

高品質音声分析変換合成法 STRAIGHT の 楽音への応用：最初の一步

河原英紀^{1,2}、片寄晴弘^{1,3}

¹和歌山大学/CREST

²ATR 人間情報通信研究所 ³イメージ情報研究所

¹〒 640-8510 和歌山市栄谷 930

kawahara@sys.wakayama-u.ac.jp

あらまし 本デモンストレーションでは、筆者らが提案した音声分析・変換・合成方法 STRAIGHT(Speech/sound Transformation and Representation using Adaptive Interpolation of weiGHTed spectrogram) を楽器音の変換に用いた場合の例を示す。尺八は、非常に生々しく再現されており、本方式が音楽の分野への応用においても高い潜在能力を持つことが示唆された。しかし、ピアノ音などでは音源情報のモデル化と抽出方法に更に工夫が必要であることが明らかとなった。

キーワード 時間周波数特性、基本周波数、時間揺らぎ、スペクトル変換、楽器

Musical instruments manipulation using a high quality sound analysis-modification-synthesis method, STRAIGHT: A first step toward music applications

Hideki Kawahara^{1,2} and Haruhiro Katayose^{1,3}

¹Wakayama University/CREST, ²ATR Human Information Processing Research Laboratories,

³Laboratory of Image Information Science and Technology

¹930 Sakaedani, Wakayama, Wakayama 640-8510

kawahara@sys.wakayama-u.ac.jp

Abstract A preliminary test to apply a new sound manipulation procedure STRAIGHT (Speech/sound Transformation and Representation using Adaptive Interpolation of weiGHTed spectrogram) is demonstrated. Vivid reproduction and manipulation of SYAKUHACHI (bamboo flute) indicated potential usefulness of the proposed procedure in music applications. However, it was also suggested that more sophisticated models of excitation source and pitch extraction, especially for piano sounds are crucially important.

key words Time-frequency Representation, F0, Jitter, Spectral Transformation, Instruments

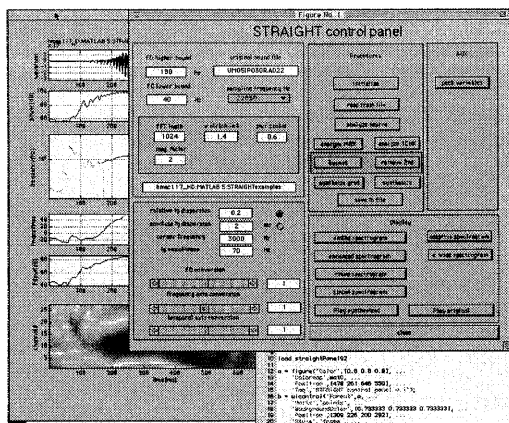


図 1: GUI interface for STRAIGHT procedure.

1 はじめに

我々が耳にする音の中で、音楽に由来するものは、無視できない部分を占めている。これらの音が我々の聴覚機構でどのように処理されて様々な感覚や情緒、芸術的印象と言ったものを生み出すのかを解析し、それらを適切に制御する方法を明らかにすることは、人間の世紀となるべき 21 世紀に向けての重要な研究課題の一つであろう。ここでは、そのための第一歩として、著者らが開発中の音声分析・変換・合成方法 STRAIGHT[2, 3] を楽器音の変換に適用することを試みる。

2 STRAIGHT の適用領域

STRAIGHT は、音声のようにほぼ周期的な信号が主体となる信号の様々なパラメタを、変換音声の高い品質を保ったまま自由に操作することを可能とする方法である。この方法は、音声をスペクトル包絡と音源とに分離し再合成する VOCODER[1] の一種に分類される。この型の方法は、音が持続的な音源の特性と伝達特性とに分離可能であることを前提としている。これは、ほぼ周期的に励振のエネルギーが供給される多くの管楽器や、擦弦楽器については妥当な仮定である。しかし、打楽器の場合には、音源は単位インパルスであり、成分音が調波関係に無い場合が多く、「ほぼ周期的な励振」という条件は満たされていない。

3 楽器音の変換

擦弦楽器の代表としてバイオリン、管楽器からは尺八、打楽器としてピアノを選び、STRAIGHT による分析と変換を試みた。元の楽器音と、分析合成音、変換された合成音の例を会場でのデモ他、以下の

URL に示しておく。いずれの素材も 48kHz 16bit でデジタル化されている。

<http://www.sys.wakayama-u.ac.jp/~kawahara/music.html>

このような実験を進めるための対話的な環境として MATLAB の上に試作した GUI を図 1 に示す。この環境はまだプロトタイプが出来たばかりであり、GUI に実装されている機能は、STRAIGHT で可能な操作を網羅している訳ではない。例えば、ピッチ、周波数軸、時間軸については、線形の伸縮のみが実装されている。STRAIGHT では、相互の干渉に煩わされずにこれらのパラメタを操作でき、任意の非定常な非線形の写像関数を用いることが可能である。音楽への応用のための GUI デザインにおいては、それらの操作に対する適切なメタファーの提供が重要となる。

4 今後の課題

楽器音の変換に STRAIGHT を適用することを試みた。その結果、「ほぼ周期的な励振」という条件を満たすバイオリンと尺八については、自然な分析合成と変換／操作が可能であることが示された。特に尺八の場合には、生々しさも再現されており、実際の音楽シーンへの利用の可能性が示された。しかし、そのような条件を満たさないピアノ音の場合には、音源情報の抽出の段階から問題を生じた。

打楽器のように異なったモードの固有周波数が調波関係に無い信号のピッチをどう定義するかは、人間のピッチ知覚の計算理論を踏まえた精密な議論が必要である。ただし、ここで取り上げたピアノの場合には、時間周波数表現を局所的に眺めると周期信号に類似した構造を有しているので、音源モデルの簡単な拡張によって対応できる可能性はある。時間周波数領域での干渉の除去自体は、局所的な演算のみに依存しているため、そのまま適用可能である。今後は、これらの拡張と、適切な操作のメタファーのデザインについて、検討を進める予定である。

謝辞 実験に用いた音素材を提供して下さった大阪芸大志村 哲 先生に感謝します。本研究の一部は、科学技術振興事業団による戦略的基礎研究推進事業 CREST の援助を受けている。

参考文献

- [1] H. Dudley. Remaking speech. *J. Acoust. Soc. Am.*, Vol. 11, No. 2, pp. 169–177, 1939.
- [2] 河原英紀, 増田郁代. 時間周波数領域での補間を用いた音声の変換について. 信学技報, Vol. EA96-28, August 1996.8.
- [3] Hideki Kawahara. Speech representation and transformation using adaptive interpolation of weighted spectrum: Vocoder revisited. In *Proceedings of IEEE Int. Conf. Acoust., Speech and Signal Processing*, Vol. 2, pp. 1303–1306, Muenich, 1997.