4-3　非言語に着目した対話時のインタラクション解析

4-3 Interaction Analysis at the Dialog by Nonverbal Behavior

善本　淳　　水上悦雄　　山下耕二　　矢野博之
YOSHIMOTO Jun, MIZUKAMI Etsuo, YAMASHITA Koji, and YANO Hiroyuki

要旨

我々が対面対話を行う場合、人のコミュニケーションの歴史と共存している非言語動作が、重要な情報を伝え、対話を調整するということはよく知られている。しかしながら、科学的な手法を用いた非言語動作の研究は、近年始まったばかりである。ビデオカメラが接続された計算機を用いて、会話の活性度を測定するためには、例えば、発話権の維持や交代というレギュレータを検出する必要がある。本文にて二者一組の被験者が対話を行っている様子のビデオ動画を計算機に取り込み、非言語動作の自動的な分割及びクラスター分析による分類方法に関して提示する。また、分類を評価する一例として、特定の動作グループが、ターン維持の動作を示唆していることに関して議論を行う。

We much know nonverbal behaviors that coexist with our communicative history may tell us important information and regulate our verbal dialogue when we have face to face conversed. Researches for nonverbal behaviors in human interaction were begun by scientific ways, however, quite recently. In order to measure the activity of conversation by the computer with video cameras, it is necessary to detect regulators that called turn-maintaining cues or turn-yielding cues. This paper presents a method of automatic segmentation and classification of nonverbal behaviors in dialogues captured on video from two subjects by cluster analysis. As an example for evaluating a classification, we discuss the specific group of behaviors had suggested turn-maintenance cue.

[キーワード]
非言語動作，レギュレータ，ターン維持
Nonverbal behavior, Regulator, Turn-maintaining

1　まえがき

非言語という単語が持つ多義性・曖昧性を回避するため、まずは緒言として対面コミュニケーションにおける非言語の役割等について以下で解説し、その後、研究の内容について論じることにする。

1.1　対面コミュニケーションで生じる三つの層

我々が他人とコミュニケーションを取る場合、そこで相互に発信・受信する情報はどのような構造を形成しているであろうか？　対面コミュニケーションを考えた場合、情報は三層構造を呈していると考えられる。

まずは言語層がある。これは、文法規則に沿うように並べられた単語による、情報の中核部分を形成するものであり、これにより他者に情報を伝達・共有することが可能となる。我々が一般的に読み書きする対象でもあり、容易にテキスト化が可能なことも、この層を特徴付ける一つの側面であると言える。

次の層はパラ言語層である。この層は発話行為により、言語層に付随して露見する層である。言い換えれば、発話によって生じる音声情報から言語層を抜いた部分であり、声の高さとその抑揚、声の大きさ、話速、間などが挙げられる。電話な

どの音声コミュニケーションツールを用いれば、言語層とパラ言語層からなる複層情報を、他人と相互交換する事が可能である。一般的に、書記行為と比べ、何ら道具を要せず、習得も容易なために、利用者の身体的負担は少ない。パラ言語層では、言語層の修飾や、発信者の状態や意図を伝えることが可能で、高モダリティー通信には欠くことのできない層である。しかしながら、言語層以外には曖昧性があり、発信者が意識的に発信しても受信されなかったり、発信者が無意図的に発信しても誤解して受信されたりすることがある（例えば、発信者が大きな声で強めに発話したつもりでも、受信者には小さな声であるととらえられたり、発信者が普通の話速で話しているつもりでも、受信者には早口で話していると誤解され、急いでいる等の印象を与えたり、というのはよくあることである）。同一言語圏、同一文化圏で共通の意味が存在する単語と比べ、意図や意味の不明確さ、発話内や発話間での相対的な比較に基づくことに由来する上述例のような少なからぬ誤解の必然的な内包、個人差、困難なテキスト化等のため、コミュニケーションに重要である層であることは理解されているが、古くから積極的に研究されてきたテーマではない。また、実際の対話を通じてパラ言語層の利用法を学習するために、話者の地域性が強く生じることも否めない。例えば、日本における方言は、語彙差とともに、アクセントを含むイントネーション差が大きいことは自明であり、これらの差異により地方在住者の音声認識を困難にさせている原因の一因にもなっている。

最後に非言語層がある。対面対話を行っているときの視覚的な情報がそれである。この非言語層には、発信者の容貌や服の印象までをも含めることがあるが、本文では発話中、または相手の話を聞いている場合の動作を主に扱う。パラ言語層の表出パターンである、いわゆる「話しぶり」と、動作表出パターンを組にして扱うポスチャという概念も非言語研究のテーマであるが、本文では前述のとおり、動作を主に扱う。この層は、パラ言語層と比較して、対象範囲が広く、複雑な層であるため、その理解には多くの困難を伴い、テキスト化がより困難である。一方、この非言語層が持つ意味性は、場合によっては言語層と同程度まで引き上げることが可能である。聴覚障がい者や音声言語を発せない人を中心に用いられている手話は、まさに発話行為の代替表現として機能している。逆に、表情一つで音声発話の意味を逆転させることも可能であることや（例えば、「ありがとうございます」と穏やかな口調で発話を行ったとしても、目や顔が笑っておらず、拳を握った手や肩が震えつつの発話など）、また、対話における流暢なターン交代などを担っているレギュレータを考えれば、パラ言語層と同様以上に、コミュニケーション及びインタラクションに重要である層であることは理解されている。しかしながら、近年までは散発的な研究は存在するが、伝統的に研究されてきたテーマではない。

1.2 非言語研究の経緯と今現在抱えている課題

我々は、我々の使っている非言語を詳しく知っているとは言えないだろう。非言語は人のコミュニケーションの歴史とともにあったと考えられるが、例えば、ジェスチャと挨拶の研究が始まったのは 19 世紀後期であり、非言語のコミュニケーションへの影響に関する正式な研究が始まったのは 20 世紀後半以降である。例えば、米国で非言語コミュニケーションに関する書籍が初めて出版されたのは 1972 年であるといわれている（これら非言語研究における歴史の解説は文献[1][2]による）。1970 年の日本万国博覧会で、非言語を交えたコミュニケーションが可能なテレビ電話が展示され、それから時間が経過して現在に至るが、今日非言語を交えた通信を一般的な国民は十分に堪能しているとはいえないだろう。非言語情報を補う技術開発は日々進められているが、各非言語情報が対話の場でどのように利用されているのか、また、それらが相手にどのような影響を与えているのかが、未だ十分には解明されていない。人と人とのコミュニケーションにおける、各非言語情報要素の詳細な運用、認知機構の解明及びそれら要素同士のコミュニケーション上での結びつきがコミュニケーションの性質に与える影響等の解明が求められており、同時に、それらのシステムへの応用が重要課題となっている。

1.3 非言語動作の分類に関する研究経緯

ボディーランゲージという概念[3]が研究者のみならず一般的な人々の中にまで浸透したが、非

言語は、言語ニ非ズ、と元より言語として破綻しているように見える。Birdwhistell によって動作学に対する構造的なアプローチ[4]があり、言語学に類似した分類手法を駆使し、非言語動作を細かく分離・分類するという手法が試みられた。異音 allophone に対して allokine、音 phone に対して kine、音素 phoneme に対して kineme 等を準備し、意味をなさない微小な音が複数連結して意味のある発話を構成するように、意味をなさない微小な動作が複数連結して意味のある動作を構成するものとして分類を行おうとした。しかしながら、その手法には反論[5][6]があり、多くの研究者に受け入れられたとは言えなかった。言語学で行われるような分類手法のみによって、すべての非言語行動を分類することは、困難が多い。

むしろ、非言語動作が発生した会話の背景抜きでは、非言語動作を分類できないという議論がある。Ekman と Friesen は、非言語動作を構造的に分類するのではなく、その動作の目的、意味、意図によって幾つかの基本的な分類を行うという、動作学に関しての外部変数的なアプローチ[7]–[11]を提案した。この方針は現在、研究者の間で広く受け入れられている。

1.4　本文における非言語動作の分類方針

前述の分類を念頭に、Birdwhistell、Ekman と Friesen らの長所を利用し、非言語動作の分類を行うべく以下のように方針を定めた。まずは、対話時の動画を記録し、そこからある一定閾値よりも大きな動作を抽出した。次に、抽出された動作同士を、表出時のパラ言語層を考慮しながら総当たりで比較し、似た動作同士を同一のカテゴリーとしてまとめて扱った。ここまでは Birdwhistell の提案に近いために同様の問題点を含んでいるが、単純な構造を持つ動作ならば、特に問題はないと考えられる。構造的なアプローチによる大きな問題点は、あまりにも細かく動作を区切ってしまった点と、動作は発話と異なり、明瞭な動作区切りが存在しないが故に、分割が困難な点にある。最後に、Ekman と Friesen らの提唱した非言語動作分類に合致しそうな特徴的なカテゴリーを吟味し、これらの分類手法の正当性を問う。

非言語の基礎的研究として、非言語の自動分類を試みるということは大きな意義があると考えられる。本文では以下、**2** で対話実験を説明し、**3** で取得されたデータの処理方法を述べる。

2　対話実験

2.1　実験準備

対話における非言語の基礎的なデータを得るために、二者一組となる被験者を用いて対話を行わせ、課題を遂行する際の対話過程を記録した。各被験者は互いに見聞きできないように壁等で囲まれた個室に入り、ビデオカメラとモニタ、マイクとヘッドホンを通じて相手被験者とインタラクションを行った。被験者の正面にはビデオカメラが設置され、被験者同士は互いの正面画像をモニタで確認しつつ対話を行うことになるが、他に側面から撮像するためのビデオカメラも設置した。これらのビデオカメラやマイクから収録された動画や音声をいったん保存し、後日計算機上に取り込み、処理を行った。収録された動画データは両被験者の正面画像及び側面画像であった。また、個室が互いに独立しているため、二者の発話はそれぞれ独立したチャネルにて収録した。

2.2　処理方針

動画は NTSC 方式で収録されたため 29.97 fps であった。極めて短時間の動作表出は相手が見落とす可能性が高く、故に本来的に見せる意図が低い点と、ある動作を見聞きし、それに応じる動作を表出するまでの反応潜時がおよそ 200〜400 ms である点の二点を考慮し、解析に用いるフレームレートは 7.49 fps（1 フレーム当たり約 133 ms）とした。これは、インタラクションにおける動作の伝搬や同時発生を考察するのに適したフレームレートであると考えられる。

3　データ処理

3.1　非言語動作の自動分割

得られた動画の各フレームにおいて、各被験者に対して測定する領域を設定し、その領域内を対象として、あるフレームとその次のフレームとの間に生じる輝度差の総量を、1 フレーム間当たりの被験者の移動量とし、移動量が一定閾値未満の場合は静止状態として処理した。図 1 は、収録さ

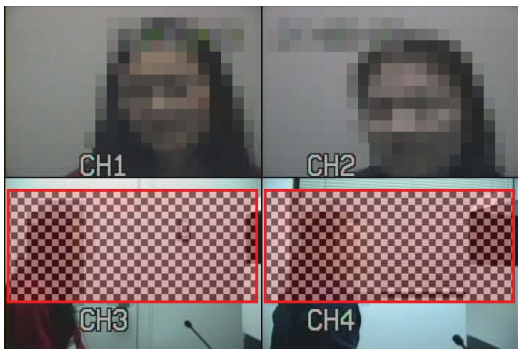

図1 収録された動画例

上段に被験者両名の正面図、下段に側面図を配してあり、下段の格子状部位は動作検出対象領域を示している。

れた動画の、あるフレームにおける画像例である。側面画像中の格子状の部位は、動作検出対象領域を示しており、この領域内での輝度が変化すれば、移動が生じていると判断した。時間軸上で移動のない状態、すなわち、静止状態に挟まれ、かつ、一定長以上の移動が連続して生じた場合、一つの動作が発生しているとみなした。

図 2 は、ある被験者組の 241 秒間の対話中に表出した動作を自動分割した非言語動作チャート例である。被験者別に対数化した移動量を縦軸に、時間軸を横軸にして図示した。各動作に通し番号を与え、各被験者にとっての奇数番目と偶数番目で描画濃度を変え、一べつで動作単位の理解が可能となるように工夫を施した。

また、図 2 では動作のみが表記されているが、図 3 では同一時間軸上に発話声量も表記し、正面から収録した対話動画と同期させた音声付動画を制作するシステムの開発も行った。これにより、静止画のみならず、動画でも対話状態を俯瞰することが可能になった。

3.2 非言語動作のクラスター分析による分類

Ekman らの提案に従うならば、分割された動作を一つ一つ手作業で吟味し、エンブレム、イラストレータ、レギュレータ、アフェクトディスプレイ、アダプタという 5 種類のカテゴリ[7]–[12]、のいずれかに帰属させるべきである。しかしながら本文ではそれを行わず、動作の特徴を元に大略的に、かつ、自動的に分類する方法で分類を試みた。この方法を用い始めた初期では動作ごとに、その動作の継続時間、移動量、動作中の音量を変数として準備した（初期型変数群）。次に、各変数は単位が不揃いであるために標準化処理を行った後、被験者ごとに動作のクラスター分析を行った。図 4 は、ある被験者の動画から得た 56 個の動作を元に、典型的なクラスター化法 UPGMA[13] を行った時に生じた樹形図である。個別の動作や動作クラスター同士が低い位置で結ばれていれば、それら諸動作は類似性が高く、反対に、高い位置で結ばれていれば、それら諸動作は類似性が低いことになる。図 4 の左半分には低い位置で結ばれた動作が数多く見られるが、これらは互いに類似した動作であることを示している。

図 5 は、114 秒の対話を行っている、ある被験者組の分類用符号付き非言語チャート例である。図 2 と同様に、被験者別に対数化した動作の移動量を縦軸に、時間軸を横軸にした図示を行い、さらにその上に、各動作が帰属するグループが理解できるように分類用の符号を重ねて図示した。分類されるグループ数は、クラスター分析で生じる樹形図の任意の高さで区切ることによって変更が可能であり、今回は 10 グループに分類した。初期型変数群で類似度が高い、すなわち、見かけ上よく似た動作は、同一グループに属した。同一被験者において、同じ高さの矩形符号で示唆された動作は、他の同じ高さの矩形符号で示唆された動

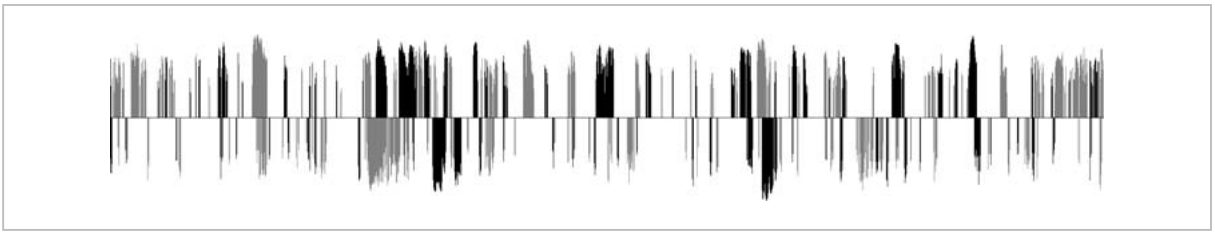

図2 非言語動作チャート

上段が被験者 A、下段が被験者 B の動作であり、縦軸は移動量、横軸は時間を表している。

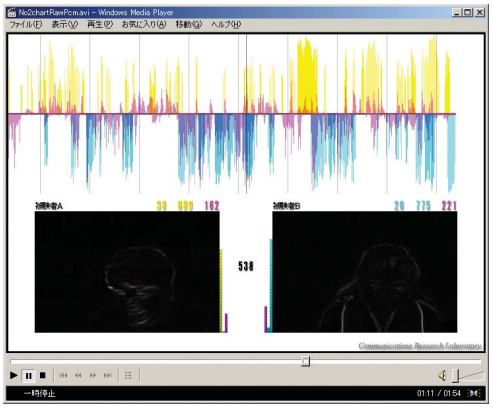

図3　対話俯瞰動画

上段は図２に発話声量を加えた図、下段は被験者正面図において、移動が生じた部位を光量で表している。

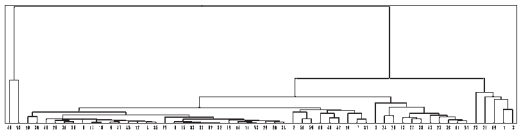

図4　頭部動作を分類する樹形図

動作は通し番号で示され、下位で接続されている動作ほど類似の動作であり、上位で接続されている動作ほど異なる動作である。

作と同一グループであることを示唆している。これは、楽譜における音符のアナロジーである。同じ高さの音符は同じ音の高さを意味するように、同じ高さの矩形符号は同じグループの動作を意味している。このようなアナロジーは一般的であり、筆者らの発表[14]のほかに、例えば、表情を楽譜のように高低で表す研究[15]も存在する。

3.3　分類された動作の検討

動作発生時と動作終了時がほぼ同じ状態（姿勢）である動作を閉動作、反対に異なる状態である動作を開動作と定義する。開動作は、ある動作の途中で停止状態を挟むことなどによって発生し、実際には不連続ながらも一連の長期動作であるが、本手法では二単位以上に分割されてしまうため、以降の議論では閉動作のみを対象とする。開動作に対応しつつ動作を解析する方法も存在するが、現在の所その方法は分析時に実装していない。

クラスター分析を行う際、その分類精度は、準備する変数群に左右される。変数群は増やせば増やすほど精度が上がるというものではなく、むしろ何の動作を高精度に分離したいのかによって、準備する変数群を定め、変数の数を絞ったほうが良い結果をもたらすことがある。動作を大略的に分類するための初期型変数群では、見かけの動作分類には適していたが、レギュレータの分類という点では、初期型変数群はあまり適した変数群ではない。

レギュレータ用変数群として、動作の継続時間、動作発生時の対象者や相手の相対的な平均発話比率、動作発生時の対象者や相手の平均単独発話比率、動作中の音量総和を準備し、その変数によってある被験者の対話時の動作をクラスター分析したところ、それによって導かれた樹形図（図 6a 参

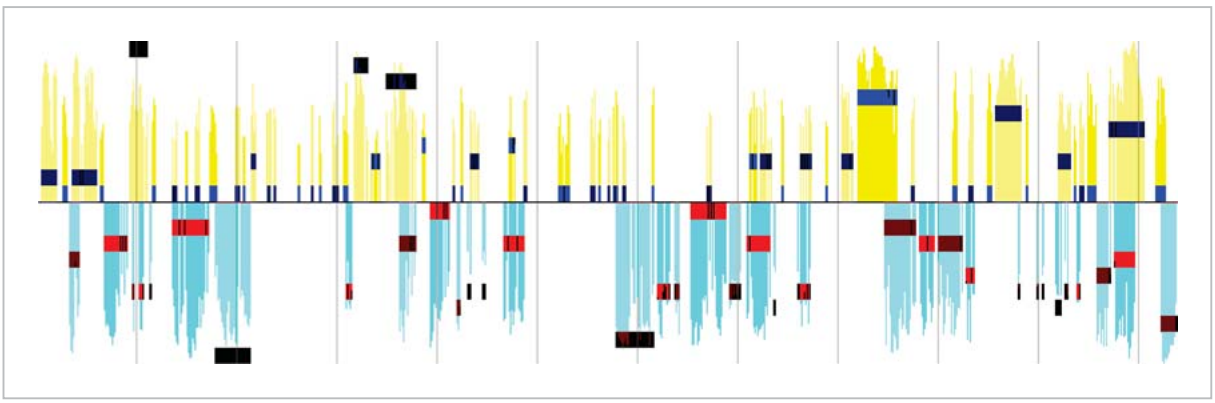

図5　非言語チャートに分類用の符号を重ねた例

図２の表現を基本に、同一グループの動作を同一高で示す分類符号を重ねた図。上段の被験者 A を例にすると、中心軸に最も近い位置の符号が付けられた動作は、主に頷き動作として観察された。

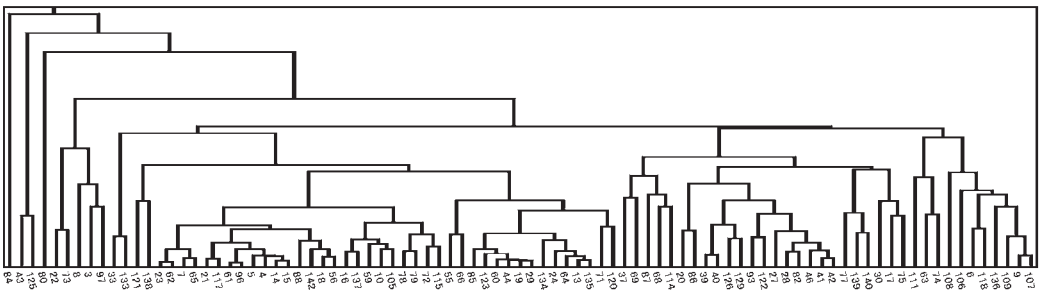

図6a 閉動作のみの頭部動作を分類する樹形図

図４と同じ様式であるが、145 動作中、81 の閉動作のみを対象にクラスター分析を行ったときに生じた樹形図（全体図）

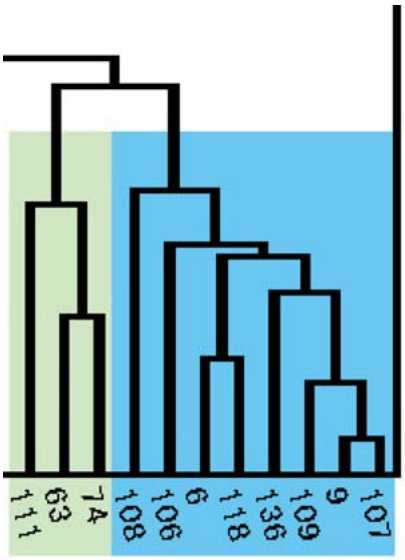

図6b 樹形図（対象拡大図）

表1 対象被験者の動作と付随した発話内容

動作#	動作長 [frame]	直前の相手被験者の発話	対象被験者の発話
6	20	何の根拠もないけど	んー
9	24	うん	えー
106	12	なんか人物１が	んー
107	24	店の経営者で	んー
108	23	なんか話してそれに対して	んー
109	32	人物２が	んー
118	22	経営者は結構裏、 裏で	んー
136	37	今１の方は	んー

照）の１グループにスポットを当て（図 6b 参照）、レギュレータ分類の成果を以下で問うことにする。

この例は、ある日本人被験者組による 251 秒間の対話であり、対象となった被験者は 145 動作を生じ、その内 81 動作が閉動作であった。さらにこの内の 8 動作（動作番号 6，9，106，107，108，109，118，136）に関し、直前の対話内容（言語層）とともに表 1 にまとめた（前述 **2.1** で述べたように、二者一組の被験者は課題を与えられており、その課題を二人共同で解いている場面が収録された動画と音声をここでは用いた。表 1 の対話での発話内容は、その課題に即したものとなっている。この時、両被験者は「ある二人の人物写真」を見ながら、写真の人物のどちらが経営者であり、どちらが店員であるかを推測するという課題[16]が与えられていた）。対象被験者は動作発生中に、「んー」や「えー」の発話を生じており、また、その発話直前の相手被験者の発話を考慮すると、このグループに分類された動作は、ほぼターン維持のレギュレータであることは明白である。

4 むすび

Ekman らが提唱した動作のカテゴリーの中で、単純な構造を持ち、類似性が高いような動作に配慮して変数を準備すれば、Birdwhistell が提唱した構造的な視点の一部を取り入れつつ、自動分割及び自動分類が可能であることを示した。これにより、人間の情報のやり取りの基本である二者一組の対話を計算機を用いて観測させることで、音声と動作の特質からその対話状況を知ることも可能になった。

対話状況理解の精度の向上や、対話時の二者の状況をより深く理解するためには、言語層・パラ言語層における対話処理も必須である。紙幅の関係上詳細に触れなかったが、二者間の合意形成対話における同意表現等の研究[17]、二者間対話の間（ま）、音声パワーのリズムに着目し、発話の仕方の適切性をモデル化した研究[18]、フィラーや情動的感動詞を発話者の心的状態を類推するための心的マーカーとして抽出・分類した研究[19][20]も平行して行った。これらの成果もまた、対話理解に貢献した。

参考文献

1 A. Kendon, "Gesture: Visible Action as Utterance", New York; Cambridge University Press, 2004.
2 V. P. Richmond and J. C. McCroskey, "Nonverbal behavior in Interpersonal Relations", Allyn and Bacon, 2003.
3 R. L. Birdwhistell, "Kinesics and Context: Essays on Body Motion Communication", Philadelphia: University of Philadelphia Press, 1970.
4 R. L. Birdwhistell, "Introduction to Kinesics: An Annotation System for Analysis of Body Motion and Gesture", Louisville, KY: University of Louisville Press, 1952.
5 A. T. Dittomann, "Review of kinesics and context by R. L. Birdwhistell", Psychiatry, 34, 34-342, 1971.
6 V. P. Richmond, "Nonverbal Communication in the classroom", Acton, MA: Tapestry Press, 1996.
7 P. Ekman, "Movements with precise meanings", Journal of Communication, 26, 14-26, 1976.
8 P. Ekman and W. V. Friesen, "Nonverbal leakage and clues to deception", Psychiatry, 21, 88-106, 1969.
9 P. Ekman and W. V. Friesen, "The repertoire of nonverbal behavior: Categories, origins, usage, and coding", Semiotica, 1, 49-98, 1969.
10 P. Ekman and W. V. Friesen, "Hand movements", Journal of Communication, 22, 353-374, 1972.
11 P. Ekman and W. V. Friesen, "Detecting deception from the body or face", Journal of Personality and social Psychology, 29, 288-298, 1974.
12 K. R. Scherer and P. Ekman, "Handbook of Methods in Nonverbal Behavior Research", New York; Cambridge University Press, 1982.
13 H. C. Romesburg, "Cluster Analysis for Researchers", Florida: Robert E.Krieger Publishing Company Inc., 1989.
14 善本淳，矢野博之，“対話動画像中の頭部動作のクラスター分析”，電子情報通信学会総合大会，2004.
15 M. Nishiyama, H. Kawashima, T. Hirayama, and T. Matsuyama, "Facial Expression Representation based on Timing Structures in Faces", IEEE International Workshop on Analysis and Modeling of Faces and Gestures (W.Zhao et al. (Eds.): AMFG 2005, LNCS 3723), pp.140-154, Oct. 2005.
16 D. Archer, "How To Expand Your S.I.Q.(Social Intelligence Quotient)", New York: M. Evans and Company Inc., 1980.（邦訳：工藤　力，市村英次，“ボディ・ランゲージ解読法”，誠信書房，1988.）
17 矢野博之，善本　淳，“合意形成対話における同意表現の言語・非言語情報の分析”，人工知能学会SLUD研究会，SIG-SLUD-A203-07，41-46，2003.

18 E. Mizukami, "How the Conversational Rhythm of 'MA' can be Constructed in Japanese Dialogue", In Proceedings of The 8th World Multi-Conference on Systemics, Cybernetics and Informatics, 14, 3-8, 2004.
19 E. Mizukami, K. Yamashita, and H. Yano, "Effects of Modality and Familiarity on Dialogue to Describe a Figure: Analysis of Speech Fillers", Progress in Asian Social Psychology Series, 6, 343-358, 2007.
20 山下耕二，水上悦雄，“心的マーカーによる心的処理プロセスの理解－図形説明課題対話におけるフィラーを中心とした分析－”，自然言語処理，14(3)，39-60，2007.

善本 淳
知識創成コミュニケーション研究センター音声言語グループ研究員（旧情報通信部門けいはんな情報通信融合研究センター社会的インタラクショングループ研究員） 博士（学術）
計算機科学、非言語コミュニケーション

水上悦雄
元情報通信部門けいはんな情報通信融合研究センター社会的インタラクショングループ長期専攻研究員
博士（理学）
フィラー、多人数インタラクション、相互行為解析

山下耕二
元情報通信部門けいはんな情報通信融合研究センター社会的インタラクショングループ長期専攻研究員 博士（人間科学）
認知心理学、教育工学、コミュニケーション（非言語、メディア）

矢野博之
総合企画部企画戦略室プランニングマネージャー（旧情報通信部門けいはんな情報通信融合研究センター社会的インタラクショングループリーダー）
博士（工学）
対話処理、対話の認知モデル