

歌詞とモチーフを入力とした作曲モデルによる作曲支援システム*

荒木大誉 (法政大学情報科学部), 伊藤克亘 (法政大学情報科学部)

1 序論

作曲に関する知識がない人が作曲をするのは難しい。部分的に旋律を考えるが可能なのだが、その続きが思い浮かばないことができないことが多い。また歌曲を作る際に、歌詞を先に作る場合と後で付け加える場合がある。前者は詞先、後者は曲先と呼ばれている。詞先の曲である場合を想定して、作曲に関する知識があまりない人でも、部分的な旋律と好きな歌詞を用いて作曲が出来るようにする。

2 歌詞とモチーフを入力とした作曲モデル

通常、歌曲では似たような旋律を繰り返す構成が良く使われる。従来研究では繰り返し構成を意識したもののがあまりみられない。そこで本研究では、動的計画法にモチーフを用いて、繰り返し構成を意識した旋律を生成する手法を提案する。

本研究では、モチーフと呼ばれる旋律片と歌詞情報を基に、動的計画法を用いてリズム・メロディの半自動生成を行う。

まずはモチーフリズムの統合・分割を行って、歌詞へのアラインメントを行う。その後、求まったリズムや小節数に合わせてコード進行を入力してもらい、音高を求める。この時、それぞれのステップで動的計画法を用いる。

2.1 アラインメント

歌詞とフレーズをアラインメントするためには、曲全体の小節数・モチーフに当てはめる音素数を決めなければならない。曲全体の小節数が4の倍数になるように以下の式で計算する。

$$\frac{\text{歌詞の音素数}}{m} > \frac{\text{モチーフ音価数}}{\text{モチーフ小節数}}$$

を満たす最大の $m = \{4n | n = 1, 2, \dots\}$ を全体小節数とする。また、

$$N > \frac{\text{モチーフ音価数}}{\text{全体小節数/モチーフ小節数}} - \text{モチーフ小節数}$$

を満たす最小の N がモチーフに当てはめる最低音素数とする。ここで、アラインメントに入力される値に、語頭データがある。これは、歌詞の切れ目を表したものである。よって、語頭データと N から各小節に入れる音節数を決定する。アラインメントはモチーフ小節数ずつ行う。

アラインメントに使われるコストは、強弱裏拍・語頭と強弱裏拍の関係・音符の種類・前回分割箇所である。「強弱裏拍」は「強拍を優先分割」という仮説を考慮したものである。よって、強拍を分割するときはコ

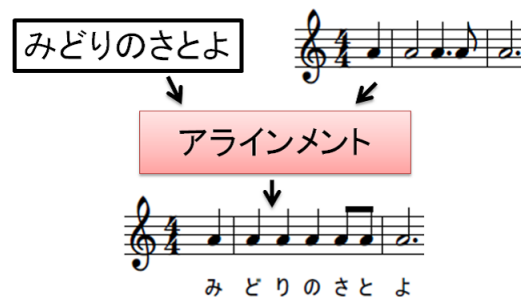


図 1. アラインメント

ストが無く、その他の箇所ではコストがかかる。ただし、弱起（弱拍から曲が始まる状態）では「弱拍を優先分割」させる。その場合はコストは逆に弱拍以外の箇所にかかる。

「語頭と強弱裏拍の関係」は前項のものとは違い、ある音価を分割したときに、語頭が強拍に来るかどうかを確認するものである。つまり、語頭が弱拍や裏拍に来る分割ではコストがかかる、ということである。ただし、弱起では語頭が弱拍に来ることが多いので、語頭が弱拍以外の箇所に来る場合にコストがかかる。これは、歌詞の内容理解へとつながる事柄である。

「音価の種類」は、より細かい音価はなるべく分割しないように考慮したものである。例えば、4分音符以外の音価が8分音符である時には、8分音符の方により高いコストをかけるということである。統合も、一部が逆になる同様のコストを用いる。

2.2 音高決定

本モデルで作る曲では、歌詞のアクセントに沿うように作ることを目的とする一方で、なるべく覚えやすい曲を作るため、モチーフの形をなるべく崩さないようにすることも考慮する。そこでこれらの事柄をコストとし、そのコストを最小化することで音高を決定する。音高決定では、歌詞のアクセント情報・調情報・音域・コード進行・アラインメント処理された音価データ・音高データ・アウトタクトの有無を入力すると、コストによって決定された音高データが出力される。

音高決定に用いるコストは、調の音階音・和声音と強弱裏拍の関係・歌詞アクセントと旋律音高の関係・モチーフの形状・モチーフ上下方向・音程・連音緩和・既存曲の音高曲線・曲の最初音と最終音である。調の音階音コストでは、決められた調で主に使われる音であればコストを低めに設定し、そうでない場合はコストを高めに設定する。

モチーフの形状コスト・モチーフ上下方向コストは、曲の覚えやすさに直結する、モチーフ繰り返しの関係するコストを設定する。モチーフの形状コストでは、アラインメントされたモチーフの該当部分と比べて、ある決められた値以上離れた場合に、離れた分コストが高くなるように設定する。モチーフ上下方向コストでは、モ

* Supporting System by Composition Model Regarded Lyrics and Motif as Input by Tomotaka Araki. (Faculty of Computer and Information Science, Hosei University) et. al.

チーフと比べて方向が違った場合にコストを設定する。

音程コストは歌曲の歌いやすさに関わるものである。ある特殊な音高の開き (例, 増4度・減5度・7度) があると, 歌が歌いづらくなる。その場合にコストを設定し, その曲中で特殊な音高の開きを少なくする。連音コストは, 後の音高曲線と併せて, 単調な旋律を生成しないためのものである。4回以上連続で同じ音高 (オクターブを考慮しない) が出現するとき, コストを設定する。

曲の開始音コストは, 曲の一番最初の音高をモチーフの最初の音と一致させるものである。よって, 最初音がモチーフの最初音と異なる場合にコストを設定する。また, 最終音コストは, 曲の最後の音に関するものである。どのような曲も基本的には主音で終わるものなので, 最終音が主音でない場合にコストを設定する。

一般的に4分の4拍子の曲において, 和声音でない音 (非和声音) は経過音や刺繍音として, 裏拍や2・4拍目に現れることが多い。該当部分が和声音か非和声音かどうか, コードから判断する。非和声音だった場合には, 禁則かどうかを判断する。ここでの禁則とは, 非和声音が該当コードの第3音と第5音と半音の関係にある状態をさす。次に強拍か弱拍・裏拍か判断する。強拍だった場合にはコストを高めに, 弱拍・裏拍だった場合にはコストを低めに設定する。ただし弱拍や裏拍だった場合に, その音が経過音・刺繍音ならば, 禁則であってもコストを低めに設定する。和声音だった場合にはコストはなしに設定する。この設定で, 非和声音が設定される時に表拍である状態を少なくでき, 刺繍音や経過音である非和声音を設定できる。

歌詞アクセントが旋律にどのような影響を与えているかを述べる。既存の詞先曲と詞先でないJ-POP曲の歌詞アクセントと旋律の関係性を調べた結果, 詞先曲では, 歌詞アクセントが変化しているとき, 旋律も同様の方向に変化あるいは変化なしである箇所が多くなっていた。逆に歌詞アクセントが変化しないときは, 旋律は特に目立って多い変化が見られなかった。以上のことから, 歌詞アクセントが変化するときは旋律音高もその方向へ変化させるか, あるいは旋律音高を変化させないべきであるといえる。歌詞アクセントが変化しないときは, どのような旋律音程の変化も認められる。

これらもコストとして取り入れる。歌詞アクセントが変化する時に, 旋律音高が逆向きに変化する場合, コストを高めに設定し, 音高が変化しない場合には, コストを低めに設定する。同方向に変化する場合には, コストは0とする。

モチーフと歌詞情報だけでは, 単調な旋律になってしまう。そこでメロディらしさを付加するために, 複数の既存曲の音高曲線を平均化したものをコストとして用いる。まず, 曲によって音域・音符数が違うので, 音高曲線を正規化する必要がある。音域は, 音高決定に用いる音域へ正規化する。音符数は, 歌詞の音素数へ正規化する。ただし, 曲線の形をなるべく変えないようにするため, 音符数の正規化にはスプライン補間を用いる。次に, 正規化された曲線の平均をとる。この平均が, メロディらしさを表す曲線となる。コストは, この曲線から音高がどれだけ離れるかで決めるこ

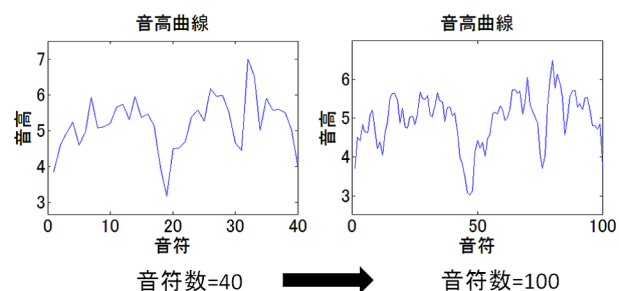


図2. 既存曲音高曲線のスプライン補間 (音符数40 → 100)

ととする。音高曲線の例を図2に示す。

3 結論

本研究では, 作曲を支援するために, 歌詞とモチーフからその先の旋律を作るモデルを作った。歌詞情報とモチーフから旋律のリズムを作り出し, それを歌詞に当てはめ, そこに和声進行を与えることで音高を決定し作曲を行った。

本モデルの問題点として, 拍子を4分の4に固定していることが挙げられる。そのため, 強拍と弱拍を固定することはできたが, 曲によっては変拍子など拍子が固定されていない事が多々ある。様々な種類の曲の生成を可能にするため, 拍子をコストとした変拍子モデルを構成する必要があると考えられる。

今後, 作曲モデルに対して評価実験を行う。原曲とモデルで作った曲をそれぞれ2回ずつ, 計6回聴いてもらい, 「覚えやすさ」, 「歌詞の内容理解」の2項目について評価する。聴く実験では, 歌詞内容が理解できているかどうか, 歌詞の穴埋め問題を解いてもらう客観評価と, 5件法尺度のアンケートによる主観評価を行う。

その後, それぞれの曲の楽譜を見せ, 実際に1曲ずつ歌ってもらい, 「覚えやすさ」・「歌いやすさ (音高・音域等)」の2項目について評価する。歌う実験では, 2曲の練習を各3回ずつ行い, その後楽譜を見ないで歌ってもらう。歌詞や旋律についてある程度正しく歌えるか, 客観評価を行う。また聴く実験と同様, 歌い終わった後に5段階評価のアンケートによる主観評価を行う。

参考文献

- [1] 深山 寛, 中妻 啓, 米林 裕一郎, 酒向 慎司, 西本 卓也, 小野 順貴, 嵯峨山 茂樹, "Orpheus 歌詞の韻律に基づいた自動作曲システム", 情報処理学会研究報告, 2008-MUS-76, pp.179-184, 2008]
- [2] 早川 和宏, 稲垣 博人, 田中 一男, "歌詞からラララ-言葉から歌への自動変換-", 人工知能学会ことば工学研究会第3回資料, 1999
- [3] 松尾 聡子, 片寄 晴弘, 井口 征士, "旋律予測のコンピュータシミュレーションに関する一検討", 情報処理学会論文誌, Vol.41, No.2, pp.498-508, 2000
- [4] M. Tomari, M. Sato and Y. Osana, "Automatic Composition System based on Genetic Algorithm and N-gram Model," Proceedings of IEEE International Conference on System, Man and Cybernetics(2008)
- [5] 住田 浩之, 林 朗, "マルコフ確率場を用いた自動作曲", 情報処理学会研究報告, Vol.2008, No.12, pp.151-156, 2008