

# ラジオドラマ生成システムと読み上げデータの作成支援

金子 つばさ<sup>†</sup>      吉田 有里<sup>§</sup>      田村 直良<sup>‡</sup>

<sup>†</sup> 横浜国立大学 教育人間科学部

<sup>§</sup> 日本ビジネスシステムズ

<sup>‡</sup> 横浜国立大学大学院 環境情報研究院

{kaneko-tsubasa-xc<sup>†</sup>, tam<sup>‡</sup>}@ynu.ac.jp

## 1 はじめに

本稿では、視覚障がい者への娯楽支援を目的とする、物語テキストから朗読音声を生自動生成するラジオドラマ生成システムの構築について述べる。

音声合成技術の発展とインターネットの普及は、視覚障がい者へのアクセシビリティに大きく貢献し、彼らが独力で自由に読みたいテキストへアクセスするための重要な手段のひとつとなった [2]。視覚障がい者への生活基盤に直接関わる部分の支援が進む一方、娯楽に関する支援は遅れがちである。今日、青空文庫<sup>1</sup>や小説投稿サイトによってインターネット上に公開されている物語や、電子書籍など、電子テキストの物語に触れる機会が増えてきている。しかし、現在の読み上げソフトは単一話者による読み上げがほとんどであり、物語のような娯楽性の高いテキストの読み上げには物足りなさを感じさせる。これらは登場人物によって読み上げる声を変えたり、雨や風など、登場人物のいる環境を表す音を入れたりすることによって、より豊かな音声になると考えられる。

我々は、視覚障がい者の娯楽支援を目的としたラジオドラマ生成システムの構築を進めている [1]。本システムでは、ラジオドラマをモデル化し、自然言語処理技術により物語テキストからセリフの話者や話者の声質、感情、効果音などを推定することによって、それらを考慮した娯楽性の高い読み上げ音声を生成することを目指している。しかし現実問題として、物語中で明示的に示されていない要素を自然言語処理で自動抽出するのは難しい。したがって、実音声を生成するにあたっては、機械による自動抽出と共に、その出力を容易に修正するための支援環境も必要となる。

本稿では、ラジオドラマのモデル化について述べ、そのモデルに基づいた音声を生成するラジオドラマ生

成システムの構成について述べる。また、機械による自動抽出結果の修正支援ツールについて述べる。

## 2 ラジオドラマ生成システム

### 2.1 ラジオドラマモデル

ラジオドラマ生成システムで定義するラジオドラマモデルでは、各登場人物への異なる声質の割り当てやセリフへの話者設定、文への効果音付与、場面への背景音楽の付与などの表現が可能である。

本モデルでは、ラジオドラマは物理的な音声データとしての音声要素と、それらの間の意味的な関係を規定する意味要素によって構成される。

#### 2.1.1 音声要素

音声要素は、内容文の読み上げ音声や効果音、背景音楽など、人間が実際に耳にする音を用途別に定義したものである。音声要素を以下に示す。

- Title : タイトルや著者などの書誌情報の読み上げ音声
- Text : 物語テキストの内容文の読み上げ音声
- SE (Sound Effect、効果音) : 扉をノックする音やガラスの割れる音のような単発的な音
- BGS (Background Sound、環境音) : 雨音や風の音のような持続的な音
- BGM (Background Music、背景音楽) : 場面の雰囲気形成する音楽

<sup>1</sup><http://www.aozora.gr.jp>

### 2.1.2 意味要素

意味要素は、音声要素の意味的なまとまりに着目してラジオドラマの構造を定義したものである。意味要素を以下に示す。

- RadioDrama：ラジオドラマ全体を表す。意味要素 Biblio と複数の意味要素 Act から構成される
- Biblio：書誌情報に関する意味的なまとまり。書誌情報の読み上げ音声である音声要素 Title から構成される
- Act：物語の場面を表す。音声要素 BGM（背景音乐）と意味要素 Utterance、Pause、Scene、Beginning、Ending から構成される。BGM は Act 内の始めから終わりまでの全要素を通して適用される
- Utterance：物語テキスト中の一文を単位としたまとまり。音声要素 SE（効果音）と文の読み上げ音声である音声要素 Text から構成される。Text には、どんな声質で読み上げられるかの話者が設定される。また、感情・語調も反映される。Utterance 内に Text は一つ、SE は 0 個以上付与することができ、SE には Text に対してどのタイミングで鳴るか（先、後、同時など）を指定する
- Pause：Utterance 間のポーズを表す
- Scene：音声要素 BGS（環境音）についてのまとまり。BGS の鳴る範囲を指定する。Utterance の始めから、Utterance の終わりを単位として付与され、Scene 同士は一部あるいは全体を重ねることができる
- Beginning：Act の発話開始までのポーズ。場面の効果的な始まりを表現する
- Ending：Act の発話終了後のポーズ。場面の余韻を表現する

ラジオドラマモデルの構造を図 1 に示す。

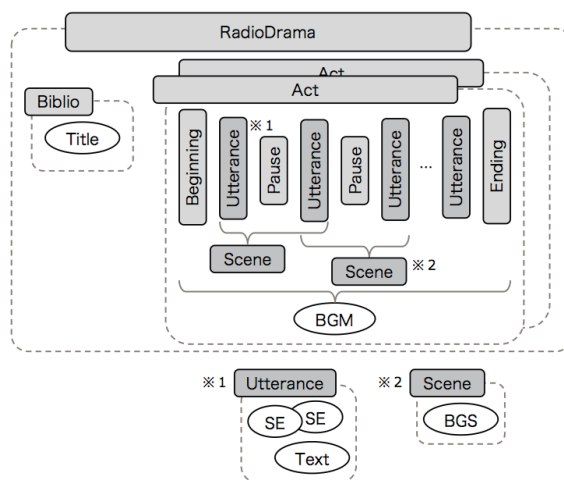


図 1: ラジオドラマモデル

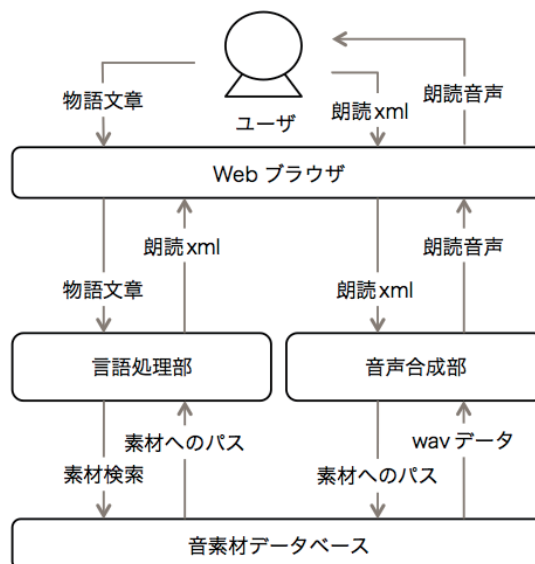


図 2: ラジオドラマ生成システムの構成

らラジオドラマモデルに則った各情報を抽出し、朗読xmlを出力する。音声合成部では、朗読xmlを読み込み、記述された内容に従って各文の合成音声の生成や音素材の取得、加工を行い、時間軸上に配置することによって朗読音声を生成する。音声要素同士が重なり合う部分に関しては pcm データの飽和加算により合成する。

### 2.3 朗読 xml

朗読 xml は、ラジオドラマモデルを機械で扱うためのデータ形式である。人間が見て大体の構造を見て取れること、種々のプログラミング言語で簡単に扱うこ

## 2.2 ラジオドラマ生成システムの構成

ラジオドラマ生成システムは、ラジオドラマモデルに基づいた朗読音声を生成するシステムである。システムの構成を図 2 に示す。

本システムは、インターフェース部と言語処理部、音声合成部から構成される。インターフェース部は Web サーバによって提供され、ユーザーと言語処理部、音声合成部間のやり取りを管理する。言語処理部では、自然言語処理技術を用いて、入力された物語テキストか

とができることにより xml 形式を採用する。

本ラジオドラマモデルにおける意味要素は、朗読 xml により記述され、音声合成部において話者や効果音などを考慮した音声を生成するために必要となる情報を記述する。

## 3 朗読 xml 修正支援ツール

### 3.1 概要

朗読 xml 修正支援ツールは、ラジオドラマ生成システムの言語処理部で抽出された朗読 xml の誤り修正を容易に行えるようにすることを目的とする。このツールでは、関連知識を持たないユーザでも扱えるよう、直感的で使いやすいと思われる GUI によって構築し、以下の機能を提供する。

- 朗読 xml のタグ構造に基づくアノテート機能
- 言語処理部の自動抽出結果の校正
- 音素材のキーワード検索機能
- 音素材や話者の声質、ラジオドラマの試聴機能

### 3.2 構成

朗読 xml 修正支援ツールの構成を図 3 に示す。

本ツールは、物語文章または修正が必要な朗読 xml を入力とし、朗読 xml として出力する。また、データベースサーバ上の音素材データベースと連携することにより、音素材のキーワード検索機能を実現する。商用の音声合成器<sup>2</sup>を内蔵することによって、合成音声の試聴機能を可能にする。

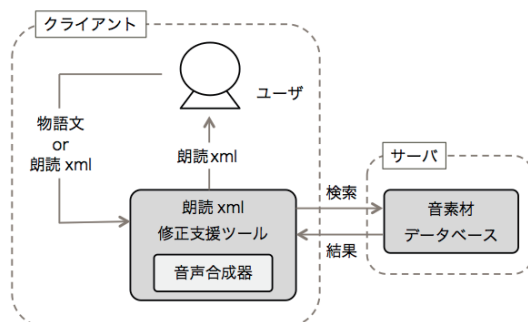


図 3: 朗読 xml 修正支援ツールの構成

<sup>2</sup> ANIMO FineSpeech SDK V3.2 を使用

### 3.3 修正の流れ

朗読 xml 修正支援ツールのメイン画面を図 4 に示す。

ユーザは、入力ファイルとして朗読 xml を開く。セリフの話者の変更や SE の付与、削除などは、画面の右側に表示される物語テキストの文を選択し、右クリックを押すと表示されるメニューから行う。セリフの話者候補の作成、削除、声質の変更は、画面の左側の「話者の設定」ボタンで表示される画面で行う。修正結果は朗読 xml 形式として保存され、再度読み込んで修正を行うことができる。

アノテーションの様子を図 5 に示す。

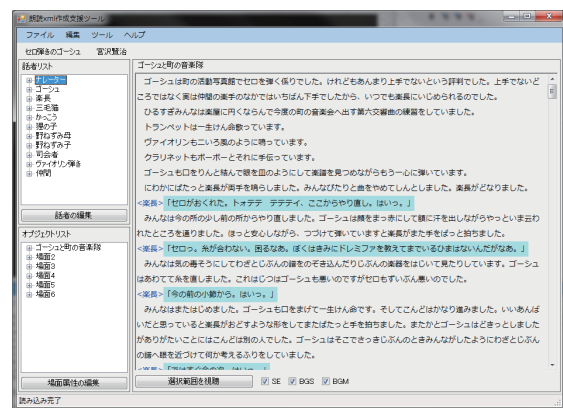


図 4: 朗読 xml 修正支援ツールのメイン画面

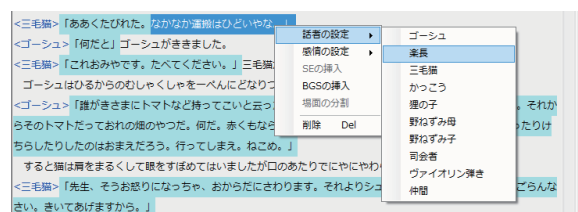


図 5: アノテーションの様子

### 3.4 音素材の検索

ラジオドラマモデルでは、合成音声以外の音素材として SE（効果音）、BGS（環境音）、BGM（背景音楽）を用いる。音素材は、特に SE において多くの種類が用意されていることが望ましい。しかし、数が多くなればなるほど目的の音を探すのが大変になる。

本ツールでは、ユーザの効率的な音素材の検索を実現するため、各音を特徴づける語とカテゴリ、擬音語を格納した音素材データベースを作成し、キーワードによる検索機能を実装する。ユーザはカテゴリから探

すか、キーワードを入力して検索を行うことで目的とする音の候補を絞り込むことができる。

### 3.5 試聴

SE や BGS、BGM といった音素材は、音名やキーワードで選択してもそれが意図した音かどうかは実際に音を聞いてみなければ分からない。また、ラジオドラマの一部を試聴するために、修正したファイルを保存してから音声合成器にかけるのでは、効率が悪い。

本ツールでは、音素材の検索結果に対して試聴を行うことができる。また、ラジオドラマ生成システムの音声合成部と同等の機能を内蔵することにより、テキストから試聴したい範囲を選択して、その場でラジオドラマを試聴することが可能である。

### 3.6 実験（予定）

朗読 xml 修正支援ツールの有効性を評価するために、実証実験を行う予定である。物語 5 編（『走れメロス』、『芋粥』、『羅生門』、『駆込み訴え』、『伊豆の踊子』）について、以下の実験を行う。

実験 1（客観評価）：

セリフ話者の誤り修正をタスクとし、テキストエディタで朗読 xml を直接修正する場合と、本ツールを用いて修正する場合について、校正作業時間を比較する。修正する朗読 xml は、正解データのセリフの話者指定の 10 % 程度について人為的に異なる話者に書き換えたものを用いる。

実験 2（主観評価）：

使いやすさ、効率性、脚色のしやすさなど、ツールへの使用感について主観的な意見を求める。

## 4 まとめと今後の課題

本稿では、ラジオドラマのモデル化とラジオドラマ生成システムの概要について述べた。また、朗読 xml の修正支援ツールについて述べた。

今後の課題としては、朗読 xml 修正支援ツールの実験を行い評価をすること、修正データを効率的に言語処理部の精度向上へと還元するための仕組みを考えることが挙げられる。

## 参考文献

- [1] 金子つばさ, 吉田有里, 田村直良. E-025 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. 情報科学技術フォーラム講演論文集, Vol. 11, No. 2, pp. 213–215, 2012.
- [2] 浅川智恵子. アクセシビリティ向上のための情報技術とユニバーサルデザイン (< 特集 > 情報サービスとユニバーサルデザイン). 情報の科学と技術, Vol. 62, No. 5, pp. 192–197, may 2012.