

(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2010-56763

(P2010-56763A)

(43) 公開日 平成22年3月11日(2010.3.11)

(51) Int.Cl.	F I	テーマコード (参考)
H04R 1/40 (2006.01)	H04R 1/40 320A	5D015
G10L 15/20 (2006.01)	G10L 15/20 370E	5D018

審査請求 未請求 請求項の数 5 O L (全 15 頁)

(21) 出願番号	特願2008-218399 (P2008-218399)	(71) 出願人	000006297
(22) 出願日	平成20年8月27日 (2008. 8. 27)		
		(71) 出願人	村田機械株式会社
			京都府京都市南区吉祥院南落合町3番地
		(71) 出願人	593006630
			学校法人立命館
			京都府京都市中京区西ノ京梅尾町1番地の7
		(74) 代理人	100101454
			弁理士 山田 卓二
		(74) 代理人	100081422
			弁理士 田中 光雄
		(74) 代理人	100098280
			弁理士 石野 正弘

最終頁に続く

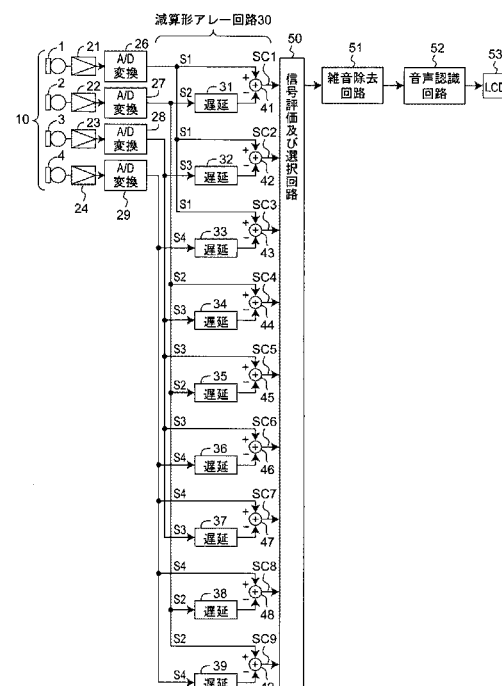
(54) 【発明の名称】 音声認識装置

(57) 【要約】

【課題】例えば工場などの大きな雑音を発生する現場において音声認識率を従来技術に比較して向上させることができる音声認識装置を提供する。

【解決手段】マイクロホンアレー10は例えば正三角錐の各頂点に配置された4個のマイクロホン1～4を備えて構成される。減算形アレー回路30は、各マイクロホン1～4からの音声信号に基づいて減算形アレー法を用いて各マイクロホンから各マイクロホンに対する方向にゼロ点を生成するカージオイドを生成してそれらに対応する複数のカージオイド信号を発生する。信号評価及び選択回路50は複数のカージオイド信号のうちより高い信号対雑音電力比を有する複数のカージオイド信号を検出して選択し、選択した複数のカージオイド信号を加算し、加算結果のカージオイド信号を出力する。音声認識回路52はカージオイド信号に基づいて音声認識を行って音声認識結果を出力する。

【選択図】 図4



【特許請求の範囲】

【請求項 1】

角錐の各頂点のうち上部頂点において、放射主軸が話者の口元に実質的に向くように設けられた第 1 のマイクロホンと、上記角錐の底面の少なくとも 2 つの頂点において、放射主軸が話者の口元方向に実質的に平行となるように設けられた複数の第 2 のマイクロホンとを備えたマイクロホンアレーと、

上記各マイクロホンからの音声信号に基づいて減算形アレー法を用いて上記各マイクロホンから上記各第 2 のマイクロホンに対する方向にゼロ点を生成するカージオイドを生成してそれらに対応する複数のカージオイド信号を発生する減算形アレー手段と、

上記複数のカージオイド信号のうちより高い信号対雑音電力比を有する複数のカージオイド信号を検出して選択し、選択した複数のカージオイド信号を加算し、加算結果のカージオイド信号を出力する評価選択手段と、

上記評価選択手段から出力されるカージオイド信号に基づいて音声認識を行って音声認識結果を出力する音声認識手段とを備えたことを特徴とする音声認識装置。

【請求項 2】

上記評価選択手段と上記音声認識手段との間に挿入して設けられ、上記評価選択手段から出力されるカージオイド信号においてスペクトルサブトラクション法を用いて雑音を除去した後、上記音声認識手段に出力する雑音除去手段をさらに備えたことを特徴とする請求項 1 記載の音声認識装置。

【請求項 3】

上記角錐は三角錐であることを特徴とする請求項 1 又は 2 記載の音声認識装置。

【請求項 4】

上記角錐は正三角錐であることを特徴とする請求項 1 又は 2 記載の音声認識装置。

【請求項 5】

上記正三角錐の底面の 3 つの頂点において、3 つの第 2 のマイクロホンを設けたことを特徴とする請求項 4 記載の音声認識装置。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、例えば所定の雑音環境下で音声認識率を向上させるために好適なマイクロホンアレーと、それを用いた音声認識装置に関する。

【背景技術】

【0002】

例えば、特許文献 1 において、話者の方向又は位置を推定して音声認識率を向上させることができる音声認識装置が開示されている。

【0003】

この従来例に係る音声認識装置では、複数のマイクロホンを所定の間隔で並置してなるマイクロホンアレーを備えた音声認識装置において、方向推定部は、各マイクロホンから出力される電気信号に基づいてマイクロホンアレーで受信される少なくとも 1 つの音源の方位角を推定し、ビームフォーミング部は、各マイクロホンから出力される電気信号に基づいて推定された少なくとも 1 つの音源の方位角の方向に対応する少なくとも 1 つのビーム信号を生成する。次いで、音源判定部は各ビーム信号に基づいて音声の HMM と雑音 HMM とを用いて各ビーム信号が音声であるか非音声であるかを判定し、音声認識部 17 は音声であると判定されたときに、当該ビーム信号に対して音声認識を行って音声認識結果を出力する。

【0004】

【特許文献 1】特開 2002-091469 号公報。

【特許文献 2】特開 2003-044092 号公報。

【特許文献 3】特開平 11-327593 号公報。

【非特許文献 1】S. E. Boll, "Suppression of Acoustic Noise in Speech Using Spect

10

20

30

40

50

ral Subtraction", IEEE Transaction on Acoustic Speech and Signal Processing, Vol . ASSP-27, pp.113-120, April 1979.

【発明の開示】

【発明が解決しようとする課題】

【0005】

しかしながら、マイクロホンアレーを、例えば工場などの大きな雑音を発生する現場において用いて音声認識する場合、当該雑音により音声認識率が大幅に低下するという問題点があった。

【0006】

本発明の目的は以上の問題点を解決し、例えば工場などの大きな雑音を発生する現場において信号対雑音電力比（以下、SNRという。）を従来技術に比較して向上させた音声信号を得ることができるマイクロホンアレーと、それを用いて音声認識することにより従来技術に比較して大きな音声認識率を得ることができる音声認識装置とを提供することにある。

【課題を解決するための手段】

【0007】

本発明に係る音声認識装置は、

角錐の各頂点のうち上部頂点において、放射主軸が話者の口元に実質的に向くように設けられた第1のマイクロホンと、上記角錐の底面の少なくとも2つの頂点において、放射主軸が話者の口元方向に実質的に平行となるように設けられた複数の第2のマイクロホンとを備えたマイクロホンアレーと、

上記各マイクロホンからの音声信号に基づいて減算形アレー法を用いて上記各マイクロホンから上記各第2のマイクロホンに対する方向にゼロ点を生成するカージオイドを生成してそれらに対応する複数のカージオイド信号を発生する減算形アレー手段と、

上記複数のカージオイド信号のうちより高い信号対雑音電力比を有する複数のカージオイド信号を検出して選択し、選択した複数のカージオイド信号を加算し、加算結果のカージオイド信号を出力する評価選択手段と、

上記評価選択手段から出力されるカージオイド信号に基づいて音声認識を行って音声認識結果を出力する音声認識手段とを備えたことを特徴とする。

【0008】

上記音声認識装置において、上記評価選択手段と上記音声認識手段との間に挿入して設けられ、上記評価選択手段から出力されるカージオイド信号においてスペクトルサブトラクション法を用いて雑音を除去した後、上記音声認識手段に出力する雑音除去手段をさらに備えたことを特徴とする。

【0009】

また、上記音声認識装置において、上記角錐は三角錐又は正三角錐であることを特徴とする。

【0010】

さらに、上記音声認識装置において、上記正三角錐の底面の3つの頂点において、3つの第2のマイクロホンを設けたことを特徴とする。

【発明の効果】

【0011】

本発明に係るマイクロホンアレーによれば、少なくとも3つのマイクロホンを用いて話者の音声を集めることにより、従来技術に比較して向上させた音声信号を得ることができる。また、当該マイクロホンアレーを用いて音声信号を収録し、減算形アレー法を用いて複数のカージオイド信号を生成し、そのうちのより高いSNRを有する複数のカージオイド信号を加算し、その加算信号に対してスペクトルサブトラクション法を用いて雑音除去をした後音声認識することにより、例えば工場などの大きな雑音を発生する現場において音声認識率を従来技術に比較して向上させることができる。

【発明を実施するための最良の形態】

【0012】

以下、本発明に係る実施形態について図面を参照して説明する。なお、以下の各実施形態において、同様の構成要素については同一の符号を付している。

【0013】

図1は本発明の一実施形態に係るマイクロホンアレー10の配置を示す斜視図であり、図2は図1のマイクロホンアレー10を備えたマイクロホン筐体11を示す側面図であり、図3は図2のマイクロホン筐体11を示す正面図である。本実施形態に係るマイクロホンアレー10は、各無指向性マイクロホン1, 2, 3, 4を正三角錐の各頂点の位置に設けたことを特徴としている。図1及びそれ以降の配置図において、XYZの3次元座標系で各マイクロホン1, 2, 3, 4の位置を示しており、各マイクロホン1～4の配置座標は以下の通りである。

【0014】

(A) マイクロホン1のXYZ座標 = $(0, 0, 0)$; 正三角錐の上部頂点の位置であって、XYZの3次元座標系の原点に位置する。

(B) マイクロホン2のXYZ座標 = $(0, \sqrt{(6)d/3}, \sqrt{(3)d/3})$; 正三角錐の底面の一頂点の位置であって、XY平面の0度及びXZ平面の55度の方位に位置する。

(C) マイクロホン3のXYZ座標 = $(d/2, \sqrt{(6)d/3}, -\sqrt{(3)d/6})$; 正三角錐の底面の一頂点の位置であって、XY平面の30度及びXZ平面の110度の方位に位置する。

(D) マイクロホン4のXYZ座標 = $(-d/2, \sqrt{(6)d/3}, -\sqrt{(3)d/6})$; 正三角錐の底面の一頂点の位置であって、XY平面の300度及びXZ平面の110度の方位に位置する。

【0015】

図1において、XYZの3次元座標系は、話者の口元先端部5から話者音声6が矢印6の音声放射方向がY軸方向となるように配置されている。すなわち、Y軸方向は話者の口元先端部5からの法線ベクトルの方向であり、X軸方向は水平方向であり、Z軸方向は垂直方向である。そして、各マイクロホン1～4の放射主軸（放射指向特性の主方向の軸であり、マイクロホンの筒形状の軸に対応する。）は音声放射方向6を向くようにかつそれに対して実質的に平行となるように配置されている。

【0016】

図2及び図3において、4つのマイクロホン1～4からなるマイクロホンアレー10はマイクロホン筐体11に収容され、当該マイクロホン筐体11は話者のヘッドホンセットのフレキシブルアーム12の先端部に取り付けられている。当該マイクロホン筐体11を正面から見ると、図3から明らかなように、4つのマイクロホン1～4の放射面が見えるが、上部頂点のマイクロホン1のみが話者の口元により近接するような配置となっている。また、マイクロホン1～4のうちの各隣接する2つのマイクロホン間の間隔は、図2から明らかなように、10mmに設定されている。これについては、音声信号をサンプリングするサンプリング周波数16kHzに対して、標本化定理と同様にマイクロホン間に許容される最大距離が音速÷サンプリング周波数により、 $340000/16000 = 21.25$ mmとなり、さらに詳細後述するカージオイドを用いた信号処理を行う場合は折り返しという減少を防ぐために、さらに半分にする必要があり、許容される最大距離が10.625mmとなることから決定されている。その中で、最大の位相差及び角度差が得られる形状として正三角錐を採用している。

【0017】

図1乃至図3の実施形態においては、正三角錐の各頂点にマイクロホン1～4を配置しているが、本発明はこれに限らず、正三角錐は三角錐、多角錐又は角錐でもよく、底面の各頂点に配置されるマイクロホン2～4は少なくとも2つのみ配置してもよい。また、多角錐の場合は、底面の各頂点に配置されるマイクロホンの数は少なくとも2つ、すなわち複数配置すればよい。

【0018】

図4は図1のマイクロホンアレー10を用いた音声認識装置の構成を示すブロック図である。

【0019】

図4において、マイクロホン1に入力された音声は音声信号に変換された後、低周波増幅器21及びA/D変換器26を介してデジタル信号S1に変換され、減算器41, 42, 43に入力される。また、マイクロホン2に入力された音声は音声信号に変換された後、低周波増幅器22及びA/D変換器27を介してデジタル音声信号S2に変換され、次いで、当該デジタル音声信号S2は、遅延器31を介して減算器41に入力され、減算器44に入力され、遅延器35を介して減算器45に入力され、遅延器38を介して減算器48に入力され、減算器49に入力される。マイクロホン3に入力された音声は音声信号に変換された後、低周波増幅器23及びA/D変換器28を介してデジタル音声信号S3に変換され、次いで、当該デジタル音声信号S3は、遅延器32を介して減算器42に入力され、遅延器34を介して減算器44に入力され、減算器45に入力され、減算器46に入力され、遅延器37を介して減算器47に入力される。マイクロホン4に入力された音声は音声信号に変換された後、低周波増幅器24及びA/D変換器29を介してデジタル音声信号S4に変換され、次いで、当該デジタル音声信号S4は、遅延器33を介して減算器43に入力され、遅延器36を介して減算器46に入力され、減算器47に入力され、減算器48に入力され、遅延器39を介して減算器49に入力される。なお、各遅延器31～39は、隣接するマイクロホン間の音声信号の到来時間差を補償するために本実施形態では、29.4マイクロ秒の遅延量を有する。

【0020】

遅延形アレー回路30は、9個の遅延器31～39と、9個の減算器41～49を備えて構成され、公知の減算形アレー法（例えば、非特許文献2参照。）を用いて、図5及び図6を参照して説明するように、雑音方向に対してゼロ点（指向性利得の最小点）を生成する所定のカーゴイドC1～C9を発生する。

【0021】

減算器41はデジタル音声信号S1から遅延されたデジタル音声信号S2を減算し減算結果のカーゴイド音声信号SC1（後述するカーゴイドC1の指向特性で検出されたデジタル音声信号である。）を信号評価及び選択回路50に出力する。減算器42はデジタル音声信号S1から遅延されたデジタル音声信号S3を減算し減算結果のカーゴイド音声信号SC2（後述するカーゴイドC2の指向特性で検出されたデジタル音声信号である。）を信号評価及び選択回路50に出力する。減算器43はデジタル音声信号S1から遅延されたデジタル音声信号S3を減算し減算結果のカーゴイド音声信号SC3（後述するカーゴイドC3の指向特性で検出されたデジタル音声信号である。）を信号評価及び選択回路50に出力する。

【0022】

減算器44はデジタル音声信号S2から遅延されたデジタル音声信号S3を減算し減算結果のカーゴイド音声信号SC4（後述するカーゴイドC1の指向特性で検出されたデジタル音声信号である。）を信号評価及び選択回路50に出力する。減算器45はデジタル音声信号S3から遅延されたデジタル音声信号S2を減算し減算結果のカーゴイド音声信号SC5（後述するカーゴイドC5の指向特性で検出されたデジタル音声信号である。）を信号評価及び選択回路50に出力する。減算器46はデジタル音声信号S3から遅延されたデジタル音声信号S4を減算し減算結果のカーゴイド音声信号SC6（後述するカーゴイドC6の指向特性で検出されたデジタル音声信号である。）を信号評価及び選択回路50に出力する。減算器47はデジタル音声信号S4から遅延されたデジタル音声信号S3を減算し減算結果のカーゴイド音声信号SC7（後述するカーゴイドC7の指向特性で検出されたデジタル音声信号である。）を信号評価及び選択回路50に出力する。減算器48はデジタル音声信号S4から遅延されたデジタル音声信号S2を減算し減算結果のカーゴイド音声信号SC8（後述するカ

ージオイド C 8 の指向特性で検出されたデジタル音声信号である。) を信号評価及び選択回路 5 0 に出力する。減算器 4 9 はデジタル音声信号 S 2 から遅延されたデジタル音声信号 S 4 を減算し減算結果のカージオイド音声信号 S C 9 (後述するカージオイド C 9 の指向特性で検出されたデジタル音声信号である。) を信号評価及び選択回路 5 0 に出力する。

【0023】

信号評価及び選択回路 5 0 は、入力される 9 つのカージオイド音声信号 S C 1 ~ S C 9 について、V A D (Voice Activity Detection) 機能を用いて、音声区間と雑音区間とを検出し、それに基づいて S N R を計算し、S N R が大きい上位 2 つ (変形例では、3 つ) のカージオイド音声信号を選択し、選択したカージオイド音声信号を加算して加算結果のカージオイド音声信号を雑音除去回路 5 1 に出力する。ここで、V A D 機能は、以下の条件で音声区間を検出する。

10

(1) 所定のしきい値以上の信号レベルを有すること。

(2) 所定のパワーレベル以上離れたカージオイド信号が存在しないこと。これは、口元方向に対応する 3 つのカージオイド信号と、顔平面方向に対応するカージオイド信号について、口元方向からの音声に対しては、前者 3 つのカージオイド音声信号はもちろん、後者 6 つのカージオイド音声信号も少しパワーが上がるのに対して、口元以外の方向からの音声信号は、1 つ以上のカージオイドの死角に入る可能性が高く、9 つの中で相対的にパワー差が開く傾向にあることを利用しようというものである。

(3) 音声区間として検出されたフレームの前後 5 0 0 ミリ秒を音声区間として扱う。

20

【0024】

次いで、雑音除去回路 5 1 は、入力されるカージオイド音声信号に対して、公知のスペクトルサブトラクション法 (以下、S S 法という。) を用いて音声信号中の雑音を除去し、処理後のデジタル音声信号を音声認識回路 5 2 に出力する。ここで、S S 法は周波数領域における雑音除去法として従来から用いられており、雑音が付加された音声信号のパワースペクトルから、別途推定した雑音のパワースペクトルを差し引き、そのパワースペクトルをフーリエ逆変換することで雑音を除去した音声信号を復元するものである (例えば、特許文献 3 及び非特許文献 1 参照。)。ここで、S S 法を用いた演算後のスペクトル成分 $X(f)$ は次式で表される。

【0025】

30

[数 1]

$$X^2(f) = \max \{ x(f) - \alpha N(f), \beta N(f) \} \quad (1)$$

【0026】

ここで、 α 、 β は所定の定数であって、例えば $\alpha = 2.0$ 、 $\beta = 0.001$ である。また、 $X(f)$ は雑音をスペクトル減算した結果のスペクトル成分であり、 $x(f)$ は収録音声データ (音声 + 雑音) のスペクトル成分であり、 $N(f)$ は雑音のスペクトル成分である。

【0027】

音声認識回路 5 2 は、入力されるデジタル音声信号に対して例えば所定の音声辞書又は音声モデル (例えば HMM) を用いて音声認識処理を実行して、音声認識結果のテキストデータを液晶ディスプレイ (L C D) 5 3 に表示出力し、もしくはパーソナルコンピュータなどの外部装置に出力する。

40

【0028】

次いで、図 3 の音声認識装置において形成されるカージオイド C 1 ~ C 9 について、図 5 及び図 6 を参照して以下に説明する。

【0029】

図 5 は図 4 の音声認識装置において実現される口元方位に対応する 3 つのカージオイド C 1、C 2、C 3 を示す斜視図である。図 5 において、カージオイド C 1 はデジタル音声信号 S 1 及び S 2 により形成されるものであり、マイクロホン 2 に向う方向にゼロ点を有する。また、カージオイド C 2 はデジタル音声信号 S 1 及び S 3 により形成されるも

50

のであり、マイクロホン 3 に向う方向にゼロ点を有する。さらに、カージオイド C 3 はデジタル音声信号 S 1 及び S 3 により形成されるものであり、マイクロホン 3 に向う方向にゼロ点を有する。

【0030】

図 6 は図 4 の音声認識装置において実現される顔水平方位に対応する 6 つのカージオイド C 4, C 5, C 6, C 7, C 8, C 9 を示す斜視図である。図 6 において、カージオイド C 4, C 5 はデジタル音声信号 S 2 及び S 3 により形成されるものであり、カージオイド C 4 はマイクロホン 3 に向う方向にゼロ点を有し、カージオイド C 5 はマイクロホン 2 に向う方向にゼロ点を有する。また、カージオイド C 6, C 7 はデジタル音声信号 S 3 及び S 4 により形成されるものであり、カージオイド C 6 はマイクロホン 4 に向う方向にゼロ点を有し、カージオイド C 7 はマイクロホン 3 に向う方向にゼロ点を有する。さらに、カージオイド C 8, C 9 はデジタル音声信号 S 4 及び S 2 により形成されるものであり、カージオイド C 8 はマイクロホン 2 に向う方向にゼロ点を有し、カージオイド C 9 はマイクロホン 4 に向う方向にゼロ点を有する。

【実施例 1】

【0031】

図 7 は本発明者らによって実行された実施例 1 に係るシミュレーション実験（3 つの定常雑音 N s t 1 1, N s t 1 2, N s t 1 3）における雑音配置を示す斜視図である。図 7 において、スピーカの記号は 3 つの定常雑音 N s t 1 1, N s t 1 2, N s t 1 3 の配置位置及び放射方向を示している。ここで、定常雑音 N s t 1 1 は、XY 平面 60 度及び XZ 平面 90 度の方位から放射され、定常雑音 N s t 1 2 は、+Y 軸から原点に向う方向で放射され、定常雑音 N s t 1 3 は、XY 平面 300 度及び XZ 平面 90 度の方位から放射される。このときに図 4 の音声認識装置により評価した各カージオイド C n (n = 1, 2, …, 9) に対する SNR (C n) は以下の通りである。

【0032】

[表 1]

SNR (C 1) = 25.8 dB
SNR (C 2) = 24.4 dB
SNR (C 3) = 24.1 dB
SNR (C 4) = 15.0 dB
SNR (C 5) = 14.8 dB
SNR (C 6) = 13.6 dB
SNR (C 7) = 13.8 dB
SNR (C 8) = 14.9 dB
SNR (C 9) = 14.9 dB

【0033】

この表 1 の SNR (C n) のうち上位 m 個 (m = 2, 3, …, 9) のカージオイド音声信号を加算したときの SNR_{A D D} (T m) を以下に示す。

【0034】

[表 2]

SNR _{A D D} (T 2) = 25.3 dB
SNR _{A D D} (T 3) = 25.9 dB
SNR _{A D D} (T 4) = 23.3 dB
SNR _{A D D} (T 5) = 21.6 dB
SNR _{A D D} (T 6) = 20.7 dB
SNR _{A D D} (T 7) = 20.0 dB
SNR _{A D D} (T 8) = 19.4 dB

$$S N R_{A D D} (T 9) = 18.7 \text{ dB}$$

【0035】

表2から明らかなように、上位3個のカージオイド音声信号を加算することで最高のS N Rの音声信号を得ている。

【実施例2】

【0036】

図8は本発明者らによって実行された実施例2に係るシミュレーション実験（1つの突発性雑音N s u 2 1）における雑音配置を示す斜視図である。図8において、スピーカの記号は突発性雑音N s u 2 1の配置位置及び放射方向を示している。ここで、突発性雑音N s u 2 1は、X Y平面60度及びX Z平面90度の方位から放射される。このときに図4の音声認識装置により評価した各カージオイドC n（n=1, 2, …, 9）に対するS N R（C n）は以下の通りである。

【0037】

[表3]

$$S N R (C 1) = 5.2 \text{ dB}$$

$$S N R (C 2) = 0.8 \text{ dB}$$

$$S N R (C 3) = 16.4 \text{ dB}$$

$$S N R (C 4) = -6.5 \text{ dB}$$

$$S N R (C 5) = 1.3 \text{ dB}$$

$$S N R (C 6) = 16.0 \text{ dB}$$

$$S N R (C 7) = -8.6 \text{ dB}$$

$$S N R (C 8) = -6.6 \text{ dB}$$

$$S N R (C 9) = 1.6 \text{ dB}$$

【0038】

この表3のS N R（C n）のうち上位m個（m=2, 3, …, 9）のカージオイド音声信号を加算したときのS N R_{A D D}（T m）を以下に示す。

【0039】

[表4]

$$S N R_{A D D} (T 2) = 16.2 \text{ dB}$$

$$S N R_{A D D} (T 3) = 9.5 \text{ dB}$$

$$S N R_{A D D} (T 4) = 7.1 \text{ dB}$$

$$S N R_{A D D} (T 5) = 6.5 \text{ dB}$$

$$S N R_{A D D} (T 6) = 5.0 \text{ dB}$$

$$S N R_{A D D} (T 7) = 2.7 \text{ dB}$$

$$S N R_{A D D} (T 8) = 1.3 \text{ dB}$$

$$S N R_{A D D} (T 9) = -0.5 \text{ dB}$$

【0040】

表4から明らかなように、上位2個のカージオイド音声信号を加算することで最高のS N Rの音声信号を得ている。

【実施例3】

【0041】

図9は本発明者らによって実行された実施例3に係るシミュレーション実験（1つの突発性雑音N s u 3 1及び1つの定常雑音N s t 3 2）における雑音配置を示す斜視図である。図9において、スピーカの記号は1つの突発性雑音N s u 3 1及び1つの定常雑音N s t 3 2の配置位置及び放射方向を示している。ここで、突発性雑音N s u 3 1は、X Y

平面 60 度及び XZ 平面 90 度の方位から放射され、定常雑音 N s t 3 2 は、XY 平面 300 度及び XZ 平面 90 度の方位から放射される。このときに図 4 の音声認識装置により評価した各カージオイド C n (n = 1, 2, ..., 9) に対する S N R (C n) は以下の通りである。

【0042】

[表 5]

S N R (C 1)	= 9. 3 d B
S N R (C 2)	= 6. 4 d B
S N R (C 3)	= 9. 4 d B
S N R (C 4)	= - 1. 5 d B
S N R (C 5)	= 0. 8 d B
S N R (C 6)	= - 0. 2 d B
S N R (C 7)	= - 2. 9 d B
S N R (C 8)	= - 1. 2 d B
S N R (C 9)	= 1. 0 d B

10

【0043】

この表 5 の S N R (C n) のうち上位 m 個 (m = 2, 3, ..., 9) のカージオイド音声信号を加算したときの S N R_{A D D} (T m) を以下に示す。

20

【0044】

[表 6]

S N R _{A D D} (T 2)	= 10. 0 d B
S N R _{A D D} (T 3)	= 7. 6 d B
S N R _{A D D} (T 4)	= 7. 0 d B
S N R _{A D D} (T 5)	= 6. 4 d B
S N R _{A D D} (T 6)	= 5. 6 d B
S N R _{A D D} (T 7)	= 4. 9 d B
S N R _{A D D} (T 8)	= 4. 3 d B
S N R _{A D D} (T 9)	= - 3. 4 d B

30

【0045】

表 6 から明らかなように、上位 2 個のカージオイド音声信号を加算することで最高の S N R の音声信号を得ている。

【実施例 4】

【0046】

図 10 は本発明者らによって実行された実施例 4 に係るシミュレーション実験 (1 つの定常雑音 N s t 4 1) における雑音配置を示す斜視図である。図 10 において、スピーカの記号は 1 つの定常雑音 N s t 4 1 の配置位置及び放射方向を示している。ここで、定常雑音 N s t 4 1 は、XY 平面 30 度及び XZ 平面 90 度の方位から背景雑音レベル 90 d B A で放射される。このときに図 4 の音声認識装置により評価した各カージオイド音声信号 S C n (n = 1, 2, ..., 9) に基づいて、上位 2 個又は 3 個のカージオイド音声信号を加算したときの S N R_{A D D} (T m) を以下に示す。

40

[表 7]

S N R _{A D D} (T 2)	= 8. 0 d B
S N R _{A D D} (T 3)	= 7. 3 d B

【0047】

50

ここで、より高い S N R を有する上位 2 個のカーゴイド音声信号を加算したときに、S S 法を用いる図 4 の雑音除去回路 5 1 を用いた場合の S N R_{s s} を以下に示す。

【0048】

[表 8]

S N R_{s s} ($\alpha = 1.0$; $\beta = 0.001$) = 8.0 d B

S N R_{s s} ($\alpha = 2.0$; $\beta = 0.001$) = 10.3 d B

【0049】

表 8 から明らかなように、S S 法を用いた雑音除去回路 5 1 を用いることにより大幅に S N R が改善されていることがわかる。

10

【0050】

以上の実施例 1 乃至 4 において、定常雑音は例えばベルトコンベヤーなどから発生するホワイトノイズであり、突発性雑音は例えば金属材料のパンチングなどから発生する突発雑音である。

【実施例 5】

【0051】

実施例 5 において、本発明者らは、各種の過酷な雑音環境下（出願人の犬山工場にて）で、以下の実験条件下で異なる 100 個の数字 4 桁を話者により読み上げ、そのときの、音声認識率を測定した。

20

【0052】

[表 9]

(A) 音声認識ソフトウェア：日本電気製音声認識テストアプリケーション

(B) 認識辞書：数字認識辞書 4 桁

(C) 使用マイクロホン：

(C 1) 日本電気製ヘッドセットマイクロホン（比較例 1；単一性音声用マイクロホンと、無指向性雑音用マイクロホンとを備えて構成される）

(C 2) ゼンハイザー製 H M D - 2 5 型マイクロホン（比較例 2）

(C 3) 本実施形態に係るマイクロホンアレー（実施形態；図 1 乃至図 3 に示すように、1 個の無指向性音声用マイクロホン 1 と、3 個の無指向性雑音用マイクロホン 2, 3, 4 とを備えて構成される。）

30

【0053】

図 1 1 は本発明者らによって実行された実施例 5 に係る雑音下音声認識実験の実験結果（音声認識率）を示す表である。図 1 1 から明らかなように、騒音レベルが 80 d B A という非常に過酷な雑音環境下において、本実施形態に係るマイクロホンアレー 10 を用いて收音することにより、従来技術に比較して大きく改善された S N R を有する音声信号を得ることができる。

【0054】

また、実施例 1 乃至 4 の結果から明らかなように、本実施形態に係るマイクロホンアレー 10 を用いて收音しかつ本実施形態に係る図 4 の音声認識装置を用いて音声認識することにより音声認識率を大幅に向上させることができる。

40

【0055】

以上の実施形態においては、減算形アレー法と S S 法とを併用しているが、本発明はこれに限らず、前者のみを用いて信号処理した後、音声認識してもよい。

【産業上の利用可能性】

【0056】

以上詳述したように、本発明に係るマイクロホンアレーによれば、少なくとも 3 つのマイクロホンを用いて話者の音声を集めることにより、従来技術に比較して向上させた音

50

声信号を得ることができる。また、当該マイクロホンアレーを用いて音声信号を収録し、減算形アレー法を用いて複数のカージオイド信号を生成し、そのうちのより高いS N Rを有する複数のカージオイド信号を加算し、その加算信号に対してS S法を用いて雑音除去をした後音声認識することにより、例えば工場などの大きな雑音を発生する現場において音声認識率を従来技術に比較して向上させることができる。

【図面の簡単な説明】

【0057】

【図1】本発明の一実施形態に係るマイクロホンアレー10の配置を示す斜視図である。

【図2】図1のマイクロホンアレー10を備えたマイクロホン筐体11を示す側面図である。

10

【図3】図2のマイクロホン筐体11を示す正面図である。

【図4】図1のマイクロホンアレー10を用いた音声認識装置の構成を示すブロック図である。

【図5】図4の音声認識装置において実現される口元方位に対応する3つのカージオイドC1, C2, C3を示す斜視図である。

【図6】図4の音声認識装置において実現される顔水平方位に対応する6つのカージオイドC4, C5, C6, C7, C8, C9を示す斜視図である。

【図7】本発明者らによって実行された実施例1に係るシミュレーション実験（3つの定常雑音N s t 1 1, N s t 1 2, N s t 1 3）における雑音配置を示す斜視図である。

【図8】本発明者らによって実行された実施例2に係るシミュレーション実験（1つの突発性雑音N s u 2 1）における雑音配置を示す斜視図である。

20

【図9】本発明者らによって実行された実施例3に係るシミュレーション実験（1つの突発性雑音N s u 3 1及び1つの定常雑音N s t 3 2）における雑音配置を示す斜視図である。

【図10】本発明者らによって実行された実施例4に係るシミュレーション実験（1つの定常雑音N s t 4 1）における雑音配置を示す斜視図である。

【図11】本発明者らによって実行された実施例5に係る雑音下音声認識実験の実験結果（音声認識率）を示す表である。

【符号の説明】

【0058】

30

1, 2, 3, 4…マイクロホン、

5…口元先端部、

6…音声放射方向、

10…マイクロホンアレー、

11…マイクロホン筐体、

12…フレキシブルアーム、

21, 22, 23, 24…低周波増幅器、

26, 27, 28, 29…A/D変換器、

30…遅延形アレー回路、

31, 32, 33, 34, 35, 36, 37, 38, 39…遅延器、

40

41, 42, 43, 44, 45, 46, 47, 48, 49…減算器、

50…信号評価及び選択回路、

51…雑音除去回路、

52…音声認識回路、

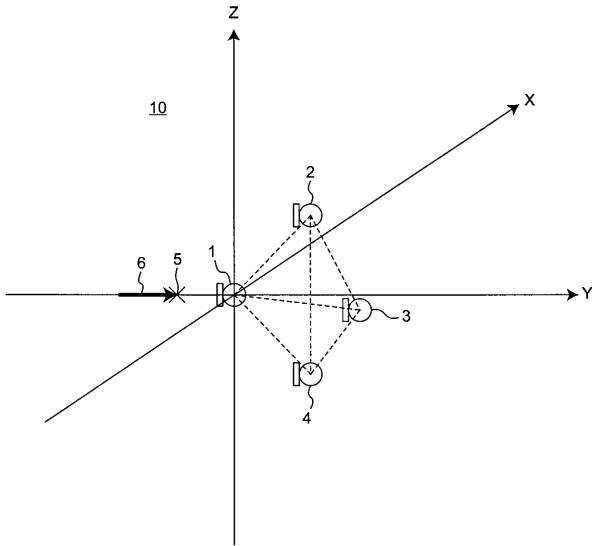
53…液晶ディスプレイ（LCD）、

C1, C2, C3, C4, C5, C6, C7, C8, C9…カージオイド、

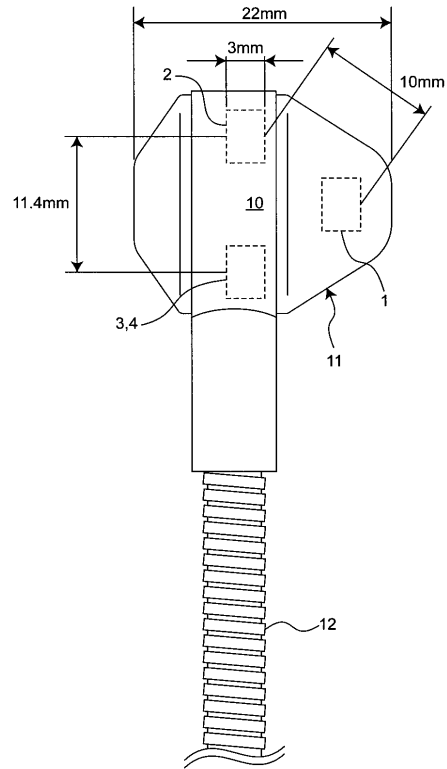
N s t 1 1, N s t 1 2, N s t 1 3, N s t 3 2, N s t 4 1…定常雑音、

N s u 2 1, N s u 3 1…突発性雑音。

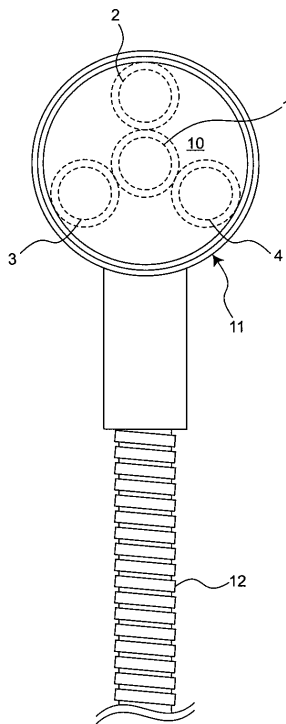
【図 1】



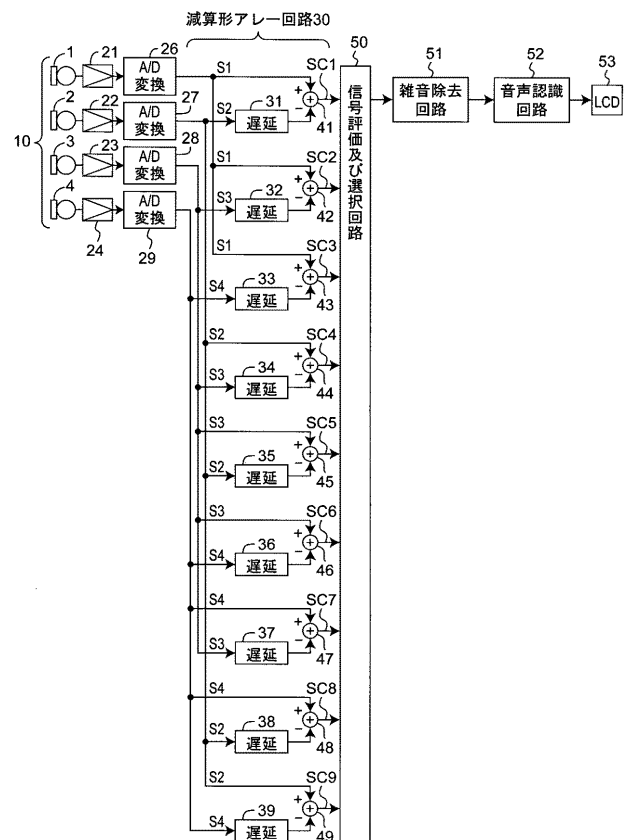
【図 2】



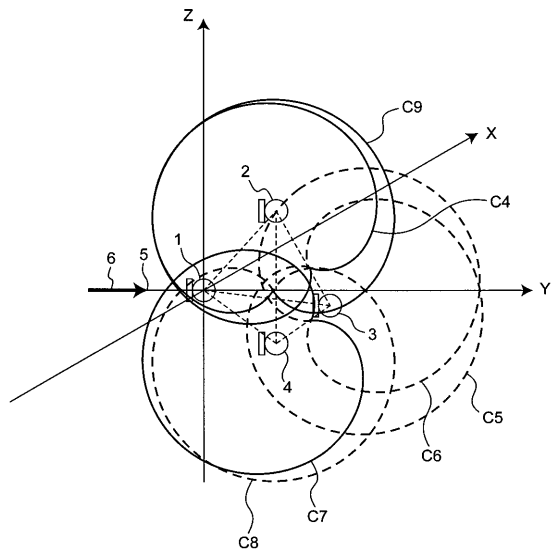
【図 3】



【図 4】

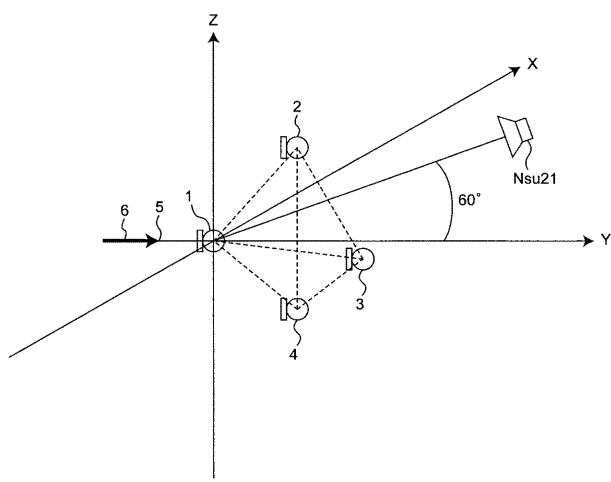


【图 6】



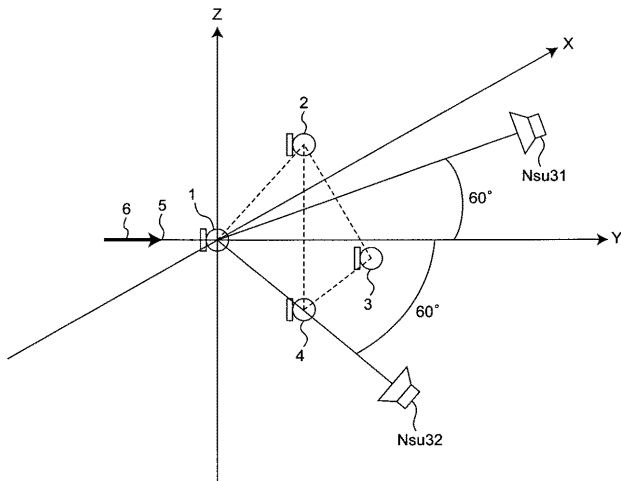
【図 8】

实施例2



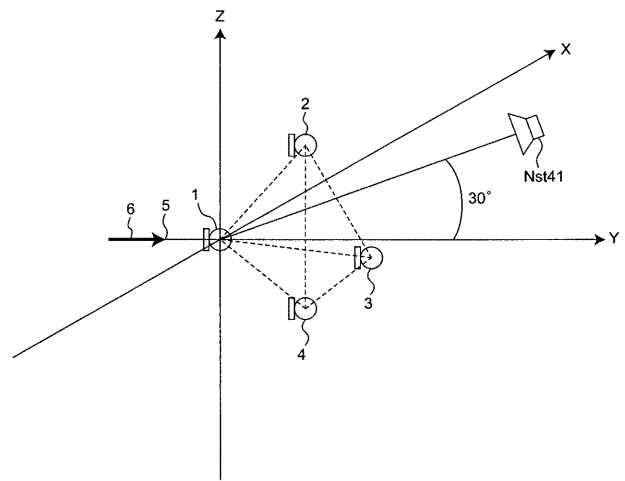
【図 9】

実施例3



【図 10】

実施例4



【図 11】

実施例5

騒音レベル	25dB	60dB	70dB	80dB
比較例1	98/100 98%	97/100 97%	96/100 96%	76/100 76%
比較例2	97/100 97%	97/100 97%	95/100 95%	88/100 88%
実施形態	97/100 97%	96/100 96%	96/100 96%	91/100 91%

フロントページの続き

(72)発明者 西浦 敬信

滋賀県草津市野路東一丁目1番1号 立命館大学びわこ・くさつキャンパス情報理工学部内

Fターム(参考) 5D015 EE05

5D018 BB23 BB25