

## 静止画特徴によるコード進行自動生成の検討

武田裕太<sup>†</sup> 相川清明<sup>†</sup>

風景写真に適合するコード進行を生成するシステムを提案する。風景写真とコード進行の関係性を調べるために、両者の特徴を8種の感情表現語から成るベクトルで表現し、1枚の風景写真に対してベクトルの類似度が高いコード進行を生成する。

以上の過程で作成した風景写真とコード進行の組合せに対する満足度を被験者30人に視聴させ、5段階で評価させた結果、70%が平均満足度3以上となった。

### On the automatic generation of chord progression depending on picture features

Yuta Takeda<sup>†</sup> Kiyooki Aikawa<sup>†</sup>

This report describes a method for generating chord progression depending on picture features.

To learn the relationship of pictures and chord progressions, picture features and chord progressions were represented by eight dimensional vectors. Experimental results showed the average MOS score of three

#### 1. はじめに

コンピュータの技術は日々発展し、音楽制作においてコンピュータを使う機会が多くなっている。メロディの打ち込み・楽器演奏の録音・ミキシング・マスタリングの一通りの工程をコンピュータ1台で済ませる事ができ、高音質な楽曲が制作できるようになった。さらにメロディやコード進行をコンピュータに作らせる自動作曲システムも開発された。たとえば、入力した歌詞の音韻からメロディを作る Orpheus[1]、既存のメロディに和声(コード)を追加する AMOR[2]などが挙げられる。

今回は自動作曲の応用として、ユーザーが選んだ画像に対する適切コード進行を生成するシステムを提案する。作曲する際に例えば「この海のような曲が作りたい」というイメージが出る。そのイメージに沿った曲を提供できるようなシステムを作れば、作曲初心者や、作曲を専門としない活動(例:映像関係)をする人たちにとってスムーズに作業が進行できると思ったからである。

#### 2. 風景写真とコード進行の特徴分析

今回扱う画像は風景写真である。海・山・雷などの風景を扱った写真は印象がつかみやすく、扱っている色彩もほぼ統一されているからである。たとえば海の写真を見ると穏やかな印象がつく事がある。その海の写真もほぼ青に統一されていれば、穏やかな印象が強くなる。このような理由から風景写真を扱う。

風景写真に適合するコード進行を提供するために、最初に風景写真とコード進行の特徴を調べることが必要である。まずは風景写真とコード進行の感情的特徴をアンケートによって調査し、アンケート集計結果からそれぞれの風景写真とコード進行の感情ベクトルを作成(ベクトル空間法3を参照)する。

##### 2.1 実験方法

18枚の風景写真と24種類のコード進行を視聴させる。8種類の感情表現語(2.2.3のヘヴナーの感情表現円環では8種扱われているため)を評価項目とし、それぞれ当てはまるかというアンケートを20代前後の男女30人に1(非常に良く当てはまらない)-5(非常に良く当てはまる)の5段階評価で答えさせた。

風景写真を18枚扱う理由に関しては、まず100枚の写真からクラスター分析を使って複数のクラスに分類するが、19クラス以上にすると、似たようなクラスができるので、今回は18クラスに分類し、各クラス代表1枚を選出した。コード進行に関しては25種類以上だと似たようなコード進行ができる(例:C→F→G→CとF→G→C→C)ので、24種類のコード進行を扱った。

<sup>†</sup> 東京工科大学  
Tokyo University of Technology

## 2.2 使用したサンプル

実験をはじめる前に、風景写真やコード進行などの実験素材をインターネットなどから集めて使用した。

### 2.2.1 風景写真

インターネット上に公開されているフリー素材である写真を100枚選択した。これらの選択基準は風景写真（山・海・空・夕焼けなど）であること、コンピュータによる加工（ペイントソフトを使用した絵画、写真の色の変換や被写体の変形など）がされていないこと（種類が多くなり分析の精度が落ちるため）、1枚の写真の色彩が分散されすぎていないことである。100枚の写真をクラスター分析し、以下の18枚の風景写真が選ばれた。



Figure 1 採用された風景写真（1列目左から砂漠，秋山，曇海，森林，盆地，野原，2列目同様青空，落葉，雷，夜景，海，紅葉，3列目同様夕焼，街，花畑，鉄道，洞窟，崖）

### 2.2.2 コード進行

コード進行に関しては、成瀬正樹氏による『コード進行スタイル・ブック』[4]に掲載されているコード進行を参考に、明るい進行，暗い進行，シンプルな進行，複雑な進行といったように幅広く採用した。キーはCまたはAmとし、主に使用頻度が高いもの、聞いた時の印象が高いものを中心に採用した。

採用したコード進行を ImageLine 社の FL STUDIO 9 で打ち込み、自動演奏したものを wav ファイルで保存した。

採用したコード進行を次の表 Table 1 に記載する。

Table 1 採用されたコード進行

音源	1st	2nd	3rd	4th
01.wav	C	Caug	C6	C7
02.wav	C	F	G	C7
03.wav	F	C	G	Am
04.wav	C	Am	C6	Am
05.wav	Dm7	Em7	Am7	Am7
06.wav	F	G	C	Am
07.wav	Am	F	G	C
08.wav	F	Fm	C	C
09.wav	Am	AmMaj7	Am7	Am6
10.wav	Am	G	F	Em
11.wav	Am	F	G	C
12.wav	Em	Am	F	G
13.wav	C	Am	F	G
14.wav	C	Cmaj7	C7	F
15.wav	F	G	C	E7
16.wav	C	Bm7	Em7	Am
17.wav	Am7	Dm7	G7	C
18.wav	C	G	Am	Em
19.wav	Am	Dm	Em	Am
20.wav	F	G	Am	Am
21.wav	Am7	Cm7	Am7	Cm7
22.wav	C	C	C	C
23.wav	A b	B b	C	C
24.wav	Am	Am	Am	Am

### 2.2.3 評価項目

ヘヴナーの感情表現円環[5]を参考にし、以下の感情表現語を採用した。

Table 2 今回使用する感情表現語

記号	感情表現語	その他の形容詞
A	力強い	たくましい, 勇ましい
B	明るい	愉快的, 喜びの
C	静かな	安らかな, 静かな
D	穏やかな	懂れの, やさしい
E	威厳ある	神聖な, 崇高な
F	情熱的な	興奮させる, 活力を与える
G	暗い	悲しい, 憂鬱な
H	優雅な	軽い, 活発な

## 2.3 実験結果

30人の被験者のアンケート回答データを集計し、各項目の評価を平均した。風景写真に関して、平均値が3.5以上のものが144サンプル中37サンプル、音源に関して、192サンプル中8サンプルである。結果の特徴として、明度の高い画像にはB,Cの評価が、低い画像にはA,E,Gの評価が多かった。コード進行に関しては、メジャーコードで終わるものにはBが、そうでない場合にはGが多く評価されている。

### 3. 風景写真からコード進行を生成する

第2章で行った実験を発展させる。前章のアンケート集計結果から風景写真とコード進行の感情ベクトルを生成し、風景写真とコード進行のベクトルの類似度が近い者同士を組み合わせ、108種の視聴覚刺激を作成した。

最終的には以下の図のとおりコード進行を生成する。

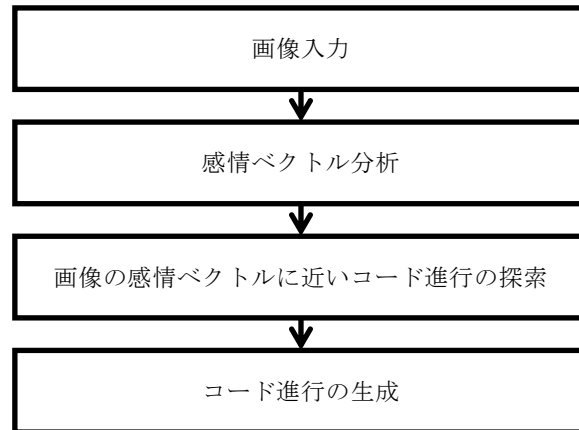


Figure 2 風景写真からコード進行生成の流れ

#### 3.1 画像とコード進行の類似度分析、視聴覚刺激の作成

第2章で行った特徴分析のデータから、特徴が近いもの同士を結合させる。今回はベクトル空間法を使って風景写真とコード進行の類似度を調べ、ひとつの風景写真につき適切な画像を模索した。風景写真 I とコード進行 S の各項目別評価を IA~IH, SA~SH とおく。x 番目の画像 I<sub>x</sub> の各評価 IA~IH で成り立つベクトル  $\vec{I}_x$  と y 番目のコード進行 S<sub>y</sub> の各評価 SA~SH で成り立つベクトル  $\vec{S}_y$  のなす角を  $\theta$  と置くと、 $\cos \theta$  は以下の式のように表される。

$$\cos \theta = \frac{\vec{I}_x \cdot \vec{S}_y}{|\vec{I}_x| |\vec{S}_y|} = \frac{I_{xA} \cdot S_{yA} + I_{xB} \cdot S_{yB} + \dots + I_{xH} \cdot S_{yH}}{\sqrt{(I_{xA}^2 + I_{xB}^2 + \dots + I_{xH}^2)(S_{yA}^2 + S_{yB}^2 + \dots + S_{yH}^2)}}$$

Figure 3 画像とコード進行の類似度を求める計算式

$\cos \theta$  の値が大きいほどふたつのベクトルのなす角度は 0 に近く、特徴が近いことを表す。18 枚の画像に対し、 $\cos \theta$  がもっと大きい上位 5 種類のコード進行と、小さいコード進行 1 種を採用し組み合わせた。 $\cos \theta$  が小さいコード進行に関しては、20~24 番目のコード進行から選んだ。最下位のコード進行だけにすると、コード進行が重複してしまうからである。

類似度の小さいコード進行を取り扱ったのは、次に説明する満足度が低くなる事を仮定したためである。

全体 108 刺激の平均類似度は 0.95 である。うち類似度の高い 90 刺激の平均, 最大, 最小類似度は 0.98, 0.99, 0.91 であり、類似度の低い 18 刺激の平均, 最大, 最小類似度は 0.88, 0.96, 0.77 である。

このように 108 の視聴覚刺激を作成し、次節の評価実験に利用する。

#### 3.2 実験方法

作成した視聴覚刺激を 30 人の被験者 (20 代の男女, 男 21 人, 音楽経験者 21 人) に視聴させる。「今映っている写真に対して、再生されたコード進行が選ばれたとしたら、写真とコード進行の組み合わせに満足できるか」を 1 (全然満足できない), 2 (あまり満足できない), 3 (かろうじて満足できる), 4 (大体満足できる), 5 (かなり満足できる) の 5 段階で評価させた。以降、各刺激の評価を被験者の数で平均したものを満足度とする。

#### 3.3 実験結果

108 刺激中 76 刺激が満足度 3 以上となり、全体的に約 70% を占めている。さらに、 $\cos \theta$  の値が大きいものだけで見ると (Figure 3~5 のグラフが青・緑の項目), 90 刺激中 72 刺激が満足度 3 以上となった。しかし満足度が 4.0 以上のものは 2 刺激しかなく、全体の平均満足度は 3.17 であり、かろうじて満足できる組み合わせが多かった。

満足度が最も大きいのは雷-3 (Am→Ammaj7→Am7→Am6) であり、4.07 となっている。

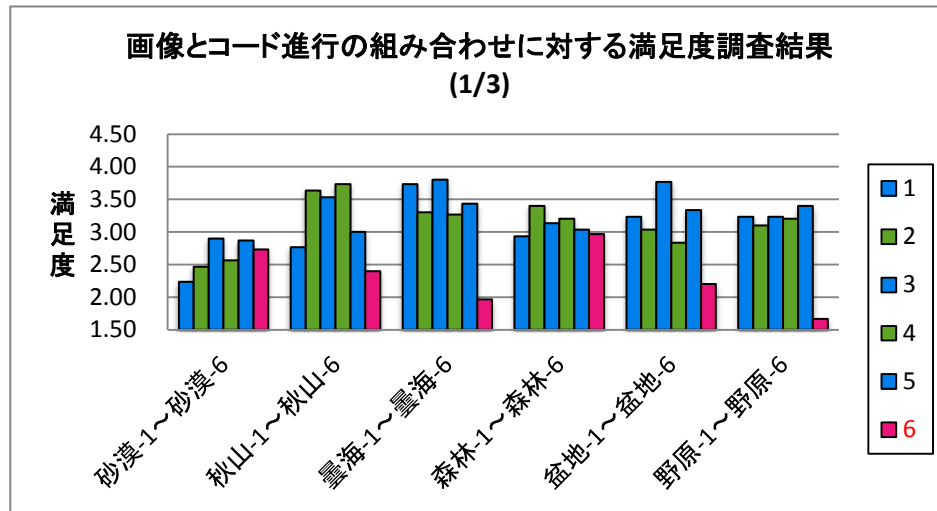


Figure 4 画像とコード進行の組み合わせに対する満足度調査結果(1/3)

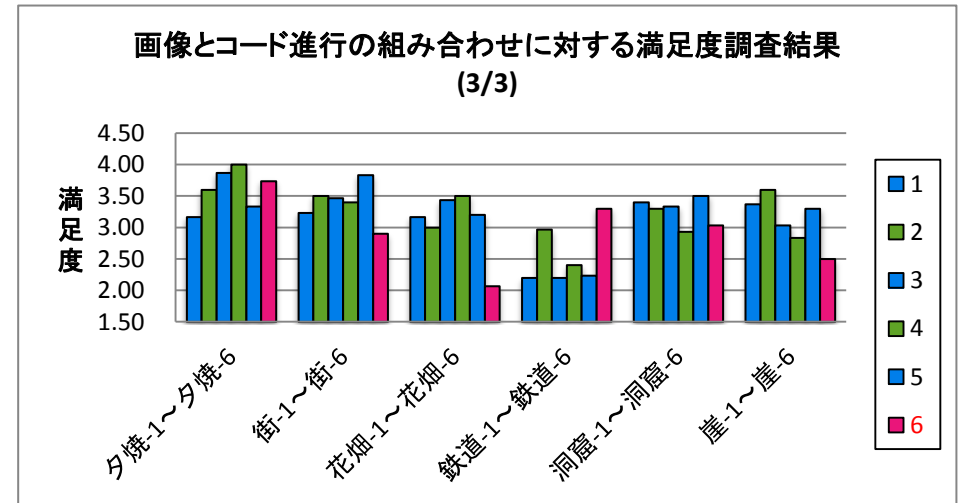


Figure 6 画像とコード進行の組み合わせに対する満足度調査結果(3/3)

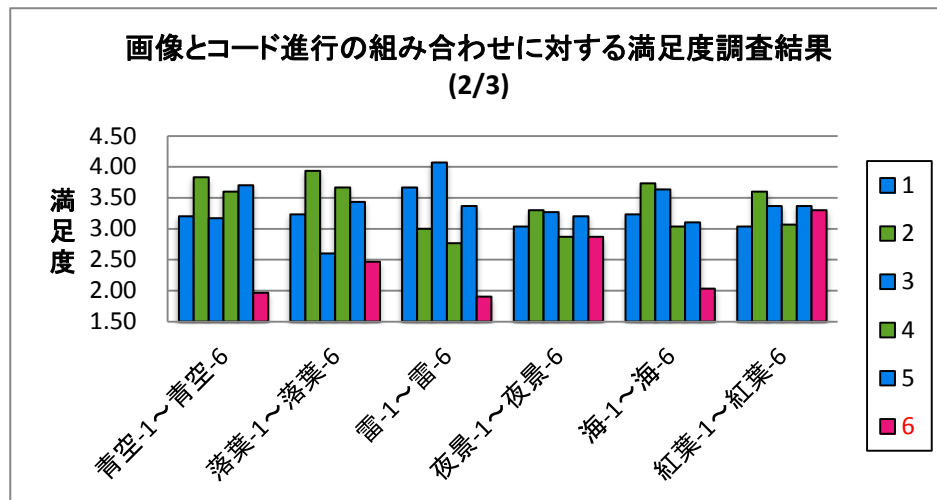


Figure 5 画像とコード進行の組み合わせに対する満足度調査結果(2/3)

### 3.4 考察

満足度が高かった3刺激と低かった4刺激の特徴を調べる。

#### 3.4.1 満足度が高かった3刺激の項目別評価

類似度が高く、満足度が高かった雷-3 (満足度 4.07, 類似度 0.97), 夕焼-4 (満足度 4, 類似度 0.98), 青空-2 (満足度 3.72, 類似度 0.97) の項目別評価を調べる。Table 3, 4, 6 の項目は Table 2 の感情表現語に該当する。太文字が写真, コード進行ともに高評価または低評価の項目である。この表から見ると, 両者とも高評価または低評価である項目が多くあることがわかる。たとえば, 雷-3 の A (力強い) が両者とも高評価であり, 青空-2 の G (暗い) が両者とも低評価である。これからわかる事は, 写真とコード進行の特徴がはっきりし, 両者ともに評価が高い (または低い) 特徴が存在しているということである。

Table 3 雷-3, 夕焼-4, 青空-2の項目別評価 (左から風景写真, コード進行, コード進行の内容は Table 1 を参照)

項目	雷-3.		夕焼-4		青空-2.	
	雷	09.wav	夕焼	06.wav	青空	13.wav
A (力強い)	<b>4.13</b>	<b>3.23</b>	2.03	2.57	2.50	3.07
B (明るい)	<b>1.77</b>	<b>1.53</b>	<b>2.7</b>	<b>2.33</b>	<b>4.63</b>	<b>3.53</b>
C (静かな)	<b>1.43</b>	<b>2.30</b>	4.17	2.8	3.90	2.83
D (穏やかな)	<b>1.17</b>	<b>2.13</b>	<b>4.1</b>	<b>3.1</b>	4.23	2.70
E (威厳ある)	<b>3.80</b>	<b>3.07</b>	2.57	2.7	2.37	2.10
F (情熱的な)	2.47	2.67	2.77	2.13	<b>2.07</b>	<b>2.33</b>
G (暗い)	<b>4.10</b>	<b>4.17</b>	2.8	3.13	<b>1.37</b>	<b>1.80</b>
H (優雅な)	<b>1.43</b>	<b>2.23</b>	2.87	3.17	2.83	3.13

### 3.4.2 満足度の低かった4刺激の項目別評価

類似度が、高く満足度が低かった砂漠-1 (満足度 2.23, 類似度 0.93) と鉄道-5 (満足度 2.23, 類似度 0.97) の項目別評価を調べる. 太文字で示したとおり, 写真・コード進行の一方が高評価なのにもう一方が低評価である項目が存在している. しかも評価が2~3あたりに多く分布されている. この事から評価が全体的にやや低めで特徴がはっきりとしていなく, 写真・コード進行ともに評価が高いまたは低い項目が少ないため, 満足度が低くなったと見る事ができる.

Table 4 砂漠-1, 鉄道-5の項目別評価 (Table 3 と同様)

項目	砂漠-1		鉄道-5	
	砂漠	22.wav	鉄道	16.wav
A	2.40	2.03	3.03	3.53
B	<b>3.23</b>	<b>2.77</b>	2.67	2.07
C	3.87	2.93	1.50	2.37
D	2.80	3.30	1.53	2.47
E	2.63	1.87	2.57	3.27
F	2.13	1.83	2.07	2.60
G	2.13	2.57	<b>2.50</b>	<b>3.63</b>
H	1.83	2.43	<b>1.60</b>	<b>3.03</b>

風景写真砂漠, 鉄道の項目 B (明るい), G の評価別回答数からみると, 評価3の回答数が多い. これは個人差が大きく, 人によっては明るく見えたり暗く見えたりする写真である事がわかる. この様な写真だと, 適切なコード進行を選ぶのが難しくなる.

Table 5 砂漠, 鉄道の項目 B, G の評価別回答数

画像	砂漠		鉄道	
	B	G	B	G
平均	3.233	2.133	2.667	2.500
評価が1の回答数	0	13	5	10
評価が2の回答数	7	6	7	4
<b>評価が3の回答数</b>	<b>11</b>	<b>6</b>	<b>13</b>	<b>8</b>
評価が4の回答数	10	4	3	7
評価が5の回答数	2	1	2	1

なお, 類似度が低く, 満足度も低い2刺激の項目別評価を調べる. ここで雷-6 (満足度 1.90, 類似度 0.83), 野原-6 (満足度 1.67, 類似度 0.77) を扱う. Table 4 と同様に, 写真・コード進行の一方が高評価なのにもう一方が低評価である項目を太文字で表した.

Table 4 よりも各項目の写真とコード進行の評価の差が大きく開いていることがわかる. たとえば Table 4 に記載されている鉄道-5の項目 G の評価の差が 1.37 に対し, Table 6 の野原-6の項目 G の評価の差は 2.87 である. 以下の2刺激の感情的特徴 (特に「暗い」, 「穏やかな」の項目) が写真とコード進行で多く離れていることが, 満足度が低くなる原因となることがわかる.

さらに2刺激の写真とコード進行をよく見てみると, 雰囲気一致していないことがわかる. たとえば野原-6の場合, 写真が全体的に明るめであることにに対し, コード進行が Am7→Cm7→Am7→Cm7 となり, マイナーコードを多用した暗めな雰囲気となっている. こういった雰囲気からわかるように, これらの組み合わせは合わないということがわかる.

Table 6 雷-6, 野原-6 の項目別評価 (Table 3 と同様)

項目	雷-6		野原-6	
	雷	08.wav	野原	21.wav
A	<b>4.13</b>	<b>2.33</b>	1.97	3.10
B	<b>1.77</b>	<b>3.37</b>	<b>4.23</b>	<b>1.57</b>
C	1.43	2.63	<b>4.07</b>	<b>1.90</b>
D	1.17	3.10	<b>4.57</b>	<b>1.87</b>
E	<b>3.80</b>	<b>2.10</b>	2.03	3.37
F	2.47	2.23	1.40	2.30
G	<b>4.10</b>	<b>2.23</b>	<b>1.33</b>	<b>4.20</b>
H	1.43	3.10	2.80	2.23

#### 4. おわりに

選択した風景写真に適切なコード進行を生成するために、風景写真とコード進行の感情的特徴を調べ、風景写真の特徴に最も近いコード進行を生成した。

風景写真と生成されたコード進行の組み合わせのうち、満足度3以上の組み合わせが約70%となった。これにより人間が選んだ写真に対しかろうじて満足できる組み合わせを提供することができる。

満足度が高い組み合わせでは、感情的特徴において写真・コード進行ともに高いまたは低い評価を出している項目が多く、逆に満足度が低い組み合わせだと、一方が高く、もう一方が低い項目が多いことがわかった。また個人差の大きい写真の場合、3（どちらとも言えない）という項目別評価が多く回答されやすく、特徴がつかみづらい。

精度を向上させるには、感情的特徴の評価の差が大きすぎる組み合わせの排除や、個人差が少なく、特徴がはっきりしている写真（多くの被験者が4, 5 または 1, 2 を答えている項目が多い写真）の厳選が必要である。

**謝辞** 被験者として協力して下さった東京工科大学の学生、助言をして下さった研究室の学生に感謝する。

#### 参考文献

1) 深山 覚, 西本 卓也, 嵯峨山 茂樹(東大院情報理工), “歌詞の韻律に基づく自動作曲のための旋律生成モデル”

2) 柳田益三ら6人, “ポップス系の旋律に対する和声システム: AMOR”, 2005/5, 情報処理学会論文誌 Vol.45 No.5 P1176-1187

3) 吉村 建慶(岐阜高等専門学校・電気情報工学科) “テキストマイニングによる文書の類似度計算に関する研究”より“ベクトル空間法”, <http://www.gifu-nct.ac.jp/elec/deguchi/sotsuron/yoshimura/node11.html>

4) 成瀬正樹, “コード進行スタイル・ブック”, 株式会社リットーミュージック発行, 2007/12/20 第9版発行, 第2章~第10章 P37~P180

5) P.N. ジェスリン&J.A.スロギダ, “音楽と感性の心理学”, 誠信書房, 2008/10/31, P137 第5章 音楽的構造の感情表現に及ぼす影響