

(19) 日本国特許庁(JP)

(12) 公表特許公報(A)

(11) 特許出願公表番号

特表2004-509362

(P2004-509362A)

(43) 公表日 平成16年3月25日(2004.3.25)

(51) Int. Cl.⁷

G 1 O L 15/20

G 1 O L 21/02

F I

G 1 O L 3/02

3 O 1 E

G 1 O L 9/00

F

テーマコード (参考)

5 D O 1 5

審査請求 未請求 予備審査請求 有 (全 48 頁)

(21) 出願番号 特願2002-512971 (P2002-512971)
 (86) (22) 出願日 平成13年7月17日 (2001.7.17)
 (85) 翻訳文提出日 平成15年1月20日 (2003.1.20)
 (86) 国際出願番号 PCT/US2001/022490
 (87) 国際公開番号 W02002/007151
 (87) 国際公開日 平成14年1月24日 (2002.1.24)
 (31) 優先権主張番号 60/219,297
 (32) 優先日 平成12年7月19日 (2000.7.19)
 (33) 優先権主張国 米国 (US)
 (31) 優先権主張番号 09/905,361
 (32) 優先日 平成13年7月12日 (2001.7.12)
 (33) 優先権主張国 米国 (US)

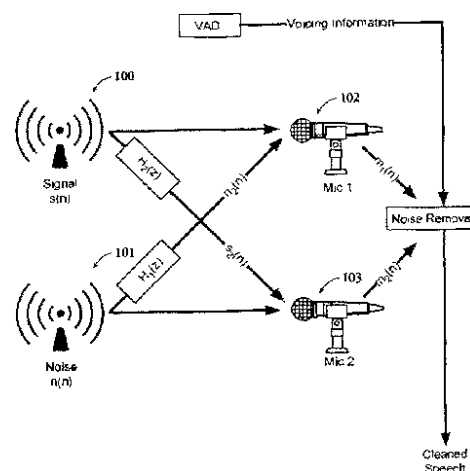
(71) 出願人 503029522
 アリフコム
 アメリカ合衆国カリフォルニア州9400
 5, プリズベン, スイート 120, マリ
 ーナ・ブールバード 8000
 (74) 代理人 100089705
 弁理士 社本 一夫
 (74) 代理人 100076691
 弁理士 増井 忠武
 (74) 代理人 100075270
 弁理士 小林 泰
 (74) 代理人 100080137
 弁理士 千葉 昭男
 (74) 代理人 100096013
 弁理士 富田 博行

最終頁に続く

(54) 【発明の名称】 電子的信号からノイズを除去するための方法および装置

(57) 【要約】

人間のスピーチからの音響ノイズの除去のための方法およびシステムを提供し、これにおいては、ノイズのタイプ、大きさあるいは方位に無関係にノイズを除去してその信号を復元する。本システムは、プロセッサに結合したマイクロホンおよびセンサを備える。マイクロホンは、音響信号を受け、そしてVADは、スピーチ（有声および無声の両方）が生起しているとき二進1であり、そしてスピーチの不在時に二進0の信号を供給する。プロセッサは、デノイズ処理アルゴリズムを備え、これは、伝達関数を発生する。これら伝達関数は、指定された時間の間受けた音響信号に発声情報が存在しないとの判定にตอบสนองして発生する伝達関数を含む。また、それら伝達関数は、指定された時間の間その音響信号に発声情報が存在するとの判定にตอบสนองして発生する伝達関数を含む。少なくとも1つのデノイズ処理した音響データ・ストリームは、それら伝達関数を使用して発生する。



【特許請求の範囲】

【請求項 1】

音響信号からノイズを除去するノイズ除去方法であって、
複数の音響信号を受けるステップと、
人間の発声活動に関連した生理学的情報を受けるステップと、
少なくとも 1 つの指定された時間の間前記複数の音響信号に発声情報が存在しないと判定したときに、前記複数の音響信号を表す少なくとも 1 つの第 1 の伝達関数を発生するステップと、
少なくとも 1 つの指定された時間の間前記複数の音響信号に発声情報が存在すると判定したときに、前記複数の音響信号を表す少なくとも 1 つの第 2 の伝達関数を発生するステップと、
前記少なくとも 1 つの第 1 の伝達関数と前記少なくとも 1 つの第 2 の伝達関数との少なくとも 1 つの組み合わせを使用することにより、前記複数の音響信号からノイズを除去して、少なくとも 1 つのデノイズ処理した音響データ・ストリームを発生するステップと、
から成るノイズ除去方法。

【請求項 2】

請求項 1 記載の方法において、前記複数の音響信号は、少なくとも 1 つの関連するノイズ・ソース信号の少なくとも 1 つの反射と、少なくとも 1 つの音響ソース信号の少なくとも 1 つの反射とを含むこと、を特徴とするノイズ除去方法。

【請求項 3】

請求項 1 記載の方法において、生理学的情報を受けるステップは、無線周波デバイス、電気グロットグラフ、超音波デバイス、音響喉当てマイクロホン、空気流検出器から成るグループから選択した少なくとも 1 つの検出器を使用することによって、人間の発声に関連する生理学的数据を受けること、を含むことを特徴とするノイズ除去方法。

【請求項 4】

請求項 1 記載の方法において、複数の音響信号を受けるステップは、複数の独立に位置付けたマイクロホンを使用して受けること、を含むことを特徴とするノイズ除去方法。

【請求項 5】

請求項 1 記載の方法において、ノイズを除去するステップは、さらに、前記少なくとも 1 つの第 1 伝達関数と前記少なくとも 1 つの第 2 伝達関数とを使用して少なくとも 1 つの第 3 の伝達関数を発生すること、を含むことを特徴とするノイズ除去方法。

【請求項 6】

請求項 1 記載の方法において、前記少なくとも 1 つの第 1 伝達関数の発生は、少なくとも 1 つの予め指定したインターバルの間において、前記少なくとも 1 つの第 1 伝達関数を再計算すること、を含むことを特徴とするノイズ除去方法。

【請求項 7】

請求項 1 記載の方法において、前記少なくとも 1 つの第 2 伝達関数の発生は、少なくとも 1 つの予め指定したインターバルの間において、前記少なくとも 1 つの第 2 伝達関数を再計算すること、を含むことを特徴とするノイズ除去方法。

【請求項 8】

請求項 1 記載の方法において、前記少なくとも 1 つの第 1 伝達関数と前記少なくとも 1 つの第 2 伝達関数の発生は、適応形技術と再帰的技術とから成るグループから選択した少なくとも 1 つの技術の使用を含むこと、を特徴とするノイズ除去方法。

【請求項 9】

電子的信号からノイズを除去するノイズ除去方法であって、
少なくとも 1 つの時間の間において発声情報の不在を検出するステップと、
前記少なくとも 1 つの時間の間において、少なくとも 1 つのノイズ・ソース信号を受けるステップと、
前記少なくとも 1 つのノイズ・ソース信号を表す少なくとも 1 つの伝達関数を発生するステップと、

音響信号とノイズ信号とを含む少なくとも1つの複合信号を受けるステップと、
前記少なくとも1つの複合信号から、前記少なくとも1つの伝達関数を使用して前記ノイズ信号を除去することにより、少なくとも1つのデノイズ処理された音響データ・ストリームを発生するステップと、
から成るノイズ除去方法。

【請求項10】

請求項9記載の方法において、前記少なくとも1つのノイズ・ソース信号は、少なくとも1つの関連するノイズ・ソース信号の少なくとも1つの反射を含むこと、を特徴とするノイズ除去方法。

【請求項11】

請求項9記載の方法において、前記少なくとも1つの複合信号は、少なくとも1つの関連する複合信号の少なくとも1つの反射を含むこと、を特徴とするノイズ除去方法。

【請求項12】

請求項9記載の方法において、検出することは、無線周波デバイス、電気グロットグラフ、超音波デバイス、音響喉当てマイクロホン、空気流検出器から成るグループから選択した少なくとも1つの検出器を使用することによって、人間の発声に関連する生理学的データを収集すること、を含むことを特徴とするノイズ除去方法。

【請求項13】

請求項9記載の方法において、受けることは、少なくとも1つのマイクロホンを使用して前記少なくとも1つのノイズ・ソース信号を受けること、を含むことを特徴とするノイズ除去方法。

【請求項14】

請求項13記載の方法において、前記少なくとも1つのマイクロホンは、複数の独立に位置付けたマイクロホンを含むこと、を特徴とするノイズ除去方法。

【請求項15】

請求項9記載の方法において、前記少なくとも1つの複合信号から前記少なくとも1つの伝達関数を使用して前記ノイズ信号を除去することは、前記少なくとも1つの伝達関数を使用して少なくとも1つの他の伝達関数を発生すること、を含むことを特徴とするノイズ除去方法。

【請求項16】

請求項9記載の方法において、少なくとも1つの伝達関数を発生することは、少なくとも1つの予め指定したインターバルの間において前記少なくとも1つの伝達関数を再計算すること、を含むことを特徴とするノイズ除去方法。

【請求項17】

請求項9記載の方法において、少なくとも1つの伝達関数を発生することは、適応形技術と再帰的技術とから成るグループから選択した少なくとも1つの技術の使用して、前記少なくとも1つの伝達関数を計算すること、を含むことを特徴とするノイズ除去方法。

【請求項18】

電子的信号からノイズを除去するノイズ除去方法であって、
有声音情報がその間において不在である少なくとも1つの無発声時間を判定するステップと

、
前記少なくとも1つの無発声時間の間において少なくとも1つのノイズ信号入力を受け、そして前記少なくとも1つのノイズ信号を表す少なくとも1つの無発声伝達関数を発生するステップと、

有声音情報がその間において存在する少なくとも1つの発声時間を判定するステップと、
前記少なくとも1つの発声時間に間において少なくとも1つの信号検知デバイスから少なくとも1つの音響信号入力を受け、そして前記少なくとも1つの音響信号を表す少なくとも1つの発声伝達関数を発生するステップと、

音響信号およびノイズ信号を含む少なくとも1つの複合信号を受けるステップと、
前記少なくとも1つの無発声伝達関数と前記少なくとも1つの発声伝達関数の少なくとも

10

20

30

40

50

1つの組み合わせを使用して、前記少なくとも1つの複合信号から前記ノイズ信号を除去することにより、少なくとも1つのデノイズ処理した音響データ・ストリームを発生するステップと、
から成るノイズ除去方法。

【請求項19】

音響信号からノイズを除去するためのノイズ除去システムであって、
少なくとも1つの音響信号を受ける少なくとも1つの受信器と、
人間の発声活動に関連する生理学的情報を受ける少なくとも1つのセンサと、
前記少なくとも1つの受信器と、複数の伝達関数を発生する前記少なくとも1つのセンサとの間に結合した少なくとも1つのプロセッサであって、前記少なくとも1つの音響信号を表す少なくとも1つの第1伝達関数は、発声情報が、少なくとも1つの指定された時間の間に前記少なくとも1つの音響信号に不在であるとの判定に応答して発生され、前記少なくとも1つの音響信号を表す少なくとも1つの第2伝達関数は、発声情報が、少なくとも1つの指定された時間の間に前記少なくとも1つの音響信号に存在しているとの判定に
10 応答して発生され、前記少なくとも1つの第1伝達関数と前記少なくとも1つの第2伝達関数との少なくとも1つの組み合わせを使用して、前記少なくとも1つの音響信号からノイズを除去することにより、少なくとも1つのデノイズ処理した音響データ・ストリームを発生する、前記の少なくとも1つのプロセッサと、
から成るノイズ除去システム。

【請求項20】

請求項19記載のシステムにおいて、前記少なくとも1つのセンサは、人間のスピーチ発生に関連した組織運動を検出する少なくとも1つの無線周波(RF)干渉計を含むこと、
を特徴とするノイズ除去システム。

【請求項21】

請求項19記載のシステムにおいて、前記少なくとも1つのセンサは、無線周波デバイス、電気グロットグラフ、超音波デバイス、音響喉当てマイクロホン、空気流検出器から成るグループから選択した少なくとも1つのセンサを含むこと、を特徴とするノイズ除去システム。

【請求項22】

請求項19記載のシステムであって、さらに、
前記少なくとも1つの音響信号の音響データを複数のサブバンドに分割し、
前記少なくとも1つの第1伝達関数と前記少なくとも1つの第2伝達関数の前記少なくとも1つの組み合わせを使用して、前記複数のサブバンドの各々からノイズを除去し、複数の
30 デノイズ処理した音響データ・ストリームが発生され、
前記複数のデノイズ処理した音響データ・ストリームを組み合わせることによって、前記少なくとも1つのデノイズ処理した音響データ・ストリームを発生すること、
を含むこと、を特徴とするノイズ除去システム。

【請求項23】

請求項19記載のシステムにおいて、前記少なくとも1つの受信器は、複数の独立に位置付けたマイクロホンを含むこと、を特徴とするノイズ除去システム。
40

【請求項24】

音響信号からノイズを除去するためのノイズ除去システムであって、少なくとも1つのマイクロホンと少なくとも1つの発声センサとの間に結合した少なくとも1つのプロセッサとを備え、前記少なくとも1つの発声センサが、発声に関連した生理学的数据を収集し、
少なくとも1つの時間の間において前記少なくとも1つの発声センサを使用して有声音情報の不在を検出し、前記少なくとも1つの時間において前記少なくとも1つのマイクロホンを使用して少なくとも1つのノイズ・ソース信号を受け、前記少なくとも1つのプロセッサが、前記少なくとも1つのノイズ・ソース信号を表す少なくとも1つの伝達関数を発生し、前記少なくとも1つのマイクロホンが、音響信号およびノイズ信号を含む少なくとも1つの複合信号を受け、前記少なくとも1つのプロセッサが、前記少なくとも1つの伝
50

達関数を使用して前記少なくとも1つの複合信号から前記ノイズ信号を除去することにより、少なくとも1つのデノイズ処理した音響データ・ストリームを発生すること、を特徴とするノイズ除去システム。

【請求項25】

少なくとも1人のユーザと少なくとも1つの電子デバイスとの間に結合した信号処理システムであって、該信号処理システムが、音響信号からノイズを除去するための少なくとも1つのデノイズ処理サブシステムを含み、該デノイズ処理サブシステムが、少なくとも1つの受信器と少なくとも1つのセンサとの間に結合した少なくとも1つのプロセッサとを備え、前記少なくとも1つの受信器を、少なくとも1つの音響信号を受けよう結合し、前記少なくとも1つのセンサを、人間の発声活動に関連した生理学的情報を受けよう結合し、前記少なくとも1つのプロセッサが、複数の伝達関数を発生し、前記少なくとも1つの音響信号を表す少なくとも1つの第1伝達関数が、発声情報が、少なくとも1つの指定された時間の間に前記少なくとも1つの音響信号に不在であるとの判定にตอบสนองして発生され、前記少なくとも1つの音響信号を表す少なくとも1つの第2伝達関数は、発声情報が、少なくとも1つの指定された時間の間に前記少なくとも1つの音響信号に存在しているとの判定にตอบสนองして発生され、前記少なくとも1つの第1伝達関数と前記少なくとも1つの第2伝達関数との少なくとも1つの組み合わせを使用して、前記少なくとも1つの音響信号からノイズを除去することにより、少なくとも1つのデノイズ処理した音響データ・ストリームを発生すること、を特徴とする信号処理システム。

10

【請求項26】

請求項25記載のシステムにおいて、前記少なくとも1つの電子デバイスは、セルラ電話、パーソナル・デジタル・アシスタント、ポータブル通信デバイス、コンピュータ、ビデオ・カメラ、デジタル・カメラ、テレマチック・システムから成るグループから選択した少なくとも1つのデバイスを含むこと、を特徴とする信号処理システム。

20

【請求項27】

実行可能な命令を含むコンピュータ読み取り可能媒体であって、前記実行可能な命令が、処理システムにおいて実行されたときに、
少なくとも1つの音響信号を受け、
人間の発声活動に関連する生理学的情報を受け、
前記少なくとも1つの音響信号を表す少なくとも1つの第1伝達関数を、発声情報が、少なくとも1つの指定された時間の間に前記少なくとも1つの音響信号に不在であるとの判定にตอบสนองして発生し、
前記少なくとも1つの音響信号を表す少なくとも1つの第2伝達関数を、発声情報が、少なくとも1つの指定された時間の間に前記少なくとも1つの音響信号に存在しているとの判定にตอบสนองして発生し、
前記少なくとも1つの第1伝達関数と前記少なくとも1つの第2伝達関数との少なくとも1つの組み合わせを使用して、前記少なくとも1つの音響信号からノイズを除去することにより、少なくとも1つのデノイズ処理した音響データ・ストリームを発生すること、
により、受けた音響信号からノイズを除去すること、を特徴とするコンピュータ読み取り可能媒体。

30

40

【請求項28】

実行可能な命令を含む電磁的媒体であって、前記実行可能な命令が、処理システムにおいて実行されたときに、
少なくとも1つの音響信号を受け、
人間の発声活動に関連する生理学的情報を受け、
前記少なくとも1つの音響信号を表す少なくとも1つの第1伝達関数を、発声情報が、少なくとも1つの指定された時間の間に前記少なくとも1つの音響信号に不在であるとの判定にตอบสนองして発生し、
前記少なくとも1つの音響信号を表す少なくとも1つの第2伝達関数を、発声情報が、少なくとも1つの指定された時間の間に前記少なくとも1つの音響信号に存在しているとの

50

判定に応答して発生し、

前記少なくとも1つの第1伝達関数と前記少なくとも1つの第2伝達関数との少なくとも1つの組み合わせを使用して、前記少なくとも1つの音響信号からノイズを除去することにより、少なくとも1つのデノイズ処理した音響データ・ストリームを発生すること、により、受けた音響信号からノイズを除去すること、を特徴とする実行可能な命令を含む電磁的媒体。

【発明の詳細な説明】

【0001】

発明の分野

本発明は、音響的な伝送または録音から望ましくない音響ノイズを除去あるいは抑制するための数学的方法並びに電子的システムに関するものである。 10

【0002】

背景

代表的な音響用途においては、人間のユーザからのスピーチは、録音されるかあるいは格納されそして様々な場所にいる受け手に送信される。そのユーザの環境においては、問題とする信号（ユーザのスピーチ）を、望まない音響ノイズで汚染する1つまたはこれより多いノイズ・ソースが存在することがある。これは、受け手が人であろうと機械であろうと、その受け手がユーザのスピーチを理解するのを困難にしたりあるいは不可能にしたりする。このことは、特に、セルラ電話およびパーソナル・デジタル・アシスタントのようなポータブル通信デバイスの普及と共に、現在特に問題となっている。これらノイズ付加物を抑制する既存の方法があるが、これらは、いずれも、長過ぎる計算時間あるいは嵩張るハードウェアを必要としたり、問題の信号を歪ませ過ぎたり、あるいは有用な性能に欠けたりするものである。これら方法の多くは、ヴァセギ（Vaseghi）による“先進のデジタル信号処理およびノイズ低減（Advanced Digital Signal Processing and Noise Reduction）”（ISBN 0-471-62692-9）の教本に記載されている。結果として、代表的なシステムの上記欠点に対処しそして歪みなしで問題の音響信号をクリーンにする新たな技術を提供するノイズ除去および低減法に対するニーズがあることになる。 20

【0003】

摘要

人間のスピーチからの音響ノイズの除去のための方法およびシステムを提供し、これにおいては、ノイズ・タイプ、大きさあるいは方位に無関係にノイズを除去してその信号を復元する。本システムは、プロセッサに結合したマイクロホンおよびセンサを備える。マイクロホンは、ノイズと人間の信号ソースからのスピーチ信号の両方を含む音響信号を受け取る。センサは、二進の発声活動検出（VAD）を発生し、これは、スピーチ（有声および無声の両方）が生起しているとき二進“1”であり、そしてスピーチが生起していないとき二進“0”の信号を供給する。このVAD信号は、種々の方法、例えば音響利得、加速度計、および無線周波（RF）センサを使用して得ることができる。 30

【0004】

このプロセッサ・システムおよび方法は、デノイズ処理アルゴリズム（denoising algorithm）を備え、これは、ノイズ・ソースとマイクロホンとの間の伝達関数、並びに人間のユーザとマイクロホンとの間の伝達関数を計算する。これら伝達関数を使用することにより、受けた音響信号からノイズを除去して、少なくとも1つのデノイズ処理した音響データ・ストリームを発生する。 40

【0005】

詳細説明

図1は、1実施形態のデノイズ処理システム（denoising system）のブロック図であって、このシステムは、発声活動に関する生理学的情報から得た、いつスピーチが生起しているかについての知識を使用する。本システムは、複数のマイクロホン10と複数のセンサ20とを備え、そしてこれらは、少なくとも1つのプロセッサ30へ信 50

号を供給する。プロセッサは、デノイズ処理を行うサブシステムまたはアルゴリズムを備えている。

【0006】

図2は、単一のノイズ・ソースとマイクロホンへの直接経路を仮定したときの、1実施形態のノイズ除去システム／アルゴリズムのブロック図である。このノイズ除去システム図は、1実施形態のこのプロセスの図式記述を含み、単一の信号ソース（100）と、単一のノイズ・ソース（101）とがある。このアルゴリズムは、2つのマイクロホン、すなわち“信号”マイクロホン（MIC1, 102）と、“ノイズ”マイクロホン（MIC2, 103）とを使用する（但し、これに限定されるものではない）。MIC1は、大部分の信号といくらかのノイズを捕獲する一方で、MIC2は、大部分のノイズといくらかの信号とを捕獲すると仮定する。これは、従来の先進の音響システムと共通の構成である。MIC1への信号からのデータは $s(n)$ で示し、MIC2への信号からのデータは $s_2(n)$ で示し、MIC2へのノイズからのデータは $n(n)$ で示し、MIC1へのノイズからのデータは $n_2(n)$ で示している。同様に、MIC1からのデータは $m_1(n)$ で、そしてMIC2からのデータは $m_2(n)$ で示し、ここで、 $s(n)$ はソースからのアナログ信号の離散的なサンプルを示している。

【0007】

信号からMIC1への伝達関数およびノイズからMIC2への伝達関数は、1であると仮定するが、信号からMIC2への伝達関数は $H_2(z)$ 、ノイズからMIC1への伝達関数は $H_1(z)$ で示す。1の伝達関数のこの仮定は、このアルゴリズムの一般性を妨げるものでなく、その理由は、信号とノイズとマイクロホンとの実際の関係が単に比率であり、そしてこの比率は、簡単のためこのようにして再定義されるからである。

【0008】

従来のノイズ除去システムにおいては、MIC2からの情報は、MIC1からのノイズを除去しようとする試みにおいて使用されている。しかし、語られていない仮定は、発声活動検出（VAD（Voice Activity Detection））が決して完全ではないことであり、したがってそのデノイズ処理は、ノイズと一緒に信号をもかなり除去してしまうことのないよう、注意深く実行しなければならない。しかし、このVADが完全であり、そしてこれが、スピーチが全くユーザによって発されていないときにゼロに等しく、そしてスピーチが発生されているときに1に等しいと仮定すると、このノイズ除去においてかなりの改善を行うことができる。

【0009】

マイクロホンへの単一のノイズ・ソースおよび直接経路の分析においては、図2において、MIC1へ入来する音響情報は、 $m_1(n)$ で示される。MIC2へ入来する情報は、同様に $m_2(n)$ で示される。 z （デジタル周波数）ドメインにおいては、これらは、 $M_1(z)$ および $M_2(z)$ として表される。このとき、

【0010】

【数1】

$$M_1(z) = S(z) + N_2(z)$$

$$M_2(z) = N(z) + S_2(z)$$

【0011】

ここで、

【0012】

【数2】

$$N_2(z) = N(z)H_1(z)$$

$$S_2(z) = S(z)H_2(z)$$

【0013】

したがって、以下となる。

10

20

30

40

50

【 0 0 1 4 】

【 数 3 】

$$\begin{aligned} M_1(z) &= S(z) + N(z)H_1(z) \\ M_2(z) &= N(z) + S(z)H_2(z) \end{aligned} \quad \text{Eq. 1}$$

【 0 0 1 5 】

これは、2 個のマイクロホン・システムの全てに対する一般的なケースである。実際のシステムでは、常に、M I C 1 へのノイズにいくらかの漏れと、M I C 2 への信号にいくらかの漏れとがある。式 1 には、未知数が 4 つで、既知の関係が 2 つしかないため、これを明快に解くことはできない。 10

【 0 0 1 6 】

しかし、式 1 中の未知数のいくつかに対し解を与える別の方法がある。この分析は、信号が発生されていないケース、すなわち V A D 信号がゼロに等しくかつスピーチが発生されていない場合を調べることから始まる。このケースでは、 $s(n) = S(z) = 0$ であり、したがって式 1 は、以下となる。

【 0 0 1 7 】

【 数 4 】

$$\begin{aligned} M_{1n}(z) &= N(z)H_1(z) \\ M_{2n}(z) &= N(z) \end{aligned} \quad 20$$

【 0 0 1 8 】

ここで、変数 M の下付文字 n は、ノイズのみを受けていることを示す。これにより、以下となる。

【 0 0 1 9 】

【 数 5 】

$$\begin{aligned} M_{1n}(z) &= M_{2n}(z)H_1(z) \\ H_1(z) &= \frac{M_{1n}(z)}{M_{2n}(z)} \quad \text{Eq. 2} \end{aligned} \quad 30$$

【 0 0 2 0 】

$H_1(z)$ は、本システムがノイズのみを受信していることが確かであれば、利用可能なシステム識別アルゴリズムおよびマイクロホン出力の任意のものを使用して計算することができる。この計算は、適応的に行うことができ、これにより、本システムは、そのノイズにおける変化に反応することができる。

【 0 0 2 1 】

式 1 中の未知数のうちの 1 つに対し、解がこれで入手可能である。別の未知数、すなわち $H_2(z)$ は、V A D が 1 に等しくしかもスピーチが発生されている場合を使用することにより、決定することができる。その場合が発生しているが、マイクロホンの最近（おそらく、1 秒未満）の履歴が低レベルのノイズを示している場合、 $n(s) = N(z) \sim 0$ とみなすことができる。このとき、式 1 は、以下となる。 40

【 0 0 2 2 】

【 数 6 】

$$\begin{aligned} M_{1s}(z) &= S(z) \\ M_{2s}(z) &= S(z)H_2(z) \end{aligned}$$

【 0 0 2 3 】

これはさらに、以下となる。

【 0 0 2 4 】

【 数 7 】

$$M_{2s}(z) = M_{1s}(z)H_2(z)$$

$$H_2(z) = \frac{M_{2s}(z)}{M_{1s}(z)}$$

【 0 0 2 5 】

これは、 $H_1(z)$ の逆である。しかし、分かるように、異なった入力を使用している、すなわち、これでは、信号のみが発生しており、そしてこれに対し、以前では、ノイズのみが発生していた。 $H_2(z)$ の計算の間、 $H_1(z)$ に対し計算した値は、一定に保持し、そして逆の場合もそうである。したがって、 $H_1(z)$ および $H_2(z)$ は、その他方を計算している間は実質上変化しない、と仮定する。 10

【 0 0 2 6 】

$H_1(z)$ および $H_2(z)$ を計算した後、これらは、信号からノイズを除去するのに使用する。もし、式 1 を以下のように書き直すと、

【 0 0 2 7 】

【 数 8 】

$$S(z) = M_1(z) - N(z)H_1(z)$$

$$N(z) = M_2(z) - S(z)H_2(z)$$

$$S(z) = M_1(z) - [M_2(z) - S(z)H_2(z)]H_1(z) ,$$

$$S(z)[1 - H_2(z)H_1(z)] = M_1(z) - M_2(z)H_1(z)$$

20

【 0 0 2 8 】

このとき、 $N(z)$ は、 $S(z)$ を解くために、示したように以下のように置換することができる。

【 0 0 2 9 】

【 数 9 】

$$S(z) = \frac{M_1(z) - M_2(z)H_1(z)}{1 - H_2(z)H_1(z)}$$

Eq. 3

30

【 0 0 3 0 】

もし、伝達関数 $H_1(z)$ および $H_2(z)$ が十分な正確さで記述することができる場合、このときには、ノイズを完全に除去することができ、そして元の信号を復元することができる。このことは、ノイズの大きさまたはスペクトル特性に無関係に当てはまる。行った仮定は、完全な VAD、十分に正確な $H_1(z)$ および $H_2(z)$ と、 $H_1(z)$ および $H_2(z)$ がその他方を計算しているときに実質上変化しない、ということのみである。実際、これら仮定は、妥当なものであると判明した。 40

【 0 0 3 1 】

記述したこのノイズ除去アルゴリズムは、任意の数のノイズ・ソースを含むように容易に一般化できる。図 3 は、 n 個の区別できるノイズ・ソースに対し一般化した、1 実施形態のノイズ除去アルゴリズムのフロントエンドのブロック図である。これら区別できるノイズ・ソースは、互いの他の反射またはエコーであることができる（但し、これに限定されるものではない）。図示したように、いくつかのノイズ・ソースがあり、その各々は、各マイクロホンに対し伝達関数または経路を有している。先に名称を付与した経路 H_2 は、 H_0 としてラベルを付与しており、これにより、MIC 1 へのノイズ・ソース 2 の経路にラベルを付与することは、より都合よくなる。各マイクロホンの出力は、 z ドメイ 50

ンに変換したときには、以下となる。

【 0 0 3 2 】

【 数 1 0 】

$$\begin{aligned} M_1(z) &= S(z) + N_1(z)H_1(z) + N_2(z)H_2(z) + \dots N_n(z)H_n(z) \\ M_2(z) &= S(z)H_0(z) + N_1(z)G_1(z) + N_2(z)G_2(z) + \dots N_n(z)G_n(z) \end{aligned} \quad \text{Eq. 4}$$

【 0 0 3 3 】

信号が全くないとき（VAD = 0）、このとき（簡単のため z を抑制する）、以下となる。

【 0 0 3 4 】

【 数 1 1 】

$$\begin{aligned} M_{1n} &= N_1H_1 + N_2H_2 + \dots N_nH_n \\ M_{2n} &= N_1G_1 + N_2G_2 + \dots N_nG_n \end{aligned} \quad \text{Eq. 5}$$

【 0 0 3 5 】

これにより、上記の $H_1(z)$ と同じように、新たな伝達関数を定義することができる。

【 0 0 3 6 】

【 数 1 2 】

$$\tilde{H}_1 = \frac{M_{1n}}{M_{2n}} = \frac{N_1H_1 + N_2H_2 + \dots N_nH_n}{N_1G_1 + N_2G_2 + \dots N_nG_n} \quad \text{Eq. 6}$$

【 0 0 3 7 】

したがって、 $\tilde{H}_1$ は、ノイズ・ソースとそれらの各々の伝達関数にのみ依存し、したがって伝送されている信号がないどのような時にも計算することができる。もう一度繰り返すと、マイクロホン入力の下付文字 n は、ノイズが検出されていることのみを示し、その一方で、下付文字 s は、信号のみをマイクロホンが受信していることを示している。

【 0 0 3 8 】

ノイズが全く発生されていないと仮定している間において、式 4 を調べると、以下となる。

【 0 0 3 9 】

【 数 1 3 】

$$\begin{aligned} M_{1s} &= S \\ M_{2s} &= SH_0 \end{aligned}$$

【 0 0 4 0 】

したがって、 H_0 は、任意の利用可能な伝達関数計算アルゴリズムを使って、前と同じように解くことができる。数学的には、以下となる。

【 0 0 4 1 】

【 数 1 4 】

$$H_0 = \frac{M_{2s}}{M_{1s}} \quad 40$$

【 0 0 4 2 】

式 6 で定義した $\tilde{H}_1$ を使って、式 4 を書き直すと、以下となる。

【 0 0 4 3 】

【 数 1 5 】

$$\tilde{H}_1 = \frac{M_1 - S}{M_2 - SH_0} \quad \text{Eq. 7}$$

10

20

30

40

50

【 0 0 4 4 】

S に関し解くと、以下となる。

【 0 0 4 5 】

【 数 1 6 】

$$S = \frac{M_1 - M_2 \tilde{H}_1}{1 - H_0 \tilde{H}_1} \quad \text{Eq. 8}$$

【 0 0 4 6 】

これは、式 3 と同じとなり、ここで、 H_0 が H_2 に取って代わり、 H_{-1} が H_1 に取って代わっている。したがって、このノイズ除去アルゴリズムは、依然として、ノイズ・ソースの多数のエコーを含む任意の数のノイズ・ソースに対し、数学的に有効である。再び、 H_0 と H_{-1} を十分に高い正確さで推定することができ、そして信号からマイクロホンに対する経路が 1 つのみであるという仮定が保たれる場合、ノイズは、完全に除去することができる。

【 0 0 4 7 】

最も一般的なケースは、多数のノイズ・ソースと多数の信号ソースが関係する場合である。図 4 は、 n 個の区別できるノイズ・ソースおよび信号反射とがある最も一般的なケースにおける、1 実施形態のノイズ除去アルゴリズムのフロントエンドのブロック図である。ここで、信号の反射は、両方のマイクロホンに入る。これは、最も一般的なケースであるが、それは、マイクロホンへのノイズ・ソースの反射が、単純な追加のノイズ・ソースとして正確にモデル化できるからである。簡単のため、信号から MIC 2 への直接経路は、 $H_0(z)$ から $H_{00}(z)$ に変えてあり、そしてマイクロホン 1 および 2 へのその反射経路は、それぞれ、 $H_{01}(z)$ および $H_{02}(z)$ として示している。

【 0 0 4 8 】

これにより、マイクロホンへの入力は、以下となる。

【 0 0 4 9 】

【 数 1 7 】

$$\begin{aligned} M_1(z) &= S(z) + S(z)H_{01}(z) + N_1(z)H_1(z) + N_2(z)H_2(z) + \dots N_n(z)H_n(z) \\ M_2(z) &= S(z)H_{00}(z) + S(z)H_{02}(z) + N_1(z)G_1(z) + N_2(z)G_2(z) + \dots N_n(z)G_n(z) \end{aligned} \quad \text{Eq. 9}$$

【 0 0 5 0 】

VAD = 0 のとき、それら入力は、以下となる（ z を再び抑制する）。

【 0 0 5 1 】

【 数 1 8 】

$$\begin{aligned} M_{1n} &= N_1 H_1 + N_2 H_2 + \dots N_n H_n \\ M_{2n} &= N_1 G_1 + N_2 G_2 + \dots N_n G_n \end{aligned}$$

【 0 0 5 2 】

これは、式 5 と同じである。したがって、式 6 における H_{-1} の計算は、予期した通り、変化しない。ノイズがないこの状況を検討すると、式 9 は、以下となる。

【 0 0 5 3 】

【 数 1 9 】

$$\begin{aligned} M_{1s} &= S + SH_{01} \\ M_{2s} &= SH_{00} + SH_{02} \end{aligned}$$

【 0 0 5 4 】

これは、 H_{-2} の定義となる。

【 0 0 5 5 】

【数 2 0】

$$\tilde{H}_2 = \frac{M_{2s}}{M_{1s}} = \frac{H_{00} + H_{02}}{1 + H_{01}} \quad \text{Eq. 10}$$

【0 0 5 6】

再び、(式 7 におけるのと同じように) H_{-1} に対する定義を使用して式 9 を書き直すと、以下となる。

【0 0 5 7】

【数 2 1】

$$\tilde{H}_1 = \frac{M_1 - S(1 + H_{01})}{M_2 - S(H_{00} + H_{02})} \quad \text{Eq. 11}$$

10

【0 0 5 8】

いくつかの代数的操作により、以下のようになる。

【0 0 5 9】

【数 2 2】

$$\begin{aligned} S(1 + H_{01} - \tilde{H}_1(H_{00} + H_{02})) &= M_1 - M_2 \tilde{H}_1 \\ S(1 + H_{01}) \left[1 - \tilde{H}_1 \frac{(H_{00} + H_{02})}{(1 + H_{01})} \right] &= M_1 - M_2 \tilde{H}_1 \\ S(1 + H_{01}) [1 - \tilde{H}_1 \tilde{H}_2] &= M_1 - M_2 \tilde{H}_1 \end{aligned}$$

20

【0 0 6 0】

最後には、以下となる。

【0 0 6 1】

【数 2 3】

$$S(1 + H_{01}) = \frac{M_1 - M_2 \tilde{H}_1}{1 - \tilde{H}_1 \tilde{H}_2} \quad \text{Eq. 12}$$

30

【0 0 6 2】

式 12 は、式 8 と同じであるが、但し、 H_0 が H_{-2} で置き換わっており、また、 $(1 + H_{01})$ の要素が左辺に追加されている。この余分な要素は、 S がこの状況では直接解くことができないということを意味しているが、解は、信号にそのエコーの全ての追加に対し生成することができる。このことは、それほど悪い状況ではないが、それは、エコー抑制を取り扱う多くの従来の方法があるからであり、そしてこれらエコーが抑制されない場合でも、それらがスピーチの理解度に意味のある程度にまで影響を与えることは起きそうにない。 H_{-2} のより複雑な計算は、マイクロホン 2 における信号エコーを考慮する必要がある。

【0 0 6 3】

40

図 5 は、1 実施形態のデノイズ処理方法のフロー図である。動作を説明すると、音響信号を受ける (502)。さらに、人の発声活動に関連する生理学的情報を受ける (504)。音響信号からの発声情報が少なくとも 1 つの指定した時間の間存在しないと判定したときに、音響信号を表す第 1 の伝達関数を計算する (506)。音響信号を表す第 2 の伝達関数は、この音響信号において発声情報が少なくとも 1 つの指定した時間の間存在すると判定したときに、計算する (508)。この音響信号からのノイズの除去は、第 1 伝達関数と第 2 伝達関数の少なくとも 1 つの組み合わせを使用して行い、これによりデノイズ処理した音響データ・ストリームを発生する (510)。

【0 0 6 4】

ノイズ除去のためのアルゴリズム、すなわちデノイズ処理アルゴリズムは、直接経路をも

50

つ単一のノイズ・ソースの最も単純なケースから、反射およびエコーをもつ多数のノイズ・ソースまでここに記述した。このアルゴリズムは、どのような環境条件下においても実行可能であることを示した。ノイズのタイプおよび量は、 H_1 および H_2 について良好な推定を行った場合で、しかもそれらが他方の計算中に実質上変化しない場合には、重要ではない。ユーザ環境が、エコーが存在するようなものである場合、それらは、ノイズ・ソースから来たものである場合には、補償を行うことができる。もし、信号エコーも存在する場合、それらは、クリーンにした信号に影響を与えるが、その影響は、ほとんどの環境においては、無視できる程度のものである。

【0065】

動作について説明すると、1実施形態のアルゴリズムは、様々なノイズのタイプ、大きさ、方位の取り扱いにおいて、優れた結果を示した。しかし、数学的概念からエンジニアリング用途へ移行するときには、常に近似および調節を行わなければならない。式3では、1つの仮定を行っており、これでは、 $H_2(z)$ は小さく、したがって $H_2(z)H_1(z) \approx 0$ と仮定し、このため、式3は、以下になる。

【0066】

【数24】

$$S(z) \approx M_1(z) - M_2(z)H_1(z).$$

【0067】

このことは、 $H_1(z)$ のみ計算しなければならないことを意味し、これにより、本プロセスをスピードアップし、そして必要な計算数をかなり減少させる。マイクロホンを選択すれば、この近似は容易に実現することができる。

【0068】

もう1つの近似は、1実施形態において使用するフィルタに関係する。実際の $H_1(z)$ は、疑いなく、極とゼロの両方を有することになるが、しかし、安定性および簡単さのためには、全ゼロの有限インパルス応答 (FIR) フィルタを使用する。十分なタップ (およそ60) では、実際の $H_1(z)$ に対する近似は、非常に良好となる。

【0069】

サブバンド選択に関しては、伝達関数を計算しなければならない周波数範囲が広がるにつれて、それを正確に計算することがより難しくなる。したがって、音響データは、16個のサブバンドに分割し、そして最も低い周波数を50Hz、最も高いものを3700とした。次に、本デノイズ処理アルゴリズムを各サブバンドに順番に適用し、そしてこの16個のデノイズ処理したデータ・ストリームを組み合わせることによって、デノイズ処理した音響データを発生した。これは、非常にうまく機能するが、サブバンドのどのような組み合わせ (すなわち、4, 6, 8, 32個の等しく離間させ、知覚できる程離間させたもの) も使用でき、そしてこれは、同様に機能することが分かった。

【0070】

ノイズの大きさは、1実施形態においては抑制することにより、使用したマイクロホンが飽和 (すなわち、線形応答領域外での動作) しないようにした。重要なことは、マイクロホンが線形に動作することによって、最良の性能を確保することである。この抑制を伴う場合でも、非常に高い信号対雑音比 (SNR) のテストを行うことができた (約-10dBまで)。

【0071】

$H_1(z)$ の計算は、最小二乗平均法 (LMS)、一般的な適応性伝達関数を使用して、10ミリ秒毎に実行した。この説明は、Prentice-Hall発行のWidrowおよびStearnsによる“適応性信号処理 (Adaptive Signal Processing)” (1985), ISBN 0-13-004029-0に見ることができる。

【0072】

1実施形態に対するVADは、無線周波数センサおよび2つのマイクロホンから得て、こ

れにより、有声のスピーチおよび無声のスピーチの両方に対し、非常に高い正確さ（＞99％）を発生した。1実施形態に対するこのVADは、無線周波数（RF）干渉計を使用して、人のスピーチ発生に関係する組織運動を検出する（但しこれに限定されるものではない）。したがって、これは、完全に音響ノイズなしであり、このため、どのような音響ノイズ環境においても機能することができる。簡単なエネルギー測定を使用することにより、有声スピーチが生起しているかどうかを判定することができる。無声スピーチは、有声部分への近さにより、または上記の組み合わせによって、従来の周波数ベースの方法を使用して判定することができる。無声スピーチには、それほど多くのエネルギーがないため、その活性化の正確さは、有声スピーチ程には重要でない。

【0073】

10

有声スピーチおよび無声スピーチを信頼性良く検出することにより、1実施形態のアルゴリズムを実現することができる。ここで再び、ノイズ除去アルゴリズムは、VADを得る方法に依存しないこと、これは有声スピーチに対しては特に正確であることのみを繰り返すことは有益である。もしスピーチが検出されずしかもトレーニングがそのスピーチに対して起きる場合、その後続のデノイズ処理された音響データは、歪むことがある。

【0074】

データは、4つのチャンネルで、すなわち、MIC1に対して1つ、MIC2に対して1つ、有声スピーチに関連する組織運動を検出する無線周波数センサに対して2つで、収集した。このデータは、40KHzで同時にサンプリングし、そして次に、デジタル的にフィルタしそして8KHzにデシメートした。この高いサンプリング・レートを使用することによって、このアナログ・デジタル・プロセスから生じることのあるどのようなエイリアシングも低減するようにした。4チャンネルのナショナル・インスツルメンツのA/Dボード（National Instruments A/D board）を、Labviewと共に使用して、上記データをキャプチャし格納した。このデータは、次にCプログラムに読み込み、そして一時に10ミリ秒デノイズ処理した。

20

【0075】

図6は、空港ターミナルのノイズ（他の多くの話しをする人および公共のアナウンスを含む）の存在下でのアメリカの英語を話す女性に対しての、1実施形態のノイズ抑制アルゴリズムの結果を示している。この話し手は、中位の空港ターミナル・ノイズの真っ只中で、番号406-5562を発している。汚れた音響データは、一時に10ミリ秒デノイズ処理し、そしてデータの10ミリ秒のデノイズ処理の前に50～3700Hzにプレフィルタ処理した。およそ17dBのノイズ低減が明らかとなった。このサンプルには、ポストフィルタ処理は行わなかったため、実現したノイズ低減は全て、1実施形態のこのアルゴリズムに起因するものである。本アルゴリズムは、ノイズに瞬時に適応し、したがって他の人の話者の非常に困難なノイズを除去する能力がある。多くの異なったタイプのノイズ（ほんのいくつかを挙げると、ストリートのノイズ、ヘリコプター、音楽、正弦波）は、その全てをテストしたが、同様の結果となった。また、ノイズの方位は、ノイズ抑制性能を有意に変化させずとも、実質的に変化させることができる。最後に、クリーンにしたスピーチの歪みは、非常に低く、スピーチ認識エンジン並びに人間の受け手に対しても同様に、良好な性能を確保する。

30

40

【0076】

1実施形態のノイズ除去アルゴリズムは、どのような環境条件の下でも実行可能であることを示した。ノイズのタイプおよび量は、 H^{-1} および H^{-2} について良好な推定が行われた場合には、取るに足りない。もしユーザ環境が、エコーが存在するようなものである場合、これらがノイズ・ソースから来たものである場合にはそれを補償することができる。もし信号エコーも存在する場合、これらは、クリーンにした信号に影響を与えるが、その影響は、ほとんどの環境においては無視できるものである。

【0077】

各種の実施形態について、図面を参照して説明したが、詳細な説明および図面は、限定を意図するものではない。記述した要素の種々の組み合わせについて示さなかったが、これ

50

らは、冒頭の特許請求の範囲に記載の本発明の範囲内にあるものである。

【図面の簡単な説明】

【図 1】

図 1 は、1 実施形態のデノイズ・システムのブロック図。

【図 2】

図 2 は、単一のノイズ・ソースとマイクロホンへの直接経路を想定したときの、1 実施形態のノイズ除去アルゴリズムのブロック図。

【図 3】

図 3 は、 n 個の区別できるノイズ・ソース（これらノイズ・ソースは、互いに他のものの反射またはエコーであることもある）に一般化した、1 実施形態のノイズ除去アルゴリズムのフロントエンドのブロック図。 10

【図 4】

図 4 は、 n 個の区別できるノイズ・ソースと信号反射とがある最も一般的なケースにおける、1 実施形態のノイズ除去アルゴリズムのフロントエンドのブロック図。

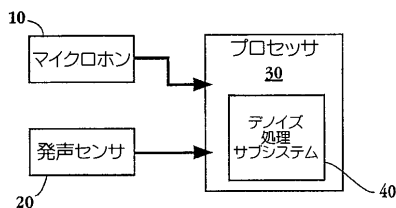
【図 5】

図 5 は、1 実施形態のデノイズ方法のフロー図。

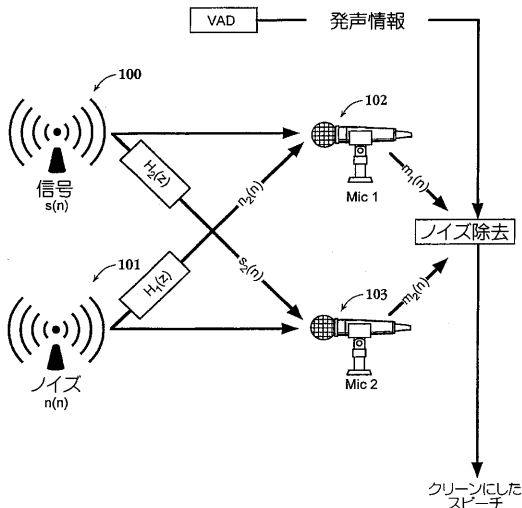
【図 6】

図 6 は、空港ターミナルのノイズ（他の多くの話しをする人および公共のアナウンスを含む）の存在下でのアメリカの英語を話す女性に対しての、1 実施形態のノイズ抑制アルゴリズムの結果を示す。 20

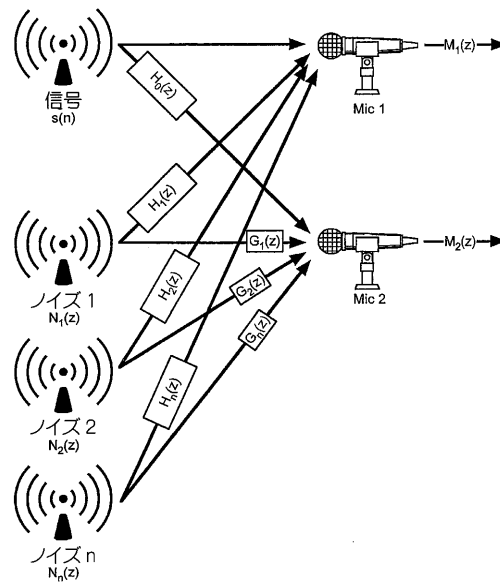
【図 1】



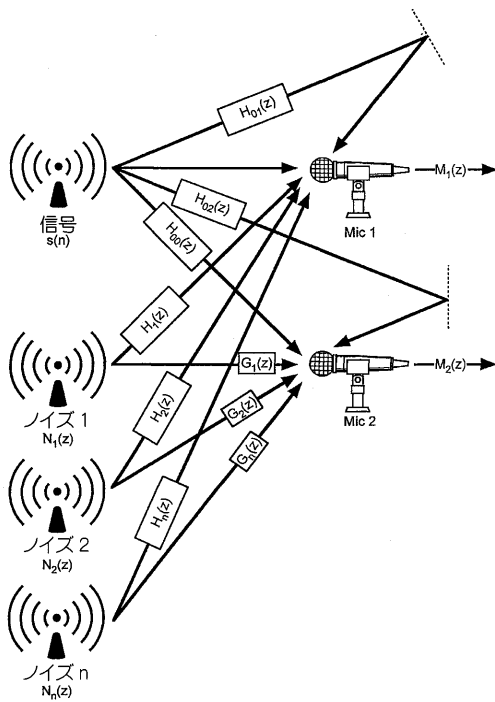
【図 2】



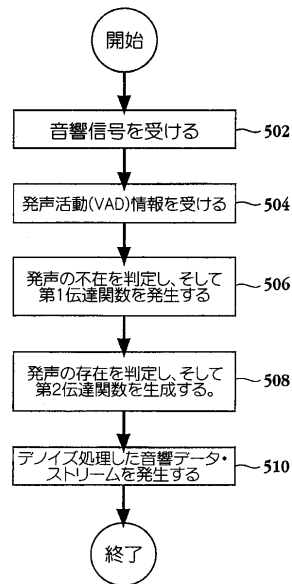
【図 3】



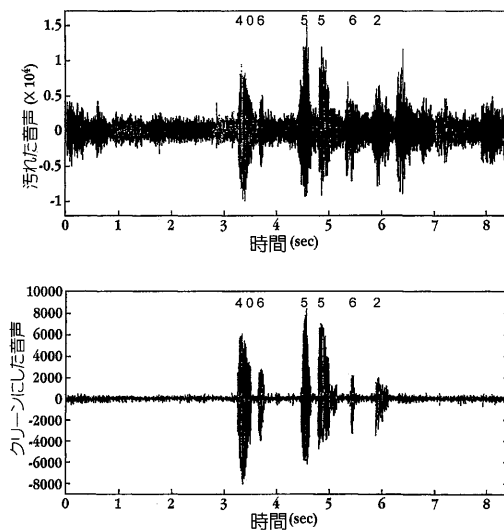
【図 4】



【図 5】



【図 6】



【国際公開パンフレット】

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau(43) International Publication Date
24 January 2002 (24.01.2002)

PCT

(10) International Publication Number
WO 02/07151 A2

- (51) International Patent Classification: **G10L 21/00** 94550 (US). BREITFELLER, Eric, F. [US/US]; 11526 Streambed Place, Dublin, CA 94568-2294 (US).
- (21) International Application Number: PCT/US01/22490
- (22) International Filing Date: 17 July 2001 (17.07.2001) (74) Agent: PERKINS COIE, LLP; P.O. Box 2168, Menlo Park, CA 94026 (US).
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
60/219,297 19 July 2000 (19.07.2000) US
09/905,361 12 July 2001 (12.07.2001) US
- (63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:
US 60/219,297 (CIP)
Filed on 19 July 2000 (19.07.2000)
US 09/905,361 (CIP)
Filed on 12 July 2001 (12.07.2001)
- (81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.
- (84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BI, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).
- (71) Applicant (for all designated States except US): ALIPH-COM [US/US]; 2nd floor, 54 Mint Street, San Francisco, CA 94103 (US). Published:
— without international search report and to be republished upon receipt of that report
- (72) Inventors; and
(75) Inventors/Applicants (for US only): BURNETT, Gregory, C. [US/US]; 637 South H Street, Livermore, CA
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.



WO 02/07151 A2

(54) Title: METHOD AND APPARATUS FOR REMOVING NOISE FROM ELECTRONIC SIGNALS

(57) Abstract: A method and system are provided for acoustic noise removal from human speech, wherein noise is removed without respect to noise type, amplitude, or orientation. The system includes microphones and a voice activity detection (VAD) data stream coupled among a processor. The microphones receive acoustic signals and the VAD produces a signal including a binary one when speech (voiced and unvoiced) is occurring and a binary zero in the absence of speech. The processor includes denoising algorithms that generate transfer functions. The transfer functions include a transfer function generated in response to a determination that voicing information is absent from the received acoustic signal during a specified time period. The transfer functions also include transfer functions generated in response to a determination that voicing information is present in the acoustic signal during a specified time period. At least one denoised acoustic data stream is generated using the transfer functions.

WO 02/07151

PCT/US01/22490

5 **METHOD AND APPARATUS FOR REMOVING NOISE FROM**
 ELECTRONIC SIGNALS

FIELD OF THE INVENTION

10 The invention is in the field of mathematical methods and electronic
systems for removing or suppressing undesired acoustical noise from acoustic
transmissions or recordings.

BACKGROUND

15 In a typical acoustic application, speech from a human user is recorded
or stored and transmitted to a receiver in a different location. In the
environment of the user, there may exist one or more noise sources that pollute
the signal of interest (the user's speech) with unwanted acoustic noise. This
makes it difficult or impossible for the receiver, whether human or machine, to
understand the user's speech. This is especially problematic now with the
20 proliferation of portable communication devices like cellular telephones and
personal digital assistants. There are existing methods for suppressing these
noise additions, but they either require far too much computing time or
cumbersome hardware, distort the signal of interest too much, or lack in
performance to be useful. Many of these methods are described in textbooks
25 such as "Advanced Digital Signal Processing and Noise Reduction" by Vaseghi,
ISBN 0-471-62692-9. Consequently, there is a need for noise removal and
reduction methods that address the shortcomings of typical systems and offer
new techniques for cleaning acoustic signals of interest without distortion.

WO 02/07151

PCT/US01/22490

SUMMARY

A method and system are provided for acoustic noise removal from human speech, wherein the noise can be removed and the signal restored without respect to noise type, amplitude, or orientation. The system includes microphones and sensors coupled with a processor. The microphones receive acoustic signals including both noise and speech signals from human signal sources. The sensors yield a binary Voice Activity Detection (VAD) signal that provides a signal that is a binary "1" when speech (both voiced and unvoiced) is occurring and a binary "0" when no speech is occurring. The VAD signal can be obtained in numerous ways, for example, using acoustic gain, accelerometers, and radio frequency (RF) sensors.

The processor system and method includes denoising algorithms that calculate the transfer function among the noise sources and the microphones as well as the transfer function among the human user and the microphones. The transfer functions are used to remove noise from the received acoustic signal to produce at least one denoised acoustic data stream.

WO 02/07151

PCT/US01/22490

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a block diagram of a denoising system of an embodiment.

Figure 2 is a block diagram of a noise removal algorithm of an
5 embodiment, assuming a single noise source and a direct path to the
microphones.

Figure 3 is a block diagram of a front end of a noise removal algorithm
of an embodiment, generalized to n distinct noise sources (these noise sources
may be reflections or echoes of one another).

10 **Figure 4** is a block diagram of a front end of a noise removal algorithm
of an embodiment in the most general case where there are n distinct noise
sources and signal reflections.

Figure 5 is a flow diagram of a denoising method of an embodiment.

15 **Figure 6** shows results of a noise suppression algorithm of an
embodiment for an American English female speaker in the presence of airport
terminal noise that includes many other human speakers and public
announcements.

WO 02/07151

PCT/US01/22490

DETAILED DESCRIPTION

Figure 1 is a block diagram of a denoising system of an embodiment that uses knowledge of when speech is occurring derived from physiological information on voicing activity. The system includes microphones 10 and sensors 20 that provide signals to at least one processor 30. The processor includes a denoising subsystem or algorithm.

Figure 2 is a block diagram of a noise removal system/algorithm of an embodiment, assuming a single noise source and a direct path to the microphones. The noise removal system diagram includes a graphic description of the process of an embodiment, with a single signal source (100) and a single noise source (101). This algorithm uses two microphones, a "signal" microphone (MIC 1, 102) and a "noise" microphone (MIC 2, 103), but is not so limited. MIC 1 is assumed to capture mostly signal with some noise, while MIC 2 captures mostly noise with some signal. This is the common configuration with conventional advanced acoustic systems. The data from the signal to MIC 1 is denoted by $s(n)$, from the signal to MIC 2 by $s_2(n)$, from the noise to MIC 2 by $n(n)$, and from the noise to MIC 1 by $n_2(n)$. Similarly, the data from MIC 1 is denoted by $m_1(n)$, and the data from MIC 2 $m_2(n)$, where $s(n)$ denotes a discrete sample of the analog signal from the source.

The transfer functions from the signal to MIC 1 and from the noise to MIC 2 are assumed to be unity, but the transfer function from the signal to MIC 2 is denoted by $H_2(z)$ and from the noise to MIC 1 by $H_1(z)$. The assumption of unity transfer functions does not inhibit the generality of this algorithm, as the actual relations between the signal, noise, and microphones are simply ratios and the ratios are redefined in this manner for simplicity.

In conventional noise removal systems, the information from MIC 2 is used to attempt to remove noise from MIC 1. However, an unspoken assumption is that the Voice Activity Detection (VAD) is never perfect, and thus the denoising must be performed cautiously, so as not to remove too much of the signal along with the noise. However, if the VAD is assumed to be perfect and is equal to zero when there is no speech being produced by the user,

WO 02/07151

PCT/US01/22490

and one when speech is produced, a substantial improvement in the noise removal can be made.

In analyzing the single noise source and direct path to the microphones, with reference to **Figure 2**, the acoustic information coming into MIC 1 is denoted by $m_1(n)$. The information coming into MIC 2 is similarly labeled $m_2(n)$. In the z (digital frequency) domain, these are represented as $M_1(z)$ and $M_2(z)$. Then

$$\begin{aligned} M_1(z) &= S(z) + N_1(z) \\ M_2(z) &= N(z) + S_1(z) \end{aligned}$$

with

$$\begin{aligned} N_1(z) &= N(z)H_1(z) \\ S_1(z) &= S(z)H_2(z) \end{aligned}$$

so that

$$\begin{aligned} M_1(z) &= S(z) + N(z)H_1(z) \\ M_2(z) &= N(z) + S(z)H_2(z) \end{aligned} \quad \text{Eq. 1}$$

This is the general case for all two microphone systems. In a practical system there is always going to be some leakage of noise into MIC 1, and some leakage of signal into MIC 2. Equation 1 has four unknowns and only two known relationships and therefore cannot be solved explicitly.

However, there is another way to solve for some of the unknowns in Equation 1. The analysis starts with an examination of the case where the signal is not being generated, that is, where the VAD signal equals zero and speech is not being produced. In this case, $s(n) = S(z) = 0$, and Equation 1 reduces to

$$\begin{aligned} M_{1n}(z) &= N(z)H_1(z) \\ M_{2n}(z) &= N(z) \end{aligned}$$

where the n subscript on the M variables indicate that only noise is being received. This leads to

$$\begin{aligned} M_{1n}(z) &= M_{2n}(z)H_1(z) \\ H_1(z) &= \frac{M_{1n}(z)}{M_{2n}(z)} \end{aligned} \quad \text{Eq. 2}$$

WO 02/07151

PCT/US01/22490

$H_1(z)$ can be calculated using any of the available system identification algorithms and the microphone outputs when the system is certain that only noise is being received. The calculation can be done adaptively, so that the system can react to changes in the noise.

5 A solution is now available for one of the unknowns in Equation 1. Another unknown, $H_2(z)$, can be determined by using the instances where the VAD equals one and speech is being produced. When this is occurring, but the recent (perhaps less than 1 second) history of the microphones indicate low levels of noise, it can be assumed that $n(z) = N(z) \sim 0$. Then Equation 1 reduces to

$$\begin{aligned} M_{1s}(z) &= S(z) \\ M_{2s}(z) &= S(z)H_2(z) \end{aligned}$$

which in turn leads to

$$\begin{aligned} M_{2s}(z) &= M_{1s}(z)H_2(z) \\ H_2(z) &= \frac{M_{2s}(z)}{M_{1s}(z)} \end{aligned}$$

15 which is the inverse of the $H_1(z)$ calculation. However, it is noted that different inputs are being used – now only the signal is occurring whereas before only the noise was occurring. While calculating $H_2(z)$, the values calculated for $H_1(z)$ are held constant and vice versa. Thus, it is assumed that $H_1(z)$ and $H_2(z)$ do not change substantially while the other is being calculated.

20 After calculating $H_1(z)$ and $H_2(z)$, they are used to remove the noise from the signal. If Equation 1 is rewritten as

$$\begin{aligned} S(z) &= M_1(z) - N(z)H_1(z) \\ N(z) &= M_2(z) - S(z)H_2(z) \\ S(z) &= M_1(z) - [M_2(z) - S(z)H_2(z)]H_1(z) \\ S(z)[1 - H_2(z)H_1(z)] &= M_1(z) - M_2(z)H_1(z) \end{aligned}$$

then $N(z)$ may be substituted as shown to solve for $S(z)$ as:

$$S(z) = \frac{M_1(z) - M_2(z)H_1(z)}{1 - H_2(z)H_1(z)} \quad \text{Eq. 3}$$

25 If the transfer functions $H_1(z)$ and $H_2(z)$ can be described with sufficient accuracy, then the noise can be completely removed and the original signal recovered. This remains true without respect to the amplitude or spectral

WO 02/07151

PCT/US01/22490

characteristics of the noise. The only assumptions made are a perfect VAD, sufficiently accurate $H_1(z)$ and $H_2(z)$, and that $H_1(z)$ and $H_2(z)$ do not change substantially when the other is being calculated. In practice these assumptions have proven reasonable.

5 The noise removal algorithm described herein is easily generalized to include any number of noise sources. **Figure 3** is a block diagram of a front end of a noise removal algorithm of an embodiment, generalized to n distinct noise sources. These distinct noise sources may be reflections or echoes of one another, but are not so limited. There are several noise sources shown, each
10 with a transfer function, or path, to each microphone. The previously named path H_2 has been relabeled as H_0 , so that labeling noise source 2's path to MIC 1 is more convenient. The outputs of each microphone, when transformed to the z domain, are:

$$\begin{aligned} M_1(z) &= S(z) + N_1(z)H_1(z) + N_2(z)H_2(z) + \dots N_n(z)H_n(z) \\ M_2(z) &= S(z)H_0(z) + N_1(z)G_1(z) + N_2(z)G_2(z) + \dots N_n(z)G_n(z) \end{aligned} \quad \text{Eq. 4}$$

15 When there is no signal ($VAD = 0$), then (suppressing the z 's for clarity)

$$\begin{aligned} M_{1n} &= N_1H_1 + N_2H_2 + \dots N_nH_n \\ M_{2n} &= N_1G_1 + N_2G_2 + \dots N_nG_n \end{aligned} \quad \text{Eq. 5}$$

A new transfer function can now be defined, analogous to $H_1(z)$ above:

$$\tilde{H}_1 = \frac{M_{1n}}{M_{2n}} = \frac{N_1H_1 + N_2H_2 + \dots N_nH_n}{N_1G_1 + N_2G_2 + \dots N_nG_n} \quad \text{Eq. 6}$$

20 Thus $\tilde{H}_1$ depends only on the noise sources and their respective transfer functions and can be calculated any time there is no signal being transmitted. Once again, the n subscripts on the microphone inputs denote only that noise is being detected, while an s subscript denotes that only signal is being received by the microphones.

Examining Equation 4 while assuming that there is no noise produces

$$\begin{aligned} M_{1s} &= S \\ M_{2s} &= SH_0 \end{aligned} \quad \text{Eq. 7}$$

Thus H_0 can be solved for as before, using any available transfer function calculating algorithm. Mathematically

WO 02/07151

PCT/US01/22490

$$H_0 = \frac{M_{2z}}{M_{1z}}$$

Rewriting Equation 4, using $\tilde{H}_1$ defined in Equation 6, provides,

$$\tilde{H}_1 = \frac{M_1 - S}{M_2 - SH_0} \quad \text{Eq. 7}$$

Solving for S yields,

$$S = \frac{M_1 - M_2 \tilde{H}_1}{1 - H_0 \tilde{H}_1} \quad \text{Eq. 8}$$

which is the same as Equation 3, with H_0 taking the place of H_2 , and $\tilde{H}_1$ taking the place of H_1 . Thus the noise removal algorithm still is mathematically valid for any number of noise sources, including multiple echoes of noise sources.

Again, if H_0 and $\tilde{H}_1$ can be estimated to a high enough accuracy, and the above assumption of only one path from the signal to the microphones holds, the noise may be removed completely.

The most general case involves multiple noise sources and multiple signal sources. **Figure 4** is a block diagram of a front end of a noise removal algorithm of an embodiment in the most general case where there are n distinct noise sources and signal reflections. Here, reflections of the signal enter both microphones. This is the most general case, as reflections of the noise source into the microphones can be modeled accurately as simple additional noise sources. For clarity, the direct path from the signal to MIC 2 has changed from $H_0(z)$ to $H_{00}(z)$, and the reflected paths to Microphones 1 and 2 are denoted by $H_{01}(z)$ and $H_{02}(z)$, respectively.

The input into the microphones now becomes

$$\begin{aligned} M_1(z) &= S(z) + S(z)H_{01}(z) + N_1(z)H_1(z) + N_2(z)H_2(z) + \dots N_n(z)H_n(z) \\ M_2(z) &= S(z)H_{00}(z) + S(z)H_{02}(z) + N_1(z)G_1(z) + N_2(z)G_2(z) + \dots N_n(z)G_n(z) \end{aligned} \quad \text{Eq. 9}$$

When the VAD = 0, the inputs become (suppressing the z 's again)

$$\begin{aligned} M_{1n} &= N_1H_1 + N_2H_2 + \dots N_nH_n \\ M_{2n} &= N_1G_1 + N_2G_2 + \dots N_nG_n \end{aligned}$$

WO 02/07151

PCT/US01/22490

which is the same as Equation 5. Thus, the calculation of $\tilde{H}_1$ in Equation 6 is unchanged, as expected. In examining the situation where there is no noise, Equation 9 reduces to

$$\begin{aligned} M_{1s} &= S + SH_{01} \\ M_{2s} &= SH_{00} + SH_{02} \end{aligned}$$

5 This leads to the definition of $\tilde{H}_2$:

$$\tilde{H}_2 = \frac{M_{2s}}{M_{1s}} = \frac{H_{00} + H_{02}}{1 + H_{01}} \quad \text{Eq. 10}$$

Rewriting Equation 9 again using the definition for $\tilde{H}_1$ (as in Equation 7) provides

$$\tilde{H}_1 = \frac{M_1 - S(1 + H_{01})}{M_2 - S(H_{00} + H_{02})} \quad \text{Eq. 11}$$

10 Some algebraic manipulation yields

$$\begin{aligned} S(1 + H_{01} - \tilde{H}_1(H_{00} + H_{02})) &= M_1 - M_2\tilde{H}_1 \\ S(1 + H_{01}) \left[1 - \tilde{H}_1 \frac{(H_{00} + H_{02})}{(1 + H_{01})} \right] &= M_1 - M_2\tilde{H}_1 \\ S(1 + H_{01}) [1 - \tilde{H}_1\tilde{H}_2] &= M_1 - M_2\tilde{H}_1 \end{aligned}$$

and finally

$$S(1 + H_{01}) = \frac{M_1 - M_2\tilde{H}_1}{1 - \tilde{H}_1\tilde{H}_2} \quad \text{Eq. 12}$$

Equation 12 is the same as equation 8, with the replacement of H_0 by $\tilde{H}_2$, and the addition of the $(1+H_{01})$ factor on the left side. This extra factor means that S cannot be solved for directly in this situation, but a solution can be generated for the signal plus the addition of all of its echoes. This is not such a bad situation, as there are many conventional methods for dealing with echo suppression, and even if the echoes are not suppressed, it is unlikely that they will affect the comprehensibility of the speech to any meaningful extent. The more complex calculation of $\tilde{H}_2$ is needed to account for the signal echoes in Microphone 2, which act as noise sources.

Figure 5 is a flow diagram of a denoising method of an embodiment. In operation, the acoustic signals are received 502. Further, physiological

WO 02/07151

PCT/US01/22490

information associated with human voicing activity is received 504. A first transfer function representative of the acoustic signal is calculated upon determining that voicing information is absent from the acoustic signal for at least one specified period of time 506. A second transfer function
 5 representative of the acoustic signal is calculated upon determining that voicing information is present in the acoustic signal for at least one specified period of time 508. Noise is removed from the acoustic signal using at least one combination of the first transfer function and the second transfer function, producing denoised acoustic data streams 510.

10 An algorithm for noise removal, or denoising algorithm, is described herein, from the simplest case of a single noise source with a direct path to multiple noise sources with reflections and echoes. The algorithm has been shown herein to be viable under any environmental conditions. The type and amount of noise are inconsequential if a good estimate has been made of $\tilde{H}_1$
 15 and $\tilde{H}_2$, and if they do not change substantially while the other is calculated. If the user environment is such that echoes are present, they can be compensated for if coming from a noise source. If signal echoes are also present, they will affect the cleaned signal, but the effect should be negligible in most environments.

20 In operation, the algorithm of an embodiment has shown excellent results in dealing with a variety of noise types, amplitudes, and orientations. However, there are always approximations and adjustments that have to be made when moving from mathematical concepts to engineering applications. One assumption is made in Equation 3, where $H_2(z)$ is assumed small and
 25 therefore $H_2(z)H_1(z) \approx 0$, so that Equation 3 reduces to

$$S(z) \approx M_1(z) - M_2(z)H_1(z).$$

This means that only $H_1(z)$ has to be calculated, speeding up the process and reducing the number of computations required considerably. With the proper selection of microphones, this approximation is easily realized.

30 Another approximation involves the filter used in an embodiment. The actual $H_1(z)$ will undoubtedly have both poles and zeros, but for stability and

WO 02/07151

PCT/US01/22490

simplicity an all-zero Finite Impulse Response (FIR) filter is used. With enough taps (around 60) the approximation to the actual $H_1(z)$ is very good.

Regarding subband selection, the wider the range of frequencies over which a transfer function must be calculated, the more difficult it is to calculate it accurately. Therefore the acoustic data was divided into 16 subbands, with the lowest frequency at 50 Hz and the highest at 3700. The denoising algorithm was then applied to each subband in turn, and the 16 denoised data streams were recombined to yield the denoised acoustic data. This works very well, but any combinations of subbands (i.e. 4, 6, 8, 32, equally spaced, perceptually spaced, etc.) can be used and has been found to work as well.

The amplitude of the noise was constrained in an embodiment so that the microphones used did not saturate (i.e. operate outside a linear response region). It is important that the microphones operate linearly to ensure the best performance. Even with this restriction, very high signal-to-noise ratios (SNR) can be tested (down to about -10 dB).

The calculation of $H_1(z)$ was accomplished every 10 milliseconds using the Least-Mean Squares (LMS) method, a common adaptive transfer function. An explanation may be found in "Adaptive Signal Processing" (1985), by Widrow and Stearns, published by Prentice-Hall, ISBN 0-13-004029-0.

The VAD for an embodiment was derived from a radio frequency sensor and the two microphones, yielding very high accuracy (>99%) for both voiced and unvoiced speech. The VAD of an embodiment uses a radio frequency (RF) interferometer to detect tissue motion associated with human speech production, but is not so limited. It is therefore completely acoustic-noise free, and is able to function in any acoustic noise environment. A simple energy measurement can be used to determine if voiced speech is occurring. Unvoiced speech can be determined using conventional frequency-based methods, by proximity to voiced sections, or through a combination of the above. Since there is much less energy in unvoiced speech, its activation accuracy is not as critical as voiced speech.

With voiced and unvoiced speech detected reliably, the algorithm of an embodiment can be implemented. Once again, it is useful to repeat that the

WO 02/07151

PCT/US01/22490

noise removal algorithm does not depend on how the VAD is obtained, only that it is accurate, especially for voiced speech. If speech is not detected and training occurs on the speech, the subsequent denoised acoustic data can be distorted.

5 Data was collected in four channels, one for MIC 1, one for MIC 2, and two for the radio frequency sensor that detected the tissue motions associated with voiced speech. The data were sampled simultaneously at 40 kHz, then digitally filtered and decimated down to 8 kHz. The high sampling rate was used to reduce any aliasing that might result from the analog to digital process.

10 A four-channel National Instruments A/D board was used along with Labview to capture and store the data. The data was then read into a C program and denoised 10 milliseconds at a time.

Figure 6 shows results of a noise suppression algorithm of an embodiment for an American English speaking female in the presence of airport terminal noise that includes many other human speakers and public announcements. The speaker is uttering the numbers 406-5562 in the midst of moderate airport terminal noise. The dirty acoustic data was denoised 10 milliseconds at a time, and before denoising the 10 milliseconds of data were prefiltered from 50 to 3700 Hz. A reduction in the noise of approximately 17 dB is evident. No post filtering was done on this sample; thus, all of the noise reduction realized is due to the algorithm of an embodiment. It is clear that the algorithm adjusts to the noise instantly, and is capable of removing the very difficult noise of other human speakers. Many different types of noise have all been tested with similar results, including street noise, helicopters, music, and sine waves, to name a few. Also, the orientation of the noise can be varied substantially without significantly changing the noise suppression performance. Finally, the distortion of the cleaned speech is very low, ensuring good performance for speech recognition engines and human receivers alike.

15

20

25

The noise removal algorithm of an embodiment has been shown to be viable under any environmental conditions. The type and amount of noise are inconsequential if a good estimate has been made of $\hat{H}_1$ and $\hat{H}_2$. If the user environment is such that echoes are present, they can be compensated for if

30

WO 02/07151

PCT/US01/22490

coming from a noise source. If signal echoes are also present, they will affect the cleaned signal, but the effect should be negligible in most environments.

Various embodiments are described herein with reference to the figures, but the detailed description and the figures are not intended to be limiting.

- 5 Various combinations of the elements described have not been shown, but are within the scope of the invention which is defined by the following claims.

WO 02/07151

PCT/US01/22490

CLAIMS

What is claimed is:

1. A method for removing noise from acoustic signals, comprising:
receiving a plurality of acoustic signals;
5 receiving physiological information associated with human voicing activity;
generating at least one first transfer function representative of the plurality of acoustic signals upon determining that voicing information is absent from the plurality of acoustic signals for at least one specified period of time;
10 generating at least one second transfer function representative of the plurality of acoustic signals upon determining that voicing information is present in the plurality of acoustic signals for the at least one specified period of time;
removing noise from the plurality of acoustic signals using at least one
15 combination of the at least one first transfer function and the at least one second transfer function to produce at least one denoised acoustic data stream.
2. The method of claim 1, wherein the plurality of acoustic signals include at least one reflection of at least one associated noise source signal and at least one reflection of at least one acoustic source signal.
- 20 3. The method of claim 1, wherein receiving physiological information comprises receiving physiological data associated with human voicing using at least one detector selected from a group consisting of radio frequency devices, electroglottographs, ultrasound devices, acoustic throat microphones, and airflow detectors.
- 25 4. The method of claim 1, wherein receiving the plurality of acoustic signals includes receiving using a plurality of independently located microphones.

WO 02/07151

PCT/US01/22490

5. The method of claim 1, wherein removing noise further includes generating at least one third transfer function using the at least one first transfer function and the at least one second transfer function.
6. The method of claim 1, wherein generating the at least one first transfer function comprises recalculating the at least one first transfer function during at least one prespecified interval.
7. The method of claim 1, wherein generating the at least one second transfer function comprises recalculating the at least one second transfer function during at least one prespecified interval.
8. The method of claim 1, wherein generating the at least one first transfer function and the at least one second transfer function comprises use of at least one technique selected from a group consisting of adaptive techniques and recursive techniques.
9. A method for removing noise from electronic signals, comprising:
detecting an absence of voiced information during at least one period;
receiving at least one noise source signal during the at least one period;
generating at least one transfer function representative of the at least one noise source signal;
receiving at least one composite signal comprising acoustic and noise signals; and
removing the noise signal from the at least one composite signal using the at least one transfer function to produce at least one denoised acoustic data stream.
10. The method of claim 9, wherein the at least one noise source signal includes at least one reflection of at least one associated noise source signal.

WO 02/07151

PCT/US01/22490

11. The method of claim 9, wherein the at least one composite signal includes at least one reflection of at least one associated composite signal.
12. The method of claim 9, wherein detecting comprises collecting physiological data associated with human voicing using at least one detector
5 selected from a group consisting of radio frequency devices, electroglottographs, ultrasound devices, acoustic throat microphones, and airflow detectors.
13. The method of claim 9, wherein receiving includes receiving the at least one noise source signal using at least one microphone.
- 10 14. The method of claim 13, wherein the at least one microphone includes a plurality of independently located microphones.
- 15 15. The method of claim 9, wherein removing the noise signal from the at least one composite signal using the at least one transfer function includes generating at least one other transfer function using the at least one transfer function.
16. The method of claim 9, wherein generating at least one transfer function comprises recalculating the at least one transfer function during at least one prespecified interval.
- 20 17. The method of claim 9, wherein generating the at least one transfer function comprises calculating the at least one transfer function using at least one technique selected from a group consisting of adaptive techniques and recursive techniques.
- 25 18. A method for removing noise from electronic signals, comprising:
determining at least one unvoicing period during which voiced
information is absent;

WO 02/07151

PCT/US01/22490

receiving at least one noise signal input during the at least one unvoicing period and generating at least one unvoicing transfer function representative of the at least one noise signal;

5 determining at least one voicing period during which voiced information is present;

receiving at least one acoustic signal input from at least one signal sensing device during the at least one voicing period and generating at least one voicing transfer function representative of the at least one acoustic signal;

10 receiving at least one composite signal comprising acoustic and noise signals; and

removing the noise signal from the at least one composite signal using at least one combination of the at least one unvoicing transfer function and the at least one voicing transfer function to produce at least one denoised acoustic data stream.

15 19. A system for removing noise from acoustic signals, comprising:

at least one receiver that receives at least one acoustic signal;

at least one sensor that receives physiological information associated with human voicing activity;

20 at least one processor coupled among the at least one receiver and the at least one sensor that generates a plurality of transfer functions, wherein at least one first transfer function representative of the at least one acoustic signal is generated in response to a determination that voicing information is absent from the at least one acoustic signal for at least one specified period of time, wherein at least one second transfer function representative of the at least one acoustic signal is generated in response to a determination that voicing information is present in the at least one acoustic signal for at least one specified period of time, wherein noise is removed from the at least one acoustic signal using at least one combination of the at least one first transfer function and the at least one second transfer function to produce at least one denoised acoustic data stream.

25

30

WO 02/07151

PCT/US01/22490

20. The system of claim 19, wherein the at least one sensor includes at least one radio frequency (RF) interferometer that detects tissue motion associated with human speech production.
21. The system of claim 19, wherein the at least one sensor includes at least one sensor selected from a group consisting of radio frequency devices, electroglottographs, ultrasound devices, acoustic throat microphones, and airflow detectors.
22. The system of claim 19, further comprising:
dividing acoustic data of the at least one acoustic signal into a plurality of subbands;
removing noise from each of the plurality of subbands using the at least one combination of the at least one first transfer function and the at least one second transfer function, wherein a plurality of denoised acoustic data streams are generated; and
combining the plurality of denoised acoustic data streams to generate the at least one denoised acoustic data stream.
23. The system of claim 19, wherein the at least one receiver includes a plurality of independently located microphones.
24. A system for removing noise from acoustic signals, comprising at least one processor coupled among at least one microphone and at least one voicing sensor, wherein the at least one voicing sensor collects physiological data associated with voicing, wherein an absence of voiced information is detected during at least one period using the at least one voicing sensor, wherein at least one noise source signal is received during the at least one period using the at least one microphone, wherein the at least one processor generates at least one transfer function representative of the at least one noise source signal, wherein the at least one microphone receives at least one composite signal comprising acoustic and noise signals, and the at least one processor removes the noise

WO 02/07151

PCT/US01/22490

signal from the at least one composite signal using the at least one transfer function to produce at least one denoised acoustic data stream.

25. A signal processing system coupled among at least one user and at least one electronic device, wherein the signal processing system includes at least one denoising subsystem for removing noise from acoustic signals, the denoising subsystem comprising at least one processor coupled among at least one receiver and at least one sensor, wherein the at least one receiver is coupled to receive at least one acoustic signal, wherein the at least one sensor is coupled to receive physiological information associated with human voicing activity, wherein the at least one processor generates a plurality of transfer functions, wherein at least one first transfer function representative of the at least one acoustic signal is generated in response to a determination that voicing information is absent from the at least one acoustic signal for at least one specified period of time, wherein at least one second transfer function representative of the at least one acoustic signal is generated in response to a determination that voicing information is present in the at least one acoustic signal for at least one specified period of time, wherein noise is removed from the at least one acoustic signal using at least one combination of the at least one first transfer function and the at least one second transfer function to produce at least one denoised acoustic data stream.

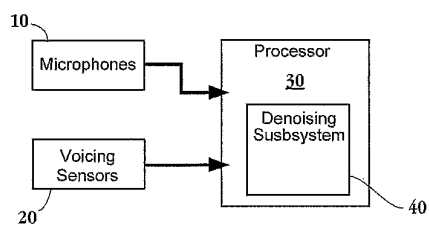
26. The signal processing system of claim 25, wherein the at least one electronic device includes at least one device selected from a group consisting of cellular telephones, personal digital assistants, portable communication devices, computers, video cameras, digital cameras, and telematics systems.

27. A computer readable medium comprising executable instructions which, when executed in a processing system, remove noise from received acoustic signals by:
- receiving at least one acoustic signal;

WO 02/07151

PCT/US01/22490

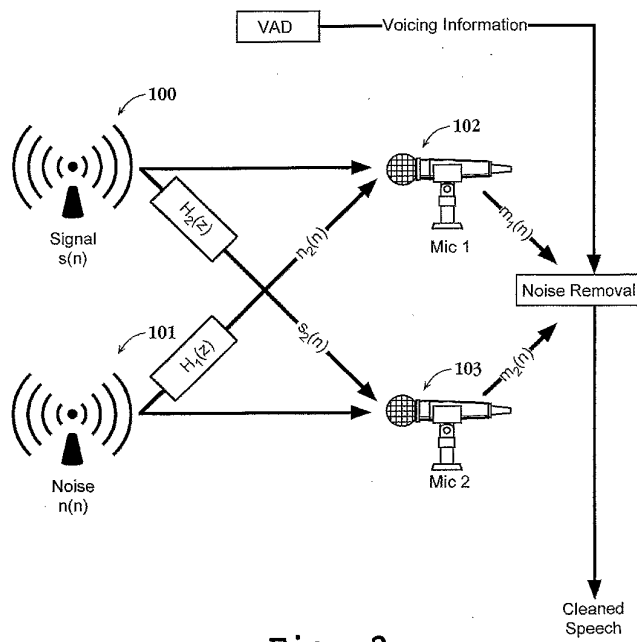
- receiving physiological information associated with human voicing activity;
- generating at least one first transfer function representative of the at least one acoustic signal upon determining that voicing information is absent from the at least one acoustic signal for at least one specified period of time;
- generating at least one second transfer function representative of the at least one acoustic signal upon determining that voicing information is present in the at least one acoustic signal for at least one specified period of time;
- removing noise from the at least one acoustic signal using at least one combination of the at least one first transfer function and the at least one second transfer function to produce at least one denoised acoustic data stream.
28. An electromagnetic medium comprising executable instructions which, when executed in a processing system, remove noise from received acoustic signals by:
- receiving at least one acoustic signal;
- receiving physiological information associated with human voicing activity;
- generating at least one first transfer function representative of the at least one acoustic signal upon determining that voicing information is absent from the at least one acoustic signal for at least one specified period of time;
- generating at least one second transfer function representative of the at least one acoustic signal upon determining that voicing information is present in the at least one acoustic signal for at least one specified period of time;
- removing noise from the at least one acoustic signal using at least one combination of the at least one first transfer function and the at least one second transfer function to produce at least one denoised acoustic data stream.

**Fig. 1**

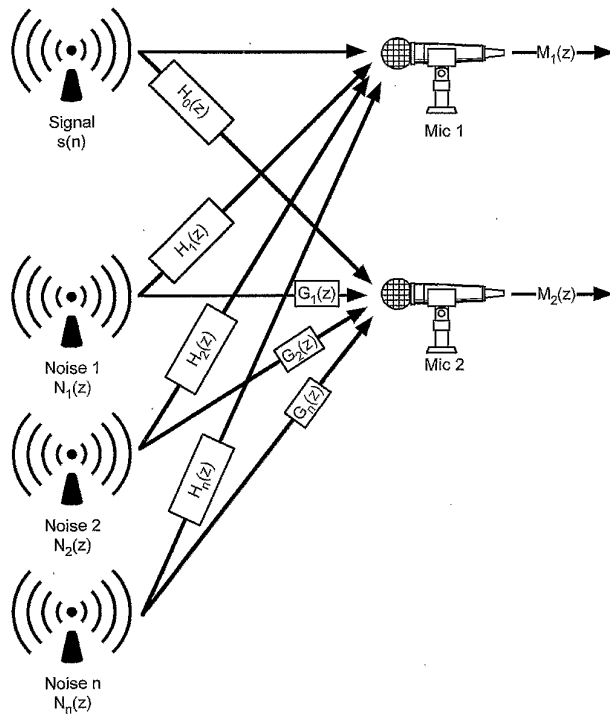
WO 02/07151

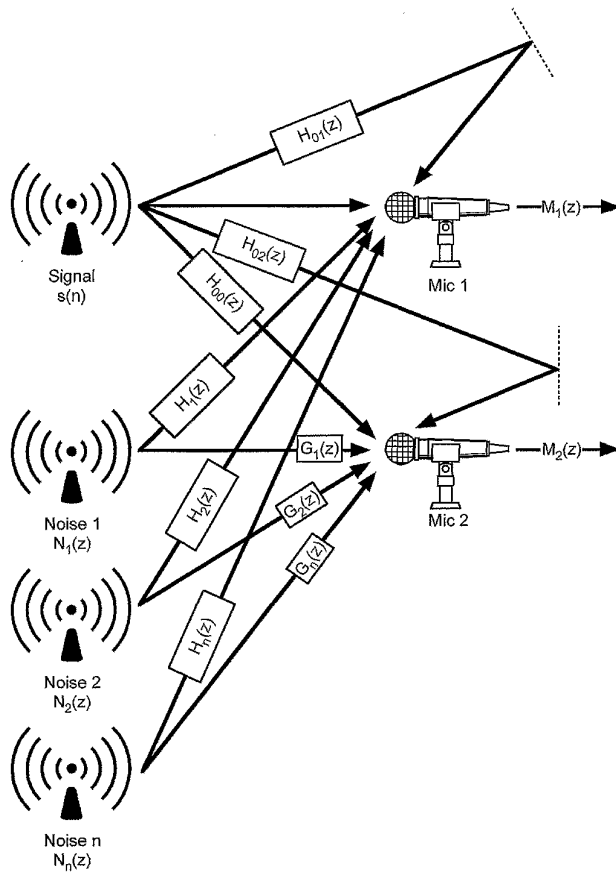
PCT/US01/22490

2/6

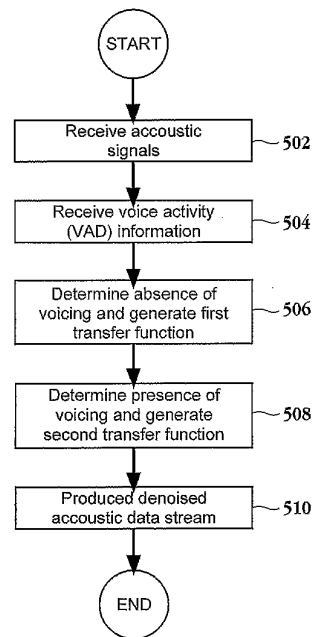
**Fig. 2**

SUBSTITUTE SHEET (RULE 26)

**Fig. 3**

**Fig. 4**

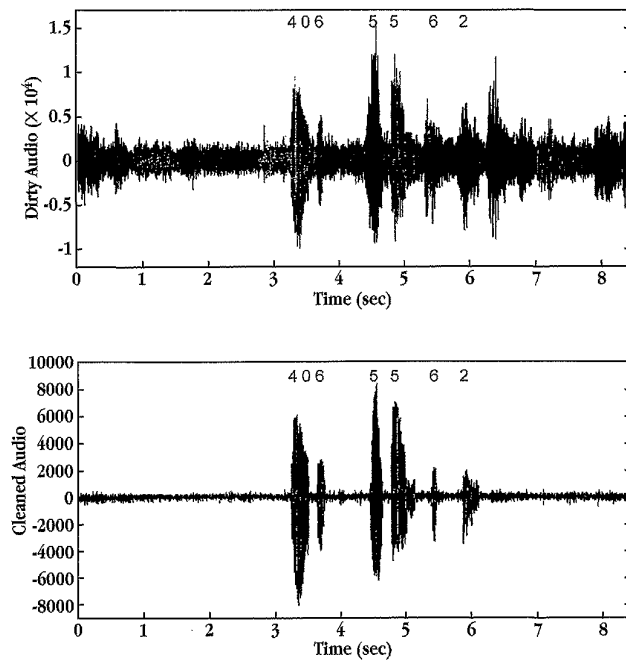
5/6

**Fig. 5**

WO 02/07151

PCT/US01/22490

6/6

**Fig. 6**

【国際公開パンフレット（コレクトバージョン）】

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau(43) International Publication Date
24 January 2002 (24.01.2002)

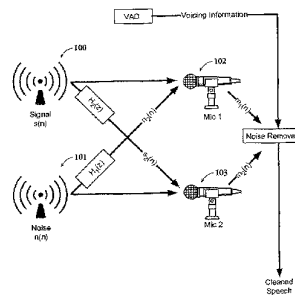
PCT

(10) International Publication Number
WO 02/07151 A3

- (51) International Patent Classification: G10L 21/02
- (21) International Application Number: PCT/US01/22490
- (22) International Filing Date: 17 July 2001 (17.07.2001)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
60/219,297 19 July 2000 (19.07.2000) US
09/905,361 12 July 2001 (12.07.2001) US
- (63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:
US 60/219,297 (CIP)
Filed on 19 July 2000 (19.07.2000)
US 09/905,361 (CIP)
Filed on 12 July 2001 (12.07.2001)
- (71) Applicant (for all designated States except US): ALIPH-COM [US/US]; 2nd floor, 54 Mint Street, San Francisco, CA 94103 (US).
- (72) Inventors; and
(75) Inventors/Applicants (for US only): BURNETT, Gregory, C. [US/US]; 637 South H Street, Livermore, CA 94550 (US). BREITFELLER, Eric, F. [US/US]; 11526 Streambed Place, Dublin, CA 94568-2294 (US).
- (74) Agent: PERKINS COIE, LLP; P.O. Box 2168, Menlo Park, CA 94026 (US).
- (81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.
- (84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

[Continued on next page]

(54) Title: METHOD AND APPARATUS FOR REMOVING NOISE FROM SPEECH SIGNALS



(57) Abstract: A method and system are provided for acoustic noise removal from human speech, wherein noise is removed without respect to noise type, amplitude, or orientation. The system includes microphones and a voice activity detection (VAD) data stream coupled among a processor. The microphones receive acoustic signals and the VAD produces a signal including a binary one when speech (voiced and unvoiced) is occurring and a binary zero in the absence of speech. The processor includes denoising algorithms that generate transfer functions. The transfer functions include a transfer function generated in response to a determination that voicing information is absent from the received acoustic signal during a specified time period. The transfer functions also include transfer functions generated in response to a determination that voicing information is present in the acoustic signal during a specified time period. At least one denoised acoustic data stream is generated using the transfer functions.

WO 02/07151 A3

WO 02/07151 A3



Published:
— with international search report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(88) Date of publication of the international search report:
30 May 2002

【国際調査報告】

INTERNATIONAL SEARCH REPORT		International Application No. PC1/US 01/22490
A. CLASSIFICATION OF SUBJECT MATTER IPC 7 G10L21/02		
According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED		
Minimum documentation searched (classification system followed by classification symbols) IPC 7 G10L		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the international search (name of data base and, where practical, search terms used) EPO-Internal, WPI Data, INSPEC		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	ZHAO LI HOFFMAN ET AL: "Robust speech coding using microphone arrays" SIGNALS, SYSTEMS & COMPUTERS, 1997. CONFERENCE RECORD OF THE THIRTY-FIRST ASILOMAR CONFERENCE ON PACIFIC GROVE, CA, USA 2-5 NOV. 1997, LOS ALAMITOS, CA, USA, IEEE COMPUT. SOC, US, 2 November 1997 (1997-11-02), pages 44-48, XP010280758 ISBN: 0-8186-8316-3	1, 2, 4-11, 13-19, 22-28
Y	the whole document --- -/--	3, 12, 20, 21
☒ Further documents are listed in the continuation of box C. ☐ Patent family members are listed in annex.		
* Special categories of cited documents: *A* document defining the general state of the art which is not considered to be of particular relevance *E* earlier document but published on or after the international filing date *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) *O* document referring to an oral disclosure, use, exhibition or other means *P* document published prior to the international filing date but later than the priority date claimed *T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention *X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art. *Z* document member of the same patent family		
Date of the actual completion of the international search 20 December 2001		Date of mailing of the international search report 04/02/2002
Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel: (+31-70) 340-2040, Tx: 31 651 epo nl, Fax: (+31-70) 340-2016		Authorized officer Quélavoine, R

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

International Application No.
PCT/US 01/22490

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	NG L C ET AL: "Denoising of human speech using combined acoustic and EM sensor signal processing" 2000 IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING. PROCEEDINGS (CAT. NO. 00CH37100), ISTANBUL, TURKEY, 5-9 JUNE 2000, pages 229-232 vol.1, XP002186255 2000, Piscataway, NJ, USA, IEEE, USA ISBN: 0-7803-6293-4 the whole document	3,12,20, 21
A	AFFES S ET AL: "A SIGNAL SUBSPACE TRACKING ALGORITHM FOR MICROPHONE ARRAY PROCESSING OF SPEECH" IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING, IEEE INC. NEW YORK, US, vol. 5, no. 5, 1 September 1997 (1997-09-01), pages 425-437, XP000774303 ISSN: 1063-6676 the whole document	2,10,11

フロントページの続き

(81)指定国 AP(GH,GM,KE,LS,MW,MZ,SD,SL,SZ,TZ,UG,ZW),EA(AM,AZ,BY,KG,KZ,MD,RU,TJ,TM),EP(AT,BE,CH,CY,DE,DK,ES,FI,FR,GB,GR,IE,IT,LU,MC,NL,PT,SE,TR),OA(BF,BJ,CF,CG,CI,CM,GA,GN,GQ,GW,ML,MR,NE,SN,TD,TG),AE,AG,AL,AM,AT,AU,AZ,BA,BB,BG,BR,BY,BZ,CA,CH,CN,CO,CR,CU,CZ,DE,DK,DM,DZ,EC,EE,ES,FI,GB,GD,GE,GH,GM,HR,HU,ID,IL,IN,IS,JP,KE,KG,KP,KR,KZ,LC,LK,LR,LS,LT,LU,LV,MA,MD,MG,MK,MN,MW,MX,MZ,NO,NZ,PL,PT,RO,RU,SD,SE,SG,SI,SK,SL,TJ,TM,TR,TT,TZ,UA,UG,US,UZ,VN,YU,ZA,ZW

(74)代理人 100120112

弁理士 中西 基晴

(72)発明者 バーネット, グレゴリー・シー

アメリカ合衆国カリフォルニア州94550, ライブモア, サウス・エイチ・ストリート 637

(72)発明者 プレイトフェラー, エリック・エフ

アメリカ合衆国カリフォルニア州94568-2294, ダブリン, ストリームベッド・プレイス

11526

Fターム(参考) 5D015 DD02 EE05