

環境音による音響的橋渡しを用いた楽曲遷移手法の提案

藤本 直樹^{1,a)} 山西 良典^{1,b)}

概要：音楽ストリーミングサービスの普及により、プレイリストによる連続的な音楽聴取が一般化している。プレイリストによる音楽聴取は大量の楽曲を時間的に組み合わせ、任意の楽曲体験を得るという音楽との新たな関わり方を生み出した。しかし、1つのプレイリスト内では隣り合う楽曲間で楽曲特徴に親和性をもたせるような曲順に偏る傾向がある。本研究では、楽曲間に環境音を“音響的橋渡し”として機能させ、遷移時の印象形成を最適化するアプローチ「環境音 Bridge」を提案する。環境音を持つ特定の風景を想起させる効果（サウンドスケープ）によって、楽曲間の楽曲特徴量（テンポ、調性、音圧等）の差異を緩和させ、本来は接続されないような楽曲から構成されるプレイリストの構築を目指す。本稿は、環境音 Bridge の実現に向けた端緒として、環境音と楽曲、印象評価の組み合わせからなるデータセット構築について、スキーム策定とデータ収集における課題の整理を述べる。

1. はじめに

音楽ストリーミングサービス (Spotify^{*1}, AppleMusic^{*2} など) の普及により、ユーザは複数の楽曲を自由に組み合わせ連続的に再生する“プレイリスト”を日常的に利用している。プレイリストによる音楽聴取は大量の楽曲を時間的に組み合わせることで多様な音楽体験をもたらす一方で、特徴が大きく異なる楽曲を接続する場合に“楽曲ペアの衝突”が発生することが指摘されている [1], [2], [3]。具体的な例として、ゆったりしたジャズと激しいメタルを連続再生するなど、ジャンルの飛躍が大きいプレイリストでは音楽体験の一貫性が損なわれ、リスナーは強い違和感や不快感を感じることが報告されている [4], [5]。こうした衝突の原因としては、テンポやキー、音色、楽器構成の不一致などが挙げられ、既存の研究ではこれらの要因を小さくすることでスムーズな遷移を実現する手法が提案されてきた [6], [7]。しかし、これらの手法は類似度の高い楽曲を集めることを企図する傾向があり、プレイリスト内での楽曲で全体の多様性が低下する問題が指摘されている [8]。

環境音（自然音や都市音など）について、既存研究 [9], [10] によれば、単に音としての要素だけでなく、特定の空間イメージ（サウンドスケープ）や心理効果を強く喚起する特徴があるとされる。これは都市空間のデザインやセラピー、

ストレス軽減の実験などで広く活用されており、自然音を聴くだけでもリラクセス効果が認められるという研究報告もある [11], [12]。

本研究では、楽曲の間に環境音を挿入して楽曲の衝突を緩和するアプローチ「環境音 Bridge (Bridge)」を提案する。Bridge では環境音を用いて楽曲の連続再生時における印象形成を効果的にデザインし、遷移時の違和感や没入感の低下を低減することを目指す。特徴が大きく異なる2曲のあいだに特定の環境音（雨音や雑踏音など）を挟むことで、楽曲が持つ要素の不一致の緩和をねらう。例として“テンポの遅いジャズ曲からアップテンポの EDM 曲へ移行するとき、ジャズの終了後にバーの店内雑音や夜の街の雑踏音を数秒入れ、その後に EDM をフェードインさせる”ことで、夜の街を移動するような一貫したストーリーを感じさせるといったユースケースを想定する。

以上の目的に向けた端緒として、Web 実験システムによる印象評価を実施し、楽曲ペア情報、環境音特性、主観評価、遷移パラメータを含む統合データセットを構築する。本稿では、このデータセット構築に向けたポジションペーパーとして位置づけられ、データセットのスキームやデータ収集アプリケーションでのインタラクションデザインを示すとともに、本研究の Scope と課題について議論する。

2. 提案手法

2.1 提案する楽曲遷移手法

本研究では、

楽曲 A → (環境音 Bridge) → 楽曲 B

のように、楽曲 A の終了部分と楽曲 B の開始部分の間に

¹ 関西大学
Ryozenji 2-1-1, Takatsuki, Osaka 569-1095, Japan
^{a)} k503587@kansai-u.ac.jp
^{b)} ryama@kansai-u.ac.jp
^{*1} <https://open.spotify.com/> (2025 年 2 月 20 日確認)
^{*2} <https://music.apple.com/> (2025 年 2 月 20 日確認)

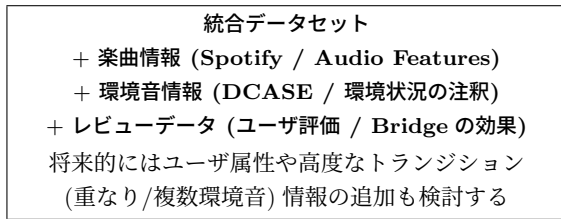


図 1 本研究で構築する分析基盤データセットの概要



図 2 実験用アプリのインターフェース

環境音を挿入し、特徴の異なる 2 曲を緩やかに接続するアプローチ (Bridge) を提案する。

- (1) 楽曲 A をフェードアウトしつつ終了させる
 - (2) 短時間 (数秒程度)、環境音のみを再生する
 - (3) 楽曲 B を開始し、環境音をフェードアウトする
- これにより楽曲間に環境音から想起される特定の情景演出を心理的緩衝材として挿入し、1 章で述べた楽曲同士の衝突要因 (テンポ差、ジャンル飛躍など) に対しても、楽曲間での印象の断絶を緩和させる。

2.2 データモデル

Bridge の効果検証や最適化には、図 1 に示すような楽曲情報、環境音情報、および遷移時の主観評価 (レビューデータ) を一元的に管理するデータスキームが必要である。これは、「どの楽曲同士を、どの環境音で繋ぎ、そのとき被験者がどのように感じたか」というレコードを蓄積・参照できる仕組みを意味する。本研究では、「どの楽曲同士を、どの環境音で繋ぎ、そのとき被験者がどのように感じたか」というレコードを蓄積・参照できる仕組みの確立を目指し、以下の 3 要素に基づくデータモデルを提案する。

- **楽曲情報**: BPM, キー, 音色, ムードなど。衝突の要因となりうる属性で、楽曲ペアごとの差分を把握するために不可欠。
- **環境音情報**: シーンラベル (bus, cafe 等), 音源ファイル (約 10 秒), 想起情景ラベルなど。どの環境音がどん

表 1 Spotify Web API から取得する主な楽曲情報の例

属性	型 / 範囲	説明
name	string	トラック名
id	string	Spotify 内でのトラック ID
popularity	integer (0-100)	トラックの人気度
duration_ms	integer (ミリ秒)	トラックの長さ
preview_url	string (URL)	試聴用オーディオへのリンク
energy	float (0-1)	エネルギー指標
valence	float (0-1)	曲のポジティブさ指標
tempo	float (BPM)	テンポ (BPM)

表 2 DCASE(CochlScene) が提供する環境音情報

属性	型 / 範囲	説明
scene_label	string	bus, cafe, car など 13 種
file_format	wav	音声データのフォーマット
duration	10 sec	クリップ長
annotations	string	Crowdsourced(Weak ラベル)

表 3 評価項目

評価対象	設問	回答形式
知覚的連続性	曲間がどの程度スムーズか	7 段階評価
感情的連続性	感情が自然に繋がったか	7 段階評価
文脈適合性	環境音が印象形成に作用したか	7 段階評価
補足チェック	違和感の有無や再利用意向など	Boolean 等

なムードや空間を演出するかを管理し、楽曲間の「物語性」や「心理的緩衝」を検討する指標とする。

- **レビューデータ (印象評価)**: 知覚的連続性, 感情的連続性, 文脈適合性など (表 3)。どのペア×環境音の組み合わせが衝突を緩和するかを定量化。将来的にはユーザ属性も個人差分析に活用。

2.3 評価実験の設定

本研究の初期段階では、図 2 に示すインターフェースを持つ Web アプリを用いた印象評価実験を行う。楽曲ペア (A, B) と複数の環境音候補を組み合わせ、ユーザ (被験者) にクロスフェード時間などのパラメータを調整しながら試聴してもらい、表 3 の項目を 1~7 段階スライダーやチェックボックスで回答させる。また、比較実験の対照として無音やホワイトノイズを挿入するパターン [13] を用意し、サウンドスケープを伴う環境音の効果との比較をおこなう。

- **楽曲情報** Spotify Web API を用いて、テンポ (tempo), エネルギー (energy), ヴァレンス (valence) などの音響特徴量やメタデータを取得する (表 1)。楽曲同士の差分 (BPM や調性など) を数値化し、衝突の度合いや相性を評価する指標として用いる。

- **環境音情報** DCASE の CochScene データセット*3を

*3 <https://dcase-repo.aalto.fi/?data=scenes/cochscene>

使用する。本データセットには bus, cafe, car など日常的なシーンを想定した 13 種のシーンラベル付きクリップが収録されており (表 2), 日常的な音楽体験を想定したプレイリストの作成に利用しやすい。環境状況についてクラウドソーシングを用いてラベルが付与されているため, 想起させる情景が明確であり, オンラインでの実験にも適している。

- **レビューデータ (印象評価)** 被験者が楽曲 A と B の間に環境音を挿入したパターンを聴取し, どの程度スムーズに感じられたか (知覚的連続性), 感情的な落差が緩和されたか (感情的連続性), 文脈として適切に感じられたか (文脈適合性) などを評価する (表 3)。どのような楽曲ペアとどの環境音の組み合わせが衝突を緩和するかを定量的に分析する。

これらの情報を統合管理し, 「楽曲 A, B の特徴」「挿入した環境音のタイプ」「ユーザ評価 (7 段階評定)」「クロスフェードパラメータ」をひとつのレコードとして保存する。このデータセットは, 後続の解析 (衝突度合いの推定, Bridge の最適化など) に活用される。

2.4 スコープ設定と制限

本研究の目的は, 環境音による楽曲間の音響的橋渡しが不適合や衝突をどの程度緩和できるかを示し, データ分析やアルゴリズム開発の基盤を整備することである。そのため以下の点は本研究では扱わないか, 極めて限定的に扱う。

- **細粒度な環境音ラベルや自動生成:** 「穏やかな波音」「荒い波音」などの細かいバリエーションの区別, および環境音を自動生成する技術はスコープ外とする。
- **個人差への最適化:** ある被験者にとって心地よい音でも, 他の被験者には不快感を与えるなど, 環境音評価には大きな個人差がある。本研究ではまず一般的な傾向の把握を目標とし, 個別最適化は行わない。
- **大規模実験の負荷:** 多様な楽曲ペア×環境音を評価すると被験者の負荷が高まるため, 本稿では小規模実験でのデータ収集設計を中心に検討する。

3. おわりに

本稿では, 楽曲間の大きな特徴量差や雰囲気差を緩和し, 連続的かつ多様な音楽体験を提供するために, 環境音を“音響的橋渡し”として用いる手法を提案した。本稿では, 音響的橋渡しの実現するためのデータセット構築に向けたポジションペーパーとして, データスキームの策定と課題の整理を行った。

今後は, 大衆音楽配信サービスから取得可能な楽曲情報と環境音分析に関する研究コミュニティが提供する環境音データを統合する。さらに, ユーザの遷移評価データを加えることで, どのような曲同士をどの環境音で繋ぐと効果的かを包括的に分析する枠組みを構築していく。

参考文献

- [1] Chen, B.-Y., Hsu, W.-H., Liao, W.-H., Ramírez, M. A. M., Mitsufuji, Y. and Yang, Y.-H.: Automatic DJ Transitions with Differentiable Audio Effects and Generative Adversarial Networks (2022).
- [2] Bittner, R. M., Gu, M., Hernandez, G., Humphrey, E. J., Jehan, T., McCurry, H. and Montecchio, N.: Automatic Playlist Sequencing and Transitions, *International Society for Music Information Retrieval Conference*, (online), available from <https://api.semanticscholar.org/CorpusID:11244249> (2017).
- [3] Karakaya, Z. and Ergül Aydın, Z.: Mathematical modelling and a greedy heuristic for harmonic mixing and popularity-based playlist generation problem, *Journal of Turkish Operations Management*, Vol. 8, No. 2, p. 487–496 (2024).
- [4] Liu, Y.: Deep Learning Based Music Recommendation Systems: A Review of Algorithms and Techniques, *Applied and Computational Engineering*, Vol. 109, pp. 17–23 (2024).
- [5] Liu, Y.: Deep Learning Based Music Recommendation Systems: A Review of Algorithms and Techniques, *Applied and Computational Engineering*, Vol. 109, pp. 17–23 (2024).
- [6] Bernardes, G., Davies, M. E. P. and Guedes, C.: A Hierarchical Harmonic Mixing Method, *Music Technology with Swing*, Cham, Springer International Publishing, pp. 151–170 (2018).
- [7] Pauws, S. and Eggen, B.: PATS: Realization and user evaluation of an automatic playlist generator, Vol. 32 (2002).
- [8] Schedl, M., Zamani, H., Chen, C.-W., Deldjoo, Y. and Elahi, M.: Current challenges and visions in music recommender systems research, *International Journal of Multimedia Information Retrieval*, Vol. 7, No. 2, p. 95–116 (2018).
- [9] Schafer, R. M.: *The Soundscape: Our Sonic Environment and the Tuning of the World*, Knopf, New York (1977).
- [10] Gaver, W. W.: Auditory icons: using sound in computer interfaces, *Hum.-Comput. Interact.*, Vol. 2, No. 2, p. 167–177 (1986).
- [11] Lin, H.-Y.: The Effects of White Noise on Attentional Performance and On-Task Behaviors in Preschoolers with ADHD, *International journal of environmental research and public health*, Vol. 19, No. 22, p. 15391 (2022).
- [12] Aletta, F., Guattari, C., Evangelisti, L., Asdrubali, F., Oberman, T. and Kang, J.: Exploring the compatibility of “Method A” and “Method B” data collection protocols reported in the ISO/TS 12913-2:2018 for urban soundscape via a soundwalk, *Applied Acoustics*, Vol. 155, pp. 190–203 (2019).
- [13] Pascoe, A. J., Haque, Z. Z., Samandra, R., Fehring, D. J. and Mansouri, F. A.: Dissociable effects of music and white noise on conflict-induced behavioral adjustments, *Frontiers in Neuroscience*, Vol. 16 (2022).