

感性情報を利用した音声合成

副題：単語音声に含まれる感情情報*

3R-2

小川 孝 重永 実 伊藤 誠 (中京大学情報科学部)

1. はじめに

音声を持つ情報には、言語情報をはじめとして個人情報や、感情情報など様々なものがある。このうち、感情情報については国内や外国において数多く研究されているが、感情そのものの特性がまだ良く分かっていないこともあり、感情情報を考慮した音声合成や認識には、まだ解決しなければならない問題が多いのが現状である。そこで、筆者らはまず、単語音声を対象として4つの感情（「喜び」「怒り」「嫌々」「悲しみ」と平静、疑問の計6通りの発声を設定し、この設定を踏まえて発声してもらった19種類の単語について、特に平静音声との変化分に注目した韻律情報の分析を行った。さらに、これらの分析結果を用いて、単語音声の任意の感情変換を試みた。

2. 音声資料と分析

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」「疑問」「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。計5回の発声で570サンプル用意したが、このうち4つの感情表現の380サンプルについて、聴取実験を行い、結果の良好であったもの147サンプルを抽出し、selectグループとして検討した。音声の分析は12kHz, 16bits, フレーム長20ms, フレーム周期10msで行った。

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
--

3. 分析結果

3.1 最高基本周波数：図1(a)に基本周波数パターン最高値の平静のそれに対する比の平均値と標準偏差を示す。

3.2 先頭基本周波数：図1(b)に基本周波数パターン先頭値の平静のそれに対する比の平均値と標準偏差を示す。ここで、「喜び」と「怒り」に注目すると、図1(a)では「喜び」と「怒り」の平均値、標準偏差の間に明確な差が無かったが、図1(b)を見ると、「喜び」と「怒り」の差は幾分大きくなっている。

3.3 持続時間：各単語の発声持続時間を平静のそれで正規化したものを図1(c)に示す。実際には、各モーラ毎に伸縮の程度に差がある。

3.4 振幅包絡：音声波形の振幅包絡の最大値を無音区間の平均値で割った比を平静のそれで正規化したものを図1(d)に示す。

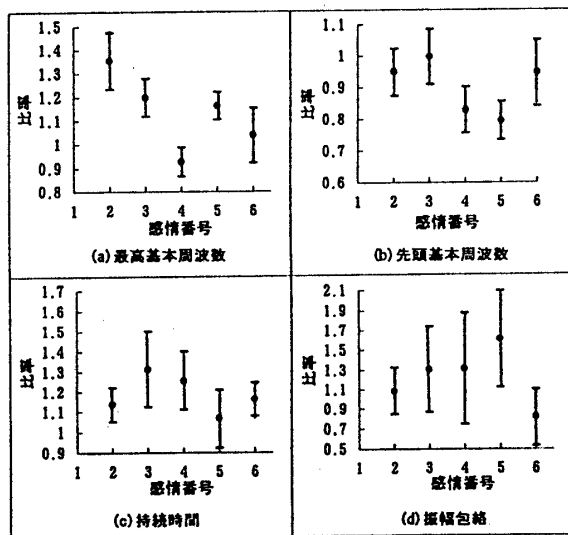


図1 各種パラメータの平静に対する比の平均値と標準偏差 (selectグループ)
感情番号：1. 平静、2. 疑問、3. 喜び、4. 嫌々、5. 怒り、6. 悲しい

* Emotional information expressed by spoken words.
By Takashi Ogawa, Minoru Shigenaga, Makoto Itoh.
School of Computer & Cognitive Sciences, Chukyo Univ.

3. 5 基本周波数パターン：各表現に対する基本周波数パターンの形状を大分類した⁷⁾。各表現に共通に表れるパターンもあるが、表現独自のパターンも存在する。但し、' ^ ' の型と言っても、変化幅の大小や上昇・下降の勾配などが感情によって変化していたり、僅かな変形も存在する。発声時に抱いたイメージや表現の仕方に個人差があり、単語の意味やアクセント型、音素の並びなどの影響もあると思われることから、単語数や発声者数を増やせば、さらに異なったパターンが現れる可能性がある。

4. 平静音声への感情の付与

3. の結果から作成した感情表現に寄与するパラメータを抽出して、平静に発声した音声波形に感情を付与することを試みた。まず、平静の音声を分析し、周波数スペクトル包絡、基本周波数パターン、振幅包絡(音源の三角波列情報)を抽出し、簡略化したABSの手法を使って、ホルマント周波数とその帯域幅を求める。これらの情報を使ってターミナルアナログ型合成器でまず平静の音声合成を行うと共に、これら3種のパターンを別々に表示する。合成パラメータ加工のツールはこれら3つのパターンを互いに独立に制御し、合成音を聞くことができる。従って、各特徴パラメータの寄与の程度を調べることができる。

5. 考察

状況も文脈もなしに、単語に感情を込めて発声してもらうことも、聴取も容易ではないが、分析の結果からは、色々な知見を得ることができた。

1. 「喜び」と「怒り」を区別する情報としては、分析結果3. 2より、先頭基本周波数を用いるのが良いようである。
2. 疑問の基本周波数パターンは語尾の母音の基本周波数が高くなる。それ以外は、平静のパターンとほぼ同じである。但し、最後の文節が無声子音で始まる場合にはこの出だしの基本周波数がかかなり高くなりがちである。また、他の感情表現パターンも最終基本周波数が高くなると疑問の感じが加わる。
3. 「嫌々」の感情には「怒り」の感情を含む場合が

多い。一方、基本周波数、振幅包絡などは「悲しい」と似ているところがあるので、「怒り」や「悲しい」への異聴が現れることが多い。

4. 例えば「怒り」の表現には疑問を含めた「怒り」があり、疑問特有の最終基本周波数の上昇または下降のないパターンが現れる。発声者の意図と異なった聴取結果が現れるものには、回答された感情特有のパターン(基本周波数だけでなく、持続時間や振幅包絡も含む)が幾らか含まれている。

5. 例えば、「駄目」「嫌い」のように意味から考えて通常楽しく発声することが殆どない音声は、発声者の努力にも関わらず、「喜び」が殆ど「怒り」に聴取されている。

6. 「嫌々」には基本周波数変化幅の小さい' ^ '形をした基本周波数パターンが、小振幅で付加されることが多い。「怒り」「喜び」にも付加されることがある。

7. 平静音声に目的の感情の韻律情報を付与することは比較的容易にできるようになったが、出発点となる平静の合成音の音質が、変換結果の感情の知覚に大きく影響するため、品質の良い平静音声の合成データの容易な作成が重要な課題となっている。

6. おわりに

基本周波数パターンの分類、感情音声への変換などに多くの問題点を残しているが、19単語に感情を込めて発声した音声を分析し、韻律情報の特性を調べ、韻律情報のモデル化、規則による感情音声の合成を試みた。今後、更に検討を続けると共に、合成音の質を良くしていきたい。本研究を手伝っていただいた中尾光志君、池田元君らに深謝します。

参考文献

1. Murray, I. R. and Arnott, J. L. : J. A. S. A. 93, pp. 1097-1108 (1993).
2. Carlson, R., Granstrom, B. and Nord, L. : ICSP92, pp. 671-674 (1992).
3. 杉藤：日本人の声, p15, 和泉書院(1994).
4. 上林, 小林, 新美：信学技報, SP92-131 (1993-02).
小林, 新美：平5秋音響学講論1, 2-8-7, pp. 233-234 (1993-10).
5. 藤崎 他：平6春音響学講論1, 2-7-18, pp. 83-84 (1994-03).
6. 富安, 有泉：平6春音響学講論1, 2-Q-15, pp. 293-294 (1994-03).
7. 小川, 中尾, 重永：平7春音響学講論1, 2-4-2, pp. 267-268 (1995-03).
重永, 小川, 中尾：信学技報, SP95-15 (1995-05).