

科学研究費助成事業 研究成果報告書

平成 28 年 6 月 21 日現在

機関番号：34412

研究種目：若手研究(B)

研究期間：2013～2015

課題番号：25750137

研究課題名(和文) 悲鳴発声時の声道特性の解析と悲鳴検出システムへの応用

研究課題名(英文) Feature analysis of scream sound and its application to scream detection system

研究代表者

早坂 昇 (Hayasaka, Noboru)

大阪電気通信大学・情報通信工学部・講師

研究者番号：50554573

交付決定額(研究期間全体)：(直接経費) 2,600,000円

研究成果の概要(和文)：犯罪の多様化・巧妙化に伴い、防犯システムに対する意識が向上している。最も代表的な防犯システムである防犯カメラはプライバシーへの配慮や暗闇で使用できないなどの課題があり、音情報を利用する防犯システムが提案されている。本研究では、悲鳴音声に焦点を当てた悲鳴検出システムに関して、悲鳴音声のデータベース構築と悲鳴音声の検出精度を向上させる手法の提案を行う。具体的には、悲鳴音声と雑音の周波数分布の違いを表すスペクトルエントロピーを特徴量として加え、雑音重畳モデルを導入する。雑音下を想定した悲鳴識別実験を行った結果、雑音既知・未知どちらの環境においても大幅な性能の向上が見られた。

研究成果の概要(英文)：Recently, as crimes have become more diverse and increasingly sophisticated, our awareness of security systems has increased. The most typical kind of security system is a security camera, but there are various issues in that there are situations in which they cannot be used, such as from consideration of privacy or in darkness. For that reason, security systems that utilize sound information have been proposed. This report relates to scream detection system that focuses on screams within various sounds, and proposes techniques increasing its noise robustness. More specifically, we use spectral entropy, which expresses the difference in frequency distribution between a scream and noise, as a feature, and also construct a model from screams with superimposed noise. In scream identification and scream detection experiments under noisy environments, the proposed method could be improved the equal error rate under all noisy environment, in comparison with the conventional method.

研究分野：音声信号処理

キーワード：悲鳴検出 帯域制限 スペクトルエントロピー

1. 研究開始当初の背景

防犯・セキュリティ分野において、新たな犯罪防止・抑止技術が望まれている。現在、主流のセキュリティシステムに監視・防犯カメラがあるが、人件費等の問題から常に監視することは稀であり、単に録画しているだけに過ぎない場合が多い。そのため、犯罪発生後の捜査にとっては有効だが、犯罪防止・抑止の効果としては限定的である。このような背景から、申請者は音声を利用した新たな犯罪防止・抑止システムとして、人(主に女性)が発する悲鳴を自動で検知するシステムの開発に取り組んでいる。本システムの利用例を図1、2に示す。

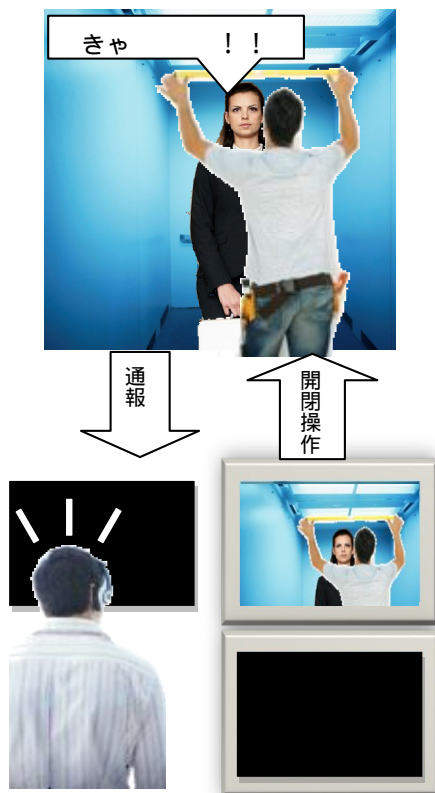


図1 使用例1

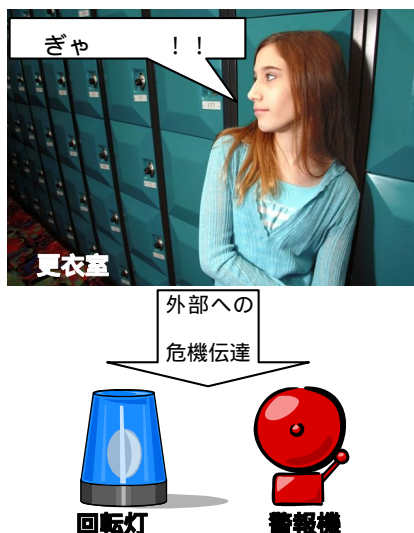


図2 使用例2

2. 研究の目的

本研究の目的は高い雑音耐性を持った悲鳴検出システムを構築することである。前述の利用例のようなシステムを運用するためには非常に高い検出性能でなければならない。また、暗い路地など屋外でも利用されることが想定されるため、雑音への耐性も望まれている。本研究では、以下の3点に取り組んだ。

- (1) 悲鳴音声データベースの構築
- (2) スペクトルエントロピーによる悲鳴検出精度の向上
- (3) 雑音重畳モデルの導入による雑音耐性の向上

3. 研究の方法

(1) 悲鳴データベースの構築

悲鳴音声の解析に先立ち、多数の女性話者から悲鳴音声を収録し、悲鳴データベースを構築する。本研究では悲鳴音声を「女性が身の危険を外部に伝達するための音声」と定義し、収録スタジオで録音する。スタジオ内にて身の危険を感じる状況を再現することは困難であるため、身の危険をイメージすることを助ける静止画像や動画を作成し、それを視聴した後、録音する。具体的な方法を以下に示す。

1. 各話者が思い浮かべる危機的場面を意識して発声
2. 悲鳴を発している最中の静止画像を見てから発声
3. 危機的場面の再現動画(映像のみ)を見てから発声
4. 危機的場面の再現動画(映像と音声)を見てから発声
5. 演技指導者による演技指導を受けてから発声

(2) スペクトルエントロピーによる悲鳴検出精度の向上

悲鳴音声は担う目的から、音量が増大し、通常よりも高い音高で発せられる。また、放射特性が良い音韻である「a」の音を含む。これらの特徴はすべて検出に有効であるが、音量や音高のような韻律的特徴は雑音の影響を受けやすい欠点を持つ。

音量は一般的に以下の式で表される対数エネルギーが用いられる。

$$E_t = \log_{10} \sum_{k=0}^{K-1} X_t[k] \quad (1)$$

ここで $X_t[k]$ は t フレーム目、周波数ビン k のパワースペクトルであり、 K はフーリエ変換点数である。また、基本周波数の抽出には、以下の式で定義される自己相関関数($R_t[i]$)が用いられる。

$$R_t[i] = \frac{1}{K} \sum_{k=0}^{K-1} \hat{X}_t[k] \exp\left(j \frac{2\pi ki}{K}\right) \quad (2)$$

ここで、 $\hat{X}_t[k]$ はケプストラム領域にてスペクトル包絡情報を除去したパワースペクトル

である。その後、サンプリング周波数を以下の規則で得られる I で除算し、各フレームの基本周波数 (FF_t) とする。

規則 1 $i_{\min} < i$ の範囲で極大値 (Local Max) およびその最大値 (Max) を抽出する。

規則 2 $\alpha \text{Max} \leq \text{Local Max}$ を満たす Local Max を取る i の中で最小のものを I とする。

($0 \leq \alpha_1 \leq 1$)

ここで、雑音下で発せられた悲鳴音声に対する E_t の FF_t を図 3、4 に示す。どちらも 200 ~ 360 フレームに悲鳴音声が発せられている。

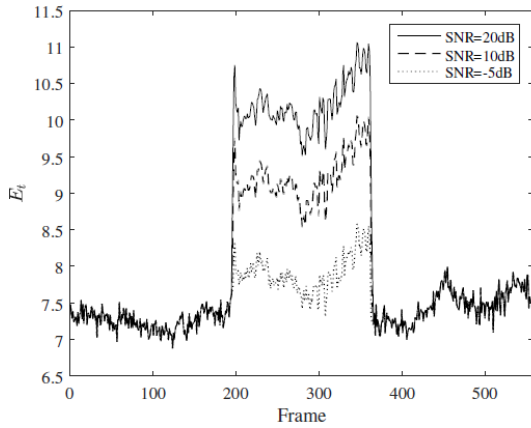


図 3 E_t のフレーム時間変化

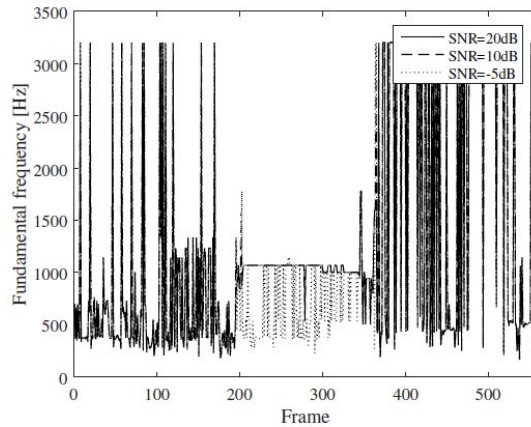


図 4 FF_t のフレーム時間変化

E_t は SNR が低下するにつれ環境音レベルと同程度に近づき、SNR=-5dB では音量に大きな差がない。また、SNR=10, -5dB で悲鳴音声の FF_t が正しく抽出されていない。

次に、音響的特徴である声道特性について述べる。声道特性を表すパラメータとして、以下の式で表される Mel-Frequency Cepstral Coefficients (以降 MFCC) ($C_t[l]$) がある。

$$C_t[l] = \sqrt{\frac{1}{M} \sum_{m=0}^{M-1} \log(X_t^{\text{mel}}[m]) \cos\left\{\frac{(2m+1)\pi l}{2M}\right\}} \quad (3)$$

$$X_t^{\text{mel}}[m] = \sum_{k=0}^{M-1} B_{m,k} X_t[k] \quad (4)$$

ここで、 $X_t[k]$ は t フレーム目、 k 番目の周波数ピンのパワースペクトル、 $B_{m,k}$ はメルフィルタバンク行列、 M はフィルタバンク数であ

る。音声認識では $1 \leq l \leq 13$ が一般的であり、本研究においても同様の値を用いる。

ここで、悲鳴音声のパワースペクトルを図 5 に示す。

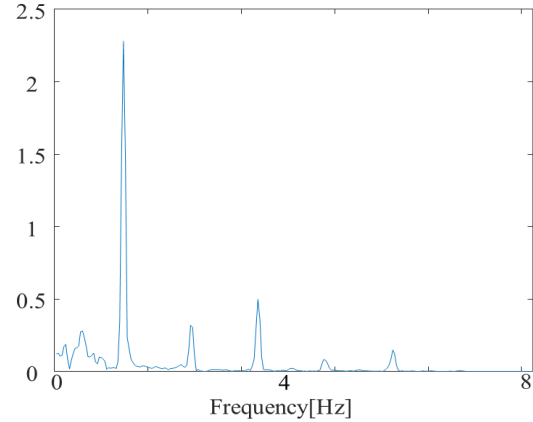


図 5 悲鳴音声のパワースペクトル

悲鳴音声は高い基本周波数を持つため、その倍音成分を含めても全周波数帯域に対して疎らにしか存在しない。一方、雑音は特定の基本周波数を持たない場合が多く、全周波数帯域に広く存在する。これらの周波数分布形状を表すパラメータとして次式で示すスペクトルエントロピー H_t がある。

$$H_t = - \sum_{n=K_{\min}}^{K_{\max}} P_t[n] \log P_t[n] \quad (5)$$

$$P_t[n] = \frac{X_t[n]}{\sum_{k=K_{\min}}^{K_{\max}} X_t[k]} \quad (6)$$

K_{\min} および K_{\max} はスペクトルエントロピーを算出する下限および上限の周波数ビンである。 H_t を悲鳴検出の特徴量に加え検出性能の向上を狙う。

(3) 雑音重畳モデルの導入による雑音耐性の向上

雑音重畳モデルとは、複数の雑音を複数の SNR で重畳した音声 (本研究では悲鳴音声) により構築したモデルである。無雑音環境と雑音環境のミスマッチを低減することができる。重畳する雑音の種類や SNR は、そのシステムが稼働する場面で想定されるものを使用することが最良であるが、実際にはそのような場面は未知であることが多い。その観点から、性質が異なる複数の雑音を複数の SNR で重畳してモデルを構築する。なお、重畳する雑音の種類や SNR の選択に関する議論は本研究では行わない。

4. 研究成果

(1) 悲鳴データベースの構築

収録はサンプリング周波数 48kHz、量子化 16 ビットで行い、その後の実験では 16kHz にダウンサンプリングした。上記手順により女性 48 名から計 922 個の悲鳴音声収録した。収録した 922 個の悲鳴音声の基本周波数の分

布を図6に示す。

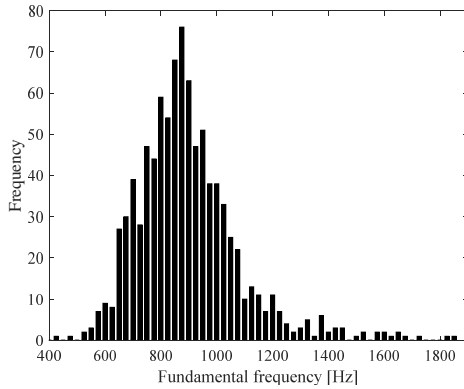


図6 基本周波数の分布

基本周波数算出時の i_{\min} は5、 α は1としている。基本周波数は450~1850Hzに分布し、900Hz前後に多く分布している。これは女性の通常発声時の基本周波数(200~300Hz)よりも遥かに大きく、外部に危険を伝達する目的を達成しようとする意図が働いたためであると考えられる。対数エネルギーに関しては、悲鳴発声時に音割れしないように調整したため、定量的な評価はできていない。

(2) スペクトルエントロピーによる悲鳴検出精度の向上

音声の分析条件を表1に示す。

表1 分析条件

Window function	Hamming window
Frame length	25 ms
Shift length	10 ms
Pre-emphasis	$1 - 0.97z^{-1}$

本実験では悲鳴および雑音のモデルとして正規分布を用いる。収録した922個の悲鳴音声の内、半分の461個の悲鳴音声を悲鳴モデルのパラメータ推定に使用した。雑音モデルのパラメータ推定には、JEITA騒音データベースより「工場」、「交差点」、「計算機室」を選択し、それぞれの雑音から悲鳴音声と同じ長さでランダムに461個切り出しそれらを用いた。悲鳴の入力には学習に用いていない悲鳴音声461個を用い、それぞれに以下の雑音をSNR=0dBで重畳した。

- 雑音モデルに用いた「工場」、「交差点」、「計算機室」(Closed condition)
- 未知の雑音「駅」、「列車」、「空調機」(Open condition)

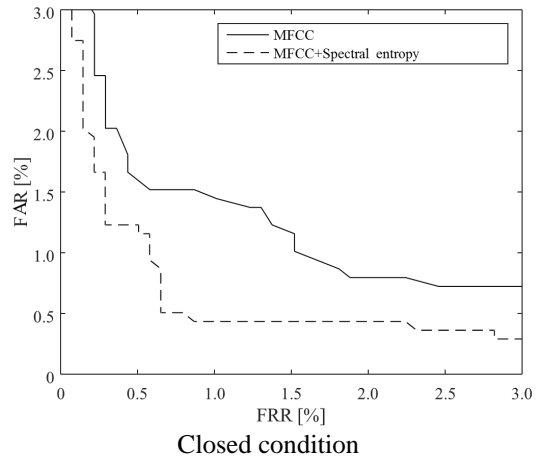
また、雑音の入力には上記雑音を用いた。

次に評価指標として、以下の式で表される False Acceptance Rate(FAR)と False Rejection Rate(FRR)によって描かれる Receiver Operating Characteristic(ROC)曲線により評価する。ROC曲線が原点に近ければ近いほど高い性能であることを意味している。

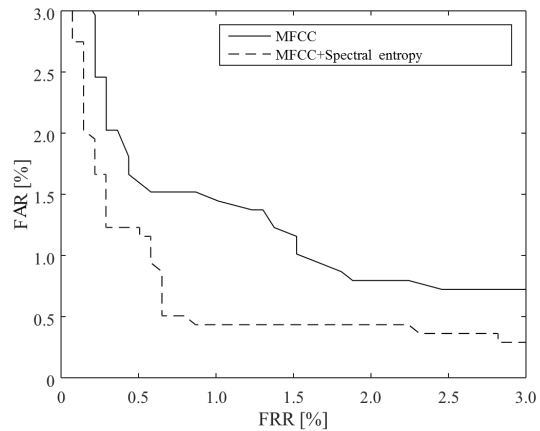
$$FAR = \frac{\text{Number of detected noise}}{\text{Number of evaluated noise}} \times 100 \quad (11)$$

$$FRR = \left(1 - \frac{\text{Number of detected screams}}{\text{Number of evaluated screams}} \right) \times 100 \quad (12)$$

結果を図7に示す。図より従来特徴量であるMFCCにスペクトルエントロピーを加えることで原点に近い曲線を描いており、スペクトルエントロピーが有効であることが示されている。



Closed condition



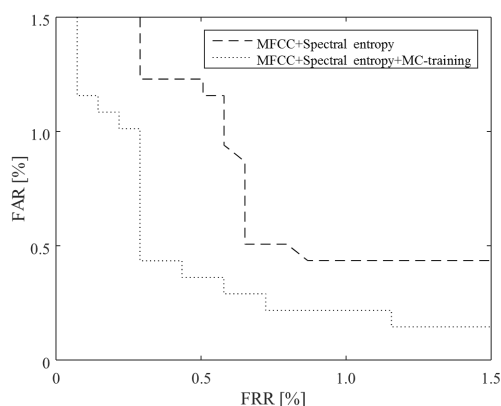
Open condition

図7 スペクトルエントロピー導入効果

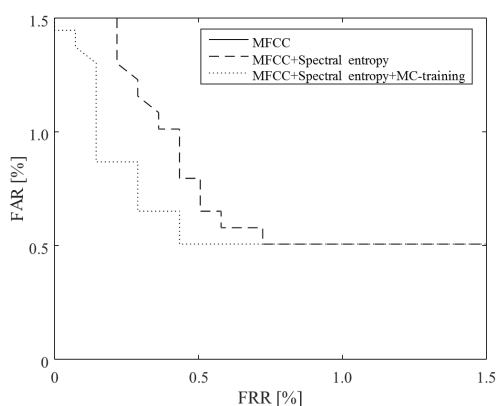
(3) 雑音重畳モデルの導入による雑音耐性の向上

雑音重畳モデルを除く実験条件は前述と同様である。雑音重畳モデルのパラメータ推定は、悲鳴モデルを構築した悲鳴音声に加え、その悲鳴音声に雑音モデルに用いた雑音3種をそれぞれSNR=0,10dBで重畳した音声により行った。

結果を図8に示す。MFCC+Spectral entropyの結果は図7と同一である。図8から、雑音重畳モデルの結果は更に原点に近い曲線を描いていることが確認できるが、スペクトルエントロピーを特徴量に加えた結果がすでに十分に高い性能(等誤り率0.6~0.8%)であり、大幅な改善には至らなかった。



Closed condition



Open condition

図8 雑音重畳モデル導入効果

<引用文献>

L. Gerosa, G. Valenzise, F. Antonacci, M. Tagliasacchi, and A. Sarti, "Scream and gunshot detection in noisy environments," in Proc. The 15th European Signal Processing Conference, Poznan, Poland, pp.1216-1220, Sep. 2007.

W. Huang, T.K. Chiew, H. Li, T.S. Kok, and J. Biswas, "Scream detection for home applications," in Proc. IEEE Conference on Industrial Electronics and Applications, Taichung, Taiwan, pp. 2115-2120, June 2010.

W. Choi, J. Rho, D.K. Han, and H. Ko, "Selective background adaptation based abnormal acoustic event recognition for audio surveillance," in Proc. IEEE International Conference on Advanced Video and Signal Based Surveillance, Beijing, China, pp.118-123, Sep. 2012.

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[学会発表](計5件)

早坂昇, "スペクトルエントロピーとマルチコンディション学習による雑音下悲鳴検出," 電子情報通信学会技術研究

報告 SIS, Osaka, Japan, pp.1-6, 3-4 Dec. 2015.

Noboru Hayasaka, "Noise-robust scream identification using spectral entropy and multi-condition GMM," Int'l Symposium on Smart Info-Media Systems in Asia (SISA), Narashino, Japan, pp.96-99, 26-28 August 2015.

Hiroyuki Sakuno, Noboru Hayasaka, and Youji Iiguni, "Noise robust speech recognition selectively using noise adapted HMM set," IEEE Int'l Conference on Electronics Circuits and Systems (ICECS), Marseille, France, pp.124-127, 7-10 Dec. 2014.

Junichiro Wakasugi, Noboru Hayasaka, and Youji Iiguni, "Robust voice activity detection using selectively energy features," IEEE Int'l Conference on Electronics Circuits and Systems (ICECS), Marseille, France, pp.359-362, 7-10 Dec. 2014.

早坂昇, "基本周波数の分散に着目した悲鳴信頼度に関する検討," 第27回回路とシステムワークショップ講演論文集, pp.197-202, Awaji, Japan, 4-5 August 2014.

6. 研究組織

(1)研究代表者

早坂昇 (HAYASAKA, Noboru)

大阪電気通信大学・情報通信工学部・講師

研究者番号: 50554573