

近接-遠隔マイクコンビネーションに基づく話者方向の逐次的同定

藤原 真志 沼波 幸 川端 豪

関西学院大学理工学部 〒669-1337 三田市学園 2-1

E-mail: {fujihara, bhz80890}@ksc.kwansei.ac.jp, kaw@kwansei.ac.jp

あらまし ロボットに話者の方向を同定させる簡易な手法として、ロボットに取り付けられた(遠隔)マイクと発話者の手元の(近接)マイクを併用する方法を前報にて提案した。この手法においては単一の発話に対して方向同定を行っていたが、走行など動作中のロボットに対して複数の呼びかけがあった場合、はじめの呼びかけでだまかな方向同定を行い動作を開始し、その後の追加音声で方向同定の精度を上げる逐次的方式が、ロボットの反応性の観点で有利と考えられる。本報告では、近接-遠隔マイクを併用する手法において、話者が連続的に発話する単語を用いて逐次的に方向を同定する。連続した10単語の発話後98%の方向同定率が得られた。

キーワード 方向同定, マイクコンビネーション, ロボット

A Sequential Method of Speaker Direction Detection based on the Close-and-Distance Microphone Combination

Masashi FUJIHARA, Tsukasa NUNAMI and Takeshi KAWABATA

School of Science and Technology, Kwansei Gakuin University, 2-1 Gakuen, Sanda City, 669-1337 JAPAN

E-mail: {fujihara, bhz80890}@ksc.kwansei.ac.jp, kaw@kwansei.ac.jp

Abstract A simple speaker-direction detection method based on the close-and-distance microphone combination was proposed in our last report. In this paper, we extend the approach to handle speaker's sequential utterances. The robot roughly estimates the speaker direction using the first word and begins to move. Quick response is promising for such interactive robots. After that, the direction is finely detected using more utterances. Experiments show that the proposed method effectively improves the direction detection accuracy from 69% to 98% after ten word utterances.

Keyword direction detection, microphone combination, robot

1. はじめに

近年、ロボットの活動の場は多岐に広がりつつあり、近い将来には我々の日常生活にも関わりを持つと思われる。このような場面において、人間とロボットが音声で自然な対話を行うことが必要になってくると思われるが、音声で自然な対話を行うためには話者方向の同定が重要な要素の一つとして挙げられる。例えば、ロボットは話しかけられた人の方向にまず顔を向け、話を聞く態度を表明するのが自然である。このように、円滑なコミュニケーションにおいて方向同定は不可欠な要素であると考えられる。

コンピュータに音声の方向を同定させる機構を考えると、人間の聴覚機能を無視することはできない。人間は音源方向を見つめる際に、代表的なキューとして両耳間の時間差(ITD: Interaural time difference)を利用していることが知られている[1]。近年の音源位置推定における研究でも、このキューを基に、音源の場所を同定する方法が研究されている。例えば、Omologo

らは2つのマイクに入力される信号の相互相関関数を計算することで、2つのマイク間の時間差を導き出し、話者方向の同定を行っている[2]。さらに、Hangらは2つのマイク間の時間差では困難であった前後間の同定について、3つのマイクを三角形に配置する事で、前後間の時間差も考慮し、360度における話者方向を一意に同定することを可能としている[3]。また、この他にも数多くの研究がなされている[4]。

さらに、方向の同定手法としては時間差に基づくもの以外にも、複数のマイクでアレイを形成し、信号処理を用いて音源の方向を同定するものがある[5,6,7]。ロボットとの音声コミュニケーションを行うのに対し、ロボット自身の発する機械雑音が大きな問題となる。駆動音、モーター音、ギア音等の機械雑音は外から入ってくる音声に比べ、大きなレベルで入力される。このような状況下では音声認識はもとより、方向同定を行うことは極めて難しくなる[8]。そこで、我々は前報にて、発話者の手元のマイク(近接マイク)とロボット

に取り付けたマイク(遠隔マイク,Distance-mic)の入力を使用し、精度よく話者の方向を同定する手法を提案した[9].

これらの多くの研究は、単一の発話に対して話者方向の同定を行っている。しかし、人間とロボットが共存するような日常生活においては、人間が呼びかけた際に、ロボットが間違った方向へ歩き出したため修正するために呼び直す(図1)、より早く移動するように何度もロボットに呼びかける、ロボットの動きを見て面白くなって何度も呼びかけるなど、ロボットに話者方向を同定させる際に、複数回ロボットへ向かって呼びかける場合も多々あると思われる。こういった場合においては、人間が一言目に発した単語について、ロボットは反応し、大まかに話者方向を同定する。そして、続いて発せられる言葉を逐次的に同定していくことで、より精密な話者方向を得ていくことが、より有用な方法であると考えられる。

そこで、本稿においては近接-遠隔マイクコンビネーションを用い、話者方向の「逐次的な」同定を行う。話者方向が連続的に発話する単語を用い、逐次的に話者方向を同定することで、さらなる精度の向上を試みる。

2. ロボットのアプリケーションイメージ

日常生活において、人間と共存するロボットにはどのような特徴が求められるだろうか。少なくとも、「安全性」、「面白さ」、「有益性」等の観点は必ず考慮すべきである。

「安全性」については、ロボットのデザインや素材、あるいは人間やロボット自身を傷つけない動作、ロボットが間違った、さらにはロボットが何らかの危険な動きをした場合など容易に制止することが出来る機能等が求められる。「面白さ」については、ロボットのデザイン性、挙動、そして人間が行動を起こした際のロ

ボットの反応等が考えられる。「有益性」については、人間のほしい情報を正確に伝える、場合によってはパートナーである人間の有益であろう情報をロボットが自主的に発信するとさらに良いかもしれない。また他にも、掃除や片付けなど人間の代わりに仕事をしてくれることが考えられる。

本稿の対象とするペットロボットやトイロボットの場合には、特に「安全性」や「面白さ」の要素が重要と考えられる。例えば、ロボットは人間に呼ばれたら反応し、人間の元へやっていく。その際に、ロボットが間違った方向へ向かおうとしていれば、人間がロボットを呼び直すかもしれない。また、人間の元へ移動した後には、人間とロボットで少しおしゃべりができたり、場合によっては、その日の天気など情報が得ることができたりすると、使っている人間も楽しくなるかもしれない。

このように、話者方向の同定はロボットとのコミュニケーションに欠かせない要素であるが、日常生活での使用を前提とするペットロボットやトイロボットの場合には、この機能の実装にあまりコストをかけられないという事態もある。コストを掛けられる場合には、様々なセンサやカメラ等を設置することも考えられるが、コストがあまり掛けられない場合においては、マイク数個にて話者方向を同定することは有意であると考える。

家庭内にペットロボットが歩き回っている状況で、ペットに声を掛けるようにロボットに呼びかけるシーンは想像に難しくない。このとき、一度の呼びかけのみで、ロボットに行動を促すこともあるが、逆にある行動をさせるためには何度も呼びかける必要がある場合もあると考えられる。例えば、部屋の中で横たわっているロボットを想像してみる。部屋の中で寝転んでいるロボットに対して、人間が呼びかけ、人間の元までロボットを来させようとしている状況があるとする。この時、人間はロボットに一言話しかけることで、発話者の方向を同定させ、行動スケジュールを完全に決定するというよりも、人間のはじめの発話に対してロボットは反応し、体を起こし、首を左右に振り周りを見渡す。さらに呼びかけると、大まかな方向を同定し、歩き出す。場合によっては、ロボットは話者方向を間違え、誤った方向に歩き出すかもしれない。それを見た人間は、さらに呼びかけることによってロボットの行動を抑制し、自分の位置まで導くといった全体のプロセスは、ペットとのコミュニケーションとしても自然であり、本来の目的を果たすものである。

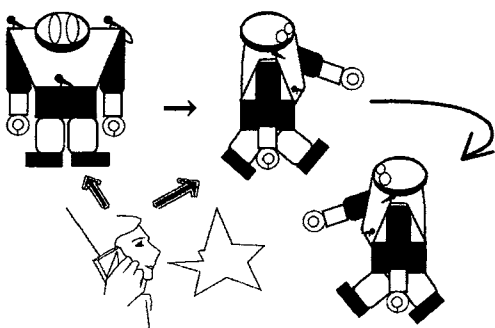


図 1: 話者方向の逐次同定のイメージ図

3. 話者方向の同定の基本的な考え

人間は音源方向を見つける際に、代表的なキューとして、両耳に入ってくる音の時間差(ITD)が挙げられる。ロボットの場合には、左右の肩にマイクを取り付け、それぞれのマイクに入力される信号について相互相関関数を計算し、ピークを探し、対応するズレ量を時間差とし、この時間差を基に左右方向の同定を行う方法が考えられる(図 2)。この方法は相互相関関数(cross correlation function)を用いることから、CC 法と呼ぶ。しかし、ロボットの駆動に伴って機械雑音が発生する場合には、ロボットに取り付けたマイクへ直接機械雑音が混入する。このため、外から発声される人間の声に比べ、機械雑音が非常に大きなレベルになることから、ロボットの駆動中における話者方向の同定は極めて困難になる。

そこで、前報ではロボットに取り付けられたマイク(遠隔マイク)の他に、発話者の手元にマイク(近接マイク)を用意することとし、この近接マイクのスペクトル要素を、遠隔マイクの中から探し出し、時間差を求めることによって、機械雑音下であっても、ある程度の精度をもった方向同定を可能とした。この時の信号処理のダイアグラムを図 3 に示す。近接マイクとロボットの左右の肩のマイクとの相互相関関数をそれぞれ計算し、ピークを探し、対応する時間差に基づく計算をすることにより話者方向を同定することとした。

さらに、2つのマイク間の時間差では原理的に前後の判定が難しいことから、Huang らと同様に左右だけではなく前後についても立体配置となるように3つのマイクを配置することで、全方位について話者方向を一意に同定することとした。

この結果、ロボットの機械雑音下において遠隔マイクのみによる CC 法では話者方向の同定精度は約 40% だったのに対し、近接マイクを併用することによって 70% 程度の精度を得ることができた。

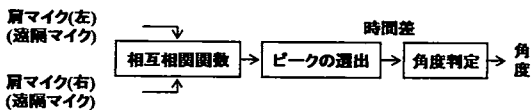


図 2: CC 法による話者方向の同定方法

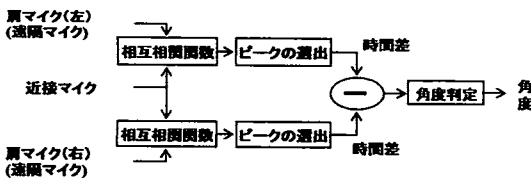


図 3: 近接マイクを利用した話者方向同定の方法

4. 話者方向の逐次的同定

前節において、機械雑音環境下であってもある程度の精度を持つ話者方向の同定方法を確立した。しかし、前節では単一の単語の発声を用いて同定を行っている。実際に、家庭内にペットロボットがいる場合に、人間がペットロボットを呼ぶ際には、複数回の呼びかけがあった際には、はじめの一言目でロボットは人間の言葉に反応し、大まかに方向を同定する。続いて言葉が発せられた場合には、前の音声入力と共に、逐次的に話者方向を同定することで、ロボットの反応性の点からも有意であると思われる。そこで、本節においては、近接-遠隔マイクコンビネーションを用いた、話者方向の逐次的同定の方法について述べる。

図 4 に連続的に単語の発話があった際に、逐次的に話者方向同定を行う手順の流れ図を示す。ロボットに対して、一言目の呼びかけに対しては、単純にマイク間の相互相関関数から時間差を求め、角度判定を行う。しかし、その後連続して発話がある場合においては、過去の発話における相関値を利用することで、過去の情報を利用し、同定を行う。

また、逐次的に話者方向の同定を行う際の、相互相関関数 $R_k(N)$ ($k=0,1,2,\dots,N-1$, N : サンプル点数) を式 1 に示す。

$$R_k(N) = \frac{1}{N} \sum_{n=p-k+1}^{N-k} \{ pR_k(p) + x_{mb}(n) \cdot y_{mb}(n+k) \} \quad (式 1)$$

ただし、入力信号 $x_{mb}(n)$, $y_{mb}(n)$ ($n=0,1,2,\dots,N-1$, N : サンプル点数) をサンプル点数とし、 p 点までの相関値 $R_k(p)$ は、過去の計算により計算されているものとする。式 1 において k の値を変化させていき、相関値が最大となる時、すなわち 2つの入力信号が最も一致すると考えられる時の k が時間差となる。

また、ここで $p=N-1$ の時、つまり 1 点ずつの逐次同定の場合、式 1 は以下のようなになる。

$$R_k(N) = \frac{N-1}{N} R_k(N-1) + \frac{1}{N} x_{mb}(N-k) \cdot y_{mb}(N) \quad (式 2)$$

さらに、両辺を N 倍することで式 3 のような漸化式で表すことが出来る。

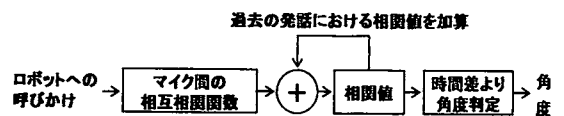


図 4: 話者方向の逐次的な同定方法

$$N \cdot R_k(N) = (N-1)R_k(N-1) + x_{TD}(N-k) \cdot y_{TD}(N) \quad (\text{式 3})$$

ここで、 $R_k(N-1)$ は事前に一度計算されている相関値となり、 $x_{TD}(N-k) \cdot y_{TD}(N)$ の計算のみが、新たに計算が必要な相関値となる。そして、これらから、 $N \cdot R_k(N)$ を得るが、データの点数 N と相関値のピークを見つける際のずれ幅 k は無関係であることから、 N の値によってピークとなる k の値は変わらない。このため、計算量削減の観点より、 $N \cdot R_k(N)$ のピークによる時間差を基に逐次的な話者方向の同定を行う。

5. 評価実験

5.1. 実験条件

評価実験にはトイロボットを使用することとした。図5のように、遠隔マイクとしてロボットの左右の肩と胸の3箇所に無指向性マイクを1つずつ取り付けた。さらに、近接マイクとして、発話者の手元にマイクを1つ用意することとした。

今回、男性2名、女性2名の計4名がロボットへ向けて単語の発声を行った。発声を行った部屋は暗騒音36dBAである防音室である。また、発声する単語は音韻の出現頻度を考慮した50単語(表1)を用いた。この50単語をロボットに対しての入射角度を12方向(0° , 30° , 60° , 90° , 120° , 150° , 180° , 210° , 240° , 270° , 300° , 330°) (図6)からそれぞれ発声している。また、本論文では、ペットロボットとの親密な会話を想定しているため、話者とロボットとの距離は約50cmとした。

人間がロボットに向かって単語を発声する際には、ロボットを動かさず機械雑音が混入しない条件下と、常時ロボットの腕を上下させることで、断続的に機械音が混入する条件下の2通りについて行った。ここで、機械雑音が混入しない条件下において“guai”と発声したときに遠隔マイクに入力される音声波形を図7に、機械雑音が混入する際に“guai”と発声したときの音声波形を図8に示す。機械雑音が混入する際において、遠隔マイクに入力される音声データの S/N 比は約10dBである。

ロボットを動かさず、機械雑音の混入が一切無い条件下の元で発声した音声データ全てを学習用データとして用い、機械雑音下の音声データを評価することとした。学習データは、12方向について、それぞれの角度における時間差の平均と分散を求めるために用いた。

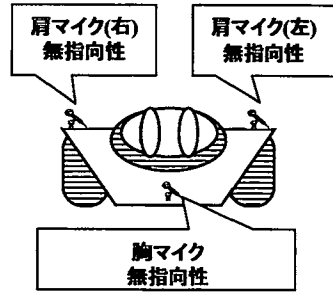


図5: ロボットに取り付けたマイクの説明

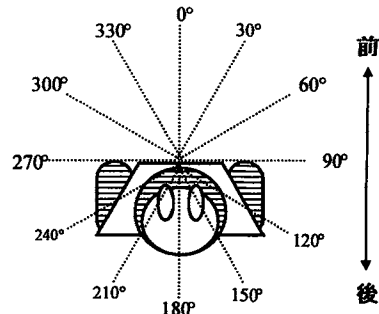


図6: ロボットに発話した入射角度



図7 機械雑音の無い場合の発話(guai)



図8 機械雑音がある場合の発話(guai)

表1: 発声した単語リスト

| | | | |
|------------|--------------|--------------|-----------|
| ikioi | iyoiyo | urayamasii | omoshiroi |
| guai | zairyoo | juuichigatsu | shuukyoo |
| juNbaN | suichoku | daijobu | daidokoro |
| chaNto | chuuoo | toriatsukau | basho |
| byooiN | hyoojuN | buNmee | poketto |
| boNyari | megane | wagamama | ukeau |
| uchiawase | kareNdaa | shuuteN | seei |
| sobieru | chichioya | chuuneN | teochi |
| deshabaru | nisemono | hahaoya | hyakusyoo |
| byooniN | furafura | hohoemu | rejaa |
| wasuremono | akachan | akegata | ajiwau |
| atarimae | apaato | aruminiumu | aNkeeto |
| iede | ichijirushii | | |

5.2. 逐次的同定による評価結果

人間が連続的に単語を発声した際、話者方向を逐次的に同定した結果を図9に示す。横軸は連続して発話を行った単語数、縦軸は正しく同定した精度を示す。図9を見ると、被験者A,B,Cの3人については、発声する単語の数が増加することにより、同定精度が著しく向上することがわかる。1単語での発話に対しては、70%前後の同定精度であったが、3単語連続で発話すると、90%前後の精度が得られ、6単語連続での発話に対しては、発話者A,B,Cの3人について約98%になり、それ以降はほぼ一定となった。これは、連続して単語を発声することで、母音の立ち上がり等の話者方向を同定する際のキューが増加したことが考えられる。

また、被験者Dに対しては発声する単語を増やすことで、同定精度の向上は若干見られたが、他の3人に比べれば、向上の度合いは低く、10単語連続で発話した場合においても、同定精度は71%となった。この被験者のデータを詳しく見ると、マイク間の時間差を求める段階で、他の被験者3人と比べ、各角度ともに非常に分散の大きいデータとなっていることがわかった。つまり、この被験者Dについては、相互相関に基づく時間差算出の過程で既に問題が起きている。今後この段階の改善も必要であることが示唆された。

被験者Aに対して、12方向についてそれぞれの角度における逐次的同定の精度を図10、図11に示す。図10については、ロボットに対して前方から発話した時の逐次的な同定精度、図11はロボットに対して後方から発話したときの逐次的な同定精度を示す。図10を見ると、前方からの発話に対しては、はじめの数回の発話に対しての各角度の精度はばらつきがあるが、発話回数を増やすことで、精度はほぼ一定となった。6単語以上連続で発話した際にはどの角度においても98%程度の精度が得られている。一方で図11を見ると、ロボットに対して前方からの発話に比べ、1単語の同定に関しては、全体的に多少精度の悪いものが多かった。ロボットに対して後方からの発話は、前方からの発話に比べ、算出される時間差について分散が多少大きくなっていることを確認した。これは、ロボットの胸に取着的けたマイクについて、前方からは直接声が入力されるのに対し、後方からだとロボットの体を回り込んでマイクに入力されるためだと考えられる。しかし、その場合においても、発話回数を重ねることで、どの角度についても飛躍的に精度は向上している。

最後に、被験者Dを除く、A,B,Cの3人に対しての1単語のみの発話について話者方向を同定した時の混同表を図10に、連続的に10単語発話したときの混同表を図11に示す。この混同表は、縦軸がロボットに示す。この混同表は、縦軸がロボットに向かって発話し

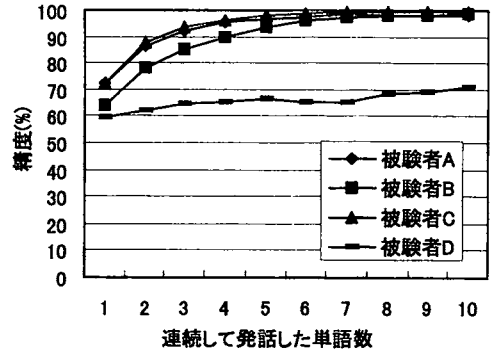


図9：逐次的な処理に基づく同定結果

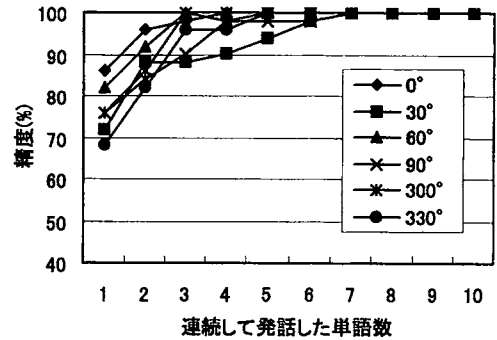


図10：各角度における逐次同定の精度 (ロボットに対して前方からの発声)

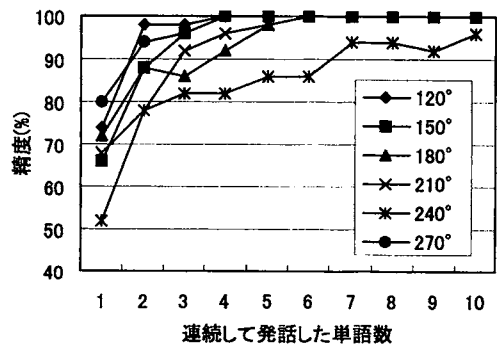


図11：各角度における逐次同定の精度 (ロボットに対して後方からの発声)

た入射角度，横軸が入力された音声から話者方向の同定を行った結果の角度を表す。このため，太枠で囲ってある箇所は正しく同定することができた回数を表す。1 単語にて話者方向の同定を行った図 12 を見ると，0°，30°，330°へ多くの誤まった同定が見られる。こういった特定の角度への引き込みは機械雑音の影響が考えられる。今回，実験にて使用したトイロボットについては人間が発声する際にロボットの左右の手を上下に動かしており，その左右の手を動かすモーターが 30°，330°の位置にあることが原因だと考えられる。しかし，10 単語連続で発声した場合の結果である図 13 を見ると，1 単語のみの場合には存在していた特定の角度への引き込みがほぼ解消された。10 発話連続で発話された場合には，98.88%と非常に高い精度での同定が可能となった。

| OUT \ IN | 0° | 30° | 60° | 90° | 120° | 150° | 180° | 210° | 240° | 270° | 300° | 330° |
|----------|-----|-----|-----|-----|------|------|------|------|------|------|------|------|
| 0° | 112 | 14 | 0 | 3 | 0 | 1 | 2 | 0 | 1 | 1 | 2 | 14 |
| 30° | 12 | 118 | 1 | 3 | 0 | 2 | 3 | 0 | 0 | 1 | 0 | 10 |
| 60° | 13 | 13 | 110 | 7 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 4 |
| 90° | 4 | 10 | 4 | 110 | 7 | 4 | 3 | 3 | 0 | 0 | 3 | 2 |
| 120° | 10 | 9 | 6 | 9 | 89 | 13 | 5 | 2 | 1 | 2 | 0 | 4 |
| 150° | 11 | 14 | 0 | 4 | 6 | 111 | 1 | 1 | 0 | 0 | 0 | 2 |
| 180° | 12 | 1 | 0 | 0 | 1 | 107 | 18 | 0 | 1 | 1 | 1 | 9 |
| 210° | 11 | 2 | 0 | 0 | 0 | 1 | 1 | 111 | 8 | 4 | 0 | 14 |
| 240° | 10 | 4 | 0 | 2 | 1 | 2 | 5 | 13 | 88 | 9 | 6 | 9 |
| 270° | 4 | 2 | 3 | 0 | 0 | 3 | 3 | 4 | 7 | 110 | 4 | 10 |
| 300° | 13 | 4 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 7 | 110 | 13 |
| 330° | 12 | 10 | 0 | 1 | 0 | 0 | 3 | 2 | 0 | 3 | 1 | 118 |

図 12：1 単語発話したときの同定結果
(太枠箇所は正しく話者方向を同定できた回数)

| OUT \ IN | 0° | 30° | 60° | 90° | 120° | 150° | 180° | 210° | 240° | 270° | 300° | 330° |
|----------|-----|-----|-----|-----|------|------|------|------|------|------|------|------|
| 0° | 144 | 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 30° | 0 | 150 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 60° | 0 | 0 | 150 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 90° | 0 | 0 | 0 | 150 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 120° | 0 | 0 | 0 | 2 | 148 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 150° | 0 | 0 | 0 | 0 | 0 | 150 | 0 | 0 | 0 | 0 | 0 | 0 |
| 180° | 0 | 0 | 0 | 0 | 0 | 0 | 149 | 1 | 0 | 0 | 0 | 0 |
| 210° | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 150 | 0 | 0 | 0 | 0 |
| 240° | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 149 | 1 | 0 | 0 |
| 270° | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 142 | 8 | 0 |
| 300° | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 150 | 0 |
| 330° | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 150 |

図 13：10 単語連続して発話したときの同定結果
(太枠箇所は正しく話者方向を同定できた回数)

6. おわりに

人間がロボットを自分の位置へ来させようと呼びかける際に，複数回の呼びかけがあった際には，はじめの一単語目で，大体の方向を同定し，連続して発声される言葉を逐次的に同定していくことがロボットの反応性の観点から望ましいと考えられる。そこで，本稿では近接-遠隔マイクコンビネーションを用いた手法を用い，話者が連続的に発話する単語について，話者方向を逐次的に同定することを試みた。結果として，発話単語数を増やすにつれて，飛躍的に精度を向上することができた。はじめの 1 単語のみの発声について，話者方向を同定したときは，69%であったのに対し，6 単語以上続けて発声することで，精度は 98%程度になり，非常に高い精度で話者方向を同定することができた。さらに，ロボットに対しての各入射角度における同定結果を見ると，はじめ数単語の同定精度については，機械雑音の影響からくる特定の角度への引き込みが見られたものが，単語数を増やすことで，機械雑音下においても極めて高い精度で同定することができることがわかった。しかし，今回一人の被験者においては，相互相関から時間差を算出する過程において，既に問題が起き，他と比べ同定精度が低いことが確認された。今後，このような場合の改善について検討していきたい。

文 献

- [1] Jeffress, L., "A place theory of sound localization". J. Comp. Physiol Psychol. 41 pp35-39 (1948)
- [2] Maurizio Omologo, P. Svaizer, "Acoustic Event Localization Using a Crosspower-Spectrum Phase Based Technique". Proc. ICASSP, vol.2, (1994)273-276
- [3] J.Huang, N. Ohnishi, and N.Sugie. "Building ears for robots: sound localization and separation. Artificial Life and Robotics", 1(4) pp157-163,1997
- [4] Nakadai, K., Matsuura, D., Okuno, H., Kitano, H., "Applying Scattering Theory to Robot Audition System. Proc". IROS-2003 pp1147-1152 (2003)
- [5] 田嶋 誠, 鈴木 基之, 他, "複数マイクロホンを用いた簡易な呼びかけ方向同定システム" 電子情報通信学会 pp19-24 (2002)
- [6] 浅野 太, 速水 悟, 松井 俊浩, "遅延和アレイを用いた音源方向同/雑音制御システムの思索とその音声認識への応用" 電子情報通信学会 pp23-30 (1996)
- [7] 菊池 隆文, 山岡 建夫, 浜田 望, "音声に調波構造を方向推定に適用したマイクロフォンアレイシステム" 電子情報通信学会 (1999)23-28
- [8] 岩沢 透, "接話マイク型同時発話音声インタフェースの開発とチャイルドケアロボットへの適用", FIT2005(第 4 回情報科学技術フォーラム), pp287-288 (2005)
- [9] 藤原真志, 川端 豪 "近接-遠隔マイクコンビネーションによる全方位型話者方向同定", 信学技法, Vol.106, No.122, SP2006-8, 13-18