

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号
特許第4764995号
(P4764995)

(45) 発行日 平成23年9月7日 (2011.9.7)

(24) 登録日 平成23年6月24日 (2011.6.24)

(51) Int.Cl.

F I

G 1 O L 15/04 (2006.01)

G 1 O L 21/02 (2006.01)

G 1 O L 15/04 3 O O C

G 1 O L 21/02 1 O 1 B

G 1 O L 21/02 1 O 3 Z

請求項の数 21 (全 15 頁)

(21) 出願番号	特願2001-517379 (P2001-517379)	(73) 特許権者	511099995
(86) (22) 出願日	平成12年8月11日 (2000.8.11)		キューエヌエックス ソフトウェア シス
(65) 公表番号	特表2003-507764 (P2003-507764A)		テムズ コーポレイション
(43) 公表日	平成15年2月25日 (2003.2.25)		カナダ国 ケー2エム 1ダブリュー8
(86) 国際出願番号	PCT/US2000/022201		オンタリオ, オタワ, テレンス マシ
(87) 国際公開番号	W02001/013364		ューズ クレセント 175
(87) 国際公開日	平成13年2月22日 (2001.2.22)	(74) 代理人	100107489
審査請求日	平成19年7月31日 (2007.7.31)		弁理士 大塩 竹志
(31) 優先権主張番号	09/375,309	(72) 発明者	ザカロスカス, ピエール
(32) 優先日	平成11年8月16日 (1999.8.16)		アメリカ合衆国・92024・カリフォル
(33) 優先権主張国	米国 (US)		ニア州・エンシニタス・ケニングトン ロ
			ード・1723
		審査官	清水 正一
			最終頁に続く

(54) 【発明の名称】 雑音を含む音響信号の高品質化

(57) 【特許請求の範囲】

【請求項 1】

入力信号を強調する方法であって、
該方法は、
雑音のある入力信号の時間 - 周波数表現を決定することと、
背景雑音レベルと信号対雑音比とを推定することと、
該時間 - 周波数表現に対する整合低雑音信号テンプレートを決定することと、
該時間 - 周波数表現の一部を、該時間 - 周波数表現と該整合低雑音信号テンプレートとの混合に置き換えることであって、該混合は、該信号対雑音比によって重み付けられている、ことと
を包含する、方法。

【請求項 2】

前記整合低雑音信号テンプレートを決定することが、整合低雑音スペクトログラムを決定することを包含する、請求項 1 に記載の方法。

【請求項 3】

前記整合低雑音信号テンプレートを決定することが、前記時間 - 周波数表現と、信号モデル内の複数の低雑音信号テンプレートであって、前記整合低雑音信号テンプレートを含む複数の低雑音信号テンプレートとの最小の平均二乗根差を決定することを包含する、請求項 1 に記載の方法。

【請求項 4】

前記複数の低雑音信号テンプレートが、低雑音スペクトログラムを含む、請求項 3 に記載の方法。

【請求項 5】

信号モデルへの複数の低雑音信号テンプレートであって、前記整合低雑音信号テンプレートを含む複数の低雑音信号テンプレートを収集することをさらに包含する、請求項 1 に記載の方法。

【請求項 6】

前記信号モデルをトレーニングすることをさらに包含する、請求項 5 に記載の方法。

【請求項 7】

前記方法が、
学習モードがアクティブであるか非アクティブであるかを決定することをさらに包含し

10

、
前記置き換えることが、

該学習モードが非アクティブである場合に、デジタル化された入力信号の一部を、前記時間 - 周波数表現と前記整合低雑音信号テンプレートとの信号対雑音比で重み付けられた混合に置き換えることをさらに包含する、請求項 6 に記載の方法。

【請求項 8】

前記トレーニングすることが、前記学習モードがアクティブである場合に、前記時間 - 周波数表現を用いて前記整合低雑音信号テンプレートを更新することを包含する、請求項 7 に記載の方法。

20

【請求項 9】

前記トレーニングすることが、前記学習モードがアクティブである場合に、前記信号モデルへの新しい低雑音信号テンプレートとして前記時間 - 周波数表現を追加することを包含する、請求項 7 に記載の方法。

【請求項 10】

プロセッサと、
該プロセッサに結合されたメモリと
を備えた信号強調システムであって、
該メモリは、命令を含み、
該命令は、

30

複数の低雑音信号テンプレートを含む信号モデルを確立することと、

入力信号を取得することと、

該入力信号に対する該信号モデル内の整合低雑音信号テンプレートを決定することと

、
該入力信号の一部を、該入力信号と該整合低雑音信号テンプレートとの信号対雑音比で重み付けられた混合に置き換えることと

を該プロセッサに行わせる、システム。

【請求項 11】

前記メモリは、前記入力信号の入力信号スペクトログラムを決定することを前記プロセッサに行わせる命令をさらに含み、

40

前記整合低雑音信号テンプレートを決定することを該プロセッサに行わせる命令は、整合低雑音スペクトログラムテンプレートを決定することを該プロセッサに行わせ、

前記入力信号の一部を置き換えることを該プロセッサに行わせる命令は、該入力信号スペクトログラムの一部を入力信号スペクトルと該整合低雑音スペクトログラムテンプレートとの信号対雑音比で重み付けられた混合に置き換えることによって、低雑音スペクトログラムを生成することを該プロセッサに行わせる、請求項 10 に記載のシステム。

【請求項 12】

前記メモリは、前記低雑音スペクトログラムから低雑音出力時系列を合成することを前記プロセッサに行わせる命令をさらに含む、請求項 11 に記載のシステム。

【請求項 13】

50

前記整合低雑音信号テンプレートを決定することを前記プロセッサに行わせる命令は、前記入力信号と前記複数の低雑音信号テンプレートの各々との間の信号対雑音比で重み付けられた距離を決定することを該プロセッサに行わせ、それによって、該入力信号における周波数帯域は、当該周波数帯域の信号対雑音比に比例して、該信号対雑音比で重み付けられた距離に寄与する、請求項 10 に記載のシステム。

【請求項 14】

前記メモリは、前記信号モデルをトレーニングすることを前記プロセッサに行わせる命令をさらに含む、請求項 10 に記載のシステム。

【請求項 15】

前記信号モデルをトレーニングすることを前記プロセッサに行わせる命令は、該信号モデル内の前記複数の低雑音信号テンプレートのうちの少なくとも 1 つを更新することを該プロセッサに行わせる命令を含む、請求項 14 に記載のシステム。

10

【請求項 16】

前記信号モデルをトレーニングすることを前記プロセッサに行わせる命令は、新しい低雑音信号テンプレートとして前記入力信号を該信号モデルに追加することを該プロセッサに行わせる命令を含む、請求項 14 に記載のシステム。

【請求項 17】

命令が記録されたコンピュータ読み取り可能な格納媒体であって、

該命令は、

複数の低雑音信号テンプレートを含む信号モデルから雑音のある入力信号に対する整合低雑音信号テンプレートを決定することと、

20

該入力信号の一部を該入力信号と該整合低雑音信号テンプレートとの信号対雑音比で重み付けられた混合に置き換えることと

をプロセッサに行わせる、コンピュータ読み取り可能な格納媒体。

【請求項 18】

前記命令は、前記雑音のある入力信号の入力信号スペクトログラムを決定することを前記プロセッサにさらに行わせ、

前記整合低雑音信号テンプレートを決定する命令は、

該入力信号スペクトログラムと前記複数の低雑音信号テンプレートの各々との間の信号対雑音比で重み付けられた距離を決定することと、

30

該整合低雑音信号テンプレートとして、該信号対雑音比で重み付けられた距離のうちの最小の距離を有する該信号モデル内の低雑音信号テンプレートを選択することと

を該プロセッサに行わせ、それによって、該雑音のある入力信号における周波数帯域は、当該周波数帯域の信号対雑音比に比例して、該信号対雑音比で重み付けられた距離に寄与する、請求項 17 に記載のコンピュータ読み取り可能な格納媒体。

【請求項 19】

前記媒体は、前記整合低雑音信号テンプレートを決定する前に前記雑音のある入力信号における過渡現象を検出することを前記プロセッサに行わせる命令をさらに格納する、請求項 17 に記載のコンピュータ読み取り可能な格納媒体。

40

【請求項 20】

前記過渡現象は、音声の過渡現象である、請求項 19 に記載のコンピュータ読み取り可能な格納媒体。

【請求項 21】

前記媒体は、

前記雑音のある入力信号における過渡現象を検索することと、

該過渡現象が存在しない場合に背景雑音推定を更新することと、

該過渡現象の検出の後、前記整合低雑音信号テンプレートを決定し、該入力信号の一部を置き換えることと

を前記プロセッサに行わせる命令をさらに格納する、請求項 17 に記載のコンピュータ読み取り可能な格納媒体。

50

【発明の詳細な説明】

【0001】

(技術分野)

本発明は、付加雑音によって劣化された音響信号の質を高めるためのシステムおよび方法に関する。

【0002】

(背景)

音響信号の高品質化を調査するいくつかの研究分野があり、スピーチ信号に重きが置かれている。これらには、音声通信、自動音声認識(ASR)、補聴器がある。各研究分野は音響信号高品質化に対して独自の手法を採用し、それらの間に何らかの重なりがある。

10

【0003】

音響信号は、雑音があることによってしばしば劣化する。たとえば、にぎやかな事務所や移動中の自動車の中では、ASRシステムの性能が実質的に劣化する。遠隔会議システム内のように音声は遠隔の聞き手に伝送される場合は、雑音があると聞き手にとって不快であるとともに気が散ることもあり、さらには語音を理解するのが困難になる可能性もある。聴覚障害を有する人は、騒々しい環境内で語音を理解するのが著しく困難であり、最新の補聴器によって信号に加えられる全利得は問題を解決する役には立っていない。古い音楽記録は、瞬間的な雑音またはヒス(hissing)が存在することによってしばしば劣化している。雑音による音響信号劣化が発生する通信の他の例は、電話、無線通信、ビデオ会議、コンピュータ記録などを含む。

20

【0004】

連続音声大語彙ASRは特に雑音妨害に弱く、これまで業界が採用している解決策は、ヘッドセット・マイクロフォンの使用であった。雑音低減は、マイクロフォンと被験者の口の近接(約1.5インチ(38.1mm))によって達成され、特別な近接効果マイクロフォンによる場合もある。しかし、ユーザは、ヘッドセットによってコンピュータに束縛されることをしばしばぎこちないと感じ、ひどく目立つ機器を着用するのを不快に感じる。ヘッドセットを使用する必要があることは、人と機械の即座の対話を妨げ、ASR技術の市場浸透にとって重大な障壁である。

【0005】

近接マイクロフォンに加えて、通信時の音響信号高品質化に対する従来の手法は、適応フィルタおよびスペクトル・サブトラクションであった。適応フィルタでは、第2のマイクロフォンが信号ではなく雑音をサンプルする。次いで雑音を信号から減ずる。この手法の1つの問題は、重要なソースを拾うために使用するものと異なる場所で位置決めする必要がある第2マイクロフォンのコストである。さらに、雑音だけをサンプルし、所望のソース信号を含まないということがほとんどできない。適応フィルタの他の形態は、信号にバンドパス・デジタル・フィルタを適用する。フィルタのパラメータは、雑音スペクトルを長期間にわたって平均して信号対雑音比(SNR)が最大になるように適合される。この方法は、低SNRの帯域内で信号が取り残されるという欠点を有する。

30

【0006】

スペクトル・サブトラクションでは、雑音を、信号がない期間中に推定し、次いで信号が存在するとき信号スペクトルから減ずる。しかし、これは「ミュージカル・ノイズ」および不自然な他の歪みの導入を引き起こす。これらの問題の根元は、SNRの非常に低い領域内で、スペクトル・サブトラクションが、信号が一定レベルより低いことを決定できるにすぎないということである。不十分な場合もある証拠に基づいて信号レベルを選択せざるを得ないことにより、本当の信号からの少なからぬ乖離が、雑音および歪みの形態でしばしば発生する。

40

【0007】

雑音低減に対する最近の手法は、マイクロフォンのアレイを使用するビームフォーミングの使用である。この技法は、複数のマイクロフォン、A/Dコンバータなど専用化されたハードウェアを必要とし、したがってシステムのコストを引き上げる。信号処理コストは

50

、マイクロフォンの数の2乗に比例して増加するため、そのコストもまた高額になる。マイクロフォン・アレイの他の限界は、ビームフォーミング・プロセス全体にわたって依然としていくらかの雑音が漏れることである。さらに、実際のアレイ利得は通常、妨害音ソースの反響および残響が依然としてアレイの主ローブおよびサイドローブを介して受け入れられるため、無響条件で測定されたもの、あるいは理論から予測されたものよりもさらに低くなる。

【0008】

本発明者は、スペクトルの一部を取り残したり、不自然な雑音を導入したり、信号を歪ませたりすることなく、またマイクロフォン・アレイに出費することなく音響信号を高品質化することができることが望ましいと考えた。本発明は、従来技法の限界を回避する音響信号強調のためのシステムおよび方法を提供する。

10

【0009】

(概要)

本発明は、雑音のレベルが非常に低い出力信号を生成するよう入力信号を処理することにより、音響信号の質を高めるための方法、装置、およびコンピュータ・プログラムを含む(「信号」は品質向上させるべき対象信号そのものを意味し、それに対して背景音および気の散る音は「雑音」と称する)。好ましい実施形態では、学習によって向上された信号モデルの使用によって高品質化する。入力信号は人の語音を表すことができるが、本発明は、楽器および鳥や人の歌声など、どのタイプの生音または記録音響データも向上させるために使用できることを理解されたい。

20

【0010】

本発明の好ましい実施形態は、入力信号を以下のように強調する。すなわち、入力信号を、時間-周波数表現に変換された2進データにデジタル化する。背景雑音を推定し、過渡的な音を隔離する。信号検出器を過渡音に適用する。信号内容のない長い過渡音および過渡音間の背景雑音を雑音推定に含める。過渡音の少なくとも何らかの部分が重要な信号(対象信号)を含む場合は、再スケール後にその信号のスペクトルを信号モデルと比較し、信号のパラメータをデータに合わせる。低雑音信号を、信号モデル・パラメータの最適組を使用して再合成する。信号モデルは低雑音信号を組み込んでいるだけであるため、出力信号もまた雑音が少ない。テンプレートが既存のテンプレートと著しく異なる場合はスペクトログラムからテンプレートを作成することによって信号モデルを低雑音信号データで調整する。既存のテンプレートが入力パターンに似ていることが判明した場合は、得られるテンプレートが過去にそのテンプレートと整合されたすべてのスペクトルの平均になるような形でテンプレートをそのパターンで平均化する。したがって、モデルに組み込まれた信号特性の知識は、信号の再現を収斂するように働き、それによって不自然な雑音または歪みの導入が回避される。

30

【0011】

本発明は、以下の利点を有する。すなわち、瞬間的および静的な雑音のない再合成信号データを出力でき、入力信号のソースとして単一のマイクロフォンが必要なだけであり、低SNRの領域内の出力信号は、ソースが生成可能なスペクトルと矛盾しないように保たれる。

40

【0012】

本発明の1つまたは複数の実施形態の詳細は、添付図面および下記の説明で述べる。本発明の他の特徴、目的、利点は、説明および図面から、また特許請求の範囲から明らかになるう。

【0013】

様々な図面内の類似の参照番号および指定は、類似の要素を示す。

【0014】

(詳細な説明)

この説明全体にわたって、図示された好ましい実施形態および例は、本発明を限定するものではなく、典型と見なすべきである。

50

【 0 0 1 5 】

動作環境の概観

図 1 は、本発明の信号向上システムを実施するために使用することができる典型的な従来技術のプログラム可能な処理システムのブロック図である。音響信号はトランスデューサ・マイクロフォン 10 部で受け取られ、これが音響信号を表す対応電気信号を生成する。次いで、トランスデューサ・マイクロフォン 10 からの信号は、アナログ・デジタル・コンバータ 14 によってデジタル化する前に、増幅器 12 によって増幅するのが好ましい。アナログ・デジタル・コンバータ 14 の出力は、本発明の品質向上技法を適用する処理システムに加えられる。処理システムは、CPU 16、RAM 20、ROM 18（フラッシュ ROM など書き込み可能であってよい）、および図のように CPU バス 23 によって結合された磁気ディスクなど任意選択の記憶装置 22 を含むのが好ましい。品質向上プロセスの出力は、ASR システムなど他の処理システムに加えることも、ファイルに保存することも、聞き手のために再生することもできる。プレイバックは一般に、処理済みデジタル出力ストリームをデジタル・アナログ・コンバータ 24 によってアナログ信号に変換し、オーディオ・スピーカ 28（たとえば、スピーカ、ヘッドホン、またはイヤホン）を駆動する出力増幅器 26 でそのアナログ信号を増幅することによって行う。

10

【 0 0 1 6 】

システムの機能概観

以下、音響信号強調システムの機能構成要素について述べる。本発明の第 1 の機能構成要素は、入力データを時間・周波数表現に変換する動的な背景雑音推定器である。雑音推定器は、信号劣化を引き起こす連続的な、または緩やかに変わる背景雑音を推定する手段を提供する。雑音推定器はまた、雑音源が活動化された（たとえば、空調システムがオンまたはオフになった）場合など、雑音レベルの突然の変化に適應することができるべきである。動的背景雑音推定機能は、過渡的な音を背景雑音から分離し、背景雑音だけを推定することが可能である。一実施形態では、パワー検出器が複数の周波数帯域のそれぞれで動作する。データの雑音だけの部分を使用して、デシベル（dB）単位で雑音の平均および標準偏差を生成する。パワーが、周波数帯域内の指定数の標準偏差を超えて平均を上回った場合は、対応する時間は信号を含むものとして示され（flagged）、雑音だけのスペクトルを推定するために使用されない。

20

【 0 0 1 7 】

動的背景雑音推定器は、第 2 の機能構成要素の過渡音検出器と密接に動作する。過渡音は、比較的短い時間内に音響パワーが上昇して下降するとき発生する。過渡音は発声された語音とすることができるが、衝撃音、ドアを激しく閉じる音など過渡的な雑音とすることもできる。過渡音の隔離は、過渡音を別々に調査し、信号事象と非信号事象に分類することを可能にする。また、新たな雑音源がオンになった場合など、パワー・レベルの上昇が永続的であるときを認識するのに有効である。これは、システムがその新たな雑音レベルに適應することを可能にする。

30

【 0 0 1 8 】

本発明の第 3 の機能構成要素は信号検出器である。信号検出器は、非信号の非静的雑音を弁別するのに有効である。高調波の場合もまた、聞き手が再現信号を聞くことが望ましい場合にこれを使用してピッチ推定を行う。以下、雑音が存在する中で音声を検出する信号検出器の好ましい実施形態を述べる。音声検出器は、周波数領域内で声門パルス検出を使用する。データのスペクトログラムを生成し（信号の時間・周波数表現）、スペクトルの対数をとった後で、信号を周波数閾値まで時間軸に沿って合計する。得られた時系列の高い自己相関は、音声化された語音を表す。音声のピッチは、自己相関が最大になるラグである。

40

【 0 0 1 9 】

第 4 の機能構成要素はスペクトル再スケーラである。入力信号は、弱いことも強いことも、近いことも遠いこともある。測定されたスペクトルをモデル内でテンプレートと整合する前に、パターン間の距離が信号の全音量に依存しないように測定されたスペクトルを再

50

スケールする。好ましい実施形態では、重み付けがデシベル (dB) 単位の SNR に比例する。重みは、それぞれ最小値および最大値によって下および上の境界となる。スペクトルは、記憶されたテンプレートそれぞれへの重み付けされた距離が最小になるように再スケールされる。

【0020】

第5の機能構成要素はパターン整合器である。テンプレートと測定されたスペクトルの間の距離は、ユークリッド距離または加重ユークリッド距離など、いくつかの適切な測定基準の1つとすることができる。測定されたスペクトルまでの最小距離を有するテンプレートが、最適な原型として選択される。信号モデルは、低雑音信号から得られた1組の短期間の原型スペクトログラムからなる。信号モデルの調整は、先に収集した原型から著しく異なるスペクトログラムを収集することによって行う。第1原型は、雑音より著しく上の信号を含む第1信号スペクトログラムである。後続の時間エポックについては、スペクトログラムが、選択された距離閾値より既存の原型に近い場合に、スペクトログラムを最も近い原型で平均化する。スペクトログラムが、選択された閾値より原型から離れている場合には、スペクトログラムを新しい原型として宣言する。

10

【0021】

第6の機能構成要素は低雑音スペクトログラム生成器である。低雑音スペクトログラムは、低SNRスペクトログラム・ビン内のデータを最適な原型の値で置き換えることにより、パターン整合器によって生成された雑音の多いスペクトログラムから生成される。高SNRスペクトログラム・ビンでは、測定されたスペクトルが変化しないままとされる。原型と測定された信号を混合したものが、中間SNRケースで使用される。

20

【0022】

第7の機能構成要素は再合成器である。出力信号は、低雑音スペクトログラムから再合成される。以下、好ましい一実施形態に移る。信号は、高調波部分と非高調波部分に分けられる。高調波部分の場合は、各成分について任意の初期位相が選択される。次いで、非ゼロ出力の各点について、各成分の振幅をスペクトログラムから補間し、基本周波数を信号検出器の出力から補間する。各成分を、それぞれ連続位相、振幅、およびその周波数間の高調波関係によって別々に合成する。高調波部分の出力は、成分の合計である。

【0023】

非高調波部分の場合は、再合成された時系列の基本周波数が、信号の基本周波数をたどる必要がない。一実施形態では、基本周波数を一定に保つことを除いて、高調波部分の場合のように連続振幅および位相再現を実行する。他の実施形態では、信号の各周波数帯域について1つずつ雑音生成器を使用し、振幅は、補間を介して低雑音スペクトログラムのものをたどっている。さらに他の実施形態では、バンドパス済み雑音の一定振幅ウィンドウを、その全振幅をその時点のスペクトログラムのものに調節した後で追加する。

30

【0024】

基本方法の概観

図2は、本発明の好ましい方法実施形態の流れ図である。図2に示す方法は、図1に示すアナログ・デジタル・コンバータ14からの出力として生成された複数のデータ・サンプルからなる着信音響信号を高品質化するために使用する。この方法は、「開始」状態で始まる(ステップ202)。着信データ・ストリーム(たとえば、先に生成された音響データ・ファイルまたはデジタル化された生音信号)が、1組のサンプルとしてコンピュータ・メモリ内に読み取られる(ステップ204)。好ましい実施形態では、本発明が通常、連続音響データ・ストリームの一部分を表すデータの「移動するウィンドウ」を高品質化するために適用されることになり、データ・ストリーム全体が処理される。一般に、高品質化すべき音響データ・ストリームは、元の音響データ・ストリームの期間にかかわらず、一連の固定長のデータ「バッファ」として表される。

40

【0025】

現在のウィンドウのサンプルは、事前フィルタ、シェーディングなど適切な条件付けオペレーションを含むことができる時間・周波数変換を受ける(ステップ206)。短時間フ

50

ーリエ変換、フィルタ・バンク解析、離散ウェーブレット変換など、いくつかの時間 - 周波数変換のいずれかを使用することができる。

【 0 0 2 6 】

時間 - 周波数変換の結果は、初期時系列 $x(t)$ が時間 - 周波数表現 $X(f, i)$ に変換されることであり、ただし t は時系列 x のサンプリング・インデックス、 f および i はそれぞれ、スペクトログラム X の周波数および時間次元を指し示す離散変数である。好ましい実施形態では、別途指定しない限り、後続のステップで X の代わりに X の大きさの対数を使用する（ステップ 207）。すなわち $P(f, i) = 20 \log_{10}(|X(f, i)|)$

【 0 0 2 7 】

時間および周波数に応じたパワー・レベル $P(f, i)$ を、今後「スペクトログラム」と称する。

【 0 0 2 8 】

次いで、個々の帯域 f 内のパワー・レベルが、過渡音隔離（ステップ 210）と結合された背景雑音推定（ステップ 208）を受ける。過渡音隔離は、静的雑音内に埋もれた過渡信号の存在を検出し、そのような過渡音の推定開始時間および終了時間を出力する。過渡音は探索信号のインスタンスとすることができるが、瞬間的雑音とすることもできる。背景雑音推定は、過渡音間で背景雑音パラメータの推定を更新する。

【 0 0 2 9 】

背景雑音推定を実行するための好ましい実施形態は、各周波数帯域について移動するウィンドウ内の音響パワーを平均化するパワー検出器を含む。所定の数の周波数帯域内のパワーが、背景雑音より上で一定数の標準偏差として決められた閾値を上回った場合は、パワー検出器が信号の存在を宣言する。すなわち、次式のときである。

$$P(f, i) > B(f) + c(f)$$

ただし、 $B(f)$ は帯域 f 内の平均背景雑音パワー、 $c(f)$ は同じ帯域内の雑音の標準偏差、 c は定数である。代替実施形態では、雑音推定が動的である必要はなく、1回で測定できよう（たとえば、本発明を実施するソフトウェアが動作するコンピュータの起動中）。

【 0 0 3 0 】

次いで、過渡音検出器を通過する変換済みデータが信号検出器機能に加えられる（ステップ 212）。このステップは、信号と同じクラスでない過渡的雑音を弁別することを可能にする。語音（スピーチ）の高品質化の場合は、音声検出器をこのステップで適用する。具体的には、好ましい音声検出器では、レベル $P(f, i)$ が最小および最大周波数、それぞれ $lowf$ および $topf$ の間で時間軸に沿って合計される。

【 数 1 】

$$b(i) = \sum_{f=lowf}^{topf} P(f, i)$$

【 0 0 3 1 】

次いで、 $b(i)$ の自己相関を、 $maxpitch$ から $minpitch$ のタイム・ラグ に応じて計算する。ただし、 $maxpitch$ は許容される最大音声ピッチに対応するラグであり、一方 $minpitch$ は許容される最小音声ピッチに対応するラグである。音声 / 無音声の決定のために基づかれる統計値は、時間 i で中心付けられたウィンドウ内で計算される $b(i)$ の正規化自己相関（自己相関係数）の値である。最大正規化自己相関が閾値より大きい場合は、音声を含むものと考えられる。この方法は、短時間スペクトログラム内に現れる声門パルスによって特徴付けられる、人の声の振動する性質を利用する。これらの声門パルスは、スペクトログラムの周波数次元に沿って並ぶ。音声周波数の少なくとも何らかの領域を占めている場合は、合計の自己相関が、その音声に対応するピッチ期間の値で最大を示す。この音声検出方法の利点は、 $b(i)$ の自己相関係数が高くなるためにスペクトル

10

20

30

40

50

の部分全体にわたって S N R が良好であることだけが必要であるため、スペクトルの大部分にわたって雑音妨害に強いことである。

【 0 0 3 2 】

音声検出器の他の実施形態は、低 S N R の周波数帯域ビンの影響を低減するため、スペクトログラム要素を合計する前に重み付けする。

【 数 2 】

$$b(i) = \sum_{f=\text{lowf}}^{\text{topf}} P(f,i)w'(f,i)$$

10

【 0 0 3 3 】

重み $w(i)$ は、時間 i の帯域 f 内の S N R $r(f,i)$ に比例し、レベルの差、すなわち各周波数帯域について $r(f,i) = P(f,i) - B(f)$ で計算される。この実施形態では、再スケール係数の各要素が、以下のように定義される重みによって重み付けされる。ただし、 w_{\min} および w_{\max} はプリセット閾値である。

$w(f,i) = w_{\min}$ 、 $r(f,i) < w_{\min}$ の場合

$w(f,i) = w_{\max}$ 、 $r(f,i) > w_{\max}$ の場合

$w(f,i) = r(f,i)$ 、その他の場合

【 0 0 3 4 】

好ましい実施形態では、重みは、各時間枠で重みの合計によって正規化される。すなわち

20

$w'(f,i) = w(f,i) / \text{sum}_f(w(f,i))$

$w'_{\min} = w_{\min} / \text{sum}_f(w(f,i))$

$w'_{\max} = w_{\max} / \text{sum}_f(w(f,i))$

【 0 0 3 5 】

次いで、ステップ 2 0 8 および 2 1 0 からのスペクトログラム P は、記憶されているテンプレートと比較できるように再スケールするのが好ましい (ステップ 2 1 4)。このステップを実行する 1 つの方法は、スペクトログラム $P(f,i)$ の各要素を定数 $k(i,m)$ で上げ、 $P(f,i) + k(i,m)$ と第 m 番目のテンプレート $T(f,m)$ との間の平方 2 乗平均差が最低になるようにすることである。これは、以下をとることによって行

30

う。ただし、 N は周波数帯域の数である。

【 数 3 】

$$k(i,m) = \frac{1}{N} \sum_{f=1}^N [P(f,i) - T(f,m)]$$

【 0 0 3 6 】

他の実施形態では、比較に先立ちテンプレートを再スケールする際に重み付けを使用する。

【 数 4 】

40

$$k(i,m) = \frac{1}{N} \sum_{f=1}^N [P(f,i) - T(f,m)]w'(f,i)$$

【 0 0 3 7 】

このような再スケールの効果は、S N R の高いテンプレートの周波数帯域を優先的に整列させるためである。しかし、再スケールは任意選択であり、すべての実施形態で使用するには及ばない。

【 0 0 3 8 】

他の実施形態では、テンプレートを再スケールするために、テンプレートの S N R ならび

50

に測定されたスペクトルのSNRが使用される。テンプレート $T(f, m)$ のSNRは、 $r_N(f, m) = T(f, m) - B_N(f)$ で定義される。ただし、 $B_N(f)$ は調整時の周波数帯域 f の背景雑音である。 r および r_N を使用する重み付け方式の一実施形態では、重み w_N が、テンプレートおよびスペクトログラムの重みの積の平方根と定義される。

【数5】

$$\begin{aligned} \sqrt{r_N(f, m)r(f, i)} < w_{\min} \text{ の場合、 } w_2(f, i, m) &= w_{\min} \\ \sqrt{r_N(f, m)r(f, i)} > w_{\max} \text{ の場合、 } w_2(f, i, m) &= w_{\max} \\ \text{その他の場合、 } w_2(f, i, m) &= \sqrt{r_N(f, m)r(f, i)} > w_{\max} \end{aligned}$$

10

【0039】

r_N と r の他の組み合わせも許容可能である。好ましい実施形態では、重みは、各時間枠で重みの合計によって正規化される。すなわち

$$\begin{aligned} w'_2(f, i) &= w_2(f, i) / \sum_f (w_2(f, i)) \\ w'_{\min} &= w_{\min} / \sum_f (w_2(f, i)) \\ w'_{\max} &= w_{\max} / \sum_f (w_2(f, i)) \end{aligned}$$

【0040】

スペクトルの再スケール後、好ましい実施形態は、現在のスペクトログラム $P(f, i)$ に最適に整合する信号モデル内のテンプレート T^* を見つけるようにパターン整合を行う（ステップ216）。「最適整合」という用語の定義、ならびに最適整合を見つけるために使用する方法にはいくらかの自由度がある。一実施形態では、 $P + k$ と T^* の間の最も小さいRMS（平方2乗平均）差 d^* を有するテンプレートを見つける。好ましい実施形態では、重み付けされたRMS距離を使用する。ただし、

20

【数6】

$$d(i, m) = \frac{1}{N} \sum_{f=1}^N [P(f, i) + k(i, m) - T(f, m)]^2 w'_2(f, i, m)$$

【0041】

この実施形態では、最低SNRの周波数帯域は、より高いSNRの周波数帯域より距離計算への影響が少ない。時間 i での最適整合テンプレート $T^*(i)$ は、 $d^*(i) = \min_m (d(i, m))$ となるように m を見つけることによって選択される。

30

【0042】

次いで、低雑音スペクトログラム C が、選択された最も近いテンプレート T^* に測定されたスペクトル P を合併することによって生成される（ステップ218）。各ウィンドウ位置 i について、低雑音スペクトログラム C が P および T^* から再現される。好ましい実施形態では、以下の形で再現が行われる。各時間-周波数ビンについて、

$$C(f, i) = w'_2(f, i) P(f, i) + [w'_{\max} - w'_2(f, i)] T^*(f, i)$$

40

【0043】

低雑音スペクトログラム C を生成した後で、低雑音出力時系列を合成する（ステップ220）。好ましい実施形態では、スペクトログラムが高調波（ y_h ）と非高調波（ y_u ）の部分に分けられ、各部が別々に再現される（ $y = y_h + y_u$ ）。高調波部分は、一連の高調波 $c(t, j)$ を使用して合成される。任意の初期位相 $\phi_0(j)$ が各成分 j について選択される。次いで、各出力点 $y_h(t)$ について各成分の大きさがスペクトログラム C から補間され、基本周波数 f_0 が音声検出器の出力から補間される。成分 $c(t, j)$ は、それぞれ連続位相、振幅、および他の成分との共通ピッチ関係によって別々に合成される。すなわち

$$c(t, j) = A(t, j) \sin[f_0 j t + \phi_0(j)]$$

50

ただし、 $A(t, j)$ は時間 t での各高調波 j の振幅である。一実施形態は、スプライン補間を使用して、スペクトログラム点の間でなめらかに変わる f_0 および $A(t, j)$ の連続値を生成する。

【0044】

出力の高調波部分は、成分の合計 $y_h(t) = \sum_j [c(t, j)]$ である。信号 y_u の非高調波部分の場合は、基本周波数が信号の基本周波数をたどる必要がない。一実施形態では、 f_0 を一定に保つことを除いて、高調波部分の場合のように連続振幅および位相再現を実行する。他の実施形態では、信号の各周波数帯域について1つずつ雑音生成器を使用し、振幅は、低雑音スペクトログラムのものをたどるようにされる。

【0045】

いずれかの入力データが処理されていない場合は(ステップ222)、音響データの次のサンプルについてプロセス全体を繰り返す(ステップ204)。そうでない場合は処理が終了する(ステップ224)。最終出力は、元の入力音響信号の質向上を示す低雑音信号である。

【0046】

背景雑音推定および過渡音隔離

図3は、図2のステップ212および208としてそれぞれ簡単に述べた背景雑音推定および過渡音検出のプロセスをさらに詳しく述べた流れ図である。過渡音隔離プロセスは、静的雑音に埋もれた過渡信号の存在を検出する。背景雑音推定器は、過渡音間で背景雑音パラメータの推定を更新する。

【0047】

このプロセスは、「プロセス開始」状態で始まる(ステップ302)。このプロセスは、十分な数の背景雑音のサンプルを必要とし、それから雑音の平均および標準偏差を使用して過渡音を検出することができる。それゆえに、ルーチンは、十分な数の背景雑音のサンプルが得られているかどうかを判定する(ステップ304)。得られていない場合は、現在のサンプルを使用して雑音推定を更新し(ステップ306)、プロセスが修了する(ステップ320)。背景雑音更新プロセスの一実施形態では、スペクトログラム要素 $P(f, i)$ がリング・バッファ内に保たれ、各周波数帯域 f 内の雑音の平均 $B(f)$ および標準偏差 $\sigma(f)$ を更新するために使用される。背景雑音推定は、インデックス i がプリセット閾値より大きい場合に準備が整ったと見なす。

【0048】

背景雑音サンプルの準備が整った場合は(ステップ304)、信号レベル $P(f, i)$ がいずれかの周波数帯域で背景雑音より著しく高いかが判定される(ステップ308)。好ましい実施形態では、所定の数の周波数帯域内のパワーが、背景雑音平均レベルより上で一定数の標準偏差として決められた閾値より大きい場合に、判定ステップが、パワー閾値を上回ったことを示す。すなわち、次式のときである。

$$P(f, i) > B(f) + c(f)$$

ただし、 c は経験的に所定の定数である。次いで、処理はステップ310で続く。

【0049】

スペクトログラム要素 $P(f, i)$ が過渡信号を含んでいるかどうかを判定するために、フラグ「In-possible-transient」が真にセットされ(ステップ310)、起こりうる過渡音の期間が増分される(ステップ312)。次いで、(起こりうる過渡音が)過渡音とするには長すぎるか否かが判定される(ステップ314)。可能な過渡期間がなおも最大期間内にある場合は、プロセスが終了する(ステップ320)。一方、過渡期間が長すぎて発声された言葉にならないと判断された場合は、背景雑音レベルの増加と考えられる。したがって、雑音推定が遡及的に更新され(ステップ316)、「In-possible-transient」フラグが偽にセットされ、かつ過渡期間が0にリセットされ(ステップ318)、処理が終了する(ステップ320)。

【0050】

ステップ308で十分強力な信号が検出されなかった場合は、背景雑音統計値がステップ

10

20

30

40

50

306で更新される。その後で、「In - possible - transient」フラグがテストされる(ステップ322)。フラグが偽にセットされている場合はプロセスが終了する(ステップ320)。フラグが真にセットされている場合は、ステップ318のように偽にリセットされ、過渡期間が0にリセットされる。次いで過渡音の期間がテストされる(ステップ324)。過渡音が短すぎて発声された言葉の一部にならないと考えられる場合は、プロセスが終了する(ステップ320)。過渡音が、可能な発声された語音とするのに十分長い場合は、過渡フラグが真にセットされ、過渡音の開始および終了が呼出しルーチンに渡される(ステップ326)。次いでプロセスが終了する(ステップ320)。

【0051】

パターン整合

図4は、図2のステップ216として簡単に述べたパターン整合のプロセスをさらに詳しく述べた流れ図である。このプロセスは、「プロセス開始」状態で始まる(ステップ402)。パターン整合プロセスは、熟考されたスペクトログラム $P(f, i)$ に最適に整合する信号モデル内のテンプレート T^* を見つける(ステップ404)。パターン整合プロセスはまた、信号モデルの学習プロセスを受け持つ。「最適整合」という用語の定義、ならびに最適整合を見つけるために使用する方法にはいくらかの自由度がある。一実施形態では、 $P + k$ と T^* の間の最も小さいRMS差 d^* を有するテンプレートを見つける。好ましい実施形態では、重み付けされたRMS距離を使用して整合の度合いを測定する。一実施形態では、RMSが次式によって計算される。

【数7】

$$d(i, m) = \frac{1}{N} \sum_{f=1}^N [P(f, i) + k(i, m) - T(f, m)]^2 w'_2(f, i, m)$$

【0052】

この実施形態では、最低SNRの周波数帯域は、より高いSNRの周波数帯域より距離計算への影響が少ない。時間 i でステップ404の出力である最適整合テンプレート $T^*(f, i)$ は、 $d^*(i) = \min_m [d(i, m)]$ となるように m を見つけることによって選択される。システムが学習モードでない場合は(ステップ406)、 $T^*(f, i)$ は最も近いテンプレートとしてプロセスの出力でもある(ステップ408)。次いでプロセスが終了する(ステップ410)。

【0053】

システムが学習モードにある場合は(ステップ406)、 $P(f, i)$ に最も似ているテンプレート $T^*(f, i)$ が使用されて信号モデルが調節される。 $T^*(f, i)$ がモデル内に組み込まれる方法は、 $d^*(i)$ の値に応じて決まる(ステップ412)。 d_{max} が所定の閾値であり、 $d^*(i) < d_{max}$ の場合は、 $T^*(f, i)$ が調節され(ステップ416)、プロセスが終了する(ステップ410)。ステップ416の好ましい実施形態は、 $T^*(f, i)$ が、 $T^*(f, i)$ を構成するために使用されるすべてのスペクトル $P(f, i)$ の平均となるように実施される。好ましい実施形態では、 $T(f, m)$ に関連する

スペクトルの数 n_m がメモリに保たれ、新たなスペクトル $P(f, i)$ を使用して $T(f, m)$ を調節する場合は、調節されたテンプレートが

$T(f, m) = [n_m T(f, m) + P(f, i)] / (n_m + 1)$

であり、テンプレート m に対応するパターンの数も次のように調節される。

$$n_m = n_m + 1$$

【0054】

ステップ412に戻り、 $d^*(i) > d_{max}$ の場合は、新しいテンプレートが作成され(ステップ414)($T^*(f, i) = P(f, i)$ 、重み $n_m = 1$)、プロセスが終了する(ステップ410)。

【0055】

10

20

30

40

50

コンピュータの実施

本発明は、ハードウェアでもソフトウェアでも、あるいは両方の組み合わせでも実施することができる（たとえば、プログラマブル・ロジック・アレイ）。別途指定しない限り、本発明の一部として含まれるアルゴリズムは、どの特定のコンピュータまたは他の装置にも本質的に関連付けられていない。具体的には、様々な汎用機を本明細書の教示に従って記述されたプログラムと共に使用することができ、あるいはより専用化された装置を構築して、必要とされる方法ステップを実行することがより好都合である可能性がある。しかし、本発明は、それぞれが少なくとも1つのプロセッサ、少なくとも1つのデータ記憶システム（揮発性および不揮発性メモリおよび/または記憶要素を含む）、少なくとも1つの入力装置、少なくとも1つの出力装置を備えるプログラム可能なシステム上で実行される1つまたは複数のコンピュータ・プログラム内で実施されることが好ましい。このようなプログラム可能なシステム構成要素はそれぞれ、一機能を実行するための手段を構成する。プログラム・コードはプロセッサ上で実行され、本明細書に記載された機能を実行する。

10

【0056】

このようなプログラムはそれぞれ、コンピュータ・システムと通信するために所望のコンピュータ言語（機械語、アセンブリ、上位手続き言語、オブジェクト指向プログラミング言語を含む）で実施することができる。いかなる場合でも、言語はコンパイラ型言語とすることもインタープリタ型言語とすることもできる。

【0057】

このようなコンピュータ・プログラムはそれぞれ、汎用または専用のプログラム可能なコンピュータ可読記憶媒体または装置（たとえば、ROM、CD-ROM、または磁気もしくは光媒体）上に記憶され、記憶媒体または装置がコンピュータによって読み取られた際にコンピュータを構成し、かつ動作させて、本明細書に記載された手順を実行することが好ましい。本発明のシステムはまた、コンピュータ・プログラムで構成されたコンピュータ可読記憶媒体として実施され、そのように構成された記憶媒体が、コンピュータを特定の事前定義された形で動作させて、本明細書に記載された機能を実行すると見なすことができる。

20

【0058】

以上、本発明のいくつかの実施形態について述べた。しかしながら、本発明の精神および範囲から逸脱することなく、様々な修正を加えることができることを理解されたい。たとえば、様々なアルゴリズムのいくつかのステップは順番に依存しないものとすることができ、したがって上述した以外の順番で実行することができる。それゆえに、他の実施形態が以下特許請求の範囲内にある。

30

【図面の簡単な説明】

【図1】 本発明の信号向上技法を実施するために適した従来技術のプログラム可能なコンピュータ・システムのブロック図である。

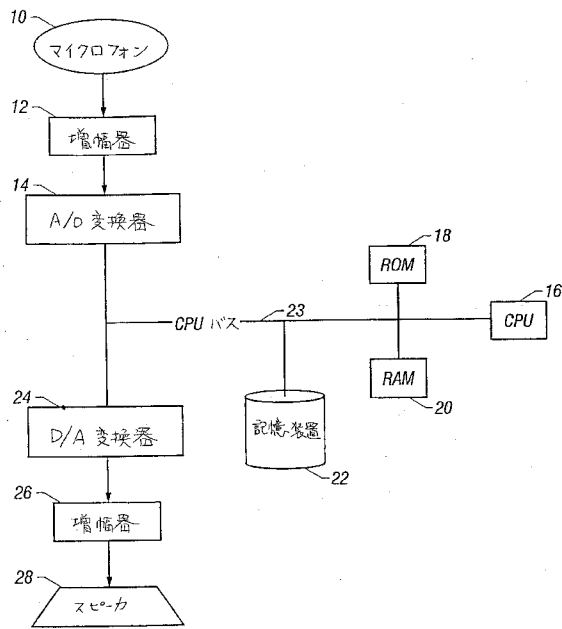
【図2】 本発明の好ましい実施形態の基本方法の流れ図である。

【図3】 入力データ内の過渡音を検出および隔離し、背景雑音パラメータを推定するための好ましい工程の流れ図である。

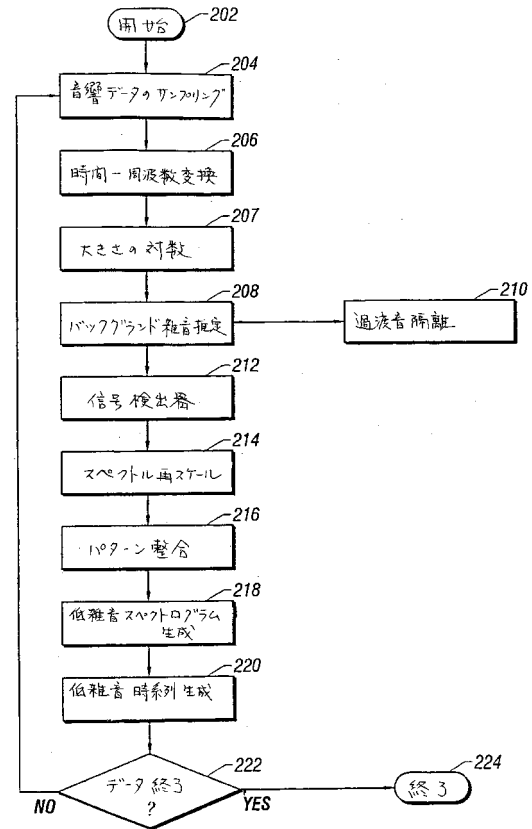
40

【図4】 信号モデル・テンプレートを生成および使用するための好ましい方法の流れ図である。

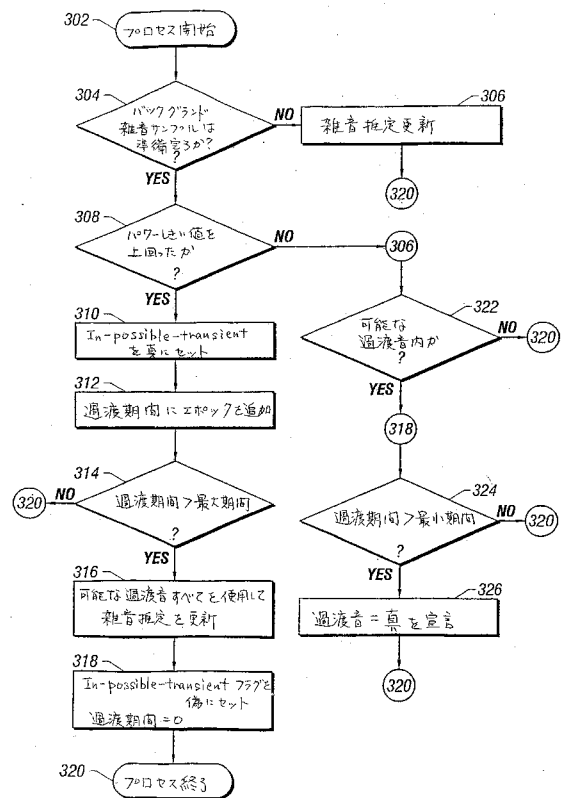
【 図 1 】



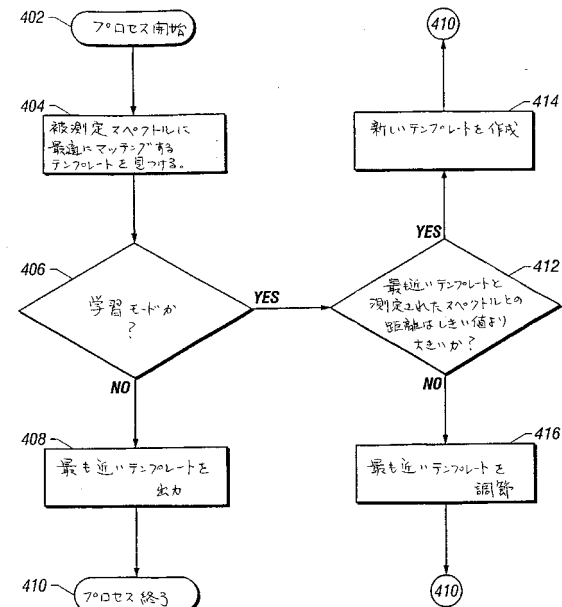
【圖 2】



【 図 3 】



【 図 4 】



フロントページの続き

- (56)参考文献 特開平 0 9 - 2 5 8 7 6 8 (J P , A)
特開平 0 9 - 2 1 2 1 9 6 (J P , A)
特開平 0 8 - 2 2 3 0 5 2 (J P , A)
特開平 0 7 - 2 6 1 8 0 0 (J P , A)
特開平 0 5 - 0 7 3 0 9 0 (J P , A)
特開平 0 4 - 2 5 7 8 9 8 (J P , A)
特開平 1 0 - 0 4 9 1 9 7 (J P , A)
特開平 1 0 - 3 0 4 0 2 3 (J P , A)
国際公開第 9 9 / 0 0 4 5 0 5 (WO , A 1)
特開平 0 9 - 2 5 8 7 8 3 (J P , A)

(58)調査した分野(Int.Cl. , D B 名)

G10L 15/00-17/00
G10L 21/00-21/06
G10L 11/00-11/06
H04R 25/00