

演奏者と「聴衆」間の 聴覚フィードバックの有無による手拍子同期の比較

水野 伸子¹ 津崎 実²

概要: 演奏者と「聴衆」間の同期に対する聴覚フィードバックの有無による影響を検討した。参加者は2部屋に分かれ、別室からリアルタイムに配信されるピアノ演奏を聞きながら拍に合わせて手拍子をした。手拍子音が演奏者へフィードバックされ聴覚情報が双方向に伝達される条件と、演奏配信のみの一方方向条件の2群間で比較をした。拍時点は音響信号から信号処理をして推定した。演奏の拍時点と手拍子の誤差の比較および拍周期の相互相関解析の結果、聴覚フィードバックを有する方が演奏の拍時点と手拍子の同期度が高く、直前1小節以内の情報を互いに利用し相互に調整し合う様子が認められ、同期に対する聴覚フィードバックの優位性が示された。

キーワード: 同期, 聴覚フィードバック, 演奏者と聴衆, 手拍子, 拍

Comparison of Beat Synchronization among a Performer and Audience with or without Auditory Feedback

NOBUKO MIZUNO^{†1} MINORU TSUZAKI^{†2}

Abstract: We investigated influences of auditory feedback on synchronization between the beats of piano performances and those of hand claps under two situations. (a) where the bidirectional, auditory communication channel was available, or (b) where the unidirectional, auditory information was available. Participants were required to clap their hands in time to piano performances under both conditions. While the pianist could hear the hand claps of the participants in the (a) condition, she could not in the (b) conditions. The onset time of each beat was estimated by a signal processing to the digital recordings of the piano sounds and those of clap sounds. The results suggested that the bidirectional, auditory information transmission enhances the degree of synchronization between the beat timings of the performer and those of the audience. The results of cross-correlation analysis of inter-beat interval of the performance and clapping suggest that the performer and the audience are mutually synchronized while using the information up to the past one bar.

Keywords: Synchronization, Auditory feedback, Performer and audience, Handclap, Beat.

1. はじめに

現代社会ではテレビなどのメディアから、YouTube による動画共有プラットフォームや BD などの高音質の電子媒体まで多様な方法で音楽を聴くことができる。再生する機器やインターネットへの接続環境さえ整えば、いつでもどこでも気軽に音楽を楽しむことができる。しかし、再生環境が充実していながらも、人は演奏者と同じ時間と空間を共有し直接に演奏を聴く生演奏を求めてコンサートやライブ会場へ足を運ぶ。生演奏と録音（録画）再生演奏との違いはどこにあるのか。生演奏を特徴づける要因の一つは、演奏者と鑑賞者とが互いに同じ時間や空間を共有し、両者のインタラクションが可能な状況にあることである[1]。先行研究では生演奏と録音演奏との差異を演奏者と「聴衆」の相互作用に見出し[2][3]、特に同期や同調の観点から検討されてきた[4][5]。合奏者間の同期に注目した先行研究[6]では「引き込み（entrainment）」と表現し知覚と運動を統合する時空間

調整であると説明した。この引き込みの方向性を決定付けるものが感覚フィードバックである。例えば、2者間の合奏における「一方的な引き込み」は一方からもう一方への伝達のみで出力側へのフィードバックはない。しかし「相互引き込み」（2奏者間）や「集団的引き込み」（3奏者以上）は出力された音響信号が他者の制御された音響信号となって元の出力者へフィードバックされ、自身の運動の制御を可能とする。先行研究[7][8]では、情報伝達が一方と双方向及び自身の音のみの3種類に条件統制して2者間の合奏やタッピングによる同期を調べ、どちらも聴覚フィードバックを有する双方向条件の方が、それ以外の条件より同期の度合いが高かったことを示した。これらの研究は、個人対個人の実験から検討された。コンサートで見られるような演奏者と「聴衆」という個人対集団の同期においても、聴覚フィードバックは影響すると予想される。フィードバックを伴う同期は、生演奏と録音（録画）再生演奏との差異につながると考えられた。

そこで、本研究では演奏者と「聴衆」間の同期に対する聴覚フィードバックの有無による影響を演奏の拍時点と参加者集団の手拍子との同期度から比較検討する。演奏に合わ

1 愛知東邦大学
†1 Aichi Toho University
2 京都市立芸術大学
†2 Kyoto City University of Arts

せた手拍子は音楽聴取を基に行われることから聴取内容を反映した行為であると判断し「聴衆」を参照することとする。なお、本論では拍の時刻を時間の流れの中のある一点という意味を明確に表現するために拍時点と表記する。

2. 実験方法

2.1 実験参加者, 実験環境, 手順

実験参加者は大学生 68 名である ($M=19.79$ 歳, $SD=0.28$, 男性 25 名, 女性 43 名)。楽器教習経験から専門的な音楽のトレーニングは認められず、リズム感覚に対して一般的な集団であるとみなすことができる。使用した音楽は計 3 曲 (W1, 2, 3), W1 はシュトラウス I 世作曲《ラデツキー行進曲》(4/4 拍子, 40 小節), W2 はモーツァルト作曲《きらきら星変奏曲ハ長調 (主題)》(2/4 拍子, 24 小節), W3 は三木たかし作曲《アンパンマンのマーチ》(4/4 拍子, 20 小節) である。実験後に DAW ソフト (Cubase Pro 10) が検出した W1, W2, W3 における計 4 セッションの平均テンポは 157 (W1), 124 (W2), 101 (W3) bpm であった。曲途中の意図的な速度変化 (加速や減速) は行っていない。

Figure 1 は実験環境の概要図である。3 室を準備し、参加者集団を半数ずつ Room 1, 3 の 2 部屋に分け、演奏者は Room 2 で電子ピアノ演奏を行った。演奏された音響信号は演奏者のヘッドフォンおよび Room 1, 3 のスピーカへ送られた。補助的に、参加者がコンサートで手拍子するイメージを持ち手拍子する行為を不自然に感じないように、ビデオカメラで撮影した演奏者の映像を Room 1, 3 のスクリーンに投影した。ただし、ビデオカメラで撮影した映像の HDMI 伝送によるディスプレイ表示は 66~100 ms の遅延が想定される[9]ため、映像はあくまで演奏会の雰囲気伝えるもので手拍子する手がかりとして用いないよう参加者へ注意を促し「コンサートで聴衆が音楽に合わせて手拍子をするように、これから演奏する音楽を聞きながら拍に合わせて手拍子してください」と教示した。参加者はリアルタイムに配信された演奏を聞きながら手拍子を行い、Room 1 の手拍子音のみ演奏者のヘッドフォンへフィードバックした。つまり、演奏者と Room 1 の参加者の音は双方向に伝わり聴覚フィードバックを有する。一方、Room 3 は演奏配信のみの方向で手拍子音の演奏者へのフィードバックはない。演奏および手拍子のデータは DAW ソフトを用いてパソコンに記録した。

実験はカウンターバランスをとり、双方向と一方向条件を入れ替えた計 2 セッションを実施した。同一内容の実験を別の参加者でも実施し計 2 実験行った。実験デザインは 2 (双方向・一方向条件) × 3 (曲数) × 2 (実験数) である。

2.2 特徴量

解析に用いた特徴量は、演奏と手拍子の拍時点の差分 (誤差) と演奏と手拍子それぞれの拍周期 (IOI: the inter onset interval) である。拍周期とはある拍時点の時刻から一つ前の

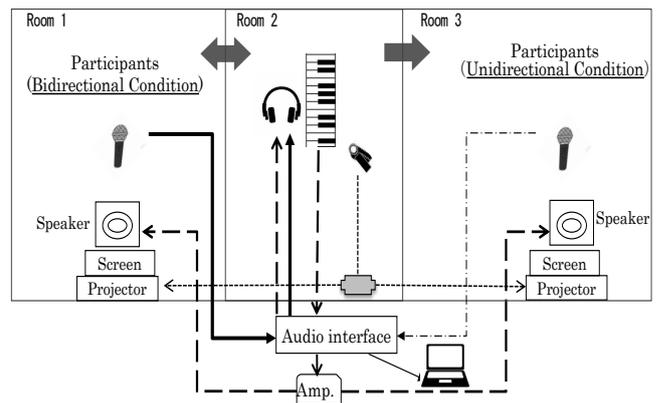


図 1 実験環境図

Figure 1 Diagram of the experimental environment.

拍時点の時刻を差し引いた 1 拍に要した時間を表す。

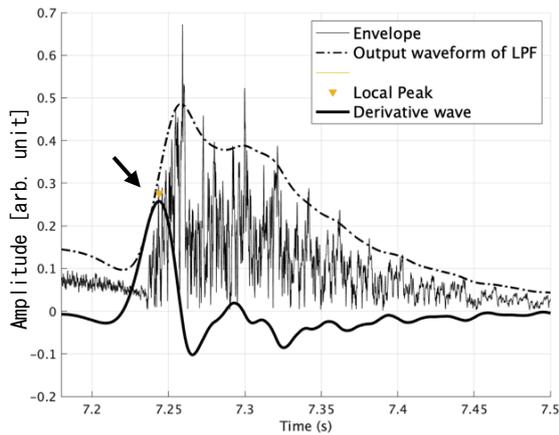
誤差からは演奏と手拍子の時間的な同期の度合い (同期度) を、拍周期 IOI の相互相関解析からは演奏者と参加者の関係、つまり、手拍子と演奏のどちらがどちらの情報を利用して合わせようとしているのかという二つの時系列信号間の関係を調べることができる。

2.3 拍時点の推定

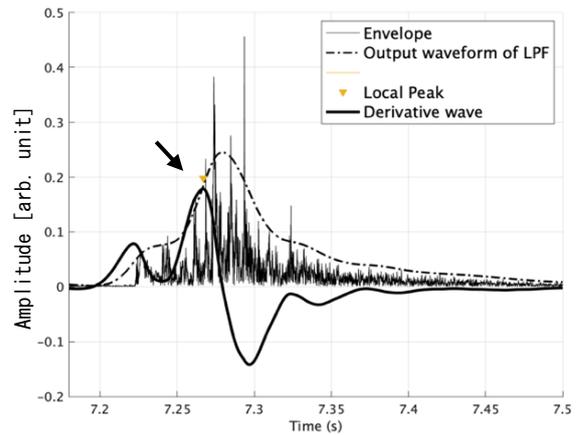
先行研究ではピアノによる演奏の拍時点は MIDI から取得し、複数の同時音に対しては最も早い音や最高音の打鍵時刻とする場合が多い[10][11]。一方、音響信号処理を用いた拍時点推定の手法も多く提案されている[12][13]。本実験の参加者は演奏者の音響情報が手がかかりに反応し、演奏の拍時点と手拍子の同期を分析対象とする。したがって、両方の拍時点は同一の方法で推定することが妥当であると考えられ、音響解析から拍時点推定した。処理は MATLAB R2020b で行った。

ピアノの音響信号から拍時点を決めるプロセスを W1 の D3, F3, G3 と F4 を打鍵した合成音からなる 1 拍の音波形を例に以下に説明する。Figure 2a はヒルベルト変換後の包絡線 (細線), カットオフ周波数 20 Hz によるゼロ位ローパスフィルタ出力波形 (一点鎖線), 時間微分波形 (太線) を任意単位で比較表示した。横軸は時間, 縦軸は振幅である。拍時点の音楽が全て単音の場合, ローパスフィルタ出力波形の音量最大時 (ローカルピーク) は拍時点推定の手掛かりになる。しかし, 本実験曲は拍時点に複数の音が演奏され, これらの合成音による波形の概形や時定数の影響を受けたローカルピークは検出位置にばらつきがみられた。人が音を知覚するタイミングは音が立ち上がった時, つまり急激に音量が増した時であると仮定される。したがって, 音量急増時を検出するために, ローパスフィルタ出力波形を時間微分しその微分波形が最大になる時点, つまり微分係数の最大値の横軸の時刻を演奏の拍時点とみなした。

手拍子の拍時点もピアノ音と同じパラメータで推定した。Figure 2b は 1 拍の集団手拍子音の合成波形を基に Figure 2a



(a) Envelope, output waveform of LPF (low pass filter), its derivative wave and local peak, based on a waveform of piano signal.



(b) Envelope, output waveform of LPF (low pass filter), its derivative wave and local peak, based on a waveform of handclap signal.

図 2 ピアノ演奏と参加者集団の手拍子における拍時点推定のプロセス

Figure 2 Onset time detection of piano and handclap signal.

と同様の処理過程を表した。Figure 2a, b の矢印位置から時間微分波形のピークは音量の急増時点を捉えられていることが確認できる。日常場面で使われる手がかりの多くは音響情報である。音響データを基にした拍時点推定は、この種の実験の多様な演奏場面への拡張が期待される。

3. 結果

3.1 演奏拍時点と手拍子の正規化誤差の比較

本論で使用する「誤差」はピアノ演奏の拍時点と手拍子の拍時点の差分を意味し、分析には BPM の違いを考慮して正規化した誤差 (ND 値) を用いた。i 番目のデータの手拍子の拍時点を a_i (ms), ピアノ演奏の拍時点 f_i (ms), BPM から計算される 1 拍あたりの時間を b (ms) と置くと計算式は次のようになる(1)。

$$ND = \frac{(a_i - f_i)}{b} \quad (1)$$

ほぼ全員の手拍子が出揃ってからの拍を分析対象とし、演奏開始後 4 拍を分析対象から外した。Figure 4a, b, c は式 (1) で計算した正規化誤差の分布を曲ごとに表したものである。全 4 セッションを合わせたデータから曲ごとの大まかな傾向を把握した。演奏者と参加者間の聴覚情報の伝達が双方向の場合 (上図: Bidirectional) と演奏配信のみの一方の場合 (下図: Unidirectional) とに分けて表示した。ヒストグラム横軸上の「0」が演奏拍時点であり、正の符号値は手拍子が演奏拍時点より遅く叩かれたことを、負の符号値は手拍子が演奏より早く叩かれたことを意味する。

3 曲ともに双方向条件の方が一方条件より分散が小さく、演奏の拍時点に集中した。W2 の一方条件は演奏より

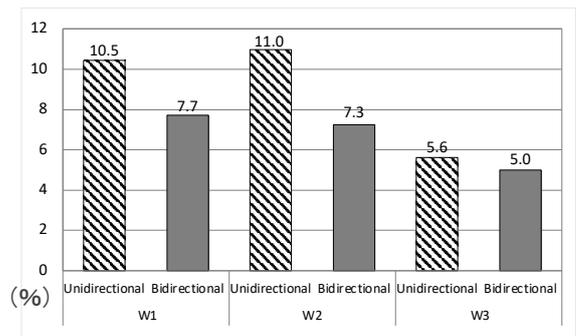


図 3 演奏の拍時点と手拍子の平均平方二乗誤差率 (RMSPE)

Figure 3 Root mean square percentage error (%), shown by the experimental conditions for auditory feedback.

早く叩く傾向が見られ、W3 は他の曲に比べ両条件ともに手拍子が演奏の拍時点に集中する傾向が認められた。

RMSPE (Root Mean Square Percentage Error) を計算し比較した。手拍子と演奏拍時点のずれを正負の符号に相殺されずに手拍子の演奏に対する精度が評価される。n 個のデータにおける i 番目のデータの手拍子の拍時点 a_i (ms) を、演奏拍時点 f_i (ms), bpm から計算される 1 拍あたりの時間 b (ms) と置く。式(2)で計算した結果を Figure 3 に表示した。

$$RMSPE = \sqrt{\frac{1}{n} \sum_{i=1}^n \left(\frac{a_i - f_i}{b} \right)^2} \times 100 \quad (2)$$

Figure 3 からわかるように RMSPE は 3 曲とも双方向条件が一方条件より小さく、双方向条件は一方条件より演奏の拍時点と手拍子の同期度が高いことが示された。W3 は

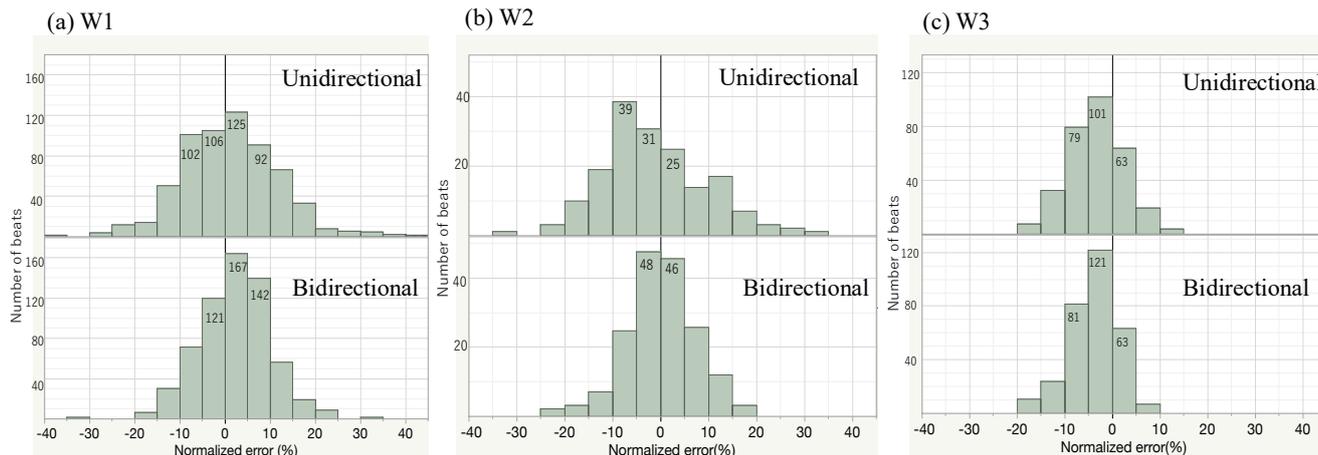


図 4 演奏拍時点と手拍子の正規化誤差の分布

Figure 4 Normalized error distribution between the beat of the piano performance and handclaps.

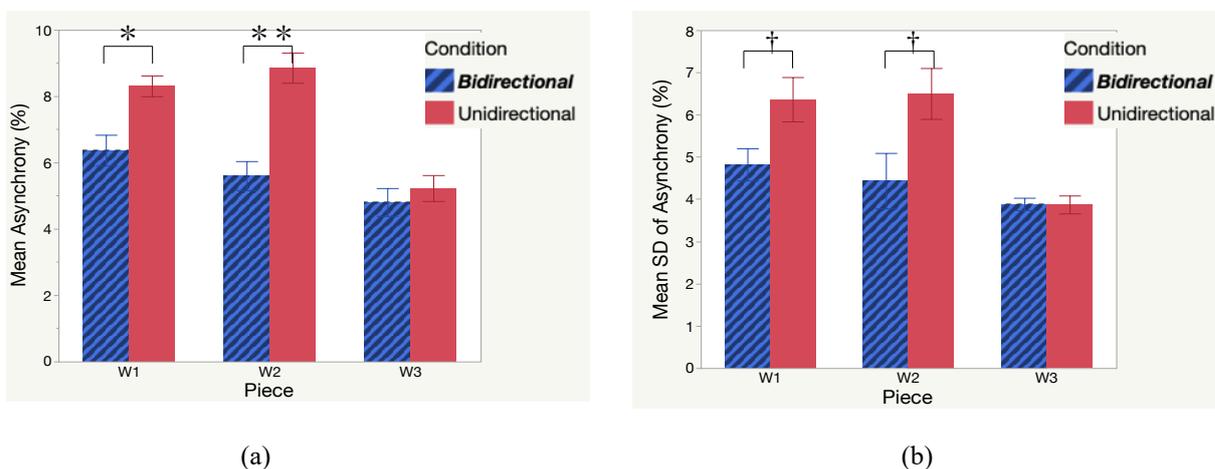


図 5 セッションごとの二乗平方根正規化誤差の平均値(a)と標準偏差(b)の聴覚フィードバックによる比較

Figure 5 Mean asynchronies (a) and mean standard deviation of the Asynchronies (b) in percentage, shown by pieces and auditory feedback as the factors. The Asynchronies indicate root square of normalized error. Error bars denote the standard error of the means. The contrasts between the bidirectional and unidirectional conditions with the asterisks reached the significant level ($p < 0.01(**)$), $p < 0.05(*)$; those with the dagger (\dagger) tended to be significant ($p < 0.1$).

一方方向条件の同期度が高く、条件間の差が小さかった。

3.2 曲要因と聴覚フィードバック要因による二元配置分散分析

演奏と手拍子の拍時点の正負の差分を全てずれと判断し、正規化誤差の二乗平方根から平均値の比較をした。セッションごとに算出した二乗平方根正規化誤差の平均値と標準偏差からなる一つの標本を作り、演奏の拍時点と手拍子の同期度と散布度を曲要因 (W1, W2, W3) と聴覚フィードバック要因 (双方向・一方方向条件) の 2 元配置分散分析から検定した。Figure 5 は平均値 (a) と標準偏差 (b) の聴覚フィードバック要因による比較を曲ごとに表したものである。実験要因の効果を確認するために各指標について 3 (W1,

W2, W3, 参加者内要因) \times 2 (双方向条件 vs. 一方方向条件, 参加者間要因) の分散分析の結果、平均値と標準偏差の両指標とも各主効果は 1%水準で有意であり、効果量 η^2 は全て大きかった (Piece: $F(5, 18) = 19.79, p < .0001, \eta^2 = .40(M)$, $F(5, 18) = 8.55, p = .003, \eta^2 = .34(SD)$, Condition: $F(5, 18) = 30.28, p < .0001, \eta^2 = .30(M)$, $F(5, 18) = 10.08, p = .005, \eta^2 = .20(SD)$). 交互作用は平均値が 5%水準で有意であり、標準偏差は有意傾向、効果量 η^2 はともに中程度であった (Mean: $F(5, 18) = 5.79, p = .011, \eta^2 = .12(M)$, SD: $F(5, 18) = 2.73, p = .092, \eta^2 = .11(SD)$). 双方向と一方方向条件の対比について曲ごとに t 検定および両側検定を行った結果、W1 と W2 において平均値では双方向条件が有意に小さく、標準偏差では有意傾向

があり、W3 はどちらも有意ではなかった (Figure 5 に表示).
セッションごとの二乗平方根正規化誤差の平均値と標準偏差を基に分析した以上の結果から、W1 と W2 において演奏の拍時点と手拍子の二乗平方根正規化誤差の同期度と散布度は双方向条件の方が一方向条件より小さく、W3 の差はほとんどなかったことが示された.

3.3 演奏と手拍子の拍周期 IOI による相互相関解析

聴覚の情報伝達が双方向の場合、演奏者は手拍子音が聞こえることから恣意的に手拍子に合わせたという可能性も客観的には排除できない. 恣意的か否かにかかわらず、手拍子に合わせることができたか否かという演奏者側の作用のみが今回得られた同期度の高さにつながる可能性も考えられた. したがって次の方法で両者の関係を検討した. 演奏と手拍子それぞれの拍周期 IOI からなる時系列を相互相関解析し、時間的關係性からリーダー/フォロワー関係[8]を推定した. 時間の先行する方がリーダー、後から模倣する方はフォロワーになる. Lag の数値はフォロワーが何拍前のリーダーの拍を模倣しているのかを示す. IOI は一つ前の拍時点との時間間隔を 1 拍あたりの平均時間で除して正規化した.

Figure 6 は前後 4 拍までの相関解析結果を曲ごとに 4 セッションの平均値から表した. エラーバーは標準誤差を表す.

双方向条件において以下の傾向が認められた. W1 は弱い正の相関が全体的に見られ、演奏者と参加者が直前 4 拍までの情報を互いに利用して追従した. W2 は Lag -1 で正の相関が見られ、演奏者が 1 拍前の手拍子の情報を利用して追従した. W3 は Lag が -3 から 3 に亘って弱い正の相関が見られ、相関係数は lag の符号が正である方が大きく、参加者の追従度の方が演奏者よりやや高かったことを表した.

以上の結果から、演奏者と参加者のリーダー/フォロワー関係は同一曲のセッション間に差異はあるものの、曲によってその関係は異なることが観察され、一概に演奏者が手拍子に合わせたという一方的な傾向は認められなかった.

4. 考察

演奏の拍時点と手拍子の正規化誤差は、聴覚情報の伝達が双方向の方が一方向より小さかった. このような同期に対する聴覚フィードバックの優位性を先行研究[7][8]では 2 重奏やタッピングの相手との 2 者間で明らかにしている. 本研究は演奏者の演奏に集団である「聴衆」が手拍子を合わせるというコンサート等で自然発生するような状況に近い条件で、これらの先行研究で見出したことと基本的には同様のことが生じることを実験的に確かめた.

演奏の拍時点と手拍子の同期の程度は曲によって異なっていた. 演奏者が聞く参加者の手拍子は 3 曲ともに拍時点で叩かれる四分音符間隔の音である. しかし、参加者は曲ごとに異なる音楽情報を受け取る (Figure 7). 特に打拍間隔内の分割したリズム情報は次の拍時点を予測する有力な手掛かりになる. 1 拍の中に 1/2 や 1/4 に分割する音価を持つ拍数の

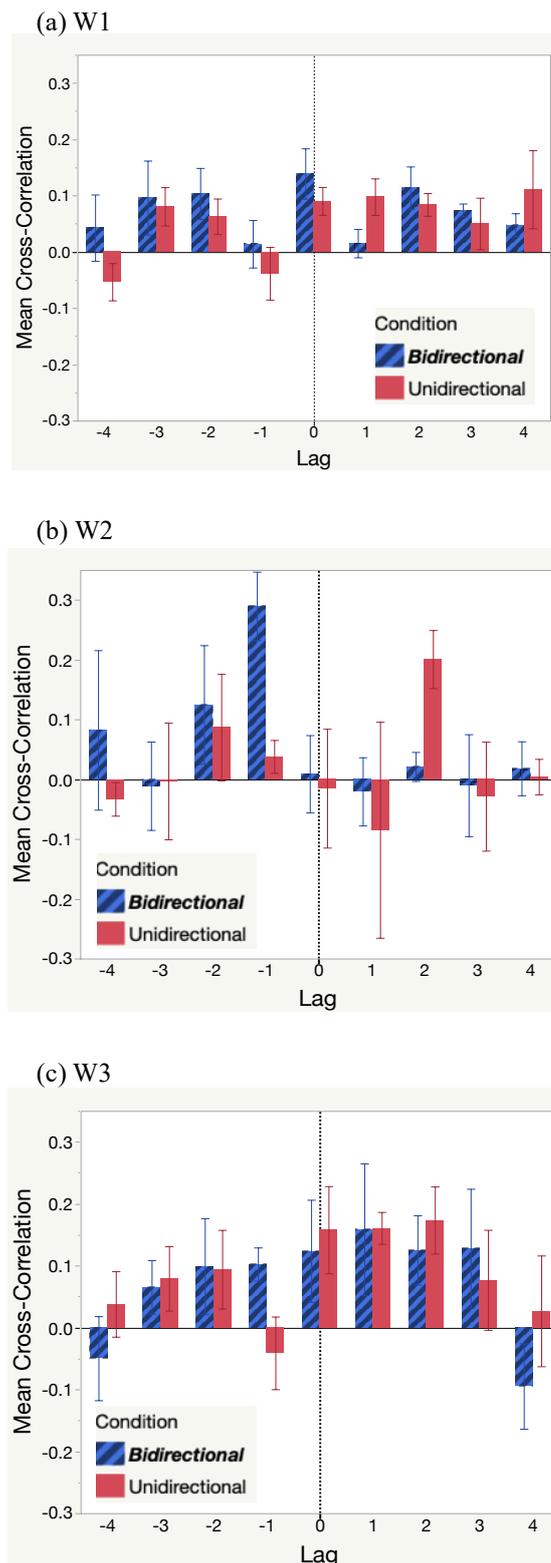


図 6 演奏の拍時点と手拍子のセッションごとの IOI の相関係数

Figure 6 Mean cross-correlation coefficients of the interonset interval timing for handclaps and Piano performance beat by auditory feedback and lag direction. Error bars denote the standard error of the means.



図7 演奏曲の譜例の一部（開始16拍）

Figure 7 The first sixteen beats of W1, W2, and W3.

一曲全体の拍数に占める割合は、W3が77%と最も高く、次いでW1の26%、W2の6%であった。合奏する2者間の音符比率と聴覚フィードバックを条件統制した先行研究[7]では、リーダー対フォロワーの音符比率が2:1条件で聴覚情報の伝達が一方向条件の場合に両者のずれが小さかったことを報告しており、本研究のW3の結果と一致する。1拍の音価の単純分割となる音の存在は、次の拍時点の予測性を高めることが示唆された。

先行研究[14]は、合奏の演奏制御に関するモデルの出力が現実的な演奏の特徴を持つためには、直前1小節程度のタイムスパンでの相手の情報の利用が必要であることを報告した。本研究の演奏者と「聴衆」の関係においても相互相関解析の結果から同様の傾向が確認された。

本実験では演奏者が演奏を発信し（リーダー）、参加者が手拍子で合わせる（フォロワー）という関係に、双方向と一方向の条件間の差はない。しかし、双方向条件に認められた互いの拍を利用して相互に調整し合う関係から、演奏者も聴覚のフィードバックが効いて自分の演奏タイミングを参加者のテンポに合わせてようとするフォロワー役割を、ある程度併せ持つようになることが示唆された。先行研究[6]から、合奏する2者間の聴覚情報の伝達が双方向の場合には、互いに相手のリズム情報がフィードバック信号となって自己の運動を調整し合う「相互引き込み」が生じることが示唆された。本研究の演奏者と「聴衆」という個対集団の関係においても、相互相関解析の結果から同様の「相互引き込み」が生じたと考えられた。

聴覚フィードバックがあるということは、出力した聴覚信号が他者の制御された聴覚信号となって元の出力者へフィードバックされ、自身の運動の制御を可能にする。したがって、聴覚フィードバックの伴う同期は互いの運動制御による相互のタイミング調整が行われることが示唆された。

5. おわりに

本研究では演奏者と「聴衆」間のコミュニケーションチャネルを聴覚に限定し、同期に対する聴覚フィードバックの有無による影響を演奏の拍時点と参加者集団の手拍子との同期度から比較した。演奏の拍時点と手拍子の誤差の比較および拍周期 IOI の相互相関解析の結果から、聴覚情報が双方向に伝達される環境の方が演奏配信のみの一方向より演奏の拍時点と手拍子の同期度が高く、直前1小節以内の情報を互いに利用し相互に拍のタイミングを調整し合う様子が認められた。これにより、演奏者と「聴衆」間の同期に対する聴覚フィードバックの優位性が示唆された。

謝辞 実験にご協力いただいた皆様に、謹んで感謝の意を表す。本研究はJSPS 科研費 19K02769 の助成を受けた。

参考文献

- [1] 正田 遥, 阪田真己子, William, A. 生演奏による聴取がヴァイオリン演奏の評価に及ぼす影響: 全体評定と連続評定. 音楽知覚認知研究. 2017, vol. 23, no. 1, p. 35-55.
- [2] Miell, D., MacDonald, R., & Hargreaves, D. J. How do people communicate using music. *Musical communication*. 2005, p. 1-21.
- [3] 河瀬論. 音楽演奏における演奏者と観客の視覚的コミュニケーション. 認知科学. 2015, vol. 22, no. 4, p. 638-648.
- [4] 山本知仁, 藤井倫雅, 三宅美博. 手拍子を付加した場合の演奏者-聴取者間相互作用の解析. 計測自動制御学会論文集. 2004, vol. 40, no. 2, p. 207-209.
- [5] Nakra, T. M., & BuSha, B.F. Synchronous Sympathy at the Symphony: Conductor and Audience Accord. *Music Perception: An Interdisciplinary Journal*. 2014, vol. 32, no. 2, p. 109-124.
- [6] Phillips-Silver, J., Aktipis, C. A., & A. Bryant, G. The ecology of entrainment: Foundations of coordinated rhythmic movement. *Music perception*. 2010, vol. 28, no. 1, p. 3-14.
- [7] Goebel, W., & Palmer, C. Synchronization of timing and motion among performing musicians. *Music Perception*. 2009, vol. 26, no. 5, p. 427-438.
- [8] Konvalinka, I., Vuust, P., Roepstorff, A., & Frith, C. D. Follow you, follow me: continuous mutual prediction and adaptation in joint tapping. *Quarterly journal of experimental psychology*. 2010, vol. 63, no. 11, p. 2220-2230.
- [9] 徳差雄太, 松谷健史, 空閑洋平, 村井純. 低遅延により自然な遠隔コミュニケーションを実現する映像配信システムの提案. マルチメディア, 分散協調とモバイルシンポジウム 2013 論文集. 2013, p. 911-917.
- [10] Flossmann, S. Grachten, M., & Widmer, G. Expressive Performance Rendering: Introducing Performance Context. *Proceedings of the SMC 2009 - 6th Sound and Music Computing Conference*. 2009, p. 155-160.
- [11] Keller, P. E., Knoblich, G., & Repp, B. H. Pianists duet better when they play with themselves: on the possible role of action simulation in synchronization. *Consciousness and cognition*. 2007, vol. 16, no. 1, p. 102-111.
- [12] Bello, J.P., Daudet, L., Abdallah, S., Duxbury, C., Davies M., & Sandler, M.B. A tutorial on onset detection in music signals. *IEEE Trans. Speech Audio Process*. 2005, vol. 13, no. 5, p. 1035-1047.
- [13] Ellis, D.P. (). Beat tracking by dynamic programming, *Journal of New Music Research*. 2007, vol. 36, no. 1, p. 51-60.
- [14] 堀内靖雄, 坂本圭司, 市川薫. 合奏時の人間の演奏制御の分析・推定. 情報処理学会論文誌. 2004, vol. 45, no. 3, p. 690-697.