

ACLプロジェクト報告書

大島正嗣

1. はじめに

本プロジェクトでは、表現のためのソフトウェアおよびプラットフォームについての探求を行った。具体的には、以下の技術について導入・試用した。

- VRトラッカー (VIVE Tracker)
- ハンドキャプチャ (ContactGlobe2)
- 超指向性スピーカ

これらの技術は、個別にも興味深いが、お互いに組み合わせることもできる。VRトラッカーは三次元的な位置や回転を取得するデバイスであるので、全身に10個程度装着し、これとハンドキャプチャを組み合わせれば、ほぼ全身+指先の動きまでもキャプチャする（フル・トラッキングと呼ばれるレベルの）モーションキャプチャが構成できる。（人間の表情や目線についてのキャプチャ技術もACLにはすでにある）。また、VRトラッカーとカメラを組み合わせれば、実在の人物をバーチャル空間内に合成するバーチャルプロダクションが実現できる。超指向性スピーカは、空間全体に音を流すのではなく、いわば音の場の構造化を可能にする技術であるが、これにVRトラッカーを組み合わせれば、音場の動的な変化をつくることが理論上可能になる。

本報告書では、これらの技術について、順に、試行と可能性について、述べる。

2. HTC VIVE Tracker

HTC VIVE Trackerは、HTC社のVRゴーグルを装着した際に、被験者が身体ごとVR空間に没入するためのモーショントラッキング装置である。モーショントラッキングには各種方式があるが、VIVE Trackerは光学式のモーショントラッキング装置に属する。各方式にはそれぞれ一長一短があるが、光学式のVIVE Tracker の場合は、比較的精度が高く、また人間以外の「物」などに取り付けることにより、3次元空間上の「物」の位置を利用できる、などの利点がある。

2.1. motion capture としての使用法

まずはモーションキャプチャとしての使用法について述べる。

モーションキャプチャには、次の3つがある： - 光学式 - 慣性式 - AIによる姿勢推定 以下順に特性を述べる。

光学式モーションキャプチャ

光学式モーションキャプチャは精度がよいことで知られる。著名なシステムには OptiTrack などがある。手足の位置などを光学的に捉えるため、例えば「両手を合わせる」などが行える（慣性式では難しい）。VIVE Trackerの動作は次のような物である。ライトハウスと呼ばれるレーザーを発する装置を数台準備し、身体側にはTrackerと呼ばれる受光部を6～10個装着する。これにより、ライトハウスからのレーザーを基準に、各Trackerの位置が計算される。

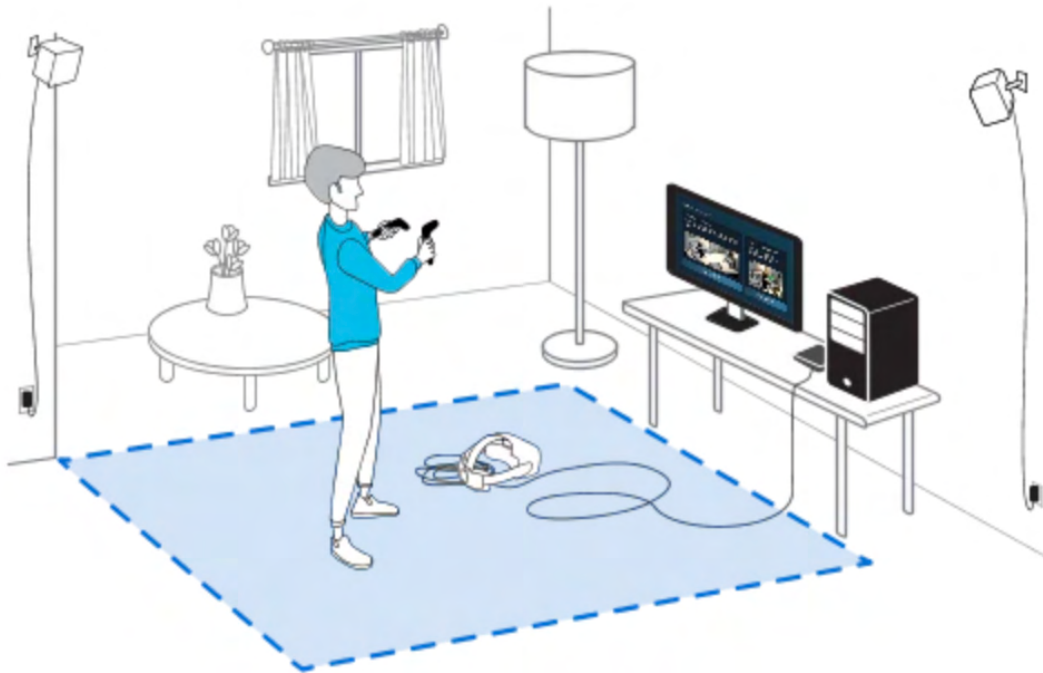


Figure 1: VIVE利用の様子

ライトハウスの光が及ぶ範囲は 7m x 7m 程であり、比較的広い範囲を移動しつつモーションキャプチャすることができる。また、Trackerの位置精度は 0.3cm~1cm程度と高い。動きのなめらかさであるfpsであるが、通常30fps 程度であるものの、最大 250fps と述べている資料もある。デメリットとしては、どうしてもライトハウスの光が受光できない場合が起こりうるため（オクルージョン）、その対策（例えばライトハウスの数を増やすなど）が必要な点である。

慣性式モーションキャプチャ

慣性式モーションキャプチャの代表的なものは Sony の mocopi などである。これは、被験者の手足に慣性式のセンサ（IMU）を配置し、センサに感じられた加速度をもとに、センサの現在位置を推定するものである。この方式のため、測定が長時間にわたると、次第に誤差が蓄積してしまうのでリセットする必要がある。しかし、メリットとしては、光学式に必要とされるライトハウスが不要な点と、更に、加速度だけが観測できれば良いため、たとえば「布団にもぐりこむ」や、「コートを羽織る」など、光学式センサーがライトハウスの光を受光できないようなシチュエーションであっても、モーションを記録できるというメリットがある。

AIによる姿勢推定

Webcamで記録された動画をもとに、AIによる3次元姿勢推定を行うことにより、モーションキャプチャを行う方式も存在する。著名なものには、ThreeDPoseTracker などがある。極めて簡便であるが、被験者の画像のみをもとに推定するためか、「被験者の足が接地している」感じが出ないことが多い。

デプロイの様子

実際に HTC VIVE Trackerを用いてモーションキャプチャを行った。適当な広さの教室にライトハウス4台を設置して測定を行った。



Figure 2: ライトハウスの設置



Figure 3: ライトハウスの設置 2

VIVE Trackerには一台ずつ、データを受信するアンテナが必要になる。Trackerは10台用意したが、電波の混線を避けるため、アンテナ間に距離を置くことが推奨されていた。実際には、20cm～30cm ずつ離して設置したが、特に問題はみられなかった。



Figure 4: アンテナ

実際にキャプチャしてみたのが次の図である。体感としてはfpsの高さによるなめらかさが印象的であった。



Figure 5: 動作画面

ライトハウス数の違い

ライトハウスはVIVEでは最低2台から動作するとされている。しかしながら、ライトハウスを2台、6台、8台と変えて、キャプチャーされた動作のなめらかさを比較したところ、8台が最もなめらかだったとの報告がある。(フルトラ研究所 2024)



Figure 6: 「ベースステーション1.0x2個と2.0x2個と2.0x4個だと動きはどう変わるか？」

ただし、ライトハウスを増やした場合、レーザーの乱反射により精度が下がるため、収録する部屋での反射対策が必要になるとされる。

トラッキングポイント数の違い

今回、VIVE Trackerは11個用意した。基本的には6点あれば最低限の動作はカバーできるとされる。

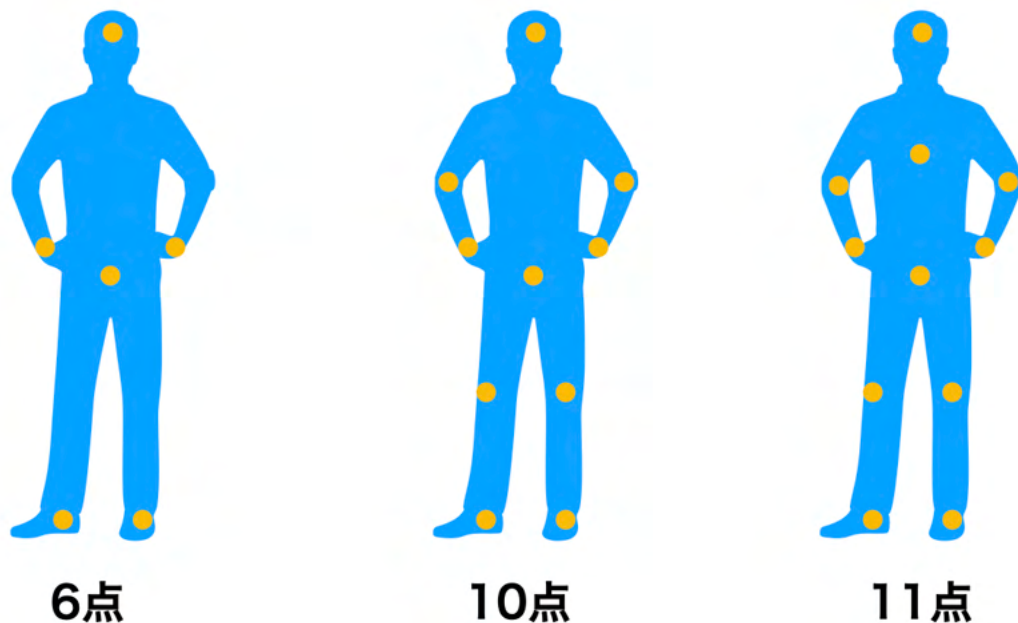


Figure 7: トラッキングポイントの違い

6点であっても、インバースキネマティクスなどを使って、手首の位置から腕の配置を推定できるからである。しかしながら、6点では肘の情報がないため、「手首の位置は同じだが肘が上下する」などの動きはキャプチャーすることができない。更に最近のVRChatなどが対応している11点トラッキングであるが、10点トラッキングに加え、「胸」にトラッキングポイントが増えているのが特徴である。この「胸」と「腰」のセンサーのずれによって、「上半身だけねじる」などの動きをキャプチャー／表現することができると思われる。(Kawasaki Silvia 2023)

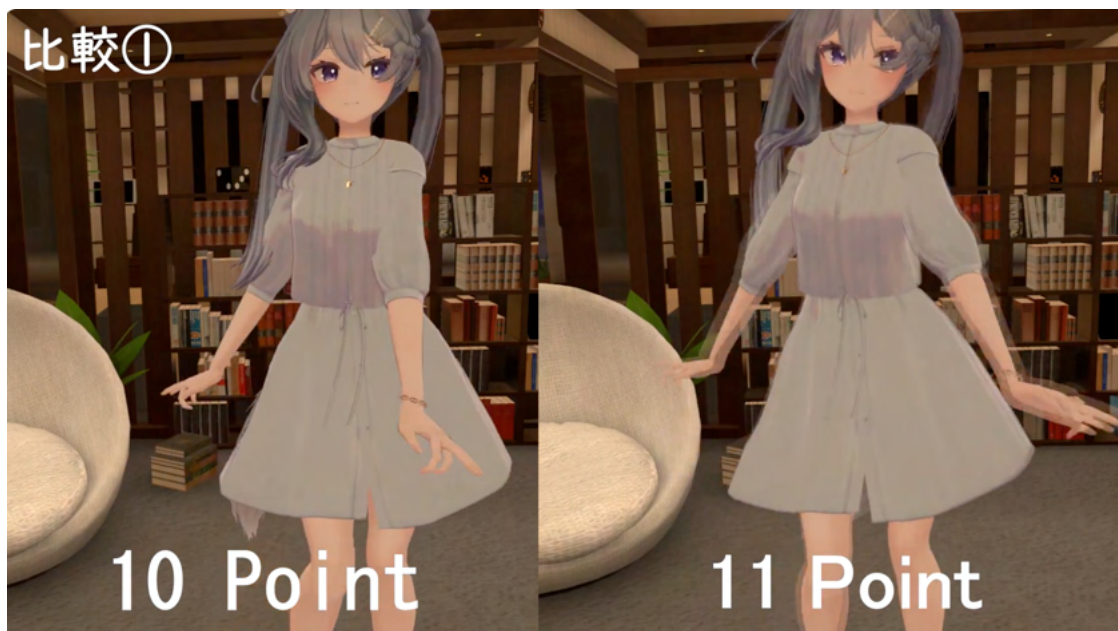


Figure 8: 「11点がシン・フルトラ！ VIVEトラッカー フルトラッキング full body tracking フルトラッキング」

2.2. virtual production としての使用の可能性

VIVE Tracker の用途はモーションキャプチャだけに限定されるものではない。ライトハウスからのLEDが届く範囲での、3次元的位置や角度を測定できるデバイスであり、HTCから提供されている“Vive Tracker developer guidelines”[@-VIVETrackerGuidelinesDeveloper] には、VRゲーム中で使用される機材（ラケット、銃、、）などの例が示されている。

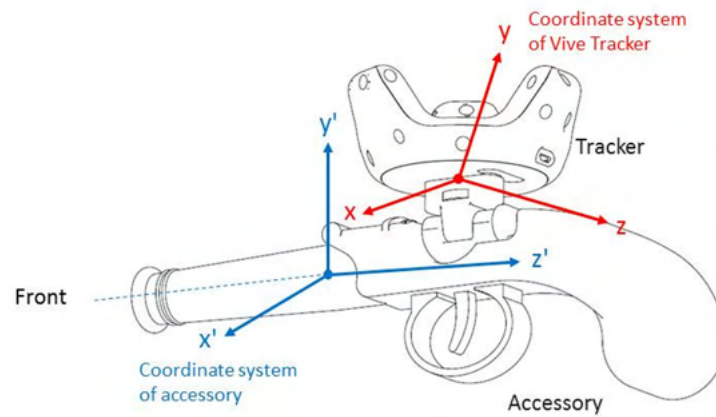


Figure: Example to integrate accessory and Vive Tracker

Figure 9: VIVEトラッカーを取り付けた銃

このVIVE Trackerをカメラに取り付けると、いわゆる Virtual Production におけるヴァーチャルカメラを実装することができる。

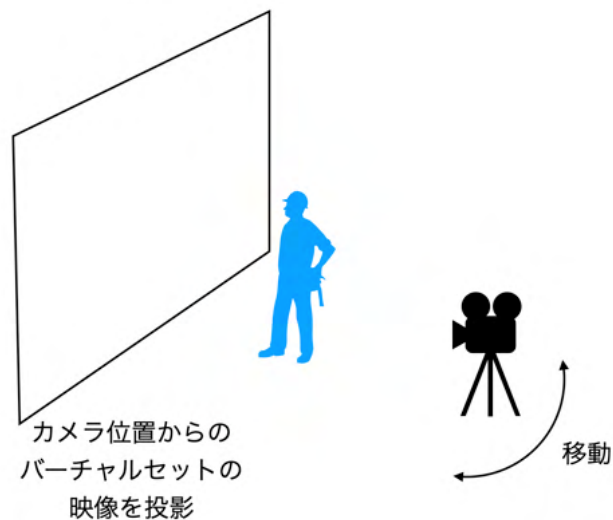


Figure 10: ヴァーチャルプロダクション

商業的な撮影ではLEDパネルで背景を組み、その前で演技者がパフォーマンスを行い、その様子を VIVEを装着したカメラで撮影する。カメラの位置・移動情報からバーチャルなセット内での映像を生成し、それをLEDパネルに表示することにより、セットの立体感やカメラが移動した時の視差をそのままに撮影を行うことができる。もっともLEDパネルは高価であるので、撮影セットを合成用のグリーンバックで覆い、バーチャルなセット内での映像を合成することにより同様の撮影ができる。

以下は BBC のコメディーターなどを手がけた Small and Clever Production による映像である。(Phillip Moss 2022)



Figure 11: ニュースの合成



Figure 12: ニュースの合成 2

2.3. 空間のデジタル化

VIVE Trackerからの情報を取得するPythonライブラリも存在する。このようなライブラリを使用すれば、VIVE Trackerは空間をデジタル化するデバイスとして使用できる。

先の“VIVE tracker developer guidelines”には、VIVE トラッカーからのデータ内容なども記載されている。この情報をもとにpythonなどのプログラミング言語から処理できるようにしたライブラリとして、libsurvive (<https://github.com/collabora/libsurvive>) などがある。

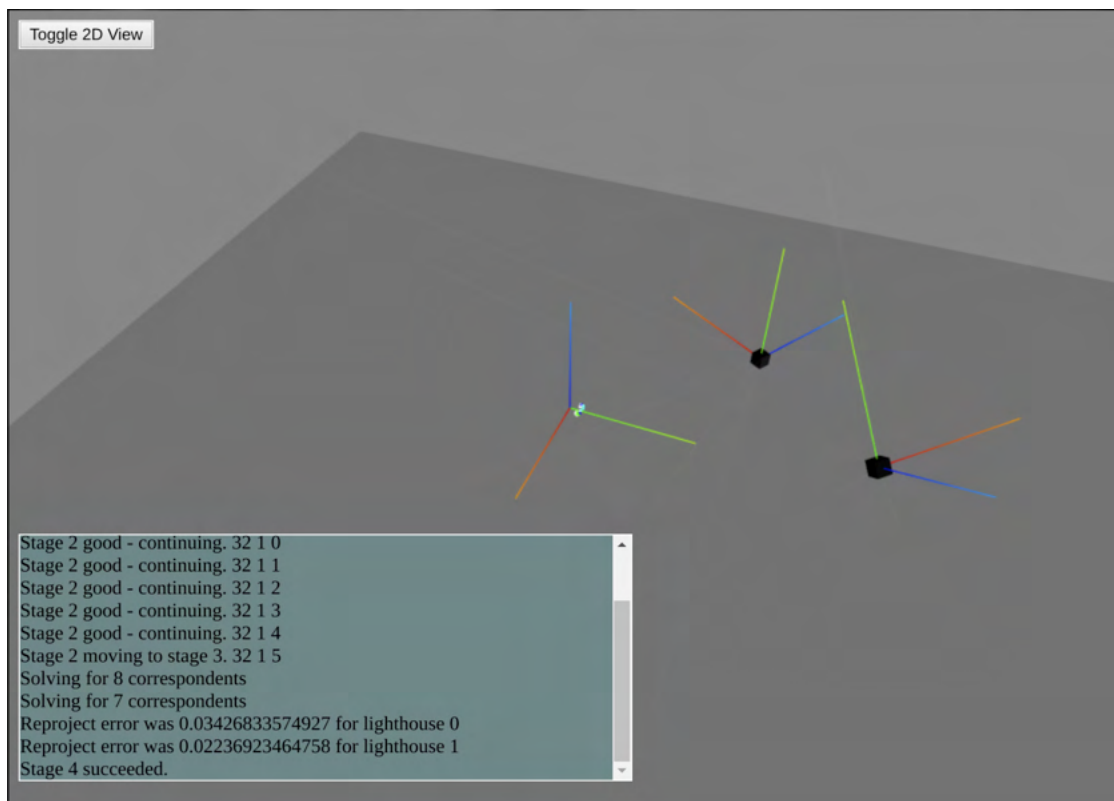


Figure 13: libsurviveの様子

このようなライブラリを使うことで、Pythonのプログラムから容易にVIVEトラッカーの位置情報を利用することができる。ロボティクス分野での応用が考えられるが、例えばドローンにVIVEトラッカーを設置し、トラッカーからの位置情報をもとにドローンをフィードバック制御するなどができる。

3. Diver-X ContactGlobe2

身体のモーションキャプチャは VIVE トラッカーをはじめとする前述の種々の方法があるが、例えば演技のようなものを再現しようと思った時には十分ではない。演技者の表情や、指先の動きを再現する必要があると思われる。

Diver-X 社の ContactGlobe2 はハンドトラッキングを行う機材である。ハンドトラッキングにも種々の方法があり、一般的なモーションキャプチャと同じく、慣性式と光学式

がある。そして ContactGlobeは 曲げセンサ式であり、安価にハンドトラッキングを行える可能性がある。

慣性式ハンドトラッキング 慣性式ハンドトラッキングは慣性式モーションキャプチャと同じく、指先に慣性センサ（IMU）を配置し、そのデータでキャプチャを行うものである。デメリットは慣性式のモーションキャプチャと全く同じで、時間が経つにつれ誤差が集積してしまう点である。ACLでは、かなり早い時期からNoitom社のHi5という機材を用意しているが、地磁気の影響なども受けやすく扱いが難しい。



Figure 14: Noitom Hi5

光学式ハンドトラッキング 光学式のハンドトラッキングとしてはGoogleの MediaPipe が有名である。このモジュールは他の様々な製品にも使用されている。

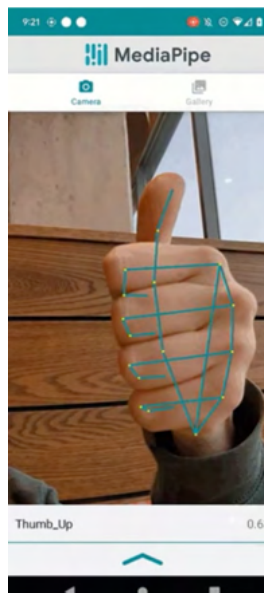


Figure 15: MediaPipeの様子

これを応用したものとして、例えば Webcam motion capture がある。



Figure 16: WebcamMotionCapture

ContactGlobe2

ContactGlobe2は「曲げセンサ」方式を取っているため、慣性式の誤差の蓄積などはないはずであり、光学式にみられるオクルージョンの影響もないはずである。製造元である Diver-X 社のホームページでは、good luckのゼスチャー（指を交差）すら再現できるとされた。

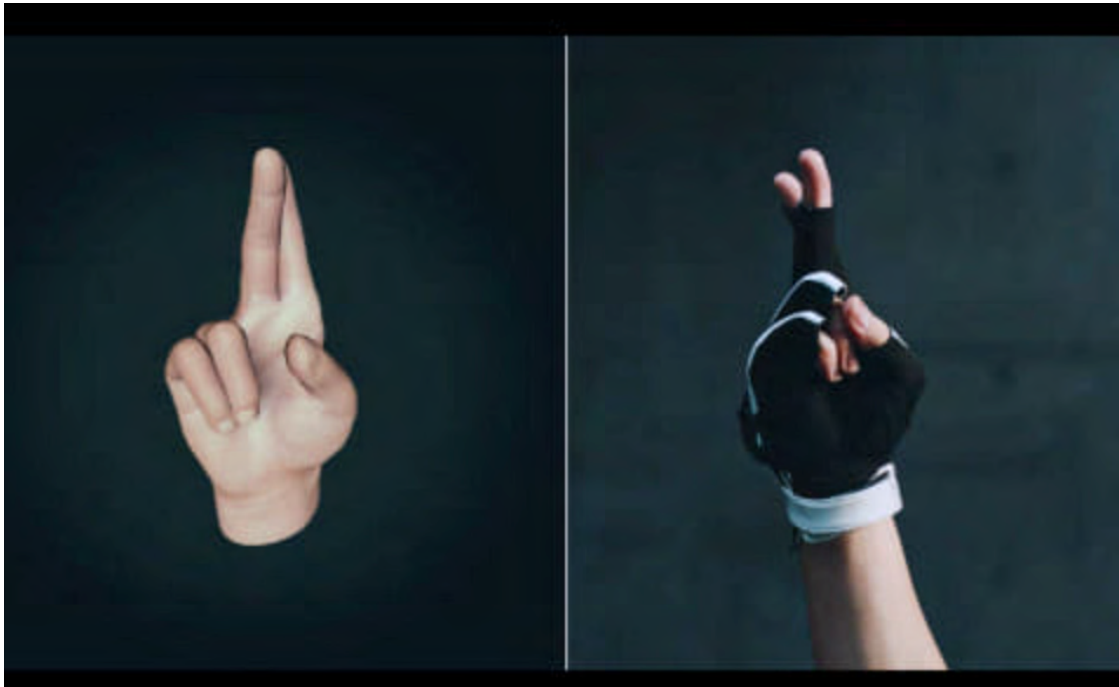


Figure 17: ContactGlobe2のイメージ

実際に動作させてみたところ、なかなか表現に富む指先のキャプチャーは難しいことがわかった。特にドライバソフト ver 1.7まででは、「指をすぼめる」のような動きができなかった。ようやく ver1.8 になって若干すぼめることができるようになった。



Figure 18: ver1.7での動作



Figure 19: ver1.8での動作

もっとも、この「すぼめる」動作も、「指の開きやすさ」パラメータを調整してできるようになったレベルであり、「曲げセンサー」によるハンドトラッキングは、個々の指の折り曲げレベルであれば安価にキャプチャできるデバイスといえるが、横方向の指の動き（ひろげる／すぼめる）についてはなかなか現時点では難しいという感想を持った。

4. 超指向性スピーカ

超指向性スピーカはパラメトリックスピーカともいい、音を直進させるスピーカである。構造としては可聴域の音声を、超音波の高さに変換し、複数の小さなスピーカ（アレイスピーカ）から音を出す。超音波は直進性が高いため、「直進する音」が作りだされる。また、空気の非線形性により超音波同士が干渉し、（超音波同士の差分が「うなり」として可聴域になるので）可聴域の音が復元される、というものである。

用途としては、たとえば美術館の特定の作品の前だけに解説音声を流す、であるとか、鉄道の券売機の前だけに操作説明の音声を流す、など、特定のユーザにむけてのみ情報を伝えたい場合に利用される。また、超音波の直進性により、音エネルギーが拡散しにくいので、同じ音量でも、より遠くまで音声を届けられるという特性もある。

さらに、超音波を壁などに反射させることにより、壁において超音波の干渉が発生するため、「音が壁から発している」ように聞こえるなどの現象もある。（この現象はたとえばディズニーランドで、ミッキーの声がエリアのあちらから聞こえたり、こちらから聞こえたり、というアトラクションに使用されていると推定される）。

このように超指向性スピーカは、音の場を構造化することができ、メディア表現に使える可能性があるため、実際に4セット購入し、効果を確認した。

購入したのは、秋月電気通商で扱っている「パラメトリックスピーカ」である。

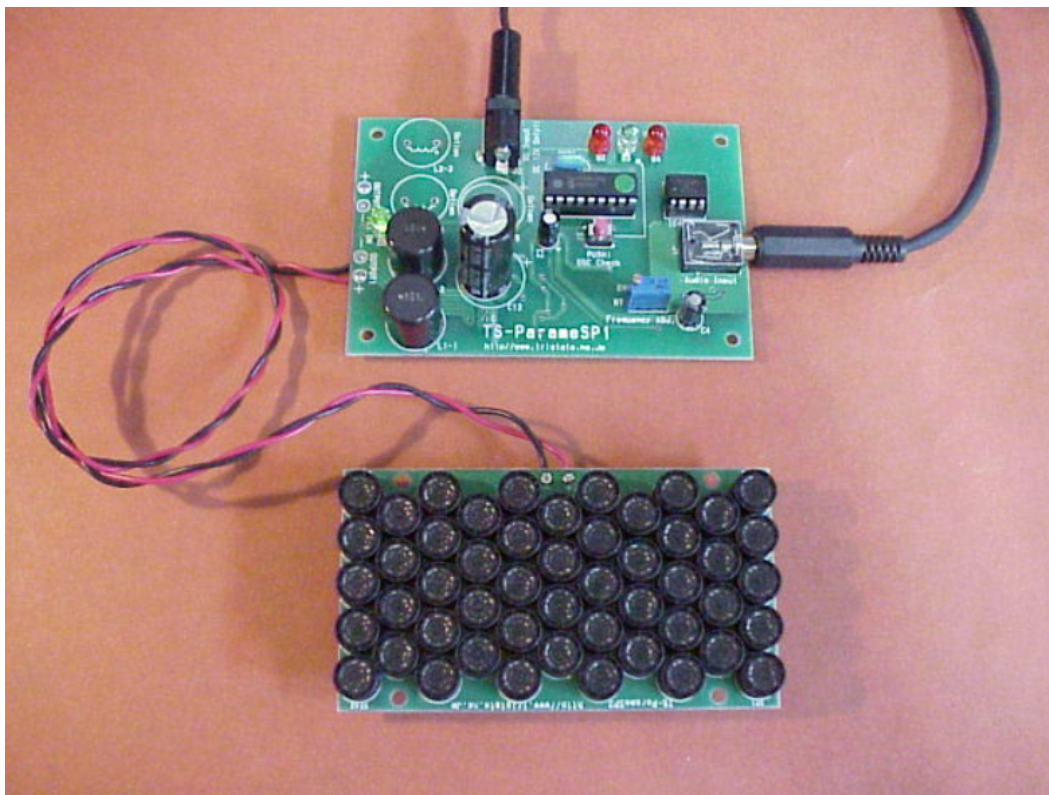


Figure 20: パラメトリックスピーカ

マニュアルによれば、約20度の範囲に可聴範囲を絞り込むことができる。

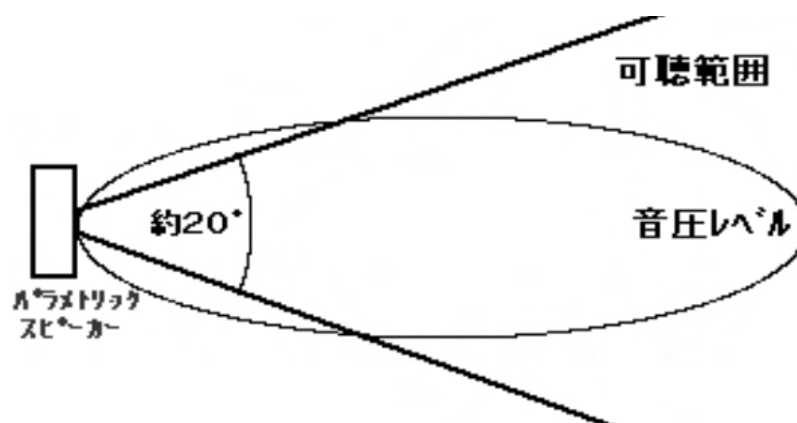


Figure 21: 可聴域

また、つぎのような利用例が掲載されている。



Figure 22: 利用例

今回は、新たな使用方法として、音声によって歩行者をガイドすることができるかを実験した。超指向性スピーカで音を浴びた場合、極めて不思議な感覚で音声聞こえることになる。その音声で、「音声の聞こえる方向へ進め」など指示することにより、被験者に任意のコースを歩ませることができないだろうか。もしこれが可能になれば、例えば、ギャラリーなどの展示室が、ほぼ暗闇の場合であっても、まるで神曲で天使がガイドするように人を導くことができる。

実験は音の反射を避けるために、短大の中庭で行った。概念図を以下に示す。このように音のビームを辿らせることにより、任意のカーブを人を誘導できるかを試した。

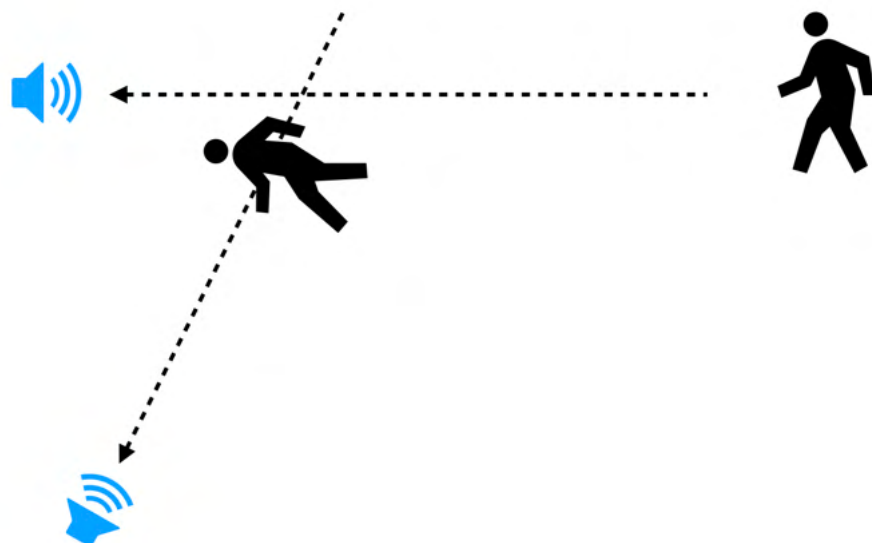


Figure 23: 誘導の構想

まず一台の超指向性スピーカーを用いて、一本の音のビームがどの程度の距離を持ちうるかを試した。その結果、今回使用した秋月電子の機材では、短大の中庭を横切る、ほぼ30m近くから音に気づかせることができた。予想外の長距離であった。



Figure 24: 設置の様子

続いて、2台の超指向性スピーカを用いて、2本の音のビームを交差させ、1本目の音のビームから2本目の音のビームに被験者を誘導できるかを実験した。その結果、確かに1本目の音も2本目の音も聞こえるが、逆に、1本目から2本目に「乗り換える」ことが伝わりにくく、被験者を誘導する目的では改善が必要であることがわかった。つまり、2本のビームの交差点に到達しても、1本目の音がまだ聞こえているため、そのまま1台目のスピーカに近づいていってしまう状況であった。改善策としては、被験者が曲がるべき位置に到達した場合に、1台目のスピーカの音を切る手立てが必要である。これは簡単なマイクロコントローラと距離センサで実現できる。安価な超音波距離センサであっても4m程度であれば検知できるので、スピーカから交差点までの距離で1台目の音を切ることができる。以上のように、超指向性スピーカを用いて、音の場の構造化・分節化ができると思われる。その際に、距離センサを用いて制御しても良いし、あるいはVIVE Trackerの節で述べたような方式をもちいて、VIVE Trackerの位置によってスピーカーを制御する方法もあり得るかと思われる。

5. まとめ

今回、試用した技術はいずれもなんらかの形で作品製作に利用できると思われる。また、今回は、ソフトウェア的なサービス（例えば、動画から登場人物をCGに置き換える Wonder Studio や、生成AI関連の様々なサービス）については試せなかった。特に生成AI関連のサービスは進歩も早いが何らかの形で試行すべきと思われる。

参考文献

Kawasaki Silvia (Director). (2023, March 26). 11 point tracking 【VRChat】 11点がシン・フルトラ！ VIVEトラッカー フルトラッキング full body tracking フルトラッキング [Video recording]. <https://www.youtube.com/watch?v=vvIINKGDKVU>

Phillip Moss (Director). (2022, January 29). How we made a Comedy Series for the BBC using Virtual Production [Video recording]. <https://www.youtube.com/watch?v=DQT0Qy856mA>

フルトラ研究所 (Director). (2024, January 5). 【比較動画】 ベースステーション1.0×2個と2.0×2個と2.0×4個だと動きはどう変わるか？ [Video recording]. <https://www.youtube.com/watch?v=3jgLr9-X4EU>