

日本音響学会 2022 年 秋季研究発表会に参加して

柴田 夏生

Natsuki SHIBATA

情報メディア学専攻修士課程 2年

1. はじめに

2022年9月14日から16日にかけて北海道科学大学にて開催された、日本音響学会第148回(2022年秋季)研究発表会に参加し、「敵対的生成ネットワーク(pix2pix GAN)を用いた環境雑音の抑圧に関する検討」という題目でポスター発表を行った。

2. 研究内容

2.1 背景・目的

環境雑音の混入した音声データから所望の音声成分のみを強調する雑音抑圧技術には多数の手法が提案されているが、その多くが定常的な雑音の抑圧を対象としている。しかし、一般に混入する雑音は必ずしも定常的であるとは限らず、会話や突発音などの非定常な性質を持つ雑音を無視することはできない。本研究では定常的でない特徴を持つ環境雑音を敵対的生成ネットワークの一種である pix2pix GAN を用いて抑圧し、所望の音声を強調する方法について検討する。

2.2 pix2pix GAN

敵対的生成ネットワーク(GAN: Generative Adversarial Networks)とは生成モデルの一つであり、CNNなどと同様に画像処理や画像生成などを目的として開発されたモデルである。GANは主に生成器G(Generator)と識別器D(Discriminator)の2つのネットワークから構成されており、それらを敵対的に競争させ交互に学習することで出力精度を高めている。

pix2pix GANはオリジナルのGANのアイデアを拡張し、画像のドメイン変換を実現するために提案

されたモデルである。ドメイン変換とは、線画のみの情報から実物に近い色を持った画像への変換(塗り絵)や、風景画像の昼夜を逆転させるといった変換のことである。本研究では pix2pix GAN への入力として、音声信号に高速フーリエ変換処理を施して得られるパワースペクトルのデータを用いることで、音声のドメイン変換をA(雑音が混入した音声)からB(雑音除去済み音声)への変換として考える。そして学習済みの生成器を音声Aから音声Bに変換する時間周波数マスクと捉えて処理することで雑音抑圧・音声強調に応用する(図1)。

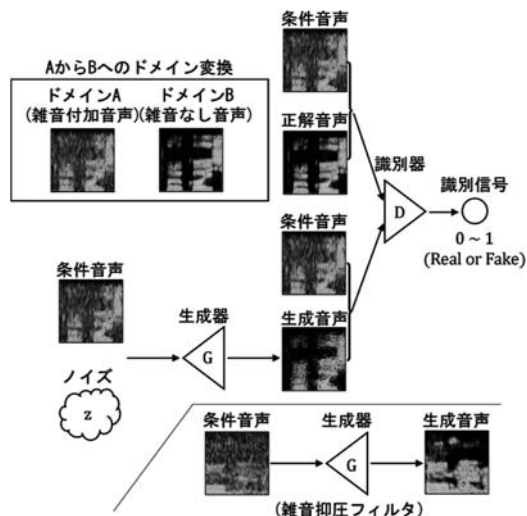


図1 pix2pix GAN 雑音抑圧処理

2.3 評価実験・結果

音声の取音時に混入する環境雑音の種類として、定常雑音に river (川のせせらぎ) と hoth (室内雑音を模擬した雑音)、準定常雑音に metro (電車の発進音+小さな会話)、非定常雑音に restaurant (屋内レストランでのざわついた食事音+会話音) と square (屋外広場での風と鳥の鳴き声+会話音) の計5種類を想定した。学習用のデータはそれぞれ100組である。処理性能の評価指標として、音声品質の歪み尺度を評価する PESQ と、音声の聞きとりやすさを評価する STOI を使用した。PESQ・STOI とともに値が大きいほど良い評価値を示す。

評価実験には SN 比が -3dB となるように各種雑音を付加した別の話者による 4 つのテストデータを採用した。雑音の種類ごとの評価値改善量の平均値を表 1 に示す。定常的な環境雑音である river 雑音と hoth 雑音については雑音が抑圧できていることがわかった。ただし river 雑音に関して、PESQ 値は平均で 0.6 ほど改善しているものの、STOI 値は若干低下する結果となった。

電車の発進音である metro 雑音は PESQ 値 STOI 値ともに処理前音声とほぼ変化がなく、準定常的な雑音でありながらも提案手法では雑音を抑圧することができなかった。モーター動作音が周期的に変化する性質を正確に学習できなかったことなどが原因であると考えられる。

restaurant 雑音は会話を含む雑音であることから、所望の音声成分と雑音である会話音声の成分を識別できず評価値の大幅な改善には至らなかった。ただし、実際に処理後の音声を試聴すると、会話成分ではないレストラン内の騒音成分は抑圧することができており、人間の会話を除く定常的な雑音成分に関しては抑圧できるものと考えられる。その例として、restaurant 雑音よりも定常成分が多く会話音が少ない square 雑音に関しては、restaurant 雑音と比

較して PESQ 値 STOI 値ともに改善量が大きいことが分かった。

3. まとめ

本研究では、機械学習の枠組みの一種の敵対的生成ネットワークの派生モデルで、画像のドメイン変換を目的として考案された pix2pix GAN を、音声信号処理に応用して環境雑音の抑圧処理に用いる手法を提案した。雑音抑圧処理前の音声と処理後の音声に対して行った客観品質評価の結果から、時間によって周波数特性に大きな変化の少ない定常的な雑音に関しては抑圧できることがわかった。

今後の課題として、metro 雑音などに代表される周期的変化の性質を持つ雑音の抑圧量向上を目指したい。本研究で用いた pix2pix GAN 単体ではなく、他の処理構造との組み合わせなどのさらなる検討が考えられる。

4. おわりに

以前にも学外発表を行った経験はあるが、自身としては初の全国規模の発表であり、外部の方々と研究内容について討論し意見交換をすることができた。数年振りの対面形式での開催のため発表方法の変更など多少の混乱もあり緊張したものの、自身の研究について可能な限りわかりやすく発表できたように思う。また、他の参加者の研究にも触れることができた。今回の発表会で得た多くの知見を今後の研究活動に活かし、さらなる学びに繋げたい。

表 1 評価実験の結果 (平均改善量)

	PESQ (0.5~4.5)	STOI (0.0~1.0)
river	0.637	-0.014
hoth	0.608	0.273
metro	0.161	0.063
restaurant	-0.001	0.036
square	0.250	0.124