

TR - A - 0161

21

順向逆向両マスクングを併用した  
動的ケプストラムによる音声認識

升方 幹雄      相川 清明

1993. 3. 1

ATR 視聴覚機構研究所

〒619-02 京都府相楽郡精華町乾谷 ☎07749-5-1411

**ATR Auditory and Visual Perception Research Laboratories**

Inuidani, Sanpeidani, Seika-cho, Soraku-gun, Kyoto 619-02 Japan

Telephone: +81-7749-5-1411

Facsimile: +81-7749-5-1408

Telex: 5452-516 ATR J

# 順向逆向両マスクングを併用した動的ケプストラムによる音声認識

升方 幹雄† 相川 清明‡

† 豊橋技術科学大学 情報工学課程

‡ ATR 視聴覚機構研究所

## 概要

聴覚系には、先行する音声で現時点の音声をマスクする順向マスクングと、後続する音声で現時点の音声をマスクする逆向マスクングがある。これらのマスクングの時間周波数特性に基づいて音声の動的特徴を強調したスペクトルパラメータを動的ケプストラムと呼ぶ。順向マスクングのみを用いた動的ケプストラムを連続型 HMM のパラメータとして用いることの有効性は既に報告されている。本報告では動的ケプストラムを用いる音声認識において、順向マスクングの他に逆向マスクングを考慮する各種の方法について検討した結果について述べる。その結果、順向マスクングを考慮した動的ケプストラムと逆向マスクングを考慮した動的ケプストラムを併用することにより認識率を向上できることがわかった。なお、本研究ではケプストラムをより聴覚的な周波数軸であるメル尺度に変換したメルケプストラムを基本パラメータとして検討を行なっている。

## 1 はじめに

音声認識においては従来からケプストラムのような静的特徴を表すパラメータの他にデルタケプストラムのような音声の動的特徴を捉えるパラメータを併用することにより認識性能を向上できることが報告されてきた。一方、正弦波を用いた聴覚心理実験により、2つの継時的な音を与えられるとマスクングが生じることが知られている。先行する音声で現時点の音声をマスクすることを順向マスクング、後続する音声で現時点の音声をマスクすることを逆向マスクングという。この順向マスクングや逆向マスクングを考慮したスペクトルパラメータを動的ケプストラムという。デルタケプストラムは動的特徴のみを表しているが、動的ケプストラムは静的な特徴と動的な特徴の両方を表している。順向マスクングのみを用いた動的ケプストラムをパラメータとして連続型 HMM に用いることの有効性は [1] に示されている。[1] では順向マスクングのみを考慮していたが、順向マスクングに加え逆向マスクングを考慮することで音声認識の性能が向上することが期待される。

人間の聴覚には、周波数分解能は低い周波数では細かく、高い周波数では粗いという特性がある。そこで、本実験では線形周波数軸に基づいたケプストラムをより聴覚的な周波数軸となるようにメル尺度に非線形変換したメルケプストラムを基本パラメータとした。

本研究では、16次のパラメータとして、

- ケプストラム
- メルケプストラム
- 順向マスクングのみの動的メルケプストラム
- 順向マスクングと逆向マスクングが同時に作用する動的メルケプストラム

の場合について比較検討し、2種類の16次のパラメータを併用した32次のパラメータの組合せとして、

- ケプストラムとデルタケプストラムの併用

- メルケブストラムとデルタメルケブストラムの併用
- 順向マスクングのみの動的メルケブストラムとデルタメルケブストラムの併用
- 順向マスクングと逆向マスクングが同時に作用する動的メルケブストラムとデルタメルケブストラムの併用
- 順向マスクングのみの動的メルケブストラムと逆向マスクングのみの動的メルケブストラムの併用

の場合について比較検討した。

## 2 順向マスクングと逆向マスクングを考慮した動的ケブストラム

動的ケブストラムとは順向マスクングの時間周波数特性に基づいて動的特徴を強調したスペクトルパラメータ、あるいは更に逆向マスクングの時間周波数特性に基づいて動的特徴を強調したスペクトルパラメータである。順向マスクングとは先行する音声によるマスクングであり、時間の経過と共に減衰し、周波数に関して平滑化される。反対に、逆向マスクングとは後続する音声によるマスクングであり、同様に時間が離れると共に減衰し、周波数に関して平滑化される。

順向マスクングと逆向マスクングの時間周波数マスクングモデルを示したのが図1である。この図は一定のガウシアン状のスペクトルピークが時間と共に周波数軸状を低周波方向に一定の移動するような音声信号により形成されるマスクングパターンと、それによりマスクされたスペクトルを表している。マスクングパターンは順向逆向とも、現時点のスペクトルから時間的に離れるほど、スペクトルの高さは低下し、周波数平滑化が行われてピークの幅が広がる。この図では、現時点のスペクトルに先行する4時点の平滑化されたスペクトルの和によって順向マスクングのパターンが設定され、現時点に後続する4時点の平滑化されたスペクトルの和によって逆向マスクングのパターンが設定されている。

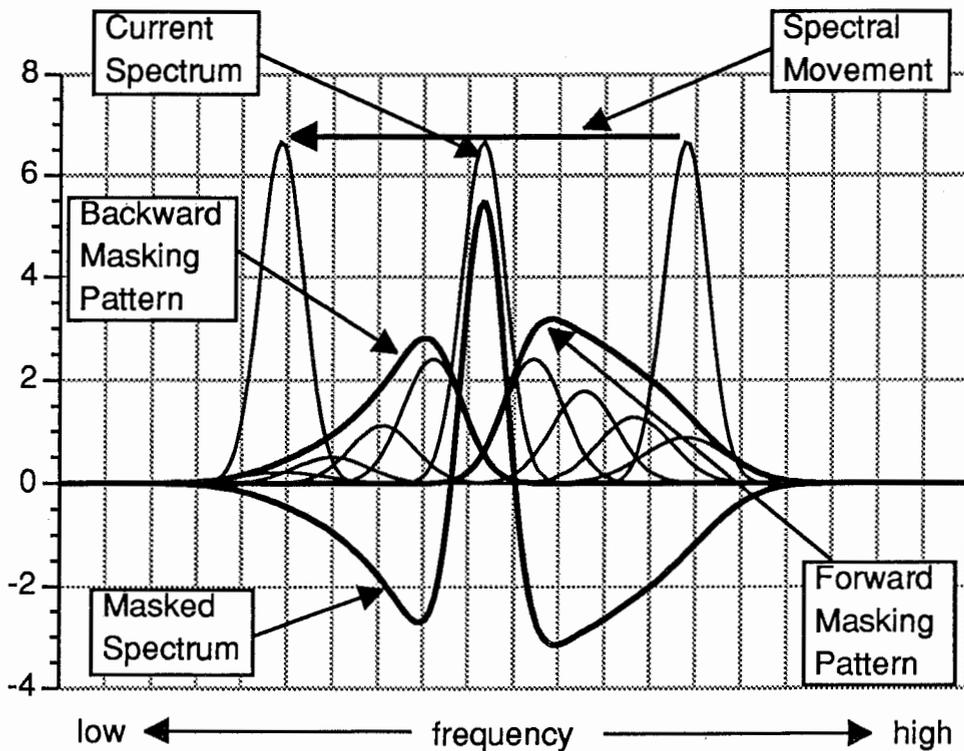


図 1: 時間周波数マスクングモデル

次に、マスクングモデルに基づいて動的ケブストラムを定式化する。時点  $i$  におけるマスクされた実効的なスペクトル  $B(\omega, i)$  は現時点のスペクトル  $C(\omega, i)$  からマスクングパターン  $M(\omega, i)$  を減算したものに

より求める。即ち、

$$B(\omega, i) = C(\omega, i) - M(\omega, i) \quad (1)$$

である。なお、 $\omega$  は周波数を表す。この式の両辺を逆フーリエ変換すると

$$b_k(i) = c_k(i) - m_k(i) \quad (2)$$

となる。 $c_k(i)$  はケプストラム係数、 $m_k(i)$  はマスキングパターンのケプストラム展開係数である。 $b_k(i)$  は、時点  $i$  における実効的なスペクトルの  $k$  次のケプストラム展開係数である。この特徴パラメータを動的ケプストラムと呼ぶ。

スペクトルを平滑化する演算はケプストラム領域ではリフタの乗算により行うことができる。従って、時点  $i$  における  $k$  次のマスキングレベル  $m_k(i)$  は、順向マスキングのマスキングレベル  $f_k(i)$  と逆向マスキングのマスキングレベル  $o_k(i)$  の和として表せる。

$$f_k(i) = \sum_{n=1}^{N_f} c_k(i-n)l_k(n) \quad (3)$$

$$o_k(i) = \sum_{n=1}^{N_b} c_k(i+n)r_k(n) \quad (4)$$

$$m_k(i) = f_k(i) + o_k(i) \quad (5)$$

$l_k(n)$  は  $n$  時点前のマスキングパターンを平滑化するリフタの  $k$  次の係数である。また、 $r_k(n)$  は  $n$  時点後のマスキングパターンを平滑化するリフタの  $k$  次の係数である。 $N_f$  は順向マスキングの影響の及ぶ最大時間長を、 $N_b$  は逆向マスキングの影響の及ぶ最大時間長を表している。周波数平滑リフタ  $l_k(n)$ 、 $r_k(n)$  の条件として以下の2つがある。

- 時間と共に通過帯域が徐々に狭くなり、利得が減衰すること
- 定常的なスペクトル時系列に対する式 (5) のマスキングレベルが現時点のケプストラムを越えない。即ち、 $k$  次の動的ケプストラムを求める時間フィルタの利得  $g_k$  が負にならないこと

この2つの条件を式で表すと、

$$l_k(n+1) \leq l_k(n) \quad (6)$$

$$r_k(n+1) \leq r_k(n) \quad (7)$$

$$g_k = 1 - \sum_{n=1}^{N_f} l_k(n) - \sum_{n=1}^{N_b} r_k(n) \geq 0 \quad (8)$$

となる。

### 3 ガウス型平滑化リフタ

本実験では、リフタとして [1] で提案された、ガウス型平滑化リフタを用いている。 $n$  時点前のスペクトルを平滑化するガウス型リフタの  $k$  次の係数は式 (9) のようになる。

$$l_k(n) = \alpha\beta^{n-1} \exp\left(-\frac{k^2}{2(p_0 - \mu(n-1))^2}\right) \quad (9)$$

同様に  $n$  時点後のスペクトルを平滑化するガウス型リフタの  $k$  次の係数は式 (10) のようになる。

$$r_k(n) = \gamma\delta^{n-1} \exp\left(-\frac{k^2}{2(q_0 - \nu(n-1))^2}\right) \quad (10)$$

ガウス型平滑化リフタでは  $p_0, q_0$  はそれぞれ 1 時点前、1 時点後の平滑化リフタのガウス分布の標準偏差を与える。  $\mu$  は順向マスクングでの標準偏差移動速度を与え、  $\nu$  は逆向マスクングでの標準偏差移動速度を与える。順向マスクングのガウス分布の標準偏差は時間を遡るほど線形に小さくなり、逆向マスクングのそれは時間を先行すると線形に小さくなるものとする。順向マスクングのリフタの利得は時間  $n$  と共に初期減衰率を  $\alpha$ 、中期減衰率を  $\beta$  として減衰する。逆向マスクングのリフタの利得は時間  $n$  と共に初期減衰率を  $\gamma$ 、中期減衰率を  $\delta$  として減衰する。ガウス型平滑化リフタは  $\mu \geq 0$ 、  $\nu \geq 0$ 、  $\alpha, \beta, \gamma, \delta \geq 0$  である限りは式 (6,7) の条件を満たす。また式 (8) の条件は

$$g_0 = 1 - \alpha \frac{1 - \beta^{N_f}}{1 - \beta} - \gamma \frac{1 - \delta^{N_b}}{1 - \delta} \geq 0 \quad (11)$$

となる。図 2 にガウス型の平滑化窓を持つマスクングリフタを示す。この図は本実験で用いるガウス型平滑化リフタは、マスクの対象となる時点から離れるに従ってリフタの利得が減衰し、また標準偏差が小さくなり通過帯域が狭くなっていくことを表している。

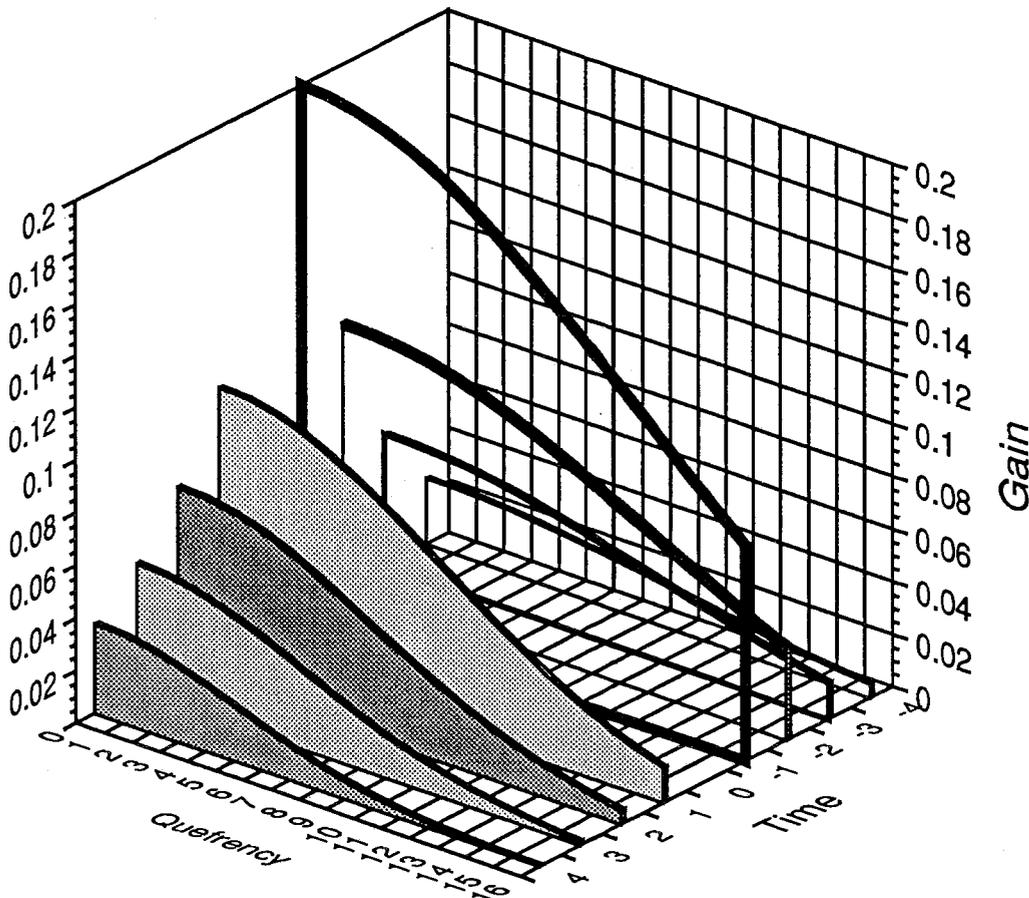


図 2: ガウス型の平滑化窓を持つ順向逆向の両方を含むマスクングリフタ

## 4 実験

### 4.1 音声資料

各種類のパラメータの性能を 6 子音認識実験で評価した。実験の対象とした 6 子音は /b,d,g,m,n,N/ である。用いた音声データベースは 1 人の男性話者 (MAU) が発声した日本語重要語 5240 単語及び文節ごとに区切って発声した 115 文章である。HMM の学習に用いた音韻は日本語重要語データベースのうち偶数番の 2620 単語からラベルに従って切り出した。語頭と語中の区別はしていない。学習された HMM の試

験用の音韻は重要語の奇数番の単語及び文節発声音声から収集した。音韻サンプル数は学習、試験共に最大 200 とした。

音声データのサンプリング周波数は 12kHz、フレーム周期は 10ms、フレーム長は 30ms である。スペクトルは 16 次の LPC(線形予測) 分析により求め、32 次または 16 次の LPC ケプストラムで表す。これらの 32 次または 16 次の LPC ケプストラムから、16 次のメルケプストラムを求める。メルケプストラム時系列から動的ケプストラム系列を求め、これを特徴パラメータとして対角型混合ガウス分布 HMM による学習・認識を行う。HMM の状態数は、最終状態を含め 4 で、ループのある状態数は 3 である。混合数は 8 とした。なお、以下の実験においてマスキングを考慮する場合は順向逆向とも標準偏差移動速度はすべて 1 とした。即ち、 $\mu = \nu = 1$  とした。

## 4.2 16 次のパラメータ 1 組による認識

### 4.2.1 従来の方法

ケプストラムをメルケプストラムに変換して得られたパラメータを用いた 6 子音認識実験結果を表 1 に示す [5]。表中の Frequency Warping=0.0 の欄はケプストラムによる結果を示している。ケプストラムに比べメルケプストラムの方が高い認識率が得られた。また表 2 には 32 次のケプストラムを 16 次のメルケプストラムに変換した場合の結果を示す。32 次のケプストラムは 16 次のケプストラムより細かなスペクトル包絡を表すことができる。32 次のケプストラムでの認識結果を表 2 に示す。実際に 32 次のケプストラムは 16 次のケプストラムより良い結果を示していることが分かる。よって 16 次のケプストラムから求めるより 32 次のメルケプストラムからメルケプストラムを求めるほうが認識率は上がると期待されたが、期待されたほどの結果は得られなかった。以下の実験では、表 1 の結果で良い結果を示した、16 次のケプストラムから frequency warping = 0.3 で変換された 16 次のメルケプストラム (表 1 中の A) を基本パラメータとした。

### 4.2.2 順向マスキングのみを考慮した動的メルケプストラム

この実験では、表 1 中の A の 16 次のメルケプストラムに順向マスキングのみを考慮して得られた 16 次の動的メルケプストラムを用いた。結果を表 5~表 7 に示す。この結果を見ると、メルケプストラムによる結果と比較して特に文節発声音声について大きな認識率の向上が見られる。よって、動的メルケプストラムは学習に用いられた音声とは異なる発話様式を持った音声に対して大きな効果があることが分かる。

### 4.2.3 順向逆向両マスキングを考慮した動的メルケプストラム

この実験では、表 1 中の A の 16 次のメルケプストラムに順向逆向両マスキングを考慮し、得られた 16 次のパラメータを用いた。まず、 $N_f=4, N_b=4$  と固定し  $\alpha, \beta, \gamma, \delta$  を変化させたときの結果を表 8~表 13 に示す。次に、 $N_b=1$  とし  $\beta, \gamma$  を変化させたときの結果を表 14~表 15 に示す。更に、 $N_b=2$  とし  $\beta, \gamma$  を変化させたときの結果を表 16~表 18、 $N_b=3$  とし  $\beta, \gamma$  を変化させたときの結果を表 19~表 21 に示す。また、 $\alpha=\beta=0.15$  と固定し、 $N_b$  を変えたときの結果を表 22~表 24 に示す。今回の実験では、順向マスキングと逆向マスキングを同時に考慮した 16 次のパラメータを用いる場合では順向マスキングのみを考慮した場合からの認識率の向上は見られなかった。

## 4.3 2 種類の 16 次のパラメータの組合せによる認識

ここでは 2 種類の 16 次のパラメータを併用する実験を行った。従って、この実験で用いるパラメータ数は 32 となる。表 3 には 16 次のケプストラムから 16 次のデルタケプストラムを求め、ケプストラムとデルタケプストラムを併用した場合の結果を示す。また、表 4 には 16 次のメルケプストラム (表 1 中の A) から 16 次のデルタメルケプストラムを求め、この 2 つをパラメータとして用いた場合の結果を示す。これらの結果よりケプストラムとデルタケプストラム、メルケプストラムとデルタメルケプストラムを組み合わせることで、ケプストラム、メルケプストラムを単独で用いる場合に比べて認識率を向上できることが分かる。特に文節発声音声に対する効果は著しい。

#### 4.3.1 動的メルケブストラムとデルタメルケブストラムの併用

表 25 に順向マスクングのみを考慮した 16 次のパラメータと 16 次のデルタメルケブストラムを併用した場合の結果を示す。この併用は、メルケブストラムとデルタメルケブストラムの併用より良い結果を示した。

また、表 26 に順向逆向両マスクングを考慮した 16 次のパラメータと 16 次のデルタメルケブストラムを併用した場合の結果を示す。この併用では更に、順向マスクングのみの動的メルケブストラムとデルタメルケブストラムの併用より良い結果を示した。

#### 4.3.2 順向マスクングのみを考慮した動的メルケブストラムと逆向マスクングを考慮した動的メルケブストラムの併用

この実験では、16 次のメルケブストラム時系列に順向マスクングのみを考慮した 16 次の動的メルケブストラムと、逆向マスクングのみを考慮した 16 次の動的メルケブストラムを併用した。順向マスクングのみを考慮したパラメータは表 6 中の C のパラメータを固定的に使用し、逆向マスクングのみを考慮する方のパラメータを変化させて認識率を調べた。表 27～表 29 は、 $N_b$  を 3 に固定し、 $q_0$  を 12, 18, 24 とそれぞれ変化させたときの結果である。表 30～表 32 は、 $N_b$  を 4 に固定し、 $q_0$  を 12, 18, 24 とそれぞれ変化させたときの結果である。表 33～表 35 は、 $N_b$  を 5 に固定し、 $q_0$  を 12, 18, 24 とそれぞれ変化させたときの結果である。更にこれらの結果の中から良い認識率を示したものに対して、逆向マスクングの初期標準偏差をより大きい値にした場合について検討を行った。それらの結果を示したものが表 36～表 37 である。表 36～表 37 中の記号は表 27～表 35 中の各パラメータに対応する。これらの実験では、2 種類の 16 次のパラメータを併用する方法の中で最も良い認識率を得ることができた。以上の結果より各実験で最も良い結果をまとめたものを図 3～図 6 に示す。図中のパラメータを表す記号の意味は、Cep: ケブストラム、Mcep: メルケブストラム、DyMc: 動的メルケブストラム、Fwd: 順向マスクング、Bwd: 逆行マスクング、 $\Delta$ Cep: デルタケブストラム、 $\Delta$ Mcep: デルタメルケブストラムである。

## 5 結論

マスクングの時間周波数特性を模擬した動的ケブストラムを用いる音声認識において、従来考慮していた順向マスクングに逆向マスクングを併用する効果について調べた。16 次のパラメータ 1 組を用いる場合の認識実験より、ケブストラムをメル尺度に変換したメルケブストラムをパラメータとすることで認識率の向上が見られた。また、メルケブストラムにおいても、順向マスクングを考慮することで、更に認識率が向上することが分かった。特に、学習に用いていない種類の音声に対する認識率の向上が著しい。しかし、順向逆向両マスクングを併用した動的メルケブストラムによる音声認識性能は順向マスクングのみの場合と同程度であった。次に、2 種類の 16 次のパラメータを組み合わせた認識実験を行った結果、順向マスクングのみを考慮した動的メルケブストラムと逆向マスクングのみを考慮した動的メルケブストラムを併用する場合、及び順向逆向両マスクングを考慮した動的メルケブストラムとデルタメルケブストラムを併用する場合の 2 種の場合で、順向マスクングのみを考慮した方法を越える認識率が得られることが分かった。

以上のように逆向マスクングを考慮することで、順向マスクングのみを考慮した従来方法よりも高い認識率が得られることが分かった。

## 参考文献

- [1] 相川、河原、東倉：“連続型 HMM を用いた動的ケブストラムによる音声認識” 信学技報 SP92-103 (1992-12).
- [2] 難波精一郎：“聴覚ハンドブック” ナカニシヤ出版 (1984).
- [3] 中川聖一：“確率モデルによる音声認識” 電子情報通信学会.
- [4] 古井貞照：“デジタル音声処理” 東海大学出版会.

- [5] Alan V. Oppenheim, Donald H. Johnson, "Discrete Representation of Signals", Proceedings of IEEE, Vol. 60, No. 6, Jun 1972.
- [6] 宮坂栄一: "急激な立ち上がり、立ち下がりをもつ正弦波信号に対する聴覚マスキングの時空間特性" 音響学会誌、Vol. 39, No. 9, pp. 614-623, (1983).
- [7] Tatsuya Hirahara, "Internal speech spectrum representation by spatio-temporal masking pattern", J. Acoust. Soc. Jpn. (E), Vol. 12, No. 2, 1991.

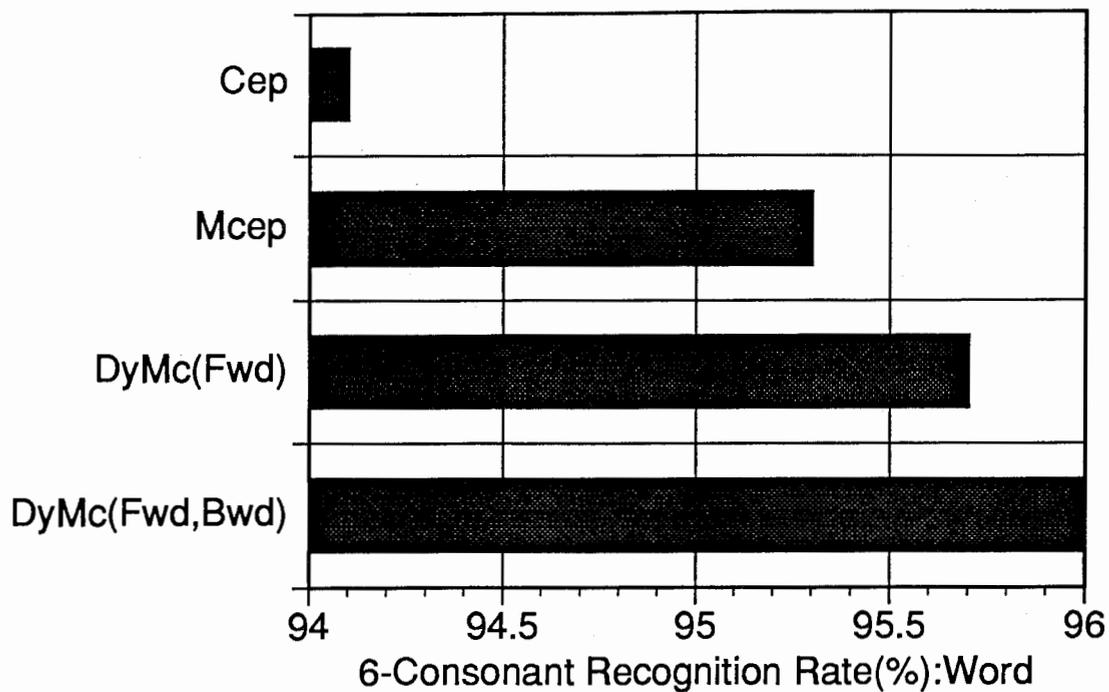


図 3: 16 次のパラメータのみを用いる各種方法の認識率の比較 : (単語単位発声)

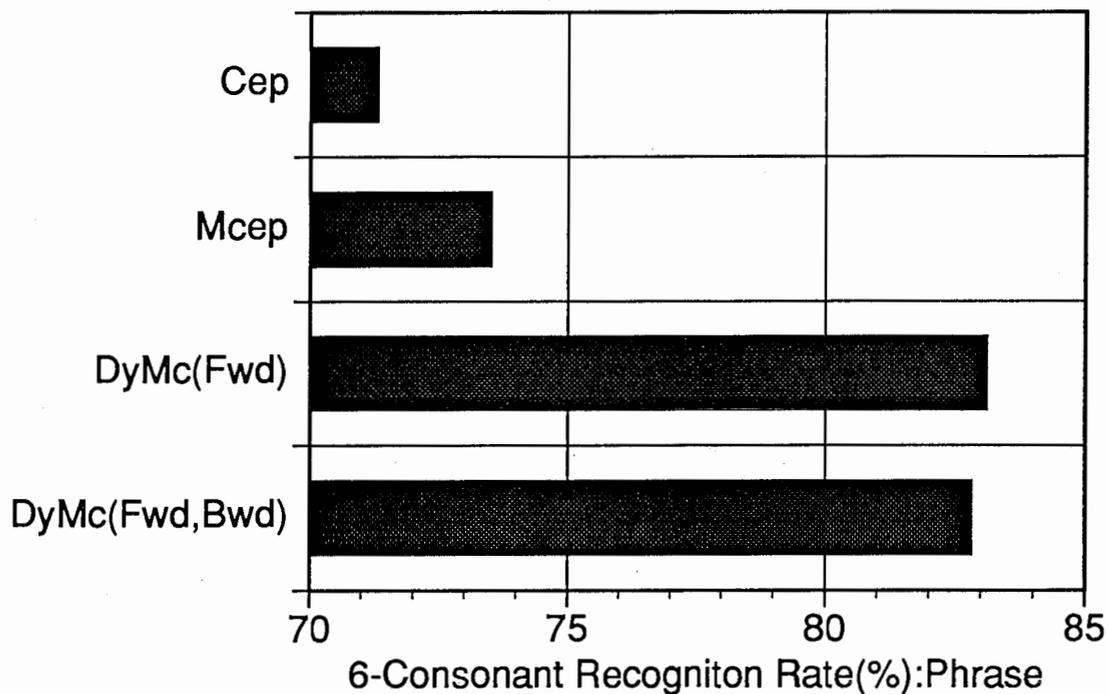


図 4: 16 次のパラメータのみを用いる各種方法の認識率の比較 : (文節単位発声)

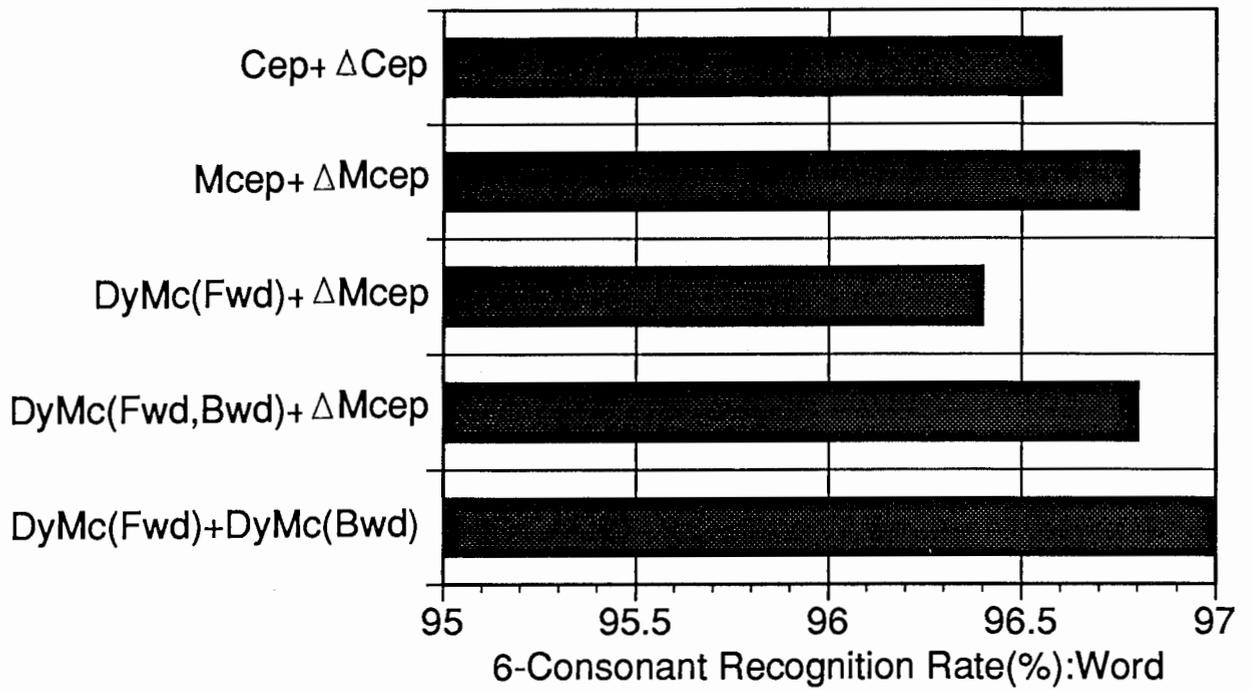


図 5: 16 次のパラメータを 2 組用いる各種の方法による認識率の比較 : (単語単位発声)

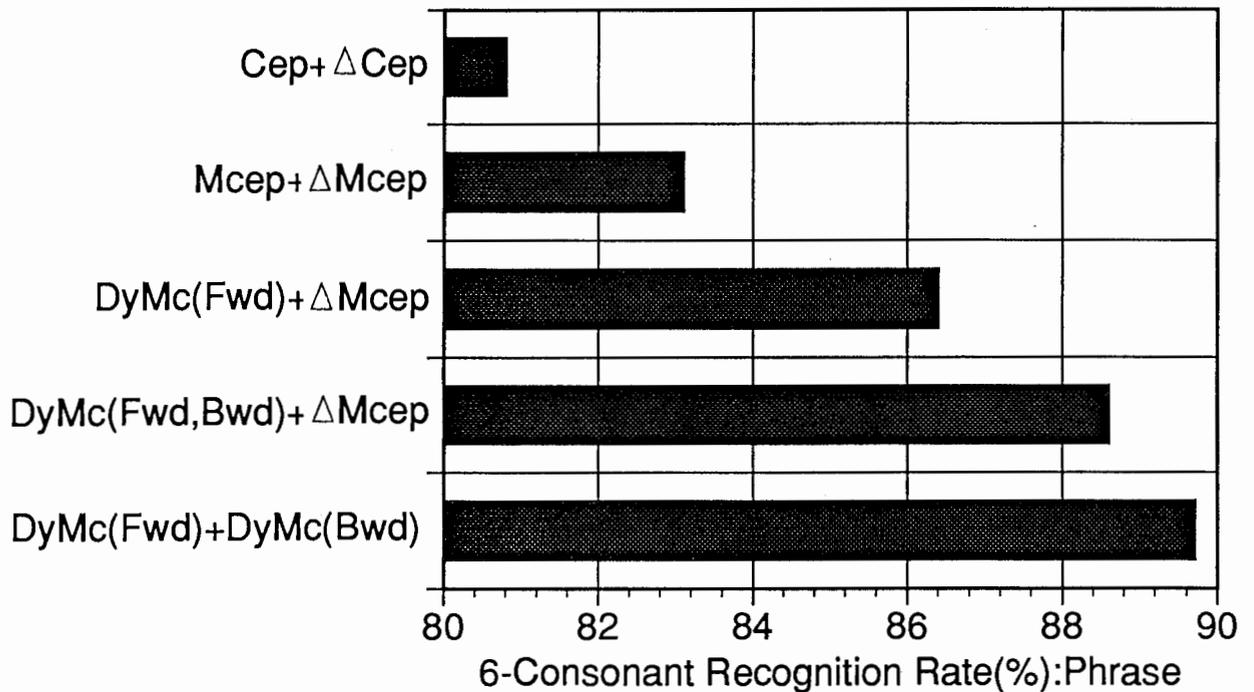


図 6: 16 次のパラメータを 2 組用いる各種の方法による認識率の比較 : (文節単位発声)

## 従来方法による認識実験結果

表 1: 16 次のケプストラムから求めた 16 次のメルケプストラムを用いた 6 子音認識実験結果

試験音声	Frequency Warping				
	0.0	0.2	0.3	0.4	0.5
単語	94.1	95.1	95.3	94.3	93.9
文節	71.3	71.8	73.5 <sup>A</sup>	73.3	65.9

表 2: 32 次のケプストラムから求めた 16 次のメルケプストラムを用いた 6 子音認識実験結果

試験音声	Frequency Warping			
	0.1	0.2	0.3	0.4
単語	95.3	94.6	95.2	94.4
文節	72.2	74.9	73.2	71.3

32 次のケプストラムによる 6 子音認識実験結果

単語… 94.7%

文節… 73.5%

表 3: 16 次のケプストラムから 16 次のデルタケプストラムを求め、ケプストラムとデルタケプストラムを併用した 6 子音認識実験結果

試験音声	RWL
	7
単語	96.6
文節	80.8

RWL … Regression Window Length

表 4: メル変換係数 0.3 の 16 次のメルケプストラムから 16 次のデルタメルケプストラムを求め、これらのメルケプストラムとデルタメルケプストラムを併用した 6 子音認識実験結果

試験音声	Regression Window Length			
	5	7	9	11
単語	96.8	97.1	96.9	96.9
文節	83.1	79.5	80.5	74.4

## 順向マスクングのみを考慮した場合の認識実験結果

表5: メル変換係数0.3のメルケプストラムに対して順向マスクングを行った場合の6子音認識実験結果:  $N_f = 3, \alpha = 0.3, \beta = 0.7$

試験音声	$p_0$					
	10	12	14	16	18	20
単語	94.6	94.7	94.8	95.1	95.3	95.3
文節	81.5	82.5	81.8	82.2	82.5	80.5

表6: メル変換係数0.3のメルケプストラムに対して順向マスクングを行った場合の6子音認識実験結果:  $N_f = 4, \alpha = 0.3, \beta = 0.7$

試験音声	$p_0$									
	8	10	12	14	16	18	20	22	24	32
単語	94.4	94.7	95.2	94.9	95.4	95.7	95.3	96.0	95.6	96.3
文節	78.6	81.2	83.4	82.3	82.5 <sup>B</sup>	83.1 <sup>C</sup>	82.9 <sup>D</sup>	82.7 <sup>E</sup>	82.3	82.1

表7: メル変換係数0.3のメルケプストラムに対して順向マスクングを行った場合の6子音認識実験結果:  $N_f = 5, \alpha = 0.3, \beta = 0.7$

試験音声	$p_0$					
	10	12	14	16	18	20
単語	94.9	95.2	95.2	95.0	94.9	95.3
文節	80.7	79.7	80.4	80.6	81.0	79.7

## 順向逆向両マスクングを考慮した場合の認識実験結果

表 8: 順向逆向両マスクングを考慮した場合の 6 子音認識実験結果:  $N_f = 4$ ,  $\alpha = 0.1$ ,  $\beta = 0.5$ ,  $p_0 = 18$ ,  $N_b = 4$ ,  $q_0 = 18$

		試験音声	Initial Backward Masking Decay						
			0.1	0.2	0.3	0.4	0.5	0.6	0.7
Medial	0.3	単語							
		文節							
Backward	0.4	単語					*		
		文節							
Masking	0.5	単語	94.7	93.9	94.8				
		文節	78.8	77.4	79.6				
Decay	0.6	単語				*			
		文節							
Masking	0.7	単語	94.5	95.3	95.2				
		文節	77.2	78.4	76.8				
Decay	0.8	単語			*				
		文節							
Masking	0.9	単語	94.2	94.7					
		文節	77.4	78.9					

表 9: 順向逆向両マスクングを考慮した場合の 6 子音認識実験結果:  $N_f = 4$ ,  $\alpha = 0.1$ ,  $\beta = 0.7$ ,  $p_0 = 18$ ,  $N_b = 4$ ,  $q_0 = 18$

		試験音声	Initial Backward Masking Decay						
			0.1	0.2	0.3	0.4	0.5	0.6	0.7
Medial	0.3	単語							
		文節							
Backward	0.4	単語					*		
		文節							
Masking	0.5	単語	94.6	94.8	94.7	*			
		文節	78.1	77.9	80.0				
Decay	0.6	単語							
		文節							
Masking	0.7	単語	94.3	94.7	*				
		文節	79.8	77.6					
Decay	0.8	単語							
		文節							
Masking	0.9	単語	94.3	94.6					
		文節	78.6	75.7					

表 10: 順向逆向両マスクングを考慮した場合の 6 子音認識実験結果:  $N_f = 4, \alpha = 0.1, \beta = 0.9, p_0 = 18, N_b = 4, q_0 = 18$

		試験音声	Initial Backward Masking Decay						
			0.1	0.2	0.3	0.4	0.5	0.6	0.7
Medial	0.3	単語							
		文節							
Backward	0.4	単語							
		文節							
Masking	0.5	単語	94.9	95.1	94.8	*			
		文節	80.0	81.2	79.8				
Masking	0.6	単語							
		文節							
Decay	0.7	単語	95.0	95.0	*				
		文節	81.9	77.5					
Decay	0.8	単語							
		文節							
Decay	0.9	単語	95.8	94.0*					
		文節	81.1	75.6*					

表 11: 順向逆向両マスクングを考慮した場合の 6 子音認識実験結果:  $N_f = 4, \alpha = 0.2, \beta = 0.3, p_0 = 18, N_b = 4, q_0 = 18$

		試験音声	Initial Backward Masking Decay						
			0.1	0.2	0.3	0.4	0.5	0.6	0.7
Medial	0.3	単語	95.2	94.4					
		文節	77.8	76.4					
Backward	0.4	単語					*		
		文節							
Masking	0.5	単語	95.3	94.7		*			
		文節	77.2	79.2					
Masking	0.6	単語							
		文節							
Decay	0.7	単語	95.1	95.3	*				
		文節	78.0	77.5					
Decay	0.8	単語							
		文節							
Decay	0.9	単語							
		文節							

表 12: 順向逆向両マスキングを考慮した場合の6子音認識実験結果:  $N_f = 4, \alpha = 0.2, \beta = 0.5, p_0 = 18, N_b = 4, q_0 = 18$

		試験音声	Initial Backward Masking Decay						
			0.1	0.2	0.3	0.4	0.5	0.6	0.7
Medial	0.3	単語	94.1	94.8					
		文節	79.1	81.8					
Backward	0.4	単語				*			
		文節							
Masking	0.5	単語	94.8	95.0	93.8				
		文節	79.5	80.2	78.1				
Masking	0.6	単語			*				
		文節							
Decay	0.7	単語	94.7	94.4					
		文節	80.5	77.1					
Decay	0.8	単語							
		文節							
Decay	0.9	単語	95.3						
		文節	81.9	*					

表 13: 順向逆向両マスキングを考慮した場合の6子音認識実験結果:  $N_f = 4, \alpha = 0.2, \beta = 0.7, p_0 = 18, N_b = 4, q_0 = 18$

		試験音声	Initial Backward Masking Decay						
			0.1	0.2	0.3	0.4	0.5	0.6	0.7
Medial	0.3	単語	95.9	95.3					
		文節	82.5	82.1					
Backward	0.4	単語			93.7				
		文節			76.4				
Masking	0.5	単語	95.8	94.4	*				
		文節	81.3	79.7					
Masking	0.6	単語	95.8	95.3					
		文節	81.3	78.8					
Decay	0.7	単語	95.4						
		文節	83.3 <sup>F</sup>	*					
Decay	0.8	単語	95.3						
		文節	82.7						
Decay	0.9	単語	95.3						
		文節	83.1						

表 14: 順向逆向両マスクングを考慮した場合の 6 子音認識実験結果:  $N_f = 4$ ,  $\alpha = 0.1$ ,  $p_0 = 18$ ,  $N_b = 1$ ,  $q_0 = 18$

		試験音声	Medial Forward Masking Decay						
			0.3	0.4	0.5	0.6	0.7	0.8	0.9
Initial	0.1	単語			94.6		95.2		94.9
		文節			76.9		79.6		80.2
Backward	0.2	単語			94.6		94.1		95.5
		文節			76.6		77.5		79.5
Masking	0.3	単語			94.7		94.1		95.0
		文節			77.4		76.5		81.4
Masking	0.4	単語			93.0		94.1		94.7
		文節			74.3		77.5		79.6
Decay	0.5	単語			94.5		94.6		95.1
		文節			78.2		80.2		79.7
Decay	0.6	単語			94.0		94.2		93.5
		文節			76.8		77.6		77.9
Decay	0.7	単語							*
		文節							

表 15: 順向逆向両マスクングを考慮した場合の 6 子音認識実験結果:  $N_f = 4$ ,  $\alpha = 0.2$ ,  $p_0 = 18$ ,  $N_b = 1$ ,  $q_0 = 18$

		試験音声	Medial Forward Masking Decay						
			0.3	0.4	0.5	0.6	0.7	0.8	0.9
Initial	0.1	単語							
		文節							
Backward	0.2	単語			94.6		95.5		94.7
		文節			81.2		81.8		81.9
Masking	0.3	単語			94.6		95.4		94.4
		文節			79.2		81.7		79.7
Masking	0.4	単語			94.1		94.5		*
		文節			79.9		80.0		
Decay	0.5	単語			94.1		*		
		文節			81.2				
Decay	0.6	単語			92.2				
		文節			74.8				
Decay	0.7	単語			*				
		文節							

表 16: 順向逆向両マスキングを考慮した場合の6子音認識実験結果:  $N_f = 4, \alpha = 0.1, \beta = 0.5, p_0 = 18, N_b = 2, q_0 = 18$

		試験音声	Initial Backward Masking Decay						
			0.1	0.2	0.3	0.4	0.5	0.6	0.7
Medial	0.3	単語							
		文節							
Backward	0.4	単語						*	
		文節							
Masking	0.5	単語	95.3	93.7	93.9	94.5			
		文節	77.2	73.7	76.7	79.8			
Masking	0.6	単語							
		文節							
Masking	0.7	単語	94.8	94.7	94.6	94.4	*		
		文節	78.4	79.1	78.6	76.9			
Decay	0.8	単語							
		文節							
Decay	0.9	単語	94.6	94.1	93.8	94.2			
		文節	77.3	77.8	79.9	78.1			

表 17: 順向逆向両マスキングを考慮した場合の6子音認識実験結果:  $N_f = 4, \alpha = 0.1, \beta = 0.7, p_0 = 18, N_b = 2, q_0 = 18$

		試験音声	Initial Backward Masking Decay						
			0.1	0.2	0.3	0.4	0.5	0.6	0.7
Medial	0.3	単語							
		文節							
Backward	0.4	単語							
		文節							
Backward	0.5	単語	95.2	95.1	94.5	94.5	*		
		文節	79.8	79.6	80.7	77.8			
Masking	0.6	単語							
		文節							
Masking	0.7	単語	94.4	94.1	94.8	93.6			
		文節	78.1	79.6	79.1	76.5			
Masking	0.8	単語							
		文節							
Decay	0.9	単語	94.4	94.9	94.3	*			
		文節	77.8	81.0	76.9				

表 18: 順向逆向両マスクングを考慮した場合の6子音認識実験結果: $N_f = 4, \alpha = 0.1, \beta = 0.9, p_0 = 18, N_b = 2, q_0 = 18$

		試験音声	Initial Backward Masking Decay							
			0.1	0.2	0.3	0.4	0.5	0.6	0.7	
Medial	0.3	単語 文節								
	0.4	単語 文節					*			
Backward	0.5	単語 文節	94.6 80.4	95.5 79.5	94.7 81.7	94.0 74.5				
	0.6	単語 文節								
Masking	0.7	単語 文節	95.1 80.3	94.7 80.2	95.5 80.8	*				
	0.8	単語 文節								
Decay	0.9	単語 文節	95.2 79.2	94.7 80.0	94.5 78.9					

表 19: 順向逆向両マスクングを考慮した場合の6子音認識実験結果: $N_f = 4, \alpha = 0.1, \beta = 0.5, p_0 = 18, N_b = 3, q_0 = 18$

		試験音声	Initial Backward Masking Decay							
			0.1	0.2	0.3	0.4	0.5	0.6	0.7	
Medial	0.3	単語 文節								
	0.4	単語 文節								
Backward	0.5	単語 文節	94.1 77.9	94.0 77.2	94.6 77.2	94.1 78.8	*			
	0.6	単語 文節								
Masking	0.7	単語 文節	94.2 76.8	94.0 79.7	94.5 79.6	*				
	0.8	単語 文節								
Decay	0.9	単語 文節	94.1 76.8	95.4 80.5	*					

表 20: 順向逆向両マスクングを考慮した場合の6子音認識実験結果: $N_f = 4, \alpha = 0.1, \beta = 0.7, p_0 = 18, N_b = 3, q_0 = 18$

		試験音声	Initial Backward Masking Decay						
			0.1	0.2	0.3	0.4	0.5	0.6	0.7
Medial	0.3	単語							
		文節							
Backward	0.4	単語					*		
		文節							
Masking	0.5	単語	94.8	93.9	95.6	94.2			
		文節	77.9	79.2	79.4	75.9			
Masking	0.6	単語				*			
		文節							
Masking	0.7	単語	94.9	94.6	94.3				
		文節	79.2	79.6	78.3				
Decay	0.8	単語							
		文節							
Decay	0.9	単語	94.5	94.6	*				
		文節	79.7	78.8					

表 21: 順向逆向両マスクングを考慮した場合の6子音認識実験結果: $N_f = 4, \alpha = 0.1, \beta = 0.9, p_0 = 18, N_b = 3, q_0 = 18$

		試験音声	Initial Backward Masking Decay						
			0.1	0.2	0.3	0.4	0.5	0.6	0.7
Medial	0.3	単語							
		文節							
Backward	0.4	単語							
		文節							
Masking	0.5	単語	95.3	95.7	93.9	*			
		文節	78.6	82.5	80.7				
Masking	0.6	単語			95.2				
		文節			78.1				
Masking	0.7	単語	94.5	94.7	*				
		文節	79.8	81.0					
Decay	0.8	単語							
		文節							
Decay	0.9	単語	95.0	95.3					
		文節	79.2	80.0					

表 22: 順向逆向両マスクングを考慮した場合の 6 子音認識実験結果:  $N_f = 4$ ,  $\alpha = 0.15$ ,  $p_0 = 18$ ,  $N_b = 4$ ,  $\gamma = 0.15$ ,  $q_0 = 18$

		試験音声	Medial Forward Masking Decay						
			0.3	0.4	0.5	0.6	0.7	0.8	0.9
Medial	0.3	単語 文節							
	0.4	単語 文節							
Backward	0.5	単語 文節			94.9 83.5 <sup>G</sup>		94.7 80.2		95.3 83.1
	0.6	単語 文節							
Masking	0.7	単語 文節			94.1 78.6		95.8 80.7		95.0 83.3 <sup>H</sup>
	0.8	単語 文節							
Decay	0.9	単語 文節			94.7 78.8		94.6 76.9		*

表 23: 順向逆向両マスクングを考慮した場合の 6 子音認識実験結果:  $N_f = 4$ ,  $\alpha = 0.15$ ,  $p_0 = 18$ ,  $N_b = 1$ ,  $\gamma = 0.15$ ,  $q_0 = 18$

試験音声	Medial Forward Masking Decay				
	0.5	0.6	0.7	0.8	0.9
単語	95.1	94.9	95.0	95.3	96.0
文節	77.1	80.4	80.7	82.2	82.8 <sup>I</sup>

表 24: 順向逆向両マスクングを考慮した場合の6子音認識実験結果:  $N_f = 4$ ,  $\alpha = 0.15$ ,  $p_0 = 18$ ,  $N_b = 2$ ,  $\gamma = 0.15$ ,  $q_0 = 18$

		試験音声	Medial Forward Masking Decay						
			0.3	0.4	0.5	0.6	0.7	0.8	0.9
Medial	0.3	単語 文節							
	0.4	単語 文節							
Backward	0.5	単語 文節			94.5 78.9		95.5 80.2		94.9 82.5
	0.6	単語 文節							
Masking	0.7	単語 文節			94.3 77.6		94.5 80.7		95.1 82.7
	0.8	単語 文節							
Decay	0.9	単語 文節			94.6 78.1		95.2 81.0		95.2 82.9

## 動的メルケプストラムとデルタメルケプストラムを併用 した場合の認識実験結果

表 25: 順向マスクングのみの動的メルケプストラムとデルタメルケプストラムを併用した場合の6子音認識実験結果

試験音声	Symbol	B			C			D			E		
	RWL	5	7	9	5	7	9	5	7	9	5	7	9
単語		96.3	96.5	97.0	96.4	96.6	96.8	96.3	96.4	96.7	95.9	96.0	96.9
文節		85.6	83.9	81.0	86.4	84.6	79.7	84.7	82.1	80.5	85.0	83.0	80.0

RWL ... Regression Window Length

表 26: 順向逆向両マスクングの動的メルケプストラムとデルタメルケプストラムを併用した場合の6子音認識実験結果

試験音声	Symbol	F			G			H			I		
	RWL	5	7	9	5	7	9	5	7	9	5	7	9
単語		96.8	96.9	96.9	97.1	96.7	96.8	95.9	96.4	96.9	96.8	96.9	96.5
文節		88.6	86.4	83.5	86.7	85.1	81.9	85.9	81.5	81.9	87.2	83.9	83.4

RWL ... Regression Window Length

# 順向マスクングと逆向マスクングを併用した場合の認識実験結果

- Forward Masking

- Time = 4
- Cutoff Quefrecny of Lifter = 18
- Initial Decay = 0.3
- Medial Decay = 0.7

(\* ...gain < 0)

表 27: 順向マスクングと逆向マスクングを併用した場合の6子音認識実験結果:  $N_b = 3, q_0 = 12$

		試験音声	Initial Backward Masking Decay						
			0.1	0.2	0.3	0.4	0.5	0.6	0.7
Medial	0.3	単語 文節							
	0.4	単語 文節							*
Backward	0.5	単語 文節		96.4 83.3	95.8 82.6	95.4 84.7		*	
	0.6	単語 文節							
Masking	0.7	単語 文節		96.6 82.1	95.8 83.1	96.3 84.6	*		
	0.8	単語 文節				96.6 82.7			
Decay	0.9	単語 文節		96.4 83.6	96.9 83.8	96.7* 86.9*			

表 28: 順向マスクングと逆向マスクングを併用した場合の 6 子音認識実験結果:  $N_b = 3, q_0 = 18$

		試験音声	Initial Backward Masking Decay						
			0.1	0.2	0.3	0.4	0.5	0.6	0.7
Medial	0.3	単語							
		文節							
Backward	0.4	単語							*
		文節							
Masking	0.5	単語		96.1	96.4	96.7		*	
		文節		83.1	84.1	83.5			
Masking	0.6	単語							
		文節							
Masking	0.7	単語		96.1	96.2	96.8	*		
		文節		83.1	83.4	85.0			
Decay	0.8	単語				96.9			
		文節				85.7			
Decay	0.9	単語		96.4	96.9	97.1*			
		文節		84.5	83.8	88.0*			

表 29: 順向マスクングと逆向マスクングを併用した場合の 6 子音認識実験結果:  $N_b = 3, q_0 = 24$

		試験音声	Initial Backward Masking Decay						
			0.1	0.2	0.3	0.4	0.5	0.6	0.7
Medial	0.3	単語						96.5	
		文節						86.5 <sup>J</sup>	
Backward	0.4	単語					96.5	96.4	*
		文節					86.0 <sup>K</sup>	86.6 <sup>L</sup>	
Masking	0.5	単語		96.0	96.4	96.6	97.1		
		文節		82.6	85.1	85.7	88.0 <sup>M</sup>	*	
Masking	0.6	単語				96.7	97.2		
		文節				85.4 <sup>N</sup>	87.5 <sup>O</sup>		
Masking	0.7	単語		96.8	96.8	96.8			
		文節		83.7	85.2	86.4 <sup>P</sup>	*		
Decay	0.8	単語				96.9			
		文節				86.2 <sup>Q</sup>			
Decay	0.9	単語		96.0	96.7	97.3*			
		文節		84.9	86.1	85.8*			

表 30: 順向マスクングと逆向マスクングを併用した場合の6子音認識実験結果:  $N_b = 4, q_0 = 12$

		試験音声	Initial Backward Masking Decay						
			0.1	0.2	0.3	0.4	0.5	0.6	0.7
Medial	0.3	単語 文節							
	0.4	単語 文節							*
Backward	0.5	単語 文節		96.6 82.8	96.4 82.8	96.4 83.5		*	
	0.6	単語 文節					*		
Masking	0.7	単語 文節		95.5 85.2	95.8 83.1	96.7* 86.7*			
	0.8	単語 文節				*			
Decay	0.9	単語 文節		96.2 83.1	96.8* 84.3*	96.4* 82.7*			

表 31: 順向マスクングと逆向マスクングを併用した場合の6子音認識実験結果:  $N_b = 4, q_0 = 18$

		試験音声	Initial Backward Masking Decay						
			0.1	0.2	0.3	0.4	0.5	0.6	0.7
Medial	0.3	単語 文節							
	0.4	単語 文節							*
Backward	0.5	単語 文節		95.8 83.1	96.2 83.6	96.4 84.7		*	
	0.6	単語 文節					*		
Masking	0.7	単語 文節		96.1 83.4	96.4 82.6	97.0* 85.9*			
	0.8	単語 文節			97.3 87.0	*			
Decay	0.9	単語 文節		96.5 83.8	96.7* 86.9*	97.3* 79.6*			

表 32: 順向マスクングと逆向マスクングを併用した場合の6子音認識実験結果:  $N_b = 4, q_0 = 24$

		試験音声	Initial Backward Masking Decay						
			0.1	0.2	0.3	0.4	0.5	0.6	0.7
Medial	0.3	単語							
		文節							
Backward	0.4	単語							*
		文節							
Masking	0.5	単語		96.1	96.4	96.4	96.9	*	
		文節		83.6	83.9	84.7	86.0		
Masking	0.6	単語				97.1	*		
		文節				87.5 <sup>R</sup>			
Decay	0.7	単語		96.9	96.5	96.9*			
		文節		83.5	83.5 <sup>S</sup>	85.9*			
Decay	0.8	単語			97.3	*			
		文節			87.3 <sup>T</sup>				
Decay	0.9	単語		96.7	97.2*	97.7*			
		文節		84.2	86.6*	82.5*			

表 33: 順向マスクングと逆向マスクングを併用した場合の6子音認識実験結果:  $N_b = 5, q_0 = 12$

		試験音声	Initial Backward Masking Decay						
			0.1	0.2	0.3	0.4	0.5	0.6	0.7
Medial	0.3	単語							
		文節							
Backward	0.4	単語							*
		文節							
Masking	0.5	単語		95.9	95.4	96.6		*	
		文節		83.5	82.9	84.6			
Masking	0.6	単語					*		
		文節							
Decay	0.7	単語		96.4	96.8	96.6*			
		文節		83.6	85.3	82.7*			
Decay	0.8	単語			*	*			
		文節							
Decay	0.9	単語		97.0	96.6*	96.8*			
		文節		83.5	78.6*	81.7*			

表 34: 順向マスクングと逆向マスクングを併用した場合の6子音認識実験結果:  $N_b = 5, q_0 = 18$

		試験音声	Initial Backward Masking Decay						
			0.1	0.2	0.3	0.4	0.5	0.6	0.7
Medial	0.3	単語 文節							
	0.4	単語 文節							*
Backward	0.5	単語 文節		95.9 83.1	96.2 85.7	96.4 84.2		*	
	0.6	単語 文節					*		
Masking	0.7	単語 文節		96.6 83.4	96.7 86.2	97.4* 85.0*			
	0.8	単語 文節			*	*			
Decay	0.9	単語 文節		97.0 84.9	97.3* 82.9*	97.5* 80.4*			

表 35: 順向マスクングと逆向マスクングを併用した場合の6子音認識実験結果:  $N_b = 5, q_0 = 24$

		試験音声	Initial Backward Masking Decay						
			0.1	0.2	0.3	0.4	0.5	0.6	0.7
Medial	0.3	単語 文節							
	0.4	単語 文節							*
Backward	0.5	単語 文節		95.9 83.4	96.9 83.7	96.6 83.3	97.2 87.2	*	
	0.6	単語 文節					*		
Masking	0.7	単語 文節		96.1 82.5	96.9 86.2 <sup>U</sup>	96.9* 85.3*			
	0.8	単語 文節			*	*			
Decay	0.9	単語 文節		97.2 84.5	96.9* 82.1*	97.5* 79.8*			

表 36: 初期標準偏差を大きくした逆向マスクングと順向マスクングを併用した場合の6子音認識実験結果

試験音声	Symbol	J		K		L		M		N		O	
	$q_0$	32	64	32	64	32	64	32	64	32	64	32	64
単語		96.6	96.8	96.9	97.3	96.4	97.0	97.4	97.0	97.2	96.8	97.5	97.3
文節		85.3	85.6	86.0	85.7	86.0	85.7	88.4	89.7	84.6	85.6	88.0	88.3

表 37: 初期標準偏差を大きくした逆向マスクングと順向マスクングを併用した場合の6子音認識実験結果

試験音声	Symbol	P		Q		R		S		T		U	
	$q_0$	32	64	32	64	32	64	32	64	32	64	32	64
単語		96.8	97.1	97.2	96.7	97.2	96.9	96.8	97.3	97.4	96.9	97.4	97.4
文節		87.2	88.2	85.9	85.1	87.3	86.2	85.8	84.6	87.0	87.0	86.9	86.9

Symbol J ~ U は表 29,32,35中の条件を示す