

クリッピングした音響信号の修復*

三浦晋 (筑波大,HRI), 中島弘史 (HRI), 牧野昭二, 山田武志 (筑波大), 中臺一博 (HRI)

1 はじめに

音声や音楽を録音する際、入力音圧レベルが過大であれば、電気回路やマイクロフォンの機構により、録音されたデータの振幅が許容値以上の場合に許容最大値として扱われ、独特のひずみが生じてしまう (Fig. 1)。この現象はクリッピングと呼ばれている。クリッピングは受聴時の音質の劣化だけでなく、音声認識、音源分離、方向推定などの性能低下も引き起こす原因の一つでもある。現在、携帯電話やカーナビゲーション等を使用されている小型マイクロフォンは、物理的な都合上過大な音圧に耐えることができず、クリッピングを引き起こす事も少なくない。本研究では、音声に限らず楽音にも対応できるクリッピング修復法の開発を目指す。

従来手法としては、一般的に行われる人手による波形編集 (波形の描画、部分的ローパスフィルタ処理)、フォルマント修復 [1]、両側波形代入法 [2][3] に基づく近似波形代入法・近似スペクトル代入法、線形予測法 [4] などが挙げられる。人手による波形編集では、音声編集の専門的な知識・経験が必要であることと、人手で行うために自動化が難しいことが問題である。また、[1] では対象音源が母音の音声信号のみに限られている。ここでは、対象音源の性質に関係なくクリッピング信号の修復をすることができるこれらの手法 [2]~[4] を従来手法として扱う。

本研究では対象音源の種類を問わないクリッピング修復法として、RVP (逐次ベクトル射影法) [5] に基づく修復手法を提案し、三種類の従来手法との性能比較を行う。

2 従来手法

クリッピングした音声や音楽波形は、元の波形情報が失われているため、完全な再現をすることは不可能である。しかし、対象とする音響信号の性質や、クリッピングしていない部分の波形を用いることでクリッピング区間をある程度予測することは可能である。ここでは、対象とする信号の性質を問わずクリッピングした音響信号の修復を行う既存の方法を紹介する。

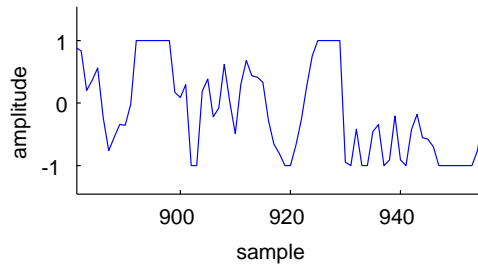


Fig. 1 クリッピング信号の例

2.1 近似波形代入法

近似波形代入法は、両側波形代入法 [2][3] を一部改良し、クリッピング信号に適用させたものである。まず、クリッピング区間の前後の波形に類似する波形をそれぞれ同音源内から検索し、それぞれの類似波形に適切な一次変換を施した後、クリッピング区間に波形をクロスフェードで代入する。類似度は波形をベクトルと考えたときのクリッピング直前または直後の波形と類似波形の2つのベクトルの余弦で評価する。

2.2 近似スペクトル代入法

近似スペクトル代入法は、近似波形代入法での処理を周波数領域に拡張したものである。クリッピング信号をフーリエ変換し、各フレームについて、時間領域でクリッピングが発生しているサンプルを含むフレームを、クリッピングフレームとして扱う。そして、その直前のフレームに最も類似するフレームの次のフレームを (類似フレーム直後のフレームのノルム)/(類似フレームのノルム) 倍クリッピングフレームに代入する。この操作を全ての周波数について行なった後、逆フーリエ変換を施して時間領域に戻す。類似度は近似波形代入法と同様にクリッピング直前のスペクトルと類似スペクトルの2つのベクトルの余弦で評価する。

2.3 線形予測法

クリッピング直前の数サンプルを用いて線形予測を行い、クリッピング部分の真の振幅の予測値を得る。クリッピング区間が2サンプル以上ある場合は既に予測値を代入した1サンプル目の値も用いて線形予測係数を決定する。

*Restoration of clipped acoustic signal, Shin Miura (University of Tsukuba and HRI), Hirofumi Nakajima (HRI), Shoji Makino (University of Tsukuba), Takeshi Yamada (University of Tsukuba), Kazuhiro Nakadai (HRI)

3 従来手法の問題点と提案手法

3.1 従来手法の問題点

前章で紹介した従来手法は、同音源内を探索することに計算時間が必要であることや、代入波形とその前後の波形に連続性がないこと、クリッピング区間が長い場合に誤差が累積してしまうことなどが問題として挙げられるため、修復の性能は実用的ではない。

3.2 RVP に基づく手法

本研究では RVP(逐次ベクトル射影法) [5] に基づくクリッピング音声修復法を提案する。

まず、信号を主要な成分の和として表現するという問題を、線型方程式の条件付最小 2 乗解を求めることとして定式化する。クリッピング前後の一定区間を含んだ区間を N 次元ベクトル y 、信号を表現する M 個の候補信号をベクトル $a_k (k = 1, 2, \dots, M)$ 、信号表現に必要なそれぞれのベクトルの成分 (展開係数) を $x(k)$ 、選択可能な候補ベクトルの数を $L < M$ とすれば

$$y = x_1 a_1 + x_2 a_2 + \dots + x_M a_M + e = Ax + e \quad (1)$$

と表すことができる。ただし、 x は $x(k)$ を成分にもつ M 次元のベクトル、 A は a_k を列にもつ N 行 M 列の行列

$$x = [x_1 x_2 \dots x_M]^T$$

$$A = [a_1 a_2 \dots a_M]$$

である。選択可能な候補ベクトルの数 L の制約条件は

$$S \leq L \quad (2)$$

$$S = \#\{x_k \mid x_k \neq 0\} \quad k = 1, 2, \dots, M$$

と表される。ここで $\#$ は集合の元の数を表し、 S は 0 でない解の要素の数を意味している。ここで、クリッピング前後の一定区間を含む区間を y とし、 y を主要な成分の和として表現することを考える。このとき、 x, A は次のようにして逐次近似的に求められる。

$$u = 0, x(u) = 0, e(u) = y \quad (3)$$

$$p_k = \frac{(e(u), a_k)}{(a_k, a_k)} a_k \quad k = 1, 2, \dots, n \quad (4)$$

$$m = \arg \max_k [\|p_k\| \mid k = 1, 2, \dots, n] \quad (5)$$

$$x_m(u+1) = x_m(u) + \frac{(e(u), a_m)}{(a_m, a_m)} \quad (6)$$

$$x_k(u+1) = x_k(u) \quad (k \neq m)$$

$$e(u+1) = e(u) - p_m \quad (7)$$

$$u = u + 1 \quad (8)$$

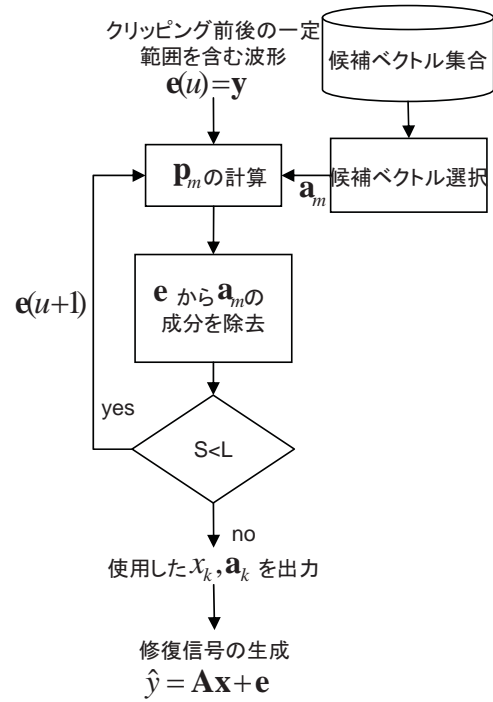


Fig. 2 RVP を用いた方法の処理フローチャート

$$S < L \text{ ならば式 (2) へ、} S = L \text{ ならば終了} \quad (9)$$

ここで、 u は反復回数、 (a, b) は a と b の内積、 $\|a\|$ は a の 2 乗ノルム、 $\max[\cdot]$ は \square 内の最大の要素を意味している。このアルゴリズムを逐次ベクトル射影法 (Recursive Vector Projection: RVP) と呼ぶ。RVP は、全ての候補ベクトルを一度に選択せずに一本ずつ逐次的に選択することによって、式 (2) の条件の下に式 (1) の誤差の 2 乗ノルムを小さくする解を求めるアルゴリズムである。この手法によって選択されるベクトルは、候補ベクトルの全組み合わせの中で最適であるとは限らないが、単純な繰り返し計算で誤差の少ない組み合わせを選択することができる。

RVP は幾何学的に解釈できる。式 (4) は、誤差ベクトル e を A の列ベクトル a_k 方向へ射影したベクトル p_k を求めることを表している。式 (5) は、誤差ベクトルの方向に最も近い方向をもつ A の列ベクトルを選択することに対応している。式 (5) は、誤差ベクトルの方向に最も近い方向を持つ A の列ベクトルを選択することに対応する。式 (6),(7) は、解と誤差ベクトルの更新を表している。式 (5) で選択されたベクトル p_m は、 A の列ベクトル a_m を $\frac{(e(u), a_m)}{(a_m, a_m)}$ 倍することによって表すことができる。言い換えれば、解 x の第 m 番目の係数 x_m に $\frac{(e(u), a_m)}{(a_m, a_m)}$ だけ加算することによって $p(m)$ の成分を Ax で表現できる。これが式 (6) に対応する。解の更新によって e から p_m の成分が減算される。これが式 (7) である。また、反復

Table 1 評価に用いた音声と楽曲

名称	内容
男性音声	数秒の台詞の読み上げ
女性音声	数秒の台詞の読み上げ
野外走行音	走行車のナンバープレートに設置したマイクで録音
メタル音楽	Lost in My Dreams のサビ部分一部
ジャズ音楽	Jive (ピアノソロ) の特に音量が大きい一部分
クラシック音楽	交響曲第 94 番ト長調「驚愕」第一楽章から特に音量が大きい一部分

回数を無限にすれば RVP 解の 2 乗誤差は最小 2 乗解に収束する。

本手法では、この RVP を用いてクリッピング波形の修復を行う。処理フローチャートを Fig.2 に示す。まず、クリッピング前後の一定区間を含む波形を y とする。このとき、クリッピング区間内の値については未知の値として扱うため値を全て 0 とした。また、 a_k については様々な周波数の複素正弦波に加え、区間中央の振幅を三角波で強調した基底を使用した。これによりクリッピング区間の真波形の振幅の再現が期待できる。これを入力として RVP による解析を行った結果出力される Ax は、クリッピング区間にも何らかの補間波形が存在する。これにより修復が可能である。RVP を用いることで高分解能での解析が可能であることや、クリッピング区間とその前後区間に連続性が保たれるという特徴、自由に作成した基底ベクトルを用いることができる特徴から、クリッピングによる波形の変形で生じた不連続性の修復が期待される。

4 修復実験

本手法の性能を評価するために、クリッピングを起こした音声の修復実験を行った。音声や楽曲データを、意図的にクリッピングしたものをいくつか用意し、修復を施す前後での差異を比較することで評価を行った。

4.1 評価に用いたデータ

本評価実験では音声と楽曲のデータを用いた。評価に用いたデータと内容は表 1 の通りである。楽曲データについては RWC 研究用音楽データベース [6] に収録されていたものを用いた。

Table 2 実験時のパラメータ

周波数刻み	0.2Hz
クリッピング区間前サンプル数	50
クリッピング区間後サンプル数	最大 30
基底ベクトル数	クリッピング長と同じ

4.2 実験条件

前節で紹介した音声と楽曲について、振幅の範囲を $[+1, -1]$ に正規化し、振幅のクリッピングは (0.05, 0.10, 0.15, 0.20) の 4 種類にて行った。また、RVP に必要なパラメータは経験に基づき決定した。パラメータを表 2 に示す。

4.3 評価尺度

評価尺度は SNR、PESQ、主観評価の 3 種類を用いた。尚、PESQ は音声の評価のみに使用した。ここでは、それぞれの評価尺度について説明する。

4.3.1 SNR

SNR は、値が大きいほど音質が良く $x(t)$ を真信号、 $\hat{x}(t)$ を修復信号とすると、次式にて表される。

$$SNR(dB) = 10 \log \frac{\sum(x(t))^2}{\sum(x(t) - \hat{x}(t))^2} \quad (10)$$

4.3.2 PESQ

PESQ[7] は、ITU-T P.862 で勧告された音声品質を評価する最も新しい客観評価手法である。原音声と劣化音声を入力することで、客観評価値を出力する。主観評価や、音声認識性能と強い相関があることが明らかになっている。評価値は 0.5~4.5 のスコアで算出され、値が大きいほど音質が良い。なお、今回は帯域を 8kHz に拡張した wideband PESQ を用いた。

4.3.3 主観評価

主観評価は被験者 1 名について行った。実際に受聴することで、波形の観察や他の評価尺度だけでは分からないような、音の様子を聴きとった。

4.4 実験結果・考察

3 章で紹介したそれぞれの手法について実際にクリッピング音声の修復を行った。SNR、PESQ 表 3 に示す。また、本手法適用後の波形の例とスペクトログラムの例を Fig. 3,4 に示す。

従来手法では、近似波形代入法であれば修復波形とその前後に位相の連続性がないこと、線形予測法で

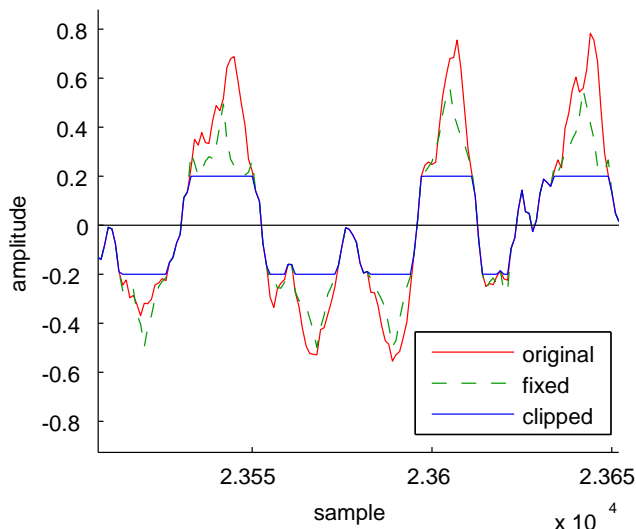


Fig. 3 本手法による修復音声の波形の一例

Table 3 処理前後の SNR と PESQ の改善量

手法	SNR	PESQ 値
近似波形代入法	-1.95	-0.21
近似スペクトル代入法	-0.28	-0.06
線形予測法	-9.91	-0.81
本手法	+2.00	-0.11

はクリッピング区間が長い場合に誤差が累積してしまうことにより SNR、PESQ 共に数値が低下してしまっただと考えられる。受聴時の音質についてもクリックノイズが目立った。本手法では SNR、PESQ 値共に従来手法より評価が上がった。しかし、PESQ 値の差は負の値である。受聴時の音質についてはクリッピングによる雑音が増加したように感じられたがその反面、ホワイトノイズのような雑音を感じられた。Fig.4 のスペクトルからも、全体に渡ってパワーが増し、白色的に変化したように見られる。

5 まとめ

本稿では、音声・楽音双方に適用できるクリッピング信号修復アルゴリズムとして RVP に基づく手法を提案した。クリッピングを起こした信号の修復を試みたところ、主観評価や PESQ についてはあまり変化がなかったものの SNR において回復が見られた。

参考文献

[1] 奥山福太郎, "振幅クリップされた音声信号の修復に関する研究," 東京工業大学機械制御システ

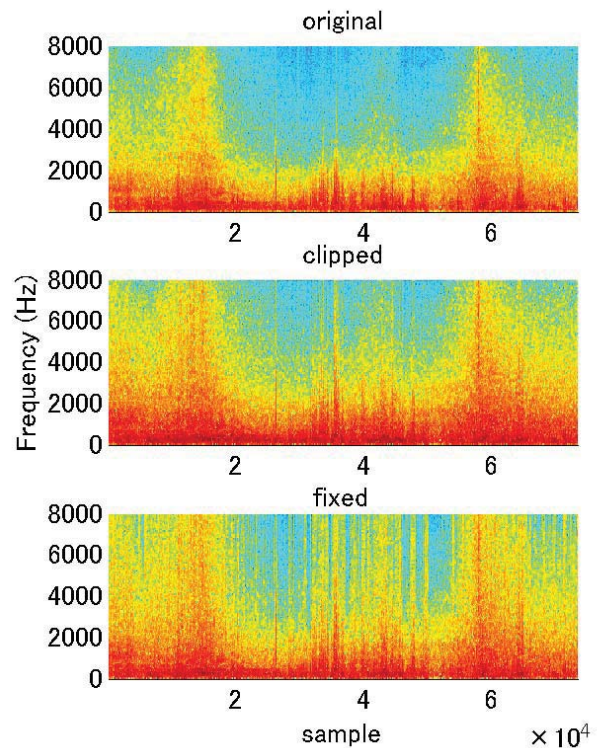


Fig. 4 本手法による修復音声のスペクトルの一例

△専攻修士論文 2004

[2] 唐駿, "雑音区間の推定と波形代入による断続雑音の抑制," 信学技報, EA88-14, 1988.

[3] 唐駿, "雑音区間の高精度検出による断続雑音の除去," 信学技報, EA89-21, 1989.

[4] 北脇信彦編, "音のコミュニケーション工学," コロナ社, 1996.

[5] 中島 弘史, 東山 三樹夫, 田中 雅史, "逐次ベクトル射影法による信号表現とその逆フィルタへの応用," 電子情報通信学会論文誌, Vol. J83-A No. 4, pp353-360, Apr. 2000.

[6] 後藤 真孝, 橋口 博樹, 西村 拓一, 岡 隆一: "RWC 研究用音楽データベース: 研究目的で利用可能な著作権処理済み楽曲・楽器音データベース", 情報処理学会論文誌, Vol.45, No.3, pp.728-738, March 2004.

[7] ITU-T Rec. P.862, "Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs," Feb. 2001.