

## 文節区分による音声の感情認識に関する一考察 A Study about Feelings Recognition of Sound by Clause Division

飯高 智美†      金子 正人‡      武内 惇‡      藤本 洋‡  
Tomomi Iitaka    Masato Kaneko    Atsushi Takeuchi    Hiroshi Fujimoto

### 1. はじめに

本研究の目的は、ユビキタス社会に向け、ユーザの個性や感性に合わせた HMI(Human Machine Interface)を提供するために、音声から感情情報を抽出することで、よりユーザに合ったインタフェースを提供することである。

これまで、ファジィ学習ベクトル量子化法(以下、FLVQ と略)<sup>[1]</sup>を用いて、学習を行った複数の短文からの感情認識や、学習を行っていない(未学習の)複数の短文からの感情認識を行ってきた。しかし、人間のコミュニケーションは、短文だけではなく長文になる場合が多いことから、長文を認識させることが重要になる。

本稿では、長文に対応させるためのアプローチとして、まずはこれまで行ってきた短文を文節単位に区切る、方法を用いて感情認識を試みた実験の結果について述べる。

### 2. 長文音声の感情認識の考え方

短文においては、パラメータ解析をする範囲が短く、音声の特徴を捉える事ができる。しかし、長文になると、パラメータ解析をする範囲が広がるため、短文よりも詳しいパラメータを採取する事が困難である。そこで、文を文節単位に区切り、音声の特徴を詳細さらに捉える文節区分が有効であると考えた。

日本語の文は、文節が連ねられて構成されているため、文節ごとの感情認識方法が確立すれば、あらゆる長文の日本語音声について感情認識が可能になる。

本稿では、その有効性について論じる。

### 3. 音声の感情認識実験

全体を4つのステートに分けた感情認識実験の流れについて、図1に示す。実験は発話実験、聴取実験、パラメータ解析、認識実験の順に行う。各実験の内容は次に記載する。

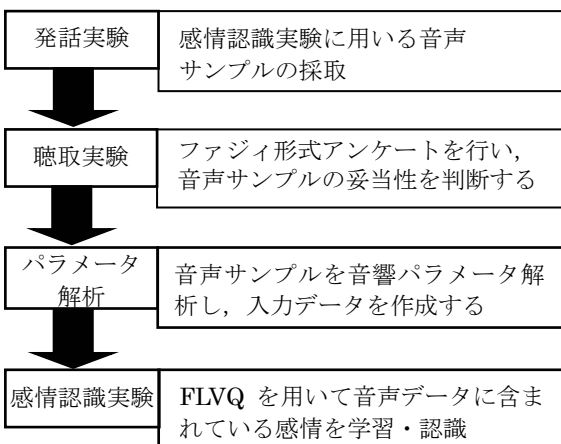


図1. 実験の一連の流れ

† 日本大学大学院工学研究科情報工学専攻

‡ 日本大学工学部情報工学科

### 3. 1 発話実験

発話実験では、実験に使用する音声サンプルの採取が目的である。本稿では、以下の3短文のサンプルを採取した。

表1. サンプルとして採取した音声

	短文	文節区分
A	ああ、雪だ	ああ
		雪だ
B	はい、わかりました	はい
		わかりました
C	ええ、聞いています	ええ
		聞いています

5つの感情『平静』『喜び』『悲しみ』『怒り』『嫌々』を付加したものをそれぞれ発話し、感情情報を含む音声サンプルを採取する。

今回は、成人男性6名、成人女性2名の計8名の音声サンプル(3短文×8人×5感情=200サンプル)を採取した。

### 3. 2 聴取実験

聴取実験において、ランダムに提示した音声サンプルを、被験者に聞いてもらい、どの感情が付加されているかファジィ形式のアンケートに記入してもらう。被験者30名に行い、音声サンプルが妥当であるか判断する。

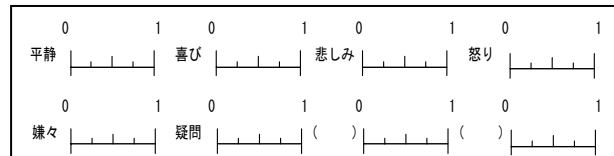


図2. ファジィ形式のアンケート用紙

### 3. 3 音響パラメータの解析

採取した音声サンプルをもとに、感情認識実験で用いる音響パラメータを得るために、ピッチ周波数や発声時間の解析を行う。

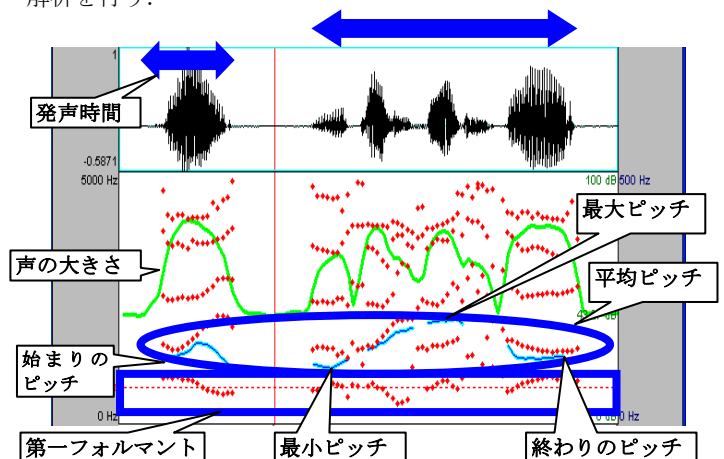


図3. 音響パラメータの解析

### 3. 4 感情認識実験

音声サンプルの音響パラメータ解析の結果から、FLVQ プログラムを用いた感情認識実験を行う。

音声の特徴パラメータは、第1フォルマント、発声時間、ピッチ周波数の最大ピッチ、最小ピッチ、平均ピッチ、始まりのピッチ、終わりのピッチ、声の大きさの8項目とし、FLVQによる学習回数は500回とする。音声サンプルの学習データは、短文A~Cまでの文節区分6つを用いる。学習では、聴取実験で行ったアンケートの結果における上位4名の音声を利用する。また、認識データは、学習データ以外の文節区分を使用する。感情認識実験で、以下について比較する。

- (1) 文節区分を用いて認識実験を行い、文節区分を用いていない場合と比較する。
- (2) 学習に用いた短文による認識と、未学習である短文の認識実験を行った場合と比較する。

### 4. 結果と考察

複数短文での感情認識結果について述べる。文節区分を用いた場合および用いていない場合での学習・認識を行った感情ごとの平均を表2、図4、図5に示す。

5感情ごとの平均を見てみると文節区分ありの場合は、学習済みおよび未学習の認識率の差が小さいことがわかる。文節区分なしの場合では、学習済みおよび未学習の認識率の差が大きい。特に未学習の認識率の差は、感情間のばらつきが非常に大きくなっている。

全体での認識率を図6に示す。文節区分ありの場合では、学習済みおよび未学習で、約8割の認識率を得ることができた。文節区分なしの場合、全体的に文節区分ありの場合より認識率は低かった。そして、未学習における認識率は、7割前半という結果であった。

表2. 分節区分による感情認識結果

	平静	喜び	悲しみ	嫌々	怒り
学習済み 文節区分なし	78.6%	72.4%	75.4%	77.2%	78.9%
未学習 文節区分なし	78.3%	76.7%	73.1%	78.0%	70.2%
学習済み 文節区分あり	81.4%	80.9%	78.1%	79.1%	79.4%
未学習 文節区分あり	82.5%	82.3%	79.9%	79.2%	78.6%

また、文節区分なしの場合の学習済みよりも文節区分ありの場合の未学習の認識率の方が高いことがわかる。

今回の実験では以下のことが分かった。

- ① 文節区分を用いた場合、学習済みおよび未学習でも高い認識率が得られた。
- ② 文節区分を用いた場合、全体の感情間の平均にばらつきがあまりなく、一定の認識率が得られた。

これは、文節区分を用いた場合では、文を文節単位に区切るため狭い範囲を詳細にパラメータ解析できることで、文節区分を用いていない場合と比べて認識率が向上したと考えられる。

### 5. おわりに

本稿では、長文に対するアプローチとして文節区分を提案し、これまで行ってきた短文での認識について有効であるかを実験し、その結果、学習済みおよび未学習の両方について文節区分が有効であることがわかった。

今後は、文節区分の有効性を更に確認するため、短文数を増やし、長い文章についても検討していく。

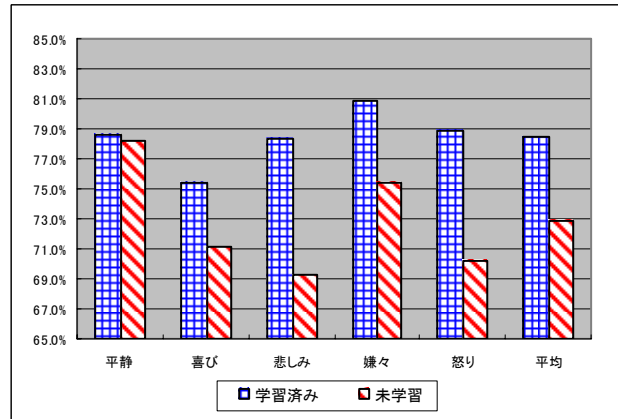


図4. 文節区分なしの感情ごとの平均認識率

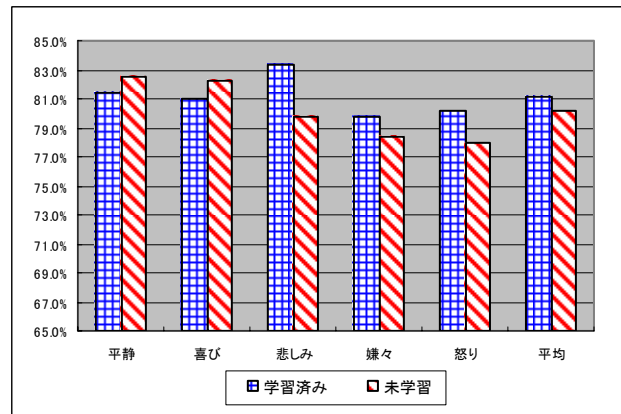


図5. 文節区分ありの感情ごとの平均認識率

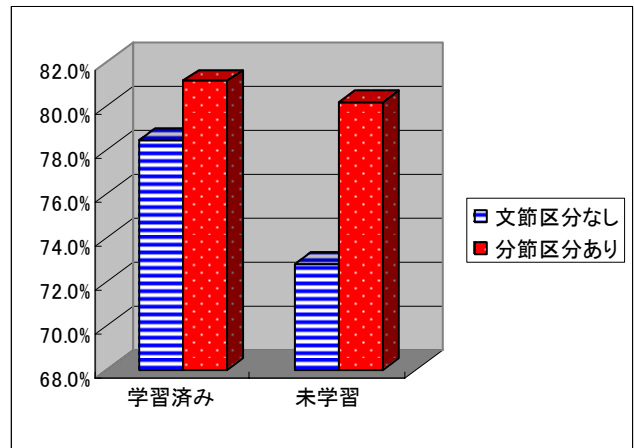


図6. 全体の平均認識率

### 【参考文献】

- [1] 櫻庭, 中本, 森泉: “ファジィ理論を用いた学習ベクトル量子化法”, 電子情報通信学会論文誌 Vol. J-73-DII No11, pp. 1863-1871 (1990)
- [2] 斎藤, 金子, 武内, 藤本: “ファジィ学習ベクトル量子化法を用いた未学習音声データの感情認識”, 平成17年度東北支部連合大会 (2005)