F-05

多人数会話におけるうなずきの会話制御機能の分類

Function Analysis of Nodding for Control of Turn-Taking in Multi-Party Conversation

齊賀　弘泰†　　角　康之†　　西田　豊明†
Hiroyasu Saiga　Yasuyuki Sumi　Toyoaki Nishida

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いる．この機能に注目し工学分野でも，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．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[3]があり，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．各クラスタ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. 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取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着目し，うなずきの会話制御機能を調べた研究として Maynard[3]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，さらに誰も発話していない間発話中に発話意図を示すために行うものもあるとした．これらの機能は Duncan ら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．

このようにうなずきには複数の機能があり，聞き手が行うあいづちとしての機能だけでなく，Maynard[3] のいう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ことが考えられる．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，Maynardのカテゴリを参考にうなずきの機能を分類することを考える．また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．そのため研究の目的として，周辺に発生する言語，非言語行動を計測し，それらの共起関係からうなずきの機能を分類することを試みる．

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．

自動抽出はモーションキャプチャや加速度センサを用いた抽出法[2][7]と画像認識を用いた抽出法[1]の二種類がある．今回簡易なセンサである加速度センサのデータを用いて抽出を行う．これは立ちながらの会話において人間の動きをカメラの範囲に固定せず，自由に動ける会話を対象とするためである．また画像認識を用いた抽出はロボットが自身の視点から人間の動きをセンシングするという目的があるため，本研究と目的が異なるため画像認識による抽出は行わない．

分類するうなずきの機能は 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．しかし前田ら[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われている．同様の研究は英語圏にも存在しており McClave[6]が同様の考察をしている．そのため本論文では聞き手のうなずきを応答的行為と発話継続を促す意思表示の機能の二つに分類した．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る．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．この二つは話し手の振る舞いに対するリアクションか，自発的に出たものかということから会話制御にも関連していることが考えられる．

このような機能分類を行うため，今回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を入力データとした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，各クラスタにおける特徴的なモダリティを調べ，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[8]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

†京都大学大学院情報学研究科，Graduation School of Informatics,Kyoto University

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，移動型会話にくら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ため，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．

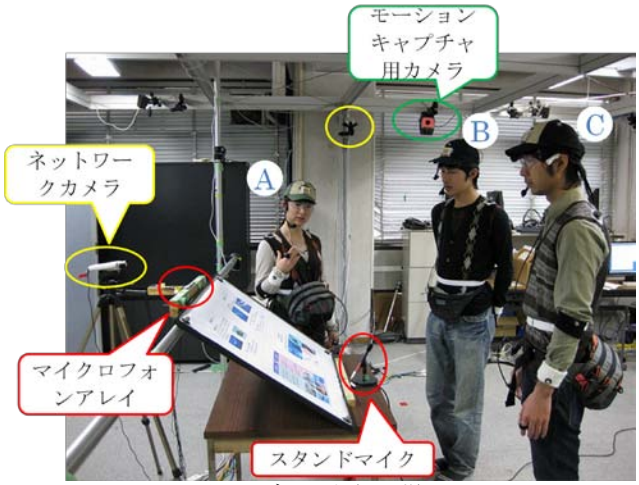

図1:データ収録の様子

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分が過ぎた時点で発表が途中でも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WAA-001を用いた．加速度センサを用いる理由は立ち会話において被験者をカメラの範囲に固定しなくても首振り動作を抽出できるようにするためである．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所に装着した．頭部以外に装着した理由は首振り動作をより精度よく抽出するためである．今回サンプリングレートは40Hzで収録を行った．

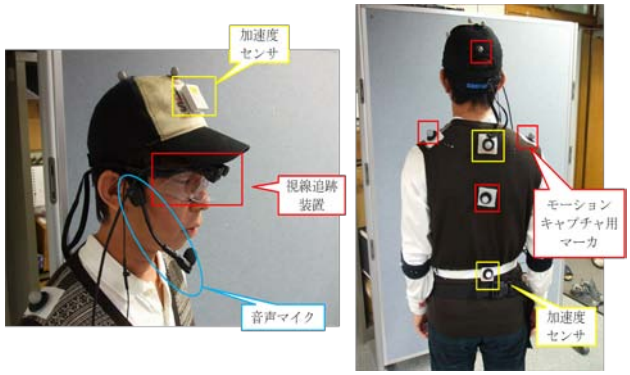

図2:各種センサの装着の様子

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．視線追跡装置はゴーグル型と帽子型の二種類を用い，得られた2次元の注視点座標を3.3.1節で述べる方法で3次元座標に変換することで，視線の自動アノテーションを行った．

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．音声マイクはヘッドセットタイプであり，各個人に装着した．ネットワークカメラは四方と非発表者の頭部の動きを撮影するためにポスターの後ろに配置した．

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　視線

自動抽出した首振り動作の精度検証のため，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iCorpusStudio[9]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．

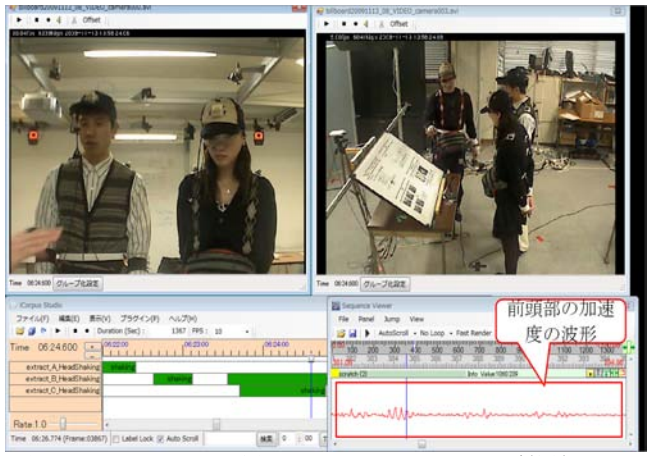

図3:アノテーション用ソフト iCorpusStudioの利用例

3.3.2　首振り動作

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[10]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
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　発話と相槌

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基に日本語話し言葉コーパス(CSJ)[11]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[12]を参考に相槌を認定し，発話から相槌を分離した．

4.　うなずきの自動抽出

4.1　抽出対象とする頭部動作

今回Maynard[3]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
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

聞き手行動のうな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4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．

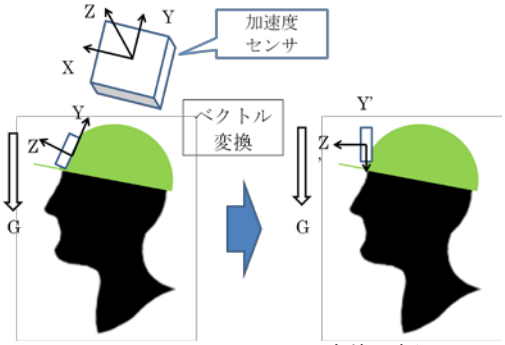

図4：ベクトル変換の概要

4.3　首振り動作の抽出法

本研究では図5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．

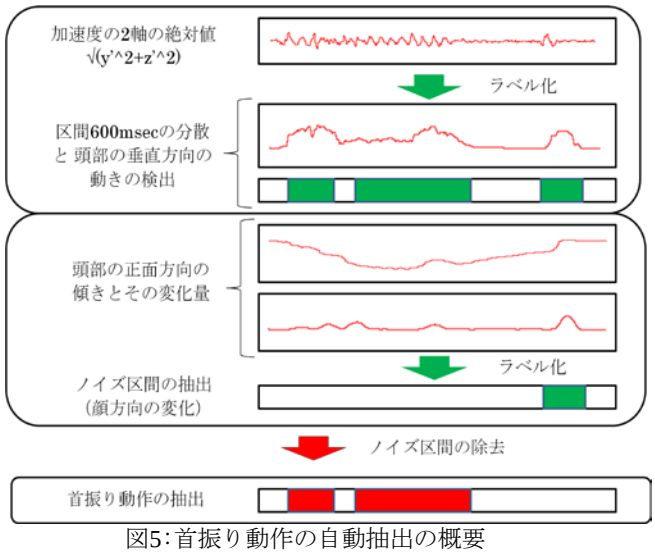

図5：首振り動作の自動抽出の概要

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，図6で示すような前頭部の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4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'軸方向に加速がか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，$T = m + fs$ 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．

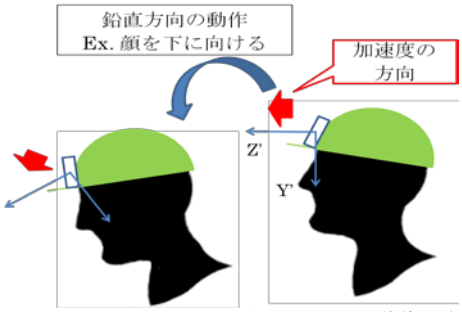

図6:Y'-Z'平面における動作の例

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7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．

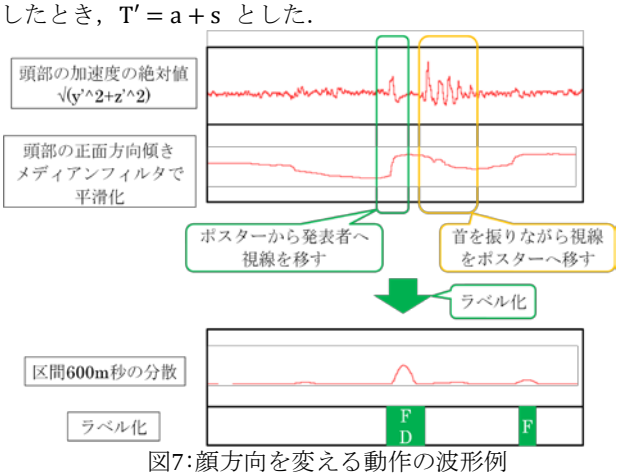

図7:顔方向を変える動作の波形例

(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用し，傾きの変化量の大きい区間とした．加速度センサの傾きデータの取得方法および，傾きの変化量の大きい区間の抽出方法は顔方向を変える動作と同様とした．

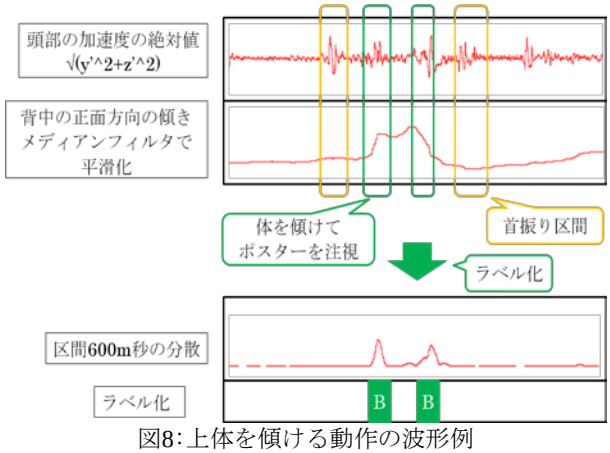

図8:上体を傾ける動作の波形例

(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図9のように周波数解析を行うことで抽出した．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$w(x) = 0.54 - 0.46cos(2\pi x)$ $(0 \leq x \leq 1)$ 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．

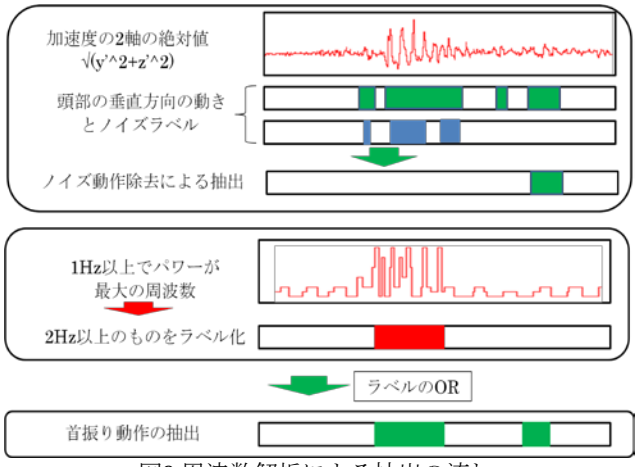

図9:周波数解析による抽出の流れ

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教示はしなかったが，立ち位置が変化することはなかった．

今回精度の評価をするため，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．session 2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．

表1: 自動抽出の再現率と適合率(%)

		発表者 A	非発表者 B	非発表者 C
session 1	再現率	83.6	83	81.3
	適合率	65	79.7	81.9
session 2	再現率	59.1	61.4	71.4
	適合率	37.5	95.3	69.8

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い，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．

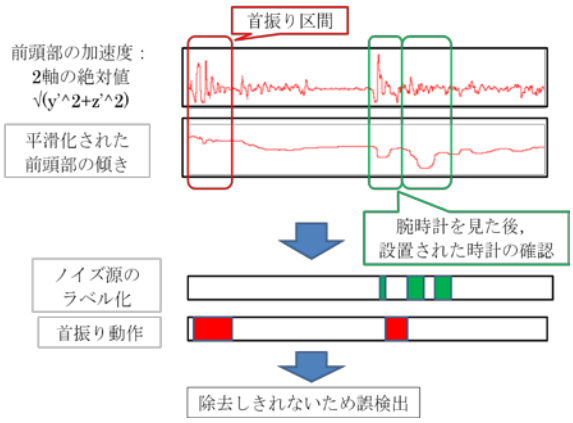

図 10:誤検出例

5. うなずきの機能分類

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 使用するモダリティと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以下とした．

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テを三人分，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．発話を用いるのは注目する被験者の状態が，聞き手であるか話し手であるかを区別するためである． 相槌を使用する理由は，相槌は聞き手の行う言語行動であり，うなずきと同様の機能を持つという先行研究[4]もあることから今回用いた．また，瀬戸口ら[8]によると，ポスター発表では質疑応答のとき質問者と発表者の間で相互注視の状態が多く，説明時はポスターへの共同注意が多いことが分かっている．また質疑応答時では聞き手のうなずき，相槌が多いことも分かっている．質疑応答のときは通常の説明に比べ聞き手の話し手に対する態度が積極的になり，生起するうなずきの機能も異なってくることが予想されるため，視線と首振り動作を加えた．さらに前田ら[5]によると話し手の行う働きかけのうなずきを行う際，聞き手は応答を返すということが言われており，機能分類に必要なモダリティであると考え入力データとした．

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11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．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．

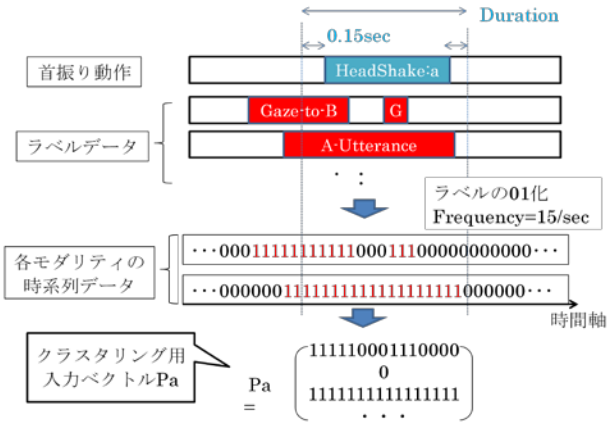

図11:入力データの作成の流れ

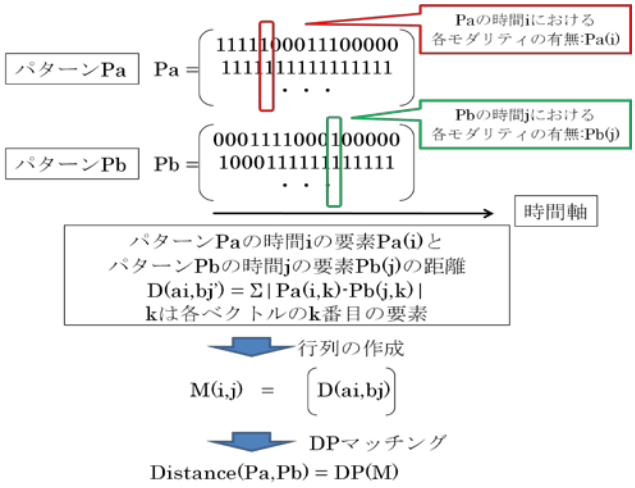

図12:パターン間の距離の算出法

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12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．

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

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．

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

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．

$$Distance_{ab} = DP(M_{ab})$$

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ウォード法を用いた．

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タとみなした．また連結したノードの高さが閾値より低くなった場合も分割を停止するようにした．

今回自動抽出を行った2sessionにおいて，上記の15モダリティを入力データとして，被験者ごとに首振り動作のクラスタリングを行った．session 1，2ともに発表者Aの認識の精度が低いため，発表者のみ手作業で修正を行った．各被験者の入力データ数は180個から240個程度である．

5.3 結果と考察

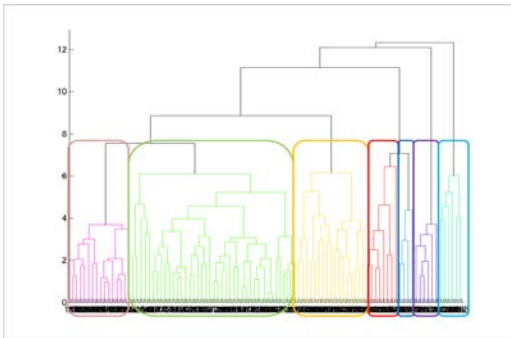

図13:樹状図の例(session 1の非発表者C)

今回 2session 計 6 人のクラスタリングを行った．各被験者のクラスタ総数を表 2 に，実際のクラスタリング結果を樹状図にしたものの例を図 13 に示す．以下では各クラスタがどの機能と対応しているか，またどのモダリティがクラスタの特徴となっているかについて分析結果を述べる．

表 2:クラスタ総数(個)

	発表者 A	非発表者 B	非発表者 C
session 1	5	10	7
session 2	8	7	4

5.3.1 非発表者のクラスタリング結果

まず非発表者のクラスタリング結果について説明する．各非発表者のクラスタと首振り動作の機能の対応を表 3 にまとめた．ここで話し手の行う混在とは，話し手の行う首振り動作と聞き手の行う首振り動作が混ざっており，機能は混在していると考えられるクラスタを指す．また聞き手の行う動作に表示している混在とは発話継続の促しと話し手に対する応答の機能が混在してしまったクラスタとした．

話し手の首振り動作の中で強調を示すクラスタについて説明する．このクラスタは非発表者自身が質問をしている際に発生しており，質問の受け手である非発表者の首振り動作が同期している．これは前田ら[5]のいう話し手の働きかけのうなずきと聞き手の応答としてのうなずきの対応関係と考えられることから強調の機能とした．各クラスタを

特徴づけるモダリティは自身の発話行動と発表者 A の首振り動作である. また session 1 の非発表者 B のクラスタ 1, 3, 4, session 2 の非発表者 C でのクラスタ 2 は話し手行動と聞き手行動の首振り動作が混在したクラスタとなった． session 1 では, 3 つのクラスタはすべて B の質問とそれに対する回答中に発生している．クラスタ 1 ではクラスタ 2 に対し，発表者 A の首振り動作が同期していなかった. クラスタ 3, 4 ではクラスタ 1 に対し, 発表者 A の B に対する視線が多く，非発表者 C の首振り動作区間と被って生起しているといった特徴があった(クラスタ 3:4 個中 4 個, クラスタ 4:8 個中 7 個)．このようなモダリティの表出の違いからクラスタがわかれてしまったと考えられる．また session2 の非発表者 C は質問を行う際，あまり発表者 A に対し視線を向けないという特徴があった．共起するモダリティが他のクラスと差がないため，混在してしまったと考えられる．

次に聞き手行動の機能のクラスタについて説明する．まず話し手に対する応答行為としたのは表3のとおりである．これらのクラスタの特徴は話し手の首振り動作と同期していることがあげられる．この中で session 1 の非発表者 B では，クラスタ 10 はクラスタ 6 に対し，自身の質問に対する回答中であった．A の首振り動作は B のみに対して行う働きかけであり，非発表者 C の首振り動作は同期していないことが特徴的である(8 個中 0 個). 非発表者 C ではクラスタ 5 はもう一人の非発表者の首振り動作も多く同期している. session 2 の各クラスタも同様に他の非発表者の首振り動作が同期しており，働きかけに対する応答の首振り動作は協調的なリズム取りの機能も持っているということが考えられる.

聞き手の行う継続の機能のクラスタにおける特徴は，話し手の行う首振り動作が少ない点である．session 1 の非発表者 B のクラスタが二つに分割されている理由は，他の非発表者の態度が異なっている点があげられる．クラスタ 7 ではクラスタ 9 に対し非発表者 C の首振り動作の同期が多い(クラスタ 7:45 個中 34 個，クラスタ 9:24 個中 10 個)．また非発表者 C では，クラスタ 7 はもう一人の非発表者の質疑応答中に生起している．発表者の視線はもう一人の非発表者に向かっており, 対話に近い状態になっている(12 個中 10 個). それに対しクラスタ 1, 2 では自身の視線は発表者 A に向かっていることが多く(クラスタ 1:8 個中 7 個，クラスタ 2:16 個中 13 個)，説明に対して積極的であることが分割されている理由であると考えられる. session 2 の非発表者 B も同様にクラスタ 3, 4 では他の非発表者の質問中，または回答中に発生している．このような状態は傍参与者の状態であり，自身の相槌の生起数は少ない状態になっている．また session 2 の非発表者 B のクラスタ 4 以外では発話数が少なくクラスタ 4 での発話も笑いであったため，発話意欲が少ないことの現れであると考えられる.

最後に聞き手の機能が混在したクラスタは半分が継続，残りは応答行為のように明確に分割されなかったクラスタである．session 1 の非発表者 B のクラスタ 5 では話し手が自身に対して視線を向けるのが多いという特徴があった．非発表者 C のクラスタ 7 ではクラスタ自体が大きく，あまり混在したクラスタとなってしまった．このようなクラスタはそれのみを取り出してクラスタリングを行い，分割するほうがよいと考えられる．また session 2 の非発表者 C のクラスタの機能が混在している理由は，今回使用したモダリティのうち視線が発生していないことが挙げられる．これは視線変換時にずれが発生していることなどが理由に挙げられる．生起している特徴量が少ないため各首振り動作の距離が全体として同じようになってしまい，大きく機能が混ざったクラスタになっていると考えられる.

表 3:非発表者のクラスタと機能の対応

		Session 1		Session 2	
		非発表者 B	非発表者 C	非発表者 B	非発表者 C
話し手行動	強調	2	6	7	
	混在	1, 3, 4			2
聞き手行動	応答行為	6, 10	5	1, 5	
	継続	7, 9	1, 2, 3, 7	2, 3, 4, 6	1
	混在	5, 8	4		3,4

5.3.2 発表者のクラスタリング結果

次に発表者のクラスタリング結果について述べる．発表者の各クラスタと機能の対応表を表 4 に示す．まず話し手のクラスタは強調的なクラスタが生成された．これらは非発表者と同様に，話し手の首振り動作が多く同期しているクラスタであった． session 1 でクラスタが分かれている理由は聞き手に対する視線配布の量が違う点にある．クラスタ 1 では発表者 A は質疑応答より説明する区間が多く，ポスターを注視することが多い. それに対しクラスタ 4 では B に対して回答を行っており，A から B に対する視線と B から A に対する視線が多いという結果が得られた(視線 A から B:16 個中 14 個，視線 B から A:16 個中 14 個).

session 1 におけるクラスタ 1 は話し手の行う首振り動作のみで構成されているが，聞き手に対して働きかけを行うような強調を示すものとリズム取りなど複数の機能が混在したクラスタとなったため，混在のところに表示した．このクラスタは全体として数が多く，機能もまとまっていないためこのクラスタを取り出し，さらにクラスタリングをかけることで分割することがよいと考えられる．同様に話し手の機能が混在しているのは session 2 のクラスタ 7 とクラスタ 8 となった．クラスタ 7 は session 1 における発表者のクラスタ 1 と同様に大きいクラスタのため，同様に分割したほうがよいと考えられる．またクラスタ 8 の特徴は質問が行われた後の回答のはじめに発生していることである. これは回答初めにおける自身の発話の強調の他，発話権の保持のような機能が考えられる.

次に聞き手の行うクラスタについて説明する．今回発話継続の促しの機能が session 1, 2 それぞれに二つ生成された. これは非発表者が質問を行うのを許可しているということが考えられるため，この機能に分類した．二つ生成された理由は，非発表者がそれぞれ個別に質問を行っているため，特徴となるモダリティが異なってしまったことが挙げられる．また session 2 のクラスタ 2 は三人で談笑を行っている区間と，非発表者 B の質問を行っている区間の二つで生起した首振り動作で構成されている．質問区間で生起した首振り動作は継続の機能を持つと考えられるが，談笑中の区間は応答や継続が混在しており，機能は混在しているとした.

最後に間発話中の機能について説明を行う．session 2 のクラスタ 7 は間発話中に大きく動かす首振り動作が含まれていた．これは間を埋め自身の発話権の保持を示している発話の前振りであると考えられる．

表 4:発表者のクラスタと機能の対応

		session 1	session 2
		発表者 A	発表者 A
話し手	強調	2, 4	3, 4
	混在	1	7, 8
聞き手	応答行為		
	継続	3, 5	1, 6
	混在		
間発話中			5

5.3.3 考察

発表者，非発表者の各クラスタを分析した結果，話し手が働きかけるような強調の機能，またそれに対して聞き手が行う応答，他にも聞き手の行う発話継続の促しといった機能がみられた．また間発話中に間を埋めるようなクラスタや回答の最初のほうに集中して出る発話権を保持するためのクラスタといったものも見られた．

しかし，大きく機能が混在したクラスタが生成されており，このようなクラスタは分割する必要がある．分割するために大きいクラスタのみを取り出して部分的にクラスタリングを行うことが考えられる．また，現在のようにモダリティを 15 次元並列に扱って距離を計算するだけでは大きなクラスタは分割できないため，モダリティの取捨選択が必要である．首振り動作にとって重要なモダリティは誰が発話いるかとその話し手の視線，首振り動作であることが考えられるため，モダリティを減らすことで，よりこれらのモダリティの発生タイミングに注目したクラスタリングができると考えられる．また，聞き手の首振り動作のクラスタリングを行う場合，会話における積極性を見るためにポスターへの共同注意や話し手の指さし，他には話し手との相互注視が行われているかといったモダリティを使用することが考えられる．

6. おわりに

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お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，また間発話中に間を埋める機能が分類できた．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，入力とするモダリティの取捨選択，他にも大きなクラスタに対してより詳細に分割を行うなど機能分析の手法も改善する必要が分かった．

今後の発展としてよりモダリティを削減して大きいクラスタに対しそのクラスタを抽出してさらなるクラスタリングを行い分割するなど，うなずきの機能とそれにかかわる多人数会話での参加者の言語，非言語行動の詳細なモデル作成をすることが考えられる．これによって多人数会話における人がどのようにふるまい，会話の流れを制御しているかの理解ができると考えられる．

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．また，本研究における実験協力や論文へのアドバイスなど多大なご協力を頂いた，京都大学河原研究室並びに西田・角研究室の皆様に感謝いたします．

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] 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).
[6] E. McClave.: Linguistic functions of head movements in the context of speech. *Journal of Pragmatics*,32(7):855{878, 2000.
[7] N. Kern, B. Schiele, H. Junker, P. Lukowicz, and G.Tr¨oster: Wearable sensing to annotate meeting recordings. Personal and Ubiquitous Computing, 7(5):263{274, 2003.
[8] 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).
[9] 來嶋宏幸，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).
[10] 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).
[11] 独立行政法人国立国語研究所：日本語話し言葉コーパス.
[12] 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).