

図-4 波形での損失計算の概念図。(a) では全ての領域で L_{range} , 黄の領域で $L_{\text{range}}, L_{\text{flat}}$ が適用される。(b) では白の領域で $\tilde{L}_{\text{range}}$, 赤の領域で L_{up} , 紫の領域で $L_{\text{up}}, L_{\text{limit}}$ が適用される。閾値 θ_n での歪み信号を緑と青の破線でそれぞれ示し、元信号を灰色の実線で示した。

学習の損失計算では、(a) 観測信号のデータセットの $\theta_n \in [0, \theta_{\text{clip}}]$ での各出力に加えて、逆発展でのスペクトログラムの過度な変動を防ぐために (b) $\theta_n \in [\theta_{\text{clip}}, 1]$ での各出力も用いる。ただし、 θ_n は $\theta = 0$ を 1 番目としたときの n 番目の閾値を表す。(a), (b) の損失 $L_{\text{clip}}, L_{\text{est}}$ を以下のように定める。

$$L_{\text{clip}} = \alpha_1 L_{\text{range}} + \alpha_2 L_{\text{flat}} + \alpha_3 L_{\text{spec}} + \alpha_4 L_{\text{weight}}$$

$$L_{\text{est}} = \beta_1 \tilde{L}_{\text{range}} + \beta_2 L_{\text{up}} + \beta_3 L_{\text{limit}}$$

ただし $\alpha_1, \dots, \alpha_4, \beta_1, \dots, \beta_3$ は調整パラメータである。 $L_{\text{range}}, \tilde{L}_{\text{range}}, L_{\text{flat}}, L_{\text{up}}, L_{\text{limit}}$ は波形での損失、 $L_{\text{spec}}, L_{\text{weight}}$ はスペクトログラムでの損失である。

波形での損失計算の概念図を図-4 に示す。(a) では、波形の忠実性を促すために、 θ_n で波形打ち切りのない領域での平均絶対誤差 (MAE) L_{range} に加えて、打ち切られた領域での MAE L_{flat} を付加する。(b) では、 θ_{clip} で打ち切りのない領域での MAE $\tilde{L}_{\text{range}}$ およびその領域で絶対値が θ_{clip} 以下の値の MAE L_{up} を付加する。さらに、波形の過度な伸長を防ぐために絶対値が θ_n を超えた値の MAE L_{limit} を付加する。

スペクトログラムでの損失は、既知の観測データがある (a) の場合のみで計算する。このとき、振幅の大きい成分での忠実性を重視するため、観測信号のスペクトログラムの各時間フレームにおける振幅上位 K 番目までのピンでの MAE L_{spec} を付加する。また、歪み成分がより多く含まれる高周波領域の振幅を抑圧するため、 $\theta = 1$ のスペクトログラムのパラボラ重み付き L_1 ノルム L_{weight} [1] を付加する。

モデルの学習は、事前学習と本学習の 2 段階で行う。本学習では $L_{\text{clip}} + L_{\text{est}}$ を損失とする一方、事前学習では L_{clip} のみを損失とする。これにより、観測データのみに基づいてモデルを安定化した後、本学習を行えるようにする。また、過度な抑圧を防ぐために、 L_{weight} のパラメータは本学習で $\alpha_4 = 0$ とする。

4 実験

音楽信号のデータセット [1] を用いて、 L_1 ノルムを用いたスパース最適化に基づくベースライン手法 [1] と提案手法を比較した。サンプリング周波数は 44.1 kHz で、0.25 s を抜粋して用いた。短時間フーリエ変換では 46 ms のハン窓を 75 % オーバーラップで用いた。最適

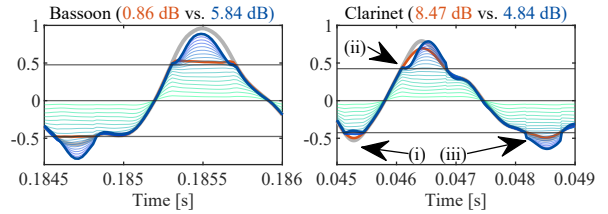


図-5 結果の波形の例。元信号を灰色で、従来手法を赤色で示した。提案手法の $\theta = 1$ の推定信号を青の太線で示し、 $\theta = 0$ から $\theta = 1$ までの推定信号を緑から青で示した。上部に各手法に対応した色で ΔSDR_c を示した。

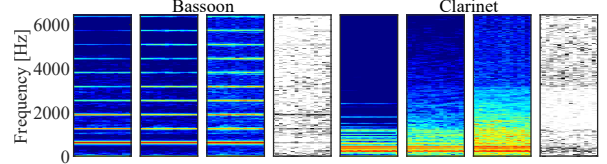


図-6 結果の振幅スペクトログラムの例。各音源で左から元信号、観測信号、推定信号、元信号と観測信号の絶対誤差に対する元信号と推定信号の絶対誤差の比を示した。比は 1 より大きいほど明るく、小さいほど暗く示した。

化では Adam を用い、ステップサイズを 0.01 として、事前学習で 100 反復、本学習で 1000 反復した。各パラメータは $\alpha_1, \alpha_3, \beta_1, \dots, \beta_3 = 10$, $\alpha_2 = 1$, $\alpha_4 = 100$, $K = 100$ とした。閾値のサンプル数は (a), (b) の場合で共に 10 個とした。評価指標は ΔSDR_c [1] とし、元信号の SDR_c は 10 dB として閾値を設定した。

代表例として Bassoon および Clarinet での結果を示す。Bassoon では、図-5 の波形において、提案手法では従来手法と比べて元信号に近い波形が得られた。一方で Clarinet では、提案手法では (i) に示す波形の停滞や、(ii), (iii) に示す矩形状の歪みが見られた。これらは、絶対値が θ_{clip} を超える波形の発展を十分にモデル化できていないことが原因と考えられる。

それぞれの音源での振幅スペクトログラムを図-6 に示す。いずれの音源でも、低周波の主要成分の振幅が観測信号よりも元信号に近くなった一方で、全体的にノイズ成分が見られた。Clarinet では、観測信号の歪み成分がさらに増加していた。これらから、単純な正弦波成分の復元は可能な一方、複雑な成分の復元にはモデルの表現力がより必要であると考えられる。

5 むすび

本稿では、スペクトログラムの変化をモデル化することで、歪み成分の抑圧と波形の伸長を統一的に実現できる手法を提案した。今後は、より表現力の高いモデルを用いて復元性能の向上を図る。

参考文献

- [1] P. Závřiska, P. Rajmic, A. Ozerov and L. Rencker, "A survey and an extensive evaluation of popular audio declipping methods," *IEEE J. Sel. Top. Signal Process.*, **15**(1), 5–24 (2020).
- [2] R. T. Chen, Y. Rubanova, J. Bettencourt and D. K. Duvenaud, "Neural ordinary differential equations," *Int. Conf. Neural Inf. Process. Syst.*, pp. 6572–6583 (2018).