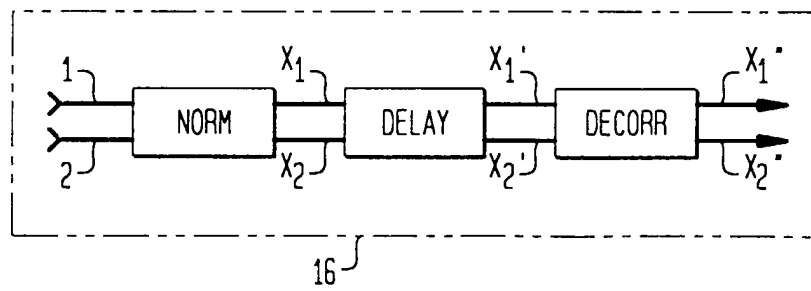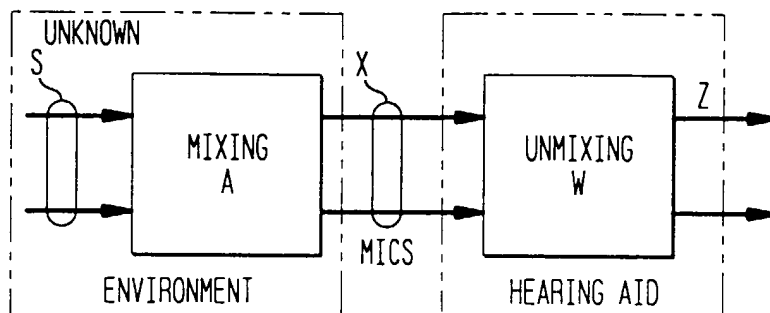
FIG. 1

FIG. 2

FIG. 3

FIG. 4