

平成 28 年度 卒業研究

機能語に注目した
音声合成朗読システムのための感情
推定

東京理科大学 理工学部 経営工学科
西山研究室 7413069 恒川 泰輝

指導教員 西山 裕之

学士論文概要

目次

学士論文概要	i
第1章 序論	1
1.1 背景	1
1.1.1 音声合成	1
1.1.2 オーディオブック	1
1.2 本論文の目的	2
第2章 関連研究	3
第3章 提案手法	4
3.1 提案手法の概要	4
第4章 実装	6
4.1 音声合成ソフト	6
4.2 書籍データ	6
4.3 開発環境	6
4.4 まとめ	6
第5章 評価	7
5.1 評価目的	7
第6章 結論	9
6.1 まとめ	9
6.2 今後の展望	9

図 目 次

3.1 正解データの作成	4
------------------------	---

表 目 次

第1章 序論

序論ではまず，本論文の背景としてオーディオブックと音声合成について解説する．

1.1 背景

1.1.1 音声合成

音声合成とは，人間の音声を人工的に作り出すことである．音声合成技術は文字を読むことが困難な障害者，外国人や幼児などに画面読み上げソフトとして長く利用されてきており，言葉を発することが困難な人が代替手段として利用することも多い．さらに，21世紀に入ってからでは家電製品の音声ガイダンスや公共交通機関のアナウンス，ロボットの発話用途などとして広く使用されるようになっている．

1.1.2 オーディオブック

オーディオブックとは主に書籍を朗読したものを録音した音声コンテンツのことである．アメリカを中心に市場規模が拡大している．もともと車社会のアメリカなどの国では早期から市場が確立していたが，近年インターネットを介して気軽にダウンロードして楽しめる環境が整ったことなどによりアメリカとカナダの市場規模が2015年には前年比21%拡大している．また日本においても定額配信サービスが開始されており，今後さらに普及する可能性がある．しかしながら，このようなオーディオブックは書籍から音声化する際に手間やコストが電子書籍にくらべて10倍ほどかかっており2～3ヶ月ほどかかると言われている．

1.2 本論文の目的

音声合成技術を用いて人手で行っている朗読作業を根源的な目的である．これまでの音声合成研究の結果，単に情報を伝達する目的では十分な音質が確立されている．しかし，従来の音声合成は一文やフレーズの読み上げでは高品質な音声を実現している一方で書籍データのような長い文章では平板で淡々とした読み上げになってしまい，感情的あるいは情緒的な表現を多く含む小説などの朗読を聞くには不十分である．近年になって，喜怒哀楽といった感情の種類をパラメタとして与え表現豊かな音声を合成できるソフトが販売されている．しかし，これらのパラメタは文または単語ごとに人手で設定する必要がある．短い文章など限られた場合は容易であるが，小説といった膨大な文章に対して都度人手でパラメタ調整を行うのは大変手間がかかる．そこで，本研究では文章から読み上がる感情として最適なものを推測することを目的とする．

第2章 関連研究

本章では音声合成や朗読システム感情推定に関する既存の研究について紹介し、それに対しての本研究の位置づけと意義を述べる。

音声合成による朗読システムとして吉田ら [6] は、文間ポーズ長の頻度分布調査実験の結果を基に文内・文末表層情報に着目している。文内情報 命令，否定，意志等と文末情報「～ある」，「～いる」，「～んだ」等をカテゴリーに分け，文間ポーズと基本周波数，話速を調整している。

文章の意味内容に着目した研究では，布目ら [8] はポーズ情報の推定と感情表現の推定を行っている。ポーズ推定では，タイトルや章立て構造といった文章の論理要素に応じてポーズ長を推定し，ポーズを挿入する。感情推定では「喜び」「怒り」「悲しみ」及び「平静」の各感情の推定モデルを作成しスコアを算出し最もスコアの高い感情を文にラベリングする。そのラベリングに応じて，文ごとに韻律辞書や音声制御用パラメタを切り替えて読み上げる。この研究では，意味内容に着目して感情の各スコアを算出しているが最終的にラベリングを行ってしまっているので，各感情の中間の表現ができない。また，各単語に分けてラベルとの類似度を算出してしまっているため，例えば文末の「でない」といった否定によって文全体の文意が反転する場合に対応できていない可能性がある。

一方で，自然言語の感情分析の分野では，Hong ら [5] は単語の繋がりを考慮して文全体のポジティブ度ネガポジ度を判定している。また，IBM の Tone Analyzer では各文の感情 *Anger, Disgust, Fear, Joy, Sadness* などとその度合を推定することができる。

第3章 提案手法

本章では提案する手法の詳細について説明する。

3.1 提案手法の概要

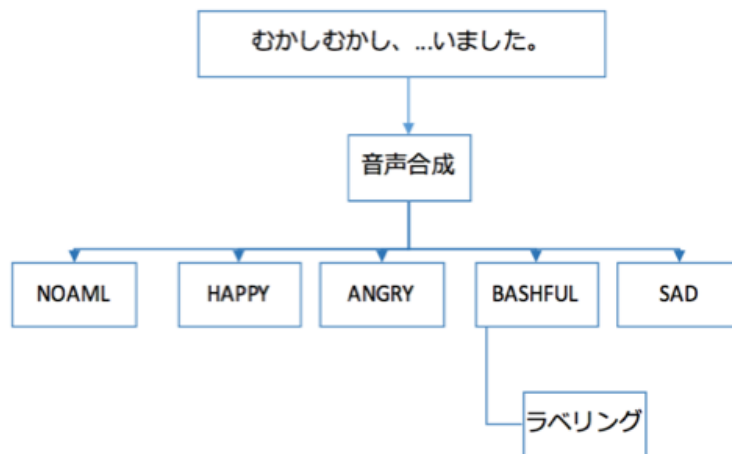


図 3.1: 正解データの作成

本手法では予め人手で各文に対して感情ラベルを与えた教師データを作成し、そのデータを機械学習することで分類機を作成する。感情ラベルは、普通、嬉しい、恥ずかしい、怒ってる、悲しいの5種類とする。未知の入力文が与えられた場合に、この中の一つを自動的に割り当てることが目的である。基本的に句

点ごとに文章を分割してすべての文がそれぞれのどの感情として音声合成されるべきかを推定する．なお，分割する際は基本的に句読点で分割し，カギカッコの始まりは句読点とみなしてそこから次の文の始まりとした．また，カギカッコ内の句読点も分割の対象とした．ここで文章に依存する内容語による推定では膨大な教師データが必要になると考えられる．そこで内容語によらず文の形式である程度，感情を推定することができると仮定して，文の機能語のみを抽出して学習を行う．具体的には名詞，動詞，形容動詞，形容詞を取り除く．その後，各文は品詞分解し基本形変換を行い bags of words に変換し正解ラベルを元に学習を行う．分類機にはナイーブベイズ，ランダムフォレスト等を用いる予定である．

第4章 実装

本章では，提案するシステムの具体的な実装について述べる．まず使用した機材や技術情報などについて明確にし，次に提案するシステムの流れを全体像とともに説明する．

4.1 音声合成ソフト

本研究では音声合成ソフトウェアとしてオープンソフトの Open JTalk[7] を用いる．Open JTalk は形態素解析部に MeCab[2]，発音辞書に NAIST Japanese Dictionary[3]，波形生成部に hts - engine API[1] を組み込んでいる．感情パラメタとして，NOMAL, HAPPY, ANGRY, BASHFUL, SAD の5種類が用意されている．

4.2 書籍データ

青空文庫 [4] とは著作権が消滅した作品や著者が許諾した作品のテキストを公開しているインターネット上の電子図書館である．この中から年代に依存しない現代的な話し言葉が用いられている童謡を中心に 30 作品 予定 を書籍データとして用いる．なお，ルビのデータが含まれているため予めタグを除いておく．

4.3 開発環境

4.4 まとめ

本章では，提案するシステムを実装するにあたり利用したハードウェア及びソフトウェア情報について述べ，具体的な手法の実装部分の詳細について述べた．

第5章 評価

本章では提案手法を具体的に実装したシステムを用いて，実験の概要と結果について述べる．

5.1 評価目的

第6章 結論

本章では，本研究のまとめ，および今後の展望について述べる．

6.1 まとめ

本研究では

6.2 今後の展望

今後の展望として，

参考文献

- [1] Hmm-base speech synthesis engine: hts engine api. <http://hts-engine.sourceforge.net/>.
- [2] Mecab: Yet another part-of-speech and morphological analyzer. <http://taku910.github.io/mecab/>.
- [3] Naist japanese dictionary. <http://naist-jdic.osdn.jp/>.
- [4] 青空文庫. <http://www.aozora.gr.jp/>.
- [5] James Hong and Michael Fang. Sentiment analysis with deeply learned distributed representations of variable length texts.
- [6] 吉田有里, 奥平康弘, and 田村直良. 音声合成による朗読システムに関する研究. 情報科学技術フォーラム講演論文集, 8(2):337–380, aug 2009.
- [7] 大浦 圭一郎, 酒向 慎司, and 徳田 恵一. 日本語テキスト音声合成システム open jtalk. 日本音響学会春季講論集, 1(2-7-6):343–344, 2010.
- [8] 布目光生, 鈴木優, and 森田眞弘. 自然で聞きやすい電子書籍読上げのための文書構造解析技術. 東芝レビュー, 66(9):32–35, 2011.

付 録 A Web ページの URL リスト