

対話における雰囲気分析

伊藤秀樹 重野真也 西本卓也 荒木雅弘 新美康永
京都工芸繊維大学工学部電子情報工学科

現在の音声対話システムにおいては、対話者間でやり取りされる情報が主な歓心事である。しかし、対話そのものを活性化することも、種々のアプリケーションにおいては重要である。本研究では、対話が活性化している状態を対話が「盛り上がっている」状態と呼び、このような状態における対話者間の発話権の移動、身振り等がどの様になっているかを分析した。更に、対話エージェントに対話が活性化している雰囲気を表現させるために基礎研究として、幾人かの被験者に指定した心的状態を演じてもらった結果の分析結果について報告した。

The Analysis of the Atmosphere in the Dialogs

Hideki ITO, Masaya SHIGENO, Takuya NISHIMOTO, Masahiro ARAKI, Yasuhisa NIIMI
Dept. of Electronics and Information Science
Faculty of Engineering and Design, Kyoto Institute Of Technology

The information exchanged between the participants is concerned in the current spoken dialog systems. It is, however, important to activate a dialog itself in several applications such as training systems. From this point of view, physical behaviors of the participants in a dialog and turn-taking in the situations where a dialog was activated and where a dialog was not activated were observed and analyzed. Furthermore, to collect basic data necessary to make dialog agents express some atmosphere of an activated dialog, we asked some subjects to perform directed psychological situations and analyzed their behaviors. This paper reported the results on these studies.

1 はじめに

人同士のコミュニケーションの場には種々の「雰囲気」が存在する。雰囲気を醸成する要因としては、コミュニケーションの場に参加している人の社会的関係、その人数とそれぞれの性格、そこで行われる会話の内容、その場の背景(場所、聴覚/視覚的背景)等が挙げられる。コミュニケーションの場で醸成される雰囲気は、そこに参加する人々によって感じ取られるだけでなく、この様な場を第三者的に観察することによってもある程度感知できる。その手掛りとなる現象としては、言語的要因(話題、文体、語彙など)と非言語的な要因(発話権の移動、発話の重なり、声の特徴、表情、姿勢、身振りなど [1][2][3][4])

が挙げられる。一方、昨今の情報技術の進歩により、人とコンピュータが対話できる音声対話システムも実現しつつある。また、コンピュータ側に擬人化エージェントを置いて対話の自然性を向上させようとする試みも多い。更にエージェントにジェスチャーを付いたりパーソナリティや感情を持たせるための基礎的な研究も開始されている [5][6]。現在この様なシステムでやり取りされる対話は、単純な目的指向型のものが多い。しかしその目的によっては、対話の場に種々の雰囲気を醸成して対話を活性化することにより、その目的の達成が容易になる場合(例えば、セールスや教育訓練を目的としたシス

テム等)もある。またコンピュータと「おしゃべり」を楽しむためのシステム等では対話の雰囲気は重要な要因となるであろう。

このような観点から、我々是对話における雰囲気分析を非言語的な側面から行っている。本稿では、人同士の対話において、対話が活性化している状態(これを俗な言葉で「盛り上がっている」状態と呼ぶ)とそうでない状態における対話者間の動作、発話権の移譲の状態などの差を調べた。また将来エージェントにその心的状態を仕草として表現させるための基礎として、幾人かの被験者に指定した心的状態を演じてもらった結果の初歩的な分析結果についても報告する。

2 対話における盛り上がった状態

2.1 盛り上がった状態とは

1. で述べたように本稿では人同士の対話において、対話が活性化し話が弾んでいる状態を「盛り上がった状態」と呼ぶ。具体的には次のような状態を扱う。

1. 話者各々が現在の会話を楽しんでいる
2. 話者各々が対話に熱中している

一般的に人が「盛り上がり」という言葉を聞くと、そこから連想される言葉は「楽しそうな」という形容詞になる。よって、一般的な盛り上がり状態としては、1 だけでも良いかもしれない。しかし、「相互が現在の対話を建設的な状態と認識している雰囲気」として盛り上がりを捉え、2 も考えられる。例えば、会議等において、建設的な議論が活発に行われている時、それを盛り上がっていると表現する。よって、これら二つを盛り上がり状態とする。

2.2 分析データ作成

分析を行う為のデータ作成を以下の順序で行った。

1. 対話収録

二人の対話の状態を収録するために、図 1 のような形式で音声と映像を収録した。

外部からの情報を遮断するために、収録は防音室内で行った。防音室内では約 15 分間、特にテーマも定めず自由な対話を行ってもらい、そ

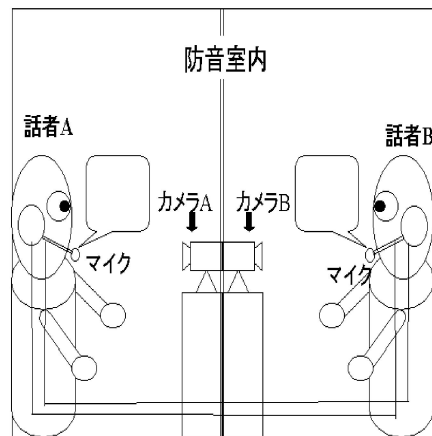


図 1: 対話動画収録方法

の状況を収録した。

2. 動作の書き出し

以上のように収録した各対話から、「盛り上がっている状態」とそうでない状態を各々 2 場面(1 分/場面) ずつを視察により抽出した。対話中の時間の流れに対する動作の流れを表現するために、収録された動画から動作を書き出しを行った。以下の動作を対話の中から視察により同定し、その動作の開始時間と終了時間を書き出した。

- 発話
- うなづく
- 笑う
- 頭の移動(前後左右上下)
- 首を振る(右、左、左右)
- 視線の移動(上下左右、見回す)
- 上体(肩)の移動(前後左右)
- 両手(腕)の移動(前後左右上下)
- 特殊な動作(サイン、指示、触れる等...)

また、音声に関しては、その短時間エネルギーをもとに発話区間を抽出した。実際の書き出し例を図 2 に示す。

2.3 データ分析

分析は以下の点に注目し、盛り上がっている状態とそうでない状態の比較を行った。

- 発話、うなづき、笑う、の重なり時間
- 発話動作の間の状況

<03:00:23-03:02:21>右手を右手前に出す
 <03:05:05-03:10:12>うなづく*3
 <03:10:27-03:13:16>笑う
 <03:14:19-03:19:16>左手で右袖を捲り上げる
 <03:19:16-03:29:16>頭を後ろに反らせる
 <03:29:28-03:35:21>発話
 <03:38:04-03:54:22>右上を見る
 <03:42:05-03:43:03>うなづき*1

図 2: 書き出しデータの一例

- 発話動作が重なり合っている区間の他の動作

分析の対象としたのは 20 対話で、盛り上がっている状態とそうではない状態それぞれ 2 × 20 (対話) = 40(場面) である。

2.3.1 発話、うなづき、笑う、の重なり時間

まず、対話中でお互いが同じ動作を同時に行っている部分を、ここでは「動作の重なり」と呼ぶ。対話の中での動作の重なりの出現頻度が高いのは発話、うなづき、笑うの順であることから、その 3 動作のそれぞれの重なり時間がその対話中に占める総時間を比較する。各対話毎に盛り上がっている状態、そうではない状態、それぞれ 2 場面の重なり時間の平均を取り比較を行った。その結果、以下の表 1 のような結果を得た。この表のデータ数は、比較を行った際に盛り上がっている状態の方が重なり時間が長い対話の数を表す。

表 1: 3 動作の重なり時間の比較

動作名	データ数 (全 20)
発話	19
笑い	12
うなづき	15

2.3.2 発話動作の間

二人の話者の一つの動作に注目すると、「動作の間」は図 3 のように 4 つのタイプに分類できる。Type1 は話者 A がある動作を行った後、間があり再び話者 A が同じ動作をするという事を表している。Type2 は話者 A がある動作を行った後、間があ

表 2: 各 Type 毎の出現頻度の比較

Type	データ数
Type1	7
Type2	8
Type3	9
Type4	19

り話者 B が同じ動作をするという事を表している。Type3 は話者 A のある動作中に話者 B が同じ動作をするという事を表している。Type4 は話者 A がある動作を行い、その終了付近で話者 B の同じ動作が重なるという事を表している。

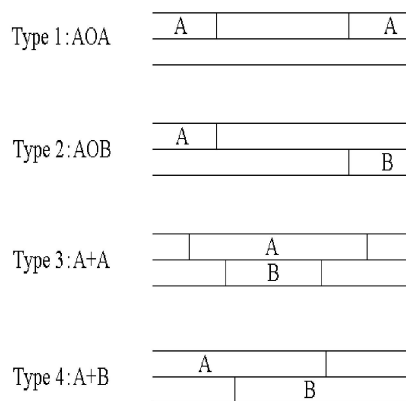


図 3: ある一つの動作に関する「間」の種類

ここでは、発話動作に注目し、各 Type 毎の対話中の総出現頻度 (回数) の比較を行う。各対話毎に盛り上がっている状態、そうではない状態、それぞれ 2 場面の総出現頻度の平均を取り比較を行う。その結果、以下の表 2 のような結果を得た。この表のデータ数は、比較を行った際に、盛り上がっている状態の出現頻度が多い対話の数を表す。

2.3.3 発話動作が重なっている部分の他の動作

対話が盛り上がっている場合は発話動作において Type4 がよく観察される。この時、同時に起こっているほかの動作の種類を調べた。Type4 部分の中にある他の動作の出現頻度 (回数) を数え、全 80 場面の合計を算出し、その頻度の大きなものから順位をつける。各状態において少なくともどちらか一人がその動作を行ってればその動作が出現したとみな

した。発生頻度の多い動作を表 3、4 に記す。

表 3: 発話の Type4 発生時の他の動作の出現回数（盛り上がり状態）

順位	動作名	データ数 (回)
1	うなづく	116
2	笑う	112
3	下を見る	34
4	左下を見る	32
5	上体を後ろに反らせる	26
6	右下を見る	26
7	右上を見る	24
8	右を見る	24

2.4 結果及び考察

動作の重なり時間の比較では、発話、うなづき、笑いの 3 動作の中では発話の重なり時間が最も盛り上がり状態と関係が深い。次に発話に注目した間の比較を行ったが、Type4 の頻度が盛り上がっている状態の時の方がより高い。つまり、対話が活性化する程、お互いが話している時に生じる重なり状態（かぶり状態）が発生する頻度が高くなる。最後に Type4 が生じた時の他の動作の出現回数を比較したが、盛り上がっている状態の方では視線の動きが活発になった。また、盛り上がっている状態では「掻く」という動作がほとんど見られないが、そうではない状態では比較的好く見られた。他にも、そうではない状態は頭の移動が比較的好く見られた。

表 4: 発話の Type4 発生時の他の動作の出現回数（そうではない状態）

順位	動作名	データ数 (回)
1	うなづく	134
2	笑う	90
3	下を見る	46
4	右上を見る	28
5	頭を後ろにそらせる	28
6	頭を右に傾ける	24
7	上体を前に出す	18
8	右手で鼻を掻く	16

3 動作による雰囲気表現

エージェントが種々の動作によってその心的状態（ここではこれを雰囲気と呼ぶ）を表現することにより、人との対話を活性化することが期待される。そのためには、人間がどのような動作によって雰囲気を表現あるいは感知しているかを調べる必要がある。以下では、被験者に指示した雰囲気を演じてもらい、その動作を分析することによって人間が雰囲気を表現する際におこなう動作について調べた。

3.1 方法

分析対象とする雰囲気は、日常生活の中で個人が比較的良く遭遇し、かつ、ある程度持続する雰囲気という観点から次の 6 つを分析の対象とした。

- 上機嫌
- 不機嫌
- 納得
- 納得いかない
- 興味がある
- 興味がない

これらの雰囲気の収録を次の方法により行った。

被験者が一人で防音室に入って椅子に座り、向かいに對話の相手がいると想像しながら、指示された雰囲気を 20 秒間演じる。その様子をデジタルビデオカメラおよびモーションキャプチャセンサ「DigitEye3D[7]*1」を用いて記録し、これを 1 サンプルとする。以下の説明において、前者をビデオ画像、後者をセンサデータと呼ぶ。なお、モーションキャプチャ用の光源ポインタは文献 [8] 等を参考にして、被験者の頭部に 2 つ、利き腕の肩と手首にそれぞれ 1 つずつ取り付けられた。被験者の数は 5 人（全員右利き）であり、5 人 × 6 雰囲気 = 30 サンプルの動画を用意した。収録の様子を図 4 に示す。

3.2 結果

以上の方法で収録したビデオ画像から、被験者の動作について頭部、上体、右手の 3 つの部位に注目して、各部位で生じる動作方向を書き出した。動作書き出しリストの一例を図 5 に示す。図 5 において、左から順に動作開始時刻、動作終了時刻、動作部位、動作方向を表している。

*1 (財) イメージ情報科学研究所
<http://www.image-lab.or.jp/>

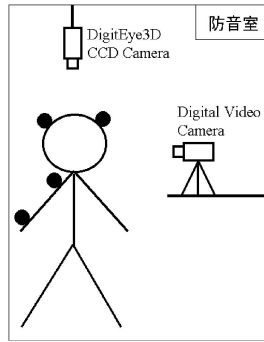


図 4: 雰囲気収録方法

被験者 A : 上機嫌

[00:00:01.15-00:00:02.11] 頭部 : 左

[00:00:02.27-00:00:04.06] 頭部 : 左

[00:00:04.11-00:00:07.25] 頭部 : 左

[00:00:07.25-00:00:10.19] 頭部 : 前後

[00:00:02.18-00:00:03.06] 上体 : 下

[00:00:07.04-00:00:09.27] 上体 : 前後

[00:00:06.05-00:00:07.11] 右手 : 顔を触る

被験者 B : 上機嫌

...

図 5: 動作書き出しリスト

動作書き出しリストをもとに、雰囲気ごとに頭部・上体・右手に対応している動作の出現頻度（回数）を調べた。収録した 30 ビデオ画像について集計した結果を表 5 に示す。また、頭部の動作開始方向で分類した各動作の出現頻度詳細を表 6 に示す。表中の空白部は、その動作が表れなかったことを示している。また、同じ動作方向において、他の雰囲気より出現頻度の高い数字は太字で表している。

この結果、他の部位に比べ頭部の動作が頻繁に発生しており、頭部の動きが雰囲気表現において重要な役割を担っていると考えられる。ただし、今回は椅子に座った状態で被験者に演技をしてもらったので下半身の制約があった。そのため、頭部の動作が頻繁に生じた可能性もある。また、表 6 より、不機嫌を除く各雰囲気では、それぞれ頻繁に発生する動作が観測された。これらの対応関係は、動作による雰囲気表現をエージェントに行わせる際のひとつの指標になると考えられる。不機嫌に関しては頭部の左右の動作頻度がやや高いものの、他の雰囲気と比

表 5: 動作出現頻度 : 単位 [回]

	頭部	上体	右手
上機嫌	22	9	8
不機嫌	18	7	6
納得	26	5	2
納得いかない	22	3	3
興味がある	22	6	1
興味がない	1	0	0
計	111	30	20

表 7: 頭部動作平均振幅 : 単位 [mm]

	頭部		
	左右	前後	うなずく
上機嫌	107.2	145	
不機嫌	196.7	205.6	142
納得	137	81	34.5
納得いかない	24	134.8	10
興味がある		131	29.7
興味がない	91.5	92	39.3

較して特徴的な動作は観測できなかった。これは、「眉間にしわを寄せる」「眉をひそめる」など表情によって不機嫌を表現している被験者が多数いたことから、不機嫌な雰囲気構成要素の大半が表情であると考えられる。

次に、センサデータを用いて、頭部の「左右・前後・うなずく」動作について、各動作の揺れ幅を調べ平均振幅を求めた。結果を表 7 に示す。表中の空白部は、その動作が表れなかったことを示している。

この結果、「不機嫌」では全ての動作において他の雰囲気よりゆれが著しく大きい値となり、全方向への振幅が大きいことがわかる。「納得」では前後の振幅が小さく、「納得いかない」では、左右とうなずきの振幅が小さい値となっている。しかし、他の雰囲気ではこれといった特徴は表れなかった。

両者をあわせると「不機嫌」では全方向への振幅が大きく、それ以外の雰囲気の場合は表 6 より、各雰囲気に対応する頭部の特徴的な動きを抽出することができた。しかし、「上機嫌」と「興味がある」に

表 6: 動作出現頻度詳細：単位 [回]

	頭部					
	上下	左右	前後	うなづく	傾ける	首を回す
上機嫌	3	6	12		1	
不機嫌		8	5	1	1	1
納得		3	2	19	2	
納得いかない		1	5	2	14	
興味がある			11	7	4	
興味がない	2	11	2	4	2	

関しては、同じ動作特徴が表れただけで、特徴を分離するまでには至らなかった。

4 おわりに

本研究ではエージェントに雰囲気を加えるために、まず対話が活性化していると、すなわち、盛り上がっていると、どのような現象が起こっているかを人同士の対話中の動作及び発話から分析を行った。発話が盛り上がっている時は、相互に発話がかぶっている部分の頻度が高かった。また、相互に発話がかぶっている時の他の動作の出現傾向を調べ、盛り上がっている部分とそうでない部分の比較を行った。今後はデータ数の増加、他の着眼点からのデータ分析を行っていく必要がある。

次に動作による雰囲気表現の分析するために、ビデオおよびモーションキャプチャセンサによる動作の記録を行い、得られたデータを分析することによって雰囲気表現の動作特徴を調べた。その結果、頭部の動作出現頻度が高く、頭部の動作に注目することによって各雰囲気に対応する特徴的な動作の分類を行った。今後は今回の分類を元に各雰囲気の特徴的な動作をデータベース化し、適切に再合成することによって動作による雰囲気表現が可能なエージェントを作成し評価していく予定である。

参考文献

[1] M.F. ヴァーガス (石丸正訳) : 非言語コミュニケーション, 新潮社 (1987)
 [2] 黒川隆夫: ノンバーバルインターフェイス, オーム社 (1994)

[3] 渋谷昌三: しぐさ・ふるまいでわかる相手の心理, 日本実業出版社 (1999)
 [4] 大坊郁夫: しぐさのコミュニケーション, サイエンス社 (1998)
 [5] Proc. of Workshop on Representing, Annotating, and Evaluating Non-Verbal and Verbal Communicative Acts to Achieve Contextual Embodied Agents, Montreal, Canada, (May, 2001).
 [6] Proc. of 2nd Workshop on Attitude, Personality and Emotions in User-Adapted Interaction, Sonthofen, Germany, (July, 2001).
 [7] 金森務, 片寄晴弘, 井口征士: 「簡易モーションキャプチャセンサ DigitEye3D 原理と基本特性」電子情報通信学会論文誌 No.10 (Oct. 1997)
 [8] 綿貫啓子, 関進, 三吉秀夫: 「発話交代時に現れる頭部の動きの特徴」電子情報通信学会ヒューマン情報処理研究会 HIP99-74.