管楽器合奏のコヒーレンス度の定式化と個別・全体録音からの推定*

☆尾上 悟嗣 (九州大・工), △ 杉本 悠 (九州大院・システム情報),
△ 金子 仁美 (東京藝大・音楽), 中村 栄太 (九州大院・システム情報)

1 はじめに

吹奏楽では各音符の音高や発音時刻などが揃った合奏の実現が最初の課題であるが，経験の浅い演奏者や指導者にとって合奏の調和度を認識することは容易ではない．従来，管楽器の個別演奏を対象とした連続音高や発音時刻の定量分析手法は数多く研究されているが [1, 2]，合奏録音からその調和度を分析するためには新たな方法の開拓が必要である．本研究では，吹奏楽の合理的な練習を支援するため，合奏の調和度を定量的に分析する手法について調べる．

具体的には，クラリネットのユニゾン合奏における演奏者の個別録音から音符単位で抽出する演奏特徴量を用いて合奏の調和度を表すコヒーレンス度を計算する方法（以下，I2C と呼ぶ）を構築する（Fig. 1）．また，こうして得られる各音符のコヒーレンス度の値を全体録音データから推定するための機械学習手法（以下，W2C と呼ぶ）の構築も行う．さらに，吹奏楽指導者による合奏録音の分析データとの照合により，これらの手法の有効性を検証する．

2 個別録音からのコヒーレンス度の計算

2.1 演奏特徴量の抽出

本研究では，音符ごとの連続音高と発音時刻，ラウドネスに関するコヒーレンス度の計算方法について考える．この準備として，まず各演奏者 $a \in \{1, \ldots, A\}$ の個別録音から各音符 n の基本周波数（F0）$x_a^{\mathrm{F0}}(n)$ と発音時刻 $x_a^{\mathrm{on}}(n)$，ラウドネス $x_a^{\mathrm{LN}}(n)$ の抽出を行う．個別録音には楽器付近に設置する通常のマイクであるピンマイク，または，楽器に接触させてその振動を記録するチューナーマイクが使用できるが，ここでは他者の演奏音の混入が少ない後者のマイクを用いる．チューナーマイク録音は，音響特性が通常のマイクと異なり，楽器のキーノイズの混入もあるため，発音時刻やラウドネスの計算に工夫が必要である．

F0 の抽出では，まず pYIN [3] によりフレームごとにヘルツ単位での F0 を取得し，これをセント単位の値に変換する．音符単位の F0 の値 $x_a^{\mathrm{F0}}(n)$ は，次に述べる MIDI 採譜で得られる各音符の発音・消音時刻により定められる音符区間においてフレーム単位の F0 を集約し，その中央値を求めることで計算する．

発音時刻の抽出には，演奏音源から各音符の（半音単位の）音高と発音・消音時刻を推定する MIDI 採譜手法を用いる．具体的には，深層学習モデルである CRNN に基づく自動採譜手法 [4] を用いる．この手法の学習にはチューナーマイク録音とそれに同期した MIDI データが必要であるが，この同期 MIDI データを作成することは容易ではない．そこで，ピンマイク録音とチューナーマイク録音の同期録音データを収集し，Fig. 2 に示す適応学習の方法により採譜モデルを構築する．約 12 時間の同期録音データを用いて

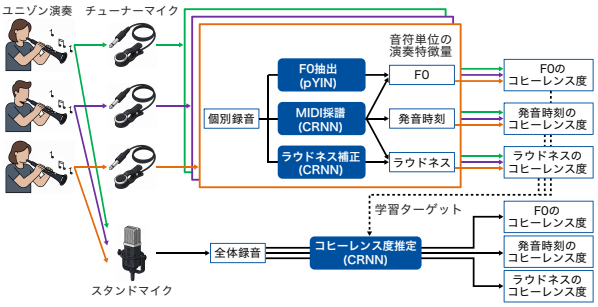

Fig. 1 Flowchart of coherence calculation.

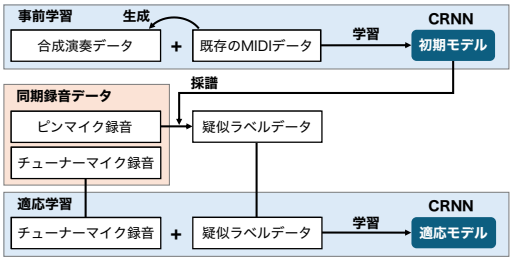

Fig. 2 Adaptive training for MIDI transcription.

学習した採譜モデルによる MIDI 採譜の評価結果を Table 1 に示す．ただし，COn，COnP，COnPOff は発音時刻のみ，発音時刻と音高，発音時刻と音高と消音時刻の一致を見る F 値である．本手法では，汎用的採譜手法 [5] や合成データを用いて学習した初期モデルに比較して精度が大幅に向上し，高精度の発音時刻推定が可能であることが示された．

ラウドネスの抽出では，パワースペクトログラムに対して A 特性周波数重み付けを行って得られるフレーム単位のラウドネス値を用いて，F0 と同様に音符区間における中央値を求める方法が考えられる．ここでもチューナーマイク録音の音響特性により，通常の計算で得られるラウドネス値は，音高や音色によって実際の値から非線形に変化する．そこで，上記の同期録音データを用いて，ピンマイク録音のラウドネス値をチューナーマイク録音のメルスペクトログラムから推定する手法を上述の CRNN を用いて構築する．本手法によるフレーム単位のラウドネスの平均推定誤差は 3.07 dB であることが評価実験で示された．

2.2 コヒーレンス度の定式化

前節で得られた各音符 n に関する演奏特徴量 $x_a^{\phi}(n)$ $\phi \in \{\mathrm{F0}, \mathrm{on}, \mathrm{LN}\}$ から，2 奏者間 a, b $(a \neq b)$ の差分 $\Delta_{ab}^{\phi}(n) = |x_a^{\phi}(n) - x_b^{\phi}(n)|$ を計算する．これらの差分から音符 n における集約化された差分を $\Delta_{\phi}(n) = g_{\phi}(\{\Delta_{ab}^{\phi}(n)\}) \in \mathbb{R}_{\geq 0}$ のように計算する．ここで g_{ϕ} は集約化を表す関数であり，具体的には平均（mean）や二乗平均平方根（RMS），最大値（max）などが考えられる．この $\Delta_{\phi}(n)$ を用いて，音符 n における合

*Formulation and estimation of coherences of wind ensemble performance, by Satoshi Onoue, Yu Sugimoto (Kyushu Univ.), Hitomi Kaneko (Tokyo Univ. of the Arts), Eita Nakamura (Kyushu Univ.)

採譜手法	COn	COnP	COnPOff
CREPE Notes [5]	80.1%	79.5%	61.6%
初期モデル	74.2%	72.0%	61.0%
適応モデル	**97.8%**	**97.8%**	**94.3%**

Table 1　Evaluation results of MIDI transcription for individual recordings.

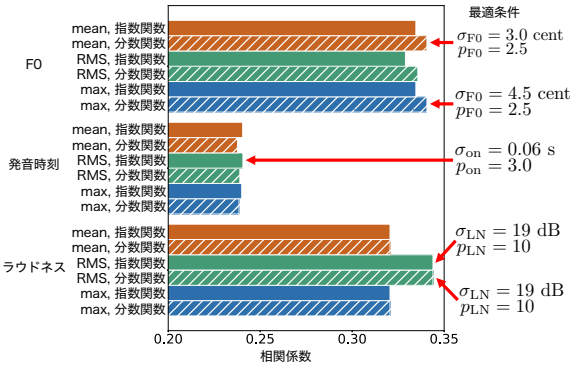

Fig. 3　Comparison of correlations between coherence values and expert annotations.

奏のコヒーレンス度を $C_\phi(n) = h_\phi(\Delta_\phi(n)) \in [0,1]$ のように定義する．ここで，h_ϕ は $\Delta_\phi(n)$ をコヒーレンス度の値域である $[0,1]$ に写す単調減少関数であり，具体的には以下の 2 つの関数形を考える．

$$h_\phi(\Delta) = \begin{cases} \exp[-(\Delta/\sigma_\phi)^{p_\phi}] & \text{（指数関数）} \\ 1/[(\Delta/\sigma_\phi)^{p_\phi} + 1] & \text{（分数関数）} \end{cases} \quad (1)$$

パラメーター σ_ϕ は差分のスケール，p_ϕ は差分に対する感度を表す．

本手法（I2C）のコヒーレンス度の定義にある関数 g_ϕ, h_ϕ の具体形およびパラメーター σ_ϕ, p_ϕ の値を求める方法として，専門家による演奏分析データとの照合により，人間の主観的な調和度を最もよく再現するものを選択する手法が考えられる．この方法を実行するため，3 名の演奏者によるクラリネット合奏を収録し，その内の 6 曲に対して吹奏楽指導者に演奏の改善点の分析を音符単位で行ってもらった（ただし，有効なデータが得られなかったラウドネスに関しては第一著者が自ら分析を行った）．この結果は各音符 n と演奏要素 ϕ に対する二値ラベル $B_\phi(n) \in \{0,1\}$（0 = 改善が必要，1 = 良好）として表される．

$C_\phi(n)$ と $B_\phi(n)$ の相関係数を最適化の尺度として関数 g_ϕ, h_ϕ の具体形を探索した結果を Fig. 3 に示す．F0 と発音時刻では，最適な σ_ϕ はこれらの知覚量の弁別閾の数倍程度であり，妥当な結果と言える．一方，最適な相関係数は比較的低い値になっており，人間の演奏分析では特徴量の差分の知覚よりも複雑な基準が用いられている可能性がある．ラウドネスではコヒーレンス度が全て 1 となる場合が最適となったため，最適化の方法を改善する必要がある．

3　全体録音からのコヒーレンス度の推定

3.1　推定手法

実際の練習現場で個別録音よりも容易に収録できる全体録音（混合音のモノラル録音）から直接コヒーレンス度を推定する手法（W2C）を深層学習を用いて構築する．入力には全体録音の各音符区間のメルスペクトログラムを用い，各演奏特徴量の特性に応じて，F0 とラウドネスでは発音時刻から消音時刻までの区間（1 s 以上の場合は先頭の 1 s の区間），発音時刻では発音時刻の前後 0.2 s の区間のデータを切り取る．出力は各特徴量のコヒーレンス度で，学習のターゲットには個別録音から計算したコヒーレンス度の値を用いる．この際，関数形は Fig. 3 の最適条件を用いる（ただし，ラウドネスでは g_ϕ = RMS, h_ϕ = 指数関数, $\sigma_\phi = 3$ dB, $p_\phi = 1$ とした）．深層学習モデルには上述の CRNN を用い，学習には合計約 25 分間の全体録音データを用いた．

比較データ	F0	発音時刻	ラウドネス
W2C-I2C	0.455	0.467	0.076
W2C-人手分析	0.276	0.128	0.098
I2C-人手分析	0.341	0.241	0.098

Table 2　Evaluation results of coherence estimation.

3.2　評価実験

W2C によるコヒーレンス度の推定値と I2C による値および人手による演奏分析結果（2.2 節）との相関係数を測定した（Table 2）．W2C と I2C の比較では，F0 と発音時刻は 0.46 程度の相関係数が得られており，W2C の手法の有効性がある程度確認できる．より多くの学習データを用いることで，W2C の誤差は低減できる可能性があると考えられる．一方，ラウドネスの相関係数は小さく，メルスペクトログラムからラウドネスのコヒーレンス度を推定することは難しいことが示唆される．W2C による推定値と人間の演奏分析データとの間の相関係数は，W2C の誤差の影響を反映した結果になっている．

4　まとめ

本研究では，管楽器のユニゾン合奏のコヒーレンス度を個別録音および全体録音から計算する方法を構築した．I2C ではチューナーマイク録音から各演奏特徴量を高精度で抽出する手法と，それらの差分に基づくコヒーレンス度の計算法を実現した．専門家の演奏分析との照合では，F0 と発音時刻に関して中程度の相関係数が得られた．今後の課題として，W2C の手法の改良，非同期の個別録音からのコヒーレンス度の計算手法の構築，音符内での連続音高やラウドネスの変化を考慮した手法の構築などを考えている．また，提案手法は多くの木管・金管楽器に適用可能であり，複数パートの合奏への拡張にも取り組みたい．

謝辞　本研究は JST FOREST No. JPMJPR226X 及び科研費 Nos. 25H01148, 25H01169 の支援を受けた．クラリネット演奏収録と合奏分析の協力者に感謝する．

参考文献

[1] G. Bandiera et al.: Proc. ISMIR, 414–419, 2016.
[2] 山田昌尚ほか: 情報科学技術フォーラム, RE-001, 9–12, 2016.
[3] M. Mauch et al.: Proc. ICASSP, 659–663, 2014.
[4] Y. Sugimoto et al.: Proc. APSIPA ASC, 305–310, 2025.
[5] X. Riley et al.: Proc. SMC, 1–5, 2023.