音声言語情報処理 44－17
(2002. 12. 19)

ピッチ周期可変伸長処理と窓掛け処理による線形予測分析の精度改善

松浦 正樹† 深林 太計志‡

†静岡大学大学院理工学研究科 ‡静岡大学工学部 〒432-8561 静岡県浜松市城北 3-5-1
E-mail: †m-masaki@tdfuka9.eng.shizuoka.ac.jp, ‡tdtfuka@ipc.shizuoka.ac.jp

あらまし 本論文では，線形予測分析・合成によりピッチ周期を可変伸長処理し，ピッチ周期毎に窓を掛け波形を分離して，フレーム内の波形を一括線形予測分析することにより精度を改善する方法を提案した．このような処理を施した波形を分析することによって，個々の窓掛けピッチ周期波形はピッチ周期無限大の一部とみなせ，隣接するピッチ周期波形に影響を及ぼさない．これによりピッチ周期の影響を軽減でき，分析精度を改善できることを示す．また，自己相関係数から求めた残差のピッチパルスと残差を求めた線形予測係数とによる合成波形の振幅の増大が改善される．そして雑音付加音声において，提案法が従来の線形予測分析と比較してどの程度の雑音量まで耐性があるかを検討した．

キーワード 線形予測分析，ピッチ周期可変伸長処理，残差，'0'系列の付加，窓掛け処理による波形分離，耐雑音

Improvement of Accuracy of Linear Prediction Analysis by Processing of the Variable Extension of Pitch Period and Processing of Windowing

Masaki MATSUURA† and Takeshi FUKABAYASHI‡

† Graduate School of Science and Engineering, Shizuoka University
‡ Faculty of Engineering, Shizuoka University 3-5-1 Jouhoku, Hamamatsu-shi, Shizuoka, 432-8561 Japan
E-mail: †m-masaki@tdfuka9.eng.shizuoka.ac.jp, ‡tdtfuka@ipc.shizuoka.ac.jp

Abstract In this paper, we proposed a method to improve the accuracy of analysis of high pitch speech by processing of the variable extension of pitch period by linear prediction analysis and synthesis, and processing of windowing to separate waveform every pitch period, and then doing linear prediction analysis of these waveforms collectively. We show that by these processing the influence of the short pitch period is reduced and the analysis accuracy is improved. And we investigated the tolerance to noise of the proposed method.

Keyword Linear prediction analysis, Variable extension of pitch period, Residual of linear prediction, Addition of '0' sequences, Separation of waveform by windowing, Tolerance to noise

1. まえがき

線形予測分析[1]，[2]は音声の分析，合成，認識など様々な音声信号処理の分野で利用されている．それは，この手法が音声波形やそのスペクトルの性質を極めて少数のパラメータで効率的かつ正確に表現でき、しかもそのパラメータが比較的簡単な計算で求まるという利点による[3]，[4]．

しかしながら，高ピッチ音声の分析精度は，ピッチの影響を受けて良くない．ピッチ周波数が高い場合，推定されたホルマント周波数にずれが生じ，推定されたパワースペクトルは真のスペクトルに一致しない．そしてホルマント周波数の帯域幅が小さく推定されると，合成音の振幅が増大するという問題が起こる．そこで線形予測分析による高ピッチ音声の分析精度を改善する方法がいくつか研究されている[5]～[7]．しかしこれらは，処理が複雑で計算量が多くなったり[5]，自己相関法に比べ分析次数等の分析条件の制約が厳しい共分散法に属する解法であったり[6]，[7]で問題点も抱えている．比較的簡単な計算で実行できるという線形予測分析の利点を損なう恐れがある．2 回の繰り返しはあるが，比較的簡単な計算で実行できる自己相関法に基づく精度改善法として，文献[8]の方法がある．この方法においてはほぼピッチ周期毎に分離した線形予測残差を線形予測分析し，その線形予測係数を利用している．

ここでは，文献[8]の方法に準じるが，更なる精度改善を目的に，線形予測残差の線形予測係数を利用する代わりに，ピッチ周期を可変伸長処理[9]し，ピッチ周期毎に窓を掛け波形を分離して，フレーム内の波形を一括線形予測分析する新しい方法を提案した．線形予

測残差の振幅の変化を基に，ピッチ抽出を行わずに，線形予測残差のほぼ1ピッチ周期毎に，適当な位置で隣接する長さの異なる‘0’の複数個からなる‘0’系列を挿入することにより，系列を後ろに伸長する．このピッチ周期を可変伸長処理した線形予測残差を用いてピッチ周期毎に独立に合成した波形に窓を掛け，窓の後に線形予測分析次数以上のサンプル数の‘0’を付加して波形を分離する．この処理により線形予測分析に必要な分析次数までの遅れの自己相関係数は，ピッチ周期毎波形の自己相関係数の，フレーム内での和になる．個々の窓掛けピッチ周期波形はピッチ周期無限大の一部とみなせ，隣接するピッチ周期波形には影響を及ぼさない．これが精度改善のポイントであり，ピッチ周期が短いことによる悪影響を軽減できると考えられる．

本論文では2.で提案法のアルゴリズムについて述べ，3.と4.で合成音，実音声の実験により提案法の有効性を示し，5.で提案法の耐雑音性について検討した．

2. 分析法

図1に提案法の概略を示す．まず音声信号から分析長L_1を取り出しハミング窓を掛け，自己相関法で線形予測分析し，線形予測係数を求める．これに共振の帯域幅の過小推定の影響を避けるために，ホルマントのバンド幅拡大操作を施し，線形予測係数を補正する．

ハミング窓を掛けた信号をx_nとし，分析次数をp，線形予測係数を$\{a_i\}(i=1,2,...,p)$とすると，補正した線形予測係数$\{a_i'\}$は，

$$a_i' = a_i \times e^{-\pi BT} \qquad (1)$$

として求められる．ここでTはサンプリング周期，Bは線形予測係数の補正係数である．

そして，この補正した線形予測係数と音声信号から線形予測残差（以下，単に残差と呼ぶ）を求める．残差y_nは次式で求める．

$$y_n = 0, \quad n \le p$$

$$y_n = x_n + a_1' x_{n-1} + a_2' x_{n-2} + \cdots + a_p' x_{n-p},$$

$$p < n \le L_1 \qquad (2)$$

この残差を2次のオールパスフィルタに通し，オールパスフィルタ出力後の残差にピッチ周期可変伸長処理を行う．オールパスフィルタに通すのは，ピッチ周期毎に残差の絶対値振幅の最大値より前にある振幅の大きい成分の位相を遅らせることにより，この後の‘0’系列挿入を容易に行えるようにするためである．

ピッチ周期可変伸長処理は，残差のほぼ1ピッチ周期毎に適当な位置に，隣接する長さの異なる‘0’の複数個からなる‘0’系列を挿入して，隣接する周期を異なる長さに伸長し，合成波形をピッチ周期毎に独立に生成することによって行っている．合成波形x_i'はピッチ周期伸長処理残差u_nとこの残差を求めた補正線形予測係数からピッチ周期毎に独立に次式で生成する．

$$x_n' = 0, \quad n \le 0$$

$$x_n' = u_n - a_1' x_{n-1}' - a_2' x_{n-2}' - \cdots - a_p' x_{n-p}',$$

$$1 \le n \le L_u \qquad (3)$$

ここでL_uは‘0’系列の挿入により伸張されたピッチ周期の長さである．

最後に，図2に示すような窓掛け，分離処理を行い，そのフレーム内の波形を自己相関法で一括線形予測分析する．窓掛け処理は，合成波形の‘0’系列付加部分に対応する波形部分に行い，その後に分析次数に等しい長さの‘0’系列を付加して波形を分離する．このようにピッチ周期毎に波形を分離することにより，隣接するピッチ周期波形の影響を受けず，悪影響を緩和できる．

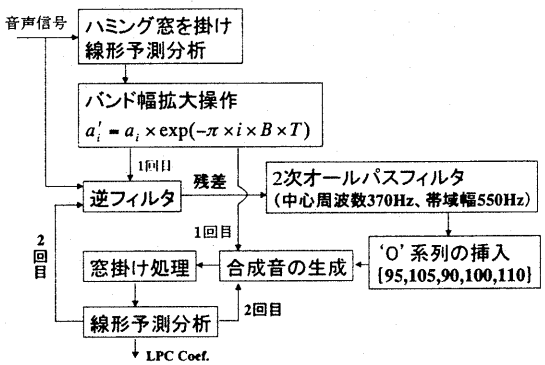

図1　提案法のアルゴリズム

Fig.1　Algorithm of proposed method.

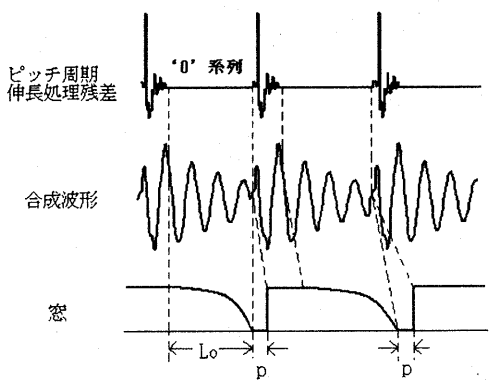

図2　窓掛け処理

Fig.2　Processing of windowing

使用した窓関数は窓を掛け始める位置を基準にして次式で表される．

$$w_i = 0.5 + 0.5\cos(i\pi / L_0), \quad 1 \le i \le L_0 \qquad (4)$$

ここでL_0は窓を掛ける部分の長さである．

この方法では，残差を求めるところからの繰り返しが何回でも可能であるが，3 回以上の繰り返しは合成音の実験で顕著な改善が見られなかったので，ここでは繰り返し 1 回と 2 回について検討した．

次のフレームの分析は，フレームシフト分シフトした波形を取り込みこの章の始めに述べた処理から繰り返す．

3. 合成音による実験

3.1 実験条件

合成音は声門開口比 0.65 のローゼンベルグ波[10]を入力として用い，サンプリング周期 0.1ms と想定して合成した．分析に用いた合成音/a/，/i/，･･･，/o/のホルマントは，表 1 に示す．放射特性は，差分特性$(1-z^{-1})$を用いた．

分析誤差はスペクトル包絡の分析誤差として次式で表す．フレームシフト 5ms で 10 フレーム分析を行い，その平均値で表す．

$$D = \sqrt{\sum_{M_1}^{M_2} \{\hat{S}(f_i) - S(f_i)\}^2 / (M_2 - M_1)} \ \text{(dB)} \qquad (5)$$

ここで，$\hat{S}(f_i)$は，推定された線形予測係数から計算したパワースペクトルの離散値を dB で表現したもので，FFT を用いて計算した．線形予測係数から計算したこのパワースペクトルを，ここではスペクトル包絡の推定値と呼ぶ．$S(f_i)$は，表 1 に示す合成音のホルマントと入力の 1 ピッチ周期分のローゼンベルグ波，放射特性から求めた真の（与えられた）スペクトル包絡である．$\hat{S}(f_i)$と$S(f_i)$は，$i = M_1$（＝5，約 78Hz 対応）とM_2（＝256，約 4980Hz に対応）の間で，平均値が 0dB になるように，それぞれ，正規化している．真のスペクトル包絡には，放射特性のために，0Hz に深いスペクトルの谷があるが，分析モデルは，全極形モデルであるために，このスペクトルの谷を良く近似しない．M_1＝5 としたのは，このような分析モデルが合致しないための誤差を除くためである．

また，第 1～第 3 ホルマント周波数推定誤差は，次式で表す．

$$E = \frac{1}{15}\sum_{j=1}^{5}\sum_{i=1}^{3} |\hat{F}_{ij} - F_{ij}| / F_{ij} \qquad (6)$$

ここでF_{ij}は，第j母音の第iホルマント周波数で，$\hat{F}_{ij}$は，第j母音の第iホルマント周波数の推定値である．

表 1　合成音のホルマント
Table.1　Formant of synthesized speech

/a/	周波数（Hz）	1160	1570	3090	4200
	帯域幅（Hz）	60	70	130	200
/i/	周波数（Hz）	340	2630	3480	4200
	帯域幅（Hz）	50	110	150	200
/u/	周波数（Hz）	340	1270	2750	4200
	帯域幅（Hz）	50	60	110	200
/e/	周波数（Hz）	500	2260	3130	4200
	帯域幅（Hz）	50	90	130	200
/o/	周波数（Hz）	580	910	3240	4200
	帯域幅（Hz）	50	60	140	200

表 2　実験パラメータ
Table.2　Experimental parameters.

サンプリング周期	0.1ms
分析長	30ms
フレームシフト幅	5ms
線形予測分析次数	10
オールパスフィルタ中心周波数	370Hz
オールパスフィルタ帯域幅	550Hz
‘0’系列	100 前後

本実験で使用したパラメータを表 2 に示す．これらは実験により最適値を求め，その値を使用した．

3.2 分析精度

図 3 にピッチ周期を 3.0～5.0ms まで変化させた場合のピッチ周期に対する分析誤差を示す．ピッチ周期毎に 5 母音の分析誤差を求め，その平均を表している．従来の線形予測分析の自己相関法を LP 法（以下これを従来法と呼ぶ）とし，提案法の繰り返し 1 回を新方法 1，繰り返し 2 回を新方法 2 としている．以下の図においてもこのように示す．図 3 から提案法により繰り返し 1 回，2 回とも誤差が減少し，分析精度が改善されたことがわかる．この分析精度は文献[8]の結果より良い（文献[8]の図 7 参照）．

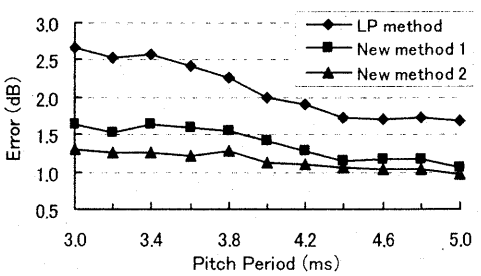

図 3　ピッチ周期に対する分析誤差
Fig.3　Analysis error on pitch period.

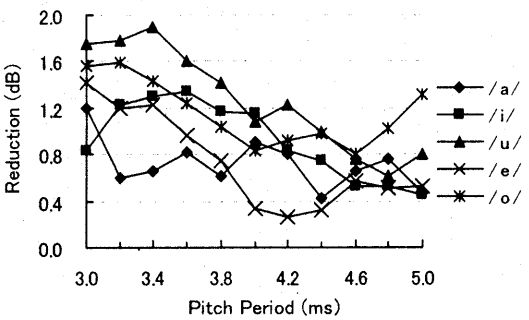

図 4　ピッチ周期に対する誤差の減少量

Fig.4　Reduction of error on pitch period.

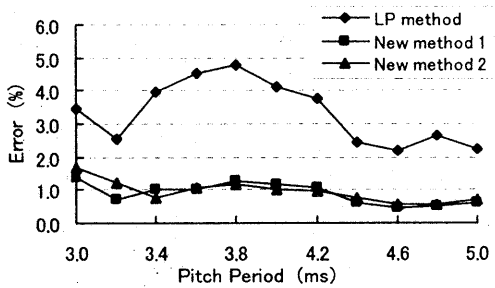

図 5　ピッチ周期に対するホルマント周波数推定誤差

Fig.5　Estimation error of formant frequency on pitch period.

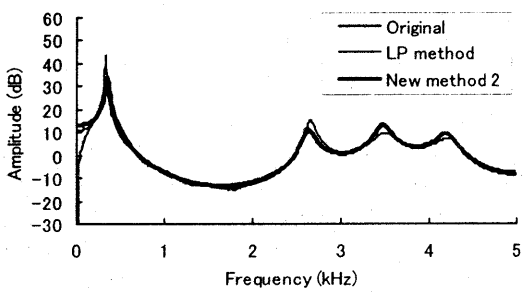

図 6　合成音/i/のスペクトル包絡の推定値

Fig.6　Estimated value of spectral envelope of synthesized speech /i/.

図 4 に各母音のピッチ周期に対する誤差の減少量を示す．ここで減少量とは，従来法による分析誤差と提案法の繰り返し 2 回による分析誤差の差である．どの母音も各ピッチ周期で提案法により誤差が減少したことがわかる．特にピッチ周期が短いときの減少量が多く，問題となっている高ピッチ音声に対する分析精度が改善されたと言える．高ピッチ音声では/u/の減少量が最も多く，ピッチ周期 3.0ms の合成音で約 1.8dB の改善がみられた．

また，図 5 にピッチ周期に対するホルマント周波数推定誤差を示す．ピッチ周期毎に 5 母音の推定誤差を求め，その平均を表している．ホルマント周波数の推定値は提案法により改善されるが，繰り返し 1 回と 2 回であまり差が無いことがわかる．この結果も文献[8]の結果より良い（文献[8]の図 8 参照）．

図 6 にピッチ周期 3.0ms の合成音/i/のスペクトル包絡の推定値を示す．比較のため基準スペクトル包絡を示してある．これは使用した合成音のホルマントから求めた基準のスペクトル包絡の推定値である．従来法では第 1 ホルマントの帯域幅が狭く推定されているが，提案法では基準スペクトルとほぼ等しくなり，帯域幅の過小推定が改善されていることがわかる．またホルマント周波数のずれも改善されていることが確認できる．図には示していないが，/u/でも第 1 ホルマントの帯域幅の過小推定が改善され，/a/，/e/，/o/では過大推定が改善された．

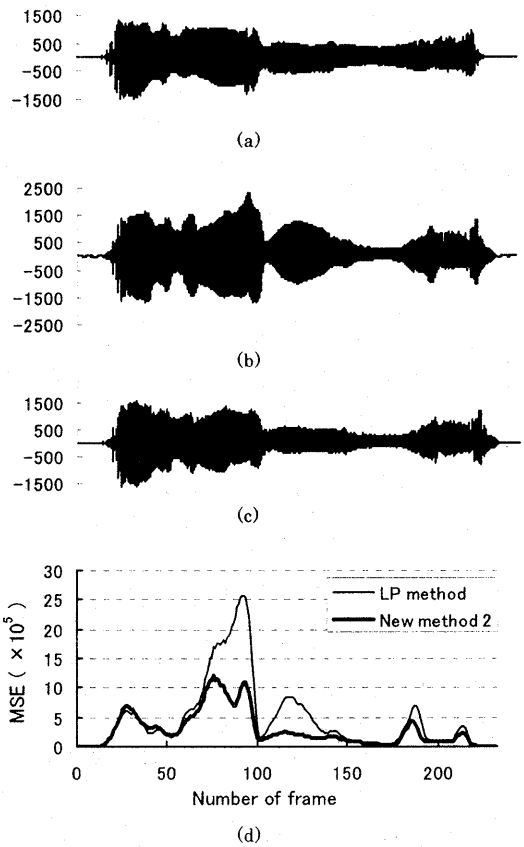

図 7　合成音声波形 (a)実音声，(b)従来法による波形，(c)新方法 2 による波形，(d)振幅の平均二乗誤差

Fig.7 Synthesized speech. (a) Original, (b) By LP method, (c) By new method 2, (d) MSE of amplitude.

4. 実音声の分析結果

図 7 の(a)に女声の/aoiue/という実音声，(b)，(c)にこの実音声を従来法と提案法により分析し，残差のピッチパルスと線形予測係数から合成した波形 1.2s をそれぞれ示す．フレームシフト幅は 5ms としている．ピッチ抽出は相関法を用いた．従来法では帯域幅の過小推定により，特に/o/から/i/（ピッチ周期約 3.6ms）の部分で振幅の増大が見られるが，提案法ではそれが改善されていることがわかる．

また，(d)にこの合成波形の分析長毎に次式で平均二乗誤差（MSE）F を求めたものを示す．

$$F = (1/L_1)\sum_{i=1}^{L_1}(\hat{s}_i - s_i)^2 \qquad (7)$$

ここで $\hat{s}_i$ は従来法及び提案法による線形予測係数により合成した波形のサンプル値で，s_i は実音声のサンプル値である．1.2s の実音声に対し，分析長 $L_1 = 300$，フレームシフト幅 5ms としているので，フレーム数は 235 である．この図から振幅の増大が改善されている部分が確認できる．

実際にこれらの 3 つの音声を聞いてみると，従来法の線形予測係数による合成音は声が篭もって聞こえ，提案法の線形予測係数による合成音は，より実音声に近いものとして聞こえる．

5. 提案法の耐雑音性

5.1 使用した雑音

実験に用いた白色雑音は，コンピュータで正規擬似乱数を生成して得たものである．有色雑音は，この白色雑音に次式で表される 1 次のフィルタを通して生成した高域で約－6dB/oct の傾きを持つ信号である．

$$H(z) = 1/(1 - e^{-\pi BT} z^{-1}) \qquad (8)$$

ここでサンプリング周期 $T = 0.1$ms，帯域幅 $B = 2000$Hz としている．

5.2 雑音付加合成音での実験結果

5.2.1 分析精度

図 8 に白色雑音を付加したピッチ周期 3.0ms の合成音の SNR に対する分析誤差を示す．5 母音の分析誤差の平均で表している．提案法の繰り返し 2 回を新方法として，従来の LP 法と比較している．ここで SNR 無限大というのは，雑音の無い合成音のことである．また，図には示していないが，有色雑音を付加した場合も同様な結果が得られた．

この図を見ると，SNR15dB 程度までは従来法より提案法の方が分析精度が良いが，それ以上の雑音が加わると精度は同程度であることがわかる．また，白色雑音と有色雑音では有色雑音を付加した場合の方が少し分析精度が良い．これは雑音スペクトルの違いによると考えられるが，それについては次の 5.2.2 で述べる．

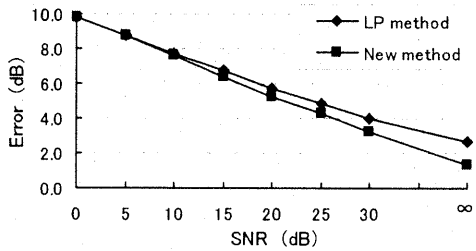

図 8　白色雑音付加合成音の分析誤差

Fig.8　Analysis error of synthesized speech with white noise.

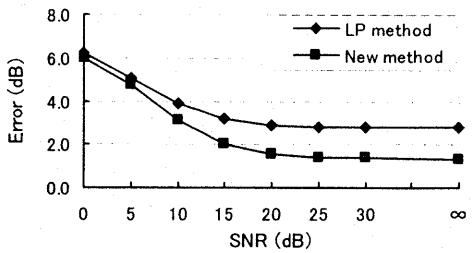

図 9　白色雑音付加合成音/e/の分析誤差

Fig.9　Analysis error of synthesized speech /e/ with white noise.

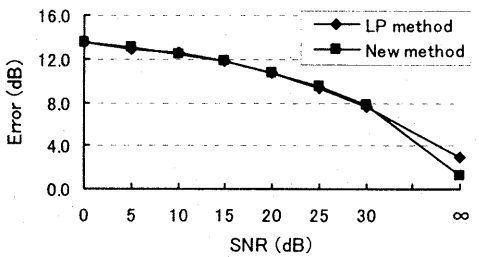

図 10　白色雑音付加合成音/u/の分析誤差

Fig.10　Analysis error of synthesized speech /u/ with white noise.

次に各母音別に分析結果を示す．図 9，図 10 はそれぞれ，白色雑音を付加したピッチ周期 3.0ms の合成音/e/，/u/についての SNR に対する分析誤差である．

図 9 から/e/の場合には，SNR10dB 程度まで提案法の方が分析精度が良く，それ以下では精度は同程度であることがわかる．/a/の場合もこの/e/の場合と同様な結果が得られた．

これに対して，図 10 から/u/の場合には，少しでも雑音が加わると分析誤差は大きくなり，雑音の影響を受けて，提案法は従来法と同程度の精度となってしま

うことがわかる．/o/の場合もこの/u/の場合と同様な結果が得られた．/i/の場合はSNR20dB程度までは提案法の方がわずかに分析精度が良いが，それ以上の雑音が加わると精度は同程度であった．この雑音の影響について具体的に次の節で考察する．

5.2.2 スペクトルの分析

ここではスペクトルを分析した結果を示し，雑音の影響について考察する．図11は雑音の無い合成音と，それにSNR15dBの白色雑音を付加した合成音のスペクトル表示と，雑音を付加した合成音での従来法と提案法によるスペクトル包絡の推定値を比較したものである．ピッチ周期3.0msの合成音/e/についての結果を表している．この図を見ると，雑音の無い合成音のスペクトルの調波成分は，どの帯域においても雑音のある合成音のスペクトルの調波成分に埋もれていないことがわかる．そして提案法によるスペクトル包絡は従来法によるものとは異なり，ホルマント周波数のずれ，第1ホルマントの帯域幅の過大推定が改善されている．

これに対して図12は/u/についての同様な結果であるが，/u/の場合は高域の振幅の小さいスペクトルの調波成分が雑音を付加した合成音のスペクトルの調波成分に完全に埋もれてしまっている．これが分析精度に影響を与えていると考えられ，提案法によるスペクトル包絡は従来法によるものとほぼ等しく，精度の改善が見られない．

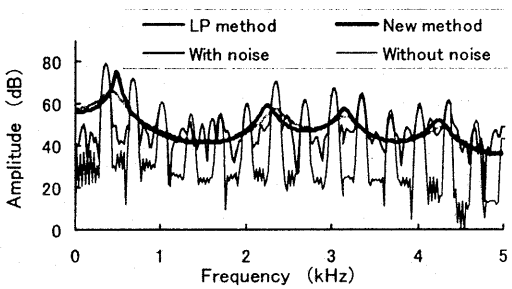

図11　白色雑音付加合成音/e/のスペクトル分析

Fig.11　Analysis of spectrum of synthesized speech /e/ with white noise.

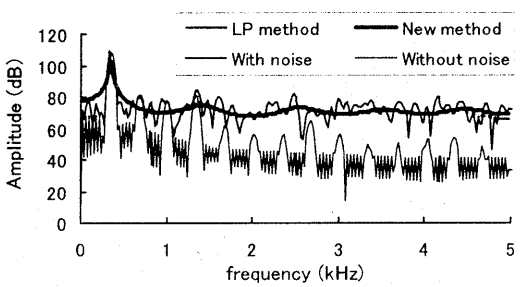

図12　白色雑音付加合成音/u/のスペクトル分析

Fig.12　Analysis of spectrum of synthesized speech /u/ with white noise.

以上のことから，スペクトルの調波成分が雑音付加成分に埋もれているかいないかが分析精度に影響を与えていると考えられる．また，これらは白色雑音を付加した合成音についての結果であるが，有色雑音を付加した合成音についても同様のことが言える．そして，有色雑音は高域のスペクトルが白色雑音よりも小さいので，有色雑音を付加した合成音の方が少し良い結果が得られる．

6. むすび

線形予測分析の精度改善について，残差に‘0’系列を挿入することでピッチ周期を伸長し，ピッチ周期毎に独立に波形を合成し，窓を掛けて分析する3回の線形予測分析を利用する方法を提案した．この方法により提案法は合成音，実音声の実験で分析精度が改善できることを示した．また，雑音付加合成音の実験で，提案法は，どの周波数帯においてもスペクトルの調波成分が雑音付加成分に埋もれていない場合には従来法より分析精度が良く，雑音が多くなり調波成分が雑音付加成分に埋もれる場合は分析精度は従来法と変わらないことを示した．具体的には，ピッチ周期3.0msの/a/，/e/ではSNR10dB程度，/i/では20dBまでの耐性があり，/u/，/o/では従来の線形予測分析と精度は同程度である．

文献

[1] 板倉文忠，斎藤収三，“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)，vol.53－A，no.1，pp.35－42，Jan. 1971.

[2] B.S. Atal and S.L. Hanauer，“Speech analysis and synthesis by linear prediction of speech wave，” J. Acoust. Soc. Am.，vol.50，no.2，pp.637－644，Aug. 1971.

[3] 古井貞熙，ディジタル音声処理，東海大学出版会，東京，p.60，1995.

[4] 守谷健広，音声符号化，電子情報通信学会，東京，p.7，1998.

[5] 佐宗晃，田中和世，“HMMによる音源のモデリングと高基本周波数に頑強な声道特性の抽出，”信学論(D－II)，vol.J84－D－II，no.9，Sep. 2001.

[6] 三好義昭，大和一晴，柳田益造，角所収，“2段標本選択線形予測法による高ピッチ音声分析，”信学論(A)，vol.J70－A，no.8，pp.1146－1156，Aug.1987.

[7] 有馬由紀，島村徹也，“システム同定法を用いた雑音にロバストな音声分析 ，”信学論(A)，vol.J83－A，no.12，pp.1455－1466，Dec. 2000.

[8] 深林太計志，“線形予測分析の残差を利用した高ピッチ音声の分析精度改善，”信学論（A），Vol.J75－A，No.3，pp.474－482，March 1992.

[9] 深林太計志，野田明弘，近藤大芝，“ピッチ周期可変伸長処理によるブロック適応ラティス型分析法の精度改善，”音響講論集，pp.233－234，March 2001.

[10] Rosenberg A. E.，“Effect of glottal pulse shape on the quality natural vowels，” J. Acoust. Soc. Am.，vol.49，no.2，pp.583－590，Feb. 1971.