



多感覚コンテンツに付加する全身振動情報の 最適振動レベル決定法に関する検討

阿部 翔太¹⁾, 坂本 修一¹⁾, 山高 正烈²⁾, 鈴木 陽一¹⁾, 行場 次朗³⁾

Shota ABE, Shuichi SAKAMOTO, Zhenglie CUI, Yôiti SUZUKI and Jiro GYOBA

1) 東北大学大学院情報科学研究科/電気通信研究所

(〒 980-8577 仙台市青葉区片平 2 丁目 1-1, shota.abe.r4@dc., saka@ais., yoh@riec.tohoku.ac.jp)

2) 愛知工科大学 (〒 443-0047 愛知県蒲郡市西迫町馬乗 50-2, yamataka-masahiro@aut.ac.jp)

3) 尚絅学院大学 (〒 981-1295 宮城県名取市ゆりが丘 4-10-1, j_gyoba@shokei.ac.jp)

概要: これまで、実測振動以外の感覚情報から視聴覚コンテンツに付加する全身振動を生成する方法が提案されてきたが、振動の大きさの決め方は不明であった。本研究では、体験者自身に音から生成された振動の大きさを調整させることで、感覚情報と最適振動量の関係性を検討した。実験の結果、最適振動量と幾つかの音響特徴量との関連が示され、最適な振動の大きさをラウドネス、シャープネス等の音響特徴量で決定できることが示唆された。

キーワード: 多感覚情報, 全身振動情報, 振動生成

1. はじめに

近年の情報科学技術の進歩により、視聴覚情報に加えて体性感覚や嗅覚などより多くの感覚情報を含んだバーチャル・リアリティ (VR) 空間の構築が可能となり、より自然でリアリティのある VR 空間の実現につながっている。その中でも、体全体を揺らすような全身振動情報は空間に関する高次感性知覚と関連が深く [1], 「臨場感」や「迫真性」といった高次感性評価指標を促進させることが報告されている [2]。更に、全身振動情報は、「人と交流している感じ」を促進する効果 [3] や、VR 体験による酔いを低減させる効果 [4] もあることが報告されており、次世代 VR システムの設計には不可欠な感覚情報であると考えられる。

しかし、現状のコンテンツには全身振動情報が含まれていないのが一般的であり、これらにおいて全身振動情報を利用するためには何らかの感覚情報から作り出す必要がある。先行研究において、音と振動の関連性が深いという知見に基づき、音情報から振動情報を生成する方法が提案されている [5][6]。提案法では、音情報の低周波数成分を振動と見立てて、感性評価実験を実施し、生成された振動が実測振動と同程度に高次感性知覚を促進し得ることが示された。また、一人称視点のコンテンツを対象とした研究において、映像情報から振動情報を生成する方法も提案されている [7]。そこでは主に画面の揺れ、人物の動きから振動を生成しており、視聴覚情報と共に生成された振動を提示することでコンテンツ体験の質を向上させ得ることが報告された。

以上のように、感覚情報から振動を生成する方法が提案されつつあるが、これらの方法は物理量を単に振動に変換しているだけで、人間の知覚情報処理過程が考慮されてい

なかった。そのため、生成された振動が体験者にとって理想的な振動となっていたかは明らかではない。更に、有効性が確認できたコンテンツも限定されており、どのようなコンテンツにおいても適用可能な方法であったかも不明である。これらの問題を解決するためには、体験者にとってコンテンツに適した振動とはどのようなものかを調査し、それに適応する振動の生成法を構築する必要があると考えられる。

そこで、本研究は振動の大きさを対象とし、体験者が最適と感じる振動の大きさの調査を行った。具体的には、様々な種類の視聴覚コンテンツに音から生成した振動を基準振動として提示した上で、そのコンテンツに適していると感じる振動の大きさをリアルタイムで調整させた。更に、体験するコンテンツを音のみ、もしくは映像のみといった感覚情報を制限した条件を用意し、音響特徴量や映像特徴量と最適な振動の大きさとの関連性を検討した。

2. 実験の内容

2.1 コンテンツの収集

動画共有サイト "Vimeo[8]" において視聴覚コンテンツ (動画) の収集を行った。限られた条件数で多様なコンテンツを用意するために、以下の流れでコンテンツを収集した。

- Vimeo にて提供されているカテゴリ・サブカテゴリ分けにおいて類似しているものを再度分類。
- 分類されたカテゴリごとに、表 1 の収集条件に合致する動画の中から再生数が多い順に複数の動画を抽出。

再分類により、「非実在」、「日常・自然環境」、「スポーツ」、「パフォーマンス」、「知識・トーク」の 5 つのカテゴリに分類した。このカテゴリから表 1 に示す収集条件を満たす動

表 1: 収集条件

1	クリエイティブ・コモンズ・ライセンス [9] (表示, 継承, 非商用のどれか) を示している.
2	撮影場所に存在する環境音以外の音を付加していない, もしくは無音でない.
3	一分以内の動画である.

画を抽出し, 最終的に 20 種類のコンテンツを得た.

2.2 実験刺激

インターネット上から収集した 20 種類のコンテンツを以下のような処理を施し統制した. 映像情報については, 解像度が 1920×1080 pixel, フレームレートが 30 fps となるように再エンコードを行った. 音情報については, 収集された動画がインターネット上にアップロードされたものであり撮影場所の実環境音の大きさが不明であるため, 全コンテンツについて同じ A 特性音圧レベル (75 dB) になるよう調整を行った. 振動情報については, 元動画の音に対してモノラル変換を行った後に, 遮断周波数 70 Hz の低域通過フィルタを適用し, その信号を振動振幅波形とみなした. その上で, 全時間の平均振動レベルが 60 dB になるようにコンテンツ毎に調整した [6]. 以降, この振動を先行研究 [6] にならって ViLA (Vibration from Low-Frequency Audio) と表記する.

2.3 実験環境

実験環境の概要を図 1 に示す. 実験は東北大学電気通信研究所本館内に設置され, 防音, 吸音処理もされたシールド暗室にて行った. 収録素材の視覚刺激 (解像度: 1920×1080 pixel, フレームレート: 30 fps), 聴覚刺激 (サンプリング周波数: 48 kHz, 量子化ビット数: 16 bit), および全身振動となる振動刺激 (サンプリング周波数: 8 kHz, 量子化ビット数: 16 bit, 上下方向) を, それぞれ DLP プロジェクタ (SANYO, PDG-DHT100JL), 密閉型ヘッドフォン (SENNHEISER, HDA-200), モーションプラットフォーム (D-BOX, MASTERING MOTION) から提示した. 被験者の立ち位置からスクリーンまでの距離は 2.5 m であり, 映像提示画角は被験者からおよそ 90 deg (水平方向) \times 50 deg (垂直方向) であった. 被験者の右側にスロットルレバー型コントローラ (SAITEK, Throttle Quadrant) を設置した. このコントローラは上下方向のみに滑らかに可動し, レバー位置を保持することもできる. 可動範囲は 0~90 deg, 最小可動角度はおおよそ 0.2 deg で, 1 s 間に 10 個のデータを取得可能であった.

2.4 実験条件, 被験者

被験者は正常な視覚 (矯正を含む) と聴覚を有する大学生, および大学院生 27 名 (男性 22 名, 女性 7 名, 平均年齢 22.4 ± 2.2 歳) であった.

本実験のコンテンツの数は 20 個であり, それぞれに 3 つの感覚条件 (Visual & Sound, Visual Only, Sound Only) が存在するため計 60 条件となった. 実験時間の長さや, 一度体験したコンテンツの視聴覚情報を被験者が記憶するこ

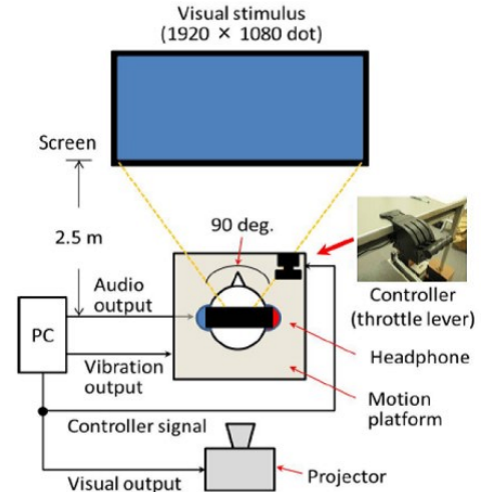


図 1: 実験環境の概要

とを考慮し, 一人の被験者はひとつのコンテンツあたり一つの感覚条件のみを体験し, かつ体験する感覚条件の数が偏らないように調整を行った. その結果, 全ての条件において 9 人の被験者が割り当てられた.

2.5 実験の流れ

実験の流れは以下のとおりである. 初めに被験者の正面方向となるスクリーン位置に十字の注視点が表示され, あるコンテンツの実験刺激が提示された. 被験者は, この間手元のレバーを操作して, 「そのコンテンツ (場面) に適していると感じる振動の大きさ」をリアルタイムで調整した. この際, 調整された振動の大きさの時間変化は随時保存された. コンテンツの再生が終わった段階で「次に進む」, 「もう一度調整する」の 2 択のどちらかを選択することができ, 被験者自身が最適な振動だと判断するまで振動の大きさの調整を行わせた. なお, 「もう一度調整する」を選択した場合は前の試行で調整された振動が基準振動として提示された.

2.6 調整された角度と提示される振動波形の関係

レバーの調整により変化する振動レベルの上昇量 $R_n(t)$ [dB] は以下のように記述できる.

$$R_n(t) = R_{n-1}(t) + I \quad (1)$$

$$R_0(t) = 0 \quad (2)$$

ここで, t は経過時間 [s], n は試行回数である. 振動を再生するハードウェアの制約上, $0 \leq R_n(t) \leq 40$ となる. I はレバーによって設定された変化量 [dB] である. 各試行の始まりにおいて, レバーは基準位置である 45 deg. に位置している. レバーが基準位置にある場合は $I = 0$ であり, そこからレバーを上へ調整すると I が増加し, 下へ調整すると I が減少する. 調整を経るごとにより細かく振動を調整できるようにするため, レバーの角度と振動量の幅は変化可能にした. 具体的には, 1~2 回目の調整においては, ± 1 dB/deg., 3~4 回目の調整では, ± 0.5 dB/deg., 5 回目以降は, ± 0.25 dB/deg. とした. なお, レバーの調整によ

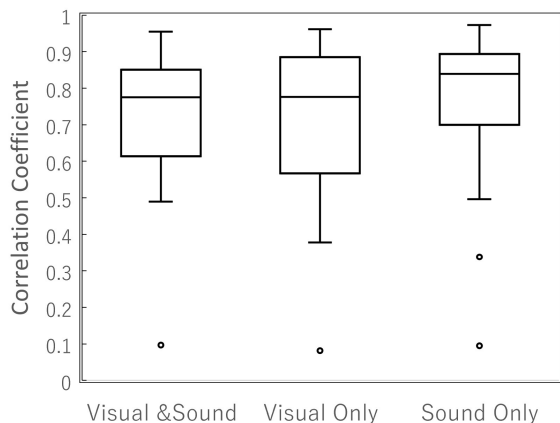


図 2: コンテンツ毎に算出した ViLA の振動レベルの時間変化と、調整後の最適振動レベルの時間変化との相関係数

り $R_n(t)$ が 0 を下回った場合においては $R_n(t) = 0$ とするようにした。

3. 実験結果

3.1 ViLA と最適振動レベルの関係

コンテンツ毎に調整後の振動波形（以降，“最適振動”と表記）における振動レベルの時間変化が調整前の振動である ViLA（音に低域通過フィルタを適用して生成した振動）の振動レベルの時間変化とどの程度異なっているのかについて検討した。各コンテンツにおいて、全被験者で得られた最適振動レベルの時間変化を平均し、その波形と ViLa の振動レベルの時間変化との相関を求めた結果を図 2 に示した。図 2 から、どの感覚条件においてもその中央値が 0.7 を超えており、半数以上のコンテンツにおいて最適振動レベルの時間変化が ViLA の振動レベルの時間変化と高い相関を示していることが分かる。特に Sound Only 条件においては、他の感覚条件よりもその中央値が高いことが伺える。

3.2 感覚条件と最適振動レベルの関係

Visual & Sound 条件の最適振動レベルと Visual Only 条件、もしくは Sound Only 条件の最適振動レベルの類似性について検討した。各コンテンツにおいて、全被験者で得られた最適振動レベルを 0.1 s 毎に平均を行った後、縦軸を Visual & Sound 条件、横軸をその他の感覚条件で得られた振動の最適振動レベルとして散布図にプロットをした（図 3）。それぞれの感覚条件間の相関係数を計算したところ、Visual & Sound 条件-Visual Only 条件間の相関係数は 0.80、Visual & Sound 条件-Sound Only 条件間の相関係数は 0.88 という結果になり、Visual & Sound 条件-Sound Only 条件間の方が相関が大きいという結果になった。

3.3 音響特微量と最適振動レベルの関係

全てのコンテンツに対して、Sound Only 条件の 0.1 s 毎の最適振動レベルを目的変数、下記に記す音響特微量を説明変数として重回帰分析を行い、音響特微量と最適振動レベルの関係性を検討した。説明変数として用いた音響特微量は、ラウドネス [10]、ラフネス [11]、シャープネス [12]、スペクトルフラックス [13] の 4 つとし、それぞれ 0.1 秒毎

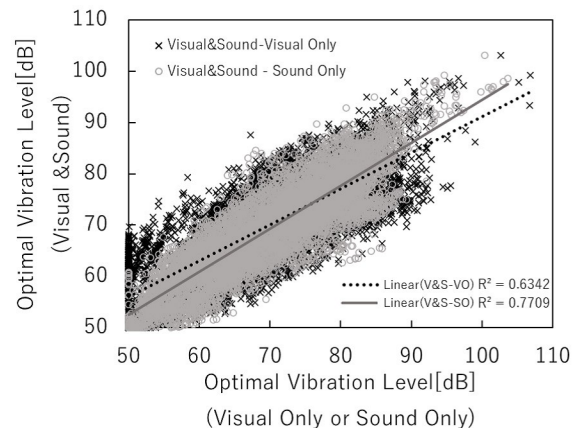


図 3: Visual & Sound 条件と Visual Only 条件、Sound Only 条件の最適振動レベルの関係

に導出した。なお、図 2 から ViLA と最適振動レベルに高い相関関係があったため、ViLA の振動レベルも説明変数に加えた。その結果を表 2 に示す。得られた標準偏回帰係数について分析したところ、ラウドネス ($\beta = 0.19$)、ラフネス ($\beta = 0.04$)、シャープネス ($\beta = -0.11$) の影響が有意であった ($p < .01$)。

得られた重回帰式を Visual & Sound 条件、Visual Only 条件の 0.1 s 毎の最適振動レベルに適用して得られた決定係数を表 3 に示す。表 3 には比較として、ViLA の振動レベルのみで単回帰分析を行った際の決定係数も示す。表 3 をみると、Visual & Sound 条件においても、Sound Only 条件と同程度に使用したパラメータを用いて説明できることがみとれる。

4. 考察

図 2 から、特に Sound Only 条件において最適振動の振動レベル時間変化と ViLA の振動レベル時間変化の相関が非常に高かった。本実験において調整前振動として用いた ViLA は音に低域通過フィルタを適用することで生成しているため、そもそも音情報と深い関連性を持った振動であった。そのため、音情報のみしか判断基準がない Sound Only 条件においては、音と何かしらの関連を持った ViLA の振動レベルの時間変化自体が最適な時間変化とみなせていた可能性が考えられる。

図 3 において、Visual & Sound 条件-Visual Only 条件間よりも Visual & Sound 条件-Sound Only 条件間の方が重回帰直線の当てはまりが良いという結果になった。このことは、Visual & Sound 条件の振動の大きさの調整は Visual Only 条件の調整よりも Sound Only 条件の調整と似通っていたことを示しており、Visual & Sound 条件の被験者は映像情報よりも音情報に重きをおいて振動の大きさを調整したことを示唆している。そのため、Sound Only 条件の判断基準の解明は、Visual & Sound 条件の判断基準の解明に有効だと考えられる。

そこで、Sound Only 条件の最適振動レベルと音響特微量の関係を表 2 の重回帰分析の結果から考察する。表 2 を

表 2: 重回帰分析の結果 (標準偏回帰係数)

説明変数	目的変数
	Sound Only 条件の 調整後の最適振動レベル
ラウドネス	0.19*
ラフネス	0.04*
シャープネス	-0.11*
スペクトルフラックス	-0.01
ViLA の振動レベル	0.68*
決定係数 R^2	0.60*

*: $p < .01$

表 3: 重回帰式, 単回帰式の感覚条件への適応

	Visual & Sound	Visual Only	Sound Only
重回帰式の 決定係数 †	0.57	0.47	0.60
単回帰式の 決定係数 ‡	0.53	0.46	0.56

† 表 2 で得られた重回帰式による線形回帰

‡ ViLA の振動レベルによる線形回帰

見ると, ViLA の振動レベル以外ではラウドネスが最も最適振動レベルに影響を与えていることが分かる。一般的に, 日常生活において物体の衝突や移動等で大きな振動が発生する場面においては大きな音も発生する場面が多い, そのため, 体験者は大きな音に対して大きな振動を期待し, 振動を大きく調整したとと考えられる。

また, シャープネス, ラフネスという音響特徴量も最適振動レベルに影響を与えた。シャープネスは「音の甲高さ」に対応した尺度であり, この特徴量が負の標準偏回帰係数を持つことは, 低い音が鳴っている場面に振動を大きく調整する傾向があったことを示す。“音の迫力”や“音の重さ”は周波数の低さと関連する [14] ことから, 被験者は音の低さから迫力を感じることで振動を大きく調整したと考えられる。一方, ラフネスは音の振幅エンベロープの変調度に関連した指標であり, 主に物体同士が擦れたり, 物体を引きずったりする場面で大きく評価される。このことから, 音の粗さが大きい場面において被験者に物体の運動が想起され振動を大きく調整することにつながったと考えられる。

表 3 から, Visual & Sound 条件においても Sound Only 条件と同程度に使用したパラメータを用いて説明できることがみてとれる。このことから, ラウドネスやシャープネス, ラフネスといった音響特徴量は, Sound Only 条件だけでなく Visual & Sound 条件での最適振動レベルの調整にも寄与していたと考えられる。

5. 結論

本研究では振動レベルを対象とし, 体験者が最適と感じる振動レベルの調査を行った。その結果, 最適と感じる振動レベルの時間変化は従来の生成法である ViLA の振動レ

ベルの時間変化と類似していた。更に, 最適な振動の大きさをラウドネス, シャープネス等の音響特徴量で決定できる可能性が示唆された。

謝辞 本研究の一部は, JSPS 科研費 19H04145, および東北大学電気通信研究所共同プロジェクト研究の助成を受けたものである。

参考文献

- [1] 寺本 渉ら: 臨場感の素朴な理解, TVRSJ, Vol. 15, No. 1, pp. 7-16, 2010.
- [2] Shuichi Sakamoto et al. : Body vibration effects on perceived reality with multi-modal contents, MTA, Vol. 2, No. 1, pp. 46-50, 2014.
- [3] Myungho Lee et al. : Exploring the Effect of Vibrotactile Feedback through the Floor on Social Presence in an Immersive Virtual Environment, Proc of IEEE VR, pp. 105-111, 2017.
- [4] Yuki Sawada et al. : Effects of synchronised engine sound and vibration presentation on visually induced motion sickness, Sci Rep, Vol. 10, No. 1, pp. 1-10, 2020
- [5] S. Merchel et al. : The influence of vibration on musical experience, AES, Vol. 62, No. 4, pp. 230-234, 2014.
- [6] 崔正烈ら: 多感覚コンテンツの音情報から生成した床振動の高次感性促進効果, JIP, Vol. 59, No. 11, 2018.
- [7] Gyeong Yun et al. : Improving Viewing Experiences of First-Person Shooter Gameplays with Automatically-Generated Motion Effects, Proc of CHI '21, pp. 1-14, 2021
- [8] <https://vimeo.com/jp>
- [9] <https://creativecommons.jp/>
- [10] Brian C. J. Moore et al. : Testing and refining a loudness model for time-varying sounds incorporating binaural inhibition, JASA, Vol. 143, No. 3, pp. 1504-1513, 2018
- [11] H. Fastl et al. : Psychoacoustics: Facts and Models, 3rd ed. (Springer, Berlin, Heidelberg), pp. 257-264., 2006.
- [12] DIN 45692: Measurement Technique for the Simulation of the Auditory Sensation of Sharpness, German Institute for Standardization, 2009.
- [13] Scheirer, E. et al. : Construction and Evaluation of a Robust Multifeature Speech/Music Discriminator, Proc of ICASSP. Vol. 2, pp. 1331-1334, 1997.
- [14] Solomon et al. : Semantic Approach to the Perception of Complex Sounds, and Research for Physical Correlates to Psychological Dimensions of Sounds, JASA, Vol. 30, pp. 421-497, 1958.