メディア操作　歌唱

TANDEM-STRAIGHTそしてその先へ

河原 英紀　森勢 将雅†　高橋 徹‡　西村 竜一　坂野 秀樹*　入野 俊夫

（河原グループ：和歌山大学システム工学部聴覚メディア研究室）
†関西学院大学理工学部 ‡京都大学大学院情報学研究科 *名城大学理工学部

概要　音声の自然で柔軟な加工を可能にするSTRAIGHTは，歌唱デザイン転写を実現するための基礎となる重要な技術である．2007年に発明されたTANDEM-STRAIGHTでは，このSTRAIGHTのアルゴリズムが根本的に書き換えられた．この発明は，コード量，計算量，調整すべきパラメタの数を大きく削減し，アルゴリズムを遥かに見通しの良いものとした．ここでは，TANDEM-STRAIGHTの背景にある幾つかの重要なアイデアを紹介し，それらがアルゴリズムと構造にどのような変化をもたらしたかについて説明する．また，この発明の応用へのインパクトと，今後検討すべき技術課題について説明する．

キーワード：音声，分析合成，周期信号，標本化定理，VOCODER.

1. はじめに

STRAIGHT[1] は，聴覚における情報表現と機能的に等価な工学的表現に基づく音声・音響処理を実現することを目標として開発されたシステムである．STRAIGHT の中核部分は，生態学的に重要な意味を持ち聴覚においても特別扱いされている，周期性のある音の処理のために開発された．聴覚との機能的な等価性を追求する戦略は，パラメタの高い操作性と加工された音声の高い品質を両立させるシステムとしての STRAIGHT の実現と応用に大きく貢献した [2]．しかし，その実現のために用いられた手段は，複雑で見通しの悪いものであり，STRAIGHT の改良と応用の拡大の障害となっていた．

TANDEM[3] は，周期性のある信号から，分析位置に依存しないパワースペクトルを求めるためのアルゴリズムである．TANDEM は，STRAIGHT で用いられていた相補的時間窓に基づく方法を置き換え，簡潔で見通しの良いものとした．

TANDEM-STRAIGHT[4] では，パワースペクトルの新しい抽出法に加え，標本化の新しい考え方も重要な役割を果たしている．これまでの標本化定理の前提を緩和した consistent sampling という枠組みが紹介されている [5]．この枠組みに基づくことにより，STRAIGHT で用いられていた複雑な平滑化の処理が簡潔で見通しの良いものに置き換えられた．

TANDEM-STRAIGHT の発明は，処理の高速化とコード量の削減という応用上のメリットだけではなく，歌唱音声のより深い理解にもつながる新たな問いを生み出している．ここでは，それらについても紹介し，将来への展望としたい．

2. 背景

聴覚は，到来した音から自動的に，どのような発音体が，どのように刺激されて音を出しているのかを分析する [6]．この機能に倣い，STRAIGHT では，音声を音源情報とスペクトル情報とに分解する．このとき，通常の分析では混在してしまう音源の周期性の影響を，スペクトルから取り除くことが必要となる．この基本的なアイデアを実現するための手段が，TANDEM と consistent sampling により，根本的に書き換えられた．

2.1 TANDEM

音声の物理的性質は，時間とともに変化する．窓関数を用いて選択した信号から求めたパワースペクトルには，この変化とともに，周期信号との相対位置に依存した変化が表れてしまう．TANDEM では，この周期信号に基づく変化が正弦波となることに注目する．正弦波状の変化は，窓関数の位置を基本周期の半分だけ移動させた場合には逆相となる．したがって，元のパワースペクトルと位置を移動させたものとの和を求めるという簡単な処理で，分析位置に依存しないパワースペクトルを求めることができる．周期信号に基づく変化が正弦波状になるために必要な条件を実質的に満たすことは，Blackman 窓のようにサイドローブのレベルが低く急速に減衰する窓関数の場合，容易である．また，基本周波数の推定誤差や変化による影響も，実用上無視することができる．

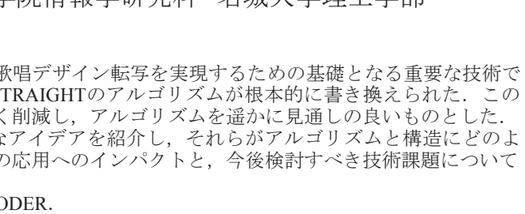

図 1　周期信号のパワースペクトル（左）と TANDEM スペクトル（右）

2.2 consistent sampling

声帯の開閉による声道の周期的な駆動は，周波数領域で声道特性を周期的に標本化することに対応する．このように見ると，音声の生成と知覚の過程を，A/D 変換と D/A 変換になぞらえて考えることができる．標本化定理では，標本化される信号が標本化周波数の半分以下に帯域制限されている場合に，元の信号の完全な復元を保証している．しかし，声道の伝達特性は，フォルマントの周辺では，この条件を満たさない成分を含む．この制限は，復元された信号の再標本化位置での一致だけを要請する consistent sampling[5] に基づくことで回避することができる．また，局所的な伝達特性の形状を仮定することにより，この過程で失われる成分を補充することもできる [7]．

2.2.1 アルゴリズム

音声分析に用いる窓関数のスペクトルと，基本周期による標本化の影響を取り除く平滑化関数の畳込みに基づいて，consistent sampling で用いられる離散領域の補償フィルタ係数が一意に決まる．この係数の絶対値は，次数とともに急速に減少する．この係数を 1 次までで打ち切り，また，スペクトルの正値性を保証するために演算を対数スペクトルの上で行うこととした．平滑化関数を，最も局所性の高い矩形とすることにより，アルゴリズム全体は，3 つの式で表すことのできる簡潔なものとなった．

2.3 音源情報

TANDEM により求められるパワースペクトル（TANDEM スペクトル）には，音源情報とスペクトル情報が含まれている．consistent sampling の考えに基づいて復元されたスペクトル（STRAIGHT スペクトル）では，周期性に起因する音源情報が取り除かれている．両者の比からバイアス分を取り除くことにより，音源情報だけを表したスペクトルが求められる．この表現に基づく周期性の評価は，基本周波数に依存しない．複数の基本周波数を仮定し評価値を合成することにより，事前知識を必要としない簡潔な分析器が構成された．ここでは，詳細を省くが，帯域毎の音源の非周期性についても，同様な分析器が構成できる．

3. 構成

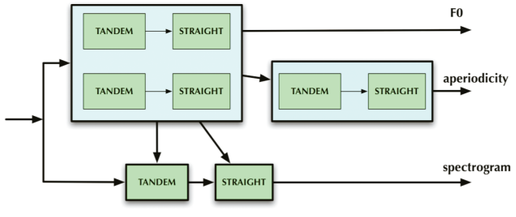

図 2　TANDEM-STRAIGHT の構成

TANDEM-STRAIGHT の構造を図 2 に示す．TANDEM と STRAIGHT と記された同一の構造の処理が，音源情報（基本周波数，非周期性）とスペクトル情報の抽出に繰り返し用いられている．さらに，それぞれの要素は FFT や線形補間など，最近のプロセッサ（あるいは GPU）で効率よく実行できる処理で構成されている．これらは，TANDEM-STRAIGHT を実時間システムとして実装する上で有利な特徴である．

4. その先へ

TANDEM-STRAIGHT には，以前の STRAIGHT に共通するコードは存在していない．当然，STRAIGHT の品質の改良のために加えられて来たノウハウに属する工夫も継承されていない．そのため，注意深く比較すると，現状では STRAIGHT の方が合成音声の品質が良い．失われた成分の回復 [7] は，その差を明確に定式化された方法で埋める試みの第一歩である．これまでの STRAIGHT の C 言語での実装の提供 [8] による応用の促進と併せて，着実に TANDEM-STRAIGHT の改良を進め，STRAIGHT を凌ぐ品質を実現する予定である．

参考文献

[1] Kawahara, H., Masuda-Katsuse, I., and Cheveigné, A.: Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction, Speech Communication., 27(3-4), pp.187-207 (1999).
[2] 河原英紀：Vocoder のもう一つの可能性を探る - 音声分析変換合成システム STRAIGHT の背景と展開 -，日本音響学会誌，Vol.63，No.8，pp.442-449 (2007).
[3] 森勢 将雅 , 高橋 徹，河原 英紀 , 入野 俊夫 : 窓関数による分析時刻の影響を受けにくい周期信号のパワースペクトル推定法，電子情報通信学会誌 D, Vol.J90-D, No.12, pp.3265-3267 (2007).
[4] Hideki Kawahara, Masanori Morise, Toru Takahashi, Ryuichi Nisimura, Toshio Irino, Hideki Banno: 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).
[5] Unser, M.: Sampling - 50 years after Shannon, Proc. IEEE, 88(4): 569—587 (2000).
[6] Toshio Irino and Roy D. Patterson: Segregating information about the size and shape of the vocal tract using a time-domain auditory model: The Stabilised Wavelet Mellin Transform, Speech Communication, 36 (3-4), pp.181-203 (2002).
[7] 河原英紀 , 森勢将雅 , 高橋 徹 , 坂野秀樹 , 西村竜一 , 入野俊夫 : TANDEM-STRAIGHT によるスペクトル包絡の近似精度の改善についてー基本周波数により定まる Nyquist 周波数以上の空間周波数成 分の復元についてー，信学技報 , 音声研究会 (2008)
[8] http://straight-suite.sys.wakayama-u.ac.jp/