

振幅精度を改善させる回帰型 WaveNet のノイズ除去法

田中 大雅[†] 荒井 秀一[†]
[†] 東京都市大学大学院総合理工学研究科

1. はじめに

ノイズ除去は近年時間域信号を直接入力する WaveNet[1]が提案された。しかしWaveNetを用いた手法はバーストノイズが生じる。図1は教師音声と合成音声の時間域信号であり、合成音には赤枠で囲まれた振幅値にバーストノイズがある。これはWaveNetが出力を振幅値のクラスとしているため教師音声と合成音声を評価する際に誤差の有無しか判定してない故に生じる。そこで誤差の大小を考慮して学習する回帰型WaveNetを提案することで合成音のバーストノイズを減らし、ノイズ除去の品質向上を狙う。

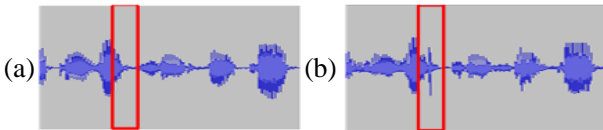


図1. (a) 教師音声 (b) 合成音声

2. WaveNet

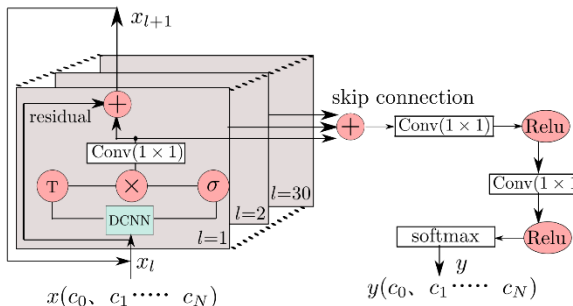


図2. WaveNet のモデル図

図2に従来手法であるWaveNetのモデル図を示す。WaveNetはDCNN(Dilated Convolution Neural Networks)を重ねることで音声波形のサンプルを畳み込む。畳み込み後はゲート付き活性化関数を通り、入力との残差をとり、skip connectionで各層から総和をとり出力層に送る。Tは tanh関数をσはシグモイド関数を表す。出力層では最後に Softmax関数を通し信号波形の事後確率を出力する。

3. 回帰型 WaveNet

図3に提案する回帰型WaveNetのモデルを示す。図3のように出力を全結合層にすることで振幅値を出力し、回帰問題に置き換えることで誤差の大小を評価する。

4. 実験

実験に用いるクリーン音声にはVoice Bank corpus[2]ノイズ音声はDEMANDデータセット[3]を用いた。学習データはSNRがそれぞれ0[dB], 5[dB], 10[dB], 15[dB], 20[dB], テストデータは2.5[dB], 5.5[dB], 7.5[dB], 12.5[dB], 17.5[dB]

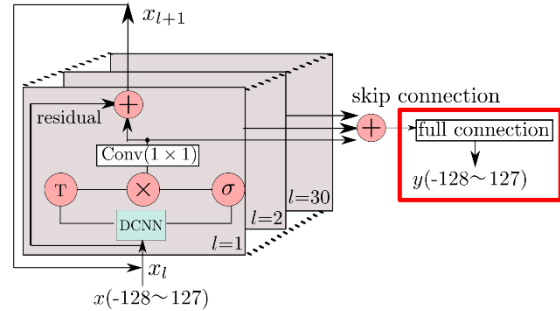


図3. 回帰型WaveNetのモデル図

ノイズをのせた。定量評価にはSNR(Signal to Noise Rate), LSD(Log Spectral Distance)を用いた。SNRは値が高い程, LSDは値が低い程評価が良い。使用したモデルのパラメーター, 実験結果を表1, 表2に示す。従来のWaveNetと比較して提案モデルのSNRは1.15pt, LSDは0.64pt向上した。

表1 使用モデルのパラメーター

層数	30
最大 Dilation	524
Residual channels	256
Skip channels	128
optimizer	Adam
学習率	0.001

表2. 各評価尺度の結果

	WaveNet	提案モデル	従来比
SNR[dB]	19.79	20.94	+1.15
LSD[dB]	5.59	4.95	+0.64

5. まとめ

従来のWaveNetと提案したモデルを2つの評価尺度で比較した。結果各評価尺度の値が向上し、提案した回帰型WaveNetの有効性を確認できた。

参考文献

[1]kaizhi Qian, et al "speech enhancement using bayesian wavenet". InterSpeech, 2017.
 [2]Christophe Veaux,"the voice bank corpus: Design, collection and data analysis of a large regional accent speech database". pp. 1-4, 2013
 [3]Joachim Thiemann, "the diverse environments multi-channel acoustic noise database: A database of multichannel environmental noise recordings". The Journal of the Acoustical Society of America, Vol. 133, No. 5, pp. 3591-3591, 2013