

平成 28 年 1 月 22 日

慶應義塾大学 理工学研究科

開放環境科学専攻 情報工学専修 御中

## 修士論文発表における質問に対する回答文

著者学籍番号：81621728

題目：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

拝啓

時下ますますご清祥のこととお慶び申し上げます。

この度は、私の修士論文発表において貴重なコメントを賜りまして誠にありがとうございました。頂戴したコメントに従い、以下のような方針に基づいて改良を加えましたので、ご確認を何卒よろしくお願ひ申し上げます。

敬具  
堀井 絵里

質問 1 不自然な例として示した手描きのアニメーションと、自動生成している 3DCG のアニメーションは本質的に違うのではないか。(斎藤英雄先生)

仰る通り、本質は違います。しかし、実際に 3DCG でそのようなシーンを再現する場合も、音と動きを 1 つずつ合わせるには時間と労力がかかるということは、アニメーション制作に詳しい者から聞いております。

実際に、本手法によるアニメーション制作の時間および労力の軽減について、アニメーション制作に詳しい者へアンケートをとったところ、実際に軽減されると予想した者がほとんどでした。そのため、本手法により 3DCG のアニメーション制作の時間および労力の軽減が可能であるといえます。こちらのアンケート結果は、4.2.4 項に示しております。

質問 2 音と指のモーションはどのようにマッピングしているのか。(斎藤英雄先生)

トランペットに関しては、それぞれの指の状態が、ピストンを押すか押さないかの 2 つとなります。一方トロンボーンに関しては、スライドを止める場所が 7 箇所存在します。いずれの楽器も、音と指や腕の状態を 1 対 1 で対応させることができるので、音が変わるとその音が鳴るような姿勢に遷移し、それをその長さだけ保つ、といった実装になっております。

上記の内容は、3.8.1 項にて詳説しております。

質問 3 金管楽器ごとにマップ表を作成することにより、どの楽器でも自動生成が可能であるという点が新規性なのか。(斎藤英雄先生)

はい、仰る通りです。マップ表とモデルさえあれば、どの金管楽器でも容易に吹奏アニメーションの自動生成が可能となっております。

こちらについては、2.5 節に記載いたしました。

質問 4 弦楽器に拡張する時に解決が必要となる課題は何か。(杉浦先生)

弦楽器は金管楽器とは異なり表情の変化は小さいですが、音と弦の押さえ方が 1 対 1 ではないため、どの指使いが適切なのかを、音を鳴らす度に考慮するような処理を加える必要があります。

弦楽器の場合は 2.1 節で紹介している先行研究が存在するため、そちらも参考に実装を行うことにより、対応が可能であると考えられます。

## 質問 5 高音を吹き続けると辛くなる様子を、呼吸のパラメタを用いて表現することについて。(杉浦先生)

頂きましたアドバイスの通り、一定時間息継ぎをしていない、あるいは高音域が続くなど、辛くなるような条件を満たした際に辛そうな表情をする、といったことは実現が可能です。こちらに関しては、モーションに関するメタデータを用意し、そちらに表情やモーションの指定を記載することを考えております。自動的に算出できる場合は算出し、算出ができない場合はそちらのデータに記載することにより、表現が可能であると考えられます。

表情を制御するためには、メッッシュを制御する必要があります。Unreal Engineでメッッシュを制御する際、モーフターゲットという機能を用いますが、この機能はブレンドシェイプが適用されていないメッッシュには使用できません。ここで、ブレンドシェイプとは、あらかじめ用意された複数の表情モデルをパラメタの調整により組み合わせることで、さまざまな表情を作成するアニメーション手法をします。今回用いたユニティちゃんは、デフォルトで口元にブレンドシェイプが設定されているため、モーフターゲットによる口元の制御が可能です。しかし辛い表情などを再現するためには、デフォルトの設定ではブレンドシェイプが不足しているため、他のソフトウェアを介して再設定する必要があると考えられます。つまり、Unreal Engineのみでは再現が不可能であるといえます。

上記の内容については、今後の課題として 5.2.1 項にまとめました。

## 質問 6 人による動きの違いを各演奏者に適用する方法について。(杉本先生)

元となるモーションは全て同じですが、そのモーションの大きさを 0 から 1 のパラメタでランダムに設定することにより、個人差を出す工夫をしています。杉浦先生からご指摘がありました、演奏者ごとに体型が異なる場合は、質問 5 で述べたメタデータにその旨を記載し、モーションを分ける必要があると考えられます。

前者の内容については 3.8.3 項に、後者の内容については今後の課題として 5.2.4 項に記載いたしました。

## 質問 7 ユニティちゃんを Unreal Engine で使用することについて。(杉本先生)

著作権的には問題ございません。こちらについては『ユニティちゃんライセンス条項』の第 3 条にて判断できます。(URL: [http://unity-chan.com/contents/license\\_jp/](http://unity-chan.com/contents/license_jp/))

上記の内容は、3.6 節に明記いたしました。

以上

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 紘里

(学籍番号 : 81621728)

指導教員 教授 藤代 一成

2018年2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。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の製作方法はさまざまであるが、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アニメーションが生成されており、楽器の輝きや形状なども忠実に再現されている。しかし、楽器を演奏するキャラクタの運指や身体の動きが、音楽に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がある。これらの違和感を解消するためには、身体の動きを1つずつ音階やリズムなどに手動で合わせる必要がある。この方法は効率が悪く、多くの時間と労力が必要となる。

上記の課題を解決するため、鍵盤楽器や弦楽器を演奏する演奏者のアニメーションを、音源から自動生成する研究が存在する。しかし、管楽器を対象とした研究は、我々の知るかぎりでは存在しない。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、音源から自動生成することを目指した。アニメータの作業支援を目的とするため、自動生成の流れは実際の演奏アニメーション制作フローに沿わせた。より具体的には、楽曲は電子楽器を用いてMIDI音源として生成した。次に、生成した楽曲を解析することにより、吹奏の情報を得た。最後に、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した。

自動生成したアニメーションについて評価実験を行った結果、提案手法により、自然なアニメーションを自動的に生成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

### Automatic Generation of Blowing Animation Taking Fingering and Expression Synchronized with Digital Music into Account

There are known many animation works with the theme of music performance. Representative televised example of these include “Nodame Cantabile”, “K-ON!”, and “Hibike! Euphonium”, all of which aim at commonly generating realistic motion of performers as well as faithfully reproducing the shape and appearance of the instruments though the underlying production method varies from each other. However, the performer’s fingering and motions are sometimes not completely synchronized with music, and thus giving a sense of incompatibility. This type of artifact occurs especially in a scene that he/she plays a fast phrase or in a complex rhythm. Also, facial expressions in animation series occasionally look unnatural. In order to eliminate these kinds of discomfort, it is necessary to manually synchronize the movement of the body one by one with the scale, rhythm, and so on. This work is inefficient because tedious work and much effort are required.

In order to address these issues, there are known works to automatically generate animations of performers playing keyboard instruments and stringed instruments from sound sources. However, to the best of the author’s knowledge, there are no previous works targeting at wind instruments. Therefore, in this thesis research, we aimed to automatically generate blowing animation of characters playing wind instruments from sound sources. For the purpose of animator’s work assistance, animation was generated in accordance with the actual animation production flow of musical performance. More specifically, the sound source was generated as a MIDI sound source using an electronic musical instrument. Next, we analyzed the generated sound source to obtain music information. Finally, by applying the obtained information to the character’s movement, facial expression, and wind instruments, we generated natural blowing animation synchronized with the sound source.

Evaluation on the resultant animation by third parties empirically proved that the proposed method can generate natural animation automatically.

#### Keywords

Animation; wind orchestra; brass instrument; and automatically generation.

# 目 次

<b>1</b>	<b>序論</b>	<b>1</b>
1.1	研究背景 . . . . .	2
1.2	研究目的 . . . . .	3
1.3	本論文の構成 . . . . .	3
<b>2</b>	<b>関連研究</b>	<b>4</b>
2.1	音からアニメーションを自動生成する研究 . . . . .	5
2.2	アニメーションから音を生成する研究 . . . . .	6
2.2.1	サウンドレンダリング . . . . .	6
2.2.2	プロシージャルオーディオ . . . . .	6
2.3	音とアニメーションを同期させる研究 . . . . .	6
2.4	ユーザの表情をキャラクタに反映する研究 . . . . .	7
2.5	吹奏楽に関連した研究 . . . . .	7
2.6	本研究の新規性 . . . . .	8
<b>3</b>	<b>提案手法</b>	<b>9</b>
3.1	自動生成の流れ . . . . .	10
3.2	入力 . . . . .	10
3.3	出力 . . . . .	11
3.4	デバイス . . . . .	11
3.5	ソフトウェア . . . . .	11
3.6	3D モデル . . . . .	12
3.7	MIDI データから音情報への変換 . . . . .	13
3.8	音情報のモーションへの適用 . . . . .	15
3.8.1	指や腕、楽器のパーツへの適用 . . . . .	15
3.8.2	口元のメッシュへの適用 . . . . .	17
3.8.3	その他の部位への適用 . . . . .	18
3.9	提案手法の使用方法 . . . . .	19
<b>4</b>	<b>結果と評価</b>	<b>20</b>
4.1	実行環境 . . . . .	21
4.2	アニメーション自動生成結果 . . . . .	22
4.3	評価 . . . . .	25
4.3.1	実際の演奏シーンとの比較による評価 . . . . .	26
4.3.2	既存のアニメーションとの比較による評価 . . . . .	29

4.3.3	アンサンブルアニメーションの評価 . . . . .	32
4.3.4	システムの有用性の予想 . . . . .	33
4.3.5	アンケート結果に対する分析 . . . . .	34
<b>5</b>	<b>結論</b>	<b>35</b>
5.1	結論 . . . . .	36
5.2	今後の課題 . . . . .	36
5.2.1	表情の豊かさの向上 . . . . .	36
5.2.2	モーションの種類の向上 . . . . .	37
5.2.3	モーションと音の関連付け . . . . .	37
5.2.4	楽器の種類および演奏者の増加 . . . . .	37
	<b>謝辞</b>	<b>38</b>
	<b>公開文献</b>	<b>39</b>
	<b>参考文献</b>	<b>40</b>

# 図 目 次

1.1 演奏アニメーションが不自然であると感じることがあるかどうか（■：ひじょうにある、■：時々ある、■：どちらともいえない、■：あまりない、■：全くない） . . . . .	2
2.1 マーチングを行う演奏者の軌跡および誤差の表示例 . . . . .	7
3.1 音源から吹奏アニメーションを自動生成する流れ . . . . .	10
3.2 ウィンドシンセサイザ「EWI5000」 . . . . .	11
3.3 MIDI シーケンスソフトウェア「domino」 . . . . .	11
3.4 使用する 3D モデル . . . . .	12
3.5 トランペットのピストンおよびトロンボーンのスライドの番号付け . . . . .	15
3.6 音域による口元の変化 . . . . .	17
3.7 息継ぎをするときの口元 . . . . .	17
3.8 息継ぎ時のモーション . . . . .	18
4.1 Unreal Engine の初期設定画面 . . . . .	21
4.2 トランペット奏者 2 名の基本姿勢 . . . . .	22
4.3 トランペット奏者がピストンを押す様子 . . . . .	22
4.4 演奏者の口元の様子 . . . . .	23
4.5 トランペット奏者 2 名とトロンボーン奏者 2 名の基本姿勢 . . . . .	23
4.6 トロンボーン奏者がスライドを動かす様子 . . . . .	24
4.7 曲の入りを合わせるために膝を使い楽器を下に向かた瞬間の様子 . . . . .	24
4.8 息継ぎ時に身体を反らしている様子 . . . . .	25
4.9 トランペット奏者の指の動きは再現できているか（■：再現できている、■：どちらかといえば再現できている、■：どちらともいえない、■：どちらかといえば再現できない、■：再現できていない） . . . . .	26
4.10 トランペット奏者の身体の動きは自然か（■：自然である、■：どちらかといえば自然である、■：どちらともいえない、■：どちらかといえば不自然である、■：不自然である） . . . . .	27
4.11 トロンボーン奏者の腕の動きは再現できているか（■：再現できている、■：どちらかといえば再現できている、■：どちらともいえない、■：どちらかといえば再現できない、■：再現できていない） . . . . .	27
4.12 トロンボーン奏者の身体の動きは自然か（■：自然である、■：どちらかといえば自然である、■：どちらともいえない、■：どちらかといえば不自然である、■：不自然である） . . . . .	28

4.13 既存のアニメーションと自動生成したアニメーションを比較したときに、トランペッ ト奏者の指の動きは、どちらの方がより音楽のリズムに合っているか（■：自動生成 したアニメーション、■：どちらかといえば自動生成したアニメーション、■：どち らともいえない、■：どちらかといえば既存のアニメーション、■：既存のアニメー ション） . . . . .	29
4.14 既存のアニメーションと自動生成したアニメーションを比較したときに、トランペッ ト奏者の身体の動きは、どちらの方がより自然か（■：自動生成したアニメーション、 ■：どちらかといえば自動生成したアニメーション、■：どちらともいえない、■： どちらかといえば既存のアニメーション、■：既存のアニメーション） . . . . .	30
4.15 既存のアニメーションと自動生成したアニメーションを比較したときに、トロンボー ン奏者の腕の動きは、どちらの方が音楽のリズムに合っているか（■：自動生成した アニメーション、■：どちらかといえば自動生成したアニメーション、■：どち らともいえない、■：どちらかといえば既存のアニメーション、■：既存のアニメーション）	30
4.16 既存のアニメーションと自動生成したアニメーションを比較したときに、トロンボー ン奏者の身体の動きは、どちらの方がより自然か（■：自動生成したアニメーション、 ■：どちらかといえば自動生成したアニメーション、■：どちらともいえない、■： どちらかといえば既存のアニメーション、■：既存のアニメーション） . . . . .	31
4.17 演奏時や息継ぎ時の口元は自然か（■：自然である、■：どちらかといえば自然であ る、■：どちらともいえない、■：どちらかといえば不自然である、■：不自然である）	32
4.18 全員でタイミングを合わせつつ演奏している雰囲気はあるか（■：ある、■：どち らかといえばある、■：どちらともいえない、■：どちらかといえばない、■：ない）	33
4.19 アニメーション制作の労力や時間が軽減されると思うか否か（■：軽減されると思 う、■：多少は軽減されると思う、■：どちらともいえない、■：あまり軽減されな いと思う、■：全く軽減されないと思う） . . . . .	33

# 表 目 次

3.1	MIDI データと音, 運指の対応 . . . . .	16
4.1	実装環境 . . . . .	21

---

## 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、製作方法はさまざまであるが、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アニメーショ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いる。しかし、楽器を演奏するキャラクタの運指や身体の動きが、音楽に完全に同期されていないことがある。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がある。

実際に、演奏アニメーションが不自然であると感じるか否か、以下の3グループに分けてアンケートをとった。

- (a) 吹奏楽・オーケストラ団体への所属経験者または金管楽器（主にトランペット、トロンボーン、ホルン、ユーフォニアム、チューバ）経験者：12名
- (b) (a) にはあてはまらない楽器経験者：16名
- (c) 楽器未経験者：14名

図1.1は、その集計結果である。楽器経験者を中心に、演奏アニメーションが不自然であると感じている者が多いことが分かった。

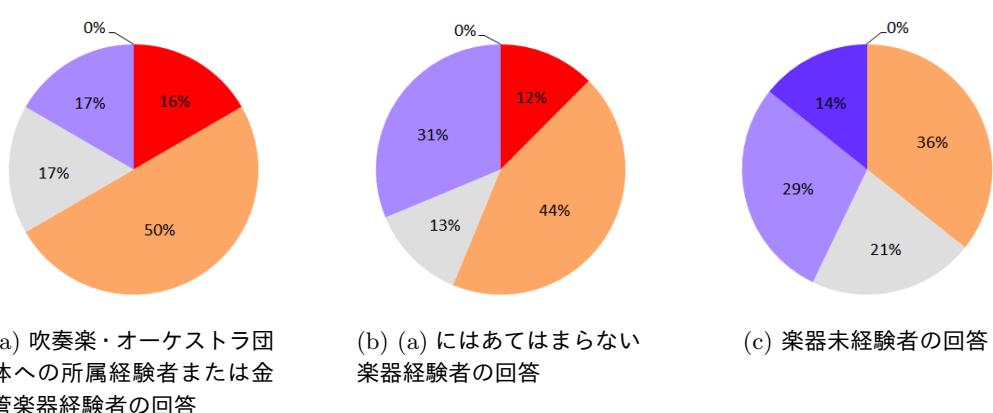


図1.1 演奏アニメーションが不自然であると感じることがあるかどうか

（■：ひじょうにある、■：時々ある、■：どちらともいえない、■：あまりない、■：全くない）

演奏アニメーションが不自然であると感じる要因は、指使いが間違っている点、指や身体の動きがリズムに合っていない点、演奏している雰囲気が感じられない点、表情など、吹奏楽やオーケストラ団体への所属経験や楽器の経験に関わらず、さまざまであるという結果も得た。

これらの不自然さを解消するためには、身体の動きを1つずつ音階やリズムなどに手動で合わせる必要がある。しかし、この方法はひじょうに効率が悪く、多くの時間と労力が必要となる。

## 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、楽曲は電子楽器を用いてMIDI（Musical Instrument Digital Interface）音源として生成する。次に、作成した楽曲を解析することにより、吹奏の情報を得る。最後に、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。提案手法の対象ユーザはアニメータとし、最終的にはアニメータの作業支援を目的とする。また、本研究は、藤代研究室学士4年の武内航と共同で進めてきた。武内が担当した部分については、その旨を明記する。

## 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。

### 第2章 関連研究

アニメーションや音を自動生成する研究や、吹奏楽に関する研究を紹介する。

### 第3章 提案手法

1.2節で示した研究目的を達成するための手法を説明する。

### 第4章 自動生成結果と評価

アニメーションの自動生成結果について述べ、この結果に対する評価をまとめる。

### 第5章 結論

本研究の結論を述べ、今後の課題に言及する。

なお、本研究の成果の一部はVisual Computing/グラフィクスとCAD合同シンポジウム2017においてポスター[iii]として発表した。そして、映像表現・芸術科学フォーラム2018（2018年3月）において発表する予定である。

---

## 第2章

---

### 関連研究

本章では関連研究に言及する。まず、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、2.2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節で音とアニメーションを同期させる研究、2.4節でユーザの表情をキャラクタに反映させる研究、2.5節で吹奏楽に関連した研究を紹介する。最後に2.6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つずつ合わせる必要がある。Langloisら[18]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同期させる、リップシンクという技術がある。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、アニメーションに合わせて台詞を収録するため、口元と台詞の同期が比較的容易である。一方、音声に合った口元のアニメーションを生成するには、口の形や動きを音声と1つずつ合わせる必要がある。また、映画の吹替え版では、役者の口と声にズレが生じることに、違和感を感じることがある。上述の課題を解消するため、リップシンクは、近年注目を集めている技術の1つである。

Breglerら[10]、Ezzatら[15]は、既存の動画に映っている対象人物の口元のみを、音声データに合わせて再構成することにより、リップシンクを図る手法を提案した。Edwardsら[13]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、顎と唇だけ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。さらに、後から表情を編集することが可能となっているため、より自然な表情のアニメーションを実現できる。

本研究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行われている。Zhuら[28]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。ピアノのような鍵盤楽器は、指と音が一対一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ときは、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。

Yinら[25]は、wave音源からバイオリンを弾く手のアニメーションを自動生成する手法を提案した。彼らは弦を押さえる左手の指元のアニメーションだけでなく、その指の動きが不自然に見えないような手首や腕の動きも再現している。加えて、弓を持つ右手の動きも自動生成している。Kimら[16]も同様に、バイオリンを引く手のアニメーションを自動生成する手法を提案した。

ElKouraら[14]は、ギターを弾く左手のアニメーションを自動生成する手法を提案した。彼らの研究はギターの弾き方の指導を目的としており、アニメーションは音源から自動生成するのではなく、指を置く位置を入力することで生成される。

上述の研究の他にも、音源に合った表情を自動生成する研究[11][12][9]や、音源に合ったダンスマニエーションを自動生成する研究[29][17]など、さまざまなジャンルに着目した研究が存在する。

## 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。以下では、その分野に関連した研究を紹介する。

### 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研究分野である。この分野は、1992年に発表されたTakalaらの論文[22]が発端となっている。彼らは、音の実態が光と同じ波動であること、音が物体の動きに依存することに着目して、音をレンダリングにより自動生成する手法を提案した。2009年には1つのシミュレーションパイプラインのなかで、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[26]がZhengにより考案された。本アルゴリズムでは、自然な流水音が、流水アニメーションから自動生成される。また、自動生成された流水音は、アニメーションと完全に同期している。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[27]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。そのため、複数の物体がぶつかり合い、破壊し合うようなアニメーションでも、そのアニメーションに同期する音が、忠実に再現された。

### 2.2.2 プロシージャルオーディオ

プロシージャルオーディオとは、作り置きした音を必要なタイミングで再生するのではなく、その都度プログラムで作り出す技術である。ゲームソフトを開発している株式会社スクウェア・エニックスは本技術を研究しており、2014年のCEDECにて詳細[2]を発表した。

ゲームでは、プレイヤーの操作に応じてシステム内で効果音が選択され、再生される。それらの音は、一般的には事前に作成し、システムに記憶させておく。しかし、ゲームは使用可能な記憶領域が限られているため、可能な限りメモリを節約したい。そのようなときに、本技術はひじょうに役立つ。また、物体の挙動に合わせて音を自動生成するため、自然な効果音を再現することができる。

## 2.3 音とアニメーションを同期させる研究

音やアニメーションを、もう一方から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[19]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。

この研究分野は、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[20]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 ユーザの表情をキャラクタに反映する研究

2017年秋に発売されたiPhoneXには、アニ文字という機能がある。この機能は、カメラで認識したユーザの表情を、デバイス内のキャラクタに反映することができる。このように、ユーザの表情をカメラで認識することにより、キャラクタのフェイシャルアニメーションを自動生成することを目的とした研究が存在する。

Weiseら[23]は、カメラを用いてユーザの顔のモーションをキャプチャすることにより、顔のモーションデータを取得し、それを仮想世界のキャラクタの顔に反映する手法を提案した。顔のモーションキャプチャにはマーカーが必要がない。また、キャプチャ中に手などにより顔が遮られた場合は、自動的に補正される仕組みとなっている。なお、反映されるキャラクタは、人型のキャラクタのみである。この研究を行っていたメンバーの1人であるPauly[21]は、2年後に、トラッキングの際に深さを考慮することにより、トラッキングの質を向上させた。そのことにより、しわなど皮膚の細かい情報もキャラクタに反映されるようになった。

同年にXuら[24]は、顔のモーションデータを人型ではないキャラクタの顔に反映する手法を提案した。

## 2.5 吹奏楽に関連した研究

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名の演奏者の軌跡を可視化することにより、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2.1)[i][ii]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。

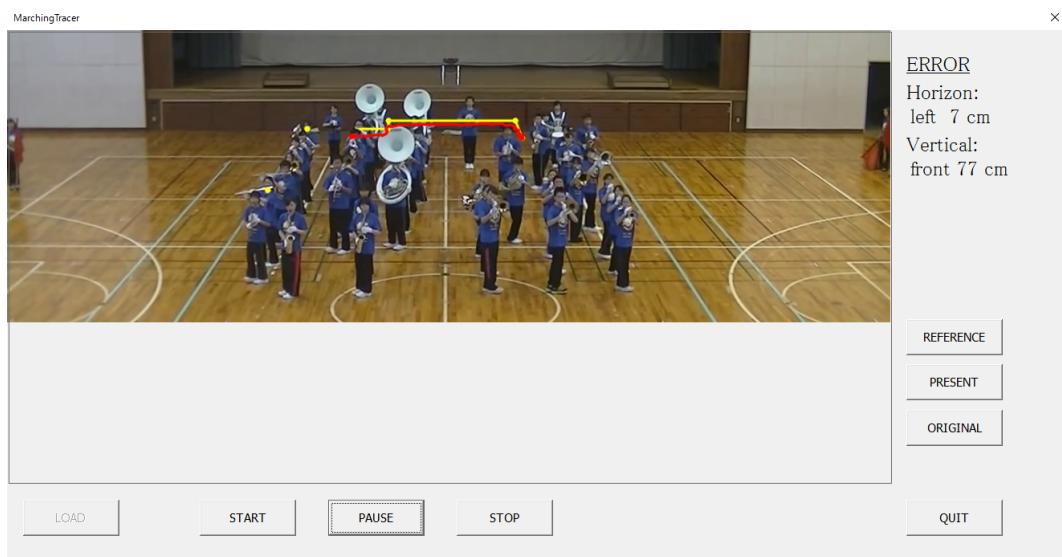


図2.1 マーチングを行う演奏者の軌跡および誤差の表示例

なお、この研究の成果は、NICOGRAPH2016においてポスター[i]にて発表した。そして、一連の研究をまとめた成果が、画像電子学会誌のVol.47[ii]に掲載された。

## 2.6 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。また、2.1節で挙げた、ピアノやバイオリンを演奏するアニメーションの自動生成を目的とした研究では、複数名での演奏は考慮されていない。したがって本研究の新規性は、管楽器を複数名で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だけを示すが、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できる。このように、あらゆる楽器への対応が可能である点も、新規性があるといえる。

---

## 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、3.1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。3.2節および3.3節にて入出力について説明し、3.4節にて使用するデバイス、3.5節にて使用するソフトウェア、3.6節にて使用する3Dモデルについて述べる。3.7節にてMIDIデータから音情報への変換方法、3.8節にて音情報のモーションへの適用方法を説明する。そして最後に、3.9節にて実際にシステムを使用する際の使用方法に言及する。

### 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。

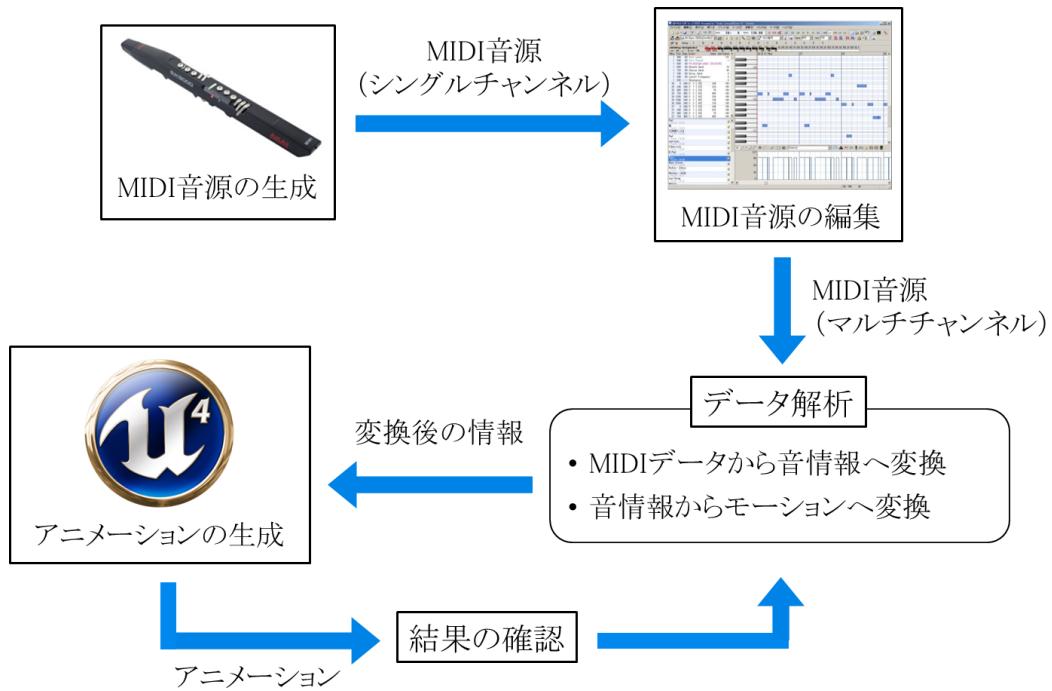


図3.1 音源から吹奏アニメーションを自動生成する流れ

実際のアニメーション制作フローに沿わせるため、音源の生成は電子楽器を用いて行う。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が動くアニメーションを自動生成する。

### 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という信号を用いて発音する音源のことである。一般的に使用される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。なお、MIDIの仕様は文献[8]に詳しく記載されている。

このMIDI音源を生成する方法は、3.4節および3.5節で説明する。

### 3.3 出力

出力は、管楽器を演奏するキャラクタのアニメーションである。今回対象とする管楽器は、トランペット、トロンボーンである。この2本の楽器は、吹奏楽ではとくに目立つ楽器であり、また3Dモデルが入手しやすかったために選んだ。

### 3.4 デバイス

MIDI音源を生成するために、電子楽器であるウインドシンセサイザ「EWI5000」(図3.2)を使用した。このウインドシンセサイザは、さまざまな楽器の音を再現することが可能である。



図3.2 ウインドシンセサイザ「EWI5000」

### 3.5 ソフトウェア

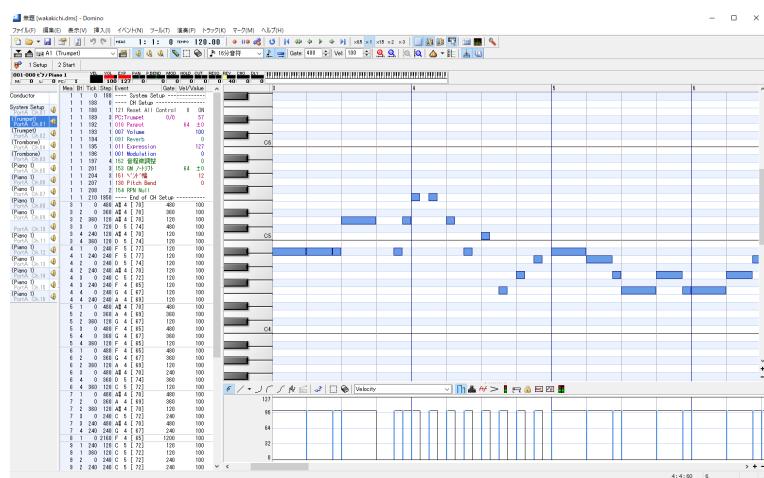


図3.3 MIDIシーケンスソフトウェア「domino」

3.4節で述べたウインドシンセサイザは、単音しか鳴らすことができないため、シングルチャンネルの音源しか生成することができない。しかし、吹奏アニメーションを自動生成するためには、複数

名で演奏しているマルチチャンネルの音源が必要となる。そこで、ウインドシンセサイザで生成したMIDI音源を、フリーソフトウェアであるMIDIシーケンスソフトウェア「Domino」(図3.3)[5]へ出力し、重ねて何度も録音することにより、マルチチャンネルの音源を生成した。

また、アニメーションの生成には、Epic Gamesより開発されたゲームエンジン、Unreal Engine[3]を用いた。

### 3.6 3D モデル

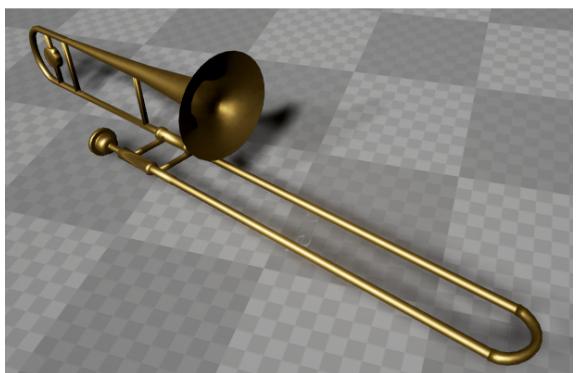
使用する3Dモデルは、ユニティちゃん(図3.4(a)[6])、トランペット(図3.4(b)[4])、トロンボーン(図3.4(c)[1])である。なお、ユニティちゃんのライセンス条約は、webサイト[7]にて確認済みである。



(a) ユニティちゃん



(b) トランペット



(c) トロンボーン

図3.4 使用する3Dモデル

ユニティちゃんは、主に以下の部位を制御する。

- 右指（トランペット演奏時）
- 右腕（トロンボーン演奏時）
- 背
- 腰
- 両足
- 口元

### 3.7 MIDIデータから音情報への変換

MIDIは、チャンクとよばれるデータブロックから構成され、先頭にヘッダチャンク、その後にトラックチャンクが続く。ヘッダチャンクには、チャンクタイプ、データ長さ、フォーマットタイプ、トラック数、タイムベース値という5つの値が格納されている。ここで、タイムベース値とは、四分音符1つ分のクロック数を表す値であり、一般的には48から960までの自然数から、96の倍数が選ばれることが多い。一方トラックチャンクには、チャンクタイプ、データ長、トラックイベントデータが格納されている。提案手法では、トラック数、タイムベース値、トラックイベントデータを楽譜データに変換した後に、アニメーションに適用できる形へとさらに変更する。

まず、音に関する情報を楽譜データに変換する方法について説明する。テンポ情報は、四分音符あたりの秒数( $\mu s$ )として格納されている。音に関する情報は、トラックイベントデータとして、以下のように16進数で格納されている。

9 0 4 8 6 4 8 1 7 0 8 0 4 8 0 0 8 3 6 0

この情報を2つずつペアにし、そこから音の種類と長さを取得する。それぞれの数字から得られる情報は、以下の通りである。

- 90: ノートオン. このタイミングで音を鳴らすことを意味する.
- 48: 音の種類 (48 はドを意味する. 対応表は次項に示す.)
- 64: 音の大きさ
- 81: 音を鳴らす長さ
- 70: 音を鳴らす長さ
- 90: ノートオフ. このタイミングで音を止めることを意味する.
- 48: 音の種類 (48 はドを意味する.)
- 00: 音の大きさ (音を止めているため, 大きさはゼロとなる.)
- 83: 音を止める長さ
- 60: 音を止める長さ

音の長さは, 以下のように取得する. ここでは, 81 70 を用いて説明する.

- それぞれを 2 進数に変換する. 81: 1000 0001, 70: 0111 0000
- それぞれの最上位ビットを取り除いて合成し, 10 進数に直す. 000 0001 111 0000 → 240

この例の場合, ドを 240 クロック分伸ばすことを意味する. 以降, この値をデルタタイムとよぶ. 最後に, このクロックを現実時間 [s] に変換する. 変換式は式 (3.1) で表される.

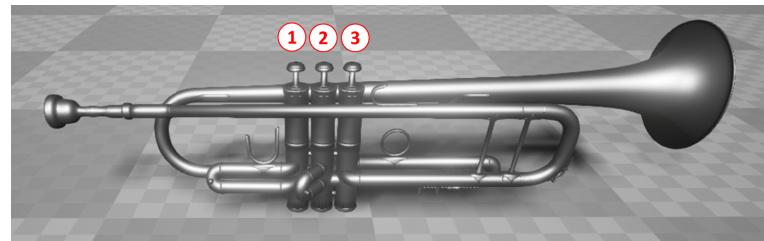
$$\text{時間 } [s] = \frac{\text{デルタタイム } [\text{ticks}] \times 60[\text{seconds}/\text{minute}]}{\text{タイムベース値 } [\text{ticks}/\text{beat}] * \text{テンポ } [\text{beats}/\text{minute}]} \quad (3.1)$$

タイムベース値, テンポをそれぞれ仮に 480, 120 とすると, 例で示した音情報は, 譜面情報に直した結果『ドを 0.25 秒伸ばし, 0.5 秒休む』ことを意味すると分かる.

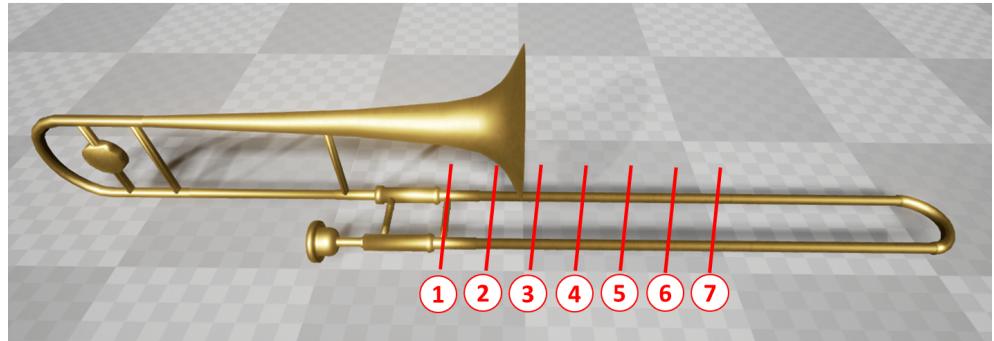
## 3.8 音情報のモーションへの適用

### 3.8.1 指や腕、楽器のパーツへの適用

楽器を演奏する様子をアニメーションで再現するには、キャラクタの指や腕を介して楽器のモデルを制御する必要がある。トランペットはピストンの操作、トロンボーンはスライドの操作により音を変えることができ、ピストンは3箇所、スライドは止める場所が大きく分けて7箇所ある。トランペットのピストン番号を、吹き口に近い方から1-3(図3.5(a))、トロンボーンのスライドの位置を、吹き口に近い方から1-7(図3.5(b))と表すと、MIDIデータと音、運指の対応は表3.1となる。



(a) ピストン番号



(b) スライドの位置番号

図 3.5 トランペットのピストンおよびトロンボーンのスライドの番号付け

表 3.1 MIDI データと音、運指の対応

MIDI データ	音名（実音）	トランペット	トロンボーン
⋮	⋮	⋮	⋮
39	A (ラ, 442Hz)	2	2
3b	B (シ)	1・2・3	4
3c	C (ド)	1・3	3
3e	D (レ)	1・2	1
40	E (ミ)	2	2
41	F (ファ)	0	1
43	G (ソ)	1・2	2
45	A (ラ)	2	2
⋮	⋮	⋮	⋮

3.7節で例として挙げた、『ドを 0.25 秒伸ばし、0.5 秒休む』という譜形を演奏する場合、トランペットの場合は、1番と3番を 0.25 秒間押し続け、その後 0.5 秒間は休みのためそのまま、トロンボーンの場合は、スライドを 3 番まで移動させ、そのまま 0.25 秒間維持、その後 0.5 秒間は休みのためそのまま、という表現方法となる。

### 3.8.2 口元のメッシュへの適用

金管楽器を演奏する際、高音域の演奏時は、低音域に比べて口元が緊張する。緊張の程度には個人差があるが、本論文では、高音演奏時と低音演奏時の口元の違いを図3.6のように、口の引き具合で表す。

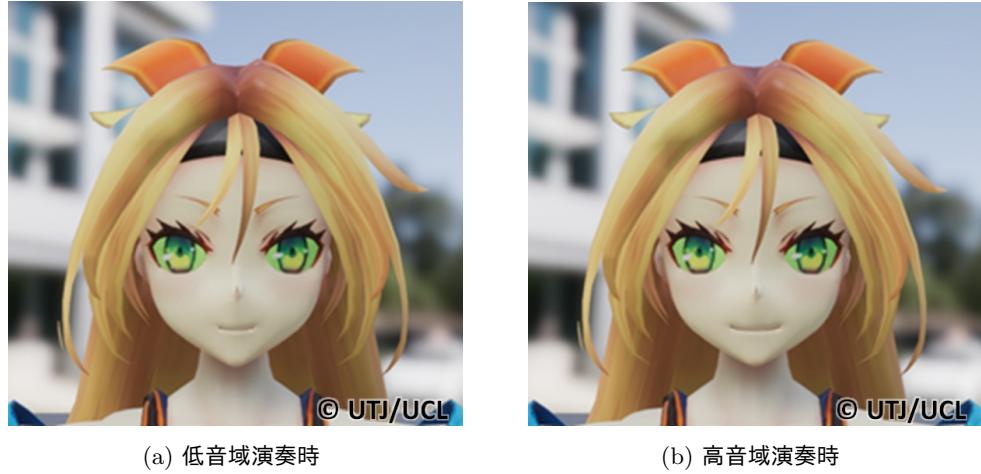


図 3.6 音域による口元の変化

休みが一定の時間以上続く場合、演奏者は息継ぎを行う。息継ぎのタイミングには個人差があり、演奏しているフレーズによっても異なるが、提案手法では息継ぎのタイミングを以下の3パターンに分ける。

- 休みが0.5拍以上～1拍未満：休みの間、息継ぎのモーションを行う。
- 休みが1拍以上～2拍未満：最後の1拍で息継ぎのモーションを行う。
- 休みが2拍以上：最後の2拍間で、息継ぎの予備モーションおよび息継ぎのモーションを行う。

息継ぎのモーションを行うときに、口元を図3.7のように変形させる。その他のモーションについては次項で述べる。



図 3.7 息継ぎをするときの口元

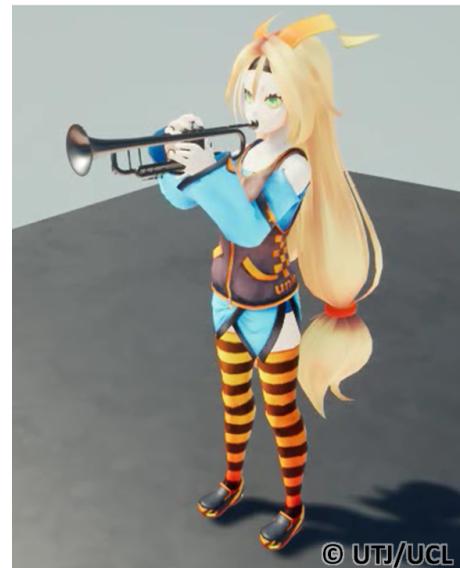
### 3.8.3 その他の部位への適用

息継ぎをするときは口元だけでなく、上半身も動く。そこで、前項で述べたタイミングで、息継ぎのモーション（図3.8(a)）、息継ぎの予備モーション（図3.8(b)）を行う。なお、比較対象として通常の直立状態を図3.8(c)に示す。楽器と床との位置関係にて差異が確認できる。



(a) 息継ぎのモーションの最中

(b) 息継ぎの予備モーションの最中



(c) 通常の直立状態

図3.8 息継ぎ時のモーション

息継ぎ時以外にも演奏者は動く。例えば、任意のタイミングで重心を移動させたり、テンポに合わせて身体や楽器を上下に揺らす。しかし実際、演奏時の動きには個人差があるため、モーションの大きさを調節するためのパラメタを指定することによる個人差の表現が可能となっている。アンサンブルアニメーションを自動生成する際は、全員に異なるパラメタをランダムで割り当てることにより、個人差を表現する。ここで、アンサンブルとは複数名で演奏することを意味し、一般的には少人数で演奏することをさす。アンサンブルには指揮者が存在しないため、曲の始まりは、リードを担当する演奏者が楽器や身体で合図をする。なお、重心移動や合図のモーションは、武内が担当した。

### 3.9 提案手法の使用方法

提案手法を用いてアニメーションを自動生成する際は、以下の順序が必要となる。ここで、キャラクタや楽器のモデルは、あらかじめセッティングされているものとする。

1. MIDI 音源の作成
2. 音源データの相対パスをソースコードに記載し、コンパイル
3. Unreal Engine を起動
4. 1. で生成した音源を BGM として設定
5. 再生
6. 結果の確認および修正

手順 4. について、本来なら MIDI 音源を解析すると同時に音を流すべきであるが、現在の実装ではそれが不可能となっているため、解析する音源とは別に、流す音源として新たに設定する必要がある。また、このとき音のずれが生じる場合がある。その場合は、音源を流すタイミングの調節が必要となるが、一度調節をするだけでアニメーションと音は完全に同期する。

---

## 第4章

### 結果と評価

---

本章では、3章で紹介した手法によって自動生成した吹奏アニメーションについて記述する。4.1節ではシステムの仕様や使用したPCについて述べ、4.2節では自動生成した吹奏アニメーションのキャプチャリング画像を示す。そして4.3節では、自動生成結果を実際の演奏シーンや既存の吹奏アニメーションと比較することにより、提案手法を評価する。

## 4.1 実行環境

事前にUnreal Engineにて専用のプロジェクトを作成し、モーションのデータが記載されているUnreal Engine専用のファイルをインポートする。そして、キャラクタと、それぞれが演奏する楽器を図4.1のようにセッティングする。なお、最初にインポートするUnreal Engine専用のファイルには、キャラクタの姿勢データも存在するため、キャラクタの姿勢のセッティングは容易にできる。

実装環境は表4.1の通りである。

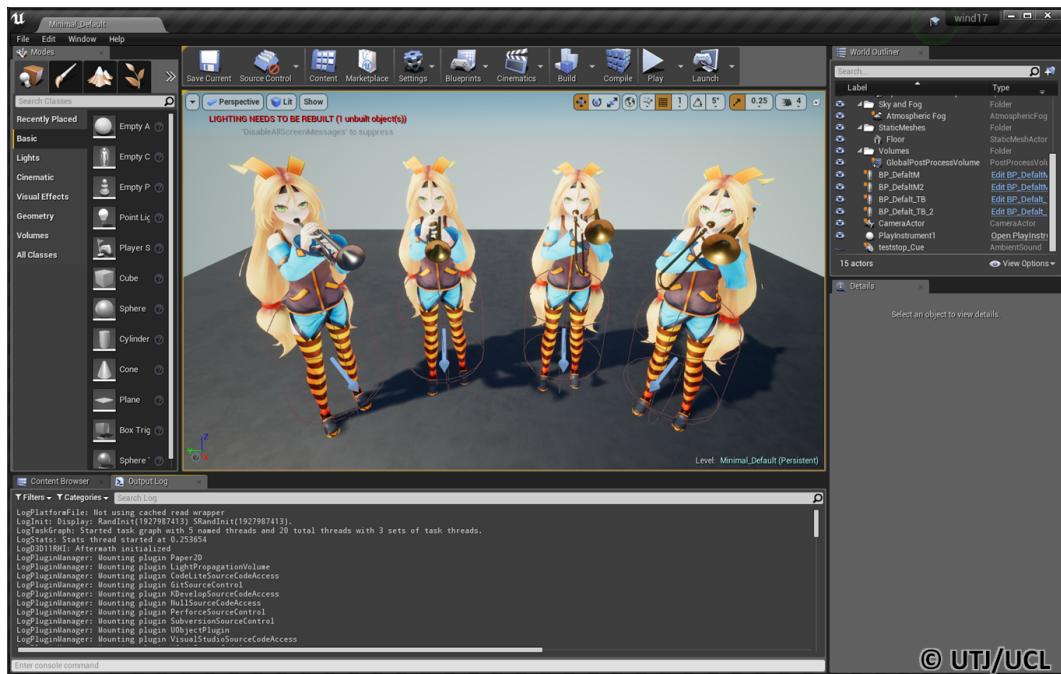


図 4.1 Unreal Engine の初期設定画面

表 4.1 実装環境

OS	Windows10 64bit
CPU	Intel®Core™i7-3930K
RAM	32.00GB
言語	c++

## 4.2 アニメーション自動生成結果

本節では、自動生成したトランペット奏者 2 名で演奏するアンサンブルアニメーションと、トランペット奏者 2 名、トロンボーン奏者 2 名の計 4 名で演奏するアンサンブルアニメーションのキャプチャリング画像を示し、自動生成結果について述べる。以下、2つのアニメーションそれぞれについて、制御している口元やボーンを示すが、前者のアニメーションでは、トランペット奏者の口元や指元がズームインされているため、主に口元や指元のボーンの制御に、後者のアニメーションでは、全体を俯瞰しているシーンであるため、前者のアニメーションにて触れなかったボーンの制御に触れる。

まず、トランペット奏者 2 名でのアンサンブルアニメーションの自動生成結果について述べる。図 4.2 は、基本姿勢である。

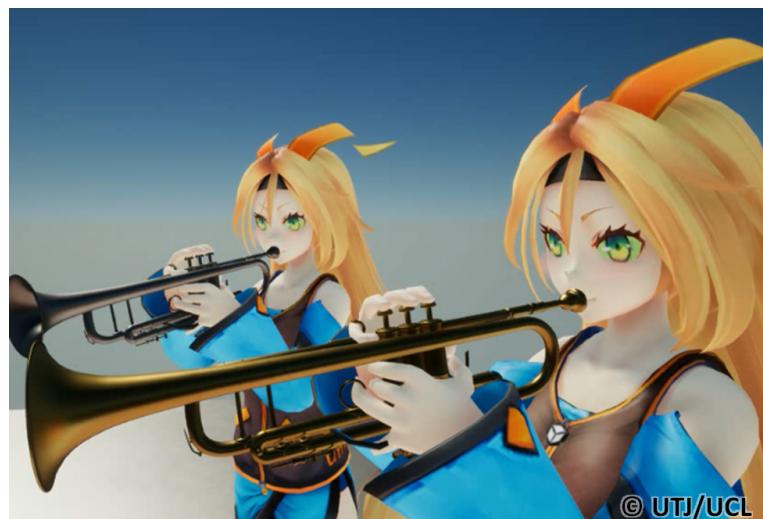


図 4.2 トランペット奏者 2 名の基本姿勢

トランペットを演奏する際は、図 4.3 のように指でピストンを操作する。

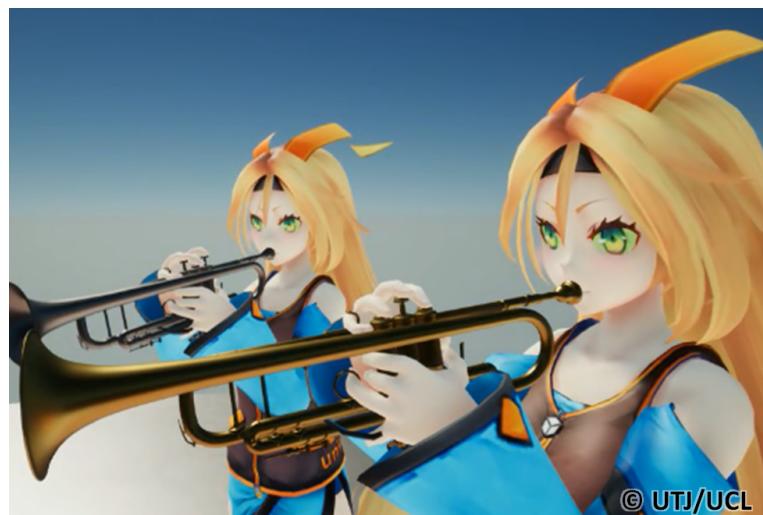


図 4.3 トランペット奏者がピストンを押す様子

息継ぎをするタイミングでは、演奏者は図4.4(a)のように口を開ける。また、高音を演奏するタイミングでは、図4.4(b)のように口元が緊張する。



図4.4 演奏者の口元の様子

次に、トランペット奏者2名、トロンボーン奏者2名でのアンサンブルアニメーションの自動生成結果について述べる。図4.5は、基本姿勢である。



図4.5 トランペット奏者2名とトロンボーン奏者2名の基本姿勢

トロンボーンを演奏する際は、図4.6のように右手でスライドを操作する。



図4.6 トロンボーン奏者がスライドを動かす様子

曲の入りでは、タイミングを合わせるために、膝を使って楽器を縦に振る。図4.7は、膝を曲げ、楽器を下に向けた瞬間の様子である。



図4.7 曲の入りを合わせるために膝を使い楽器を下に向けた瞬間の様子

息継ぎをするタイミングでは、演奏者は身体を反らす。図4.8は、トランペット奏者2名が息継ぎをしている瞬間の様子である。



図 4.8 息継ぎ時に身体を反らしている様子

### 4.3 評価

自動生成した吹奏アニメーションに対して、以下の 3 つの評価を行った。

- 実際の演奏シーンとの比較による評価
- 既存のアニメーションとの比較による評価
- アンサンブルアニメーションの評価

以下、評価結果を、1.1 節と同様に以下の 3 グループに分けて示す。

- (a) 吹奏楽・オーケストラ団体への所属経験者または金管楽器（主にトランペット、トロンボーン、ホルン、ユーフォニアム、チューバ）経験者：12 名
- (b) (a) にはあてはまらない楽器経験者：16 名
- (c) 楽器未経験者：14 名

### 4.3.1 実際の演奏シーンとの比較による評価

トランペット奏者やトロンボーン奏者が実際に演奏しているシーンと、自動生成した吹奏アニメーションを比較することによる評価を、以下の4項目について行った。

- トランペット奏者の指の動きは再現できているか
- トランペット奏者の身体の動きは自然か
- トロンボーン奏者の腕の動きは再現できているか
- トロンボーン奏者の身体の動きは自然か

図4.9は、トランペット奏者の指の動きが再現できているか否かを評価した結果である。いずれのグループも、提案手法に対して肯定的な回答をした者がほとんどであり、否定的な回答をした者はいなかった。

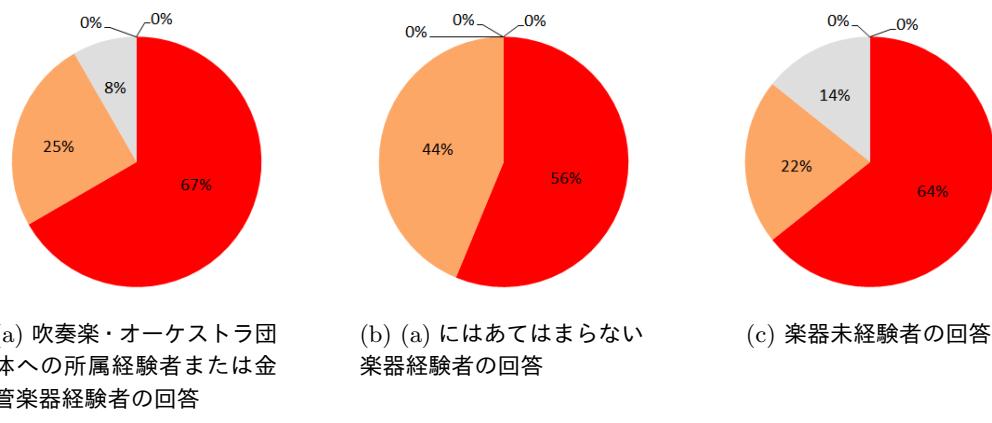


図4.9 トランペット奏者の指の動きは再現できているか

(■: 再現できている, ■: どちらかといえば再現できている, ■: どちらともいえない, ■: どちらかといえば再現できていない, ■: 再現できていない)

図4.10は、トランペット奏者の身体の動きが自然か否かを評価した結果である。いずれのグループも、提案手法に対して否定的な回答をした者の比率が高かった。

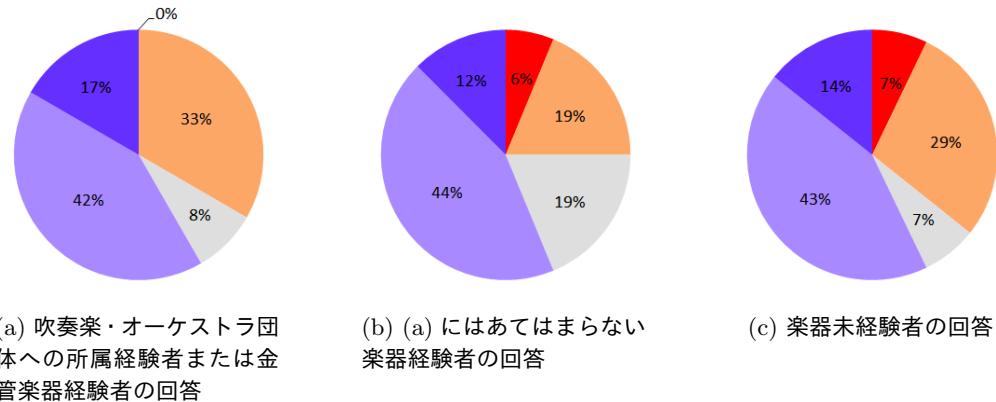


図4.10 トランペット奏者の身体の動きは自然か  
(■: 自然である, ■: どちらかといえば自然である, ■: どちらともいえない, ■: どちらかといえば不自然である, ■: 不自然である)

図4.11は、トロンボーン奏者の腕の動きが再現できているか否かを評価した結果である。吹奏楽やオーケストラ団体への所属経験者や金管楽器奏者は、全員提案手法に対して肯定的な回答をし、否定的な回答をした者はいなかった。また、その他のグループも、大半が肯定的な回答をした。

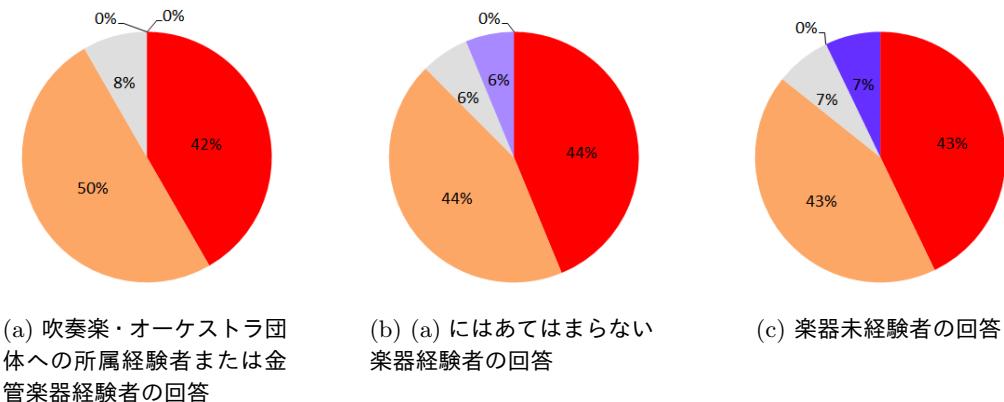
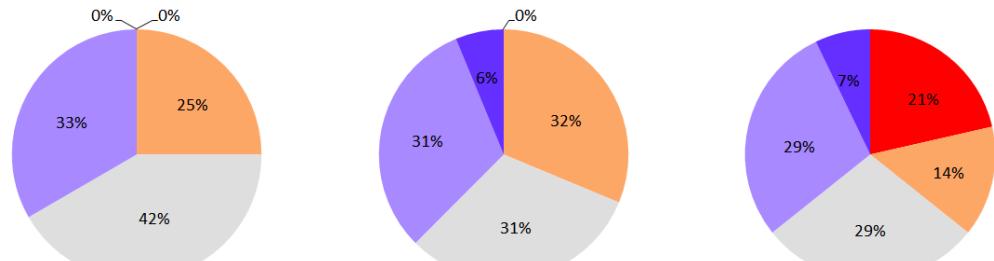


図4.11 トロンボーン奏者の腕の動きは再現できているか  
(■: 再現できている, ■: どちらかといえば再現できている, ■: どちらともいえない, ■: どちらかといえば再現できない, ■: 再現できない)

図4.12は、トロンボーン奏者の身体の動きが自然か否かを評価した結果である。提案手法に対して肯定的な回答をした者も否定的な回答をした者も、それぞれのグループにおいて半分以下という結果になった。



(a) 吹奏楽・オーケストラ団体への所属経験者または金管楽器経験者の回答  
(b) (a) にはあてはまらない  
樂器経験者の回答  
(c) 楽器未経験者の回答

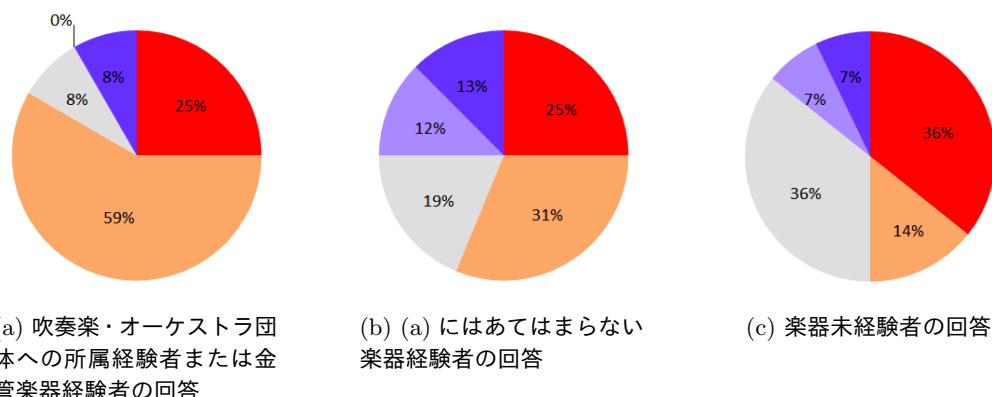
図4.12 トロンボーン奏者の身体の動きは自然か  
(■：自然である, □：どちらかといえば自然である, ▨：どちらともいえない, ▢：どちらかといえば不自然である, ▤：不自然である)

### 4.3.2 既存のアニメーションとの比較による評価

既存の吹奏アニメーションにおける、トランペット奏者やトロンボーン奏者が実際に演奏しているシーンと、自動生成した吹奏アニメーションを比較することによる評価を、以下の4項目について行った。

- トランペット奏者の指の動きは、どちらの方がより音楽のリズムに合っているか
- トランペット奏者の身体の動きは、どちらの方がより自然か
- トロンボーン奏者の腕の動きは、どちらの方がより音楽のリズムに合っているか
- トロンボーン奏者の身体の動きは、どちらの方がより自然か

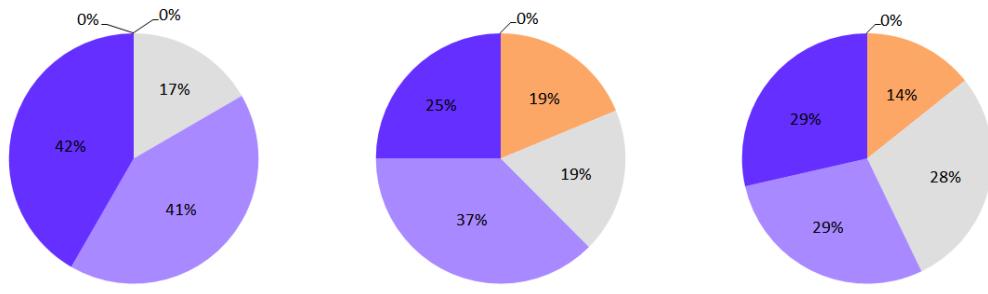
図4.13は、既存のアニメーションと自動生成したアニメーションを比較したときに、トランペット奏者の指の動きはどちらの方がより音楽のリズムに合っているのかを評価した結果である。いずれのグループも、半分以上が提案手法に対して肯定的な回答をした。特に、吹奏楽やオーケストラ団体への所属経験者や金管楽器奏者は、ほとんどが肯定的な回答をした。



(a) 吹奏楽・オーケストラ団体への所属経験者または金管楽器経験者の回答  
 (b) (a) にはあてはまらない楽器経験者の回答  
 (c) 楽器未経験者の回答

図4.13 既存のアニメーションと自動生成したアニメーションを比較したときに、トランペット奏者の指の動きは、どちらの方がより音楽のリズムに合っているか  
 (■: 自動生成したアニメーション, □: どちらかといえば自動生成したアニメーション, ■: どちらともいえない, □: どちらかといえば既存のアニメーション, ■: 既存のアニメーション)

図4.14は、既存のアニメーションと自動生成したアニメーションを比較したときに、トランペット奏者の身体の動きはどちらの方がより自然なのかを評価した結果である。提案手法に対して肯定的な回答をした者の割合は小さく、どのグループも半数以上が否定的な回答をした。

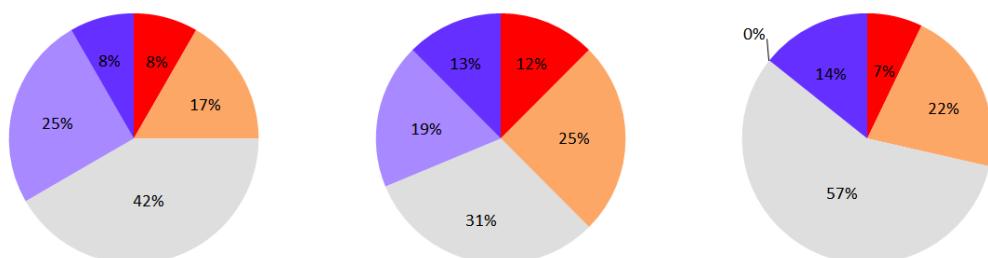


(a) 吹奏楽・オーケストラ団体への所属経験者または金管楽器経験者の回答  
(b) (a) にはあてはまらない楽器経験者の回答  
(c) 楽器未経験者の回答

図4.14 既存のアニメーションと自動生成したアニメーションを比較したときに、トランペット奏者の身体の動きは、どちらの方がより自然か

(■：自動生成したアニメーション、■：どちらかといえば自動生成したアニメーション、■：どちらともいえない、■：どちらかといえば既存のアニメーション、■：既存のアニメーション)

図4.15は、既存のアニメーションと自動生成したアニメーションを比較したときに、トロンボーン奏者の腕の動きはどちらの方がより音楽のリズムに合っているのかを評価した結果である。いずれのグループも、半分以上が提案手法に対して肯定的な回答をした。また図4.13とは異なり、肯定的な回答をした者の比率は、吹奏楽やオーケストラ団体への所属経験や楽器の経験に依存しなかった。



(a) 吹奏楽・オーケストラ団体への所属経験者または金管楽器経験者の回答  
(b) (a) にはあてはまらない楽器絏験者の回答  
(c) 楽器未経験者の回答

図4.15 既存のアニメーションと自動生成したアニメーションを比較したときに、トロンボーン奏者の腕の動きは、どちらの方が音楽のリズムに合っているか

(■：自動生成したアニメーション、■：どちらかといえば自動生成したアニメーション、■：どちらともいえない、■：どちらかといえば既存のアニメーション、■：既存のアニメーション)

図4.16は、既存のアニメーションと自動生成したアニメーションを比較したときに、トロンボーン奏者の身体の動きはどちらの方がより自然なかを評価した結果である。吹奏楽やオーケストラ団体への所属経験や楽器の経験がないほど、提案手法に対して肯定的な回答をした者の比率が高かった。

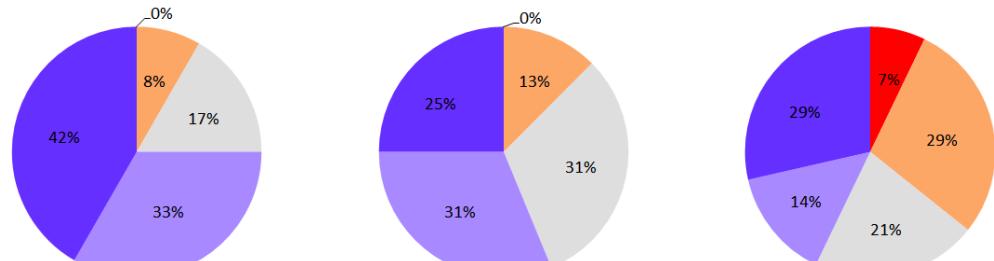


図4.16 既存のアニメーションと自動生成したアニメーションを比較したときに、トロンボーン奏者の身体の動きは、どちらの方がより自然か  
(■：自動生成したアニメーション、□：どちらかといえば自動生成したアニメーション、■：どちらともいえない、■：どちらかといえば既存のアニメーション、■：既存のアニメーション)

### 4.3.3 アンサンブルアニメーションの評価

自動生成したアンサンブルアニメーションの評価を、以下の2項目について行った。

- 演奏時や息継ぎ時の口元は自然か
- 全員でタイミングを合わせつつ演奏している雰囲気はあるか

図4.17は、演奏時や息継ぎ時の口元は自然か否かを評価した結果である。この項目については、金管楽器演奏者の口元の状態に関する知識のある者に対して評価を行ったため、回答者数はそれぞれ(a)10名、(b)10名、(c)9名となっている。いずれのグループも、提案手法に対して肯定的な回答をした者の比率は、否定的な回答をした者の比率よりも高かった。

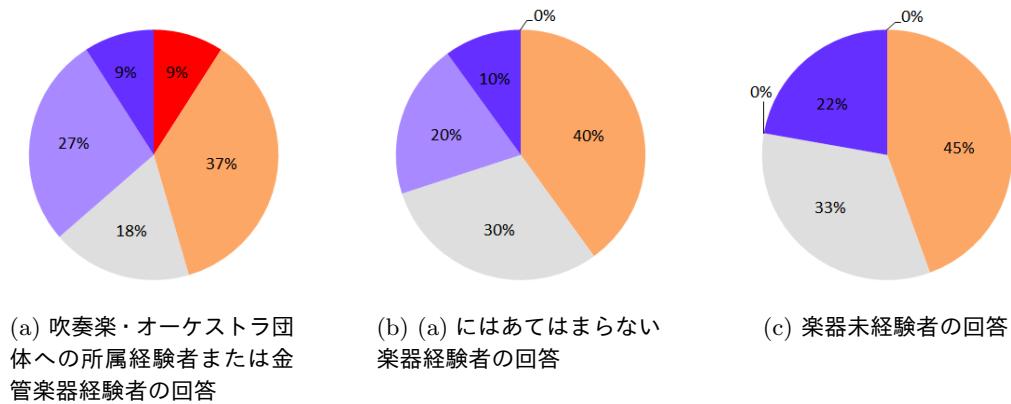
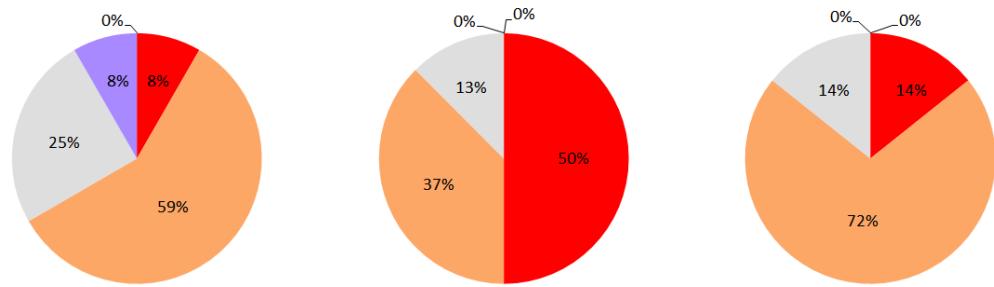


図4.17 演奏時や息継ぎ時の口元は自然か  
 (■: 自然である, ■: どちらかといえば自然である, ■: どちらともいえない, ■: どちらかといえば不自然である, ■: 不自然である)

図4.18は、全員でタイミングを合わせつつ演奏している雰囲気があるか否かを評価した結果である。吹奏楽やオーケストラ団体への所属経験者または金管楽器経験者のグループには、提案手法に対して否定的な回答をした者が少數いたが、それ以外のグループにはいなかった。



(a) 吹奏楽・オーケストラ団体への所属経験者または金管楽器経験者の回答  
(b) (a) にはあてはまらない 楽器経験者の回答  
(c) 楽器未経験者の回答

図4.18 全員でタイミングを合わせつつ演奏している雰囲気はあるか  
(■：ある、 ■：どちらかといえばある、 ■：どちらともいえない、 ■：どちらかといえばない、 ■：ない)

#### 4.3.4 システムの有用性の予想

アニメーション制作やCGに関する知識がある者15名に、本手法によりアニメーション制作時間や労力が軽減されると思うか否かを聞いた。その結果、図4.19のような結果を得た。全員が肯定的な回答をし、そのうち13名が軽減されると思うと回答したことから、本手法により、アニメータの制作支援が可能であると考えられる。

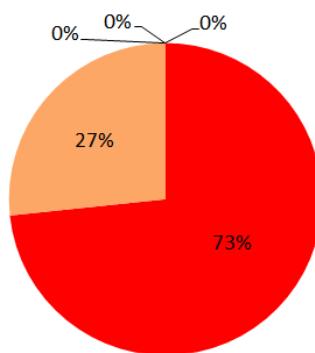


図4.19 アニメーション制作の労力や時間が軽減されると思うか否か  
(■：軽減されると思う、 ■：多少は軽減されると思う、 ■：どちらともいえない、 ■：あまり軽減されないと思う、 ■：全く軽減されないと思う)

#### 4.3.5 アンケート結果に対する分析

指や腕の動きのように、音楽情報から一意に決まる動きに関しては、本手法により音源に同期した自然なアニメーションが作成可能であるといえる。一方、全身の動きのような個人差がある動きに関しては、現段階では本手法による自然なアニメーションが作成可能であるとはいえない結果となつた。しかし、アニメータの作業支援は可能であると予想できる。

吹奏楽やオーケストラ団体への所属経験や楽器の経験がない者ほど、本手法に対して否定的な回答をした者の比率が高かった項目がある。その要因として、グループによって自然であると感じる演奏の動作が異なっていることが挙げられると考えられる。

---

## 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 結論

本論文では、音源から演奏アニメーションを自動生成することにより、音源に同期したアニメーションを自動生成する手法を提案した。本手法により、トランペット奏者の指元やトロンボーン奏者の腕の動きは、音と完全に同期した自然な動きを再現できた。一方、身体の動きは、本手法だけでは自然な動きが再現できなかった。これらのことから、楽譜情報と一対一で結びつけることが可能な動きは、本手法により自然なアニメーションとして再現できるが、個人差のある動きは、現状では再現できないことが結論としていえる。

## 5.2 今後の課題

今後の課題として挙げられるのは、表情やモーションの種類の向上、モーションと音の関連付け、楽器や演奏者の増加である。本節では、それぞれについて詳しく説明する。

### 5.2.1 表情の豊かさの向上

3.8.2 項で口元の制御について述べたが、演奏する際に変化する表情は口元だけではない。他の演奏者と目配せをすることや、音が長くて息継ぎできない場合、音が高い場合に辛そうな表情をすることがある。楽器に息を入れる際に頬が膨らむ演奏者もいる。これらの表情の変化もモデルに適用することにより、表情の豊かさが向上し、より自然なアニメーションが完成すると考えられる。方法としては、音の長さなどから判断できる情報については、音情報から自動的に算出、感情の変化のような、音情報から判断しづらい情報については、表情の遷移を記載するメタデータを用意することが挙げられる。

しかし、目配せや辛い表情など、あらゆる表情の再現は、現状では Unreal Engine のみでは不可能である。口元を制御する際、Unreal Engine のモーフターゲットという機能を用いているが、この機能はブレンドシェイプが適用されていないメッシュには使用できない。ここで、ブレンドシェイプとは、あらかじめ用意された複数の表情モデルをパラメタの調整により組み合わせることで、さまざまな表情を作成するアニメーション手法をさす。今回用いたユニティちゃんは、デフォルトで口元にブレンドシェイプが設定されているため、モーフターゲットによる口元の制御は可能であるが、目配せや辛い表情などを再現するためのブレンドシェイプは、デフォルトの設定では不足している。そのため、あらゆる表情を再現するためには、他のソフトウェアを介してブレンドシェイプを設定する必要がある。

### 5.2.2 モーションの種類の向上

5.1節で述べたように、本手法により、現状では個人差のある動きを自然なアニメーションとして再現できない。その原因の1つとして、モーションの種類が少ないことが挙げられる。楽器を演奏する際の身体のモーションについては3.8.3項で触れたが、ここで述べたモーションだけでは足りない。例えば、音の高低に合わせて楽器を上下に揺らす演奏者や、円を描くように楽器を揺らす演奏者がいる。これらのモーションを追加し、モーションの種類を増やすことにより、さまざまな表現の実現が可能になり、より自然な吹奏アニメーションの自動生成が可能になると考えられる。

### 5.2.3 モーションと音の関連付け

本手法により個人差のある動きを自然なアニメーションとして再現できない原因として他に考えられるのが、モーションと音の関連付けが完璧でない点である。現在は、曲のテンポに従って身体が動く仕組みとなっており、モーションの種類や大きさは、ランダムに選択される仕様となっている。より自然な吹奏アニメーションを自動生成するためには、音の遷移とモーションを1対1で対応させたり、モーションを事前に指定するためのメタデータを用意することにより、モーションと音を関連付ける必要がある。

### 5.2.4 楽器の種類および演奏者の増加

本論文では吹奏楽の1つの演奏形態であるアンサンブルを想定し、楽器はトランペットとトロンボーンを選択したが、アンサンブルで使用される楽器はこの2本だけではない。また、将来的には大人数で曲を合奏している吹奏アニメーションの再現を目指す。そのため、楽器のモデルを増やす必要がある。3.8.1項の表3.1に示したような音と運指の対応表を作成することにより、トランペット、トロンボーンと同じように吹奏アニメーションの再現が可能となる。また、今回は演奏者としてユニティちゃんを選択しているため、全員同じ体型であるが、実際演奏者それぞれの体型は異なっている。体型に関わらず自然な吹奏アニメーションを生成するには、リターゲティングのための追加実装や、動きの差を記述するメタデータが必要となる。

## 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、さまざまな相談も聞いていただきました、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。特に、修士1年生の途中で研究テーマを変更するか否か迷っていた際に、多大な後押しをしていただきました。藤代教授の後押しがあったからこそ、後悔のない研究室生活を送ることができました。藤代研究室の一員として受け入れていただけたことを、改めて嬉しく思います。本当にありがとうございました。

本論文の執筆にあたり、貴重なご意見を頂きました、慶應義塾大学理工学部情報工学科の斎藤英雄教授、杉本麻樹准教授、杉浦裕太助教に深く感謝いたします。お陰様で、広い視野を持って本論文を執筆することができました。

本研究を進めるにあたり、実装面において多くのアドバイスを頂いただけでなく、コンピュータグラフィックス分野におけるさまざまな課題に出会う機会や、実際に業務に携わる機会を頂きました、株式会社デジタル・フロンティア開発部の皆様に深く感謝いたします。

本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く感謝いたします。特に、共同で研究を進めてくれた研究室学士4年の武内氏は、吹奏楽の経験がないにも関わらず、何度もディスカッションを重ねることにより知識を取得し、重心移動や合図などのモーションを作ってくれました。また同期の皆様は、時には優しく、時には厳しく、私の研究室生活を支えてくれました。このメンバーでなければ、ここまで楽しく充実した研究室生活を送ることはできませんでした。

最後に、本研究の評価アンケートにご協力いただきました、藤代研究室の皆様、株式会社デジタル・フロンティア開発部の皆様、お茶の水女子大学伊藤研究室の皆様、株式会社スクウェア・エニックス2018年新卒の皆様、東京個別指導学院高円寺教室の講師の皆様、応援指導部の皆様、慶應義塾大学塾生会館受付アルバイトの皆様、高校時代のクラスメイトの皆様に、厚く御礼申し上げます。

## 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, No. 1, pp.66–70, 2018 年 1 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, 2017 年 6 月.

## 参考文献

- [1] 3D EXPORT, 最終アクセス : 2018 年 1 月 22 日. [Online]<https://jp.3dexport.com/>
- [2] 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 1 月 22 日. [Online]<http://www.4gamer.net/games/032/G003263/20140903036/>
- [3] Epic Games: Unreal Engine, 最終アクセス : 2018 年 1 月 22 日.  
[Online]<https://www.unrealengine.com/ja/what-is-unreal-engine-4>
- [4] Free3D, 最終アクセス : 2018 年 1 月 22 日. [Online]<http://tf3dm.com/>
- [5] TAKABO SOFT: Domino, 最終アクセス : 2018 年 1 月 22 日.  
[Online]<http://takabosoft.com/domino>
- [6] unity-chan!, 最終アクセス : 2018 年 1 月 22 日. [Online]<http://unity-chan.com/>
- [7] unity-chan!, “ユニティちゃんライセンス条項”, 最終アクセス : 2018 年 1 月 22 日.  
[Online][http://unity-chan.com/contents/license\\_jp/](http://unity-chan.com/contents/license_jp/)
- [8] 井上 謙次, “Windows API による MIDI プログラミング”, 最終アクセス : 2018 年 1 月 22 日.  
[Online][http://www.deqnotes.net/midi/winapi\\_midiprog/winapi\\_midiprog.pdf](http://www.deqnotes.net/midi/winapi_midiprog/winapi_midiprog.pdf)
- [9] Wakana Asahina, Narumi Okada, Naoya Iwamoto, Taro Masuda, Tsukasa Fukusato, and Shigeo Morishima, “Automatic Facial Animation Generation System of Dancing Characters Considering Emotion in Dance and Music,” in *Proceedings of SIGGRAPH Asia 2015 Posters*, pp.11:1, 2015.
- [10] Christoph Bregler, Michele Covell, and Malcolm Slaney, “Video Rewrite Driving Visual Speech with Audio,” in *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques*, pp. 353–360, 1997.
- [11] Cosker Darren and James Edge, “Laughing, Crying, Sneezing and Yawning: Automatic Voice Driven Animation of Non-Speech Articulations,” in *Proceedings of Computer Animation and Social Agents*, pp. 225–234, 2009.
- [12] Steve DiPaola and Ali Arya, “Emotional Remapping of Music to Facial Animation,” in *Proceedings of the 2006 ACM SIGGRAPH Symposium on Videogames*, pp. 143–149, 2006.
- [13] Pif Edwards, Chris Landreth, Eugene Fiume, and Karan Singh, “JALI An Animator-Centric Viseme Model for Expressive Lip Synchronization,” *ACM Transactions on Graphics*, Vol. 35, No. 4, pp. 127:1–127:11, 2016.

- [14] George ElKoura and Karan Singh, “Handrix: Animating the Human Hand,” in *Proceedings of the 2003 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, pp. 110–119, 2003.
- [15] Tony Ezzat, Gadi Geiger, and Tomaso Poggio, “Trainable Videorealistic Speech Animation,” in *Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques*, pp. 388–398, 2002.
- [16] Junhwan Kim, Frederic Cordier, and Nadia Magnenat-Thalmann, “Neural Network-based Violinists Hand Animation,” in *Proceedings of the 17th Computer Graphics International*, pp. 37–42, 2000.
- [17] Tae-Hoon Kim, Sang Il Park, and Sung Yong Shin, “Rhythmic-Motion Synthesis Based on Motion-Beat Analysis,” *ACM Transactions on Graphics*, Vol. 22, No. 3, pp. 392–401, 2003.
- [18] Timothy R. Langlois and Doug L. James, “Inverse-Foley Animation: Synchronizing Rigid-body Motions to Sound,” *ACM Transactions on Graphics*, Vol. 33, No. 4, pp. 44:1–44:11, 2014.
- [19] Hyun-Chul Lee and In-Kwon Lee, “Automatic Synchronization of Background Music and Motion in Computer Animation,” *Computer Graphics Forum*, Vol. 24, No. 3, pp. 353–361, 2005.
- [20] Zicheng Liao, Yizhou Yu, Bingchen Gong, and Lechao Cheng, “Audeosynth: Music-Driven Video Montage,” *ACM Transactions on Graphics*, Vol. 34, No. 4, pp. 68:1–68:10, 2015.
- [21] Mark Pauly, “Realtime Performance-Based Facial Avatars for Immersive Gameplay,” in *Proceedings of Motion on Games*, pp. 23:23–23:28, 2013.
- [22] Tapio Takala and James Hahn, “Sound Rendering,” in *Proceedings of the 19th Annual Conference on Computer Graphics and Interactive Techniques*, pp. 211–220, 1992.
- [23] Thibaut Weise, Sofien Bouaziz, Hao Li, and Mark Pauly, “Realtime Performance-Based Facial Animation,” *ACM Transactions on Graphics*, Vol. 30, No. 4, pp. 77:1–77:10, 2011.
- [24] Yuyu Xu, Andrew W. Feng, Stacy Marsella, and Ari Shapiro, “A Practical and Configurable Lip Sync Method for Games,” in *Proceedings of Motion on Games*, pp. 109:131–109:140, 2013.
- [25] Jun Yin, Ankur Dhanik, David Hsu, and Ye Wang, “The Creation of a Music-driven Digital Violinist,” in *Proceedings of the 12th Annual ACM International Conference on Multimedia*, pp. 476–479, 2014.
- [26] Changxi Zheng and Doug L. James, “Harmonic Fluids,” *ACM Transactions on Graphics*, Vol. 28, No. 3, pp. 37:1–37:12, 2009.
- [27] Changxi Zheng and Doug L. James, “Rigid-body Fracture Sound with Precomputed Soundbanks,” *ACM Transactions on Graphics*, Vol. 29, No. 4, pp. 69:1–69:13, 2010.

- [28] Yuanfeng Zhu, Ajay Sundar Ramakrishnan, Bernd Hamann, and Michael Neff, “A System for Automatic Animation of Piano Performances,” in *Proceedings of Computer Animation and Virtual Worlds*, Vol. 24, No. 5, pp. 445–457, 2012.
- [29] 二宮 啓, 中澤 篤志, 竹村 治雄, “音楽に合った舞踊動作の自動生成”, 情報処理学会 *Computer Vision and Image Media* 研究会, No. 34, pp. 1–8, 2009.