2U-05　LSP-VCV 規則音声合成における子音残差信号パワーの符号化法[*1]

山口倫廣 清水忠昭 吉村宏紀 西田博充 井須尚紀 菅田一博[*2]
鳥取大学工学部知能情報工学科[*3]

1. はじめに

分析合成系を用いた規則音声合成方式において，予測残差を声道フィルタの駆動音源として用いる方式は，インパルス駆動方式に比べ肉声に近い音質が得られるため品質の向上が期待できる．しかし，残差駆動の規則音声合成への適用に際しては，ピッチ周期変更時に歪の発生を伴う，駆動音源の記憶容量が多い等の課題が残されている[1,2,3].

予測残差信号は,周波数スペクトルが白色化されている．従って，周波数スペクトル上では，符号化すべき情報は残っていない．特に，無声子音については，残差信号のパワーの時間変動だけが子音の特徴を持っていると考えられる．

本研究では，LSP-VCV 規則音声合成のために情報量の削減と合成音声の明瞭度の向上を目的として，残差信号のパワーの時間変動を符号化する手法を提案する．本手法により，比較的小規模なシステムで残差駆動に近い音質の規則音声合成を実現することが可能となる．

2. 本研究で提案する残差信号パワーの符号化手順

本研究で提案した符号化の手順を以下に示す．

(1) 音声資料を収集し，LSP 分析を行い，音声資料の LSP パラメータを求める．

(2) 音声信号と LSP パラメータを用いて予測残差信号を求め，子音の破裂点を中心として前後 128 点，計 256 点切り出す．

(3) 微少区間(16 点)毎に平均パワー $p_i(i=0,1,\cdots\cdots,255)$ を以下の式により求める．

$$p_i=\sqrt{\frac{1}{16}\sum_{k=0}^{15}s_{i+k}^2} \quad \cdots\cdots (1)$$

s_j : 切り出した予測残差$(j=0,1,\cdots,255)$

(4) 求まった平均パワー(256 点)の中から，音質的に重要な点だけを取り出し符号化する．

破裂子音は，破裂部分にパワーが集中しているため，音韻情報も破裂部分に多く含まれている．このため合成音声における破裂子音の明瞭度(主観的品質)は，破裂部分の再現の正確さに左右される．従って，残差信号の符号化法では，破裂点近傍のパワーを重点的に記憶し，残りの部分を直線補間してパワーを再生すれば，合成音声の品質を低下させることなく，情報量を削減できる．

本符号化法では，予測残差信号(256 点分)のパワー変動に対して破裂点(128 点)を中心に前後N 点，計$2N+1$ 点のパワーを記憶する．記憶するパワーは以下に示す関数により決定する．$x_i(i=0,1,\cdots,N)$,$y_j(j=0,1,\cdots,N)$ は記憶するパワーの離散時間軸上の位置を表している．

$$x_i=128-\alpha\left\{e^{\frac{1}{N}\log\left(\frac{128}{\alpha}+1\right)\cdot i}-1\right\} \quad \cdots\cdots (2)$$

$$(0\le x_i\le 128)$$

$$y_j=128+\alpha\left\{e^{\frac{1}{N}\log\left(\frac{128}{\alpha}+1\right)\cdot j}-1\right\} \quad \cdots\cdots (3)$$

$$(128\le y_j\le 255)$$

尚，破裂点(128 点),子音の開始点(0 点),終了点(255 点)のパワーは必ず記憶するように設定してある．また，符号化パラメータα は，$\alpha\ge 1$ なる定数で，値を小さくすれば破裂点近傍のパワーを集中的に記憶し，逆に値を大きくすれば等間隔に記憶する．

上式において，$N=4,\alpha=4$ とした場合の x_i, y_j の値を表 1 に示す．この場合 $p_0, p_{77}, p_{109}, p_{122}, p_{128}, p_{135}, p_{147}, p_{179}, p_{255}$ を記憶することになる．従って，記憶するパワーの総数は $2N+1$ で 9 点となる．

表 1．$N=4,\alpha=4$ とした場合の x_i, y_j

x_4	x_3	x_2	x_1	x_0,y_0	y_1	y_2	y_3	y_4
0	77	109	122	128	135	147	179	255

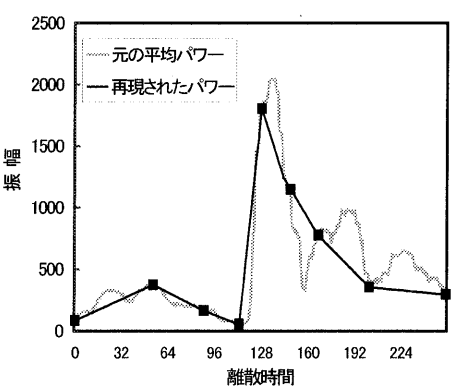

図 1．原残差信号の平均パワーと再現されたパワー
(例:/egi/ a=4,N=4)

*1 Coding method of residual signal for LSP-VCV speech synthesis

*2 Tomohiro Yamaguchi, Tadaaki Shimizu, Hiroki Yoshimura, Hiromitsu Nishida, NaokiIsu, Kazuhiro Sugata
*3 Tottori University,Factory of engineering,Department of information and knowledge engineering

3. 復号化と最適な符号化パラメータの決定

本研究では，提案した符号化手法により再現されたパワーにM系列信号の振幅変動を与え，作成した駆動音源と，LSPパラメータをもとに，音声合成を行った．原残差信号の平均パワーと再現されたパワーの一例を図1に示す．

記憶するパワーの数Nと符号化パラメータαを可変にし，原残差信号の平均パワーと再現されたパワーとの対数概形誤差を求めることにより，さまざまなNについて最適なαの値を実験的に決定した．図2に示すように，記憶するパワーの総数が9点($N=4$)の時は，$\alpha=30$，17点($N=8$)，25点($N=12$)，33点($N=16$)の時は$\alpha=60$で，対数概形誤差が最小となった．

4. 本手法により作成された合成音声の品質評価

本研究では，それぞれの記憶するパワーの数において対数概形誤差が最小となる符号化パラメータαの値を用いて合成音声を作成し，明瞭度評価実験を行うことで，合成音声の聴感上の評価を行った．

音声資料として，男性話者1名の発声により採取した破裂子音/b/,/d/,/g/,/p/,/t/,/k/,/c/を含むVCV素片計140種類を用いた．LSP分析の予測残差波形から生成した以下の6種類の駆動音源を用いて，得られた合成音声を音声資料とした．

A) 原残差信号
B) 原残差信号の平均パワー
C) 再現されたパワー($\alpha=60, N=16$)に
　M系列信号を付加したもの
D) 再現されたパワー($\alpha=60, N=8$)に
　M系列信号を付加したもの
E) 再現されたパワー($\alpha=30, N=4$)に
　M系列信号を付加したもの
F) M系列信号

被験者は，男性6名，女性4名とした．聴覚実験は，音声資料を1資料ずつランダムに提示し，聞こえた言葉(字数制限なし)を回答するという手法で行った．

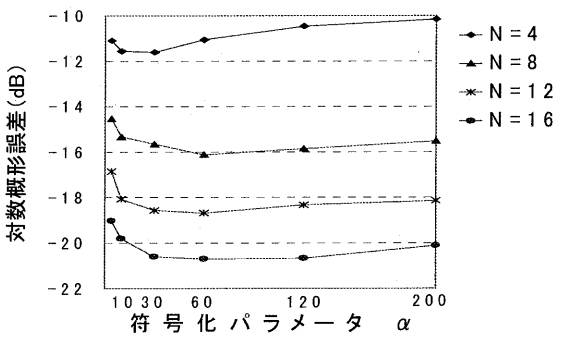

図2. 原残差信号の平均パワーと再現されたパワーとの対数概形誤差

5. 実験結果と考察

表2に，提示した素片別の誤聴率を示す．これは，提示した各音源・子音について，被験者が素片の子音部分を誤って別の子音と回答した割合である．表2より，各音源による誤聴率の傾向が，試料に含まれる子音によって異なることが分かった．

破裂子音/b/,/d/,/g/,/t/については，音源C)～E)のいずれかの誤聴率が音源F)の誤聴率より低く，提案した手法が合成音声の品質向上に有効であることが判明した．これは，これらの破裂子音の原残差の破裂点が顕著で，提案した手法の音源波形が，原残差の概形により近づくためだと考えられる．

破裂子音/k/,/c/については，全体的に誤聴率が低く，音源による差が見られなかった．従って，合成音声の品質改善の必要はない．

破裂子音/p/については，音源C)～音源E)のいずれの誤聴率も音源F)の誤聴率に差がなく，提案した手法で合成音声の品質改善ができなかった．その原因についてはわかっていないため，今後その原因を解明する必要がある．

6. おわりに

本研究では，VCV規則音声合成システムを前提に，破裂子音を含むVCV素片の予測残差信号から，破裂部分の特徴を抽出し，情報圧縮を図る残差符号化法を提案した．本符号化法により，破裂子音/b/,/d/,/g/,/t/について，比較的小規模なシステムで残差駆動による規則音声合成を実現でき，高品質な合成音声を得ることが可能となった．

【参考文献】

[1] 清水忠昭,吉村宏紀,西田博充,井須尚紀,菅田一博,"LSPベクトルVCV規則音声合成方式のための合成単位素片数と素片選択法",電学論,Vol.119-C,1060-1067,(1999).
[2] 西田博充,清水忠昭,吉村宏紀,井須尚紀,菅田一博",LPCケプストラム距離を用いたLSPベクトルVCV規則音声合成方式",情報・システムソサイエティ大会講演論文集,D-14-7,(1998).
[3] 清水忠昭,吉村宏紀,隅田庸市,井須尚紀,菅田一博,"LSPパラメータにベクトル量子化を適用した小規模応用のためのVCV規則音声合成", 電学論(C), Vol.120-C, No.3, 420-427(2000).

表2. 提示した素片別の誤聴率（単位：%）

子音	音源A (原残差)	音源B (パワー)	音源C (パワー) (N=16)	音源D (パワー) (N=8)	音源E (パワー) (N=4)	音源F (M系列)
/k/	2.0	2.8	2.4	3.6	2.8	3.2
/c/	2.0	0.5	2.5	1.0	2.5	1.0
/b/	13.2	13.6	14.4	14.2	13.6	16.0
/d/	7.0	7.66	8.66	8.66	11.66	11.33
/g/	5.6	5.2	6.0	5.4	4.8	7.2
/t/	3.33	4.33	3.66	5.0	5.33	8.33
/p/	7.8	20.2	22.8	18.6	21.2	20.4