

演技感情音声における演技感情と他者評価の関係の分析*

○目良和也, △黒澤義明, 竹澤寿幸 (広島市立大院)

1 はじめに

近年, 機械学習を用いた音声感情認識や, 感情音声合成に関する研究が盛んに行われている. また多くの感情音声コーパスが公開されており, これらの機械学習の学習データとしてよく用いられている.

感情音声コーパスを用いた機械学習手法を人間とのコミュニケーションシステムに適用する場合, Fig. 1 に示すように音声感情認識でも感情音声合成でも想定される感情ラベルは「相手にとってどのような感情に聞こえるか (他者評価感情)」である.

とはいえ, 全音声データに対して複数の評価者による話者感情アノテーションを行うことは非常にコストがかかる. そのため多くの感情音声コーパスでは, 他者評価感情の代わりに話者が演じた感情 (演技感情) をそのまま感情ラベルとして採用している.

しかし, 演技感情をそのまま感情ラベルとした場合, 演じた通りの感情が常に相手に正しく伝わっていると言えるのだろうか? もし演技感情ラベルが他者評価感情ラベルと大きく異なるのであれば, 演技感情をそのまま他者評価感情とみなして機械学習データに使うことは望ましくない.

そこで本研究では, 演技感情音声における演技感情と他者評価感情の関係について, (1) 各感情における演技感情と他者評価感情の一致度, (2) 演技感情以外のどの感情として認識されやすいかの傾向, (3) 他者評価感情との一致度が高い音声と低い音声における音響的特徴の違い, の3つの観点から分析を行う.

2 広島市立大学感情音声データベース

本研究では, 各演技感情音声に対して演技感情ラベルと他者評価感情ラベルの両方が付与されている広島市立大学感情音声データベース (HCUDB1) を用いて分析を行う.

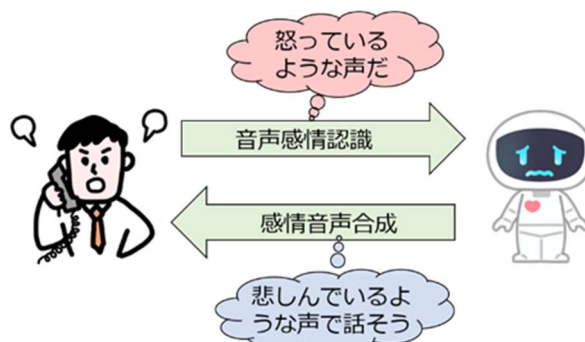


Fig. 1 Perceived emotions in communication

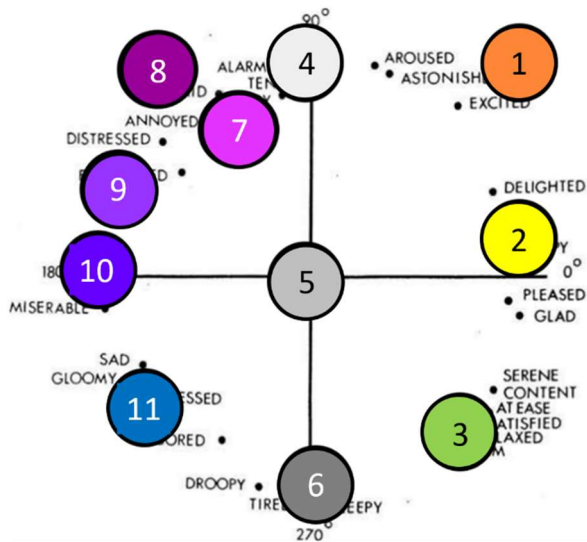
2.1 感情音声データの収録手順

音声提供者は, 声優・俳優・ナレーターなど声を使った演技を行う職業に就いている14名 (男性6名, 女性8名) である. 年齢は20代~60代, 職歴は4年~40年, 出身地は14名中12名が広島県である.

音声提供者は1名ずつ録音スタジオのブースに入り, 事前に指示された10種類のセリフを, それぞれ11種類の感情で演じる. その際, 各セリフと感情の組み合わせにおいて, 可能な範囲で違いをつけて3テイク収録する. 演じてもらう感情は, Russellの感情円環モデル^[1]に基づき“狂喜・楽しい (以降「狂喜」と表記)”, “余裕・嬉しい (嬉しい)”, “リラックス・気楽 (リラックス)”, “驚き”, “冷静”, “眠い・疲れた (疲れた)”, “怒り”, “恐れ”, “嫌い”, “軽蔑”, “憂鬱・悲しい (悲しい)”の11種類である. Fig. 2に感情円環モデルにおける各感情の位置を示す. なお図中の横軸は快不快度, 縦軸は活性度を表す. セリフは, 「そうなんですか」, 「どうなってるの」, 「わかりました」, 「ありがとうございます (ございます)」, 「覚えてますか」, 「全然嬉しくない」, 「それはできない(よ)」, 「気にしないで」, 「仕方ないな」, 「ごめんなさい」の10種類である.

以上の手順により収録された感情音声の総数は, 10セリフ×11感情×3テイク×14話者=4,620件である.

* Analysis of the Relationship between Acting Emotions and Peer Evaluation in Acted-Emotional Speech, by MERA, Kazuya; KUROSAWA, Yoshiaki; and TAKEZAWA, Toshiyuki (Hiroshima City University).



①狂喜, ②嬉しい, ③リラックス, ④驚き, ⑤冷静, ⑥疲れた, ⑦怒り, ⑧恐れ, ⑨嫌い, ⑩軽蔑, ⑪悲しい

Fig. 2 Defined emotion classes on valence and arousal dimensions

2.2 感情ラベルの付与

HCUDB1 では、各感情音声に対して演技感情ラベルと他者評価感情ラベルの両方を付与している。他者評価感情ラベルは 2.1 節の 11 種類の感情のうち話者感情に該当すると判断したものを複数選択可で回答したものである。HCUDB1 における話者感情評価実験の参加者は 21 歳～24 歳の男子大学生 10 名である。

3 演技感情と他者評価の関係の分析

本節では、音声提供者が演じた演技感情が評価者にどのように認識されているかについて分析を行う。本稿では、演技感情が他者にどれだけ伝わっているかの指標として、各演技感情音声に対して“演技感情伝達度（他者評価感情として演技感情を含む選択をした評価者数／全評価者数）”を定義する。

分析を行う前に以下のデータクレンジング処理を行う。

- 「快不快度＝中、活性度＝高」の感情クラス名として、音声収録の際は「驚き」、他者評価の際は「不安・緊張」と異なる名称を提示していたため、今回の分析からは除外する
- (1)を除外した各演技感情音声に対して、演技感情伝達度の平均値が全音声提供者の平均値より 2σ 以上低い音声提供者が 1 名居たため、当該提供者の演技音声を今回の分析から除外する

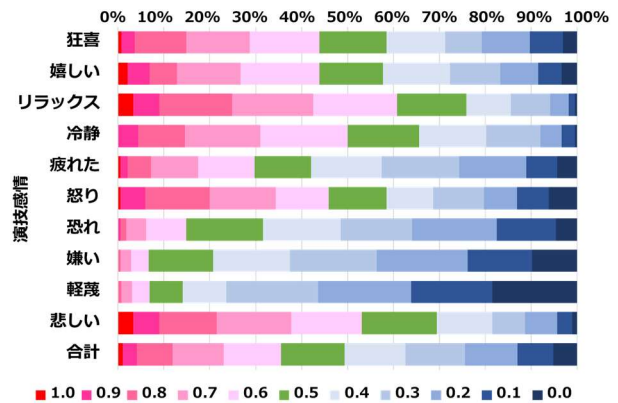


Fig. 3 Degree of agreement between acted emotion and perceived emotion

これらの前処理を行った後、演技感情と他者評価感情との一致傾向および、演技感情と異なる他者評価感情の選択傾向について分析する。

3.1 演技感情と他者評価の一致度

Fig. 3 に各演技感情音声に対する演技感情伝達度の分布を示す。各演技感情の音声の数は、10 セリフ×3 テイク×13 話者＝390 件である。

Fig. 3 より、演技感情伝達度が 0.5 以上の音声の数は全体の 49% (1,927 件) であったことから、一概に演技感情を他者評価感情とみなすのは適切でないといえる。しかし、感情によって演技感情伝達度に大きな差がみられた。感情別に分析すると、リラックス、悲しい、冷静、怒り、狂喜、嬉しいの演技感情音声の半数以上が 0.5 以上の伝達度となっており、比較的相手に正しく演技感情が伝わりやすい感情といえる。一方、嫌い、軽蔑、恐れ、疲れたの演技感情音声は、演じた感情があまり相手に伝わっていない。このことから、日本人に馴染み深い喜怒哀楽と冷静に関連する感情については演技感情と他者評価感情に比較的相関があることが示された。

3.2 演技感情と異なる他者評価感情の傾向

次に、各演技感情音声が演技感情以外のどの感情と認識されるのかについて分析を行った。本節では、特定の感情を表現する音声が他者から評価された際に、演技された感情とは判断されなかったケースを考察対象とした。そして、演技感情とは見なされなかった時に、どのような他の感情が選択されたかについて

分析を行った。

Fig. 4 に、10 種類の感情を演じた音声について、演技感情以外に選択された他者評価感情の割合を示す。HCUDB1 では一つの演技感情音声に対して複数の評価者が評価を行っているため、同じ音声でも評価者によっては演技感情と評価される場合とされない場合がある。そこで他者評価感情のカウントは評価者ごとに行っている。

Fig. 4 で黄緑と灰色の領域が多いことから分かるように、全体的にリラックス（黄色）および冷静（灰色）との誤推定が多くみられた。一方、狂喜、疲れた、恐れ、軽蔑との誤推定は少なかった。

快不快度に注目すると、①狂喜、②嬉しい、③リラックスといった快度が高い感情は、相対的に他の快度の高い感情（狂喜（橙色）、嬉しい（黄色）、リラックス（黄緑））と誤認識されやすい傾向にあった。そして⑥疲れた、⑦怒り、⑩軽蔑といった感情は、快度の低い感情（⑦怒り（桃色）、⑨嫌い（紫色）、⑪悲しい（青緑色））と誤認識されやすい傾向にあった。⑤冷静も⑦怒りと似た誤認識傾向にあるが、これは感情表出を抑えた冷静な演技が cold anger と判断されやすいためと考えられる。なお⑪悲しいは快度が低い感情であるにも関わらず、快度の低い他の感情との誤認識はあまりみられなかった。

次に活性度に注目すると、全体的に活性度の高い感情に誤推定されることは少なかった。また演技感情の活性度と誤推定感情の活性度の間に相関はあまりみられなかった。

3.3 演技感情伝達度による音響特徴量の違い

本節では、各演技感情について、演技感情伝達度が高い音声と低い音声間の音響特徴量の違いについて分析する。なお、今回の分析では、演技感情伝達度が 0.8 以上の音声を演技感情伝達度が高い音声、0.2 以下の音声を演技感情伝達度が低い音声とする。

分析対象とする音響特徴量は、有本らの研究²⁾を参考に、平均基本周波数（Fmean）、基本周波数の発話内標準偏差（Fstdv）および平均発話速度（Drate）の 3 種類とする。Fmean については、声の高さの個人差の影響を除去するため、話者ごとに平均値=0、標準偏差=1

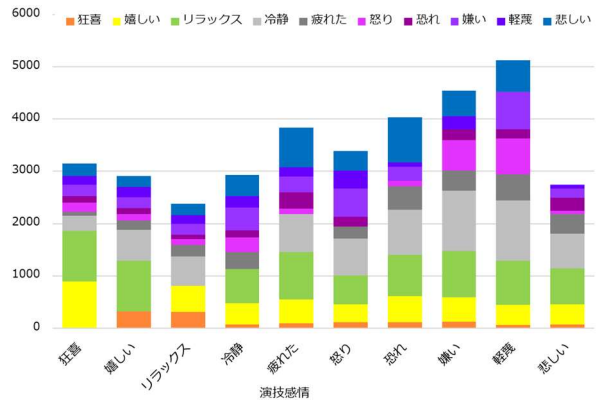


Fig. 4 Trends in misrecognition of acted-emotion

Table 1 Acoustic Features Test: High vs Low Emotion Transmission Groups

	p<0.05 の演技感情
Fmean	狂喜(+), 悲しい(+)*
Fstdv	狂喜(-)*, 嬉しい(-), 怒り(-), 恐れ(-), 悲しい(-)
Drate	怒り(-), 軽蔑(+)

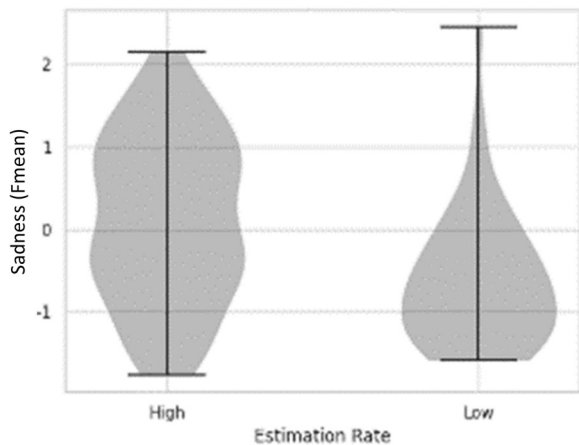
*p<0.05/N

となるよう z-score normalization を適用した。Drate については、音響分析ツール openSMILE 3.0¹⁾で算出される'VoicedSegmentsPerSec' (1 秒あたりの発声領域の数)を話速の代用とした。

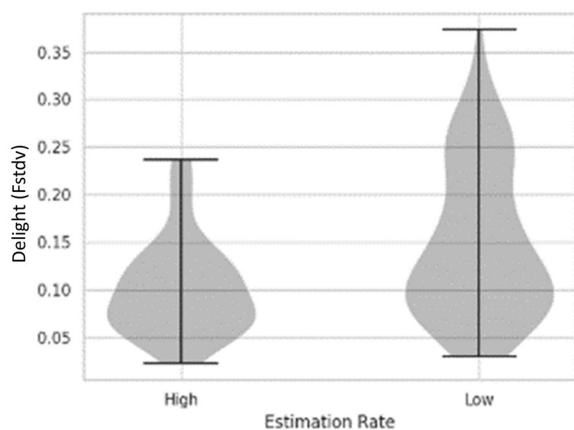
各演技感情音声における Fmean, Fstdv, Drate のうち、有意水準 $\alpha=0.05$ で有意差があるとされた感情クラスを Table 1 に示す。p 値はマンホイットニーの U 検定を用いて算出している。表中の(+)は高群の平均特徴量が高いこと、(-)は高群の平均特徴量が低いことをそれぞれ表す。全体的な傾向として、演技感情伝達度が高い音声は声の高さの平均がより高く、変化がより小さいことが示された。また、話速については感情ごとに傾向が異なり、怒りっぽく聞こえる音声は話速が遅く、軽蔑っぽく聞こえる音声は話速が速いという傾向がみられた。

さらに、今回は有意差検定試行回数が 30 回 (N=10 感情×3 特徴量)であったため、ボンフェローニ補正により有意水準を 0.05/N に変更したところ、有意差があるのは“悲しいの Fmean”と“狂喜の Fstdv の 2 種類の特徴量となった。これらの音響特徴量の分布を Fig. 4 に示す。各図中の左が演技感情伝達度の高い音声、右が演技感情伝達度の低い音声である。Fig. 5(a)より、悲しいの演技感情音声では、

¹ <https://www.audeering.com/research/opensmile/>



(a) Sadness (Fmean)



(b) Delight (Fstdv)

Fig. 5 Differences in the Distribution of Acoustic Features in Acted-Emotional Speech

平均的に高い声の音声は演技感情伝達度が高い傾向にあることがわかる。悲しいの感情は活性度が低いとされており (Fig. 2 参照), 悲しみを演じる際には落ち込んだ暗い声になると想定していたが, 声のトーンに頼らない悲しみの表現が出来ていれば相手に悲しみという感情が伝わると考えられる。

一方 Fig. 5(b)より, 狂喜の演技感情音声では, 声の高さのばらつきが大きい音声は感情伝達度が低い傾向にあることがわかる。こちらは逆に, 活性度の高い狂喜の感情 (Fig.1 参照) を演じる際に, 声の抑揚に頼った演技では狂喜は伝わりにくかったためと考えられる。

4 おわりに

本研究では, 演技感情をそのまま他者評価感情とみなして扱うことに問題が無いか調べるため, 演技感情音声における演技感情と他者評価の関係について分析を行った。

まず, 演技感情伝達度を指標として演技感情ごとに他者評価感情との一致度を比較した。

その結果, 半数以上の評価者に正しく演技感情が伝わった音声は全体の半数程度 (0.49) に留まったが, 日本人になじみ深い喜怒哀楽と冷静については比較的演技感情伝達度が高いことが示された。

次に, 各演技感情音声, 演技感情以外のどの感情として認識されやすいかについての分析を行った。その結果, 快度が高い感情は他の快度が高い感情と, 快度が低い感情は他の快度が低い感情とそれぞれ誤推定されやすい傾向が確認された。活性度については, 演技感情と誤推定される感情間の相関はみられなかった。

さらに, 演技感情伝達度が高い音声と低い音声間で, Fmean, Fstdv, Drate の3種類の音響特徴量の比較を行った。その結果, 演技感情伝達度が高い音声は, 声の高さの平均がより高く, 変化がより小さいことが示された。

今回は音声提供者, 評価者ともに日本人であったため, 日本人に馴染みの深い「喜怒哀楽」が表現しやすく, また判別しやすかった可能性が考えられる。今後より一般的な分析を行うには多様な評価者による他者評価結果について分析する必要がある。さらに, より多くの音響特徴量による比較や, 話者の属性による違いについても分析する予定である。

謝辞

本研究は国立研究開発法人科学技術振興機構(JST)の研究成果展開事業「センター・オブ・イノベーション(COI)プログラム」の助成を受けています。

参考文献

- [1] Russell *et al.*, Journal of Personality and Social Psychology 39, 1161-1178, 1980.
- [2] 有本, 河津, 音講論 (秋), 385-388, 2013.