

(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2014-122939

(P2014-122939A)

(43) 公開日 平成26年7月3日(2014.7.3)

(51) Int.Cl.	F I	テーマコード (参考)
G 1 O L 15/10 (2006.01)	G 1 O L 15/10 4 O O R	
G 1 O L 15/20 (2006.01)	G 1 O L 15/20 3 7 O D	
G 1 O L 21/0232 (2013.01)	G 1 O L 21/02 1 O 1 B	
G 1 O L 25/90 (2013.01)	G 1 O L 11/04	

審査請求 未請求 請求項の数 19 O L (全 28 頁)

(21) 出願番号	特願2012-277662 (P2012-277662)	(71) 出願人	000002185 ソニー株式会社 東京都港区港南1丁目7番1号
(22) 出願日	平成24年12月20日 (2012.12.20)	(74) 代理人	100082131 弁理士 稲本 義雄
		(74) 代理人	100121131 弁理士 西川 孝
		(72) 発明者	東山 恵祐 東京都港区港南1丁目7番1号 ソニー株式会社社内

(54) 【発明の名称】 音声処理装置および方法、並びにプログラム

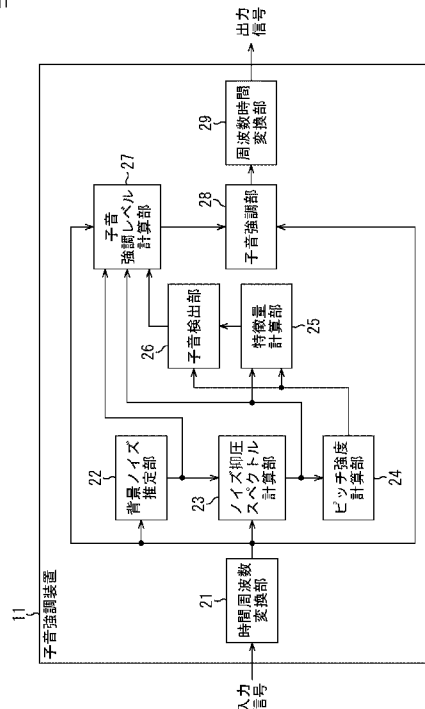
(57) 【要約】

【課題】より精度よく子音を検出することができるようにする。

【解決手段】時間周波数変換部は、供給された入力信号を入力スペクトルに変換し、背景ノイズ推定部は、入力スペクトルから背景ノイズを推定する。ノイズ抑圧スペクトル計算部は、背景ノイズの推定結果に基づいて、入力スペクトルの背景ノイズを抑圧し、ノイズ抑圧スペクトルを算出する。特徴量計算部は、ノイズ抑圧スペクトルから特徴量を算出し、子音検出部は、特徴量に基づいて入力信号から子音フレームを検出する。このように、背景ノイズが抑圧されたノイズ抑圧スペクトルから特徴量を算出し、その特徴量に基づいて周波数領域で子音を検出することで、精度よく子音を検出することができる。本技術は、子音強調装置に適用することができる。

【選択図】 図1

図1



## 【特許請求の範囲】

## 【請求項 1】

入力信号の背景ノイズを推定する背景ノイズ推定部と、  
前記背景ノイズの推定結果に基づいて、前記入力信号の前記背景ノイズを抑圧するノイズ抑圧部と、  
前記背景ノイズが抑圧された前記入力信号に基づいて特徴量を算出する特徴量計算部と、  
前記特徴量に基づいて、前記入力信号から子音を検出する子音検出部とを備える音声処理装置。

## 【請求項 2】

前記背景ノイズ推定部は、周波数領域で前記背景ノイズを推定し、  
前記ノイズ抑圧部は、前記入力信号から得られた入力スペクトルに含まれる前記背景ノイズを抑圧することでノイズ抑圧スペクトルを求め、  
前記特徴量計算部は、前記ノイズ抑圧スペクトルに基づいて前記特徴量を算出する請求項 1 に記載の音声処理装置。

## 【請求項 3】

前記背景ノイズ推定部は、過去の前記入力スペクトルの平均値を求めることで前記背景ノイズを推定する請求項 2 に記載の音声処理装置。

## 【請求項 4】

前記ノイズ抑圧スペクトルの値が、前記背景ノイズの推定により得られた背景ノイズスペクトルを定数倍した値よりも大きい周波数について、前記入力スペクトルの強調を行なう子音強調部をさらに備える請求項 3 に記載の音声処理装置。

## 【請求項 5】

前記子音強調部は、予め定められた強調量で前記入力スペクトルの強調を行なう請求項 4 に記載の音声処理装置。

## 【請求項 6】

前記入力信号の過去の母音部分のパワーの平均値と、前記入力信号の現在のパワーとの比に基づいて強調量を算出する子音強調レベル計算部をさらに備え、  
前記子音強調部は、前記強調量で前記入力スペクトルの強調を行なう請求項 4 に記載の音声処理装置。

## 【請求項 7】

周波数方向に対して前記強調量の補間が行なわれる請求項 6 に記載の音声処理装置。

## 【請求項 8】

前記ノイズ抑圧部は、スペクトルサブトラクション法により前記ノイズ抑圧スペクトルを求める請求項 2 に記載の音声処理装置。

## 【請求項 9】

前記入力信号のピッチ強度がさらに前記特徴量として用いられ、  
前記子音検出部は、前記特徴量としての前記ピッチ強度と、前記ノイズ抑圧スペクトルに基づいて算出された前記特徴量とに基づいて、前記入力信号から子音を検出する請求項 2 に記載の音声処理装置。

## 【請求項 10】

前記ピッチ強度は、前記ノイズ抑圧スペクトルのピークが、ピッチ周波数位置と、ピッチ周波数の倍音周波数位置とに発生している度合いで表される請求項 9 に記載の音声処理装置。

## 【請求項 11】

前記ピッチ強度は、前記入力信号の自己相関係数値である

10

20

30

40

50

請求項 9 に記載の音声処理装置。

【請求項 1 2】

前記特徴量計算部は、前記ノイズ抑圧スペクトルの周波数帯域を複数のバンドに分割し、前記バンドにおける前記ノイズ抑圧スペクトルの代表値に基づいて前記特徴量を算出する

請求項 2 に記載の音声処理装置。

【請求項 1 3】

前記ノイズ抑圧スペクトルはパワースペクトルである

請求項 1 2 に記載の音声処理装置。

【請求項 1 4】

前記ノイズ抑圧スペクトルは振幅スペクトルである

請求項 1 2 に記載の音声処理装置。

【請求項 1 5】

前記代表値は、前記バンドにおける前記ノイズ抑圧スペクトルの平均値である

請求項 1 2 に記載の音声処理装置。

【請求項 1 6】

前記代表値は、前記バンドにおける前記ノイズ抑圧スペクトルの最大値である

請求項 1 2 に記載の音声処理装置。

【請求項 1 7】

前記特徴量計算部は、前記ノイズ抑圧スペクトルにおける前記バンドの前記代表値の時間差分値を前記特徴量として算出する

請求項 1 2 に記載の音声処理装置。

【請求項 1 8】

入力信号の背景ノイズを推定し、

前記背景ノイズの推定結果に基づいて、前記入力信号の前記背景ノイズを抑圧し、

前記背景ノイズが抑圧された前記入力信号に基づいて特徴量を算出し、

前記特徴量に基づいて、前記入力信号から子音を検出する

ステップを含む音声処理方法。

【請求項 1 9】

入力信号の背景ノイズを推定し、

前記背景ノイズの推定結果に基づいて、前記入力信号の前記背景ノイズを抑圧し、

前記背景ノイズが抑圧された前記入力信号に基づいて特徴量を算出し、

前記特徴量に基づいて、前記入力信号から子音を検出する

ステップを含む処理をコンピュータに実行させるプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本技術は音声処理装置および方法、並びにプログラムに関し、特に、より精度よく子音を検出することができるようにした音声処理装置および方法、並びにプログラムに関する

。

【背景技術】

【0002】

近年、高齢社会が進むにつれ、加齢性の難聴に悩まされる人が増えている。加齢性難聴では高い周波数帯域の聴覚感度から悪化する傾向があるため、音声のうち特に子音の聞き取りが困難になる例が多い。

【0003】

そこで、子音をより聞き取りやすくする技術として、イコライザで入力信号の子音が存在する周波数帯域を強調するものがある。しかし、この技術では、入力信号が何であるかに関わらず常にその周波数帯域の強調が行なわれるため、子音は強調されるものの、それ以外の音の音質が変わり、音声の聞きにくさが発生してしまう。

10

20

30

40

50

## 【 0 0 0 4 】

また、携帯型電話機の発達により、いつでもどこでも携帯型電話機を用いた会話ができるようになったが、これは発話側の環境がうるさい場所である機会が増えたということでもある。音声信号のうち、子音部は母音部よりも相対的にパワーが小さいため、ノイズに子音部が埋もれてしまうことにより音声の明瞭度が下がり、会話音声の聞き取りにくくなる事態が発生する。特に加齢性難聴者にとっては、音声の聞き取りやすさは背景ノイズの大きさに影響されることが知られており、うるさい環境下でも音声を聞き取りやすくする技術が望まれている。

## 【 0 0 0 5 】

例えば、ノイズが大きい場合には、ノイズ抑圧技術でノイズを低減することで相対的に音声のSN比 (Signal to Noise ratio) を上げる技術があるが、この技術では音声自体の音質も変わってしまうことが多く、語音明瞭度や了解度が下がってしまう傾向がある。また、そもそも発話者自体がもごもごとした不明瞭な話し方をしている場合には、ノイズ抑圧技術は何の役にも立たない。

10

## 【 0 0 0 6 】

以上のような状況から、子音を検出して強調する技術が要望されており、これまでもいくつかの子音検出強調技術が提案されている。

## 【 0 0 0 7 】

例えば、そのような技術として、複数の時間フレームによって複数のフレーム信号を抽出し、このフレーム信号の平均パワーを計算して比較するだけで子音を検出し、強調する技術が提案されている (例えば、特許文献 1 および特許文献 2 参照)。

20

## 【 先行技術文献 】

## 【 特許文献 】

## 【 0 0 0 8 】

【 特許文献 1 】 特開 2 0 1 0 - 0 9 1 8 9 7 号 公 報

【 特許文献 2 】 特許第 0 4 8 7 6 2 4 5 号 公 報

## 【 発明の概要 】

## 【 発明が解決しようとする課題 】

## 【 0 0 0 9 】

ところで特許文献 1 や特許文献 2 に記載の技術では、子音区間や音節の長さを予め定義し、その定義に合うフレームのみを子音としているが、実際の音声はその定義にしたがうとは限らない。特に、言語によってそれらの定義は異なるため、アルゴリズムの言語依存性が高くなってしまふ。

30

## 【 0 0 1 0 】

また、子音の検出をフレーム信号のパワーの比較だけで行っているため、背景ノイズによりパワーが変動した場合には、子音を正しく検出することができなくなってしまう。

## 【 0 0 1 1 】

以上のように、上述した子音検出方法では、子音の背景にノイズが存在するような信号については、子音を精度よく検出することが困難であった。

## 【 0 0 1 2 】

本技術は、このような状況に鑑みてなされたものであり、より精度よく子音を検出することができるようにするものである。

40

## 【 課題を解決するための手段 】

## 【 0 0 1 3 】

本技術の一側面の音声処理装置は、入力信号の背景ノイズを推定する背景ノイズ推定部と、前記背景ノイズの推定結果に基づいて、前記入力信号の前記背景ノイズを抑圧するノイズ抑圧部と、前記背景ノイズが抑圧された前記入力信号に基づいて特徴量を算出する特徴量計算部と、前記特徴量に基づいて、前記入力信号から子音を検出する子音検出部とを備える。

## 【 0 0 1 4 】

50

前記背景ノイズ推定部には、周波数領域で前記背景ノイズを推定させ、前記ノイズ抑圧部には、前記入力信号から得られた入力スペクトルに含まれる前記背景ノイズを抑圧することでノイズ抑圧スペクトルを求めさせ、前記特徴量計算部には、前記ノイズ抑圧スペクトルに基づいて前記特徴量を算出させることができる。

【0015】

前記背景ノイズ推定部には、過去の前記入力スペクトルの平均値を求めることで前記背景ノイズを推定させることができる。

【0016】

音声処理装置には、前記ノイズ抑圧スペクトルの値が、前記背景ノイズの推定により得られた背景ノイズスペクトルを定数倍した値よりも大きい周波数について、前記入力スペクトルの強調を行なう子音強調部をさらに設けることができる。

10

【0017】

前記子音強調部には、予め定められた強調量で前記入力スペクトルの強調を行なわせることができる。

【0018】

音声処理装置には、前記入力信号の過去の母音部分のパワーの平均値と、前記入力信号の現在のパワーとの比に基づいて強調量を算出する子音強調レベル計算部をさらに設け、前記子音強調部には、前記強調量で前記入力スペクトルの強調を行なわせることができる。

【0019】

周波数方向に対して前記強調量の補間が行なわれるようにすることができる。

20

【0020】

前記ノイズ抑圧部には、スペクトルサブトラクション法により前記ノイズ抑圧スペクトルを求めさせることができる。

【0021】

前記入力信号のピッチ強度がさらに前記特徴量として用いられ、前記子音検出部には、前記特徴量としての前記ピッチ強度と、前記ノイズ抑圧スペクトルに基づいて算出された前記特徴量とに基づいて、前記入力信号から子音を検出させることができる。

【0022】

前記ピッチ強度は、前記ノイズ抑圧スペクトルのピークが、ピッチ周波数位置と、ピッチ周波数の倍音周波数位置とに発生している度合いで表されるようにすることができる。

30

【0023】

前記ピッチ強度が、前記入力信号の自己相関係数値であるようにすることができる。

【0024】

前記特徴量計算部には、前記ノイズ抑圧スペクトルの周波数帯域を複数のバンドに分割させ、前記バンドにおける前記ノイズ抑圧スペクトルの代表値に基づいて前記特徴量を算出させることができる。

【0025】

前記ノイズ抑圧スペクトルをパワースペクトルとすることができる。

【0026】

前記ノイズ抑圧スペクトルを振幅スペクトルとすることができる。

40

【0027】

前記代表値を、前記バンドにおける前記ノイズ抑圧スペクトルの平均値とすることができる。

【0028】

前記代表値を、前記バンドにおける前記ノイズ抑圧スペクトルの最大値とすることができる。

【0029】

前記特徴量計算部には、前記ノイズ抑圧スペクトルにおける前記バンドの前記代表値の時間差分値を前記特徴量として算出させることができる。

50

## 【 0 0 3 0 】

本技術の一側面の音声処理方法またはプログラムは、入力信号の背景ノイズを推定し、前記背景ノイズの推定結果に基づいて、前記入力信号の前記背景ノイズを抑圧し、前記背景ノイズが抑圧された前記入力信号に基づいて特徴量を算出し、前記特徴量に基づいて、前記入力信号から子音を検出するステップを含む。

## 【 0 0 3 1 】

本技術の一側面においては、入力信号の背景ノイズが推定され、前記背景ノイズの推定結果に基づいて、前記入力信号の前記背景ノイズが抑圧され、前記背景ノイズが抑圧された前記入力信号に基づいて特徴量が算出され、前記特徴量に基づいて、前記入力信号から子音を検出される。

10

## 【 発明の効果 】

## 【 0 0 3 2 】

本技術の一側面によれば、より精度よく子音を検出することができる。

## 【 図面の簡単な説明 】

## 【 0 0 3 3 】

【 図 1 】 子音強調装置の構成例を示す図である。

【 図 2 】 時間周波数変換について説明する図である。

【 図 3 】 背景ノイズの推定について説明する図である。

【 図 4 】 ノイズ抑圧スペクトルの算出について説明する図である。

【 図 5 】 特徴量の算出について説明する図である。

20

【 図 6 】 入力スペクトルの強調について説明する図である。

【 図 7 】 入力信号の強調結果の一例を示す図である。

【 図 8 】 子音強調処理を説明するフローチャートである。

【 図 9 】 子音検出処理を説明するフローチャートである。

【 図 1 0 】 強調量算出処理を説明するフローチャートである。

【 図 1 1 】 子音強調装置の他の構成例を示す図である。

【 図 1 2 】 子音強調装置の他の構成例を示す図である。

【 図 1 3 】 子音強調装置の他の構成例を示す図である。

【 図 1 4 】 子音検出装置の構成例を示す図である。

【 図 1 5 】 子音検出装置の他の構成例を示す図である。

30

【 図 1 6 】 コンピュータの構成例を示す図である。

## 【 発明を実施するための形態 】

## 【 0 0 3 4 】

以下、図面を参照して、本技術を適用した実施の形態について説明する。

## 【 0 0 3 5 】

## 第 1 の実施の形態

## 子音強調装置の構成例

本技術は、背景ノイズの抑圧を行った信号を基に子音の検出を行なうことで、背景にノイズがある場合にも精度よく子音を検出することができるようにするものである。また、入力信号、推定した背景ノイズ、およびノイズ抑圧信号のレベルを基にした強調量の決定を行うことで、ノイズを考慮した適切な子音強調を行なうことができるようにするものである。

40

## 【 0 0 3 6 】

図 1 は、本技術を適用した子音強調装置の一実施の形態の構成例を示す図である。

## 【 0 0 3 7 】

子音強調装置 1 1 は、音声信号である入力信号を入力とし、入力信号から子音部分を検出するとともに、その検出結果に基づいて子音を強調し、その結果得られた音声信号を出力信号として出力する。

## 【 0 0 3 8 】

子音強調装置 1 1 は、時間周波数変換部 2 1、背景ノイズ推定部 2 2、ノイズ抑圧スベ

50

クトル計算部 2 3、ピッチ強度計算部 2 4、特徴量計算部 2 5、子音検出部 2 6、子音強調レベル計算部 2 7、子音強調部 2 8、および周波数時間変換部 2 9 から構成される。

【 0 0 3 9 】

時間周波数変換部 2 1 は、供給された入力信号に対して時間周波数変換を行い、その結果得られた入力スペクトルを背景ノイズ推定部 2 2、ノイズ抑圧スペクトル計算部 2 3、子音強調レベル計算部 2 7、および子音強調部 2 8 に供給する。

【 0 0 4 0 】

背景ノイズ推定部 2 2 は、時間周波数変換部 2 1 から供給された入力スペクトルに基づいて背景ノイズ推定を行い、その結果得られた背景ノイズスペクトルをノイズ抑圧スペクトル計算部 2 3、および子音強調レベル計算部 2 7 に供給する。

10

【 0 0 4 1 】

なお、背景ノイズとは、入力信号の音声のうちの発話者の音声等とは異なる、環境音などのノイズ成分である。また、背景ノイズスペクトルは、背景ノイズのスペクトルである。

【 0 0 4 2 】

ノイズ抑圧スペクトル計算部 2 3 は、時間周波数変換部 2 1 から供給された入力スペクトルと、背景ノイズ推定部 2 2 から供給された背景ノイズスペクトルとに基づいて入力スペクトルに含まれる背景ノイズ成分を抑圧し、ノイズ抑圧スペクトルを求める。ノイズ抑圧スペクトル計算部 2 3 は、得られたノイズ抑圧スペクトルをピッチ強度計算部 2 4、特徴量計算部 2 5、および子音強調レベル計算部 2 7 に供給する。

20

【 0 0 4 3 】

ピッチ強度計算部 2 4 は、ノイズ抑圧スペクトル計算部 2 3 から供給されたノイズ抑圧スペクトルに基づいて、入力信号のピッチ強度を算出し、特徴量計算部 2 5 および子音検出部 2 6 に供給する。なお、ここでは、ノイズ抑圧スペクトルからピッチ強度を求める場合について説明するが、ノイズ抑圧前のスペクトルや、時間領域の信号である入力信号からピッチ強度が求められてもよい。

【 0 0 4 4 】

特徴量計算部 2 5 は、ノイズ抑圧スペクトル計算部 2 3 から供給されたノイズ抑圧スペクトルに基づいて、またはノイズ抑圧スペクトルと、ピッチ強度計算部 2 4 から供給されたピッチ強度とに基づいて特徴量を算出し、子音検出部 2 6 に供給する。特徴量計算部 2 5 で算出される特徴量は、入力信号からの子音検出に用いられる。

30

【 0 0 4 5 】

子音検出部 2 6 は、ピッチ強度計算部 2 4 から供給されたピッチ強度と、特徴量計算部 2 5 から供給された特徴量とに基づいて、入力信号の子音区間を検出し、その検出結果を子音強調レベル計算部 2 7 に供給する。

【 0 0 4 6 】

なお、より詳細には、子音区間の検出では処理対象となっている入力信号のフレームが、子音のフレームであるか、母音のフレームであるか、またはその他のフレーム、つまり子音でも母音でもないフレームであるかが特定される。以下では、特に子音のフレームを子音フレームと称し、母音のフレームを母音フレームと称することとする。

40

【 0 0 4 7 】

子音強調レベル計算部 2 7 は、時間周波数変換部 2 1 からの入力スペクトル、背景ノイズ推定部 2 2 からの背景ノイズスペクトル、ノイズ抑圧スペクトル計算部 2 3 からのノイズ抑圧スペクトル、および子音検出部 2 6 からの検出結果に基づいて強調量を算出する。すなわち、子音検出により子音フレームであるとされたフレームの強調量が算出されて、子音強調レベル計算部 2 7 から子音強調部 2 8 に供給される。

【 0 0 4 8 】

子音強調部 2 8 は、子音強調レベル計算部 2 7 から供給された強調量を、時間周波数変換部 2 1 から供給された入力スペクトルに乗算することで、入力スペクトルの子音部分を強調し、周波数時間変換部 2 9 に供給する。

50

## 【 0 0 4 9 】

周波数時間変換部 2 9 は、子音強調部 2 8 から供給された入力スペクトルに対して周波数時間変換を行なって、その結果得られた出力時間波形を出力信号として出力する。

## 【 0 0 5 0 】

入力スペクトルの算出について

次に、図 1 に示した子音強調装置 1 1 を構成する各部の処理について説明する。

## 【 0 0 5 1 】

まず、入力信号を入力スペクトルに変換する時間周波数変換部 2 1 の処理について説明する。

## 【 0 0 5 2 】

例えば、時間周波数変換部 2 1 に図 2 の矢印 A 1 1 に示す波形の入力信号が入力されたとする。なお、矢印 A 1 1 に示す入力信号において、横方向は時間を示しており、縦方向は振幅を示している。

## 【 0 0 5 3 】

このような矢印 A 1 1 に示す入力信号が供給されると、時間周波数変換部 2 1 は、入力信号を構成する所定の連続する複数のサンプルをまとめてフレーム化する。この例では、入力信号の区間 L 1 1 乃至区間 L 1 9 のそれぞれが 1 つのフレームとされる。

## 【 0 0 5 4 】

また、時間周波数変換部 2 1 は、入力信号の各フレームに対して矢印 A 1 2 に示す形状の窓、つまり窓関数を用いて窓がけを行なう。ここで、矢印 A 1 2 に示す窓関数において縦方向は窓関数の値を示しており、横方向は時間、つまり窓関数の値が乗算される入力信号のサンプル位置を示している。

## 【 0 0 5 5 】

なお、窓がけには、例えばサイン窓を用いてもよいし、ハニング窓やハミング窓などを用いるようにしてもよい。但し、周波数信号を時間信号に戻す逆変換時との整合が取れるようにする必要がある。

## 【 0 0 5 6 】

時間周波数変換部 2 1 は、入力信号のフレームを構成する各サンプルに窓関数を乗算することで窓がけを行なうと、その結果得られた信号に対して零詰めを行なう。例えば、入力信号の区間 L 1 1 の部分に対して矢印 A 1 2 に示した窓関数により窓がけを行い、その結果得られた信号に零詰めを行なうと、矢印 A 1 3 に示す信号が得られる。

## 【 0 0 5 7 】

矢印 A 1 3 に示す信号において、縦方向は振幅を示しており、横方向は時間を示している。この矢印 A 1 3 に示す信号では、区間 L 3 1 の部分が、零詰めが行なわれた部分であり、この部分では信号の振幅が 0 となっている。また、後段の時間周波数変換における周波数分解能を高くするために、零詰め後の信号の長さは、例えば窓の長さの 2 倍や 4 倍などとされるとよい。

## 【 0 0 5 8 】

零詰めが行なわれると、さらに時間周波数変換部 2 1 は、零詰めにより得られた信号に対して離散フーリエ変換などの時間周波数変換を行い、時間信号を周波数信号である入力スペクトルへと変換する。例えば、矢印 A 1 3 に示す信号に対して離散フーリエ変換が行なわれると、矢印 A 1 4 に示す入力スペクトルが得られる。なお、矢印 A 1 4 に示す入力スペクトルにおいて横軸は周波数を示しており、縦軸はパワーまたは振幅を示している。

## 【 0 0 5 9 】

このようにして入力信号のフレームから得られる入力スペクトルは、パワースペクトルでもよいし、振幅スペクトルや対数振幅スペクトルであってもよい。また、入力スペクトルを得るための時間周波数変換として、離散フーリエ変換に限らず、離散コサイン変換などが行なわれるようにしてもよい。

## 【 0 0 6 0 】

さらに、図 2 の例では、周波数分解能を高くするために、零詰めによるオーバーサンブ

10

20

30

40

50



リングにより周波数変換長が窓の長さより長くされているが、特に零詰めが行なわれないようにしてもよい。

【 0 0 6 1 】

以上のような処理により、入力信号の各フレームについて入力スペクトルが得られる。

【 0 0 6 2 】

背景ノイズ推定について

続いて、背景ノイズ推定部 2 2 により行なわれる背景ノイズ推定について説明する。

【 0 0 6 3 】

例えば図 3 に示すように、時間周波数変換部 2 1 において、矢印 A 2 1 に示す入力信号を構成する 5 つのフレーム  $F(t-1)$  乃至フレーム  $F(t-5)$  から、それぞれ矢印 A 2 2 乃至矢印 A 2 6 に示す入力スペクトル  $X(t-1, f)$  乃至入力スペクトル  $X(t-5, f)$  が得られたとする。ここで、矢印 A 2 1 に示す入力信号の縦方向は振幅を示しており、横方向は時間を示している。また、各フレームの入力スペクトル  $X(t, f)$  において、 $t$  は時間インデックスを示しており、 $f$  は周波数を示している。

10

【 0 0 6 4 】

背景ノイズ推定部 2 2 は、時間周波数変換部 2 1 で得られた入力スペクトル  $X(t-1, f)$  乃至入力スペクトル  $X(t-5, f)$  の平均値を求め、得られた入力スペクトルの平均値を背景ノイズスペクトル  $N(t, f)$  とする。図 3 の例では、矢印 A 2 7 に示されるスペクトルが、入力スペクトル  $X(t-1, f)$  乃至入力スペクトル  $X(t-5, f)$  の平均を求めることで得られた背景ノイズスペクトル  $N(t, f)$  を表している。

20

【 0 0 6 5 】

このように、背景ノイズ推定部 2 2 では、入力信号の過去の所定フレーム数分の入力スペクトルの平均値が背景ノイズであるとして背景ノイズの推定が行なわれる。一般的に比較的長い期間について、音声信号の各フレームのスペクトルの平均をとれば、ほぼノイズのスペクトルとなることが知られている。

【 0 0 6 6 】

例えば、過去  $M$  フレームの入力スペクトルの平均が背景ノイズスペクトルであるとされる場合には、背景ノイズ推定部 2 2 は次式 (1) を計算することで、時間インデックスが  $t$  であるフレームの背景ノイズスペクトル  $N(t, f)$  を算出する。

【 0 0 6 7 】

【 数 1 】

$$N(t, f) = \frac{1}{M} \sum_{\tau=1}^M X(t-\tau, f) \quad \dots (1)$$

30

【 0 0 6 8 】

なお、式 (1) において、 $X(t, f)$  は、時間インデックスが  $t$  であるフレームの入力スペクトルを示している。

【 0 0 6 9 】

また、背景ノイズスペクトルの算出時において、レベル変動の大きなフレームについてはノイズではなく音声信号であるとみなして、そのフレームの入力スペクトルが背景ノイズスペクトルを算出するための平均値計算処理から除かれるようにしてもよい。

40

【 0 0 7 0 】

レベル変動が大きいフレームの特定は、例えば隣接フレームとの入力スペクトルのパワーの比率に基づいて行なわれてもよいし、入力スペクトルに対する閾値処理などにより行なわれるようにしてもよい。

【 0 0 7 1 】

また、背景ノイズスペクトルの算出は、式 (1) の計算に限らず、他の方法により算出されるようにしてもよい。例えば、過去の所定数のフレームの入力スペクトルの平均値を背景ノイズスペクトルとするのではなく、過去のフレームの影響を継続して受けるように、フレームごとに背景ノイズスペクトルが更新されてもよい。

50

【 0 0 7 2 】

そのような場合、例えば背景ノイズ推定部 2 2 は、次式 ( 2 ) を計算することで背景ノイズスペクトル  $N(t, f)$  を算出する。

【 0 0 7 3 】

【 数 2 】

$$N(t, f) = \frac{\alpha_n(f)N(t-1, f) + \alpha_x(f)X(t, f)}{\alpha_n(f) + \alpha_x(f)} \quad \dots (2)$$

【 0 0 7 4 】

なお、式 ( 2 ) において、 $\alpha_n(f)$  および  $\alpha_x(f)$  は、所定の係数を示している。

10

【 0 0 7 5 】

したがって、式 ( 2 ) では、直前のフレームの背景ノイズスペクトルと、現フレームの入力スペクトルとを重み付き加算することで、現フレームの背景ノイズスペクトルが算出される。例えば、レベル変動の大きなフレームの寄与度を小さくしたい場合には、レベル変動の大きなフレームについては係数  $\alpha_x(f)$  の値を、0 などの小さい値とすればよい。

【 0 0 7 6 】

なお、以下、特に時間インデックスを区別する必要がない場合には、背景ノイズスペクトル  $N(t, f)$  を単に背景ノイズスペクトル  $N(f)$  とも称することとする。同様に、以下、特に時間インデックスを区別する必要がない場合には、入力スペクトル  $X(t, f)$  を単に入力スペクトル  $X(f)$  とも称することとする。

20

【 0 0 7 7 】

ノイズ抑圧スペクトルの算出について

次に、ノイズ抑圧スペクトル計算部 2 3 により行なわれるノイズ抑圧スペクトルの算出について説明する。

【 0 0 7 8 】

例えば、ノイズ抑圧スペクトルは、図 4 に示すように、スペクトルサブトラクション法により算出される。

【 0 0 7 9 】

図 4 では、矢印 A 4 1 乃至矢印 A 4 3 に示される各スペクトルが、それぞれノイズ抑圧スペクトル  $S(f)$ 、入力スペクトル  $X(f)$ 、および背景ノイズスペクトル  $N(f)$  を示している。なお、図 4 に示す各スペクトルにおいて、縦軸はパワーまたは振幅を示しており、横軸は周波数を示している。

30

【 0 0 8 0 】

スペクトルサブトラクション法では、音声部分のスペクトルであるノイズ抑圧スペクトル  $S(f)$  と、背景ノイズの成分である背景ノイズスペクトル  $N(f)$  の和が、入力スペクトル  $X(f)$  であると仮定される。

【 0 0 8 1 】

したがって、入力スペクトル  $X(f)$  から背景ノイズスペクトル  $N(f)$  を減算して得られるスペクトルが、推定により得られたノイズ抑圧スペクトル  $S(f)$  となる。なお、図 4 では、入力スペクトル  $X(f)$  における斜線部分は、入力スペクトル  $X(f)$  に含まれている背景ノイズ成分を表している。

40

【 0 0 8 2 】

より具体的には、ノイズ抑圧スペクトル計算部 2 3 は、例えば入力スペクトル  $X(f)$  と背景ノイズスペクトル  $N(f)$  に基づいて次式 ( 3 ) を計算することで、ノイズ抑圧スペクトル  $S(f)$  を算出する。

【 0 0 8 3 】

【数 3】

$$S(f) = X(f) \left( 1 - \beta(f) \left( \frac{|N(f)|}{|X(f)|} \right)^i \right)^{\frac{1}{i}} \quad \dots (3)$$

【0084】

なお、式(3)において  $\beta(f)$  はノイズ抑圧量を決める係数であり、周波数ごとに異なる値を用いても、全ての周波数で同じ値を用いてもよい。また、 $i$  はノイズ抑圧のドメインを決定する値である。

【0085】

以上のようにして求められるノイズ抑圧スペクトル  $S(f)$  は、パワースペクトルでもよいし、振幅スペクトルでもよい。

【0086】

ピッチ強度の算出について

さらに、ピッチ強度計算部 24 によるピッチ強度の算出について説明する。

【0087】

ピッチ強度は、ノイズ抑圧スペクトル  $S(f)$  から算出される。

【0088】

ここで、ピッチ強度は、ピッチ周波数と、そのピッチ周波数の倍音周波数において、パワースペクトルまたは振幅スペクトルであるノイズ抑圧スペクトルのピークがどの程度存在するかで表される。すなわち、ピッチ強度は、ノイズ抑圧スペクトルのピークが、ピッチ周波数位置と、そのピッチ周波数の倍音周波数の位置に発生している度合いで表される。

【0089】

したがって、ピッチ強度は、ピッチ周波数位置にピークが存在するか否かと、その倍音周波数位置にもピークが存在するか否か、つまりピークが存在する倍音周波数がいくつ存在するかに基づいて定められる。

【0090】

なお、ピークであるか否かは、例えばピーク周波数近辺のスペクトルの曲率に基づいて、ピークらしさを求めることで決定される。また、ピーク周波数におけるスペクトルと、その周囲のスペクトルまたは周囲のスペクトルの平均値との割合や差などに基づいてピークらしさを求めることで、ピークであるか否かが決定されてもよい。

【0091】

特徴量の算出について

続いて、特徴量計算部 25 による特徴量の算出について説明する。

【0092】

なお、特徴量は、ノイズ抑圧スペクトルとピッチ強度とに基づいて算出されるようにしてもよいが、ここではノイズ抑圧スペクトルに基づいて特徴量が算出される例について説明する。

【0093】

例えば、ノイズ抑圧スペクトル計算部 23 から特徴量計算部 25 に、図 5 に示すノイズ抑圧スペクトル  $S(f)$  が供給されたとする。なお、図 5 において、縦軸はパワーまたは振幅を示しており、横軸は周波数を示している。

【0094】

また、ノイズ抑圧スペクトル  $S(f)$  における 1 つの長方形は、1 つの周波数(周波数ビン)におけるスペクトルの値を表している。この例では、ノイズ抑圧スペクトル  $S(f)$  には、17 個の周波数ビンにおけるスペクトルの値が含まれている。

【0095】

このようなノイズ抑圧スペクトル  $S(f)$  が供給されると、特徴量計算部 25 は、ノイズ抑圧スペクトル  $S(f)$  の周波数帯域を複数のバンドに分割する。すなわち、ノイズ抑

10

20

30

40

50

圧スペクトル  $S(f)$  の周波数帯域が、点線の矩形で表される 7 つのバンド BD 1 1 乃至バンド BD 1 7 に分割される。例えば、最も低周波数側の 2 つの周波数ビンが束ねられて、バンド BD 1 1 とされている。

【0096】

なお、バンドの分割方法は、各バンドが均一幅となるように分割されてもよいし、聴覚フィルタを模した不均一幅で分割が行なわれるようにしてもよい。図 5 の例では、バンド BD 1 1 乃至バンド BD 1 4 は、2 つの周波数ビンから構成されているが、バンド BD 1 5 乃至バンド BD 1 7 は、3 つの周波数ビンから構成されている。

【0097】

さらに、特徴量計算部 2 5 は、ノイズ抑圧スペクトル  $S(f)$  を構成する各バンドについて、バンド内のスペクトルの最大値をそのバンドの代表値とし、各バンドの代表値を組み合わせて得られるベクトルを、ノイズ抑圧スペクトル  $S(f)$  の特徴量とする。

10

【0098】

例えば、バンド BD 1 1 乃至バンド BD 1 7 の代表値が 55、50、40、30、20、25、および 20 である場合、それらの値を順番に並べて得られるベクトル  $b = \{55, 50, 40, 30, 20, 25, 20\}$  が特徴量とされる。

【0099】

なお、ここでは、バンド内のスペクトルの最大値が代表値とされる例について説明したが、バンド内のスペクトルの値の平均値が代表値とされるようにしてもよい。また、子音の立ち上がりを検出するための特徴量として、ノイズ抑圧スペクトル  $S(f)$  の各バンドの代表値の時間差分値、つまり時間方向に隣接するフレームについての同じバンドの代表値の差分値が用いられるようにしてもよい。

20

【0100】

子音フレームの検出について

次に、子音検出部 2 6 による子音フレームの検出について説明する。

【0101】

例えば、子音検出部 2 6 は、特徴量計算部 2 5 から供給された特徴量に基づいて線形判別を行なうことにより、入力信号の処理対象となっている現フレームが子音フレームであるか否かを判定する。

【0102】

具体的には、例えば子音検出部 2 6 は、次式 (4) に示される線形判別式  $Y$  に特徴量を代入して、判別を行なう。

30

【0103】

【数 4】

$$Y = \sum_{n=1}^N a_n b_n + a_0 \quad \dots (4)$$

【0104】

なお、式 (4) において、 $a_n$  (但し、 $1 \leq n \leq N$ ) および  $a_0$  は、予め学習された係数および定数を示しており、子音検出部 2 6 は、これらの係数と定数からなる係数ベクトルを保持している。また、 $b_n$  (但し、 $1 \leq n \leq N$ ) は、特徴量計算部 2 5 で算出された特徴量であるベクトルの各要素を示している。

40

【0105】

子音検出部 2 6 は、特徴量計算部 2 5 から供給された特徴量を式 (4) に示す線形判別式  $Y$  に代入し、その結果得られた値が負であった場合、つまり  $Y = a_n b_n + a_0 < 0$  である場合、現フレームは子音フレームであるとする。

【0106】

また、子音検出部 2 6 は、線形判別式  $Y$  の値が 0 以上である場合、さらにピッチ強度が閾値より大きいかが否かを判定することで、現フレームが母音フレームであるか否かを判定する。例えば、ピッチ強度が閾値より大きい場合には、現フレームは母音フレームである

50

とされ、ピッチ強度が閾値以下である場合、現フレームは子音フレームでも母音フレームでもない、その他のフレームであるとされる。

【0107】

子音検出部26は、このようにして判別された現フレームの種別を示す情報を、子音の検出結果として子音強調レベル計算部27に供給する。

【0108】

例えば母音フレームのスペクトルでは、ピークが周期的に出現することが知られており、入力信号のピッチ強度に基づいて、母音フレームらしいか否かを特定することが可能である。

【0109】

子音強調装置11では、周波数領域において入力信号のピッチ強度を求めることで、例えばピークが出現しやすい低域側の周波数帯域のみを用いるなど、特定周波数帯域を選択的に用いてピッチ強度を算出することができる。これにより、母音の検出精度を向上させることができる。

【0110】

また、子音強調装置11では、背景ノイズが抑圧されたノイズ抑圧スペクトルがピッチ強度の算出に用いられるが、ノイズ抑圧スペクトルは背景ノイズが抑圧されたスペクトルであるので、より高精度にピークを検出することができるようになる。したがって、ノイズ抑圧スペクトルを用いれば、より高精度にピッチ強度を算出することができる。

【0111】

なお、以上においては、子音フレームの判別に用いる特徴量として、ノイズ抑圧スペクトル $S(f)$ から得られた特徴量を用いる例について説明したが、ノイズ抑圧スペクトル $S(f)$ から得られる特徴量だけでなく、ピッチ強度も特徴量として用いられてもよい。

【0112】

そのような場合、例えば特徴量として用いられるピッチ強度は、線形判別式 $Y$ の項として含まれるようにしてもよいし、ピッチ強度単独による子音の判別結果を線形判別式 $Y$ にカスケード接続するようにしてもよい。このように子音フレームの判別にピッチ強度を用いることで、さらに子音の検出精度を向上させることができる。

【0113】

また、子音の判別方法として、線形判別の他にサポートベクターマシンやニューラルネットワークなどの判別方法を用いるようにしてもよい。

【0114】

強調量の算出と入力スペクトルの強調について

さらに、子音強調レベル計算部27による強調量の算出と、子音強調部28による入力スペクトルの強調について説明する。

【0115】

例えば、子音強調レベル計算部27は、入力信号の過去の母音フレームのパワーの平均値を母音部パワーとして算出し、保持している。母音フレームのパワーは、例えば母音フレームの入力スペクトルにおける各周波数のパワーの平均値などとされる。

【0116】

子音強調レベル計算部27は、処理対象の現フレームが母音フレームである場合に、保持している母音部パワーの更新を行なう。

【0117】

具体的には、子音強調レベル計算部27は、子音検出部26から供給された子音の検出結果から、現フレームが母音のフレームであると特定された場合、保持している母音部パワーと、時間周波数変換部21から供給された現フレームの入力スペクトルとに基づいて、母音部パワーを更新する。

【0118】

そして、子音強調レベル計算部27は、子音検出部26から供給された子音の検出結果から、現フレームが子音のフレームであると特定された場合、保持している母音部パワー

10

20

30

40

50

を用いて強調量を算出する。

【0119】

例えば、子音強調レベル計算部27は、時間周波数変換部21から供給された現フレームの入力スペクトルにおける各周波数のパワーの平均値を求めて現フレームパワーとする。現フレームパワーは、入力スペクトルの全体のパワーである。そして、子音強調レベル計算部27は、次式(5)を計算することで、現フレームの強調量を算出する。

【0120】

【数5】

$$\text{強調量} = \frac{\text{母音部パワー}}{\text{現フレームパワー}} \dots (5)$$

10

【0121】

式(5)では、現フレームの入力スペクトルのパワーと、過去の母音フレームのパワーの平均値との比(割合)が強調量として算出される。これは、子音部分のパワーが母音部分のパワーと同程度となるように強調すれば、子音が十分聞きとりやすくなるからである。

【0122】

なお、入力スペクトルの強調量は、式(5)で得られる値に限らず、その他、例えば予め定めた定数とされてもよいし、式(5)で得られる値と、予め定めた定数とのうちの大きい方または小さい方の値などとされてもよい。

20

【0123】

また、強調量は、実際の子音強調音を再生する環境に応じて変更されるようにしてもよい。例えば高域が出にくい環境で再生する場合には強調量が多めにされ、高域が元々大きめに再生される環境では強調量が少なめにされるようにしてもよい。

【0124】

子音強調部28では、以上のようにして算出された強調量を用いられて、入力スペクトルの強調が行なわれる。

【0125】

例えば、入力信号を強調する際に、入力信号の全ての帯域、または特定の定められた帯域に対して同じ強調量でスペクトルの強調を行なうと、子音成分だけでなくノイズ成分も同時に強調されてしまう。そうすると、強調後の音声は、ノイズ感の大きい不快な音声となってしまう。

30

【0126】

そこで、子音強調装置11では、背景ノイズが優位なスペクトルについては強調が行なわれないようにする。

【0127】

具体的には、子音強調レベル計算部27は、例えば図6に示すように、ノイズ抑圧スペクトル $S(f)$ の値が、背景ノイズスペクトル $N(f)$ の値の定数倍よりも大きい場合にのみ強調が行なわれるようにする。

【0128】

なお、図6において、折れ線C11乃至折れ線C13は、ノイズ抑圧スペクトル $S(f)$ 、背景ノイズスペクトル $N(f)$ 、および定数が乗算された背景ノイズスペクトル $N(f)$ を示している。また、図中、横軸は周波数を示しており、縦軸はパワーまたは振幅を示している。

40

【0129】

図6の例では、折れ線C13により示される、背景ノイズスペクトル $N(f)$ に予め定められた定数が乗算された値と、曲線C11に示すノイズ抑圧スペクトル $S(f)$ の値とが周波数ごとに比較される。すなわち、子音強調レベル計算部27は、定数が乗算された背景ノイズスペクトル $N(f)$ と、ノイズ抑圧スペクトル $S(f)$ とを比較し、その比較結果と強調量を子音強調部28に供給する。

50

## 【 0 1 3 0 】

この例では、図中、上方向を向く矢印が記されている周波数において、ノイズ抑圧スペクトル  $S(f)$  が、背景ノイズスペクトル  $N(f)$  の定数倍の値よりも大きくなっており、この部分のスペクトルが強調される。上向きの矢印は、周波数成分が強調される様子を表している。

## 【 0 1 3 1 】

このように背景ノイズスペクトル  $N(f)$  との比較を行なうのは、子音フレームにおいて背景ノイズよりもパワーまたは振幅が大きい周波数帯域は子音成分が含まれている周波数帯域、つまり子音が関係している周波数帯域であるはずだからである。

## 【 0 1 3 2 】

また、ノイズ抑圧スペクトル  $S(f)$  が背景ノイズスペクトル  $N(f)$  の定数倍の値以下である周波数帯域は、子音等の音声よりも背景ノイズが優位な周波数帯域であるので、スペクトルの強調は行なわれない。

## 【 0 1 3 3 】

子音強調部 28 は、子音強調レベル計算部 27 からの比較結果に基づいて、ノイズ抑圧スペクトル  $S(f)$  の値が、定数が乗算された背景ノイズスペクトル  $N(f)$  の値よりも大きい周波数についてのみ、入力スペクトルに強調量を乗算する。

## 【 0 1 3 4 】

このように、背景ノイズが優位なスペクトルについては強調を行わないようにすることで、強調後の音声の音質が、子音のみが強調されたように聞こえるものとなるように音声の子音部分を強調することができる。

## 【 0 1 3 5 】

但し、強調を行うスペクトルが歯抜けになるとミュージカルノイズと呼ばれる耳障りなノイズが発生する可能性があるため、周波数方向に強調量の補間を行うことが望ましい。例えば、強調量の補間は、ノイズ抑圧スペクトル  $S(f)$  の値と、定数が乗算された背景ノイズスペクトル  $N(f)$  の値との比較結果に基づいて行なえばよい。

## 【 0 1 3 6 】

なお、図 6 では、定数が 1 よりも大きい値である例について説明したが、定数の値は、1 より小さい値であってもよい。また、定数の値は周波数ごとに異なる値とされるようにしてもよい。

## 【 0 1 3 7 】

以上のようにして、子音フレームにおける、背景ノイズが優位でない帯域のみスペクトルの強調を行なうと、強調後の入力スペクトルから、例えば図 7 に示す出力信号が得られる。なお、図 7 において縦軸は振幅を示しており、横軸は時間を示している。

## 【 0 1 3 8 】

図 7 では、矢印 A 6 1 は、子音部分が強調される前の入力信号の時間波形を示しており、矢印 A 6 2 は、子音部分が強調された出力信号の時間波形を示している。

## 【 0 1 3 9 】

この例では、矢印 Q 1 1 乃至矢印 Q 1 9 に示すように、入力信号の子音部分のレベルが強調されて、出力信号における、それらの子音部分と同じ部分のレベルが入力信号と比べてより大きくなっていることが分かる。

## 【 0 1 4 0 】

上述した従来技術では、背景にノイズがある場合には、精度よく子音を検出することができなかった。これに対して、子音強調装置 11 では、背景ノイズを抑圧したノイズ抑圧スペクトルを求め、少なくともノイズ抑圧スペクトルを用いて得られた特徴量に基づいて、周波数領域で子音を検出することで、より高精度に子音を検出することができる。

## 【 0 1 4 1 】

また、従来技術では、音声信号の時間領域での増幅を行なっているため、背景にノイズがある場合には、子音だけでなくノイズも増幅されてしまう。そうすると、増幅後の音声を再生すると、子音ではなくノイズが強調されたように聞こえてしまう。つまり、従来

10

20

30

40

50

の技術では、ノイズを考慮した強調が行なわれていないため、増幅により得られる音声はノイズ感だけが強くなったように聞こえるものになってしまう。

【0142】

これに対して、子音強調装置11では、周波数領域において子音フレームの背景ノイズが優位な周波数帯域以外の周波数帯域を強調することで、子音のみを強調したような音声を得ることができる。つまり、より効果的に音声の強調を行なうことができる。

【0143】

また、子音強調装置11では、周波数領域で母音部パワーや現フレームパワーを算出するので、パワーの算出を行う際に全帯域を用いるのではなく、音声が含まれない帯域を除外するなど、特定周波数帯域を選択的に用いてパワーを算出することができ、より自由度の高い処理を行なうことができる。

10

【0144】

子音強調処理について

ところで、子音強調装置11に入力信号が供給され、入力信号の子音部分の強調が指示されると、子音強調装置11は子音強調処理を行い、出力信号を生成する。

【0145】

以下、図8のフローチャートを参照して、子音強調装置11による子音強調処理について説明する。なお、子音強調処理は、入力信号のフレームごとに行われる。

【0146】

ステップS11において、時間周波数変換部21は、供給された入力信号に対して時間周波数変換を行い、その結果得られた入力スペクトルを背景ノイズ推定部22、ノイズ抑圧スペクトル計算部23、子音強調レベル計算部27、および子音強調部28に供給する。

20

【0147】

例えば、入力信号の処理対象のフレームである現フレームに対して窓関数が乗算され、さらに窓関数が乗算された信号が離散フーリエ変換されて入力スペクトルとされる。

【0148】

ステップS12において、背景ノイズ推定部22は、時間周波数変換部21からの入力スペクトルに基づいて背景ノイズ推定を行い、得られた背景ノイズスペクトルをノイズ抑圧スペクトル計算部23、および子音強調レベル計算部27に供給する。

30

【0149】

例えば、上述した式(1)または式(2)の計算が行なわれ、背景ノイズスペクトル $N(f)$ が求められる。

【0150】

ステップS13において、ノイズ抑圧スペクトル計算部23は、時間周波数変換部21からの入力スペクトルと、背景ノイズ推定部22からの背景ノイズスペクトルとに基づいてノイズ抑圧スペクトルを求め、ピッチ強度計算部24、特徴量計算部25、および子音強調レベル計算部27に供給する。例えば、上述した式(3)の計算が行なわれて、ノイズ抑圧スペクトル $S(f)$ が求められる。

【0151】

ステップS14において、ピッチ強度計算部24は、ノイズ抑圧スペクトル計算部23からのノイズ抑圧スペクトルに基づいて入力信号のピッチ強度を算出し、特徴量計算部25および子音検出部26に供給する。

40

【0152】

ステップS15において、特徴量計算部25は、少なくともノイズ抑圧スペクトル計算部23から供給されたノイズ抑圧スペクトルを用いて特徴量を算出し、子音検出部26に供給する。例えば、特徴量計算部25は、図5を参照して説明したようにノイズ抑圧スペクトルを複数のバンドに分割し、各バンドの代表値を並べて得られるベクトルを特徴量とする。

【0153】

50



ステップ S 1 6 において、子音検出部 2 6 は子音検出処理を行なって現フレームの種別を特定し、その結果を子音強調レベル計算部 2 7 に供給する。

【 0 1 5 4 】

ここで、図 9 のフローチャートを参照して、図 8 のステップ S 1 6 の処理に対応する子音検出処理について説明する。

【 0 1 5 5 】

ステップ S 5 1 において、子音検出部 2 6 は、特徴量計算部 2 5 から供給された特徴量を線形判別式に代入する。例えば、上述した式 ( 4 ) に示す線形判別式に特徴量を構成する各要素  $b_n$  が代入される。

【 0 1 5 6 】

ステップ S 5 2 において、子音検出部 2 6 は、線形判別式への特徴量の代入結果が負の値であるか否かを判定する。

【 0 1 5 7 】

ステップ S 5 2 において代入結果が負であると判定された場合、ステップ S 5 3 において、子音検出部 2 6 は、現フレームは子音フレームであるとして、その旨の子音検出結果を子音強調レベル計算部 2 7 に供給する。子音検出結果が子音強調レベル計算部 2 7 に供給されると、子音検出処理は終了し、その後、処理は図 8 のステップ S 1 7 に進む。

【 0 1 5 8 】

これに対して、ステップ S 5 2 において代入結果が負でないと判定された場合、ステップ S 5 4 において、子音検出部 2 6 は、ピッチ強度計算部 2 4 から供給されたピッチ強度が所定の閾値よりも大きいか否かを判定する。

【 0 1 5 9 】

ステップ S 5 4 において、ピッチ強度が閾値よりも大きいと判定された場合、ステップ S 5 5 において、子音検出部 2 6 は、現フレームは母音フレームであるとして、その旨の子音検出結果を子音強調レベル計算部 2 7 に供給する。子音検出結果が子音強調レベル計算部 2 7 に供給されると、子音検出処理は終了し、その後、処理は図 8 のステップ S 1 7 に進む。

【 0 1 6 0 】

また、ステップ S 5 4 において、ピッチ強度が閾値以下であると判定された場合、ステップ S 5 6 において、子音検出部 2 6 は、現フレームは子音フレームでも母音フレームでもない、その他のフレームであるとする。そして、子音検出部 2 6 は、現フレームはその他のフレームである旨の子音検出結果を子音強調レベル計算部 2 7 に供給する。子音検出結果が子音強調レベル計算部 2 7 に供給されると、子音検出処理は終了し、その後、処理は図 8 のステップ S 1 7 に進む。

【 0 1 6 1 】

図 8 のフローチャートの説明に戻り、ステップ S 1 6 において子音検出が行なわれると、ステップ S 1 7 において、子音強調レベル計算部 2 7 は強調量算出処理を行なって、その結果得られた強調量を子音強調部 2 8 に供給する。

【 0 1 6 2 】

ここで、図 1 0 のフローチャートを参照して、図 8 のステップ S 1 7 の処理に対応する強調量算出処理について説明する。

【 0 1 6 3 】

ステップ S 8 1 において、子音強調レベル計算部 2 7 は、子音検出部 2 6 からの子音検出結果に基づいて、現フレームは子音フレームであるか否かを判定する。

【 0 1 6 4 】

ステップ S 8 1 において、子音フレームではないと判定された場合、ステップ S 8 2 において、子音強調レベル計算部 2 7 は、子音検出部 2 6 からの子音検出結果に基づいて、現フレームは母音フレームであるか否かを判定する。

【 0 1 6 5 】

ステップ S 8 2 において、現フレームは母音フレームではないと判定された場合、つま

10

20

30

40

50

り現フレームはその他のフレームであると判定された場合、入力スペクトルの強調量は出力されずに強調量算出処理は終了し、その後、処理は図8のステップS18に進む。なお、この場合、現フレームは子音フレームではないので、ステップS18では入力スペクトルの強調は行なわれない。

【0166】

また、ステップS82において、現フレームは母音フレームであると判定された場合、ステップS83において、子音強調レベル計算部27は、保持している母音部パワーと、時間周波数変換部21からの入力スペクトルとに基づいて、母音部パワーを更新する。例えば、現フレームを含む過去の母音フレームの入力スペクトルのパワーの平均値が、更新後の母音部パワーとされ、子音強調レベル計算部27に保持される。

10

【0167】

母音部パワーが更新されると強調量算出処理は終了し、その後、処理は図8のステップS18に進む。この場合においても、現フレームは子音フレームではないので、ステップS18において入力スペクトルの強調は行なわれない。

【0168】

さらに、ステップS81において、現フレームが子音フレームであると判定された場合、ステップS84の処理が行なわれる。

【0169】

すなわち、ステップS84において、子音強調レベル計算部27は、保持している母音部パワーと、時間周波数変換部21からの入力スペクトルとに基づいて強調量を算出し、子音強調部28に供給する。例えば、上述した式(5)の計算が行なわれ、強調量が算出される。

20

【0170】

ステップS85において、子音強調レベル計算部27は、背景ノイズ推定部22からの背景ノイズスペクトルと、ノイズ抑圧スペクトル計算部23からのノイズ抑圧スペクトルとを比較して、その比較結果を子音強調部28に供給する。

【0171】

例えば、図6を参照して説明したように、背景ノイズスペクトル $N(f)$ に定数  $K$  が乗算された値と、ノイズ抑圧スペクトル $S(f)$ の値とが周波数ごとに比較される。

【0172】

背景ノイズスペクトルとノイズ抑圧スペクトルとの比較結果が子音強調部28に供給されると、強調量算出処理は終了し、その後、処理は図8のステップS18に進む。

30

【0173】

図8のフローチャートの説明に戻り、ステップS18において、子音強調部28は、子音強調レベル計算部27から供給された強調量を、時間周波数変換部21から供給された入力スペクトルに乗算して入力スペクトルを強調し、周波数時間変換部29に供給する。

【0174】

より具体的には、子音強調部28は、子音強調レベル計算部27から供給された比較結果に基づいて、入力スペクトルのうち、背景ノイズが優位であるとされた周波数帯域以外の周波数帯域に強調量を乗算する。

40

【0175】

なお、現フレームが子音フレームではない場合には、入力スペクトルの強調は行なわれず、子音強調部28は、時間周波数変換部21から供給された入力スペクトルをそのまま周波数時間変換部29に供給する。

【0176】

ステップS19において、周波数時間変換部29は、子音強調部28から供給された入力スペクトルに対して周波数時間変換を行なって、入力スペクトルを時間信号である出力信号に変換し、出力する。出力信号が出力されると、子音強調処理は終了する。

【0177】

以上のようにして、子音強調装置11は、背景ノイズを抑圧したノイズ抑圧スペクトル

50

を求め、ノイズ抑圧スペクトルから得られた特徴量に基づいて周波数領域で子音を検出するとともに、その検出結果に応じて子音フレームを強調する。

【0178】

このようにノイズ抑圧スペクトルを用いて周波数領域で子音を検出することで、より高精度に子音を検出することができる。また、周波数領域において子音フレームの背景ノイズが優位な周波数帯域以外の周波数帯域を強調することで、より効果的に音声の強調を行なうことができる。

【0179】

第1の実施の形態の変形例1

子音強調装置の構成例

なお、以上においては、入力スペクトルに基づいて強調量が算出されると説明したが、入力信号に基づいて、時間領域で強調量が算出されるようにしてもよい。

【0180】

そのような場合、子音強調装置11は、例えば図11に示すように構成される。なお、図11において、図1における場合と対応する部分には同一の符号を付してあり、その説明は適宜省略する。

【0181】

図11に示す子音強調装置11は、供給された入力信号が子音強調レベル計算部27にも供給される点で図1の子音強調装置11と異なり、その他の点では図1の子音強調装置11と同じ構成とされている。

【0182】

図11に示す子音強調装置11では、子音強調レベル計算部27は、供給された入力信号に基づいて、時間領域で母音部パワーや、子音フレームであるとされた現フレームの入力信号のパワーを算出する。したがって、式(5)に示した強調量は、時間信号である入力信号から算出されることになる。なお、例えば入力信号のパワーは、RMS (Root Mean Square) などとすればよい。

【0183】

また、時間周波数変換部21は、時間周波数変換により得られた入力スペクトルを背景ノイズ推定部22、ノイズ抑圧スペクトル計算部23、および子音強調部28に供給する。

【0184】

第1の実施の形態の変形例2

子音強調装置の構成例

さらに、以上においてはノイズ抑圧スペクトルに基づいて、入力信号のピッチ強度が算出される例について説明したが、入力信号に基づいて時間領域でピッチ強度が算出されるようにしてもよい。

【0185】

そのような場合、子音強調装置11は、例えば図12に示すように構成される。なお、図12において、図1における場合と対応する部分には同一の符号を付してあり、その説明は適宜省略する。

【0186】

図12に示す子音強調装置11は、供給された入力信号がピッチ強度計算部24にも供給される点で図1の子音強調装置11と異なり、その他の点では図1の子音強調装置11と同じ構成とされている。

【0187】

図12に示す子音強調装置11では、ピッチ強度計算部24は、供給された時間信号である入力信号の自己相関を求めることでピッチ強度を算出し、特徴量計算部25および子音検出部26に供給する。つまり、ピッチ強度計算部24では、時間領域において、入力信号に基づいて算出された自己相関係数値が、そのままピッチ強度とされる。

【0188】

10

20

30

40

50

また、ノイズ抑圧スペクトル計算部 23 は、ノイズ抑圧により得られたノイズ抑圧スペクトルを特徴量計算部 25、および子音強調レベル計算部 27 に供給する。

【0189】

第 1 の実施の形態の変形例 3

子音強調装置の構成例

さらに、強調量とピッチ強度の両方が時間領域で算出されるようにしてもよい。そのような場合、子音強調装置 11 は、例えば図 13 に示すように構成される。なお、図 13 において、図 1 における場合と対応する部分には同一の符号を付してあり、その説明は適宜省略する。

【0190】

図 13 に示す子音強調装置 11 は、供給された入力信号が時間周波数変換部 21 の他に、ピッチ強度計算部 24、および子音強調レベル計算部 27 にも供給される点で図 1 の子音強調装置 11 と異なり、その他の点では図 1 の子音強調装置 11 と同じ構成とされている。

【0191】

図 13 に示す子音強調装置 11 では、時間周波数変換部 21 は、時間周波数変換により得られた入力スペクトルを背景ノイズ推定部 22、ノイズ抑圧スペクトル計算部 23、および子音強調部 28 に供給する。

【0192】

ピッチ強度計算部 24 は、供給された時間信号である入力信号に基づいてピッチ強度を算出し、特徴量計算部 25 および子音検出部 26 に供給する。また、ノイズ抑圧スペクトル計算部 23 は、ノイズ抑圧により得られたノイズ抑圧スペクトルを特徴量計算部 25、および子音強調レベル計算部 27 に供給する。

【0193】

さらに、子音強調レベル計算部 27 は、供給された入力信号に基づいて母音部パワーや、子音フレームであるとされた現フレームの入力信号のパワーを算出する。つまり、強調量が時間領域で計算される。

【0194】

第 2 の実施の形態

子音検出装置の構成例

さらに、以上においては、入力信号から子音部分を検出し、子音のスペクトルを強調する子音強調装置に本技術を適用する例について説明したが、本技術は、入力信号から子音フレームを検出する子音検出装置に適用されるようにしてもよい。

【0195】

そのような場合、子音検出装置は、例えば図 14 に示すように構成される。なお、図 14 において、図 1 における場合と対応する部分には同一の符号を付してあり、その説明は適宜省略する。

【0196】

図 14 の子音検出装置 61 は、時間周波数変換部 21、背景ノイズ推定部 22、ノイズ抑圧スペクトル計算部 23、ピッチ強度計算部 24、特徴量計算部 25、および子音検出部 26 から構成される。

【0197】

時間周波数変換部 21 は、供給された入力信号に対して時間周波数変換を行い、その結果得られた入力スペクトルを背景ノイズ推定部 22 およびノイズ抑圧スペクトル計算部 23 に供給する。

【0198】

背景ノイズ推定部 22 は、時間周波数変換部 21 からの入力スペクトルに基づいて背景ノイズ推定を行い、その結果得られた背景ノイズスペクトルをノイズ抑圧スペクトル計算部 23 に供給する。ノイズ抑圧スペクトル計算部 23 は、時間周波数変換部 21 からの入力スペクトルと、背景ノイズ推定部 22 からの背景ノイズスペクトルとに基づいてノイズ

10

20

30

40

50

抑圧スペクトルを求め、特徴量計算部 2 5 に供給する。

【 0 1 9 9 】

ピッチ強度計算部 2 4 は、供給された時間信号である入力信号に基づいて、時間領域においてピッチ強度を算出し、特徴量計算部 2 5 および子音検出部 2 6 に供給する。

【 0 2 0 0 】

特徴量計算部 2 5 は、ノイズ抑圧スペクトル計算部 2 3 からのノイズ抑圧スペクトルに基づいて、またはノイズ抑圧スペクトルと、ピッチ強度計算部 2 4 からのピッチ強度とに基づいて特徴量を算出し、子音検出部 2 6 に供給する。

【 0 2 0 1 】

子音検出部 2 6 は、ピッチ強度計算部 2 4 からのピッチ強度と、特徴量計算部 2 5 からの特徴量とに基づいて、入力信号の子音区間を検出し、その検出結果を後段に出力する。すなわち、子音検出部 2 6 では、例えば図 9 のフローチャートを参照して説明した子音検出処理と同様の処理が行なわれる。

10

【 0 2 0 2 】

このように子音検出装置 6 1 においても、子音強調装置 1 1 における場合と同様に、より精度よく入力信号から子音を検出することができる。

【 0 2 0 3 】

第 2 の実施の形態の変形例 1

子音検出装置の構成例

また、図 1 4 に示した子音検出装置 6 1 では、ピッチ強度が時間領域で求められる例について説明したが、ピッチ強度が周波数領域で求められるようにしてもよい。

20

【 0 2 0 4 】

そのような場合、子音検出装置 6 1 は、例えば図 1 5 に示すように構成される。なお、図 1 5 において、図 1 4 における場合と対応する部分には同一の符号を付してあり、その説明は適宜省略する。

【 0 2 0 5 】

図 1 5 に示す子音検出装置 6 1 は、入力信号が時間周波数変換部 2 1 のみに供給され、ノイズ抑圧スペクトル計算部 2 3 からピッチ強度計算部 2 4 にノイズ抑圧スペクトルが供給される点で図 1 4 の子音検出装置 6 1 と異なり、その他の点では図 1 4 の子音検出装置 6 1 と同じ構成とされている。

30

【 0 2 0 6 】

ノイズ抑圧スペクトル計算部 2 3 は、背景ノイズの抑圧により得られたノイズ抑圧スペクトルをピッチ強度計算部 2 4、および特徴量計算部 2 5 に供給する。

【 0 2 0 7 】

ピッチ強度計算部 2 4 は、ノイズ抑圧スペクトル計算部 2 3 から供給されたノイズ抑圧スペクトルに基づいて、周波数領域で入力信号のピッチ強度を算出し、特徴量計算部 2 5 および子音検出部 2 6 に供給する。

【 0 2 0 8 】

ところで、上述した一連の処理は、ハードウェアにより実行することもできるし、ソフトウェアにより実行することもできる。一連の処理をソフトウェアにより実行する場合には、そのソフトウェアを構成するプログラムが、コンピュータにインストールされる。ここで、コンピュータには、専用のハードウェアに組み込まれているコンピュータや、各種のプログラムをインストールすることで、各種の機能を実行することが可能な、例えば汎用のパーソナルコンピュータなどが含まれる。

40

【 0 2 0 9 】

図 1 6 は、上述した一連の処理をプログラムにより実行するコンピュータのハードウェアの構成例を示すブロック図である。

【 0 2 1 0 】

コンピュータにおいて、CPU (Central Processing Unit) 3 0 1, ROM (Read Only Memory) 3 0 2, RAM (Random Access Memory) 3 0 3 は、バス 3 0 4 により相互に接続され

50

ている。

【0211】

バス304には、さらに、入出力インターフェース305が接続されている。入出力インターフェース305には、入力部306、出力部307、記録部308、通信部309、及びドライブ310が接続されている。

【0212】

入力部306は、キーボード、マウス、マイクロホン、撮像素子などよりなる。出力部307は、ディスプレイ、スピーカなどよりなる。記録部308は、ハードディスクや不揮発性のメモリなどよりなる。通信部309は、ネットワークインターフェースなどよりなる。ドライブ310は、磁気ディスク、光ディスク、光磁気ディスク、又は半導体メモリなどのリムーバブルメディア311を駆動する。

10

【0213】

以上のように構成されるコンピュータでは、CPU301が、例えば、記録部308に記録されているプログラムを、入出力インターフェース305及びバス304を介して、RAM303にロードして実行することにより、上述した一連の処理が行われる。

【0214】

コンピュータ(CPU301)が実行するプログラムは、例えば、パッケージメディア等としてのリムーバブルメディア311に記録して提供することができる。また、プログラムは、ローカルエリアネットワーク、インターネット、デジタル衛星放送といった、有線または無線の伝送媒体を介して提供することができる。

20

【0215】

コンピュータでは、プログラムは、リムーバブルメディア311をドライブ310に装着することにより、入出力インターフェース305を介して、記録部308にインストールすることができる。また、プログラムは、有線または無線の伝送媒体を介して、通信部309で受信し、記録部308にインストールすることができる。その他、プログラムは、ROM302や記録部308に、あらかじめインストールしておくことができる。

【0216】

なお、コンピュータが実行するプログラムは、本明細書で説明する順序に沿って時系列に処理が行われるプログラムであっても良いし、並列に、あるいは呼び出しが行われたとき等の必要なタイミングで処理が行われるプログラムであっても良い。

30

【0217】

また、本技術の実施の形態は、上述した実施の形態に限定されるものではなく、本技術の要旨を逸脱しない範囲において種々の変更が可能である。

【0218】

例えば、本技術は、1つの機能をネットワークを介して複数の装置で分担、共同して処理するクラウドコンピューティングの構成をとることができる。

【0219】

また、上述のフローチャートで説明した各ステップは、1つの装置で実行する他、複数の装置で分担して実行することができる。

【0220】

さらに、1つのステップに複数の処理が含まれる場合には、その1つのステップに含まれる複数の処理は、1つの装置で実行する他、複数の装置で分担して実行することができる。

40

【0221】

さらに、本技術は、以下の構成とすることも可能である。

【0222】

[1]

入力信号の背景ノイズを推定する背景ノイズ推定部と、  
前記背景ノイズの推定結果に基づいて、前記入力信号の前記背景ノイズを抑圧するノイズ抑圧部と、

50

- 前記背景ノイズが抑圧された前記入力信号に基づいて特徴量を算出する特徴量計算部と、
- 前記特徴量に基づいて、前記入力信号から子音を検出する子音検出部とを備える音声処理装置。
- [ 2 ]  
前記背景ノイズ推定部は、周波数領域で前記背景ノイズを推定し、  
前記ノイズ抑圧部は、前記入力信号から得られた入力スペクトルに含まれる前記背景ノイズを抑圧することでノイズ抑圧スペクトルを求め、  
前記特徴量計算部は、前記ノイズ抑圧スペクトルに基づいて前記特徴量を算出する  
[ 1 ]に記載の音声処理装置。 10
- [ 3 ]  
前記背景ノイズ推定部は、過去の前記入力スペクトルの平均値を求めることで前記背景ノイズを推定する  
[ 2 ]に記載の音声処理装置。
- [ 4 ]  
前記ノイズ抑圧スペクトルの値が、前記背景ノイズの推定により得られた背景ノイズスペクトルを定数倍した値よりも大きい周波数について、前記入力スペクトルの強調を行なう子音強調部をさらに備える  
[ 2 ]または[ 3 ]に記載の音声処理装置。
- [ 5 ] 20  
前記子音強調部は、予め定められた強調量で前記入力スペクトルの強調を行なう  
[ 4 ]に記載の音声処理装置。
- [ 6 ]  
前記入力信号の過去の母音部分のパワーの平均値と、前記入力信号の現在のパワーとの比に基づいて強調量を算出する子音強調レベル計算部をさらに備え、  
前記子音強調部は、前記強調量で前記入力スペクトルの強調を行なう  
[ 4 ]に記載の音声処理装置。
- [ 7 ]  
周波数方向に対して前記強調量の補間が行なわれる  
[ 5 ]または[ 6 ]に記載の音声処理装置。 30
- [ 8 ]  
前記ノイズ抑圧部は、スペクトルサブトラクション法により前記ノイズ抑圧スペクトルを求める  
[ 2 ]乃至[ 7 ]の何れかに記載の音声処理装置。
- [ 9 ]  
前記入力信号のピッチ強度がさらに前記特徴量として用いられ、  
前記子音検出部は、前記特徴量としての前記ピッチ強度と、前記ノイズ抑圧スペクトルに基づいて算出された前記特徴量とに基づいて、前記入力信号から子音を検出する  
[ 2 ]乃至[ 8 ]の何れかに記載の音声処理装置。
- [ 10 ] 40  
前記ピッチ強度は、前記ノイズ抑圧スペクトルのピークが、ピッチ周波数位置と、ピッチ周波数の倍音周波数位置とに発生している度合いで表される  
[ 9 ]に記載の音声処理装置。
- [ 11 ]  
前記ピッチ強度は、前記入力信号の自己相関係数値である  
[ 9 ]に記載の音声処理装置。
- [ 12 ]  
前記特徴量計算部は、前記ノイズ抑圧スペクトルの周波数帯域を複数のバンドに分割し、前記バンドにおける前記ノイズ抑圧スペクトルの代表値に基づいて前記特徴量を算出する 50

[ 2 ] 乃至 [ 1 1 ] の何れかに記載の音声処理装置。

[ 1 3 ]

前記ノイズ抑圧スペクトルはパワースペクトルである

[ 1 2 ] に記載の音声処理装置。

[ 1 4 ]

前記ノイズ抑圧スペクトルは振幅スペクトルである

[ 1 2 ] に記載の音声処理装置。

[ 1 5 ]

前記代表値は、前記バンドにおける前記ノイズ抑圧スペクトルの平均値である

[ 1 2 ] 乃至 [ 1 4 ] の何れかに記載の音声処理装置。

10

[ 1 6 ]

前記代表値は、前記バンドにおける前記ノイズ抑圧スペクトルの最大値である

[ 1 2 ] 乃至 [ 1 4 ] の何れかに記載の音声処理装置。

[ 1 7 ]

前記特徴量計算部は、前記ノイズ抑圧スペクトルにおける前記バンドの前記代表値の時間差分値を前記特徴量として算出する

[ 1 2 ] 乃至 [ 1 6 ] の何れかに記載の音声処理装置。

【符号の説明】

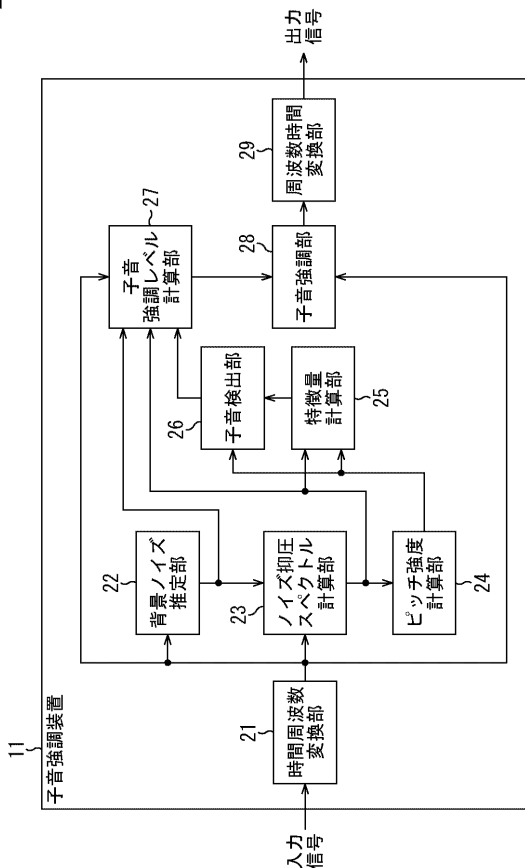
【 0 2 2 3 】

11 子音強調装置, 21 時間周波数変換部, 22 背景ノイズ推定部, 23 ノイズ抑圧スペクトル計算部, 24 ピッチ強度計算部, 25 特徴量計算部, 26 子音検出部, 27 子音強調レベル計算部, 28 子音強調部, 29 周波数時間変換部, 29 周波数時間変換部, 61 子音検出装置

20

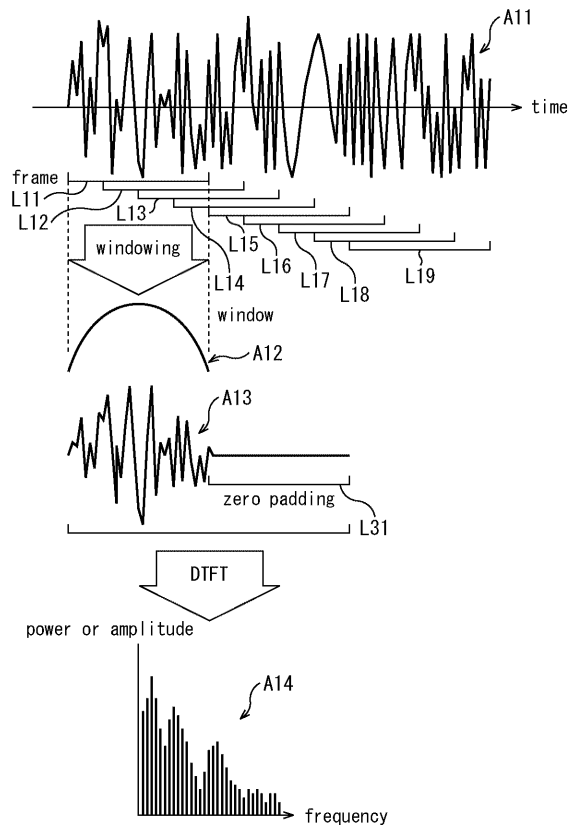
【 図 1 】

図1



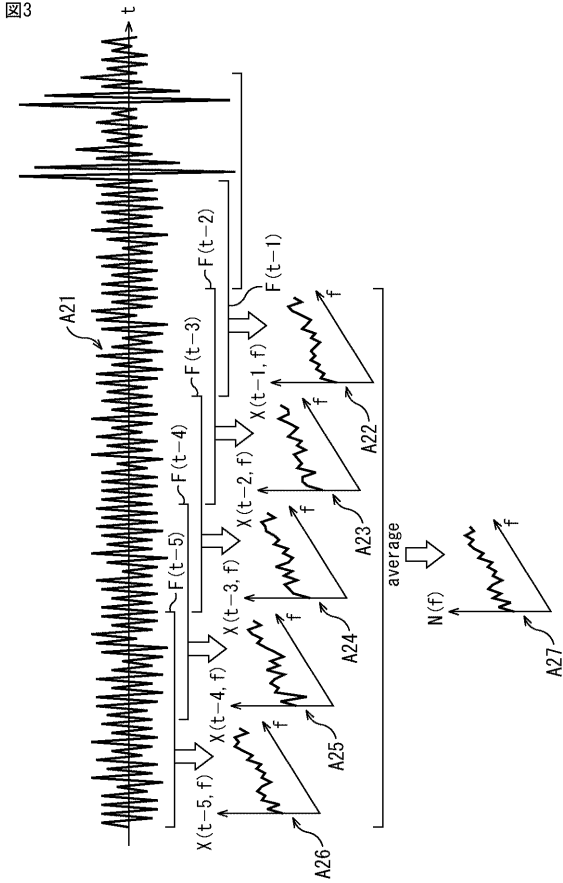
【 図 2 】

図2

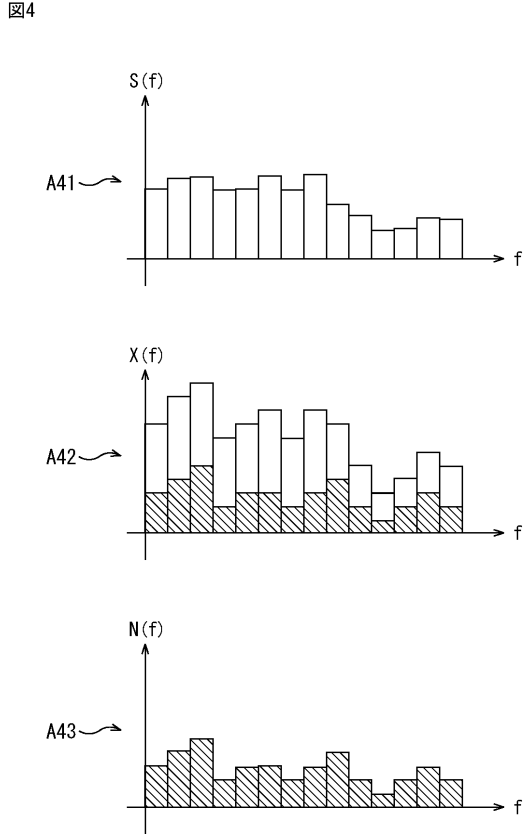




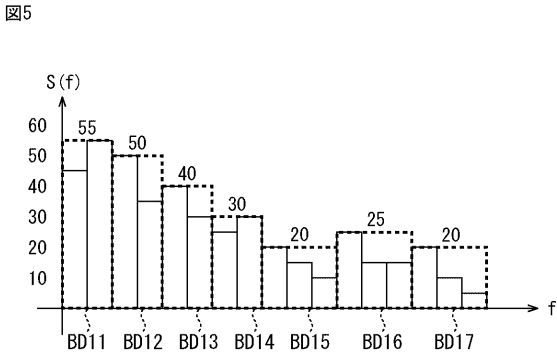
【 図 3 】



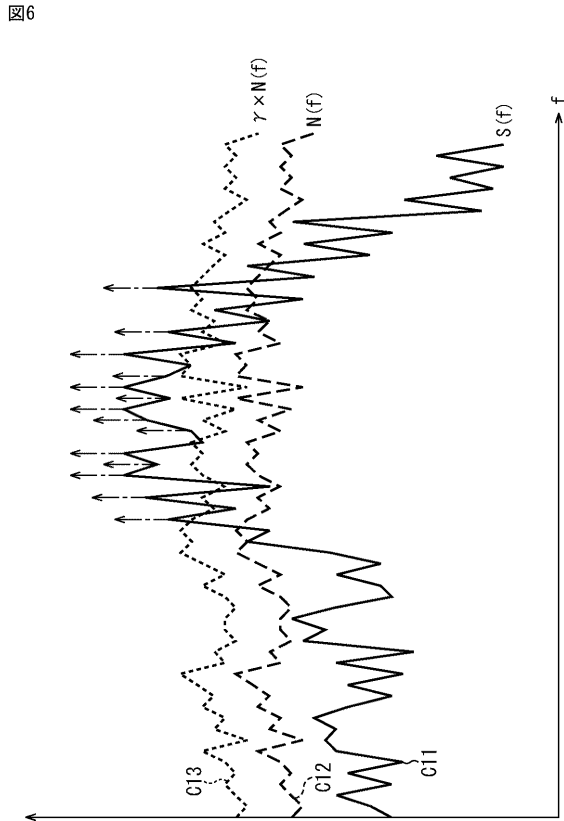
【 図 4 】



【 図 5 】

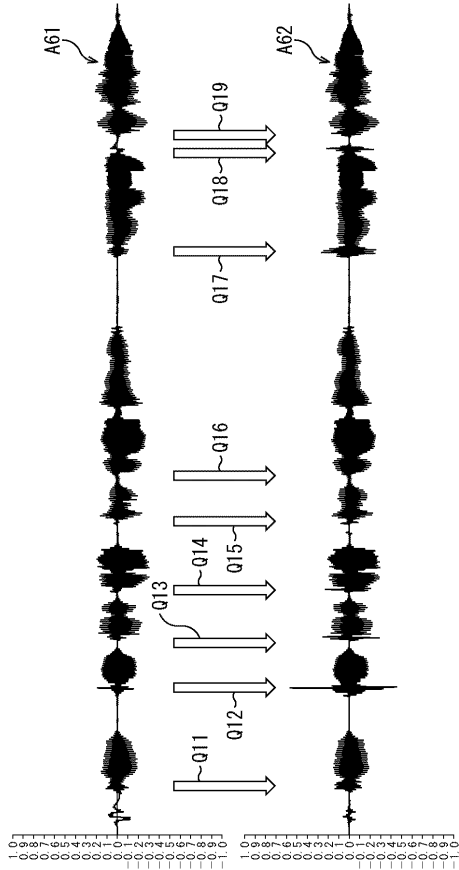


【 図 6 】



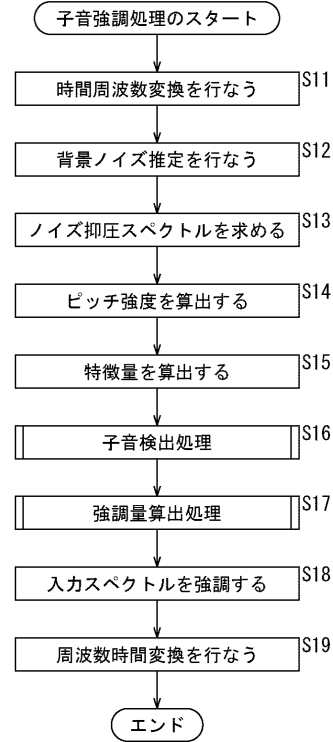
【図7】

図7



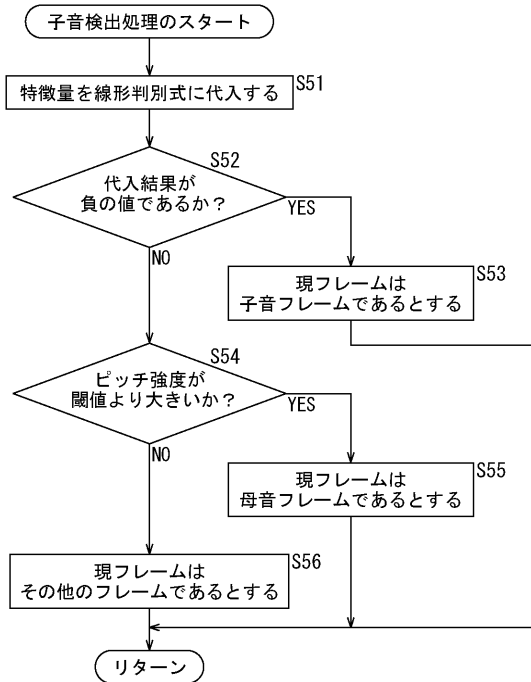
【図8】

図8



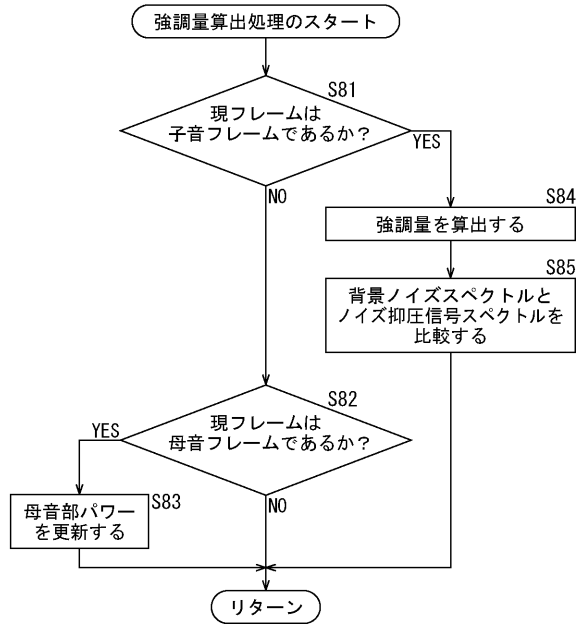
【図9】

図9

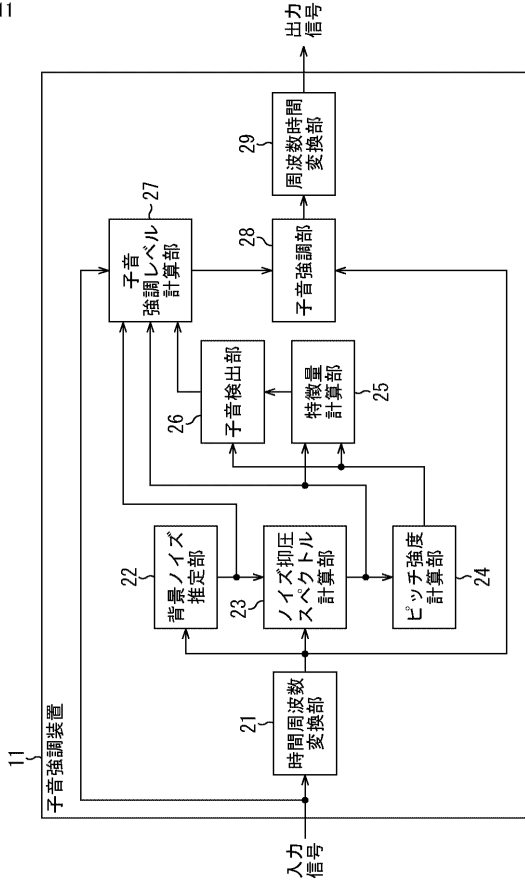


【図10】

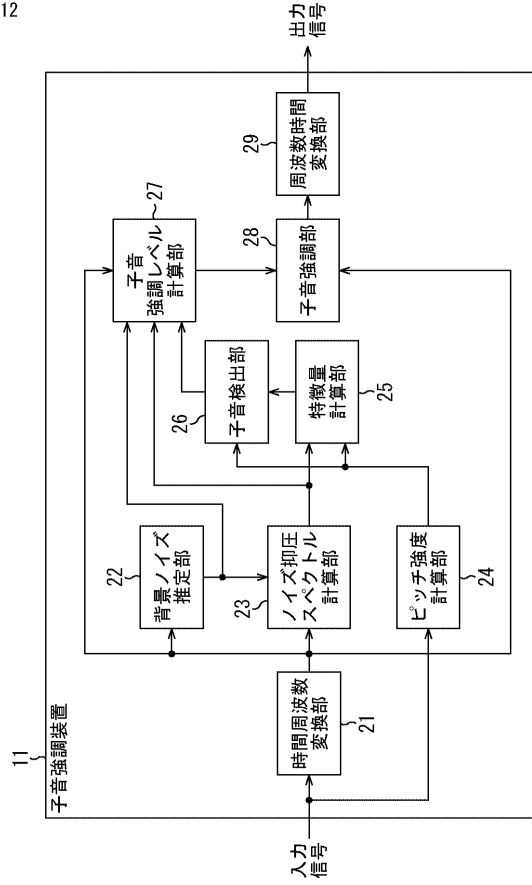
図10



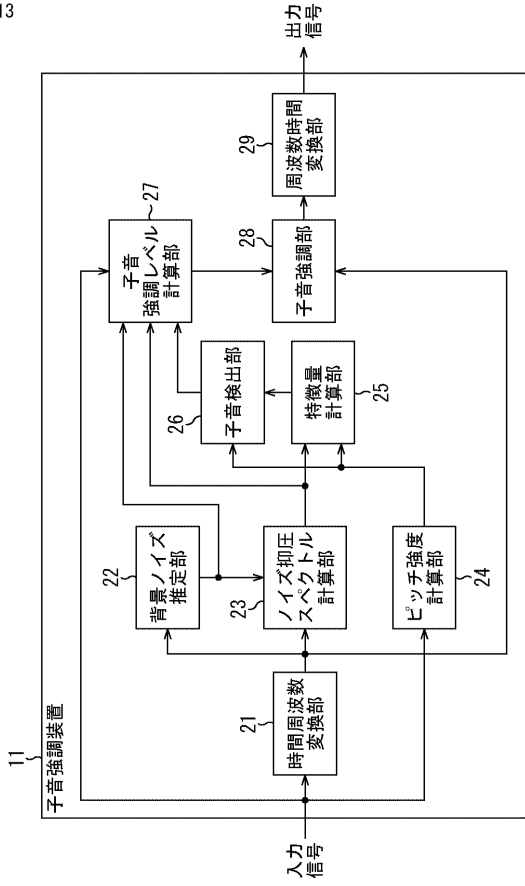
【図 1 1】  
図 11



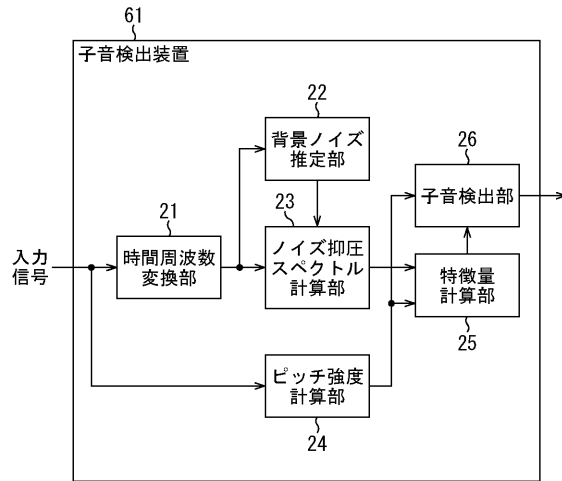
【図 1 2】  
図 12



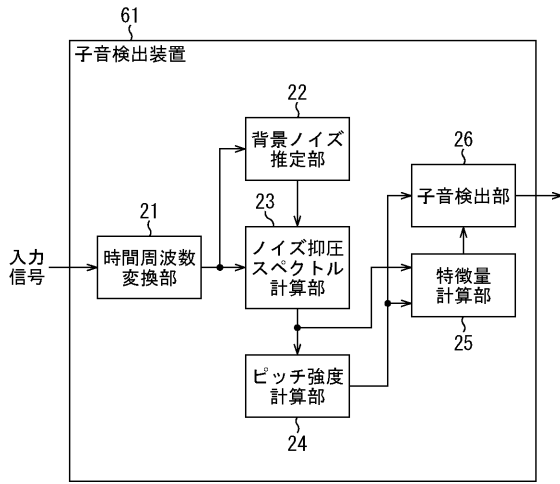
【図 1 3】  
図 13



【図 1 4】  
図 14



【図15】  
図15



【図16】  
図16

