

機能音声ラベルデータの作成と統計解析*

川上大輔¹, 金子仁美², 嵯峨山茂樹¹ (¹東大・情報理工, ²桐朋学園大学音楽学部)

1 はじめに

本稿では、多数の曲に新しい観点から和声ラベル付けを行い、その統計解析からの知見を述べ、perplexityの観点から和声言語モデルにおける意義を示す。

和声認識と音声認識は同型の問題と見なせ、音声認識で言語モデルが性能に大きな役割を果たしている[2]のと同様に、和声認識に関しても前後の文脈を表現する和声言語モデルが和声の認識に大きく寄与することが期待できる。さらに、和声言語モデルには、旋律への自動和声づけ、自動編曲、音楽情報検索、ムード分析、ジャンル分析等に幅広い応用が考えられる。

そこで本研究では和声言語モデル設計に関する知見を得るために、クラシック音楽60曲の詳細な情報を含む和声のラベルデータを作成し、和声の種類ごとの出現頻度、遷移確率等、基本的な統計情報を抽出し、様々な条件におけるN-gram言語モデルの性能をパープレキシティ(PP)により評価した。この評価の観点から、確率値の平滑化法、N-gramモデルにおける最適なNの値の検討、調を考慮しない和音言語モデルに対する、調を考慮した機能音声言語モデルの優越性について検討した。

2 和声ラベルデータの作成

2.1 従来の和声表記に基づいたラベルデータの問題

統計学習によって良い和声言語モデルを得るには、和声データが必要である。しかし、その目的には、従来の和声の表記法にはいくつかの問題がある。

日本で多く使われている島岡らによる和声学(いわゆる芸大和声)テキストにおける和声表記は、西欧の伝統的和声表記に対して教育上の多くの利点はあるが、例えばFig. 1(a)に示すように記法が複雑であり、コンピュータで扱うためのテキスト形式としては作成にも入力にも不向きであり、またFig. 1(b)に示すように長三和音と短三和音などと和音種別を陽に区別しないなど、工学的目的には不満が多い。

また、ポピュラー和声は、F, Am, Gdim7, 等のように和声の構成音に関する情報しか含まれず、転回形や借用和音、調構造などの表記には不便である。

2.2 KSN 和声ラベルデータ

上記のような背景から、以下のような観点で、コンピュータで扱うための新しい和声表記法を策定し、KS表記法(KSN: KS notation)と呼ぶことにした。

- 従来の和声学教程を学習した作業員にも、和声解析結果を記述しやすい
- ASCIIテキスト形式として扱え、特殊な入力環境を必要としない
- 記述文法に曖昧さがなく、機械で扱いやすい

Fig. 2にKSNを例示する。KSNは機能和声を基礎としているため、ポピュラー音楽の和音表記と異なり、調における位置、和声の解釈がより明確で、ポピュラー和声におけるm(minor), M(major), aug, sus, dim, 7, M7, mM7, dim7, 9等全て表記可能であり、転回形、各構成音の変位や省略の情報も扱える。

(a) 2nd inversion of ii. (b) German aug. 6th.

Fig. 1 Examples of traditional, Shimaoka's and KS chord notations.

```
I | ii7'''' | V7' | I | vi' | V: V7'''' | I'
| IV7'''' | ii7 | V7 | V | ii: !V9-'''' |
i' | !V9-'''' | I' | IV7'''' | ii7 | V7 | I
| IV: V7 | I7+ | V: !V9-' | !V9-'''' | V7
| I'' | V73!&[I] | V7 | V: !V9-''''/V | I''
| V73!&[I] | V7 | IV: V7 | I'' IV&6'' |
V7/I | I ||
```

Fig. 2 An example of KS notation for J.S.Bach's "Wohltemperierte Klavier, Teil I, Preludium 1."

RWC音楽データベースのクラシック音楽編に含まれる60曲に関して、KSNデータを作成した。作曲年代は17世紀から20世紀前半で、作曲者は24人でやや偏りがあり、バッハ、モーツァルト等が多い。

3 和声ラベルデータの統計解析

3.1 ラベルデータの統計解析

作成したKSNデータに対して統計解析を行った。和声系列をなす和声ラベルの合計個数は17764であった。和声ラベルの種類数は1026であり、種類の出現頻度から計算したエントロピーは7.08bitであった。

3.2 和声の種類に関する出現頻度と遷移確率等に関する統計解析

主音に対する和声の根音の音程のみに着目し、和声の出現頻度とその間の関係を遷移確率として計算した。ドミナントであるVが最も多く、次にトニックであるI、更に次にサブドミナントであるIV、そしてIIの順に出現頻度が高い。ほとんどの7割近くの和声がIとVのみによって占められていた。更に、和声学で示される基本的な進行であるドミナント進行やサブドミナント進行が、主要な和声II, V, I, IVの間で5割以上の高い確率で生起していた。V-Iとなる確率に関しては8割近くであった。

音楽音響信号から和声を推定する際、和声の種類相対度数や遷移確率は、明らかに調に依存して偏っているため、調が推定出来れば和声の推定精度も向上すると考えられる。

そのほかの統計として、転回形は基本形が最も多く第三転回形が少ないこと、転調の傾向に関して、属調や下屬調へ転調する確率が20%弱であり、他の調に比べ頻度が高いこと等がわかった。以上の結果から、実曲であるクラシック音楽60曲に関して、和声学が

* "Developping a functional harmony labeled database and its statistical analysis," by Daisuke Kawakami¹, Hitomi Kaneko² and Shigeki Sagayama¹ (¹University of Tokyo, ²Toho Gakuen School of Music).

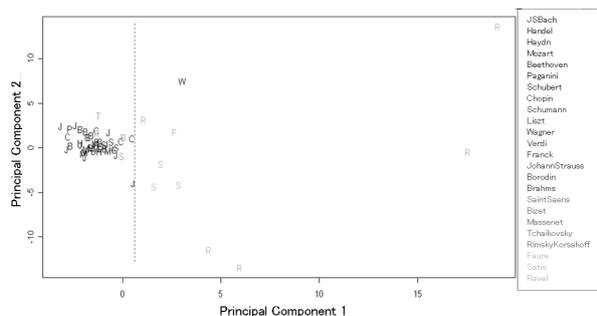


Fig. 3 Result of PCA where Faure, Satie, Ravel and Wagner scatter in the right side of the dashed line while others form a cluster in the left.

ら大きく逸脱しているということはないという知見が得られた。

3.3 主成分分析を用いた和声と年代の關係の分析

各曲に関して、和声の種類の出現率の分布を考え、各和声の出現割合を要素軸とする多次元空間上の点に対して主成分分析を行った。結果を Fig. 3 に示す。

この結果から、第一主成分軸上で、フォーレ、サティ、ラヴェル、ワーグナーの曲が第一主成分のある閾値によって弁別されていた。対して、それ以前の作曲家による曲は、一つのクラスを形作っていることを見出した。この結果はある年代までの曲は、共通の和声言語モデルで網羅できる可能性を示している。

4 和声言語モデル

4.1 N -gram 言語モデル

和声認識問題は入力としての音響信号だけでなく、推定対象である和声系列の確率言語モデルを利用することにより推定精度を向上させることができる [3, 4]。

和声認識問題は音声認識問題と類似しているため、音声認識分野の技術を認識率の向上のために流用できる可能性がある。音声認識では言語モデルとして N -gram 言語モデルが広く使われている [2]。そこで、和声系列を N -gram モデルとしてモデル化した。

N -gram モデルによる和声系列 $X = x_1 \dots x_T$ の出力確率値は以下のように計算される。

$$P\{x_1 \dots x_T\} = \prod_{t=1}^T P\{x_t | x_{t-1}, x_{t-2}, \dots, x_{t-n+1}\} \quad (1)$$

言語モデルの性能は以下の perplexity (PP) によって評価される [2]。テストデータ $X = x_1 \dots x_T$ に対して PP は小さいほど言語モデルの性能は良い。

$$PP = \sqrt[T]{P\{X\}} \quad (2)$$

式 1 の右辺の状態付き確率は学習データから最尤推定して求めるが、学習データに出現しない系列に対しては確率が 0 になってしまい、言語モデルとして使えなくなってしまう。そこでグッド・チューリング法 (GT 法)、ウィッテンベル法 (WB 法)、絶対法等の確率のスムージング [2] を行った。

4.2 N -gram モデルにおける N 値の評価

WB 法、絶対法いずれにおいても、 N の値が 3 の時に、最もテストデータにおける PP が低くなった。和声の言語モデルとして 3-gram の有効性が示された。

4.3 調を考慮した機能と声の有意性の検証

節 3.2 で、調を考慮した機能と声の方が、構成音のみに着目する和音よりも、言語モデルの語の単位として有用であると考察をした。言語モデルの PP を比

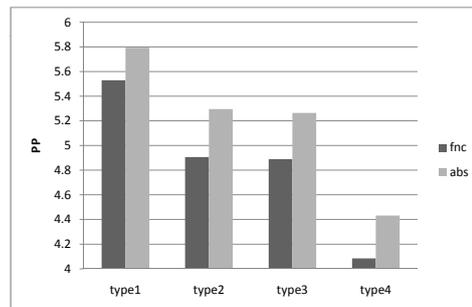


Fig. 4 Perplexities of functional harmony and (Popular) harmony models calculated for 3-gram models.

較することにより、より直接的に機能と声モデルの有用性を示した。

Fig. 4 に、機能と声の言語モデルと和音の言語モデルの PP を比較した結果を示す。和声ラベルのもつ情報の詳細さを段階的に type1 から type4 まで変えた。type1 から type4 の何れの場合についても、機能と声言語モデルの PP が和音言語モデルよりも PP が小さく、言語モデルの性能が良いことが示されている。これは節 3.2 で述べたように主要和音、特に I, V の出現確率、遷移確率に偏りがあるためであると考察できる。

5 おわりに

和声の言語モデルである和声モデルを検討するために得るために詳細な和声ラベルデータの仕様を決定し、クラシック 60 曲に関するラベルデータを作成した。ラベルデータの統計的特徴について解析することで、実曲を構成する和声の出現頻度、遷移確率、転回形の相対度数、転調の傾向が確かめられた。

更に和声ラベルデータから N -gram モデルを作成し様々な条件における PP を計算することで、和声言語モデルの 3-gram モデルの有効性を示した。また、確率値のスムージングとして検証した方法の中ではウィッテン・ベル法が有効であることが示された。機能と声言語モデルが調を考慮しない和音言語モデルよりも有効であることが確かめられた。

N -gram 言語モデルには様々な類似モデルがあり [1, 5]、今後は、より和声系列のモデル化に適した言語モデルについて検討していく方針である。

謝辞 本研究の一部は、文部科学省科学研究費補助金 基盤研究 (A) (課題番号 00303321) と科学技術振興機構 CrestMuse プロジェクトの支援を受けて行われた。

参考文献

- [1] F. Pereira, Y. Singer, N. Tishby: "Beyond Word N -grams," *Proc. 3rd Workshop on Very Large Corpora*, MIT, pp. 95-106, Mar. 1995.
- [2] 鹿野, 伊藤, 河原, 武田, 山本: 音声認識システム, 情報処理学会, 2001.
- [3] 諸岡孟: "確率的音楽生成モデルに基づく自動和声解析の研究," 東京大学情報理工学系研究科修士論文, Mar. 2008.
- [4] 上田他: "調波音/打楽器音分離手法とチューニング補正手法を用いた音楽音響信号からの自動和音認識," 情報処理研究会報告, 2009-MUS-81, pp. 1-6, 2009.
- [5] 森: "クラスに基づく言語モデルのための単語クラスタリング," 情報処理論文誌, Vol. 38, No. 11, pp. 2200-2208, 1997.