

異なる会話環境におけるうなずきの機能分析

齊 賀 弘 泰[†] 角 康 之[†] 西 田 豊 明[†]

会話において、うなずきは聞き手の行うフィードバックとされている。他方で話し手が肯定や強調、発話の前振りを行うなど会話の流れを調節しているという研究がある。そこで本研究ではうなずきを話し手、聞き手両方が行う行為であるとし、首振り動作と定義する。そして首振り動作に対して会話調節の機能を分類するため、言語・非言語行動の発生パターンによってクラスタリングを行い、機能分析を行った。結果として同じ聞き手としての首振り動作でも、受け手的に行うものと傍参与者的に行うものが得られた。また異なる会話環境間でも異なった機能のクラスが生成された。

Functional Analysis of Head Nodding in Multi-party Conversations

HIROYASU SAIGA,[†] YASUYUKI SUMI[†] and TOYOAKI NISHIDA [†]

In face-to-face communication, nodding has important functions of visual feedback to a speaker. On the other hand, previous studies showed that speakers nod in order to agree or emphasize an important point, and to show that they start speaking in order to adjust conversation. In this paper, we defined head nodding as “vertical head movements” that both speakers and listeners use. To analyze “vertical head movements”, we used clustering against patterns of verbal and non-verbal behavior. As a result, we acquired several clusters of listeners’ vertical head movements, for example, doing as an addressee and as a side participant. Moreover, between two conversation environments, we detected common and different clusters.

1. はじめに

人は会話において、言語以外にも身振りや視線移動といった非言語的な情報で意図を伝える他にも、会話のリズムを調整している。その中でもうなずきは聞き手が相手に対して行うフィードバックとして、非常に重要な機能を持っているとされる。他方で、うなずきは話し手の行う強調やリズム取りの他、誰も話していない状況での発話の前振り行動など、聞き手以外も使う行動であるという研究¹⁾がある。しかしこれらの機能は頭部の動きだけでは理解できないと考えられる。

そこで本論文では話し手、聞き手を問わず頭部の鉛直な動きを“首振り動作”と定義し、うなずきの機能を機械的に理解することを試みる。今回は発話したタイミングだけでなく、相槌、首振り動作の発生パターンを入力とし、クラスタリングを行う。異なる会話環境において共通する機能を持つクラスや環境に依存したクラスを見つけ、それに寄与した周囲の振る舞いを分析する。それにより首振り動作が周囲の状況よりどのように理解できるかを調べる。

2. うなずきの機能と本研究の目的

会話分析やジェスチャー研究において、従来うなずきは聞き手行動としての頭部ジェスチャーの1つとして考えられていた。他方で話し手の行ううなずきに注目し、うなずきの会話の流れの調節機能を調べた研究として Maynard¹⁾があげられる。Maynard はうなずきを聞き手の行う相槌としての機能だけでなく、話し手の行う肯定、強調、リズム取り、さらに発話前に発話意図を示す機能があるとした。

しかし Maynard はうなずきの機能のカテゴリは提供しているが、どのような言語、非言語行動が共起しているかといったことには注目していない。また発話のタイミングのみに注目しているため、聞き手のうなずきは1つのカテゴリにまとめられている。これに対し前田ら⁵⁾によると、話し手は聞き手の反応を得るため働きかけの機能を持ったうなずきを行い、聞き手はそれに対して、応答的に頷き返すという分析をしている。さらにこういったうなずきの他、聞き手には発話継続を促すうなずきを行うといわれており、同じ聞き手のうなずきでも他の共起現象により機能が異なると考えられる。このようにうなずきの機能を理解するには周囲の振る舞いを知る必要がある。

[†] 京都大学大学院情報学研究科
Graduate Informatics of Kyoto University

そこで本論文では Maynard と同様に、うなずきは話し手も行う非言語行動であるとし、発話だけでなく、視線や他者のうなずきとの共起関係から、うなずきの機能を機械的に理解することを試みる。本論文ではうなずきを聞き手、話し手問わず行う行為として「首振り動作」と定義し、首振り動作に対し言語、非言語行動の共起パターンを入力データとしたクラスタリングを行う。各クラスタの機能を分析し、クラスタリングに特徴を与えた振る舞いを調べ、機械的に理解を行えるかを検証する。

3. 多人数会話のデータ収録

3.1 収録会話設定

分析対象とする会話データは図 1 左のようなポスター発表型の会話と、図 1 右のようなポスターを対象物とした自由環境型の会話の 2 種類を収録し、分析を行った。これは異なる環境において、共通してみられる機能や、環境に依存した機能が見られることを期待したためである。本研究ではポスター発表型、自由環境型会話ともに 2 セッションを分析対象とした。

3.1.1 ポスター発表型会話

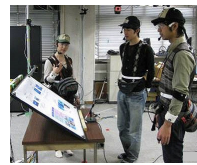
図 1 左に示すよう、非発表者 2 人に対し発表者が 1 枚のポスターを説明する会話を収録した。非発表者役の被験者は理解を深めてもらうため、分からないところは積極的に質問を行い、発表者はそれについて回答をするよう教示した。ポスター発表型会話を使用する理由は、瀬戸口ら²⁾が指摘するように、会話の役割が明確であり、会話状況が説明と質疑応答に分かれる点である。会話状況の切り替わりからうなずきと共起する振る舞いの変化が期待される。

3.1.2 自由環境型会話

自由環境型会話とは、図 1 右のように三人の被験者にポスターについて会話をしてもらうものである。参加者には「描かれた年代を推測する」などのいくつかの話題を与えた上で、博物館の展示をみるような形で自由に雑談を行うよう教示した。ポスターについては京都の古地図を 6 枚に分割し、周囲に配置した。自由環境型会話を使用する理由は、自由環境会話では発話交代が頻繁に起こるため、ポスター発表型会話と違ったクラスタが生成されることを期待したためである。

3.2 使用したセンサ機器

非言語行動を抽出するため以下のセンサを使用した。まず首振り動作を検出するため、小型無線加速度センサ WAA-001 を用いた。今回は前頭部、背中上部、腰の計 3 所に装着した。次に視線データを自動的に作成するため、モーションキャプチャと視線追跡装置を用



ポスター発表会話



自由環境会話

図 1 収録した会話環境

Fig. 1 Scenes of recorded conversations

いた。モーションキャプチャマーカを帽子に数点の他、肩と背中に装着した。視線追跡装置はゴーグル型と帽子型の 2 種類を用いた。その他にヘッドウォンタイプの音声マイクと複数の環境カメラも用いて会話状況を収録した。

3.3 アノテーション方法

首振り動作の機能分析において非言語行動のデータを使用するために、いくつかの非言語行動に対しアノテーション作業を行った。

3.3.1 首振り動作

本研究ではうなずきを、垂直に頭を上または下に動かした後に元の位置に戻る動作と定義し、前頭部に付けた三軸加速度センサによって抽出する。抽出の方法は以前筆者らが行った手法³⁾を用いる。加速度の変化量の大きい区間を抽出後、ノイズとなりうる区間を除去する。最後にポスター発表会話における発表者、自由環境会話の参加者の抽出区間は精度が悪かったため、手作業で修正を行った。

3.3.2 視線

視線情報は福岡らの方法⁷⁾をもとに、会話参加者が装着したモーションキャプチャ、および視線計測装置のデータから、自動的にラベルを生成した。まずアイマーカーレコーダと頭の座標データから視線ベクトルを計算し、その後他の人の頭部をモデル化した球体との衝突判定を行った。

3.3.3 発話および相槌

発話区間の認定、および相槌と通常発話の認定については手作業で行った。まず、無線マイクによって収録した発話音声に基づき日本語話し言葉コーパス⁴⁾を基準に発話区間の認定と発話内容の書き起こしを行った。次に、吉田らの研究⁶⁾を参考にして通常発話と相槌の分離を行った。

4. クラスタ分析

4.1 入力データ

今回入力データは多次元時系列データとし、それに対し階層的クラスタリングを行った。使用したモダリティは、以下の 5 つとした。

異なる会話環境におけるうなずきの機能分析

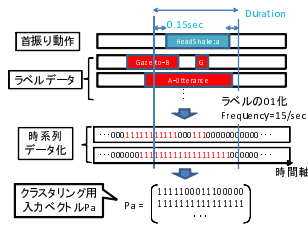


図2 入力データの作成法
Fig. 2 Method of input data

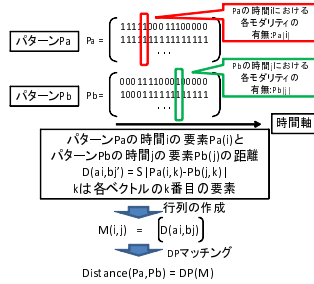


図3 時系列パターン間の距離の算出法
Fig. 3 Distance between time sequence data

- (1) 相槌を除いた発話 (On, Off)
- (2) 相槌 (On, Off)
- (3) 他の被験者 1 に対する視線 (On, Off)
- (4) 他の被験者 2 に対する視線 (On, Off)
- (5) 首振り動作 (On, Off)

各被験者の首振り動作の区間に対し、これらのモダリティを3人分、計15モダリティを入力として用いた。これらのデータを用いてクラスタリングを行う理由は、話者の視線と首振り動作は同期すること²⁾や話者と話し手の首振り動作は同期する⁵⁾など先行研究より首振り動作前後に発生するものを時系列で調べることによって機能分析が行えると期待したためである。

次に入力データの作成法について述べる。区間は各首振り動作の開始時間0.15秒前から終了時間の0.15秒後までの区間とした。入力データの作成法は図2のように1秒間を15の区間に分割し、区間内で各モダリティが発生していれば1、していなければ0となる時系列データを作成した。

4.2 クラスタリング手法

まず各パターン間の距離の算出法について述べる。図3に示すように各時系列パターン間の距離をDPマッチングにより算出した。各時系列パターンの要素は15次元の特徴量で構成されている。時系列パターン P_a の時系列 i の特徴ベクトル p_{ai} と P_b の時系列 j の特徴ベクトル p_{bj} の距離 D_{ij} をハミング距離で算出した。

$$D_{ij} = \sum_{k=1}^n |p_{ai}(k) - p_{bj}(k)| \quad (1)$$

上記式において、 n は入力とした特徴ベクトルの次元数であり、今回は15とする。次に各時間における時系列パターンの特徴ベクトルの距離 D_{ij} を要素とする行列 M を作成した。

$$M_{ab}(i, j) = D_{ij} \quad (2)$$

行列 M_{ab} の要素 $M_{ab}(i, j)$ は P_a の時間 i と P_b の時

間 j の生じているモダリティの違いの多さを意味する。この行列に対して動的計画法 (DP マッチング) を適用することで、最小コストが算出される。これは P_a と P_b の非類似度とみなせるので、これを距離とする。以上のように算出した各時系列パターン間の距離を要素とした距離行列を作成し、それをもとに階層的クラスタリングを行った。クラスタ間の距離計算はクラスタ内の平方和が最も小さくなるようにするワード法を用いた。

最後にクラスタの分割法を述べる。まずノード間の連結距離を高さの標準偏差を求めた。次にツリーのルートから高いノードの標準偏差を見ていき、標準偏差が閾値よりも高いノードを分割し、低いノードを発見した時点でノードの分割を停止するようにした。

4.3 分析結果

今回はポスター発表型会話、自由環境型会話ともに2つを対象にクラスタリングを行った。また入力データは各参加者それぞれで作成し、計12人のクラスタリング結果について分析を行った。以下では各会話環境で見られた特徴的なクラスタでの周囲の行動がどのように共起するか分析する。

4.3.1 ポスター発表型会話のクラスタ

聞き手で見られた機能として、まず話者の首振り動作に対して応答するように首振りを返すものが見られた。応答的な機能は、他の参加者の発話およびうなずきが特徴的な振る舞いとなっている⁵⁾。この中でポスター発表型会話では、質疑応答中に質問を行った人が発表者と対話的に1人で返すものが見られた（もう一人は首振り動作8回中0回）。またこのとき聞き手は発表者を注視していた（8回中8回）。次に話者と聞き手2人が同時に首振り動作を行うというクラスタが見られた。これらは協調的なリズム取りとしての機能もあると考えられる。

話し手の首振り動作を伴わない聞き手行動のクラスタで、得られたものは他の2人が対話状態になっているクラスタがある。これは自身が参与構造における傍参与者の状態になっており、話を聞いていることを示す、理解していることを示していると考えられる。周囲に起こる振る舞いでは自身に対し視線が来ないことが特徴的である。ポスター発表型会話では、他の聞き手の質疑応答中であり、話者が他の聞き手に対し視線配布が多くなされている（12回中10回）。

話し手の行う機能では、相手の応答を要求するように働きかける強調の機能⁵⁾を持つクラスタが生成された。これは自身の発話のとき、視線配布を行いながら首を振り、他の参加者の首振り動作が同期することが

特徴的となっており、内容を強調したい場合に相手の様子を確認しながら強調を行っていると考えられる。質疑応答時に、対話的に1人の聞き手に対し行うクラスタと、2人の聞き手に行うクラスタの2種類が生成された。

最後に誰も発話をしていない中、次に発話を行う参加者が間を埋めるように首振り動作を行うクラスタが生成された。これは発表者が次の発話までの準備のため、首振り動作を行い、間を埋めていることが考えられる。

4.3.2 自由環境型会話のクラスタ

聞き手で見られたクラスタではポスター発表型会話と同様の傾向が見られた。話者の首振り動作に対して応答するクラスタでは、聞き手自身は話者を注視していなかった点が異なる。また他の聞き手と同時に首振り動作を行う協調的なクラスタも見られた一方で、他の聞き手が代わりに相槌を行うクラスタも見られた(10回中6回)。

次に話し手の首振り動作を伴わない聞き手のクラスタでは、ポスター発表型会話と同様に他の聞き手と話者が対話状態になっているクラスタが見られた。話し手は他の聞き手に対し視線配布が多くなされている(16回中13回)。

話し手の行う機能では、相手に働きかける強調の機能を持つクラスタが生成された。自由環境型会話も同様に対話的に1人の聞き手が首振り動作を返すクラスタと、2人の聞き手が返すクラスタが生成された。

4.4 考 察

今回話し手、聞き手の行うクラスタにそれぞれ分け、共通して見られたクラスタや、そのクラスタにおける特徴的な振る舞いを比較した。結果として、聞き手では同じ機能でも会話状態が異なるクラスタが見られ、話し手の行う機能ではポスター発表型会話でのみ見られたクラスタがあった。また共起する振る舞いの違いとして、自由環境型会話では、視線配布があまり起きなかった点が挙げられる。これは周囲にポスターがあり、常に参照しており、他の被験者の振る舞いを認識しづらいからではないかと考えられる。またポスター発表型会話では、発表者は聞き手が理解できるように発表をするよう教示したため、聞き手の振る舞いを注視していたことが理由に挙げられる。

また自由環境型会話では話し手、聞き手が混在したクラスタが多い点が特徴的であった。これは自由環境型会話では発話のオーバーラップが多いことが理由に考えられる。さらに首振り動作が同期するクラスタが生成されない被験者もいた。このように異なる会話環

境ではうなずきを誘発する振る舞いに違いがあり、入力するデータを考える必要がある。

5. 終わりに

本論文では、うなずきという非言語行動の持つ多様な機能に注目し、聞き手話し手を問わず首振り動作として分析を行った。言語、非言語行動の発生パターンからクラスタリングを行い、生成されたクラスタの機能を調べた。

その結果、聞き手、話し手ともに共通した機能を持つクラスタが生成された。また共通するクラスタでも同期する振る舞いが違い、受け手的に頷き返すものや協調的に頷くなど会話状況の違うものが見られた。さらに間を埋めるクラスタなど会話環境による違いも現れ、同期する振る舞いによって機能が理解できることが示唆された。

今後の展望としてうなずきの機能に対しどのような振る舞いが影響しているのかを定量的に評価し、入力データを変更していく必要がある。

謝辞 本研究は、文部科学省科学研究費補助金「情報爆発時代に向けた新しいIT基盤技術の研究」の一環で実施されました。

参 考 文 献

- 1) Maynard・K・泉子：会話分析，くろしお出版(1993).
- 2) 瀬戸口久雄，高梨克也，河原達也：ポスター会話における聞き手反応のマルチモーダルな分析，人工知能学会研究会資料，SIG-SLUD-A703，pp. 65-70 (2008).
- 3) 齊賀弘泰，角 康之，西田豊明：多人数会話におけるうなずきの会話制御機能の分析，情報処理学会研究報告(ユビキタスコンピューティングシステム)，No.2010-UBI-26 (2010).
- 4) 独立行政法人国立国語研究所：日本語話し言葉コーパス.
- 5) 前田真季子，堀内靖雄，市川 嘉：自然対話におけるジェスチャーの相互的關係の分析，情報処理学会研究報告，HI，ヒューマンインタフェース研究会報告，Vol.2003，pp.39-46 (2002-HI-102).
- 6) 吉田奈央，高梨克也，伝 康晴：対話におけるあいづち表現の認定とその問題点について，言語処理学会第15回年次大会発表論文集，pp.430-433 (2009).
- 7) 福岡良平，角 康之，西田豊明：人のインタラクシオ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(ユビキタスコンピューティングシステム)，No.2009-UBI-23 (2009).