CNN を用いたハウリング防止のための研究

2018SE015 永田龍佑
指導教員：野呂昌満

1　はじめに

デジタル化された音源を用いる場合, それに追加されてしまうノイズを抑制することは旧来からの課題であった.

「ノイズの処理」とは広義の意味であり, 処理を目的とする場合はノイズごとの種類に分けて処理を行うことが一般的である. そこで本研究では扱うノイズの種類をハウリングが日常生活で発生しやすく不快であることからハウリングに限定し、目的を「ハウリング周波数の大まかな特定」とする.

デジタルフィルタを用いたノイズ処理の既存研究では荒川らが行った研究 [3] 等が挙げられるが, ハウリングに関するものは見受けられなかった.

画像処理分野において利用されることの多い CNN を用い, 音源を周波数スペクトル波形化することで,「ある音源が低, 中, 高音域でハウリングを起こす可能性があるか, あるいはそうでないか」を判断する.

本研究ではハウリングが起こる可能性のある周波数を持つ音源の推定を CNN の画像認識技術を用いて行う手法を提案し, その有効性、妥当性を確認する.

上記の研究背景を踏まえ、本研究の研究課題を以下の二点とする.

1. ハウリング周波数帯判断のためのニューラルネットワークの設計
2.（1）を用いた手法の妥当性, 有効性の評価

また本研究を発展させることで得られると考えられる期待効果は以下の二点である.

1. ハウリングの発生頻度の抑制
2. 発生後の周波数特定の簡易化

2　関連研究, 関連技術

2.1　ハウリング（鳴音,feedback）

ハウリングとはある音に対して生じる現象である. 日常生活ではその音を耳にする機会が多く, ピー, といった耳を劈くような高音やボー, といった地鳴りのような音や, 反響音であったりとその形は様々である. 音楽界ではその現象を表現方法の一つとして利用することもあるが, 意図せず生じることが発生割合の多くを占め, 多くの人にとって不快な音であることは明白である.

2.2　ハウリングの発生原理

多くのハウリングの発生原理は「ある音をマイクが広い、その音をアンプで増幅し, それをスピーカーで出力し, それをマイクが拾い…」のようなある種のループによって生じる. 音の波形とは図 1 のような折れ線となる. そのうちのとがっている部分をピークと呼ぶが, そのピークが大きくなりすぎるとハウリングが始まり, その周波数帯によって生じる音が変化する. ハウリングを抑制は, ピークの部分の周波数帯を口述するイコライザー等によってその音量レベルを抑える方法や, 物理的なマイク等の位置を変更する方法, マイクの本数を減らす方法等が様々である.

本研究では音を伝えることを考えた場合, 物理的な解決方法であるマイク位置の変更等は万人にとって容易であると考え,「ハウリング処理は周波数に操作を加えること」を主として考える.

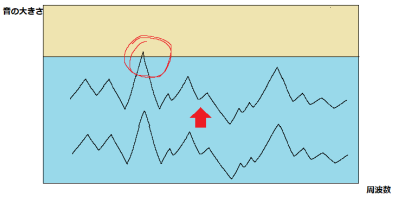

図 1　ハウリング発生のイメージ図

2.3　イコライザー（EQ）

イコライザーとは, オーディオ機器や楽器などに搭載され, 音楽制作によく用いられる特定の周波数帯域をレベルを増減することで理想とする音に近づけるための機器である. 物理的な機器としてはもちろん, コンピュータ上で用いるプラグインとしても様々な種類がある.

本研究では EQ を用いることによって低, 中, 高音域帯をブーストし, 意図的にハウリングを起こした音源と元の音源を用いることで判断を行う.

3　技術的課題の解決

3.1　ニューラルネットワークの設計

研究課題の解決のため, その中間層を決定する必要がある. 波形から特徴抽出を行う際に,「色の濃さ」,「色の場所」の二つの点が特徴抽出に重要な 2 つであると考えた. そこで VGG16 モデルの一部を参考に図 2 のような CNN を構築した. 図 2 は本研究で用いた CNN のモデル図である.

3.2　データセットの作成

CNN で教師あり学習を行う際にはカテゴリ分けされたデータセットが必要になる。表 1 に従ってカテゴリ分けを行う。表 1 のカテゴリごとに周波数帯をブーストした音源をスペクトログラム波形画像に変換する. 学習量の確保の為, 画像反転, ネガポジ反転の処理を行った画像を用意する. 図 2 に比べ図 3 の高音域帯の色がブーストされ明るくなっており, ハウリングが起こっている, または起こりうる

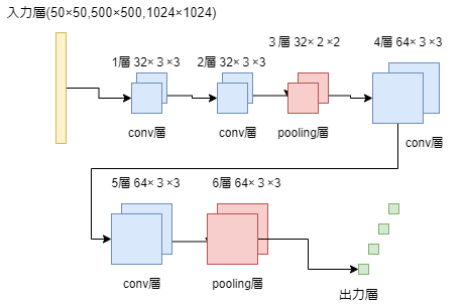

図 2　作成した CNN のモデル図

表 1　データセットのカテゴリ

カテゴリ	周波数帯
原音	
低音域	20hz — 100hz
中音域	100hz — 1khz
高音域	1khz — 5khz

状態に変化していることがわかる．図 2，3 は横軸が時間，色の明るさが音の大きさ，縦軸が周波数である．

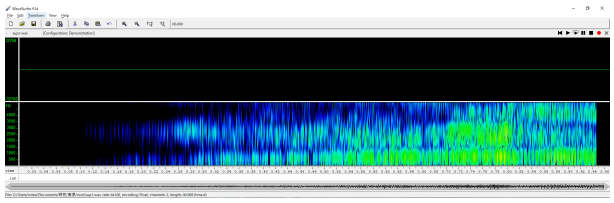

図 3　サンプル：原音

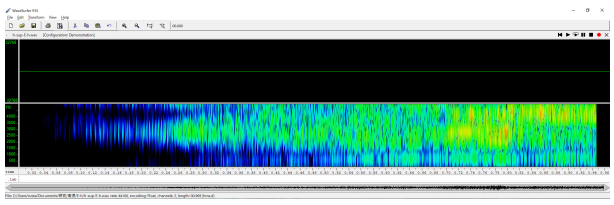

図 4　サンプル：EQ 高音域ブースト

4　実験結果

入力サイズを 50 × 50(epoch50,100),500 × 500,1024 × 1024 とし，テストデータを用い実験を行った．表 2 はその結果をまとめたものである。

表 2　テストデータを用いた際の結果

入力サイズ	loss	accuracy
50 × 50	0.837096631	0.6333333253860474
epoch100	1.0802426338	0.6555555462837219
500 × 500	0.6920004487	0.6666666865348816
1024 × 1024	0.5369381904	0.7333333492279053

表 2 より，本研究におけるスペクトログラム波形を用いて学習させた CNN は入力サイズが文字の識別レベルの入力サイズでは accuracy の値が低くなっているのがわかる．また入力サイズが小さな状態のまま epoch 数を増加させた所,accuracy の値があまり変動しておらず，本研究で用いた CNN においては epoch 数の増加は精度を上昇させづらいことがうかがえる．

5　考察

課題であった「ハウリングが生じうる，または生じている画像の特徴抽出」は，実験結果から精度こそは高いものではないが成功したといえる．また，その精度の向上には 500 × 500 を超えるまでは epoch 数よりも入力サイズの増加が有効なことが分かった．入力サイズ 1024 × 1024 は 500 × 500 の入力サイズに比べおよそ 0.08 程の精度の上昇を可能とした．その精度は約 75 %と高いとは言えない数字ではあるが，提案手法によって構築された本 CNN は音楽等に精通していない人に対して，ハウリング防止として一定の役割を果たせる余地があると考える．しかしながら本研究での周波数帯のカテゴリ分けは 3 つしかなく，その精度の妥当性には発展の余地がみられる．

6　おわりに

本研究ではある波形画像を対象として「ハウリングが起こりうるかどうか」「ハウリングが起こるとすればどのあたりの周波数帯なのか」を判断するため研究を行った．作成したデータセットを用い，自身で構築した CNN を用いた結果,CNN を用いることでピーク周波数帯の特定が容易になる可能性が示された．より周波数帯の範囲を狭め，特定の制度を向上させることが今後の課題であり．よりデータセットのカテゴリを細分化し，そのデータセットを学習させ，その有効性を調査することによって実用に足る能力を得ていくと考える．また、今後の課題は以下の二つであると考えられる．

1. 判断精度の上昇
2. 複雑な音声等のバリエーションに富んだ学習
3. 時系列データを追加したリアルタイム処理

参考文献

[1] Karen Simonyan ,Andrew Zisserman, VERY DEEP CONVOLUTIONAL NETWORK FOR LARGE-SCALE IMAGE RECOGNITION, arxiv:1409.1556v6 [cs.CV] 10 Apr 2015

[2] KTH https://www.speech.kth.se/wavesurfer/

[3] 荒川薫，松浦浩平，渡部宏明，荒川奉彦“成分分離型フィルタを用いた音声の雑音遮減法“電子情報通信学会論文誌 vol. J85-A No.10 Oct 2002.