(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第3626492号
(P3626492)
(45)発行日　平成17年3月9日(2005.3.9)　(24)登録日　平成16年12月10日(2004.12.10)

(51)Int.Cl.[7]　F I
G10L 21/02　G10L 9/00　F
G10L 13/00　G10L 7/00　D

請求項の数 13　(全 15 頁)

(21)出願番号　特願平7-504026
(86)(22)出願日　平成6年6月6日(1994.6.6)
(65)公表番号　特表平9-503590
(43)公表日　平成9年4月8日(1997.4.8)
(86)国際出願番号　PCT/US1994/006367
(87)国際公開番号　WO1995/002288
(87)国際公開日　平成7年1月19日(1995.1.19)
審査請求日　平成13年6月6日(2001.6.6)
(31)優先権主張番号　08/086,707
(32)優先日　平成5年7月7日(1993.7.7)
(33)優先権主張国　米国(US)

(73)特許権者　500080720
ポリコム・インコーポレイテッド
アメリカ合衆国、９５０３５　カリフォルニア州　ミルピタス　バーバー・レーン　１５６５
(74)代理人　100070150
弁理士　伊東　忠彦
(72)発明者　ヘルフ　ブラント　マーティン
アメリカ合衆国　マサチューセッツ州　０２１７６　メルローズ　マウント　ヴァーノン　ストリート　８３
(72)発明者　チュー　ピーター　エル
アメリカ合衆国　マサチューセッツ州　０２１７３　レキシントン　ハードリー　ロード　７

最終頁に続く

(54)【発明の名称】会話の品質向上のための背景雑音の低減

(57)【特許請求の範囲】

【請求項１】

雑音成分を有する入力音響信号に背景雑音の知覚されるリアルタイム抑制を行うための装置であって、

前記入力音響信号を一連の音響信号フレームに分割するフレーム化装置、

現在のウィンドウ化された音響信号フレームを生成し、一つの音響信号フレームのすべてを前記一つの音響信号フレームに時間的に直前に先行する前記音響信号フレーム内のいくつかの信号と結合するウィンドウ化装置、

前記現在のウィンドウ化された音響信号フレームから１つのグループの周波数スペクトル成分を得る変換器、

前記周波数スペクトル成分を使用して、前記周波数スペクトル成分の雑音量の雑音評価値を生成する雑音評価器、

前記雑音評価値および前記周波数スペクトル成分に基づいて、現在の利得乗法因数を生成する雑音抑制スペクトル修正器、

人間の耳の感度に応じた固定数のフレーム分、前記一連の音響信号フレームの前記周波数スペクトル成分を遅延させ、遅延した周波数スペクトル成分を生成する遅延器、

前記現在のフレームを用いて生成された前記現在の利得乗法因数に基づいて一連の先行のフレームのうち直前の一フレームの前記遅延した周波数スペクトル成分を減衰させ、雑音の低減された周波数成分を生成する制御減衰器、および

前記雑音の低減された周波数成分を時間領域に変換する逆変換器、

を備えている装置。
【請求項２】
請求項１記載の装置であって、
前記雑音抑制スペクトル修正器が、
前記周波数スペクトル成分の各周波数成分に対して、その周波数成分が雑音であるかどうかについての決定を行うグローバル判定機構、
前記周波数スペクトル成分の各周波数成分に対して、周波数成分が雑音成分である確度レベルを導出するローカル雑音判定機構、
前記確度レベルに基づいて、各周波数成分に対して前記利得乗法因数を決定する検出器、
前記利得乗法因数をスペクトル的および時間的に調整する分散機構、ならびに
前記周波数成分のスペクトルの谷間を検出および充填するスペクトル・バレー充填器、
を備えていることを特徴とする装置。
【請求項３】
請求項２記載の装置であって、
前記背景雑音評価器が、各周波数スペクトル成分に対して、対応する雑音評価値を生成するものであり、
前記ローカル雑音判定機構が、
（ａ）　周波数成分のそれぞれとその対応する雑音評価値との比、および
（ｂ）　前記グローバル判定機構によって行われた決定、
に基づいて、確度レベルを導出するものである、
ことを特徴とする装置。
【請求項４】
請求項２記載の装置であって、
前記分散機構が、前記利得乗法因数を前記確度レベルに基づいて調整するものである、ことを特徴とする装置。
【請求項５】
請求項１乃至３のいずれか一項記載の装置であって、
滑らかにされた時間領域の成分を生成して、前記雑音の低減された時間領域の成分の不連続性を最小化する後置ウィンドウ、および
前記滑らかにされた時間領域の成分の第１の部分を、滑らかにされた時間領域の成分の先に格納された部分と組み合わせて出力し、前記滑らかにされた時間領域の成分の前記第１の部分に含まれない部分からなる残りの部分を格納する重ね合わせ加算器、
をさらに備えている、ことを特徴とする装置。
【請求項６】
雑音成分を有する入力音響信号に背景雑音の知覚される抑制を行うための装置であって、
前記入力音響信号から抽出される信号のフレームから周波数スペクトル成分を得る変換器、
各周波数成分に対して乗法利得因数を決定する検出器、
前記乗法利得因数を調整して時間的及びスペクトル的に分散させる分散機構、ならびに
前記周波数成分を変換して、雑音の修正されたスペクトル信号を取り出す制御減衰器、
を備えている装置。
【請求項７】
雑音成分を有する入力音響信号に背景雑音の知覚される抑制を行うための装置であって、
前記入力音響信号を一連の音響信号フレームに分割するフレーム化装置、
現在のウィンドウ化された音響信号フレームを生成し、一つの音響信号フレームのすべてを前記一つの音響信号に時間的に直前に先行する前記音響信号フレーム内のいくつかの信号と結合するウィンドウ化装置、
前記ウィンドウ化された音響信号フレームから１つのグループの周波数スペクトル成分を得る変換器、
周波数スペクトル成分を使用して、前記周波数スペクトル成分の雑音量の雑音評価値を生

成する雑音表価器、
前記雑音評価値および前記周波数スペクトル成分に基づいて、現在の利得情報因数を生成する雑音抑制スペクトル修正器、
前記一連の音響信号フレームの前記周波数スペクトル成分を遅延させ、遅延した周波数スペクトル成分を生成、する遅延器、
前記現在のフレームを用いて生成された前記現在の利得乗法因数に基づいて、一連の先行のフレームうち直前の一フレームの前記遅延した周波数スペクトル成分を減衰させ、
前記雑音の低減された周波数成分を時間領域に変換する逆変換器、
を備え、
前記雑音抑制スペクトル修正器が更に、
前に生成された利得乗法因数を用いて現在の利得乗法因数を決定する手段と、
フレームの１つのグループの周波数スペクトル成分に対して、当該グループが雑音であるかどうかについての決定をするためのグローバル決定メカニズムと、
前記周波数スペクトル成分の各周波数成分、各成分の確度レベルに対して、当該周波数成分が雑音であるかどうかを導出するローカル雑音決定メカニズムと、
前記確度レベルに基づいて、各周波数成分についての初期利得乗法因数を決定する検出装置と、
スペクトル的および時間的に前記初期利得乗法因数を調整する分散メカニズムとを備えていることを特徴とする装置。
【請求項８】
入力音響信号の背景雑音の知覚を低減するための方法であって、
前記入力音響信号を一連の音響信号フレームに分割し、
ウィンドウ化された音響信号フレームを生成し、
一つの音響信号フレームの全てと前記一つの音響信号フレームに時間的に直前に先行する音響信号フレーム内のいくつかの信号とを結合し現在のフレームを獲得し、
前記ウィンドウ化された音響信号フレームから１つのグループの周波数スペクトル成分を獲得し、
前記周波数スペクトル成分を用いて、周波数スペクトル成分中の雑音の量の雑音評価値を生成し、
前記雑音評価値および前記周波数スペクトル成分に基づいて、現在の利得乗法因数を生成し、
人間の耳の感度に応じた固定数のフレーム分、前記一連の周波数スペクトル成分を遅延させて、遅延した周波数スペクトル成分を生成し、
前記現在のフレームを用いて生成された現在の利得乗法因数に基づいて一連の先行のフレームのうち直前の一フレームの前記遅延した周波数スペクトル成分を減衰させて、雑音の低減された周波数成分を生成し、そして
前記雑音の低減された周波数成分を時間領域に変換する、
ステップを備えている方法。
【請求項９】
請求項８記載の方法であって、
前記利得乗法因数を生成するステップが、
前記周波数スペクトル成分の各周波数成分に対して、その周波数成分が雑音であるかどうかについての決定を行い、
前記周波数スペクトル成分の各周波数成分に対して、周波数成分が雑音成分である確度レベルを導出し、
前記確度レベルに基づいて、各周波数成分に対して前記利得乗法因数を決定し、
前記利得乗法因数をスペクトル的および時間的に調整し、
前記周波数成分のスペクトルの谷間を検出および充填する、
ステップを備えている、ことを特徴とする方法。
【請求項１０】

請求項８又は９記載の方法であって、
後置ウィンドウ化を行って、滑らかにされた時間領域の成分を生成し、
前記滑らかにされた時間領域の成分の第１の部分を、滑らかにされた時間領域の成分の先に格納された部分と組み合わせて出力し、
前記滑らかにされた時間領域の成分の前記第１の部分に含まれない部分からなる残りの部分を格納する、
ステップをさらに備えている、ことを特徴とする方法。
【請求項１１】
雑音成分を有する入力音響信号の背景雑音の知覚を低減するための方法であって、
前記入力音響信号から抽出される信号のフレームから周波数スペクトル成分を得、
各周波数成分に対して乗法利得因数を決定し、
前記乗法利得因数を時間的及びスペクトル的に調整し、
前記周波数成分を変換して、雑音の修正されたスペクトル信号を取り出す、
ステップを備えている方法。
【請求項１２】
入力音響信号の背景雑音の知覚を低減するための方法であって、
前記入力音響信号を一連の音響信号フレームに分割し、
ウィンドウ化された音響信号フレームを生成し、
一つの音響信号フレームの全てと前記一つの音響信号フレームに時間的に直前に先行する音響信号フレーム内のいくつかの信号とを結合し現在のフレームを獲得し、
前記ウィンドウ化された音響信号フレームから１つのグループの周波数スペクトル成分を獲得し、
前記周波数スペクトル成分を用いて、周波数スペクトル成分中の雑音の量の雑音評価値を生成し、
前記雑音評価値および前記周波数スペクトル成分に基づいて、現在の利得乗法因数を生成し、
固定数のフレームによって、前記一連の周波数スペクトル成分を遅延させて、遅延した周波数スペクトル成分を生成し、
前記現在のフレームを用いて生成された現在の利得乗法因数に基づいて前のフレームの前記遅延した周波数スペクトル成分を減衰させて、雑音の低減された周波数成分を生成し、
そして
前記雑音の低減された周波数成分を時間領域に変換する、ステップとを備えており、
前記現在の利得乗法因数を生成するステップが、前に生成した前記現在の利得乗法因数を用いて前記現在の利得乗法因数を決定し、
フレームの前記周波数スペクトル成分の１つのグループに対して、そのグループが雑音かどうかについての決定をし、
前記周波数スペクトル成分の各周波数成分、各成分に対する確度レベルに対し、その周波数成分が雑音性分かどうかを導出し、
前記確度レベルに基づいて各周波数成分に対する初期利得乗法因数を決定し、そして
前記初期利得乗法因数をスペクトル的および時間的に調整するステップを備えている方法。
【請求項１３】
雑音成分を有する入力音響信号における背景雑音の知覚を低減するための方法であって、
前記入力音響信号から導出される音響信号フレームから周波数スペクトル成分を獲得し、
各周波数成分に対する乗法利得因数を決定し、該乗法利得因数を調整して時間的及びスペクトル的に分散させ、
前記調整された乗法利得因数に従って前記周波数成分を減衰させて雑音修正スペクトル信号を導出するステップとを備えたことを特徴とする方法。
【発明の詳細な説明】
発明の背景

本発明は、電話通信チャネルのようなチャネル上の音声情報の通信に関する。
音声送信システムに使用されるマイクロフォンは、典型的には、拾い集めるべき音声とともに、雑音と呼ばれる周囲の音、すなわち背景音を拾い集める。マイクロフォンから話し手まである距離がおかれている音声送信システム、例えば、映像音声電話会議環境で使用されるシステムにおいては、マイクロフォンによって拾い集められる会話に背景雑音が付加されるので、背景雑音は、音声品質の劣化の原因となる。それらの性質および使用目的によって、これらのシステムは、それらのマイクロフォンの周囲にある全ての位置からの音を拾い集めなければならず、これらの音は背景雑音を含んでいる。
HVACシステム、コンピュータ、および他の電子機器から発するファン雑音（fan noise）は、ほとんどの電話会議環境における支配的な雑音源である。
優れた雑音抑制技術は、背景雑音が知覚されることを低減する一方、同時に会話の質、すなわち会話の明瞭性に悪影響を与えない。一般に、本発明の目的は、一つのマイクロフォンによって拾われる会話に付加される狭帯域または広帯域の一定の雑音を抑制することにある。本発明の他の目的は、一つのマイクロフォンによって拾い集められる会話に付加されるファン雑音を低減することにある。

発明の要約

本発明の一特徴において、一般に、本発明は、入力音声信号の背景雑音を低減するための装置に関する。この装置は、入力音響信号を複数の信号フレームに分割するフレーム化装置、および前記信号フレームのそれぞれから雑音成分を除去して、フィルタリングされた信号フレームを生成するノッチ・フィルタ・バンクを特徴として備えている。乗算器は、結合された信号フレームを乗じ、ウィンドウ化された信号フレームを生成する。ここで、結合された信号フレームは、一つのフィルタリングされた信号フレームに時間的に直前に先行するフィルタリングされた信号フレーム内のいくつかの信号と結合された前記一つのフィルタリングされた信号フレーム内の全信号を含んでいる。変換器は、前記ウィンドウ化された信号フレームから周波数スペクトル成分を得、背景雑音評価器は、前記周波数スペクトル成分を使用して、前記周波数スペクトル成分の雑音量である雑音評価値を生成する。雑音制御スペクトル修正器は、前記雑音評価値および前記周波数スペクトル成分に基づいて利得乗法因数を生成する。遅延器は、前記周波数スペクトル成分を遅延させ、遅延した周波数スペクトル成分を生成する。制御減衰器は、前記利得乗法因数に基づいて前記周波数スペクトル成分を減衰させ、雑音の低減された周波数成分を生成する。逆変換器は、前記雑音の低減された周波数成分を時間領域に変換する。
好ましい実施形態においては、前記雑音抑制スペクトル修正器は、グローバル判定機構、ローカル判定機構、検出器、分散機構およびスペクトル・バレー充填器を備えている。
グローバル判定機構は、前記周波数スペクトル成分の各周波数成分に対して、その周波数成分が主に雑音であるかどうかについての決定を行う。ローカル雑音判定機構は、前記周波数スペクトル成分の各周波数成分に対して、周波数成分が主として雑音成分である確度レベルを導出する。検出器は、前記確度レベルに基づいて、各周波数成分に対して前記利得乗法因数を決定する。分散機構は、前記決定された利得乗法因数の影響をスペクトル的および時間的に分散し、スペクトル・バレー充填器は、その結果の周波数成分のスペクトルの谷間を検出し、充填する。
また、この好ましい実施形態の変形例として、前記背景雑音評価器は、各周波数スペクトル成分に対して雑音評価値を生成し、前記ローカル雑音判定機構は、周波数成分のそれぞれとその対応する雑音評価値との比、および前記グローバル判定機構によって行われた決定に基づいて確度レベルを導出する。
他の実施形態において、本発明は、後置ウィンドウおよび重ね合わせ加算機構を特徴としてさらに備えている。後置ウィンドウは、滑らかにされた時間領域の成分を生成して、前記雑音の低減された時間領域の成分の不連続性を最小化する。重ね合わせ加算機構は、前記滑らかにされた時間領域の成分の第１の部分を、滑らかにされた時間領域の成分の先に格納された部分と組み合わせて出力し、前記滑らかにされた周波数成分の前記第１の部分に含まれない部分からなる残りの部分を格納する。

本装置の好ましい実施形態において、背景雑音評価器は、それぞれが背景雑音評価値を生成する少なくとも二つの評価器、およびこれらの背景雑音評価値を比較し、その一つを選択する比較器を含んでいる。これらの評価器の一方は稼働最小評価器であり、他方の評価器は定常評価器である。
好ましい実施形態において、本装置は、ノッチ・フィルタ・バンクのためのノッチの位置を決定するノッチ・フィルタ機構も含んでいる。

【図面の簡単な説明】

図１は、本発明による雑音抑制システムのブロック図である。
図２～図４は、図１のブロック図の部分部分を実現する詳細なブロック図である。

好ましい実施形態の説明

世界中の何百万人もの人々によって日常使用されている最もシンプルな雑音抑制装置は、いわゆる「スケルチ」回路である。スケルチ回路は、ほとんどの市民バンド・２ウェイ・ラジオの標準である。この回路は、受信信号のエネルギーがあるしきい値より下がると、システムのラウドスピーカを単純に接続解除することにより作動する。このしきい値の値は、たいていは、マニュアルの制御ノブを使用して、遠端（far end）が静寂になるときに背景雑音がスピーカに渡らないようなレベルに固定される。この種の回路の問題は、遠端スピーカがスタートし、続いてストップするように回路がオン・オフすると、雑音が存在し、続いて消えることが明瞭に聞き取れるということである。雑音は広帯域で、小さな会話のエネルギーの存在する周波数をカバーするので、人が話しているのと同時に雑音が聞き取られる。どんなものでも雑音抑制の行われることが好ましいが、スケルチ・ユニットの動作は、非常に混乱した効果を生み出す。
本発明の雑音抑制方法は、音の会話と非会話の部分の双方における背景雑音を低減することによって、「スケルチ」コンセプトを著しく優れたものとする。
本発明によるアプローチは、人間の知覚作用に基づいている。スペクトル・マスクおよび時間マスク（双方ともに以下に定義する）の原理を使用して、本発明は、会話信号に付加または混合される雑音の知覚される大きさを低減する。
このアプローチは、例えば、抑制システムの処理後の会話出力と会話成分のみ（無雑音会話）との間の平均自乗誤差を最小化することが目的である他のアプローチとは異なる。
本発明で使用される方法は、チャネルのエネルギーがしきい値を超えると、そのチャネルの利得を上げ、チャネルのエネルギーがしきい値より下がると、利得を下げるという「スケルチ」の概念を利用するが、本方法は、異なる周波数領域において個別に処理を実行する。チャネルの利得は、入力信号のボリュームとそれに対応する出力信号のボリュームとの比であるとみなすことができる。
さらに、本方法は、スペクトル・マスクを行うさまざまな心理音響学の原理を利用する。特に、本方法は、ある周波数において大きな音があると、その周波数の回りには、他の信号を聞き取ることができない所与の周波数帯域（臨界帯域と呼ばれる）が存在することを基本的に述べている原理を利用する。すなわち、臨界帯域にある他の信号は聞こえない。本発明の方法は、会話が遠端から受信されている間、雑音の知覚作用を低減させる簡単な「スケルチ」回路よりもはるかに効果的である。
また、本発明の方法は、一時的なマスクを行う特性を利用する。大きな音のバーストが発生すると、そのバースト後の200ミリ秒までの期間中、バーストのスペクトル領域において耳の感度が減少する。別の音響効果は、バースト前の20ミリ秒までの時間の間、耳の感度が減少するというものである（このように、人間のヒアリングは、約20ミリ秒のパイプライン遅延を有する）。したがって、本発明の一つのキーとなる要素は、十分大きな信号の発生前後の期間中、雑音に対する耳の感度が減少するので、所与の帯域の利得をそれ以下に下げる信号しきい値を、その帯域における十分大きな信号の発生前後において、一定期間小さくすることができるというものである。

システムの概要

図１のブロック図を参照して、入力信号１は、まず、フレーマ（フレーム化装置）２によって分割され、20ミリ秒のサンプル・フレームにされる（入力信号は16kHzのレートでサ

ンプリングされるので、各20msのフレームは320個のサンプルを含む）。本方法の計算の複雑さは、一時に個々のサンプルに処理を行わずに、一時にサンプル・フレームのグループに処理を行うことにより大幅に減少する。続いて、フレームにされた信号は、ノッチ・フィルタ４のバンクに入力する。ノッチ・フィルタ４は、典型的にはモータの回転周波数で発生するモータ雑音のような狭帯域の雑音成分を除去するためのものである。ノッチが、十分疎らなスペクトル密度で十分狭いならば、会話の音質は悪影響を受けない。続いて、ディジタル信号の各フレームは、直前のディジタル信号フレームの終端からの一部と組み合わされ、ウィンドウ化フレームが生成される。

好ましい実施形態においては、ディジタル信号の各フレーム（20ms）は、先行するフレームの最後の12msと結合されて、32msの間隔を有するウィンドウ化フレームにされる。つまり、各ウィンドウ化フレームは、あるディジタル信号フレームからの320個のサンプルと、直前のフレームのフィルタリングされたサンプルのうちの最後の192個のサンプルとを組み合わせて備えている。続いて、この512個のサンプルからなる会話セグメントは、乗算器６においてウィンドウを乗じられて、512サンプル・フレームの開始部と終端部において信号の不連続性から生じる問題が軽減される。続いて、高速フーリエ変換（FFT）８が、512個のサンプルからなるウィンドウ化フレームに施され、257個の成分の周波数スペクトルが生成される。

変換される信号の最低周波数（D.C.）成分と最高周波数（サンプリング周波数を２で割った値、すなわち8kHz）成分は、実数部のみを有し、他の255個の成分は、実数部と虚数部の双方を有する。このスペクトル成分は、背景雑音評価器20に与えられる。この評価器は、背景雑音のスペクトル・エネルギーを評価し、ノッチ・フィルタ４のノッチを配置すべき背景雑音のスペクトルのピークを見つけ出すためのものである。各周波数成分に対する信号強度のスペクトル評価器である定常評価器（stationary estimator）24および背景雑音のスペクトル評価器である稼働最小評価器（running minimum estimator）22の評価値は、比較器28によって評価される。そして、特定の周波数成分が主に雑音からなるのか、信号と雑音の混ざったものからなるのかについての判定が、判定機構32により各周波数に対して行われることにより、さまざまな確度レベルが抽出される。これらの確度レベルに基づいて、周波数帯域の利得が、利得設定器34によって決定される。続いて、利得は、分散機構36によって、臨界帯域の周波数領域に分散され、スペクトル的にも時間的にも、心理音響学的なマスク効果が利用される。スペクトル・バレー充填器（spectral valley filler）38は、周波数成分利得関数におけるスペクトルの谷間（バレー:valley）を検出し、その谷間を埋めるために使用される。雑音圧縮スペクトル修正器30からの最終的な周波数成分利得関数は、減衰器12において、512ポイントFFTのスペクトル成分の大きさを修正するために使用される。減衰器12におけるフレームは、主として利得を生成するために使用される信号の後に続く１つの時間ユニットである。続いて、逆FFT（IFFT）14が、信号を周波数領域から時間領域に逆写像する。その結果として雑音が低減された信号の512ポイント・フレームは、乗算器16においてウィンドウを乗じられる。続いて、その結果は、加算器18において先行フレームの信号に重ね合わせおよび加算が施され、20ミリ秒、すなわち320個のサンプルの出力信号がライン40に出力される。

この信号処理の流れにおける各ブロックの詳細な説明は、入力から出力へその生起順に以下に示される。

上述したように、フレームにされた入力信号は、ノッチ・フィルタ４のバンクを介して引き出される。

図１および図２を参照して、ノッチ・フィルタ・バンク４は、無限インパルス応答（IIR: Infinite Impulse Response）ディジタル・フィルタのカスケードからなる。各フィルタは、以下の式の応答を有する。

$$H(z) = \frac{1-2\cos\theta z^{-1}+z^{-2}}{1-2r\cos\theta z^{-1}+r^2 z^{-2}} \qquad (1)$$

ここで、θ＝π/8000×（ノッチの周波数）であり、ｒはノッチの幅を反映した１より小さな値である。－3dB幅のノッチがωHzであるならば、ｒ＝１－（ω/2）（π/8000）とな

る。例示する好ましい実施形態において使用される帯域幅ωは、20Hzである。ノッチは、公称周波数に近い背景雑音エネルギーの最大ピークにおいて、約100Hzごとに配置される。

ノッチ・フィルタリングは、新たな信号フレームの320個のサンプルに施される。フィルタリング結果の320個のサンプルは、先行するフレームからのノッチ・フィルタリングされた出力の最後の192個のサンプルに追加されて、全部で512個のサンプルからなる拡張フレームが生成される。

図１および図２を参照して、フィルタ・バンク４から取り出されたノッチ・フィルタリング後の512サンプル・フレームは、次の式を用いてウィンドウを乗じられる。

$$w(i) = f(i)\sqrt{0.5-0.5\cos\left(\pi\frac{i}{191}\right)} \qquad for\ i=0,1,\ldots,191$$

$$w(i) = f(i) \qquad for\ i=192,193,\ldots,319$$

$$w(i) = f(i)\sqrt{0.5-0.5\cos\left(\pi\frac{511-i}{191}\right)} \qquad for\ i=320,321,\ldots,511 \qquad (2)$$

ここで、ｆ（ｉ）は、フィルタ・バンク４からの512サンプル・フレームのノッチ・フィルタリングされた第ｉ番目のサンプルの値である。ｗ（ｉ）は、FFT8に次に与えられる512サンプルのウィンドウ化された出力結果の第ｉ番目のサンプルの結果値である。乗算器６によってもたらされるウィンドウは、拡張フレームの開始部と終了部におけるエッジ効果および不連続性を最小化するためのものである。

時間ウィンドウ化された512個のサンプル・ポイントは、FFT8に与えられる。FFTの偏在性により、多くのディジタル信号処理（DSP）チップの製造者が、FFTを実現するための高度に最適化されたアセンブリ言語コードを供給している。

１フレーム・ディレイ10は、FFTの信号周波数成分を増幅でき、かつ、後に発生する信号値に基づいて減衰器12で処理できるように導入されている。これは、知覚される雑音を導入しない。その理由は、信号成分は、上述したように、それが実際に発生する前のそのスペクトルの20ミリ秒の近傍部分の周波数をマスクするからである。また、会話音のボリュームは、零の振幅から開始してしだいに増大するので、この１フレームのディレイは、会話の発声開始のクリッピングを防止する。

雑音によるFFTの成分は、減衰器12によって減衰するが、一方、信号による成分の減衰量は少ないか、もしくは減衰せず、または増幅されることもある。上述したように、各周波数に対して、実数成分と虚数成分がある。両成分は、雑音抑制スペクトル修正モジュール30から見つけ出された単一の因数（factor）を乗じられ、位相は、周波数成分の大きさが変更されても、その周波数成分に対して維持される。

逆FFT14（IFFT）は、大きさの修正されたFFTに施され、長さが512個のサンプルからなる周波数処理された拡張フレームが生成される。

乗算器16で使用されるウィンドウ化処理は、上述した乗算器６のウィンドウ化処理と全く同じである。その目的は、周波数成分の減衰により発生する不連続性を最小化することである。例えば、全ての周波数成分が１つを除いて零に設定されていると仮定する。IFFTを施した結果は正弦波になる。この正弦波は、大きな値で始まり、大きな値で終わることがある。隣接するフレームには、この正弦波成分が存在しないことがある。したがって、適切なウィンドウ化処理を行うことなく、出力加算器18でこの信号に重ね合わせおよび加算が行われると、クリック音がフレームの開始部および終了部で聞こえることがある。しかしながら、例えば、式２で定めたパラメータを使用して、正弦波を適切にウィンドウ化することによって、大きさがスムーズに増加しスムーズに減少する正弦波が聞こえることになる。

乗算器６および16によるフレームの前置ウィンドウ化処理および後置ウィンドウ化処理の

ために、フレームの重ね合わせおよび加算処理は、フレームの開始部および終了部において出力の大きさが減少することを防止するために必要となる。したがって、512個のサンプルからなる現在の拡張ウィンドウ化フレームの最初の192個のサンプルは、先行の拡張ウィンドウ化フレームの最後の192個のサンプルに加えられる。続いて、現在の拡張フレームの次の128個のサンプル（８ミリ秒）が出力される。続いて、現在の拡張ウィンドウ化フレームの最後の192個のサンプルは、次のフレームとの重ね合わせ／加算処理等に使用するために格納される。

好ましい実施形態において、使用されるウィンドウ関数Wは、変調時間の超過を避けるために、次の特性を有する。

W^2＋（重ね合わせ量だけシフトしたW^2）＝１

例えば、重ね合わせ量がフレームの1/2ならば、ウィンドウ関数Wは次の特性を有する。

W^2＋（1/2だけシフトしたW^2）＝１

［背景雑音評価器20］

図１および図３を参照して、背景雑音評価器20および雑音抑制スペクトル修正モジユール30は、次のように動作する。

背景雑音評価器20は、FFTの各周波数成分の評価を行い、背景雑音によるエネルギーの大きさの平均値を求めるためのものである。この背景雑音評価器により、環境が変わるごとに、ユーザがシステムをマニュアルで調整または操作する必要はなくなる。背景雑音評価器は、信号／雑音環境を絶えず監視し、例えば、エアコンのファンがオン・オフする等に応答して、自動的に背景雑音の評価値を更新する。２つのアプローチが、ある特定の状況で使用される一方のアプローチまたは他方のアプローチの結果とともに使用される。第１のアプローチは、より正確ではあるが、背景雑音のみに対して１秒間隔を必要とする。第２のアプローチは、正確さでは劣るが、どのような状況であっても、10秒で背景雑音評価値を求める。

［定常評価器24］

図１および図３を参照して、第１のアプローチは、定常評価器24を使用して、各フレームのスペクトル形状が他のフレームのスペクトル形状と非常に類似する長いフレーム・シーケンスを探す。たぶん、このような状況は、部屋にいる人間は静かにしており、ファンや回路の雑音による一定の背景雑音が主な信号源である場合にのみ生じ得る。このようなシーケンスが検出されると、各周波数の平均の大きさは、FFTシーケンスの中央部分のフレームから取られる（シーケンスの開始部および終了部のフレームは、低レベルの会話成分を含んでいることがある）。この方法は、第２のアプローチ（後述）よりもはるかに正確に背景雑音のスペクトルを測定するが、背景雑音が比較的一定であることと、部屋にいる人間がある一定期間会話をしていないことを必要とし、このような状況は、実際にはあまり見られない状況である。

この評価器の詳細な処理は、次に示すとおりである。

1. 図３を参照して、第１のアプローチによる方法は、現在の20msフレームのスペクトル形状が先行フレームのスペクトル形状と類似しているかどうかを判断する。まず、この方法は、ステップ240において、先行フレームのスペクトル形状を計算する。

$$N_i(f_c) = 0.25 \sum_{f=f_c-3}^{f=f_c} \left(\sum_{k=k_i}^{k=k_i+31} (R^2(k,f) + I^2(k,f)) \right) \qquad (3)$$

ここで、f_cは、現在の20msフレームのフレーム番号である（この番号は、連続フレームの一つごとに進められる）。ｉは、1000Hzの周波数帯域を示し、k_i＝ｉ×32である。ｋは、512ポイントFFTの256個の周波数成分のインデックスである。Ｒ（k,f）とＩ（k,f）は、それぞれ、フレームｆの第ｋ番目の周波数成分の実数成分と虚数成分である。

2. 次に、現在のフレームのスペクトル形状S_i（f_C）が、ステップ242で決定される。

$$S_i(f_c) = \sum_{k=k_i}^{k=k_i+31} (R^2(k,f_c) + I^2(k,f_c)) \qquad (4)$$

ここで、この式は、上記式（３）と同じ意味を有する。S_iは、現在のフレームf_Cの第ｉ番目の周波数成分の大きさである。

3. 続いて、評価器24は、ステップ244および246において、以下の不等式が成立するかどうかをチェックする。

$$N_i(f_c) > t_l S_i(f_c) \quad (5)$$

または

$$S_i(f_c) > t_l N_i(f_c), \ for \ i = 0,1,\ldots,7 \quad (6)$$

ここで、t_lは低い方のしきい値である。好ましい実施形態においては、t_l＝３である。不等式（５）または（６）が、ｉの４つの値よりも多い値に対して成立するならば、現在のフレームf_cは、信号として分類される。そうでなければ、評価器は、次の不等式が成立するかどうかをチェックする。

$$N_i(f_c) > t_h S_i(f_c) \quad (7)$$

または

$$S_i(f_c) > t_h N_i(f_c), \ for \ i = 0,1,\ldots,7 \quad (8)$$

ここで、t_hは高い方のしきい値である。N_iは、背景雑音評価値の第ｉ番目の周波数成分の大きさを示す。好ましい実施形態においては、t_h＝4.5である。ｉの１つまたは２つ以上の値に対して、いずれかの不等式が成立すると、現在のフレームf_cは信号フレームとして分類される。そうでなければ、現在のフレームは、雑音として分類される。

4. ステップ252において、雑音に分類されたフレームが50個連続して発生すると（１秒間の雑音に相当する）、評価器24は、第10番目から第41番目のフレームの周波数エネルギーを合計することにより、背景雑音の評価値を求める。このシーケンスの開始部分のフレームおよび終了部分のフレームを無視することにより、信号が残りのフレームに存在しないという確度が増加する。評価器は、ステップ254において、次の式の値を求める。

$$B_k = \frac{1}{32} \sum_{f=f_s+31}^{f=f_s} (R^2(k,f) + I^2(k,f)) \quad (9)$$

ここで、ｋ＝0,1,2,...,255であり、f_Sは、第10番目の雑音に分類されたフレームの開始インデックスである。他の変数は、式（３）と同一の表記を有する。値B_kは、第ｋ番目の周波数に対する信号の雑音成分スペクトルの平均の大きさを表す。

図１および図４を参照して、ノッチ・フィルタ・バンクのノッチを配置する場所を決定するために、32個の雑音に分類されたフレームのみに対応するウィンドウ化されていない20msの時間領域サンプルが互いに結合され、連続シーケンスを形成する（ステップ260）。このシーケンスに対して、長いFFTが施される（ステップ262）。約100Hzごとに最大の大きさを有する成分が、突き止められる（ステップ264）。この局所的に最大の大きさが生じる周波数が、ノッチの中心周波数が置かれる位置に対応する（ステップ266）。ノッチは、1500Hz辺りまでのファン雑音のみを減少させるのに有益である。その理由は、ファン雑音スペクトルは、より高い周波数に対して強いピークが存在せず、かなり平坦になる傾向を有するからである。

［稼働最小評価器22］

会話信号が１秒間以上存在しないということがないか、背景雑音自体が一定のスペクトル形状でない場合がある。この場合には、定常評価器24（上述）は、背景雑音評価値を発生しない。このような場合のために、稼働最小評価器22が、正確さの点では劣るものの、背景雑音評価値を生成する。

稼働最小評価器によって使用される処理は、次のとおりである。

1. 10秒間隔に渡って、各周波数成分ｋに対して、その周波数成分についての８つの連続フレームのエネルギーを最小にする８つの連続フレームを見つけ出す。すなわち、すべての周波数成分ｋに対して、以下のM_k（f_k）を最小にするフレームf_kを見つけ出す。

$$M_k(f_k) = \frac{1}{8}\sum_{f=f_k}^{f=f_k+7}\left(R^2(k, f) + I^2(k, f)\right) \qquad (10)$$

ここで、f_kは、上記10秒間隔内で発生する任意のフレーム番号である。一般に、式（10）を最小にするf_kは、異なる周波数成分ｋに対して異なる値をとる。

2．以下の２つの状況が両方とも成立すると、先のステップで背景雑音スペクトル評価値として求められた最小値M_kを使用する。

（ａ）　定常評価器による背景雑音スペクトル評価値の最後の更新から10秒以上が経過している。

（ｂ）　過去の背景雑音評価値（定常評価器または稼働最小評価器から求められたもの）と現在の稼働最小評価器との差Ｄが大きい。差Ｄを定義するために使用される基準は、式（11）で与えられる。

$$D = \sum_{k=0}^{k=255}\left(\max\left(\frac{M_k}{N_k}, \frac{N_k}{M_k}\right) - 1\right)^2 \qquad (11)$$

ここで、max関数は、その２つの引数の大きい方の値を返す。N_kは、（稼働最小評価器または定常最小評価器のいずれかからの）前の背景雑音評価値である。M_kは、稼働最小評価器からの現在の背景雑音評価値である。

Ｄが、あるしきい値（例えば、好ましい実施形態においては3000）よりも大きく、上記条件（ａ）を満足しているならば、M_kは、新しい背景スペクトル評価値として使用される。M_kを雑音評価値として使用することは、ノッチ・フィルタを不作動にすべきことを示す。その理由は、ノッチの中心周波数の優れた評価が可能でないからである。

［雑音抑制スペクトル修正器30］

図１を参照して、背景雑音評価値が求められると、現在のフレームのスペクトルを背景雑音評価値のスペクトルと比較し、この比較に基づいて、出力信号の雑音の知覚を減少させるために、現在のフレームのFFTの各周波数成分に対して減衰処理を行わなければならない。

［グローバル会話信号対雑音検出器32］

任意の所与のフレームは、会話信号とそれ以外のもののいずれかを含んでいる。グローバル会話信号対雑音検出器（Global speech versus noise detector）32は、フレームが雑音であるかどうかについて二値の判定を行う。

会話信号が存在する場合に、マスク効果が不正確な信号対雑音宣言を目立たなくさせる傾向にあるので、しきい値を低くすることができる。一方、フレームが雑音のみであるならば、周波数成分が雑音によるものか信号によるものかを決定することの僅かな誤りが、いわゆる「きらめき」音（twinkling sound）を引き起こすことになる。

例示の実施形態によると、フレーム内に会話信号が存在するかどうかを判断するために、システムは、現在のフレームの第ｋ番目の周波数成分の大きさS_kと背景雑音評価値の第ｋ番目の周波数成分の大きさC_kとを比較する。次に、（１フレームに対して）ｋの７つの値より多い値について$S_k > T \times C_k$ならば、そのフレームは会話信号のフレームと宣言される。そうでなければ、そのフレームは雑音フレームと宣言される。ここで、Ｔは一定のしきい値である（好ましい実施形態ではＴ＝３）。

［個別周波数成分用ローカル会話信号対雑音検出器34］

上記節で説明したグローバル会話信号対雑音検出器32は、各周波数成分が雑音かどうかについて二値判定を行うものである。一方、ローカル会話信号対雑音検出器（Local speech versus noise detector）34は、各周波数成分が雑音であるかどうかについて幅のある判定を行う。これらの判定は、第ｋ番目の周波数成分が雑音であるという高い確度の判定から、第ｋ番目の周波数成分が信号であるという高い確度の判定までの幅を有する。

この判定は、現在のフレームの第ｋ番目周波数成分の大きさの、背景雑音スペクトル評価値のうちのこれに対応する成分の大きさに対する比に基づいてなされる。判定値をD_kで示す。この実施形態において、判定値D_kは０から４まで変動し、D_k＝０の判定は、「その成

分が雑音である確度が高い」ことに対応し、D_k＝４の判定は、「その成分が信号である確度が高い」ことに対応する。

D_kの値は以下のように定められる。

$$\begin{aligned} &if\ \frac{S_k}{N_k} > t_4,\ D_k = 4,\\ &else\ if\ \frac{S_k}{N_k} > t_3,\ D_k = 3,\\ &else\ if\ \frac{S_k}{N_k} > t_2,\ D_k = 2,\\ &else\ if\ \frac{S_k}{N_k} > t_1,\ D_k = 1,\\ &else\qquad\qquad D_k = 0 \end{aligned} \tag{12}$$

ここで、現在のフレームに対して、$S_k = R^2(k) + I^2(k)$である。N_kは成分ｋの背景雑音評価値である。t_1、t_2、t_3、t_4に使用される値は、グローバル会話信号検出器32が先行フレームを会話信号と宣言したか雑音と宣言したかに依存して変化する。例示の実施形態において、雑音と宣言した場合には、t_1＝6.3、t_2＝9.46、t_3＝18.9、t_4＝25.2であり、信号と宣言した場合には、これらのしきい値は1/2に下げられ、t_1＝3.15、t_2＝4.73、t_3＝9.45、t_4＝12.6にされる。

これらのD_kは、制御される減衰器の利得乗法因数（gain multiplicative factor）を調整する以下の処理に使用される。

［臨界帯域付近の周波数ビン（frequency bin）利得の時間スペクトル分散器36］

配列A_kは、すべてのFFT周波数成分ｋに対する乗法因数（multiplicative factor）を格納する。A_kの要素は、FFT8から１フレーム・ティレイ10を介して与えられるスペクトル成分を修正するために、制御減衰器12によって使用される。A_kの値は、ローカル会話信号対雑音検出器34で求められた判定値D_kに基づいて変更される。

A_kの値は、Ｌ＜A_k＜１の範囲に制限される。ここで、Ｌは、雑音低減量の下限である（以下に説明する）。Ｌの値が小さいほど、より多くの雑音を低減することができる。しかしながら、一般に、雑音を多く低減するほど、人為的なもの（artifact）が同時に発生してくる。信号の信号対雑音比（SN比）を高くすると、反対すべき人為的なものを会話信号内に生成することなく、Ｌの値を小さく設定することができる。適度な14dBのSN比に対するＬの良好な値は、0.25である。会話の明瞭性に影響する反対すべき人為物を減らすために、SN比が低くなるにつれて、Ｌの値を大きくすべきである。例えば、6dBのSN比では、Ｌの値を0.5にする必要がある。会話のSN比が、システムの処理中に測定され、測定値は、Ｌの値を決定するために使用される。

A_kは、上記式（12）から導出されるように、先行フレームにおけるA_kと現在のフレームのD_kの値の関数として、各新フレームごとに変化する。第ｉ番目のフレームからのA_kをA_k^iとすると、$A_k^i = G(A_k^{i-1}, D_k)$となる。ここで、関数Ｇは、以下の式（13）で定義される。

$$\begin{aligned} G(A_k^i, D_k) = &if\ D_k = 0\ then\\ &A_k^i = A_k^{i-1} \times \beta_0\\ &if\ D_k \geq 1\ then\\ &A_k^i = A_k^{i-1} \times \beta \end{aligned} \tag{13}$$

ここで、β＞１であり、D_kとともに増加する。β_0＜１である。

つまり、判定値D_k≧１ならば、先行フレームからのA_kは、D_kの値の増加とともに増加する１より大きな乗法因数を乗じられる。判定値D_k＝０ならば、先行フレームからのA_kは、１より小さな値（代表的には0.8）の乗法因数を乗じられる。

これは、時間的分散（拡張、拡散:temporal spreading）である。好ましい実施形態では、時間的分散は、現在のフレーム前20msから現在のフレーム後200msまで存在する。

判定値D_k＝４は、スペクトル成分ｋが会話信号を含む確度が高いことを意味し、A_kはその最大値１に設定される。
次に、A_kのスペクトル分散（拡張、拡散:spectral spreading）が500Hzよりも大きな周波数に対して実行される。この分散は、この例示の実施形態においては中心周波数の1/6に等しい臨界帯域幅上で起こる。この理由は、心理音響学の実験からきている。この実験は、所与の周波数において強い信号成分があると、この信号成分は、その周波数の1/6の帯域幅における雑音に対してマスク効果を有する、ということを示している。スペクトル分散は、次のようにして達成される。
$D_k \geqq 1$の判定値に対して、上述したように計算される臨界帯域幅のA_kは、Ｆ（D_k）より小さくなることは許されない（Ｆは、以下で定義する）。本質的に、雑音（および会話信号）の低減量がスペクトル領域で少なくなるように、臨界帯域幅におけるA_kに対する下限Ｌは大きくされる。典型的には、Ｌは0.25に等しく、Ｆ（D_k）は、次のように定められる。
Ｆ（４）＝0.5
Ｆ（３）＝0.4
Ｆ（２）＝0.333
500Hzより小さな周波数に対して、スペクトル分散は実行されない。実験では、背景雑音スペクトルは、低い周波数において、回転音響効果により、多くのピークと谷間を有し、耳障りな粗い人為的なものは、これらの周波数においてスペクトル分散を行うことが原因で生じることが示されている。
［スペクトル・バレー充填器38］
雑音低減方法の人為的なものの一つとして、残響感の増加がある。これは、残響による信号のスペクトル消失点（spectral nulls）の深さが、利得乗法因数の配列A_kを導出するプロセスにおいて増加するという事実から起因する。この人為的なものとの戦いを補助するために、残響に伴うスペクトル消失点（null）に対応するA_kの局所的最小値が増加させられる。500Hzから上の周波数の範囲に対して、$A_k < A_{k-4}$かつ$A_k < A_{k+4}$であるならば、ｋ＝16,17,18,…,251に対して、次のようになる。

$$A_k = \frac{A_{k-4} + A_{k+4}}{2} \qquad (14)$$

［減衰器12］
乗法因数配列の値がある特定のフレームについて決定されると、雑音対会話信号の決定に反映するように周波数成分を調整することができる。
制御される減衰器12において、乗法因数配列A_kは次のように使用される。遅延変換された信号の実数および虚数の各周波数成分は、次のように増減される。

$$\text{for } k = 0, 1, 2, \ldots, 256$$

$$R_n(k) = A_k R(k)$$

$$I_n(k) = A_k I(k) \qquad (15)$$

続いて、図１のブロック図に示すように、修正されたフーリエ成分R_n（ｋ）およびI_n（ｋ）には、逆FFT、ウィンドウ化および重ね合わせが行われ、最終的に、雑音が低減された出力信号フレームが生成される。
これにより、審美的に満足され、雑音が低減されて知覚されない信号が生成される。
本発明の好ましい特定の実施形態に対して付加し、減らし、削除する修正、および他の修正を行うことは、この技術分野の専門家（当業者）には明らかであり、以下の請求の範囲に含まれるものである。

【図４】

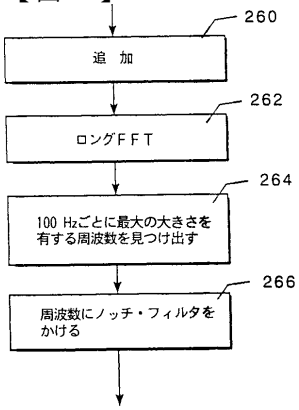

FIG. 4

【図１】

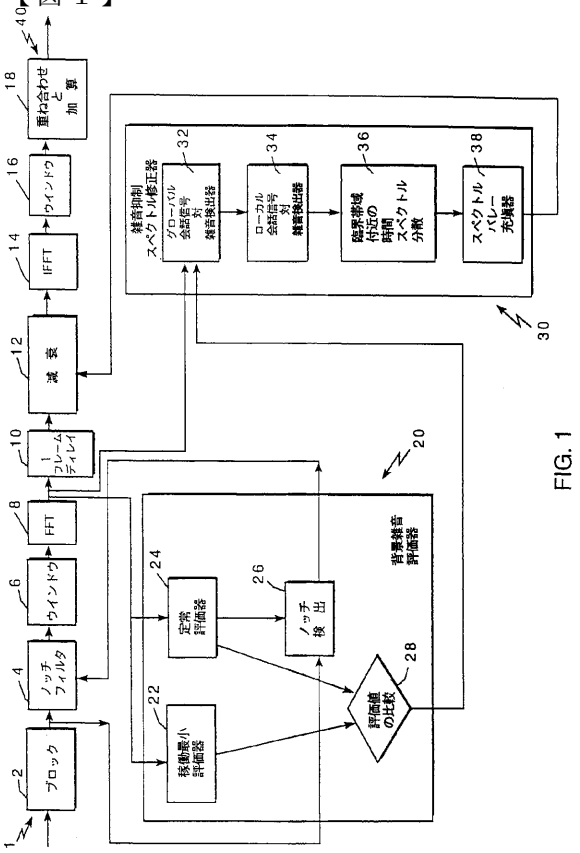

FIG. 1

【図２】

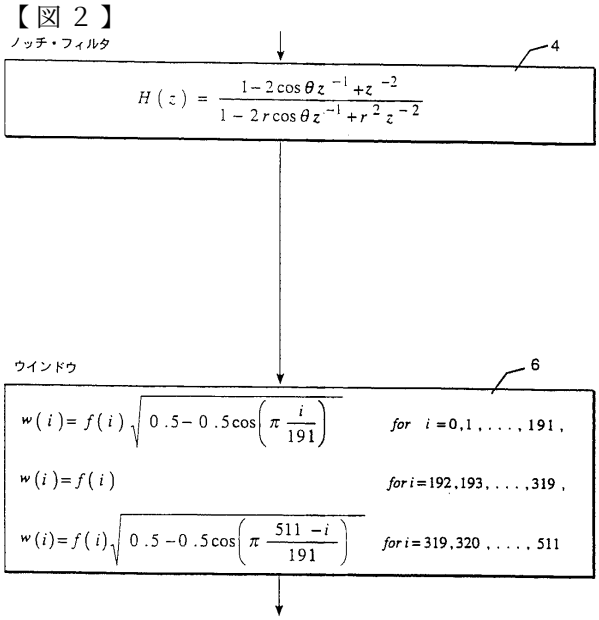

FIG. 2

【図３】

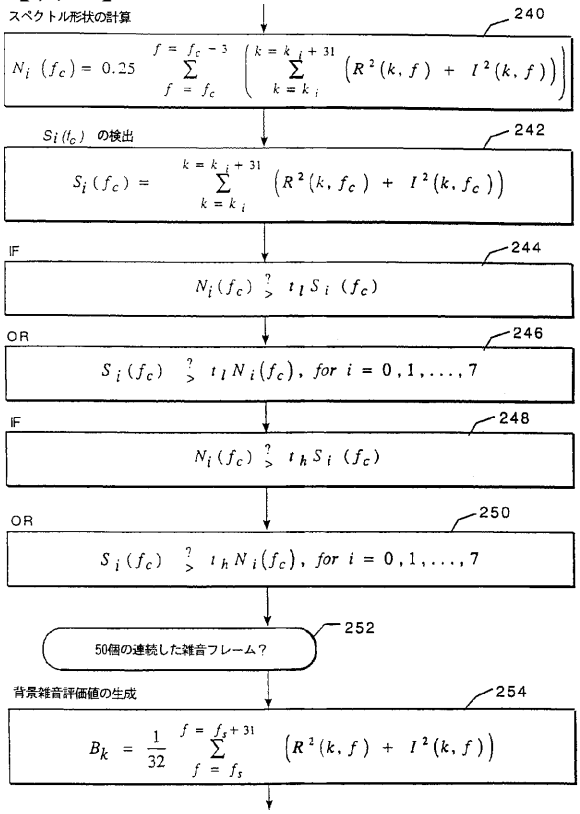

FIG. 3

フロントページの続き

審査官　渡邊　聡

(56)参考文献　特公昭６１－００２９６０（ＪＰ，Ｂ１）
特開平０４－３４０５９９（ＪＰ，Ａ）
斎藤収三、中田和男，音声情報処理の基礎，オーム社，１９８１年１１月３０日，１３、１４、５６

(58)調査した分野(Int.Cl.[7]，ＤＢ名)
G10L 21/02