音声言語情報処理 19－3
(1997. 12. 11)

ロバストな音声認識実現を目的とした変調スペクトル特性の検討

金寺 登[1] Hynek Hermansky[2] 荒井 隆行[3] 船田 哲男[4]

[1]石川高専 〒929-03 石川県河北郡津幡町北中条 kane@i.ishikawa-nct.ac.jp
[2]Oregon Graduate Institute of Science & Technology, Portland, Oregon, U.S.A.
[3]International Computer Science Institute, Berkeley, California, U.S.A.
[4]金沢大学 〒920 石川県金沢市小立野 2-40-20

あらまし CMS法や動的特徴は変調周波数特性を操作することにより音声認識性能が向上することが知られているが、どの変調周波数がどの程度重要であるのかという定量的な検討は行われていない。そこで本研究では、様々な変調周波数特性を持った入力に対し、音声認識性能の違いを種々の雑音環境、認識方式、特徴量のもとで調べた。その結果、以下のことが分かった: 1) 言語情報のほとんどが1～16 Hzの変調周波数帯域に存在し、その中でも4 Hz付近が最も重要である。2) 変調スペクトルにおいては位相情報も重要である。3) 4 Hz付近の変調周波数を含む特徴量を用いることで動的特徴量と同等以上の結果が得られる。4) 適切な中心周波数と帯域幅をもつ複数のサブバンドを変調周波数上で用いることにより、認識性能がさらに向上する。

キーワード: 音声認識・変調周波数

On Properties of the Modulation Spectrum for Robust Automatic Speech Recognition

Noboru Kanedera[1] Hynek Hermansky[2] Takayuki Arai[3] Tetsuo Funada[4]

[1]Ishikawa National College of Technology, Tsubata, Ishikawa
[2]Oregon Graduate Institute of Science & Technology, Portland, Oregon, U.S.A.
[3]International Computer Science Institute, Berkeley, California, U.S.A.
[4]Faculty of Engineering, Kanazawa University, Kanazawa, Ishikawa

Abstract We report on the effect of band-pass filtering of the time trajectories of spectral envelopes on speech recognition. Several types of recognizers, several types of features, and several types of filters are studied. Results indicate the relative importance of different components of the modulation spectrum of speech for ASR. General conclusions are: (1) most of the useful linguistic information is in modulation frequency components from the range between 1 and 16 Hz, with the dominant component at around 4 Hz, (2) it is important to preserve the phase information in modulation frequency domain, (3) The features which include components at around 4 Hz in modulation spectrum outperform the conventional delta features, (4) The features which represent the several modulation frequency bands with appropriate center frequency and band width increase recognition performance.

Key words: Automatic Speech Recognition, Modulation Frequency

1 はじめに

現在広く用いられているCMS (cepstral mean subtraction) 法や動的特徴は、いずれも特徴パラメータの時間変化に注目している。この時間変化を周波数次元で表したものが変調スペクトルであり、その周波数次元は変調周波数と呼ばれる。CMS [1] はケプストラムの時間軌跡の直流成分を取り除く。これにより マイクの周波数特性や通信伝送路におけるチャンネル特性などによる乗法性ノイズの影響を軽減することができる。動的特徴量 [2] の計算においては、ケプストラム係数の時間軌跡をフィルタリングし、10 Hz 付近の変調周波数成分が強調されるのに対し、その他の成分は軽減される。RASTA (RelAtive SpecTrAl processing) [3] においては、約 1～12 Hz の変調周波数成分が使用されている。

このように現在広く用いられているこれらの処理は、音声の変調スペクトルを効果的に修正している [3]。しかし ASR にとって どの変調周波数がどの程度重要であるのかという定量的な検討は行われていない。もし変調スペクトル間の相対的な重要性を知ることができれば、「音声の特徴の中で どこの部分が言語情報を担うのか？」「どうすればもっとロバストな ASR システムを実現できるのか？」といった疑問への手がかりを得ることができるかもしれない。従って人間同士、または人間と機械との音声コミュニケーションに対する変調スペクトル間の相対的な重要性を知ることは重要である。

知覚実験 [5, 6] により、一部の変調スペクトル成分が他に比べて重要であることが知られている。この事実は日本語 [7] や英語 [9] においても確認されている。Drullman ら [5, 6] は、16 Hz 以下の低域通過フィルタリングや 4 Hz 以上の高域通過フィルタリングによって、音声の明瞭度が低下しないことを示している。荒井ら [7] は、Drullman らの研究を対数領域に拡張し、低域／高域通過フィルタばかりでなくバンドパスフィルタを適用した。この結果、明瞭度を保持するために必要なほとんどの情報が 1～16 Hz の変調周波数帯域に存在することが明らかとなった。

ASR に対して、我々は変調スペクトル間の相対的な重要性を調査した [14]。この結果、ASR にとって重要な言語情報のほとんどが 1～16 Hz の変調周波数帯域に存在し、その中でも音声の音節速度 (syllabic rate) に対応する 4 Hz 付近が最も重要であることがわかった。また雑音環境においては、2 Hz 以下や 16 Hz 以上の変調スペクトル成分が認識性能を劣化させることがあることがわかった。特に 1 Hz 以下の変調スペクトル成分は認識性能を著しく低下させる。

このような変調スペクトル間の相対的な重要性の傾向は、認識方式等の各種要因によって変化する可能性がある。そこで本研究では各種要因として (1) 認識方式 (DTW, HMM), (2) 特徴量 (フィルタバンク出力, MFCC, PLP), (3) 変調スペクトルを抽出するフィルタ特性 (線形位相 FIR フィルタ, DFT) を取り上げ、これらの要因により 変調スペクトル間の相対的な重要性を定量的に調査した結果を報告する。

2 変調周波数間の相対的な重要性

本節では、変調周波数間の相対的な重要性を定量的に調査するため、認識性能への貢献度を定義する。

図1に示すように、時間軌跡処理を伴う ASR システムは、対数スペクトル等の抽出部、対数スペクトルの時間軌跡をフィルタリングするフィルタ部、及び認識器から成る。フィルタ部をバンドパスフィルタとしたとき、低域遮断周波数 f_L と高域遮断周波数 f_U を与えることによってシステムの認識率 $p(f_L, f_U)$ が求められ、図2のように図示できる。(実験条件等は3節で示す。)

各変調周波数の相対的な重要性を知るために、まず隣接する認識率の差を取った例を図3に示す。図中の矢印の長さは隣接する認識率の差の大きさを示しており、与えられた変調周波数帯 (図中では 2～4 Hz) の重要性を表すと考えられる。与えられた変調周波数帯に対応する認識率の差 (矢印の大きさ) を平均することにより、その周波数帯の平均的な重要性を知ることができる。ここでは、これを認識性能への貢献度と呼ぶことにする。

一般に $f_L \sim f_U$ を含むことによる認識性能への貢献度 $I(f_L, f_U)$ は、

$$I(f_L, f_U) = \frac{1}{N-1}\Big[\sum_{l<f_L}\{p(l, f_U) - p(l, f_L)\} + \sum_{u>f_U}\{p(f_L, u) - p(f_U, u)\}\Big] \quad (1)$$

と定義される。ここで N は、変調周波数の範囲の数である。

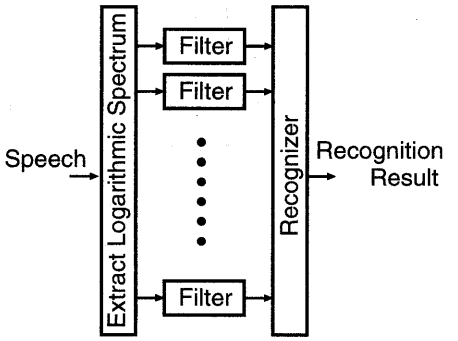


図 1: Block diagram of the ASR system with temporal processing.

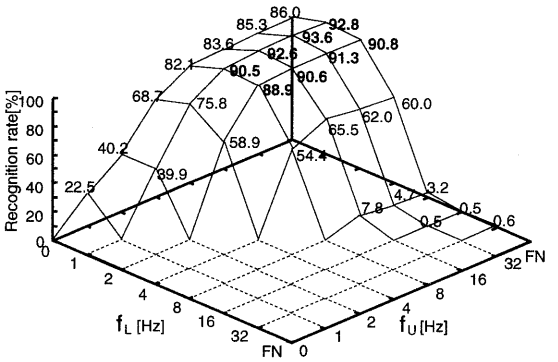


図 2: Recognition results for the band-passed time trajectories.

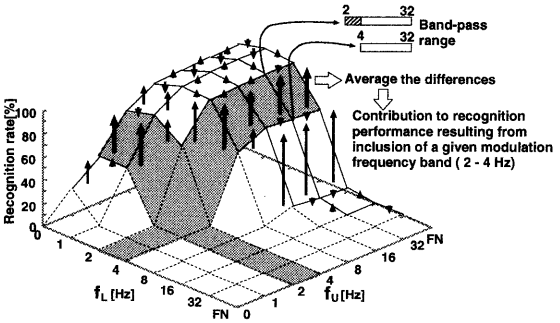


図 3: Deriving the relative importance of a given modulation frequency band.

表 1: 実験条件 1

Task		216 words ATR Japanese database set C
Feature		FFT-based filter bank output
Filter		511-tap linear phase FIR filter
Recognizer		DTW
Training	Speaker Sampling Recording	10 male speakers 10 kHz, 16 bit an anechoic room
Test	Speaker Sampling Recording	5 male speakers 11.025 kHz, 16 bit a computer room
Window length		25 ms
Frame period		12.5 ms

3 認識実験

3.1 コンピュータ室内での音声認識における変調周波数間の相対的重要性

表 1に示す条件で 216 単語音声認識実験を行った。学習にはクリーンな音声 (ATR 日本語音声データベース) を使用したのに対して評価データにはコンピュータ室 (約 55 dB の背景雑音環境) においてハンドヘルドマイクロホンを使用して直接パソコンに収録された音声を使用した。

分析には FFT によるフィルタバンクを使用し、16 チャネルの等価 Q バンド出力を抽出した。各スペクトル係数は lin-log 関数 [8]

$$y = \log(1 + Jx), \tag{2}$$

によって、対数的な値に変換された。ここで J は正定数である。振幅変換関数 (2) 式は $J \ll 1$ のとき線形的であり、$J \gg 1$ のとき対数的である。

次に各チャネル毎に 511 点の線形位相 FIR フィルタを通過させることにより、対象とする変調周波数のみを抽出した。

図 2は 変調周波数領域において、いろいろなバンドパス特性を持つフィルタを適用した場合の認識結果を示している。縦軸は単語認識率、その他の軸は低域遮断周波数 f_L 及び 高域遮断周波数 f_U を示

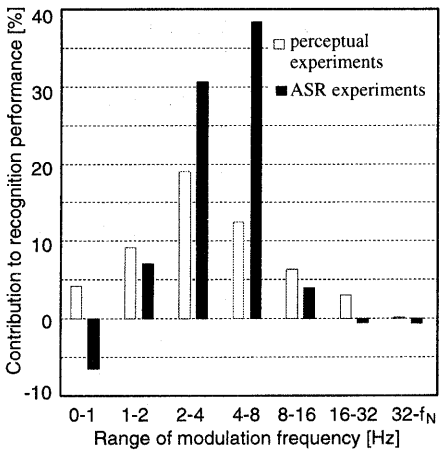


図 4: Comparison with the perceptual experiments.

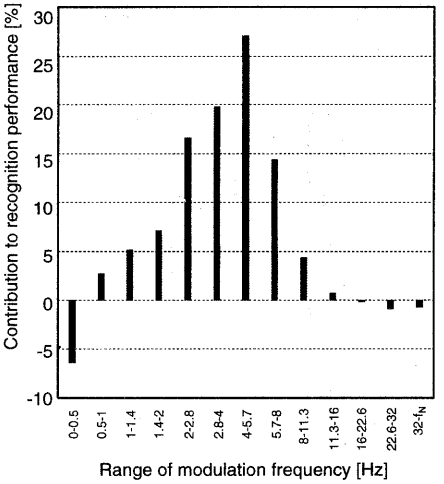


図 5: Improvement of recognition accuracy by including each modulation frequency.

している。太字で示されているいくつかのバンドパス条件による認識率は、フィルタリングなしの認識率 (86%) に比べ明らかに改善されていることがわかる。

図 4は、2節の (1) 式により導出された認識性能への貢献度を示している。すなわち縦軸は与えられた変調周波数帯を含めたことによる認識率の平均的な増加量を示している。比較のため、知覚実験 [7] によるデータを用いて、同様の評価を行った結果もまた 図 4 に示してある。これらの結果より、ASR にとって有用な言語情報のほとんどが 1～16 Hz の変調周波数帯に存在し、その中でも 音声の音節速度 (syllabic rate) に対応する 4 Hz 付近が最も重要であることがわかる。図 5は 変調周波数のレンジをより細かくして評価した場合の結果を示している。明らかに図 4と図 5の傾向は類似している。

3.2 認識方式の違いによる影響

変調周波数間の相対的な重要性は、認識方式によって変化する可能性がある。そこで表 2 及び 表 3の実験条件で、認識方式として DTW または HMM を用いた場合の変調周波数間の相対的な重要性を調査した。図 6は、認識方式として DTW または HMM を用いた場合の認識性能への正規化貢献度を示している。ここで、いろいろなケースを比較するため、貢献度を最大値で正規化した。HMM を用いた実験においては、HMM tool kit (HTK) を使用した。また各単語を 8 状態 (始終端を含む) にモデル化し、各状態毎に 2 混合正規分布を用いた。

表 2: 実験条件 2

Task	13 words Bellcore digit database (0–9, zero, oh, yes, no)
Recognizer	DTW
Training	20 speakers (10 males and 10 females)
Test	50 speakers (25 males and 25 females)
Sampling frequency	8 kHz
Window length	25 ms
Frame period	12.5 ms

DTW と HMM による 貢献度の傾向は類似している。すなわち 2～8 Hz の変調周波数帯は、クリーンな環境 (図 6(a)) 及び雑音環境 (図 6(b)) の両方において有用であるのに対して、その他の区間の雑音中での貢献度はクリーンな場合に比べて低下している。

3.3 特徴量の影響

変調周波数間の相対的な重要性は、使用する特徴量によっても変化する可能性がある。図 6は、3.1節で述べたフィルタバンク出力 (FB) とメルケプスト

表 3: 実験条件 3

Task	13 words Bellcore digit database (0–9, zero, oh, yes, no)
Recognizer	HMM
Training	150 speakers (75 males and 75 females)
Test	50 speakers (25 males and 25 females)
Sampling frequency	8 kHz
Window length	25 ms
Frame period	12.5 ms

ラム (MFCC), PLP [10] を用いた場合の認識性能への正規化貢献度もまた示している。MFCC の次数は 12 次とし、PLP の次数は 8 次とした。また HMM 中の共分散行列の内、MFCC と PLP においては対角要素のみを使用したのに対し、FB では全共分散を使用した。

クリーンな環境に対しては いずれの特徴量についても傾向は類似している。雑音環境においては、FB 特徴量はその他の特徴量と (特に 1〜2 Hz において) 多少傾向が異なっている。いずれの特徴量についても、2〜8 Hz の変調周波数帯は、クリーンな環境及び雑音環境の両方において有用である。

3.4 フィルタ特性の影響

これまでの実験においては、鋭い変調周波数特性を得るため、長いタップの線形位相 FIR フィルタを使用した。しかしながら長いタップの FIR フィルタは長い時間遅れを生ずる。従って実際の ASR 環境においては、短いタップを持つフィルタが望ましい。一般に短いタップのフィルタは鋭い周波数特性を得ることが難しい。よって、緩やかな周波数特性を持つフィルタを使用したとき、変調周波数の相対的な重要性が変化するかどうかを確認するため、短いタップのフィルタの一例として少ない点数の DFT を用いて実験を行った。

具体的には まず 8 次の PLP と対数パワーを求め、これらの各時間軌跡について、64 フレームを切り出し、ハミング窓を適用後、64 点の DFT を計算した。次に 対象とする変調周波数帯に対応する成分のみを抽出し、その時刻における新しい特徴量とした。さらに 切り出し位置を 1 フレームずつシフトすることにより、すべての時刻において対象とする変調周波数帯に対応する新しい特徴量を抽出した。なお、端点での計算では 0 を挿入した。対象とする変調周波数帯を様々に変化させ、対応するシステムの認識率を求めれば、2節の (1) 式により、各変調周波数が認識性能に寄与する貢献度を求めることができる。

図 7は、ハミング窓を伴った 64 点 DFT フィルタリングを用いた場合の正規化貢献度を示している。横軸は各 DFT フィルタの中心変調周波数を表している。これらの実験には、表3の英語データベースを使用した。HMM の学習には HMM tool kit (HTK) を用い、各単語毎に 8 状態 2 混合 HMM を学習した。各混合の共分散は対角とした。図 7(a) は 雑音が少ない環境での結果を示しているのに対し、図 7(b) においては、評価データが 加法性雑音 (SNR 10 dB) と乗法性雑音 (HPF, 6 dB/oct) によって劣化された場合の結果を示している。この加法性雑音は、コンピュータ室 (約 55 dB の背景雑音環境) においてダイナミックマイクロホンを使用し Sound Blaster 互換ボードを介して、直接パソコンに収録された。

フィルタの変調周波数特性が大きく異なるにもかかわらず、図 7の変調周波数間の相対的な重要性の傾向は、以前の実験の傾向と類似している。すなわち、最も重要な変調周波数は 約 4 Hz であり、4 Hz 周辺は クリーンな環境と雑音環境の両方で重要であった。雑音環境では 2 Hz 以下あるいは 10 Hz 以上の変調周波数成分の重要性は低くなった。特に 1 Hz 以下の変調周波数成分は著しく認識率を劣化させることがわかった。

3.5 選択された一部の変調スペクトルを用いた 2-D ケプストラム

ケプストラムの時間軌跡のフーリエ変換を横軸が時間、縦軸がケフレンシになるように 2 次元的に並べたものは 2-D ケプストラム [11, 12] と呼ばれているので、3.4節の実験は また 特徴量として 2-D ケプストラムを使用することに相当する。北村ら [12] は孤立単語音声認識において、各単語毎に 1 つの 2-D ケプストラムを使用し、優れた結果を得た。一方 Ben Milner [13] は 連続音声にも適用できるように

各フレーム毎に1つの 2-D ケプストラムを使用している。図 7の結果は ASR にとって重要な言語情報の大部分が 1～16 Hz (特に 2～10 Hz) の変調周波数帯に存在することを示唆しているので、我々は [13] に比べて より長い時間軌跡窓を使用し、重要な言語情報が存在する変調周波数帯のみを選択して使用した。

表 4は、変調スペクトルの選択された一部分のみを使用した 2-D ケプストラムと 一般に使用されている動的特徴量の認識結果を比較している。表中の「clean」は 評価環境が学習データと同一で雑音が少ない環境での結果を示しており、「noisy」は 評価データが 加法性雑音 (SNR 10 dB) と乗法性雑音 (HPF, 6 dB/oct) によって劣化された場合の結果を示している。表中の DCT を用いた 2-D ケプストラムは、8 次の PLP と対数パワーの各係数について 16 フレームの時間軌跡を対称に配置後、ハミング窓を適用し、32 点 DFT 後の実部に相当する。一方 DFT を用いた 2-D ケプストラムは、32 フレームの時間軌跡にハミング窓を適用し 32 点 DFT したもので、実部と虚部を持つ。この実験においては 2-D ケプストラムの内、中心変調周波数 5, 7.5 Hz に対応する 2 番目と 3 番目の成分を使用した。これらは約 3～9.5 Hz の変調周波数帯域を持つ。DCT を使用した場合の特徴量の数は 18 (2 成分 × 1 (実部のみ) × 9) であり、DFT を使用した場合の特徴量の数は 36 (2 成分 × 2 (実部, 虚部) × 9) であった。フィルタリングなしの静的な特徴量も同時に使用した場合の特徴量の数は 27 (DCT 使用時), 45 (DFT 使用時) であった。

これらの結果は、4 Hz 周辺の変調周波数成分を含む 2-D ケプストラムにより、広く使用されている動的特徴量と同等以上の結果が得られることを示している。特に学習と評価の環境が異なる場合に 2-D ケプストラムが有利である。また 400 ms の時間軌跡情報を持つ DFT が 200 ms の時間軌跡情報を持つ DCT よりも優れている。

3.6 位相の影響

表 5は、変調スペクトルにおける位相の影響を示している。実験には 3.4節と同様の実験条件を用いた。絶対値のみを用いた結果は、絶対値と位相の両方

表 4: Recognition results of conventional delta features and 2-D cepstrum using selected part of the modulation spectrum. (WER: Word Error Rate).

Feature	Static feature	Feature size	WER [%] Clean	WER [%] Noisy
Δ, Δ^2	yes	27	1.7	21.7
2-D cepstrum (DCT)	yes	27	0.9	12.8
2-D cepstrum (DFT)	yes	45	0.9	4.6
Δ, Δ^2	no	18	2.8	28.8
2-D cepstrum (DCT)	no	18	4.2	4.8
2-D cepstrum (DFT)	no	36	3.4	3.5

表 5: Recognition results in various phase conditions. (WER: Word Error Rate).

Feature	Static feature	Feature size	WER [%] Clean	WER [%] Noisy
2-D cepstrum (DFT)	yes	45	0.9	4.6
Real part	yes	27	1.5	14.2
Imaginary part	yes	27	1.1	10.3
Absolute values	yes	27	5.2	24.2
2-D cepstrum (DFT)	no	36	3.4	3.5
Real part	no	18	7.7	10.5
Imaginary part	no	18	8.0	11.8
Absolute values	no	18	22.5	38.8

(つまり実部と虚部の両方) を用いた 2-D cepstrum (DFT) の結果よりも悪くなった。従って変調スペクトルにおいては 位相情報を保持することが重要であると考えられる。実部のみあるいは虚部のみを用いた場合は、絶対値のみを用いた場合と絶対値と位相の両方を用いた場合 (つまり実部と虚部の両方を用いた場合) の中間的な結果を得た。

3.7 Multi-resolution ASR

この節では multi-resolution の効果を検討する。表 6は、16 点, 32 点, 64 点の DFT による特徴量を組み合わせた場合の結果を示している。実験条件は 3.4節と同様である。それぞれの解像度に対して、約 2～10 Hz の変調周波数帯の中から使用する成分を選択した。例えば (g) の場合には、64 点 DFT の第 2 成分、32 点 DFT の第 2,3 成分を使用した。これらの成分の中心変調周波数はそれぞれ 2.5, 5, 7.5 Hz である。これらの変調周波数はそれぞれ 単語速度、音節速度、半音節速度に対応すると考えられる。これらの結果は、multi-resolution の有効性を示している。

表 6: Recognition results using multi-resolution. (WER: Word Error Rate).

DFT size	16	32	64	Feature size	WER [%] Clean	WER [%] Noisy	Case
Order of DFT components	1	—	—	18	2.4	24.9	(a)
	—	2, 3	—	36	3.4	3.5	(b)
	—	—	2–6	90	2.8	2.3	(c)
	1	2, 3	—	54	1.7	6.2	(d)
	—	2, 3	2–6	126	2.0	2.5	(e)
	1	2, 3	2–6	144	1.5	2.2	(f)
	—	2, 3	2	54	1.4	1.9	(g)

4 まとめ

ASR における変調スペクトル間の相対的重要性について調査検討した。その結果、以下のことがわかった。

1) 言語情報のほとんどが 1〜16 Hz の変調周波数帯域に存在し、その中でも音声の音節速度 (syllabic rate) に対応する 4 Hz 付近が最も重要である。調査した認識方式 (DTW, HMM), 特徴量 (FB, MFCC, PLP), フィルタ特性 (線形位相 FIR, DFT) の違いにもかかわらず、同様の傾向が見られた。
2) 変調スペクトルにおいては位相情報も重要である。
3) 4 Hz 付近の変調周波数を含む特徴量を用いることで動的特徴量と同等以上の結果が得られる。
4) 適切な中心周波数と帯域幅をもつ複数のサブバンドを変調周波数上で用いることにより、認識性能がさらに向上する。

以上の結果は、一般的に使用されている時間軌跡の窓長よりも もっと長い窓を使用し、変調スペクトルの一部のみを使用することが重要であることを示している。

謝辞 Oregon Graduate Institute of Science and Technology の Sangita Tibrewala, Narendranath Malayath, Sarel van Vuuren、そして University of California, Davis の Carlos Avendano の協力に深く感謝致します。また有益なご助言を頂きました東京理科大学の藤崎博也教授、Indian Institute of Technology の B. Yegnanarayana 教授に深く感謝致します。

参考文献

[1] B. S. Atal (1974), "Effectiveness of linear prediction characteristics of the speech wave for automatic speaker identification and verification," J. Acoust. Soc. Amer., Vol. 55, No. 6, pp. 1304 – 1312.

[2] S. Furui (1986), "Speaker-independent isolated word recognition using dynamic features of speech spectrum," IEEE Trans. Acoust. Speech Signal Process., Vol. ASSP-34, No. 1, pp. 52 – 59.

[3] H. Hermansky and N. Morgan (1994), "RASTA processing of speech," IEEE Trans. Speech and Audio Process., Vol. 2, No. 4, pp. 578 – 589.

[4] T. Houtgast and H. J. M. Steeneken (1985), "A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria," J. Acoust. Soc. Amer., Vol. 77, pp. 1069 – 1077.

[5] R. Drullman, J. M. Festen, and R. Plomp (1994), "Effect of temporal envelope smearing on speech perception," J. Acoust. Soc. Amer., Vol. 95, pp. 1053 – 1064.

[6] R. Drullman, J. M. Festen, and R. Plomp (1994), "Effect of reducing slow temporal modulations on speech perception," J. Acoust. Soc. Amer., Vol. 95, pp. 2670 – 2680.

[7] T. Arai, M. Pavel, H. Hermansky and C. Avendano (1996), "Intelligibility of speech with filtered time trajectories of spectral envelopes," In Proc. of the ICSLP, Philadelphia, pp. 2490 – 2493.

[8] H. Hermansky, N. Morgan and H. Hirsch (1993), "Recognition of speech in additive and convolutional noise based on RASTA spectral processing," Proc. IEEE ICASSP, Minneapolis, MN, pp. II-83 – II-86.

[9] S. Greenberg (1996), "Understanding speech understanding — Towards a unified theory of speech perception," In Proc. of the ESCA Tutorial and Advanced Research Workshop on the Auditory Basis of Speech Perception, Keele, England, pp. 1 – 8.

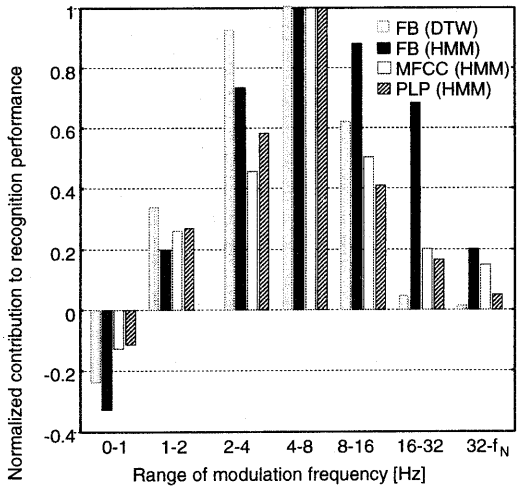
[10] H. Hermansky (1990), "Perceptual linear predictive (PLP) analysis for speech," J. Acoust. Soc. Amer., Vol. 87, No. 4, pp. 1738 – 1752.

[11] 今井 聖, 北村 正 (1976), "2 次元ケプストラムを利用する音声分析," 電子通信学会論文誌, Vol. J59-A, No. 12, pp. 1096 – 1103.

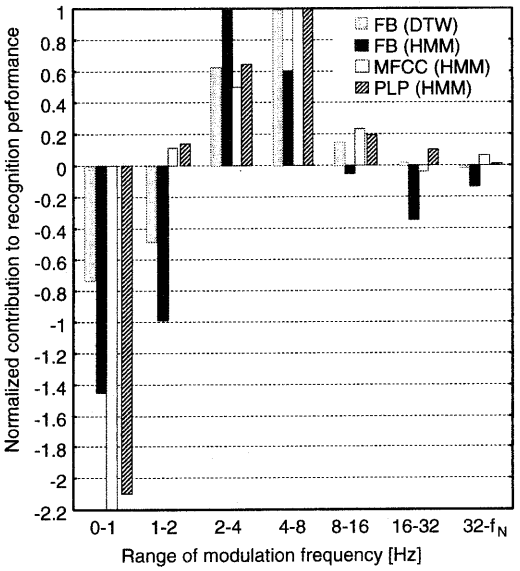
[12] 北村 正, 片柳 恵一 (1989), "2 次元メルケプストラムの静的・動的特徴を用いる数字音声認識," 電子通信学会論文誌, Vol. J72-A, No. 4, pp. 640 – 647.

[13] B. Milner (1996), "Inclusion of temporal information into features for speech recognition," In Proc. of the ICSLP, Philadelphia, pp. 256 – 259.

[14] N. Kanedera, T. Arai, H. Hermansky, and M. Pavel (1997), "On the importance of various modulation frequencies for speech recognition," Proc. Eurospeech '97, Rhodes, Greece, pp. 1079 – 1082.

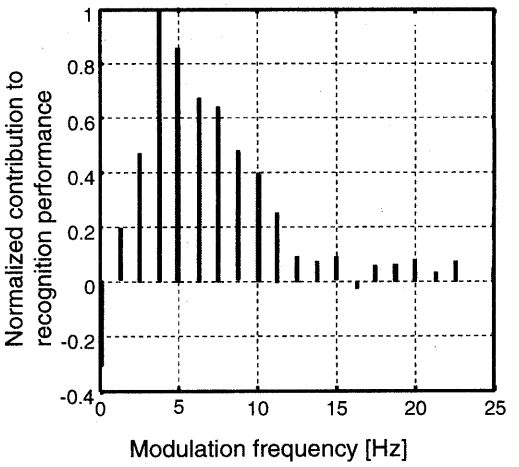


(a) Bellcore digit database.

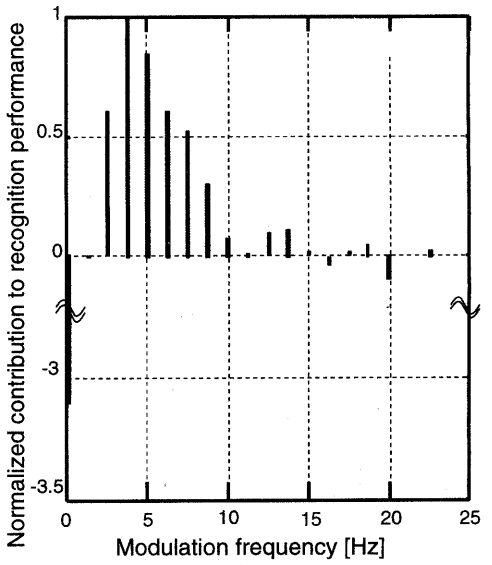


(b) Bellcore digit database degraded by additive noise (10 dB) and convolutional noise (HPF, 6 dB/oct).

図 6: Normalized contributions to recognition performance for different recognizers and different features.



(a) Clean



(b) Noisy

図 7: Normalized contribution to recognition performance for DFT filtering.