

感情音声を活用した親しみやすい音声対話システムのための感情合成音の検討

M1 加藤 大地 (岡山大学 ヒューマンセントリック情報処理研究室)



↑ 合成音声



OKAYAMA UNIVERSITY

1. 研究背景

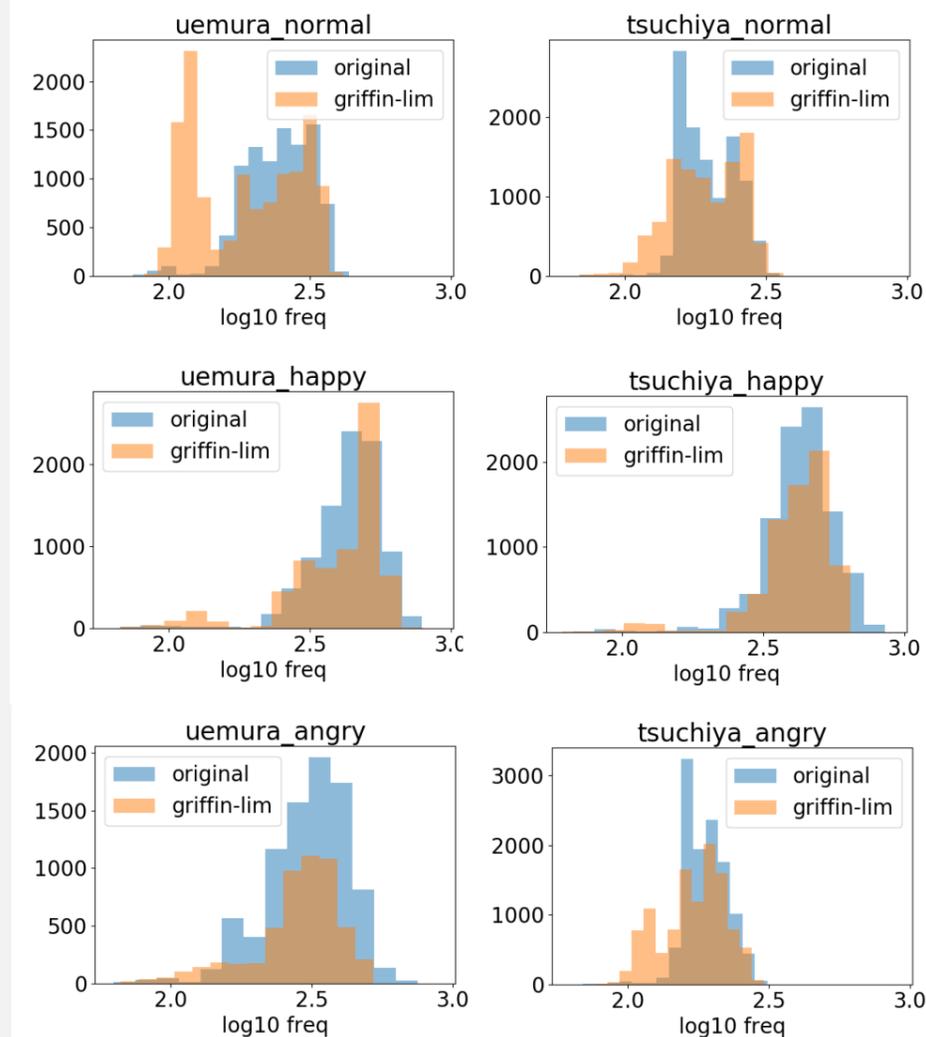
- 音声対話システム (SDS)
 - 人間同士の対話：感情的な表現
 - 例：明るい声, 沈んだ声
 - 対話システム：単調的な応答が一般的
 - 感情表現により人間らしく, 親しみやすく

2. 感情表現を使うSDSの要件

- 感情音声の合成
 - 中立的, 楽しい, 悲しい
- 応答文の感情推定
 - 間違った感情では不快に感じる可能性
 - 対話開始直後からいきなり感情豊かに話すべき?
 - ユーザ発話 or システム応答の単語から推定 [Chiba+,2018]
- リアルタイム合成
 - できなければあらかじめ合成
 - RTF (音声の長さに対する合成にかかる時間の比) が1未満
 - FastSpeech[Ren+,2019]+Parallel WaveGAN[Yamamoto+,2020]なら速い

3. 感情音声の合成

- 音響モデルの学習
 - 声優統計コーパス
 - 100文×3話者×3感情
 - normal, happy, angry
 - Transformer-TTS[Li+,2019]+Griffin-Lim
 - ESPnet[Watanabe+,2018] JVSレシピ
 - JSUT学習済みモデルを再学習
- 合成結果
 - 声質の再現
 - アクセントは不完全
 - tsuchiya_happy は語尾が上がる⇒再現
 - F0の分布は類似



4. 今後の課題

- 本稿の内容
 - 感情音声を使うSDSの要件
 - 感情音声の合成
- 今後の課題
 - ニューラルボコーダによる高品質な合成
 - 合成音声の主観評価
 - 応答文の感情推定
 - リアルタイム合成