

**科学研究費助成事業 研究成果報告書**

平成 28 年 6 月 14 日現在

機関番号：12701

研究種目：基盤研究(C) (一般)

研究期間：2012～2015

課題番号：24500640

研究課題名(和文) 視覚障がい者のための自動朗読システム

研究課題名(英文) Automatic Radiogram Generation System for Visually Impaired People

研究代表者

田村 直良 (Tamura, Naoyoshi)

横浜国立大学・環境情報研究科(研究院)・教授

研究者番号：20179906

交付決定額(研究期間全体)：(直接経費) 3,300,000円

研究成果の概要(和文)：当課題では、出来るだけ高品質なラジオドラマを自動的に生成すること、人間によるコンテンツ作成支援を前提とした利用モデルを構築することを目的とした。前者では、発話者ごとに個性を持たせ、同一の音響的パラメータを用いて同一の声色の登場人物として音声合成し、自動生成するラジオドラマを表現力豊かなものにすることが出来る。後者では、背景音楽用や効果音用の音声素材を視聴しながら効率的に選択でき、GUIをベースとした編集機能を実現されている。解析システムの不完全さを後処理システムで補うことになる。実際のシステムの公開、利用促進については今後の課題となった。

研究成果の概要(英文)：There are two sub-goals in this research subject: to improve automatic radio drama generation aiming for high quality output, and to construct utilization models of contents production support system. On the former sub-goal, the system became able to generate expressive radio dramas by assigning same phonetic parameters for utterances spoken by same person in order to characterize utterances in story text. On the latter sub-goal, we construct a GUI-based annotation tool as the post-editing environment for text analyzer which has music/sound database reference and replaying facility. Releasing and utilization of practical system remains as a further research.

研究分野：自然言語処理

キーワード：福祉情報処理 ラジオドラマ自動生成 自動朗読 自然言語処理 音声合成 視覚障がい者支援 娯楽

### 1. 研究開始当初の背景

音楽や娯楽など、生活の基盤に直接関わらない領域は、障がい者へのボランティア活動の対象としてはとかく軽視されがちであったが、インターネット諸技術の発達により、障がい者はボランティアを介さずに、気兼ねなくそれらを扱える可能性が開けてきた。音声合成技術が進歩し、生成された合成音声が多様な場面で使用されるようになり、我々の生活に身近な存在となった。音声合成技術の応用例のひとつに、視覚障がい者や高齢者のための代読支援がある。実際、神奈川県内で一か月に300件のリーディングサービスの利用があり、視覚障がい者からの利用が増加している(2000年)。しかし、現状では音声合成を用いた機械による音声をそのまま用いて朗読を図るとき、人間が朗読するような自然で聞きとりやすいような発話は難しい。一方、自然言語処理技術の研究が進み、発話文の解析、段落の解析、文章の文脈解析など、様々な対象が研究対象になってきている。これら諸技術を応用、活用し、本課題では、物語など文字で書かれたテキスト情報から、その朗読音声を自動合成するシステムの構築を目的とする。なお、図書のデジタル化・音声化に関してはDAISY(Digital Accessible Information SYstem)と呼ばれる国際規格があるが、当システムが生成する音声はWAV形式であり、容易にDAISY規格のコンテンツも自動生成可能である。

### 2. 研究の目的

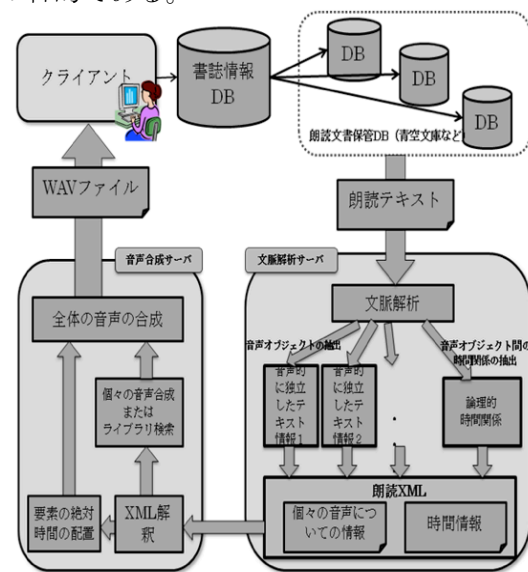
本課題では、視覚障がい者への娯楽提供を想定し、音声合成技術、自然言語処理技術を用いて物語文章から朗読音声を自動生成するシステムの構築を目的とする。

このために、物語から語り手などの地の文、登場人物ごとの発話文、場面の切り替えや状況描写などを分類する技術、それぞれの文に対して人物の特徴付け、登場人物の動作や行動にふさわしい効果音を推定する技術、適切な発声や擬音、シーンにふさわしい背景音楽の選択、抽出された要素に対して、具体的に音声として実現する技術、これら音声の断片について文章から相対的な時間関係を抽出し一つの音声ファイルとして統合する技術が必要になる。さらに、インターネットから様々な電子文書(物語)が入手できるが、ダウンロード操作や文書の管理を考慮したシステム活

用技術が必要となる。

また、視覚障がい者がボランティアの助けなく本システムを利用するために、マウス操作を前提としたGUIでは使いにくい物になってしまう。視覚障がい者の利用に適したユーザインタフェースについても検討する必要がある。

以上のように当課題は、単に福祉目的というだけではなく、福祉的な利用をタスクとした自然言語処理の研究としての側面が強い。文脈処理というとかなり抽象的なテーマであるが、上記のように考慮しなければならない様々な要素を分割して、構造としてとらえることが当課題における文脈処理の目的である。



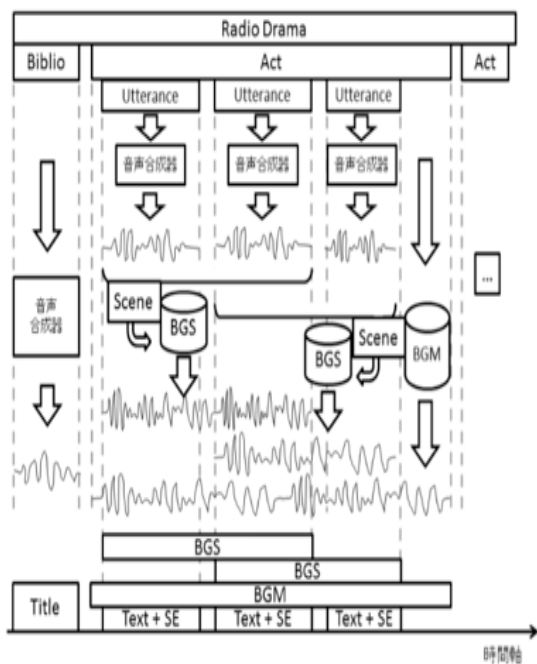
### 3. 研究の方法

(1)文脈解析サーバでは、入力された物語文章を解析し、文章中の発話文についての人物同定や、その他の文について文の種類の自動分類を実現するなど、前述の音声化するための情報を入力文章から抽出する。さらに必要な抽出すべき情報として、事象発生時間の関係解析については記述順として単純化する。発話文と発話文の間の無音の部分である文間ポーズ長の自動判定については、これまでの成果を発展させる。すなわち、文脈に関する素性を十分に取り入れ、分類器による「長」、「短」の判定精度の向上や、可能であれば、数値での推定を目指す。

また、文脈解析サーバからの情報を音声合成サーバに渡すために、XML形式での記述方法を設計する必要があるが、単に記法の問題だけではない。音声劇に関連する本質的な構

成要素を認識し、それらの間の関係をモデル化することにより音声劇全体をモデル化必要がある。このモデルは、朗読XMLとしてとられ、XML形式であるために、同時に伝達情報の表現形式としても利用される。

(2) 音声合成サーバは、文脈解析サーバで得られた情報をもとに朗読音声を合成する。音声合成サーバでは、主に、話速や声質を変化させることにより朗読音声としてのクオリティを確保する。具体的には、平叙文については基本周波数やパワーを文末で下げることで比較的自然な合成音声を得られているが、疑問文等に関しては、合声器自体が行う変調と競合して不自然な発話となりがちであり、これに対処する。各音の素材については、単純に重み付きで飽和加算することで統合できる。なお、感情の表出については音声合成ソフトのオプション機能により実現できる。韻律の変化については、より自然な発話となるよう変調の仕方を工夫する。



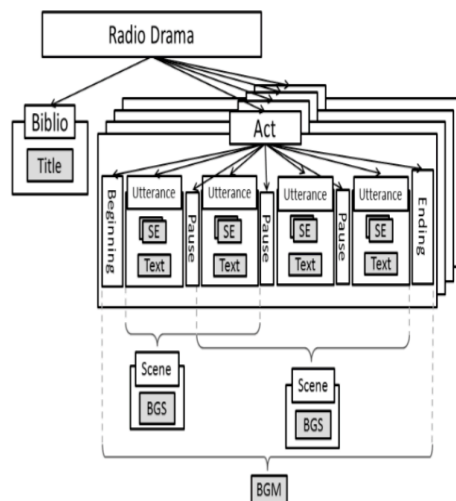
(3) 異なる人物として発話させるために、発話者ごとに音響的なパラメータの組を設定し、「プロフィール」として個性化する。話者個々の特徴をとらえるためには、その前にそれぞれの発話文の話者を同定する必要がある、会話文の話し手の推定機能を実現する。文章中の各種素性の組み合わせに関連するルールを収集し、ルールベースの話者推定システムを構築する。目標として、精度70%で、物語中の発話文について発話者を推定する。生成された音声を聞いて物語を理解するには、話者

同定は重要な要素である。しかし話者推定の精度は70%程度に留まり、この精度ではむしろ聞き手に混乱をもたらす。話者推定の判定精度として95%以上をめざす。そのためには、文章中の代名詞の利用などの照応解析、主語等の省略解析を取り入れる必要がある。物語中で、同一の登場人物が様々な表現で参照され、また、ときには省略されることも多いからである。

(4) 言語処理部(文脈解析サーバ)では、話者推定の不完全さは回避できそうもないことと、感情指標の付与の表出が未対応であったことにより、人手による入力も可能となるようにシステムを拡張し、入力支援の方法を検討する。さらに人手による入力を発展させ、背景音楽、効果音のデータベースを開発し、人手の入力の際にこれらのデータも付与できるようにする。

#### 4. 研究成果

(1) 学会発表④では、ラジオドラマ自動生成システムについて発表した。システムは自然言語解析部と音声合成部に分けられている。両者を連携させるために、朗読XMLの仕様を決定した。朗読XMLは、朗読劇をモデル化し、巨視的なシーンの構造、微視的な登場人物、ナレータの発話の構造を表現できる。また、それらに対してBGMや効果音を割り当てること出来る(下図参照)。



また、同発表④において自然言語処理部を補助/代替する入力支援システムが有効であろうとのコメントを受け、これをヒントに同支援システムを構築し、有効性/効率性を確認した(学会発表②)。支援システムでは、背景音

楽用の音声素材や効果音用の音声素材を視聴しながら効率的に選択できる音素材データベースを含んでおり、GUI をベースとした十分な編集機能が実現されている。さらに、雑誌論文①、学会発表③⑤では、文章の文脈解析の手法が検討されており、物語文を解析して各種の情報を抽出する際の基礎となる事項について検討されている。

プロトタイプ・システムにより生成されたドラマ音声を精査したところ、発話文の話者の同定の精度が、生成された朗読音声の品質向上に必須の要素であることが分かった。誤った話者同定により別人格での発声となり、物語理解に本質的に影響する。照応の問題や省略された主格の推定など、自然言語処理の典型的な課題として検討されているこれらの問題に取り組み、学会発表した(学会発表①)(2)課題の研究機関終了に際して、幾つかの課題が残された。

・物語に登場する話者の同定、話者の特徴抽出は、自然言語処理の問題としても困難な問題であり、代名詞等の照応の問題、名詞や指示対象の同一性指示の問題を含んでいる。さらに話者の行動、言動から話者の性格などを推定し、適した声質をわりあてる。自然言語処理の問題として、今後も継続的に取り組みたい。

・登場人物やナレータの発話に感情を含ませることは、生成されるラジオドラマの質を高めるために有用である。感情を込めた発話合成については、商用の音声合成システムの感情オプション等の利用で対処することが出来るが、発話者、ナレータにどのような感情で発話させるかを文章テキストから判断させることは挑戦的な課題である。

・視覚障がい者の利用を前提とした時、利用者にとどの程度の操作を要求するかを検討が重要である。また、生成された音声を配布する際の(音声合成器の)ライセンスの問題も未着手である。未着手である理由としては、被験者(=視覚障がい者)が見つげづらいことと、学術的な位置づけから優先順位が低いことがあげられる。

<引用文献>

## 5. 主な発表論文等

[雑誌論文] (計 1 件)

- ① 藤田彬、藤田央、田村直良、国語教育的評価項目を考慮した機械学習による日本語文章の自動評価と評価モデルの構築、自然言語処理、査読有、Vol. 19、No. 4、pp. 281-301、2012 年

[学会発表] (計 4 件)

- ① 平川大樹、田村直良、物語テキストにおける登場人物の同一指示解析、言語処理学会第 22 回年次大会、査読無、A6-2、2016 年 3 月
- ② 金子つばさ、吉田有里、田村直良、視覚障がい者の娯楽使用を想定した物語朗読システムの開発、電子情報通信学会福祉情報工学会研究会、査読無、71-7、2013 年 12 月
- ③ 勝又大介、藤田彬、田村直良、文章構造解析に基づく小論文の論理構成における整然さの自動評価、言語処理学会第 19 回年次大会、査読無、B3-1、2013 年 3 月
- ④ 金子つばさ、吉田有里、田村直良、視覚障がい者のための物語テキストの自動朗読システム、第 11 回情報科学技術フォーラム、E-025、2012 年 9 月
- ⑤ Akira Fujita, Naoyoshi Tamura, Automated Evaluation of Texts by Individual Teacher's Model, 7<sup>th</sup> International Conference on e-Learning, ICEL-2012, 香港中文大学, 2012 年

## 6. 研究組織

### (1) 研究代表者

田村 直良 (TAMURA, Naoyoshi)  
横浜国立大学・大学院環境情報研究院・教授  
研究者番号：20179906

### (2) 研究分担者

後藤 敏行 (GOTO, Toshiyuki)  
横浜国立大学・大学院環境情報研究院・教授  
研究者番号：30234991