マルチモーダル音声対話コーパスの収録とうなずきの分析
Multimodal Dialogue Corpus and Analysis of Speaker's Nod

千葉大学大学院自然科学研究科
Graduate School of Science and Technology, Chiba University

市川熹
Akira Ichikawa

＜研究協力者＞

堀内靖雄，西田昌史
Yasuo Horiuchi, Masafumi Nishida

People use gestures like gaze and nod for smooth communication in spontaneous speech dialogue. In this paper, we will report our recordings of multi-modal dialogue corpus and introduce the analysis of speaker's nod as an example of analyses about gestures. 36 dialogues by four 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. Usual dialogue continues exchanging interlocutor's information with each other using speech and gestures and therefore it is supposed that there is correlation between speakers' gestures and listeners' reaction. We focused the speaker's nods in the final part of an utterance and the listener's reaction like nods or backchannels for the analyses of gestures. As a result, it was suggested that speakers' nod is caused more frequently in the final part of utterances than the middle part and that listeners show reaction like nods or backchannels frequently when the speaker nodded or said some typical words in the final part of utterance.
Key words: Spontaneous Speech, Gesture, Multimodal Dialogue Corpus, Prompter

1 はじめに

音声対話において表出する韻律（プロソディ）として、一般には基本周波数、パワー、時間情報などの音声情報が主に研究されているが、韻律の概念をもっと広くとらえた場合、視線やうなずきなどのジェスチャーも韻律情報と考えることができ、音声コーパスにこのようなジェスチャーをタグ付けする試みも盛んに行なわれている [1, 2, 3, 4, 5]。

本研究ではマルチモーダル音声対話の収録、ジェスチャー情報のアノテーション手法について検討を行なうと同時に自然対話におけるジェスチャーの分析を行なってきた [6, 7, 8, 9]。本稿ではマルチモーダル対話コーパスの構築、および、これまでに得られた分析結果のうち、話し手のうなずき・言語・韻律と聞き手の反応に関する分析結果 [10] を紹介する。

2 マルチモーダル対話コーパスの構築

2.1 対話収録環境

本研究ではまず始めにマルチモーダルな状況での自然対話の収録方法について検討を行なった。マルチモーダル対話の収録において、検討すべき条件として以下のものが挙げられる。

1. 自然な環境での対話
2. 両者の分離したクリアな音声の収録
3. 正面からの顔および上半身の画像収録

本来なら、これらすべての条件を満足する収録方法が望ましいが実現は困難である。そこで本研究では両者の分離したクリアな音声を録音するとともに、正面からの顔および上半身の画像収録を行なうため、独立した二つの防音室において、話者を正面から撮影できるプロンプタを用いて対話収録を行なった。

このような収録条件により、2 と 3 の条件は満たされるが、別室でのプロンプタを介した対話という特殊な収録環境により、1 の条件における自然さがそこなわれる可能性がある。しかし実際に収録を行なってみたところ、大学生の実験参加者はそのような環境にもすぐ適応し、自然な会話が収録された[1]。

プロンプタとはテレビ局などで利用される機器であり、カメラの直前に配置されたガラス（ハーフミラー）上に原稿を投影でき、カメラ目線のまま原稿を読むことが可能となる。本研究ではプロンプタを

[1] [1] でも同様の傾向が示されている。

二台用意し、お互いに相手の上半身の映像がプロンプタ上に投影される環境での対話収録を行なった。

本研究で作成したプロンプタは、下部に設置された液晶ディスプレィ上の相手話者の画像をハーフミラーを介して呈示すると同時に、ハーフミラーの背面から話者の正面画像を二台のカメラで収録することが可能である（図 1 参照）。プロンプタの外観（前面）とハーフミラー背面から対話者を見た様子を図 2 に示す。

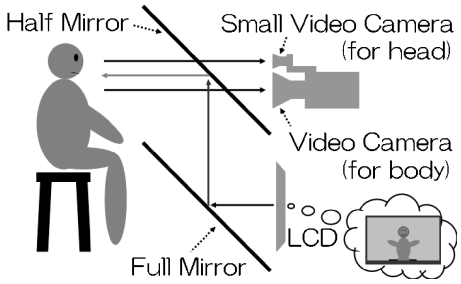

図 1: プロンプタ

図 2: プロンプタの外観（前面）と背面からの様子

プロンプタのハーフミラーの背面に設置された下側のカメラで上半身の映像、上側の小型カメラで顔の映像を収録することができる。相手話者には下側のカメラで撮影された上半身の映像が呈示される。

プロンプタでは各話者の顔の映像と上半身の映像が収録できるため、対話において画像情報は全部で四つとなるが、各画像の同期を正確にとるため、4 画面を一つに合成し、デジタルビデオデッキ (DV-CAM) で記録した。記録した映像の例を図 3 に示す。

なお、プロンプタではハーフミラー越しでの撮影となるため、高感度なカメラでの撮影が必要である。また、顔の部分に影が生じると視線の動きなどが不明確になるため、プロンプタの左右に照明を配置し、顔に影が生じることを防いだ。

2.2 音声情報の収録

今回の実験では各対話者が別々の独立した防音室に入って対話を行なうことにより、相手の音声がマイクに回りこむことを防ぎ、二人の話者の音声を完全に分離して録音した。音声のやり取りは片耳の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、相手の声だけをイヤホンから再生し、自分の音声はイヤホンをしていない耳から直接に聞こえるようにした。

図 3: 対話収録画像例

2.3 画像情報の収録

アイコンタクトやうなずき等のジェスチャー情報を記録するため、各被験者の正面にプロンプタを配置した。プロンプタは相手話者の画像を下部に設置された液晶ディスプレィからハーフミラーを介して呈示すると同時に、ハーフミラーの背面から話者の画像を二台のカメラで収録することが可能である。

2.4 実験参加者

実験参加者（大学生）は 4 人で 1 グループ（四人組）とした。各四人組は二組の話者ペアからなり、今回の収録ではこの四人は同性とした。各ペア内のメンバー二人はお互いによく知っている間柄である（親近性がある）が、同じ四人組の他方のペアのメンバーとはどちらの話者とも互いに面識がない（親近性がない）。

収録ではまず始めに親近性条件での対話を収録し、その後、相手を替えて非親近性条件（初対面条件）で対話を収録した。同じ話者同士での対話収録は 3 回ずつ行なった。各話者が三人の相手と対話を行なうため、四人組 1 組あたり 18 対話の収録が行なわれることとなる。数回の予備収録の後、本収録として男性と女性の四人組各一組ずつ収録を行なった。

2.5 対話内容

本研究では活発な会話を自然に行なえるようにするため、ディベート形式の対話を収録することとした。対話収録に先立ち、アンケート調査を実施し、

図 4: アノテーションツール ANVIL

被験者二人に意見が分かれそうな二者択一の質問にいくつか答えてもらい、二人の意見が分かれたテーマに関して自由に討論するよう教示した[2]。途中で議論が収束するか、あるいは約 5 分経過した段階で実験者が対話を打ち切り、対話の終了とした。

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表われ、視線一致やうなずき、身振り手振り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。

2.6 アノテーション

ジェスチャーのアノテーションはアノテーションツール ANVIL[4] を用いて行なった。このツールは自分の目的にあった情報を自由にアノテートでき、それらの時間的相互関係が明確に表記できる（図 4 参照）。本研究では「うなずき」、「視線（相手を見ている／そらしている）」をアノテートした。

しかしながら、ANVIL では映像のフレームレートである約 1/30 秒単位でしかアノテートできないため、音声情報のアノテーションには適さない。そのため、音声情報のアノテーションに関しては千葉大学で開発したツール群を利用した [11]。このツールにより、発話単位の正確な開始・終了時刻やオーバーラップの有無などが容易にアノテート可能である。

[2]質問の例としては、「男と女どっちが得だと思いますか？」「神様の存在を信じますか？」「機械が知能を持つことに賛成ですか？」「男女間に友情は成り立つと思いますか？」など。

3 話し手のうなずきの分析

近年の計算機能力の急速な発展により、人間と対話するコンピュータシステムの研究が盛んになってきている。ユーザフレンドリーなマンマシンインタフェースを目指すためには、人間らしい受け答えが可能となることが重要である。そのためには人間同士が行なっている自然対話における現象を分析し、システム上に実装する必要がある。しかしながら、文献 [12] などで指摘されているように、とくに日本語の場合、話し手と聞き手がお互いに密接に協調しながら対話を行なっていることが示唆されている。たとえば日常の会話では聞き手は話し手の発話を聞きながら、適切なタイミングであいづちやうなずきを返したりしている。このような聞き手のあいづちやうなずきなどの反応がないと話し手は相手が自分の発話を理解しているのか不安になってしまう。

このようにあいづちなどの現象は自然対話によるマンマシンインタフェースを考える上で非常に重要である。日本語のあいづちに関しては古くからさまざまな研究が行なわれており、興味深い結果が得られている [13, 12, 14, 15, 16, 17, 18, 19, 20]。また、分析して得られた知見を用いて、音声対話システムが適切なタイミングであいづち発話を生成する研究も数多く行なわれている [21, 22, 23, 24, 25, 26, 27, 28]。

さらに、あいづちなどの音声現象だけでなく、うなずきや視線などの非言語情報が対話でどのような機能を有しているのかを調べるため、マルチモーダル

な状況における対話収録が行なわれ[2, 4, 3, 5, 6]、うなずきなどのジェスチャーの研究も盛んになってきている[12, 29, 1, 30, 31, 32, 33, 7, 34, 35, 8, 9, 36, 37]。うなずきはあいづち同様、日本語の自然対話において頻出する現象であり、人間とのマンマシンインタフェースを考えるときに非常に重要なものとなる。

これまでのあいづちに関する研究では、おもに言語的な情報や韻律的な情報により、どのようなタイミングであいづちがうたれるのか、というようなことが中心に分析されてきている。一方、うなずきに関する研究ではうなずきと話者交替の関係などが論じられている。このように過去の多くの研究では、あいづちやうなずきが別個に扱われていたが、うなずきもあいづちの機能を有することがあり、文献[12]ではうなずきもあいづちとして分析されている。そこで本研究では、日本語の自然対話において、うなずきがどのような状況で発現し、それぞれがどのような機能を持っているかを分析し、これまであいづち研究で言われてきた言語的な情報や韻律的な情報との関係を明らかにすることを目的とする。

3.1 分析データ

以下の分析では収録された対話中から親近性条件の四対話約13分の対話データを分析対象とした。

3.2 発話単位

自然発話には句点や読点などの明確な区切りがなく、発話の単位を決定するのは難しい。そこで本研究では200ミリ秒以上の無声区間で分割された音声区間を分析単位IPU (Interpausal Unit) [19, 11]とした。

また、自然発話には笑いが多く表われるが、本研究では笑いは分析対象外とした。しかしながら笑いはあいづちなどと似た機能がある、という指摘もあるため、今後の検討課題としたい。

3.3 うなずき

本研究で対象とするジェスチャーはうなずきである。うなずきとは顔の縦方向の運動で典型的には下降上昇する動きとした。また、うなずきが一連の動作として、連続して複数回おこなわれた場合にはそれらを一つのうなずきとして分析を行なった。うなずきのタグ付けは二人の熟練した作業者が共同で行ない、不明な点については両者相談の上、決定した。

3.4 分析１：うなずきはいつ起こるのか

まず始めに日本語対話におけるうなずきがいつ起こるのか検討してみたい。うなずきが話し手、聞き手のどのような状況で発現するのかを調査した結果を表1に示す。ここで、発話直後とはそれぞれ話し手が発話の直後にうなずいた場合、聞き手が相手の発話の直後にうなずいた場合を表わしている。また、無音区間中とは、それぞれ、話し手が発話終了後しばらくたってから無音区間でうなずいた場合と聞き手が相手の発話終了後しばらくたってからうなずいた場合を表わしている。なお、表中、重複発話とは二人が同時に話している状況であり、うなずきが各話者の発話とどのような関係があるのかが不明なため、分析対象外とした。

表 1: 話し手／聞き手の状況とうなずきの出現頻度
(表中の数値は出現回数。太字は状況ごとの合計。)

	話し手	聞き手
発話途中	173	50
発話末	196	69
あいづち発話と同時	—	79
発話中のうなずき	**369**	**198**
発話直後	8	15
無音区間中	12	8
無音中のうなずき	**43**	
重複発話中のうなずき	47	
うなずきの合計	**657**	

この表からほとんどのうなずきはどちらかの発話中に行なわれていることがわかる。また、一般的には、うなずきはあいづちなどと同様、聞き手のときに表出されることが多いと考えられがちであるが、この表から、話し手のうなずきの方が多いことが示唆される。さらに、話し手のうなずきの中でも、話し手の発話末のうなずき頻度が高いことがうかがえる。何故なら、もし話し手が発話中、無作為にうなずきを行なっているとすると、発話末よりも発話途中の方がうなずきは多く発生すると考えられるが、それに比べると表1の発話末のうなずきの頻度は高い。すなわち、話し手は発話末にピンポイントでうなずきを発生させていることが示唆される。この発話末のうなずきはメイナードによるうなずきのカテゴリー分類[12]では3および4に相当する。文献[12]ではこのようなうなずきは発話末を明確にしたり、発話順番の終了を示すのではないか、というような考察が与えられているが、本研究では、この発話末のうなずきに着目し、その機能について、より詳細な分析を試みる。

3.5 分析２：発話末のうなずきと聞き手の反応

前節で示したように、話し手の発話末のうなずきの頻度は非常に高い。日本語の会話において、話し手の発話末のうなずきはどのような役割を果たしているのであろうか。そこで、発話末に話し手のうな

ずきがある場合とない場合に対し、聞き手の反応（あいづち、または、うなずき）がどのように行なわれるのかについて分析を行なった。

表 2: 話し手の発話末のうなずきに対する聞き手の反応（うなずき・あいづち）の出現頻度

		話者交替	聞き手の反応 うなずきあいづち	うなずきのみ	あいづちのみ	反応なし	合計
話し手のうなずき	あり	交替	— —	35 (44%)	— —	45 (56%)	80
		継続	31 (27%)	40 (35%)	11 (10%)	32 (28%)	114
	なし	交替	— —	20 (33%)	— —	40 (67%)	60
		継続	13 (10%)	17 (13%)	11 (8%)	91 (69%)	132

表 2 に話し手の発話末のうなずきに対する聞き手の反応を示す。表から、話し手の発話末にうなずきがある場合とない場合を比較すると、話者交替と話者継続、どちらの状況でも、うなずきがない場合よりもうなずきがある場合の方が聞き手の反応の頻度が高い。このことから、話し手の発話末のうなずきは聞き手への何らかの働きかけの機能があり、聞き手はそれに対して反応しているように見える。また、話者交替に比べ、話者継続時の方がより一層、その傾向が強いことがわかる。

3.6 分析３：言語情報・韻律情報と聞き手の反応

前節の分析より、話し手の発話末のうなずきには、話し手と聞き手の会話制御に関する何らかの機能が含まれていることが示唆されたが、ここで、あいづちに関する過去の研究 [12, 18] を考慮すると、言語情報や韻律情報も発話末のうなずきと同様に聞き手の反応との関係が深いと考えられる。

そこでまず、言語情報として、メイナード [12] らが指摘しているように、「ね」「さ」「よ」などの終助詞や間投助詞、あるいは、「じゃない」「でしょ（う）」などの助動詞の文末表現などがあいづちの直前のコンテクストに表われやすいということを考慮して、発話末の言語情報と聞き手の反応について分析を行なった。

表 3 に発話末にこれらの言語情報がある場合とない場合の聞き手の反応の違いを示す。なお、表中、「言語情報あり」とは、終助詞・間投助詞として「さ」「ね」「の」「よ」「な」「じゃん」、助動詞として「でしょ」と「だ」を分析対象とした[3]。

表 3 から、ある特定の助詞や助動詞が発話単位末に出現する場合としない場合を比較すると、ここでも上述の発話末のうなずき同様、話者交替と話者継続、どちらの状況でも、言語情報がない場合よりもある場合の方が聞き手の反応の頻度が高い。また、話者交替時に比べ、話者継続時の方がその傾向が強い。

[3]これら 8 個の助詞と助動詞は、今回の分析データに出現したもののみを列挙しており、それ以外に該当する助詞や助動詞は今回収録したデータには見受けられなかった。

表 3: 話し手の発話末の言語情報に対する聞き手の反応（うなずき・あいづち）の出現頻度

		話者交替	聞き手の反応 うなずきあいづち	うなずきのみ	あいづちのみ	反応なし	合計
発話末の言語情報	あり	交替	— —	28 (47%)	— —	32 (53%)	60
		継続	17 (24%)	22 (32%)	6 (9%)	24 (35%)	69
	なし	交替	— —	27 (34%)	— —	53 (66%)	80
		継続	27 (15%)	35 (20%)	16 (9%)	99 (56%)	177

次に発話末の韻律的特徴について検討を行なう。以前行なった研究 [16, 17, 18, 19] では、発話末の韻律的特徴が話者の交替や継続、あいづち現象などと関係があることが示唆された。そこで本研究では、韻律的特徴の中でもとくに顕著な山型イントネーション（尻上がりイントネーション）と発話末のモーラの音引き現象（山型イントネーションは除く）に焦点をあて、これらの韻律的特徴と聞き手の反応について分析を行なった。山型イントネーションとは最後のモーラが音引きされ、上昇下降調の独特のイントネーションを形成するものである。音引き現象とは発話単位の最後のモーラがそれ以前の平均的なモーラ長に比べ、引き伸ばされる現象である。おもに助詞などにその現象が見られるが、名詞の最後のモーラを伸ばす現象なども多々見られた。

表 4 に発話末の韻律的特徴（山型イントネーション・音引き現象・特徴なし）と聞き手の反応の関係を示す。

表 4: 話し手の発話末の韻律情報に対する聞き手の反応（うなずき・あいづち）の出現頻度

		話者交替	聞き手の反応 うなずきあいづち	うなずきのみ	あいづちのみ	反応なし	合計
発話末の韻律的特徴	山型	交替	— —	7 (32%)	— —	15 (68%)	22
		継続	21 (32%)	18 (27%)	4 (6%)	23 (35%)	66
	音引	交替	— —	7 (39%)	— —	11 (61%)	18
		継続	5 (14%)	4 (11%)	3 (8%)	24 (67%)	36
	なし	交替	— —	41 (41%)	— —	59 (59%)	100
		継続	18 (13%)	35 (24%)	15 (10%)	76 (53%)	144

表 4 を見ると、山型イントネーション、音引き現象が生じた場合、話者交替に比べ、話者継続の頻度が高くなっていることがわかる。このような現象は従来研究の結果と一致している。

また、山型イントネーションが生じた場合には聞き手の反応が非常に強くなっており、聞き手の反応をうながしていることが示唆される。一方、音引き現象に関しては韻律的特徴がない場合とほとんど差が見られなかった。これは、今回の分析では、音引き

表 5: 話し手のうなずき・言語情報・韻律情報と聞き手の反応の関係。各欄は反応のあり／なしの頻度と反応ありの比率

		韻律情報				合計	
		あり		なし			
		うなずき		うなずき		うなずき	
		あり	なし	あり	なし	あり	なし
言語情報	あり	13/5 (72%)	9/6 (60%)	21/8 (72%)	2/5 (29%)	34/13 (72%)	11/11 (50%)
		22/11 (67%)		23/13 (64%)		45/24 (65%)	
	なし	26/7 (79%)	8/29 (22%)	22/12 (65%)	22/51 (30%)	48/19 (72%)	30/80 (27%)
		34/36 (49%)		44/63 (41%)		78/99 (44%)	
合計		39/12 (77%)	17/35 (33%)	43/20 (68%)	24/56 (30%)	82/32 (72%)	41/91 (31%)
		56/47 (54%)		67/76 (47%)		123/123 (50%)	

現象を 1 パターンのみとしてまとめて扱ってしまったためではないかと考えられる。文献 [18, 19] などで指摘したように、音引き現象も様々なパターンがあるため、それらをいくつかのカテゴリーに分けて分析した方が傾向が出たのではないかと考えられる。

3.7 分析 4 ：話し手のうなずき・言語情報・韻律情報の相互的関係の分析

上述の分析では話者交替と話者継続をともに扱ったが、話者交替時に生じるうなずきはメイナード [12] のうなずきのカテゴリー 2 に相当し、聞き手が次に発話権を獲得し、発話を行なうことを示すものである。一方、話者継続時のうなずきはメイナード [12] のカテゴリー 1 に相当し、あいづちと同様に扱われる。すなわち、話者継続時のうなずき、あいづちは聞き手のあいづち的反応ととらえることができる。ところで、これまでの分析から、発話末の話し手のうなずき、特定の言語情報、特定の韻律的特徴がそれぞれ聞き手の反応をうながすことが示唆された。では、これら三つの要因間の相互的関係はどのようになっているのだろうか。そこで、話者継続時における上記三要因と聞き手の反応を表 5 にまとめた。なお、ここでは分析の見通しをよくするため、うなずき、あいづちの区別をせず、聞き手の反応として、うなずきかあいづちのどちらかがあった場合を反応あり、どちらもなかった場合を反応なしとして記述することとする。

表 5 から、おおむね各要因が存在するときに聞き手の反応が顕著であることがわかる。

表 5 にはすべての要因が書かれているが、各要因ごとの違いを明らかにするため、ここでは二要因ずつ別々に分析することとする。まず始めに、表 6 に発話末の話し手のうなずき、言語情報と聞き手の反応を示す。

表 6 から、話し手の発話末にうなずきがある場合には、非常に高い頻度で聞き手が反応していることがわかる (31%→ 72%)。一方、言語情報については、言語情報なしの 44%に比べ、言語情報ありの場合、65%となっているため、聞き手の反応に影響を与えていることは示唆されるが、うなずきほど強い要因とはなっていないようである。また、うなずきがある場合にはともに 72%となっていることから、うなずきと同時に発現してもその効果は薄れてしまうのかもしれない。逆にうなずきが生じていなければ、27%から 50%へと向上しており、やはり何らかの影響を及ぼしていると考えるのが妥当であろう。

表 6: 話し手のうなずき・言語情報と聞き手の反応の関係。各欄は反応のあり／なしの頻度と反応ありの比率

		うなずき		合計
		あり	なし	
言語情報	あり	34/13 (72%)	11/11 (50%)	45/24 (65%)
	なし	48/19 (72%)	30/80 (27%)	78/99 (44%)
合計		82/32 (72%)	41/91 (31%)	123/123 (50%)

次に、表 7 に発話末の話し手のうなずき、韻律情報と聞き手の反応を示す。

表 7: 話し手のうなずき・韻律情報と聞き手の反応の関係。各欄は反応のあり／なしの頻度と反応ありの比率

		うなずき		合計
		あり	なし	
韻律情報	あり	39/12 (77%)	17/35 (33%)	56/47 (54%)
	なし	43/20 (68%)	24/56 (30%)	67/76 (47%)
合計		82/32 (72%)	41/91 (31%)	123/123 (50%)

話し手の発話末にうなずきがある場合には、非常に高い頻度で聞き手が反応しているが、韻律情報に

ついては、聞き手の反応にはほとんど影響を及ぼしていないように見受けられる。しかしながら、うなずきがある場合に着目すると、韻律情報がない場合は 68% であるのに対し、韻律情報が付加されると 77%まで上昇している。このことは韻律が補助的に聞き手の反応をうながしている可能性を示唆しているが、それほど大きな差ではないので、今後、より詳細な検討が必要であると考えられる。

では、言語情報と韻律情報を比べるとどうなるであろうか。表 8 に発話末の話し手の言語情報、韻律情報と聞き手の反応を示す。

表 8: 話し手の言語情報・韻律情報と聞き手の反応の関係。各欄は反応のあり／なしの頻度と反応ありの比率

		韻律情報		
		あり	なし	合計
言語情報	あり	22/11 (67%)	23/13 (64%)	45/24 (65%)
	なし	34/36 (49%)	44/63 (41%)	78/99 (44%)
合計		56/47 (54%)	67/76 (47%)	123/123 (50%)

表 8 を見ると、どちらもそれほど高い要因とはなっていないが、お互いに補完しあって、影響を強めているようにも見える。また、その程度の違いから、韻律情報よりも言語情報の方が聞き手の反応に与える影響はわずかに大きいようである。

3.8 考察

これらの結果から、話し手のうなずきに関して、重要な示唆を得ることができた。まず第一に、話し手のうなずきにはメイナードの指摘 [12] にもあるように、さまざまな機能があり、その使われるコンテクストにより、その機能は異なると考えられる。とくに今回の分析結果を見ると、話し手の発話末のうなずきは発話単位の区切り、発話権の放棄の機能だけではなく、積極的に聞き手に反応をうながす機能を持っていることが推測される。また、発話末のうなずきは従来のあいづち研究などで言われてきた言語情報や韻律情報などと補いあって、聞き手の反応をうながしているが、その中でも発話末のうなずきはその要因が強いということが示唆される。

最後にひとつ検討事項が残っている。今回の分析では話者交替時についての詳細な分析を行なっていないが、話者交替時にも同じように話し手の発話末にうなずきが生じ、それを受けて、うなずいてから発話を開始する場面が非常に多く見受けられた。すなわち、発話末のうなずきは話者継続、話者交替の状況を問わず、ともに話し相手（聞き手）への積極的な会話への参加をうながすシグナルになっているのではないかとも考えられるが、今回のデータからではこれ以上のことは言えないので、今後、より詳細な分析を進めていきたいと考えている。

4 おわりに

本稿ではマルチモーダル対話コーパスの構築について述べるとともに、これまでに得られた分析結果のうち、話し手のうなずき・言語・韻律と聞き手の反応に関する分析結果を紹介した。分析においては、日本語の自然対話における話し手のうなずきに注目し、話者継続時における発話末の話し手のうなずきと聞き手のあいづちやうなずきとの関連性を分析した。また、従来のあいづち研究などから指摘されてきた言語情報や韻律情報との比較・検討を行なった。結果として、話し手の発話末のうなずきは聞き手の反応を強くうながす機能を有していることが示唆された。今後は考察で指摘したように、発話末のうなずきの後、聞き手のうなずきに続いて話者交替するケースなども含め、さまざまな状況におけるうなずきについて、より詳細な分析を行なっていく予定である。

参考文献

[1] 綿貫啓子, 関進, 三吉秀夫. 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. 情報処理学会研究報告 HI-84-5, 1999.

[2] Satoru Hayamizu, Shigeki Nagaya, Keiko Watanuki, Masayuki Nakazawa, Shuichi Nobe, and Takashi Yoshimura. A multimodal database of gestures and speech. In *Proceedings of the 6th European Conference on Speech Communicational Technology*, pp. 2247–2250, 1999.

[3] Tony Bigbee, Dan Loehr, and Lisa Harpr. Emerging requirements for multi-modal annotation and analysis tools. In *Proceedings of the 7th European Conference on Speech Communicational Technology*, 2001.

[4] Michael Kipp. Anvil a generic annotation tool for multimodal dialogue. In *Proceedings of the 7th European Conference on Speech Communicational Technology*, 2001.

[5] Ulrich Turk. The technical processing in smartkom data collection: a case study. In *Proceedings of the 7th European Conference on Speech Communicational Technology*, 2001.

[6] 前田真季子, 堀内靖雄, 市川熹. 音声対話コーパスにおける画像情報のアノテーション手法の検討. 情報処理学会研究報告, Vol. 2001, No. 38, pp. 21–28, 2001.

[7] 前田真季子, 堀内靖雄, 市川熹. 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-A201-09, pp. 53–58, Jun. 2002.

[8] 前田真季子, 堀内靖雄, 市川熹. 自然対話におけるジェスチャーの相互的関係の分析. 情報処理学会 研究報告, Vol. 2003, No. 9, pp. 39–46, Jan. 2003.

[9] 前田真季子, 西田昌史, 堀内靖雄, 市川熹. 自然対話における発話者のうなずきに対する聞き手の反応. 人工知能学会研究会資料 SIG-SLUD-A302-07, pp. 35–42, Nov. 2003.

[10] 堀内靖雄, 庵原彩子, 西田昌史, 市川熹. 自然対話における聞き手の反応と話し手のうなずき・言語情報・韻律情報との関係に関する予備的検討. 情報処理学会 研究報告, 2004-SLP-52-18, Jul. 2004.

[11] 堀内靖雄. 音声対話コーパス作成, 分析ツールについて:日本語地図課題対話コーパス作成の経験から. 人工知能学会研究会資料 SIG-SLUD-9902-5, pp. 23–28, Oct. 1999.

[12] 泉子 K メイナード. 会話分析. くろしお出版, 1993.

[13] 小坂直敏. あいづちを中心とした会話音声の呼応関係の分析. 電子情報通信学会 信学技報 SP87-107, pp. 61–66, 1987.

[14] 島津明, 川森雅仁, 小暮潔. 対話の分析:間投詞的応答に着目して. 電子情報通信学会 信学技報 NLC93-9, pp. 567–574, 1993.

[15] 片桐恭弘, 川森雅仁, 島津明. あいづちの分散システムモデル. 言語処理学会第 1 回年次大会予稿集, 1995.

[16] 小磯花絵, 堀内靖雄, 土屋俊, 市川熹. 下位発話単位の音声的特徴と「あいづち」との関連について. 人工知能学会研究会資料 SIG-J-9501-2, pp. 9–16, Dec. 1995.

[17] 堀内靖雄, 小磯花絵, 土屋俊, 市川熹. 自発的音声対話における話者交替の制御に関わる発話末の統語的・韻律的特徴. 情報処理学会 研究報告, Vol. 96, No. 21, pp. 45–50, Feb. 1996. (96-SLP-10-9).

[18] 小磯花絵, 堀内靖雄, 土屋俊, 市川熹. 先行発話断片の終端部分に存在する次発話者に関する言語的・韻律的要素について. 電子情報通信学会 信学技報, Vol. 95, No. 600, pp. 25–30, Mar. 1996. (NLC95-72).

[19] Hanae Koiso, Yasuo Horiuchi, Syun Tutiya, Akira Ichikawa, and Yasahuru Den. An analysis of turn-taking and backchannels based on prosodic and syntactic features in japanese map task dialogues. *Language and Speech*, Vol. 41, No. 3-4, pp. 291–317, 1998.

[20] 野口広彰, 片桐恭弘, 伝康晴. 尤度付きあいづち生起文脈コーパスの提案. 人工知能学会研究会資料 SIG-SLUD-A101-6, pp. 25–32, 2001.

[21] 菊池英明, 工藤育男, 小林哲則, 白井克彦. 音声対話インタフェースにおける発話権管理による割込みへの対処. 電子情報通信学会論文誌 D-II, Vol. J77-D-2, No. 8, pp. 1502–1511, 1994.

[22] Nigel Ward. In japanese a low pitch means "back-channel feedback please". 情報処理学会 研究報告 SLP-11-2, pp. 7–12, 1996.

[23] 平沢純一, 川端豪. わかってうなずくコンピュータの試作. 情報処理学会 研究報告 SLP-19-20, pp. 131–138, 1997.

[24] 平沢純一, 川端豪. 音声対話システム noddy : ユーザ発話途中でのうなずき・相槌生成. 情報処理学会 研究報告 SLP-20-9, pp. 51–52, 1998.

[25] 岡登洋平, 加藤佳司, 山本幹雄, 板橋秀一. 韻律情報を用いた相槌の挿入. 情報処理学会論文誌, Vol. 40, No. 2, pp. 469–478, 1999.

[26] 向井理朗, 関進, 中沢正幸, 綿貫啓子, 三吉秀夫. 力学系モデルに基づくマルチモーダル対話システムの試作. 1999 年度人工知能学会全国大会予稿集, 1999.

[27] 向井理朗, 関進, 中沢正幸, 綿貫啓子, 三吉秀夫. 感情を持つマルチモーダル対話エージェントシステムの提案. 人工知能学会研究会資料 SIG-SLUD-9902-3, 1999.

[28] M. Takeuchi, N. Kitaoka, and S. Nakagawa. Timing detection for realtime dialog systems using prosodic and linguistic information. In *Proceedings of Speech Prosody 2004*, pp. 1177–1180, 2004.

[29] 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-9401-6, pp. 39–46, 1994.

[30] 綿貫啓子, 関進, 三吉秀夫. 発話時の人間の振舞い : 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.

[31] 渡辺富夫, 大久保雅史. 身体的コミュニケーション解析のためのバーチャルコミュニケーションシステム. 情報処理学会論文誌, Vol. 40, No. 2, pp. 670–676, 1999.

[32] 綿貫啓子, 関進, 三吉秀夫. 対話過程における身体の動きと音声との関係. 2000 年度人工知能学会全国大会予稿集, 2000.

[33] Keiko Watanuki, Susumu Seki, and Hideo Miyoshi. Turn taking and multimodal information in two-people dialog. In *Proceedings of International Conference on Spoken Language Processing*, pp. 657–660, 2000.

[34] 江尻康, 小林哲則. 対話中における頭部ジェスチャの認識. 電子情報通信学会 信学技報 PRMU2002-61, pp. 31–36, Jul.

[35] 松坂要佐, 江尻康, 小林哲則. 合意形成型対話における声・顔の機能的表情の分析と認識. 電子情報通信学会 信学技報 PRMU2002-118, pp. 31–36, Nov. 2002.

[36] 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. 人工知能学会研究会資料 SIG-SLUD-A203-07, pp. 41–46, 2003.

[37] 重本佳孝, 内山智之, 森本一成, 黒川隆夫. 対面対話およびビデオ対話における頷き・瞬目・指差しの生起特性. ヒューマンインタフェースシンポジウム 2003 論文集, pp. 321–324, 2003.