

超指向性スピーカを用いた 仮想オブジェクトの存在感向上手法の検討

下田 桂輔^{1,a)} 福島 拓^{2,b)}

概要： マーカーレス AR を用いて仮想オブジェクトが仮想空間内を自由に移動するコンテンツが存在する。ユーザは動き回る仮想オブジェクトに追従して、AR デバイスのカメラを向ける必要があるが、手がかりがカメラ映像のみのため仮想オブジェクトを見失うことがある。本研究では、仮想空間上の音像を超指向性スピーカによって現実空間に提示することで、AR 体験中に仮想オブジェクトの定位感や実在感の向上が可能かどうかを評価した。本稿の貢献は次の通りである。(1) 仮想オブジェクトはその場で静止しているものより動き回るもののほうが、実在感と定位感が得られやすい可能性があることを示した。(2) 超指向性スピーカの音が反射する位置とユーザの位置とが近距離の場合に実在感が得られやすい可能性があることを示した。(3) 短時間の音を瞬間的に提示するとき、超指向性スピーカはヘッドホンを用いたバイノーラル音源より定位感が得られにくい可能性があることを示した。

キーワード： AR, 音源定位, 実在感,

1. はじめに

近年、AR 市場は拡大し続けており、AR 対応デバイスもそれに伴い普及していくと思われる [1]。スマートフォンゲーム [2] やバーチャルライブ [3] などのエンターテインメント用途や商品のプロモーション用途 [4]、教育・訓練用途 [5] などに利用される事例も多く、一般的なユーザも AR に触れる頻度が高まってきた。最近では、カメラで取得した画像の特徴量から空間認識を行うマーカーレスの AR システムが主流となっており、自由度の高い表現が可能となりつつある。

このとき、ユーザは自身のスマートフォンを用いて AR 体験をする場合が多いが、スマートフォンの画面サイズでは視野角が狭く、仮想空間上に存在する仮想オブジェクトを十分に見ることが出来ないといった問題がある。特に、空間内を動き回るような仮想オブジェクトが出現する場合、その物体がどこに出現しているのか分からなかったり、見失ったりすることがある。

そこで、適切な方向に AR デバイスを向けるよう促すために「音の AR (Sound Augmented Reality)」がよく用い

られる。これは、適切な場所に音像を提示することにより、仮想オブジェクトの存在感や定位感を表現することが出来るものである。音の AR で用いられる手法としてヘッドホンやイヤホンによるバイノーラル方式や、多チャンネルスピーカによるサラウンド方式があるが、移動するユーザや複数人のユーザへの音像提示が難しい。また、ヘッドホン等をつけたり、スピーカの位置を調整したりする手間がかかることで、手軽さに欠けることが課題である。さらにヘッドホン等をつけたユーザは現実空間上の周囲の音が聞こえないため、他者とコミュニケーションをとったり、周囲の状況に気づいたりすることが難しい。特に周囲の音が聞こえない状況でデバイスの画面を注視しているため危険が伴う。

本研究では、移動するユーザや複数人ユーザで音像を共有することで、エリア内の任意の場所で手軽に AR コンテンツを楽しめるシステムの開発を目指している。本稿では、仮想オブジェクトを現実空間に重畳表示する AR アプリと超指向性スピーカを用いて、仮想オブジェクトの位置に音像を提示するシステムを作成した。また、このシステムで AR 体験を行うユーザが仮想オブジェクトの定位感や実在感を得られるかどうかを評価した。

¹ 大阪工業大学大学院情報科学研究科

² 大阪工業大学情報科学部

^{a)} m1m19a15@st.oit.ac.jp

^{b)} taku.fukushima@oit.ac.jp

2. 関連研究

移動する仮想オブジェクトと任意の場所に位置するユーザに対応する音像提示手法として、サラウンド音響を動的に再現するものと超指向性スピーカを用いたものが存在する。

2.1 サラウンド音響による音像提示

一般的なサラウンド再生システムとして、映画館などのサラウンド音響システムがある。このシステムでは、スピーカの位置の固定や視聴者の位置を想定した最適化などがされているため、特定の環境下でしか効果を発揮しない。また、ヘッドホンなどで用いられるバイノーラル手法では、パーソナライズした音を提示するため、視聴者の位置に合わせた音源をそれぞれ生成しなければならない。さらに、それぞれがヘッドホンをするため、同じ体験をしている人同士でコミュニケーションが取りにくい、認識の齟齬が生じやすいといった問題がある。そこで金子らは複数のスマートフォンを同期し現実空間への音像定位を動的に行うシステム SymPhonicAudio[6] を提案している。評価実験において、提案手法とバイノーラル手法で比較した結果、音の方向性のわかりやすさはバイノーラル手法の方が優れていたが、コミュニケーションの取りやすさや一体感があるという点は提案手法のほうが優れていた。SymPhonicAudioの特徴として「体験の共有」「多人数に対応」「手軽さ」を挙げており、この観点は本研究でも共通している。

2.2 超指向性スピーカを用いた音像提示

超指向性スピーカ [7] は超音波を使って音に直進性をもたせることができるスピーカである。可聴音を直進性の高い超音波に変調して空気中に放射すると、空気中で復調されスピーカを向けた方向に可聴音を届けることができるものである。

伊藤ら [8] の研究では、聴覚的 MR の提示方式として音像ブラネタリウム方式を用いた際の移動音源の実現方法を提案している。これは、ユーザの周囲に複数の反射面を用意し、超指向性スピーカで反射面へ音声を投射する構造となっており、反射面から反射した音声は仮想オブジェクトの位置から聞こえるように補完する研究である。しかし、ユーザの移動や複数人での利用を想定しておらず、スマートフォンによる AR 体験には向かない。

北岸ら [9] は、講師・受講生間などの一人対多人数コミュニケーションを目的として、影と移動軌跡の描画による視覚定位と、移動する足音音源による音源定位で表現される講師の分身エージェントを提案している。プロジェクタで影と移動軌跡の描画、超指向性スピーカとサーボモータで足音音源の提示を行っているため、仮想オブジェクトと音



図 1 システム構成

源の動的な提示が可能となっている。この研究では、講義での利用を想定しているために影と移動軌跡、足音のみで分身エージェントの位置を表現している。本研究ではその利用シーンを拡張し、ARで仮想オブジェクトが表現された場合や、足音以外での音源でも音像定位が可能かを確認する。

3. 提案システム

3.1 システム概要

本研究では、仮想オブジェクトの実在感を感じられるAR体験が手軽にできるシステムを目指している。よって、次の3つの要件を元にシステムを設計した。

- (1) スマートフォン（カメラ・ディスプレイ）のみで利用できる
- (2) エリア内の任意の場所で利用できる
- (3) 複数人で同時に利用できる

この要件を満たすために、超指向性スピーカを用いてARオブジェクトの位置に音声を提示する。これにより、手軽にAR体験が可能となるだけでなく、同時に体験している複数のユーザとコミュニケーションをとりながらの体験や、様々な場所へ移動しながらの体験ができるようになる。また、1つの音源のみを超指向性スピーカで再生すればよいので、ユーザに最適化された個別の音を生成する必要がない。

システム構成を図1に示す。作成したシステムのAR部分はiOSアプリとして動作する。まずiPhoneのカメラで空間キャリブレーションを行うことで、任意の場所に仮想オブジェクトを出現させることができる。仮想オブジェクトにはアニメーションが設定されており、アニメーションが再生されるタイミングで音声を超指向性スピーカか

ら出力する。また、画面上に表示される仮想オブジェクトの位置を取得し、超指向性スピーカの向きを随時変更することで、音像が仮想オブジェクトの移動に合わせて追従する。これにより、移動するユーザや複数人ユーザで音像を共有することができる。iOS アプリはゲームエンジンである Unity で作成し、ARFoundation[10] から ARKit[11] を利用することで、AR を動作させるための機能を実装した。超指向性スピーカは K-02617 (有限会社トライステート) を用いた。超指向性スピーカの向きは、舞台照明で使われるムービングヘッド LM70S (BETOPPER) にスピーカを取り付け、TouchDesigner[12] で制御することで実現した。システムの各機能について以降の各節で説明する。

3.2 仮想オブジェクト生成機能

予め用意した 3D モデルを任意の座標に生成する機能である。仮想オブジェクトの出現位置 (アニメーションの開始位置) と実空間に設置したスピーカの位置を指定する。キャリブレーション時に床面や壁面を認識し、タップした平面部分に生成されるようにした。実験システムでは、生成できる仮想オブジェクトは 1 種類につき 1 箇所のみとしている。

3.3 アニメーション再生機能

生成した仮想オブジェクトのアニメーションを再生する機能である。仮想オブジェクトには予めアニメーションを設定している。再生する仮想オブジェクトを選択した後、Play ボタンを押すことで再生される。アニメーションはループせず、Play ボタンを押すたびに 1 回再生される。このとき、音声の再生指示を TouchDesigner に送るため、音声も同時に再生される。

3.4 超指向性スピーカでの音声再生と方向制御

超指向性スピーカの方向を制御するために、舞台照明などで利用されるムービングヘッドにスピーカ本体と制御基板を取り付け、ムービングヘッドに DMX 信号を送信することで、任意の向きにスピーカの音声を出力することを可能にした。AR アプリケーション上で計算した仰角と方位角を、PC 上で動作する TouchDesigner に OSC 通信で送信し、DMX 信号に変換してムービングヘッドに送出している (図 1)。ムービングヘッドの方位角は 540 度まで回転できるが、実験では 180 度に制限した。超指向性スピーカの座標を O 、仮想オブジェクトの座標を P 、 P の y 座標を O_y まで移動した座標を Q 、 O の x 座標を P_x まで移動した座標を R とすると、超指向性スピーカの仰角 (俯角) Yaw と方位角 $Pitch$ は次の各式で算出できる。

$$Q = (P_x, O_y, P_z) \quad (1)$$

$$R = (P_x, O_y, O_z) \quad (2)$$

$$Yaw = \arctan \frac{OQ}{PQ} \quad (3)$$

$$Pitch = \arctan \frac{QR}{OR} \quad (4)$$

4. 実験

今回は本システムを用いることで、従来手法であるヘッドホンでの音像提示 (バイノーラル手法) と比較して、仮想オブジェクトの定位感と実在感が得られるかどうかを評価する。今回複数人での利用は評価しない。バイノーラル手法では Google の 3D オーディオ SDK である ResonanceAudio[13] を使用し、AR デバイスからヘッドホンを通じて立体音響を提示する。立体音響は AR デバイスの位置や向きに合わせて変化する。超指向性スピーカ (以降の 4 章, 5 章では単にスピーカと表記する) は高さ 2.3m の位置に設置した。また、音や仮想オブジェクトの種類によって定位感の実在感の違いがあるかどうかを検証するため、4 種類の仮想オブジェクト A~D (表 1) を用意して実験を行った。仮想オブジェクトは仮想オブジェクト移動エリア内 (図 2) を動き回るアニメーションとその場に留まって動いているアニメーションをそれぞれ用意した。アニメーションはいずれも音声付きで、アニメーションに連動して再生される。実験ではユーザが所持する AR デバイスとして iPhone7 を用いた。遅延を最小限に抑えるため、AR デバイスとスピーカ制御用 PC を同一 LAN に接続した。実験室内は再生される音以外の影響を最小限にするため空調を停止した状態で行った。静音時の騒音レベルは 46dB で、音声再生時の騒音レベルが最大 60dB となるように調節した。仮想オブジェクトとスピーカの位置は、実験開始前に予めシステム内で設定しておく。

実験の被験者は大阪工業大学情報科学部の学生 8 名 (男性 6 名, 女性 2 名) とし、被験者に対し 2 種類の実験を行った。実験 1 は被験者がスピーカから 4m 離れた場所に座った状態で、4 種類の仮想オブジェクトそれぞれ対して、表 2 の実験条件の音声をそれぞれ一対比較し、アンケートに回答する実験である。1 人の被験者に対して実験条件全対の比較、すなわち仮想オブジェクト 1 種類につき 6 回の比較を行う。実験 2 は被験者がユーザ移動エリア内 (図 2) を自由に移動し、キャラクター (歌声) の仮想オブジェクトを再生したときの評価をアンケートに回答する実験である。比較の回数に制限を設けないため、順序効果は考慮しない。アンケートの設問を表 3 に示す。これは仮想オブジェクトに対して定位感と実在感を評価する質問となっている。また、実験 2 では表 3(3),(4) に対する理由も質問した。事後アンケートでは実験を通して違和感があった部分や改善点なども質問した。

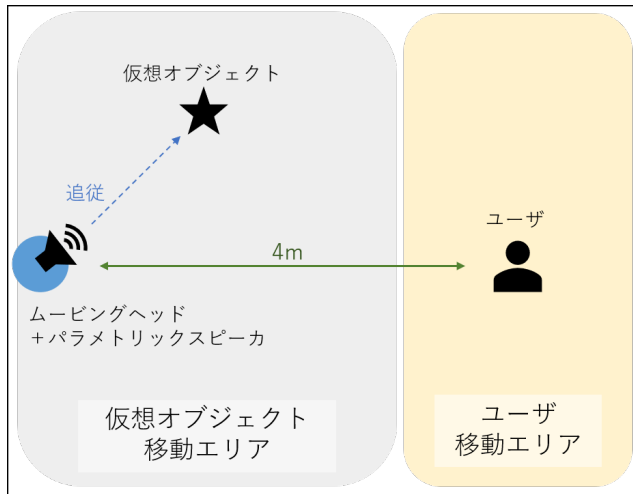


図2 実験のレイアウト

表1 仮想オブジェクトの種類

種類	3D モデル	提示音源
A	キャラクター	歌声
B	キャラクター	足音
C	ボール	バウンド音
D	ロボット掃除機	吸引音

表2 実験条件

パターン	音源提示方式	仮想オブジェクトの動作
ア	ヘッドホン	動き回る
イ	ヘッドホン	静止する
ウ	スピーカ	動き回る
エ	スピーカ	静止する

表3 アンケートの設問

	質問
(1)	音が聞こえた方向とスマートフォンに表示された物体の位置が一致していたように感じたのはどちらですか。(定位感)
(2)	あたかもその場に実在するかのよう感じたのはどちらですか。(実在感)
(3)	自身が移動しても音が聞こえた方向とスマートフォンに表示された物体の位置が一致していた。(定位感)
(4)	自身が移動しても表示された物体から実在感があった。(実在感)

- ・(1), (2) は実験1において、実験条件全対の24回分の回答を得た。
- ・(3), (4) は実験2でヘッドホンの場合とスピーカの場合との2回分の回答を得た。
- ・いずれも5段階評価
- ・(1)(2) は評価対象を A・B とすると、
1:A, 2:やや A, 3:どちらともいえない, 4:やや B, 5:B のような評価段階を設定した。
- ・(3)(4) の評価段階: 1:強く同意しない, 2:同意しない, 3:どちらともいえない, 4:同意する, 5:強く同意する

5. 実験結果と考察

各仮想オブジェクトの定位感と実在感に関するアンケート結果から一対比較を行い、順位付けをした結果を図3



図3 実験1-定位感の順位付け

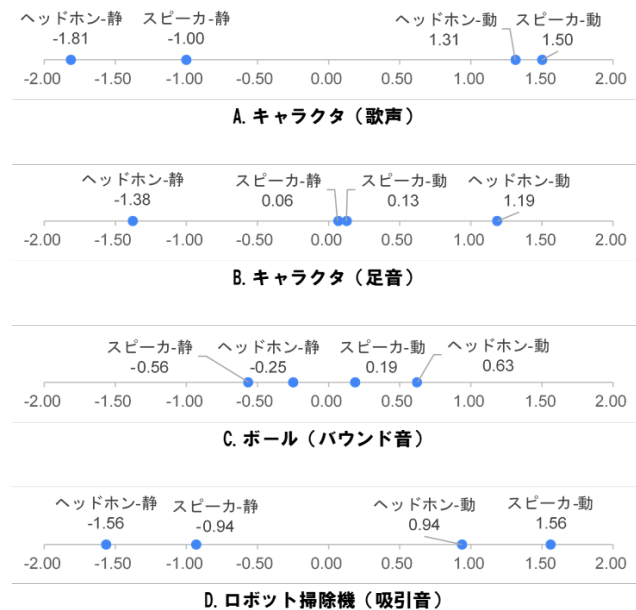


図4 実験1-実在感の順位付け

と図4に示した。順位付けは、5段階の回答結果から、(-4, -2, 0, 2, 4)で重み付けを行い、平均嗜好度を算出したものである。また、「スピーカ-動」と「ヘッドホン-動」の比較のアンケート結果を表4に、自身が移動した場合のアンケート結果を表5に示した。

5.1 定位感について

まず、仮想オブジェクトの動作の違いから考察する。図3より、すべての仮想オブジェクトにおいて、その場で静止しているものより、動き回っているもののほうが定位感があると評価された。これは音像が動くことで聞こえ方に変

表 4 「スピーカ-動」と「ヘッドホン-動」の比較のアンケート結果

	評価感覚	仮想オブジェクト	評価段階					中央値	最頻値
			1	2	3	4	5		
(1)	定位感	キャラクタ（歌声）	1	1	1	4	1	4	4
(2)	定位感	キャラクタ（足音）	0	1	1	5	1	4	4
(3)	定位感	ボール（バウンド音）	0	3	2	2	1	3	3,4
(4)	定位感	ロボット掃除機（吸引音）	1	4	3	0	0	2	2
(5)	実在感	キャラクタ（歌声）	0	3	1	2	2	3.5	4,5
(6)	実在感	キャラクタ（足音）	1	0	1	3	3	4	4,5
(7)	実在感	ボール（バウンド音）	2	1	1	1	3	3.5	5
(8)	実在感	ロボット掃除機（吸引音）	3	1	2	2	0	2.5	1

・表中の評価段階の数字は人数を表す。

・評価段階：1:「スピーカ-動」、2:やや「スピーカ-動」、3:どちらともいえない、4:やや「ヘッドホン-動」、5:「ヘッドホン-動」

表 5 自身が移動した場合のアンケート結果

	音声提示手法	評価感覚	評価段階					中央値	最頻値
			1	2	3	4	5		
(1)	スピーカ	定位感	0	0	0	5	3	4	4
(2)	スピーカ	実在感	0	0	0	4	4	4.5	4,5
(3)	ヘッドホン	定位感	1	0	0	4	3	4	4
(4)	ヘッドホン	実在感	0	1	1	3	3	4	4,5

・表中の評価段階の数字は人数を表す。

・評価段階：1:強く同意しない、2:同意しない、3:どちらともいえない、4:同意する、5:強く同意する

化が生まれ、その変化によって場所を捉えることができるためだと考えられる。

次にヘッドホンによる立体音響と超指向性スピーカの違いから考察する。図 3 によると、D. ロボット掃除機（吸引音）では、比較的、ヘッドホン-動よりスピーカ-動のほうが定位感があると評価されている。また、A. キャラクタ（歌声）では、スピーカ-動とヘッドホン-動で比較すると定位感の差はわずかで、同程度の定位感が得られていた。キャラクタ（歌声）とロボット掃除機（吸引音）は音が連続しているという特徴がある。連続的な音を出す仮想オブジェクトが動くとき、スピーカのほうが定位感が得られやすい傾向にあると考えられる。一方で、音が断続的な B. キャラクタ（足音）と C. ボール（バウンド音）について分析すると、動群ではヘッドホンよりスピーカのほうが定位感について劣っていることがわかる。その理由として、音を出す時間が瞬間的であること（足音 1 回につき約 200ms、バウンド音 1 回につき約 400ms）に加え、空間的にも一定の距離間隔を空けて音を出す。そのため、スピーカでは仮想オブジェクトの動きが予測しづらいことが定位感に影響を及ぼしているのではないかと考えられる。つまり、スピーカはヘッドホンに比べて瞬間的な音像提示では定位感が得られにくい可能性がある。

5.2 実在感について

図 4 によると、定位感と同じく、静止した仮想オブジェクトより動き回る仮想オブジェクトのほうが実在感があると評価された。特に、連続音である A. キャラクタ（歌声）と D. ロボット掃除機（吸引音）において、静群と動群の差が顕著であった。

ヘッドホンとスピーカの比較においても、音が断続的な

A. キャラクタ（歌声）と D. ロボット掃除機（吸引音）の 2 つはスピーカのほうが実在感があるとの評価を得た。事後アンケートでは、スピーカ再生時において、「（仮想オブジェクトが）自分の前を通り過ぎるときに実在感を感じた」「（ロボット掃除機が）近くにきた時に足を少し引いてしまった」との記述があり、これは超指向性スピーカの音が被験者のすぐ近くの床面で反射し、反射音がダイレクトに被験者に届くことでその場にいるという感覚が増大したと考えられる。実験では動き回るキャラクタ（歌声）とロボット掃除機（吸引音）には、被験者の約 50cm 前を通過するようなアニメーションを設定していた。一方で、ボール（バウンド音）は被験者の目の前を通過せず、実験室の端から端までをバウンドしながら移動するアニメーションであったため、スピーカで実在感が得られにくかったように考えられる。これらから、スピーカの音が反射する位置とユーザの位置とが近距離の場合に、スピーカにおいて実在感が得られやすい傾向にあることが分かる。

また、スピーカ再生時において、ヘッドホンをしていないことが、視聴時の違和感を感じさせない要因である可能性がある。これは、周囲の自然音（自身が出した音や実験室外から鳴る自然な騒音）も同時に聞こえることから、仮想空間上の音場と現実空間の音場が混ざり合い、より実在感を感じたのではないかと考えられる。

ただし、表 4(5)～(8) によると、評価段階の分布が広いことがうかがえる。そのため、実在感を感じる音像提示手法が人によって分かれる可能性がある。今後追加実験を行い詳しく分析する必要がある。

5.3 被験者が移動する場合の存在感と実在感

表 5 によると、スピーカとヘッドホンによる違いはほとんどみられなかった。よって、超指向性スピーカはヘッドホンと同程度の効果がある可能性を示せた。しかし、(3) で強く同意しないと回答した被験者は「歩いてみるとどこから聞こえるのか分かりにくかった」と回答しており、ヘッドホンによるバイノーラル手法では定位感に個人差がある可能性がある。最終アンケートでは「実際の位置とのズレを感じた」といった意見があった。超指向性スピーカの特性として、音声仮想オブジェクトの位置で反射した後、その反射音が別の場所でも乱反射していることがあり、実際の位置とは違う場所で音がしているように聞こえることがある。その影響を受けやすい位置に被験者が移動した場合、定位感が得られない可能性が考えられる。

5.4 応用例と今後の展望

本稿では評価していないが、仮想空間を複数のユーザと共有することで、ユーザは現実空間でも仮想空間でも同一の空間で存在することになる。さらに、本システムを使うことで、現実空間の音場と仮想空間の音場も共有されるた

め、仮想オブジェクトの存在感と実空間での周辺環境を同時に再現できる。この特徴は、仮想空間と現実空間が相互作用するコンテンツと相性が良い。例えば、現実空間上にステージと照明などの演出機構を用意し、ステージ上に仮想空間を設定することで、仮想オブジェクトがステージ上で演奏するARバーチャルライブができる。また、ARのため演出の場がステージ上にとどまらず、ステージを超えて仮想オブジェクトを観客の近くまで移動させることも可能である。通常のバーチャルライブでは、バーチャルキャラクターの存在感が課題だったが、音像を動き回る仮想オブジェクトに追従させることで存在感を演出でき、ヘッドホンも不要なためライブ会場特有の観客同士の一体感も阻害することはない。今回の実験結果を踏まえ、音楽は通常のサラウンドスピーカ、歌声は超指向性スピーカ、観客からバーチャルキャラクターへのアクション（仮想空間での拍手や声援など）はARデバイスのスピーカから再生するなど、状況に合わせて動的に音声の提示手法を変更し同時併用することで、より効果が高い演出が期待できると思われる。

6. おわりに

本研究は、移動するユーザや複数人ユーザで音像を共有することで、エリア内の任意の場所で手軽にARコンテンツを楽しめるシステムの開発を目指している。本稿では前段階の調査として、超指向性スピーカを用いた音像とヘッドホンを用いたバイノーラル手法の音像を比較して、AR体験中での仮想オブジェクトの定位感や実在感を評価した。また、仮想オブジェクトや音の種類による影響についても考察を行った。本稿の貢献は以下の通りである。

- (1) 仮想オブジェクトはその場で静止しているものより動き回るもののほうが、実在感と定位感が得られやすい可能性があることを示した。
- (2) 超指向性スピーカの音が反射する位置とユーザの位置とが近距離の場合に実在感が得られやすい可能性があることを示した。
- (3) 短時間の音を瞬間的に提示するとき、超指向性スピーカはヘッドホンを用いたバイノーラル音源より定位感が得られにくい可能性があることを示した。

参考文献

- [1] 情報通信白書平成30年版, 総務省, <https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h30/html/nd111350.html> (参照 2019/12/7)
- [2] PokémonGO, Niantic/Pokémon/Nintendo/GAME FREAK, <https://www.pokemongo.jp/> (参照 2019/12/7)
- [3] ホロリー, hololive, <https://www.hololive.tv/> (参照 2019/12/7)
- [4] ARビュー, Amazon, <https://www.amazon.co.jp/b?ie=UTF8&node=6350135051> (参照 2019/12/7)

- [5] Expeditions, Google For Education, https://edu.google.com/intl/ja/products/vr-ar/expeditions/?modal_active=none (参照 2019/12/7)
- [6] 金子翔麻, 渡邊恵太: SymPhonicAudio: 複数のスマートフォンを同期し現実空間への音像定位を動的に行う手法の提案, 情報処理学会インタラクション 2018, 1B39(2018).
- [7] 鎌倉友男, 酒井新一: パラメトリックスピーカの原理と応用, 信学技報, Vol.105, No.556, pp.25-30(2005).
- [8] 伊藤仁一, 中山雅人, 西浦敬信, 木村朝子, 柴田史久, 田村秀行: 複数の超音波スピーカを用いた音像位置の補間と視覚補助に基づく移動音像の実現, 電子情報通信学会技術研究報告, Vol.112, No.385, pp.93-98(2013).
- [9] 北岸佑樹, 田中友樹, 米澤朋子: 多人数講義における講師分身エージェントを用いたコミュニケーションの促進, 電子情報通信学会技術研究報告, Vol.118, No.404, pp.107-114 (2018).
- [10] AR Foundation, Unity, <https://docs.unity3d.com/Packages/com.unity.xr.arfoundation@3.0/manual/index.html> (参照 2019/12/9)
- [11] ARKit3, Apple, <https://developer.apple.com/jp/augmented-reality/arkit/> (参照 2019/12/9)
- [12] TouchDesigner, Derivative Inc., <https://derivative.ca/> (参照 2019/12/9)
- [13] ResonanceAudio, Google, <https://resonance-audio.github.io/resonance-audio/> (参照 2019/12/9)