ISS-P-37

感情音声を用いた 韻律制御音声合成システムの検討

栗原　大樹　　加藤　正治　　小坂　哲夫
山形大学大学院 理工学研究科

1. はじめに

現在，入力音声を用いて合成音声の韻律情報を制御するシステムが提案されている[1][2]．このシステムではモデルからのスペクトル情報と，韻律制御用の入力音声からの韻律情報を用いて音声が合成される．音声による韻律制御はユーザが直感的に合成音声の発話表現を制御することが可能であり，従来の方法では生成できなかった多種多様な音声を合成することが可能となる．応用を考えると，韻律情報の変動が激しい感情音声による制御が期待されるが，入力音声に感情音声コーパス等を用いた場合の実験は行われておらず，合成される音声に対しどの程度の問題や劣化が生じるのか明らかになっていない．本稿では韻律制御用の入力音声として OGVC[3]を利用し，感情音声を用いた場合の問題点について調査・検討を行う．

2. 実験条件

合成音声用 HMM の学習データには，ATR 音素バランス文日本人男性・女性話者各 1 名の A-I セット 450 文を使用する．韻律情報の内，音素時間長を分析する際に用いる音声認識用 HMM は学習データに CSJ 学会講演及び模擬講演の男性・女性話者 2667 講演を用いて，不特定話者モデルを作成する．韻律制御用の入力音声として，「感情評定値付きオンラインゲーム音声チャットコーパス」[3]の演技音声を使用して実験を行う．

本稿では二種類の評価実験を行った．一つはピッチの変換精度に対する評価である．本稿では入力音声の対数ピッチの平均を合成音声の対数ピッチ平均に合わせる処理を行っている．この処理が適切に入力音声の韻律情報を変換できているかを調べるため，ピッチ変換を行わなかった場合と行った場合の主観評価を行った．また，音素時間長分析精度が合成音声に与える影響を調べるため，システムによって自動的に得られた自動音素長と，手動によって作成した正解音素長それぞれで音声を合成し，主観評価を行った．主観評価では，自然性に対しては MOS 評価を，了解性と再現性についてはDMOS評価を行った．了解性は合成音声が入力音声と発声内容が一致しているかを，再現性は合成音声が入力音声の韻律情報を再現できているかを評価する．

3. 実験結果

実験結果を図 1，2 に示す．図 1 では再現性の項目においてピッチ無変換での音声がスコアが高く，ピッチの変換

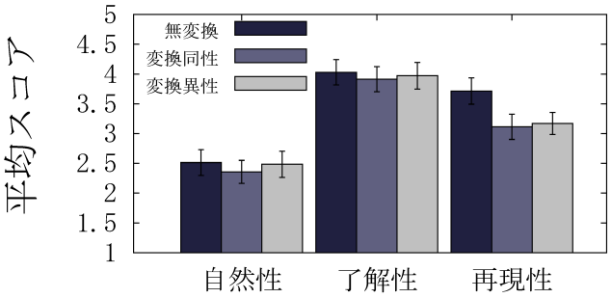

図1. ピッチの変換精度に対する主観評価実験結果

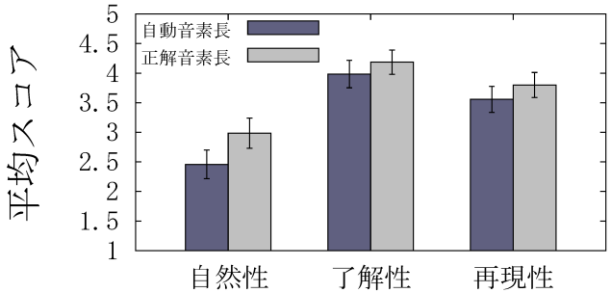

図 2. 音素時間長分析精度の主観評価実験結果

が韻律情報をうまく再現できていないことがわかる．図 2 では正解音素長を用いて合成した方が自然性のスコアが高く，音素時間長の分析精度が合成音声の自然性に影響することがわかる．また，別途行った客観評価の結果，感情がある音声の方が音素時間長がずれる傾向にあった．

4. まとめ

本稿では，入力音声で韻律情報を制御する音声合成システムにおいて，感情音声を入力とした場合の問題点について調べた．実験結果では，ピッチ変換が入力音声の韻律情報を再現できていないこと，音素時間長の分析精度が合成音声の自然性に影響し，感情音声では分析精度が低下することがわかった．今後の展望として，他のピッチ変換法の検討や音素時間長分析精度の改善，F0 量子化コンテキストを用いた声質変換との比較を行う．

謝辞　音声合成に関し種々ご教示くださった東北大の能勢隆講師に感謝致します．

参考文献

[1] 栗原　他，IPSJ 東北支部研究会，12-6-B3-4，2013.
[2] 西垣　他，信学技報，Vol. 114，No. 365，SP2014-115，pp. 81-86，2014.
[3] 有本　他，音講論(秋)，1-P-46a，pp. 385-388，2013.