

E-042

## 音源定位機能を持つロボットシステムの試作

## Trial Development of Robot Julie Having Sound Source Localization

中澤竜次† 真坂直樹† 齋藤竜一† 畑岡 信夫†

Ryuuji Nakazawa Naoki Masaka Ryuuiti Saitou Nobuo Hataoka

## 1 まえがき

最近、人間型ロボット(ヒューマノイドロボット)の研究が活発化している。その中で、我々が日常的に使用している音声を使ってロボットとの意思疎通を図ることが試みられており、音声研究者にとっては究極的な夢となっている[1]。我々は話者の位置を同定(音源定位)して、話者の方向にロボットを振る向かせることを目的に研究を進めている。今回は、2つのマイクロフォンを使用した場合の音源定位処理の詳細と評価実験結果に関して報告する。

## 2 音源定位概要

音源定位処理、音声認識、全体の制御を行う制御システムはPCと無線LAN、及びロボット(Julie)の3つで構成されている[2]。ロボットは、市販のネットタンサーを使用した。

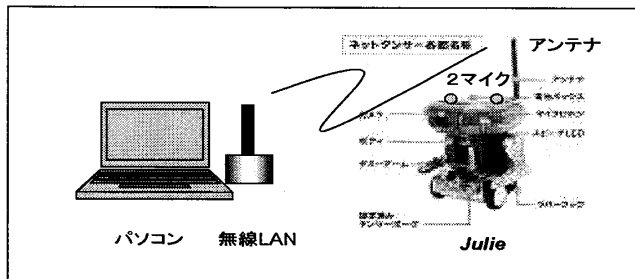


図1 システム構成(パソコン、無線LAN、Julie)

以下、音源定位の原理を述べる。

2つのマイクを使ってマイク1に入る音声とマイク2に入る音声を収録し、サンプル  $i$  ずれた相互相関関数を算出する。

$$\sum_{j=0}^N \text{wave}(j) \times \text{wave}1(j+i) \quad (1)$$

$i$  を変化させて求めた相互相関関数の最大値から  $\text{wave}, \text{wave}1$  の配列のズレを求める。

求めたズレ  $a$  より  $b$  秒のズレをだす。

$$b = a \times 1.0 / f$$

ただし、 $f = 22.050 \text{kHz}$

$b$  秒のズレより  $c$  cmのズレをだす。

$$c[\text{cm}] = b[\text{s}] \times 340[\text{m/s}] \times 100$$

マイク1から話者の口元までの高さを  $21 \text{cm}$  マイク1から話者までの距離  $90 \text{cm}$  とし高さをふまえたマイク1から話者の口元までの距離  $e$  を求める。

$$e = \sqrt{21^2 + 90^2}$$

高さをふまえたマイク1から話者の口元までの距離を図2に示す。

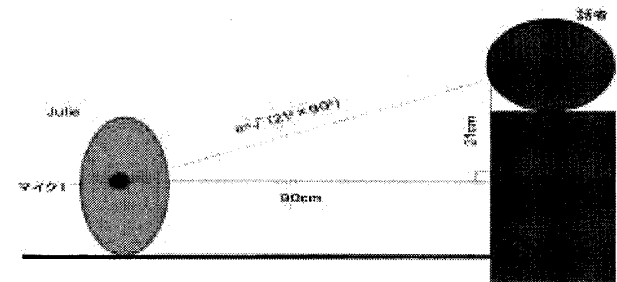


図2

高さをふまえたマイク1から話者の口元までの距離を求めたマイク1から話者の口元までの距離と  $i$  を変化させて求めた相互相関関数の最大値か求めた  $c$  cmを使い、マイク2から話者の口元までの距離  $(e-c)$  を求める。

マイク1からマイク2までの距離  $16 \text{cm}$  と求めた  $e$ ,  $(e-c)$  を使い角度  $x$  を求める。

$$\cos(x) = \frac{e^2 + 16^2 - (e-c)^2}{2 \times 16 \times e}$$

求めた  $x$ ,  $e$ , マイク1からJulieの中心までの距離  $8 \text{cm}$  を使い、Julieの中心までの距離  $h$  を求める。

† 東北工業大学

〒982-8577 仙台市太白区八木山 香澄町 35-1

$$h = \sqrt{e^2 + 8^2 - (2 \times 8 \times \cos(x))}$$

求めた  $h$ ,  $(e-c)$ , 中心からマイク 2 までの距離 8cm を使い角度  $y$  を求める。

$$\cos(y) = \frac{h^2 + 8^2 - (e-c)^2}{2 \times 8 \times h}$$

求めた角度  $y$  より振り向かせる角度  $k$  を求める。

$$k = 90 - y$$

話者とマイク 1, 2 より振り向かせる角度を算出する。マイク 1, 2 と話者との関係を図 3 に示す。

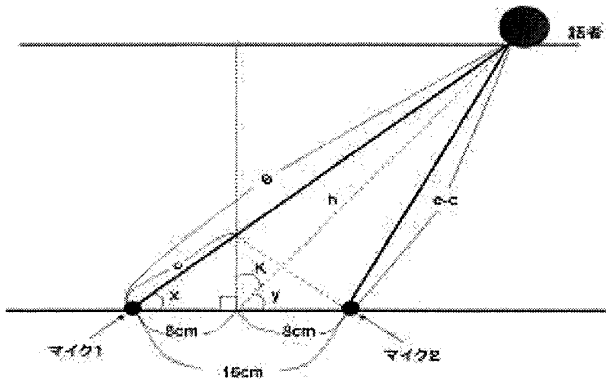


図3 マイク 1, 2 と話者との関係

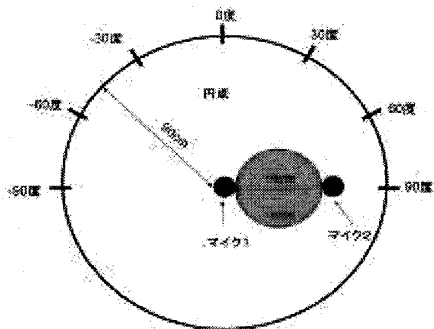


図4 評価実験方法

### 3 評価実験結果

音源定位処理をし、得た値を表 1 に示す。

また、30度の位置からマイク 1 へ発話された音声の波形を図 5、マイク 2 の音声波形を図 6 に示す。

### 4 考察

90度、-90度が理想と約15度違うのは、最大のズレがマイク 1 とマイク 2 の距離 16cm なのでこの音源定位処理で算出される最大のズレが 10 ポイント 15.42cm で 74.6 度、ズレが 11 ポイントで計算すると、16.96cm になり最大のズレ 16cm をこえてしまい計算できないため実際はサンプリングのズレ 10 と 11 の間にあるためである。0 度、30 度、60 度、-30 度、-60

度は、評価実験の音源定位率が約 88% と理想に近い結果となった。

表 1 実験結果

話者	話者の位置						
	90[度]	60[度]	30[度]	0[度]	-30[度]	-60[度]	-90[度]
A	74.6	60.3	28.9	0.0	-28.9	-60.2	-74.6
B	74.6	74.6	35.4	0.0	-35.4	-50.5	-74.6
C	74.6	60.3	35.4	0.0	-35.4	-60.2	-74.6
D	74.6	60.3	28.9	0.0	-28.9	-60.2	-74.6
E	74.6	74.6	28.9	0.0	-28.9	-60.2	-74.6

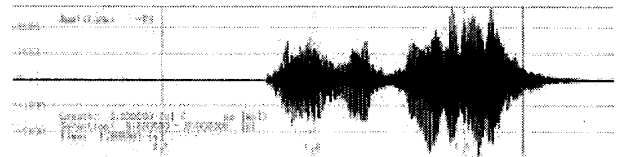


図5 30度の位置からマイク 1 へ発話された音声波形

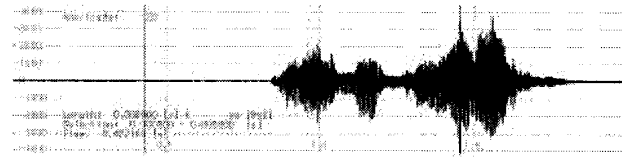


図6 30度の位置からマイク 2 へ発話された音声波形

### 5 まとめ

2つのマイクロフォンを使用して、音源定位方式を検討し、実際にロボット Julie に搭載して、正確に音源定位をすることを確かめた。

### 6 今後の予定

現在 2 個のマイクロフォンを 3 個に増設し、現在の音源定位の有効角度を 180 度から 360 度まで対応させ、距離の指定がなくても正確な音源定位処理が行なえるようにする。

### 参考文献

- [1] 奥乃博、他：ロボット聴覚のための情報統合の現状と課題、計測と制御、Vol. 46, pp. 415-419 (2007)
- [2] 菅野博文：音源定位による振り向きロボットの製作、平成 20 年東北地区若手研究者研究発表会講演資料 pp.117-118(2008 年)