

スマートフォン操作中に収録した環境音に含まれるタップ音の検出・除去方式の検討

The study of the detection and elimination method of the tapping noise included in the environmental sound that recorded on using smartphones

朝田 興平

Kohei Tomoda

岡山大学 阿部研究室

Abe Laboratory, Okayama University

概要 本研究では、操作中のスマートフォンで収録された環境音信号に含まれるタップ音の検出・除去方式について検討を行った。静かな環境下ではタップ音は、環境音と比べると極めて大きな音で収録されるため除去する必要がある。本研究では収録された環境音からタップ音を検出する方法と、タップ音の検出された環境音からタップ音を除去する方法の検討を行った。

1 はじめに

環境音はその中に様々な情報を含んでいるため、その中に含まれた様々な情報を解析することで、收音した付近の様子を推測することができる。近年スマートデバイスの急速な高機能化を背景に、クラウドソーシングのアプローチを用いて、より多くの環境音を収集しようという試みがある [1]。このとき、ユーザが他のアプリケーションを実行している際にバックグラウンドで收音を行うことで収録機会を増やすことができるが、端末を操作中に環境音収録を行うと、収録した環境音に操作時のタップ音が混入する。タップ音は周囲の環境音ではないが、大きな音として記録されてしまうため、タップ音が混入した環境音をそのまま解析に用いると、収録された音量が実際より大きくなり、その後の解析で悪影響が出るおそれがある。

そこで、本研究では端末を操作しながら収録した環境音からでも正確な分析が行えるよう、端末で收音した環境音からタップ音を除去する方式を検討する。

2 提案方式の概要

提案方式では、まず入力信号からタップ音を検出する。提案方式のフローチャートを図 1 に示す。まず、図の左側の破線で囲まれた部分で、タップ音の検出に必要な識別関数と、検出したタップ音を除去するために必要なタップ音の平均スペクトルを求める。その後、図の右側の破線で囲まれた部分で、タップ音の含まれる環境音信号からタップ音の含まれる区間を検出し、その区間からタップ音を除去する。

3 実験に用いるデータと特徴量

3.1 実験に用いる特徴量

本実験では MFCC12 次元とパワーと、それら一次微分をとった Δ MFCC12 次元、 Δ パワーの合計 26 次元

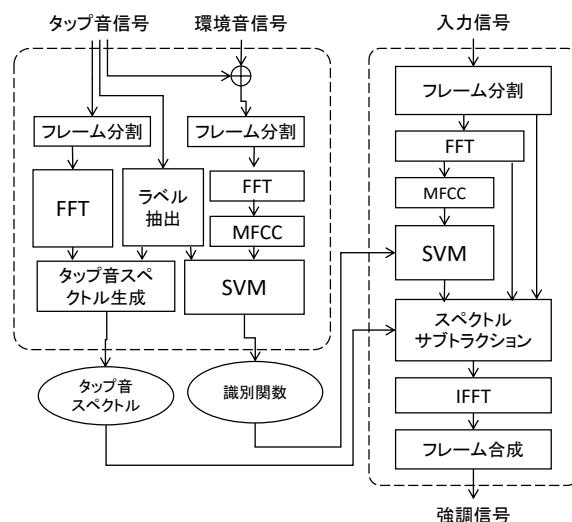


図 1: 提案方式のフローチャート

を特徴量とした。

3.2 環境音とタップ音の収録

本研究では、2015 年 6 月 29 日から 2015 年 12 月 22 日までの期間に通学や通勤の往復中に津島本町、自宅周辺、帰省中などに収録した。収録は 1 つのデータにつき 1 分を目安に収録した。

環境音収録は、端末を手に持ち、胸の高さに保持した状態で行う。収録中は画面を触ったり端末を持ち替えたりせず、ノイズが入らないように注意する。ただし徒歩移動中に収録する場合に足音や服の擦れる音などが入るのはやむをえないものとする。タップ音は環境音収録と同じデバイスを用いて防音室内で収録した。タップ音の収録は片手で持った状態で、端末と手が擦れる音が発生しないように注意しながら、毎秒 1 回を目安に画面の中央付近を親指でタップして収録する。

3.3 タップ音区間の自動抽出

SVM の学習、タップ音の平均スペクトルパターンの生成とタップ音検出器の性能評価には、タップ音区間の情報が必要であるため、タップ音区間の自動抽出を行う。まず、収録したタップ音の波形の振幅値を二乗して移動平均をとる。次に、それに閾値を設定して閾値以下のものを無音区間、閾値以上のものをタップ音区間として、フレーム単位でタップ音区間を抽出する。

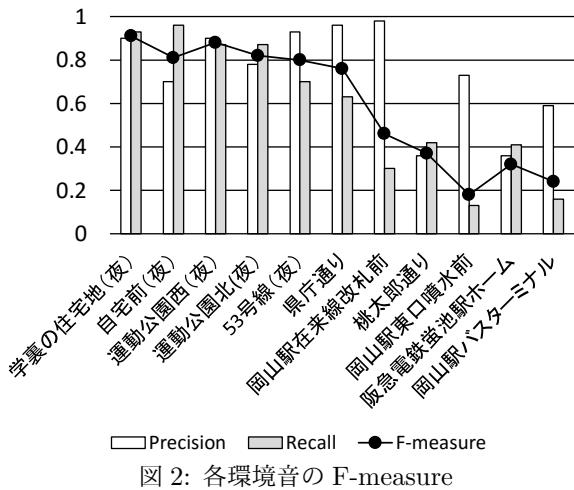


図 2: 各環境音の F-measure

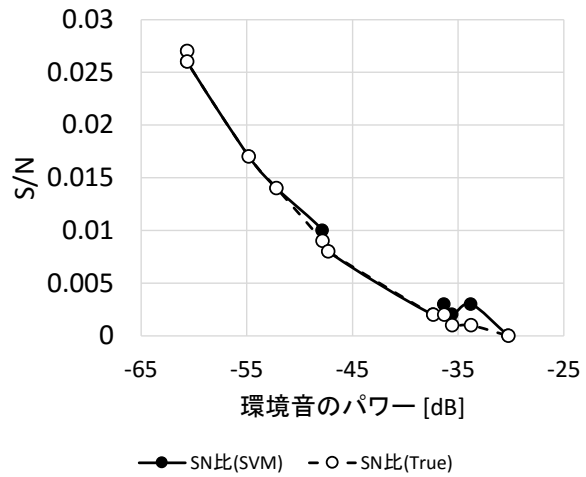


図 3: 各環境音の SN 比

4 タップ音の検出・除去実験

本稿ではタップ音の検出器を構築し、検出器から得られたタップ音区間からタップ音を除去する実験を行う。実験では 10 分割交差検証を用いてタップ音の検出性能と除去性能を評価する。

4.1 タップ音の検出器の性能評価

タップ音検出器に用いた SVM は $C = 8.85$, $\gamma = 0.05$ を用いて識別関数を求め、SVM 識別器を構築する。また、タップ音ラベルは無音区間に対してラベル数が少ないため、正しく学習できない。そこで、本実験ではタップ音ラベルが境界面を超えた場合のペナルティーは無音区間に比べて 10 倍となる重みづけを行い、データのバランスをとる。また、本実験では LIBSVM[3] を用いて識別関数を求める。SVM に入力する MFCC は AuditoryToolbox[4] を用いて求める。タップ音検出器の性能評価は F-measure によって行う。

各環境音毎に求めた F-measure を図 2 に示す。図 2 より、夜間に住宅地で収録した静かな環境音では F-measure は 0.8 を上回った。しかし、それよりも騒がしくなると F-measure は低下することが分かった。

4.2 タップ音除去フィルターの性能評価

タップ音除去の性能評価は、信号対雑音比 (SN 比) を用いて行う。本稿では、タップ音の付加された信号の電力を P_T 、タップ音除去処理後の信号の電力を P_E としたときの SN 比を式 1 のように定義する。

$$SNR = \frac{P_T - P_E}{P_T} \quad (1)$$

除去効果が高いほど数値が高くなる。各環境音の SN 比を図 3 に示す。図 3 より、正解ラベルと SVM の予測ラベルを比較して、両者の除去効果に大きな差が無いことが分かった。また、静かな環境音ではタップ音が除去されている一方で、騒がしい環境音ではほとんど除去が行われていない。静かな環境音ではタップ音

検出器が正しくタップ音を検出しているため、正解ラベルを用いて除去した場合と除去性能に差がない。このため、静かな環境音では正しく検出・除去が行われているといえる。一方で、騒がしい環境音ではタップ音検出器の性能が静かな環境音のときより低くなるにも関わらず、正解ラベルを用いて検出した場合と除去性能に大きな差はみられない。騒がしい環境音で除去効果が低くなる原因として、騒がしい環境音では環境音のパワーが大きいため、相対的にタップ音の及ぼす影響が小さくなると考えられる。

5 まとめ

本稿では、タップ音の影響を受けない環境音の分析を目的として、環境音とタップ音の収録、タップ音識別器の構築とタップ音の除去実験を行った。収録して集まった環境音とタップ音を用いてタップ音検出器を作成した。タップ音検出器は、汎化性能の高い二値分類器である SVM を用いて構築し、その検出精度の測定は 10 分割交差検証法を用いて F-measure で評価した。タップ音除去フィルターは検出された区間にスペクトルサブトラクション法を適用することで実現した。除去効果の測定は SN 比によって行い、静かな環境音ほど除去効果が高いことが分かった。

参考文献

- [1] 原他, “クラウドソーシングによる環境音収集システムを用いた予備収録実験,” 2015 年日本音響学会秋季研究発表会, pp.147-148, 3-Q-3, Sept. 2015.
- [2] C.M. ビショップ, 『パターン認識と機械学習 下』, 丸善出版株式会社, May. 2012.
- [3] C.-C. Chang and C.-J. Lin. LIBSVM : a library for support vector machines. ACM Transactions on Intelligent Systems and Technology, 2:27:1-27:27, 2011.
- [4] Slaney, Malcolm. ”Auditory toolbox.” Interval Research Corporation, Tech. Rep 10 (1998): 1998.