ISS-A-074

敵対生成ネットワークを使った音声合成に関する研究

西山　蒼太†　　川波　弘道††
† 津山高専専攻科　電子・情報システム工学専攻　　†† 津山高専　総合理工学科

1. はじめに

近年，合成音声の需要が増えている．より良い合成音声を目指す上でより明瞭な音声とより自然な音声がある．本研究では音声合成においてより自然に生成することを目指す．

画像の分野では敵対生成ネットワーク(GAN)[1]という技術を使った生成モデルによって自然な画像を生成することに成功した．本研究では音声においてどのように GAN を効率的に扱えるかについて調査する．GAN の弱点として学習の進み具合が不明瞭な点が挙げられる．そこで，データに対してダウンサンプリングやアップサンプリングを行う AutoEncoder[2]で GAN の学習方法を適用することで学習の精度を測れないか検証する．

2. 実験内容

Donahue らの研究[3]内で紹介された音声波形から学習をする WaveGAN から WaveAutoEncoder を作成する．WaveGAN には Phase Shuffle という波形の周期をランダムにずらす層が採用されていたが，現時点では技術的な観点から実装することができず，WaveAutoEncoder に実装することができなかった．これにより過学習が起きやすくなるが学習自体は可能であると考えられる．

3. 結果・考察

WaveAutoEncoder で学習を行ったところ，学習の進行具合である損失値が下がり収束する様子が確認された．WaveAutoEncoder は入力された音声と同じ音声を出力するように学習しているため，学習の結果を得るために音声を入力して変換を行った．変換された音声を図 1 に示す．見た目から音声波形の様な画像が出力されている．これを聞いてみると息を吹きかけられるようなノイズとそれに埋もれたように少し音声の様なものが聞ける．ノイズは一定の音ではなく抑揚の様な特徴がある．AutoEncoder の機能として入力の特徴を抽出し，その部分を出力するようになっているが，WaveAutoEncoder は音声の特徴をあまり抽出することができず，ノイズが多く生成されている．つまり，比較的低い周波数成分を持つ声の判別ができていないと考えられる．また，図 2 に変換前の音声波形(下)と変換後の音声波形(上)を示す．2 つの波形からは関連が見受けられなく，また変換後の音声波形に明確な周期は見

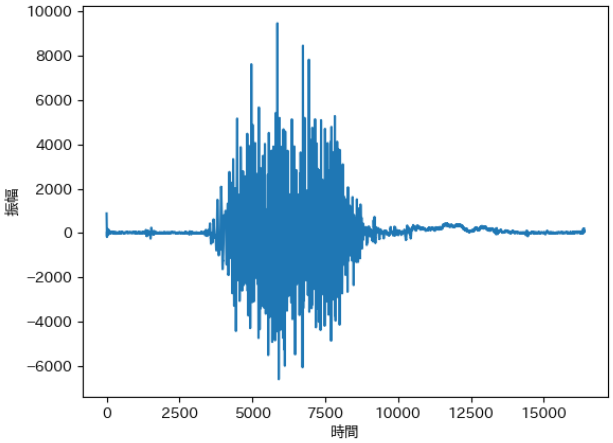

図 1 WaveAutoEncoder により変換された音声

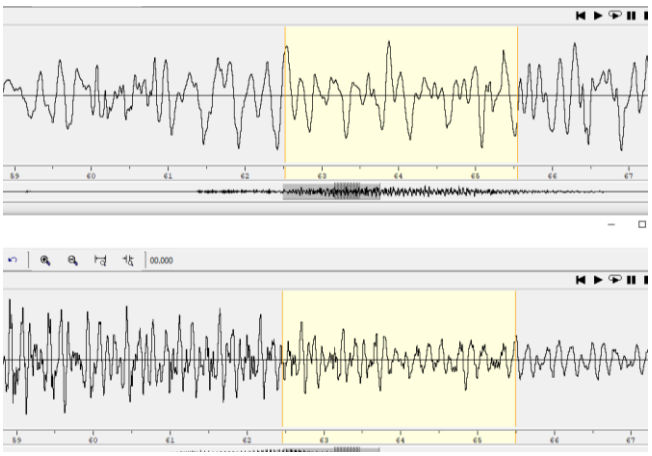

図 2 WaveAutoEncoder への入力(下)と出力(上)の詳細

られない．これらのことから，より低い周波数を扱える学習モデルを検討する必要があると考えられる．

4. 今後の課題

今後は他の GAN でも比較を行いより適した学習モデルを検討する．

参考文献

[1] Ian J. Goodfellow ほか，Generative Adversarial Networks, arxiv.org/abs/1406.2661, 2014.

[2] Geoffrey E. Hinton, R. R. Salakhutdinov, Reducing the Dimensionality of Data with Neural Networks, Science 313, pp.504-507, 2006.

[3] Chris Donahue ほか Adversarial Audio Synthesis, arxiv.org/abs/1802.04208, 2018.