

## GAN を用いた補聴器向け風雑音除去法の検討

## Study on Wind Noise Reduction Method using GAN for Hearing Aids

有元克<sup>†</sup>      中藤良久<sup>†</sup>  
Katsuya Arimoto      Yoshihisa Nakatoh

## 1. はじめに

現在、日本では補聴器装用者が 340 万人程度であり、片耳装用者はその約 6 割といわれている [1]。一方、補聴器を屋外で使用する際に聞き取りを妨げる原因の 1 つとして風雑音あげられるが、様々な風雑音除去法が提案されている。中でも、両耳補聴器の入力信号の相関性を用いて風雑音の有無を判定し周波数帯域で風雑音除去を行う手法が提案されている [2]。

そこで本研究では、片耳補聴器における風雑音除去を目的とし、GAN を用いて風雑音を推定し、スペクトルサブトラクション法 (SS 法) に基づいて時々刻々に風雑音を除去する方法を提案し、その有効性を評価する。

## 2. 提案法の概要

提案法のフローチャートを図 1 に示す。提案法は風雑音推定部と風雑音除去部に分けられる。提案法の処理は風雑音下の音声信号を入力し、風雑音推定部で GAN により風雑音を生成する。その後、風雑音除去部で元の音声信号から生成した風雑音を周波数領域で差し引くといった流れとなる。

## 2.1 風雑音推定部

本研究では、風雑音を推定する際に機械学習手法の 1 種である Generative Adversarial Networks (GAN) を用いた。GAN は学習データの実サンプルを模倣し、入力データから疑似サンプルを生成するモデルである。本研究で用いた GAN のネットワーク構造を図 2 に示す。GAN は Generator(G)・Discriminator(D) と呼ばれるニューラルネットワークのペアで構成される。G はオートエンコーダの構造をしており風雑音下の音声信号を元に学習データの風雑音と類似した風雑音の生成を目的とする。D は識別器と

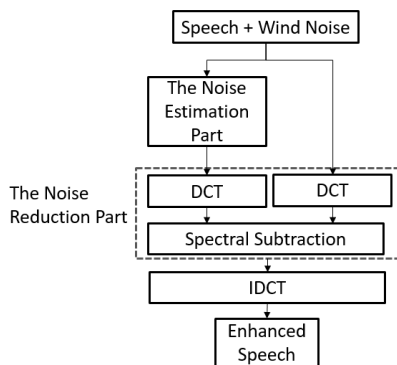


図 1 提案法のフローチャート

して動作し、入力データを学習データの風雑音と G が生成した疑似風雑音に正確に分類することを目的とする。その後、D の分類結果と実際の学習データのラベルを比較し、誤差逆伝播法により G 及び D の重みを更新することで学習を行う。G と D を競合させた学習を行うことで、G による学習データと類似した精度の高い風雑音信号の生成が可能となる。

## 2.2 風雑音除去部

本研究では風雑音除去に SS 法、周波数領域への変換手法には離散コサイン変換 (DCT) を用いた。SS 法とは周波数領域で推定雑音スペクトルを差し引く音声強調法の 1 つである。2.1 で記した G により入力信号から推定風雑音を生成する。

## 3. 提案法の有効性評価

## 3.1 実験条件

GAN の学習と評価に用いる風雑音のデータは図 3 の条件で録音する。学習データの風雑音はダミーヘッドの左耳にマイクを着用し、あらゆる角度からの風雑音を録音したものであり、図 2 の Wind Noise 部に使用する。一方、Speech + Wind Noise 部はコンピュータ上で風雑音の学習データに Noisy speech database [3] (男女 28 人 : 11572 種) の音声が付与したものを使用する。評価データの風雑音は

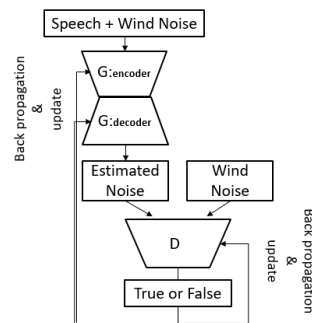


図 2 GAN のネットワーク構造

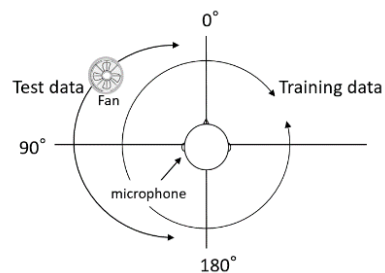


図 3 実験条件

<sup>†</sup> 九州工業大学 Kyusyu Institute of Technology

ダミーヘッドの左耳に無指向性の耳穴型補聴器(ITE)を装着し、正面の角度を  $0^\circ$  とし、左回りに  $180^\circ$  まで  $15^\circ$  刻みで録音したものに Noisy speech database の男性音声 10 種をコンピュータ上で付与したものを用いた。また、ITE 補聴器のマイクは 2 チャンネルである。使用した信号の SNR (S: 音声信号, N: 風雑音) は  $-30 \sim 15\text{dB}$  であった。また、学習は G と D を  $100 \times 420$  epoch 行い、使用した GAN のモデルは先行研究 (SEGAN: Speech Enhancement Generative Adversarial Networks<sup>[4]</sup>) と同様のものを用いた。DCT の分析条件としては、フレーム数は 256, フレームシフトは 50% オーバーラップ, 窓関数はハニング窓を使用する。また、実験の条件を表 1 に示す。

### 3.2 評価実験

先行研究 (SEGAN) として GAN に音声を学習させることで、音声信号に対する雑音除去が行えることが報告されている。そこで本稿では SEGAN で風雑音除去を行った音声と、提案法で風雑音を行った音声の STOI<sup>[5]</sup> (Short-Time Objective Intelligibility measure) による評価を行った。

評価データの ITE 補聴器における風雑音の音圧レベルと風の角度の相関図を図 4 に示す。その結果、ITE 補聴器は正面から  $75^\circ$  方向からの風の影響が最も低く、風雑音は全体的に  $74\text{dB}$  程度であることが示された。一方、SEGAN 及び提案法の評価値の平均結果をまとめ、図 5, 6 に示す。図の縦軸は理想的な音声との相関性の評価値、横軸は風の角度を示しており、実線が SEGAN, 破線が提案法の結果である。その結果、提案法による風雑音除去を行うことで、理想的な音声との相関性が 70% 以上の音声を生成できることが示された。以上より、いずれのマイク、風の角度(風雑音の音圧)においても提案法の有効性が高いことが示された。

### 3.3 考察

先行研究では音声信号を生成出力するために、GAN の学習データ内の音声の性別、言語、基本的周波数、音圧等の特徴量に大きく影響を受けるが、提案法においては、除去を行う特定の雑音を GAN の学習データとしているため、音声信号の特徴に影響を受けにくいと考えられる。以上より、提案法による風雑音除去は風の角度(風雑音の音圧)に影響を受けにくく、音声信号に影響が少ない効率的な手法と考えられる。

表 1 実験条件

Condition	Noise data for Training	Noise data for Test
Wind Speed	2~5 m/s	4 m/s
# of data	11572	13
Microphone	Roland CS-10EL	ITE mic1 ITE mic2

## 4. まとめ

本研究では、片耳補聴器における風雑音除去の検討をニューラルネットワーク(GAN)を用いて行った。ITE 補聴器において、いずれの風の角度による風雑音に対しても、精度の高い雑音除去を実現できた。今後の課題としては、音声の性別による雑音除去率の評価や補聴器に実装した際に必要とされるシステムの高速度化(リアルタイム処理)について検討する。

### 参考文献

- [1] 日本補聴器工業会, JapanTrak2015, pp17, (2015)
- [2] Phonak Insight / Speech in Wind / pp.1,(2012)
- [3] Valentini-Botinhao, Cassia. Noisy speech database for training speech enhancement algorithms and TTS models, [dataset]. University of Edinburgh. School of Informatics. Centre for Speech Technology Research (CSTR). (2016)
- [4] SEGAN : Speech Enhancement Generative Adversarial Network: Santiago Pascual, Antonio Bonafonte1, Joan Serra 1Universitat Politècnica de Catalunya, Barcelona, Spain Telefónica Research, Barcelona, Spain,(2017).
- [5] Cees H. Taal, Richard C. Hendriks, Richard Heusdens, Jesper Jensen, "Short-Time Objective Intelligibility Measure for Time-Frequency Weighted Noisy Speech", in Proc. ICASSP, pp.4214-4217, (2010)

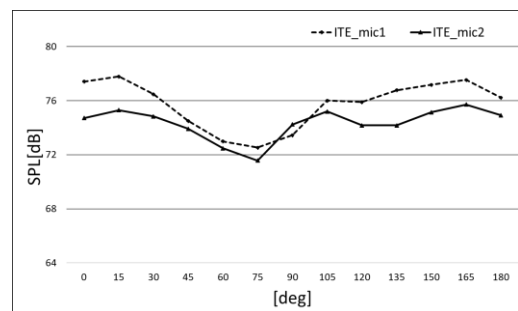


図 4 風の角度と音圧レベルの相関図

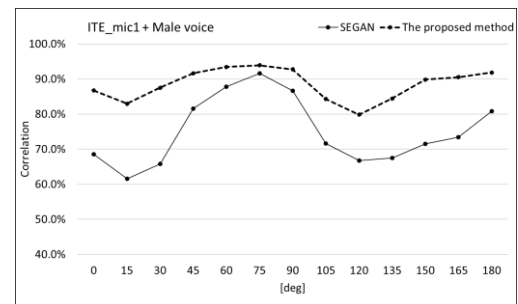


図 5 実験結果 (マイク 1)

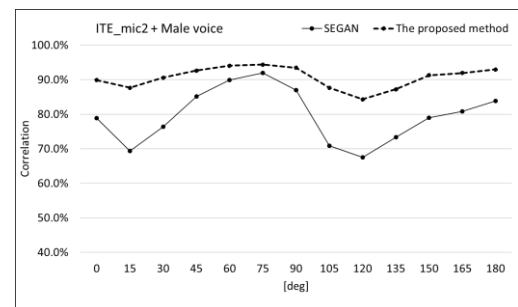


図 6 実験結果 (マイク 2)