

日常的な発話からの話者の感情状態の推定*

佐竹智幸, 森大毅 (宇都宮大・工)

1 はじめに

人間同士のコミュニケーションにおいて、音声に付随して伝達されるパラ言語情報の役割は非常に重要である。そのため、感情や発話意図、態度などのパラ言語情報を識別する研究が近年盛んに行なわれている [1][2]。本研究ではその中でも豊富なパラ言語情報が含まれている日常的な発話を分析対象とする。これまでの分析は少数の特徴的な話者を対象としてきたが [3]、本研究では複数の話者のデータを用いて、音響特徴量からパラ言語情報も含めた感情状態の推定を行ない、不特定話者モデルを構築した。さらに、話者ごとの分析結果と比較することで、構築したモデルの妥当性を検討した。

2 音声データ

分析には宇都宮大学パラ言語情報研究向け音声対話データベース [4] を使用する。これは親しい関係にある 7 ペア、14 名の大学生が「4 コマまんが並べ替え課題」と呼ばれる課題を行なった対話音声からなる。

このデータベースの各発話には評価者 3 名分の感情状態のラベルが付加されている。感情状態は快-不快、覚醒-睡眠、支配-服従、信頼-不信、関心-無関心、肯定的-否定的の 6 つの評価項目をそれぞれ 1 から 7 までの 7 段階で評価することで表現する。

14 名の話者、4716 発話を分析の対象とし、評価者 3 名の平均評価値を正解ラベルとする。

3 音響特徴量

以下の 4 つの音響特徴量を分析に用いた。

f_0 レンジ

1 発話中の対数基本周波数の最大値と最小値の差で f_0 レンジを定義する。

強度最大値

1 発話中の強度の最大値で定義する。

発話速度

発話のモーラ数を持続時間で割った 1 秒当たりのモーラ数で定義する。ただし、発話に笑いやポーズが含まれている場合、それらの区間は持続時間から除いた。

$f_{aperiodic}$ 平均値

$f_{aperiodic}$ は周期的/非周期的成分境界周波数 [5] を表わし、声の気息性に関するパラメータである。気息性が高くなると、非周期的成分が多くなり、 $f_{aperiodic}$ は低くなる。1 発話中の有声区間の平均値で定義する。

4 分析結果

4.1 不特定話者モデルの構築

不特定話者モデルを構築するために、全話者のデータを使い、音響特徴量を独立変数、感情状態の各項目を従属変数として重回帰分析を行なった。結果を Table 1 に示す。表の各音響特徴量の値は標準偏回帰係数を表わしている。

表よりすべての項目で強度最大値の係数が一番大きな値となっており、推定の際に強く影響することが分かる。加えて、覚醒-睡眠、支配-服従、関心-無関心では f_0 レンジの係数が大きく、快-不快、信頼-不信、肯定的-否定的は $f_{aperiodic}$ 平均値の係数が大きい。発話速度については支配-服従でやや係数が大きく、符号が正であることから、速い発話は支配的と推定されるといえる。

また、覚醒-睡眠、支配-服従、関心-無関心に関しては重相関係数が 0.7 以上となり精度が高いことが分かる。一方、快-不快、信頼-不信、肯定的-否定的は 0.5 前後とやや精度が低い。

4.2 推定精度の検証

次に、不特定話者モデルの open データに対する推定精度の検証を行なった。10-fold のクロスバリデーション法による推定値と正解ラベルとの RMS 誤差を Table 2 に示す。表より RMS 誤差は 0.6 前後の値となることが分かる。先行研究 [6] での評価者の同一の発話に対する評価の標準偏差が 0.5 程度になることを考えると、このモデルは比較的精度よく推定できているといえる。

しかし、支配-服従に関しては他の項目に比べ RMS 誤差が大きい。支配-服従の RMS 誤差が大きくなった原因の一つとして、支配-服従は分布が双峰性になり [6]、評価値のばらつきが他の項目に比べて大きいことが考えられる。また、誤差の大きくなる発話を調

* Estimation of speaker's emotional states from conversational utterances. by SATAKE, Tomoyuki, MORI, Hiroki (Utsunomiya University)

Table 1 全話者データでの重回帰分析結果

	快-不快	覚醒-睡眠	支配-服従	信頼-不信	関心-無関心	肯定的-否定的
f_0 レンジ	-0.02	0.21	0.36	-0.07	0.21	-0.10
強度最大値	0.38	0.60	0.48	0.36	0.54	0.30
発話速度	0.07	0.05	0.11	0.11	0.02	0.12
$f_{aperiodic}$ 平均値	0.21	0.14	-0.08	0.20	0.00	0.22
重相関係数	0.52	0.84	0.76	0.48	0.70	0.43

Table 2 予測値と正解ラベルとの RMS 誤差

	快-不快	覚醒-睡眠	支配-服従	信頼-不信	関心-無関心	肯定的-否定的
RMS 誤差	0.61	0.56	0.79	0.66	0.59	0.69

査したところ、評価の際に言語情報の影響を受けたと思われるものが多かった。

4.3 特定話者モデルとの比較

最後に、特定話者モデルとの比較を行なった。ここで特定話者モデルは各話者ごとに重回帰分析を行ない 10-fold のクロスバリデーション法により評価したものとす。また、不特定話者モデルは前述したのではなく対象話者以外の 13 名のデータでモデルを構築し、対象話者のデータを評価したものである。

Fig. 1 に一例として話者 FJK, FMS, FUE のデータを特定話者モデルと不特定話者モデルのそれぞれで求めた推定値と正解ラベルとの相関係数を示す。話者 FMS の肯定的-否定的や話者 FUE の覚醒-睡眠、関心-無関心のように、やや相関係数の差が大きいものもあるが、全体として、同一の話者では 2 つのモデルで相関係数の値に大きな違いは見られない。他の話者に関しても同様の結果となった。

このことから、感情状態の推定は話者による依存が小さく、不特定話者モデルで十分な推定が行なわれているといえる。

5 まとめ

本稿では重回帰分析による不特定話者モデルでの感情状態の推定を試みた。その結果、韻律及び声の気息性に関する音響特徴量を用いることで比較的高い精度での推定が可能であった。また、特定話者モデルと比較することで、不特定話者モデルの妥当性を確認した。

しかし、支配-服従については他の項目に比べ誤差が大きくなった。今後は、新たな音響関連量の調査や言語情報を考慮した分析を行なう予定である。

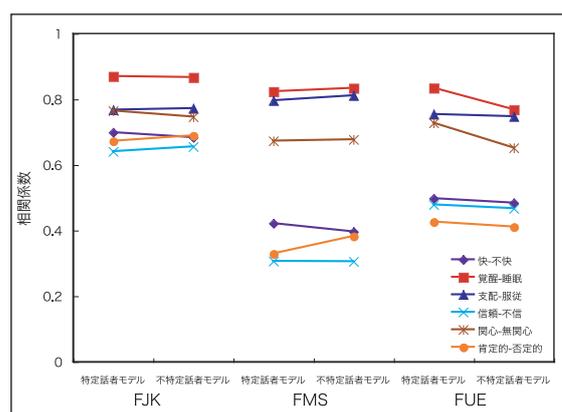


Fig. 1 各モデルでの重相関係数

参考文献

- [1] 吉川哲史, 牧本慎平, 柏岡秀紀, ニックキャンベル, “音響的特徴を用いた機械学習によるノンバーバル発話の意図自動識別の検討,” 2008 年秋季音講論, 1-11-11, 2008.
- [2] 藤江真也, 江尻康, 菊池英明, 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答,” 電子情報通信学会論文誌, Vol. J88-DII, No.3, pp. 489-498, 2005.
- [3] 森大毅, 相田千尋, 粕谷英樹, “活性-評価次元に基づくパラ言語情報ラベルの音響関連量,” 2005 年秋季音講論, 1-6-5, 2005.
- [4] 森大毅, 粕谷英樹, 中村真, “宇都宮大学パラ言語情報研究向け音声対話データベースの構築,” 2007 年秋季音講論, 1-4-10, 2007.
- [5] 大塚貴弘, “ARX 音声生成モデルに基づく音声分析合成法に関する研究,” 宇都宮大学博士論文, 2002.
- [6] 森大毅, 相澤宏, 粕谷英樹, “対話音声のパラ言語情報ラベリングの安定性,” 日本音響学会誌, 61 巻 12 号, pp. 690-697, 2005.