

# Denoising Wavenet を使用した ディザ量子化信号の復号

品川 皓亮<sup>†</sup> 間野 一則<sup>†</sup>  
<sup>†</sup> 芝浦工業大学大学院 理工学研究科

## 1. はじめに

近年、様々な場面で議事録のための録音など、音声を録音し、利用する機会が増えている。録音ではPCMという符号化方式が用いられているが、量子化により波形が劣化してしまう問題点がある。とくに、会議などの場でマイクから遠くの人々の音声を聞き取れないときに、音量を増大させて聞いてしまうと、波形の量子化雑音による劣化が拡大され品質低下は避けられない。

本研究では、Denoising WaveNet にディザ量子化信号を用いた、量子化音声信号の量子化雑音低減方法を提案する。従来法[1]と比較して、Denoising WaveNet を学習することで、通常の量子化手法で復号した音声よりも、量子化後の音声の品質が量子化前の音声の品質に近づくことが期待される。

## 2. 提案手法

本研究では、録音音声はすべての信号が観測されているという条件から、図1のDenoising WaveNetモデルによる、量子化音声データからの復号手法を提案する。

学習部分では、ディザを加えた音声を8bitの場合は256次元に、6bitの場合は64bitのように指定されたbit数で変換し、one-hot-vectorに変換してWaveNetに入力として渡す。また、補助特徴量としてメルスペクトログラムを足すことで話者を特定して音声を出力させる[2]。最終的に出力音声からディザを引いたものを最終的な出力音声とする。ディザは  $\mu$ -law アルゴリズムの段階ごとの量子化ステップ幅の半分の値を足すものとする。

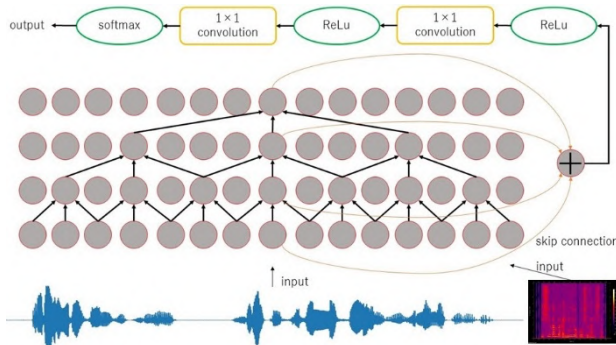


図1. Denoising WaveNet

## 3. 実験

学習データについては、話者4人の合計11,573文を用いた。テストデータについては、825文を用いて、サンプリング周波数は16kHzとした。Epoch数は十分に学習させる

ために1,000とする。評価方法は、客観的評価手法であるPESQを用いて評価した。

## 4. 実験結果

図2は、単純な量子化/逆量子化、ディザなしとありのDenoising WaveNetを使用して生成した音声のPESQの結果になっている。客観的評価手法であるPESQ(図2)では、単純な量子化/逆量子化から生成した音声より、Denoising WaveNet で生成した音声およびディザを加えて生成した音声の値が高いことから、品質は向上していることがわかる。

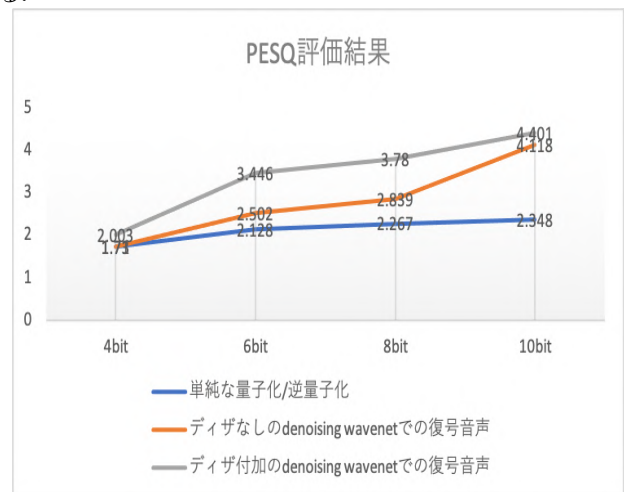


図2. PESQ 評価結果

## 5. まとめ

本研究では、Denoising WaveNet とディザを用いて、量子化雑音の低減を行った。PESQ が向上したことからDenoising WaveNet とディザを用いた手法が有効であることが示した。今後の課題としては、音声データを十分に与えられるデータセットを探すなどして学習させるデータを増やしていくことで、さらに低いbit数での量子化でもクリーンな音声を出力できるようにすることが今後の課題となる。

## 参考文献

- [1] 大内, 間野, “Denoising Auto-encoder を用いた量子化雑音低減”, 信学技報, vol.116, no.378, SP2016-59, pp. 57-58, 2016-12.
- [2] W. B. Kleijn, et al., “Wavenet Based Low Rate Speech Coding”, Proc. IEEE ICASSP-2018, pp. 676-680, 2018.