

自然対話中の発話対における音響特徴量に基づく感情の程度推定

松澤 直之[†] 政倉 祐子^{††} 大野 澄雄^{††}

[†] 東京工科大学大学院 バイオ・情報メディア研究科 ^{††} 東京工科大学 コンピュータサイエンス学部

1 はじめに

感情認識において、話者の感情状態を音声から推定するための様々な手法が研究されている。既存研究 [1] では、演技音声を対象として韻律的特徴を用いて発話の時系列から話者の感情状態を推定する手法が提案され、個人差に影響しにくい感情判別が可能であるという結果が得られている。また、韻律的特徴以外に声質などのスペクトル特徴量を音声全体以外に文頭や文末から抽出することで、各感情の特徴を得やすいことが既存研究 [2] により明らかとなった。

我々は普段会話をしているときに単独発話で相手の感情を推定している場合もあるが、先行発話から後続発話はどう変化したのかを自然に読みとってコミュニケーションを行っていると考えられる。そのため、先行研究において単独提示と連続提示によって感情の受容が異なることを調べた [3]。

本研究では重回帰分析による感情の程度推定で、発話全体の特徴量に加えて、文頭文末の局所的特徴量と先行発話の特徴量を推定パラメータに含めることで程度推定の精度にどの程度影響するかを調べた。その結果、文頭文末の特徴量を加えることで各感情の推定精度が向上することと、先行発話の特徴量を加えることでさらに精度を向上できることを確かめた。

2 評価実験

2.1 音声資料

音声は大学生のペアを話者としたゲームプレイ中の自然対話を用いた。ゲームは非対面で Skype を用いて会話し、戦略を立てながら協力して攻略を目指すものを用いた。音声はサンプリング周波数 16kHz、量子化ビット 16bit で録音を行った。評価実験に用いた音声は発話対中の 200ms 以上の無音区間を境界として発話を切り出し、単独発話と、先行/後続発話を結合した連続発話でそれぞれ 420、210 発話用意した。評価実験で用いた各発話の構成を図 1 に示す。

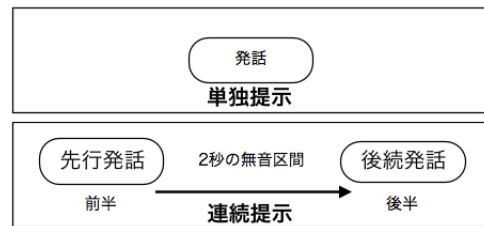


図 1: 評価実験に用いた発話の構成

2.2 評価方法

本研究では“怒り”、“恐れ”、“喜び”、“悲しみ”の 4 感情を扱った。20 代の男子大学生 10 名に音声を単独及び連続提示し、4 つの感情すべてに対してそれぞれ 0(なし)~5(強い)の 6 段階で程度の評価をさせた。

2.3 EWE 法による評価値の統合

得られた感情の程度評価値は、音声試料に対して評価者ごとに評価値のばらつきが異なる。より信頼のできる評価値を導出するために、EWE[4] を用いて評価者に対して重み付け統合を行った。

3 音響特徴量の抽出

今回扱った特徴量を表 1 に示す。特徴量の抽出は音声分析ソフト praat を用いて、発話単位の音声から抽出を行った。 F_0 は semitone を単位とし、話者ごとに平均値がゼロとなるような正規化を行った。また、文頭文末の特徴として文頭及び文末の 500ms の傾斜や平均値などを用いた。

表 1: 扱った音響特徴量

F_{0max}	F_0 の発話内最大値
F_{0min}	F_0 の発話内最小値
F_{0mean}	F_0 の発話内平均値
F_{0sd}	F_0 の発話内標準偏差
$F_{0Lslope}$	F_0 の文頭傾斜
$F_{0Fslope}$	F_0 の文末傾斜
F_{0Lmean}	F_0 の文頭平均値
F_{0Fmean}	F_0 の文末平均値
P_{max}	パワーの発話内最大値
P_{min}	パワーの発話内最小値
P_{mean}	パワーの発話内平均値
P_{sd}	パワーの発話内標準偏差
P_{Lslope}	パワーの文頭傾斜
P_{Fslope}	パワーの文末傾斜
P_{Lmean}	パワーの文頭平均値
P_{Fmean}	パワーの文末平均値
C_{1max}	第 1 次メル周波数ケプストラム係数の発話内最大値
C_{1min}	第 1 次メル周波数ケプストラム係数の発話内最小値
C_{1mean}	第 1 次メル周波数ケプストラム係数の発話内平均値
C_{1sd}	第 1 次メル周波数ケプストラム係数の発話内標準偏差

Estimate the degrees of emotion based on the acoustic features of the speech in natural dialogue

[†] Naoyuki MATSUZAWA (Graduate School of Bionics, Computer and Media Sciences, Tokyo University of Technology)

^{††} Yuko MASAKURA and Sumio OHNO (School of Computer Science, Tokyo University of Technology)

4 重回帰分析による推定

本研究では感情の程度推定を行うため、感情の評価値(単独提示及び連続提示)を目的変数、音響特徴量を説明変数とし、AICによる変数選択を行った上で重回帰分析を行った。また、文頭文末の特徴量と先行発話の特徴量の有無により精度がどの程度変わるかを調べた。推定精度は決定係数 R^2 及び実測値と予測値の残差 RMS、特徴量の標準化偏回帰係数によって評価を行った。そのうちの R^2 を表 2、実測値と文頭文末の特徴量と先行発話の特徴量の両者を加えた場合の予測値の推定結果を図 2、標準化偏回帰係数を図 3 に示す。

5 先行発話の特徴量による影響

単独提示においては先行発話という概念が無いため、本報告では連続提示の後続発話を取り上げる。

表 2 より、文頭文末の局所的な特徴量を加えることで推定結果の精度がわずかに向上し、さらに先行発話の特徴量を加えることで、より精度を向上させることができた。精度は全体的に向上が見られるが、特に“恐れ”と“悲しみ”に関しては両者の特徴量を加えない場合に比べ、決定係数の上昇は顕著であった。

図 4 では推定に用いた特徴量の標準化偏回帰係数を見る事でそれぞれの重みを知ることができる。標準化偏回帰係数の絶対値が高ければ高いほど推定に強く影響しているため、文頭文末の特徴量や先行発話の特徴量が推定精度の向上に大きく関係していると言える。“恐れ”と“悲しみ”においては推定にそれらの特徴量が多く扱われたため、精度が向上したと考えられる。

6 まとめ

本研究では文頭文末の特徴量と先行発話の特徴量が感情の程度推定にどのような影響を与えるかを調べるため、重回帰分析を行った。その結果、文頭文末のような局所的な特徴量と先行発話の特徴量の両者を推定のパラメータに加えることで精度の向上が期待できることを確かめた。今後は、今回取り扱っていない感情においても精度が向上するかなど、さらなる検証が必要であると考えられる。

参考文献

- [1] 多田和彦, 矢野良和, 道木慎二, 大熊繁, “感情遷移における急激な韻律的特徴変化の検出による感情遷移判別方法”, 日本知能情報学会誌, Vol.22, No.1, pp.90-101, 2010.
- [2] Atassi H, Esposito A, Smekal Z, “Analysis of High-level Features for Vocal Emotion Recognition”, Telecommunications and Signal Processing(TSP), 34th International Conference, pp361-366, 2011.
- [3] 松澤直之, 政倉祐子, 大野澄雄, “自然対話中の発話対における単独提示と連続提示の感情受容の比較”, 情報処理学会, 情報処理学会第 74 回全国大会講演論文集, Vol.2, pp.595-596, 2012.
- [4] G.Michael, K.Kristian, “Evaluation of natural emotions using self Assessment manikins”, IEEE Workshop on Automatic Speech Recognition and Understanding, pp.381-385, 2005.

表 2: 各モデルの決定係数 R^2 (連続提示後続発話)

文頭末特徴	先行発話特徴	怒り	恐れ	喜び	悲しみ
なし	なし	0.071	0.008	0.089	0.007
あり	なし	0.079	0.009	0.089	0.007
あり	あり	0.087	0.015	0.101	0.037

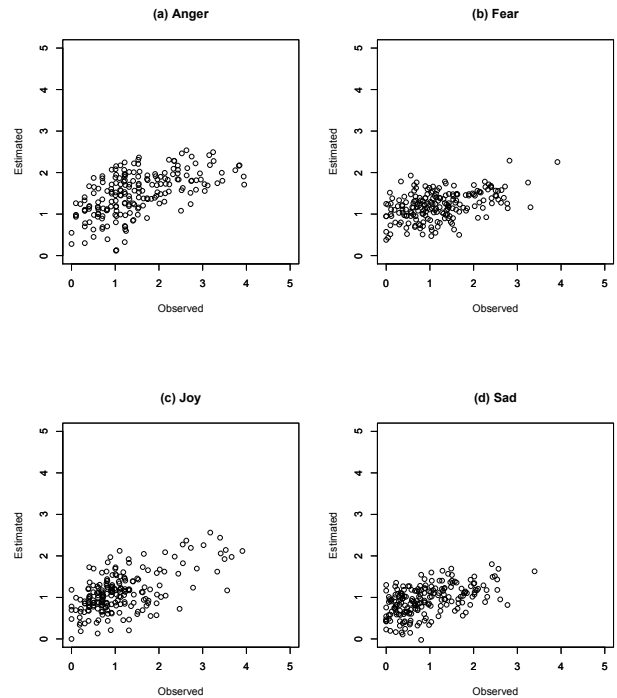


図 2: 重回帰分析による推定結果(連続提示後続発話)

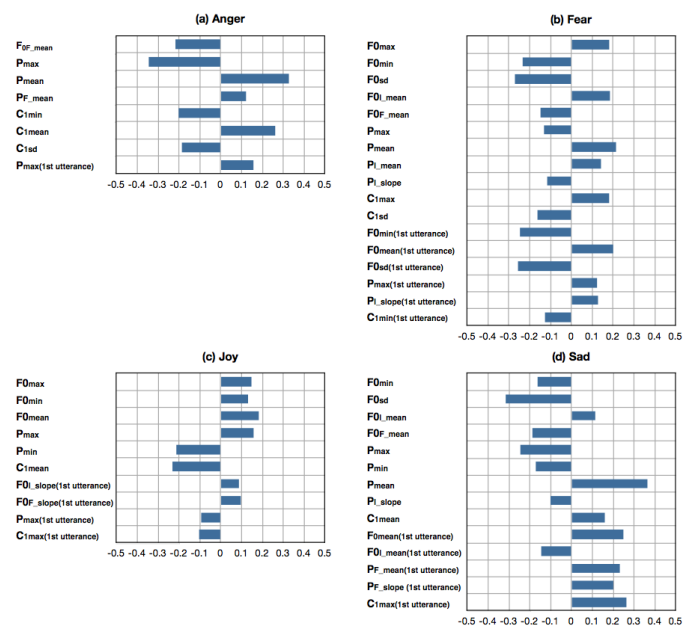


図 3: 各感情の標準化偏回帰係数(連続提示後続発話)