

泣き声による児童虐待検知装置の提案

The propose of the embedded system to detect child crying
for the care of abused children

中野萌[†]
Megumu NAKANO

矢野良和[†]
Yoshikazu YANO

江口一彦[†]
Kazuhiko EGUCHI

Abstract

In recent years, the number of cases on child abuse was increased. Eliminating child abuse and preventing children from child abuse are needed. In order to reduce child abuse, ongoing monitoring on abused children is awaited, which is without case workers. Therefore we propose the monitoring system for abused children, as a wearable computer which detects child's crying. We propose both software algorithm and embedded hardware system. First, the method of cry detection by sound is proposed. The sound of cry consists of 2 major mode, cry with words and cry without words. The latter presents a melody, a same pitch keeps with some periods. Its pitch varies in order. Therefore, variation of pitch shows the period of melody cry. Experimental results shows its effectiveness. Next, we approach the hardware implementation. ARM based processor with DSP will execute extraction method of prosodic features. We tested the microphone control using microprocessor, and confirmed the captured data can be treated as a sound data.

1 はじめに

近年、経済的不安や育児不安などの要因により児童への虐待が増加している。図 1 は児童相談所に寄せられる虐待相談の件数の推移を示している。相談件数は平成 11 年度で 1 万件程度に対し、平成 21 年度では 4 万件程度と 10 年間で 3.8 倍に増加した。児童虐待の抑制や被虐待児童の早期発見、早期保護が求められている。

虐待が疑われる場合、発見者から児童相談所に通報され、児童福祉司による調査が行われる。児童福祉司は、対象家庭への聞き取り調査という直接的な調査や、周辺住民への聞き込みなどによる間接的な調査を行う。これら調査から虐待の事実関係の確認を行い、児童保護の必要性を判断する。また緊急性のある場合は法律上の権限で強制的に立ち入り調査を行い、児童を保護する権限を持っている。申告により対処をする仕組みから通報が遅れる問題がある。対象家庭との関係悪化や誤通

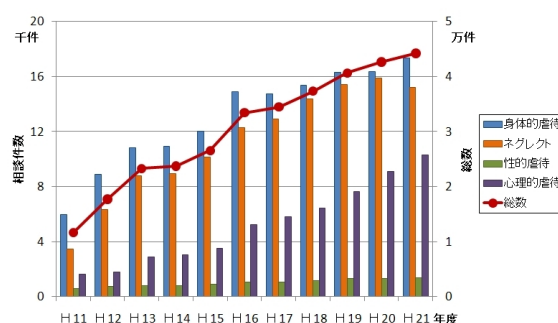


図 1: 児童相談所への児童虐待相談件数の推移

報時の責任を理由に、発見者が通報をためらい発見や調査が遅れるケースがある。一方で通報があった場合も、児童福祉司の人員数やその他の案件への対応のため、児童相談所は通報に対し即座に調査できないなどの問題がある。

図 1 の示すとおり、相談件数は年々増加し平成 21 年度では 4 万件を超える事態となっている。それに伴い児童福祉司の人材不足も指摘されている。現在、児童福祉司は人口 4 万から 7 万程度に一人

[†]愛知工業大学 工学部 電気工学科 (豊田市)

を標準として配置されている [1] . そのため児童福祉司が一人で複数案件を処理する必要がある . それにより , 虐待を受けている児童の発見が遅れ , 虐待が事件化することなどが指摘されている . 虐待で亡くなった児童も平成 22 年 4 月 ~ 平成 23 年 3 月の間に 50 人程度 [3] と高い水準で推移している . 児童福祉司を増員することなどの対応をとることも考えられるが , 人件費の負担が増えるため即対応が難しい .

調査においても人による調査のため , 担当する児童福祉司の熟練度の差が調査結果に強く影響を与える . 十分な情報収集ができず , 虐待の実態を把握できないまま調査が終了することもある . また , 過剰な情報収集から周辺住民に虐待という不名誉なうわさを生じさせ , 状態をさらに悪化させることも考えられる . そこで , 十分な情報を漏れなく収集するため機械的な常時監視を考える . 常時監視は情報収集に極めて有用な方法であり , 児童福祉司への負担軽減や熟練度差の是正にも効果を発揮すると考える . 一方でプライバシーや運用で情報収集への障害について検討をする .

カメラを使った動画の取得は , 異常状態の記録に極めて有効である . 画像処理技術などで , 被撮影者の行動分析から虐待の早期発見ができる可能性もある . しかし , カメラによる常時監視では様々な問題が発生する . 日常生活をすべて監視することはプライバシーや人権保護の観点から望ましい方法ではない . また , カメラの死角での虐待を監視することはできない . 得られた動画の目視による確認作業も時間がかかる処理で , 現実的な時間での応用は難しい .

次に音声による機械的監視について考えてみる . 音声は画像と比べて死角がなく広範囲の情報が取得できる . プライバシーについては視聴を伴わない虐待状況の確認が必要となる . 画像処理に比べて処理する情報量が少ないため , 単純な音声処理で虐待が検出できればリアルタイムでの虐待状況の監視へ応用ができる . 加えて , 処理の負荷が少ないならばマイコンによる実現も可能となり , ウェアラブルな装置での解析や記録へも期待できる .

日本で児童虐待は , 「児童虐待の防止等に関する法律」第 2 条において , 暴行などを受ける身体的虐待 , 性的ないやがらせをする性的虐待 , 育児放棄などのネグレクト , 精神的苦痛を与える心理的虐待として定義されている . 図 1 から身体的虐待とネグレクトが高い割合を占めていることが分かる .

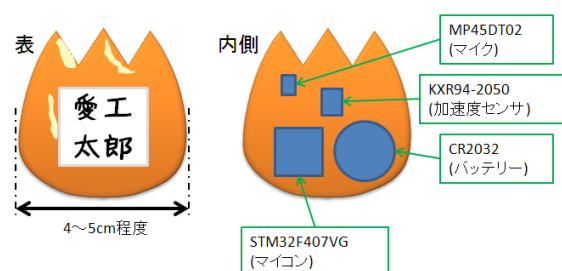


図 2: 提案するパッチ型虐待監視装置

身体的虐待を受けている児童やネグレクトを受けている児童は泣いている可能性が高いと考えられるため , 泣き声から虐待を検知できると考える .

そこで , 本研究では音声特徴を用いて泣き声を検知することで , 虐待の早期発見を補助する装置の提案をする . この手法では私生活を覗き見することがないためプライバシーの問題は発生しないと考えられる . また , この装置は児童が身につけられるパッチのような小型のものとする . これにより常時監視や監視の欠落に対する問題も解決できると考えられる . 本研究では泣き声の解析を行い , 泣き声検知アルゴリズムの提案とその有効性について調査する . またパッチのような装置の開発する前段研究としてマイコンでマイクを制御し , 音声の取得を行い , 取得音声の有効性についても述べる .

2 ハードウェアシステム

常時監視装置としてウェアラブルな装置を提案する . 図 2 に提案する装置を示す . 児童が日常的に付ける名札パッチと同サイズのシステムを目標とする .

この装置では , 次章で提案する泣き声を検出アルゴリズムをリアルタイム処理させるため , 高速なマイコンを搭載する . 一方で小型軽量のため , バッテリーによる長時間駆動が重要となる . そこで , ARM 系マイコンや DSP を利用した処理系を採用する .

通常 , パッチは胸位置や肩位置に取り付けられる . そこで , パッチ上方にマイクを取り付け , 集音を行う . 小型省電力を実現するため , 本研究ではデジタルタイプのマイクを採用する .

パッチタイプの集音システムだが , 虐待時の音声を集音させないようにするなどの理由から外したままにされることが考えられる . 録音した音が

無音が続いたり、パッチ自体が動かない状況は、児童もしくはパッチが異常状態にあるものと考え、これを監視する必要がある。そこで、泣き声検知とは別の目的で加速度センサを搭載する。このセンサは今後、児童への外力と発話などの相関性検証にも利用できるものとする。

3 泣き声検出アルゴリズム

泣き声検出アルゴリズムについて述べる。一般児童話者における泣き声と通常の発話では、音声の特徴量に明確な境界を定義することが難しい。そこで、まず泣き声を分析的に考え、通常発話と区別可能な特徴の定義を行う。次に泣き声の検出方法についての提案と、実験結果を示す。

3.1 音声の分類

音声は音声特徴を用いてその発話内容を分析する。音声特徴には、声の大きさや高さを表現する韻律特徴と、音を発する声道のパラメータを抽出した声質特徴に分類することができる。声質特徴は口や喉の形状を表現するため、発話内容を認識するのに有用であることが知られている。また、個人差も特徴に現れやすいため個人認証にも利用される。一方で通常発話と泣き声の分類の観点では、互いに複雑な特徴分布となり分離が困難となる。そこで、韻律特徴を用いて泣き声と通常発話との分離を試みる。

泣き声は、発話を伴う泣き声と発話を伴わない泣き声に大きく分類できる。発話を伴う泣き声は様々な発話バリエーションが存在するため、通常発話との違いを明確に表現することができない。この泣き方を本稿では「話し泣き」と呼ぶ。一方で、発話を伴わない音声は喉や口を響かせる泣き方で、メロディを奏でるように声の高さが変化する。この泣き方を「音階泣き」と呼ぶ。

音階泣きは、話し泣きや通常発話と異なり、単調なメロディを持つ。音の高さが連続的かつ普遍的に変化する発話と比べると、音階泣きは一定の音の高さを持続する特徴を持つ。泣き声は通常、話泣きの中に音階泣きが高頻度で出現し、話泣きと音階泣きは連続して観測される。そのため話し泣き部分は抽出することなく、泣き声区間を特定することが可能となる。そこで本研究では音階泣き

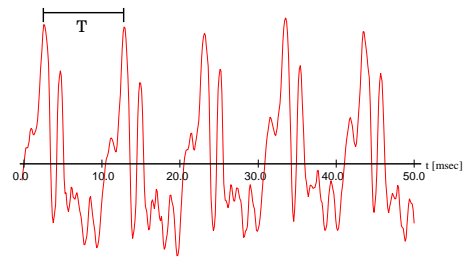


図 3: 「あ」の発話波形

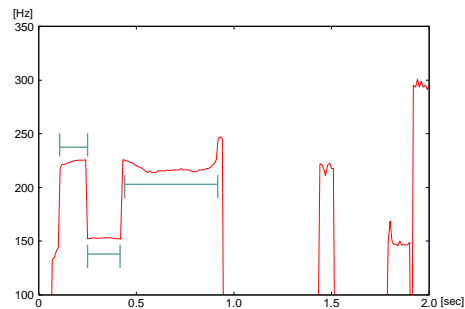
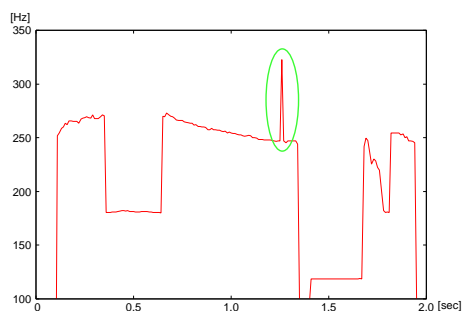


図 4: 泣き声要素のピッチ

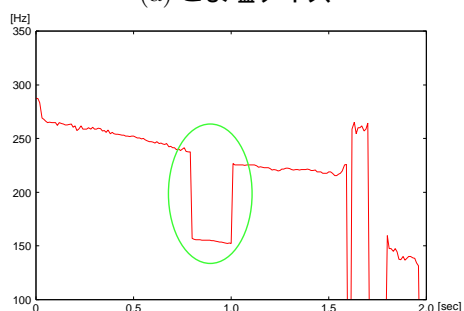
の要素を「泣き声要素」と定義し、泣き声要素の検出による泣き声の検知を目的とする。

3.2 泣き声要素の検出

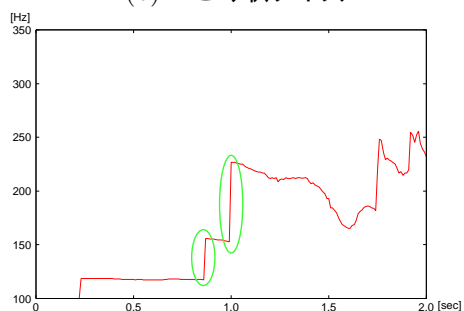
まず、検出対象である泣き声要素の分析を行う。音声解析では音声特徴量であるパワー、ピッチ、ホルムントについて解析を行った。解析を行った結果、パワーでは泣き声要素と通常発話との間に特徴的な違いが見られなかった。ホルムントでは泣き声要素と通常発話とで特徴的な違いが現れた。この特徴量は他の特徴量と比べて情報量が多く、抽出や判別にも時間がかかる。そのためハードウェア実装の観点から利用は困難と考えられる。これらに対しピッチは、泣き声要素と通常発話に特徴的な違いがあり、抽出や分離も比較的容易である。そのため、本研究ではピッチを特徴量として使用し泣き声要素の検出を行うものとする。ピッチは基本周波数とも呼ばれ、人間の知覚において音の高さに対応する。音声波形においては図3のように最も大きな波形の周期である T の逆数がピッチの周波数となる。図中の T は基本周波数に対応して基本周期と呼ばれる [4]。音声は大きな分類として有声音と無声音に分けられる。有声音は声帯の振動を伴い、無声音は声帯の振動を伴わない音声となる。ピッチは声帯の振動に対応する特徴量の



(a) ごま塩ノイズ

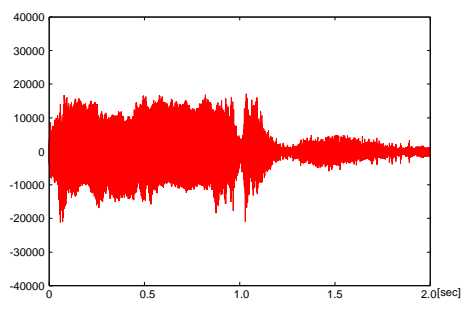


(b) へこみ状ノイズ

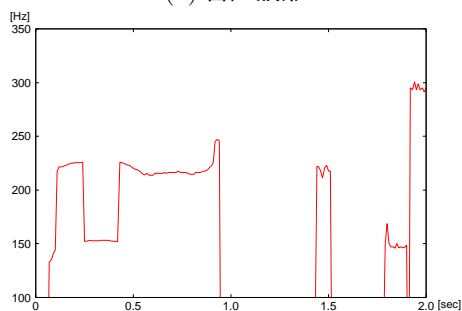


(c) 階段状の変化

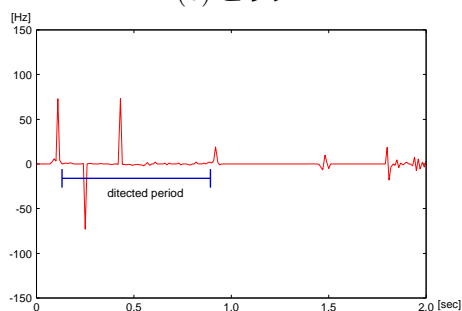
図 5: ピッチの例外的変化



(a) 音声波形



(b) ピッチ



(c) ピッチ変化量

図 6: 泣き声 1

ため、有声音のみに存在する。

泣き声要素は図 4 に示すようにピッチの変化の少ない部分が持続する部分が多くみられた。ここで、ピッチの時間微分をピッチ変化量と定義する。泣き声要素ではピッチ変化量が小さい状態の音声区間が比較的長く持続することが分かった。そこで本研究では、ピッチ変化量が小さい状態が比較的長時間持続する音声区間を泣き声要素として定義する。このとき、図 5(a)~(c) に示すような変化はメロディーの変化で連続した音声区間と考えられるため連続しているものとして処理をする。この泣き声要素の分布を分析することで、泣き声を検知する。

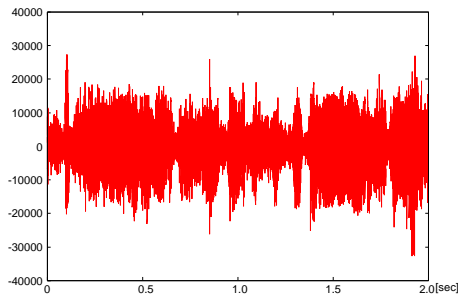
3.3 実験結果

提案手法での検出の成功例を図 6~10 に示す。実験結果ではそれぞれ音声波形、ピッチ、ピッチ変化量を示した。図中で泣き声として検出された部分にはマーカーを描画した。図 6, 7 には実際の音声示す。

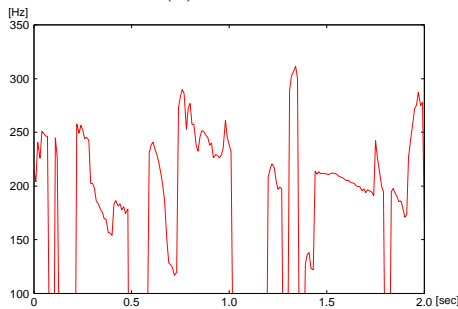
図 6 ではピッチ変化量がしきい値以下で持続した部分を泣き声として検出している。ステップ変化した部分は連続した区間として抽出されたこともわかる。図 7 の検出部分と音声データの部分を再生した音を比較検証した結果、泣き声部分の検出に成功したことがわかった。図 7 では、ピッチの変化が大きくピッチ変化量が閾値以下で長時間持続しないため、泣き声要素が検出されなかった。

提案手法での誤検出例と検出漏れの例を図 8~10 に示す。成功例と同様、図中で泣き声部分とし

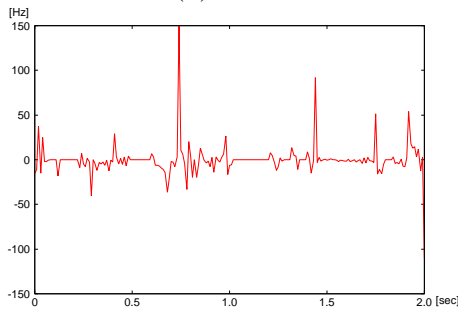
泣き声による児童虐待検知装置の提案



(a) 音声波形



(b) ピッチ

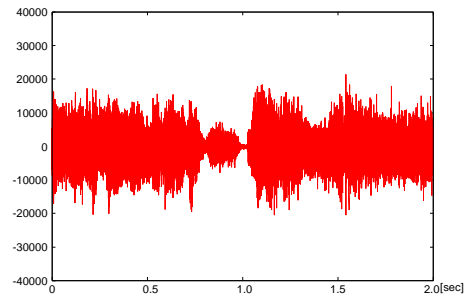


(c) ピッチ変化量

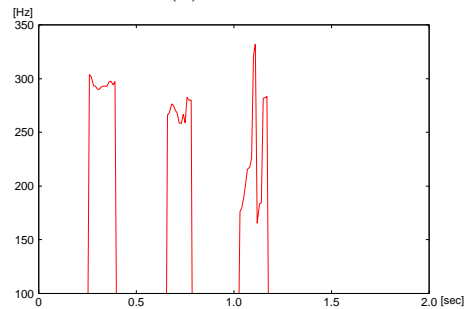
図 7: 研究室でのレクリエーション時の会話 1

て検出された部分にはマーカーを描画している。

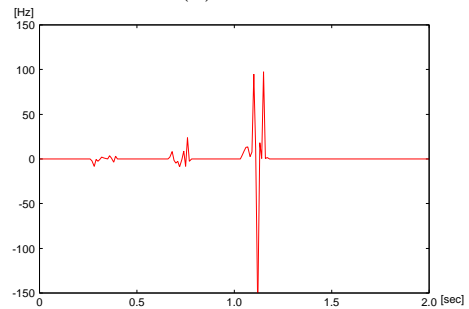
図 8 は全体としてかすれたような泣き声になっており、音の高さが明瞭に判断できない音のため、ピッチの抽出ができない部分が多く、音声データを再生したとき泣き声と認識される部分の検出に失敗したと考えられる。図 9 の誤検出部分は同じ音を長く発話する状態が持続したためピッチ変化量が小さい状態が持続し泣き声部分として検出されてしまった。図 10 での検出部分は発話の速度を速くした音声、いわゆる早口状態の部分を泣き声部分として検出した。続けて発話する時、ピッチは急激な変化を起こさないためピッチ変化量が小さい状態が持続し泣き声部分として検出されたと考えられる。しかし、このような音声群はごくまれにしか現れないことと、身体的虐待を受けている児童は長時間泣いていることが考えられるため、



(a) 音声波形



(b) ピッチ



(c) ピッチ変化量

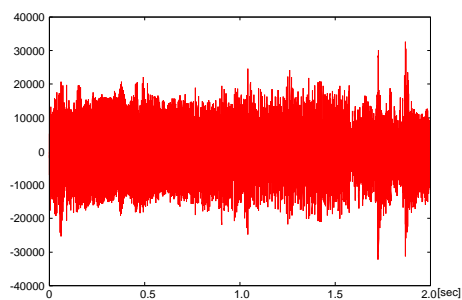
図 8: 泣き声 2

表 1: アルゴリズム検証結果

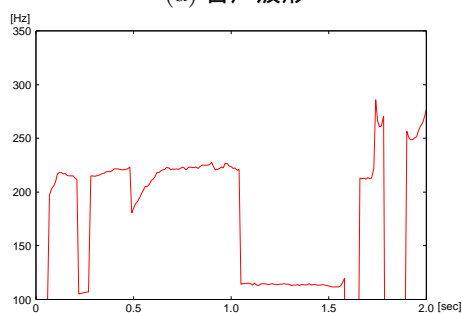
音声の種類	収録時間	検出時間	検出率
実際の泣き声	456sec	168sec	36.8%
自然な会話音声	370sec	16sec	4.3%

泣き声区間の発生する頻度を用いて改善することができると考える。

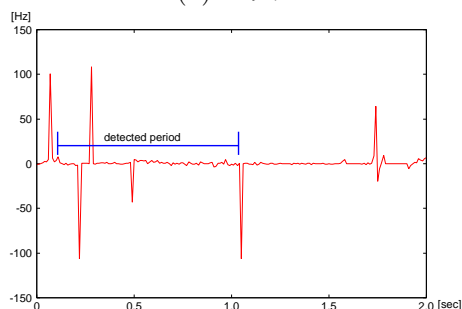
実験結果のまとめを表 1 に示す。表中で音声の種類は人間があらかじめ聞いて分類したもののラベル、収録時間はラベル分けされた音声の全時間を収録時間、検出時間は提案したアルゴリズムで泣き声として検出された音声区間、検出率は収録時間中に占める割合を示している。実際の泣き声と擬似的な泣き声では検出率は 30%程度と低い値を示したが、本アルゴリズムでは泣いている頻度を取得することが可能なので、児童虐待の検知には一定の効果があるといえる。このことから提案し



(a) 音声波形

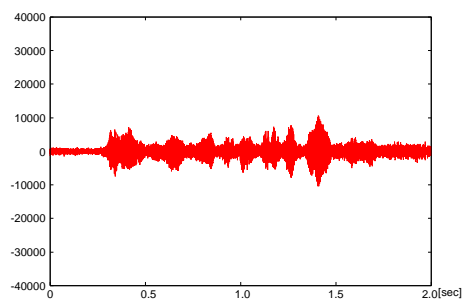


(b) ピッチ

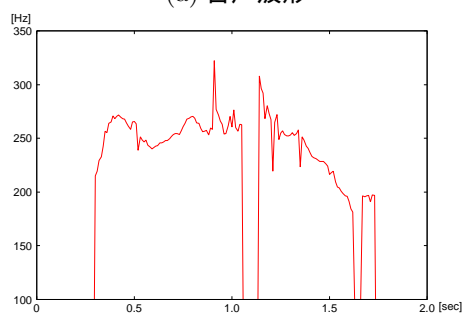


(c) ピッチ変化量

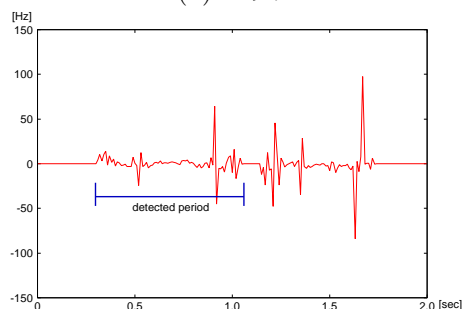
図 9: 研究室でのレクリエーション時の会話 2



(a) 音声波形



(b) ピッチ



(c) ピッチ変化量

図 10: 研究室でのレクリエーション時の会話 3

たアルゴリズムをハードウェアに実装することを考える。

4 ハードウェア実装

今回は、児童が身につけられるパッチのような装置開発の前段研究として、マイコンでマイクを制御しマイコンに音声を取り込む実験を行った。ここでは、使用したハードウェアの詳細について述べ、ハードウェアを制御し音声取得を行った結果を述べる。なお、音声波形は確認がしやすいよう PC 上に転送し音声波形としてグラフ化したものを掲載する。マイコン PC 間の通信にはシリアル通信を採用した。実際の通信信号を 11 に示す。

次に、使用したハードウェアの詳細について述べる。使用したマイコンは ARM Cortex M4 アーキテ

クチャを採用した STM32F407VG マイコンをマイクには MEMS 技術によって作られた MP45DT02 を使用した。STM32F407VG マイコンは内部に FPU を内蔵していることや DSP 命令を実装しているため実数の演算を高速におこなうことができる。音声処理において、ピッチの抽出時には実数の演算を多重に行う必要があるため、STM32F407VG マイコンは今回の研究において最適なマイコンと言える。MP45DT02 は MEMS 技術によって作られているため非常に小型なことがあげられる。また PDM 変調という変調方式を採用し、デジタル値の読み取りのみで音声データを複合することができる。PDM 変調とは音声などアナログ値をデジタル値にする方式で、パルス幅一定のパルスが発生させる。発生させるパルスはアナログ値の振幅の大きさに応じて一定周期間でのパルス

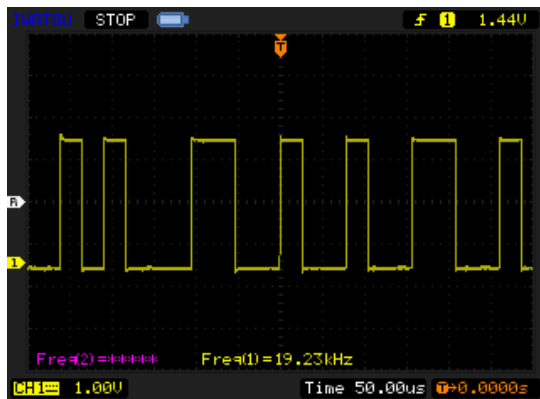


図 11: USART 信号波形

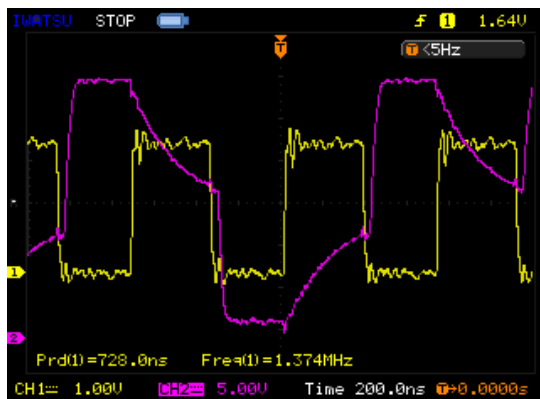


図 12: マイク出力とクロック信号

密度の疎密を変化させて出力することにより、アナログ値をパルスの密度によって表現する変調方式である。実際のマイクの信号とマイコンからのマイク駆動用のクロック信号を合わせた図を図 12 に示す。図中で黄色い信号がマイコンからのクロック信号、赤い線がマイクからの出力をそれぞれ示している。図 12 から分かる通りクロック信号の low 状態のとき信号を出力し、クロック信号が high 状態のときハイインピーダンス状態になっている。MP45DT02 ではハードウェア上でクロックとマイク出力の状態を設定できる。今回はクロックが low 状態のときマイクの信号を出力するように設定しているため、マイコンでマイクの制御が実現できていると考えられる。

最後にマイコンで取得した音声波形を図 13, 14 にそれぞれ示す。図 13 はピッチが一定になるように発話し、マイクからマイコンに取り込んだ音声、図 14 はピッチが一定にならないように発話しマイクからマイコンに取り込んだ音声をそれぞれグラフ化したものを示している。図 13 は類似した波形

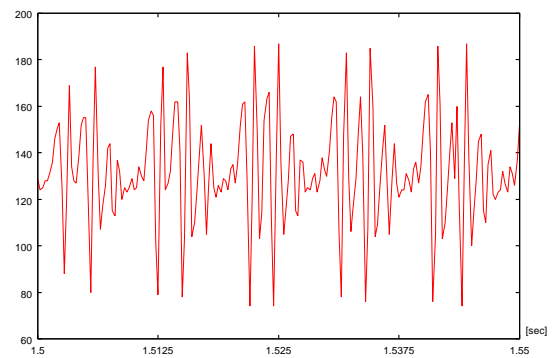


図 13: ピッチ一定の発話

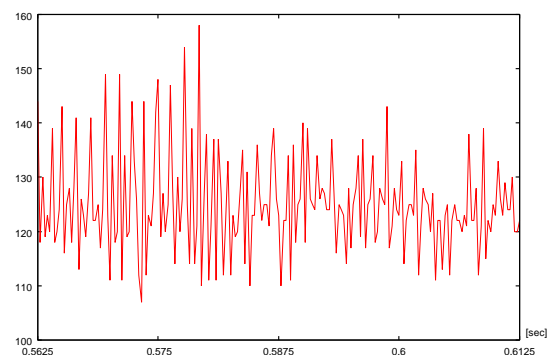


図 14: ピッチが一定ではない発話

が一定周期で繰り返している。このことからピッチが一定の音声として処理されていることが分かる。一方、図 14 は類似した波形が見られないので、ピッチが一定ではない音声として取り込まれていることが分かる。以上のことから、マイコンでマイクを制御し正しい音声波形の取得に成功したと考えられる。

5 まとめと今後の課題

音声からピッチを抽出しピッチの変化から泣き声を検出する手法を提案した。提案した泣き声検知手法で泣き声検出実験を行いその有効性を示した。次に、ハードウェア実装の前段階研究としてマイコンによりマイクを制御し、音声波形を取得することが可能なことを示した。

今後の課題として、提案アルゴリズムではしきい値の設定が音声の分離に大きく影響するため、しきい値の設定を機械学習で設定していく手法を考える必要が出てきた。今回はハードウェア実装について基本となる音声の取得のみが行えなかった。今後は周波数解析をマイコンに実装しピッチ

をマイコン上で抽出し, 提案アルゴリズムを実装することで, 泣き声の検出が可能かどうか検討していく.

参考文献

- [1] 児童福祉法施行令 第一章 第二条,
<http://law.e-gov.go.jp/htmldata/S23/S23SE074.html>
- [2] 厚生労働省, “雇用均等・児童家庭局資料”, 児童相談所における虐待相談の内容別件数の推移, <http://www.crc-japan.net/contents/situation/pdf/10011303.pdf>
- [3] 厚生労働省, “子ども虐待による死亡事例等の検証結果等について (第 8 次報告)”, http://www.mhlw.go.jp/bunya/kodomo/dv37/index_8.html
- [4] 広瀬 啓吉, “韻律と音声言語情報処理”, 丸善株式会社, 2006.
- [5] 板橋 秀一, 赤羽 誠, 石川 泰, 大河内正明, 粕谷 英樹, 桑原 尚夫, 田中 和世, 新田 恒雄, 矢頭 隆, 渡辺 隆夫, “音声工学”, 森北出版株式会社, 2008.
- [6] 久野 和宏, 野呂雄一, 井 研治, 成瀬治興, 吉久 光一, 大石弥幸, 岡田恭明, 佐野泰之, “音響学 ABC”, 技報堂出版, 2010.

(受理 平成 25 年 3 月 19 日)