

修士論文

2017 年度

音源に同期する運指および表情に注目した 吹奏アニメーションの自動生成

堀井 絵里
(学籍番号：81621728)

指導教員 教授 藤代 一成

2018 年 3 月

慶應義塾大学大学院理工学研究科
開放環境科学専攻

論文要旨

音楽演奏を題材としたアニメーションは多く存在する。そこでは、実際に演奏する演奏者の身体の動きに近い演奏シーンが生成されている。しかし、楽器を演奏するキャラクタの運指に注目すると、音楽と指の動きが完全に同期されていないことがあり、強い違和感を感じる。また、完全に同期されているアニメーションを作成するには、多くの労力が必要となる。そこで本研究では、管楽器を演奏するキャラクタの吹奏アニメーションを、音楽に容易に同期させるため、音源から自動的に運指のアニメーションを生成することを目指す。より具体的には、ウインドシンセサイザを用いて作成した MIDI 音源を解析することにより、運指が音源に同期した、自然な吹奏アニメーションを実現する。

近年、コンピュータの処理速度が向上したことで、高品質な画像の高速な描画が可能となった。この影響により、ゲームや映像をユーザが完成版に近い品質で編集できる技術が発達している。このような CG 技術の台頭や IT の発達で、最新のコンピュータ上に仮想の社会や人間をつくる技術が研究者や CG 関係者の注目を集めている。しかし人間のパーツの多くは目や肌、髪の毛のようにやわらかい、小さい、半透明といった特徴をもつパーツで構成され、コンピュータには簡単に描画できないものが多い。特に頭髪は、人間にとって重要なパーツでありながら、その複雑さが原因で高速な描画が難しい物体として知られている。また頭髪状物体が落とす影も同様に、一般的な描画手法を使うとちらつきやありえない濃度の影が生じるため、計算が難しい描画対象として知られている。

頭髪は半透明で、かつ細長い円筒の集合であり、このような幾何形状をもつ物体は頭髪の形状、位置や視点、照明に変化がある場合、影を描画する際にちらつきが生じる。このちらつきのうち、頭髪が半透明で遮蔽判定が難しいことによるちらつきは、影を描く場所に光が届くまでに通過する頭髪密度の積算値を使い影を描くことで解決できる。一方、髪が細すぎてコンピュータが髪を認識できないことによるちらつきは、投影後にちらつきが生じている部分と周囲の色を平均化して抑制されてきた。しかしこの平均化処理は、結果の品質が投影後の単位面積あたりの画素数に依存してしまう。

そこで本論文では、頭髪の密度を積算する既存手法を拡張して、大きさが可変で不透明度分布をもつ正方形小板で頭髪の三次元形状を近似することにより、結果の品質が画素数に依存しないアンチエイリアシング手法を提案する。これに加えて既存研究では積極的に行われてこなかった、出力画像の定量的な品質評価をするために新たな評価方法を作成した。この評価方法を使用して既存手法と提案手法の出力画像の品質を比較した結果、提案手法は計算速度を大きく減少させることなく効果的にちらつきを抑制できることを確認した。

キーワード

アニメーション, 吹奏楽, 金管楽器, 自動生成.

Thesis Abstract

An Object-Space Anti-Aliasing Method for High-Quality Shadowing of Hair-Shaped Objects

Recent improvement in CPU and GPU performance made it possible to render high-quality computer graphics in real time. Accordingly, technologies to edit video games or movies with nearly-final quality previews have been developed. With support of such rapid growth of CG technology as well as other IT developments, it has begun attracting much attention from related researchers and practitioners to construct realistic and dynamic virtual societies and humans on a modern computer. However, making virtual humans is a very difficult task because they are made up of many soft, small or transparent parts such as skin, **hairs** or eyes. Especially, human hairs are well-known as intricate objects which require much executing time to render and cause **aliasing**, even though they are essential parts of humans. Shadows cast from hairs are also known as difficult objects to render because they generate aliasing and too far dense shadows with commonly-used shadowing methods.

Hairs are **semi-transparent** and can be presented as mass of thin cylinders, but this style of geometric modeling causes aliasing to rendered shadows when users update a shape or a position of hair, camera angle and/or the position of lights. As for a type of aliasing caused by difficulty to detect occlusion due to the transparency of hairs, even previous shadowing methods can decrease it by calculating the accumulated **amount of density values of all hairs** that light goes through. On the other hand, aliasing caused by the problem that a computer cannot recognize too-thin hairs could be alleviated by image-space post-processing such as neighboring pixel averaging, though the quality depends strongly on the density of projected pixels.

Therefore, this thesis proposes an object-space method which realizes resolution-independent **anti-aliasing**. The method is an extension of the previous shadowing method that calculates the accumulated amount of density values of hairs, and **approximates a three-dimensional shape of hair** as a cloud of variable-size square splats with concentric opacity distribution. In this study, a novel evaluation method was also proposed to compare the extent of antialiasing in a quantitative manner. Comparison based on the novel evaluation method revealed that the proposed method outperforms the existing image-space anti-aliasing with minimum loss of computational speed.

Keywords

Hair; real-time rendering; anti-aliasing; object space; semi-transparent objects.

目次

1	序論	1
1.1	研究背景	2
1.2	研究目的	2
1.3	本論文の構成	2
2	関連研究	3
2.1	音からアニメーションを自動生成する研究	4
2.2	アニメーションから音を生成する研究	4
2.2.1	サウンドレンダリング	4
2.2.2	プロシージャルオーディオ	5
2.3	音とアニメーションを同期させる研究	5
2.4	マーチングを行う演奏者の軌跡を解析する研究	5
2.5	本研究の新規性	6
3	提案手法	7
3.1	自動生成の流れ	8
3.2	入力	8
3.3	出力	9
3.4	デバイス	9
3.5	ソフトウェア	9
3.6	3D モデル	9
3.7	MIDI データの解析	9
3.7.1	楽譜データへの変換	9
3.7.2	モーションへの適用	9
4	結果と分析	10
4.1	アニメーション生成結果	11
4.2	評価	11
4.2.1	実際の演奏シーンとの比較による評価	11
4.2.2	既存のアニメーションとの比較による評価	11
4.2.3	アンサンブルアニメーションの評価	11
5	結論	12
5.1	まとめ	13
5.2	今後の課題	13

5.2.1	表情の豊かさの向上	13
5.2.2	モーションの質と種類の向上	13
5.2.3	モーションと音の関連付け	13
5.2.4	楽器の種類および演奏者の増加	13
	謝辞	14
	公開文献	15
	参考文献	16

目 次

2.1	複数の剛体がぶつかり破壊される様子	5
2.2	マーチングを行う演奏者の軌跡および誤差の表示例	6
3.1	音源から吹奏アニメーションを自動生成する流れ	8

表 目 次

第 1 章

序論

本章では、本研究の研究背景と研究目的について説明する。1.1 節で研究背景、1.2 節で研究目的、1.3 節で本論文の構成について述べる。

1.1 研究背景

音楽演奏を題材としたアニメーションは多く存在する。テレビで放映されたアニメーションの例を挙げると、『のだめカンタービレ』、『けいおん!』、『響け! ユーフォニアム』が相当する。これらのアニメーションは、セル画であったり 3DCG であったり、製作方法がさまざまであるが、いずれも実際に演奏する演奏者の身体の動きに近い演奏シーンが生成されている。楽器の輝きや形状なども忠実に再現されており、評判が高い。

一方、楽器を演奏するキャラクタの運指や身体の動きに注目すると、音楽と動きが完全に同期されていないことがあり、違和感を感じる。特に速いフレーズや、複雑なリズムを演奏するシーンで、このようなアーティファクトが起きやすい。また、表情からも不自然さを感じることもある。しかし、1 つ 1 つの音と動きや表情を合わせるには、多くの労力が必要となる。

1.2 研究目的

1.1 節で述べた課題を解消するため、本研究では、管楽器を演奏するキャラクタの吹奏アニメーションを、音源から自動的に生成することを目指す。より具体的には、実際の演奏アニメーション制作フローに沿わせるため、楽曲は楽器（ウインドシンセサイザ）を用いて、MIDI 音源として生成する。そして、作成した楽曲を解析することにより、吹奏の情報を得る。得られた情報をキャラクタの身体の動きや表情、そして管楽器に適用することにより、音源に同期した自然な吹奏アニメーションを実現する。

1.3 本論文の構成

本論文は次章以降、以下のように構成される。

次章では関連研究を説明する。第 3 章では本手法の仕組みを説明する。そして、第 4 章では自動生成結果を述べ、この結果に対する評価をまとめる。最後に第 5 章で結論を述べ、今後の課題に言及する。

なお、本研究の成果は Visual Computing/グラフィクスと CAD 合同シンポジウム 2017 にてポスター発表 [iii] を行った。そして、映像表現・芸術科学フォーラム 2018（2018 年 3 月）、Cyberworlds（2018 年 8 月）において発表を行う予定である。

第 2 章

関連研究

本章では 2.1 節で音からアニメーションを自動生成する研究を紹介し、?? 節でアニメーションから音を自動生成する研究を紹介する。2.3 節では音とアニメーションを同期させる研究、2.4 節で本論文と同様、吹奏楽に関連した研究を紹介し、最後に 2.5 節にて本研究の新規性を述べる。

2.1 音からアニメーションを自動生成する研究

音とアニメーションを同期させることを目的として、音からアニメーションを自動生成する研究は多く存在している。

物が落下するアニメーションを生成するには、物の落下音と物が落下するタイミングを 1 つ 1 つ合わせる必要がある。Langlois ら [1] は、それらを正確に合わせるために、物が落下する音が収録されている音源から、物が落下するアニメーションを自動生成する手法を提案した。

キャラクタの口の動きと音声を合わせる、リップシンクも難しい課題となっている。キャラクタの口の動きに合わせて後から音声を録音するアフレコは比較的容易であるが、音声から、その言葉を発している口元のアニメーションの生成は、言葉と口の形、そして発するタイミングと口の動きを合わせる必要があり、ひじょうに難易度が高い。そこで、Edwards ら [2] は、台詞が収録されている音源および台詞が記載されているテキストファイルを入力すると、その台詞を話している口元のアニメーションを自動生成する手法を提案した。本手法では、顎と唇のみを制御することにより、自然な口元のアニメーションを自動生成している。更に、後から表情を編集することが可能となっているため、最終的には自然な表情のアニメーションが完成する。

本論文の目的と同じように、音源から楽器を演奏するアニメーションを自動生成する研究も多く行われている。Zhu ら [3] は、MIDI 音源からピアノを弾く指元のアニメーションを自動生成する手法を提案した。ピアノのような楽器は、指と音が 1 体 1 で対応していないため、フレーズにより指使いを変える必要がある。ある指が他の指の上や下を通るような指使いの時は、指同士が衝突しないように考慮する必要もある。彼らの手法では、これらの課題を解消することが可能である。

2.2 アニメーションから音を生成する研究

音とアニメーションを同期させることを目的として、物理シミュレーションの結果を音に反映させる研究分野が存在する。

2.2.1 サウンドレンダリング

サウンドレンダリングという分野は、物理アニメーションを生成すると同時に音を生成することにより、双方を同期させることを目的とする分野である。CG-ARTS の教育リポート、『音を描き出す夢の実現』[11]によると、本分野は 1990 年代前半に登場し、未だ実験的にとどまっている分野となっている。2009 年には 1 つのシミュレーションパイプラインの中で、CG アニメーションと、それに呼応した音を同時に生成するアルゴリズム??が考案された。本アルゴリズムは Zheng らにより考案され、水が流れる様子を再現したアニメーションと音の見事な同期が達成された。翌年には、同氏らにより破壊音を自動生成するアルゴリズム [5] が発表された。本アルゴリズムでは、さまざまな

剛体の破壊音だけでなく、その衝撃が間接的におよぼす影響も考慮されているため、図2.1のような複数の物体がぶつかり、破壊されるような複雑なシーンでも、忠実に音が再現された。

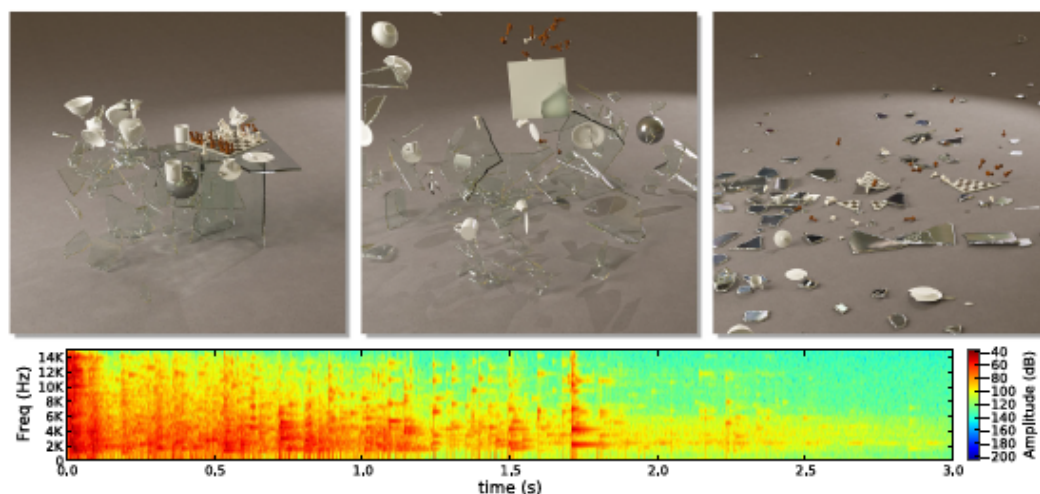


図 2.1 複数の剛体がぶつかり破壊される様子

2.2.2 プロシージャルオーディオ

アニメーションから音を自動生成する技術は、ゲームのような、使用可能なメモリが限られている場合には、できる限りメモリを節約したい。ゲームソフトを開発している株式会社スクウェア・エニックスは、メモリを節約するため、プロシージャルという技術を研究している。この技術は2014年のCEDEC[12]にて発表されており、あらかじめ作っておいた音をゲーム内で使用するのではなく、音をプログラムで作り出す技術をさす。ゲームでは、プレイヤーの操作に合わせて音を鳴らすタイミングを内部で調節する必要がある。そのことから、この技術はゲームにおいて特に必要な技術であると考えられる。

2.3 音とアニメーションを同期させる研究

片方からもう一方を自動生成する研究だけでなく、既に生成されている双方を編集することにより同期させる研究も存在する。Leeら[6]は、MIDI音源およびキャラクタのモーションデータを入力とし、双方を最小限編集することにより、MIDI音源と同期したキャラクタのダンスアニメーションを生成する手法を提案した。

このような研究は、複数のカットを組み合わせて作成する、モンタージュの作成にも応用されている。Liaoら[7]は、BGMとなる音源および複数のビデオクリップを入力することにより、BGMに合ったモンタージュを作成する手法を提案した。

2.4 マーチングを行う演奏者の軌跡を解析する研究

マーチングとは、演奏者が演奏しながら定められた経路に従って歩くことにより隊形を作る、吹奏楽の演奏形態である。本論文の筆者は、マーチングを行う1人の演奏者に注目し、その演奏者が歩く

べき経路と実際の経路の誤差を表示するシステム図2.2を提案した。本システムにより、演奏者は自身の歩行の改善点を知ることができ、その結果マーチングの隊形の完成度を高めることが期待できる。なお、本研究の成果は NICOGRAPH2016 にてポスター発表 [?] を行った。そして、一連の研究

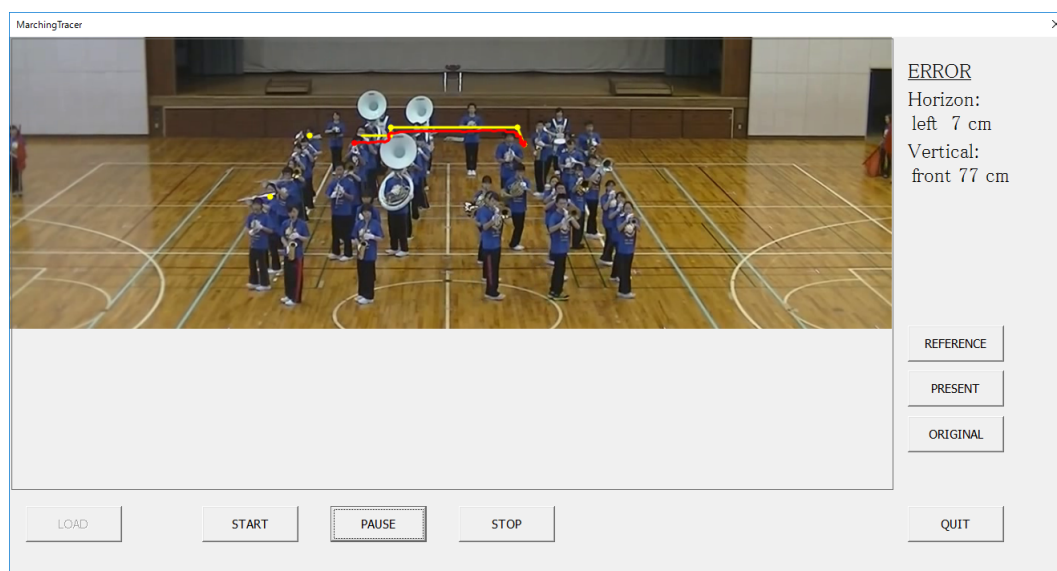


図 2.2 マーチングを行う演奏者の軌跡および誤差の表示例

をまとめた成果が画像電子学会誌の Vol.47[ii] に掲載される予定である。

2.5 本研究の新規性

アニメーションと音の同期や、吹奏楽をテーマとした研究は多く存在する。しかし、管楽器を演奏するキャラクタと音の同期に注目した研究は存在しない。したがって本研究の新規性は、管楽器を演奏するアニメーションを、音から自動生成するという点である。さらに、本論文ではトランペット奏者およびトロンボーン奏者に適用した例のみを示すが、音と指使いの対応表を用意することにより、すべての管楽器の演奏アニメーションが入力音源から自動生成が可能である。このように、あらゆる楽器への対応が可能である点も新規性である。

第 3 章

提案手法

本章では提案手法を詳しく説明する．まず，??節にて自動生成の大まかな流れを説明し，その後にそれぞれの工程について詳しく述べる．??節および??節にて入出力について説明し，??節にて使用するデバイス，??節にて使用するソフトウェア，??節にて使用する 3D モデルについて述べる．??節にて MIDI データの解析からモーションへの適用方法を説明する．

3.1 自動生成の流れ

図 3.1 に，音源から吹奏アニメーションを自動生成する流れを示す．実際のアニメーション制作フ

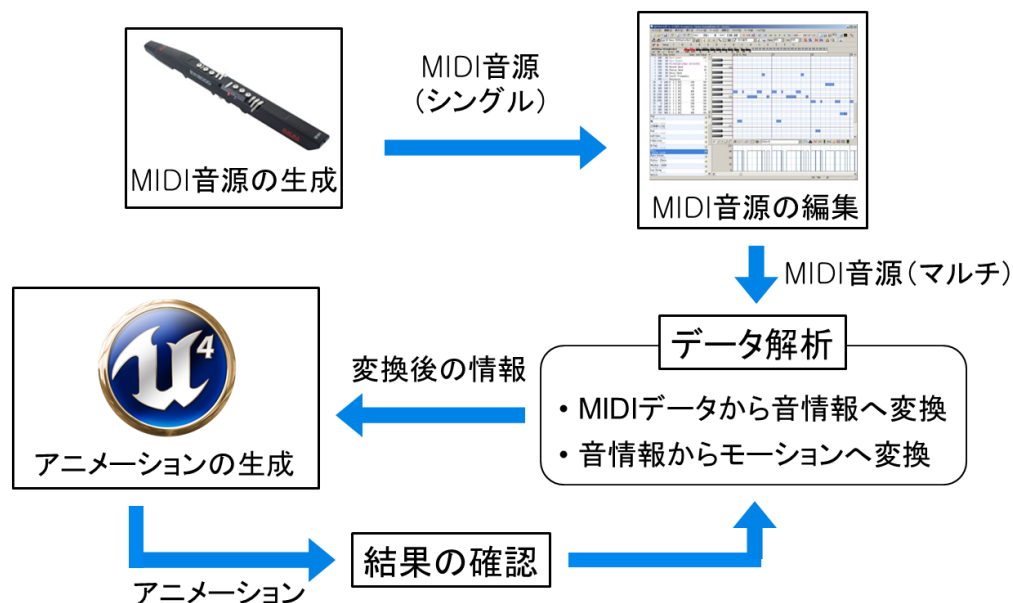


図 3.1 音源から吹奏アニメーションを自動生成する流れ

ローに沿わせるため，音源の生成は楽器を用いて行う．次に，生成した音源を解析し，譜面データへ変換する．そして，アニメーション生成と同時に音源を流すことにより，音源に合わせてキャラクターが動くアニメーションが完成となる．

3.2 入力

入力する音源は，MIDI 音源とする．ここで，MIDI 音源は，MIDI という信号を受信して発音する音源のことである．よく使用される mp3 や wave などの形式とは異なり，中身が譜面データとなっているため，音の解析が比較的容易である．

3.3 出力

3.4 デバイス

3.5 ソフトウェア

3.6 3D モデル

3.7 MIDI データの解析

3.7.1 楽譜データへの変換

3.7.2 モーションへの適用

第 4 章

結果と分析

4.1 アニメーション生成結果

4.2 評価

4.2.1 実際の演奏シーンとの比較による評価

4.2.2 既存のアニメーションとの比較による評価

4.2.3 アンサンブルアニメーションの評価

第 5 章

結論

本章では本論文の結論を述べ，今後の課題に言及する．

5.1 まとめ

5.2 今後の課題

5.2.1 表情の豊かさの向上

5.2.2 モーションの質と種類の向上

5.2.3 モーションと音の関連付け

5.2.4 楽器の種類および演奏者の増加

謝辞

本論文の執筆にあたり、ひじょうに多くのアドバイスを頂いたばかりでなく、私からの様々な相談もお聞きいただいた慶應義塾大学理工学部情報工学科の藤代 一成教授に深く感謝いたします。私の至らなさから様々なご負担やご迷惑をおかけしたにも関わらず、辛抱強くご指導いただきありがとうございました。今後いかなる場で働く際にも、先生のご指導を思い出しながら仕事に励んでいきたいと思えます。

本研究に取り組むにあたり、実装面において多くのアドバイスをいただいただけでなく、コンピュータグラフィクス分野における様々な課題に出会う機会をいただいた、現在は株式会社 DeNA に所属しておられる松岡 徹氏、株式会社デジタル・フロンティア開発部の福田 啓氏、斎藤 弘氏、高山 耕平氏並びに開発部の皆様に深く感謝いたします。

研究を進めるにあたり時には楽しく話し合い、時には真面目に議論を行い、共に助け合った研究室の皆様に深い感謝をいたします。皆様の存在が無ければ、私の研究はここまで進めることができませんでした。

研究をするにあたり、科研費基盤研究 (B)25280037 及び慶應義塾大学博士課程教育リーディングプログラム オールラウンド型「超成熟社会発展のサイエンス」に多大なるご支援をいただいたことを感謝しております。

最後に、研究を進めるにあたりお世話になりました、すべての方に感謝を申し上げたいと思います。

公開文献

- [i] 堀井 絵里, 藤代 一成, 「マーチングバンドにおける演奏者個人を対象とした移動経路の計量可視化」, 芸術科学会 NICOGRAPH2016 予稿集, pp.119-120, 2016 年 11 月.
- [ii] 堀井 絵里, 藤代 一成, 「マーチングバンドの演奏者個人に注目した移動経路の計量可視化」, 画像電子学会誌, Vol.47, 未決定, 2017 年 12 月.
- [iii] 堀井 絵里, 藤代 一成, 「音源に同期する運指に注目した吹奏アニメーションの自動生成」, Visual Computing/グラフィクスと CAD 合同シンポジウム 2017 DVD 予稿集, No.41, 一橋講堂, 2017 年 6 月.

参考文献

- [1] T. R. Langlois and D. L. James, “Inverse-Foley Animation: Synchronizing Rigid-body Motions to Sound,” *ACM Transactions on Graphics*, Vol. 33, No. 4, Article No. 44, pp. 44:1-44:11, July 2014.
- [2] P. Edwards, C. Landreth, E. Fiume, and K. Singh, “JALI An Animator-Centric Viseme Model for Expressive Lip Synchronization,” *ACM Transactions on Graphics*, Vol. 35, No. 4, pp. 127:1-127:11, July 2016.
- [3] Y. Zhu, A. S. Ramakrishnan, B. Hamann, and M. Neff, “A System for Automatic Animation of Piano Performances,” in *Proceedings of Computer Animation and Virtual Worlds*, Vol. 24, No. 5, pp. 445-457, September 2012.
- [4] C. Zheng and D. L. James, “Harmonic Fluids,” *ACM Transactions on Graphics*, Vol. 28, No. 3, pp. 37:1-37:12, July 2009.
- [5] C. Zheng and D. L. James, “Rigid-body Fracture Sound with Precomputed Soundbanks,” *ACM Transactions on Graphics*, Vol. 29, No. 4, pp. 69:1-69:13, July 2010.
- [6] H. C. Lee and I. K. Lee, “Automatic Synchronization of Background Music and Motion in Computer Animation,” *Computer Graphics Forum*, Vol. 24, No. 3, pp. 353-361, September 2005.
- [7] Z. Liao, Y. Yu, B. Gong, and L. Cheng, “Audeosynth: Music-Driven Video Montage,” *ACM Transactions on Graphics*, Vol. 34, No. 4, pp. 68:1-68:10, July 2015.
- [8] <http://takabosoft.com/domino> (最終アクセス日 : 2018/02/05).
- [9] <http://tf3dm.com/> (最終アクセス日 : 2018/02/05).
- [10] <https://jp.3dexport.com/> (最終アクセス日 : 2018/02/05).
- [11] CG-ARTS, “音を描き出す夢の実現,” 最終アクセス : 2018 年 2 月 5 日. [Online]https://www.cgarts.or.jp/report/rep_kr/rep0901.html
- [12] 4Gamer.net, “ [CEDEC 2014] 物理シミュレーションの結果に合わせて音を生成する方法とは。スクエニの開発者が語るプロシージャルオーディオの応用例,” 最終アクセス : 2018 年 2 月 5 日. [Online]<http://www.4gamer.net/games/032/G003263/20140903036/>