2006－SLP－64（2）
2006／12／21

Multi-class AdaBoostを用いた雑音検出

三宅 信之† 滝口 哲也† 有木 康雄†

† 神戸大学工学部情報知能工学科 〒657-8501 神戸市灘区六甲台町1－1
E-mail: †miyake@me.cs.scitec.ac.jp, ††{takigu,ariki}@kobe-u.ac.jp

あらまし 雑音が混入することで音声認識率は低下してしまう．そのため雑音抑圧を行うことが重要であるが，雑音抑圧には雑音を推定する必要がある．しかしながら，発話中に突如雑音が発生する場合，雑音を推定するのは困難であり抑圧も行いにくい．本稿では発話中に突如発生したような雑音（突発性雑音）をMulti-class AdaBoostで検出し，同時に雑音の種類を識別する方法を提案する．評価実験の結果，音声に重畳した信号対雑音比（SNR）5 dB～-5 dBの継続時間が200 ms程度の雑音を高い精度で検出，識別できることを確認した．

キーワード 雑音検出，AdaBoost，Multi-class

Noise Detection with Multi-class AdaBoost

Nobuyuki MIYAKE†, Tetsuya TAKIGUCHI†, and Yasuo ARIKI†

† Department of Computer and System Engineering, Kobe University Nada, Kobe 657-8501, Japan
E-mail: †miyake@me.cs.scitec.ac.jp, ††{takigu,ariki}@kobe-u.ac.jp

Abstract A noise signal decreases speech recognition rate. Therefore, noise reduction is important, and it needs to estimate the noise signal. However, estimating noise is difficult when the noise happens suddenly in a speech. We proposed the method for detecting and identifying the noise in a speech, where it happens suddenly. Its effectiveness is confirined at SNR -5 ～ 5dB for the noise duration time 200 ms.

Key words noise detection, AdaBoost, Multi-class

1. はじめに

音声認識技術を使用するとき、発話に雑音が重畳することで誤認識を引き起こすことが少なくない．そのためスペクトラルサブトラクションをはじめとした雑音を除去する研究が数多くなされている．

雑音除去は基本的に雑音を推定し，その推定された雑音を雑音重畳音声から減算するという手順で行われる．雑音の推定には発話直前の雑音のみの区間や，その情報を確率的に追跡していくものが用いられることが多い．雑音は時間的に緩やかに変化するものだと考えると，発話付近の雑音の情報を使用することで雑音抑圧は高い効果が得られると期待できる．

しかし例えば家の中のような実環境で使用することを考えるとき，雑音には電話の音など中には突然発生するものも少なくない．発話中に雑音が発生した場合，雑音の情報のみを取り出すのが困難であり，抑圧も難しい．図1は極端な例だが，音声中に電話のコール音が重畳している波形である．

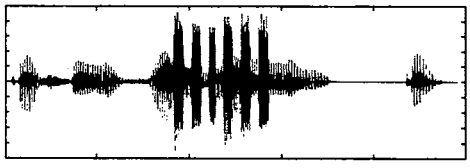

図1 音声に電話音が重畳した波形

このように発話中に雑音が突如混入した時，たとえ雑音が短時間しか継続しない場合であっても音声認識率は低下する．

本稿ではこのように発話中に発生する雑音に対応

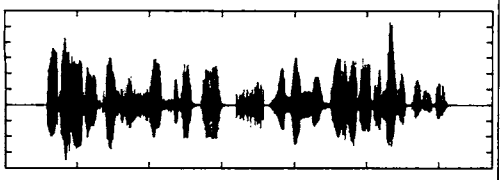

図 2　音声に各雑音が重畳した波形 (SNR 5 dB)

するための前処理として，対数メルフィルタバンク (FBANK) に Multi-class AdaBoost を用いることで，発生した雑音とその区間を検出し，同時に雑音の種類を識別する手法を提案する．

2. 雑音対策の研究

従来雑音除去に用いられる推定雑音は，非音声区間や，または非音声区間と抑圧するフレームから推定されることが多い．また，追従能力を有する雑音推定法として，最小統計量に基づく雑音推定法もよく用いられる [1]．定常的な雑音，時間的に緩やかに変化する雑音には非常に有効な手法であり，雑音除去によく使用される推定法であるが，これらの手法を用いて図 1 のように音声波形上のみに存在しているような雑音を推定するのは困難である．

このような雑音下での認識を行う場合 HMM 合成を行うことも考えられる [2] [3]．しかし HMM 合成法を行うときも，あらかじめどのような雑音が重畳するかを特定しておかなければ組み合わせの数が増え，認識に時間がかかるといった問題がある．

3. AdaBoost による雑音検出手法

音声信号中に混入した雑音を検出する手法について述べる．図 1 のように極端に SNR が悪い場合，パワーを調べるだけである程度の検出はできる．しかし図 2 の波形には SNR 5dB のスプレー，紙を破る，電話のコール音の 3 種類の雑音が重畳している区間があるが，それらをパワーで検出するのは不可能である．またスプレー，電話のコール音は完全に音声区間中に存在している．本研究ではそのような雑音を検出することを目的としている．

3.1 AdaBoost

AdaBoost は二値判別問題に対して強力な手法であり，判別性能の低い複数の弱識別器の重み付き多数決によって最終的な結果を出力する Boosting と呼ばれる

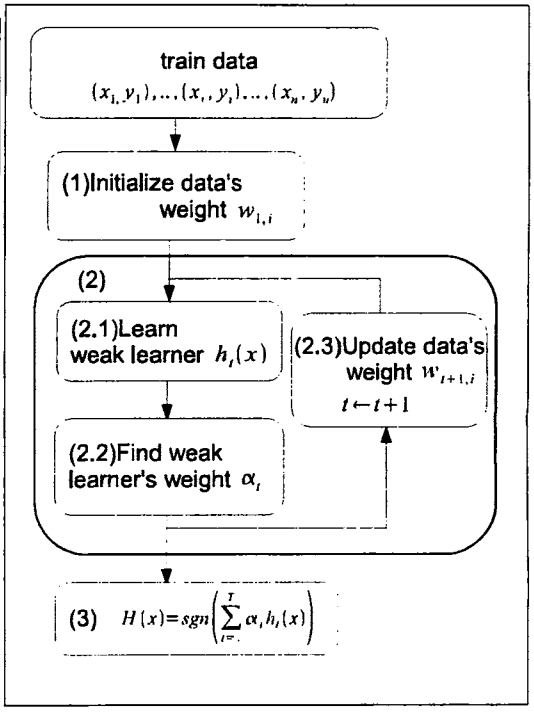

図 3　AdaBoost の概要

手法のひとつである．AdaBoost では弱識別器を学習した後，その識別器で誤識別を起こしたデータの重みを大きくする．その重みを用いて新しい弱識別器を作成し，再び重みを更新する．こうして弱識別器を自動で生成していき，最後に弱識別器の重み付き多数決で，最終的な識別器を作成するという手順となる．

高精度で高速のため，画像情報から顔などのオブジェクト検出によく用いられている [4]．また音声分野でも音声区間検出などに適応されている例もある [5] [6]．

本稿では AdaBoost を用いて突発性雑音の検出を試みる．以下にそのアルゴリズムを示す (図 3)．

特徴ベクトルを x，ラベルを $y = \{-1, 1\}$ として，学習データ　$(x_1, y_1), \ldots, (x_n, y_n)$ を与える．

(1) 重みの初期化

for　$i = 1, \ldots, n$

$if\ y_i = +1$

$$w_{1,i} \leftarrow \frac{1}{2m}$$

$else$

$$w_{1,i} \leftarrow \frac{1}{2l}$$

ここで m は $y_i = 1$ となるデータの数，l は $y_i = -1$ となる数である．$w_{t,i}$ は t 個目の弱識別器を学習する時の i 番目のデータの重みになる．

(2) 学習

for $t = 1, ..., T$

(2.1) エラーが最小となるよう弱識別器 $h_t(x)$ を学習
ただし $h_t(x) = \{-1, 1\}$

(2.2)$h_t(x)$ の重み付き誤差 ϵ_t の計算

$\epsilon_t \leftarrow 0$

$for \quad i = 1, ..., n$

$if \quad h_t(x_i) \neq y_i \quad \epsilon_t \leftarrow \epsilon_t + w_{t,i}$

(2.3)t 個目の識別器の重み α_t 決定

$$\alpha_t \leftarrow \frac{1}{2} \log \left(\frac{1 - \epsilon_t}{\epsilon_t} \right)$$

(2.4) データの重み $w_{t,i}$ の更新

$for \quad i = 1, ..., n$

$$w_{t+1,i} \leftarrow \frac{w_{t,i} \exp\left(-\alpha_t(x) y_i h_t(x)\right)}{\sum_{i=1}^{n} w_{t,i} \exp\left(-\alpha_t(x) y_i h_t(x)\right)}$$

(3) 最終出力

$$H(x) = \mathrm{sgn}\left(\sum_{t=1}^{T} \alpha_t h_t(x) \right) \qquad (1)$$

弱識別器 $h_t(x)$ は x の各次元において重みつきエラーが最小になるようしきい値を設定し，その中でさらに重みつきエラーが最小となる次元を選択した．

3.2 Multi-class AdaBoost

上記の AdaBoost は基本的に二クラス判別である．雑音の除去まで考えるのであれば，雑音の区間検出を行うだけではなく，どのような雑音が混入したのかまでを知ることで，雑音抑圧時にあらかじめ保存しておいた雑音ごとのデータを用いて雑音抑圧が可能と考えられる．そこで AdaBoost を多クラス問題に適応できるように拡張し，雑音の種類の識別まで行えるようにする．

1 クラス対その他のクラスの二値判別器を複数作成し，最も結果の高かったものを識別結果とすることで Multi-class を実現した．具体的には以下のアルゴリズムになる．

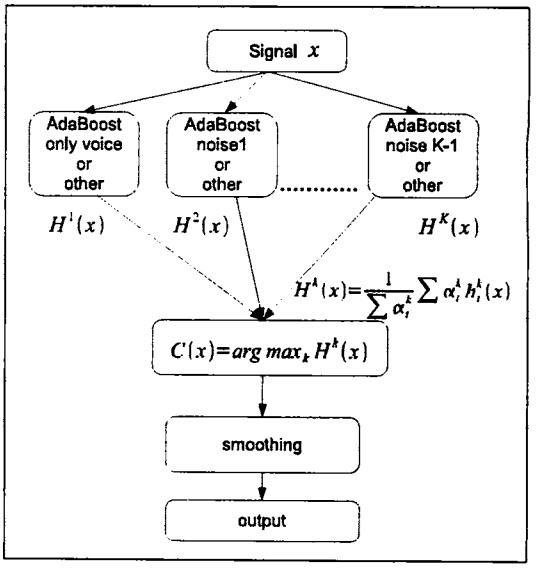

図 4 提案手法の概要

(1) 学習

K クラスの学習データ $(x_1, y_1), ..., (x_n, y_n)$
$y_i = \{1, ..., K\}$

$for \quad k = 1, ..., K$

(1.1) データのラベル付け

$if \quad k = y_i$

$y_i^k \leftarrow +1$

$else$

$y_i^k \leftarrow -1$

(1.2)$(x_1, y_1^k), ..., (x_n, y_n^k)$ を用い AdaBoost を学習
ただし式 (1) を以下の式に変更する

$$H^k(x) = \left(\frac{1}{\sum_{t=1}^{T} \alpha_t^k} \sum_{t=1}^{T} \alpha_t^k h_t^k(x) \right)$$

(2) クラス識別

$$C(x) = \underset{k}{\mathrm{argmax}}\, H^k(x)$$

この Multi-class AdaBoost にフレームごとに特徴量である対数メルフィルタバンクを入力し，音声かその他か，雑音 1 かその他か，雑音 2 かその他か，というように識別していく．

3.3 平 滑 化

本稿では継続時間が 200 ms 以上の雑音を対象としているため，それより明らかに小さい継続時間で雑音

と判定されるものは湧き出しと考える．また雑音と判定されるフレームが連続している中で，ごくわずかに音声だと判定されるフレームが存在した場合，そのフレームは音声ではなく誤識別を起こした雑音だと考えられる．

そのようなフレームに対し，平滑化をかけてやることでミスを取り除く．

平滑化処理は前後3フレームとそのフレームの計7フレームの中で最も多い出力結果をそのフレームの出力結果とし、変更がなくなるまで繰り返す．

連続して雑音と判定されている区間はフレームごとに出力結果が違ってもひとつの雑音とみなす．その後，継続時間が200 msの半分である100 ms以下の雑音を切り取る．提案手法の概要を図4に示す．

4. 評価実験

提案手法の再現率，適合率および雑音の識別率を調べるため評価実験を行う．

4.1 実験条件

発話データにはASJから提供されている研究用連続音声データベースから学習データとして男性話者21人×10発話，評価データとして男性話者5人×平均240発話を用いた．

また雑音はRWCPの提供する非音声ドライソースの中から，電話の音，紙をやぶく音，スプレーの三種類を使用した．

AdaBoostの学習には学習データとして，音声データと，さらにその音声データに各雑音をSNRを調整して重畳させたものを用いた．学習データのSNRは-5 dBから5 dBの間でランダムに変化させた．

評価データには，1発話に200 ms以上の継続時間のSNRを調整した雑音を1〜3つ重畳させる．ただし雑音が重畳した区間に，さらに別の雑音が重畳するようなデータは存在しない．評価用データのSNRは-5 dB，0 dB，5 dBの3つである．

SNRは今回の実験では，以下のように求めた．

音声 $s(t)t = 1,....,T_s$，雑音 $n(t)t = 1,....,T_n$ に対し

$$sMeanPower = \frac{\sum_{t=1}^{T_s} |s(t)}{T_s}$$

$$nMeanPower = \frac{\sum_{t=1}^{T_n} |n(t)}{T_n}$$

表1 提案手法の検出率，再現率，適合率

SNR	検出率(%)	適合率(%)	再現率(%)
-5dB	97.7	95.7	99.9
0dB	97.5	95.7	99.7
5dB	95.5	95.8	98.9

$$SNR = 10\log\left(\frac{sMeanPower}{nMeanPower}\right)$$

特徴量には対数メルフィルタバンクを使用した．学習，テストサンプルともにフレーム幅20 msフレームシフト10msであり，$1-0.97z^{-1}$ のプリエンファシス，ハミング窓を用いている．

4.2 雑音検出

検出という観点のみから，区間が正しく検出できているものは，雑音の種類が異なっていたとしても正解と判定する．また，誤差のマージンを決めておき正解データとの誤差がそのマージン以内であるものも正解とする．なおマージンは今回の実験では30 msとした．

また検出区間が大きすぎるものは誤検出，検出区間が小さすぎるものは未検出とした．

評価には検出率 (Detection rate)，再現率 (Recall rate)，適合率 (Precision rate) の3つを使う．それぞれ検出した区間の中で正解した数 Tp，誤検出数 Fp，未検出数 Tn，雑音の総数 Ta を用いて以下の式で計算した．

$$DetectionRate = \frac{Tp}{Ta}$$

$$RecallRate = \frac{Ta - Tn}{Ta}$$

$$PrecisionRate = \frac{Tp}{Tp + Fn}$$

本来，検出率と再現率は等しいものだが本稿では区間を大きく取りすぎた雑音を誤検出として評価しているために異なる値が出ており，その両方を示した．結果を表1に示す．

すべてのSNRに対して検出率，再現率，適合率95%以上と良好な結果がでており5dB以上の強さの雑音であれば検出できることが確認できた．

4.3 雑音識別

4.2節においては区間さえ正しければ雑音の種類が違っても正解と判定した．ここでは区間が正しく判定された雑音の中での雑音の識別率を評価し，さらに検

表 2　雑音正解率

SNR	ノイズ識別率(%)	検出率(%)	正解率(%)
-5 dB	99.7	97.7	97.4
0 dB	99.6	97.5	97.1
5 dB	99.7	95.5	95.2

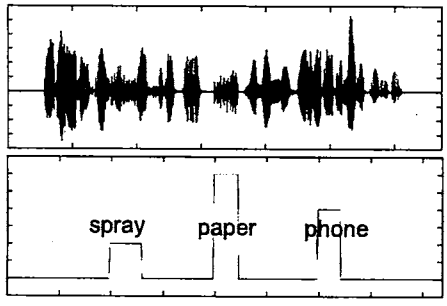

図 5　雑音検出の例

出率とあわせて区間が正しく，かつ雑音の識別結果も正しいものを雑音の正解率として求めた．その結果を表 2 に示す．

表 2 からすべての SNR において，99.5%を超える高い雑音識別率を得ることができていることがわかる．すなわち検出できたもののほとんどは正しく識別できていることになる．

正しく検出，識別できている例を図 5 に示す．波形からでは電話の音が見分けられないが全て正しく検出できている．

4.4　ミスマッチモデルによる検出精度の変化

4.2 節では検出する雑音と学習する雑音の SNR が等しかったが，ここでは学習データとテストデータの SNR を変化させ精度がどの程度変化するかを調べる．

4.2 節と同様に SNR -5～5 dB で学習したモデル，SNR -5 dB のみ，0 dB のみ，5 dB のみのデータを用いて AdaBoost を学習したモデルのそれぞれについて-10 dB～10 dB のテストデータに対し検出率，再現率，適合率を算出し違いを比較する．AdaBoost の学習回数は同様に 1000 回である．結果を表 3 に示す．

適合率は，テストデータの雑音の SNR を変えても誤検出数には影響しなかったため，学習時に使用する SNR によってほぼ決定する．学習データの SNR が低くなるほど高くなる結果となった．

表 3　学習 SNR による検出精度の変化

学習SNR		テストデータのSNR -10	-5	0	5	10
SNR-5～5 (4.2と同様)	検出率	97.6	97.7	97.5	95.5	82.7
	適合率	95.7	95.7	95.7	95.8	95.4
	再現率	99.9	99.9	99.7	98.9	84.2
SNR-5	検出率	99.6	99.3	96.6	76.7	39.0
	適合率	99.6	99.6	99.6	99.5	99.2
	再現率	99.8	99.6	96.8	76.7	39.2
SNR0	検出率	98.6	98.7	98.5	93.4	69.9
	適合率	97.7	97.7	97.8	97.7	97.2
	再現率	99.7	99.7	99.5	94.4	70.7
SNR5	検出率	95.5	95.7	95.8	95.5	87.7
	適合率	91.4	91.5	91.7	91.7	91.1
	再現率	100.0	99.8	99.7	99.4	91.3

表 4　学習 SNR による識別精度の変化

学習SNR		テストデータのSNR -10	-5	0	5	10
SNR-5～5	検出率	97.6	97.7	97.5	95.5	82.7
	識別率	98.0	99.7	99.6	99.7	98.5
	正解率	95.6	97.4	97.1	95.2	79.8
SNR-5	検出率	99.6	99.3	96.6	76.7	39.0
	識別率	99.6	99.9	99.5	97.7	94.4
	正解率	99.1	99.2	96.1	74.9	36.8
SNR0	検出率	98.6	98.7	98.5	93.4	69.9
	識別率	93.7	99.3	99.5	98.5	95.9
	正解率	92.4	97.9	98.0	92.0	67.0
SNR5	検出率	95.5	95.7	95.8	95.5	87.7
	識別率	80.1	93.3	98.4	98.5	96.6
	正解率	76.5	89.2	94.3	94.1	84.6

また，テストデータの SNR が高くなるほど未検出数が増える傾向がある．表 3 より学習 SNR -5 dB の識別器を用いたとき，テスト SNR5 dB において検出率は 76.7%，学習 SNR 0 dB の識別器では，テスト SNR 10 dB のにおいて検出率 69.9%まで下がる．4.2 節のように全てについて学習した識別器では-5 dB，0 dB と比べ適合率が下がるが，検出率，再現率の減少量は少なかった．

4.5　ミスマッチモデルによる雑音識別精度の変化

4.4 節と同様の条件で雑音の識別率，正解率を評価する．結果を表 4 に示す．

表 4 より学習した雑音の SNR とテストデータの SNR の差が大きいほど，識別率は低下するという結果になった．SNR -5 dB で学習した識別器を用いたときのテスト SNR 10 dB の識別率は 94.4%と減少はするものの高い値を示しており，SNR 5 dB で学習した識別器を用いたとき，テスト SNR -10 dB の識別率は 80.1%と比較的低い値となった．また 0 dB で学習した識別器を

見ると，テストSNR 10 dBでは95.9%，テストSNR -10 dBでは93.7%となった．

また雑音の正解率はモデルマッチのものが高い値を示すが，平均的に見ると-5〜5 dBで学習したものが一番高い値を示した．

5. 終わりに

本稿ではAdaBoostを用いた音声中に存在する雑音の検出手法について提案した．雑音の種類が少なく，未知の雑音のない環境下であったため検出，識別ともに良好な結果を得ることができた．今後は検出させる雑音の増加，さらに短時間の突発性雑音の検出，検出した雑音の除去を行っていく予定である．

文　献

[1] V.Stahl, A.Fischer, and R.Bippus, "Quantile based noise estimation for spectral subtruction and Wiener filtering", Proc. ICASSP 2000, pp.1875-1878, May 2000

[2] 三木一浩, 西浦敬信, 中村哲, 鹿野清宏, "HMMを用いた環境音識別の検討", 電子情報通信学会音声研究会, SP99-106, pp.79-84 (1999-12)

[3] 伊田政樹、中村哲、"雑音DBを用いたモデル適応化HMMのSN比別マルチパスモデルによる雑音下音声認識" 電子情報通信学会技術報告, Vol.101, No.522, pp.51-56, 2001-12

[4] Paul Viola and Michael Jones:"Rapid Object Detection using a Boosted Cascade of Simple Features". IEEE CVPR, vol.1, pp.511-518, 2001.

[5] Kwon, O., Lee, T.: "Optimizing speech/non-speech classifier design using adaboost" Proc. IEEE ICASSP 2003, pp I-436–I-439. pp. Apr. 2003

[6] 松田博義, 滝口哲也, 有木 康雄: "Real Adaboostによる音声区間検出", 日本音響学会2006年秋季研究発表会, 2-P-12, pp.117-118, 2006-09.