

## 音声にみられる感情表現について:聴取実験結果の分析

飯田 朱美<sup>†</sup>, 伊賀 聡一郎<sup>†</sup>, 安村 通晃<sup>†</sup>

慶應義塾大学大学院政策・メディア研究科<sup>†</sup>

〒 252 神奈川県藤沢市遠藤 5322

akeiida@sfc.keio.ac.jp

<http://www.mag.keio.ac.jp/~akeiida/>

### Abstract

我々はユーザー・フレンドリーなマルチモーダル・インタフェースの実現に向けて、その基礎となる脳波、音声の感情解析に着手した。本研究では聞き手がどの程度話者の感情を理解できるかを検討するために5種の感情の元で発声された音声の音響的特徴を分析するとともに聴取実験を行なった。聴取者に感情の多肢選択による判定を行なってもらった後、聴取者の発話ごとの印象の違いを分析するためにSD法による実験を行なった。評定データは因子分析し、得られた因子と感情音声の音響的特徴との関連を考察した。本稿では実験方法と分析結果を報告するとともに、問題点を検証する。

## Study of Emotion in Speech: Findings from Perceptual Experiments

Akemi Iida<sup>†</sup>, Soichiro Iga<sup>†</sup>, Michiaki Yasumura<sup>†</sup>

Graduate School of Media and Governance, Keio University<sup>†</sup>

### Abstract

Towards the realization of a user friendly multimodal interface, emotion analysis is under going. So far, one of the authors has reported the effectiveness of EEG(Electroencephalogram) measurement. Authors began emotion analysis of speech of which extension is the emotion analysis combining EEG and speech analysis. In this paper, emotion analysis on speech is reported. We have attempted an acoustic analysis of phrases uttered with 5 emotion types and conducted a perceptual experiment where subjects were told to identify emotion types. Succeedingly, SD(Semantic Differential) method experiment was conducted in order to define subjects' mental images for each emotion types.

## 1 はじめに

現在コンピュータはハード面の技術発展と激しい市場競争の結果、高性能のものが一般消費者にも入手可能になり、処理できるメディアも、画像や音声など多彩になったが、その反面、操作が複雑になったなど様々な問題も抱えている。現在使いやすいユーザーインターフェースの代名詞のようになっている GUI(Graphic User Interface) もマウス使用が困難な身体障害者やアイコンやマウスという概念に親しんでいない高齢者にとってはユーザフレンドリーとは言い難い。しかしながら、社会との接触を望みながらもなかなか実現が困難な彼らにこそ、ネットワークにつながったコンピュータは格好の社会参加のできる機会を提供してくれる。そのためにも、「誰にでも使用できて親しみやすいインターフェース」が必要不可欠である。

では、親しみやすいインターフェースとは何か。それは人間らしい、心のこもったコミュニケーションを可能にするインターフェースである。「心」とは即ち、感情である。人間は時として言葉がなくても、顔の表情や「え〜」という発声だけで、コミュニケーションすることができるが、それは感情が伝わるからである。感情発生と脳機能の働きの研究は以前から盛んだが、表情・音声・脳波を含めた生体信号と感情の関係についての研究は、感情が主観的で、客観的に扱うのが困難だったことから、取り組みが遅れていたが、近年、様々な測定方法が考案され、多くの機関が着手している [1][2][3][4]。

我々のグループでも脳波計測による感情解析を行なったのを始めとして、音声の感情表現の研究に着手した [5]。研究の目指すところは音声・脳波・顔の表情・しぐさなど生体信号から感情を理解し、人間のコミュニケーションをより円滑にするシステムの構築である。具体的には障害者のための感情表現が可能なコミュニケーション・エイドとしての text-to-speech システムであり、音声や脳波から計測された感情が通信先の相手に伝えられる。

本稿では上述の研究の一貫としての音声の感情表現の分析について報告する。本研究では聞き手がどの程度話者の感情を理解できるかを検討するために 5 種の感情の元で発声された音声の音響的特徴を分析するとともに聴取実験を行なった。聴取者に

感情の多肢選択による判定を行なってもらった後、聴取者の発話ごとの印象の違いを分析するために SD 法による実験を行なった。評定データは因子分析し、得られた因子と感情音声の音響的特徴との関連を考察した。本稿では実験方法と分析結果、そして今後の課題を報告する。

## 2 音声サンプルと音響的特徴

### 2.1 音声サンプル

5 種類の感情(喜び、怒り、驚き、悲しみ、嫌悪)を設定し、平静な状態で発声された音声と合わせて、男性 1 名女性 1 名に 5 モーラ語、「ありがとう、こんにちは、ちばたつや」とフレーズ、「何言ってるの」を発声してもらった合計 48 音声を 16bit 量子化、48kHz でサンプリングした。

### 2.2 音声的特徴

参考として「こんにちは」の音響特徴(発話時間、パワー平均値、ピッチレンジと形状)を図 1 に示す。

- **発話時間:** 前回のプレ実験では話者に限らず、怒りが最も短かったが [6]、今回の実験では両話者の発話には感情による違いはみられなかった(図 1 参照)。
- **パワーの平均値:** 全体的には驚き、喜びが大きく、悲しみ、嫌悪が小さい(図 1 参照)。怒りは話者間、単語サンプル間にばらつきがあり、より多くのサンプルの分析が必要である。「こんにちは、ありがとう」の 2 単語については怒りと嫌悪の発声ではアクセント核が移動し、前者では最終モーラ「は」に、後者では第 4 モーラに移動し、パワーも最大だった。
- **ピッチレンジと形状:** 喜び、驚きの場合にピッチレンジが最大になった。形状としては、「ありがとう」「こんにちは」はそれぞれ単語固有のアクセント型である起伏型(2 型)、平版型(0 型)を保持して、全体のピッチ形状としては下降している。「ちばたつや」に関しては起伏型(3 型)を保持して、ピッチ形状としては下降型と上昇型に分かれる。男性は驚

きのみ上昇、女性は喜び、驚き、嫌悪が上昇であった。フレーズ、「何言ってんの」については両話者とも下降型のイントネーションだった。

### 3 聴取実験

#### 3.1 多肢選択による判別実験

##### 3.1.1 目的

判別実験の目的はSD法に先立ち、話者が5感情を想定して発声した音声サンプルが、聴覚的にも話者が意図した感情と合致するか、つまり、SD法で感情を判別できる刺激語になり得るデータを抽出することである。

##### 3.1.2 被験者

被験者は、大学生・大学院生7名だった。

##### 3.1.3 方法

音声サンプルは単語ごとに被験者に提示される。被験者は最初に平静のサンプルを聞かされ、その後のピープ音に続いて、5感情の音声サンプルがランダムに提示される。被験者はそれぞれの感情のアイコンをクリックすることで、提示された音声の感情を判別する。

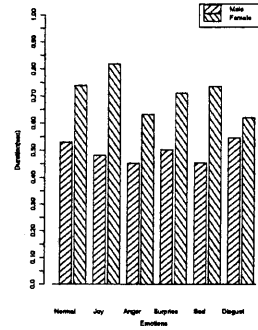
##### 3.1.4 実験結果

単語別得点は両話者の発話に対して、「ちばたつや、こんにちは、ありがとう、何言ってんの」の順だった。表1に示す通り、認識率は全体で65.7%で、5%の有意水準で検定を行った結果、有意であった。

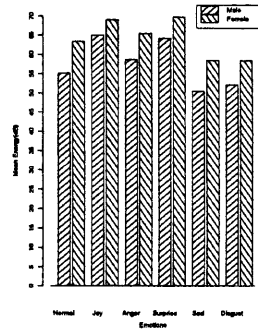
表 1: 発話別の判別結果

	有難う	今日は	ちばたつや	何言ってんの
男性	62.8%	74.2%	88.5%	48.5%
女性	68.5%	77.1%	85.7%	51.4%
Total	65.7%	75.7%	87.1%	50.0%

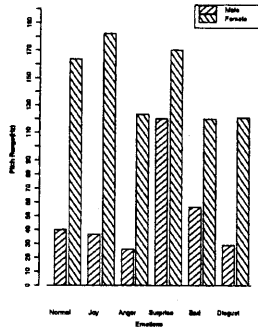
感情別の判別実験の結果を表2に示す。



(a) 発話時間:左から平静、喜、怒、驚、悲、嫌悪



(b) パワー平均値:左から平静、喜、怒、驚、悲、嫌悪



(c) ピッチレンジ:左から平静、喜、怒、驚、悲、嫌悪

図 1: 感情別の音響パラメータ

ここで得点の高かった上位3単語「ちばたつや、こんにちは、ありがとう」についてSD法での聴取実験を行なった。

表 2: 感情別の判別結果

	喜び	怒り	悲しみ	驚き	嫌悪
喜び	66.1%	1.8%	0.0%	32.1%	0.0%
怒り	5.4%	67.9%	0.0%	0.0%	26.8%
悲しみ	1.8%	1.8%	91.1%	0.0%	5.4%
驚き	35.7%	7.1%	1.8%	51.8%	3.6%
嫌悪	3.6%	12.5%	10.7%	0.0%	73.2%

## 3.2 SD法による実験

### 3.2.1 目的

被験者の音声サンプルごとの印象の違いを分析するためにSD(Semantic Differential)法を用いて実験を行なった。評定データは主因子法により因子分析し、ヴァリマックス法により回転した。その結果、得られた因子と感情音声の音響的特徴との関連を考察した。

### 3.2.2 方法

判別実験結果の中から4発話中3発話について実験を行なった。刺激の提示はランダムに行なわれた。被験者には5段階評価の11の評価語(形容語対)が計算機のウィンドウ上に提示され、一定時間(約30秒)内にすべての項目に関する選択を行なってもらう。刺激は5感情、2話者、3単語で合計30刺激を提示し、実験の所要時間は平均15分だった。評価語対を表3に示す。

本実験は被験者の聴覚的印象と音響的特徴との関連づけが目的なため、評価語は速い・遅い、強い・弱いなど本研究で計測した音響的特性を示唆するような形容詞を選択した。それ以外にも、本研究では対象としなかったが音声の特徴づける重要な要素としての呼気流や声道特性にも関連するような形容詞対を選択した。

### 3.2.3 被験者

被験者は、大学生・大学院生25名だった。

表 3: 評価語一覧

遅い(slow)	-	速い(fast)
暖かい(warm)	-	冷たい(cold)
沈んだ(depress)	-	弾んだ(lively)
緊張感に満ちた(strain)	-	のどかな(calm)
こもった(confined)	-	澄んだ(clear)
平板な(monotonous)	-	起伏のある(undulate)
強烈な(intense)	-	柔和な(gentle)
低い(low)	-	高い(high)
力強い(powerful)	-	弱々しい(weakly)
明るい(cheerful)	-	暗い(gloomy)
鋭い(sharp)	-	丸みのある(roundish)

### 3.2.4 結果

評定データを因子分析の結果、「強さに関する因子(強烈-柔和因子)」、「明るさに関する因子(明るさ-暗さ因子)」の2因子が得られた。それぞれのパラメータの因子得点と得られた因子との関連から、今回対象にした音響的特性の中ではパワーが「強さに関する因子」、ピッチレンジが「明るさに関する因子」に関連すると考えられる。

図2は感情音声別の評定データと評価語との関係を表している。図中の各々の折れ線は5感情を表している。ここで採用した感情音声別の評定データは、判定実験で全員が正しく判定したものである。評価軸の1と5がそれぞれ「強く思う」、3が「どちらでもない」、2と4がそれぞれの中間を表す。

また、図3は因子1、2平面における刺激音声の分布である。y軸が「強さにおける因子」、x軸が「明るさにおける因子」であり、喜び、驚きは「明るい・柔和」という空間に分布し、悲しみ、嫌悪は「柔和・暗い」という空間に分布している。怒りはy軸方向に細長く分布している。

## 4 考察

音響特徴はサンプル数も少ないため、確実に特徴を断言するのは難しく、より多くのデータが必要である。また音声サンプルの収集についても、今回は発話者に設定された5感情の元で発声してもらうように依頼したが、それによって自然さが損なわれてしまうことは否定できない。自然な会話からのサンプルの抽出は困難だが、必要である。また、

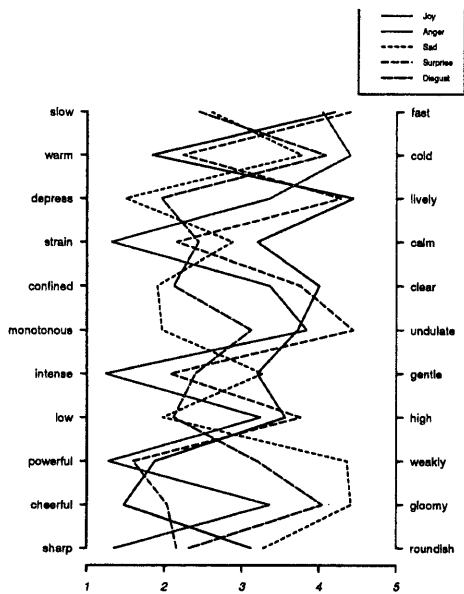


図 2: 感情別の平均得点

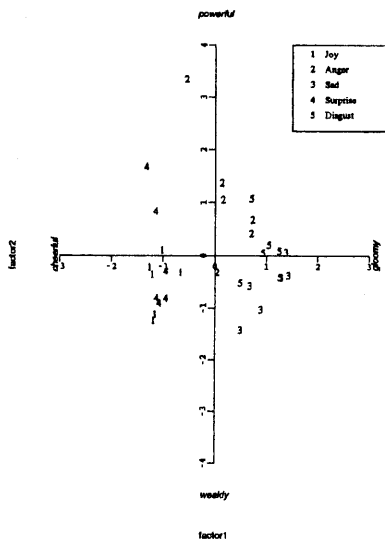


図 3: 因子 1、2 平面図

完全に自然な発話が無理であっても、状況設定をしたり、また長めの発話を発声してもらって、その中から抽出するなど工夫が必要であろう。

感情タイプの数も研究者によって様々であり、多いものでは14感情の分析を試みている[7]。感情の数も検討が必要である。我々の目的はコミュニケーション・エイドへの応用であるので、その用途では、どのくらいの感情の識別が可能であればよいか、という観点から検討したい。

今回の実験では発話する側、聞き取る側の双方の感情の表現、理解には個人差があり、感情タイプごとのパラメータを決定するには、より多くの話者の音声サンプルの検証を必要とする。また、SD法の実験を通して、判別テストの重要性を再認識した。今回、因子1、2の平面図で喜び・驚きと悲しみ・嫌悪のペアが非常に接近している。これは音響特性に似た部分があるという観測も成り立つが、判別実験で、かなりの混同があったことから、誰にでもすぐに特定の感情と判定できないサンプルをSD法の刺激語としたということも否定できない。怒りのサンプルについても強烈から柔和な印象まで、かなり広い分布になっており、嫌悪と判別した被験者も多く、判別実験の結果と合わせて検討する必要がある。その反面、男性話者の「こんにちは」の喜びと驚きのように、SD得点上は全く同じであるにもかかわらず、被験者が両者を全く混同することなく判別しているものもあり、聴覚的には確かに前者が明るく、後者がこもっているような印象を受けるが、この違いが、どんな音響特徴で表されるのか、より深い分析が望まれる。今回、行なえなかったLPCスペクトル解析、フォルマント分析などを次回には行なう予定である。

音声サンプル、「ちばたつや」においては、感情を変えて発声されると、イントネーション・パターンも変化すると同時に発話の含む意図も変化する。女性の発話の場合は怒りと悲しみが下降型となっているが、それ以外は上昇型である。男性の場合は驚きのみ上昇型である。これは語彙の担っている意味の他に自分の印象を感情表現によって伝達していると考えられる。感情は状況や話し手-聞き手との関係、発話意図、態度、性別や年齢などの個人情報に依存する。「ありがとう」の怒りと嫌悪の場合に見られるように語彙のアクセント型が感情によ

て移動するという現象も感情を計測する上で重要な研究課題である。

## 5 まとめ

本研究では聞き手がどの程度話者の感情を理解できるかを検討するために5種の感情の元で発声された音声の音響的特徴を分析するとともに聴取実験を行なった。聴取者に感情の多肢選択による判定を行なってもらった後、聴取者の発話ごとの印象の違いを分析するためにSD法による実験を行なった。評定データは因子分析し、得られた因子と感情音声の音響的特徴との関連の考察を試みた。

今後はより効果的なLPCスペクトル解析、フォルマント分析などの音響分析法を採用しながら、研究対象とする感情の種類の設定、より自然な音声サンプルの収集を行ない、音声の感情分析に取り組むと同時に発話時の脳波計測を行なうなど、脳波の感情の研究と同時進行して行く予定である。

最後に、実験に御協力頂いた慶應義塾大学大学院政策・メディア研究科マルチモーダルインタフェースプロジェクトの皆さんに感謝申し上げます。

## 参考文献

- [1] 山田奨治, “脳波キーボードの入力速度向上手法と評価”, 電子情報通信学会論文誌, Vol.J79-A No.2, 1996.
- [2] 広瀬啓吉, 高橋登, 藤崎博也, 大野澄雄, “音声の基本周波数パターンにおける話者の意図・感情の表現”, 信学技報, HC94-41, pp.33-40, 1994.
- [3] 志村洋子, 今泉敏, “乳児は伝える”, 日本音響学会誌, Vol.52, No.7, pp. 547-551, 1996.
- [4] 土佐尚子, 中津良平, “芸術とテクノロジー”, 第4回 日立中研研究会予稿集, pp.75-80, 1996.
- [5] 坂巻資浩, 武者利光, 安村通晃, 脳波計測による感情解析を用いたヒューマンインターフェイス評価手法について, 第12回ヒューマンインタフェースシンポジウム, pp.665-670, 1996.
- [6] 飯田朱美, 伊賀聰一郎, 安村通晃, “韻律にみられる感情表現について”, 第54回 情報処理学会全国大会発表予定, 1997.
- [7] Banse, R. and Scherer, K.R., “Acoustic Profiles in Vocal Emotion Expression”, In *Journal of Personality and Social Psychology*, Vol. 70, No. 3, pp. 614-636, 1996.
- [8] 伊藤英一, 大橋正洋, 玉垣努, 北村啓, “在宅重度障害者の遠隔コミュニケーションによる生活と意識の変化”, 情報処理学会論文誌, Vol.38, No.5, 1996.