情報処理学会論文誌 Vol. 52 No. 3 1185–1194 (Mar. 2011)

情動・感情判別のための自然発話音声データベースの構築

酒造正樹[†1] 山本泰史[†1] 志村誠[†2]
門間史晃[†2] 光吉俊二[†2] 山田一郎[†1]

コミュニケーションの研究領域において，音声情報を用いた「情動」または「感情」の認識が有用とされている．関連する従来研究では，「情動」「感情」の定義が混同して用いられることがあったが，本論文において「情動」とは不随意的反応で「快」と「不快」の 2 値を持つものとし，「感情」とは情動から引き起こされる随意的反応で「怒り」「喜び」「悲しみ」「不安」「苦痛」などの言葉で表されるものと定義した．本論文の目的は，これら「情動」と「感情」の組合せからなる発話音声を大量に収集し，今後の判別分析の基礎データとして用いられるようその関係性を整理することである．音声データベース構築の要件としては，実験参加者から自然な感情を引き出し，その収録音声に適切な心理状態のラベルを付与すること，分析に十分なサンプル数があることとした．結果，約 100 人の実験参加者からの自然発話音声を約 16,000 ファイル集め，本人および 10 人の第三者の主観評価による快不快情動と個別感情のラベルの付いたデータベースを構築することができた．

Construction of Natural Voice Database for Analysis of Emotion and Feeling

Masaki Shuzo,[†1] Taishi Yamamoto,[†1]
Makoto Shimura,[†2] Fumiaki Monma,[†2]
Shunji Mitsuyoshi[†2] and Ichiro Yamada[†1]

In the human communication research area, it is often said to be useful to recognize “emotion” or “feeling” from voice. As “emotion” and “feeling” were sometimes confused in previous studies, we clearly defined, in this paper, “emotion” as an involuntary response in the human brain which has two states of “pleasant” and “unpleasant”; “feeling” (e.g., anger, enjoyment, sorrow, fear, and distress) as a state voluntarily resulting from an emotion. Our objective is to collect sufficient amount of voice data which contains “emotions” and “feelings”. Requirements for voice database are to extract the natural “feelings” from participants, to add appropriate psychological labels for the recorded voice data and to have enough samples for the recognition analyses. As a result, about 16,000 natural voice data ware collected from about 100 participants, and the database with “emotion” and “feeling” labels based on self assessments by themselves and 10 third-party members were constructed.

1. はじめに

コミュニケーションの円滑化には，言語情報のほかに表情や身振り，周囲の環境といった複合的な非言語・非明示情報つまりは雰囲気情報の共有が重要となる．筆者らは，遠隔コミュニケーションで失われがちな雰囲気情報を伝達する端末「障子」を作成し，その効果を検証してきた[1)]．「障子」が伝達する人間に関する情報（感情，存在，移動）と環境に関する情報（体感温度，騒音，照度）の中でも，ユーザスタディの結果から感情情報の共有が最も効果的であるという結果を得ている．「障子」における感情情報の取得に関しては，身体への拘束性や情報取得の容易性を考慮し，音声から感情を認識する手法をとっている．

音声による感情判別に関する研究は従来から多々あるが，実用化においてはいまだ精度において十分でなく，課題を残している．たとえば，どのように感情のラベルを音声に付与するのかという問題や，発話者の意思によって音声への感情の発現が抑制されるなどの問題は広く一般に知られている．また，「怒り」と「喜び」，「平常」と「悲しみ」が誤判別されやすいという問題（たとえば文献 2)–4) など）がある．そこで，本研究では脳における感情処理の前段にあるとされる情動処理に着目した．

情動の定義には複数あるが，本研究では脳科学などの文献（たとえば文献 5)，6) など）に基づき，快不快情動のモデルを定義した（**図 1**）．情動は感情よりも根源的であり，快情動と不快情動に大別される．情動の発生は声帯の緊張など不随意な生理反応をともなうため，声帯に由来する音声特徴量から快不快情動が観測できると考える．

よって，感情よりも根源的な快不快情動の概念を取り入れ，音声による感情判別の精度向上について議論したい．そのため，音声からの判別分析を行うに十分な質と量を持つ発話

†1 東京大学
The University of Tokyo
†2 株式会社 AGI
AGI. Inc.

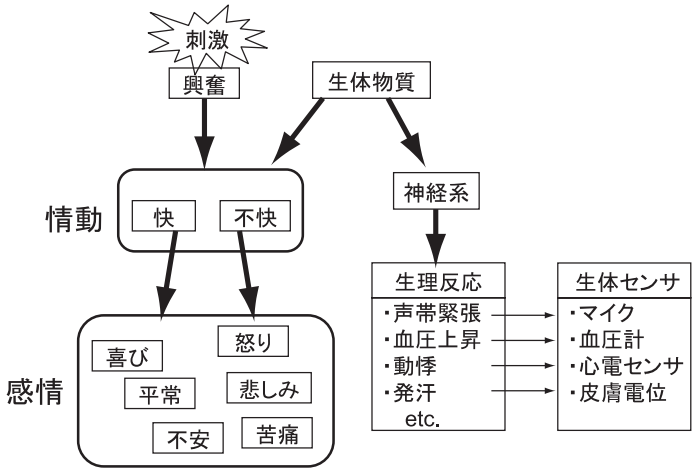

図 1　情動と感情
Fig. 1　Emotion and feeling in this paper.

音声データベースを構築することが重要であり，本論文ではこれを目的とした．要件としては，実験参加者から自然な感情を引き出すこと，収録音声に適切な心理状態のラベルを付与すること，分析に十分なサンプル数があることなどがあげられる．具体的な量については，判別対象となる個々の感情ラベルに対して，学習データとして少なくとも 100 以上のサンプルが経験的に必要である．人間の評価のばらつきも加味するならば，さらに 10 倍程度の音声ファイルを収集することが望ましい．よって本論文においては情動と感情について分類を行うので，約 10,000 以上の音声ファイルを収録することを目標値とした．

以降，2 章では感情判別の関連研究について述べ，3 章では発話音声の収録実験とその主観評価手法について述べる．4 章では取得した発話音声データの記述分析を行い，5 章では音声実験に対する考察を述べる．6 章では情動・感情判別分析の例を用い，収録音声の量的評価を行う．最後に 7 章にまとめを行う．

2. 関 連 研 究

本章においては，従来の感情判別の関連研究について言及し，情動と感情の関係について議論する．

近年，人間の感情に関する研究について注目が高まっている．たとえば Norman[7] が指摘するように，人々が道具を使うといった日常的なことがらでさえも，そのときにどのような感情が喚起されるかで，得られる心理的効用が大きく異なってくるのである．

脳波，心電，表情など，人間の感情を反映していると考えられる指標は数多く存在するが，最も取得しやすいと考えられる指標として音声が考えられる．音声はマイクがあれば取得することが可能であり，人体に対しての侵襲性が少ない．そのため，音声によって感情を認識することができれば，幅広い分野に役立てることが可能となる．具体的な技術の応用先としては，たとえばコールセンタが考えられる．顧客の問合せを受ける際に，オペレータが顧客の感情を把握することによって，よりきめ細かな対応を行うことができる．また，感情判別技術を遠隔コミュニケーションに用い，離れた相手の感情情報を伝達しあうことで，親しい人同士のコミュニケーションを支援するといった試みがなされている[1]．

音声波形を分析して感情を判別する研究は，すでに多く研究者が行っている（たとえば文献 2)–4)，8)–12) など）．柴崎ら[8] は，50 人の被験者から音声ファイルを収録し，その音声に含まれる感情を 20 人の別の被験者に評価させることで，感情ラベルのついた音声データを取得した．そして音声データから得られた基本周波数（ピッチ）や音量（パワー）に関する多くの音声特徴量を解析することで感情判別を行った．佐藤ら[9] は，男性 12 人，女性 5 人の学生被験者から「怒り」「悲しみ」「喜び」「平常」の 4 感情の音声を取得し，ニューラルネットワークにより分析を行った．門谷ら[10] は，男性 7 人，女性 6 人の被験者から「怒り」「悲しみ」「喜び」「平常」の 4 感情の音声を取得し，正準判別分析を行った．白澤ら[2] は，演劇経験者やアナウンサからなる女性 9 人から音声を取得し，「平静」「怒り」「悲しみ」「喜び」「驚き」「嫌悪」の 6 感情をマハラノビス距離により判別している．個人差を含むことが容易に予想される感情研究において，これらの従来研究は被験者の数が十分とはいえない．また，Mitsuyoshi[12] は，2 人の被験者から感情を込めた音声を発話している際の脳活動状況を，マイクと fMRI を用いて分析している．

しかし，ここで問題となるのが，「感情」をどう定義するかということである．感情の定義については，感情を表す単語を言語学的に分類，分析することで，感情概念を明らかにする研究が行われてきた．代表的な感情概念の分類として，次の 2 つがあげられる．

(1)　感情の基本次元として「快–不快」「覚醒–睡眠」という軸を用いたもの[13]
(2)　「喜び」「驚き」「恐れ」「悲しみ」「怒り」「嫌悪」など個別感情を用いたもの[14]

音声から感情を認識する試みも，「快–不快」軸を用いたもの[15),16)]，個別の感情を用いたもの[17),18)] について，これまでそれぞれが独立に行われてきた．

これに対して光吉[11] は，別々に検討されてきた感情の定義を 1 つにまとめあげ，人間の

基本情動として「快–不快」が存在し，情動が生じた結果として，「悲しみ」「怒り」など個別の感情が生じるというモデルを提示した．これによれば，人の感情処理では，外的刺激を受けると脳内化学物質やホルモンなどの生理的な反応がまず起きる．この生理反応によって引き起こされる根源的な「快」もしくは「不快」の感覚が情動と定義される．さらに，生理反応から生じた情動を本人が認知することによって，より複雑な「怒り」「喜び」「悲しみ」などの個別感情が生じるものと考えられる．この情動–感情モデルによって，これまで別々に扱われてきた「快–不快」と，「悲しみ」「怒り」などの個別感情を統合した形で，音声から判別できる可能性が示された．

本論文において，音声における快不快情動と個別感情の関係を整理する．快不快情動と個別感情を組み合わせて感情を分類することで，音声からの感情判別精度を向上させることが可能になると考えられる．

より具体的には，快情動に基づいたネガティブな感情の発生，不快情動に基づいたポジティブな情動の発生の可能性があげられる．一般的には「喜び」はポジティブな感情であり，快情動と結び付いていると考えられるが，嫌な相手に対する愛想笑いのように，不快情動を含んだ喜びというものがありうる．また「怒り」はネガティブな感情ではあるけれども，激しく感情を発露して怒ることですっきりする場合もある．このように，情動における「快」「不快」と個別感情における「ポジティブ」「ネガティブ」は必ずしも一致しない．よって，情動の快–不快と個別感情のポジティブ–ネガティブの関係に特に注目する．つまり「怒り」「喜び」といった感情を，「快情動を含んだ怒り」「不快情動を含んだ怒り」「快情動を含んだ喜び」「不快情動を含んだ喜び」に分割するのである．情動の快–不快と感情のポジティブ–ネガティブの一致度をみてみることが必要になる．

そのため本研究では，従来研究にある被験者に感情を込めて決められた台詞を発話（演技発話）してもらう方法をとるのではなく，感情や情動を喚起する刺激映像を提示し，それに対して被験者に自由に会話をしてもらい，その自然発話音声を収録する形をとった．収録された音声発話に快不快情動と個別感情のラベリングをするために，本人と第三者が快不快情動と個別感情の判定を行う．それと同時に，演技発話も行い，自然状況での発話と，演技状況での発話を比較することで，本研究の音声収録手法の妥当性を検討できるものと考えた．

3. 音声収録実験と主観評価

発話音声データベースの構築は，概略として，音声を収録する実験と，収録した音声へ心理状態の主観評価からなる．まず，実験参加者に感情を喚起しそうな動画を見せ，その際の発話音声を収録した．男女計 96 人に対して収録を行い，約 16,000 発話を収録した．次に，発話者が自らの発話音声を聞き，発話時の情動・感情に関する主観評価を行った（自己評価）．さらに，複数の第三者が発話音声を聞き，発話者の情動・感情に関する主観評価を行った（他者評価）．

3.1 音声収録実験

3.1.1 自然発話音声

感情を含んだ大量の自然な状況の音声データを取得するために，刺激映像を見ながら実験者と実験参加者が自由に会話を行う，音声収録実験を行った．以下，これで得られたデータを自然発話音声とする．実験参加者は 10 代から 40 代の男性 53 人と，女性 43 人の計 96 人で，2009 年 9 月から 10 月にかけて東京大学工学部の研究室において実施した．また，本実験は東京大学医学部倫理委員会の定める同意書にサインを得て進めた．

実験は，1 人のインストラクタと 1 人のサクラ（実験誘導者）および，1 または 2 人の実験参加者で行われた．実験誘導者は自身も実験参加者の 1 人であるかのように振る舞い，実験参加者と一緒に実験を受けた．実験誘導者が実験参加者に対して積極的に話しかけ，友好関係を築くことで，感情を含んだ音声を多く取得することを目指した．実験の手順は以下のとおりである．

(1) インストラクタが，プロジェクタを設置した部屋の中に実験参加者と実験誘導者を案内し，音声収録用のインカムマイクをつけてもらう．
(2) インストラクタが，これから動画を見て自由に感想を話しあってもらう実験を行う旨を，実験参加者と実験誘導者に説明する．
(3) 実験参加者と実験誘導者，インストラクタの間に会話をしやすい雰囲気を作るために，実験参加者と実験誘導者にそれぞれ自己紹介をさせる．
(4) 感情を喚起させる数分程度の動画をプロジェクタに提示し，実験参加者と実験誘導者に感想をいいあってもらう．

刺激映像によって喚起させる感情として，従来研究[11] に従い「怒り」「喜び」「悲しみ」「不安」「苦痛」「平常」の 6 種類の感情を取り上げた．実験に使用した動画の概要を下に示す．

- 怒り：スポーツ選手への失礼なインタビュ，戦争による理不尽な暴力
- 喜び：お笑い芸人の言動
- 悲しみ：戦争や貧困を扱った悲しいアニメのダイジェスト
- 不安：精神病院をテーマにしたゲームの CM

図 2　評価用アプリケーション
Fig. 2　Appearance of the voice evaluation application.

- 苦痛：嫌悪感をいだく昆虫が大量発生している映像

なお，音声収録には，ヘッドセット型マイク（C555L，AKG）と IC レコーダ（H4n, ZOOM）を用いて行い，その音声ファイルは PC に保存される．記録条件は，44.1 kHz, 16 bit で，非圧縮 wave 形式で保存した．また所要時間は約 1 時間であった．

3.1.2　演技発話音声

さらに，比較対象となる演技発話音声も取得した．「怒り」「喜び」「悲しみ」「不安」「苦痛」「平常」の 6 感情それぞれについて，40 個の短文または単語を提示して，感情を込めて発話してもらった．実験参加者は自然発話音声の場合と同じで，男性 53 人，女性 43 人であった．

3.2　主 観 評 価

3.2.1　自己評価（本人による主観評価）

音声収録終了の直後に，実験参加者が発した自然発話音声を，発話単位に区切って順番に提示し，そのときに自分がどのように感じていたかを記憶に従って評価させた．評価用アプリケーションの画面を**図 2** に示す．実験参加者が評価するのは「快不快情動」「個別感情」「興奮度合い」の 3 種類であった．各指標の項目を下に示す．

- 快不快情動：「快」「不快」「どちらでもない」
- 個別感情：「怒り」「喜び」「悲しみ」「不安」「苦痛」「平常」「どれにもあてはまらない」
- 興奮度合い：「興奮していない」「やや興奮している」「興奮している」

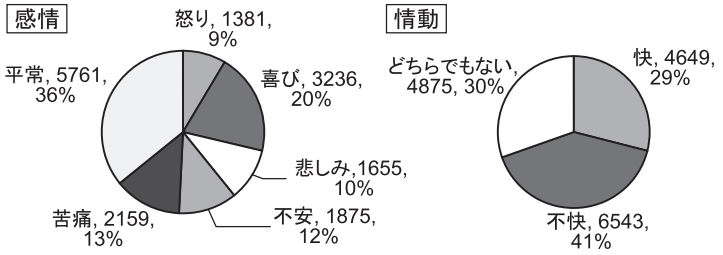

図 3　収録音声のファイル数とその割合（自己評価）
Fig. 3　The number and the rate of recorded voice data (subjective assessment by him/herself).

音声発話は，無音区間によって分割し，無音区間と無音区間の間の有音区間を 1 発話ファイルとして扱った．上記手順に従って実験を行った結果，その結果 16,067 ファイルを得た．「快」「不快」「どちらでもない」は，それぞれ 4,649，6,543，4,875 ファイルであった．また「怒り」「喜び」「悲しみ」「不安」「苦痛」「平常」は，それぞれ 1,381，3,236，1,655，1,875，2,159，5,761 ファイルを得られた（**図 3**）．感情を引き出しにくいと予想された「怒り」「悲しみ」「不安」「苦痛」に関して，「喜び」より少ないもののそれぞれ約 10%ずつ得られた．

なお，演技発話音声については，指示した感情をラベルとして用いるので，自己評価を実施しない．

3.2.2　他者評価（第三者による主観評価）

96 人のすべての自然発話および演技発話の音声データを取得し終えた後，音声データに対して第三者が心理状態のラベル付けを行う音声評価実験を行った．1 つの音声ファイルに対して，男性 5 人，女性 5 人の計 10 人が自己評価と同じく「快不快情動」「個別感情」「興奮度合い」の 3 点を評価した．評価のばらつきを抑えるため，すべての音声データからランダムに抜き出されたデータを評価者に順々に提示し，それを評価していく形をとった．評価に用いたアプリケーションインタフェースは，自己評価に用いたのと同じものであった．評価者は，大学生を中心に約 50 人で，2009 年 12 月から 2010 年 3 月にかけて行った．

4.　発話音声データの記述分析

4.1　自然発話音声

4.1.1　他者評価結果

個別感情それぞれについて，快不快情動の「快」「どちらでもない」「不快」にあてはま

表 1　自然発話音声における感情ごとの快不快情動の分布
Table 1　Distribution of emotion by each feeling in natural voice.

	快	どちらでもない	不快	計
怒り	404	509	5,603	6,516
喜び	27,623	1,583	1,019	30,225
悲しみ	322	997	7,009	8,328
不安	1,016	2,912	12,653	16,581
苦痛	528	752	9,703	10,983
平常	18,813	42,729	9,173	70,715
計	48,706	49,482	45,160	143,348

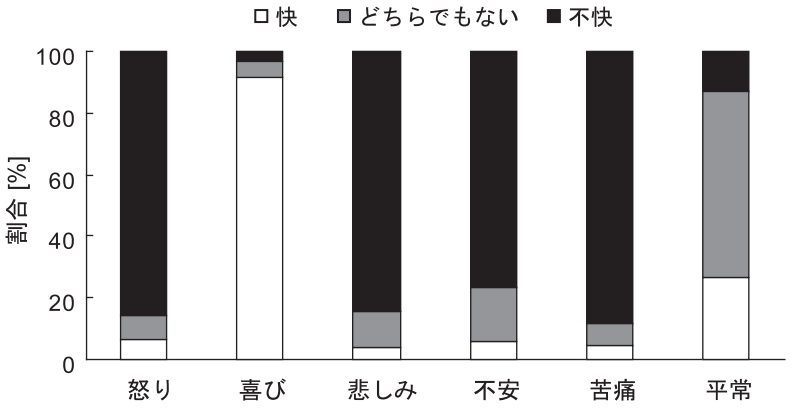

図 4　自然発話音声における感情ごとの快不快情動の分布（他者評価）
Fig. 4　Distribution of emotion by each feeling in natural voice (subjective assessment by third persons).

る自然発話音声ファイルの個数を示したものが **表 1** および **図 4** である．1 つの音声データに対して，10 人が評価をしているため，実際の音声データ数の約 10 倍のデータ数がある（一部のデータについては，評価者の評価漏れがあるため，実際には 10 倍よりは少ない数になっている）．「平常」以外の 5 感情については，「喜び」は快情動，「怒り」「悲しみ」「不安」「苦痛」は不快情動と強く結び付いていると考えられる．図表に示されたとおり，得られた音声データも，その大半は快不快情動と個別感情が一致していた．各感情において，平均して 80〜90%程度の音声データが，快不快情動と一致した方向性であった．逆に，快不快情動と個別感情の方向が不一致のもの（「喜び」なのに「不快」，「怒り」「悲しみ」「不安」「苦痛」なのに「快」）も 3〜5%程度の割合で存在した．「悲しみ」と「不安」で，「どちらでもない」の割合がやや高かったものの，全体として感情ごとの快不快情動の比率に大きな違いはなかった．

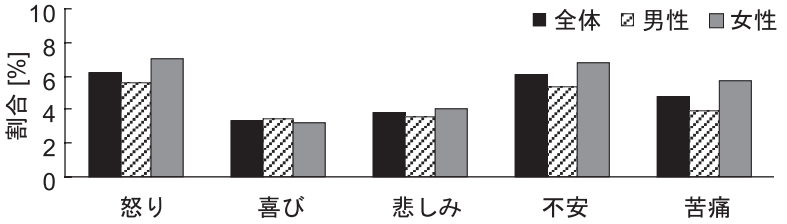

図 5　快不快情動と個別感情が不一致の割合（他者評価）
Fig. 5　Discordance rate between emotion and feeling (subjective assessment by third persons).

次に，音声データを男性と女性で分割して検討した結果を示す．これは，男性と女性とでは声の質が大きく異なっており，その違いが評価者の評価に影響を及ぼす恐れがあるためである．そこで男女別に，快不快情動と個別感情が不一致の割合を示したものが **図 5** である．男性と女性とでは，全体的に女性の方がやや不一致の比率が高かったが，それほど大きな差は見られなかった．

4.1.2　自己評価結果

自然発話音声に関しては，第三者からの評価だけではなく，発話者自身の評価も取得している．そこで，他者評価結果の妥当性を検討する意味で，発話者の自己評価についても同様に結果を示す．まず個別感情それぞれの，快不快情動の分布を示したものが **図 6** である．全体的な傾向としては，他者評価結果と大きな違いはなく，大半のデータにおいて快不快情動と個別感情の一致が見られた．また，男性と女性の音声データに分割して，不一致の割合を比較したものを **図 7** に示す．自己評価の場合も他者評価と同様，男女で大きな違いは見られなかった．

4.1.3　他者評価結果と自己評価結果の比較

他者評価と自己評価の結果について，比較した結果を **図 8** に示す．快不快情動と個別感情が不一致の割合は，おおむね 2〜6%程度に収まっていた．他者評価の不一致率の方が，自己評価に比べるとやや高かった．また「怒り」「喜び」「苦痛」では「悲しみ」「不安」に比べ差があるように見えるが，両者の間に大きな差は見られなかった．その意味で，第三者による評価の結果は，発話者と個人の主観ともおおむね一致しており，双方の評価結果に妥当

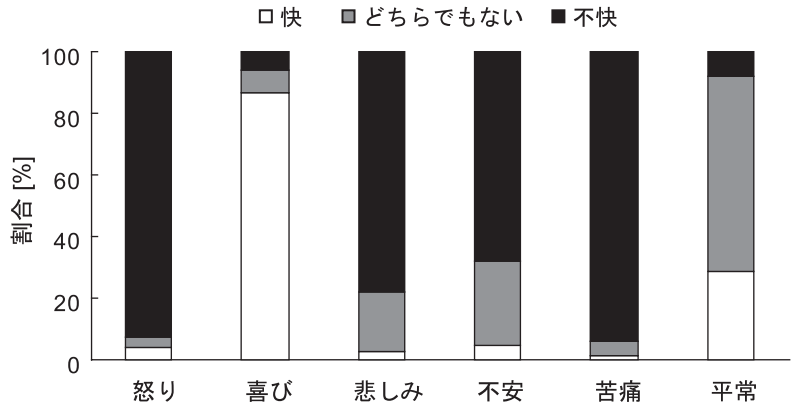

図 6　自然発話音声における感情ごとの快不快情動の分布（自己評価）

Fig. 6　Distribution of emotion by each feeling in natural voice (subjective assessment by him/herself).

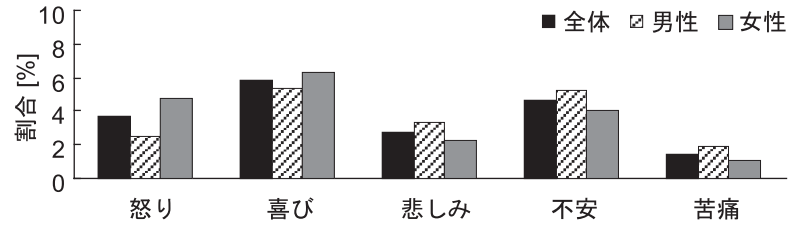

図 7　快不快情動と個別感情が不一致の割合（自己評価）

Fig. 7　Discordance rate between emotion and feeling (subjective assessment by him/herself).

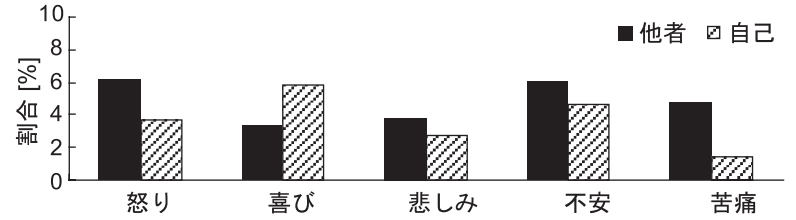

図 8　自己評価と他者評価の不一致比率の比較

Fig. 8　Comparison of discordance rate between subjective assessments by him/herself and third persons.

表 2　演技発話音声における感情ごとの快不快情動の分布

Table 2　Distribution of emotion by each feeling in acting voice.

	快	どちらでもない	不快	計
怒り	1,326	1,926	43,076	46,328
喜び	29,167	1,631	781	31,579
悲しみ	487	2,573	24,310	27,370
不安	1,028	4,962	29,334	35,324
苦痛	805	2,060	32,488	35,353
平常	13,153	40,269	10,121	63,543
計	45,966	53,421	140,110	239,497

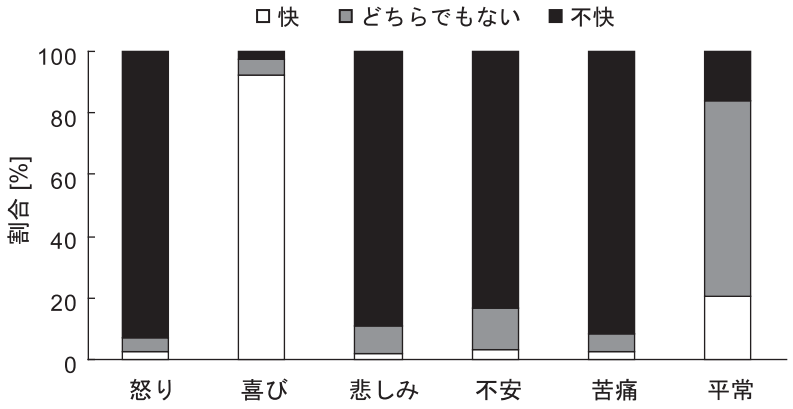

図 9　演技発話音声における感情ごとの快不快情動の分布

Fig. 9　Distribution of emotion by each feeling in acting voice.

性があるものと考えられる．

4.2　自然発話音声と演技発話音声の比較

続いて，自然状況で発話された音声と，演技で感情を込めて発話された音声との違いを検討した．演技発話について，10 人の評価者がラベル付けを行った結果として得られた，個別感情ごとの快不快情動の分布を示したものが**表 2** および**図 9** である．その結果として，おおむね自然発話と同様に，大半の音声において快不快情動と個別感情が一致するという結果が得られた．

これを自然発話と比較した結果を**図 10** に示す．「怒り」「喜び」「悲しみ」「不安」「苦痛」の 5 つすべての感情において，自然発話の方が不一致の比率が高かった．演技発話の場合は，

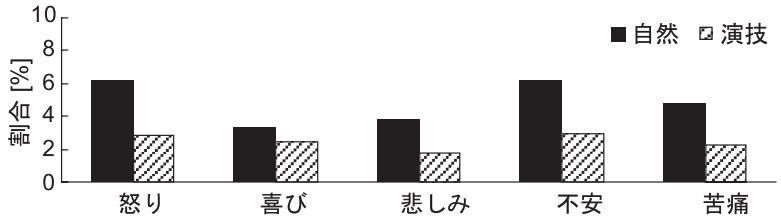

図 10　自然発話音声と演技発話音声の不一致比率の比較
Fig. 10　Discordance rate of feeling between natural voice and acting voice.

強制的に感情を込めて発話しているため，快不快情動とのズレが少なくなったものと考えられる．逆に自然発話の場合は，会話の文脈によって，面白くないけれど相手に合わせて笑っておく，といった状況が生じやすいため，不一致の比率が高くなったものと推測される．

5. 音声実験に関する考察

本研究では，音声における快不快情動と個別感情の関係を整理することを目的として，音声収録実験を実施し，取得したデータに対して記述的な分析を行った．それにより，まず本研究では，これまで別々に検討されていた，快不快情動と個別感情を統合する試みを行った．その結果，基礎的な知見として，情動の快–不快と感情のポジティブ–ネガティブは，方向性が一致しない場合があることが明らかになった．一致する割合は 85%，不一致の割合が 5%，どちらでもない場合が 10%であった．平均して 15%の音声は情動と感情の方向性が一致していない．そして，この情動と感情の方向性の不一致は，他者評価と自己評価，自然発話と演技発話のいずれにおいても，一貫して見られた．つまり，個別感情と快不快情動は必ずしも一致するわけではないということである．そして，他者評価と自己評価，また自然発話と演技発話のすべてにおいて，一貫して安定的な結果が得られた．また，特定の感情だけが不一致率が高いといったこともなく，全体としてロバストな傾向が見られているといえる．

その一方で，今回の実験の結果として，問題点が浮かび上がった．たとえば，発声される感情が偏っていたことである．表 1 に示したように，自然発話音声の大半は「平常」であり，その次に「喜び」と続き,「怒り」と「悲しみ」は非常にデータ数が少なかった．そのため，このデータをそのまま感情判別した場合には，データ数に偏りがありすぎるため，良い分析精度が得られない．この原因は，初対面の人同士が映像を見ながら会話するという音声収録状況によるものと考えられる．初対面の相手に対して「怒り」や「悲しみ」の感情を見せる状況は，一般的に想定しにくく，逆に相手と友好な関係を築きたいと実験参加者が考えることで，「喜び」の発話が多くなってしまうものと推測される．

この問題を解決するためには，「喜び」以外の感情を喚起させる新たな実験状況を作り出す必要がある．具体的には，親しい友達同士を実験参加者として会話させる，また議題を与えてディベートを行ってもらうなどの方法が考えられる．

また情動の快–不快と感情のポジティブ–ネガティブの関係についても，大きな偏りが見られた．感情判別の精度を高めるという観点から見れば，情動と感情が不一致になりやすい状況というものを人工的に作り出す必要がある．情動と感情が不一致になる状況がどういうものかは，まだ明らかになってはいない．今後はまず本研究で得られたデータについて，情動と感情が不一致であったデータの特徴を抽出することで，どういった状況下で情動と感情が不一致になりやすいのかを明らかにする必要がある．それによって，情動と感情が不一致のデータを効率的に取得し，分析することで感情判別の精度を高めていくことが可能になると考えられる．

6. 情動・感情判別分析によるデータベースの評価

収録した音声ファイル数の妥当性については，実際に判別分析を行い，その結果から評価することができる．たとえば，ある判別手法を用いて良い結果であれば，データベースを構成するファイルの量が十分であるといえる．一方で，ファイル数がいくら多くても判別結果がともなわなければ，量的に十分でないデータベースとなる．本章においては，快不快情動の判別および個別感情の判別の結果の例を紹介しつつ，量的評価を行いたい．本章で用いる判別手法は新たに開発したものでなく，ほかに最適な方法が存在する可能性もあるが，データベースの量的評価を行ううえでは 1 つの指標として十分に議論が可能であろう．

6.1 快不快情動の判別

まず取得したラベル付き音声の一部を抽出し，その特徴量を算出して，快および不快の学習データを作成した．次に，別途用意した試験データを用い，快–不快の各群との距離判別により 2 群の判別を行った．

発話に含まれる言語情報の影響が少ない特徴量として，従来研究（たとえば文献 2)–4), 8)–12) など）の知見をもとにピッチとパワーに関する音声特徴量 57 個（平均，分散，最大値，最小値，レンジ，発話値など）を採用した．ここで，パワーは実験参加者によって口とマイクの距離が異なることを考慮し，パワーの平均値によって正規化を行った．

学習データを作成する際に，自己評価と他者評価でラベルが異なることが生じる．自己評

価と他者評価のラベルが一致した音声ファイルを学習データとして使用した方が，判別に適したものになると予想される．これを検証するため，「自己評価のみ」「自己評価と他者評価が一致したもの（一致評価）」の 2 種類の学習データを作成し，快不快情動の判別を行った．なお，他者評価によるラベルは，1 つの音声に対して，10 人の評価者のうち 80%以上が一致したものを割り当てた．

ここで，それぞれの学習データを構成する音声ファイルの数は 436 に揃えた．これは，自明ではあるが，「一致評価」のファイル数が「自己評価のみ」に比べ減少することが理由であり，両者を均一な数にすることで適切な比較となる．

試験データにも「一致評価」のデータを用い，線形判別関数による判別分析を試した．判別の結果，「一致評価」を学習データとした方が判別率は約 5%高く 77%であり．この学習データが判別に適していることを確認した．

続いて，音声は個人間・個人内での変動が大きいため，変動が小さく判別に有効な特徴量を選定する必要がある．その指標として，相関比を用いた特徴量選定法を導入した．この選定法を「一致評価」のデータベースに適用し，線形判別を用いて判別率を算出した．特徴量選定により，特徴量の数は 57 から 17 となり，80%の判別率を得た．

以上の結果から，音声特徴量を用いた機械学習により，快不快情動が判別可能といえる．その際，学習データの作成には本人と第三者の一致評価を用いることが有効であることを示した．また，これまでに述べてきた快情動がポジティブ感情の「喜び」と強い関係を持つという仮定においては，感情判別の前段に快不快情動判別を行うことで，「怒り」と「喜び」の誤判別の改善に貢献すると予想される．

6.2　個別感情の判別

音声特徴量から個別感情を判別する分析を行った．使用したデータは自然発話のデータであり，判別対象とした感情は，他者評価によりラベル付けされた感情であった．他者評価によるラベルは，1 つの音声に対して，10 人の評価者のうち 50%以上が一致した感情を割り当てた．そのため一致率 50%未満の音声データに関しては，分析対象から除外した．その結果，「喜び」のデータ数が 2,430 ファイルに対して，「怒り」「悲しみ」「不安」「苦痛」がそれぞれ 224，232，585，580 ファイルとなった．「喜び」のデータが抜きん出て多いことにより，感情判別の結果が「喜び」に過剰にフィットしてしまう事態を防ぐために，最も少ない「怒り」のファイル数に合わせる形で，他の感情データをランダム抽出により削減し分析を行った．

判別手法としては，C5.0 による決定木学習アルゴリズムを用いた．分析にあたっては，ブースティングを 10 回，またサンプルを 10 分割した形のクロスバリデーションを行った．音声特徴量としては，ピッチとパワーから算出した計 273 個を用いた．

自然発話音声を用いて個別の 6 感情を判別した結果，平均で 60%であった．参考として，演技発話音声に関しては 67%であった．この結果は，判別する感情の数が 6 であることを考慮すれば，悪くはない結果といえよう．

感情の研究においては，適切な心理ラベルを付与することが難しく，本論文では複数の第三者による主観評価（他者評価）が一致した音声ファイルのみを扱うことで，質の良い学習データとなる．今回の報告では，実験参加者 100 人の規模で 16,000 の音声を収集したといっても，個別感情の学習データ作成時には本節においては約 200 ファイルに減少している．この数は判別分析を行うに必要最低限の数であるが，データベース構築の目標を達したといえる．継続して発話音声データの収集を行えば，判別率が改善可能であろう．

7. おわりに

本研究では，多様な自然感情を含む音声発話を，現在までに約 100 人の実験参加者から約 16,000 ファイルを取得した．それらに自己評価と他者評価からなる情動・感情ラベルを付与した大規模な音声データベースを構築し，情動と感情の関係について整理した．

構築した音声データベースを用いて，一例として判別分析を行った．音声認識エンジンを用いた意味内容の理解に基づく判別分析を行うとさらに判別率の向上が見込めるが，本論文ではピッチやパワーに関する特徴量のみに基づく判別分析を行った．快不快情動の判別では 80%程度，個別感情の判別では 60%程度の結果を示し，このデータベースが判別分析に有用であることを示した．とりわけ情動判別の結果からは，音声から不随意の情報である情動が判別可能であることを示している．

また，各音声ファイルに付与したラベルについて，第三者の主観評価とできるだけ一致した音声ファイルを用いることで，情動判別，感情判別ともに質の良い学習データとなりうることが分かった．とりわけ感情判別においては，感情の数を 6 個としたこともあり，一致評価採用後の最小ファイル数の感情が「怒り」の約 200 ファイル，最大ファイル数の「喜び」が約 2,000 ファイルと大きな差が生じた．当初の目標であった 100 を上回っていたため，今回は最小ファイル数に合わせたデータ構成としたが，継続して音声ファイルの収集を行いデータベースの規模を拡大し，調整していくことで判別率の改善が予想できる．

また，このような研究は，他の論文の手法と比較して判別率などを公平に議論していく必要があると思われる．今後，我々が入手可能な手法があれば，本データベースにあてはめて

分析してみたい．あるいは，取得したデータベースを公開し，他の研究者が考案した判別手法を自らで評価できるようにしておけば，意義が深まるであろう．

参 考 文 献

1) Shuzo, M., Shimura, M., Delaunay, J.-J. and Yamada, I.: Shoji: A Communication Terminal for Sending and Receiving Ambient Information, *Proc. ASME 2009 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference* (*IDETC/CIE 2009*), DETC2009-86314 (2009).
2) 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，Technical Report of IEICE, HIP, Vol.96, No.499, pp.79–84 (1997).
3) 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，Technical Report of IEICE, SP, Vol.97, No.396, pp.73–80 (1997).
4) 直井克也，松本哲也，竹内義則，工藤博章，大西　昇：感情に関係する特徴量の検討，Technical Report of IEICE, HIP, Vol.105, No.99, pp.37–42 (2005).
5) Ledoux, J.: *The Emotional Brain: The Mysterious Underpinnings of Emotional Life*, Simon & Schuster, New York (1996). 松本　元，小幡邦彦，湯浅茂樹，川村光毅，石塚典生（訳）：エモーショナル・ブレイン―情動の脳科学，東京大学出版 (2003).
6) Damasio, R.A.: *Looking for Spinoza: Joy, Sorrow and the Feeling Brain*, Harcourt, New York (2003). 田中三彦（訳）：感じる脳情動と感情の脳科学よみがえるスピノザ，ダイヤモンド社 (2005).
7) Norman, D.A.: *Emotional Design: Why We Love* (*Or Hate*) *Everyday Things*, Basic Books (2004).
8) 柴崎晃一，光吉俊二：抑揚からの感情認識の評価：感性制御技術（ST）の評価と，人間の感情の評価法について，Technical Report of IEICE, TL, Vol.105, No.291, pp.45–50 (2005).
9) 佐藤秀明，赤松則男：ニューラルネットワークによる感情音声の分類，Technical Report of IEICE, NC, Vol.101, No.154, pp.85–90 (2001).
10) 門谷信愛希，阿曽弘具，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，Technical Report of IEICE, SP, Vol.100, No.522, pp.43–48 (2000).
11) 光吉俊二：音声感情認識及び情動の脳生理信号分析システムに関する研究，徳島大学博士論文 (2006).
12) Mitsuyoshi, S., Tanaka, Y., Shibasaki, K., Kato, M., Minami, M. and Murata, T.: Emotion Voice Analysis System Connected to the Human Brain, *Proc. 2007 IEEE International Conference on Natural Language Processing and Knowledge Engineering* (*IEEE NLP-KE'07*), pp.479–484 (2007).
13) Russell, J.A.: A Circumplex Model of Affect, *Journal of Personality and Social Psychology*, Vol.39, pp.1161–1178 (1980).
14) Ekman, P.: *Emotions Revealed: Understanding Faces and Feelings*, Phoenix (2003).
15) Zhang, S., Xu, Y.-J., Jia, J. and Cai, L.-H.: Analysis and Modeling of Affective Audio Visual Speech Based on Pad Emotion Space, *Proc. 6th International Symposium on Chinese Spoken Language Processing* (*ISCSLP 2008*), DOI: 10.1109/CHINSL.2008.ECP.82 (2008).
16) Eadie, T.L. and Doyle, P.C.: Direct Magnitude Estimation and Interval Scaling of Pleasantness and Severity in Dysphonic and Normal Speakers, *Journal of the Acoustical Society of America*, Vol.112, No.6, pp.3014–3021 (2002).
17) Ren, F., Matsumoto, K., Mitsuyoshi, S., Kuroiwa, S. and Lin, Y.: Researches on the Emotion Measurement System, *Proc. IEEE International Conference on System, Man and Cybernetics*, pp.1666–1672 (2003).
18) Ren, F. and Mitsuyoshi, S.: To Understand and Create the Emotion and Sensitivity, *International Journal of Information*, Vol.6, No.5, pp.547–556 (2003).

(平成 22 年 5 月 21 日受付)
(平成 22 年 11 月 5 日採録)

酒造　正樹（正会員）

2003 年東京大学大学院工学系研究科博士課程修了．博士（工学）．2003 年より東京大学大学院情報理工学研究科産学官連携研究員（2003〜2004 年），日本学術振興会特別研究員（2004〜2007 年）．この間，MEMS を用いた生体情報計測の研究に従事．2007 年より東京大学大学院工学系研究科助教として，生体・環境情報処理基盤の研究開発に従事．ロボット学会，機械学会，電気学会等の会員．

山本　泰史

2010 年東京大学工学部機械工学科卒業．現在，同大学院工学系研究科機械工学専攻修士課程に在籍，人間環境情報（感情や匂い等）の研究に興味を持つ．

志村　誠

2005 年東京大学大学院人文社会系研究科修士課程修了．2007 年同大学院工学系研究科修士課程修了．日本学術振興会特別研究員（2009〜2010 年）．2010 年東京大学大学院人文社会系研究科博士課程中退．同年（株）AGI に入社．現在は感情認識を主とした研究開発に従事．

門間　史晃

2002 年帝京平成大学情報学部情報システム学科卒業．同年（株）エイ・ジー・アイ（現 AGI）入社．音声からの感情認識アルゴリズムの研究開発を中心として，音声感情受付システム，会話型ロボット，コールセンタにおける感情認識システム等の開発に従事．

光吉　俊二

1988 年多摩美術大学美術学部彫刻科卒業．1999 年まで彫刻家として活動．同年（株）エイ・ジー・アイ（現 AGI）代表取締役就任．2006 年徳島大学大学院工学系研究科博士課程修了．博士（工学）．スタンフォード大学バイオロボティクスラボラトリ客員研究員（2003〜2004 年）．慶應大学上席研究員（2009 年〜）．その間，感情認識を中心として，工学，心理学，脳科学の横断的研究を行う．電子情報通信学会，日本音響学会等の会員．

山田　一郎

1974 年東京大学大学院工学系研究科修士課程修了．同年日本電信電話公社（現在 NTT）に入社．1995 年通信エネルギー研究部長，2000 年生活環境研究所長．この間，情報機器の運動制御の研究，大容量光記憶システムの研究開発，クリーンエネルギーシステムの研究開発等に従事．2002 年から東京大学大学院工学系研究科教授として，環境問題の解決やライフスタイルの革新に資する生活環境 ICT の研究に従事．日本機械学会，精密工学会，計測自動制御学会，電子情報通信学会等の会員．工学博士．