

# 音楽の知見に基づく音声合成を用いた感情表現手法の提案

齋藤将吾<sup>†</sup> 樽松理樹<sup>†</sup> 羽倉淳<sup>†</sup> 藤田ハミド<sup>†</sup>

岩手県立大学ソフトウェア情報学部<sup>†</sup>

## 1. はじめに

近年、音声合成技術の利用範囲は文章の読み上げを行う従来の用途から、人間のように豊かな感情表現を行わせる用途へと拡大している。それに伴い、感情を伝える音声合成の研究<sup>[1]</sup>も試みられるようになってきた。多くの研究では、人の感情と音声特徴の対応付けを何らかの形で合成音声の音声特徴に反映し感情表現を試みるアプローチ<sup>[1]</sup>が用いられている。しかし、この方法では大量の音声データが必要となり、その収集や解析の負荷は大きい。また感情表出能力は依然として低く、さらなる改善が必要である。

このような問題に対し、我々は音楽知見を用いた合成音声による感情表出手法<sup>[2]</sup>に試みている。本方式では、感情表出する際の合成音声の音声特徴を、音楽と感情の関係に関する知見を用いて決定している。評価実験を通し、手法の有用性を示すことはできたが、細やかな感情表出までには至っていない。その理由として、処理している音声特徴の種類が少ないことなどが考えられる。

以上の背景から、本研究では前述の手法に改良を加え、より細やかな感情表現の実現を試みる。さらに、その有用性を評価する。

## 2. 提案手法

### 2.1 概要

本章では音楽の知見に基づく音声合成手法について説明する。本手法では、はじめに合成音声を作成し、その後、感情毎に設定されたパラメータを加えることで感情表出を目的とする合成音声を生成する。以後、入力、それぞれの処理について説明する。

### 2.2 入力

入力としては、表出を試みる感情を付与した文章を与える。ここで表出を試みる感情としては、Ekman らが提唱した怒り・喜び・悲しみ・驚き・恐れ・嫌悪の基本6感情<sup>[3]</sup>を用いる。

### 2.3 処理1：合成音声の生成

与えられた文章に対し、アクセント辞書を用いて単語ごとの音素の高低を決定し、音声を合

成する。この時、高い音、低い音は固定値を用いる。この音声を基本音声とよぶ。

### 2.4 処理2：パラメータの適用

文献[4]を参考に、基本音声に対し、①テンポ、②ボリューム、③音高、④音程、⑤音程差、⑥間（ま）の6つのパラメータを変更していく。以下、各パラメータについて説明する。なお各パラメータ値には変化を加えない“変化なし”も含まれるが、ここでは省略する。

①テンポ：音声全体に対し、10%増加する“速く”と10%減少する“遅く”を用意する。

また、発話の後半部のテンポを変更する。変化量については楽譜に着目し、5小節単位の音符の長さの平均を用いる。実装に用いる楽譜としては、表1に示す感情との対応付けが示された6曲<sup>[5]</sup>を用いる。なお表1に含まれない感情に対しては、後半部の変更を行わない。

②ボリューム：一定の周期でボリュームを変化させる“変化あり”を用意する。

③音高：音声全体の音高を半音上げる“高く”と逆に半音下げる“低く”を用意する。

また、各音素に対し、Charpentier<sup>[6]</sup>が提言した感情に対するコードスケールを設定し、発話に含まれる音素の高さが決定する。その結果、コードスケールに含まれない音素については、コードスケールのベースとなるコード音、最近傍の高さの音、より高い音の優先順位で音高を変更する。

④音程：音声に含まれる各文節に対し、音節数÷2（切捨て）だけの音節を1全音上げる“上昇”と1全音下げる“下降”を用意する。

⑤音程差：次の音節との音程差が1音以下なら1音より大きくする“広く”と次の音節との音程差が3全音以上なら3全音より小さくする“狭く”を用意する。

⑥間：文献[7]で示された感情と間の平均長の関係を再現するように、間を一定間隔で挿入する。間の長さについては、文献[7]で示された長さを用いる。また間を挿入する際、文節を分断する場合は、前後の文節の区切りへ位置を変更する。

これらのうち、後半部のテンポの変動および間の挿入が今回新たに加えたものである。これらの特徴は、音声での感情表出においては重要である<sup>[7]</sup>と考えられる。しかし、先行手法では

A proposal for expressing emotion in synthetic speech based on musical knowledge

<sup>†</sup>Shogo SAITO, Masaki KUREMATSU, Jun HAKURA, Hamido FUJITA · Iwate Prefectural University, Software and Information Science

実現していなかったため改良を加えた。

以上の考えに基づき、各感情に対して付与した主だったパラメータ設定を表 2 に示す。なおコードスケールが 1 感情に対し複数あるため、感情毎にパラメータ設定パターン数は異なっている。また間は、感情付音声の間の平均÷基本音声の間の平均を意味している。

パラメータの適用においては、互いの関連性を考慮し、全体のテンポ⇒後半部のテンポ⇒全体のボリューム⇒全体の音高⇒文節単位の音程⇒文節単位の音程差⇒音素単位の高さ⇒間の順番で変更を加える。

表 1: 利用する楽曲と感情

作曲家	楽曲	感情
ホルスト	惑星より火星	恐れ
ムスグルスキー	禿山の一夜	恐れ
ヴィヴァルディ	四季より春	喜び
アルフヴェン	真夏の夜の目覚め	喜び
アルビノーニ	アダジオ	悲しみ
バーバー	管楽のためのアダジオ	悲しみ

表 2: 意図する感情と各パラメータの組み合わせ

No	表出感情	テンポ	ボリューム	音高	コードスケール	音程	音程差	間
1	怒り	速く	変化なし	高く	F	上昇	狭く	2.00
2					E			
3	嫌悪	遅く	変化なし	変化なし	Fm	変化なし	狭く	1.43
4	恐れ	速く後半有	変化あり	高く	Bbm	上昇	狭く	1.43
5					Em			
6	喜び	速く後半有	変化なし	高く	D	上昇	広く	1.25
7					G			
8					A			
9	悲しみ	遅く後半有	変化なし	低く	Cm	下降	狭く	0.43
10					Bm			
11					Fm			
12	驚き	遅い	変化なし	高く	Bbm	上昇	広く	1.00
13					Gm			
14					G			

### 3. 評価・考察

本手法の有用性、すなわち従来手法に対する感情表出能力の向上を評価するために、次の実験を行う。

従来手法に加え、後半のテンポ変更のみ、間の挿入のみ、両方の変更を加えた方法で音声を合成する。これらの音声を実験参加者に聴取してもらい、最も強く感情を感じた音声を選択してもらおう。その結果、提案手法が従来手法より上回れば、新規手法によって感情表出能力が向上したと評価する。

実験に用いる音声の内容は、感情が推定に影響が少ない天気予報の記事を用いている。また、クリプトン社の歌声合成ソフト・巡音ルカ<sup>[8]</sup>を

用いて音声合成を作成した。

現在評価実験を実施しており、学会発表時にその結果および評価について報告する。

### 5. おわりに

本稿では、音楽の知見に基づいた合成音声による感情表出手法を提案した。今後は、評価実験を通し、新たに提案したパラメータの有用性、一般性の検証を行う。さらに、評価実験で得た結果を基にパラメータの追加、修正、重み付け等を行うことで、感情表出能力の向上を図る。

### 参考文献

- [1] M. Schröder: “Emotional Speech Synthesis : A Review”, Proc. Of Eurospeech 2001, Vol.1, pp. 561 - 564 (2001)
- [2] M. Kurematu 他: “A Framework of Emotional Speech Synthesize Using Musical Knowledge”, Proc. of The 10th Internatinal Conference on Intelligent Software Methodologies, Tools and Techniques , pp.305-312(2011)
- [3] A. Gabrielsson 他: “EMOTIONAL EXPRESSION IN MUSIC”, HANDBOOK OF AFFECTIVE SCIENCE, Oxford University Press, pp. 503-534(2009)
- [4] P. Ekman 他: “表情分析入門”, 誠信書房(1987)
- [5] 大串: “音楽と感情”, バイオメカニズム学会誌, Vol. 30, No. 1(2006)
- [6] H. Fujita 他: “Cognitive Modeling in Software and Relation to Human Emotional Reasoning”, Proc. of The 6th Internatinal Conference on Intelligent Software Methodologies, pp. 145-165(2007)
- [7] Angelica Braun 他: “Speaking Tempo in Emotional Speech - a Cross-Cultural Study Using Dubbed Speech”, Proc. Of International workshop on Paralinguistic Speech - between models and data, pp. 77-82(2007)
- [8] クリプトン社: “巡音ルカ”, <http://www.crypton.co.jp/mp/pages/prod/vocaloid/cv03.jsp>, (2012/01/12)