

音響的特徴と言語的特徴をパラメータ化して 統合的に扱う感情表現の程度推定手法

飯田 仁
有本 泰子

This article aims to present a set of parameters relating both acoustic and linguistic features of natural speech, which can estimate the degree of emotion in natural speech. Focusing on 'angry' speech, we collected examples of correct angry speech data through simulated speech dialogs. After analyzing the data, 16 parameters for acoustic and linguistic features were decided upon. Experiments to verify those parameters showed that the question-type parameter and parameters for fast speech, loud voice, and high pitch tone can estimate the degree of anger.

キーワード：音声対話、感情音声、パラ言語、怒り、接続助詞中止型

1. はじめに

インターネットが普及し、ユビキタス社会が浸透するなか、情報システムとしてのコンピュータと人間との対話が増加する。これまでの情報システムとの対話では、言語とは独立した指差しを利用したり、キーボードあるいは音声認識システムを介した言語情報を主に扱い、パラ言語情報を扱うことは少ない。状況に依存する言語理解の難しさに輪をかけて、パラ言語情報を十分に扱わないため、使用者の状態を適切に判断して、満足のいく対話を実現することができない。本研究では、言語情報およびパラ言語情報により推定可能な感情を識別することを狙いとする。そのため、感情の程度による言語表現の特徴および音声の音響的变化を分析し、コンピュータと人間とのインタラクションにおける人間の感情および態度表出を捉えることを目指す。それにより、両者の円滑なコミュニケーションを図ることができる仕組みを実現していく。さまざまな感情を取り扱う点で、網羅的に、また相互関係的に捉えることが望まれるが、具体的には特定の感情に的を絞って取り組む。まず、応用対象として近

い将来考えられる対話を想定し、コールセンターなどへの自動音声応答システムにおける「苛立ち」や、真意が伝わらないことに対する「腹立ち」などの表現に着目して、ユーザの内的状態としての「怒り」の程度をその発話の言語形式および音響的な特徴により推定する手法について検討する。

本論では、形態・形式にまわり着くパラ言語を含む音声言語による「怒り」の感情表現の程度を推定することを試みる。そのために、「怒り」の音響的な特徴ならびに言語形式的な特徴による特性をパラメータとして設定し、それらパラメータが「怒り」の特徴付けに関与する可能性があることを確かめる評価実験を行う。それにより選定されたパラメータを使って、決定木を使った「怒り」の度合いを推定する実験を行う。その実験を通して、甲高くまくし立てる怒り声や慇懃無礼に不満を表すような話し方を音響的な特性のみにより推定できることを示す。さらに、疑問や接続助詞中止型の言語形式に対し、音響的な特性が一体となって、「怒り」の程度を推定することが可能であることを示す。

2. 研究のアプローチ

音声として発せられる言葉を音声言語と呼ぶことにして、音声言語による基本感情として直感的に捉えられる、「怒り」、「喜び」、「驚き」、「悲しみ」を感情の識別対象とする。このとき、音声言語としての発声の抑揚などに関わる音響情報をパラ言語とした言語形式に特徴があるか、音響情報と言語形式との間に関連があるか、さらにそれらの情報が感情の程度を示す特徴になりえるか、これらのことを明らかにすることを狙いとする。これらの狙いを実現するために、つぎの4段階の進め方をとる。

<1> [正解と見なすデータの収集] 基本感情表現を含む音声データの収集：

正解となるデータが示されていない、かつ特定の研究者による言語直感に基づいた判定では感情の表現分布をより良く捉える理論的ならびにモデル論的考察がまだ充分でないと考えられる。

<2> [正解の確からしさ検証のためのデータ加工] 主観評価法による感情情報を付与した音声データの作成：

感情情報を定量化することで、収集した各データを情報空間上に配置する。

<3> [特徴となる性質の特定] 音響上ならびに言語形式上の特徴分析：

音声処理、ならびに言語処理等で扱われている特定の特徴について、目標とする特徴付けに関与し得るか否かを個別に検定し、特徴パラメータとして選定する。

<4> [選定した特徴の検証] 選定したパラメータを使った感情の程度推定実験：

決定木による判別手法を使って、「怒り」の感情の程度を推定する検証実験を行い、各パラメータの特徴付けへの寄与の様子を探る。

3. 関連研究

関連する研究分野は多岐に渡るが、主たる4分野について概要をつぎに示す。

感情の心理学的なモデルに関して、現状では広く支持されているモデルが確立しておらず、感情にまつわる研究でも様々な概念化が行われ、その独自のモデルに基づいて研究が行われている。感情の程度推定に関わる心理学的な尺度について、基本感情8感情 (acceptance, anger, anticipation, disgust, joy, fear, sadness, surprise) の設定とその立体構造モデルを定めた Plutchik (1980) の研究、水平軸に快 (pleasure) - 不快 (displeasure)、垂直軸に覚醒 (arousal) - 眠気 (sleep) の2次元で表される平面状に日常使用する一般用語で分類した感情を円環に並べた circumplex model を提唱する Russell (1989)、pleasantness と activity の2軸の他に円錐状に neutral に向かって上昇する intensity の次元を設けた3次元モデルを提唱する Daly ら (1983) の研究がある。

音響的特徴による感情の推定および識別手法に関する研究は多感情を扱い、感情間の識別を行う研究として広く行われている。本論文では、怒りの感情の認識に絞って研究を進めるため、怒りの音声研究について触れておく。武田ら (2002) は平板、中高、頭高の3種類のアクセント型を持つ4モーラと6モーラの有意味単語および無意味単語をNHKアナウンサー4人および音声研究者1名の計5人により、平常、軽い怒り、怒り、激怒の4段階で発声されている音声試料を用意し、音声パワー、時間構造 (持続時間、発話速度) および基本周波数を韻律の特徴パラメータとして分析を行った。その結果、怒りの度合いが大きくなるにしたがって、最高音圧および最高基本周波数の増大が認められている。また、時間構造でみる特徴では、怒りの度合いが強くなるにしたがって早口になる傾向が見られる一方、逆に激怒時には発話が遅くなる傾向が見られたことも確認している。また、Ang ら (2002) の研究は DARPA Communicator Project のもとで開発された電話による航空券予約システムを利用して、演技ではない実際の音声の収録を行い識別実験に利用している。怒りのなかでも frustration と annoyance に限定し、話速、休止、基本周波数、音圧、スペクトログラムなどの音響的特徴だけでなく、対話中の発話の位置や繰り返しや言い直しなどの情報もパラメータとして加え、怒りの音声の特徴を分析している。一方、Banse ら (1996) は怒りを Hot Anger と Cold Anger に分け、異なる感情として扱い、複数の識別実験間の比較を行い、Hot Anger に対しては70%程度の識別率を得たものの、Cold Anger に対しては50%以下の識別率しか得られていない。

つぎに、言語形式と感情あるいは話者の心的態度との関係について論じた研究事例について見る。言語表現と話者の心的状態および態度に関する認知的なアプローチとして横森 (2006) の接続助詞ケドの文末用法に関する考察がある。文末のケドに符号化された意味は従属節から予期されることに対して、聞き手が認知していることが逸脱的であるという認知的

評価であるとしている。心的態度表出はこの認知的評価により決定付けられるとしている。さらに、終助詞「よ」または「ね」の語用論的機能とイントネーションとのかかわりについての研究も多い（片桐 1997, 杉藤 2001, 犬飼 2001 など）。

4. 「怒り」の発話の音声収録

怒りを表現した発話を含むできる限り自然な対話音声を収録し、各収録発話に対し5段階の主観評価に基づいて、感情の程度の実測値を付与した。以下に、実際に行った音声収録方法・主観評価法について概説する。詳細の説明は有本ら（2006b）の報告に詳しい。

4.1. 音声収録手法

コンピュータ対人間および人間対人間の電話による擬似対話を設定し、収録を行った。発話内容を限定せず、(a) ユーザ役の音声提供者がコンピュータ役の人間と対話をし、コンピュータの認識ミスによる聞き返しにより怒りを誘発させる方法と、(b) あらかじめ、クレームの内容を提示し、電話にて人間のオペレータとその内容について話し合ってもらう方法の2通りの対話を実施した。その際、音声提供者には対話に必要な最低限の情報のみを与え、オペレータの指示に従って自由に発話するようにした。実際の対話例を図1に示す。

(a) 認識ミスによる聞き返しの例

O: お問い合わせの内容はどのようなものですか

U: あの料金についてちょっと聞きたいと思ってんですけど

O: 設備でよろしいですか

U: いえちょっと 料金ですね 料金

O: 料金でよろしいですか

U: はい

(b) クレームに対応している例

U: いや でも発行されてるんで

O: いや 8万代の予約番号というのは

U: え でもされてるんで

O: 会議室では発行されないんですよ

U: ジャ どこでされるんスカね

図1 「怒り」の収録対話例

注：Oはオペレータの発話、Uはユーザの発話

収録は一般の大学生10名（男性5名、女性5名）がユーザ役の音声提供者となり、研究の趣旨を熟知しているオペレータ役1名が対応した。オペレータ役とユーザ役とは非対面に対話を行い、それぞれの発話をDATレコーダにて左右チャンネルに分けて録音した。収録した音声発話（男性3名、女性2名）から、159発話を音声試料として使用した。各々の発話の長さは様々で、最短の発話は「はい」（112ms）、最長の発話は「だから その... そのデー

タベースに全部残っているっていう証明はどうやってするんですか」(6.26s)であった。

4.2. 主観評価法

音声資料に対して、含まれている怒りの程度を定量化するため、主観評価実験を行った。被験者は大学生12名(男性11名、女性1名)で、ヘッドホンを通して1発話につき2回ずつランダムに提示される各発話について、怒りの程度を1(ぜんぜん)から5(すごく)までの5段階で評価させた。発話ごとに評価された値の平均値を求め、それを怒りの程度を表す実測値として各発話に付与した。

さらに、音響的な特徴を分析する際に、怒りの程度の大まかな指標を設けるため、全159発話の音声資料を主観評価値をもとに4つのクラスタにクラスタリングした。クラスタリングには距離尺度にWard法¹⁾を用いた階層型分類法を利用した。以後、分類した4つのクラスタを平静から順に怒りの程度が大きくなるようA、B、C、Dと称する。各クラスタのデータ数は順に41、52、55、11である。このクラスタリングは、特徴パラメータの選定に際し、パラメータとして有効であることを検証するために必要となる。言語形式に関するパラメータ化の説明などに使っている。

なお、評価値、すなわち怒りの程度は連続に分布しており、このクラスタリングは以後の分析のために便宜上行ったもので、相隣り合うクラスタの境界は必ずしも明確ではない。

5. 「怒り」を含む発話の分析

5.1. 音響的分析

本研究では音声の音響情報を反映するようなパラメータを怒りの推定のために利用する。検討したパラメータは感情表現に関する先行研究(武田他2002, Cowie et al. 2001, Banse and Scherer 1996, Ang et al. 2002, など)を参考に定めた。それらのパラメータは4種の特性に関する性質: 声の高さ、強さ、速さ・長さ、声質を示す。多くの特徴量については、発話内の代表的な値として平均値を、時間的な変動の指標として発話内での標準偏差の値を用いた。

(1) 高さに関連するパラメータ

音声の高さに関連するパラメータとして、基本周波数(F_0)から求められる発話内での統計量F-1~F-7: F-1; F_0 の発話内平均(対数軸)、F-2; F_0 の発話内最低値、F-3; F_0 の発話内最高値、F-4; 男女差正規化 F_0 の発話内平均(対数軸)、F-5; 男女差正規化 F_0 の発話内最低値、F-6; 男女差正規化 F_0 の発話内最高値、F-7; F_0 の発話内標準偏差(対数軸)を用いる。なお F_0 は10ms間隔で自動抽出した後、視察により修正を施した。また、 F_0 の観測値そのものを対象とする特徴量(F-1、F-2、F-3)については、男女間の差の影響を除去するため、簡易的

に男女それぞれについて全データの平均値を求め、その値を差し引くことで男女差の正規化を行った特徴量 (F-4、F-5、F-6) を用意した。

(2) 強さに関連するパラメータ

音声の強さに関連するパラメータとして、P-1～P-3：P-1; 短時間平均パワーの発話内標準偏差、P-2; 短時間平均パワーの発話内最大値、P-3; 短時間パワーの変動量を用いる。

(3) 長さ・長さに関連するパラメータ

発話の長さ・長さに関連するパラメータとして、D-1; 息継ぎまでの発声単位である呼気段落の平均モーラ数、ならびに D-2; 平均発話速度 (mora/s) の2種の値を用いる。D-1 は呼気段落のモーラ数を発話内で平均したものであり、D-2 は休止区間を除外して求めたモーラ数に基づく平均発話速度である。

(4) 声質に関連するパラメータ

声質に関連するパラメータとして、スペクトル包絡の大局的な傾きに着目して、2種のパラメータを設定した。C-1; 第1次ケプストラム係数²⁾の発話内平均、ならびに C-2; 第1次ケプストラム係数の発話内標準偏差である。いずれも、音源のスペクトル包絡の傾きに対応すると考えられるもので、共に声の質の違いに大きく関係している特徴である。

5.2. 実測値の分布から判断するパラメータの選定

上述の計14の特徴量のほかに10種近い音響的性質の関わりを調べたが、これら14種が怒りを特徴つける要因となり得ることを確認し、パラメータとして選定した。そのために、それぞれ単独でどの程度怒りの推定能力を有するかについて検証した。

いくつかの特徴量について、怒りの程度に関する4つのクラス内での平均値、標準誤差、標準偏差の変化の様子を調べた。クラス間での変化を見ると、特徴的な現象として、一つは怒りの程度が大きくなるに従って、ほぼ単調に値が増加あるいは減少するもの (D-2 および F-2)、もう一つは怒りの程度が最大のクラス D について A から C までの傾向と反する変化を呈するもの (P-2 および F-3) を挙げる事が出来る。

選定したパラメータ間の相関について見ておくことも重要である。複数の特徴量を組み合わせる際に相互に相関の高い特徴量の対を把握する必要がある。上述の14種の特徴量に関して相関係数を求めたところ、いずれの特徴量も同種のカテゴリを超えて、高い相関が見られたものはなかった。同一のカテゴリ内で特に正の相関が大きいもの ($r \geq 0.75$) として、高さに関連する特徴量では、 F_0 対数平均、 F_0 最高値、 F_0 最低値のそれぞれの組み合わせであった。

5.3. 言語形式に関する分析

収録した音声データのうち怒りの強いクラスタ内に疑問表現、ならびに終助詞「よ・ね」を伴う言い切りの表現が多いことが分かる。さらに、特徴的な表現として、「ので・けど」で終止する接続助詞中止型と称している表現も頻出する(有本他 2006a, 2006b, など)。ここでは、こうした言語特徴を利用し、怒りの推定に貢献するパラメータの設定を試みるため、語用論的機能の観点から見た怒りとの関係や音響特徴による機能の変化などに着目し、分析を行う。

(5) 文構成に関するパラメータ

怒りの強いクラスタである C と D における具体的な典型例は、「それはどうやって証明するんですか」、「でも全部残っているって証明はされてないじゃないですか」などの疑問文である。

疑問表現は一般的に未知の部分の情報を相手に求める質問型と自分自身に問いかける自問型に分類される。

本論で扱った対話は、e-ラーニングシステムなどにおける学習者の発声とは異なり、目的をもったシステムとの対話を想定しているため、質問型の疑問表現が現れている。この質問型では一般的に上昇調のイントネーションが使われると言われるが、怒りが伴うときの発声においても上昇調であるか即断はできない。そして、それらの言語形式は相手の主張の正当性に対する強い反意を疑問形式として表現していると捉えられる。

(6) 文末形式を捉えたパラメータ

話し言葉の一つの典型的な発話タイプである「予約したんですけど／予約したんで／残ってるんですけど／でもされてるんで」などの接続助詞中止型は、一般的に主節部分を明示せずに、聞き手にその解釈を委ねる。こうした表現は、その主節で聞き手側の対応の悪さに対する怒りの内容を聞き手に想起・自覚させることで、その事態への再考を促す役目を果たすと見なせる。本論での対話収録においては、ユーザーに敢えて怒りを誘発させるような発話をオペレータが発しているため、既知の確認済みの事柄を提示するに留まる接続助詞中止型で終わる発話が多く収録されていると考えられる。

終助詞「ね・よ」で終止する発話は、いずれも言語形式上は同意要求あるいは確認としての使用と見なせるが、話者が納得できない条件、事態について同意させられていることに対する話者の内的状態の表出と捉えることができる。その意味で、「怒り」の感情への解釈につながると思なせる。終助詞「ね・よ」に関しては多くの研究者(片桐 1997, 杉藤 2001, 犬飼 2001 など)がその意味や機能、イントネーションとのかかわりについて研究を行っているが、片桐(1997)によれば、イントネーションの位置によって不満・強調・強引な押し付けなどさまざまな含みを持たせることが可能であるとしている。収録音声に見られる「ね」で終止

する発話についても、イントネーションとの関わりにより、その言語機能に変化があるか否かを探るため、やや怒りの強いクラスターCで見られる発話に限定して、文末30フレームの F_0 の傾きと手前30フレームの F_0 の傾きの差分を捉えてみた(図2)。一般に句末の F_0 は下降調になるといわれるが、終助詞「ね」で終わる発話はその他の発話に比べ、緩やかな変化ではあるものの上昇傾向を示している(図2中の終助詞に関する変位が正の側に位置している)と考えられる。ここでの発話の機能は確認要求である。

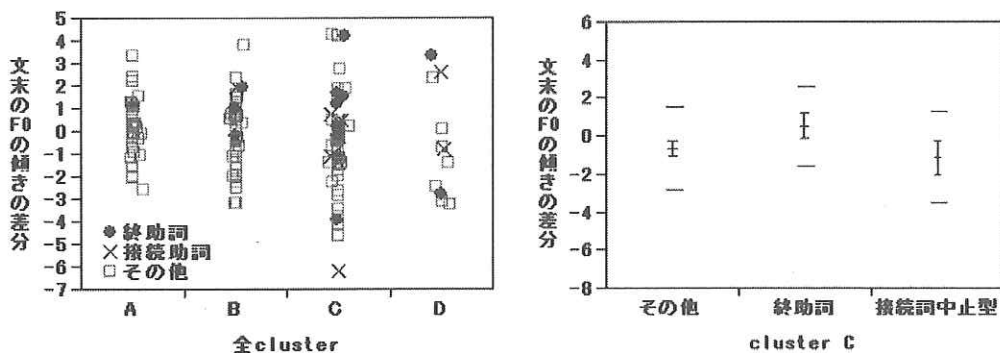


図2 文末30フレームの F_0 の傾きと手前30フレームの F_0 の傾きの差分とクラスターとの分布(ただし、左図の大きな丸点は終助詞、バツ印は接続助詞、ボックスはその他を示す。)

5. 4. 言語形式のパラメータ化

言語パラメータとして、各発話をまず平叙文、あるいは疑問文という括りで捉えることにし、文構成を特徴付けるパラメータL-1(平叙文または疑問文)を設定した。このL-1は独立する二つの言語形式を共に特徴付けるパラメータであり、特徴付けに一見ふさわしくないが、語尾のイントネーションを上げる平叙文が頻出することから、上昇イントネーションを疑問の助詞の代用と捉え、こうした平叙文も形式が異なるが疑問文とした。

さらに、文末形式というパラメータL-2を設定し、接続助詞中止型「ので・けど」と終助詞「ね・よ」を特徴付けることとした。接続助詞中止型「ので・けど」/終助詞「ね、よ」で終了している発話、ならびにその他の発話末形式に対してこのパラメータを付与した。ただし、「したいんですけど」を発話末形式とする発話については、慣用化した間接的要求を示す言語形式として捉え、接続助詞中止型とは区別し、パラメータL-2の対象とはみなさなかった。

言語パラメータに対する怒りの程度1から5の実測値の分布を見ると、パラメータL-1においては、平叙文より疑問文のほうが怒りの程度が高い発話が多く、怒りの度合い自体もほぼ3以上の値を示している。一方、パラメータL-2で捉えられる文末表現に対する分布では、すべての項目でバラツキが大きいものの、平均値をみると、その他・終助詞・接続助詞の順に怒りの度合いが上昇している。図3および図4にそれらの様子を示す。

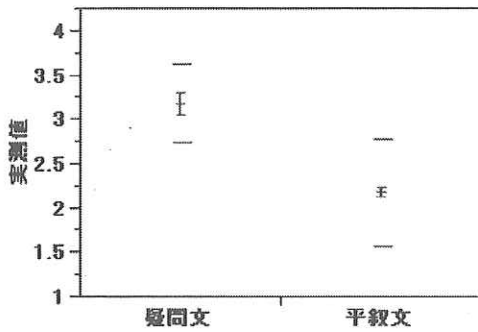


図3 文構成のパラメータに対する実測値の分布

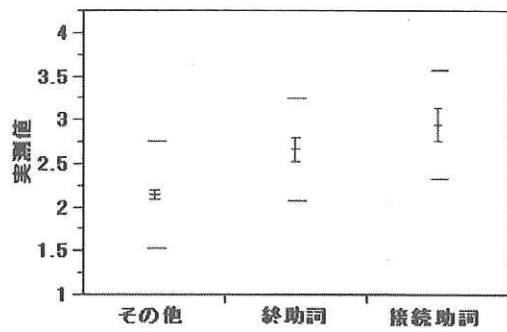


図4 文末表現に対する実測値の分布

6. パラメータ検証のための「怒り」推定実験

収録した音声データを用い、音響パラメータと言語パラメータを使用して、決定木による怒りの程度を推定する実験を行った。ここではその実験手法とその結果について述べる。

6.1. 音声データと特徴パラメータの組み合わせに基づく推定実験手法

収録した音声データにはデータ数に男女の偏りがあることが分かっている（有本他 2004, Arimoto et al. 2005 など）ため、全 159 発話を含む data set A の他に男性の発話のみ（141 発話）を含む data set B を用意した。各 data set に対し、 F_0 の絶対的な値を用いるパラメータ（F-1, F-2, F-3）を除いた 11 種の音響パラメータを用いた実験と 2 種の言語パラメータを加えた 13 種のパラメータによる実験を行った。さらに、data set A には女性の発話が含まれていることから、男女差の影響の程度を比較するため、14 種のすべてのパラメータを用いることにした。data set A による 14 種のパラメータを用いた実験を実験 I、11 種のパラメータを用いた実験を実験 II、言語パラメータを加えた 13 種のパラメータを用いた実験を実験 III とする。data set B による 11 種のパラメータを用いた実験を実験 IV、言語パラメータを加えた 13 種のパラメータを用いた実験を実験 V とする。

推定実験は決定木を用いて行った。交差検定を行うため、各 data set の発話を 3 つのグループに分割した。各 data set は発話数が限られているため、ランダムに三分割を行うと、グループ間に怒りの程度の偏りが生じてしまう。そのため、1 発話ずつを主観評価値の高いデータから順に 3 つのグループに振り分け、グループ間の怒りの程度の偏りを排除し、実験の信頼性の確保に努めた。決定木のための学習用データとして 2 グループを、評価用データとして 1 グループを使用する。この際、交差検証法を用いた最適木を求めるため学習用と評価用の各データに使用するグループを入れ替えて実験を行う。

学習用ならびに評価用データに含まれる怒りの種類に影響されない木を求めるため、評価

用データの分岐回ごとの予測値と主観評価による実測値の標準誤差を用いて、データに依存しない分岐回数を求めた。実験の評価方法として主観評価により付与された怒りの実測値と決定木による予測値との推定誤差を求め、推定性能の指標とした。

6.2. 結果と考察

実験の結果、音響パラメータのみを用いた実験Ⅰ、Ⅱ、Ⅳでは8回目、7回目、6回目が最適な分岐回となり、その際の平均推定誤差は0.54、0.60、0.53となった。音響パラメータに言語パラメータを加えた実験Ⅲ、Ⅴでは5回目、4回目が最適な分岐回となり、その際の平均推定誤差はともに0.50となった。さらに、音響パラメータのみの実験と言語パラメータを加えた実験の結果を比べると、男女混合の data set A では平均推定誤差に0.10差、男性のみの data set B で0.03差となり、音響的特徴のみの実験よりも言語的特徴を含めた実験のほうが推定精度が若干向上していることが確認できた。また、分岐回が早くなっていることも確認でき、情報識別能力が高いパラメータが加わったことが見て取れる。このことは、言語形式に関するパラメータによって目標とするデータが早めに選別され、その後、音響的特徴による仔細な特性が選別されていくことを示す。

実験結果を整理して、以下に特徴とその説明を列挙する。それにより、設定した音響パラメータと言語形式によるパラメータが相俟って共に「怒り」の程度推定に寄与していることが示せる。

<特徴1> 共通に選択されたパラメータ (F-7、P-2、D-2) :

音響パラメータのみを使用した実験Ⅰ、Ⅱ、Ⅳにおいて明確に現れた特徴であり、誰もが感じ取っていると思われる「怒り」の声の特徴に大きく関わる特性として現れている。発話内の声の高さのバラつきを示す F-7、瞬間的に大きな声を示す P-2、概して早口であることを示す D-2 が共通して登場することが分かる。つまり、男女差に関係なく、男女の発話が混合するデータ、ならびに男性のデータに対して、発話内で声の高さに大きな変化があり声が大きく、速い発話であることが共通に確認できる。

<特徴2> 怒りの発話は、はやくて声の高い発話と、ゆっくりで声の低い発話:

怒りの程度が高い発話には相反する傾向を示す2種の発話が含まれるという特性が決定木に現れていると言える。平均発話速度 D-2 が速くとゆっくりとに分別し、さらに F_0 の発話内最高値 F-6 に関して、声の低い発話と声の高い発話を分別する決定木の様子がそれを示す。

<特徴3> 疑問文である発話は怒りが強い:

言語パラメータ L-1 がほぼすべての実験で1回目あるいは2回目という早い分岐回で現れている。これは疑問文の文構成が怒りの程度の推定に大きく寄与したと言える。L-1 が2回目の分岐回に現れるときは、1回目の分岐回には F-7 あるいは P-2 が現れている。

つまり、声の高さあるいは大きさに係わるパラメータと疑問文とした L-1 の言語形式とが共に、「怒り」を識別する要因の中核にあり、音声言語という点で両者が不可分に連関していると解釈できる。

＜特徴 4＞文末形式も怒りを識別するパラメータとなる：

文末形式に関わる言語パラメータ L-2 が 3 回目という早い分岐に現れる決定木があり、怒りを選別する大きな要素として寄与していると思わせる。この L-2 が現れる状況は、実験Ⅲの男女混合データの実験において、まず L-1 による選別が行われ、2 回目の分岐で P-2 が現れ、それに続いて L-2 が現れている。このことは、声の大きさと言語形式とがそれぞれ「怒り」の推定に大きな識別能力を持っていることが分かる。ただし、男性のみのデータに対しては、L-2 の寄与度は小さく、L-1 と声の高さ、大きさ、速さに係わるパラメータが大きく寄与していた。

7. 課題と展望

「怒り」の発話を対象とした感情表現の程度を含めた推定実験について、その手法と実験結果を述べたが、詳細を充分吟味して、パラメータの設定に努める必要がある。とくに、言語形式に関するパラメータの選定においては、現状 L-2 に間接的な要求を示す接続助詞中止型や確認を表す終助詞「ね・よ」の用法も含まれているため、それらを区別した詳細な検証実験が必要である。また、現状のデータ数を確保する点から、発話末の上昇イントネーションをもつ平叙文を形式が異なるものの疑問文とした。また、収録した「怒り」の発話には、疑問や質問の終助詞「か」が現れることはまれで、反駁あるいは念押しを示す終助詞として多く使われていた。この点では、パラメータ L-1 を再整理して、言語形式から決まるパラメータと機能上の用法として使われるパラメータとを明確に設定し、相互の関わりを明らかにしていくための推定実験をしていくことが急務となる。

しかし、現状の音響パラメータのみを使うだけでも、意味ある音声対話の現象を捉えることが出来ると考える。例えば、今回の実験で使用していない発話を含めた約 1,500 発話を対象にして「怒り」の程度推定実験を行ってみると、「怒り」の程度が高い発話の言語形式には一定の傾向が見えてくる。「怒り」の度合いが高い順に、形式上の荒いタイプ分けに従って、上位の発話例を整理した結果を表 1 に示す。表 1 には「怒り」の推定値が 3.5 以上という強い印象を受ける発話があり、言語形式の特徴を探る上で貴重な資料になると考える。ただ、現時点での音響的な推定値は言語形式上のどこで計量されているかまでは特定できない。今後は発声時間に沿った声質の変化や拍単位と同期した音響情報までを考慮した「怒り」の正解データの作成が必要となる。

表1 「怒り」の強い発話例（「怒り」の度合いを score で、発話長を duration（秒）で示す）

utterance	score	duration (s)
・・・んで／なんで（終止形十ので／名詞類十なので）		
だいたい機械だってわからないじゃないですかしょっちゅうそういうのおこるんで	4.40	3.3
確かに予約したはずなんで	4.00	1.7
いやこっちはちゃんと予約したんでー	3.90	1.5
でも実際予約したわけなんで	3.45	1.8
いやでも確かにメモしたんで	3.30	1.7
・・・よ（です十よ十（ね））		
システムに名乗れて言っても名乗らなかつたんですよ	4.30	2.8
自分から見たらもうシステムのミスなんですよ	4.10	1.9
とにかくあのほんとあしたで困るんですよ	3.80	2.4
でもシステムが多分おかしいんですよ	3.80	1.7
いやでもホントにね 88542 って昨日聞いたんですよ	3.60	2.9
・・・ですよ／ですね十（副詞／接続詞）		
いやでもいつもはこんなことなかつたんですよ本当に	3.90	2.7
予約したんすよだって	3.87	1.2
そういわれてもこっちも困るんですよ本当に	3.66	2.6
それは嫌ですねもう	3.63	1.0
・・・か（終止形十終助詞か）		
それは機械がおかしいんじゃないんですか	3.88	1.2
こっちは引き下がるしかないじゃないですか	3.33	1.5
他の人のがちゃんと残っているってことは証明できるんですか	3.33	3.4
・・・たら／から／ば／（終止形十接続助詞）		
なんであのほんと用意できないんでしたら	3.80	2.3
そっちが一方的に言えば	3.66	1.2
予約したんすけど	3.50	1.5
こうやって現にメモってるんですから	3.33	2.5
平叙文・終止形		
こっちは苦情を入れてます	3.30	1.2
たしかにここに掛けさせていただきました	3.00	1.8

8. おわりに

本論では、パラ言語を含む音声言語による「怒り」の感情表現の程度を推定することを試みた。そのために、「怒り」の音響的な特徴ならびに言語形式的な特徴による特性を14種と2種のパラメータとして設定した。それらパラメータが「怒り」の特徴付けに関与する可能性があることを確かめる評価実験を行い、それにより選定されたパラメータを使って、決定

木を使った「怒り」の程度を推定する実験を行った。その結果、発話内での声の高さに大きな変化があり、声が大きく、速い発話が「怒り」が強く現れていることを示すことが出来た。さらに、疑問や接続助詞中止型の言語形式において、強い「怒り」が現れていることを示した。とくに、声の高さあるいは大きさに係わるパラメータと疑問文としたL-1の言語形式とが共に、「怒り」を識別する要因の中核にあり、音声言語という点で両者が不可分に連関していることを示した。また、文末形式に関わる言語パラメータが決定木において高い識別能力をもつことが分かり、怒りを選別する大きな要素として寄与していることを示した。

注

- 1) Ward 法とは、多変量の正規混合分布などにおける各クラスタでの尤度を最大にするようにクラスタを結合する手法であり、クラスタ間の分布を平滑化する。
- 2) 1次ケプストラム係数とは、スペクトル全体の傾きの大まかな傾向を表すもので、対数パワースペクトルの逆フーリエ変換で求めた次係数のうちの第1次係数として算出する。

参考文献

- Ang, J., R. Dhillon, A. Krupski, E. Shriberg, and A. Stolcke. 2002. "Prosody-Based Automatic Detection of Annoyance and Frustration in Human-Computer Dialog." *Proceedings of International Conference on Spoken Language Processing*, Vol. 3, 2037-2040.
- Arimoto, Y., S. Ohno, and H. Iida. 2005. "A Method for Discriminating Anger Utterances from Other Utterances Using Suitable Acoustic Features." *Proceedings of SPECOM2005*, 613-616.
- Banse, R. and K. R. Scherer. 1996. "Acoustic Profiles in Vocal Emotion Expression." *Journal of Personality and Social Psychology*, 70: 3, 614-636.
- Cowie, R., E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz, and J. Taylor. 2001. "Emotion Recognition in Human-Computer Interaction." *IEEE Signal Processing Magazine*, 18: 1, 32-80.
- Daly, E., W. J. Lancee, and J. Polivy. 1983. "A Conical Model for the Taxonomy of Emotional Experience." *Journal of Personality and Social Psychology*, 45: 2, 443-457.
- Plutchik, R. 1980. "Chapter 11: A Structural Model of the Emotions." In R. Plutchik (ed.) *Emotions: A Psychoevolutionary Synthesis*, 152-172, Harper and Row.
- Russell, J. A. 1989. "Chapter 4: Measures of Emotion." In R. Plutchik and H. Kellerman (eds.) *Emotion: Theory, Research, and Experience*, 83-111, Academic Press.
- 有本泰子・大野澄雄・飯田仁. 2004. 「『怒り』識別のための音声の特徴量の検討」、『人工知能学会研究会資料』、SIG-SLUD-A303 巻、13-19.
- 有本泰子・大野澄雄・飯田仁. 2006a. 「『怒り』の感情表現とその言語表現 - 音響的特徴の関係」、『言語処理学会第12回年次大会ワークショップ「感情・評価・態度と言語」論文集』1巻、53-56.
- 有本泰子・大野澄雄・飯田仁. 2006b. 「感情音声の収録と言語表現に現れる感情表現の分析」、『電子情報通信学会2006年総合大会講演論文集』、シンポジウム、42-43.
- 犬飼隆. 2001. 「低く短く付く終助詞『ね』」、『文法と音声Ⅲ』東京：くろしお出版.
- 片桐恭弘. 1997. 「終助詞とイントネーション」、『文法と音声』東京：くろしお出版.

- 杉藤美代子. 2001. 「終助詞『ね』の意味・機能とイントネーション」、『文法と音声Ⅲ』東京：くろしお出版。
- 武田昌一・大山玄・朽谷綾香・西澤良博. 2002. 「日本語音声における『怒り』を表現する韻律的特徴の解析」、日本音響学会誌, 58: 9、561-568.
- 横森大輔. 2006. 「接続助詞ケドの文末用法と話し手の態度」、『言語処理学会第12回年次大会ワークショップ「感情・評価・態度と言語」論文集』37-40.