Vol.2020-MUS-127 No.28
Vol.2020-SLP-132 No.28
2020/6/7

Life-M: ランドマーク画像を題材としたフリーの音楽コーパス

大森 陽[2,a)]　小口 純矢[2,b)]　高道 慎之介[3,c)]

概要：本稿では，画像を題材とした音楽コーパスを構築する．機械学習に恩恵を受けたモーダル間の深い情報交換に関する研究が期待される．画像と音楽における情報交換研究を見据え，本稿では，ランドマーク画像を題材としたフリー音楽コーパス Life-M (Landmark Imaged-themed Free Music Corpus) を設計する．本コーパスは画像と音楽のペアから成り，音楽データは単一作曲者により作曲される．本稿ではその設計と構築結果を述べる．

キーワード：フリーコーパス， 画像，音楽情報処理

1. はじめに

深層学習技術の発展により，複数のモーダルの情報を統一的に扱えるようになりつつある．これにより，言語・音・画像などのモーダル間における情報交換が加速すると予想される．相互作用 [1] が古くから知られている動画像と音においては，人工現実感（バーチャルリアリティ）向上のために，深層学習を用いて背景画像から単純な音響シーン [2]・音響イベント [3] を生成する研究が行われている．他方，モーダル間情報交換ではなく純粋な楽音生成の文脈においても，深層学習の貢献は大きい．音楽特有の構造，例えば，音声と異なり複数のノートが同時に発生しうること，中・長期的な繰り返しを持つことを考慮した手法が多数提案 [4] されている．本稿では，画像からの楽音生成，すなわち，音楽的構造を持った音信号を画像から生成するタスクを考える．これは，単純な音響シーン・音響イベントに限定されない視聴覚印象や人工現実感制御に向けたタスクとみなされる．

その実現に向け本研究では，フリーの画像コーパスである Google landmark dataset [5] のランドマーク画像を題材とした音楽コーパス Life-M (Landmark Imaged-themed Free Music Corpus) を設計する．本コーパスは画像と音楽のペアから成る．さらに本コーパスは，以下のような特徴を持つ．

単一作曲者による作曲：全ての楽曲は，単一作曲者により作曲されているため，統一された音楽性をもつ．これは，より精度の高い音楽構造学習や，機械学習を用いた作曲者適応などに利用できる．

世界遺産画像に限定した作曲：画像ドメインを限定した作曲により，画像ドメイン適応を利用した機械学習に利用できる．

作曲情報： 楽譜情報だけでなく作曲時の付随情報（音楽ジャンルやスケール）を含むことで, 音楽情報処理研究にも利用できる. 音楽においては, 音響的なコーパスとしても記号論的な楽曲分析にも用いることができる．

研究用途に限りフリーで使用可：本コーパスは，研究用途であれば学術機関だけでなく民間企業においても利用可能である．

以降では，本コーパスの設計方針と構築結果を述べる．本コーパスはウェブページ上からダウンロード可能である [6].

2. コーパス作成

2.1 画像の選択

ランドマークは，国・地域を象徴する重要性のある，もしくはシンボル的なモニュメント・建造物や，山や湖などの地形が多く選ばれる．しかし本来は，地理学上の特徴物として，初めてその地を訪れる人のための目印としての役割が主である．例えば，方角を知るための鉄塔のような，特

1 早稲田大学大学院
Waseda University, , Tokyo 164–8525, Japan
2 明治大学大学院
Meiji University, Nakano, Tokyo 164–8525, Japan
3 東京大学 大学院情報理工学系研究科
The University of Tokyo, Hongo, Tokyo, 113-8656, Japan.
a) yo_ohmori@suou.waseda.jp
b) ev50552@meiji.ac.jp
c) shinnosuke_takamichi@ipc.i.u-tokyo.ac.jp

ⓒ 2020 Information Processing Society of Japan

徴や理由が無いもの，ほとんど知られていないようなマイナーなランドマークが多数ある．それらは，作曲者にとって連想されるイメージが少ないため楽曲制作が難しく，加えてより有名なランドマークが多い方がコーパスを利用するにあたってもより有益である．上記の理由から我々は，世界遺産に登録されているランドマークを選び楽曲を制作した．また，ランドマークにおいては多数の種類があるため，建築物のみに絞った．

2.2 作曲者に与えられる情報

本研究では，画像と音楽のペアデータ作成を目的としているため，画像と音楽以外の背景情報の影響を可能な限り排除すべきである．そのため，楽曲制作にあたっては，画像に加え，以下の必要最低限の背景情報のみを利用した．詳細は，2.3 節で述べる．

- ランドマーク画像
- 国名及び地名
- 建造された時代
- 建築物の用途

2.3 作曲手順

作曲者は上記の情報から以下のようなイメージを連想し作曲を行う．

- ランドマーク画像
  - 荘厳，華美，規模の大きさなどの素直に感じられた建造物自体への印象．
  - 建築様式．ルネサンス様式やゴシック建築など歴史的に変遷のある建築様式は，視覚的に特徴的であり，芸術史においても音楽史と深く結びついている [7–10] ためイメージを膨らませるために有益である．
  - 建造物周辺のその国や地名特有の街並みや地形，昼夜や天候による印象．
- 国名及び地名
  - 国や地域にはそこから想起される音楽性が存在する．ランドマークから想起される音楽性はそういった国や地域の音楽性の影響を少なからず受けているはずであるため，画像の印象に合う楽曲を制作するにあたり，当時のクラシック音楽及び作曲家, 大衆音楽などが連想されることが多い．連想された音楽を理解するために，その国・地方特有の舞曲，旋法や和声進行，使用楽器などをサーベイ [11] し，なおかつ現在の画像建築物に齟齬が生じないように，それらの要素を組み合わせて作曲を行う．
- 建造された年号
  - 一部の歴史的建造物は，災害や戦争，政治的な利権などにより，取り壊しや建て直しがあるため [10,12–14]，画像に写っている現在の形に至った年号を情報として与えられる．
  - グレゴリオ聖歌などのモノフォニーから古典対位法，バッハ以降の対位法，和声法，そして後期ロマン派や近現代クラシックなど，歴史的変遷の中で音楽理論は変化している [15–17]．ランドマーク画像の建築物が建造された時代の理論を用いることが， イメージを表すために有益な場合がある．例えば，我々の時代の教程としてはとても強い禁則である 2 声部の平行 5 度などが，古代ギリシャ音楽で用いられている [18]．
- 建築物の用途
  - 例えば，宗教に関わる教会や修道院などの建築物の場合，様式の違いが音楽にも関わってくる．例えば，正教会においては，楽器を使用しない文化がある [10,12]．この特徴は楽曲制作において強い制約になり，イメージに沿いつつも，確立された音楽を見出だすことができる．

また，楽式については，コーパスの使いやすさを考慮して 30 秒から 1 分という制約を設けた．そのため，その時代の楽式を踏襲しつつ，例えば西洋音楽などにおいては，曲の長さが短くても比較的作曲しやすい三部形式や変奏曲などが主である．しかし，将来的な展望として，例えば純邦楽における序破急など様々な国や文化の楽曲の形式を考慮する必要があると考えられる．本稿では，歴史が深く教程が確立されている西洋のランドマークを中心に選んだ．

本来は，ランドマーク画像からイメージされる楽曲を制作するには，その時代における音楽理論・規則はもちろん，背景知識の深い調査が必要になる．しかし，本コーパスの設計方針が，あくまでも作曲者がランドマーク画像から受けた印象による楽曲制作と，単一作曲者の楽曲制作による統一性の重要視である．そのため本稿では，あくまでも作曲者がその画像から受けた印象を楽曲にする際の補助としての調査 [10–17,19] に留めた．

3. コーパスデータ

3.1 ディレクトリ構造

本コーパスのディレクトリ構造を以下に示す．ディレクトリ及びファイル名は，Google のデータに付随するランドマーク ID を用いた．

```
│ 🗀 000001
│  └─🗀 000001
│      ├─🗎 000001.jpg
│      ├─🗎 000001.wav
│      ├─🗎 000001.mid (SMF Format:1)
│      └─🗎 000001.txt
├─🗀 000002
├─🗀 ...
└─🗀 000010
```

3.2 楽曲例

- エッフェル塔
  - ランドマーク番号：40333
  - 国名：フランス
  - 建設された時代：1889 年
  - 調：イ短調
  - 楽式：三部形式，ミュゼット
  - 使用楽器とプラグイン：
    - アコーディオン (Native Instruments KONTAKT 6 Factory library)
    - ギター (Prominy HummingBird)
    - ベース (Native Instruments KONTAKT 6 Factory library)
  - 参考楽曲及び作曲者・CD：
    - ”Paris Musette 2: Swing & manouche (French Accordion)/The Paris Musette” [20]
    - ”Paris Musette 3/Paris Musette” [21]
    - ”De Clichy A Broadway/Gus Viseur” [22]
    - ”Valse Musette/Valses De L’accordeon” [23]

ランドマーク画像であるエッフェル塔（Fig. 1）と，そこから作曲した楽曲（Fig. 2）を以下に示す．画像のイメージからフランスの地方の民族音楽であるミュゼットを連想した．画像のイメージに合う明るい曲調と，エッフェル塔の外見的特徴から，A では鉄骨での組みあがりを意識し，稲妻形のベースラインにし，中間部（B）でのアコーディオンは塔を成すシンメトリカルなメロディラインを考慮し楽曲を制作した．

また連想したミュゼットを楽曲分析して多く見られた特徴，文献 [11] による調査を統合して以下の特徴を反映した．

- 四分の三拍子
- 三部形式
- アコーディオン・ギター・ベースの三種類の楽器編成
  - より多くの楽器を用いた編成もあったが，建造物の規模を考慮し，上記の編成にした．
- 和声進行
- 半音音階を含む順次進行のメロディ
- 楽器の奏法とリズム，定型的フレーズ

3.3 課題

作曲のための手掛かりが少ないため，与えられた画像だけではなく，国名や宗教的特質，その地方で用いられる楽器などの情報に楽曲が強く影響を受けてしまう．また西洋に関しては，国に限らずキリスト教関連の建造物が多く，教会や修道院，墓地など，外見的がとても類似したものが多いため，楽曲制作にあたっての区別化が難しい．以上のように，画像と音楽をペアとしたコーパス構築に関しては多くの問題があり，加えてコーパス構築として本稿はまだ始まったばかりの試みである．前述の通り，非常に幅広い知見が必要であるため，一人の人間が全ての楽曲を担当していることから知識的にも不十分であることが大きな課題である．

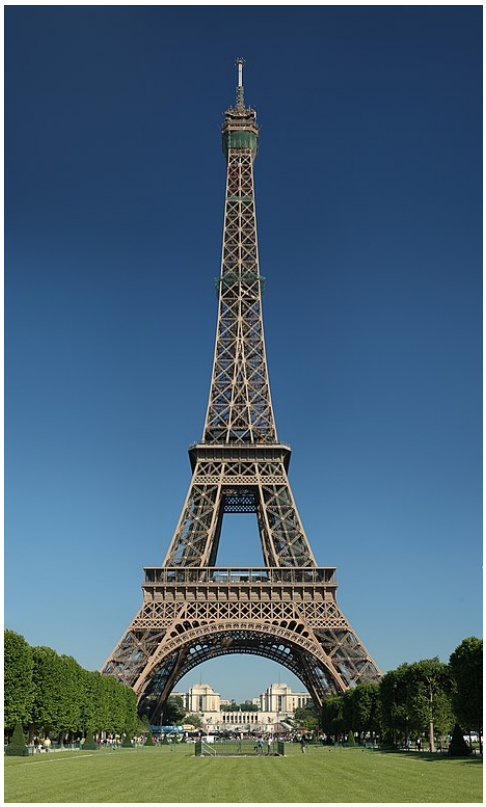

Figure 1 エッフェル塔 (Google landmark dataset No. 40333)

4. おわりに

ランドマーク画像及びそれを題材とした単一作曲者による楽曲の，画像と音楽のペアからなる作曲情報の付随したフリーのコーパスを構築し，本稿ではその構成とデザイン法を解説した．

本コーパスは，プロジェクトページ [6] にて公開しており，以下の場合に限り使用可能である．

- アカデミック機関での研究
- 非商用目的の研究（営利団体での研究も含む）
- 個人での利用（ブログなどを含む）

謝辞 本稿は，東京大学 GAP ファンドプロジェクト「音声合成技術の研究開発・商用利用を加速させる音声コーパスの設計・構築」の支援を受けて実施した．

参考文献

[1] 岩宮眞一郎：音楽と映像によるマルチモーダル・コミュニケーション，日本音響学会誌，Vol. 52, No. 1, pp. 40–45 (1995).
[2] Kajihara, Y., Dozono, S. and Tokui, N.: Imaginary Soundscape: Cross-Modal Approach to Generate Pseudo Sound Environments, *NIPS 2017 Workshop*, California, U.S.A. (2017).
[3] Zhou, Y., Wang, Z., Fang, C., Bui, T. and Berg, T. L.:

ⓒ 2020 Information Processing Society of Japan

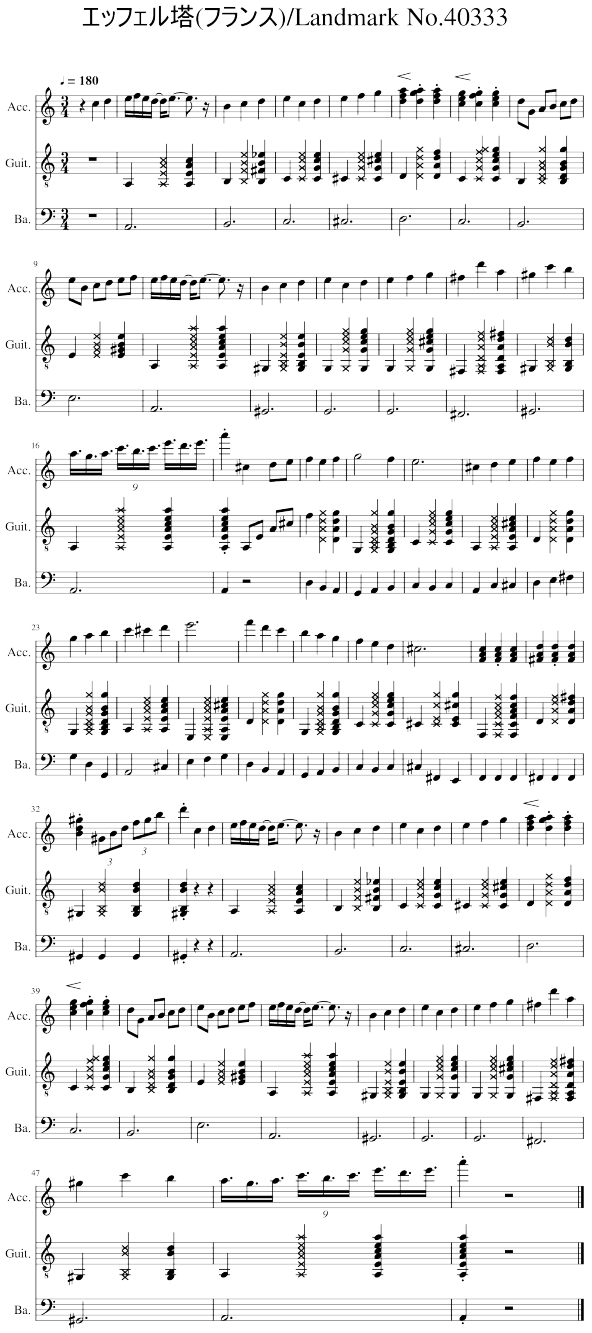

Figure 2 Fig. 1 を題目として制作した楽曲

Visual to Sound: Generating Natural Sound for Videos in the Wild, *CVPR*, Utah, U.S.A., pp. 3550–3558 (2018).

[4] Dong, H.-W., Hsiao, W.-Y., Yang, L.-C. and Yang, Y.-H.: MuseGAN: Multi-track Sequential Generative Adversarial Networks for Symbolic Music Generation and Accompaniment, *AAAI*, Louisiana, U.S.A. (2018).

[5] Weyand, T., Araujo, A., Cao, B. and Sim, J.: Google Landmarks Dataset v2 - A Large-Scale Benchmark for Instance-Level Recognition and Retrieval, *CVPR*, Utah, U.S.A., (online), available from ⟨https://arxiv.org/abs/2004.01804⟩ (2020).

[6] : Life-M: Landmark imaged-themed free music corpus, `https://sites.google.com/site/shinnosuketakamichi/research-topics/life-m_corpus`.

[7] 八代秀夫：イタリアに於ける初期バロック音楽，論集，Vol. 27, No. 2, pp. 43–54 (1980).

[8] 菅野裕子：建築と音楽におけるロマネスク期からゴシック期への変化の類似性について，日本建築学会計画系論文報告集，Vol. 446, pp. 155–163 (1993).

[9] 菅野裕子：ルネサンス期における建築と音楽の単位に関する考察：モドゥルスとタクトゥスの類似性について，日本建築学会計画系論文集， Vol. 70, No. 589, pp. 213–220 (2005).

[10] 木下康彦，木村靖二，吉田　寅：詳説世界史研究，山川出版社 (2008).

[11] 菊池有恒：楽典: 音楽家を志す人のための，音楽之友社 (1987).

[12] 高階秀爾：カラー版西洋美術史，美術出版社 (2002).

[13] 世界遺産アカデミー，世界遺産検定事務局：きほんを学ぶ世界遺産 100: 世界遺産検定 3 級公式テキスト，世界遺産アカデミー (2017).

[14] 世界遺産アカデミー，世界遺産検定事務局：くわしく学ぶ世界遺産 300: 世界遺産検定 2 級公式テキスト，世界遺産アカデミー (2017).

[15] Jeppesen, K.: *Kontrapunkt (vokalpolyfoni)*, W. Hansen (1930).

[16] Piston, W.: *Counterpoint*, W. W. Norton (1947).

[17] Piston, W. and DeVoto, M.: *Harmony*, W. W. Norton (1987).

[18] Ensemble, P. T.: Ancient Greek Musical Instruments/Music of Ancient Greece (2009).

[19] 長谷川良夫：対位法: 線的作曲技法及びカノン・フーガ，音楽之友社 (1955).

[20] Musette, T. P.: Paris Musette 2 : Swing & manouche (French Accordion) (2009).

[21] Musette, P.: Paris Musette 3 (1998).

[22] Viseur, G.: De Clichy A Broadway (2000).

[23] L'accordeon, V. D.: Valse Musette (2019).