

複数回にわたる楽器指導の特徴分析 —ギターレッスンを対象として—

Analysis of Instrumental Instruction Across Multiple Times - A Focus on Guitar Lessons -

飯野なみ^{1,2*} 三浦寛也² 武田英明¹
Nami Iino^{1,2} Hiroya Miura² Hideaki Takeda¹

¹ 国立情報学研究所

¹ National Institute of Informatics

² 理化学研究所

² RIKEN Center for Advanced Intelligence Project

Abstract: In instrumental instruction, there exists a unique form of interaction where musical performance and verbal communication intersect. Particularly, when multiple lessons are conducted for a piece of music, there may be specific trends and patterns in the process in terms of the student's progress and the teacher's perspective. In this study, we analyzed audio data recorded from one-on-one classical guitar lessons to visualize the implicit aspects of instructional processes.

1 はじめに

本稿では、楽器指導における非明示的な指導過程の可視化を目的として、複数回にわたって実施されたギターレッスンの音声データを対象とした分析について述べる。通常、楽器指導(以降、レッスンと呼ぶ)は生徒の目的に合わせて選定された特定の楽曲に対して何度も行われる。そのレッスン過程では、生徒の上達や教師の着眼点などの側面で特徴的なパターンや変化が見られる可能性がある。特に、レッスン特有の演奏と発話によるインタラクションや教師の指導内容に着目することによって、レッスンの様相を明らかにすることが期待できる。

このような背景から、我々はこれまでクラシックギターの個人レッスンを対象として、実際のレッスン現場で録音された音声データの収集と分析を進めてきた[1, 2]。現在、音声データは対面レッスンとオンラインレッスンを合わせて100件を超えるが、簡易的に収録されたデータであるために、音声区間や発話の情報を適切に抽出することが難しいという問題がある。中でも人手による発話の書き起こしは人的コストが大きく、これまでの先行研究で扱ったデータ数は極めて少なかった。

そこで本研究では、データの拡充とレッスン内の状

態を可視化することを試みた。データの拡充では、教師と生徒のペアを増やし、既存の音声区間検出フレームワークや音声認識モデルを用いて分析データを作成した。レッスンの分析では、教師の発話に焦点を当て、指導内容の意味的要素を表現したラベルを定義・付与し、レッスン内で起こる発話と演奏の状態遷移を調査した。さらに状態遷移については、要素間の因果関係や包含関係を体系化するために、ISM法を適用した構造分析も実施した。結果として、レッスンのフェーズや楽曲によって様相が変化することが明らかとなった。

楽器の個人レッスンにおけるインタラクションに関する既存研究は、ピアノを対象としたものが多く見られる。グループレッスンに比べて個人レッスンは教師が支配的であり、生徒の相互作用は限定的であることを示した研究[3]や、教師と生徒によって演奏表現が協働的に構築されていくことを明らかにした研究[4]などがある。また、ジャズピアノの即興レッスンでは、教師によってフレーズが再構成されていくことを示されている[5]。このような教師と生徒のバランスについて、複数回のレッスンを考慮した研究は少ない。そのため、本研究は教師と生徒のインタラクションの変遷を明らかにするという点で新しく、かつ挑戦的な試みと言える。

*連絡先：国立情報学研究所
〒101-0003 東京都千代田区一ツ橋2丁目1-2
E-mail: nami-iino@nii.ac.jp

表 1: 分析対象であるレッスンデータの概要

ID	ペア	形態	回数	合計時間	曲名/作曲者	時代	フェーズ
1	T1-S1	対面	2	0:11:14	8 度のエチュード/A. Cano	古典派	初期
2			0:07:38	6 度のエチュード/F. Sor	古典派	初期	
3	T2-S2	対面	4	1:35:00	アメリカの遺言/M. Llobet	ロマン派	後期
4			0:57:47	魔笛の主題による変奏曲/F. Sor	古典派	後期	
5			2:00:49	スペインの城よりオリーテ/F. M. Torroba	ロマン派	初期	
6	T2-S3	オンライン	5	2:17:19	盗賊の歌/M. Llobet	ロマン派	後期
7			3	1:15:46	マリアルイサ/J. S. Sagreras	ロマン派	後期

2 分析用データの作成

本章では、クラシックギターの個人レッスンで収録された音声データから、音声区間や発話情報といった基本的な情報抽出を行う。さらに、指導の変遷を調査するために、抽出した発話情報に対して意味付加を行う。

2.1 音声区間検出と発話の書き起こし

表 1 に、データの概要を示す。1 曲につき 2 回以上のレッスンが実施された 7 曲を対象として、全 24 件の音声データである。各ペアの音声データは、T1-S1 は Youtube に公開されているレッスン動画¹から、T2-S2 は対面レッスン時に IC レコーダーで録音された音声ファイルから、T2-S3 は Zoom で記録された動画から、それぞれ収集した。ここでのフェーズとは、レッスンが実施された時期、あるいは生徒の楽曲に対する取り組み具合を意味し、初期は新しく取り組み始めた楽曲を扱ったレッスン、後期は生徒が過去に演奏したことがある楽曲のレッスン、あるいは本データ以前にも指導が行われていた場合のレッスンと位置付けた。

本研究ではまず、新たに追加した ID1, 2, 6, 7 の音声区間検出を行った。レッスンにおいて主要素である演奏区間を検出するために、CNN ベースの音声区間検出ライブラリである inaSpeechSegmenter[6] を用いて、演奏 (Music)、発話 (Speech)、ノイズ (Noise)、無音 (NoEnergy) の 4 つの区間を検出した。ただし、楽器レッスンのような演奏と発話が重複する場面においては、Speech が Music に分類されるという問題があった。

そこでより適切な Speech 情報を獲得すべく、次に発話の書き起こしと区間 (セグメント) 情報の抽出を行った。書き起こしの自動化を進めるために、予備調査として YouTube, Google Cloud Speech-to-Text, Whisper, Amazon Transcribe の精度を比較した。その結果、Whisper の精度が 80% を超え、それ以外は 5 割を下回っ

た。Whisper は、OpenAI が開発した音声認識モデルであり、セグメント情報の取得も可能である。そのため、本研究ではこのモデルを採用して書き起こしを実施した。ただし、話者分離には対応しておらず、かつ既存モデルと組み合わせた場合も精度が低かったことから、手動で教師と生徒の発話を分離した。結果として、総セグメント数は、教師 (T-Speech) は 1669 件、生徒 (S-Speech) は 981 件となった。

2.2 教師の発話に対する意味的要素の付加

教師は、演奏あるいは演奏者の現状を整理し、目標とする演奏に向けた改善点や練習方法を示す。我々は、このような教師の指導における思考過程を体系化すべく、指導コンテンツラベル (Instructional Content Label, ICL) と名付けたラベルを定義した。これは、既存研究 [7] に基づいて定義したものであり、以下の 6 つのラベルで構成され、発話の内容がどのような意味を持つかを分類できる。

- Giving Objective Information (GSI) : 教師の主観に基づく一般的、あるいは特定の概念の情報提供。感想や印象に加え、独り言や雑談を含む。
- Giving Subjective Information (GOI) : 楽譜や音楽関連の知識など、参照可能な事象や概念に基づく客観情報の提供。
- Asking Question (AQ) : 質問、予測、問いかけ。「ここから弾いて」などの指示も含む。
- Giving Feedback (GF) : 生徒の演奏に対する具体的な情報提供。くせなどに関する評価・分析・見立て。
- Giving Practice (GP) : 特定のパッセージの練習方法の提示。部分的な改善・改良に関するコメント。
- Giving Advice (GA) : 目標提示、アドバイス。特定の目標達成へ導くための具体的な意見や推奨。演奏全体へのコメント。

¹ 【クラシックギター レッスン】のえレッスンより Vol.3, 4
https://www.youtube.com/watch?v=ZPjGTsHYwho&list=PL_1N47mmXpTad57CmFKvXIZU9LJMqR0U79&index=7

表 2: 各種音声区間における発話と指導コンテンツラベル (ICL) の一例

種類	開始時間	終了時間	発話内容	ICL
Music	18:30	20:22		
T-Speech	18:35	18:43	あとねもうちょっとなんかこの、あの、はっきり音を出せない、欲しいな。	GSI
T-Speech	19:00	19:05	これもちゃんとあの4分音符分伸ばしてください。	GP
S-Speech	19:06	19:09	あそうか、伸びてない。	
T-Speech	19:13	19:14	そうそう。	GSI
T-Speech	19:56	20:00	今までより音確りクリアで良かったです。	GSI
T-Speech	21:32	21:34	変な音出さない。	GF
S-Speech	21:41	21:43	消えちゃうな。	
T-Speech	21:48	21:50	3ですよ。	GOI
T-Speech	22:00	22:19	もうちょいね、あの、調整を、していただければ。いやでも素晴らしい、いやだからあの、弾くのがどうこうじゃなくて、音楽的な部分、伝えなきゃいけないので、聴衆の人に、	GA
Music	22:08	22:19		

ICL をセグメントごとに付与した一例を表 2 に示す。セグメントの一部には、相槌だけなど明確に分類できない発話や、一つのセグメント内に複数のラベルが該当する発話があった。前者に関しては、前のトピックを引き継いでいたり、暗に示している場合があるため、全てのセグメントに対して前後を考慮してラベルを付与した。後者に関しては、ラベルの優先順位を GSI < GOI < AQ < GF < GP < GA と定め、優先されるラベルを付与した。そのほか、(a) GP は変化を促すものとして、「～してみるのもあり」といった表現の提案を含む、(b) 右手の弾き方は部分的な要素が強いため GP とする、などいくつかのルールを設けた。ラベルの付与は、本著者を含む 2 名が吟味し一意に定め、二者間のアノテーション一致度は 94.5% であった。

(GA) をする様子が多かった。また、1 回目のレッスンではラベルが短い区間で頻繁に変化する傾向が見られたが、レッスンが進むにつれてラベルの変化が少なくなった。このようなレッスンの回数に関する変化は、レッスンのフェーズに着目することで特徴を捉えることができると考えられる。

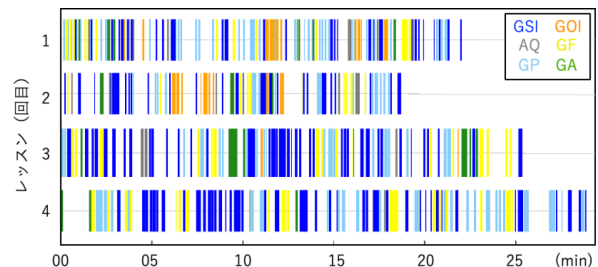


図 1: 教師の発話における ICL の時系列変化の一例

3 ギターレッスンの状態遷移の観察

前節で作成した分析用データを用いて、レッスンの様相を可視化する。

3.1 レッスン内の ICL の変化

図 1 は、ID3 における指導コンテンツラベル (ICL) の時系列変化を示している。例えば、GF (例: なってしまっている) と GP (例: ~したほうがよい) が隣接するパターンが多くみられた。また、レッスンの最初と最後は、GSI になる傾向があった。他のレッスンにおいても、生徒がまず楽曲を通して演奏することが多く、それに対して教師はいきなり具体的な指導 (GP) をするのではなく、パフォーマンス全体に関するコメント

3.2 演奏と発話の状態遷移

本節では、レッスンのフェーズにおける演奏および発話区間の共起関係を調査した。教師の発話 (T-Speech) を 6 種に分類した ICL に加え、演奏 (Music) および生徒の発話 (S-Speech) による合計 8 つの要素を対象として、bigram を抽出し、Python のライブラリである NetworkX を用いて有向グラフを生成した (図 2)。矢印に傍近の数値は、割合である。S-Speech については、発話の割合が少ないことや指導の意味的要素を含まないことから、ラベルによる分類は行わなかった。

初期では、Music, S-Speech, GSI の3つの要素間が強い共起関係にあり、次いで S-Speech と GF, Music と GP, Music と GF の結びつきが強かった。Music と S-Speech は、教師の発話のように分類されていないため、必然的に共起する割合が高くなる。しかし、GF との共起関係も強かったことや、GF と GP それぞれが連続する割合が高かったことから、初期のレッスンは演奏に対する問題提起 (GF) や解決方法の提示 (GP) が中心の指導であることが示唆された。

後期では、GF と他の要素が共起する割合が少なくなった一方で、Music, S-Speech, GSI, GP の共起関係が強くなった。また、Music 間の割合が 13.16% と高く、かつ Music を中心として上記の3要素との強い結びつきがあった。つまり、レッスンが進むにつれて、演奏 (Music) を主軸に細かな指摘 (GP) や印象評価 (GSI) の指導に変化することが明らかとなった。

以上のように、状態遷移の可視化によってレッスンの変遷を捉えることができた。しかし、有向グラフではレッスンの様相を適切に読み取ることが難しく、より直感的な表現方法が必要である。

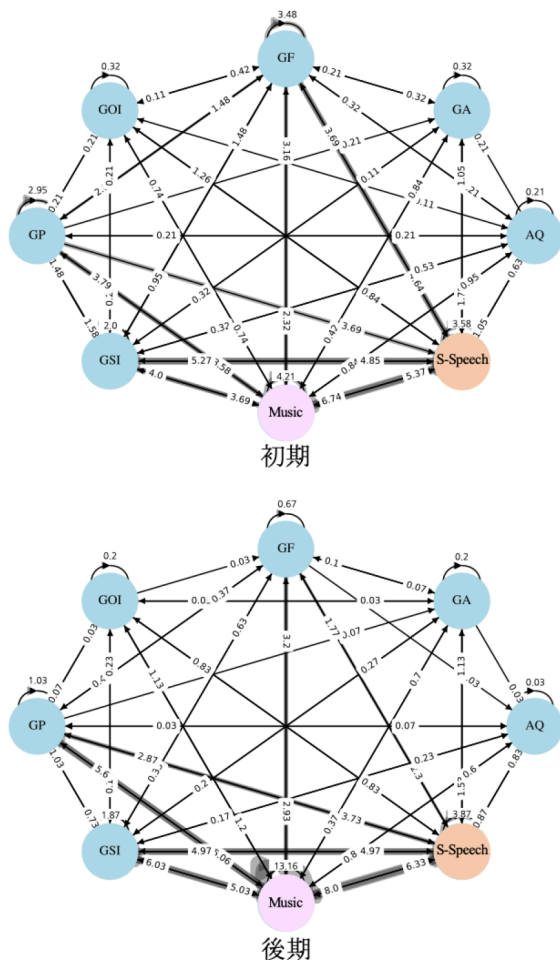


図 2: 教師の発話における ICL の有向グラフ

4 状態遷移結果の構造分析

本節では、前章で述べた指導コンテンツラベル (ICL) の状態遷移結果に対する構造分析をおこなう。我々は、ISM 法 (Interpretive Structural Modeling) による構造化モデリング手法を用いて、ラベル間の因果関係を客観的かつ視覚的に構造化することを試みた。ISM 法とは、グラフ理論の応用として、J.N.Warfield によって提唱された手法の1つであり [8], システムのユーザビリティの問題解決手法や設計仕様の因果関係の分析などで利用されている。要素間の相互関係を一対比較による binary 値の置き換えにより、様々な要素間の相互関係が多階層有向グラフとして図示できる。

4.1 ISM 法の適用手順

ISM 法の適用手順は、(1) 要素の抽出、(2) 隣接行列の作成と行列計算、(3) 構造化である。本手法を用いて、教師の発話における要素間の因果関係や包含関係を可視化する。

(1) 要素の抽出

前章で生成した有向グラフの8つの要素 $S = \{s_1, s_2, \dots, s_8\}$ を対象として、状態遷移を表現する “binary relation” R を定義する。例えば、要素 s_i が要素 s_j に直接関係している場合、 $s_i R s_j$ と表現される。そして、任意の要素 s_i, s_j, s_k の間に、 $s_i R s_j$ and $s_j R s_k \Rightarrow s_i R s_k$ が成立する場合、関係 R は “transitive” である。グラフ理論を用いると、各要素は頂点 (vertex) に対応し、任意の要素間に関係 R が存在する場合は、辺 (edge) でこれを表現できる。ただし、8要素の場合は有効辺の数が多いため、特徴的な関係性の正確な理解が難しい。そのため、本分析では状態遷移の割合に関する閾値を 5.0% と設定し、この閾値を満たす上位項目に対して ISM 法の適用した。

(2) 隣接行列の作成と行列計算

関係 R は、二値行列として表現できる。集合 S の各要素を行と列に持つ行列を考えると、 $s_i R s_j$ が成立していれば、行 i と列 j の交点に “1” を記入し、そうでない場合 “0” を記入する。閾値を設定した有向グラフ D に等価な二値行列 A は次の通りである。

$$A = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \end{bmatrix}$$

この行列 A は、有向グラフ D の随伴行列 (adjacency matrix) と呼ばれる。隣接行列 A のべき乗 $A^2, A^3 \dots A^m$ から直接、間接的な因果関係の全てをあらわす可到達行列 T をブール代数演算則に従い、次式により求める。

$$T = \sum_{k=1}^n A^k + I, A^m = O$$

この時、 O は零行列、 I は単位行列、 m は n 次元の正方向行列で $A^m = O$ を満たす $n-1$ より小さい自然数である。

$$T = (A + I)^2 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1^* & 1^* & 1 & 1^* & 1 & 1 \\ 0 & 0 & 1^* & 1^* & 1^* & 1 & 1 & 1 \\ 0 & 0 & 1 & 1^* & 1 & 1 & 1 & 1 \\ 0 & 0 & 1^* & 1^* & 1^* & 1 & 1 & 1 \end{bmatrix}$$

ここで、*印のついた“1”は、もとの行列 $A + I$ にはなかった要素を示している。例えば、要素 s_3 から s_5 には直接的な関係はないが、要素 s_7 を通じて間接的に関係しているため、その transitive な関係を示している。従って、行列 T には、もとの行列 A における全ての直接的関係と共に、関係 R の推移性によって得られる間接的關係も同時に表現される。

(3) 構造化

行列計算で得られた可到達行列 T の各項目について行和 d と列和 r を求め、関連度 ($d+r$) および影響度 ($d-r$) を算出する。関連度は他項目に与える影響 d と他項目から受ける影響 r の和であるため、当該項目と他項目との関係の深さを示す。一方、影響度は他項目に与える影響 d より他項目から受ける影響 r を差し引いたものであるため、当該項目が他項目に与える影響の強さを示す。これらを 2次元の散布図で示し、クラスター分析により、関連度と影響度に基づく類似性を分類する。最後に隣接行列の関係に従い有向枝で連結する。

4.2 分析結果と考察

前節で詳述した ISM 法の適用により、前章の状態遷移結果に対する多階層有向グラフを獲得した(図3, 4)。図3に示すように、グラフは階層構造を持ち、上位の階層 (Level 3) が他の要素へ影響をより強く与え、下位の階層 (Level 2, 1) が他の要素から影響をより強く受けることを表現している。具体的には、全要素の中で Level 3 の Music, S-Speech, GSI, GP が要素間で

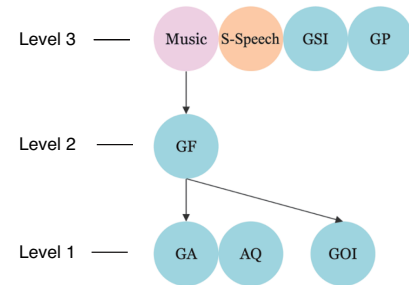


図 3: すべてのレッスンにおける多階層有向グラフ

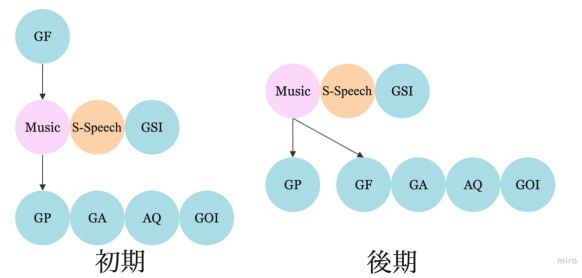


図 4: フェーズごとの多階層有向グラフ

互いに影響を及ぼし合い Level 2 の GF に影響を与え、さらに Level 2 から Level 1 の GA, AQ, GOI に影響を与えるという関係を示している。

図4のフェーズごとの構造を見ると、初期では GF が起点となり、Music、そして GP 他に影響を与えていることがわかった。そして、後期では Music が上位になり、GF は影響を受ける要素に変化した。また、Music, S-Speech, GSI の3つの要素は常に影響を及ぼし合うことが示された。これらは、前章の状態遷移結果から推察した結果と同じである。つまり、ISM 法を適用することによって、有向グラフだけでは不明瞭な要素間の関係を適切かつ直感的に構造化できることがわかった。

上記の結果を受けて、新たに作曲時代とペアによる構造化を試みた結果を図5, 6に示す。作曲時代に関しては、古典派では GSI と GP が上位であったが、ロマン派は教師の発話がすべて下位の階層に分類され、両者ともに Music と S-Speech が他の要素に影響を与えることが明らかになった。ロマン派の楽曲の場合は、より演奏中心のレッスンになることを示唆している。ペアによる比較では、3ペアとも類似しており、T1-S1 と T2-S3 においては同じ構造であった。これらの特徴として、S-Speech あるいは GP が起点となりレッスンが展開されることが示唆されるが、基本的にほとんどの要素が同一階層だったことから、ペアによる特徴的な傾向はないと言える。今後、データ数やバリエーションを増やした分析を実施することで、楽器の指導過程の定式化が期待できる。

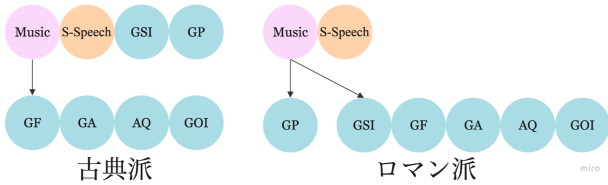


図 5: 作曲時代ごとの多階層有向グラフ

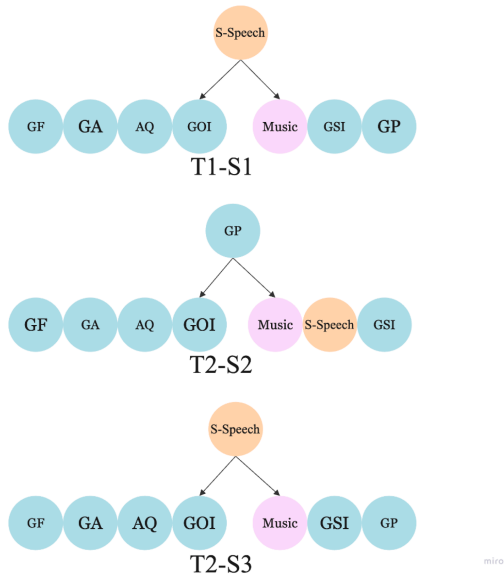


図 6: ペアごとの多階層有向グラフ

5 おわりに

本稿では、複数回にわたる楽器の指導過程の可視化を目的として、クラシックギターの個人レッスンを対象とした音声データの分析結果について報告した。収録音源から音声区間や発話情報を抽出し、さらに教師の発話に対する意味的要素を表現したラベルを付加した結果に基づいてレッスン内の状態遷移を観察したところ、レッスンのフェーズによって指導の様相が異なることが示された。また、構造化モデリング手法を用いた状態遷移結果の構造分析を試みたところ、レッスンの特徴をより直感的に表現できることがわかった。今後は、より詳細な議論を通じて、レッスンの振り返りや指導ノウハウの体系化につなげたい。

謝辞

本研究の一部は、公益財団法人柏森情報科学振興財団 K33 研 XXVI 第 598 号の助成を受けたものである。

参考文献

- [1] 三浦寛也, 飯野なみ: 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, 第 25 回インタラクティブ情報と可視化マイニング研究会 (2020).
- [2] Iino, N., Miura, H., Takeda, H., Hamanaka, M., and Nishimura, T.: Structural Analysis of Utterances during Guitar Instruction, *Proceedings of the 16th International Symposium on Computer Music Multidisciplinary Research*, pp. 692–695 (2023).
- [3] Daniel, R.: Exploring music instrument teaching and learning environments: video analysis as a means of elucidating process and learning outcomes, *Music Education Research*, Vol. 8, No. 2, pp. 191–215 (2006).
- [4] 山本敦, 古山宣洋, ピアノレッスンにおける演奏表現のマルチモーダルな協働的構築, *社会言語科学*, Vol. 23, No. 1, pp. 84–99 (2020).
- [5] 蓮見絵里, ジャズの即興演奏学習場面における演奏者の「音の協働探索」, *日本認知科学会第 33 回大会*, pp. 657–665 (2016).
- [6] 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).
- [7] Matsubara, M., Kagawa, R., Hirano, T., and Tsuji, I.: Analysis of Usefulness of Critique Documents on Musical Performance: Toward better Instructional Document Format, *The 23rd International conference on Asia-Pacific Digital Libraries, ICADL*, pp. 344–353 (2021).
- [8] J.N.Warfield: *Societal Systems; Planning, Policy and Complexity*, John Wiley & Sons Inc (1976).