

マイクロホンアレイを用いた音声入力 インタフェース

Speech Input Interface with Microphone Array

あらまし

本稿では、マイクロホンアレイによる音源位置検出処理と目的音強調処理を用いた音声入力インタフェースについて説明する。

このインタフェースは、雑音のある環境において、(a) あらかじめ定めた位置で話者が発声しているか否かを検出し、(b) 発声しているときにのみ、その音声を明瞭に受音し、発声していない場合には無音にするという特長を持つ。今回は、3個のマイクロホンから成るアレイを用いて、音源位置検出処理による音声スイッチと目的音強調処理による指向性マイクロホン構築し、これらを結合して必要な機能を実現した。実験の結果、音源位置検出処理により話者が発声しているかどうかを検出し、発声している場合に目的音強調処理により音声強調を行い、発声していない場合には無音にできることを確認した。

Abstract

This paper describes a speech input interface designed for speech recognition in a noisy environment.

In a noisy environment, the interface detects and enhances the speech of a person speaking at a specified location. When the person stops speaking, the interface reduces its output level to zero. The interface uses a voice switch based on a sound source position detection algorithm and a directional microphone system consisting of a three-microphone array and a signal enhancement algorithm. The interface has performed well in various tests we have conducted.



松尾直司 (まつお なおし)

1989年九州工業大学大学院工学研究科電気工学専攻修士課程了。同年(株)富士通研究所入社。以来音声・オーディオ信号処理およびマルチモーダルインタフェースの研究に従事。パーソナルシステム研究所ヒューマンインタフェース研究部



北川博紀 (きたがわ ひろき)

1992年大阪大学工学部電子制御機械工学科卒。同年(株)富士通研究所入社。以来マルチモーダルインタフェースの研究に従事。パーソナルシステム研究所ヒューマンインタフェース研究部



長田茂美 (ながた しげみ)

1977年東京工業大学大学院総合理工学研究科システム科学専攻修士課程了。同年(株)富士通研究所入社。以来コンピュータビジョン、ニューロコンピューティング、センサフュージョンの研究に従事。その間1988年人工知能学会優秀論文賞受賞。パーソナルシステム研究所ヒューマンインタフェース研究部

まえがき

現在、パーソナルコンピュータやカーナビゲーションなどにおいて、音声入力インタフェースと音声認識システムを組み合わせることにより、音声入力を用いた操作が可能になりつつある。このインタフェースには、接話型または非接話型マイクロホンが用いられているが、ユーザが使いやすいようにするためには、非接話型マイクロホンを用いることが望ましい。

一般に、この非接話型マイクロホンには、雑音のある環境における認識率低下を防ぐため、指向性マイクロホンが用いられている。しかし、従来の指向性マイクロホンでは、話者が発声していない場合でも雑音を受音するため、あるレベル以上の雑音があるとき、それを音声と誤って認識処理を行い、パーソナルコンピュータなどが誤動作を行うことがある。したがって、話者が発声しているときのみ、その音声を明瞭に受音し、雑音をできるだけ受音しない音声入力インタフェースが必要になる。マイクロホンアレイ⁽¹⁾は、このようなインタフェースを実現するのに有効な受音装置である。

今回は、マイクロホンアレイを用いた音源位置検出処理による音声スイッチと目的音強調処理による指向性マイクロホンを組み合わせて、あらかじめ定めた位置にいる話者が発声しているときのみ音声の強調を行い、それ以外のときには無音にするインタフェースの開発を行った。

構 成

● 概要

マイクロホンアレイとは、複数のマイクロホンから構成される受音装置であり、各マイクロホンで入力した信号を用いて、つぎに示す三つの空間的な信号処理を行うことができる。

- ・音源位置検出処理
- ・目的音強調処理
- ・雑音抑制処理

今回は、図-1に示す3個のマイクロホンで構成されるアレイを用いて、音源位置検出処理と目的音強調処理を組み合わせて音声入力インタフェースを開発した。

処理ブロック図と話者位置を表す座標 (x_i, y_j) を図-2に示す。あらかじめ複数の話者位置を定め、各マイクロホンを用いて入力した信号を基に音源位置検出処理を行い、そのいずれかで話者が発声しているか否かを検出して音声スイッチを操作する。また、検出した話者位置に合わせた目的音強調処理によりマイクロホンアレイに指向性を持た

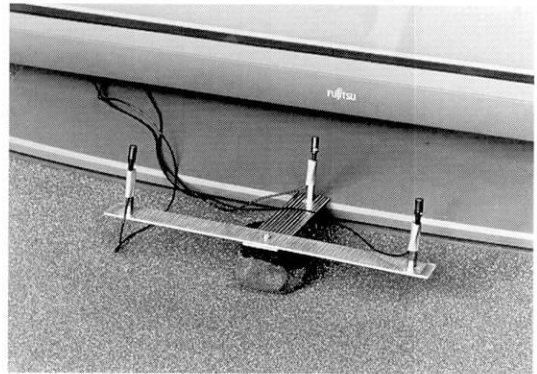


図-1 マイクロホンアレイ

Fig.1-Microphone array.

せ、目的音である音声を強調する。

● 音源位置検出処理

図-2に示すように、3個のマイクロホンからの入力信号の相関係数⁽²⁾を計算し、その値からあらかじめ定めた座標に話者がいて発声しているか否かを検出する。ここで、相関係数とは、二つの信号の相関関係を示し、-1以上1以下の値となり、独立な信号の相関係数は0となる。各処理の内容は、つぎのとおりである。

(1) 相関係数計算

3個のマイクロホン mic a, mic b, mic cからの入力信号 $a(t_g)$, $b(t_g)$, $c(t_g)$ の相関係数 $R_{ac}(k)$ と $R_{bc}(k)$ を式(1)と式(2)に示すように計算する。

$$R_{ac}(k) = \frac{\sum_{g=0} \{a(t_g - k) c(t_g)\}}{\sqrt{\sum_{g=0} a(t_g - k)^2} \sqrt{\sum_{g=0} c(t_g)^2}} \quad \dots\dots(1)$$

$$k = -n_{ac}, \dots, 0, \dots, n_{ac}$$

$$R_{bc}(k) = \frac{\sum_{g=0} \{b(t_g - k) c(t_g)\}}{\sqrt{\sum_{g=0} b(t_g - k)^2} \sqrt{\sum_{g=0} c(t_g)^2}} \quad \dots\dots(2)$$

$$k = -n_{bc}, \dots, 0, \dots, n_{bc}$$

t_g はサンプリング時刻を示す。 n_{ac} と n_{bc} は、式(3)と式(4)に示すように、マイクロホンの間隔 l_{ac} と l_{bc} により決まる。

$$n_{ac} = \frac{l_{ac}}{v} F_s \quad \dots\dots(3)$$

$$n_{bc} = \frac{l_{bc}}{v} F_s \quad \dots\dots(4)$$

v は音速、 F_s はサンプリング周波数を示す。

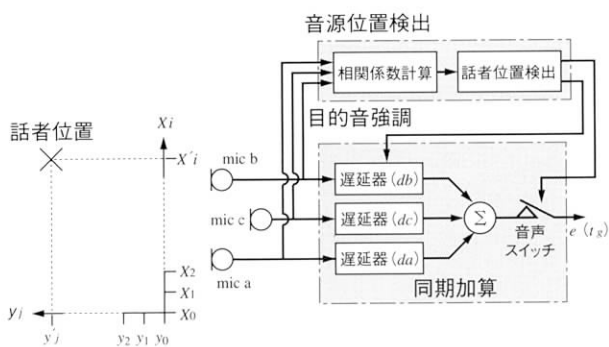


図-2 処理ブロック図
Fig.2-Block diagram.

(2) 話者位置検出

式(5)に示すように、図-2の座標であらかじめ定めた位置 (x_i, y_j) における相関係数 $R_{ac}(k)$ と $R_{bc}(k)$ の積 $r(x_i, y_j)$ を計算する。

$$r(x_i, y_j) = R_{ac}(k_{ac}) R_{bc}(k_{bc}) \quad \dots\dots(5)$$

ただし

$$k_{ac} = \frac{F_s l_{ac} \sin \left\{ \tan^{-1} \left(\frac{y_j - y_c}{x_i - x_c} \right) - \theta_{ac} \right\}}{v}$$

$$k_{bc} = \frac{F_s l_{bc} \sin \left\{ \tan^{-1} \left(\frac{y_j - y_c}{x_c - x_i} \right) - \theta_{bc} \right\}}{v}$$

(x_c, y_c) は mic c の座標、 θ_{ac} は mic a と mic c を結ぶ線に垂直な線と x 軸のなす角、 θ_{bc} は mic b と mic c を結ぶ線に垂直な線と x 軸のなす角を示す。

この相関係数の積に関する閾値^{しきい値}を定めておき、 $r(x_i, y_j)$ の値が閾値以上の場合、この座標に話者がいて発声していると判定する。

● 目的音強調処理

図-2 に示すように、3 個のマイクロホンの入力信号 $a(t_g)$ 、 $b(t_g)$ 、 $c(t_g)$ について、話者位置 (x_i, y_j) に合わせた同期加算⁽¹⁾ を行い、音源位置検出処理で座標 (x_i, y_j) に話者がいて発声していると判定された場合、音声スイッチをオンにして同期加算の結果を出力する。各処理の内容は、つぎのとおりである。

(1) 同期加算

座標 (x_i, y_j) の話者から 3 個のマイクロホン mic a、mic b、mic c までの各々の距離が異なるため、目的音である音声⁽¹⁾ が各マイクロホンに到達するまでの時間に差が生じる。このとき、式(6)に示すように、この時間差を示すサンプル数 da 、 db 、 dc を式(5)の k_{ac} と k_{bc} より計算

し、このサンプル数を用いた遅延器により同相化を行う。一方、座標 (x_i, y_j) 以外の音源からの信号、すなわち雑音は、同じ遅延器により同相化されない。したがって、式(7)に示すように加算することにより目的音を相対的に強調することができる。これにより、高い感度の位置を話者位置に合わせた指向性マイクロホンを実現する。

$$\left. \begin{aligned} da &= k_{ac} \\ db &= k_{bc} \\ dc &= 0 \end{aligned} \right\} \quad \dots\dots(6)$$

$$e(t_g) = a(t_g - da) + b(t_g - db) + c(t_g - dc) \quad \dots\dots(7)$$

(2) 音声スイッチ

座標 (x_i, y_j) における相関係数の積 $r(x_i, y_j)$ が閾値より大きく、この座標に話者がいて発声していると判定された場合にのみ、式(7)の同期加算の結果 $e(t_g)$ を出力する。

実 験

● 実験方法

開発を行った音声入力インタフェースの特性を調べるために、音源位置検出処理による相関係数の積と、目的音強調処理による指向性パターンを測定する実験を行った。また、雑音がある場合に、音源位置検出処理と目的音強調処理を組み合わせて、目的音である音声がある場合にのみそれを強調し、それ以外ときには無音にできることを確認する実験を行った。

今回は、話者位置と各マイクロホンの座標を次のように定めた。ただし、 $x_{i+1} - x_i = y_{j+1} - y_j = 10 \text{ cm}$ とした。

- ・話者位置： (x_5, y_{15})
- ・mic a の位置： (x_9, y_1)
- ・mic b の位置： (x_{11}, y_1)
- ・mic c の位置： (x_{10}, y_0)

また、サンプリング周波数を $F_s = 12.0 \text{ kHz}$ とした。

これらの値より、式(6)の遅延サンプル数はつぎの値になる。

- ・ $da = 4$ 、 $db = 2$ 、 $dc = 0$

また、雑音については、つぎのように定めた。

- ・音源位置： (x_{15}, y_{15})
- ・信号の種類：白色雑音

● 実験結果

(1) 音源位置検出処理による相関係数の積

3 個のマイクロホンで入力した信号を基に式(5)の相関係数の積 $r(x_i, y_j)$ を計算すると、図-3 に示すようにな

る。この図において、座標 (x_5, y_{15}) 付近で $r(x_i, y_j)$ が最大値になっており、相関係数の積 $r(x_i, y_j)$ を用いて、話者の存在を正しく検出できることを確認した。

(2) 目的音強調処理による指向特性

式 (7) の $e(t_g)$ のパワーより求めた指向性パターンを 図-4 示す。この図において、座標 (x_5, y_{15}) 付近で感度が

高くなっており、同期加算を用いて音声を約 3 dB 強調できることを確認した。

(3) 音源位置検出処理と目的音強調処理の組合せによる音声の検出

結果を 図-5 に示す。(a) が目的音である音声、(b) が mic c からの入力信号、(c) が相関係数の積 $r(x_5, y_{15})$ 、

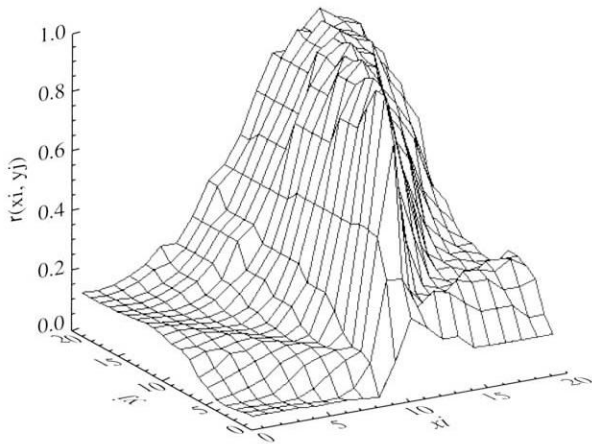


図-3 相関係数の積 $r(x_i, y_j)$

Fig.3-Products of correlation coefficients $r(x_i, y_j)$.

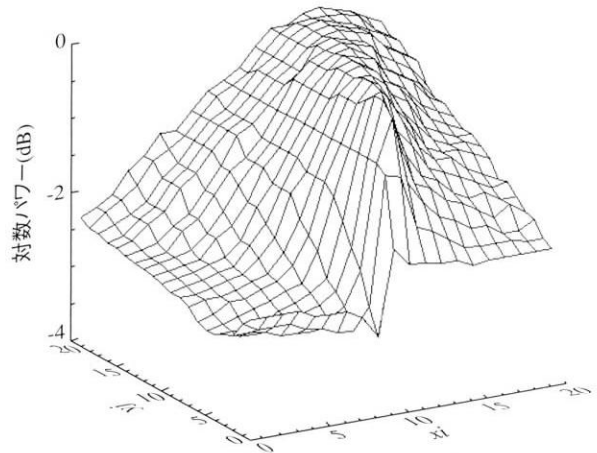


図-4 指向性パターン

Fig.4-Directivity pattern.

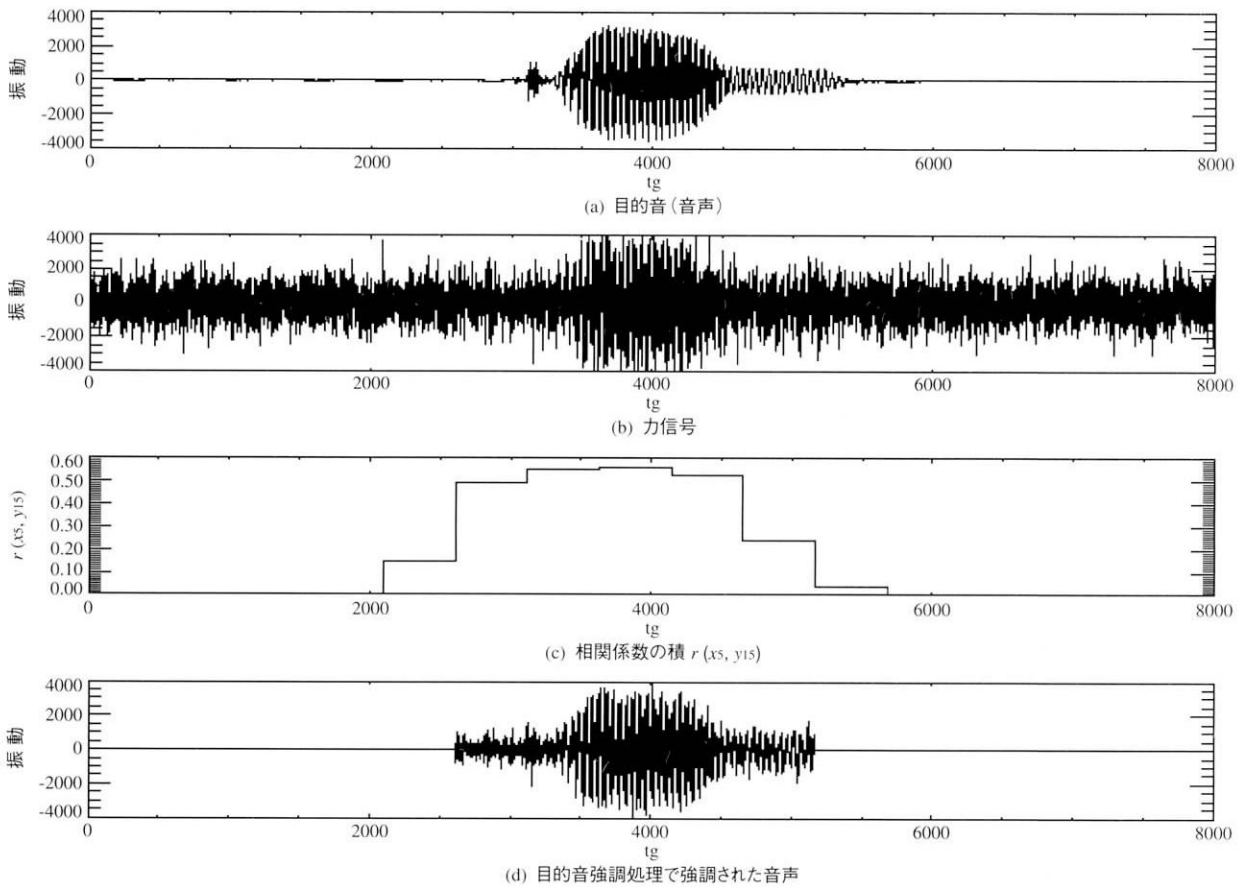


図-5 音声の検出とその強調

Fig.5-Speech detection and enhancement.

(d) は音源位置検出処理で検出され目的音強調処理で強調された音声を示す。この結果から、雑音の混ざった入力信号から音声を検出して強調できること、および音声がない場合には無音にできることを確認した。

む す び

今回、マイクロホンアレイを用いた音声入力インタフェースの開発を行った。雑音のある環境において、あらかじめ定めた位置に話者がいるときにのみ入力した音声

を強調し、それ以外は無音にする機能を実現した。また実験の結果、これらの機能を実現できることを確認した。

参考文献

- (1) 金田豊：マイクロホンアレイによる指向性制御，日本音響学会誌，51，5，pp.390-394 (1995)。
- (2) (社)日本音響学会：音響用語辞典，初版，東京，コロナ社，pp.331 (1988)。

