

## 母音に着目した事例ベース感情音声合成

朴冠宇 吉富康成 浅田太郎 田伏正佳

京都府立大学大学院生命環境科学研究科<sup>†</sup>

### 1. はじめに

近年の音声合成技術では、自然で肉声感の高い合成音声期待されるだけでなく、感情も伝える音声合成のニーズが高まっている[1]-[4]。感情音声を作成するためには、韻律情報を変更する必要がある。日本語など自然言語は、主に母音と子音で構成されており、母音の種類は少ない（日本語の場合、5種類）。かつ、母音は子音より発話時間が長いため、聞き手の印象に大きな影響を与える。本研究では、感情音声における「感情と母音の発声時間の関係」および「感情と母音の最大振幅の関係」の特徴を見出し、感情音声合成に反映させた。

### 2. 提案手法

まず、被験者に、「怒り」、「喜び」、「無感情」、「悲しみ」、「驚き」の各感情で発声してもらったときの音声データを WAV 形式ファイルとして得る。そして、各感情音声について、各母音の発声時間を測定し、あわせて、各母音発声時の最大振幅を測定する。

次に、被験者が発声した音素列の音声合成を行う。そして、以下の処理を順次行う。

- (1) 「無感情」音声での各母音および隣接する直前の子音の発声時間と一致するように、合成した音声の各母音および隣接する直前の子音の持続時間を変更（変更後の合成音声を、以下では、「無感情」合成音声」と表記）
- (2) 感情音声における各母音の発声時間の「無感情」音声との比率と一致するように、「無感情」合成音声の各母音および隣接する直前の子音の持続時間を変更（変更後の合成音声を、以下では、「発声時間変更合成音声」と表記）
- (3) 感情音声における各母音の最大振幅の「無感情」音声との比率を変換率として、発声時間変更合成音声における、各母音および隣接する直前の子音の振幅を変換（変換後の合成音声を、以下では、「感情合成音声」と表記）

Synthesized Speech with Emotion Using Vowel's Features in Cases

<sup>†</sup>Kanu Boku, Yasunari Yoshitomi, Taro Asada, and Masayoshi Tabuse

Graduate School of Life and Environmental Sciences  
Kyoto Prefectural University

ここで、母音に隣接する直前の子音がない場合（その母音の直前に隣接するのが母音の場合、または、その母音で発声が始まる場合）は、上記の(1), (2)における「母音および隣接する直前の子音」は、母音だけを意味する。

### 3. 実験

#### 3.1 条件

音声認識には、汎用大語彙連続音声認識エンジン Julius[5]を用い、音声合成には、汎用知的音声合成システム「ボイスソムリエ ネオ」プレミアムモデル[6]を用い、2章記載の(3)の処理には、波形編集ソフトを用いた。50代男性1名(被験者 A)に、「怒り」、「喜び」、「無感情」、「悲しみ」、「驚き」の各感情で中立的な言葉（日本名「taro」）を発声してもらったときの音声データを WAV 形式ファイルとして作成し、2章記載の方法で、男性 1(明るい声)[6]の条件で、感情合成音声を作成した。この場合、2章における、各母音は、/a/, /o/, 各母音および隣接する直前の子音は、/ta/, /ro/である。また、各感情カテゴリにおける各母音の発声時間および最大振幅には、6~20個のデータの平均値(表 1,2)を用いた。

そして、被験者 14名(50代男性 2名(被験者 A, B), 30代男性 1名(被験者 C), 20代男性 9名(被

表 1 感情音声の母音の発声時間

感情カテゴリ	平均発声時間(s)		「無感情」の値を基準にした平均発声時間相対値	
	/a/	/o/	/a/	/o/
怒り	0.035	0.073	0.565	0.241
喜び	0.087	0.314	1.403	1.036
無感情	0.062	0.303	1.000	1.000
悲しみ	0.127	0.273	2.048	0.901
驚き	0.070	0.157	1.129	0.518

表 2 感情音声の母音の最大振幅

感情カテゴリ	測定値(平均)		「無感情」の平均値を基準にした測定値(平均)の相対値	
	/a/	/o/	/a/	/o/
怒り	809.7	456.3	0.905	0.623
喜び	1726.7	1305.0	1.931	1.782
無感情	894.3	732.2	1.000	1.000
悲しみ	702.2	429.6	0.785	0.587
驚き	1520.0	886.0	1.700	1.210

験者 D~L), 20 代女性 2 名(被験者 M,N))が, 被験者 A の, ①感情合成音声, ②感情音声, を, この順で試聴し, 「怒り」, 「喜び」, 「無感情」, 「悲しみ」, 「驚き」のいずれの感情が込められているかの主観評価を行った.

### 3. 2 結果と考察

感情音声には, 図 1 に例示したように感情毎の波形の違いが見られ, その違いが, 母音の発声時間(表 1)と母音の最大振幅(表 2)の違いに表れている. 一方, 「無感情」の感情音声と「無感情」合成音声にも波形の違いが見られる(図 1). このため, 感情合成音声と感情音声には波形の違いが生じた(図 1). 表 3 に, 感情音声および感情合成音声の主観評価結果を示す. 感情音声の認識率は 100%で, 感情合成音声の平均認識率は,

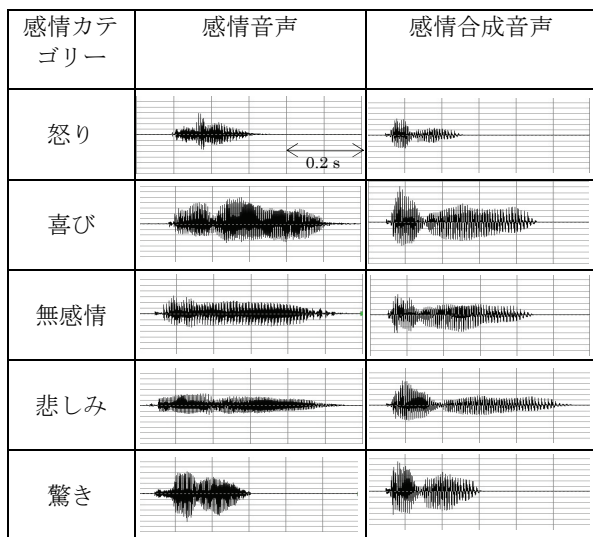


図 1 感情音声および感情合成音声の波形

表 3 主観評価結果(認識率)

#### (1)感情合成音声

		入力				
		怒り	喜び	無感情	悲しみ	驚き
認識	怒り	35.7	0.0	0.0	0.0	64.3
	喜び	0.0	42.9	28.6	21.4	0.0
	無感情	14.3	35.7	71.4	21.4	0.0
	悲しみ	0	21.4	0.0	57.1	0.0
	驚き	50.0	0.0	0.0	0.0	35.7

(%)

#### (2)感情音声

		入力				
		怒り	喜び	無感情	悲しみ	驚き
認識	怒り	100.0	0.0	0.0	0.0	0.0
	喜び	0.0	100.0	0.0	0.0	0.0
	無感情	0.0	0.0	100.0	0.0	0.0
	悲しみ	0.0	0.0	0.0	100.0	0.0
	驚き	0.0	0.0	0.0	0.0	100.0

(%)

48.6%であった. 感情合成音声の誤認識の特徴は, (1)「怒り」を「驚き」に誤認識(50.0%), (2)「驚き」を「怒り」に誤認識(64.3%), (3)「喜び」を「無感情」に誤認識(35.7%), (4)「無感情」を「喜び」に誤認識(28.6%), であった. (1)および(2)の主な原因は, 表 2 および図 1 からわかるように, 「驚き」より「怒り」の方が, 音声波形の振幅が小さいことによると考えられる. 「怒り」の感情音声を録音する際に, 被験者 A が下向きとなり, マイクから被験者 A の口が遠くなったために, 「怒り」の感情音声の振幅が小さくなったものと考えられる. (3)および(4)の主な原因は, 本法では, 「喜び」の感情音声の特徴を感情合成音声に反映できていないためと考えられる. 「無感情」以外の感情合成音声は, 「無感情」に平均 17.9%誤認識されていることから, 本法では感情音声の特徴を感情合成音声に十分反映できていないことがわかる. 今回, 事例ベースで感情合成音声を作成した. 今後は, 事例の感情音声における感情特徴を強調して感情合成音声を作成する予定である.

### 4. まとめ

感情音声における「感情と母音の発声時間の関係」および「感情と母音の最大振幅の関係」の特徴を見出し, 感情音声合成に反映させた. 「怒り」, 「喜び」, 「無感情」, 「悲しみ」, 「驚き」の 5 つの感情について, 感情合成音声を作成したところ, 被験者による平均認識率は, 48.6%であった. 今後は, 事例の感情音声における感情特徴を強調して感情合成音声を作成するなど, 感情合成音声の改善を行う予定である.

### 参考文献

- [1]片江伸之, 木村晋太, “感情音声合成における声質と韻律の制御の効果”, 日本音響学会研究発表会講演論文集 2000(2), pp.187-188, 2000.
- [2]緒方信, 四倉達夫, 森島繁生, “韻律情報の制御による感情音声合成のための声質変換”, 電子情報通信学会技術研究報告, HIP, ヒューマン情報処理, 99(582), pp.53-58, 2000.
- [3]飯田朱美, 伊賀聡一郎, 樋口文人, CAMPBELL Nick, 安村通晃, “対話支援のための感情音声合成システムの試作と評価”, ヒューマンインタフェース学会論文誌, 2(2), pp.63-70, 2000.
- [4]森山 剛, 森 真也, 小沢 慎治, “韻律の部分空間を用いた感情音声合成”, 情報処理学会論文誌, 50(3), pp.1181-1191, 2009.
- [5]<http://julius.sourceforge.jp/>
- [6]<http://hitachibusiness.com/products/package/sound/voice/index.html>