実世界システムプロジェクト

音声合成部の高品質化

リサーチアシスタント 米田隆一

情報理工学系研究科システム情報学専攻

概 要

ユーザのグループを引率し案内をする案内ロボットのシステム開発において、ユーザとの対話に必要である音声合成部の開発を行った。合成用モデルの学習においては、韻律ラベル、音素ラベル等の言語情報の修正を行い、再学習を行った。話者適応モデルの学習においては、特定話者の音声データを収集し、学習を行う予定である。また、大学院の講義である実世界システム講究に参加してプロジェクト進行のための議論を進めた。

1 はじめに

案内ロボットのシステム開発において、音声合成部はユーザとの対話に必要であり、人間らしい自然な振舞いを要求される。人間同士が対面で会話をする場合には、音声による言語的な情報のやりとりに加えて、非言語情報やパラ言語情報から相手の感情や態度を読み取ったり、自分と相手の発話のリズムから息が合っているかの判断を行ったりしている。本研究では、話者モデルのカスタマイズ、音色制御、韻律制御などの機能を用いた表情豊かな音声の合成を目指す。これには感情および態度 (疑問、強調、自信の有無など) の表現が必要であるが、既存の音声合成システムではこれらの機能を実現することが困難である。そこで、我々の研究グループが開発している Galatea Talk を使用し、話者モデルのカスタマイズ、音色制御、韻律制御などの機能を用いて表情豊かな合成音声を実現する。

2 音声合成技術

日本語の音声合成の場合、漢字仮名混じり文が入力となる。テキスト音声合成は大きくテキスト解析処理と波形生成処理にわかれる。テキスト解析では文を単語に分割する通常の形態素解析処理に加え、読みの付与、単語のアクセント型の決定、それらが連結した句としてのアクセント型の決定が行われる。波形生成処理では、入力されたテキストと、それに付随した韻律・言語情報を実際に音声波形に変換する。

本手法では、表情豊かな音声合成を目標とし、そのための基礎を実現する。この機能は、我々が開発している音声合成エンジン Galatea Talk を使用し、話者モデルのカスタマイズ、音色制御、韻律制御などの機能を用いることで可能である。この音声合成システムには、Galatea Project で開発された隠れマルコフモデルに基づく音声合成法を用いており、モデルパラメータの変換により、柔軟な音色の制御が行える。今回は、合成音声の高品質化と、情緒を表現できる音声合成の実現に向けて本手法を応用する。

3 音声合成手法

3.1 合成用モデル

前節で述べたように、テキスト文は、テキスト解析処理と波形生成処理を経て合成音声へと変換される。テキスト解析処理の段階では、アクセント、音素境界 (ポーズ)、音素ラベルが正しく付与

されるとは限らず、波形生成処理の段階で、学習時に誤ったラベルが用いられることが問題である。以下の節では、これらの誤りが表情豊かな音声を合成するにあたってどう障害となるかを示し、どう解決したかを示す。

3.1.1 アクセント部

日本語はアクセントにより区別される単語の存在する言語である (例: 箸 vs. 橋)。アクセントの区別がなければ音声合成の出力は当然不自然になってしまい、対話の相手に違和感がのこる。アクセントに関する修正の結果、合成音の韻律変化の自然性が大幅に改善された。ラベルの修正の実行画面を図??にしめす。

3.1.2 話者の独自性

話者の独自性をより明確に再現するために、分析合成系 (HMM から出力された音響パラメータから波形を生成する部分) に話者ごとの分析パラメータ (ポストフィルタ係数) を導入する。この係数を話者ごとに調整することで、明瞭性の高い合成音声が得られることを確認した。

3.1.3 息継ぎ

誤った息継ぎの挿入があると、継続長モデルの学習に問題が生じる可能性が高い。例えば、「今日は」と合成するときに「今日、は」のような不自然な息継ぎが入りやすくなる。このような不具合を起こさないよう息継ぎ位置の修正を行い、案内ロボットによる不自然な対話の改善を確認した。

3.1.4 音韻ラベル

日本語において文を単語に分割することを形態素解析といい、形態素解析の入力は漢字仮名混じりであることは前節で述べた。形態素解析による読みの出力は実際の音声と異なる場合がある (例: ニホン vs. ニッポン)。部分的な誤りであっても、受け取る側では全体としての品質を大きく損ねることになる。このような、読みラベルと実際の音声波形との食い違いを修正した。

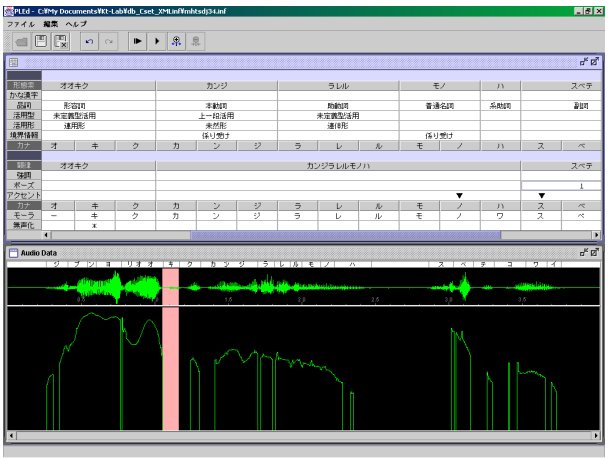

図 1: ラベル修正ツールの実行画面

3.2 話者適応モデル

話者適応モデルの作成にはデータの収集が必要となる。話者モデルの対象としての話者数はまだ十分とはいえない。音声ファイルの収集などにより多数の話者に対応していく予定である。

4 まとめ

本手法では、人間と案内ロボットとの快適な対話を目指し、音声合成部の開発を行った。合成用モデルの学習においては、通常の言語的な情報のやりとりに加えて、発話のリズム、アクセント、読みを修正した再学習を行い、案内ロボットの対話の改善効果を確認した。話者適応モデルの学習においては、特定話者の音声データの収集を予定している。今後の展望としては、人間それぞれの性格に合った話し方を案内ロボットに盛りこむ予定である。