

FPGA を用いた低消費電力な防犯システムの実装に向けた 窓ガラスの割れる音の認識手法

A Method for Recognizing a Breaking Sound of a Window Glass for Implementing a Low-power Security System Using FPGA

寺藤 凌[†] 越智 裕之[†]
Ryo Terafuji Hiroyuki Ochi

1. はじめに

近年の犯罪や事件事故の増加に伴い、無人の防犯システムの需要が高まっている。防犯システムはリアルタイム処理が必要であり、かつ長時間連続して稼働させる必要があり、消費電力が大きいシステムでは安定した電源供給が前提となる。我々は、設置場所の制約や、停電時および災害時などにも適用できることを目指し、電池駆動で長時間稼働できる防犯システムの実現を目指している。

防犯システムには画像を用いたものと音声を用いたものがあるが、後者は前者に比べ、(1) 消費電力が比較的少ないことや、(2) 夜間など真っ暗な環境にも適用可能であることなどの利点がある。本研究では、音声を用いた低消費電力な防犯システムについて検討する。

これまでにも音声認識システムの計算量を削減し計算速度をあげるためのアルゴリズムの改良や新たな手法の提案、認識率の向上のための雑音除去手法の提案がされている [1][2][3] が、本研究では消費電力を大幅に削減するべく、特定の音（例えば窓ガラスの割れる音）の認識に特化したハードウェアの実現を検討する。本論文ではまず、窓ガラスの割れる音に固有の振幅や周波数の時間変化を明らかにし、次にウェーブレット変換 [4] を用いて解析された入力音声の情報を用いて窓ガラスの割れる音を識別するために線形 SVM を用いた学習が有効であることを示す。最後に提案アルゴリズムで最も計算量が多いウェーブレット変換を FPGA 上に実装する方法について検討し、FPGA 上の ROM を活用してこの計算量を大幅に削減し、低消費電力化を達成する手法を提案する。

以下、2 章でウェーブレット変換の結果から捉えた窓ガラスの割れる音の特徴を述べ、3 章で提案システムの構成を述べる。4 章では提案した窓ガラスの割れる音を認識するシステムを適用した実験結果を述べる。5 章で提案システムの FPGA への実装可能性を検討し、最後に 6 章で結論と今後の課題を述べる。

2. 窓ガラスの割れる音の特徴解析

窓ガラスの割れる音を認識するシステムを開発するにあたり、まず窓ガラスの割れる音の特徴を明らかにする

必要がある。本章では、窓ガラスの割れる音の時系列変化を解析して得られた特徴を述べ、認識手法を提案する。

2.1 窓ガラスの割れる音の時系列変化

図 1 は、ガラスが割れ始めてから 1.5s の期間の振幅 (図 1 上) および周波数 (図 1 下) の時系列変化である。これらのグラフは、ガラスの割れ始め 1.5s を 0.1s 毎に区切り、その 0.1s の区間での振幅および周波数の平均値をプロットしたものである。周波数の平均値は、入力された音の 17 種類の周波数成分それぞれの振幅をウェーブレット変換を用いて求め、これらの重み付き平均として得ている。グラフ中の緑の線は、鳴り始めから最大値までの最小二乗法による近似直線であり、赤の線は、最大値から鳴り終わりまでの最小二乗法による近似直線である。

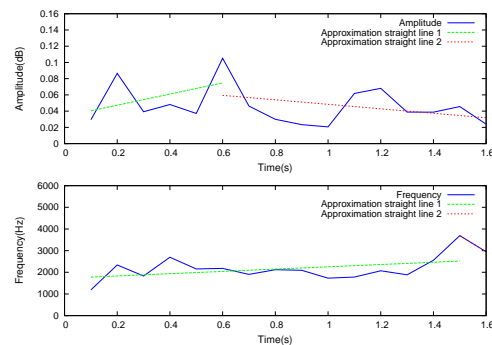


図 1: ウェーブレット変換での窓ガラスの割れる音の解析結果

図 1 より、窓ガラスの割れる音には以下のような特徴があることがわかる。

- 特徴 1: 振幅は徐々に小さくなる。
- 特徴 2: 周波数は徐々に高くなる。
- 特徴 3: 振幅の最大値の時刻が周波数の最大値の時刻よりも早い。
- 特徴 4: 周波数の平均値及び最大値が大きい。

[†] 立命館大学, Ritsumeikan University

上に挙げた特徴がみられる原因は、以下のように定性的に説明できる。

特徴1は、ガラスに与えられた衝撃エネルギーが時間の経過とともに減衰することによるものと考えられる。但し、この性質はガラスが割れる場合以外でも比較的よくみられる性質であり、これだけで窓ガラスの割れる音であることを特定することはできない。

特徴2は、窓ガラスのような1枚の大きなガラス板が割れた場合、破片の平均的な大きさは時間の経過と共に小さくなり、これに伴い、破片の共振する周波数が低い周波数から高い周波数へと推移していくためであると考えられる。特徴3は、特徴1および特徴2の2つを時間変化の観点で捉えたものである。すなわち、特徴1および2で述べたように振幅は徐々に小さく周波数は徐々に高くなるならば、振幅が最大となる時刻は周波数が最大となる時刻よりも早くなるはずである。

特徴4は、ガラスの材質に由来する性質である。経験的にわかるように、ガラスは割れると比較的高い周波数の音を発する。

2.2 ガラスの風鈴の鳴る音の特徴

次に、認識対象と非認識対象を識別するために認識対象である大きな窓ガラスの割れる音以外の音の特徴を捉える。例として、同材質であるガラスの風鈴の鳴る音を解析し、得られた振幅及び周波数の結果を図2に示す。

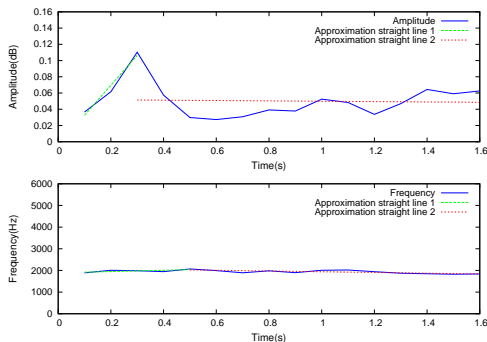


図2: ウェーブレット変換でのガラスの風鈴の鳴る音の解析結果

図2よりガラスの風鈴の鳴る音には以下の特徴があることがわかる。

- 特徴: 周波数がほぼ一定である。

風鈴は特定の周波数に共鳴するように作られているため、窓ガラスが割れるときのように周波数が徐々に高くなるということはない。そのため、2.1節の特徴2から識別が可能である。

3. 提案システムの構成

提案システムのフローチャートを図3に示す。

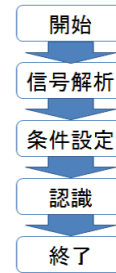


図3: 提案システムのフローチャート

信号解析部では、入力信号をウェーブレット変換で解析する。ここでの入力信号は wav ファイル (44,100Hz, 16bit) の音とする。今回ウェーブレット変換で用いるマザーウェーブレットは、周波数情報と時間情報を得ることができる Gabor 関数を使用する。また、マザーウェーブレットは 1,119Hz を 2 倍, 3 倍としていき, 17 倍の 19,023Hz までの計 17 個の周波数の波を用いて周波数解析を行う。解析範囲は音の鳴り始めから 0.1s を 1 つの範囲として、範囲内の振幅及び周波数の平均値を 1 範囲の値とし, 1.5s まで解析を行う。

条件設定部では、信号解析部で解析して捉えた特徴を取り出して認識部に与える。次の認識部では、認識対象である窓ガラスの割れる音と非認識対象である窓ガラスの割れる音以外の日常音を識別する。その識別には線形 SVM のツールである LIBLINEAR [5] を使用する。識別に使用する特徴量として 2 章で述べた窓ガラスの割れる音の特徴を含む以下のものを使用する。

- 特徴1: 最大周波数
- 特徴2: 振幅と周波数の最大時の時刻の除算。
- 特徴3: 振幅の最大時までの傾き
- 特徴4: 振幅の最大時からの傾き
- 特徴5: 周波数の最大時までの傾き
- 特徴6: 周波数の最大時からの傾き
- 特徴7: 最大周波数までの周波数の上昇値
- 特徴8: 平均周波数

特徴2は、2.1節の特徴3が満たされているかの判断をするためのものであり、満たしている場合は値が1よりも小さくなり、満たしていない場合は値が1よりも大きくなる。特徴3~6は、2.1節の特徴1および2が満た

されているかの判断をするためのものであり、近似直線が上昇傾向である場合は値が正の値、下降傾向である場合は値が負の値となる。

この認識システムの中で信号解析部のウェーブレット変換の計算量が最も多く、そのまま FPGA に実装した場合の回路規模やクロック周波数、消費電力が大きくなる。これについては 5 章で検討する。

4. 認識実験

4.1 実験方法

認識実験で用いる音声データは、サンプリング周波数 44,100Hz、量子化 16bit の wav ファイルを用いる [6], [7]。認識区間は、音声が鳴り始めてから 1.5s の間とする。ただし、最大音量の 10 分の 1 以下の音量となった場合はその時点で音が鳴り終わったものとみなす。これは、音が鳴り終わった後のノイズが識別に影響を与えないようにするためである。

4.2 パラメータ設定のための機械学習

実験に先立ち、LIBLINEAR を用いた識別のための学習を行う必要がある。線形 SVM の学習は、窓ガラスが割れる音とそれ以外の音をそれぞれ 45 ファイル、55 ファイルの計 100 ファイル分のデータを用いて行う。

4.3 認識実験結果

窓ガラスの割れる音 30 サンプル、日常音 70 サンプルを識別した結果を表 1 に示す。

表 1: 認識実験結果

入力 \ 結果	窓ガラスの割れる音	日常音
窓ガラスの割れる音	28/30 = 93.3 %	2/30 = 6.7 %
日常音	8/70 = 11.4 %	62/70 = 88.6 %

全体の認識率として、 $(28 + 62)/100 = 90\%$ を達成した。

5. FPGA への実装可能性の検討

提案手法では入力信号を解析するためにウェーブレット変換を用いるが、様々な周波数のマザーウェーブレットを用意し、入力信号との畳み込み乗算を行う必要がある。このため計算量が非常に大きくなり、通常のプロセッサでリアルタイム処理を行うと消費電力が大きくなってしまふ。そのため、この信号解析部を FPGA に実装することで低消費電力化を図る。その場合でも、マザーウェーブレットを求めるためには平方根、三角関数、指数関数などが必要となり、これをそのまま FPGA に実装しても回路規模が大きくなってしまい、低消費電力化は望めない。

提案手法におけるウェーブレット変換で使用するマザーウェーブレットの周波数は予め決まっており、入力データに依存しない。このことに着目し、本手法では、マザーウェーブレットを事前に計算しておき、FPGA 内の ROM に格納して参照する方法を用いる。これにより、必要な演算は ROM から読み出されたマザーウェーブレットと入力信号との畳み込み乗算を行うための乗算器と加算器だけとなり、大幅な計算量の削減が可能となる。

上に述べた方法でウェーブレット変換を FPGA に実装した場合に必要なハードウェア量及びリアルタイム動作に必要なクロック周波数を表 2 に示す。

表 2: 1 区間のウェーブレット解析ハードウェアを提案する方法で FPGA に実装した場合の諸元

乗算回数	約 120 万回
ROM の語数	約 6,000 語
リアルタイム処理のためのクロック周波数の下限 (乗算器 17 個のとき)	約 3.3MHz

表 2 の見積もりについて説明する。まず、各周波数のマザーウェーブレットが解析するサンプル数を表 3 に示す。例えば、1,119Hz の場合は 867 サンプルであり、マザーウェーブレットが解析するサンプル番号を m 、解析の開始位置を x とすると、マザーウェーブレットは 0 を中心とする正負両方の範囲に渡って解析するので、解析する範囲は $\max(0, -867 + x) \leq m \leq \min(867 + x, 4, 410)$ となる。ここで 4,410 は 1 区間 (0.1 秒) のサンプル数である。また、1 回の解析につき実数部と虚数部でそれぞれ 1 回の乗算が行われるため、解析サンプル数の 2 倍が乗算が行われる。このようにして、各周波数のマザーウェーブレットで 1 区間を解析するために何回の乗算を行うか求めたものを表 4 に示す。全周波数の乗算の回数の合計は約 120 万回となった。

1 個の乗算器ハードウェアでリアルタイム処理する場合、0.1 秒間の区間の解析につき約 120 万回の乗算が必要であることから、乗算器を約 12MHz で動作させればよいことがわかる。また、17 個の周波数の解析は独立しているため、17 個の乗算器を用いれば、並列処理が可能である。この場合、最も乗算回数が多い 1,119Hz の解析のための乗算回数が 0.1 秒につき約 33 万回であることから、乗算器を約 3.3MHz で動作させればリアルタイム処理が可能である。

ROM の語数は、各周波数のマザーウェーブレットが最初に解析を行うサンプル数をすべて足した数と等しいため、 $2,974 + 1 \times 17 = 2,991$ となる。実数部と虚数部の計算でそれぞれ 1 つのテーブルがいることを考慮すると、ROM の語数は $2,991 \times 2 = 5,982$ となる。

表 3: 各周波数に適したマザーウェーブレットが解析するサンプル数

番号	解析周波数 [Hz]	解析サンプル数
1	1,119	867
2	2,238	433
3	3,357	289
4	4,476	216
5	5,595	173
6	6,714	144
7	7,833	123
8	8,952	108
9	10,071	96
10	11,190	86
11	12,309	78
12	13,428	72
13	14,547	66
14	15,666	61
15	16,785	57
16	17,904	54
17	19,023	51
		計 2,974

この推定結果から、市販の FPGA に十分に実装可能であり、リアルタイム処理が低周波数で可能であるといえる。

6. おわりに

本稿では、電池駆動で長時間の稼働が可能となる防犯システムの実現を目的とし、窓ガラスの割れる音に特化した音声認識システムを提案した。提案システムはウェーブレット変換で抽出された入力音の音量や周波数の変化などの特徴量を線形 SVM で識別するものであり、シミュレーションにおいて認識率 90 % 以上を達成した。また、提案システム中で最も計算量が多いウェーブレット変換の実装については、事前計算したマザーウェーブレットを FPGA 上の ROM に格納して参照することにより、約 6 千語の ROM と 17 個の乗算器があれば 3.3MHz の動作周波数でリアルタイム処理が可能となることを示した。今後は実機での認識実験を行い、認識率や消費電力の評価を行いたい。

参考文献

- [1] 吉沢真吾, 和田直哉, 早坂 昇, 宮永喜一, “RSF/DRA 手法を用いた雑音ロバスト音声認識システムのハードウェア実現,” 信学技報, pp.127-132, 2003.
- [2] 中山仁史, 重川直紀, 横内孝史, “瞬時処理・瞬時認識のためのハードウェア音声認識システム,” 通信技報. EA, 応用音響, vol.110, no.331, pp.13-18, 2010.

表 4: 各周波数に適したマザーウェーブレットが行う乗算回数

番号	周波数 [Hz]	乗算回数 [回]
1	1,119	330,232
2	2,238	173,972
3	3,357	118,172
4	4,476	89,122
5	5,595	71,776
6	6,714	59,966
7	7,833	51,388
8	8,952	45,208
9	10,071	40,264
10	11,190	36,144
11	12,309	32,826
12	13,428	30,330
13	14,547	27,834
14	15,666	25,754
15	16,785	24,090
16	17,904	22,842
17	19,023	21,594
		計 1,201,514

- [3] 松野裕之, 友利記昌, 宮崎崇, 西村英樹, 神戸尚志, “音声認識システムのハードウェア化の一手法: HMM 出力確率計算のハードウェア化 (ソフトコア・ハードウェア化組込技術とネットワークに関するワークショップ),” 信学技報. CPSY, コンピュータシステム, vol.104, no.737, pp.79-84, 2005.
- [4] C. Torrence and G.P. Compo, “A practical guide to wavelet analysis,” Bulletin of the American Meteorological society, vol.79, no.1, pp.61-78, 1998.
- [5] M.L.G. at National Taiwan University, “LIBLINEAR – a library for large linear classification,” <http://www.csie.ntu.edu.tw/~cjlin/liblinear/>, Nov. 2014.
- [6] キングレコード, “効果音大全集 2 時計/時報/破壊/衝撃,” March 1999.
- [7] 株式会社こだまプロダクション, “効果音源,” <http://koukaongen.com/>, Aug. 2014.