豊かな感情を表現する音声合成の研究

Research on speech synthesis that expresses rich emotions

魏　超然

指導教員　大野　澄雄

東京工科大学　大学院　バイオ・情報メディア研究科　コンピュータサイエンス専攻　大野研究室

キーワード：感情音声合成、機械学習

１．初めに

音声合成技術の研究開発により、合成音声の音質は大きく向上した。しかし、現在の音声合成技術の研究はまだ中立的な音声が中心であり、感情音声合成の研究は少ない。人間生活における音声の需要は、基本的なテキスト内容だけでなく、豊かな感情的な情報も求められている。そのため、感情音声合成の研究は、音声研究の分野でも重要な研究対象であると考えられる。

End-to-end に基づく音声合成 Tacotron は、2017 年 Google から英語によるシステムが提案された。従来の HMM 統計的パラメトリック音声合成と比較して、end-to-end に基づく音声合成は合成音声の音質が向上される。本研究では、日本語による end-to-end に基づく Tacotron2 の手法を用いて、豊かな感情を表現する音声合成の研究を行っている。

２．関連研究

2.1 Tacotron モデルに基づく end-to-end 音声合成[1]

2017 年、Google チームは、英語による end-to-end 音声合成 Tacotron モデルを提案した。Tacotron は文字から音声を直接合成する end-to-end のテキスト読み上げモデルである。

2.2 Tacotron2、テキストから直接音声を合成するためのニューラルネットワーク アーキテクチャ[2]

このシステムは、文字列をメルスケールのスペクトログラムにマッピングする sequence-to-sequence 特徴予測ネットワークと、それらのスペクトログラムから時間領域の波形を合成するボコーダーとして動作する修正 WaveNet モデルで構成される。

2.3 Tacotron２に基づく日本語音声合成の実装 [3]

「Python で学ぶ音声合成」では、JSUT コーパスを使用して、Tacotron２に基づく日本語音声合成システムを作成した。

3. データベース

本研究では、JVS の日本語コーパスと OGVC の感情音声コーパスを使用する。データ量が不足しているため、JVS で学習したモデルを元に転移学習を行い、OGVC コーパスを追加する。

JVS コーパスは、100 人のプロフェッショナル話者（声優・俳優など）による、高音質(スタジオ収録)・高サンプリングレート(24 kHz)・多数の(30 時間) 音声ファイルである。

OGVC コーパスは、話者 13 人で、オンラインゲーム中のプレーヤーに音声チャットを利用させ、自然に感情が表出した音声を収集して、感情のない中立に加えて、Pluchik の立体構造モデルに基づくの８種類の感情種別ラベルが付与されたものである。

4. 研究方法

JVS コーパスの複数話者の音声データを利用して、Tacotron2 の中立発話モデルを構築した。OGVC 感情コーパスは転移学習を利用し、日本語データベースに複数感情データを追加した。追加したデータベースで複数感情をトレーニングし、複数感情の音声モデルを作成した。最後に合成音声の精度を検証した。

5. 研究経過

①JVS の日本語データをトレーニングして、OGVC の感情コーパスを追加した、転移学習を行う。

②複数感情データをトレーニングして、複数感情の音声モデルを作成し、Tacotron2 に基づく音声合成システムを構築する。

③学習データ(OGVC)と学習量(学習回数)に基づく、合成音声の精度を検証する。

6. おわりに

本研究では、日本語による、end-to-end に基づく Tacotron2 の手法を用いて、豊かな感情を表現する音声合成の研究である。最後は合成した複数感情の音声に対して、精度の検証と評価を行う。

参考文献

[1] Wang, Y., Skerry-Ryan, R. J., Stanton, D., Wu, Y., Weiss, R. J., Jaitly, N., ... & Saurous, R. A. (2017). Tacotron: Towards end-to-end speech synthesis. arXiv preprint arXiv:1703.10135.

[2] Shen, Jonathan, et al. "Natural tts synthesis by conditioning wavenet on mel spectrogram predictions." 2018 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE, 2018.

[3] 山本龍一, 高道慎之介(2021).Python で学ぶ音声合成 インプレス出版.