連続的な環境音の認識に関する研究

2016 年 3 月修了　人間環境学専攻

47-146706　佐々木 逸士
指導教員　佐々木 健 教授

Recognition of non-speech environmental sound is expected to be used in security, product inspection and complement of image information. Because non-speech sound is different from the vowel-based voice, voice recognition technology cannot be applied directly. This paper describes recognition of the steady sound, such as sound of water flow, by using the method of detecting a repeat of sound mass.

Key words: Environmental Sound, Spectrogram,

1 概要

今日,一人暮らしの見守りや工業製品の非破壊検査,果てには画像の情報補完等においてまで,非音声の環境音利用の需要が高まってきている [1].人間の声を分析する音声認識はこれまでさまざまな研究がなされており,実際にスマートフォンの音声認識などさまざまな場所で活用されている.一般的に音声認識は声道特性を利用して分析を行っており,また言語の仕組みと組み合わせて実際の言葉とのマッチングを行っている.しかし環境音はその種類が多く,また特徴量となる特徴も多種多様であるためそのような音声認識の一般的な分析方法をそのまま適応してもうまくいかないことがままある.本研究では人間が環境音を認知する際にその音の揺らぎのような成分を聞き取って分類しているのではないかという観点から環境音に着目し,その成分を検出,解析する手法について解説する.

環境音でない音声認識については,Mel Frequency Cepstral Coefficient (MFCC) を用いて特徴量を抽出するのが一般的となっている.MFCC は人間の知覚特性を模したフィルタバンクを用いてケプストラム解析を行った物で,人の声道の特性を得ることに秀でている.一方で突発的であったり,声道特性の存在しない環境音には適応しにくい.一般的な音声解析では,MFCC を用いて特徴量を抽出した後に隠れマルコフモデルやニューラルネット等を用いて特徴量と言語のマッチングを行う.

音声認識技術以外にも幾つか環境音を認識する技術が研究されている.しかし既存の環境音認識技術では特定の目的に特化し,特徴的な波形に注目して解析するといった手法を取っている.既存の環境音認識技術をまとめた物が Table.1 となる.波形の自己相関関数のピーク間隔を特徴量とするものと,波形の 0 点間の距離とその間のピークを特徴量とするものである.本研究では人間が耳で聞いた際に感じる環境音の揺らぎのような特徴を利用して,環境音の特徴量を抽出できないかと考えた.

実際の環境音の特徴を観察するためには流れる水の音を用いることとした.環境音には衝突音のような過渡的な音と,そうではない風の音といったような連続的な音が存在するが,本研究で実験をするにあたってはまずスペクトル分析が行いやすい連続音を扱うことにした.

連続音の中でも水の音を用いた理由としては,身近な環境音であるため観測が容易である,ファンの音などの回転体由来の物ではないので単一の周波数成分のみが含まれる純音[4] ではなく,通常の環境音の大半と同じ複合音である,といった事が挙げられる.

2 水の流れる音を用いた解析

水の音を扱うに当たって水の音の基礎的な特徴を検出するため実験を行った.Fig.2 のように水の音を切った物を繰り返ししたサンプルを作成し,聞き比べを行うことで水の音は 0.3 秒以上の長さの繰り返しの際に水の音であると認識できるという結果を得た.また逆再生を行った所,前後にはあまり関係がない事が分かった.また,実際に人間が聞き分けている水の音の要素がどの周波数帯に存在するのかを確かめるために 100Hz 刻みでバンドパスフィルタを適応して聞き比べを行った所,1300Hz～4300Hz の範囲に水の音の要素が存在することが明らかとなった

Table.1 Existing environmental sound recognition technology

year	Feature of sound	use
2012	Distance of peak of ACF (Autocorrelation function) of soundwave[2]	Identification of a number of environmental sounds
2007	Distance of the point amplitude is 0 of soundwave and extreme value[3]	Car models discrimination

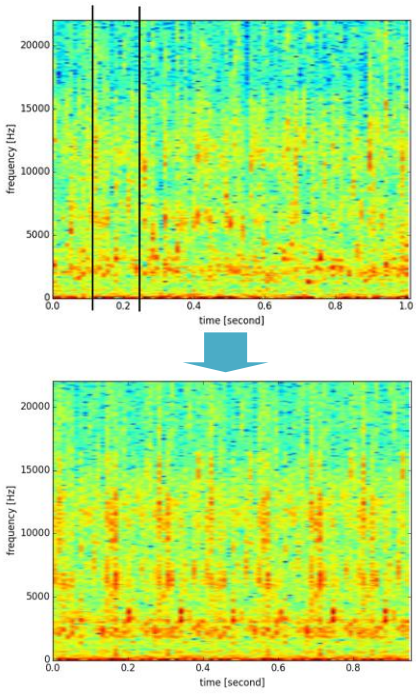

Fig2 Repeated sample

水の音のスペクトログラムを目視で観測したところ,Fig.4 のようにパワースペクトルの強い箇所を容易に観測することが出来る.そこで,水の音の揺らぎはそのようなパワースペクトルの強い箇所が周期的に表れることによって観測されていると仮説を立てた.スペクトルの強い箇所が周期的に表れることで揺らぎが表現されているのならば,ある程度の周波数幅を取ってパワースペクトルを足し合わせ,それに高速フーリエ変換を適用することで変化の周波数を捉えることが出来るはずである(Fig.3).そこで実際に FFT を行った出力例が Fig.5 となる.聞き比べによる実験で得られた水の音の特徴が存在する 1300Hz~4300Hz を更に聞こえ方に差がある 2600Hz で区切り,両周波数帯に対して解析を行った.しかしながら 1300Hz~2600Hz, 2600Hz~4300Hz 共に目立った繰り返しの周波数を観測することはできなかった.つまり水の音の揺らぎは,単純に一つの繰り返しで表現されている物ではないという結論を得た.

3 音のパワースペクトルの集合を仮定した解析

水の音の揺らぎは単純に一つの繰り返しとして抽出することはできない.そこで水の音はパワースペクトルの強い箇所の小さな固まりの繰り返しがそれぞれ独立に複数重なり合うことによって表現されている(Fig.6)と仮定を行った.その固まりを抽出するために,スペクトログラムにメディアンフィルタを適用しノイズ除去後閾値を設け二値化を行った物を,画像データとして領域抽出を用いてパワースペクトルの強い箇所をそれぞれ独立した音の固まりとして抽出を行ったものが Fig.7 となる.[5]ノイズが取り除かれ,ラベル付が行われたためグラフ上ではグラデーションのように表されている.

抽出された固まりのパラメータとして,その重心の周波数情報,時系列情報,そして固まりの大きさが挙げられる.

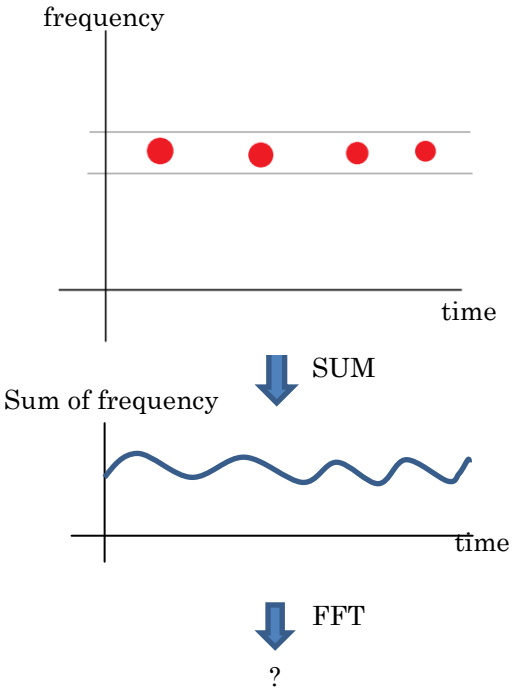

Fig.3 Sum of the power spectrum and FFT with a bandwidth

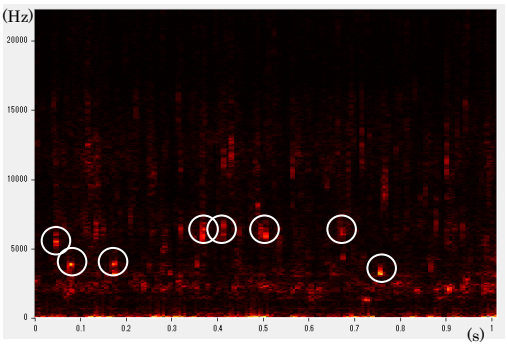

Fig.4 Strong point of the power spectrum

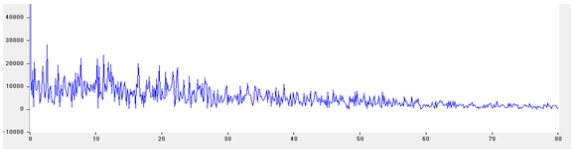

Fig.5 FFT:1300Hz~2600Hz

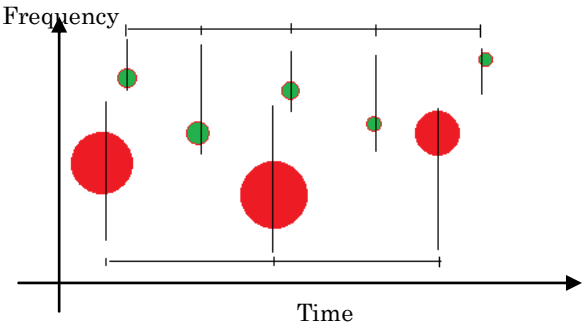

Fig.6 Assumptions of environmental sound configuration

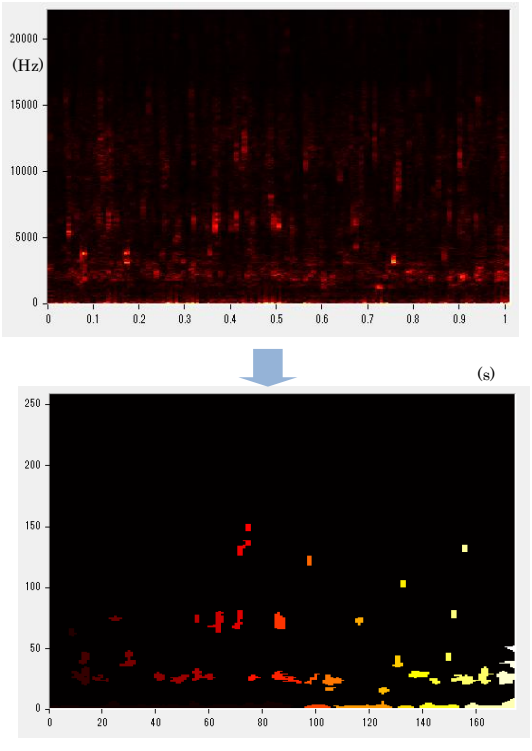

Fig7. Blob Detection on the spectrogram

抽出された音の固まりの特徴量の中で,まずは面積と重心の周波数に注目し,その二つを用いて音の分類を行う事を試みた.面積と周波数の分布をプロットしたものの一例が Fig.8 上図のようになる.そこでさらにその分布の密度が高い箇所を抜き出して Fig.8 下図のようなクラスタ分類を行った.この図を便宜上 Size-Frequency 図と呼ぶ.

更に,前項にて得られた水の音の要素は 1300Hz～4300Hz の範囲において存在しているという結果と照らし合わせその範囲において複数の水の音と水の音以外のサンプルを用いてそれぞれに関し Size-Frequency 図の作成を行った.

作成した Size-Frequency 図より,水の音に関してその Size-Frequency 図上の形状が定性的に類似して観測されたため,その形状の類似度によって水の音とそうでない物を見分けることを試みた.水の音の Size-Frequency 図の合計を取った物(Fig.9)と各サンプルについて排他的論理和を取り差分を観測することで,水の音の合計クラスタ形状との差異を検出した(Fig.10).その結果をまとめた物が Fig.11 となる.X 軸が元のサンプルサイズ,Y 軸が排他的論理和適用後のサイズとなる.この結果より水の音とそれ以外の物で Size-Frequency 図上において明らかな差がみられることがわかる.

また,クラスタ分けされた音の固まりそのものの形についても水の音とそうでないもので定性的に違いが観測された(Fig.12, Fig.13).これは水の音には過渡的な音の要素が多く含まれている事に由来すると考察でき,音の種類を区別する一つの指針とすることが出来ると考えられる.

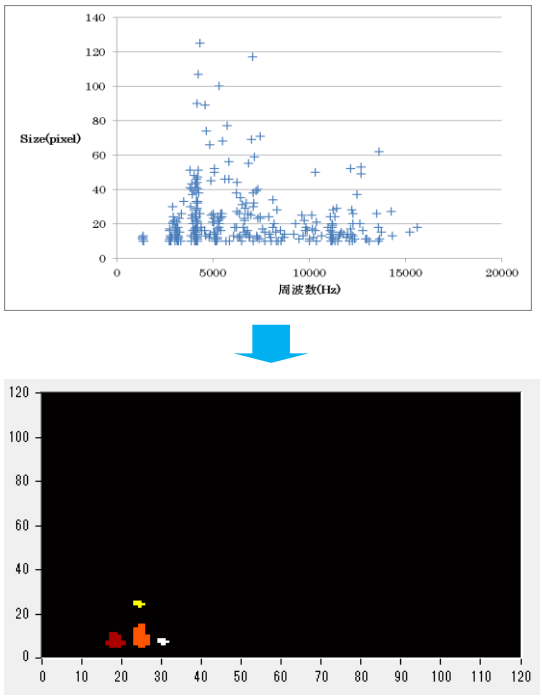

Fig.8 High density places of distribution

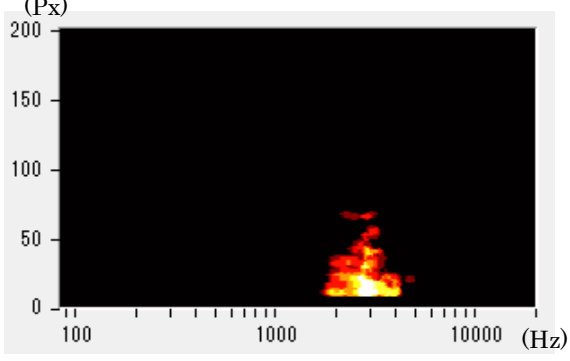

Fig.9 Sum of cluster of water

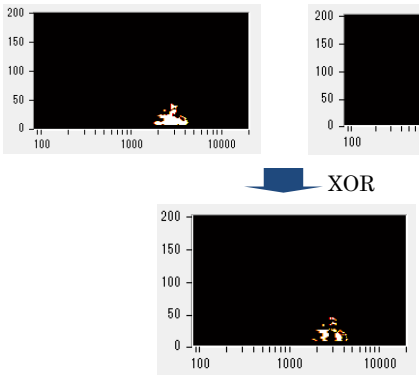

Fig.10 XOR of Size-Frequency figure

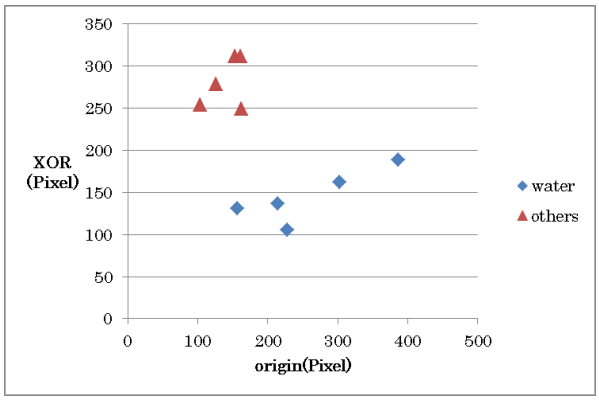

Fig.11 XOR of water sound and others

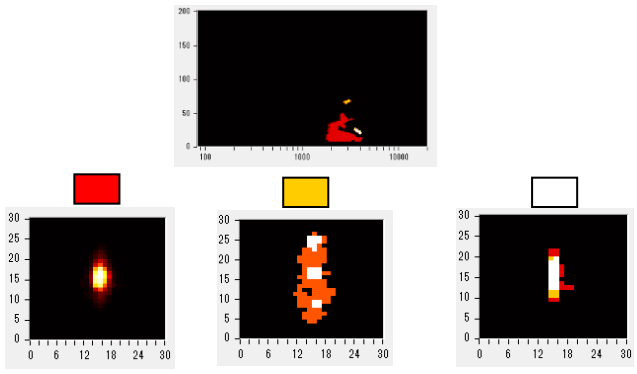

Fig.12 Mass shape of the water sound

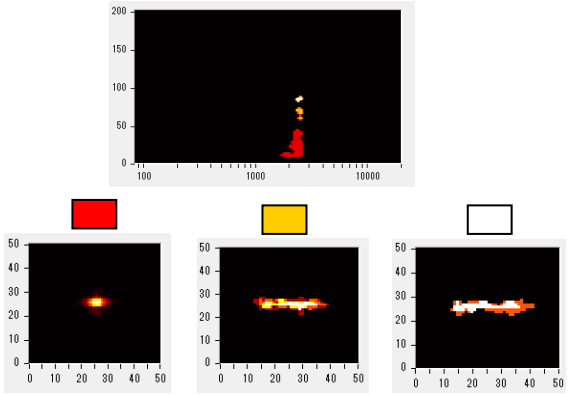

Fig.13 Mass shape of the train sound

4 音の固まりの時系列情報へ統計的処理の適用

Size-Frequency 図を用いて音の分類の一つの指針を得ることが出来たが,Size-Frequency 図は音の固まりの時系列情報を考慮していない.そこで音の時系列情報を用いて統計処理を行う事とした.Size-Frequency 図で分類されたクラスタ内での音の固まりの相対的な時間軸情報(Fig.14)を用い,その分布を求めた分布(Fig.15).さらに F 検定を用い,各サンプルの間隔分布同士が等分散であるかどうかを確かめた.その結果 Table.16 のような結果を得た.検定値が 0.05 以上のものが等分散となる.[6)]

この結果より,水同士と水-水以外の場合で明らかに分散に差が出ていることがわかる.よって音の固まり間隔の分布を用いた統計処理による結果もまた音の分類の一つの指針となることが分かる.

5 結論

人間が環境音を感じるに当たって音の揺らぎのような物を特徴として捉えているのではないかと仮設を立て,実際に水の音を観測しスペクトログラム解析を行い分析した.その結果水の音には 0.3 秒程度の繰り返しが存在し,またその特徴は 1300Hz~4300Hz の周波数帯に存在することが分かった.

水の音の揺らぎの周波数の検出を試み,FFT での解析を行った結果水の揺らぎは単純な繰り返しとて扱う事はできないと分かった.そこで水の音の揺らぎは複数の種類の音の固まりの独立した繰り返しの重ね合わせにより構成されているという仮説を立てた.そして固まりの大きさと周波数帯を用いて Size-Frequency 図を作成し,クラスタ形状で分類を行った.その結果水の音とそれ以外で得られるクラスタ形状に差がみられることが分かった.また,クラスタ分けされた音の固まりの形状にも差異を観測することが出来た.

固まりの時系列データを用いた統計解析によっても固まり同士の間隔の分散に差異を見出すことが出来ることが分かった.

これらの結果より,水の音の揺らぎは複数の種類の音の固まりの繰り返しの重ね合わせであるという仮定の元に環境音の有意な分類を行う事ができるという結論を得た.

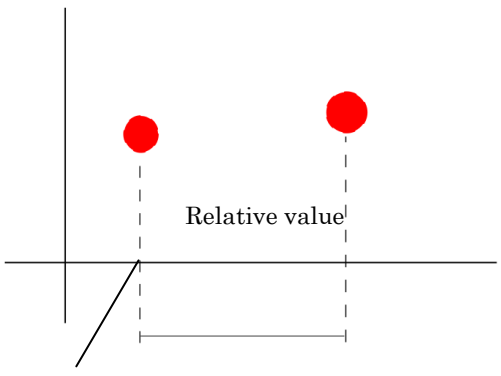

Fig.14 Choice of time-series information

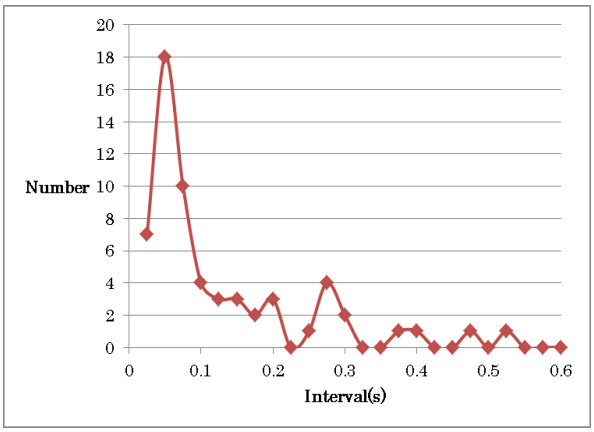

Fig.15 Interval of masses of sound

Table.16 Equal distribution survey by F-test

water_water	0.6072
water_water	0.4318
water_water	0.7820
water_other	0.0000
water_other	0.0024
water_other	0.0000
other_other	0.8132
other_other	0.0000

文献

1) Ellis, Daniel P. W., "Environmental Sound Recognition and Classification", Presented at Third Joint Workshop on Hands-free Speech Communication and Microphone Arrays, 30 May - 1 June 2011, Edinburgh, Scotland.
2) Burak Uzkent, Buket D. Barkana and Hakan Cevikalp:"NON-SPEECH ENVIRONMENTAL SOUND CLASSIFICATION USING SVMS WITH A NEW SET OF FEATURES", International Journal of Innovative Computing, Information and Control Volume 8, Number 5(B), May 2012
3) Georgios P. Mazarakis, John N. Avaritsiotis: "Vehicle classification in Sensor Networks using time-domain signal processing and Neural Networks", Microprocessors and Microsystems 31 (2007) 381-392
4) 平原達也, 蘆原郁, 小澤賢司, 宮坂榮一: "音と人間," コロナ社, 2013
5) 村上 伸一: "画像処理工学," 東京電機大学出版局, 2004
6) 涌井 貞美, 涌井 良幸: “統計解析がわかる,” 技術評論社, 2010