

楽音の統計的性質を利用した音楽信号の高次bit推定方法

橋本 秋彦 山田 辰美 下原 勝憲

NTTヒューマンインタフェース研究所

アナログ信号を有限長量子化によってデジタル符号化すると情報量の欠落に伴って量子化歪みが発生する。本論文はこの問題が補間の細工だけでは解決できず原信号の推定ができてのみ解決できることを指摘し、楽音の統計的性質を仮定しそれを用いた標本点における真値の推定方法について検討したものである。

Original signals estimation method from digitalized musical data

Akihiko Hashimoto Tatsumi Yamada Katsunori Shimohara

NTT Human Interface Laboratories

1. はじめに

現在、音声の記録方式はCD, DAT, MD等のデジタル記録方式が主流である。ところで、有限長量子化によるデジタル記録においては量子化に伴う情報の欠落によって量子化歪みの発生が避けられない。

本文では、対象信号の統計的性質について検討し、この性質を利用することによる有限長量子化の際に失われた情報の推定方法について報告する。

2. 従来技術

図1に示す原信号1aを有限長量子化すると図1の黒点で示す標本点列となる。この点列データから通常の補間で曲線を補間生成すると1bが得られる。この1aと1bの差が量子化歪みである。

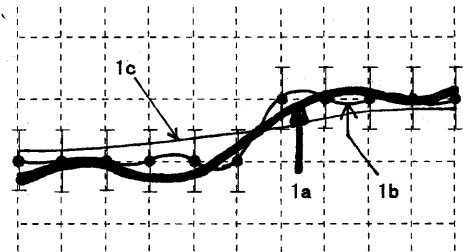
ところで、1cのようにより滑らかな曲線で補間する方式[1]が量子化歪みの低減を売り物に商品化されている。このような補間を行うと有限長量子化に伴って発生する階段状の波形がなまるためにあたかも真値に近づくかのような錯覚を目に与える効果はある。

しかしながら、原信号の性質を適切に反映する工夫をしない限り、補間式の表現形態(多項式/有理式, 次数等)を変えて滑らかさを増すとこと、補間結果の原信号に対する近似精度を向上させることは無関係であることが数値計算の初歩的な常識[2]として知られている。これは補間式が標本点を通過するタイプか否かには依存しない。

上記[1]の技術は原信号の性質を適切に反映するものではなく、本質的には標本点間の変位が1LSBの区間だけに作用する適応型ローパスフィルタに過ぎない。この時のナイキスト周波数は1LSB変位するまでの標本点数の逆数で与えられ、概周波数以下の帯域にある量子化歪みは除去されない。概

周波数以上の帯域にある量子化歪みも除去されるのではなく単に低域に折り返されるだけである。つまり、量子化歪みの周波数分布を低域に集めるだけの結果に終わり量子化歪み総量は変わらない。

従って、このような技術?は単に滑らかな曲線を生成するだけで、量子化歪みの低減、原信号に対する近似精度の向上、失われた情報の復元等観点では全く無意味な曲線を補間生成しているに過ぎなかった。



→標準化時刻
図1 有限長量子化標本点列と原信号

3. 楽音の統計的性質に対する仮説

失われた情報を復元することなしに量子化歪みの低減、原信号に対する近似精度の向上は不可能である。このことを波形の生成で言い換えると、図1の与えられた有限長量子化標本点列データより、量子化前の原信号が1aであったなら1aの波形を作り出して初めて量子化歪みが低減できることを意味する。しかしながら、各標本点の $\pm 1/2$ LSB内を通過する曲線は無数に存在し、既存の弱い拘束条件(ex. 原信号の帯域20-20kHz)では真値の推定ができるほど解空間を縮小することはできない。

我々は、対象信号の統計的性質を拘束条件とし

て解空間を縮小することによって推定を行う超解像のアプローチに基づき、有限長量子化の際に失われた情報の復元を試み量子化歪みの低減を図る。

ところで対象信号について、音声記録物の大半は楽音であり音質向上が重要視されるのもまた楽音である。この楽音は、現実存在する楽器（含む声音）、あるいは器楽音のPCM音源方式シンセサ

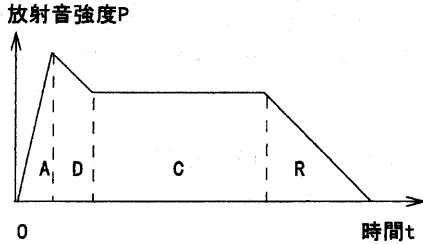


図2 器楽音の時間推移

A: Attack time D: Decay time
C: Continous time R: Release time

イザ等である場合がほとんどである。そこで、器楽音の統計的性質について検討する。器楽音の時間推移は図2のように示される。自励楽器は図2によく従うがADにおける強度ピークがない場合もある。減衰楽器ではCの過程がなくADRと変移する（利便上、RをCと見なし、強制的に振動体を停止させる過程をRとすることがある）。大半の楽器はADで示されるトランジェント区間は数十～数百ms秒しかなく、CまたはRの区間が時間の大半を占める。

器楽音の楽器の物理現象[3][4][5]より図2の変移過程で起きていると推定される物理現象を表1に示す。

なお、ADの過程は表1に示すような楽器固有の物理要因以外に、演奏者が楽器の音を立ち上げるために共鳴体にエネルギーを送る初期充足過程、およびその後の音量／音程調整過程と言った人間の過度応答特性も実際には含まれると思われるが、本論で考慮する必要があるほどの応答速度で進行

楽器の種類	A	D	C	R
減衰楽器 (打楽器、オルガンを除く鍵盤楽器)	打撃点で振動体の変移、その変移が全体に波及する過程	変位波が境界で反射し、振動体の各モードに合う周波数の振動成分が定常波を形成していく過程		定常波が音響放射損失、摩擦熱損失等によって指数関数的に減衰する過程
リード系自励楽器 (金管楽器、フルートを除く木管楽器)	リード振動源の音が共鳴管に供給される過程	共鳴管で定在波が形成されていく過程	流れる空気から定常波に変換されるエネルギーと音響放射損失、熱損失、粘性損失により失われるエネルギーの平衡状態	音響放射損失、熱損失、粘性損失により共鳴管内空気の振動が指数関数的に減衰する過程
エアリード系自励楽器 (フルートを含む笛の仲間、パイプオルガン等)	フルーからのジェット流と静止空気間で発生する圧力変動が共鳴管に供給される過程	共鳴管内での定在波の形成と、フルーのジェット流の圧力変動が共鳴によって同時形成されていく過程	同上	同上
弦楽器系自励楽器 (弦楽器)	弓によって弦にヘルムホルツ波が発生し、その振動が駒を通して共鳴体に供給される過程	共鳴体の定在波と弦の共鳴が同時形成されていく過程	弦振動から共鳴体に供給されるエネルギーと音響放射損失、熱損失等により失われるエネルギーの平衡状態	音響放射損失、熱損失等により共鳴体の振動が指数関数的に減衰する過程

表1 種類別に見た器楽音の変移過程の物理的意味

するわけではないのでここでは無視する。

ここで、時間的な大半を占めるGとRの過程について考える。

Gは共鳴体または共鳴管内の空気の定常振動であり、非線形性がなければ音響スペクトル形状は一定であると言える。Gはエネルギー流入量とエネルギー流出量の平衡状態でもあるが、流出量（ \propto 音響パワー）は内部振動エネルギーの定数倍（ $\ll 1$ ）なので内部エネルギーが一定なら音響パワーも一定である。一方、エネルギー流入量は時間毎に変動するが、内部エネルギーはエネルギー流入量の積分量であるため、エネルギー流入量の変化に対して内部エネルギーの変化は特に変動量の高周波成分に対しては小さい。Rは内部エネルギーが指数関数的に減衰していく過程であり、その時定数は一般の楽器では数百m秒から10秒前後のオーダーである。

このようなGとRの性質に基づき、器楽音の周波数スペクトルは強度、位相共に1~10m秒のオーダーでは線形変位で表現できるとの仮説を立てる。本仮説は次のような場合には成立しない、もしくは近似精度が悪くなる。

- P1. ADで示されるトランジェント区間
- P2. 管弦楽器等で音程を変える瞬間
- P3. 振動体の非線形効果が著しく不連続な挙動を示す場合
- P4. 蓄積される内部エネルギー量と放射エネルギー量にあまり差がない場合

しかしながら、P1やP2が時間的に占める割合は非常に小さい。また、P3やP4も問題になるような楽器は一部（打楽器、特に南方系民族楽器に見られる響きをほとんど伴わない打楽器）に限られるので問題は少ないと考えられる。

4. 量子化歪みに対する仮説

有限長量子化によって発生する量子化誤差は原信号、標本周波数、標本bit数が与えられれば一意に決まる歪みであり、量子化歪みと呼ばれる。しかし、原信号が単純かつ規則正しい場合は容易に

原信号との相関を認めることができるが、楽音のように多彩な周波数スペクトルで構成される複雑な波形上に生じる量子化誤差は低次の相関が不明瞭となる。音聴感上においても、量子化歪みであるにも関わらずノイズのように聞こえるため量子化ノイズと呼ばれることもある。そこで、楽音を音源としたときに生じる量子化歪みの性質は、1~10m秒のオーダーで見た場合、相関を有さないノイズと同等に扱えると言う仮説をおく。

5. 真値の推定手法

真値の推定は、設定した仮説の元では周波数スペクトル変位の平滑化を行うことによって実現できる。周波数スペクトルが強度、位相共に微小時間において線形変位する器楽音は平滑化処理に影響されることがないのに対して、量子化歪みは平滑化処理によって処理対象となる標本点数の平方根分の1に減少させることができるからである。この真値推定システムのブロック図を図3に示す。

図3の動作を説明する。周波数スペクトル検出部は標本点列データから各標本点における周波数スペクトルを逐次算出する。ここで時刻Tにおける周波数スペクトルは、標本点列データを D_i ($i=0 \dots n$)として、 D_{T-M} から D_{T+M} の2M個の標本点データよりフーリエ変換した値を用いる。Mは16~256まで2のべき乗倍で可変に設定できる。また、窓関数にはハニング窓を使用した。

変換処理によって時刻Tにおける周波数スペクトルのデータ群が以下の記述で得られる。

$$P_{ij} \quad (j=1 \dots M) \quad \omega_{ij} \quad (j=1 \dots M)$$

P_{ij} は各周波数スペクトルの強度、 ω_{ij} は位相である。

次に推定部において、真の周波数スペクトルの推定値を求める。仮説に従えば与えられたデータに直線を当てはめ、該直線の $i=T$ における値を求めることとなるが、Tが与えられたデータの中心に位置するのでこの場合は平均値を求めることと等価である。そこで、 $i=T$ 、 $j=X$ における推定値 P'_{TX} 、 ω'_{TX} を次に示す平均式で求める。

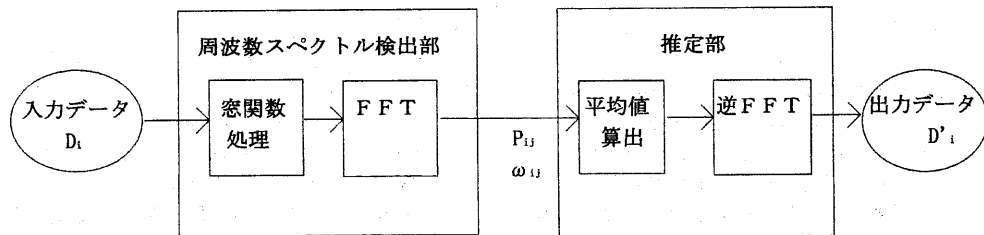


図3 推定システムのブロック図

$$P'_{ik} = \left(\sum_{l=k}^{i=k} W_l P_{(i+l)k} \right) / 2k$$

$$\omega'_{ik} = \left(\sum_{l=k}^{i=k} W_l \omega_{(i+l)k} \right) / 2k$$

ここで W は重みづけ関数である。kは1から32まで任意に設定できる。

真値算出部は推定した上記スペクトル値の組から逆フーリエ変換によって $i=T$ における推定値 D'_i を算出する。各標本点毎に上記の処理で推定値を求めていくことによって、真値を推定する標本点列 $D'_i (i=1 \dots n) T$ が得られる。

予備実験として20kHzまでの周波数スペクトラムを有する定常波の16bit量子化信号を生成し本推定処理を行った。その結果、 $M=128$, $k=3$, $w=1$ (重みづけなし)の条件で生成した推定値の真値に対する誤差量は元の16bit量子化信号の約1/4に減少し、18bit量子化信号に相当する精度が得られた。

6. 本実験に対する検討と今後の進め方

提案した手法が定常波に対しては有効に動作する場合もあることが確認できた。次に実際の楽音に対する有効性の評価を行うが、本実験にあたって以下に示す問題を現在検討中である。

実験に用いるデータについて

推定の有効性の評価において、原信号に対する仮説(拘束条件)が厳密に成立することが言えず統計的にしか成立しない場合、有効性の評価も統計的に行わなければならない。この時、実験データ数が多いほど、また、実験データの母集団に対する統計的な偏りがないほど実験結果の信頼性が高まることは言うまでもない。

ところが、16bitの音源データは豊富にあり容易に入手できるが20bitの音源データの豊富な入手は困難である。そこで16bit音源のデータの上位12bitから下位4bitを推定する問題に置き換えて有効性を評価することを検討している。

次に実験データの偏りについて、CDは膨大な種類が出回っているが、特定の尺度から見て統計的に偏らないサンプルの抽出はできても、全ての尺度から見て統計的に偏らないサンプルの抽出は不可能である。また、他者が比較実験を試みる場合も同じデータを入手することは大変であるばかりか後日には入手不可能となる可能性もある。これについては器楽音の録音物そのものではなく、器楽音のサンプリング波形を収録したシンセサイザ

音源を実験データに用いる方法も考えられる。

誤差の評価方法について

誤差の評価方法は最小二乗法で行うことが一般的であり、本論文で示した推定器の動作も最小二乗法に基づいている。しかし、最小二乗法が最尤推定であるのは誤差が独立で正規分布に従い標準偏差が一定である場合に限られる。量子化誤差が独立であることは本仮説で前提条件としていることであり、標準偏差が一定であると見なすことも自然である。しかし、量子化誤差が正規分布に従うと言うのは非常に不自然であり一様分布に近いと見るべきであろう。この問題は誤差の評価方法ばかりでなく、推定システム内の処理内容にも直接影響するので、誤差の評価方法についてはより厳密な検討が必要と思われる。

適応処理の問題について

本論文で提案した推定方法には M, K, W 等のパラメータが含まれる。これらの最適な値を調べることが実験目的の一つでもあるが、最適値は異なる実験データによって、また同一の実験データでも時刻によって刻々変動することが予想される。また、前節で述べたように実際の信号内には本推定が前提としている仮説の成立が期待できない部分も存在する。仮説が成立しない部分にその仮説を前提とした推定を行っても無意味な結果しか得られないばかりか、条件によっては元より悪化する場合もある。このような問題の対処方法の一つに適応的動作がある。

そこで意味のある適応的動作を実現するために、信号の状態に応じた最適なパラメータ値の決定と共に、信号から適切な適応動作を行う条件の抽出方法についても実験および理論面から検討していきたい。

参考文献

- [1] 林英昭, 後木信吾, [新DA変換システム「ALPHAプロセッシング」の動作原理と実際, ラジオ技術, pp. 61-66, 1993
- [2] 補間の教科書! NUMERICAL RECIPES in C [日本語版] (技術評論社) はシニカルかつユーモアある解説
- [3] Arthur H. Benade (小暮陽三訳), [音と楽器一波動と聴覚の基礎知識], 河出書房新社
- [4] [楽器の科学], 日経サイエンス社
- [5] 吉川茂, [ピアノの音色はタッチで変わるか—楽器の中の物理学], 日経サイエンス社