

会話における非言語行動のイベント間関係に着目した 解析支援ツールの作成

A Tool for Supporting Analysis of Relations between Events of Nonverbal Behaviors in Conversations

野尻 篤* 平田 純也 高橋 和子 巳波 弘佳
Atsushi Nojiri Junya Hirata Kazuko Takahashi Hiroyoshi Miwa

関西学院大学大学院 理工学研究科
Graduate School of Science&Technology, Kwansai Gakuin University

Abstract: This paper describes mechanical analysis of relations between nonverbal behaviors in human conversation. These behaviors are taken as events of time sequence. We develop a tool for supporting analysis of these relations focusing on the beginning and ending of an event, as well as its length. It can reduce a user's burden and enables an objective analysis by handling events as numerical data. We apply this tool on an analysis of our conversation data. As a result, several new relationships between events and precise timing of their occurrences could be found. In addition, several existing relationships were confirmed, that shows the effectiveness of our approach.

1 はじめに

会話は一般に発話内容以外にも非言語行動に影響されながら進行し、会話ロボットの小さな仕草でも人間に与える影響は大きいことがわかっている [1]。また、非言語行動は一定の構造を持って行われていると考えられており、それを明らかにすることで、人工物との自然な会話が可能であると思われる [2]。本研究では非言語行動に着目した会話の解析支援ツールについて述べる。

会話時の非言語行動の取得にはカメラ、マイク、センサ、アイトラッカーなどが使われ、これらの性能向上とともに非言語行動に焦点を当てた研究も多くなってきた。たとえば、うなずきと発話の関係に関する市川の研究 [3]、視線と話者交替の関係に関する榎本らの研究 [4]、指差しと発話の関係に着目した中田ら [2] や福間らの研究 [5] などがある。しかし、得られたデータから非言語行動の発生時刻を切り出すのは、アノテーションツールを使ったタグ付けなどによる方法が多く、この方法では手間がかかる上に解析者の主観が入ってしまう。そこで、本研究では取得したデータからイベントという形で非言語行動の開始、終了時刻を自動的に切り出し、その後、ツールを使ってイベント間の関係を解析するという手法をとる。

また、従来の研究では非言語行動が具体的にどのようなタイミングで発生しているのか、例えば、同期して発生している中でもどの部分で発生していることが多いのかについて言及されているものは少ない。さらに、非言語行動の関係性がイベントの長さなどに依存するものかどうかについても明らかにされていない。本研究ではこのような点に着目して解析ができるような支援ツールを作成する。

本手法の妥当性と作成したツールの効用を調べるために研究室で収録した会話データに対して、ツールを使った解析を行なった。その結果、イベントの長さの特性、イベントの発生タイミングの特性などを新たにいくつか発見することができた。また、既存の研究結果も得られたことから提案手法の妥当性が示された。

本論文の構成は以下の通りである。第2章では非言語行動を取得するデータ収集の方法やイベントの抽出方法について説明する。第3章では会話時イベントとリレーションについて述べる。第4章では開発したツールについて述べる。第5章でこのツールを使用した解析実験、およびその結果について述べる。第6章で結論を述べる。

*連絡先：関西学院大学大学院理工学研究科情報科学専攻
〒669-1337 兵庫県三田市学園2丁目1番地
E-mail: cmg68379@kwansai.ac.jp

2 モーションキャプチャを使ったデータ収集

2.1 データ収集と補正

モーションキャプチャを使って会話時の非言語行動データの収録実験を行った。3人の内1人にモーションキャプチャのマーカーを装着して、各マーカーの位置データを取得した。非言語行動として、うなずき、視線、ジェスチャーのデータを取得するため、両額、後頭部、あご、首（鎖骨の上辺り）、両肩、両肘、両手首、両膝の計13ヶ所にマーカーを取り付けた。3人で1グループとし、1グループにつき10分間の実験を行い、10グループ分、計100分間の会話を収録した。各グループ内のメンバーはお互いによく知っている間柄であり、日常的话题を複数用意し、その中から1つ選択して、それについて自由に会話を行ってもらった。

1秒を60フレームとし、各実験で360,000フレーム分の数値データを得た。各マーカーごとに3次元の座標をとっているため各フレームには $3 \times 13 = 39$ 個のデータが含まれる。このデータに対して3次元補完と線形補完により欠損データについて修正を行った。また、現状において、モーションキャプチャによる非言語行動データの自動抽出は精度が良いものとは言えず、ノイズが発生しているものも多い。そのため、本実験においては自動抽出されたデータのうち、明らかなノイズを手作業において修正する作業を行った。予備実験から統計的にうなずきと視線のイベントを数値的に定義し、ビデオ画像で妥当性を検証したのち、これらの定義を確定した。取得したデータをプログラムにかけて各イベントの列を抽出し、これを解析ツールへの入力データとした。なお、ジェスチャーは出現回数が少なかったため、今回は解析対象から除外した。

2.2 イベントの定義

確定したイベントの定義を以下に示す。

2.2.1 うなずき

フレーム f_i における顎と首の距離を d_i とすると、列 $s = d_b, d_{b+1}, \dots, d_{e-1}, d_e$ が以下の条件を満たすとき、この列をうなずきというイベントと定義する。

$$(\text{条件}) d_{b-1} \leq d_b, d_b > d_{b+1}, d_{e-1} < d_e, d_e \geq d_{e+1}$$

2.2.2 視線

フレーム f_i における左(右)の額と首の距離をそれぞれ Lfn_i, Rfn_i とし、 $l_i = Lfn_i - Lfn_1, r_i = Rfn_i -$

Rfn_1 とする。1フレーム目を基準とするのは、体格によって個人差があるためである。

列 $s = l_b, l_{b+1}, \dots, l_{e-1}, l_e$ が以下の条件を満たすとき、この列を左方向を向いているイベントと定義する。

$$(\text{条件}) l_{b-1} \leq 17, l_b < 17, l_{e-1} < 17, l_e \geq 17$$

右方向についても同様に定義する。

2.2.3 発話

発話については、ビデオデータから手作業で発話の開始フレームと終了フレームを抽出し、開始から終了までをイベントとして定義する。

3 イベント間の関係

本研究では会話時イベントとして「発話、視線、うなずき」をとりあげる。イベントの開始時刻を begin、終了時刻を end とする。視線は右方向、左方向の2通り、発話は3人のものを別個のイベントとする。

リレーションとはイベント E1 とイベント E2 のそれぞれのイベントの起こっている時間の関係性のことである。本研究では Allen の分類に基づき [6]、2つのイベントのリレーションを時間的順序関係を考慮して以下の13種類に分ける(図1)。特に重要な概念として、同期性と引き起こしを定義する。

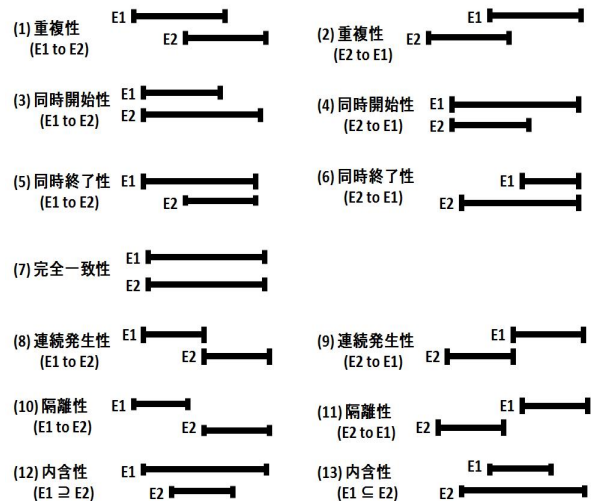


図1: イベントのリレーション

- 同期性: E1 が発生している最中に E2 が発生しているとき、E1 と E2 が同期するという。13種類のリレーションの中の(1),(4),(5),(7),(12)が相当する。

- 引き起こし：E1 の発生後に E2 が発生しているとき、E1 が E2 を引き起こすという。13 種類のリレーションの中の (8),(10) が相当する。

4 解析支援ツール

4.1 概要

複数のイベントが記述されたデータファイルを読み込み、そこに記述されたイベント間の関係について解析を行うツールを Windows Vista 上に Java で実装した。本ツールを使って、イベントの begin、end に着目することでイベントの発生のタイミングや発生順序など会話時イベントの詳細な部分を調べることができる。

まず、ユーザはダイアログ画面から入力データファイルと解析パラメータを指定する。システムはそれに対して第 1 解析、第 2 解析と 2 段階の解析を行い、計算して結果を出力する。結果をグラフで表示するとともにイベントの図も表示し、視覚的に理解することも可能である。

入力データファイルには対象イベント名のリストと各イベントの begin、end のフレーム番号の列が記述されている。解析パラメータとして、リレーション対象範囲と一致許容条件の 2 つを指定する。

リレーション対象範囲とは、イベント同士がリレーションがあると見なす範囲である。イベント同士があまりに離れすぎているとそれらは隔離性のリレーションではなく、関係性がないと捉えるべきである。したがって、どこまでの範囲をリレーションの対象とするのかを指定する。一致許容条件とは、begin や end のフレーム番号の誤差の許容範囲である。例えば、図 2 のようなリレーションを考えると、図のイベントだけに注目すると隔離性のリレーションであると考えられることができる。しかし、時系列上で見たときこれが非常に短い間隔であれば、このイベント同士は隔離性ではなく、end と begin が一致した連続発生性であると考えられるべきである。このように、どこまでの範囲を時系列上で一致しているかを見なすのかを指定する。

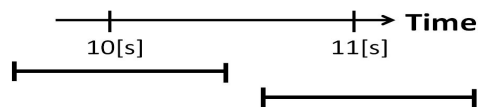


図 2: 一致許容条件 例

4.2 第 1 解析

第 1 解析は各イベントの発生回数、イベントのペアに対する 13 種類の各リレーションの発生回数などを出力するもので、ユーザが結果を概観し、解析の指針を決めるためのものである。これにより、発生タイミングを算出したり、どのイベント同士がどのような関係があるのかを推測したりすることができる。図 3 は第 1 解析の出力画面である。

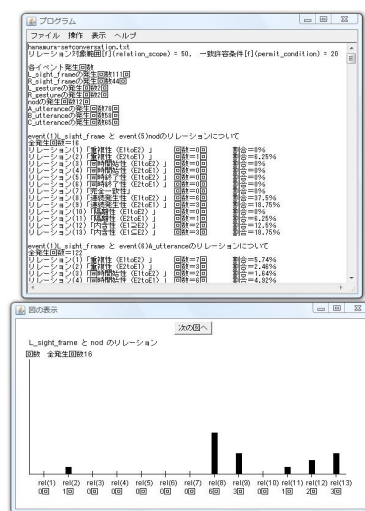


図 3: 第 1 解析 出力例

4.3 第 2 解析

第 2 解析では解析内容が異なる解析項目がいくつかあり、ユーザは目的に合った項目を選択し、解析対象イベントや比較対象イベントとそのリレーションなどを指定する。これによりユーザの調べたいイベントやイベントの特性など目的に絞った解析をすることができ、会話分析において解析頻度の高い解析を自動的に行うことが可能である。比較対象を指定することで、うなずきと発話の関係性の中で同期性である確率やどのイベントのペアが同期性を持つリレーションとして多く出現しているのかなどを調べることができる。また、イベントの長さに着目した解析も可能である。イベントとリレーションを指定すれば、それぞれのイベントの平均の長さとそのリレーションを満たすときの平均の長さを入力する。これによりリレーションの発生がイベントの長さ依存するのかを調べることができる。図 4 は同期性を調べる時の入出力画面であり、図 5 はイベントの長さを調べる時の入出力画面である。

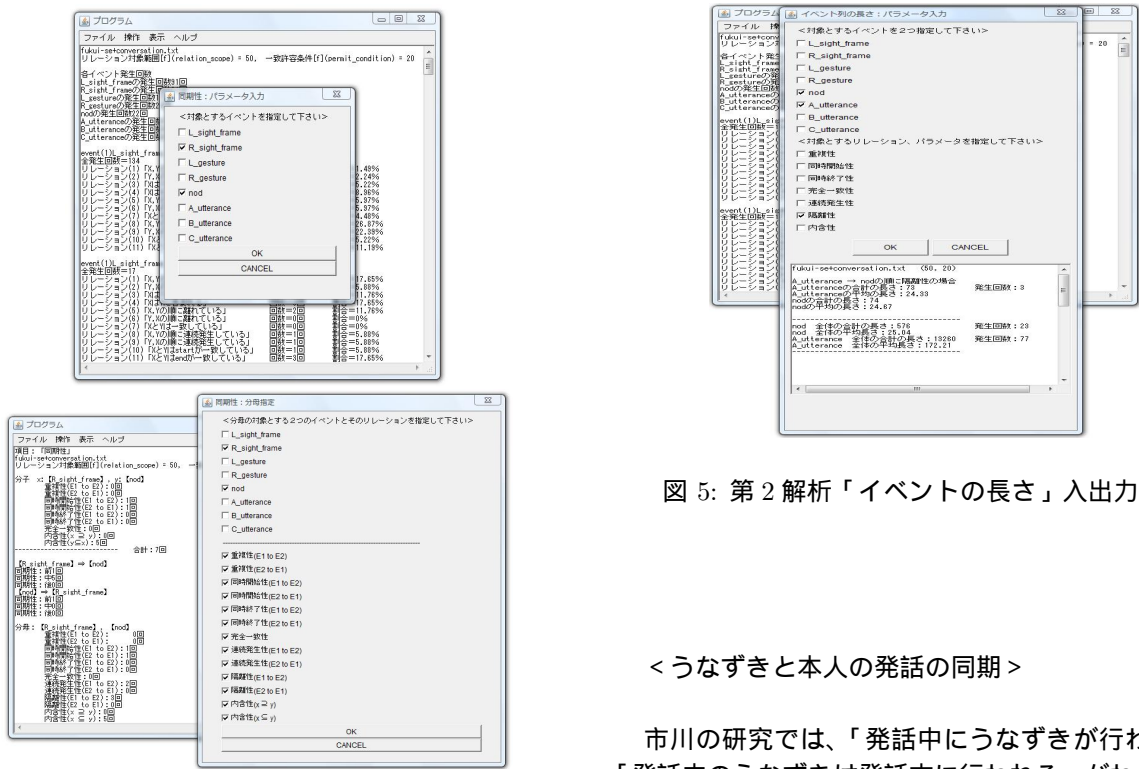


図 5: 第 2 解析「イベントの長さ」入出力画面

〈うなずきと本人の発話の同期〉

市川の研究では、「発話中にうなずきが行われる」、「発話中のうなずきは発話末に行われる」がわかっている [3]。そこで、うなずきと本人の発話の全てのリレーションの中で同期性がどの程度見られるか、うなずきが発話中のどの場所で発生しているのかを調べた。その結果を表 1, 2 に示す。

図 4: 第 2 解析「同期」入出力画面

5 ツールの応用

5.1 解析パラメータ

リレーション対象範囲と一致許容条件の 2 つのパラメータの値を設定するにあたり、2 つの方法を検討した。1 つは会話の様子を観察して行う目視による設定、もう 1 つは既存研究の結果から逆算して算出する既存研究から設定である。既存研究から算出されたパラメータの値でリレーションの様子を会話動画で確認したところ、不自然であると感じた。非言語行動の関係性や発生タイミングなど、会話に関わるものは人間が違和感を感じないものにするのを優先するべきことだと考えるので本研究においては、目視によるパラメータ付けを採用し、リレーション対象範囲を 50 フレーム、一致許容条件を 20 フレームとすることにした。

5.2 結果：既存研究の関係性の自動抽出の確認

本研究では数値データによってイベントを定義し、自動的にイベントを抽出している。この手法の妥当性を確認するため、既存の研究で判明している非言語行動間の関係性が得られるかどうかを調べた。

表 1: うなずきと発話「同期」

	回数
うなずきと発話の同期	70 回
うなずきと発話の同期以外	27 回

表 2: 発話とうなずきの発生場所

	回数
発話イベントの開始付近のうなずき	21 回
発話イベントの中央付近のうなずき	13 回
発話イベントの終了付近のうなずき	36 回

表 1, 2 より、うなずきと発話の関係性は市川の研究結果とほぼ一致する。

<視線と本人の発話と本人以外の発話>

榎本らの研究では、話者は次話者に視線を向けている場合が多いことが示されている [4]。そこで、発話者の視線と次の発話者の関係を調べた。その結果を表 3 に示す。

表 3：視線交替時の視線の方向

	回数
次話者	301 回
非次話者	152 回
視線をそらしている	175 回

表 3 よりこの結果は榎本らの研究結果と一致する。

<まとめ>

これらの結果により、従来の研究で得られた解析結果を数値データに基づく手法で自動抽出する手法の妥当性が確認された。

5.3 結果：新たな非言語行動の関係性

次に今回新たに得られた関係性について述べる。

<うなずきと発話の関係性 1 (話し手の場合)>

話し手本人のうなずきと発話の関係性について調べた。発話イベントの長さに着目したところ、本人の発話イベントが短い時に引き起こしの関係、つまり、発話が行われた後にうなずきが発生する関係が多く発生していることが判明した (表 4)。

表 4：本人のうなずきと発話

	イベントの長さ
全発話の平均	112 フレーム
引き起こしの発話の平均	51 フレーム

<うなずきと発話の関係性 2 (聞き手の場合)>

聞き手のうなずきと発話の関係性について調べた。発話イベントの長さに着目したところ、本人以外の発話イベントが短い時に引き起こしの関係が多く発生していることが判明した (表 5)。

表 5：本人のうなずきと本人以外の発話

	イベントの長さ
本人以外の平均発話	123 フレーム
引き起こしの A 以外の平均発話	73 フレーム

<視線と発話の関係性>

話し手の発話と視線の関係性について調べた。視線イベント begin と発話イベント begin に着目して解析したところ、発話イベントの begin は視線イベント begin よりも先に発生していることがわかった (表 6)。

表 6：視線 begin と発話 begin のタイミング

	回数	割合
発話 begin	496 回	53 %
視線 begin	297 回	32 %
begin の一致	144 回	15 %
合計	937 回	

さらに、視線イベント begin と発話イベント end に着目して解析したところ、発話イベント end と視線イベントの begin が同時に起きていることがわかった (表 7)。

表 7：視線 begin と発話 end のタイミング

	回数
発話 end	75 回
視線 begin	116 回
begin、end の一致	143 回

<発話同士の関係性>

同一グループのメンバーそれぞれの発話の関係性について調べた。あるメンバーの発話イベント end と別のメンバーの発話イベント begin に着目して解析したところ、これらは引き起こしの関係であり、その中でも発話同士が連続して発生 (約 0.33 秒) に発生していることが多いことがわかった (表 8)。

さらに、引き起こしの関係の中の隔離性について調べたところ、これらの発話間の距離のほとんどが 30 フレームほど (約 0.5 秒) であることがわかった (表 9)。

表 8 :各発話のリレーション

	回数
連続発生性	673 回
隔離性	521 回
合計	1194 回
全リレーション	1659 回
引き起こし割合	72 %

表 9 :各発話間の隔離性のフレーム距離

	フレーム距離
隔離性平均距離	30.2 フレーム

5.4 新たに発見された関係性のまとめ

前節で述べた結果から新たに発見された関係性についてまとめる。

<うなずきと発話の関係性>

話し手である場合でも、聞き手である場合でも発話のイベントが短い場合、発話とうなずきは引き起こしの関係、つまり、発話が行われた後にうなずきが発生するケースが多いことがわかった。従来の会話分析の研究では、発話中にうなずきが発生していることが多い、つまり同期性については指摘されているが [3]、本研究ではうなずきと発話の引き起こしの関係性の特性、そして、発話イベントの長さの特性を明確にすることができた。

<視線と発話の関係性>

発話の開始時、発話を行ってから相手に視線を向け、発話の終了と同時に別の相手に視線を向けていることがわかった。従来の会話分析の研究において、視線配分を行いながら発話を行なっていることは示されている [4]。しかし、その発生順序や発生タイミングについては曖昧な表現のままであった。本研究ではそれらを具体的な数値として示すことができた。

<発話同士の関係性>

ある人物の発話が行われてから別の人物の発話が行われるまでの時間はその多くが 0.33 秒以内であり、時間がかかっていたとしても 0.5 秒ほどで発話が行われていることがわかった。つまり、円滑な会話では 0.5 秒

以内に次の発話が行われていることがわかった。従来の会話分析において、それらが引き起こしの関係性であることは自明として扱われており、それ以上の研究は見当たらない。本研究では、発話間の離れ具合を具体的な数値として示すことができた。

6 まとめ

本研究では会話時イベントの開始、終了時刻に着目し、イベント間の関係性を調べるツールを作成した。これを使って解析実験を行い、今までの会話分析では得られなかった特性や現象を発見することができた。特にイベントの発生のタイミングや順序、イベント間の関係性がイベント間の距離や長さに依存するものが存在することを明らかにした。

これらの結果は会話分析の研究を促進させるだけでなく、得られた関係性をロボットに取り込むことで、より自然な会話をするロボットの開発にも役立つと考えられる。今後は対象イベントの種類を拡張し、着目すべき事象をツールに取り入れることで、ツールの利便性を向上させることを考えている。

参考文献

- [1] 上田博唯, “対話型ロボットの小さな仕草がひとに与える印象”, 第 5 回身体知研究会, pp.23-28, (2009)
- [2] 中田篤志, 角康之, 西田豊明, “非言語行動の出現パターンによる会話状況の特徴抽出”, 情報処理学研究所 Vol.2009-HCI-135 No.13, (2009)
- [3] 市川薫, “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”, 千葉大学大学院 自然科学研究科, (2004)
- [4] 榎本美香, 伝康晴, “3 人会話における参与役割の交替に関わる非言語行動の分析”, 人工知能学会研究会資料, SIG-SLUD-A301, pp.25-30, (2003)
- [5] 福間良平, 角康之, 西田豊明, “人のインタラクションに関するマルチモーダルデータからの時間構造発見”, 情報処理学会研究報告 Vol.2009-UBI-23 No.23, (2009)
- [6] James F.Allen and George Ferguson “Actions and Events in Interval Temporal Logic” Technical Report 521 The University of Rochester Computer Science Department Rochester, (1994)