

## スペクトルサブバンドセントロイドを用いた雑音下での音声認識

柘植 覚†‡ 深田 俊明† シンガー ハラルド†‡ パリワル クルディップ †‡

†ATR 音声翻訳通信研究所  
‡徳島大学  
‡Griffith 大学

†〒 619-02 京都府相楽郡精華町光台 2-2

Tel.: 0774-95-1397 e-mail: {stsuge, fukada, singer}@itl.atr.co.jp, K.Paliwal@me.gu.edu.au

あらまし 音声認識システムを実環境下で使用する場合、背景雑音等で音響モデルの学習環境と認識環境とのミスマッチが生じることが多い。このような環境のミスマッチに対し、ケプストラム等の特徴パラメータは大きく変動してしまうため、認識性能の劣化を引き起こす。本報告では、スペクトル・サブバンド・セントロイド(SSC)を補助的パラメータとして用いることにより、環境のミスマッチによる認識性能の劣化が低減できることを示す。SSCは、ナイキスト周波数をサブバンドに分割したときの各々のサブバンド内のパワースペクトルのセントロイド周波数として定義される。このセントロイド周波数は、雑音下においても変動が少ないため、環境のミスマッチに対してロバストな特徴パラメータであると考えられる。自然発話を用いた音声認識実験結果より、Mel-Frequency Cepstral Coefficients (MFCC)の補助的パラメータとしてSSCとMFCCの二次回帰係数を併用した場合、SNR = 10dBにおいて14.3%、SNR = 20dBにおいて19.9%の誤り改善率を得ることができた。

キーワード スペクトルサブバンドセントロイド, 補助パラメータ, 雑音環境, 音声認識

## Noise environment speech recognition using spectral subband centroids

Satoru Tsuge†‡ Toshiaki Fukada† Harald Singer† Kuldip K. Paliwal†‡

†ATR Interpreting Telecommunications Research Laboratories  
‡Tokushima University  
‡Griffith University

†2-2 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-02

Tel.: 0774-95-1397 e-mail: {stsuge, fukada, singer}@itl.atr.co.jp, K.Paliwal@me.gu.edu.au

**Abstract** This paper investigates the effectiveness of a novel feature for speech recognition called spectral subband centroids (SSC). SSC are computed as frequency centroids for each subband using the power spectrum of the speech signal. This feature can be obtained reliably even under noisy conditions because SSC are mainly computed from spectral peaks such as formants whose positions are almost unchanged in a noisy environment. Therefore, we can expect SSC to provide here useful information. Experimental results on Japanese spontaneous speech recognition showed that SSC produced significant improvements at SNR=10dB and 20dB when used as a supplemental feature to the conventional Mel-Frequency Cepstral Coefficients (MFCC).

**key words** Spectral subband centroids, Supplement parameter, Noise environment, Speech recognition

## 1 はじめに

音声認識システムは、音声波形から後の認識に有益な特徴パラメータを抽出する信号処理部(フロントエンド)を含んでいる。特徴パラメータは、直接音声認識システムの認識性能に影響を与えるため、特徴パラメータを適切に選択することは認識システム構成において重要な問題である [1][2]。

現在、音声認識システムに有益な特徴パラメータとして、線形予測分析、フィルタバンクから計算されるケプストラム係数が広く用いられている [3]。これらの特徴パラメータは、音響モデル作成のための学習データとそのモデルを用い認識を行う評価データの発声環境が同一の条件の場合、認識に非常に有効な情報を与えることがわかっている。しかし、実環境において、音声認識システムを使用する場合、背景雑音、マイクからの位置のずれ等の影響により、学習データの発声環境と評価データの発声環境の間にミスマッチが生じてしまう。このような環境のミスマッチにより、入力音声のケプストラムと学習データから構築された音響モデルとの間にはずれが生じ、認識性能の劣化を引き起す。このため、これらの環境のミスマッチに対して、頑健な音声認識手法の研究が多数行われている。例えば、加算性雑音に対する雑音対策手法として、スペクトルサブトラクション法(SS)[4]やParallel Model Combination (PMC)法 [5]などが提案されている。また乗算性雑音に対する手法として、ケプストラム平均を減算するケプストラム平均正規化手法等 [6]が挙げられる。SSは入力音声のスペクトルから推定雑音のスペクトルを除去する方法であり、PMC法はモデル側において、雑音モデルを重畳することにより雑音処理を行っている。また、これらの手法以外に環境のミスマッチに対応する手法として、環境のミスマッチに対しても頑健な特徴量を音声波形から抽出し、それを特徴パラメータとして用いる手法がある [7]。また、このような環境のミスマッチに頑健な特徴量を補助的に併用し、環境のミスマッチに対応することにより、認識性能の劣化を抑えることができると考えられる。

近年、音声認識に有効な特徴パラメータとしてスペクトル・サブバンド・セントロイド(Spectral Subband Centroid: SSC)が提案されている [8][9]。SSCは、各サブバンド内に含まれる音声波形のパワースペクトルを用い、サブバンド内のセントロイド周波数として計算される。このセントロイド周波数は、各サブバンド内のパワースペクトルのピークを与える周波数をおおまかにとらえるため、雑音下においても変動が少ないと考えられる。つまり、環境のミスマッチが生じて、学習時の発音環境

下に近い特徴量が抽出できる。このため、SSCを補助的なパラメータとして用いることにより、認識性能の劣化が低減できると考えられる。本稿では、環境のミスマッチに対するSSCの有効性を検証するために、雑音環境下における連続音声認識実験を行っている。

以下、2ではSSC、3では認識実験の結果について、4では実験結果に基づく考察、5では本稿のまとめを述べる。

## 2 スペクトル・サブバンド・セントロイドを用いた雑音下での音声認識

### 2.1 スペクトル・サブバンド・セントロイド

SSC[8][9]は、周波数帯 $[0, F_s/2]$ ( $F_s$ : サンプリング周波数)を $M$ 個のサブバンドに分割を行い、各サブバンドに含まれる音声信号のパワースペクトルを用い、各サブバンドのセントロイド周波数として次式により定義される。

$$C_m = \frac{\int_{l_m}^{h_m} f P^\gamma(f) df}{\int_{l_m}^{h_m} P^\gamma(f) df} \quad (1)$$

ここで、 $f$ は周波数、 $P(f)$ は周波数 $f$ におけるFFTパワースペクトルを示す。 $\gamma$ はパワースペクトル・ダイナミックレンジ・コントロール変数であり、 $0 < \gamma$ の定数である。例えば、 $\gamma = 0.5$ とすると、 $C_m$ は振幅スペクトルから計算される各サブバンドのセントロイド周波数となる。また、 $l_m$ 、 $h_m$ は $m$ 番目のサブバンドの開始および、終了周波数を表す。各サブバンドの開始、終了周波数は、重なりなく周波数帯を等分割した場合、 $l_1 = 0$ 、 $h_M = F_s/2$ 、 $l_{m+1} = h_m = m * F_s / (2 * M)$  ( $m = 1, 2, \dots, M - 1$ )となる。本稿では、周波数帯を等分割したサブバンドからSSCを計算したが、メルスケール、バークスケールを用い分割したサブバンドからの計算も可能である。また、FFTパワースペクトル以外にスペクトル包絡からSSCを計算することも可能である。

### 2.2 分析例

図1に、成人男性が発声した日本語母音/u/をサンプリング周波数16kHzで標準化し、その標準化された音声に対するパワースペクトルを示す。図中の(a)は、クリーンな発声のパワースペクトル、(b)は、その発声にSNR = 10dBの計算機雑音を付加したパワースペクトルである。ここで、SSCは、周波数帯 $[0, 8\text{kHz}]$ を $M = 8$ に重なりなく等分割を行ったサブバンド( $l_1 = 0\text{Hz}$ 、 $h_1 = l_2 = 1000\text{Hz}$ 、 $h_2 = l_3 = 2000\text{kHz}$ 、...)から、 $\gamma =$

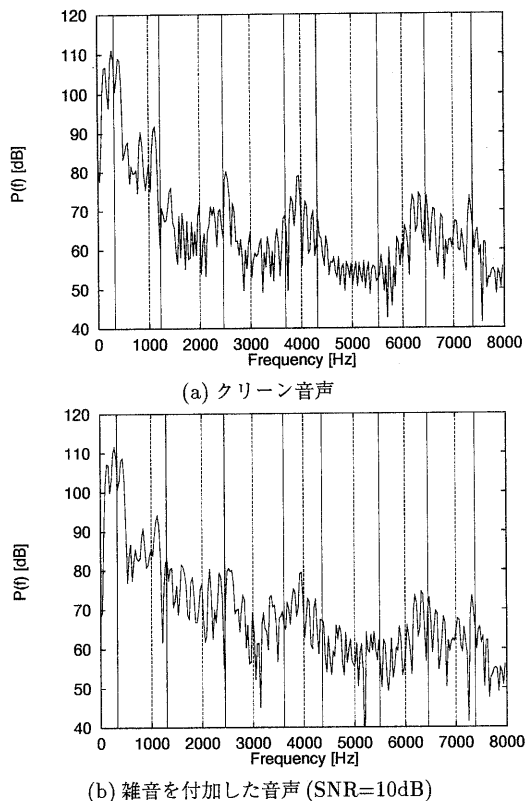


図 1: 母音 /u/ に対する SSC( $C_m$ )、( $M = 8, \gamma = 0.5$ )

0.5 とし式 (1) を用い計算を行った。FFT パワースペクトルは、図 1 に示されるように、1000 ~ 4000Hz あたりにおけるスペクトルが大きく異なることがわかる。その結果、従来、音声認識システムにおいて広く用いられるケプストラムなどの特徴パラメータの変動は大きくなると考えられる。このように環境の変動に対して特徴パラメータが大きく変わると一般に認識性能が大きく劣化する。この図に示した FFT スペクトルから計算した、クリーンな音声、雑音を付加した音声 (SNR = 10dB) の低次の SSC ( $C_1, C_2, C_3$ ) を表 1 に示す。この表から、スペクトルは大きく変化しているにもかかわらず、SSC は変動が非常に少ないことがわかる。これは、SSC はパワースペクトルのレベルが大きい周波数に対して大きな重みをかけ計算され、これらのスペクトルのピークは、雑音下においても変動が少ないためと考えられる。よって、SSC は雑音付加のような環境のミスマッチが生じて、クリーンな環境下と同様な特徴量が抽出できると考えられ、その結果、認識性能の劣化を低減することができると思わ

表 1: 図 1 における SSC (Hz)

SNR (dB)	$C_1$	$C_2$	$C_3$
clean( $\infty$ )	324.5	1223.0	2461.8
10	326.3	1299.4	2445.3

表 2: 音響分析条件

サンプリング周波数	16kHz
プリエンファシス	0.98
フレーム周期	10 msec
フレーム長	20 msec (ハミング窓)
ケプストラム次数	12
フィルタバンク次数	16
SSC 次数	6

れる。すなわち、SSC は雑音下におけるロバストな特徴パラメータであることが期待できる。SSC は、ケプストラムと同様に単独で特徴パラメータとして用いても効果があることが確認されているが [9]、本稿では、SSC を従来の特徴パラメータに併用した場合の効果について調べることにする。

### 3 音声認識実験

SSC の雑音下での効果を調べるため、Travel Arrangement をタスクとする自由発話音声データベース [10] を用いた連続音声認識実験を行った。

#### 3.1 実験条件

学習データとして、230 名 (男性 100 名、女性 130 名)、総音素数約 12 万 (約 200 分) を用い、評価データには、学習に用いられていない 42 名 (男性 17 名、女性 25 名)、総音素数約 2 万 (約 40 分) を用いた。音響分析条件を表 2 に示す。特徴ベクトルは、この音響分析条件で分析した表 3 に示す 4 種類を用いた。この 4 種類の特徴ベクトルは以下である。表 3 の 1. の MFCC(+ $\Delta$ ) は、MFCC と対数パワーに、各々の一次回帰係数を加えた合計 26 次元の特徴ベクトルであり、2. の MFCC(+ $\Delta$ + $\Delta\Delta$ ) は 1. に加え、

表 3: 特徴ベクトル

特徴ベクトル	合計次元数
1. MFCC(+ $\Delta$ )	26
2. MFCC(+ $\Delta$ + $\Delta\Delta$ )	39
3. 1.+ SSC(+ $\Delta$ )	38
4. 2.+ SSC(+ $\Delta$ )	51

それぞれの二次回帰係数も使用した合計 39 次元の特徴ベクトルである。3. は 1. に、SSC 6 次元とその一次回帰係数を加えた合計 38 次元の特徴ベクトルであり、4. は 2. に、SSC 6 次元とその一次回帰係数を加えた合計 51 次元の特徴ベクトルとする。SSC は、ナイキスト周波数 8kHz を 6 個のサブバンドに重なりなく等分割 ( $M = 6$ ) を行い、 $\gamma = 0.5$  として式 (1) より計算を行った。音響モデルのトポロジーは、1. ~ 4. においてすべて共通であり、有音モデルは、MFCC(+ $\Delta$ ) を特徴ベクトルとして ML-SSS アルゴリズム [11] により分割した、総状態数 800、各 5 混合の HMnet (音素環境依存 HMM) を用い、無音モデルは、3 状態 10 混合の HMM を用いた。

音声認識システムを実環境下で使用する場合、学習に用いた音声の収録環境と認識を行う環境とのミスマッチが生じることが多い。そのため、本稿では、音響モデル学習時には雑音を付加しないクリーンな音声を用い、評価データには様々な雑音比となるように加算性雑音を付加した音声を用いた。この評価データに対し、日本語の音素接続規則の制限下において、日本語の 26 音素 (無音を含む) を使用した連続音素認識によって評価を行った [12]。雑音は、電子協雑音データベース [13] から計算機室雑音 (ワークステーション) を用い、重量雑音比は発声ごとに決定をした。

## 3.2 認識結果

### 3.2.1 SSC の効果

音素認識率を表 4 に示す。まず、ミスマッチがない場合 ( $\text{SNR} = \infty$ ) では、SSC を用いたことによる認識性能の改善はみられなかった。しかしながら、低 S/N 比になるに従い、認識性能の劣化がより低減されていくことがわかる。特に、 $\text{SNR} = 10\text{dB}$  においては、音素認識率が 4.4% 向上する (表 4 の 1., 3. 参照)。さらに、従来よく使用されている MFCC の二次回帰係数と比較しても、低 S/N 比になるに従い、SSC を用いることによる認識性能の改善が二次回帰係数の改善を上回ることがわかる。例えば、 $\text{SNR} = 10\text{dB}$  において、音素認識率が 2.3% 向上する (表 4 の 2., 3. 参照)。よって、SSC は低 S/N 比の場合において認識性能の劣化の低減に有効なパラメータであることがわかる。

### 3.2.2 SSC と二次回帰係数の併用の効果

SSC と MFCC の二次回帰係数は、各々 MFCC(+ $\Delta$ ) の補助的パラメータとして用いた場合、認識性能を改善することがわかった。各々単独で使用したこの 2 つの補助的パラメータを併用した場合の認識結果を表 5 に示す。

表 4: 各特徴ベクトルに対する音素認識率 (%)

特徴ベクトル	SNR			
	10	20	30	$\infty$
1. MFCC(+ $\Delta$ )	33.8	61.0	72.8	74.0
2. MFCC(+ $\Delta + \Delta\Delta$ )	35.9	66.1	74.5	75.7
3. 1.+SSC(+ $\Delta$ )	38.2	66.7	72.8	73.3

表 5: SSC と MFCC の二次回帰係数の併用の効果 (音素認識率 (%))

特徴ベクトル	SNR			
	10	20	30	$\infty$
4. 2.+SSC(+ $\Delta$ )	43.3	68.8	74.9	75.1

この結果より、両補助的パラメータを併用した場合は、MFCC(+ $\Delta$ ) と比較して、 $\text{SNR} = 10\text{dB}$  において音素認識率 9.5% (誤り改善率 14.3%) 向上、 $\text{SNR} = 20\text{dB}$  において音素認識率 7.8% (誤り改善率 19.9%) 向上の効果を示した (表 4 の 1., 表 5 の 4. 参照)。また、MFCC の二次回帰係数のみを補助的パラメータとして使用した場合と比較しても、音素認識率が  $\text{SNR} = 10\text{dB}$  において 7.3% (誤り改善率 11.4%) 向上、 $\text{SNR} = 20\text{dB}$  において 2.7% (誤り改善率 7.9%) 向上を示した (表 4 の 2., 表 5 の 4. 参照)。この結果より、MFCC(+ $\Delta$ ) の補助的パラメータとして SSC と MFCC の二次回帰係数を併用すると、各々単独で補助パラメータとして使用するよりも、環境のミスマッチに対して認識性能の劣化が低減できることがわかった。

## 4 考察

ここでは、認識時間の分析、誤り傾向の分析、サブバンドオーバーラップ分割の観点から考察を行う。

### 4.1 次元数増加による認識時間の比較

補助的パラメータの使用は、特徴ベクトルの次元数の増加になるために認識時間の増加を引き起こすと考えら

表 6: 発声時間に対する認識時間の比率

特徴ベクトル	SNR			
	10	20	30	$\infty$
1. MFCC(+ $\Delta$ )	1.07	0.70	0.48	0.37
2. MFCC(+ $\Delta + \Delta\Delta$ )	1.00	0.70	0.51	0.40
3. 1.+SSC(+ $\Delta$ )	1.09	0.68	0.52	0.41
4. 2.+SSC(+ $\Delta$ )	1.19	0.75	0.59	0.45

れる。各々の特徴ベクトルを用いた場合の発声時間に対する認識時間の比率を表6に示す。この表より、次元増加に対して、認識時間はさほど増加していないことがわかる。これは、本稿では、ゆう度幅一定による仮説数の枝刈りを行っているため、適切な音響モデルを用いることにより、枝刈りされる仮説の数が増加し認識時間の増加が抑えられたと考えられる。しかし、認識時間以外に記憶装置などで特徴ベクトルの次元数増加が音声認識システムにおいて問題となるならば、Linear Discriminant Analysis (LDA)[14]などの次元圧縮技法を用いることにより対処できると考えられる。

## 4.2 誤り傾向の分析

ここでは、SSCを補助的パラメータとして用い、特に認識性能の改善が大きかったSNR = 10dBについて、認識誤りの分析を行う。表7に、挿入、脱落、置換の誤り数を示す。この表より、SSCは挿入誤りを著しく減少させる効果があることがわかる。MFCC(+ $\Delta$ )、MFCC(+ $\Delta$ ) + SSC(+ $\Delta$ )に対する挿入誤りの過半数は、母音による挿入誤りであることがわかった。よって、この両特徴ベクトルに対する日本語5母音の挿入誤り数を表8に示す。表7、8より、SSCをMFCC(+ $\Delta$ )と併用することによる挿入誤りの減少は971個であり、そのうちの約45%(433個)は母音の挿入誤りの減少であることがわかる。また、SSCは全母音の挿入誤り数を減少させていることがわかる。さらに、母音別にみると、母音/i/、/e/は誤りが50%以上減少している。しかし、母音/a/、/o/に関してはさほど誤りが減少していないことがわかる。これは、母音の第一ホルマント(F1)、第二ホルマント(F2)周波数の位置に関わりがあると考えられる。一般に、母音/i/、/e/のF1は200~700Hz、F2は1500~3000Hzであり、F1とF2が比較的離れた位置にあるのに対し、母音/a/、/o/はF1は300~1200Hz、F2は800~2000Hzと隣接している。3の実験で用いたSSCは、周波数帯8kHzを6個のサブバンドに分割を行った。このとき、F1とF2が離れている母音/i/、/e/のF1は第一サブバンド(0~1333Hz)、F2は第二サブバンド(1334~2667Hz)内に含まれる。このため、SSCが各サブバンドのパワースペクトルのピークを与える周波数として、これらのホルマント周波数を近似することができたため、著しく挿入誤りが減少したと考えられる。しかし、母音/a/、/o/のF1とF2は隣接しているため、共に第一サブバンド内に含まれてしまい、1つのサブバンド内にパワースペクトルのピークが2つ存在することになる。このため、SSCは、それらのピークを一つの周波数で表現

表 7: 認識誤りについての詳細 (全認識音素数 21014 音素)

特徴ベクトル	挿入	削除	置換
1. MFCC(+ $\Delta$ )	3027	2293	8597
2. MFCC(+ $\Delta$ + $\Delta\Delta$ )	2821	2395	8249
3. 1.+SSC(+ $\Delta$ )	2056	2620	8322
4. 2.+SSC(+ $\Delta$ )	1663	2561	7702

表 8: SNR=10dB における日本語母音の挿入誤り数

特徴ベクトル	/a/	/i/	/u/	/e/	/o/
1. MFCC(+ $\Delta$ )	1011	73	145	382	87
3. 1.+SSC(+ $\Delta$ )	918	33	91	152	71

しようとし、その両ピークを与える周波数の間をセントロイド周波数としてしまい、サブバンド内のパワースペクトルのピークを与える周波数を適切に表現することができなかったと考えられる。その結果、母音/a/、/o/に対しては、挿入誤りがさほど減少しなかったと考えられる。よって、さらなる認識性能改善のためには、各音素ごとにサブバンド幅を決定するなどの適切なサブバンド分割手法の決定が重要であると考えられる。

## 4.3 サブバンドオーバーラップ分割

3の認識実験に用いたSSCは、重なりのない等分割を行ったサブバンドを用い計算を行っている。この場合、サブバンド境界上をホルマントが移動した時には、SSCの値は大きく変動すると考えられる。これは、サブバンド境界上をパワースペクトルのピークが移動した場合、このピークを与える周波数は異なるサブバンドに含まれ、各サブバンドのセントロイド周波数が大きく変動するためである。このとき、類似した特徴を示すパワースペクトルであっても、大きくSSCの値が異なり、安定に特徴量を抽出できなくなると考えられる。この問題は、サブバンド境界上に別のサブバンドをオーバーラップさせるサブバンドのオーバーラップ分割を用いることにより回避できると考えられる。

そこで、3で用いた周波数帯を6つに重なりなく分割を行ったサブバンドに、新たに各サブバンドに半分ずつオーバーラップするサブバンドを加えサブバンド数11 ( $M = 11$ )としてSSCを計算をした。このオーバーラップしたサブバンドから計算をしたSSCをSSC<sub>ov</sub>とする。オーバーラップ分割の効果を調べるため、3と同様の音素認識実験を行った。結果を表9に示す。過去の報告では、サブバンド分割の効果は見られたが[8]、ここでは、オーバーラップ

表 9: オーバーラップの効果 (音素認識率 %)

特徴ベクトル	SNR			
	10	20	30	$\infty$
MFCC(+ $\Delta$ )+SSC(+ $\Delta$ )	38.2	66.7	72.8	73.3
MFCC(+ $\Delta$ )+SSC <sub>ov</sub> (+ $\Delta$ )	38.4	66.7	72.7	73.4

ブ分割による認識性能の改善は見られなかった。

## 5 むすび

音声認識のための新しい補助的パラメータとしてスペクトル・サブバンド・セントロイド(SSC)を用いた雑音環境下における連続音声認識実験を行った。SSCは、サブバンド内に存在するパワースペクトルのセントロイド周波数として定義され、雑音下においても変化が少ない。このため、認識時に環境のミスマッチが生じて、学習時の発音に近いロバストな特徴量が得られる。自然発話を用いた音声認識実験を行った結果、MFCCの二次回帰係数とSSCを併用し、MFCCとその一次回帰係数の補助的パラメータとして用いた場合、SNR = 10dBにおいて14.3%、SNR = 20dBにおいて19.9%の誤り改善率を得ることができた。本稿で用いたSSCは、雑音除去手法(例えば、[4])を使用することなく、認識性能を改善することができたが、雑音除去手法と組み合わせることにより、さらに認識性能の改善が期待できる。また、今後は(1)計算機雑音以外の種々の雑音下における有効性、(2)単語 *bigram* や可変長 *N-gram*[15]を用いた単語認識率による評価を行う予定である。

## 謝辞

日頃熱心にご討論頂くATR音声翻訳通信研究所の皆様へ感謝をします。また、ATR音声翻訳通信研究所にて、研究を行う機会を与えて頂いた徳島大学工学部北研二助教授に深く感謝をします。

## 参考文献

- [1] J. Picone. Signal modeling techniques in speech recognition. In *Proceedings of the IEEE*, Vol. 81, pp. 1215-1247, 1993.
- [2] Ch. Jankowski Jr, H. Vo, and R. Lippmann. A comparison of signal processing front ends for automatic word recognition. *IEEE Trans. Speech and Audio Processing*, Vol. 3, No. 4, pp. 286-293, July 1995.
- [3] L. Rabiner and B-H. Juang. *Fundamentals of Speech Recognition*. Prentice Hall, Englewood Cliffs, NJ, 1993.
- [4] S. Boll. Suppression of acoustic noise in speech using spectral subtraction. In *IEEE Trans. Acoustics, Speech and Signal Processing*, pp. 113-120, April 1979.
- [5] M. J. F. Gales and S. Young. An improved approach to the hidden markov model decomposition of speech and noise. In *Proc. ICASSP*, pp. 233-236, 1992.
- [6] F. H. Liu, R. M. Stern, X. Huang, and A. Acero. Efficient cepstral normalization for robust speech recognition. In *Proc. DARPA Workshop*, pp. 69-74, March 1993.
- [7] H. Hermansky and N. Morgan. Rasta processing of speech. In *IEEE Trans. Speech and Audio Processing*, Vol. 2, pp. 578-589, October 1994.
- [8] K. Paliwal, 柘植寛, Harald Singer, 深田 俊明, 北研二. 自然発話音声認識における音響分析の比較. 音講論, pp. 5-6, September 1997.
- [9] K. Paliwal. Spectral subband centroids as features for speech recognition. In *Proc. IEEE ASRU Workshop97*, 1997.
- [10] A. Nakamura, S. Matsunaga, T. Shimizu, M. Tonomura, and Y. Sagisaka. Japanese speech databases for robust speech recognition. In *Proc. ICSLP*, pp. 2199-2202, Philadelphia, 1996.
- [11] M. Ostendorf and H. Singer. HMM topology design using maximum likelihood successive state splitting. *Computer Speech and Language*, Vol. 11, No. 1, pp. 17-41, 1997.
- [12] 大脇浩, Harald Singer, 鷹見淳一, 樽松明. 音素配列構造の制約を用いた音素タイプライタ. 信学技報, Vol. SP93-113, pp. 71-78, December 1993.
- [13] 板橋秀一. 騒音データベースと日本語共通音声データ DAT版. 日本音響学会誌, Vol. 47, No. 2, pp. 951-953, 1991.
- [14] R. Duda and P. Hart. *Pattern Classification and Scene Analysis*. John Wiley., New York, 1973.
- [15] 政瀧浩和, 松永昭一, 匂坂芳典. 連続音声認識のための可変長連鎖統計言語モデル. 信学技報, Vol. SP95-73, pp. 1-6, November 1995.