





図 1: レッスン音声の時系列変化

できる可能性があることが分かった。

表 1: 各音声区間の割合

レッスン	教師の発話	生徒の発話	演奏音
1 回目	23.5%	21.6%	54.9%
2 回目	27.1%	26.3%	46.6%
3 回目	26.5%	12.8%	60.7%
4 回目	26.3%	10.6%	63.1%

### 3.2 指導ラベル

前節の音声区間の結果を踏まえて、教師の発話に対して観点を表す“指導ラベル”を付与した。指導ラベルは、4つの大分類と18つの中分類で構成されている。詳細は[三浦 21]を参照されたい。付与の際は、中分類は2つまで付与できるようにした。なお、図1の教師の発話では大分類の変化も示している。発話区間と指導ラベルの変化から、次の傾向が見えた。(a)レッスンの初期段階では大分類は局所的に変化することが多いが、回数を重ねるごとに、指導ラベル内容の変化は大域的に変化することが多い。図1では大分類Bを中心とした指導に変化している。(b)大分類と中分類の組み合わせにもパターンが存在する。大分類Bに限定した場合、初期段階では速度や奏法が多く、後期になると音色や音楽的なまとまりといったラベルが頻出された。ただし、指導ラベルだけでは教師のどの発言が重要であるか、といった要点を掴むには情報が不十分である。効率的に重要なポイントを把握し、振り返りをしやすくするためには、教師の発話のさらなる詳細な分析が必要である。

## 4. 教師の発話における構造的分析

前章では、教師の各発話における重要度の同定が必要であることが示唆された。我々が目指している本研究の最終成果は、レッスンの要約を達成することである。そこで本章では、教師の発話を階層的に捉えるための第一歩として、GTTMを援用した発話の構造化を行う。

### 4.1 観点による木構造の違い

GTTM [Lerdahl 83]は数学的な演算ができる音楽理論であり、分析することで階層的な時間構造および枝の主従関係を持つ木構造を生成することができる。GTTMの最大の特長は、基本的なルールを保持しながらも分析者の解釈によって木構造が変化することである。また、木構造の上位階層から抽象化されたグループを抽出する、簡約と呼ばれる操作も可能である。つまり、GTTMの利用はレッスンの情報集約と可視化に貢献し得る。

そこで本研究では研究者2名によってそれぞれ構造化を試みた。まず、前章で分析した4回のレッスンのうち1回目のレッスンの教師の発話を取り上げ、前後の音声区間と発話内容に基づいて5つ程度のセクションに分けた。次に、各セクションごとにより詳細なグループ(発話のまとまり)を作り、その中で階層化を行なった。

図2に、発話と木構造の一例を示す。木構造A, Bは分析者A, Bの分析結果に対応している。比較すると、グループに大きな違いは見られなかった。図2では3つのグループに分かれており、大局的な認識には共通性があることが示唆された。一方で、グループ内の階層には違いが見られた。この要因として、分析者Aは要約を強く意識しており、発話内容を時系列的に見る傾向が強かったこと、分析者Bは中分類の数や後半の発話が重要であるという観点で分析していたことが挙げられる。どちらの木構造も“間違いではない”が、汎用的かつ説得力のある木構造を作るためには階層を判断するためのルール作りが求められる。そのためには、我々が定義したラベル以外の要素を考慮することで、有用な指標が見つかる可能性がある。

### 4.2 SOAP ラベルの有用性

自然言語を対象として文章の意味的要素を考慮した分類に、SOAPがある。これは医療現場における医師の思考過程を、自然文を用いてカルテに科学的に記載する枠組みであり、次の4つの項目から成る: Subjective data (S), Objective data (O), Assessment (A), Plan (P)。松原らの研究では、SOAPを音楽用に適用させて、以下のように定義した[Matsubara 21]。本研究ではこれらが指導内容を特徴づけるものであると考え、新たに教師の発話に対して付与した。1つの発話が複数のラベルを持つと判断した場合は、2つ以上付与した。

- S: 演奏の特定箇所の批評のうち根拠が講評者の主観に基づく場合
- O: 演奏の特定箇所の批評のうち根拠が楽譜を参照できる場合または客観的な音楽的知識からなる場合
- A: 演奏全体や演奏者に関する批評(音楽認知や奏法の癖)
- P: あるべき演奏の目標提示や目標に対する練習方法の提示

5つのセクションにおいてラベル付けできた割合を表2に示す。また、図2にはSOAPを付与した結果も示している。注目すべき点は、付与できなかった「該当なし」の発話が約30%もあったことである。例えば、図2の冒頭のようにリズムを

音声	発話内容	SOAP	中分類1	中分類2	大分類
演奏					
教師	うん、このくらいで。ワンツースリー。		速度		B
教師	あそうあのPortamentoで急いじゃうんですね。	S	奏法	速度	B
教師	そう。		奏法	速度	B
教師	ギリギリまでちゃんと溜めましょう。これも。	P	奏法	音価	B
生徒	あそうか。				
教師	あとシーの後速いです。	S	速度		B
教師	しっかりこの音価分溜めてください。	P	音価		B
教師	そう、それだったら間に合うと思うんですね。急ぎすぎなんじゃないかな。これが走って。	S・A	速度		B
生徒	そうかこの辺が、これがこれが走ってこれが走っているからここで、タラグシャーンってなっちゃう。				
教師	バタッと。		速度		B
教師	そう、一个一个ちゃんと、のせていきましょう。	P	速度	音価	B
生徒	なるほどここですかあ、ここ益々ぐちゃぐちゃになってきてる。				
教師	うん、急がないで。弾けるようになると速くなるので気をつけな いとイケないですね。	A	速度		D

図 2: 発話情報と木構造

カウントしているものや、「なるほど」といった短い単語による応答などが挙げられる。つまり、教師の発話の中で明確に“意味を持つ”の発話は限られており、SOAP ラベルが重要度の同定に有用であることが示唆された。さらに階層構造に着目すると、S よりも P が上位に位置するというパターンが多く抽出された。ただし、特定のラベルが連続しているケースもあることから、より多くのパターンを収集する必要がある。

ラベル	個数 (割合)
S	8 (16.7%)
O	9 (18.6%)
A	1 (2.1%)
P	18 (37.5%)
S&A	1 (2.1%)
該当なし	11 (22.9%)

#### 4.3 自動化に向けて

本章で述べた木構造の自動生成のための取り組みとして、生成方式とそのルールを試案している。この木構造は、構造的に重要な発言が幹になるような2分木であり、レッスン音声に含まれる発話やそのメタデータを含む会議記録の情報に基づき、以下2項目のルールに基づいてボトムアップに生成されることが期待できる。(1) レッスンデータに含まれるグループ(ゲシュタルト)を抽出するルール(Grouping Preference Rules: GPR)、(2) あるグループ全体の時間幅を代表する重要な発言を同定するルール(Significance Preference Rules: SPR)。(1)では、発言時間長、発話量などの尺度をもとにルールが構成される。また(2)では、発話の間隔や発話量、重要単語の初出箇所、3.2節で詳述した指導ラベルの変化(中・大分類)、演奏中の発話であるかどうか、4.2節で述べたSOAPラベルの付与情報をもとにした、ルール構成を検討している。

## 5. まとめ

本稿では、音声区間と発話内容を統合的に分析するために、GTTMを援用した教師の発話の構造的分析を試みた。まず対象レッスンの音声的な特徴と、我々が定義した指導ラベルにおける課題を明らかにした。次に、GTTMを援用して生成した

2つの木構造からパターンを抽出し、より精緻な分析に向けた新たなラベルの追加と自動化に向けた課題の整理を行なった。

本研究のような極めて固く複雑なリアルタイム情報を扱うためには、膨大なデータから特徴を見つけるのではなく、個別のデータの精密な分析による共通項への落とし込みが重要であると考えている。それによって、分析結果の解釈や説明可能性の向上に繋がることが期待できる。

## 謝辞

この研究の一部は栢森情報科学振興財団の助成を受けて遂行された。

## 参考文献

- [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, 7, pp.103–122 (2018).
- [Lerdahl 83] Lerdahl, F. and Jackendoff, R.: A Generative Theory of Tonal Music, MIT Press (1983).
- [Matsubara 21] Matsubara, M., Kagawa, R., Hirano, T., and Tsuji, I.: CROCUS: Dataset of Musical Performance Critiques Relationship between Critique Content and Its Utility, AProc. of the 15th International Symposium on CMMR, pp.279–288 (2021).
- [三浦 19] 三浦 寛也, 竹川 佳成, 寺井 あすか, 平田 圭二: 言語・非言語情報に基づく議論構造化と要約生成を繰り返しながら議事録を生成するシステムの実現, 知能と情報(日本知能情報フuzzy学会誌), Vol.31, No.1, pp.1–10 (2019).
- [三浦 21] 三浦寛也, 飯野なみ, 浜中雅俊, 武田英明, 西村拓一: 楽器演奏場面における指導知識のモデル化の試み, 第35回人工知能学会全国大会, 1E4-OS-8b-03 (2021).