

拡張音響現実感を用いた仮想音声会議クライアントの検討

代表研究者 近藤 和弘 山形大学 大学院理工学研究科 准教授

1. はじめに

遠隔コミュニケーションにおける音空間の3次元化がもたらす高臨場感知覚への効果とそれを実現するためのシステム全体の仕様を検討し、音声を中心とした次世代マルチメディア・コミュニケーション環境の構築を検討している。

高い臨場感を備えた遠隔コミュニケーション・システムでは、参加する人間同士があたかもその場を共有している感覚が必要になるが、これがコミュニケーションを活性化、円滑にするか明らかになっていない。例えば音声のみを用いたコミュニケーションでも各発話者の音像を水平面で 45° 以上話せば -6 dBのSN比でも音声了解度が70%以下に低下しないことが分かっている。しかし遠隔コミュニケーション・システムとして本格活用するためには、さらに全体の遅延の影響、音声と映像の同期ならびに3次元空間内の相対配置、音像定位に用いる頭部伝達関数の精度の検討等、単に空間を提示するだけでなく、他の影響についても総合的に検討する必要がある。この中で、本検討では特に頭部回転時、他者音声位置を変更し相対的に不動とし、自然な音声提示を実現し、その効果を検証する。

2. 拡張音響現実感を用いた仮想音声会議システムの概要

2-1 提案システムの目標

本提案では使用者の周囲の音も自然に聞こえるように供給しつつも、会議参加者の音声を音声了解度が保たれる位置から聞こえるように音像定位し供給するシステムを構築する。会議参加者音声は使用者の頭部回転に追従して音源位置を実時間で変更し、使用者から見て音源が相対的に動かないようにする必要となる。これは周囲の音と、仮想的に配置した音源を混合する拡張音響現実感では特に重要となる。なぜなら頭部運動に応じて音源位置を変更しないと、周囲の実音源は使用者から見て相対的に不変な位置を保つのに対し、仮想音源が頭部運動に追従して移動し、極めて不自然な状況が発生するためである。

空間内の特定位置から放射される音源の位置知覚は、音源と受聴者の間の空間伝達関数により音源特性が変形を受け、この変形音の両耳入力により得られる[1, 2]。この伝達関数には空間伝搬特性のほか、受聴者の頭部、外耳、肩などの反射や回折特性も含まれる。音源から両耳への経路差により乗じる両耳入力間の時間差(位相差)、レベル差、相関、周波数特性の差などから発生音源位置を知覚する。

あらかじめこの方向からの伝達関数を任意音信号に畳み込んで両耳に入力することにより、受聴者はこの音が同方向から到来するように知覚する。そこで受聴者の多数の方位角、仰角からの伝達関数を測定しておき、このデータベースの内特定の伝達関数を選択して畳み込むことにより音源位置を自由に選ぶことができる。また連続した位置の伝達関数を切り替えて畳み込むことにより音源位置移動も模擬することができる。

一方、受聴者が頭部を回転させると、実空間内の音源に対しては両耳への伝達関数が連続的に変化し、両耳への音の変化から実空間内の固定した絶対位置からこの音源が知覚される。これを模擬するためには、頭部運動に応じて畳み込む頭部伝達関数を切り替えることにより音源の絶対位置が不変であるようにすることが必要となる。この様子を図1に示す。

この図ではまず音源が正面にあるように模擬されているものとする。その後、(b)に示すように頭部が 45° に向いたことを方位センサーで感知し、頭部伝達関数を右 45° のものに切り替えて音源に畳み込むことにより、音源は受聴者から見て右 45° から到来するように知覚され、空間内の絶対位置としては不変として知覚される。

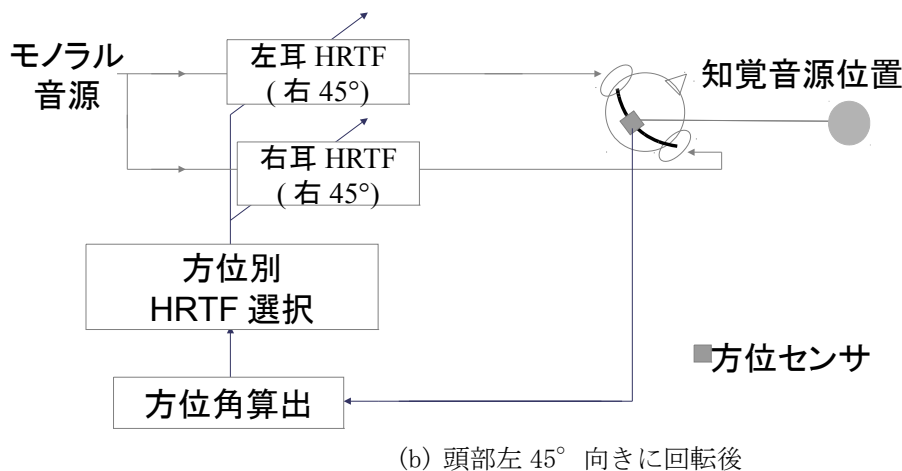
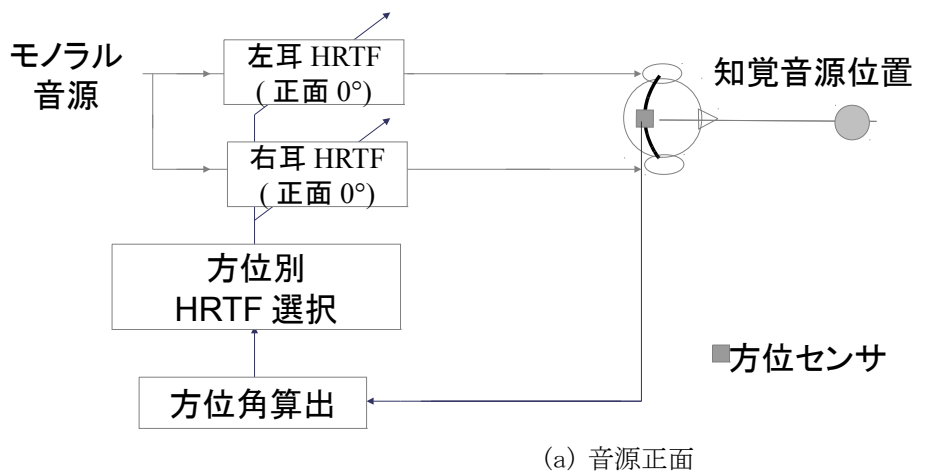


図 1. 頭部運動に応じた頭部伝達関数の切り替え

このシステムでは、最終的に以下を目標とする。

- (1) 長時間に渡り被験者に実際に遠隔地点との会話に利用した場合、80%以上の被験者が音声の提供方法が不自然ではないと評価されること。特に被験者ではシステムを使用して、実際の周囲音環境にも接しながらも仮想音響空間内の音源を聴取する拡張音響現実感会話システムの使用感を明らかにする。
- (2) 会議音声の了解度も同時に評価し、音声了解度 90%以上が達成されること。

2-2 拡張音響現実感を応用した音声会議システムの構成

図 2 に示すに拡張音響現実感を用いた音声会議システムを目指す。このシステムの入出力部は主に以下の要素よりなる。

- (1) 音響入出力ユニット：イヤホン型の音響入出力一体型装置に音声入力用ブームマイクを組み合わせたユニットである。イヤホン型入出力は左右耳入口で周囲音を入力するマイクと、この周囲音に音像定位した他地点の音声を混合した出力を再生するスピーカを組み合わせたものである。
- (2) 頭部運動検出装置（ヘッドトラッカー）：頭部の方位、および位置を測定するユニットであり、測定値に応じて他地点からの音声の音像位置を変更する。
- (3) 音声送受信ユニット：周囲音や頭部方位などを受信し、他地点より入手した音声信号を他地点の方位をもとに音像定位し、周囲音に混合して、左右イヤホンに出力する。また、入力した音声信号と頭部位置をネットワークを介して他地点にも送信する。

以下このシステムの動作を仮想的な地点 1 にいるユーザー A を中心に説明する。地点 1 には別のユーザー B も存在するが、このユーザーの音声は周囲音とともにイヤホン・マイクより収録し、イヤホンスピーカより再生する。これによりユーザー B の実際に見える位置と音像位置が一致するため、自然である。一方仮想的な地点 2 にユーザー C が、また地点 3 にユーザー D が存在するとする。これらの仮想的な地点の位置は実際の位置と一致させる必要はなく、各地点のユーザーの音声は先行研究成果に従いお互いに方位 45° 以上離れるように配置さ

れる。この配置に従い、また各話者の方位に従い、使用者 A が向いている方向を加味して各使用者の音声を定位して周囲音と多重化して使用者 A に提供する。これにより同地点に存在する使用者 B とともに、遠隔地にいる使用者 C, D も同じ地点にいるように会話できる。

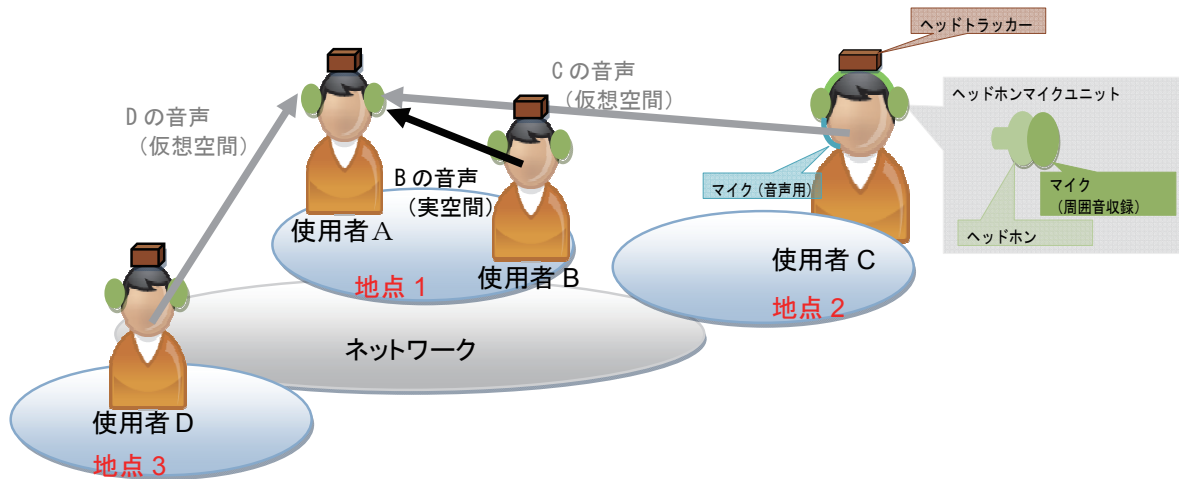


図 2. 拡張音響現実感を用いた仮想音声会議システムの構成

3. 拡張音響現実感を用いた仮想音声会議システム・クライアントの試作

図 3 に試作したシステムの構成を示す。この図は 2 地点間の音声通信を模擬し、その片側のシステム概要を表している。近端のシステム利用者をユーザ A、遠隔地の相手のシステム利用者をユーザ B とする。

まず、ユーザ B の発話をマイクでモノラル録音し、ユーザ A のコンピュータに送信する。ユーザ A のコンピュータで、ユーザ B の音声をバイノーラルシミュレーションして立体音声にする。ユーザ B の音声は被験者の正面（コンピュータ画面の位置）に来るように定位する。

このときユーザ A は作業の関係で頭部を動かすことがあるが、ユーザ B の音声は被験者の正面にではなく、コンピュータ画面の位置に固定したい。そのため、ユーザ A の頭部運動を追跡する必要がある。この目的のためヘッドトラッキングを用いる。ヘッドトラッキングで得た情報を基に、頭部伝達関数を逐次更新することで、ユーザ A が頭部運動しても、音像位置を固定する。

ヘッドトラッキングに用いる 3 次元センサーには様々なものがあるが、本研究ではバーチャルリアリティシステムに広く用いられている Polhemus 社の位置センサー-FASTRAK-TX4 を使用した[3]。このセンサーは、磁気発信装置(トランスミッタ)、磁気受信装置(レシーバ)、演算ユニットとなっており、トランスミッタより発生した微弱磁界をレシーバで受信して 3 次元空間内の位置を検出する。表 1 に FASTRAK-TX4 の仕様を示す。

位置情報は固定されたトランスミッタを原点としたレシーバの位置が x , y , z の 3 自由度で計測される。方向情報は、オイラー角と呼ばれる方位角、仰角、回転角の 3 自由度の計 6 自由度が計測される。磁気を用いたこのセンサーの長所としては、位置、姿勢/方位角同時検出が可能であること、センサーが人体などによって遮られた状態でも位置と姿勢が検出できること、構造が単純であることが挙げられる。逆に短所としては、信号強度が距離の 3 乗に反比例して低下してしまうこと、計測領域は半球内であること、周囲に導電性金属がある場合、渦電流による磁界歪み発生し、位置、姿勢情報に誤差が生じることがあげられる。

ヘッドトラッキングによって得た頭部の方位角に応じて頭部伝達関数を畳み込まれた音声はイヤホンを通じて両耳へ音声提示される。この時、イヤホンによって両耳がふさがれるため、被験者の周囲音がイヤホンによって遮断される。そこで、イヤホンの外装にマイクを備えたバイノーラル・イヤホン・マイク (Roland CS-10EM) を用いて周囲音を収集し、バイノーラルシミュレーションされた音声に加算して被験者へ提示する。これらにより、仮想音響空間と実音響空間を共有する拡張音響現実感を実現し、自然な音声提示の実現をする。

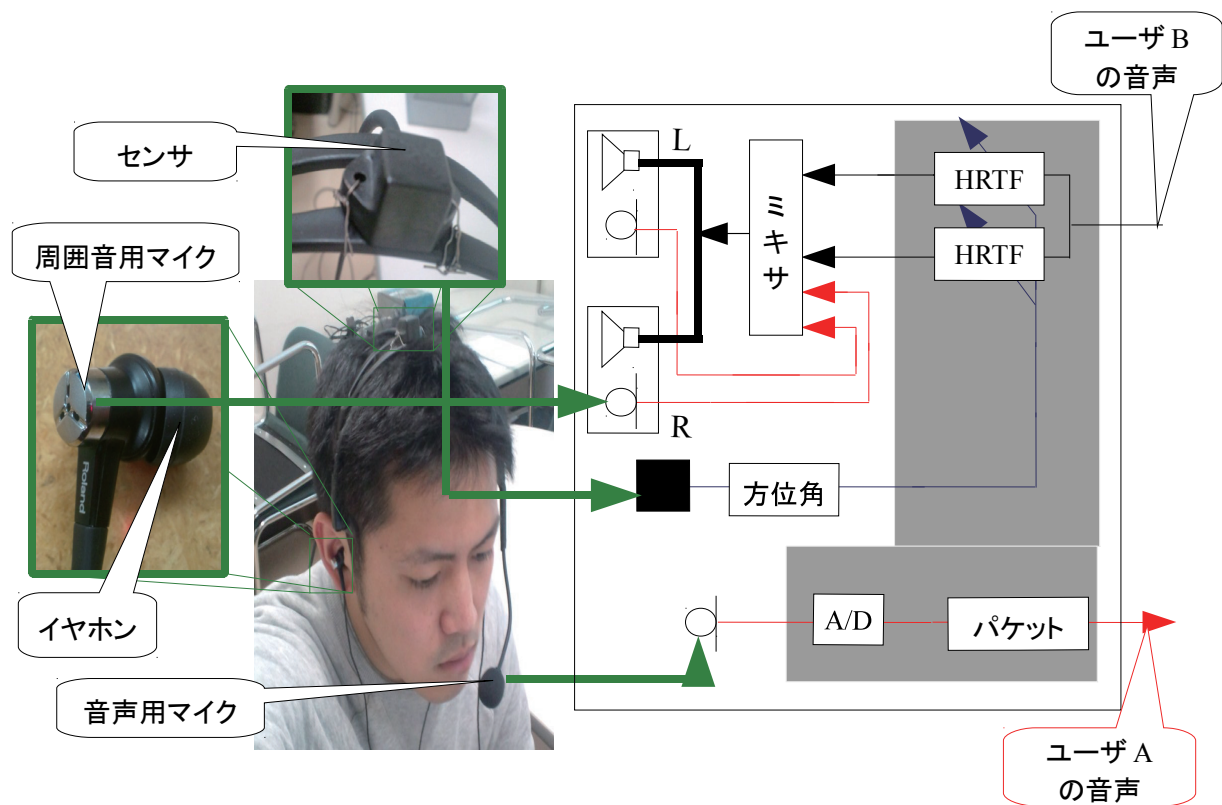


図 3. 試作システムの構成

表1.FASTRAK-TX4の仕様

項目	仕様
位置分解能(76cm以内)	8 mm
角度分解能(76cm以内)	0.15°
標準操作範囲	76 cm
操作可能範囲	305 cm
更新レート (レシーバ1つの時)	120 Hz
センサー更新時間 (レシーバ1つの時)	8.3 ms
最大レシーバ数	4

4. 試作システムの自然性評価

試作システムを実際に使用して、他地点にいる人物と会話をして、ヘッドトラッキングの有無で自然性や会話のしやすさなどに、どのような違いが出るか、どちらが使い易いかなどを比較した。実験は、実際の音声会議の使用状況を想定して、試作システムを利用した会話実験を2種類行った。ここでは、ユーザ A を主要被験者とする。

まず第1の実験ではユーザ A である被験者と、相手側のシステム利用者のユーザ B との2名による対向会話実験を行った。実験1で被験者が聴取する音源はユーザ B の音声の仮想音1つである。

一方、第2の実験として、被験者、第1実験と同様に別室の話者 B と、被験者と同じ部屋の話者 C の3名による会話実験を行った。実験2で被験者が聴取する音源は、ユーザ B の音声の仮想音と話者 C の音声の実音の2つを混合したものであり、仮想音響現実感を実装した形態となっている。

本研究では、ヘッドトラッキング導入の効果を検証したい。よって、被験者には実験中、積極的に頭部運

動を行う必要がある。そこで、会話実験で使用する会話のタスクは、頭部運動を積極的に行うものを使用した。実験1の会話タスクは旅行会社員と顧客のやり取り、実験2では軍艦ゲームを設定した。このとき、ヘッドトラッキングの有る場合と無い場合のユーザビリティを比較する。

この実験では前節で述べた試作音声会議システムではなく、NASAで開発したSlab3dを転用することにした[4]。Slab3dは3次元音響の研究用にNASA Ames Research CenterのSpatial Auditory Displays Labで開発された環境であり、実時間で3次元音源レンダリングを行う機能があり、Polhemus FASTRAKを用いて頭部運動に適応することができ、また音声信号をLANを用いてVoice over IP (VOIP)を用いて伝送することもできる。頭部伝達関数が開発者のもので固定されていること、細かいパラメータの設定が困難であるなどの欠点はあるが、とりあえずこの実験に必要な機能が安定して実装されていることを考慮し、本実験ではこのSlab3dを応用することとした。

実験時の提示音圧レベルは60 dB、サンプリング周波数は44.1 kHz、量子化ビット数は16 bitとした。ユーザAとユーザBの接続はハブなどを介さずに、LANケーブルとクロスケーブルでPCを直結し、お互いのIPアドレスをあらかじめ固定しておく。その他の機器はアップリンクを含めて接続せず使用し、トラフィックの影響を受けないようにした。音声信号の部屋間の伝送にはSlab3dのSlabCallを用いてVOIPで転送している。

4-1. 話者2名による対向会話実験

図4に実験1の実験系と音源配置を示す。試験開始時の音源の配置位置は、受聴者から距離1.0mの正面に配置する。1対1の会話なので、正面に配置するのが妥当と考え、この配置とした。所要時間はインストラクションに10分、会話実験が20分(10分×2回)、評価に10分(5分×2回)の40分程度である。インストラクションで試験の内容の説明、バイノーラルシミュレーションを使用した音の受聴、ヘッドトラッキングの効果の体験、他地点の相手の会話音声の音量調節手順、イヤホン・マイクの周囲音の音量の調節手順を指示する。

話者2名による対向会話試験に用いる会話のタスクは、温泉旅行を企画したい顧客と、旅行会社の会社員という設定で会話を行う。会話試験では被験者が知りたい情報を相手を持っている、逆に相手が知りたい情報を被験者が持っているように設定している。よって、お互いに情報交換の会話を行い、記入用紙に記録しながら会話試験を行う。試験後に主観評価シートに評点を記入させ、ヘッドトラッキングの有無を切り替えて、他は同じ条件で再度試験を行う。

また、本研究では、ヘッドトラッキングの導入の効果を調べるため、被験者が会話試験中に自然な形で頭部運動を行う必要がある。そのため、複数の紙をホチキスで綴じるなど、情報を一か所にひとまとめにしておく、視線の変化が少ないため、頭部運動を積極的に行わない可能性がある。そこで会話のタスクの中で用いる情報をいくつかの紙に書き、クリアファイルに分けて入れた。会話のタスクに用いる情報が記載された用紙を被験者の $\pm 45^\circ$ 、 $\pm 90^\circ$ に配置し、部屋内の複数の掲示板に分散した状態を模擬した。これで、水平方向に情報を分散させる。これにより相手の要望を聞いてうえて、関連する情報を探し、情報を見ながら相手と会話をする機会が増える。このように、視線の変更による頭部運動が積極的に行われるように工夫した。情報を探す動作により、主に左右方向の頭部運動がおこなわれる。また、記入用紙の記録により、主に上下方向の頭部運動がおこなわれる。

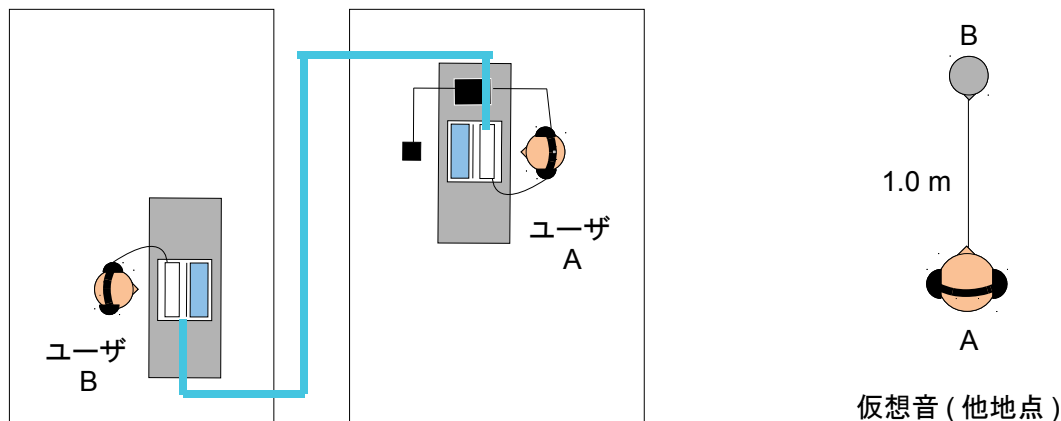


図4. 実験1の実験系と音源配置

4-2 評価方法

評価にはオピニオン評価 MOS(Mean Opinion Score)を用いた。主観評価の評点は、項目ごとに1~5点の評点をつける形式で行った。複数の評価者の5段階評価の値の平均を取ったものがMOSの評点値となる。MOSの評点は高いほど評価がよいことになる。

本研究ではユーザ体感品質をMOSを利用して評価した。本実験に用いるMOSの評価項目は「相手の会話音声の音質」、「会話のしやすさ」、「話しやすさ」、「聞きやすさ」、「使いやすさ」、「周囲音の聞こえ方」、「違和感を感じる時があったか」、「音声の遅延」、「音声の途切れ」の9項目である。各項目につきそれぞれ5段階評価を行い、全被験者の平均値を各項目の評価値とした。被験者は聴力健常な20代の男性5名、女性1名である。

4-3 音圧校正法

本研究の実験でバイノーラルイヤホンから再生される音声は、LANケーブルを介し、システムでバイノーラルシミュレーションされた相手の話者音声と耳介入口に装着されたイヤホン・マイクから録音された周囲音の2つである。実験を行う中で、この2つの音声の音圧が統一されていないと、評価値に影響すると考えられる。そこで実験開始前に、被験者には用意したイヤホン校正音を聴きながら、イヤホン・マイクとイヤホンの音圧レベルを校正する。各被験者に主観でスピーカとイヤホンの音圧レベル、スピーカとイヤホン・マイクの音圧レベルを合わせ、主観的に等価となるように音量調整をしてから、実験を行った。

スピーカと被験者の位置関係を図5に示す。被験者とスピーカ間の距離は1500mm、スピーカの高さ1140mmである。実験を行った環境の暗騒音は47dBである。この環境下で、椅子に座った被験者から受聴するスピーカの校正音を60dBとする。校正音の音圧を騒音計で測定する。人が普通に会話する際の音圧は60dBであるので、校正音を騒音計で測りながら、60dBに設定した。イヤホン校正の際は被験者の基準軸上正面にスピーカを配置し、校正を行う。

校正音は、各機器と被験者の周波数帯域を考慮し、適度な周波数帯域幅を持つ刺激音を用いる必要がある。そこで、オクターブ帯域幅でエネルギーの等しいピンクノイズを、オクターブバンドパスフィルタに通し、作成した1/1オクターブバンドノイズを使用する。校正音には下限遮断周波数を2kHz、上限遮断周波数を4kHzに設定したフィルタを適用し、中心周波数が3kHzとなるオクターブバンドノイズを使用した。これは人間の聴覚が2~4kHzで最も感度がよいためである。この構成音を被験者正面から60dBで再生し、この音を聴きながら、イヤホン・マイクとイヤホンの音量を調整した。

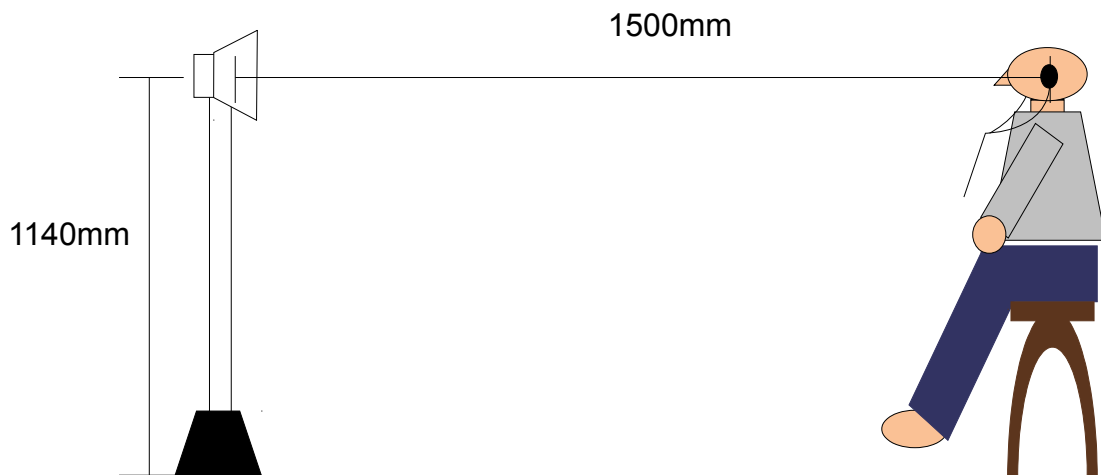


図5. 音圧校正の実験系

イヤホンとイヤホン・マイク校正の流れは図6に示す。イヤホン・マイクの校正は以下の手順による。

- (1) 被験者は、実験室に設置された椅子に座り、音響ユニットを装着する。
- (2) イヤホン・マイクのみキサの音圧レベルを最小にした状態で、被験者正面のスピーカから校正音を再生する。
- (3) 被験者はみキサのボリュームを徐々に上昇させ、イヤホン・マイクの音圧レベルを調整する。スピーカから再生された音とイヤホン・マイクで録音され、イヤホンから再生される音の音圧レベルが等しいと感じる場所でレベルを固定する。
- (4) (3) のレベルを記録する。
- (5) 校正音の再生を一度止め、みキサの音圧レベルメータを最大にした状態で校正音を再度再生する。
- (6) 被験者は音圧レベルを徐々に下降させ、(3)、(4)と同じ手順を行う。
- (7) 測定した2つのレベルの値の平均を取り、その音圧レベル値で再生音圧を固定して実験を行う。

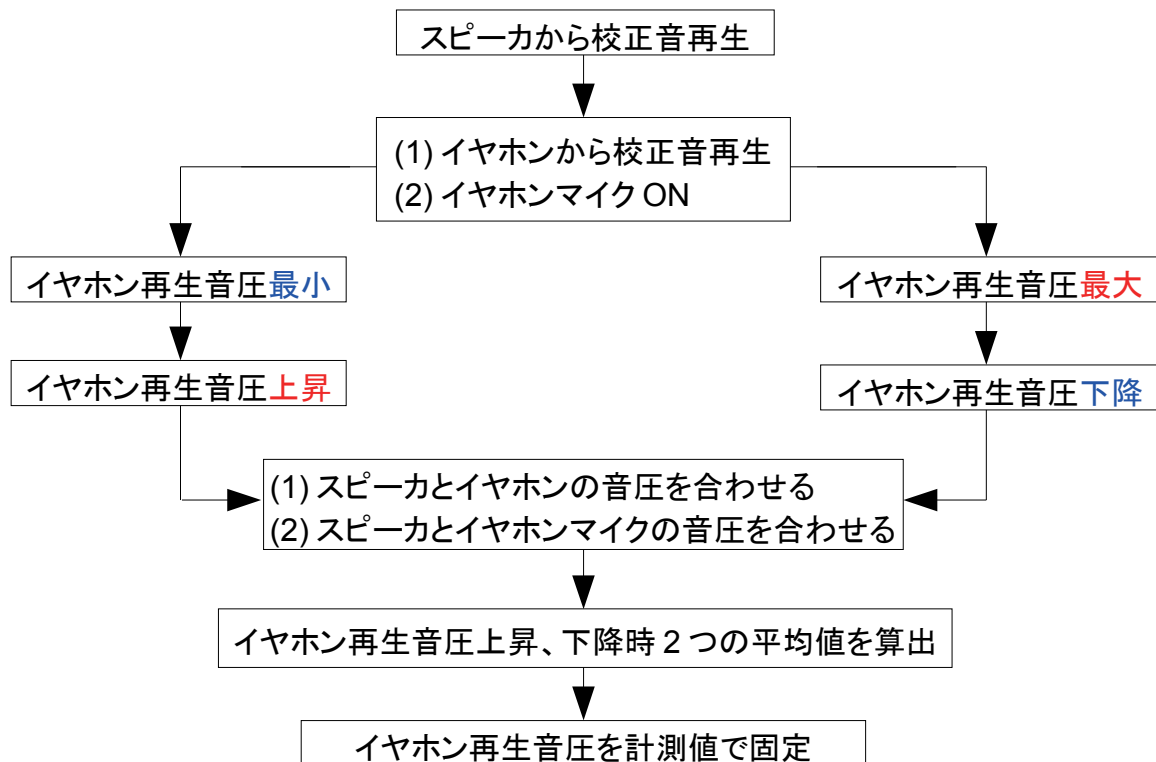


図 6. 音圧校正の流れ

4-4 話者 2 名の対向会話実験の結果

表 2 と図 7 に話者 2 名による対向会話実験結果を示す。被験者は、20 代の 6 名である。四角のプロット点がヘッドトラッキング有の平均値、ひし形のプロット点がヘッドトラッキング無の平均値を表している。ヘッドトラッキングの有無で比較してみると、会話のしやすさ以外の項目でヘッドトラッキング有よりもヘッドトラッキング無の方が高いという結果になった。しかし、「聞きやすさ」、「違和感を感じるがあったか」という項目以外では、MOS 値の差異が 0.5 以上なく、ほとんど差が出ないという結果になった。ちなみに、この実験結果で有意差検定を行ったものの、どの項目でも、有意差は見られなかった。

表 2. 話者 2 名による会話実験の MOS 評価結果

評価項目	ヘッドトラッキング有	ヘッドトラッキング無
相手の会話音声の音質	3.50	3.50
会話のしやすさ	3.83	3.50
聞きやすさ	3.33	3.83
話しやすさ	3.33	3.40
使いやすさ	3.00	3.17
違和感を感じる時があったか	3.83	4.67
音声の途切れ	4.50	4.83
音声の遅延	4.67	4.83

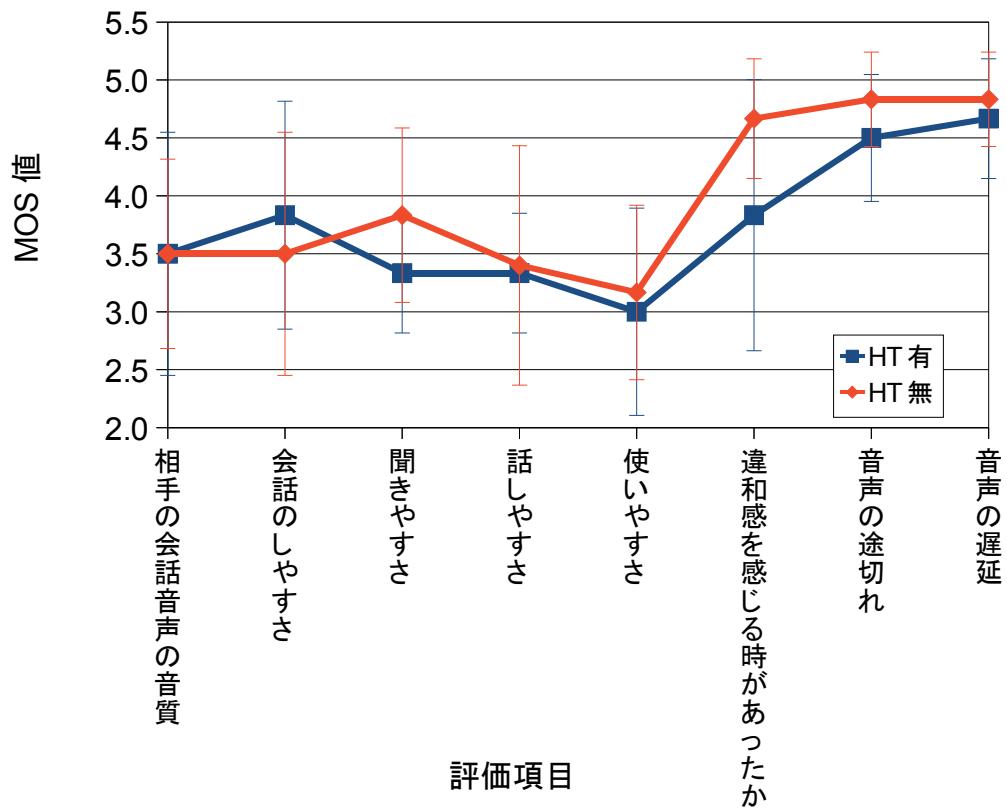


図 7. 話者 2 名による会話実験の MOS の評価結果

4-5 話者 3 名による会話実験

図 8 に 3 名による会話実験の実験系と音源配置を示す。実験開始時は、被験者は PC がある方向を向いて、これを正面とする。同じ部屋にいる人物は正面を見ている被験者から、距離 1.5m の左 90° の位置とする。試験開始時の仮想音源の配置位置は、被験者から距離 1.5m の正面に配置する。

所要時間はインストラクションに 10 分、会話実験が 20 分(10 分×2 回)、評価に 10 分(5 分×2 回)の 40 分程度である。インストラクションで行う内容は、実験 1 と同様である。

3 人での会話のタスクとして軍艦ゲームを行う。軍艦ゲームとは、相手の船を沈めるのが目的のゲームである。この実験では、被験者は、ゲームの中継役および進行役を務める。他の 2 人は、ゲームを行い、順番に盤面上の砲弾着地位置を被験者を通して相手に伝え、その着弾を同じく被験者から教わることとした。

プレイヤーは、まずゲーム開始前に 6×6 のマス目の中に自軍の船をシート上に記入する。プレイヤーは 2 人で、1 人は被験者と同じ部屋にいる人物、もう 1 人は隣の部屋にいる人物とした。記入後、同じ部屋にいるプレイヤーが先攻でゲームを開始する。ゲーム開始後、攻撃側のプレイヤーに順にどの位置を攻撃するかを聞く。被験者は攻撃側のプレイヤーが指定したマスを防御側のプレイヤーに伝える。そして、攻撃の成否を防御側のプレイヤーに聞く。攻撃が成功した場合、攻撃側のプレイヤーが続けて攻撃する。攻撃が失敗した場合、ターン終了となり、攻守が入れ替る。ゲームの終了条件は、どちらかが敵の船を全機沈めることとした。

試験後に主観評価シートに記入してもらい、ヘッドトラッキングの有無を切り替えて、他は同じ条件でゲームを繰り返すこととした。ヘッドトラッキングの有無の試験順番はランダムとした。

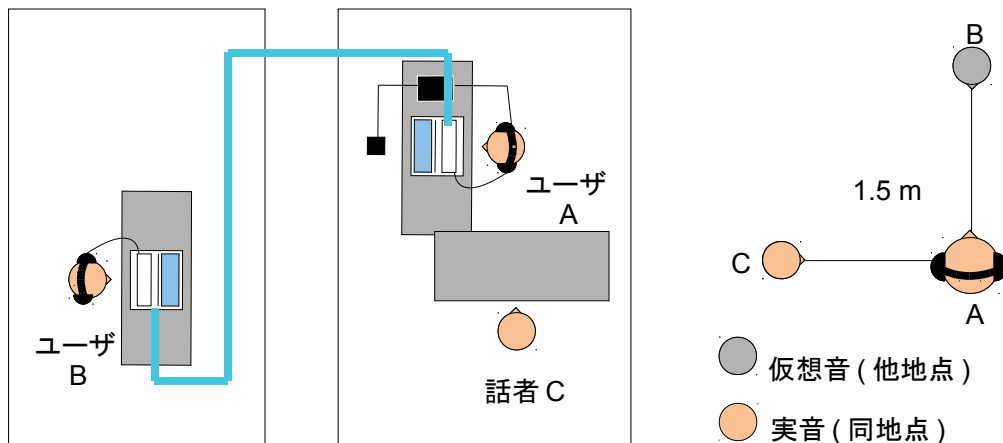


図 8. 実験 2 の実験系と音源配置

4-6. 話者 3 名による会話実験の評価結果

表 4 と図 9 に 3 名による会話実験結果を示す。被験者は、20 代の 6 名である。

ヘッドトラッキングの有無で比較してみると、実験 1 の結果とは違い、「違和感を感じるがあったか」でヘッドトラッキング無よりもヘッドトラッキング有の方が 0.50 高いという結果が得られた。しかし、「聞きやすさ」の項目は実験 1 と同様に、ヘッドトラッキング有よりもヘッドトラッキング無の方が高いという結果になった。他の項目では、MOS 値にほとんど差が出ないという結果になった。実験 1 と同様に、この実験結果で有意差検定を行ったものの、どの項目でも、有意差は見られなかった。

表 4. 話者 3 名による会話実験の MOS 評価結果

評価項目	ヘッドトラッキング有	ヘッドトラッキング無
相手の会話音声の音質	4.33	4.33
会話のしやすさ	4.17	4.33
聞きやすさ	3.67	4.00
話しやすさ	3.67	3.67
使いやすさ	4.17	4.17
周囲音の聞こえ方	4.00	4.00
違和感を感じる時があったか	4.33	3.83
音声の途切れ	4.17	4.17
音声の遅延	4.33	4.33

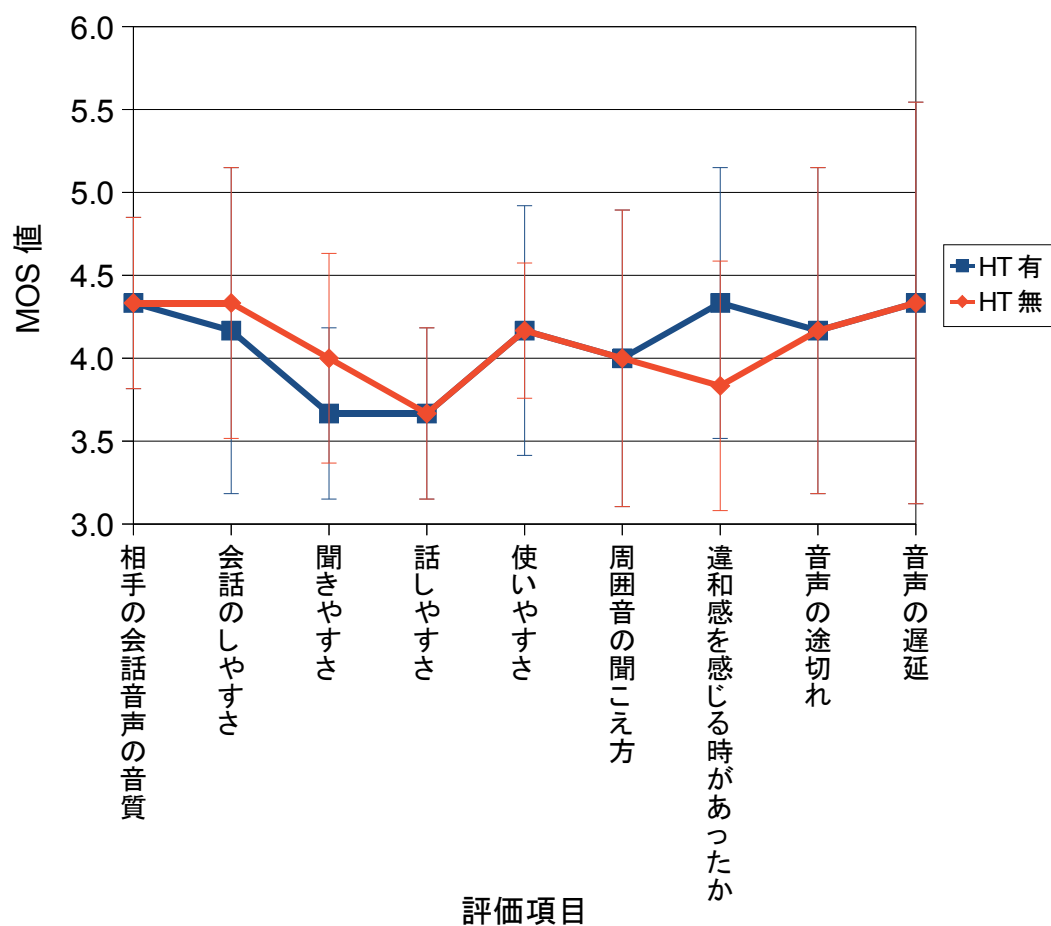


図 9. 話者 3 名による会話実験の MOS の評価結果

4-7 考察

聞きやすさの項目は、実験 1、2 ともに「ヘッドトラッキング有り」が「ヘッドトラッキング無し」を下回った。実験後の意見聴取でヘッドトラッキング無しの会話が、他地点の人との会話で普段使用している電話に聞こえ方が類似していたという意見が多数あった。そのため、ヘッドトラッキングされた音声を聞き慣れていなかったことが、ヘッドトラッキング無の評点が高くなった理由の一つと思われる。

「違和感を感じるがあったか」の項目は、実験 1 では、ヘッドトラッキング無しがヘッドトラッキング有りを、0.84 上回った。一方、実験 2 では、「ヘッドトラッキング有り」が「ヘッドトラッキング無し」を、0.50 上回った。実験 1 ではヘッドトラッキング無のほうが評点が高かったのに対し、実験 2 で「ヘッドトラッキング有り」のほうが評点が高くなった。これは、ヘッドトラッキングによる音像固定が効果的に働き、通常会話と同じ状況を構築できたためと考えている。

実験 1、2 ともに有意差検定を行ったものの、どの項目でも、有意差はなかった。被験者数が少ないことが原因の一つとして考えられるので、被験者数を増やして有意差が出るかを検証する必要がある。

5. 結論

本研究では、3次元音響会議システムにヘッドトラッキングを導入した時の効果を検証するため、実際の音声会議使用を想定し、システムを利用した会話実験を2名の場合と3名の場合の2種類を行った。どちらの場合も、被験者に会話実験中、頭部運動を積極的に行う会話タスクを使用した。ユーザ体感品質を検証するために、MOS 試験を実施した。ヘッドトラッキングが有る場合と無い場合で、選定した9項目に対してMOS 評点を集計し、比較した。そのうち、聞きやすさの項目は実験1、2ともにヘッドトラッキングが有る場合がヘッドトラッキングが無い場合を下回った。「違和感を感じるがあったか」の項目は、実験1では、「ヘッドトラッキング無し」が「ヘッドトラッキング有り」を、0.84 上回った。一方、実験2では、「ヘッドトラッキング有り」が「ヘッドトラッキング無し」を、0.50 上回った。

実験結果を見る限り、仮想音のみを用いた2名による対向会話では、音像固定の比較対象となる実音が存在しなかったため、ヘッドトラッキングの効果が見られなかった。しかし、「違和感を感じる時があったか」の項目で、実験1に比べ、実験2の方がヘッドトラッキング有の評点が0.50 向上したため、違和感が改善され、会話の自然さを向上させたと考えられる。よって、実音と仮想音が混在する会話において、ヘッドトラッキングが有効であることがわかった。

今回は、ヘッドトラッキングの導入効果を調べるための実験であったため、パケット損失による音声の途切れや雑音による評価の影響を考慮してバッファサイズを使用した3次元音響シミュレーションシステム Slab3d 内で最も大きい8192 byte を採用した。そのため、システム遅延を考慮していない。ちなみに、システムの遅延時間は、バッファサイズとサンプリング周波数の比で決定される。本研究の実験1、2ではサンプリング周波数を44.1 kHz としているため、システムの遅延時間は185.8ms である。一般に会話システム遅延の検知限は80 ms とされている[5]。本研究のシステム遅延はこの検知限を大幅に越しているため、システム遅延によって、会話の自然性が損なわれる可能性が生じる。

よって、短期的な課題として、システムの遅延も考慮してサンプルレートやバッファサイズを調整した上で実験をする必要がある。また、Slab3d より処理時間の少ないシステムが求められる可能性もある。その場合には、Slab3d を参考にして自ら音声会議システムの開発を行う予定である。

また、本実験では、被験者数が6名となっているが、MOS 評価の人数としてはやや少ない。そのため、実験ごとのグラフに表した標準偏差も大きく、有意差がない状態である。よって、追加実験を実施して、評価人数を増やす必要がある。

さらに、今回は音声のみを立体化したので、相手の位置、相手が会話中にどんな作業をしているか、どんな表情をしているか、などの視覚情報が入ってこない。人間の得る情報は視覚によるものが多く、臨場感や自然性を向上させる可能性があるため、視覚情報と聴覚情報を複合した実験環境の検討が必要かと思われる。

3次元音響ナビゲーションは、視覚情報に加えて、音の到来方向から目的位置を認識させるものである。例えば、全地球測位システム(GPS: Global Positioning System)などによって得られる位置情報を利用し、人ごみの中で相手の姿は見えなくても、声が実際にその人がいる方向から聞こえてくるように感じることで、相手の居る位置が分かる、というような待ち合わせナビゲーションも考えられる[6]。

また、美術館や博物館において、あるエリアに入ると、そのエリアの展示物の説明が、各展示物の位置から聞こえてくるようにすることで、どこに何が展示されているかが音により分かる、というようなバーチャルガイドや、繁華街において、広告の音情報とその店や商品の方向から聞こえてくるタウンガイドのような用

途も広がる[6]。いずれも周囲音と同時に音源定位した音声案内を混合提示する拡張音響現実感が有効なアプリケーションである。これらのアプリケーションでは本研究の実験と同様にイヤホン外側にマイクを取り付けたイヤホン・マイクにより、システム利用者の周囲音も聞こえるように構成する。また、システム利用者の頭部方向を考慮して、案内対象の方向に立体音声を定位させて音声情報を提供する。これにより、限定的な音声による伝達に対象物のおよその方向に関する情報も同時に与えるモバイル用3次元音響ナビゲーションが考えられる。

【参考文献】

- [1] 飯田、森本（編著）、「空間音響学」、コロナ社、ISBN 978-4-339-01322-1（2010年8月）
- [2] Jens Blauert, “Spatial Hearing: The Psychophysics of Human Sound Localization, Revised Edition,” MIT Press, Cambridge, MA, ISBN 0-262-02413-6（1997）
- [3] Polhemus FASTRAK, http://www.polhemus.com/?page=Motion_Fastrak
- [4] J. D. Miller, Slab3d, <http://slab3d.sonisphere.com>
- [5] 矢入 聡、岩谷 幸雄、鈴木 陽一、「頭部運動と聴覚ディスプレイのシステム遅延の関係に関する一考察」信学技報、E A2005-38、pp.23-28（2005,8）
- [6] 安田泰代、大矢智之「リアリティ音声音響通信技術」NTT DoCoMo テクニカルジャーナル、Vol. 11, No. 1, pp. 55-62（2003）

〈発表資料〉

題 名	掲載誌・学会名等	発表年月
三次元音響会議システムにおけるヘッドトラッキング導入効果の評価	東北地区若手研究者研究発表会予稿集	2012年3月
Evaluation of localized speech intelligibility from bone-conduction headphones with competing noise for augmented audio reality	Proceedings of the 40th International Congress and Exposition on Noise Control Engineering (Internoise)	2011年9月