

複数人対話における頭部運動に基づく次話者の予測

石井 亮^{1,a)} 大塚 和弘¹ 熊野 史朗¹ 大和 淳司¹

受付日 2015年6月30日, 採録日 2016年1月12日

概要: 複数人対話において, これまで着目されていなかった参加者の頭部運動と次話者との関連性を明らかにし, 頭部運動の情報を用いて誰が次話者になるかを予測するモデルを提案する. まず, 4者の対話のデータ収録を行い, 発話情報, および頭部計測センサから取得される頭部位置, 回転角からなる頭部運動情報を含む会話コーパスを構築した. コーパスデータを用いて, 現話者の発話末の頭部運動と話者交替の関連性を分析した結果, 話者継続時と交替時で, 現話者の頭部運動が異なることが示された. また, 非話者の発話末の頭部運動と次話者の関連性を分析した結果, 話者継続時の非話者, 話者交替時の非次話者と次話者の3者間で, 頭部運動の特徴が異なることが示された. 分析で差が見られた現話者, 非話者の頭部運動情報を用いて, 話者継続と話者交替のどちらが起るか, さらに, 話者交替時に非話者の中で誰が次話者になるかを2段階で予測する次話者予測モデルを構築した. その予測精度の評価の結果, 現話者と非話者の頭部運動情報が話者継続/交替の予測に有用であることが示唆された. また, 非話者の頭部運動情報が話者交替時の次話者の予測に有用であることが示唆された.

キーワード: 頭部運動, 次話者予測, 話者交替, 複数人対話, 会話分析

Predicting Next Speaker Based on Head Movement in Multi-party Meetings

RYO ISHII^{1,a)} KAZUHIRO OTSUKA¹ SHIRO KUMANO¹ JUNJI YMATAO¹

Received: June 30, 2015, Accepted: January 12, 2016

Abstract: We proposed a model for predicting the next speaker in multi-party meetings by focusing on the participants' head movements measured by using a six degrees-of-freedom head tracker. Results of an analysis of head movements collected from multi-party meetings revealed differences in the amounts and amplitude of movement of the head position and rotation of the speaker near the end of an utterance in turn-keeping and turn-changing. The results also revealed the differences in the amounts of movement, amplitude, and frequency of head position movement and rotation between the listeners in turn-keeping, turn-changing, and the next speaker in turn-changing. We then built a next speaker prediction model that features two processing steps to predict whether turn-changing or turn-keeping will occur and who the next speaker will be in turn-changing. The evaluation results for the model suggest that the speaker's and listeners' head movements contribute to predicting the next speaker.

Keywords: head movement, next-speaker prediction, turn-taking, multi-party meetings, meeting analysis

1. はじめに

対面会話は, 社会生活において他者との情報交換, 意思疎通, 意思決定を行ううえで, 最も基本的なコミュニケーション

形態である. 対面会話のような円滑なコミュニケーションを, 遠隔コミュニケーションや, 人と会話エージェントのコミュニケーションにおいても実現することが望まれている. そのためには, そもそも人間の対面会話の構造やメカニズムがどのようなになっているのかを明らかにし, モデル化することは重要である. 近年, 2者対話だけでなく, 3者以上の対話 (以後, 複数人対話と呼ぶ) に対して

¹ 日本電信電話株式会社 NTT コミュニケーション科学基礎研究所
Communication Science Laboratories, NTT Corporation,
Atsugi, Kanagawa 243-0198, Japan

^{a)} ishii.ryo@lab.ntt.co.jp

も、研究対象が拡大している [1], [2].

話者交替は、会話における重要な局面の1つであり、話者交替に関する多くの研究が行われている [3], [4], [5]. 本研究では、以降、話者継続時において、発話を行っている人物を“現話者”、発話を行っていない人物を“非話者”と呼ぶ。また、話者交替時において、発話を行っていた人物を“現話者”、非話者のうち、次の発話を行う人物を“次話者”、次の発話を行わない人物を“非次話者”と呼ぶ。2人対話では、非話者はつねに1人であるため、次話者は自明である。しかしながら、複数人対話では、複数の非話者が存在し次話者の候補が複数いるため、話者交替がより複雑になる。そのため参加者は、現話者の発話の終了のタイミングや、複数人の非話者のうち次に誰がいつ発話を開始するかを予測して、参加者自身が発話をいつすべきか、または対話相手に発話を促すかといった調整を行っている。もし、複数人対話において次話者や発話開始のタイミングを予測可能な予測モデルが構築できれば、適切なタイミングで発話を開始、終了できる会話エージェントや、遠隔コミュニケーションにおいて参加者に誰が次話者となるかを通知するなどして、映像遅延による発話衝突を回避するシステムの実現に向けた基盤的な技術となることが期待される。

心理学・社会学・言語学などの人文諸分野のコミュニケーション研究において、言語情報に加えて視線や韻律などの非言語情報が次話者や発話のタイミングの規定に重要な役割を担っていることが示されている [3], [4], [5]. 工学分野では、主に音声処理の分野において、複数人対話における話者交替の有無について、特に韻律情報を用いて推定を行う研究が行われてきた [6], [7], [8]. これに加えて、近年、音声情報以外の視線などの身体的な非言語情報もあわせて取り扱われる試みがなされている [9], [10], [11], [12]. さらに、複数人対話を対象に、話者交替時に誰が次話者になるか [13], [14], [15], さらに、前の発話に対して、いつその発話が開始されるかのタイミング [14], [16] を予測可能な予測モデルの構築が取り組まれている。しかしながら、これらの話者交替の有無および次話者の予測モデルは、予測精度が必ずしも十分であるとはいえない。そのため、話者交替の有無や次話者の選定に有用な新たな非言語情報を発見し、その情報を利用したより頑健かつ高精度な予測モデルの構築が望まれている。

本研究では、複数人対話における話者交替の有無や次話者の選定と非言語行動の関連性を明らかにし、より高精度な次話者の予測モデルの構築を目指して、これまで着目されていなかった頭部位置、回転角からなる頭部運動情報に着目する。頭部運動は、傾きなどの頭部の姿勢変化はもちろんのこと、視線変化や体全体の姿勢変化によっても連動して変化するため、頭部運動は多様な情報を保持しており、次話者との関連性は深いと考えられる。

本稿では、最初に4者の対話データの収録を行い、発話情報と頭部センサを用いた位置・回転角からなる頭部運動情報の収集について述べる。次に、現話者の発話末の頭部運動と話者交替の関連性を分析した結果、話者継続時と話者交替時で、現話者の頭部運動が異なることを示す。また、発話末の非話者の頭部運動と次話者の関連性を分析した結果、話者継続時の非話者、話者交替時の非次話者と次話者の3者間で、頭部運動が異なることを示す。分析で差が見られた現話者、非話者の頭部運動情報を用いて、話者継続と話者交替のどちらが起こるか、さらに、話者交替時に非話者の中で誰が次話者になるかを2段階で予測する次話者予測モデルを構築した。その予測精度の評価の結果、現話者と非話者の頭部運動情報が話者継続/交替の予測に有用であること、非話者の頭部運動情報が話者交替時の次話者の予測に有用であることが示唆されたので報告する。

以降、2章では関連研究について、3章では収集した4人対話のコーパスデータについて、4章では、参加者の頭部運動と次話者の関連性についての分析結果について、5章では、分析結果に基づく頭部運動の情報を用いた予測モデルの構築、および予測精度の評価結果について述べる。6章では、頭部運動の分析、および予測モデルの評価結果についての議論を行い、最後に7章で結論と今後の予定を述べる。

2. 関連研究

工学分野では、複数人対話において、話者交替が起こるか否かの推定を行う試みがなされている。この話者交替の有無の推定は、当初、主に音声処理の分野で研究が始められた。Ferrerら [6] は韻律情報、Schlangen [7] は語彙および韻律情報、Levow [8] は声調 (tonal language) を用いて話者交替の有無の推定を行っている。しかしながら、音声情報だけでは話者交替の有無の推定を高精度に行うことは困難であり、近年、音声情報に加えて視線などの情報もあわせて取り扱われる試みがなされている。Dielmannら [9] は身体動作、Jokinenら [10] は注視行動、Chenら [11] はジェスチャと注視行動、de Kokら [12] は話者の頭部ジェスチャ、注視行動を用いて話者交替の有無の推定に取り組んでいる。

話者交替の有無の推定に加えて、話者交替時において、複数の非話者のうち、誰が次話者になるかを推定する試みが、3者のポスター対話を対象に取り組まれている [13]. また我々は、複数人の自然な対話を対象に、視線の遷移パターンや呼吸動作の情報を用いて次話者の予測モデルの構築を行ってきた [14], [15].

しかしながら、これらの話者交替の有無、および次話者の予測モデルは、予測精度が必ずしも十分であるとはいえない。そのため、話者交替の有無や次話者の選定に有用な新たな非言語情報を発見し、その情報を利用したより頑健



図 1 (左) 4 人対話シーンの例, (右) 分析に用いる各参加者の頭部の位置 (x, y, z), 回転角 ($yaw, roll, pitch$) の座標系

Fig. 1 Sample scene of multi-party meetings and coordinate system with origin at center of seated position.

かつ高精度な予測モデルの構築が望まれている。これに対して本研究では、次話者予測のためにこれまで取り扱われていなかった新たな情報である、頭部位置、回転角からなる頭部運動情報に着目をする。

2 者対話において、頭部運動は話者交替と関連性が高いことが知られている。Duncan ら [17], [18] によると、現話者は発話を継続する際、頭を対話相手の方向から外す傾向があるとされている。また、非話者は発話ターンを取得しようとする際に頭の向きを変える傾向にあるとされている。これらの頭部運動は、6 自由度の頭部運動の座標系を図 1 (右) のようにとったとき、主に回転角 yaw の変化として現れると考えられる。また、Maynard によると、頷きの機能の一部として、ターンが終了することの表示、ターン取得の前ぶれの表示、放棄・譲渡されたターンを受け取ることの承認したという表示などがあるとしている [19], [20]。頷きを 6 自由度の頭部運動から抽出すること試みた研究 [21] では、頷きは主に回転角 $pitch$ の変化として現れることが報告されている。これらの知見は、本研究の頭部運動による次話者予測のアプローチの妥当性を支持するものである。

次に本研究に関連して、大まかな頭部姿勢や動作の情報を用いて、話者交替の有無を推定することに取り組んでいる研究を紹介する。Jokinen らは、ビデオを観察して人手でラベル付けされた 9 種類の大まかな頭部の姿勢の情報を利用して、話者交替の有無の判定を試みている [10]。また、de Kok らは、頭部姿勢が意図する同調、強調といった意図に関する 6 種類のラベルを、人手でラベル付けして、それらの情報を話者交替の有無の判定に利用することを試みている [12]。これらの研究では、大まかな頭部姿勢や動作の意図を、ビデオから判断して人手ラベリングを行った情報を利用しており、結果として、これらの情報は話者交替の有無の判別に有効的に機能しなかった。

これに対して本研究では、頭部位置、回転角からなる頭部運動情報に着目をする。頭部運動情報を扱う利点として、先行研究 [10], [12] が扱っていた、頭部イベントには

含まれない頭部の微細な揺れといった詳細な運動情報を扱えることがあげられる。たとえば、頷きなどの同じ頭部イベントであっても、6 自由度の詳細なパラメータはそのつど、異なるはずであり、それらの違いを定量的に扱える利点がある。また、本研究では、頭部の情報を用いた先行研究 [10], [12] が行っている話者交替の有無の推定だけでなく、話者交替時の次話者予測にも焦点を当てる。そして、頭部運動とそれらの関連性を明らかにし、頭部運動の情報を用いた次話者予測の可能性を探る。

3. 複数人対話コーパスの構築

複数人対話における頭部運動などの非言語情報と発話データのコーパスを構築するために、4 人対話を収録した。参加者は 20–30 代の女性であり、対話収録時が初対面であった。参加者は図 1 (左) のように、着席し会話を行った。会話内容は、「血液型と性格に深い関連はあると思うか?」のように、参加者間で意見が大きく割れるような議題について議論をするように会話開始時に教示した。参加者の言語、非言語行動の計測手段として、発話は参加者の胸に装着されたピンマイクを用いて録音し、外観や参加者の挙動はビデオカメラで撮影を行った。頭部運動の計測には、FASTRAK [23] を利用した。また、頭部動作に加えて、NeXus-10 MARKII [24] を利用し、参加者の呼吸 (肺気量) を計測した。本実験には、異なる 4 人の参加者の組が 4 組で、計 16 人が参加した。1 回の対話時間は約 10 分間であり、各組 1 対話ずつ対話を行った。よって、4 対話、計約 40 分の対話を収録した。収録された音声、映像、各種計測情報から下記の言語・非言語データを構築した。

- 発話：音声情報から発声言語の書き越しを行った後、ある一定の音声の無音区間が発生した際に発話の切れ目として扱い、それまでの有音区間を 1 つの発話単位と規定する IPU (Inter-Pausal Unit) [25] を利用して、発話区間を生成した。IPU の具体的な生成方法として、200 ms 以上の無音区間で囲まれる発話部分

表 1 各参加者の発話数 (IPU 数) と発話時間

Table 1 Numbers of IPU and utterance time of each person.

グループ ID	1				2				3				4			
参加者 ID	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
発話数 (IPU 数)	97	39	48	53	46	61	46	92	90	67	99	51	97	54	50	64
発話時間 (秒)	127.5	70.6	115.3	81.9	78.1	90.3	121.1	139.7	178.1	122.0	140.73	106.6	170.8	124.5	65.7	88.9

を 1 つの IPU とした。作成された IPU から、あいつち [26] を除外し、同一の人物による継続した IPU を 1 つの発話ターンとした*1。そして、時間的に連続した IPU の組を作成し、話者継続および話者交替時の IPU の組に分類した。このとき、発話中にオーバーラップした場合や、話者交替時に複数の参加者が同時に発話を開始した場合のデータは除去した。最終的に作成された IPU の組は、話者継続時で 906、話者交替時で 148 であった。計 1,054 の IPU の組のうち、前 IPU がそれぞれの参加者でどれくらいの数であったか、また発話の合計時間を集計したものを表 1 に示す。全体的に、参加者の IPU 数や発話時間には多少の差はあるものの、特定の参加者のみの IPU 数が極端に多く、他の参加者の IPU がほとんどないといった極端な偏りは見られなかった。

- 頭部運動：各参加者は、FASTRAK の小型センサが取り付けられたヘアバンドを、センサが後頭部に位置するように頭に巻いた。このセンサから、6 自由度の頭部位置・回転角の情報が 30 Hz で取得される。取得されるデータの誤差は、トランスミッタとセンサの位置関係などにも影響されるが、おおむね位置で 1 mm 以下、回転角で 0.15° 程度の非常に小さなものであり、高精度に頭部位置、回転角を取得可能である。また、センサから得られるこの位置、回転角の情報は、FASTRAK のトランスミッタを基点とした座標系で表される。分析にあたって、図 1 (右) のように、各参加者の頭部の位置・回転角の座標系を、参加者が椅子の中心に着座した際の頭部位置を原点とし、参加者の横方向を X 軸、前後方向を Y 軸、上下方向を Z 軸、 Z 軸中心の回転を yaw 、 Y 軸中心の回転を $roll$ 、 X 軸中心の回転を $pitch$ となるように変換した。
- 視線：我々が構築した NTT MM-Viewer [29] を用いて、熟練したアノテータが撮影された映像を観察して、各参加者の注視対象のラベリングを行った。ラベリングされた注視対象は、“4 人の各参加者 (人物 P1, P2, P3, P4)” および “人物以外” の 5 種類である。さらに、各注視対象から次の注視対象までに視線が変化してい

る箇所についても、視線変化のラベリングを行った。

- 呼吸動作：参加者の腹部に、NeXus-10 MARKII のバンド式の呼吸センサを装着し、128 Hz で肺気量を計測した。この肺気量は、バンドの伸縮度合いに応じて、参加者ごとに異なる大きさをとる。そこで、個人ごとの肺気量の差を排除するために、各参加者の肺気量の平均値 μ と標準偏差値 δ を用いて、 $\mu + \delta$ が 1、 $\mu - \delta$ が -1 になるように参加者ごとに肺気量を正規化した。これによって、すべての参加者の肺気量を同一に扱うことが可能である。

これらのデータと音声、ビデオデータはすべて 30 Hz の時間分解能に変換し、時間的な同期をとった。

4. 頭部運動の分析

4.1 分析方針

現話者は発話を継続する際、頭を対話相手の方向から外す傾向や、非話者は発話ターンを取得しようとする際に頭の向きを変える傾向にあるといった知見 [17], [18], [19], [20] から、現話者と非話者の発話末付近の頭部運動は話者継続、交替時でそれぞれ異なると考えた。そこで、発話末付近の現話者と非話者の頭部運動をそれぞれ話者継続/交替の状況ごとに別々に分析を行った。具体的に、現話者の頭部運動においては、話者継続時と話者交替時でどのような差があるかを分析した。非話者の頭部運動においては、話者継続時の非話者、話者交替時の非次話者と次話者の 3 つの状況に分けて、どのような差があるかを分析した。頭部運動の分析区間を決定するうえで、実際に頭部運動の波形を観察した結果、おおむね発話末の約 3 秒前から、参加者の頭部運動が大きく変化することが観察された。また、これまでの我々の研究 [14], [15] においても発話末の時刻付近の情報を利用して、予測モデルを構築することを試みていた。以上のことから、実験的に、発話末の 3 秒前から発話末までの 3 秒間を頭部運動の分析区間として設定した。また分析パラメータとして、頭部の x , y , z の各座標位置と、 yaw , $roll$, $pitch$ の各回転角に対して、変化量、揺れの振幅と周期に関する下記の 3 つの分析パラメータを定義した。

- MO ：1 秒間あたりの変化量の平均。
- AM ：位置または角度の変化を波の振動と見なしたときの波の振幅の平均。
- FQ ：位置または角度の変化を波の振動と見なしたときの 1 秒間あたりの波の周波数の平均。

*1 IPU では一連の発話の途中であっても、音声の一定区間の無音区間をもって発話を区切るため、本研究で扱っている話者継続時における IPU の切れ目は、ターン末で話者交替が可能であった箇所と、単に無音区間が発生したがために発生した箇所の分離は行われていない。

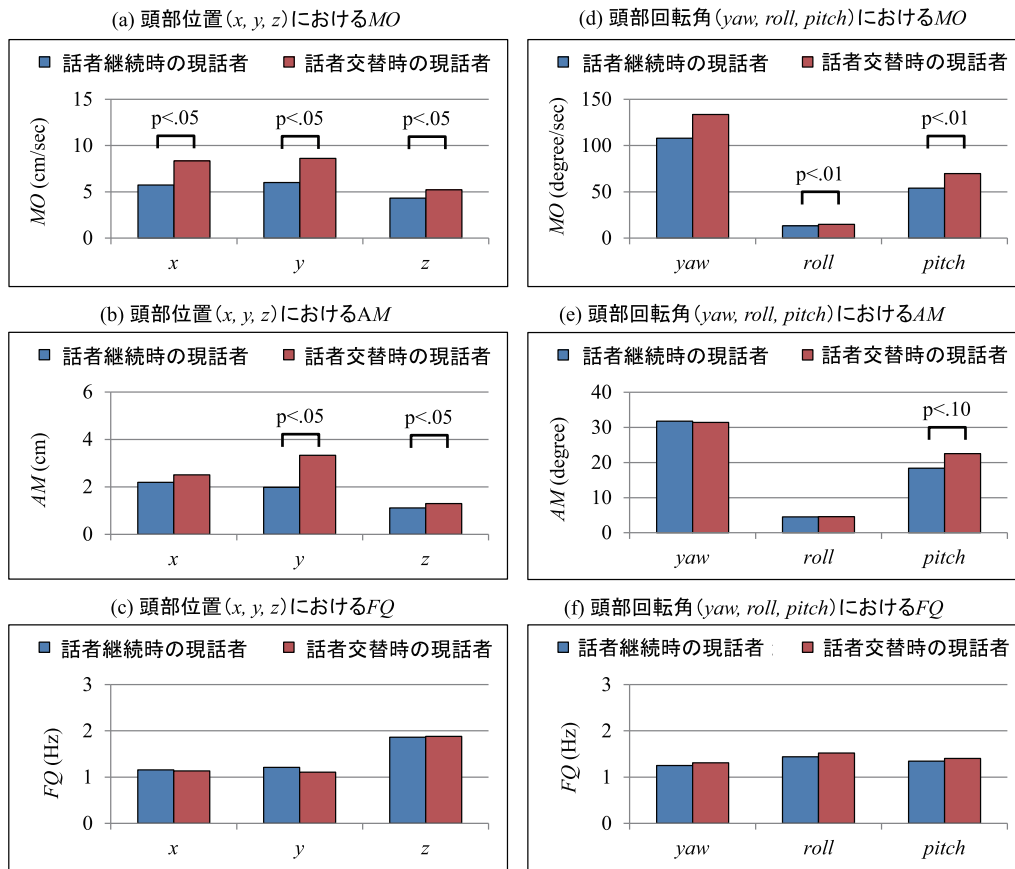


図 2 現話者の頭部位置 (x, y, z) と回転角 ($yaw, roll, pitch$) の各パラメータの MO, AM, FQ の分析結果

Fig. 2 Results of analysis of $MO, AM,$ and FQ of speaker's head position (x, y, z) and rotation ($yaw, roll, pitch$).

4.2 現話者の頭部運動の分析

話者継続時で 906、話者交替時で 148 のデータを用いて、現話者の頭部位置 (x, y, z) と回転角 ($yaw, roll, pitch$) の各パラメータにおける、 MO, AM, FQ の平均値を算出した。その結果を、図 2 に示す。話者継続、話者交替の条件間で、現話者の頭部運動の各パラメータに有意な差があるかを検証するために、参加者 16 人の各パラメータの平均値を用いて、対応のある t 検定を行った。その結果、 $x, y, z, roll, pitch$ における MO と、 $y, z, pitch$ における AM において、有意差または有意傾向が認められた*2。以上の結果から、現話者の頭部位置・回転角において、 $x, y, z, roll, pitch$ における $MO, y, z, pitch$ における AM は、話者継続時よりも話者交替時で大きくなる傾向にあることが示唆された。

*2 t 検定の結果の詳細を以下に示す。 x における $MO : t(30) = -2.58, p < .05, y$ における $MO : t(30) = -2.22, p < .05, z$ における $MO : t(30) = -2.38, p < .05, roll$ における $MO : t(30) = -3.28, p < .01, pitch$ における $MO : t(30) = -3.24, p < .01, y$ における $AM : t(30) = -2.83, p < .05, z$ における $AM : t(30) = -2.34, p < .05, pitch$ における $AM : t(30) = -1.89, p < .10$

4.3 非話者の頭部運動の分析

話者継続時で 906、話者交替時で 148 のデータを用いて*3、話者継続時の非話者、話者交替時の非次話者と次話者の頭部位置 (x, y, z) と回転角 ($yaw, roll, pitch$) における、 MO, AM, FQ の平均値を算出した。その結果を、図 3 に示す。非話者における 3 つの条件が、頭部運動の各パラメータに有意な影響を与えるかを検証するために、一元配置の分散分析を行った。その結果、すべてのパラメータにおいて有意差が認められた*4。

次に、非話者における 3 つの条件の要因に効果が確認されたすべてのパラメータについて、Bonferroni 法による多重比較を行った。その結果、 z における FQ 以外のすべてのパラメータにおいて、話者継続時の非話者と話者交替時の非次話者の条件間、および話者継続時の非話者と話者交替時の次話者の条件間で有意差または有意傾向が認められた。また、 x における MO と、 z における MO, AM, FQ において、話者交替時の非次話者と次話者の条件間で有意

*3 このとき、話者継続時は 3 人の非話者、話者交替時は 2 人の非次話者と 1 人の次話者がいることから、分析に用いたデータ数は、話者継続時の非話者で 2,718、話者交替時の非次話者で 296、次話者で 148 となる。

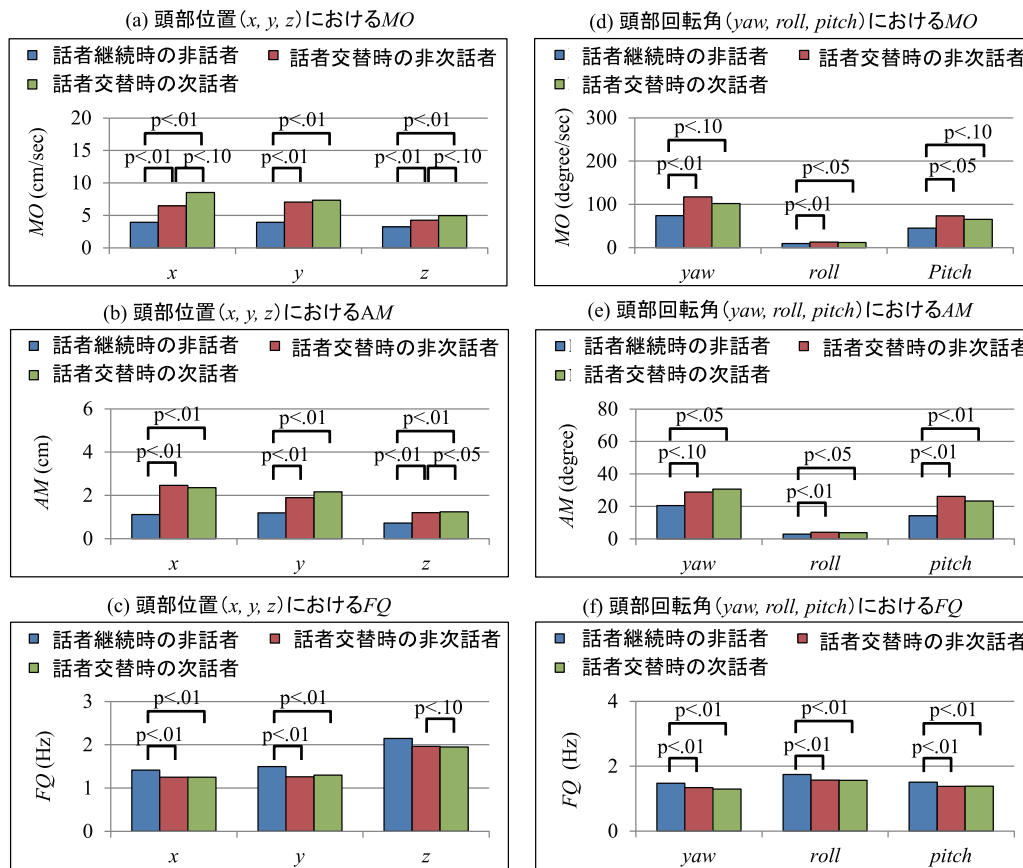


図 3 話者継続時の非話者，話者交替時の非次話者と次話者の頭部位置 (x, y, z) と回転角 ($yaw, roll, pitch$) の各パラメータの MO, AM, FQ の分析結果
Fig. 3 Results of analysis of $MO, AM,$ and FQ of head positions (x, y, z) and rotations ($yaw, roll, pitch$) of listeners in turn-keeping and turn-changing and next speaker in turn-changing.

差または有意傾向が認められた。以上の結果から、非話者の頭部運動について下記のことが明らかになった。

- $x, y, z, yaw, roll, pitch$ における MO と AM は、話者継続時の非話者に比べて、話者交替時の非次話者と次話者の方が大きい。逆に、 $x, y, yaw, roll, pitch$ における FQ は、話者継続時の非話者に比べて、話者交替時の非次話者と次話者の方が小さい傾向にある。
- x と z における MO 、 z における AM は、話者交替時の非次話者に比べて、次話者の方が大きい。逆に、 z

における FQ は、話者交替時の非次話者に比べて、次話者の方が小さい傾向にある。

5. 頭部運動に基づく予測モデルの構築

前章で、現話者の発話末の頭部運動が話者継続時と交替時で異なること、話者継続時の非話者と話者交替時の非次話者と次話者の間で、発話末の頭部運動が異なることが明らかとなった。本章では、これらの頭部運動を用いて、次話者の予測モデルを構築し、頭部運動の情報が予測に有用であるかを検証する。本研究では、次話者を予測するにあたり、先行研究 [14], [15] と同様に、2段階で処理を行う予測モデルを構築する。具体的に、1段階目として、話者継続と話者交替のどちらが起こるか、さらに、2段階目として、話者交替時に非話者の中で誰が次話者になるかを順に予測する、また、先行研究 [14], [15] で有用であった、視線と呼吸動作をそれぞれ用いた予測モデルもあわせて構築し、予測精度の比較を行った。

*4 一元配置の分散分析の結果の詳細を以下に示す。 x における $MO: F(2, 3060) = 16.8, p < .01$, y における $MO: F(2, 3060) = 18.3, p < .01$, z における $MO: F(2, 3060) = 19.3, p < .01$, yaw における $MO: F(2, 3060) = 8.5, p < .01$, $roll$ における $MO: F(2, 3060) = 10.7, p < .01$, $pitch$ における $MO: F(2, 3060) = 9.9, p < .01$, x における $AM: F(2, 3060) = 31.6, p < .01$, y における $AM: F(2, 3060) = 19.1, p < .01$, z における $AM: F(2, 3060) = 32.2, p < .01$, yaw における $AM: F(2, 3060) = 6.4, p < .01$, $roll$ における $AM: F(2, 3060) = 15.9, p < .01$, $pitch$ における $AM: F(2, 3060) = 29.7, p < .01$, x における $FQ: F(2, 3060) = 12.8, p < .01$, y における $FQ: F(2, 3060) = 20.8, p < .01$, z における $FQ: F(2, 3060) = 3.8, p < .10$, yaw における $FQ: F(2, 3060) = 201.1, p < .01$, $roll$ における $FQ: F(2, 3060) = 12.6, p < .01$, $pitch$ における $FQ: F(2, 3060) = 8.3, p < .01$

5.1 現話者と非話者の頭部運動に基づく話者継続/交替の予測

4.1節の分析結果から、現話者の頭部位置と回転角の x , y , z , $roll$, $pitch$ における MO と、 y , z , $pitch$ における AM は、話者継続時よりも話者交替時で大きくなる傾向にあることが示唆された。また、4.2節の分析結果から、話者継続時の非話者と、話者交替時の非次話者および次話者の間で、頭部運動のうち、 z における FQ 以外のすべてのパラメータに差が見られた。よって、これらの現話者、非話者の頭部運動のパラメータは、話者継続・交替の予測に利用できると考えられる。そこで、それらの情報が実際に予測に有用であるかを検証するために、非線形識別およびソフトマージン法による SVM (Support Vector Machine) を利用して話者継続/交替の予測モデルを構築し、その予測精度を評価した。SVM における二次計画問題 (QP) を解くうえで、Weka [27] に実装された SMO (Sequential Minimal Optimization) アルゴリズム [28] を用いた。また、SVM のカーネルには RBF (Radial Basis Function) を用いた。クラスとして、話者継続、話者交替の 2 クラスを設定した。頭部運動の情報を用いた予測モデルの特徴量として、下記の計 61 種類のパラメータを使用した。

- 現話者の頭部運動：現話者の頭部運動の x , y , z , $roll$, $pitch$ における MO と、 y , z , $pitch$ における AM の 8 種類の情報。
- 非話者の頭部運動：3 人の非話者の x , y , z , yaw , $roll$, $pitch$ における MO と AM , x , y , yaw , $roll$, $pitch$ における FQ の 51 種類の情報。
- 現話者の人物情報：現話者が誰であったかの 1 種類の情報。

また、視線、呼吸動作の情報を用いた予測モデルで利用した特徴量を下記に示す。

- 視線の情報：IPU 末の 1,000 ms 前から IPU 末までの 1,000 ms の間に、注視対象がどのように遷移したかという情報である注視遷移パターン [14] を利用する。注視遷移パターンは、対話の参加者を、現話者と非話者に分類し、分析区間でどのような順序で参加者を注視したかという時間的な注視対象の遷移パターンを N-gram で表現したものである。また、注視した際の相互注視の有無の情報も、注視遷移パターンには含まれている。注視遷移パターン生成の詳細については、先行研究 [14] を参照されたい。具体的に使用したのは、現話者および 3 人の非話者の注視遷移パターンが何であったかの 4 種類の情報である。
- 呼吸動作の情報：IPU 末の直前に起きた吸気の特徴を利用する。具体的に、3 人の非話者の吸気量、吸気終了時の肺気量、吸気波形の傾き、吸気開始時刻と IPU 末の時間間隔 [15] の 12 種類の情報を使用した。なお、現話者の IPU 末直後の吸気の特徴も話者継続/交替の

表 2 話者継続/交替の予測結果

Table 2 Results of evaluation of prediction model of turn-keeping and turn-changing.

使用した情報	正答率
チャンスレベル	50.0%
視線 [14]	79.5%
呼吸動作 [15]	58.8%
頭部運動	62.5%

予測に有用であるが、本研究では、IPU 末までの情報で予測を行うため、IPU 末直後の吸気の特徴は利用しない。

視線および呼吸動作の情報を利用した予測モデルは、これらのパラメータに加えて、頭部運動情報を用いた予測モデルの特徴量と同様に、現話者の人物情報を利用した。よって、視線情報を用いた予測モデルで計 5 種類、呼吸動作を用いた予測モデルで計 13 種類のパラメータを使用した。

予測モデル構築のために使用した話者継続/交替のデータは、4 章の分析で使用した 4 対話から得られた 906 の話者継続データと、148 の話者交替データである。このうち、3 対話分のデータを学習データとして用い、残り 1 対話分のデータをテストデータとして用いることで、構築した予測モデルが、他の人物に適用してどれだけ汎用的に予測ができるかを検証した。学習およびテストに用いる対話データは入れ替えを行い、4 交差検定を実施した。またこのとき、話者継続と交替のデータ数を均等にするために、データ数の少ない話者交替のデータ数に合わせて、話者継続のデータの中からダウンサンプリングを行って、話者継続と話者交替のデータ数を揃えた。なおこのとき、データの偏りを緩和するために、ランダムにダウンサンプリングを 10 回繰り返す、モデルの構築と評価を 10 施行した。SVM におけるパラメータ C および γ は、毎回、グリッドサーチを行い最適な値を探索した。具体的に、学習データ集合のみを用いて、10 交差検定で評価を行ったうえで、最も予測精度が高くなる最適なパラメータを取得した。パラメータの探索範囲は、 C で $-0.00001 \sim 100000$, γ で $0.01 \sim 100$ とした。

予測モデルの精度の評価結果を表 2 に示す。頭部運動を用いた予測モデルの正答率は 62.5% であった。話者継続と交替のデータ数は同じであったためチャンスレベルは 50.0% であり、頭部運動の情報を用いた予測モデルの予測精度はチャンスレベルを上回る結果となった。よって、現話者と非話者の頭部運動が話者継続、話者交替の予測に有用であることが示された。また、視線情報を用いた予測モデルは、正答率 79.5% と最も高い結果となった。また、呼吸動作を用いた予測モデルは、正答率 58.8% であり、頭部運動を用いた予測モデルよりも、正答率が低かった。

表 3 話者交替時の次話者の予測結果

Table 3 Results of evaluation of prediction model of next speaker in turn-changing.

使用した情報	正答率
チャンスレベル	33.3%
視線 [14]	52.5%
呼吸動作 [15]	53.4%
頭部運動	57.0%

5.2 非話者の頭部運動に基づく話者交替時の次話者の予測

4.2節の分析結果から、頭部位置 x , z における MO , および z における AM は、話者交替時の非次話者に比べて次話者の方が大きく、 z における FQ は、話者交替時の非次話者に比べて次話者の方が小さい傾向にあることが明らかになった。よって、これらの非話者の頭部運動のパラメータは話者交替時の次話者の予測に利用できると考えられる。そこで、実際にそれらの情報が予測に有用であるかを検証するために、SVM (Support Vector Machine) を用いた予測モデルを構築し、モデルの予測精度を評価した。利用したアルゴリズムや手順は、前節と同様である。頭部運動の予測モデルで使用する特徴量は、話者交替時に非話者であった3人の人物の3クラス、特徴量として下記の計16種類のパラメータを使用した。

- 非話者の頭部運動：3人の非話者の頭部位置 x における MO , z における MO , AM , FQ の12種類の情報
- 現話者の人物情報：現話者が誰であったか (3人の非話者が誰であったか) の1種類の情報

視線、呼吸動作情報を用いた予測モデルの特徴量は、5.1節と同様である。使用したデータは、4章の分析で使用した話者交替の148のデータを用いた。5.1節の話者継続/交替の予測モデル構築時と同様の手順で、話者交替時の次話者の予測モデルを構築し、その予測精度を評価した。

予測モデルの予測精度の評価結果を表3に示す。頭部運動を用いた予測モデルの正答率は57.0%であり、視線、呼吸動作を用いた予測モデルに比べて、最も正答率が高かった。また、3人の非話者から次話者を予測するため、チャンスレベルは33.3%であり、構築した頭部運動を用いた予測モデルの予測精度はチャンスレベルを上回る結果となった。よって、非話者の頭部運動が話者交替時の次話者の予測に有用であることが示された。

6. 議論

6.1 頭部運動の分析結果について

4章の話者継続/交替時における現話者と非話者の頭部運動の分析結果と、話者交替時における頭部行動の従来知見 [17], [18], [19], [20] とを比較し議論を行う。

まず、4.1節の現話者の頭部運動の分析結果に関連して、Duncan ら [17], [18] が示したような、現話者が発話を継続

する際に見られる頭を対話相手の方向から外すといった挙動は、主に頭部回転角 yaw に現れると期待された。しかしながら、話者継続時と話者交替時で有意な差は見られなかった。この原因については、Duncan ら [17], [18] の実験が2者対話を対象にしていたのに対して、本研究が4人対話を扱っていることや、座席配置による影響など様々な要因が考えられる。それらの要因についての検証は、今後さらなる検討が必要である。

また、Maynard [19], [20] が示した、現話者によるターンが終了することの傾きの情報は、主に回転角 $pitch$ に現れると考えられる。分析の結果、現話者の頭部の回転角 $pitch$ の変化量 MO と振幅 AM は、話者交替時の方が話者継続時よりも大きかった。よって、話者継続時に比べて話者交替時で、現話者が大きく傾く傾向にあることを示しており、これは Maynard [19], [20] が報告したような、現話者によるターン終了の合図と考えられる傾きを、回転角 $pitch$ の変化量や振幅の大きさとして定量的に示したものであると考えられる。

一方、現話者の頭部の位置 x , y , z においても、変化量 MO や振幅 AM が、話者継続時よりも交替時で大きくなる結果となった。これらの変化量の差は、平均して5cm未満の微細な差である。このような微細な頭部運動が話者交替や次話者を予測するうえで重要な情報である可能性がある。この結果は、6自由度の頭部運動を分析したことにより得られた、従来知見では明らかにされていない新たな知見であると考えられる。

次に、4.2節の非話者の頭部運動の分析結果についても同様に議論する。Duncan ら [17], [18] が示したような、非話者が発話ターンを取得しようとする際に頭の向きを変える傾向は、6自由度の頭部運動においては、主に回転角 yaw の変化の大きさとして現れると考えられる。また、Maynard [19], [20] によると、傾きの機能の一部として、非話者によるターン取得の前ぶれの表示、放棄・譲渡されたターンを受け取ることを承知したという表示などがあるとしている。これらの傾きは、6自由度の頭部運動において、主に回転角 $pitch$ の変化として現れると考えられる。

分析結果では、頭部回転角 yaw , $pitch$ の変化量 MO と振幅 AM は、話者継続時の非話者よりも、話者交替時の次話者の方が大きかった。また、頭部回転角 yaw , $pitch$ の周波数 FQ は、話者継続時の非話者よりも、話者交替時の次話者の方が小さかった。これらの結果は、話者継続時の非話者よりも、話者交替時の次話者の方が、頭の向きの変化や傾きを大きく行っていたことを裏付ける結果であり、前述した従来知見 [17], [18], [19], [20] と一致する。

また、話者交替時の非次話者と次話者の頭部運動の特徴を比較した結果として、頭部位置 x と z における変化量 MO と、 z における振幅 AM は、話者交替時の非次話者に比べて、次話者の方が大きい。逆に、 z における周波数 FQ

表 4 視線変化にともなう/ともなわない頭部運動の情報を用いた予測モデルの評価結果
 Table 4 Results of evaluation of prediction model with/without head movement associated with gaze shift.

使用した情報	話者継続/交替の正答率	話者交替時の次話者の正答率
チャンスレベル	50.0%	33.3%
視線変化にともなう頭部運動	54.2%	35.3%
視線変化にともなわない頭部運動	60.7%	54.2%
すべての頭部運動 (5.1, 5.2 節)	62.5%	57.0%

は、話者交替時の非次話者に比べて、次話者の方が小さい傾向にあった。これは、話者交替時の次話者が非次話者に比べて、横や上下方向に頭部運動を大きく行っていることを示している。これらの頭部位置の変化量 MO と振幅 AM の差は、それぞれ平均して 5 cm 以下と 1 cm 以下であり、周波数 FQ の差も平均して 0.2 Hz 以下であった。よって、話者交替時の非次話者と次話者の間で、非常に小さな頭部運動の揺れの差が存在することが明らかとなった。この結果についても、6 自由度の頭部運動を分析したことにより得られた、従来知見 [17], [18], [20] では明らかにされていない新たな知見であると考えられる。

6.2 頭部運動を用いた次話者の予測モデルについて

5 章において、視線、呼吸動作、頭部運動をそれぞれ用いて予測モデルを構築した結果、話者継続/交替の予測においては、頭部運動を用いた予測モデルは、視線の情報を用いた予測モデルに予測精度は及ばないものの、呼吸動作を用いた予測モデルよりも精度が高かった。また、話者交替時の次話者の予測には、頭部運動の情報を用いた予測モデルの予測精度が最も高いことが示唆された。

これらの予測モデルを構築し、その予測精度の評価を行ううえで、3 組の参加者の対話データを用いて学習を行い、その他の参加者の組のデータを用いてモデルの予測精度の評価を実施したことから、予測モデルの精度の評価結果は、特定の参加者の個性も学習データに含んだうえで、他者に適用してどれだけ有効であったかを示している。頭部運動の特徴量が話者継続/交替、および話者交替時の次話者予測に有効であったことから、本研究で用いた頭部動作の特徴量は特定の人物の個性に大きく依存をしていないものと考えられる。

また、本研究で用いた対話データは、参加者の座席は半円状に配置されており、身体方向や頭部方向に対称性が保たれていないため、座席配置によっても頭部運動が異なる可能性が考えられた。これについては、予測モデルを構築するうえで、座席配置の情報は特徴量として用いていなかったにもかかわらず、構築した予測モデルが話者継続/交替、および話者交替時の次話者予測に有効であったことから、頭部運動の情報は座席配置に、大きく影響を受けない可能性が考えられる。頭部運動の情報が、予測にどれだ

け汎用的に利用できるかは、検討の余地が多くあり、今後、様々な参加者や座席配置による対話データを用いて、頭部運動への影響を詳しく検証することが望まれる。

また、本研究では、頭部運動を 6 自由度の頭部位置・回転角の変化として扱った。そのため、頷きや、視線変化にともなう頭部行動といった、特定の頭部イベントを取り扱ってはいない。言い換えれば、このような頭部イベントが、本研究で扱った頭部運動の情報には含まれていると考えられる。一方で、視線の情報を用いた予測モデルは、話者継続/交替の有無、および話者交替時の次話者の予測に有効であり、特に話者継続/交替の予測には高い予測精度を示した。そのため、本研究で用いた頭部運動の特徴量の中で、“視線変化にともなう頭部運動”が予測の特徴量として大きく寄与しており、本研究が利用した頭部運動の情報が単に視線変化を代替して抽出されたものであった可能性も考えられる。そこで、本研究で用いた頭部運動の情報のうち、“視線変化にともなう頭部運動”と、“視線変化にともなわない頭部運動”を分離し、各々の特徴量が予測にどの程度有用であったかを検証した。もし、視線変化にともなわない頭部運動の特徴が予測に有用であれば、本研究が利用した頭部運動の情報が、単に視線変化を代替して抽出されたものではなく、視線変化とは異なる新たな特徴量であったと考えられる。

視線変化にともなう頭部運動と、視線変化にともなわない頭部運動の特徴量をそれぞれ用いた予測モデルを構築し、その予測精度を検証した。具体的には、5 章において予測モデルの特徴量として用いた、発話末 3 秒前から発話末までの頭部運動において、視線変化のラベリングがされた区間での頭部運動を“視線変化に付随する頭部運動”と定義し、視線変化のラベリングがされていない区間での頭部運動を“それ以外に付随する頭部運動”と定義した。そして、それぞれ抽出された区間での頭部運動の特徴量を算出し、それらの特徴量を用いて予測モデルを構築した。予測モデルの構築方法は、5.1, 5.2 節と同様である。予測精度の評価結果を表 4 に示す。

視線変化にともなう頭部運動を用いた予測モデルは、話者継続/交替の予測で正答率が 54.2%、話者交替時の次話者の正答率が 35.3%と、チャンスレベルと比べて、予測精度はあまり高くなかった。一方、視線変化にともなわない

頭部運動を用いた予測モデルは、話者継続/交替の予測で正答率が60.7%、話者交替時の次話者の正答率が54.2%であり、5.1, 5.2節で構築したすべての頭部運動を用いた予測モデルと比べて、若干、精度が落ちるものの大きく精度を落とすことはなかった。よって、本研究が扱った6自由度の頭部運動には、視線変化や傾きなどにもなる様々な頭部運動が含まれているが、視線変化にもなる頭部運動の成分が予測に有効であったわけではなく、それ以外の頭部運動が話者継続/交替、話者交替時の次話者の予測に有効であった可能性が示唆された。

今後、本研究で扱った頭部運動の情報の中で予測に有用であった情報が、具体的にどのような頭部イベントにあたるのかを、検証することも興味深い。

7. まとめと展望

複数人対話において、これまで明らかにされていなかった参加者の6自由度の頭部運動と話者交替および次話者との関連性を分析し、頭部運動の情報をを用いて次に誰が次話者になるかを予測するモデルを構築した。現話者の頭部運動と話者継続/交替の関連性を分析した結果、話者継続時と話者交替時で、現話者の発話末の頭部位置・回転角の変化量と振幅のパラメータが異なることが示された。また、非話者の頭部運動と次話者の関連性を分析した結果、話者継続時の非話者、話者交替時の非次話者と次話者の3者間で、発話末の頭部位置・回転角の変化量、振幅、周波数が異なることが示された。分析によって差が見られた現話者、非話者の頭部情報のパラメータを用いて、話者継続と話者交替のどちらが起るか、さらに、話者交替時に非話者の中で誰が次話者になるかを予測する予測モデルを構築した。予測モデルの精度を評価した結果、現話者の頭部位置・回転角の変化量と振幅および、非話者の頭部位置・回転角の変化量、振幅、周波数の情報が話者継続/交替の予測に有用であることが示された。また、非話者の頭部位置の変化量、振幅、周波数が、話者交替時の次話者の予測に有用であることが示唆された。

本研究の成果は、次話者予測における頭部位置・回転角の利用の有用性を初めて示唆したものであり、会話分析において、参加者の頭部位置・回転角の情報が様々な事象を分析するうえで重要な情報である可能性が考えられる。また、頭部運動はカメラ画像の処理によるトラッキング技術[22]を用いることで、近年、比較的簡単に計測が可能である。従来、次話者予測のために用いられてきた視線行動や呼吸動作は、アイトラッカや呼吸計測装置が必要であり[14], [15]。それに比べて、提案した頭部運動を用いた予測手法は、将来的により汎用的に利用できる可能性を秘めている。また、本研究の次話者予測の成果の一適用先として、人間との複数人対話において自然に発話交替を行うことのできる会話エージェントシステムの実現を例にとる

と、2つの観点から貢献がなされる可能性が考えられる。1点目は、人間の頭部運動から次話者が誰になるかという予測を行うことで、たとえば、会話エージェント自身が適切なタイミングで発話を行うことを支援する。2点目は、発話交替の状況に合わせた頭部運動の制御を支援する。たとえば、会話エージェントが次話者となる際には、本研究の分析結果を基に、頭部の上下方向(z 方向)の変化量を大きくすることで、人間と同じような自然な頭部運動を行うことができる可能性も考えられる。これらの具体的な制御方法については、今後、検討を行う必要がある。

また今後は、より詳細な頭部運動のパラメータを検討し、次話者の予測モデルの予測精度を向上させるとともに、頭部運動と発話の開始タイミングとの関連性を調査し、発話開始タイミングの予測モデル構築に取り組む予定である。また、視線行動や呼吸動作などの他のモダリティと頭部運動を用いた、マルチモーダル処理による予測モデルの構築にも取り組む予定である。

参考文献

- [1] Gatica-Perez, D.: Analyzing group interactions in conversations: A review, *Proc. IEEE Int'l Conf. Multi-sensor Fusion and Integration for Intelligent System*, pp.41–46 (2006).
- [2] Otsuka, K.: Conversational scan analysis, *IEEE Signal Processing Magazine*, Vol.28, pp.127–131 (2011).
- [3] Kendon, A.: Some functions of gaze direction in social interaction, *Acta Psychologica*, pp.26:22–63 (1967).
- [4] Sacks, H., Schegloff, E. and Jefferson, G.: A simplest systematics for the organisation of turn taking for conversation, *Language*, pp.50:696–735 (1974).
- [5] 榎本美香, 伝 康晴: 話し手の視線の向け先は次話者になるか, *社会言語科学*, Vol.14, pp.97–109 (2011).
- [6] Ferrer, L., Shriberg, E. and Stolcke, A.: Is the speaker done yet? Faster and more accurate end-of-utterance detection using prosody in human-computer dialog, *Proc. International Conference on Spoken Language Processing (ICSLP)*, Denver, CO (2002).
- [7] Schlangen, D.: From reaction to prediction experiments with computational models of turn-taking, *Proc. International Conference on Spoken Language Processing (ICSLP)* (2006).
- [8] Levow, G.: Turn-taking in mandarin dialogue: Interactions of tones and intonation, *Proc. SIGHAN Workshop* (2005).
- [9] Dielmann, A., Garau, G. and Boulard, H.: Floor holder detection and end of speaker turn prediction in meetings, *INTERSPEECH*, pp.2306–2309 (2010).
- [10] Jokinen, K., Furukawa, H., Nishida, M. and Yamamoto, S.: Gaze and turn-taking behavior in casual conversational interactions, *J. Transactions on Interactive Intelligent Systems*, Vol.3, No.2, Article 12 (2013).
- [11] Chen, L. and Harper, M.: Multimodal floor control shift detection, *Proc. Int. Conf. on Multimodal Interface (ICMI)*, pp.15–22 (2009).
- [12] de Kok, I. and Heylen, D.: Multimodal end-of-turn prediction in multi-party meetings, *Proc. Int. Conf. on Multimodal Interface (ICMI)*, pp.91–98 (2009).
- [13] Kawahara, T., Iwatate, T. and Takanashi, K.: Prediction

of turn-taking by combining prosodic and eye-gaze information in poster conversations, *Proc. INTERSPEECH* (2012).

- [14] 石井 亮, 大塚和弘, 熊野史朗, 松田昌史, 大和淳司: 複数人対話における注視遷移パターンに基づく次話者と発話開始タイミングの予測, *信学論*, Vol.J97-A, No.6, pp.453–468 (2014).
- [15] Ishii, R., Otsuka, K., Kumano, S. and Yamato, J.: Analysis of Respiration for Prediction of “Who Will be Next Speaker and When” in Multi-Party Meetings, *ACM International Conference on Multimodal Interaction (ICMI 2014)*, pp.18–25 (2014).
- [16] Ishii, R., Otsuka, K., Kumano, S. and Yamato, J.: Analysis and Modeling of Next Speaking Start Timing based on Gaze Behavior in Multi-party Meetings, *Proc. Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)* (2014).
- [17] Duncan S. and Niederehe, G.: On signalling that it’s your turn to speak, *J. Experimental Social Psychology*, Vol.10, pp.234–247 (1974).
- [18] Duncan, S. and Fiske, D.W.: *Face-to-face interaction: Research, methods and theory*, Hillsdale, New Jersey, Lawrence Erlbaum (1977).
- [19] Maynard, S.: Interactional functions of a nonverbal sign: Head movement in Japanese dyadic casual conversation, *Journal of Pragmatics*, Vol.11, pp.589–606 (1987).
- [20] Maynard, S.: *Japanese conversation: Self-contextualization through structure and interactional management*, Norwood, New Jersey, Ablex Publishing Corporation (1989).
- [21] 山本 剛, 坂根 裕, 竹林洋一: マルチモーダルヘッドセットを用いたうなずき検出と会話の重要箇所把握, *情報処理学会研究報告*, Vol.2003, No.94, pp.13–19 (2003).
- [22] 三上 弾, 大塚和弘, 大和淳司: メモリベースパーティクルフィルタ: 状態履歴に基づく事前分布予測を用いた頑健な対象追跡, *電子情報通信学会論文誌 D*, Vol.J93-D, No.8 (2010).
- [23] Polhemus FASTRAK: available from (<http://polhemus.com/motion-tracking/all-trackers/fastrak/>)
- [24] MIND MEDIA “NeXus-10 MARKII”, available from (<http://www.mindmedia.info/CMS2014/products/systems/nexus-10-mkii>)
- [25] Koiso, H., Horiuchi, Y., Tutiya, S., Ichikawa, A. and Den, Y.: An analysis of turn-taking and backchannels based on prosodic and syntactic features in Japanese Map Task dialogs, *Language and Speech*, Vol.41, pp.295–321 (1998).
- [26] Jovanovic, N., Akker, R. and Nijholt, A.: Addressee identification in face-to-face meetings, *Proc. Conf. Euro. Chapter of the Association for Computational Linguistics* (2006).
- [27] Bouckaert, R.R., Frank, E., Hall, M.A., Holmes, G., Pfahringer, B., Reutemann, P. and Witten, I.H.: WEKA—Experiences with a Java Open-Source Project, *Journal of Machine Learning Research*, Vol.11, pp.2533–2541 (2010).
- [28] Keerthi, S.S., Shevade S.K., Bhattacharyya, C. and Murthy, K.R.K.: Improvements to Platt’s SMO Algorithm for SVM Classifier Design, *Neural Computation*, Vol.13, No.3, pp.637–649 (2001).
- [29] Otsuka, K., Araki, S., Mikami, D., Ishizuka, K., Fujimoto, M. and Yamato, J.: Realtime meeting analysis and 3D meeting viewer based on omnidirectional multimodal sensors, *Proc. Int. Conf. on Multimodal Interface (ICMI)*, pp.219–220 (2009).



石井 亮

2008年東京農工大学大学院工学府情報工学専攻修士課程修了。同年日本電信電話株式会社入社。現在、NTTコミュニケーション科学基礎研究所研究員。2013年京都大学大学院情報学研究科博士後期課程修了。2011～2013年成蹊大学客員研究員。コミュニケーション科学、ヒューマンコンピュータインタラクションの研究に従事。ACM 15th Int. Conf. Multimodal Interaction (ICMI2014) Outstanding Paper Award ほか受賞。電子情報通信学会、人工知能学会各会員。博士(情報学)。



大塚 和弘 (正会員)

1995年横浜国立大学大学院工学研究科修士課程修了。同年日本電信電話株式会社入社。現在、NTTコミュニケーション科学基礎研究所主幹研究員。2007年名古屋大学大学院情報科学研究科博士後期課程修了。2010年スイス Idiap Research Institute 特別招聘研究員。コミュニケーション科学、コンピュータビジョンに興味を持つ。IAPR 10th Int. Conf. ImageAnalysis and Processing Best Paper Award, ACM 9th Int. Conf. MultimodalInterfaces (ICMI2007) OutstandingPaper Award, H22年度電信情報通信学会論文賞, 同喜安善市賞, 経産省 Innovative Technologies, 同特別賞 (2012), ほか受賞。電子情報通信学会, IEEE 各会員。博士(情報科学)。



熊野 史朗

2009年東京大学大学院情報理工学系研究科博士課程修了。同年日本電信電話株式会社入社。現在、NTTコミュニケーション科学基礎研究所研究員。表情認識や共感状態推定等、コンピュータビジョンや感性情報工学に関する研究に従事。8th Asian Conf. Computer Vision (ACCV2007) Honorable Mention Award 等, 各賞受賞。情報理工学博士。電子情報通信学会, IEEE 各会員。



大和 淳司

1990年東京大学大学院工学系研究科精密機械工学専攻修士課程修了。1998年マサチューセッツ工科大電気工学及びコンピュータサイエンス学科修士課程修了。1990年日本電信電話株式会社入社。画像認識等の研究に従事。

NTT ヒューマンインタフェース研究所，MIT 人工知能研究所，NTT 第三部門を経て，現在，NTT コミュニケーション科学基礎研究所協創情報研究部部長，兼インタラクション対話研究グループリーダー。IEEE，ACM 各会員。博士（工学）。