様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

令和　５ 年　 ６ 月 ２０ 日現在

機関番号：６２６１５
研究種目：挑戦的研究（萌芽）
研究期間：2021～2022
課題番号：２１Ｋ１９８０８

研究課題名（和文）AI噺家は人を楽しませる事ができるのか？－落語音声合成の表現力向上と噺の自動生成

研究課題名（英文）Can AI Rakugoka entertain people? -Improved expressiveness of rakugo speech synthesis and automatic generation of storytelling

研究代表者
山岸　順一（Yamagishi, Junichi）

国立情報学研究所・コンテンツ科学研究系・教授

研究者番号：７０７０９３５２

交付決定額（研究期間全体）：（直接経費）　　4,900,000円

研究成果の概要（和文）：我々は落語の実演データからニューラルネットワークを学習し、プロの落語家風に噺をし聞き手を楽しませる事が可能なAI噺家の実現を目指し研究を行ない、以下の業績を挙げた。まず我々の落語DB上でTacotron、 Transformer、VITS、FastPitchという音声合成モデルを構築した。また落語で多用される笑い等の非言語情報の明示的モデリング法の開発にも取り組み、音声波形の概形を利用する新たな手法を提案した。 さらに落語の噺が毎回完全同一では聞き手を楽しませる事は不可能であることから、GPT-2、BART、T5といったニューラル言語モデルにより落語の噺を自動生成する枠組みについても検討した。

研究成果の学術的意義や社会的意義
伝統話芸である落語を深層学習で再現し、AI噺家を実現しようと言う、本研究の試み自体が、情報伝達や質問回答を目的とする従来の音声対話システムとは目的が全く異なり、ユニークでかつ学術的意義のある試みである。構築された音声合成システムの比較実験からは、AI噺家が人を楽しませるためには、従来の音声合成の自然性に関する評価指標のみでは解決できない事も判明し、音声合成のモデリングのみならず評価体系を抜本的に変化させる必要があることも判明した。また同時に、Tacotron、 Transformer、FastPitchという種々のEnd-to-end音声合成モデルの中でどれが落語音声に適しているかも判明した。

研究成果の概要（英文）：We have conducted machine learning research to construct a DNN-based rakugo performer’s speech synthesis model, which can generate natural-sounding audio that entertains listeners by performing rakugo like a professional performer. First, we constructed speech synthesis models called Tacotron, Transformer, VITS, and FastPitch on our rakugo database. We also developed an explicit modeling method for nonverbal information such as laughter, which is frequently used in rakugo, and proposed a new method that uses the approximate shape of speech waveforms as input units. Furthermore, since it is impossible to entertain listeners if rakugo stories are exactly the same every time, we also studied a framework for automatic generation of rakugo stories using neural language models such as GPT-2, BART, and T5.

研究分野：音声情報処理

キーワード：音声合成　落語　深層学習　言語生成

様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９（共通）

１．研究開始当初の背景

研究開始当時、入力文章から音声を生成する音声合成技術は、深層学習の発展により大きく進展し、新聞記事の読み上げ等ごく限られた条件下ではあるが、人間に非常に近い自然な音声を生成可能となっていた。その読み上げ音声は情報伝達を目的とした音声の発話様式のごく一例であり、実際の音声の発話様式は多種多様であり、話者の感情、個性、意図なども同時に聴取者に伝える事も可能である。当時の音声合成研究においては、この様な情報伝達以外の観点はこれまで全く重要視されてこなかった。

そこで我々は、落語音声を機械学習の対象として取り上げ、噺家の口調を緻密に真似させ、情報伝達以外の観点も聞き手に適切に送り届ける事で、機械が人を楽しませる事が可能になるのか？と言う科学的な問いを調査することにした。

まず、江戸落語の真打（最高位の格付け）の柳家三三師匠の協力と助言のもと、古典落語の収集、アノテーション、分析を行い、また深層学習による音声合成のモデル化も行った。次に、落語の演目を実際に合成させ、前座・二ツ目と言う階級の異なる噺家と比較することで、どの程度、話し上手か、聞き手を楽しませる事が可能か、そのスキルをベンチマークする枠組みも提案した。これらの緻密な分析の結果、音声は自然だが、聞き手が十分に楽しむことができていない事が判明し、その理由として、噺に登場する複数人物の区別や、笑い・動物の鳴き真似等の非言語情報など音声の表現のモデリングにまだ問題があり、更なる改善が必要であることが示された。さらに、噺家のスキルや聞き手の満足度は、合成音声の自然性よりも、内容理解度や登場人物の識別性と相関が高い事も判明した。これが本研究の提案に至った背景と経緯である。

２．研究の目的

本研究の目的は、1.に記載した問題点を改善し、落語音声合成システムの音響的表現力を向上させ、あたかもプロの噺家の様に、噺を読み上げる改良版システムを実現することを目標に、基礎となっている音響モデルを改良することである。また落語の噺が毎回完全同一では聞き手を楽しませる事は不可能であることから、演目名を指定すれば落語音声が都度異なる形で生成される生成方法の開発にも取り組む。

３．研究の方法

本研究では、上記研究目標の実現に向け、二つの研究課題に取り組んだ。それぞれの課題の具体的方法は以下の通りである。

<u>課題１：非言語情報の明示的モデル化による合成音声の表現力向上</u>

複数の落語家との比較実験から、現在の落語音声合成システムの音声は非常に自然だが、噺に登場する複数人物の区別や、笑い・咀嚼音・咳払い・動物の鳴き真似等の非言語情報などのモデリングに未だ問題があり、聞き手が十分に楽しむことができていない事が判明しているため、落語で多用される笑い・咀嚼音・咳払い等の非言語情報の明示的モデル化に取り組む事が課題１である。笑い声のみに特化した音声合成システムの先行研究はあるが、非言語情報一般を音声合成においてどう扱い、学習・予測させるかその枠組みを検討した論文は一切無い状況であり先駆的かつ挑戦的課題である。

<u>課題２：ニューラル言語モデルによる噺の自動生成</u>

落語の噺が毎回完全同一では聞き手を楽しませる事は不可能である。古典落語の噺は序破急と呼ばれる構造があるが、詳細は噺家毎に異なり実際の表現も毎回異なる。そこで流暢な文章や新聞記事を生成可能な GPT といったニューラル言語モデルを参考に、噺を演目名から自動生成する枠組みを検討する。具体的には非線形自己回帰型 Transformer モデルである GPT 等を落語書き起こしテキストを元に転移学習させることを検討する。また課題１の音声合成技術と統合し、演目名を指定すれば落語音声が都度異なる形で生成される事を目指す。

４．研究成果

合計７名の研究者が課題１と課題２に貢献し、２年間の研究期間に以下の成果を挙げた。

課題１：非言語情報の明示的モデル化による合成音声の表現力向上

課題１の達成に向け、令和３年度は、まず、落語音声合成システムの音響モデル単位を複数の呼気段落に変更し、また同時に、前後の呼気段落を連結することで学習データを擬似的に増やす学習法を試み、評価を行なった。

続いて令和４年度は、さらに Tacotron、Transformer、VITS、FastPitch という様々な音響モデルによる落語音声合成システムの構築と評価を行なった。

また同時に、音声合成システムにおいて複数の非言語情報を学習し予測させる研究も行い成果を挙げた。具体的には、下図の右側の例の様に、複数の非言語情報をある固定長のセグメント毎にアノテーションし、笑い声は１、泣き声は２という様にモデル化したい非言語情報クラス毎にユニークな番号を割り当てるという比較的シンプルなラベリングスキームで非言語情報をモデル化・制御可能である事を示した。非言語情報の音声を生成するモデルは、エンコーダーデコーダモデルとニューラルボコーダを組み合わせたモデルを利用した。本成果は音声情報処理のトップカンファレンスである Interspeech2023 に採択され、2023 年８月に発表を行う予定である。

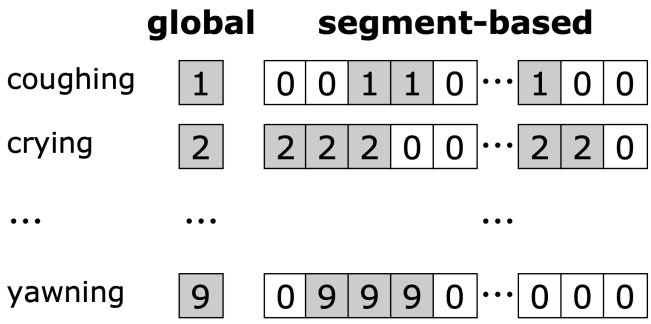

図１非言語音声情報の提案ラベリング方法

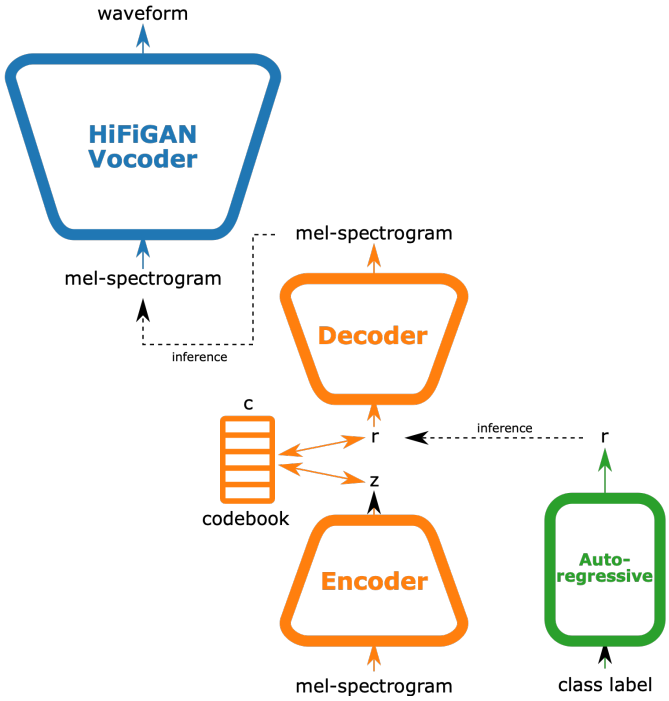

図２非言語音声を生成するネットワーク

課題２：ニューラル言語モデルによる噺の自動生成

課題２の達成に向け、令和３年度はまずは種々の落語音源の書き起こしを行い、本課題に必要なデータ整備を行なった。　続いて令和４年度は、日本語 GPT-2、日本語 BART、日本語 T5 といったニューラル言語モデルを落語の書き起こし文章セット、および、あらすじを記載した文章セットにより Fine-tuning し、題目もしくは最初の数文から残りの噺を生成する実験も行なった。そして、比較実験を通して、日本語 BART モデルが本用途に適している事、および、パラメータ数が多い日本語 BART モデルが最も良い性能を出すことを確かめた。しかし、生成された噺は、必ずしも最後まで噺の内容が一貫していないこともあり、噺としての完成度を高めるためにはさらなる改善が必要である事も確認した。

以下に日本語 BART モデルから生成した結果の１例を示す。

ある所に、そそっかしい殿様がおり、その家老（用人）の三太夫もまたかなりの粗忽者であった。そこで、家老をそそのかすと、その日のうちに殿様の屋敷へ押しかけ、というのがあるという。そしてその翌日にもう一度殿様の屋敷に忍び込んだ。そのとき「何をやってるんだ？」とお目付けが付いてしまった。実は「無頼漢がお城へ帰ってくる」というのだ。（続く）

５．主な発表論文等

〔雑誌論文〕　計2件（うち査読付論文　2件／うち国際共著　0件／うちオープンアクセス　2件）

１．著者名 Cooper Erica、Huang Wen-Chin、Toda Tomoki、Yamagishi Junichi	４．巻 -
２．論文標題 Generalization Ability of MOS Prediction Networks	５．発行年 2022年
３．雑誌名 ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)	６．最初と最後の頁 8442-8446
掲載論文のDOI（デジタルオブジェクト識別子） 10.1109/ICASSP43922.2022.9746395	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 －

１．著者名 Hieu-Thi Luong ,　Junichi Yamagishi	４．巻 -
２．論文標題 Controlling Multi-Class Human Vocalization Generation via a Simple Segment-based Labeling Scheme	５．発行年 2023年
３．雑誌名 Interspeech 2023	６．最初と最後の頁 -
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 －

〔学会発表〕　計2件（うち招待講演　2件／うち国際学会　0件）

１．発表者名 Erica Cooper
２．発表標題 The VoiceMOS Challenge 2022
３．学会等名 Special Interest Group on Spoken Language Processing, Information Processing Society of Japan（招待講演）
４．発表年 2022年

１．発表者名 Junichi Yamagishi
２．発表標題 Speech Synthesis Research 2.0
３．学会等名 34TH CONFERENCE ON COMPUTATIONAL LINGUISTICS AND SPEECH PROCESSING (Rocling 2022)（招待講演）
４．発表年 2022年

〔図書〕　計0件

〔産業財産権〕

〔その他〕

Synthesizing laughter from waveform silhouettes https://arxiv.org/abs/2110.04946

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
研究分担者	Ｃｏｏｐｅｒ　Ｅｒｉｃａ (Cooper Erica) (30843156)	国立情報学研究所・コンテンツ科学研究系・特任助教 (62615)	

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