

動画から顔の動きを抽出する試み —対話解析・修学行動評価への適用を目指して—

Extraction of Facial movement from movie

伊藤 敏^{*1}, 大塚容子^{*2}, 鷲野 嘉映^{*3}
Satoshi ITOU^{*1}, Yoko OTSUKA^{*2}, Kaei WASHINO^{*3}

^{*1} 岐阜聖徳学園大学経済情報学部

^{*1}Fac. Economics and Information, Gifu Shotoku University

^{*2} 岐阜聖徳学園大学外国語学部

^{*2}Fac. Foreign Languages, Gifu Shotoku University

^{*3} 岐阜聖徳学園大学看護学部

^{*3}Fac. Nursing, Gifu Shotoku University

Email: itous@gifu.shotoku.ac.jp

あらまし：会話解析をするためのツールとして、動画から顔検出をし、顔の特徴点、顔の向きなどの情報を抽出することで会話解析に必要と思われる、うなずきや発話状態を取り出す手法開発を試みた。また、カメラ1台で複数者の会話中の正面動画を取得する方法を提案した。その結果、会話中の聞き手や話者の顔を中心とした行動を抽出し、目視による結果と良く一致することを示した。

キーワード：会話、録画、顔検出、うなずき

1. はじめに

対面による教育活動は、言語及び視覚的行動により行われている。講義や講演における聴衆の反応（うなずきや疑問を示す行動）や、あるいはグループ活動での討論、数名での会話などにおける聞き手の言語活動以外の行動（ノンバーバルコミュニケーション）が、話者にとって「話しやすさ」などに関係すると思われる⁽¹⁾。

そこで、会話時のノンバーバルコミュニケーションのうち「うなずき」の検出をする試みがなされている。会話者の頭部に慣性センサを取り付け、動きを検出する試みや⁽²⁾、カメラを用いて動画から検出する試みがある⁽³⁾。

我々は学習者にセンサを装着して、学習者の行動を観測する試みを行ってきた⁽⁴⁾。課題中の頭部の動きや、脈波、呼吸数を取得することで負荷の状態などを推測した。しかし、検出装置を装着することによる学習者への身体的・精神的負担は測定に影響を及ぼす。よって、非接触で学習負荷時の行動を記録・解析する方法の検討も重要であると考える。

学習者の行動情報を非接触で取得する手段として、ビデオカメラによる観測が考えられる⁽⁵⁾。これは、学習者に装置を装着するのではなく、非接触で行動を観測可能であり、学習者への負荷が少ない。一方で、記録された動画を目視により解析した場合、解析者に大きな負担がかかる。

本稿の目的は、会話行動を動画観測する支援ツールを提供することである。動画より、顔を検出し、顔の位置、特徴点座標を取得する。それらより、顔の位置変動および特徴点変動から顔の向き推定を行う。会話による実験時間は数分間とした。

本稿では、2章で動画録画の方法と顔検出の方法を記述し、3章で応用例として、二者対話での会話を録画し、行動の抽出を試み、4章でまとめる。

2. 方法

会話をカメラで録画して、その動画から顔を検出し、顔の特徴点座標を得た。それらの特徴点座標より (1) 顔の中心(鼻の頂点) (2) 口の開閉 (3) 顔の向き を計算することで、会話中の行動を抽出した。

2.1 録画方法の工夫

会話の録画に汎用のカメラを用いた場合、対話では、聞き手の動画を撮影した。会話研究には聞き手だけでなく話し手の動きも知る必要がある。そのため全会話参加者の動画を撮影するために、全天球360度カメラを用いた。一般に会話は2名の場合対面で、3名以上の場合疑似的な円陣で行われる。それらの場合、各会話参加者の正面を録画するために複数のカメラが必要であり、各カメラの同期を取る必要がある。これらの煩雑さを避けるため、全天球360度カメラ1台を用い、会話参加者にはカメラを正面に見るように伝えた。得られた動画を平面に展開することで全会話参加者が正面動画として得られ、画像解析が可能になる。

2.2 顔情報の抽出と数値化

動画からフレーム毎に機械学習ライブラリであるdlib⁽⁶⁾を用いて顔検出と顔の特徴点68点を検出した。得られた顔器官の両目じり、鼻の頂点、両口角、顎最下部の6点を用いて2次元座標から向きの算出などの行列演算を行い、や画面表示にはOpenCVのラ

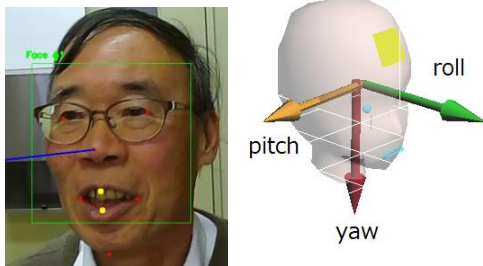


図 1. 顔検出と特徴点の座標, 顔の向き

イブラリ⁶⁾を用いた。顔の向きは, yaw, pitch, roll として求めた⁷⁾。

図 1 に, 用いた顔の特徴点 6 点を赤色で, 唇の中央上下を黄色で表示した。顔の向きを鼻の頂点からの青色の線分で表した様子, yaw, pitch, roll を示す。これらより得られた座標は動画の 1 フレーム毎 (30fps) に数値として保存され, 会話の分析に用いられた。

3. 会話中の行動分析への適応

人は会話中に話者は口を動かし, 聞き手はうなずきなど様々な行動をする。うなずき行動は会話の制御に寄与すると言われる¹⁾。「うなずき」を顔の短時間での上下運動とみなし, 顔の中心部に位置する鼻頂点の上下動と顔の向きの pitch の変動から推測する。

3.1 対話へ 聞き手の行動

二者による会話の聞き手のみを録画して, 目視に

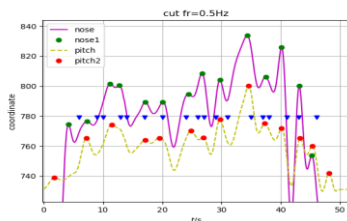


図 2. 聞き手の鼻の位置と pitch の変化, 目視によるうなずき位置

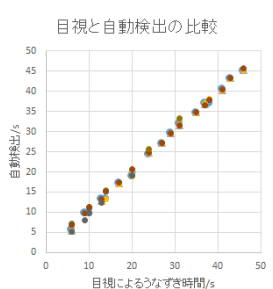


図 3. 顔座標の変動と目視による比較



図 4. 全天球カメラ撮影の二者による会話

よる「うなずき」検出と, 鼻の頂点座標, および顔の向き pitch からの自動検出した結果の比較をした。図 2 に鼻の座標移動と顔の向き pitch の変動の時間経過を示す。各グラフの頂点の時刻と目視により観測した時刻の比較を図 3 に示す。ほぼ一致する結果が得られた。

対話中に体を前後の大きく揺らす聞き手の場合も, 目視と顔座標変動において同じ傾向を示した。

3.2 二者の対話

二者の対面による会話を全天球 360 度カメラで録画し平面へ展開した一例を図 4 に示す。展開により前述の手法を顔検出などに利用可能である。二者は横軸座標値で区別可能であり, 話者と聞き手の区別は唇の上下距離から推測が可能であった。

4. まとめ

会話時間が数分という制限下では, 動画の顔検出解析より, 鼻の位置などから得たうなずき行動は目視による観測と良い一致を示した。また, カメラ 1 台の利用で, 複数者の正面動画を取得可能であることを示した。これらの結果より, 本方法は会話分析のツールとして利用可能であると考えられる。

本研究の一部は科研費 (16K01083) の助成を受けたものである。

参考文献

- (1) 泉子・K・メイナード: “会話分析”, くろしお出版, 東京 (1993)
- (2) 斎賀弘泰, 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としての機能分析”, 情報処理学会研究報告, Vol.2010-UBI-26 No.1, pp1-8 (2010)
- (3) 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).
- (4) 伊藤敏, 王琳琳, 鷺野嘉映, 井上祥史: “慣性センサを用いた行動検出試行”, 教育システム情報学会 2016 年度第 2 回研究会, pp9-13 (2016)
- (5) <http://dlib.net/> 2018 年 6 月 11 日確認
- (6) <https://opencv.org> 2018 年 6 月 11 日確認
- (7) <https://www.leanopencv.com/head-pose-estimation-using-opencv-and-dlib/> 2018 年 6 月 11 日確認