

## 多人数会話におけるうなずきの自動検出と機能分析

齊賀 弘泰†      角 康之‡      西田 豊明‡

† 京都大学 工学部 ‡ 京都大学 情報学研究科

### 1 はじめに

人は会話的インタラクションにおいてうなずきや視線、指さしなどの非言語行動を行うことで、感情の表出や強調、同意や否定、他にも発話交代の調整などを行っているといわれる。

その中でもうなずきは相手に対する視覚的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。Maynard[1]はうなずきのような頭部の垂直な動きを9つの機能に分類した。聞き手行為の機能としてのうなずきだけではなく、話し手の行う強調や肯定の機能、さらに誰も発話していない状況において、発話権をとる前振りといった会話管理にも用いられるとした。

そこで本論文では頭部の垂直な動きに注目し、聞き手行為のうなずきだけではなく話し手の動きも含めた首振り動作の自動抽出を行う。次に自動抽出された首振り動作に対して他の非言語行動データをもとに機能分析を行う。

### 2 データ収録

分析対象とする会話データとして三人によるポスター発表を題材とした(図1)。これは発表者役の被験者が聞き手役の被験者二人に対して、あらかじめ作成したポスターをもとに自身の研究内容を発表するものである。今回は20分の長さの1会話分を対象とした。

センサ機器は三軸無線加速度センサ、モーションキャプチャシステム、視線追跡装置、音声マイク、ネットワークカメラを用いた。加速度センサは首振り動作の検出のため被験者の前頭部、背中上部、腰の三点、左右の手首に装着した。モーションキャプチャ、視線追跡装置、音声マイクは視線、発話、あいづちラベルを作成するため使用した。ネットワークカメラは会話状況の記録に用いた。

### 3 首振り動作の自動抽出

対象とする首振り動作はMaynard[1]の定義した頭の縦振り動作を用いて以下のように定義した。頭部を顔

Automatic detection and function analysis of nodding in multi-party conversation

†Hiroyasu Saiga and Yasuyuki Sumi†and Toyooki Nishida‡

†Faculty of Engineering, Kyoto University

‡Graduate School of Informatics, Kyoto University

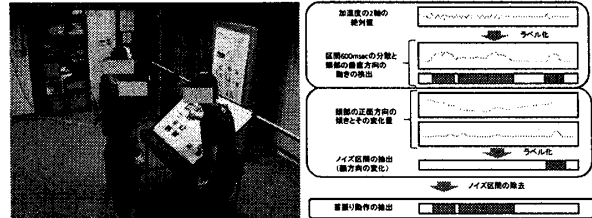


図 1: ポスター発表会話      図 2: 首振り動作の検出

の正面方向に上または下に動かした後正常位置に戻る動作を首振り動作とし、正常位置に戻らず顔方向を変えただけのものは首振り動作とみなさない。また顔の左右方向に振る横振りは抽出対象としない。

首振り動作の自動抽出の概要を述べる。前頭部の加速度センサの水平の正面方向と、垂直方向の二軸の絶対値をとったものをもとに、区間 600m 秒を窓幅として区間内のデータの分散をとった。図2のように一段目の絶対値の波形で振動のある区間は二段目に表示されているとおりに分散値が大きくなるので、全体の中央値を  $M$ 、標準偏差を  $s$  として、閾値  $T$  を  $T = M + s/10$  のように定め、それ以上大きい区間を頭部が垂直に動いた区間とした。

次にノイズ源となる動作の抽出を行った。顔方向を変えただけの動作、体を傾げる動作、立ち位置を変えるなど体全体が動く動作をノイズ源とみなした。

顔方向を変える動作は加速度センサの傾きを特徴量として用いた。これは重力加速度の方向から計測するもので、今回は正面の水平方向に対する傾きの変化量から抽出した。また首振り動作がノイズとなり傾きも振動するため、メディアンフィルタをかけ、振動を消し大まかな傾きをとった。

体を傾げる動作は背中上部の加速度センサの傾きを利用して抽出した。この動作の抽出法は顔方向を変える動作と同様の手段で、体の正面方向と左右方向の二方向について検出した。

最後に立ち位置を変えるなどの動作の抽出を行った。これは腰に付けた加速度センサの三軸の絶対値をとったものに対して、区間 600m 秒を窓幅として区間内のデータの分散をとり、分散値の大きい区間は体全体が動いている区間とみなしノイズとした。三軸の絶対値に対してメディアンフィルタをかけ、首振り動作によ

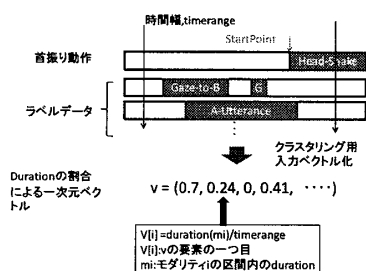


図 3: クラスタリングの入力ベクトルの作成

る振動ノイズを消した。

以上のノイズ動作区間を除去した垂直動作の区間を首振り動作の区間として抽出を行った。

## 4 首振り動作の機能分析

### 4.1 分析方法

前章で述べた抽出法により自動抽出した首振り動作に対して、視線やあいづちなど首振り動作と共起しておこるモダリティの出現の組み合わせによってクラスタリングを試みた。

入力する特徴量としては、会話参加者がどこを見ているか (他の被験者二人とポスター)、発話の有無を使用した。発話データはあいづちを別のモダリティとみなして、あいづちとその他の発話の二つに手作業で分けた。視線データに関してはモーションキャプチャシステムと視線追跡装置による三次元座標データより自動的に作成したデータを用いた。

また入力ベクトルは、図 3 に示すように各首振り動作開始点前後の一定の区間で、各モダリティが発生している時間の割合を要素とした一次元ベクトルを作成し、入力ベクトル間の距離はユークリッド平方距離を用いた。

### 4.2 分析

まず上記の三人分の視線、発話、あいづちを用いてクラスタリングを行った。なお視線に関しては誰を見ていたか、またはポスターを見ていたかの 3 モダリティとして全 15 モダリティを入力とした。瀬戸口 [2] の研究より視線とうなずき、うなずきとあいづちは同期するといわれていることから視線とあいづちを選択した。これに発話モダリティを加え、入力ベクトル時間幅は開始点 2 秒前から開始点までの 2 秒間とした。首振り動作直前の行動パターンから首振り動作の機能を推定するために、各モダリティに対して重みは付けずにクラスタリングを行い、各クラスがどのような機能

クラス	人数	発話者のあいづち	聞き手 A の発話	重心への距離の平均
1	6	0	0	0.21
2	6	5	6	0.13
3	38	0	0	0.14
4	25	0	0	0.24
5	5	3	4	0.43
6	5	0	0	0.14
7	6	5	6	0.30
8	19	0	0	0.25
9	6	0	0	0.29

表 1: 聞き手 A の結果

を持つのかを検証した。

はじめに発生している時間の割合を用いた一次元ベクトルのクラスタリングを行った。被験者の首振りデータに対してクラスタリングを行い、各クラスを分析すると、いくつかのクラスは自分自身の発話が直前に多く、残りのクラスは発話がほとんどないものに明確に分かれた。これはポスター発表会話においては発表者が説明を行っている区間と聞き手が発話権を取得し質問を行っている区間では、発話以外にも質問者と発表者が視線を合わせるなど通常の発表時にくらべ、発生するモダリティの種類が変わってくるものと思われる。

また、話し手の首振り動作のクラスの中には他の聞き手二人に対して視線を多く送り、相手へ発話内容を強調しているようなものも見られた。

## 5 おわりに

本論文ではうなずきという非言語行動が会話において重要な機能を持つものとして、話し手、聞き手両方の頭部の垂直な動きを首振り動作として自動抽出を行い、機能分析を行った。その結果首振り動作の開始点前後における発話や視線などのモダリティの発生状況をみることで首振り動作の持ついくつかの機能が推定された。

今後はうなずきに共起する他のモダリティを入力データに追加することでより詳細なうなずきの自動検出を行うことともに、機能からコンテキストの推定をするなどが考えられる。

### 謝辞

本研究は、科研費補助金「情報爆発時代に向けた新しい IT 基盤技術の研究」の一環で実施された。

### 参考文献

- [1] 泉子・K・Maynard. 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.