

多人数会話におけるうなずきの会話制御としての機能分析

齊 賀 弘 泰^{†1} 角 康 之^{†1} 西 田 豊 明^{†1}

会話において、うなずきは相手の意見に対する同意など、話し手に対するフィードバックとして重要な機能を持っているとされる。また話し手が肯定や強調として用いたり、発話の前振りを行うなど会話の流れを制御しているという研究もある。そこで本論文では、頭部に装着した加速度センサのデータを用いて、うなずきの候補となりうる縦方向の首振り動作の自動抽出を試みる。その際、聞き手だけでなく話し手についても首振り動作の抽出対象とした、次に抽出された首振り動作に対して会話制御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結果を検証した。結果として話し手の強調とそれに対する聞き手の応答的な機能、他にも続けてよいという意味表示の機能といったクラスが得られた。

Function analysis of nodding for conversation adjustment in multi-party conversation

HIROYASU SAIGA,^{†1} YASUYUKI SUMI^{†1}
and TOYOAKI NISHIDA^{†1}

In face-to-face communication, nodding has important functions of visual feedback to speaker. Additionally, in some studies, speakers nod to agree or emphasize their important part, and to show that they start to speak in order to adjust conversation. In our thesis, first, we define nodding as head shaking that both speakers and hearers use. Second, we detect head shaking automatically from three-party conversation about poster presentation. Finally we analyze the functions of conversation adjustment using clustering against patterns of verbal and non-verbal behavior. We acquire clusters where listeners use head shaking to show in order to allow others to speak, react to speakers' behavior, and where speakers use to emphasize.

^{†1} 京都大学
Kyoto University

1. 序 論

人は会話において、言語以外にも身振りや視線移動と言った非言語的な情報で意図を伝えたり、会話のリズムを調整している。その中でもうなずきは相手に対するフィードバックとして非常に重要な機能を持っており、擬人化エージェントとの自然な対話¹⁾ や会議録のインデックス作成²⁾ などさまざまな用途で利用しようと自動検出が研究されている。また、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があり³⁾、会話の流れを制御しているものと考えられる。

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、うなずきの会話制御の機能を自動的に分類する手法を提案する。そのためまず多人数会話を収録し、センサデータより自動抽出を行った。次に抽出された首振り動作に対して機能分類を行うため、発話、相槌、首振り動作の発生パターンによって、クラスタリングを行った。各クラスにおける機能を分析し、会話制御にどのような入力データが特徴となるのかを検証した。

2. 多人数会話におけるうなずき

2.1 うなずきに関する先行研究

2.1.1 会話分析におけるうなずきの機能

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一つとして考えられていた。Duncan らの研究⁴⁾ ではうなずきを視覚的な相槌と定義し、非言語行動としての相槌の機能を持つとした。この研究では相槌とは発話権をとる意思がないものであり、うなずきには発話意図を示す機能はないとされていた。

これらの研究に対して、話し手の行ううなずきに注目し、うなずきの会話制御機能を調べた研究として Maynard³⁾ があげられる。Maynard は話し手の行ううなずきも含め、うなずきを9つの機能に分けた。この中には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、リズム取りといった機能、さらに間発話中に発話意図を示す機能があった。これらの機能は Duncan ら⁴⁾ がいう視覚的な相槌だけでは説明できないものがあることが分かる。

また話し手のうなずきに対する聞き手の反応についても研究されている。前田ら¹⁰⁾ は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

し応答的に頷き返すとした。

2.1.2 うなずきの自動検出の先行研究

うなずきの自動検出には大きく分けてモーションキャプチャやヘッドセットのようなセンサ機器を使用して頭部の動きを検出するものと、画像認識によって頭部の動きを抽出するものの二種類の方法がある。

まずセンサ機器を使用した自動検出として山本ら²⁾があげられる。この研究では会話、会議での重要シーン抽出を行うことを目的とし、うなずきを興味度の指標として抽出を行った。これに対し画像認識によるうなずきの自動検出として Morency ら¹⁾がある。この研究の主な目的はロボットなどの機械に非言語行動を認識させ、文脈を理解し、より人と円滑なコミュニケーションを行うことである。そのため相手の話への興味度や相手の質問に対する同意をうなずきの機能としている。

2.2 うなずきの機能と本研究の目的

上記よりうなずきには複数の機能があり、聞き手のうなずきに注目し工学分野でも自動抽出が研究されている。しかし、うなずきには聞き手が行うあいづちとしての機能だけでなく、Maynard のような話し手の行う行為、さらには間発話中において発話意図を示すなど会話制御としての機能もある。そこで本論文ではうなずきの機能を Maynard³⁾ のカテゴリを参考に分類した。

Maynard はうなずきを会話管理としての機能に注目し、どの時点でうなずいているかでカテゴリ分けを行っていたため、聞き手のうなずきについては一つのカテゴリにまとめられていた。そのため本論文では聞き手のうなずきを発話継続を促す意思表示、応答的行為の機能の二つに分類した。発話継続を促す意思表示の機能は自身の発話意図がないことを示す機能であり、会話の流れを制御していると思われる。応答的行為とは話し手の働きかけの行為に対し反応を返す機能であり、話し手の振る舞いに対するリアクションという点で続けてよいという機能と分けた。

このような分類を行うため、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」と定義し自動抽出を行う。次に首振り動作を機能の分類を行う。そのため、自動抽出した首振り動作に対し言語、非言語行動の発生パターンよりクラスタリングを行う。そして各クラスタの機能を分析し、適切な入力データについて検証を行う。

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 のような三人によるポスター発表を題材とした。ポスター発表を題材とした理由は話し手と聞き手が明確である点と移動が少ない点の二点である。瀬戸口ら⁵⁾によるとポスター発表では、被験者の役割が発表者と非発表者に明確に分かれていることがいわれている。さらに会話の構造も発表者が説明を行う状態と質疑応答を行う状態の二つが明確に分かれているといわれている（以下説明モードと質疑応答モードとする）。説明モードでは発表者が発話をしているのに対し、質疑応答モードでは非発表者が質問し、それに発表者が答えるという通常の対話に近いものとなる。この二つのモードでは、視線の動きなど非言語行動に違いがあることも瀬戸口らによって指摘されており、首振り動作の機能分析において違いが出ることが予想される。また、移動型会話に比べ立ち位置があまり変わらないため、動くことによる誤検出が起こりにくいという点も、自動抽出の精度を向上させることができると考えられる。

次に会話設定の説明を行う。会話状況は、発表者役の被験者が非発表者役の被験者二人に対して、作成したポスターをもとに自身の研究内容を発表するというものである。非発表者役には発表者の研究分野に詳しくない人もいるため、発表者には非発表者役の被験者が理解しやすいような内容で発表するよう教示を行った。また、非発表者役の被験者は理解を深めようするため、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。ポスター発表の時間は 15 分を目安としてもらい、20 分と 25 分を過ぎた時点で紙により合図を行い、30 分を過ぎた時点で発表が途中で終了することとした。結果 30 分を過ぎた会話はなかった。

今回 8 つの会話を収録し、モーションキャプチャや視線追跡装置などデータの欠損の少ない二つの会話を対象に自動抽出を行った。

3.2 使用したセンサ機器

首振り動作の検出を行うため、小型無線加速度センサを用いた。これは図 2 のように X、Y、Z 三軸の加速度を無線で PC に送信するもので、今回は前頭部、背中上部、腰の計 3 か所に装着した。

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いた。モーションキャプチャは頭部の位置を推定するため、頭部にマーカを装着した帽子を被験者に被ってもらった。その他に肩と背中にも装着した。

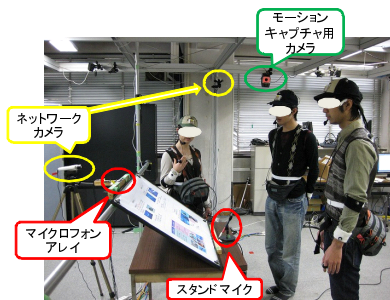


図 1 ポスター発表会話の様子
Fig. 1 Scene of presentation

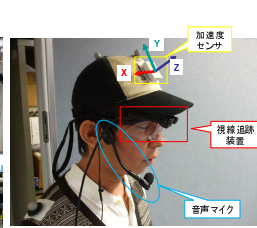


図 2 各種センサの装着位置
:頭部
Fig. 2 Sensors position
around head

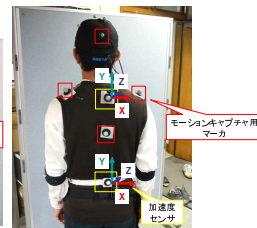


図 3 各種センサの装着位置
:背中
Fig. 3 Sensors position on
back

また会話状況を記録するため音声マイクとネットワークカメラも用いた。

3.3 アノテーション作業

抽出した首振り動作の検証と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

自動抽出した首振り動作の精度検証のため、2.2 節で述べた首振り動作に対し手作業でアノテーションを行った。首振り動作の定義は 4.1 節で述べる。アノテーション区間は各セッションの開始 5 分から 10 分までの 5 分間とした。また連続した首振り動作は一つの首振り動作区間とみなし、アノテーションを行った。アノテーション環境としては iCorpusStudio⁶⁾ を使用した。iCorpusStudio は図 4 に示すように複数の動画、音声を同時に再生しながら、アノテーションを行えるソフトである。今回は画像、音声に加えて、頭部の加速度の波形の三つを参照しながら行った。

3.3.2 視線

視線データの作成法としては視線データのモーションキャプチャデータの三次元座標への変換と、視線ベクトルの衝突判定の二段階に分けられる。視線追跡装置データの三次元座標への変換は福間らの研究⁷⁾ の手法を用いて、各被験者につけた視線追跡装置とモーションキャプチャシステムのデータを利用して行った。実験前に取得したキャリブレーションデータをもとに視線を三次元座標に変換する変換行列を作成した。次にモーションキャプチャより被験者の頭部および右目の三次元座標を作成し、右目を始点とした三次元の視線ベクトルへ変換した。このようにして三次元座標化した視線ベクトルと各被験者の頭部、ポスターと

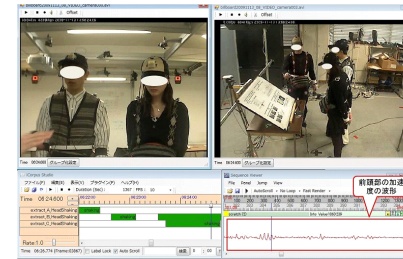


図 4 アノテーション作業環境
:iCorpusStudio の画像例
Fig. 4 Annotation
Environment:iCorpusStudio

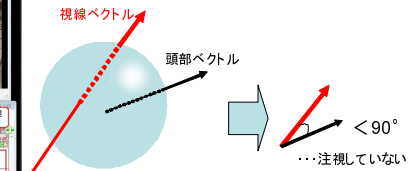


図 5 後ろから見た場合の誤認識除去
Fig. 5 Error nooise cut

の衝突判定を以下の条件で行い、視線ラベルを作成した。

- (1) 人の頭部を仮想的な球体とみなし、視線ベクトルとの交点が球体の場合頭部を注視しているとする。
- (2) 視線ベクトルと、モーションキャプチャの頭部から取得した注視対象の頭部方向ベクトルの角度が 90 度よりも小さい場合は、人を注視しているとはみなさない(図 5)
- (3) 2 人が一列に並ぶなどで 2 人の頭部と交差した場合は、視線ベクトルの始点に近いほうを注視しているとする

3.3.3 発話および相槌

厳密な発話区間および相槌の認定は自動で行うのは難しいため、手作業で行った。

本研究における発話ラベルは、無線マイクによって収録した発話音声をもとに日本語話し言葉コーパス (CSJ)⁸⁾ の基準に準拠したタグ付き書き起こしを作成し、そこで認定された発話区間を用いた。

次に相槌と通常発話の分離のため、相槌部分の認定を行い、作成した書き起こしを基に通常発話と相槌の分離を行った。吉田らの研究⁹⁾ を参考に相槌を認定し、発話から相槌を分離した。

4. うなずき動作の自動検出

4.1 抽出対象とする頭部動作

今回 Maynard³⁾ を参考に、以下の動作を首振り動作と定義し抽出対象とした。

- (1) 垂直に頭を上、または下に動かしたのち元の位置に戻る動作

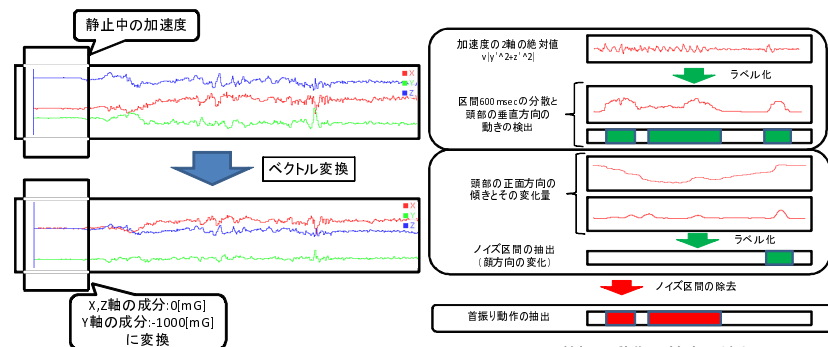


図 6 ベクトル変換の概要

Fig. 6 Vector converting image

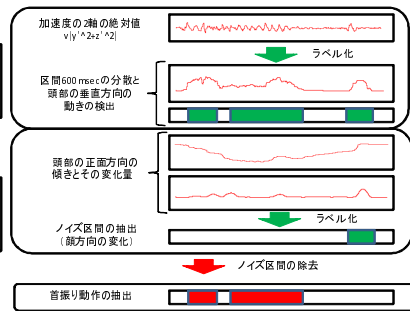


図 7 首振り動作の抽出の流れ

Fig. 7 Flow of extracting headshake

- (2) 顔を上げただけ、下げただけのような顔方向を変えたものは対象としない
- (3) 顔を左右に振るものも対象としない
- (4) 動かしてから元に戻るまでの時間、動かす速さについて考慮しない

聞き手行動のうずきだけでなく話し手行動も含めた動作として自動抽出を行った。

4.2 加速度データの前処理

使用する加速度センサは装着した際に傾きが人によって異なってしまう。そこで図 6 のように取得した 3 軸加速度情報に対して、静止中において加速度センサの Y 軸の負の方向が重力加速度と一致するよう、静止中の加速度情報から作成した回転行列をかけて正規化を行った。

次に加速度センサの傾きを出した。変換後の加速度データから、直立時からの頭部の傾きを検出できる。変換後の座標軸を X', Y', Z' 軸とおくと、直立していないときに Y' 軸以外でも加速度が検出されるためである。傾きの方向は加速度センサの装着が図 2 より、X' 軸と水平面との角度を被験者の左右方向の傾き、Z' 軸と水平面との角度を被験者の正面方向の傾きとした。

4.3 首振り動作の抽出法

本研究では図 7 に示す通り、頭部の Y'-Z' 平面における動作を抽出後、単に顔の向きを変えただけのような首振り動作ではない動作を除去する手法をとった。以下ではその抽出法を説明する。

4.3.1 頭部の Y'-Z' 平面における動作区間の抽出

頭部の縦振りを認識するため前頭部の Y'-Z' 平面に動く動作を前頭部に付けた三軸加速度センサによって抽出する。まず特徴量として、加速度センサの図 2 の Y' 軸方向と Z' 軸方向の加速度の絶対値 $\sqrt{y'^2 + z'^2}$ を使った。y', z' は Y' 軸, Z' 軸の加速度の値である。2 軸の値のみを使用した理由は正規化したため、垂直方向の動作に対して X' 軸方向に加速がわからないためである。

次にこの絶対値のデータに対し 600 ミリ秒の区間で分散をとり、分散の時系列データを作成した。この時系列データの中央値 m と標準偏差 s を求め、 $T = m + f * s$ となるよう閾値 T を定めた。T より値の大きい区間を垂直方向に動いた区間として抽出した。なお、 f は全被験者での再現率と適合率の積の平均が最大となる値とした。

4.3.2 他の身体動作の抽出

前節で頭部が Y'-Z' 平面における動作区間を推定したが、これだけでは首振り行為以外の動きも多く取ってしまうため、別途に誤検出しやすい動作をノイズ源として抽出、除去を行った。また、一回の首振り動作の最小時間を 600 ミリ秒と仮定し、ノイズ区間を除いた区間で 600 ミリ秒に満たない区間は誤検出として除去した。

(1) 顔方向を変える動作

顔方向を変える動作とは、ポスターから他の被験者に顔を向けるなど元の位置に戻らない動作である。図 8 で示すように、この動作のうち前頭部の Y'-Z' 平面に加速のかかるものはノイズ源となるため抽出を行った。元の位置に戻るかを顔の傾きより判定するため、今回は前頭部の加速度センサの傾きを使用した。まず被験者の正面方向の傾きに対してメディアンフィルタをかけ、傾きの微振動する区間を除去することとした。これによって首振り動作の区間を誤検出することを防ぐ。

平滑化された傾きに対して前頭部の Y'-Z' 平面での動きの抽出と同じく、区間 600 ミリ秒の分散が大きい区間を顔方向の変化した区間として抽出をした。またこの閾値は閾値を T 、セッション全体での傾きの平均値を a 、標準偏差を s としたとき、 $T = a + s$ とした。

(2) 上体を傾ける動作

上体を傾ける動作とはポスターに対して上半身を曲げながら注視するなど下半身はあまり動かない上体の動作である。この動作は図 9 で示す通り同時に顔も Y'-Z' 平面で加速がつくため、ノイズとなりやすい。この動作の抽出には背中上部に装着した加速度センサの傾きを用いた。また傾きは被験者の正面方向および左右方向の二方向を使

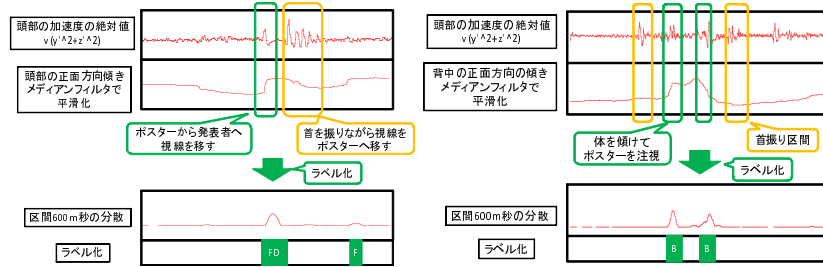


図 8 ノイズ波形:顔方向の変化
 Fig. 8 Changing head direction

図 9 ノイズ波形:上体を傾ける動作
 Fig. 9 Bowing

(3) 体全体の動作

立ち位置を変えるなど下半身を含めた全身が動く動作を体全体の動作と定義した。この動作の抽出には腰につけた加速度センサを使用した。まず加速度センサの3軸の絶対値 $\sqrt{x'^2 + y'^2 + z'^2}$ を計算した。これに対し首振り動作による振動を消すため同様にメディアンフィルタをかけた。また体全体の動作の抽出は平滑化した加速度データに対して区間 600 ミリ秒を窓幅として分散をとり分散値が大きい区間とした。閾値は顔方向を変える動作と同様とした。

4.3.3 周波数解析による連続した首振り動作の抽出

上記のように、簡単にメディアンフィルタによって平滑化したデータから誤検出しやすい動作を検出する場合、連続した大きな首振り動作は平滑化しきれずノイズ区間として誤検出されてしまう恐れがある。そのため、連続した大きな首振り動作区間を周波数解析を行うことで抽出した(図 10)。またノイズ区間を除去した首振り動作区間と OR をとることで、首振り動作区間を認定した。

次に具体的な方法について述べる。まず前頭部の加速度センサの X,Y の 2 軸の絶対値をとったものに対し、窓幅を 128 データとして自己相関をとった。次に自己相関をとったものに対し、窓関数をかけ高速フーリエ変換により各周波数成分におけるパワーを調べた。なお今回窓関数には Hamming 窓 $w(x) = 0.54 - 0.46\cos(2\pi x), 0 \leq x \leq 1$ を用いた。

また人間の頭部は一定の周期で常に動いてはいないため、1Hz 以下の低周波領域にパワーが集中する。そのため今回は 1Hz 以上の周波数成分におけるパワーのみを参照し、各時間

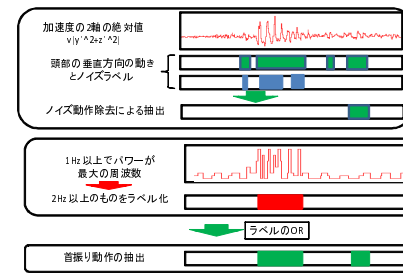


図 10 周波数解析による首振り動作の抽出
 Fig. 10 Detection of head shaking by frequency analyzing

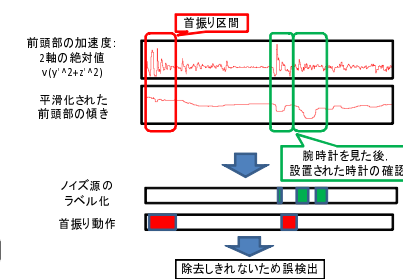


図 11 除去失敗による誤検出
 Fig. 11 Error:removing

表 1 首振り動作の自動抽出の結果 (%)

Table 1 Result for automatic detection of head shaking

| | 発表者 A | | 非発表者 B | | 非発表者 C | |
|----------|-------|------|--------|------|--------|------|
| | 再現率 | 適合率 | 再現率 | 適合率 | 再現率 | 適合率 |
| sessionA | 83.6 | 65.0 | 83.0 | 79.7 | 81.3 | 81.9 |
| sessionB | 59.1 | 37.5 | 61.4 | 95.3 | 71.4 | 69.8 |

における最大パワーをもつ周波数を取得した。連続したうなずきは高周波領域にあるため、周波数をデータとした時系列データに対し 2.5Hz 以上の区間の抽出を行った。次に 2.5Hz 以上の周波数を持つ時間において、前後 200 ミリ秒の区間を連続した領域として抽出した。

4.4 結果と考察

この節では自動検出の結果と考察を述べる。まず結果について、表 1 に再現率と適合率を示す。また、被験者 A, B, C は図 1 に示したように発表者を A, 発表者の近くに立っている非発表者を B, 遠くに立っている非発表者を C とした。また今回立ち位置は固定するよう指示はしなかったが、立ち位置が変化することはなかった。

今回精度を評価をするため、3.3.1 節で述べた環境を用いて手作業で正解データを作成した。また対象とする動作は 4.1 節で述べた首振り動作とした。sessionB の発表者 A で特に再現率適合率が低い結果となってしまった。これは A は頻繁に顔方向を変えたり、首を横に振る動作など誤検出しやすい動作を多く行い、それらを除去できていないことが理由だと考えられる。

次に実際に起こった未検出、誤検出の例を参照しながら、検出法の改善点を考察する。まずノイズ源となる動作の除去に失敗したために誤検出を起こした例を図 11 に示す。これは時計を見るために顔を一度下げたあと、すぐに上げた動作を誤検出してしまったものであ

る．このような動作に対しメディアンフィルタで平滑化を行うと，顔の傾きの変化量が少なくなってしまう，ノイズの区間に対して一部しか除去ができなくなってしまう．またこのような動作は顔を元の位置に戻すまでの時間が短く，首振り動作に近い動作となっており，除去は大変難しいと考えられる．そのほかの誤検出としては，体の重心を変える動作や体勢を変える動作があげられる．このような動作は頭部や背中の傾きが変わらないためノイズ源として抽出が行えなかった．

以上より，個別にノイズ源となる動作抽出法において，データを平滑化し分散をとる方法では，首振り動作を誤検出する可能性がある．そのためいくつか傾きや加速度データを複数個入力データとして，どの動作が起こったかを判別する手法を用いるほうがよいと考えられる．

5. うなずき動作の機能分析

この章では発話交替に加え，視線，他者の首振り動作を加えた言語，非言語の行動の発生パターンよりクラスタリングを行い，首振り動作の機能分析を行った．

5.1 入力データの作成法

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，

- (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)

とした．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三人分，計 15 モダリティを入力として用いた．各動作は 3.3 節で認定されたものを使用した．また今回は比較的精度が高く，加速度センサーデータの欠落がなかったセッション A のデータを対象とした．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⁵⁾ や話者と話し手の首振り動作は同期する¹⁰⁾ 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で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 12 のように 1 秒間を 15 の区間に分割し，各区間でのモダリティ発生の有無を見る．各モダリティが発生していれば

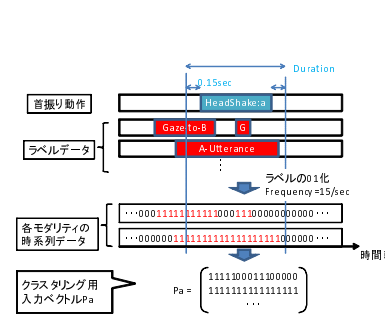


図 12 入力データの作成法
Fig. 12 Making input data

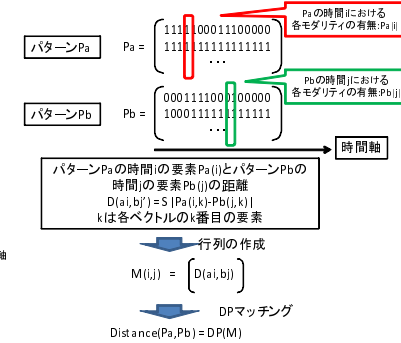


図 13 時系列パターン間の距離の算出法
Fig. 13 Distance between time sequence data

1, していなければ 0 となる時系列データを作成した．

5.2 クラスタリングの手法

まず各パターン間の距離の算出法について述べる．図 13 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，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．

$$Distance_{ab} = DP(M_{ab}) \quad (3)$$

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としたクラスタツリー

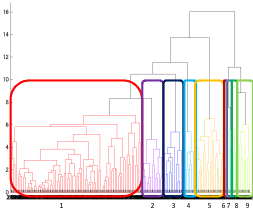


図 14 発表者 A
:樹状図とクラス対応
Fig. 14 Dendrogram:A

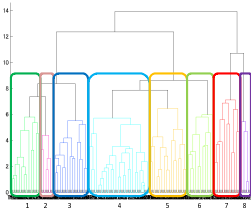


図 15 非発表者 B
:樹状図とクラス対応
Fig. 15 Dendrogram:B

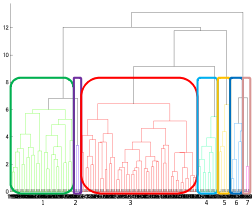


図 16 非発表者 C
:樹状図とクラス対応
Fig. 16 Dendrogram:C

表 2 発表者 A:各クラスのモダリティの生起数
Table 2 Modalities occurrence:Presenter A

| クラス | 総数 | A 相槌 | B 発話 | C 発話 | 視線 A | B | 視線 A | C | 視線 B | A | 視線 C | A | Bnod | Cnod |
|-----|-----|------|------|------|------|----|------|----|------|----|------|---|------|------|
| 1 | 153 | 8 | 18 | 8 | 41 | 16 | 18 | 36 | 57 | 48 | | | | |
| 2 | 26 | 1 | 2 | 0 | 2 | 3 | 1 | 1 | 8 | 19 | | | | |
| 3 | 22 | 1 | 2 | 1 | 8 | 8 | 1 | 19 | 11 | 11 | | | | |
| 4 | 14 | 8 | 0 | 11 | 3 | 6 | 0 | 8 | 2 | 9 | | | | |
| 5 | 36 | 1 | 4 | 0 | 7 | 8 | 3 | 10 | 34 | 24 | | | | |
| 6 | 4 | 1 | 1 | 0 | 1 | 0 | 3 | 3 | 1 | 2 | | | | |
| 7 | 6 | 1 | 3 | 0 | 4 | 2 | 6 | 5 | 5 | 1 | | | | |
| 8 | 6 | 3 | 5 | 0 | 6 | 2 | 1 | 5 | 3 | 3 | | | | |
| 9 | 14 | 6 | 11 | 0 | 6 | 2 | 9 | 2 | 6 | 2 | | | | |

を生成した．次に連結されたノードの高さの標準偏差を求めた．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．分割されなかったノードの下にいる葉はすべて同じクラスとみなす．

5.3 結果と考察

実際に分割されたクラスに対してどのような機能を持つかをビデオと音声とを参照しながら検証した．第 4 章と同じく発表者を A，発表者に近い非発表者を B，遠い非発表者を C とする．

5.3.1 発表者 A の結果

まず発表者 A の首振り動作について述べる．図 14 はクラスタリング結果の樹状図であり，表 2 は各首振り動作中に発生した各動作の表である．発表者の聞き手として行うクラスは 4, 8, 9 となった．クラス 8, 9 はほとんどのものが B の質問中に発話継続を促す意思表示の機能として使われた．クラス 8, 9 の違いはクラス 9 が A, B が相互注視の状態であるのに対し，クラス 8 では B はポスターを注視している点である．また一部質問

表 3 非発表者 B:各クラスのモダリティの生起数
Table 3 Modalities occurrence:Listener B

| クラス | 総数 | A 相槌 | B 相槌 | B 発話 | 視線 A | B | 視線 B | pos | 視線 B | A | Anod | Cnod |
|-----|----|------|------|------|------|----|------|-----|------|---|------|------|
| 1 | 23 | 3 | 3 | 6 | 11 | 20 | 7 | 9 | 7 | | | |
| 2 | 10 | 0 | 2 | 9 | 0 | 9 | 1 | 5 | 2 | | | |
| 3 | 30 | 3 | 18 | 7 | 8 | 29 | 3 | 24 | 20 | | | |
| 4 | 50 | 0 | 7 | 4 | 15 | 50 | 7 | 35 | 16 | | | |
| 5 | 32 | 0 | 8 | 2 | 9 | 31 | 1 | 16 | 24 | | | |
| 6 | 22 | 2 | 7 | 3 | 4 | 21 | 3 | 14 | 10 | | | |
| 7 | 22 | 3 | 11 | 9 | 19 | 17 | 5 | 11 | 11 | | | |
| 8 | 8 | 1 | 2 | 2 | 3 | 2 | 8 | 5 | 0 | | | |

表 4 非発表者 C:各クラスのモダリティの生起数
Table 4 Modalities occurrence:Listener C

| クラス | 総数 | A 相槌 | C 相槌 | C 発話 | 視線 A | C | 視線 C | A | 視線 C | pos | Anod | Bnod |
|-----|----|------|------|------|------|----|------|----|------|-----|------|------|
| 1 | 51 | 1 | 10 | 3 | 5 | 10 | 50 | 34 | 36 | | | |
| 2 | 4 | 0 | 0 | 0 | 4 | 2 | 2 | 1 | 3 | | | |
| 3 | 95 | 1 | 11 | 4 | 13 | 24 | 87 | 54 | 34 | | | |
| 4 | 18 | 0 | 2 | 1 | 8 | 14 | 6 | 13 | 8 | | | |
| 5 | 9 | 2 | 2 | 0 | 4 | 3 | 9 | 4 | 6 | | | |
| 6 | 12 | 6 | 0 | 11 | 6 | 7 | 8 | 7 | 1 | | | |
| 7 | 4 | 0 | 0 | 4 | 1 | 0 | 4 | 0 | 0 | | | |

に対する肯定を示す首振り動作も含まれていた．クラス 4 は C の質問中に行われており機能は同じである．それに対しクラス 7 は質問に対し応答的な反応を行った後，話し始めるクラスである．クラス 7 は話者の行うものと聞き手が行うものが混在したクラスとなった．まず B の質問に対し相槌と同時にいった後，そのまま回答を行う際に首振り動作を行っている．これは発話意図を示すための首振り動作とみなせる．

残りのクラスは話し手の行うクラスとなった．クラス 3, 5 はデータ上では視線配布は少ないが，B, C に対して視線配布を頻繁に行っており，B, C も首振り動作を行い反応している．これは前田ら¹⁰⁾のいう働きかけの機能を持った首振り動作であり，強制的な機能を持っていると考えられる．クラス 6 は同様に強制的な機能を持つが 3, 5 に対して，非発表者 B の質問区間であるため，B に対して視線を送り働きかけているため，B のみがそれに応答しているという点で異なる．クラス 1, 2 は拍子取りや強調といった機能が混在したクラスとなった．

5.3.2 非発表者の結果

先ほどと同様に樹状図を図 15, 16, 発生した動作を表 3, 4 に示す．まず非発表者 B の結果について述べる．話し手の行うクラスとなったのはクラス 2 となった．2 は質問を行っている際に発生しており，強調を示すクラスである．クラス 7 は話し手の行うクラスと聞き手の行うクラスが混在したものとなった．話し手の行うクラスはクラス 2 と同様の機能を持つが，聞き手の行うクラスは発話継続を促す意思表示の機能であった．

残りのクラスタはすべて聞き手の行うクラスタとなった。まず、クラスタ 3, 4 は発表者 A の首振り動作と同期した応答的な行為の機能を持つクラスタとなった。またクラスタ 4 は 3 に対し、C の首振り動作の同期も多く、協調的なリズム取りの意味合いを持っていると考えられる。クラスタ 1, 5, 6, 8 は発話継続を促す意思表示の機能のクラスタだと考えられる。この中でクラスタ 8 は自身の質問に対する回答での首振り動作であり、興味度が高い状態であると考えられる。クラスタ 5 は非発表者 C との同期が多く協調的なリズム取りとなっている。

最後に非発表者 C の結果について述べる。まず話し手の行うクラスタはクラスタ 6, 7 となった。これらのクラスタは強調の機能を持つと考えられる。2 つのクラスタの違いはクラスタ 7 では発表者 A の相槌や首振り動作といった聞き手反応が少ない点にある。またこれらのクラスタは非発表者 B のクラスタ 2 と対応すると考えられる。

残りのクラスタはすべて聞き手の行うクラスタとなった。クラスタ 1, 4 は発表者 A の首振り動作に対する応答行為の機能である。クラスタ 1 は非発表者 B の首振り動作の同期が多く、協調的なリズム取りの意味合いがある点で 4 と異なる。またクラスタ 1 は非発表者 B のクラスタ 4, クラスタ 4 は非発表者 B のクラスタ 3 とそれぞれ対応していると考えられる。次にクラスタ 2, 3, 5 は発話継続を促す意思表示の機能を持つと考えられる。この中でクラスタ 5 は非発表者 B の質問、回答の区間であり、C はポスターを注視し、興味度が低い状態だと考えられる。

6. 結 論

本論文では、うなずきという非言語行動の持つ多様な会話制御機能に注目し、加速度センサによって話し手聞き手問わず首振り動作と定義して自動抽出を行った。今回誤検出を起こしやすい動作をノイズ源として除去を行うことで、頭部の加速度情報では除去できない誤検出を防ぐことができるようになった。しかし、いまだに誤検出が多い被験者もあり抽出法に改善が必要であることが分かった。

次に自動抽出された首振り動作に対して言語、非言語行動の発生パターンからクラスタリングを用いて機能分類を行った。その結果、実際に会話制御にかかわる、話し手の行う相手に対する強調とそれに対する聞き手の応答や続けてよいという意思表示の機能が分類できた。しかし複数の機能が混在した大きいクラスタも生成されており、入力データに対して改善が必要であると考えられる。

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) Morency, L., de Kok, I. and Gratch, J.: Context-based recognition during human interactions: automatic feature selection and encoding dictionary, *Proceedings of the 10th international conference on Multimodal interfaces*, ACM New York, NY, USA, pp.181-188 (2008).
- 2) 山本 剛, 坂根 裕, 竹林洋一: マルチモーダルヘッドセットを用いたうなずき検出と会話の重要箇所把握, 情報処理学会ヒューマンインタフェース研究会, Vol.2003, No.94, pp.13-19 (2003).
- 3) Maynard・K・泉子: 会話分析, くろしお出版 (1993).
- 4) Duncan, S. and Fiske, D.: *Face-to-face interaction: Research, methods, and theory*, Halsted Press (1977).
- 5) 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).
- 6) 来嶋宏幸, 坊農真弓, 角康之, 西田豊明: マルチモーダルインタラクション分析のためのコーパス環境構築, 情報処理学会研究報告. HCI, ヒューマンコンピュータインタラクション研究会報告, Vol.2007, No.99, pp.63-70 (2007).
- 7) 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 (ユビキタスコンピューティングシステム), No.2009-UBI-23 (2009).
- 8) 独立行政法人国立国語研究所: 日本語話し言葉コーパス.
- 9) 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).
- 10) 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関係の分析, 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).