



3人以上から成る会話グループが存在する場合でも、複数の話者の中から高い精度で会話グループを抽出することができる音声処理装置。この装置(400)は、音響信号から複数の話者の発話音声(410)を個別に検出する自発話検出部(420)および方向別音声検出部(430)と、検出された発話音声に基づいて、複数の話者のうちの2人の組み合わせの全てについて、判定対象時間を区切ったセグメント毎に会話成立度を算出する会話成立度計算部(450)と、組み合わせ毎に、判定対象時間における会話成立度の長時間特徴量を算出する長時間特徴量計算部(460)と、算出された長時間特徴量に基づいて、複数の話者の中から、会話を形成する会話グループを抽出する会話相手判定部(470)とを有する。

明 細 書

発明の名称：音声処理装置および音声処理方法

技術分野

[0001] 本発明は、複数の話者から発話音声を検出する音声処理装置および音声処理方法に関する。

背景技術

[0002] 従来、補聴器や電話会議装置における指向性制御等を目的として、複数の話者の中から、会話を形成するグループ（以下「会話グループ」という）を抽出する技術が存在している（例えば特許文献1参照）。

[0003] 特許文献1記載の技術（以下「従来技術」という）は、2人の話者が会話を行っているとき、それぞれの話者から有音区間が交互に検出される、という現象を前提としている。従来技術は、この前提の下、有音区間／無音区間が交互となっているか否かに基づいて、2人の話者の会話成立度合いを算出する。

[0004] 具体的には、従来技術は、単位時間ごとに、2人の一方が有音であり他方が無音であれば、会話成立度合いを加点し、両方が有音または両方が無音であれば、会話成立度合いを減点する。そして、従来技術は、判定対象区間における加点減点の累積結果が閾値以上となっている場合に、該当する2人の間で会話が成立していると判定する。

[0005] このような従来技術を用いることにより、複数の話者の中から、互いに会話を行っている2人を抽出することができる。

先行技術文献

特許文献

[0006] 特許文献1：特開2004-133403号公報

発明の概要

発明が解決しようとする課題

[0007] しかしながら、従来技術は、3人以上から成る会話グループが存在する場

合、会話グループの抽出の精度が低くなるという課題がある。

[0008] なぜなら、3人以上の会話の場合は、ほぼ全ての単位時間において、1人の発話している話者と、複数人の無言の話者とが存在することになるが、この複数人の無言の話者の間で、会話成立度が低くなるためである。また、3人以上の会話の場合は、聞き役で発話をほとんど行わない話者が存在すると、その無言の話者と他の話者との間で、会話成立度が低くなるためである。

[0009] 本発明の目的は、3人以上から成る会話グループが存在する場合でも、複数の話者の中から高い精度で会話グループを抽出することができる音声処理装置および音声処理方法を提供することである。

課題を解決するための手段

[0010] 本発明の音声処理装置は、音響信号から複数の話者の発話音声を個別に検出する音声検出部と、検出された前記発話音声に基づいて、前記複数の話者のうちの2人の組み合わせの全てについて、判定対象時間を区切ったセグメントごとに会話成立度を算出する会話成立度計算部と、前記組み合わせごとに、前記判定対象時間における前記会話成立度の長時間特徴量を算出する長時間特徴量計算部と、算出された前記長時間特徴量に基づいて、前記複数の話者の中から、会話を形成する会話グループを抽出する会話相手判定部とを有する。

[0011] 本発明の音声処理方法は、音響信号から複数の話者の発話音声を個別に検出するステップと、検出された前記発話音声に基づいて、前記複数の話者のうちの2人の組み合わせの全てについて、判定対象時間を区切ったセグメントごとに会話成立度を算出するステップと、前記組み合わせごとに、前記判定対象時間における前記会話成立度の長時間特徴量を算出するステップと、算出された前記長時間特徴量に基づいて、前記複数の話者の中から、会話を形成する会話グループを抽出するステップとを有する。

発明の効果

[0012] 本発明によれば、3人以上から成る会話グループが存在する場合でも、複数の話者の中から高い精度で会話グループを抽出することができる。

図面の簡単な説明

- [0013] [図1]本発明の一実施の形態に係る音声処理装置を含む補聴器の構成を示す図
[図2]本実施の形態に係る補聴器が使用される環境の例を示す図
[図3]本実施の形態に係る音声処理装置の構成を示すブロック図
[図4]本実施の形態における会話成立度と会話グループとの関係を説明するための第1の図
[図5]本実施の形態における会話成立度と会話グループとの関係を説明するための第2の図
[図6]本実施の形態に係る音声処理装置の動作を示すフローチャート
[図7]本実施の形態におけるマイクロホンアレイの指向性のパターンの例を示す図
[図8]本実施の形態における会話相手判定処理を示すフローチャート
[図9]本発明の実験のために簡略化した会話相手判定処理を示すフローチャート
[図10]本発明の実験結果を示すプロット図

発明を実施するための形態

- [0014] 以下、本発明の一実施の形態について、図面を参照して詳細に説明する。
本実施の形態は、本発明を、補聴器の指向性制御のための会話相手特定手段に適用した例である。
- [0015] 図1は、本発明に係る音声処理装置を含む補聴器の構成を示す図である。
- [0016] 図1に示すように、補聴器100は、両耳型の補聴器であり、ユーザの左右の外耳に掛けるための補聴器筐体110L、110Rを有する。
- [0017] 左右の筐体110L、110Rのそれぞれの上部には、周囲の音を拾う2つずつのマイクロホンが前後に並べて設置されている。左右2個ずつあわせて4個から成るこれらのマイクロホンは、マイクロホンアレイ120を構成している。4個のマイクロフォンは、補聴器100を装着したユーザに対して、それぞれ所定の位置に配置される。
- [0018] また、左右の筐体110L、110Rのそれぞれには、補聴した音を出力

するスピーカ130L、130Rが設置されている。左右のスピーカ130L、130Rには、それぞれ内耳に嵌めるためのイヤーチップ140L、140Rが、チューブを介して接続されている。

[0019] また、補聴器100は、補聴器マイクロホンアレイ120およびスピーカ130L、130Rと有線で接続された、リモートコントロール（以下「リモコン」という）装置150を有する。

[0020] リモコン装置150には、CPU160およびメモリ170が内蔵されている。CPU160は、マイクロホンアレイ120にて收音された音声を入力し、メモリ170に予め格納された制御プログラムを実行する。これにより、CPU160は、マイクロホンアレイ120から入力された4チャンネルの音響信号に対し、指向性制御処理および補聴処理を行う。

[0021] 指向性制御処理は、会話相手の発話音声をユーザが聞き易くなるように、マイクロホンアレイ120からの4チャンネルの音響信号の指向方向を制御する処理である。補聴処理は、会話相手の発話音声をユーザが聞き易くなるように、ユーザの聴力が低下した周波数帯域のゲインを増幅し、スピーカ130L、130Rから出力させる処理である。

[0022] このような補聴器100を装着することにより、ユーザは、会話相手の発話音声が聞き易くなった音声を、イヤーチップ140L、140Rから聞くことができる。

[0023] 図2は、補聴器100が使用される環境の例を示す図である。

[0024] 図2Aおよび図2Bに示すように、両耳型の補聴器100を装着したユーザ200は、例えば、レストラン等の賑やかな環境において、友人等の話者300と会話をする。図2Aは、ユーザ200が前方に位置する話者300Fと2人だけで会話しているケースである。また、図2Bは、ユーザ200が前方に位置する話者300Fおよび左方に位置する話者300Lと3人で会話しているケースである。

[0025] 図2Aのケースでは、補聴器100は、左方や右方に位置する他人の発話音声をできるだけ除外し、前方の話者300Fの発話音声を聞き取りやすくす

るために、前方の狭い範囲に指向性を向けるべきである。

[0026] 一方、図2Bのケースでは、補聴器100は、前方の話者300Fだけでなく左方の話者300Lの発話音声を聞き取り易くするために、前方と左方とを含む広い範囲に指向性を向けるべきである。

[0027] このような指向性制御が行われることにより、ユーザ200は、周りが騒がしい環境であっても、会話相手の声を明瞭に聞くことが可能となる。会話相手の方向に応じて指向性を制御するためには、その方向を指定する必要がある。この指定は、例えば、ユーザ200が手動で行うことが考えられる。

[0028] ところが、その操作は煩雑であり、特に高齢者や子供の場合、誤った操作が行われ、誤った方向に指向性が向けられて、逆に聞き取りにくくなるおそれがある。

[0029] このため、補聴器100のCPU160は、周囲の話者の中から自動でユーザ200の会話相手を抽出する会話相手抽出処理を行う。そして、補聴器100のCPU160は、抽出した会話相手の方向に、マイクロホンアレイ120による音声入力の指向性（以下「マイクロホンアレイ120の指向性」という）を向ける。

[0030] 会話相手抽出処理は、会話相手が2人以上である場合でも会話相手の抽出を高精度に行う処理となっている。ここでは、会話相手抽出処理を実現する機能を、音声処理装置というものとする。

[0031] 以下、音声処理装置の構成と、会話相手抽出処理の詳細について説明する。

[0032] 図3は、音声処理装置の構成を示すブロック図である。

[0033] 図3において、音声処理装置400は、AD変換部410、自発話検出部420、方向別音声検出部430、発話量和計算部440、会話成立度計算部450、長時間特徴量計算部460、会話相手判定部470、および出力音制御部480を有する。自発話検出部420および方向別音声検出部430は、合わせて、音声検出部435とする。

[0034] AD変換部410は、マイクロホンアレイ120において、マイクロホン

ごとに収録された4チャンネルの音響信号であるアナログ信号を、それぞれデジタル信号に変換する。そして、A/D変換部410は、変換後の4チャンネルのデジタル音響信号を、自発話検出部420、方向別音声検出部430、および出力音制御部480へ出力する。

[0035] 自発話検出部420は、A/D変換された4チャンネルのデジタル音響信号から低域の振動成分を強調し（つまり低域の振動成分を取り出し）、自発話パワー成分を求める。自発話検出部420は、A/D変換された4チャンネルのデジタル音響信号を用いて、細かい時間間隔ごとに発話音声の有無を判定する。そして、自発話検出部420は、フレームごとの自発話の有無を示す音声／非音声情報を、発話量和計算部440および会話成立度計算部450へ出力する。

[0036] ここで、自発話とは、補聴器100を装着したユーザ200の発話をいう。また、以下、発話音声の有無が判定される時間間隔を「フレーム」という。1フレームは、例えば10msec（ミリ秒）である。なお、自発話の有無の判定は、前後2チャンネルのデジタル音響信号を用いてもよい。

[0037] 本実施の形態では、例えば、ユーザ200から見て前方、左方、および右方の各位置を発話者が位置する可能性のある位置（以下「音源」という）として説明する。

[0038] 方向別音声検出部430は、マイクロホンアレイ120からのA/D変換後の4つのデジタル音響信号から、前方、左方、右方の各方向の音声を抽出する。より具体的には、方向別音声検出部430は、4チャンネルのデジタル音響信号に対して公知の指向性制御技術を用いる。これにより、方向別音声検出部430は、ユーザ200の前方、左方、右方の方向別に指向性を形成し、これにより得られる音声を、前方、左方、右方の各方向の音声とする。そして、方向別音声検出部430は、抽出された各方向の音声のパワー情報から、細かい時間間隔で発話音声の有無を判定し、その判定結果に基づいて、各方向の他発話の有無をフレームごとに判定する。そして、方向別音声検出部430は、フレームごとおよび方向ごとの他発話の有無を示す音声／

非音声情報を、発話量和計算部440および会話成立度計算部450へ出力する。

[0039] ここで、他発話とは、補聴器100を装着したユーザ200以外の者の発話（自発話以外の発話）をいう。

[0040] なお、自発話検出部420および方向別音声検出部430は、同一の時間間隔で発話音声の有無の判定を行うものとする。

[0041] 発話量和計算部440は、自発話検出部420から入力される自発話の音声／非音声情報と、方向別音声検出部430から入力される音源ごとの他発話の音声／非音声情報とに基づいて、セグメントごとに発話量和を算出する。具体的には、発話量和計算部440は、4つの音源のうちの2つの組み合わせ（以下「ペア」という）の全てについて、その2音源のセグメント内の発話量の和を、セグメントごとの発話量和として検出する。そして、発話量和計算部440は、算出したペアごとおよびセグメントごとの発話量和を、会話成立度計算部450へ出力する。

[0042] ここで、発話量とは、ユーザが発話音声を発した時間長さの合計を表す。また、セグメントは、2人の話者の間で局所的に会話が成立するかどうかを表す会話成立度を求めるための固定長の時間窓である。したがって、その長さは2人の話者の間で局所的に会話が成立するとみなされる程度の長さにする必要がある。セグメントが長ければ長いほど、会話成立度の正解の精度は高くなるが、発言権を有するペアの交代への追従の精度は低くなる。また、セグメントが短ければ短いほど、会話成立度の正解の精度は低くなるが、発言権を有するペアの交代への追従の精度は高くなる。本実施の形態では、1セグメントは、例えば40秒とする。これは、予備実験により得られた、会話成立度は1分程度で飽和するとの知見と、会話の流れへの追従とを考慮して、決定される。

[0043] 会話成立度計算部450は、発話量和計算部440から入力された発話量和と、自発話検出部420および方向別音声検出部430から入力された音声／非音声情報とに基づいて、ペアごとおよびセグメントごとに、会話成立

度を算出する。そして、会話成立度計算部450は、入力された発話量と算出した会話成立度とを、長時間特徴量計算部460へ出力する。

[0044] ここで、会話成立度とは、会話成立度と同様の指標値であり、一方が有音で他方が無音となっている区間の割合が高いほど高くなり、両方が有音または無音となっている区間の割合が高いほど低くなる値である。但し、本実施の形態では、発話量とが閾値未満となっているセグメントについては、いずれも聞き役となっているものとして、その会話成立度を後述の長時間特徴量の算出対象に含めないようにする点が、従来技術とは異なる。

[0045] 長時間特徴量計算部460は、入力された発話量および会話成立度に基づいて、ペアごとに、長時間特徴量を算出する。そして、長時間特徴量計算部460は、算出した長時間特徴量を、会話相手判定部470へ出力する。

[0046] 長時間特徴量とは、判定対象時間における会話成立度の平均値である。なお、長時間特徴量は、会話成立度の平均値に限定されるものではなく、会話成立度の中央値あるいは最頻値など、他の統計量としてもよい。また、長時間特徴量は、時間的に最近の会話成立度ほど重みを大きくして求めた重み付け平均値や、会話成立度の時系列にある程度長時間の時間窓をかけて求めた移動平均値などとしてもよい。

[0047] 会話相手判定部470は、入力された長時間特徴量に基づいて、複数の音源に位置する複数の話者（ユーザ200を含む）の中から、会話グループを抽出する。具体的には、会話相手判定部470は、長時間特徴量が似通っており、かつ、いずれも閾値以上となっている1つまたは複数のペアが存在するとき、その1つまたは複数のペアを構成する複数の話者を、1つの会話グループと判定する。本実施の形態では、会話相手判定部470は、ユーザ200の会話相手の位置する方向を抽出し、抽出した方向を示す情報を、指向性を向けるべき指向方向情報として、出力音制御部480へ出力する。

[0048] 出力音制御部480は、入力された音響信号に対し、上述の補聴処理を行い、処理後の音響信号を、スピーカ130L、130Rへ出力する。また、出力音制御部480は、入力された指向方向情報が示す方向に指向性が向く

ように、マイクロホンアレイ 120 に対する指向性制御処理を行う。

[0049] このような音声処理装置 400 は、ペアごとの発話量および会話成立度に基づいて、複数の話者の中から会話グループを抽出することができる。

[0050] ここで、発話量、会話成立度、および長時間特徴量について説明する。

[0051] 図 4 および図 5 は、会話成立度と会話グループとの関係を説明するための図である。図 4 および図 5 において、横軸は判定対象時間におけるセグメント（つまり時間）を示し、縦軸は各ペアを示す。また、灰色部分は、発話量とが閾値未満のセグメントを示す。白色部分は、発話量とが閾値以上であり会話成立度が閾値未満であるセグメントを示す。そして、黒色部分は、発話量とが閾値以上であり会話成立度が閾値以上であるセグメントを示す。

[0052] まず、ユーザと左方に位置する話者とが 2 人で会話しており、前方に位置する話者と右方に位置する話者とが 2 人で会話している場合を想定する。この場合、図 4 に示すように、ユーザ 200 と左方の話者とのペア（上から 2 行目）と、前方の話者と右方の話者とのペア（上から 5 行目）では、発話量とが閾値以上かつ会話成立度が閾値以上であるセグメントは多くなる。ところが、他のペアでは、発話量とが閾値以上かつ会話成立度が閾値以上であるセグメントは少なくなる。

[0053] 次に、ユーザ 200 と左方、前方、右方に位置する 3 人の話者とで会話している場合を想定する。3 人以上の会話では、1 人が発話した後に別の 1 人が発話するとき、残りの話者は聞き役となる。すなわち、話者は、短い時間においては、発話権を有する 2 人と、その他の聞き役とに分けることができる。そして、長い時間においては、発話権を有する 2 人の組み合わせが交代していきながら、会話が進行することになる。

[0054] すなわち、3 人以上の会話グループでは、発話権を有する 2 人の間で局所的に会話成立度が高くなるといえる。この結果、図 5 に示すように、発話量とが閾値以下のセグメントと、発話量とが閾値以上かつ会話成立度が閾値以上であるセグメントは、全てのペアにおいてそれぞれまんべんなく存在することになる。

- [0055] そこで、音声処理装置400は、発話量和が閾値以上のセグメントのみから長時間特徴量を算出し、長時間特徴量がまんべんなく高くなっている話者のグループを、会話グループと判定する。
- [0056] したがって、音声処理装置400は、図4の場合、左方に位置する話者のみをユーザ200の会話相手と判定し、マイクロホンアレイ120の指向性を左方に狭める。また、音声処理装置400は、図5の場合、左方、前方、右方に位置する3人の話者をユーザ200の会話相手と判定し、マイクロホンアレイ120の指向性を左方から右方までの広い範囲に広げる。
- [0057] 図6は、音声処理装置400の動作を示すフローチャートである。
- [0058] まず、ステップS1100において、AD変換部410は、マイクロホンアレイ120から入力された1フレーム分の4チャンネルの音響信号をそれぞれA/D変換する。
- [0059] そして、ステップS1200において、自発話検出部420は、4チャンネルのデジタル音響信号を用いて、現在のフレームについて、自発話音声の有無を判定する。この判定は、デジタル音響信号の低域成分を強調することにより得られる自発話パワー成分に基づいて行われる。つまり、自発話検出部420は、自発話の有無を示す音声/非音声情報を出力する。
- [0060] なお、音声処理装置400は、処理開始時に、会話が行われているかどうかを判定することが望ましい。そして、音声処理装置400は、会話が行われている場合、ユーザ200の後方からの音声を抑制するように、マイクロホンアレイ120の指向性を制御することが望ましい。会話が行われているかどうかの判定は、例えば、自発話パワー成分に基づいて行うことができる。また、音声処理装置400は、後方からの音声が発話音声であるか判断し、発話音声が到来する方向のみを抑制の対象としてもよい。また、音声処理装置400は、周囲が静かな環境では、これらの制御を行わないようにしてもよい。
- [0061] そして、ステップS1300において、方向別音声検出部430は、A/D変換された4チャンネルのデジタル音響信号を用いて、現在のフレームに

ついて、前方、左方、右方の各方向の他発話音声の有無をそれぞれ判定する。この判定は、方向別に指向性を形成し、方向ごとの音声帯域（例えば2000Hz～4000Hz）のパワー情報に基づいて行われる。つまり、方向別音声検出部430は、方向ごとである音源ごとの他発話の有無を示す音声／非音声情報を出力する。

[0062] なお、方向別音声検出部430は、自発話による影響を軽減するために、方向別パワーの対数値から自発話パワーの対数値を差し引いた値に基づいて、他発話音声の有無を判定してもよい。また、方向別音声検出部430は、左方および右方からの他発話音声について、自発話音声や前方からの他発話音声との分離度を上げるために、左右のパワー差を利用してもよい。また、方向別音声検出部430は、パワーに対して時間方向に平滑化を行ってもよい。また、方向別音声検出部430は、短時間の音声区間を非音声区間として扱ったり、音声時間が長時間継続している間に短時間の非音声区間が含まれている場合、その非音声区間を音声区間として扱ってもよい。これらの後処理を行うことにより、最終的なフレームごとの有音無音の検出精度を向上させることができる。

[0063] そして、ステップS1400において、発話量計算部440は、所定の条件が満たされているか否かを判断する。この所定の条件は、音響信号の入力開始から1セグメント（40秒）分の時間が経過し、かつ、後述の会話相手判定を前回行ってから1シフト間隔（例えば10秒）分の時間経過したことである。発話量計算部440は、1セグメント分の処理がまだ終了していない場合（S1400：NO）、ステップS1100へ戻る。この結果、次の1フレーム分に対する処理が行われる。また、発話量計算部440は、最初の1セグメント分の処理が終了した場合（S1400：YES）、ステップS1500へ進む。

[0064] すなわち、音声処理装置400は、1セグメント（40秒）分の音響信号が用意されると、その後、1シフト間隔（10秒）ごとに1セグメントの局所的な時間窓をシフトさせながら、以降のステップS1500～S2400の

処理を繰り返す。なお、シフト間隔は、時間長さではなく、フレーム数またはセグメント数で定義されてもよい。

[0065] なお、音声処理装置400は、計算処理のための変数として、フレームカウンター t と、セグメントカウンター p と、音源のペアごとの発話量 and の多いセグメントの数を表す多発話セグメントカウンター $g_{i,j}$ とを用いる。

[0066] 音声処理装置400は、判定対象時間の開始時に $t=0$ 、 $p=0$ 、 $g_{i,j}=0$ とする。そして、音声処理装置400は、処理がステップS1100へ進むごとに、フレームカウンターを1進め、処理がステップS1400からステップS1500へと進むごとに、セグメントカウンター p を1進める。すなわち、フレームカウンター t は、処理開始からのフレームの数を示し、セグメントカウンター p は、処理開始からのセグメントの数を示す。また、音声処理装置400は、処理が後述のステップS1800へ進むごとに、該当するペアの多発話セグメントカウンター $g_{i,j}$ を1進める。すなわち、多発話セグメントカウンター $g_{i,j}$ は、ペアごとの、後述の発話量 $H_{i,j}(p)$ が所定の閾値 θ 以上となっているセグメントの数を示す。

[0067] また、以下の説明において、現在のセグメントは、「Seg(p)」と表す。また、ユーザ200自身を含む4つの音源を示す記号は、「S」を用い、音源を識別する記号は、「 i 」、「 j 」を用いる。

[0068] ステップS1500において、発話量 and 計算部440は、複数の音源の中からペア $S_{i,j}$ を1つ選択する。以降のステップS1600~S1900の処理は、ユーザ200自身を含む4つの音源の組み合わせの全てに対して行われることになる。4つの音源は、自発話の音源、他発話のうち前方の音源、他発話のうち左方の音源、他発話のうち右方の音源とする。そして、自発話の音源は S_0 とし、前方の音源は S_1 とし、左方の音源は S_2 とし、右方の音源は S_3 とする。この場合は、 $S_{0,1}$ 、 $S_{0,2}$ 、 $S_{0,3}$ 、 $S_{1,2}$ 、 $S_{1,3}$ 、 $S_{2,3}$ の6通りの組み合わせについて、処理が行われることになる。

[0069] そして、ステップS1600において、発話量 and 計算部440は、音源 $S_{i,j}$ のペア(i,j)に対して、過去1セグメント分の音源別の音声/非音声情

報を用いて、現在のセグメント $Seg(p)$ の発話量 $H_{i,j}(p)$ を算出する。発話量 $H_{i,j}(p)$ は、音源 S_i の発話音声があると判定されたフレームの数と、音源 S_j の発話音声があると判定されたフレームの数との和である。

[0070] そして、ステップ S1700 において、会話成立度計算部 450 は、算出された発話量 $H_{i,j}(p)$ が所定の閾値 θ 以上であるか否か判断する。会話成立度計算部 450 は、発話量 $H_{i,j}(p)$ が所定の閾値 θ 以上である場合 (S1700: YES)、ステップ S1800 へ進む。また、会話成立度計算部 450 は、発話量 $H_{i,j}(p)$ が所定の閾値 θ 未満である場合 (S1700: NO)、ステップ S1900 へ進む。

[0071] ステップ S1800 において、会話成立度計算部 450 は、ペア $S_{i,j}$ の両方ともが発言権を有していたものとして、その音声/非音声情報から、現在のセグメント $Seg(p)$ の会話成立度 $C_{i,j}(p)$ を算出する。そして、会話成立度計算部 450 は、ステップ S2000 へ進む。

[0072] 会話成立度 $C_{i,j}(p)$ は、例えば以下のようにして算出される。過去 40 秒分のフレームから成る現在のセグメント $Seg(p)$ に対応するフレームは、1 フレーム = 10 msec のとき、直前の 4000 フレームとなる。このため、会話成立度計算部 450 は、セグメント内のフレームを k ($k = 1, 2, 3, \dots, 4000$) とすると、例えば以下の式 (1) を用いて、会話成立度 $C_{i,j}(p)$ を算出する。

[数1]

$$C_{i,j}(p) = \frac{\sum_{k=1}^{4000} V_{i,j}(k)}{4000} \quad \dots\dots (1)$$

但し、 S_i が発話音声有り、かつ、 S_j が発話音声有りのとき、

$$V_{i,j}(k) = -1$$

S_i が発話音声有り、かつ、 S_j が発話音声無し のとき、

$$V_{i,j}(k) = 1$$

S_i が発話音声無し、かつ、 S_j が発話音声有りのとき、

$$V_{i,j}(k) = 1$$

S_i が発話音声無し、かつ、 S_j が発話音声無しのとき、

$$V_{i,j}(k) = -1$$

[0073] なお、会話成立度計算部450は、ペア(i, j)ごとに異なる重み付けを、加点減点の値である $V_{i,j}(k)$ に対して行ってもよい。この場合、会話成立度計算部450は、例えば、ユーザ200と前方の話者とのペアに対してより高い重み付けを行う。

[0074] また、ステップS1900において、会話成立度計算部450は、ペア(i, j)の少なくとも一方が発言権を有していないものとして、現在のセグメント $Seg(p)$ の会話成立度 $C_{i,j}(p)$ を0に決定する。そして、会話成立度計算部450は、ステップS2000へ進む。

[0075] すなわち、会話成立度計算部450は、現在のセグメント $Seg(p)$ の会話成立度が実質的に評価に用いられないようにする。これは、少なくとも一方が聞き役となっているようなセグメントの会話成立度を評価に用いないことが、3人以上の会話の抽出においては重要だからである。なお、会話成立度計算部450は、ステップS1900において、単に、会話成立度 $C_{i,j}(p)$ の決定を行わないようにしてもよい。

[0076] そして、ステップS2000において、会話成立度計算部450は、全てのペアについて会話成立度 $C_{i,j}(p)$ 算出の処理が終了したか否かを判断する。会話成立度計算部450は、全てのペアの処理が終了していない場合(S2000:NO)、ステップS1500へ戻り、未処理のペアを選択して処理を繰り返す。また、会話成立度計算部450は、全てのペアの処理が終了した場合(S2000:YES)、ステップS2100へ進む。

[0077] ステップS2100において、長時間特徴量計算部460は、ペアごとに、判定対象時間における会話成立度 $C_{i,j}(p)$ の長時間における平均である長時間特徴量 $L_{i,j}(p)$ を、例えば以下の式(2)を用いて算出する。ここで、パラメータ q は、判定対象時間の累積のセグメント数であり、現在のセ

グメント $Seg(p)$ のセグメントカウンター p の値である。また、多発話セグメントカウンター $g_{i,j}$ の値は、上述の通り、発話量 $H_{i,j}(p)$ が所定の閾値 θ 以上となっているセグメントの数を示す。

[数2]

$$L_{i,j}(p) = \frac{\sum_{q=1}^p C_{i,j}(q)}{g_{i,j}} \quad \dots\dots (2)$$

[0078] なお、音声処理装置 400 は、連続する所定数のフレームにおいて全ての音源に対して発話音声が無いと判定した場合、セグメントカウンター p および多発話セグメントカウンター $g_{i,j}$ を初期化してもよい。すなわち、音声処理装置 400 は、会話が行われていない状態が一定時間継続した時点で、初期化する。この場合、判定対象時間は、最後に会話が始まった時刻から現在時刻までの時刻までとなる。

[0079] そして、ステップ S2200 において、会話相手判定部 470 は、ユーザ 200 の会話相手を判定する会話相手判定処理を実行する。会話相手判定処理の詳細については後述する。

[0080] そして、ステップ S2300 において、出力音制御部 480 は、会話相手判定部 470 から入力される指向方向情報に基づき、イヤータップ 140L、140R からの出力音を制御する。すなわち、出力音制御部 480 は、判定されたユーザ 200 の会話相手の方向にマイクロホンアレイ 120 の指向性を向ける。

[0081] 図 7 は、マイクロホンアレイ 120 の指向性のパターンの例を示す図である。

[0082] まず、指向方向情報が左方、前方、右方を示す場合、または指向方向情報が左方および右方を示す場合を想定する。この場合、出力音制御部 480 は、図 7A に示すように、前方に広指向性を有するように、マイクロホンアレイ 120 を制御する。同様に、出力音制御部 480 は、会話開始時や、会話相手を判定できなかった場合も、前方に広指向性を有するように、マイクロ

ホンアレイ 120 を制御する。

- [0083] また、指向方向情報が左方と前方とを示す場合を想定する。この場合、出力音制御部 480 は、図 7 B に示すように、左斜め前方にやや広めの指向性を有するように、マイクロホンアレイ 120 を制御する。
- [0084] また、指向方向情報が前方と右方とを示す場合を想定する。この場合、出力音制御部 480 は、図 7 C に示すように、右斜め前方にやや広めの指向性を有するように、マイクロホンアレイ 120 を制御する。
- [0085] また、指向方向情報が前方のみを示す場合を想定する。この場合、出力音制御部 480 は、図 7 D に示すように、前方に狭指向性を有するように、マイクロホンアレイ 120 を制御する。
- [0086] また、指向方向情報が左方のみを示す場合を想定する。この場合、出力音制御部 480 は、図 7 E に示すように、左方に狭指向性を有するように、マイクロホンアレイ 120 を制御する。
- [0087] また、指向方向情報が右方のみを示す場合を想定する。この場合、出力音制御部 480 は、図 7 F に示すように、右方に狭指向性を有するように、マイクロホンアレイ 120 を制御する。
- [0088] そして、図 6 のステップ S 2400 において、音声処理装置 400 は、ユーザ操作等により処理の終了を指示されたか否かを判定する。音声処理装置 400 は、処理の終了を指示されていない場合 (S 2400 : NO)、ステップ S 1100 へ戻って、次のセグメントの処理に移る。また、音声処理装置 400 は、処理の終了を指示された場合 (S 2400 : YES)、一連の処理を終了する。
- [0089] なお、音声処理装置 400 は、会話が行われているかどうかを逐次判定し、会話が終了した場合、マイクロホンアレイ 120 の指向性を徐々に解除してもよい。この判定は、例えば、自発話パワー成分に基づいて行うことができる。
- [0090] 図 8 は、会話相手判定処理 (図 6 のステップ S 2200) を示すフローチャートである。

[0091] まず、ステップS2201において、会話相手判定部470は、全てのペアの長時間特徴量 $L_{i,j}(p)$ がまんべんなく高いか否かを判断する。具体的には、全てのペアの長時間特徴量 $L_{i,j}(p)$ の最大値および最小値をそれぞれMAX、MINとすると、会話相手判定部470は、所定の閾値 α 、 β について、以下の式(3)が満たされるか否かを判断する。

$$\text{MAX} - \text{MIN} < \alpha \quad \text{かつ} \quad \text{MIN} \geq \beta \quad \dots \dots (3)$$

[0092] 会話相手判定部470は、全てのペアの値がまんべんなく高い場合(S2201: YES)、ステップS2202へ進む。また、会話相手判定部470は、全てのペアの値がまんべんなく高くなっていない場合(S2201: NO)、ステップS2203へ進む。

[0093] ステップS2202において、会話相手判定部470は、4人(ユーザ200、左方の話者、前方の話者、および右方の話者)で会話を行っているかと判定し、図6の処理へ戻る。すなわち、会話相手判定部470は、左方の話者、前方の話者、および右方の話者を、ユーザ200の会話相手と判定し、左方、前方、および右方を示す指向方向情報を、出力音制御部480へ出力する。この結果、マイクロホンアレイ120は、前方に広指向性を有するように制御される(図7A参照)。

[0094] ステップS2203において、会話相手判定部470は、ユーザ200と他の各話者とから成る3つのペアのうち、特定の話者とのペアの長時間特徴量 $L_{i,j}(p)$ が突出して高くなっているか否かを判断する。具体的には、会話相手判定部470は、所定の閾値 γ について、以下の式(4)が満たされるか否かを判断する。ここで、SMAX1は、ユーザ200を含むペアの全ての長時間特徴量 $L_{i,j}(p)$ うちの最大値であり、SMAX2は、次に大きい値である。

$$\text{SMAX}1 - \text{SMAX}2 \geq \gamma \quad \dots \dots (4)$$

[0095] 会話相手判定部470は、特定の話者とのペアの値が突出して高くなっている場合(S2203: YES)、ステップS2204へ進む。また、会話相手判定部470は、特定の話者とのペアの値が突出して高くなっていない

場合（S 2 2 0 3 : N O）、ステップS 2 2 0 5へ進む。

[0096] ステップS 2 2 0 4において、会話相手判定部4 7 0は、長時間特徴量 $L_{i,j}(p)$ が突出して高くなっている相手が前方の話者であるか否かを判断する。すなわち、会話相手判定部4 7 0は、S M A X 1が、ユーザ2 0 0と前方の話者とのペアの長時間特徴量 $L_{0,1}(p)$ であるか否かを判断する。会話相手判定部4 7 0は、前方の話者との長時間特徴量 $L_{0,1}(p)$ が突出して高い場合（S 2 2 0 4 : Y E S）、ステップS 2 2 0 6へ進む。また、会話相手判定部4 7 0は、前方の話者との長時間特徴量 $L_{0,1}(p)$ が突出して高くない場合（S 2 2 0 4 : N O）、ステップS 2 2 0 7へ進む。

[0097] ステップS 2 2 0 6において、会話相手判定部4 7 0は、ユーザ2 0 0と前方の話者との2人で会話を行っているとは判定し、図6の処理へ戻る。すなわち、会話相手判定部4 7 0は、前方の話者をユーザ2 0 0の会話相手と判定し、前方を示す指向方向情報を、出力音制御部4 8 0へ出力する。この結果、マイクロホンアレイ1 2 0は、前方に狭指向性を有するように制御される（図7 D参照）。

[0098] ステップS 2 2 0 7において、会話相手判定部4 7 0は、長時間特徴量 $L_{i,j}(p)$ が突出して高くなっている相手が左方話者であるか否かを判断する。すなわち、会話相手判定部4 7 0は、S M A X 1が、ユーザ2 0 0と左方話者とのペアの長時間特徴量 $L_{0,2}(p)$ であるか否かを判断する。会話相手判定部4 7 0は、左方話者との長時間特徴量 $L_{0,2}(p)$ が突出して高い場合（S 2 2 0 7 : Y E S）、ステップS 2 2 0 8へ進む。また、会話相手判定部4 7 0は、左方話者との長時間特徴量 $L_{0,2}(p)$ が突出して高くない場合（S 2 2 0 7 : N O）、ステップS 2 2 0 9へ進む。

[0099] ステップS 2 2 0 8において、会話相手判定部4 7 0は、ユーザ2 0 0と左方話者との2人で会話を行っているとは判定し、図6の処理へ戻る。すなわち、会話相手判定部4 7 0は、左方話者をユーザ2 0 0の会話相手と判定し、左方を示す指向方向情報を、出力音制御部4 8 0へ出力する。この結果、マイクロホンアレイ1 2 0は、左方に狭指向性を有するように制御され

る（図7E参照）。

[0100] ステップS2209において、会話相手判定部470は、ユーザ200と右方の話者との2人で会話を行っているとは判定し、図6の処理へ戻る。すなわち、会話相手判定部470は、右方の話者をユーザ200の会話相手と判定し、右方を示す指向方向情報を、出力音制御部480へ出力する。この結果、マイクロホンアレイ120は、右方に狭指向性を有するように制御される（図7F参照）。

[0101] ステップS2205へ処理が進んだ場合、全員参加の会話でも2人会話でもない。すなわち、前方、左方、右方の話者のいずれか一人は、ユーザ200とは無関係な話者である可能性が高い。

[0102] そこで、まず、ステップS2205において、会話相手判定部470は、ユーザ200と前方の話者とのペアの長時間特徴量 $L_{0,1}(p)$ が、所定の閾値 η 以上となっているか否かを判断する。会話相手判定部470は、長時間特徴量 $L_{0,1}(p)$ が閾値 η 未満である場合（S2205：YES）、ステップS2210へ進む。また、会話相手判定部470は、長時間特徴量 $L_{0,1}(p)$ が閾値 η 以上である場合（S2205：NO）、ステップS2211へ進む。

[0103] ステップS2210において、会話相手判定部470は、ユーザ200と、左方の話者と、右方との3人で会話を行っているとは判定し、図6の処理へ戻る。すなわち、会話相手判定部470は、左方の話者および右方の話者をユーザ200の会話相手と判定する。そして、会話相手判定部470は、左方と右方とを示す指向方向情報を、出力音制御部480へ出力する。この結果、マイクロホンアレイ120は、前方に広指向性を有するように制御される（図7A参照）。

[0104] ステップS2211において、会話相手判定部470は、ユーザ200と左方の話者とのペアの長時間特徴量 $L_{0,2}(p)$ が、所定の閾値 η 以上となっているか否かを判断する。会話相手判定部470は、長時間特徴量 $L_{0,2}(p)$ が閾値 η 未満である場合（S2211：YES）、ステップS2212へ

進む。また、会話相手判定部470は、長時間特徴量 $L_{0,2}(p)$ が閾値 η 以上である場合(S2211:NO)、ステップS2213へ進む。

[0105] ステップS2212において、会話相手判定部470は、ユーザ200と、前方の話者と、右方との3人で会話を行っているとは判定し、図6の処理へ戻る。すなわち、会話相手判定部470は、前方の話者および右方の話者をユーザ200の会話相手と判定する。そして、会話相手判定部470は、前方と右方とを示す指向方向情報を、出力音制御部480へ出力する。この結果、マイクロホンアレイ120は、右前方にやや広めの指向性を有するように制御される(図7C参照)。

[0106] ステップS2213において、会話相手判定部470は、ユーザ200と右方の話者とのペアの長時間特徴量 $L_{0,3}(p)$ が、所定の閾値 η 以上となっているか否かを判断する。会話相手判定部470は、長時間特徴量 $L_{0,3}(p)$ が閾値 η 未満である場合(S2213:YES)、ステップS2214へ進む。また、会話相手判定部470は、長時間特徴量 $L_{0,3}(p)$ が閾値 η 以上である場合(S2213:NO)、ステップS2215へ進む。

[0107] ステップS2214において、会話相手判定部470は、ユーザ200と、前方の話者と、左方との3人で会話を行っているとは判定し、図6の処理へ戻る。すなわち、会話相手判定部470は、前方の話者および左方の話者をユーザ200の会話相手と判定し、前方と左方とを示す指向方向情報を、出力音制御部480へ出力する。この結果、マイクロホンアレイ120は、左前方にやや広めの指向性を有するように制御される(図7B参照)。

[0108] ステップS2215において、会話相手判定部470は、ユーザ200の会話相手を判定不能として、指向方向情報を出力せずに、図6の処理へ戻る。この結果、出力音の指向性は、デフォルトの状態または最後の判定結果に応じた状態が維持されることになる。

[0109] 上述の通り、全ての話者が同じ会話に参加している場合、全てのペアの長時間特徴量 $L_{i,j}(p)$ は、まんべんなく高くなる。また、2人会話の場合、ユーザ200と会話相手とペアの長時間特徴量 $L_{0,j}(p)$ のみが突出して高

くなり、ユーザ200と残りの音源とのペアの長時間特徴量 $L_{0,j}(p)$ は低くなる。

[0110] したがって、以上説明した動作により、音声処理装置400は、ユーザ200の会話相手を精度良く判定し、ユーザ200が属する会話グループを高精度に抽出することができる。

[0111] また、音声処理装置400を含む補聴器100は、ユーザ200の会話相手を精度良く判定することができるので、ユーザ200が会話相手の音声を聴き取り易くなるように、出力音を制御することができる。また、補聴器100は、途中で会話グループが変化する場合でも、その変化に追従して指向性を制御することができる。途中で会話グループが変化する場合とは、例えば、2人会話の途中で参加者が増えて3人、4人になったり、4人の会話から参加者が減って3人、2人の会話になる場合である。

[0112] なお、マイクロホンアレイ120の指向性を急激に切り替えると、ユーザ200に大きな違和感を与えてしまう可能性がある。したがって、出力音制御部480は、ある程度の時間を掛けて、徐々に指向性を変化させるようにしてもよい。また、会話相手の人数を判定するには、後述の通り、ある程度時間が掛かる。したがって、補聴器100は、会話開始から一定時間が経過してから、指向性の制御を行うようにしてもよい。

[0113] また、マイクロホンアレイ120に指向性が形成された場合、指向方向以外の発話音声聞き取り辛くなる。例えば、3人会話が行われているときに2人会話であると誤判定されると、一人の話者の発話が聞こえなくなってしまうという問題が生じる。3人会話を2人会話と誤判定するよりも、2人会話を3人会話と誤判定する方が、ユーザ200にとって会話をする上での弊害が少ないといえる。したがって、閾値 α 、 β 、 γ は、実際よりも少ない人数に判定することを防ぐことができる値に設定されることが望ましい。すなわち、 γ および α は高めに設定し、 β は低めに設定すればよい。

[0114] 以下、本発明の効果について、実験結果を参照して説明する。

[0115] 実験は、2人から成る会話グループと、3人から成る会話グループとをそ

れぞれ5組ずつ用意し、各会話グループから収録した10分間の会話の音声データに対して行った。会話内容は、日常会話（雑談）である。発話区間を定義する発話開始時刻および発話終了時刻については、予め、試聴によりラベル付けを行った。また、実験は、簡単のため、2人会話であるか3人会話であるかの判定についての精度を測定するものとした。

[0116] 本実験に係る音声処理方法は、2人から成る会話グループに対しては、話者のうちの1人をユーザ200とし、1人を前方の話者とした。そして、本実験では、別の会話グループを形成する2人の話者を更に用意し、そのうちの一人を、ユーザ200の左方の話者とした。

[0117] また、本実験では、3人から成る会話グループに対して、話者のうちの1人をユーザ200とし、1人を前話者とし、残りの1人の話者を左話者とした。

[0118] 本発明に係る音声処理方法（以下「本発明」という）は、発話量を考慮したセグメントごとの会話成立度に基づくものであり、10秒ごとに会話相手判定を行った。

[0119] 図9は、実験のために簡略化した会話相手判定処理を示すフローチャートであり、図8に対応するものである。図8と同一部分には同一ステップ番号を付し、これについての説明を省略する。

[0120] 図9に示すように、実験においては、本発明は、会話相手判定部470は、全てのペアの長時間特徴量 $L_{i,j}(p)$ がまんべんなく高い場合、3人全員での会話であると判定するようにした。また、本発明は、3人会話ではない場合、左方、前方のいずれかの方向の話者との2人会話であると判定するようにした。更に、音声処理装置400は、どの話者との2人会話であるかが判定されない場合、高指向性とするために、3人全員での会話であると判定するようにした。

[0121] また、実験において、閾値 α 、 β は、それぞれ $\alpha=0.09$ 、 $\beta=0.54$ とした。また、抽出精度の指標値は、会話相手を正しく検出する割合と非会話相手を正しく棄却する割合との平均値である会話相手検出率を定義した。

[0122] 本発明に対しては、2人会話の場合においては前方の話者との2人会話であると判定されたときに正解とし、3人会話の場合においては3人会話であると判定されたときに正解とした。

[0123] 一方、比較のために採用した従来技術に係る音声処理方法（以下「従来法」という）は、特許文献1の実施の形態に開示されている方法を拡張したものであり、具体的には、以下のような方法とした。

[0124] 従来法は、会話開始からの会話成立度をフレームごとに求め、10秒ごとに、会話相手との会話成立度が閾値 T_h を超えていたときに正解と判定し、非会話相手との会話成立度が閾値 T_h を下回っていたときに正解と判定するようにした。なお、従来法は、時定数を用いて会話成立度を更新し、フレーム t における会話成立度 $C_{i,j}(t)$ を、以下の式(5)を用いて算出するようにした。

$$C_{i,j}(t) = \varepsilon \cdot C_{i,j}(t-1) + (1-\varepsilon) [R_{i,j}(t) + T_{i,j}(t) + (1-D_{i,j}(t)) + (1-S_{i,j}(t))] \dots (5)$$

但し、 S_j が発話音声有りのとき、 $V_j(t) = i$

S_j が発話音声無しの場合、 $V_j(t) = 0$

$$D_{i,j}(t) = \alpha \cdot D_{i,j}(t-1) + (1-\alpha) V_i(t) \cdot V_j(t)$$

$$R_{i,j}(t) = \beta \cdot R_{i,j}(t-1) + (1-\beta) (1-V_i(t)) V_j(t)$$

$$T_{i,j}(t) = \gamma \cdot T_{i,j}(t-1) + (1-\gamma) V_i(t) \cdot (1-V_j(t))$$

$$S_{i,j}(t) = \Delta \cdot S_{i,j}(t-1) + (1-\delta) (1-V_i(t)) (1-V_j(t))$$

$$\alpha = \beta = \gamma = 0.99999$$

$$\delta = 0.999995$$

$$\varepsilon = 0.999$$

[0125] 図10は、従来法による会話相手正解率と本発明による会話相手正解率との比較を示すプロット図である。図10において、横軸は会話開始からの時間を示し、縦軸は会話開始時から現在時刻までの会話相手判定正解率の累積平均値を示す。白丸印(○)は、2人会話に対する従来法の実験値を示し、白三角印(△)は、3人会話に対する従来法の実験値を示す。黒丸印(●)は、2人会話に対する本発明の実験値を示し、黒三角印(▲)は、3人会話に対する本発明の実験値を示す。

[0126] 図10から分かるように、従来法に比べて、本発明の会話相手検出正解率は大幅に向上している。特に、本発明は、3人会話において、従来法に比べて非常に早い段階において、会話相手を精度良く検出している。このように、本発明は、3人以上から成る会話グループが存在する場合でも、複数の話者の中から高い精度で会話グループを抽出することができる。

[0127] なお、従来法では、時間的により新しい情報により高い重み付けを行うために時定数を用いている。ところが、3人以上の会話において1対1の会話関係が成り立つのは、通常、2～3発話程度の比較的短い期間である。したがって、従来法では、ある時点での会話成立を検出するためには、時定数を小さくする必要がある。ところが、そのような短い期間では、発話を行っていない聞き役の話者を含むペアの会話成立度は低くなるため、2人会話であるか3人会話であるかの区別は困難となるだけでなく、会話相手の判定精度は低くなる。

[0128] 以上のように、本実施の形態に係る補聴器100は、各ペアの局所的な会話成立度を時間的にシフトさせながら求め、発話量と高いセグメントの会話成立度を長時間観測することにより、ユーザ200の会話相手を判定する。これにより、本実施の形態に係る補聴器100は、ユーザ200が2人会話を行っている場合のみならず、3人会話を行っている場合でも、その会話が成立していることを正しく判定することができる。すなわち、本実施の形態に係る補聴器100は、3人以上から成る会話グループであっても、高い

精度で抽出することができる。

- [0129] また、補聴器 100 は、会話グループを高い精度で抽出することができるので、マイクロホンアレイ 120 の指向性を適切に制御し、ユーザ 200 に会話相手の発話を聴き取り易くすることができる。また、補聴器 100 は、会話グループに対して高い追従性を有するので、会話相手の発話を聴き取り易い状態を、会話開始の早い段階で確保し、また、それを維持することができる。
- [0130] なお、音源分離のための指向性の方向は、上述の、前方、左方、右方の 3 方向の組み合わせに制限されない。例えば、補聴器 100 は、マイクロホンの個数を増やす等して指向性の角度を狭められる場合、更に多数の方向について指向性を制御し、4 人よりも多くの話者を対象として会話相手の判定を行うようにしてもよい。
- [0131] また、補聴器 100 の筐体 110 L、110 R とリモコン装置 150 とは、有線ではなく、無線により通信可能に接続されていてもよい。また、補聴器 100 は、筐体 110 L、110 R に DSP (digital signal processor) を備え、制御処理の一部または全てを、リモコン装置 150 においてではなくこの DSP において実行するようにしてもよい。
- [0132] また、補聴器 100 は、発話の検出を、方向別に音を分離することによってではなく、独立成分分析 (ICA) 等の他の音源分離方法を用いて行うようにしてもよい。また、補聴器 100 は、発話者ごとに個別マイクを配置して、各発話者から発話音声を入力してもよい。
- [0133] また、補聴器 100 は、いわゆるウェアラブルマイクではなく、卓上に置かれたマイクロホンアレイを用いて、音源分離を行ってもよい。この場合、ユーザ 200 の方向を予め設定しておくことにより、自発話検出の処理が不要となる。
- [0134] また、補聴器 100 は、自発話と他発話とを音響信号における音響的な特徴の違いから識別してもよい。この場合、同じ方向に複数の話者がいる場合にも、話者ごとに音源を分離することが可能となる。

[0135] なお、以上説明した実施の形態では、本発明を補聴器に適用した例について説明したが、本発明の適用はこれに制限されない。例えば、本発明は、音声レコーダ、デジタルスチルカメラ、デジタルビデオカメラ、電話会議システム等、複数の話者の発話音声を入力する各種装置やアプリケーションソフトウェアに適用することができる。会話グループの抽出結果は、出力音の制御以外の各種用途に供することができる。

[0136] 例えば、本発明を電話会議システムに適用した場合、発言者の音声を明瞭に出力および記録するためにマイクロホンの指向性を制御したり、参加者の人数の検出とその記録を行ったりすることができる。例えば、2拠点間で行われる電話会議において、一方の拠点の入力音声に妨害音がある場合に、他方の拠点の発話者に対する会話相手の発話音声を特定し、これを抽出することにより、スムーズに会議を行うことが可能となる。また、両拠点において妨害音がある場合、例えば、マイクロホンに入力される発話音声のうち最も音量の高い発話音声を検出し、その会話相手を特定することにより、同様の効果を得ることができる。

[0137] また、本発明を音声レコーダ等のデジタル記録機器に適用した場合、他人の会話等の、会話相手の発話音声に対する妨害音が抑制されるようにマイクロホンアレイを制御することができる。

[0138] また、いずれの用途においても、全ての方向からの音声を方向ごとに記録しておき、後から会話成立度が高くなる組み合わせの音声データを抽出し、所望の会話を再生してもよい。

[0139] 2010年9月28日出願の特願2010-217192の日本出願に含まれる明細書、図面および要約書の開示内容は、すべて本願に援用される。

産業上の利用可能性

[0140] 本発明は、3人以上から成る会話グループが存在する場合でも、複数の話者の中から高い精度で会話グループを抽出することができる音声処理装置および音声処理方法として有用である。

符号の説明

- [0141] 100 補聴器
- 110L、110R 筐体
 - 120 マイクロホンアレイ
 - 130L、130R スピーカ
 - 140L、140R イヤーチップ
 - 150 リモコン装置
 - 160 CPU
 - 170 メモリ
- 400 音声処理装置
- 410 AD変換部
 - 420 自発話検出部
 - 430 方向別音声検出部
 - 435 音声検出部
 - 440 発話量和計算部
 - 450 会話成立度計算部
 - 460 長時間特徴量計算部
 - 470 会話相手判定部
 - 480 出力音制御部

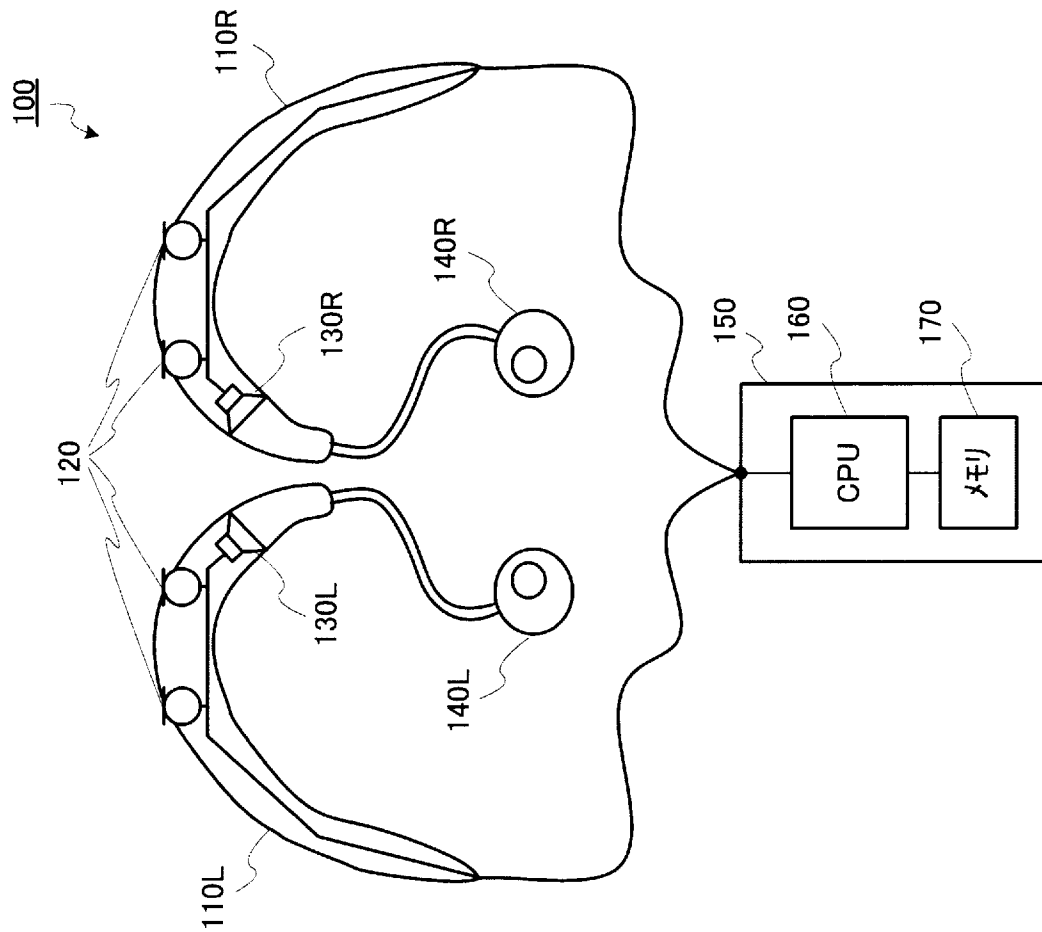
請求の範囲

- [請求項1] 音響信号から複数の話者の発話音声を個別に検出する音声検出部と、
、
検出された前記発話音声に基づいて、前記複数の話者のうちの2人の組み合わせの全てについて、判定対象時間を区切ったセグメント毎に会話成立度を算出する会話成立度計算部と、
前記組み合わせ毎に、前記判定対象時間における前記会話成立度の長時間特徴量を算出する長時間特徴量計算部と、
算出された前記長時間特徴量に基づいて、前記複数の話者の中から、会話を形成する会話グループを抽出する会話相手判定部と、を有する、
音声処理装置。
- [請求項2] 前記会話成立度は、2人の前記話者のうち一方が発話し他方が発話していない時間の割合の高さを示す値である、
請求項1記載の音声処理装置。
- [請求項3] 前記組み合わせの全てについて、前記セグメント毎に、前記話者の発話量の和である発話量とを算出する発話量計算部、を更に有し、
前記会話成立度計算部は、
前記発話量とが所定の閾値未満となる前記セグメントの前記会話成立度を、前記長時間特徴量の算出に関して無効化する、
請求項1記載の音声処理装置。
- [請求項4] 前記音響信号は、前記複数の話者の一人であるユーザの近傍に配置され可変の指向性を有する音声入力手段において入力された音声の音響信号であり、
抽出された前記会話グループに前記ユーザが含まれるとき、当該会話グループの前記ユーザ以外の前記話者の方向に指向性が向くように、前記音声入力手段を制御する出力音制御部、を更に有する、
請求項1記載の音声処理装置。

- [請求項5] 前記出力音制御部は、
前記音響信号に対して所定の信号処理を行い、前記所定の信号処理が行われた前記音響信号を、前記ユーザが装着する補聴器のスピーカへ出力する、
請求項4記載の音声処理装置。
- [請求項6] 前記音声検出部は、
前記ユーザを基準とした所定の複数の方向毎に、その方向に位置する話者の発話音声を検出し、
前記出力音制御部は、
抽出された前記会話グループに属する前記ユーザ以外の話者が位置する方向に前記指向性が向くように、前記音声入力手段を制御する、
請求項4記載の音声処理装置。
- [請求項7] 前記会話相手判定部は、前記組み合わせのうちの複数組において前記長時間特徴量がまんべんなく高いとき、当該複数組を構成する複数の前記話者を、同一の会話グループに属すると判定する、
請求項1記載の音声処理装置。
- [請求項8] 前記会話相手判定部は、
前記ユーザを含む前記組み合わせにおいて、最も高い前記長時間特徴量と次に高い前記長時間特徴量との差が所定の閾値以上であるとき、前記最も高い前記長時間特徴量に該当する前記ユーザ以外の話者を、前記ユーザの唯一の会話相手と判定する、
請求項1記載の音声処理装置。
- [請求項9] 前記判定対象時間は、前記ユーザが参加する会話が最後に開始されてから現在までの時間である、
請求項1記載の音声処理装置。
- [請求項10] 音響信号から複数の話者の発話音声を個別に検出するステップと、
検出された前記発話音声に基づいて、前記複数の話者のうちの2人の組み合わせの全てについて、判定対象時間を区切ったセグメント毎

に会話成立度を算出するステップと、
前記組み合わせ毎に、前記判定対象時間における前記会話成立度の
長時間特徴量を算出するステップと、
算出された前記長時間特徴量に基づいて、前記複数の話者の中から
、会話を形成する会話グループを抽出するステップと、を有する、
音声処理方法。

[図1]



[図2]

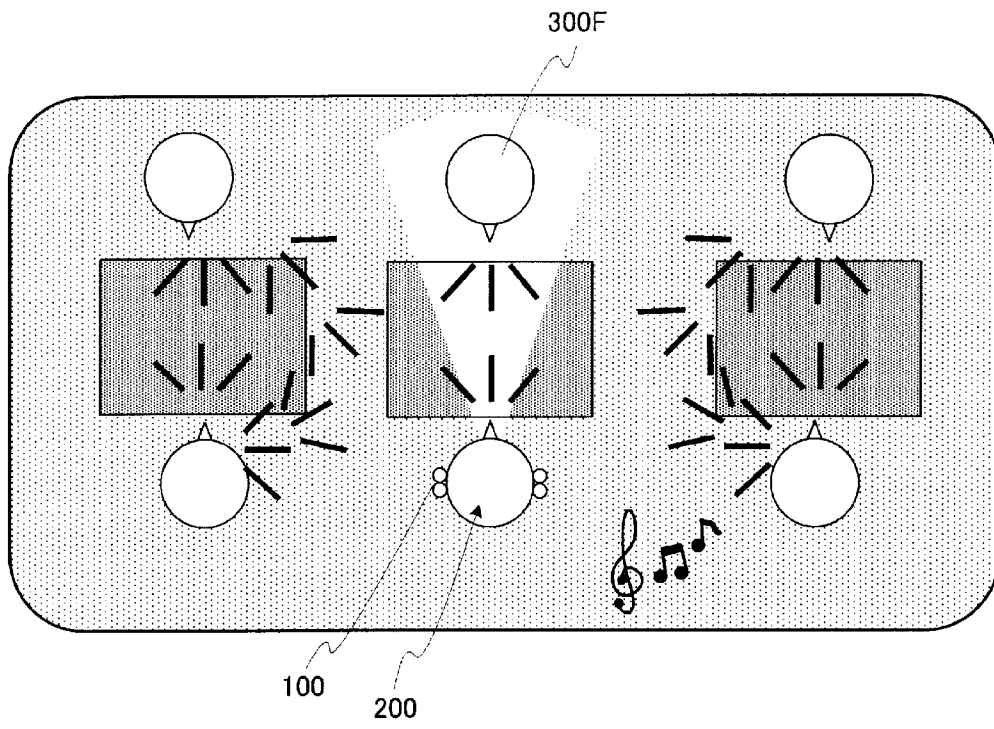


図2A

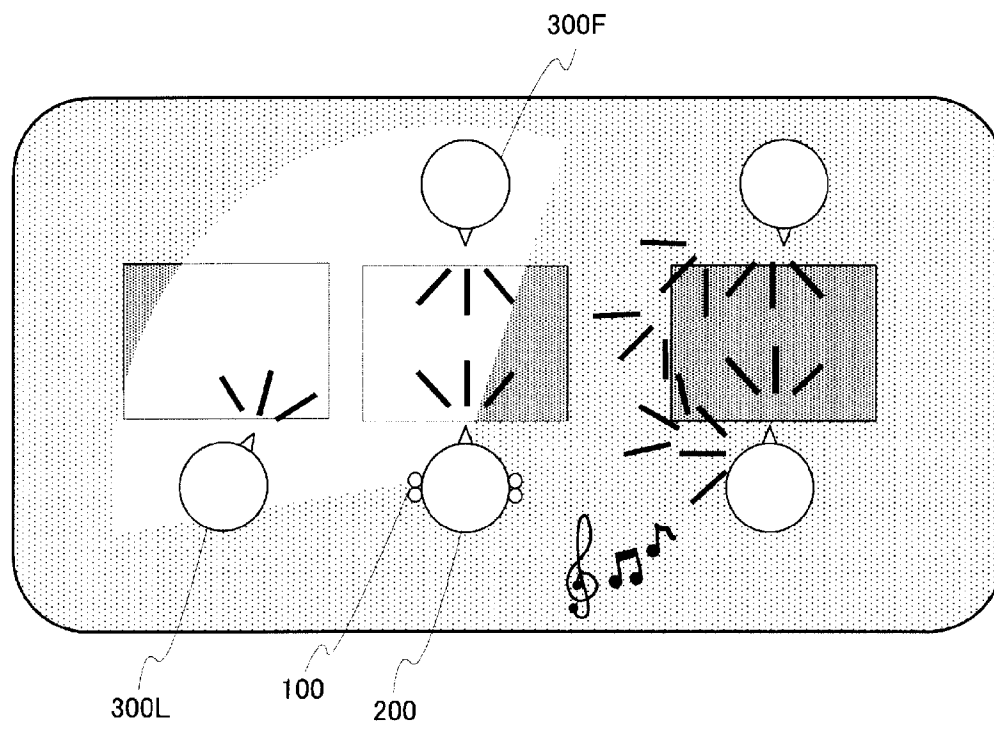
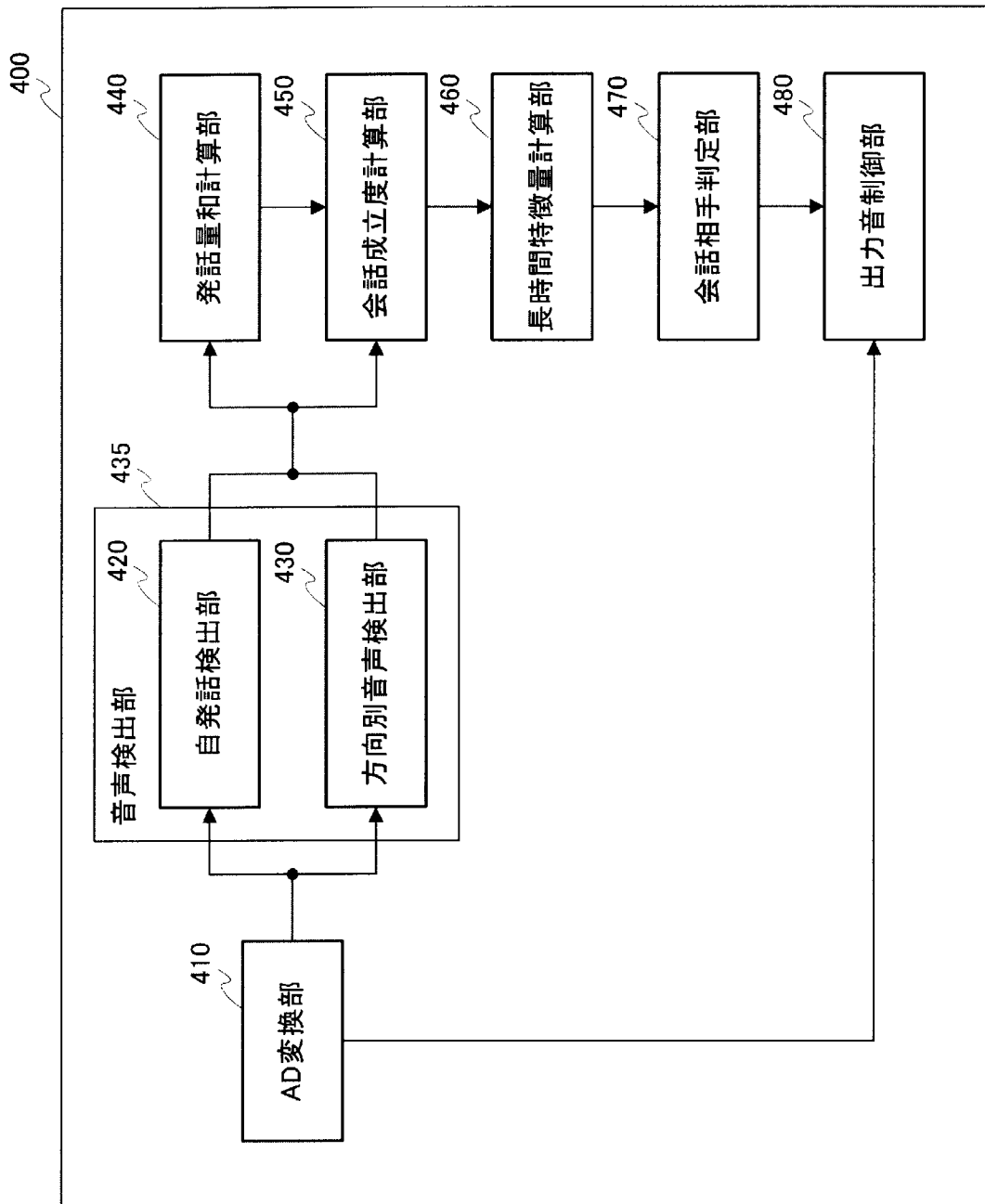
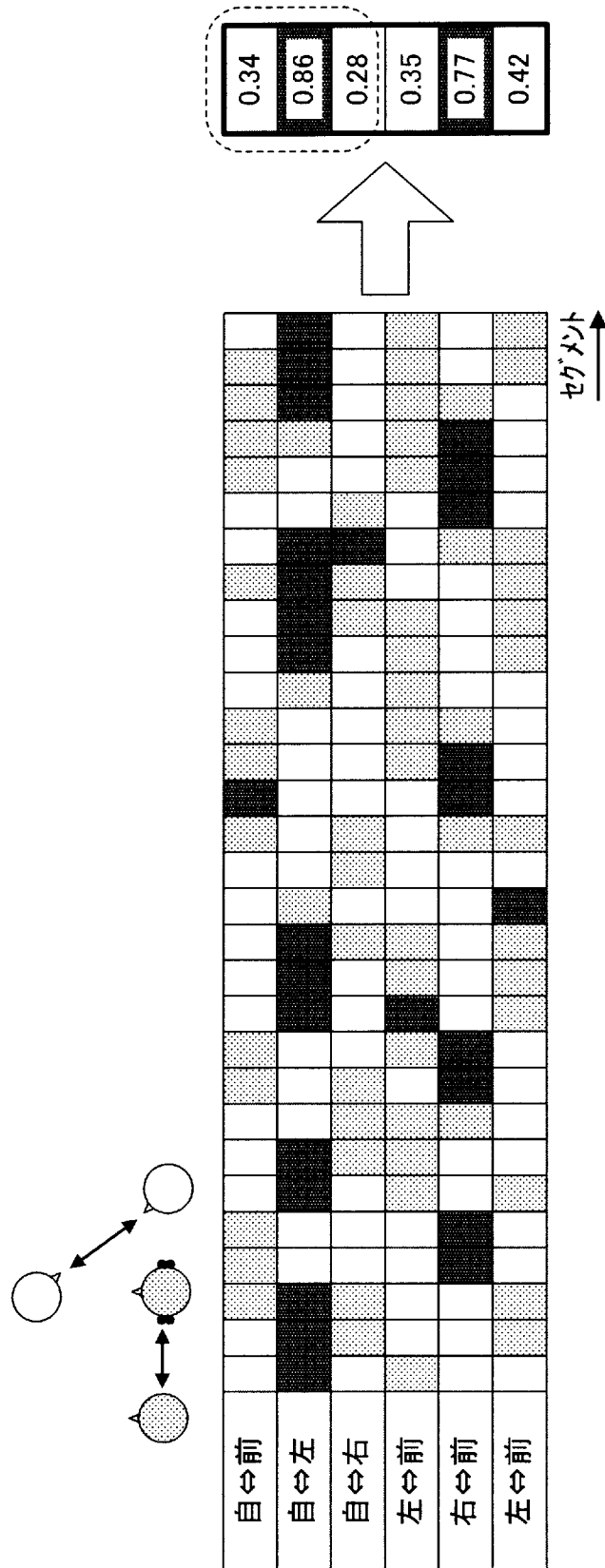


図2B

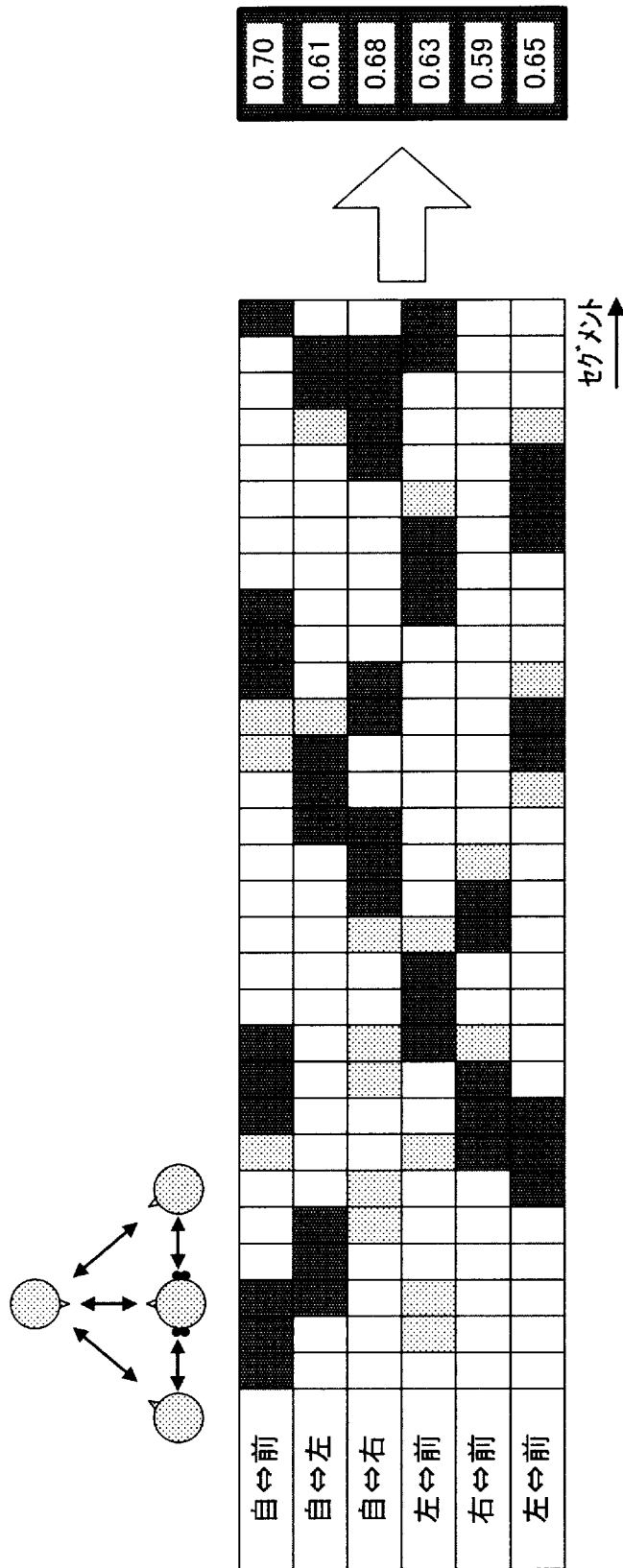
[図3]



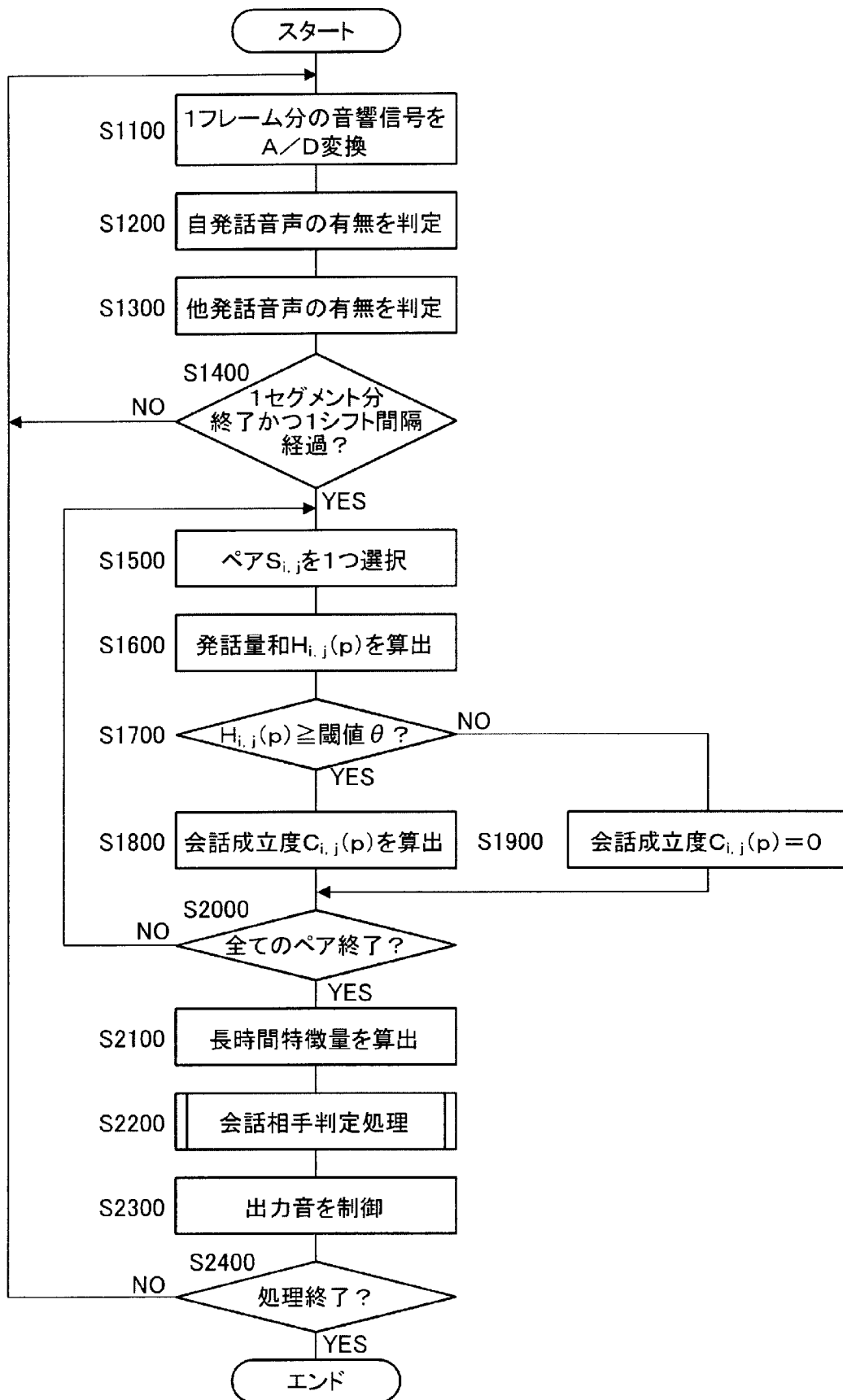
[図4]



[図5]



[図6]



[図7]

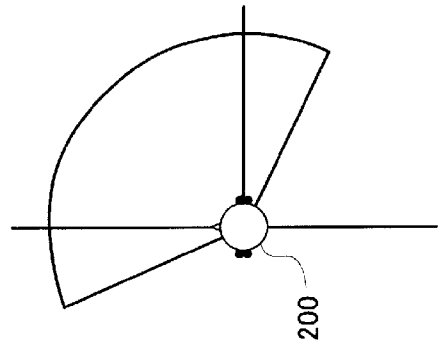


図7A

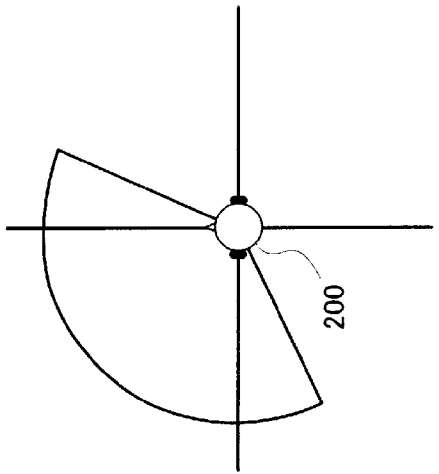


図7B

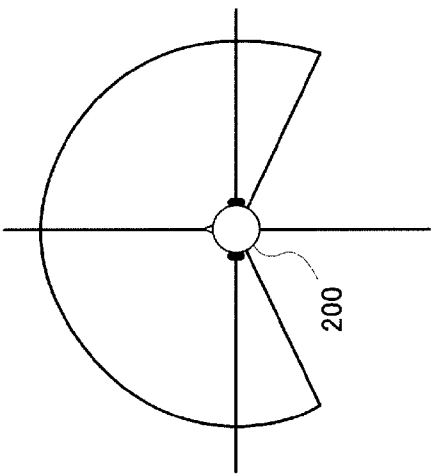


図7C

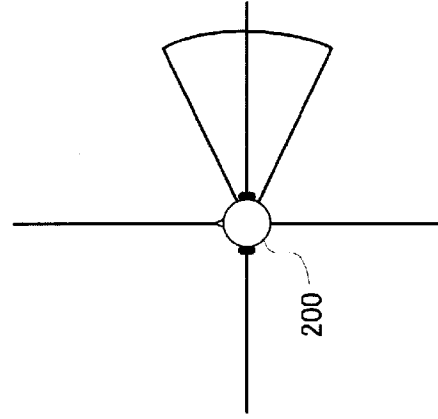


図7D

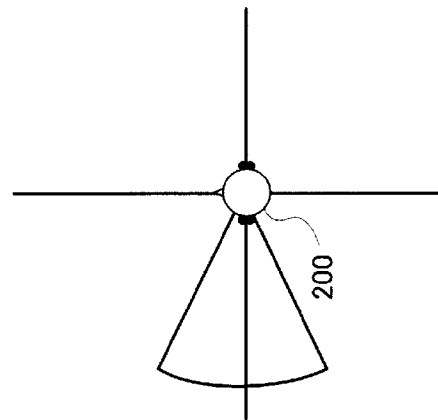


図7E

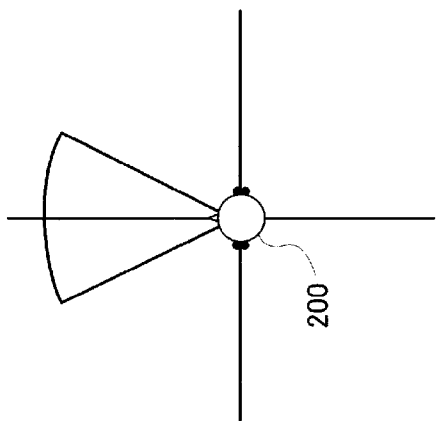
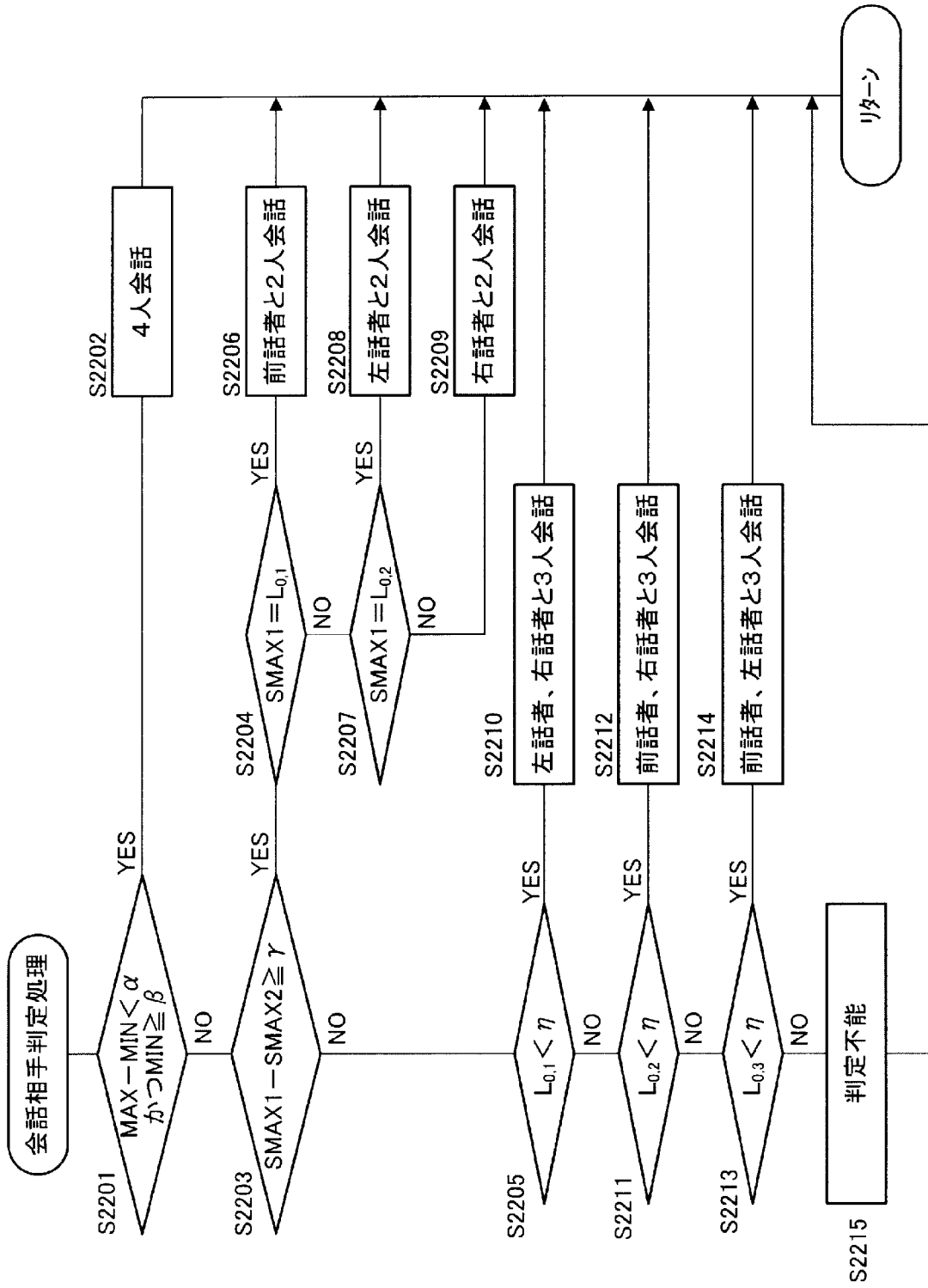
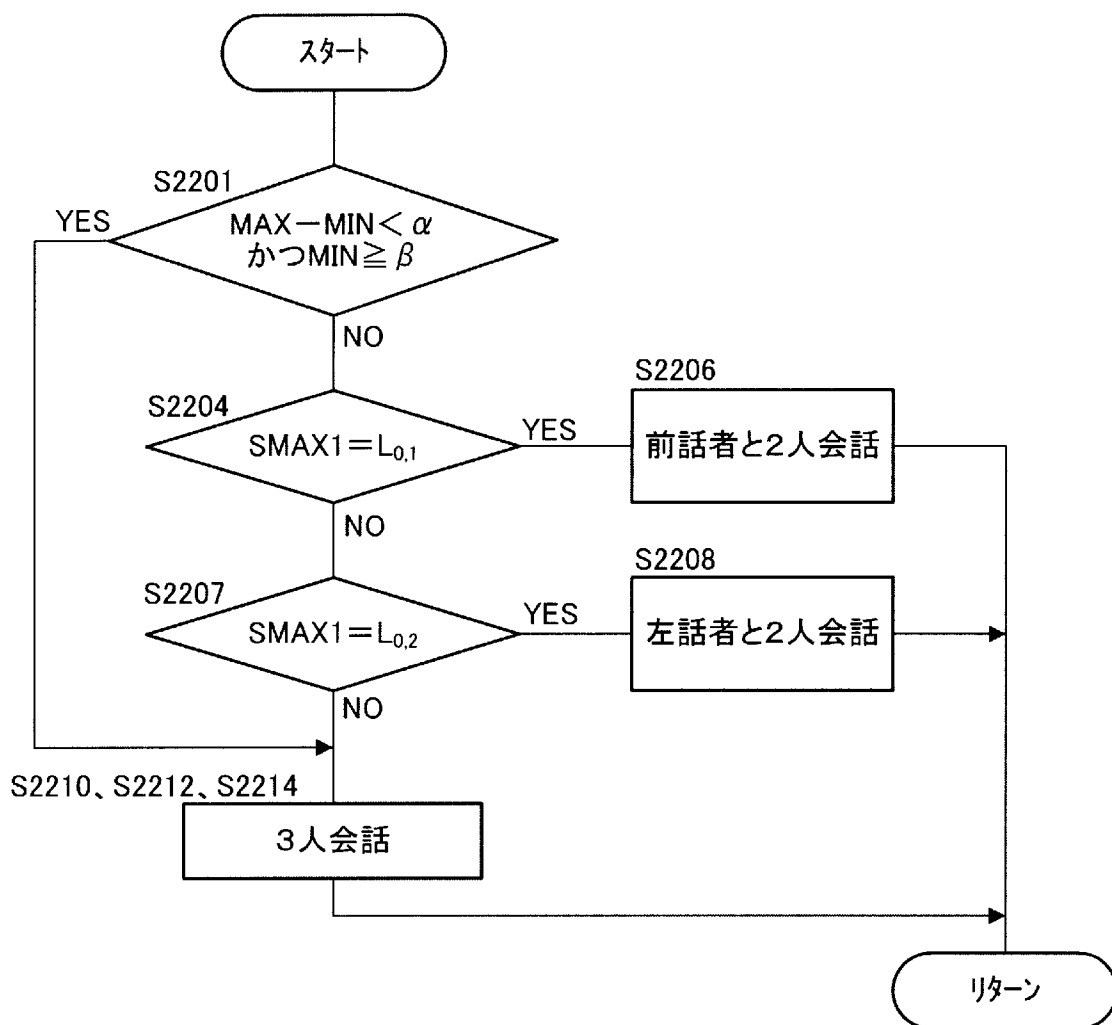


図7F

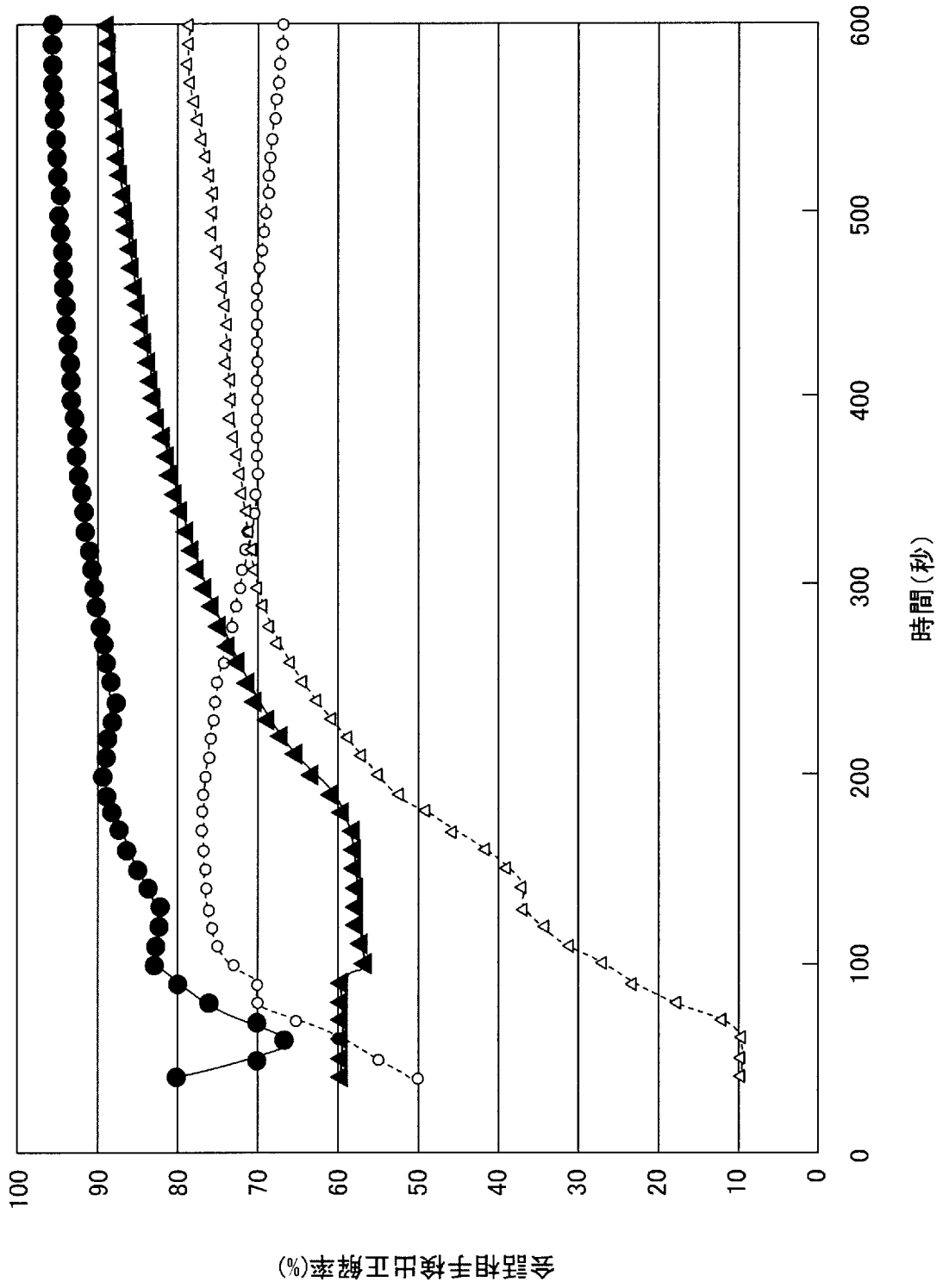
[図8]



[図9]



[図10]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2011/005173

A. CLASSIFICATION OF SUBJECT MATTER

G10L11/00(2006.01)i, G10L11/02(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)

G10L11/00-21/06

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 1996-2011

Kokai Jitsuyo Shinan Koho 1971-2011 Toroku Jitsuyo Shinan Koho 1994-2011

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

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 2004-133403 A (Kobe Steel, Ltd.), 30 April 2004 (30.04.2004), entire text; all drawings (Family: none)	1-10
A	JP 2005-202035 A (Toshiba Corp.), 28 July 2005 (28.07.2005), entire text; all drawings (Family: none)	1-10
A	WO 2009/104332 A1 (NEC Corp.), 27 August 2009 (27.08.2009), entire text; all drawings (Family: none)	1-10

 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
14 October, 2011 (14.10.11)Date of mailing of the international search report
25 October, 2011 (25.10.11)Name and mailing address of the ISA/
Japanese Patent Office

Authorized officer

Facsimile No.

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2011/005173

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 2008-242318 A (Toshiba Corp.), 09 October 2008 (09.10.2008), entire text; all drawings & US 2008/0243494 A1	1-10
A	JP 2005-157086 A (Matsushita Electric Industrial Co., Ltd.), 16 June 2005 (16.06.2005), entire text; all drawings (Family: none)	1-10
A	WO 02/085066 A1 (WIDEX A/S), 24 October 2002 (24.10.2002), entire text; all drawings & CA 2440233 A & EP 1380187 A & JP 2004-527177 A & AT 410901 T & DE 60229227 D & DK 1380187 T	1-10
P,A	WO 2011/105003 A1 (Panasonic Corp.), 01 September 2011 (01.09.2011), entire text; all drawings (Family: none)	1-10

A. 発明の属する分野の分類 (国際特許分類 (IPC))
 Int.Cl. G10L11/00(2006.01)i, G10L11/02(2006.01)i

B. 調査を行った分野
 調査を行った最小限資料 (国際特許分類 (IPC))
 Int.Cl. G10L11/00-21/06

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

日本国実用新案公報	1922-1996年
日本国公開実用新案公報	1971-2011年
日本国実用新案登録公報	1996-2011年
日本国登録実用新案公報	1994-2011年

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)

C. 関連すると認められる文献

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 請求項の番号
A	JP 2004-133403 A (株式会社神戸製鋼所) 2004.04.30, 全文、全図 (ファミリーなし)	1-10
A	JP 2005-202035 A (株式会社東芝) 2005.07.28, 全文、全図 (ファミリーなし)	1-10
A	WO 2009/104332 A1 (日本電気株式会社) 2009.08.27, 全文、全図 (ファミリーなし)	1-10

C欄の続きにも文献が列挙されている。 パテントファミリーに関する別紙を参照。

* 引用文献のカテゴリー	の日の後に公表された文献
「A」特に関連のある文献ではなく、一般的な技術水準を示すもの	「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
「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	「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
「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)	「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
「O」口頭による開示、使用、展示等に言及する文献	「&」同一パテントファミリー文献
「P」国際出願日前で、かつ優先権の主張の基礎となる出願	

国際調査を完了した日 14.10.2011	国際調査報告の発送日 25.10.2011	
国際調査機関の名称及びあて先 日本国特許庁 (ISA/JP) 郵便番号100-8915 東京都千代田区霞が関三丁目4番3号	特許庁審査官 (権限のある職員) 井上 健一	5Z 9373
	電話番号 03-3581-1101 内線 3591	

C (続き) . 関連すると認められる文献		
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 請求項の番号
A	JP 2008-242318 A (株式会社東芝) 2008. 10. 09, 全文、全図 & US 2008/0243494 A1	1-10
A	JP 2005-157086 A (松下電器産業株式会社) 2005. 06. 16, 全文、全 図 (ファミリーなし)	1-10
A	WO 02/085066 A1 (WIDEX A/S) 2002. 10. 24, 全文、全図 & CA 2440233 A & EP 1380187 A & JP 2004-527177 A & AT 410901 T & DE 60229227 D & DK 1380187 T	1-10
P, A	WO 2011/105003 A1 (パナソニック株式会社) 2011. 09. 01, 全文、全 図 (ファミリーなし)	1-10