

発話文における意図識別に有効な音響特徴の検討

A Study of Effective Acoustic Features for Discrimination of Intention in Utterance Sentence

伊藤 大介[†]
Daisuke Ito

酒向 慎司[†]
Shinji Sako

北村 正[†]
Tadashi Kitamura

1. はじめに

計算機が人の意図を理解する事で、人と機械のコミュニケーションはより円滑なものになると期待される。近年では、音声認識によって得られた言語情報から意図を推定する研究が数多く行われている [1]。しかし、言語情報だけでは的確な意図の推定は難しいといえる。たとえば、「はい」という発話は、相手の発話に同意する場合にも、相手の発話が聞き取れなかった場合にも用いられるが、それらの示す意図が違うのは明らかである。人同士のコミュニケーションにおいて、意図を伝える視聴覚情報の割合に関して諸説があるが、文献 [2] によると、言語情報によって伝わる話し手の意思は全体の約 35% であり、残りの 65% は話し方、動作、その人の身体的特徴などの非言語情報によって伝わりとされている。

このような背景から、我々は、言語情報に加えて複数の視聴覚情報から意図を推定する手法を検討している。本稿では、その基礎検討として、音響特徴である基本周波数パターンからの意図推定の可能性について検討した。自由対話における音声から、意図に寄与するであろう特徴量を分散分析によって調査した。

2. 発話文における意図推定

2.1 人同士における意図理解

人間同士のコミュニケーションにおける意図理解のプロセスについて、文献 [4] では次のようにモデル化している。人は、相手と情報を交換する機能と、得られた情報を解釈して意図を理解する機能を用いて意図理解を行っている。前者では、言葉や動作などの人間から発信された多チャンネルの情報を体の受容器に入力している。一方、後者では、入力された多チャンネルの情報を、人が予め持っている知識と参照しながら、総合的に解釈している。

計算機における意図理解において、このモデルを適用すると、前者は人の発信している情報から意図に関係するものを抽出する問題として置き換えられ、後者は入力された複数の情報を総合的に解釈する識別問題に置き換えられる。本稿では、前者の特徴量抽出の問題の一つとして、どのような音響特徴が意図に寄与しているかを明らかにする。

2.2 発話意図の分類

発話意図推定を識別問題として捉えた場合、意図を何らかの基準で分類する必要がある。

対話は 1 つ以上のやりとりから構成され、そのやりとりは “Initiate-Response-(Follow-up)” から成る。Initiate は、新しいやりとりを開始する働き掛けの機能を持ち、Response は働き掛けに対する反応の機能を持つ。Follow-up は、現在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。これらの構成要素は、役割に応じてさらに複数の発話意図に分類することができる。本研究で

表 1: 発話意図の分類

構造	発話意図
Initiate	情報伝達, 確認, 提案, 示唆, 依頼, 未知情報要求, 真偽情報要求, 約束・申し出, 希望
Response	肯定・受諾, 否定・拒否, 未知情報応答, 保留
Follow-up	了解, 相槌, 復唱

表 2: 収録データ

収録データ	3分 × 16 セット
収録人数	6人
総センテンス数	403文

は荒木らが提案している発話単位タグ [3] を発話意図と定義する。本研究で扱う発話意図の分類を表 1 に示す。

3. 意図識別に有効な音響特徴の分析

発話意図識別の基礎検討として、音声からいくつかの特徴量を抽出し、それらの意図への寄与に関して考察する。

3.1 音声データ収録

自然な意図を含んだ音声データを得る為、読み上げや演技などではなく、自由対話による音声収録を行った。本研究では、レストランにおいて二名のうちどちらかが相手に晩御飯を御馳走するという設定の下、二者間の自由対話を収録した。その際、メニューを見ながら対話を行う事で、自然な対話を模倣した。また、収録した発話データには、センテンスの分割、発話意図のタグ付け、書き起こしを手動で行った。収録したデータ規模について表 2 に示す。

3.2 特徴量抽出

一般的に、人の感情や意図を表す特徴量として基本周波数パターン (F0 値) が用いられる。しかし、F0 値は意図だけでなくアクセントなどの発話内容にも依存する。F0 値から、発話内容の情報と意図とを分離して扱う事は困難である為、まずは、発話内容が同じ時の F0 値の違いについて検討する。発話内容が同じ時に意図間で F0 値の違いが見られれば、発話内容の言語列とセグメンテーションが既知という制限の下で、意図推定が可能であるといえる。

しかし、自由対話で収録した場合、センテンス単位で同じ発話内容が表れる事は少ない。そこで、本研究では、先行研究 [5] を元に、センテンスの語頭と語尾に意図関与する特徴が表れると考え、語頭や語尾の一区間に対して特徴量抽出を行った。一区間の単位には、様々考えられるが、本研究では音素を一区間の単位として、語頭と語尾の第一モーラにおける母音部分と語尾に多く見られる /N/ の区間を用いた。

ここで、音素区間の検出には Julius を用いた音素セグ

[†]名古屋工業大学, Nagoya Institute of Technology

表 3: 特徴量の種類と分散分析の結果

特徴量番号	説明	語頭					語尾					
		/a/	/i/	/u/	/e/	/o/	/a/	/i/	/u/	/e/	/o/	/N/
1	F0 値の傾き		*		*				*		*	
2	F0 値のレンジ (最大値-最小値)				*				*			
3	F0 値の平均	*			*				*		*	
4	F0 値の分散				*				*			
5	F0 値の最高値				*				*		*	
6	F0 値の最低値	*							*		*	

メンテーションで行い、F0 値の抽出には YIN を用いた。また、特徴量を抽出する前には人毎に抑揚と声の高さの正規化を行っている。

3.3 分散分析

語頭か語尾が同じであるグループごとに分散分析を行い、意図毎の特徴量の平均の差の検定を行った。特徴量抽出において得た特徴量と分散分析の結果を表 3 に示す。

この結果、表中の*が付与されている音素と特徴量の組み合わせにおいて有意差 (有意水準 5%) が確認された。これにより、語頭や語尾の一区間の F0 値に、意図を識別する要素があるとわかる。

次に、この結果を元に多重比較を行い、どの意図間に有意な差があるかを調査した。多重比較には Tukey 検定を用いた。

多重比較の結果、いくつかの特徴量では、データ数が明らかに少ない意図において有意差が見られていたが、信憑性は薄い。しかし、語尾が/N/の区間の F0 値の平均値においては、真偽情報要求に対して肯定・受諾、保留、了解、相槌の四つに有意差を確認できた。ここで、真偽情報要求とは相手に「はい」または「いいえ」で答えられる質問をするものである。それに対し保留・相槌は相手の発言に対して回答を先延ばしにする時や相手の発言を促す時に使われ、了解はやりとりを締めくくるときに使われるものである。この三つの意図はどれも強い意思は含まれていないと考えられる。真偽情報要求とこれら三つの意図間に有意差が見られたという事は、語尾の F0 値の平均値で疑問文と強い意思が含まれていない意図を識別できる可能性を示している。

3.4 クラスタリング

語尾が/N/の区間の F0 値の平均値に対してクラスタリングを行う事で、分散分析・多重比較で得られた結果の妥当性を検討する。クラスタリングには LBG 法を用いた。クラスタリングの結果を表 4 に示す。ここで、クラスタリングは全意図で行ったが、本稿には有意差の見られた 5 つの意図の分類のみを示す。

その結果、保留・了解・相槌はほぼ一つのまとまりとしてクラスタに割り当てられ、それぞれ真偽情報要求とは違うクラスタに割り当てられている場合が多い。これは、語尾の F0 値の平均値を見る事で、疑問文と強い意思を含まない意図とを識別できるという多重比較で得られた仮定通りの結果となっている。

4. むすび

複数の視聴覚情報からの発話意図推定の基礎検討として、どのような音響特徴が意図に寄与しているのかを調査した。その結果、疑問文と強い意思が含まれていない相槌などとの間で有意差が見られた為、語尾の一区間における意図推定の可能性を示した。しかし、多くの特徴

表 4: 語尾が/N/区間の F0 値の平均値における出現意図

クラス	意図	出現回数
1	肯定・受諾	5
	保留	1
	了解	1
	相槌	1
2	真偽情報要求	1
	相槌	1
3	保留	1
	了解	2
	相槌	2
4	肯定・受諾	3
5	真偽情報要求	1
	肯定・受諾	1
	相槌	3
6	真偽情報要求	1
7	肯定・受諾	1
	保留	5
	了解	1
	相槌	2
8	真偽情報要求	1

量の意図への寄与を示す結果とはならなかった。これは、特徴量の問題以外にもデータ数やセグメンテーションの信頼性の問題が考えられる。今後の課題として、F0 値の変動を連続値でなく、上昇や下降などの離散値で扱う事や、F0 値だけでなく、パワーや時間方向の特徴を検討する事等が挙げられる。また、今回は収録時に頭部動作等も音声と整合して取得しており、音響特徴だけでなく、発話者の動作に基づいた意図推定についても今後検討していく。

参考文献

- [1] 松本 宗也, 傳松 明, 白井克彦, “音声対話システムにおける発話意図推定”, 社団法人情報処理学会, 第 70 回平成 20 年 (2), 2-141, 2008.
- [2] Marjorie Fink Vargas, 石丸正 訳, “非言語コミュニケーション”, 新潮選書, 1987.
- [3] 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, Vol.14, No.2, Mar. 1999.
- [4] 佐藤知正, 西田桂史, 市川純理, 畑村洋太郎, 溝口博, “ロボットによる人間の意図の能動的な理解機能”, 日本ロボット学会誌, Vol. 13, No. 4, pp. 545-552, 1995.
- [5] 藤江真也, 江尻康, 小林哲則, “肯定的/否定的発話態度の認識とその音声対話システムへの応用”, 信学論, (D-II), Col.J880D-II, No.3, pp. 489-498, 2005.