

音声の印象評定を用いた感情強度推定モデル性能の検討

背景・目的

- 発話内容に依らない音声のみからの感情強度推定は研究段階
- 先行研究[1]: 深層学習を用いた ①個別推定 ②並列型推定 ③直列型推定の3つのモデルを作成し、直列型において平均正解率 57.1%の結果を得た
- この先行研究ではカテゴリとして感情強度分類を行っているが、回帰による感情強度推定の方が柔軟な表現が可能であり、適していると考えられる

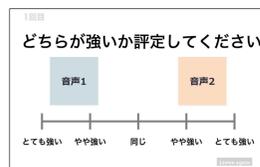
- 目的**
- 感情知覚評定実験を通して算出した各感情強度カテゴリの知覚強度値の、感情強度カテゴリによる違いの検討
 - 知覚強度値を利用した回帰モデルによる感情強度推定可能性の検討

手法

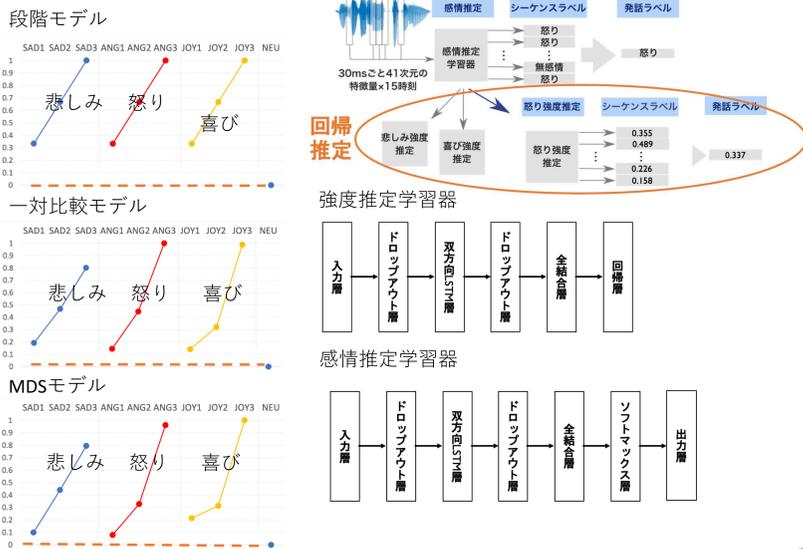
- 音声データ**
- 感情評定値付きオンラインゲーム音声チャットコーパス (OGVC)
 - ゲーム中の対話を声優4人が「怒り」「悲しみ」「喜び」の3感情について「無感情」の強度0を含む4段階の強度で収録 (NII提供)
 - 感情強度数: 3感情(怒り, 喜び, 悲しみ) × 3強度 + 無感情の10種 (計991発話)

- 特徴量**
- 生の音声データでなく、30 ms (オーバーラップ15 ms) ごと41次元の特徴量を抽出 (括弧内の数字は次元数)
 - ガンマトーンケプストラム係数(13)、ガンマトーンケプストラム係数の微分(13)、メル周波数ケプストラム係数(13)、Spectral Crest(1)、基本周波数(1)

- 感情強度評価実験**
- 2つの音声を聴きどちらの音声の感情が強いかを5段階で評価
 - 20発話 (男性女性それぞれによる10発話ずつ) を用いて同じ発話同士の比較を含んだ20 × 20 = 400の比較対を評価
 - 実験結果からScheffeの一対比較法 (浦の変法)、多次元尺度構成法 (MDS) により算出した2種の知覚強度値を0~1の値に正規化し推定モデルに使用

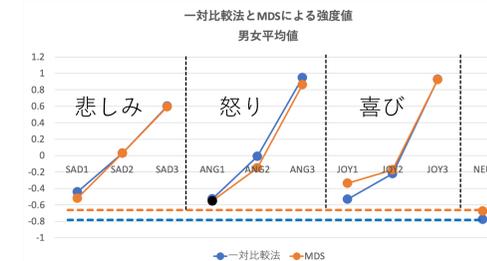


- 回帰推定モデル**
- 3感情で別の強度推定学習器を作成
 - 感情推定により4感情いずれかに分類し、その後感情ごとの強度回帰推定学習器で強度推定
 - 3つのラベル強度値を設定
 - 隣り合う強度カテゴリの強度値差が等しくなるよう設定 (段階モデル)
 - Scheffeの一対比較法により得られた知覚強度値を正規化 (一対比較モデル)
 - 多次元尺度構成法 (MDS) により得られた知覚強度値を正規化した値 (MDSモデル)



結果

- 強度値算出結果**
- 算出した2種の知覚強度値の相関係数は0.99となり、強い相関を持つ
 - 一対比較法による強度値と相関係数の高い音響特徴量を調べるためOpenSmileを用いて88の音響特徴量 (eGeMAPSv02[2]) を抽出し、強度値と音響特徴量の相関係数が0.8以上の音響特徴量を表示
 - loudnessに関する音響特徴量が全体の半分を占め、loudnessと感情強度値に関連性があると考えられる



音響特徴量	相関係数
loudness sma3 amean	0.906
loudness sma3 percentile50.0	0.830
loudness sma3 percentile80.0	0.878
loudness sma3 pctlrange0-2	0.874
logRelF0-H1-A3 sma3nz amean	-0.800
mfcc1V sma3nz amean	-0.835
equivalentSoundLevel dBp	0.889

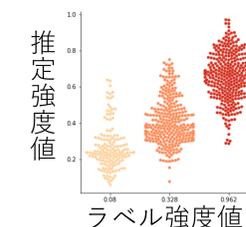
知覚強度値を用いた感情強度推定

- 各感情強度の設定強度値を以下3つのように設定した3つのモデル
- ①段階モデル ②一対比較モデル ③MDSモデル
- いずれのモデルにおいても怒り, 悲しみ, 喜びの順に高い相関係数
- 平均相関係数は0.73→回帰による感情強度推定が可能であると考えられる
- 感情ごとに高い相関係数を持つモデルは異なる

	段階	一対比較	MDS
怒り	0.77	0.78	0.77
悲しみ	0.72	0.72	0.74
喜び	0.69	0.67	0.66
平均	0.73	0.73	0.73

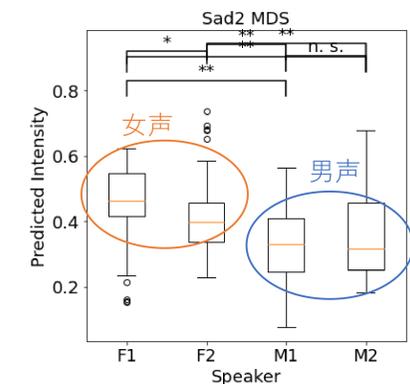
各感情, 設定強度値ごとのラベル強度値と推定強度値の相関係数

怒り強度値, MDS (r = 0.77)



考察

- 強度値推定結果の音声話者による違いの検討**
- ウェルチのt検定を用いて悲しみ2のMDSモデルによる推定強度値の話者間での違いを調べた (*p<0.05, **p<0.01)
 - どのモデルにおいても女性話者の音声の方が有意に大きな推定強度値→悲しみの感情は女性話者の音声の方が強く表出しやすい
 - 5回分の実験による推定強度値を音声話者ごと, 各感情強度カテゴリごとに分け, 話者間での違いがあるかウェルチのt検定を用いて調査→「喜び3」以外の感情強度においては推定強度値の平均値が音声話者により有意に異なる話者のペアが存在



- 回帰モデルによる強度値推定を行うことにより推定強度値の分布が得られ, カテゴリ分類よりも柔軟に感情強度を表現可能なモデルが作成可能

References

- [1]川瀬恵, 中山実, “音声感情強度推定のための感情強度ラベルシーケンス分析”, 信学技報, vol. 121, no. 100, HIP2021-20, pp. 17-22, 2021年7月.
[2] The Geneva Minimalistic Acoustic Parameter Set (GeMAPS) for Voice Research and Affective Computing

今後の展開

強度値と音響特徴量 (loudness) の関係及び音響特徴量 (loudness) による強度値の推定可能性についての検討