

音声対話システムにおける韻律修正合成音声の利用

木澤 宏 白木 克英 趙 國 山下 洋一

立命館大学・情報理工学部

1 はじめに

音声対話技術を用いたヒューマンインターフェイスの期待は高まっているが、十分に世間一般に広まっているとは言いがたい。そのため、親しみやすさや利便性を向上させることが求められる。より使いやすく親しみやすい音声対話システムを構築していくために、合成音声の韻律情報を修正し感情があるような合成音声を作成し、音声対話システムで利用すると効果があると考えられる。

本報告では、感情豊かな合成音声を作成するための韻律修正ツールの開発、そのツールによって作成された合成音声の音声対話システムでの利用について述べる。

2 擬人化音声対話ツールキット Galatea

音声認識、音声合成、顔画像合成の機能を持ち、対話制御の元でユーザと音声で対話する機能を備えたインターフェイスが擬人化音声対話エージェントである。本研究では、音声対話技術コンソーシアム(ISTC)によって開発されている擬人化音声対話エージェントツールキット Galatea[1][2][3]を利用する。

Galatea は音声認識 音声合成 顔画像合成のモジュールを統合し音声対話システムを提供する。

顔画像合成モジュール(FSM)

音声合成モジュール(SSM)

音声認識モジュール(SRM)

の三つのモジュールをエージェントマネージャ(AM)が統合し、タスクマネージャ(TM)、あるいはダイアログマネージャ(DM)の下で動作する[2]。基本構成を図1に示す。

3 音声合成モジュール

3.1 Galatea Talk

音声合成モジュール SSM は、日本語テキスト音声合成システムとして単体で動作する GalateaTalk として開発されている。入力コマンドの解析部、音声波形を合成する音声合成エンジン(波形生成部)、音声出力を音声出力部を

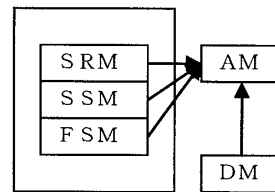


図1 Galatea の基本構成

実装した gtalk が、形態素解析を行う chasen および音韻交替処理やアクセント結合処理などを行う chaone を内部で呼び出すことによって、GalateaTalk は実現されている。

GalateaTalk の波形生成部では、HMM に基づいた音声合成[4]を用いている。音声合成のための形態素解析では、発音形およびアクセントに関する素性の情報を得ることが必要であることから、UniDic プロジェクト[5]によって開発された辞書を利用している。

GalateaTalk では合成された音声の韻律情報をファイルに書き出したり、ファイルに書かれている韻律情報を読み込んで音声を合成することができる。ここで、韻律情報とは、音素の時間長、声の大きさ、声の高さをいう。

3.2 Prosedit

韻律情報の手作業による修正を行うために、GalateaTalk の生成した韻律情報ファイルに対する修正ツール Prosedit を開発している。

Prosedit は GalateaTalk にコマンドを送ることにより、音声の合成、韻律情報ファイルの生成、合成音声の出力などを行うことができる。Prosedit は、GalateaTalk の生成した韻律情報ファイルを読み込み、GUI インターフェイスを利用して、グラフィカルに基本周波数、パワー、音素継続時間長を修正することができる。さらに、GalateaTalk にコマンドを送ることにより、修正された韻律情報ファイルを用いた音声を合成することができる。図2に動作画面の例を示す。赤線が音の大きさ、青線が音の高さを表す。韻律情報の修正は、画面左のパネルでパラメータの種類を選び、右のグラフ中でマウスで操作

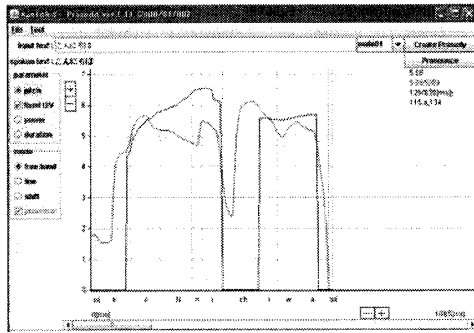


図 2 : Prosedit の動作例

することにより行う。

対話中の発話は、対話の流れに応じて動的にその内容や感情が決定される場合もあるが、多くの対話では、あらかじめ内容や感情を決定しておくことのできる発話も多く存在すると考えられる。そのような発話に対しては Prosedit を利用することにより、発話の韻律情報を修正した合成音声をあらかじめ作成しておき、対話システムで利用することができる。

4 音声対話システム作成

4.1 交通安全教育プログラム

感情ある合成音声を使うシステムとして効果があるシステムとして子供に対する交通安全教育システムを作成した。交通安全に関わる問題に間違えたなら怒る、論すといった感情、正解したなら喜ぶ、褒めるといった感情表現が効果的だと考えたからである。交通安全教育システムの対話例を図 3 に、実行画面を図 4 に示す。

4.2 韻律を修正した合成音声作成

問題に正解した時に出力する明るい合成音声、不正解の時に出力する合成音声、理解したかどうかを問う合成音声を Prosedit を用いて作成する。合成音声の韻律情報の修正する指針として、表 1 を参考にしながら編集する[6]。

表 1 : 韻律情報修正の指針

	喜び	怒り	悲しみ
パワー	全体的に大きくする	全体的に大きくする	全体的に小さくする
継続時間	文の始め短くし、最後長くする。	全体的に短くする	全体的に長くする
ピッチ	文の最初の語を上げ、最後を下げる	上げ下げを激しくする	文の最初の語を下げ最後の語を戻す

システム：交通安全について学びましょう。始めますか？

ユーザ：はい。

システム：それでは始めます。

システム：第一問目です。自転車で歩道を走っても良いでしょうか？

ユーザ：いけない。

システム：正解です。自転車で歩道を走ってはいけませんね。

システム：次の問題に進みますか？

図 3 : 交通安全教育システムの対話例

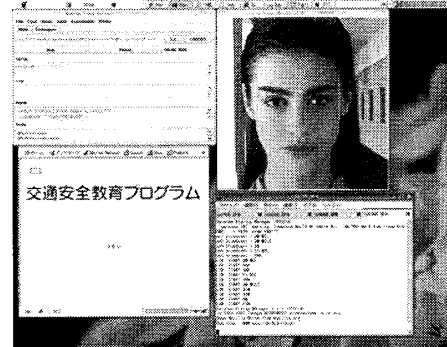


図 4 : 交通安全教育システムの動作画面

5 まとめ

本研究では感情豊かな合成音声を作成するツールの開発と、そのツールにより作られた合成音声を使用した音声対話システムを開発した。韻律情報修正の難しさの問題が課題として挙げられる。今後は、評価実験を行い、更に使いやすく親しみやすい音声対話システムとなるよう改善していく。

参考文献

- [1] 擬人化エージェントツールキット GalateaToolkit <http://hil.t.u-tokyo.ac.jp/~galatea/index-jp.html>
- [2] 嵯峨山 茂樹, 伊藤 克己, 宇津呂 武仁, 甲斐 充彦, 小林 隆夫, 下平 博, 伝 康晴, 徳田 恵一, 中村 哲, 西本 卓也, 新田 恒雄, 広瀬 啓吉, 峯松 信明, 森島 繁生, 山下 洋一, 山田 篤, 李 晃伸, "擬人化音声対話エージェント基本ソフトウェアの開発プロジェクト報告" 情処学研報, 2003-SLP-49-56 (2003)
- [3] 山下 洋一, 李 晃伸, 河原 達也, 四倉 達夫, 西本 卓也, 桂田 浩一, 新田 恒雄, "音声対話技術コンソーシアム(ISTC)の活動成果報告", 情処学研報, 2008-SLP-73, pp.47-52(2008)
- [4] 益子 貴史, 徳田 恵一, 小林 隆夫, 今井 聖, "動的特徴を用いた HMM に基づく音声合成", 信学論(D-II), J79-D-II, 12, pp.2184--2190 (1996)
- [5] <http://www.tokuteicorpus.jp/dist/>
- [6] 緒方 信, 四倉 達夫, 森島 繁生, "韻律情報の制御による感情音声合成のための声質変換", "電子情報通信学会技術研究報告, Vol.99, No.582(20000121), pp.53-58,(2000)