

令和 4 年 6 月 22 日現在

機関番号：32706

研究種目：基盤研究(C)（一般）

研究期間：2018～2021

課題番号：18K11381

研究課題名（和文）読み聞かせ戦略を反映できるような表現力の高い音声合成手法の研究

研究課題名（英文）Study on Expressive Speech Synthesis Methods Reflecting Reading Strategies

研究代表者

斉藤 隆（Saito, Takashi）

湘南工科大学・工学部・教授

研究者番号：90633636

交付決定額（研究期間全体）：（直接経費） 2,600,000 円

研究成果の概要（和文）：音声合成の物語読み聞かせへの応用を目指し、実際の発話に基づく基礎的な検討を進めた。日本の昔話8話を6名が読み聞かせスタイルで朗読した音声から、読み聞かせ音声のデータベースを構築した。フリーの音声分析ツール上で、基本的な音声区分から、アクセント句、イベント層、人物層等の物語構造まで、読み聞かせ音声に関する階層的な情報を可視化できるようにした。読み聞かせスタイルの全般的な韻律特徴を、話者の共通性・固有性も考慮しつつ、ニュートラルなスタイルと対比することで明らかにした。さらに、物語の展開における様々な状況変化をイベントと定義し、各イベントでの話者間共通にみられる韻律的なパターンの存在を確認した。

研究成果の学術的意義や社会的意義

本研究で取り上げたような、発話戦略の分析を目的とした音声データベースの構築は、先行研究は少ないものの、音声合成の高度化には意義深いものと考えられる。研究目的に即したデータベース構築のメリットを活かして、ここで取り上げた音声イベントのような、発話戦略に関する様々な知見の獲得が期待できる。また得られた知見をもとにした音声合成が実現すれば、応用面での期待も大きい。例えば、コミュニケーションロボットと組み合わせた、子供向けの教育・娯楽用の対話システムとしての活用である。こういった環境での音声合成はとりわけ、親しみやすさや表現力が重要であり、本研究の成果はそういった場面での音声合成への応用につながる。

研究成果の概要（英文）：Aiming to apply speech synthesis to story reading, a basic study based on actual storytelling speech was conducted. A prosodic database of storytelling speech was constructed from the voices of six people reading eight Japanese fairytales in a storytelling style. The system allows users to visualize hierarchical information about the read-aloud voice on a free speech analysis tool, from the basic speech segments of phonemes to the layered structures of the story, including accented phrases, event layers, character layers, and so on. The general prosodic features of the storytelling style were clarified by contrasting it with the neutral style, while also considering commonalities and individualities among speakers. Furthermore, we defined events as the various changes in the stream of the story plot, and confirmed the existence of common prosodic patterns among speakers at each event.

研究分野：音声言語情報処理

キーワード：音声合成 読み聞かせ 韻律データベース

1. 研究開始当初の背景

現在の音声合成システムでは、基本的に個々の文単位の情報で文ごとに独立した合成処理を行うことを前提としている。ニュースや天気予報など情報の伝達を主目的とする用途では、合成音声に多少違和感があったとしても情報の内容は伝わるため文単位の合成で支障をきたすことにはならない。ただ、長いまとまった文章の合成音声を聞いていると、単調で物足りないと感じたり、あるいは、受聴に対する疲労感を受けてしまったりすることは多々ある。これは受け取る人間側が、情報を単文ごとではなく、文の流れ、関係性として受け取ろうとしていることに起因する。たとえば、物語の読み聞かせのように、文の流れにさらに情感などが加わってくると現在の音声合成ではまったく太刀打ちできない。このように、従来の単文の音声合成では対応しきれない領域が明らかに存在している。しかも、コンピュータとの音声によるコミュニケーションが今後、教育、娯楽、福祉など様々な分野に広がるにつれ、その領域が拡大していくことは必然と考えられるものの、文レベルを超えて制御が行える音声合成技術に関する研究への取り組みはまだ少ないのが現状である。

2. 研究の目的

上述のような背景から、本研究では、音声合成で伝達する対象を一つの文ではなく、一つの話題としてとらえ、話の流れや物語の展開を制御できるような音声合成の枠組みの導入を検討する。それによってより表現力の高い音声合成技術の確立を目指していく。本研究においては、特に、物語の「読み聞かせ」への応用を具体的な目標として設定する。人間のナレーターが行う実際の読み聞かせにおいては、様々な技法を駆使して物語の情感を効果的に伝え、聞き手を話の中に引き込んでいく。本研究ではそのようなプロのナレーターの読み聞かせの技法に着目し、彼らの「読み聞かせ戦略」の詳細な分析と体系化を進める。そのために、読み聞かせ戦略の分析を主目的とした音声データベースを設計・構築する。本研究のテーマとする「読み聞かせ戦略」のような話者の高次の発話戦略の分析を目的とした独自のデータベースの構築は、先行研究ではあまり実施されていない。一方で、オーディオブック等の既存のデジタルコンテンツから表現のバリエーションを活用する音声合成の研究が増えているが、本研究の目的に合致したものを見つけないのは困難であった。今回、研究の目的に即したデータベース構築のメリットを活かすことで、その構築・分析の過程から、発話戦略に関する様々な観点からの新たな知見が得られることも期待できる。

3. 研究の方法

プロのナレーターの読み聞かせ戦略の詳細な分析を行うために、読み聞かせ音声を収集し、本研究の目的に即したデータベースを構築する。今回の研究では読み聞かせの物語素材を日本の昔話から選定することとした。日本の昔話の多くは起承転結が明快で感情表現もわかりやすく、読み聞かせ戦略に関する検討が比較的行いやすいと考えられる。今回のデータベース構築では、読み聞かせ戦略を表現していると考えられる音声上の「イベント」、たとえば、登場人物の入れ替わり、感情や場面の移り変わりなどの抽出、それをさらに進めて体系化も考えた設計を検討していく。

(1) 読み聞かせ音声の収録

ナレーターが同一の物語を、読み聞かせのスタイルに加え、ニュートラルのスタイル(ニュースを読むような通常の読み上げ方を指示)の2つのスタイルで、朗読した音声を収録する。物語の題材としては著作権のない日本の昔話を利用する。データ数としては10種類程度用意し、5~10名程度のナレーターで収録を行うことを想定した。また、音声収録を完了したナレーターに対して、それぞれの物語の解釈と読み聞かせの具体的な戦略、また一般的な読み聞かせ技法、読み聞かせのための発話訓練等についてもヒアリングを実施しておく。

(2) 音声データベース構築用ツールの開発

申請者はこれまでも音声合成のためのデータベース構築の実績がある。その経験をもとに、今回はオープンソースやフリーのツールを活用して新たなデータベース構築環境を作成する。データベース情報の編集ツールとしては、フリーで提供されている wavesurfer (<http://www.speech.kth.se/wavesurfer/>)を用いる。ピッチ分析は wavesurfer の機能を利用し、ピッチエラーを除去するプログラムは別途用意する。また、音素、単語、アクセント句の分割は別のツールを活用して行い、その出力を wavesurfer のラベルとして表示、および、視察修正ができるようにする。

(3) 読み聞かせ音声の韻律特徴、イベントの分析

まず、読み聞かせ音声の全般的な韻律特性を、韻律の基本属性である F0 値、速度、パワー、ポーズ等について調査し、ニュートラル発声との違いから読み聞かせスタイルの傾向性をみていく。この際、話者間の違い、共通性についても注意しておく。

さらに、ナレーターへのヒアリングの結果も参考にして、音声合成の観点から読み聞かせ制御のためのイベントカテゴリーを体系化する。たとえば、そのカテゴリーとしては、キャラクター

ベント(語り手、主人公、おじいさん、...)、感情イベント(願い、安堵、恐怖、...)、場面イベント(導入部、盛り上がり、エンディング、...)などが考えられる。これらのイベントに対する音声データ上のタグ付けを wavesurfer 上で視察により実施する。2つのスタイル(読み聞かせとニュートラル)のデータベースを参照し、読み聞かせイベントでの韻律パラメータ(F0、時間長、パワー、ポーズ)についての統計的な性質の分析を行い、モデル化のための基礎データとする。

4. 研究成果

(1) 読み聞かせ音声の収録とデータベースの構築

日本の昔話8話(一寸法師、浦島太郎、この正直者が、カチカチ山、花咲か爺さん、桃太郎、お結びころりん、鶴の恩返し、笠地蔵)について、上述のとおり、読み聞かせとニュートラルの2つのスタイルで収録した。話者としては、まだプロではないが声優のトレーニングを受けている専門学校生の中から、事前収録を行って6名(女性4名、男性2名)を選定した。

収録した音声について、音素、アクセント句、呼吸段落等についてのラベル付けを実施した。なお、その作業過程で必要となるポーズ検出と音素アラインメントについては、自動処理を行った後に視察修正する方法をとった。ラベリングの結果は wavesurfer で、表示・編集できるようにした。Fig. 1 にその表示例を示す。図の下部から、F0、スペクトル、波形、音素区分、アクセント句(Phrase layer)、イベント層(Event layer)、人物層(Impersonation layer)を表示している。図からは小さくて読み取ることが難しいが、この表示区間で、声優は2つの声(語り手:「鬼はびっくりです」、鬼:「たっ、たっ、助けてくれー」)を切り替えている。このように、読み聞かせ音声の分析データを、物語の展開に関する上位層情報まで含めた形で可視化できるようにした。

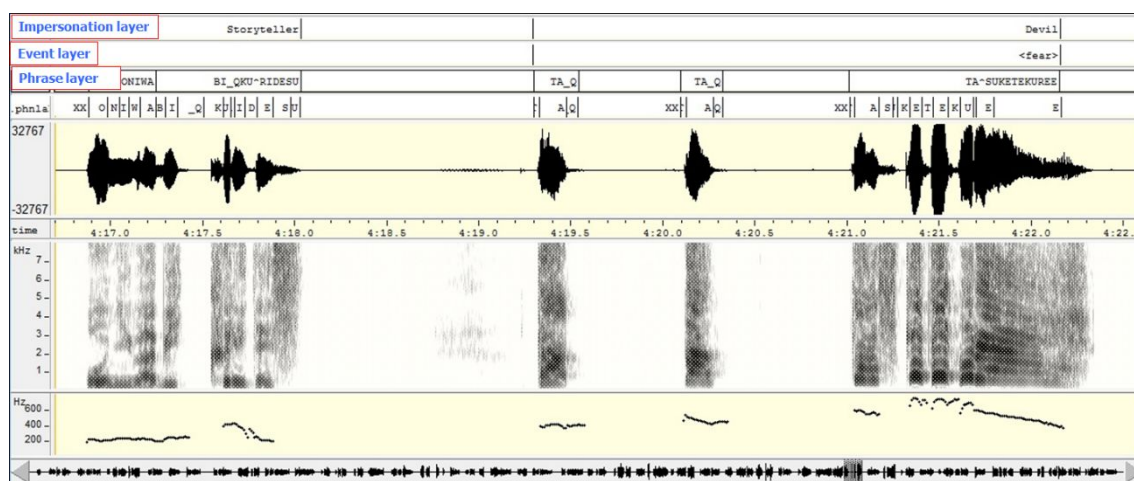


Fig. 1 Example of the storytelling speech labeling on Wavesurfer

(2) 読み聞かせ音声の韻律特徴

読み聞かせ音声での韻律特徴(F0、速度、パワー、ポーズ)についてニュートラルな発声と比較してまとめると、以下ようになった。

・アクセント句のF0 最大値

読み聞かせスタイルではニュートラルに比べ大幅に増加し、同時に振れ幅も拡大する。これは話者全員に共通する傾向であるが、その度合いについては個人差がみられる。

・呼吸段落単位の数

話者間でほぼ共通の傾向として、読み聞かせでは平均的な速度は低下し分散は拡大する。各話者の上限値は大きく変化せず、遅い方に広がる傾向にあることがわかった。

・パワー

2つのスタイル間での、少なくとも全体的な傾向としての顕著な違いはみられない。

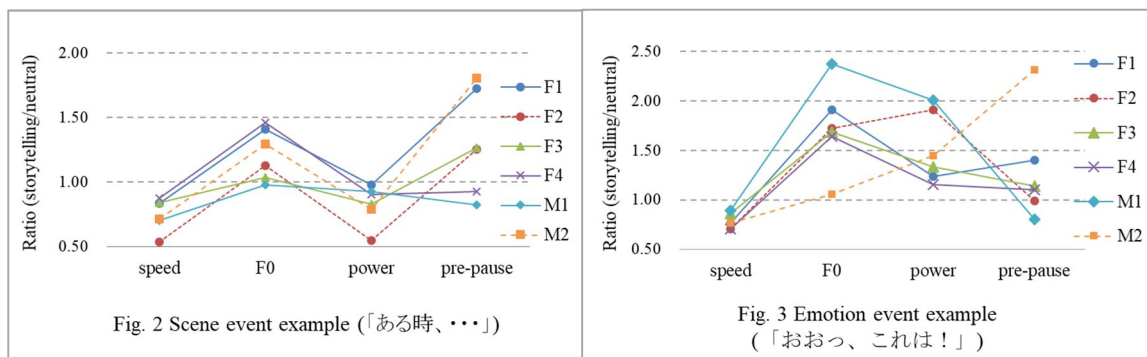
・呼吸段落間のポーズ

平均値としてはニュートラルと比べて大きな変化はなく話者間で共通の特徴も見られないが、振れ幅には増加の傾向がみられる。なお、ポーズ挿入の頻度も増加する傾向にある。

(3) 読み聞かせイベントの特徴

物語の展開における様々な状況変化「イベント」、ここでは、3種のイベント、場面の切り替わり「シーンイベント」、話者の感情の変化「感情イベント」、登場人物の交代「キャライベント」について、韻律特徴(速度、F0、パワー、前接ポーズ)を調べた。「シーンイベント」と「感情イベント」の具体的例を、Fig. 2、Fig. 3にそれぞれ示す。縦軸は、同じ話者の同じテキスト部分についての、ニュートラルに対する読み聞かせ音声の比率である。これは、読み聞かせのスタイルがニュートラルなスタイルとどう異なる傾向にあるかを図示したものである。両イベントとも、話者間でほぼ共通するパターンが確認できる。ただ、Fig. 3からわかるように、感情

イベントで 1 名だけ全く別のパターンを示している。そこで、この話者の発声を確認したところ、情感を表現する上手さはあるが不自然さは全く感じられない。このことは、表現技法として話者共通に表れる基本的なものの他に、当然、話者に固有のものも含まれることを示唆している。ただ、収録したナレーターは専門学校生で声優の技法を学んでいる段階にあることから、個人差が比較的少なかったとも考えられる。



(4) 読み聞かせトレーナーの試作

収録した 6 名の話者について韻律特徴の分析を進めていく中で、読み聞かせの技量の違いがデータに現れていることがわかってきた。つまり、発話法の習熟、あるいは、もともと本人に備わっている技量の違いを韻律特徴の差からつかめる可能性があるということである。この考え方を利用すれば、読み聞かせ初心者者の自己学習ツール、さらに進めて、声優専門学校での教材への応用も考えられる。これは当初の研究計画にはなかったものであるが、このアイデアを進めるために読み聞かせトレーナーという、ウェブベースの読み聞かせの練習用ツールを試作した。声優のお手本となる発声と自分の発声をシステムが比較評価してくれるというもので、抑揚、話速、間の取り方の違いを確認しつつ、繰り返して発声練習ができる。

< 引用文献 >

斉藤隆、小椋蓮、高橋宏和、守谷能宣、渡辺直貴、“物語読み聞かせ音声における韻律特徴の分析,” 日本音響学会 春季研究発表会講演論文集, 1-P-23, Mar. 2019.

5．主な発表論文等

〔雑誌論文〕 計0件

〔学会発表〕 計1件（うち招待講演 0件 / うち国際学会 0件）

1．発表者名 齊藤隆、小椋蓮、高橋宏和、守谷能宣、渡辺直貴
2．発表標題 物語読み聞かせ音声における韻律特徴の分析
3．学会等名 日本音響学会2019年春季研究発表会
4．発表年 2019年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6．研究組織

	氏名 （ローマ字氏名） （研究者番号）	所属研究機関・部局・職 （機関番号）	備考
--	---------------------------	-----------------------	----

7．科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8．本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------