

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平11-305792

(43) 公開日 平成11年(1999)11月5日

|                           |       |              |
|---------------------------|-------|--------------|
| (51) Int.Cl. <sup>6</sup> | 識別記号  | F I          |
| G 1 0 L 3/02              | 3 0 1 | C 1 0 L 3/02 |
| 9/00                      |       | 9/00         |
|                           |       | 3 0 1 E      |
|                           |       | F            |

審査請求 未請求 請求項の数37 O L (全 22 頁)

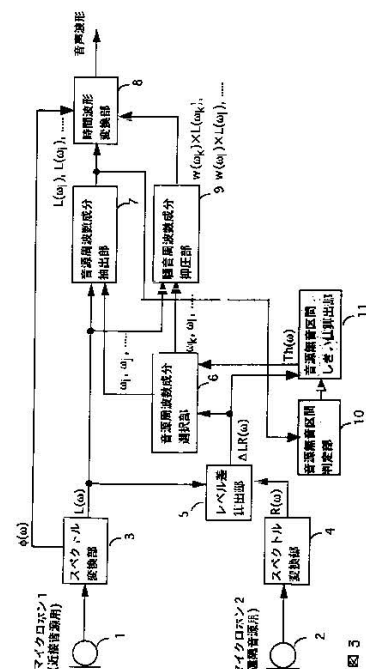
|              |                    |          |                               |
|--------------|--------------------|----------|-------------------------------|
| (21) 出願番号    | 特願平10-232565       | (71) 出願人 | 000004226                     |
| (22) 出願日     | 平成10年(1998) 8 月19日 |          | 日本電信電話株式会社                    |
| (31) 優先権主張番号 | 特願平10-39206        |          | 東京都千代田区大手町二丁目3番1号             |
| (32) 優先日     | 平10(1998) 2 月20日   | (72) 発明者 | 高野 智大                         |
| (33) 優先権主張国  | 日本 (J P)           |          | 東京都新宿区西新宿三丁目19番2号 日本電信電話株式会社内 |
|              |                    | (72) 発明者 | 松井 弘行                         |
|              |                    |          | 東京都新宿区西新宿三丁目19番2号 日本電信電話株式会社内 |
|              |                    | (74) 代理人 | 弁理士 草野 卓 (外1名)                |

(54) 【発明の名称】 収音装置、音声認識装置、これらの方法、及びプログラム記録媒体

(57) 【要約】

【課題】 正しく騒音を推定でき、かつ非定常な騒音も抑圧可能とし、更に音声認識率を向上させる。

【解決手段】 目的信号音源に近いマイク1と、これより遠いマイク2の各出力をスペクトルにそれぞれ変換し(3、4)、その各対応成分 $L(\omega_i)$ と $R(\omega_i)$ とのレベル差 $\Delta LR(\omega_i)$ を検出し(5)、 $\Delta LR(\omega_i)$ がしきい値 $Th(\omega_i)$ 以上の $\omega_i$ を選択し(6)、 $L(\omega)$ からその成分 $L(\omega_i)$ を抽出し、これを時間波形に変換する(8)、 $\Delta LR(\omega_i)$ が $Th(\omega_i)$ 以下の成分 $L(\omega_k)$ に対し、重み $w(\omega_k)$ を与えて(9)、時間波形に変換し、音源無音区間を判定し(10)、その時 $\Delta LR(\omega) > Th(\omega)$ であれば、その $\Delta LR(\omega)$ を $Th(\omega)$ とする(11)。変換された時間波形を蓄積し、音声区間を骨導マイクで検出して、その区間のみ蓄積波形に対し音声認識させる。



**【特許請求の範囲】**

【請求項1】 目的信号の音源に近い位置に設置された第1マイクロホンと、

前記位置より前記目的信号の音源から離れた位置に設置された第2マイクロホンと、

前記第1、第2マイクロホンの各々の出力信号を振幅スペクトルと位相スペクトルに変換する第1、第2スペクトル変換手段と、

前記第1、第2スペクトル変換手段から出力される、各対応する周波数成分ごとの振幅スペクトルのレベル差を計算するレベル差算出手段と、

前記レベル差算出手段より出力される各周波数成分ごとのレベル差と、予め設定されたしきい値とを比較し、前記レベル差がしきい値よりも大きい周波数成分を目的信号が支配的な周波数成分と判定する音源周波数成分選択手段と、

前記第1マイクロホンの出力信号の振幅スペクトルから、前記判定にもとづき前記目的信号が支配的と判定された周波数成分を抽出する音源周波数成分抽出手段と、前記音源周波数成分抽出手段より抽出された振幅スペクトルを、前記第1スペクトル変換手段により算出される位相スペクトルを用いて時間波形に変換する時間波形変換手段を有することを特徴とする收音装置。

【請求項2】 請求項1記載の收音装置において、前記音源周波数成分選択手段において目的信号が支配的と判定されなかった周波数成分の前記第1スペクトル変換手段の出力周波数成分に対して予め設定された重み係数を乗ずる騒音周波数成分抑圧手段を具備し、前記時間波形変換手段においては、前記音源周波数成分抽出手段および前記騒音周波数成分抑圧手段により処理された周波数成分を、前記スペクトル変換手段で算出される対応周波数の位相スペクトルを用いて時間波形に変換することを特徴とする收音装置。

【請求項3】 請求項1または請求項2記載の收音装置において、

前記音源周波数成分抽出手段の出力振幅スペクトルの大きさと、予め設定された無音区間判定しきい値とを比較し、前記振幅スペクトルが前記無音区間判定しきい値よりも小さいとき目的信号の音源が無音状態であると判定する音源無音区間判定手段と、

前記音源無音区間判定手段により目的信号の音源が無音状態と判定された場合において、前記音源周波数成分選択手段に用いるしきい値を、前記レベル差算出手段より出力されるレベル差以上に更新するしきい値算出手段を具備することを特徴とする收音装置。

【請求項4】 請求項2または請求項3記載の收音装置において、

前記騒音周波数成分抑圧手段において目的信号が支配的と判定されなかった周波数成分に乗ずる重み係数を、前記レベル差算出手段よりの各周波数成分のレベル差と前

記音源周波数成分選択手段のしきい値の大小関係を比較して算出する重み係数算出手段を具備することを特徴とする收音装置。

【請求項5】 請求項3または請求項4記載の收音装置において、

前記音源無音区間判定手段により目的信号の音源が無音状態と判定された場合において、前記音源周波数成分抽出手段と前記騒音周波数成分抑圧手段の出力または、前記時間波形変換手段の出力を減衰させる音源無音区間減衰手段を具備することを特徴とする收音装置。

【請求項6】 請求項1、請求項2、請求項3、請求項4、または請求項5記載の收音装置において、

前記第1マイクロホンと前記第2マイクロホンのうち、前者のマイクロホンが後者のマイクロホンに比べて使用状態において、その使用者の口元に近い位置になるようにハンドセット、ヘッドセット、イヤーマイクセットなどに組み込まれていることを特徴とする收音装置。

【請求項7】 目的信号の音源に近い位置に第1マイクロホンを設置し、

前記位置より前記目的信号の音源から離れた位置に第2マイクロホンを設置し、

前記第1、第2マイクロホンの各々の出力信号を第1、第2振幅スペクトルと第1、第2位相スペクトルにそれぞれ変換し、

前記第1、第2振幅スペクトルのレベル差を、各対応する周波数成分ごとに計算し、

前記計算された各周波数成分ごとのレベル差と、予め設定されたしきい値とを比較し、前記レベル差がしきい値よりも大きい周波数成分を目的信号が支配的な周波数成分と判定し、

その判定にもとづき前記第1振幅スペクトルから前記判定にもとづき、前記目的信号が支配的と判定された周波数成分を抽出し、

前記抽出された周波数成分を、前記第1位相スペクトルを用いて時間波形に変換することを特徴とする收音方法。

【請求項8】 請求項7記載の收音方法において、前記しきい値との比較において目的信号が支配的と判定されなかった周波数成分の前記第1振幅スペクトルに対して予め設定された重み係数を乗じ、

前記時間波形変換時に、前記重み係数が乗じられた周波数成分に対しても対応周波数の第1位相スペクトルを用いて時間波形に変換することを特徴とする收音方法。

【請求項9】 請求項7または請求項8記載の收音方法において、

前記目的信号が支配的と判定された周波数成分の第1振幅スペクトルの大きさと、予め設定された無音区間判定しきい値とを比較し、前記第1振幅スペクトルが前記無音区間判定しきい値よりも小さいとき目的信号の音源が無音状態であると判定し、

目的信号の音源が無音状態と判定された場合において、前記目的信号が支配的な周波数成分か否かの判定に用いるしきい値を、前記レベル差以上に更新することを特徴とする收音方法。

【請求項10】 請求項8または請求項9記載の收音方法において、前記目的信号が支配的と判定されなかった周波数成分に乗ずる重み係数を、前記各周波数成分のレベル差と、前記レベル差と比較するしきい値の大小関係とを比較して算出することを特徴とする收音方法。

【請求項11】 請求項9または請求項10記載の收音方法において、前記目的信号の音源が無音状態と判定された場合において、前記目的信号が支配的と判定されて抽出された周波数成分と、前記重み係数が乗じられた目的信号が支配的と判定されなかった周波数成分、または、前記時間波形変換された信号を減衰させることを特徴とする收音方法。

【請求項12】 請求項7、請求項8、請求項9、請求項10、または請求項11記載の收音方法において、前記第1マイクロホンと前記第2マイクロホンのうち、前者のマイクロホンが後者のマイクロホンに比べて使用状態において、その使用者の口元に近い位置になるようにハンドセット、ヘッドセット、イヤーマイクセットなどに組み込むことを特徴とする收音方法。

【請求項13】 目的信号の音源に近い位置に設置された第1マイクロホンと前記位置より前記目的信号の音源から離れた位置に設置された第2マイクロホンとの各々の出力信号をそれぞれ第1、第2振幅スペクトルと第1、第2位相スペクトルに変換する処理と、前記第1、第2振幅スペクトルのレベル差を各対応する周波数成分ごとに計算する処理と、前記計算されたレベル差を各周波数成分ごとに、予め設定されたしきい値とを比較し、前記レベル差がしきい値よりも大きい周波数成分を目的信号が支配的な周波数成分と判定する第1判定処理と、前記第1振幅スペクトルから、前記第1判定処理にもとづき前記目的信号が支配的と判定された周波数成分を抽出する処理と、前記抽出された第1振幅スペクトルを、前記第1位相スペクトルを用いて時間波形に変換する時間波形変換処理とを收音装置のコンピュータに実行させるプログラムを記録した記録媒体。

【請求項14】 請求項13記載の記録媒体において、前記第1判定処理において目的信号が支配的と判定されなかった周波数成分の前記第1振幅スペクトルに対して予め設定された重み係数を乗ずる騒音抑圧処理と、前記時間波形変換処理においては、前記重み係数が乗じられた周波数成分に対しても対応周波数の第1位相スペクトルを用いて時間波形に変換することを前記コンピュ

ータに行わせるプログラムを前記プログラムが有することを特徴とする記録媒体。

【請求項15】 請求項13または請求項14記載の記録媒体において、前記目的信号が支配的と判定された周波数成分の第1振幅スペクトルの大きさと、予め設定された無音区間判定しきい値とを比較し、前記第1振幅スペクトルが前記無音区間判定しきい値よりも小さいとき目的信号の音源が無音状態であると判定する第2判定処理と、前記第2判定処理により目的信号の音源が無音状態と判定された場合において、前記第1判定処理に用いるしきい値を、前記レベル差以上に更新する処理とを前記コンピュータに行わせるプログラムを前記プログラムが有することを特徴とする記録媒体。

【請求項16】 請求項14または請求項15記載の記録媒体において、前記騒音抑圧処理で乗ずる重み係数を、前記各周波数成分のレベル差と前記第1判定処理で用いるしきい値の大小関係を比較して算出する処理と前記コンピュータに行わせるプログラムを前記プログラムが含むことを特徴とする記録媒体。

【請求項17】 請求項15または請求項16記載の記録媒体において、前記第2判定処理により目的信号の音源が無音状態と判定された場合に、前記目的信号が支配的と判定されて抽出された周波数成分と前記騒音抑圧処理の出力または、前記時間波形変換処理の出力を減衰させる処理を前記コンピュータに行わせるプログラムを前記プログラムが有することを特徴とする記録媒体。

【請求項18】 目的信号（音声信号）の音源に近い位置に設置された第1マイクロホンと、前記位置より目的信号の音源から離れた位置に設置された第2マイクロホンと、前記第1、第2マイクロホンの各々の出力信号を第1、第2振幅スペクトルと第1、第2位相スペクトルに変換する第1、第2スペクトル変換手段と、前記第1、第2スペクトル変換手段から出力される、各対応周波数成分ごとの第1、第2振幅スペクトルのレベル差を計算するレベル差算出手段と、前記レベル差算出手段より出力される各周波数成分ごとのレベル差と、予め設定されたしきい値とを比較し、前記レベル差がしきい値よりも大きい周波数成分を目的信号が支配的な周波数成分と判定する音源周波数成分選択手段と、前記第1マイクロホンの出力信号の第1振幅スペクトルから、前記判定にもとづき前記目的信号が支配的と判定された周波数成分を抽出する音源周波数成分抽出手段と、前記音源周波数成分抽出手段により抽出された第1振幅スペクトルを前記第1位相スペクトルを用いて時間波形

に変換する時間波形変換手段と、  
前記目的信号の音源が発声状態か無発声状態かのいずれかを判定する発声状態判定手段と、  
前記時間波形変換手段の出力を蓄積する音声蓄積手段と、  
前記音声蓄積手段の出力を利用して音声認識を行う音声認識処理手段と、  
前記発声状態判定手段の出力結果を利用して、発声区間の音声データのみを前記音声認識処理手段に入力する制御、及び前記音声認識処理手段で処理された認識結果を出力する制御を行う音声認識制御手段とを有することを特徴とする音声認識装置。

【請求項19】 請求項18記載の音声認識装置において、  
前記音源周波数成分選択手段において目的信号が支配的と判定されなかった周波数成分に対して予め設定された重み係数を乗ずる騒音周波数成分抑圧手段を具備し、  
前記時間波形変換手段においては、前記音源周波数成分抽出手段および前記騒音周波数成分抑圧手段により処理された周波数成分を前記第1位相スペクトルを用いて時間波形に変換することを特徴とする音声認識装置。

【請求項20】 請求項18または請求項19記載の音声認識装置において、  
前記発声状態検出手段により目的信号の音源が無音状態と判定された場合において、前記レベル差算出手段より出力されるレベル差以上となるように、前記音源周波数成分選択手段に用いるしきい値を算出し更新するしきい値算出手段を具備することを特徴とする音声認識装置。

【請求項21】 請求項19または請求項20記載の音声認識装置において、  
前記騒音周波数成分抑圧手段において目的信号が支配的と判定されなかった周波数成分に乗ずる重み係数を、前記レベル差算出手段よりの各周波数成分のレベル差と前記音源周波数成分選択手段のしきい値の大小関係と比較して算出する重み係数算出手段を具備することを特徴とする音声認識装置。

【請求項22】 請求項18、19、20、または21記載の音声認識装置において、  
前記第1マイクロホンと前記第2マイクロホンのうち、前者のマイクロホンが後者のマイクロホンに比べて使用状態で、その使用者の口元に近い位置になるようにハンドセット、ヘッドセット、イヤーマイクセットに組み込まれたことを特徴とする音声認識装置。

【請求項23】 請求項18乃至22の何れかの音声認識装置において、  
前記発声状態判定手段は、目的信号の音源の発声状態を検出する発声検出手段を備え、その発声検出手段の出力結果により、前記目的信号の音源が発声状態か無発声状態かのいずれかを判定することを特徴とする音声認識装置。

【請求項24】 請求項18乃至22の何れかの音声認識装置において、  
前記発声状態判定手段は、前記音源周波数成分抽出手段の出力振幅スペクトルの大きさと、予め設定された無音区間判定しきい値とを比較し、前記振幅スペクトルが前記無音区間判定しきい値よりも小さいとき前記目的信号の音源が無発声状態であると判定する手段であることを特徴とする音声認識装置。

【請求項25】 目的信号（音声信号）の音源に近い位置に第1マイクロホンを設置し、  
前記位置より目的信号の音源から離れた位置に第2マイクロホンを設置し、  
前記第1、第2マイクロホンの各々の出力信号を第1、第2振幅スペクトルと第1、第2位相スペクトルにそれぞれ変換し、  
前記第1、第2振幅スペクトルのレベル差を、各対応する周波数成分ごとに計算し、  
前記計算された各周波数成分ごとのレベル差と、予め設定されたしきい値とを比較し、前記レベル差がしきい値よりも大きい周波数成分を目的信号が支配的な周波数成分と第1判定手段で判定し、  
前記第1振幅スペクトルから、前記第1判定手段の判定により、目的信号が支配的と判定された周波数成分を抽出し、  
前記目的信号が支配的な周波数成分の第1振幅スペクトルを前記第1位相スペクトルを用いて時間波形に時間波形変換手段により変換し、  
前記目的信号の音源が発声状態か無発声状態かのいずれかを判定し、  
前記時間波形変換手段の出力を音声蓄積手段に蓄積し、  
前記音声蓄積手段の出力を利用して音声認識を音声認識処理手段で行い、  
前記発声状態の音声データのみを前記音声認識処理手段に入力し、及び前記音声認識処理手段で処理された認識結果を出力することを特徴とする音声認識方法。

【請求項26】 請求項25記載の音声認識方法において、  
前記しきい値との比較において目的信号が支配的と判定されなかった周波数成分の第1振幅スペクトルに対して予め設定された重み係数を乗じ、  
前記時間波形変換手段において、前記重み係数が乗じられた周波数成分に対しても、対応周波数の前記第1位相スペクトルを用いて時間波形に変換することを特徴とする音声認識方法。

【請求項27】 請求項25または請求項26記載の音声認識方法において、  
前記目的信号の音源が無音状態と判定された場合において、前記レベル差以上となるように、前記第1判定手段に用いるしきい値を更新することを特徴とする音声認識方法。

【請求項28】 請求項26または請求項27記載の音声認識方法において、

前記目的信号が支配的と判定されなかった周波数成分に乗ずる重み係数を、各周波数成分のレベル差と、前記第1判定手段のしきい値の大小関係を比較して算出することを特徴とする音声認識方法。

【請求項29】 請求項25、26、27、または28記載の音声認識方法において、

前記第1マイクロホンと前記第2マイクロホンのうち、前者のマイクロホンが後者のマイクロホンに比べて使用状態で使用者の口元に近い位置になるようにハンドセット、ヘッドセット、イヤーマイクセットに組み込んで用いることを特徴とする音声認識方法。

【請求項30】 請求項25乃至29の何れかに記載の音声認識方法において、

前記発声状態判定は、目的信号の音源の発声状態を検出する発声検出手段を設け、その発声検出手段の出力結果により、前記目的信号の音源が発声状態か無発声状態かのいずれかを判定することを特徴とする音声認識方法。

【請求項31】 請求項25乃至29の何れかに記載の音声認識方法において、

前記発声状態判定は、前記第1判定手段の判定にもとづき抽出された目的信号が支配的と判定された第1振幅スペクトルの大きさと、予め設定された無音区間判定しきい値とを比較し、前記第1振幅スペクトルが前記無音区間判定しきい値よりも小さいとき前記目的信号の音源が無発声状態であると判定することを特徴とする音声認識方法。

【請求項32】 目的信号（音声信号）の音源に近い位置に設置された第1マイクロホンと、前記位置より目的信号の音源から離れた位置に設置された第2マイクロホンの各々の出力信号を第1、第2振幅スペクトルと第1、第2位相スペクトルにそれぞれ変換する処理と、各周波数ごとに前記第1、第2振幅スペクトルのレベル差を計算し、

前記計算されたレベル差と、予め設定されたしきい値とを比較し、前記レベル差がしきい値よりも大きい周波数成分を目的信号が支配的な周波数成分と判定する第1判定処理と、

前記第1振幅スペクトルから、前記第1判定処理にもとづき、目的信号が支配的と判定された周波数成分を抽出する処理と、

前記抽出された周波数成分の第1振幅スペクトルを、対応周波数の前記第1位相スペクトルを用いて時間波形に変換する時間波形変換処理と、

目的信号の音源が発声状態か無発声状態かのいずれかを判定する第2判定処理と、

前記時間波形変換処理の出力を音声蓄積手段に蓄積する処理と、

前記音声蓄積手段の出力を利用して音声認識処理手段に

より音声認識を行う処理と、

前記第2判定処理の結果を利用して、発声区間の音声データのみを前記音声認識処理手段に入力し、及び前記音声認識処理手段で処理された認識結果を出力する処理とをコンピュータにより実行させるプログラムを記録した記録媒体。

【請求項33】 請求項32記載の記録媒体において、前記第1判定処理において目的信号が支配的と判定されなかった周波数成分に対して予め設定された重み係数を乗ずる処理と、

前記時間波形変換処理においては、前記重み係数が乗じられた周波数成分に対しても、対応周波数の前記第1位相スペクトルを用いて時間波形に変換することを前記コンピュータに実行させるプログラムを前記プログラムが有することを特徴とする記録媒体。

【請求項34】 請求項32または請求項33記載の記録媒体において、

前記第2判定処理により目的信号の音源が無音状態と判定された場合において、前記レベル差以上となるように、前記第1判定処理に用いるしきい値を算出し更新する処理を前記コンピュータに実行させるプログラムを前記プログラムが有することを特徴とする記録媒体。

【請求項35】 請求項33または請求項34記載の記録媒体において、

前記目的信号が支配的と判定されなかった周波数成分に乗ずる重み係数を、前記レベル差と前記第1判定処理のしきい値の大小関係を比較して算出する処理を前記コンピュータに実行させるプログラムを前記プログラムが有することを特徴とする記録媒体。

【請求項36】 請求項32乃至35の何れかに記載の記録媒体において、

前記第2判定処理は、目的信号の音源の発声状態を検出する発声検出手段の出力結果により、前記目的信号の音源が発声状態か無発声状態かのいずれかを判定することを特徴とする記録媒体。

【請求項37】 請求項32乃至35の何れかに記載の記録媒体において、

前記第2判定処理は、前記第1判定処理の判定にもとづき抽出された目的信号が支配的と判定された第1振幅スペクトルの大きさと、予め設定された無音区間判定しきい値とを比較し、前記第1振幅スペクトルが前記無音区間判定しきい値よりも小さいとき前記目的信号の音源が無発声状態であると判定することを特徴とする記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】この発明は、周囲騒音が混在した音源信号に対し、周囲騒音成分を抑圧し、目的信号を抽出する收音装置、その收音機能を有し、高い認識率を得る音声認識装置、これらの方法、およびプログラム

記録媒体に関するものである。

【0002】

【従来の技術】騒音下で、SN比よく目的信号を抽出する従来技術として、騒音が重畳した目的信号の振幅スペクトルから騒音信号（推定値）の振幅スペクトルを差し引くスペクトルサブトラクション法がある。（例えば、“S. F. Boll, ‘Suppression of acoustic noise in speech using spectral subtraction’, IEEE Trans., ASSP, Vol. 27, No. 2, p113-120, (1979)”）

図20は、従来技術の一例を説明する流れ図である。まず、気導マイクロホン（音響信号を空気を媒体として受音する）の出力信号Lをスペクトルに変換する（S01）。つぎに、気導マイクロホンの出力信号Lの振幅スペクトルから推定された騒音の振幅スペクトルを差し引くことで騒音の除去されたスペクトル $S(\omega_i)$ を求める（S02）。そして、スペクトル $S(\omega_i)$ を気導マイクロホンの出力信号Lの位相スペクトルを用いて時間波形に変換して、騒音が抑圧された目的信号を得る（S03）。

【0003】上記の処理で騒音スペクトルを推定する方法として、例えば、①騒音収音用のマイクロホンを設置して騒音スペクトルを求める方法、②目的音源の無音区間から得た騒音スペクトルを用いる方法、が利用されている。しかし、これらの騒音推定方法には騒音スペクトル推定誤差を引き起こす以下のような問題点が含まれている。

【0004】

【発明が解決しようとする課題】①の方法では目的信号収音用マイクロホンと騒音収音用マイクロホンが配された位置が離れすぎていると、2つのマイクロホンに収音される騒音スペクトル特性に違いが生じてしまう。また、2つのマイクロホンの構成された位置が近すぎると騒音収音用マイクロホンに目的信号の寄与が含まれてしまう。すなわち、①の方法では正確な騒音スペクトルを推定するための2つのマイクロホン構成が取りにくいという問題点がある。

【0005】②の方法では、目的音源の有音区間において騒音のスペクトル特性が変化した際に生じる騒音スペクトルの推定誤差は免れない。特に、交通騒音など非定常性の強い騒音下では、騒音抑圧処理の良好な動作を期待することはできない。このように、現状のスペクトルサブトラクション法では正確な騒音スペクトルの推定が困難であるのが現状である。

【0006】騒音下で音声認識を行う場合、周囲騒音による音声信号の変形に対処する手段として、音声の入力段階で騒音除去処理を行うことが有効である。（例えば、“飛田、菅村、“音声認識における周囲騒音の影響”、音響学会誌、Vol. 51、No4、p331-33

5、(1995)”）

しかしながら、現状の騒音抑圧処理では前述のようにあらゆる環境で適用できる手法は確立されていない。

【0007】また、音声認識を行うためには、まず発声区間の検出を行う必要がある。しかしながら、騒音下においては、発声区間を正しく検出するのは困難である。前述のスペクトルサブトラクション法等で騒音除去を行った信号より、区間検出を行う方法が考えられるが、そのためには現状の問題点である、正確な騒音スペクトルの推定を実現しなければならない。

【0008】この発明の目的は、従来手法のように正確な騒音信号の推定を必要とせず、かつ非定常な騒音に対しても適用が可能な新しい騒音抑圧処理機能を有する収音装置及び前記騒音抑圧処理機能と騒音下における正しい発声区間検出機能を有する音声認識装置を提供することにある。

【0009】

【課題を解決するための手段】この発明の第1形態は、目的信号の音源に近い位置に設置された第1マイクロホンと、前記位置より目的信号の音源から離れた位置に設置された第2マイクロホンとが設けられ、これら第1、第2マイクロホンの各々の出力信号は、それぞれ第1、第2スペクトル変換手段により、振幅スペクトルと位相スペクトルに変換され、これら第1、第2スペクトル変換手段から出力される、各周波数成分ごとの振幅スペクトルについて、レベル差がレベル差算出手段で計算される。この各周波数成分ごとのレベル差と、予め設定されたしきい値とが音源周波数成分選択手段により比較され、前記レベル差がしきい値よりも大きい周波数成分は目的信号が、音源周波数成分選択手段により支配的な周波数成分と判定され、第1スペクトル変換手段の出力振幅スペクトルから、前記目的信号が支配的と判定された周波数成分が音源周波数成分抽出手段により抽出される。この抽出された目的信号が支配的な周波数成分の振幅スペクトルは、前記第1スペクトル変換手段により算出された位相スペクトルを用いて時間波形に時間波形変換手段により変換される。

【0010】この発明の第2形態は、第1形態の収音装置において、前記音源周波数成分選択手段で目的信号が支配的と判定されなかった周波数成分の第1スペクトル変換手段の振幅スペクトルに対して予め設定された重み係数（1より小）が騒音周波数成分抑圧手段で乗算され、その乗算出力は、前記第1スペクトル変換手段で算出された対応周波数成分の位相スペクトルを用いて時間波形に前記時間波形変換手段で変換される。

【0011】この発明の第3形態は第1又は第2形態の収音装置において、前記音源周波数成分抽出手段において目的信号が支配的と判定された周波数成分の振幅スペクトルの大きさと、予め設定された無音区間判定しきい値とが音源無音区間判定手段で比較され、前記振幅スペ

クトルが前記無音区間判定しきい値よりも小さいとき目的信号の音源が無音状態であると判定され、目的信号の音源が無音状態と判定された場合において、しきい値算出手段で、前記音源周波数成分選択手段に用いるしきい値が前記レベル差算出手段より出力されるレベル差以上となるように算出更新される。

【0012】この発明の第4形態は第2又は第3形態の收音装置において、前記騒音周波数成分抑圧手段において目的信号が支配的と判定されなかった周波数成分に乗ずる重み係数が、前記レベル差算出手段による各周波数成分のレベル差と前記音源周波数成分選択手段のしきい値の大小関係を比較して重み係数算出手段で算出される。

【0013】この発明の第5形態は第3又は第4形態の收音装置において、前記音源無音区間判定手段により目的信号の音源が無音状態と判定された場合において、音源無音区間減衰手段により前記音源周波数成分抽出手段と前記騒音周波数成分抑圧手段の出力または、前記時間波形変換手段の出力が減衰される。この発明の第6形態は第1乃至第4形態の何れかを前提として、目的信号の音源の発声状態が発声検出手段により検出され、その検出結果より目的信号の音源が発声状態か無発声状態かのいずれかであるかが発声状態判定手段で判定される。この発声状態判定手段は第3形態の音源無音区間判定手段を用いてもよく、これとは独立に設けてもよい。前記時間波形に変換された信号は音声蓄積手段に蓄積され、その蓄積された信号を利用して音声認識処理手段により音声認識され、音声認識制御手段により、発声状態判定手段の判定結果を利用して、発声区間の音声データのみが音声認識処理手段に入力制御され、その認識処理された認識結果が出力制御される。

【0014】この発明の第7形態は第1乃至第6形態の何れかにおいて、前記目的信号の音源に近い位置に設置された第1マイクロホンと前記目的信号の音源から離れた位置に設置された第2マイクロホンのうち、前者のマイクロホンが後者のマイクロホンに比べて使用状態において使用者の口元に近い位置になるようにハンドセット、ヘッドセット、イヤーマイクセットなどに組み込まれている。

【0015】「作用」第1形態においては、目的信号の音源に近い位置と、前記位置より目的信号の音源から離れた位置に設置されたマイクロホンの振幅スペクトルのレベル差によって目的信号が支配的な周波数成分であるか否かの判定が行われる。この発明の装置構成では、目的信号について、その音源と各々のマイクロホンの位置関係はほとんど変化せずそれらの距離も短いため、2つのマイクロホンの間で安定したレベル差が生じる。一方、騒音については、その音源とマイクロホンの間の距離は、目的信号の音源とマイクロホンの間の距離に比べて長くなると考えてよい。このため、目的信号によって

生じる2つのマイクロホンの出力レベル差は、騒音によって生じる2つのマイクロホンの出力のレベル差よりも常に大きくなると考えられる。この発明では、上記のように2つのマイクロホンに生じるレベル差が目的信号と騒音とで異なる点に着目して目的信号が支配的な周波数成分の抽出処理を行う。このような周波数成分抽出処理によれば、正確な騒音スペクトルの推定を必要としない騒音抑圧処理が可能になる。

【0016】第2形態においては、第1形態において目的信号が支配的と判定されなかった周波数成分に対して予め設定された重み係数を乗ずることにより、対応する周波数成分が減衰される。この騒音周波数成分抑圧処理によって、目的信号が支配的と判定されなかった周波数成分に重畳している目的信号の成分が欠落してしまうことによる処理信号の音質劣化を防ぎ、音質を向上させる。

【0017】第3形態においては、第1又は第2形態において目的音源が無音状態であることが判定されると、設定しきい値が算出され、音源周波数成分選択手段において目的信号が支配的な周波数成分の判定精度が向上し、音質が向上する。第4形態においては、第2、第3形態において目的信号が支配的でない周波数成分の減衰量が騒音の重畳度に応じて制御され、音質が向上する。

【0018】第5形態においては第3又は第4形態において目的信号の音源が無音状態と判定された場合において、前記音源周波数成分抽出手段と前記騒音周波数成分抑圧手段の出力または、時間波形変換手段の出力が減衰され騒音抑圧効果が向上する。第6形態においては、第1形態と同様に上記のように2つのマイクロホンに生じるレベル差が目的信号と騒音とで異なる点に着目して目的信号が支配的な周波数成分の抽出処理を行うため、正確な騒音スペクトルの推定を必要としない騒音抑圧処理が可能になり、その結果SN比を改善した音声信号による音声認識処理が可能となる。

【0019】また、発声区間検出には、例えば、骨導マイクロホンの出力外耳道マイクロホンの出力、あるいはカメラによる口元映像の分析結果を用いる。骨導音あるいは外耳道音は、高周波数成分の減衰のため音質が充分でないが、耐騒音性に優れているため、それらの出力信号は騒音の影響による変形が小さいと考えられる。図21は、騒音下（ピンクノイズ）90 dBAの実験室環境下で発声したときの気導音と骨導音の時間波形を比較したものである。この図より、骨導マイクロホンは、気導マイクロホンと比較して発声／無発声区間が明確になっていることが確認できる。この骨導マイクロホンあるいは外耳道マイクロホンの出力信号を使うことによって騒音下において精度の高い発声区間検出が可能となる。また、テレビ電話などカメラの映像を利用するシステムでは、話者の口元の映像情報を利用することによって周囲騒音に依存しない発声区間検出が可能となる。

【0020】第2形態を前提とする場合はその騒音周波数成分抑圧処理によって、目的信号が支配的と判定されなかった周波数成分に重畳している目的信号の成分が欠落してしまうことによる処理信号の音質劣化を防ぎ、音声認識精度を向上させる。第3形態を前提とする場合はその無発声区間しきい値算出手段によって、音源周波数成分選択部において目的信号が支配的な周波数成分の判定精度を向上させ、音質劣化を防ぐことによって音声認識精度を向上させる。

【0021】第4形態を前提とする場合はその重み係数算出手段によって、目的信号が支配的でない周波数成分の減衰量を騒音の重畳度に応じて制御し、音質劣化を防ぐことによって音声認識精度を向上させる。第7形態では目的信号の音源に近い位置に設置されたマイクロホンと目的信号の音源から離れた位置に設置されたマイクロホンのうち、前者のマイクロホンが後者のマイクロホンに比べて口元に近い位置になるようにハンドセット、ヘッドセット、イヤーマイクセットを組み込むことによって、各々の送受話器において送話信号の耐騒音性能を向上させ、また音声コマンド等を利用する際の音声認識精度を向上させることが可能となる。

【0022】

【発明の実施の形態】「実施例1」図1はこの発明の第1形態の実施例を示すブロック図である。マイクロホン1は近接音源用であって、マイクロホン2（遠隔音源用）よりも目的信号の音源に近い位置に配置されてある。マイクロホン1、2の出力は目的信号の音源からの音に対してレベル差が大であるが、騒音の音源はマイクロホン1、2から十分遠くにあって、騒音によるマイクロホン1、2の出力のレベル差はゼロ乃至わずかである。

【0023】図7はこの発明の第1形態の実施例の処理手順を示す流れ図である。まず、マイクロホン1、2に騒音が重畳した目的信号が各々取り込まれ、それをデジタル信号として読み込む（S02）。読み込まれたマイクロホン1、2の信号を以下では、L、Rとする。スペクトル変換部3、4では、取り込んだ信号L、Rをスペクトル $L(\omega_i)$ 、 $R(\omega_i)$ に変換する（S03）。この変換は、例えば離散的フーリエ変換によって実行される。

【0024】レベル差算出部5では、 $L(\omega_i)$ 、 $R(\omega_i)$ の各周波数成分について、以下の式で与えられるレベル差 $\Delta LR(\omega_i)$ を計算する（S04）。
$$\Delta LR(\omega_i) = 20 \log_{10} (|L(\omega_i)| / |R(\omega_i)|)$$
上式中の $\omega_i$ は周波数（ $i = 1, 2, \dots, n$ ）、 $|L(\omega_i)|$ 、 $|R(\omega_i)|$ は、各々L、R信号の振幅スペクトル成分を表わす。

【0025】音声周波数成分選択部6では、各周波数成分について $\Delta LR(\omega_i)$ と予め設定されたしきい値T

$h(\omega_i)$ の大小関係より、目的信号が支配的な周波数の選択を行う。目的信号が支配的な周波数か否かの判定条件は例えば以下の式によって決定される（S05）。
$$\Delta LR(\omega_i) > Th(\omega_i) \rightarrow \text{目的信号が支配的}$$
$$\Delta LR(\omega_i) \leq Th(\omega_i) \rightarrow \text{目的信号が支配的でない}$$

音源周波数成分抽出部7では、 $L(\omega_i)$ から、音源周波数成分選択部6で選択した周波数について、目的信号が支配的な周波数成分のみを抽出する（S06）。

【0026】最後に、（S06）の出力を時間波形変換部8において信号Lの位相スペクトル $\Phi(\omega_i)$ を用いて時間波形に変換し、時間波形信号を出力する（S07）。しきい値 $Th(\omega_i)$ はマイクロホン1、2と目的信号音源（例えばマイクロホン1、2の使用者の口）との相対関係から、実験的に決めることができる。

「実施例2」第1形態では、目的信号が支配的な周波数成分と判定されなかったものについては完全に除去する。しかし、音声と騒音の周波数成分が互いに重なっているような成分についても完全に除去してしまうと音質の歪みの原因となり好ましくない。第2形態は、音声と騒音の重畳した周波数成分において上記の「騒音成分の抽出しすぎ」の問題を改善することによって、処理後の音質を向上させるものである。

【0027】図2は第2形態の実施例を示すブロック図である。図2においてスペクトル変換部3と音源周波数成分選択部6の各出力側に騒音周波数成分抑圧部9を接続したことを除けば、図1と同じ動作である。以下で、第2形態の実施例を示す図8の流れ図を用いて騒音周波数成分抑圧部9における処理について説明する。騒音周波数成分抑圧部9では、音源周波数成分選択部6において目的信号が支配的でないと判定された周波数成分、すなわち、

$$\Delta LR(\omega_i) \leq Th(\omega_i)$$

を満たす、スペクトル変換部3の出力中の周波数成分 $L(\omega_i)$ について、外部より設定された重み係数 $w(\omega_i)$ を乗じた値を騒音抑圧処理後のスペクトル成分 $S(\omega_i)$ とする（S07）。

$$【0028】S(\omega_i) = w(\omega_i) \times L(\omega_i)$$

$w(\omega_i)$ の値は、騒音抑圧量を重視したい帯域で小さな値（ $w(\omega_i) \approx 0$ ）に設定し、音質を重視したい帯域では大きな値（ $w(\omega_i) \approx 1$ ）に設定すればよい。目的信号が支配的と判定された周波数成分は第1形態の実施例の場合と同様に音源周波数成分抽出部7で処理され、抽出された周波数成分は $S(\omega_i)$ とされる（S06）。

$$【0029】S(\omega_i) = L(\omega_i)$$

騒音周波数成分抑圧部9と音源周波数成分抽出部7の出力 $S(\omega_i)$ は、時間波形合成部8において逆フーリエ変換し、時間波形信号を出力する（S08）。

「実施例3」第1、第2形態では、音源周波数成分選択

部6において、ある周波数成分が目的信号が支配的であるか否かを判定するしきい値 $Th(\omega_i)$ を外部より設定している。第3形態は、目的音源が無音状態であるときの周囲騒音に生じているマイクロホン1、2間の各周波数成分におけるレベル差を利用して、しきい値 $Th(\omega_i)$ を算出し、修正することにより音源周波数成分選択部6において目的信号が支配的であるか否かの判定精度を向上させ、音質を向上させるものである。

【0030】図3は第3形態の実施例を示すブロック図である。図3において、音源周波数成分抽出部7の出力側に音源無音区間判定部10を接続し、その音源無音区間判定部10の出力側に音源無音区間しきい値算出部11を接続したことを除けば、図2と同じ動作である。以下で、第3形態の実施例を示す図9の流れ図を用いて音源無音区間判定部10、および音源無音区間しきい値算出部11における処理について説明する。

【0031】音源無音区間判定部10では、音源周波数成分抽出部7からまず第一に目的信号が支配的な振幅スペクトルの和 $P$ を求め(S08)、 $P$ と外部より設定したしきい値 $PTh$ との大小関係より目的信号の音源の無音状態を検出する(S09)。

$P > PTh \rightarrow$  目的信号の音源が有音状態

$P \leq PTh \rightarrow$  目的信号の音源が無音状態

音源無音区間判定部10において、目的信号の音源が無音状態と判定された場合には、音源無音区間しきい値算出部11においてしきい値 $Th(\omega_i)$ を算出する。例えば、新しいしきい値を以下の式により算出する(S10, S11)。 $Th(\omega_i) = \Delta LR(\omega_i)$  ( $\Delta LR(\omega_i) > Th(\omega_i)$  のときのみ)つまり、マイクロホン1、2の出力の周波数成分は、騒音に対しては本来は互いにはほぼ等しく $\Delta LR(\omega_i)$ はゼロであるが、反射などの関係である周波数成分については、 $\Delta LR(\omega_i)$ が比較的大となる。この音源無音区間での $\Delta LR(\omega_i)$ は、目的信号が存在する場合も、その騒音周波数成分が存在し、この成分ではしきい値が高められ、騒音が有効に除去される。なお必要に応じて $Th(\omega_i)$ がある程度以上大きくなるとリセットし、一時的に騒音の $\Delta LR(\omega_i)$ が大きくなったことがいつまでも影響しないようにすることができる。

【0032】「実施例4」第2形態および第3形態では、騒音周波数成分抑圧部9において目的信号が支配的でない周波数成分に乘じる重み係数 $w(\omega_i)$ を外部より設定している。第4形態は、レベル差算出部5により算出された各周波数成分のレベル差 $\Delta LR(\omega_i)$ としきい値 $Th(\omega_i)$ を利用して、重み係数 $w(\omega_i)$ を算出することにより、騒音周波数成分抑圧部9において各周波数成分の減衰量を騒音成分の重畳度に応じて制御し、音質を向上させるものである。

【0033】図4は第4形態の実施例を示すブロック図である。レベル差算出部5及び音源周波数成分選択部6

の各出力側に重み係数算出部12を接続したことを除けば、図3と同じ動作である。以下で、第4形態の実施例を示す図10の流れ図を用いて重み係数算出部12における処理について説明する。重み係数算出部12では、音源周波数成分選択部6において目的信号が支配的でない判定された周波数成分に対して乗じる重み係数 $w(\omega_i)$ を例えば次式で算出する(S07)。

【0034】

$w(\omega_i) = (\Delta LR(\omega_i) / Th(\omega_i))^{npw}$   
 $npw$ の値としては騒音抑圧効果を高めたい場合に大きな値とすればよい。この式で求められた $w(\omega_i)$ は、騒音の重畳度が大きいときには $\Delta LR(\omega_i)$ は騒音スペクトルのレベル差に支配されるため小さな値をとり、騒音の重畳度が小さいときには $\Delta LR(\omega_i)$ は目的信号のスペクトルのレベル差に支配されるため大きな値をとる。なお、上式で $w(\omega_i)$ を計算した値が負となった場合には、騒音が支配的な周波数成分と見なして、その値を小さな値、例えば0とする。

【0035】「実施例5」第5形態は、第3形態または第4形態において音源無音区間検出部10により目的信号の音源が無音状態と判定された場合に、音源周波数成分抽出部と騒音周波数成分抽出部の出力または、時間波形変換部の出力を減衰させ、騒音抑圧効果を向上させるものである。

【0036】図5は第5形態の実施例を示すブロック図である。図5において、音源周波数成分抽出部7及び騒音周波数成分抑圧部9と時間波形変換部8との間に音源無音区間減衰部13を挿入したことを除けば、図4と同じ動作である。以下で、第5形態の発明の実施例を示す図11の流れ図を用いて音源無音区間減衰部13における処理について説明する。

【0037】音源無音区間減衰部13では、音源無音区間判定部10において目的信号の音源が無音状態と判定された場合には(S10)、音源周波数成分抽出部7と騒音周波数成分抑圧部9の出力 $S(\omega_i)$ を全帯域にわたり減衰させる(S13)。なお、音源無音区間減衰部13の処理は、時間波形変換部8の出力信号に対して行ってもよく、その効果は(S13)の処理による効果と同等である。

【0038】「実験例1」第4形態を適用した実験例を以下に示す。目的信号は音声、騒音信号は駅のホームでの周囲騒音を用い、マイクロホン1とマイクロホン2の入力信号は、図12に示すように計算機上で作成した。つまり目的信号(音声)はマイクロホン2にはマイクロホン1よりも6dB低下して入力されるが、騒音はマイクロホン1、2に同レベルで入力される。S/N比は目的信号の平均電力と騒音信号の平均電力の比で定義し、マイクロホン1におけるその値を-9dBとした。信号のスペクトル分解における周波数分解能は22Hz、分析フレームは46ms、分析フレームの更新周期は23

msとした。

【0039】図13は、マイクロホン1の処理前の目的信号A、騒音信号B、騒音信号+目的信号C、そして処理後の信号Dである。この図から、処理後の信号が処理前の目的信号をよく復元していることが確認できる。

「実施例6」図14に第6形態の実施例を示し、図1と対応する部分に同一符号を付けてある。この実施例では発声検出部14の出力が発声状態判定部15へ供給され、発声状態判定部15の出力は音声認識制御部16に入力される。時間波形変換部8の出力は音声蓄積部17に蓄積され、その蓄積された信号が音声認識処理部18で音声認識処理され、その認識結果は音声認識制御部16を介して出力される。

【0040】図15に第6形態の実施例の処理手順を示す。この処理においてステップS01～S06は図7の処理と同一である。ステップS05において $\Delta LR(\omega_i) \leq Th(\omega_i)$ である目的信号は支配的でない周波数成分については騒音が支配的な周波数成分とみなして小さな値 $\epsilon ps$ に変更する(S07)。 $\epsilon ps$ の値は例えば0としてステップS08で時間波形に変換することは図7と同様である。

【0041】発声検出部14では、骨導あるいは外耳道マイクロホン、またはカメラを接続して動作させることが可能である。図15には、骨導あるいは外耳道マイクロホンを接続して、発声状態判定部15で目的信号の音源の発声/無発声状態を判定する方法を示している。まず第一に骨導あるいは外耳道マイクロホンの出力信号のパワーPBを求める(S09)。次に、PBと外部より設定されたしきい値PTHBとの大小関係より目的信号の音源の発声/無発声状態を判定する(S10)。

【0042】

$PB > PTHB \rightarrow$  目的信号の音源が発声状態  
 $PB \leq PTHB \rightarrow$  目的信号の音源が無発声状態  
 なお、発声検出部14にカメラを利用した場合には、例えば口元の映像信号の変動を監視して発声状態判定を行えばよい。音声認識制御部16では、 $\Phi$  音声蓄積部17から音声認識処理部18への音声データ入力、 $\Phi$  音声認識結果の出力、の2つの制御を行う。 $\Phi$  の制御については、以下の $\Phi-1$ 、 $\Phi-2$ に示す二つの方法が考えられる。

【0043】 $\Phi-1$ ：発声区間に相当する時間波形変換部8のデータのみを音声蓄積部17に蓄積させ、音声認識処理部18に音声蓄積部17のデータの全てを入力させる方法

$\Phi-2$ ：音声蓄積部17に時間波形変換部8のデータを全て蓄積させ、その中から発声区間に相当するデータを音声認識処理部18に入力させる方法この実施例では $\Phi-1$ の制御方法の具体例を示す。

【0044】音声認識制御部16では、発声が始まった瞬間、即ち、発声状態判定部15において前処理フレー

ムが無発声状態と判定され、かつ現在のフレームが発声状態と判定された瞬間に(S10、S11)音声蓄積部17のデータ入力端子をオンにする。この瞬間から、音声蓄積部17に時間波形変換部8の出力信号の蓄積が開始される(S12)。この蓄積は、発声状態判定部15において発声状態と判定されている間行われ、発声状態から無発声状態に移した瞬間(発声状態判定部15において、前処理フレームが発声状態と判定され、かつ現在のフレームが無発声状態と判定された瞬間(S13))に終了する(S14)。

【0045】音声蓄積部17での音声データ蓄積が終了となったとき(S14)、音声認識処理部18では音声蓄積部17で蓄積されたデータが入力され、音声認識制御部16の命令を受けて音声認識処理が行われる(S15)。音声認識結果は音声認識制御部16を介して外部に出力される(前記 $\Phi$ の制御に相当する)。音声蓄積部17に蓄積された音声データについては音声認識処理部18に出力した直後にクリアされる(S16)。

【0046】なお、使用する音声認識アルゴリズムが音声のスペクトル情報のみを用い、音声の時間波形を必要としない場合には、記憶装置等で蓄積される処理音声データは、(S08)の時間波形出力でなく、処理されたスペクトル(S06)と(S07)としてもよい。その場合、(S08)の時間波形変換処理は不要となる。

「実施例7」実施例2に実施例6の構成14～18を付加し、音声と騒音の重畳した周波数成分において上記の「騒音成分の抽出しすぎ」の問題を改善することによって、処理後の音質を向上させ、音声認識精度を向上させるものである。図2に図14の構成14～18が付加され、図8中のS08の次に図15のステップS09～S18が付加されたものとなる。

【0047】「実施例8」実施例3に実施例6の構成14～18を付加して、目的信号の音源が無音状態(無発声区間)であるときの周囲騒音に生じているマイクロホン1、2間の各周波数成分におけるレベル差を利用して、しきい値 $Th(\omega_i)$ を算出し、修正することにより音源周波数成分選択部6において目的信号が支配的であるか否かの判定精度を向上させ、音質劣化を防ぐことによって、音声認識精度を向上させるものである。この場合は図16に示すように図3に対し図14中の構成14～18を付加し、かつ図3中の音源無音区間判定部10を省略して発声状態判定部15の出力を無発声(音源無音)区間しきい値算出部11へ入力する。この実施例の処理の流れは図17に示すように図15に対し、発声状態判定部15において、目的信号の音源が無音状態と判定された場合には(S20)、無発声区間しきい値算出部11においてしきい値 $Th(\omega_i)$ を算出する。例えば、新しいしきい値を以下の式により算出し(S17、S18)、ステップS02に戻る。 $Th(\omega_i) = \Delta LR(\omega_i)$  ( $\Delta LR(\omega_i) > Th(\omega_i)$ )の

きのみ) 発声検出部14と発声状態判定部15を省略し、代りに音源無音区間判定部10を用いてもよい。

【0048】「実施例9」この実施例は、レベル差算出部5により算出された各周波数成分のレベル差 $\Delta L R(\omega_i)$ としきい値 $Th(\omega_i)$ を利用して、重み係数 $w(\omega_i)$ を算出することにより、騒音周波数成分抑圧部9において各周波数成分の減衰量を騒音成分の重畳度に応じて制御し、音質劣化を防ぐことによって音声認識精度を向上させるものである。図16の構成に対し、図4中の重み係数算出部12が付加され、これに応じて処理の流れは図10中のステップS07が図17に対し加わる。

【0049】「実験例2」実施例6を適用した実験例を以下に示す。発声区間検出部14の入力信号として骨導マイクロホンの出力を用いた。図18Aに示す收音条件、つまり騒音源から150cm離れ、第1、第2マイクロホン1、2の間隔が10cmであり、かつ図18Bに実験条件を示す。

【0050】図19は、マイクロホン1の出力(処理前)、実施例6の騒音抑圧処理を行った出力(処理後)の音声認識率を求めた結果である。この図から、高騒音下において、処理後の信号が処理前の信号よりも音声認識率を改善していることがわかる。図6に第7形態の実施例を示す。図6Aはハンドセット21にマイクロホン1とマイクロホン2を取付けた場合である。ハンドセット21の使用状態においてマイクロホン1はその使用者の口22、つまり目的信号の音源近くに位置され、マイクロホン2はハンドセット21の受話器部分、つまり耳23の近くに位置するようにされている。

【0051】図6Bはヘッドセット25にマイクロホン1、2を取付けた場合でヘッドセット25を使用者の頭部26に装着した使用状態で、その耳23に対接される受話器27の部分にマイクロホン2が取付けられ、この受話器27の部分から、支持アーム28が延長され、支持アーム28の遊端部が口22の近くに位置し、ここにマイクロホン1が取付けられる。

【0052】図6Cはイヤーマイクセット31に取付けた場合で、イヤーマイクセット31が耳23の部分に取付けられた状態で、アーム32が口22側に延長され、これにマイクロホン1が取付けられ、このアーム32と反対にアーム33が延長され、これにマイクロホン2が取付けられる。

【0053】

【発明の効果】以上、説明したように、第1形態の発明によれば、目的信号の音源に近い位置に設置されたマイクロホンと、前記位置より目的信号の音源から離れた位置に設置されたマイクロホンと、各々の出力信号を振幅スペクトルと位相スペクトルに変換し、その各周波数成分ごとの振幅スペクトルについて、レベル差を計算し、この各周波数成分ごとのレベル差と、予め設定されたし

きい値とを比較し、前記レベル差がしきい値よりも大きい周波数成分を目的信号が支配的な周波数成分と判定し、これら周波数成分を前記目的信号の音源に近い位置に設置されたマイクロホンの出力信号の振幅スペクトルから抽出し、これら抽出された目的信号が支配的な周波数成分の振幅スペクトルを前記目的信号の音源に近いマイクロホンの前記スペクトル変換の際に算出された位相スペクトルを用いて時間波形に変換することにより、正確な騒音スペクトル成分を推定することなく騒音抑圧処理が可能である。したがって、非定常騒音など騒音スペクトルの推定が難しい場合でも目的信号の抽出が可能な收音装置が得られる。

【0054】第2形態の発明によれば、第1形態の收音装置において、前記目的信号が支配的と判定されなかった周波数成分に対して予め設定された重み係数を乗じて騒音周波数成分を抑圧し、この抑圧された成分も、前記時間波形変換においては、前記目的信号の音源に近いマイクロホンにより算出される位相スペクトルを用いて時間波形に変換することにより、目的信号として抽出されなかった周波数帯域に存在していた目的信号の周波数成分が欠落することによる音質劣化を改善することが可能な收音装置が得られる。

【0055】第3形態の発明によれば、第1形態または第2形態の收音装置において、前記音源周波数成分抽出において目的信号が支配的と判定された周波数成分の振幅スペクトルの大きさと、予め設定された無音区間判定しきい値とを比較し、前記振幅スペクトルが前記無音区間判定しきい値よりも小さいとき目的信号の音源が無音状態であると判定し、無音状態と判定された場合には、前記音源周波数成分選択に用いるしきい値を、前記レベル差以上となるように更新するので、音源周波数成分選択において目的信号が支配的な周波数成分抽出精度が向上し、処理後の信号の品質向上が可能な收音装置が得られる。

【0056】第4形態の発明では、第2形態または第3形態の收音装置において、前記騒音周波数成分抑圧において目的信号が支配的と判定されなかった周波数成分に乗ずる重み係数を、前記各周波数成分のレベル差と前記音源周波数成分選択時のしきい値との大小関係を比較して算出しているので、騒音周波数成分抑圧において、各周波数成分の減衰量を騒音の重畳度に応じて決定することが可能な收音装置が得られる。

【0057】第5形態の発明では、第3形態または第4形態の收音装置において、目的信号の音源が無音状態と判定された場合において、前記音源周波数成分抽出出力と前記騒音周波数成分抑圧の出力または、前記時間波形変換手段の出力を減衰させているので、目的信号の音源が無音状態のときは信号が減衰され、これにより騒音が抑圧され、さらに騒音の少ない收音装置が得られる。

【0058】第6形態の発明では、第1乃至第4形態の

何れか1つとはほぼ同様の構成を含み、更に音声蓄積手段、音声認識処理手段を有し、前記発声状態判定手段の出力結果を利用して、発声区間の音声データのみを前記音声認識処理手段に入力し、及び前記音声認識処理手段で処理された認識結果を出力するようにしているので、正確な騒音スペクトル成分を推定することなく騒音抑圧処理が可能である。したがって、非定常騒音など騒音スペクトルの推定が難しい場合でも目的信号の抽出が可能であり、音声認識精度を向上させた音声認識装置が得られる。また、発声区間検出には、例えば骨導あるいは外耳道マイクロホン、またはカメラを利用することにより、周囲騒音の影響をほとんど受けないか、あるいは全く受けない発声区間検出が可能となり、それによってさらに音声認識精度を向上させた音声認識装置が得られる。

【0059】第7形態の発明は、第1乃至第6形態の收音装置又は音声認識装置において、前記目的信号の音源に近い位置に設置されたマイクロホンと前記目的信号の音源から離れた位置に設置されたマイクロホンのうち、前者のマイクロホンが後者のマイクロホンに比べて使用時に口元に近い位置になるようにハンドセット、ヘッドセット、イヤーマイクセットに組み込まれているので、従来のハンドセット、ヘッドセット、イヤーマイクセットにおいて送話信号の耐騒音性能を向上させることが可能となり、また各々の送受話器において音声コマンド等を利用する際の音声認識精度を向上させることが可能となる。従来、耐騒音性に優れた送話信号を得るイヤーマイクセットとして骨導マイクロホンとレシーバを一体化したものがある。しかし骨導マイクロホンによって收音された音声は周波数成分が低周波成分に偏っており、高周波成分が少ないため、音質が悪い。また骨導マイクロホンとレシーバとの音響結合の問題もある。しかしこの第7形態の発明では、気導音をベースとした收音であり、レシーバとマイクロホン間の距離も確保できるため、上記の問題を持たないイヤーマイクセットの提供が可能となる。

【0060】なお、以上の説明で使用したマイクロホンは、無指向性マイクロホンに限定されるものではなく、例えば、マイクロホン1は、目的信号の音源の方向に指向性を有するマイクロホンを使用し、マイクロホン2は、目的信号の音源と反対の方向に指向性を有するマイクロホンを使用してもよい。この場合、目的信号の音源方向のみに鋭い指向性を有する超指向性マイクロホンとして利用できる。

【0061】この発明における收音装置は、騒音抑圧が必要な各種收音装置のほか、通話を目的とした電話装置や、音声認識の入力装置にも利用できる。また図1乃至

図5、図14、図16は機能構成を示すものであって、マイクロプロセッサを主体とする手段により各機能を構成させることもでき、つまりコンピュータによりプログラムを読み出し、解読実行させることによって機能させることもできる。

【図面の簡単な説明】

【図1】請求項1の発明の実施例の機能的構成を示すブロック図。

【図2】請求項2の発明の実施例の機能的構成を示すブロック図。

【図3】請求項3の発明の実施例の機能的構成を示すブロック図。

【図4】請求項4の発明の実施例の機能的構成を示すブロック図。

【図5】請求項5の発明の実施例の機能的構成を示すブロック図。

【図6】この発明の第7形態の概略構成例を示す図。

【図7】請求項1の発明の実施例の動作手順を示す流れ図。

【図8】請求項2の発明の実施例の動作手順を示す流れ図。

【図9】請求項3の発明の実施例の動作手順を示す流れ図。

【図10】請求項4の発明の実施例の動作手順を示す流れ図。

【図11】請求項5の発明の実施例の動作手順を示す流れ図。

【図12】請求項4の発明を適用したシミュレーション実験例における信号と騒音の印加例を示す図。

【図13】請求項4の発明を用いた実験例における処理前の目的信号、騒音信号、騒音+目的信号、及び処理後の信号のそれぞれを示す図。

【図14】請求項23の発明の実施例の機能的構成を示すブロック図。

【図15】図14の実施例の動作手順を示す流れ図。

【図16】請求項20の発明の実施例の機能的構成を示すブロック図。

【図17】図16の実施例の動作手順を示す流れ図。

【図18】Aは請求項23の発明を適用したシミュレーション実験例における騒音と信号の印加例を示す図、Bはその実験条件を示す図。

【図19】図18の実験結果を示す図。

【図20】従来の技術を説明する流れ図。

【図21】騒音（ピンクノイズ）90 dBAの実験室環境下で発声した気導音と骨導音の時間波形の比較例を示す図。

【図1】

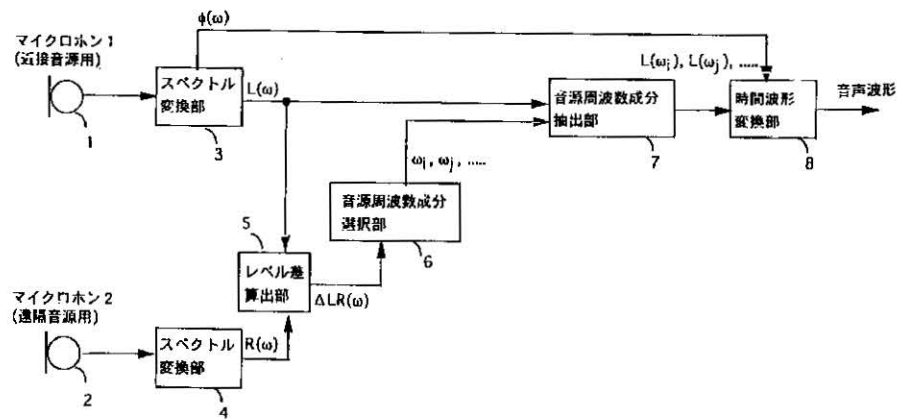


図 1

【図2】

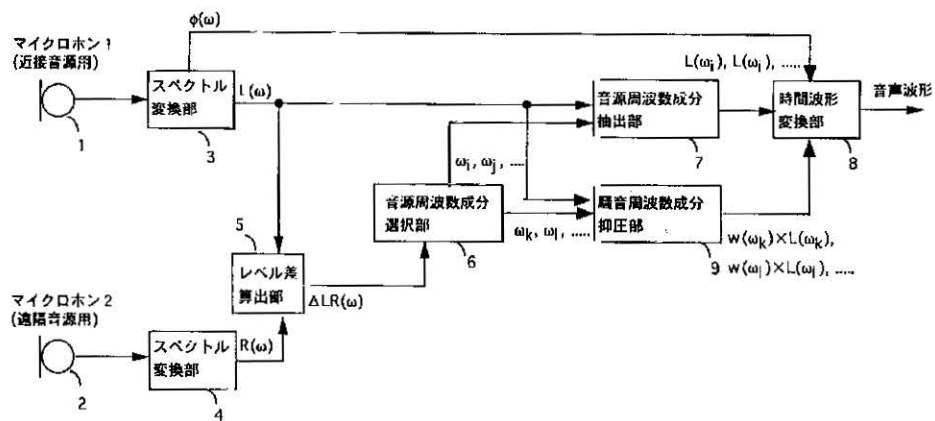


図 2

【図3】

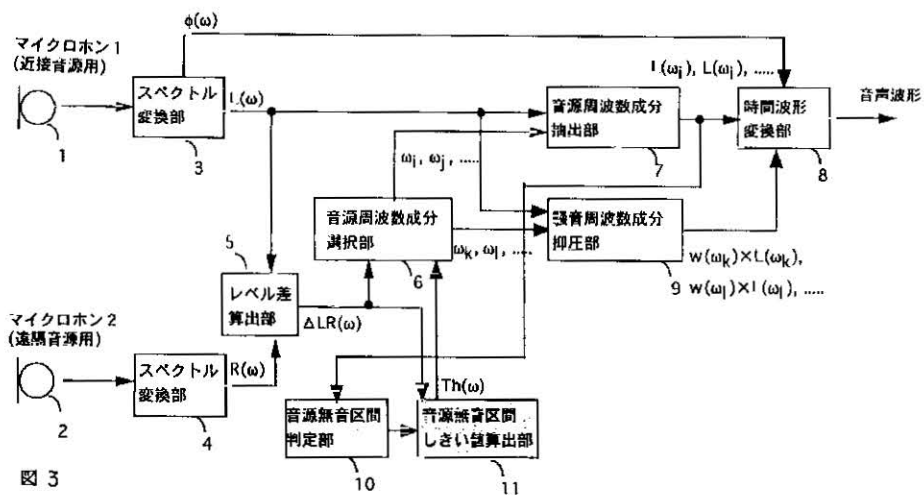


図 3

【図4】

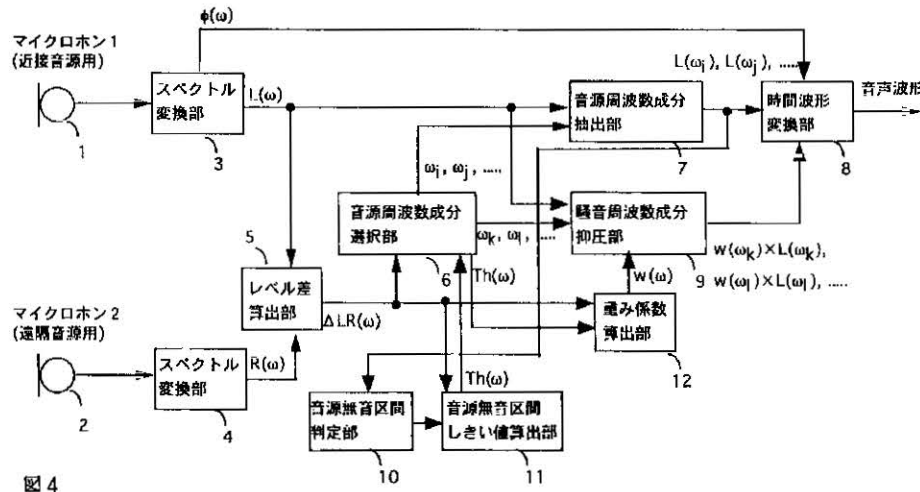


図4

【図5】

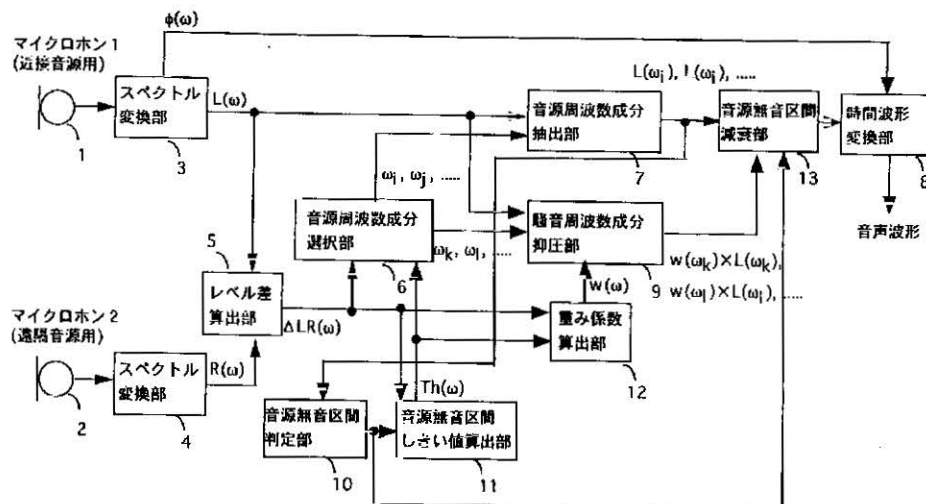


図5

【図6】

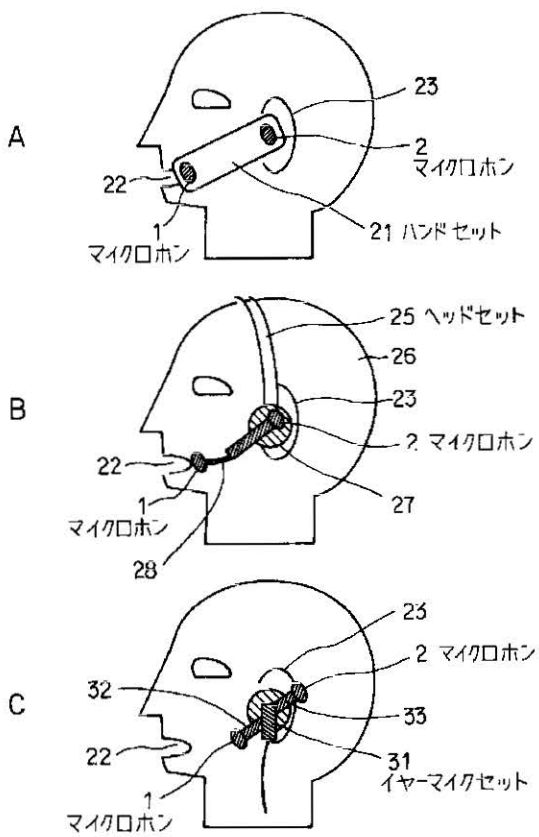


図 6

【図7】

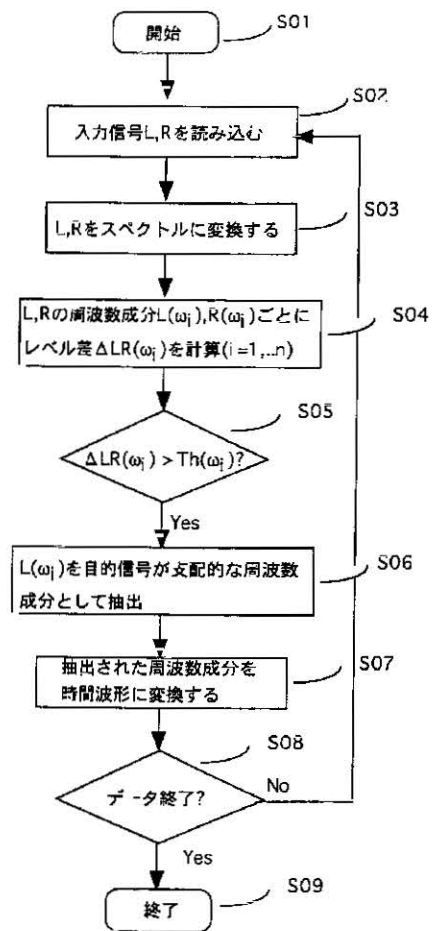


図 7

【図12】

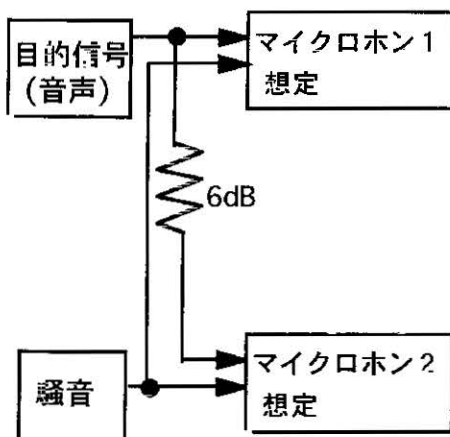


図12

【図19】

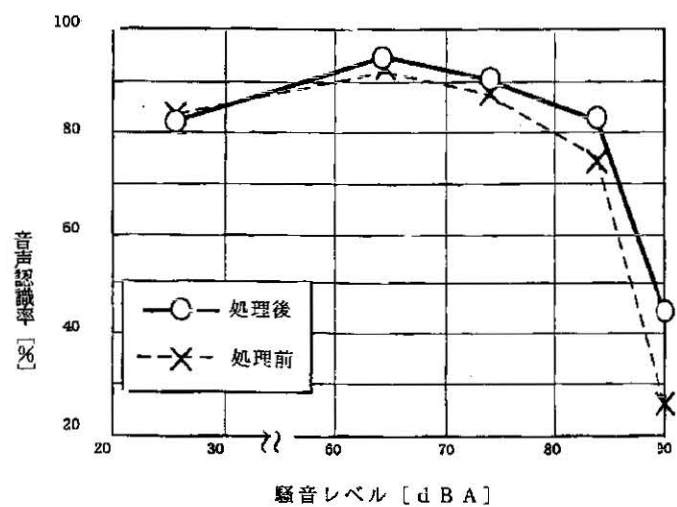
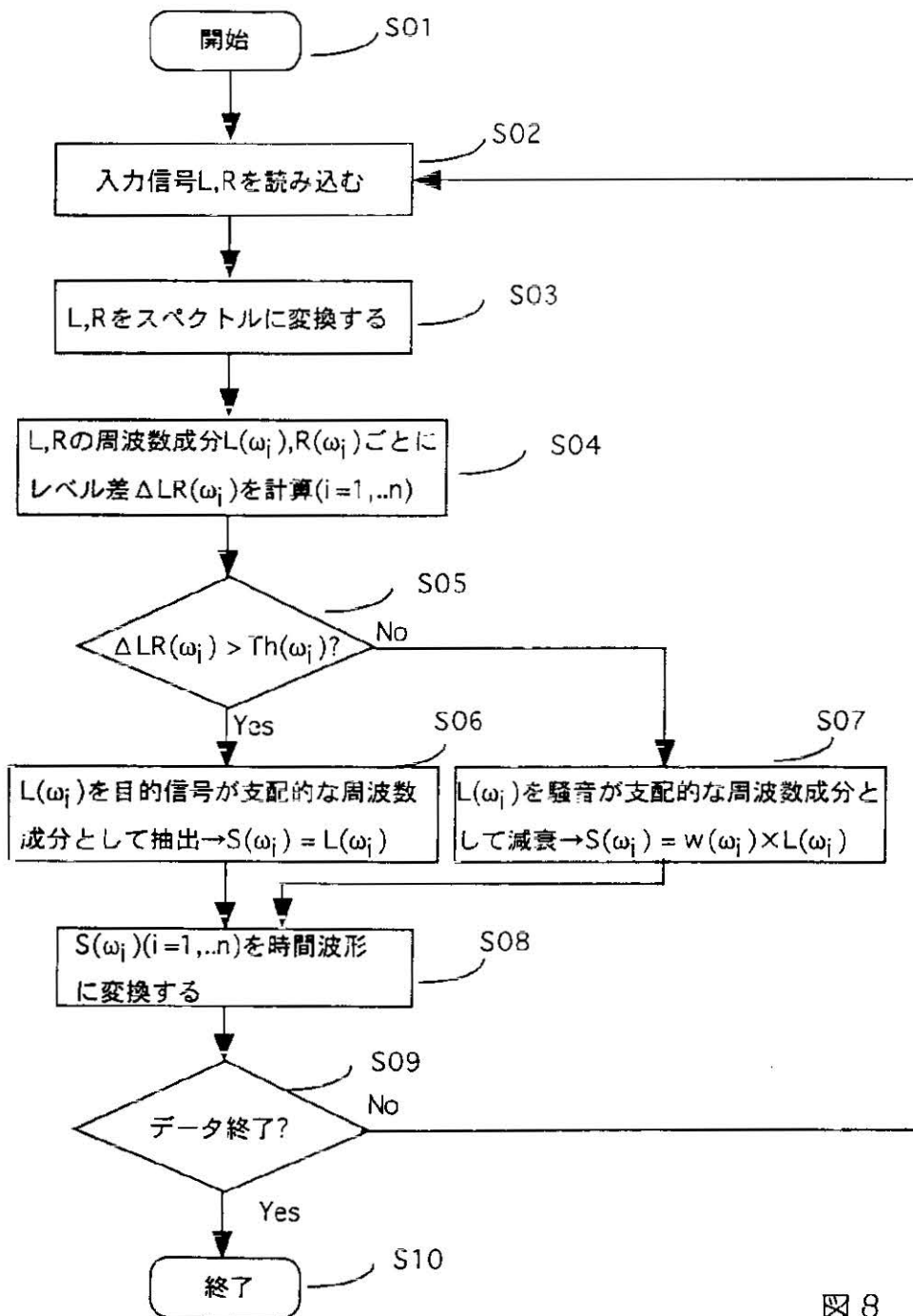
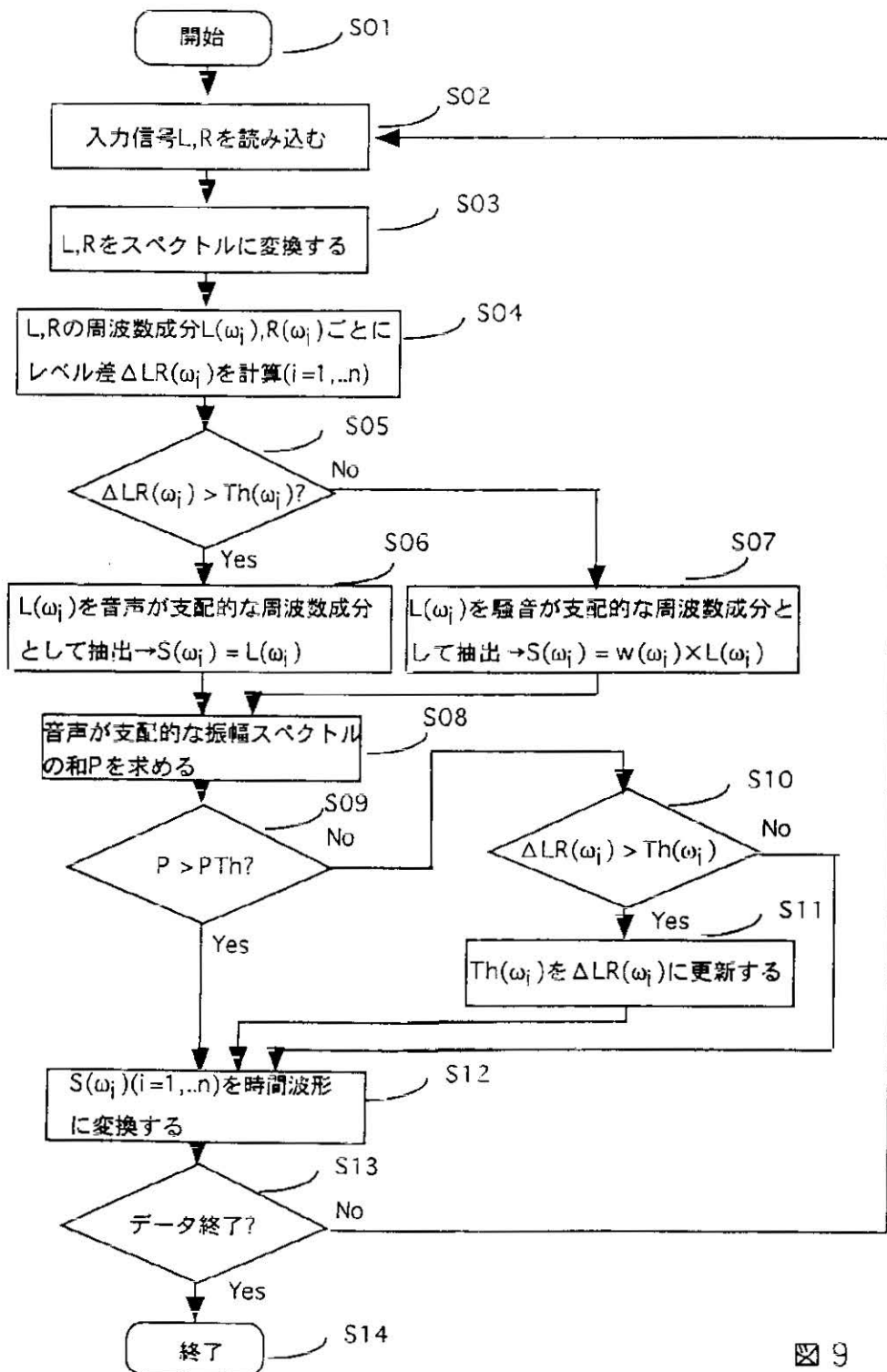


図19

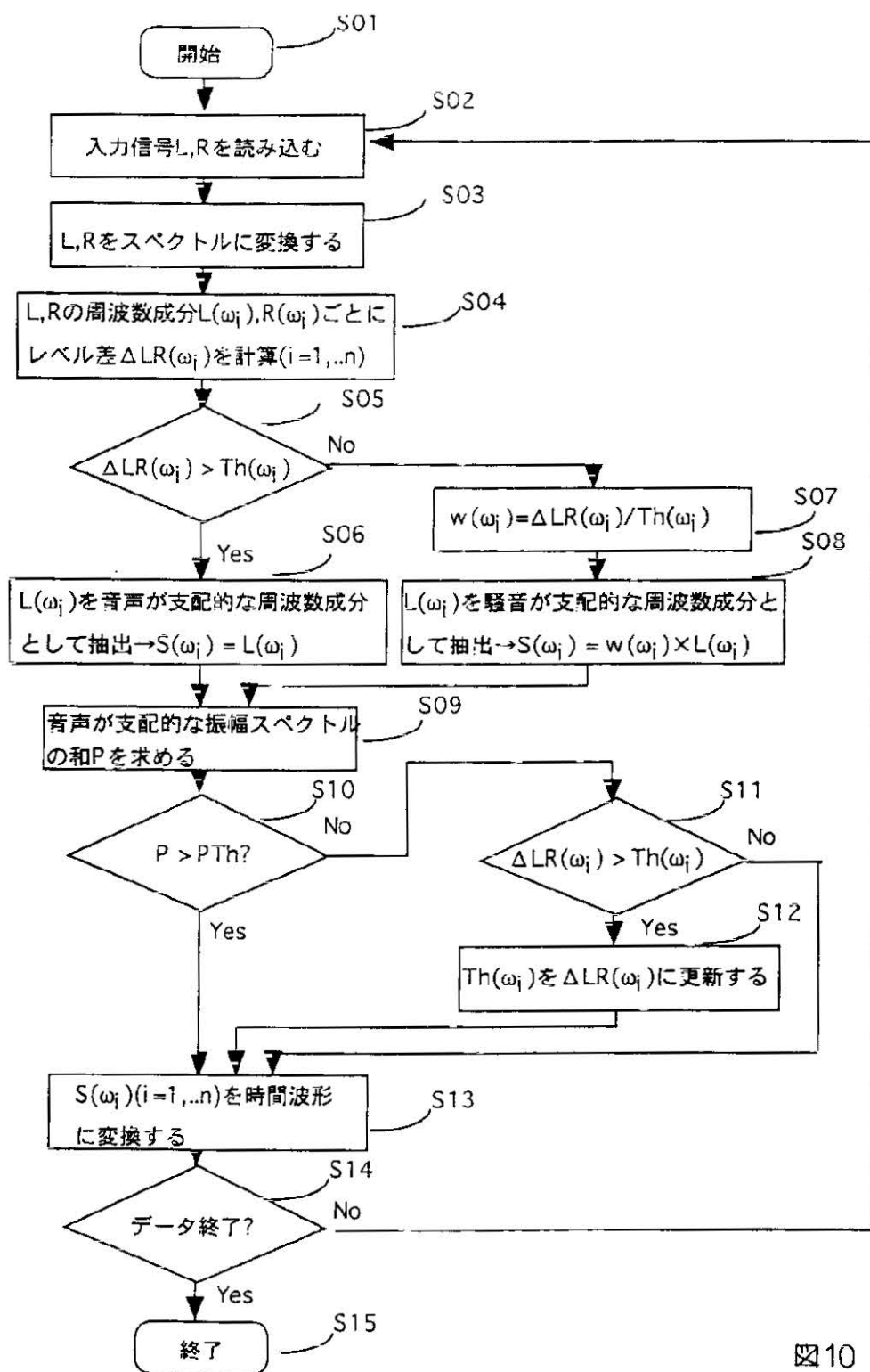
【図8】



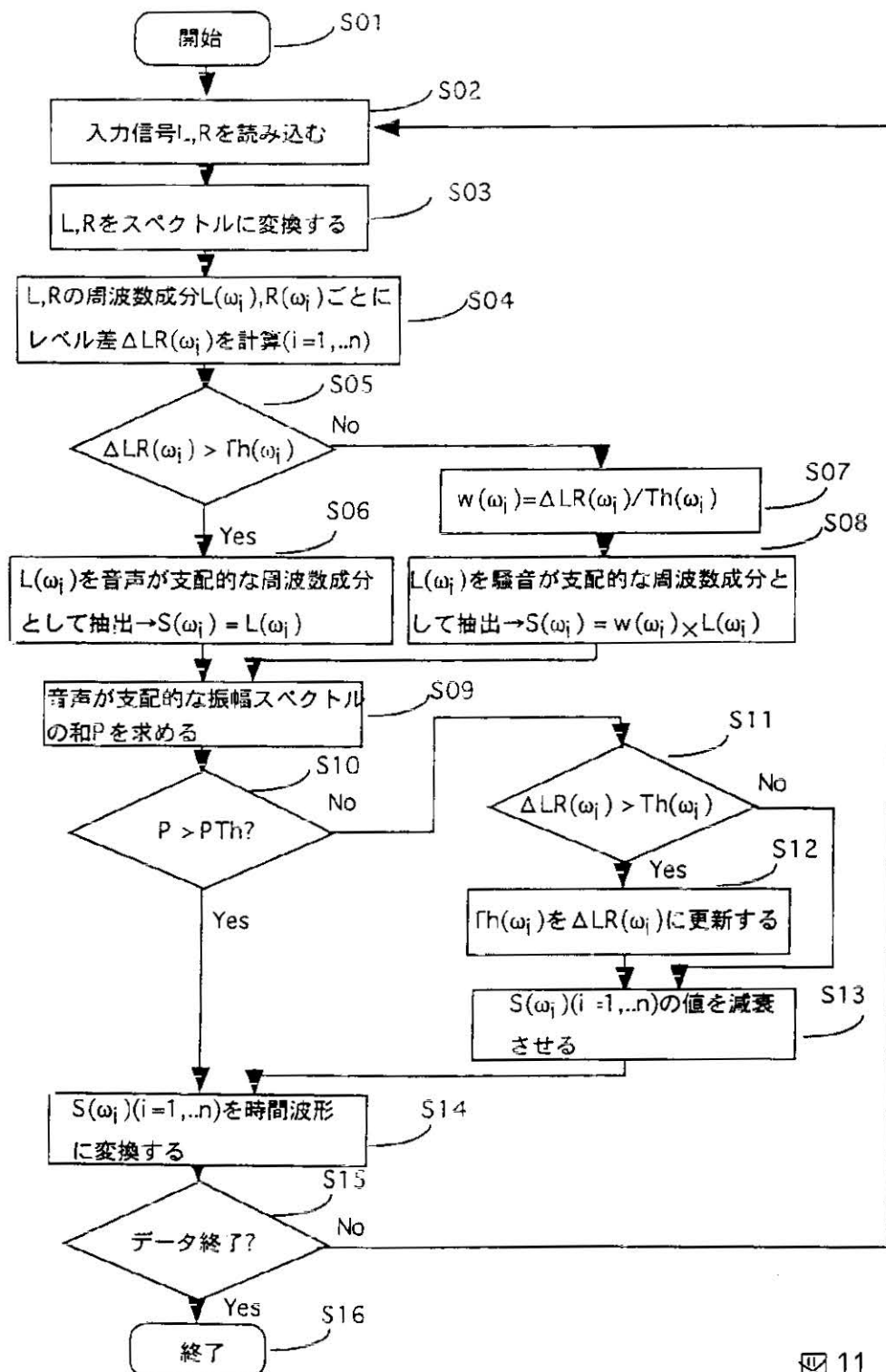
【図9】



【図10】



【図11】



【図13】

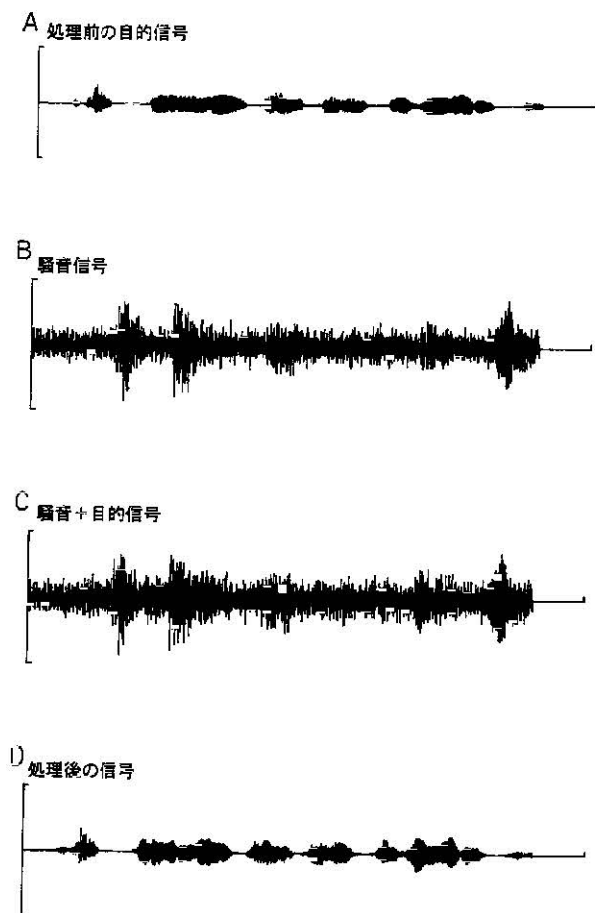


図13

【図15】

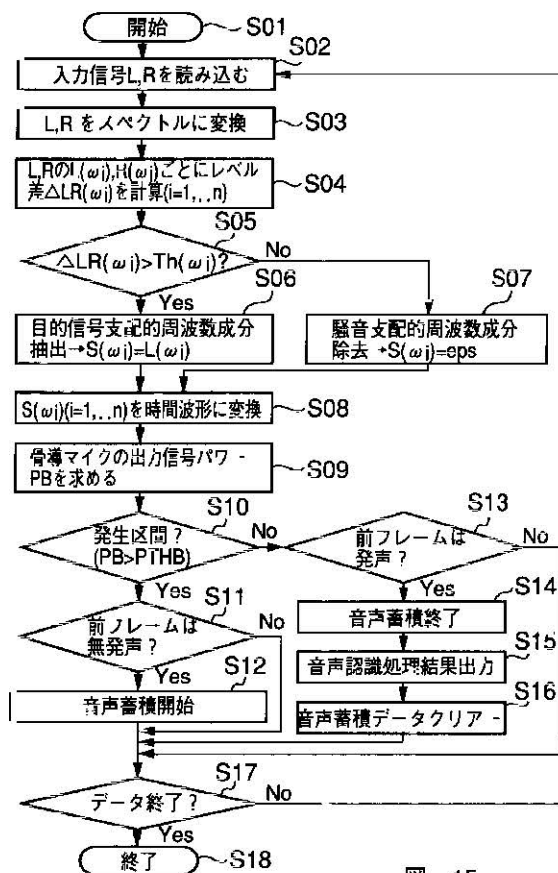


図 15

【図14】

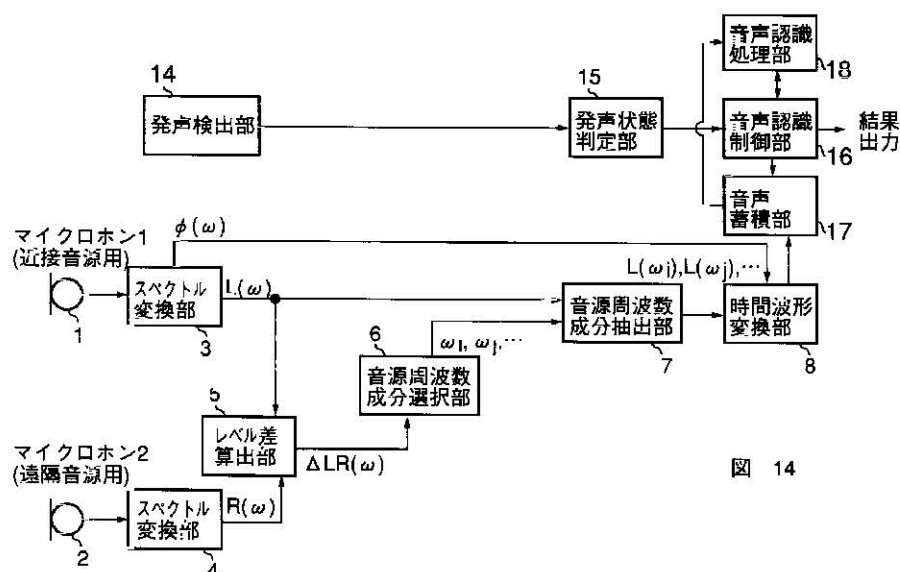


図 14



【図20】

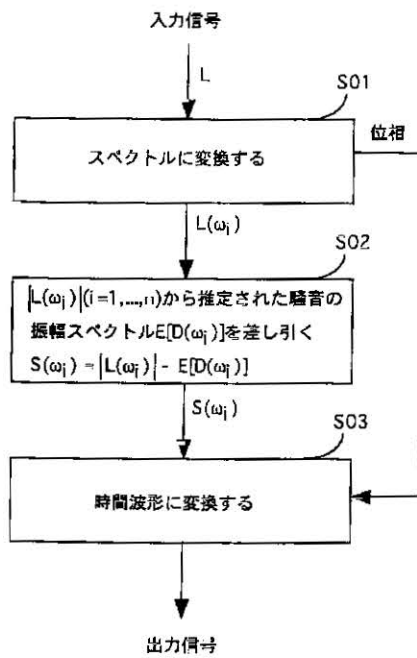


図20

【図21】

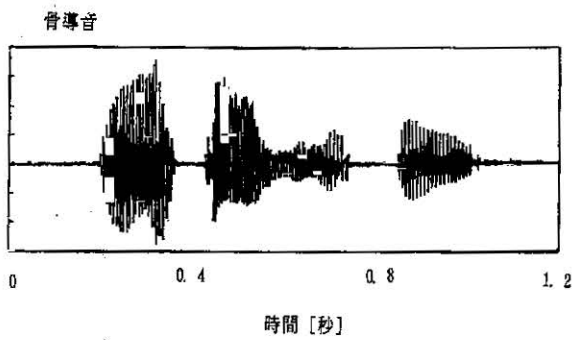
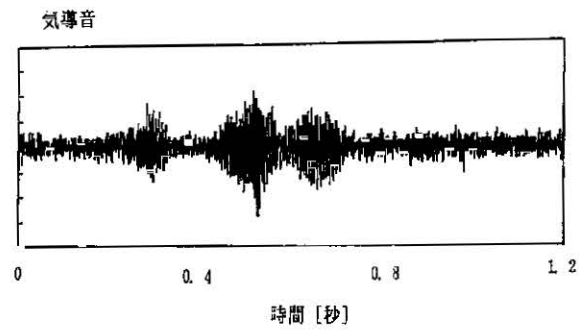


図21