

自然対話におけるジェスチャーの相互的関係の分析

前田 真季子 堀内 靖雄 市川 薫
千葉大学

概要

人は視線の動きやうなずきなどのジェスチャーを用いて、対話の円滑なやり取りを行なっている。自然対話は話者同士の音声情報、視覚情報を用いた相互作用によって進行していくものであるため、音声におけるあいづち現象などと同様に、ジェスチャー同士にも話者間に相互作用が生じていることが推測される。そこで、本論文では、特にうなずきに注目し、ジェスチャーによる相互作用を分析した。分析に用いたデータは、6組の親しい友人同士による対話であり、収録には正面映像を撮ることが可能な、2つのプロンプターを使用した。そして、その収録データを一般に公開されているアノテーションツール“ANVIL”を用いて、アノテートし、分析を行なった。分析の結果、うなずきは、あいづちと同様に相手話者の発話に対する何らかの応答動作として生じる場合よりも、自己発話内の方が多く生じる傾向が見られた。また、うなずきが二人の話者で同時に発生する現象が多いことも示唆された。

Analysis of Correlation between Interlocutors' Gestures in Spontaneous Speech

Makiko Maeda Yasuo Horiuchi Akira Ichikawa
Chiba University

Abstract

People use gestures like gaze and nod for smooth communication in dialogue. Usual dialogue continues exchanging interlocutor's information with each other using speech and gestures and therefore it is supposed that there is correlation between interlocutors' gestures as backchannels in speech. In this paper, we focused nods for the analyses of gestures. 18 dialogues by six pairs of good friends were recorded, where they can look at each other via two prompters. The prompter can record the interlocutors' gesture on videotape and project the partner's image through a half mirror. We annotated recorded dialogue by using the annotation tool "ANVIL" developed by Michael Kipp and the transcription tool developed by ours. As a result, it was suggested that gestures are caused more frequently when an interlocutor is speaking than listening, and interlocutors tend to nod simultaneously with considerable frequency.

1. はじめに

近年急速に情報化が進み、コンピュータはわれわれにとって、より身近な存在となってきた。例えば、人に代わって、道案内、受付の対応をするロボットや、掃除をするロボットが開発されているし、パソコン上でも擬人化エージェントが対話形式でユーザの要望に受け答えするといった開発がされている[1]。そして、現在でも多くの研究者がロボットと人とのコミュニケーションの研究[2]や、擬人化エージェントの開発を行なっている[3][4][5][6]。そこで我々は、そのようなロボットや、擬人化エージェントが、ユーザの要望に対して、必要最低限の受け答えを返すのではなく、より親しみやすい存在として実現できないだろうか考えた。現在、開発されているロボットの中で、親しみやすいものの例として、視覚、聴覚、臭覚、皮膚感覚をもとに、様々な表情を見せるロボットがある[7]。しかし、このロボットは表情を様々に変化させることができるが、人とのやり取りは行なえない。また、うなずきを生じるロボットの研究[2]もあるが、このロボットは音声情報のパワーをもとにうなずきを返しており、相手の画像情報を利用していない。しかし、人は対話をする際、視覚情報をもとに、ジェスチャーを生じているのではないかと考える。さらに、擬人化エージェントにおいては、音声情報の相互的やり取りは考慮しているが、そこにジェスチャー情報の認識が含まれていないもの[3]や、ジェスチャー認識をおこなっていても、対象の動きの大きさによって、あいづちを返す動作にとどまっているものがある[1]。また、これらのエージェントはどれも、受付対応システムであり、ユーザの要求に答えるという、ある程度相手とのやり取りが明確であるといえる。一方、自然対話では、音声情報のみならず、お互いのジェスチャー情報の相互的作用がコミュニケーション手段としてより重要であると考えられる。よって、ロボット側、エージェント側も、人と同様に、相手に合わせた動きを返すことが、より円滑なコミュニケーションを行なうために重要なのではないだろうか。そこで、本研究では、対話によって生じるジェスチャー情報に、明確な相互的關係が存在する

かを分析することを考えた。今回焦点を当てたジェスチャーは視線とうなずきの2つである。これまでの研究でもいくつか、ジェスチャーに関する研究は行なわれているが、それらの研究は自然対話ではなく事務応対のデータを用いた分析であり、さらにジェスチャー同士の相互關係を分析していない[8][9]。一方、視線の分析については[10][11]といった報告がある。そのうち、[10]の研究報告では、視線は話し手の時の方が、聞き手の時よりも視線をそらすことが多いとしている。そこで、今回の収録データとこの研究結果との比較をおこなう。また、[12]では、道案内での身ぶりやアイコンタクトなどの身体的同調が空間情報の伝達を円滑に行なう役割を果たすといえるという研究結果を報告している。そこで我々は、自然対話においても同様な現象が、うなずきや、視線において生じるのではないかと考えた。よって、それらの同調性について見当する。

本論文では、まず、対話収録についての説明を行ない、続けてそれらのデータを分析するために用いたアノテーションツールの説明をする。そして、それらのデータを基に分析した結果について報告する。

2. 対話収録

本研究では、自然な対話におけるジェスチャーを分析するため、自然対話の動画像による収録を検討した。以下に行なった収録内容について概説する。

2.1 収録設備

ジェスチャーを分析するためには、正面から画像を収録することが望ましいと言えるが、正面にカメラをおくことは、対面での会話の際には不可能である。そこで、本研究では、プロンプターを用いることにより、正面映像を収録することとした。両者は互いに、仕切りによって区切られた防音室の中に入り対話を行なう。そこでは、もう一方の部屋にいる相手の映像がプロンプターのハーフミラー上に投影され、同時に自分の映像はハーフミラー後方のカメラにより収録され、相手のプロンプターに送られる。図1にプロンプターの仕組みを示す。

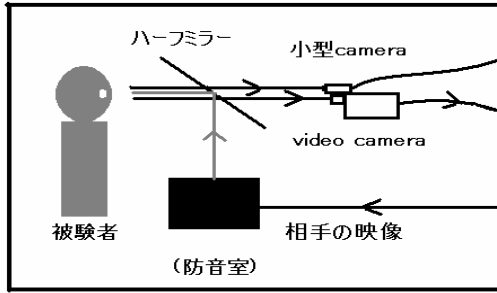


図1 プロンプター

2. 2 音声情報の収録

今回の実験では、お互いのいる部屋を別々に仕切り、相手の音声マイクに回りこむのを防ぐことで、二人の被験者の音声を完全に分離して録音した。音声のやり取りはイヤホンとピンマイクを用いた。ヘッドセットの使用も検討されたが、ヘッドセットでは頭にある程度の重さのものをかぶることになるので、自然対話の妨げになり、自然な顔の表情が出にくくなる恐れがあると考えた。イヤホンは片耳だけの装着とし、相手の声だけが聞こえてくるようにした。

2. 3 画像情報の収録

プロンプターの内部には、2つのビデオカメラを設置した。1つは上半身の映像を収録するカメラで、その映像はリアルタイムで相手のプロンプターに映し出される。さらにもう1つは、アイコンタクトやうなずきといった顔の詳細の情報を記録するための、顔のアップ画像を収録する小型カメラである。画像情報は全部で4つあるが、画面を4分割しそれぞれの映像を一つの画像として録画した。これにより、各動きを同時に観察する事が可能となった(図2)。



図2 収録風景

画像、音声はともにデジタルビデオデッキで記録された。画像は DV 形式、音声は量子化ビット数 16bit(線形)ステレオ、サンプリング周波数 48kHz である。

また、顔の部分に影が生じると、分析対象である、視線の動きについて明確な情報を得ることができない。そこで、プロンプターの左右に照明を配置し、顔に影が生じることを防いだ。

2. 4 実験参加者

今回参加してもらった被験者は、男性同士のペア 3組、女性同士のペア 3組の計 6組であり、各組 3対話合計 24対話の収録を行なった。また、収録という環境においても、ある程度自然な対話を行なうことを期待し、話者同士は友人関係であることを条件とした。

2. 5 収録実験の内容

特定のテーマによるディベート形式での対話を収録した。あらかじめ、意見が分かれそうな二者択一の質問が書かれたアンケート用紙(図3参照)を作成し、収録前に、被験者二人にそのアンケートに答えてもらった。

氏名	
以下のアンケートに答えて下さい。	
Q1. 男と女どっちが得だと思いますか?	(男・女)
Q2. 付き合うなら相手の顔と性格どちらを重視しますか?	(顔・性格)
Q3. 結婚はお見合いと恋愛どちらがうまくいくと思いますか?	(お見合い・恋愛)
Q4. 結婚に一番必要なのはお金と愛どちらだと思いますか?	(お金・愛)
Q5. 男女間に友情は成り立つ(親友はできる)と思いますか?	(思う・思わない)
Q6. 恋人と親友を計りにかけたらどちらを選びますか?	(恋人・親友)
Q7. 学生の自分 or 彼女に子供ができればおろす(妊娠中絶は認められる)と思いますか?	(思う・思わない)
Q8. 飼うとしたら犬と猫どちらがいいですか?	(犬・猫)
Q9. 日本でも自己防衛の手段として銃の所持を認めるべきだと思いますか?	(思う・思わない)
Q10. 死刑制度はあるべきだと思いますか?	(思う・思わない)
Q11. 機械が知能を持つこと(人間に近いロボットができること)に賛成ですか?	(賛成・反対)
Q12. 阪神は今年優勝すると思いますか?	(思う・思わない)
Q13. 幽霊はいると思いますか?	(思う・思わない)
Q14. 神様存在を信じますか?	(信じる・信じない)
Q15. 超能力や催眠術を信じますか?	(信じる・信じない)
ご協力ありがとうございました。	

図3 アンケート用紙

この時、互いに相談することは禁止し、きちんと理由を考えて答えを選んでもらうように指示した。

そして、二つのアンケートを比較し、意見が分かれたものに対して二人で討論してもらうようにした。今回ディベート形式を試みたのは、テーマを与えることによって、会話が途切れず、自分の考えを相手に訴えかける姿勢や、相手の意見を聞く姿勢が現れ、視線の動きやうなずきの動作が数多く表出されることが期待できるという意図からである。時間は1テーマ5分程度とし、5分経過後に実験者が対話を打ち切った。

3. アノテーション方法

収録したデータのうち、各組の2対話目を分析した。また、分析のためのアノテーションを Michael Kipp 氏が開発しているアノテーションツール ANVIL を用いて行なった[13]。このツールは自分の目的にあった情報をアノテートでき、それらの時間的相互関係が明確に表記できる。

本研究でのアノテーション内容を説明する。

表記した情報

- 音声情報・・・words
 - 発話単位を無音区間 200ms 以上に設定
- 視線情報・・・eye
 - 相手を見ている場合を look で表記
- 頭部情報・・・head
 - nod・・・うなずき
 - tilt・・・首を傾げる
 - shake・・・上下、左右、その他の動き

音声情報において、以前の話者交替を分析した研究[14]では、発話単位を 400ms に設定したのに対し、200ms と短くしたのは、発話単位を短く設定するほうが、話している時間と話していない時間のより詳細な区別がつくためである。また、200ms 以下にすると、促音により発話が分割されてしまうため、200ms を妥当であると考えた。さらに、ANVIL では音声情報を正確にアノテートすることが困難であるため、音声情報は千葉大学で開発した書き起こし支援ツールを用いて表記した。一方、ジェスチャー情報の表記は、ジェスチャーの起こり始めから、元に戻るまでを一連の動きとして表記した。また、うなずきは上下の動きを一連の動作で行なったものを定義し、上を向いたまましばらく動かなかったものなどは shake の動きとして区別した。また、動きの表記基準として、視線においては、相手を見していない状態から、次のフレームで相手を見たフレームを始点とし、相手を見ている状態から、次のフレームで目がそれた場合の前フレームを終点とした。また、頭部動作においては、何も動きがない状態から次のフレームで動きの変化が起こる時に、そのフレームを始点とし、動きが連続して起こった後、次のフレームで動きに変化がない場合の前フレームを終点とした。図4に ANVIL の例を示す。

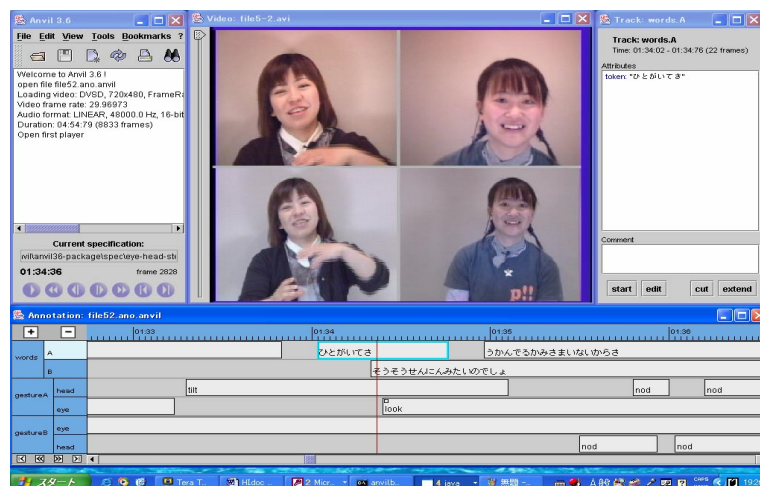


図4 ANVIL の例

表1 話し手、聞き手、全体における視線の状態の合計時間と比率(単位 sと%)

話者	話し手				聞き手				合計			
	見ている		そらしている		見ている		そらしている		見ている		そらしている	
1	67.201	66	34.301	34	106.240	92	9.476	8	173.441	80	43.777	20
2	121.489	88	16.717	12	81.249	98	1.435	2	202.738	93	18.152	7
3	95.697	78	26.360	22	88.923	94	5.672	6	184.620	85	32.032	15
4	100.401	87	14.415	13	97.865	93	7.508	7	198.266	90	21.923	10
5	49.150	42	66.734	58	93.928	90	11.011	10	143.078	58	77.745	42
6	74.441	73	28.162	27	116.284	96	5.205	4	190.725	77	33.367	23
7	109.744	79	29.296	21	92.727	84	17.017	16	202.471	73	46.313	27
8	127.962	83	25.893	17	88.756	92	7.774	8	216.718	77	33.667	23
9	74.642	78	21.588	22	148.683	99	1.568	1	223.325	91	23.156	9
10	114.416	68	53.888	32	72.139	94	4.771	6	186.555	73	58.659	27
11	83.417	48	88.790	52	62.330	71	25.659	29	145.747	54	114.449	46
12	86.954	68	41.008	32	117.552	90	13.547	10	204.506	79	54.555	21
全体	1,105.514	71	447.152	29	1,166.676	91	110.643	9	2,272.190	80	557.795	20

4. 分析

4.1 視線の分析

[10]の分析手法と同様に、話し手、聞き手、全体での相手を見ている状態と、相手を見ていない状態の各合計時間(左)と比率(右)を表1に示す。なお、話者1~6は男性、7~12は女性の被験者であり、話者1と2、3と4、5と6、7と8、9と10、11と12がそれぞれ対話したペアである。

4.2 頭部の分析

表2 うなずき、あいづちの発生頻度(単位 回)

話者	あいづち以外	あいづち	うなずき
1	22	5	27
2	20	10	30
3	40	20	60
4	4	9	13
5	26	13	39
6	43	11	54
7	55	38	93
8	65	33	98
9	36	52	88
10	52	22	74
11	49	17	66
12	51	41	92
合計	463	271	734

表2は各話者のうなずきの発生頻度を示し、あいづちのうなずきとあいづち以外のうなずきを別々に表

記した。なお、あいづちのうなずきとは、「うん」や「ああ」などのあいづち発話に付随したうなずきと、言葉を発せずに相手の発話中に起こしたうなずきのことであり、あいづち以外のうなずきとは自分の発話中に生じたうなずきのことである。

次に、各うなずきの単位時間(1分)あたりのうなずき回数を表3に示す。

表3 単位時間(1分)あたりのうなずき回数

(単位 回)

話者	あいづち以外	あいづち	うなずき
1	15.53	3.23	18.76
2	12.90	7.06	19.96
3	32.43	13.79	46.23
4	2.76	7.30	10.06
5	28.36	14.72	43.08
6	48.68	12.00	60.68
7	42.31	20.00	62.31
8	34.21	25.38	59.60
9	38.57	31.20	69.77
10	31.20	23.57	54.77
11	32.31	14.57	46.88
12	43.71	27.03	70.75
合計	29.06	17.01	46.07

次に、各話者のうなずき前後150ms以内において、相手話者がうなずきを生じたものの頻度と比率を表4に示す。なお、前後150ms以内と幅

を持たせたのは、もしも、話者が動作を生じる場合、相手話者の動作後にいくらか遅れて生じる可能性がある」と推測したからである。

表4 ある話者のうなずき前後150msにおける相手話者のうなずきの頻度と比率(単位 回と%)

話者	うなずき	
	頻度	比率
1	8	30
2	5	17
3	4	7
4	5	38
5	11	28
6	11	20
7	54	58
8	57	58
9	39	44
10	42	57
11	40	61
12	40	43
合計	316	43

続いて、ある話者のうなずきにおける相手話者のうなずきが発生するタイミングごとの発生頻度と比率を表5に示す。

表5 発生タイミング別うなずきの発生回数と比率

表5-1 全体(単位 回と%)

	あいづち以外		あいづち		全体	
	頻度	比率	頻度	比率	頻度	比率
前(150ms)	25	7	7	2	32	8
前	94	25	21	6	115	30
全体	28	7	14	4	42	11
後	56	15	62	16	118	31
後(150ms)	13	3	15	4	28	7
内部	33	9	12	3	45	12
合計	249	66	131	35	380	100

表5-2 あいづち以外(単位 回と%)

	あいづち以外		あいづち		全体	
	頻度	比率	頻度	比率	頻度	比率
前(150ms)	10	4	5	2	15	6
前	37	15	18	7	55	22
全体	20	8	12	5	32	13
後	37	15	59	24	96	39
後(150ms)	8	3	13	5	21	8
内部	20	8	10	4	30	12
合計	132	53	117	47	249	100

表5-3 あいづち(単位 回と%)

	あいづち以外		あいづち		全体	
	頻度	比率	頻度	比率	頻度	比率
前(150ms)	13	10	2	2	15	11
前	59	45	3	2	62	47
全体	10	8	2	2	12	10
後	18	14	3	2	21	16
後(150ms)	5	4	2	2	7	5
内部	12	19	2	2	14	11
合計	117	89	14	11	131	100

表5-1は当該話者のあいづちとあいづち以外の両方を含んだうなずきに対する相手話者のうなずきの出現頻度と比率であり、表5-2はあいづち以外のうなずきに対する頻度と比率を、表5-3はあいづちのうなずきに対する頻度と比率を示している。表中の分類の意味を以下に説明する。

「前(150ms)」

…うなずきの開始よりも以前に相手の頭部動作が発生しており、うなずきの開始時刻よりも前(150ms以内)でその動作が終了していることを意味する。

「前」

…うなずきの開始時刻より前から相手の頭部動作がおこり、うなずきの開始時刻と終了時刻の間でその動作が終了したことを意味する。

「全体」

…うなずきの開始時刻より前に相手の頭部動作が開始され、うなずきの終了時刻より後にその動作が終了したことを意味する。

「後」

…うなずきの開始時刻と終了時刻の間で相手の頭部動作が起こり、うなずきの終了時刻より後にその動作が終了したことを意味する。

「後(150ms)」

…うなずき終了後150ms以内で、相手の頭部動作が開始したものを意味する。

「内部」

…うなずきの動作の開始時刻以降に相手の頭部動作が開始され、うなずきの終了時刻以前にその

動作が終了することを意味する。

4. 3 うなずきと視線の分析

うなずきを生じる時の視線一致比率を分析した。

うなずきの開始から終了まで完全に視線が一致していたのは、全体の72%であった。また、うなずきを相手話者が完全に見ていなかったものは5%、相手を見ずにうなずきを行なったものは3%であった。

5. 考察

まず始めに、視線の分析結果について考察する。

表1より、話し手のときの方が聞き手のときよりも、視線をそらす割合が高くなっていることがわかる。これは[10]で報告された分析と同様の結果である。[10]の収録では2人の話者の自己紹介がテーマとして与えられており、知らない者同士の対話であったが、今回の結果から、仲の良い友人同士においてもこのような現象は見られるということがわかった。また、個人ごとの比率を見ると、ほとんどの話者が全体を通して、相手を見る比率が高いのに対し、話者5と話者11は比較的相手を見ない傾向がある。これは、個人差の影響であると考えられる。

続いて、頭部の分析結果について考察する。表2から、うなずきを発生する頻度には個人差があることがわかる。さらに、相対的に女性の方がうなずきを発生する頻度が高いという傾向が見られた。収録映像を見ても、女性のペア同士のほうが上半身や、頭部の動きが頻繁に生じていた。また、表3より、話者1, 2, 4を除く9話者が自己発話内で1分間に28回以上の割合でうなずきを生じており、相手の発話に対しては、話者7, 8, 9, 10, 12の5話者が20回以上の割合でうなずき(あいづち)を生じている事がわかる。このことから、本来なら、うなずきはあいづち的な機能を持つと考えられるが、実際には、あいづちなうなずきよりも自己発話内でのうなずきのほうが多く生じる傾向にあることが解かった。

続いて、表4について考察する。ある話者のうなずきが生じた際、相手話者は比較的高い割合でほぼ同時にうなずきを生じる事がわかる。中でも特に、話者7~12において高い割合でうなずきが生じており、逆に話者1~6においては比較的割合が低い。これは、

表2のうなずき頻度の高さが影響しているといえる。つまり、うなずきの頻度が高い話者同士はうなずきの同期も起こりやすく、低い話者同士は同期が起こりにくい。また、話者3や、話者6は比較的うなずき頻度が高いが、相手話者の頻度が低いために、互いのうなずきの発生比率は低くなっていると考えられる。この結果から、互いのうなずき頻度が高い場合、生じたうなずきには相手話者のうなずきの同期現象が伴われることが比較的多いことが示唆できる。

続いて、表5について考察する。表5-2から、あいづち以外のうなずきにおいて、あいづちが同期する場合、「後」「後(150ms)」のタイミングで生じることが多いといえる。また、同様に表5-3より、あいづちに対して、あいづち以外が同期する場合の比率を見ても「前」「前(150ms)」のタイミングで生じることが多いという結果になっている。また、あいづち同士のうなずきの同期は起こりにくいことがいえる。

これらの分析結果から総合して考察すると、うなずきは本来、相手の発話への同意の意味として、あいづちなどで生じることが多いと考えられるが、実際は、自己発話中に生じることが多く、さらに、そのうなずきに対して、相手話者も比較的多くの割合でうなずきを生じる事がわかった。つまり、[10]の分析であったような笑いの同期現象や、[12]の道案内で同じジェスチャーの動作をおこなう現象と同様に、うなずきにおいても同期現象は生じることが示唆された。このような、うなずきの同期現象の背景として、話者同士の会話のやり取りに対するコミュニケーションが考えられる。4.3章のうなずきと視線の分析からも示唆されるように、うなずき時には、互いが非常に高い割合で相手を見ている。このため、お互いの会話において、発話中にうなずきを生じることが、一種のコミュニケーションの役割をもっていると考えられる。つまり、話者の自己発話内のうなずきは、話に対する相手への投げかけを無意識的に示しており、それに対し、相手話者は同じうなずきを生じることで応えている。このような同期現象のやり取り

が、お互いの会話をスムーズに進めることに役立っていると考えられる。

6. まとめ

本論文では、プロンプターを用いた対面映像による自然対話の収録を行なった。そして、ANVIL という時間的相互関係を明確に記述できるアノテーションツールを用い、視線の動きやうなずきの相互関係を分析した。そしてその結果、うなずきは、自己発話内で生じることが多く、さらに、話者間での同期現象が起こりやすいことが示唆された。また、このような視線の動き、頭部動作の発生頻度には、個人差が見られた。その中で、女性のペアでは一貫してうなずく頻度が高くなっていった。よって、対話では、女性の方がうなずきを生じやすい傾向にあると考えられる。また、それ以外にも個人の性格の違いが影響したと考えられる。

今後はさらに収録データを増やし、分析結果の信頼性及び個人性について検討する。また、視線やうなずき以外のジェスチャーについて、また音声情報についても、それぞれの相互関係を分析する予定である。

参考文献

- [1]綿貫 啓子：マルチモーダル対話システムMA I C Oの試作，人工知能学会研究会資料，SIG-SLUD-A102(2001)
- [2]東條 剛史，松坂 要佐，小林 哲則：身振りによる対話調整機能を持つ対話ロボット，情報処理学会，2000-SLP-30，pp.33
- [3]川本 真一，下平 博，新田 恒雄，西本 卓也，中村 哲，伊藤 克，森島 繁生，四倉 達夫，甲斐 充彦，李 晃伸，山下 洋一，小林 隆夫，徳田 恵一，広瀬 啓吉，峰松 信明，山田 篤，伝康 晴，宇津 呂 武仁，嵯峨山 秀樹：擬人化エージェントツールの基本設計，情報処理学会，2002-HI-97，2002-SLP-40，pp.61-66
- [4]山下 洋一，喜多 竜二，峰松 信明，吉村 貴克，徳田 恵一，田村 正統，益子 貴史，小林 隆夫，広瀬 啓吉：マルチモーダルコミュニケーションのための音声合成プラットフォーム，情報処理学会，2002-HI-97，2002-SLP-40，pp.67-72
- [5]西本 卓也，新田 恒雄，足立 裕秋，桂田 浩一：対話システムにおけるタスク記述とプロトタイプ作成支援，情報処理学会，2002-HI-97，2002-SLP-40，pp.73-78
- [6]四倉 達夫，森島 繁生：擬人化音声対話エージェントのための表情合成技術，2002-HI-97，2002-SLP-40，pp.79-84
- [7]高西 淳夫：ヒューマノイド・ロボットの技術とその展望，人工知能学会研究会，SIG-Challenge-0216，pp.1-5
- [8]坂本 憲治，綿貫 啓子，外川 文雄：マルチモーダル対話解析，人工知能学会研究会資料，SIG-FAI-9401，pp.39-46(1994)
- [9]綿貫 啓子，関 進，三吉 秀夫：発話時の人間の振る舞い-マルチモーダル対話データの解析-人工知能学会研究会資料，SIG-SLUD-9902-9，pp.49-54(1999)
- [10]ADAM KENDON： Conducting interaction Some function of gaze direction in two-person conversation(3章)，Cambridge University(1990)
- [11]GEOFFREY BEATTIE： An Analysis of Speech and Non-Verbal Behaviour in Conversation, The organization of speaker's nonverbal behaviour in conversation(4章) Open University(1983)
- [12]中津 良平，石黒 浩，今井 倫太，小野 哲雄：身体表現を用いた人とロボットの共創対話，情報処理学会論文誌，vol.42.No.6，pp.1348-1358(2001)
- [13]Michael Kipp： ANVIL A Generic Annotation Tool for Multimodal Dialogue ，Eurospeech 2001，pp.1367-137
- [14]前田真季子，堀内靖雄，市川薫：話者交替における視線とうなずきの分析，人工知能学会研究会資料，SIG-SLUD-A201(2002)，pp.53-58