Vol.2021-MUS-130 No.33
Vol.2021-EC-59 No.33
2021/3/17

配信ライブの同時視聴における ヘッドバンギング同期のための動作推定手法

二宮洸太[1]　中村聡史[1]

概要：音楽ライブに参加する観客は演奏に合わせてサイリウムを振る，ヘッドバンギングをするなど，アーティストや他の観客との一体感や非日常感を楽しんでいる．また，ライブの模様を，インターネットを通じて配信する配信ライブも多く行われているが，自宅でひとりで鑑賞することが多く，アーティストや他のファンとのかかわりや一体感が希薄化する問題がある．そこで，配信ライブ中の視聴者間の一体感を向上させることを目的に，ライブ中に行われるヘッドバンギングを媒介として，その動作を検知し，タイミングを視聴者間で共有するシステムを提案する．本研究ではポケットに入れたスマートフォンのセンサデータを使い，ヘッドバンギングの予備動作からヘッドバンギングの推定に関する検討を行った．具体的には，ヘッドバンギング中のセンサデータに関するデータセット構築を行い，機械学習により，予備動作からヘッドバンギングの推定を行った．その結果，93.5%の精度で推定を行うことができた．

キーワード：ライブ，配信ライブ，ロック，ライブモーション，ヘッドバンギング，スマートフォン，センサ，COVID-19

1. はじめに

Apple Music や Spotify といったストリーミングサービスの普及，並びに YouTube やニコニコ動画などの動画配信サービス上のミュージックビデオやライブ映像等の音楽コンテンツの増加によって，音楽が身近なものとなった．その中で音楽への価値観も変わりつつあり，アーティストは音源だけでなく，音楽を介した体験を様々な形で発表し，エンタテインメントとして提供している．特に，ライブは 2019 年には観客動員数が 5497 万人に達し，過去 10 年で最大となっている[1].

ライブは会場にそのアーティストのファンが集まっているという特殊な環境によって，一体感や空気感などを感じることができ，通常の音楽聴取では得られない音楽体験がある．ここで，観客同士で行うライブ独自の楽しみ方に「ライブモーション」がある．アイドルのライブでは，光る棒状のデバイス（サイリウム，ペンライト）や好きなメンバーへのメッセージが書かれたうちわをライブ中に振る行為や，楽曲中の決まった区間で掛け声を叫ぶ「コール」といったライブモーションがある．また，ラウドロックやヘヴィメタルといった激しい楽曲の多いアーティストのライブでは，観客同士が体をぶつけあう「モッシュ」や，観客たちが円状に走り回る「サークルピット」，4 つ打ちのビートにあわせて独特のステップを踏む「ツーステップ」，楽曲に合わせて頭を振る「ヘッドバンギング（以下ヘドバン）」などがある．このようなライブ独特の行動や気持ちの高まった観客たちの空気感によって，ライブ空間でしか体験できない一体感や非日常感をアーティストとファンは共有することができ，ライブ体験をよりよいものにしている．

一方でこのライブの模様をインターネット上で配信する配信ライブも広がりをみせている．これは会場に行かなくても手軽にライブを視聴できることや，アーカイブとしてライブ動画を残すことで一定の期間，自由に視聴できることから需要が高まっている．また COVID-19 の流行により，会場に多くの人を集められない状況下でもライブを行う方法としてさらに脚光を浴びている．しかし，会場で開催されるライブに足を運ぶときとは異なり，配信ライブで視聴者が行えることは限られている．YouTube Live や SHOWROOM といったライブ配信プラットフォームでは，アーティストや他のファンへのメッセージを発信する「コメント」機能や，特定のアイテムなどを購入することでアーティストを支援できる「投げ銭」といった機能が提供されている．これらの機能によって視聴者は自身の熱意を，アーティストや他の視聴者に発信することが可能であるが，会場で開催されるライブのように他の観客とライブモーションを行うことで得られる一体感を感じることは難しい．

ここで，配信ライブ中に視聴者がひとりでライブモーションを行うことも珍しくない．例えば Crossfaith[2]のメンバーの Teru は視聴者に対して配信ライブ中にヘドバンを煽るツイートを行っており [3]，その配信ライブ中には「パソコンに向かってヘドバンしている」というツイートが多数投稿されるなど，実際にライブ会場で行うヘドバンを配信ライブ中にひとりで行っていることがわかる．しかし，既存のライブ配信システムが提供する機能では，他者が行っているヘドバンの様子を知ることができず，他者の存在が希薄になってしまうため，一体感を得ることは難しい．

そこで本研究では，配信ライブでも視聴者が行いやすいライブモーションであるヘドバンを通じた一体感の向上を支援することを目的とし，ヘドバン共有システムの実現を目指す．このシステムの実現には視聴者のヘドバン動作を

1 明治大学
Meiji University

2 http://crossfaith.jp/
3 https://twitter.com/terucrossfaith/status/1302824740940201984

推定し，その動作を視覚や聴覚を介して他者と一緒にヘドバンしていると感じられる形で共有･伝達する必要がある．その中でも本稿では，ヘドバンのタイミングを機械学習によって推定する手法を実現し，その有用性を検証する．なお，本研究ではできるだけ多くの人が手軽に参加できるようにするため，ズボンのポケットに入れたスマートフォンのセンサ情報のみを用い，配信ライブ中のヘドバン動作を推定する．またヘドバンのリアルタイム共有のためにはヘドバン後に推定するのでは間に合わないため，ヘドバンの予備動作からの推定を行う．

2. 関連研究

ライブに関する研究はこれまで多数行われている．Brown ら[2]は，ライブに行きたい理由に関する調査を行い，ライブ参加者はライブの雰囲気や演奏が毎回変化するユニークな体験を求めていること，アーティストを生で見たいこと，ファンとアーティスト並びにファン同士のコミュニケーションやインタラクションを求めてライブに参加していることを明らかにした．また，ライブ中の観客の動きに着目した研究もあり，Swarbrick ら[3]は，あるアーティストのファンとファンではない人にライブ体験と音源聴取をそれぞれさせ，動きを比較する実験を行った．その結果，ライブを体験させた場合，音源聴取のみの場合に比べ観客は活発に動くこと，特にファンの方がファンではない人よりも活発に動くことを明らかにした．また，ファンとファンではない人は曲によって動きが異なる事例が見られ，ファン独自の盛り上がり方があることも明らかにした．このようなライブ中の観客の動きに関して，Silverberg ら[4]は，ヘヴィメタルコンサート中の観客の動き，特に「モッシュ」という観客同士が体をぶつけあう動きに関する研究を行っている．ここでは，主体的にモッシュを行う Active Mosher と，主体的には行わず他者のモッシュによって巻き込まれる Passive Mosher に観客を分け，流体力学的にシミュレートすることで，観客の動きをモデル化している．このようにライブの観客にまつわる研究は多く行われている．本研究ではライブで観客が行うヘドバンを対象に，その特性の解明を目指すものである．

ライブの支援を目的とした研究も数多く行われている．Freeman ら[5]は，大人数の観客がそれぞれのスマートフォンを利用し，演奏者とインタラクションを可能とする massMobile を開発した．これは観客の投票によって楽曲のテンポやダイナミクスが変化する TeamWork と，観客が描いた図形が映像として演奏者の背後に投影される Sketching の 2 つの機能により，観客のライブへの参加感を高めている．Kato ら[6]は，音楽再生をインターネット経由で同期し，機器を制御する大規模音楽連動制御プラットフォーム Songle Sync を研究開発した．これは音楽理解技術 Songle[7]のデータを用いて音楽に連動した映像や照明効果を，スマートフォンや IoT デバイスといった複数の端末で同期可能とすることで，一体感のある演出を行うことができる．これらは実際にライブ会場で利用され，演出等を変化させることで観客の体験を向上させている．本研究では配信ライブを対象に，ヘドバンの共有を行うことで，視聴者間のつながりを向上させるものである．

ヘドバンを用いた研究として，Bardos ら[8]は，ギター型コントローラで音源を選択し，ヘドバンによって音源の再生を認識することで，ミュージシャンでなくとも作曲できるシステム Bangarama を提案している．また，Merrill[9]は，ギター演奏時のエフェクトの変更を頭の動きのトラッキングとジェスチャ認識で行うシステムを開発している．さらに Moll ら[10]は，ヘドバンを撮影し，その映像を PoseNet[11]を用いて取得した骨格情報からヘドバンのリズムを推定し，ビートとヘドバンのマッチによって得点の得られる音楽ゲーム HeadbangZ を作成している．このようにヘドバンを用いたシステムは多く存在し，本研究でのデータセット構築や機械学習での推定において，これらのヘドバン技術を参考にしているが，事後推定であることや，配信ライブでの利用を想定している点で，本研究はこれらの研究と異なる．

センサ情報を用いた音楽にまつわる研究として，安永ら[12]は，音楽からの舞踊動作の生成に音楽特徴量に加えて，ユーザの動きの盛り上がりを加速度センサとして特徴量化し，キャラクタの動きをインタラクティブに制御できるシステムを開発している．Kanke ら[13]は，ドラムの中で使用頻度の低い打楽器の仮想化を目的として，ドラムスティックに加速度センサを取り付け，その情報を用いて仮想化された打楽器の叩打を推定している．このように，センサ情報は舞踊動作やドラム演奏といった楽曲に同期した動作の取得，推定にも用いられており，ヘドバン動作の手掛かりとしてもセンサ情報が利用できると考える．

3. ヘドバン共有システム

本研究では，ライブモーションの中で，配信ライブ中に視聴者が行いやすいヘドバンを通じた一体感の向上を支援することを目的とし，ヘドバン共有システムの実現を目指す．その実現イメージを図 1 に示す．

このシステムでは，視聴者はそれぞれの家で，パソコンでライブ映像を視聴しながらヘドバンを行う．また，そのヘドバンをシステムが推定し，他者に共有する．ここで，ヘドバンの推定に関しては，動画から骨格情報を取得できる OpenPose[14]などの手法を用いて，パソコンのカメラデータから，ヘドバン推定を行う方法も考えられるが，計算量が膨大であり，配信ライブのような環境で大人数の情報をリアルタイムに処理することは難しい．そこで，多くの人が所有しているスマートフォンのセンサを用いるだけで，

Vol.2021-MUS-130 No.33
Vol.2021-EC-59 No.33
2021/3/17

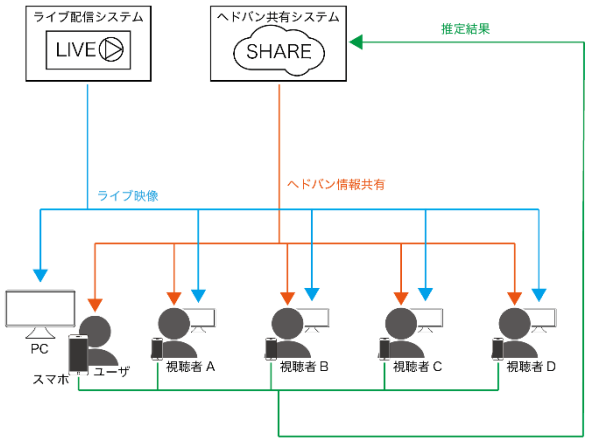

図 1 ヘドバン共有システムの概要図

多くの配信ライブ視聴者が利用できるようにする．ヘドバンは腰から上半身を前後に動かす動作であり，ヘドバン中に腰や太ももは頭の動きに連動し前後する．この特性を利用すると，ズボンのポケットにスマートフォンを入れておくことで，ヘドバンの動きをセンシングできると期待できる．

ここで，実際のヘドバンのタイミングに合わせてヘドバンの推定結果を共有システムに送信し，複数人の情報を集約し，視聴者にフィードバックするためには，配信ライブ中にリアルタイムに推定する必要があり，ヘドバン後に認識して送信するような仕組みは使えない．また，認識にも時間がかかることや，通信遅延も考慮すると，ヘドバンを行おうとしている予備動作から推定する必要がある．そこで，ある時点までのデータを用いてその先の予測を行う．その概略を図 2 に示す. ある推定したいヘドバンに対して，予測するまでの時間となるインターバルを設け，その前のフレームを特徴量区間とし，このデータを用いて特徴量化を行う．また，そうした特徴量を用い，機械学習によってヘドバンのタイミングを推定する．

4. ヘドバンデータセット構築

4.1 データセット構築設計

配信ライブにおけるヘドバンを介した視聴者の一体感向上を実現するために，視聴者の動きをセンシングし，ヘドバン動作をリアルタイムに推定する必要がある．そこで

図 2 推定位置と特徴量の関係

4 http://www.lasvegas-jp.com/,
5 http://coldrain.jp

まず，データ収集システムを開発し，ヘドバンセンサデータセットの構築を行った．

データセット構築においても，実現するヘドバン共有システムの環境を想定し，スマートフォンを用いて行う．また, ユーザの動作に対してアノテーション付与を行うため, 実際のヘドバンの動きをカメラによって撮影する．具体的にはスマートフォンで楽曲の再生とセンサ情報の収集を行い, カメラを搭載したパソコンで収集中の様子を撮影する．ヘドバンは音楽を聴きながら行うため，データセット構築においても実際のライブでヘドバンが行われる楽曲を聴きながらデータの収集を行う．

今回使用した楽曲を表 1 に示す．これらは実際に著者がライブに行き，ヘドバンを行った曲の中で，ヘドバン時間が長いものを選定した．一方で，ヘドバンは視聴者が自身の判断で行うものであるため，データセット構築実験でヘドバンを行う区間に関しては実験協力者に任意で決めてもらうこととした．どこでヘドバンを行っているかについては，データ収集中の様子の動画から手作業で決定するものとした．

4.2 データ収集システム

センサデータと動画データの対応づけを行うために，実験時に聴いてもらう楽曲に同期した状態でセンサデータと動画データの収集を行う Web システムを開発した．システムの概要図を図 3 に示す．実装において，バックエンドは Express.js，MySQL，フロントエンドは Next.js を用いた．なお，撮影用のパソコンとスマートフォン間の同期通信は

表 1 データ収集に用いた楽曲

歌唱アーティスト	曲名	時間
Fear, and Loathing in Las Vegas[4]	Twilight	3:54
Crossfaith[2]	Monolith	3:43
Coldrain[5]	The Revelation	4:22
SiM[6]	JACK. B	4:04
Survive Said The Prophet[7]	T R A N S l a t e d	3:42

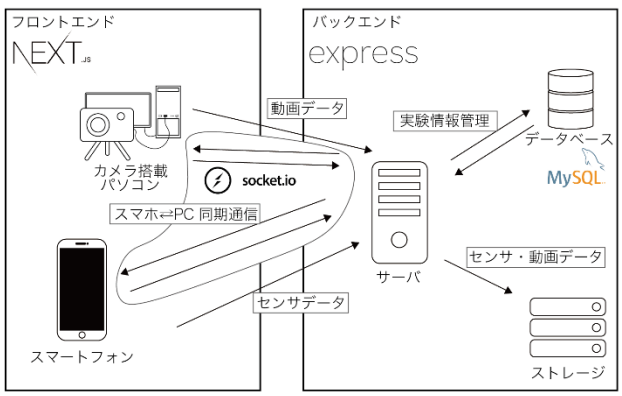

図 3 データ収集システムの概要図

6 https://sxixm.com/
7 https://survivesaidtheprophet.com/

Vol.2021-MUS-130 No.33
Vol.2021-EC-59 No.33
2021/3/17

socket.io を用いた．実験では，パソコンとスマートフォンそれぞれでシステムへのアクセスを行う．

システムは，実験協力者がスマートフォンで楽曲を選択し，データ収集を開始すると，パソコンに開始時間が送られ，スマートフォンとパソコンが時刻を合わせて同時にデータ収集が行われる仕組みとなっている．なお，スマートフォンとパソコンは実験協力者が各自所有しているものを利用してもらった．ただし実験システムの都合上スマートフォンは iOS 端末に，パソコンのブラウザは Google Chrome に限定した．また，センサ情報として，3 軸加速度（x， y，z）と 3 軸デバイス方向（alpha， beta， gamma）は 60fps，動画は 30fps で収集した．

4.3 データセット構築実験の実施

実験協力者は，ヘドバンを行うライブによく行く大学生 6 名（男性 5 名，女性 1 名，著者含む）である．実験協力者には事前にデータセット構築で使用する楽曲を聴きこんでもらい，どこでヘドバンを行うかについて各自の判断で決めてもらった．また，実験前にシステムの動作確認と実験の流れを把握してもらった．なお，実験は各実験協力者の自宅で行ってもらうこととした．これはセンサデータのみをデータセットとして利用することから，場所の影響を受けにくいと判断したためである．

実験ではイヤホンで音楽を聴きながら，表 1 に示した楽曲 1 曲に対し，それぞれ 5 回ずつデータ収集を行ってもらった．そのためデータセットは実験協力者 6 名×使用楽曲 5 曲×5 回の 150 データが集まった．

ヘドバン動作の確認をわかりやすくするため，実験中はカメラに対し，横向きでヘドバンを行うように指示をした．またスマートフォン（システムの都合上 iPhone）はイヤホンジャックを上，画面を肌と逆側，ポケットの位置を右ポケットにするよう指示し，向きや位置を統制した．さらにヘドバンをしていない時はジャンプ等の過度な動きは控えてもらったうえで自由に音楽を聴くよう指示した．なお，ヘドバンは首や腰に負担がかかるため，首や腰に配慮したうえで，複数日に分けて実施することを推奨した．

4.4 ヘドバン区間のラベリングとデータ変換

実験で得られたセンサデータに対して，実験で収集した動画データを手掛かりとして，ヘドバン区間とその周期に関する情報を付与する．ここで，ヘドバンには頭を横に振るものや全身を使わず首だけで行うものもあるが，本研究で扱うヘドバンは全身を使った上下に振るものに限定する．

ラベルとしては非ヘドバン時を 0，ヘドバン中の頭を上から下に振り下げる動作（振下動作）を 1，ヘドバン中の頭を下から上に振り上げる動作（振上動作）を 2 と設定する．ヘドバンは振下動作と振上動作が交互に繰り返される運動であり，このタイミングが一致することで一体感が得られる．そのため，それぞれの動作を認識する必要があり，振下動作と振上動作に分けてラベリングを行った（図 4 参照）．また，実際のライブで一緒にヘドバンをしていると感じるのは振下・振上動作の開始タイミングが一致しているときである．そのため，振下・振上開始地点を推定する．開始地点は振下・振上区間の開始から 3 フレーム目の位置とした．これは，頭の振りにおいて上半身に力を加え，初速を与える地点であり，このタイミングが一致することで，一体感が得られると考えたためである．

機械学習による推定を行ううえで，加速度，デバイス方向それぞれについて値の変換を行う．まず，加速度については，x，y，z それぞれの向きの値を合成した合成加速度を算出する．これにより，端末の傾き等による方向の変化の影響を除外する．

デバイス方向は alpha，beta，gamma の 3 種類を取得している．これは，それぞれ z 軸，x 軸，y 軸を中心としたデバイスの動きであり，角度として取得できる．このうち取得した alpha 値の一部を図 5 に示す．alpha 値の範囲は 0° 以上 360°未満となっているため，360°以上となった場合は 360°引かれた値が記録される．そのため 359°から 360°へ変化した場合，記録上は 359°，0°と記録されてしまう．これではデバイスの動きを正しくとらえているとは言えない．そこで取得されたデバイス方向の角度をθとして，$x = \cos\theta, y = \sin\theta$の形に変換し，2 次元座標系にマッピングすることによって補正を行う．マッピング後の値を図 6 に示す．なお alpha 以外の beta，gamma についても同様に変換する．

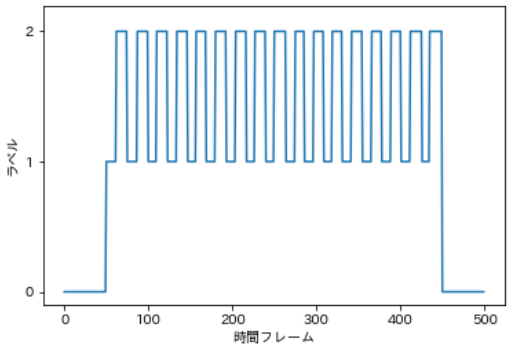

図 4 ラベリング結果

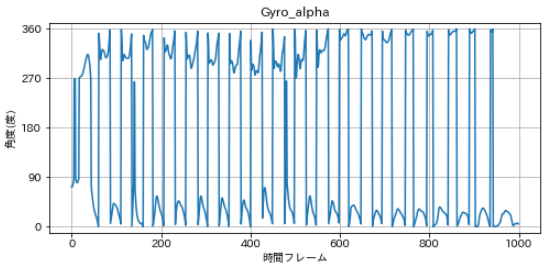

図 5 取得されたデバイス方向の値の一部

Vol.2021-MUS-130 No.33
Vol.2021-EC-59 No.33
2021/3/17

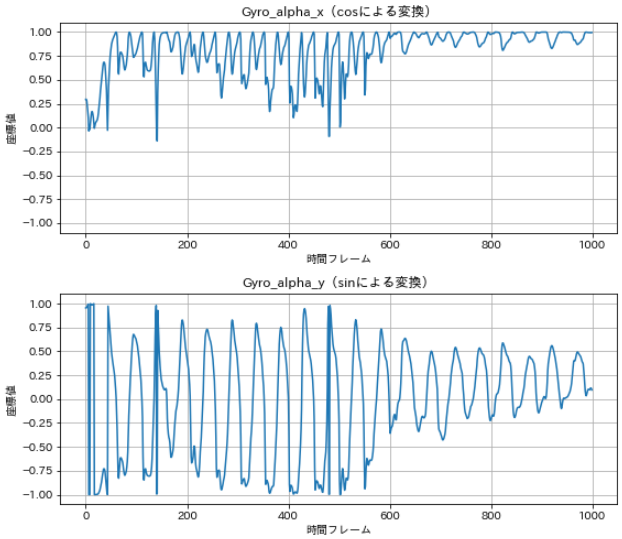

図 6 変換後のデバイス方向の値のグラフ

4.5 結果

実験で得られたデータをグラフにプロットしたものを，図 7，図 8 に示す．ここで図の横軸は時間フレーム，縦軸はセンサの値である．なおセンサは 60fps，ラベルは 30fps で収集したため，時間によって位置を合わせたうえで，センサ 2 フレームに対してラベル 1 フレームとしてプロットしている．図 7 はあるユーザの 1 試行全体でのセンサ値の変動の様子を示したものである．青色の線がセンサの値，背景色が灰色の部分は振下動作と振上動作をまとめたヘドバン区間を表している．合成加速度，デバイス方向の形状は，ヘドバン区間と，非ヘドバン区間で異なっており，ヘドバンの身体的な特性がセンサに表れていることがわかる．ここからズボンのポケットでも十分にヘドバンの特徴量となるデータが取得できていると言える．

次にあるヘドバン区間において，振下動作，振上動作とそのセンサの値をプロットしたものを図 8 に示す．青色の線がセンサの値，背景色が赤色の区間が振下動作，緑色の区間が振上動作を示している．合成加速度，デバイス方向ともに周期的な波形をしており，ヘドバンの頭の動きの周期性がセンサに表れていることがわかる．また，振下動作と振上動作でそれぞれの区間でのグラフ形状に着目すると，それぞれのセンサ値において類似性が見られる．これらのことからヘドバンの振下動作，振上動作に関しても，ズボンのポケット内のスマートフォンから取得できていることがわかる．

5. 機械学習による推定

4 章で構築したデータセットを用いてヘドバンの推定を行う．5.1 節ではデータセットから学習するうえでの特徴量抽出の方法について述べ，5.2 節では予測までの時間と特徴量の切り出す時間を変化させ学習を行い，推定精度の変化について考察する．

5.1 特徴量抽出とラベリング

今回構築したデータセットはラベルが 30fps，センサデータが 60fps となっている．そのため，フレームレートの違いに関しては，事前に時間基準でそろえたうえで，ラベル

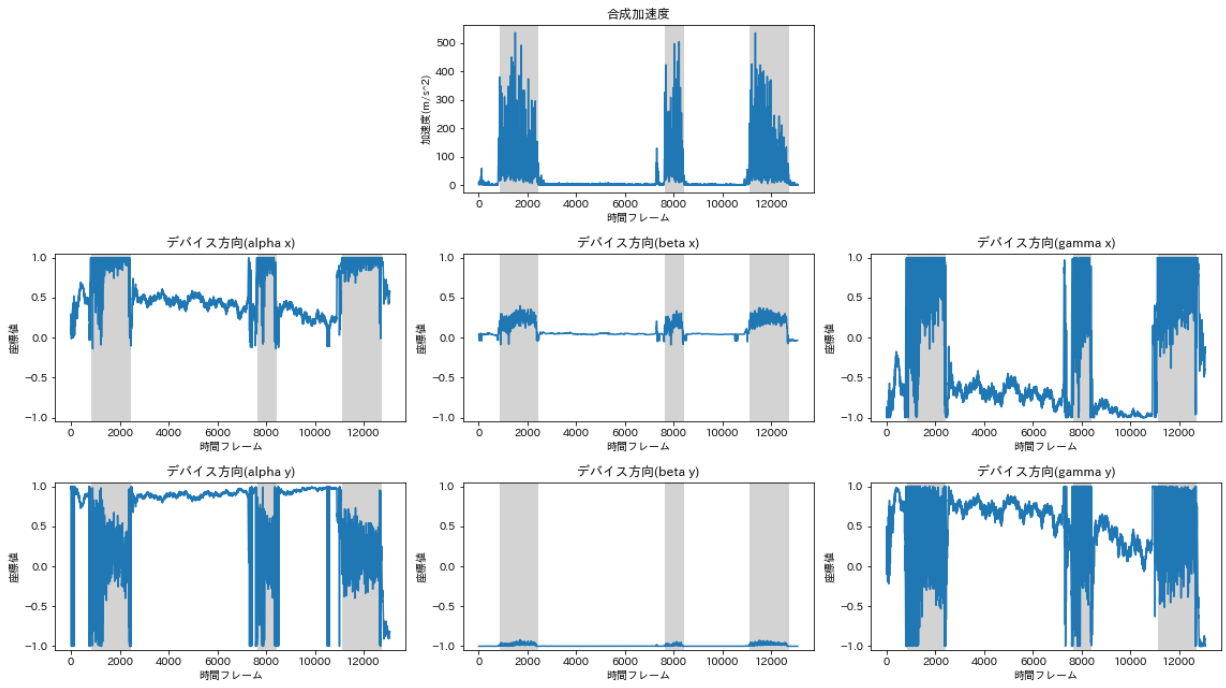

図 7 あるユーザの 1 試行分のセンサとラベルの対応のグラフ

Vol.2021-MUS-130 No.33
Vol.2021-EC-59 No.33
2021/3/17

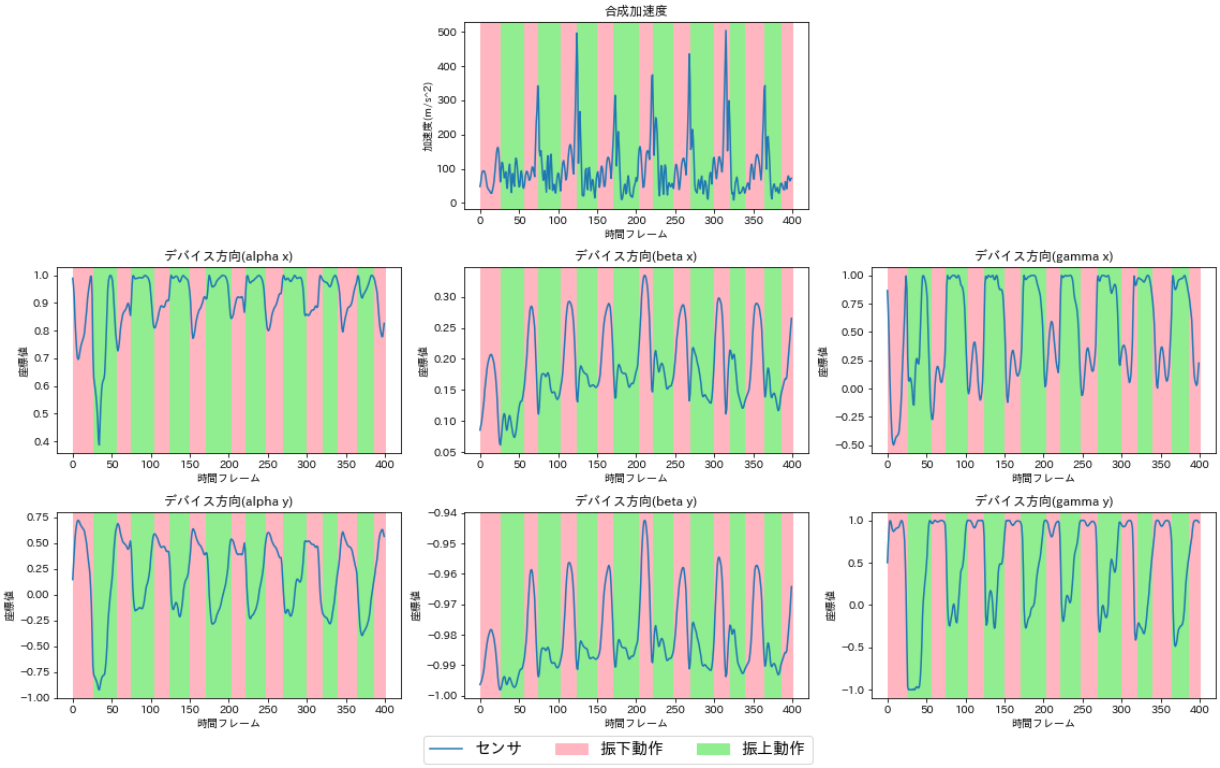

図 8 あるヘドバン区間におけるセンサと振下・振上動作の対応のグラフ

1 フレームに対してセンサ 2 フレームを切り取ることで統制する．

本研究ではヘドバンの振下・振上動作の開始地点の推定を行うため，タスク設定としてはヘドバンを行っていない非ヘドバンと振下開始，振上開始の 3 値分類とする．振下・振上開始は開始地点のラベルを起点としてデータを利用する．ヘドバンを行っていない区間では，1 秒ごとにデータをずらした位置を開始地点とし，利用する．

次に特徴量化は，先述の特徴量区間内でデータを切り取り統計量化することで行う．データは合成加速度と 3 軸デバイス方向に加え，それぞれの差分データを利用する．差分データは隣接するデータをx_n, x_{n+1}とすると，差分$d_n = x_{n+1} - x_n$となる．これらのデータに対し，センサの種類ごとに，平均，標準偏差，最大，最小を求め，特徴量化する．そのためデータの次元は，センサ 7 種×特徴量とその差分の 2 種×統計量 4 種の 56 次元となる．

5.2 予測までの時間と推定精度

本研究ではヘドバンのリアルタイム推定を想定しているため，なるべく短い時間で推定することが求められる．しかし，時間が短いほど特徴を捉えることが難しくなり，推定精度は低下すると考えられる．そこで，実際に予測までの時間となるインターバルと特徴量区間の時間の 2 つのパラメータを変化させ，推定精度への影響を検証する．特徴量は 5.1 節の方法で抽出し，インターバルと特徴量区間の時間を変化させる．インターバルは 50，75，100 ミリ秒の 3 種，特徴量区間は 100，150，200，250，300 ミリ秒の 5 種を採用する．4 章で構築したデータに対し，特徴量化を行い，そのうちの 80%を Train データ，20%を Test データとしてランダムに分割した．機械学習アルゴリズムは Random Forest を用い，実装には Python の機械学習ライブラリである scikit-learn[15]を用いた．

各インターバルでの精度について，表 2 に正解率を，表 3 に適合率，表 4 に再現率，表 5 に F 値を示す．横軸がインターバル，縦軸が特徴量区間の時間，値が各モデルの精度となっている．どのモデルにおいても特徴量区間の時間が短くなるほど精度が下がっており，特に再現率が下がっていることから，振下・振上開始の推定が難しくなっていることがわかる．これは特徴量として使える時間が短くなるために，特徴をうまく捉えられないことが原因と考えられる．次に同じ特徴量区間の時間内でインターバルの時間の違いによる変化を見る．どのインターバル時間においても，特徴量時間 300 ミリ秒のモデルでは正解率 93%前後，100 ミリ秒のモデルでは 87%前後であり，インターバル時間の違いによる差異は見られない．これは，インターバル時間の変化が 25 ミリ秒と短いこと，インターバルのずれに応じて捉えられる特徴も変化していることなどが理由として考えられる．

以上の結果から，実際の運用において時間パラメータを決定する場合は，インターバルの時間より，特徴量区間として利用する時間が重要であると考えられる．

6. リミテーション

今回構築したモデルで，最大 93.5%の精度でヘドバンの

Vol.2021-MUS-130 No.33
Vol.2021-EC-59 No.33
2021/3/17

推定ができることが明らかになった．しかし，配信ライブでヘドバンの共有を行うためには課題が残されている．

まず，今回使用したモデルでは，振下・振上開始以外のヘドバン中のデータを含んでいないことが挙げられる．そのため，ヘドバン中の開始地点かそうでないのかの分類ができない．実際に，開始地点以外のデータを含んでモデル構築を行ったところ，開始地点の推定精度は約 60%であり，正しく推定できているとは言えない．この問題の解決には，ヘドバンの周期が役に立つと考える．ヘドバンの周期は著者の経験上，楽曲の区間に依存する．そのため，各楽曲のヘドバン区間とその周期を考慮して，ユーザの振下・振上開始を推定することで，高精度な推定が行えると考える．ここで，今回構築したデータセットの各実験協力者のヘドバン区間を楽曲ごとにまとめたものを図 9 に示す．実験では実験協力者の任意の区間でヘドバンを行ってもらったが，多くの区間でヘドバン箇所が一致していることがわかる．

表 2 インターバルと特徴量区間ごとの正解率

		インターバル（ms）		
		0.050	0.075	0.100
特徴量区間（ms）	0.10	0.870	0.871	0.870
	0.15	0.900	0.901	0.901
	0.20	0.912	0.913	0.913
	0.25	0.921	0.924	0.930
	0.30	0.933	0.933	0.935

表 3 インターバルと特徴量区間ごとの適合率

		インターバル（ms）		
		0.050	0.075	0.100
特徴量区間（ms）	0.10	0.844	0.844	0.846
	0.15	0.883	0.886	0.882
	0.20	0.896	0.898	0.901
	0.25	0.911	0.914	0.918
	0.30	0.920	0.923	0.924

表 4 インターバルと特徴量区間ごとの再現率

		インターバル（ms）		
		0.050	0.075	0.100
特徴量区間（ms）	0.10	0.795	0.798	0.795
	0.15	0.841	0.843	0.845
	0.20	0.865	0.867	0.865
	0.25	0.878	0.881	0.892
	0.30	0.897	0.897	0.900

表 5 インターバルと特徴量区間ごとの F 値

		インターバル（ms）		
		0.050	0.075	0.100
特徴量区間（ms）	0.10	0.816	0.818	0.817
	0.15	0.860	0.862	0.862
	0.20	0.879	0.882	0.882
	0.25	0.893	0.896	0.904
	0.30	0.908	0.909	0.912

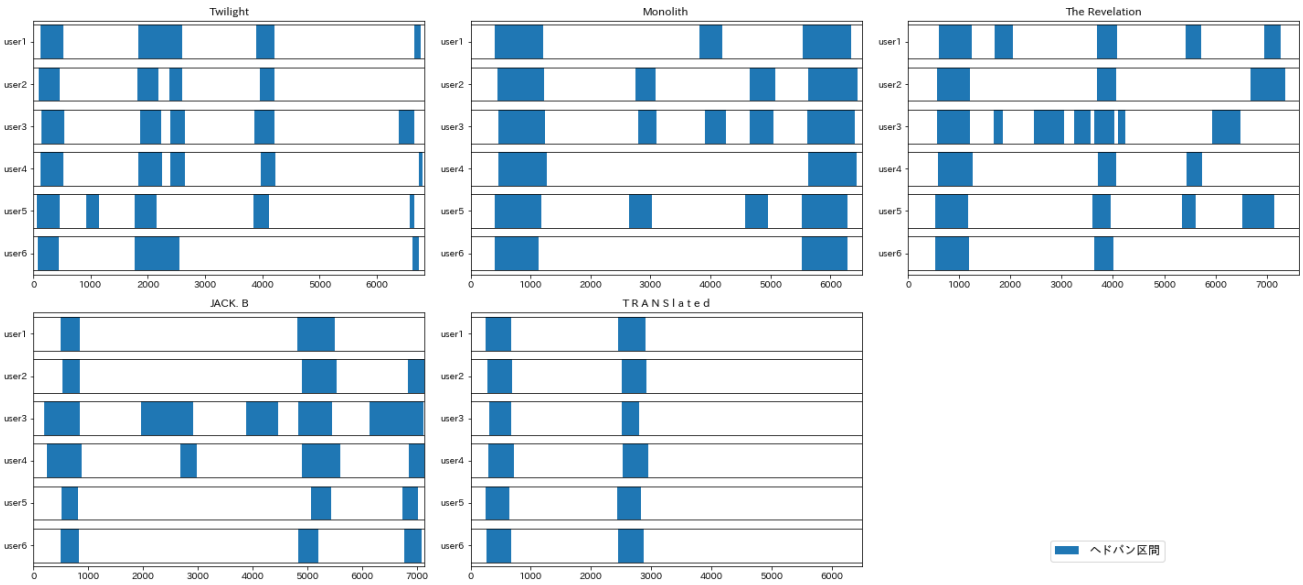

図 9 楽曲ごとに実験協力者のヘドバン区間を可視化した様子

このことから，ヘドバンを行うライブに行く人は，楽曲の特徴からヘドバンを行う区間を判定し，ライブにおいては参加者が暗黙的に了解している可能性が考えられる．そのため，ヘドバンを行う区間の楽曲的な特徴を明らかにすることで,楽曲からのヘドバンの推定が可能であると考える．

また，配信ライブでは遅延が生じることが想定される．本研究の目指すシステムは，大人数の視聴者が同じタイミングでヘドバンを行い，他の視聴者のヘドバンを体感することで一体感を向上させるものである．そのため，遅延によってタイミングがずれてしまうと一体感の低減や違和感につながってしまうと考えられる．5.2 節の結果から，遅延具合にあわせて予測までのインターバルを変化させることでこの問題の解決ができると期待される．しかし，インターバルの変化では大きな遅延に対応することは難しい．この場合は，遅延具合が近い視聴者間でヘドバン情報の共有を行うことが考えられる．どの程度遅延しているかは配信開始からの時間を記録する等の方法で取得可能であると思われる．なお，この手法では共有できる視聴者の人数の低減につながるため，どの程度の視聴者数がいれば一体感を感じられるかについても調査を行う必要がある．

7. おわりに

本研究では，配信ライブ中の視聴者のヘドバン動作のリアルタイム事前推定を目指し，ヘドバンデータセットの構築を行った．具体的には，スマートフォンを用いて楽曲聴取中にヘドバン動作を行ってもらい，3 軸加速度，3 軸デバイス方向のセンサデータを収集する実験を行った．また Random Forest を用いて機械学習による推定を行い，最大 93.5%の精度でヘドバンの振下開始，振上開始，ヘドバンを行っていない状態の分類を行うことができた．

今後は 6 章で挙げた楽曲情報を利用した振下，振上開始の高精度な推定と，配信中に発生する遅延を考慮した推定の 2 つに取り組むとともに，ヘドバン情報の共有・伝達方法について検討を行う．ヘドバン情報の共有を行うにあたって，ヘドバン中は画面を見ることができないため，視覚情報以外の方法で提示することが望ましい．また，配信ライブでは各個人がイヤホンを装着可能であること，センサデータ取得にスマートフォンをポケットに入れていることを考慮すると，ライブ音声を利用して聴覚情報として共有する方法や，スマートフォンを振動させ触覚情報として伝達する方法などが考えられる．さらに，ヘドバンの共有の前には複数人のヘドバン情報を統合する必要があるため，どのような方法で統合を行うかについても，今後検討していく．

謝辞 本研究の一部は，JST ACCEL（グラント番号 JPMJAC1602）の支援を受けたものである．

参考文献

[1] “2020 ライブ・エンタテインメント白書”. https://live-entertainment-whitepaper.jp/, (参照 2021-02-14).
[2] Brown, S. C., Knox, D.. Why go to pop concerts? The motivations behind live music attendance, Musicae Scientiae, 2017, vol. 21, no. 3, pp. 233-249.
[3] Swarbrick, D., Bosnyak, D., Livingstone, S. R., Bansal, J., Marsh-Rollo, S., Woolhouse, M. H., Trainor, L. J.. How Live Music Moves Us: Head Movement Differences in Audiences to Live Versus Recorded Music, Frontiers in Psychology, 2019, vol. 9, pp. 2682.
[4] Silverberg, J. L., Bierbaum, M., Sethna, J. P., Cohen, I.. Collective Motion of Moshers at Heavy Metal Concerts, Physical Review Letters, 2013, vol. 10, no. 22, pp. 228701.
[5] Freeman, J., Xie, S., Tsuchiya, T., Shen, W., Chen, Y., Weitzner, N.. Using massMobile, a flexible, scalable, rapid prototyping audience participation framework, in large-scale live musical performances, Digital Creativity, 2015, vol. 26, no. 3-4, pp. 228-244.
[6] Kato, J., Ogata, M., Inoue, T., Goto, M.. Songle Sync: A Large-Scale Web-based Platform for Controlling Various Devices in Synchronization with Music, Proc. of MM 2018, 2018, pp. 1697-1705.
[7] Goto, M., Yoshii, K., Fujihara, H., Mauch, M., Nakano, T.. Songle: A Web Service for Active Music Listening Improved by User Contributions, Proc. of ISMIR 2011, 2011, pp. 311-316.
[8] Bardos, L., Korinek, S., Lee, E., Borchers, J.. Bangarama: Creating Music With Headbanging, Proc. of NIME 2005, 2005, pp. 180-183.
[9] Merrill, D.. Head-tracking for gestural and continuous control of parameterized audio effects, Proc. of NIME 2003, 2003, pp. 218-219.
[10] Moll, P., Leibetseder, A., Kletz, S., Lux, M., Muenzer, B.. Alternative inputs for games and AR/VR applications: deep headbanging on the web, Proc. of MMSys 2019, 2019, pp. 320-323.
[11] PoseNet. https://github.com/tensorflow/tfjs-models/tree/master/posenet, (参照 2021-02-14).
[12] 安永卓哉，中澤篤志，竹村治雄. 加速度センサによるユーザコントロールを導入した音楽に合った舞踊動作の自動生成, 研究報告音声言語情報処理（SLP）, 2012, vol. 2012-SLP-90, no. 25, pp. 1-6.
[13] Kanke, H., Takegawa, Y., Terada, T., Tsukamoto, M.. Airstic Drum: A Drumstick for Integration of Real and Virtual Drums, Proc. of ACE 2012, 2012, pp. 57-69.
[14] Cao, Z., Hidalgo, G., Simon, T., Wei, S., Sheikh, Y.. OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields, Proc. of CVPR 2017, 2017.
[15] scikit-learn. https://scikit-learn.org/, (参照 2021-02-14).