

聴取実験を考慮した自動感情判別と感情音声合成システム開発の試み\*

7N-8

小川 孝 重永 実 伊藤 誠  
(中京大学 情報科学部)

1. はじめに

音声に含まれる感情情報についての研究は、国内外において広く行われているが、筆者らは音声に含まれる感情情報の解明を目指し、計算機による感情評価と、感情を込めた音声を高品質で合成できるシステムの開発を試みた。

まず、サンプル感情音声について聴取実験を行い、感情表現の評価を求める。その求めた評価と、感情情報を持つと思われる音響パラメータに対し、多変量解析の手法[1]を用いることで、複数の感情が込められた音声の感情評価を試みた。また、これらの音響パラメータを制御し、原音声に様々な感情を付加することの出来る音声合成システムの構築も試みた。

2. 音声資料と分析

(a) 音声資料

大学の演劇クラブに所属している学生、男女各1名には2回、さらに男性1名には1回の計3名に、19種類の単語(表1)を、「平静」「喜び」「怒り」「嫌々」「悲しみ」のそれぞれについて、状況を自由にイメージした上で、発声してもらった。

この感情音声を12kHz, 16bitsでA/D変換し、フレーム長20ms, フレーム周期10msで分析を行った。

表1

はい、美しい、うれしい、すてき、綺麗、 素晴らしい、いや、嫌い、汚い、大好き、 大嫌い、早く、痛い、駄目、大きい、 小さい、可愛い、走って、静かに
--

(b) パラメータ抽出

分析により求めた音響パラメータは、最高基本周波数、先頭基本周波数、持続時間、最大振幅などで、

全て平静音声のパラメータで正規化した。なお、これらのパラメータには、何らかの感情情報が含まれていることが確認されている。[2], [3]

図1に先頭基本周波数と感情の関係を示す。

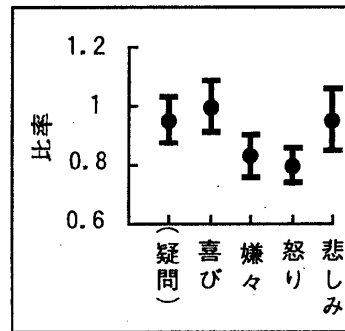


図1：先頭基本周波数の平静に対する比の平均値と標準偏差

(c) 聴取実験

また、これらの感情音声データについて聴取実験を行い、各感情毎の評価パラメータ(点数)を得た。実験は単語毎に平静の音声を提示したのち、異なる4種類の感情音声をランダムに、一定間隔をあけて2回ずつ提示した。この提示した音声について、感情表現の度合いを各感情毎に4段階評価で回答してもらった。

3. 計算機による感情評価方法

(a) 主成分分析

先に挙げた音響パラメータ  $O(o_1, \dots, o_m)$  は、個別でも感情毎にある程度の特徴が表れている。しかし、各パラメータ間の関係は独立したパラメータではなく、組み合わせ方によっては何らかの感情情報をもつと仮定して、主成分分析を行う。これにより、音響パラメータ  $O$  は新たにファクター  $C(c_1, \dots, c_m)$  とい

\* Automatic emotional evaluator considering listening test and emotional speech synthesizer.

う  $m$  個から成るパラメータに変換される。ここで、 $C$  は感情の説明変数として解釈できる。

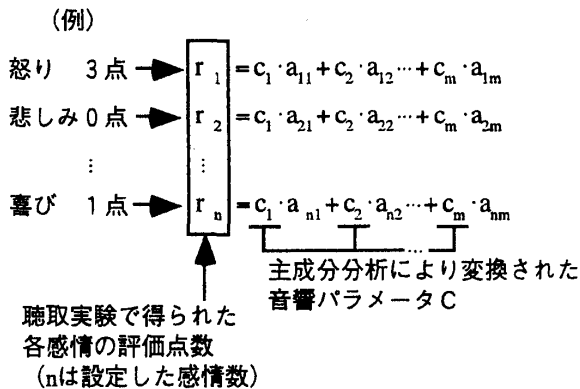


図2：重回帰分析による感情評価法

#### (b) 重回帰分析

次に変換されたパラメータ  $C$  を説明変数とし、先の聴取実験により得られた各感情毎の評価パラメータ (点数)  $R(r_1, \dots, r_n)$  を目的変数として、図2の式に代入する。この式で係数  $A(a_1, \dots, a_m)$  を求めることにより、音響パラメータと感情評価点数の間に対応関係が得られる。

なお、これらの分析は全て SAS を使用して行った。

## 4. 感情付加可能な音声合成システム

先の分析時に抽出した音響パラメータを用いて、平静音声に感情を付加できるような音声合成システムの構築を試みた。

以前はターミナルアナログ型音声合成を用いていたが、今回は高品質の合成音声を得る目的のため、波形編集方式を選択した。

変換音声の元となる入力音声には、あらかじめラベル付け処理を施しておく。ただし、このシステムでのラベル情報は厳密である必要はなく、無声、有声、無音区間の区別が出来る程度で良い。そしてこのラベル情報毎に、図3で示した制御パラメータ (基本周波数パターン、各音素の継続時間、振幅包絡) を与えて時間伸縮、基本周波数変更、振幅包絡変更などの波形編集処理を行う。

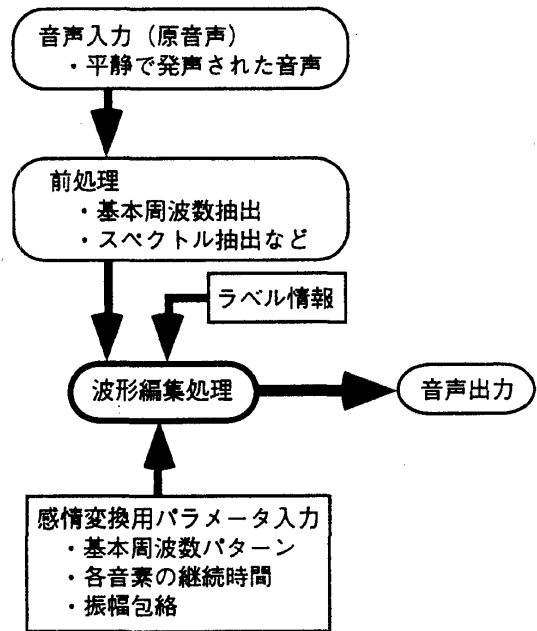


図3：感情付加型音声合成システム概略図

## 5. おわりに

今回は多変量解析の手法を用いることにより、計算機に感情を評価させることを試みたが、まだ問題は多い。特に聴取実験の評価方法や設定した感情の種類、重回帰分析の方法などは今後も検討が必要である。

また、音声合成に関しても、現在はラベル情報を前提として処理を行っている。今後は音響パラメータからの波形編集部分の自動抽出機能を追加したり、音質改善索として制御パラメータにフォルマント周波数や帯域幅などを加えてみる必要があると思われる。

## 参考文献

- [1] 森山, 斎藤, 小沢: "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", 信学技報, SP95-67 (1995-10)
- [2] 小川, 重永, 伊藤: "文音声に含まれる感情表現の基本周波数パターンについて", 平7秋音響学講論 1, 1-1-4, pp. 205-206, (1995-09)
- [3] 小川, 重永, 伊藤: "感性情報を利用した音声合成", 第51回情報処理全国大会, pp. 2-119-120, (1995)
- [4] 中村ほか: "高品質リアルタイム話速変換システム", 信学技報, SP92-55 (1992-09)
- [5] 新居, 西村ほか: "高品質音声編集システムに関する検討", 信学技報, SP93-141 (1994-02)