音声合成におけるPARCOR係数とLPCケプストラムの次数

北村 正・寒川賢太・今井 聖*・チャッチャバリ・サラバリ*
電子工学科

（1983年9月3日受理）

Relation between Orders of PARCOR Parameters and LPC-cepstral Parameters in Speech Synthesis

Tadashi KITAMURA, Kenta SOKAWA, Satoshi IMAI*, Chatchavari SARAVARI*
Department of Electronics

(Roceived September 3, 1983)

This paper describes the relationship between orders of PARCOR parameters and LPC-cepstral parameters in speech synthesis. We used paired comparison experiments for subjective quality measurement and spectral distortion for objective quality measurement respectively. According to the resulting preference scores, it has been found that the necessary order of LPC-cepstral parameters in speech synthesis is about two times larger than that of PARCOR parameters. Furthermore it has been found that necessary order of PARCOR parameters are 14 for a male speaker, and 12 for a female speaker, and the necessary order of LPC-cepstral parameters is 25 in 10 kHz sampling frequency.

1. まえがき

コンピュータなどによる情報化の時代がますます進歩するに伴ない，人間と機械の間の情報交換の手段として音声が注目されるようになってきた。そのための技術として，音声合成，音声認識などがある。

このうちの音声合成は，録音編集方式，分析合成方式，法則合成方式などに大別されるが，前の2方式は最近のLSI技術，音声情報処理技術などの進歩により，実用化される例も多くみられるようになってきた。特に，計算機向けの音声分析合成法として，線形予測法[1]に基づく方法が良く知られており，その原理に基づくLSI音声合成器も製品化されている。この方法は，音声の生成モデルとして全極形モデルを用いており，線形予測係数PARCOR係数[2]，LPCケプストラム係数などの音声の特徴パラメータが得られる。これらのパラメータは音声合成や音声認識において広く用いられている。

音声合成において，出力しようとする内容が長くなったり，任意の音声を発生させようとするときは，前述の3方式のうち法則合成方式を用いる必要が出てくる。この方式で，LPC法によるPARCOR係数などでCV（子音－母音）連鎖，VCV（母音－子音－母音）連鎖などを音声の基本単位とした場合[3]，接続部において，必ずしも十分自然な合成音声が得られるとは限らない。音声の対数スペクトルのフーリエ係数で定義されるケプストラムは，音の大きさに対する人間の聴覚特性との対応が良く，CVを基本単位とした場合，接続部に不自然な音が発生することは少ない[4]などの特長をもっている。LPCケプストラムは上述のケプストラムとはモデルが異なるが同様の特長をもっている。音声合成におけるこのLPCケプストラムの性質はまだあまり調べられておらず，その性質を調べるのは有用と思われる。

ここでは，音声合成におけるPARCOR係数とLPCケプストラム係数の等価次数を合成音声の品質の主観的評価尺度である一対比較試験によって検討し，更に客観的評価尺度であるスペクトルひずみにより音声合成に必要な基準次数について検討を行っている。

2. PARCORとLPCケプストラムの等価次数

2. 1 PARCOR係数とLPCケプストラム係数

離散化された音声信号 s_n の過去のサンプル値 s_{n-k} の線形結合から，現在のサンプル値を予測するのが線形予測法であり，予測値 $\hat{s}_n$ は

* Tokoy Institute of Technology

$$\hat{s}_n = -\sum_{k=1}^{N} \alpha_k^{(N)} s_{n-k} \tag{1}$$

で与えられる。その予測誤差 e_n は

$$e_n = s_n - \hat{s}_n = s_n + \sum_{k=1}^{N} \alpha_k^{(N)} s_{n-k} \tag{2}$$

となる。予測係数 $\alpha_k^{(N)}$ は，予測誤差 e_n の2乗平均値を最小にすることに決定される。(2)式で与えられる予測誤差 e_n を入力とし，音声波形 s_n を出力とするようなシステムが音声の生成モデルとなり，その伝達性 $H(z)$ は

$$H(z) = \frac{1}{1+\sum_{k=1}^{N} \alpha_k^{(N)} z^{-k}} \tag{3}$$

のような全極モデルとなる。

この線形予測法により音声の分析合成系が構成できるが，係数 $\alpha_i^{(N)}$ の量子化誤差により合成フィルタが不安定となり発振しやすい。そこで線形予測法による分析合成系は，これを係数変換した PARCOR 係数 k_i を用いて行うことが多い。係数 $\alpha_i^{(n)}$ と k_i との間には次式の関係がある。

$$k_n = \alpha_n^{(n)} \tag{4}$$

$$\alpha_i^{(n+1)} = \alpha_i^{(n)} - k_{n+1} \cdot \alpha_{n+1-i}^{(n)} \tag{5}$$

この PARCOR 係数 k_i をフィルタパラメータとする格子型フィルタは安定となる。

(3)式の音声生成モデルから得られる LPC ケプストラム $c_m^{(N)}$ は

$$c_m^{(N)} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \ln |H(e^{jm\Omega})|^2 \, d\Omega \tag{6}$$

で定義される。ケプストラムは，音声の対数スペクトルの線形変換なので，音の大きさに対する人間の聴覚特性にほぼ合っており，対数パラメータの時間方向の補間によるスペクトル誤差が小さく，対数振幅近似（LMA）フィルタ[5]を用いることにより直接ケプストラムパラメータをフィルタパラメータとして音声合成ができ，法則合成や音声認識にも適しているという特徴をもっている。通常のケプストラムは，音声波形を DFT したスペクトルの対数の IDFT で定義されるが，LPC ケプストラムは DFT の操作なしに線形予測係数 $\alpha_i^{(N)}$ から次のような漸化式で容易に得られる。

$$c_m^{(N)} = -\alpha_m^{(N)} - \sum_{k=1}^{m-1} \frac{m-k}{m} c_{m-k}^{(N)} \alpha_k^{(N)} \tag{7}$$

但し，

$1 \leqq m \leqq M$, $\alpha_k^{(N)} = 0 \ (k > N)$, $c_0 = \ln |H(e^0)|^2$

である。

2. 2 PARCOR と LPC ケプストラムの等価次数

アナログの音声信号は，標本化周波数10 kHz，語長12ビットで AD 変換される。使用した音声資料は「南部では東の風」（男声1.4秒長），「明日は北の風」（女声1.2秒長）で，分析長25.6 ms，フレーム周期 5 ms で分析を行った。被験者は17名である。

音声合成の際，標本化周波数が 8 kHz の場合，女声に対して PARCOR 係数は 8 次以上にしても合成音声の品質向上が少ないという報告がある[6]。従って，PARCOR 係数と LPC ケプストラム等価次数を検討するために標本化周波数が10 kHz であることを考えて，PARCOR 係数の次数が 8，10，12次の 3 種，LPC ケプストラムの次数が15，20，25次の 3 種，計 6 種類の合成音声を作製した。それぞれの合成フィルタは PARCOR 係数をフィルタパラメータとする 2 乗算格子フィルタ，LPC ケプストラムをフィルタパラメータとする LMA フィルタを用いた。LPC ケプストラムは $N=12$ の $\alpha_i^{(N)}$ から(7)式により求めた。合成フィルタの音源は，有声音に対してはピッチ周期の間隔のパルス列，無声音に対しては M 系列を白色雑音源として発生している[6]。

ここでは，合成音声の品質の主観的評価尺度の一つである対比較試験を行い，6 種類の合成音声のプレファレンススコアを求めた。Fig. 1 にプレファレンススコアを示す。図から明らかなように，プレファレンススコアは，

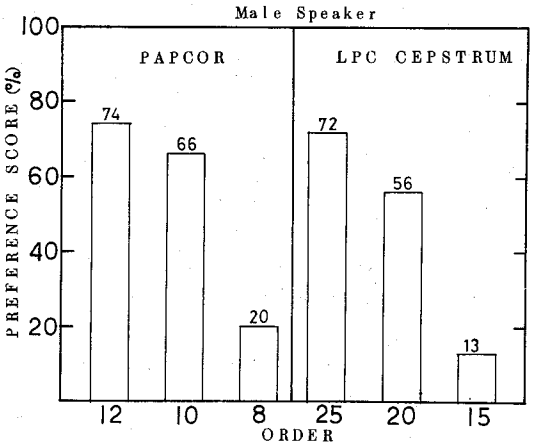

Fig. 1 (a) Preference scores of synthesized speech for a male speaker (synthesis orders are 8, 10, 12 for PARCOR and 15, 20, 25 for LPC-cepstral parameters).

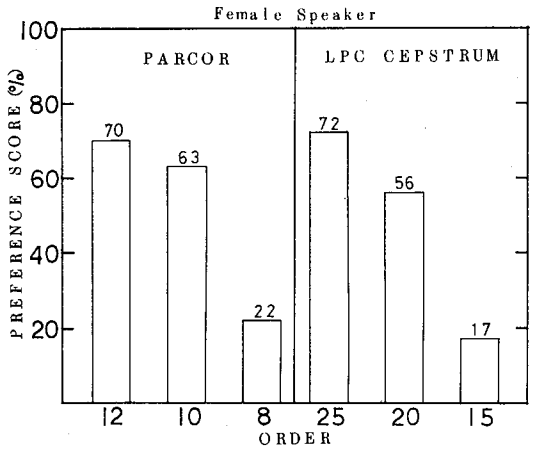

Fig. 1 (b) Preference scores of synthesized speech for a female sphaker (synthesis orders are 8, 10, 12 for PARCOR and 15, 20, 25 for LPC-cepstral parameters).

PARCOR-12 ≃ LPCCEP-25 ≳ PARCOR-10 ≳ LPCCEP-20 > PARCOR-8 ≳ LPCCEP-15という順序になっている。このことからPARCOR-12, 10, LPCCEP-25, 20の合成音声の品質差はあまりないことがわかる。又, LPCケプストラム係数はPARCOR係数のほぼ2倍程度の次数が必要であることがわかる。

Fig.2に男声の母音/a/の一部の音声波形と上述の6種類の合成音声のためのスペクトル包絡を示す。PARCOR-8, LPCCEP-15のスペクトル包絡は, 第1, 第2ホルマントもはっきりせず他の包絡とかなり異なっていることがわかる。

3. スペクトルひずみと次数

3. 1 スペクトルひずみ

合成音声の品質の客観的評価尺度としては, SN比, スペクトルひずみなどがある。ここで検討している音声

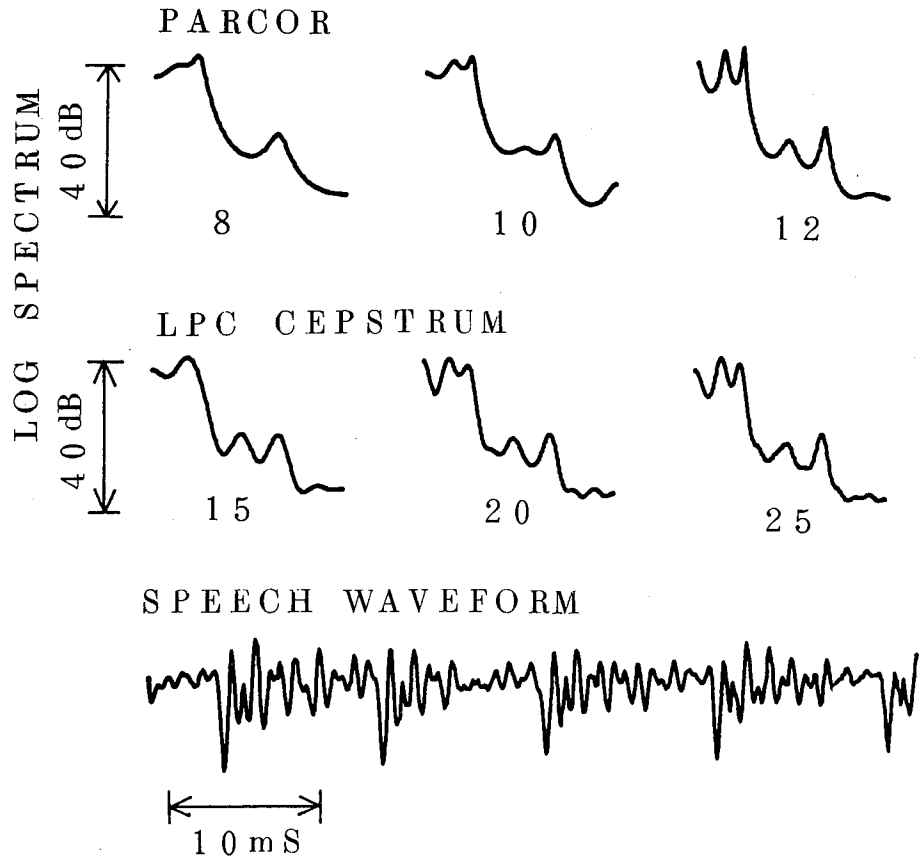

Fig. 2 Speech waveform of a portion of a vowel /a/ and its log spectral envelopes (analysis orders are 8, 10, 12 for PARCOR and 15, 20, 25 for LPC-cepstral paramethers).

の分析合成系は，音声の波形ではなく音声のスペクトル包絡情報を保存し伝送する方式なので，スペクトル領域での評価尺度であるスペクトルひずみが適当である[7]。スペクトル包絡特性のひずみ尺度として幾つかの尺度が考えられる[8]が，ここでは次式の対数スペクトル距離尺度 DS を考える。

$$DS^2=\frac{1}{L}\sum_{l=0}^{L-1}\frac{1}{N}\sum_{k=0}^{N-1}\left\{\ln|H^{(p)}(k,l)|^2-\ln|H^{(q)}(k,l)|^2\right\}^2 \quad (8)$$

ここで $\ln|H^{(p)}(k,l)|^2$ は，p 次の線形予測分析から得られる音声の対数スペクトル，k は周波数番号，l はフレーム番号を示す。又，Perseval の定理から，これはケプストラム距離に等しく

$$DS^2=\frac{1}{L}\sum_{l=0}^{L-1}\sum_{m=0}^{N-1}A^2\left\{c^{(p)}(m,l)-c^{(q)}(m,l)\right\}^2 \quad (9)$$

と書くこともできる。($A=10/\ln(10)$)

3. 2 スペクトルひずみと次数

前述の男女各一名の音声資料を用いて，(8)式での次数 p, q をそれぞれ 6 次から16次までの 6 種類に選んだときの PARCOR 係数のスペクトルひずみを Fig. 3 に，15次から35次までの 5 種類に選んだときの LPC ケプストラムのスペクトルひずみを Fig. 4 に示す。

PARCOR 係数：合成音声間のスペクトルひずみが 1 dB 以下になるとその品質差はほとんど感じることができない[7]ので，8 次対10次のように隣り合う次数間のスペクトルひずみに注目すると，スペクトルひずみが 1 dB 以下になるのは，男声で12次対14次以上，女声では10次対12次以上である。

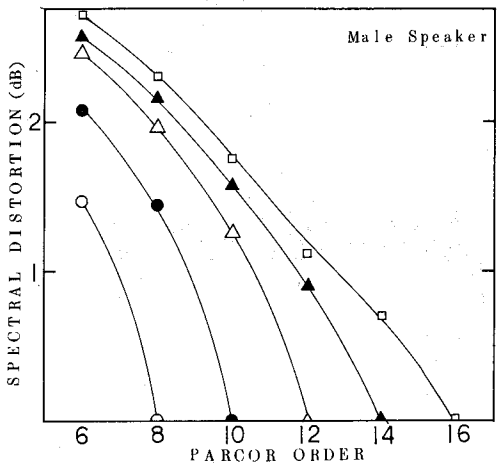

Fig. 3 (a) Relation between spectral distortion and orders of PARCOR parameters for a male speaker.

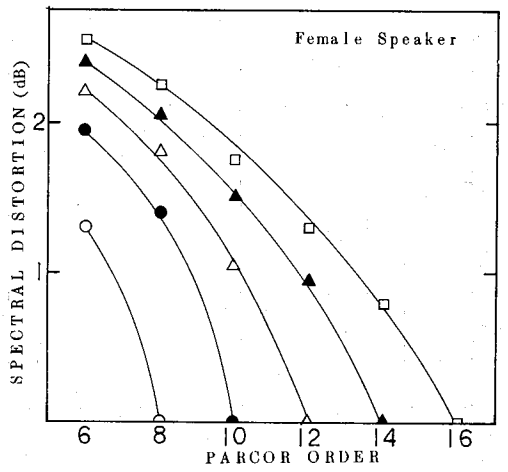

Fig. 3 (b) Relation between spectral distortion and orders of PARCOR parameters for a female speaker.

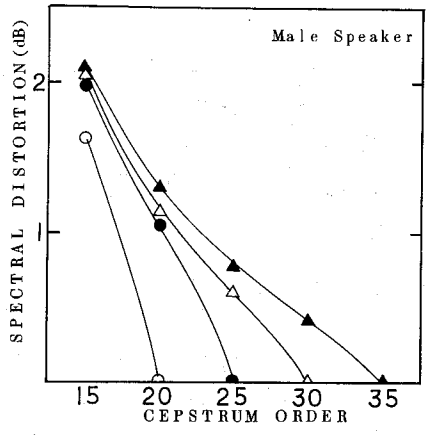

Fig. 4 (a) Relation between spectral distortion and orders of LPC-cepstral parameters for a male speaker.

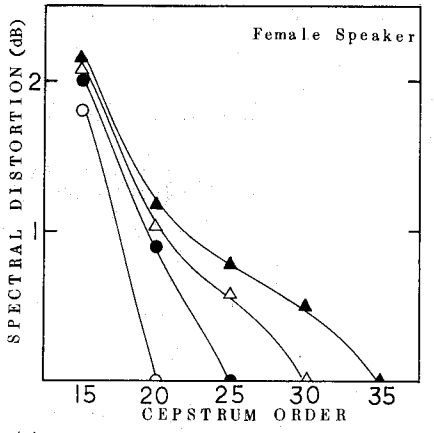

Fig. 4 (b) Relawlon between spectral distortion and orders of LPC-cepstral parameters for a female speaker.

LPC ケプストラム係数：上述の結果から，男声の場合14次，女声の場合12次の線形予測係数 $\alpha_i^{(p)}$ を基にして(7)式から LPC ケプストラム係数を求めた。PARCOR 係数の場合と同様に考えると，スペクトルひずみが 1 dB 以下となるのは男声女声共に，20次対25次以上である。

以上のことから，PARCOR 係数では男声14次，女声12次，LPC ケプストラム係数では25次以上あれば，合声音声の品質はほぼ飽和状態になると考えられる。従って，これらの次数を各係数の合成音声の基準次数と考えてもよいと思われる。又，PARCOR係数では男声12次，女声10次，LPC ケプストムでは20次までの合成音声の品質差は少なくよい品質と考えられるが，これは Fig. 1 のプレファレンススコアともほぼ合っている。

これらの結果からスペクトルひずみにより対比較試験によるプレファレンススコアはある程度予想できるものと考えられる。合成音声の品質評価のための主観評価実験は，被験者に対する高負担，訓練，バラつきなどの問題があるが，スペクトルひずみによりこれらの問題を減少できるものと考えられる。そのためには，更にスペクトルひずみとプレファレンススコアとの関係を詳細に検討する必要がある。

ここでの結果は，標本化周波数が10 kHz の場合であるが，他の標本化周波数の場合は上述の次数の値が変わると考えられる。例えば 8 kHz の場合，20％程度次数の値の減少が考えられる。

4．むすび

線形予測分析から得られる PARCOR 係数と LPC ケプストラム係数の等価次数と基準次数について検討を加えた。

合成音声の品質の主観的評価尺度として対比較試験によるプレファレンススコアを用いた。その結果，LPC ケプストラム係数は PARCOR 係数の約 2 倍の次数が音声合成においては必要であることが明らかとなった。又，客観的評価尺度として，対数スペクトル距離で与えられるスペクトルひずみを用いた。その結果，PARCOR 係数では，男声14次，女声12次，LPC ケプストラム係数では25次を基準次数とすればよいと考えられる。

今後は，合成音声の品質の主観的評価尺度であるプレファレンススコアと客観的評価尺度であるスペクトルひずみとの対応について検討していくことを考えている。

文　献

1）B.S.Atal and S.L.Hanauer：Speech Analysis and Synthesis by Linear Prediction of the Speech wave, J.Acoust. Soc. Amer., 50, 637（1971）

2）板倉，斉藤：偏自己相関係数による音声分析合成系，日本音響学会講論集，2－2－6（昭44－10）

3）佐藤：PARCOR-VCV 連鎖を用いた音声合成方式，信学論(A)，J61-D，11，pp.858－865（昭53－11）

4）阿部，今井：CV 音節のケプストラムパラメータからの音声合成，信学論(A)，J64-D，9，pp.861－868（昭56－09）

5）今井：対数振幅近似(LMA)フィルタ，信学論(A)，J63-A，12，pp.886－893（昭55－12）

6）今井，北村：対数振幅特性近似フィルタを用いた音声の分析合成系，信学論(A)，J61-A，6，pp.527－534（昭53－06）

7）北脇，板倉，斉藤：PARCOR 形音声分析合成系における最適符号構成，信学論(A)，J61-A，2，pp.119－126（昭53－02）

8）伊藤，北脇，筧：音声のディジタル波形符号化方式の客観的品質評価尺度の検討，信学論(A)，J66-A，3，pp.274－281（昭58－03）