社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.

# 弱識別器にSVMを用いたAdaBoostの検討

松田　博義†　　滝口　哲也††　　有木　康雄††

† 神戸大学工学研究科　〒 657-8501 兵庫県神戸市灘区六甲台町 1-1
†† 神戸大学工学部　〒 657-8501 兵庫県神戸市灘区六甲台町 1-1
E-mail: †matsuda@me.cs.scitec.kobe-u.ac.jp, ††{takigu,ariki}@kobe-u.ac.jp

**あらまし**　雑音が重畳されている音声から，音声・非音声の識別を行ない，音声区間のみを検出する，音声区間検出 (VAD: Voice Activity Detection) を行なうことは，音声認識を行なううえで非常に重要である．本研究では，音声区間検出法において，音声・非音声の識別を行なう識別器に，SVM を弱識別器とした AdaBoost を提案する．AdaBoost とは弱識別器を線形結合する事により，より高い識別率をもつ強識別器を構成する手法である．その弱識別器に，カーネルトリックやマージン最大化により高度な識別を行うことができる SVM を用いることにより，SVM のもつ汎化能力を保ったまま，より高度な識別を行なうことが期待できる．提案手法と，単一で SVM を用いた場合，CART を弱識別器とした AdaBoost を用いた場合とを，区間検出評価用データベース CENSREC-1-C 上で比較し報告する．
**キーワード**　AdaBoost，SVM，音声区間検出，CENSREC

# An Investigation for AdaBoost using SVM as Weak Learner

Hiroyoshi MATSUDA†, Tetsuya TAKIGUCHI††, and Yasuo ARIKI††

† Graduated School of Science and Technology, Kobe University　Rokkodaicho1-1, Nada-ku, Kobe, Hyogo 657-8501 Japan
†† Faculty of Engineering, Kobe University　Rokkodaicho1-1, Nada-ku, Kobe, Hyogo 657-8501 Japan
E-mail: †matsuda@me.cs.scitec.kobe-u.ac.jp, ††{takigu,ariki}@kobe-u.ac.jp

**Abstract**　VAD (Voice Activity Detection) by separating of speech and non-speech from noisy speech is an important probrem for speech recognition. The proposed method constructs AdaBoost using SVM as weak learners for separation of speech and non-speech. AdaBoost is an iterative algorithm that combines simple classification rules with ‘ mediocre ’performance in terms of misclassification error rate to produce a highly accurate classification rule. Though AdaBoost generally takes CART as weak learners, the proposed method takes SVM with high robustness and classification capacity for speech detection as weak learners. We report the experimental results that compared single SVM, AdaBoost with CART and the proposed method.
**Key words**　AdaBoost, SVM, Voice Activity Detection, CENSREC

## 1. はじめに

観測された音声信号から，目的音声の区間を検出する音声区間検出 (VAD : Voice Activity Detection) は，音声認識など，音声情報処理の前処理として行なわれる重要な技術である．一般に VAD は，音声信号からの音声特徴量抽出部と，得られた特徴量をもとに音声・非音声の判定を行なう識別部からなる．音声特徴量としては，音声・非音声信号のエネルギー比，ゼロ交差数が用いられるが，非音声が存在する環境では音声・非音声の違いを表現することができず、音声区間の検出を行なうことができない．このため，非音声に対して頑健な音声特徴

量がいくつか提案されている (***ここにいくつか参考文献***). しかし，これらの手法も一定の効果を上げているものの，高いエネルギーを持つ非音声が入力された際に，本来抽出されるべき音声の特徴が非音声の中に埋もれることにより，特徴量による音声・非音声の差別化が曖昧になり，VAD の性能が劣化してしまう．そこで，識別部を改善し曖昧な特徴量を正しく音声・非音声に識別することできれば，VAD の精度向上を期待できることから、本研究では音声・非音声の識別器の構成について検討を行なう．

VAD は，識別器が事前に雑音重畳音声・雑音のモデルをもっており，観測信号がそれぞれのモデルに属する確率の比 (尤度比) を求め，閾値処理を行なうことにより音声・非音声の識別を行なうことが一般的である．統計学一般に良く用いられている識別器としては二分木，k 近傍識別器，正規分布, GMM, SVM, AdaBoost 等様々な手法があるが，音声区間検出においては正規分布，GMM が用いられることが多い．しかし，正規分布及び GMM は，学習データをすべて均一に学習する為，識別が困難なサンプルの情報が他のデータに埋もれてしまう，識別ベースでの学習を行なわないため，識別とは無関係の領域に重点を置いてしまう可能性がある等の問題がある．我々は以前，これらの問題を解決するため，識別ベースの学習を行い，識別が困難なデータに重み付けをして学習を行なう識別器，AdaBoost を用いた音声区間検出法 [5] を提案した．AdaBoost は，Boosting と呼ばれる手法の一つで，いくつもの弱識別器を組み合わせてひとつの高度な識別器 (強識別器) を構成するアンサンブル学習法として近年注目を集めている．図 1 は AdaBoost の概要図である．

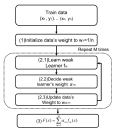

図 1　AdaBoost の概要

AdaBoost では $m$ 段目に誤識別を起こしたデータに対して，$m$ 段目の識別木の精度に応じて高い重みを与え，$m+1$ 段目で重点的に学習を行なう．これにより最終学習仮説の学習誤差を小さくすることができる．弱識別器としては CART (Classification And Regression Trees) を用いることが一般的である．しかし，CART には，一つあたりの識別精度が悪い，汎化能力が低いといった問題点が挙げられる．AdaBoost において，前者は必ずしも問題であるとは言えないが，後者は VAD の実環境での使用を考えた際には問題となる．

AdaBoost 以外の識別ベースの識別器としては，SVM (Support Vector Machine) が挙げられる．SVM は識別能力，汎化能力ともに高いが，計算量が多い，学習データに対する識別率を高くするよう設計すると汎化能力が低下してしまうといった問題点がある．そこで SVM の汎化能力を低下させず，識別率を向上させる手法として，SVM を AdaBoost の弱識別器として用いることを提案する．提案手法により，SVM を普通に用いた場合や，弱識別器に CART を用いた AdaBoost よりも，識別能力・汎化能力が共に高く，環境や突発的な雑音に対して頑健な VAD を実現することが期待できる．区間検出評価用データベース CENSREC-1-C を用いることにより，提案手法の効果を評価する．

## 2. 提 案 手 法

### 2.1 SVM (Support Vector Machine)

SVM (Support Vector Machine) [9] は Vladimir N. Vapnik によって提案された 2 クラス分類器の一種である．SVM は $x$ を $y \in \{+1, -1\}$ に分離する平面を構成し，その分離平面に最も近い例 (Support Vector) 同士のマージン (Support Vector と分離平面の最小距離) を最大化にするよう学習が行なわれる．しかしこのままでは線形分離不可能なデータに適用することができず用途が限られてしまうため，元のデータ $x \in R^D$ を高次空間 $\phi(x) \in R^H$ に写像することにより，高次空間上にて線形分離を行なう．これにより，もとの空間では非線形な識別を行なっていることになる (図 4, 5 参照).

具体的には以下の様に記述される．二つのクラスに属する学習データのベクトル集合を

$$(x_1, y_1), \ldots, (X_n, Y_n) \quad x_i \in R^D, y_i \in \{+1, -1\} \tag{1}$$

とする．この時，SVM の分離関数は次式で与えられる．

$$y = sgn(\sum_{i=1}^{n} y_i \alpha_i \phi(x_i) \cdot \phi(x) + b) \tag{2}$$

ただし，

- $\phi$ は $R^D$ から $R^H$(一般に $D \ll H$) への写像関数
- $\alpha_i, b \in R, 0 \leqq \alpha_i \leqq C$

である．式 (1) は，テストデータを高次元空間に写像した

データ $\phi(x)$ と $n$ 個の学習データを写像したデータ $\phi(x_i)$ との内積を全事例についてそれぞれ計算し，重み $y_i\alpha_i$ と線形結合した形となっている．高次元空間への写像関数 $\phi$ はすべての事例が，$R^H$ における超平面で分離できるように設計する．

式 (1) より，実際に SVM を構成する場合に写像関数は，$\phi(x_i)\cdot\phi(x)$ という内積の形で出現する．SVM における学習の式も，写像関数の内積のみで表現できるので，高次空間上での内積 $\phi(x)\cdot\phi(x')$ を効率よく計算するため，カーネル関数 $K(x_1,x_2)=\phi(x_1)\cdot\phi(x_2)$ を用いる．カーネル関数を用いて式 (1) は以下のように書き換えられる．

$$y = sgn(\sum_{i=1}^{n} y_i\alpha_i K(x_i,x)+b) \tag{3}$$

式 (3) より，SVM の学習，分類には，事例間の類似度 (内積) を与えるカーネル関数のみを定義できれば良いことが分かる．この手法は，実際にはどのような空間への写像を行なっているか分からないことから，カーネルトリックと呼ばれる．カーネル関数は多くの種類があり，中でも代表的な手法としてはシグモイドカーネル $k(x,y)=1/(1+\exp(-\beta x\cdot y))$ や多項式カーネル $(x\cdot y+c)^p$，RBF (Radial Basis Function) カーネル $k(x,y)=\exp(-\beta\parallel x-y\parallel^2)$ 等がある．

SVM における分離平面は，サポートベクターによってのみ決定される．すなわち，SVM はクラス間のマージンを最大にするサポートベクターを推定する手法として捉えることができる．これにより，$y=\{+1,-1\}$ の分離に必要なデータのみを見て，学習データの識別に最適な識別平面を引くことができる．

これより，SVM は適切な分類を行ないつつ，汎化誤差を小さくすることができるが，問題点がないわけではない．まず，SVM であっても識別誤りを避けることはできない．しかし，学習時における識別誤りを最小化する様に SVM の設計を行なうと，未知のデータに対する汎化性能が大きく低下してしまう．そして，二つ目の問題点は識別速度である．カーネル関数を用いた場合の識別速度はサポートベクターの数に依存するため，大規模なデータの解析には非常に多くの時間がかかってしまい適用が難しくなる．そこで，汎化性能を低下させず識別誤りを減少させ，かつ識別速度を大きく低下させないよう SVM を設計する手法として，AdaBoost を用いて SVM の学習，識別を行なう手法を提案する．まず，次の章にて AdaBoost に関する説明を行なう．

### 2.2 AdaBoost

AdaBoost とは，Boosting と呼ばれる手法の一つで，精度の低い学習器械 (弱学習器) を組み合わせることで精度の良い学習器械 (強学習器) を構成するための手法である．AdaBoost は Boosting のなかでも顕著な性能を示す手法として注目されている．AdaBoost の特徴は，

- 逐次的に学習器械を構成
- 学習データへの重み付け
- 弱学習器の重みつき結合

が挙げられる．$X$ を特徴空間として $x\in X$ から $y\in\{+1,-1\}$ を予測する器械を例題から学習することを目的としている．Schapire らが提案している，Discrete AdaBoost と呼ばれる学習のアルゴリズム [6] を以下に示す．

（1） $n$ 個の学習サンプル集合 $Z$ を用意する．$Z=\{(x_1,y_1),\ldots,(x_n,y_n)\}$

（2） 各学習サンプルに対する重みを $w_{1i}=\frac{1}{n}$ で初期化する．

（3） $m=1,\cdots,M$ まで，以下を繰り返す．

- 重み付き学習サンプル $x_i$ に対して弱識別器 $f_m(x)\in\{+1,-1\}$ を構成する．
- $f_m(x)$ を用いて学習誤り $err_m$，及び弱識別器への重み $\alpha_m$ を計算する．

$$err_m=\sum_{i=1}^{n} w_{mi}I\{f_m(x_i)\neq y_i\} \tag{4}$$

$$\alpha_m=\log\frac{1-err_m}{err_m} \tag{5}$$

- 各データ $x_i$ の重みを更新する．

$$w_{(m+1)i}=\frac{w_{mi}e^{\alpha_m I\{f_m(x_i)\neq y_i\}}}{\sum_{r=1}^{n} w_{mr}e^{\alpha_m I\{f_m(x_r)\neq y_r\}}} \tag{6}$$

（4） 弱識別器の線形結合により，強識別器 $F(x)$ を得る．

$$F(x)=\sum_{m=1}^{M}\alpha_m f_m(x) \tag{7}$$

式 (5) より，学習サンプルの過半数を正しく識別することができれば，すなわち式 (4) における $err_m$ が 0.5 未満であれば，$\alpha_m$ は正の値をとり，$err_m$ が小さくなるほど，$\alpha_m$ は大きな値をとるようになる．これにより誤りの少ない，正確な弱学習器ほど最終的な貢献度が高くなる．さらに式 (6) により，$m$ 段目で誤ったデータに対して $\alpha_m$ に応じて高い重みを与え，$m+1$ 段目で重点的に学習を行なうことにより，最終的に正しく識別できるようにする．また，未知のサンプルに対する汎化誤差も小さくできることが実験的に報告されている [6] [8]．AdaBoost にはここで示した Discrete AdaBoost 以外にも，弱識別器が実数値を返すものに変更した Real AdaBoost，他 Gentle AdaBoost，LogitBoost など様々なものが考案されている [7].

一般に弱識別器には CART (Classification and Regression Trees) や Decision Stump が用いられることが

多い．CART とは次元の値の大小に基づき，決定木を作成し分類を行なう手法である．CART を弱識別器に用いた AdaBoost では，データへの重み付けを変えて複数個の CART の学習を行い，それらを精度に応じた重みを与えて線形結合することで最終的な出力を得る．

AdaBoost を用いた音声区間検出法は，音声データを $y = 1$，非音声データを $y = -1$ として学習を行う．そして，式 (7) において，テストデータ $x$ が入力された際の $F(x)$ の値を閾値処理することにより，音声・非音声の判定を行う．

### 2.3 SVM を弱識別器に用いた AdaBoost

ここで，CART と SVM の識別手法を図例で示す．ここでは，もっとも単純な 2 種類の分類問題について考えてみる．すなわち，図 2 のように，□のデータと○のデータを分類する問題について考える．分類の様子を分かりやすく示す為，これらのデータは，$x = [x_1, x_2]$ と表されるような 2 次元のデータとする．

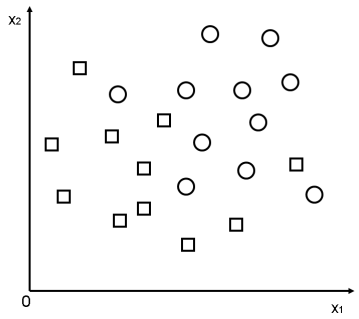


図 2　2 クラス分類問題

この時，CART では最もよくデータを分離できる次元から閾値を設定していき，入力されたデータがどの領域に属するかで，識別を行なう．図 3 に識別例を示す．

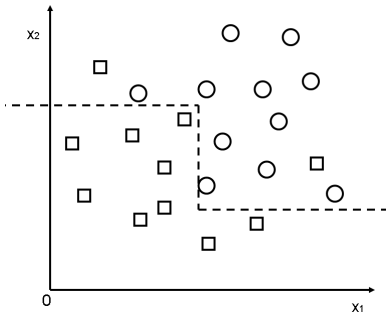


図 3　CART による識別境界

次に，SVM では図 4 のように高次元にデータを写像し，そこでマージン最大化による識別平面が引かれる．もとの次元に戻したときには図 5 のように非線形の分類を行なっていることになる．尚，図 4 において塗りつぶしてある点はサポートベクトルである。

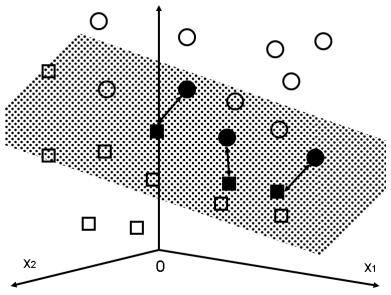


図 4　高次空間における SVM による識別境界

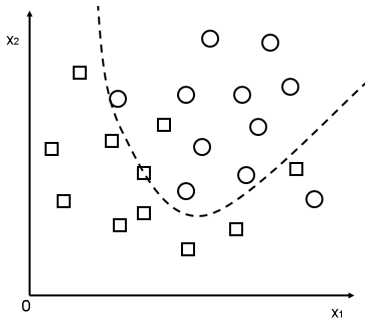


図 5　元の入力空間における SVM による識別境界

これらより SVM は CART より高度な識別・分類を行なえることがわかる．そこで，AdaBoost の弱識別器として SVM を使うことを考える．具体的なアルゴリズムは以下のようになる．

（1） $n$ 個の学習サンプル集合 $Z$ を用意する．$Z = \{(x_1, y_1), \ldots, (x_n, y_n)\}$

（2） 各学習サンプルに対する重みを $w_{1i} = \frac{1}{n}$ で初期化する．

（3） $m = 1, \cdots, M$ まで，以下を繰り返す．

- $Z$ から $\eta$ 個の $x_i$ を重みに応じてランダムにリサンプリングし $Z'$ とする．$Z'$ を用いて $f_m(x) \in \{+1, -1\}$ の学習を行なう．ここで $f_m(x)$ は SVM である．
- $f_m(x)$ を用いて学習誤り $err_m$，及び SVM への重み $\alpha_m$ を計算する．

$$err_m = \sum_{i=1}^{n} w_{mi} I\{f_m(x_i) \neq y_i\} \tag{8}$$

$$\alpha_m = \log \frac{1 - err_m}{err_m} \tag{9}$$

- 各データ $x_i$ の重みを更新する．

$$w_{(m+1)i} = \frac{w_{mi} e^{\alpha_m I\{f_m(x_i) \neq y_i\}}}{\sum_{r=1}^{n} w_{mr} e^{\alpha_m I\{f_m(x_r) \neq y_r\}}} \tag{10}$$

（4） SVM の線形結合により，強識別器 $F(x)$ を得る．

$$F(x) = \sum_{m=1}^{M} \alpha_m f_m(x) \tag{11}$$

SVM には AdaBoost が与える重みを直接用いることができないため，ここでは重みに応じたリサンプリングを行なうことにより，間接的に重みを用いた．これにより，前段で検出できなかったデータを優先して用いることになる．SVM を AdaBoost の枠組みで用いることにより，汎化誤差を保ったまま，より高い識別能を持つことができる．さらに，$\eta < n$ とすることにより，SVM 一つあたりの学習，及び識別時にかかる計算量を減らすことができる．

## 3. 評価実験

### 3.1 実験条件

提案手法の評価データには，VAD の評価用に設計されたデータベース CENSREC-1-C [10] を用いた．CESREC-1-C には人工的に作成したシミュレーションデータと，実環境にて収録されたデータの 2 種類が含まれている．ここでは，実際の環境での VAD 性能を評価するため，実環境にて録音されたデータを用いて実験を行なった．実データは，男声 4 名・女性 5 名が 1 12 桁の連続数字を 1 ファイルにつき 8 10 回，約 2 秒間の間隔で発声したものが各話者につき計 4 ファイル (一人につき 38 39 発話) ある．これがそれぞれ，高速道路付近 (高 SNR)，高速道路付近 (低 SNR)，学生食堂 (高 SNR)，学生食堂 (低 SNR) の計 4 環境で録音されている．

雑音として学習に用いたデータは，CENSREC-1-C より評価に用いられなかった男声の発話データ [10] より非音声部分を切り出したものを用いた．音声として学習に用いたデータは，音声用として雑音環境下連続英語数字音声認識タスクの共通評価データベースである AURORA-2J よりクリーントレーニングに用いられる学習データに，雑音の学習に用いた非音声データを重畳させたもの 8440 発話を用いた．尚，学習データ・評価データともにデータは全て 12,000 Hz にリサンプリングしている．

音声特徴量には MFCC (Mel Frequency Cepstrum Coefficient) を用いた．特徴量計算の際の主なパラメータは，フレーム長 32 ms，シフト長 8 ms である．MFCC と前後 5 フレームから得られた Δ-MFCC，あわせて 32 次元を用いた．

評価は，あらかじめ与えられた音声始端終端の時間ラベルとフレーム単位で照合することにより行なった．評価尺度は次式の $FRR$ (False Rejection Rate) と $FAR$ (False Acceptance Rate) である．

$$FRR = N_{FR}/N_s \times 100[\%] \tag{12}$$

$$FAR = N_{FA}/N_{ns} \times 100[\%] \tag{13}$$

$N_s$ は音声フレーム数，$N_{FR}$ は音声を非音声と検出したフレーム数，$N_{ns}$ は非音声フレーム数，$N_{FA}$ は非音声を音声と検出したフレーム数である．FAR と FRR は，互いにトレードオフの関係にあるので尤度比判定の閾値を調整して複数の実験結果を得，ROC (Receiver Operating Characteristics) 曲線を描くことにより評価を行なう．

### 3.2 実験結果

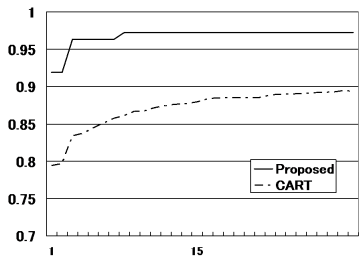


図 6 学習データに対する識別率

図 6 に，学習時における学習データの識別率を示す．横軸がイタレーション数，縦軸が識別率となっており，実線が提案手法の，破線が従来手法である CART を弱識別器に使った際の結果となっている．図 6 より，学習データに対する識別率は提案手法が従来手法を大きく上回っている事が分かる．

次に，実際に評価実験を行なってみたときの実験結果を図 7,8 に示す．ここで，点線にて表されている実験結果は CENSREC-1-C に添付されていた実験結果で，波形の強さを用いて区間検出を行なった場合の実験結果となっている．ROC 曲線は原点からの距離が遠いほど高い性能を示しており，図 7,8 の結果より，二つの雑音環境において提案手法は単一で SVM を用いた場合より良い結果を示していることが分かる．これは単一の SVM だけでは推定しきれない領域を AdaBoost と組み合わせることにより，推定することが可能になったためである．すなわち，SVM はサポートベクターのみから識別平面を得るが，学習データにない，サポートベクターの外側に存在するデータがあった場合，単一の SVM では誤識別してしまう．しかし，AdaBoost にてリサンプリングを行なった際，やや偏ったデータ集合にて識別平面を計算するので，単一の SVM では誤識別してしまうようなデータに対しても，他の SVM に誤りを吸収された結果，正しく

識別できるようになったためであると考えられる．しかし，CART を用いた AdaBoost と比較した場合，図 7 の食堂環境では提案手法が最良の結果を得ているが，図 8 の道路環境において CART に劣ってしまっている．この理由は，道路環境においては音声の領域と非音声の領域が明瞭に分かれるため，CART のようなシンプルな識別器であっても容易に識別が行なえる．その際，学習データに対する重みをダイレクトに反映できる CART のほうが，より適切な分類を行なえる為である．また，実験データに対する識別率が，学習データに対する識別率ほど向上しなかった理由は，AdaBoost の後段において与えられる，識別が困難なデータ群のみの持つ識別境界は，未知のデータの分離を行なう際にあまり重要ではない為である．これは，学習データに対する重みを直接用いることができれば解決できると考えられる．

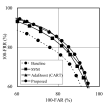

図 7　食堂環境における実験結果

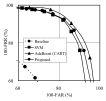

図 8　道路環境における実験結果

## 4. ま と め

本研究では，音声区間検出法として SVM を弱識別器に持つ AdaBoost を識別器として用いることを提案した．そして，提案手法と，SVM を単体で用いたもの，CART を弱識別器とした AdaBoost との比較を行なった．結果，提案手法は，全環境において SVM を単一で用いた場合の識別率を上回ることはできたが，道路環境においては，CART を用いた AdaBoost の結果を上回ることはできなかった．これは，AdaBoost が学習データに与える重みを直接用いることができなかったことに要因があると考えられる．今後は，SVM と CART を組み合わせて AdaBoost を構成する，CART に変わる学習データへの重みを生かすことのできる弱識別器の考案すること等を予定している．

### 文　献


[1] Sohn, J., Kim, N.S., and Sung, W.: “ A statistical model-based voice activity detection,” IEEE Signal Process. Lett., pp.1-3, 1999, 16, (1).

[2] Juan E. Rubio, Kentaro Ishizuka, Hiroshi Sawada, Shoko Araki, Tomohiro Nakatani, and Masakiyo Fujimoto, “Two-microphone voice activity detection based on the homogeneity of the direction of arrival estimates,” ICASSP, pp.385-388, 2007, IV.

[3] J.M. Gorriz, C.G. Puntonet, J. Ramirez, and J.C. Segura: “Bispectrum Estimators for Voice Activity Detection and Speech Recognition ”, Lecture Notes in Artificial Inteligence, pp. 174-185, 2005, No. 817.

[4] Norbert Binder, Konstantin Markov, Rainer Gruhn, Satoshi Nakamura: “ SPEECH NON-SPEECH SEPARATION WITH GMMS ”, 日本音響学会講演論文集, pp. 141-142，2001 年 10 月.

[5] 松田博義，滝口哲也，有木康雄,“ 3 次キュムラントバイスペクトラム特徴と Real AdaBoost による音声区間検,” 日本音響学会講演論文集，pp.187-188, 2007 年 9 月.

[6] R. Schapire, Y. Freund, P. Bartlett, and W. Lee, “Boosting the margin: A new explanation for the effectiveness of voting methods,” Annals of Statistics, pp. 1651-1686, 1998, Oct. vol.26, no.5.

[7] Jerome Friedman, Trevor Hastie, and Robert Tibshirani:“Additive logistic regression: A statistical view of boosting,”The Annals of Statistics, pp.337-407, 2000, Vol. 28, No. 2.

[8] H.Schwenk and Y.Bengio, “Adaboosting neural networks,” Proc. ICANN’97, vol.1327 of LNCS Berlin Springer, pp.967-972, Oct. 1997.

[9] Vladimir N. Vapnik:“ The nature of statistical learning theory,” 2nd ed. New York, p.314, 2000, Springer, xix.

[10] 北岡教英他，“雑音下音声認識評価ワーキンググループ活動報告：認識に影響する要因の個別評価環境，” 電子情報通信学会技術研究報告，pp. 1-6, 2006 年 12 月 21 日.