

会話中における話者の方向と 聴取者の頭部方向との関連性の検討

田中 章悟^{1,a)} 荻木 禎史^{1,b)}

概要：人間の聴覚が持つ能力の一部として、人は雑音環境下においても目的とする音源のみを選択的に聴取することが可能である。この両耳聴の機能の一部をモデル化したものの一つである周波数領域両耳聴モデルは、音源方向推定とそれに基づく音源信号強調を行うことができる。しかし問題点として、複数音源が存在する際の方向誤推定が挙げられる。そこで頭部動作を含むユーザの体の動きが音源方向推定の補助に役立つことが期待される。本稿では、小型多機能センサを用いて地磁気と加速度を計測し、それらを用いて会話中の頭部動作の回転角度の推定を行い、会話中における対話者の方向と聴取者の頭部正面方向との関連性、聴取者による個人差について検討を行った。

キーワード：マルチモーダルインタフェース、ウェアラブルコンピューティング

Abstract: Estimation of sound source directions and separation of the sound sources are implemented on many products widely, and one of the applications is binaural hearing aids. However, estimation accuracy against multiple sound sources is degraded by interference between sound sources. In case of binaural hearing aids, users' body, including a head, movements would be expected to be helpful for estimation of sound source direction. In this paper, examination was carried out about relevance between speakers' direction and subject's head direction during conversation. And estimation of speakers' direction by head direction during conversation.

Keywords: Multimodal interface, Wearable computing

1. はじめに

人間の聴覚が持つ能力の一部として、人は雑音環境下においても目的とする音源のみを選択的に聴取することが可能である。これは左右耳で観測された信号に生じる時間差やレベル差から音源の方向を捉え、その方向からの音響信号に着目し、他方向からの信号を抑圧するような処理を行っているためと考えられている。この両耳聴の機能の一部をモデル化することで、音源方向推定や音源分離に応用する研究が従来より行われている。このようなモデル化の1つとして周波数領域両耳聴モデル (Frequency Domain Binaural Model : FDBM) が提案されており、音源方向推定とそれに基づく音源信号強調を行うことができる [1]。FDBM は複数の音源方向を同時に推定することもできる [2]。しかし FDBM の問題点として、複数音源が存在する

際の方向誤推定が挙げられる。これは、他の音源から到来する信号により、位相情報、振幅情報が干渉をうけることに起因する。本稿ではこの問題を解決するため、音声以外の情報を用いることを考えた。

音声以外で、知的コミュニケーションにおいて重要な役割を占めているのは、身体で表現される表情や姿勢、身振りなどの非言語情報である [3]。その中でも、無意識に現れるものは、本音を表している場合が多く、ユーザインターフェースに必要なだけでなく、計算機を通じて他の人間とのコミュニケーションを円滑に行う上でも必要である [4]。そこで補聴システムを実現する際に、頭部動作を含むユーザの身体動作が音源方向推定とともに用いることで精度改善が期待される。昨今の多軸モーションセンサの小型化・高精度化により、情報機器のみならず日常生活で利用する装身具にセンサを装着することが可能になり、身体動作の検出が容易になってきている。この技術を用いて、対話者の方向を向くという頭部運動を計測することで話者方向をおおまかに推定することができ、その方向の音を強調することが可能になると考えられる。以上のことから、補聴

¹ 熊本大学大学院自然科学研究科
Kumamoto University, 2-39-1, Kurokami, Chuo-ku,
Kumamoto, 860-8555, Japan

^{a)} shogo@hicc.cs.kumamoto-u.ac.jp

^{b)} chisaki@cs.kumamoto-u.ac.jp

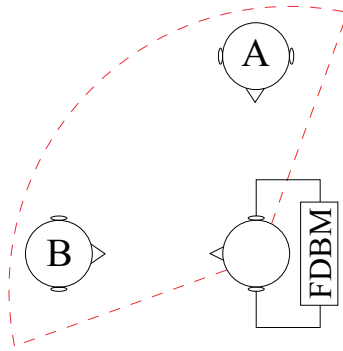


図 1 Spatial area of a conversation for sound source A and B.

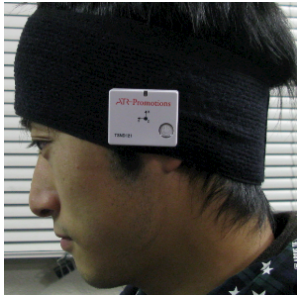


図 2 A subject wearing a small multifunction sensor TSND121 manufactured by ATR-Promotions.

システムを実現する際に、小数の小型センサから得られた頭部動作を含むユーザの身体動作の情報を用いることで、FDBM による音源方向推定の精度改善が期待される。

本稿では、センサによる頭部回転動作の角度推定を行い、会話中の被験者の頭部回転動作と対話者との関連性、被験者の頭部正面方向による対話者の方向推定についての検討を行った。

2. 加速度、地磁気による頭部回転角度検出

本節では、センサによる頭部動作検出方法について述べる。2.1 項では用いたセンサと加速度・地磁気の検出について述べる。2.2 項ではセンサから得られた加速度、地磁気データによる方位の算出方法について説明する。

2.1 小型無線多機能センサ

本稿では、ATR-Promotions 製の小型無線多機能センサ TSND121 を図 2.1 に示すように左耳の位置に装着し、3 軸の加速度と地磁気を検出した。図 2.1 に示す座標系で、加速度 Ax , Ay , Az , 地磁気 Mx , My , Mz が得られる。

2.2 加速度、地磁気による方位算出

身体動作における回転角は、身体に装着したセンサにより計測された重力加速度と地磁気から方位を算出することで得られる [5]。本稿ではこの手法を用いて頭部の回転角を検出した。

頭部が傾いているとき、図 4 のように XYZ 軸が $X'Y'Z'$

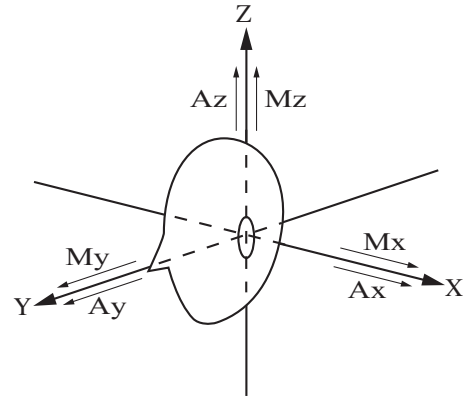


図 3 Relations of acceleration, geomagnetism and the coordinate system when a subject attached a sensor to the left ear.

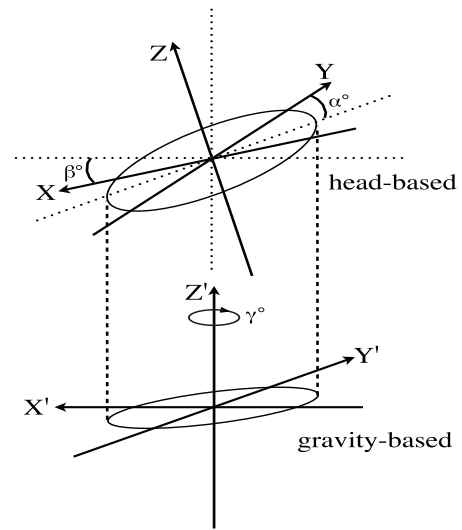


図 4 Mapping onto a gravity-based horizontal plane.

軸に対してそれぞれ α° , β° 回転し、頭部動作により、水平方向に γ° 回転したとすると、 α , β は加速度より下記の式で得られる。

$$\alpha = \arctan\left(\frac{Ay}{Az}\right) \quad (1)$$

$$\beta = \arctan\left(\frac{Ax}{\sqrt{Ay^2 + Az^2}}\right) \quad (2)$$

センサが傾斜すると、求める方位に誤差が生じてしまう。そこで、加速度により α , β を求め、下記の式 (3) を用いることでセンサから得られる地磁気の値 Mx , My , Mz から傾斜誤差を計算し、重力場を基準とした座標系における水平面に写像した地磁気の値 Mx' , My' , Mz' を算出する。

$$\begin{bmatrix} Mx' \\ My' \\ Mz' \end{bmatrix} = \begin{bmatrix} \cos\beta & 0 & \sin\beta \\ \sin\alpha\sin\beta & \cos\alpha & -\sin\alpha\cos\beta \\ -\cos\alpha\sin\beta & \sin\alpha & \cos\alpha\cos\beta \end{bmatrix} \begin{bmatrix} Mx \\ My \\ Mz \end{bmatrix} \quad (3)$$

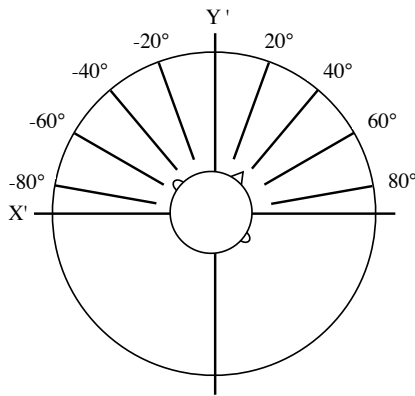


図 5 A subject and target degrees.

写像した地磁気の値を用いて式(3)を計算することで重力場を基準とした座標系における γ を算出する。この γ が頭部回転角度の情報となる。

$$\gamma = \arctan\left(\frac{My'}{Mx'}\right) [\text{deg.}] \quad (4)$$

3. 実験

本説では対話者の方向推定に関する実験と結果について述べる。3.1 項では、目標角度とセンサデータから得られる値を関連付ける関数について説明する。3.2 項では、今回対話者の方向として推定する領域について説明する。3.3 項では、会話実験の実験条件と、会話中の頭部動作の分析について説明する。

3.1 頭部動作計測

センサから得られる加速度、地磁気の値により方位を算出し、頭部の向いている角度を得るが、対話者の方向と被験者の向いている方向が必ずしも一致するとは限らない。そこで予備実験として、被験者が座っている状態で目標角度に向かって頭部を向けるという動作のデータを集めた。Y' 軸の正の方向を被験者の正面とし、目標角度を図 5 に示すように、 -80° から 80° の間で 20° 毎に設定した。被験者には目標角度を向いた後、再び正面を向くように指示をした。被験者は男性健常者 5 名で、各目標角度につき 50 サンプル、5 人で 250 サンプルのデータを集めた。

被験者は椅子の上に座り、下半身を動かさないように指示した。また頭部動作の際も、出来る限り頭部を X'Y' 軸の中心から動かさないように指示した。首の可動域は一般的に $\pm 60^\circ$ といわれ、 80° 方向に対応するために、被験者には上半身を動かすことを許可した。そのため、センサから得られるデータは、上半身の動きを含む頭部回転動作の加速度、角速度となる。

表 1 に、それぞれの目標角度に対する頭部回転角度 γ のデータの中央値を示す。

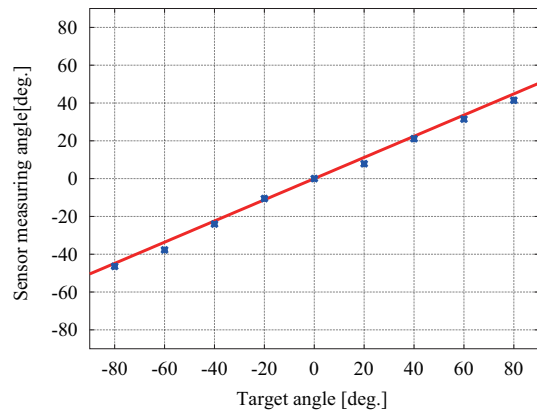


図 6 A straight line to correlate target angles with the sensor data.

表 1 より、目標角度と頭部回転角度は異なることがわかる。この違いは、視線の動きなどが原因であると考えられる。実際に FDBM に頭部回転角度を入力する場合、信号強調したい空間領域と、センサから得られた頭部回転角度が異なることが予想されるため、これらを関連付ける関数を定義した。今回は各角度のデータの中央値から、最小 2 乗法を用いて、目標角度を x 、頭部回転角度を y として回帰式を求め、これら 2 変数が対応する一次関数を求める。また、この関数の妥当性を示す方法として、決定係数 R^2 を求める。決定係数とは、得られた回帰式とデータ全体との近さを表す。 i 番目のサンプルを目標角度 x_i 、頭部回転角度 y_i で表す。 y_i のすべての値の平均値を $\bar{y}$ 、 x_i における一次関数上の値を $\hat{y}_i$ とすると、決定係数 R^2 は式(5)で求められる。

$$R^2 = 1 - \frac{(y_i - \hat{y}_i)^2}{(y_i - \bar{y})^2} \quad (5)$$

図 6 のような、 $y = 0.56x$ の関数が得られた。この関数の決定係数 R^2 は、 $R^2 = 0.99$ となり、データに近いことが分かる。

3.2 推定される領域

対話者の方向を推定する際、会話中の人の動作は頭部回転動作だけでないことを考慮する必要がある。そこでもう一つの予備実験として、座っている状態で人は会話中にどの程度動くかを調べる実験を行った。被験者の後方に、対話者から $\pm 10^\circ$ 、 $\pm 15^\circ$ の方向になるようにマーカーを配置した。その状態で会話を行い、被験者の正面にいた対話者から、被験者の頭部がどの程度振れるのかをカメラで記録した。その結果、会話中において、人は対話者から見て $\pm 15^\circ$ 以内の範囲で動くことがわかった。この結果より、

表 1 The median values of head rotational angle for each target angle.

Target angle	-80°	-60°	-40°	-20°	20°	40°	60°	80°
Median value	-54.4	-43.7	-28.0	-12.6	9.87	25.1	37.5	49.5

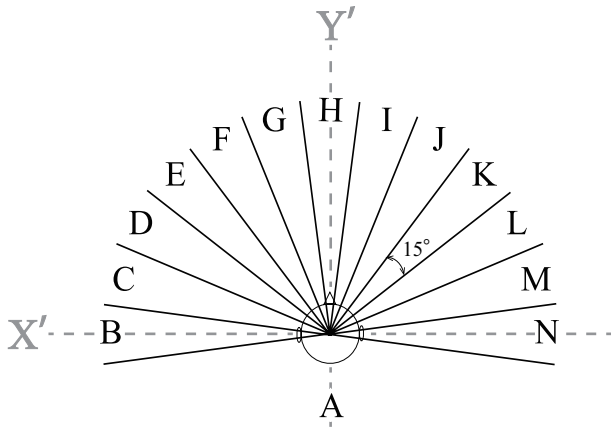


図 7 Partition of zone.

30° 以内の範囲で対話者の方向を推定することが望ましいということが示された。本稿では対話者の方向として推定する領域を、図 7 に示すように、領域 B から N まで 15° 毎に区切り、その他の後方に当たる部分を A とした。また、ある領域 P の左右境界をそれぞれ θ_{Pl} , θ_{Pr} , ある時刻 t における、センサから得られたデータを $S(t)$ とおくと、 $S(t)$ がどの領域にあるかという判定を下記の式 (6) で行った。3.3 項の会話実験では、式 (6) を基に対話者の方向推定を行う。

$$S(t) = \{Z | \theta_{Zl} < \gamma(t) < \theta_{Zr}\} \\ Z \in \{a, b, \dots, h\} \quad (6)$$

3.3 会話実験

3.3.1 実験条件

実験は会議の状況を想定し、一つのトピックを与え、そのトピックについてフリーディスカッションをする実験を行った。その際、被験者が装着したセンサから加速度、地磁気を取得し、頭部正面方向の推定を行った。条件は以下の通りである。

- 椅子に座って 3 分間会話をする
- 会話に集中し、他の行動をしない
- 正面を 0° とし、±30° 方向に対話者が 2 人いる状況と、0°, ±45° 方向に対話者が 3 人いる状況の 2 パターン行う
- 被験者は 5 人
- 対話者は 7 人の中から無作為に選ばれた 2 人もしくは 3 人である

3.3.2 実験結果

図 8 のように、対話者が ±30° にいる場合の、被験者 5 名 (I ~ V) の推定された頭部正面方向の領域を図 9 ~ 13 にそれぞれ示す。横軸が時間、縦軸が領域として、点線の枠が対話者の発話区間、実線が被験者の頭部正面方向の領域となっている。また、右側の数値は時刻 t で推定された $S(t)$ の各領域における頻度を示している。被験者 V は対話者が存在する領域 F, J やその隣の領域が多く推定され

ており、対話者の方向を向いている時間が多いことがわかる。しかし、被験者 I は領域 F にいる対話者が発話している区間が多いにも関わらず、領域 F 周辺の頻度はあまり高くない。一方、領域 J にいる発話区間の短い対話者方向の頻度が高い。これは発話者が、被験者に対してではなく、もう一人の対話者に向けて発話している時間が長いことが原因であった。また被験者 IV は、対話者の方向ではなく、正面にあたる領域 H の方向を向いていることも多かった。このように、対話者方向をあまり向いていない被験者も見られた。また、図 14 のように、対話者が 0°, ±45° にいる場合の、被験者 I ~ V の推定された頭部正面方向の領域を同様に図 15 ~ 19 に示す。被験者 I, III は発話者の領域周辺をある程度向いていることが示されているが、発話者がいる領域とは違う領域を向いている被験者も見られた。これらの結果より、2 つの会話状況において、被験者が必ずしも発話者の方向を向いているわけではないことが示された。

実験結果を定量的に評価するために、評価式を以下のよう考えた。発話者の方向に被験者が頭部正面を向けることの頻度が高ければ、音から推定した発話者の方向の補助的な役割を果たすことができると考え、発話者が発話している区間において、被験者が発話者の方向を向いているかどうかの評価を行った。正答となる領域を U 、発話区間におけるセンサデータのサンプルを i 、正答率を P とすると、評価式は式 (7) になる。 n はサンプルの個数を表す。

$$P = \frac{n(S(i) \subseteq U)}{n(i)} \times 100 \quad [\%] \quad (7)$$

正答となる領域 U を発話区間において発話者がいる領域とした場合の正答率 P の結果を表 2 に示す。表 2 より、2 つの会話状況のどちらにおいても、全ての被験者で正答率 P は 50% に満たなかった。よってこの手法では、音源方向推定の精度改善に寄与することが困難であると考えられる。

そこで、発話していない間も加味し、会話中に被験者が対話者の方向に頭部正面を向けることの頻度が高ければ、頭部正面方向より対話者の方向を推定でき、音源方向推定の補助ができるのではないかと考え、発話をしていない時間帯も含め被験者が対話者の方向を向いているかどうかの評価を行った。正答となる領域を T 、センサデータの全サンプルを j 、正答率を Q とすると、評価式は式 (8) となる。

$$Q = \frac{n(S(j) \subseteq T)}{n(j)} \times 100 \quad [\%] \quad (8)$$

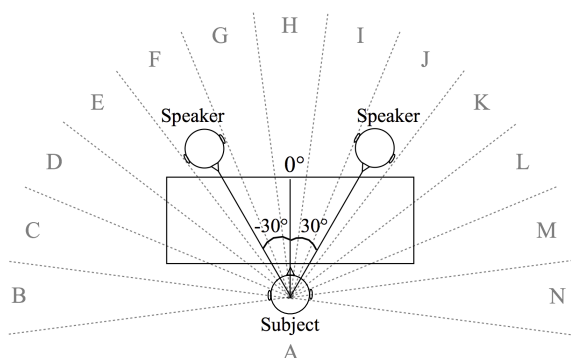


図 8 The experiment placement in case of two speakers.

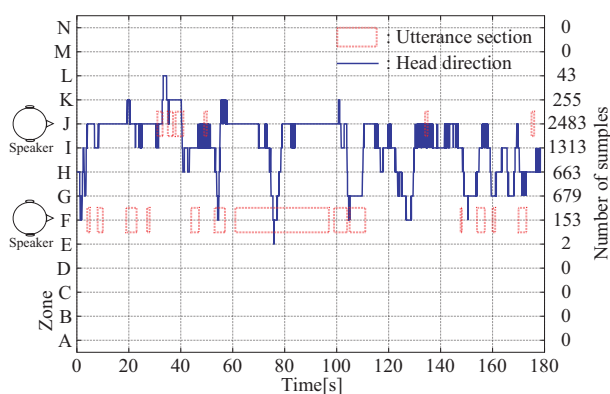


図 9 Subject I's head direction zone when the speakers are in zone F and J.

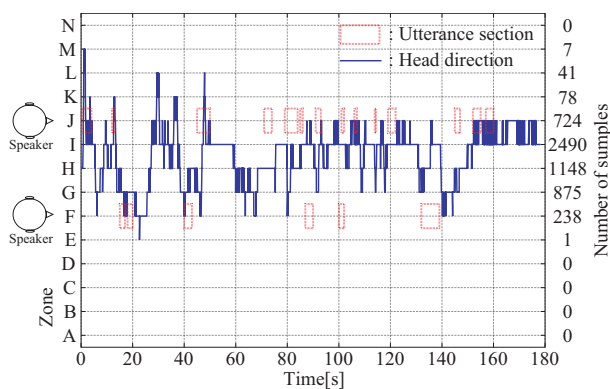


図 10 Subject II's head direction zone when the speakers are in zone F and J.

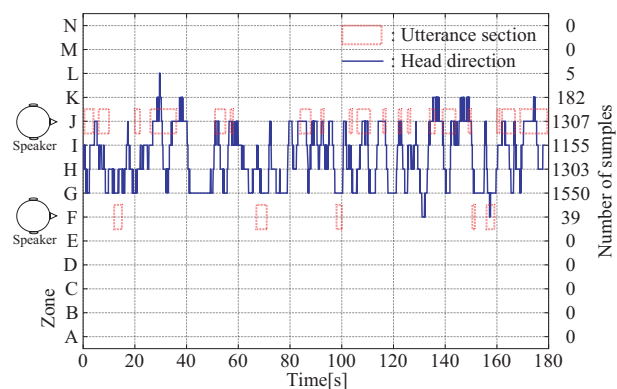


図 11 Subject III's head direction zone when the speakers are in zone F and J.

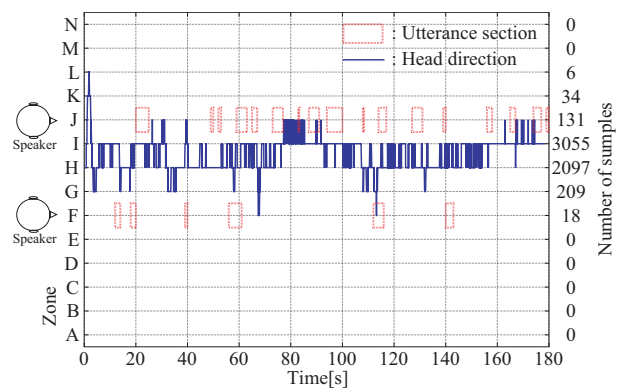


図 12 Subject IV's head direction zone when the speakers are in zone F and J.

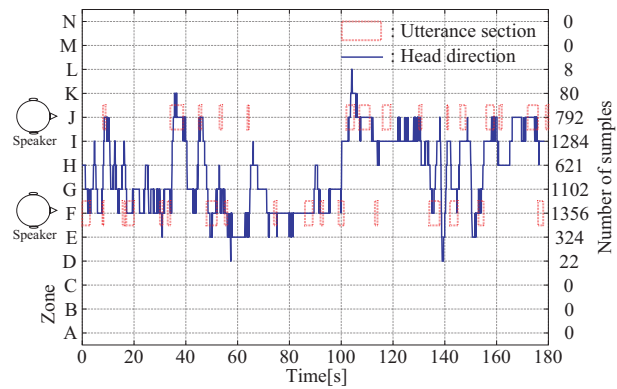


図 13 Subject V's head direction zone when the speakers are in zone F and J.

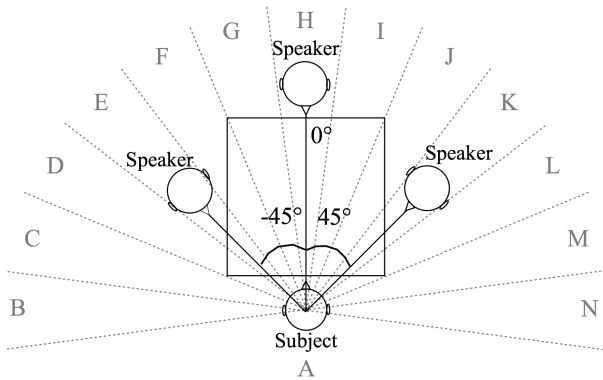


図 14 The experiment placement in case of three speakers.

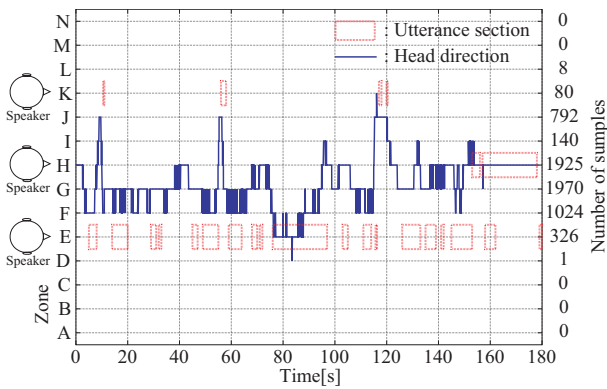


図 15 Subject I's head direction when the speakers are in zone E, H and K.

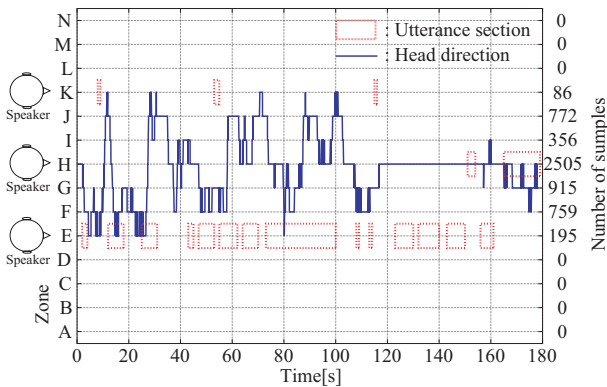


図 16 Subject II's head direction zone when the speakers are in zone E, H and K.

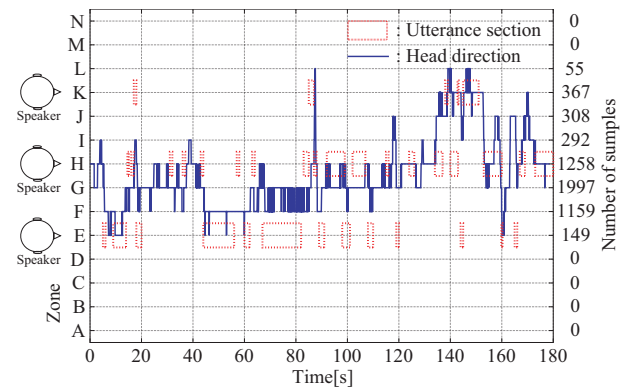


図 17 Subject III's head direction zone when the speakers are in zone E, H and K.

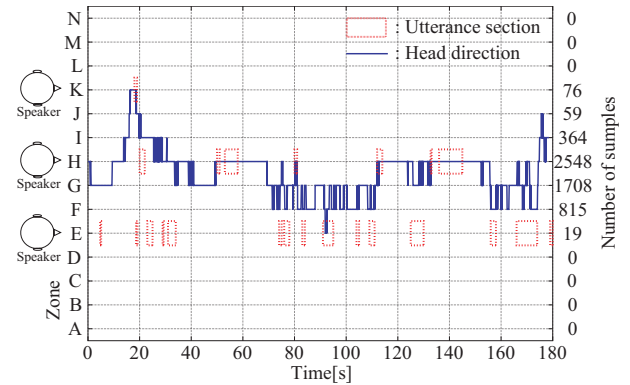


図 18 Subject IV's head direction zone when the speakers are in zone E, H and K.

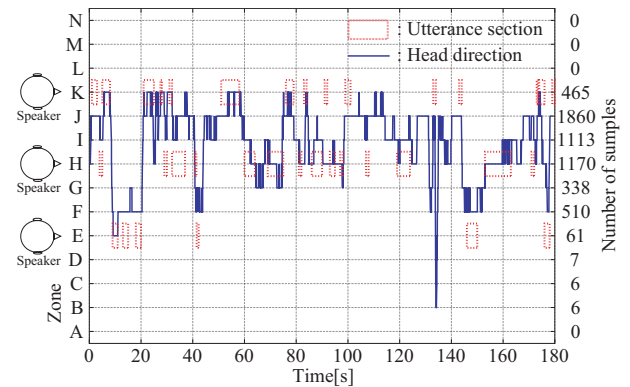


図 19 Subject V's head direction zone when the speakers are in zone E, H and K.

表 2 The rate (P) that a subject's head direction to the zone of utterer while utterance section.

Subject	Two speakers (Zone F, J) [%]	Three speakers (Zone E, H, K) [%]
I	6.63	31.9
II	43.7	8.17
III	14.6	34.5
IV	4.93×10^{-1}	23.6
V	29.7	35.4

表 3 The rate (Q) that a subject's head direction to the zone of speakers.

Subject	Two speakers (Zone F, J) [%]	Three speakers (Zone E, H, K) [%]
I	47.1	59.8
II	38.4	50.1
III	17.2	52.7
IV	2.68	68.2
V	24.3	69.9

正答となる領域 T を対話者がいる領域とした場合の正答率 Q の結果を表 3 に示す。表 3 より、会話中に被験者が対話者の方向を向いている割合は、2 つの会話状況のどちらでも、全ての被験者で 50% に満たなかった。よってこの手法でも、音源方向推定の精度改善に寄与することが困難であると考えられる。

ここで、3.2 項で示したように、人は会話中に対話者から見て $\pm 15^\circ$ の範囲で頭部回転動作以外の動作を行うことを考慮し、対話者がいる領域の両隣の領域も正答となる領域として評価を行う。つまり対話者が 2 人の場合、領域 EFG, IJK, 3 人の場合、領域 DEF, GHI, JKL が正答となる領域である。その場合の、式 (7) を用いた正答率 P の結果を表 4 に示す。対話者が 2 人の場合、被験者 II は 80.1% と高い正答率を示したが、被験者 I は 24.3% と被験者によって差がみられた。被験者 I が発話者の方向を向いている割合が小さいのは、発話者が被験者ではなく、他の対話者に向けて発話しているためであった。対話者が 3 人の場合、被験者 V は 83.18% と高い正答率を示したが、被験者 II は 24.3% と、こちらも被験者によって差がみられた。被験者 II が発話者の方向を向いている割合が小さい理由も、発話者が被験者ではなく、他の対話者に向けて発話しているためであった。式 (8) を用いた正答率 Q についての結果を表 5 に示す。対話者が 2 人の場合、被験者 5 人の正答率の平均は 78.6% であった。ある程度対話者の方向を向いていることが示されているが、被験者 IV は 62.1% と割合が小さく、ここでも被験者によって差があった。被験者 IV が対話者の方向を向いていないとき、対話者 2 人の間を向いていることが多かった。対話者が 3 人の場合、ほぼ 100% の割合で対話者の方向を向いていることが示された。これは、領域 D ~ L が全て正答となり、対話者 2 人の場合での誤推定の主な理由である対話者の間の

表 4 The rate (P) that a subject's head direction to the zone of utterer and on both sides while utterance section.

Subject	Two speakers (Zone EFG, IJK) [%]	Three speakers (Zone DEF, GHI, JKL) [%]
I	24.3	55.3
II	80.1	27.3
III	65.9	63.1
IV	58.6	69.9
V	64.8	83.8

表 5 The rate (Q) that a subject's head direction to the zone of speakers and on both sides.

Subject	Two speakers (Zone EFG, IJK) [%]	Three speakers (Zone DEF, GHI, JKL) [%]
I	87.4	100
II	88.4	100
III	78.7	100
IV	62.1	100
V	76.4	99.8

表 6 The rate (Q) that a subject's head direction to the zone of speakers and between them.

Subject	Two speakers (Zone E ~ K) [%]	Three speakers (Zone D ~ L) [%]
I	99.2	100
II	99.5	100
III	99.1	100
IV	99.9	100
V	99.9	99.8

領域が、全て正答となるためであった。このことから、会話中に被験者は、左端と右端の対話者の外側の領域に頭部正面を向けることが少ないということがわかった。

よって会話中の被験者の頭部正面方向を推定することで対話者の方向を推定することは、被験者によって差があり困難だが、対話者がいる範囲を推定することは可能であると考えられる。そこで、対話者の間の領域を、全て正答である領域として、式 (8) を用いて再度評価した。つまり、対話者が $\pm 30^\circ$ にいる場合は領域 E ~ K, 対話者が 0° , $\pm 45^\circ$ にいる場合は領域 D ~ L が正答となる領域となる。正答率 Q を表 6 に示す。表 6 より、2 つの会話状況において、会話中に被験者が左端の対話者がいる領域の一つ外側の領域から、右端の対話者がいる領域の一つ外側の領域までの間を向いている割合が 99.1% 以上であるということが示された。

4. まとめ

本稿では、FDBM の音源方向推定の精度改善を目的とした対話者の方向推定のため、モーションセンサによる加速度・地磁気の検出、頭部回転角度の推定をし、目標角度と頭部回転角度を対応づける一次関数を求めた。また会話実験を行い、被験者の会話中の頭部正面方向の推定による

対話者の方向推定についての検討を行った。その結果、対話者の方向を推定することは個人差があり、推定することが困難な場合がある。しかし、対話者がいる範囲の領域を推定することは可能であるということが示された。これにより、頭部正面方向を推定することで、複数の対話者がいる範囲の領域を推定することができ、音源方向推定の補助ができるということが示唆された。今後の課題として、短い時間幅での対話者がいる領域を推定する方法の検討が必要である。

謝辞 本研究の一部は、科学研究費補助金（基盤 C No.24500147）の支援を受け行ったものである。

参考文献

- [1] Hidetoshi Nakashima, Yoshifumi Chisaki, Tsuyoshi Usagawa and Masano Ebata: *Frequency domain binaural model based on interaural phase and level differences*, Acoustical Science & Technology, Vol.24, No.4, pp.172-178, (2003).
- [2] Yoshifumi Chisaki, Sho Kawano, Kyoko Nagata, Kotaro Matsuo, Hidetoshi Nakashima and Tsuyoshi Usagawa: *Azimuthal and elevation localization of two sound sources using interaural phase and level differences*, Acoustical Science & Technology, Vol.29, No.2, pp.139-148, (2008).
- [3] 會田卓也, 西本卓也, 大川茂樹, 嵯峨山茂樹: 頭部モーションセンサと音声を用いた対話インタフェースの検討, 電子情報通信学会技術研究報告, pp.85-90, (2007).
- [4] 青山宏, 河越正弘: 無意識動作の検出による直観の推定, 電子情報通信学会総合大会, p.320, (1997).
- [5] 廣瀬圭, 土岐仁, 小田紳介, 永作清: 拡張カルマンフィルタを用いたスキー・ターンの運動計測と力学解析に関する研究, 日本機械学会論文集, 77 巻 774 号, pp.470-480, (2011).