

## 自由対話におけるマルチモーダル情報を用いた意図識別

平松 晃成<sup>†</sup> 酒向 慎司<sup>††</sup> 北村 正<sup>††</sup><sup>†</sup>名古屋工業大学工学部情報工学科<sup>††</sup>名古屋工業大学

## 1. はじめに

近年では様々な場面で音声認識システムが導入されている。それらは丁寧な発話を入力対象にしているため、実際の会話のような話し言葉を入力としたシステムの実用化は長年研究され続けてきた。しかし会話中で話し手の言語情報だけでは意図や感情を理解したとは言えず、それに付随する話し手の声の抑揚、顔の表情、ジェスチャーなどの情報を活用して理解していると考えられる。このような背景から頑健な音声認識を実現するために、対話で現れる言語情報に付随して現れる非言語情報を用い、意図を考慮することが重要になると考えられる。

先行研究[1]では自由対話中で頭部動作と、音声から得られる非言語情報を組み合わせて意図の識別を行った。しかしタスクが意図に寄与しない頭部動作が多く含まれる内容であった。そのため音響特徴のみの識別率と頭部動作を合わせた時の識別率に変化が見られなかった。

そこで本研究では先行研究での識別率と、タスク内容を頭部動作が意図に反映させるように考慮しなおした場合の識別率を比較し、また意図に寄与すると思われる顔の特徴量も新たに追加し識別を行う。

## 2. 発話意図の分類

対話は、働き掛け、応答、了解の3つから構成されており、これらは意図の機能に相当すると考えられる。働き掛けは相手に応答を求める機能であり、応答は働き掛けに対する返答の機能を持つ。了解は現在の働き掛けと応答のやりとりを終了させる機能や、次のやりとりを促す機能がある。これらの機能は、役割に応じてさらに複数の発話意図に分類することが出来る。本研究では荒木ら[2]が提案した発話単位タグに基づき表1の発話意図の識別を目的としている。

## 3. 自由対話における識別方法

本研究では、発話中に意図が随時変化し得る自由対話を対象とする。自由対話では意図の種類とその変化は一般的に一つの意図区間が既知であれば、単一発話区間の文頭や文末から意図の違いを判断できるとされている。文頭は発話開始から最初の母音区間終了まで、文末は最後の母音区間開始から発話終了までとしている。

表 1. 発話意図の機能と詳細の分類

機能	発話意図
働き掛け	情報伝達、確認、提案、示唆、依頼、未知情報要求、真偽情報要求、勧誘、約束・申し出、希望
応答	肯定・受諾、否定・拒否、未知情報応答、保留
了解	了解、相槌、復唱

表 2. 特徴量の種類

音響	基本周波数 (フレーム内の平均)
	短時間平均パワー (フレーム内の平均・傾き)
頭部動作	頭部速度 (フレーム内の平均・傾き)
意図機能	相手発話者の意図ラベル
	前発話の意図ラベル
顔の特徴	目と眉毛の距離 (フレーム毎の傾き)

## 3.1. 発話意図推定手法

本研究では、C4.5を用いて決定木で学習し、人間同士の対話から得られる発話や動作のデータからフレーム単位 (0.1 sec ~ 1.0 sec) で特徴量を抽出し、それらを決定木学習に用いる説明変数とする。説明変数に用いる特徴量を表1に示す。音声から得られる非言語情報として、基本周波数パターンや単時間平均パワーを用い、頭部動作や顔表情は意図の寄与する特徴と考え識別に用いる。

## 3.2. 対話データ

先行研究[1]では意図に寄与すると思われる頭部動作の特徴が正確にとれなかったことをふまえ、本研究ではクイズ形式という対話でお互いが顔を向き合うような形で収録を行った。収録の際には顔の各部位の点を取得することが可能であり、音声を同期に録音可能であるKinect センサーを用いた。意図のラベルづけは音声と映像を見ながら主観で行った。本研究では複数のラベルを持つ場合は識別の対象外としている。

## 4. 評価実験

音声から得られる特徴と頭部動作の特徴を追加した場合の発話意図推定実験 (実験1) と、顔の表情から得られる特徴量を追加した場合の実験 (実験2) を行う。実験1で対話タスクの変更により意図に寄与しない場合の頭部動作の特徴を取り除き、寄与する場合の頭部動作の特徴を追加し向上を確認する。実験1では顔の表情が意図に寄与する特徴かを検討する。

## 5. まとめ

本研究では頭部動作と顔の特徴点を用いて意図識別を行った。しかし非言語情報には視線情報やジェスチャーなどがあり、今後として新たにそれらの特徴量を用いて識別を行っていくつもりである。

## 参考文献

- [1] 伊藤大介, 酒向慎司, 北村正, ¥自由対話における視聴覚特徴を用いた発話意図推定の検討", 情報処理学会研究報告, 2010-SLP-84(23), 1-6.  
 [2] 荒木雅弘, 伊藤敏彦, 熊谷智子, 石崎雅人, ¥発話単位タグ標準化案の作成", 人工知能学会誌, 1999, Vol.14, No.2. 251-260.