ＶＣＶ規則音声合成の素片接続に影響する音韻環境の長さ

4Q−4

深坂淳一　木本雅也　清水忠昭　井須尚紀　菅田一博
鳥取大学工学部知能情報工学科

1. はじめに

規則音声合成法として，ＶＣＶ音声合成が現在よく用いられている．ＶＣＶ規則音声合成では，先ず文章(合成目的文)が持つＶＣＶ音韻連鎖素片(ＶＣＶ素片)を予め複数採取し，素片データベースを作成する．音声合成時には採取したＶＣＶ素片から，より自然で明瞭な音声が合成されるようにＶＣＶ素片を規則に従って選択する．

ＶＣＶ素片選択法の１つであるＰＥＲスコア最適選択法は，合成目的文中のＶＣＶ素片と素片データベース中のＶＣＶ素片との音韻環境一致度(Phonemic Environmental Resemblance Score：ＰＥＲスコア)が高くなるようにＶＣＶ素片を選択し，接続する手法である．ＰＥＲスコア最適選択法では，考慮する音韻環境の範囲が合成音声の品質と記憶容量の大きさに多大な影響を及ぼす．

本研究では，ＰＥＲスコア最適選択法における音韻環境の範囲と，合成音声の品質との関係について調べた．

2. ＰＥＲスコア最適選択法

ＰＥＲスコアを評価するためＶＣＶ素片を収集する際，ＶＣＶ素片の前後に音韻環境情報を付加しておく．ＰＥＲスコアは，素片データベース中のＶＣＶ素片の音韻環境と，合成目的文中のＶＣＶ素片の音韻環境との一致度を表す指標である．図１は，ＰＥＲスコア最適選択法における音韻環境一致度の評価の仕方を示す．

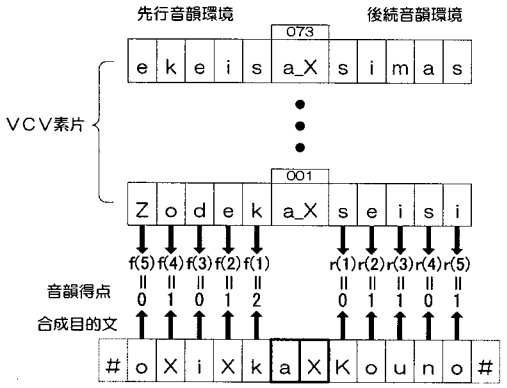

図1　ＰＥＲスコア最適選択法概念図

$f(i)$はＶＣＶ素片の先行するi番目の音韻について，$r(j)$はＶＣＶ素片の後続するj番目の音韻について，素片データベース中のＶＣＶ素片の音韻と合成目的文中の音韻との一致度を表す得点である．音韻得点は音韻が一致すれば２点，音韻種別(母音，摩擦子音等)が一致すれば１点，いずれにも該当しない場合には０点を与える．ＰＥＲスコアは次式で定義される．

$$PER=\frac{\sum_{i=1}^{F}\frac{1}{3^{(i-1)}}f(i)+\sum_{j=1}^{R}\frac{1}{3^{(j-1)}}r(j)}{\sum_{i=1}^{F}\frac{2}{3^{(i-1)}}+\sum_{j=1}^{R}\frac{2}{3^{(j-1)}}} \quad (1)$$

ＰＥＲスコアが最大となるようにＶＣＶ素片を選択する手法がＰＥＲスコア最適選択法である．式(1)で，Fは先行する音韻環境の長さ，Rは後続の音韻環境の長さである．ＰＥＲスコアは$f(i)$と$r(j)$をＶＣＶ素片の前後の音韻につ

Minimum Range of Phonemic Environment to Select VCV Instances for Speech Synthesis
Junichi Fukasaka, Masaya Kimoto, Tadaaki Shimizu, Naoki Isu, Kazuhiro Sugata
Dept. of Information and Knowledge Engineering, Tottori Univ., 4-101 Koyama-minami, Tottori 680, Japan

いて，時間経過を考慮した重み付きで合計し，ＶＣＶ素片の前後２つの音韻得点の最高点で除して正規化したものである．音韻得点の重みはＶＣＶ素片により近い音韻得点の差が，より遠い音韻得点の差によって逆転されないよう，３の指数の逆数としている．

3. 実験方法

合成音声の品質と音韻環境の長さとの相互関係を調べるため，合成目的文100文に対してＰＥＲスコア最適選択法で素片選択を行った．

我々は先の研究[1),2)]では，音韻環境の長さを先行5音韻,後続5音韻で実験を行いその結果,十分な合成音声が得られた．それ故，本研究ではＶＣＶ素片の前後の音韻環境の長さを 0～5の組み合わせ，35通りのＰＥＲスコア最適選択法で素片選択を行った(先行0，後続0ではランダム接続となるので省く)．この素片選択により，合成音声の評価値であるＬＳＰ距離とＰＥＲスコアを計算した．標準化のため，算出した２つの評価値を z - スコア[3)]に変換し，全ての選択結果についての平均値を計算した．

4. 実験結果

音韻環境の長さと各評価値との関係を示す図2，図3を得た．図のＦは先行，Ｒは後続する音韻環境の長さを表している．ＬＳＰ距離の平均値は先行1,後続0で値が大きく変化している．またＰＥＲスコアの平均値は先行0，後続0で値が極端に悪くなっている．以上の結果より，音韻環境は先行２音韻，後続音韻の長さがあれば合成音声の品質を落とさずに音声合成を行え，記憶容量も小さく出来ることが判った．

5. おわりに

本研究では，ＰＥＲスコア最適選択法における音韻環境の長さと合成音声の品質の関係について調べた．音韻環境の長さは先行２音韻，後続１音韻という少ない数でも高品質な合成音声が得られることが判った．これより，音韻は後続より先行の音韻の影響を，より大きく受けると考えられる．

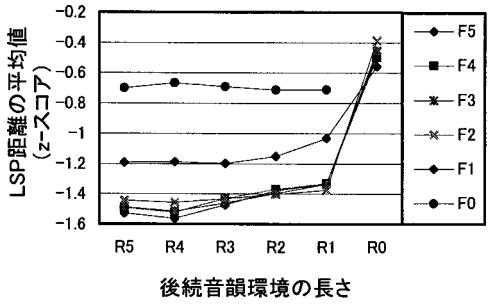

図2　音韻環境の長さとＬＳＰ距離の平均値

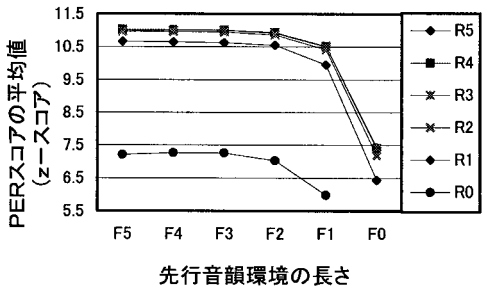

図3　音韻環境の長さとＰＥＲスコアの平均値

参考文献

1) 清水忠昭，吉村宏紀，木本雅也，並木寿枝，井須尚紀，菅田一博，“ＶＣＶ規則音声合成における音韻環境指標と接続歪み指標の関係”，電気学会論文誌(Ｃ)，Vol.121-C, No.3 (2001)

2) 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”，電気学会論文誌(Ｃ)，Vol.119-C, No.8/9, 1060-1067 (1999)

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”，電気学会論文誌(Ｃ)，Vol.120-C, No.3, 420-427 (2000)