Linear Prediction Analysis by Considering Amplitude Distribution of Speech

音声の振幅分布を考慮した線形予測分析

三 好 義 昭

Linear Prediction Analysis by Considering Amplitude Distribution of Speech

Yoshiaki MIYOSHI

1．まえがき

音声研究の長い歴史の中で，音声生成過程を構成している要因を定量的に記述するためのいろいろなモデルが提案されてきたが，その中でも音声の音響学的性質について最もよくできたモデルに，Fantの線形音声生成モデルがある[1]。このモデルは，喉から唇までの空間（これを声道と称する）が重要な役割を担っており，声道の形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，その声道伝達特性，特にその極周波数であるホルマント周波数を正確に推定することは音声認識を行なう上で非常に重要なことと言える。

このホルマント周波数推定手法として，今日，線形予測分析[2],[3]が広く用いられているが，この分析手法は，現時点の音声振幅値を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，その線形結合係数（線形予測係数と称する）を予測誤差の自乗平均最小の条件より求めるものである。したがって，従来の線形予測分析では線形予測係数を求めるのに，予測誤差の大きさにのみ注目していることに若干の問題があると言える。すなわち，予測誤差が零となるような線形予測係数が得られれば問題はないが，現実には予測誤差が零となるようなことはないので，予測の良さを予測誤差の大きさだけでなく，予測時点の波形の特性を総合的に考慮して評価する必要があると考える。例えば，たまにしか観測されない振幅値の比較的大きい部分での予測誤差はあまり重視しない，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等の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，音声波の振幅分布に基づき，線形予測分析における予測誤差の評価に重み付けする方法を提案し，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，振幅分布を基にして，線形予測分析の予測誤差の評価に重み付けする方法を定式化し，3．において，今回用いた重み付け手法を具体的に示す。そして，4．において，合成音のシミュレーションにより，本方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，本方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．振幅分布を考慮した線形予測分析

線形予測分析における，予測誤差の評価に重みを付けた場合の最小自乗解を求める線形予測分析の定式化を行う。

音声波の第n標本値をy_n，$n=1,2,3,\cdots,N$とし，y_nの予測値$\hat{y}_n$を

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

平成8年9月17日受理

とする。今，観測値と予測値との差（予測誤差 $\varepsilon_n = y_n - \hat{y}_n$）に重み付けした重み付き予測誤差 e_n を，

$$e_n = w_n (y_n - \hat{y}_n)$$

$$= w_n (y_n + \sum_{k=1}^{p} \alpha_k y_{n-k})$$

$$= w_n \sum_{k=0}^{p} \alpha_k y_{n-k} \qquad (2)$$

但し，$\alpha_0 = 1$，$w_n = f(y_n)$

とすると，この自乗平均は，

$$\overline{e_n^2} = \overline{(w_n \sum_{k=0}^{p} \alpha_k y_{n-k})^2}$$

$$= \frac{1}{N} \sum_{n=1}^{N} (w_n \sum_{k=0}^{p} \alpha_k y_{n-k})^2 \qquad (3)$$

となる。これは，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。したがって，重み付き予測誤差の自乗平均を最小にするような予測係数 $\{\alpha_k\}$ は，式(3)を各 α_k について偏微分した値を0，すなわち，

$$\frac{\partial \overline{e_n^2}}{\partial \alpha_j} = \frac{2}{N} \sum_{k=0}^{p} \alpha_k \sum_{n=1}^{N} w_n^2 y_{n-k} y_{n-j}$$

$$= \frac{2}{N} \{ \sum_{n=1}^{N} w_n^2 y_n y_{n-j} + \sum_{k=1}^{p} \alpha_k \sum_{n=1}^{N} w_n^2 (y_{n-k} y_{n-j}) \}$$

$$= 0$$

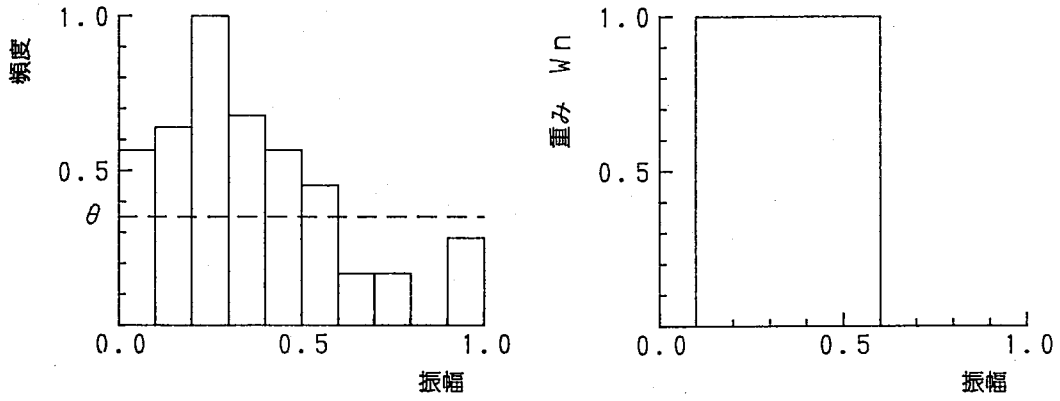
$$\therefore \sum_{k=1}^{p} \alpha_k \sum_{n=1}^{N} w_n^2 (y_{n-k} y_{n-j}) = - \sum_{n=1}^{N} w_n^2 y_n y_{n-j} \qquad (4)$$

なる連立1次方程式の解として得られる。そして従来の線形予測分析と同様，こ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]。

ここで，$w_n = 1$ とすると従来の線形予測分析と等しくなる。すなわち，従来の線形予測分析は，式(2)の w_n を1とした予測誤差 ε_n の自乗平均を最小とするような予測係数を求めるものである。したがって，従来の線形予測分析では，予測の良さを予測誤差の大きさのみで評価し，音声振幅値 y_n の大きさ，あるは音声波形の特性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，予測の良さを評価する際に予測誤差の大きさだけでなく，例えば，たまにしか観測されない振幅値の比較的大きい部分での予測誤差はあまり重視しない，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等の考慮が必要と考えられる。このような考慮を可能とする一方法として，本論文では，音声波の振幅分布に基づき，式(2)の w_n を設定する手法を提案し，その有効性を検討したものである。以下，式(2)の w_n を重み係数と称する。

3．　振幅分布に基づく重み係数の設定

図1（a）に合成母音／u／の振幅分布の例を示す。但し，音声波を絶対値の最大値で正規化し，それを10等分（以下，分割数と称する）した範囲のヒストグラムを求め，その最大値を1に正規化したものである。この振幅分布の例によると，振幅の大きさが0～0.6の比較的振幅値の小さいところの頻度が多く，振幅の大きさが0.7～1.0，つまり振幅値の大きいところの頻度が少なくなっている。この振幅分布に基づき，各々の予測誤差に適切な重みを付けることが可能であると考えられる。具体的には，しきい値 θ を設定し，頻度が θ 未満の振幅に対して $w_n = 0$ とし，頻度が θ 以上となる振幅に対しては，$w_n = 1$ とすれば良いと言える。すなわち，頻度の少ない振幅値が比較的大きい部分の予測誤差は，重視する必要がないと考え，その部分での予測誤差の評価を行わず，逆に，頻度の多い振幅値の小さいところの誤差を重視することができる。ところで，実際の音声では，振幅値の小さな部



（a）振幅分布　　（b）重み係数

図1　振幅分布と重み係数の例

分は雑音の影響をより強く受けていると言えるので，ここでは頻度がθ以上であっても，振幅値が0～1/k(但し，k：分割数)の範囲の音声波の重み係数を$w_n=0$とした。

図1（b）に重み係数の例を示す。但し，図1（a）の振幅分布に対して，しきい値$\theta=0.35$とした場合の例である。この例では，振幅値が0.1～0.6の音声波に対してのみ線形予測分析することになる。

4．合成音によるホルマント周波数推定精度の検討

合成音ではそれぞれのホルマント周波数が既知であるので，その既知のホルマント周波数と推定ホルマント周波数との誤差Eを式(5)で定義し，ホルマント周波数推定精度の観点より本方法の有効性を検討する。

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

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：第iホルマント周波数

以下，本方法と従来の方法による合成母音のホルマント周波数推定誤差の比較実験結果を示す。但し，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5母音を用い，分析条件はいずれも，分析次数p=12，分析窓長Ta=25.6ms，式(5)のM=20（0.4ms間隔で1ピッチ周期8.0msに渡り分析）であり，前処理として，一階差分を用いた。

図2に合成母音/a/のホルマント周波数推定誤差の分割数依存性を示す。但し，図中○印：本方法において，しきい値$\theta=0.04$とした場合の誤差，×印：従来の方法による誤差である。

図2より，合成母音/a/の場合，従来の方法ではホルマント周波数推定誤差が3.19Hzに対して，分割数6のとき0.02Hzと誤差がほとんど零で高精度にホルマント周波数が推定できることが分かる。合成母音の場合，いわゆる声門閉止区間分析を行えば，ホルマント周波数が誤差

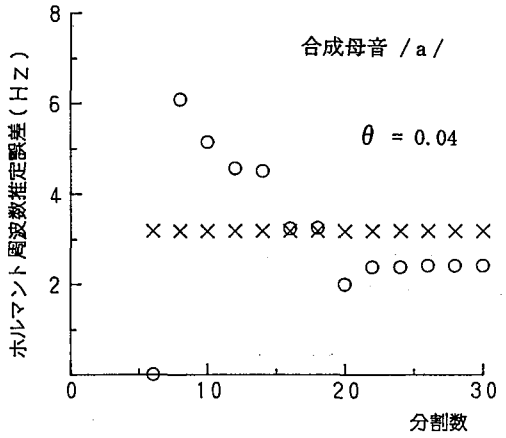


図2　ホルマント周波数推定誤差の
分割数依存性

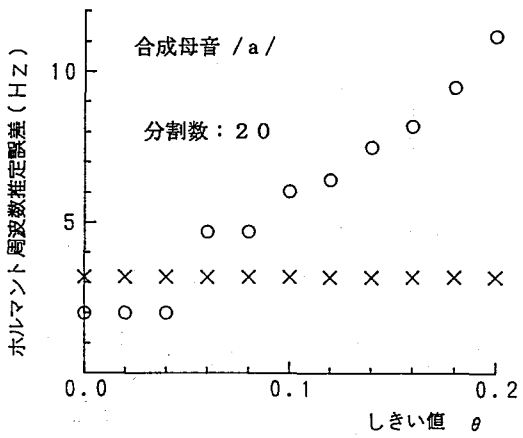


図3　ホルマント周波数推定誤差の
しきい値依存性

なしで推定可能であり[6]~[9]，図2の結果より，本方法の分割数及びしきい値θを適切に設定すれば(今の場合，分割数：6，θ=0.04)，実効的に声門閉止区間分析が可能と言える。一方，分割数が8～18では，本方法の誤差が従来の方法の誤差より大きくなり，推定精度が悪くなるが，分割数が20～30では2.0Hz代の誤差となり，従来の方法より改善されていると言える。

図3に合成母音/a/のホルマント周波数推定誤差のしきい値依存性を示す。但し，図中○印：本方法において，分割数20とした場合の誤差，×印：従来の方法による誤差である。

図3より，従来の方法による誤差が3.19Hzに対して，本方法では$\theta \leqq 0.04$において2.00Hzとなり，ホルマント周波数推定精度が改善されるが，$\theta \geqq 0.06$において，本方法による誤差の方が従来の方法による誤差より大きくなっていることが分かる。この結果より，しきい値はθ=0.0～0.04が適切と考えられる。

表1に，合成5母音におけるホルマント周波数推定誤差を従来の線形予測分析と比較して示す。但し，分割数を2から50まで1間隔，しきい値θを0.0～0.4まで0.02間隔でそれぞれ変化させて得られた最小誤差を示したもので，その時の分割数及びしきい値θを表中に示す。

表1より，合成母音/a/～/o/すべてにおいて誤差が0.2Hz以下となり，特に，合成母音/i/では，本方法により声門閉止区間分析となり，本方法による誤差が零となっている。また，合成母音/e/では，従来の方法の誤差が4.89Hzに対して，本方法の誤差が0.18Hzとなり，かなり改善されていると言える。5母音平均では，従来の方法による誤差が3.45Hzに対し，本方法による誤差は0.06Hzと大幅に改善し，本方法の有効性が示されていると言える。

表1　ホルマント周波数推定誤差
－　合成5母音　－

母音	振幅分布を考慮した線形予測分析			従来の線形予測分析による誤差(Hz)
	分割数	しきい値	誤差(Hz)	
/a/	6	0.04	0.02	3.19
/i/	5	0.08	0.00	2.49
/u/	17	0.38	0.05	2.07
/e/	7	0.04	0.18	4.89
/o/	15	0.40	0.04	4.63

次に雑音を付加した合成母音/a/を用いて，本方法の有効性を検討する。図4に，合成母音/a/の場合のSN比依存性を示す。但し，分割数及びしきい値θを表1と同様の範囲で変化させた場合の各SN比における最小誤差を示したもので，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。なお，S/N＝∞は雑音がない場合である。

図4より，S/N＝10dBの場合，ホルマント周

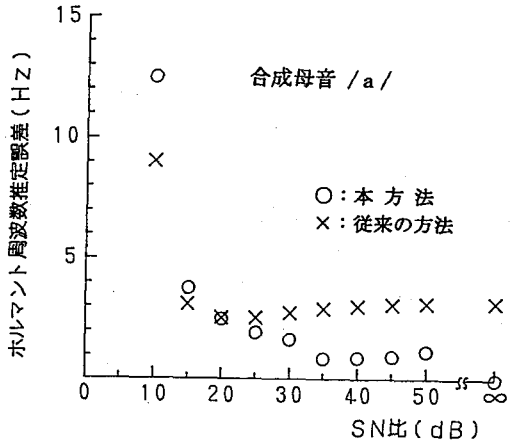


図4　ホルマント周波数推定誤差の
SN比依存性

波数推定誤差が従来の方法では9.04Hzに対して，本方法では12.50Hz，またS/N＝15dBの場合，従来の方法では3.12Hzに対して，本方法では3.79Hzと推定精度が本方法の方が悪くなっているが，S/N＝20dB～∞の実用的な広範囲の雑音領域において，従来の方法より小さい誤差でホルマント周波数が推定できることが分かる。

5．自然有声破裂音への適用例

自然音声では真のホルマント周波数が未知であるので，その誤差を定量的に評価できない。したがって以下，自然有声破裂音の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たホルマント空間での有声破裂音識別に適用することにより，本方法の有効性を検討する。具体的には，第2－第3ホルマント空間での識別率の良さでその最適性について評価し，また識別率の同じものについては分散比で評価した。ここで言う識別率とは，第2－第3ホルマント空間での/b/,/d/,/g/それぞれの重心からの距離で識別した場合に正しく識別できた率であり，分散比は，/b/,/d/,/g/　各々の類内分散と各重心の類間分散の比で，式(6)により定義する。

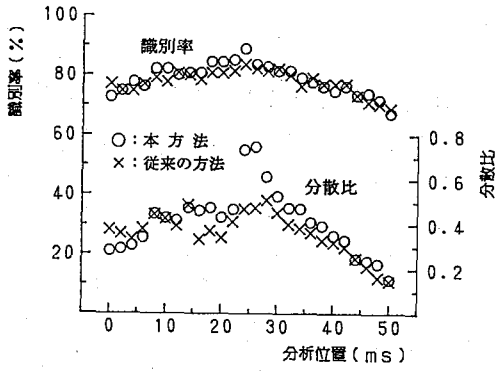


図5　識別率及び分散比の分析位置依存性
自然有声破裂音（後続母音/a/）

$$\text{分散比}=\frac{\frac{1}{3}\sum_{i=1}^{3}(G_i-G)^T(G_i-G)}{\frac{1}{3L}\sum_{i=1}^{3}\sum_{j=1}^{L}(X_{ij}-G_i)^T(X_{ij}-G_i)} \qquad (6)$$

$$G_i=\frac{1}{L}\sum_{j=1}^{L}X_{ij},\qquad G=\frac{1}{3}\sum_{i=1}^{3}G_i$$

但し，G_i：クラスiの重心ベクトル，X_{ij}：クラスiのj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L＝45となる。

図5に，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：本方法において，分割数12，しきい値θ＝0.02とした場合の識別率及び分散比，×印：従来の線形予測分析による識別率及び分散比で，破裂時点を時間原点として示す。

図5より，従来の方法で最も良いのは，分析位置24msのとき識別率83.7%，分散比0.469であったのが，本方法の場合，分析位置24msのとき識別率88.9%，分散比0.734と識別率，分散比とも向上し，本方法の有効性が伺える。また，後続母音/a/の自然有声破裂音識別を行うに

表2　第2－第3ホルマント空間における有声破裂音識別
－　後続母音/a/，分析位置：24ms　－

(a) 振幅分布を考慮した線形予測分析

	/b/	/d/	/g/	識別率
/b/	42	1	2	93.3 %
/d/		39	6	86.7 %
/g/		6	39	86.7 %
分散比 0.734		平　均		88.9 %

(b) 従来の線形予測分析

	/b/	/d/	/g/	識別率
/b/	40	2	3	88.9 %
/d/	1	37	7	82.2 %
/g/		9	36	80.0 %
分散比 0.469		平　均		83.7 %

は，分析位置は破裂時点後24ms付近が適切であると言える。

本方法ならびに従来の方法で最も高い識別率が得られる分析位置24msでの識別結果の内訳を表2に示す。表2より，有声破裂音/b/,/d/,/g/のいずれの識別率も，従来の線形予測分析より向上しており，特に/g/においては，従来の方法の場合80.0%に対して，本方法では86.7%と向上し，本方法の有効性が示されていると言える。

6．むすび

音声波の特性を考慮した線形予測分析の方法として，音声振幅分布に基づき予測誤差の評価に重み付けする手法を提案し，従来の線形予測分析と比較した。

本方法は，線形予測分析における予測の良さを評価する際に，音声波の振幅分布に注目すれば，たまにしか観測されない振幅値での予測誤差，ならびに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ことを考慮し，振幅の小さい部分での予測誤差を評価の対象外と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3.45Hzであるのに対し，本方法では0.06Hzと大幅に改善し，本方法により，いわゆる声門閉止区間分析が実効的に可能となることが明らかとなった。また破裂時点付近の第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による識別率が破裂時点後24msの分析位置で83.7%であったのが，本方法により88.9%に向上するとの結果が得られ，本方法の有効性が明かとなった。

なお，これらの結果は，本方法の分割数及びしきい値θを最適に設定した場合に得られるのであり，これらのパラメータ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).

(6) S.Chandra and W.C.Lin:" Experimental comparison between stationary and nonstationary formulations of linear prediction applied to voiced speech analysis",IEEE Trans. Acoust., Speech & Signal Process.,ASSP-22,pp.403-415(1974).

(7) K.Steiglitz and B.Dickinson: " The use of time-domain selection for improved linear predic-

tion",IEEE Trans.Acoust.,Speech & Signal Process.,ASSP-25,pp.34-39(1977).
(8)河原，栃内，永田："小区間の線形予測分析とその誤差評価"，日本音響学会誌，33，9，pp.470-479(1977-09).
(9)片桐，松井，牧野，城戸："高ピッチ音声に対する短区間線形予測分析の検討"，信学技報，EA80-31(1980-08).