

## クロスコーパス感情ラベリングによる対話音声の比較\*

○有本泰子 (理研), 森大毅 (宇都宮大)

### 1 はじめに

大規模音声コーパスを必要とする音声認識や感情認識などの人間と機械との対話の円滑化を念頭にした研究では、複数のコーパスを併用することが求められる。しかし、コーパスごとに独自基準で感情ラベリングを行っているため、感情ラベルをコーパス間で等価とみなすことができず、現状では複数コーパスの併用は不可能である。複数コーパスを用いて、音声コーパス間で共通した感情ラベルを基準として研究を進めるためには、音声に対する感情ラベルの標準化が求められる。音声に表出される感情を表現する適切なラベルを規定することを念頭に、まずは既存のふたつの感情音声コーパス（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) [1] および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 (UADB) [2]) に付与されている異なる基準のラベルを相互に付与した。ラベルを付与した結果、対話ドメインの異なるコーパス間および異なる感情ラベル間にどのような関係があるかを検証するとともに、音響的な差についても検証した。

### 2 クロスコーパス感情ラベリング

#### 2.1 ラベリング手法

本研究で用いた感情ラベルはOGVCで採用している感情カテゴリラベル（喜び (JOY), 受容 (ACC), 恐れ (FEA), 驚き (SUR), 悲しみ (SAD), 嫌悪 (DIS), 怒り (ANG), 期待 (ANT), その他 (OTH), 平静 (NEU)) とUADBで採用している7段階 (1-7) の感情次元ラベル（不快-快 (pleasantness), 睡眠-覚醒 (arousal), 服従-支配 (dominance), 不信-信頼 (credibility), 無関心-関心 (interest), 否定的-肯定的 (positivity)) の二つである。この二つの感情ラベルをふたつのコーパスに相互に付与した。安定した感情ラベルを得るため、まずは被験者スクリーニングのための聴取実験を行った。実験参加者はコーパスの話者と同年代の10名（男性4名, 女性6名, 平均年齢: 21.8歳 ( $SD = 0.84$ )) である。評価対象音声は108発話 (OGVC: 9感情 × 3段階 (弱・中・強) × 各2発話 = 54発話, UADB: 3段階 (弱・中・強) × 各18発話 = 54発話) である。一人の被験者に対し、発話の提示順序を変えて8回試行を行った。

被験者スクリーニングにより、3名（男性1名, 女性2名）のラベラーを選定し、本実験を行った。ラベラーの平均年齢は22歳 ( $SD = 0.82$ ) である。評価対象の音声は11,418発話 (OGVC: 6578発話, UADB: 4840発話) で、聴取回数は1回とした。付与したラベルは

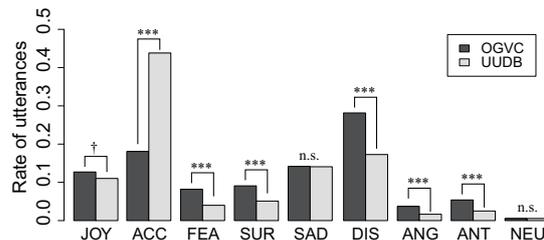


Fig. 1 Result of cross-corpus emotional category labeling

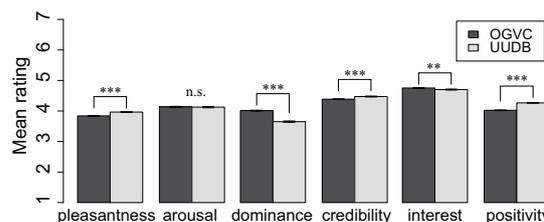


Fig. 2 Result of cross-corpus emotional dimension labeling

被験者スクリーニングと同じである。

#### 2.2 結果と考察

Fig. 1は感情カテゴリを3名のラベラーに評価させた結果である。3名のラベラーのうち2名以上が同じ感情と判定した発話の数を、コーパス内の比率に換算し、その差を検定したものである。SADおよびNEUの発話はコーパス間で有意な差を示さなかったが、それ以外の7感情の発話は有意差あるいは有意傾向を見せ、異なる対話ドメインで収集された発話に感情の偏りがあることを示唆している。特に、UADBはACCの発話が非常に多い。UADBが話者間で課題解決を行う対話であるため、対話相手の発言を受容する発話が多くなったのではないかと考える。一方で、OGVCはFEA・DIS・ANGの発話がUADBより多く含まれる。OGVCでは、対話中の注意をオンラインゲームに向けさせているため、社会的な場面で表出を控えているネガティブ感情の制御が弱まったと考える。

Fig. 2は感情次元の評価結果である。3名のラベラーが付与した感情次元の平均値を各発話の感情次元強度とみなし、感情次元ごとにコーパス間の平均値の差を検定したものである。arousal以外の感情で、その強度にコーパス間で差が見られた。OGVCは不快傾向、支配傾向、不信傾向、高い関心度、否定的な傾向が見られ、感情カテゴリ評価の結果で示唆されたネガティブ感情の発話が多い傾向と一致する。UADBは快傾向、服従傾向、高い信頼度、中程度の関心度、肯定的な傾向がみられた。

\* Perceptive and acoustical comparison of dialog speech by cross-corpus emotional labeling. by ARIMOTO, Yoshiko (RIKEN BSI), MORI, Hiroki (Utsunomiya university)

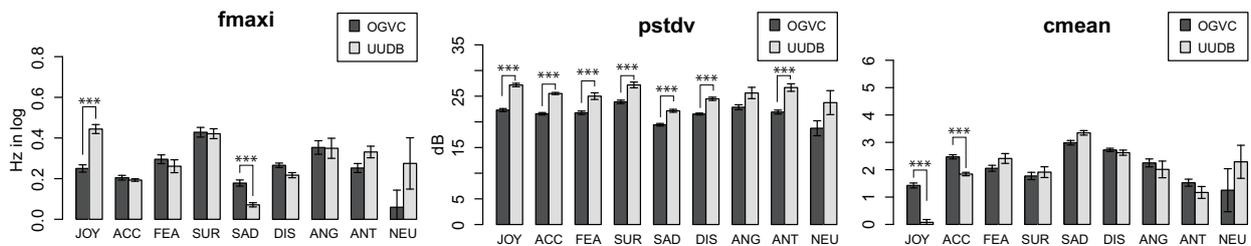


Fig. 3 Mean acoustic features (fmaxi, pstdv, and cmean) in emotional categories for each corpus

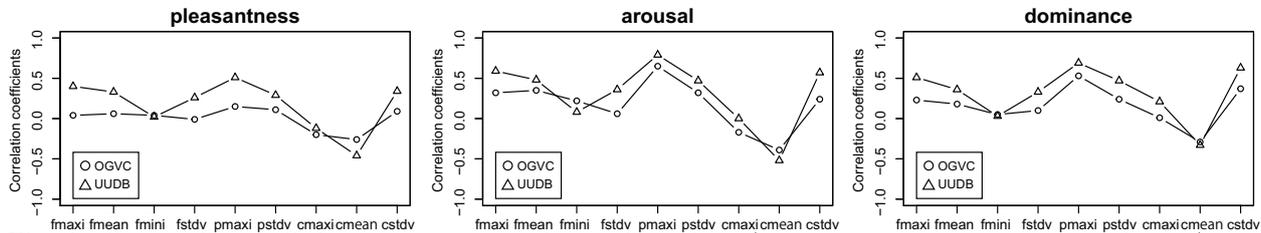


Fig. 4 Correlation coefficients between emotional dimensions (pleasantness, arousal, and dominance) and acoustic features for each corpus

### 3 音響分析によるコーパス間比較

#### 3.1 特徴量抽出と分析手法

音響特徴量はこれまでの研究 [1] を参考に、基本周波数 ( $F_0$ )・音圧・ケプストラムの 1 次係数 ( $C_1$ ) の発話内基本統計量を用いた。 $F_0$  に関する特徴量は発話内の最高値 (fmaxi)・平均値 (fmean)・最低値 (fmimi)・標準偏差 (fstdv)、音圧に関する特徴量は発話内の最大値 (pmaxi)・標準偏差 (pstdv)、 $C_1$  に関する特徴量は発話内の最高値 (cmaxi)・平均値 (cmean)・標準偏差 (cstdv) である。 $F_0$  は対数に変換し、話者ごとに発話内平均  $F_0$  を求め、各発話の  $F_0$  値から差し引くことで話者正規化を行った。 $C_1$  は  $F_0$  抽出時に算出される発話内の有声・無声の判定結果を参考に、有声区間のみを用いて特徴量を計算した。

各特徴量に対し、二元配置分散分析 (コーパス × 感情カテゴリ) を行うとともに、感情次元ラベルとの相関係数をコーパスごとに求め、その差を検証した。

#### 3.2 結果と考察

Fig. 3 に感情カテゴリごとの音響特徴量のコーパス内平均値を示す (fmaxi・pstdv・cmean のみ)。fmaxi・pstdv・cmean のいずれも、コーパスと感情カテゴリによる主効果および交互作用が有意となった。さらに、Tukey の多重比較によりその差を検証した。感情間の差については、JOY・SUR・ANG は感情を含まない NEU よりも fmaxi の値が高く、JOY・SUR は NEU よりも pstdv の値が大きく、SAD は NEU よりも cmean の値が大きいことが示された。コーパス間の差については、fmaxi は JOY と SAD で、pstdv は ANG と NEU 以外の多くの感情で、cmean は JOY と ACC で有意な差を認めた。対話ドメインが異なると、喜びと悲しみにおける声の高さの表現も異なること、喜びと受容の声質の表現が異なることが示された。

Fig. 4 に各コーパスの感情次元と各特徴量との相関係数を示す (pleasantness・arousal・dominance のみ)。pleasantness においては、コーパス間で相関係数の値に差があることが分かる。OGVC (図中の白丸) の特

徴量と pleasantness との相関係数は多くが 0 付近であり、音響特徴量と OGVC の pleasantness との相関は強くないことが示される。UUDB は fmaxi・fmean・pmaxi・pstdv・fmean・cstdv などが相関係数の絶対値が 0.5 付近であり、pleasantness との中程度の相関があることが示される。arousal および dominance では多くの特徴量でコーパスごとの相関係数が同じ傾向を示している。pleasantness と同様の傾向は credibility に、arousal および positivity と同様の傾向は interest にも見られた。positivity と各特徴量との相関係数もコーパス間で同じ傾向がみられるが、相関係数の絶対値が小さく、関係は強くないと考えられる。

### 4 おわりに

音声に表出される感情を表現する適切なラベルを規定し、音声に対する感情ラベルを標準化すること念頭に、まずは既存のふたつの感情音声コーパスに付与されている異なる基準のラベルを相互に付与し、コーパス間の感情ラベルおよび音響特徴量の差について検証した。その結果、異なるコーパス間で付与される感情カテゴリおよび感情次元の傾向に差があること、同じ感情ラベルが付与された発話でもコーパス間で音響特徴量に差があることが明らかとなった。今後は、感情カテゴリと感情次元間との関係について精査しながら、適切な感情ラベルの規定する手法について検討する。

謝辞 本研究は財団法人立石科学技術振興財団研究助成 (A) および JSPS 科研費 (26280100) の助成を受けた。

### 参考文献

- [1] Arimoto *et al.*, Acoust. Sci. Tech., 33 (6), 359–369, 2012.
- [2] Mori *et al.*, Speech Commun., 53(1), pp. 36–50, 2011.