

# 滑舌の良い Cross synthesis VOCODER

西 大 輝<sup>†</sup> 赤 桐 隼 人<sup>†</sup> 西 村 竜 一<sup>†</sup>  
入 野 俊 夫<sup>†</sup> 河 原 英 紀<sup>†</sup>

音声の狭帯域伝送技術として発明された VOCODER は、特有の品質劣化とクロス合成が簡単にできることを活かして、今では音楽におけるエフェクターとしての応用の方が有名になっている。ここでは、元の音声の品質をほとんど劣化させない分析合成を可能とする VOCODER の一種である TANDEM-STRAIGHT を用いた Cross synthesis VOCODER を提案する。会場では、Matlab を用いたプロトタイプとその GUI を紹介する。

## Making cross synthesis VOCODER intelligible

TAIKI NISHI,<sup>†</sup> HAYATO AKAGIRI,<sup>†</sup> RYUICHI NISIMURA,<sup>†</sup>  
TOSHIO IRINO<sup>†</sup> and HIDEKI KAWAHARA<sup>†</sup>

“VOCODER” is regarded as an effector for music applications nowadays, while it was originally invented as a technology to enable narrow-band telecommunications in 1930s. One of the reasons that promoted applications in music might be VOCODER’s unique quality/timbre degradations. Simplicity in concept and implementation of VOCODER-based cross synthesis might also be an important facilitating factor. We propose a new cross synthesis VOCODER based on TANDEM-STRAIGHT, which is a cutting edge VOCODER enabling close-to-the-original reproduction of speech sounds. A prototype system that is implemented using Matlab and its GUI tools will be demonstrated.

### 1. はじめに

“VOCODER” は、楽器の音に声を乗せるエフェクターとして、J-POP などでも広く用いられている。しかし、VOCODER<sup>1)</sup> は、本来は音声の狭帯域伝送のために発明された技術であり、特有の劣化やクロス合成の容易さが『キャラクター』として利用されている現状は、VOCODER にとって不本意に違いない。ここでは、VOCODER を現代的な技術として蘇らせた STRAIGHT<sup>2)</sup> の後継技術である TANDEM-STRAIGHT<sup>3)</sup> を用いた、『滑舌の良い』クロス合成を提案する。

### 2. 背景と狙い

クロス合成では、入力された音声のスペクトル情報をフィルタゲインの設定に用い、楽器の音をそのフィルタを用いて処理することにより、楽器の音に声を乗せるという効果が実現されている。Channel VOCODER<sup>1)</sup>

を応用した実装では、音声の分析に帯域フィルタ群を用い、楽器音の処理にも帯域フィルタ群が用いられる。乗せられる音声の明瞭度は、このフィルタの精度に依存する。本提案では、入力音声の分析に TANDEM-STRAIGHT を用い、求められたスペクトルにより設計されたフィルタを用いて楽器音を処理することで、明瞭度の高いクロス合成の実現を試みる。

### 3. TANDEM-STRAIGHT を用いた実装

以下では、まず TANDEM-STRAIGHT の概要を紹介し、それに基づいて、クロス合成にどのように応用し実装するかについて説明する。

#### 3.1 TANDEM-STRAIGHT

短時間 Fourier 変換 (STFT) を用いて求められるスペクトログラムには、分析対象の信号の周期性に起因する周期的な変動が、周波数方向と時間方向の双方に認められる。TANDEM-STRAIGHT では、窓関数の工夫により時間方向の変動を取り除き、基本周波数に適応した平滑化により周波数方向の変動を取り除いている。基本周波数に適応したこれらの処理により、調波位置での値を保存しかつ滑らかなスペクトルを求め

<sup>†</sup> 和歌山大学システム工学部

Faculty of Systems Engineering, Wakayama University

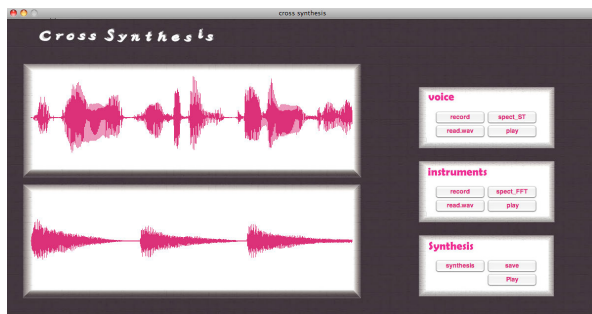


図 1 クロス合成 VOCODER 用の GUI  
Fig. 1 GUI for cross synthesis VOCODER

ることができる。

### 3.2 OLA によるフィルタ処理

こうして求められた滑らかなスペクトルを、FIR フィルタとして用いる。実際のフィルタ処理は、FFT による高速畳み込みを用いて実装した。具体的には、窓長 10 ms の Hanning 窓を用いて切出した楽器音を 5 ms 毎に処理し、OLA(overlap and add)<sup>4)</sup> によりクロス合成された信号の波形を作成する。フィルタ処理に用いるスペクトルは、フレーム毎に読み出され更新される。

なお、音声の長時間スペクトルは約 -6 dB/oct の傾斜を有しているため、そのままでは、楽器音の高域を大きく減衰させてしまう。ここでは、出力信号を差分処理することにより、この高域の低下を補償する。

## 4. プロトタイプの GUI

提案した方式の動作を確認するため、図 1 に示すインタフェースを持つプロトタイプを作成した。アルゴリズムの実装と GUI の作成には、Matlab が用いられている。この GUI では、以下の機能が提供される。

- 音声の録音、音声ファイルの読み込み、スペクトル (STRAIGHT) の計算と表示、再生
- 楽器音の録音、音声ファイルの読み込み、スペクトル (STFT) の計算と表示、再生
- クロス合成、再生、保存
- 音声波形の表示
- 楽器音の波形の表示

### 4.1 GUI の操作と予備検討

これらの機能は、GUI 上のボタンのクリックにより実行される。入出力にオーディオインタフェースを用いる場合、第一チャンネルに音声の入力が、第二チャンネルに楽器音の入力が割り当てられている。なお、読み込み可能なファイルは、WAVE 形式のもののみである。入力された信号波形は、直ちに GUI の左側に表示される。

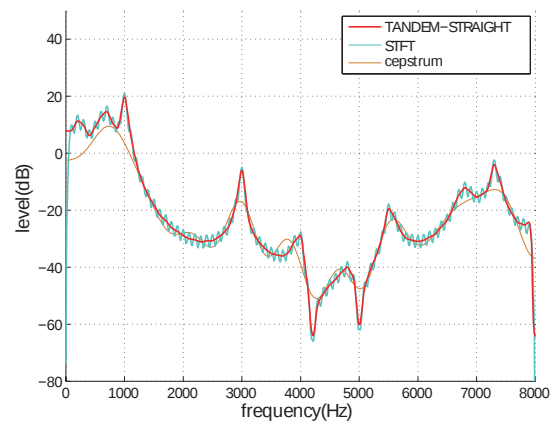


図 2 STFT, cepstrum 平滑化および STRAIGHT スペクトル  
Fig. 2 STFT, cepstrum-smoothed and STRAIGHT spectra

提案方法の効果を確認するため、音声のスペクトルとして STFT で求められるパワースペクトルをそのまま用いる場合、cepstrum による平滑化を行った場合の二種類の処理を比較対象とした予備実験を行った。図 2 に、これらにより求められた母音/a/のスペクトルの例を示す。用いる楽器音の種類によって効果は異なるが、提案法により処理された楽器音に寄せられた音声は、最も明瞭に感じられる。

## 5. ま と め

音声の分析に TANDEM-STRAIGHT を用いた『滑舌の良い』クロス合成を提案し、プロトタイプを実装した。今後は、提案法の高速化、用いる楽器音と処理音声の明瞭度の関係の検討を進める予定である。

## 参 考 文 献

- 1) H.Dudley: Remaking speech, *J. Acoust. Soc. Am.*, Vol.11, No.2, pp.169-177 (1939).
- 2) H.Kawahara, I.Masuda, and A.deCheveigné: Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction, *Speech Communication*, Vol.27, No.3-4, pp.187-207 (1999).
- 3) H. Kawahara, M. Morise, T. Takahashi, R. Nisimura, T. Irino, H. Banno: Tandem-STRAIGHT: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, F0, and aperiodicity estimation, *Proc. ICASSP 2008*, Las Vegas, pp.3933-3936 (2008).
- 4) A.V.Oppenheim, R.W.Shafer: *Discrete Time Signal Processing*, Prentice Hall (1998), (旧版の訳は、伊達玄訳: デジタル信号処理 (上), コロナ社 (1978)) .