VTuber

森田　恭平
Kyohei MORITA

1　はじめに

近年，動画投稿サイトが広く普及している．特に，YouTube は日本人の 72 %が利用しており，動画内に広告を挿入することで収益をあげることができる広告プラットフォームとしても注目を集めている [1)]．

YouTube に動画を投稿している人のことを YouTuber と呼ぶ．国内 YouTuber 市場規模は，2017 年には 219 億円となっており，前年と比べ 2.2 倍伸びている．さらに 2022 年には 579 億円に達すると予想されており，これからも成長を続けると考えられる．近年,YouTuber の新しい形として注目が集まっているのが，2D，3D のアバターを用いて動画出演を行う VTuber（Virtual YouTuber）である．この VTuber を支える技術として，モーションキャプチャーやトラッキング，リップシンクといった技術が存在し，これらの技術は日々進化している．本稿では，VTuber の技術，モーションキャプチャー，VTuber の撮影方法，今後の展望について述べる．

2　VTuber

VTuber とは，YouTube 上で活動を行う 2D や 3D のアバターのことを言う．VTuber が広がった背景には，アバターの作成やモーションキャプチャー機能，トラッキング機能，リップシンク機能を搭載したソフトが手軽かつ安価に使えるようになったことがある．

近年，個人だけでなく企業や自治体も，マーケティングとしての活用から VTuber に注目している．企業や自治体独自で VTuber を用いるメリットは二つある．一つは，タレントを起用することに比べ，企業や自治体のイメージを損ねるリスクが少ないことである．VTuber は，基本的にアバターの演者が非公開であり，起用したタレントが問題を起こすことで企業や自治体のイメージを損ねることがない．もう一つは，幅広い企画に採用できることである．タレントやアニメキャラクターには，既存のイメージがあるため，イメージに合わない企画に起用することは難しい．それに対し，VTuber はアバターの外見やイメージを自由に構成できるため，各企業や自治体に適切なイメージをアバターに持たせることができる．

3　VTuber の作成技術

3.1　モーションキャプチャー

モーションキャプチャーとは，人間の動きをディジタルデータ化し，コンピュータに取り込むための技法である．目や口，頭部，手足などにマーカやセンサを取りつけ，カメラでそれらの位置を計測し記録することで，トラッキングができる．トラッキングとはマーカやセンサを用い，演者の関節の動きや位置を追跡する技術である. 演者の動きを認識し，アバターと演者の動きをリンクするために用いる．各センサの動き方をコンピュータ上のアバターの部位に割り当てることで，実際の人間の動きをそのままアバターに反映できる．モーションキャプチャーの普及により，従来ではアニメータが数ヶ月かけて作成していたアバターの動作 CG を，一日で作成できるようになった．Table. 1 に各モーションキャプチャー手法の比較を示す．以降の章では各項目について詳細を述べる．

Table.1　各手法の比較

手法	価格	機材の多さ	拘束度	精度
iPhone	低	少	少	×
Web カメラ	低	中	少	△
VR	中	中	大	○
モーション キャプチャー	高	多	中	◎

3.1.1　iPhone を用いる方式

iPhoneX シリーズは TureDepth カメラが搭載されており，三次元空間を正確に読み取ることができる．赤外線で 3 万個の点を顔に照射し，その反射を読み取ることで顔の形を認識する．この方式では，iPhone のセンサを用いて手軽にフェイストラッキングができ，専用のアプリを使用することで，演者の表情をリアルタイムにアバターに反映できる．さらに手を動かすことや握ることも表現でき，ライブ配信やボイスチェンジ，背景の変更もできる．したがって，iPhone とアプリだけで動画の配信が手軽に行える．

3.1.2　Web カメラを用いる方式

Web カメラを用いる方式では，PC に接続した Web カメラから映像を取得し，トラッキングを行う方式である．Web カメラで顔のパーツを認識し，コンピュータ上で顔にマーカをつけるため，Web カメラのみでアバターを操作できる．この方式のメリットは，低コストで導入できることである．デメリットは，カメラの性能が低い場合に，アバターの動きが止まる可能性があることである．また，ソフトウェアのみで顔のトラッキングを行うため，他の手法に比べ精度は劣る．

3.1.3　VR 機器を用いる方式

VR 機器を用いる方式では，手持ちのセンサと HMD を用いてアバターの全身を動かす．演者の関節にトラッカーと呼ばれるセンサを装着すると足腰の詳細な動きをアバターに反映できる．VR の機器を Fig. 1 を示す. メリットは，後述のマーカやセンサを用いる方式より低コストで導入できることである．デメリットは，HMD をつけるた

め顔の表情の取得ができないことである．しかし，アイトラッキング機能を備えた HMD を用いると視線の検出ができ，アバターに表情をつけることができる．

Fig.1　VR の機器

3.1.4　マーカやセンサを用いる方式

マーカやセンサを用いる方式では，演者の全身にセンサを付け，全身のトラッキングを行い，アバターに反映する．Fig. 2 にモーションキャプチャーの様子を示す．この方式のメリットは，表情や全身の動きを細かくアバターに反映できることである．デメリットは，センサの価格が高価であることや，広いスペース，高性能の PC，大規模なスタジオ，多数のカメラ，専用機器，人件費が必要なことである．そのため，この方式は企業や自治体が用いることが多い．

Fig.2　マーカやセンサを用いる方式

3.2　光学式のモーションキャプチャー

光学式では，反射マーカをつけた演者を，複数の方向からカメラで撮影し，反射マーカの光からトラッキングを行う．このため，広い空間が必要であり，専用のスタジオや大学，研究機関で使用することが多い．この方式以外に磁気コイルを用いてトラッキングを行う磁気式や，撮影した映像からトラッキングを行うビデオ式などがある．光学式は，データ精度，汎用性，拡張性，演者の拘束性などから，ビデオ式，磁気式などより，総合的に優れており，モーションキャプチャーの主流方式である．この方式は，演者の拘束性が低く，同時に複数人のキャプチャーができるが，反射マーカがカメラから見えていなければデータが取れない．現在では，隠れた反射マーカの位置を他のマーカの位置から補間する機能も開発され，反射マーカの見えない位置のトラッキングが可能となりつつある．

3.3　リップシンク

リップシンクとは，映像上のアバターの口の動きとセリフの音声を合わせることである．リップシンクの方式には，大きく分けて画像認識，音圧の大小，口形素推定の三つがある．一つ目の画像認識のよる方法は，演者の口の映像から口の形を認識し，アバターに反映する方法である．口の形が正確に認識できるため，口の動きを正確に表現できる．二つ目の音圧の大小による方法は，音声の音量を口の大きさに反映する方式である．この方式は，話していないときは口を閉じ，声を出すと音量に応じて口を開くため，収録された音声のみで口の動きを表現できる．三つ目の口形素推定は，音声認識によりどのような音を発しているのかを推定し，口の形に反映する方法である．どの音を発しているかを認識し，その音を発する口の形を決める．収録した音声のみで口の動きを表現でき，音圧による方法より口の動きが正確である．

4　今後の展望

2018 年にグリー株式会社が VTuber 事業の 40 億円の投資を行ったことから，VTuber 市場は今後も注目を集め，ユーザーを増やしていくことが考えられる．さらに，2019 年には，視線追跡機能を搭載した VIVE Pro Eye により，HMD をつけたまま，アイトラッキングができる環境も整っている．したがって，今後も多くの企業や自治体が独自の VTuber を作成することが予想される．

また，有名な VTuber は，トークＬＩＶＥの開催や他企業とのタイアップイベントも開催し，YouTube 以外の場でも活動の場を広げていると考えられる．VTuber は日本独自のコンテンツとして世界から注目されており，今後はエンターテイメントやマーケティングの分野で世界に進出していくと予想される．

参考文献

1) 17　年の国内　YouTuber　市場規模　https://www.itmedia.co.jp//news/articles/1801/30/news109.html 参照　Apr.23.2019
2) VTuber などに使われる主要なリップシンク方式の比較メモ https://qiita.com/nkjzm/items/5fb4f4dbcdaa9bc7cc34 参照　Apr.23.2019
3) データで見る VTuber https://biz-arts.com/2018/06/12/vtuber/ 参照　Apr.23.2019
4) 【3 分でわかる！】モーションキャプチャの仕組み https://xr-hub.com/archives/866 参照　Apr.23.2019