Vol.2021-MUS-132 No.23
2021/9/17

ラウドなポピュラー音楽のダイナミクス復元

尾関 日向[1]　酒向 慎司[1]

概要：ポピュラー音楽の制作では，マスタリングの際に曲の音量レベルを過剰に高めようとする傾向がみられる．しかし，このようにして作られたダイナミクスの小さな曲は，近年のリスニングスタイルに適していないことが多いと考えられる．そこで本研究では，ラウドなポピュラー楽曲のスペクトログラムからマスタリング前のラウドネスを推定することで，ダイナミクスの復元を目的とする．

Dynamics Restoration for "Loud" Popular Music

Abstract: In the production of popular music, mastering engineers tend to excessively increase the volume level of songs. However, those loud songs with low dynamics are often unsuitable for recent listening styles. Therefore, in this study we attempt to restore the dynamics of them, by estimating the short-term loudness before mastering from the spectrogram after mastering.

1. 背景と目的

ポピュラー音楽業界では録音音楽の制作にあたり，楽曲全体のラウドネスの大きさを競い合う傾向がみられる．メディアで放送される際に他の曲よりも大きな音で目立つことで，高い売り上げを狙うためである．こうした競争はラウドネス・ウォー [1] と呼ばれ，1990 年代後半から 2000 年代初頭にかけて急速に広まったとされている [13].

より大きなラウドネスを追求した結果，曲中で常に大音量を維持するような楽曲が多く生産された．ラウドネスの増大は，主にマキシマイジングと呼ばれる，録音後のデジタル信号処理によってなされる．本稿では**図 1(a)** に示す波形のように，過剰なマキシマイジングによってダイナミクス（曲中の音量変化）が失われた状態を「ラウドである」と表現する．

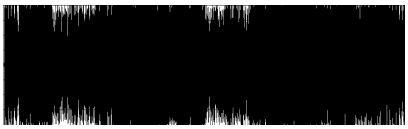
ラウドな曲の問題点として，聴取する際の音楽的な情感・興奮の乏しさや，継続的に大きな音にさらされることによる聴覚疲労などが指摘されており [1]，また音楽の売り上げとラウドネスの間に関係がないことも示されている [1]. ノイズレベルの大きな視聴環境においてはラウドな楽曲のほうが好まれることがわかっている [2] が，近年はノイズキャンセル機能付きのオーディオ機器や遮音性の高いヘッドホンが普及し，静かなリスニング環境が容易に手に入るようになった．さらに，主要な音楽ストリーミングサービスでラウドネス正規化 [4][5] が行われるようになり，再生時に曲どうしの聴感的な音量差が抑制されることで，「他の曲より目立つ」というラウドな曲の持つ強みは失われてしまった．

このように，ラウドネス・ウォー期に生まれたラウドな楽曲は現在の静かでストリーミング主体のリスニングスタイルに適しているとはいえない．2010 年 3 月にダイナミクスを重視した音楽制作を勧めるキャンペーンとして「ダイナミックレンジデー [3]」が開催されていることが示すように，近年はラウドネスよりもダイナミクスのある音楽体験への需要が高まってきているといえる．そこで著者は既に発表されたラウドな曲に対し，過剰なマキシマイジングが施される前のダイナミクスを復元できる技術が特に有効であると考えた．その取り組みの最初の段階として，本研究ではデジタル信号処理によって，ラウドなポピュラー楽曲に対する自動的なダイナミクス復元を試みる．

2. 提案手法

既発表曲のダイナミクスを復元するにあたって，音楽音響信号から元のダイナミクスの手がかりとなるような音響特徴量を取り出し，短時間フレームごとに音量を再調整するという手法を検討する．これまでに音楽のダイナミクス復元に関する直接的な研究は著者の知る限りでは存在しな

[1] 名古屋工業大学　大学院工学研究科
Graduate School of Engineering，Nagoya Institute of Technology.


(a) ラウドネスの大きな曲の波形（一例）

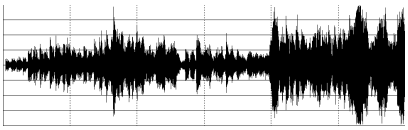

(b) ダイナミクスの大きな曲の波形（一例）

図 1 波形でみたときのラウドネスとダイナミクス

いため，より広く，音楽情報処理全般に従来用いられてきた手法を参考とした．

2.1 音の大きさを表す尺度

ラウドネスとは一般的には音の大きさを表す語であるが，ヒトが感じる音の大きさとして ITU（国際電気通信連合）や EBU（欧州放送連盟）によって規格化された知覚量 [6] の名称でもある．ラウドネスは信号の単純なピークレベルや RMS 値と違いヒトの聴覚特性を考慮した尺度であるため，音楽や音声の大きさを測るのに広く利用されている．先述したラウドネス正規化も，この規格化されたラウドネスに基づくものである．

本研究でもラウドネスに基づいて音量を計測する．また以降では，ラウドネスは規格化されたものを指す．

2.2 音響特徴量

音楽にダイナミクスが生まれる要因として
(1) 演奏技法に基づく，個々の楽器の音量変化
(2) 楽曲構造に基づく，楽器の数や種類の変化
の 2 つが挙げられる．一般的にポピュラー音楽において (1) は録音時や編集の初期段階の信号処理によって抑えられるため，(2) がダイナミクスをもたらす主要因だと考えられる．楽器の音はその種類ごとに時間周波数領域において特有のパターンをもつため，(2) はスペクトログラムに顕現するはずである．実際に楽曲構造解析では特徴量としてスペクトルの類似度行列が主に用いられている [8][11][12]．

よって本研究ではダイナミクス復元の手がかりとなる音響特徴量としてスペクトログラムに着目し，なかでもヒトの聴覚特性を考慮したメルスペクトログラムを使用する．

2.3 ダイナミクス復元システム

本システムの要となる，音響特徴量からダイナミクスを復元する箇所には，現在の音楽情報処理で主流といえるディープニューラルネットワークを用いる [7][14]．ダイナミクスの変化はある程度の長時間のなかに表れる特徴であるため，長期的な時系列データの処理に長けた LSTM (Long short-term memory) ネットワークが適している．

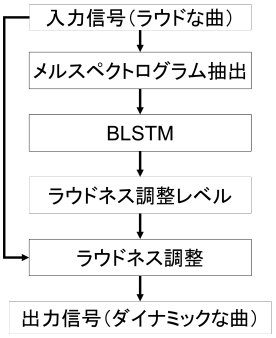


図 2 ダイナミクス復元システムの構成

表 1 ラウドネス調整ラベル

クラス	ラウドネス調整分 (dB)
0	-0.0
1	-1.0
2	-2.0
3	-4.0
4	-8.0

そのなかでも本システムはリアルタイム処理の必要がなく，未来の楽曲展開がダイナミクス復元にあたり有益と考えられるため，BLSTM (Bidirectional LSTM) を採用する．

システムの構成を**図 2** に示す．ラウドな曲のメルスペクトログラムのブロック列を BLSTM の入力とし，BLSTM は短時間フレームごとのラウドネス調整レベルを示す配列を出力する．その結果に基づき入力信号のラウドネスを調整して最終的な出力信号を得る．ただし本稿の評価実験段階では簡単のためラウドネス調整レベルを**表 1** のように 5 段階に限定し，マルチクラス分類問題に帰着させている．

3. データセット

ダイナミクス復元モデルの学習にあたって，正解データであるダイナミックな曲を用意する必要がある．しかし実際に売られているラウドな曲に対し，正解となるマキシマイジング前のデータは一般的には入手できない．そこで今回はもともとダイナミックな曲に対しマキシマイジングを行うことで，入力と正解の出力となるマキシマイジングの前後の楽曲データのペアを作成する．

ダイナミクスのあるポピュラー楽曲群として，RWC 研究用音楽データベースに含まれるポピュラー楽曲 100 曲（RWC-Popular）[9] を用いる．ただし今回は簡単のため固定長の系列に限定し，各曲の冒頭 90 秒のみを使用する．RWC-Popular の楽曲に対し，ダイナミクスの大きさの指標の一つである DR (Dynamic Range) 値 [8] を計測すると**図 3** のようになる．DR 値は 0 から 14 までの 15 段階のスケールを持ち，一般にラウドな曲の DR 値が 6〜7 以下とされている [3] ことから，RWC-Popular の楽曲群は比較的ダイナミックであることがわかる．

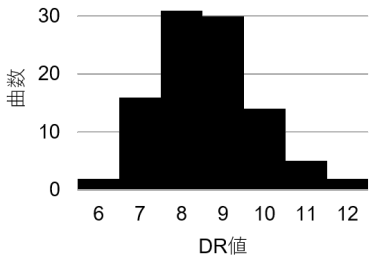
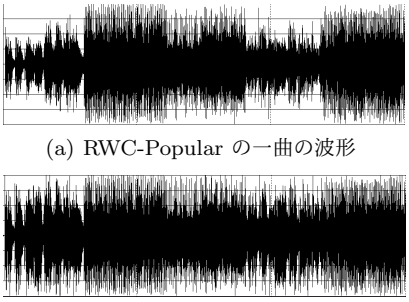


図 3 RWC-Popular データセットの DR 値のヒストグラム



(a) RWC-Popular の一曲の波形

(b) マキシマイジング後の波形

図 4 DynAudNorm によるマキシマイジング例

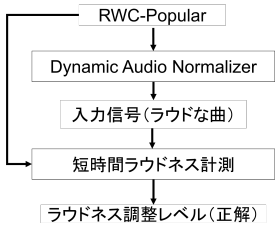


図 5 データセットの作成フロー

3.1 マキシマイジング

人手で 100 曲分のマキシマイジングを行うのは困難なため，音量正規化フィルタとして DynAudNorm (Dynamic Audio Normalizer)[10] を用いて自動化する．このフィルタは音声動画変換ソフトウェアとして広く利用されている ffmpeg に実装されているものである．DynAudNorm によって**図 4** に示すように，曲中の音量が小さい箇所が上がり，曲全体のラウドネスを高めることができる．コンプレッサーやリミッターと呼ばれる一般的なマキシマイジングツールによる自動処理と比較して波形の歪みが発生しないため，人手で丁寧にマキシマイジングした結果と近い出力が得られると考えられる．

3.2 正解ラウドネス調整レベルの取得

図 5 のように DynAudNorm によるマキシマイジング後の曲と，対応するもとの RWC-Popular の曲で短時間ラウドネスの差分を取り，正解となるラウドネス調整レベルを得る．

4. 評価実験

提案するダイナミクス復元システムの有効性を試験的に測るにあたり，後述のような設定で実験を行った．

4.1 実験条件

メルスペクトログラム

FFT の窓幅を約 0.74 秒，ホップ幅をその 1/4，メルフィルタ数を 64 とし，前処理として 8 kHz のローパスフィルタをかけた．

ラウドネス調整レベル

時間フレーム幅を 1 秒，ホップ幅をその 1/2 とした．

ニューラルネットワーク構成

48 次元の隠れ状態ベクトルを持つ 1 層の BLSTM 層と，その後段の 1 層の全結合層から成る．

学習条件

損失関数に平均二乗誤差，オプティマイザに学習率 0.001 の ADAM を使用した．またデータセットは 3 : 1 の比率で学習データとテストデータに分け，さらに学習データの 1/5 を検証データとした．振り分けの際にはサブセット間で DR 値に偏りが出ないよう調整した．

4.2 実験結果

推定されたラウドネス調整レベルの正解率は 58.5 %，隣接したクラスへの誤りを許容すると 94.6 %であった． 混同行列は**図 6** のようになった．もともと正解のラウドネス調整レベルには「調整の必要なし」を示すクラス 0 に該当するフレームが全体の 85 %以上を占めており，逆に「ラウドネスを大きく下げる必要あり」のクラス 3 とクラス 4 は合計で僅か 2.5 %である．そのため正解率の高さの割には実際の分類精度は決して良いものとは言えない．

また，ダイナミクス復元前後の短時間ラウドネスの比較結果の例を**図 7** に示す．RWC が もともとの RWC-Popular のラウドネス，DAN が DynAudNorm によってマキシマイジングした後のラウドネス，DynGen が提案システムによって推定されたラウドネスを示す．各曲のイントロに注目すると，マキシマイジングによるラウドネスの増加の有無にかかわらず DynGen のラウドネスが減少している．また**図 7(a)** や**図 7(b)** を見ると，イントロより後にあるラウドネスの要調整箇所では DynGen に変化が見られない．ここから，今回の評価実験では，一様にイントロ部分のラウドネスを下げるようなネットワークが生成されてしまい，ラウドネスの過大な部分の音響的特徴を正しく学習できなかったことがわかる．原因としては先述のようなクラス 3 やクラス 4 に該当するデータの不足や，シンプルなニューラルネットワークの構成のため，スペクトルからダイナミクス復元のてがかりとなる十分な特徴を発見できな

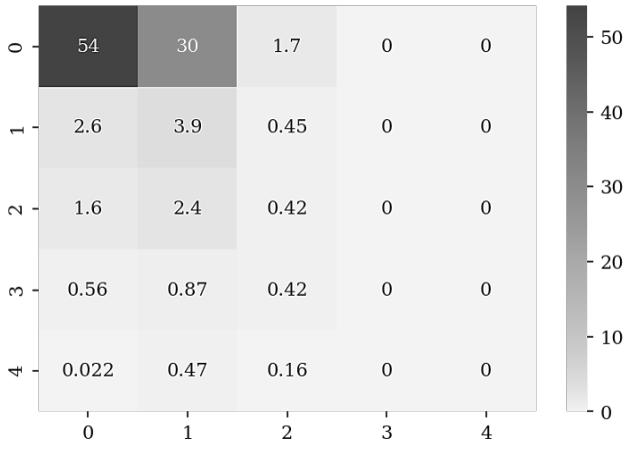


図 6 ラウドネス調整レベルの混同行列

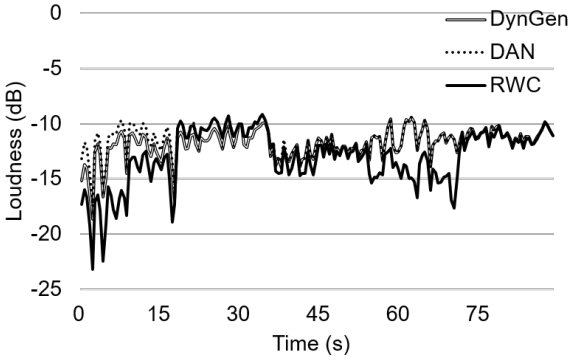
かったことが考えられる．

5. むすび

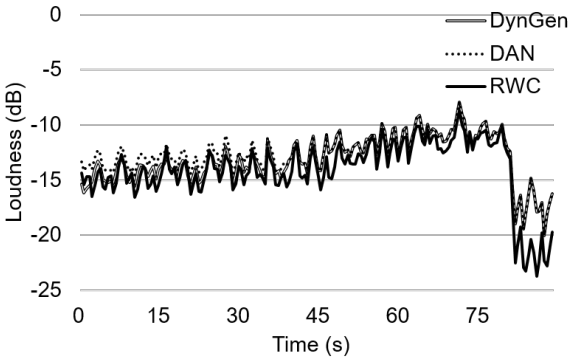
本稿ではダイナミクスの自動復元システムとして，マルチクラス分類問題として楽曲のスペクトルからラウドネス調整ラベルを推定する手法を提案し，評価実験を行った．結果は満足できるものではなく，本研究はまだまだ手探りの段階だといえる．今後は分類精度向上のため，学習に適した入力データの選別・抜粋やニューラルネットワーク構成・学習条件の調整を行う必要がある．ダイナミクスの変化点となる可能性の高い楽曲構造情報を取り入れることも検討している．また，主観評価のための被験者実験を実施し，正解の存在しない曲に対する本システムの有効性を調査したい．

参考文献

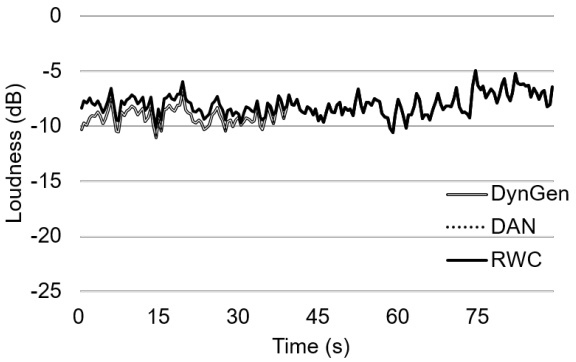
[1] Earl Vickers: *The Loudness War: Background, Speculation and Recommendations*, 129th Audio Engineering Society Convention, pp.135-161, 2010.
[2] Earl Vickers: *Loudness Normalisation: Paradigm Shift or Placebo for the Use of Hyper-Compression in Pop Music?*, ICMC—SMC—2014, pp.14-20.
[3] Mastering Media (Production Advice) Ltd: *Dynamic Range Day*, `https://dynamicrangeday.co.uk/`, 2021.
[4] Apple Computer, Inc.: *Apple Digital Masters: Music as the Artist and Sound Engineer Intended*, 2021.
[5] Spotify AB: *Spotify for Artists: Loudness normalization*, 2021.
[6] ITU-R: *Recommendation ITU-R BS.1770-4: Algorithms to measure audio programme loudness and true-peak audio level*, 2015.
[7] 阪上大地: 音楽情報処理のための深層学習, 研究報告音楽情報科学（MUS）, 2018-MUS-119(2), pp.1-13.
[8] *Dynamic Range DB*, `https://dr.loudness-war.info/`
[9] 後藤真孝: *RWC* 研究用音楽データベース, `https://staff.aist.go.jp/m.goto/RWC-MDB/index-j.html`.
[10] LoRd_MuldeR: *Dynamic Audio Normalizer*, `https://github.com/lordmulder/DynamicAudioNormalizer`.



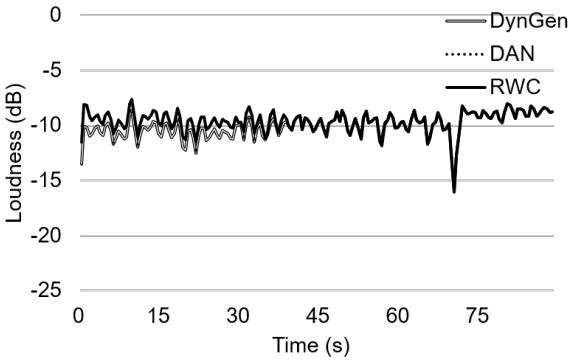
(a) マキシマイジングによるラウドネス増加が大きい曲 (1)



(b) マキシマイジングによるラウドネス増加が大きい曲 (2)



(c) マキシマイジングによるラウドネス増加が小さい曲 (1)



(d) マキシマイジングによるラウドネス増加が小さい曲 (2)

図 7 ダイナミクス復元前後の短時間ラウドネスの比較（例）

[11] Jouni Paulus, Meinard Muller and Anssi Klapuri, *Audio-based music structure analysis*, 11th International Society for Music Information Retrieval Conference, ISMIR

2010.
[12] 亀岡弘和, 中村友彦, 高宗典玄: 音楽音響信号処理技術の最先端, 電子情報通信学会誌, vol.98, pp.467-474, 2015.
[13] Emmanuel Deruty and Damien Tardieu: *About dynamic processing in mainstream music*, AES: Journal of the Audio Engineering Society, vol.62, pp.42-56, 2014.
[14] Keunwoo Choi, György Fazekas, Kyunghyun Cho and Mark Sandler, *A Tutorial on Deep Learning for Music Information Retrieval*, arXiv:1709.04396v2, 2018.