

SVMを用いた自発対話音声の 感情認識における学習データの検討

阿部 渉^{†1} 真壁 大介^{†1} 小坂 哲夫^{†1}

感情認識の分野においては、収集が容易である演技音声を用いて音響モデルの学習を行うことが一般的である。一方で、自然に感情が表出した音声は収集が難しく、これを使用した研究の例は非常に少ない。本研究では自発対話音声を用いて感情認識を行い、その際に自発対話音声の感情認識における認識率向上を図るため、学習データについて検討を行う。

Study on Training Data for Emotion Recognition in Spontaneous Dialogue Speech Using SVM

WATARU ABE,^{†1} DAISUKE MAKABE^{†1}
and TETSUO KOSAKA^{†1}

In the field of emotion recognition, it is common to train an acoustic model by using acted speech data which are easy to collect. On the other hand, it is difficult to collect speech including natural emotions, and there are very few study examples using this type of speech. In this research, emotion recognition is performed using spontaneous dialogue speech. In order to improve the recognition performance of emotion recognition in spontaneous dialogue speech, we make a study on training data.

1. はじめに

人間が発する音声には発話内容だけではなく、感情といった心理的な非言語情報も含まれている。音声対話システムなどにおいて、発話内容だけではなく感情についても認識することができれば、人間と会話をするような自然な対話が可能になり、より親しみやすく使いやすいものになると考えられる。

感情認識において音声から感情の特徴量を抽出するためのツールキットとして openEAR¹⁾ が広く用いられている。また、認識手法として一般的には SVM(Support Vector Machine) が用いられている^{2),3)}。近年では、DNN(Deep Neural Network) による感情認識も行われている⁴⁾。

このような感情認識では音声コーパスとして俳優や一般人が指定された文章を、指示された感情が相手に伝わるようにして発せられた演技音声を使用されている。一方で自然に感情が表出した自発音声は収集が容易ではない。また、演技音声のようにあらかじめ感情が指定されていないため、あとから第三者によって感情のラベル付けが必要となる。この場合、評価者によって感じ方が異なる場合があり、感情のラベル付けが難しいという問題がある。

近年、自然に感情が表出した音声の感情認識を可能にするために、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) が登場した⁵⁾。OGVC はオンラインゲームをプレイしている最中の音声チャットの様子を収録した自発対話音声と、それを元に俳優が発声した演技音声の 2 種類の音声を用意されている。これらはあらかじめ 9 種類の感情ラベルと感情強度 (感情の表出度合い) の評価が行われており、「怒り」や「喜び」といったカテゴリ分けによる感情認識に向いている。

自発音声による感情認識の例として、発話内容は指定されているが、感情については自分なりの表現で発してもらった音声によるネガティブ感情の認識実験を行った研究⁷⁾がある。自発音声について「怒り」や「喜び」といったカテゴリ分けされた感情認識の例は少ない。一方、カテゴリ分けをした自発音声の感情認識の例として、OGVC の自発対話音声を使用し、声門特性に基づいた特徴量の検討を行った研究⁶⁾がある。そこでは特徴量についての検討は行っているが、どのような音声を学習データとして使用すればよいかについては検討されていない。本研究では学習データについて焦点をあて、OGVC の自発対話音声を使用し、自発対話音声の感情認識における学習データの検討を行う。

^{†1} 山形大学
Yamagata University

2. 実験手法

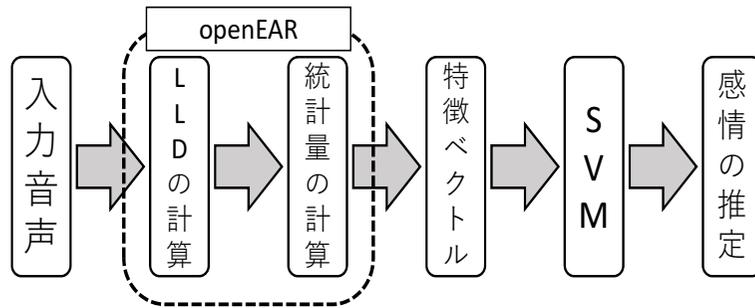


図1 感情認識の流れ

Fig.1 Flow chart of emotion recognition

2.1 感情認識における特徴量

本研究では特徴量の抽出に openEAR を使用する。抽出手順を図1に示す。特徴抽出では、はじめに入力音声から LLD(Low Level Descriptors) が抽出される。LLD とは、基本周波数やパワーといった特徴量のことである。LLD は音声認識と同様にして窓関数を用いてフレームごとに分割し、フレームごとに計算を行う。これによって得られる LLD は時系列であるが、感情は音声全体で表されているものと考えられるため、LLD から最大値や平均値などの統計量を計算し、特徴ベクトルへと変換される。本研究では、INTERSPEECH 2009 の特徴量セット⁸⁾ である 384 次元の特徴ベクトルを用いる。ここで使用される特徴量について以下に示す。

- RMSenergy (Root-Mean-Square energy)
信号の強さで音量の二乗平均平方根値で表される。
- MFCC (Mel-Frequency Cepstral Coefficients) 1-12
メル周波数ケプストラム係数のうち低次の 1~12 次元を使用する。
- 零交差率
時間信号の零交差率 (フレームベース) であり、音声波形の検出などに使用される。
- VoiceProb (Voice Probability)
その時点での音が声であるかどうかの確率である。全パワーに占める調波成分の割合を

表す。

● F0

ケプストラムから計算された基本周波数で、有声部で値をとり、無声部では 0 になる。なお、本研究の実験で用いる F0 はスムージングしたものを用いる。ここでは、指数関数的減衰によるスムージングが行われており、時刻 t における減衰する量を $N(t)$ 、崩壊定数を λ とすると

$$N(t) = N_0 e^{-\lambda t} \quad (1)$$

で表される。ここで、 $N_0 = N(0)$ は初期値、また定数は $\lambda = 0.75$ で計算されている。

2.2 SVM(Support Vector Machine)

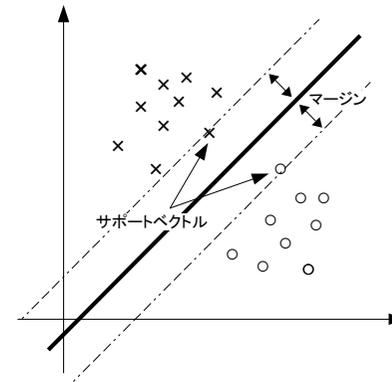


図2 SVM のイメージ

Fig.2 Image of SVM

感情認識における認識手法として SVM(Support Vector Machine) を用いることが一般的であり、本研究でも SVM による感情認識を行う。SVM は 2 クラス間を線形分離するパターン認識手法であり (図 2)、2.1 節で述べた特徴ベクトルを感情ごとのクラスとして分類する。本研究では WEKA⁹⁾ と呼ばれるツールを使用する。openEAR では WEKA で読み込むことが可能なファイルとして特徴ベクトルが出力されるため、容易に感情認識実験の学習、評価を行うことができる。SVM による学習方法について下記に示す。なお、詳細につ

いては文献¹⁰⁾に記載されている。

サポートベクトルと決定境界との距離は、

$$\min_{i=1,\dots,n} \frac{|\mathbf{w}^t \mathbf{x}_i + w_0|}{\|\mathbf{w}\|} = \frac{1}{\|\mathbf{w}\|} \quad (2)$$

となり、右辺分母にある $\|\mathbf{w}\|$ を最小化することで、マージンが最大となる (後の展開のために $\|\mathbf{w}\|^2$ を最小化するという問題に置き換える)。

しかし、誤識別が起きない範囲で行う必要があり、識別関数の正負と正解クラスラベルの正負を合わせることで、以下のような不等式で表現できる。

$$y_i(\mathbf{w}^t \mathbf{x}_i + w_0) \geq 1 \quad (i = 1, \dots, n) \quad (3)$$

このような条件のもとで最小化問題を考える場合は、ラグランジュの未定乗数法を用いて求める。具体的には下記の (4) 式ようになる。ここで α_i はラグランジュ係数である。

$$\sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) \quad (4)$$

また、 $K(\mathbf{x}_i, \mathbf{x}_j)$ はカーネル関数であり、本研究では、下記の多項式カーネル関数 (5) 式を用いる。

$$K(\mathbf{x}, \mathbf{x}') = (\mathbf{x}^t \mathbf{x}' + 1)^p \quad \text{ただし } p \text{ は自然数} \quad (5)$$

3. 音声データと実験条件

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)⁵⁾ はオンラインゲームをプレイしている際の 2~3 人程度の音声チャットの会話を収録している。その音声を自発対話音声として男性 9 名、女性 4 名の 13 話者、約 1 万発話が収録されており、感情のラベルや強度が付与されている (ただし、男女それぞれ 1 名の話者について、振幅が小さいなどの音響的分析に影響を及ぼすなどという理由で、感情のラベルや強度は付加されていないため、実際には 11 話者分の 6578 発話のデータが使用可能である)。また、自発対話音声の発話の 2656 発話について、男女 2 名ずつのプロの俳優に感情を含まない「平静」と、感情強度 1~3 で演技した音声 (演技音声) も収録されている。なお、音声データは 40kHz, 16bit の WEVE 形式で保存されている。

3.1 自発対話音声

自発対話音声は、オンラインゲームをプレイ中の 2~3 人の計 6 組の音声チャットを録音

したものであり、11 話者の 6578 発話に感情のラベルが付加されている。感情のラベルは、3 名の評価者が発話を聞き、「喜び (JOY)」「受容 (ACC)」「恐れ (FRA)」「驚き (SUR)」「悲しみ (SAD)」「嫌悪 (DIS)」「怒り (ANG)」「期待 (ANT)」「平静 (NEU)」「その他 (OTH)」の 10 種類から感情を選ぶ。2 名以上が一致した場合、それが発話の感情として決定される。次に、前述で感情が決定した発話について男性 13 名、女性 5 名の 18 名の評価者が 1(弱)-3(中)-5(強) の 5 段階で感情強度の評価を行う。このように、自発対話音声の感情ラベル、感情強度は異なる評価者の主観的な評価によって決定している。

3.2 演技音声

演技音声は、自発対話音声の発話のうち、「平静」「その他」を除いた 8 つの感情の一部発話を、男性 2 名、女性 2 名の計 4 名のプロの俳優によって演技した音声 2656 発話を収録している。感情数は「喜び」「受容」「恐れ」「驚き」「悲しみ」「嫌悪」「怒り」「期待」の 8 感情分ある。また、1 発話に対して感情強度 0-3 の計 4 段階で発声した音声があり、感情強度 0 は「平静」として用いる事ができる。

3.3 感情ラベル

本研究では OGVC に収録されている音声について、次の表 1 のような基準でラベルが付加されている。なお、本研究では表 1 の上部にある「怒り」「喜び」「悲しみ」「驚き」「平静」の 5 感情のみを用いる。

3.4 実験条件

本実験で使用する特徴ベクトルの生成条件を表 2 に示す。なお、本研究では入力音声である OGVC 演技音声に付与されている発話前後の 400ms の無音区間が感情認識の結果に影響を及ぼすと考えられたため、事前に取り除いている。また、今回の実験においては INTERSPEECH 2009 の特徴量セット⁸⁾ である 32 特徴量 \times 12 統計量の計 384 次元の特徴ベクトルを使用している。音声データから openEAR によって特徴ベクトルを生成し、それを WEKA の SVM を用いて学習、評価を行う。

4. 演技音声による感情認識

自発対話音声の感情認識実験を行う前に、OGVC の演技音声を用いて「怒り」「喜び」「悲しみ」「驚き」「平静」の 5 感情の認識実験を行う。OGVC 演技音声 4 話者のうち、3 話者

表 1 OGVC の感情ラベル
 Table 1 Emotion labels of OGVC

感情	記号	説明
怒り	ANG	許しがたい事柄に接し、不快感を抑えきれず、いらだった状態の感情
喜び	JOY	良いことに会って非常に満足し、うれしい、ありがたいと思う感情
悲しみ	SAD	不幸なことに会った時など、取り返しのつかない事を思い続けて泣きたくなる感情
驚き	SUR	意外なことを見聞きして心が強く動揺し、平静を失う、どう判断すべきか戸惑う感情
平静	NEU	全く感情が表れていない
受容	ACC	心がひきつけられ、積極的に受け入れよう、接し続けようとする感情
恐れ	FEA	危害が及ぶことを心配してびくびくし、その人やその物と接することを避けたがる感情
嫌悪	DIS	その状態・行為をすんなりと受け入れることができず、避けようとする感情
期待	ANT	望ましい事柄の実現、好機の到来を心から待つ感情
その他	OTH	ノイズが大きい場合など 8 種類の感情に分類不能なもの

1032 発話を学習データとし、残り 1 話者 344 発話を評価データとするクロスバリデーションによって実験を行う。「平静」については「怒り」「喜び」「悲しみ」「驚き」の感情強度 0 のものである。表 4 に感情ごとの発話数を示す。

4.1 実験結果および考察

各評価話者別の認識率を表 3 に示す。なお、F から始まる話者は女性、M から始まる話者は男性を表す。この結果より、OGVC 演技音声の認識率はどの感情も 50%前後であり、平均すると 51.4%となった。表 3 を見ると、評価話者によって認識率にばらつきが見られる。「悲しみ」に関しては男性より、女性の方が認識率が高い傾向にある。また、「平静」に関しては逆に女性より、男性の方が認識率が高く、性別によって認識率の傾向が異なっている。また、同じ女性同士でも FOY と FYN では「怒り」の認識率が大きく異なる。このとき、FOY の「怒り」が「平静」として認識されてしまった発話が多かった。これは、FOY の話者がコールドアンガーであったことにより、怒っているにも関わらず「平静」として認識されてしまったからであると考えられる。

表 2 実験条件
 Table 2 Experimental conditions

特徴ベクトル	
特徴量	RMSenergy, MFCC (1~12 次), 零交差率, Voice Probability, スムージングした F0, 及びこれら 16 特徴量の Δ (合計 32 特徴量)
統計量	最大, 最小, 範囲, 最大の位置, 最小の位置, 平均, 線形近似式の傾き, 線形近似式の切片, 線形近似式の二乗誤差, 標準偏差, 歪度, 尖度 (合計 12 統計量)
特徴量分析	
窓幅	0.025s
シフト長	0.01s
デルタ窓	2
WEKA の設定	
分類器	SMO(Sequential Minimal Optimization)
使用カーネル関数	多項式カーネル関数

表 3 各話者の演技音声の認識率 [%]
 Table 3 Recognition rate of acted speech for each speaker [%]

評価話者	感情					平均
	怒り	喜び	悲しみ	驚き	平静	
FOY	18.3	58.7	68.3	66.7	37.2	49.7
FYN	76.7	36.5	61.9	55.6	18.6	47.7
MOY	63.3	50.8	44.4	38.9	84.9	57.8
MTY	55.0	47.6	41.3	56.9	50.0	50.3
平均	53.3	48.4	54.0	54.5	47.7	51.4

5. 自発対話音声による感情認識

4 章では OGVC 演技音声による認識実験を行った。演技音声は、あらかじめ指定された感情を俳優が演じているものであるが、俳優によっても感情の演技の仕方が異なり、認識率に大きな差が生まれることがある。自発対話音声は演技音声と異なり、オンラインゲームをプレイしている最中に自然に発せられた音声であり、そのときの場面や状況によって発せられる発話や感情が予測できない。したがって、自発対話音声の感情認識は非常に難しいと考えられる。ここで、OGVC 自発対話音声 11 話者の 5 感情の発話数について表 4 に示す。

5.1 学習データの検討

一般にパターン認識においては、学習データと評価データの性質が類似しているときは高

表 4 5 感情の発話数
Table 4 Number of utterances of 5 emotions

感情	自発対話音声	演技音声
怒り	237	240
喜び	595	252
悲しみ	243	252
驚き	565	288
平静	798	344
総発話数	2438	1376

い認識率を得られるが、逆の場合は認識率が低下する。感情音声における演技音声と自発対話音声は感情の表出に大きな違いがあり、演技音声では大げさに感情が表出している場合が多い。このため、感情の表出の度合いが少ない自発対話音声の認識には、学習データとして演技音声より自発対話音声を用いる方が向いていると考えられる。よって、本章では学習データに自発対話音声を使用し、性能向上を図る。

一方、自発対話音声の収録にあたっては、各感情の種類ごと発話数を均等に集めることは難しい。表 4 に示すように、感情によって発話数に大きなばらつきが生じる。学習データが少ない場合、パラメータ推定を精度良く行うことが難しくなり、認識率の低下が懸念される。この問題を解決するために、自発対話音声と演技音声を混合し学習に使うことを提案する。この方法によって学習データの量の問題と、類似した性質を持つ学習データの使用の両面に対応する。

なお、自発対話音声と演技音声を混合する際に、演技音声における感情強度を考慮する。感情強度の強い発話を混合した場合、自発対話音声の特徴との乖離が懸念される。したがって、どの程度の強度の演技音声を混合すればよいかについて実験的に検討する。

5.2 実験条件

自発対話音声による感情認識実験の特徴ベクトルの生成条件などの実験条件を表 2 に示す。表 5 には使用する学習データと評価データ、表 6 には実験別の学習データを示す。なお、演技音声の感情強度 0 は「平静」である。

学習データについて、実験 1 では演技音声のみ、実験 2 では自発対話音声のみ、実験 3 では自発対話音声と演技音声の感情強度 1~3、実験 4 では自発対話音声と演技音声の感情強度 1 と 2、実験 5 では自発対話音声と演技音声の感情強度 1 とする。

表 5 使用データ
Table 5 Used data

学習データ	
演技音声 (感情強度 0)	4 話者 344 発話
演技音声 (感情強度 1)	4 話者 344 発話
演技音声 (感情強度 2)	4 話者 344 発話
演技音声 (感情強度 3)	4 話者 334 発話
自発対話音声	10 話者 (1 話者あたり 162~329 発話)
評価データ	
自発対話音声	11 話者 2438 発話

表 6 実験別の学習データ
Table 6 Training data for each experiment

	学習データ
実験 1	演技音声 (感情強度 0~3)
実験 2 ^{*1}	自発対話音声
実験 3 ^{*1}	自発対話音声+演技音声 (感情強度 1~3)
実験 4 ^{*1}	自発対話音声+演技音声 (感情強度 1, 2)
実験 5 ^{*1}	自発対話音声+演技音声 (感情強度 1)

*1 クロスバリデーションを使用

5.3 実験結果および考察

OGVC 自発対話音声の認識結果を学習データ別に示す。学習データが演技音声、自発対話音声、自発対話音声+演技音声 (感情強度 1~3)、自発対話音声+演技音声 (感情強度 1, 2)、自発対話音声+演技音声 (感情強度 1) の認識結果 (認識率 [%]) をそれぞれ表 7~表 11 に示す。

これら 5 つの実験結果を表 12 にまとめた。以上の結果より、学習データに演技音声を用いた場合、「平静」の認識率が非常に高いが、その他の感情については 20%以下となっており、全体的な認識率は低くなっている。一方で、学習データに自発対話音声を用いた場合、「喜び」「悲しみ」「驚き」の 3 感情について認識率が大きく向上している。しかし、「怒り」と「平静」については認識率が低下している。表 7 を見ると、「平静」だけでなくそのほかの感情について 6~7 割程度が「平静」として認識されてしまう傾向にある。この原因として、自発対話音声の感情強度が演技音声の感情強度に比べると低いことが理由として考えられた。自発対話音声の感情強度が全体的に低いことによって、学習データに演技音声を用いた場合に「平静」(感情が表れていない)として認識されてしまったと考えられる。また、

表 7 学習データ：演技音声 [%]

Table 7 Training data : Acted speech [%]

	認識された感情				
	怒り	喜び	悲しみ	驚き	平静
正解					
怒り	12.2	5.9	3.4	3.8	74.7
喜び	20.3	6.4	3.2	6.4	63.7
悲しみ	3.3	3.3	12.3	11.1	70.0
驚き	7.4	4.2	3.0	19.6	65.7
平静	7.1	3.0	2.1	2.4	85.3
平均	27.2				

表 8 学習データ：自発対話音声 [%]

Table 8 Training data : Spontaneous dialogue speech [%]

	認識された感情				
	怒り	喜び	悲しみ	驚き	平静
正解					
怒り	7.2	27.4	2.5	19.8	43.0
喜び	5.9	44.9	2.2	11.9	35.1
悲しみ	1.2	18.5	21.8	11.5	46.9
驚き	5.7	16.8	3.2	52.0	22.3
平静	3.0	17.5	4.1	8.4	66.9
平均	38.6				

表 11 学習データ：自発対話音声+演技音声 (感情強度 1) [%]

Table 11 Training data : Spontaneous dialogue speech and Acted speech (Emotional strength : 1) [%]

	認識された感情				
	怒り	喜び	悲しみ	驚き	平静
正解					
怒り	9.3	27.4	3.4	20.7	39.2
喜び	7.9	46.6	6.2	13.9	25.4
悲しみ	2.5	15.2	30.0	14.0	38.3
驚き	4.1	17.9	4.1	53.8	20.2
平静	4.0	19.0	7.0	8.6	61.3
平均	40.2				

表 9 学習データ：自発対話音声+演技音声 (感情強度 1~3) [%]

Table 9 Training data : Spontaneous dialogue speech and Acted speech (Emotional strength : 1 to 3) [%]

	認識された感情				
	怒り	喜び	悲しみ	驚き	平静
正解					
怒り	7.6	20.7	2.1	16.9	52.7
喜び	6.6	38.0	4.7	11.6	39.2
悲しみ	0.4	11.1	24.3	15.6	48.6
驚き	4.6	14.0	4.4	49.6	27.4
平静	2.5	15.3	3.5	6.9	71.8
平均	38.2				

表 10 学習データ：自発対話音声+演技音声 (感情強度 1, 2) [%]

Table 10 Training data : Spontaneous dialogue speech and Acted speech (Emotional strength : 1 and 2) [%]

	認識された感情				
	怒り	喜び	悲しみ	驚き	平静
正解					
怒り	13.1	27.4	2.1	19.0	38.4
喜び	9.1	48.7	5.4	12.4	24.4
悲しみ	2.1	15.6	30.0	16.0	36.2
驚き	2.1	16.8	4.6	52.9	20.9
平静	5.0	20.2	5.8	9.1	59.9
平均	40.9				

表 12 学習データ別の認識率 [%]

Table 12 Recognition rate for each training data set [%]

感情	学習データ				
	演技	自発	自発+演技 (強度 1~3)	自発+演技 (強度 1, 2)	自発+演技 (強度 1)
怒り	12.2	7.2	7.6	13.1	9.3
喜び	6.4	44.9	38.0	48.7	46.6
悲しみ	12.3	21.8	24.3	30.0	30.0
驚き	19.6	52.0	49.6	52.9	53.8
平静	85.3	66.9	71.8	59.9	61.3
平均	27.2	38.6	38.2	40.9	40.2

学習データに自発対話音声を用いることで、より小さな変化で他の感情として認識されてしまったため、「平静」の認識率が低下したと考えられる。

自発対話音声に演技音声を加えた認識実験では、演技音声の感情強度 1 と 2 を加えた場合が「怒り」「喜び」「悲しみ」「驚き」の 4 感情について認識率の向上が見られた。演技音声の感情強度 1 のみを加えた場合とそれほど大きな差は見られないが、「怒り」の認識率が向上したことが全体的な認識率の向上に繋がっていると考えられる。「怒り」の認識率を向上させるためには、ある程度の感情強度をもつ演技音声が必要であると考えられる。

以上のことより、自発対話音声の認識率を向上させるためには次の 2 つの学習データが必要であると考えられる。1 つ目は自発対話音声である。学習データにも評価データと同じ自発対話音声を用いることで、「喜び」「悲しみ」「驚き」の 3 感情の認識率が大きく向上する

と考えられる。2 つ目は学習データの量を増やすことである。今回は演技音声を加えたが、その際に演技音声の感情強度 1 と 2 の音声を加えることで認識率が 40.9% という結果が得られた。このことより、演技音声であったとしても、加える演技音声の感情強度を考慮することで認識率の向上に繋がると考えられる。今回加えた感情強度 1 と 2 というものは、評価データである自発対話音声の感情強度と近いものであったと考えられる。演技音声の感情強度 3 となると感情の表出度合いが大きく、大げさな演技となってしまう、それと自発対話音声を比べることで自発対話音声が「平静」として認識されてしまったと考えられる。

6. まとめ

本研究では、感情音声コーパスの OGVC を用いて自発対話音声の感情認識における学習データの検討を行った。5 感情の認識実験において、学習データとして自発対話音声と演技音声の感情強度 1 と 2 を用いることが一番高い認識率を得た。このことより、自発対話音声の感情認識には自発対話音声を含めるだけではなく、学習データの量も増やすことが必要

であると考えられる。その際、演技音声であったとしても演技音声の感情強度を考慮し、感情強度3の(感情の表出度合いの強い) 音声を除くことによって認識率の向上が見られた。

しかし、これでも4節のような学習データ、評価データともに演技音声とした感情認識実験と比べると自発対話音声の感情の認識率は低く、自然に感情が表出した音声の感情認識は難しいことがわかる。こうした自発対話音声などの感情の認識率を向上させる方法として、より多くの自発対話音声を集め、学習データとすることが考えられる。しかし、自発対話音声は収集だけでなく、その音声の感情を評価することが難しい。したがって、今回使用した特徴ベクトルを384次元から増やすことや、音声認識によって言語的な情報からも感情を推定するといった別な方法を用いることで自発対話音声の感情の認識率が向上すると考えられる。

参 考 文 献

- 1) Florian Eyben, Martin Wöllmer, Björn Schuller “openEAR - Introducing the Munich Open-Source Emotion and Affect Recognition Toolkit”, Proc. ACII, Amsterdam, Netherlands, pp. 576-581, (2009).
- 2) Ali Hassan, Robert I. Dampier “Multi-class and hierarchical SVMs for emotion recognition”, Proc. Interspeech 2010, pp. 2354-2357, (2010).
- 3) 内田正洋, 篠崎隆宏, 堀内靖雄, 黒岩真吾 “発話中の一部区間を用いた感情認識”, 電子情報通信学会論文誌 D Vol.J97-D No.1, pp. 236-238, (2014).
- 4) 真壁大介, 小坂哲夫 “DNN を用いた日本語音声の感情認識の検討”, 情報処理学会東北支部研究会, 15-6-B1-3, (2016).
- 5) 有本泰子, 河津宏美 “音声チャットを利用したオンラインゲーム感情認識コーパス”, 日本音響学会 2013 年秋季研究発表会講演論文集, 1-P-46a, pp. 385-388, (2013).
- 6) 竹部真晃, 山本一公, 中川聖一 “音声感情認識における声門特性に基づく特徴量の検討”, 日本音響学会講演論文集 1-Q-17, pp. 113-116, (2015).
- 7) 山本 雄登, 新妻 雅弘, 山下 洋一 “サポートベクターマシンを用いた音声におけるネガティブ感情の認識”, 日本音響学会 2016 年秋季研究発表会講演論文集, 2-Q-7, pp. 53-56, (2016).
- 8) Björn Schuller, Stefan Steidl, Anton Batliner “The INTERSPEECH 2009 Emotion Challenge”, Proc. Interspeech 2009, pp. 312-315, (2009).
- 9) WEKA <http://www.cs.waikato.ac.nz/ml/weka/>
- 10) 荒木雅弘 “フリーソフトでつくる音声認識システム”, 森北出版株式会社, pp. 63-71, (2007).