

## F<sub>0</sub>分布に基づく親近感特徴量の話者適応

渋谷 貴紀 益永 祐吾 川端 豪

関西学院大学理工学部 〒669-1337 兵庫県三田市学園2-1

E-mail: {t.shibutani, bju84828}@ksc.kwansei.ac.jp, kaw@kwansei.ac.jp

あらまし 人間と音声対話エージェントが円滑に会話を行うためには相互関係の理解が重要な要素であり、その一例として韻律情報を用いて親近感の高低を判定する取り組みを行っている。前報において、特定の被験者に対して基本周波数のヒストグラム形状が親近感の判定に有効であることを示した。しかし、被験者によって常用基本周波数は異なるので、この結果をそのまま不特定話者に適用することはできない。そこで、本報告では会話のある段階において、発話の初期部分のヒストグラムの最頻値を初期値とするEMアルゴリズムによって、未知被験者に対するヒストグラム特徴量を適切に抽出する。提案する手法を7名の被験者の対話音声に適用し、約88%の親近感判定精度を得た。

キーワード 親近感, 韻律, 感情分析, 音声対話システム

## Speaker Adaptation of Intimacy Factors based on F<sub>0</sub> Histograms

Takanori SHIBUTANI Yugo MASUNAGA and Takeshi KAWABATA

School of Science and Technology, Kwansei Gakuin University

2-1 Gakuen, Sanda City, Hyogo, 669-1337 Japan

E-mail: {t.shibutani, bju84828}@ksc.kwansei.ac.jp, kaw@kwansei.ac.jp

**Abstract** For realizing the dialog agent which gets along with human beings, it is necessary to detect the user's mental state to the agent, e.g. the intimacy factor with it. In our previous work, we showed that speaker's intimacy factors reflected to the histograms of the fundamental frequency (F<sub>0</sub>). Because the frequently-used F<sub>0</sub> range strongly depends on the speaker, the shape of F<sub>0</sub> histogram related to the intimacy factor may differ from that of another speaker. In this paper, speaker adaptation of the intimacy factors based on F<sub>0</sub> histograms is proposed. Roughly speaking, the histogram feature for any persons can extract appropriately according to EM algorithm whose initial value is embraced F<sub>0</sub> histogram's mode value in the early time of utterances. The intimacy extraction accuracy for 7 persons is about 88% based on the proposed approach.

**Keyword** intimacy features, prosody, emotion analysis, spoken dialog systems

### 1. はじめに

昨今、コンピュータの急速な普及により、人間がコンピュータと接する機会は増え、ありとあらゆる場所で使用することが求められている。しかしながら、コンピュータの操作は、初めて利用するユーザにとって負担になることがある。そのため、誰でも簡単に使用できるという観点から、音声インタフェースは有効であり、音声認識、合成、理解、対話など様々な分野で研究が盛んに行われている。それらの研究の最終的な目標の一つとして、音声対話によって人間とコンピュータが相互的なコミュニケーションを行うことが挙げられる。そのようなシステムでは、コンピュータが紋切り型の応答を行うのではなく、人間同士が行うようなより自然で円滑な対話が可能でなければならない。そのような対話の一例として、エージェントがユーザ

の心的要因を理解し、エージェントが自己の心的要因などの内部状態を動的に表出させることが挙げられる。

そのため、音声に含まれる韻律情報に基づく、様々な心的要因に関する研究がなされてきた。例えば、韻律情報から感情分析を行っている先行研究として、平静からのずれを用いて、判別する方法がある[1]。また、平静状態との比較が困難な場合を考え、多くの研究においては過去に観測された音声に基づいて正規化する方法が用いられる[1,2]。これらの研究では基本周波数やパワーなどが特徴量として分析が行われているが、基本周波数の時間的変化に注目した分析[3]や発話速度に注目した分析[4,5]も行われている。また、人間がラベル付けを行い、予め分類した音声に基に分析されている例もある[6]。このような感情分析に用いられる韻律特徴量については、A.Kieblingらの研究において詳しく言及されている[7]。分析方法については様々な

統計的手法が試みられており、独立成分分析や主成分分析などを用いた分析[8-11]、HMMやニューラルネットワークを用いた分析[12,13]などがある。ニューラルネットワークを用いて感情分析を実装した例としてはニューロベビーがある[14]。その他の実装例として、話者の発話態度が肯定的か否定的かに基づいて動的に対話を進めるロボットについても研究が行われている[15]。また、話者の快・不快に注目した例もある[16]。この他にも話者の心的要因に関する研究については多岐にわたる研究がなされている[17-19]。さらに、これらを包括的に扱った研究成果も報告されている[20]。

人間同士が会話を行う時、これら以外にも様々な要素を考慮して会話を行う。その一つに人間の「親近感」がある。筆者らは前報において、特定被験者に対して基本周波数のヒストグラム形状が親近感の判定に有効であることを示した[21]。しかし、前報において提案した二つの特徴量のうち、基本周波数の分布形状に基づく親近感特徴量は被験者によって常用基本周波数が異なり、そのまま不特定話者へは適用できない。

そこで、本報告では、提案する親近感特徴量を未知の話者の音声から適切に抽出する検討を行った。具体的には、ある段階において、発話の初期部分のヒストグラムの最頻値をEMアルゴリズムの初期値とする。それ以降については逐次的に初期値の更新を行うことによって、未知被験者に対するヒストグラム特徴量を適切に抽出することで、不特定話者に対して有効な親近感判定を行えるようになる。

## 2. 親近感判定の必要性

人間同士が会話を行う時、会話相手に対する心的な要因を考慮することで、適切な会話を行う。例えば、親近感のようにお互いの間柄を考慮することは、円滑な会話を行う上で重要な要素になると考えられる。

人間とコンピュータが音声対話を行う場合についても、人間同士の会話で行われるように、対話エージェントがユーザの親近感に応じて動的に振舞うことで、より柔軟な会話を行うことができると考えられる。そのような一例として、親近感が高いユーザに対しては、さらに身近に感じてもらえるように、そのような場合にしか得られない情報を教えてくれることが考えられる。また、親近感が低いユーザに対しては、会話の初期段階ではやや堅い言動を行うが、徐々に親近感を持ってもらえるような振舞いをすることが考えられる。このような知的な振舞いは、インタラクティブな対話システムを実現するために考慮すべきである。

上記のような対話システムを実現するためには、適切な親近感判定が必要である。筆者らは前報において、特定被験者に対して基本周波数の分布形状が親近感の判定に有効であることを示した[21]。しかし、被験者によって常用基本周波数は異なるので、この結果をそ

のまま不特定話者に適用することはできない。

本研究では、前報において提案した親近感特徴量について、不特定話者へ適応することを目指す。親近感特徴量の一つとして、基本周波数の分布形状がある。会話が進むに従って蓄積されるヒストグラムに対して、二つのピークからなる混合正規分布を仮定し、特徴量抽出を行っている。特徴量抽出の過程においてEMアルゴリズムを用いるが、適切な初期値を設定する必要がある。そのため、会話のある段階において、発話の初期部分のヒストグラムの最頻値を初期値とし、それ以降については直前の平均値をEMアルゴリズムの初期値とすることで、未知被験者に対する適切な特徴量抽出を目指す。

## 3. $F_0$ 分布に基づく親近感特徴量の判定

### 3.1. 親近感の分類

親近感には瞬間的に感じる局所的親近感と長期に渡って変わらない大局的親近感の異なる二つの性質の親近感が存在していると思われる。局所的親近感とは発話単位あるいは数発話間で瞬間的に現れるものであり、大局的親近感とは数年来の友人のように長期的に不変な親近感であるとする。

伊藤らは、馴れ馴れしい声と緊張した声によって数発話から親近感判定を行っている[2]。これは局所的親近感に分類できると思われる。しかしながら、人間同士の雑談において、親近感を高く感じている者同士が常に馴れ馴れしい声で会話を行い、低く感じている者同士が常に緊張した声で会話を行っているとは限らない。局所的に現れる親近感を捉えることは必要であるが、会話全体を通して、大局的に親近感を判定することが必要である。本報告では、この大局的親近感を以降「親近感」と記す。

### 3.2. 親近感を表わす特徴量

本節では、前報において提案した  $F_0$  分布のヒストグラム形状に基づく親近感特徴量について簡潔に述べる。

#### 3.2.1. 継時的観点の導入

大局的親近感を捉える特徴量を抽出するために、継時的な観点を導入する。具体的には発話単位の韻律情報から判定を行うのではなく、発話単位の韻律情報を蓄積して得られた分布形状(ヒストグラム形状)から特徴抽出を行う。次項以降に前節で提案した二つの特徴量を説明する。

### 3.2.2. 基本周波数の分布形状

第一の特徴量は、基本周波数のヒストグラムを二峰の混合正規分布で表現した時、常用基本周波数に対するピークの分散値である。

継続的な観点から十分に会話が進行した時点における、実際に観察された特定被験者の基本周波数のヒストグラム形状を図1に示す。図1は横軸が度数(使用頻度)、縦軸が基本周波数を表わしている。親近感が高い被験者との会話においては、ある基本周波数(100Hz)をピークに、高い周波数に向けてなだらかに減少する傾向があった(図1(a))。親近感が低い場合では、高い場合とほぼ同じ基本周波数(105Hz)にピークがあるが、少し高い基本周波数(125Hz)にもピークができる傾向があり、そのピークは度数が低くなっている(図1(b))。親近感の高低に関わらず現れる基本周波数(100~105Hz)が常用基本周波数であり、親近感が低い場合にはそのピークに対する分散値が小さくなる。

本特徴量は抽出過程においてEMアルゴリズムを用いて、各ピークに対する値を計算するが、適切な初期値を与えなければ局所解となる可能性がある。話者によって常用基本周波数は異なるため、未知話者へ適用する場合、適切な初期値の設定(話者適応)を行う必要がある。

### 3.2.3. 発話内における基本周波数の変化量に対する分布形状

第二の特徴量は発話内の基本周波数の最大値と最小値の差に対するヒストグラムが、指数分布に従うかどうかである。本特徴量は、基本周波数の相対値であるため、話者の影響を受けにくいと思われる。そのため、今回の報告では、本特徴量の話者適応は行わないが、後述する評価実験では本特徴量も併用しているため、本節で簡潔に説明を行う。

実際に観測された特定被験者の発話内における基本周波数の最大値と最小値の差に対するヒストグラム形状を図2に示す。親近感が高い場合には基本周波数の変化量が0~20Hzの割合が低く、20~50Hzに高くなっている(図2(a))。また、親近感が低い場合には高い場合とは対照的に変化量が0~20Hzの割合が高くなっており、指数分布のような形状になった(図2(b))。

そこで、このような形状の違いを定量化するために、指数分布を仮定し、式(1)に基づいて指数分布とヒストグラムの二乗誤差を求めた。

$$E = \sum \left( \frac{1}{\theta} e^{-\frac{1}{\theta} y_i} - \frac{F_i}{C} \right)^2 \quad (1)$$

$y_i$ : ヒストグラムのクラスiの級中値

$F_i$ : クラスiの度数

C: 正規化定数

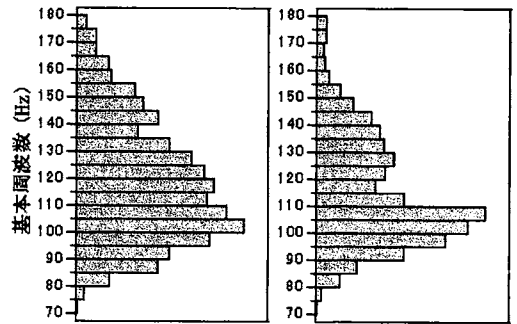
その結果、親近感が低い場合は高い場合に比べ、二乗誤差の値が小さくなる傾向があった。つまり、親近感が低い場合は、より指数分布に近似される。

## 4. 親近感特徴量の話者適応

前節で述べたように、基本周波数の分布形状に基づく親近感特徴量は、推定の初期値の影響を受けやすいため、標準的な基本周波数を仮定し、不特定の話者に適用するには無理がある。そこで、本報告ではこの特徴量を未知話者に対して、適応することを考える。

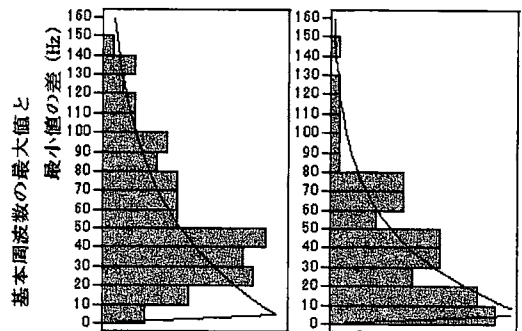
### 4.1. EM アルゴリズムの適切な初期値設定

3.2.2 節で述べた基本周波数の分布形状に基づく特徴量は、二つからなる混合正規分布を仮定し、主として利用する基本周波数に対する正規分布(第1ピーク)の分散値を特徴量としている。特徴量抽出にはEMアルゴリズムを用いるため、平均値、分散値、混合比について初期値を与える必要がある。



(a) 親近感が高い場合 (b) 親近感が低い場合

図1 基本周波数の分布形状



(a) 親近感が高い場合 (b) 親近感が低い場合

図2 発話内の基本周波数の変化量に対する分布形状

ここで式(2)は混合正規分布を表わし、式(3)~(7)はEMアルゴリズムを表わしている。式(2)~(7)において、 $k$ は正規分布の数(ここでは $k=2$ )、 $\xi_i$ は混合比、 $\mu_i$ は平均値、 $\sigma_i^2$ は分散値、 $x_j(j=1, \dots, N)$ は観測値を表わしている。また、 $\omega_{ij}$ は各観測値 $x_j$ に対する未知変数 $z$ の条件付き確率である。式(2)で示される混合正規分布は、式(3)~(7)で示されるEMアルゴリズムによって解くことができる。

$$p(x; \theta) = \sum_{i=1}^k \xi_i \phi(x; \mu_i, \sigma_i^2) \quad (2)$$

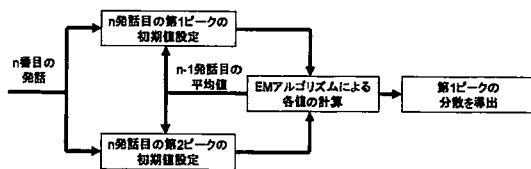


図3 EMアルゴリズムの初期値の逐次更新

表1 主観評価による親近感判定にかかる平均時間

|       | person1 | person2 | person3 | all   |
|-------|---------|---------|---------|-------|
| first | 16.8    | 56.8    | 39.3    | 37.6  |
| last  | 119.3   | 211.5   | 128.9   | 153.2 |

(単位: 秒)

| EMアルゴリズム |  |
|----------|--|
| STEP 1   | $\omega_{ij}(t) = p(z   x_j; \theta^{(t)})$ $= \frac{\xi_i \phi(x_j; \mu_i^{(t)}, \sigma_i^2{}^{(t)})}{\sum_{i=1}^k \xi_i \phi(x_j; \mu_i^{(t)}, \sigma_i^2{}^{(t)})} \quad (3)$ |
| STEP 2   | $\xi_i^{(t+1)} = \frac{1}{N} \sum_{j=1}^N \omega_{ij}^{(t)} \quad (4)$   |
| STEP 3   | $\mu_i^{(t+1)} = \frac{1}{N \xi_i^{(t+1)}} \sum_{j=1}^N \omega_{ij}^{(t)} x_j \quad (5)$   |
| STEP 4   | $\sigma_i^2{}^{(t+1)} = \frac{1}{N \xi_i^{(t+1)}} \sum_{j=1}^N \omega_{ij}^{(t)} x_j^2 - \mu_i^2{}^{(t+1)} \quad (6)$  |
| STEP 5   | $\text{if } \sum_{i=1}^k (\mu_i^{(t)} - \mu_i^{(t+1)})^2 \geq \varepsilon \text{ then STEP 1} \quad (7)$   |

## 5. 評価実験

### 5.1. 実験条件

本報告においては、被験者間で明確に親近感が異なることが予想される様々な被験者の会話を収録した。被験者はそれぞれ異なる防音室に入り二者間で会話を行い、直接相手の姿は見えないようになっている。被験者の発話はマイクに向かって行い、相手の声のみがヘッドフォンから聞こえるようになっている。

防音室の暗騒音は36dBである。マイクは単一指向性のダイナミックマイクロフォンを使用し、A/D変換機、ミキサーを経て、被験者のヘッドフォン・録音用のPCへ出力されている。PCで録音される音声のサンプリング周波数は44.1kHzである。また、被験者同士の声は、ヘッドフォンを通して聞こえるようになっているので、録音された音声には話者ごとの音声は独立して収録されている。

会話はあるテーマに沿って約5分間ずつ行い、2ないし3テーマについて、約10~15分間の雑談を行ってもらった。これを一つの会話セッションとする。会話が開始してから5分経過したことは外部より合図を送り被験者に知らせ、現在のテーマが一段落したら、次のテーマに移るように指示した。テーマの設定については、事前に様々なテーマについて興味の有無を5段階で評価してもらい、調査したアンケート結果に基づき、興味の差があまりないように選択した。

分析対象とする音声は、次節で述べる人間による主観評価の結果、親近感の高低について評価者全員の最終的な意見が一致した音声を対象とする。本報告においては、男性6名、女性1名の合計7名10セッションについて分析を行う。10セッションの内訳は、親近感が高いと判断された会話6セッション、親近感が低いと判断された4セッションとなっている。

このような手順において、EMアルゴリズムに適切な初期値を設定することにより、不特定話者への適応を行う。

特徴量は継続的に抽出されるため、会話開始後10発話目までを学習期間とする。11発話目が入力されると、10発話目が終わった時点での累積された基本周波数のヒストグラムから、最も頻度の高い級中値を第1ピークの平均値として初期値に設定する。他方の正規分布(第2ピーク)については、被験者ごとのヒストグラムを観察したところ、平均して約25Hz高い基本周波数に第2ピークの平均値が存在したため、推定した第1ピークから25Hz高い値を平均値として採用した。分散値については、特定被験者の分散値を参考に設定し、混合比については両ピークに対して0.5と設定した。それ以降のn発話目の特徴量抽出には、図3に示すように、n発話目が入力されると、n-1発話目に導出された平均値を初期値として設定し、EMアルゴリズムによって特徴量の抽出を行う。また、分散値と混合比については11発話目と同様の設定を行った。

表 2 評価実験結果

| utterances | (common)     | before adaptation |                  | after adaptation |                  | ideal case (*1) |                  |
|------------|--------------|-------------------|------------------|------------------|------------------|-----------------|------------------|
|            | $\Delta F_0$ | $F_0$             | $F_0+\Delta F_0$ | $F_0$            | $F_0+\Delta F_0$ | $F_0$           | $F_0+\Delta F_0$ |
| 11-20      | 45           | 67                | 69               | 54               | 57               | 72              | 75               |
| 21-30      | 75           | 72                | 95               | 67               | 90               | 68              | 91               |
| 31-40      | 79           | 49                | 87               | 60               | 98               | 60              | 98               |
| 41-50      | 81           | 49                | 86               | 69               | 99               | 69              | 99               |
| 51-60      | 85           | 42                | 74               | 63               | 91               | 62              | 90               |
| 61-70      | 81           | 32                | 69               | 52               | 89               | 52              | 89               |
| 71-80      | 82           | 48                | 78               | 60               | 90               | 60              | 90               |
| 81-90      | 80           | 50                | 80               | 58               | 88               | 60              | 90               |
| all        | 76.00        | 51.13             | 79.75            | 60.38            | 87.75            | 62.88           | 90.25            |

(\*1) 人間が目視により常用基本周波数の初期値を与える。

(単位：パーセント)

## 5.2. 人間による主観評価

今回の評価実験の正解ラベル(親近感の高低)は、人間による主観評価を基準にすることにした。

前節の実験条件下で収録された音声について、被験者3名が親近感の高低について主観評価を行った。評価を行う際、音声は二者間で行われた会話のうち、一方のみの音声しか聞こえないようになっている。評価はパソコン上で行い、音声の流れると共に、評価を行うためのツールの値が起動するようになっている。評価ツールの値は5段階になっており、「unknown(わからない)」「very low(とても低い)」、「little low(少し低い)」、「little high(少し高い)」、「very high(とても高い)」のように設定した。初期状態は「unknown(わからない)」となっており、被験者はキーボードの矢印キーから適宜、値を変更することができる。また、被験者が値を変更した時刻は、ファイルに出力され、操作した時刻と評価した値が記録されるようになっている。

主観評価の結果を表1にまとめる。表1は被験者ごとに、会話開始から初めて判定を行うまでに要した平均時間が「first」欄、最終的な判定を行うまでに要した平均時間が「last」欄となっている。また、被験者3名が判定までに要した時間の平均についても合わせて「all」欄に表記する。この結果より、第三者が他人の会話を聞いた時、親近感の高い相手と会話しているか、あるいは低い相手と会話しているか判定するためには、分単位の時間が必要であると考えられる。

## 5.3. 結果と考察

本研究で利用する親近感特徴量は、継時的な観点に基づいているため、各時点における瞬時的な値からだけでは判定できない。そこで、学習データは前もって

会話の開始時点からn発話目までの該当する特徴量を記憶しておき、各時点でのユークリッド距離に基づいて、親近感の高低を判定する。さらに二つの特徴量を併用するため、式(8)に基づき、学習データの特徴量*i*の全データに対する平均値で補正したユークリッド距離を用いて判定を行った。

$$\sum_{i=1}^N \frac{(x_i - \mu_i^{(j)})^2}{\mu_i^{(all)2}} \quad (8)$$

$x_i$ : 判別データの特徴量 *i*

$\mu_i^{(j)}$ : 学習データの特徴量 *i* のカテゴリ *j* に対する平均値

$\mu_i^{(all)}$ : 学習データの特徴量 *i* の全データに対する平均値

学習データは特定被験者のものを使用する。データ数は親近感が高い被験者との会話6セッションと親近感が低い被験者との会話4セッションとなっている。判別データの中には、特定被験者のものは含まれていない。

一発話ごとに判別データの特徴量を更新しつつ、親近感の高低についての判定を行った結果を表2に示す。会話の進行に伴い、判定精度がどう変化していくかを観察するために、表は10発話ごとの判定率を示している。また、基本周波数の変化量に基づく特徴量を利用した場合を「 $\Delta F_0$ 」欄、基本周波数の分布形状に基づく特徴量のみを利用して判定した場合を「 $F_0$ 」欄、二つの特徴量を利用して判定した場合を「 $F_0+\Delta F_0$ 」欄と表記している。

基本周波数の分布形状に基づく特徴量については、話者適応を行わない場合を「before adaptation」欄、本報告で提案した話者適応を行った場合を「after adaptation」欄、人間が目視により常用基本周波数の初

期値を与えた場合を「ideal case」欄の三つに分類し、二つの特徴量を用いて判別を行った結果「 $F_0 + \Delta F_0$ 」を合わせて表記している。また、一発話内での基本周波数の変化量に基づく特徴量「 $\Delta F_0$ 」は、基本周波数の相対値になっているため、話者適応は行わず、共通して判定に利用している。

本報告において提案した手法は、11~20発話では不安定になっているが、それ以降は安定して特徴量を抽出していることが分かる。また、話者適応を行わない場合、会話が進むにつれ判定率が低下し、提案手法と比べると、全体の判定率は低くなった。さらに、提案手法と人間が目視により常用基本周波数の初期値を与えた場合を比べると、提案手法は判定初期段階(11~20発話)では判定率が低くなっているが、それ以降ではほぼ同等の判定率を得ることができた。そのため、本手法による基本周波数の分布形状に基づく特徴量の適切な導出が、全体として親近感判定の精度向上に寄与していることが確認された。

## 6. むすび

本研究では将来、人間と音声対話エージェントの会話を円滑に行う手段として「親近感」に注目し、特定話者から得られたデータを基に、不特定話者への適応を行った。

まず、本研究で扱う親近感の定義を行った。これは前報において不明確だった部分の補足となっている。本研究では、数年来の友人のように不変的な親近感を大局的親近感と定義した。このような大局的親近感を表わす特徴量として、特定話者の親近感判定に有効であった二つの特徴量、基本周波数の分布形状に基づく特徴量と発話内における基本周波数の変化量の分布形状に基づく特徴量について簡潔に説明した。

そのうち、基本周波数の分布形状に基づく特徴量は、基本周波数の絶対値を使用している。しかし、被験者ごとに常用基本周波数は異なるため、本報告において適切な抽出方法を提案した。

その結果、話者適応を行わない場合に比べて、提案手法を用いて評価を行うと約9%判定精度が向上し、最終的には約88%の親近感判定精度が達成された。

## 文 献

- [1] 重永實, "感情の判別分析からみた感情音声の特性," 信学論(A), Vol. J83-A, No. 6, pp. 726-735, Jun. 2000.
- [2] 伊藤亮介, 駒谷和範, 河原達也, 奥乃博, "ロボットとの音声対話におけるユーザの心的状態の分析," 情処研報, Vol. 2003, No. 14, 2003-SLP-45, pp. 107-112 Feb. 2003.
- [3] 前川喜久雄, "韻律によるパラ言語情報の表出に関する準備的考察," 人工知能学会研究会資料, SIG-SLUD-9602-06, pp. 31-36, Oct. 1996.
- [4] 川波弘道, 広瀬啓吉, "態度・感情音声における韻律的特徴の考察," 信学技報, SP97-67, pp. 73-80, Nov. 1997.
- [5] 武田昌一, 栢谷綾香, 橋澤保輝, 加藤修一, 大山玄, "感情を含む案内音声の韻律的特徴の解析," 音響誌, Vol. 60, No. 11, pp. 629-638, Nov. 2004.
- [6] 小林聡, 北澤茂良, "日本語の自然対話音声におけるパラ言語情報の検討," 音響誌, 56巻7号, pp. 467-476, Jul. 2000.
- [7] Kiebling, A., Kompe, R., Batliner, A., Niemann, H. and Noth, E., "Classification of Boundaries and Accents in Spontaneous Speech," In Proceedings of the CRIM/FORWISS Workshop, pp. 104-113 Aug. 1996.
- [8] 森山剛, 斎藤英雄, 小沢慎治, "音声における感情表現語と感情表現パラメータの対応付け," 信学論(D-II), Vol. J82-D-II, No. 4, pp. 703-711, Apr. 1999.
- [9] 廣瀬陽介, 平原誠, 永野俊, "教師付き独立成分分析による音声の感情認識," 信学技報, NC2002-152, pp. 113-118, Mar. 2003.
- [10] 金城良治, 土本悠治, 長山格, "音声インターフェイスにおける感情判別の試み," 信学技報, HIP2001-86, pp. 49-54, Jan. 2002.
- [11] 白澤敏行, 山村毅, 田中敏光, 大西昇, "音声に込められた感情の判別," 信学技報, HIP96-38, pp. 79-84, Jan. 1997.
- [12] 高橋和彦, 中津良平, "ニューラルネットワークによる感情音声認識に関する一考察," 日本機械学会論文集(C編), 68巻672号, pp. 123-129, Aug. 2002.
- [13] 刀根優子, 萩原昭夫, 柴田浩, "音声対話システムのためのHMMに基づく感情判別," 信学技報, SP2000-22, pp. 47-53, Jun. 2000.
- [14] 柿本正憲, 土佐尚子, 森淳一, 真田麻子, "ニューロベビー," 音声から感情を認識して反応する表情合成システム," テレビ学技報, Vol. 16, No. 33, pp. 7-12, Jun. 1992.
- [15] 藤江真也, 江尻康, 菊池英明, 小林哲則: 肯定的/否定的発話態度の認識とその音声対話システムへの応用, 電子情報通信学会論文誌(D-II), vol. J88-D2, no. 3, pp. 489-498 Mar. 2005.
- [16] 目良和也, 市村匠, 相沢輝昭, 山下利之, "語の好悪度に基づく自然言語発話からの情緒生起手法," 人工知能学会論文誌, Vol. 17, No. 3, pp. 186-195 May. 2002.
- [17] 曾我部優子, 寛一彦, 河原英紀, "感性情報に曖昧さがある場合の音声の心理的評価とその物理的特性," 信学技報, SP2002-185, pp. 37-42, Mar. 2003.
- [18] 門谷信愛希, 阿曾弘具, 鈴木基之, 牧野正三, "音声に含まれる感情の判別に関する検討," 信学技報, NLC2000-34, SP2000-82, pp. 43-48, Dec. 2000.
- [19] 佐藤貴志, 遠藤祐樹, 金子正人, 竹内惇, 藤本洋, "音声に含まれる感情情報の認識に関する研究," 信学技報, HIP2001-60, pp. 25-29, Dec. 2001.
- [20] 広瀬啓吉 編著, "韻律と音声言語情報処理 アクセント・イントネーション・リズムの科学," 丸善株式会社(第1版), pp. 9-64, Jan. 2006.
- [21] 渋谷貴紀, 川端豪, "音声対話システムのための親近感特徴量の探索," 信学技報, Vol. 106, No. 122, SP2006-10, pp. 25-30 Jun. 2006.