



映像に基づく運動体感の自動生成に関する研究

第 1 報：姿勢推定に基づく跳躍動作の体感生成

赤井峻真¹⁾, 松原亨¹⁾, 昆陽雅司¹⁾, 田所諭¹⁾

Ryoma AKAI, Toru MATSUBARA, Masashi KONYO and Satoshi TADOKORO

1) 東北大学 情報科学研究科 (〒 980-8579 宮城県仙台市青葉区荒巻字青葉 6-6-01, konyo@rm.is.tohoku.ac.jp)

概要: 動画を視聴している際の臨場感を増すために、演者の運動に連動した体感を振動刺激により提示する方法を提案する。本稿では跳躍動作を対象とし、跳躍の速さや高さに応じた体感を自動生成する。姿勢推定ライブラリを用いて人物の骨格を追い、跳躍の速さや高さに応じて振動刺激を生成する。振動刺激は効果音に基づいて生成し、著者らが提案する音響-触覚変換手法を利用する。被験者実験により提案手法の有効性を検証する。

キーワード: 触覚, 振動提示, 音響-触覚信号変換, 姿勢推定

1. はじめに

人の動きに基づいて振動刺激を生成すると視聴体験が向上すると考えられる。動きに基づいて音を出す方法としてゲームやアニメでは効果音が用いられている。そこで、振動刺激を生成する方法として、効果音を利用する方法を検討する。

しかし、振動子の周波数応答が限られているため、従来では聴覚で利用するような高周波帯の周波数は振動刺激として利用することが困難であった。そこで、高周波数の振動刺激を提示可能な低周波数の振動刺激に変換する方法として、山口らが提案したヒトの高周波振動に対する知覚特性に着目した感覚等価変換手法 Intensity Segment Modulation (ISM) [1] を用いる。ISM とは、ヒトの高周波振動の強度知覚と包絡線情報の知覚の 2 つの知覚特性に着目した変換手法であり、高周波振動を 5ms のセグメントに分割し、セグメント毎に知覚インテンシティを維持して任意のキャリア周波数の振動に変換する手法である。ISM 変換を用いて効果音を振動刺激に変換することで、本来提示できない周波数の振動刺激を提示することができる。

また、振動刺激は運動に連動して生成することが望ましい。これまで人の動きを計測する際にはモーションキャプチャを用いたり、身体に多数の慣性センサを取り付ける方法が一般的であった。しかし、これらの方法は、コストや手間がかかり、簡単に計測することができない。そこで、スマートフォンなどで撮影した動画から姿勢推定ライブラリを用いて人の動きを推定する方法を利用する。筆者らは、スマートフォンを利用した体感付き動画を配信する方法も提案しており、運動の撮影から振動提示まで、スマートフォン 1 台で完結できると期待される [2]。

本研究では、映像内の人の動きに基づいて振動刺激を自動生成する手法を提案する。運動から振動刺激を生成する

方法として、運動に効果音をつけて、その効果音を振動刺激に変換する手法をとる。また、体感を向上させるために、運動に応じて効果音の音量を変化させる。

2. 提案手法

2.1 映像から運動を抽出する方法

映像から運動を抽出する方法として、姿勢推定ライブラリである BlazePose[3][4] を用いた人の骨格の姿勢推定を行った。BlazePose では機械学習を使用して推測を行っている。BlazePose では全身の 33 点のキーポイントを取得することができる。これは、標準的な COCO データセットよりも顔と手足のキーポイントが多い。スマートフォン上でもリアルタイムで姿勢推定を行うことができ、30FPS 以上で動作する。

推論を行うパイプラインには detector-tracker を使用した。detector がフレーム内の姿勢の関心領域を特定し、tracker がキーポイントを予測する。画像内に頭部が写っているという仮定をすることで、高速で軽量の detector を実現している。また、人の腰の中心点、人体全体に外接する円の半径、肩の中心点と腰の中心点を結ぶ線の傾斜角度を予測しており、複雑なポーズでも一貫した追跡が可能である。

実際に姿勢推定を行った例の一部を図 1 に示す。本稿では簡易的に跳躍動作の高さを推定するために、左足首の y 座標を取得した。

2.2 運動から振動刺激を生成する方法

運動から振動刺激を生成する方法として、運動に効果音をつけて、効果音を ISM 変換して振動刺激に変換した。ISM を使うことで効果音を感じやすい触覚信号に変換することができる。また、体感をよくするために、運動に応じて効果音の音量を変化させる。

ゲームやアニメなどで用いられているように、跳躍動作には離陸時と着地時の 2 回に効果音をつけた。離陸時の風



図 1: BlazePose で姿勢推定した例の一部

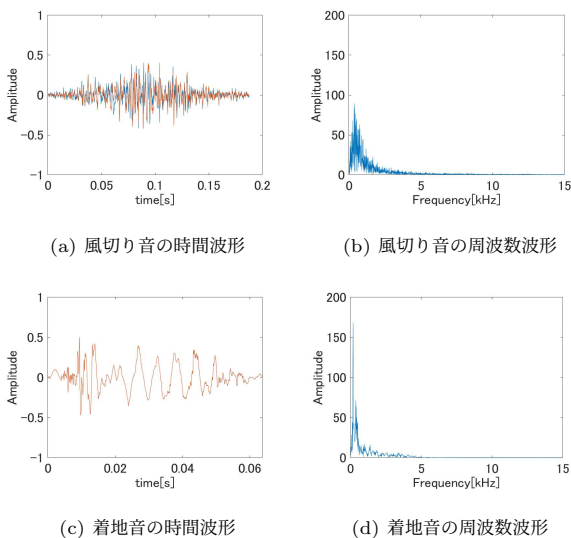


図 2: 効果音の波形

切り音は効果音辞典 [5] の効果音を、着地音は OtoLogic [6] の効果音を利用した。効果音の生波形と FFT をかけたグラフを図 2 に示す。風切り音は 200~800Hz の音の振幅が大きく、400Hz の音の振幅が最も大きい。4kHz ほどまでの周波数の音が含まれており、4kHz 以上の周波数の音も振幅がわずかなもの含まれている。着地音は 150~500Hz の音の振幅が大きく、200Hz の音の振幅が最も大きい。5kHz ほどまでの周波数の音が含まれている。

風切り音、着地音ともに高周波の音が含まれていた。触覚の検出閾には周波数依存性があり高周波の振動は感じにくい [7]。そこで、体感をよりよくするために高周波成分の振幅を大きくした。具体的には、風切り音は 800Hz~4kHz の振幅を 2 倍にし、着地音は 800Hz~6.3kHz の振幅を 2 倍にした。

姿勢推定して得られた左足首の y 座標のうち、足が地面についているときの座標と跳躍の最高到達点の座標から跳躍の高さを求めた。足が地面についているときの座標及び跳躍の最高到達点の座標は、座標の極大値及び極小値により求め、それぞれの跳躍の高さに応じて着地音の音量を変

更した。音量の調整については、複数の跳躍動作の中で、最も高く飛んだ跳躍の着地音を最大とし、跳躍の高さが小さくなるにつれて音量が小さくなるようにした。跳躍の高さが大きいほど着地したときの速さが大きくなる。着地したときの速さが大きいほど力積が大きくなり、地面から受ける力が大きくなる。地面から受ける力が大きいほど感じる振動刺激が大きいと考えた。

また、座標から速さを求め、速さに応じて風切り音の音量を変更した。跳躍動画の FPS とフレーム数から動画の経過時間を求め、座標と時間を用いて速さの極大値から速さを求めた。複数の跳躍動作の中で、最も速く飛んだ跳躍の風切り音を最大とし、跳躍の速さが小さくなるにつれて音量が小さくなるようにした。地面を蹴ったときの速さが大きいほど力積が大きくなるため、地面を蹴る力が大きくなり、感じる振動刺激が大きいと考えた。

振動刺激も効果音と同様に、離陸時の振動の大きさを速さに応じて、着地時の振動の大きさを高さに応じてそれぞれ変更した。

効果音を発生させるタイミングとしては、y 座標の正を上向きにとったとき、速度が最大になったときに風切り音を発生させて、最小になったときに着地音を発生させた。振動刺激も効果音と同様のタイミングで発生させた。

効果音が出ないところは、環境音を除去して無音とした。次節の実験で用いる効果音のみの条件の動画は、生成した音声信号はと元の跳躍動画の音声と置き換えて生成した。効果音と振動をつける条件では、効果音が L チャンネルで、振動刺激が R チャンネルとする音声ファイルを作成し、元の跳躍動画を結合した。

3. 実験

3.1 実験概要

提案する振動刺激生成が妥当であることおよび振動刺激を加えることで迫力・臨場感が増すことを確かめるために被験者実験を行った。被験者は 20 代から 40 代の男性 3 名、女性 3 名の計 6 名である。

実験で使う動画として跳躍動画に、離陸時のみ、着地時のみ、両方の 3 種類それぞれに効果音のみつけたものと、効果音と振動刺激をつけたものの計 6 条件の刺激を用意した。各条件の効果音と振動刺激の設定を表 1 にまとめる。

動画は被験者の前に置かれた液晶モニターで表示した。効果音はイヤホンを用いて提示した。また、外部の音を遮断するためにイヤマフをつけさせた。振動刺激を提示するパイプレーターにはアクーヴ・ラボの Vp408 を用いる。振動刺激は USB オーディオインターフェース (Roland, Rubix24)、オーディオアンプ (S.M.S.L, SA-36A PRO) を介して出力する。

はじめに効果音も振動刺激もついていない動画を被験者に提示した。その後、6 つの条件をそれぞれ示し、各動画がはじめに示した動画よりも迫力・臨場感を感じるかどうかを 7 段階のリッカート尺度 (0:まったくそう思わない, 1:

表 1: 刺激条件

条件	効果音	振動刺激
A	離地時のみ	なし
B	離地時のみ	あり
C	離地・着地の両方	なし
D	離地・着地の両方	あり
E	着地時のみ	なし
F	着地時のみ	あり

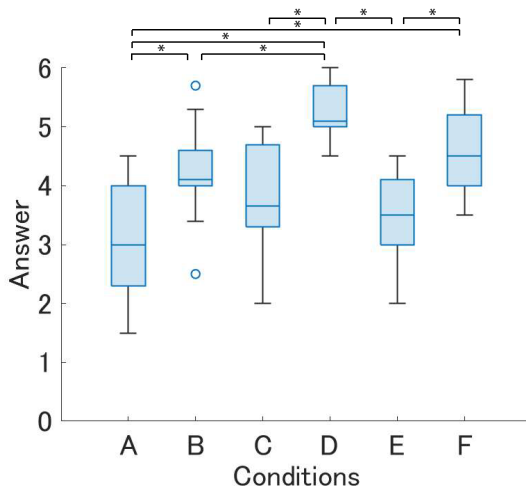


図 3: 無音・無振動の動画よりも迫力・臨場感を感じるかの主観評価

そう思わない, 2:あまりそう思わない, 3:どちらでもない, 4:少しそう思う, 5:そう思う, 6:とてもそう思う)に小数点第1位までつけて回答してもらった。刺激を提示する順番はラテン字格法によりカウンターバランスをとった。この一連の流れを一人当たり3回行った。

3.2 実験結果と考察

被験者の回答結果を図3に示す。回答の中央値を見ると、いずれの条件も基準条件と同程度か基準条件よりも迫力・臨場感を感じる傾向が見られた。最も迫力・臨場感を感じた条件は、効果音が離陸時と着地時の両方ついていて振動刺激がある条件Dであった。この結果から、人の動きに合わせて効果音をつけると迫力・臨場感を感じる事が確認された。

振動刺激がある条件 B, D, F の回答の中央値は、振動刺激がない条件 A, C, E の回答の中央値よりも高く、いずれも少しそう思うより高かった。効果音がそれぞれ同じである条件 A と B, 条件 C と D, 条件 E と F を比較すると、いずれも振動刺激がある条件の方が迫力・臨場感を感じていた。

Steel-Dwass 検定を用いて条件 A~F のすべての群間について多重比較を行った結果、有意水準 0.05 で条件 A と B, A と D, A と F, B と D, C と D, D と E, E と F の

群間に有意差が認められた。図3に有意差が認められたものを*で示した。効果音のみの条件間ではいずれも有意差があるとは言えなかった。振動刺激もついた条件間では B と D でのみ有意差が認められた。また、条件 D に関しては、F との間でのみ有意差が確認されなかった。

これらの結果から、振動刺激を加えた方が動画の迫力・臨場感が増すことが確認された。

効果音のみの条件に着目すると、迫力・臨場感を感じる順が、振動刺激を加えた条件で迫力・臨場感を感じる順と同じであった。この結果は、良い効果音を生成できれば、良い振動刺激を生成できることを示しており、対象とした跳躍動作においては効果音から振動を生成するアプローチの有効性が示された。

4. 結言

映像内の人の跳躍動作に基づいて振動刺激を自動生成する手法を提案した。被験者実験の結果から、跳躍動作に合わせて効果音と振動刺激をつけると映像の迫力・臨場感が増すことが確認された。また、効果音の評価が高い良いほどから振動の評価も高くなることから、効果音を用いて振動体感を生成するアプローチの有効性が確認された。今後、他の運動動作にも提案するアプローチの有効性を検証する予定である。

参考文献

- [1] K. Yamaguchi, M. Konyo, and S. Tadokoro, "Sensory Equivalence Conversion of High-Frequency Vibrotactile Signals using Intensity Segment Modulation Method for Enhancing Audiovisual Experience", 2021 IEEE World Haptics Conference (WHC), pp. 674-679, 2021.
- [2] 星裕也, 昆陽 雅司, 田所 諭, スマートフォンで利用可能な高臨場感振動変換法を用いた体感付き動画の提案, 第27回日本バーチャルリアリティ学会, 2022
- [3] V. Bazarevsky, I. Grishchenko, K. Raveendran, T. Zhu, F. Zhang and M. Grundmann, BlazePose: On-device Real-time Body Pose tracking, [online] Available: <http://arxiv.org/abs/2006.10204>, 2020.
- [4] Google Research, <https://ai.googleblog.com/2020/08/on-device-real-time-body-pose-tracking.html> (閲覧 2023/07/20)
- [5] 効果音辞典, <https://sounddictionary.info/life-1/> (閲覧 2023/07/20)
- [6] OtoLogic, <https://otologic.jp/free/se/gamefighting01.html> (閲覧 2023/07/20)
- [7] S. J. Bolanowski, G. A. Gescheider, R. T. Verrillo and C. M. Checkosky, "Four channels mediate the mechanical aspects of touch", J. Acoust. Soc. Am., vol. 84, pp. 1680-1694, 1988.