TR－A－0104

固定Q型蝸牛フィルタを用いた単語認識
－耐雑音性、耐残響性、話者変動の評価－

小原　和昭　　平原　達也

1991. 3.19

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

目次

第１章　はじめに

人間の聴覚系は音声認識装置として非常に優れた性能を持っている。例えば代表的な音声認識装置はＳＮ比が２５［ｄＢ］以下のノイズ環境下ではその認識性能は大きく劣化してしまう。一方我々はＳＮが５～１０［ｄＢ］程度の悪い環境下であっても話者の発声内容を容易に認識できる［1］。この様な聴覚系での信号処理機構を反映した特徴抽出を音声認識フロントエンドとして用い音声認識の性能を改善しようとする試みがなされてきている。

本報告は聴覚固定Ｑ型蝸牛フィルタを音声認識のフロントエンドとして用いた場合の雑音、残響,話者変動に対する認識耐性をDTWによる単語認識を通してＤＦＴフロントエンドと比較したものである。第２章では聴覚モデルを音声認識のフロントエンドとして用いた従来の認識実験結果をまとめるとともに、本実験で用いた実験条件、音声データの作成方法について述べる。第３章では実験内容とその結果について述べる。第４章では実験に用いたフロントエンドについて考察する。第５章では本研究のまとめと今後の課題について整理する。

第２章　特徴抽出のための信号処理

２－１　聴覚末梢系のモデルをフロントエンドとした音声認識

聴覚末梢系のモデルをフロントエンドとして音声認識を行なう試みがいくつかなされている。ここではそれらの試みを概観し、その結果をまとめた。聴覚モデルをフロントエンドとして音声認識を行った従来の実験について第２－１図にまとめた［2］［3］［4］［5］［6］［7］［8］［9］。これらの結果を整理すると聴覚末梢系のモデルをフロントエンドとして用いて特徴抽出を行いる効果については、

１）耐雑音性の向上［3］［4］［7］。

２）認識率の改善［5］［6］［8］

等が報告されている。

雑音下での聴覚モデルをフロントエンドとした音声認識の検討はいくつかなされているが、残響環境下での認識性能や話者依存性について,このようなフロントエンドを用いた検討は見あたらない。本研究は様々な雑音、残響環境下および話者変動下において,固定Ｑ型蝸牛フィルタをベースにしたモデルによる特徴抽出と従来の代表的なＤＦＴメルスケールによる特徴抽出とを,DTWによる単語認識を通して比較検討したものである。

２－２　ＤＦＴによる特徴抽出

ＤＦＴによる特徴抽出のブロック図を第２－２図に示した。メルスケールＤＦＴとバークスケールＤＦＴによる特徴抽出をおこなった。メルスケールＤＦＴでは２０［kＨｚ］サンプリングされた音声を２０［ｍｓ］のハミング窓を１０［ｍｓ］毎シフトし５１２点ＦＦＴし、１０［kＨｚ］の帯域をメル周波数軸上で６４［ｃｈ］に分割した各帯域に含まれるＦＦＴ離散スペクトルパワーを加算し、各帯域に含まれるＦＦＴ離散スペクトルパワー数で平均して対数をとることにより６４［ｃｈ］の特徴ベクトルを求めた。Ｈｚ周波数からメル周波数への変換は次式を用いた。

$$\mathrm{MEL}=\frac{1000}{\log_{10}(2.0)}*\log_{10}\left(1.0+\frac{\mathrm{Hz}}{1000.0}\right)$$

バークスケールＤＦＴでは２０［kＨｚ］サンプリングされた音声を２０［ｍｓ］のハミング窓を１０［ｍｓ］毎シフトし５１２点ＦＦＴし、１９.５[Bark]帯域をバーク周波数軸上で６４［ｃｈ］に分割した各帯域に含まれるＦＦＴ離散スペクトルパワーを加算し、各帯域に含まれるＦＦＴ離散スペクトルパワー数で平均して対数をとることにより６４［ｃｈ］の特徴ベクトルを求めた。Ｈｚ周波数からバーク周波数への変換は以下に示したZwickerの変換式[13]を用いた。

$$\mathrm{Bark}=13.0*\mathrm{atan}(0.76*\mathrm{kHz})+3.5*\mathrm{atan}\left(\frac{\mathrm{kHz}}{7.5}\right)^2$$

２－３　固定Ｑ型蝸牛フィルタ

本実験で用いた固定Ｑ型蝸牛フィルタの特性の一例を第２－３図に示した。低域の減衰特性が高域のそれよりもゆるやかなことが固定Ｑ型蝸牛フィルタの特徴である。１．５［Ｂａｒｋ］から１９．５［Ｂａｒｋ］までを６４分割して求めたＢａｒｋ周波数Ｂ［１］～Ｂ［６４］をもとに６４［ｃｈ］の固定Ｑ型蝸牛フィルタを求めた。特性はノッチフィルタとバンドパスフィルタの従続接続によって生成している［１０］。

ノッチフィルタの伝達関数Ｎ［ｓ］は

$$N[s]=\frac{\omega_p}{\omega_z}*\frac{s^2+(\omega_z/Q_z)*s+\omega^2{}_z}{s^2+(\omega_p/Q_p)*s+\omega^2{}_p} \qquad (2-1)$$

で与えた。ここで$\omega p, \omega z$はそれぞれ極と零点の角周波数でＢ［１］～Ｂ［６４］をＨｚ周波数に変換した後、ｐｒｅ－ｗａｒｐｉｎｇした周波数により設定した。

すなわち

$$\omega_p=\frac{\tan(F_d[n]*T_s*\pi)}{(T_s*\pi)}*\sqrt{R} \qquad (n=1 \quad \text{to} \quad 64)$$

$$\omega_z=\frac{\tan(F_d[n]*T_s*\pi)}{(T_s*\pi)}\Big/\sqrt{R} \qquad (n=1 \quad \text{to} \quad 64)$$

ここで

Ｆｄ[ｎ]は［Ｂａｒｋ］周波数Ｂ［ｎ］に対応する［Ｈｚ］周波数、
Ｔｓはサンプリング周期（１／２０［ｋＨｚ］）、
Ｒは極と零点の角周波数の比ωp／ωzで０．９７５と設定した。

またＱｚ，Ｑｐはそれぞれ極と零点のＱの値で今回の設計ではＱｚ＝７、Ｑｐ＝５と固定した。（２－１）式を双一次変換して２次のデジタルノッチフィルタＤＮ［ｚ］を求めた。すなわち

$$DN[z]=\frac{B_0*z^2+B_1*z+B_2}{z^2+A_1*z+A_2}$$

但し

$$B_0 = \left(\frac{\omega_p}{\omega_z}\right)^2 * \left\{\left(\frac{2}{T_s}\right)^2 + \left(\frac{\omega_z}{Q_z}\right)^2 * \left(\frac{2}{T_s}\right) + \left(\frac{2}{\omega_z}\right)\right\} \Big/ A$$

$$B_1 = \left(\frac{\omega_p}{\omega_z}\right)^2 * 2 * \left\{\omega_z{}^2 - \left(\frac{2}{T_s}\right)^2\right\} \Big/ A$$

$$B_2 = \left(\frac{\omega_p}{\omega_z}\right)^2 * \left\{\left(\frac{2}{T_s}\right)^2 - \left(\frac{\omega_z}{Q_z}\right)^2 * \left(\frac{2}{T_s}\right) + \left(\frac{2}{\omega_z}\right)\right\} \Big/ A$$

$$A_1 = 2 * \left\{\omega_p{}^2 - \left(\frac{2}{T_s}\right)^2\right\} \Big/ A$$

$$A_2 = \left\{\left(\frac{2}{T_s}\right)^2 + \left(\frac{\omega_p}{Q_p}\right)^2 * \left(\frac{2}{T_s}\right) + \left(\frac{2}{\omega_p}\right)\right\} \Big/ A$$

$$A = \left\{\left(\frac{2}{T_s}\right)^2 + \left(\frac{\omega_p}{Q_p}\right)^2 * \left(\frac{2}{T_s}\right) + \omega_p{}^2\right\}$$

である。

また　バンドパスフィルタの伝達関数Ｈ［ｓ］は

$$H[s] = \frac{\omega_b}{Q_b} * \frac{s^2 + (\omega_b / Q_z) * s + \omega^2{}_z}{s^2 + (\omega_p / Q_p) * s + \omega^2{}_p} \qquad (2-2)$$

で与えた。ここでω_bはバンドパスフィルタの共振角周波数でＢ［ｎ］（ｎはチャネル番号でｎ＝１～６４）をＨｚ周波数に変換した周波数Ｆｄ［ｎ］をｐｒｅｗａｒｐｉｎｇした周波数により設定した。

すなわち

$$\omega_b = \frac{\tan(F_d[n] * T_s * \pi)}{(T_s * \pi)} * \sqrt{R} \qquad (n = 1 \quad \text{to} \quad 64)$$

またQbはバンドパスフィルタのＱの値で今回の設計では各ＢＰＦの中心周波数Ｂ［ｎ］から低域カットオフ０．５［Ｂａｒｋ］、高域カットオフ０．５［Ｂａｒｋ］それぞれ離れた角周波数ＢＬ［ｎ］とＢＨ［ｎ］とをＨｚ周波数に変換した後ｐｒｅｗａｒｐｉｎｇした周波数により設定した。すなわち

$$Q_b = \frac{F_c[n]}{F_H[n] - H_L[n]}$$

Fc[n]はＢ［ｎ］に対応する［Ｈｚ］周波数をｐｒｅｗａｒｐｉｎｇした周波数、ＦＬ[ｎ]は（Ｂ［ｎ］－０．５）に対応する［Ｈｚ］周波数をｐｒｅｗａｒｐｉｎｇした周波数、ＦＨ[ｎ]は（Ｂ［ｎ］＋０．５）に対応する［Ｈｚ］周波数をｐｒｅｗａｒｐｉｎｇした周波数である。

（２－２）式を双一次変換して２次のデジタルバンドパスフィルタＤＨ［ｚ］を求めた。すなわち

$$DH[z] = \frac{B_0 * z^2 + B_1 * z + B_2}{z^2 + A_1 * z + A_2}$$

ここで

$$B_0 = \left(\frac{\omega_p}{Q_b}\right) * \left(\frac{2}{T_s}\right) \Big/ A$$

$$B_1 = 0.0$$

$$B_2 = \left(\frac{\omega_p}{Q_b}\right)^2 * \left(\frac{2}{T_s}\right) \Big/ A$$

$$A_1 = 2 * \left\{ {\omega_b}^2 - \left(\frac{2}{T_s}\right)^2 \right\} \Big/ A$$

$$A_2 = \left\{ \left(\frac{2}{T_s}\right)^2 - \left(\frac{\omega_b}{Q_b}\right) * \left(\frac{2}{T_s}\right) + \left(\frac{2}{\omega_p}\right) \right\} \Big/ A$$

但し

$$A = \left\{ \left(\frac{2}{T_s}\right)^2 + \left(\frac{\omega_b}{Q_b}\right) * \left(\frac{2}{T_s}\right) + {\omega_p}^2 \right\}$$

である。

以上の設計でＢａｒｋ周波数からＨｚ周波数への変換には以下の関係式を用いた。

```
if (Bark<=5)  then
    Hz=100*Bark
else  if (Bark>10.04) then
    Hz=1000.*exp{(Bark-8.85)/6.}
else
Hz=1000.*(Bark-1.5)/7.
```

２－４　モデル１

モデル１は固定Ｑ型蝸牛フィルタの特徴抽出とＤＦＴによる特徴抽出で認識率の差異がどの程度生じるかを検討するモデルである。すなわちモデル１は６４［ｃｈ］固定Ｑ型蝸牛フィルタでフィルタリングしてその出力をウインド長２０［ｍｓ］シフト１０［ｍｓ］のハミングウインドウで切出し、フレーム内に含まれるデータの平均パワーの対数を取り特徴ベクトルとするモデルである。

２－５　モデル２

モデル２はモデル１の時間軸方向の変動を滑らかにする平均化処理に加えて周波数方向の変動を滑らかにする処理を行なったモデルである。周波数軸方向の処理は６４［ｃｈ］出力に対しＬＰＦ処理を行った（ＬＰリフタリング）。このＬＰＦ処理はスペクトログラム上の周期の速い変動を減衰させることを目的に行った。ＬＰリフタリングはＢａｒｋ周波数方向にＬＰＦ処理を行うことで行った（周波数方向のデータを７６「ＫＨｚ］サンプリングとしてカットオフ周波数を１８「ＫＨｚ］に設定した）。
モデル１、２での信号処理を第２－４図にまとめた。

２－６　音声データ

音声の仕様を第２－５図にまとめた。サンプリングは２０［kＨｚ］で音声の開始点から終了点までをラベルデータに基づき切り出しを行った。単語セットとしてはＡＴＲ音声データベースの音韻バランス単語２１６語を使用した。話者は男性２名（ＭＳＴ，ＭＹＳ）女性２名（ＦＳＴ，ＦＮＹ）を用いた。

２－７　雑音の重畳

雑音による単語認識の性能劣化を調べるために音声データに種々のＳＮの雑音を重畳し雑音音声データを作成した。雑音源として２０［ｋＨｚ］でサンプリングされた５秒のピンク雑音を用いた。初めに音声のトータルパワーをもとめる。その音声と同じ時間長のノイズを開始点がランダムになるよう雑音源から切り出し、設定されたＳＮからノイズの振幅を設定し、音声に重畳した。ここでのＳＮは音声のトータルパワーと雑音のトータルパワーの比で定義した（グローバルＳＮ）。ＳＮが∞（ピュア），４０，２０，１０，５，０［ｄＢ］の雑音音声データを作成した。すなわち音声データをＳ［ｎ］，ランダムに切り出したノイズデータをＮ［ｎ］，雑音重畳された音声データをＳｎ［ｎ］とすると以下の式によって雑音重畳された音声データを作成した。

$$Sn[n] = S[n] + N[n]\frac{\sqrt{\sum_{n}(S[n]*S[n])}}{\sqrt{\sum_{n}(N[n]*N[n])}}*10^{-\frac{SN}{20}}$$

２－８　残響音声の生成

残響音声は原音声とＡＴＲ可変残響室で収録したインパルスレスポンスの畳み込みを行うことで生成した。インパルスレスポンスとしては残響時間の異なる５種類を５０［ｍｓ］の時間長で切り出して用いた。各々のインパルスレスポンスを第２－６図に示した。各々のインパルスレスポンスの５００［Ｈｚ］における残響時間はそれぞれ、約２００、２１０、４６０、６２０、１１００［ｍｓ］であった。残響下での音声データ作成のブロック図を第２－７図に示した。音声と残響のインパルスレスポンスを２０４８サンプルのＦＦＴとｏｖｅｒｌａｐ－ａｄｄ法によって残響音声を生成した［１１］。畳み込みによって生成された音声は原音声と同じ時間長になるように音声の終端部を切り捨てた。

第３章　認識実験

第２章で述べたモデル１、モデル２、ＤＦＴにより音声の特徴抽出を行いＤＴＷによる単語認識を行った。ＤＴＷとしてはスタッガードアレーＤＰ［１２］を用いた。整合窓ｒはテンプテートのフレーム長をＬｔ、認識単語のフレーム長をＬｗとすると

$$r = min(Lt, Lw) / 4 + 3$$

とした。格子点の間引き間隔を３として，文献［１２］のＤＰ３－１のアルゴリズムによって認識を行った。認識実験の構成図を第３－１図に示す。認識実験結果を以下に示す。

３－１　認識実験とその結果

実験１）ウインドウの影響について（１話者＝ＭＳＴ）

特徴抽出に与えるウインドウの影響を調べるために方形、ハミングの２種類の切り出しウインドウを用いて、モデル２でのノイズによる認識率変動を調べた。ハミングウィンドウ長は２０［ｍｓ］で１０［ｍｓ］シフトで、また方形ウィンドウ長は１０［ｍｓ］で１０［ｍｓ］シフトで特徴抽出を行った。結果を第３－２図に示した。ノイズによる認識率の変動はウインドウによらず同様な傾向を示した。また認識率には大きな差はみられなかった。以下の実験にはＤＦＴとの比較のためにハミングウインドウを用いた。

実験２）モデル１、２とＤＦＴの耐雑音性の比較（４話者）

各話者の一回目の発声をテンプテートとし二回目の発声に雑音を重畳して、その単語の認識実験を行った。４話者の単語認識でのＤＦＴ、モデル１、２の認識結果を第３－３図～第３－５図に示した。また４話者認識結果の平均を第３－６図に示した。ＳＮ４０［ｄＢ］以上ではＤＦＴの性能がモデルの性能と同等か５［%］程度上回ったが、ＳＮ２０［ｄＢ］以下ではモデル１、２の性能がＤＦＴによる認識率を５～２５［%］上回った。ＳＮの劣化に対してモデル１、２の認識率の劣化の方がＤＦＴの認識率の劣化よりも少なかった。またバークスケールDFTのほうがメルスケールDFTよりも低SNでの劣化が少なかった。

実験３）モデル１、２とＤＦＴの耐残響性の比較（１話者　ＭＳＴ）

モデル１、２とＤＦＴによる耐残響性能を調べるために、種々の残響時間を持つ音声データを作成して認識実験を行った。テンプレートは残響付加もノイズ付加もないクリーンな環境での発声を用いた。その結果を第３－７図に示した。残響環境下では残響時間によらずＤＦＴによる特徴抽出の性能がモデルの性能を上回った。また残響時間の変化に対しＤＦＴの認識率の変動はモデルの変動よりも少なく、実験に用いた条件内ではクリーンな環境から約５％の認識率の劣化があった。一方モデル１、２の認識率の劣化はクリーンな環境から約１６％であった。今回用いた最も残響の長い環境ではＤＦＴの方がモデル１、２に比べ１２［％］程度認識率が高かった。またバークスケールDFTのほうがメルスケールDFTよりも劣化が大きかった。

実験４）モデル１、２とＤＦＴの耐雑音性、耐残響性の比較（１話者　ＭＳＴ）

雑音、残響下でのＤＦＴとモデル１、２の認識率をクリーンなテンプレートを用いて比較した。残響インパルスＲ１（第２－６図　Ｒ１）を用いて残響下での音声を生成して、それに雑音を加えＳＮが∞（ピュア），４０，２０，１０，５，０［ｄＢ］の残響付加雑音音声データを作成した。認識結果を第３－８図に示した。残響下かつ雑音下での認識ではモデル１、２のほうがＤＦＴよりも２０［ｄＢ］以下の低ＳＮ下では認識性能が良いが、絶対的な認識性能はＳＮが２０［ｄＢ］で８０［％］程度で，どちらのモデルでも大きく劣化した。

実験５）正規化特徴量による雑音下認識（１話者　ＭＳＴ）

各話者の一回目の発声をテンプテートとし二回目の発声に雑音を重畳して、その単語の認識実験を行った。各単語の特徴ベクトルの最大値で特徴ベクトルの正規化を行い認識を行った。すなわち各単語の特徴ベクトルをS(m,n)とすると正規化特徴ベクトルS'(m,n)を次式で求めた。

$$S'(m,n)=\frac{S(m,n)}{\max\{S(m,n)\}}$$

$$\left[\max\{S(m,n)\}:\ \text{maximum of } \{S(m,n)\} \text{ for all m and n}\right]$$

メルＤＦＴ、モデル１、２による認識結果を第３－９図に示した。特徴量の正規化をおこなうことによりモデル１、２の認識率はＳＮによらずＤＦＴよりも高くなった。また低ＳＮでの認識率の大きな改善がみられた。ＤＦＴは特徴量の正規化を行うことで低ＳＮでの認識率が低下したがモデル２では全てのＳＮ

で認識率が向上した。

実験６）正規化特徴量による残響下認識（１話者　ＭＳＴ）

各話者の一回目の発声をテンプテートとし二回目の発声に残響付加を行った後、各単語の特微ベクトルの最大値で特徴ベクトルの正規化を行い認識を行った。メルＤＦＴ、モデル1,2による認識結果を第３－１０図に示した。特徴量の正規化をおこなうことによりモデル1,2の耐残響性が向上し認識率はＤＦＴと同程度になった。ＤＦＴでも特徴量の正規化を行うことで認識率が数%向上した。

実験７）テンプレートと異なる話者による単語認識率

男性話者ＭＳＴの一回目の発声をテンプテートとし、他の話者の発声した単語の認識実験を行った（テンプレート　ＭＳＴ、話者ＭＹＳ、ＦＳＴ、ＦＮＹ）。各単語の特微ベクトルの最大値で特徴ベクトルを正規化した場合としない場合とについて、ＤＦＴ、モデル１、２による認識結果を第３－１１、１２図に示した。異なる話者のテンプレートを用いた場合、ＤＦＴの特徴量による劣化１０［%］に対しモデル特徴量による劣化は３０［%］程度でＤＦＴ特徴量による認識の方がモデル１、２の特徴量による認識率よりも性能がよかった。特徴量の正規化をおこなうことによりモデル１、２の認識率は向上したがＤＦＴによる認識率の向上は少なかった。正規化を行ってもＤＦＴ特徴量による認識率の方がモデルによる認識率よりもよかった。。モデル１、２の特徴量による正規化なしの場合の特徴量による認識ではとくに女性話者に対する認識率の劣化が男性話者の１０［%］に対し３０［%］程度と大きかった。特徴量を正規化することにより話者間の変動が押えられた。

第４章 考察

この章ではＤＦＴによる特徴抽出とモデル１、２による特徴抽出の差異についてまとめた。

４－１　モデルとＤＦＴの特徴抽出の差

本実験で用いた４つの特徴抽出モデル（ＤＦＴメル、ＤＦＴバーク，モデル１、モデル２）の１０００［Ｈｚ］でのスペクトログラムを第４－１図に示した。これらを比べると、モデル２はＤＦＴに比べかなり滑らかな特徴抽出になっておりまた応答チャネルがＤＦＴより広くなっている。またモデル１はモデル２との差は大きくはないがモデル２はモデル１よりも周波数方向の変動が少ない特徴抽出となっている。

モデルとＤＦＴの特徴抽出の差を以下に示した。

（Ａ）フィルタ特性：

モデルの周波数分析は基底膜のフィルタリング特性を反映して低域と高域の減衰特性が非対象であるが、ＤＦＴは通過帯域内での特性は一様でフラットである。また固定Ｑ型蝸牛フィルタの帯域とＤＦＴの帯域を比べるとＤＦＴのほうが狭帯域である。固定Ｑ型蝸牛フィルタの周波数軸はｂａｒｋスケールであり分析帯域は約５.７［ｋＨｚ］でありメルＤＦＴの周波数軸はメルスケールで分析帯域が１０［ｋＨｚ］となっている。そのため最大応答を示すチャネルが異なっているがモデル１、２の固定Ｑ型蝸牛フィルタのほうが広帯域になっているのが分かる。またバークスケールDFTでは分析帯域はモデル１、２とほぼ等しいが各チャネルの周波数分解能はバークスケールDFTのほうがよい。

（Ｂ）フィルタ帯域ののオーバーラップ：

モデル１、２の固定Ｑ型蝸牛フィルタは各チャネルが約０．２８[bark]間隔に設定してあり隣接チャネル間で通過帯域が重なっている。一方ＤＦＴでは隣接チャネルの重なりはない。２００［Ｈｚ］＋５００［Ｈｚ］＋１０００［Ｈｚ］の入力に対するＤＦＴ，モデル１、２の出力を第４－２図に示した。ＤＦＴに比べ固定Ｑ型蝸牛フィルタの周波数分析はブロードになっている。

（Ｃ）時間積分による通過帯域：

モデル１、モデル２では２０［ｍｓ］のハミングウインドウで音声を切り出

しその区間の４００サンプルの時間軸の加算平均を１０［ｍｓ］毎に行った。そのため各フィルタ出力のカットオフ周波数が５０［Ｈｚ］程度となっている。またＤＦＴでもフレーム周期を１０［ｍｓ］としているため各チャネル出力のカットオフ周波数は５０［Ｈｚ］となっているので時間分解能は等しくなっている。

４－２　ノイズによるテンプレートと認識単語との距離の変動

ノイズによる認識性能の劣化の原因を調べるために、ノイズ重畳によるテンプレートと認識単語の距離の変動を調べた。クリーンな環境での１回目の発声と２回目の発声の距離を基準１として、テンプレートとノイズ付加による認識単語の距離の変化を測定した。その結果を第４－３図に示した。距離の変動はメルＤＦＴによる特徴抽出の方がＳＮが２０［ｄＢ］程度まではモデル１、２での距離変動よりも少なかった。これはＤＦＴに比べモデル１、２の各チャネルの周波数分析が広帯域であるためＳＮがＤＦＴよりも悪いためであると考えられる。ＳＮが比較的良いときにはこのＳＮの変化に対する距離の変化がＳＮの劣化に伴う認識率の変化に現れていると考えられる。すなわちＳＮが４０［ｄＢ］程度の時にはＤＦＴでの距離の変動対し、モデル１、２での距離の変化は大きく、認識率の劣化はモデル１、２の方が大きくなっている。

一方ＳＮが２０［ｄＢ］以下になるとモデル２での距離の変動はＤＦＴの距離変動よりも少なくなっている。このためにＳＮが２０［ｄＢ］以下ではモデル２の認識率がメルＤＦＴよりも良くなっていると考えられる。またモデル１では距離の変動がメルＤＦＴよりも大きくなっているが認識率はＤＦＴより良くなっている。この点を調べるためにＳＮの変化に対するＤＦＴとモデル１、２のスペクトログラムの変化を第４－４、５、６図に示した。これらの図からモデル１、２のスペクトログラムではＳＮが２０［ｄＢ］程度になっても大まかなスペクトログラム上の特徴は保存されているが、ＤＦＴではスペクトログラム上の山谷の特徴がつぶれてきているのがわかる。モデル１、２がノイズにあまり影響されない比較的滑らかな特徴量を取り出しているの対しＤＦＴによる特徴量はノイズによって大きく変化してしまっていると考えられる。すなわちＤＦＴの特徴抽出は細かな特徴をよく取り出している反面ノイズによってその特徴が失われやすい。一方モデル１、２の特徴抽出は比較的大まかな特徴を取り出しているためノイズによってもその特徴が失われ難くなっていると考えられる。そのためにＳＮが２０［ｄＢ］以下ではモデル１、２による認識率がＤＦＴによるそれよりも良くなっていると考えられる。

すなわちSNが悪くなったときモデル1,2の特徴ベクトルのクラス間距離がDFTよりも大きくなっていると考えられる。また特徴量の正規化によってテンプレートと認識単語間の距離がどう変化するかを第４－７図に示した。正規化によってモデル１、２の距離変動がＤＦＴよりも押えられているのがわかる。このことによりモデル１、２の認識率が正規化により向上した考えられる。一方ＤＦＴ特徴量を正規化すると２０［ｄＢ］以下の低ＳＮではモデル１、２よりも距離変動が大きくなっているのがわかる。

４－３　残響付加によるスペクトログラムの変化

第４－８，４－９、４－１０図に残響付加によるスペクトログラムの変動とその差分の一例を示した。モデル１、２に比べメルＤＦＴは残響付加によりスペクトログラムの変動が少なくなっている。この点からも残響環境下でのＤＦＴの認識率の劣化がモデル１、２よりも少ないことが推察できる。

４－４　特徴抽出にかかる時間の比較

モデル１、モデル２での処理時間はＤＦＴ処理時間の１７倍であった。

４－５　誤りパターン

第４－１１図にＳＮ４０［ｄＢ］における単語認識の誤りの一例を示した。この図は例えばＤＦＴによる特徴抽出において７５番目の単語（きゅーげき　付録A参照）が第１候補から第５候補までそれぞれ２０１番目（とーひょー），１２０番目（にょーぼー），１８番目（ひょーじゅん），５３番目（いらしゃる），２３番目（わがまま）の単語と誤認識されたことを示している。モデル１、２、ＤＦＴとも同じ単語について似たような認識誤りをおかしているのがわかる。

第５章　まとめと今後の課題

本研究では以下の点が明らかになった。

１．モデル１、２とＤＦＴの特徴抽出を単語認識を通じて比較検討した結果、認識すべき単語のＳＮが４０［ｄＢ］以上では両者の認識率の差はほとんどなかった。しかしノイズや残響がある環境での認識率はＳＮが２０［ｄＢ］程度以下では固定Ｑ型蝸牛フィルタを用いたモデル１、２のほうがＤＦＴよりも性能が良かった。

２．モデル１、２のＳＮによる認識率の劣化は特徴量のパワー正規化によって大きく改善され、ＳＮ１０［ｄＢ］においてＤＦＴの認識率１６［%］に対して７３［%］の認識率を得た。この認識率の改善は正規化によりテンプテートと認識単語間の距離変動を押えたことによる改善である。

３．残響環境下の認識率はモデル１、２いずれよりもよりもＤＦＴの方が良好であった。また残響環境の変化に対してもＤＦＴの認識率の変動がモデルの変動よりも少なかった。これはモデルによる特徴抽出よりもＤＦＴによる特徴抽出が残響環境での特徴量の変動が少ないためである。

４．モデル1,2の特徴量を正規化することにより対残響性は向上したがＤＦＴの性能は越えなかった。

５．テンプレートと話者が異なる場合の認識率の変動はＤＦＴの性能がモデル１、２の性能よりも良かった。この場合にもモデル１、２の特徴量を正規化することにより認識率は改善されたがＤＦＴの性能をこえることはなかった。

６．総じてＤＦＴによる特徴抽出は今回の考慮した範囲での残響環境の変動に対し良好であるが低ＳＮでの認識率の劣化が著しい。またモデル１、２による特徴抽出は低ＳＮで効果が認められたが、その効果は低ＳＮでの認識時に限られ、残響環境下での認識率の劣化はＤＦＴよりも大きかった。

７．モデル特徴量の正規化により耐雑音性,耐残響性が向上し,話者変動による認識率の変動も押えられることを確認した。

ＤＦＴやモデル１、２による特徴量が環境によって大きく変動してしまうため環境変動によって大きな認識率の劣化が生じる。今回のモデルで考慮した末梢系での信号処理のみでは人間の認識能力に匹敵する特徴抽出を行なうことは難しい。今後の課題としては、音声の認識においてどの様な特徴量を人が利用して様々な環境の変化に対応しているのかを検討し、この様な環境変動の影響を受けずらい音声特徴量の抽出方法とその特徴量を用いた認識の検討が必要である。

参考文献

[1] Pollack,Pikett:"Masking of speech by noise at high sound levels", J.Acoust.Soc.Am.30, pp127-130(1958)

[2] Blomberg,M.,et.al(1984):"Auditory Models and isolated word recognition," Q Prog.Stat. Rep.,SpeechTransmiss. Lab.(Royal Institute of Technology,Stockhom),pp.1-15

[3] Ghiza,O.(1988):"Temporal Non-place information in the auditory-nerve firing patterns as a front-end for speech recognition in a noisy environment,"J.of Phonetics Vol.16,No.1 pp.109-123

[4] 佐久嶋、磯、吉田、渡辺：聴覚モデル分析を用いた音声認識”、音学講論,１－３－９，１９８８．１０

[5] Hunt,M.J,C.Lefebvre(1988):"Speeker Dependent and Independent Speech RecognitionExperiments with an Auditory Model," Proc. IEEE Int. Conf. Acoustics,Speech&Signal Processing,ICASSP-88,pp215-218(1988)

[6] Cohen,J.R.(1989):"Application of an Auditory model to speech recognition,"J.Acoust.Soc.Am.85(6),pp.2623-2629

[7] Paterson,R. Hirahara,T.(1989):"HMM Speech Recognition using DFT and Auditory Spectrogram,"ATR Tech.Report TR-A-0063

[8] H.Hamada,T.Hirahara,A.Imamura,T.Matsuoka,R.Nakatu(1989):"Auditory-Based Filter-Bank Analysis as a front-end Processor for Speech Recognition," Proc. Eurospeech 89, Vol.2 pp.396-399

[9] T.Hirahara,H.Iwamida(1990):"Auditory Spectrograms in HMM Phoneme Recognition,"1990 Int. Conf.on Spoken Language Processing,ICSLP-90, pp.381-384

[10] 駒木根, 平原："蝸牛の周波数分解機能を模擬するフィルターバンクの一構成法", 信学技報 SP87-45 (昭61)

[11] A.V.Oppenheim & R.W.Schafer: "ディジタル信号処理（下）", (伊達訳) コロナ社 (昭53)

[12] 鹿野、相川:"Staggerered Array DP マッチング”、信学論誌,61-D,9, pp.657-544（昭５７）

[13] Zwicker, et al.(1980):"Analytical expression for critical-band rate and critical band width as a function of frequency", J.Acoust.Soc.Am.74(3),pp1523-1525

Blomberg et al.(1984)[2]
Model: 5 different spectral scaling (FFT, Bark, Phon, sone, domin)
Task: 9 Swedish vowels and 18 Swedish consonants
Recognition: linear normalization of time
Results: FFT gives the best score

Ghiza(1988) [3]
Model: 85 ch log scale spacing cochlea filter+EIH
Task: 39-word alpha-degits(2male2female)
Recognition: DTW
Results: Same recognition score comparing DFT front end ,Better noise immunity

Sakushima et al.(1988)[4]
Model: Seneff model
Task: speaker dependent 250 Words
Recognition: DTW
Result: Better noise immunity

Hunt et al.(1988)[5]
Model: CBF+ Onset detector ,Periodicy detector +Linear Discriminant analysis
Recognition: DTW
Task: talker independent and dependent, Number
Result: In all cases ,Better score comparing with Mel Cepstrum

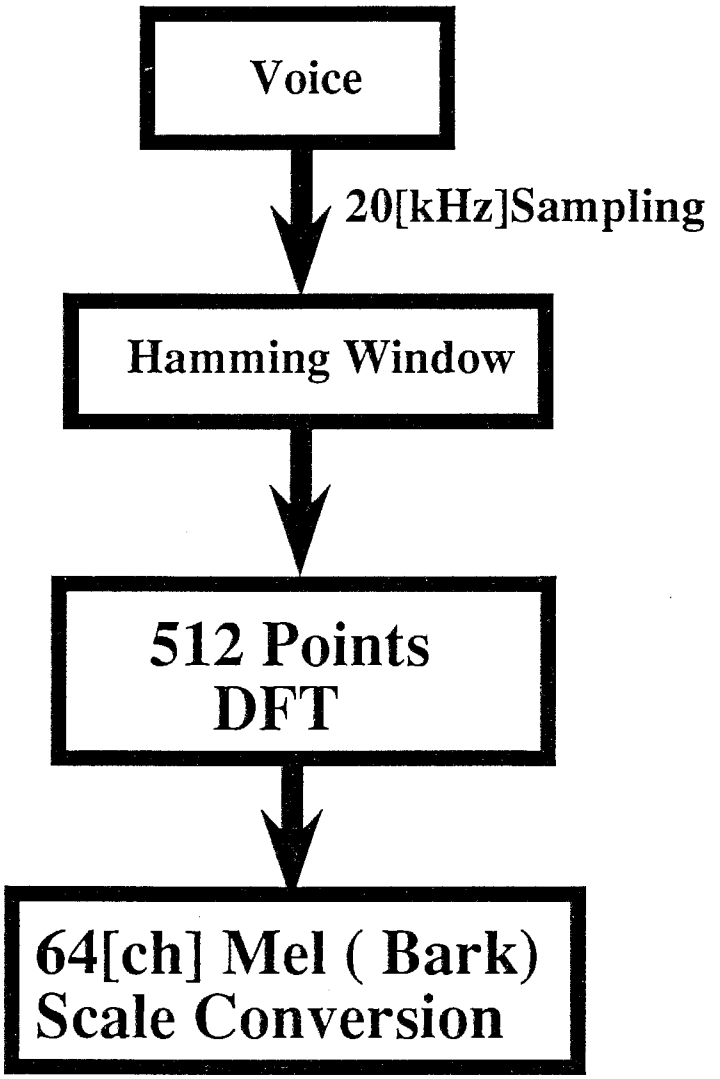
Cohen(1989)[6]
Model: 20 ch CBF+loudness scaling+ adaptation
Recognition: HMM (Training 100 Words, Test 50 Words)
Task: Words
Results: Better score comparing to Filter Bank Only (Error 6.3-[%]>3.9[%])

Patterson et al.(1989)[7]
Model: S AS
Recognition: HMM
Task: 14 consonants
Results: When code book size is small, better recognition score comparing DFT front end

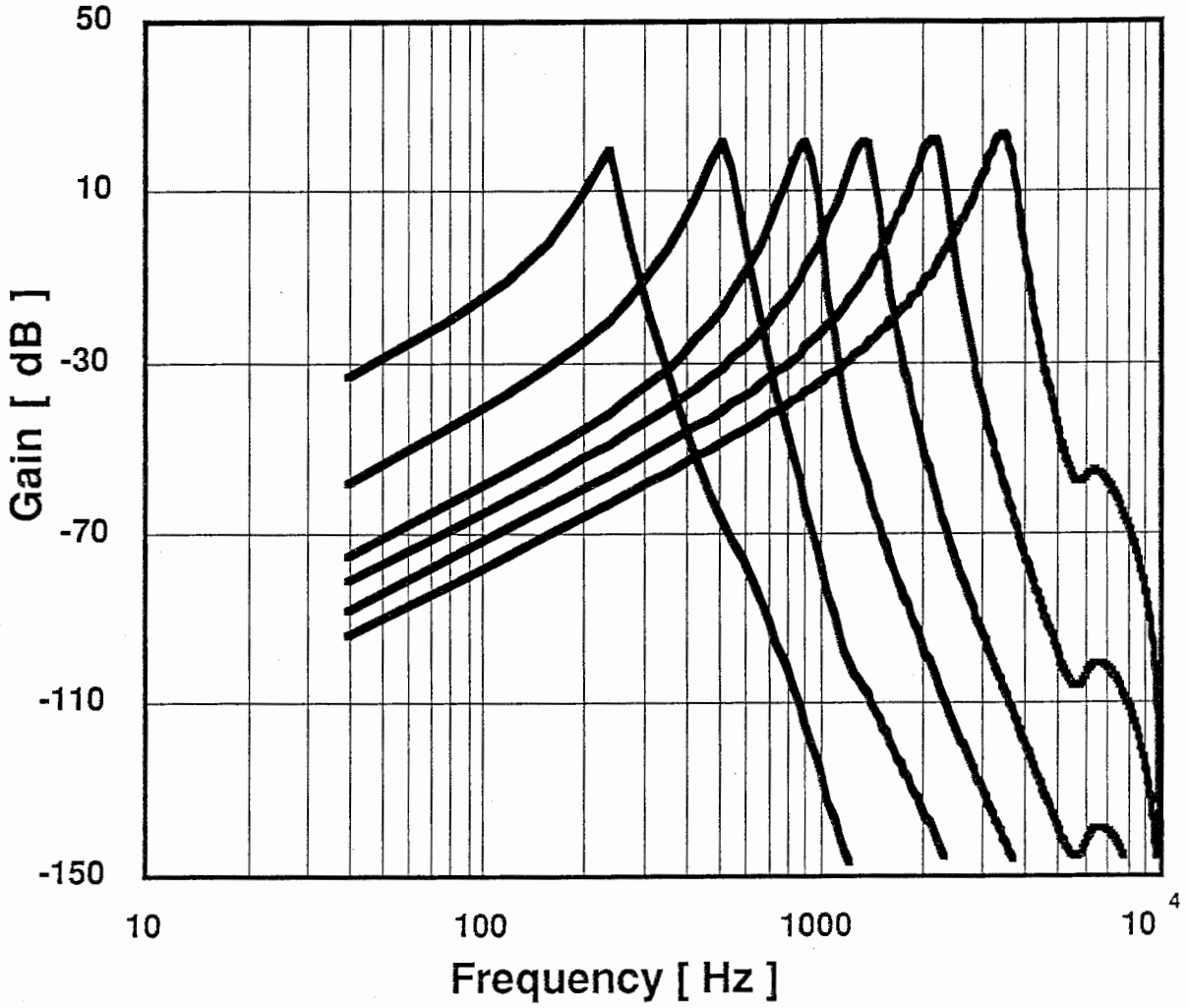
Hamada et al.(1989)[8]
Model: 35 ch CBF+HWR+Integration+Log+LIN
Recognition: DTW, HMM
Task: 14 consonants
Results: Same or better recognition score comparing LPC analysis

Hirahara et al.(1990)[9]
Model: Adaptive Q Chchlea Filter+Adaptation Model+LIN
Recognition: HMM
Task: 18 phoneme
Results: Without LINH,DFT gives the best score
With LINH „Adaptive Q Model gives the best score

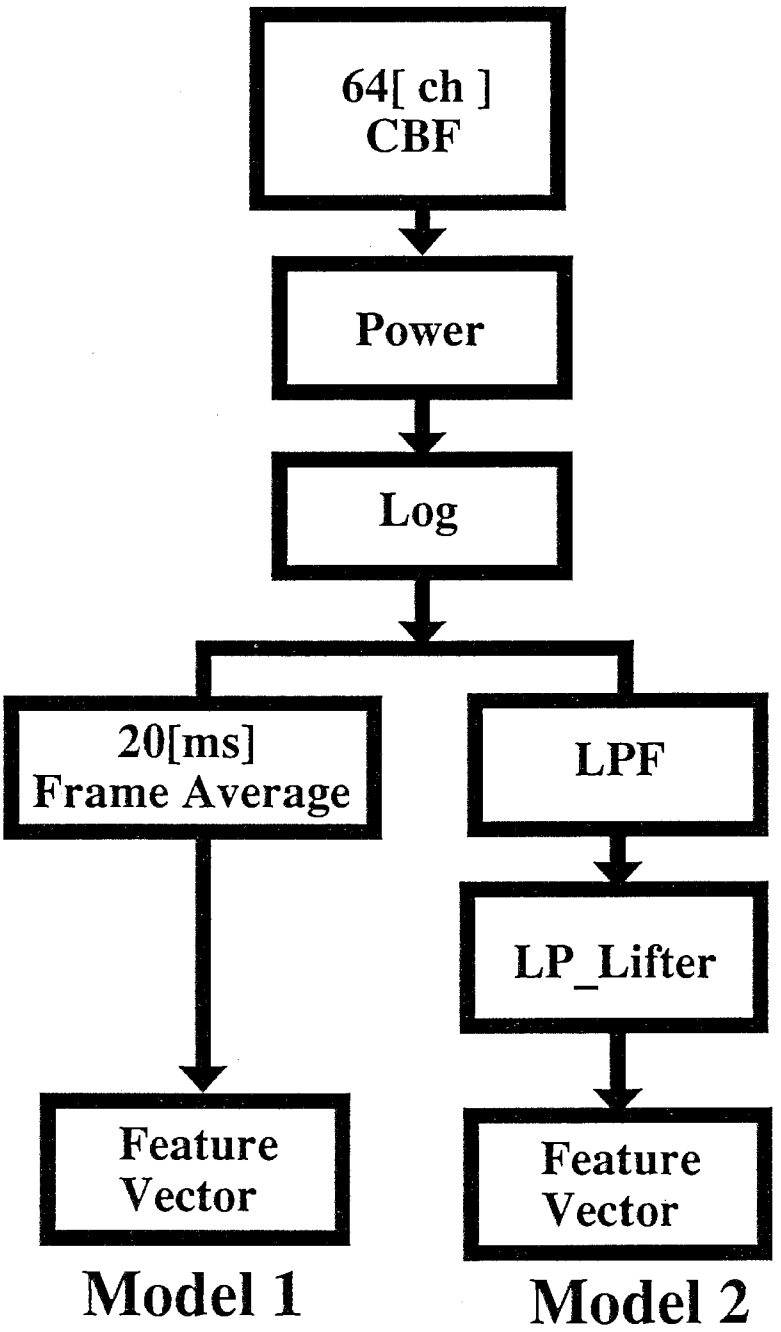
第２－１図　聴覚モデルによる音声認識の研究



第２－２図ＤＦＴによる特徴抽出



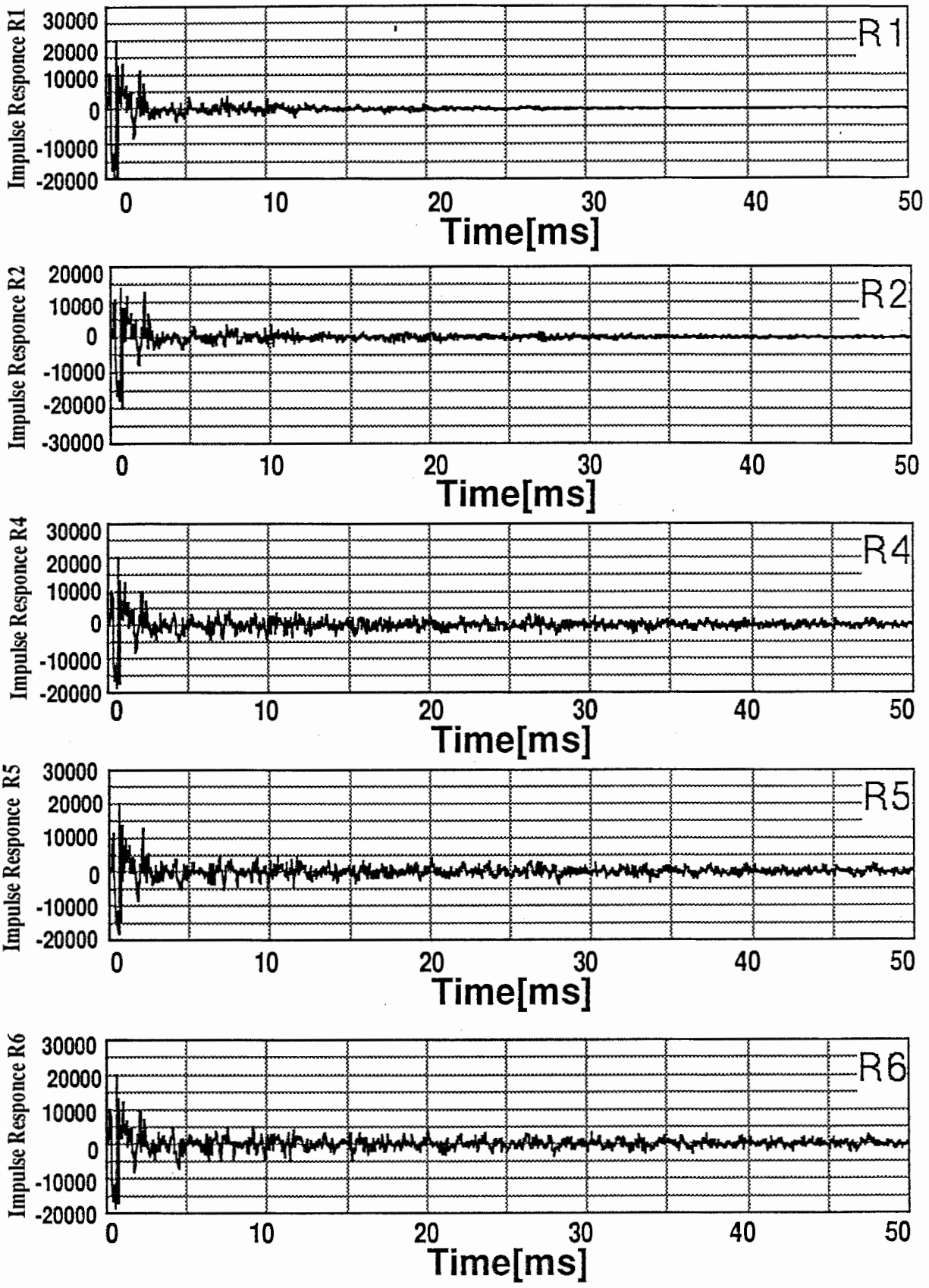
第２－３図 基底膜フィルタの特性の一例



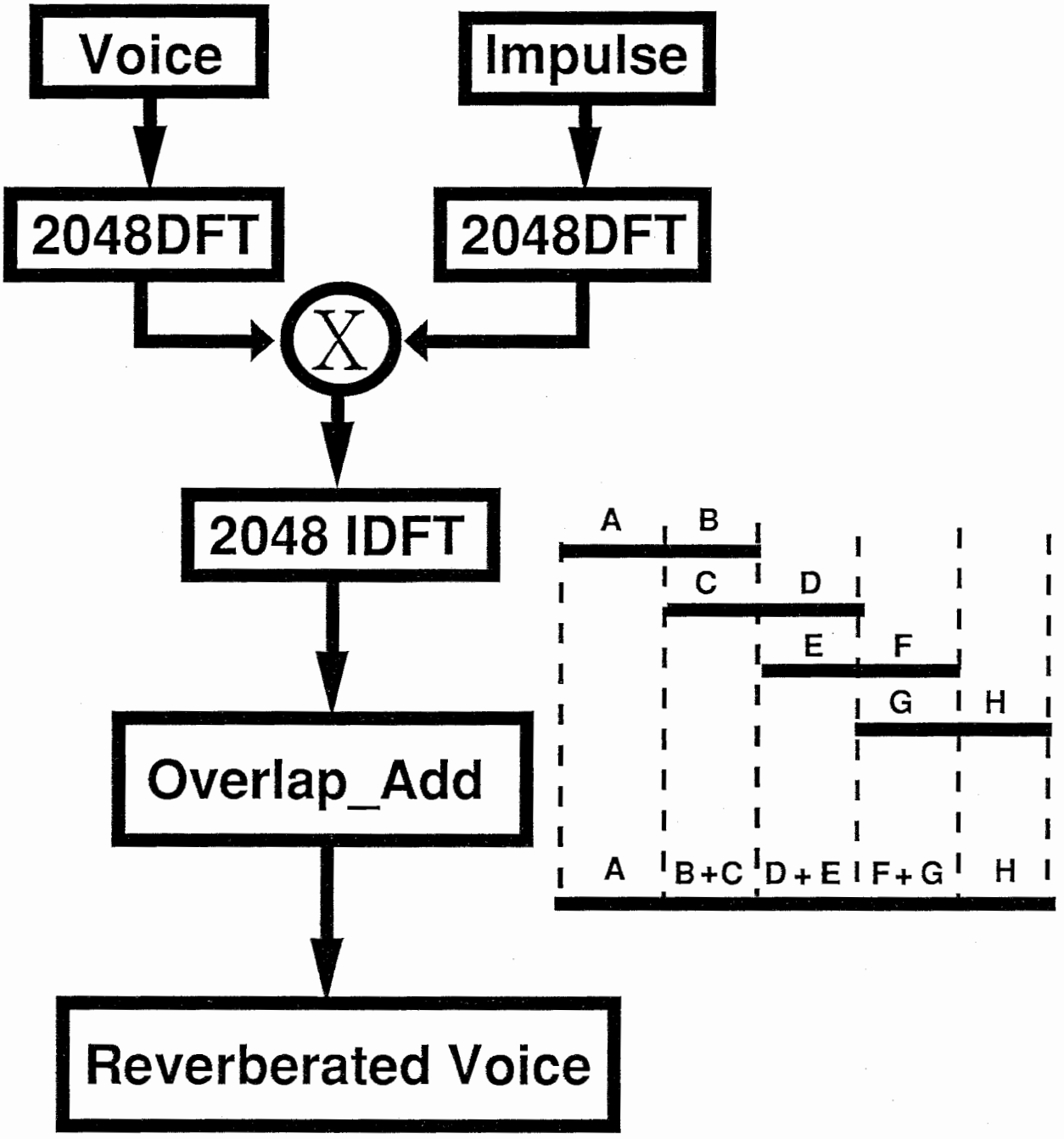
第２－４図　モデル１とモデル２による特徴抽出

サンプリング	２０［ｋＨｚ］
話者	男性：ＭＳＴ，ＭＹＳ　女性ＦＳＴ，ＦＮＹ
単語セット	ＡＴＲ　音声データベース　音韻バランス単語２１６語
プリエンファシス	なし
雑音音声生成	雑音の音声への加算
残響音声生成	残響インパルスと音声の畳み込み
特徴抽出	モデル１　モデル２　メルスケールＤＦＴ
距離尺度	ユークリッド距離
認識部	スタッガードアレイＤＴＷ

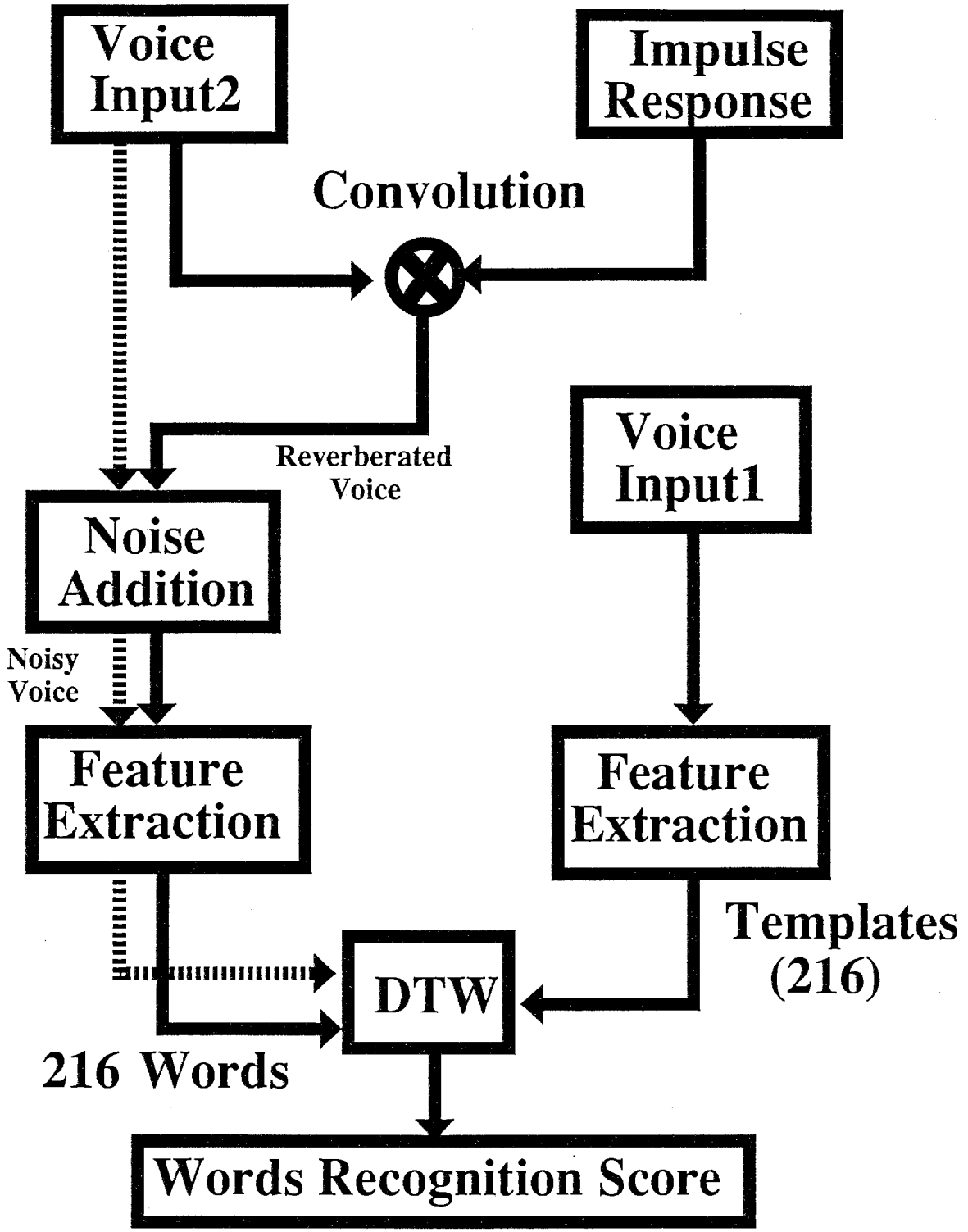
第２－５図　音声データの仕様



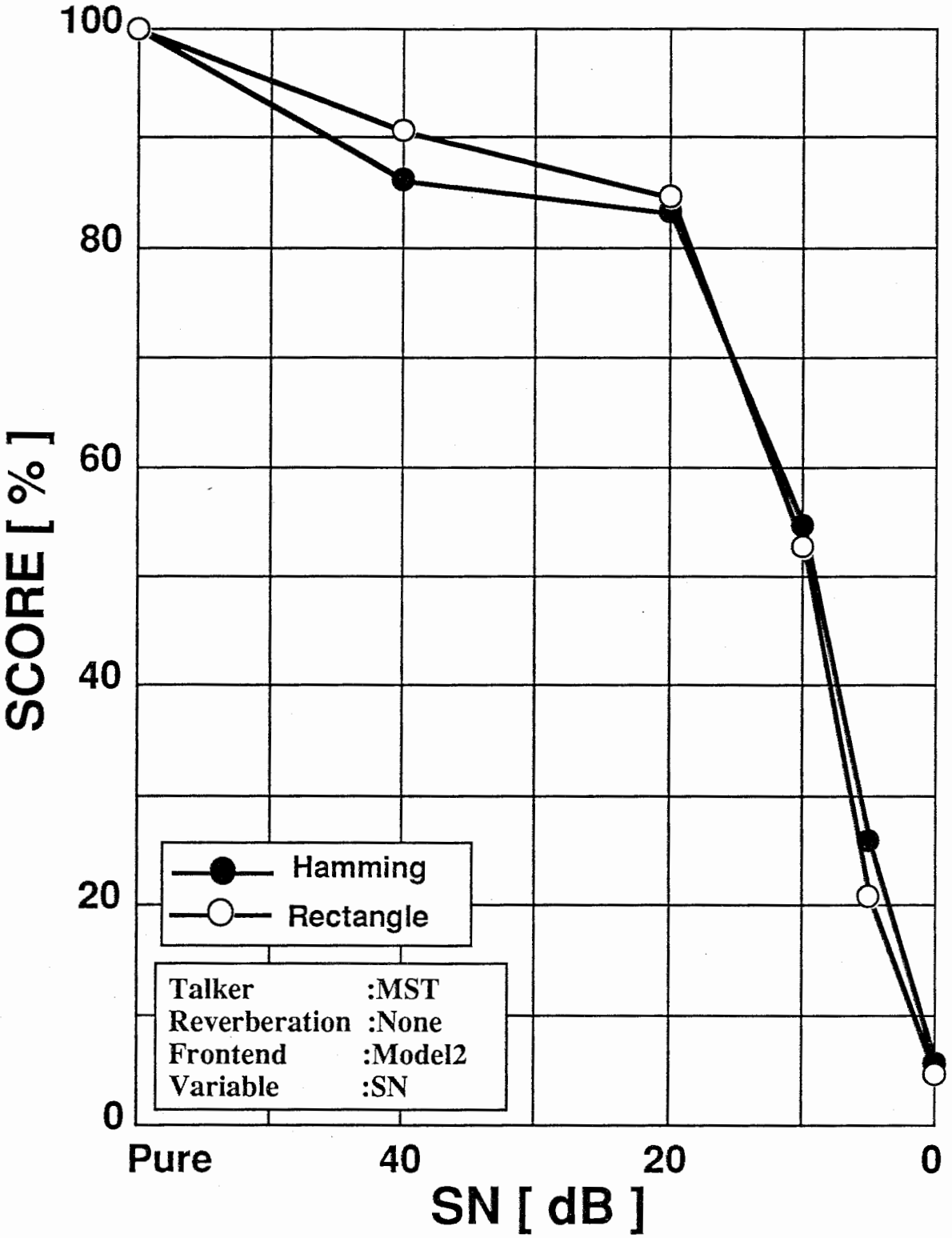
第２－６図　残響インパルスデータ



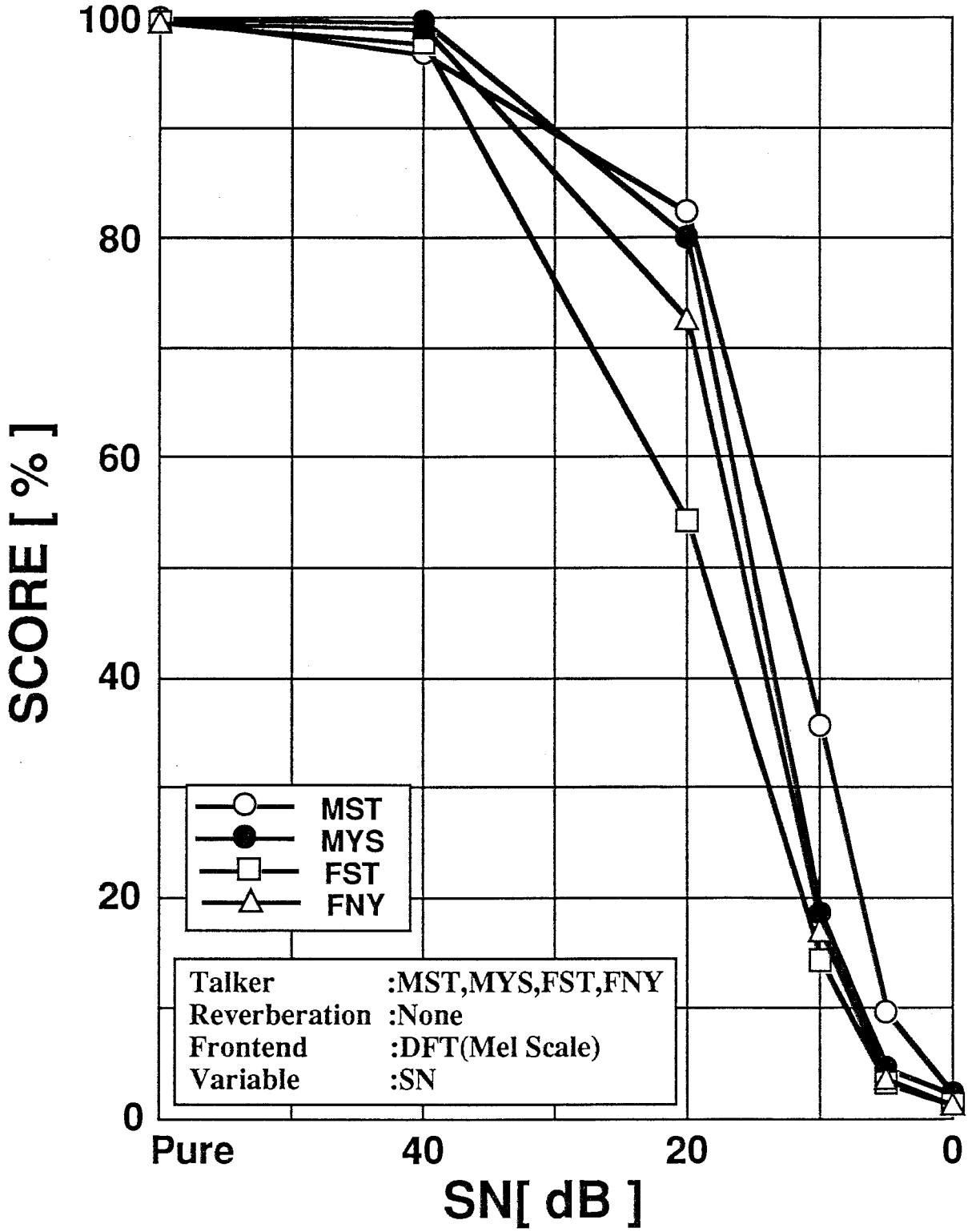
第２－７図　残響特性の畳み込み方法


Voice Input2
Impulse Response
Convolution
Reverberated Voice
Voice Input1
Noise Addition
Noisy Voice
Feature Extraction
Feature Extraction
Templates (216)
DTW
216 Words
Words Recognition Score

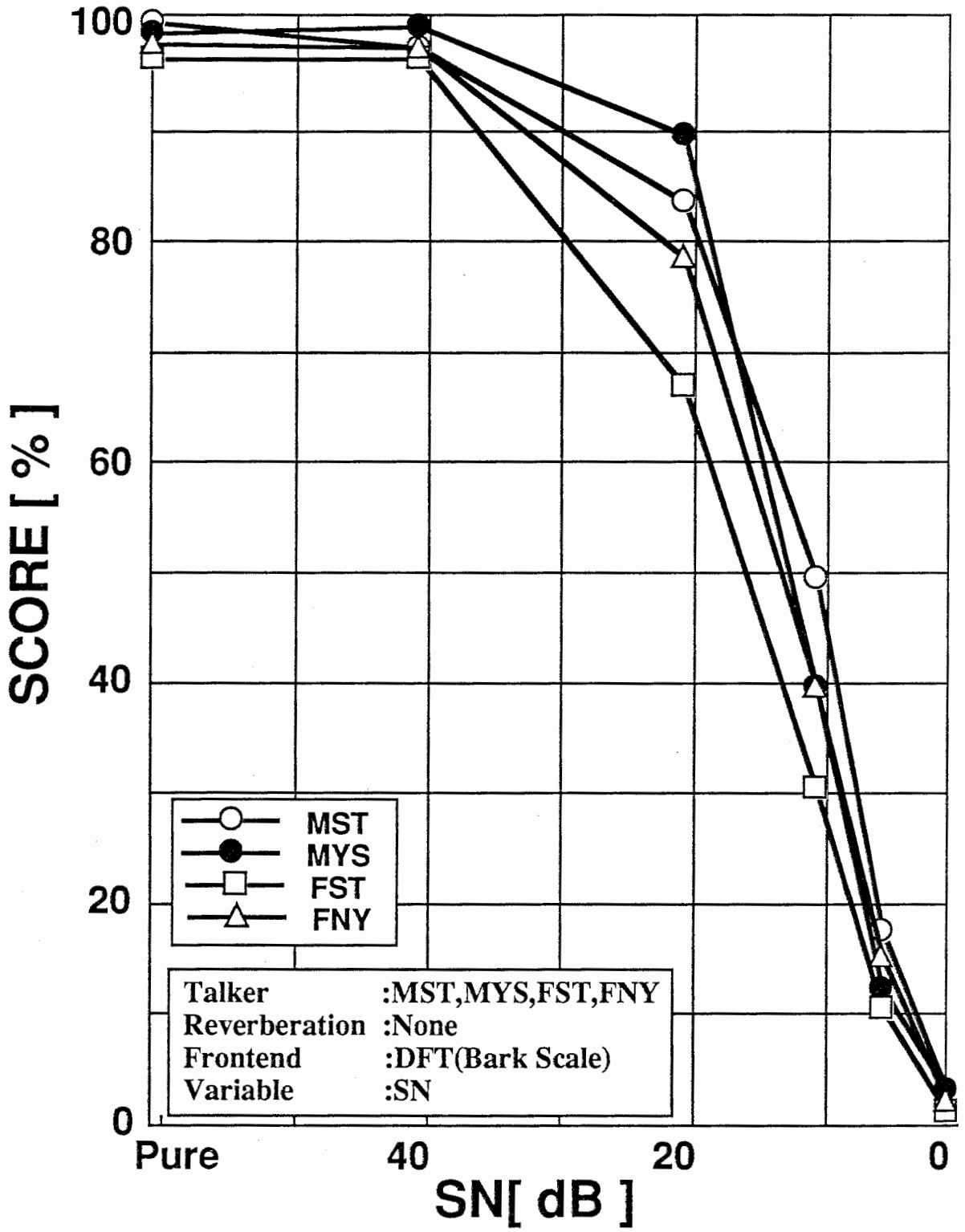
第３－１図　実験構成図



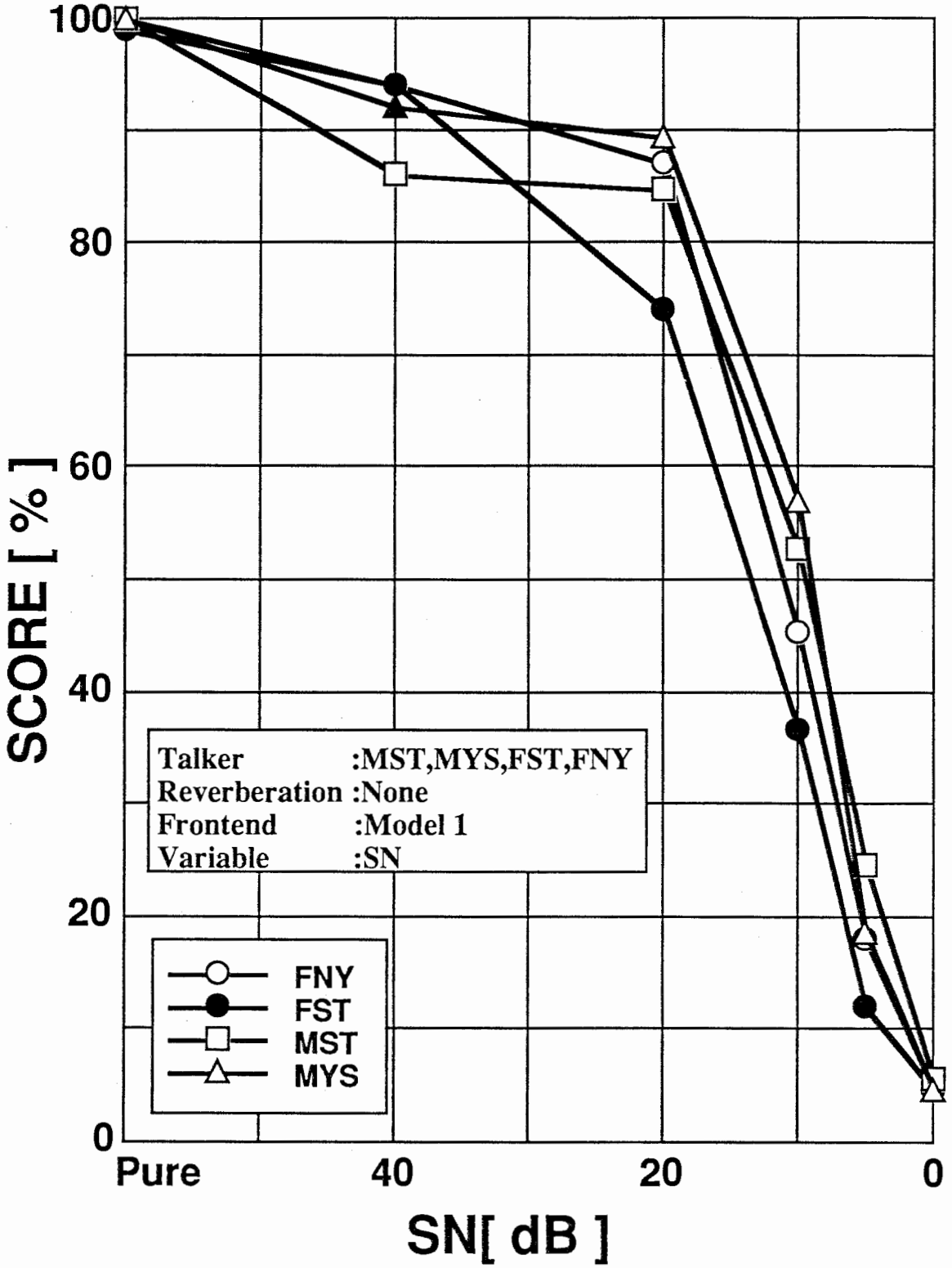
第３－２図　モデル２でのウインドウの認識率への影響



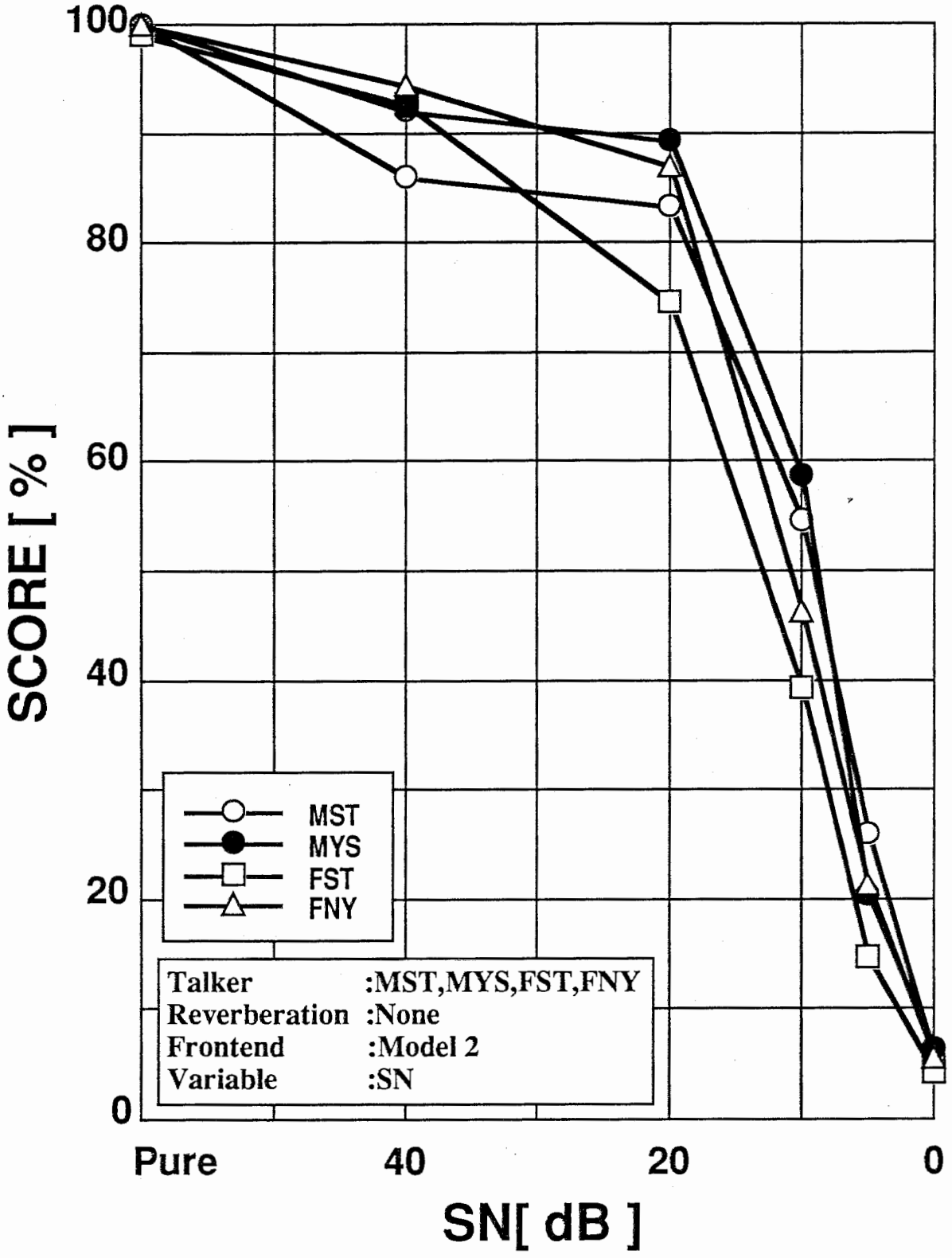
第３－３－１図　ＤＦＴ(Mel) 特徴量による単語認識



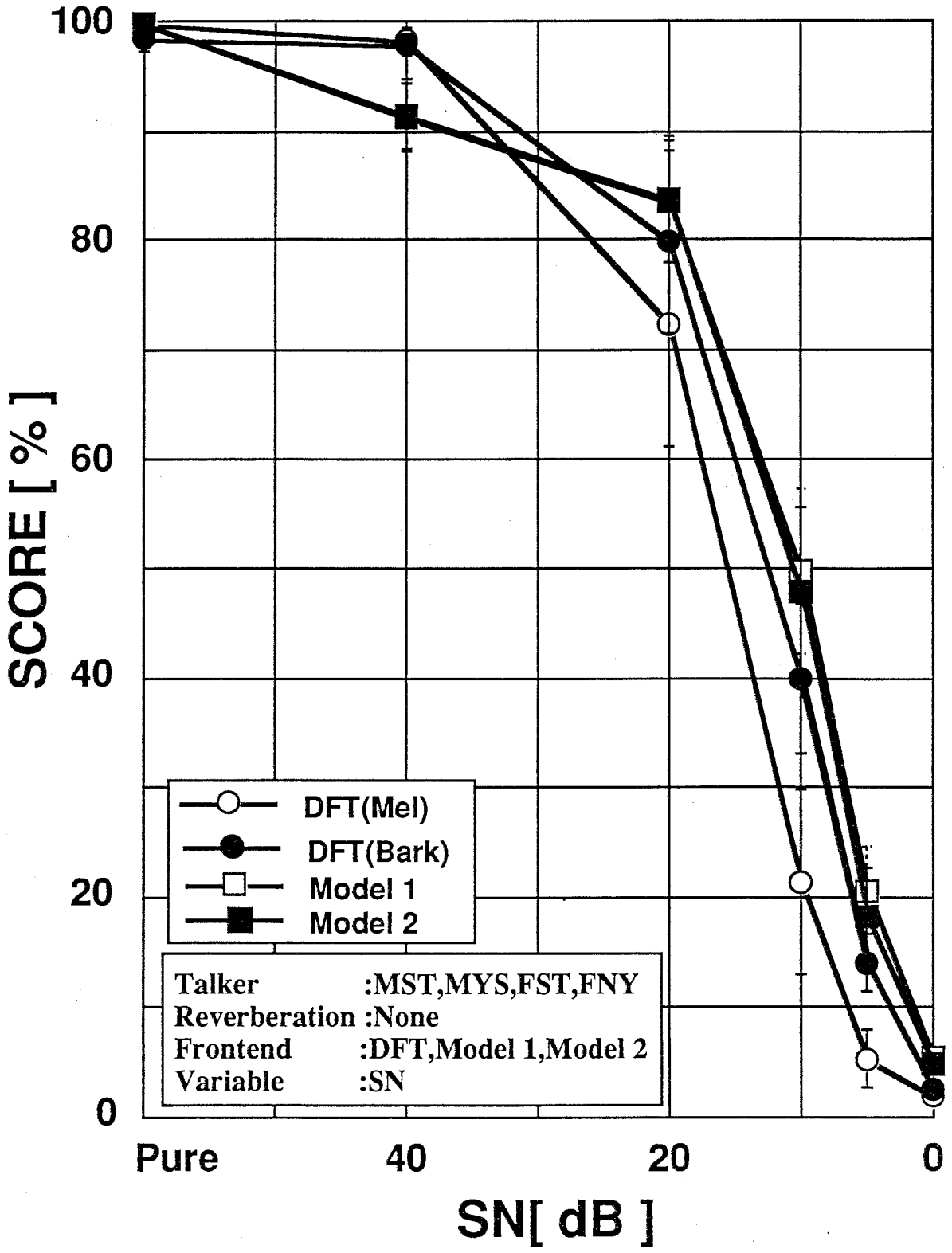
第３－３－２図　ＤＦＴ(Bark) 特徴量による単語認識



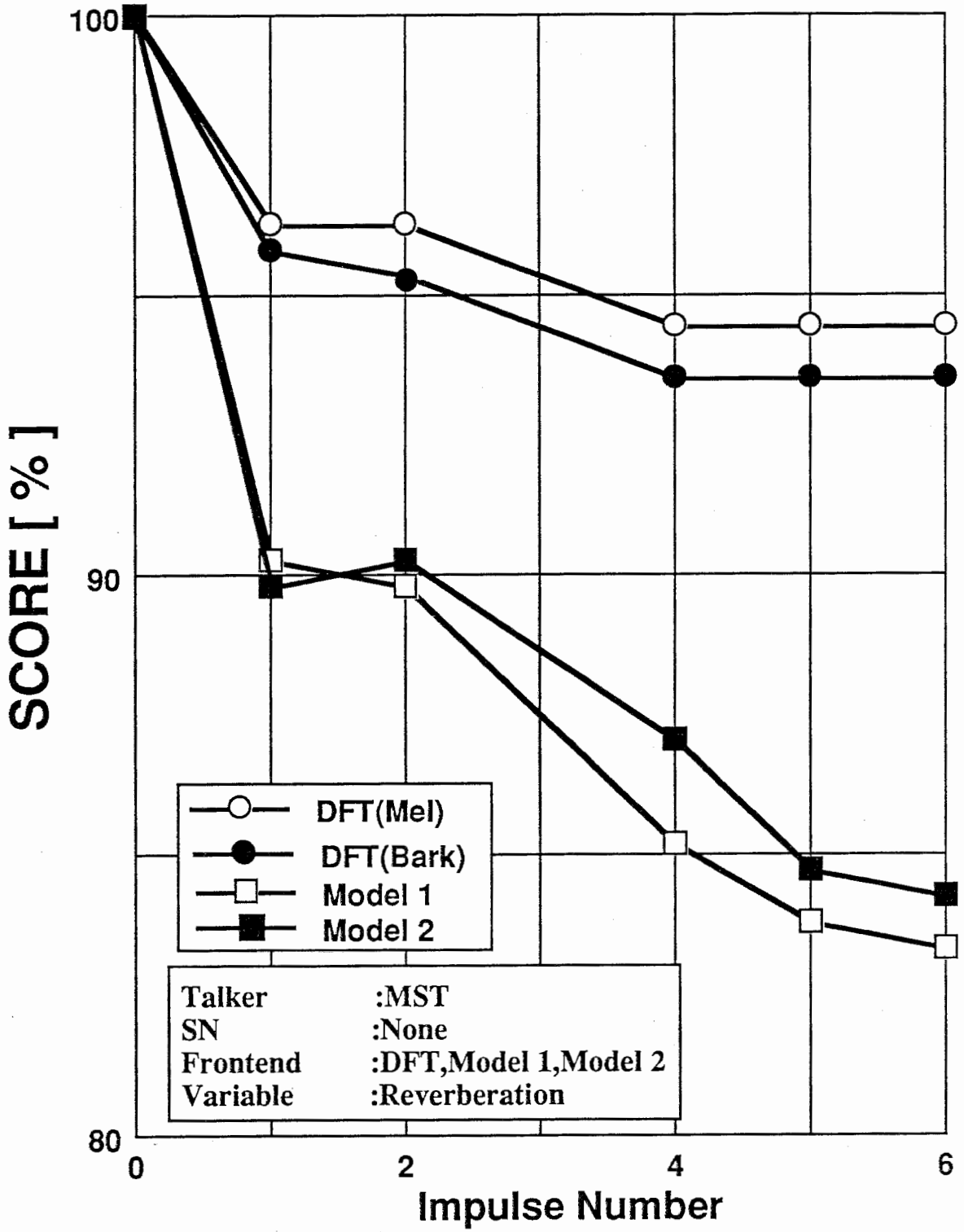
第３－４図　モデル１特徴量による単語認識



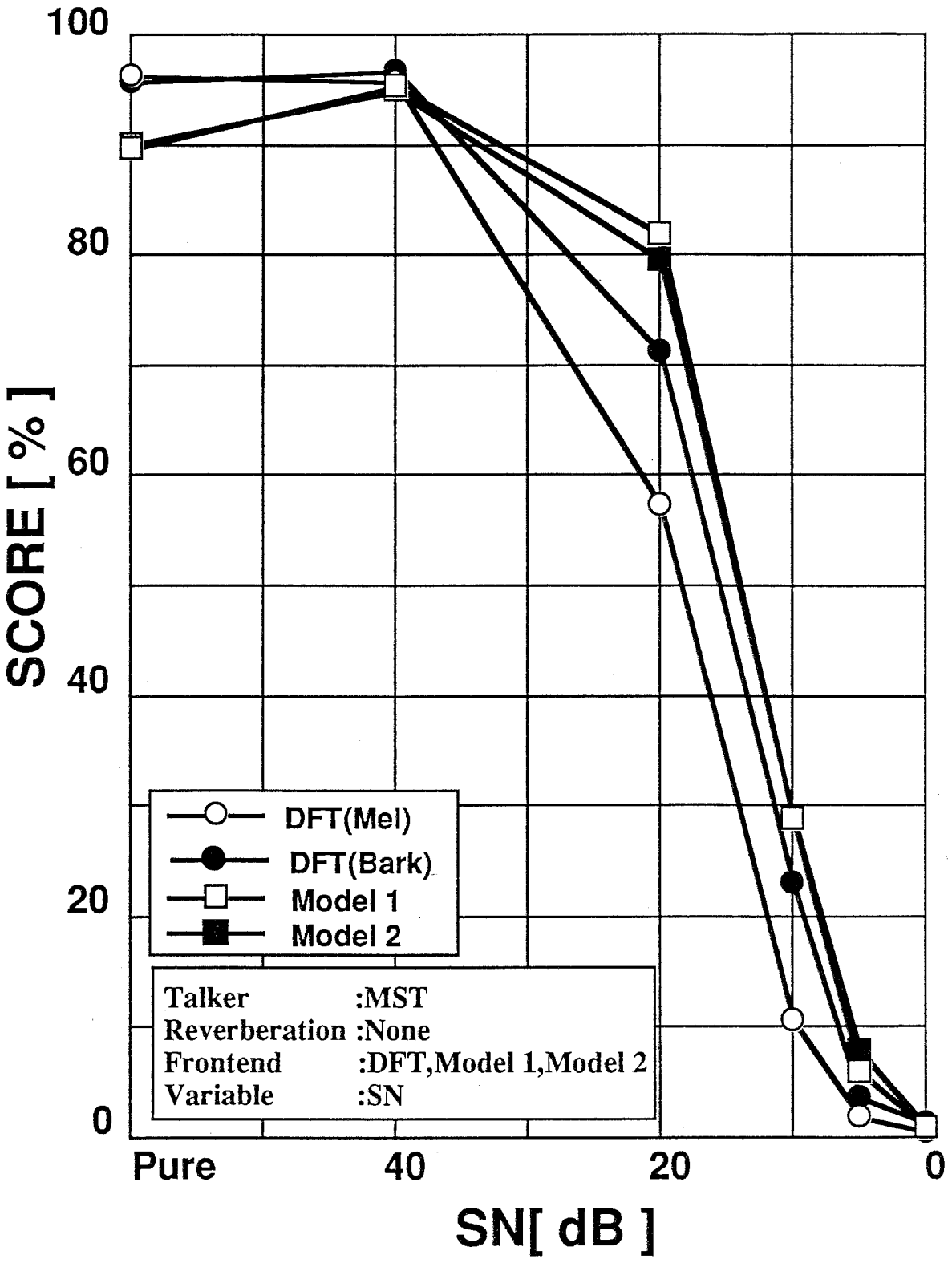
第３－５図　モデル２特徴量のよる単語認識



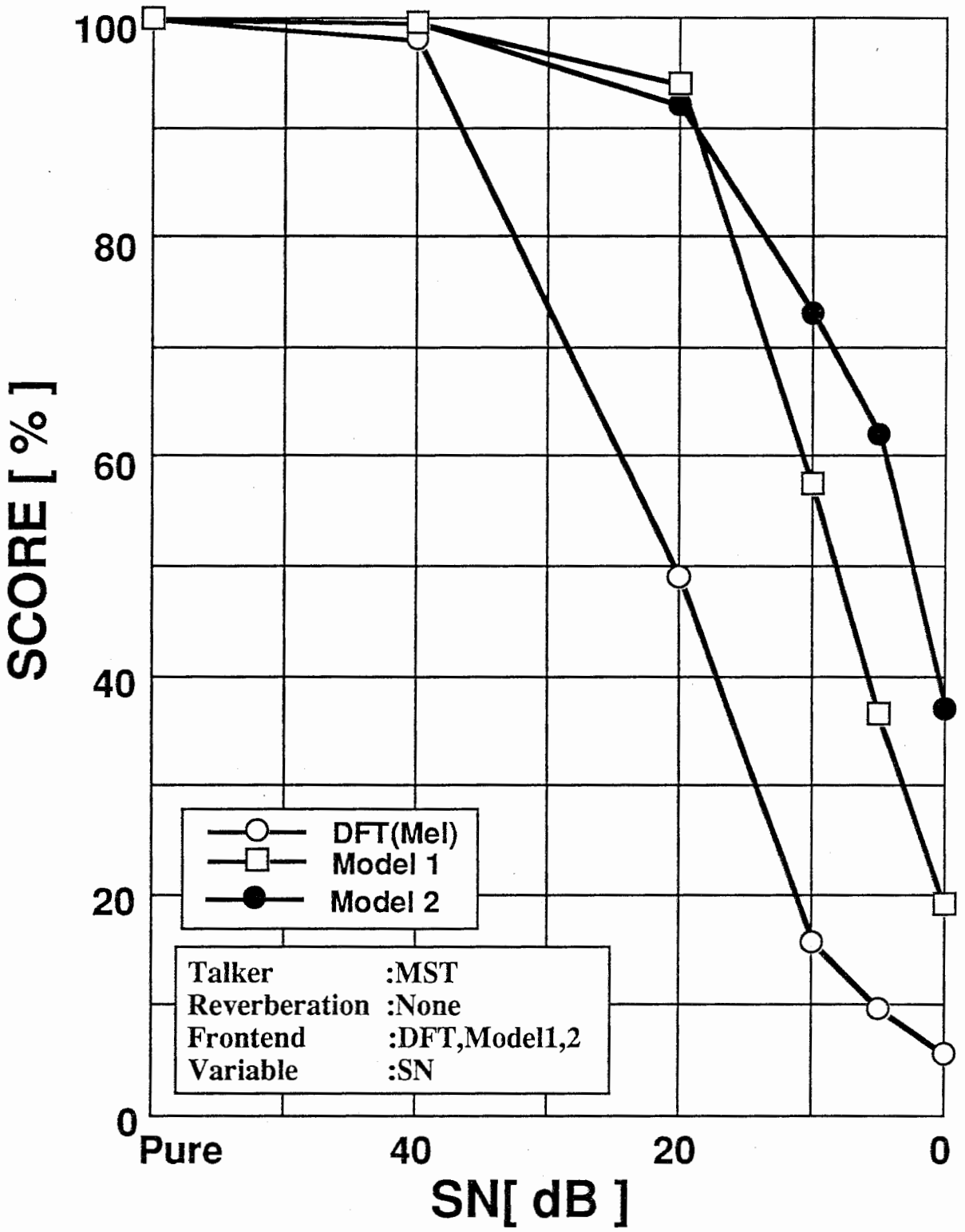
第３－6図　雑音下での異なるフロントエンド特徴量による単語認識(4話者平均)



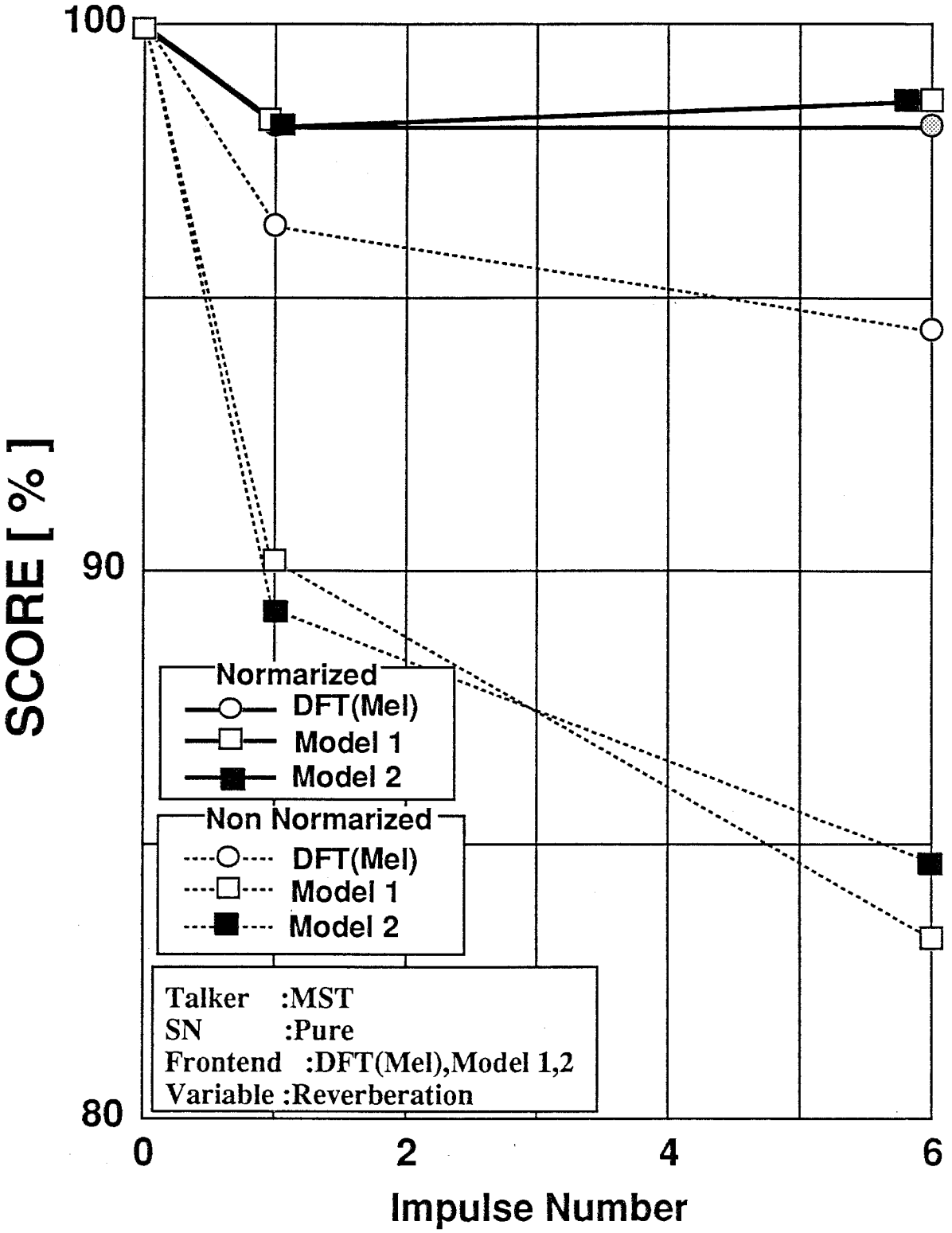
第３－7図　残響下での異なるフロントエンド特徴量による単語認識(4話者平均)



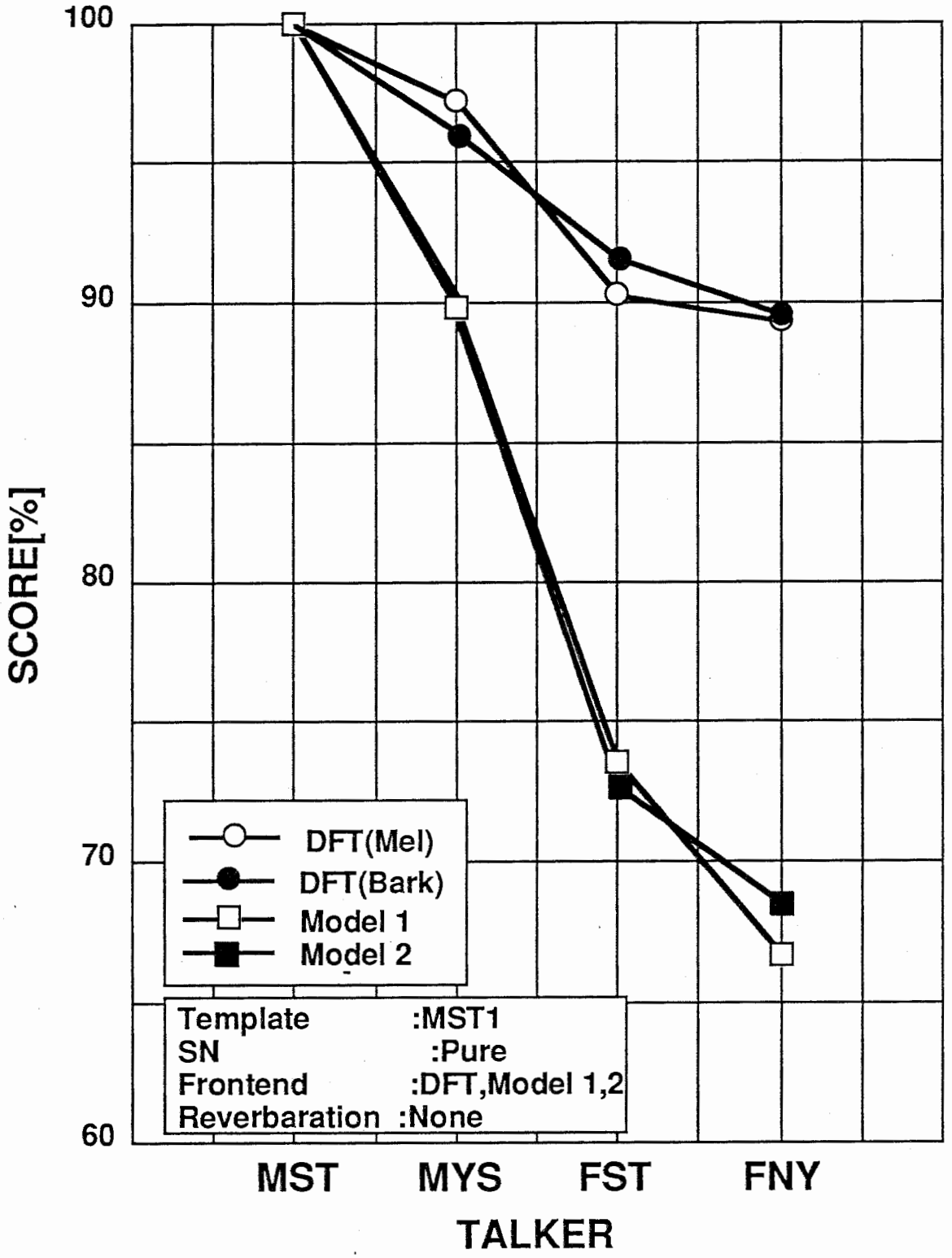
第３－8図　残響下,雑音下での異なるフロントエンド特徴量による単語認識(4話者平均)



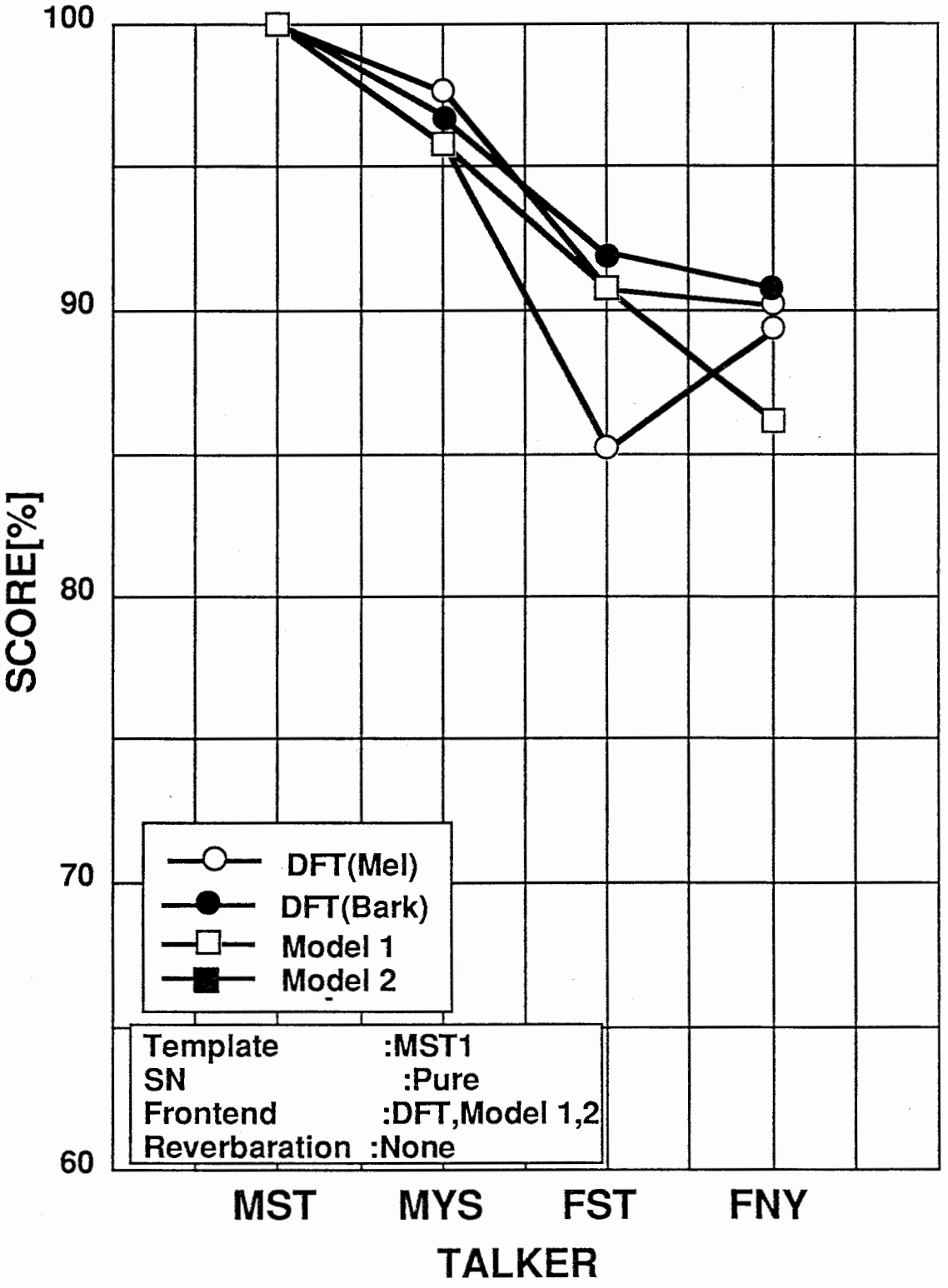
第３－9図　パワー正規化特徴量による単語認識



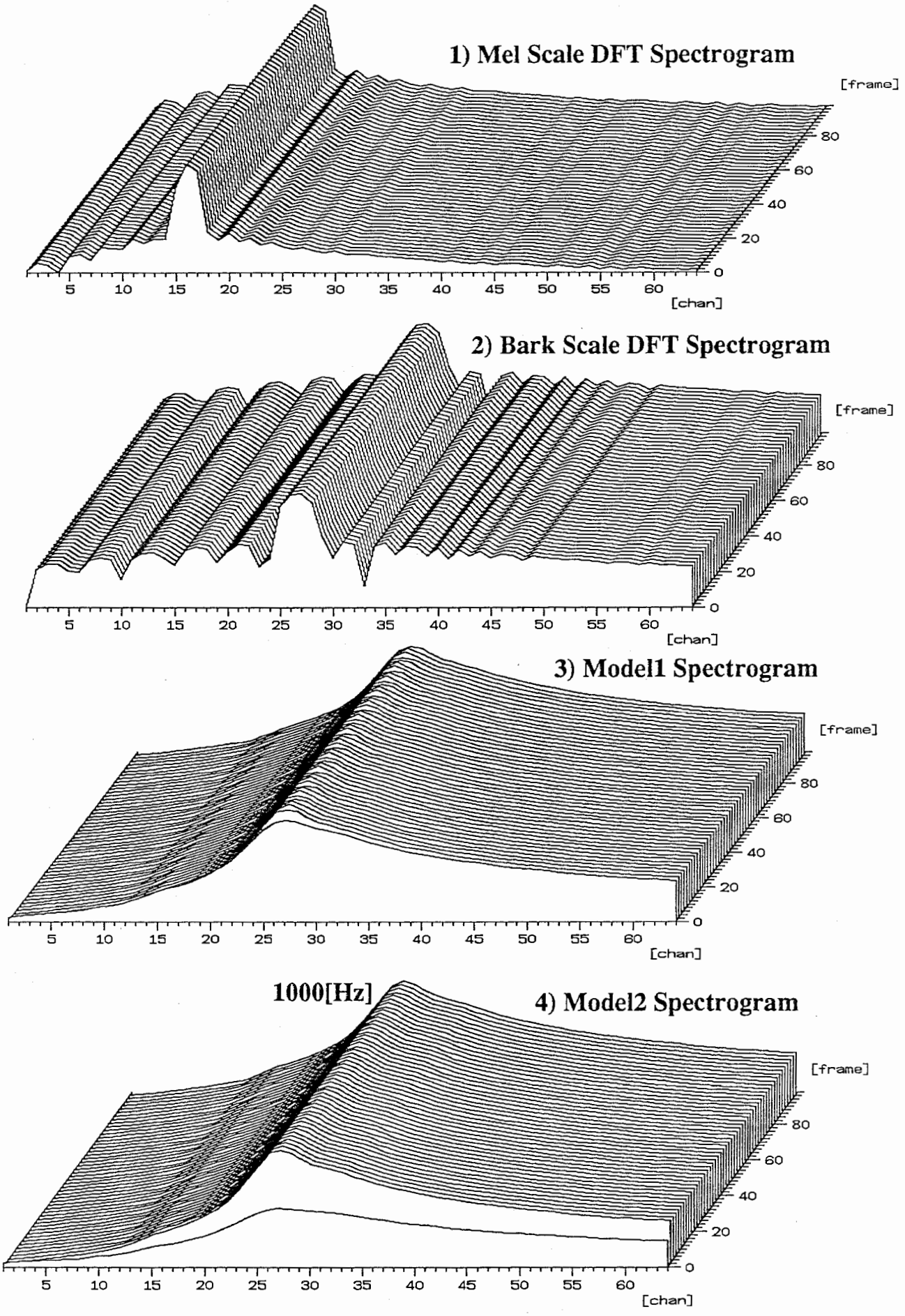
第３－１０図　残響下での単語認識(正規化)



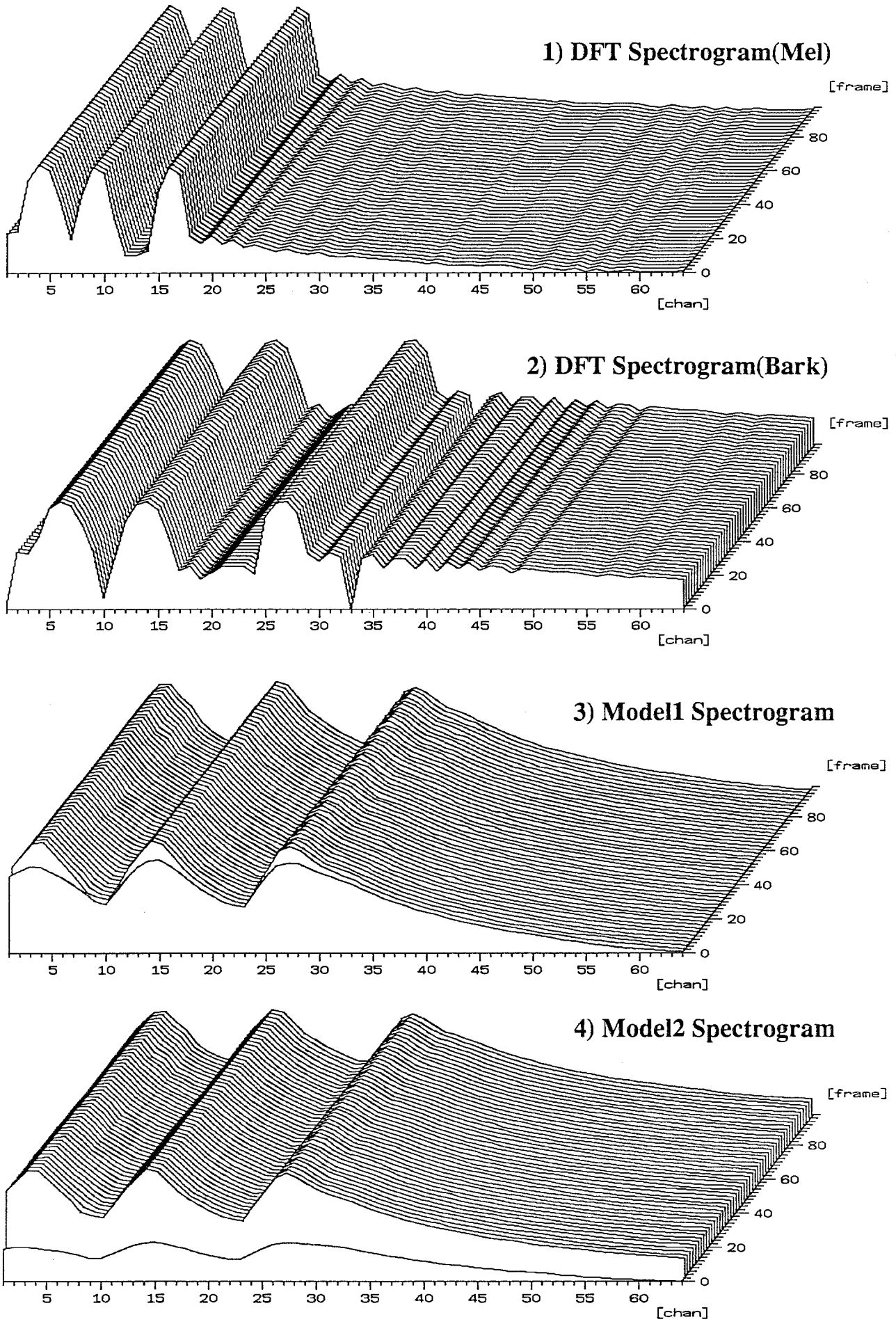
第３－１１図　異なる話者間の単語認識



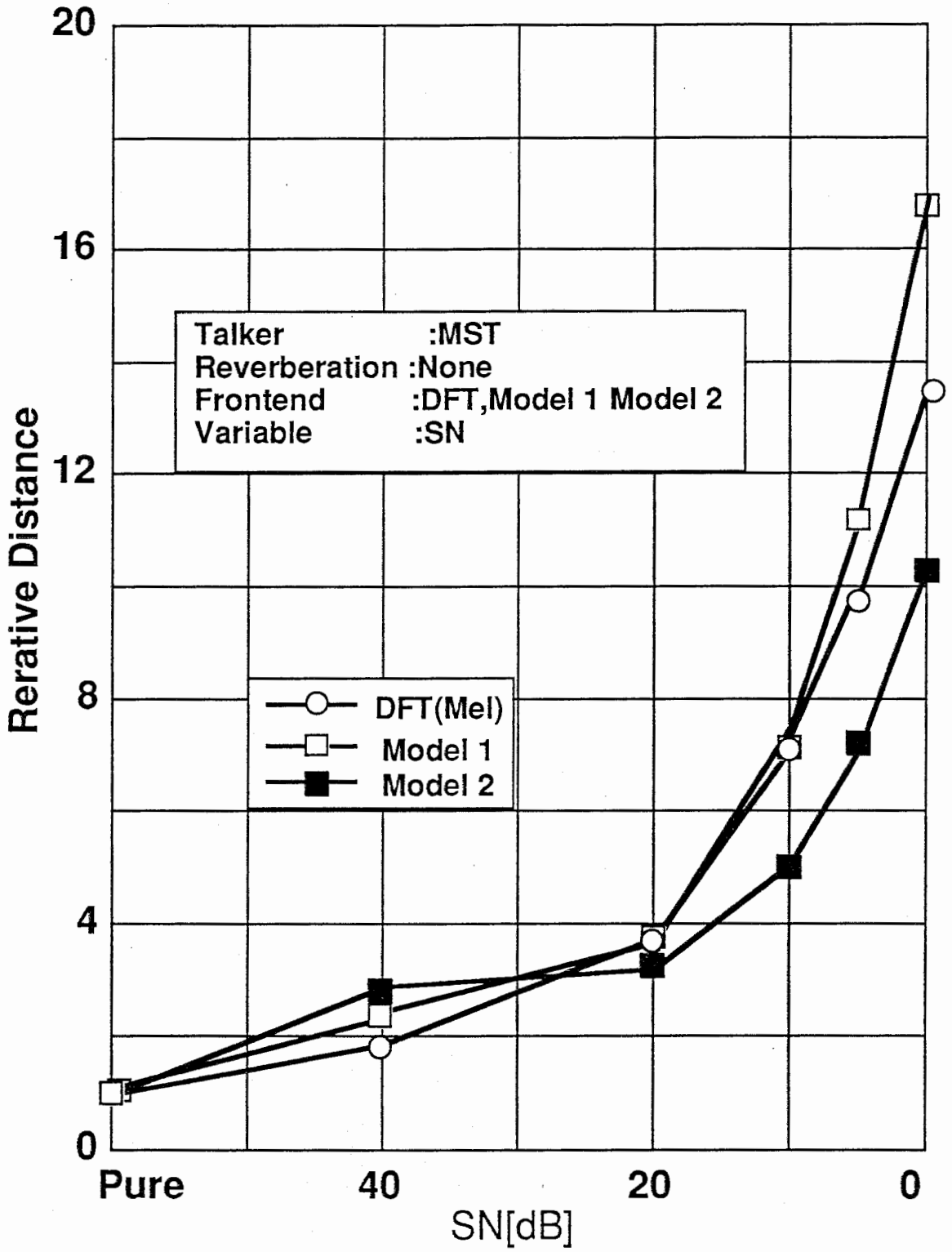
第３－１２図　異なる話者間の単語認識（特徴量正規化）



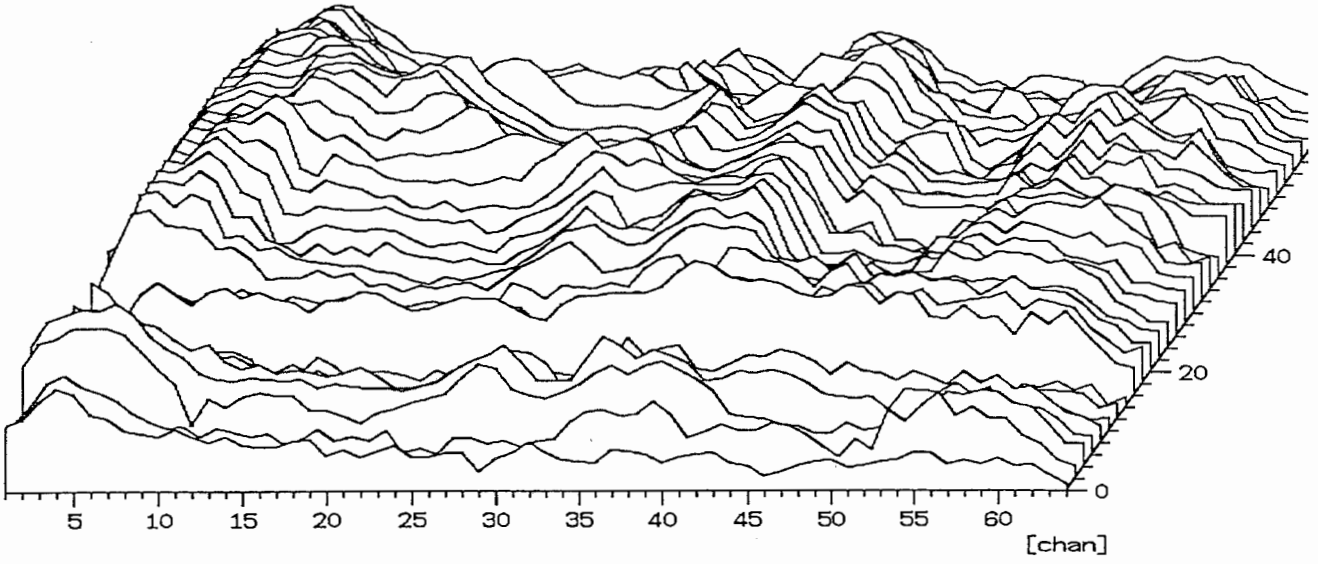
第４－１図　ＤＦＴ、モデル１，モデル２スペクトログラム



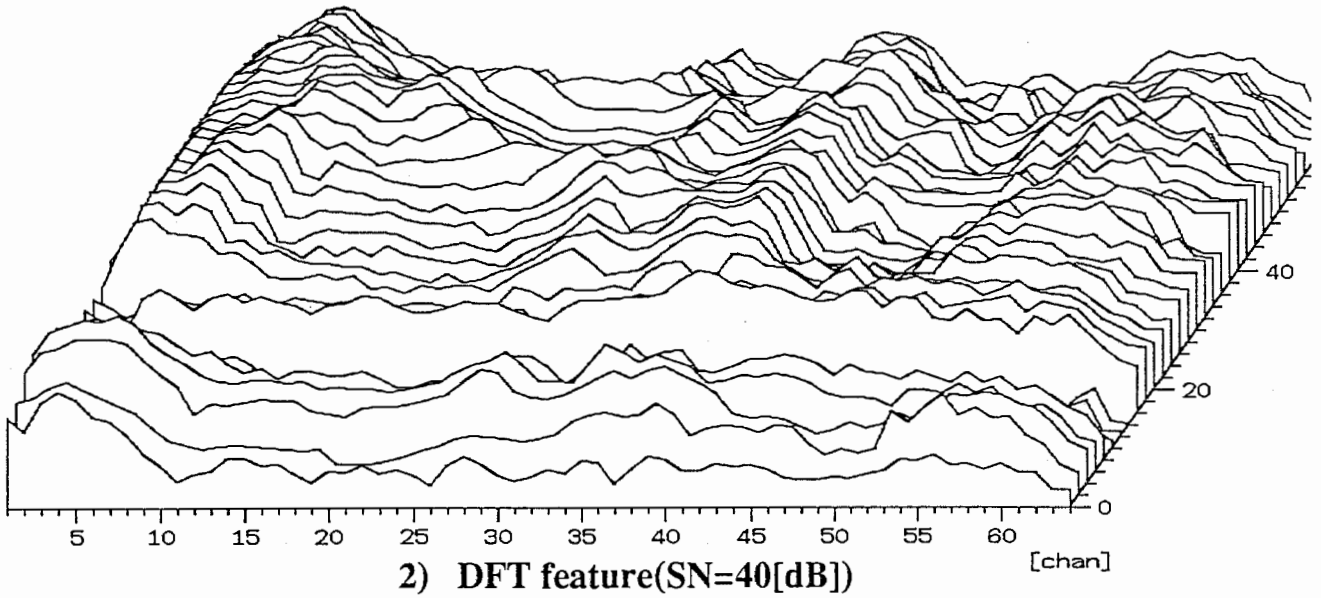
第４－２図　２００＋５００＋１０００［Ｈｚ］のスペクトログラム



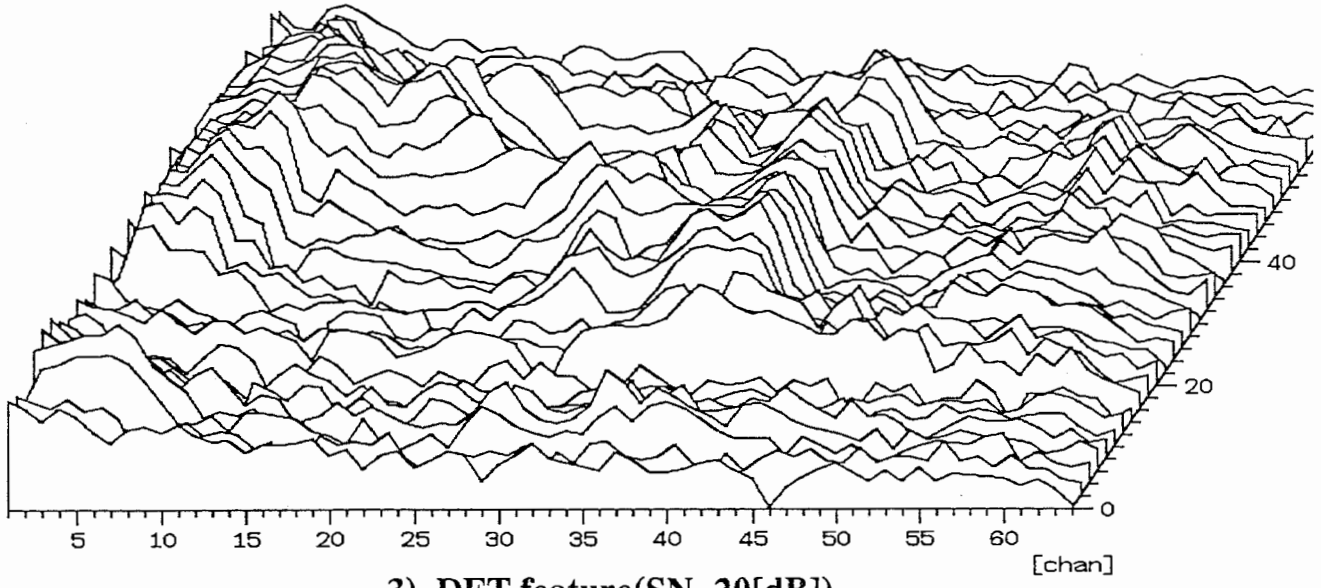
第４－３図　ノイズによるテンプレート単語間距離変動



1) DFT feature(Pure)

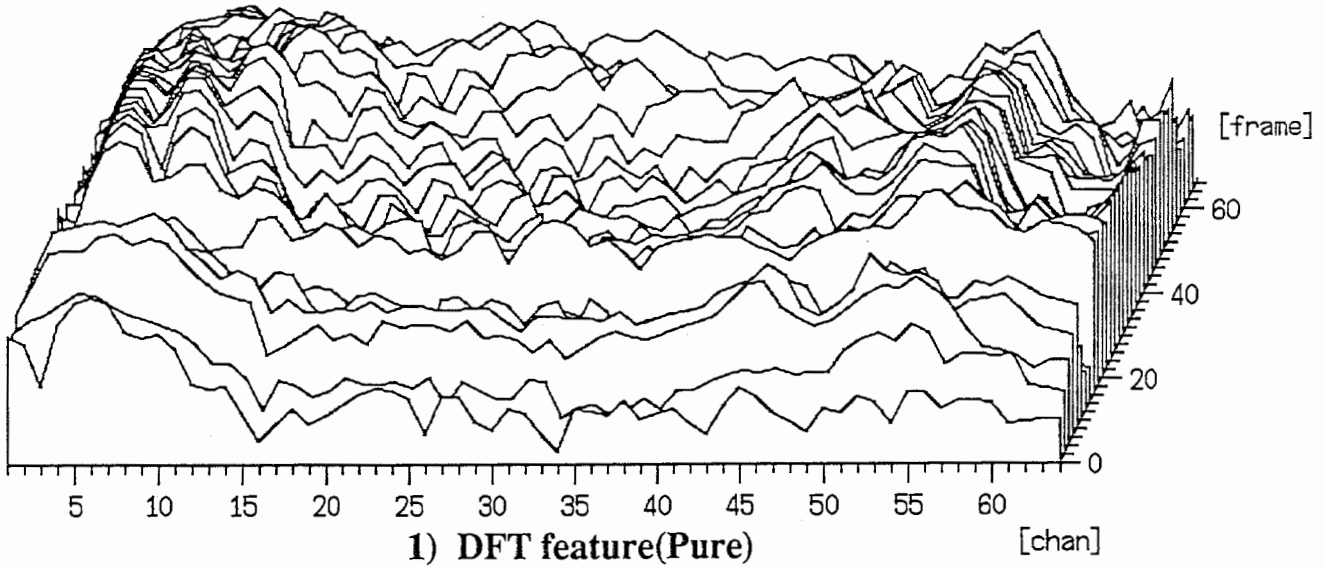


2) DFT feature(SN=40[dB])

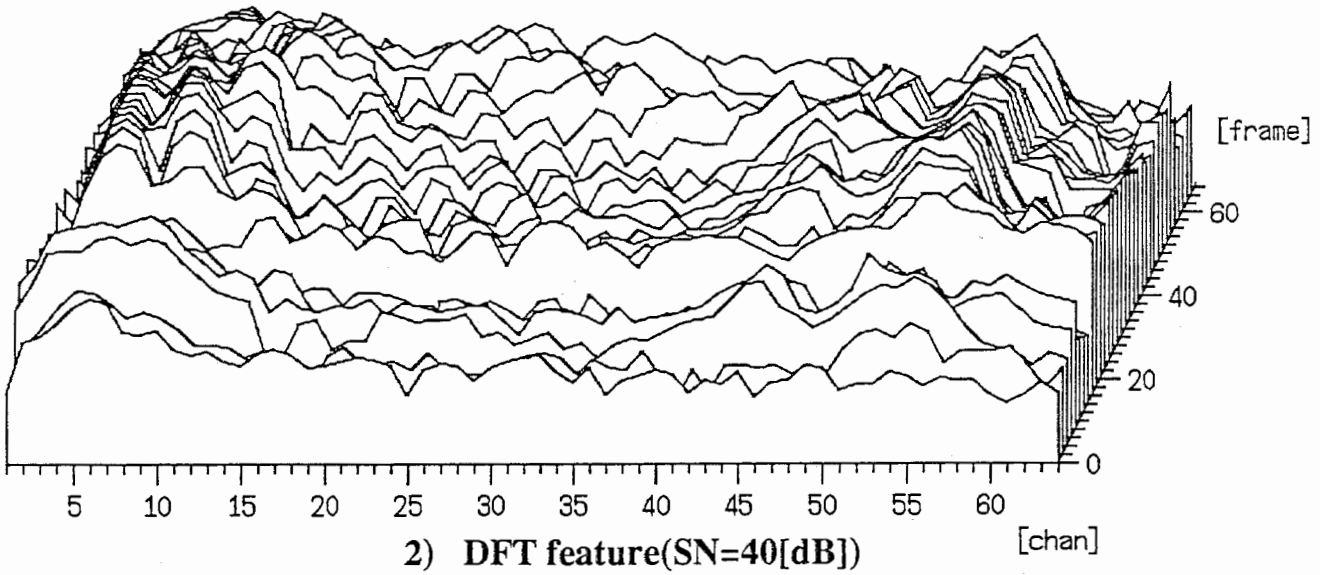


3) DFT feature(SN=20[dB])

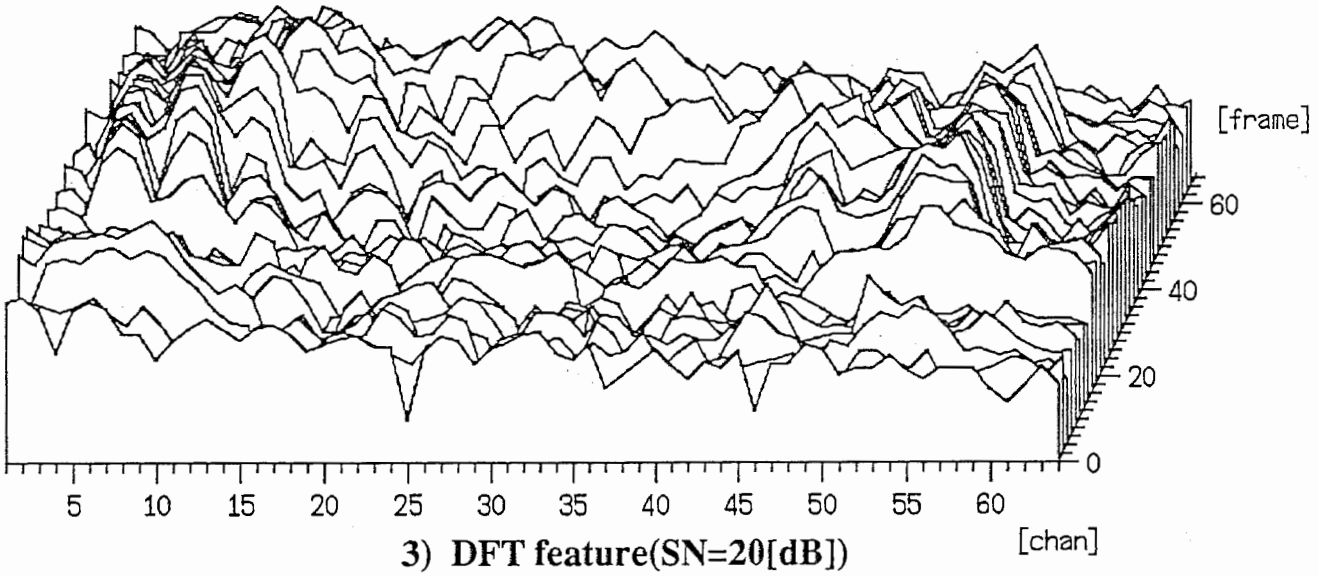
第４－４－1図　ＳＮ変動によるスペクトルの変化（Mel ＤＦＴ）



1) DFT feature(Pure)

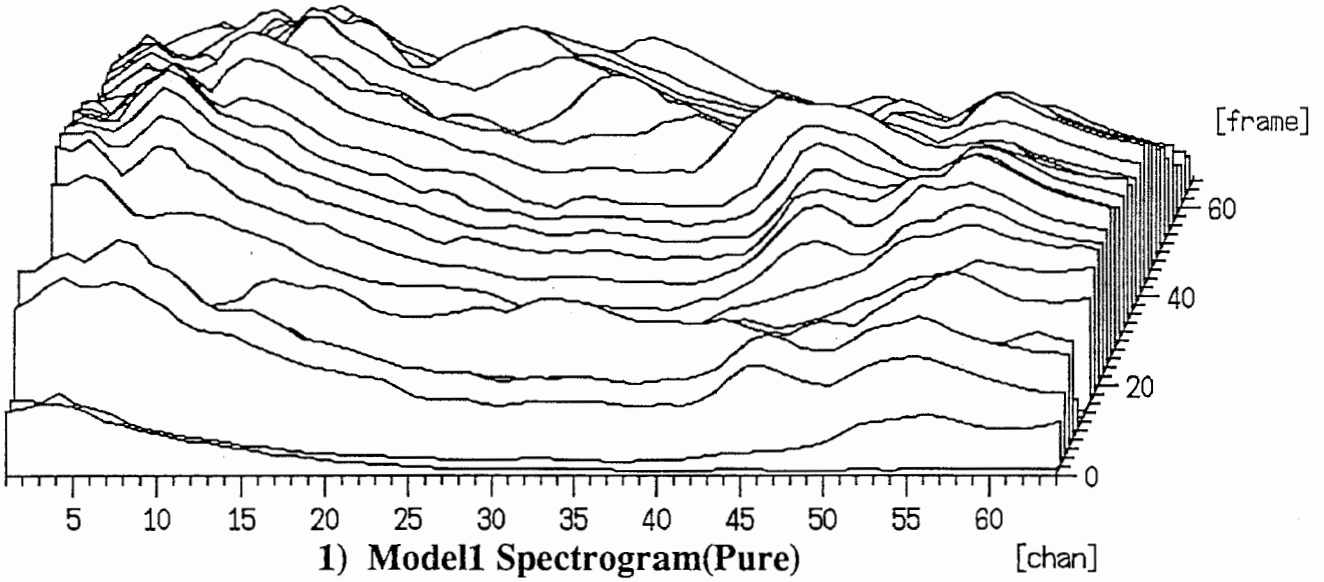


2) DFT feature(SN=40[dB])

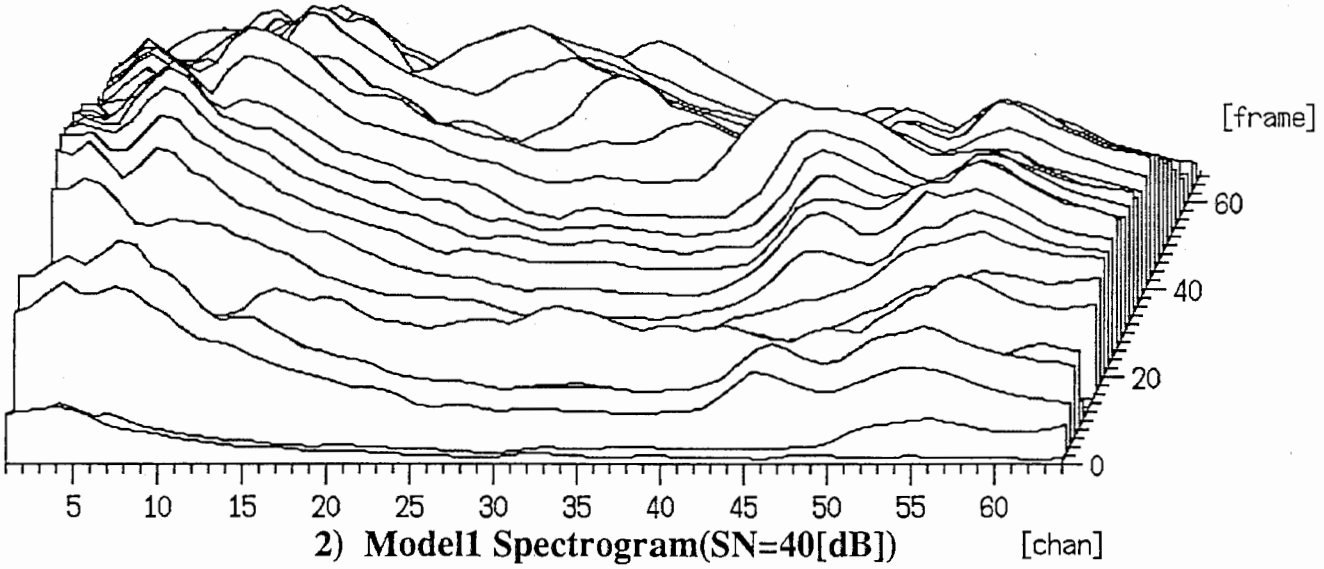


3) DFT feature(SN=20[dB])

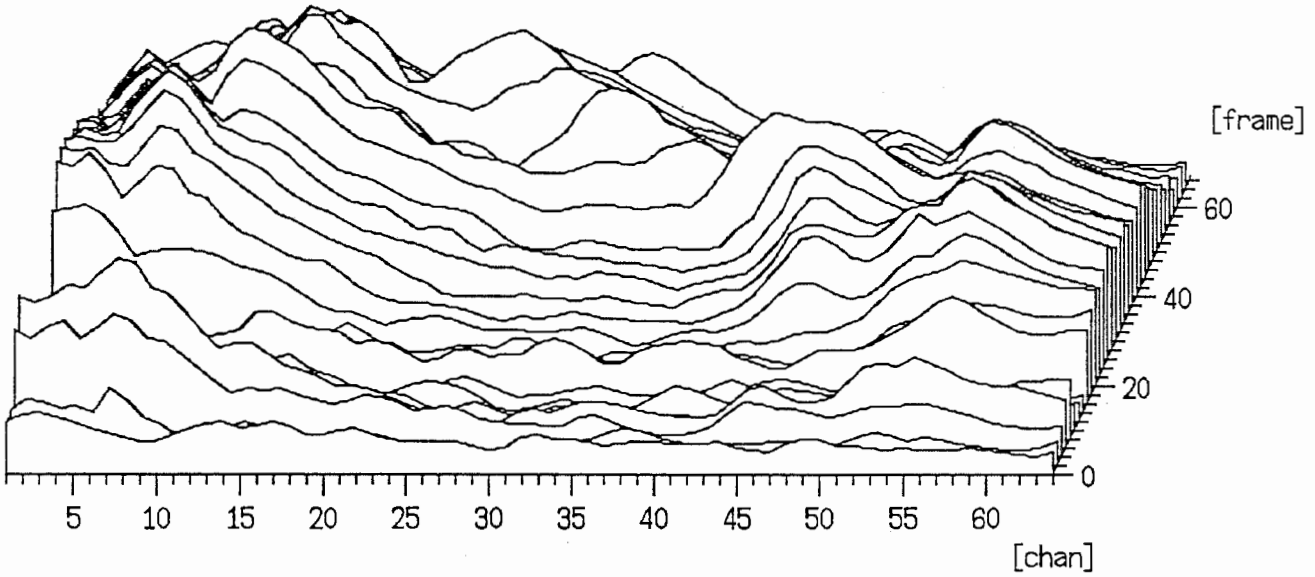
第４－４－2図　ＳＮ変動によるスペクトルの変化（Bark ＤＦＴ）



1) Model1 Spectrogram(Pure)

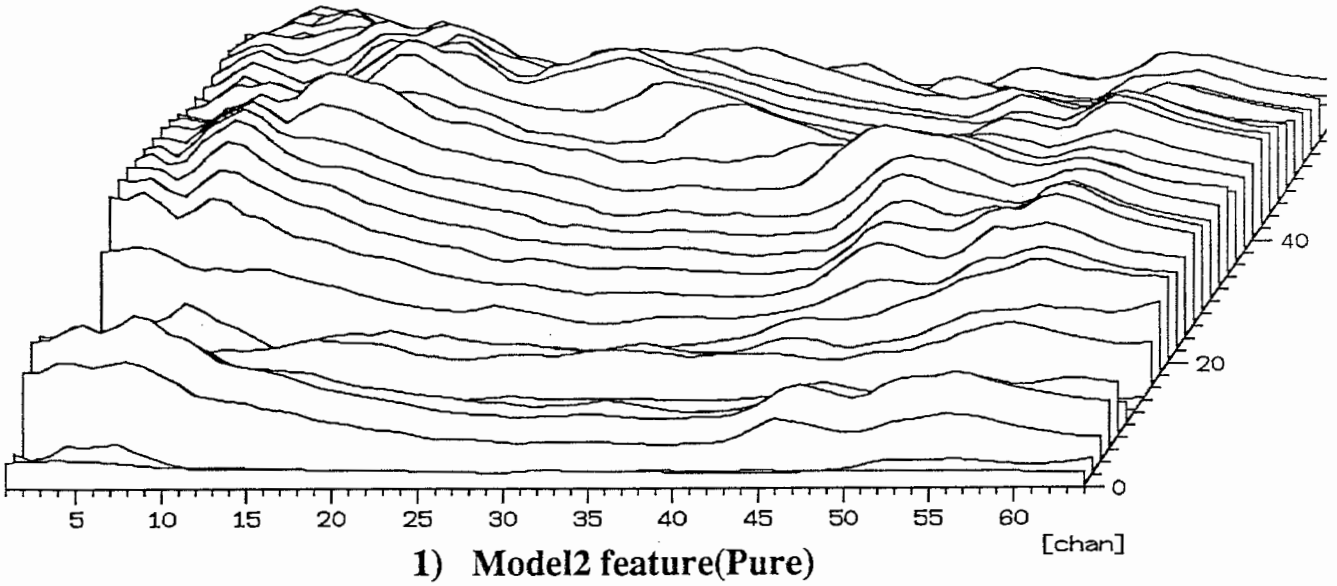


2) Model1 Spectrogram(SN=40[dB])

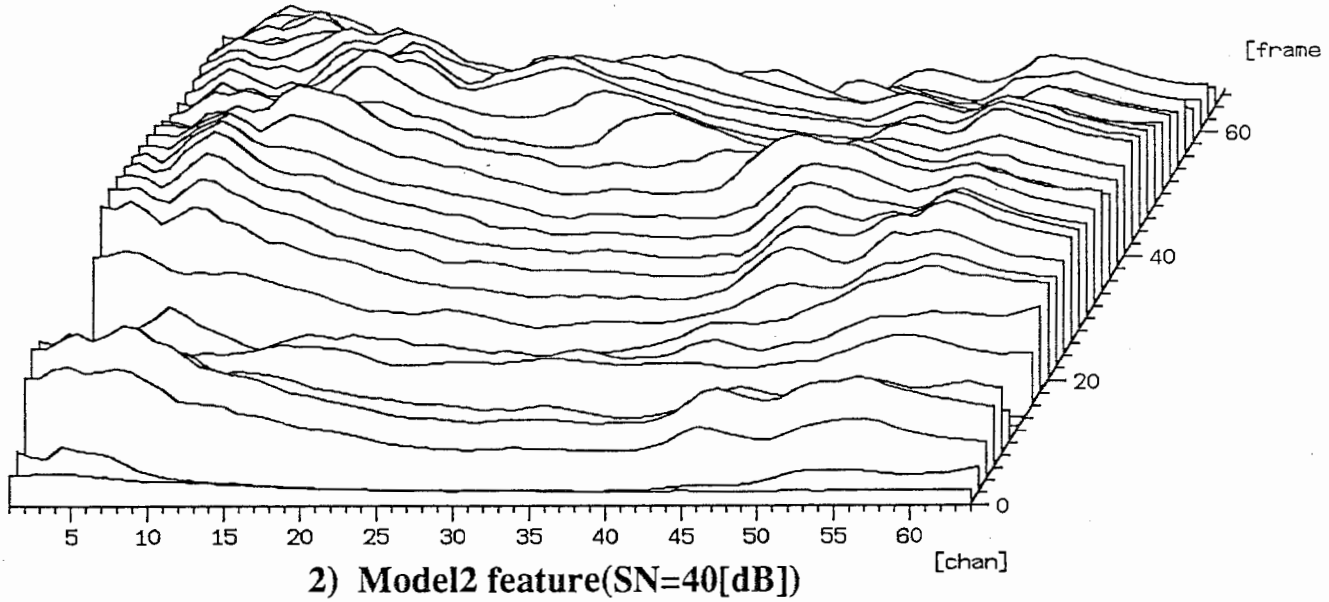


3) Model1 Spectrogram(SN=20[dB])

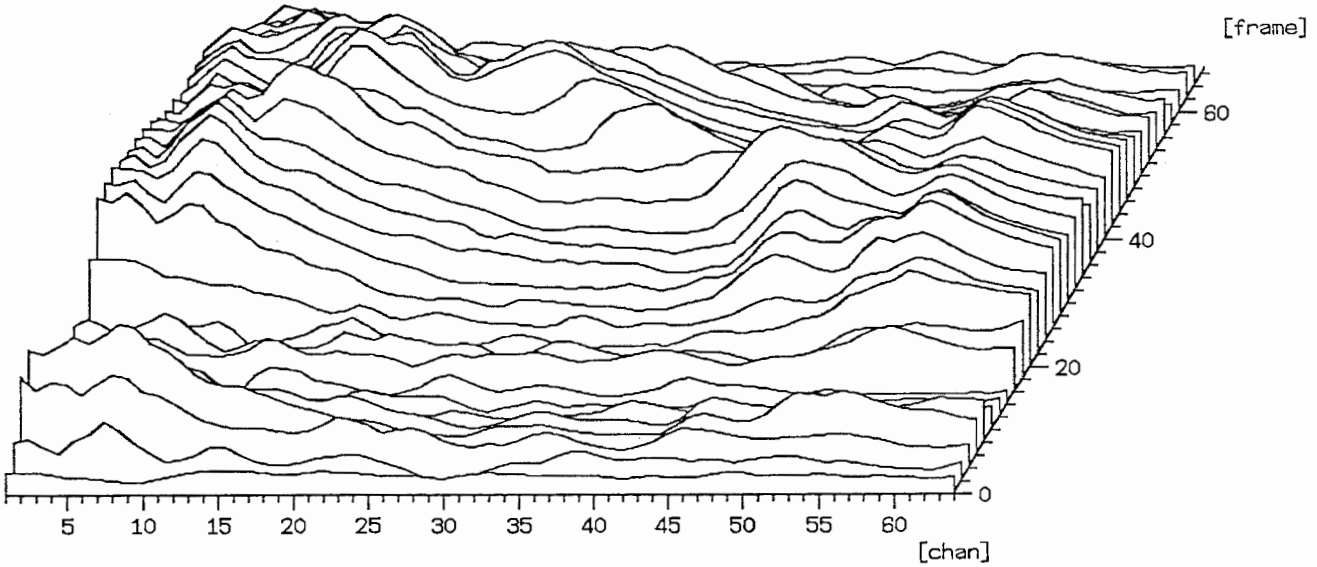
第４－５図　ＳＮ変動によるスペクトルの変化（モデル１）



1) Model2 feature(Pure)

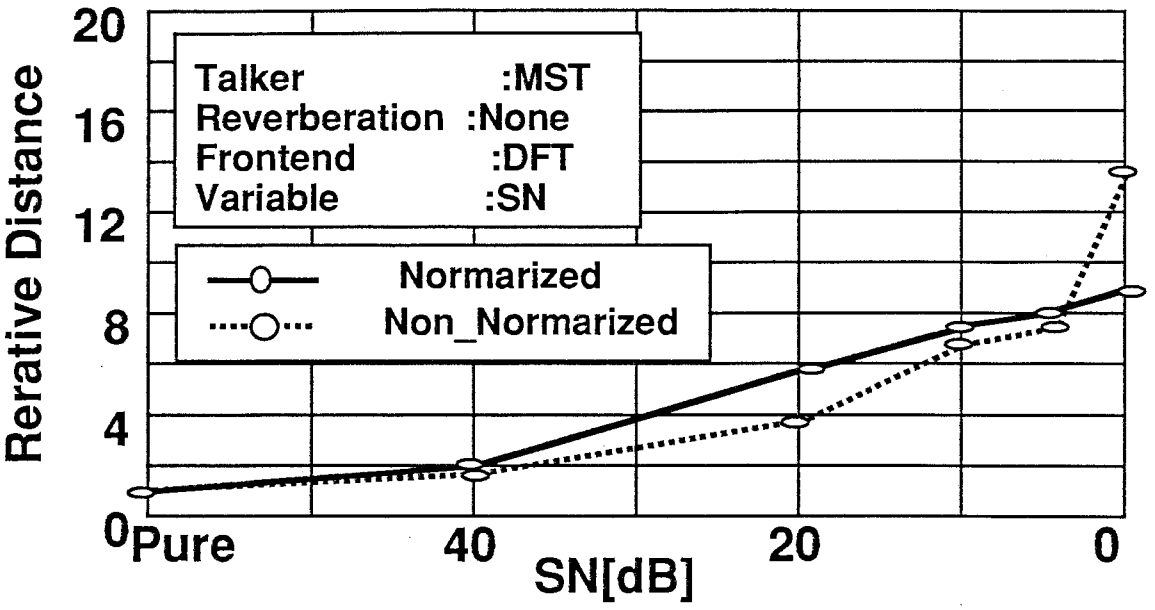


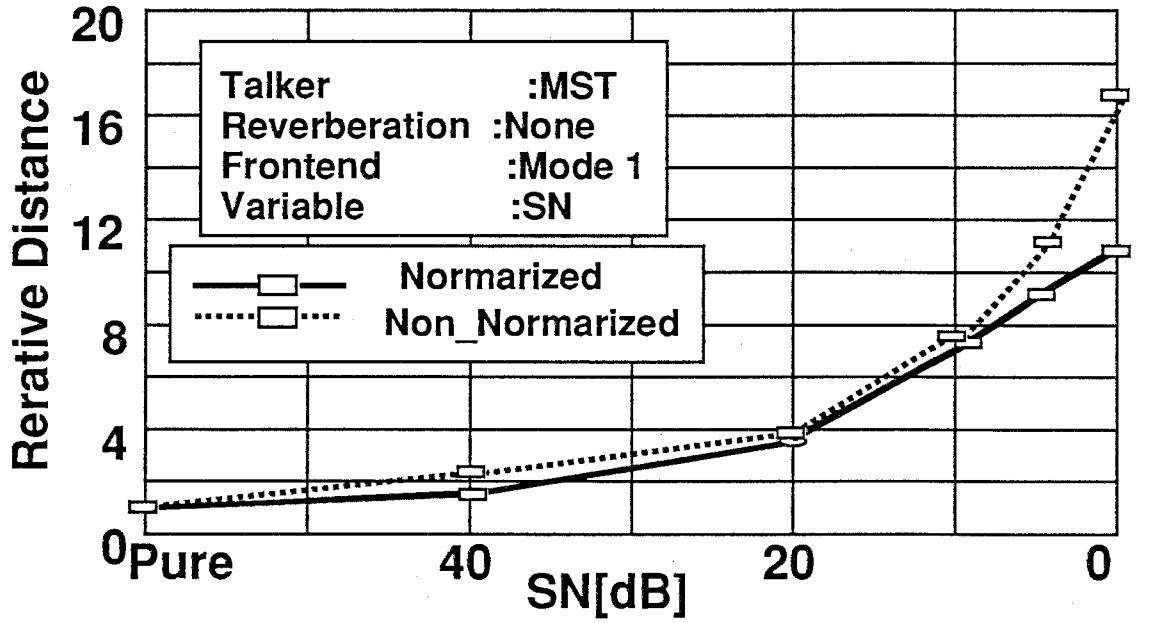
2) Model2 feature(SN=40[dB])

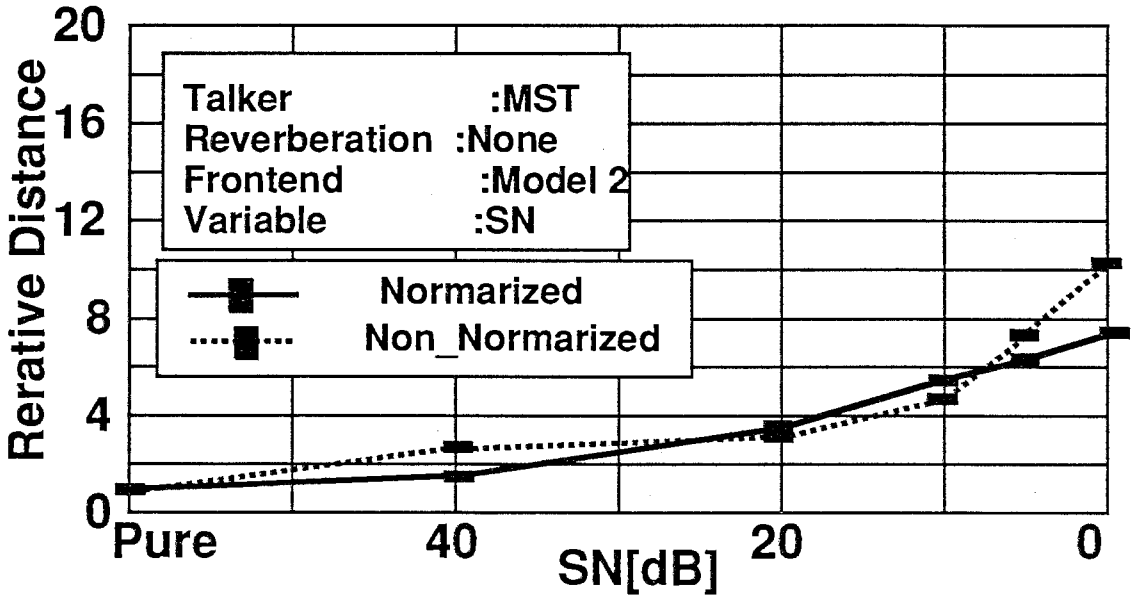


3) Model2 feature(SN=20[dB])

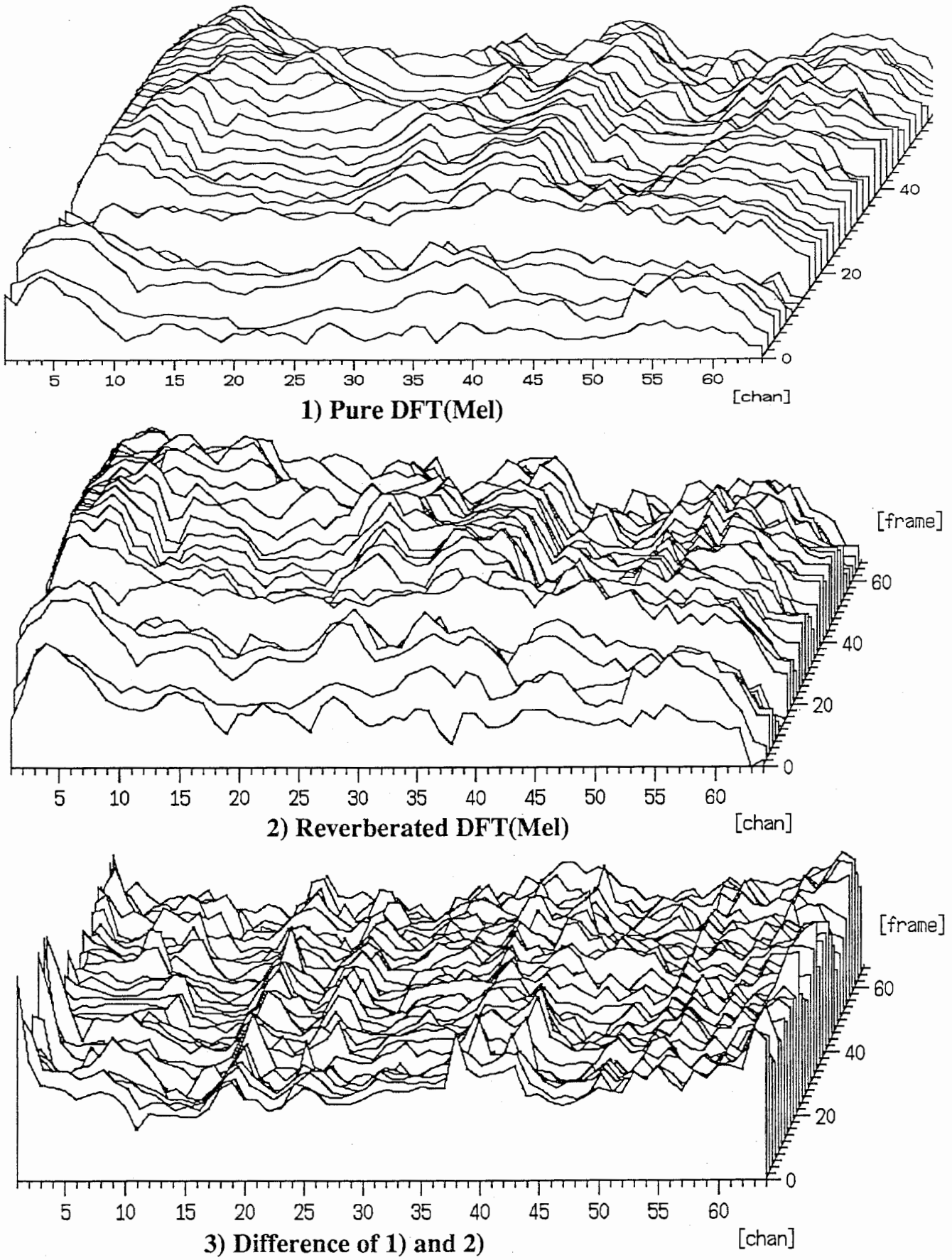
第４－６図　ＳＮ変動によるスペクトルの変化（モデル２）



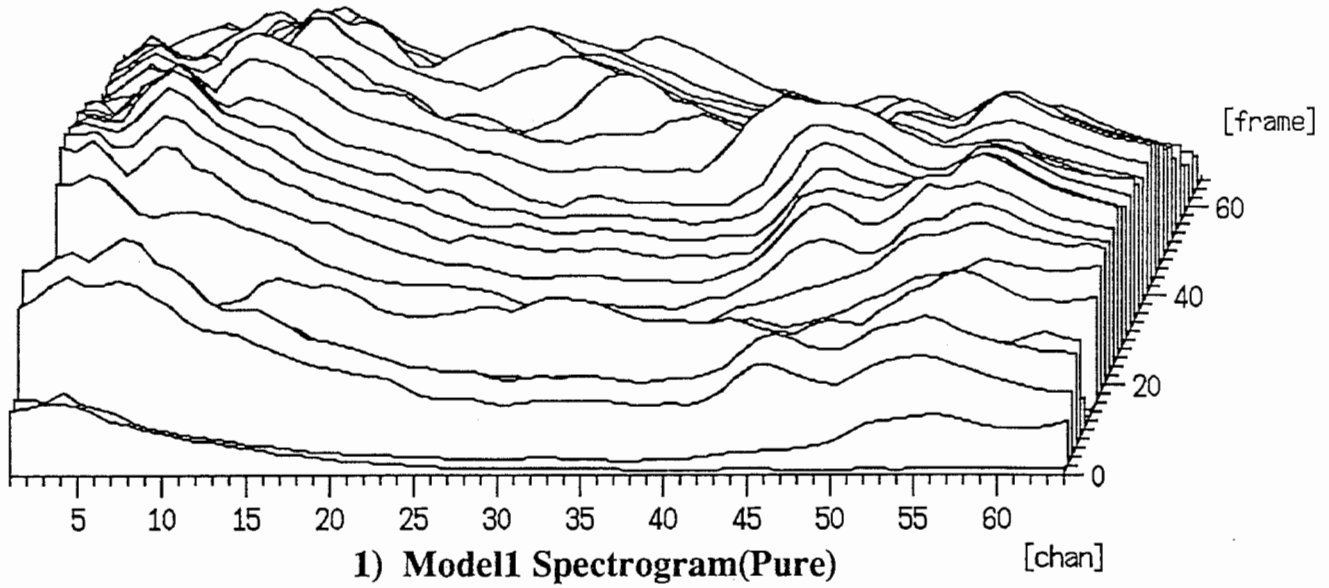




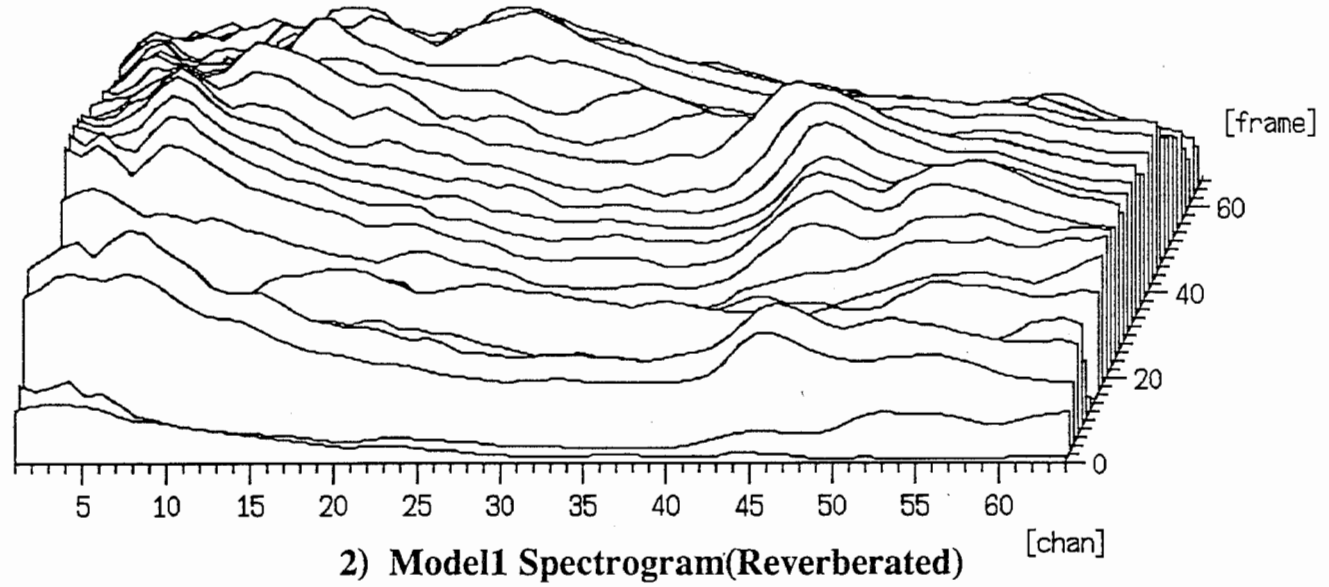
第４－７図　ノイズによるテンプレート単語間距離変動



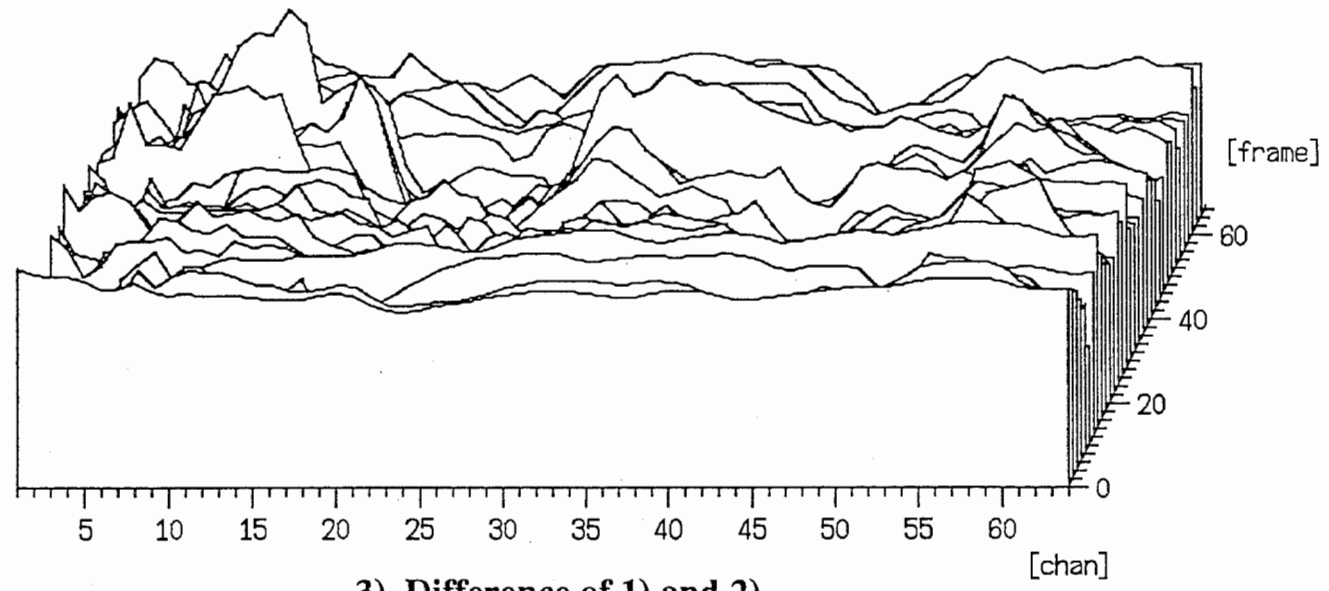
第４－８図　残響付加によるスペクトル変化（Mel ＤＦＴ）



1) Model1 Spectrogram(Pure)

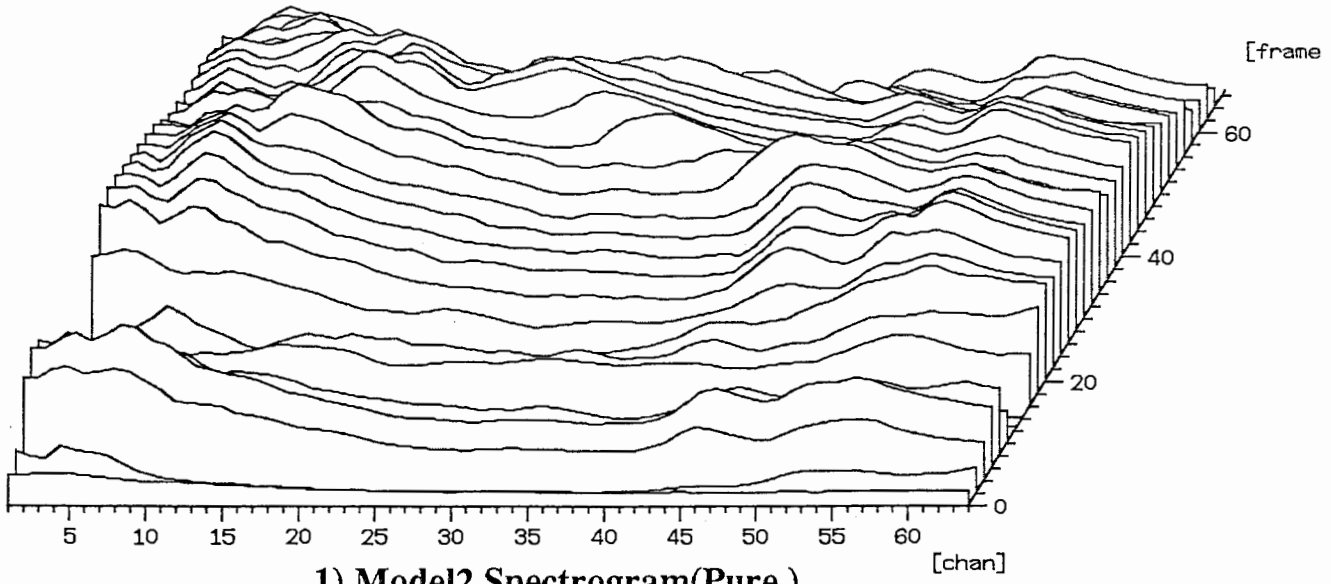


2) Model1 Spectrogram(Reverberated)

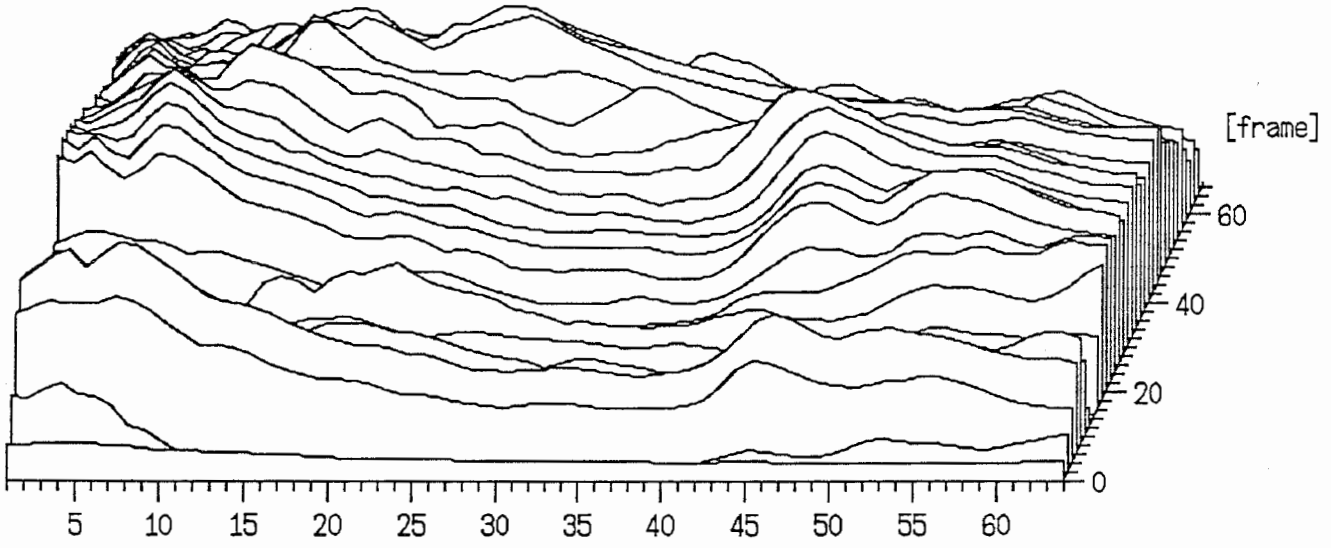


3) Difference of 1) and 2)

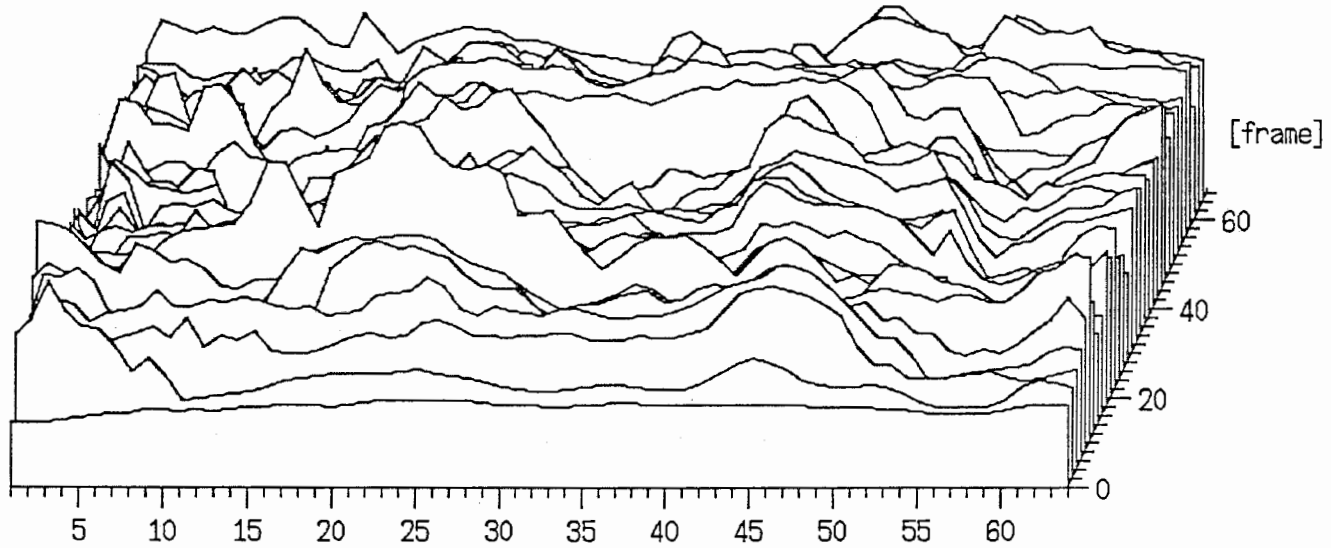
第４－９図　残響付加によるスペクトルの変動（モデル１）



1) Model2 Spectrogram(Pure)



2) Model2 Spectrogram(Reverberated)



3) Difference of 1) and 2)

第４－１０図　残響付加によるスペクトル変化（モデル２）

****

Feature Extraction:DFT
Score:96.296295 Error 8

****

Word 75 recognized as 201 120 18 53 23
Word 76 recognized as 73 201 13 204 76
Word 134 recognized as 110 159 134 18 104
Word 154 recognized as 82 135 1 154 86
Word 155 recognized as 117 94 191 199 39
Word 195 recognized as 178 118 209 201 181
Word 204 recognized as 78 201 204 202 203
Word 212 recognized as 208 212 57 192 213

****

Feature Extraction:Model 1
Score 89.814812 Error 22

****

Word 7 recognized as 130 94 7 26 103
Word 36 recognized as 16 36 187 78 56
Word 50 recognized as 33 50 166 197 8
Word 66 recognized as 38 66 173 174 210
Word 73 recognized as 203 13 73 184 201
Word 75 recognized as 18 20 201 36 184
Word 76 recognized as 73 130 201 78 204
Word 90 recognized as 8 78 90 56 93
Word 91 recognized as 30 91 132 13 27
Word 134 recognized as 116 126 36 110 20
Word 147 recognized as 157 207 201 184 56
Word 154 recognized as 82 55 16 135 21
Word 155 recognized as 117 191 200 39 94
Word 186 recognized as 13 186 203 27 204
Word 187 recognized as 184 187 36 38 120
Word 191 recognized as 201 184 191 117 51
Word 195 recognized as 118 20 82 181 51
Word 204 recognized as 78 204 203 202 21
Word 205 recognized as 117 202 203 205 8
Word 208 recognized as 57 17 101 208 117
Word 211 recognized as 17 211 91 27 130
Word 212 recognized as 57 212 208 193 192

****

Feature Extraction:Model 2
Score: 90.27777 Error 21

****

Word 7 recognized as 130 12 94 205 106
Word 50 recognized as 33 50 98 197 75
Word 75 recognized as 201 16 120 53 18
Word 76 recognized as 201 73 184 13 78
Word 90 recognized as 21 90 8 104 53
Word 98 recognized as 16 98 36 116 58
Word 116 recognized as 16 116 36 58 187
Word 134 recognized as 116 16 110 36 38
Word 147 recognized as 207 201 120 157 56
Word 154 recognized as 135 82 1 55 21
Word 155 recognized as 117 94 39 200 191
Word 186 recognized as 13 202 27 30 120
Word 187 recognized as 184 187 38 201 120
Word 191 recognized as 201 191 5 53 184
Word 195 recognized as 118 20 21 201 178
Word 196 recognized as 173 184 182 38 187
Word 197 recognized as 210 197 200 180 121
Word 203 recognized as 202 203 73 201 13
Word 204 recognized as 202 201 120 21 78
Word 211 recognized as 17 58 100 91 93
Word 212 recognized as 57 208 17 212 135

第４－１１図　認識誤りの例

付録A　ATR音声データベース 音韻バランス単語一覧

単語番号	単語	
1	いきおい	勢い
2	いよいよ	いよいよ
3	うらやましい	羨ましい
4	おもしろい	面白い
5	ぐあい	具合
6	ざいりょー	材料
7	じゅーいちがつ	十一月
8	しゅーきょー	宗教
9	じゅんばん	順番
10	すいちょく	垂直
11	だいじょーぶ	大丈夫
12	だいどころ	台所
13	ちゃんと	ちゃんと
14	ちゅーおー	中央
15	とりあつかう	取り扱う
16	ばしょ	場所
17	びょーいん	病院
18	ひょーじゅん	標準
19	ぶんめー	文明
20	ぽけっと	ポケット
21	ぼんやり	ぼんやり
22	めがね	眼鏡
23	わがまま	我儘
24	うけあう	請け合う
25	うちあわせ	打ち合わせ
26	かれんだー	カレンダー
27	しゅーてん	終点
28	せーい	誠意
29	そびえる	そびえる
30	ちちおや	父親
31	ちゅーねん	中年
32	ておち	手落ち
33	でしゃばる	でしゃばる
34	にせもの	贋物
35	ははおや	母親
36	ひゃくしょー	百姓
37	びょーにん	病人
38	ふらふら	ふらふら
39	ほほえむ	微笑む
40	れじゃー	レジャー
41	わすれもの	忘れ物
42	あかちゃん	赤ちゃん
43	あけがた	明け方
44	あじわう	味わう
45	あたりまえ	当たり前
46	あぱーと	アパート
47	あるみにゅうむ	アルミニュウム
48	あんけーと	アンケート
49	いえで	家出
50	いちじるしい	著しい

単語番号	単語	
51	いちぶぶん	一部分
52	いまごろ	今頃
53	いらっしゃる	いらっしゃる
54	うでまえ	腕前
55	うぬぼれる	自惚れる
56	うめあわせる	埋め合わせる
57	うやまう	敬う
58	えーきゅー	永久
59	えーゆー	英雄
60	えすかれーたー	エスカレーター
61	えねるぎー	エネルギー
62	えぷろん	エプロン
63	えれべーたー	エレベーター
64	おいはらう	追い払う
65	おごそか	厳か
66	おしゃべり	お喋り
67	おぼえ	覚え
68	おもちゃ	おもちゃ
69	おもわず	思わず
70	がいしゅつ	外出
71	ぎこちない	ぎこちない
72	きむずかしい	気難しい
73	ぎゃくたい	虐待
74	きゅーぎょー	休業
75	きゅーげき	急激
76	ぎゅーにゅー	牛乳
77	きゅーりょー	給料
78	ぎょーせー	行政
79	ぎょぎょー	漁業
80	きょくたん	極端
81	きょぜつ	拒絶
82	げひん	下品
83	げんえき	現役
84	こぎって	小切手
85	こころよい	快い
86	ことづて	言伝
87	このあいだ	この間
88	こぴー	コピー
89	ごぶさた	御無沙汰
90	さきほど	先程
91	じぐざぐ	ジグザグ
92	しゅーへん	周辺
93	しゅっせ	出世
94	じゅんじゅんに	順々に
95	じゅんちょー	順調
96	じょーきゃく	乗客
97	しょーふだ	正札
98	しょーりゃく	省略
99	じんいん	人員
100	すぴーど	スピード

単語番号	単語	
101	ぜんしゅー	全集
102	ぜんてー	前提
103	それぞれ	それぞれ
104	それでは	それでは
105	ぞんざい	ぞんざい
106	だいぶぶん	大部分
107	ちゅーしゃ	駐車
108	ちょーえつ	超越
109	つけくわえる	付け加える
110	できごと	出来事
111	てぬぐい	手拭
112	てのひら	手のひら
113	でぱーと	デパート
114	とりあえず	取り敢えず
115	なおさら	なおさら
116	なかなおり	仲直り
117	にゅーいん	入院
118	にゅーじょー	入場
119	にゅーよく	入浴
120	にょーぼー	女房
121	ねあげ	値上げ
122	ねさげ	値下げ
123	のみこむ	飲み込む
124	ぱいぷ	パイプ
125	はなはだ	はなはだ
126	はなばなしい	華々しい
127	ばんぐみ	番組
128	ひきうける	引き受ける
129	ひっくりかえす	ひっくりかえす
130	びょーしゃ	描写
131	ひょーじょー	表情
132	びょーどー	平等
133	ひょーほん	標本
134	ぷろぐらむ	プログラム
135	ぶんるい	分類
136	ぺーじ	ページ
137	べっど	ベッド
138	ぽんぷ	ポンプ
139	みうしなう	見失う
140	みすぼらしい	みすぼらしい
141	みせびらかす	見せびらかす
142	みゃく	脈
143	みょーじ	名字
144	みょーにち	明日
145	めうえ	目上
146	もちぬし	持ち主
147	ものすごい	物凄い
148	ゆーもあ	ユーモア
149	ゆきずまる	行き詰まる
150	ゆびさす	指差す

単語番号	単語	
151	よこづな	横綱
152	よつかど	四つ角
153	よっぱらい	酔っ払い
154	りゃくする	略する
155	りゅーいき	流域
156	りゅーちょー	流暢
157	りゅっくさっく	リュックサック
158	りょーがえ	両替
159	れくりえーしょん	レクリエーション
160	れんあい	恋愛
161	ろくおん	録音
162	ろっかー	ロッカー
163	ろんそー	論争
164	わざわざ	わざわざ
165	わりあてる	割り当てる
166	わりびき	割引き

人為的な追加分

単語番号	単語	
167	せきらんうん	積乱雲
168	たんおんひょーじ	単音表示
169	かんふる	カンフル
170	まえひょーばん	前評判
171	おんなみょーり	女冥利
172	ふせーみゃく	不整脈
173	かりゅー	下流
174	がびょー	画鋲
175	けびょー	仮病
176	ぽぴゅらー	ポピュラー
177	とっきょ	特許
178	そっちょく	率直
179	はんぎゃく	反逆
180	まじょ	魔女
181	ざひょー	座標
182	こんにゃく	こんにゃく
183	かにゅー	加入
184	めにゅー	メニュー
185	こーみょー	巧妙
186	みゅーじっく	ミュージック
187	きにゅー	記入
188	こーにゅー	購入
189	さんみゃく	山脈
190	はっぴゃく	八百
191	けわしい	険しい
192	めーりょー	明瞭
193	かんぴょー	干瓢
194	ついきゅー	追及
195	おう	追う
196	いんりょく	引力
197	ごびゅー	誤謬
198	てちょー	手帳
199	さぎょー	作業
200	だいひょー	代表

単語番号	単語	
201	とーひょー	投票
202	えーぎょー	営業
203	びみょー	微妙
204	じゅみょー	寿命
205	じぎょー	事業
206	じゅーびょー	重病
207	かんびょー	看病
208	でんわ	電話
209	けんきゅー	研究
210	けいびゃく	啓白
211	だきょー	妥協
212	ひゅーひゅー	ひゅーひゅー
213	ぴゅーぴゅー	ぴゅーぴゅー
214	おやぴゅーま	親ピューマ
215	たいぴゅーた	タイピューター
216	かくれぴゅーりたん	隠れピューリタン