HOKKAIDO UNIVERSITY

Title	小区間線形予測分析法と母音音声の変動に関する検討
Author(s)	村上, 寛之; 斎川, 勝男; 栃内, 香次; 永田, 邦一
Citation	北海道大學工學部研究報告, 132, 119-128
Issue Date	1986-07-31
Doc URL	http://hdl.handle.net/2115/41998
Type	bulletin (article)
File Information	132_119-128.pdf

Instructions for use

Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

北海道大学工学部研究報告
第132号（昭和61年）

Bulletin of the Faculty of Engineering,
Hokkaido University, No. 132 (1986)

小区間線形予測分析法と母音音声の変動に関する検討

村上　寛之　斎川　勝男　栃内　香次　永田　邦一

（昭和61年3月31日受理）

Studies on Short-term LPC Analysis and Variability of Vowel Sounds

Hiroyuki MURAKAMI, Katsuo SAIKAWA, Koji TOCHINAI and Kuniichi NAGATA

(Received March 31, 1986)

Abstract

Over long periods, words uttered by the same speaker, do not always have the same spectrum time patterns. This variability is the main cause for increases in error rates in automatic speakers varification.

A study on short-term LPC analysis was made using synthetic vowel sounds.

Using the short-term LPC analysis, the spectra of vowelsounds can be analyzed into the transfer-function of the vocal tract and the glottal source spectra.

Vowel sounds of Japanese spoken number digits uttered over one year period were analyzed, and it was observed that variability of the transfer-function is dominant in general.

1．まえがき

同一話者の音声のスペクトルが，発声時期により変動することは周知の事実であり，それが話者照合における照合率低下の主な原因となっている[1)2)]。すでに音声スペクトルの概形の傾斜を用いて，声道伝達特性と音源特性を分離して音声パラメーターの変動を調査した結果が報告されているが[3)]，ここでは音声の母音区間に着目し，変動の原因が声道伝達特性の変動と声帯波形の変動のいずれに起因するかを調べるために，小区間分析法を用いて両者の分離を行った。小区間分析法については，すでに報告もなされているが[4)]，さらに各種の合成音声を用いて小区間分析法に関する検討を行い，次に実音声母音部に小区間分析法を適用し，音声スペクトルを声道伝達特性と声帯波形とに直接分離してその変動の調査，検討を行った。その結果について述べる[5)6)7)]。

2．小区間分析法の検討

小区間分析法を用いて音声を分析し，その声道の伝達特性と声帯波形に分離する場合に，その推定精度を確かめるには，実際にその音声が生成された声道伝達特性と声帯波形を知っておく必要がある。したがって既知の声道伝達特性と声帯波形を用いて合成音声を作り，それを音声サンプルとして用いて分析精度の検討を行った。

電子工学科　電子機器工学講座

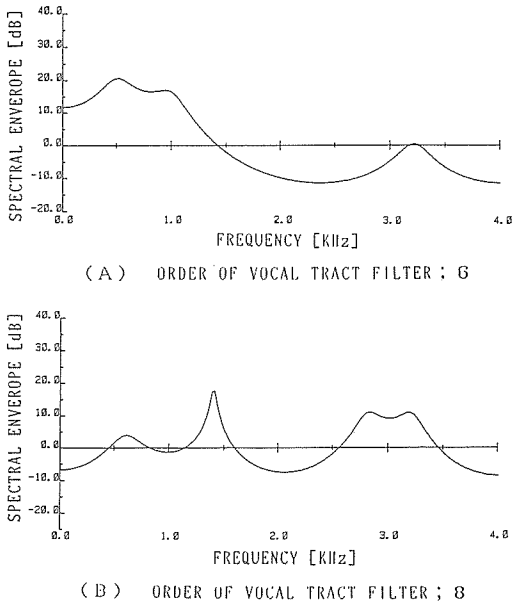

Fig. 1 Examples of transfer-functions of vocal tract filters for synthetic vowel sounds.

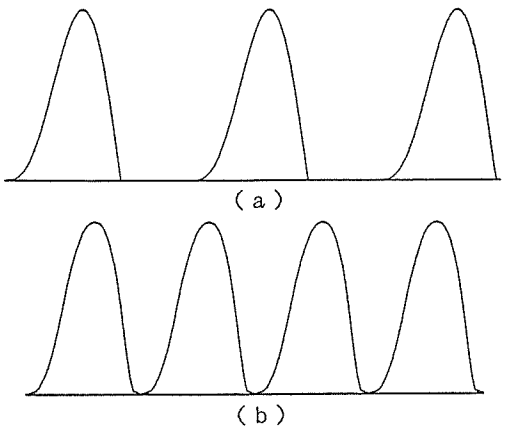

Fig. 2 Examples of glottal waveforms for synthetic vowel sounds.

2．1　音声サンプル

合成音声サンプルは，Fig. 1に示すような伝達特性を持つ全極形フィルター((A)は6次，(B)は8次）を擬似声帯波で駆動し，その出力を唇の放射特性の近似として1回差分して得られたものである。Fig. 2に擬似声帯波の一例を示す。Fig. 3で (a) はローゼンベルグ波，(b) は実音声からとり出された声帯波形である。実音声からとり出されたという意味は，実音声に対して，分析位置を1サンプル点ずつ移動させながら小区間分析を行い，得られた声帯波形の中から比較的なめらかなもの（声帯波形として不自然でないもの）を選出したという意味であり，これが正しい声帯波形であるという事ではない。(b) の波形には休止区間に相当する部分が存在しないが，これには測定系の位相回転の影響などが考えられる。合成音声の記述法として，例えばFig. 1(A)のフィルターをFig. 2(a)の波形で駆動した出力を1回差分して得られた合成音声を以後"合成音声A−a"のように記述することにする。

分析は合成音声の1回差分信号に対して行われるが，これは擬似声帯波の2回差分波形でフィルターを駆動して得られる出力信号を分析することと等価であり，擬似声帯波に直線的に変化する区間が存在すれば，その区間も休止区間とみなすことができる。尚，擬似声帯波の2回差分波形をexcitation波形，excitation波形に現われるメインピークをmain excitationと呼ぶことにする。

2．2　一致性の評価

推定された線形予測係数の正確さの評価には，合成に用いたフィルターの伝達特性と推定された伝達特性のボード線図上の表示による一致度の視察，ならびに，合成に用いたフィルターの係数と，それによる合成音声から推定された線形予測係数からそれぞれ50次までのケプストラム係数を求め[8]，（1）式によって与えられる値 d^2 を尺度として用いた。

$$d^2=\sum_{k=1}^{50}(C_k-C_{ko})^2 \qquad (1)$$

C_k : the kth order cepstrum coefficient estimated from the synthetic vowel sound.

C_{k0} : the kth order cepstrum coefficient of the transfer function of the vocal tract filter.

d^2は推定された伝達特性（対数スペクトル）とフィルターの伝達特性（対数スペクトル）との差の2乗平均に相当する量である。また，推定された声帯波形の正確さの評価には，その時間波形の視察，ならびに，合成に用いた声帯波形と推定された声帯波形をそれぞれ分析長30 msecで14次の線形予測分析を行い，50次までのケプストラム係数を求めて(1)式と同様にd^2を求め尺度とした。このd^2は，合成に用いた擬似声帯波形のスペクトルエンベロープと推定された声帯波形のスペクトルエンベロープとの差の2乗平均に相当する量である。

2．3　各種合成音声と分析結果

合成音声に対して分析位置を1サンプル点ずつ移動させながら，1周期にわたって分析長20サンプル点の小区間分析を行った。Fig. 3 (a) は合成音声A−aを6次（合成次数と同次）で分析した結果である。図で縦軸は(1)式で計算される量d^2，横軸は分析開始サンプル点を表しており，0はmain excitationがかかっているサンプル時点である。図中，実線（TFUNC），破線（GLOTT），点線(EXCIT)はそれぞれ声道伝達特性，声帯波形，声帯波形の2回差分波形のd^2の変化を表している。Fig. 3 (b) は同じ合成音声を8次（合成次数より高次）で分析した結果である。この合成音声には自由応答区間が存在し，6次分析の場合は自由応答区間内で正確な推定が行われているが，8次分析の場合はmain excitationのかかっている点の直後でのみ正しい推定が行われ，他の分析位置では自由応答区間内であっても正しい推定が行われていない。このよう

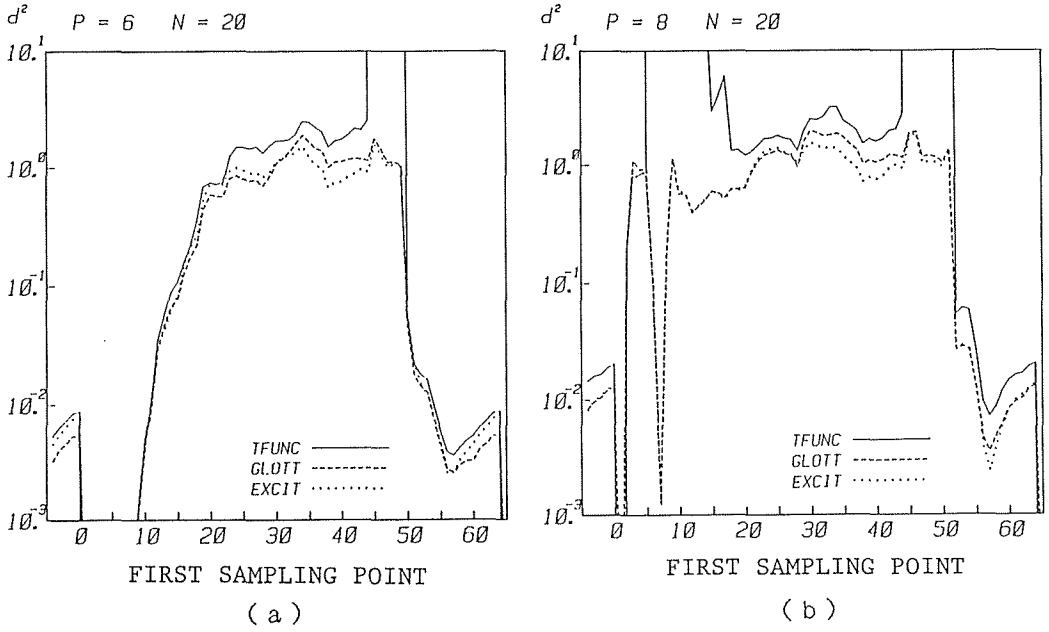

Fig. 3　d^2 versus location of the first sampling point of analysis segments for the synthetic vowel sound A−a.

な現象は，一般に完全な自由応答区間内で合成次数よりも高次で分析したときに現われる。

このことをさらに明確にするために，6次のフィルターAを1個のインパルスで駆動して合成音声をつくり，同様の実験を行った結果をFig. 4に示す（ただし，この場合のみ差分処理は行っていない）。図中，0の時点でインパルス（excitation）が加わっており，インパルスによる自由応答区間内を，分析位置を1サンプル点ずつうしろに移動させながら分析を行った結果である。図から分かるように，6次で分析した場合は正確な推定が可能で，信号の減衰及び量子化誤差に伴いd^2の値が徐々に大きくなってゆくが，8次で分析した場合はインパルスが加わっている点の直後の分析区間でのみ正確な推定が可能で，それより1点でもずれると正確な推定が行われなくなる。これは，完全な自由応答区間内で高次分析を行うと，共分散行列のランク落ちなど非正則性の影響が現われるためと考えられる[9]。しかし，excitationのかかっている点の直後の自由応答区間に分析区間を設けたときは，高次分析の場合でも共分散行列の正則性が保たれるため正確な推定が可能になる。尚，このとき推定される線形予測係数は，合成に用いたフィルタの係数とほぼ等しく，分析により得られる合成次数より高次の線形予測係数はほぼ0に推定される。以上のことから，自由応答区間のある合成音声の場合は，excitation直後の自由応答区間内に分析区間を設けると，合成次数の決定も可能であるといえる。

次に，Fig. 2 (a) の波形にノイズ（ピーク値比0．1）を重畳した波形（これを擬似声帯波a'とする）で6次のフィルターを駆動した合成音声A−a'に対する同様の実験の結果をFig. 5に示す。図中，pは分析次数である。また，声帯波形のd^2は声道伝達特性のd^2とほぼ同じような変化をするので省略する。この合成音声には完全な自由応答区間が存在しないため，いずれの分析位置においても正確な推定は行われていないが，6次，8次分析ともにmain excitation直後の分析区間でd^2が最小となっており，最も良好な推定が行われていると考えられる。

Fig. 6は合成音声B−bに対しての結果である。この場合もmain excitationのかかっている時

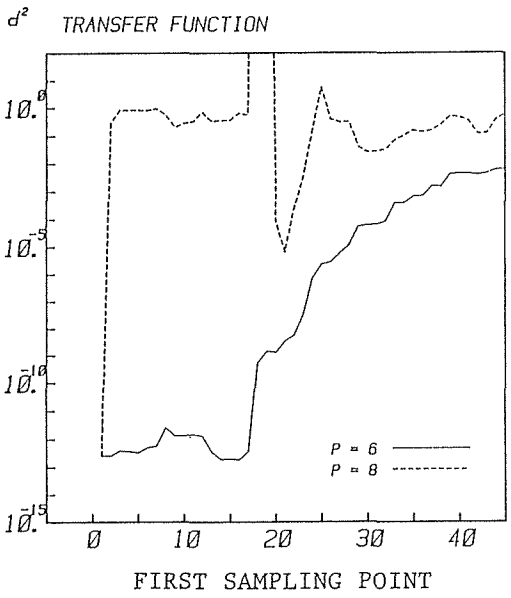

Fig. 4 d^2 versus location of the first sampling point of analysis segments.

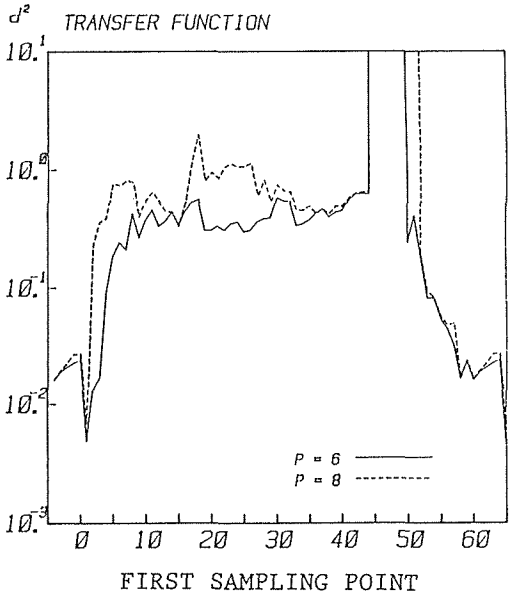

Fig. 5 d^2 versus location of the first sampling point of analysis segments for the synthetic vowel sound A−a'.

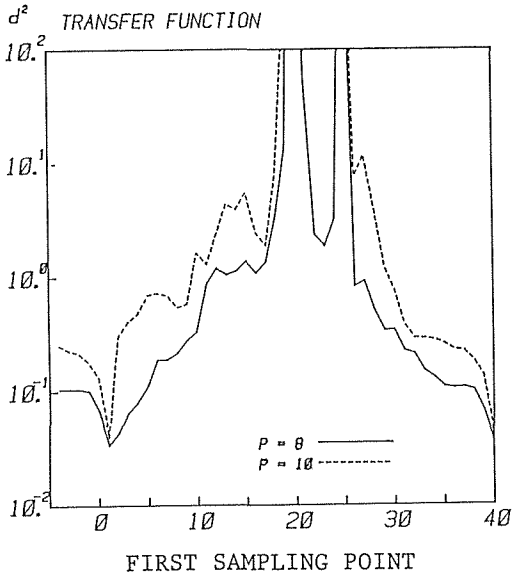

Fig. 6 d^2 versus location of the first sampling point of analysis segments for the synthetic vowel sound B−b.

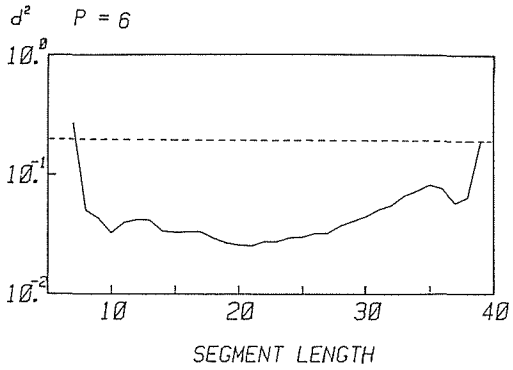

Fig. 7 d^2 versus length of analysis segment for the synthetic vowel sound A−b.

点の直後の分析区間で最も良好な推定がなされている。この他にも実音声からとり出された数個の擬似声帯波で駆動した合成音声を用いて実験を行ったが，ほぼ同様の結果が得られた。

次に，小区間分析の分析長と d^2 との関係を調べるために，分析開始点を main excitation 直後に固定し，分析長を変化させた場合の d^2 の変化の様子が調べられた。Fig. 7 は合成音声 A−b についての結果である。図で，破線は分析長が 1 周期のサンプルにわたった時の d^2 の値である。他の合成音声についても同様の実験を行ったところ，分析長と分析精度 d^2 の関係は各合成音声ごとに異なり，最適な分析長を決定することはできなかった。

各合成音声の推定された声道伝達特性と声帯波形の数例を Fig. 8 に示す。Fig. 8 (a) は合成音声A−aをmain excitation直後の自由応答区間で 6 次で分析した結果で，正確な特性の推定が行われている。この分析位置では高次で分析を行っても同様の結果が得られる。(b) は同じ合成音声を main excitation の 2 点後ろに分析開始点を設け 8 次で分析を行った場合である。分析区間は自由応答区間に含まれているが，正確な推定は行われていない。(c) (d)はそれぞれ合成音声 A−a', B−b を main excitation 直後で分析した結果（分析次数はそれぞれ 6 次，10 次）であり，ほぼ正確な推定がなされている。

2.4 結果のまとめ

各種合成音声による小区間分析法の検討の結果，フィルターの特性や駆動に用いた声帯波形の形によって多少の差は見られるが，main excitation のかかっている時点の直後に分析区間を設けたときに最も良好な推定が行われることが確かめられた。特に実音声からとり出された擬似声帯波（図 2 (b) など）は実音声の声帯波形に近いものと考えられるので（ただし，測定系の位相特性の影響を含む），それらによって駆動された合成音声に関する実験結果は，実音声の場合にも適用できるものと考えられる。したがって，本分析による分離精度はすべての母音音声に対して十分とは言えないが，同一人物の同一音声について分析を行い，声道伝達特性と声帯音源波形に

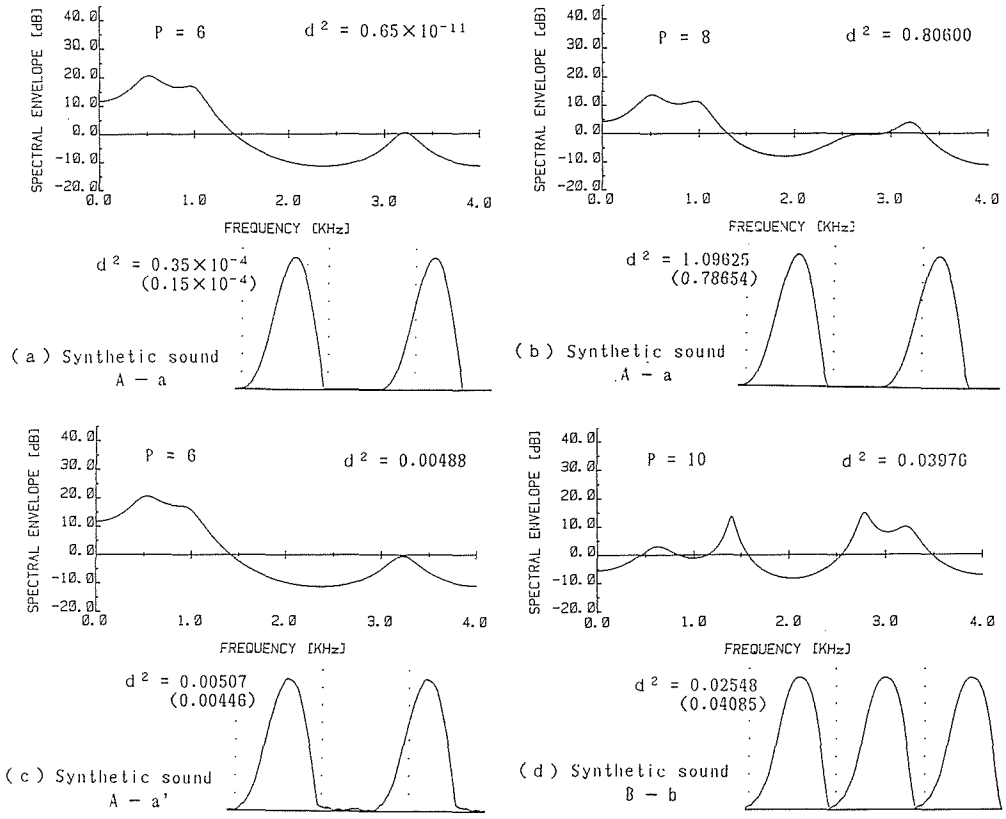

Fig. 8 Estimated vocal tract transfer-functions and glottal waveforms.

分離し，それぞれの変化を検討することは十分可能であると思われる。

3．発声時期による音声スペクトルの変動

発声時期による音声スペクトルの変動の原因を調査するため，スペクトル変動の比較的大きい音声試料について以下に述べる分析実験を行った。

3．1　音声サンプル

実音声のサンプルは，男性話者7名が約1年間にわたる7時期に発声した数字音声の中から，母音部分の変動が比較的大きい話者4名の数字音声0，4，7に含まれる母音/e/,/o/,/a/について，第1時期からリファレンスを選び，同じ第1時期の中で変動の大きいものと，他の時期で変動の大きいものを選んで分析を行った。

3．2　母音部分の切り出し

まず，各話者，各数字音声のリファレンス音声から，波形及びスペクトルの視察ならびにスペクトルの変化の計量によって母音定常部を求める。次に，リファレンス音声とテスト音声（同一話者，同一数字音声）とのDPマッチングを行い[10]，得られたDPパスからリファレンス音声の母音定常フレームに対応するテスト音声のフレームを求めた。この様子をFig.9に示す。このようにして得られたリファレンス及びテスト音声の対応するフレームの1次差分信号のスペクトルエン

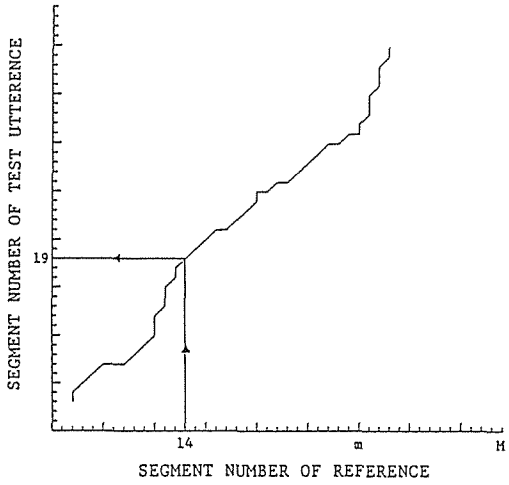

Fig. 9 An example of DP path.

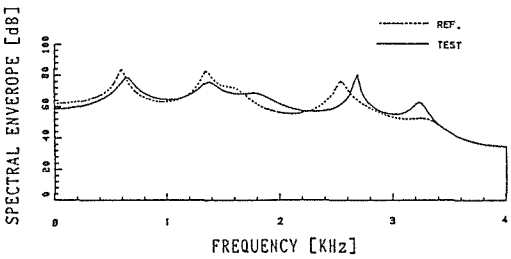

Fig. 10 Spectrum envelopes of time differential signals of corresponding vowel segments.

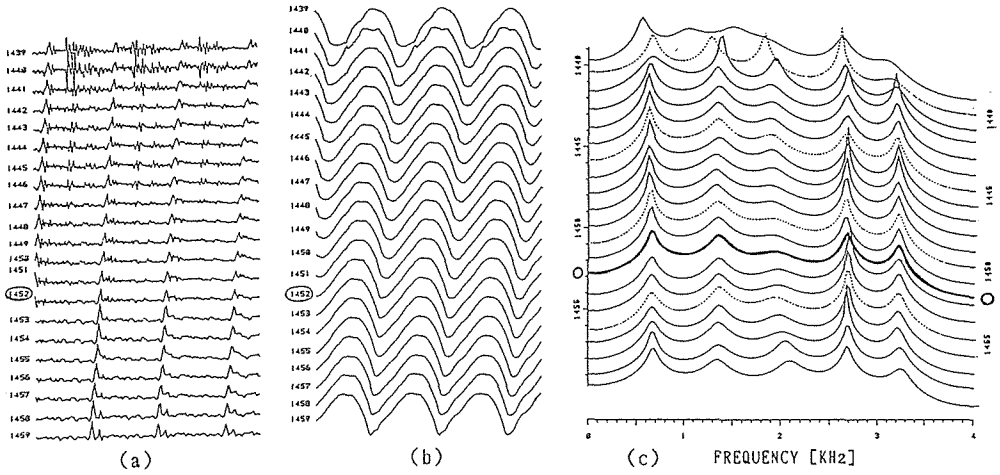

Fig. 11 (a) prediction residuals ; (b) estimated glottal waveforms ; (c) estimated transfer-functions.

ペロープ（自己相関法，分析長 30 msec，ハニング窓）の一例を Fig. 10 に示す。

3．3 母音スペクトルの変動

以上のようにして得られた 2 つの定常フレームに対して小区間分析法を適用し，声道伝達特性と声帯波形との分離を行った。小区間分析法（分析次数 12 次，分析長 24 サンプル）は，音声信号を差分処理したのち，分析位置を 1 サンプル点ずつ移動させながら行い，残差波形に現われるメインピークの直後の分析区間で得られた声道伝達特性および声帯波形を最も正しいものとして採用した。ここで，excitation 波形は実際には未知のため，その近似として残差波形を用いている。

Fig. 11 は実音声の分析の一例で，(a) は予測残差波形，(b) は推定された声帯波形，(c) は推定された声道伝達特性である。(a) の予測残差波形の左端の点は分析開始点と一致させてあるので，この残差波形に現われるピークの直後から分析を開始している○印のものを最良のものとして採用した。以上のような処理をリファレンス及びテスト音声に対して施し，得られた声道伝達

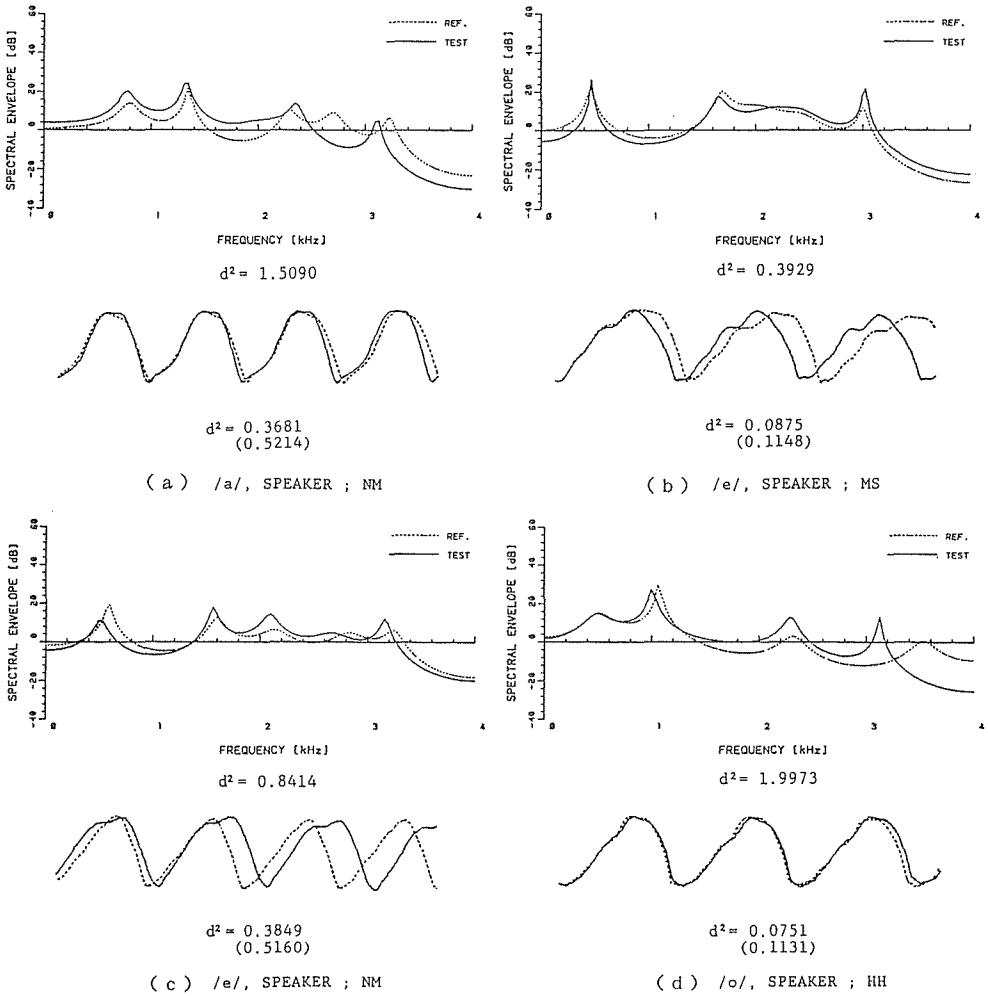

Fig. 12 Examples of estimated vocal tract transfer-functions and glottal waveforms.

特性と声帯波形について比較，検討を行った。

Fig. 12 に，特に異時期の音声サンプルの分析結果について数例を示す。図で上段が推定された声道伝達特性，下段が推定された声帯波形でともに点線がリファレンス音声，実線がテスト音声を表している。d^2 はリファレンス音声とテスト音声の推定された声道伝達特性間，及び推定された声帯波形のスペクトルエンベロープ間の差の 2 乗平均に相当する量である。声帯波形の d^2 で( ) 内は 2 回差分波形（予測残差波形）における値を表している。Fig. 12 (a) (c) は声道伝達特性，声帯波形ともに比較的変動の大きな例で，(b) はともに比較的変動の小さい例である。(d) は声道伝達特性の変動が大きく，声帯波形の変動が小さい例である。

3．4　結果のまとめ

以上の変動の結果を TABLE I に示す。表は特性変化量の大小を，d^2 の値を参考にして視察により分類したもので，同一時期・異時期発声のいずれの母音音声に対しても声道伝達特性が変動

TABLE 1 SPECTRAL VARIATIONS OF VOWEL SEGMENTS OF UTTERANCE OF EACH SPEAKER.

		Utterances at the Same Time		Utterances More Than 1 Week After	
		Number of Samples	d^2	Number of Samples	d^2
Transfer-function	Varied	5	Mean 0.415	22	Mean 0.945
	Varied Slightly	13	Max 1.022	4	Max 1.997
	Almost Same	6	Min 0.079	2	Min 0.393
Glottal Source	Varied	2	Mean 0.164 (0.160)	5	Mean 0.181 (0.249)
	Varied Slightly	5	Max 0.418 (0.515)	10	Max 0.537 (0.718)
	Almost Same	17	Min 0.009 (0.016)	13	Min 0.035 (0.049)

した音声サンプル数は声帯波形の変動したサンプル数より多く，また変動量を示す d^2 の平均値も大きいことが示されている。また，同一時期に発声された母音音声間の変動と異時期に発声された母音音声間の変動とを比較した場合，声道伝達特性が異時期間で大きく変動しているが，声帯波形はあまり変動していないことが示されている。これより音声の発声時期変動の主要因は声道伝達特性の変動であることが推察できる。

4．結　　論

各種合成音声による小区間分析法の検討の結果，音声信号に差分処理を施し，main excitation のかかっている時点直後の分析区間で分析を行うことにより最も良好な推定が行われることが明らかになった。

小区間分析法によって音声スペクトルの特性分離を行い，各特性の変動の様子を調査，検討した結果，時期的変動の主な原因は声道伝達特性の変動であるという結果が得られた。これは従来から言われていた時期的変動の主な原因は声帯音源スペクトルの変動であるという推定[3]とは相反する結果を与えるものである。今後はさらに小区間分析法の精度の明確化を図るとともに，音声サンプルを増やして，声道伝達特性の変動の様態についてさらに調査，検討を行い，話者照合等における誤り率改善への利用を図りたいと考えている。

参考文献

1) W. A. Hargreaves and J. A. Starkweather, Lang. & Speech, 6, (1963), 2, P. 63
2) J. E. Luck : JASA, 46, (1969), 4, P.1026
3) 古井：電子通信学会論文誌，57-A，(1974)，12，P. 880
4) 河原，栃内，永田：日本音響学会誌，33巻，(1977)．9号，P. 470
5) 村上，斎川，栃内，永田：日本音響学会講演論文集，(1985-8，10)，3-3-9
6) 村上，服部，斎川，栃内，永田：日本音響学会講演論文集，(1985-9，10)，3-4-3
7) 村上，斎川，栃内，永田：日本音響学会音声研究会資料，S 58-76 (1986-1)
8) 村上，栃内，永田：昭和59年度電気関係学会北海道支部連合大会講演論文集，207，P. 237
9) 柳田，木村，角所：電子通信学会論文誌，63-A，(1980-10)，10，P. 665
10) 小田，田中，前田，斎川，栃内，永田：北海道大学工学部研究報告，119号，(1984)，P. 127