

和声ラベルデータの作成と 和声進行の統計解析

川上大輔^{†1} 金子仁美^{†2} 嵯峨山茂樹^{†1}

和声は西洋音楽の重要な要素であり、特に音楽音響信号からの和声推定や自動採譜などにおいては、精密な和声進行の統計的モデルが必要である。筆者らは、和声に関する研究推進のため、人手による和声ラベル作業の容易さと、コンピュータ可読性の両立を主眼にして、和声記述仕様を策定し、それに基づいて和声の18世紀から20世紀初頭までのクラシック音楽作品60曲に機能と和声ラベルデータを付与した。その和声系列を統計解析し、音楽的な知見から説明を試みる。また、統計的和声モデルとして N -gram モデルに関して、 N の値、スムージング法等を検討する。調や機能と和声などを反映した詳細な和声進行のモデルは、従来の和声モデルより perplexity を低くできることを示す。

Development of functional harmony labels and statistical analysis of chord progressions

DAISUKE KAWAKAMI,^{†1} HITOMI KANEKO^{†2}
and SHIGEKI SAGAYAMA^{†1}

Harmony is an important element of Western music, and a statistical model of precise harmony progression is especially necessary in harmony estimation and an automatic record in a musical note, etc. from the music acoustic signal. Authors settled on the harmony description specification with a main objective of the easiness of the harmony label work by people and coexisting of the computer readability for research promotion concerning harmony, and gave the function harmony label data to 60 classical music works from the 18th century of harmony to the 20th beginning of the century based on it. Harmony sequences is taking statistics analyzed, and the explanation is tried from musical knowledge. Moreover, the value N and the smoothing method of N -gram model, as a statistical harmony model, are examined. The model of detailed harmony progression that prepares and reflects key and function harmony etc. shows that perplexity can be lowered more than a past harmony model.

1. はじめに

wav,mp3等の音楽音響データから和声を自動で認識する技術、自動和声認識は、自動和声づけ、自動編曲のみに限らず、認識された和声ラベルを利用した音楽情報検索、ムード分析、ジャンル分析等、幅広い応用が考えられ、この問題について取り組むことは工学的に十分な意義がある。

和声認識問題と音声認識問題は数学的定式化の上で同一の問題であり、音声認識問題は推定したい語に関する言語モデルを利用し推定精度を高めている。²⁾ 言語モデルを利用することで、例えばある時刻に関する音響データが明瞭に得られない場合でも言語に関する前後の文脈から和声や語の系列を推定することが出来る。推定精度には言語モデルの性能が大きく関わっており、よりよい和声の言語モデルを設計することは和声認識問題において重要である。

そこで本研究は和声の言語モデル設計に対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、基本的な統計データを算出し、更に、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)を計算することで評価した。この結果を用いて、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、調を考慮した機能と和声言語モデルの有用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

和声言語モデルを設計するためには、モデル化の対象となる和声系列の、統計的特徴を把握する必要がある。しかし、研究目的の観点から、主要な従来の和声の表記法にはいくつかの問題がある。

Paul Hindemith、Arnold Schoenbergによる機能と和声の表記には和音種別の区別がなく、和音変位の詳細も区別されない。

島岡和声による表記(あるいは芸大和声)は、日本を中心に使われている表記法であり、長三和音と短三和音等の、和音種別の区別がない。記法が複雑であり、テキストデータ入力を中心としたラベルデータ作成、及び計算機による読み込みには難があると考えられる。

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演奏者のための和声表記といえる。調が不明確であり、転調の位置も不明確である。

^{†1} 東京大学大学院情報理工学系研究科

Graduate School of Information Science and Technology, The University of Tokyo

^{†2} 桐朋学園大学音楽学部

Toho Gakuen School of Music

■ 伝統的記法

```
C: | ii2 | V56 | | vi6 | G: V2 | I6 | IV2 | ii7 | V7 | C: V  
| d: vii7o | I6 | C: vii7o | I6 | IV2 | ii7 | V7 | | F: V7 |  
I7+ | C: V: vii7o | V346 | V7 | I46 | V37+G | V7 | II9 | I46  
| V37 | V7 | F: V7 | I46 | C: V7 | | ]
```

■ KS notation:

```
C: | | ii7''' | V7' | | vi' | { V: V7''' | I' | IV7''' | ii7 |  
V7 } | V | { ii: !V9-' | I' } | !V9''' | I' | IV7''' | ii7 | V7 |  
| | { IV: V7 | I7+ } | V: !V9' | V9''' | V7 | I' | V73!&[I]  
| V7 | [V!&!!I9 | I' | V73!&[I] | V7 | { IV: V7 | I' } |  
[I]&V7 | | ]
```

図 1 J.S. バッハ: 平均律第一巻第一番前奏曲ハ長調
従来の伝統的記法と、KSN 和声ラベルの比較。

2.2 和声ラベルデータ

従来の和声表記に準じたラベルデータの問題点を解消するため、和声の専門家と計算機処理の専門家による観点から次の点に留意して和声ラベルデータの表記法 (KS notation : KSN) を定めた。

- 和声学の教程に合致し、かつ世界的に受け入れられやすい
- 曖昧さがなくコンピュータプログラムが作成しやすい
- 表示や入力に特殊な環境を必要とせず、ASCII テキスト形式として扱える
- 人手で容易に記述でき、読みやすい

これらの特徴を持つ和声ラベルには、音大の学生によるラベルデータの作成や修正が効率的になる、計算機による読み込みが容易であることによって統計解析や言語モデルの作成の効率が向上する等の長所があると考えられる。

図 1 に一般的に用いられる和声表記と、KSN ラベルデータとの対応を例示する。

KSN 記法が持つ情報について説明する。まず、機能和声を用いているため、ポピュラー音楽の和声表記と異なり、調における位置、和声の役割が明確である。各構成音の変位を扱え、ポピュラー和声における Minar, Major, arg, dim, 7, M7, mM7, dim7, 9 等全て表記可能であり、更にどの音が省略されているかという情報も扱える。転回形の情報が含まれ、どの構成音が和声の最低音になっているのかわかる。直接的に最低音を記述することができ分数コード、保続音などが表現出来る。同時に鳴らされる音を直接表記でき、テンションも扱うことが出来る。

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の意義

この節で、実曲の和声の統計的特徴について検証する。和声の出現確率や遷移確率等の統計的特性を把握することは、より性能のよい言語モデルを設計するために重要である。多くの統計的特徴は和声学の知識からいくらか類推できるものであるが、広く知られているように実曲の和声は和声学の規則に全て厳密に従っているわけではなく多少の差異があり、厳密な数値の形で、その程度を把握することには十分な意義があると考えられる。

3.2 ラベルデータの概要

ラベルデータの概要について述べる。曲の年代は 17 世紀から 20 世紀前半である。曲データの作曲家は 24 人で、やや偏りがあり、バッハ、モーツァルト等が多い。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。2^{7.08} ≒ 135 であるため、これは 135 種類の異なる事象が当確率に生起する事象のあいまいさに等しい。和声の種類の数と比較して 1026 > 135 であることから、出現する和声の頻度に偏りがあることが分かる。

3.3 和声の種類相対度数分布

主音に対する和声の根音の音程のみに着目し、和声の出現頻度を計算したのが図 2 である。細かい情報を無視して、例えば、I や I7 等を同一視した。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ることが分かる。

ここから調を推定する必要性が裏付けされる。調が分かれば、I と V の構成音が分かり、I と V の和声が推定出来れば少なくとも 7 割の正答率が達成できる。

更に、図 2 の相対度数分布における I, II, IV, V の各々が、どのような和声によって占められているか算出したところ、I, II, IV の和音は安定な響きを持つ基本的な三和音が最も多いのに対し、V の和音は 6 割が属七の和音であった。これは V の和音のドミナントとしての役割を与えるため不安定な響きを持たせる傾向が強いことを示している。このドミナントの響きは減五度音程を含むものであり、他の三和音と構成音が異なり特徴的であるため、調推定に有益な情報である可能性があると考えられる。

3.4 和声の遷移相対度数分布

和声の根音度数に関する遷移確率を図 3 に示した。ここから和声学で示される基本的な進行、トニックやドミナントを中心とするドミナント進行やサブドミナント進行が、どの程度起こりえるのかが確率値としてわかる。この結果から、実曲であるクラシック音楽 60 曲に関して、和声学か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。

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。従って調を推定

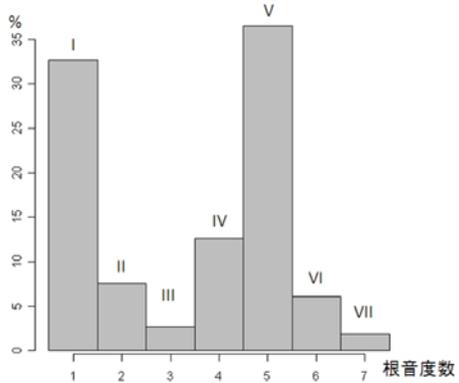


図2 根音の音程ごとの和声の相対度数分布。出現する6割以上の和声はIとVで占められる。

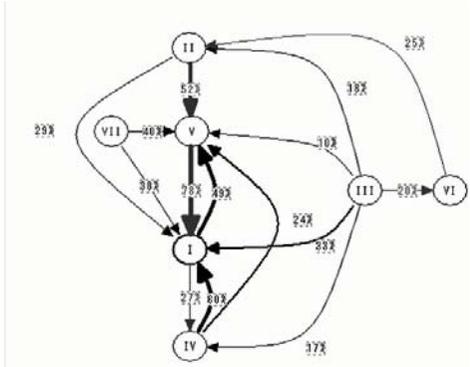


図3 和声の間の遷移確率。II, V, I の間でドミナント進行が、IV, I, V の間でサブドミナント進行が高い頻度で生じている。

する必要性の裏付けが示される。調が推定できていれば、どの和音がIやVの和声と対応しているのかが分かり、IやVの相対度数、遷移確率の偏りから、IやVの予測が容易になりえる、

3.5 転回形の統計解析

根音の度数ごとの転回形の出現頻度を図4に示す。どの根音度数に対しても基本形が一番多く、この傾向は特にIとVで顕著であった。基本形が一番響きが安定しているといわれ、対して第二転回は4度音程を含むため不安定である。基本形が一番多い理由として、基本形の響きの上での安定性が関係していると考えられる。

転回形は、和声を構成する一番最低部の音で決定される。この情報は、オクターブの違いを無視し音響情報を12次元のベクトルとして表すクロマ情報を用いて和声認識を行う場合⁴⁾、無視されてしまう。しかし、図4の結果を見る限り、和声の転回形は基本形が多い傾向にあるので、最低音の情報を和声の尤度計算に用いることにより、和声認識の推定精度を向上させる可能性がある。

3.6 転調の統計解析

転調の際、下属調、属調に転調する確率がそれぞれ、20%弱であり、他の調に比べ頻度が高いことを確認した。更に、転調先と転調元が長調か短調かに応じて、例えば平行調の関係から、転調先の分布が異なる事を確認した。

3.7 主成分分析を用いた和声と年代の関係の分析

各曲の作曲年代と和声系列の関係について主成分分析を行いその関係性を評価した。各曲に関して、どの種類の和声がどの割合だけ出現するかを示す相対度数分布を求めるこ

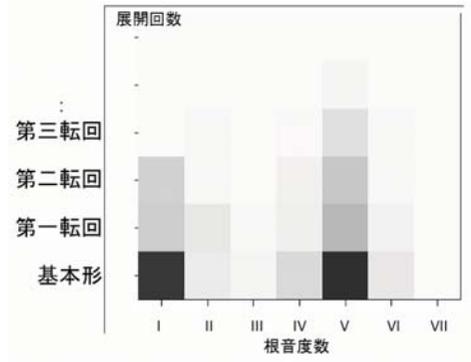


図4 各根音度数に対する転回回数の分布。濃い色ほど頻度が高い。全ての根音度数に関して基本形が最も多い。ドミナントであるVは、V7等の四和音であることが多いため、他と比べて第三転回形も多く見られる。

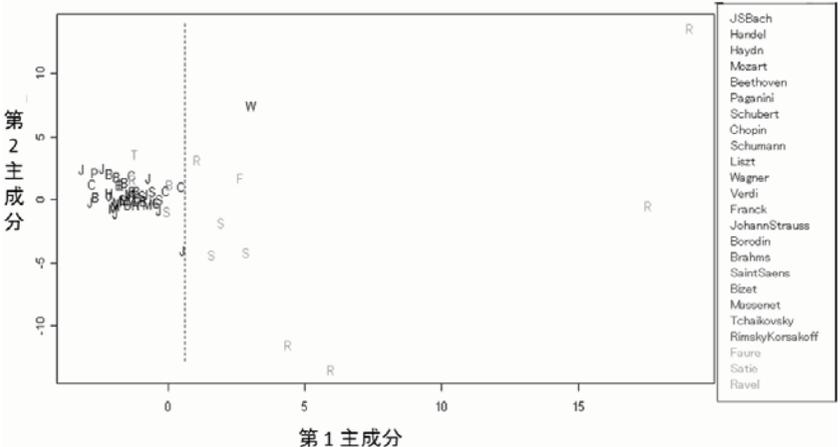


図5 各曲ごとの主成分軸上での分布。近代の作曲家の曲が破線を境に分離されている。

とができる。各曲はこの相対度数分布を特徴量とすることにより、各和声の出現割合を要素軸とする多次元空間上の点とみなせる。この点の集合に対して主成分分析を行い、第一主成分と第二主成分軸上にプロットしたのが図5である。

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る。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。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が示唆される。

4. 和声言語モデル

4.1 和声認識問題の定式化と言語モデル

この節では、様々な条件における言語モデルの性能をパープレキシティ(PP)を用いて評価、考察した。まず和声認識問題における言語モデルの位置づけについて述べる。

和声認識問題は、音響信号系列 Y が与えられたとき、和声時系列 X を条件付き確率 $P\{X|Y\}$ を最大化するように求める問題として定式化される。ベイズの定理から以下が成り立つ。

$$\operatorname{argmax}_X P\{X|Y\} = \operatorname{argmax}_X P\{Y|X\} P\{X\} \quad (1)$$

$P\{Y|X\}$ は音響モデル、 $P\{X\}$ は言語モデルと呼ばれ、和声認識システムの設計者が設計する。 $P\{X\}$ は和声系列 X が実際に起こり得るものに対しては大きく、稀なものに対しては小さくなる。

言語モデルは和声の推定精度を向上させるのに必要不可欠である。例えば、ある時刻の音響信号にノイズが加わり、その時刻の音響信号からでは和声の推定が困難な場合でも、言語モデルから導かれる和声系列の前後の文脈から、他の時刻で推定された和声を参考にすることで推定することができる。

4.2 N-gram 言語モデル

和声系列 $X = x_1 \dots x_T$ に対して、その尤もらしさ (実際にその和声系列が起こりやすいか否かの度合い) を確率値として評価するのが言語モデルである。

和声系列 $X = x_1 \dots x_T$ に対して、条件付き確率の性質から以下の式が成り立つ。

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

ここで、各 x_t が、高々 $n-1$ ステップ前の時刻の和声にしか依存しないという近似を施したモデルが N -gram モデルである。 N -gram モデルにおいて、確率値は以下のように計算される。

$$P\{x_1 \dots x_T\} = \prod_{t=1}^T P\{x_t|x_{t-1}, x_{t-2}, \dots, x_{t-n+1}\} \quad (3)$$

条件付き確率値 $P\{x_t|x_{t-1}, x_{t-2}, \dots, x_{t-n+1}\}$ の値は、学習データ系列中に $x_t, x_{t-1}, \dots, x_{t-n+1}$ と $x_{t-1}, x_{t-1}, \dots, x_{t-n+1}$ が現れた回数 N の比 f から最尤推定される。

$$P\{x_t|x_{t-1}, \dots, x_{t-n+1}\} = \frac{N(x_t, x_{t-1}, \dots, x_{t-n+1})}{N(x_{t-1}, \dots, x_{t-n+1})} \equiv f(x_t|x_{t-1}, \dots, x_{t-n+1}) \quad (4)$$

4.3 言語モデルのパープレキシティ

言語モデルの性能を測る尺度としてテストパープレキシティ(PP)が用いられる。PP は

テストデータ系列 $X = x_1, x_2, \dots, x_T$ に対して、以下の値を計算することで求められる。

$$PP = \sqrt[T]{P\{x_1, x_2, \dots, x_T\}} = 2^{\frac{1}{T} \log_2 \sum_{t=1}^T P\{x_t|x_{t-1}, x_{t-2}, \dots\}} \quad (5)$$

PP は単位文字あたりの平均情報量 (bit) を 2 の指数部分に乗せた形になっている。この情報量は言語モデルを用いて系列 X を圧縮した時のデータサイズを表しており、言語モデルが系列 X の統計的特徴をよくとらえているほど小さくなる。従って PP が小さいほど、言語モデルの性能は良い。

厳密には言語モデルの性能は実際に和声認識システムを構築し動作させ式 1 により推定された和声系列の正答率を用いて評価されるべきであるが、音響モデルの影響が無視できないこと、和声認識システム全体の構築に労力がかかることから、言語モデルの性能は PP を用いて評価されることが多い。音声認識の分野ではこの方法がとられている。⁵⁾

本研究では、60 曲分の和声ラベルのうち、ランダムに選択した 30 曲を学習データとして言語モデルの学習に用い、残りの 30 曲を言語モデルの評価用データとして PP の計算に用いた。評価用データ 30 曲から算出された PP の平均値から、言語モデルの PP としている。

4.4 確率値のスムージング

実際の和声認識において和声系列 X の確率を計算する時、あるいは PP を計算する時、 $P\{X\}$ の値が 0、もしくは不定形になってしまうことがある。これは式 4 の計算において、学習データの不足から $N(x_t, x_{t-1}, \dots, x_{t-n+1}) = 0$ であるときに生じる。ただし、どんなに学習データが豊富であってもこのような場合が起こり得ることは避けられない。

従って、場合分けを行い確率値が 0 や不定形になることを避ける。 $N(x_t, x_{t-1}, \dots, x_{t-n+1}) = 0$ であるときには、一つ N の小さい $(N-1)$ -gram を用いて確率値を推定し計算する。その際、元の N -gram から確率値をディスカウント値 λ に従って引き、その総和分を正規化係数 α として $N-1$ -gram に分配する。 λ は $0 < \lambda \leq 1$ の範囲でとる。

$$P\{x_t|x_{t-1}, \dots, x_{t-n+1}\} \quad (6)$$

$$= \begin{cases} \lambda(x_t, x_{t-1}, \dots, x_{t-n+1}) f(x_t|x_{t-1}, \dots, x_{t-n+1}) & \text{if } N(x_t, \dots, x_{t-n+1}) > 0 \\ \alpha P\{x_t|x_{t-1}, \dots, x_{t-n+2}\} & \text{otherwise.} \end{cases} \quad (7)$$

λ は $x_t, x_{t-1}, \dots, x_{t-n+1}$ に応じてそれぞれ変化させる方が言語の性能が良くなる。²⁾ この決め方にはグッド・チューリング法 (GT)、ウィッテン・ベル法 (WB)、絶対法 (Abs) があり、それぞれ以下のようにディスカウント値 λ の計算を行う。

$$\lambda_{GT}(x_t, x_{t-1}, \dots, x_{t-n+1}) = \frac{(N+1)R_{N+1}}{NR_N} \quad (8)$$

$$\lambda_{WB}(x_t, x_{t-1}, \dots, x_{t-n+1}) = \frac{N(x_{t-1}, \dots, x_{t-n+1})}{N(x_{t-1}, \dots, x_{t-n+1}) + R(x_{t-1}, \dots, x_{t-n+1})} \quad (9)$$

$$\lambda_{Abs}(x_t, x_{t-1}, \dots, x_{t-n+1}) = \frac{N(x_{t-1}, \dots, x_{t-n+1}) - \beta}{N(x_{t-1}, \dots, x_{t-n+1})} \quad (10)$$

$$(11)$$

ここで $N = N(x_t, x_{t-1}, \dots, x_{t-n+1})$, R_N は N 回出現した n つ組の種類数、 $R(x_{t-1}, \dots, x_{t-n+1})$ は $x_{t-1}, \dots, x_{t-n+1}$ の後に出現する単語の種類数、 $\beta = \frac{R_1}{R_1 + 2R_2}$ である。

いずれの場合についても、出現回数 N の値が小さく最尤推定された条件付き確率値 $f(x_t|x_{t-1}, \dots, x_{t-n+1})$ の信頼性が低い時に、 λ は小さくなる。

λ_{GT} は N の関数と見ることが出来るが、この N がある k 以上のところで $\lambda_{GT} > 1$ となる可能性がある。この場合は、 k 以上の N に関してはすべて $\lambda_{GT} = 1$ とし、 k 未満の λ_{GT} に若干の補正を施すことが行われる。

4.5 和声系列 N -gram モデルの設計

従来の和声言語モデルは⁴⁾³⁾ 転調を扱えない、もしくは、調の認識をしていない等の問題がある。しかし、節 3.4 で考察したように、調の推定が和声認識の性能向上に有効である可能性がある。そこで、転調をモデル化するために、和声系列の語を、0 から 11 までの値を取る調のクロマ k と機能と和声 h の組 (k, h) で定めた。例えば二長調もしくは二短調の 5 の属七の和音ならば $(k, h) = (2, V7)$ と表される。

データスパースネスの問題を軽減するため、式 4 で推定している条件付き確率は、調を移動しても変わらないという仮定を置くことで、以下のように計算される。ここで Δk は 0 から 11 まで動かして総和をとり、 $k + \Delta k$ の計算は 12 の剰余類における加算とする。

$$P\{(k_t, h_t)|(k_{t-1}, h_{t-1}), \dots, (k_{t-n+1}, h_{t-n+1})\} = \frac{\sum_{\Delta k} N((k_t + \Delta k, h_t), (k_{t-1} + \Delta k, h_{t-1}), \dots, (k_{t-n+1} + \Delta k, h_{t-n+1}))}{\sum_{\Delta k} N((k_{t-1} + \Delta k, h_{t-1}), \dots, (k_{t-n+1} + \Delta k, h_{t-n+1}))} \quad (12)$$

4.6 N -gram モデルの N の値の検証

確率値の平滑化法を GT 法、WB 法、Abs 法それぞれについて N -gram モデルの N を様々な値に変えつつ PP を算出したものを図 6 に示す。WB 法、Abs 法共に $N = 3$ で PP が極小になった。このことから、和声解析において 3-gram モデルが他の N -gram モデルに比べて有効である可能性が示された。

ただし、GT 法に関してはある事前条件 $x_{t-1}, \dots, x_{t-n+1}$ の時、すべての x_t に対して $\lambda = 1$ となるのが原因で、確率値が 0 になる場合が見られ、PP が算出できなかった。

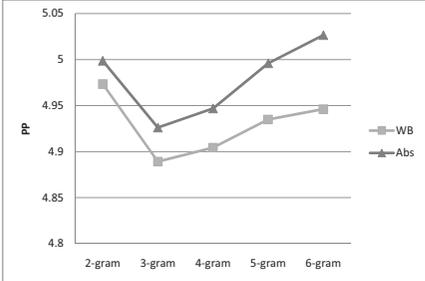


図 6 各 N -gram モデルにおける PP。WB, Abs どちらの場合でも 3-gram モデルで PP が極小になっている。

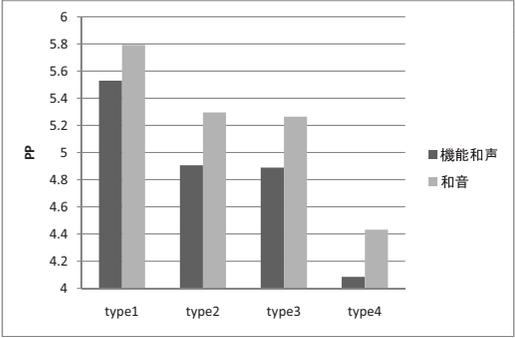


図 7 機能と和声言語モデルと調を考慮しない和音言語モデルの PP 比較。3-gram モデルによる PP で type1 から type4 までの各々算出した。バックオフスムージング法はウィッテン・ベル法を用いた。type1 から type4 全ての場合で、機能と和声言語モデルが、調を考慮しない和音言語モデルの性能を上回っている。

4.7 各曲に対する PP

4 段階に和声の種類を簡略化した場合について PP を計算した。

- (type1) 和声に関して記述された全ての情報を使用
 - 例: (0, I ' ') (0,iv&6 ') (0,iv7 ') (7, V7 ' ') (0, V7) (0, !V7 ' ') (5, V7') (2, V9-')
- (type2) 転回形の情報を無視
 - 例: (0, I) (0,iv&6) (0,iv7) (7, V7) (0, V7) (0, !V7) (5, V7) (2, V9-)
- (type3) type2 に対して更に、11 度 13 度音の付加/変位/省略、テンションを無視
 - 例: (0, I) (0,iv) (0,iv7) (7, V7) (0, V7) (0, !V7) (5, V7) (2, V9-)
- (type4) type3 に対して更に、ベース、根音、5 度音、9 度音の付加/変位/省略を無視
 - 例: (0, I) (0,iv) (0,iv7) (7, V7) (0, V7) (0, V7) (5, V7) (2, V7)

3-gram モデルを用いて評価表データ 30 曲の PP を計算したものを図 8 に示す。大多数の曲に関して、type1 から type4 にかけて PP が減少している。これは和声の種類そのものが減少すると、曖昧性が減少する傾向にあるためである。

サティのジムノペディ 1 番は type4 においても PP が高い。これはこの曲中の I,IV の和音がほぼすべて四和音から構成されている、学習データを基準として見た時珍しい和声で構成されているためであると考えられる。

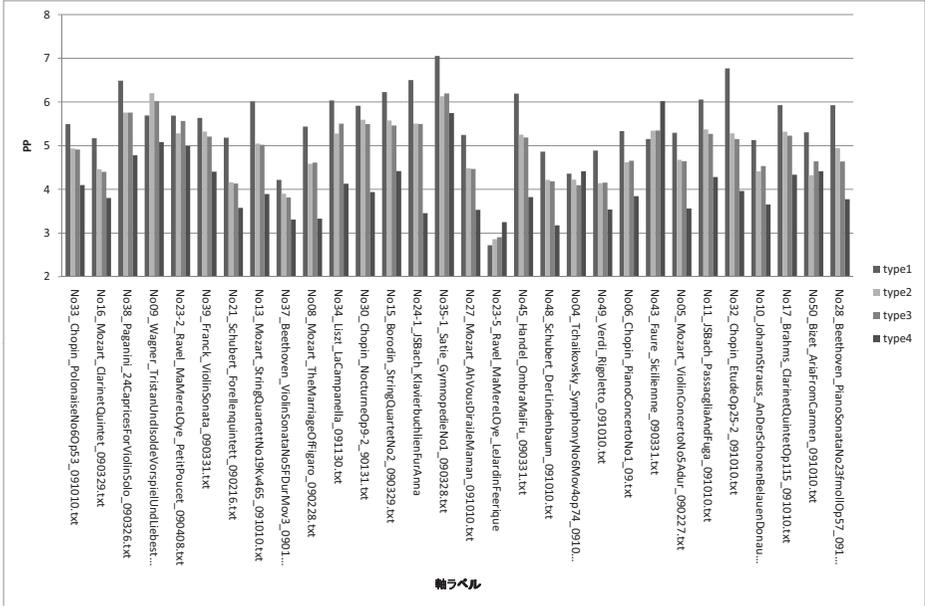


図 8 3-gram モデルによる各曲の PP。バックオフスムージング法はウィッテン・ベル法を用いた。

type4 の PP が約 5 以上の曲はすべて近代の作曲家による曲、例えばワーグナーのトリスタンとイゾルデ、ラヴェルのマ・メール・ロワ、サティのジムノペディ1 番、フォーレのシシリエンヌである。type4 は、和声の基本的な情報であり、例えば、ドミナント進行、属七の和音の現れ方などを見ていると考えられる。PP を評価する言語モデルの学習データの大部分が、和声学に比較的好く従う近代以前の曲によって占められているとすれば、これらの近代の曲は古典的な和声学からの逸脱する傾向にあるため、PP が高くなったと考察出来る。

対して、type1 ではこのような時代の差による PP の明確な違いは見いだされなかった。type1 では、より詳細な和声の情報に基づき PP を評価している。詳細な和声の情報には、和声学に厳密に縛られにくい情報が含まれやすいと考えられるならば、type1 による PP の評価は時代背景よりも各曲固有の和声的特徴に強く左右されやすいと考察出来る。

4.8 調を考慮した機能と声の有意性の検証

節 3.4 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。この節では言語モデルの PP を比較することにより、より直接的に機能と声モデルの有用性を示す。

図 7 に、機能と声の言語モデルと和音の言語モデルの PP を比較したものを示す。語の単位が 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 3.4 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. 結 論

和声言語モデルによるモデル化の対象である和声系列に対する知見を得るために詳細な和声ラベルデータの仕様を決定し、実曲 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。

参 考 文 献

- 1) Pereira, F., Y. Singer, N. Tishby. "Beyond Word N-grams." Proceedings Third Workshop on Very Large Corpora, MIT, Cambridge Mass., 95-106. March 1995.
- 2) 鹿野清宏, 伊藤克亘, 河原達也, 武田一哉, 山本幹雄: 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, March, 2008.
- 4) 上田雄他, "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) 森信介, "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.