音響防犯システムのための SVMを用いた叫び声の検出と音声認識

外村 淳[1]　南條 浩輝[1]　西浦 敬信[2*]

[1] 龍谷大学理工学部　[2] 立命館大学情報理工学部

e-mail: tonomura@nlp.i.ryukoku.ac.jp

1 はじめに

安全・安心な社会基盤の確立を目的として，音響防犯システムの研究を行う．具体的には，音声認識技術を用いて緊急性の高い叫び声を認識し，自動で警備会社に通報するシステムの実現のための研究を行う．本稿で扱う叫び声は，言語的な意味を持つ音声を対象とする．そのため，単に大きな声を出しただけなのか，助けを求める叫び声なのかを判断し，緊急性が高いと判断されたものだけを通報することを考える．我々はすでに男性10名のデータベースを用いて研究を行ってきたが [1][2]，今回は男性20名，女性20名のデータを用いた結果を報告する．また，頑健な叫び声検出のためにSVMを用いた実験を行ったので，その結果についても報告する．

2 叫び声のデータベース

緊急時に実際にどのような単語が叫ばれるか十分に検討すべきである．我々は男女各20名の平静発話，叫び声の各50単語を収録した [3][4]．このデータベースには，様々な環境での認識率を調査するために，ヘッドセットマイク（set-1），遠距離マイク（set-2）で収録した音声も含まれている．また，天井マイク収録音声（set-3）として，set-1の音声にインパルス応答を畳み込んだデータも作成している．このうち，人手でも叫び声の識別が困難であったため，今回は男性19名，女性18名の音声データを実験対象とした．

3 叫び声のモデル化

我々はこれまでに叫び声の音声認識は平静発話用の音響モデルでは音声認識が困難であることを明らかにしており [5][6]，今回も叫び声に適した音響モデルの作成を行う．平静発話で学習した音響モデルをMLLR適応により，叫び声モデルを学習する．

*Shout detection with SVM and automatic speech recognition for acoustic-based security system by A.TONOMURA, H.NANJO（Ryukoku University）, and T. NISHIURA（Ritsumeikan University）

表1: ゆう度最大基準による誤報率（%）

	男性	女性
set-1	10.3	8.47
set-2	14.8	13.2
set-3	22.3	18.0
平均	15.8	13.2

※誤棄却率を1%以下に制御

4 叫び声の検出

4.1 音響ゆう度最大基準による識別

はじめにゆう度最大基準による識別について説明を行う．これは，12次元のMFCCとその一次差分（ΔMFCC）およびΔ powerを特徴量としてHMM（Hidden Markov Model; 隠れマルコフモデル）を平静発話と叫び声それぞれで学習し，音声が入力されると音響ゆう度を算出し，ゆう度最大基準による識別を行うものである．具体的には，環境ごと，すなわち性別（男女の2種類），発話タイプ（平静，叫び声の2種類），マイクとの距離（set-1から3の3種類）の，合計12種類のHMM音響モデルを学習し，入力音声を，ゆう度が最も高いモデルのカテゴリ名とした．

識別誤りには，平静発話が入力されたときに叫び声と誤る割合（誤報率）と，叫び声が入力されたときに平静発話と誤る割合（誤棄却率）がある．音響防犯システムにおいては，誤棄却，すなわち叫び声を平静発話と誤ることは，平静発話を叫び声と誤ることよりも，致命的なミスである．そのため，誤棄却率を低く抑える必要がある．これは，叫び声モデルから算出されるゆう度に重みを与えることで対応できる．本研究では重み付けにより誤棄却率を1%以下になるように制御したときの誤報率で評価を行う．

結果を表1に示す．なおHMMの学習は4分割交差検定で行い，重みは，誤棄却率が1%以下になるように事後的に決定した*．表1よりマイクと話者との距離が

*ただし女性1名については，誤棄却率1%以下にできなかったため，ここではデータから除いた．

Copyright © 2010 Information Processing Society of Japan.
All Rights Reserved.

表 2: 誤報率（%）

男性	ゆう度最大	SVM (ゆう度)	SVM(ゆう度 +power,F0)
set-1	10.3	10.7	5.26
set-2	14.8	12.7	10.3
set-3	22.3	13.6	14.3
平均	15.8	12.3	9.95

※誤棄却率を 1%以下に制御

識別に影響していることがわかる．特に set-3 では，平静発話が入力されたとき，約 20%を叫び声と誤ることがわかった．これは，さらに識別率の向上が必要であることを示している．

4.2 音響ゆう度を特徴量とした SVM による識別

次に，識別精度向上のため，2 値分類器である SVM の利用を考える．はじめに，SVM の特徴量として音響ゆう度を用いることを考える．具体的には，先程と同じ音声特徴量（MFCC+Δ MFCC+Δ power）に対する 12 種類の HMM 音響モデルのゆう度を算出し，そのゆう度を SVM の 12 次元の特徴量とする．SVM は，識別面のどちらにあるかで 2 クラスに分類を行う．識別面は 2 クラスのデータとの距離が等しくなるように生成されるが，片方の領域を広くするため，識別面に重みを与え，平行移動させた．これにより，叫び声と分類されやすくなり誤棄却率を抑えることができる．ここでは，男性 19 名のデータを用いて評価を行った．SVM の学習データには，男性の平静，叫び声のデータと女性の叫び声を用いた．これは，男性の叫び声と女性の平静発話を間違うことが多いためである．ここでも，4 分割交差検定を行った．結果を表 2 に示す．誤報率を低くできた．特に，set-3 では，誤報率を約 40%削減することができた．

4.3 音響ゆう度，パワー，F0 を特徴量とした SVM による識別

ゆう度を特徴量とした SVM において，SVM の有用性が確認できた．次に特徴量の追加を行う．具体的には発話ごとの音声データから，パワー，基本周波数（F0），パワーの速度，F0 の速度，パワーの加速度，F0 の加速度を 10 ミリ秒ごとに抽出して，それらの平均値，最大値，標準偏差，尖度，歪度を算出し，これらと音響ゆう度の合計 42 次元を特徴量として SVM で識別を行った．結果は，表 2 に示されている．特徴量を増やしたことにより，さらに誤報率を低くできた．今後，さらなる特徴量の検討が必要である．

5 叫び声の音声認識

適応前の音響モデルと MLLR 適応により作成した叫び声音響モデルを用いて，叫び声の音声認識を行う．

表 3: 音響モデル適応後の叫び声の音声認識率（%）

		適応前	適応後
男性	set-1	51.9	88.9
	set-2	47.9	89.8
	set-3	42.5	85.9
女性	set-1	26.7	87.1
	set-2	20.3	86.7
	set-3	19.2	82.1

MLLR 適応は 4 分割交差検定で行った．

語彙サイズ 50 単語の叫び声の音声認識の結果を表 3 に示す．叫び声については，適応を繰り返すことで，それぞれ認識率の向上が見られた．しかし，男性 10 名のときの認識実験の結果 [1] よりも低い精度であり，特に，女性に対して高い精度が得られなかった．この結果は，学習方法を検討する余地があることを示している．また，これらは防音室で収録した音声を対象としており，防犯システムが必要な実環境でも実験を行う必要がある．

6 おわりに

音響防犯システムのために叫び声の検出と音声認識を行った．SVM を用いた識別により，より高い識別精度を得た．今後はさらに精度向上のため，特徴量の検討が必要である．認識においては，学習方法の検討が必要である．

参考文献

[1] Hiroaki Nanjo, Hiroshi Kawano, Hiroki Mikami, Eri Ohmura, and Takanobu Nishiura. Shouting speech detection and understanding for acoustic-based security system. *Journal of Information Assurance and Security (JIAS)*, Vol. Vol.5, Issue 1, pp.256-264, , 2010.

[2] Hiroaki Nanjo, Takanobu Nishiura, and Hiroshi Kawano. Acoustic-based security system: Towards robust understanding of emergency shout. *In Proc. Fifth International Conference on Information Assurance and Security (IAS 2009)*, 2009. (invited).

[3] Eri Ohmura, Hiroaki Nanjo, Hiroshi Kawano, and Takanobu Nishiura. Fundamental study of automatic gender detection from shout for acoustic-based security system. *10th Western Pacific Acoustics Conference (WESPAC X)*, 2009.

[4] Hiroshi Kawano, Masanori Morise, Takanobu Nishiura, and Hiroaki Nanjo. Fundamental study of radiation characteristics of shouted speech for shouted speech detection towards acoustic-based security system. *10th Western Pacific Acoustics Conference (WESPAC X)*, 2009.

[5] 南條浩輝, 三上紘輝, 川野弘, 西浦敬信. 音響防犯システムのための叫び声認識の基礎的検討. 日本音響学会研究発表会講演論文集, 1-11-18, 秋季, 2008.

[6] 南條浩輝, 国松卓, 川野弘, 中山雅人, 西浦敬信. 音響防犯システムのための叫び声の基礎的検討. 日本音響学会研究発表会講演論文集, 1-Q-17, 春季, 2008.

Copyright © 2010 Information Processing Society of Japan.
All Rights Reserved.