

(19) 世界知的所有権機関
国際事務局



(43) 国際公開日
2007年2月15日 (15.02.2007)

PCT

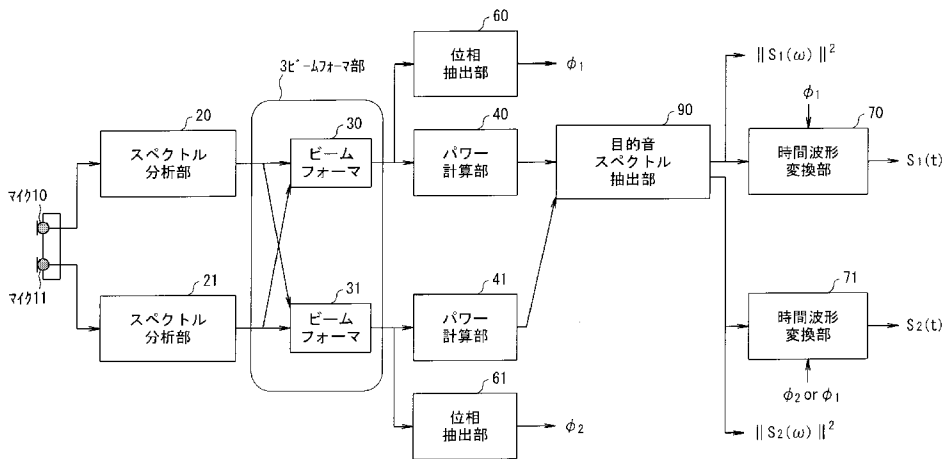
(10) 国際公開番号
WO 2007/018293 A1

- (51) 国際特許分類: *G10L 21/02* (2006.01) *G10L 15/22* (2006.01) *G10L 15/00* (2006.01) *H04R 1/40* (2006.01) *G10L 15/20* (2006.01) *H04R 3/00* (2006.01)
- (21) 国際出願番号: PCT/JP2006/315953
- (22) 国際出願日: 2006年8月11日 (11.08.2006)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (30) 優先権データ: 特願2005-233195 2005年8月11日 (11.08.2005) JP
- (71) 出願人 (米国を除く全ての指定国について): 旭化成株式会社 (ASAHI KASEI KABUSHIKI KAISHA)
- [JP/JP]; 〒5308205 大阪府大阪市北区堂島浜1丁目2番6号 Osaka (JP).
- (72) 発明者; および
(75) 発明者/出願人 (米国についてののみ): 長濱 克昌 (NAGAHAMA, Katsumasa) [JP/JP]; 〒1008440 東京都千代田区有楽町一丁目1番2号 Tokyo (JP). 松井 信也 (MATSUI, Shinya) [JP/JP]; 〒1008440 東京都千代田区有楽町一丁目1番2号 Tokyo (JP).
- (74) 代理人: 森 哲也, 外 (MORI, Tetsuya et al.); 〒1010032 東京都千代田区岩本町二丁目3番3号 友泉岩本町ビル8階 日栄国際特許事務所 Tokyo (JP).
- (81) 指定国 (表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HN,

[続葉有]

(54) Title: SOUND SOURCE SEPARATING DEVICE, SPEECH RECOGNIZING DEVICE, PORTABLE TELEPHONE, AND SOUND SOURCE SEPARATING METHOD, AND PROGRAM

(54) 発明の名称: 音源分離装置、音声認識装置、携帯電話機、音源分離方法、及び、プログラム



10, 11... MICROPHONE
 20, 21... SPECTRUM ANALYZING SECTION
 3... BEAM FORMER SECTION
 30, 31... BEAM FORMER
 60, 61... PHASE EXTRACTING SECTION
 40, 41... POWER CALCULATING SECTION
 90... TARGET SOUND SPECTRUM EXTRACTING SECTION
 70, 71... TEMPORAL WAVEFORM TRANSFORMING SECTION

(57) Abstract: The sound source signal from a target sound source is separated from the mixed sound in which the sound source signals from sound sources are mixed without being influenced by the variations of the sensitivity of microphone elements. A beam former section (3) of the sound source separating device (1) performs beam formation to attenuate the sound source signals arriving from directions symmetrical with respect to the vertical line of the line connecting two microphones (10, 11) by spectrum-analyzing the output signals from the microphones (10, 11) and multiplying the signals after the weighting factors complex conjugated with the signals. Power calculating sections (40, 41) calculate power spectrum information. Target sound spectrum extracting sections (50, 51) extract spectrum information on target sound sources according to the difference between the power spectrum information from one beam former and that from the other.

(57) 要約: マイクロホン素子の感度のばらつきに影響を受けることなく、複数の音源から発せられた音源信号が混合された混合音の中から、目的音源からの

[続葉有]

WO 2007/018293 A1



HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

添付公開書類:

— 国際調査報告書

(84) 指定国 (表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), ヨーロッパ (AT, BE, BG,

2文字コード及び他の略語については、定期発行される各PCTガゼットの巻頭に掲載されている「コードと略語のガイダンスノート」を参照。

音源信号を分離することを可能とする。音源分離装置1のビームフォーマ部3は、スペクトル分析後のマイクロホン10、11からの出力信号に対して複素共役の関係にある重み係数を乗算することにより、2つのマイクロホン10、11を結ぶ直線の垂線に対称な方向から到来する音源信号を各々減衰させるためのビームフォーマ処理を行う。パワー計算部40、41はパワースペクトル情報を計算し、目的音スペクトル抽出部50、51はパワースペクトル情報同士の差分に基づいて目的音源のスペクトル情報を抽出する。

明 細 書

音源分離装置、音声認識装置、携帯電話機、音源分離方法、及び、プログラム

技術分野

[0001] 本発明は、複数のマイクロホンを使用し、複数の音源から発せられた複数の音声信号や各種環境雑音など複数の音響信号が混ざった信号から、目的とする音源から到来する音源信号を分離する音源分離装置、音声認識装置、携帯電話機、音源分離方法、及び、プログラムに関する。

背景技術

[0002] 種々の環境下において特定の音声信号などを収録したい場合、周囲環境にはさまざまな雑音源があるため、目的音とする信号のみをマイクロホンで収録することは困難であり、何らかの雑音低減処理あるいは音源分離処理が必要となる。

これらの処理が特に必要となる例として、例えば自動車環境下が挙げられる。自動車環境下において、携帯電話の普及により運転中の携帯電話を使用しての通話は車内に離れて設置されたマイクを使用するのが一般的あり、通話品質を著しく劣化させている。また、自動車環境下で運転中に音声認識を行う場合も同様の状況で発話するため、音声認識性能を劣化させる原因になっている。現在の音声認識技術の進歩により、定常雑音に対する音声認識率の劣化の問題に対して、劣化した性能のかなりの部分を回復することが可能となっている。しかし、現状の音声認識技術で対応が難しいものとして、複数発話者の同時発話時の認識性能の劣化の問題がある。現在の音声認識の技術では同時に発話された二人の混合音声を認識する技術が低いため、音声認識装置使用時には発話者以外の同乗者は発話を制限され、同乗者の行動を制限する状況が発生している。これらの音源分離の手法として独立主成分分析法などを使用する方法もあるが、演算量、音源数の変動などの問題があり実用化されていない。

[0003] 上記の問題を解決するために、車内の複数のマイクを取り付け特定方向からの音声のみを収録する方法が各種提案されているが、自動車内に多数のマイクを取り付

けるためのスペースを確保することは困難であり、加えてコスト的な問題から特性の揃ったマイクを使用することも難しい。このため、出来るだけ少数で特性バラつきがあるマイクを使用して動作をする方式が望まれている。

一般に複数のマイクロホンを使用する場合、低コストなマイクロホンほどその感度特性のバラつきが大きく約±3dB程度各周波数特性にバラつきがあると言われている。これらの特性のバラつきがマイクロホンアレイ技術の中でも遅延和アレイなどの加算型アレイ処理ではマイクロホンアレイ性能が設計通りの特性が実現できないだけであるが、適応アレイなどのいわゆる減算型アレイでは一つのマイクを使用する場合より特に1kHz以下程度の低域において性能を悪化させる場合がある。

[0004] マイクロホンアレイ技術にとってセンサとしてのマイクロホンの特性のバラつきは大きな問題であり、これらの対策として、特許文献1～5などに複数のマイク素子の感度を揃える方法が提案されている。

従来、少数のマイクロホン数で大きな効果が得られる適応ビームフォーマ処理技術を利用したマイクロホンアレイに関しては、例えば非特許文献1や、非特許文献2に述べられているように、一般化サイドローブキャンセラ(GSC)、フロスト型ビームフォーマおよび参照信号法など、種々の方法が知られている。

[0005] 適応ビームフォーマ処理は、基本的には雑音源の方向に死角を有する指向性ビームを形成したフィルタにより雑音を抑圧する処理であり、その中でも特に一般化サイドローブキャンセラは、比較的性能が良いことで知られている。しかし、GSCでは、設定された目的音源の方向とずれた方向から目的信号が到来した際に、目的信号がキャンセルされて劣化するという問題点がある。これに対し、特許文献6、7において、これを周波数領域で動作させることで計算量を削減すると共に、周波数領域のフィルタ係数から話者方向と特定雑音方向とを逐次検知し、目的音と目的音以外の雑音とをある程度分離し、スペクトルサブトラクションと併用することで到来方向不明な雑音や拡散性の雑音を低減する方法が開示されている。

[特許文献1]特開平5-131866号公報

[特許文献2]特開2002-99297号公報

[特許文献3]特開2003-153372号公報

[特許文献4]特開2004-343700号公報

[特許文献5]特開2004-289762号公報

[特許文献6]特開2001-100800号公報

[特許文献7]特開2000-47699号公報

[非特許文献1]電子情報通信学会編「音響システムとデジタル処理」

[非特許文献2]H.Aykin著、“ADaptive Filter Theory(Prentice Hall)”

発明の開示

[0006] [発明が解決しようとする課題]

この特許文献6、7に記載されているような適応ビームフォーマとスペクトルサブトラクションを組み合わせた技術において、マイクロホンの素子感度のバラつきが存在する場合、適応フィルタ部の入力信号として目的音を抑圧した参照信号を作成する必要があるが、目的音を十分に抑圧することが出来ず、適応フィルタ部の参照信号と目的信号双方に目的音が含まれるため、適応フィルタ処理を施した結果目的音が歪み音質が劣化する。同様の現象は、初期反射の大きな閉空間でも参照信号中に目的音信号が漏れてしまい発生する。

これを解決するために、マイクロホンの素子感度補正をするために製品組み立て時にバラつきを測定し補正データを作成して使用時に補正を行ったり、使用時に特定方向からの基準信号を利用してマイクロホンの素子感度の個体差、周囲温度や部品の経年変化による逐次変化を測定し補正する方法があるが、製造コストの増加や、基準信号をいつ再生させるか、あるいは収録した信号が本当に基準信号のみかどうかをどうやって判断するかなどの問題がある。

[0007] また適応フィルタの動作に着目すると、一般的に適応フィルタで参照信号と目標信号が高い相関性を持つ場合、適応フィルタの推定動作が上手く行えず推定精度を著しく劣化させてしまう。一例を挙げると、目的音と雑音の双方が音声信号の場合などである。

同様の問題はエコーキャンセラでも発生し、遠端信号から近端信号に混入したエコーを推定する適応フィルタにおいて、遠端、近端とも同時に2人の人が発話をした場合、いわゆるダブルトーク状態が発生しエコーキャンセラの動作が著しく劣化する。こ

のため、ダブルトーク状態を如何に正確に検出し、適応フィルタのパラメータ推定を停止させたり推定速度も遅くしたりするかが重要な性能維持の条件となっている。

[0008] したがって、本適応フィルタ動作においても、目的音と特定雑音が双方とも音声である場合、双方とも発話している状態を検出する必要があると同時に、双方とも発話している頻度が高いほど適応フィルタ推定を停止させる頻度があがり、結果として適応フィルタの雑音除去性能が劣化するという問題がある。

また、自動車の走行雑音など大きな拡散性雑音などが発生する状況においては拡散性雑音の影響で適応フィルタの推定精度が劣化するため、結果として、適応フィルタの推定したフィルタパラメータを使用した話者方向、特定雑音方向の到来方向推定精度が劣化し、システム全体の性能が劣化するという問題がある。

[0009] また、上記の特許文献6、7に記載された方式では、自動車向けなどに低コストのマイクを使用して雑音成分抑圧装置を構成しようとした場合、車内音源の初期反射が大きい、拡散性の雑音成分が大きい、使用するマイク素子のバラつきが大きい(±3dB程度)、運転者と同乗者とが同時に発話し相関性の強い目的音と特定雑音とが同時に存在するなどの場合に、適応フィルタ部が所望の動作をせず、全体として所望の動作を実現できない。

本発明は上記問題点に鑑みてなされたものであり、マイクロホン素子の感度のばらつきに影響を受けることなく、複数の音源から発せられた音源信号が混合された混合音の中から、目的音源からの音源信号を分離することを可能とする音源分離装置、音声認識装置、携帯電話機、音源分離方法、及び、プログラムを提供することを目的とする。

[課題を解決するための手段]

上記課題を解決するために、請求項1に記載の発明は、互いに離して配置された少なくとも2つのマイクロホンに対して複数の音源から発せられた音源信号が混合された混合音を入力し、該混合音から目的音源からの音源信号を分離する音源分離装置において、前記マイクロホンからの出力信号に対して第1の係数を用いた演算を行うことにより、所定の方向から到来する音源信号を減衰させるための第1のビームフォーマ処理を行い、かつ、前記マイクロホンからの出力信号に対して、前記第1の係

数と周波数領域で複素共役の関係にある第2の係数を用いた演算を行うことにより、2つのマイクロホンを結ぶ直線の垂線に対して前記所定の方角と対称な方角から到来する音源信号を減衰させるための第2のビームフォーマ処理を行うビームフォーマ手段と、前記ビームフォーマ手段により得られた音源信号各々についてのパワースペクトル情報を計算するパワー計算手段と、前記パワー計算手段により計算されたパワースペクトル情報同士の差分に基づいて、目的とする音源のスペクトル情報を抽出する目的音スペクトル抽出手段とを備えることを特徴とする音源分離装置を提供する。

[0010] この発明によれば、音源分離装置は、指向特性がマイクロホン素子の感度に影響を受けないという性質を実現することができ、マイクロホン素子の感度のばらつきに影響を受けることなく、複数の音源から発せられた音源信号が混合された混合音の中から、目的音源からの音源信号を分離することが可能となる。

請求項2に記載の発明は、請求項1に記載の音源分離装置において、前記ビームフォーマ手段は、互いに離して配置された3つのマイクロホンのうち、何れか2つのマイクロホンの組合せ及び他の2つのマイクロホンの組合せ各々について、前記第1のビームフォーマ処理及び前記第2のビームフォーマ処理を行うことを特徴とする。

[0011] この発明によれば、3つのマイクロホンのうち、2つのマイクロホンを結ぶ直線の垂線を境界とした各領域に存在する目的音源からの音源信号を抽出することができる。

請求項3に記載の発明は、請求項1又は2に記載の音源分離装置において、マイクロホンからの出力信号に対して遅延を与える指向性制御手段をさらに備えることを特徴とする。

この発明によれば、遅延を与えることで指向特性を最適な状態に補正することができ、音源の分離性能を向上させることができる。

[0012] 請求項4に記載の発明は、請求項3に記載の音源分離装置において、前記指向性制御手段は、2つのマイクロホンのうち少なくとも一方のマイクロホンからの出力信号に対して遅延を与えることにより、仮想的に3つのマイクロホンからの出力信号を生成することを特徴とする。

この発明によれば、仮想的に3つのマイクロホンからの出力信号を生成することがで

きるため、2つのマイクロホンのみを用いて、2つのマイクロホンを結ぶ直線の垂線方向から到来する音源信号を分離抽出することが可能となる。

- [0013] 請求項5に記載の発明は、請求項3又は4に記載の音源分離装置において、前記音源信号が到来する方向を推定する到来方向推定手段をさらに備え、前記指向性制御手段は、前記到来方向推定手段により推定された到来方向に基づいて、2つの音源の位置が仮想的に2つのマイクロホンを結ぶ直線の垂線に対して対称となるように、該マイクロホンからの出力信号に対して遅延を与えることを特徴とする。

この発明によれば、音源の高い分離性能が得られるように遅延操作を行うことが可能となる。

- [0014] 請求項6に記載の発明は、請求項1から5の何れか1項に記載の音源分離装置において、前記目的音抽出手段により抽出されたパワースペクトル情報に対して、スペクトルサブトラクション処理を行うスペクトルサブトラクション手段をさらに備えたことを特徴とする。

この発明によれば、スペクトルサブトラクション処理を行うことにより到来方向不明の定常雑音や拡散性の雑音などを除去することが可能となる。

- [0015] 請求項7に記載の発明は、請求項1から6の何れか1項に記載の音源分離装置において、前記ビームフォーマ手段による処理を行う前に、雑音を低減する処理を行う定常雑音低減手段をさらに備えることを特徴とする。

この発明によれば、ミュージカルノイズのような聴感上不快な雑音の発生を低減することが可能となる。

請求項8に記載の発明は、請求項1から7の何れか1項に記載の音源分離装置により分離された音源信号の音声認識を行う音声認識手段を備えた音声認識装置を提供する。

- [0016] この発明によれば、精度よく分離された音源信号に基づいて、精度のよい音声認識を行うことができる。

請求項9に記載の発明は、請求項8に記載の音声認識装置において、車両の運転席側から発せられる語彙の候補の一覧である運転席側認識語彙リスト及び助手席側から発せられる語彙の候補の一覧である助手席側認識語彙リストを記憶する認識語

彙リスト記憶手段をさらに備え、前記音声認識手段は、前記認識語彙リスト記憶手段に記憶されている運転席側認識語彙リスト及び助手席側認識語彙リストに基づいて、前記音源分離装置により分離された音源信号の音声認識処理を行うことを特徴とする。

[0017] この発明によれば、音声認識装置は、認識語彙リスト記憶手段に記憶されている助手席側認識語彙リスト及び運転席側認識語彙リストに基づき音声認識処理を行うため、運転席、助手席別に語彙リストの中から最適な語彙を選択することができ、精度よく音声認識を行うことが可能となる。

請求項10に記載の発明は、請求項8又は9に記載の音声認識装置において、現在の車両の状態を管理する状態遷移手段と、車両の状態に応じた助手席側及び運転席側における有効な語彙リストを記憶する有効語彙リスト記憶手段と、前記状態遷移手段で管理されている現在の車両の状態と、前記有効語彙リスト記憶手段に記憶されている語彙リストとに基づいて、前記音声認識手段により認識された語彙が有効か否かを判断し、該判断結果に応じて制御を行う制御手段とをさらに備えることを特徴とする。

[0018] この発明によれば、現在の車両の状態と有効語彙リストとに基づいて、認識された語彙が有効か否かを判断し、判断結果に応じて制御を行うことができるため、車両内の搭乗者にとって快適な制御を行うことができる。また、有効語彙リストや制御内容を自由に設計することができるため、音声認識を使用したアプリケーション設計に自由度を与えることができる。

請求項11に記載の発明は、請求項1から7の何れか1項に記載の音源分離装置を備えた携帯電話機を提供する。

[0019] この発明によれば、携帯電話機を中規模の会議室などで集音マイクとして利用することが可能である。

請求項12に記載の発明は、互いに離して配置された少なくとも2つのマイクロホンに対して、複数の音源から発せられた音源信号を入力する音源信号受信ステップと、前記マイクロホンからの出力信号に対して、周波数領域で複素共役の関係にある2つの重み係数各々を用いて演算を行うことにより、2つのマイクロホンを結ぶ直線の

垂線に対して対称な所定の方向から到来する音源信号を各々減衰させるための第1のビームフォーマ処理及び第2のビームフォーマ処理を行うビームフォーマ処理ステップと、前記ビームフォーマ処理ステップにおいて得られた音源信号各々についてパワースペクトル情報を算出するパワー計算ステップと、前記パワー計算ステップにおいて計算されたパワースペクトル情報同士の差分に基づいて、目的音源のスペクトル情報を抽出する目的音スペクトル抽出ステップとを備えることを特徴とする音源分離方法を提供する。

[0020] 請求項13に記載の発明は、コンピュータに、互いに離して配置された少なくとも2つのマイクロホンから、複数の音源から発せられた音源信号が混合された出力信号を取得する出力信号取得ステップと、前記出力信号取得ステップにおいて取得された出力信号に対して、周波数領域で複素共役の関係にある2つの重み係数各々を用いて演算を行うことにより、2つのマイクロホンを結ぶ直線の垂線に対して対称な所定の方向から到来する音源信号を各々減衰させるための第1のビームフォーマ処理及び第2のビームフォーマ処理を行うビームフォーマ処理ステップと、前記ビームフォーマ処理ステップにおいて得られた音源信号各々についてパワースペクトル情報を算出するパワー計算ステップと、前記パワー計算ステップにおいて計算されたパワースペクトル情報同士の差分に基づいて、目的音源のスペクトル情報を抽出する目的音スペクトル抽出ステップとを実行させるためのプログラムを提供する。

[発明の効果]

本発明によれば、周波数領域で複素共役の関係にある2つの重み係数各々を用いて演算を行うことにより、2つのマイクロホンを結ぶ直線の垂線に対して対称な所定の方向から到来する音源信号を各々減衰させるための第1のビームフォーマ処理及び第2のビームフォーマ処理を行うことにより、2つのマイクロホンを結ぶ直線の垂線に対して対称な方向から到来する音源信号を各々減衰させ、前記第1のビームフォーマ処理及び前記第2のビームフォーマ処理により得られた音源信号各々についてのパワースペクトル情報同士の差分に基づいて目的音源のスペクトル情報を抽出することで、指向特性がマイクロホン素子の感度に影響を受けないという性質を実現することができ、マイクロホン素子の感度のばらつきに影響を受けることなく、複数の音源

から発せられた音源信号が混合された混合音の中から、目的音源からの音源信号を分離することが可能となる。

図面の簡単な説明

- [0021] [図1]本発明の第1実施形態に係る音源分離システムの基本的構成を示す図である。
- 。
- [図2]同実施形態に係るマイクの種類の一例を示す図である。
- [図3]同実施形態に係るビームフォーマ部の構成を示す図である。
- [図4]同実施形態に係るパワー計算部の構成を示す図である。
- [図5]同実施形態に係る目的音スペクトル抽出部の構成を示す図である。
- [図6]同実施形態に係る死角制御型ビームフォーマを説明するための図である。
- [図7]同実施形態に係る死角制御型ビームフォーマを説明するための図である。
- [図8]同実施形態に係る差分結果の指向特性の一例を示す図である。
- [図9]同実施形態に係るマイクの素子感度差を変動させた場合の共役ビームフォーマの指向特性を示す図である。
- [図10]同実施形態に係るマイクの素子感度差を変動させた場合のビームフォーマ単体の指向特性を示す図である。
- [図11]同実施形態に係る音源分離装置の360度の指向特性を求めたグラフを示す図である。
- [図12]同実施形態に係る音源分離装置の指向特性を利用した2話者個別録音の実施例を示す図である。
- [図13]同実施形態に係る音源分離装置の指向特性を利用した簡易ハンズフリー装置及び音声認識装置用マイクの実施例を示す図である。
- [図14]同実施形態に係る評価実験におけるマイク取り付け箇所を示す図である。
- [図15]同実施形態に係る評価実験の条件を示す図である。
- [図16]同実施形態に係る評価実験結果を示す図である。
- [図17]第2実施形態に係る音源分離システムの構成を示す図である。
- [図18]同実施形態に係るビームフォーマの構成を示す図である。
- [図19]第3実施形態に係る音源分離システムの構成を示す図である。

- [図20]同実施形態に係る他の音源分離システムの構成を示す図である。
- [図21]同実施形態に係る目的音スペクトル抽出部の構成を示す図である。
- [図22]第4実施形態に係る音源分離システムの構成を示す図である。
- [図23]第5実施形態に係る音源分離システムの構成を示す図である。
- [図24]同実施形態に係る指向性制御の例を示す図である。
- [図25]同実施形態に係る指向性制御部の構成を示す図である。
- [図26]第6実施形態に係る音源分離システムの構成を示す図である。
- [図27]第7実施形態に係る音源分離システムの構成を示す図である。
- [図28]同実施形態に係る分離される音源の位置を説明するための図である。
- [図29]同実施形態に係る単一指向性マイクの設置例を示す図である。
- [図30]同実施形態に係る目的音抽出部の構成の一例を示す図である。
- [図31]同実施形態に係る目的音抽出部の構成の一例を示す図である。
- [図32]同実施形態に係る目的音抽出部の構成の一例を示す図である。
- [図33]同実施形態に係る音源分離装置を利用したパソコンへの音声入力の例を示す図である。
- [図34]第8実施形態に係る目的音範囲と雑音範囲とを説明するための図である。
- [図35]同実施形態に係る遅延操作を説明するための図である。
- [図36]同実施形態に係る指向性制御手段の構成例を示す図である。
- [図37]同実施形態に係る音源分離装置システムの構成例を示す図である。
- [図38]同実施形態に係る目的音抽出部における処理方式の一例を示す図である。
- [図39]同実施形態に係る目的音抽出部における処理方式の一例を示す図である。
- [図40]第9実施形態に係る車載機器制御用音声認識システムの構成を示す図である。
- 。
- [図41]第10実施形態に係る携帯電話機を示す図である。
- [図42]第11実施形態に係る音源分離システムにおけるマイク配列を示す図である。
- [図43]第12実施形態に係る音源分離システムが適用される環境を示す図である。
- [図44]同実施形態に係る音源分離システムが適用される発話音の入力状況を示す図である。

[図45]同実施形態に係るガイダンス音声削除部の構成を示す図である。

[図46]同実施形態に係る目的音声抽出部の構成を示す図である。

[図47]同実施形態に係るガイダンス音声削除部の他の構成を示す図である。

[図48]同実施形態に係るガイダンス音声削除部の他の構成における目的音声抽出部の構成を示す図である。

[図49]同実施形態に係る音源分離システムが適用される他の環境を示す図である。

[図50]同実施形態に係る音源分離システムが適用される発話音の他の入力状況を示す図である。

[図51]同実施形態に係る音源分離システムの他の構成を示す図である。

発明を実施するための最良の形態

[0022] 以下、図面を参照して、本発明に係る実施の形態について説明する。

[第1実施形態]

図1は、本発明の第1実施形態に係る音源分離システムの基本的構成を示す図である。このシステムは、2つのマイクロホン(以下「マイク」という)10、11と、音源分離装置1とで構成されている。この音源分離装置1は、図示せぬ、全体を制御し演算処理を実行するCPUと、ROM、RAM、ハードディスク装置等の記憶装置を含むハードウェアと、記憶装置に記憶されたプログラム、データ等を含むソフトウェアとを備えている。これらのハードウェア及びソフトウェアによって、図1に示す機能ブロックが実現される。

[0023] 2つのマイク10、11は無指向性のマイクであり、平面上に互いに数cm程度離して設置されている。なお、マイク10、11は無指向性であることが基本であるが、図2に示すような単一指向性マイクも使用可能である。マイク10、11は2つの音源R1、R2から発せられた信号を受信する。このとき、これら2つの音源R1、R2は、2つのマイク10、11を結ぶ直線に対して引かれた垂線を境界として分割された2つの領域(以下「垂線の左右」という)にそれぞれ位置するものとするが、必ずしも垂線に対し左右対称の位置に存在する必要はない。

[0024] このマイク10、11で得た2つの音源信号を、スペクトル分析部20、21においてそれぞれマイク出力毎に周波数分析し、ビームフォーマ部3においてこれらの周波数分析

された信号を2つのマイク10、11を結ぶ直線に対して引かれた垂線の左右に対称に死角を形成したビームフォーマ30、31でフィルタリングを行い、パワー計算部40、41においてそのフィルタ出力のパワーを計算し、目的音スペクトル抽出部50、51においてそれぞれの差分を計算した結果に対しある一定値以上の値を出力しそれ以下をゼロとする処理を行う。これらの処理は、通常の特定雑音に関して死角を形成する処理とは異なり、ビームフォーマ部3をいくつかの条件の下で形成し上記の処理を実施することで、従来問題であったマイクの素子感度のバラつきに起因するマイクロホンアレイの特性の劣化の問題を解決すると同時に、上記垂線を中心として左右からの音を分離する指向特性を幅広い周波数帯域に対して実現することが可能となる。以下、各機能ブロックについて詳細に説明する。

[0025] [ビームフォーマ部]

まず、図3を参照して、ビームフォーマ部3の構成を説明する。図3では、スペクトル分析部20、スペクトル分析部21で周波数成分毎に分解された信号 $x_1(\omega)$ 、 $x_2(\omega)$ を入力として、乗算器100a、100b、100c、100dにて、フィルタの重み係数 $w_1(\omega)$ 、 $w_2(\omega)$ 、 $w_1^*(\omega)$ 、 $w_2^*(\omega)$ (*は複素共役の関係にあることを示す)と乗算をそれぞれ行い、加算器100e、100fにて2つの乗算結果を加算し、その出力としてフィルタリング処理結果 $ds_1(\omega)$ 、 $ds_2(\omega)$ を出力する。このように、ビームフォーマ部3は、複素共役フィルタ係数を使用することにより、マイク10、11間を結ぶ直線の垂線を中心として対称な位置に死角を形成している。

[パワー計算部]

次に、図4を参照して、パワー計算部40、41について説明する。パワー計算部40、41は、ビームフォーマ30、ビームフォーマ31からの出力 $ds_1(\omega)$ 、 $ds_2(\omega)$ を、以下の計算式により、パワー計算部40、パワー計算部41にてパワースペクトル情報 $ps_1(\omega)$ 、 $ps_2(\omega)$ に変換する。

$$ps_1(\omega) = [\text{Re}(ds_1(\omega))]^2 + [\text{Im}(ds_1(\omega))]^2$$

$$ps_2(\omega) = [\text{Re}(ds_2(\omega))]^2 + [\text{Im}(ds_2(\omega))]^2$$

[0026] [目的音スペクトル抽出部]

次に、図5を参照して、目的音スペクトル抽出部50、51に関して説明すると、パワー

計算部40、41の出力 $ps_1(\omega)$ 、 $ps_2(\omega)$ は、目的音スペクトル抽出部50、51の2つの入力として使用される。目的音スペクトル抽出部50、51は、2つのビームフォーマ30、31出力のパワースペクトル情報を入力として、出力として左右の目的音をそれぞれ出力するが、内部的には差分計算部500、510と係数変換部501、511で構成される。

差分計算部500は、減算器500aにてビームフォーマ30のパワースペクトル情報からビームフォーマ31のパワースペクトル情報の減算を行い、同様に、差分計算部510は、減算器501aにてビームフォーマ31のパワースペクトル情報からビームフォーマ30のパワースペクトル情報の減算を行い、その結果をそれぞれ、係数変換部501、係数変換部511へ入力する。係数変換部501、係数変換部511は、それぞれ左右の音を分離するためのブロックであり、それぞれ、値が特定のしきい値以上の値を、目的とする方向からの信号としてスペクトル情報を出力する。ここで、しきい値の値は一般的には“0”であるが、使用環境に応じて最適値を実測より求め別々に設定しても良い。

[0027] [動作]

次に、図1を参照して、音源分離装置システム全体の動作を説明する。

まず、2つの無指向性あるいは指向性のマイク10、11を互いに数cm程度離して設置し、2つの音源から発せられた信号をマイク10、マイク11で受信する。このとき、各マイク10、11で受信された2つの音源信号が重畳された信号は、スペクトル分析部20、スペクトル分析部21で周波数分析が行われる。ここでは、一般的には高速フーリエ変換などの手法が用いられるが、フィルタバンクなどの周波数分析手法でも良い。周波数分析処理は10msec程度の固定周期毎に行われる。

周波数分析された2つの信号は、マイク10、11を結ぶ直線への垂線に対して対称に死角が形成されたビームフォーマ30、ビームフォーマ31にてフィルタリングされ、特定方向からの信号が減衰させられる。しかし、ここでは特定の音源到来方向を正確に推測し、正確に推測された音源方向に死角を向けることを目的とはしていない。この2チャンネルの入力を使用したフィルタリングは各周波数成分毎に行われ、ビームフォーマ30、ビームフォーマ31の出力はパワー計算部40、パワー計算部41にてス

ペクトルパワー情報に変換されると同時に、ビームフォーマ30、ビームフォーマ31の出力から位相抽出部60、位相抽出部61にて位相情報 Φ_1 、 Φ_2 が抽出される。次に、パワー計算部40、パワー計算部41にてスペクトルパワー情報に変換されたビームフォーマ30、ビームフォーマ31の出力は、目的音スペクトル抽出部50、目的音スペクトル抽出部51に送られ、目的音スペクトル抽出部50にて右方向(0~90°)から来た音源信号のパワースペクトル情報が抽出され、目的音スペクトル抽出部51にて左方向(-90°~0)から来た音源信号のパワースペクトル情報が抽出される。

[0028] 目的音スペクトル抽出部51から抽出されたパワースペクトル情報を音声認識処理の前処理として使用する場合は、当該パワースペクトル情報は図示していない音響パラメータ解析部に送られ、音響分析処理が行われる。一方、抽出した音源信号のパワースペクトル情報を時間信号に戻す必要がある場合は、位相抽出部60、位相抽出部61で抽出した位相情報と目的音スペクトル抽出部50、目的音スペクトル抽出部51で抽出したスペクトル情報とを時間波形変換部70、時間波形変換部71に入力し時間信号情報に戻す処理を行う。

[死角制御型ビームフォーマの設計例]

次に、ビームフォーマ部3中のビームフォーマ30、ビームフォーマ31において、2つのマイク10、11を結ぶ直線の垂線に対して対称な位置に死角を形成することで、指向特性(指向性)がマイク素子の感度の影響を受けないことについて証明する。

[0029] 2つのマイク素子を使用する場合において、図6に示すように、目的方位 θ_1 に対するゲインを1とし、他方向 θ_2 に1つの死角(ゲイン0)を形成する死角制御型ビームフォーマの設計例を以下に示す。

死角制御型ビームフォーマの出力信号を $S(f) = [s_1(f), s_2(f)]'$ 、観測信号を $X(f) = [x_1(f), x_2(f)]'$ としたとき、ある周波数 f における死角制御型ビームフォーマの重み係数ベクトル $W(f, \theta_1, \theta_2) = [w_1(f), w_2(f)]'$ は以下の計算で求めることが出来る(ただし、'は転置操作を示す)。

[0030] [数1]

$$S(f, \theta_1, \theta_2) = W(f, \theta_1, \theta_2)X(f, \theta_1, \theta_2)$$

$$W(f, \theta_1, \theta_2) = \begin{bmatrix} w_1(f, \theta_1, \theta_2) \\ w_2(f, \theta_1, \theta_2) \end{bmatrix}$$

このとき、

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} \exp[j2\pi fd_1 \sin \theta_1 / c] & \exp[j2\pi fd_2 \sin \theta_1 / c] \\ \exp[j2\pi fd_1 \sin \theta_2 / c] & \exp[j2\pi fd_2 \sin \theta_2 / c] \end{bmatrix} \begin{bmatrix} w_1(f, \theta_1, \theta_2) \\ w_2(f, \theta_1, \theta_2) \end{bmatrix}$$

$$\begin{bmatrix} w_1(f, \theta_1, \theta_2) \\ w_2(f, \theta_1, \theta_2) \end{bmatrix} = \begin{bmatrix} \exp[j2\pi fd_1 \sin \theta_1 / c] & \exp[j2\pi fd_2 \sin \theta_1 / c] \\ \exp[j2\pi fd_1 \sin \theta_2 / c] & \exp[j2\pi fd_2 \sin \theta_2 / c] \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

[0031] 一方、図7に示すように、目的方向、死角方向を2つのマイク10、11を結ぶ直線の垂線を中心として、図6に示す位置とは線対称な位置に設定するとき、重み係数ベクトル $W(f, -\theta_1, -\theta_2) = [w_1(f), w_2(f)]'$ は以下の計算で求めることが出来る。

[0032] [数2]

$$S(f, -\theta_1, -\theta_2) = W(f, -\theta_1, -\theta_2)X(f, -\theta_1, -\theta_2)$$

$$W(f, -\theta_1, -\theta_2) = \begin{bmatrix} w_1(f, -\theta_1, -\theta_2) \\ w_2(f, -\theta_1, -\theta_2) \end{bmatrix}$$

[0033] このとき、

[0034] [数3]

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} \exp[-j2\pi fd_1 \sin \theta_1 / c] & \exp[-j2\pi fd_2 \sin \theta_1 / c] \\ \exp[-j2\pi fd_1 \sin \theta_2 / c] & \exp[-j2\pi fd_2 \sin \theta_2 / c] \end{bmatrix} \begin{bmatrix} w_1(f, -\theta_1, -\theta_2) \\ w_2(f, -\theta_1, -\theta_2) \end{bmatrix}$$

$$\begin{bmatrix} w_1(f, -\theta_1, -\theta_2) \\ w_2(f, -\theta_1, -\theta_2) \end{bmatrix} = \begin{bmatrix} \exp[-j2\pi fd_1 \sin \theta_1 / c] & \exp[-j2\pi fd_2 \sin \theta_1 / c] \\ \exp[-j2\pi fd_1 \sin \theta_2 / c] & \exp[-j2\pi fd_2 \sin \theta_2 / c] \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

[0035] と計算され、両者の関係は、

[0036] [数4]

$$\begin{bmatrix} w_1(f, -\theta_1, -\theta_2) \\ w_2(f, -\theta_1, -\theta_2) \end{bmatrix} = \begin{bmatrix} w_1(f, \theta_1, \theta_2) \\ w_2(f, \theta_1, \theta_2) \end{bmatrix}^*$$

[0037] となり、各重み係数は複素共役の関係にある。

次に、パワー計算部40、41、目的音スペクトル抽出部50、51において指向特性の導出を行う。指向特性の計算のため、重みベクトルW、及び、方位ベクトルVを以下の式で定義する。

[0038] [数5]

$$W = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}, \quad V = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$$

[0039] すると、パワー計算部40の出力 $ps_1(\omega)$ 、パワー計算部41の出力 $ps_2(\omega)$ の指向特性は以下のように表現することが出来る。

[0040] [数6]

$$\begin{aligned} ps_1(\omega) &= [V^T W^*] [V^T W^*]^H = V^T W^* W^T V^* \\ ps_2(\omega) &= [V^T (W^*)^*] [V^T (W^*)^*]^H = V^T W W^H V^* \end{aligned}$$

[0041] ただし、*は共役操作、Hは共役転置操作を表す。これより、目的音スペクトル抽出部50内の差分操作部500の出力 $dr_1(\omega)$ は以下のように求めることが出来る。

[0042] [数7]

$$\begin{aligned} dr_1(\omega) &= V^T W^* W^T V^* - V^T W W^H V^* = V^T (W^* W^T - W W^H) V^* \\ &= 2 \times \text{Re}[w_1^* w_2 v_1 v_2^*] - 2 \times \text{Re}[w_1 w_2^* v_1 v_2^*] \end{aligned}$$

[0043] ここで、マイクの素子感度のバラつきを表現するパラメータとして α を導入し、一方のマイクの素子感度が一方のマイクの素子感度の α 倍であると仮定する。このときに、片方のマイク出力が α 倍であるということと、片方のチャンネルに掛ける重みが α 倍であることとは等価であるので、マイク素子感度のバラつきを考慮し、 $w_2 = \alpha w_{org2}$ とすると、

[0044] [数8]

$$dr_1(\omega) = \alpha \left\{ 2 \times \operatorname{Re} \left[w_1^* w_{org2} v_1 v_2^* \right] - 2 \times \operatorname{Re} \left[w_1 w_{org2}^* v_1 v_2^* \right] \right\}$$

[0045] となり、マイク素子感度が変わっても指向特性は変化しない。

このとき音源がマイクに対して十分離れている場合、すなわち平面波の場合は、方位ベクトルは、

[0046] [数9]

$$V = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 1 \\ \exp(-j \frac{2\pi}{\lambda} d \sin \theta) \end{bmatrix}$$

[0047] と表現されるので、結局、

[0048] [数10]

$$dr(\omega)_1 = \alpha \left\{ 2 \times \operatorname{Re} \left[w_1^* w_{org2} \exp(j \frac{2\pi}{\lambda} d \sin \theta) \right] - 2 \times \operatorname{Re} \left[w_1 w_{org2}^* \exp(j \frac{2\pi}{\lambda} d \sin \theta) \right] \right\}$$

[0049] となる。ただし、上記に述べた手法は球面波の場合も同様の性質が維持される。

図8は、死角を前記垂線から $\pm 45^\circ$ 方向に形成する際に、ビームフォーマ30、31のゲインを維持する拘束条件を垂線に対称な位置($\pm 45^\circ$)に指定して設計した場合の差分結果の指向特性例である。図8から分かるように、 0° 方向を境界として、右方向(マイク10、11から音源R1、R2の方向を見た場合の右方向、以下同様)($0 \sim 90^\circ$)から来た音源信号は正の値をとり、左方向($-90^\circ \sim 0$)から来た音源信号は

負の値をとる。これにより、各周波数成分において左右どちらの方向から到来したのかを判断することが出来る。

[0050] [指向特性の実験結果]

以上述べてきたように、ビームフォーマ30、ビームフォーマ31の乗算器で使用する重み係数をそれぞれ複素共役の関係として上記の処理を行うことで、マイクの素子感度によりアレイマイクの指向特性が影響を受けないことを示すために、マイクの素子感度差 α を0db、+6db、-6dbと変動させて指向特性の計算を行った例を図9に示す。同図にはマイク10、11を結ぶ直線への垂線に対して対称な方向の指向特性が示されていないが、当該指向特性は図9に示した特性と対称な特性を持つ。図9を見て分かるように、ゲインの変動がある場合、アレイマイクの出レベルの変動は発生するが、指向特性は変動を受けない。これにより、安価なマイクを使用しマイクの素子感度にばらつきがある場合でも安定した指向特性を実現することが可能となる。また、図中に示した指向特性は死角を $\pm 45^\circ$ 方向に作成したものであるが、同図から分かるように指向特性に幅があるため、死角を実際の目的音源に対して正確に作成する必要はない。また、マイクの素子感度差 α を0db、+6db、-6dbと変動させた場合のビームフォーマ単体の指向特性を図10に示すが、同図よりマイクの素子感度が6db異なるると特定の方向に死角を形成するというビームフォーマに望まれている所望の動作を殆ど行えないことがわかる。一方、本発明の特徴として最も注目すべき点として、図10に示すような指向特性が発生するような指向特性の劣化したビームフォーマを使用しても、結果として得られる指向特性はマイクの素子感度が均一な場合と同一であることである。

特定方向にビームフォーマで鋭い死角を形成する手法は少数のマイクでも理論上実現可能であるため、測定した信号はSN比を上げるために使用されたり、適応フィルタとして頻繁に使用される一般化サイドローブキャンセラの参照信号生成部であるブロッキングマトリックス部分に使用されるが、上述したマイクの素子感度差により設計どおりの性能で死角を生成することが出来ず、量産を目的とした場合、実環境で所定の性能が実現出来ない大きな要因の一つになっている。

[0051] このような音源分離装置1を用いて、360度の指向特性を求めた例を図11に示す。

同図から分かるように、音源分離装置1は180° 毎の指向特性を持ち、2つの指向特性がその境界において重なることなく分離されている。しかも、その指向特性がマイクの素子感度の影響を受けないことを更なる特徴としている。リニアアレイの場合、 $\pm 90^\circ$ 以上では0～ $\pm 90^\circ$ までの特性と対称となる指向特性となる。このように、指向特性は、2つのマイクを結ぶ線に垂直に引いた線を境界として2つのゾーンに分割することが可能となる。

[0052] この特性を利用した実施例を、図12及び図13に示す。図12は音声メモ装置への応用である。従来、会議や打ち合わせの際に会議のメモの目的で音声メモ装置が使用されてきた。このような使用例の場合、周りの騒音や2人の音声と同時に録音されているため、録音された内容が聞き取りにくい場合がある。この様な場合、2つのマイク10、11をそれぞれ2人の発話者の方に向け、音源分離装置1でそれぞれ片方の会話の内容を強調して録音することで後日の聞き取りを容易にする。このように音源分離装置1を利用することによって、対向する2名の音声を分離し収集することが可能となり、会議の議事録用音声メモとして利用することや、会議議事録で2名の会話を個別に同時音声認識することが可能となる。

[0053] 図13は、簡易ハンズフリー装置、及び、音声認識装置用マイクへの応用例を示す図である。近年、パーソナルコンピュータ(以下「パソコン」という)を使用してネット会議などが行われるようになって来ているが、自宅などでパソコンを使用してネット会議を行う場合、パソコン内部にエコーキャンセラ機能がないためヘッドセットマイクなどを利用しスピーカ音声からマイクへのエコーの回り込み量を小さくするなどの対策が行われているが、ヘッドセットマイクのように身体の一部に器具を密着させることがなかなか受け入れられていない。この対策として、パソコン内部にエコーキャンセラ機能をソフトウェアとして実装したり、パソコン外部にエコーキャンセラ機能を内蔵したハンズフリー装置を接続することも可能である。パソコン内部にエコーキャンセラ機能を実装する場合、スピーカへの再生音声とマイクからの入力音声との同期をとる必要があるが、再生音声と入力音声間の遅延が大きく、また、パソコン機種ごとに遅延の大きさが異なるなど実装上の問題がある。また、外部にエコーキャンセラ機能を内蔵したハンズフリー装置を接続する場合、コストがかかる。

[0054] 一方、本実施例では、マイク10、11をスピーカと発話者との間に置く必要があるが、スピーカ再生信号との同期をとる必要もない。2つのマイク10、11を用意し、マイク10、11からの信号をパソコンに取り込み、パソコンに記憶されているソフトウェアに基づいて音源分離を実行することが出来るため、実装が容易である。

また、住宅内で音声認識を行う状況で一番頻繁に発生するであろう環境として、テレビその他被制御機器に向かって音声コマンドを使用することが考えられる。この場合、テレビその他被制御機器のスピーカは被制御機器側から発話者側に向かって設置しており、スピーカから音声が流れていたりさまざまなガイダンスが流れている状況で、発話者は音声でマイク装置に制御コマンドを入力したい場合がある。この場合、何らかの手段でスピーカの音声がマイク装置に回り込む量を減衰させる必要があるが、音源分離装置1を使用することで容易に被制御機器からの音声と発話者からの音声を分離することができ、音声認識性能が向上する。

[0055] [評価実験結果]

次に、評価実験結果について説明する。従来、自動車内での機器制御その他の目的で音声認識が使用されている。従来においては、エアコンの送風音、ロードノイズ、エンジン音などの雑音を低減する技術は確立されているが、音声などの影響を受けない音声認識技術は確立されていない。これらのことを踏まえ、以下の点を実現可能とする音声認識技術を提供することが重要となる。(1)運転席に居る運転者が発した音声、及び、助手席に居る助手席者が発した音声(以下「運転席と助手席の音声」という)の分離、(2)ヘッドポジションの移動を許容可能、(3)広いビーム幅と分離性能の両立、(4)少数マイクでの機能提供、(5)音声認識使用時、運転者又は助手席者に黙っていることを課さない

[0056] 音源分離装置1の有効性を示すために、車内において2人の発話者に同時発話させ、2つのマイクで収録して音源分離装置で音源分離を行い、音声認識実験を行った。この音源分離装置は、2つの音源の分離を目的に開発されたものであり、自動車内の利用方法としては、例えば運転席と助手席の音声を分離することが可能となる。このためのマイクの設置箇所としては、図14に示すように、取り付け箇所L1または取り付け箇所L2のように自動車内中央部が適している。また、取り付け箇所L2に設置

した場合には、ルームミラー400が運転者の方を向くため、マイク取り付け時に、マイクの取り付け方向が大体正面になるように設置してもよいし、後述する指向性制御機能を持たせてもよい。音源分離装置1では本来ビーム幅を広く取ってあるので正確な位置あわせは必要としない。また、取り付け箇所L2に設置する場合には、背面からの反射を抑えるためにマイクモジュールを工夫したり指向性マイクを利用することが有効な場合がある。

評価実験の条件を図15に、音源分離後の音声認識実験結果を図16に示す。図16(a)に示すように、車内での2話者発話時(車内同時発話)においては、1チャンネルマイクのみを使用した従来法1における未処理時には29%(停車時)、27%(60km/h走行時)であったが、本音源分離手法を適応することにより、78%(停車時)、78%(60km/h走行時)と改善された。また、図16(b)に示すように、助手席側の音声を運転席側の音声と誤って音声認識結果を出してしまった、あるいは運転席側の音声を助手席側の音声と誤って音声認識結果を出してしまった割合を評価すると、従来の1チャンネルマイクのみ使用時(車内1人発話)では、全発話の93%(リジェクション性能7%)について何らかの音声認識結果が出力されてしまったが、本手法を適応した場合は、何らかの音声認識結果が出力されたのは0%(リジェクション率100%)であった。また、2マイクを使用した従来例として“雑音成分抑圧処理装置および雑音成分抑圧処理方法(特許第3484112号)”を従来法2として停止状態での性能比較を行った。従来法2は目的音、雑音到来方向を推定しながら適応ビームフォーミング処理を行い目的音および雑音を強調し、目的音が強調された信号から雑音が強調された信号を周波数領域でスペクトルサブトラクションする方式であるが、到来方向推定エラーの影響を省くため目的音、雑音とも到来方向を既知(固定方向から到来するもの)とし、適応ビームフォーマ部の最適値を求めた後に、目的音(発話者1)、雑音(発話者2)を同時に再生し目的音声を抽出する処理をおこなった。(図16(a)参照のこと)

[0057] [第2実施形態]

次に、第2実施形態について説明する。図17には、第2実施形態に係る音源分離システムの構成を示す。上述した第1実施形態においては、マイク10、11からの入力

を、まずスペクトル分析部20、スペクトル分析部21にて周波数成分に変換したが、本実施形態においては、まず時間領域にてビームフォーマ80及びビームフォーマ81にて死角を生成して特定の到来方向からの信号を減衰させた信号を生成し、その後、スペクトル分析部20及びスペクトル分析部21にて周波数成分に変換している。なお、図17では、図1と同一の機能を持つものには同一の番号が付けてある。ビームフォーマ80、ビームフォーマ81の構成は図18に示すようなFIRフィルタなどの形式で構成されたフィルタ処理を実行することで実現される。このとき、FIRフィルタの係数は、図3で示した周波数領域での複素共役の関係にある重み係数を、時間領域のフィルタ係数に変換することで求めることができる。

[0058] [第3実施形態]

次に、第3実施形態について説明する。図19及び図20は、第3実施形態に係る音源分離システムの構成を示す図である。上述したように、図1及び図17に示す目的音スペクトル抽出部50、目的音スペクトル抽出部51は、図5で示した構成にて実現されており、実験により求めた最適なしきい値を使用して音源分離処理を行う構成となっている。一方、図8に示したように、目的音スペクトル抽出部50、51中の差分計算部500、510の出力である $dr_i(\omega)$ ($i=1, 2$)は正面 0° を中心として点対称な関係になっていることが分かる。これより、目的音スペクトル抽出部50、51中の係数変換部501、511においてしきい値を“0”とすると差分計算部500、510の符号のみを見ておくことで、正の時に右方向($0\sim 90^\circ$)から来た音源信号のパワースペクトル情報として抽出され、負のときに左方向($-90^\circ\sim 0$)から来た音源信号のパワースペクトル情報として抽出される。このため、図1及び図17で示した全体構成は、図19及び図20に示すように簡略化することが可能となる。図19及び図20中の目的音スペクトル抽出部90は、図21に示す構成で実現される。

[0059] 図21において、パワー計算部40、パワー計算部41にて計算されたビームフォーマ30、ビームフォーマ31のパワースペクトル情報は、目的音スペクトル抽出部90内部の差分計算部900に入力される。そして、減算器900aにて減算処理が行われ、係数変換部910、係数変換部920においてそれぞれ目的とする方向からの音源信号のみが抽出される。具体的には、係数変換部910は、右方向($0\sim 90^\circ$)からの音源

抽出を行うためのブロックであり、入力为正の場合は、そのスペクトル情報は右方向($0\sim 90^\circ$)から来たものとして出力し、負の場合は目的方向外から到来した音源のスペクトル情報として出力しない。一方、係数変換部920は、左方向($-90^\circ\sim 0$)からの音源抽出を行うためのブロックであり、入力を負の場合は、そのスペクトル情報は左方向($-90^\circ\sim 0$)から来たものとして出力し、正の場合は目的方向外から到来した音源のスペクトル情報として出力しない。以上の動作により、2つのマイク10、11を結ぶ直線の垂線を中心として左右の方向から到来する音源信号を分離することが可能となる。

なお、図19に示す音源分離システムと図20に示す音源分離システムとは、ビームフォーマ処理を周波数領域で行うか時間領域で行うかの違いが構成として異なっている。図19ではビームフォーマ処理を周波数領域で行っており、図20では時間領域で行っている。

[0060] [第4実施形態]

次に、第4実施形態について説明する。図22は、第4実施形態に係る音源分離システムの構成を示す図である。同図に示すスペクトルサブトラクション部100、101が、目的音スペクトル抽出部50、目的音スペクトル抽出部51で抽出した各目的音に重畳している到来方向不明の定常雑音や拡散性の雑音などを除去するためのスペクトルサブトラクション処理を行う。このような構成は、会議室中の空調やプロジェクタなどのファンの音などが存在する環境における使用でも有効であるが、自動車内での使用時に特に有効である。自動車内において運転者と助手席に同乗者が搭乗している場合など、例えば運転席と助手席の音声は、前述の方式を使用することで別々に取り出すことが可能である。しかし、エアコンの送風音、ロードノイズ、風きり音など、到来方向が不明なものや拡散性の雑音は前述の方法では除去することが出来ない。これらの雑音の影響は、処理の後段にスペクトルサブトラクション処理を入れることで除去することが可能である。スペクトルサブトラクション処理は、1つのマイクの入力信号を使用して発話区間検出を行い、無発話区間で雑音スペクトルを推定し、発話区間でその前に推定した雑音成分をスケールリングして減算するタイプと、雑音が優位に収録された信号と音声優位に収録された信号を用いて、音声が優位に収録され

た信号から雑音が優位に収録された信号のスペクトルをスケーリングして差分するタイプとがある。音声認識の場合、1マイク方式を基本にした処理で十分なことが多いが、本実施形態においてもこれを採用しており、本実施形態に係る音源分離システムは、第1実施形態に係る音源分離システムに対して、発話区間検出部110とスペクトルサブトラクション部100、101をあらたに追加することで対応している。

- [0061] 図22において、音源R1、音源R2は目的音であり、音源R3が到来方向不明の定常雑音や拡散性の雑音を表している。これらの雑音の多くは明確な指向性を持たないものが多い。このような雑音の場合、目的音スペクトル抽出部の出力において、弱い指向性をもつものはその方向の音源を抽出する目的音スペクトル抽出部に多く現れ、まったく指向性を持たないものや道路の継ぎ目を踏んだ衝撃音などは左右の抽出されたスペクトルに交互に検出されることが多いが、これらの雑音をスペクトルサブトラクション部100、101で除去可能である。なお、スペクトルサブトラクションとしては、発話区間検出の不要な連続型スペクトルサブトラクションを使用してもよい。

[第5実施形態]

次に、第5実施形態について説明する。図23には、第5実施形態に係る音源分離システムの構成を示す。本実施形態においては、分離したい2つの目的音源R1、R2が2つのマイク10、11を結ぶ直線への垂線に対して対称といえる位置から大きく外れている場合の対策が示されている。本実施形態では、この対策のために2つの目的音源R1、R2のおおまかな位置を検出するための到来方向推定部120と、到来方向推定部120が推定した2つの目的音源R1、R2のおおよその音源到来方向情報を使用し、指向性制御部140において分離したい2つの目的音源R1、R2が仮想的に出来るだけ2つのマイク10、11を結ぶ直線への垂線に対して対称となるように、片方のマイク入力に遅延操作を与える。

- [0062] 図24には、2つの音源R1、音源R2がマイク10、マイク11を結ぶ直線の垂線に対して θ_τ だけ回転した直線に対して左右対称となる状況を示している。このような場合、片方のマイクで取得した信号に一定遅延量 τd を与えることで、 θ_τ だけ回転したと等価な状況を実現可能である。また、あらかじめいくつかの2つの音源について、2つのマイクを結ぶ直線の垂線に対する位置関係から、2音源を分離するのに最適

なビームフォーマ30、31のフィルタパラメータを用意しておき、到来方向推定部120からのおおまかな2つの音源の到来方向情報を元に現在の状況の2つの音源を分離するのに最適と思われるビームフォーマのフィルタパラメータをビームフォーマ制御にて選択して、選択したビームフォーマのフィルタパラメータをビームフォーマ30、ビームフォーマ31に設定するようにしてもよい。

[0063] 図25に指向性制御部140の構成例を示す。図25(a)に示す構成例は、周波数領域で各周波数成分毎に片方のマイクで取得した信号に一定遅延量 τd を与える時の例を示している。図25(a)に示す構成例の場合、 $x_1(\omega)$ に対して乗算器100aにて $e^{-j\omega \tau d}$ との乗算を行うことで遅延操作を実現している。また、時間領域でビームフォーマを行いたい場合は、図25(b)に示すようなFIRフィルタの形式の遅延器100bによりフィルタリング処理を行えばよい。

なお、片方のマイク入力に遅延を与えるのではなく、双方のマイク入力にそれぞれ半分ずつの遅延を与え全体として同量の遅延操作を実現しても良い。つまり、片方のマイクで取得した信号に遅延量 τd を与えるのではなく、片方のマイクで取得した信号に遅延量 $\tau d/2$ 、もう片方のマイクで取得した信号に遅延量 $-\tau d/2$ を与えることで、全体の遅延差が τd になるようにしてもよい。

[0064] [第6実施形態]

次に、第6実施形態について説明する。図26は、第6実施形態に係る音源分離システムの構成を示す図である。本実施形態に係る音源分離システムは聴感を重視するための構成となっており、自動車内でのハンズフリー通話などへの応用を考慮している。自動車内において運転者と助手席の同乗者とが搭乗している場合などは、例えば運転席と助手席の音声は前述の方式を使用することで別々に取り出すことが可能であるが、エアコンの送風音、ロードノイズ、風きり音など到来方向が不明なものや拡散性の雑音は前述の方法では除去することが出来ない。これらの場合、第4実施形態で述べた様に、処理の後段にスペクトルサブトラクション処理を入れることでこれらの雑音の影響を除去することが可能であり、音声認識などの聴感上の音質を問題としない用途には最適であるが、ハンズフリー通話装置用のマイクなどに使用した場合に、聴感上ミュージカルノイズと呼ばれる雑音の消し残りが問題になることがある。

本発明においては、2つのマイク10、11を結ぶ直線への垂線を中心にして左右の音を分離するため、拡散性の雑音などその音の指向性が絶えず変動するような雑音の場合、左右の分離結果に不規則に振り分けられてしまい音質を劣化させる場合がある。

[0065] このため、本実施形態では、本発明に係る音源分離方式がマイクゲインの時間変化に影響を受けないことを利用して、通常アレイマイク処理の後段に使用するポストフィルタ処理をビームフォーマ処理の前段に入れて、拡散性雑音、定常雑音などを低減し、音源分離後のミュージカルノイズの発生を防いでいる。

[第7実施形態]

次に、第7実施形態について説明する。図27は、第7実施形態に係る音源分離システムの構成を示す図である。この音源分離システムは、3つのマイク10、11、12を使用して3音源を分離する構成を示している。同図において、マイク10とマイク12、マイク10とマイク12を使用して、それぞれ2つのマイクを結ぶ直線への垂線を中心として左右の音源信号を分離し、2組のマイクの組を使用して分離された音源信号合計4つを使用して、目的音抽出部160にてマイク10、マイク11の正面方向近辺から到来する音源R3を最終的に分離する。

[0066] 図28を使用して本構成で分離される音源の説明を行うと、図28に示すように2つのマイクを結ぶ直線への垂線aと垂線bを中心として、垂線の左右から到来する音源を分離することが可能となる。図28においては、ゾーンA、B、Cの各領域に音源が存在すると仮定すると、垂線aを使用して、ゾーンAとゾーンB、Cの領域から到来する音源信号を分離することができ、垂線bを使用して、ゾーンA、BとゾーンCから到来する音源の分離をすることができる。図27において、これらの分離を行うブロックが分離部b0と分離部b1である。分離部b0は、3つの音源が重畳した信号から、ゾーンAの領域からの音源信号 $S_A(\omega)$ と、ゾーンB、Cの領域からの音源信号 $S_{Bmix}(\omega)$ 、 $S_{Cmix}(\omega)$ が混合した信号とを分離可能であり、分離部b1は同様に、ゾーンA、Bの領域からの音源信号 $S_{Amix}(\omega)$ 、 $S_{Bmix}(\omega)$ が混合した信号と、ゾーンCの領域からの音源信号 $S_C(\omega)$ とを分離可能で、この段階で、 $S_A(\omega)$ と $S_C(\omega)$ が分離出来ている。この様にして得られた4つの信号からパワースペクトル領域で目的音抽出部160にて所定

の操作を行うことで、 $S_B(\omega)$ を求めることが出来る。ただし、ここで、 $S_{Amix}(\omega)$ 、 $S_{Bmix}(\omega)$ 、 $S_{Cmix}(\omega)$ は他信号と混合された時の各信号を表す。

[0067] 上記では、ゾーンDに音源情報が無いかあるいはあってもレベル的に小さいことが前提であるが、もしゾーンDに音源がある場合には、3つのマイク10、11、12に指向性マイクを使用することでゾーンDからの音源信号の混入を大幅に軽減することが可能である。

図29には単一指向性マイクの設置例を示すが、一般的にこのように指向性マイクを使用するとアレイマイクのビームフォーマ部において各マイク間の指向性のバラツキなどで設計時の性能が出ないことが起こるが、本方式でもともとマイクの素子感度バラつきに影響を受けないのと同様、指向特性のバラツキにも影響を受けない一定の性能を実現可能である。

[0068] 図30～32を参照して、目的音抽出部160における処理方式について詳細に説明する。図30に示す処理方式は、2チャンネルのスペクトルサブトラクション処理と同様の処理方式である。つまり、片方のチャンネル信号としても目的音と妨害音とのパワースペクトルの和が求まっており、もう片方のチャンネル信号としても妨害音のパワースペクトルが求まっているため、これらを減算器100aで減算することにより目的音 $S_B(\omega)$ を求めることが出来る。

[0069] 図31に示す処理方式は、目的音に妨害音が重畳されたものを2つ求めることが可能なため、加算器101bにてこれらを加算することで目的音のパワースペクトル情報を2倍の大きさとし、一方、加算器101aにて妨害音同士を加算することで妨害音のパワースペクトルを求め、これに乗算器101cで一定ファクタ(O. E. F 1～2)を掛けて、減算器101dで加算器101bの出力から差分を計算することにより、目的音を抽出するものである。なお、減算器101dの出力の段階で原音信号より音量が大きいのでレベル調整部101eでレベル調整を行う。

[0070] 図32に示す処理方式は、図31に示す加算器101a、101bの代わりに最小値計算102a、102bを使用するものである。尚、図31では、O. E. Fは1.0より大きい方が音質が良い場合が多いが、図32ではO. E. Fは1.0前後が良い。なお、図32では最小値計算を行ったが、最小値計算の代わりに最大値計算でも良い。

また、分離したい目的音源の位置が本手法で最適な分離性能が得られる位置より大きく外れている場合があるが、第5実施形態で説明したように、マイクから出力される入力信号に遅延を与え音源の到来方向を仮想的に変えることで、出来るだけ最適な分離性能が得られるように操作することが可能である。

[0071] 図33には、本実施形態に係る音源分離システムの使用例を示す。パソコンへの音声入力の用途において、3つのマイク10、11、12を使用して、小さい実装面積で、パソコン正面からの音声を、指向性を持たせて取得する例を示す。

[第8実施形態]

次に、第8実施形態について説明する。上述した実施形態においては、
(1)2つのマイクを使用してマイク間を結ぶ直線を中心として、左右からの音を分離する実施形態(2)3つのマイクを使用して正面からの音とその左右からの音を分離する実施形態 について説明したが、図34に示すように、2つのマイク10、11を使用して2つのマイク10、11を結ぶ直線の正面方向からの音を分離抽出したい場合がある。

[0072] この場合、指向性制御手段により2つのマイク10、11のうち片方マイクのからの出力信号に図35(b)に示すように遅延を与え、仮想的に3チャンネル目のマイク位置を作成することで、仮想的に図35(a)に示す3マイク入力を実現することが可能である。図36には、図35で示した遅延操作を行う指向性制御手段の構成例を示す。図中において D_i ($i=1, 2, 3, 4$)は遅延素子を表すが、実際の遅延操作は時間領域で遅延操作を行っても良いし、スペクトル分析後に周波数領域で遅延操作を行っても良い。

[0073] 図37には、本実施形態に係る音源分離装置システムの構成例を示す。指向性制御手段141、142は、スペクトル分析部20、21と、遅延処理を行う遅延素子とで構成されている。処理順としては、スペクトル分析処理を行ってから遅延処理を行ってもよい(図中のType1)、遅延処理を行ってからスペクトル分析処理を行ってもよい(図中のType2)。指向性制御手段141、142の出力信号は、本手法NBF以降のブロックにおいてビームフォーマ30、31、パワー計算部40、41等での処理が行われ、処理後の信号が目的音抽出部52に入力される。

[0074] 図38及び図39には、目的音抽出部52における処理方式の一例を示す。図38は

、 θ_1 と θ_2 とがマイク11、12を結ぶ直線の垂線に対して対称な角度である場合の処理方式の一例を示し、図39は、 θ_1 と θ_2 とが前記垂線に対して対象な角度でない場合の処理方式の一例を示す。

[第9実施形態]

次に、第9実施形態について説明する。図40は、第9実施形態に係る車載機器制御用音声認識システムの構成を示す図である。本実施形態においては、自動車等の車両に設けられた車載機器制御用音声認識システムに本発明に係る音源分離装置1を応用した例を示している。本応用例では、運転席と助手席の音声を2つのマイク10、11で取得し、音源分離装置1で運転席と助手席の音声を分離し、分離した運転席と助手席の音声をそれぞれ発話区間検出、音声認識処理、音声認識結果と自動車走行状態やその他運転状態によって有効な認識結果のみを使用して、機器の制御、システムの応答などを行うことで、車載機器制御用音声認識システムの信頼性の向上と車載機器制御用音声認識システムの応答の自由性の拡張などを提供するものである。

[0075] 車載機器制御用音声認識システムは、本システムに特徴的なデータとして、記憶装置に助手席側認識語彙リスト190、運転席側認識語彙リスト191、助手席側有効語彙リスト210、及び、運転席側有効語彙リスト211を記憶している。運転席側認識語彙リスト191は運転席側から発せられる語彙の候補の一覧であり、助手席側認識語彙リスト190は助手席側から発せられる語彙の候補の一覧である。運転席側有効語彙リスト211は、車両の状態(自動車走行状態やその他運転状態)に応じた運転席側における有効な語彙リストである。助手席側有効語彙リスト210は、車両の状態に応じた助手席側における有効な語彙リストである。ここで、「有効」とは、語彙(音声コマンド)に応じた制御命令を出力することが許されている状態をいう。

[0076] 本システムの動作を図40を用いて説明すると、運転者および助手席の同乗者が発話した音声は2つのマイク10、11で収録され、音源分離装置1で運転席と助手席との音声中に分離されたのち、それぞれ運転者用、助手席の同乗者用に用意された発話区間検出部170、171、音声認識部180、181に入力される。このとき、本発明に係る音源分離装置1の出力において2人の音声は精度よく分離されているために、

助手席側発話区間検出部170、運転席側発話区間検出部171で両者の発話区間を精度よく分離出来ると同時に、助手席側音声認識部180、運転席側音声認識部181に相手の音声を抑圧した情報を与えることが可能で、相手の発話に影響を受けない音声認識処理を精度良く行うことが出来る。

[0077] 本応用例では、各音声認識部180、181にそれぞれ専用にシステムの状態に関係なくどのような語彙を認識すべきかを指示するための助手席側認識語彙リスト190、運転席側認識語彙リスト191が提供されており、各音声認識部180、181はこの語彙リストに従って音声認識処理を行い、音声認識結果を制御部・状態遷移部200に出力する。

制御部・状態遷移部200が備える状態遷移部201は、音声認識結果及び現在の状態に基づいて次の状態に遷移できるようになっている。制御部・状態遷移部200が備える制御部202では、状態遷移部201から得られる現在の状態と、音声認識部180、181からの音声認識結果とに基づいて、助手席側、運転席側それぞれに用意された助手席側有効語彙リスト210、運転席側有効語彙リスト211に基づいて、どの音声コマンドに応答(制御命令を出力)して良いかを調べる。同図に示す有効語彙リスト210、211においては、状態と音声コマンドの組合せに対して“○”が対応付けられている場合に、音声コマンドに応答可能なことを示している。例えば、運転席側においては、状態が“運転中”の場合に応答が許されている音声コマンドは“もっと”、“ライトつけて”、“すずしく”であり、また、“エアコン操作”、“CD”、“MD”は禁止されている。

[0078] そして、許されている音声コマンドのみに応答し、エアコンを操作したりライトをつけたりする制御を行うことで、車両の搭乗者は車内において快適に過ごすことができる。また、車載機器制御用音声認識システムの信頼性を上げることができると同時に、音声認識を使用したアプリケーション作成に仕様設計の自由度をより多く提供することが可能となる。

以上の使用例によれば、運転席と助手席の同乗者とから同時に発話された音声を同時に認識することが可能になったり、どちらか1人が発話した場合でもそれが運転席側からの発話なのか助手席側からの発話なのかを確実に検出して認識することが

可能となるため、同乗者の行動を制限せず、発話者とその発話者の音声コマンドに対する応答を個別に設計することが出来るようになる。

[第10実施形態]

次に、第10実施形態について説明する。図41には、本実施形態に係る携帯電話機300を示している。携帯電話機300には、マイク10、11及び図示せぬ音源分離装置が搭載されている。この携帯電話機300は、通常はテレビ電話用であるが、モードを切り替えて集音マイクとしても使用可能である。図41(a)はマイク10、11がテレビ電話用マイクとして動作している時の様子を示す図であり、図41(b)はマイク10、11が集音マイクとして動作している時の様子を示す図である。中規模の会議室などで、発表者がマイクを使うほど大規模な会議室でもなく、しかし、部屋が広く小さな声だと発表者の声が聞き取り難い場面で使用することができる。

[0079] 以上説明したように、最小2本のマイクを互いに離して設置し、2本のマイクを結ぶ直線の垂線を中心として左右対称な角度に時間領域あるいは周波数領域にてビームフォーマにより死角を形成し、時間領域で死角を作成した場合には周波数領域に変換し、双方のビームフォーマのパワースペクトルの差分を計算し、得られた結果の係数変換を行うことで、左右の死角を中心として幅を持たせた指向特性を形成し、音源分離を行うことが可能となる。このように、マイクの素子感度に指向特性が影響を受けないという性質を実現することができ、マイクの素子感度のバラつきに影響を受けず、音源到来方向の想定方向とのずれや大きな初期反射に関しても適度に広い指向特性によりカバーし、安定した2つの音源の分離特性を実現することが可能となる。

[第11実施形態] 次に、第11実施形態について説明する。図42には抽出したい目的音がゾーンA(例えば運転席のゾーン)に存在し、それ以外の場所(ゾーンB, ゾーンC, ゾーンD)に妨害音が存在する状況下において、ゾーンAの目的音を抽出する例が示してある。本手法を使用したアレイマイク2001(例えば車室内で前方に配置(例えばルームミラーに設定))を使用した場合、ゾーンA/C(例えば運転席及びその後部座席のゾーン)とゾーンB/D(例えば助手席及びその後部座席のゾーン)に存在する音を分離することは可能であるがゾーンA(例えば運転席のゾーン)とゾーンC

(例えばその後部座席のゾーン)とに存在する音を分離することは出来ない。しかし、本手法を使用したアレイマイク2002を同図に記載されるようにゾーンA/BとゾーンC/Dの境界の位置に配置することでゾーンA/BとゾーンC/Dに存在する音を分離することが可能となるためゾーンAの音のみを抽出することが可能となる。

[0080] 具体的には、ゾーンA, B, C, Dに居る発話者A, B, C, Dが同時に発話した場合、まずゾーンA/BとゾーンC/Dの境界に配置したアレイマイク2002を使用することでゾーンA/Bからの音とゾーンC/Dからの音を分離することが可能となる。次にアレイマイク2001によりゾーンA/Cからの音とゾーンB/Dからの音を分離することが出来る。最後にアレイマイク2001を使用して求めたゾーンA/Cからの音とアレイマイク2002を使用して求めたゾーンA/Bからの音を各周波数領域で比較し双方に共通にある周波数成分をゾーンAからの音として分離することが可能となる。同様な処理により、各ゾーンB, C, Dからの音も個別に求めることが可能である。

[第12実施形態] 次に、第12実施形態について説明する。図43は、自動車などの環境下において音声認識により機器操作を行う状況を想定している。図44は、そのときの機器操作のガイダンス音声と発話者の音声との関係を示す。

[0081] この様な場合、スピーカ15から発話者Aの発話のタイミングを促す目的で“目的地をどうぞ”などのガイダンス音の流れの後、機械音たとえば“ピッ”などの様な音の流れ、その後に発話者Aは音声コマンドを発声する。しかし、ユーザが音声コマンドによる機器操作に慣れてくるにつれ、図44に示すように、ガイダンス音声中に発話者Aが発話を始めてしまう状況が発生するようになり、これが音声認識性能を下げる要因になっている。

この様な状況への対策として、一般的にはエコーキャンセラを使用しマイク10からの収録音に混在するガイダンス音声を適応的に推定し取り除くことで対応している。また別の対応策として、図45～図48に示すように、周波数分析後のマイク10への1入力信号に関して、スペクトルサブトラクションを行ったり(図45、図46)、各周波数成分毎にガイダンス音声、発話者Aの音声のどちらが多く含まれるかを推定し発話者Aの音声のみが多く含まれる周波数成分のみを発話者Aの音声として取り出すことが行われている(図47、図48)。

[0082] ここで、図45、図47において、フィルタ部1001は、スピーカ15からマイク10までの音響的な反射パスを模擬したフィルタであり、予め求めておいたスピーカ15からマイク10までのインパルス応答を使用しても良いし、適応フィルタ処理により動的に求めておいても良い。

図45における、ゲイン操作部1002は、スペクトルサブトラクションを行う際に使用するオーバーサブトラクションファクタを決定する部分であり、スピーカ15の音量に応じて1~10程度のゲインのなかから選択され使用する。

[0083] また、図45における、目的音声抽出部1003は、ゲイン操作部1002及びスペクトル分析部21の出力に基づいて、図46に示すような処理を行い、処理結果の信号を、時間波形変換部1004に出力する。

図47における、閾値計算部1011は、ガイドンス音声の平均エネルギーを元に閾値 th を決定する。

また、図47における、目的音声抽出部1012は、閾値計算部1011及びスペクトル分析部21の出力に基づいて、図48に示すような処理を行い、処理結果の信号を、時間波形変換部1004に出力する。なお、図48に示す th_{min} は、同図に示す $X_{Far}(\omega)$ が有効な入力であると判断するための閾値となる。

[0084] また、時間波形変換部1004は、第1の実施形態における時間波形変換部70、71と同様な処理を行う。

しかし、従来方においては図43のように、発話者Aのみが発話する状況においては上述のような構成により対応可能であるが、図49に示すように発話者A(例えば運転席の乗員)のみならず発話者B(例えば助手席の乗員)がいる場合に、図50に示すように、発話者Aのみならず発話者Bも何らかの発話をする場合が起こりえるが、このような状況には対応することが出来ない。

[0085] このような状況に対し、図51に示すように本手法と図45または図47で示したガイドンス音声削除部1021、1022とを組み合わせることで対応が可能となる。

具体的には、図51においてガイドンス音声は車内のスピーカ15から再生され、同時に発話者A、発話者Bとも発話している状況下において、マイク10およびマイク11には、ガイドンス音声、発話者Aの発話音声、発話者Bの発話音声为重畳されて入

力されている。このとき、ガイドンス音声削除部1021およびガイドンス音声削除部1022では図45または図47に示した方法により、ガイドンス音声を除去し、結果として発話者Aの音声と発話者Bの音声の両者の音声为重畳された信号を結果として出力する。また、ここでは後処理として使用する本手法への入力時の計算の無駄を省くために、時間波形には戻さず、周波数成分のまま本手法(図1)へ入力する。後段の本手法においても入力として周波数成分情報が入力されるためスペクトル分析部の処理を省き、ビームフォーマ部へ直接入力を行い、本手法を適用し処理を行うことにより出力結果として、発話者Aの音声と発話者Bの音声を個別に得ることが可能となり、音声認識装置の信頼性、性能の向上とアプリケーションの自由度を大幅に拡大することが出来る。

[0086] また、上述した各種機能の組合せやマイクからの信号の遅延操作により、正面方向に狭い指向特性を実現したり、特定方向からの音源信号のみを検出することが可能となる。

加えて、マイク間隔を広げなくとも低域まで高い分離性能を確保できるため、実装スペースを削減することが可能であり、携帯機器などへの使用も可能となる。

なお、上述した実施形態においては、音源分離システムの各機能ブロックをプログラムで実現するとして説明したが、回路等を用いることによりハードウェアで実現することも可能である。

産業上の利用の可能性

[0087] 音声認識装置、カーナビゲーション、集音装置、録音装置、音声コマンドによる機器の制御等、音源を精度よく分離する必要があるあらゆる産業に利用可能である。

請求の範囲

- [1] 互いに離して配置された少なくとも2つのマイクロホンに対して複数の音源から発せられた音源信号が混合された混合音を入力し、該混合音から目的音源からの音源信号を分離する音源分離装置において、
- 前記マイクロホンからの出力信号に対して第1の係数を用いた演算を行うことにより、所定の方向から到来する音源信号を減衰させるための第1のビームフォーマ処理を行い、
- かつ、
- 前記マイクロホンからの出力信号に対して、前記第1の係数と周波数領域で複素共役の関係にある第2の係数を用いた演算を行うことにより、2つのマイクロホンを結ぶ直線の垂線に対して前記所定の方向と対称な方向から到来する音源信号を減衰させるための第2のビームフォーマ処理を行うビームフォーマ手段と、
- 前記ビームフォーマ手段により得られた音源信号各々についてのパワースペクトル情報を計算するパワー計算手段と、
- 前記パワー計算手段により計算されたパワースペクトル情報同士の差分に基づいて、目的音源のスペクトル情報を抽出する目的音スペクトル抽出手段と
- を備えることを特徴とする音源分離装置。
- [2] 前記ビームフォーマ手段は、
- 互いに離して配置された3つのマイクロホンのうち、何れか2つのマイクロホンの組合せ及び他の2つのマイクロホンの組合せ各々について、前記第1のビームフォーマ処理及び前記第2のビームフォーマ処理を行うことを特徴とする
- 請求項1に記載の音源分離装置。
- [3] マイクロホンからの出力信号に対して遅延を与える指向性制御手段をさらに備えることを特徴とする
- 請求項1又は2に記載の音源分離装置。
- [4] 前記指向性制御手段は、
- 2つのマイクロホンのうち少なくとも一方のマイクロホンからの出力信号に対して遅延を与えることにより、仮想的に3つのマイクロホンからの出力信号を生成することを特

徴とする

請求項3に記載の音源分離装置。

- [5] 前記音源信号が到来する方向を推定する到来方向推定手段をさらに備え、
前記指向性制御手段は、
前記到来方向推定手段により推定された到来方向に基づいて、2つの音源の位置が仮想的に2つのマイクロホンを結ぶ直線の垂線に対して対称となるように、該マイクロホンからの出力信号に対して遅延を与えることを特徴とする

請求項3又は4に記載の音源分離装置。

- [6] 前記目的音抽出手段により抽出されたパワースペクトル情報に対して、スペクトルサブトラクション処理を行うスペクトルサブトラクション手段をさらに備えたことを特徴とする

請求項1から5の何れか1項に記載の音源分離装置。

- [7] 前記ビームフォーミング手段による処理を行う前に、雑音を低減する処理を行う定常雑音低減手段をさらに備えることを特徴とする

請求項1から6の何れか1項に記載の音源分離装置。

- [8] 請求項1から7の何れか1項に記載の音源分離装置により分離された音源信号の音声認識を行う音声認識手段を備えた音声認識装置。

- [9] 車両の運転席側から発せられる語彙の候補の一覧である運転席側認識語彙リスト及び助手席側から発せられる語彙の候補の一覧である助手席側認識語彙リストを記憶する認識語彙リスト記憶手段をさらに備え、

前記音声認識手段は、

前記認識語彙リスト記憶手段に記憶されている運転席側認識語彙リスト及び助手席側認識語彙リストに基づいて、前記音源分離装置により分離された音源信号の音声認識処理を行うことを特徴とする

請求項8に記載の音声認識装置。

- [10] 現在の車両の状態を管理する状態遷移手段と、
車両の状態に応じた助手席側及び運転席側における有効な語彙リストを記憶する有効語彙リスト記憶手段と、

前記状態遷移手段で管理されている現在の車両の状態と、前記有効語彙リスト記憶手段に記憶されている語彙リストとに基づいて、前記音声認識手段により認識された語彙が有効か否かを判断し、該判断結果に応じて制御を行う制御手段とをさらに備えることを特徴とする

請求項8又は9に記載の音声認識装置。

[11] 請求項1から7の何れか1項に記載の音源分離装置を備えた携帯電話機。

[12] 互いに離して配置された少なくとも2つのマイクロホンに対して、複数の音源から発せられた音源信号を入力する音源信号受信ステップと、

前記マイクロホンからの出力信号に対して、周波数領域で複素共役の関係にある2つの重み係数各々を用いて演算を行うことにより、2つのマイクロホンを結ぶ直線の垂線に対して対称な所定の方向から到来する音源信号を各々減衰させるための第1のビームフォーマ処理及び第2のビームフォーマ処理を行うビームフォーマ処理ステップと、

前記ビームフォーマ処理ステップにおいて得られた音源信号各々についてパワースペクトル情報を算出するパワー計算ステップと、

前記パワー計算ステップにおいて計算されたパワースペクトル情報同士の差分に基づいて、目的音源のスペクトル情報を抽出する目的音スペクトル抽出ステップとを備えることを特徴とする音源分離方法。

[13] コンピュータに、

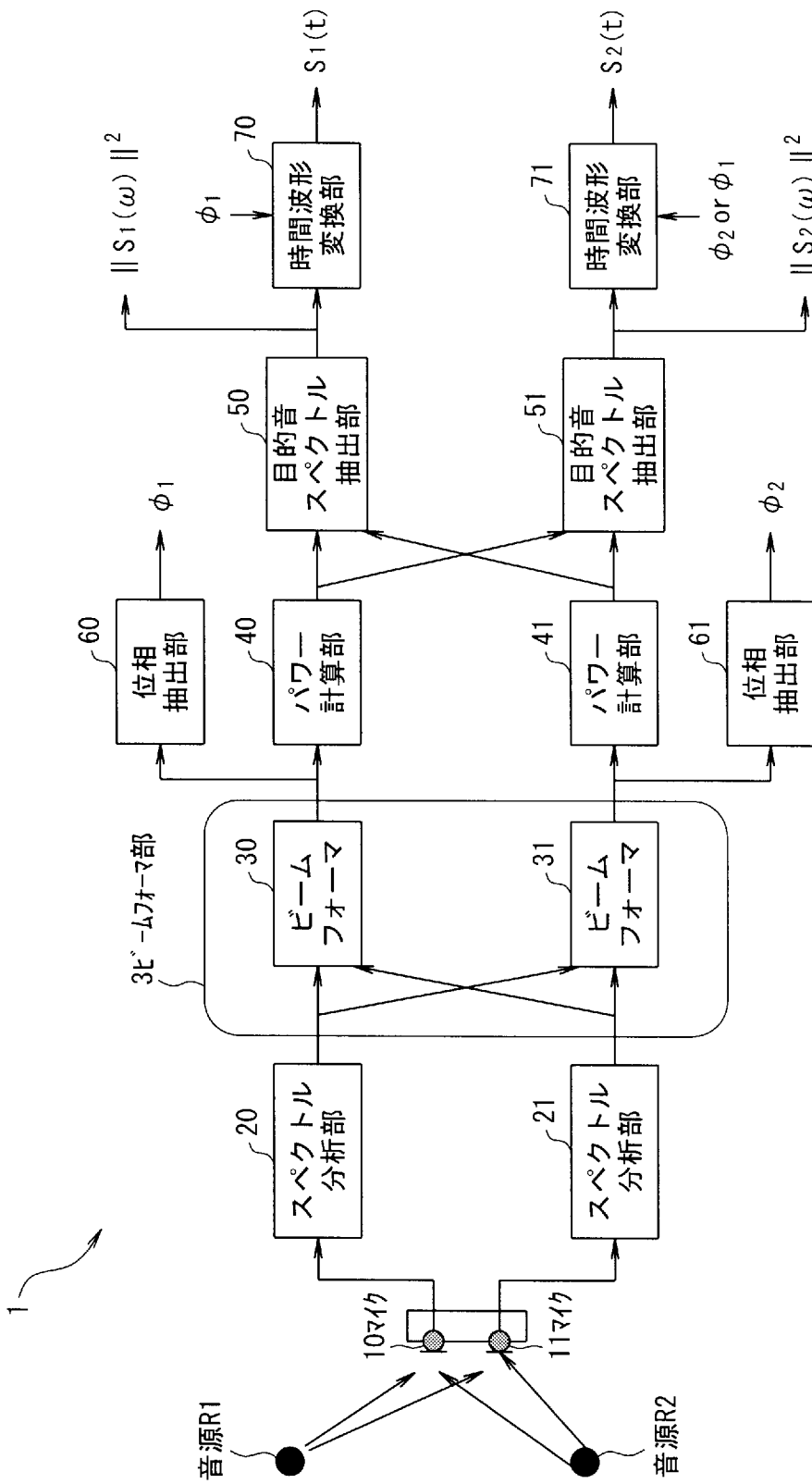
互いに離して配置された少なくとも2つのマイクロホンから、複数の音源から発せられた音源信号が混合された出力信号を取得する出力信号取得ステップと、

前記出力信号取得ステップにおいて取得された出力信号に対して、周波数領域で複素共役の関係にある2つの重み係数各々を用いて演算を行うことにより、2つのマイクロホンを結ぶ直線の垂線に対して対称な所定の方向から到来する音源信号を各々減衰させるための第1のビームフォーマ処理及び第2のビームフォーマ処理を行うビームフォーマ処理ステップと、

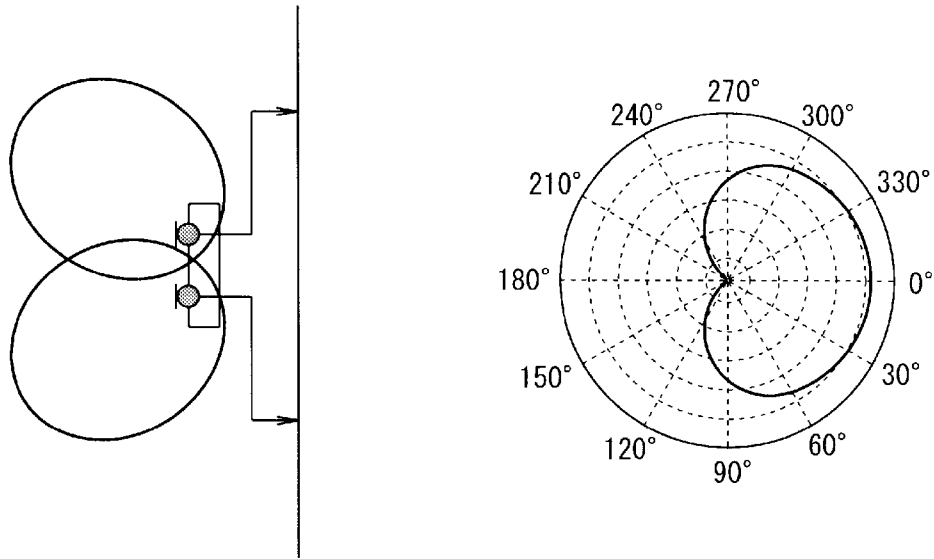
前記ビームフォーマ処理ステップにおいて得られた音源信号各々についてパワースペクトル情報を算出するパワー計算ステップと、

前記パワー計算ステップにおいて計算されたパワースペクトル情報同士の差分に基づいて、目的音源のスペクトル情報を抽出する目的音スペクトル抽出ステップとを実行させるためのプログラム。

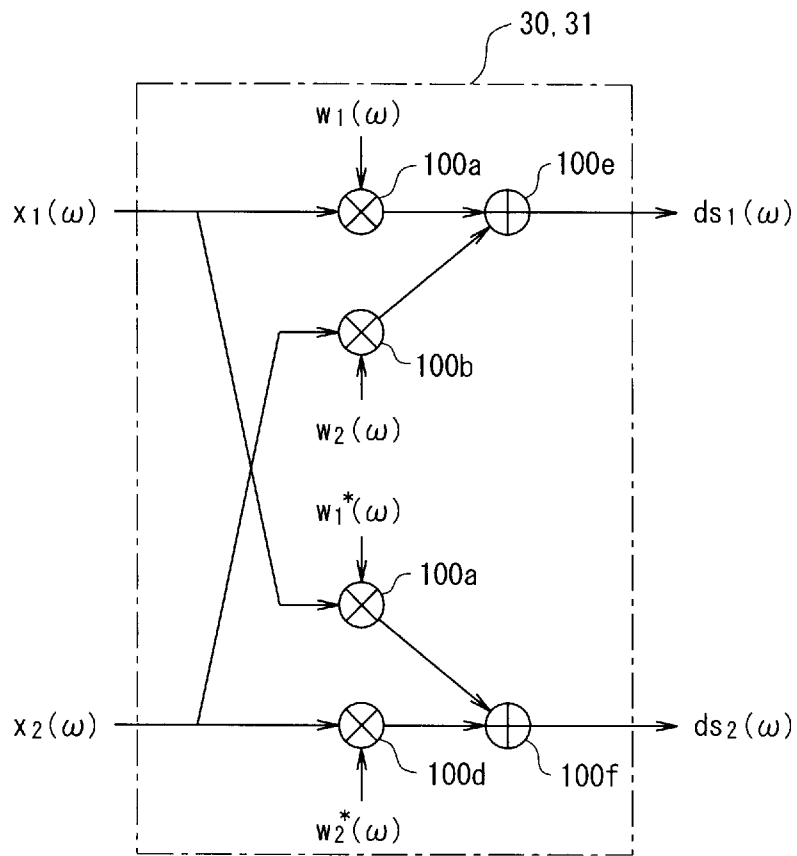
[図1]



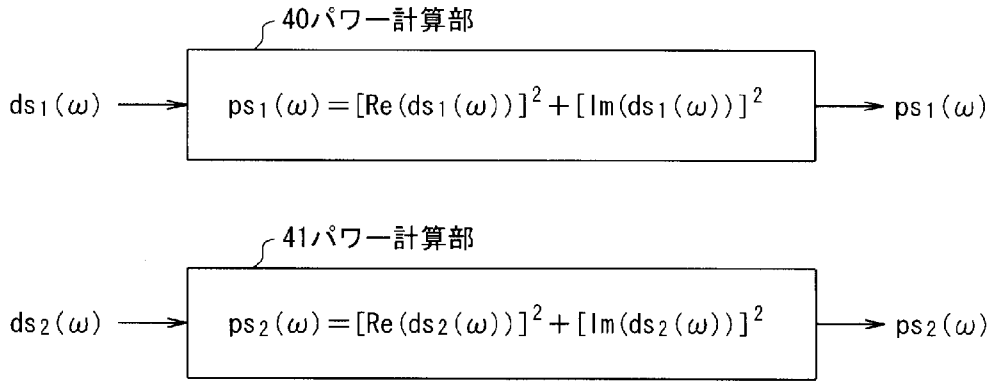
[図2]



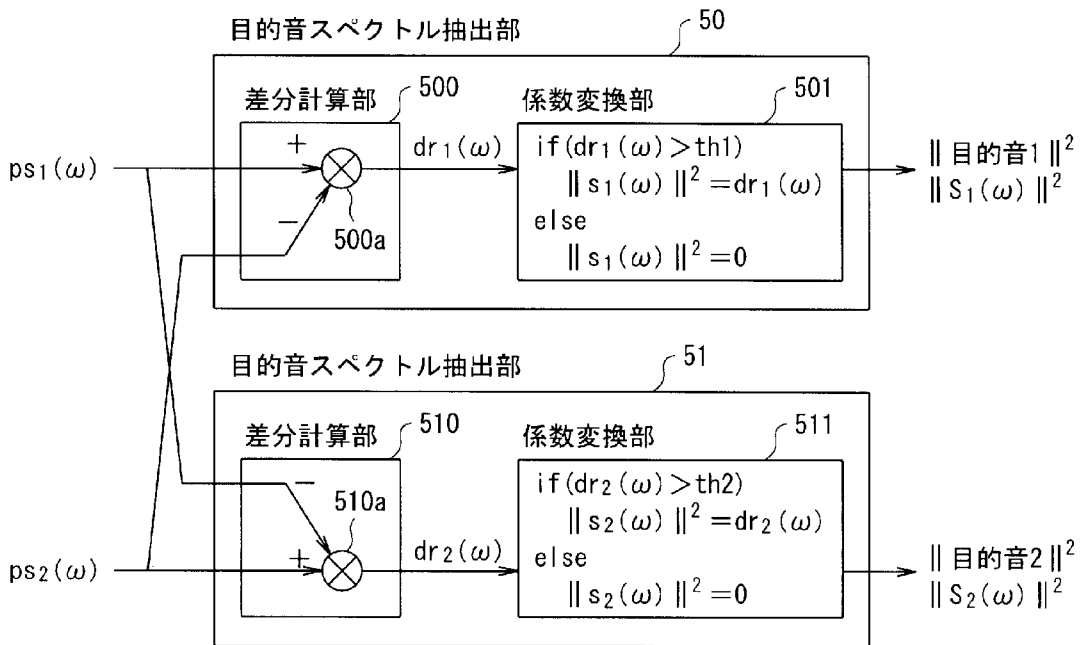
[図3]



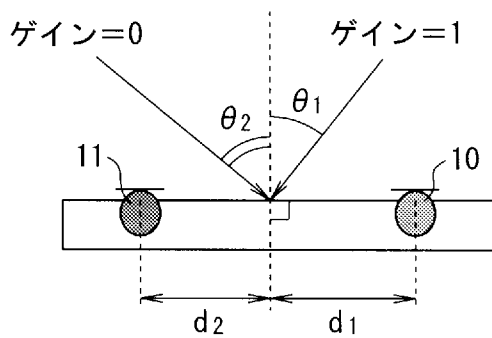
[図4]



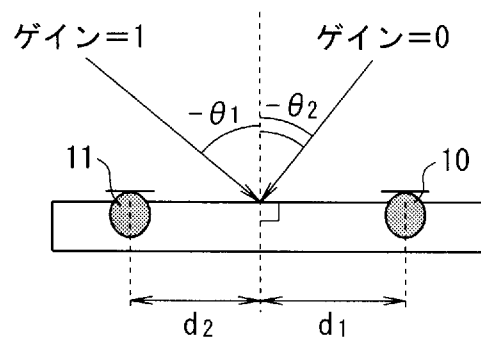
[図5]



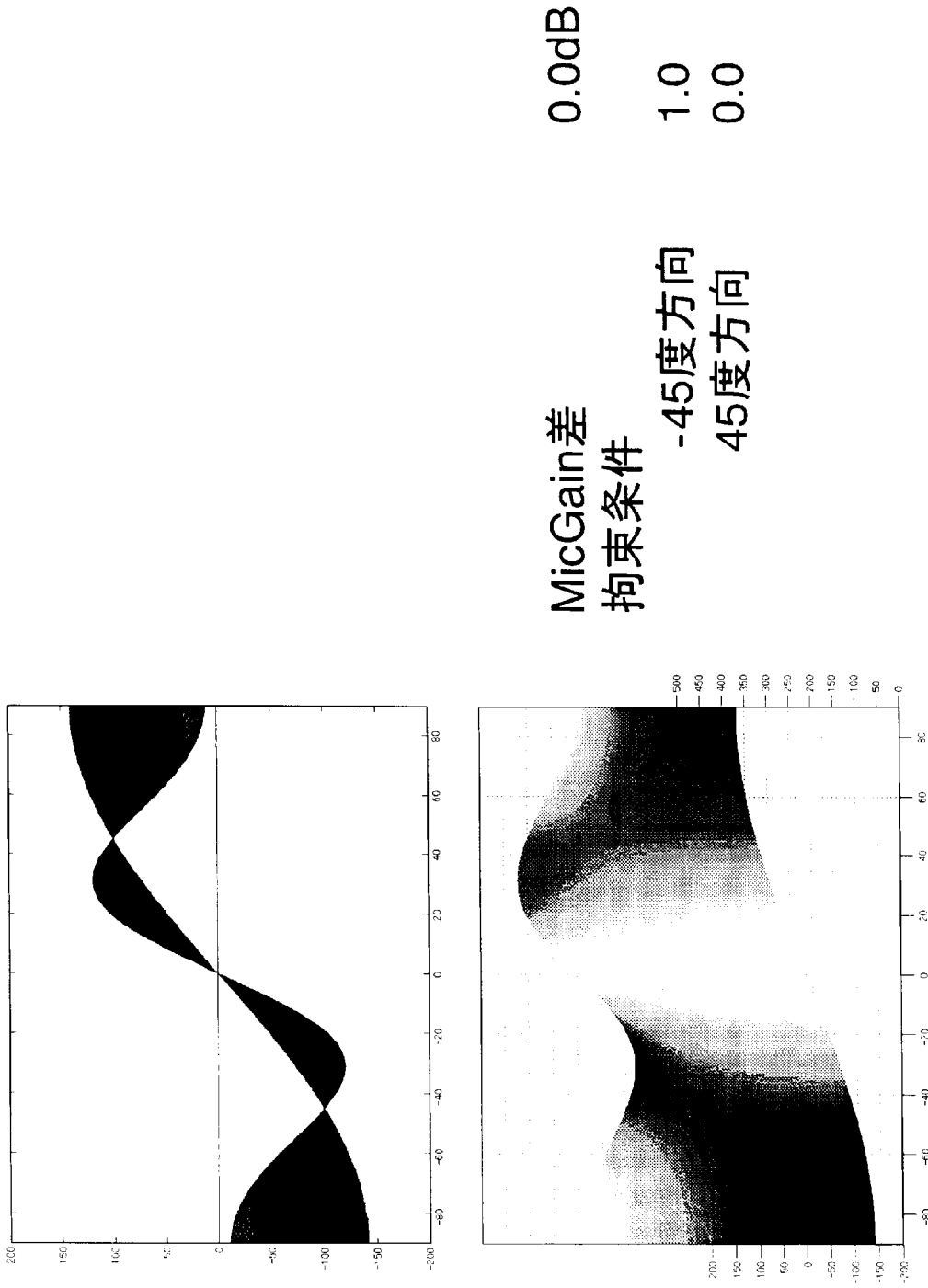
[図6]



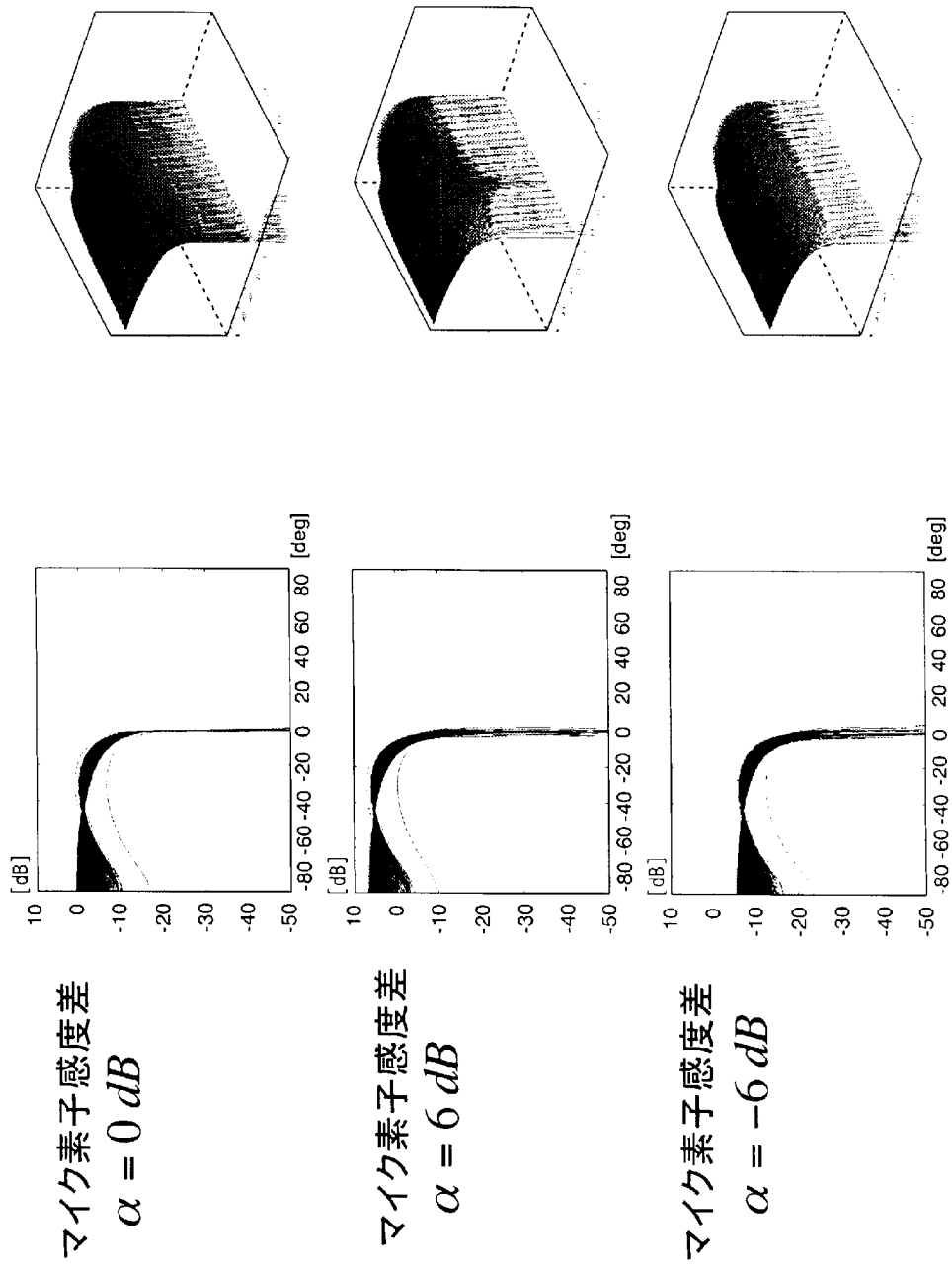
[図7]



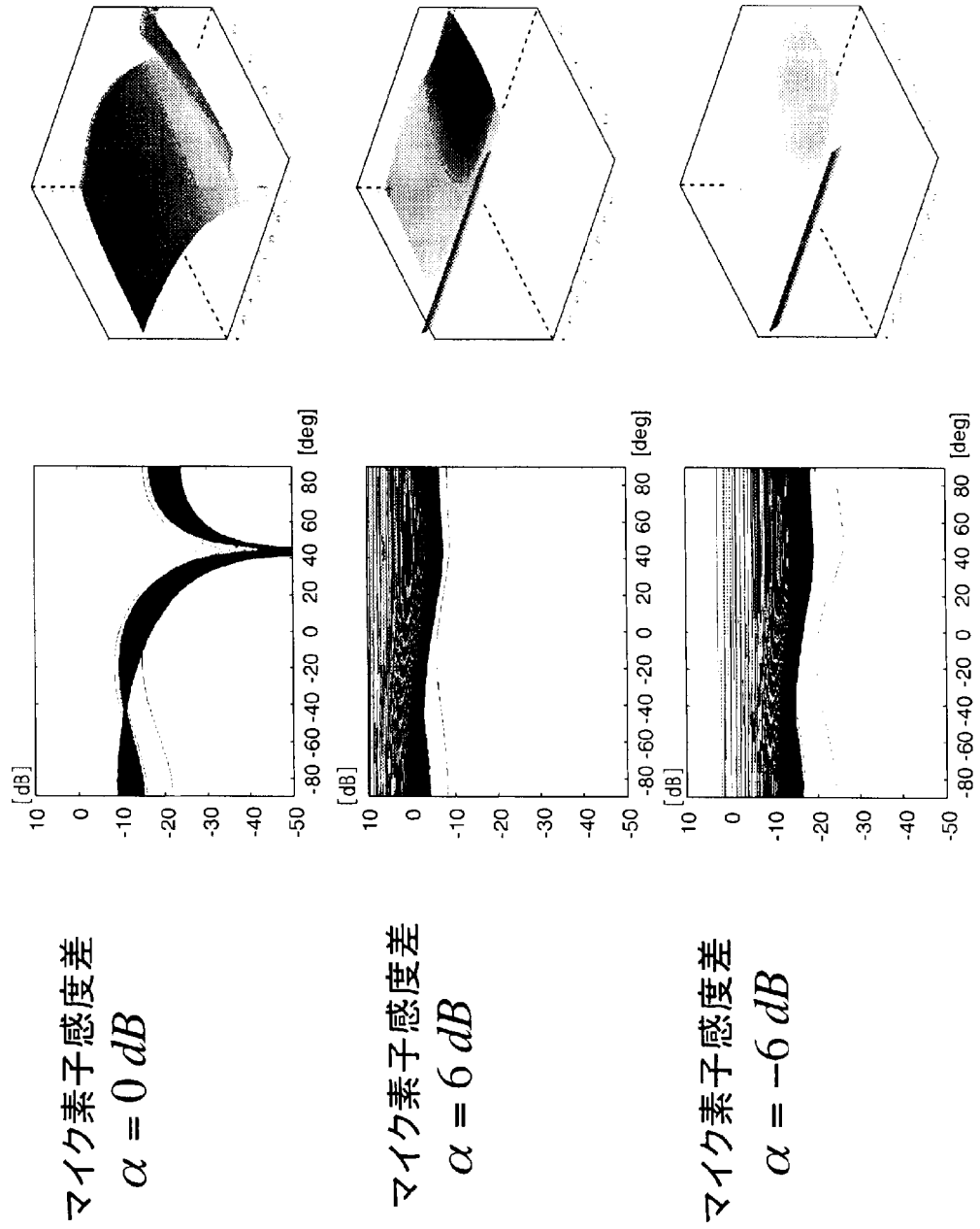
[図8]



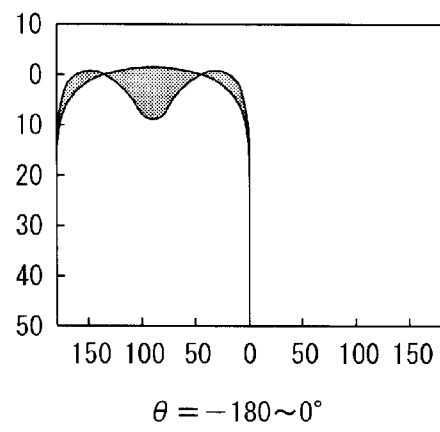
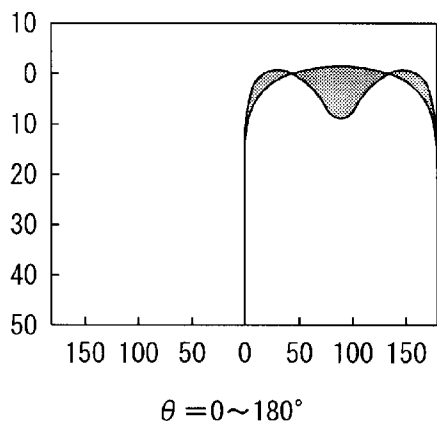
[図9]



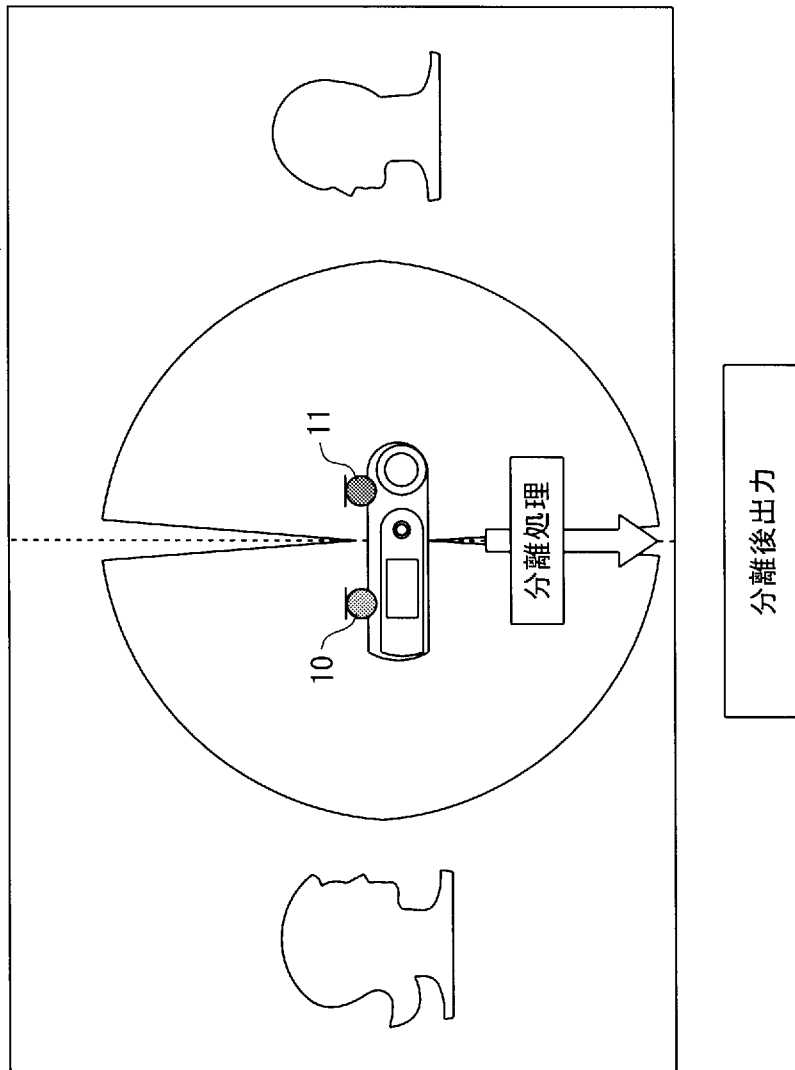
[図10]



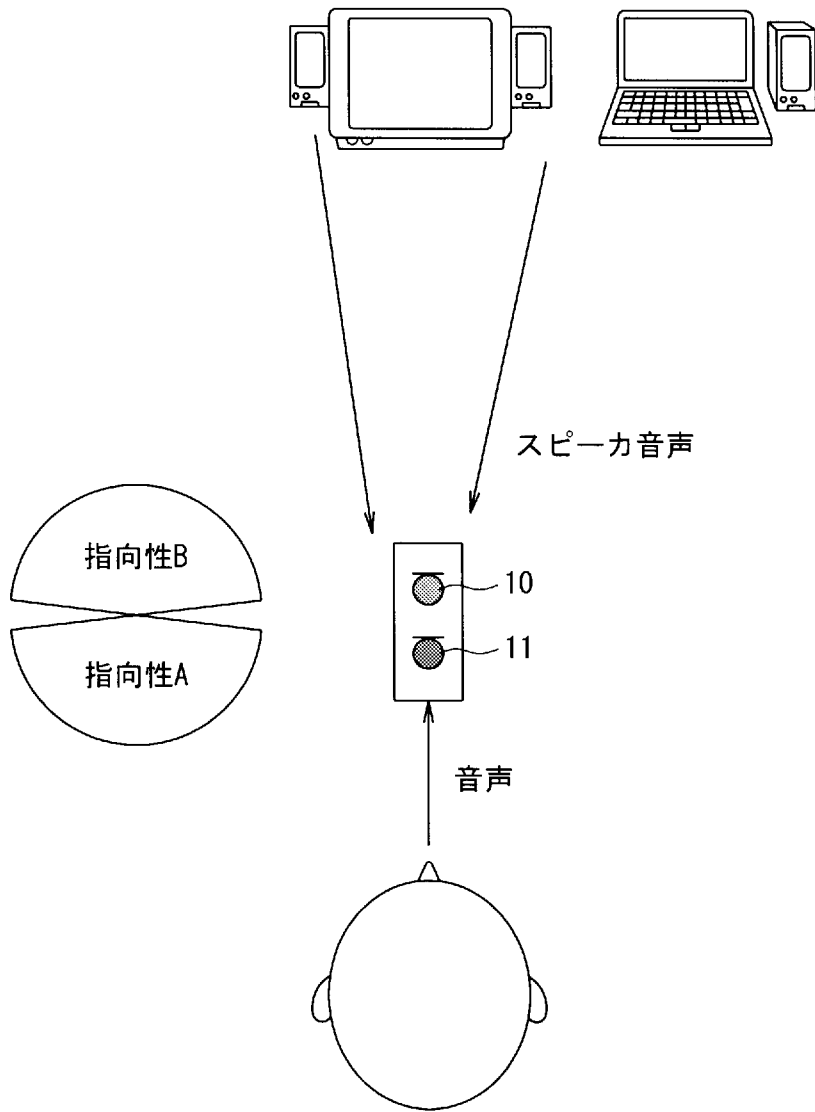
[図11]



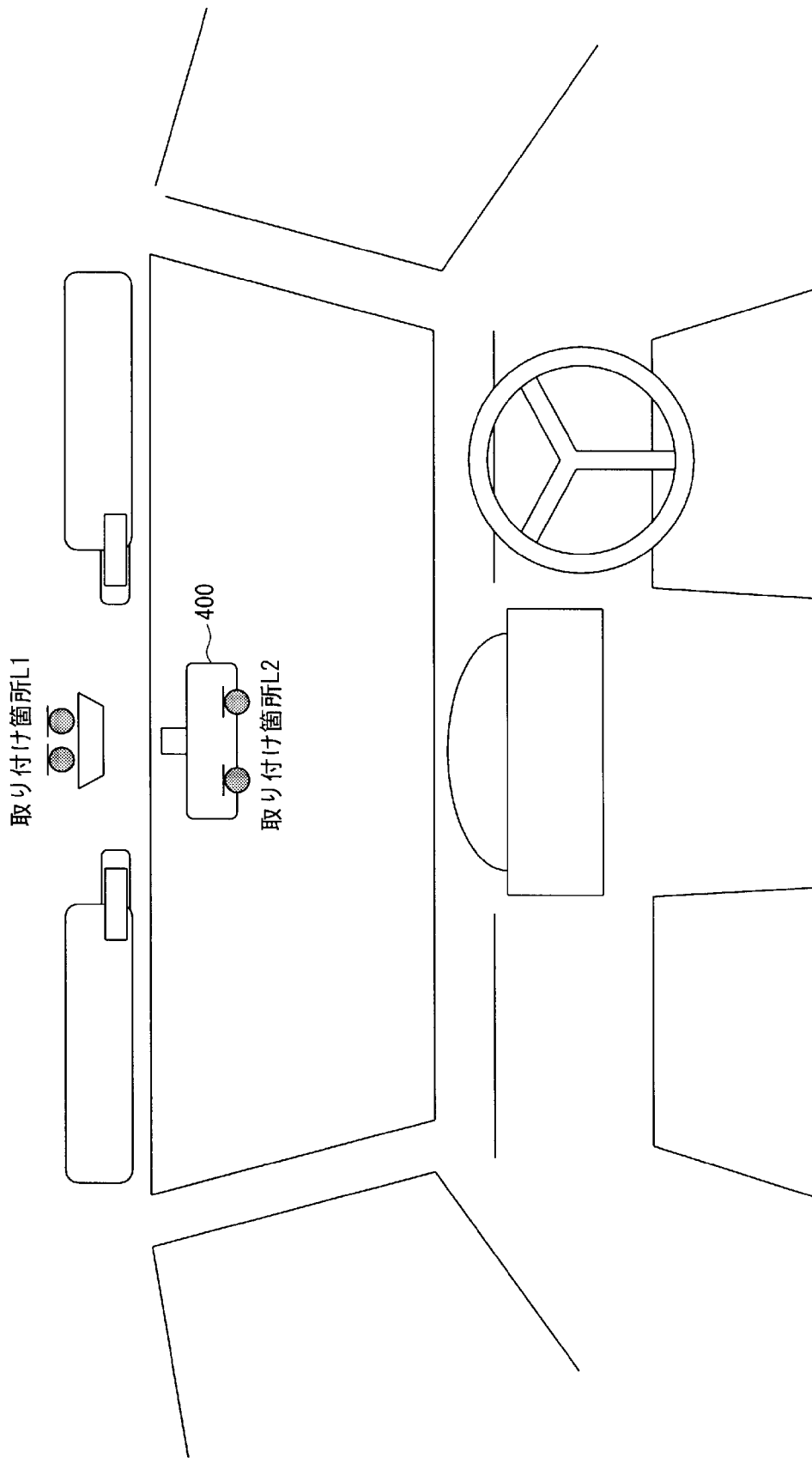
[図12]



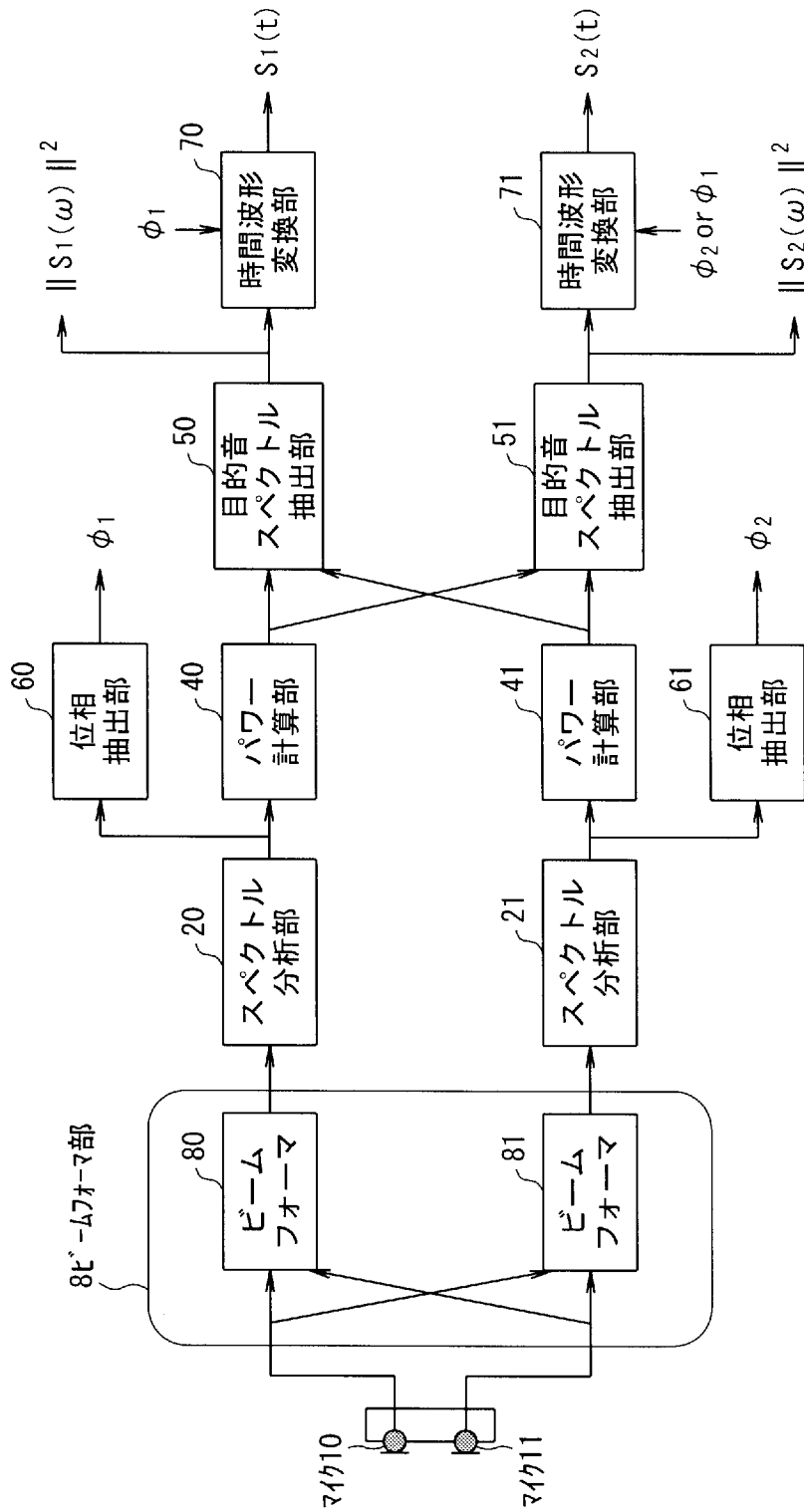
[図13]



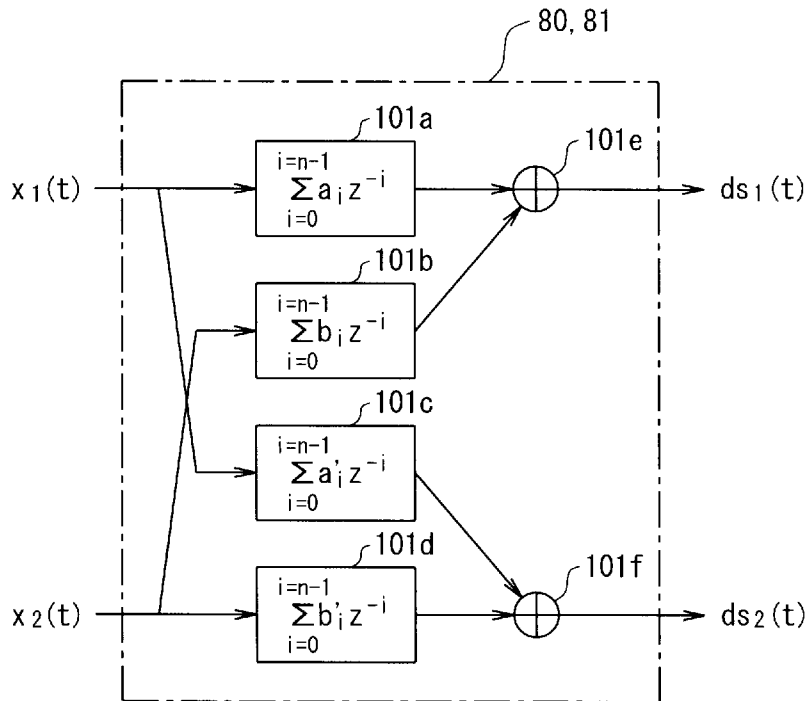
[図14]



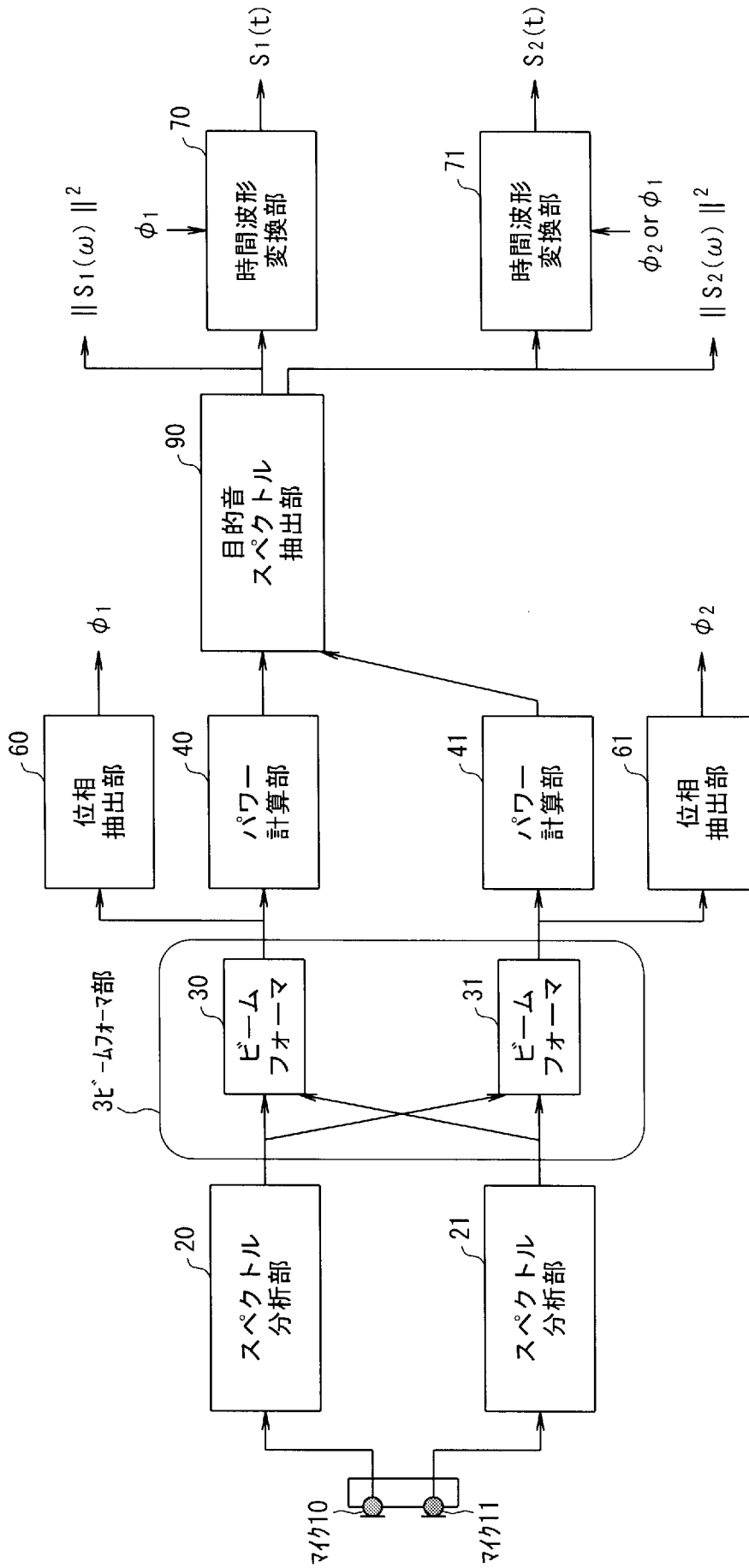
[図17]



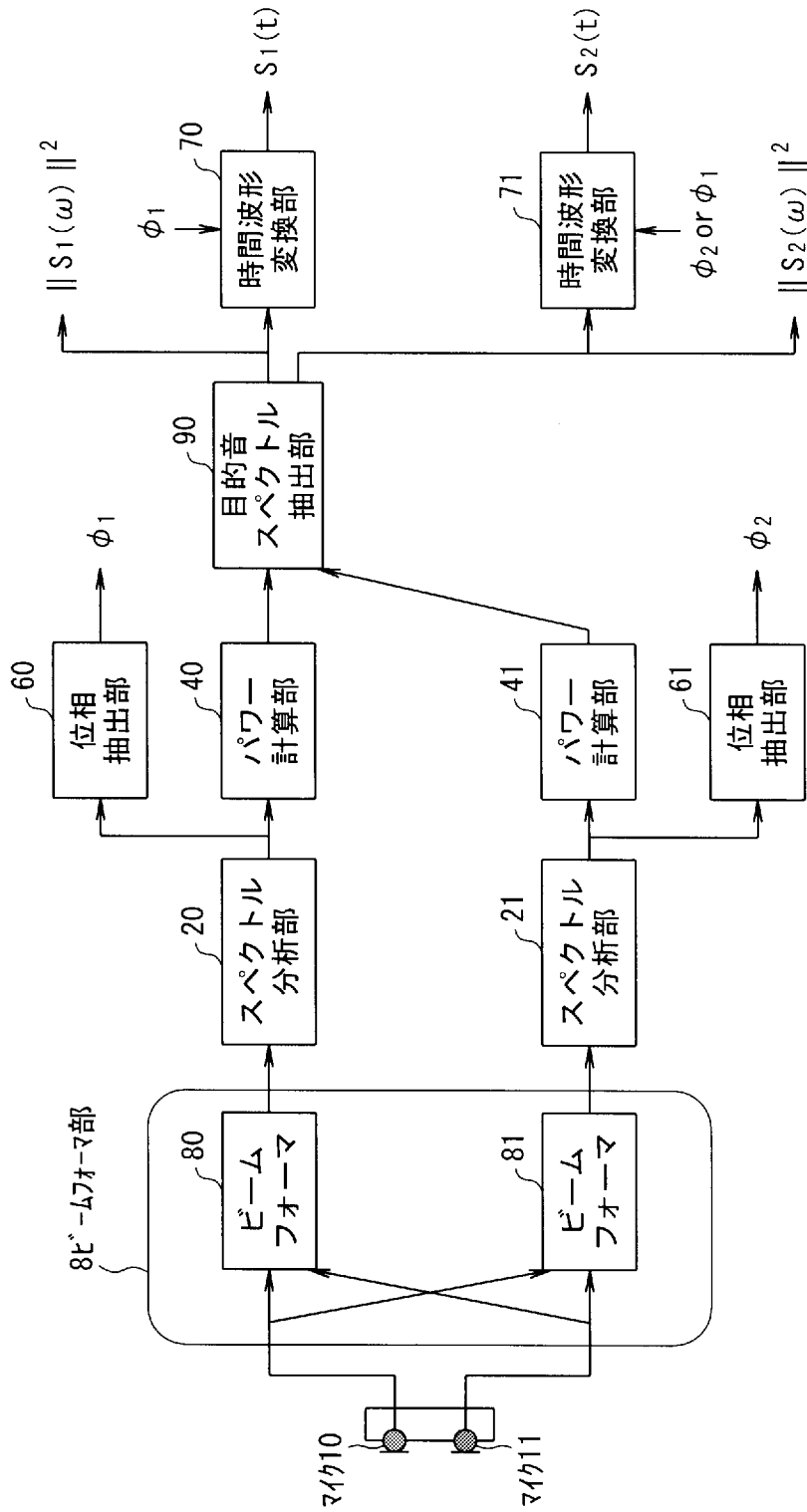
[図18]



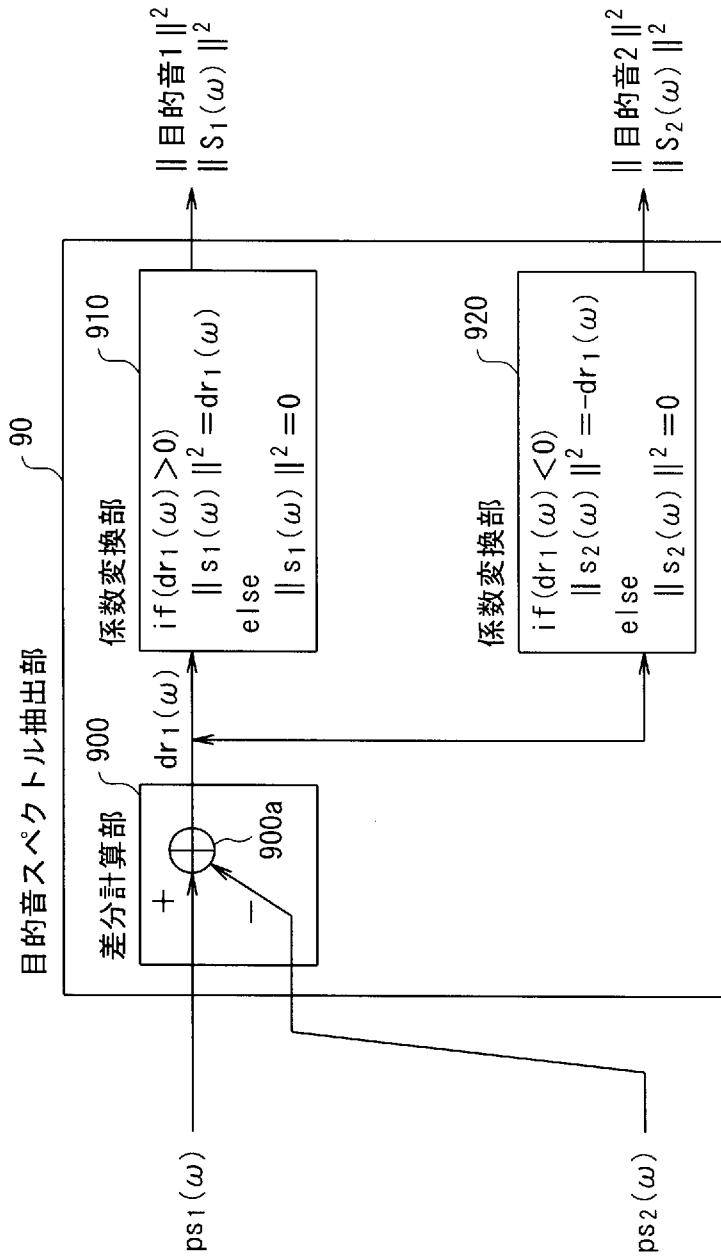
[図19]



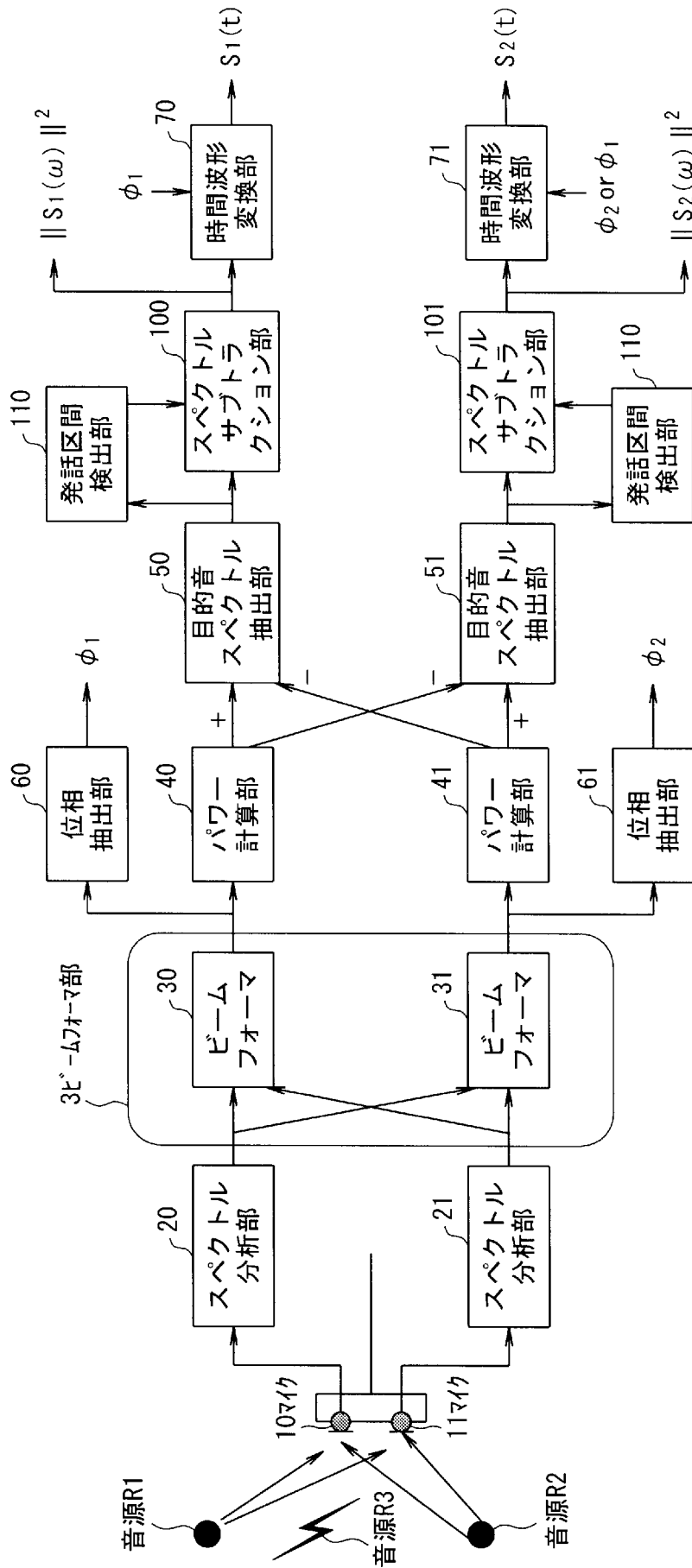
[図20]



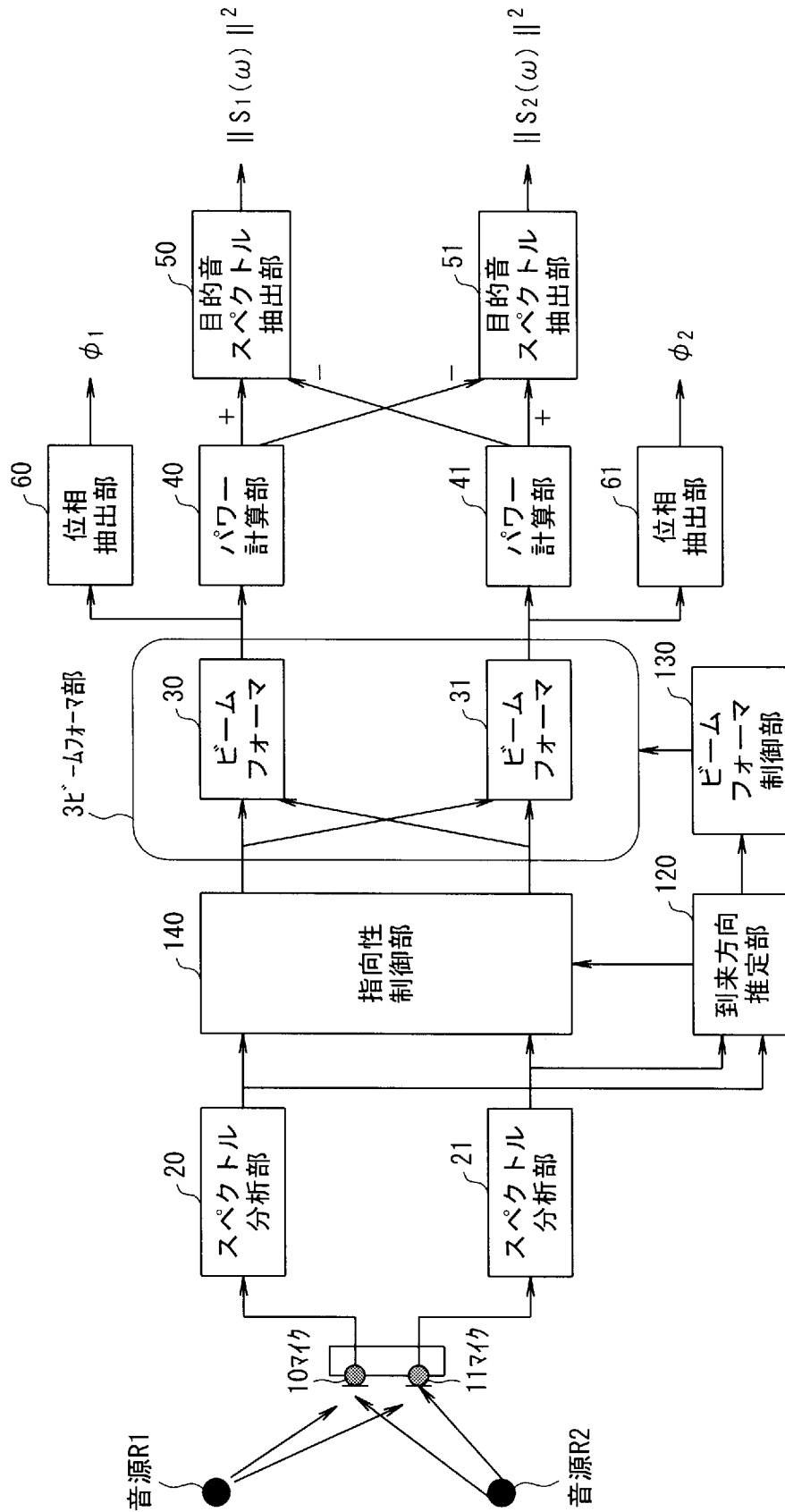
[図21]



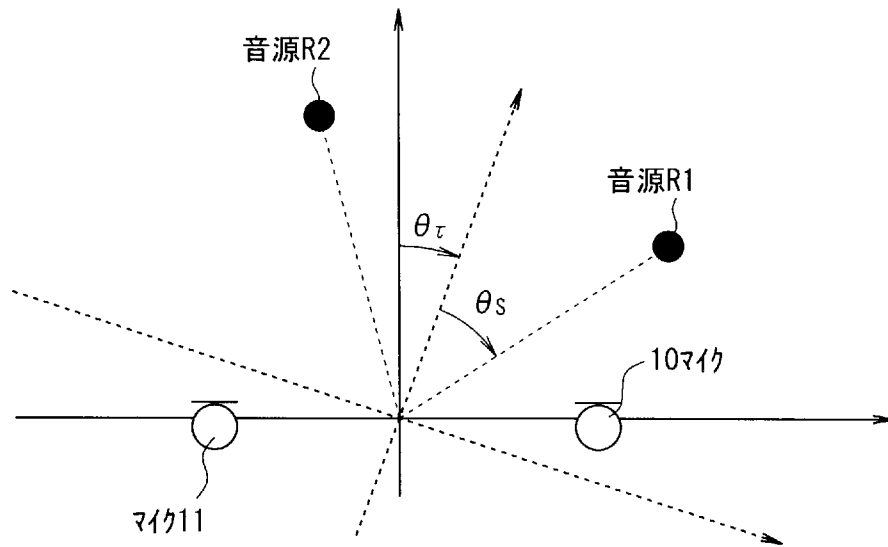
[図22]



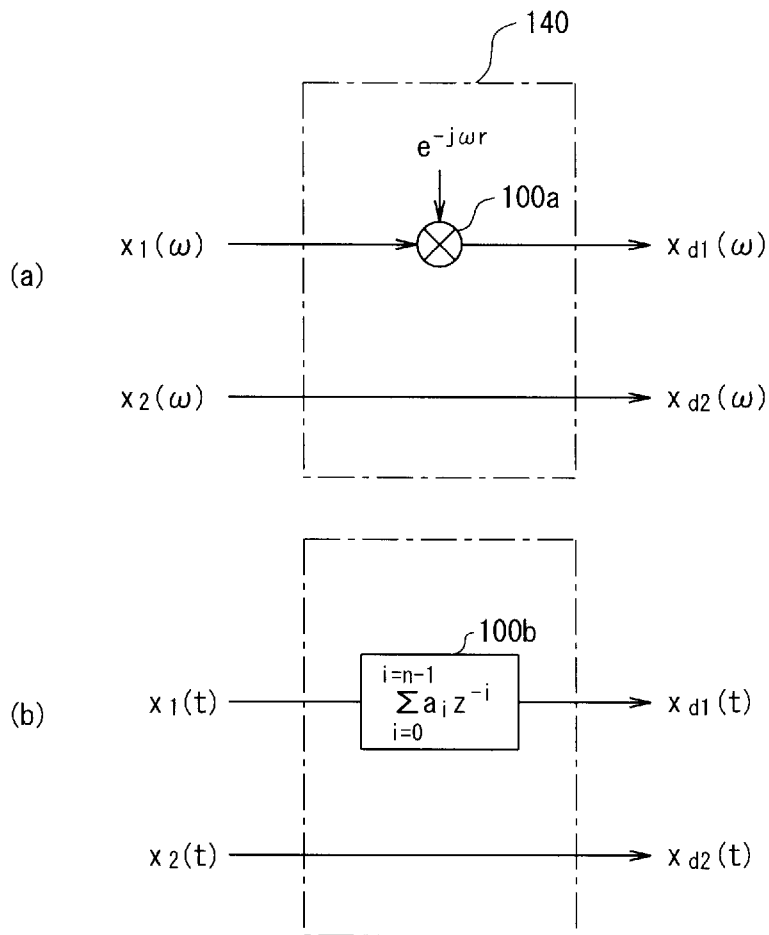
[図23]



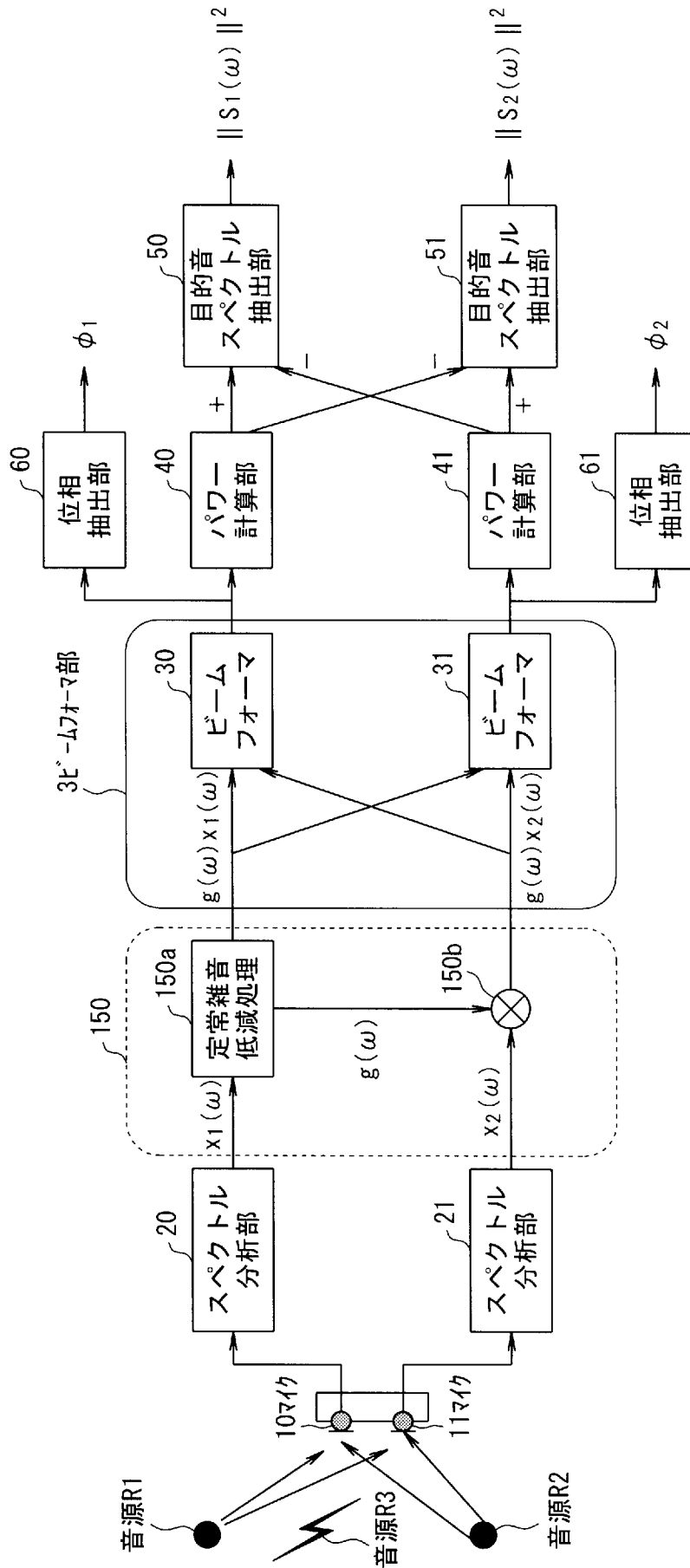
[図24]



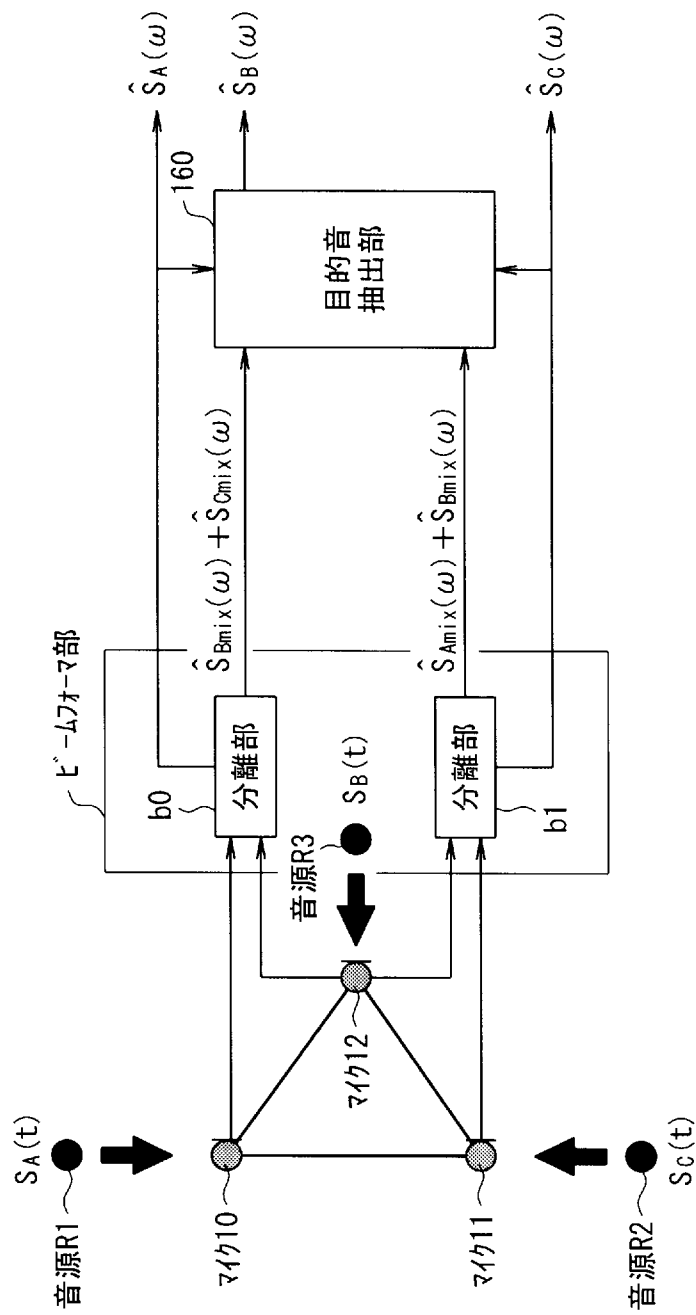
[図25]



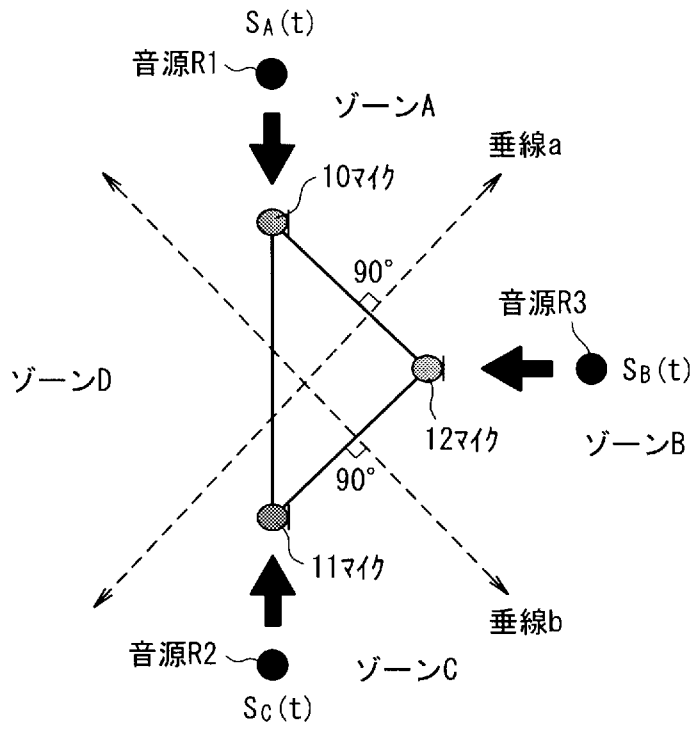
[図26]



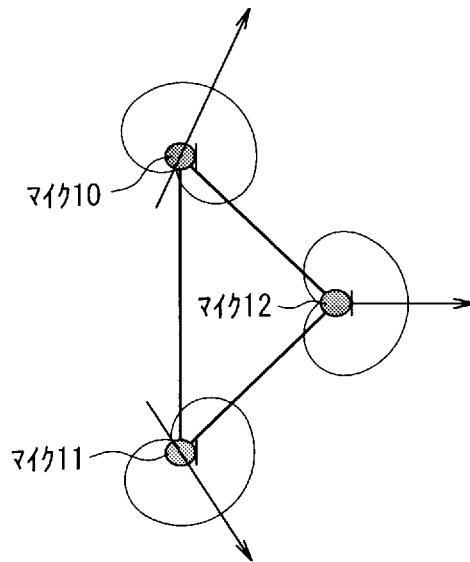
[図27]



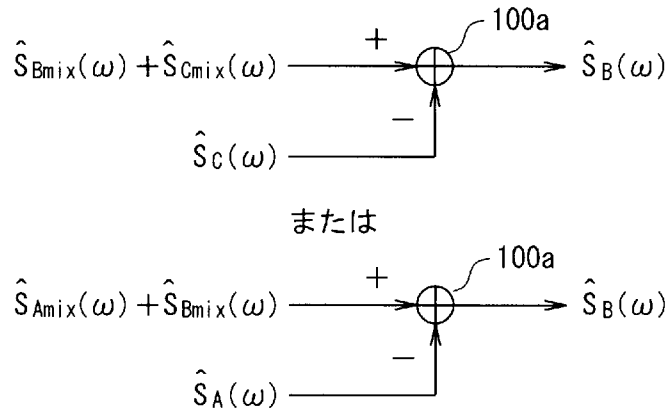
[図28]



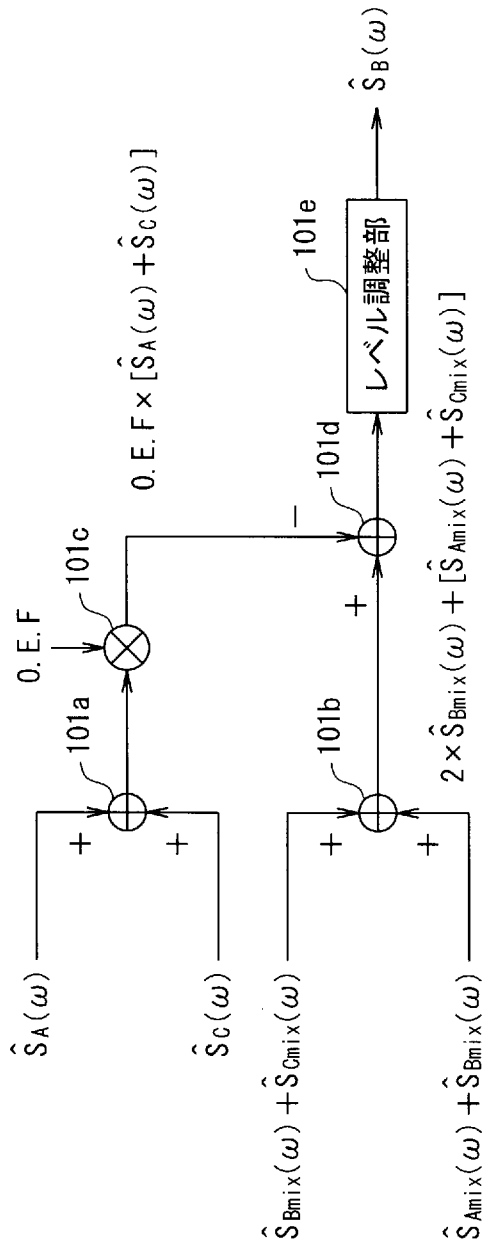
[図29]



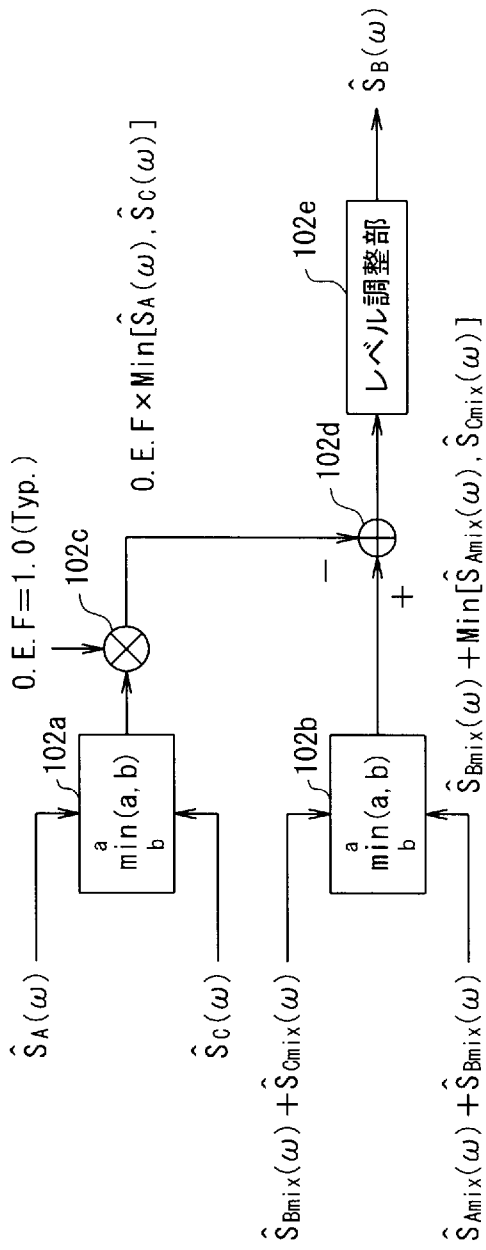
[図30]



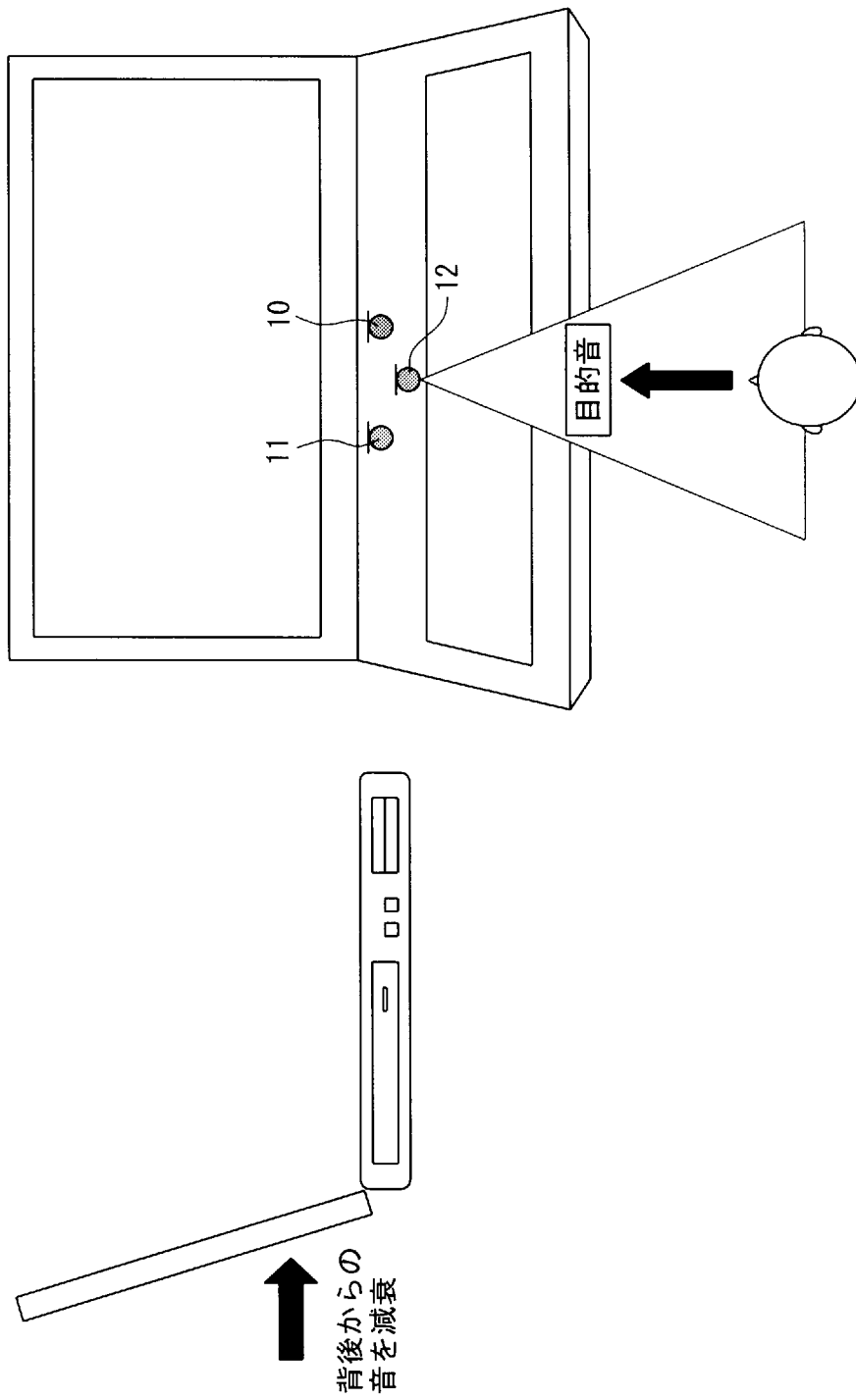
[図31]



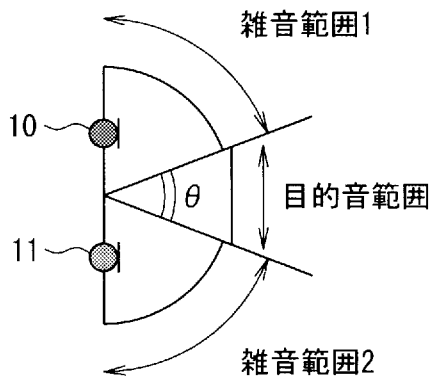
[図32]



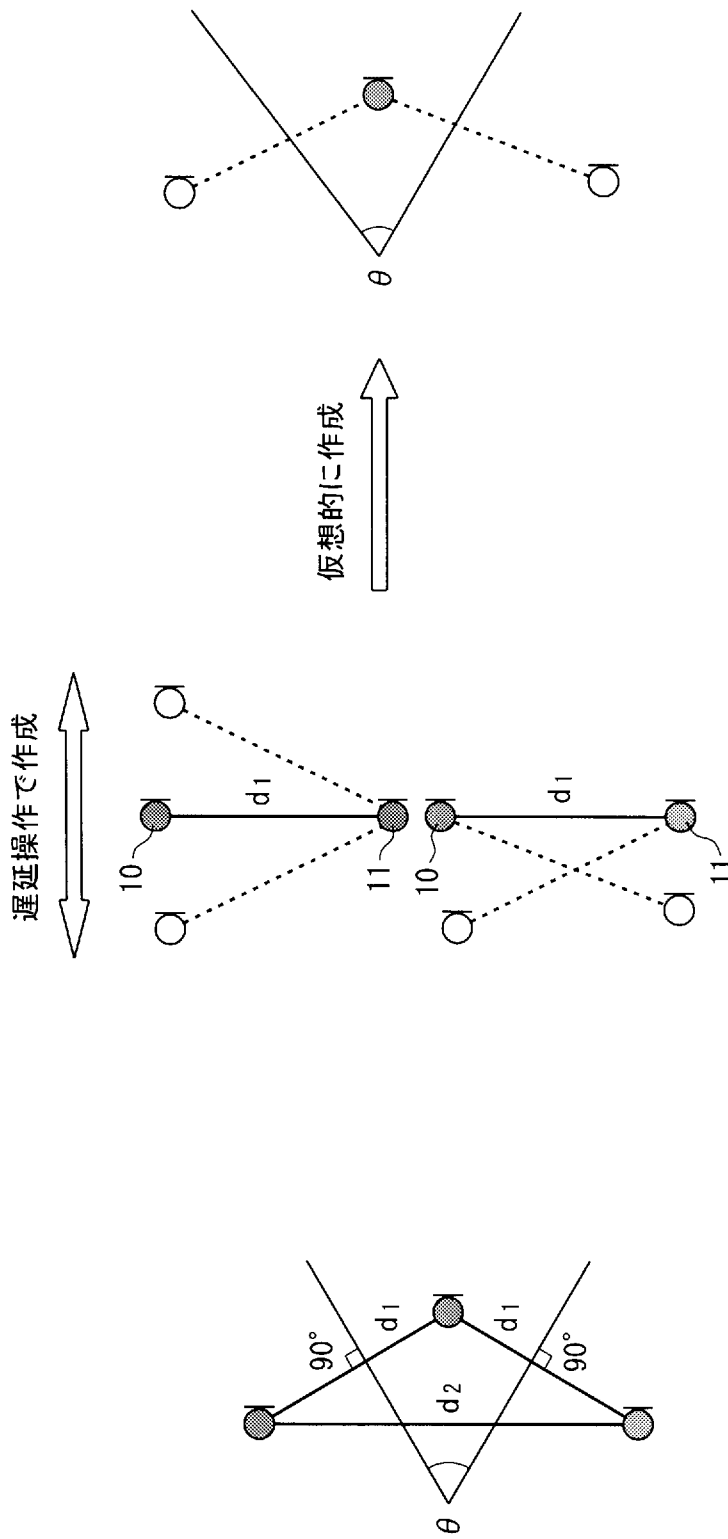
[図33]



[図34]



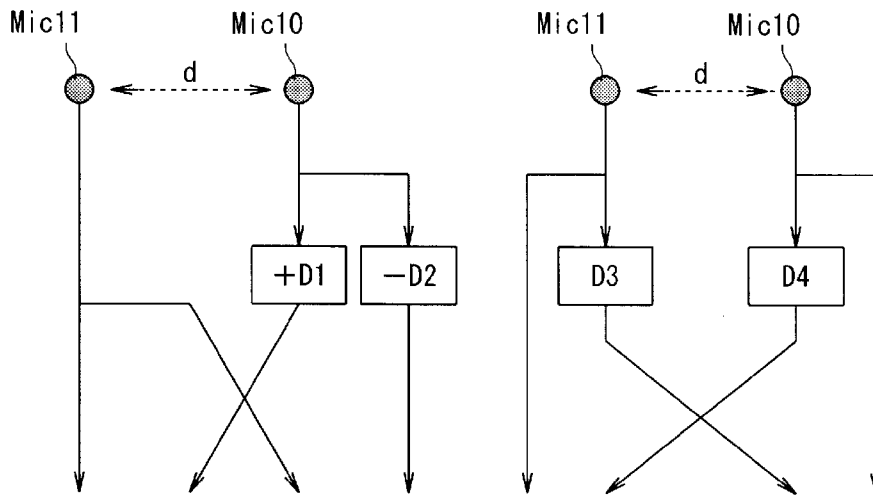
[図35]



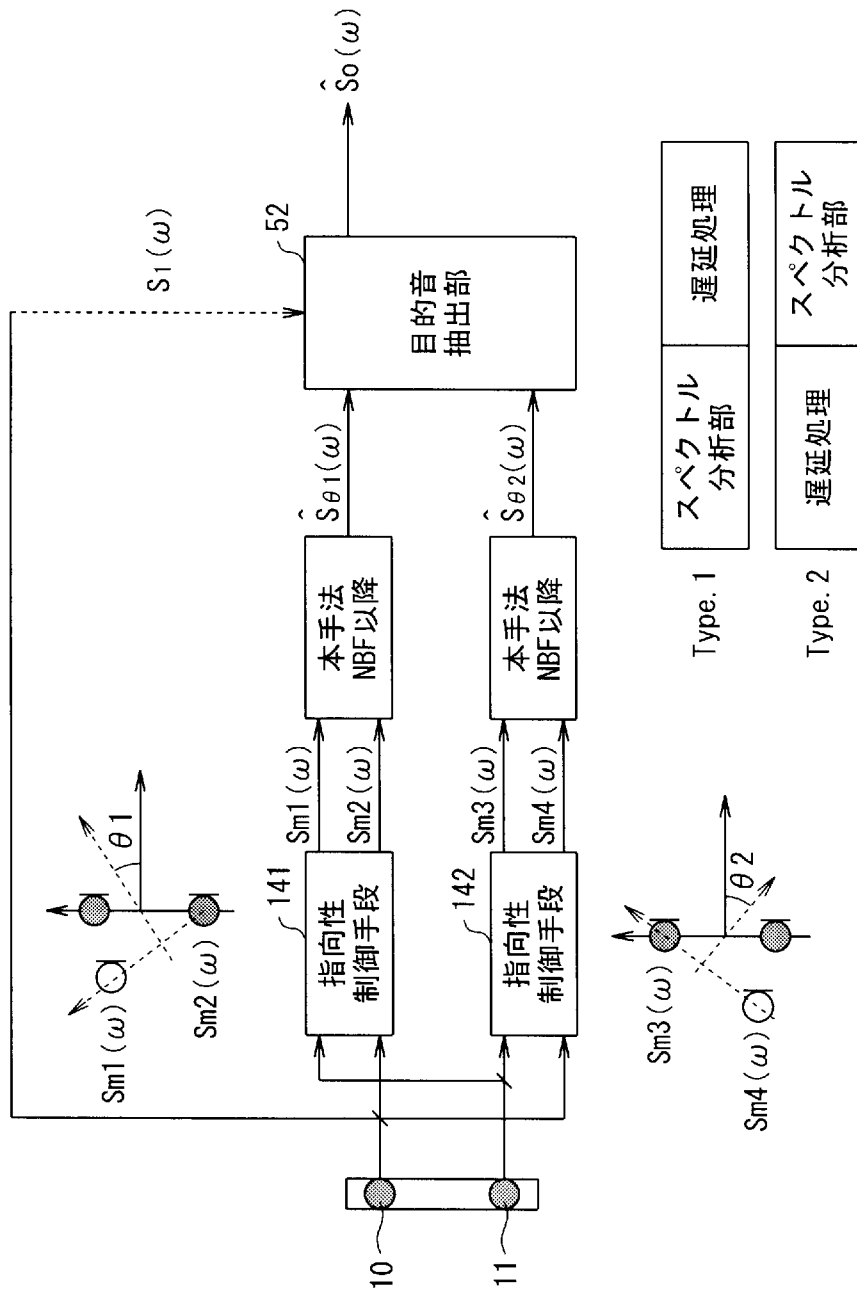
(b) 2マイク構成

(a) 3マイク構成

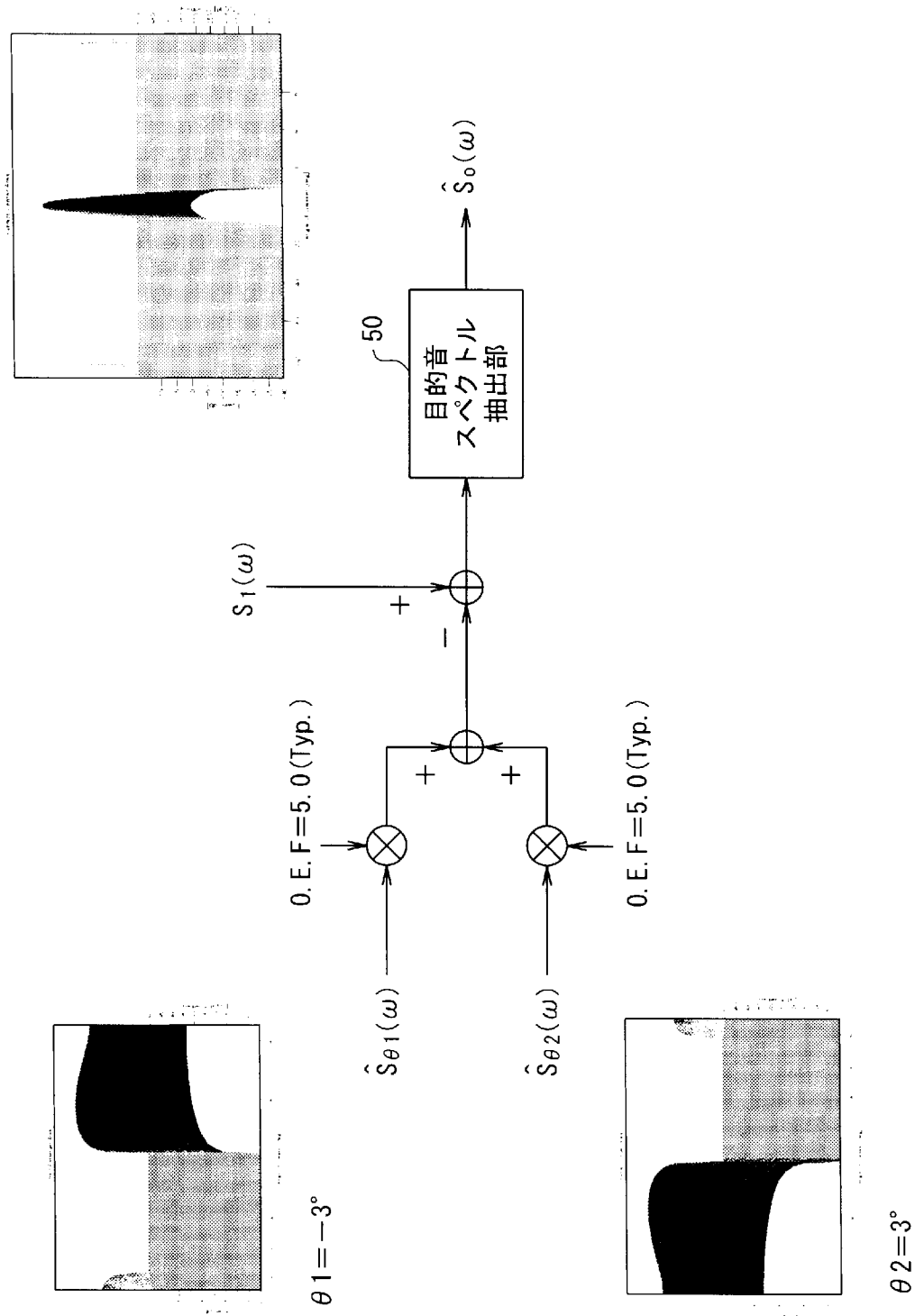
[図36]



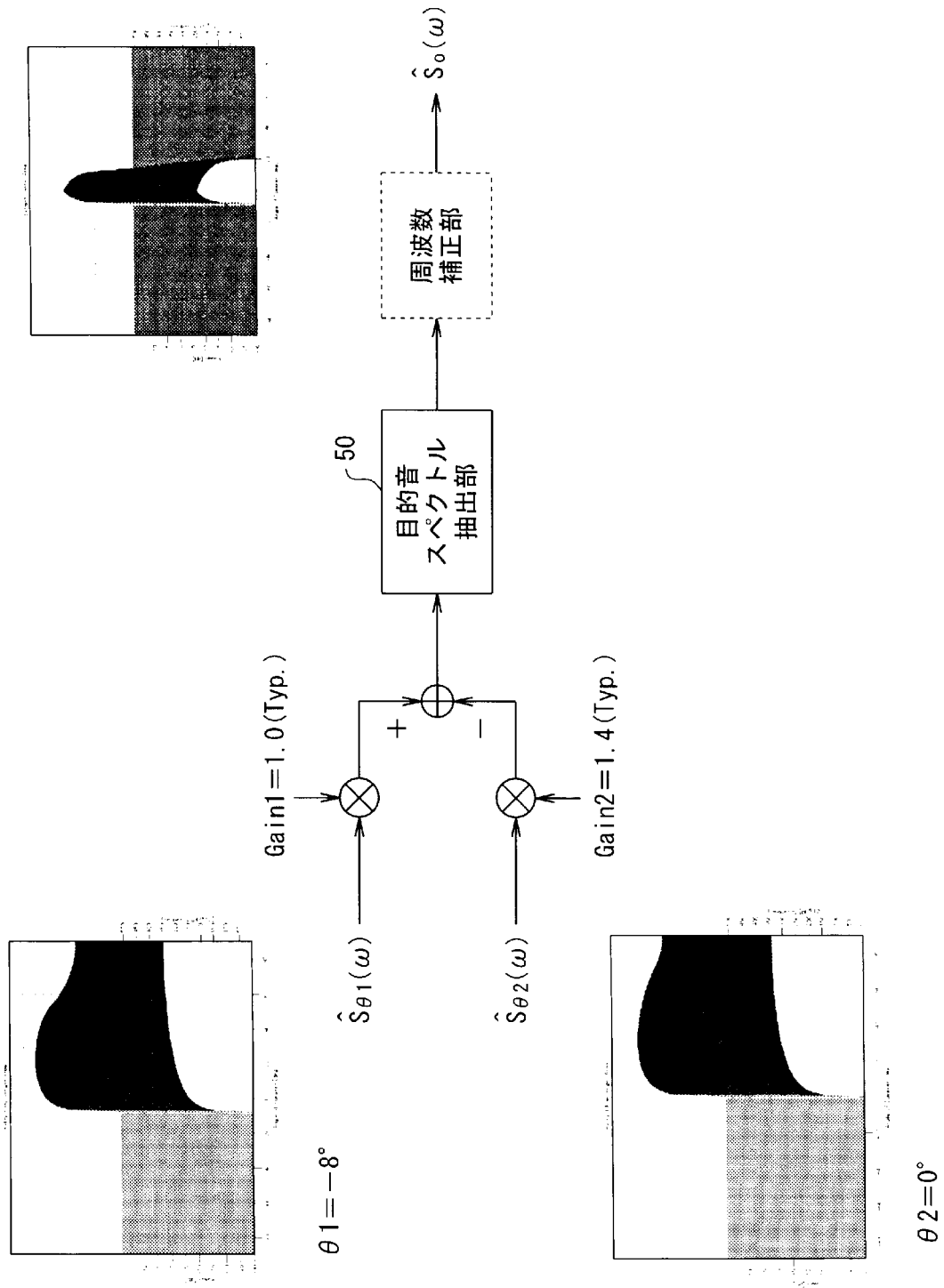
[図37]



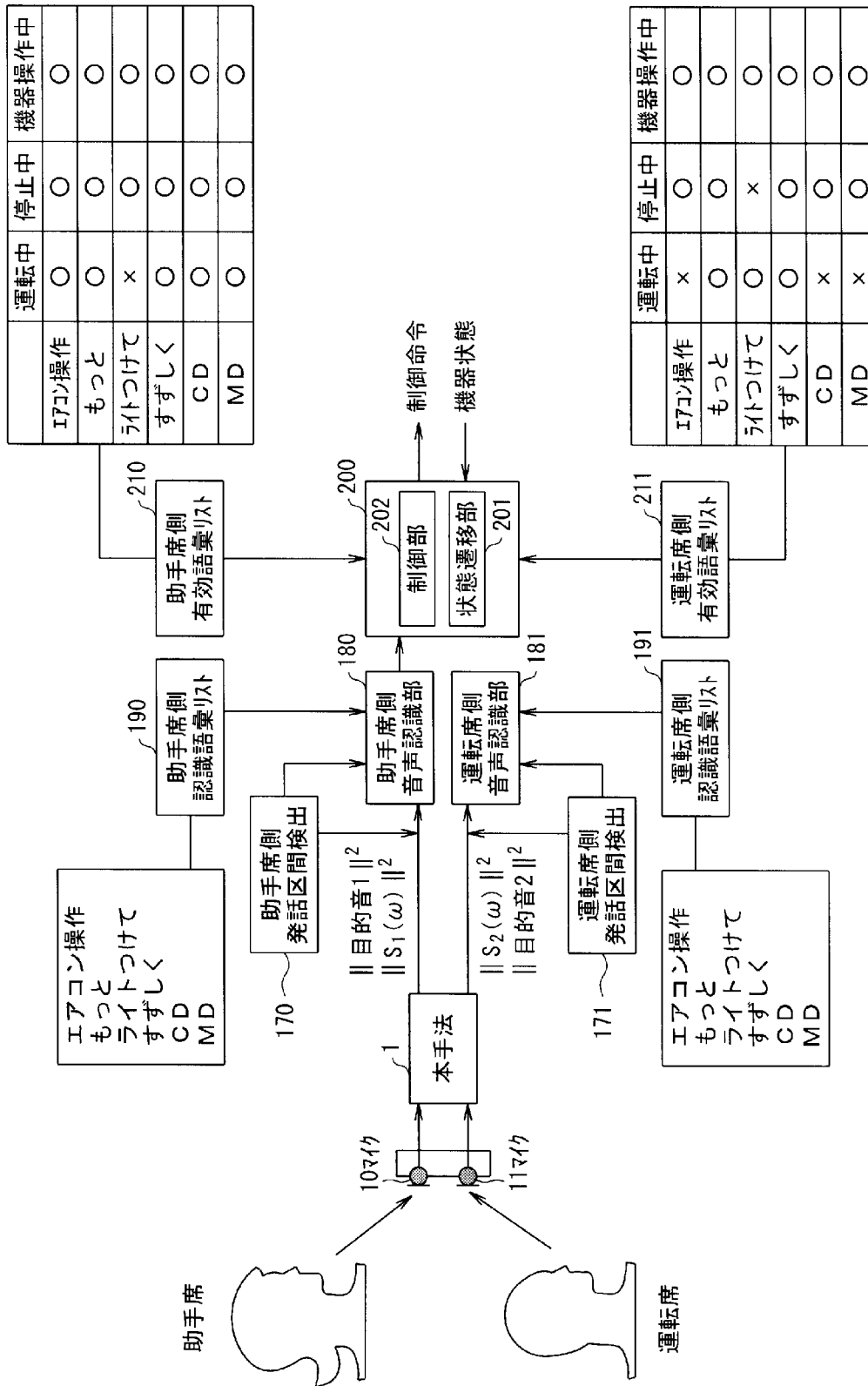
[図38]



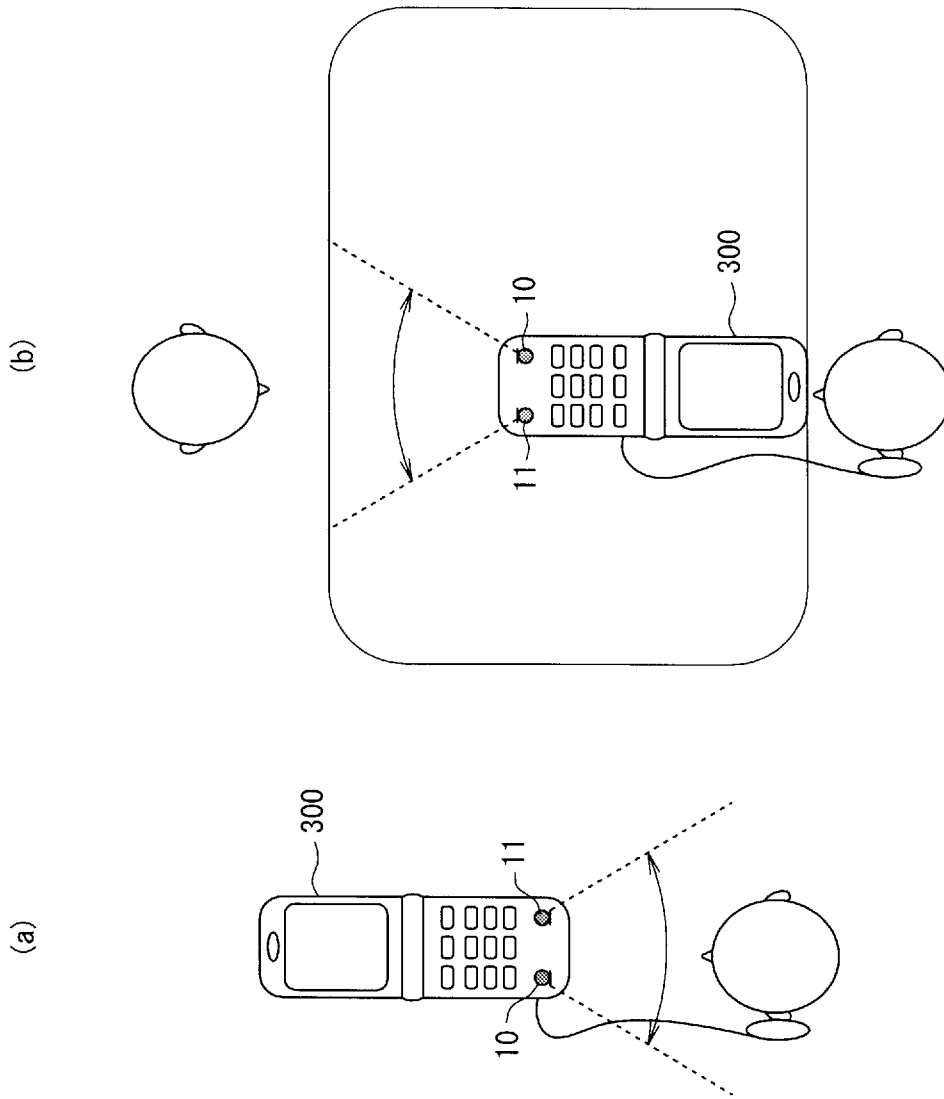
[図39]



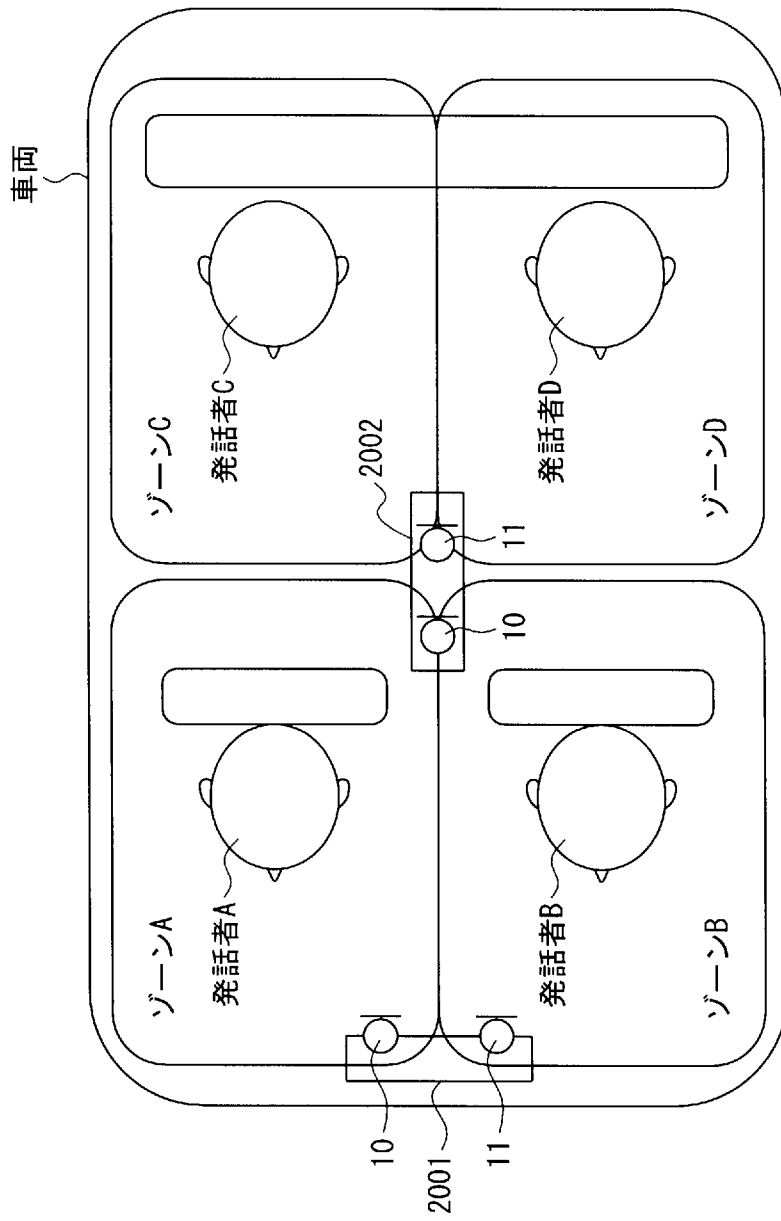
[図40]



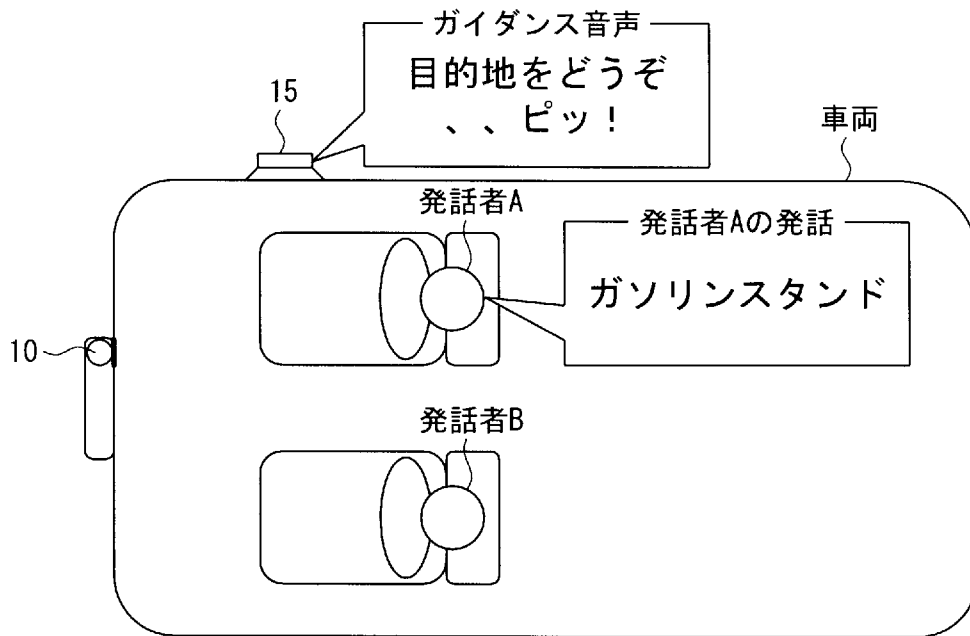
[図41]



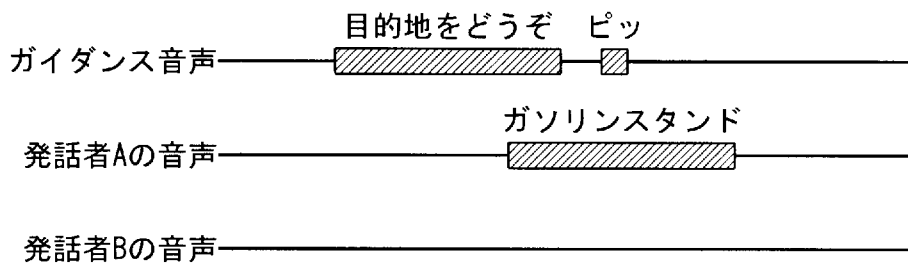
[図42]



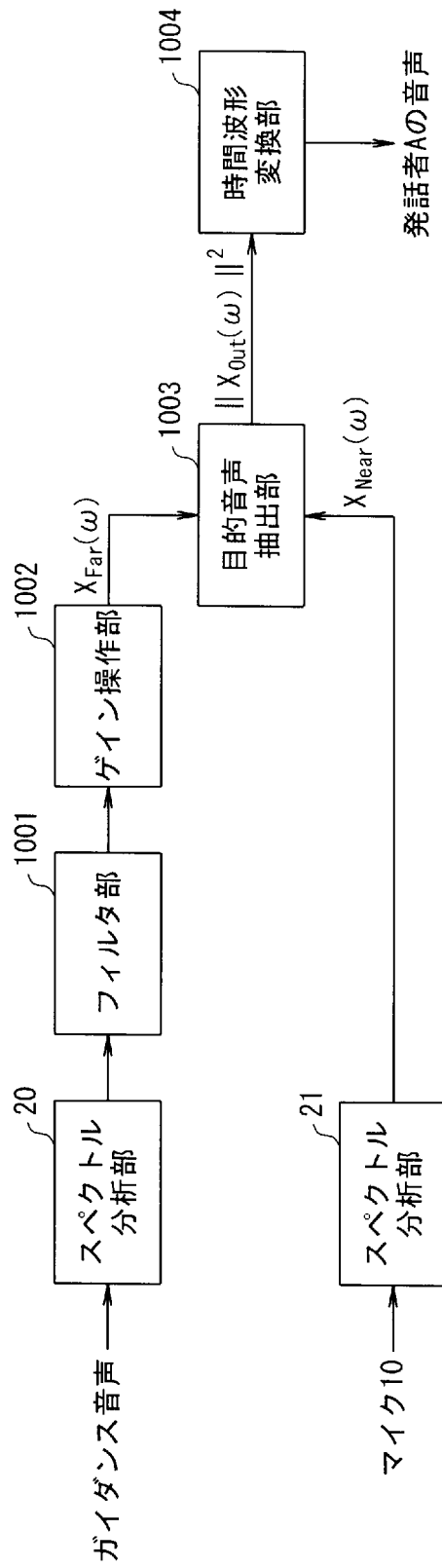
[図43]



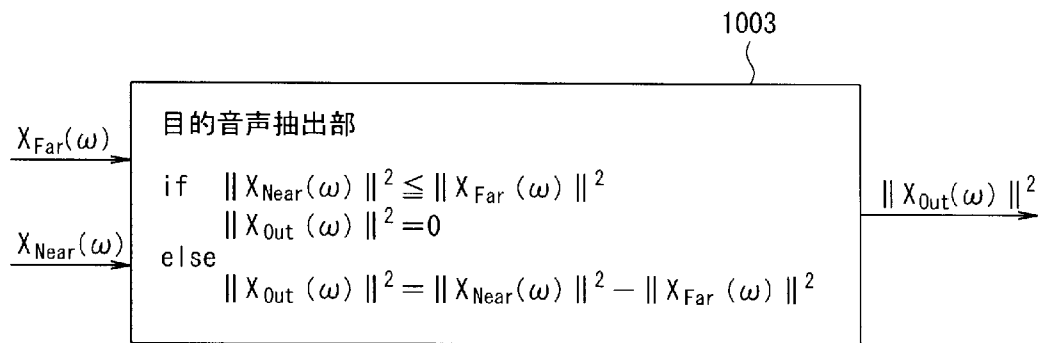
[図44]



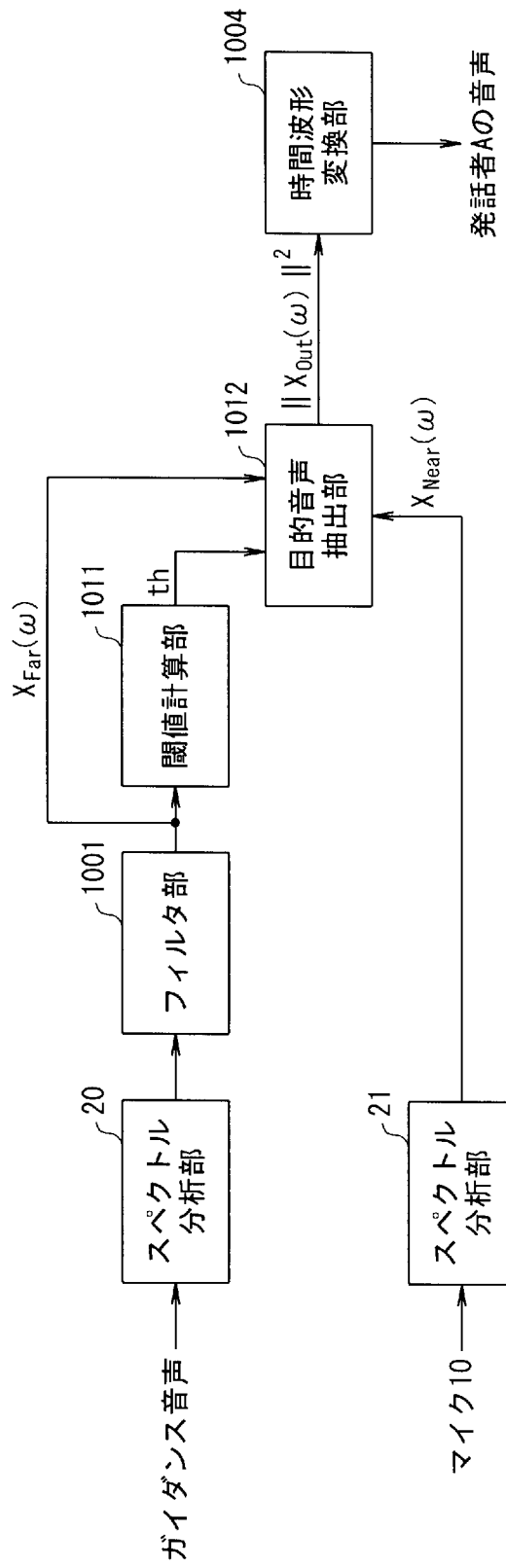
[図45]



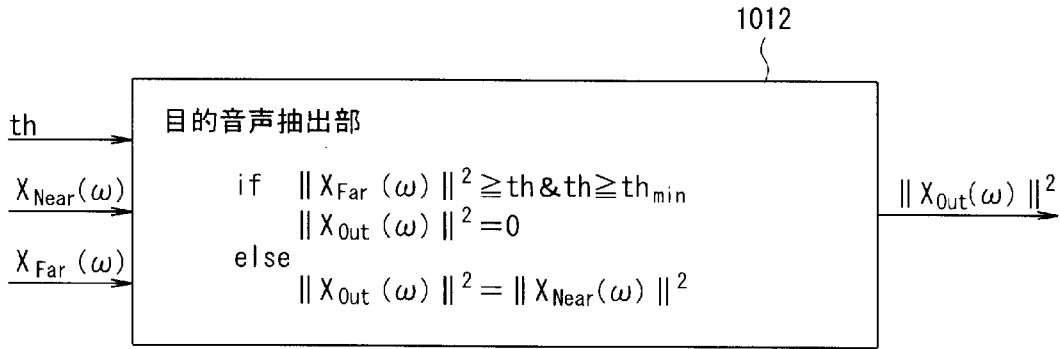
[図46]



[図47]

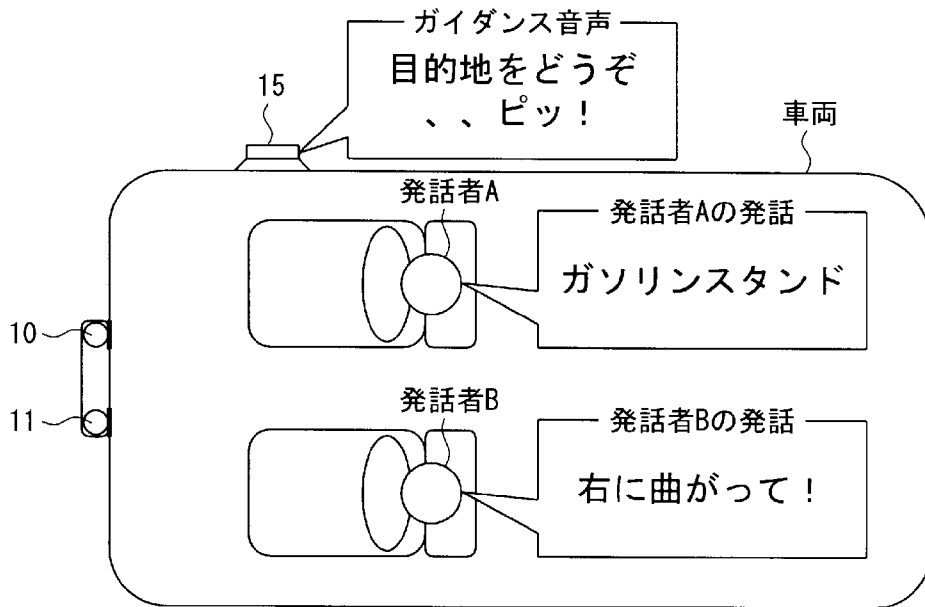


[図48]

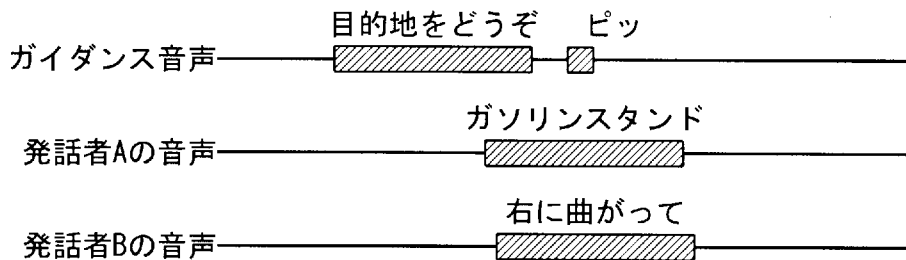


th_{min} : $X_{Far}(\omega)$ が有効な入力であると判断するための閾値

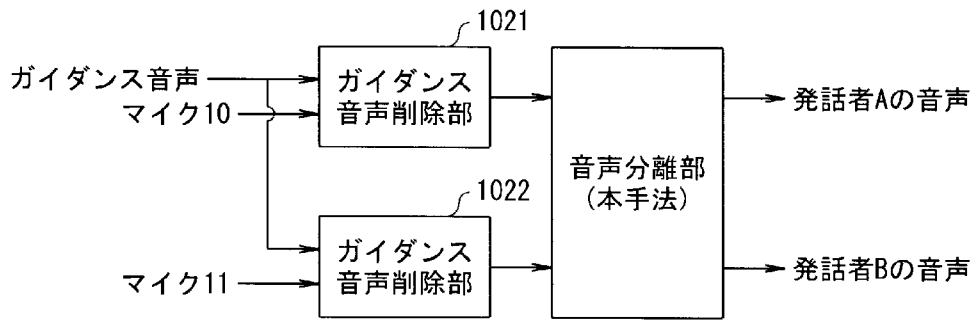
[図49]



[図50]



[図51]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2006/315953

A. CLASSIFICATION OF SUBJECT MATTER

G10L21/02(2006.01)i, G10L15/00(2006.01)i, G10L15/20(2006.01)i, G10L15/22(2006.01)i, H04R1/40(2006.01)i, H04R3/00(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G10L21/02, G10L15/00, G10L15/20, G10L15/22, H04R1/40, H04R3/00

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Jitsuyo Shinan Koho	1922-1996	Jitsuyo Shinan Toroku Koho	1996-2006
Kokai Jitsuyo Shinan Koho	1971-2006	Toroku Jitsuyo Shinan Koho	1994-2006

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

IEEE, JSTPlus (JDream2)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 3484112 B2 (Toshiba Corp.), 17 October, 2003 (17.10.03), Full text; all drawings (Family: none)	1-13
A	JP 2004-289762 A (Toshiba Corp.), 14 October, 2004 (14.10.04), Full text; all drawings (Family: none)	1-13
A	JP 2004-187283 A (Matsushita Electric Industrial Co., Ltd.), 02 July, 2004 (02.07.04), Full text; all drawings & US 2004/0185804 A1	1-13

Further documents are listed in the continuation of Box C.

See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance
 “E” earlier application or patent but published on or after the international filing date
 “L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
 “O” document referring to an oral disclosure, use, exhibition or other means
 “P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
 “X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
 “Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
 “&” document member of the same patent family

Date of the actual completion of the international search
23 August, 2006 (23.08.06)

Date of mailing of the international search report
05 September, 2006 (05.09.06)

Name and mailing address of the ISA/
Japanese Patent Office

Authorized officer

Facsimile No.

Telephone No.

<p>A. 発明の属する分野の分類 (国際特許分類 (IPC))</p> <p>Int.Cl. G10L21/02(2006.01)i, G10L15/00(2006.01)i, G10L15/20(2006.01)i, G10L15/22(2006.01)i, H04R1/40(2006.01)i, H04R3/00(2006.01)i</p>			
<p>B. 調査を行った分野</p> <p>調査を行った最小限資料 (国際特許分類 (IPC))</p> <p>Int.Cl. G10L21/02, G10L15/00, G10L15/20, G10L15/22, H04R1/40, H04R3/00</p>			
<p>最小限資料以外の資料で調査を行った分野に含まれるもの</p> <p>日本国実用新案公報 1922-1996年 日本国公開実用新案公報 1971-2006年 日本国実用新案登録公報 1996-2006年 日本国登録実用新案公報 1994-2006年</p>			
<p>国際調査で使用した電子データベース (データベースの名称、調査に使用した用語)</p> <p>IEEE, JSTPlus(JDream2)</p>			
<p>C. 関連すると認められる文献</p>			
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求の範囲の番号	
A	J P 3 4 8 4 1 1 2 B 2 (株式会社東芝) 2003.10.17, 全文, 全図 (ファミリーなし)	1-13	
A	J P 2 0 0 4 - 2 8 9 7 6 2 A (株式会社東芝) 2004.10.14, 全文, 全図 (ファミリーなし)	1-13	
A	J P 2 0 0 4 - 1 8 7 2 8 3 A (松下電器産業株式会社) 2004.07.02, 全文, 全図 & U S 2 0 0 4 / 0 1 8 5 8 0 4 A 1	1-13	
<input type="checkbox"/> C欄の続きにも文献が列挙されている。		<input type="checkbox"/> パテントファミリーに関する別紙を参照。	
<p>* 引用文献のカテゴリー</p> <p>「A」特に関連のある文献ではなく、一般的技術水準を示すもの</p> <p>「E」国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの</p> <p>「L」優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献 (理由を付す)</p> <p>「O」口頭による開示、使用、展示等に言及する文献</p> <p>「P」国際出願日前で、かつ優先権の主張の基礎となる出願</p>		<p>の日の後に公表された文献</p> <p>「T」国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの</p> <p>「X」特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの</p> <p>「Y」特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの</p> <p>「&」同一パテントファミリー文献</p>	
<p>国際調査を完了した日</p> <p>23.08.2006</p>		<p>国際調査報告の発送日</p> <p>05.09.2006</p>	
<p>国際調査機関の名称及びあて先</p> <p>日本国特許庁 (ISA/J P)</p> <p>郵便番号100-8915</p> <p>東京都千代田区霞が関三丁目4番3号</p>		<p>特許庁審査官 (権限のある職員)</p> <p>榎本 剛</p>	<p>5Z 9379</p>
		<p>電話番号 03-3581-1101 内線 3541</p>	