自然発話音声に基づく快・不快の判別手法の研究

山本 泰史[1]　ロペズ ギヨーム[2]　酒造 正樹[2]　ドロネー ジャンジャック[2]　山田 一郎[2]
門間 史晃[3]　光吉 俊二[3]

[1]東京大学工学部　[2]東京大学大学院工学系研究科　[3]株式会社AGI

1. 緒言

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（雰囲気情報）が重要となる．比較的容易に取得可能な音声は，雰囲気情報の一つである感情情報の抽出が可能であり，その研究に注目が集まっている．本研究では感情判別と関連して，音声情報をもとにした快不快情動の判別アルゴリズムの開発を目標としている．

音声による感情判別の従来研究では 4 つの感情を60％程度の精度で判別するものや[1]，個人差の問題に取り組むもの[2]などが行われている．これらの研究において，どのように感情のラベルを音声に付与するかという問題や，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問題で 判別精度が向上しないという課題が残されている．

一方，本研究で着目する快不快情動は，発話の際に意図的に抑制が可能な感情よりも，評価ラベルが定まりやすく，また人の心理状態を知るための指標として期待できる．快・不快を判別する従来の研究には脳波[3]や，顔の表情[4]をもとにする研究はあるが，音声情報に基づく研究はあまりなされていない．よって，音声から特徴量を抽出し，どのパラメータが快・不快の判別に有効であるか明らかにすることは意義がある．

本研究では，脳科学等の文献[5]を参考に，快・不快の発生モデルの仮説を立て，判別手法の検討を行う．大勢の実験参加者から自然発話音声を取得し，判別に用いる音声とした．その音声から，主観に基づく快・不快のラベルを付与して，データベースの構築を行った．

2. 快不快と感情のモデル

人は喜怒哀楽など様々な感情を感じるが，それらの感情は快と不快の二つの情動が発生し，それらが派生した結果として生じるものである．本研究で用いる快不快と感情の発生モデルを Fig. 1 に示す．まず，人間は知覚や身体反応などの刺激を受け，興奮が発生し，そこから快と不快という２つの情動に分岐する．次に，それらの情動が喜び・怒り・悲しみといった感情へと派生する．また，情動の発生には，セロトニンやノルアドレナリンなどの複数の生体物質が関係しており，それらに伴って生理反応を示す．

例えばプレゼン時では緊張や一種の不快情動を持つが，予備検討を行った結果，声帯の緊張や，心電の変化，心拍数の上昇などが起きる．これらの生理反応は，マイク，心電センサなどを用いて，物理量として観測可能である．これらの物理量（生体情報）を分析することによって，快不快情動や感情情報の判別が可能になると仮説を立てた．

複数のセンサにより多角的に解析を行えば，より確度の高い分析結果が得られるが，本研究においては音声情報のみを扱うこととする．この選択理由は，音声情報をマイクで容易に取得可能であるとともに，実験参加者へのセンサ装着の負担をできる限り軽減するためである．

3. 音声収録実験

精度の高い判別を行うための要件として，
① 実験参加者からうまく感情を引き出すこと，
② その際の音声を収録すること，
③ 音声に適切な心理状態のラベルを付与すること，
④ 分析に十分なサンプル数があること，
などが挙げられる．このため，以下の手順で音声収録とラベル付与を行った（Fig. 2）．

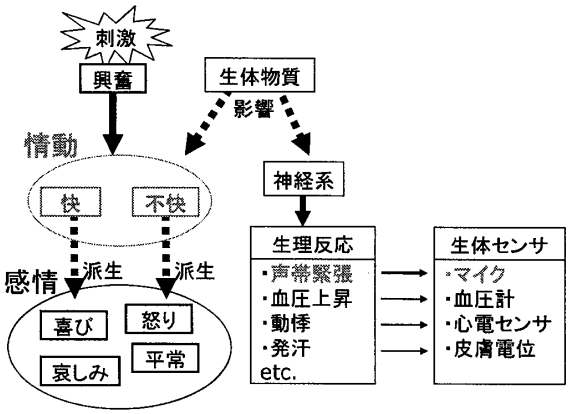

Fig. 1 快不快と感情の発生モデル

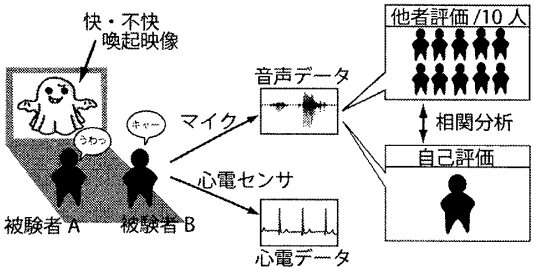

Fig. 2 音声収録実験の概要

Preliminary Study on Discrimination Techniques of Comfort and Discomfort from Natural Utterance Sound Analysis

Taishi Yamamoto[1], Guillaume Lopez[2], Masaki Shuzo[2], Jean-Jacques Delaunay[2], Ichiro Yamada[2], Humiaki Monma[3], and Shunji Mitsuyoshi[3]

1 Faculty of Engineering, The University of Tokyo
2 School of Engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan
3 AGI Inc., 6-3-13 Akasaka, Minato-ku, Tokyo 107-0052, Japan

Table 1 音声から抽出した特徴量

情報	特徴量
パワー	分散，最大値，最小値，対数分散， 対数最大値，Δ平均，Δ分散，Δ最大値， Δ最小値，Δ対数平均，Δ対数分散， Δ対数最大値，Δ対数最小値
ピッチ	平均，分散，最大値，最小値，レンジ， 発話値，対数平均，対数分散，対数最大値， 対数最小値，Δ平均，Δ分散，Δ最大値， Δ最小値，Δ対数平均，Δ対数分散， Δ対数最大値，Δ対数最小値，正規化最大値， 正規化最小値，正規化Δ平均

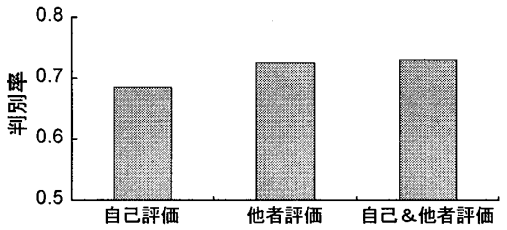

Fig. 3 快・不快の判別結果

まず，コントロールされた実験室環境下において，実験参加者に感情を誘発する動画を見せ，直後に感想を自由に述べてもらい，その発話音声を収録した．音声収録に際し，実験参加者にヘッドウォーン型マイクを装着し，音量に閾値を設けて発話を検知した．続いて，別の感情に関連する動画についても，同様に繰り返し操作を行った．音声収録実験を10~40代の男女計96名に対して行い，計19,149発話を収録した．

次に，発話者が自らの発話音声を聞き，発話時の心理状態の評価を行った（自己評価）．さらに，複数の第3者が発話音声を聞き，発話者の心理状態の評価を与えた（他者評価）．自己評価，他者評価ともに，評価の手法としては，快不快情動については「快」「不快」「どちらでもない」，感情については「平常」「喜び」「怒り」「悲しみ」「苦痛」「不安」「不明」のラベルの中からそれぞれ1つずつ選択させた．他者評価については5名以上の評価者で行い，8割以上が一致したものを他者評価として扱った．

4. 判別手法

まず取得したラベル付き音声の一部を学習データとして採用し，特徴量を算出して，快および不快のデータベースを作成した．次に，学習データ以外の音声の一部を試験データとして用い，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，先行研究の知見をもとにTable 1に示す基本周波数（ピッチ）と音量（パワー）に関する特徴量を採用した．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．

学習データからデータベースを作成する際に，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データを学習データとして使用した方が，判別に適したデータベースになると予想される．これを検証するため，「自己評価のみ」「他者評価のみ」「自己評価と他者評価が一致したもの」をラベルとした3種類の音声データベースを作成し，快・不快の判別を行った．各データベースを構成する音声ファイルの数は，それぞれ788，502，295である．

判別率の評価に用いた試験データには，評価の真値を一致させるために「自己評価と他者評価が一致したもの」を採用した．ファイル数は89である．なお，判別に用いた距離関数としては，分散を考慮してマハラノビス距離Dを採用した．

$$D=\sqrt{\sum_{i=1}^{p}\sum_{j=1}^{p}\sigma^{ij}(x_i-u_i)(x_j-u_j)} \qquad (1)$$

ここで，pは特徴量の数，xは判別対象データの特徴量，u，σ^{ij}はそれぞれ学習データの特徴量の平均値，分散共分散行列の逆行列における要素を表す．

5. 結果と考察

「自己評価のみ」「他者評価のみ」「自己評価と他者評価が一致したもの」を学習データとした場合，それぞれ68.4%，71.8%，72.2%の判別率を得た（Fig. 3）．このことから，「自己評価のみ」よりも，他者評価が一致した方が快・不快の音声が分離していると考えられる．また，「他者評価のみ」と「自己評価と他者評価が一致したもの」の判別率にあまり差が見られなかった．この原因は，「自己評価と他者評価が一致したもの」を学習データとして使用した場合の音声ファイルの数が他と比べて十分でないために，判別率が予想に比べ向上しなかったという可能性もある．そのため，今後はデータ数の増加や，他にも音声特徴量の選定，性別の考慮などを行う必要があると考える．

6. 結言

本研究では，約100人の実験参加者から多様な自然発話を取得し，現段階までにそのうちの1割の音声に自己評価と他者評価を付与した音声データベースを作成している．判別の結果，自己評価だけではなく，他者評価が加わったデータベースを学習データとして用いることによって快・不快の判別率が上がる可能性を示した．今後は快・不快の識別精度の向上を目指し，感情の判別へとつなげてゆく．

参考文献

[1] 門谷信愛希他，“音声に含まれる感情の判別に関する検討，”信学技報，Vol. 96, No. 1051, pp. 79-84, 1997.
[2] 志村誠他，“雰囲気コミュニケーション端末における音声を用いた感情抽出手法の研究，”ヒューマンフェースシンポジウム2007講演論文集，pp. 593-596, 2007.
[3] 高橋和彦，“SVMによるEEGからの感情識別に関する一考察，”人間工学，Vol. 39, No. 2, pp. 90-92, 2003.
[4] 坂本博康他，“顔画像解析による人間の快・不快の計測手法，”情報処理学会研究報告，2006-CVIM-155, 2006.
[5] J.E. LeDoux 著 松本元 他訳，“エモーショナル・ブレイン，”東京大学出版会，2003.