

研究概要報告書

(1/4)

研究題目	局所変化率に基づく音声・楽器音の高精度分析技術の研究	報告書作成者	伊藤 仁
研究従事者	伊藤 仁		
研究目的	<p>音声の有声音部や一部の楽器音は、振幅と周波数が連続的に変化する正弦波信号(sinusoid)の和として表現できる。このような表現は Sinusoidal Modeling と呼ばれ、オーディオ信号の符号化・分析・変調など広範囲の応用の基礎となる重要なアプローチである。Sinusoidal Modeling における入力信号の波形 <math>x(t)</math> であり、出力は各調波成分の瞬時振幅 <math>A_k(t)</math> と位相 <math>P_k(t)</math> になる(次式)。入力よりも出力の方が自由度が高いという意味で、これは一種の不良設定問題であり、何らかの拘束条件を設定しなければパラメータを計算する事ができない。</p> $x(t) = \sum_{k=1}^K \exp(A_k(t)) \cos(P_k(t))$ <p>既存の Sinusoidal Modeling の研究では、この拘束条件として、局所的な分析時間内において振幅と周波数が定常とみなせるという局所定常仮説を採用している。だが、実際の音声や楽器音の音響特性は時々刻々変化するものであり、このような仮説が妥当とみなせる区間は限られている。従って、局所定常仮説に基づく Sinusoidal Modeling では、特に入力信号の時間変化が大きい区間で不可避的な誤差が発生することになる。この誤差は、推定されたパラメータから再合成した音の品質に強く影響するため、既存技術を用いて原音と匹敵する品質を持つ合成音を生成するのは非常に難しい。</p> <p>本研究の目的は、物理的特性が時間変化する音声や楽器音を高精度で分析する技術を確立することにある。その為に、局所変化率符号化と呼ぶ新しい手法を提案する。この手法の特徴は、局所的な分析時間内における入力音の振幅や周波数を、単純な時間変化関数としてモデル化する点にある。即ち、上で述べた局所定常仮説に起因する誤差が原理的に発生しない。これらの信号パラメータは、入力波形の周波数応答から一意に決定することが可能である。この手法により、波形とパラメータ間の可逆変換が可能となるため、パラメータの推定精度自体も定量的に評価することができる(図1)。</p> <p>平成17年度は、対象を有声音と管弦楽器音に絞り、S/N比で60dB以上のパラメータ推定精度を実現することを目標とする。この研究の成果は、短期的には音声・楽音の高精度符号化、少データ高品質の音声合成、楽器音の音響的品質評価等に応用できる。また長期的には、我々が音をより深く理解し、有効に活用する為の基盤技術になると考えられる。</p>		

研究内容

1. 提案手法

局所変化率符号化のシステム構成を図2に示す。まず入力信号  $x(t)$  に対して、等間隔の時間ステップで周波数応答  $S_0$  とその角周波数に対する偏微分  $S_1, S_2$  を計算する。これらの応答から、次式の局所変化率変換を用いて、信号の成分パラメータ  $a_0, a_1, a_2$  を計算する。ここで各パラメータの実部が、対数振幅の瞬時値、変化速度、変化加速度にそれぞれ対応し、虚部が位相の瞬時値、変化速度 (瞬時角周波数)、変化加速度にそれぞれ対応する。

$$\alpha_0(\omega) = \log \left( \frac{S_0^2}{\sqrt{S_1^2 - S_0 S_2}} \right) + \frac{S_1^2}{2(S_1^2 - S_0 S_2)}, \quad \alpha_1(\omega) = j \left( \omega - \frac{S_0 S_1}{S_0 S_2 - S_1^2} \right), \quad \alpha_2(\omega) = \frac{S_0^2}{S_0 S_2 - S_1^2} + \gamma$$

局所変化率図3に自然音声に対する局所変化率変換の例を示す。この音声の基本周波数は約140Hzであり、各成分の振幅ピーク近傍の周波数では、入力スペクトルと推定スペクトルが良く一致していることが分かる(図3a,b)。これは局所変化率変換により各成分の信号パラメータが正しく推定できていることを意味する。このようにして得られた信号パラメータに基づいて、各成分に対する瞬時振幅関数と位相関数を計算する。基本周波数が大きく変化しない信号に対しては、ここで計算されたパラメータ関数が、そのままシステムの出力となる。

だが、自然発話音声の多くは、基本周波数の変化が激しい。このような場合、特に高次の調波成分の応答ピークがブロードになるため、成分間の干渉が発生する。図4に成分間干渉の例を示す。干渉が大きい周波数成分に対しては、局所変化率符号化を用いても正確なパラメータ推定ができない。この問題に対応するため、時間/位相変換を導入する。成分間干渉は主に高次の調波成分において発生し、低次の成分では殆ど発生しない。従って、自然発話音声の第一成分は局所変化率変換により正確に推定することが可能である。この性質を利用して、入力信号の時間  $t$  を、推定された第一成分の位相 で変換した信号を計算する。この変換により、成分間干渉を抑圧することが可能である(図4b,d)。

時間/位相変換を施した信号の全ての成分に対して、局所変化率変換を用いて成分パラメータを推定する。得られたパラメータ関数の位相 を時間  $t$  に逆変換することで、各成分のパラメータ関数を得ることができる。図5に自然音声入力に対して、推定された瞬時振幅と瞬時周波数の例を示す。これらのパラメータから再合成された信号のスペクトログラムは、入力信号のスペクトログラムに非常に近いものになっている事が確認できる(図5a,c)。

## 研究内容

## 2. 性能評価

上で述べた局所変化率符号化の性能を定量的に評価した。まず入力信号のパラメータが既知である合成二連母音を用いて評価を行った。局所変化率符号化では、信号に含まれる調波成分の瞬時位相、周波数、振幅を推定することが可能である。このうち第一成分の瞬時周波数は、基本周波数 (ピッチ) と呼ばれる重要な特徴量であり、様々なピッチ推定アルゴリズムが提案されている。同一の合成音声を入力した場合の基本周波数のピッチ推定誤差は、代表的な既存アルゴリズムであるケプストラムで 0.16%、提案手法では 0.00065% である。また全成分のパラメータ推定精度は、入力対残差信号比 (Signal to Residual ratio: S/R) で定量化できるが、実験で用いた全ての合成音声に対する平均 S/R は 46.3dB であった。

次に自然音声を入力として性能を評価した。自然音声は ATR デジタル音声データベースの有声音音声サンプル 384 個を用いる (男性 16 名 + 女性 16 名、単語数 12)。これらの入力に対して局所変化率符号化により信号パラメータを推定し、それらに基づいて波形を再合成する。再合成された信号と入力信号の差から残差信号を求めて S/R を計算する。図 6 に自然音声に対する入力、再合成、残差信号の例を示す。図から残差信号のエネルギーが非常に小さく、スペクトrogram 上では調波成分のエネルギーが殆ど無くなっていることが確認できる。これら全ての自然音声に対する平均 S/R は 22.6dB であった。合成音声と自然音声の性能の差は、音源の性質に起因すると考えられる。有声音声を発話する場合の音源は、主に声帯振動を反映した準周期信号だが、ごく低いエネルギーで乱流雑音も発生している。自然音声では、この雑音成分が残差として現れている (図 6)。だが、このような問題にもかかわらず、提案手法により再合成された音声の自然性は非常に高く、しばしば入力音と区別するのが難しかった。

また局所変化率符号化は、管弦楽器音など調波構造に近い性質を持つ信号に対しても適用可能である。図 7 に、これらの楽器音に対するパラメータ推定の例を示す。これらの結果は、入力信号の性質から三つに分類できる。第一のグループは、立ち上がり比較的緩やかで、調波成分以外のエネルギーが少ない信号であり、トランペット、アルトサクソ、ハーモニカなどがこれに対応する。このような入力音に対する提案手法の推定精度は非常に高く、S/R で 30dB 以上の性能が得られた。第二のグループは、立ち上がりに急激な振幅変化 (アタック) を有する信号で、ピアノやマンドリンに対応する。これらの信号では、立ち上がりの推定誤差が大きくなった。第三のグループは、調波成分以外のエネルギーが大きいもので、バイオリンや尺八などが対応する。例えばバイオリンの場合は、弦をこする音が、尺八の場合は呼気の音が残差として現れる。だが、これらの問題が聴取した音の印象に及ぼす影響は小さく、自然音声の場合と同様、楽器音においても入力音と再合成音を区別するのは難しかった。

研究概要報告書

( 4/4 )

<p>研究のポイント</p>	<p>本研究では、音響特性が時間変化する信号を高精度で分析する手法を提案した。この手法のポイントは、入力信号の周波数応答から時間変化を含む信号パラメータを推定する局所変化率変換と、時間変化により不可避免的に発生する成分間干渉をキャンセルする時間/位相変換の二つである。前者は、単純な時間変化信号の応答から理論的に導出したものであるが、性能評価実験により自然音声など実環境に存在する音に対しても有効であることが示された。この変換は、音信号の音響分析の分野で広く受け入れられてきた局所定常仮説の問題を解決し、時間変化する信号を精度良く分析できる点にその価値がある。また後者の時間/位相変換は、成分間干渉の問題を解決する方法のひとつである。ここでは信号の時間軸を基本周期に対応する位相軸に変換することで、話者による基本周波数の差を正規化できる点が重要である。</p>
<p>研究結果</p>	<p>提案手法による自然音声のパラメータ推定精度は、S/R で平均 22.6 dB であった。また楽器音の推定精度は、楽器の種類により異なるが概ね 20 ~ 30 dB 程度であった。これらは、本年度の目標である60dB には達していないが、既存の Sinusoidal Modeling の性能を大きく上回るものであった。実際、推定されたパラメータに基づいて再合成された信号は非常に自然で、しばしば入力信号と区別するのが難しいレベルに達していた。よって、提案手法は、自然発話音声や楽器音を高精度に分析する技術として有効であり、今後この分野の研究の発展に大きく寄与するものと考えられる。本研究の成果は、応用音響研究会 (信学技法 EA2005-2, pp.7-12)、米国音響学会 (J. Acoust. Soc. Am. Vol. 118(3), pp.2024) で発表され、東北大学博士学位論文 (工学研究科、平成 17 年度) としてまとめられた。</p>
<p>今後の課題</p>	<p>特に楽器音の分析において、音の急激な立ち上がり(アタック)に対するパラメータ推定が課題のひとつとして挙げられる。ここで生じている問題は、振幅の急激な変化による成分間の干渉であり、本研究で解決した基本周波数の急激な変化による干渉と同種のものであると考えられる。この振幅変化による干渉は、時間/位相変換のような手法では正規化できない為、これを解決する新たなアイデアが必要である。例えば、各成分のスペクトルを再合成し、成分間の干渉を含む合成スペクトルと入力スペクトルの差を最小化することにより、アタック部の信号パラメータを推定する手法などが挙げられる。また提案手法は、複数の音声と同時に存在する場合には対応できない。実環境で動作するシステムへの応用等を考える場合には、この問題も重要になると考えられる。</p>

図 1. 局所変化率符号化の概念

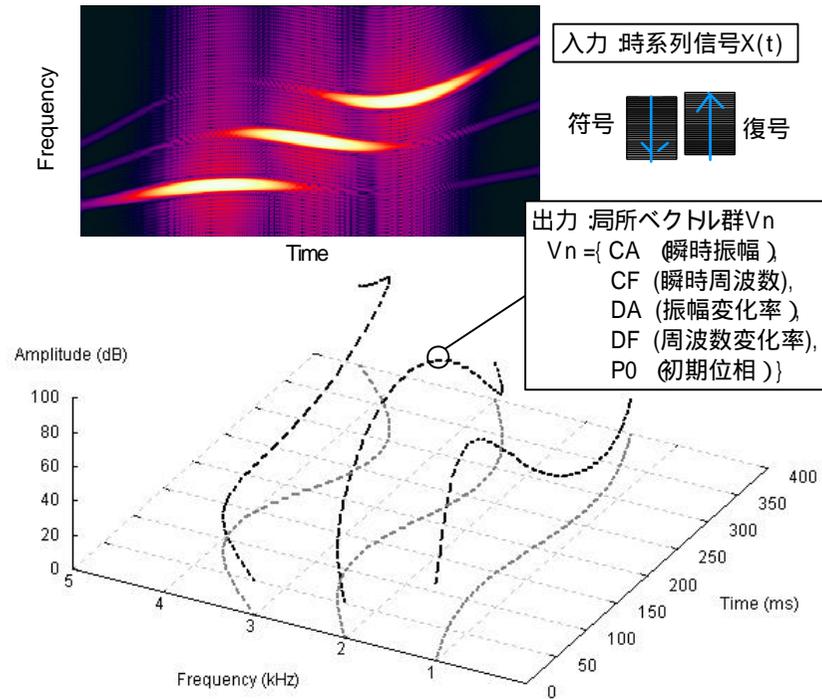
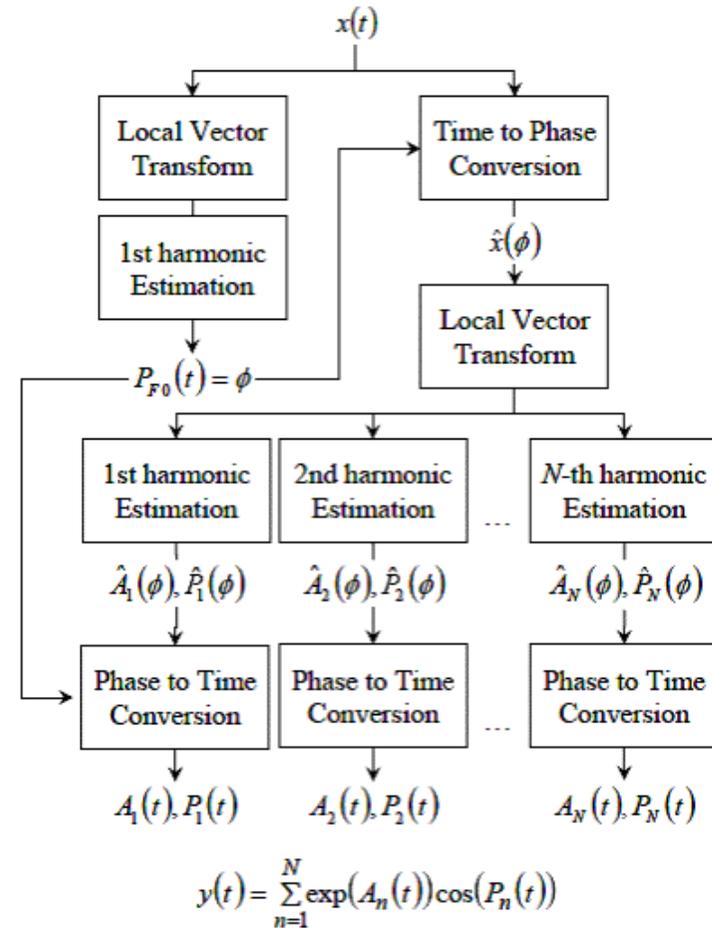
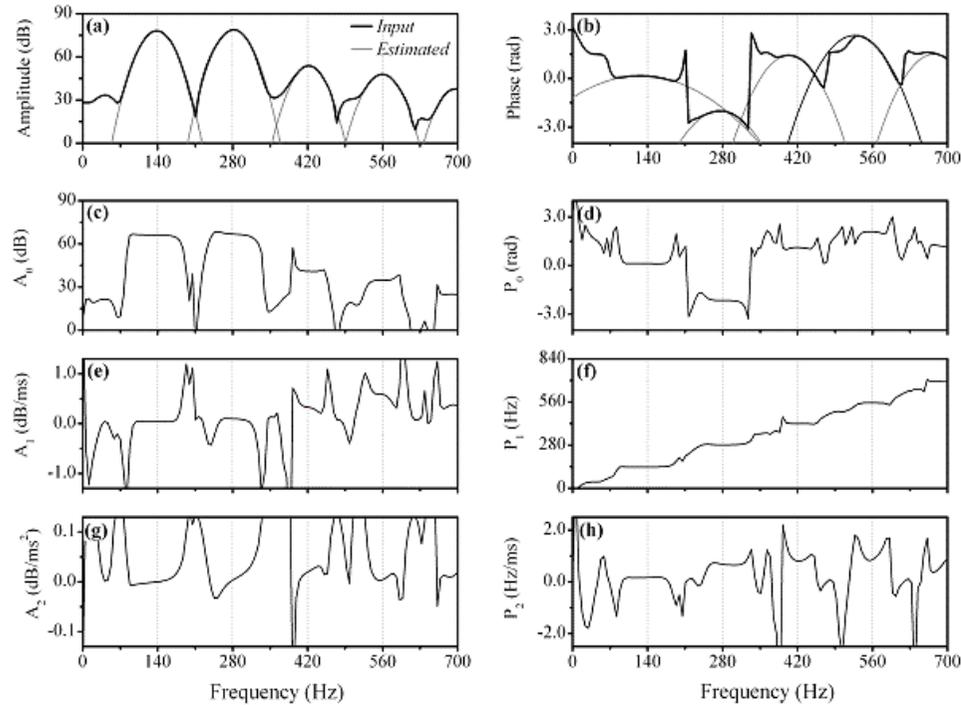


図 2. システム構成



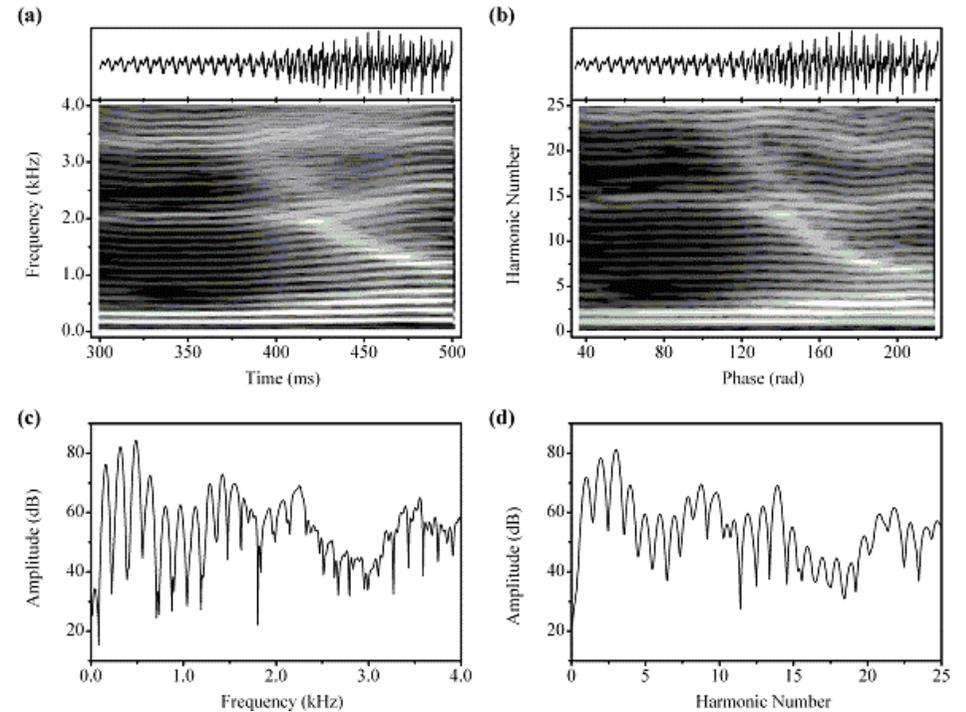
(注 :フローチャート図 ,ブロック図 ,構成図 ,写真 ,データ表 ,グラフ等 研究内容の補足説明にご使用下さい。)

図 3. 局所変化率変換の例



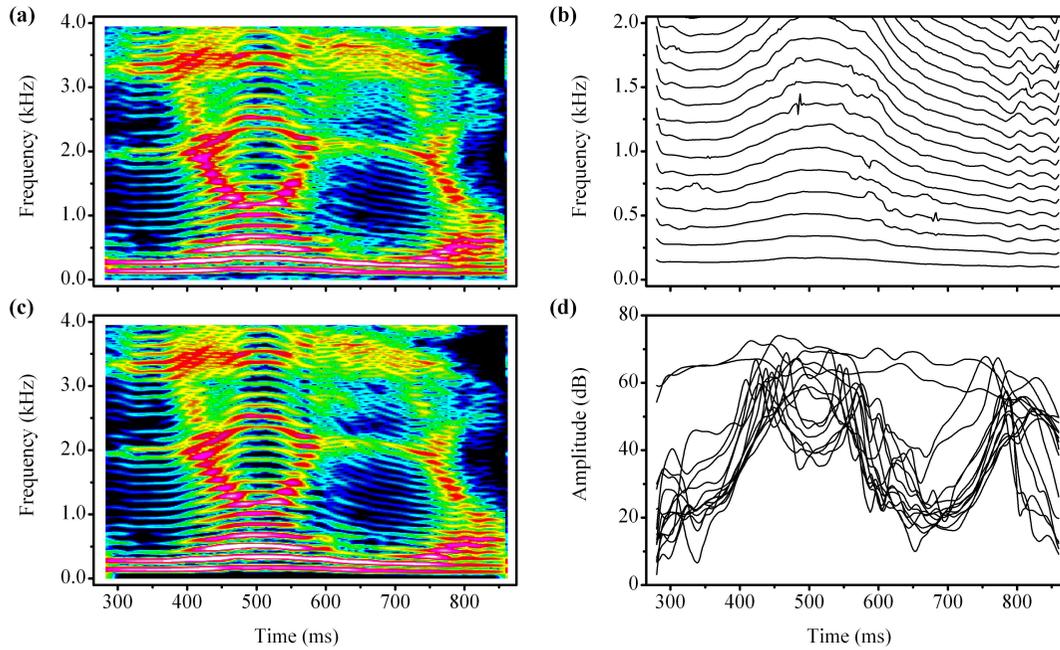
(a)実線：入力信号の振幅スペクトル (男性話者/iyoio/) 薄線：LVT により推定された信号パラメータから再合成した振幅応答 (b)位相スペクトル (c)推定された瞬時対数振幅 (d)推定された瞬時位相 (e)推定された振幅変化速度 (f)推定された瞬時周波数 (g)推定された振幅変化加速度 (h)推定された瞬時周波数変化速度

図 4. 時間/位相変換の例



(a)男性話者/iyoio/の/i/から/y/への遷移部分の波形とスペクトログラム。エネルギーの強い部分を白で表示 (b)時刻 450 ms における周波数振幅応答。F0 の変化速度が大きく、高次の調波成分 (2.0 kHz 以上) 間で干渉が強い (c)時間/位相変換した後の波形とスペクトログラム。スペクトログラムの横軸は、第一調波成分の位相、縦軸は調波番号を表す (d)振幅スペクトル

図 5. パラメータ推定の例



(a)入力信号の振幅スペクトログラム (b)推定された瞬時周波数 (c)推定パラメータから再合成された信号のスペクトログラム (d)推定された瞬時対数振幅

(a), (b) 上段 :入力信号 中段 :再合成信号 下段 :残差信号  
(c),(d) 上段 :入力信号のスペクトログラム、中段 :再合成信号のスペクトログラム 下段 :残差信号のスペクトログラム

図 6. 入力、再合成、残差信号の例

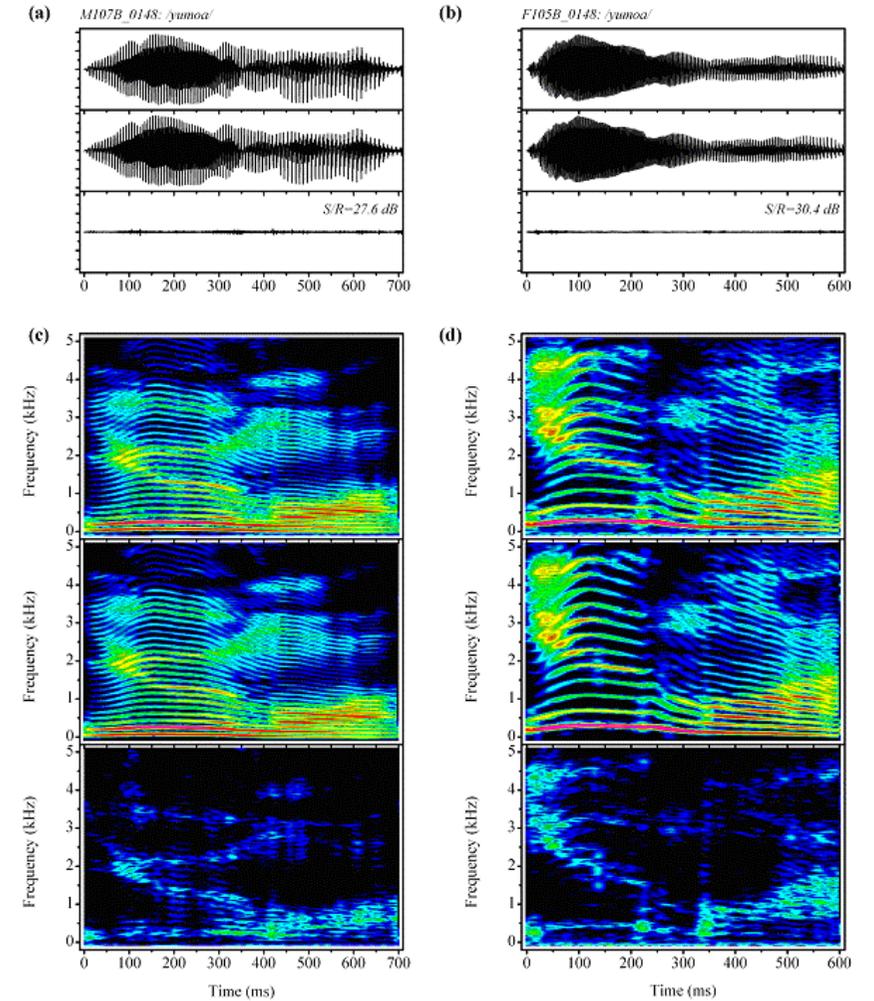
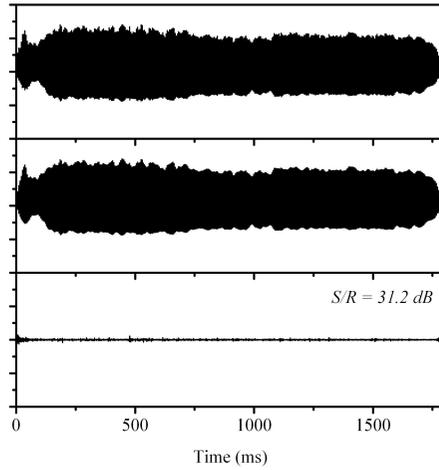
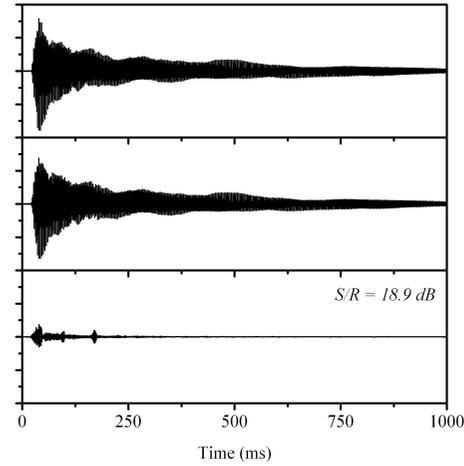


図7. 楽器音に対するパラメータ推定 (上段:入力、中段:再合成、下段:残差)

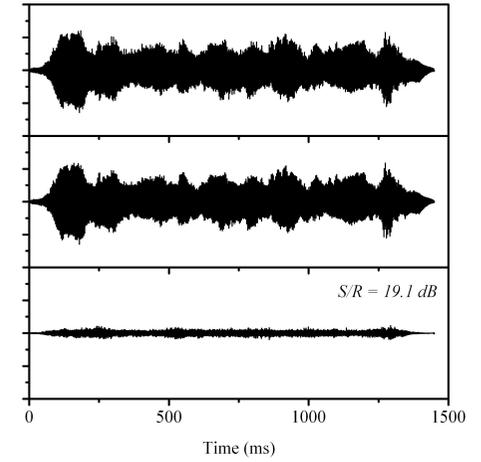
(a) トランペット



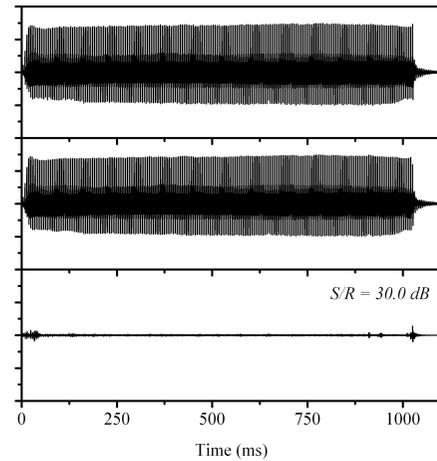
(c) マンドリン



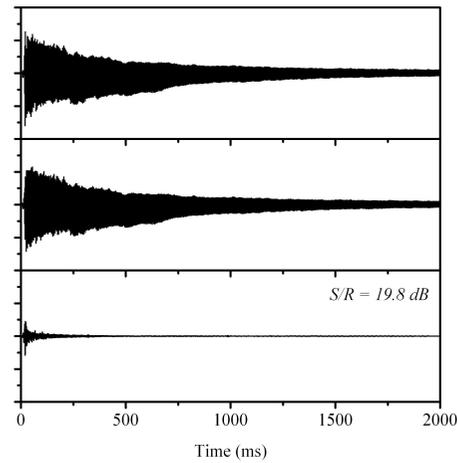
(e) バイオリン



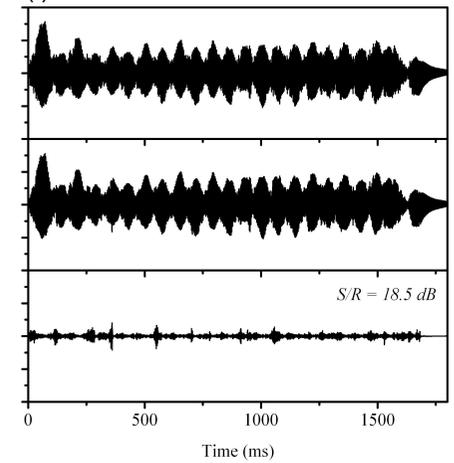
(b) アルトサクソ



(d) ピアノ



(f) 尺八



(注:フローチャート図,ブロック図,構成図,写真,データ表,グラフ等 研究内容の補足説明にご使用下さい。)