

対話における感情の変化の解析

綿貫啓子 外川文雄

RWC P 新機能シャープ研究室

シャープ株式会社 応用システム研究所

〒261 千葉市美浜区中瀬1-9-2

Email: watanuki@iml.mkhar.sharp.co.jp

対話の進行とともに、人の感情は、気分が乗ってきたり退屈したりとさまざまに変化し、その感情が表情や音声に表われる。このような感情の変化をとらえることを目的に、本稿では、二人の対話過程で、いわゆる話が乗ってきて気分が高まる様子を、マルチモーダル対話データベースを基に解析した。その結果、気分の高まりとともに、音声や動作に以下の変化が現われることを確認した： 1) 二人の発話量（時間）が等しくなってくる、2) 二人の発話の重複が多くなり、一方、沈黙の時間が減少してくる、3) 音声の平均ピッチが上昇する、4) 頭の縦振りとアイコンタクトが相補って現われる、5) まばたきの間隔が長くなる。また、二人の感情の度合いが同期し、協調しながら対話が進んでいくことが確かめられた。

ANALYSIS OF THE LEVEL OF INTEREST IN CONVERSATIONS

Keiko Watanuki and Fumio Togawa

Real World Computing Partnership

Novel Functions Sharp Laboratory in

Integrated Media Laboratories, Sharp Corporation

1-9-2, Nakase, Mihama-ku, Chiba 261, Japan

Email: watanuki@iml.mkhar.sharp.co.jp

When communicating with computers, users may display a variety of emotions in faces and voices. Thus, in realizing more flexible and natural communications between humans and computers, we consider that computers need to know about their user's emotional state: whether the user is interested or not.

In this paper, we describe some features which would convey the level of interest in the process of conversations: 1) the proportion of speaking time becomes nearly equal; 2) the proportion of time for utterance overlap increases, whereas the proportion of time for non-speaking (silence) becomes lower; 3) the mean F0 increases; 4) nodding and eye-contact serve to acknowledge the partner in a complementary way; and 5) the time intervals between blinks become longer. Further, we describe that there seems to be a shared level of interest between the two subjects in order to perform coordination.

1. はじめに

人とコンピュータの間に使いやすいコミュニケーション環境を実現するために、我々は音声や身振り、表情などを統合したマルチモーダル・ヒューマン・インタフェースの開発を目指している。コンピュータとの対話を目指した研究は従来から行われているが[1]、我々は、マルチモーダル・ヒューマン・インタフェースを実現するにあたり、人と人之间で交わされる対話過程を模範とするアプローチをとっている。人は相手の音声内容のみならず、身振り、表情といったものも利用して対話している。そこで人と人之間で交わされる対話を収集し、データベースを構築して、そのデータベースをもとに音声や身振りなどの情報を詳細に観察し、そこから得られた結果をインタフェースの設計に役立てる方法を提案してきている。

これまでに我々は、頭の振りや顔の向き、視線方向が話し手や聞き手の意図を表わしていること、また、これら頭の振りやあいつちのタイミングにおいて、二人が同期をとりながら協調して対話をしていることを述べてきた[5][6][7]。人間同士の対話ではさらに、感情もコミュニケーションされていると考えられる[8]。対話において、人の感情は、気分が乗ってきたり退屈したりと、さまざまに変化し、その感情が表情や音声に表われる。このような感情の変化を理解することができれば、人とコンピュータとの間に、より自然なコミュニケーションが実現すると考えられる。

本稿では、対話が進んでいくにしたがって感情が変化する様子を定量的に解析するため、2つの実験を行った。以下では、第2章で実験に用いたマルチモーダル対話解析システムについて述べ、次に、第3章で二人の自由対話における感情の変化を解析した結果を、第4章では、対話の途中で刺激を与えた前後の感情の変化を解析した結果を述べる。

2. マルチモーダル対話データ解析システム

我々は、人と人之間で交わされる音声やジェスチャなどを詳細に記録・解析できるマルチモーダル対話解析システム(図1参照)を構築した[4]。このシステムを用いて、対話に伴う音声やジェスチャなどを記録し、収集した対話データにラベリングを行ってデータベース化している。ラベリングデータとそれに対応する映像データを合わせてマルチモーダル対話データベースと呼ぶ。以下の解析では、このラベリングデータに基づいて解析を行った。

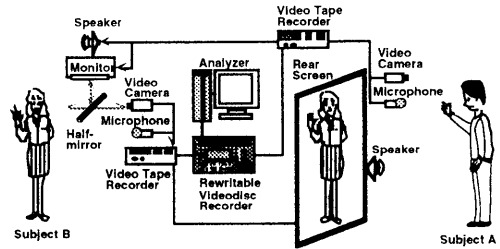


図1 マルチモーダル対話解析システム

3. 対話の流れと感情の変化

感情を表情や音声から解析しようという試みはこれまで多くなされてきている。感情の種類についてもさまざまな分類があり、たとえば、Ekman et. al. [2] は、喜び、驚き、恐れ、悲しみ、怒り、嫌悪、興味の7種類を基本感情としている。本章では、対話が進むにつれて話しが盛り上がっていく様子、すなわち、気分が高まる様子を、音声や動作の変化量で定量的に捉えようと試みた。

3.1 データ収集

対話過程において感情が変化する様子を解析するために、第1の実験では、二人に自由に対話してもらおうというタスクでデータを収集した。被験者は2組。各ペアは互いに相手をよく知っており、事前に興味あるテーマを選んでもらった。データ収集は、マルチモーダル対話解析システムで行い、互いにスクリーンを介して約10分間話してもらった。各々のペアが選んだテーマは、「パドミントン」と「秋葉原の店さがし」であった。

3.2 解析結果と考察

このようにして得られた10分の対話において、解析の便宜上、どこで被験者の気分が高まっているかを示す指針を得るため、対話の構造をいくつかの話題の集合とし、各話題は「導入部」「継続部」「終了部」の3つの部分から構成されるものとした(図2参照)。すなわち、選んだテーマについて、まず第1話題の導入から始まる。たとえば「秋葉原の店さがし」のテーマでは、「最近、秋葉原に行ったか」という第1話題が提示されて対話が始まっている。これを導入部とする。続いて、その第1話題について発話のやり取りがあり(継続部)、やがてその第1話題が終了して(終了部)、関連した別の第2話題に転換して、その話題について発話のやり取りされる、というように対話が続いていく。

本データの解析では、10分の対話のうち、一連の導入部、継続部、終了部から成る1つの話題を含む70秒の対話データを用いた。そして、対話が進むにつれて、継続部において被験者の気分が高まると仮定し、この感情の変化とともに被験者の音声や動作にどのような変化が現われるかを解析することにした。

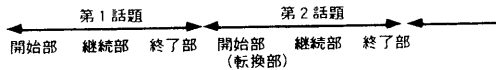


図2 対話の構造

解析にあたり、70秒のデータを10秒ずつのブロックに分け、各ブロックにおいて1) 被験者の発話時間(発話量)、2) 二人の発話に重複が生じた時間と二人とも発話しなかった沈黙の時間、3) 音声の平均ピッチ、4) 頭の縦振りの総時間、5) アイコンタクトの総時間を測定した。なお、被験者AとBの対話では、開始部はブロック1~2、継続部はブロック3~6、終了部はブロック6~7であった。また、被験者CとDの対話では、開始部はブロック1~2、継続部はブロック2~6、終了部はブロック6~7であった(図3、4参照)。

以下に、各要素についてマルチモーダル対話データベースをもとに解析した結果を述べる。

3.2.1 発話時間(発話量)

まず、感情の変化とともに、二人の発話時間の割合が変化するかどうかを調べた。図3(1)は各ブロックに占める被験者AとBの発話時間の割合を示したものである。

図3(1)に見られるように、始めの3ブロックは被験者Bが被験者Aよりも多く発話しているが、4ブロック目で逆転して、被験者Aの方がBよりも多く発話している。さらに対話が進んで5ブロック目からは両者の発話時間が近づいてきている。被験者Cと被験者Dとの対話(図4(1))では、ブロック5で両者ともに発話時間が増え、その後、両者の発話時間が等しくなってきた。これらのブロックは、それぞれ先に示した対話の継続部と終了部に当たることから、話しが進んで気分が高まってくるとともに対話のやり取りが活発になり、二人の発話時間がだんだん近づいてくると推測される。

3.2.2 発話の重複と沈黙

発話の重複は、二人の対話のやり取りが活発であるために生じると推測され、また、対話における沈黙は、やり取りがスムーズでなかったり、あ

るいは、次に発話すべき内容を考えているために生じると考えられる。図3(2)、図4(2)は各ブロックにおいて二人の発話を重ねた時間の割合と、二人とも発話しなかったために生じた沈黙の時間の割合を示したものである。

図3(2)では、ブロック4から6にかけて被験者AとBの発話の重複が増えている一方、二人とも発話していない沈黙の時間が減少している。図4(2)で示されている被験者CとDの対話においては、ブロック5において、発話の重複が増え、逆に沈黙の時間が減っている。ブロック6においては発話の重複が少ないが、沈黙の時間は依然少ない。これらのブロックは対話の継続部に含まれることから、気分が高まるとともに発話の重複が増え、沈黙の時間が減少することが確かめられた。

3.2.3 音声ピッチ

感情の変化とともに呼吸や発声の仕方が変わることが予測される。感情の変化とともに起こる音声の変化を見るために、ここでは基本周波数(以下 F_0 と呼ぶ)を測定した。

図3(3)、図4(3)は各ブロックにおける平均 F_0 を示している。まず、図3(3)を見ると、被験者Aではブロック4と5において、また被験者Bでは3~5において、 F_0 の上昇が見られる。図4(3)を見ると、被験者Cではブロック3~5で、また、被験者Dでは変動が少ないものの、ブロック3あたりから F_0 の上昇が見られる。これら F_0 の上昇が見られるブロックは、それぞれ対話の継続部に相当しており、気分の高まりとともに平均 F_0 が上昇すると推測される。さらに、終了部に向かうに従って、 F_0 が下降していることもわかる。

3.2.4 頭の縦振り

次に、頭の縦振りが、感情の変化とどのような関わりがあるのかに焦点をあてて解析した。

図3(4)、図4(4)は各ブロックで被験者が頭を縦に振る動作に占める時間の割合を示したものである。頭を縦に振る動作は個人差があり、頻繁に振る人もいれば、それ程でもない人もいる。しかし、程度の差こそあれ、対話過程において、頭の振りが多くなったり、少なくなったりする変化の様子を見て取ることができる。

被験者AとBではともに継続部のブロック4で頭の縦振りが減少し、継続部の終わりから終了部にかけての、ブロック5~6で増加している(図3(4))。一方、被験者CとDでは、継続部の始まりのブロック2と継続部のブロック5で頭の縦振りが増加している(図4(4))。したがって、頭を縦に振る動作は、気分の高まりとともに必ず

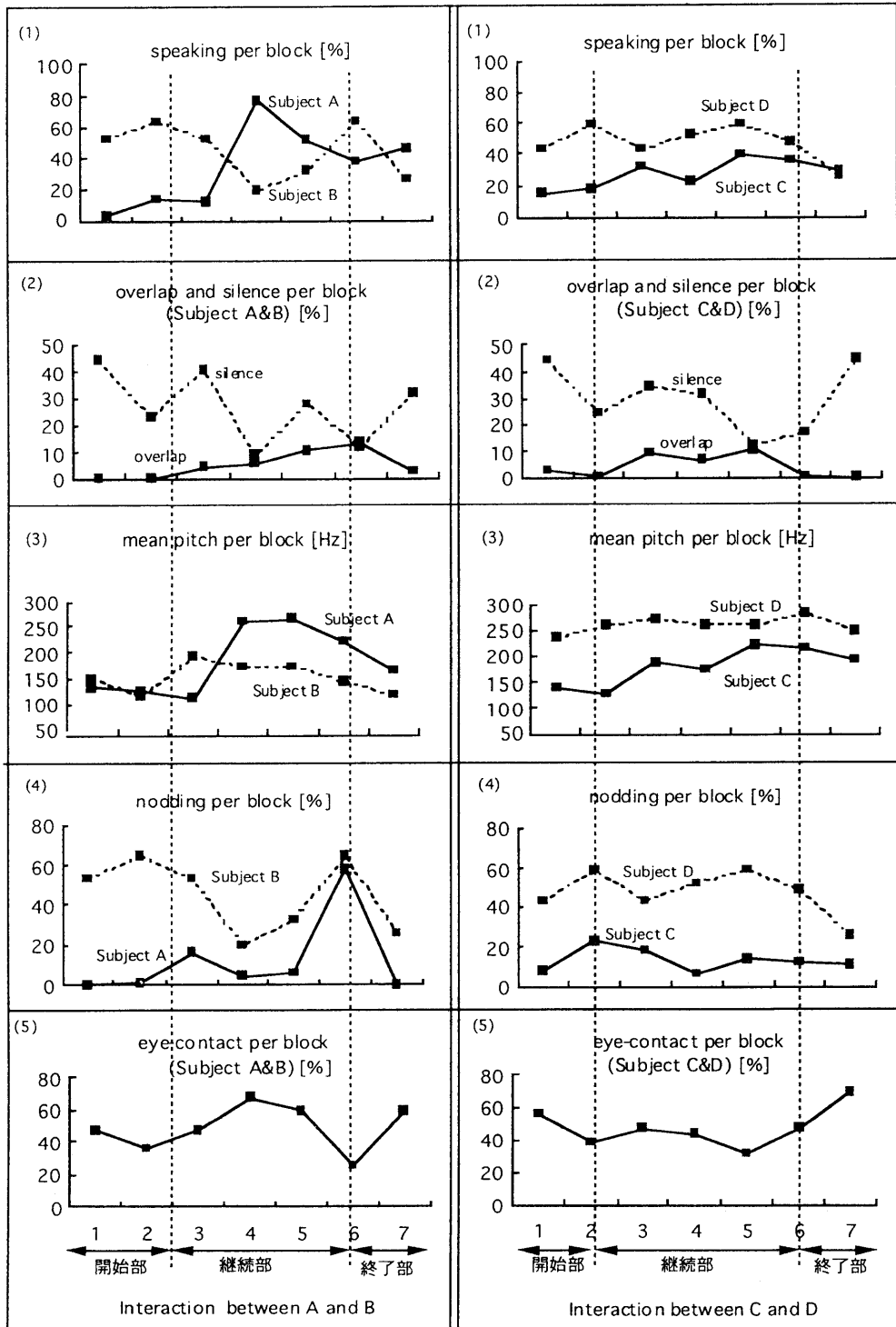


図3 (左図) 被験者AとBの対話； 図4 (右図) 被験者CとDの対話

しも増えるわけではないようである。

3.2.5 視線

一方、図3(5)、図4(5)は対話している二人が互いに相手を見る、アイコンタクトに占める時間の割合を示したものである。

これらの図を前述の頭の縦振りの図(図3(4)、図4(4))と比べて見ると、アイコンタクトが増えたと頭の縦振りが減るといように、互いに反比例して変化していることがわかる。たとえば、被験者AとBの対話(図3(4)、図3(5))では、継続部のブロック4で頭の縦振りが減少していることを先に述べたが、ここでのアイコンタクトの割合は高くなっている。被験者CとDとの対話においては(図4(4)、図4(5))、継続部のブロック5で頭の縦振りが増加していたが、アイコンタクトの割合は低くなっている。

視線を相手に向けるという行為は、相手の発話に興味を引かれたり、相手の関心を引くときに起こると考えられている[3]。しかし、継続部において必ずしもアイコンタクトの割合が高くなるわけではないことから、気分の高まりとの関係においては、頭の縦振りとアイコンタクトが相補的に働いているものと推察される。

3.2.6 感情の変化に見られる協調

上記では対話過程における感情の変化とともに、音声や動作がどのように変化するかを解析した結果を述べた。ところで、我々はこれまで、二人の対話において、発話単位の区切れやあいづち挿入時に両者が同時にうなずいたり、また、お互いの視線が合うときがあることを述べてきた[5][6][7]。そこで、二人の感情の変化の関係を調べてみた。

まず図3(3)のF0の変化を見てみると、一方のF0が高まると、他方のF0も呼応するように高まっており、被験者AとBの間に相関があることがわかる。同じように、図4(3)では、被験者CとDのF0の変化に相関が見られる。気分の高まりが二人の間で同じように起こっていることがわかる。

一方、頭の縦振りについても、図3(4)の頭の縦振りが占める時間の変化を見てみると、一方の頭の縦振り頻度が多くなると、同じように、他方も頭の縦振りが多くなっており、被験者AとBの間に相関があることがわかる。被験者CとDの頭の縦振りの変化(図4(4))にも相関がある。これらのことから、二人の感情の変化の度合いが同期し、協調しながら対話が進んでいくことが確かめられた。

4 刺激と感情の変化

第3章では、自然な対話の進行とともに音声や動作に変化が観察されるということを示した。本章では、感情の変化の様子をより客観的に解析するため、ある刺激を恣意的に被験者に与え、その前後で、どのような変化が現われるかを観察した。

4.1 データ収集

被験者がディスプレイに映し出されたCGの疑似人間と対話するというタスクでデータを収集した。被験者は2名(男女各1名)。被験者に与えた課題は以下のようなものである：

このCGは、あなたが「こんにちは」と呼びかけると、「こんにちは」と挨拶するはずなので、それまで「こんにちは」と呼びかけ続けて下さい。

CGは最初の40秒間はまったく反応せず、静止している。そして、40秒たった時にCGがまばたきをするようにコンピュータで制御してある。

4.2 解析結果と考察

本実験では、「こんにちは」と呼びかけても始めはまったくCGが反応しないことから、だんだんイライラしてくるが、CGのまばたきという直接的な刺激により、被験者が関心を抱くはずであるという仮説の下に、CGのまばたき前後の感情の変化を定量的に見ることを目的としている。そこで、本実験で得られた対話データについて、1) 刺激直前・直後に発話された「こんにちは」の音声の平均F0(Hz)、2) 刺激直前・直後の各10秒間に被験者が視線をはずした回数(回)、3) 刺激直前・直後各10秒間における被験者のまばたきの間隔(sec.)を測定した(表1参照)。

表1 刺激前後の変化量

マルチモーダル情報	被験者	刺激前	刺激後
F0 [Hz]	A (男性)	148-96 (121)	163-117 (129)
最大-最小 (平均)	B (女性)	186-140 (161)	195-166 (175)
視線の外れ [回]	A	4	0
	B	5	1
まばたきの間隔 [秒]	A	1.2	2.1
	B	1.7	2.7

4.2.1 音声ピッチ

表1の上段はCGがまばたきをした直前・直後に被験者が発話した「こんにちは」の最大・最小・平均F0を示している。図から、いずれも刺激後のほうが刺激前よりも上昇していることがわかる。

すなわち、関心を抱くとF0が高くなることが確かめられた。

4.2.2 視線の方向

表1の中段はCGがまばたきをした直前・直後各10秒間に被験者がCGから視線を外した回数である。表から、刺激前には頻繁に視線がはずれていたにもかかわらず、刺激後には、被験者Aは0回、被験者Bは1回と激減している。このことから、刺激を受けて、CGを凝視するようになったと考えられ、関心が高まるとともに、相手に視線を向けるアイコンタクトが出現することが確かめられた。

4.2.3 まばたきの間隔

次に、被験者のまばたきの間隔を測定した。表1の下段は、CGからの刺激直前・直後各10秒間における被験者のまばたきの間隔の平均を示したものである。刺激後の方が刺激前よりもまばたきの間隔が長くなっており、被験者がCGを凝視している様子がわかる。したがって、まばたきの間隔もまた、関心の高さを示すサインとなると考えられる。

5. まとめ

対話の進行とともに、人の感情は、気分が乗ってきたり退屈したりとさまざまに変化し、その感情が表情や音声に表われる。このような感情の変化をとらえることを目的に、本稿では、いわゆる話が乗ってくるという気分の高まりを、マルチモーダル対話データベースを基に解析した。その結果、気分の高まりとともに、以下の傾向が現われることを確認した：

- 1) 二人の発話量(時間)が等しくなってくる
- 2) 二人の発話の重複が多くなり、一方、沈黙の時間が減少してくる
- 3) 平均F0が上昇する
- 4) 頭の縦振りとアイコンタクトが相補って現われる。
- 5) まばたきの間隔が長くなる。

また、二人の感情の度合が同期し、協調しながら対話が進んでいくことが確かめられた。

今後は、心拍数など他のマルチモーダル情報を含め、感情の変化をさらに解析していく予定である。

謝辞

本研究を行うに際して、日頃ご指導いただく応用システム研究所中島隆之所長に感謝いたします。また、ご討議いただいた新機能シャープ研究室の皆様、および実験にご協力いただいた皆様に感謝いたします。

なお、本研究は、RWCP新機能シャープ研究室においてなされたことを付記いたします。

参考文献

- [1] R. A. Bolt, "The Integrated Multi-Modal Interface", The Transactions of the Institute of Electronics, Information and Communication Engineers (Japan), Vol. J-70-D, No. 11, 1987.
- [2] P. Ekman and W. V. Friesen, "The Repertoire of Nonverbal Behavior", *Semiotica* 1, pp. 49-98, 1969.
- [3] A. Kendon, "Some Functions of Gaze-Direction in Social Interaction", *Acta Psychologica* 26, pp. 22-63, 1967.
- [4] 坂本、綿貫、外川：マルチモーダル対話解析、人工知能学会研究会資料、Vol.1, SIG-SLUD-9401, pp.39-46, 1994.
- [5] K. Watanuki, K. Sakamoto and F. Togawa, "Analysis of Multimodal Interaction Data in Human Communication", *Proc. of ICSLP94*, pp.899-902, 1994.
- [6] 綿貫、坂本、外川：マルチモーダル対話データベースにもとづく対話解析、情報処理学会研究会報告、95-SLP-5, pp.17-22, 1995.
- [7] K. Watanuki and F. Togawa, "Multimodal Interaction in Human Communication", (to appear), *IEICE Transactions on Information and Systems, Special Section on Spoken Language Processing*, 1995.
- [8] K. Watanuki and F. Togawa, "Some Signals of Emotional Arousal: Analysis of Conversations Using a multimodal Interaction Database", (to appear), *Proc. of Eurospeech '95*, 1995.