

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

森山 剛[†] 齋藤 英雄[†] 小沢 慎治[†]

Evaluation of the Relation between Emotional Concepts
and Emotional Parameters in Speech

Tsuyoshi MORIYAMA[†], Hideo SAITO[†], and Shinji OZAWA[†]

あらまし 本論文では、感情を含むことによって音声に生ずる物理的変動と、そこから知覚される感情とを線形に対応づけるモデルを提案する。本モデルでは、日常の感情語と物理パラメータを直接対応づける代わりに、それぞれから抽出した物理的基底及び心理的基底を対応付けに用いているため、感情語や物理パラメータの選び方に依存しないという特長を有する。また、話し手の抱いた感情ではなく聞き手側に存在する感情のステレオタイプを基準とすることで、感情を可観測で一般性を有するものとしている。本研究では統計的な手法を用い、まず種々の感情が含まれた音声の韻律パラメータと心理実験によって得た主観評価値を用いて、それぞれ物理的基底（主成分）及び心理的基底（因子）を求めた。更にこれらの基底空間に写像した物理量及び心理量に重回帰分析を施すことにより、音声の物理量と感情を双方向に変換することの可能な対応情報を獲得した。

キーワード 音声処理, 感情, 因子分析, 主成分分析, 重回帰分析

1. ま え が き

近年、マルチメディア技術のめざましい進歩と情報インフラストラクチャの整備によって、多種多様な情報通信サービスが一般家庭に広く浸透し始めている。このような中で導入されるモノは、面倒な仕事の代行だけでなく、精神的代謝の代行をも担っているといえる [1]。人々が、これらのモノとのインタラクションを通して豊かさを享受するためには、感情などのいわゆるヒューマンファクタをモノが解し、またモノ自身がこれを模擬することが必要であると考えられる。

従来からこのような技術を目指して、音声における感情情報に関する研究が行われてきた。その多くは、音声に含まれる感情と音声パラメータとの間の関係のモデル化や、合成音声への感情付与を目的としていたが、その成果はまだ十分とはいえない。感情という心理量は文化的背景や状況的文脈、音声の個性や音韻性といった要素を切り離して考えることはできないために、工学的に扱うためにある種の制約を設ける必要があり、そこに立場の違いが反映される。従来の研

究では、この制約に関する考察があまりなされていなかったために、相互の研究成果の関連性が不明確になってしまったのではないかと考えられる。

制約として明らかにしなければならないのは、一つは工学的に扱い得る感情の定義であり、また二つ目として、どのような感情をどのような物理パラメータで計測するかという観測量の選択の方法であり、更に三つ目は音声に含まれる感情と音声の物理量の間に対応関係のモデル化の方法であると考えられる。

本論文では、このような制約に関する一つの解を示すとともに、その制約に基づいて、感情を含むことによって音声に生ずる物理的変動と、そこから知覚される感情とを対応づけるモデルを提案する。本研究においては、音声に含まれる感情と生じる物理的变化とを関係づける情報を対応情報と呼ぶこととする。次章では提案するモデルについて述べ、3. では本モデルにおける対応情報を統計的な手法を用いて獲得する方法について述べる。4. では実際に対応情報を獲得し、モデルを実装する対応付け処理の結果について述べる。5. では対応付け処理の結果について考察し、更に対応情報の有用性を調べ、本モデルの有効性について検討を行っている。

[†] 慶應義塾大学理工学部電気工学科, 横浜市

Department of Electrical Engineering, Keio University, 3-14-1 Hiyoshi, Kouhoku-ku, Yokohama-shi, 223-8522 Japan

2. 理 論

2.1 対応付けモデル

音声に含まれる感情には、コミュニケーションにおける話し手が実際に抱いている主観的なものと、感情表現を客観的に観察する側（聞き手）に共有されるステレオタイプの二つが考えられる。ここで話し手の感情状態は、環境などの外的要因に加えて生理的条件などの内的要因、更に直前の感情状態などが関与していると考えられるが、それらの間の因果関係（情報の流れる方向）を特定することができないために観測不能である。しかし従来の研究の多くは、この主観的な感情を実験の対象としているものが多かった [2]~[7]。これに対して聞き手に共有される感情のステレオタイプは、主観評価実験などによる統計的な方法により可観測であると考えられる。したがって、聞き手の観点から音声に含まれる感情を評価するという立場が、感情という心理量を計測する上で有効であると考えられ [8]、本研究においてはこちらを感情の定義とする。

感情語や物理パラメータといった観測対象は、実験の制約から有限数を選択する必要がある、選択されたものはそれぞれ構築するシステムの目的に応じて感情空間及び物理空間を代表していると考えられる。従来の研究では、感情語に関しては心理学などで提案されている基本感情 [9], [10] の中からいくつかを選んで調査しており [7], [11], 感情語を感情のステレオタイプを観測するという観点から選択したものはなかった。そこで本研究では、過去の文献や予備実験の結果から感情語として挙げられたものすべてを用いて主観評価実験を行い、互いにほぼ独立な感情語の組をステレオタイプの類型として求める。また物理パラメータに関しては、従来の研究では感情を含んだ音声の韻律的特徴が調べられてきた [2], [6], [11]~[14]。音声の韻律的特徴は、発話意図などのパラ言語情報や感情情報などの非言語情報を伝達する担い手といわれており、本研究でも感情音声 (Emotional speech) の物理的様相を計測するために韻律成分をパラメータ分析することとする。

また従来の研究では、分析する韻律パラメータ間や感情語間の相関関係などについてあまり考慮されていなかった。筆者らが行った統計的実験の結果、韻律パラメータ間に高い相関が認められた [15]。更に感情語についても、これを評価語として種々の音声についての主観評価実験を行った結果、感情語間に高い相関が

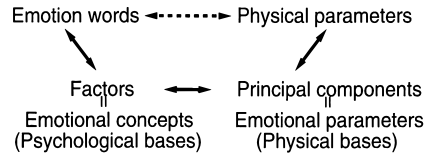


図 1 対応付けモデル
Fig. 1 Proposed relation model.

認められた。したがって、従来観測していたパラメータや感情語は冗長な成分を含んでいると同時に、それぞれ物理空間及び感情空間全体の中で局所的である可能性があった。そこで、これらのパラメータや感情語が、直交空間である物理空間及び感情空間内の特定の位置に与えられた符号であると仮定して、これらの符号から空間の基底ベクトルを抽出し、観測されるパラメータや感情語をこの基底に要約することが有効であると考えられる。また、十分なパラメータや感情語の組を用意することによって、得られる基底がそれぞれ物理空間及び感情空間を表現し、物理パラメータや感情語の選び方に依存しない対応情報の獲得を実現することが可能となると考えられる。提案する対応付けモデルを図 1 に示す。本研究では、図の実線の矢印の経路で対応情報を獲得する。

2.2 対応情報

従来の研究では、感情音声に対するいくつかの感情語 (Emotion words) についての主観評価値 e と物理パラメータ p (Physical parameters) との間の対応情報が問題とされていたが、本研究ではまず e に対して因子分析 (Factor analysis) を行うことにより、感情空間の直交基底 (Psychological bases) である因子 (Factors) を抽出し、 e を基底空間に写像した因子得点 f 及び f に乗ずることによって e を求めることのできる因子負荷行列 A_f を推定する。また、 p について主成分分析 (Principal component analysis) を行うことにより物理空間の直交基底 (Physical bases) である主成分 (Principal components) を抽出し、 p を基底空間に写像する固有ベクトル行列 A_c 及び p を基底空間に写像した主成分得点 c を推定する。そして、ベクトル c を説明変数として、因子得点 f を目的変数とした重回帰分析 (Multiple regression analysis) を行うことにより、偏回帰係数行列 A_m を推定し、これを対応情報とする (図 1)。

対応情報を用いることにより、算出した韻律パラメータからその音声に含まれる感情情報を求めること

ができ、また感情情報から音声の韻律パラメータに生ずる変化量を求めることができる。以下で対応情報を求める際に用いる統計的手法に関して述べる。

2.3 用いる統計的手法

2.3.1 因子分析

本研究では、感情語の表している感情の概念がいくつかの基底となる概念の線形和で成り立っていると仮定し、この基底概念を抽出する手法として分解の分析である因子分析 [19] を用いる。因子分析のモデルを式 (1) に示す。

$$e_i = \sum_j \alpha_{ij} f_j + o_i \tag{1}$$

$e = \{e_i | i = 1, \dots, s\}$ は観測される変数であり、共通因子 $f = \{f_j | j = 1, \dots, m; m \leq s\}$ と独自因子 o の線形結合で表される。 α_{ij} は因子負荷行列 $A_f (s \times m)$ の要素を表し、 s と m はそれぞれ変数の数と因子の数を表している。因子分析の目的は、この因子負荷行列 A_f と共通因子得点 f を推定することである。

本研究における観測変数は、音声に含まれる感情情報に対する主観評価値である。また、抽出された共通因子 f_j を、音声の聞き手に存在する感情のステレオタイプの基底概念として感情表現語と呼ぶこととする。本研究では、共通因子の推定に主因子法、因子得点の推定に回帰推定法を用いる。

2.3.2 主成分分析

本研究では、感情を含むことで音声に生ずる韻律的变化の基底は、いくつかの韻律パラメータの線形和で成り立つと仮定し、このような主成分を抽出する方法として主成分分析を用いる。主成分分析のモデルを式 (2) に示す。

$$c_k = \sum_l \beta_{kl} p_l \tag{2}$$

$p = \{p_l | l = 1, \dots, t\}$ は観測される変数であり、主成分得点 $c = \{c_k | k = 1, \dots, n; n \leq t\}$ は p_l の線形和で表される。 β_{kl} は固有ベクトル行列 $A_c (n \times t)$ の要素を表し、 t と n はそれぞれパラメータの数と主成分の数を表している。主成分分析では、固有ベクトル行列 A_c を求めることを目的としている。

本研究では、音声の韻律パラメータの平静時からの変化量を観測変数とする。また主成分得点 c_k を、音声において感情情報を担う物理的基底成分として感情表現パラメータと呼ぶこととする。本研究では、主成分の推定に主因子法を用いる。

2.3.3 重回帰分析

本研究では、音声によってある感情表現語が伝達されるのに、いくつかの感情表現パラメータが寄与していると考え、前者を目的変数、後者を説明変数とした重回帰分析を行い、得られた偏回帰係数を対応情報とする。重回帰分析のモデルを式 (3) に示す。

$$f_j = \sum_k \gamma_{jk} c_k \tag{3}$$

γ_{jk} は、偏回帰係数行列 $A_m (m \times n)$ の要素を表している。重回帰分析は結合係数行列である A_m を推定することを目的としている。

3. 対応情報の獲得

対応情報獲得の流れを図 2 に示す。左半分には感情空間の基底を抽出する処理を、右半分には物理空間の基底を抽出する処理を、そして下に感情空間の基底と物理空間の基底を対応づける処理を示す。対応情報を獲得するために、まず平静音声と感情音声(互いにほぼ独立な感情を含むもの)を用意する必要がある。

3.1 感情音声収録

感情音声の標準的なデータベースはいまだに存在しないため、独自に感情音声の収録を行わなければならない。本研究では、男性声優 4 名に「おまえ」「そんな」「みろよ」「はやく」の四つの言葉を用いて、表 1 に示す感情語(主観評価実験 (Subjective evaluation) において用いる評価語)を示しながら、できるだけ

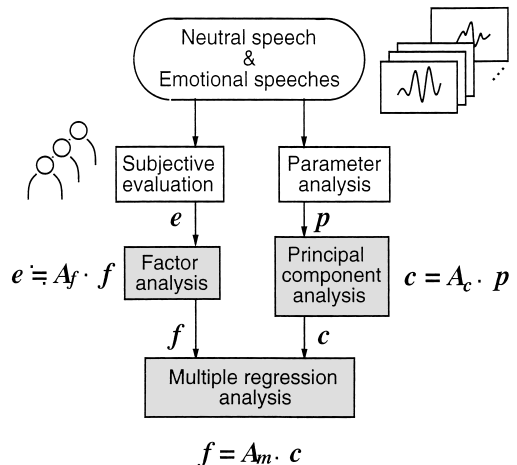


図 2 対応情報獲得の流れ
Fig. 2 Block diagram of obtaining the relation informations.

表1 感情語(アンケートの評価項目)

Table 1 Emotion words.(terms in the questionnaire)

1. 怒り	17. 寛容	33. 満足
2. 喜び	18. ほくそえむ	34. 退屈
3. 嫌悪	19. 失望	35. 苦しい
4. 悔り	20. 叱責	36. 期待
5. おかしい	21. 悲しい	37. 幸福
6. 心配	22. 恐れ	38. 好き
7. 優しい	23. 憎い	39. 嫌い
8. 安堵	24. 軽蔑	40. いや
9. 憤慨	25. 嬉しい	41. 落胆
10. 羞恥	26. 皮肉	42. 非難
11. 穏やか	27. 無関心	43. 不安
12. 憧れ	28. 賞賛	44. 驚き
13. 苛立ち	29. 誇り	45. 慌て
14. 不平	30. 愛	46. あきれ
15. 切望	31. 嘆き	
16. 気の毒な	32. こび	

	非常に 含まれて いない	かなり 含まれて いない	やや 含まれて いない	どちらとも いえない	やや 含まれて いる	かなり 含まれて いる	非常に 含まれて いる
喜び	1	2	3	4	5	6	7
怒り	1	2	3	4	5	6	7
⋮				⋮			

図3 アンケート書式

Fig. 3 Sheet format for evaluation of emotion.

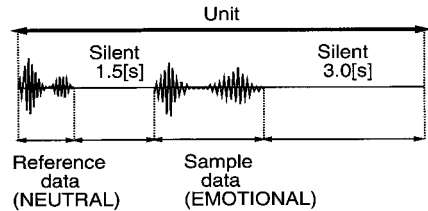


図4 音声データ構成

Fig. 4 Structure of speech data.

いろな感情に分散するようにたくさん発話してもらった。感情に関する研究では、単文を用いたものなどがあるが、本研究では文脈の影響を排除するために、1文節の言葉を用いた。またこれらの言葉は、なるべく日常生活において様々な感情を含んで発話されるようなものを選んだ。

それぞれの言葉について平静音声と感情音声約10種類ずつ、計445音声を収録した。各音声は、標本化周波数44.1kHz、16bit線形量子化でA-D変換し、ファイルに保存した。

3.2 感情表現語の抽出

3.2.1 主観評価実験

抽出される因子の信頼性を確保するためには、評価項目の数に比べてそれよりも多くの音声刺激について評価値を得る必要があるが、主観評価実験において、445音声について46項の日常感情語それぞれについて評価を行うとすれば、被験者は約 2×10^4 回評価しなくてはならず、疲労による評価のひずみを免れない。したがって現実的には、有意な評価項目に数を減らす必要がある。

そこで、ここでは主観評価実験を2段階に分けて行うこととする。まず第1段階では、表1に挙げたすべての語を用いて予備実験を行い、有意な評価語を選択する。次に第2段階では選択された評価語を用いて心理実験を行い、得られた心理データから因子の抽出を行う。また第1段階で評価語の選択を行った結果から、因子の統計的有意性を保つように第2段階で用いる音声データも選択する。

被験者は研究室の男子学生16名、女子学生2名の

計18名である。「各日常感情語について、感情音声が平静音声に比べてどのような感情を含んでいるかを感じるかを評価してください」と指示し、図3のアンケート用紙に示す7段階評定尺度を用いて評価してもらう。

感情音声をはじめに一度だけ聞かせるだけでは、すべての評価語について最後まで均等な評価を行うのが困難であると考えられる。またこれに対処するために、感情音声のみを繰り返し聞かせると、慣れの効果が生じる可能性がある。そこで聞かせる音声は、毎回平静音声を聞かせることによって、被験者は常に平静音声に感情音声の相対的な評価が行えると考え、図4に示すように平静音声と感情音声を無音区間を挟んで対にして、被験者がすべての日常感情語について評価を終えるまで何度も繰り返し聞かせることとする。

また、アンケート用紙上での評価語の並び順及び聞かせる音声対の順番は、慣れの効果を回避するために、前者については全員について、後者については被験者を3グループに分けて各グループ間で一様乱数を用いて順番をランダムにする。またサマリデータ作成時には、被験者間中央値を用いる。

3.2.2 感情表現語の抽出

主観評価実験で得られた心理データに対して因子分析を施し、因子(感情表現語)を抽出する。因子分析で抽出される因子の数については、一般に累積寄与率が60~70%までのところまでとれば十分とされてお

り、本研究でも累積寄与率 70%を超えるところまでをとることとする。

3.3 感情表現パラメータの抽出

3.3.1 韻律パラメータ

音声の韻律的特徴とは、声の高・低、強・弱、リズム・テンポを指す [11] が、本研究では韻律パラメータとして F_0 、短時間平均パワー及び発話速度を算出する。ここで、 F_0 及び短時間平均パワーは音声データ全体に関して軌跡を求め、振幅値の平均値とダイナミックレンジ、更に 1 次差分の平均値とダイナミックレンジを算出した。 F_0 の推定には、FFT ケプストラム分析を用い、軌跡はフレーム長 1024 point、フレーム周期 512 point で計算した。 F_0 は、有声/無声の自動判別 [14] を行った結果、有声部に対してのみ推定を行った。また F_0 、パワーの振幅値は、Weber-Fechner の法則 [18] により対数スケールに変換した。

3.3.2 感情表現パラメータの抽出

感情音声について算出した韻律パラメータを平静音声のそれで正規化したものを主成分分析し、主成分 (感情表現パラメータ) を抽出する。主成分についても、一般に累積寄与率が 70% を超える程度までをとれば十分とされており、本研究においても累積寄与率 70% を超えるところまでとることとする。

3.4 基底間の対応付け

抽出された因子 (感情表現語) が、その伝達の物理的な担い手である主成分 (感情表現パラメータ) の線形和によって表されるとの仮説に基づき、各因子の因子得点を目的変数、主成分得点ベクトルを説明変数とする重回帰分析を行う。

4. 処理結果

4.1 選択された感情語及び音声データ

表 1 に示す感情語すべてを用いて行った主観評価実験の結果、表 2 に示す 9 語が選択された。これらの感情語は、表 1 に挙げた感情語それぞれに対する評価値ベクトルの相関を計算し、正の相関の大きな感情語同士は似たものとして一方を削除することで統合し、また負の相関の大きなものは心理空間上で対極にあるものとして残す、という考え方で選ばれたものである。

また 9 語が選択されたことで、因子の信頼性を保証するためには 10 以上の音声を用意すればよいことがわかったので、主観評価実験の第 2 段階では被験者の労力を軽減するために、収録した 445 音声のうち、含まれている感情の評価が容易と思われる、かつなるべく

表 2 選択された感情語 (因子抽出のための評価語)

Table 2 Selected emotion words.

1.	怒り
2.	喜び
3.	皮肉
4.	恐れ
5.	悲しい
6.	驚き
7.	こび
8.	穏やか
9.	おかしい

表 3 対応付け処理に用いた音声データ

Table 3 Speech for obtaining relation information.

#	Speaker	Word
1.	A	みろよ
2.	A	そんな
3.	A	そんな
4.	B	はやく
5.	B	みろよ
6.	B	みろよ
7.	C	みろよ
8.	C	おまえ
9.	D	みろよ
10.	D	おまえ
11.	D	おまえ
12.	D	そんな

表 4 抽出された因子 (感情表現語)

Table 4 Extracted factors. (Emotional concepts)

因子#	寄与率 [%]	累積寄与率 [%]
1	44.757	44.757
2	30.444	75.202
3	12.655	87.857

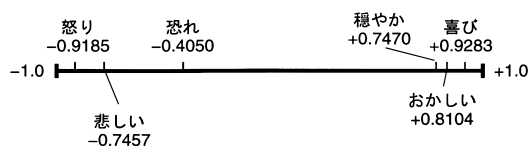


図 5 第 1 因子の感情語布置図

Fig.5 Positioning of evaluation terms for the first emotional concept.

含まれている感情がお互いに似ていないものを選択して用いることとした。表 3 に選択された音声を示す。

4.2 感情表現語の抽出結果

表 2 の 9 評価語を用いて行った心理実験データに対して因子分析を行った結果、表 4 に示すような因子が抽出された。各因子に関して、因子負荷量の比較の大きな評価語を因子負荷量の軸に布置したものを図 5 及び図 6 に示す。ここで因子負荷量は、バリマックス法によって因子軸を回転して得た値である。

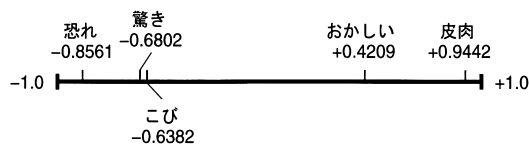


図 6 第 2 因子の感情語布置図

Fig. 6 Positioning of evaluation terms for the second emotional concept.



図 7 第 3 因子の感情語布置図

Fig. 7 Positioning of evaluation terms for the third emotional concept.

表 5 抽出された主成分 (感情表現パラメータ)
Table 5 Extracted principal components.
(Emotional parameters)

主成分#	寄与率 [%]	累積寄与率 [%]
1	44.63	44.63
2	25.58	70.21

図 5 を見ると、+1.0 側には「喜び」や「おかしい」、
「穏やか」などがあり、-1.0 側には「怒り」や「悲しい」
があることから、第 1 因子軸は心理学で従来報告
されている「快—不快」軸に相当すると考えられる。
また第 2 因子については、図 6 に示すように +1.0 側
に「皮肉」や、やや絶対値は小さいが「おかしい」が
あり、-1.0 側に「恐れ」や「驚き」などがあること
から「弛緩—緊張」軸に相当すると考えられる。また、
従来の報告ではこれらの因子の他に「注目—拒否」軸
も報告されている [9] ことから、第 3 因子について見
ると、図 7 に示すように「恐れ」や「驚き」といった
評価語と「穏やか」や「おかしい」などといった評価
語が両極に布置されていることから「注目—拒否」軸
と解釈でき、本実験の結果が心理学で報告されている
基底概念 [10] とほぼ一致していることがわかった。

4.3 感情表現パラメータの抽出結果

3.3.1 の韻律パラメータを表 3 のすべての音声につ
いて算出し、それぞれ同じ言葉で発話された平静音声
のパラメータで正規化したものを主成分分析した結果
を表 5 に示す。また、比較的大きな主成分負荷量をも
つ韻律パラメータを主成分負荷量の軸に布置したもの
を図 8 及び図 9 に示す。

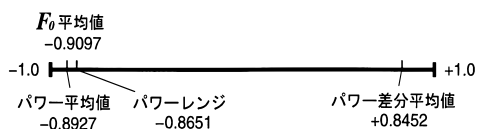


図 8 第 1 主成分の韻律パラメータ布置図

Fig. 8 Positioning of prosodic parameters for the primary emotional parameter.

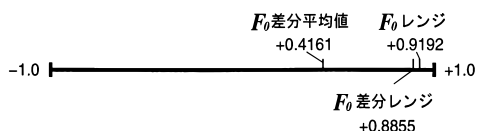


図 9 第 2 主成分の韻律パラメータ布置図

Fig. 9 Positioning of prosodic parameters for the secondary emotional parameter.

表 6 偏回帰係数行列 (() 内は標準偏回帰係数)
Table 6 Partial regression coefficients.

	第 1 主成分	第 2 主成分
第一因子	-0.1717(-0.3441)	+0.1123(+0.1704)
第二因子	+0.3039(+0.6091)	-0.0849(-0.1288)

図 8 を見ると、第 1 主成分に対して F_0 の平均値
やパワーの平均値が負に大きく寄与しており、またパ
ワー軌跡の微分値の平均値が正に寄与していることか
ら、全体的な「声高さ」(正負逆)を表す主成分である
と解釈できる。図 9 では、第 2 主成分に対して大きな
負の負荷量を有するパラメータはなく、 F_0 軌跡のダ
イナミックレンジや微分値が大きく正に寄与している
ことから「抑揚」を表すパラメータであると解釈でき
る。これらの主成分の累積寄与率が 70% を超えること
から、音声感情を含むことによる物理的变化は「声
高さ」及び「抑揚」の成分が支配的であるということ
を示している。

4.4 基底間の対応付け結果

表 3 のすべての音声について因子得点及び主成分得
点を推定し、これに対して重回帰分析を行って推定し
た偏回帰係数行列 A_m を表 6 に示す () 内は、説明変
数の重要度を表す標準偏回帰係数である。表 6 の偏回
帰係数を見ると、第 1 因子すなわち「快」感情は、平
静音声が高声で抑揚のある声に変化することでより
強くなり、また第 2 因子すなわち「弛緩」感情は、
平静音声が高声でなくなり、かつ抑揚がなくなると知
覚されるということを表していると考えられる。更に
標準偏回帰係数の値から、どちらの因子 (感情表現語)
も声高さにより重みがあることがわかる。

表 7 重回帰分析の評価
Table 7 Evaluation of the multiple regression analysis.

因子#	重相関係数	決定係数
1	0.38	0.15
2	0.62	0.39

また、重回帰分析の精度を評価するために、重相関係数及び決定係数（音声の数による自由度に関して修正済み）を算出した結果を表 7 に示す。重相関係数は、各音声について韻律パラメータから推定した因子得点と、主観評価値から推定した因子負荷量の相関関係を表しているが、表 7 に示すとおり、第 1 因子に関しては精度は十分ではないが、第 2 因子に関してはやや良好な分析が行えていることが示されている。

5. 検討及び考察

5.1 対応付け処理結果に関する考察

対応付けの精度が十分でなかった理由として、用いた音声データが複数の話者や言葉によるものであり、得られる心理データの分散に、話者性情報や音韻情報による分散が含まれてしまっていたことが考えられる。また、今回は線形モデルを仮定しており、感情の伝達に関する物理パラメータ間の非線形性を無視していることも一因と考えられる。したがって今後精度を向上させるためには、感情音声として単一話者の無意味発話を用いたり、また、可観測な感情語や物理パラメータをそれぞれ基底に変換するという本モデルの枠組みは変えずに、ニューラルネットワークなどの非線形手法を用いたりすることが有効であると考えられる。

5.2 因子の解釈の妥当性

因子に対して与えた解釈の妥当性を示すために、音声に含まれる感情を因子の解釈に基づいて再評価し、因子得点との相関を調べることにする。ここでは被験者にとって評価しやすい解釈の与えられている第 1 因子について実験を行った。実験では、第 1 因子に与えられた解釈である「快—不快」という対語を 7 段階評定尺度の両端に配置したアンケート用紙を作成し、表 3 の音声について、3.2.1 の被験者とは別の被験者男性 5 名に評価させた。図 10 に 5 名の評価値の平均値（Subjective evaluation）を横軸、4.2 で各音声に対して推定された第 1 因子得点（Estimated factor score）を正規化して縦軸にプロットしたものを図 10 に示す。図 10 の横軸と縦軸の相関係数は 0.6065 であり、「快—不快」という対語は因子に与えた解釈として

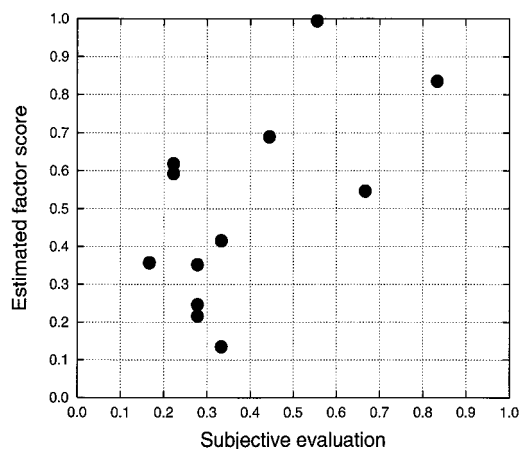


図 10 「快—不快」軸への主観評価値と 4.2 の第 1 因子得点の関係

Fig. 10 Relation between the evaluation values on 'pleasant-unpleasant' axis and the estimated 1st factor scores in 4.2.

ほぼ良好に第 1 因子を表しているといえる。

また、本研究で実装した対応付けモデルを用いて求められる因子得点との相関についても調べるために、図 10 の横軸に対して、音声の物理パラメータから対応情報を用いて求めた第 1 因子得点（Transformed factor score）を縦軸にプロットしたものを図 11 に示す。図 11 において、主観評価値と本モデルで物理パラメータから変換して求めた因子得点の間には、弱いながらも正の相関（相関係数 0.4434）が認められる。

以上の結果から、第 1 因子に対して与えた解釈は妥当であることが示された。

6. むすび

本論文では、聞き手が音声から知覚する感情と音声感情が感情を含むことによって生ずる物理的变化の間を、双方を基底空間に写像してから対応づけるモデルを提案した。本モデルは、音声によって伝達される感情情報と、音声に生ずる物理的变化の情報とを関係づけるという、音声情報処理において工学的に感情を扱う上で避けることのできない問題に対して解を与えるものである。

本モデルを実装した結果次のことがわかった。すなわち、音声の聞き手の感情のステレオタイプは 9 感情語に集約され、更にそれらに対して因子分析を施した結果得られた基底概念は、「快—不快」「緊張—弛緩」

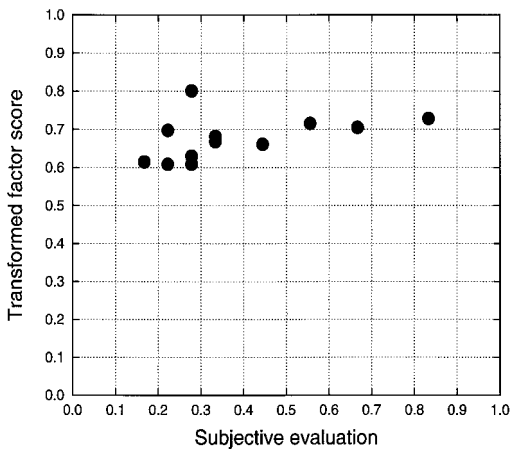


図 11 「快—不快」軸への主観評価値と第 1 因子得点の関係

Fig. 11 Relation between the evaluation values on 'pleasant-unpleasant' axis and the 1st factor scores.

そして「注目—拒否」といった, Schlosberg らによって従来提案されていた基本感情 [10] とほぼ一致する。Schlosberg らによって発見された因子は顔表情に対して得られたものであり, モダリティの違いを超えて同様の因子が抽出されたことは興味深い。また, 音声感情を含んだときに生ずる物理的な変化として, 「声高さ」「抑揚」といった成分が支配的であることもわかった。

今後, 5.1 において述べたように, 対応情報の精度を向上させ, また, 音声を合成し得る物理パラメータを用いることによって, 音声に含まれる感情の認識及び平静音声への感情付与を行うことのできるシステムを構築する。

謝辞 感情音声データの収録に快く協力してくださった阿曾沼正明氏, 堀江宏仁氏, 住吉慶彦氏にこの場をお借りして深く感謝の意を表します。

文 献

- [1] 大橋 力, 小田 晋, 日高敏隆, 村上陽一郎, “情緒ロボットの世界” 講談社, 東京, 1985.
- [2] 上床弘幸, 小林 豊, 新美康永, “音声の感情表現の分析とモデル化” 信学技報, SP92-131, Jan. 1993.
- [3] 新村貞彦, 村上憲也, “平静音声と感情音声の音韻パラメータの比較” 音響講論 3-P-19, pp.319-320, Oct. 1993.
- [4] 小林 豊, 新美康永, “音声の感情を反映する韻律情報制御方式について” 音響講論 2-8-7, pp.233-234, Oct. 1993.
- [5] 羽尻公一郎, 山下崇晴, 小川 均, 天白成一, “日本語母音単音節における感情表現と韻律情報との関係” 音響講

論 2-4-1, p.265, March 1995.

- [6] 重永 実, 小川 孝, 中尾光志, “単語音声による感情表現について” 信学技報, SP95-15, May 1995.
- [7] 鈴木 司, 有泉 均, “感情表現の分析と合成規則” 信学技報, SP96-131, March 1997.
- [8] L.A. Streeter, N.H. Macdonald, R.M. Krauss, W. Apple, and K.M. Galotti, “Acoustic and perceptual indicators of emotional stress,” J. Acoust. Soc. Am., vol.73, no.4, pp.1354-1360, April 1983.
- [9] I.R. Murray and J.L. Arnott, “Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion,” J. Acoust. Soc. Am., vol.93, no.2, pp.1097-1108, Feb. 1993.
- [10] 福井康之, “感情の心理学” 川島書店, 東京, 1990.
- [11] 藤崎博也, “音声の韻律的特徴における言語的・パラ言語的・非言語的情報の表出” 信学技報, HC94-09, Sept. 1994.
- [12] 北原義典, 東倉洋一, “音声の韻律情報と感情表現” 信学技報, SP88-158, March 1989.
- [13] 宮武正典, 勾坂芳典, “種々の発話様式における韻律パラメータの性質について” 信学技報, SP87-62, 1987.
- [14] 平賀 裕, 斉藤善行, 森島繁生, 原島 博, “音声に含まれる感情情報抽出の一検討” 信学技報, HC93-66, Jan. 1994.
- [15] T. Moriyama, H. Saito, and S. Ozawa, “Evaluation of the relationship between emotional concepts and emotional parameters on speech,” Proc. of ICASSP, vol.2, pp.1431-1434, April 1997.
- [16] 今井憲一, 都木 徹, 勾坂栄一, “声質変換における感情付与を目的としたニューラルネットによるピッチパターン制御” 音響講論 1-8-23, pp.189-190, March 1993.
- [17] 藤崎博也, 大野澄雄, 富田 修, 丸山英晃, “話者の感情が音声の韻律的特徴に及ぼす影響” 音響講論 2-4-3, pp.269-270, March 1995.
- [18] 古井貞照, “音響・音声工学” 近代科学社, 東京, 1992.
- [19] 菅 民郎, “多変量解析の実践上・下” 現代数学社, 京都, 1993.

(平成 9 年 12 月 8 日受付, 10 年 8 月 14 日再受付)

森山 剛 (学生員)



平 6 慶大・理工・電気卒。現在, 同大学院博士課程在学中。音声情報処理, 感情情報処理に関する研究に従事。日本音響学会, 言語処理学会各会員。



齋藤 英雄 (正員)

昭 62 慶大・理工・電気卒・平 4 同大大学院博士課程了。同年、慶大・理工・電気・助手、平 7 専任講師。パターン認識、画像処理の研究に従事。工博。計測自動制御学会、情報処理学会、AVIRG、IEEE 等各会員。



小沢 慎治 (正員)

昭 42 慶大・工・電気卒。昭 47 同大大学院博士課程了。昭 45 慶大・工・電気助手。同大専任講師・助教授を経て現在、同大電気工学科教授。その間、昭 59 米メリーランド大訪問助教授。デジタル通信及び画像・音声のデジタル信号の処理の研究に従事。著書「デジタル信号処理」、「基礎通信工学」(実教出版)。工博。電気学会、情報処理学会、IEEE 各会員。