

音声における感情の強弱による韻律制御規則を適用した合成音声の評価

宮坂 直樹[†] 政倉 祐子^{††} 大野 澄雄^{††}

[†] 東京工科大学大学院 バイオ・情報メディア研究科 ^{††} 東京工科大学 コンピュータサイエンス学部

1 はじめに

音声により感情情報を伝達する際に、それを表現する重要かつ主要な特徴といえる韻律上の各特徴量との関係を明らかにし、表現豊かな音声合成の実現を目指している。本稿では、感情の種類として“喜び”、“悲しみ”を取り上げ、弱・中・強の3段階の程度で感情を表現した発話を分析対象とした。韻律的特徴の分析結果から統計的手法により程度及び感情の韻律制御規則の導出を行った。また、導出した規則を適用した合成音声を作成し、聴取実験により感情情報の伝達性を評価した。

2 分析結果

2.1 韻律制御規則の導出

F_0 パターンの生成には、その生成過程に基づくモデルを利用し、感情表現の程度に応じてモデルへの入力パラメータを決定することとした。筆者らは既報 [1] において、中立な発声に対する喜びと悲しみの感情表現におけるモデルへの入力パラメータの変化の規則性を求めた。図 1~5 は、その結果を簡潔にまとめたものであり、本報ではここで得られた結果を各感情の韻律制御規則として適用する。

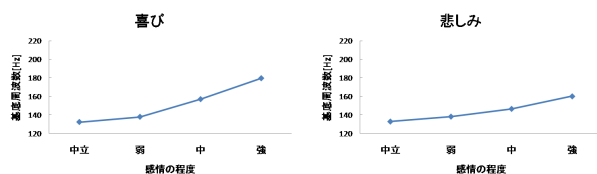


図 1: 感情の強弱に対する基底周波数の平均値

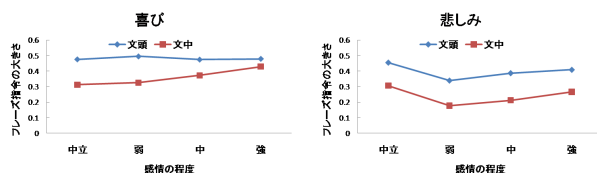


図 2: 感情の強弱に対するフレーズ指令の大きさの平均値

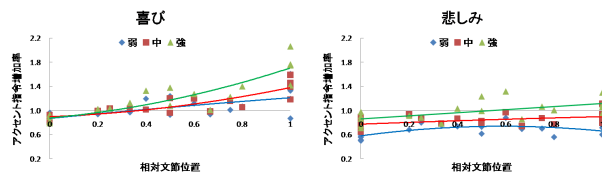


図 3: 相対文節位置に対するアクセント指令の大きさの増加率

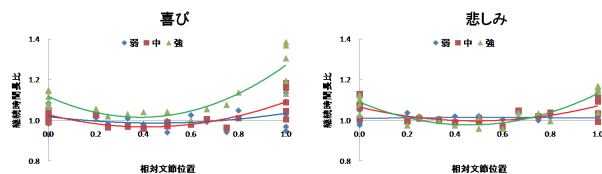


図 4: 相対文節位置に対する継続時間長比

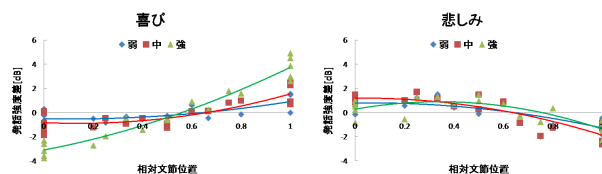


図 5: 相対文節位置に対する発話強度差

2.2 韻律生成のためのモデル化

韻律的特徴である F_0 パターンや発話速度、発話強度は、感情の種類とその程度のほか、表 1 にあげる文の言語的要因に影響を受けると考えた。そこで、感情の程度の違いが韻律的特徴にどのように影響するかについて、言語的要因との関連から検討した。

表 1: 韻律パラメータの変化に影響を与える要因

制御要因
1. 感情の程度 (無、弱、中、強)
2. 前置品詞 (17 種類)、後置品詞 (18 種類)
3. 前回のフレーズ指令生起からのモーラ数 (3 ~ 23)
4. 文節のモーラ数 (3 ~ 13)
5. 文の枝分かれ境界 (左枝分かれ、右枝分かれ)
6. フレーズ指令の生起 (有無)
7. アクセント型 (平板型、頭高型、起伏型)
8. 相対文節位置 (0.0 ~ 1.0)

日本語テキストから音声を合成する際に、特に感情を表現しない中立発話の韻律情報に対して、与えられた感情情報に従い韻律情報を制御することにより、感情を表現した音声を合成しようとするモデルを作成した。構築するモデルでは、声の高さ、長さ、強さといった韻律の 3 要素の制御を行う。

Evaluation of synthetic speech that applies prosody control rule based on a model for various emotional intensities of speech.

[†] Naoki MIYASAKA (Graduate School of Bionics, Computer and Media Sciences, Tokyo University of Technology)

^{††} Yuko MASAKURA and Sumio OHNO (School of Computer Science, Tokyo University of Technology)

声の高さでは、基本周波数パターン生成過程モデルの入力パラメータである基底周波数 F_b 、フレーズ指令の生起とその大きさ A_p 、アクセント指令の大きさ A_a について制御を行う。声の長さ、強さに関しては、音声の継続時間長 R_p 、発話強度 P_p について、その局所的な変化を捉えるために文節単位での制御を行った。制御を行う方法として、フレーズ指令の生起に関しては分類木を、その他の特徴量に関しては回帰木によりモデル化を行った。それぞれの韻律を制御するためのモデルへの入力パラメータに対する言語的要因の順位付けを表 2 にまとめた。

表 2: 韻律モデルパラメータに対する要因の順位付け

Joy							
	F_b	A_p 生起	A_p 文頭	A_p 文中	A_a	R_p	P_p
感情の強弱	1	5	1	1	2	1	3
前置品詞		4				2	2
後置品詞							4
モーラ数						5	
文節モーラ		1				3	
係り受け		2					
A_p 生起					1		
アクセント型		6			3		
相対文節位置		3				4	1

Sad							
	F_b	A_p 生起	A_p 文頭	A_p 文中	A_a	R_p	P_p
感情の強弱	1		1	2	1	1	2
前置品詞		6					3
後置品詞		3			3	2	4
モーラ数							
品詞モーラ		1		1			
係り受け		5			5		1
A_p 生起					2		
アクセント型		2					
相対文節位置		4			4		

3 合成音による評価

3.1 合成音声の作成

2 節で導出した規則を適用した合成音声を作成し、その合成音声の感情の程度を主観評価することにより規則の妥当性を確かめる。音声資料には、着目した特徴以外の変動要因を排除するため、フリーの波形合成エンジンである MBROLA[2] を用いて、 F_0 パターンと発話速度、発話強度を指定することにより合成音を作成した。以上、 F_0 パターンと時間構造及び発話強度をコントロールした 2 文節から 6 文節までの合成音声を感情及び程度ごとに計 4 セット用意した。感情を込めた音声として、それぞれの感情の程度に対応する合成音声が含まれる。

3.2 作成した合成音声に対する聴取実験

作成した合成音声資料を刺激音声として、各合成音声が表現している感情の程度を評価するための聴取実験を行った。被験者は大学 4 年生 6 名である。実験では、聴取した音声に、指定した感情がどの程度表れていると感じたかを「まったく感情が表れていない、他の感情に聞こえる」という場合の“ 0 ”と、「僅かに感情が表れている」という場合の“ 1 ”から「強く感情が表

れている」という場合の“ 5 ”までの 6 段階で評価し、その数字をマークさせた。

3.3 実験結果および考察

聴取実験の結果から、「喜び・悲しみ」の合成音声に対する主観評価値の分布を表したものを図 6 に示し、感情の程度に対する主観評価値の分布がそれぞれの程度間で有意な差を持っているかについて t 検定 (片側) を行った結果を表 3 に示す。

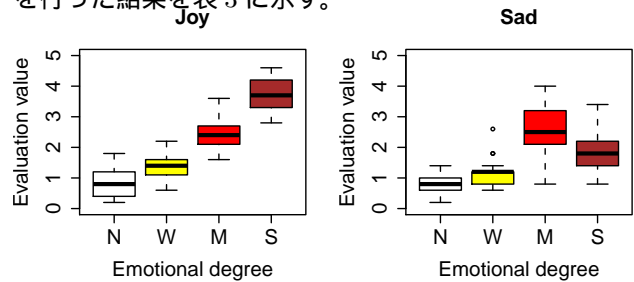


図 6: 合成音声に対する主観評価値の分布

表 3: 感情の程度に対する主観評価値の分布の有意差††

	喜び			悲しみ		
	弱	中	強	弱	中	強
中立	***	***	***	**	***	***
弱		***	***		***	***
中			***			—

喜びでは弱・中・強の 3 段階の感情の程度は、聴き手の観点でほぼ聞き分けられていることが分かった。一方、悲しみでは中立と比べ中では聞き分けられていることがわかったが、それ以外の程度ではうまく聞き分けられていないことから、程度区別の実現はうまくいかなかった。今回の聴取実験で、話者の程度の意図は伝わっているのにも関わらず、今回の韻律的特徴のパラメータでは、悲しみは程度の区別がうまくいかないという結果が出た。

4 おわりに

本報では、感情の種類として“ 喜び ”、“ 悲しみ ”の 2 つを取り上げ、それらの感情が表現された発話の韻律的特徴を分析した。これらの分析結果から感情制御規則を導出した。さらに、導出した感情制御規則を音声合成時に適用し、合成したそれぞれの音声が表示する感情の程度の主観を評価した。その結果、特に“ 喜び ”については作成したモデルの精度が良いことが分かった。今後は、本稿では取り扱わなかった韻律的特徴量が感情の程度に及ぼす影響を検討する予定である。

参考文献

- [1] 宮坂直樹, 大野澄雄, “ 音声における感情の強弱を表現する韻律制御規則 ”, 日本音響学会 2010 年春季研究発表会講演論文集, pp. 391-392(2010).
- [2] The MBROLA Project, <http://tcts.fpms.ac.be/synthesis/mbrola.html>

†† ***/**/*: 有意水準 1%/5%/10%で有意差あり