

# クラシックギターの指導現場における 知識の体系化とその実現に向けた検討

## A Study of Classical Guitar Lessons for Knowledge Systematization

三浦寛也<sup>1\*</sup> 飯野なみ<sup>1,2</sup>  
Hiroya Miura<sup>1</sup> Nami Iino<sup>1,2</sup>

<sup>1</sup> 理化学研究所革新知能統合研究センター

<sup>1</sup> RIKEN Center for Advanced Intelligence Project

<sup>2</sup> 産業技術総合研究所

<sup>2</sup> National Institute of Advanced Industrial Science and Technology

**Abstract:** This paper presents the features of the one-on-one classical guitar lessons of a particular pair. The aim of our research are clarify the patterns of interactions between speeches and performances and visualize the change of student's performances. In this study, we report the results of the acoustic analysis of the audio data and the conversation analysis of actual contents of speeches.

## 1 はじめに

本研究の目的は、演奏指導現場における音声データを用いた知識構造化データの獲得と楽譜へのアノテーションに基づく知識発見を支援するシステムの開発である。近年、音楽教育の場において、より細かな指導が重視される個人レッスンの需要が高まっている。当該場面においては、演奏に加え議論を交わすことがあるため、演奏練習の音声を収録することが多く、記録したデータを有効活用するためには、練習時の発話や演奏状況の効果的な記述が必要となる。このような背景から、筆者らは、これまで楽器演奏における教授学習場面を対象とした、より持続的に知識を蓄積し獲得していくためのナレッジマネジメントの基盤構築のための研究を進めてきた。具体的な取り組みとして、これまでおよそ5年に渡り、個人レッスンに着目してクラシックギターの個人レッスン内で起こる演奏や発話などの音声データを収録している。

楽器演奏の個人レッスンには、極めて個人差が大きい内容を孕んでいる。しかし共通して生徒の上達支援と指導の質的向上を目的としている。積み上げられていった

その過程から、学習者や指導者の思考を知ることは不可欠である。このような個別の事象を扱う研究では、ボトムアップ的に複数の事象の中から、共通となる特徴を発見していくことが期待できる。そこで本稿では、先述した目標達成のための第一歩として、特定のペアにおけるクラシックギターレッスンの複数回にわたる音声データを対象に、教授学習場面での発話と演奏のインタラクションのパターンや、実際の発話内容に着目した分析を進め、指導内容の変遷やその特徴、そして協働的な演奏構築過程がいかに行われるのかを明らかにする。

楽器指導に関する先行研究はこれまで教育工学や芸術学の観点から述べられることが多かったが、近年では、コンピュータを用いた演奏練習の支援が数多く行われてきている [坂本 15]。具体的な事例として、遠隔隔で同期式のアンサンブル練習を行うためのプロトタイプツールキット [Akoumianakis 08] や、音楽初心者向けのリアルタイムな音楽コラボレーションシステム [Gurevich 06] などが開発されてきた。また、音楽理解や音楽学に関する支援のための研究では、演奏を録音し音程の情報を多角的に表示することでピッチの揺れやずれを指摘するシステム [Lim 09] や、オーケストラスコアの理解を支援するシステム [松原 12]、共同演奏における演奏者間のコミュニケーションモデルを調査により構築した研究 [Kawase 07] などがある。しかしこれらの研究の多くは、大型ディスプレイや専用アプリケーションなどの環

\* 連絡先: 理化学研究所 革新知能統合研究センター, 〒103-0027  
東京都中央区日本橋1丁目4-1 日本橋一丁目三井ビルディング 15 階, E-mail: hiroya.miura@riken.jp

境に関する制約が多く、実際の指導現場を科学的な観点から分析・支援する研究はあまりない。

## 2 本研究の最終目標

本研究の最終目標として、演奏指導現場におけるマルチメディア情報を用いた情報構造化と楽譜へのアノテーションによる、再利用可能な知識の抽出を実現するシステムの開発を実現する。この目標を達成するためには、楽器演奏の指導現場で記録された発話や演奏状況を効果的に記述するための枠組みを提案し、以下の機能を包含するようなシステムの開発が求められる；レッスン中に記録された発話を抽出し譜面上に表示する機能、各発話と楽曲との対応関係を特定する機能、タブレット端末等で撮影した写真や動画を読み込み、楽譜上の適切な箇所に関連付ける機能、発話情報とオントロジーとの関連付けをおこなう機能(図1)。

目標達成のための具体的なアプローチとして、指導者と学習者の発話内容やそタイミング、演奏音と発話との前後関係から、演奏指導に起こる発話がどの演奏音に対するものかを自動推定することで、構造化された知識データを生成し、そこから再利用可能な知識の体系化を目指す。また、レッスンで起こる発話には、同一あるいは類似テーマという関係付けを同定し、類似性計算やアノテーションの付与によって、レッスン集合を構造的に表現することができると考えている。本研究では、その初期検討として、主に音声区間分析および発話分析について取り組み、そこで得られた成果や知見をもとにシステムの構築を目指す。

## 3 教授学習場面の音声区間分析

本節では、レッスン中に起こる事象を正確に記述するための方法として、音声データの分析からレッスン中に発せられる音声情報の同定とその区間検出を実施した。一般に、演奏指導を実施する際に考慮する必要がある主要素は、演奏音と発話である。これらを統合的に組み合わせることによって、演奏者は独自の演奏表現を構築する。楽器演奏における教授学習の相互行為を記述した研究では、指導者と学習者両者の発話と演奏音に着目した分析が主流となっている [Paul 94, Suchman 87, Sudnow 78]。これらの先行研究に倣い、本研究で対象とする音声データは、演奏音 (music)、発話 (speech)、ノイズ (noise)、無音 (noEnergy) の4つの要素に着目した分析をおこなった。

まず、演奏音に関して、一般的にクラシックギターの

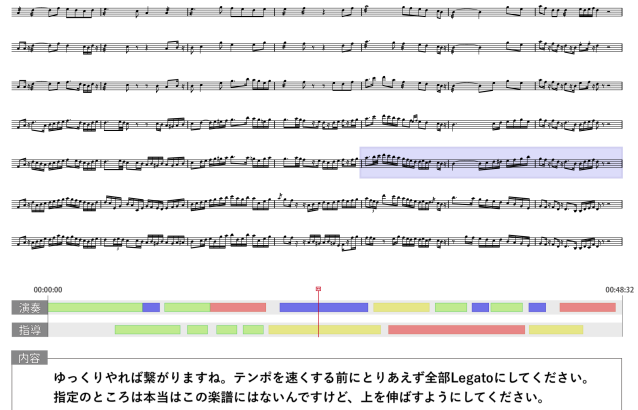


図1 本研究で開発するシステムの想定図

指導現場では、指導者と学習者のためにギターが2台必要となる。2つの楽器音の違いとして、楽器の音響的な特徴や演奏者の弾き方などが複雑に混ざっている。次に発話に関して、個人の特徴があるが、声道長や声帯などの先天的な発声器官の個人差によるものと、アクセントなどの後天的な話し方のくせによるものが相互にからみ合っている。上述以外の音声として、メトロノームや環境音などがノイズの一例である。また、クラシックギターには非常に多くの演奏法があり、中でも効果音楽法<sup>\*1</sup>は、楽器の胴体をタップしたり弦を引っ掻いたりするため、通常のクラシックギター音とは区別されたり雑音として扱われることが予想される。このような多種多様な音声混在するデータから、音声情報と演奏区間の同定によって指導者と学習者の演奏活動を正確に記述することで、より効果的なレッスン指導への応用や、指導者と学習者の演奏構築過程の解明に寄与できる。

### 3.1 対象データ

本研究で対象とするレッスン形態は、クラシックギターのマンツーマンレッスンである。指導者はプロの演奏家として活動している成人女性であり、国内トップレベルのギターコンクールを複数優勝した経験を持つ。指導歴は7年で大学生から高齢者まで幅広い年齢層の指導にあたっている。学習者はアマチュアの成人男性であり、コンサートやアマチュア向けのギターコンクールの出場に向けて積極的に取り組んでいる。このペアの受講歴は5年と比較的長く、良好な関係が構築されている。

音声データは、学習者が個人の録音機材 TASCAM

<sup>\*1</sup> ギター奏法オントロジー [Iino 19] で定義されている、効果音を出すことを目的とした演奏法。<https://github.com/guitar-san/Guitar-Rendition-Ontology>

表1 対象データと音声区間分析結果の一覧

ID	楽曲 / 作曲者	日付	時間	演奏 (music)	発話 (speech)	ノイズ (noise)	無音 (noEnergy)
01	El Testament d'Amelia / F. Tarrega	2018/07/19	22:02	55.9%	29.5%	12.9%	1.7%
02		2018/07/26	18:42	62.7%	27.0%	2.7%	7.5%
03		2018/08/03	25:20	<b>91.6%</b>	<b>6.1%</b>	1.5%	0.8%
04		2018/08/09	28:56	84.0%	12.0%	1.3%	2.7%
05	Introduction and Variations on a Theme by Mozart / F. Sor	2018/07/26	14:06	77.7%	17.2%	0.5%	4.6%
06		2018/08/03	18:25	<b>94.0%</b>	<b>2.1%</b>	1.1%	2.8%
07		2018/08/09	25:16	79.5%	15.3%	0.9%	4.3%
08	Olite from Castillos de Espana / F. M. Torroba	2018/10/04	15:13	78.0%	17.5%	2.1%	2.3%
09		2019/01/17	24:48	86.3%	5.7%	8.0%	0.0%
10		2019/02/15	28:10	87.5%	10.0%	0.4%	2.1%
11		2019/03/08	21:23	93.9%	2.3%	0.7%	3.1%
12		2019/05/30	31:15	<b>95.8%</b>	<b>1.8%</b>	2.2%	0.1%
平均				82.2%	12.2%	2.9%	2.7%

DR-05 を用いて約 1 年間に及ぶレッスンを記録した結果であり、これまでに約 100 件以上のデータを収録している。本稿では、その中から本著者らがそれらを楽曲ごとに分類し選定したものである。具体的には、1 曲につき 3 回以上のレッスンを実施した次の 3 曲を対象とし、全 12 件分の WAV データの分析を行った；(a) El Testament d'Amelia (アメリカの遺言), (b) Introduction and Variations on a Theme by Mozart (魔笛の主題による変奏曲), (c) Olite from Castillos de Espana (スペインの城よりオリーテ)。

### 3.2 分析手法とその結果

対象とする音声データの解析手法として、Doukhan が提案した inaSpeechSegmenter[Doukhan 18] を用いて、音声情報の同定とその区間検出をおこなった。inaSpeechSegmenter は、男女の発話時間割合推定に基づく大規模な男女平等研究を行うために設計された CNN ベースの音声区間検出フレームワークであり、音声信号を音楽、音声、雑音の均質なゾーンに分割することができる。inaSpeechSegmenter による分析から、先述した演奏音 (music)、発話 (speech)、ノイズ (noise)、無音 (noEnergy) の 4 要素に分類した結果を表 1 に示す。

本分析結果では、music が音声情報の中で最も割合が大きかった要素であった。一方で speech は非常に小さく、一部は noise や noEnergy とほぼ同等であった。本結果を概観すると、各レッスンの初期段階 (ID:01, ID:05, ID:08 など) においては、発話区間 (speech) の割合が多く、レッスン回数を重ねるごとに (ID:03, ID:06, ID:12 など) 減少していくという一定の傾向が見られた。これは、教授学習場面の序盤では、口頭による指導が多いが、終

盤になるにつれて実演する機会が増していくことが推測できる。このように音声区間分析の結果から、特定の楽曲を熟練する過程で発話と演奏のインタラクションやそのパターンに変化があることが伺える。一方で、技術的課題として、本分析では演奏中 (music) の発話 (speech) は全て music としてタグ付けされるため、演奏と同時に発せられた発話について考慮されていない可能性があると考えられる。また、対象データでは、指導者と学習者の歌唱頻度が高かったことも、その結果、music の割合が大きくなった一要因である考えられる。

音声区間の分割において、music の割合が大きかったことを受けて、手動による書き起こし結果をもとに音声区間の再分類を行った。図 2 は、ID: 03 の結果を可視化したものである。図内の最上部は inaSpeechSegmenter によって分割された項目の変化、中央は指導者の変化を、最下部は学習者の変化をそれぞれ示している。青色は music、赤色は speech、水色は noise と noEnergy である。薄い赤色は、inaSpeechSegmenter では music に分類されたが、実際にはその間に発話をしてきた箇所として、music&speech と分類した。また黒色は、inaSpeechSegmenter で speech に分類されたが、発話を聞いていた場合を表しており、no-speech と分類した。

図 2 の結果から、指導者と学習者の変化に着目すると、music&speech の割合が非常に多く見られた。特に多くの音楽レッスンでは発話が大半であり、学習者は傾聴と演奏を繰り返している。しかし今回の結果からは、学習者の music&speech も非常に多く、活発な会話が行われていたことを示している。これは、両者が互いに発言しやすい良好な関係が築けていることを示唆している。このように、音声区間の情報が指導者と学習者の関

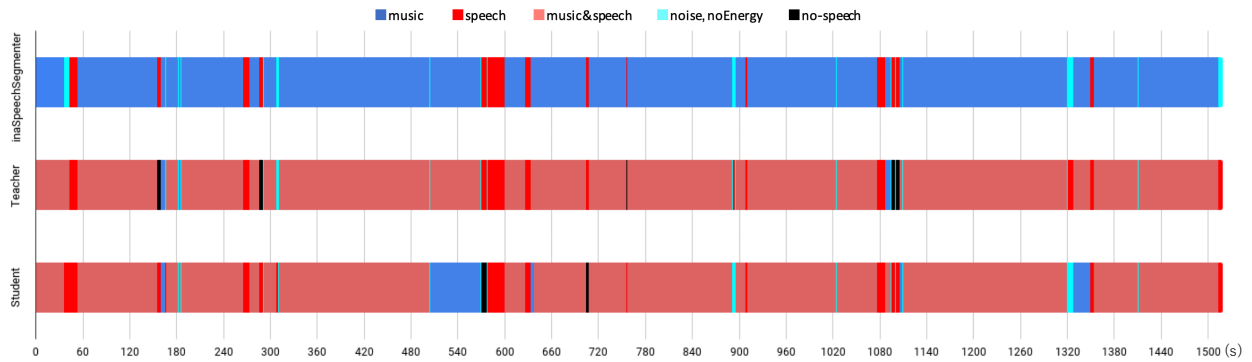


図2 ID:03 (El Testament d'Amelia) を対象とした音声区間分析の結果

係性を表す一つの指標になる得ることが推測される。

## 4 書き起こしデータに基づく発話分析

教授学習場面における発話と演奏の状況やそのパターンを明らかにするため、3.1節で詳述した音声データ全12件を対象に、発話の書き起こしとその分析をおこなった。書き起こし作業は筆者ら2名が実施した。各発話は、前後の発話との時間間隔や意味的なまとまりに考慮した上でセグメント化し、発話IDや発話者、発話の開始/終了時刻、発話長などの情報も同時に記録した(表2)。

### 4.1 形態素解析結果から見る教授学習場面の傾向

書き起こしたID:01からID:04までの4件のデータに対して、形態素解析を行った。その結果、名詞は、指導者：523, 168, 217, 268, 学習者：510, 166, 124, 91, 動詞は、指導者：445, 132, 190, 213, 学習者：377, 95, 86, 49, 形容詞は、指導者：110, 43, 72, 66, 学習者：68, 20, 17, 16 (ID: 01~04の順)の単語数が抽出された。顕著な特徴としては、名詞に分類された単語の中で「これ」「それ」といった指示語が多くみられたことである。名詞の約20%が指示詞であることがわかった。これは、楽譜を用いた指導であることや、弾き方や音色に関する内容が多いためだと考えられる。特にID:01では、音色の話から「爪の削り方<sup>\*2</sup>」に関する指導が行われ、爪の箇所や形を表す際に指示詞を多く用いた。このように、

<sup>\*2</sup> クラシックギターは、基本的に演奏者の地爪で弦を弾くため、爪の形状や磨きの細かさによって音色が変化する。

表2 ID:03 (El Testament d'Amelia) を対象とした書き起こしデータと意味ラベルの付与

ID	発話内容	意味ラベル
43	あと、タッタタって切れるのはやめましょう。	Intonation
44	あくまで、レーファーっていう旋律なので、その中にレーレフファーって息吸うことはないと思います。	Intonation
45	あとこう振りかぶらないで。	Technical Facility
46	弾けなくなっちゃう。	Technical Facility
97	でもあのスピードで弾けるようになってるのであれば本番ね、あの、上がってもそこまでしかいかないので。	Tempo
98	あの、だからこれからはちょっと落として弾くようにして、上がってもそれくらいになるようにしていきましょう。	Tempo
99	あとね、なんかね、ちっちゃくなったりするので、もうちょっとはつきり一個一個出していいと思います、音がちっちゃい。	Dynamics

指示詞を抽出することで、楽譜のどの情報に対して発言されているのか、あるいはどのような指導が行われているのかを推測できる可能性がある。

また、重要単語の同定として、1レッスンデータごとに含まれる名詞と形容詞を抽出し、TF-IDF法の適用によって得られる単語群の上位10%を抽出した。分析結果から、各レッスンの初期段階(ID:01, ID:05, ID:08など)では、“伸び、溜め、雰囲気、ストーリー、コント

ロール”といった単語が得られる一方で、レッスン回数を重ねるごとに (ID:04, ID:07, ID:12 など) “Harmonics, Vibrato, Portamento, Ceja” といった音楽用語やクラシックギター奏法などの固有表現を含む単語が上位に選定されることが分かった。これは、広義の意味を含む内容から、より詳細かつ専門的な技法に関する指導へと教授内容が変化していることが推測できる。このように、形態素解析によって、レッスン固有の特徴や傾向を正しく理解するための手掛かりになると考えられる。

## 4.2 教授場面における意味ラベルの推移

本研究で取り扱うレッスンでは、具体的な指摘や助言などは、主に指導者の発話に現れる。そのため、本研究では、各発話の内容を把握する指標として、先行研究 [Ellen 03] を参考に、指導者の各発話に対して、テンポ (Tempo)、技法 (Technical Facility)、音楽の表現 (Intonation)、リズム (Rhythm) 音の強弱 (Dynamics)、その他 (Other) の6項目の意味ラベル (以下、“意味ラベル”と呼ぶ) を手動で付与した (表2)。本ラベルの付与により、教授学習場面において、どのようなタイミングでこういった指示が行われているのかを定量的に判断することができる。

ID:01~04 (El Testament d’Amelia) を対象とした意味ラベルの付与結果として、表3に意味ラベルの割合と、図3に全体の推移を示す。表3より、テンポ (Tempo) に関する意味ラベルが 32.6% と最も割合が大きかった要素であった。続いて、音楽の表現 (Intonation) の意味ラベルが 31.5% と高い割合で出現した。一方で、音の強弱 (Dynamics) についての割合は低く、全体を通して教授されることが少ないことが分かった。また、図3の結果から、音楽の表現 (Intonation): 10, 5, 65, 49, リズム (Rhythm): 1, 7, 1, 21 (ID: 01~04 の順) とレッスン回数を重ねるごとにその値が増加していくことが分かった。この結果から、全体的な傾向として、技術的な教授内容から、より音楽的な表現や全体のリズム感を練っていく教授内容に変化していることが示唆される。以上より、意味ラベルの付与とその解析によって、複数のレッスンを横断する演奏構築過程の解明に寄与できると考えられる。

## 5 おわりに

本稿では、クラシックギターの指導時の収録音源を分析し、指導現場における様々な特徴や問題点について議論した。具体的には、畳み込みニューラルネットワーク (CNN) ベースのフレームワークを用いて音声区間の検

表3 ID:01~04 (El Testament d’Amelia) に含まれる意味ラベルの割合

意味ラベル	回数	割合
テンポ (Tempo)	133	32.6%
技法 (Technical Facility)	77	18.8%
音楽の表現 (Intonation)	129	31.5%
リズム (Rhythm)	30	7.3%
音の強弱 (Dynamics)	13	3.2%
その他 (Other)	27	6.6%
合計	409	100.0%

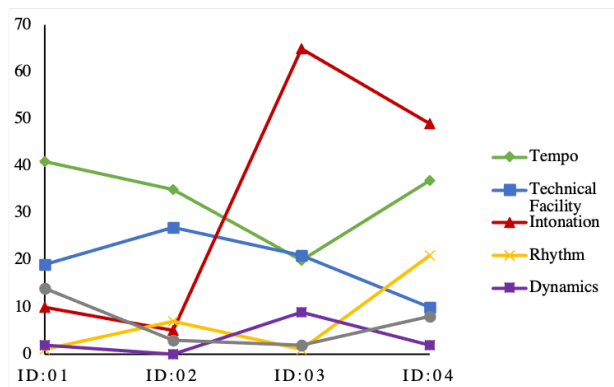


図3 ID:01~04 (El Testament d’Amelia) を対象とした意味ラベルに関する推移

出とその可視化を行った。さらに実際の音声データから発話を書き起こし、形態素解析による会話分析、そして意味ラベルの付与とその解析により、楽器指導における現場固有の特徴や傾向を提示することができた。

今後、第2章で先述したシステムや協働的な演奏構築過程の可視化に向けて、(a) 一人称研究 [諏訪 13] と身体知研究 [古川 06] のような個の事象に焦点を当てたより主観的な研究、(b) 機械学習のような客観的かつ普遍的な研究を、並行して進めていく。これにより、様々な事象に対応しうる知識の体系化を目指す。

## 参考文献

- [Akoumianakis 08] Akoumianakis, D., Vellis, G., Milolidakis, L., Kotsalis, D. and Alexandraki, C.: Distributed collective practices in collaborative music performance, ACM DIMEA: Digital Interactive Media in Entertainment and Arts, pp. 368–375 (2008).
- [Doukhan 18] Doukhan, D., Carrive, J., Vallet, F., Larcher, A., and Meignier, S.: An Open-Source

- Speaker Gender Detection Framework for Monitoring Gender Equality, IEEE International Conference on Acoustics Speech and Signal Processing, ICASSP, Vol. 7, pp. 103–122 (2018).
- [Ellen 03] Ellen, M.: A Descriptive Analysis of Error Correction in Instrumental Music Rehearsals, Journal of Research in Music Education, Vol. 51, No. 3, pp. 218–230 (2003).
- [古川 06] 古川康一: 身体知としての弦楽器演奏スキル, バイオメカニズム学会誌, Vol. 30, No. 1, pp. 17–20 (2006).
- [Gurevich 06] Gurevich, M.: JamSpace: A networked Real-time Collaborative Music Environment, ACM CHI Extended Abstracts on Human Factors in Computing Systems (2006).
- [Iino 19] Iino, N., Nishimura, S., Nishimura, T., Fukuda, K., Takeda H.: The Guitar Rendition Ontology for Teaching and Learning Support, The 13th IEEE International Conference on Semantic Computing (ICSC), Resource track, Vol. 1, pp. 404–411 (2019).
- [Lim 09] Lim, K. A. and Raphael, C.: InTune: a musician’s intonation visualization system, ACM SIGGRAPH (2009).
- [松原 12] 松原正樹, 諏訪正樹, 斎藤博昭: インタラクティブな楽譜色付けによるオーケストラスコア理解支援システム, 人工知能学会論文誌, Vol. 27, No. 5, pp. 281–295, (2012).
- [Kawase 07] Kawase, S., Nakamura, T., and Draguna, M.: Communication channels performers and listeners use: a survey study, Proceedings of the International Conference on Music Communication, pp. 76–79 (2007).
- [難波 97] 難波精一郎 編集: コンピュータと音声, 音の科学, 朝倉書店, pp. 103–118 (1997).
- [Paul 94] Paul F. Berliner: Thinking in Jazz: The Infinity Art of Improvisation, Chicago, University of Chicago Press (1994).
- [坂本 15] 坂本雄彦: 演奏練習における音声ログの楽譜への対応付けと可視化および演奏に関する議論への活用, 筑波大学大学院博士課程 システム情報工学研究科修士論文 (2015).
- [Suchman 87] Suchman, L. A.: Plans and situated action, New York: Cambridge University Press (1987).
- [Sudnow 78] Sudnow, D.: Ways of the hand, Cambridge, Harvard University Press (1978).
- [諏訪 13] 諏訪正樹, 堀浩一: 特集「一人称研究の勧め」にあたって, 人工知能学会誌, Vol. 28, No. 5, pp. 688 (2013).