

ウェアラブルデバイスにおける音声プライバシー保護システム

厚田和也[†] 孫 慈禧[‡] 浜中雅俊^{††} 李 昇姫[‡]

筑波大学第三学群工学システム学類[†]

筑波大学大学院人間総合科学研究科[‡]

筑波大学大学院システム情報工学研究科^{††}

1. はじめに

本研究では周囲の音を抑制する差分マイクでの録音により，目的音以外の周囲の会話が入力された際，そのプライバシーを保護する手法を構築する．本研究では，周囲の音を抑制し，目的音を強調する差分マイクの技術[1]を用いて得られた録音データを分析することで，マイクを装着した人物が発声した区間（以下，発声区間とする）の検出を可能とする．

ここで，本研究を進めるに至った背景として我々が現在進めているプロジェクトがある．このプロジェクトでは，複数のセンサを搭載したウェアラブルデバイスを子どもに装着し，そこに記録された子どもの行動や状態を親が確認することで子どもの安全を守るシステムを構築している．このデバイスには子どもの声を録音する機能が搭載されているが，保育園等の園内で使用する際，実際に録音データを確認する親以外の他の親同士の会話が入力される可能性がある．そこで本研究で構築するシステムにより，そのような会話のプライバシーを保護する．以下，本論文では差分マイクを搭載したウェアラブルデバイスを園内で全園児に装着することを前提とする．

従来より，複数のマイクが並んだマイクロホンアレイにより音源の方向を推定し，話者とその発声を認識する手法が提案されているが，その対象は室内における移動しない話者であった[2]．一方，差分マイク[1]や，それを応用した手法[3]では，マイクから離れた周囲の音の抑制や，目的音以外の方向への死角生成により，目的音の強調を可能としていた．本研究では差分マイクを用いて園児の発声が強調された録音データを獲得することで，その発声区間の検出を可能とする．

尚，本システムの評価実験において，非差分マイク（差分をしない普通のマイク）による録音に比べ，周囲の会話が有効に抑制されることを確認した．

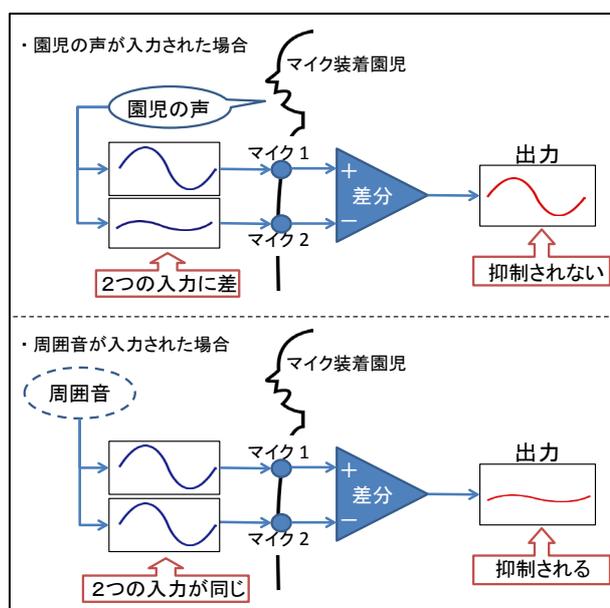


図1. 差分マイクによる録音

2. 音声プライバシー保護システム

本研究で構築するシステムでは，差分マイクによる録音で，マイクを装着した園児から離れた周囲の音が抑制され，結果的に園児の発声が強調された録音データを得る（図1）．これより，そのデータを分析し，音声信号エネルギー（振幅の2乗）があらかじめ設定する閾値より大きい区間を園児の発声区間とする．そして検出された発声区間のみを親に提供することで，園児以外の人物の会話に対するプライバシー保護が可能となる．

発声区間の検出：録音データをフレーム分割し，フレーム毎に音声信号エネルギーの平均を求め，それが閾値より大きいフレームを発声区間とする．

フレーム長：発声区間が正確に検出されるために，録音データを分割するフレームの長さは，録音データ中の各発声の長さ以下であることが望ましい．そこで録音データを調べ，咳払いなどの瞬間的なものを除いた発声の中で，その時間が最小である発声の長さは2秒であったため，それをフレーム長とする．

"Speech Privacy Protection Systems for Wearable Devices"

[†]Kazuya Atsuta [‡]JaHee Son, SeungHee Lee ^{††}Masatoshi Hamanaka (University of Tsukuba, [†]College of Engineering Systems, Third Cluster of Colleges [‡]Graduate School of Comprehensive Human Sciences ^{††}Graduate School of Systems and Information Engineering)

閾値の設定：発声区間の正確な検出のためには、フレーム長だけでなく、信号エネルギーの平均に対する閾値も適切に設定する必要がある。そこで、録音実験で得た録音データをフレーム分割し、フレーム毎に平均エネルギーを求め、このうち最大の平均エネルギーの2分の1、3分の1、4分の1を閾値としてそれぞれ発声区間の検出を行った。その結果、3分の1のときが最も正確に発声区間が検出されたため、これを閾値とする。

発声区間の訂正：2人以上の園児が至近距離にいる状況で、このうち1人が大声を出した際、他の園児の差分マイクがこの声を抑制し切れず、発声していない園児の発声であると検出される可能性がある。そこで発声区間の検出後、発声の始点と終点がともに一致する園児が複数いる場合には、その区間における信号エネルギーの平均を園児毎に求める。そして、エネルギー平均が最大である園児の発声として、それ以外の園児についてはこの発声区間を削除する。尚、この性能を評価するための実験について3.3節で述べる。

3. 実験

本研究において使用する録音データを得るため、実際に5歳の子ども4名と親の役をする成人2名を被験者として室内で録音を行った。子どもたちは差分マイクと非差分マイクを装着し、自然に会話が生まれるような作業を行った。また親役は子どもの周囲に配置し、文章の音読を行った。その結果、園内において園児の周囲で親同士が会話している状況に近い環境で録音をすることができた。

以下ではこの録音データを用いて、本研究において構築した音声プライバシー保護システムの性能評価を行った。

3.1 発声区間検出の評価

子どもの発声区間の検出性能を評価するため、録音実験で得た差分マイクと非差分マイクの録音データについて比較した。その結果、発声区間の検出率は、差分マイクが96%、非差分マイクが84%であった。これは、両マイクを胸元に装着したことで、非差分マイクについても、子どもの声が周囲の音より大きく入力される状態であったため、高い検出率となった。しかし、10%以上の開きにより、差分マイクの有効性が確認できた。

3.2 プライバシー保護の評価

親役の声を抑制し、プライバシーの保護が達成できているか評価するために、差分マイクと

非差分マイクの録音データについて比較した。その結果、親役の声が入力された区間中、差分マイクが92.3%、非差分マイクが81.2%で親役の声が抑制できた。ここで差分マイクに関して、親役の声を抑制できなかった区間は全て、子どもがマイク付近で使用していた道具の音が連続的に入力されており、これが原因で発声区間として検出された。したがって、このような音が入力されにくい工夫を施すことで、より高精度の抑制が期待できる。これに対し、非差分マイクについては、道具の音などが入力されていないにも関わらず、親役の声により子どもの発声区間として検出されることがあった。

3.3 発声区間訂正の評価

録音実験での録音データにおいて、ある子どもが発した大声が、近くにいた他の子どもの差分マイクに入力された部分が2か所あった。そこで、この録音データに対し、2章で述べた発声区間の訂正処理を施した場合と施さない場合について比較した。その結果、処理を施した場合は2か所とも発声区間が正確に検出されたが、施さない場合は正確に検出されなかった。

4. おわりに

差分マイクによる録音データを分析することで、録音したい声以外の周囲の会話が入力された際にそのプライバシーを保護するシステムを構築した。性能評価実験においても有効に周囲の会話が抑制できることがわかった。

謝辞 本研究の一部は総務省による戦略的情報通信研究開発推進制度（SCOPE）の支援を受けて行われた。

参考文献

- [1] (株) 船井電機新応用技術研究所, “話者音声のみを捉える微小で高性能な雑音カットマイクの開発に成功,” 報道発表資料-2008.7.22
- [2] 浅野太, 緒方淳, 松坂要佐, 山田実一, 中村雅巳, “会議収録データにおける発話イベントの構造化と分離について,” 日本音響学会研究発表会講演論文集, 秋季, pp. 1-2-15, Sep. 2006.
- [3] 高田晋太郎, 勘場智之, 小川哲司, 赤桐健三, 小林哲則, “減算型アレイ処理とスペクトラルサブトラクションを用いた音源分離技術とその携帯電話への応用,” 電子情報通信学会技術研究報告. SP, 音声, 106(123), pp. 7-12, June. 2006