

音波センサを用いた低消費電力な環境音認識に関する検討

山田 蒼平¹ 前川 卓也² 原 隆浩² 尾原 和也³ 岸野 泰恵³ 村上 友規⁴ アベセカラ ヒランタ⁴

概要:近年、計算機環境や機械学習理論の発展により、音認識の研究が進歩している。音声・楽音認識の技術はスマートスピーカーなどの様々な製品の基盤技術として用いられている。また、スマートデバイスの普及やセンサの小型化に伴い、ウェアラブルデバイスを用いた環境音認識の研究が盛んにおこなわれている。環境音認識とは録音されたサウンドクリップに対応する環境音クラスを分類結果として出力するタスクであり、主に(1)環境音収集、(2)音響特徴量抽出、(3)環境音分類の3つのプロセスで構成される。(2)のプロセスでは高いサンプリング周波数で得られたサウンドクリップを周波数解析処理するため、消費電力が最も大きい。ウェアラブルデバイスを用いた環境音認識では、バッテリーによる駆動時間を延伸するため、このプロセスの消費電力を削減することが重要である。そこで、本研究では消費電力の大きい周波数解析処理を必要としない環境音認識手法を提案する。提案手法では、ある特定の周波数(共振周波数)に反応を示す低消費電力な音波センサに着目する。異なる共振周波数を持つ音波センサを組み合わせたセンサアレイを用いることで、環境音の様々な周波数成分に関する情報を得ることができる。そのため、周波数解析処理をおこなわずとも、少ない消費電力で環境音の認識が可能である。提案手法で用いるセンサアレイは低消費電力な音波センサと単純な電子回路で構成されるため消費電力が小さい。音波センサは共振周波数付近の音を受けたとき、圧電効果によって微小な電圧を発生させるが、この反応を特徴量として用いるために必要となる回路は消費電力が少ない増幅回路とコンパレータ回路のみである。本研究では、市販の超音波センサを改造することで様々な共振周波数をもつ音波センサを製作し、実環境にて環境音分類実験をおこない、その認識性能を評価した。

1. はじめに

我々が暮らす世界は人の声や楽音、さまざまな環境音で溢れている。音声によってコミュニケーションを取ったり、音楽を聴いて楽しんだり、車の音などの環境音から周囲の状況を把握するなど、人間は音から情報を得て、それを活用して生活している。計算機を用いて音から情報を得る研究は古くから行われている[1]。近年、Graphics Processing Unit (GPU) やクラウド計算基盤などの計算機環境、深層学習などの機械学習理論の発展により、音認識の研究が進歩している。特に音声・楽音認識の技術はスマートスピーカー(例: Google Home, Alexa)などの様々な製品の基盤技術として用いられている。スマートスピーカーは音声認識技術により、ハンズフリーでの音楽再生、情報検索や機器のコントロールを可能としている。また、音声・楽音以外の環境音の認識の研究も活発におこなわれており[1]、スマートフォンなどのスマートデバイスの普及やセンサの小型化に伴い、ウェアラブルデバイスを用いた環境音認識

の研究が盛んにおこなわれている。スマートデバイスを用いた環境音認識の研究例として、スマートフォンに内蔵されている加速度センサとマイクを用いた屋内行動認識システム[2]や聴覚障害者支援のためのタブレットを用いた環境音可視化システム[3]などがある。専用のウェアラブルデバイスを用いた環境音認識の研究例として、カメラ、加速度センサ、マイクを搭載するウェアラブルデバイスによる行動認識の研究[4][5]がある。

これらの研究では、ユーザの日常生活にまつわる環境音を認識することで、生活行動に関するコンテキストを推定することを目的としており、掃除機がけや皿洗いなどの行動の際に発生する環境音を対象にしていることが多い。本研究でも、同様に生活行動に関わる環境音の認識を対象とする。

環境音認識で扱われる諸問題に音響シーン分類、音響タグ付けや音響イベント検出などがあるが、本研究では音響シーン分類を取り扱う。音響シーン分類とは、入力されたサウンドクリップに対応するシーンを分類結果として出力する問題である[6]。音響シーン分類は主に(1)環境音収集、(2)音響特徴量抽出、(3)環境音分類の3つのプロセスで構成される。(1)のプロセスでは、スマートデバイス内

¹ 大阪大学工学部

² 大阪大学大学院情報科学研究科

³ NTTコミュニケーション科学基礎研究所

⁴ NTTアクセスサービスシステム研究所

蔵のマイクや測定用のマイクを用いて環境音を録音する。(2)のプロセスでは、サウンドクリップから扱う問題に適した音響特徴量を抽出する。音響シーン分類でよく用いられている音響特徴量の例として、メル周波数ケプストラム係数(MFCC: MelFrequency Cepstrum Coefficient) [7], ケプストラムや低次特徴量(LLD: Low Level Descriptor) [8]などが挙げられる。これらの特徴量の抽出において、周波数解析のために高速フーリエ変換(FFT: Fast Fourier Transformation)処理がおこなわれる。(3)のプロセスでは、混合正規分布(GMM: Gaussian Mixture Model), 隠れマルコフモデル(HMM: Hidden Markov Model) [9][10], サポートベクターマシン(SVM: support vector machine) [11]などのモデルを学習データから学習し、学習済みモデルを用いて音響シーン分類をおこなう。(2)のプロセスでは高いサンプリング周波数で得られた音データを常時FFT処理をするため、(2)のプロセスの消費電力が最も大きい。ウェアラブルデバイスを用いた環境音認識では、バッテリーによる駆動時間を延伸するため、このプロセスの消費電力を削減することが重要である。そこで、本研究では消費電力の大きいFFT処理を必要としない環境音認識手法を提案する。提案手法では、ある特定の周波数(共振周波数)に反応を示す低消費電力な音波センサに着目する。異なる共振周波数を持つ音波センサを組み合わせたセンサアレイを用いることで、環境音の周波数成分に関する情報を得ることができるため、FFT処理をおこなわずとも少ない消費電力で環境音の認識が可能であると考えられる。例えば、共振周波数がそれぞれ40k, 30k, 20kHzの3つの音波センサをアレイ状に配置する。環境音に40kHz, 20kHzの音が含まれている場合、共振周波数が40kHzと20kHzの音波センサのみが反応するため、その反応パターンから環境音に含まれる周波数成分の情報を獲得できる。獲得した周波数成分を音響特徴量として、分類器の学習と分類に用いる。

提案手法で用いるセンサアレイは低消費電力な音波センサと単純な電子回路で構成されるため消費電力が小さい。音波センサは共振周波数付近の音を受けたとき、物質に圧力を加えると電圧が発生する圧電効果によって微小な電圧を発生させる。この反応を特徴量として用いるために必要となる回路は電圧を増幅する増幅回路、反応の有無を判定するコンパレータ回路であり、これらは2つのオペアンプで構成可能な回路であるため消費電力が小さい。音波センサとこれらの回路からなるセンサアレイを用いることで、FFT処理を行わずに周波数成分に関する情報を得ることができるため、消費電力を抑えた環境音認識が可能である。

本論文の構成は以下の通りである。2節では環境音認識や音波センサに関する関連研究を紹介する。3節では消費電力の小さい環境音認識を実現する提案手法について詳しく説明する。4節では提案手法を用いて評価実験を行った結果について報告する。最後に、5節で本研究をまとめる。

2. 関連研究

近年、計算機環境や機械学習理論の発展により、環境音認識の研究が盛んにおこなわれている。環境音認識の研究ではウェアラブルデバイスを用いる研究や設置型デバイスを用いる研究が存在する。ここでは、ウェアラブルデバイスを用いる環境音認識について紹介する。本研究では環境音認識をおこなう手法として音波センサを用いた手法を提案するため、音波センサの仕組みについても紹介する。

2.1 ウェアラブルデバイスを用いた環境音認識

既存研究ではマイクとその他のセンサを組み合わせることで行動・コンテキスト認識をおこなっているものが多い。大内ら[2]はスマートフォンに内蔵されている加速度センサとマイクを用いた屋内行動認識システムを提案している。まず、加速度センサを用いて歩行、作業、静止の3つの状態の分類をおこない、作業と推定された際には、マイクを起動し、周囲の音を分析して作業内容の推定をおこなう。前川ら[4]はカメラ、加速度センサ、マイク、照度センサ、方位センサを搭載するウェアラブルデバイスによる行動認識手法を提案している[4]。前川らの研究では、照度センサや方位センサよりもカメラ、加速度センサ、マイクの3種類のセンサの方が行動認識精度の寄与が大きいことが確かめられている。同様にカメラ、加速度センサ、マイクのセンサを用いた行動・コンテキスト認識の研究も存在する[5]。各センサから得られたデータから特徴を抽出してユーザの行動やコンテキストを認識し、ユーザが特異な状況にいるときに自動的にマルチメディアクリップを収集するというシステムが提案されている。浅井ら[3]はタブレットを用いた聴覚障害者支援のための環境音可視化システムを提案している。環境音の録音と情報の表示をタブレット端末でおこない、環境音の認識処理をサーバ側でおこなうクライアントサーバシステムを採用している。聴覚障害者が日常的に利用する場所で必要となる環境音を事前にシステムに登録することで、利用者によるカスタマイズと一般の環境音認識システムより高い認識精度の実現を可能にしている。

2.2 超音波センサ

本研究では提案する手法である音波センサアレイに超音波センサを改造して用いる。そこで、まず超音波センサが音を受信する原理である圧電効果を解説する。その後、超音波センサの詳しい構造について解説する。

2.2.1 圧電効果

圧電効果(piezoelectric effect)とは特定の物質に力を加えると電荷が発生する現象で、機械電気エネルギー変換の一種である。圧電効果の種類を図1に示す。(a)のように電極に対して垂直に力・変位を加えたときに電荷が発生する現象を圧電縦効果と呼ぶ。同様に(b)のように電極

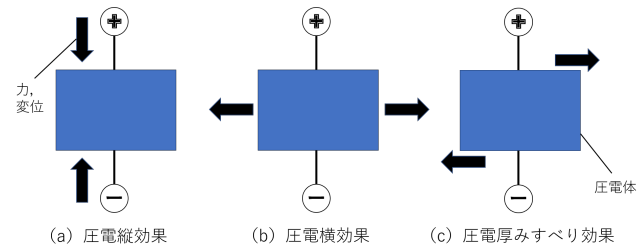


図 1: 圧電効果

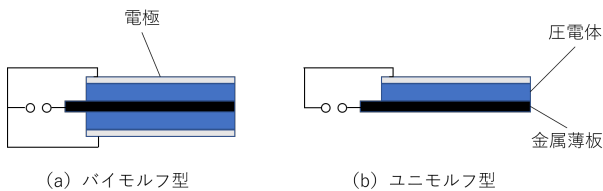


図 2: 圧電素子の形状

に対して平行に力・変位を加えた場合を圧電横効果, (c) のようにせん断方向へ力・変位を加えた場合を圧電厚みすべり効果と呼ぶ. 圧電効果は主にセンサの用途に用いられる. 電極に電圧を加えた場合に力・変位を発生する現象を逆圧電効果と呼ぶ. 逆圧電効果はアクチュエータの用途に用いられる. 圧電効果が起こる圧電体の例として, 水晶, ニオブ酸リチウム, チタン酸バリウムやジルコン酸チタン酸鉛などのセラミックスと PVDF (ポリフッ化ビニリデン) のような高分子圧電材料が挙げられる.

測定でよく用いられている圧電素子の形状を図 2 に示す. 圧電体は基本的に脆く割れやすい性質であるため金属薄板で補強される. (a) のバイモルフ型は金属薄板の両面に圧電体と電極が貼り付けられている構造である. (b) のユニモルフ型は金属薄板の片面に圧電体が貼り付けられ, 圧電体と金属薄板に電極が設けられている. ユニモルフ型では圧電体の直径が伸縮しても金属板の直径は変わらないため, たわみ変形, すなわち変位が生じる. バイモルフ型では 2 枚の圧電体が逆方向に分極されているため同じ向きの電界に対して伸縮が逆になるので, ユニモルフの 2 倍の大きさのたわみ変形が生じる [12].

圧電効果を活用したセンサには, 力センサ, 圧力センサ, 加速度センサ, 衝撃センサ, 滑り覚センサ, 角速度センサ, 形状測定用タッチプローブ, 電界強度・表面電位センサ, 化学・バイオセンサや超音波センサなどの多くのセンサがある [13].

2.2.2 超音波センサの構造

本研究で使用する開放型超音波センサ (村田製作所 MA40S) の構造を紹介する [14]. その構造を図 3 に示す. 直径はおよそ 1cm, 高さはおおよそ 1.7cm である. 共振子, 金属板, 圧電セラミックスで構成される振動子構造が電極端子につながっており, それらを樹脂ケースが覆っている. 振動子構造を図 4 に示す. 矩形型の圧電セラミック板

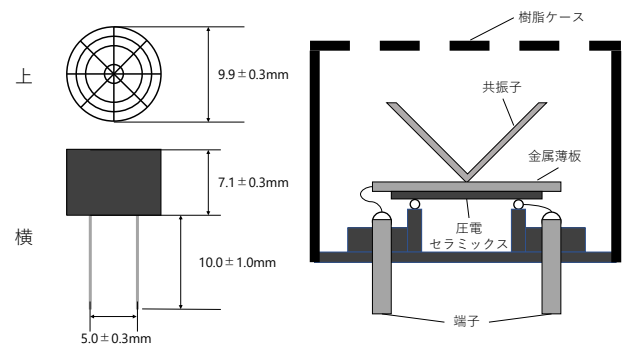


図 3: 音波センサ構造 (MA40S)

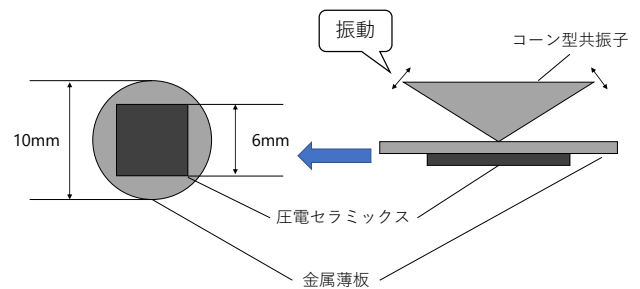


図 4: 振動子構造

に円形の金属薄板を貼り付けた図 2 のようなユニモルフ構造をとっていることがわかる. ユニモルフ構造の振動子が音波を受けて振動するとき, 節円の内側と外側では位相が 180° 異なり, 超音波の受信効率が低下してしまう. そこで図 4 の右図のように振動板の中央部に音波を受けるためのコーン型の共振子を搭載している.

3. 提案手法

3.1 想定環境

ユーザがウェアラブルデバイスを携帯しており, ユーザの周辺の環境音をデバイスが検知・認識し, その環境音のクラスを推定することを想定する. 本研究では, 特にユーザが掃除機やドライヤーなどの音の発生を伴うオブジェクトを利用している際に, その環境音のクラスを推定する.

3.2 異なる共振周波数をもつセンサアレイ製作の事前検討

提案手法で用いるセンサアレイを作成するには様々な共振周波数の音波センサを用意する必要がある. しかし, 市販されている多くの音波センサの共振周波数は不可聴帯 (超音波帯) の 40kHz に設定されている. これは, 音波センサの用途の多くが通信や距離測定などであり, 環境音に干渉されず生活の邪魔にならないような帯域の需要が主であるためと考える. そこで, 本研究では, 共振周波数が超音波帯 (40kHz) の超音波センサを加工し, 様々な共振周波数の音波センサの作成を試みた. 共振周波数は主に振動子構造の金属薄板の厚さ, 圧電セラミックスの厚さや振動



図 5: コーン型共振子と 3D プリンタで製作した共振子成形用の型

表 1: コーン型共振子の材質・厚さと共振周波数の関係

厚さ \ 金属の種類	アルミ	真鍮 (黄銅)
0.1mm	39kHz-41kHz	39kHz-41kHz
0.5mm	31kHz-34kHz	26kHz-29kHz
1.0mm	29kHz-31kHz	25kHz-27kHz

面積の大きさで決まる [14].

まず本研究では、圧電効果の根幹となる圧電セラミックスの加工を試みた。本研究で用いた超音波センサ (村田製作所の MA40S4R) の圧電セラミックスの面積は $6 \times 6\text{mm}$ の正方形である。これを削ることで $6 \times 5\text{mm}$ の長方形にしたものに 40kHz の単周波数音源を当てて、オシロスコープでセンサの出力波形を確認した。未加工の音波センサは正弦波を出力するが、加工後のものは処理が不安定な波形が出力されており、共振周波数の変更は困難であった。

次に、比較的取り外しと加工が簡単なコーン型共振子に着目した。加工前のコーン型共振子は材質がアルミで厚さは 0.1mm 程である。共振子の共振周波数はその質量や形状に依存する。そこで、この材質や厚さが異なる共振子を作成し (図 5) それらの共振周波数を調査した結果を表 1 に示す。アルミの密度は $2.7\text{g}/\text{cm}^3$ で、真鍮 (黄銅) は $8.4\text{g}/\text{cm}^3$ であるので、コーン型共振子の重さと共振周波数には関係があると考えられる。この仮説を基に、ピークが $39\text{kHz}-41\text{kHz}$ (デフォルト), $31\text{kHz}-34\text{kHz}$, $29\text{kHz}-31\text{kHz}$, $26\text{kHz}-29\text{kHz}$, $25\text{kHz}-27\text{kHz}$ となる 5 種の音波センサを作成した。なおこのピークの調査は改造した音波センサに各周波数の単周波音を 1000Hz 単位で変更して放射し、オシロスコープでその出力を確認することでおこなった。その結果を図 6 に示す。 V_{p-p} が 1 番大きい周波数から 3 番目に大きい周波数の帯域をピークとした。

3.3 手法概要

提案手法の概要を図 7 に示す。本研究では消費電力の大

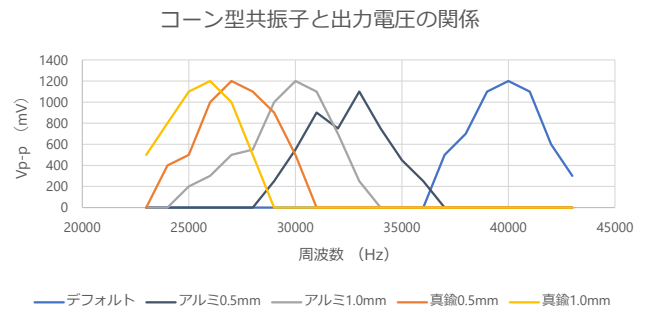


図 6: コーン型共振子と出力電圧の関係

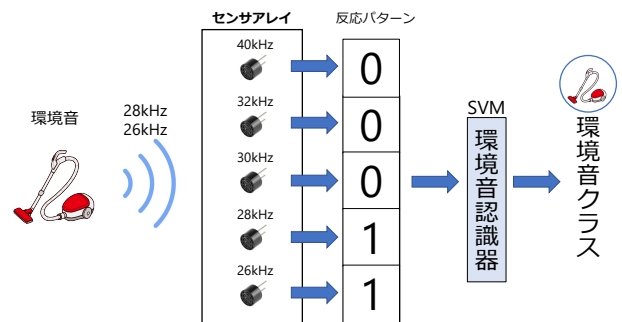


図 7: 提案手法の概要

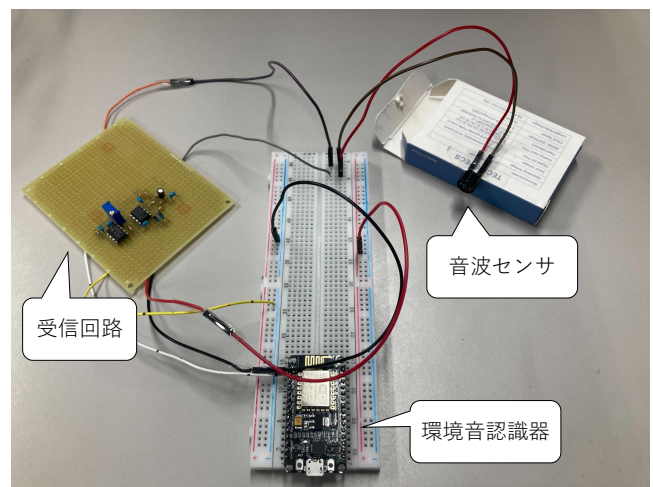


図 8: 提案手法のデバイス

きい FFT 処理をおこなわずに周波数成分に関する情報を得るために、異なる共振周波数を持つ音波センサを組み合わせたセンサアレイを用いる。センサアレイに含まれる各音波センサはそれぞれの共振周波数を含む環境音に反応する。その反応の有無を 1 または 0 で示すバイナリデータをセンサアレイが出力する。そのバイナリデータを環境音認識のための音響特徴量とし、環境音認識器で環境音のクラスを推定する。センサアレイはユニバーサル基板上に実装し、環境音認識器はマイコン (Arduino*1) 上に実装した (図 8)。

*1 <https://www.arduino.cc/>

3.4 センサアレイの構成

本研究で用いる音波センサは共振周波数付近の音を受信したとき、圧電効果により、電圧が発生する。発生する電圧はごく微小であり、環境音認識器が実装される Arduino の digitalRead 関数で読み取ることは困難である。digitalRead 関数は指定したピンの値を読み取り、2.5V を超えると 1 を出力、それ以下だと 0 を出力する関数である。そこで増幅回路を用いて音波センサが発する電圧を増幅する。増幅回路の概要を図 9 に示す。電源は 5V である。0.01 μ F のコンデンサと 3.9k Ω の抵抗で構成されるハイパスフィルタ回路、3 つの 100k Ω の抵抗と 10 μ F のタンタル電解コンデンサで構成されるバイアス回路、1k Ω 、100k Ω の抵抗とオペアンプ NJM4556A で構成される非反転増幅回路の 3 つの回路で音波センサの増幅回路は構成される。増幅率は 32db (約 40 倍) である。増幅回路を用いて、共振周波数 40kHz の音波センサが 40kHz の単周波数音を受信したときの出力波形を図 10 に示す。波形の最大から最小までの振幅である V_{p-p} が約 1.5V で、バイアス回路が 2.5V であるので、電圧の振れ幅は 1.75V から 3.25V になり、Arduino の digitalRead 関数への入力には十分な電圧となっている。しかし、音波センサが反応しているときに常に 1 を出力する必要があるが、図 10 のように出力が正弦波であるため、閾値を下回ってしまう瞬間が存在する。そこで、コンパレータ回路と RC 回路を利用して出力波形を変換する。コンパレータ回路と RC 回路を用いた変換器を図 11 に示す。この変換器の目的は RC 回路を用いて、立ち下がりにディレイをかけることで閾値を下回らないようにすることである。RC 回路を用いるためにはコンパレータ回路で正弦波を矩形波に事前に変換することが必要となる。

コンパレータ回路とは 2 つの電圧を比較するための回路である。増幅回路の出力電圧が閾値の電圧以上になるとコンパレータ回路に接続されている電源電圧の矩形波が出力される。増幅回路の出力 V_{in} を + に入力し、閾値となる比較電圧 V_c を - に入力すると、その 2 つの比較がおこなわれ、 V_{in} が V_c を上回ればコンパレータ IC, LM2903N から電源電圧 V_{cc} が出力される。 V_{in} が V_c を下回ると 0 が出力される。周波数 40kHz の正弦波が入力されると振幅 5V、周波数 40kHz の矩形波が出力される。図 11 でコンパレータ IC, LM2903N の右に位置するのが RC 回路である。RC 回路の入力である連続的な矩形波はピークから 0 まで瞬間的に変化する。RC 回路はその変化にディレイをかけることで、音波センサ反応時の電圧を digitalRead 関数の閾値より下に落ち切らないようにする。矩形波の立ち下がり時のディレイの幅は時定数 τ と呼ばれ、立ち上がりの場合は入力電圧の 63.2% の値に達する時間、立ち下がりの場合は入力電圧の 36.8% の値に落ちる時間を表す。時定数 τ は $\tau = RC$ で表せ、抵抗 R とコンデンサ C の大きさに依存する。立ち下がりの時の電圧 V_{out} は

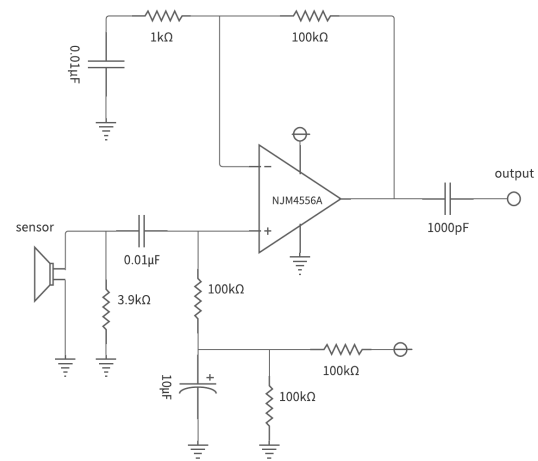


図 9: 増幅回路の概要

$$V_{out} = V_{cc}e^{-\frac{t}{\tau}} \quad (1)$$

で表せる。 t は立ち下がりからの経過時間を表す。共振周波数 40kHz の音波センサを使用する場合の抵抗 R とコンデンサ C の大きさについて考える。コンパレータの出力電圧 V_{cc} は 5V、digitalRead 関数の閾値は 2.5V、時定数は $\tau = RC$ である。これらの値を式 1 に当てはめて計算すると、

$$t = 0.693RC \quad (2)$$

となる。周波数が 40kHz であるので、ピークから次のピークが来るまでの時間は 25 μ s である。 $t_{0.5}$ は出力電圧が 2.5V に落ちるまでの時間を表すので、25 μ s の間に出力電圧が 2.5V を下回らないようにする条件は

$$t > 25 \times 10^{-6} \quad (3)$$

で表せる。式 2, 3 より、

$$RC > 36.067 \times 10^{-6}$$

となる。立ち上がりが遅いと最初の反応も遅れるので、条件を満たし、RC の値ができるだけ小さくなるような抵抗 R とコンデンサ C の組み合わせを選ぶ必要がある。

以上のように設計した回路を音波センサごとに用意し、それらの出力を Arduino の digitalRead 関数を用いて読み取り、分類器に入力するバイナリデータを作成する。

3.5 環境音認識

センサアレイの反応パターン (各時刻のバイナリ配列) を音響特徴量として、分類器の入力とする。分類器として、python のライブラリである scikit-learn で実装されている SVC (Support Vector Classifier) を使用した。ハイパーパラメータは SVC のデフォルトを使用する。

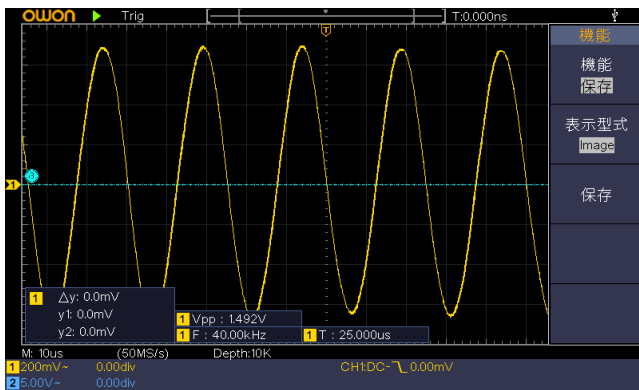


図 10: 音波センサの増幅出力波形

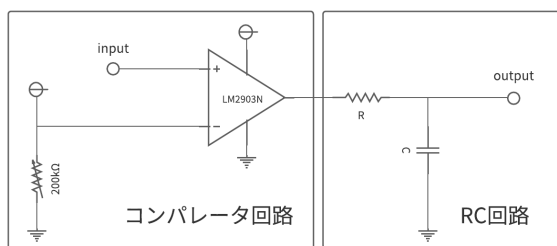


図 11: コンパレータ・RC回路からなる変換器

4. 評価実験

4.1 データ収集方法

本実験では実験を容易とするため、一般的な日常生活で使われるような5種類の機器の音をあらかじめ録音し、それをハイレゾスピーカーで再生することで環境音認識をおこなう。機器ごとに利用時の5秒間の wave データを5回録音する。録音機器であるハンディレコーダーには TASCAM の DR-07X を用いた。センサアレイは超音波帯の音に反応するため、音源として使用する音声データを録音する際、サンプリング周波数を 96kHz とした。使用した機器を以下に示す。

- シェーバー
- ドライヤー
- 電動歯ブラシ
- ハンディクリーナー
- ドリルドライバー

5 個の wave データを順に再生してセンサアレイの反応を示すバイナリデータを5系列取得する。5つの系列をまとめて1つのセッションのデータとする。上記のようなセッションを wave データとその再生する順番を変更して5セッションおこなった。ただし本稿の実験では、録音された音データに対して PC 上で FFT をおこない、疑似的なバイナリデータを再現して評価をおこなった。

4.2 評価手法

認識性能を評価するために以下の手法を用意した。

- FFT (96k) : 96kHz の録音データに対して FFT をおこない、その周波数成分を分類器 (SVC) の入力とする。この手法は単純に周波数成分を用いる分類における精度の上界を示すために用意したものである。
- FFT (20k-40k) : FFT をおこなった結果の 20kHz から 40kHz までの超音波帯の周波数成分を分類器の入力とする。この手法は超音波帯での分類精度を示すために用意したものである。
- FFT (4k) : FFT をおこなった結果の 4kHz までの周波数成分を分類器の入力とする。これはマイコン (ESP8266) でおこなう FFT の処理限界である。
- FFT (高周波数帯 バイナリデータ) : FFT をおこなった結果の 39kHz-41kHz, 31kHz-34kHz, 29kHz-31kHz, 26kHz-29kHz, 25kHz-27kHz の周波数成分の振幅と音波センサが反応する振幅を参照して設定した閾値を基に形成したバイナリデータを入力とする。予想される理想的なセンサアレイの反応データでの分類精度を確認するために用意したものである。
- FFT (低周波数帯 バイナリデータ) : FFT をおこなった結果の 4kHz-6kHz, 9kHz-11kHz, 14kHz-16kHz, 19kHz-21kHz, 24kHz-26kHz の周波数成分の振幅と閾値を基に形成したバイナリデータを入力とする。共振周波数が 5kHz, 10kHz, 15kHz, 20kHz, 25kHz の音波センサを組み合わせたセンサアレイの理想的な反応データでの分類精度を確認するために用意したものである。
- FFT (全周波数帯 バイナリデータ) : FFT をおこなった結果の周波数成分と閾値を基に形成したバイナリデータを入力とする。

5セッションのデータのうち、3セッションのデータで分類器の学習をおこない、2つのデータでテストをおこなった。

評価指標は F1-score を用いる。F1-score とは適合率 (Precision) と再現率 (Recall) の調和平均であり、2つの指標の総合的な評価指標である。

4.3 評価結果

各データの F1 スコアを表 2 に示す。また、FFT (高周波バイナリ) の混同行列を図 12 に示す。マイコンでの実行を想定した FFT (4k) では、76%の精度を達成していた。一方で、バイナリ化したデータを用いるものは、精度がそれに比べて精度が低下していた。全周波数帯を用いたものは、70%程度の高い精度を保っていたが、高周波数帯のみを用いたものは認識精度が 0 であった。一方、低周波数帯を用いたものは 60%程度の精度を達成していた。

表 2: 手法の F1-score

FFT (96k)	FFT (20k-40k)	FFT (4k)	FFT (高周波バイナリ)	FFT (低周波バイナリ)	FFT (全バイナリ)
0.796	0.561	0.759	0	0.661	0.655

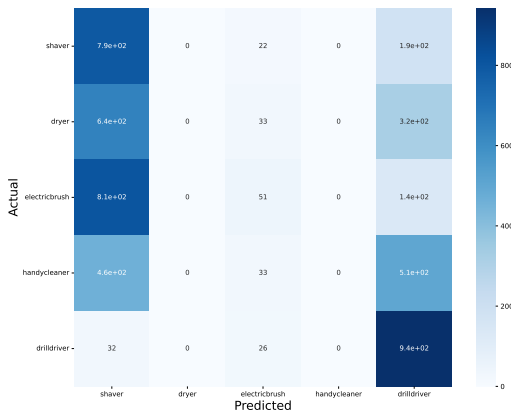


図 12: FFT (高周波バイナリ) の混同行列

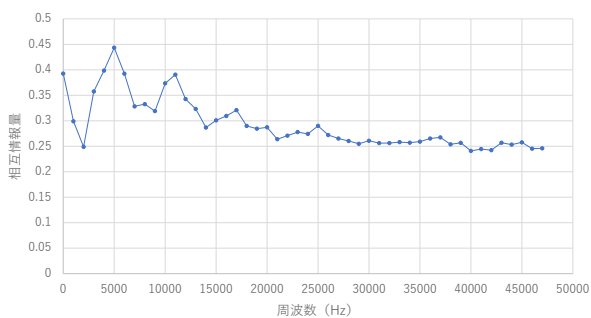


図 13: FFT (96k) の学習データの特徴量 (周波数ごとの振幅) の相互情報量

4.4 考察

分類に有効である周波数帯を確かめるため、FFT (96k) の学習データの各周波数成分の相互情報量を検証した (図 13)。20000Hz より高い周波数帯は相互情報量が 0.3 以下であるのに対し、20000Hz より低い周波数帯は相互情報量が 0.3 を超えているものが多い。特に 5000Hz は相互情報量が 0.45 程度あった。表 2 より、FFT (20k-40k) の F1-score が 0.561 であるのに対し、FFT (4k) の F1-score は 0.759 と低周波数帯のデータのほうを用いたほうが分類精度が高いことがわかる。よって、今回の実験で使用した家電機器の分類では低周波数帯の情報が重要であることがわかる。

FFT (バイナリデータ) の F1-score が 0 となったのは分類器がドライヤーとハンディクリーナーを誤ってシェーバー、ドリルドライバーと予測したためである。これは図 14 より、シェーバーとドリルドライバーが他の機器と比べて高周波数帯の振幅が大きいためだと考えられる。

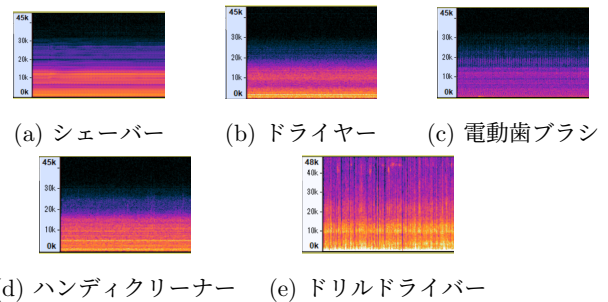


図 14: 実験で使用した機器が発する音のスペクトログラム

5. おわりに

本研究では消費電力の大きい周波数解析処理を必要としない環境音認識手法を提案した。提案手法では、ある特定の周波数 (共振周波数) に反応を示す低消費電力な音波センサに着目し、異なる共振周波数を持つ音波センサを組み合わせたセンサアレイを用いることで、環境音の様々な周波数成分に関する情報を獲得する。そのため、周波数解析処理をおこなわずとも、少ない消費電力で環境音の認識が可能である。本研究では、市販の超音波センサを改造することで様々な共振周波数をもつ音波センサを製作し、実環境にて収集した音響データを用いて評価検証実験を行った。

謝辞

本研究の一部は科研費 JP16H06539, JP17H04679, JP21H03428, 21H05299, JP21K19769 の支援によるものです。

参考文献

- [1] 井本桂右: 音響イベントと音響シーンの分析, 日本音響学会誌, Vol. 74, No. 4, pp. 198-207 (2018).
- [2] 大内一成: ウェアラブル/ユビキタス環境におけるパターン計測技術, 計測と制御: journal of the Society of Instrument and Control Engineers, pp. 579-585 (2014).
- [3] 浅井研哉, 綱川隆司, 西田昌史, 西村雅史: 聴覚障害者支援のための環境音可視化システムの開発, 技術報告 5, 静岡大学大学院総合科学技術研究科情報学専攻 (2019).
- [4] Maekawa, T., Yanagisawa, Y., Kishino, Y., Ishiguro, K., Kamei, K., Sakurai, Y. and Okadome, T.: Object-Based Activity Recognition with Heterogeneous Sensors on Wrist, *8th International Conference, Pervasive 2010 on Pervasive Computing*, pp. 246-264 (2010).
- [5] Blum, M., Pentland, A. and Troster, G.: InSense: Interest-Based Life Logging, *IEEE MultiMedia*, Vol. 13, No. 4, pp. 40-48 (2006).
- [6] 井本桂右: 環境音分析の研究動向, 日本音響学会誌, Vol. 75, No. 9, pp. 512-518 (2019).
- [7] Cowling, M. A.: Non-Speech Environmental Sound Classification System for Autonomous Surveillance, *Bachelor of Information Technology(Honours)* (2004).
- [8] Hyung - GookKim, Moreau, N., Sikora, T.: MPEG-7 Audio and Beyond: Audio Content Indexing and Retrieval, *MPEG-7 Audio and Beyond: Audio Content*

Indexing and Retrieval (2005).

- [9] Habshush, A., Rahman, A. S. and Sang, C.: Detection and Classification of Acoustic Scenes and Events, *IEEE Transactions on Multimedia*, Vol. 17, No. 10 (2015).
- [10] Eronen, A., Peltonen, V., Tuomi, J., Klapuri, A., Fagerlund, S., Sorsa, T., Lorho, G. and Huopaniemi, J.: Audio-based context recognition, *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 14, No. 1, pp. 321–329 (online), DOI: 10.1109/TSA.2005.854103 (2006).
- [11] Geiger, J. T., Schuller, B. and Rigoll, G.: Large-scale audio feature extraction and SVM for acoustic scene classification, *2013 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 1–4 (online), DOI: 10.1109/WASPAA.2013.6701857 (2013).
- [12] 大賀寿郎: 圧電材料を用いた音響部品のバラエティ, 電子情報通信学会基礎・境界ソサイエティ Fundamentals Review, Vol. 1, No. 4, pp. 46–61 (2008).
- [13] 古谷克司: 圧電効果を用いたセンサ, 計測と制御, Vol. 45, No. 4, pp. 296–301 (2006).
- [14] 浅田隆昭: 空中超音波トランスデューサの概要, 日本音響学会誌, Vol. 76, No. 5, pp. 271–278 (2020).