敵対的学習を用いた量子化音声品質の改善

山崎 拓†　　間野　一則†
† 芝浦工業大学大学院理工学研究科システム理工学専攻

1. はじめに

近年，音声を録音して利用する機会が増加しているが，録音時に音声は量子化されるため波形の劣化が生じる．本稿では，劣化した音声の品質を改善する手法を提案する．音声波形の劣化は，量子化する際に限られた桁数の符号で表すために実際の振幅値を切り上げ，あるいは切り捨てて最も近い数値で近似することにより発生する．また，音声のデータ分布は複雑であり，詳細な構造を再現することが困難なため実際の音声と比較して区別がつかないぐらい高品質な音声を作成することができない．そこで本研究では，機械学習により音声の 複雑なデータ分布を学習し，音声品質の改善を行う．

2. 敵対的学習に基づく Denoising Auto-Encoder

械学習を用いた音声品質を向上する手法として，DAE を用いた量子化雑音低減がある．この方法では，音声品質が向上することが確認されたが，中域や高域成分の推定ができなかった[1]．そこで，音声の帯域を分割して学習することにより，詳細な構造を再現する手法[2]を応用し，提案する音声品質の改善では DAE によって低ビット量子化された音声から量子化雑音を除去し，敵対的学習を用いて音声データの詳細な構造を再現する． 図 1 に提案するシステムの概要図を示す．提案システムでは短時間フーリエ変換により対数振幅スペクトルを抽出する．DAE を生成器として使用するため，雑音を除去しクリーンな音声を生成するように学習する．この時，生成された音声と原音声を識別機 D へと入力し，識別機 D は入力される音声が，原音声か生成された音声のどちらであるかを判断する．その後，生成された音声に窓かけを行って滑らかに帯域を接続したのち，逆正則化を行い，出力音声を得る．

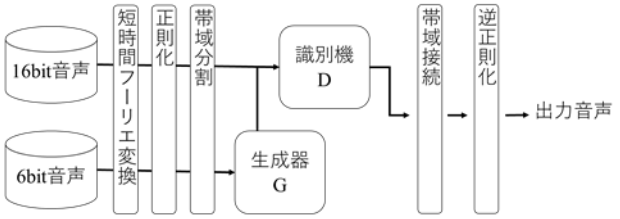

図 1. 提案システムの概要図

3. 実験

提案手法の有効性を証明するために，ATR 音声データベース 503 文を用いて音声品質の改善を行った．実験条件は表 1 のとおりである．

表 1. 実験条件

量子化ビット数	原音声:16 ビット 雑音付き:6 ビット
フレーム幅	2048 点
シフトサイズ	1024 点
正則化	L2 ノルム
帯域分割	1: $(f_1^s, f_1^e) = (1,320)$ 2: $(f_2^s, f_2^e) = (257,576)$ 3: $(f_3^s, f_3^e) = (513,832)$ 4: $(f_4^s, f_4^e) = (769,1024)$

学習の過程として生成器 G では，100 次元の全結合層を経て， 畳み込み層に入力され，BatchNormalization でバッチサイズごとに正規化する．その後，deconvolution 層へと入力され，識別機 D へ入力される．また，生成器 G および識別機 D は leaky rectified linear unit(LeakyReLU)を用いた活性化を行う．最適化は Adam で行い，バッチサイズは 5 とし，学習率は識別器 D に対しては，0.0001，β を 0.1 とし，生成器 G に対しては 0.0002，β を 0.5 とした．図 2 に実験結果を示す．

4. まとめ

実験結果より，学習させるデータの多様性がありすぎるため，それらを学習するには学習用のデータが不足していると考えられる．今後は，音素ごとのラベル付けを行い，生成器の学習を効率的に行うこと，さらに，対象のデータを増やし，ネットワーク自体の汎化性能を向上させ，さらなる高性能化を目指す．

参考文献

[1] Denoising Auto-Encoder による量子化誤差低減，信学技報，vol.116，no.378，SP2016-59，pp.57-58，2016 年 12 月

[2] 敵対的学習に基づく STFT スペクトログラムのポストフィルタリング，信学技報，vol.117，no.106，SP2017-4，pp.17-22，2017 年 6 月

Copyright © 2018 IEICE