

(19) 日本国特許庁(JP)

(12) 公開特許公報(A)

(11) 特許出願公開番号

特開2012-181561  
(P2012-181561A)

(43) 公開日 平成24年9月20日(2012.9.20)

(51) Int.Cl.	F I	テーマコード (参考)
<b>G 1 0 L 21/04 (2006.01)</b>	G 1 0 L 21/04 1 3 0 A	
<b>G 1 0 L 21/02 (2006.01)</b>	G 1 0 L 21/04 2 0 0 C	
	G 1 0 L 21/02 4 0 0	

審査請求 有 請求項の数 12 O L (全 31 頁)

(21) 出願番号	特願2012-144135 (P2012-144135)	(71) 出願人	000003078 株式会社東芝
(22) 出願日	平成24年6月27日 (2012. 6. 27)		東京都港区芝浦一丁目1番1号
(62) 分割の表示	特願2009-72886 (P2009-72886) の分割	(74) 代理人	110001092 特許業務法人サクラ国際特許事務所
原出願日	平成21年3月24日 (2009. 3. 24)	(72) 発明者	須藤 隆 東京都港区芝浦一丁目1番1号 株式会社東芝内
		(72) 発明者	長田 将高 東京都港区芝浦一丁目1番1号 株式会社東芝内

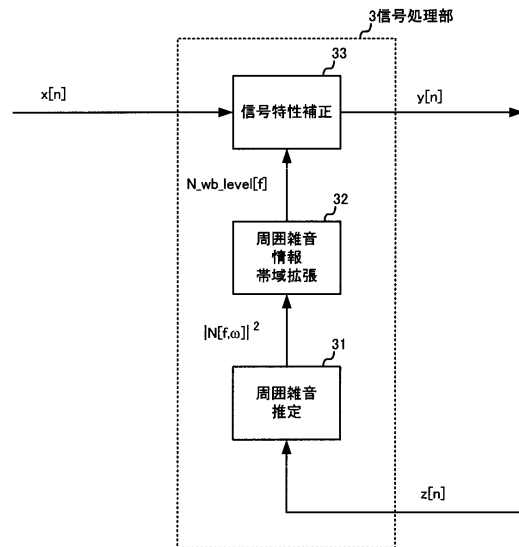
(54) 【発明の名称】 信号処理装置

(57) 【要約】

【課題】再生する目的信号と周囲雑音とで、制限される周波数帯域が異なっているために信号成分が存在する周波数帯域が異なっていたり、サンプリング周波数が異なっていたりする場合でも、明瞭度を向上させることが可能な信号処理装置を提供する。

【解決手段】第1の周波数範囲に帯域制限された入力信号に対して周波数特性を変化させる信号処理装置であって、集音信号に含まれる周囲雑音を抽出する周囲雑音抽出手段と、前記周囲雑音抽出手段によって抽出された周囲雑音から第2の周波数範囲の周波数特性情報を抽出する情報抽出手段と、前記情報抽出手段によって抽出された周波数特性情報に対して、前記第1の周波数範囲へ周波数特性情報を周波数方向に拡張する周波数特性情報拡張手段と、前記周波数特性情報拡張手段によって得られた周波数特性情報に応じて、前記入力信号の周波数特性を変化させる信号補正手段と、を具備して構成するようにした。

【選択図】 図2



**【特許請求の範囲】****【請求項 1】**

第 1 の周波数範囲に帯域制限された入力信号に対して周波数特性を変化させる信号処理装置であって、

集音信号に含まれる周囲雑音を抽出する周囲雑音抽出手段と、

前記周囲雑音抽出手段によって抽出された周囲雑音から第 2 の周波数範囲の周波数特性情報を抽出する情報抽出手段と、

前記情報抽出手段によって抽出された周波数特性情報に対して、前記第 1 の周波数範囲へ周波数特性情報を周波数方向に拡張する周波数特性情報拡張手段と、

前記周波数特性情報拡張手段によって得られた周波数特性情報に応じて、前記入力信号の周波数特性を変化させる信号補正手段と、

を備えた信号処理装置。

10

**【請求項 2】**

前記情報抽出手段は、前記第 2 の周波数範囲よりも狭い第 3 の周波数範囲に制限された周波数特性情報を抽出することを特徴とする請求項 1 に記載の信号処理装置。

**【請求項 3】**

事前に取得した信号の第 2 の周波数範囲の周波数特性情報と第 1 の周波数範囲の周波数特性情報とを対応づけて記憶する記憶手段を更に有し、

前記周波数特性情報拡張手段は、前記記憶手段に記憶された第 2 の周波数範囲の周波数特性情報と第 1 の周波数範囲の周波数特性情報との対応を用いて周波数特性情報の拡張を行うことを特徴とする請求項 1 または請求項 2 に記載の信号処理装置。

20

**【請求項 4】**

前記周波数特性情報拡張手段は、

前記情報抽出手段によって抽出された周波数特性情報から、前記第 1 の周波数範囲のうち前記第 2 周波数範囲を除いた第 4 の周波数範囲における周波数特性情報を推定し、この推定した周波数特性情報と前記情報抽出手段により抽出された前記周波数特性情報とが、前記第 2 の周波数範囲と前記第 4 の周波数範囲との境界において連続するように補正を行うことを特徴とする請求項 1 乃至請求項 3 のいずれか 1 項に記載の信号処理装置。

**【請求項 5】**

前記情報抽出手段によって抽出される周波数特性情報は、周波数ごとのマスキングレベルであることを特徴とする請求項 1 乃至請求項 4 のいずれか 1 項に記載の信号処理装置。

30

**【請求項 6】**

前記情報抽出手段によって抽出される周波数ごとのマスキングレベルは、多項式で近似表現されることを特徴とする請求項 5 に記載の信号処理装置。

**【請求項 7】**

第 1 の周波数範囲に帯域制限された入力信号に対して周波数特性を変化させる信号処理装置であって、

集音信号に含まれる周囲雑音を抽出する周囲雑音抽出手段と、

前記周囲雑音抽出手段によって抽出された周囲雑音から前記第 1 の周波数範囲よりも狭い第 2 の周波数範囲の周波数特性情報を抽出する情報抽出手段と、

40

前記情報抽出手段によって抽出された周波数特性情報に対して、前記第 1 の周波数範囲へ周波数特性情報を周波数方向に拡張する周波数特性情報拡張手段と、

前記周波数特性情報拡張手段によって得られた周波数特性情報に応じて、前記入力信号の周波数特性を変化させる信号補正手段と

を備えた信号処理装置。

**【請求項 8】**

前記情報抽出手段は、前記第 2 の周波数範囲よりも狭い第 3 の周波数範囲に制限された周波数特性情報を抽出することを特徴とする請求項 7 に記載の信号処理装置。

**【請求項 9】**

事前に取得した信号の第 2 の周波数範囲の周波数特性情報と第 1 の周波数範囲の周波数

50

特性情報とを対応づけて記憶する記憶手段を更に有し、

前記周波数特性情報拡張手段は、前記記憶手段に記憶された第2の周波数範囲の周波数特性情報と第1の周波数範囲の周波数特性情報との対応を用いて周波数特性情報の拡張を行うことを特徴とする請求項7または請求項8に記載の信号処理装置。

【請求項10】

前記周波数特性情報拡張手段は、

前記情報抽出手段によって抽出された周波数特性情報から、前記第1の周波数範囲のうち前記第2周波数範囲を除いた第4の周波数範囲における周波数特性情報を推定し、この推定した周波数特性情報と前記情報抽出手段により抽出された前記周波数特性情報とが、前記第2の周波数範囲と前記第4の周波数範囲との境界において連続するように補正を行うことを特徴とする請求項7乃至請求項9のいずれか1項に記載の信号処理装置。

10

【請求項11】

前記情報抽出手段によって抽出される周波数特性情報は、周波数ごとのマスキングレベルであることを特徴とする請求項7乃至請求項10のいずれか1項に記載の信号処理装置

。

【請求項12】

前記情報抽出手段によって抽出される周波数ごとのマスキングレベルは、多項式で近似表現されることを特徴とする請求項11に記載の信号処理装置。

【発明の詳細な説明】

【技術分野】

20

【0001】

この発明は、音声や音楽・オーディオなどの信号に対して明瞭度を向上させる信号処理装置に関する。

【背景技術】

【0002】

音声や音楽・オーディオなどの信号を再生するときに、音声や音楽・オーディオなどの所望の信号（以降、目的信号と称する）以外の周囲雑音などの影響を受けて目的信号の明瞭度が低下する場合がある。そこで、目的信号の明瞭度を向上させるために、集音した信号に含まれる周囲雑音に応じた信号処理を施す必要がある。従来、このような信号処理方法としては、周囲雑音の音量を用いる手法、周囲雑音の周波数特性を用いる手法（例えば、特許文献1）があった。

30

【先行技術文献】

【特許文献】

【0003】

【特許文献1】特開2001-188599号公報

【発明の概要】

【発明が解決しようとする課題】

【0004】

しかしながら、目的信号と周囲雑音とで、制限される周波数帯域が異なっているために信号成分が存在する周波数帯域が異なっていたり、サンプリング周波数が異なっていたりする場合がある。このような場合、従来信号処理装置では、周囲雑音の音量や周波数特性が高精度に求められないために音質劣化を招き、明瞭度を向上させることができないという課題があった。

40

【0005】

また、音声信号や音楽・オーディオ信号などの目的信号に対して、エイリアシングを用いたり非線形関数を用いたり線形予測分析を用いたりするような帯域を拡張する従来技術をそのまま用いて、集音した周囲雑音の帯域を拡張しても、周囲雑音の周波数特性を高精度に推定することはできないという課題があった。

【0006】

この発明は上記の問題を解決すべくなされたもので、再生する目的信号と周囲雑音とで

50

、制限される周波数帯域が異なっているために信号成分が存在する周波数帯域が異なっていたり、サンプリング周波数が異なっていたりする場合でも、明瞭度を向上させることが可能な信号処理装置を提供することを目的とする。

【課題を解決するための手段】

【0007】

上記の目的を達成するために、この発明は、第1の周波数範囲に帯域制限された入力信号に対して周波数特性を変化させる信号処理装置であって、集音信号に含まれる周囲雑音を抽出する周囲雑音抽出手段と、前記周囲雑音抽出手段によって抽出された周囲雑音から第2の周波数範囲の周波数特性情報を抽出する情報抽出手段と、前記情報抽出手段によって抽出された周波数特性情報に対して、前記第1の周波数範囲へ周波数特性情報を周波数方向に拡張する周波数特性情報拡張手段と、前記周波数特性情報拡張手段によって得られた周波数特性情報に応じて、前記入力信号の周波数特性を変化させる信号補正手段と、を具備して構成するようにした。

10

【発明の効果】

【0008】

本発明によれば、再生する目的信号と周囲雑音で、制限される周波数帯域が異なっているために信号成分が存在する周波数帯域が異なっていたり、サンプリング周波数が異なっていたりする場合でも、明瞭度を向上させることが可能な信号処理装置を提供することができる。

【図面の簡単な説明】

20

【0009】

【図1】この発明に係わる信号処理装置の第1の実施例を適用した通信装置の構成を示す回路ブロック図。

【図2】この発明に係わる信号処理部の第1の実施例の構成を示す回路ブロック図。

【図3】図2に示した信号処理部の周囲雑音推定部の構成例を示す回路ブロック図。

【図4】図2に示した信号処理部の周囲雑音情報帯域拡張部の構成例を示す回路ブロック図。

【図5】図4に示した周囲雑音情報帯域拡張部の辞書格納部における辞書の生成方法の動作を説明するための処理フロー図。

【図6】図2に示した信号処理部の信号特性補正部の構成例を示す回路ブロック図。

30

【図7】この発明に係わる信号処理装置の第1の実施例を適用した通信装置およびデジタルオーディオプレイヤーの構成を示す回路ブロック図。

【図8】この発明に係わる信号処理部の変形例1の構成を示す回路ブロック図。

【図9】図8に示した信号処理部の周囲雑音情報帯域拡張部の構成例を示す回路ブロック図。

【図10】図9に示した周囲雑音情報帯域拡張部の辞書格納部における辞書の生成方法の動作を説明するための処理フロー図。

【図11】広帯域マスキング閾値の例を示す図。

【図12】図8に示した信号処理部の信号特性補正部の構成例を示す回路ブロック図。

【図13】この発明に係わる信号処理部の変形例3の構成を示す回路ブロック図。

40

【図14】図13に示した信号処理装置の周囲雑音情報帯域拡張部の構成例を示す回路ブロック図。

【図15】図14に示した周囲雑音情報帯域拡張部の辞書格納部における辞書の生成方法の動作を説明するための処理フロー図。

【図16】図14に示した周囲雑音情報帯域拡張部の閾値補正部の動作を説明するための例を示す図。

【図17】図14に示した周囲雑音情報帯域拡張部の辞書格納部における辞書の他の生成方法の動作を説明するための処理フロー図。

【図18】図14に示した周囲雑音情報帯域拡張部の辞書格納部における辞書の他の生成方法の動作を説明するための処理フロー図。

50

【図19】この発明に係わる信号処理部の第2の実施例を適用した通信装置およびデジタルオーディオプレイヤーの構成を示す回路ブロック図。

【図20】この発明に係わる信号処理装置の第2の実施例の構成を示す回路ブロック図。

【図21】図20に示した信号処理装置の周囲雑音推定部と周囲雑音抑圧処理部の構成例を示す回路ブロック図。

【発明を実施するための形態】

【0010】

以下、図面を参照して、この発明の実施形態について説明する。

(第1の実施例)

図1は、この発明の一実施形態である通信装置の構成を示すものである。この図に示す通信装置は、例えば携帯電話などの無線通信装置の受信系を示すものであって、無線通信部1と、デコーダ2と、信号処理部3と、デジタル・アナログ(D/A)変換器4と、スピーカ5と、マイク6と、アナログ・デジタル(A/D)変換器7と、ダウンサンプリング部8と、エコー抑圧処理部9と、エンコーダ10とを備えている。本実施形態では、再生する目的信号は、受信した入力信号に含まれる遠端話者の音声信号であるとして説明する。

10

【0011】

無線通信部1は、移動通信網に収容される無線基地局と無線通信し、そしてこの無線基地局および移動通信網を通じて通信相手局との間に通信リンクを確立して通信する。

【0012】

デコーダ2は、無線通信部1が通信相手局から受信した受信データを、事前に決められた時間単位である1フレーム(=20[ms])ごとに復号して、デジタルの入力信号 $x[n]$  ( $n=0,1,\dots,2N-1$ )を得て、フレーム単位で信号処理部3に出力する。ただし、この入力信号 $x[n]$ は、サンプリング周波数は $f_s'$  [Hz]で $f_{s\_wb\_low}$ [Hz]から $f_{s\_wb\_high}$ [Hz]までに帯域制限された広帯域の信号である。ここでは、後述する集音信号 $z[n]$ のサンプリング周波数 $f_s$ [Hz]との関係を、 $f_s'=2f_s$ とする。また、サンプリング周波数 $f_s'$  [Hz]のときの1フレームのデータ長は $2N$ サンプルする。つまり、 $N=20[\text{ms}] \times f_s[\text{Hz}] \div 1000$ とする。

20

【0013】

信号処理部3は、後述するエコー抑圧処理部8においてエコー低減された集音信号 $z[n]$  ( $n=0,1,\dots,N-1$ )に応じて、1フレーム単位で入力信号 $x[n]$  ( $n=0,1,\dots,2N-1$ )に対して信号補正処理を施し、音量または周波数特性を変化させて、その出力信号を $y[n]$  ( $n=0,1,\dots,2N-1$ )としてD/A変換器4とダウンサンプリング部8に出力する。なお、信号処理部3の具体的な構成例については後に詳述する。

30

【0014】

D/A変換器4は、上記信号補正された出力信号 $y[n]$ をアナログ信号 $y(t)$ に変換して、スピーカ5に出力する。スピーカ5は、アナログ信号である出力信号 $y(t)$ を音響空間へ出力する。

【0015】

マイク6は、音を集音してアナログ信号である集音信号 $z(t)$ を取得し、A/D変換器7に出力する。このアナログ信号には、近端話者の音声信号と、それ以外の周囲環境に起因するノイズ成分、出力信号 $y(t)$ と音響空間に起因するエコー成分などが混在する。例えばノイズ成分としては、電車などが出す騒音、車などのカーノイズ、人ごみでのストリートノイズなどが挙げられる。本実施形態では、通信装置として近端話者の音声信号は通信相手局との間での通信にて所望される必要な信号であるため、近端話者の音声信号以外の成分を周囲雑音として扱う。

40

【0016】

A/D変換器7は、アナログ信号である集音信号 $z(t)$ をデジタル信号に変換して、デジタルの集音信号 $z'[n]$  ( $n=0,1,\dots,N-1$ )を得て、 $N$ サンプル単位でエコー抑圧処理部8に出力する。ただし、この集音信号 $z[n]$ は、サンプリング周波数は $f_s$ [Hz]で $f_{s\_nb\_low}$ [Hz]から $f_{s\_nb\_high}$ [Hz]までに帯域制限された狭帯域の信号である。ただし、 $f_{s\_wb\_low}$

50

$s_{nb\_low} < fs_{nb\_high} < fs/2$      $fs_{wb\_high} < fs' / 2$  を満たすものとする。

【 0 0 1 7 】

ダウンサンプリング部 8 は、出力信号  $y[n]$  をサンプリング周波数  $fs'$  [Hz] から  $fs$  [Hz] にダウンサンプリングして、 $fs_{nb\_low}$  [Hz] から  $fs_{nb\_high}$  [Hz] までに帯域制限した信号を  $y'[n]$  ( $n=0,1,\dots,N-1$ ) としてエコー抑圧処理部 9 に出力する。

【 0 0 1 8 】

エコー抑圧処理部 9 は、ダウンサンプリングされた出力信号  $y'[n]$  ( $n=0,1,\dots,N-1$ ) を利用して、集音信号  $z'[n]$  ( $n=0,1,\dots,N-1$ ) に含まれるエコー成分を低減する処理を行い、そのエコー低減された信号を  $z[n]$  ( $n=0,1,\dots,N-1$ ) として、信号処理部 3 とエンコーダ 10 に出力する。ここでエコー抑圧処理部 9 は、例えば、特登 4 0 4 7 8 6 7 号公報や、特開 2 0 0 6 - 2 0 3 3 5 8 号公報や、特開 2 0 0 7 - 6 0 6 4 4 号公報などに記載される既存の技術で実施してよい。

10

【 0 0 1 9 】

エンコーダ 10 は、エコー抑圧処理部 8 においてエコー低減された集音信号  $z[n]$  ( $n=0,1,\dots,N-1$ ) を  $N$  サンプルごとに符号化して無線通信部 1 に出力し、無線通信部 1 によって送信データとして通信相手局へ送信される。

【 0 0 2 0 】

次に、信号処理部 3 の実施例について説明する。以下の説明では、例えば、 $fs=8000$  [Hz]、 $fs'=16000$  [Hz]、 $fs_{nb\_low}=340$  [Hz]、 $fs_{nb\_high}=3950$  [Hz]、 $fs_{wb\_low}=50$  [Hz]、 $fs_{wb\_high}=7950$  [Hz] とする。帯域制限の周波数帯域やサンプリング周波数については、これに限らない。また、ここでは  $N=160$  とする。

20

【 0 0 2 1 】

図 2 は、信号処理部 3 の構成例を示すものである。信号処理部 3 は、周囲雑音推定部 3 1 と、周囲雑音情報帯域拡張部 3 2 と、信号特性補正部 3 3 とを備える。これらは、1 つのプロセッサと、図示しない記憶媒体に記録されたソフトウェアによって実現することも可能である。

【 0 0 2 2 】

周囲雑音推定部 3 1 は、エコー抑圧処理部 8 においてエコー低減された信号から近端話者の音声信号以外の信号を周囲雑音と推定し、この周囲雑音を特徴付ける特徴量を抽出する。なお、集音信号  $z[n]$  が狭帯域の信号であるため、周囲雑音も狭帯域の信号である。そこで、周囲雑音を特徴付ける特徴量を、狭帯域信号情報と称する。狭帯域信号情報は、パワースペクトル、振幅スペクトルや位相スペクトル、PARCOR 係数や反射係数、線スペクトル周波数、ケプストラム係数、メルケプストラム係数など、周囲雑音を特徴付ける特徴量であればどのようなものでも構わない。

30

【 0 0 2 3 】

周囲雑音情報帯域拡張部 3 2 は、狭帯域信号情報を用いて、周囲雑音を入力信号  $x[n]$  の周波数帯域と同じ周波数帯域（広帯域）に拡張した場合にこの周囲雑音を特徴付ける特徴量を推定する。この特徴量を、広帯域信号情報と称する。

【 0 0 2 4 】

信号特性補正部 3 3 は、周囲雑音情報帯域拡張部 3 2 を用いて、目的信号の信号特性を補正する。

40

【 0 0 2 5 】

このように、周囲雑音が狭帯域の信号であっても、広帯域に拡張した場合の特徴量を推定することによって、信号特性補正部 3 3 での補正処理によって明瞭度を向上させることができる。

【 0 0 2 6 】

以下の説明では、信号処理部 3 の具体的な構成について説明する。なお、以下の説明では、狭帯域信号情報は周囲雑音のパワースペクトル、広帯域信号情報は周囲雑音を広帯域の信号に拡張した場合のパワー値（広帯域パワー値）であるとして説明する。

【 0 0 2 7 】

50

図3に周囲雑音推定部31の構成例を示す。周囲雑音推定部31は、周波数領域変換部311と、パワー算出部312と、周囲雑音区間判定部313と、周波数スペクトル更新部314とを備える。

【0028】

周囲雑音推定部31は、エコー抑圧処理部8においてエコー低減された集音信号 $z[n]$  ( $n=0,1,\dots,N-1$ )から近端話者の音声信号以外である周囲雑音を推定してこの信号のパワースペクトル $|N[f,w]|^2$ を抽出して、周囲雑音情報帯域拡張部32へ出力する。

【0029】

周波数領域変換部311は、現在のフレーム $f$ の集音信号 $z[n]$  ( $n=0,1,\dots,N-1$ )が入力される。そして、このフレームの1フレーム前の集音信号から窓掛けによるオーバーラップサンプル数分のサンプルを抽出し、現在のフレームの入力信号と時間方向に結合し、適宜零詰めなどを行って、周波数領域変換に必要なサンプル分の信号を取り出す。次のフレームでの集音信号 $z[n]$ のシフト幅と集音信号 $z[n]$ のデータ長の比であるオーバーラップは50%である場合が考えられるが、ここでは例として、1フレーム前とのオーバーラップのサンプル数を $L=48$ として、1フレーム前の集音信号 $L$ サンプルと当該フレームの集音信号 $z[n]$ の $N=160$ サンプル分と $L$ サンプル分の零詰めから、 $2M=256$ サンプルを用意とする。この $2M$ サンプルに対して正弦波窓による窓関数を乗じることで窓掛けを行う。そして、窓掛けを行った $2M$ サンプルの信号に対して、周波数領域変換を行う。周波数領域への変換は、例えばFFTの次数を $2M$ としFFTによって行うことができる。なお、周波数領域変換を施す信号に零詰めすることによってデータ長を2のべき乗( $2M$ )にし、周波数領域変換の次数を2のべき乗( $2M$ )にするとしたが、周波数領域変換の次数はこれに限らない。

【0030】

集音信号 $z[n]$ が実信号である場合は、周波数領域変換を施して得られた信号から冗長な $M=128$ ピンを除くと、周波数スペクトル $Z[f,w]$  ( $w=0,1,\dots,M-1$ )が得られる、これを出力する。ただし、 $w$ は、周波数ピンを表す。なお、実信号のとき冗長なのは本来 $M-1$  ( $=127$ )ピンであり、最高域の周波数ピン $w=M$  ( $=128$ )を考慮するべきである。しかしながら、ここで周波数領域変換する信号は、帯域制限された音声信号を含むデジタル信号を前提としており、帯域制限によって最高域の周波数ピン $w=M$ を考慮しなくても音質に影響を及ぼさない。そこで、これ以降説明の簡略化のために、最高域の周波数ピン $w=M$ を考慮しない記述にする。勿論、最高域の周波数ピン $w=M$ を考慮しても構わない。その際、最高域の周波数ピン $w=M$ は、 $w=M-1$ と同等に扱うか、単独で扱うようにする。

【0031】

なお、窓掛けに用いる窓関数は、ハミング窓に限定せず、他の対称窓(ハニング窓、ブラックマン窓、正弦波窓など)あるいは音声符号化処理で用いられるような非対称窓などに適宜変更してよい。また、周波数領域変換は、DFT(Discrete Fourier Transform)や離散コサイン変換(DCT: Discrete Cosine Transform)などの周波数領域に変換する他の直交変換を代用することも可能である。

【0032】

パワー算出部312は、周波数領域変換部311から出力された周波数スペクトル $Z[f,w]$  ( $w=0,1,\dots,M-1$ )における実部と虚部の2乗和であるパワースペクトル $|Z[f,w]|^2$  ( $w=0,1,\dots,M-1$ )を算出して出力する。

【0033】

周囲雑音区間判定部313は、集音信号 $z[n]$  ( $n=0,1,\dots,N-1$ )と、パワー算出部312から出力されるパワースペクトル $|Z[f,w]|^2$  ( $w=0,1,\dots,M-1$ )と、周波数スペクトル更新部314から出力される1フレーム前の各周波数帯域の周囲雑音のパワースペクトル $|N[f-1,w]|^2$ を用いて、集音信号 $z[n]$ に周囲雑音が支配的に含まれている区間(周囲雑音区間)であるか、周囲雑音には含まない近端話者の音声信号と周囲雑音が混在している区間(音声区間)のどちらであるかの判別をフレーム毎に行い、フレーム毎に判定結果を表すフレーム判定情報 $vad[f]$ を出力する。ここでは、周囲雑音区間であるときフレーム判定情報 $vad[$

10

20

30

40

50

$f]=0$ とし、音声区間であるとき $\text{vad}[f]=1$ とする。なおこれ以降、当該成分のみしか存在しないか、あるいは当該成分が他の成分よりも非常に多く含まれる場合（所定の閾値以上含まれる場合）を「支配的に含まれる」と表現する。

【0034】

具体的には、集音信号 $z[n]$  ( $n=0,1,\dots,N-1$ )とパワースペクトル $|Z[f,w]|^2$ と1フレーム前の周囲雑音のパワースペクトル $|N[f-1,w]|^2$ を用いて、複数の特徴量を算出し、フレーム判定情報 $\text{vad}[f]$ を出力する。ここでは複数の特徴量として、1次自己相関係数 $\text{Acorr}[f,1]$ 、自己相関係数最大値 $\text{Acorr\_max}[f]$ 、周波数別SN比総和 $\text{snr\_sum}[f]$ 、周波数別SN比分散 $\text{snr\_var}[f]$ を例に挙げて説明する。

【0035】

まず、式(1)に示すように、フレーム単位でのパワーで正規化されて絶対値をとった $k$ 次自己相関係数 $\text{Acorr}[f,k]$  ( $k=1,\dots,N-1$ )を計算する。

【数1】

$$\text{Acorr}[f,k] = \frac{\left| \sum_{n=0}^{N-1-k} z[n] \cdot z[n+k] \right|}{\left| \sum_{n=0}^{N-1} z[n] \cdot z[n] \right|} \quad \dots(1)$$

このとき併せて、 $k=1$ である1次自己相関係数 $\text{Acorr}[f,1]$ も計算する。1次自己相関係数 $\text{Acorr}[f,1]$ は0から1の値を取り、0に近づくほどノイズ性が強い。つまり、1次自己相関係数の値が小さいほど、集音信号に周囲雑音が多く含まれ、周囲雑音には含まない音声信号が少ないと判断される。そして、正規化された $k$ 次自己相関係数 $\text{Acorr}[f,k]$  ( $k=1,\dots,N-1$ )から式(2)に示すように、最大となる自己相関係数 $\text{Acorr}[f,k]$ を計算して、自己相関係数最大値 $\text{Acorr\_max}[f]$ とする。自己相関係数最大値 $\text{Acorr\_max}[f]$ は0から1の値を取り、0に近づくほどノイズ性が強い。つまり、自己相関係数の値が小さいほど、集音信号に周囲雑音が多く含まれ、周囲雑音には含まない音声信号が少ないと判断される。

【数2】

$$\text{Acorr\_max}[f] = \max_{k=1}^{N-1} \{ \text{Acorr}[f,k] \} \quad \dots(2)$$

次に、パワースペクトル $|Z[f,w]|^2$ と周囲雑音のパワースペクトル $|N[f,w]|^2$ とを入力として、それらの比である各周波数帯域のSN比を、ここではdB表現した $\text{snr}[f,w]$  ( $w=0,1,\dots,M-1$ )として式(3)で算出する。

【数3】

$$\text{snr}[f,\omega] = 10 \log_{10} \left( \frac{|Z[f,\omega]|^2}{|N[f,\omega]|^2} \right) \quad \dots(3)$$

そして、各周波数帯域のSN比 $\text{snr}[f,w]$  ( $w=0,1,\dots,M-1$ )の和を式(4)で算出し、周波数別SN比総和値 $\text{snr\_sum}[f]$ とする。周波数別SN比総和値 $\text{snr\_sum}[f]$ は0以上の値を取り、この値が小さいほど集音信号中にノイズ成分である周囲雑音が多く含まれ、周囲雑音には含まない音声信号が少ないと判断される。

【数4】

$$\text{snr\_sum}[f] = \sum_{\omega=0}^{M-1} \text{snr}[f,\omega] \quad \dots(4)$$

また、各周波数帯域のSN比 $\text{snr}[f,w]$  ( $w=0,1,\dots,M-1$ )の分散を式(5)で算出し、周波数別SN比分散値 $\text{snr\_var}[f]$ とする。周波数別SN比分散値 $\text{snr\_var}[f]$ は0以上の値をと

10

20

30

40

50

り、この値が小さいほどノイズ成分である周囲雑音が多く含まれ、周囲雑音には含まない音声信号が少ないと判断される。

【数 5】

$$snr\_var[f] = \sum_{\omega=0}^{M-1} snr[f, \omega] - \frac{\sum_{i=0}^{M-1} snr[f, i]}{M} \quad \dots(5)$$

最後に、複数の特徴量である、1次自己相関係数 $Acorr[f, 1]$ 、自己相関係数最大値 $Accor\_max[f]$ 、周波数別SN比総和値 $snr\_sum[f]$ 、周波数別SN比分散値 $snr\_var[f]$ を用いて、これらにそれぞれ所定の重み付けによる重み付けを行い、複数の特徴量の重み付け和として周囲雑音度合 $type[f]$ を算出する。ここでは、周囲雑音度合 $type[f]$ が小さいほど周囲雑音が支配的であると、大きいほど周囲雑音には含まない音声信号が支配的であるとしているので、例えば、線形識別関数による判定を用いた学習アルゴリズムなどで重み $w_1$ 、 $w_2$ 、 $w_3$ 、 $w_4$ （ただし $w_1 > 0$ 、 $w_2 > 0$ 、 $w_3 > 0$ 、 $w_4 > 0$ ）を設定して、式(6)で算出する。そして、周囲雑音度合 $type[f]$ が所定の閾値 $THR$ よりも大きければ $vad[f]=1$ とし、周囲雑音度合 $type[f]$ が所定の閾値 $THR$ 以下であれば $vad[f]=0$ とする。

【数 6】

$$type[f] = w_1 \cdot Acorr[f, 1] + w_2 \cdot Accor\_max[f] + w_3 \cdot snr\_sum[f] + w_4 \cdot snr\_var[f] \quad \dots(6)$$

以上の説明では、複数の特徴量を求める際に、周波数ビンごとに処理するとして説明したが、周波数領域変換による隣接する複数の周波数ビンをもとめてグループを作り、そのグループ単位で処理を行っても構わない。また、フィルタバンクなどの帯域分割フィルタなどの周波数領域変換によって実現してもよい。

【0036】

なお、前述した複数の特徴量を全て使わなくてもよいし、他の特徴量を追加して用いてもよい。また、無線通信部1あるいはデコーダ2から出力されるコーデック情報、例えば、無音挿入記述子(SID)や音声検出器(VAD)による音声であるか音声でないかを表す音声検出情報や擬似背景雑音を生成したかどうかの情報などを用いてもよい。

【0037】

周波数スペクトル更新部314は、周囲雑音区間判定部313から出力されるフレーム判定情報 $vad[f]$ と、パワー算出部312から出力されるパワースペクトル $|Z[f, w]|^2$  ( $w=0, 1, \dots, M-1$ )を用いて、各周波数帯域の周囲雑音のパワースペクトルである $|N[f, w]|^2$  ( $w=0, 1, \dots, M-1$ )を推定して出力する。例えば、フレーム判定情報 $vad[f]$ を0として周囲雑音が支配的に含まれる区間(周囲雑音区間)であると判別されたフレームのパワースペクトル $|Z[f, w]|^2$ をフレーム単位で忘却させて平均的なパワースペクトルを算出し、これを各周波数帯域の周囲雑音のパワースペクトル $|N[f, w]|^2$  ( $w=0, 1, \dots, M-1$ )として出力する。なお、具体的には、各周波数帯域の周囲雑音のパワースペクトル $|N[f, w]|^2$ の算出は、式(8)に示すように1フレーム前の各周波数帯域の周囲雑音のパワースペクトル $|N[f-1, w]|^2$ を用いて再帰的に行う。ただし、式(7)の忘却係数 $\alpha_N$ は1以下の係数であって、好ましくは0.75~0.95程度である。

【数 7】

$$|N[f, \omega]|^2 = \alpha_N[\omega] \cdot |N[f-1, \omega]|^2 + (1 - \alpha_N[\omega]) \cdot |Z[f, \omega]|^2 \quad \dots(7)$$

周囲雑音情報帯域拡張部32は、各周波数帯域の周囲雑音のパワースペクトル $|N[f, w]|^2$ を用いて、入力信号 $x[n]$ には存在して集音信号 $z[n]$ には存在しない周波数帯域成分を含めた信号のパワー値を生成する。

10

20

30

40

50

## 【 0 0 3 8 】

図 4 は周囲雑音情報帯域拡張部 3 2 の構成例を示す図である。周囲雑音情報帯域拡張部 3 2 は、パワー正規化部 3 2 1 と、辞書格納部 3 2 2 と、広帯域パワー算出部 3 2 3 とを備える。

## 【 0 0 3 9 】

周囲雑音情報帯域拡張部 3 2 では、狭帯域信号情報から狭帯域特徴量データを算出し、狭帯域信号情報から算出される狭帯域特徴量データと広帯域特徴量データとの対応を事前にモデル化しておき、このモデルと取得した狭帯域特徴量データとの対応を用いて広帯域特徴量データを算出し、広帯域特徴量データから広帯域信号情報を生成する。前述のとおり、ここでは、狭帯域信号情報は周囲雑音のパワースペクトルである。またここでは、広帯域特徴量データと広帯域信号情報は同じであるとし、広帯域信号情報は広帯域パワー値  $N_{wb\_level}[f]$  で示される音量である。狭帯域特徴量データと広帯域特徴量データとの対応のモデル化には、GMM(Gaussian mixture model)を利用する手法を用いる。ここでは、狭帯域パワー値  $Pow\_N[f]$  と周囲雑音の正規化したパワースペクトル  $|N_n[f, w]|^2$  ( $w=0, 1, \dots, M-1$ ) を次数方向に連結して  $D_{nb}$  次の狭帯域特徴量データとして用い、広帯域パワー値  $N_{wb\_level}[f]$  を  $D_{wb}$  次の広帯域特徴量データとして用いる ( $D_{nb}=M+1$ 、 $D_{wb}=1$ )。

10

## 【 0 0 4 0 】

まず狭帯域信号情報から狭帯域特徴量データを算出するために、パワー正規化部 3 2 1 には、周囲雑音推定部 3 1 から出力される周囲雑音のパワースペクトル  $|N[f, w]|^2$  ( $w=0, 1, \dots, M-1$ ) が入力され、この周囲雑音のパワースペクトルを用いて狭帯域特徴量データを算出する。狭帯域特徴量データの 1 つは、式 ( 8 ) に基づいて算出される、パワースペクトルの各周波数ピンの総和である狭帯域パワー値  $Pow\_N[f]$  である。

20

## 【 数 8 】

$$Pow\_N[f] = \sum_{\omega=0}^{M-1} |N[f, \omega]|^2 \quad \dots(8)$$

また他の狭帯域特徴量データとしては、狭帯域パワー値  $Pow\_N[f]$  を用いて式 ( 9 ) に従って各周波数ピンのパワースペクトル  $|N[f, w]|^2$  を正規化したパワースペクトル  $|N_n[f, w]|^2$  を算出する。

30

## 【 数 9 】

$$|N_n[f, \omega]|^2 = \frac{|N[f, \omega]|^2}{Pow\_N[f]} \quad \dots(9)$$

辞書格納部 3 2 2 は、事前に集音した周囲雑音に基づいて  $D_{nb}$  次の狭帯域特徴量データと  $D_{wb}$  次の広帯域特徴量データとの対応をモデル化して学習された混合数  $Q$  (ここでは  $Q=64$ ) の GMM の辞書  $\mathcal{1}_q = \{w_q, \mu_q, \Sigma_q\}$  ( $q=1, \dots, Q$ ) を格納する。なお、 $w_q$  は  $q$  番目の混合正規分布の混合重みを示し、 $\mu_q$  は  $q$  番目の混合正規分布の平均ベクトル、 $\Sigma_q$  は  $q$  次数目の混合正規分布の共分散行列 (対角共分散行列または全共分散行列) を表している。なお、平均ベクトル  $\mu_q$  と共分散行列  $\Sigma_q$  の成分の数である次数は、 $D_{nb} + D_{wb}$  である。

40

## 【 0 0 4 1 】

辞書格納部 3 2 2 における事前の辞書  $\mathcal{1}_q$  の学習生成方法について、フローチャートを図 5 に示し、説明する。

## 【 0 0 4 2 】

GMM の生成に用いる信号は、入力信号  $x[n]$  と同様のサンプリング周波数  $f_{s'} [Hz]$  で  $f_{s\_wb\_low}[Hz]$  から  $f_{s\_wb\_high}[Hz]$  までに帯域制限された広帯域な信号を別途事前に集音した信号群である。この信号群は、多数の様々な環境、様々な音量であることが望ましい。以下では、GMM の生成に用いる広帯域信号の信号群をまとめて広帯域信号データ  $wb[n]$  と表記する。  $n$  は時刻 (サンプル) を表す。

50

## 【 0 0 4 3 】

まず、広帯域信号データ $wb[n]$ を入力として、ダウンサンプリングフィルタによってサンプリング周波数 $f_s[\text{Hz}]$ にダウンサンプリングし、 $f_{s\_nb\_low}[\text{Hz}]$ から $f_{s\_nb\_high}[\text{Hz}]$ までの狭帯域に帯域制限された狭帯域信号データ $nb[n]$ を得る（ステップS 1 0 1）。このようにして、集音信号 $z[n]$ と同じように帯域制限された信号群を生成する。なお、図示しないが、上記ダウンサンプリングフィルタや帯域制限処理でアルゴリズム遅延が生じる場合には、狭帯域信号データ $nb[n]$ を広帯域信号データ $wb[n]$ と同期を合わせる処理を行う。

## 【 0 0 4 4 】

次に、上記狭帯域信号データ $nb[n]$ からフレーム $f$ 単位で狭帯域特徴量データ $P_{nb}[f, d]$  ( $d=1, \dots, D_{nb}$ ) を抽出する（ステップS 1 0 2）。狭帯域特徴量データ $P_{nb}[f, d]$ は、所定の次数の狭帯域信号情報を表す特徴量データである。ステップS 1 0 2ではまず、狭帯域信号データ $nb[n]$ からフレーム毎に前述の周波数領域変換部3 1 1における処理と同様に周波数領域変換処理を行い、 $M$ 次の狭帯域信号データ $nb[n]$ のパワースペクトルを得る（ステップS 1 0 2 1）。次に、前述のパワー正規化部3 2 1における処理と同様の処理によって、狭帯域信号データ $nb[n]$ からフレーム毎にパワー算出を行い、1次のパワー値を得る（ステップS 1 0 2 2）。そして、これらのパワースペクトルとパワー値から $M$ 次の狭帯域信号データ $nb[n]$ の正規化されたパワースペクトルを得る（ステップS 1 0 2 3）。そして、 $M$ 次の正規化されたパワースペクトルと1次のパワー値をフレーム単位で次数方向（次元方向）に連結して、次数 $D_{nb}$  ( $=M+1$ ) の狭帯域特徴量データ $P_{nb}[f, d]$  ( $d=1, \dots, D_{nb}$ ) を生成する（ステップS 1 0 2 4）。

## 【 0 0 4 5 】

一方、上記に並行して、広帯域信号データ $wb[n]$ からフレーム $f$ 単位で広帯域特徴量データ $P_{wb}[f, d]$  ( $d=1, \dots, D_{wb}$ ) を抽出する（ステップS 1 0 3）。広帯域特徴量データ $P_{wb}[f, d]$ は、所定の次数の広帯域信号情報を表す特徴量データである。ステップS 1 0 3ではまず、広帯域信号データ $wb[n]$ からフレーム毎に前述の周波数領域変換部3 1 1における処理のFFT点数を倍の $4M$ 点にして、同様に周波数領域変換処理を行い、 $2M$ 次の広帯域信号データ $wb[n]$ のパワースペクトルを得る（ステップS 1 0 3 1）。次に、前述のパワー正規化部3 2 1における処理と同様の処理によって、広帯域信号データ $wb[n]$ からフレーム毎にパワー算出を行って1次のパワー値を得る。このパワー値を次数 $D_{wb}$  ( $=1$ ) の広帯域特徴量データ $P_{wb}[f, d]$ とする（ステップS 1 0 3 2）。

## 【 0 0 4 6 】

次に、狭帯域特徴量データ $P_{nb}[f, d]$  ( $d=1, \dots, D_{nb}$ ) と広帯域特徴量データ $P_{wb}[f, d]$  ( $d=1, \dots, D_{wb}$ ) の時間的に同期が取れた2つの特徴量データをフレーム単位で次数方向（次元方向）に連結して、次数 $D_{nb}+D_{wb}$ の連結特徴量データ $P[f, d]$  ( $d=1, \dots, D_{nb}+D_{wb}$ ) を生成する（ステップS 1 0 4）。

## 【 0 0 4 7 】

そして、連結特徴量データ $P[f, d]$ から混合数 $Q=1$ の初期GMMを生成し、各GMMの平均ベクトルをわずかにずらして別の混合分布を生成することで混合数 $Q$ を2倍に増やす処理と、連結特徴量データ $P[f, d]$ を用いてEMアルゴリズムにより収束するまでGMMの尤度最大化学習を行う処理とを交互に繰り返し行い、混合数 $Q$ （ここでは $Q=64$ ）のGMM  $\theta_q = \{w_q, \mu_q, \sigma_q\}$  ( $q=1, \dots, Q$ ) を生成する（ステップS 1 0 5）。EMアルゴリズムについては、D.A. Reynolds and R.C. Rose, "Robust text-independent speaker identification using Gaussian mixture models", IEEE Trans. Speech and Audio Processing, Vol.3, no.1, pp.72-83, Jan.1995. などの文献に詳細な記述がある。

図4の説明に戻る。広帯域パワー算出部3 2 3には、パワー正規化部3 2 1から出力された狭帯域パワー値 $Pow\_N[f]$ と周囲雑音の正規化した $D_{nb}$ 次のパワースペクトル $|N_n[f, w]|^2$  ( $w=0, 1, \dots, M-1$ )が連結されて、狭帯域特徴量データ $P_{n\_nb}[f]$  ( $d=1, \dots, D_{nb}$ )として入力される。また、広帯域パワー算出部3 2 3は、辞書格納部3 2 2からGMMの辞書  $\theta_q = \{w_q, \mu_q, \sigma_q\}$  ( $q=1, \dots, Q$ )を読み出して、最小平均2乗誤差（MMSE: Minimam Mean Square Error）推定に従って、式（10）に示すように、複数の正規分布モデルによるソフトクラ

10

20

30

40

50

スタリングと連続的な線形回帰によって、周波数帯域が拡張された広帯域に対応する特徴量データへの変換を行い、狭帯域特徴量データ $Pn\_nb[f]$ から広帯域特徴量データである広帯域パワー値 $N\_wb\_level[f]$ を算出して出力する。式(10)は、次元( $d=1, \dots, Dnb+Dwb$ )方向のベクトルとして記載している。また、平均ベクトル $\mu_q$  ( $d=1, \dots, Dnb+Dwb$ )は次元方向で、 $\mu_q^N$  ( $d=1, \dots, Dnb$ )と $\mu_q^W$  ( $d=Dnb, \dots, Dnb+Dwb$ )に分割し、 $(Dn+Dw) \times (Dn+Dw)$ 行列である共分散行列 $\Sigma_q$ も以下のように、 $Dn \times Dn$ 行列である $\Sigma_q^{NN}$ と $Dn \times Dw$ 行列である $\Sigma_q^{NW}$ と $Dw \times Dn$ 行列である $\Sigma_q^{WN}$ と $Dw \times Dw$ 行列である $\Sigma_q^{WW}$ とに分割する。

【数10】

$$N\_wb\_level[f] = \sum_{q=1}^Q h_q(Pn\_nb[f]) \cdot \left\{ \mu_q^W + \Sigma_q^{WN} (\Sigma_q^{NN})^{-1} (Pn\_nb[f] - \mu_q^N) \right\} \dots (10) \quad 10$$

$$\text{ただし、} \quad h_q(x) = \frac{\frac{w_q}{(2\pi)^{\frac{Dnb+Dwb}{2}} \cdot |\Sigma_q^{NN}|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(x - \mu_q^N)^T (\Sigma_q^{NN})^{-1} (x - \mu_q^N)\right\}}{\sum_{j=1}^Q \frac{w_j}{(2\pi)^{\frac{Dnb+Dwb}{2}} \cdot |\Sigma_j^{NN}|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(x - \mu_j^N)^T (\Sigma_j^{NN})^{-1} (x - \mu_j^N)\right\}},$$

$$\Sigma_q = \begin{bmatrix} \Sigma_q^{NN} & \Sigma_q^{NW} \\ \Sigma_q^{WN} & \Sigma_q^{WW} \end{bmatrix}, \quad 20$$

$$\mu_q = \begin{bmatrix} \mu_q^N \\ \mu_q^W \end{bmatrix} \text{とする。}$$

周囲雑音情報帯域拡張部32では、広帯域特徴量データと広帯域信号情報は同じであるとしたため、このようにして、狭帯域信号情報である周囲雑音のパワースペクトル $|N[f, w]|^2$ から、広帯域信号情報である広帯域パワー値 $N\_wb\_level[f]$ が得られる。

【0048】

図6に信号特性補正部33の構成例を示す。信号特性補正部33は、周波数領域変換部331と、補正度合決定部332と、補正処理部333と、時間領域変換部334とを備える。信号特性補正部33には、入力信号 $x[n]$  ( $n=0, 1, \dots, 2N-1$ )と広帯域パワー値 $N\_wb\_level[f]$ が入力され、入力信号 $x[n]$ が集音信号に含まれる周囲雑音に埋もれてしまわないよう明瞭化する信号補正処理を行い、その補正後の出力信号 $y[n]$  ( $n=0, 1, \dots, 2N-1$ )を出力する。

【0049】

周波数領域変換部331には、周波数領域変換部311における集音信号 $z[n]$  ( $n=0, 1, \dots, N-1$ )の代わりに入力信号 $x[n]$  ( $n=0, 1, \dots, 2N-1$ )が入力される。周波数領域変換部331は、周波数領域変換部311と同様の処理によって、入力信号 $x[n]$ の周波数スペクトル $X[f, w]$ を出力する。例えば、周波数領域変換部331は、1フレーム前とのオーバーラップのサンプル数を $L=96$ とし、1フレーム前の入力信号 $L$ サンプルと当該フレームの入力信号 $x[n]$ の $2N=320$ サンプル分と $L$ サンプル分の零詰めから、 $4M=512$ サンプルを用意する。そして、この $4M$ サンプルに対して正弦波窓による窓関数を乗じることで窓掛けを行った信号に対して、FFTの次数を $4M$ としてFFTによる周波数領域変換を行う。

【0050】

補正度合決定部332には、周囲雑音情報帯域拡張部32から出力された広帯域パワー値 $N\_wb\_level[f]$ が入力される。そして、式(11)で補正ゲイン $G[f, w]$  ( $w=0, 1, \dots, 2M-1$ )を算出して出力する。

30

40

## 【数 1 1】

$$G[f, \omega] = \frac{N_{wb\_level}[f]}{N_0} \quad \dots(11)$$

式(11)の $N_0$ は、通常の利用環境における周囲雑音のパワーを入力信号 $x[n]$ と同じサンプリング周波数・同じ帯域制限で事前に測定しておいて設定した周囲雑音の基準パワー値である。このようにすることで、通常の利用環境よりも周囲雑音のパワー値が大きい環境(すなわち、周囲雑音が多く含まれる環境)でも、補正ゲイン $G[f, \omega]$ をその分大きく設定することで、入力信号 $x[n]$ を明瞭化することができる。

## 【0051】

補正処理部333には、入力信号 $x[n]$ の周波数スペクトル $X[f, \omega]$  ( $\omega=0, 1, \dots, 2M-1$ )と補正度合決定部332から出力された補正ゲイン $G[f, \omega]$  ( $\omega=0, 1, \dots, 2M-1$ )が入力される。そして、式(12)によって入力信号 $x[n]$ の周波数スペクトル $X[f, \omega]$ を補正し、その補正結果である出力信号 $y[n]$ の周波数スペクトル $Y[f, \omega]$  ( $\omega=0, 1, \dots, 2M-1$ )を出力する。

## 【数 1 2】

$$Y[f, \omega] = G[f, \omega] \cdot X[f, \omega] \quad \dots(12)$$

時間領域変換部334は、補正処理部333から出力された周波数スペクトル $Y[f, \omega]$  ( $\omega=0, 1, \dots, 2M-1$ )に対して時間領域変換(周波数逆変換)を行って、周波数領域変換部331における窓掛けを考慮しオーバーラップを戻す処理を適宜行い、補正された信号である出力信号 $y[n]$ を算出する。例えば、周波数スペクトル $Y[f, \omega]$  ( $\omega=0, 1, \dots, 2M-1$ )に対して、入力信号 $x[n]$ が実信号であることを考慮に入れて周波数スペクトル $Y[f, \omega]$ を $\omega=0, 1, \dots, 4M-1$ まで復元した上で、 $4M$ 点のIFFT(Inverse Fast Fourier Transform)を行い、窓掛けを考慮し1フレーム前の補正された信号である出力信号 $y[n]$ を用いてオーバーラップを戻し、出力信号 $y[n]$ を算出する。

## 【0052】

以上のように、再生される入力信号と集音信号で、信号成分が存在する周波数帯域が異なっていたり、サンプリング周波数が異なっていたりしていても、集音信号の音量について入力信号の周波数帯域を加味して拡張し推定することで、集音信号の音量が高精度に求められ、入力信号の明瞭度を向上させることができる。

## 【0053】

なお、上述の説明では、本発明を通信装置に適用した場合について説明したが、図7(a)に示すように、デジタルオーディオプレイヤーに本発明を適用することも可能である。このデジタルオーディオプレイヤーは、フラッシュメモリやHDD(Hard Disk Drive)を用いた記憶部11を備え、この記憶部11から読み出した音楽・オーディオデータをデコーダ2が復号する。このとき復号して再生したい所望の信号である目的信号は音楽・オーディオ信号である。このデジタルオーディオプレイヤーのマイク6で集音される集音信号 $z(t)$ は、近端話者の音声や周囲環境に起因するノイズ成分、出力信号 $y(t)$ と音響空間に起因するエコー成分などで構成され、音楽・オーディオ信号は含まれない。この場合は、通信装置と異なって近端話者の音声は不要であるため、近端話者の音声を含んだこれら全ての成分を周囲雑音として取り扱う。

## 【0054】

また、図7(b)に示すように、本発明を通信装置に適用して、音声帯域拡張通話装置に適用することも可能である。この音声帯域拡張通話装置は、デコーダ2Aを備え、デコーダ2Aと信号処理部3の間に信号帯域拡張処理部12を備えた構成である。そして、この場合の信号処理部3は、帯域が拡張された入力信号 $x'[n]$ に対して前述の処理をする。

## 【0055】

なお、信号帯域拡張処理部12で行われる処理は、 $fs\_nb\_low$ [Hz]から $fs\_nb\_high$ [Hz]までに帯域制限された狭帯域の入力信号を $fs\_wb\_low$ [Hz]から $fs\_wb\_high$ [Hz]までの広帯

10

20

30

40

50

域の信号に帯域を拡張する処理であって、例えば、特登 3 1 8 9 6 1 4 号公報や特登 3 2 4 3 1 7 4 号公報や特開平 9 - 5 5 7 7 8 号公報などに記載される既存の技術で実施してよい。

【 0 0 5 6 】

( 信号処理部の変形例 1 )

次に、信号処理部にて用いる狭帯域信号情報は周囲雑音のパワースペクトル、広帯域信号情報は周囲雑音を広帯域の信号に拡張した場合のマスキング閾値 ( 広帯域マスキング閾値 ) である場合を例にして説明する。

【 0 0 5 7 】

図 8 は、その構成を示すものである。信号処理部 3 0 は、信号処理部 3 で用いていた周囲雑音情報帯域拡張部 3 2 および信号特性補正部 3 3 に代わって、周囲雑音情報帯域拡張部 3 4 と、信号特性補正部 3 5 とを備えて構成される。

【 0 0 5 8 】

図 9 に周囲雑音情報帯域拡張部 3 4 の構成例を示す。周囲雑音情報帯域拡張部 3 4 は、パワー正規化部 3 2 1 と、辞書格納部 3 4 2 と、広帯域パワースペクトル算出部 3 4 3 と、広帯域マスキング閾値算出部 3 4 4 と、パワー制御部 3 4 5 とを備える。

【 0 0 5 9 】

周囲雑音情報帯域拡張部 3 4 は、周囲雑音情報帯域拡張部 3 2 と同様に、周囲雑音のパワースペクトルを入力として、入力信号  $x[n]$  には存在して集音信号  $z[n]$  には存在しない周波数帯域成分を含めた情報 ( 広帯域信号情報 ) を生成する。つまり、周囲雑音情報帯域拡張部 3 4 では、狭帯域信号情報から狭帯域特徴量データを算出し、狭帯域信号情報から算出される狭帯域特徴量データと広帯域特徴量データとの対応を事前にモデル化しておき、このモデルと取得した狭帯域特徴量データとの対応を用いて広帯域特徴量データを算出し、広帯域特徴量データから広帯域信号情報を生成する。ただし、周囲雑音情報帯域拡張部 3 4 では、狭帯域特徴量データと広帯域特徴量データとの対応のモデル化に、ベクトル量子化によるコードブックを利用する手法を用いる。ここでは、周囲雑音の正規化したパワースペクトル  $|N_n[f, w]|^2$  ( $w=0, 1, \dots, M-1$ ) を  $D_{nb}$  次の狭帯域特徴量データとして用い、周囲雑音の正規化した広帯域のパワースペクトル  $|N_w[f, w]|^2$  ( $w=0, 1, \dots, 2M-1$ ) を  $D_{wb}$  次の広帯域特徴量データとして用いる ( $D_{nb}=M$ 、 $D_{wb}=2M$ )。具体的には、周囲雑音情報帯域拡張部 3 4 は、周囲雑音のパワースペクトル  $|N[f, w]|^2$  ( $w=0, 1, \dots, M-1$ ) を入力として、周囲雑音のパワースペクトル  $|N[f, w]|^2$  について入力信号  $x[n]$  には存在して集音信号  $z[n]$  には存在しない周波数帯域成分のパワースペクトルを周波数帯域拡張によって生成して、その帯域拡張されたパワースペクトルに対してマスキング閾値を求め、その結果である広帯域マスキング閾値  $N_{wb\_th}[f, w]$  ( $w=0, 1, \dots, 2M-1$ ) を出力する。

【 0 0 6 0 】

辞書格納部 3 4 2 は、 $D_{nb}$  次の狭帯域特徴量データと  $D_{wb}$  次の広帯域特徴量データとの対応をモデル化して事前に学習されたサイズ  $Q$  (ここでは  $Q=64$ ) のコードブックの辞書  $2_q = \{ \mu x_q, \mu y_q \}$  ( $q=1, \dots, Q$ ) を格納している。なお、 $\mu x_q$  は  $q$  番目のコードブックにおける狭帯域特徴量データのセントロイドベクトル、 $\mu y_q$  は  $q$  番目のコードブックにおける広帯域特徴量データのセントロイドベクトルを表している。なお、コードブックのコードベクトルの次数は、狭帯域特徴量データのセントロイドベクトル  $\mu x_q$  と広帯域特徴量データのセントロイドベクトル  $\mu y_q$  の成分の和である  $D_{nb} + D_{wb}$  である。

【 0 0 6 1 】

辞書格納部 3 4 2 における事前の辞書  $2_q$  の学習生成方法について、フローチャートを図 1 0 に示し、説明する。以下の説明では、前述の辞書  $1_q$  の学習生成方法と同じ処理については同じ番号を付番し、説明を簡明にするために必要に応じて重複する説明を省略する。

【 0 0 6 2 】

コードブックの辞書生成に用いる信号は、入力信号  $x[n]$  と同様にサンプリング周波数  $f_s$  [Hz] で  $f_{s\_wb\_low}$  [Hz] から  $f_{s\_wb\_high}$  [Hz] までに帯域制限された広帯域な信号を別途事

10

20

30

40

50

前に集音した信号群である。この信号群は、多数の様々な環境、様々な音量であることが望ましい。以下では、コードブックの辞書生成に用いる広帯域信号の信号群をまとめて広帯域信号データ $wb[n]$ と表記する。また、 $n$ は時刻(サンプル)を表す。

#### 【0063】

まず、広帯域信号データ $wb[n]$ を入力として、サンプリング周波数 $f_s[\text{Hz}]$ にダウンサンプリングし狭帯域信号データ $nb[n]$ を得る(ステップS101)。そして、狭帯域信号データ $nb[n]$ から狭帯域信号情報を表す特徴量データである狭帯域特徴量データ $Pnb[f,d]$ ( $d=1, \dots, Dnb$ )を抽出する(ステップS202)。このステップS202では、狭帯域信号データ $nb[n]$ のパワースペクトル(M次)を得て(ステップS1021)、狭帯域信号データ $nb[n]$ のパワー値を得て(ステップS1022)、これらのパワースペクトルとパワー値から狭帯域信号データ $nb[n]$ の正規化されたパワースペクトルを得て(ステップS1023)、これを次数 $Dnb(=M)$ の狭帯域特徴量データ $Pnb[f,d]$ ( $d=1, \dots, Dnb$ )とすることによって狭帯域特徴量データの抽出を行う。

10

#### 【0064】

一方、広帯域信号データ $wb[n]$ から広帯域信号情報を表す特徴量データである広帯域特徴量データ $Pwb[f,d]$ ( $d=1, \dots, Dwb$ )を抽出する(ステップS203)。このステップS203では、広帯域信号データ $wb[n]$ のパワースペクトルを得て(ステップS1031)、広帯域信号データ $wb[n]$ から広帯域信号データ $wb[n]$ のパワー値をフレーム単位で得て(ステップS2032)、これらのパワースペクトルとパワー値から広帯域信号データ $wb[n]$ の正規化されたパワースペクトルをフレーム単位で得て(ステップS2033)、これを次数 $Dwb(=2M)$ の広帯域特徴量データ $Pwb[f,d]$ ( $d=1, \dots, Dwb$ )とすることによって広帯域特徴量データの抽出を行う。

20

#### 【0065】

次に、狭帯域特徴量データ $Pnb[f,d]$ ( $d=1, \dots, Dnb$ )と広帯域特徴量データ $Pwb[f,d]$ ( $d=1, \dots, Dwb$ )を連結して、次数 $Dnb+Dwb$ の連結特徴量データ $P[f,d]$ ( $d=1, \dots, Dnb+Dwb$ )を生成する(ステップS104)。

#### 【0066】

上記連結特徴量データ $P[f,d]$ からサイズ $Q$ (ここでは $Q=64$ )のコードブックの辞書 $2_q = \{\mu x_q, \mu y_q\}$ ( $q=1, \dots, Q$ )を $k$ - $m$ - $e$ - $a$ - $n$ - $s$ アルゴリズムや $L$ - $B$ - $G$ アルゴリズムなどによるクラスタリング手法を用いて生成する(ステップS205)。ステップS205では、まず狭帯域セントロイドベクトル $\mu x_1$ を狭帯域特徴量データの全部の平均とし、広帯域セントロイドベクトル $\mu y_1$ を広帯域特徴量データの全部の平均としてサイズ $Q=1$ の初期コードブックを生成する(ステップS2051)。そして、コードブックのサイズ $Q$ が所定数(ここでは64)に達したかどうかを判定する(ステップS2052)。コードブックのサイズ $Q$ が所定数に達していなければ、コードブック $2_q$ の各コードベクトルにおける狭帯域セントロイドベクトル $\mu x_q$ と広帯域セントロイドベクトル $\mu y_q$ をわずかにずらして別のコードベクトルを生成することでコードブックのサイズ $Q$ を2倍に増やす処理を行う(ステップS2053)。そして、次数 $Dnb+Dwb$ の連結特徴量データ $P[f,d]$ について、コードブック $2_q$ の各コードベクトルにおける狭帯域セントロイドベクトル $\mu x_q$ との所定の距離尺度(例えばユークリッド距離やマハラノビス距離)が最小となるコードベクトルを求めて、連結特徴量データ $P[f,d]$ をその該当するコードベクトルに割り当てる。その後、コードブック $2_q$ のコードベクトルごとに割り当てられた連結特徴量データ $P[f,d]$ を用いて、コードベクトルごとに新しい狭帯域セントロイドベクトル $\mu x_q$ と広帯域セントロイドベクトル $\mu y_q$ を求めて、コードブック $2_q$ を更新する(ステップS2054)。コードブックのサイズ $Q$ が所定数に達していれば、そのコードブック $2_q = \{\mu x_q, \mu y_q\}$ ( $q=1, \dots, Q$ )を出力する。

30

40

#### 【0067】

広帯域パワースペクトル算出部343は、パワー正規化部321から出力された周囲雑音の正規化したパワースペクトル $|Nn[f,w]|^2$ ( $w=0, 1, \dots, M-1$ )を $Dnb$ 次の特徴量データとして入力し、辞書格納部342からコードブックの辞書 $2_q = \{\mu x_q, \mu y_q\}$ ( $q=1, \dots, Q$ )を読

50

み出して、Dnb次の狭帯域特徴量データとDwb次の広帯域特徴量データとの対応から広帯域パワースペクトル $|N_w[f, w]|^2$  ( $w=0, 1, \dots, 2M-1$ )を求める。具体的には、Q個ある狭帯域セントロイドベクトル $\mu x_q$  ( $q=1, \dots, Q$ )から、周囲雑音の正規化したパワースペクトル $|N_n[f, w]|^2$  ( $w=0, 1, \dots, M-1$ )と所定の距離尺度で一番距離が近いものを求めて、一番距離が近いコードベクトルにおける広帯域セントロイドベクトル $\mu y_q$ を広帯域パワースペクトル $|N_w[f, w]|^2$  ( $w=0, 1, \dots, 2M-1$ )とする。

【 0 0 6 8 】

広帯域マスキング閾値算出部 3 4 4 は、広帯域パワースペクトル算出部 3 4 3 から出力される広帯域パワースペクトル $|N_w[f, w]|^2$  ( $w=0, 1, \dots, 2M-1$ )を入力として、周波数成分ごとに周囲雑音のマスキング閾値である広帯域マスキング閾値 $N\_wb\_th1[f, w]$  ( $w=0, 1, \dots, 2M-1$ )を算出する。

10

【 0 0 6 9 】

一般にマスキング閾値は、spreading functionと呼ばれる関数を信号のパワースペクトルに畳み込むことで算出することができる。すなわち、周囲雑音の広帯域マスキング閾値 $N\_wb\_th1[f, w]$  ( $w=0, 1, \dots, 2M-1$ )は、spreading functionを関数 $sprdnf()$ として、式(13)の式で算出される。周囲雑音の広帯域パワースペクトル $|N_w[f, w]|^2$ が広帯域マスキング閾値 $N\_wb\_th1[f, w]$ 以下であるならば、周波数ビン 以外の周波数帯域の周囲雑音の広帯域パワースペクトルによってマスクされる。図 1 1 に、横軸を周波数[Hz]、縦軸をパワー[dB]として、屋外など様々な環境で採取した周囲雑音の広帯域マスキング閾値の例を示す。

20

【 数 1 3 】

$$N\_wb\_th1[f, \omega] = \sum_{j=0}^{2M-1} |N_w[f, j]|^2 \cdot sprdnf(bark[j], bark[\omega]) \quad \dots(13)$$

ここで、 $bark[w]$ は周波数ビン をバーク尺度に変換したバーク値を表し、spreading functionでは、バーク尺度 $bark[w]$ に適宜変換する。バーク尺度は、聴覚の分解能を考慮して、低域ほど細かく、高域ほど粗く設定された尺度である。

【 0 0 7 0 】

ここでは、spreading functionを関数 $sprdnf()$ として、ISO/IEC13818-7で定義されている方式を用いるとする。spreading functionは、例えばITU-R1387、3GPP TS 26.403といった文献で説明されている他の方式を用いても良い。なお、バーク尺度でなくても、メル尺度、ERB尺度など人間の音の高さの知覚特性や聴覚フィルタから得られた尺度を用いたspreading functionを適宜用いても構わない。

30

【 0 0 7 1 】

パワー制御部 3 4 5 は、パワー正規化 3 2 1 から出力された狭帯域パワー値 $Pow\_N[f]$ と広帯域マスキング閾値算出部 3 4 5 から出力された広帯域マスキング閾値 $N\_wb\_th1[f, w]$  ( $w=0, 1, \dots, 2M-1$ )を入力として、広帯域マスキング閾値 $N\_wb\_th1[f, w]$ の $fs\_nb\_low[Hz]$ から $fs\_nb\_high[Hz]$ におけるパワーを狭帯域パワー値 $Pow\_N[f]$ と同じになるように、広帯域マスキング閾値 $N\_wb\_th1[f, w]$ を増幅あるいは減衰させることで制御し、このパワー制御された $N\_wb\_th1[f, w]$ を広帯域マスキング閾値 $N\_wb\_th[f, w]$ として出力する。

40

【 0 0 7 2 】

このようにして、周囲雑音情報帯域拡張部 3 4 では、狭帯域信号情報である周囲雑音のパワースペクトル $|N[f, w]|^2$ から、広帯域信号情報である広帯域マスキング閾値 $N\_wb\_th[f, w]$ を求める。

【 0 0 7 3 】

図 1 2 に信号特性補正部 3 5 の構成例を示す。信号特性補正部 3 5 は、周波数領域変換部 3 3 1 と、パワー算出部 3 5 2 と、マスキング閾値算出部 3 5 3 と、マスキング判定部 3 5 4 と、パワー平滑化部 3 5 5 と、補正度合決定部 3 5 6 と、補正処理部 3 3 3 と、時間領域変換部 3 3 4 とを備える。

50

## 【 0 0 7 4 】

信号特性補正部 3 5 は、入力信号 $x[n]$  ( $n=0,1,\dots,2N-1$ )と広帯域マスキング閾値 $N_{wb\_th}[f,w]$ を入力とし、入力信号 $x[n]$ が集音信号に含まれる周囲雑音に埋もれてしまわないよう明瞭化する信号補正処理を行い、その補正後の出力信号 $y[n]$  ( $n=0,1,\dots,2N-1$ )を出力する。

## 【 0 0 7 5 】

パワー算出部 3 5 2 は、周波数領域変換部 3 3 1 から出力された入力信号 $x[n]$ の周波数スペクトル $X[f,w]$  ( $w=0,1,\dots,2M-1$ )における実部と虚部の 2 乗和であるパワースペクトル $|X[f,w]|^2$  ( $w=0,1,\dots,2M-1$ )を算出して出力する。

## 【 0 0 7 6 】

マスキング閾値算出部 3 5 3 は、パワー算出部 3 5 2 から出力された入力信号 $x[n]$ のパワースペクトル $|X[f,w]|^2$  ( $w=0,1,\dots,2M-1$ )を入力として、spreading functionを関数 $sprdnf()$ として、式 ( 1 4 ) の式で入力信号 $x[n]$ の広帯域マスキング閾値 $X_{th}[f,w]$  ( $w=0,1,\dots,2M-1$ )を算出し出力する。広帯域マスキング閾値 $X_{th}[f,w]$ は、入力信号 $x[n]$ のパワースペクトル $|X[f,w]|^2$  が入力信号 $x[n]$ の広帯域マスキング閾値 $X_{th}[f,w]$ 以下であるならば、周波数ビン 以外の周波数帯域の入力信号 $x[n]$ のパワースペクトル $|X[f,w]|^2$  によってマスクされることを表す。

## 【 数 1 4 】

$$X_{th}[f,\omega] = \sum_{j=0}^{2M-1} |X[f,j]|^2 \cdot sprdnf(bark[j],bark[\omega]) \quad \dots(14)$$

マスキング判定部 3 5 4 は、パワー算出部 3 5 2 から出力されたパワースペクトル $|X[f,w]|^2$  ( $w=0,1,\dots,2M-1$ )とマスキング閾値算出部 3 5 3 から出力された広帯域マスキング閾値 $X_{th}[f,w]$ とを入力とし、周波数帯域ごとに入力信号 $x[n]$ 自身によってマスクされるか否かを表すマスキング判定情報 $X_{flag}[f,w]$  ( $w=0,1,\dots,2M-1$ )を出力する。具体的には、パワースペクトル $|X[f,w]|^2$  と広帯域マスキング閾値 $X_{th}[f,w]$ の大小比較を行い、パワースペクトル $|X[f,w]|^2$  が広帯域マスキング閾値 $X_{th}[f,w]$ 以上ならば、その周波数成分は入力信号 $x[n]$ 中の他の周波数成分にマスクされないとして $X_{flag}[f,w]=0$ とする。また、パワースペクトル $|X[f,w]|^2$  が広帯域マスキング閾値 $X_{th}[f,w]$ 未満ならば、その周波数成分は入力信号 $x[n]$ 中の他の周波数成分にマスクされるとして $X_{flag}[f,w]=1$ とする

パワー平滑化部 3 5 5 は、パワー算出部 3 5 2 から出力されたパワースペクトル $|X[f,w]|^2$  ( $w=0,1,\dots,2M-1$ )とマスキング判定部 3 5 4 から出力されたマスキング判定情報 $X_{flag}[f,w]$ とを入力として、パワースペクトル $|X[f,w]|^2$  を式 ( 1 5 ) の式による移動平均によって平滑化して、平滑化されたパワースペクトル $|X_s[f,w]|^2$  を出力する。なお、 $K$ は平滑化を計算する範囲であり、 $\alpha_x[j]$ は、 $j$  が 0 に近いほど係数が大きくなるようなスムージング係数である。例えば、 $K=3$  で、 $\alpha_x[j]$ は  $[0.1, 0.2, 0.4, 0.8, 0.4, 0.2, 0.1]$  とする。

## 【 数 1 5 】

$$|X_s[f,\omega]|^2 = \frac{\sum_{j=-K}^K \alpha_x[j] \cdot |X[f,\omega+j]|^2}{\sum_{j=-K}^K \alpha_x[j]} \quad \dots(15)$$

補正度合決定部 3 5 6 は、パワー平滑化部 3 5 5 から出力された平滑化されたパワースペクトル $|X_s[f,w]|^2$  ( $w=0,1,\dots,2M-1$ )とマスキング判定部 3 5 4 から出力されたマスキング判定情報 $X_{flag}[f,w]$  ( $w=0,1,\dots,2M-1$ )と周囲雑音情報帯域拡張部 3 2 から出力された $N_{wb\_th}[f,w]$  ( $w=0,1,\dots,2M-1$ )とを入力として、補正ゲイン $G[f,w]$  ( $w=0,1,\dots,2M-1$ )を算出して出力する。補正ゲイン $G[f,w]$ の具体的な算出は、まずマスキング判定情報 $X_{flag}[f,w]$

10

20

30

40

50

により入力信号 $x[n]$ 中の他の周波数成分にマスクされる ( $X\_flag[f,w]=1$ ) と判定された周波数帯域であれば、 $G[f,w]=1$ とし補正による増幅も減衰も行わないようにする。そして、マスキング判定情報 $X\_flag[f,w]$ により入力信号 $x[n]$ 中の他の周波数成分にマスクされない ( $X\_flag[f,w]=0$ ) と判定された周波数帯域について、パワースペクトル $|X[f,w]|^2$ と周囲雑音の広帯域マスキング閾値 $N\_wb\_th[f,w]$ との大小比較を行う。ここで、パワースペクトル $|X[f,w]|^2$ が周囲雑音の広帯域マスキング閾値 $N\_wb\_th[f,w]$ 以上ならば、その周波数成分は集音信号 $z[n]$ 中の他の周波数成分にマスクされないで $G[f,w]=1$ とし、補正による増幅を行わないようにする。一方で、パワースペクトル $|X[f,w]|^2$ が周囲雑音の広帯域マスキング閾値 $N\_wb\_th[f,w]$ 未満ならば、集音信号 $z[n]$ 中の周囲雑音が少なければ知覚できるにも関わらず、周囲雑音があるためにマスクされていると判断し、式(16)の式のように補正ゲイン $G[f,w]$ を周囲雑音の広帯域マスキング閾値 $N\_wb\_th[f,w]$ と平滑化されたパワースペクトル $|X_s[f,w]|^2$ との比に基づいて算出する。なお、関数 $F$ は、平滑されたパワースペクトル $|X_s[f,w]|^2$ のスペクトル傾斜を周囲雑音の広帯域マスキング閾値 $N\_wb\_th[f,w]$ の形状と平行に近くなるように増幅するような関数である。ここで、 $\alpha$ は正の定数であり、 $\beta$ は正負いずれかの定数である。これらの定数は、入力信号 $x[n]$ の増幅度合いを調整するために用いられる。

10

【数16】

$$G[f,\omega]=F\left(\frac{N\_wb\_th[f,\omega]}{|X_s[f,\omega]|^2}\right) \quad \dots(16)$$

20

【数17】

$$F(x)=\alpha \cdot A^{\beta \cdot x + \gamma} \quad \dots(17)$$

補正度合決定部356において、このように求めた補正ゲイン $G[f,w]$ をさらに式(22)の式による三角窓による移動平均によって平滑化して、平滑化された補正ゲイン $G_s[f,w]$ を用いてもよい。なお、 $K$ は平滑化を計算する範囲であり、 $\alpha_g[j]$ は、 $j$ が0に近いほど係数が大きくなるようなスムージング係数である。例えば、 $K=3$ で、 $\alpha_g[j]$ は $[0.1, 0.2, 0.4, 0.8, 0.4, 0.2, 0.1]$ とする。

30

【数18】

$$G_s[f,\omega]=\frac{\sum_{j=-K}^K \alpha_g[j] \cdot G[f,\omega]}{\sum_{j=-K}^K \alpha_g[j]} \quad \dots(18)$$

以上のように、再生される入力信号と集音信号で、信号成分が存在する周波数帯域が異なっていたり、サンプリング周波数が異なっていたりしていても、集音信号の周波数特性であるパワースペクトルについて入力信号の周波数帯域を加味して帯域拡張して推定することで、集音信号の周波数特性が高精度に求められ、入力信号の明瞭度を向上させることができる。

40

【0077】

なお、図7(b)に示す音声帯域拡張通話装置に本変形例を適用する場合は、信号帯域拡張処理部12において事前に設定した周波数 $f\_limit$  ( $f\_limit$ は500~1200[Hz]程度で、例えば $f\_limit=1000$ [Hz]とする)以下の低い周波数帯域が拡張されるとき、つまり $f_{s\_low} < f_{s\_nb\_low}$ かつ $f_{s\_wb\_low} < f\_limit$ であるときは、信号特性補正部35で $f\_limit$ 以下の周波数帯域について信号補正処理をしないようにする。低域( $f\_limit$ 以下の周波数)においては、集音する環境やノイズ成分の種類によって、周囲雑音のバラツキが大きいため、このようにすることで、信号帯域拡張処理部12において拡張した低い周波数帯

50

域での周囲雑音のバラツキによって信号補正処理が不安定になることを防止できる。

【0078】

(信号処理部の変形例2)

本変形例では、図8に示す信号処理部30にて用いる狭帯域信号情報を周囲雑音のパワースペクトルとし、広帯域信号情報を周囲雑音の広帯域パワースペクトル(周囲雑音を広帯域の信号に拡張した場合のパワースペクトル)とした場合を例にして説明する。この場合、周囲雑音情報帯域拡張部34では、狭帯域信号情報である周囲雑音のパワースペクトルを入力として、狭帯域特徴量データとして周囲雑音の正規化されたパワースペクトルを算出し、広帯域特徴量データである周囲雑音の正規化された広帯域パワースペクトルを事前にモデル化された狭帯域特徴量データと広帯域特徴量データとの対応を用いて算出し、この広帯域特徴量データから広帯域信号情報である周囲雑音の広帯域パワースペクトルを生成するようにする。なお、狭帯域特徴量データと広帯域特徴量データとの対応のモデル化には、図5に示すGMMを利用する手法を用いる。これによれば、再生される入力信号と集音信号で、信号成分が存在する周波数帯域が異なっていたり、サンプリング周波数が異なっていたりしていても、集音信号の周波数特性であるパワースペクトルについて入力信号の周波数帯域を加味して帯域拡張して推定することで、集音信号の周波数特性が高精度に求まり、入力信号の明瞭度を向上させることができる。

10

【0079】

(信号処理部の変形例3)

次に、信号処理部にて用いる狭帯域信号情報は周囲雑音のパワースペクトル、広帯域信号情報は周囲雑音を広帯域の信号に拡張した場合のマスキング閾値(広帯域マスキング閾値)である場合を例にして説明する。

20

【0080】

図13は、その構成を示すものである。信号処理部300では、信号処理部30で用いていた周囲雑音情報帯域拡張部34に代わって、周囲雑音情報帯域拡張部36を用いた構成になっている。

【0081】

図14に周囲雑音情報帯域拡張部36の構成例を示す。周囲雑音情報帯域拡張部36は、パワー正規化部321と、狭帯域マスキング閾値算出部362と、帯域制御部363と、辞書格納部364と、広帯域マスキング閾値算出部365と、閾値補正部366と、パワー制御部345とを備える。

30

【0082】

周囲雑音情報帯域拡張部36は、周囲雑音情報帯域拡張部34と同様に、集音信号 $z[n]$ の周波数帯域成分における情報(狭帯域信号情報)を入力として、入力信号 $x[n]$ には存在して集音信号 $z[n]$ には存在しない周波数帯域成分を含めた情報(広帯域信号情報)を生成する。つまり、周囲雑音情報帯域拡張部36では、狭帯域信号情報から狭帯域特徴量データを算出し、狭帯域特徴量データと広帯域特徴量データとの対応を事前にモデル化しておき、このモデルと取得した狭帯域特徴量データとの対応を用いて広帯域特徴量データを算出し、広帯域特徴量データから広帯域信号情報を生成する。この際、周囲雑音情報帯域拡張部36では、狭帯域特徴量データと広帯域特徴量データとの対応のモデル化に、ベクトル量子化によるコードブックを利用する手法を用いる。ここでは、周囲雑音の帯域制御された狭帯域マスキング閾値 $N_{th}[f,w]$  ( $w=0,1,\dots,M_c-1$ )の $D_{nb}$ 次の狭帯域特徴量データとして用い、周囲雑音の広帯域マスキング閾値 $N_{wb\_th1}[f,w]$  ( $w=0,1,\dots,2M-1$ )の $D_{wb}$ 次の広帯域特徴量データとして用いる( $D_{nb}=M_c$ 、 $D_{wb}=2M$ )。具体的には、周囲雑音情報帯域拡張部36は、周囲雑音のパワースペクトル $|N[f,w]|^2$  ( $w=0,1,\dots,M-1$ )を入力として、周囲雑音のマスキング閾値を求め、このマスキング閾値を帯域制限し、帯域制限されたマスキング閾値について入力信号 $x[n]$ には存在して集音信号 $z[n]$ には存在しない周波数帯域成分を周波数帯域拡張して生成して、この帯域拡張されたマスキング閾値である広帯域マスキング閾値 $N_{wb\_th}[f,w]$  ( $w=0,1,\dots,2M-1$ )を出力する。

40

【0083】

50

狭帯域マスキング閾値算出部 3 6 2 は、パワー正規化部 3 2 1 から出力される周囲雑音の正規化したパワースペクトル  $|N_n[f, w]|^2$  ( $w=0, 1, \dots, M-1$ ) を入力として、周波数成分ごとに周囲雑音のマスキング閾値である狭帯域マスキング閾値  $N\_th1[f, w]$  ( $w=0, 1, \dots, M-1$ ) を算出する。前述した広帯域マスキング閾値算出部 3 4 4 と同様にして、データ長である  $2M$  を  $M$  で置き換え、周囲雑音の狭帯域マスキング閾値  $N\_th1[f, w]$  ( $w=0, 1, \dots, M-1$ ) は、spreading function を関数  $sprdnf()$  として、式 (19) の式で算出される。狭帯域マスキング閾値  $N\_th1[f, w]$  は、周囲雑音の正規化したパワースペクトル  $|N_n[f, w]|^2$  が狭帯域マスキング閾値  $N\_th1[f, w]$  以下であるならば、周波数ビン 以外の周波数帯域の周囲雑音の正規化したパワースペクトルによってマスクされることを示す。

【数 19】

10

$$N\_th1[f, \omega] = \sum_{j=0}^{M-1} |N_n[f, j]|^2 \cdot sprdnf(bark[j], bark[\omega]) \quad \dots (19)$$

帯域制御部 3 6 3 は、狭帯域マスキング閾値算出部 3 6 2 から出力された周囲雑音の狭帯域マスキング閾値  $N\_th1[f, w]$  ( $w=0, 1, \dots, M-1$ ) を入力として、帯域制御する下限周波数  $limit\_low$  [Hz] から帯域制御する上限周波数  $limit\_high$  [Hz] までの周波数帯域の信号情報のみを用いるように制御し、帯域制御された狭帯域マスキング閾値である  $N\_th[f, w]$  を出力する。ただし、 $fs\_nb\_low \quad limit\_low < limit\_high \quad fs\_nb\_high < fs/2$  とする。例えば  $limit\_low=1000$  [Hz]、 $limit\_high=3400$  [Hz] とするとき、これらの周波数帯域を式 (24) で周波数ビン に変換して考慮すると、狭帯域マスキング閾値  $N\_th1[f, w]$  ( $w=0, 1, \dots, M-1$ ) のうち  $w=32, 33, \dots, 108$  のみを用いるようにする。  $M_c$  を  $N\_th[f, w]$  の配列の個数として、帯域制御された狭帯域マスキング閾値  $N\_th[f, w]$  ( $w=0, 1, \dots, M_c-1$ ) は、狭帯域マスキング閾値  $N\_th1[f, w]$  ( $w=32, \dots, 108$ ) そのものを代入する。この場合は  $M_c=108-32+1=77$  である。

20

【0084】

図 1 1 にも示す通り、低域においては、集音する環境やノイズ成分の種類によって、周囲雑音のマスキング閾値の分散・バラツキが大きいことが分かる。周囲雑音の主要な成分はノイズ成分であるため、狭帯域マスキング閾値  $N\_th1[f, w]$  も低域においては、分散・バラツキが大きくなる。そこで、狭帯域特徴量データと広帯域特徴量データとの対応のモデル化をベクトル量子化によるコードブックを利用する手法を用いて、広帯域マスキング閾値を高精度に求めるために、分散・バラツキが大きい低域を用いないように帯域制御する。つまり、ここで、帯域制御する下限周波数  $limit\_low$  [Hz] は、狭帯域マスキング閾値の分散・バラツキが所定の値よりも小さいような周波数帯域の下限に設定することが望ましい。こうすることによって、広帯域マスキング閾値を高精度に求めることができ、入力信号の明瞭度を向上させることができる。

30

【0085】

また、マスキング閾値は、その周波数帯域のパワースペクトルだけではなくて周囲の周波数帯域のパワースペクトルを加味して算出される。そのため、マスキング閾値を求める元々の信号の帯域制限されている周波数帯域付近では、正確にマスキング閾値が算出できない。つまり、帯域制御する上限周波数  $limit\_high$  [Hz] は、帯域制限を加味してもマスキング閾値が正確に求まる周波数帯域の上限に設定することが望ましい。こうすることによって、広帯域マスキング閾値を高精度に求めることができ、入力信号の明瞭度を向上させることができる。

40

【0086】

辞書格納部 3 6 4 は、 $Dnb$  次の狭帯域特徴量データと  $Dwb$  次の広帯域特徴量データとの対応をモデル化して事前に学習されたサイズ  $Q$  (ここでは  $Q=64$ ) のコードブックの辞書  $3_q = \{\mu x_q, \mu y_q\}$  ( $q=1, \dots, Q$ ) を格納している。なお、 $\mu x_q$  は  $q$  番目のコードブックにおける狭帯域特徴量データのセントロイドベクトル、 $\mu y_q$  は  $q$  番目のコードブックにおける広帯域特徴量データのセントロイドベクトルを表している。なお、コードブックのコードベクトルの次数は、狭帯域信号情報のセントロイドベクトル  $\mu x_q$  と広帯域信号情報のセントロ

50

イドベクトル  $\mu_{y_q}$  の成分の和である  $D_{nb} + D_{wb}$  である。

【 0 0 8 7 】

辞書格納部 3 6 4 における事前の辞書  $3_q$  の学習生成方法の一手法について、フローチャートを図 1 5 に示し、説明する。以下の説明では、上述した変形例 1 における辞書  $2_q$  の学習生成方法と同じ処理については同じ番号を付番し、説明を簡明にするために必要に応じて重複する説明を省略する。

【 0 0 8 8 】

まず、広帯域信号データ  $wb[n]$  を入力として、サンプリング周波数  $f_s$  [Hz] にダウンサンプリングし狭帯域信号データ  $nb[n]$  を得る (ステップ S 1 0 1)。そして、狭帯域信号データ  $nb[n]$  から狭帯域信号情報を表す特徴量データである狭帯域特徴量データ  $P_{nb}[f, d]$  ( $d=1, \dots, D_{nb}$ ) を抽出する (ステップ S 2 0 2)。このステップ S 2 0 2 では、狭帯域信号データ  $nb[n]$  のパワースペクトル ( $M$  次) を得て (ステップ S 1 0 2 1)、狭帯域信号データ  $nb[n]$  のパワー値を得て (ステップ S 1 0 2 2)、これらのパワースペクトルとパワー値から狭帯域信号データ  $nb[n]$  の正規化されたパワースペクトルを得て (ステップ S 1 0 2 3)、式 (2 3) と同様にして狭帯域信号データ  $nb[n]$  のマスキング閾値を算出する (ステップ S 3 0 2 4)。そして、狭帯域信号データ  $nb[n]$  のマスキング閾値に対して、帯域制御部 3 6 3 での処理と同様に帯域制御する (ステップ S 3 0 2 5)。これを次数  $D_{nb}$  ( $=M_c$ ) の狭帯域特徴量データ  $P_{nb}[f, d]$  ( $d=1, \dots, D_{nb}$ ) とすることによって狭帯域特徴量データの抽出を行う。

【 0 0 8 9 】

一方、広帯域信号データ  $wb[n]$  から広帯域信号情報を表す特徴量データである広帯域特徴量データ  $P_{wb}[f, d]$  ( $d=1, \dots, D_{wb}$ ) を抽出する (ステップ S 3 0 3)。このステップ S 3 0 3 では、広帯域信号データ  $wb[n]$  のパワースペクトル ( $2M$  次) を得て (ステップ S 1 0 3 1)、広帯域信号データ  $wb[n]$  から広帯域信号データ  $wb[n]$  のパワー値を得て (ステップ S 2 0 3 2)、これらのパワースペクトルとパワー値から広帯域信号データ  $wb[n]$  の正規化されたパワースペクトルをフレーム単位で得て (ステップ S 2 0 3 3)、式 (2 3) の次数を  $M$  から  $2M$  にして同様にして広帯域信号データ  $wb[n]$  のマスキング閾値を算出する (ステップ S 3 0 3 4)。これを次数  $D_{wb}$  ( $=2M$ ) の広帯域特徴量データ  $P_{wb}[f, d]$  ( $d=1, \dots, D_{wb}$ ) とすることによって広帯域特徴量データの抽出を行う。

【 0 0 9 0 】

次に、狭帯域特徴量データ  $P_{nb}[f, d]$  ( $d=1, \dots, D_{nb}$ ) と広帯域特徴量データ  $P_{wb}[f, d]$  ( $d=1, \dots, D_{wb}$ ) を連結して、次数  $D_{nb} + D_{wb}$  の連結特徴量データ  $P[f, d]$  ( $d=1, \dots, D_{nb} + D_{wb}$ ) を生成する (ステップ S 1 0 4)。

【 0 0 9 1 】

そして、連結特徴量データ  $P[f, d]$  からコードブックの各コードベクトルにおける狭帯域セントロイドベクトル  $\mu_{x_q}$  と広帯域セントロイドベクトル  $\mu_{y_q}$  を求め、サイズ  $Q$  (ここでは  $Q=64$ ) のコードブックを  $k$ - $m$ - $e$ - $a$ - $n$ - $s$  アルゴリズムや  $L$ - $B$ - $G$  アルゴリズムなどによるクラスタリング手法を用いて生成する (ステップ S 2 0 5)。コードブックの各コードベクトルにおける広帯域セントロイドベクトル  $\mu_{y_q}$  である広帯域信号データ  $wb[n]$  のマスキング閾値を近似多項式係数で表現して、近似多項式係数を広帯域セントロイドベクトル  $\mu'_{y_q}$  として辞書に格納して、辞書  $3_q = \{ \mu_{x_q}, \mu'_{y_q} \}$  ( $q=1, \dots, Q$ ) を生成する (ステップ S 3 0 7)。近似多項式係数  $m_p$  ( $p=0, \dots, P$ ) とはここでは、縦軸をパワー値  $X$  [dB]、横軸を周波数  $Y$  [Hz] として、式 (2 0) のようにマスキング閾値を所定の次数 (ここでは  $P$  とし、例えば  $P=6$  とする) の多項式で近似した、その多項式の係数のことであり、これ以降そのように呼ぶ。

【 数 2 0 】

$$Y \cong \sum_{i=0}^P m_i \cdot X^i = m_0 + m_1 X + m_2 X^2 + \dots + m_P X^P \quad \dots (2 0)$$

このように、マスキング閾値を近似多項式係数で表現して辞書として格納しておくこと

で、マスキング閾値を辞書として格納しておくよりも、辞書の格納に掛かるメモリ量を削減することができ、辞書の配列の数を小さくなるため辞書の利用時の処理量を削減することができる。

#### 【0092】

広帯域マスキング閾値算出部365は、帯域制御部363から出力された帯域制御された狭帯域マスキング閾値 $N_{th}[f,w]$  ( $w=0,1,\dots,M_C-1$ )を $D_{nb}$ 次の特徴量データとして入力し、辞書格納部364からコードブックの辞書 $3_q=\{\mu x_q, \mu' y_q\}$  ( $q=1,\dots,Q$ )を読み出して、 $D_{nb}$ 次の狭帯域特徴量データと $D_{wb}$ 次の広帯域特徴量データとの対応から周囲雑音の広帯域マスキング閾値 $N_{wb\_th1}[f,w]$  ( $w=0,1,\dots,2M-1$ )を求める。具体的には、 $Q$ 個ある狭帯域セントロイドベクトル $\mu x_q$  ( $q=1,\dots,Q$ )から、帯域制御された狭帯域マスキング閾値 $N_{th}[f,w]$  ( $w=0,1,\dots,M_C-1$ )と所定の距離尺度で一番距離が近いものを求めて、一番距離が近いコードベクトルにおける広帯域セントロイドベクトル $\mu' y_q$ をそのまま広帯域マスキング閾値の近似多項式係数として設定し、式(20)と同様にして広帯域マスキング閾値 $N_{wb\_th1}[f,w]$  ( $w=0,1,\dots,2M-1$ )を算出する。

#### 【0093】

閾値補正部366は、狭帯域マスキング閾値算出部362から出力された周囲雑音の狭帯域マスキング閾値 $N_{th1}[f,w]$  ( $w=0,1,\dots,M-1$ )と広帯域マスキング閾値算出部365から出力された周囲雑音の広帯域マスキング閾値 $N_{wb\_th1}[f,w]$  ( $w=0,1,\dots,2M-1$ )を入力として、狭帯域と広帯域における境界帯域付近での不連続性あるいは微分不連続性を解消するように補正し、その補正された広帯域マスキング閾値 $N_{wb\_th2}[f,w]$  ( $w=0,1,\dots,2M-1$ )を出力する。図16(a)に、境界帯域 $f_s/2$ [Hz]前後の周波数において、狭帯域マスキング閾値 $N_{th}[f,w]$ と広帯域マスキング閾値 $N_{wb\_th1}[f,w]$ とに不連続性が生じ、それを解消するように補正された広帯域マスキング閾値 $N_{wb\_th2}[f,w]$ の例を示す。図16(b)に、境界帯域 $f_s/2$ [Hz]前後の周波数において、狭帯域マスキング閾値 $N_{th}[f,w]$ と広帯域マスキング閾値 $N_{wb\_th1}[f,w]$ とに不連続性と微分不連続性の両方が生じ、それを解消するように補正された広帯域マスキング閾値 $N_{wb\_th2}[f,w]$ の例を示す。両図共に、実線は狭帯域マスキング閾値 $N_{th}[f,w]$ を、破線は広帯域マスキング閾値 $N_{wb\_th2}[f,w]$ を、太実線は補正された広帯域マスキング閾値 $N_{wb\_th2}[f,w]$ における補正箇所を表す。ただし、 $adjust\_low$ [Hz] <  $f_s/2$  <  $adjust\_high$ [Hz]とする。ここで、 $adjust\_low$ は周波数ビン $L-1$ に対応する周波数以上で周波数ビン $L$ に対応する周波数未満であり、 $adjust\_high$ は周波数ビン $H$ に対応する周波数以上で周波数ビン $H+1$ に対応する周波数未満であるとする。例えば $f_s=8000$ [Hz]であるとき、 $adjust\_low=3600$ [Hz]、 $adjust\_high=4400$ [Hz]とする。具体的には、少なくとも境界帯域 $f_s/2$ [Hz]前後の周波数において不連続あるいは微分不連続が検出された場合に、 $adjust\_low$ [Hz]以上かつ $adjust\_high$ [Hz]以下であるような境界帯域付近について、周波数ビン $L, L+1, \dots, L+S$ と $H, H-1, \dots, H-S$ における広帯域マスキング閾値 $N_{wb\_th1}[f,w]$ を用いて、周波数ビン $L+S+1$ から $H-S-1$ までの広帯域マスキング閾値を $(2S-1)$ 次関数で模擬し、スプライン補間を行うことで、補正された広帯域マスキング閾値 $N_{wb\_th2}[f,w]$ を求める。ここで、狭帯域マスキング閾値 $N_{th1}[f,M-1]$ と広帯域マスキング閾値 $N_{wb\_th1}[f,M]$ との中点を通過するように模擬する関数を設定してスプライン補間を行ってもよい。

#### 【0094】

このように閾値補正部366において広帯域マスキング閾値を補正することで、広帯域マスキング閾値における不連続性あるいは微分不連続性が解消され、信号補正においても周波数方向の不連続性が無くなって違和感のない自然な信号補正にすることができ、高い明瞭感を得ることができる。

#### 【0095】

以上のように、再生される入力信号と集音信号で、信号成分が存在する周波数帯域が異なっていたり、サンプリング周波数が異なっていたりしていても、集音信号のマスキング閾値について入力信号の周波数帯域を加味して帯域拡張して推定することで、集音信号のマスキング閾値が高精度に求まり、入力信号の明瞭度を向上させることができる。

## 【 0 0 9 6 】

( 信号処理部の変形例 4 )

信号処理部 3 0 0 の辞書格納部 3 6 4 における事前の辞書  $3_q$  の学習生成方法の他の手法について、フローチャートを図 1 7 に示し、説明する。ここでは、狭帯域信号データ  $nb[n]$  を生成しないで広帯域信号データ  $wb[n]$  のみから辞書  $3_q$  を学習生成する方法について説明する。以下の説明では、上述した変形例 2 における辞書  $3_q$  の学習生成方法と同じ処理については同じ番号を付番し、説明を簡明にするために必要に応じて重複する説明を省略する。

## 【 0 0 9 7 】

まず、ステップ S 3 0 3 で広帯域信号データ  $wb[n]$  から広帯域信号情報を表す特徴量データ (ここではマスキング閾値) である広帯域特徴量データ  $Pwb[f, d]$  ( $d=1, \dots, Dwb$ ) を抽出する。この広帯域特徴量データ  $Pwb[f, d]$  ( $d=1, \dots, Dwb$ ) のみを用いて、ステップ S 2 0 5 でサイズ  $Q$  のコードブックを作成する。そして、コードブックの各コードベクトルにおける広帯域セントロイドベクトル  $\mu y_q$  である広帯域信号データ  $wb[n]$  の広帯域マスキング閾値に対して、帯域制御する下限周波数  $limit\_low[Hz]$  から帯域制御する上限周波数  $limit\_high[Hz]$  までの周波数帯域の広帯域マスキング閾値のみを用いるように制御する (ステップ S 3 0 2 5)。これにより狭帯域に帯域制御された狭帯域マスキング閾値が求まり、これをコードブックの各コードベクトルにおける狭帯域セントロイドベクトル  $\mu x_q$  ( $q=1, \dots, Q$ ) とする (ステップ S 3 0 6)。その後、ステップ S 3 0 7 で広帯域信号データ  $wb[n]$  のマスキング閾値の近似多項式係数である広帯域セントロイドベクトル  $\mu' y_q$  と併せて辞書に格納して、辞書  $3_q = \{ \mu x_q, \mu' y_q \}$  を生成する。

## 【 0 0 9 8 】

狭帯域特徴量データを併用してクラスタリングする図 1 5 における手法では、狭帯域特徴量データに狭帯域と広帯域における境界帯域付近で誤差を含む。このように、広帯域特徴量データのみを用いてクラスタリングして、広帯域セントロイドベクトルを帯域制限して狭帯域セントロイドベクトルを求めることで、理想的なデータである広帯域特徴量データのみを用いてクラスタリングするため、図 1 5 における手法よりも、高精度にクラスタリングを行うことができる。

## 【 0 0 9 9 】

( 信号処理部の変形例 5 )

信号処理部 3 0 0 の辞書格納部 3 6 4 における事前の辞書  $3_q$  の学習生成方法の他の手法について、フローチャートを図 1 8 に示し、説明する。以下の説明では、上述した変形例 2 における辞書  $3_q$  の学習生成方法と同じ処理については同じ番号を付番し、説明を簡明にするために必要に応じて重複する説明を省略する。

## 【 0 1 0 0 】

ステップ S 2 0 5 でサイズ  $Q$  のコードブックを作成した後、コードブックの各コードベクトルにおける狭帯域セントロイドベクトル  $\mu x_q$  である狭帯域信号データ  $nb[n]$  のマスキング閾値を式 ( 2 0 ) のように近似多項式で表現して、近似多項式係数を狭帯域セントロイドベクトル  $\mu' x_q$  ( $q=1, \dots, Q$ ) とする (ステップ S 3 0 6 A)。その後、ステップ S 3 0 7 で広帯域信号データ  $wb[n]$  のマスキング閾値の近似多項式係数である広帯域セントロイドベクトル  $\mu' y_q$  と併せて辞書に格納して、辞書  $3_q = \{ \mu' x_q, \mu' y_q \}$  を生成する。

## 【 0 1 0 1 】

一方で、この手法においては、広帯域マスキング閾値算出部 3 6 5 では、帯域制御部 3 6 3 から出力された帯域制御された狭帯域マスキング閾値  $N\_th[f, w]$  ( $w=0, 1, \dots, M_C-1$ ) を  $Dnb$  次の特徴量データとして入力し、辞書格納部 3 6 4 からコードブックの辞書  $3_q = \{ \mu' x_q, \mu' y_q \}$  ( $q=1, \dots, Q$ ) を読み出して、 $Dnb$  次の狭帯域特徴量データと  $Dwb$  次の広帯域特徴量データとの対応から周雑音の広帯域マスキング閾値  $N\_wb\_th1[f, w]$  ( $w=0, 1, \dots, 2M-1$ ) を求めるようにする。具体的には、 $Q$  個ある狭帯域セントロイドベクトル  $\mu' x_q$  ( $q=1, \dots, Q$ ) の近似多項式から、帯域制御された狭帯域マスキング閾値  $N\_th[f, w]$  ( $w=0, 1, \dots, M_C-1$ ) と所定の距離尺度で一番距離が近いものを近似多項式に代入していくことで求めて、一番距

10

20

30

40

50

離が近いコードベクトルにおける広帯域セントロイドベクトル  $\mu' y_q$  をそのまま広帯域マスクング閾値の近似多項式係数として設定し、式(20)と同様にして広帯域マスクング閾値  $N_{wb\_th1}[f, w]$  ( $w=0, 1, \dots, 2M-1$ ) を算出する。

【0102】

このように、狭帯域マスクング閾値も近似多項式係数で表現して辞書として格納しておくことで、マスクング閾値を辞書として格納しておくよりも、図15における手法と比較しても、辞書の格納に掛かるメモリ量を削減することができ、辞書の配列の数を小さくなるため辞書の利用時の処理量を削減することができる。

【0103】

(第2の実施例)

図19(a)は、本発明の第2の実施形態に係わる通信装置の構成を示すものである。

【0104】

この図に示す通信装置は、例えば携帯電話などの無線通信装置の受信系を示すものであって、無線通信部1と、デコーダ2と、信号処理部3Aと、デジタル・アナログ(D/A)変換器4と、スピーカ5と、マイク6と、アナログ・デジタル(A/D)変換器7と、ダウンサンプリング部8と、エコー抑圧処理部9と、エンコーダ10とを備えている。

【0105】

なお、第1の実施例と同様に、本発明は、図19(a)のような通信装置だけでなく、図19(b)に示すデジタルオーディオプレーヤに適用することも可能である。また、図19(c)に示す音声帯域拡張通話装置に適用することも可能である。

【0106】

次に、信号処理部3Aについて説明する。図20は、その構成を示すものである。信号処理部3Aは、第1の実施例にて説明した信号処理部3に周囲雑音抑圧処理部37を追加して構成される。以下の説明では、上述した実施例と同じ構成については同じ番号を付番し、必要に応じて重複する説明を省略する。

【0107】

図21に周囲雑音抑圧処理部37の構成例を示す。周囲雑音抑圧処理部37は、抑圧ゲイン算出部371と、スペクトル抑圧部372と、パワー算出部373と、時間領域変換部374とを備える。

【0108】

周囲雑音抑圧処理部37は、周囲雑音推定部31から出力される周囲雑音のパワースペクトルと集音信号  $z[n]$  のパワースペクトルと集音信号  $z[n]$  の周波数スペクトルを用いて、集音信号  $z[n]$  に含まれる周囲雑音であるノイズ成分を抑圧して、周囲雑音であるノイズ成分が抑圧された信号  $s[n]$  をエンコーダ10に出力する。エンコーダ10では周囲雑音抑圧処理部37から出力された信号  $s[n]$  を符号化して無線通信部1に出力する。

【0109】

抑圧ゲイン算出部371は、パワー算出部312から出力される集音信号  $z[n]$  のパワースペクトル  $|Z[f, w]|^2$  ( $w=0, 1, \dots, M-1$ ) と、周波数スペクトル更新部314から出力される周囲雑音のパワースペクトル  $|N[f, w]|^2$  ( $w=0, 1, \dots, M-1$ ) と、パワー算出部373から出力される1フレーム前の抑圧処理された信号のパワースペクトル  $|S[f-1, w]|^2$  ( $w=0, 1, \dots, M-1$ ) とを用いて、各周波数帯域の抑圧ゲイン  $G[f, w]$  ( $w=0, 1, \dots, M-1$ ) を出力する。例えば、抑圧ゲイン  $G[f, w]$  の算出は、以下のアルゴリズムまたはそれらの組み合わせによって行う。すなわち、一般のノイズキャンセラであるスペクトル・サブトラクション (Spectral Subtraction) 法 (S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction", IEEE Trans. Acoustics, Speech, and Signal Processing, vol. AS-SP-29, pp.113-120 (1979).)、ウィナー・フィルター (Wiener Filter) 法 (J. S. Lim, A. V. Oppenheim, "Enhancement and bandwidth compression of noisy speech", Proc. IEEE Vol.67, No.12, pp.1586-1604, Dec.1979.) 及び最尤推定 (Maximum Likelihood) 法 (R. J. McAulay, M. L. Malpass, "Speech enhancement using a soft-decision noise suppression filter", IEEE Trans. on Acoustics, Speech, and Signal Processi

10

20

30

40

50

ng, vol. ASSP-28, no.2, pp.137-145, Apr.1980.) などである。ここでは一例としてウィナー・フィルター法を用いて、抑圧ゲイン $G[f,w]$ を算出するとする。

【0110】

スペクトル抑圧部372は、周波数領域変換部311から出力された集音信号 $z[n]$ の周波数スペクトル $Z[f,w]$ と、抑圧ゲイン算出部371から出力された抑圧ゲイン $G[f,w]$ とを入力として、集音信号 $z[n]$ の周波数スペクトル $Z[f,w]$ を集音信号 $z[n]$ の振幅スペクトル $|Z[f,w]|$  ( $w=0,1,\dots,M-1$ )と位相スペクトル $\angle z[f,w]$  ( $w=0,1,\dots,M-1$ )に分け、集音信号 $z[n]$ の振幅スペクトル $|Z[f,w]|$ に抑圧ゲイン $G[f,w]$ を乗じることで周囲雑音であるノイズ成分を抑圧し、その抑圧処理された信号の振幅スペクトル $|S[f-1,w]|$ とし、位相スペクトル $\angle z[f,w]$ をそのまま抑圧処理された信号の位相スペクトル $\angle s[f,w]$ として、抑圧処理された信号の周波数スペクトル $S[f,w]$  ( $w=0,1,\dots,M-1$ )を算出する。

10

【0111】

パワー算出部373は、スペクトル抑圧部372から出力された抑圧処理された信号の周波数スペクトル $S[f,w]$  ( $w=0,1,\dots,M-1$ )から抑圧処理された信号のパワースペクトル $|S[f,w]|^2$  ( $w=0,1,\dots,M-1$ )を算出し出力する。

【0112】

時間領域変換部374は、スペクトル抑圧部372から出力された抑圧処理された信号の周波数スペクトル $S[f,w]$  ( $w=0,1,\dots,M-1$ )を入力として、周波数領域を時間領域に変換する処理(例えば、IFFT)を施し、周波数領域変換部311における窓掛けによるオーバーラップ分を考慮して1フレーム前の抑圧処理された信号 $s[n]$ を適宜加算して、抑圧処理された時間領域の信号 $s[n]$  ( $n=0,1,\dots,N-1$ )を算出する。

20

【0113】

以上のように、周囲雑音推定処理に周囲雑音抑圧処理を併用することで処理量の増加を抑えつつ、入力信号を明瞭化すると同時に、集音信号における周囲雑音成分を抑圧して高音質な集音信号を得ることができる。

【0114】

なお、この発明は上記実施形態そのままに限定されるものではなく、実施段階ではその要旨を逸脱しない範囲で構成要素を変形して具体化できる。また上記実施形態に開示されている複数の構成要素を適宜組み合わせることによって種々の発明を形成できる。また例えば、実施形態に示される全構成要素からいくつかの構成要素を削除した構成も考えられる。さらに、異なる実施形態に記載した構成要素を適宜組み合わせてもよい。

30

【0115】

例えば、入力信号(あるいは目的信号)のサンプリング周波数は、集音信号(あるいは周囲雑音)のサンプリング周波数の2倍に限定されるものではなく、整数倍でも、非整数倍でもよい。また、入力信号(あるいは目的信号)のサンプリング周波数は、集音信号(あるいは周囲雑音)のサンプリング周波数と等しい上で、入力信号(あるいは目的信号)の周波数帯域制限の範囲と集音信号(あるいは周囲雑音)の周波数帯域制限の範囲が異なっている場合であっても構わない。さらに、入力信号(あるいは目的信号)の周波数帯域制限の範囲が、集音信号(あるいは周囲雑音)の周波数帯域制限の範囲を包含していなくても構わない。さらにまた、入力信号(あるいは目的信号)の周波数帯域制限の範囲は、集音信号(あるいは周囲雑音)の周波数帯域制限の範囲と隣接していなくても構わない。

40

【0116】

また、入力信号がモノラル信号ではなくステレオ信号であったとしても、例えばL(左)チャンネルとR(右)チャンネルにそれぞれ上記信号処理部3における信号処理を施したり、和信号(LチャンネルとRチャンネルの信号の和)と差信号(LチャンネルからRチャンネルの信号の差)にそれぞれ上記の信号処理を施したりすることで同様の効果が得られる。勿論、マルチチャンネル信号であったとしても例えば同様にそれぞれのチャンネル信号に対して上記の信号処理を施したりすることで同様の効果が得られる。

【0117】

その他、この発明の要旨を逸脱しない範囲で種々の変形を施しても同様に実施可能であ

50

ることはいうまでもない。

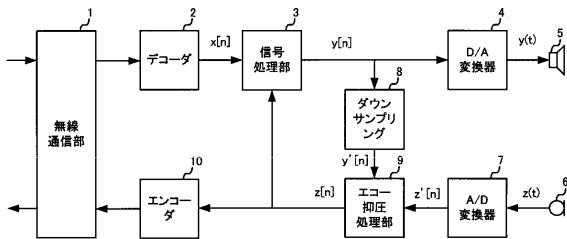
【符号の説明】

【0118】

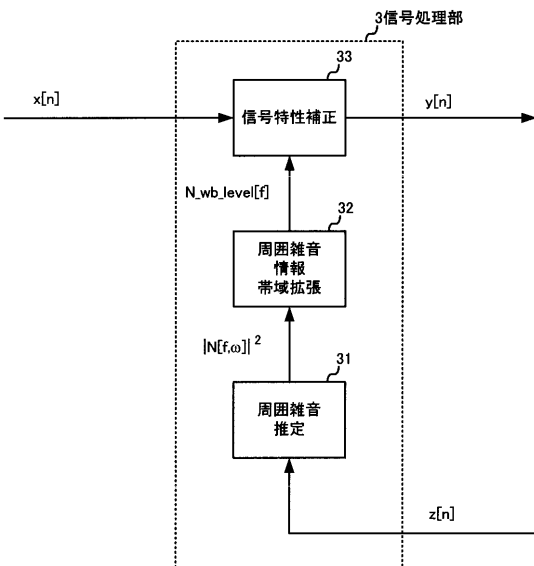
1...無線通信部、2, 2A...デコーダ、3, 30, 300, 3A...信号処理部、4...デジタル・アナログ(D/A)変換器、5...スピーカ、6...マイク、7...アナログ・デジタル(A/D)変換器、8...ダウンサンプリング部、9...エコー抑圧処理部、10...エンコーダ、11...記憶部、12...信号帯域拡張処理部、31...周囲雑音推定部、32, 34, 36...周囲雑音情報帯域拡張部、33, 35...信号特性補正部、37...周囲雑音抑圧処理部、311, 331...周波数領域変換部、312, 352, 373...パワー算出部、313...周囲雑音区間判定部、314...周波数スペクトル更新部、321...パワー正規化部、322, 342, 364...辞書格納部、323...広帯域パワー算出部、332, 356...補正度合決定部、333...補正処理部、334, 374...時間領域変換部、343...広帯域パワースペクトル算出部、344, 365...広帯域マスキング閾値算出部、345...パワー制御部、353...マスキング閾値算出部、354...マスキング判定部、355...パワー平滑化部、362...狭帯域マスキング閾値算出部、363...帯域制御部、366...閾値補正部、371...抑圧ゲイン算出部、372...スペクトル抑圧部。

10

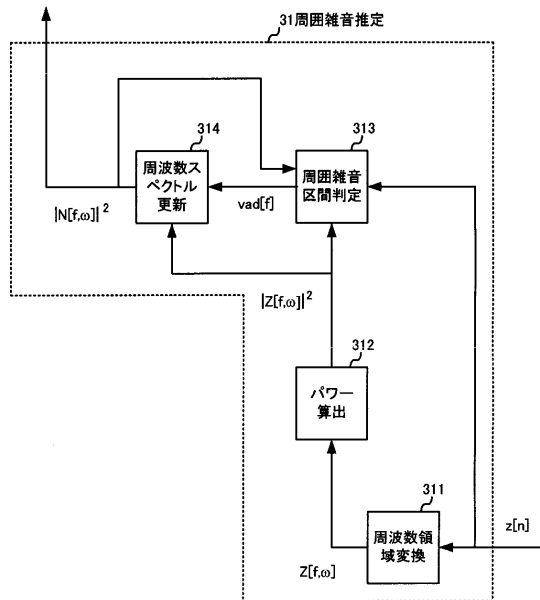
【図1】



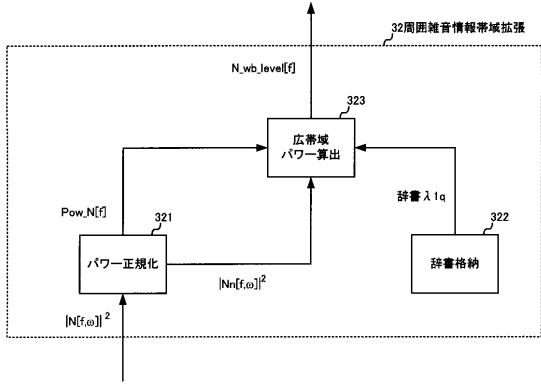
【図2】



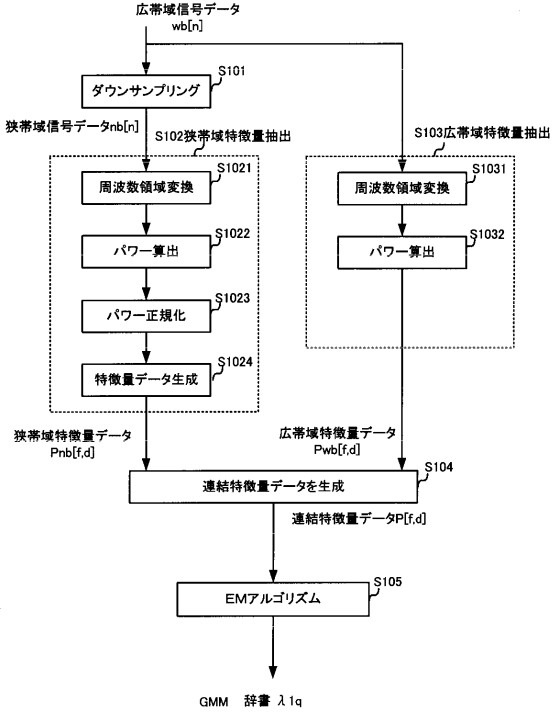
【図3】



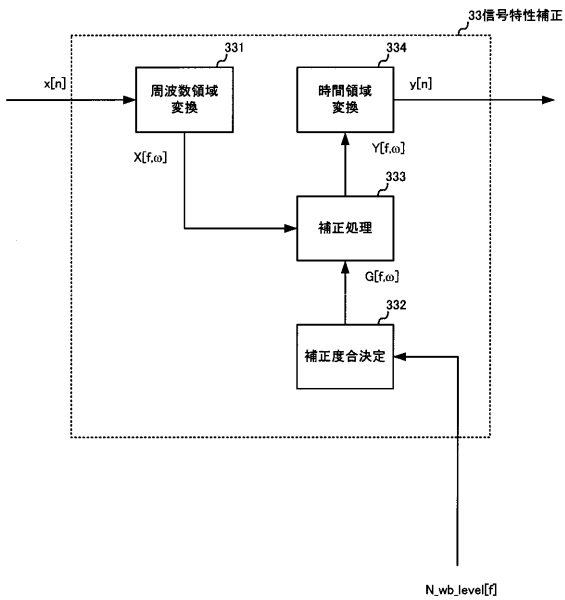
【 図 4 】



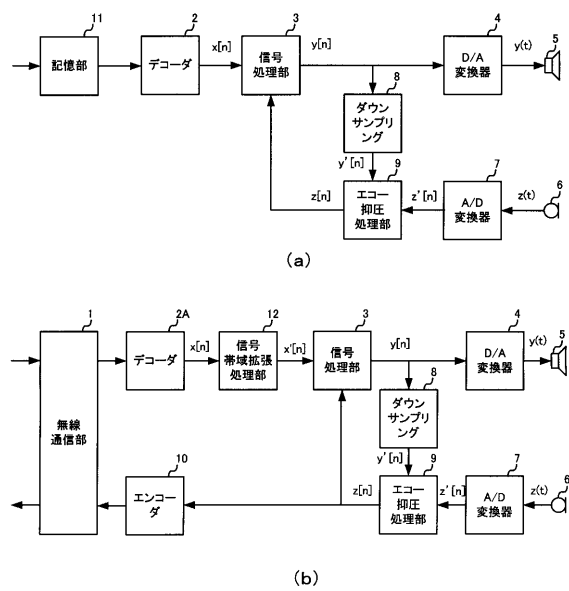
【 図 5 】



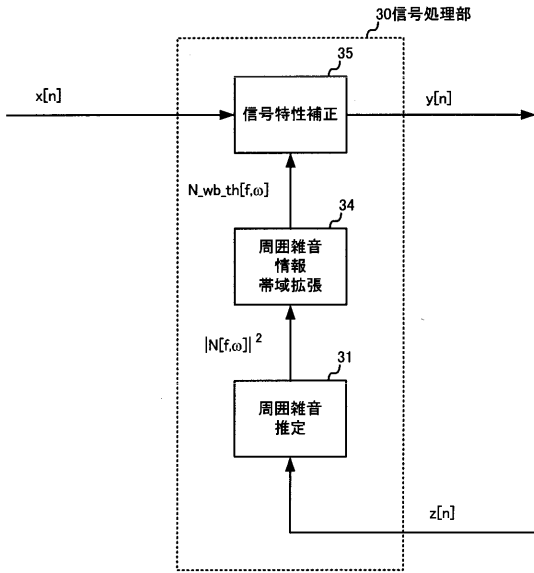
【 図 6 】



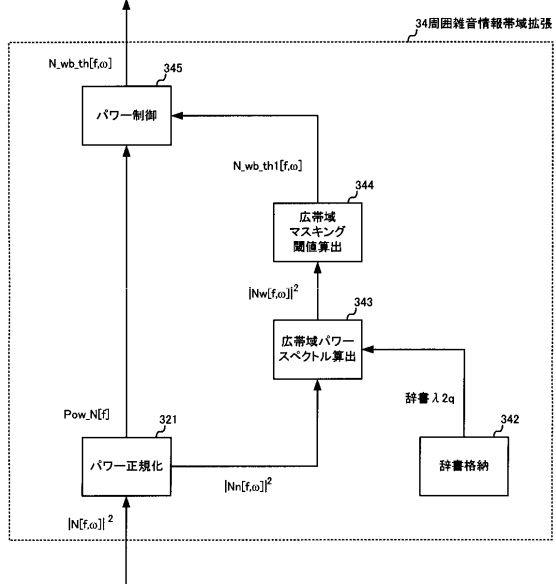
【 図 7 】



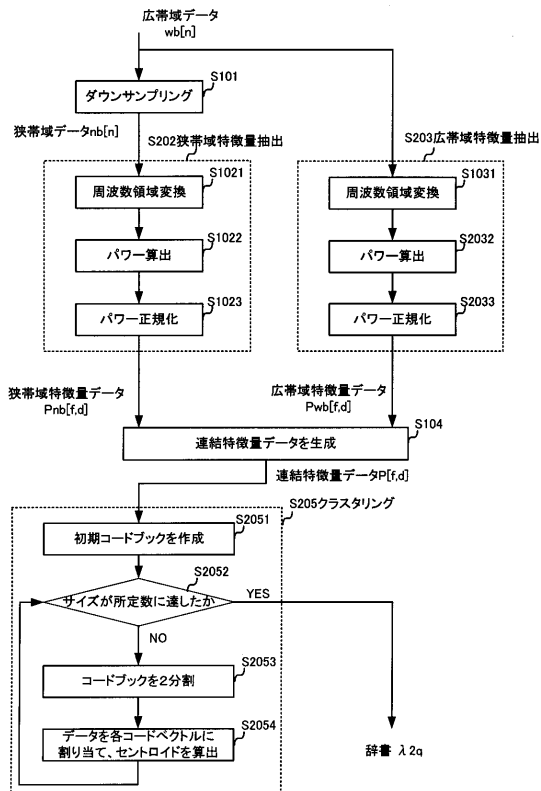
【 図 8 】



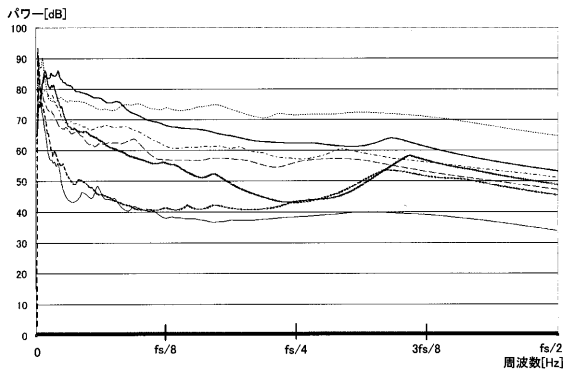
【 図 9 】



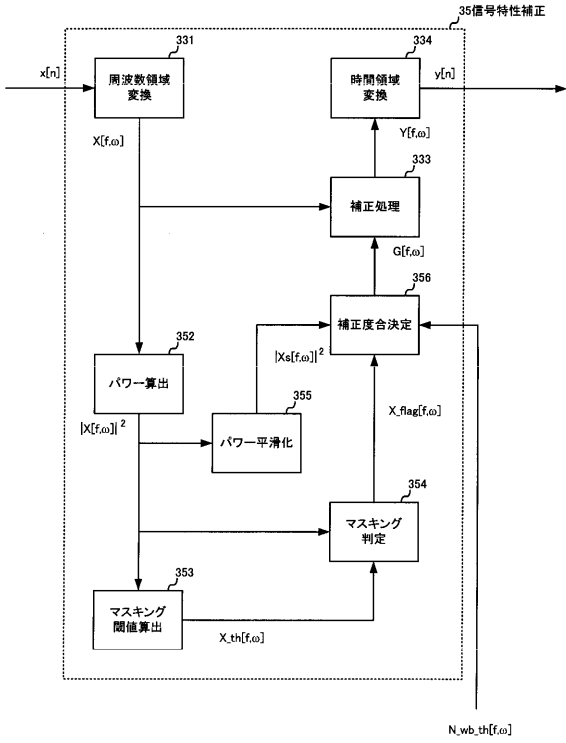
【 図 10 】



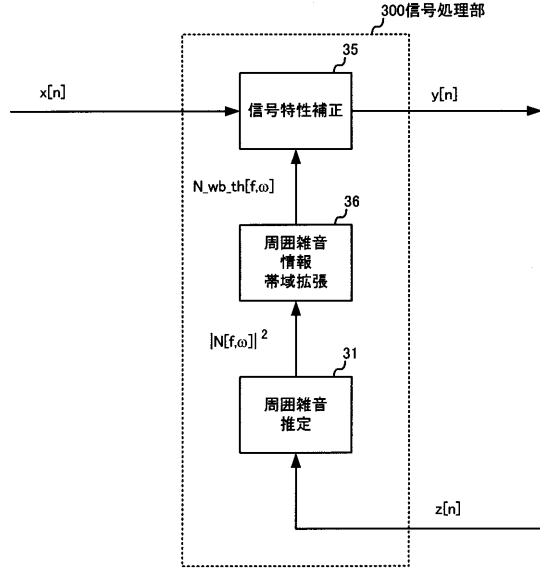
【 図 11 】



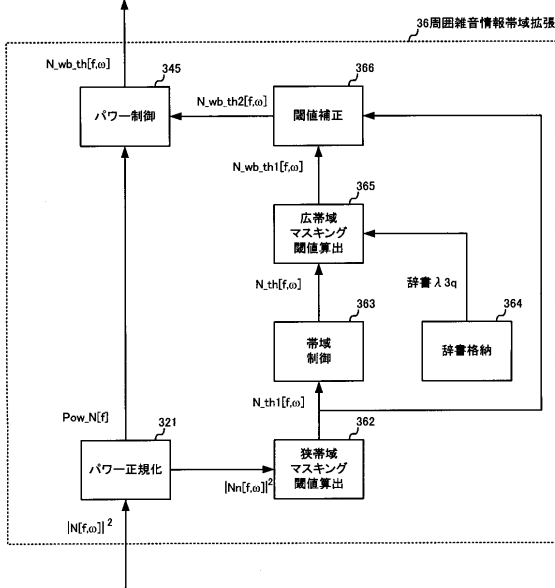
【図12】



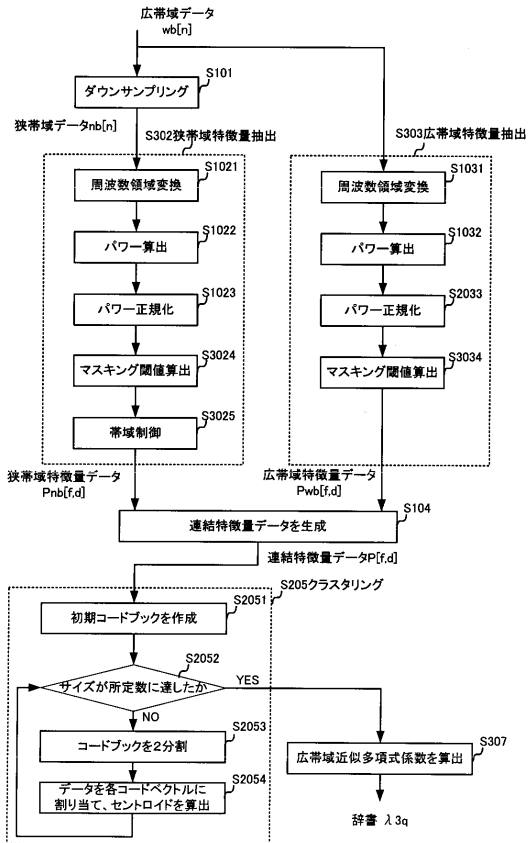
【図13】



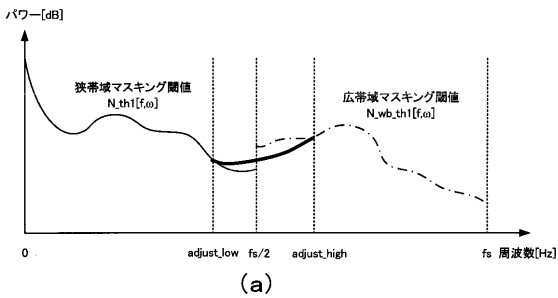
【図14】



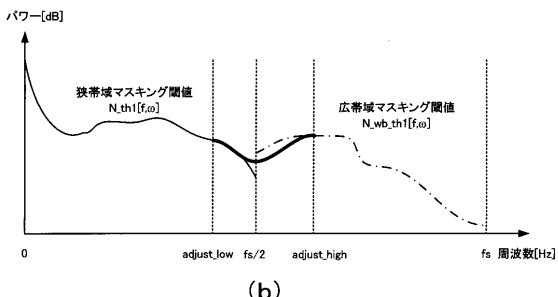
【図15】



【図16】

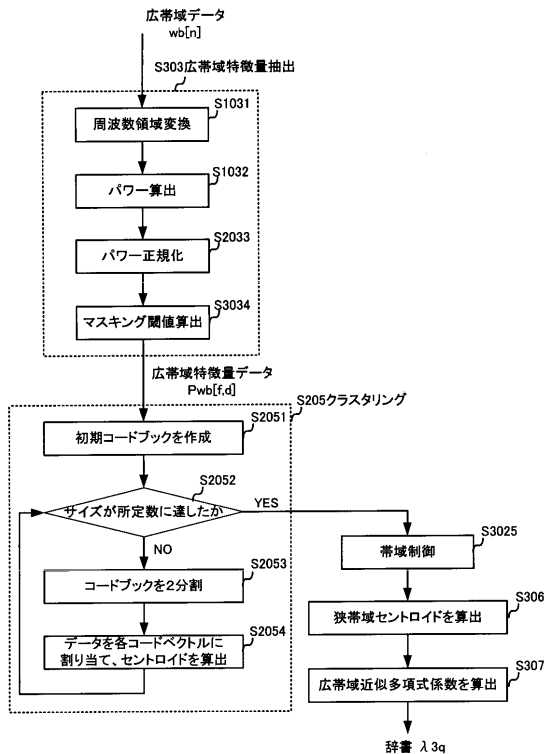


(a)

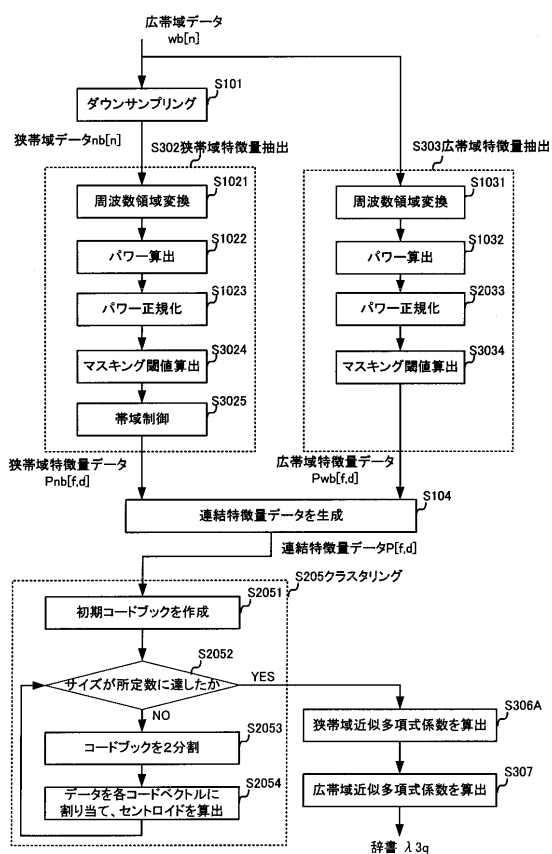


(b)

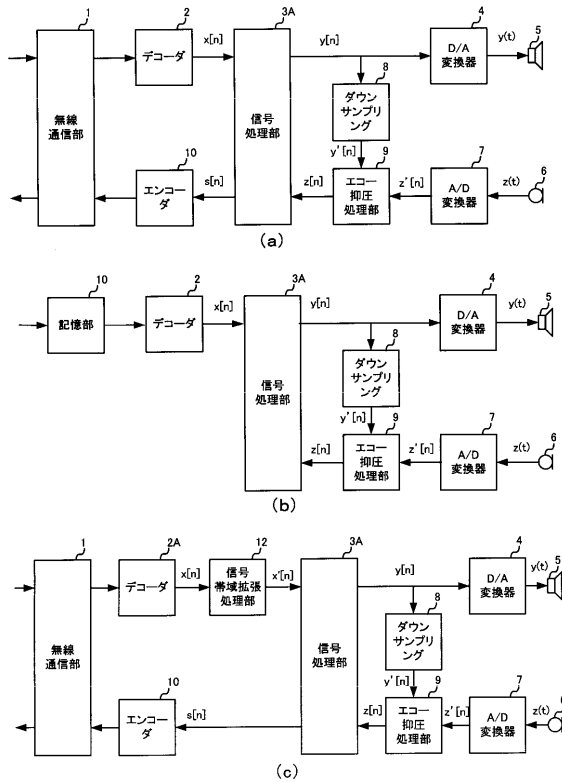
【図17】



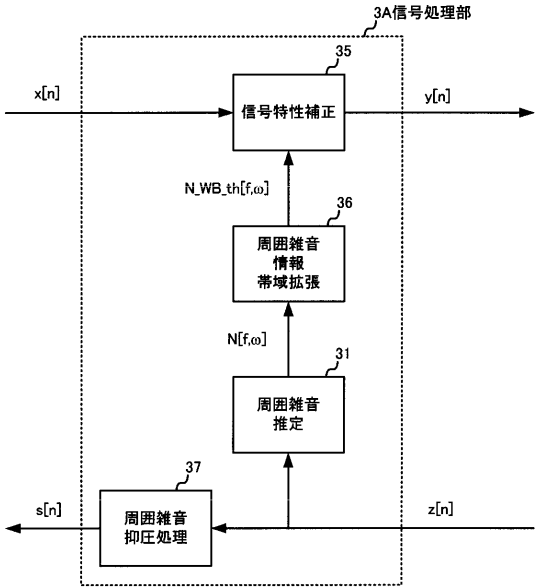
【図18】



【図19】



【 図 2 0 】



【 図 2 1 】

