2—2—3

話者感情認識における音響的特徴についての検討*

◎影山良之 (奈良先端大) 坊農真弓 (神戸大学, CREST)
ニック・キャンベル (奈良先端大, ATR, CREST)

1 はじめに

音声には言語情報の他に話者の個人性や発話意図, 感情などのパラ言語情報などが含まれており, それらは音響的特徴に様々な影響を与えている. 音声の中に含まれている感情に関するパラメータを抽出, 分析することができれば, 感情を含んだ音声合成や, コンピュータによる人間の感情理解などの実現が期待できる.

音声情報に含まれる話者の感情についての研究はこれまで心理学や言語学, 音響学などで多岐にわたってなされてきた. PHYSTA プロジェクトでは, 顔や声から人間の感情を認識するための実験システムが提案されている [1]. 中でも"FEELTRACE" は心理学の理論を取り入れ, 音声に感情情報のラベリングを行い, 今後の感情音声の分析をする時に感情情報の抽出に利用するシステムとして提案されている [2,3].

本稿では, このシステムを用いて実際に話者感情ラベリングを行い, 話者感情ラベリングシステムとしての有効性を検討する.

2 システム

FEELTRACE のシステムを図 1 に示す. 被験者は音声を聞き, リアルタイムでその音声から受ける印象に相当する位置にポインタを移動させる. ポインタの移動範囲は画面上の円内に限定される.

円内の軸は x 軸 y 軸ともに-255 から 255 の値をとり, x 軸のプラス方向を肯定的 (positive), マイナス方向を否定的 (negative), y 軸のプラス方向を積極的 (active), マイナス方向を消極的 (passive) とし, また円の内外には被験者のポインタの移動の手助けとなる感情用語が布置されている. 今回用いた軸の名前と感情用語は PHYSTA で提案されていたものを日本語に訳したものを用いた. 入力のインタフェースにはマウスを用いる. マウスの動きに対して, システムは 2 つの視覚的なフィードバックを与える. 1 点はポインタの動きの軌跡であり, もう 1 点ポインタの色の位置により変化である. ポインタの色は x 軸が最大のとき緑, 最小のとき赤, y 軸が最大のとき黄, y 軸に最小のとき青, 原点では白となる. このシステムはラベル情報として時間情報とポインタの位置情報を保存する.

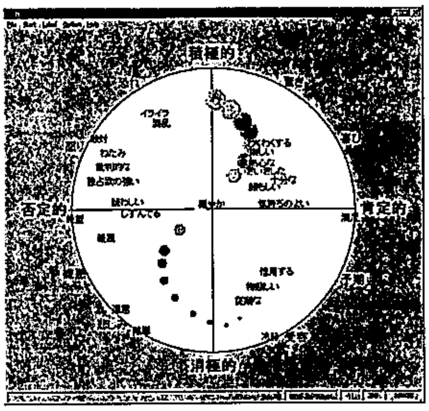

図 1: FEELTRACE

3 実験

音声資料に 4 分間の男女の自然対話を収録した. 提示する音声として音声資料 A に収録した音声をそのまま用い, 音声資料 B に発話順序を入れ替えて文脈情報をなくしたものを作成し用いた. 17 名の大学生, 大学院生 (男性 3 名, 女性 14 名) に対して本システムを用いて男性が発話している部分についての感情ラベリング実験を行った. なお, 本実験に先立ち被験者にはこのシステムの使用方法の説明と練習を行っている.

4 結果及び考察

図 2 に音声資料 A を聞いたときの全被験者の x 軸に対するポインタの変位を示す. この図より 100 秒の辺りから 140 秒の辺りにかけてすべての被験者がプラスの方向, つまり音声が肯定的であるという評価を行っている.

図 3 に音声資料 A と音声資料 B とで対応する部分を聞いたときの全被験者の x 軸に対するポインタ

*Labelling acoustic characteristics of emotional speech. By Yoshiyuki Kageyama(NAIST), Mayumi Bono(Kobe Univ., CREST), Nick Campbell(NAIST, ATR, CREST)

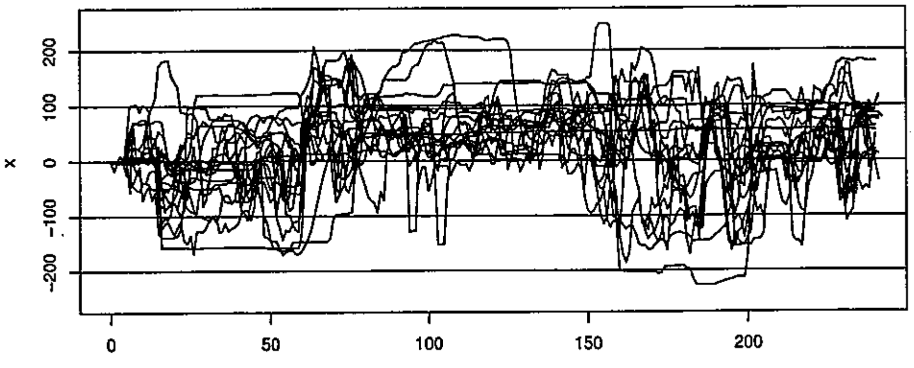

図 2: 音声資料 A に対する x 軸の軌跡 (17 名分)

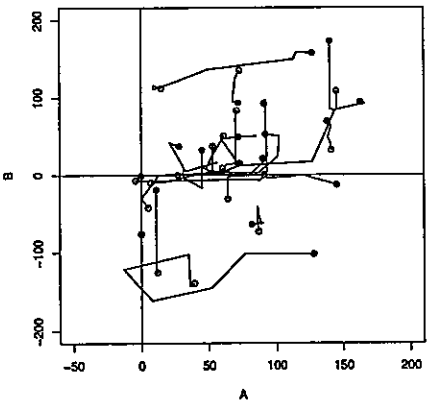

図 3: 音声資料 A,B の x 軸の軌跡

の変位を示す．このグラフの x 軸は音声資料 A を聞いたときの変位，y 軸は音声資料 B を聞いたときの変位をあらわす．また，○が始点，●が終点を表す．この音声は音声資料 A の 130 秒から 140 秒の辺りに相当する．つまり図 2 で全被験者の評価が一致した辺りである．この図より一見すべての被験者で評価が一致してた部分でも順序を入れ替えて提示してみることにより評価がばらついてしまうことが見て取れる．このような結果は他の場所で比較しても同様に得られた．

以上の結果より，今回実験で用いた FEELTRACE による音声情報から感情情報へのマッピング手法では文脈情報の影響が強くでていると考えることができる．被験者が音声から受ける印象を決定する前に次の音声が提示されてしまい，正確なマッピングが行えなかったのではないかと考えることができる．

また，話者間の系統的なマッピングが困難である原因に今回用いたシステムが音声と感情とのマッピングを 2 次元空間に対して行ったことも考えられる．音声から感情へのマッピングを 2 次元空間上に行うということは，被験者が音声を聞いてその音声から受ける印象を 2 つの軸で解釈しなければならなくなる．そのため被験者は 2 つの事柄を同時に考える必要となり，これがマッピングの難易度を上げてしまい，有益な結果を得ることができなくなったのではないかと考えることができる．

5　まとめ

本稿では実験を音声に感情要素のラベリングを試みた．今回の実験では有効な結果を得ることができなかったが，今後の実用的なラベリングシステムの開発において今回の結果を利用していきたい．

謝辞

本研究の一部は科学技術振興事業団戦略的基礎研究推進事業 (CREST/JST) の援助により行われた．

参考文献

[1] http://www.image.ece.ntua.gr/physta/

[2] Cowie R., Douglas-Cowie E., Tsapatsoulis N., Votsis G., Kollias S., Fellenz, W., & Taylor J.G., "Emotion Recognition in Human-Computer Interaction.", textitIEEE Signal Processing Magazine, Vol.18, No.1, pp.32-80(2001)

[3] Cowie R., Douglas-Cowie E., Savvidou S., McMahon E., Sawey M & Schröder M., "'FEELTRACE': An Instrument For Recording Perceived Emotion In Real Time.", *In Proceedings of the ISCA Workshop on Speech and Emotion*, pp.19-24(2000)