

# VCV 規則音声合成における素片選択法の比較

2U-06

## ～LSP接続歪みと音韻環境得点の関係～

木本雅也 清水忠昭 吉村宏紀 並木寿枝 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

VCV規則音声合成において、合成音声の品質向上を目指し、様々なVCV素片の選択法が考案されてきた。我々は、VCV素片接続部の接続歪み(以下、LSP接続歪み)を最小化するLSP距離最小化選択法(Minimal LSP Distance method: MLD選択法)と、音韻環境得点(Phonemic Environmental Resemblance Score: PERスコア)を最大化するPERスコア最適選択法についての検討を行ってきた。<sup>1),2)</sup> 両選択法の間には、合成音声の品質に聴感上の差が無い事を先の研究で示した。<sup>1)</sup>

上記2つの選択法で接続したVCV系列について、LSP接続歪みとPERスコアの関係を実験的に調べた。その結果、LSP接続歪みとPERスコアの2つの選択基準の関係を定量的に明らかにした。

### 2. MLD選択法

VCV素片を選択する際、図1のようにVCV素片をノードとし、VCV素片間の接続部におけるLSPパラメータの距離をアークの重みとして、多数の経路から最短経路を選ぶ手法がMLD法である。接続部における先行VCV素片の最終フレームと、後続VCV素片の先頭フレームのLSPパラメータを各々、 $\omega' = (\omega'_1, \omega'_2, \dots, \omega'_p)$ ,  $\omega'' = (\omega''_1, \omega''_2, \dots, \omega''_p)$  とする時、LSP距離  $LD$  は以下の式で定義される( $p$ : 分析次数)。

$$LD = \sqrt{\sum_{i=1}^p (\omega'_i - \omega''_i)^2} \quad (1)$$

接続経路上での  $LD$  の平均をLSP接続歪みとする。LSP接続歪みが最小になるようにダイナミック・プログラミングの手法を用いて経路探索を行う。最短経路上のVCV素片の系列を最適なVCV系列として接続する。

### 3. PERスコア最適選択法

VCV素片収集時、前後5つずつの音韻環境情報を付加しておく。PERスコアとは、記憶しているVCV素片の音韻環境となるべく一致させるように音声合成するための指標である。以下の式で定義するPERスコアが最大となるVCV素片を選択する手法がPERスコア最適選択法である。

$$PER = \sum_{i=1}^5 \frac{1}{3^{(i-1)}} (f(i) + r(i)) \quad (2)$$

ここで、図2に示すように  $f(i)$  はVCV素片の先行する  $i$  番目の音韻について、VCV素片収集時の音韻と、合成する文(以下、合成目的文)の音韻の一致度を表す音韻得点である。 $f(i)$  には、音韻の一致で2点、音韻種別(母音、摩擦子音等)の一致で1点、そのいずれにも該当しない場合で0点を与える。 $r(i)$  は、VCV素片の後続する  $i$  番目の音韻について  $f(i)$  と同様な得点である。(2)式は、 $f(i)$  と  $r(i)$  をVCV素片の前後各5つの音韻について時間経過を考慮した重み付きで合計したものを  $PER$  としている。重みは、VCV素片のより近い音韻得点の差が、より遠い音韻得点によって逆転されないよう、3の指数の逆数としている。

### 4. LD-PER平面

新聞文書から、一息で発話できる文を1文とした合成目的文を100文抽出した。それらを上記2つの選択法で接続したVCV系列を得た。比較のため、最悪選択したVCV系列も求めた。得られたVCV系列から、LSP接続歪みと平均PERスコアを算出した。合成目的文の各々が各評価値での分布を持つため、100文について評価値を絶対値で比べても意味が無い。そこで、LSP接続歪みと平均PERスコアを  $z$ -スコアとして取り扱った。<sup>3)</sup>  $z$ -スコアとは、以下の(3)式のように平均を0、標準偏差を1に揃える事(標準化)で得られる無名数である。

$$z = \frac{x - \bar{x}}{s} \quad (3)$$

ここで、 $\bar{x}$  は平均、 $s$  は標準偏差である。

LSP接続歪みとPERスコアの関係を探るために、標準化したLSP接続歪みとPERスコアを軸とする2次元平面(以下、LD-PER平面)上で100文につ

Relation between LSP connective distortion and phonemic environmental resemblance score on LSP VCV Method  
Masaya Kimoto, Tadaaki Shimizu, Hiroki Yoshimura, Toshie Namiki, Naoki Isu, Kazuhiro Sugata  
Dept. of Information and Knowledge Engineering Tottori Univ., 4-101 Koyama-minami, Tottori, 680 Japan

いて選択結果の分布を調べた。また、最適値／最悪値分布におけるVCV系列長と、VCV辞書平均参照個数でクラス分けをして、分布上の関係を定量的に調べた。

### 5. 評価実験

実験の結果、図3に示すLD-PER平面での分布を得た。LSP距離最適選択ではPERスコアが準最適に、PERスコア最適選択ではLSP接続歪みが準最適に分布した。いずれの最適選択法も選択基準としない他方の値も最適に近くなっており、2つの選択基準が強い関係を持ち、合成音声はほぼ同品質である事がいえる。最適／最悪選択において、VCV系列長、及びVCV辞書平均参照個数に関係は見られなかった。これらは合成音声の品質に影響を与えないと考えられる。

### 6. おわりに

2種類のVCV素片選択法における各々の基準、LSP接続歪みとPERスコアの関係について調べた。合成目的文100文を最適選択した結果、LD-PER平面からLSP接続歪みとPERスコアには強い関係がある事が示された。この事から、MLD選択法とPERスコア最適選択法を用いて合成した音声に聴感上の差が無いという報告結果<sup>1)</sup>を定量的に裏付ける事ができた。

#### 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,” 電気学会論文誌(C), Vol.120-C, No.8/9, 1060-1067 (1999)
- 2) 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)
- 3) 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規則合成における接続歪みと音韻環境得点の関係に関する研究,” 信学技報, SP2000-5, 33-40 (2000)

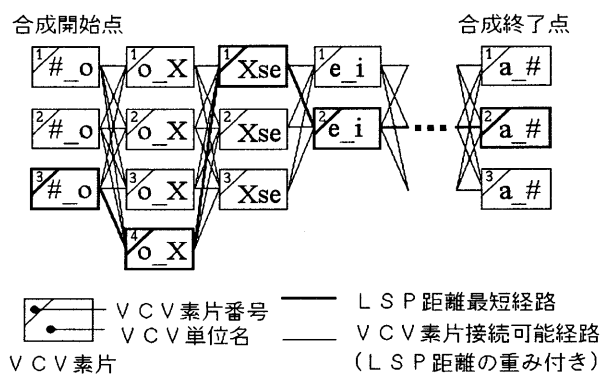


図1 MLD選択法概念図

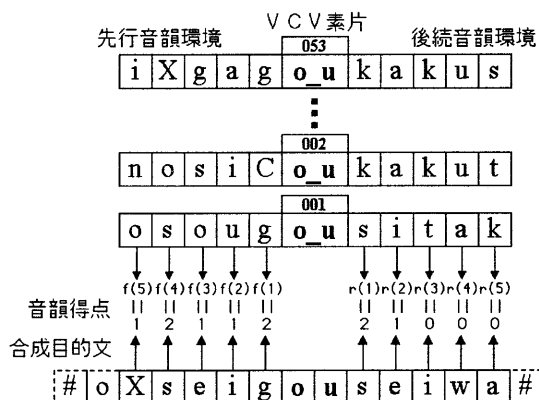


図2 PERスコア最適選択法概念図

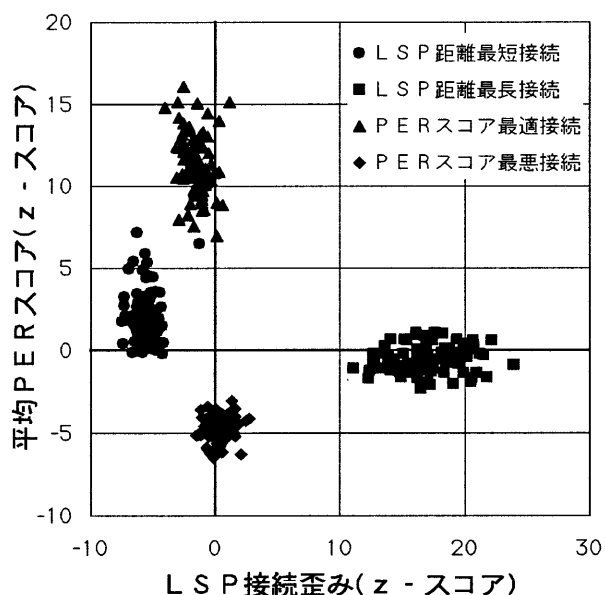


図3 LD-PER平面上の最適/最悪選択分布