

発話時の韻律的特徴の変化を考慮した発話のテキスト化

熊谷 亮介† 安井 浩之† 吉野 邦生†

東京都市大学‡

1. はじめに

感情には声の高ぶりなど、それぞれを特徴づける韻律的特徴があり、そのような特徴から人は相手の感情を推察している。しかし、テキストのみで行うコミュニケーションでは、文字以外の情報がないため書き手の感情と読み手の受ける印象にずれが起きるといった問題がある。

本報告では、発話時の韻律的特徴の変化をテキストに反映することで、人が音声から感情を推察するのと同様の効果が得られると仮定し、その表現方法について調査、検討する。具体的には、音声認識の際に韻律的特徴の変化を抽出し、認識されたテキストに、変化に対応した装飾・表現を行う。さらに装飾したテキストに対し評価を行い、表現方法の検討を行う。

2. 関連研究

2.1 印象空間における音声と文字の対応関係の分析

宮島ら[1]はパラ言語・非言語情報が付加されたテキスト(パラテキスト)を生成するために、文字と音声の印象空間における対応関係について分析及び考察を行っている。パラテキスト生成のために、「すみません」という文に対して「怒り、悲しみ、無感情、驚き、喜び」の5感情で発話データ収録し、その発話データに対してプロのグラフィックデザイナーが感情に合わせたパラテキスト化を行った(表 1)。印象調査実験の結果、音声データとパラテキストとの間には受ける印象のずれが生じることがあることが分かった。また、パラテキストだけではなく言葉自体の意味が、受ける印象に影響することがあることが分かった。

表 1. 手動で生成したパラテキスト[1]

性別: 演技	パラテキスト
男性: 喜び	すみませえん
男性: 無感情	すみません。
男性: 悲しみ	すみません
男性: 怒り	すみません

Speech transcription in consideration of changes in prosodic features during utterance

†Ryosuke Kumagai, Hiroyuki Yasui, Kunio Yoshino

‡Tokyo City University

2.2 態度・感情音声における韻律的特徴の考察

川波ら[2]は複数話者による「怒り、喜び、悲しみ、無感情」の4感情で収録した発声データを分析した結果から、感情ごとの一般的な特徴を示し、次に感情の大きさを変化させた場合に韻律的特徴がどのように変化するかを調査した。その結果、感情ごとに特有の韻律的特徴があることがわかり、感情の大きさの違いに比例して一部の特徴量の大きさにも変化が見られた。

3. 提案手法

3.1 概要

本報告では、発話データに対して音声認識と韻律的特徴の抽出を行い、韻律的特徴の基準値と比較を行うことで、韻律的特徴の強度に応じてテキストに装飾・表現を自動付加するシステムの構築を行う。また、その評価を行い、表現方法についての検討を行う。

3.2 システム構成

本システムは大きく分けて、音声認識、特徴量抽出、特徴量の強度判別、装飾付加の4つの部分に分けられる(図 1)。処理の流れは、まずマイクから入力された音声に対し、音声認識ツールの Julius[3]を用いて発話ごとの音声データと認識結果のテキストデータを生成する。その後、praat[4]を使用して音声データから特徴量を抽出する。抽出する特徴量については3.3に記述する。

次に、抽出した特徴量と基準値を比較し、特徴量の強度を決定する。その結果を装飾付加プログラムに送り、認識結果のテキストデータを用いて、特徴量の強度を反映した装飾を付加したテキストを生成する。

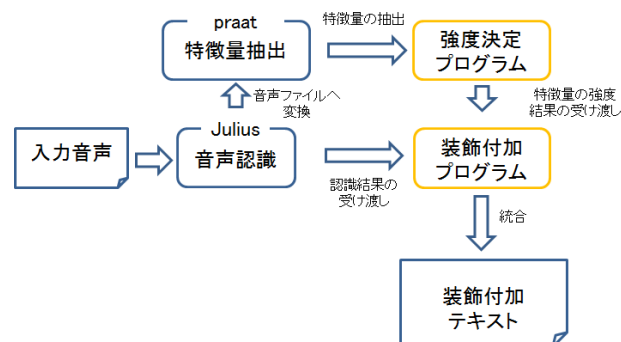


図 1. システム概要

3.3 特徴量の抽出

3.3.1 予備実験

本システムでは特徴量を決定するために、国立情報学研究所 音声資源コンソーシアム発行の研究用感情音声データベース[5]を用いて、予備実験を行った。ここで調査した特徴量は、基本周波数(最高, 最低, 平均, ピッチレンジ), 音声パワー(最大, 最小, 平均, パワーレンジ)の8つで、韻律的特徴の変化が顕著になるとされる怒り, 喜び, 悲しみの感情について感情の強さごとに分析を行った。

3.3.2 結果

予備実験の結果から感情および感情の強度ごとに大きな差が確認できた最大パワー, 平均パワーを本報告で用いることとする(表 2, 3)。なお, 複数の強度を考慮して特徴量の基準値を設定すると, 他の感情の強度と範囲が重複してしまい, 特徴量で強度の判別がうまくできないことがあるため, 本報告では, 強度 3 の場合のみを対象とし, 特徴量(基本周波数とパワー)を 1 フレーム(0.1 秒)ごとに抽出した時系列情報として用いる。最大パワーは 1 発話内のパワー上位 10%を平均したもの, 平均パワーは 1 発話のパワーを平均したものとする。

表 2. 各感情の強度別最大パワー比率

	強度 0	強度 1	強度 2	強度 3
怒り	1 (無感情)	1.55	2.30	3.48
喜び		1.50	2.19	2.90
悲しみ		0.95	1.40	1.83

表 3. 各感情の強度別平均パワー比率

	強度 0	強度 1	強度 2	強度 3
怒り	1 (無感情)	1.40	1.90	2.54
喜び		1.34	1.88	2.43
悲しみ		0.99	1.41	1.78

3.4 特徴量変化の強度測定

強度測定は基準値との比率でどの程度特徴量の変化があるかを算出し, その変化量に応じて強度を -3~3 の整数離散値でランク付けする。基準値は音声データベース[5]の音声データの中で無感情の状態のものから発話全体に対しての平均基本周波数, 最大パワー, 平均パワーを算出することで求める。

フレームごとの基本周波数とパワーの強度測定は, 1 フレームごとにそれぞれの基準値と比較することで, 時間ごとに強度測定を行い, 最大パワー, 平均パワーは 1 発話ごとにそれぞれの基準値と比較し, 強度測定を行う。

3.5 装飾テキストの生成

装飾テキストの生成は, それぞれの特徴量の変化に対して, 1 文字ずつ対応する装飾を付加することで実現する(図 2)。装飾には, 文字の位置(高さ), 色, 大きさ, 字体を用い, HTML タグを用いて実現する(図 3)。ただし, 字体は読みにくくなることを避けるため, 1 つのテキストデータに対して 1 つの字体だけを適用する。

特徴量の変化		小	大
装飾 の 変 化	表示位置	低い	高い
	色	青	赤
	大きさ	小さい	大きい
	字体	明朝体	ポップ体 ゴシック体

図 2. 特徴量と装飾の変化

だからポーションがぶ飲みしながら、倒さなきゃいけないんだよ
すげえ。なんか懐かしいな
せっかく頑張ったのに

図 3. 装飾付加テキストの例

4. まとめ

各感情における韻律的特徴の違いから, 特徴量変化の強度の違いを示した。しかし, 感情の強度を考慮して設定を行うと, 強度 1 の怒りの音声に対して悲しみを思わせるような装飾が付加されてしまうことがあった。そのため, 強度 3 のみを対象としたが, 実際の感情の強さは一意ではないため, 数値の違いだけではなく, 各感情に固有の特徴量を発見, 設定することが必要である。また, 今後, 特徴量の変化と装飾の変化の組み合わせを変えて評価を行い, 適当な組み合わせを設定していく必要がある。

参考文献

- [1] 宮島 崇浩, 菊池 英明, 樽松 明, 白井 克彦 “印象空間における音声と文字の対応関係の分析” 言語・音声理解と対話処理研究会 53,7-14, 2008-07-19
- [2] 川波 弘道, 広瀬 啓吉 “態度・感情音声における韻律的特徴の考察” 信学技報 SP97-67, 1997, p.73-80
- [3] Julius <http://julius.sourceforge.jp/>
- [4] praat <http://www.fon.hum.uva.nl/praat/>
- [5] 国立情報学研究所 音声資源コンソーシアム 25. 感情評定値付きオンラインゲーム音声チャットコーパス (OGVC) <http://research.nii.ac.jp/src/OGVC.html>