

# クロスコーパスラベリングによる感情空間への 感情カテゴリのマッピング

–音声知覚および音響分析による検討–

有本 泰子<sup>†</sup> 森 大毅<sup>††</sup>

<sup>†</sup> 芝浦工業大学工学部 〒337-8570 埼玉県さいたま市見沼区深作 307

<sup>††</sup> 宇都宮大学大学院工学研究科 〒321-8585 宇都宮市陽東7丁目1-2

E-mail: <sup>†</sup>ar@shibaura-it.ac.jp, <sup>††</sup>hiroki@speech-lab.org

あらまし 感情研究で用いられる感情理論には、主に感情カテゴリと感情次元の二種類がある。音声に含まれる感情については未整備であり、音声で知覚可能な感情カテゴリおよび感情次元の関係は明らかにされていない。本報告では、音声から知覚可能な感情について整備し、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）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（UUDB））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、音声から知覚される感情カテゴリと感情次元との関係を示すとともに、感情の知覚に利用される音声の音響的特徴との関係を明らかにする。

キーワード クロスコーパス感情ラベリング, 感情音声, 感情知覚, 音響分析

## Emotion category mapping to emotional space by cross-corpus labelling

– Psychological and acoustical examination of emotion perception –

Yoshiko ARIMOTO<sup>†</sup> and Hiroki MORI<sup>††</sup>

<sup>†</sup> College of Engineering, Shibaura Institute of Technology Fukasaku 307, Minuma-ku, Saitama-shi,  
337-8570 Japan

<sup>††</sup> Graduate School of Engineering, Utsunomiya University Yoto 7-1-2, Utsunomiya-shi, 321-8585 Japan

E-mail: <sup>†</sup>ar@shibaura-it.ac.jp, <sup>††</sup>hiroki@speech-lab.org

**Abstract** Psychological classification of emotion has two main viewpoints. One is emotion category that emotion is classified into discrete and fundamental groups and the other is emotion dimension that emotions is characterized with multiple continuous scales. Focusing on emotion that is expressed and perceived by speech, classification of emotion is not sufficiently studied and discussed and relationship between emotion category and emotion dimension presented in speech is not demonstrated. Aiming to establish the emotion label to present emotional speech, this report elucidated the relationship between the emotion category and the emotion dimension presented in speech by conducting two experiments with the result of cross-corpus emotion labelling with two Japanese dialogue corpora (Online Gaming Voice Chat Corpus with Emotional Label (OGVC) and Utsunomiya University Spoken Dialogue Database for Paralinguistic Information Studies (UUDB)). The first experiment revealed that the relationship between the perceptive emotion category and the emotion dimension. The second experiment revealed that the relationship between acoustical information of emotional speech and either the perceptive emotional category or the emotional dimension.

**Key words** Cross corpus emotion labelling, emotional speech, emotion perception, acoustical analysis

## 1. ま え が き

大規模音声コーパスを必要とする感情認識などの研究では、感情ラベルが付与された複数の音声コーパスを使用することが求められる。しかし、感情を対象とした音声の研究では、研究ごとに独自基準の感情ラベルが音声に付与されているため、そのラベルをコーパス間で共通のものとしなすことはできず、コーパスを併用することはできない。また、各音声コーパスに付与されている感情ラベルの種類が大きく異なる場合もある。感情ラベルに用いられる感情の定義としては大きく分けて二種類ある。ひとつは Ekman が提案する基本 6 感情 [1] や Plutchik が提案する基本 8 感情 [2] などの喜びや悲しみといった感情カテゴリであり、もうひとつは Russell などが提案する快-不快や覚醒-睡眠などの感情次元 [3] である。一方の音声コーパスには感情カテゴリ、もう一方の音声コーパスには感情次元が付与されていれば、これらのコーパスを併用することができないのは明らかである。また、たとえ同じ感情ラベルが付与された音声コーパスがあったとしても、ラベリング作業は個々のコーパス内で閉じているため、そのコーパス内での相対的な判断によって付与された感情ラベルとなってしまう、コーパス間の感情ラベルを等価としてみなすことができない。感情音声コーパスを自発対話音声から収集して作成した場合には、これが顕著となる可能性がある。コーパス間で対話ドメインが異なるので、表出する感情の種類も異なる可能性があるからである。

感情ラベルを等価とみなせない問題を抱えつつも、複数の感情音声コーパスを用いた感情認識や感情音声合成を試みる研究が行われている [4]~[9]。Schuller らの研究では、8 つの異なる感情音声コーパスを利用している [6]~[9]。付与されているラベルは感情カテゴリであったり、感情次元であったり、その種類もバラバラである。コーパスに付与されていたこれらの感情ラベルを、直行するふたつの感情次元（快-不快軸と覚醒-睡眠軸）の 4 象限のいずれかに強制的に当てはめ、それを正解として実験を行っている。このような方法では、コーパス間のラベルが等価であることは保証されない。また、Zong らは、4 つの異なる感情音声コーパスのうち、感情カテゴリラベルがコーパス間で同一の発話のみ抽出し、感情認識研究に利用している [4] が、この方法でも、コーパス間のラベルの等価性は保証されないし、コーパスに含まれるすべての発話を利用することはできない。音声コーパスに対する感情ラベルを共通化することが求められる。

本研究では、音声コーパスに対する感情ラベルを共通化することをめざし、まずは既存コーパスに付与されている感情ラベルの関係性について検証する。まずは、公開されているふたつの感情音声コーパスを用いて、相互に感情ラベルを付与するクロスコーパス感情ラベリングを行う。これらのコーパスを用いて、感情カテゴリと感情次元の関係を音声知覚の観点から、検証するため、感情次元で構成される感情空間上に配置される各感情カテゴリの分布の独立性を、尤度比検定によって検証する。さらに、コーパスに含まれる音声の音響的特徴量から、音響特徴のみで構成される次元を抽出し、それが感情ラベルとどのよ

うな関係があるかを相関係数および尤度比検定によって検証を行う。

以下、第 2 節では、使用する音声コーパスについて、第 3 節ではクロスコーパス感情ラベリングの手法について説明する。第 4 節では感情ラベルを用いた感情次元および感情カテゴリの関係について検証した結果について、第 5 節では音響特徴量から抽出した次元と感情ラベルとの関係について検証した結果について述べ、第 6 節で本研究についてまとめる。

## 2. 音声コーパス

現在、日本国内で公開されている自発対話音声コーパスには、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) [10] と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 (UUDB) [11] のふたつがある。本研究では、このふたつのコーパスを使用する。

### 2.1 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)

OGVC には、自発対話対話とこの対話に基づいた演技対話のふたつのコーパスが存在する。本論文では自発対話音声のコーパスを使用する。このコーパスは、Massively Multi-player Online Role-Playing Game (MMORPG) を利用することで感情を喚起している。別室に配置された対話者は MMORPG にともに参加し、音声チャットを通じて対話を行う。このパートには、13 名の日本大学生（女性 4 名、男性 9 名）の対話、全 9114 発話が収録されている。このうち、音声の振幅レベルが小さい、または、音響的分析に影響をおよぼす転記用タグが付与されている発話を除いた 6578 発話に対して、感情ラベルが付与されている。付与されているラベルはプルチックの基本 8 感情（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)) と、感情が表出していない平静 (NEU)、これらのどの感情にも分類できないその他 (OTH) を加えた 10 種類である。

### 2.2 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 (UUDB)

UUDB は自発音声対話を収録したコーパスである。「4 コマまんが並べ換え課題」を与えられた日本人大学生 7 ペア（女性 12 名、男性 2 名）の対話、全 4840 発話が収録されている。この課題は、対話者が、バラバラにされた 4 コマまんがをそれぞれ 2 コマずつ持ち、対話により相手の持っている 2 コマと自身の持つ 2 コマを合わせた本来の正しい順番を決定する。この時、対話者は相手の持つ 2 コマと音声以外の情報を知ることができない。UUDB の全発話には、以下の 6 感情次元で発話者の感情ラベルが付与されている。

- (1) 快-不快 (pleasant-unpleasant)
- (2) 覚醒-睡眠 (aroused-sleepy)
- (3) 支配-服従 (dominant-submissive)
- (4) 信頼-不信 (credible-doubtful)
- (5) 関心-無関心 (interested-indifferent)
- (6) 肯定的-否定的 (positive-negative)

表 1 スクリーニングテストにおける評価の一致率と相関係数

| Evaluator | agreement | $r$  |
|-----------|-----------|------|
| E01       | 0.42      | 0.82 |
| E02       | 0.52      | 0.89 |
| E03       | 0.44      | 0.85 |
| E04       | 0.60      | 0.90 |
| E05       | 0.50      | 0.86 |
| E06       | 0.53      | 0.83 |
| E07       | 0.51      | 0.81 |
| E08       | 0.42      | 0.83 |
| E09       | 0.36      | 0.59 |
| E10       | 0.48      | 0.87 |

### 3. クロスコーパス感情ラベリング

OGVC および UADB に付与されている感情ラベルは、2. 節で説明した通り、一方が感情カテゴリ、もう一方が感情次元となっており、共通に扱うことはできない。したがって、本研究ではコーパスに付与されている感情ラベルを使用せず、感情カテゴリおよび感情次元のラベルを新たに二つのコーパスに付与しなおし、これを用いることとした。

#### 3.1 ラベラースクリーニング

安定した感情ラベルを付与するラベラーを 3 名選定するため、スクリーニングテストを聴取実験により行った。評価者は、コーパスの話者と同年代とするため、大学生および大学院生 10 名（男性 6 名、女性 2 名、平均年齢 21.8 歳 ( $SD=0.84$ )) とした。評価対象音声は、OGVC が 54 発話（9 感情  $\times$  3 段階（弱・中・強） $\times$  各 2 発話）、UADB が 54 発話（感情強度 3 段階（弱・中・強） $\times$  各 18 発話）の 108 発話である。この 108 発話をひとつのブロックとし、発話の提示順序をランダムに変えて 8 回試行を行った。評価内容は配布されているコーパスに付属している感情ラベルと同様とした。評価者に、10 種の感情ラベルを強制選択でひとつ選択させたのち、その強度（5 段階（1~5））を評価させ、さらに、6 種の感情次元ラベル（7 段階（1~7））を評価させた。

[11] を参考に、評価者の評価の一貫性を示す指標として評価者内標準偏差を求めた。さらに、感情音声に対して同じ感性をもつ評価者を選ぶため、感情カテゴリ評価では各発話の正解ラベルを 10 名の評価者の多数決により決定し、正解ラベルと各評価者のラベルとの一致率を求めた。また、感情次元評価では、各次元ごとに全評価者のラベルの平均値を求めて正解ラベルとし、正解ラベルと各評価者のラベルの相関係数を求めた。その結果を表 1 に示す。

評価者内標準偏差を求めた結果、基準値 ( $SD<0.8$ ) を下回る標準偏差を示した評価者は 4 名（E04, E08, E09, E10）となった。この 4 名の評価者のうち、感情カテゴリ評価の一致率が高く、感情次元評価の相関係数が高い E04, E08, E10 を本実験のラベラーとして採用した。

#### 3.2 本実験

評価者スクリーニングにより選定した 3 名のラベラー（男性 1 名、女性 2 名）に本実験に参加させた、ラベラーの平均年

表 2 感情カテゴリラベリング後の発話数.

| Emotion | OGVC | UADB | Total |
|---------|------|------|-------|
| JOY     | 438  | 259  | 697   |
| ACC     | 623  | 1030 | 1653  |
| FEA     | 282  | 94   | 376   |
| SUR     | 313  | 120  | 433   |
| SAD     | 488  | 331  | 819   |
| DIS     | 970  | 406  | 1376  |
| ANG     | 128  | 39   | 167   |
| ANT     | 186  | 59   | 245   |
| NEU     | 18   | 13   | 31    |
| Total   | 3446 | 2351 | 5797  |

表 3 感情次元間の相関係数.

|              | arousal | dominance | credibility | interest | positivity |
|--------------|---------|-----------|-------------|----------|------------|
| pleasantness | 0.46    | 0.27      | 0.72        | 0.56     | 0.66       |
| arousal      |         | 0.70      | 0.45        | 0.69     | 0.22       |
| dominance    |         |           | 0.32        | 0.60     | -0.01      |
| credibility  |         |           |             | 0.65     | 0.72       |
| interest     |         |           |             |          | 0.38       |

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、試行回数は 1 回とした。付与したラベルは評価者スクリーニングと同じである。

感情カテゴリラベリングの結果、3 名のラベラーのうち 2 名が一致した感情をその発話の感情カテゴリラベルとみなす。各感情カテゴリの発話数を表 2 に示す。また、感情次元ラベリングの結果、ラベラー 3 名のラベルの平均値を発話ごとに求めて、各発話の感情次元強度とした。表 3 に感情次元間の相関係数を示す。

### 4. 実験 1: 感情次元空間への感情カテゴリのマッピング

#### 4.1 分析手法

各発話の感情次元 (pleasantness, arousal, dominance) を確率変数とし、この変数が三次元混合正規分布 (GMM) に従うと仮定する。感情カテゴリが異なる発話が異なる分布に従うか否かを検証するため、2 つの感情カテゴリ  $E_1, E_2$  の発話の感情次元の分布の同一性を、以下の帰無仮説  $H_0$  により検定する。

$H_0$ : この変数は感情カテゴリに依存しない GMM  $M_0$  から生成される

$H_1$ : この変数は感情カテゴリが  $E_1$  のときは GMM  $M_1$  から、感情カテゴリが  $E_2$  のときは GMM  $M_2$  から生成される

$H_0$  と  $H_1$  を尤度比検定により比較する。 $H_0$  を仮定したとき、感情カテゴリに依存しないモデル  $M'_0$  の最大対数尤度から感情カテゴリに依存するモデル  $M'_1, M'_2$  の最大対数尤度の和を引いた検定統計量 (逸脱度の差) の分布は、パラメトリックブートストラップ法により求められる。実際のデータの逸脱度を求め、それがパラメトリックブートストラップ法で求めた棄却域 ( $\alpha = 5\%$ ) に入るならば、 $E_1$  および  $E_2$  の感情次元は同一の

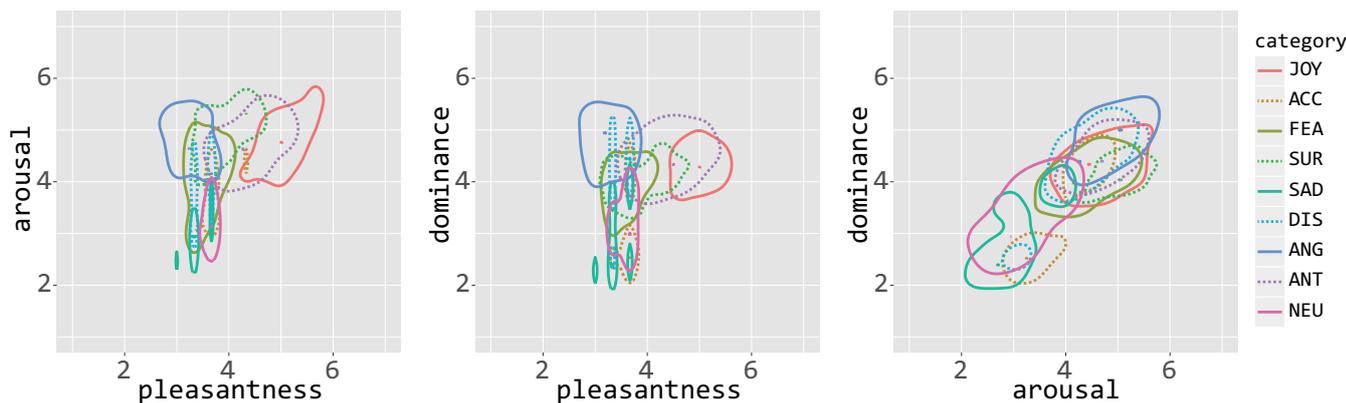


図1 感情次元で構成される二次元空間上の感情カテゴリの分布

表4 感情次元による三次元空間内の感情カテゴリの逸脱度の差

|     | ACC     | FEA    | SUR    | SAD     | DIS     | ANG    | ANT    | NEU    |
|-----|---------|--------|--------|---------|---------|--------|--------|--------|
| JOY | 1187.3* | 762.7* | 680.1* | 1406.7* | 1660.8* | 679.6* | 178.0* | 159.1* |
| ACC |         | 501.8* | 585.3* | 1248.5* | 1169.5* | 765.3* | 368.4* | 367.2* |
| FEA |         |        | 98.3*  | 342.9*  | 123.7*  | 215.2* | 268.4* | 41.7*  |
| SUR |         |        |        | 802.0*  | 482.7*  | 287.7* | 253.2* | 31.0   |
| SAD |         |        |        |         | 534.4*  | 603.8* | 678.8* | 38.2   |
| DIS |         |        |        |         |         | 108.6* | 463.0* | 11.6   |
| ANG |         |        |        |         |         |        | 361.6* | 101.6* |
| ANT |         |        |        |         |         |        |        | 99.8*  |

GMM から生成されるとは言えない。本方法により、全感情カテゴリのすべての組合せについて尤度比検定を行った。

#### 4.2 結果

図1に、pleasantness-arousal空間と pleasantness-dominance空間における感情カテゴリの実データの分布を示す。さらに、尤度比検定の結果として、各感情カテゴリの逸脱度の差を表4に示す。表4中の\*は、尤度比検定の結果、仮説  $H_0$  を棄却した組合せを示す。多くの感情カテゴリの組合せで仮説を棄却した。仮説を棄却しなかった感情カテゴリの組合せは、SAD-NEU, DIS-NEU, SUR-NEUであった。

#### 4.3 考察

図1に示した pleasantness-arousal 空間では、JOY は高覚醒度で快状態、SUR は高覚醒度、SAD は低覚醒度、ANG が高覚醒度で不快状態の象限に配置され、Russell が提案した感情次元中の配置 [3] とほぼ同じ結果を示した。感情ラベリングの結果がおおむね妥当であることが分かる。また、NEU は感情が含まれない平静の発話と定義しているが、図1に示した NEU の分布をみると、pleasantness は4 (どちらでもない) 付近に分布しているが、それ以外の感情次元では2~4に分布していることが分かる。平静状態の発話は快-不快状態が中立であるが、他の感情次元は必ずしも中立ではないことが示された。

また、感情の三次元空間上における感情カテゴリの分布の尤度比検定では、NEU-SAD, NEU-DIS, NEU-SUR 以外のすべて組合せで帰無仮説を棄却し、NEUを除いた感情カテゴリは他の感情カテゴリとは独立した分布であることが示された。音声による感情次元知覚でも平静以外の8つの感情カテゴリの情報は消失しないことが示唆された。

## 5. 実験2：音響特徴量と感情次元・感情カテゴリの関係

### 5.1 分析手法

音響的に感情が表現される情報で抽出される次元を求めるため、ふたつのコーパスに含まれる分析対象の発話の音響特徴量を算出した。発話の音響特徴量を用いて主成分分析を行い、各主成分の寄与率から音響特徴量で感情を表現する主要な成分を抽出する。抽出した成分と感情次元の相関係数を求め、音響特徴量と感情次元との関係について検証する。

さらに、抽出した成分を用いて多次元空間を構成し、そこに配置されるふたつの感情カテゴリのサンプルについて実験1と同様の手順で、尤度比検定を行った。実験2では、各発話における音響特徴量の主成分を確率変数とし、この変数が多次元混合正規分布 (GMM) に従うと仮定して、尤度比検定を行った。

### 5.2 音響特徴量

音響特徴量はこれまでの研究 [10] 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 (fnimi)・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 (cmmaxi)・平均値 (cmean)・標準偏差 (cstdv) である。 $F_0$  は STRAIGHT [12] を用いて抽出したのち、対数に変換した。 $C_1$  の抽出には VOICEBOX [13] を用いた。音圧は次式で求めた。

$$p(m) = 10 \log \frac{1}{N} \sum_{n=0}^{N-1} S^2(Nm + n).$$

$S(i)$  は  $i$  番目のサンプルの振幅値を、 $m$  はフレーム番号を示す ( $m = 0, 1, 2, \dots$ )。  $F_0$  は対数に変換し、話者ごとにこの対話内平均  $F_0$  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。

### 5.3 結果

音響特徴量より求めた主成分は、各主成分の固有値が1以上であることを基準に、4つの成分を抽出した (累積寄与率 99.86%)。各主成分と感情次元との相関係数を表5に示す。第一主成分 (PC1) は arousal との相関が強く、その係数は

表 5 感情次元と主成分との相関係数.

|              | PC1   | PC2   | PC3   | PC4  |
|--------------|-------|-------|-------|------|
| pleasantness | -0.32 | 0.02  | -0.29 | 0.10 |
| arousal      | -0.71 | -0.06 | -0.31 | 0.16 |
| dominance    | -0.57 | -0.06 | -0.12 | 0.32 |
| credibility  | -0.33 | 0.02  | -0.22 | 0.19 |
| interest     | -0.54 | -0.02 | -0.26 | 0.21 |
| positivity   | -0.15 | 0.04  | -0.17 | 0.04 |

表 6 音響特徴量による四次元空間内の感情カテゴリの逸脱度の差.

|     | ACC    | FEA   | SUR    | SAD    | DIS    | ANG    | ANT    | NEU    |
|-----|--------|-------|--------|--------|--------|--------|--------|--------|
| JOY | 391.7* | 90.8* | 180.0* | 832.8* | 440.4* | 214.1* | -57.2  | 104.4* |
| ACC |        | -81.7 | 299.6* | 520.2* | 246.6* | 100.8  | -2.5   | -39.7  |
| FEA |        |       | -6.0   | 124.1* | 5.2    | 8.9    | -43.4  | -49.9  |
| SUR |        |       |        | 515.7* | 206.3* | 57.6   | 35.2   | 51.8   |
| SAD |        |       |        |        | 362.8* | 319.2* | 332.6* | 17.0   |
| DIS |        |       |        |        |        | -28.7  | 69.7*  | 35.6   |
| ANG |        |       |        |        |        |        | -27.3  | 45.4*  |
| ANT |        |       |        |        |        |        |        | -6.7   |

-0.71 となった。第二主成分 (PC2) はいずれの次元も相関係数が低い結果となった (-0.06-0.04)。第三主成分 (PC3) は pleasantness および arousal との相関係数が -0.30 付近を示した。第四主成分 (PC4) は dominance との相関係数が最も高く、0.32 となった。

尤度比検定の結果として、各感情カテゴリの逸脱度の差を表 6 に示す。表中の \* は検定の結果、仮説  $H_0$  を棄却した組合せを示す。JOY は ANT 以外の 7 個の感情カテゴリとの検定で仮説  $H_0$  を棄却した。また、SAD も NEU 以外の 7 個の感情カテゴリとの検定で有意差を示した。DIS は 5 個の感情カテゴリ (JOY, ACC, SUR, SAD, ANT) と有意差を示した。また、4 個の感情カテゴリと有意差を示したのは ACC (JOY, SUR, SAD, DIS) と SUR (JOY, ACC, SAD, DIS) であった。ANG は 3 個の感情カテゴリと、FEA/ANT/NEU はそれぞれ 2 個の感情カテゴリとしか有意差を示さなかった。

#### 5.4 考察

音響特徴量より抽出した主成分は感情次元との強い負の相関を示すものもある反面 (PC1 と arousal), いずれの感情次元とも無相関である成分 (PC2) も抽出された。覚醒度の高い発話は、PC1 の値が小さくなることを示している。また、PC2 は感情以外の情報を表していることが示唆される。音響情報が含む感情以外の情報としては話者性があげられる。音響特徴量を算出する際に  $F_0$  関連の特徴量は話者で正規化を行ったが、音圧やケプストラムの 1 次係数は話者正規化を行っていない。そのため、話者を特徴づけるような情報が特徴量に含まれたままであった可能性がある。音圧とケプストラムの 1 次係数の話者正規化が求められる。他の二つの主成分 (PC3 と PC4) については、相関係数が低めだが、ある程度の相関があるとみなせる。特に PC4 は dominance との弱い正の相関を示し、支配傾向が強い発話は PC4 の値が高くなることが示めされた。

音響的特徴量より抽出した主成分で構成した四次元空間上で各感情カテゴリの尤度比検定を行った結果をみると、いずれの感情も他の 8 個の感情と独立した分布を示すことはなかったも

の、JOY と SAD が 7 個の感情と独立した分布を示している。喜びと悲しみのふたつの感情は音響的に独立した感情カテゴリであることが示唆される。また、FEA は JOY と SAD と、ANT は SAD と DIS と独立した分布を示したが、それ以外の感情とは尤度比検定において有意な差を示さなかった。恐れと期待は音響的に独立性の低い感情カテゴリであることが示唆される。他の多くの感情カテゴリとの独立性が低い感情カテゴリは、音声からの感情知覚においては、ひとつの感情カテゴリとして存在させず、有意差を示さなかった感情カテゴリとともに再定義しなおす必要がある。

## 6. おわりに

音声コーパスに対する感情ラベルを共通化することをめざし、まずは既存コーパスに付与されている感情ラベルの関係性について検証を行った。公開されているふたつの感情音声コーパスを用いて、相互に感情ラベルを付与するクロスコーパス感情ラベリングを行い、感情カテゴリと感情次元の関係を音声知覚の観点から、感情次元で構成される感情空間上に配置される各感情カテゴリの分布の独立性を、尤度比検定によって検証した。さらに、コーパスに含まれる音声の音響的特徴量から、音響特徴のみで構成される次元を抽出し、それが感情ラベルとどのような関係があるかを相関係数および尤度比検定によって検証した。

その結果、感情の三次元空間上における感情カテゴリの分布の尤度比検定では、平静の発話を除いた各感情カテゴリは他の感情カテゴリとは独立した分布であることが示され、音声による感情次元知覚でも平静以外の 8 つの感情カテゴリの情報は消失しないことが明らかとなった。また、音響特徴量より抽出した主成分のうち第一主成分は arousal との強い負の相関を示し、覚醒度の高い発話ほど第一主成分の値が小さくなることが明らかとなった。さらに、第四主成分は dominance との弱い正の相関を示し、支配傾向が強い発話は第四主成分の値が高くなることが示めされた。また、音響的特徴量より抽出した主成分で構成した四次元空間上で各感情カテゴリの尤度比検定を行った結果からは、喜びと悲しみのふたつの感情は他の 7 つの感情と音響的に独立した感情カテゴリであることが明らかとなった。また、恐れと期待は、多くの感情カテゴリと音響的に独立していないこと明らかとなった。

今回の音響特徴量を用いた実験では、話者性を完全に排除せずに用いた音響的特徴量が存在した。今後は、各音響的特徴量に適切な話者正規化の手法を検討するとともに、分析の精度を向上させる予定である。

## 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。

## 文献

- [1] P. Ekman and W. V. Friesen: "Unmasking the Face: A Guide to Recognizing Emotions From Facial Expressions", Prentice Hall, New Jersey (1975).
- [2] R. Plutchik: "Emotions: A psychoevolutionary synthesis",

- Harper & Row, New York (1980).
- [3] J. A. Russell: "A circumplex model of affect", *Journal of Personality and Social Psychology*, **39**, 6, pp. 1161–1178 (1980).
  - [4] Y. Zong, W. Zheng, T. Zhang and X. Huang: "Cross-Corpus Speech Emotion Recognition Based on Domain-Adaptive Least-Squares Regression", *IEEE Signal Processing Letters*, **23**, 5, pp. 585–589 (2016).
  - [5] P. Song, W. Zheng, S. Ou, X. Zhang, Y. Jin, J. Liu and Y. Yu: "Cross-corpus speech emotion recognition based on transfer non-negative matrix factorization", *Speech Communication*, **83**, pp. 34–41 (2016).
  - [6] B. Schuller, Z. Zhang, F. Wenyinger and F. Burkhardt: "Synthesized speech for model training in cross-corpus recognition of human emotion", *International Journal of Speech Technology*, **15**, 3, pp. 313–323 (2012).
  - [7] Z. Zhang, F. Wenyinger, M. Wöllmer and B. Schuller: "Un-supervised learning in cross-corpus acoustic emotion recognition", 2011 IEEE Workshop on Automatic Speech Recognition and Understanding, ASRU 2011, Proceedings, pp. 523–528 (2011).
  - [8] B. Schuller, B. Vlasenko, F. Eyben, M. Wöllmer, A. Stuhlsatz, A. Wendemuth and G. Rigoll: "Cross-Corpus acoustic emotion recognition: Variances and strategies", *IEEE Transactions on Affective Computing*, **1**, 2, pp. 119–131 (2010).
  - [9] B. Schuller, B. Vlasenko, F. Eyben, G. Rigoll and A. Wendemuth: "Acoustic emotion recognition: A benchmark comparison of performances", Proceedings of the 2009 IEEE Workshop on Automatic Speech Recognition and Understanding, ASRU 2009, pp. 552–557 (2009).
  - [10] Y. Arimoto, H. Kawatsu, S. Ohno and H. Iida: "Naturalistic emotional speech collection paradigm with online game and its psychological and acoustical assessment", *Acoustical Science and Technology*, **33**, 6, pp. 359–369 (2012).
  - [11] H. Mori, T. Satake, M. Nakamura and H. Kasuya: "Constructing a spoken dialogue corpus for studying paralinguistic information in expressive conversation and analyzing its statistical/acoustic characteristics", *Speech Communication*, **53**, 1, pp. 36–50 (2011).
  - [12] H. Kawahara, M. Morise, T. Takahashi, R. Nisimura, T. Irino and H. Banno: "Tandem-STRAIGHT: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, F0, and aperiodicity estimation", 2008 IEEE International Conference on Acoustics, Speech and Signal Processing, IEEE, pp. 3933–3936 (2008).
  - [13] M. Brookes: "VOICEBOX: Speech Processing Toolbox for MATLAB" (1999).