音声の線形予測分析における非線形前処理の効果

音声の線形予測分析における非線形前処理の効果

三好　義昭・斉藤　尚弘*

Effect of a Nonlinear Preemphasis on Linear Prediction Analysis of Speech

Yoshiaki MIYOSHI, Naohiro SAITOH

1．まえがき

我々は日常生活において「音声」を用い，情報を互いに伝えあっており，音声は人間の意思の疎通に欠くことのできない重要な媒体の一つである。そのため，音声情報処理技術に関する研究は極めて重要であると言える。

音声研究の長い歴史の中で，音声の音響学的性質について最もよくできた音声生成モデルは，Fantの線形音声生成モデル[1]である。このモデルは，喉から唇までの空間（これを声道と称する）の変化による固有の共鳴作用により，音声の言語的情報をもたらせると考えるものである。したがって，声道の伝達特性，特にその極周波数であるホルマント周波数が音声の重要な特徴を担っており，この周波数を正確に推定することは音声情報処理を行なう上で非常に重要なことと言える。

今日，このホルマント周波数推定手法として，線形予測分析[2][3]が広く用いられているが，従来の線形予測分析では，現時点の音声振幅値は過去の音声振幅値の線形結合で予測できるものと仮定し，この線形結合係数（線形予測係数と称する）を予測した値と実際の値との差（予測誤差と称する）の自乗平均最小の条件より求めている。したがって，従来の線形予測分析は予測誤差の大きさのみに注目し，予測時点の波形の大きさを全く考慮していないことに若干の問題があると言える。例えば，予測誤差の大きさが同じであったとしても，予測時点の振幅の大小によって予測の度合いが異なると考えられる。あるいは，実際の音声では振幅の小さい部分は雑音の影響をより強く受けていると言えるので，振幅の小さい部分での予測誤差はあまり重視しないといった考慮が必要と考えられる。このような観点から，本論文では，音声波形に非線形な波形変換を施した後の音声波を通常の線形予測分析することにより，予測時点の波形の大きさも考慮した線形予測分析法を提案し，そのホルマント周波数推定精度を従来の線形予測分析と比較検討したものである。以下，2．において，波形の大きさも考慮した線形予測分析法の概略を示し，3．において，今回用いた波形変換形を具体的に示す。4．において，合成音のシミュレーションにより本手法のホルマント周波数推定精度の改善度合いを示し，5．では，本手法を実際に自然有声破裂音のホルマント周波数推定に適用して，その有効性を示す。

2．振幅値を考慮した線形予測分析

音声波形の標本値の間に高い相関関係があることは実験的によく知られている。従来の線形予測分析は，この標本値間の高い相関関係に注目し，音声波の第n標本値y_nの予測値$\hat{y}_n$をそれより以前のp個の標本値の線形結合，

$$\hat{y}_n=\alpha_1 y_{n-1}+\alpha_2 y_{n-2}+\cdots+\alpha_p y_{n-p} \qquad (1)$$

で表せると仮定している。ここで，式(1)の係数$\{\alpha_k\}$，$k=1,2,\cdots,p$を線形予測係数と称する。

平成7年9月12日受理

*本学大学院教育学研究科在学中

従来の線形予測分析では，この予測係数を実際の標本y_nと予測した値$\hat{y}_n$との差(予測誤差)，

$$\varepsilon_n = y_n - \hat{y}_n = y_n - \sum_{k=1}^{p} \alpha_k y_{n-k} \quad (2)$$

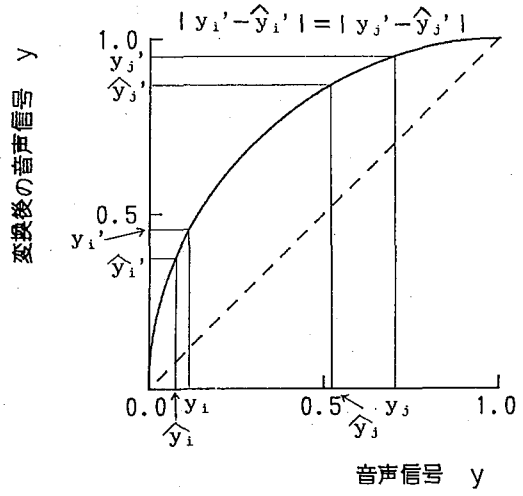
の自乗平均，

$$\overline{\varepsilon_n^2} = \frac{1}{N} \sum_{n=1}^{N} \left(y_n - \sum_{k=1}^{p} \alpha_k y_{n-k} \right)^2 \quad (3)$$

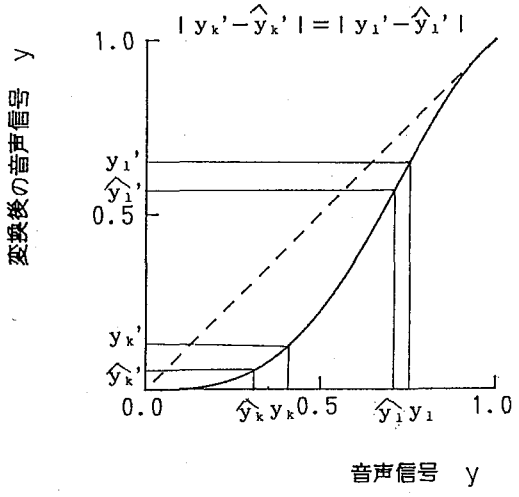
最小の条件より求めている。すなわち，式(3)は正もしくは0の量であり，極値が1つしかなければそれが最小値であるので，線形予測係数$\{\alpha_k\}$は，式(3)の$\overline{\varepsilon_n^2}$を各$\alpha_k$について偏微分し，それを0とおくp元連立1次方程式の解として求められる。そして，その予測係数を係数とするp次方程式を解くことによってホルマント周波数が得られる[4]。

以上のように，従来の線形予測分析では，予測誤差ε_nを評価する際に，音声振幅値y_nの大きさを全く考慮していない点に問題があると言える。すなわち，予測誤差が同じであったとしても，予測時点の振幅値y_nの大小によって，予測の度合いが異なると言える。また，実際の音声では振幅の小さい部分は雑音の影響をより強く受けていると言えるので，振幅の小さい部分での予測誤差はあまり重視しないと言った考慮も必要だと考えられる。

そこで本論文では，音声振幅値の大きさも考慮した線形予測分析の一方法として，音声波形に非線形変換を施す手法を導入する。例えば，図1(a)に示すような非線形変換を音声波形に施した場合，この変換後の波形y_n'を線形予測分析すれば，変換後の波形上では同じ大きさの予測誤差であっても，元の波形y_nから見れば振幅の小さい部分ではより小さい予測誤差となり，振幅の大きい部分ではより大きい予測誤差となるようにすることができ，予測誤差を音声信号に対して相対的に評価したことになる。一方，図1(b)に示すような非線形変換を音声信号に施した音声信号y_n'を線形予測分析すれば，予測誤差を音声信号が小さい比較的振幅レベルの小さな



(a) 模式図I



(b) 模式図II

図1　波形変換の模式図

部分では余り重要視せずに分析することになり，雑音がある場合に有効であると思われる。

実際の音声波は多種多様であるので，振幅レベルの小さい部分を重要視して分析したほうがよい場合や，そうでないほうがよい場合があり，分析する際の非線形変換形（以後，波形変換形と称する）の設定が重要となる。

本論文では，このことを考慮して音声波形に様々な波形変換を施し，合成音声によるシミュレーション及び自然音声に適用することにより本手法の有効性を検討したもので，以下，詳細に述べる。

3．波形変換形

今回用いた非線形変換を式(4)に示す。

$$y_n' = \{a y_n^3 + b y_n^2 + (1-a-b) y_n\}^q \qquad (4)$$

但し，y_nは絶対値の最大値で正規化されているものとする。この非線形変換の例を図2に示す。図2は，変換前の正規化された音声信号y_nと変換後の音声信号y_n'との関係を示したもので，破線は変換を行なわない場合に対応する。残りの実線は，それぞれ(a=3.0, b=-5.0, q=0.3)，(a=-1.0, b=1.5, q=1.0)，(a=2.0, b=-1.5, q=2.0) とした場合の例である。

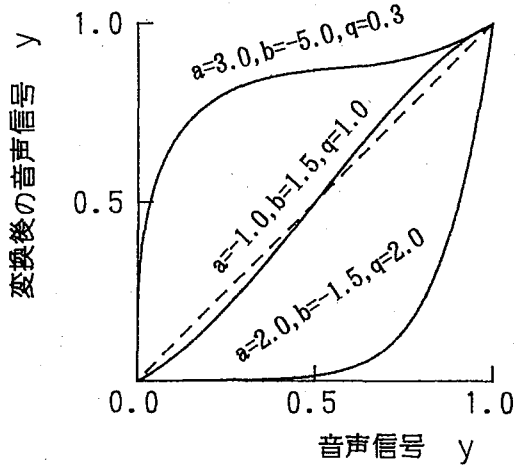


図2　非線形変換の例

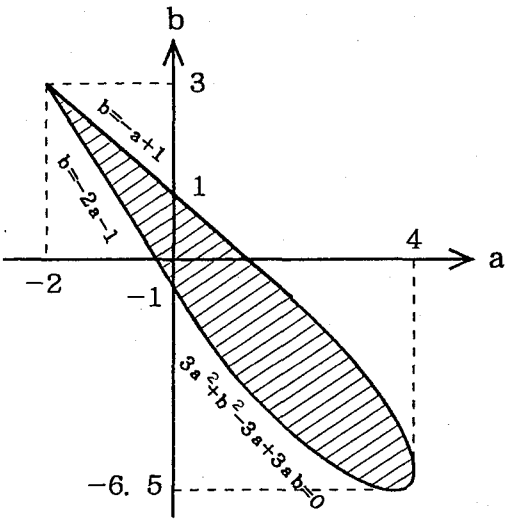


図3　変換係数a，bの範囲(斜線部)

なお，係数a，bの範囲は，式(4)が$0 \leqq y_n \leqq 1$に対し，$0 \leqq y_n' \leqq 1$であり，かつ単調増加関数となる条件から導出される，$a+b \leqq 1$，$2a+b-1 \geqq -1$，$3a^2+b^2-3a+3ab \leqq 0$の条件を満たす図3に示す範囲である。また係数qの範囲は，変換形状の視察に基づき$0.1 \leqq q \leqq 2.5$とした。

以後，式(4)中の係数a，b及びqを変換係数と称する。

4．合成音によるホルマント周波数推定精度の検討

標本化周波数 10kHz，励振源：ピッチ周期8msのRosenberg波[5]，ホルマント周波数：/a/(F_1=812.5Hz, F_2=1312.5Hz, F_3=2562.5Hz)，/i/(F_1=312.5Hz, F_2=2187.5Hz, F_3=2937.5Hz)，/u/(F_1=312.5Hz, F_2=1187.5Hz, F_3=2187.5Hz)，/e/(F_1=562.5Hz, F_2=1812.5Hz, F_3=2562.5Hz)，/o/(F_1=562.5Hz, F_2=1062.5Hz, F_3=2562.5Hz)，F_4=3437.5Hz及びF_5=4437.5Hz，放射特性：6dB/octとして作成した合成五母音を用いて，本方法のホルマント周波数推定精度の改善を明らかにする。

式(5)で定義する第1～第3ホルマント周波数推定誤差Eを従来の方法と比較して表1に示

表1　合成五母音におけるホルマント周波数推定誤差

母音	非線形変換				従来の線形予測分析による誤差 (Hz)
	最適な変換パラメータ			最小誤差 (Hz)	
	a	b	q		
/a/	-1.0	1.3	1.0	2.4	3.2
/i/	-0.1	1.0	0.6	0.6	2.5
/u/	3.1	-5.3	2.5	1.0	2.1
/e/	-0.2	0.9	0.8	2.6	4.9
/o/	-0.1	0.6	0.8	3.3	4.6

す。但し，前処理として一階差分を行い，分析次数p=12，分析窓長 Ta=25.6ms，N=20(0.4 ms 間隔で1ピッチ周期 8.0ms に渡り分析)とした場合の結果である。また，本手法の変換係数a及びbは図3に示す範囲内をそれぞれ0.1の精度で変化させ，また，係数qは0.1～2.5までを0.1の精度で変化させて得られた最小誤差を示したもので，その時の各変換係数値を表中に示す。なお，各変換係数の範囲ならびに精度は以後全て同じとする。

$$E=\frac{1}{3N}\sum_{j=1}^{N}\sum_{i=1}^{3}|F_{ij}-F_i| \qquad (5)$$

但し，F_{ij}：第j分析フレームでの第i推定ホルマント周波数

F_i：合成音の第iホルマント周波数

表1より，合成母音/a/においては，本方法における変換係数をa=-1.0, b=1.3, q=1.0とした時に，ホルマント周波数推定誤差が3.2Hzから2.4Hzに改善することが分かる。そして，特に母音/i/, /u/の改善度合いが著しく，五母音平均では，ホルマント周波数推定誤差が3.5Hzから2.0Hzと約40%改善することが分かる。

雑音を付加した合成母音/a/のホルマント周波数推定誤差のSN比依存性を図4に示す。但し，図中○印：本方法による誤差，×印：従来の方法による誤差である。

図4より，S/N≧14dBでは本方法による顕著な改善効果は得られないが，S/N≦12dBにおいて改善効果が得られ，特にS/N=10dBでは，ホルマント周波数推定誤差が従来の方法では9.0Hzであったのが本方法により4.8Hzとなり，誤差が大幅に改善することが分かる。なお，この時の本手法における変換係数a，b及びqは，それぞれa=-0.4，b=0.7，q=1.0であった。

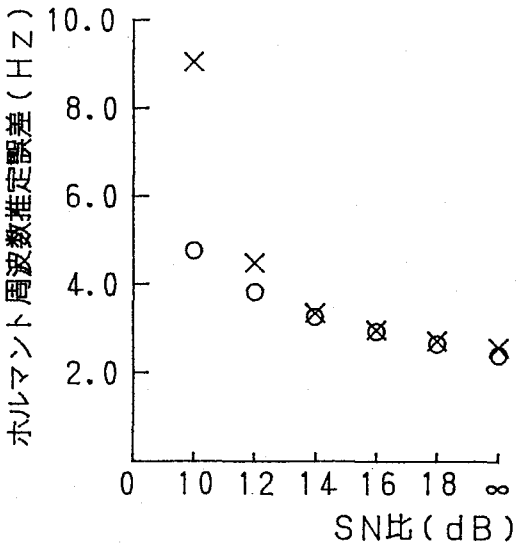


図4　ホルマント周波数推定誤差のSN比依存性(合成母音/a/)

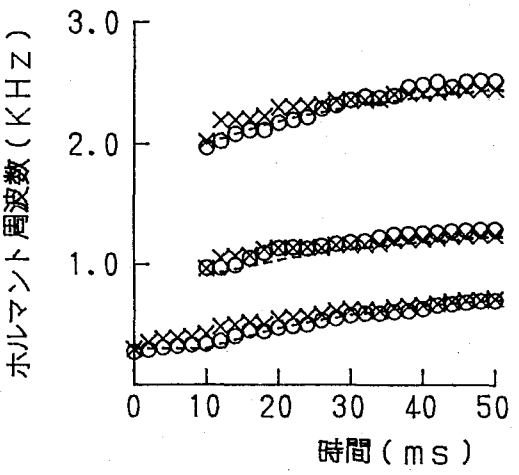


図5　合成有声破裂音/ba/のホルマント追尾

合成有声破裂音 /ba/ のホルマント追尾結果を従来の線形予測分析と比較して図5に示す。但し，前処理として一階差分を行い，分析次数 p=12, 分析窓長 Ta=25.6ms とし，○印：本方法の変換係数をそれぞれ a=3.0, b=-6.0, q=0.4と設定した場合に推定されたホルマント周波数，×印：従来の線形予測分析により推定されたホルマント周波数，破線は合成音のホルマント周波数を示す。

図5より，従来の線形予測分析では，破裂時点（今の場合，10msの時点）付近のホルマント周波数推定誤差が大きいのに対して，本方法では，より正確なホルマント軌跡が推定でき，特に有声破裂音を相互に識別する際に重要となる破裂時点付近の誤差が大幅に改善されていると言える。

合成有声破裂音 /ba/, /da/, /ga/ の破裂時点付近の第1～第3ホルマント周波数推定誤差（式(5)で定義する誤差E）を従来の線形予測分析と比較して表2に示す。但し，分析条件等は図5と同じとし，破裂時点から破裂時点後10msまでを2ms間隔で分析した場合（式(5)において，N=6）の結果である。なお，本方法の各変換係数の値を表中に示す。

表2より，例えば /ba/ では，本方法の変換係数を a=3.0, b=-6.0, q=0.4に設定すれば，ホルマント周波数推定誤差が35.4Hzとなり，従来の方法による誤差99.3Hzより少ない誤差でホルマント周波数を推定できることが分かる。

表2　合成有声破裂音におけるホルマント周波数推定誤差

有声破裂音	非線形変換				従来の線形予測分析による誤差 (Hz)
	最適な変換パラメータ			最小誤差 (Hz)	
	a	b	q		
/ba/	3.0	-6.0	0.4	35.4	99.3
/da/	3.3	-3.7	1.0	75.1	95.7
/ga/	3.5	-6.3	0.5	83.5	104.1

同様に /da/, /ga/ 各合成有声破裂音においても従来の方法より，少ない誤差でホルマント周波数推定されるとの結果が得られた。但し，最適な変換係数は /ba/, /da/, /ga/ それぞれ異なる結果となった。

5．自然有声破裂音への適用例

成人男性が発声した自然有声破裂音 /ba/ のホルマント追尾結果を図6に示す。但し，前処理として一階差分を行い，分析次数 p=12, 分析窓長 Ta=25.6ms とし，○印：本方法の変換係数をそれぞれ a=3.4, b=-3.9, q=1.2と設定した場合に推定されたホルマント周波数，×印：従来の線形予測分析により推定されたホルマント周波数である。図6より，従来の線形予測分析の結果と比較すると，第1及び第2ホルマント周波数についての追尾ではさほど差はないが，第3ホルマント周波数については，破裂時点（今の場合，10.0msの時点）付近のホルマント周波数変化に差が見受けられる。しかしながら，自然音声では真のホルマント周波数が未知であるので，その誤差を数量的に評価できない。

したがって以下，破裂時点付近での第2，第3ホルマント周波数を特徴パラメータとしてホ

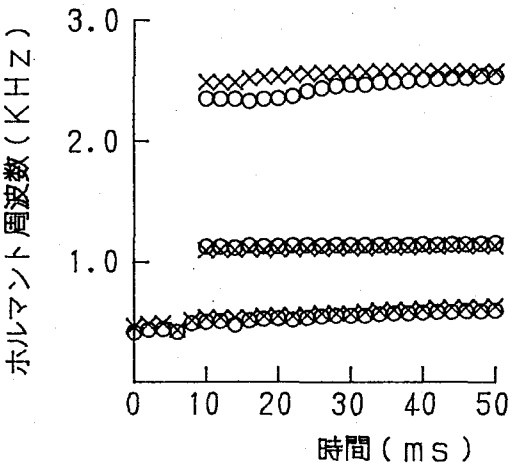


図6　自然有声破裂音/ba/のホルマント追尾

ルマント空間での自然有声破裂音識別を行い，本手法の有効性を検討する。具体的には，第2－第3ホルマント周波数空間での /b/, /d/, /g/ 各クラスの重心からの距離による識別率ならびに式(6)で定義する類間分散と類内分散の比である分散比Dの良さで評価した。

$$D = \frac{\frac{1}{3}\sum_{i=1}^{3}(G_i-G)^T(G_i-G)}{\frac{1}{3N}\sum_{i=1}^{3}\sum_{k=1}^{N}(X_{ik}-G_i)^T(X_{ik}-G_i)} \quad (6)$$

$$G_i = \frac{1}{N}\sum_{k=1}^{N} X_{ik}\ ,\quad G = \frac{1}{3}\sum_{i=1}^{3} G_i$$

但し，G_i：クラスiの重心ベクトル，X_{ik}：クラスiのk番目の資料の特徴ベクトル(今の場合，第2及び第3ホルマント周波数) である。

なお，音声資料としては，電子協日本語共通音声データベース中の20代及び30代の男性計45人の単音節 /ba/, /da/, /ga/ (但し，2回目の発声) 計135個を用いた。したがって，式(6)中のNはN＝45となる。

本方法による識別率及び分散比の分析位置依存性を従来の線形予測分析と比較して図7に示す。但し，前処理として一階差分を行い，分析次数p＝12，分析窓長 Ta＝25.6ms，○印：本方法による識別率及び分散比，×印：従来の線形予測分析による識別率及び分散比である。なお，分析位置 0ms が破裂時点であり，本方法の変換係数はそれぞれ，a＝3.4, b＝-3.9, q＝1.2 に設定した結果である。

図7より，分析位置12msにおいて，従来の線形予測分析では識別率80.7%，分散比0.39であったのが，本方法により，識別率89.6%，分

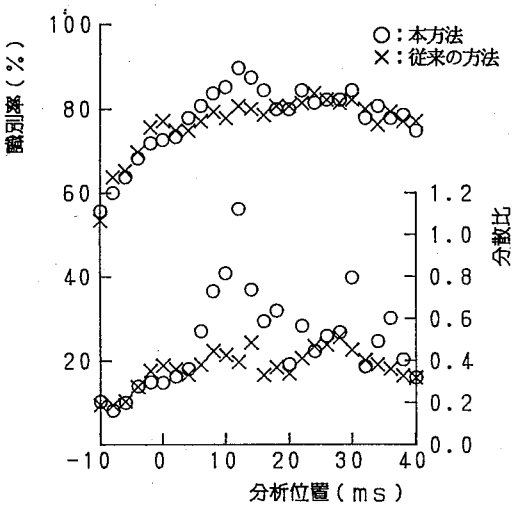
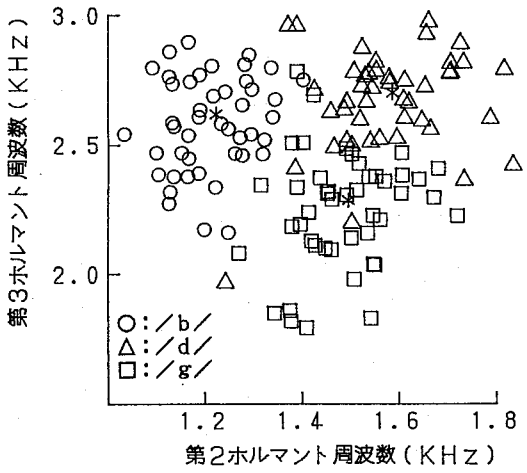
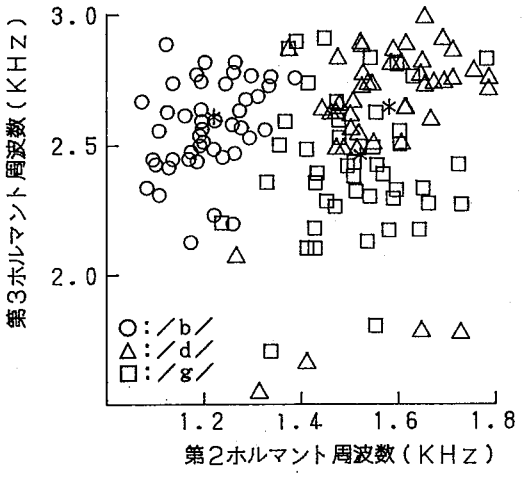


図7　識別率及び分散比の分析位置依存性



(a)　本方法



(b)　従来の方法

図8　ホルマント空間における有声破裂音(後続母音/a/)の分布

散比1.12と識別率ならびに分散比ともに向上し，本手法の有効性が示されていると言える。

図7の分析位置12msにおける第2－第3ホルマント周波数空間での/b/, /d/, /g/の分布の様子を図8に示す。但し，分析条件等は図7と同じであり，図中の○，△，□及び＊印はそれぞれ/b/, /d/, /g/及び各有声破裂音の重心位置を示す。

図8より，従来の線形予測分析では，/d/と/g/の分布がオーバーラップしているのに対して，本方法では/d/, /g/がそれぞれよりまとまった分布となっていると言える。この第2－第3ホルマント空間において，/b/, /d/, /g/相互を各重心からのユークリッド距離により識別した結果を表3に示す。表3より，本手法における/b/の識別率が従来の方法より若干悪くなっているが，/g/においては，従来の線形予測分析では，正しく/g/として識別できているのは音声資料45個中30個しかなく，13個が/d/と誤識別され，識別率が66.7%にしかならないのに対して，本方法では5個が/d/，2個が/b/に誤識別されてはいるが，他の38個は正しく/g/と識別され，識別率が84.4%と大幅に改善することが分かる。そして，/d/においても識別率が改善しており，/b/, /d/, /g/平均して80.7%から89.6%に改善し，かつ分散比も0.39から1.12に良くなり，本方法の有効性が示されていると言える。

表3　音声破裂者の識別結果
（後続母音/a/）

(a)　本方法

	/b/	/d/	/g/	識別率(%)
/b/	42	1	2	93.3
/d/		41	4	91.1
/g/	2	5	38	84.4
分散比　1.12			平均	89.6

(b)　従来の方法

	/b/	/d/	/g/	識別率(%)
/b/	43	1	1	95.6
/d/		36	9	80.0
/g/	2	13	30	66.7
分散比　0.39			平均	80.7

6．むすび

通常の線形予測分析を行う場合，入力音声波の周波数スペクトル平坦化を主な目的として，一般に一階差分等の前処理が用いられているが，本論文では，さらに一歩押し進めて，音声波に非線形な前処理を施した後の音声波を通常の線形予測分析することにより，実効的に予測誤差を振幅値との相対値として評価する，あるいは雑音が存在する場合，雑音の影響をより強く受けていると言える振幅値の小さい部分での予測誤差を余り重視しないで分析することができることを示し，その有効性を合成音のホルマント周波数推定ならびに自然有声破裂音のホルマント周波数推定に適用して検討した。その結果，雑音等が存在しない理想的な合成母音では，第1～第3ホルマント周波数推定誤差が従来の線形予測分析法では平均して3.5Hzであるのに対し，本方法では2.0Hzと約40%改善し，また雑音が存在する場合，特にS/N＝10dBにおいて，従来の線形予測分析法による誤差が9.0Hzであったのが，本方法により，4.8Hzとほぼ半減するとの結果が得られた。さらに破裂時点付近の第2，第3ホルマント周波数による自然有声破裂音識別では，従来の線形予測分析法による識別率が80.7%であったのが，本方法により，89.6%に向上するとの結果が得られ，本手法の有効性が明かとなった。

なお，これらの結果は，本手法の変換係数を最適に設定した場合に得られるのであり，変換係数の設定手法の検討が今後の課題と言える。

文 献

(1) G. Fant : "Acoustic theory of speech production", Mouton (1960).
(2) 板倉, 斉藤 : "統計的手法による音声スペクトル密度とホルマント周波数の推定", 信学論(A), 53-A, 1, pp.35-42 (1970-01).
(3) B. S. Atal and S. L. Hanauer : "Speech analysis and synthesis by linear prediction of the speech wave", J. Acoust. Soc. Amer., 50, pp. 637-655 (1971).
(4) J. Makoul : "Linear prediction : a tutorial review", IEEE Proc., 63, 4, pp.561-580 (1975).
(5) A. E. Rosenberg : "Effect of glottal pulse shape on the quality of natural vowels", J. Acoust. Soc. Amer., 49, pp.583-590 (1971).