

# 同期された AR 空間における 3 次元音声通信システムの開発

松村克広<sup>1</sup> 今村弘樹<sup>2</sup>

創価大学大学院工学研究科情報システム学専攻  
〒192-8577 東京都八王子市丹木町 1-236  
E-mail:e18m5220@soka-u.jp, imamura@soka.ac.jp

**概要**：近年，通信技術や映像技術の進化に伴い多くの 3 次元ビデオ通信システムが開発されている．しかし，これらのシステムは通信ツールとして様々な問題を抱えている．これらの問題を解決するために本研究室では，Kinect と HMD (Head Mounted Display) を用いて空間を共有することはできる 3 次元ビデオ通信システムを提案した．しかし現在のシステムは，他のユーザ情報を 3D で描画することができるが，音声を伝えることはできない．コミュニケーションシステムでは，明確な意思表示を表現できる音声が非常に大きな役割を果たす．そこで本研究では，よりリアルな 3 次元ビデオ通信システムを開発するために，通信相手の位置に応じて左右のスピーカーの出力を変えるステレオ機能のある音声通信システムを開発した．

**キーワード**：3 次元ビデオ通信システム，Kinect，HMD (ヘッドマウントディスプレイ)，サーバー，AR (拡張現実)

## 1. はじめに

本稿は，「同期された AR 空間における音声通信システムの開発」に関する研究成果である．この章では，この調査の背景と目的，および以下の章の構成について説明する．

### 1.1 研究背景

近年，PC，スマートフォン，タブレットの普及などネットワーク技術の発展により，通信ネットワークを利用したシステムが私たちの生活に必要な不可欠なものとなっている．例えば，電子メールや SNS，ビデオ通話などがある．これらのシステムでは，文字などの 2 次元情報により画像，音声，およびビデオを含む情報を伝えることが可能である．また近年では，ビデオ技術の発展に伴い，3 次元情報を活用し，3D で他のユーザの体やオブジェクトを映すことで，より円滑かつ現実的なコミュニケーションを可能とした 3 次元ビデオ通信システムが開発されている．

既存の 3 次元ビデオ通信システムには様々なシステムが存在する．例えば，10 台の Kinect を使用したテレ・ヒューマン [1] といったホログラムシステムや AR マーカーを使用した 3D 再構成システム[2]，ユーザの動きをアバターに反映するアバターキネクト[3] などがある．しかし従来の手法では，人の位置を限定される，AR マーカーを使用するなどの制限，またアバターを介することで現実感が失われてしまうなどの欠点がある．

これらの問題を解決するため，本研究室では Kinect と HMD (Head Mounted Display) を使用した 3 次元ビデオ通信システムを提案し開発した[4]．このシステムでは，遠隔地の複数のユーザが AR 空間で相互に再構成される．しかし，現在のシステムは図 1 に示すように 3D で他のユーザ情報を画面に映し出すことはできるが，音声を伝えることができない．通信システムにおいて，明確な意思表示を



図 1 システムのイメージ

伝えることのできる音声を非常に大きな役割を持っている．そこで本研究において，より実用的な 3 次元ビデオ通信システムを実現するために，現在の通信システムの音声通信機能の拡張を行う．

### 1.2 本研究の目的

本研究では，現在の 3 次元ビデオ通信システムをより現実的なコミュニケーションツールとして発展させるため，音声通信システムを構築していく．

音声通信機能の拡張によって現在のシステムは以下のよう  
なことが実現される．

まず，より実用的なコミュニケーションツールとしての使用が可能となる．また，現在の 3 次元ビデオ通信システムでは，通信相手の位置を限定されることがないため，音源の方向が一定ではない．よって，より現実感を提供するために通信相手の位置によって左右のスピーカーの出力を変化させるステレオ方式を実現する．

## 2. 提案手法

この章では，より現実的な 3 次元ビデオ通信システムを実現するための音声通信システムの手法について説明する．また，このシステムで使用する Kinect と HMD の概要についても説明する．さらに，このシステムの概要について簡単に説明する．

## 2.1 Kinectの概要

この研究で使用する Kinect の概要について説明する。図 2 に、Kinect の構造と外観を示す。Kinect には IR デプスセンサー、RGB センサーであるカラーセンサー、4本の単指向性マイクからなるマイクアレイがある。これらのカメラの垂直視野角は 43 度、水平視野角は 57 度、解像度は 640 × 480 画素である。



図 2: Kinect の構造と外観

## 2.2 HMDの概要

Oculus Rift CV1 と Ovrvision Pro の組み合わせである HMD の概要について説明する。図 3 に HMD の構造を示す。Oculus Rift CV1 は、バーチャルリアリティ (VR) の HMD である。さらに、ヘッドトラッキングにより 3 次元空間上の HMD の向きを得ることができる。HMD を VR から AR に変換するには、Oculus Rift CV1 と Ovrvision Pro を組み合わせる。Ovrvision Pro は、2つの RGB 画像を取得できるステレオカメラである。Oculus Rift CV1 は、目あたり 1080 × 1200 の解像度で表示される。Ovrvision Pro は最大 60 フレーム/秒でビデオをキャプチャする。

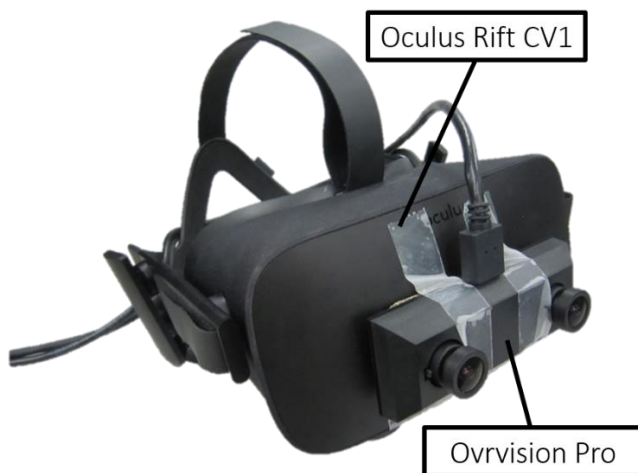


図 3: HMD の構造

## 2.3 同期された 3 次元ビデオ通信システムの概要

図 4 にこのシステムの構成要素を示す。また図 5 と図 6 に、このシステムの処理の流れを示す。このシステムでは、サーバー PC と複数のクライアント PC が使用される。Kinect と HMD は各クライアント PC に接続されている。まず、各クライアント PC は、Kinect のセンサーでユーザの身体情報を取得し、得られたデータを UDP (User Datagram Protocol)

を用いてサーバー PC に送信する。サーバー PC はサーバー共有空間を作成し、受信したデータをサーバー空間に統合する。また、サーバー PC は UDP を用いて各クライアント PC に統合した情報を送信する。各クライアント PC は、受信した情報に基づいてサーバー空間の視野を投影する。

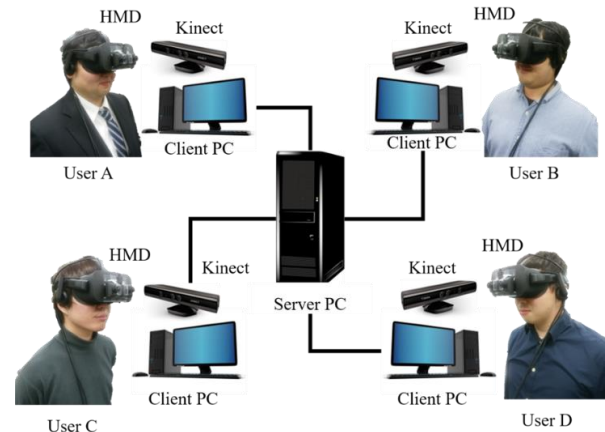


図 4: 提案システムの構成要素

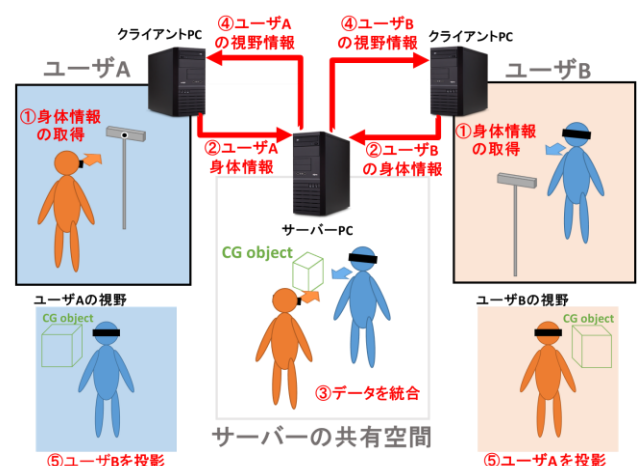


図 5: 提案システムの流れ

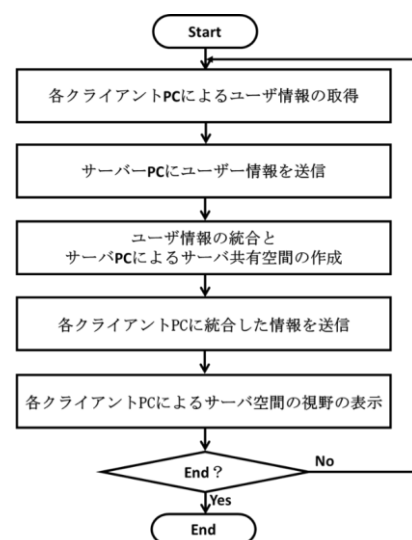


図 6: 提案システムのフローチャート

## 2.4 音声通信システムの概要

本研究の構成を簡潔に説明すると、マイクなどのデバイスから取得した音声データを、インターネット網を使用して逐次的に送受信を行い、スピーカーからの再生をリアルタイムで実行できる通信システムである。また、より現実感を伝えるため、通信相手の位置によって左右のスピーカーの出力を変化させる。

音声を入出力するためのデバイスとして、Kinect のマイクと HMD のスピーカーを使用する。

図 7 に、本研究の処理の流れを示す。このプロセスは、「送信者の処理」(図 7 (a))と「受信者の処理」(図 7 (b))に分かれている。

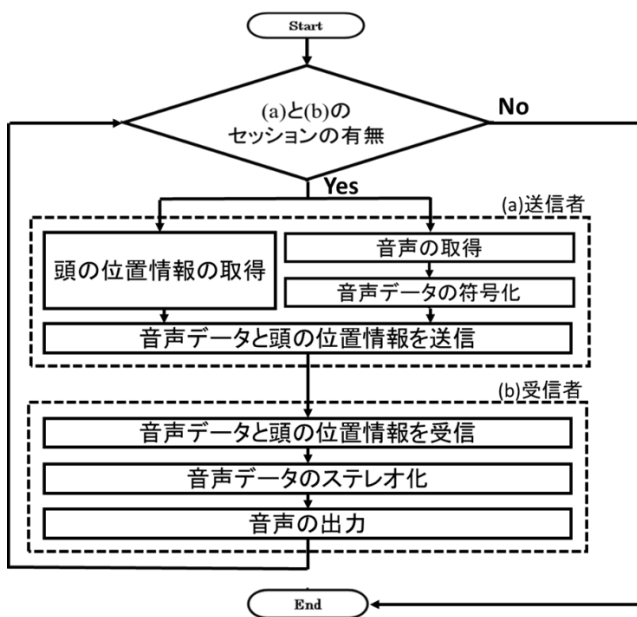


図 7: 音声通信システムのフローチャート

### 2.4.1 送信者の処理

送信者は、まず Kinect から音声と頭部の位置情報を取得する。

音声を取得する際、アナログ信号である生の音声をデジタル信号へ変換する必要がある。この処理をするのに PCM(Pulse Code Modulation)を用いた。PCM は、生の音声つまりはアナログ信号の波形の値を一定時間ごとに取得し(標本化)、この標本化した数値を 0 と 1 で表される 2 進数に変換(量子化)することで、アナログ信号の波形をビット配列で表されたデジタル信号に変換することができる。また、1 秒間に何回数値化するかを表す「サンプリング周波数」と、音声データを何ビットで表現するかを表す「量子化ビット数」によって音質が決定される。このサンプリング周波数が大きくなるほど音の解像度が上がるが、その分データ量も増加してしまう。音声通信の場合、リアルタイム性が重視されるので、本研究ではデータ量の軽量化を図るため、サンプリング周波数を 11025Hz、量子化ビット数を 8bit とした。

次に、符号化された音声データを送信する。音声のようなリアルタイム性の高い通信では、途中でパケットが紛失するなどして相手に伝達されなくても、あとで再送することは意味がない。したがって伝送方法として、IP パケットの再送機能などを備えた TCP ではなく、処理の軽い UDP を使用する。

また、現在の 3 次元ビデオ通信システムでも画像データを送る手法として UDP が使用されているので、音声データと頭部の位置情報も同様に UDP 通信によって伝送を行う。

### 2.4.2 受信者の処理

受信者は、送信者から UPD によって音声データと頭部の位置情報を受信する。

次に、送られてきた音声データのステレオ化を行う。

現在の 3 次元ビデオ通信システムでは、図 8 に示すように、自分の頭の向きや通信相手が動くことによって音源位置が変化してしまうため音源の位置が一定ではない。図 9 に示すように、L チャンネルの音が R チャンネルよりも大きい場合、聴取者は音源が左方向にあるように知覚する。逆に、R チャンネルの音が L チャンネルよりも大きい場合、聴取者は音源が右方向にあるように知覚する。これを利用して、通信相手の位置により左右のスピーカーの出力を変化させる。

図 10 の示すように、Kinect のディスプレイの一番左上を原点(0, 0)とし、一番右下の座標を(Xmax, Ymax)とする。また、Kinect で取得した送信者の頭部座標を Head(X, Y)とし、送られてきたモノラルの音声データ (INPUT DATA) に対して(1)と(2)のような操作を行う。

$$L: \text{INPUT DATA} \times \left( \frac{X_{\text{max}} - \text{Head}(X)}{X_{\text{max}}} \right) \dots (1)$$

$$R: \text{INPUT DATA} \times \left( \frac{\text{Head}(X)}{X_{\text{max}}} \right) \dots (2)$$

L と R はそれぞれ左右のスピーカーに対する出力であり、頭部座標の変化によって左右の出力の比率が変化する。

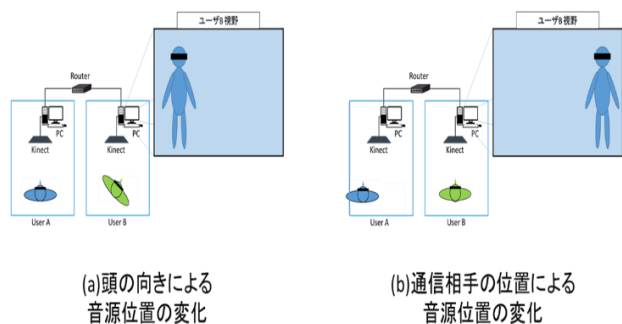


図 8: 音源位置の変化

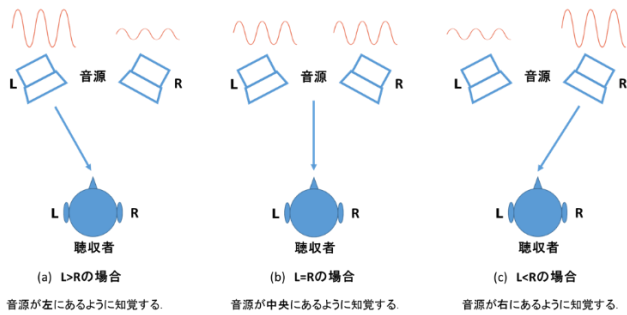


図 9: 左右の出力変化による位置表現

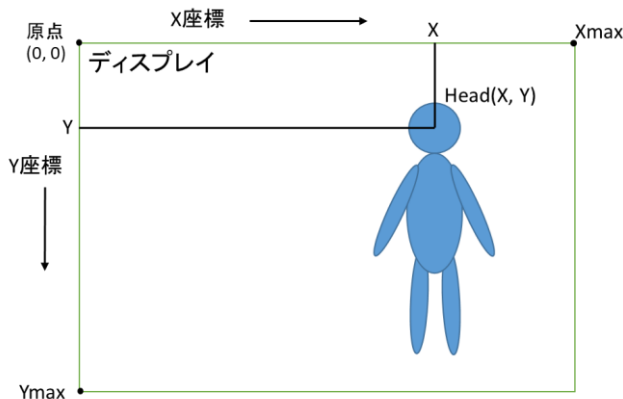


図 10: Kinect で取得したディスプレイ画面

### 3. 評価実験

この章では、本研究を評価するために提案された音声通信システムの評価実験の結果を報告する。

#### 3.1 実験の概要

本実験では実際に、本研究で提案された音声通信システムを使用して以下の評価実験を行った。2名のユーザーに提案された音声通信システムを使用し実際に会話をしてもらう。また、相手の位置による音源方向変化を評価するため、通信相手に左右に移動して会話してもらう。

10人が実験を行った。その後、以下の項目を評価した。

- ① 通信による会話の遅れは気になるか。
- ② 声のみで相手の位置を把握できるか。
- ③ 言葉を違和感なく聞き取れたか。

それぞれについての評価を、最低を1、最高を5とする5段階評価で判定した。表1に実験結果を示す。

表 1: 実験結果

NO.	①	②	③
平均値	4.10	4.80	4.60
標準偏差	0.54	0.40	0.49

#### 3.2 実験結果の考察

項目①~③の平均スコアはすべて 4.0 を上回っている。

さらに、標準偏差はすべて 1.0 より小さい。すべての項目の中で項目②は最も高い平均スコアであり、最も低い標準偏差である。項目①は平均スコアが最も低く、標準偏差が最も高い項目となった。

項目①と項目③は、会話に関する項目である。これら2つの項目のうち、項目③に比べ項目①の平均値は少し低く、標準偏差の値も少し高くなった。通信相手の言葉は違和感なく聞き取れるが、会話となると若干タイムラグが気になるようである。この結果から、タイムラグを更に減らすために伝送方式や音声の圧縮などを改善していく必要があると考える。項目②は通信相手の位置による左右の出力に関する項目である。この項目はほかの項目に比べ、非常に高い結果となった。しかし、Kinectの視野範囲からユーザが外れてしまうと、外れる前の最後の出力表現が一定で出力されるため、その場合の出力表現を新たに考える必要がある。

## 4. おわりに

本稿では、同期されたAR空間における音声通信システムの概要を述べ、実際に実験を行い評価した。評価実験では3つの項目を用意し、作成した音声通信システムを使用してコミュニケーションが可能か、また左右のスピーカ出力表現により通信相手の位置を把握できるか検証した。その結果、全体的に平均値は高く、非常に実用的なコミュニケーションツールであることがわかった。しかし、実験結果から、通信のリアルタイム性に関してや、Kinectの視野範囲からユーザが外れてしまう場合の問題も改善していくがある。今後は問題を改善し、精度を高め、より現実的なコミュニケーションツールを目指す。

## 謝辞

本研究を進めるに当たり、指導教官の今村弘樹准教授からは多大な助言を賜りました。厚く感謝を申し上げます。また進んで被験者を務めてくれた創価大学今村研究室の皆さまと卒業された藤林勇希さんにも感謝の意を表します。最後の謝辞として、私は私の恩師である大学創設者の池田大作と池田金子に最大限の感謝を捧げます。

## 参考文献

- [1] “ヒューマン・メディア研究所”. <http://www.hml.queensu.ca/telehuman/>, (参照 2018-01-15).
- [2] Ryo Jozaki . AR Chat: remote communication support system that uses augmented reality. In The national convention of IPSJ, 2012.
- [3] AvatarKinect. <http://marketplace.xbox.com/enUS/Product/Avatar-Kinect/66acd000-77fe-1000-9115d8025848081a>.
- [4] Yuki Fujibayashi and Hiroki Imamura, “同期されたAR空間における3次元映像通信システムの開発”