

Title	音声中の感情表現に関連する物理量とその制御に関する研究
Author(s)	杉本, 隆
Citation	
Issue Date	2000-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/1357
Rights	
Description	Supervisor: 赤木 正人, 情報科学研究科, 修士

音声中の感情表現に関連する物理量と その制御に関する研究

杉本 隆

北陸先端科学技術大学院大学 情報科学研究科

2000 年 2 月 15 日

キーワード: 感情情報, 合成音声, STRAIGHT, Lombard Effect.

1 はじめに

人間らしく聞きやすい合成音声実現のためには、言語情報だけでなく、人間同士において重要な役割を担っている感情情報を扱う必要がある。

これまで、音声中の感情情報に関する特徴量として基本周波数・振幅・発話時間が挙げられており、基本周波数の時系列変化などに関する研究が行なわれてきた [1, 3]。しかし、これらの研究で作成された合成音声は感情をうまく表現できていない。この一要因として、感情同士、または音声の生成機構に起因する物理量同士の関係が、合成音声で記述されていないことが考えられる。

そこで、本研究では音声中の物理量の変化が感情識別に与える影響を調べ、その結果と音声の生成機構に起因する物理量間の関係を組み合わせることにより、感情制御のための物理量変換ルールを構築した。また、平静音声を物理量変換ルールによって感情制御し、感情知覚への影響を調べた。

2 分析

2.1 音声データ

分析に用いる音声データについて説明する。

発話者は 20 歳代の男性 2 名、女性 3 名であり、演劇経験者、ヴォーカルスクール通学者、声優養成所通学者のいずれかである。このような人選を行なったのは、演劇経験者な

どが一般人(演劇経験なし)に比べ、音声によって感情状態の表現を行なう手法を的確に心得ているからである[2]。

言語情報は感情に対して、かなりの影響を及ぼすことが想像できる。また、本研究では基本周波数を扱う。よって音声データに用いる文章は、感情的自由度が高く、母音または有声子音で構成された「いいじゃない」という言葉を用いた。

様々な文脈で「いいじゃない」が発声される対話文を用意し、発話者には文脈を考慮した上で自由に感情を込めて発声してもらった。合計 167 サンプルの音声を採用した。録音は防音室にて行なった。対話全体を防音室の外の DAT レコーダに入力し、標本化周波数 48kHz で録音した。これを標本化周波数 20kHz にダウンサンプリングして、特定音声部分を切り出し、ワークステーション (WS) に保存し音声データとした。

2.2 感情分類実験

本研究では、話し手ではなく、聞き手に存在する感情を対象とするため、採取した音声の感情分類のための聴取実験を行なった。

女性話者 1 名の音声を呈示し、その音声が平静・歓喜・怒り・悲哀のどの感情のものを強制判断させた。練習 4 サンプル、本番 55 サンプルをランダムに呈示した。被験者は大学院生 11 名であり、その全てが正常聴力を有する。被験者は防音室内でヘッドホンにより受聴した。受聴はモノラルの両耳受聴である。被験者には聞き直しを許し、PC を用いて回答させた。音声データは防音室の外に設置された WS 内に保存されており、被験者の応答に応じて呈示される。

聴取実験の結果、各感情において、高い識別率¹のサンプルが多数存在しており、分析するにあたり十分な結果が得られた。

2.3 分析結果

音声分析変換合成法 STRAIGHT[4] により音声中の各物理量を抽出し、2.2 で分類した感情間の相違を調べた。

分析の結果、各感情の表現に対して、過去の研究[1]とほぼ同じ傾向が得られた。但し、怒りの感情表現で発話時間が平静より長いという、過去の研究結果と異なる点もあった。

3 感情制御のための物理量変換ルール

分析結果をフィードバックさせ、物理量変換ルールを構築した。また、どの感情表現においても子音部は時間的に伸縮されないことから、スペクトルを手がかりに、時間軸変換マップを作成し、発話時間を伸縮した場合に子音部があまり伸縮されないようにした。

¹あるサンプルについて、同一の感情を表現していると回答した被験者の割合を識別率と呼ぶことにする。

さらに、Lombard Effect を利用し、物理量変換ルールに、音声の生成機構に起因する物理量同士の関係を組み入れる。

3.1 Lombard Effect

雑音の存在で発声者の発声様式が変化し、その結果音声波形が変形することを Lombard Effect と呼ぶ [5]。これは、声量が大きくなることによって他の物理量も変化する例である。平静よりパワーが大きい怒りの感情制御において Lombard Effect のモデル化は有効だと考えられる。Lombard Effect によれば、声量が大きくなることによって、基本周波数の上昇、ホルマント周波数のシフト、高域スペクトル成分の増加が生じる。

この効果を組み込んで構築される、感情制御のための物理量変換ルールを以下に示す。

歡喜：発話時間・基本周波数の長時間平均・パワーの長時間平均は平静と同じ。

基本周波数の変化率を 2 倍に増加させる。

語尾のみ基本周波数を上げる。

怒り：発話時間を 1.2 倍に伸長させる (子音部伸縮なし)。

基本周波数の長時間平均を 1.2 倍に増加させる。

パワーの長時間平均を 1.4 倍に増加させる。

基本周波数の変化率を 1.4 倍に増加させる。

ホルマント周波数をシフトさせる。

高域スペクトル成分を 5dB 増加させる。

悲哀：発話時間を 1.5 倍に伸長させる (子音部伸縮なし)。

基本周波数の長時間平均を 0.9 倍に減少させる。

パワーの長時間平均を 0.8 倍に減少させる。

基本周波数の変化率を 0.5 倍に減少させる。

4 聴取実験

感情制御した合成音声、感情を含んでいるのか確認するために、聴取実験を行なった。

Close Test として、分析に用いた音声の平静音声を物理量変換ルールを適用し作成した音声刺激を用いて聴取実験を行なった。さらに Open Test として ATR 音声データベースより「けっこうです」「そうですか」について同様にルールを適用し作成した音声刺激を用いて聴取実験を行なった。実験方法は感情分類実験と同じである。

聴取実験の結果、Close Test では、全ての感情において高い認識率が得られた。しかし Open Test では認識率が低い感情も存在した。これは言語情報に影響を受けたためだと考えられる。

しかし、Close Test、Open Test を通じて多くの共通項を見いだすことができた。文章に依らず、各感情表現において、基本周波数の変化率が強く関わっており、特に悲哀については、重要な役割を担っていることが分かった。

また、怒りは、発話時間が長くても十分に表現されていた。これは、怒りの感情表現に対して発話時間という物理量が重要でないということではなく、Hot Anger と、Cold Anger との違いに起因するものだと考える。

5 おわりに

本研究では音声中の物理量の変化が感情識別に与える影響を調べた。その結果、過去の研究結果とほぼ合致した。

調べた結果と、音声の生成機構に起因する物理量間の関係より、感情制御のための物理量変換ルールを構築した。その中で、パワーと、基本周波数・ホルマントのシフト・高域スペクトルの増加を結び付けた。

平静音声を、物理量変換ルールによって感情制御し、感情知覚への影響を調べた。その結果、各感情表現において、基本周波数の変化率が強く関わっており、特に悲哀については、重要な役割を担っていることが分かった。怒りの感情表現に対し、Lombard Effect を利用して、物理量同士を関連づけたところ、認識率の向上が得られた。

参考文献

- [1] 北原義典, 東倉洋一: “音声の韻律情報と感情表現”, 信学技報, SP88-158, Mar.1989.
- [2] 平賀裕, 斎藤善行, 森島繁生, 原島博: “音声に含まれる感情情報抽出の一検討”, 信学技報, HC93-66, pp.1-8, Jan.1994.
- [3] J.E.H.Noad, S.P.Whiteside and P.D. Green: “A MACROSCOPIC ANALYSIS OF AN EMOTIONAL SPEECH CORPUS”, Eurospeech, pp.517-520, 1997.
- [4] 河原英紀: “聴覚の情景分析と高品質音声分析変換合成法 STRAIGHT”, 音学講論, 1-2-1, pp.189-192, Sep.1997.
- [5] 吉田勝, 小畑秀文: “ロンバート効果を考慮した低品質単語認識の一手法”, 音学講論, 1-Q-21, pp.183-184, Oct.1994.