ヒューマンインタフェース 84－5
情 報 メ デ ィ ア 35－5
（1999．8．20）

ヒューマンインタフェースのための人間の振舞いの解析 –マルチモーダル対話データの解析-

綿貫 啓子、関 進、三吉 秀夫
RWCPマルチモーダル機能シャープ研究室
シャープ（株）システム開発センター内

ビデオカメラ、マイク、およびモーションキャプチャシステムを備え、音声・画像と人体各部位の位置の3次元数値データを収集できるマルチモーダル対話データ収集システムを構築し、対話データを収集した。また、このデータを基に、人の対話過程での3次元空間上の頭部、胴体、手の動きの解析を行った。頭部、胴体、手の動きはいずれも、聞き手の時よりも、話し手の時により強く現れる傾向がある。また、その傾向が手の動きにより強く現れていることから、手の動きが発話内容そのものに関与しており、その機能が頭部や胴体と異なると考えられる。

Analysis of Human Behavior for Developing User Interface – Based on Multimodal Interaction Data –

Keiko Watanuki, Susumu Seki and Hideo Miyoshi
RWCP Multimodal Functions Sharp Laboratory
System Technology Development Center, Sharp Corporation

We have implemented a data collection system which utilizes motion capture system as well as video and audio recording equipment, to capture verbal and nonverbal information in interpersonal communications. Using this system we have collected conversation data between two persons, and analyzed the data with respect to head, body and right hand movements. This paper introduces the Multimodal Interaction Data Collection System, and presents some data showing that hand movements have different functions from head and body movements.

1. はじめに

コンピュータと人間が円滑で快適な対話を行うためには言語情報のみならず非言語情報も重要な役割を果たすとの考えに基づき、人間同士の円滑な対話を手本として、新しい対話型のユーザインタフェースの実現を目指している。人間同士の対話データを解析した結果、これまでに、対話におけるタイミングや間（ま）など時間情報の重要性を見い出してきた[7]。また、それが如実に示される相槌を、時間を取り扱うのに適した対話処理モデルによって近似し、その有効性を示してきた[6]。これを人間同士で行われているような双方向に複数の情報が伝達する自然な対話機能を持つシステムに拡張していくには、自然な対話時にマルチモーダル情報がどのように交換されているかを調べることが必要である。そこで、対話過程に現れるマルチモーダル情報の特性を調べるため、音声・画像及び人体各部位の位置の3次元数値データを収集できるマルチモーダル対話データ収集システムを構築し、対話データを収集した。このデータを基に、人の対話過程での3次元空間上の体や頭の動き情報の解析を行った結果、頭部、胴体、手の動きが発話時にそれぞれ特徴的な振る舞いをしていることを確認した。

本稿では、マルチモーダル対話データ収集システムの概要を紹介するとともに、本システムで収集した対話データの解析を基に、対話過程における人間の特徴的な振舞いについて述べる。

2. マルチモーダル対話データの収集環境

人間同士は言語情報だけでなく、身ぶりや手振り、表情など、言葉以外の情報（非言語情報）も用いてコミュニケーションしている。このような人と人の間で交される対話過程を

解析するために、我々はこれまで、音声と画像のデータを含む対話データを収集し、利用してきた[2]。しかし、あいづちなどの頭の動きや手の動きなど、身体の動きの情報にタグをつけるには、VTR映像を何度も再生しながら人手で行なっていたために膨大な労力と時間を要していた。頭の動きや手の形、動きの方向などをコード化して記号で記述する試みもあるが[4]、観察者によりタグの付け方にばらつきが出るという問題点がある上に、動きの強度をコード化することが困難であった。

そこで、従来の対話データ収集システムに、光学式のモーションキャプチャシステムを導入し、人手をほとんど介さない客観的なデータを得ることができるようにした[8]。本システムは、着座して対話をしている被験者の様子を収録するもので、音声・画像データと共に、赤外線カメラで被験者の身体につけたマーカー（図 1）を光学的にとらえて、その3次元数値データを自動的に収集する。対話しているときに身を乗り出したり、強調の動作をしたり、物の形や大きさの説明をしたり、あるいはうなずいたりしているときのマーカー位置の情報が得られる。これまで人手で行っていた動きの解析を効率的かつ客観的に行うことが可能になった。

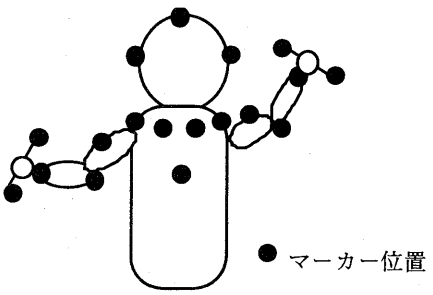

図 1　マーカー装着位置

3.マルチモーダル対話データ

マルチモーダル対話データ収集システムを用いて、これまでに、自由対話をタスクとするデータを収集している。マーカーを装着した状態での対話であるので自然な対話が行なわれるか危惧したが、被験者はすぐに順応し、ほとんど支障はなかったと思われる。

収集したマルチモーダル対話データには、以下の情報が含まれる：

●画像データ：1 フレーム=1/30 sec.

●音声データ：サンプリングレート＝8kHz

●2次データ：

音声パワー：1 フレーム=1/30 sec.

音声ピッチ：1 フレーム=1/30 sec.

人体各部位の位置（3次元座標）情報：
1 フレーム=1/60 sec.

テキスト情報（発話内容書き起こし）

図 2は、モーションキャプチャシステムによって抽出された位置データの1例である。これは、ある発話区間での頭部のマーカーの位置データで、位置座標が、x 軸、y 軸、z 軸ごとに表示されている。ここで、x 軸は被験者の横方向、y 軸は上下方向、z 軸は前後方向である。

さらに、モーションキャプチャシステムによって、人間を剛体としてモデル化し、たとえば、頭部につけた 3 つのマーカーの位置データから回転角も計算できる。これらの位置や回転角を分析することにより、身体の各区域の動きの速度や角速度、加速度等を計算により求めることができ、うなずきの大きさや速さを解析することが可能になる。

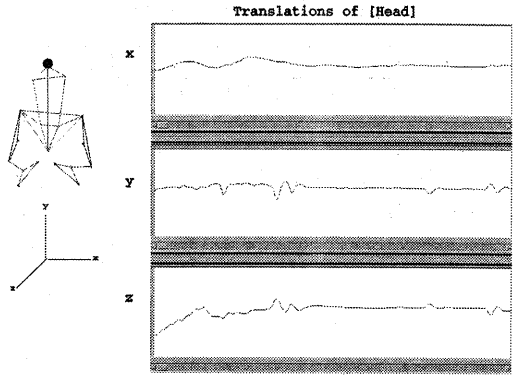

図 2　位置データの例

4.対話における動作の解析

非言語情報がどれほどコミュニケーションに重要な役割を果たしているのかを明確に計測した研究はないが、人の間で交されるメッセージの 65%は非言語が担っているという報告がある[3]。また、発話交代において、頭や手の動き、視線が重要な役割を果たしていることがわかっている[1][5][7]。さらに、頭の動き（縦振り、横振り）の大半があいづちと発話順番末に現われるとする報告がある[5]。手の動きは、その核(stroke）となる部分の 90%が発話とともに出現するとの報告もある[4]。このように、発話交代やあいづちといった、対話のある時点での機能に着目して動作の重要性を論じた研究はあるが、対話全体を通じての身体の動きを定量的に扱うことは難しかった。そこで、本稿では対話を担っている話し手と聞き手の観点から、人間の振舞いを解析した。

4.1.解析データ

活発で自然な対話の様子を解析するために、日頃よくおしゃべりをする親しい2人のひと同士の自由対話を収集し、表1に示す 3 組の20台女性被験者（計 6 人）のデータを解析の対象とした。

表1　解析対象データ

データ	データ 1		データ 2		データ 3	
データ長	300sec.		600sec.		600sec.	
被験者	SNF	MOF	MTF	HUF	YTF	YHF
話し手区間 (sec.)	165	141	375	230	261	269

4.2.解析方法

4.2.1 音声情報のラベリング

まず、各被験者の発話について、音声パワー情報を利用し、10フレーム（約333msec.）以上の無音区間で区切られた発話区間を基本単位(utterance unit = UU)としてラベリングしたのち、さらに発話権を保持している連続したUU区間を話し手区間としてラベリングし、これを解析の対象とした（図 3参照）。なお、「ああ」「うん」などのいわゆるあいづちや、「ホント」などの感嘆詞、独り言のようなオウム返しは、発話権を獲得していないと見なして、話し手区間とはしなかった。また、一方が発話権を放棄したにもかかわらず、他方が発話権を獲得しないために、両者とも無音区間になる状況が発生することがあるが、このような状況で発せられる「うん」などもやはり発話権を獲得していないと見なして、話し手区間から除外した。なお、ここで、聞き手区間とは、話し手区間で挟まれた区間をさし、必ずしも相手が話し手区間であるわけではない。

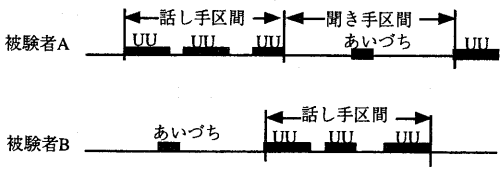

図 3　話し手区間

4.2.2 動き情報の抽出

本稿では、頭部、手（右手）および胴体の動きの大きさを解析対象とした。なお、身振

り・手振りとしては、その向き、置かれた位置、形状などの静的な情報と、その動きの大きさやタイミングなど、動的な情報とがあるが、今回の対話データの解析では、特に時間が重要なはたらきを果たす速度・角速度の情報を利用することにした。

ところで、本データでは、被験者は着座して対話している。したがって、身体の構造から鑑みると、頭部は剛体であるため、首の関節（頚椎）の一点を中心とした動きをすると見なせるので、動きの特徴を調べるには角速度が有効な数値であると考えた。また、胴体は、一塊にして剛体と見なし、単純化して扱うことにしたため、頭部と同じように回転角が重要になってくるので、やはり、角速度が有効な数値であると考えられる。一方、手（掌）の動きは、肩、ひじ、手首の3つの関節の自由度を持つため、頭部や胴体に比べ移動量が大きい。そこで、よりプリミティブであり、表現力も高いと思われる速度の方が手の動きの特徴量をよく表すと考えた。このように、頭部、胴体については角速度の大きさを、また、手の動きについては速さを用いて解析することにした。

5. 解析結果と考察

5.3.1 発話と動きの関係

話し手区間と聞き手区間とでの、頭部、胴体の角速度の大きさ、および手の動きの速さの平均をそれぞれ表 2、表 3、表 4に示す。

なお、これらの動きには、頭の縦振りや横振りの他に、さまざまな様態が含まれる。たとえば、姿勢を直したり、髪を触ったりする手の動作も含まれる。また、横を向いたり、感嘆や笑いに伴なう身体全体の動きも含まれる。

表 2　頭部の角速度の大きさ（平均）

被験者	SNF	MOF	MTF	HUF	YTF	YHF
話し手 (10^{-2}rad /sec)	25.6	27.1	16.6	20.8	19.0	26.8
聞き手 (10^{-2}rad /sec)	13.5	22.1	13.3	15.9	17.2	22.9

表 3　胴体の角速度の大きさ（平均）

被験者	SNF	MOF	MTF	HUF	YTF	YHF
話し手 (10^{-2}rad /sec)	12.2	17.1	7.8	10.3	6.6	11.5
聞き手 (10^{-2}rad /sec)	7.5	14.3	5.3	8.1	7.4	9.0

表 4　手の速さ（平均）

被験者	SNF	MOF	MTF	HUF	YTF	YHF
話し手 (mm/sec)	53.8	65.3	33.9	37.1	26.5	36.7
聞き手 (mm/sec)	26.7	30.4	18.5	20.4	19.1	16.2

表2、3、4の結果は、被験者 YTF の胴体の角速度のデータを除き、すべての場合で、話し手の時の方が聞き手の時よりも動きが大きいことを示している。表1に示すように、被験者の各ペアにおいて、発話権を取っている時間は必ずしも均等ではない。被験者 SNF、MTF、YHF の方が長く発話権を握っていて、被験者 MOF、HUF、YTF はどちらかというと、聞き手役になっている時間が長い。それにもかかわらず、先に挙げた1例を除き、話し手の時の方が、より大きく身体を動かしている。

これらの数値について、それぞれ正規分布に従うと仮定し、話し手区間と聞き手区間で大きさに差があるかを検定した。

話し手区間と聞き手区間の差がないと仮説を立てて、自由度5の t 検定（片側検定）をおこなったところ、頭部（t=3.52）と手（t=5.18）については危険率 1%でも棄却できた。胴体（t=3.20）については、危険率1%では棄却できなかったが、危険率 5%では棄却できた。このことから、頭部および手に関しては、話し手区間と聞き手区間で明確な振舞いの違いがあり、胴体に関しても有意な差が出ていると言える。

さらに、表2、表3の結果と表4の結果を比較してみると、手の速さにおける話し手と聞

き手の差は、頭部や胴体の角速度の大きさの差に比べてかなり大きい。この結果は、手の動きが話し手区間でより多く出現する傾向があることを示しており、手は「話す」という動作そのものと深く関与していて、頭部や胴体と異なる機能を担っていることを示唆している。

5.3.2　話し手区間と動きの相関

前記5.3.1の結果から、頭部、胴体、手の動きがいずれも話し手区間と深く関わっていることが示唆されたので、このことを確かめるために、これらの動きの大きさと話し手区間との相関を調べた。

相関を調べるには、図 4に示すように、2番目の動きの大きさのデータを固定して、1番目の発話のデータをマイナス方向からプラス方向へ1フレームずつスライドさせて一致するかどうかを調べる方法をとった。

$$c(\tau)=\sum_{t}\frac{h(t-\tau)\cdot m(t)}{|h|\cdot|m|}$$

ここで、h は話し手区間で 1、聞き手区間で 0 の値をとる。m は速さや角速度の大きさである。すなわち、τが負では話し手区間前、τ=0 では話し手区間中、τが正で話し手区間後の関係である。

ここでは、各ペアにおいて被験者SNF, MTF, YHFの結果を図5に示す。それぞれの特徴に関して以下の傾向が指摘できる。(1)頭部：個人差はあるものの、いずれの被験者についてもピークが見られることから、話し手区間と頭部の動きには相関があると考えられる。また、ピークの山が、話し手区間前も後も急である。すなわち、話し手区間が始まる近辺に頭部の急な動きが始まり、話し手区間の終わり近辺で急に動きが終わっているものと考えられ、発話交代のシグナルになっている可能性がある。 (2)胴体：被験者MTF,YHFの場合は、ピークの山が左右非対称であり、話し手区間前が急で、後がなだらかである。話し手区間の始まり近辺で胴体がゆっくり動き始め、話し手区間の終了後も動き続けている可能性があるが、発話交代との関わりに関してデータ数を増やし、さらに詳細に解析する必要がある。(3)手：ピークの形にばらつきがあり、被験者MTFの場合はピークが見られない。前記5.3.1の結果では、話し手区間にかなりの手の動きが出現していることがわかっているので、この結果の解釈についてもさらに詳細な解析が必要である。

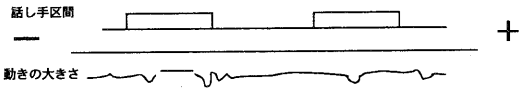

図 4　話し手区間と動きの大きさの相関

6. まとめ

データ数が3組と少ないので、今回の報告はまだ推測の域を出ないが、以下の結果を得た。第一に、頭部、胴体、手の動きは話し手、聞き手いずれの区間にも出現するが、話し手区間の方で、より大きい傾向があることが分かった。第二に、その傾向が、手の動きにより強く現れていることから、手の動きは、発話内容そのものに関与しており、その機能が頭部や胴体と異なることが示唆された。

今回は速さや角速度の大きさなど各時刻での動きの大きさと話し手区間、聞き手区間との関係について調べた。今後は、動き情報として速度や角速度など、向き情報も含めた取扱いをすることで頷きの抽出など詳細について検討が可能になると思われる。また、発話権交代と動きの関係をさらに詳細に解析していく。これら得られる知見はマルチモーダルヒューマンインタフェースシステムに反映していく予定である。

7. 参考文献

[1] Duncan,S., "Some Signals and Rules for Taking Speaking Turns in Conversations", J. of Personality and Social Psychology, Vol.23, No.2, pp.283-292, 1972.

[2] Kiyama, J., Watanuki, K. and Togawa F., "Multimodal Interaction Database and Analysis Environment", Proc. of 1997 RWC Symposium, pp.23-30, 1997.

[3] 黒川、ノンバーバルインタフェース、オーム社、1994.

[4] McNeill, D., *Hand and Mind*, Chicago: The University of Chicago Press, 1992.

[5] メイナード, S.K., *会話分析*, くろしお出版, 1993.

[6] 関 他: 力学系対話処理モデルを用いた相槌システム, 情報処理学会全国大会, 1999.

[7] Watanuki, K., Sakamoto, K. and Togawa, F., "Analysis of Multimodal Interaction Data in Human Communication", Proc. of ICSLP94, pp.899-902, 1994.

[8] 綿貫: モーションキャプチャシステムを用いた対話データの収集, 第3回社会言語科学会学会シンポジウム, 1999.

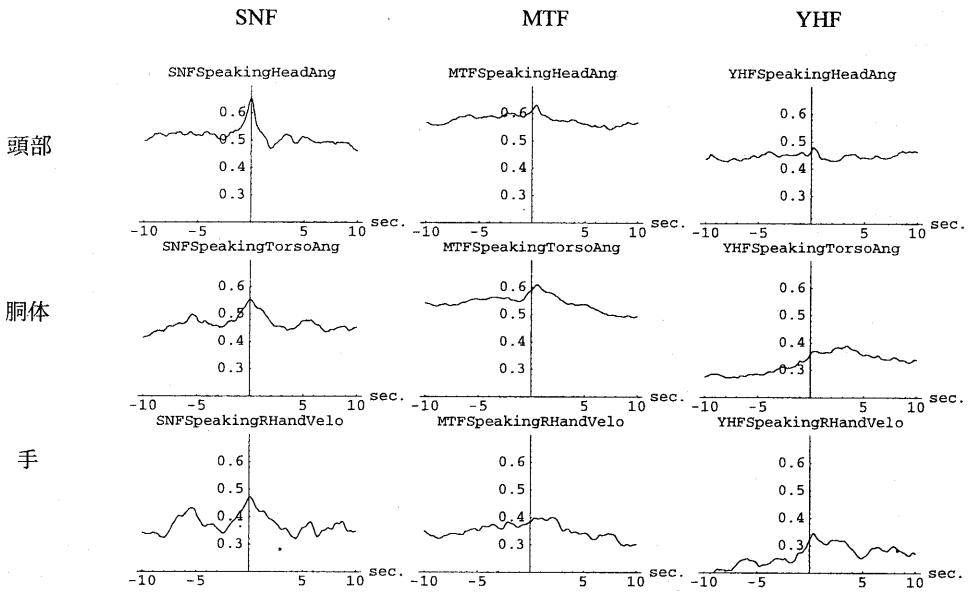

図 5 話し手区間と頭部・胴体・手の動きの相関