

感情表現を意図した音楽理論に基づく音声合成手法の提案

○千葉央輝† 樽松理樹† 羽倉淳† 藤田ハミド†

岩手県立大学ソフトウェア情報学部†

1. はじめに

音声は、大きく言語情報と非言語情報から構成されている。従来の音声合成は、言語情報を中心に行われてきたが、音声合成の利用範囲の拡大とともに非言語情報にも着目されるようになった。結果、感情表現を試みる音声合成手法の研究が活発化^[1]している。しかし、評価実験などにおける人の認識精度は、実験参加者や発話内容などの実験条件に依存する点が多い。よりの確かな合成音声による感情表出を行うためには更なる改善が必要である。

一方、音楽と感情の関係に関してもいくつかの知見^[2]が得られており、それらを利用したシステム開発^[3]も進められている。

以上のような背景から、本稿では、音楽と感情の関係を音声合成に利用することを考え、音楽理論に基づいた新たな音声合成手法を提案する。本手法は従来手法と異なり、理論を活用することから学習データを必要とする従来手法よりも開発効率が向上し、より実用的な手法となることが期待できる。

2. 音楽理論に基づく音声合成手法

本章では音楽理論に基づく音声合成手法について説明する。

2.1 入力データ

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^[4]を用いる。

2.2 パラメータ設定

EMOTIONAL EXPRESSION IN MUSIC^[2]を参考に、基本旋律に対し、①テンポ、②ボリューム、③音高、④音程、⑤音程差、⑥コード・スケールの6つのパラメータを順に変更していく。以下、各パラメータについて説明する。

①テンポ：読み上げ速度を10%増加する“速く”と10%減少する“遅く”を用意する。

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”と変化しない“変化なし”を用意する。

③音高：文全体の音高を半音上げる“高い”と逆に半音下げる“低い”および“変化なし”を用意する。

④音程：1文節において音節数÷2(切捨て)の数だけの音節を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“広げる”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“狭める”を用意する。

⑥コード・スケール：1642年にMarc-Antoine Charpentierが”Rules of Composition”として提案したコード・スケールと感情との関係^[6]を基に設定する。コード・スケール^[7]を設定することにより、1オクターブ中で利用する音が確定する。⑤までで変更した基本旋律の音のうち、設定したコード・スケールに含まれない音については、最近傍の高さの音に変更する。最近傍の高さの音が複数あった場合、高い音を選択する。

以上の考えに基づき、各感情に対して付与したパラメータ設定を表1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。

表1: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 | 音程 | 音程差 | コード・スケール |
|----|------|-----|-------|------|------|-----|----------|
| 1 | 怒り | 速く | 変化なし | 高い | 上昇 | 狭める | F |
| 2 | | | | | | | E |
| 3 | 嫌悪 | 遅く | 変化なし | 変化なし | 変化なし | 狭める | Fm |
| 4 | 恐れ | 速く | 変化あり | 高い | 上昇 | 狭める | B♭m |
| 5 | | | | | | | Em |
| 6 | 喜び | 速く | 変化なし | 高い | 上昇 | 広げる | D |
| 7 | | | | | | | G |
| 8 | | | | | | | A |
| 9 | 悲しみ | 遅く | 変化なし | 低い | 下降 | 狭める | Cm |
| 10 | | | | | | | Bm |
| 11 | 驚き | 遅い | 変化なし | 高い | 上昇 | 広げる | Fm |
| 12 | | | | | | | B♭m |
| 13 | | | | | | | Gm |
| 14 | | | | | | | G |

3. 評価実験

3.1 実験概要

本手法の妥当性を評価するために次に示す聴取実験を行った。

実験において、実験参加者は、基本旋律の合成音声とパラメータ設定を加えた合成音声を聴き、各感情に対し、「弱くなった」「少し弱く

Research on Emotional Speech Synthesis Using Musical Theories
†Hiroki CHIBA · Iwate Prefectural University, Software and Information Science

なった」「同じ」「少し強くなった」「強くなった」の5段階で回答する。これらの回答結果を比較することで表出される本手法の有用性、感情に対するパラメータ設定の妥当性を評価する。今回の実験では、10代から40代の男15名、女26名、合計41名が参加した。また合成音声としては、内容から感情が推定しにくいニュース記事をクリプトン社の歌声合成ソフト・巡音ルカ^[8]を用いて作成したものを利用した。

3.2 実験結果

実験結果の概要を表2に示す。ここでNoは表1のパターンのNo、表出感情は、表出を意図した感情を示す。各感情の列の値は、各パターンに対し、「弱くなった」から「強くなった」までに、-2から2の値を割り当て、全回答の合計を求めたものである。値が大きいほどその感情を強く感じた人が多かったと考えられる。また▲が最大値、▼が最小値、黒抜きが表出を意図した感情に該当する部分である。

表2: 実験結果の概要

| No | 表出感情 | 怒り | 嫌悪 | 恐れ | 喜び | 悲しみ | 驚き |
|----|------|-----|------|-----|------|------|-----|
| 1 | 怒り | -12 | -15 | -14 | 37 | ▼-18 | ▲38 |
| 2 | 怒り | 6 | -9 | -6 | 23 | ▼-15 | ▲40 |
| 3 | 嫌悪 | ▼-8 | -4 | -3 | -1 | ▲2 | 1 |
| 4 | 恐れ | 10 | 4 | 9 | 12 | ▼-5 | ▲38 |
| 5 | 恐れ | 6 | -4 | 7 | 22 | ▼-12 | ▲42 |
| 6 | 喜び | -7 | ▼-25 | -13 | ▲56 | -19 | 44 |
| 7 | 喜び | -1 | ▼-17 | -4 | 44 | ▼-17 | ▲47 |
| 8 | 喜び | 1 | -18 | -22 | 33 | ▼-25 | ▲42 |
| 9 | 悲しみ | 18 | 41 | 30 | ▼-40 | ▲42 | -37 |
| 10 | 悲しみ | 14 | 20 | 19 | -30 | ▲39 | -23 |
| 11 | 悲しみ | 10 | 38 | 30 | ▼-46 | ▲55 | -36 |
| 12 | 驚き | -4 | -20 | -8 | 35 | ▼-21 | ▲39 |
| 13 | 驚き | -10 | -15 | -13 | 48 | ▼-20 | ▲49 |
| 14 | 驚き | -1 | ▼-17 | -4 | 44 | ▼-17 | ▲47 |

3.3 評価・考察

意図した感情と回答の関係を見ると「悲しみ」「驚き」については最高得点であり、「喜び」も比較的高い点を出している。一方、その他の感情では意図した感情の得点は低くなっている。このことから、「悲しみ」「驚き」「喜び」についてはパラメータ設定が適切といえるが、他の感情では不適切であると考えられる。

変化の認知（「同じ」以外の回答数が多い）に対しては、「悲しみ」が最も多く「嫌悪」が低くなっている。このことから「嫌悪」の表現が難しいことが予想される。ただし、今回の発話内容に影響を受けている可能性もある。

感情ごとの相関を見た場合、「喜び」と「驚き」は強い正の相関（相関係数 0.96）がある。同様に、「怒り」「悲しみ」「恐れ」「嫌悪」間も相関係数が 0.7 以上と正の相関が強い。一方、これらの組の間では、8組中7組が相関係

数が-0.7以下であり、強い負の相関となった。これらのことから、「喜び」「驚き」とそれ以外の感情とは傾向が逆転していると予想できる。

以上の事から、音楽理論に基づくパラメータ設定は、音声合成には有効な場合もあるが、すべての場合には適応できないことがわかった。この点を考慮し、意図する感情への回答数が最大であったものによるパラメータ設定を表3に示す。ここで黒抜きになっている部分に変更された点である。

表3: 実験結果に基づく感情と対応する各パラメータ

| 表出感情 | テンポ | ボリューム | 音高 | 音程 | 音程差 | コードスケール |
|------|-----|-------|----|------|-----|------------|
| 怒り | 遅く | 変化あり | 低い | 下降 | 狭める | Fm, Em |
| 嫌悪 | 遅く | 変化なし | 低い | 下降 | 狭める | Fm |
| 恐れ | 遅く | 変化あり | 低い | 下降 | 狭める | B♭m, Em |
| 喜び | 速く | 変化あり | 高い | 上昇 | 広げる | D, G, A |
| 悲しみ | 遅く | 変化なし | 高い | 変化なし | 狭める | Cm, Bm, Fm |
| 驚き | 速く | 変化あり | 高い | 上昇 | 広げる | G |

4. おわりに

本稿では、合成音声による感情表出を行うために音楽理論に基づいた新たな手法を提案した。評価実験を通し、有用である可能性を示した。今後の課題としては、表3で示したパラメータ設定の有用性の検証、パラメータの修正と追加を行うことである。また手法を幅広く利用するために本手法の自動化を進める必要がある。

謝辞

本研究の一部は、科学研究補助費・基盤研究B（課題番号：20300078）の助成を受けています。また本研究の評価実験に参加していただきました方に感謝の意を示します。

参考文献

- [1] M. Schröder: "Emotional Speech Synthesis: A Review", Proc. Eurospeech 2001, Vol. 1, pp. 561-564, 2001.
- [2] A. Gabrielsson 他: "EMOTIONAL EXPRESSION IN MUSIC", HANDBOOK OF AFFECTIVE SCIENCE, Oxford University Press, pp. 503-534, 2009
- [3] 三浦: "「怒り、喜び、悲しみ…」MIDI 楽曲に感情表現を付加するシステム「MOR2ART」", DTM MAGAZINE, Vol. 178, pp. 98-99, 2009.
- [4] P. Ekman, W. V. Friesen 著: "表情分析入門", 誠信書房, 1987.
- [5] 松崎, 河野: "よくわかる音声", アルク, 1998.
- [6] H. Fujita 他: "Cognitive Modeling in Software and Relation to Human Emotional Reasoning", New Trends in Software Methodologies, tools and Techniques, pp. 145-165, 2007
- [7] 北川: "ポピュラー音楽理論", リットーミュージック, 2004.
- [8] クリプトン社: "巡音ルカ", <http://www.crypton.co.jp/mp/pages/prod/vocaloid/cv03.jsp>, (2011/01/07 アクセス)