(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

特開平11－305792

(43)公開日　平成11年(1999)11月５日

(51)Int.Cl.[6]		識別記号	ＦＩ		
Ｇ１０Ｌ	3/02	３０１	Ｇ１０Ｌ	3/02	３０１Ｅ
	9/00			9/00	Ｆ

審査請求　未請求　請求項の数37　ＯＬ　（全 22 頁）

(21)出願番号　特願平10－232565

(22)出願日　平成10年(1998)８月19日

(31)優先権主張番号　特願平10－39206

(32)優先日　平10(1998)２月20日

(33)優先権主張国　日本（ＪＰ）

(71)出願人　000004226
日本電信電話株式会社
東京都千代田区大手町二丁目３番１号

(72)発明者　高野　智大
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（ω_i）とＲ（ω_i）とのレベル差ΔＬＲ（ω_i）を検出し（５）、ΔＬＲ（ω_i）がしきい値Ｔｈ（ω_i）以上のω_iを選択し（６）、Ｌ（ω）からその成分Ｌ（ω_i）を抽出し、これを時間波形に変換する（８）、ΔＬＲ（ω_i）がＴｈ（ω_i）以下の成分Ｌ（ω_k）に対し、重みｗ（ω_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ΔＬＲ（ω）＞Ｔｈ（ω）であれば、そのΔＬＲ（ω）をＴｈ（ω）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。

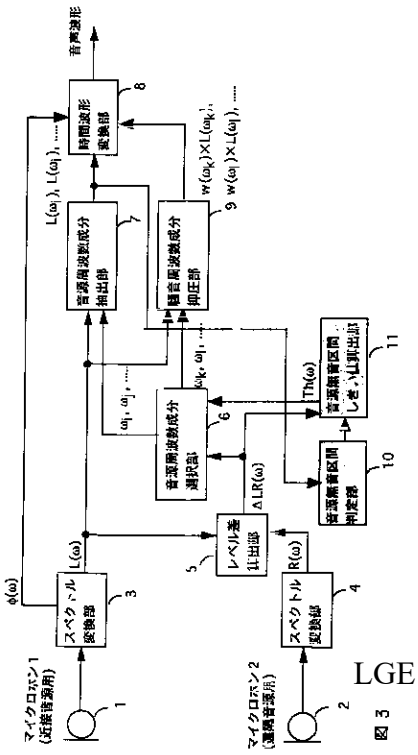


図３

LGE EXHIBIT 1018

Amazon v. Jawbone U.S. Patent 7,246,058
Amazon Ex. 1018

【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収音装置。
【請求項２】　請求項１記載の収音装置において、
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収音装置。
【請求項３】　請求項１または請求項２記載の収音装置において、
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収音装置。
【請求項４】　請求項２または請求項３記載の収音装置において、
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収音装置。
【請求項５】　請求項３または請求項４記載の収音装置において、
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収音装置。
【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記載の収音装置において、
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収音装置。
【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収音方法。
【請求項８】　請求項７記載の収音方法において、
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収音方法。
【請求項９】　請求項７または請求項８記載の収音方法において、
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収音方法。

【請求項１０】　請求項８または請求項９記載の収音方法において、
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収音方法。

【請求項１１】　請求項９または請求項１０記載の収音方法において、
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収音方法。

【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記載の収音方法において、
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収音方法。

【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収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収音装置、その収音機能を有し、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（ω_i）を求める（Ｓ０２）。そして、スペクトルＳ（ω_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、❶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、❷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】❶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、❶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】❷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、ｐ３３１－３３５、（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形態の収音装置において、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形態の収音装置において、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）９０ｄＢＡ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、それをディ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（ω_i）、Ｒ（ω_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（ω_i）、Ｒ（ω_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ΔＬＲ（ω_i）を計算する（Ｓ０４）。

ΔＬＲ（ω_i）＝２０ｌｏｇ１０（｜Ｌ（ω_i）｜／｜Ｒ（ω_i）｜）

上式中のω_iは周波数（ｉ＝１，２，…，ｎ）、｜Ｌ（ω_i）｜、｜Ｒ（ω_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ΔＬＲ（ω_i）と予め設定されたしきい値Ｔｈ（ω_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）。

ΔＬＲ（ω_i）＞Ｔｈ（ω_i）　→　目的信号が支配的

ΔＬＲ（ω_i）≦Ｔｈ（ω_i）　→　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（ω_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Φ（ω_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（ω_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、

ΔＬＲ（ω_i）≦Ｔｈ（ω_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（ω_i）について、外部より設定された重み係数ｗ（ω_i）を乗じた値を騒音抑圧処理後のスペクトル成分Ｓ（ω_i）とする（Ｓ０７）。

【００２８】Ｓ（ω_i）＝ｗ（ω_i）×Ｌ（ω_i）

ｗ（ω_i）の値は、騒音抑圧量を重視したい帯域で小さな値（ｗ（ω_i）≒０）に設定し、音質を重視したい帯域では大きな値（ｗ（ω_i）≒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（ω_i）とされる（Ｓ０６）。

【００２９】Ｓ（ω_i）＝Ｌ（ω_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（ω_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（ω_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（ω_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　→　目的信号の音源が有音状態

Ｐ≦ＰＴｈ　→　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（ω_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（ω_i）＝ΔＬＲ（ω_i）　（ΔＬＲ（ω_i）＞Ｔｈ（ω_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ほぼ等しくΔＬＲ（ω_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、ΔＬＲ（ω_i）が比較的大となる。この音源無音区間でのΔＬＲ（ω_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（ω_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ΔＬＲ（ω_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乗じる重み係数ｗ（ω_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ΔＬＲ（ω_i）としきい値Ｔｈ（ω_i）を利用して、重み係数ｗ（ω_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と判定された周波数成分に対して乗じる重み係数ｗ（ω_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（ω_i）は、騒音の重畳度が大きいときにはΔＬＲ（ω_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ΔＬＲ（ω_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（ω_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（ω_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ΔＬＲ（ω_i）≦Ｔｈ（ω_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ｅｐｓに変更する（Ｓ０７）。ｅｐｓ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　→　目的信号の音源が発声状態

ＰＢ　≦　ＰＴＨＢ　→　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、❶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、❷音声認識結果の出力、の２つの制御を行う。❶の制御については、以下の❶－１、❶－２に示す二つの方法が考えられる。

【００４３】❶－１：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

❶－２：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❶－１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遷移した瞬間（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❷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ーされる（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（ω_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（ω_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（ω_i）＝ΔＬＲ（ω_i）　（ΔＬＲ（ω_i）＞Ｔｈ（ω_i）のと

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ΔＬＲ（ω_i）としきい値Ｔｈ（ω_i）を利用して、重み係数ｗ（ω_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に示す収音条件、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収音装置が得られる。

【００５４】第２形態の発明によれば、第１形態の収音装置において、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収音装置が得られる。

【００５５】第３形態の発明によれば、第１形態または第２形態の収音装置において、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収音装置が得られる。

【００５６】第４形態の発明では、第２形態または第３形態の収音装置において、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収音装置が得られる。

【００５７】第５形態の発明では、第３形態または第４形態の収音装置において、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収音装置が得られる。

【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形態の収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収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収音であり、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】この発明における収音装置は、騒音抑圧が必要な各種収音装置のほか、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ュミ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）９０ｄＢＡ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】

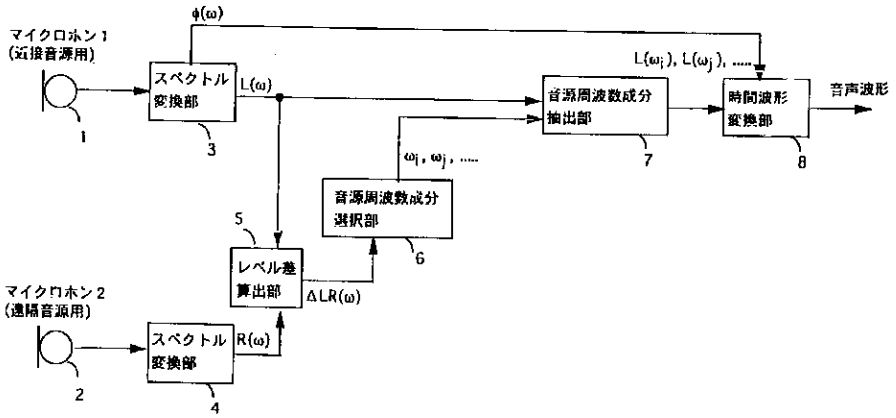


図１

【図２】

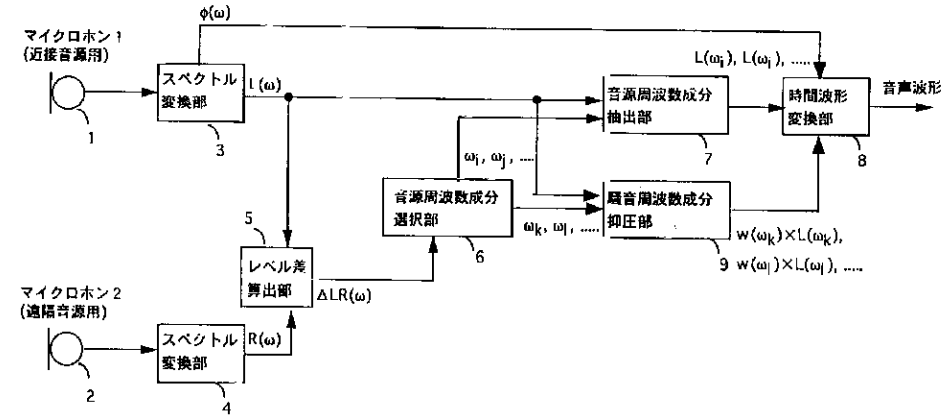


図２

【図３】

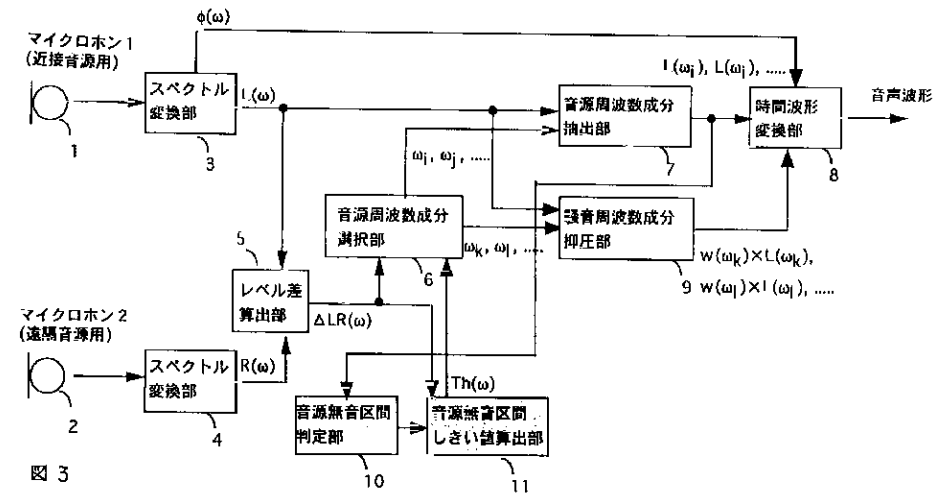


図３

【図４】

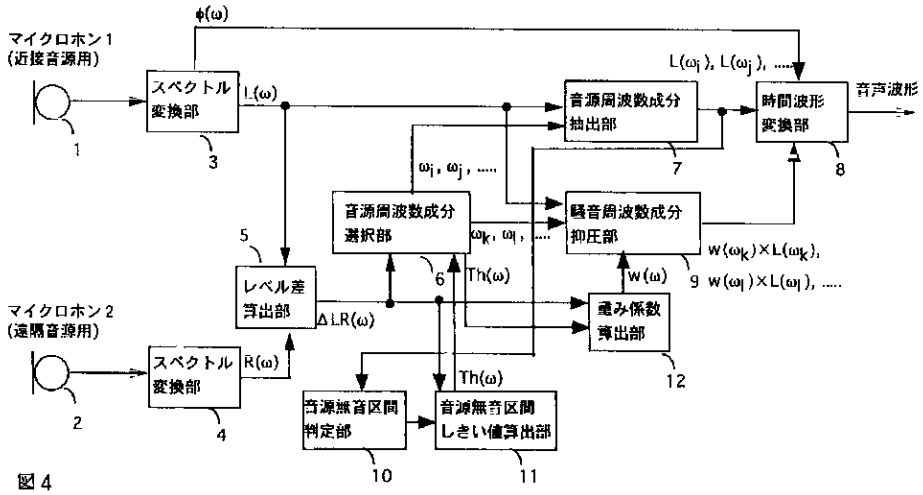

φ(ω)
マイクロホン１
(近接音源用)
スペクトル
変換部
L(ω)
1
3
L(ωi), L(ωj), ...
音源周波数成分
抽出部
7
時間波形
変換部
8
音声波形
ωi, ωj, ...
音源周波数成分
選択部
ωk, ωl, ...
6
5
騒音周波数成分
抑圧部
w(ωk)×L(ωk),
9 w(ωl)×L(ωl), ...
w(ω)
Th(ω)
レベル差
算出部
ΔLR(ω)
重み係数
算出部
12
マイクロホン２
(遠隔音源用)
スペクトル
変換部
R(ω)
2
4
Th(ω)
音源無音区間
判定部
音源無音区間
しきい値算出部
10
11

図４

【図５】

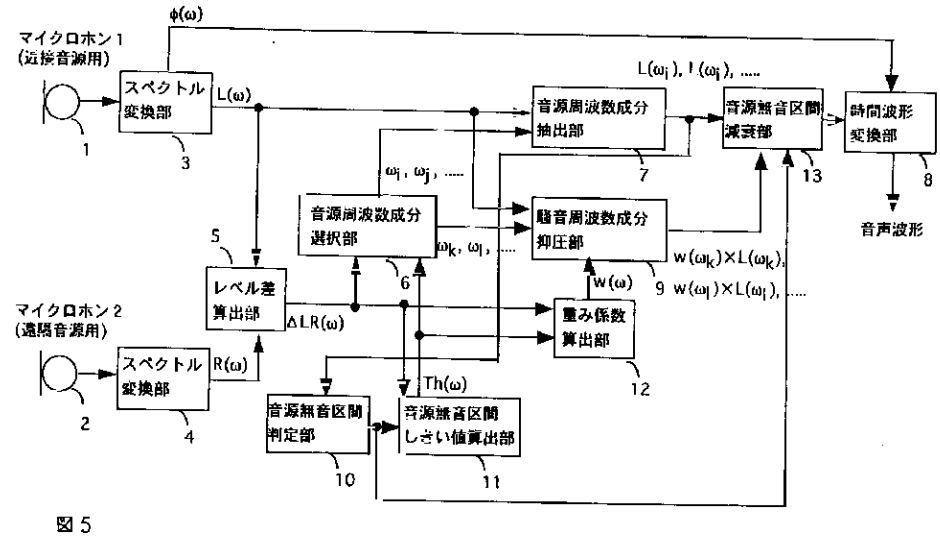

φ(ω)
マイクロホン１
(近接音源用)
スペクトル
変換部
L(ω)
1
3
L(ωi), L(ωj), ...
音源周波数成分
抽出部
7
音源無音区間
減衰部
13
時間波形
変換部
8
音声波形
ωi, ωj, ...
音源周波数成分
選択部
ωk, ωl, ...
6
5
騒音周波数成分
抑圧部
w(ωk)×L(ωk),
9 w(ωl)×L(ωl), ...
w(ω)
レベル差
算出部
ΔLR(ω)
重み係数
算出部
12
マイクロホン２
(遠隔音源用)
スペクトル
変換部
R(ω)
2
4
Th(ω)
音源無音区間
判定部
音源無音区間
しきい値算出部
10
11

図５

【図６】

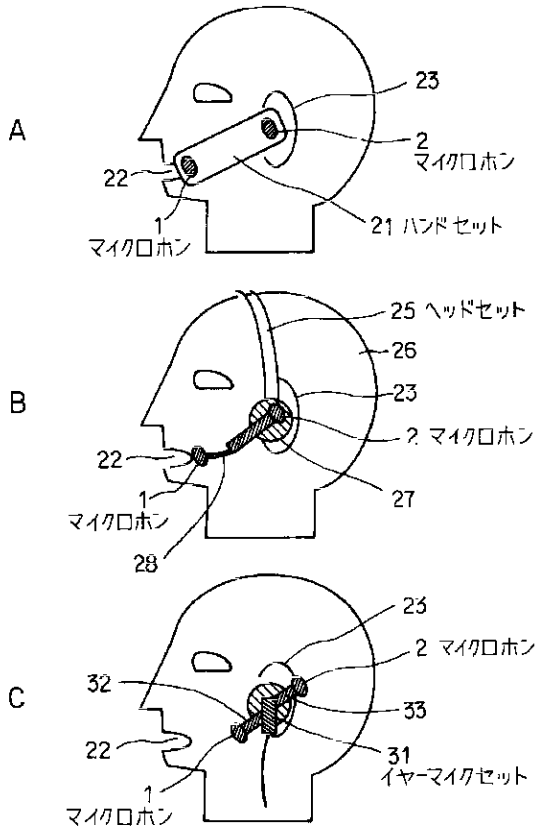


図６

【図７】

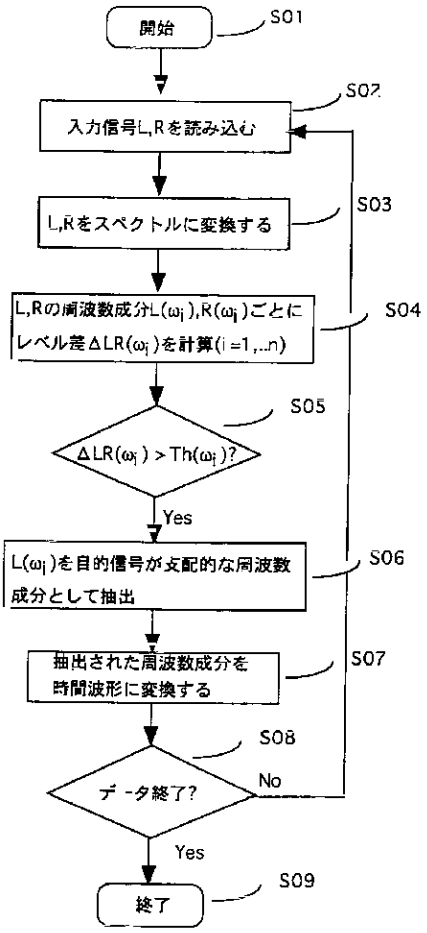


図７

【図１２】

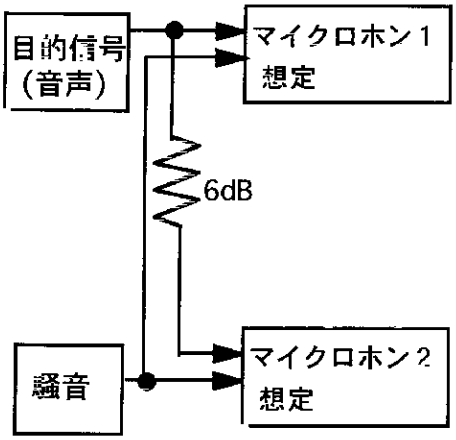


図12

【図１９】

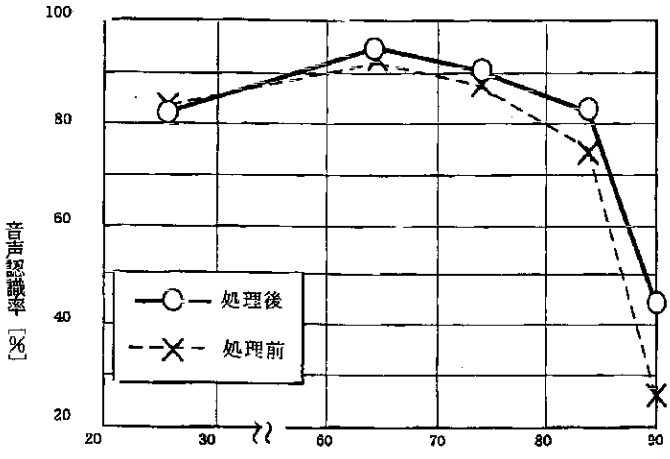


図１９

【図8】

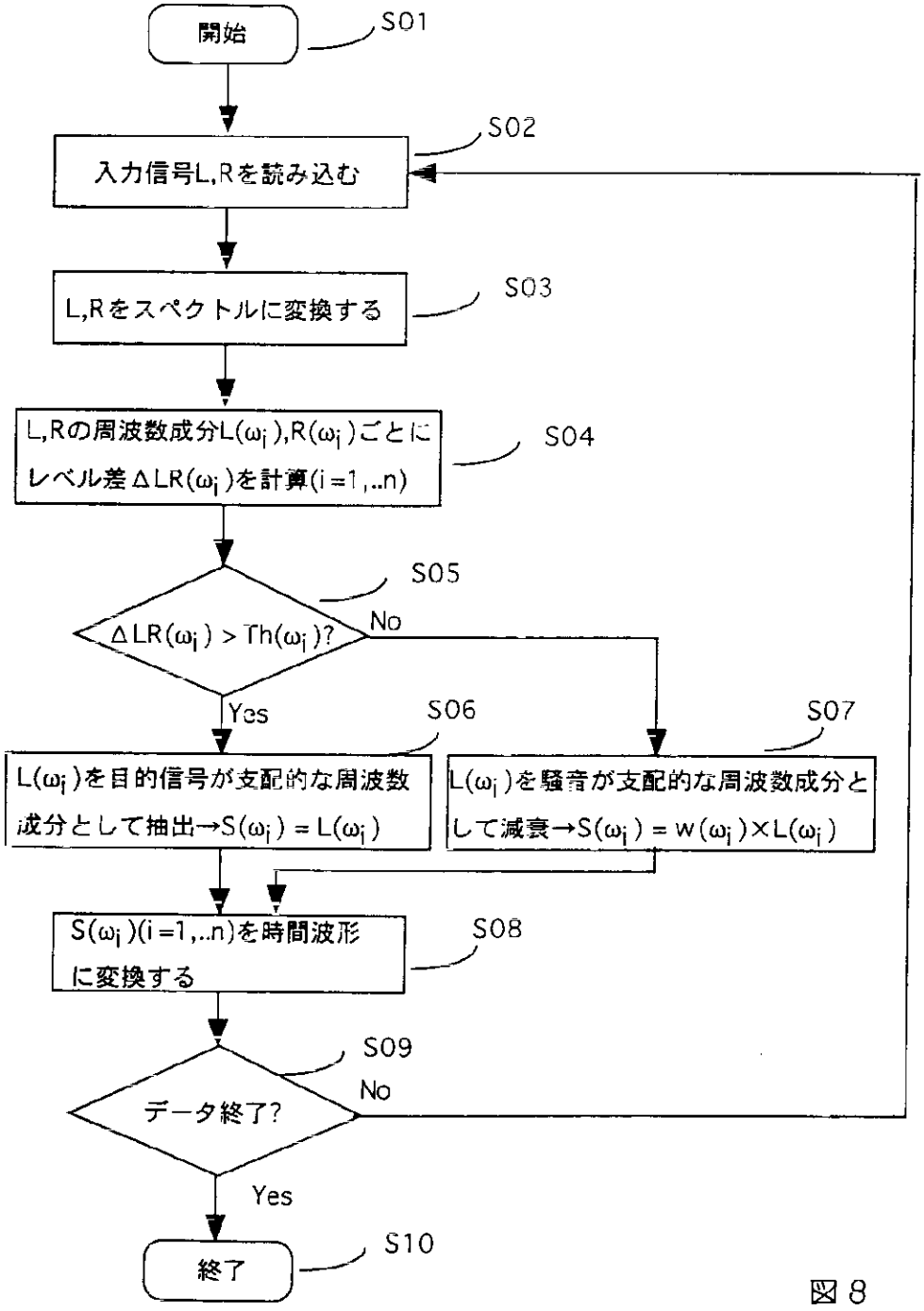

開始
S01
入力信号L,Rを読み込む
S02
L,Rをスペクトルに変換する
S03
L,Rの周波数成分L(ωi),R(ωi)ごとに
レベル差ΔLR(ωi)を計算(i=1,..n)
S04
ΔLR(ωi) > Th(ωi)?
S05
No
Yes
L(ωi)を目的信号が支配的な周波数
成分として抽出→S(ωi) = L(ωi)
S06
L(ωi)を騒音が支配的な周波数成分と
して減衰→S(ωi) = w(ωi)×L(ωi)
S07
S(ωi)(i=1,..n)を時間波形
に変換する
S08
データ終了?
S09
No
Yes
終了
S10

図8

【図９】

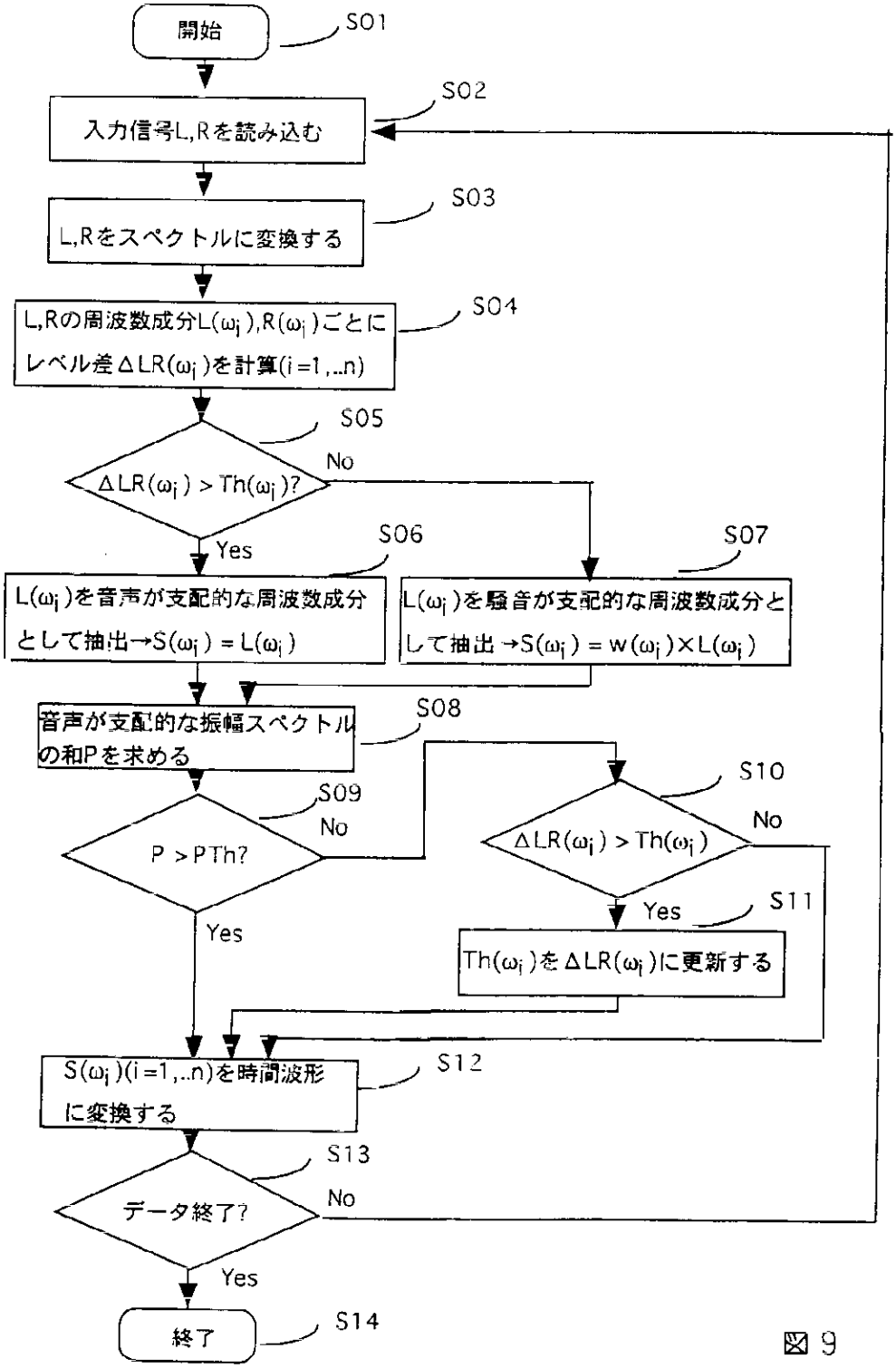

開始
S01
入力信号L,Rを読み込む
S02
L,Rをスペクトルに変換する
S03
L,Rの周波数成分L(ωi),R(ωi)ごとにレベル差ΔLR(ωi)を計算(i=1,..n)
S04
ΔLR(ωi) > Th(ωi)?
S05
No
Yes
L(ωi)を音声が支配的な周波数成分として抽出→S(ωi) = L(ωi)
S06
L(ωi)を騒音が支配的な周波数成分として抽出→S(ωi) = w(ωi)×L(ωi)
S07
音声が支配的な振幅スペクトルの和Pを求める
S08
P > PTh?
S09
No
Yes
ΔLR(ωi) > Th(ωi)
S10
No
Yes
Th(ωi)をΔLR(ωi)に更新する
S11
S(ωi)(i=1,..n)を時間波形に変換する
S12
データ終了?
S13
No
Yes
終了
S14

図９

【図１０】

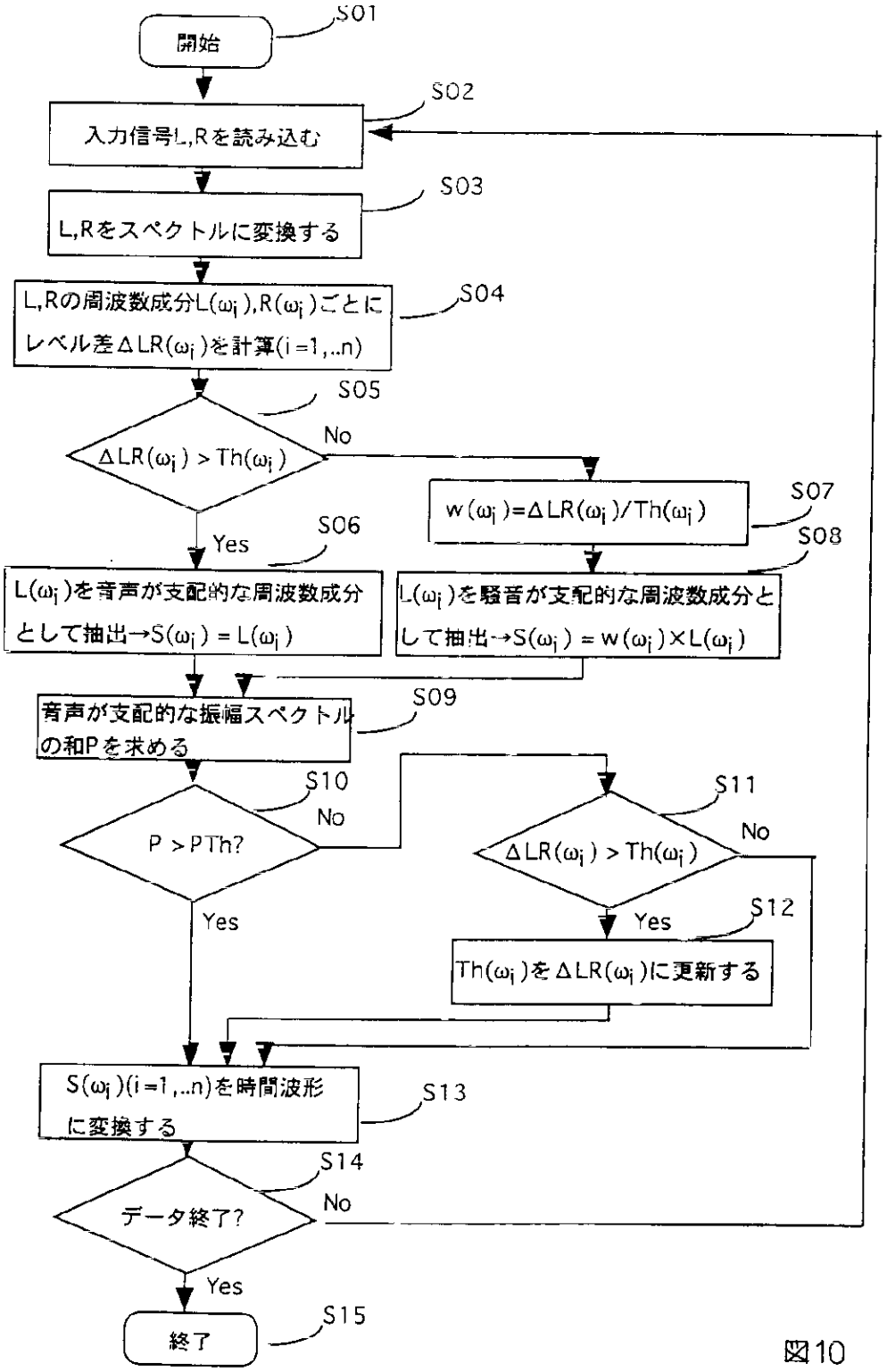

開始
S01
入力信号L,Rを読み込む
S02
L,Rをスペクトルに変換する
S03
L,Rの周波数成分L(ωi),R(ωi)ごとにレベル差ΔLR(ωi)を計算(i=1,..n)
S04
ΔLR(ωi) > Th(ωi)
S05
No
Yes
w(ωi)=ΔLR(ωi)/Th(ωi)
S07
S06
S08
L(ωi)を音声が支配的な周波数成分として抽出→S(ωi) = L(ωi)
L(ωi)を騒音が支配的な周波数成分として抽出→S(ωi) = w(ωi)×L(ωi)
音声が支配的な振幅スペクトルの和Pを求める
S09
P > PTh?
S10
No
Yes
ΔLR(ωi) > Th(ωi)
S11
No
Yes
S12
Th(ωi)をΔLR(ωi)に更新する
S(ωi)(i=1,..n)を時間波形に変換する
S13
データ終了?
S14
No
Yes
終了
S15

図10

【図１１】

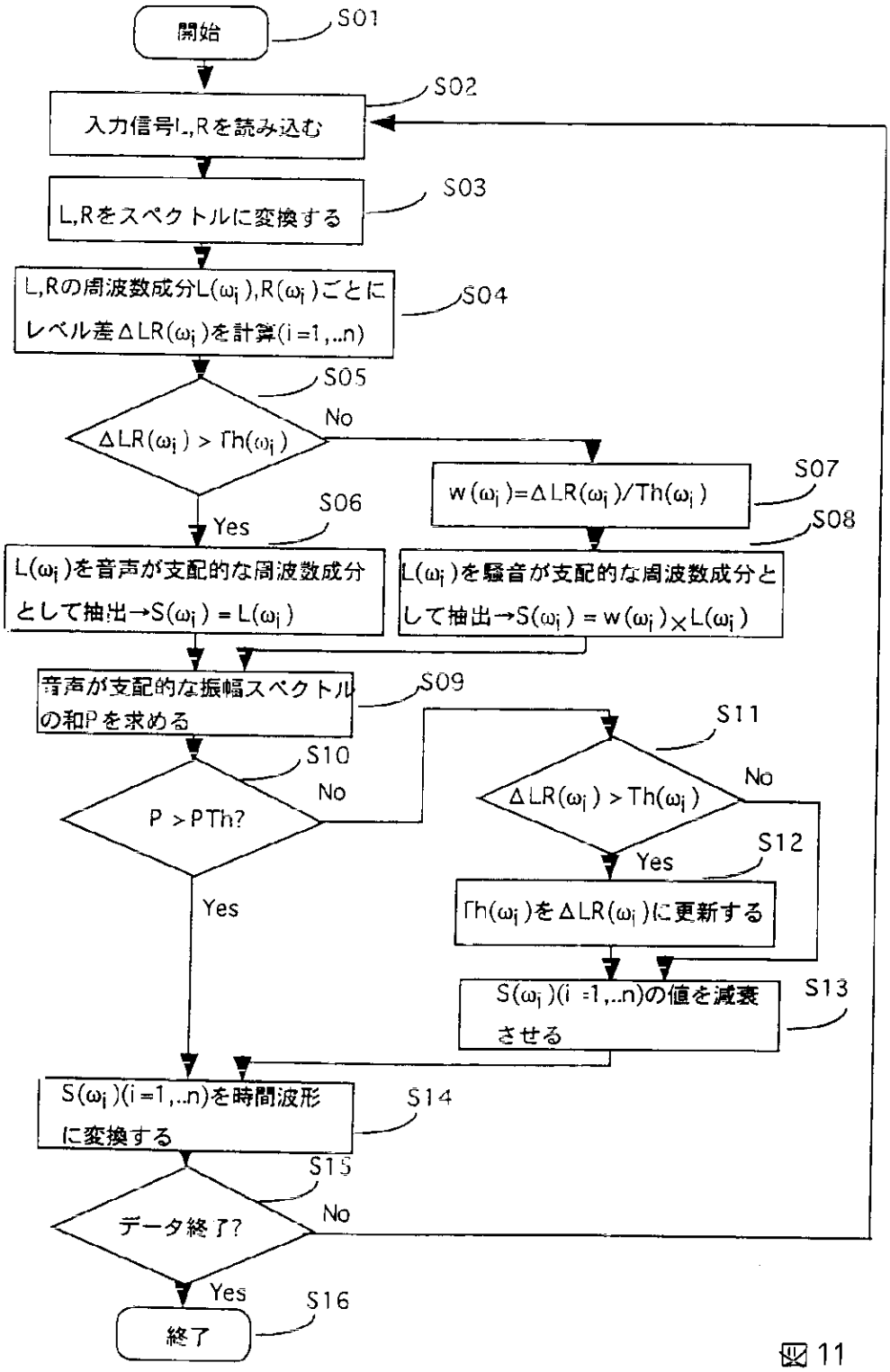

開始
S01
入力信号L,Rを読み込む
S02
L,Rをスペクトルに変換する
S03
L,Rの周波数成分L(ωi),R(ωi)ごとにレベル差ΔLR(ωi)を計算(i=1,..n)
S04
S05
ΔLR(ωi) > Th(ωi)
No
Yes
S06
S07
w(ωi)=ΔLR(ωi)/Th(ωi)
S08
L(ωi)を音声が支配的な周波数成分として抽出→S(ωi) = L(ωi)
L(ωi)を騒音が支配的な周波数成分として抽出→S(ωi) = w(ωi)×L(ωi)
音声が支配的な振幅スペクトルの和Pを求める
S09
S10
P > PTh?
No
Yes
S11
ΔLR(ωi) > Th(ωi)
No
Yes
S12
Th(ωi)をΔLR(ωi)に更新する
S(ωi)(i =1,..n)の値を減衰させる
S13
S(ωi)(i=1,..n)を時間波形に変換する
S14
S15
データ終了?
No
Yes
S16
終了

図11

【図１３】

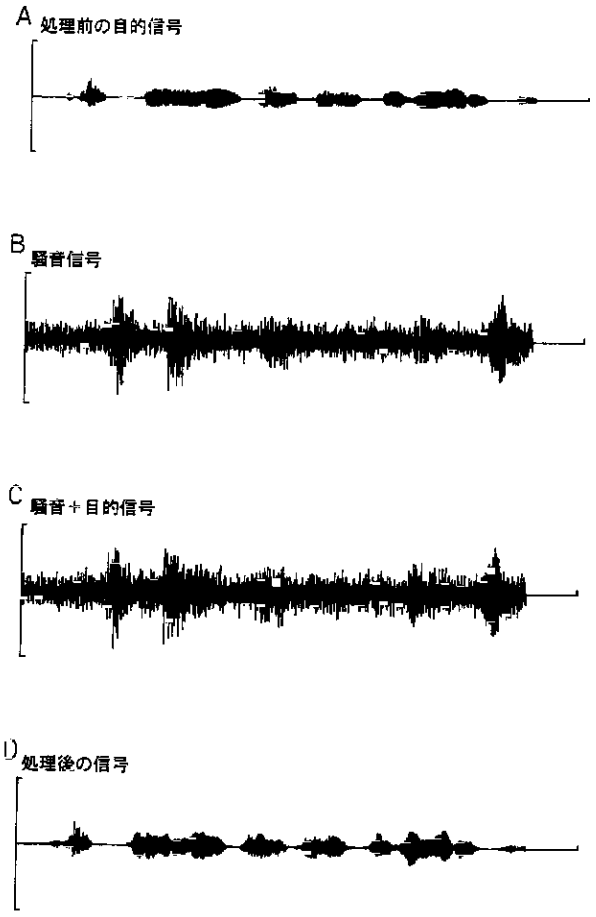


図13

【図１５】

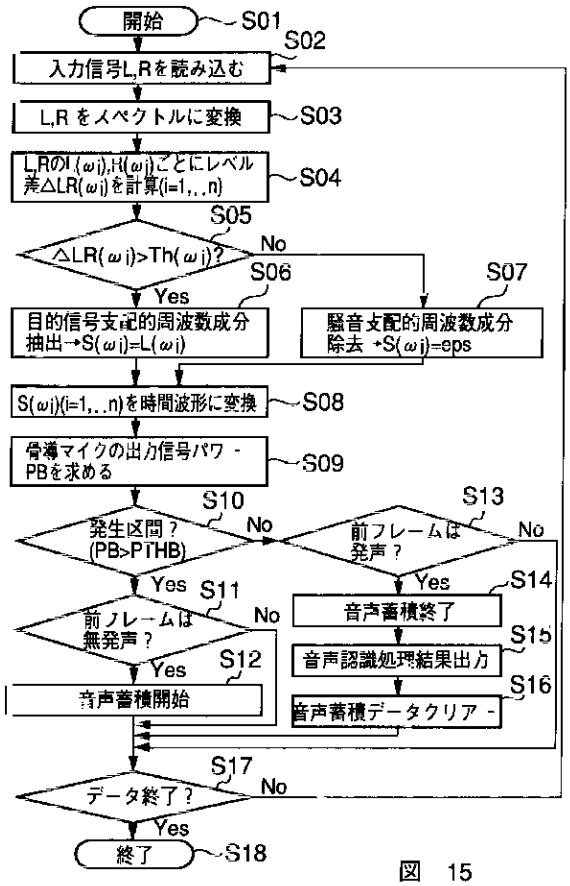


図 15

【図１４】

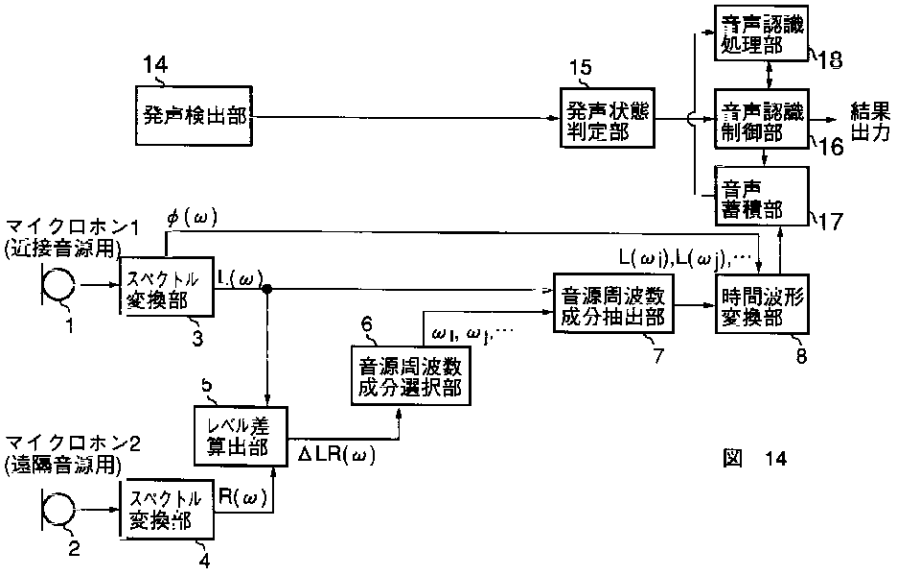


図 14

【図１６】

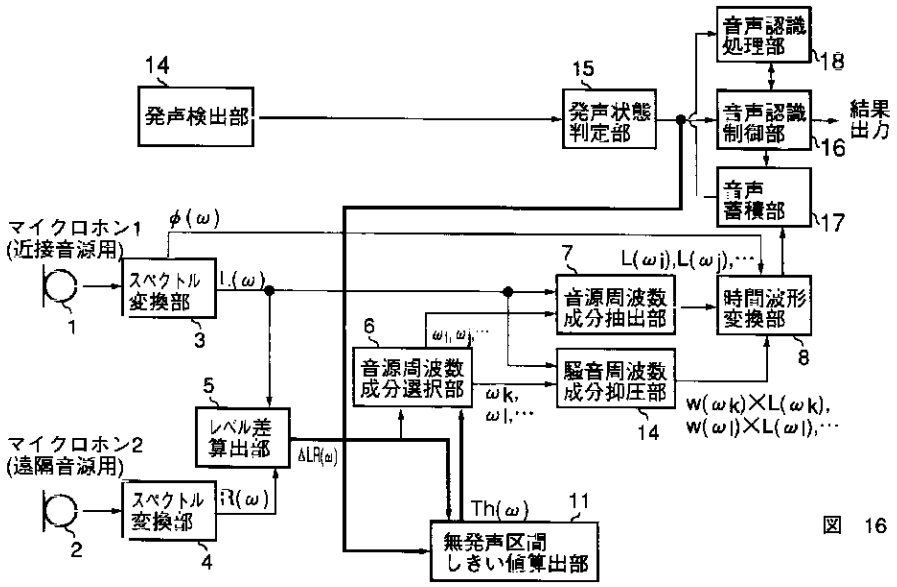


図　16

【図１７】

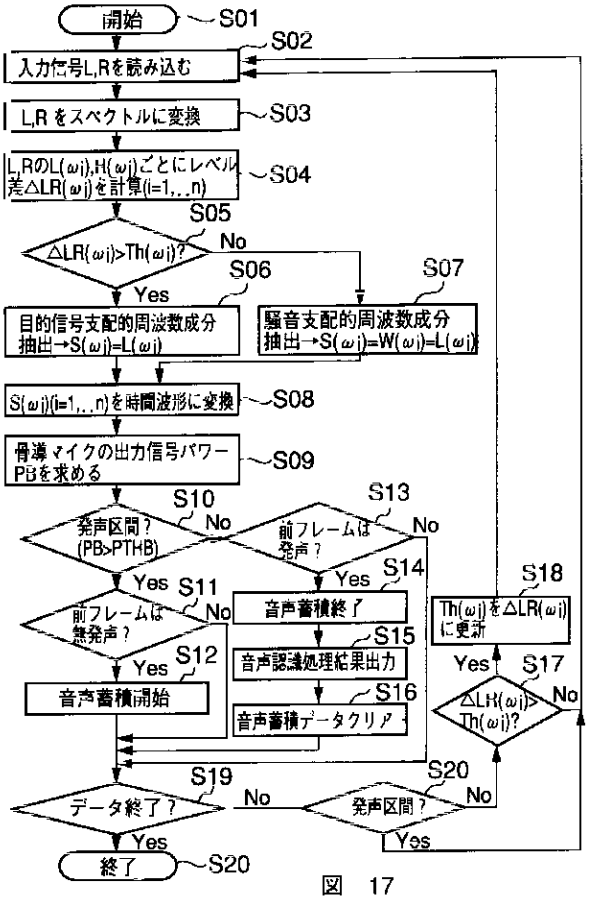


図　17

【図１８】

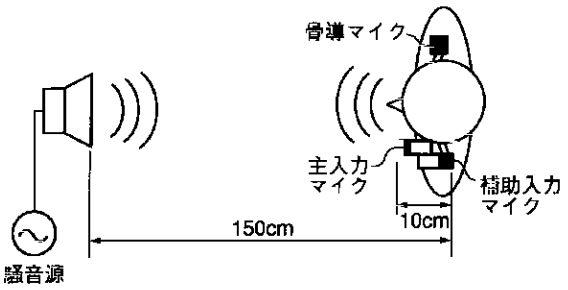


図　18A

発声条件	・被験者: 男性3名 ・日本の都市名:20単語
騒音条件	・騒音源:ピンクノイズ ・騒音レベル:26(騒音源停止時),64,74,84,90dBA
音声認識処理	・簡易型DPマッチングを利用した単語音声認識[1] ・LPCケプストラム分析を使用 ・登録時の騒音レベルは64dBA

[1]中台,杉本,桜井,西野,信学総合大,D-675,1996.3を参照

図　18B

【図２０】

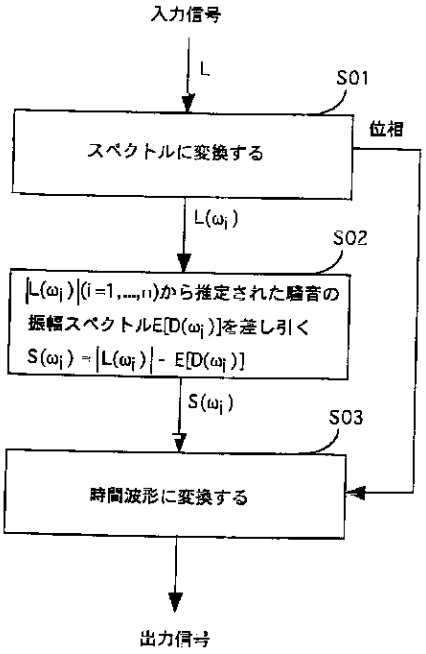


図20

【図２１】

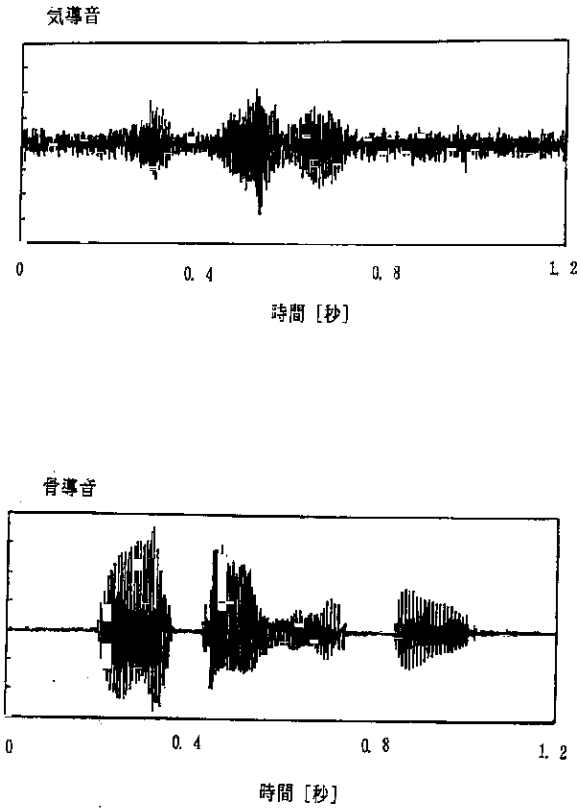


図２１