2005－SLP－57（17）
2005／7／16

ユーザ感情理解に向けた実環境音声情報案内システムの収集発話分析

大前 壮司 †　西村 竜一 †　河原 英紀 †　入野 俊夫 †

あらまし　音声対話システムにおいて，ユーザがシステムに抱く感情を理解することは円滑な対話を実現する上で重要となる．本稿では，奈良県生駒市北コミュニティセンターの音声情報案内システム「たけまるくん」のフィールドテストを通じて収集したユーザ発話を分析することで，システムによる感情理解の実現性を検討する．まず，収集発話を 16 個の基本感情を用いて被験者 2 名により 5 段階評定した．評定結果を因子分析したところ，ネガティブ及びポジティブな感情を示す因子の存在を確認することができた．続いて，ユーザ感情理解の実現に向け，因子分析から算出した因子得点と音声特徴量との相関を調査している．今回，基本周波数及びパワーを特徴量として用いたが，顕著な相関を得ることはできなかった．

キーワード 感情理解，音声対話，実環境，因子分析

Analysis for emotion understanding with utterance collection in spoken dialogue system

Souji OMAE† **Ryuichi NISIMURA**† **Hideki KAWAHARA**† **Toshio IRINO**†

Abstract　Understanding emotions that users hold is becoming important for realizing smooth conversations in spoken dialogue systems. This study discusses the actualities of an automatic emotion understanding by analyzing actual users' utterances collected via field testing our spoken dialogue system "Takemaru-kun". Two testers have carried out the five grade rating with 16 basic emotions to the collected utterances. The factor analysis on the rating result indicated the existence of two factors concerning negative or positive emotions. For realization of the emotions understanding, we have been investigating the correlation between the factors and acoustic features in user's voices. In this paper, the results showed that the factors have no remarkable correlation with the fundamental frequency and the power.

Key words　Emotion understanding, Spoken dialogue, Real environment, Factor analysis

1　はじめに

近年，音声認識技術の向上により，音声インタフェース・擬人化エージェント・ロボットへの応用が注目されている．例えば，音声対話技術コンソーシアムでは，音声認識・音声合成・顔画像合成・対話統合などの音声対話に不可欠な基本ソフトウエア群を提供している [1]．また，奈良県生駒市北コミュニティーセンターの音声情報案内システム「たけまるくん」[2] では，一問一答形式による音声対話により，センターの来訪者にセンター内やセンター周辺，天気，時間などの案内を行っている．たけまるくんには，大人と子供の発話音声の識別，咳，笑い声などの非音声識別を応用した対話機能が実装されており，年々，システムの利用者が増えている．以上のように，一般の人々が音声認識を利用する機会は増えており，ユーザとシステムとの親和性をはかることや利用者の発話に対する柔軟な応答を実現することが今後の重要な検討事項となっている．

そこで，我々は，ユーザとシステムの親和性を高めることを目的に，ユーザが音声対話システムに抱く感情について注目した．人間は，対話相手の感情を認識する際，音声以外に，顔の表情，身振り手振り，間，言語的な内容，意図などを手がかりにしている．Mehrabian[3] によると，相手にメッセージを伝える際の感情は，バーバル情報から 7 %，音声に含まれる非言語情報や顔の表情 (ノンバーバル情報) から 93 %が伝達される．本研究では，音声に含まれる代表的な非言語情報である感情が対話の円滑化を実現する上で重要なファクターであると考える．

本研究の目的は，生駒市北コミュニティーセンターの音声情報案内システム「たけまるくん」[2] を研究プラットホームとして用いて，ユーザがシステムに抱く感情を理解しながら対話することの実現性と有用性の検証することである．

これまでの音声を媒介とする感情についての研究では，感情を込めて発話された音声が含む特徴

† 和歌山大学大学院 システム工学研究科
Graduate School of Systems Engneering, Wakayama University

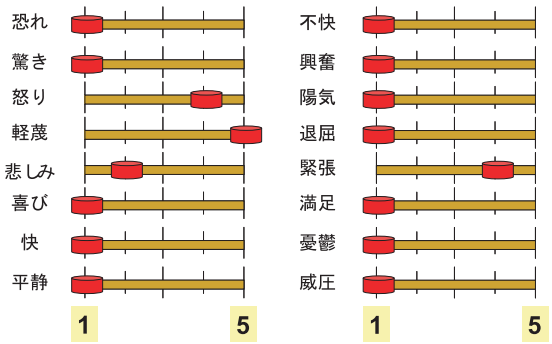

図 1: 多次元評定の例

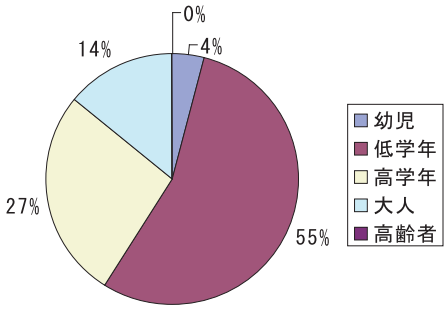

図 2: 発話における年齢層分布 (評定者 A)

量の振る舞いについての調査 [4] や特徴量とそこから知覚される感情とを対応づけたモデル等が提案されてきた [5]．また，Ververridis らは，音声の特徴量として，基本周波数やパワーの統計的特徴を利用し，音声の感情識別を行っている [6]．これらの研究では，声優が「喜び」「悲しみ」などの感情が浮かぶシーンをイメージして発話した音声を分析して個々の感情の特徴を調べている．しかし，声優の発話音声は，自然な対話の中での発話と異なるため，実際の感情音声の実態を反映していない可能性があることが問題であった．

一方，本研究では，実環境音声情報案内システム「たけまるくん」の録音機能で収録した来訪者発話を用いる．これらの発話は，ユーザとシステムとの自然な対話の中で収録されたものであり，声優が演技した感情音声では得ることができない感情音声の実態を見ることができる．

本稿では，ユーザがシステムに抱く感情を理解するための予備的な検証として，「たけまるくん」が過去に収集した来訪者発話を 16 個の基本感情で多次元評定し，因子分析した結果を報告する．また，因子分析により得られた因子得点と音声の特徴量である基本周波数とパワーの相関を調査した．

2　多次元評定

我々は，これまでにユーザがシステムに抱く感情を調査するために，「たけまるくん」が過去に収

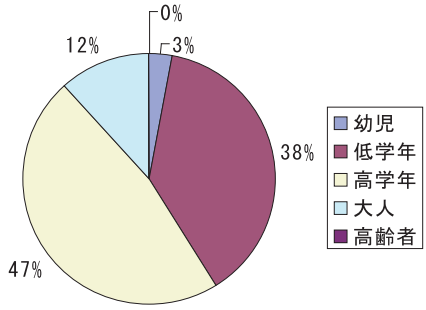

図 3: 発話における年齢層分布 (評定者 B)

表 1: 対象とする収集発話

評定者	A	B
発話数	3,134	515
収集期間	2003 年 4 月 1 日 ～ 2003 年 4 月 19 日	2003 年 4 月 1 日 ～ 2003 年 4 月 2 日

集した発話を Ekman[7] の基本 6 感情 (「喜び」「恐れ」「驚き」「軽蔑」「怒り」「悲しみ」) で分類した．多くの心理学者は，「本質的な感情として基本感情があり，それが結合したり，混合することで違った感情が生起する」と考えており，現代感情研究の主流的立場となっている．基本感情の種類は複数の説が存在し，心理学者によって様々であるが，Ekman の基本 6 感情は比較的多くの研究で利用されているため我々の研究では採用した．しかし，先行研究の結果，収集した発話は，Ekman の基本 6 感情では一意に分類することが不可能であった [8]．

そこで，Ekman の基本 6 感情と共に Schlosberg[7] や Russel[7] の感情モデルに含まれる基本感情も加えた 16 個の基本感情による多次元評定を新たに実施した．Ekman, Schlosberg, Russell の 3 つの感情モデルを用いたのは，それぞれが異なる基本感情を主張しており，様々な説の基本感情で多次元評定することで，より実態に即した分析が可能になると考えたためである．

多次元評定は，評定者が発話を聞いた時に感じた感情で，16 項目の各項目について 5 段階尺度で行った．評定者は成人男性 A, B の 2 人である．評定に使用した 16 個の基本感情の一覧を図 1 に示す．評定者が 1 つの発話を聞く毎に，ある項目の感情を含むと感じた場合には評定値 5，まったく含まないときには評定値 1 とした．

今回，対象とした発話音声は，雑音，不明瞭なものを除いたものを使用した．表 1 にその収集期間と発話数を示す．次に，図 2，図 3 に発話の年齢層分布を示す．この年齢層は，録音された音声を 1 人の作業者が聞き，主観的に判断したものである．なお，評定者 A, B で対象となる発話は同

じであるが，期間は異なる．この中で幼児は小学校入学前の子供の発話である．低学年は小学 3 年生ぐらいまであり，高学年は中学生ぐらいまでとしている．また，大人は高校生以上を想定している．高齢者は図 2，図 3 では 0 %となっているが，評定者 A では 3 発話，評定者 B では 1 発話存在していた．

3　因子分析による発話分析

3.1　因子分析

16 個の基本感情で多次元評定したデータを因子分析した．因子分析は多くの変量の持っている情報を小数個の潜在的因子によって説明する方法である [9]．因子分析のモデルを式 (1) に示す．

$$X_i = \sum_j \alpha_{ij} F_j + e_i \qquad (1)$$

$X_i(i = 1, \cdots, s)$ は観測変数であり，共通因子 $F_j(j = 1, \cdots, m)$ と独自因子 $e_i(i = 1, \cdots, s)$ の線形結合で表される．α_{ij} は因子負荷行列 $A(s \times m)$ の要素を表し，s と m はそれぞれ変数の数と因子数を表す．因子分析の目的は，因子負荷行列 A と共通因子得点 f を推定することである．

3.2　分析条件

本研究における観測変数は，1 発話毎に 16 個の基本感情で 5 段階評定した主観評価値である．共通因子の推定には最尤法を用いた．また，因子負荷行列の回転法にはバリマックス回転を用いた．共通因子推定際に必要である因子数 m は，多次元評定したデータの相関行列 R の固有値の中で 1 より大きいものを選び，低学年，高学年共に評定者 A, B それぞれ $m = 8, 6$ とした．

なお，大人の発話は，「平静」以外の感情の評定値が 1 のみであるため，因子分析することができなかった．このため，今回は大人の発話は対象外とした．子供は自分の感情に忠実に発話で表現し，感情音声を発話していた．一方，大人では感情を込めて発話することが少なく，分析が困難な結果となった．これは，マイクの前に立つ緊張感などに起因すると考えられる．

3.3　分析結果

各年齢別の分析結果を図 4，図 5, 図 6, 図 7 に示す．横軸, 縦軸はそれぞれ第 1 因子，第 2 因子の因子負荷量である．図 8 は評定者 A, B の累積寄与率を示す．

評定者 A, B の低学年，高学年の全てに，よく似た傾向の基本感情が，第 1 因子と第 2 因子の軸上に表れていることが分かる．第 1 因子には「不快」,「怒り」が表れており，第 2 因子は「快」,「陽気」,「喜び」を表現している．この結果から，第 1 因子，第 2 因子は潜在的にそれぞれネガティブな感情の因子，ポジティブな感情の因子を含んでいると考えられる．「平静」は第 1 因子，第 2 因子両方の負の因子負荷に配置している．その他の基本感情は原点周辺に表れており，第 1 因子，第 2 因子に影響を与えていない．

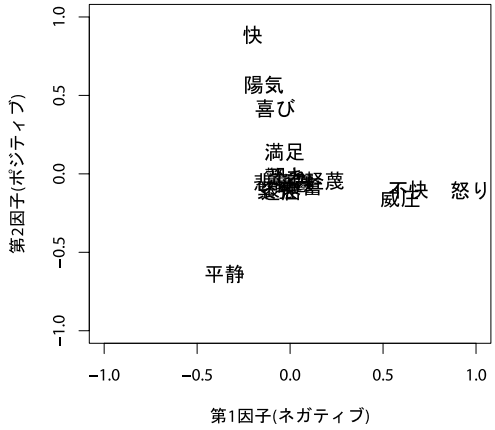

図 4: 因子分析結果 (評定者 A・低学年)

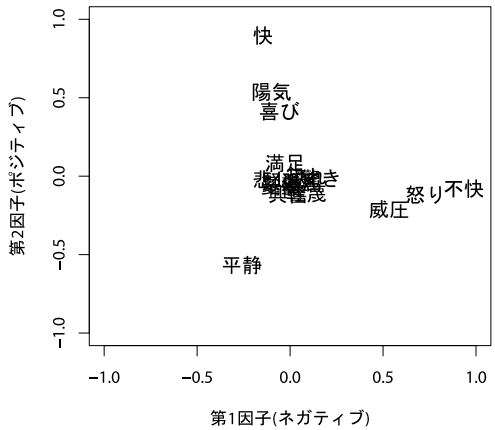

図 5: 因子分析結果 (評定者 A・高学年)

評定者 B において，低学年では「恐れ」「驚き」「喜び」「興奮」「憂鬱」「威圧」，高学年では「恐れ」「驚き」「興奮」「憂鬱」の評定値が 1 しか見られず，これらの基本感情は表出されなかった．これは，評定した発話数が少ないためであると考えられる．評定する発話数を増やした多次元評定をする必要がある．

以上より，今回のケースにおいては，第 1 因子，第 2 因子に見られたネガティブな感情の因子，ポジティブな感情の因子の 2 つに集約することができる可能性がある．また，因子得点を見ても，同様に第 1 因子，第 2 因子の軸周辺に配置していることを確認した．しかし，第 2 因子までの累積寄与率が低い．このため，第 3 の因子が存在する可能性もある．外的ノイズが分析に悪影響している可能性があるので，その改善の検討も必要である．

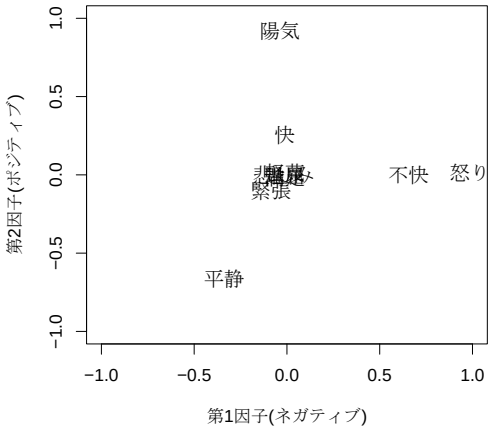

図 6: 因子分析結果 (評定者 B・低学年)

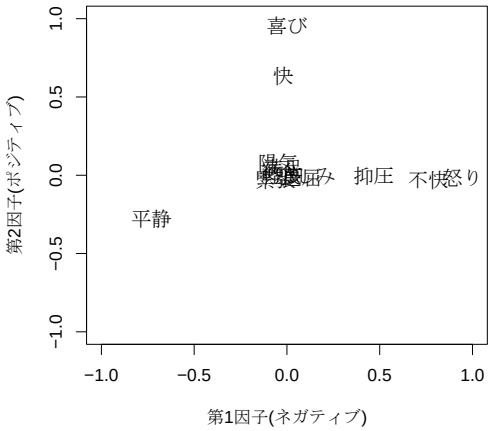

図 7: 因子分析結果 (評定者 B・高学年)

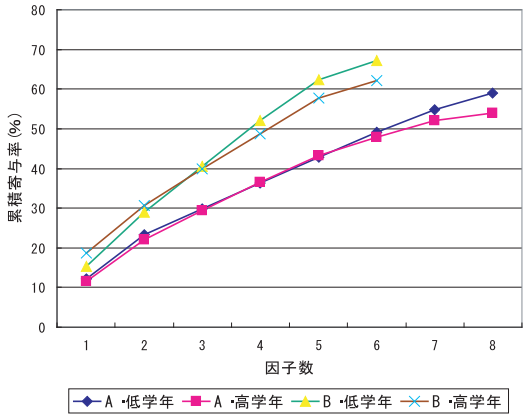

図 8: 累積寄与率 (%)

4 因子得点と音声特徴量の相関調査

本項では，音声情報案内システム「たけまるくん」が過去に収集した来訪者発話の特徴量と因子得点の相関を調査した．用いた因子得点は，評定者 A の多次元評定したデータを因子分析した際に算出した第 1 因子得点と第 2 因子得点である．

4.1 音声特徴量

音声の特徴量として，今回は，基本周波数とパワーの一次差分の発話内平均値を用いた．基本周波数とパワーの推定には，YIN[10] を用いた．YIN の出力結果を直接利用せず，フィルタ長 17ms のメディアンフィルタで平滑化している．基本周波数は，平滑化後に，有声・無声の判定のために，パワーの最大値から 30dB 低い部分を無声音部とし，有声部に対してのみ推定を行った．パワーの振幅値は，対数スケールに変換した．

4.2 結果

因子得点には，評定者 A がデータを因子分析して算出した第 1 因子得点，第 2 因子得点を使用した．

図 9, 10, 11, 12 は基本周波数の 1 次差分平均と因子得点の散布図である．この中で図 9 は，年齢層が低学年であり，横軸に第 1 因子得点，縦軸が基本周波数の 1 次差分平均とし，図 10 は，年齢層が低学年で，横軸に第 2 因子得点，縦軸は基本周波数の 1 次差分平均である．図 11 は，年齢層が高学年で，横軸に第 1 因子得点, 縦軸を基本周波数の 1 次差分平均としている．図 12 は，年齢層が高学年であり，横軸に第 2 因子得点，縦軸は基本周波数の 1 次差分平均を表している．また，図 13, 14, 15, 16 はパワーの 1 次差分平均と因子得点の散布図である．図 13 は, 年齢層が低学年で，横軸が第 1 因子得点，縦軸はパワーの 1 次差分平均を表し，図 14 は，年齢層が低学年で，横軸に第 2 因子得点であり，縦軸はパワーの 1 次差分平均である．図 15 は, 年齢層が高学年で，横軸に第 1 因子得点であり，縦軸はパワーの 1 次差分平均としている．図 16 は，年齢層が高学年で，横軸に第 2 因子得点，縦軸はパワーの 1 次差分平均である．

散布図を見ると，低学年・高学年共に，因子得点と基本周波数の 1 次差分平均・パワーの 1 次差分平均に顕著な相関はないことが分かる．

今回，因子得点と音声の特徴量に相関がなかった理由として，用いた来訪者発話データが複数の話者や様々な言葉によるものであることや言語的な内容などからも感情を判断していることが要因として考えられる．Mehrabian[3] は，人間理解には音声の特徴量の様々な統計的特徴以外に，言語的な内容，もしくは，評定者のこれまでの経験の差異が影響するとしている．今回の多次元評定では，音響的な特徴や言語的な内容も含めた総合的な判断で行っているため，音声の特徴量だけでは相関が得られなかったと考えられる．また，多次元評定は，音声の特徴量，もしくは，言語的な内容を基準に評定するが，発話ごとにそれぞれ判断比率が異なったのではないかと考えられる．

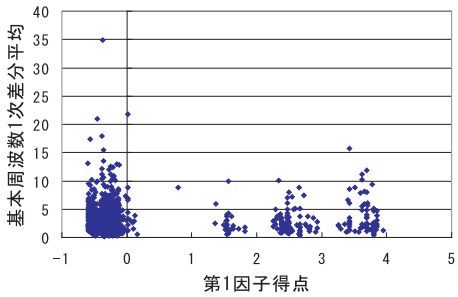

図 9: 第 1 因子得点と基本周波数 1 次差分平均 (低学年)

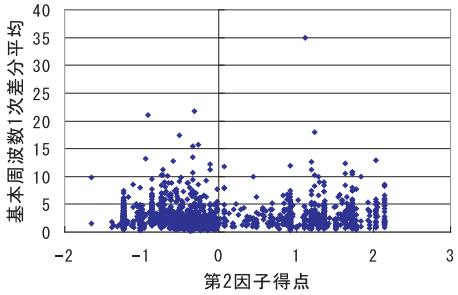

図 10: 第 2 因子得点と基本周波数 1 次差分平均 (低学年)

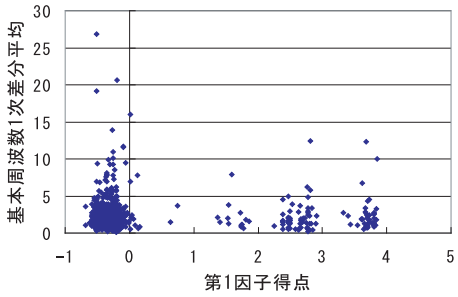

図 11: 第 1 因子得点と基本周波数 1 次差分平均 (高学年)

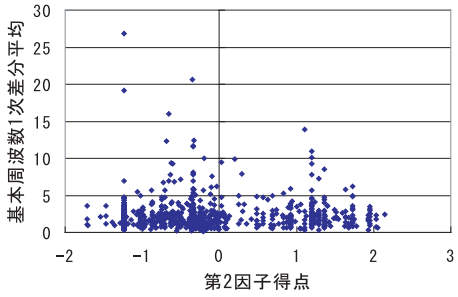

図 12: 第 2 因子得点と基本周波数 1 次差分平均 (高学年)

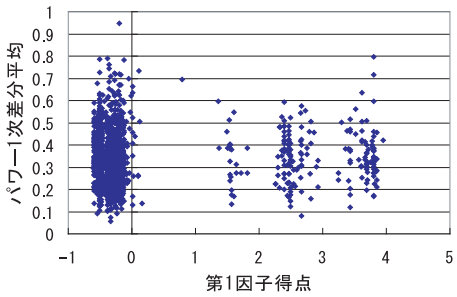

図 13: 第 1 因子得点とパワー 1 次差分平均 (低学年)

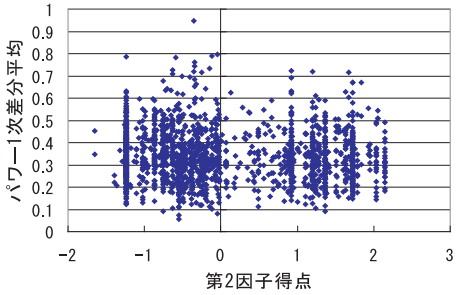

図 14: 第 2 因子得点とパワー 1 次差分平均 (低学年)

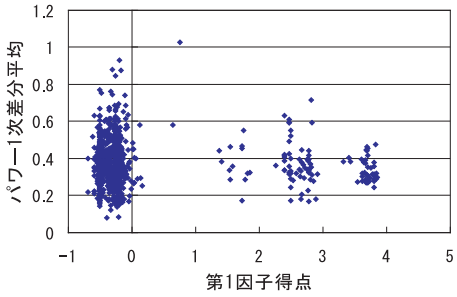

図 15: 第 1 因子得点とパワー 1 次差分平均 (高学年)

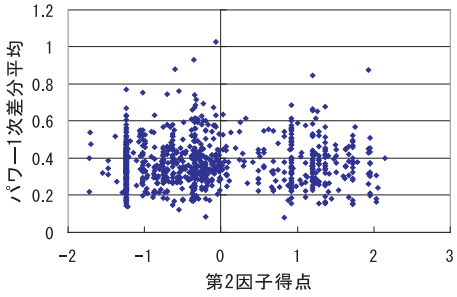

図 16: 第 2 因子得点とパワー 1 次差分平均 (高学年)

5 まとめ

本研究では，ユーザがシステムに抱く感情を理解するため，声優が意図的に発話した感情音声ではなく，実環境でユーザがシステムに対して発話した発話データの分析を行った．ここでは，生駒市北コミュニティーセンター音声情報案内システム「たけまるくん」が過去に収集した来訪者発話を 16 個の基本感情により多次元評定し，因子分析した．その結果，第 1 因子，第 2 因子には，それぞれ，ネガティブな感情因子，ポジティブな感情因子の潜在的な因子を確認することができた．

因子分析により算出した因子得点と音声の特徴量である基本周波数及び，パワーの 1 次差分平均に相関があるかを調査した．その結果，因子分析により算出した因子得点と使用した音声特徴量には相関を得ることができなかった．これは，来訪者発話を多次元評定した際，音声の特徴や言語的な内容も含めた総合的な判断で表したため，使用した音声特徴量では十分に表現できなかったためと考えられる．

今後の課題として，第 1 因子，第 2 因子までの累積寄与率が低学年，高学年共に低いので評定方法の再検討を行う．今回は，基本周波数とパワーの 1 次差分平均値を音声特徴量として用いたが，その他の統計的な特徴で調査する必要がある．Ververridis[6] らは，基本周波数，パワーや基本周波数, パワー 1 次差分平均の四分位範囲，メディアン値，基本周波数の局所的な最大値，最小値の平均，メディアン値など様々な統計的な特徴を利用したりしている．音声の非言語的な情報について着目したが，今後は言語的な情報についても考慮したい．

最終的な目標は，ユーザがシステムに抱く感情を理解する対話の実現である．今回得られた知見を活かし，ユーザの感情理解の実現を目指したい．

謝辞 本研究の一部は，文部科学省のリーディングプロジェクト e-Society 基盤ソフトウェアの総合開発「ユーザ負担のない話者・環境適応性を実現する自然な音声対話処理技術」によって行われた．

参考文献

[1] 嵯峨山，川本，下平，新田，西本，中村，伊藤，森島，四倉，甲斐，李，山下，小林，徳田，広瀬，峯松，山田，伝，宇津呂: ”擬人化音声対話エージェントツールキット Galatea”, 情報処理学会研究報告, 2002-SLP-45-10, pp.57-64, 2003.

[2] 西村, 西原, 鶴身, 李, 猿渡, 鹿野: ”実環境研究プラットホームとしての音声情報案内システムの運用”, 電子情報通信学会論文誌, Vol.J87-D-II, No.3, pp.789-798, 2004.

[3] A. Mehrabian 著，西田，津田，岡村，山口 訳: ”非言語コミュニケーション”，聖文社，1986.

[4] 重永: ”感情の判別分析からみた感情音声の特性”，電子情報通信学会論文誌，Vol.J83-A, No.6，pp.726-735，2000.

[5] 森山, 斎藤, 小沢: ”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”, 電子情報通信学会論文誌, Vol.J82-D-II, No.4, pp.703-711, 1999.

[6] D. Ververidis, C. Kotropoulos, I. Pitas: ”AUTOMATIC EMOTIONAL SPEECH CLASSIFICATION”, in Proc.2004 IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP 2004), vol. 1, pp. 593-596,Montreal, Canada，2004.

[7] 濱，鈴木，濱，”感情心理学への招待ー感情・情緒へのアプローチー”，サイエンス社, 2001.

[8] 大前, 西村，河原，入野: ”実環境音声情報案内システムにおける発話感情理解についての検討”, 日本音響学会 2004 年秋季研究発表会講演論文集, vol.1, pp.205-206, 2004.

[9] 田中，脇本，”多変量統計解析”，現代数学社, 1998.

[10] Alain de Cheveigne, Hideki Kawahra: ”YIN, a fundamental frequency estimator for speech and music”, Journal of the Acoustical Society of America, Vol.111, No.4, pp.1917-1930, 2002.