

(7) リアリティ音声音響通信技術

モバイルブロードバンドの実現により、移動通信環境下でも高臨場な通信が可能となる。伝送する音声音響の品質を究極まで高め、立体音場を再現することにより、移動通信における新しいコミュニケーションスタイルを実現する技術について述べる。

やすだ やすよ おおや ともゆき
安田 泰代 大矢 智之

1. まえがき

高速データ通信インフラの発達と、その高速性を活かしたアプリケーションの浸透は、車の両輪のようにお互いに依存関係にあり、今日のインターネットアプリケーション発展のダイナミクスを支える源となってきた。移動通信においても、第3世代移動通信（IMT・2000：International Mobile Telecommunications・2000）や無線LAN（Local Area Network）をはじめとする高速データ通信のインフラが急速に整いつつあるが、モバイルブロードバンドの利点を十分活用するアプリケーションの登場は、今一歩、出遅れているのが現状である。

音声を伝送するのが携帯電話の出発点であったが、インフラの進歩によってブロードバンド環境が安価に提供できる時代になると、より豊かな基本音声機能を誰もが利用で

きるようになると期待できる。図1に示すように、移動通信においては、旧来、音声通信のみのモノラル・モノメディアであったが、M・Stageのステレオ音楽伝送やFOMA（Freedom Of Mobile multimedia Access）のビデオホンに代表されるように、徐々にマルチ次元化／マルチメディア化に向かって進化してきた。このマルチ次元化を極限まで高めた形態がリアリティ音声音響通信である。立体（3D）音場そのものを伝送することにより、あたかもその場所にいるような感覚を再現した、仮想音空間を構築することが可能となる。これがさらに発展し、視覚情報を含めたマルチメディア環境に究極までリアリティを持たせたものが、バーチャルリアリティ通信である。

ドコモでは、これらの究極のリアリティの実現が新たな

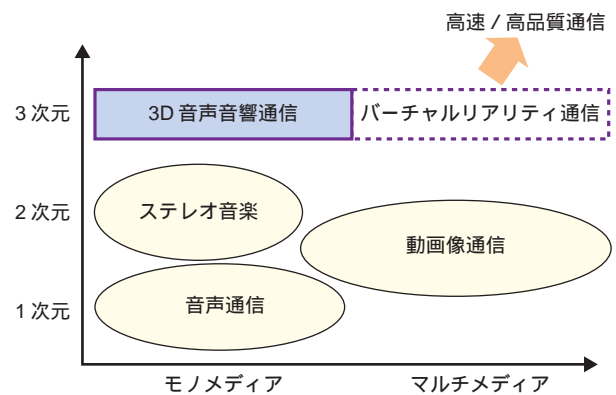


図1 マルチ次元・マルチメディア通信

アプリケーションや、ひいては新たなモバイルコミュニケーションスタイルを生み出し、モバイルブロードバンド通信の可能性が新しいスパイラルに突入することを目指して研究を進めている。

本稿では、特に音メディアの伝送品質を極限まで高めるモバイル3D音響音響通信技術に焦点をあて、その原理を解説するとともに、具体的なサービスイメージおよびプロトタイプシステムを紹介する。

2. バーチャル音響技術

2.1 劇場向けサラウンド音響技術

最も身近な立体音場は、映画館のマルチスピーカシステムであろう。Dolby Surround [1], Dolby Digital [2], DTS [3]などのフォーマットが広く実用化されており、記録されたマルチチャンネル音源を、劇場に備え付けられた複数のスピーカから再生することによって、サラウンド音響空間を再現する。同様のシステムはホームシアター向けにも実用化されており、DVDの普及に伴い、5.1チャンネル(図2)に対応した民生用サラウンドスピーカシステムも数多く生産されている。

しかしながら、これらのシステムは、例えば受聴者の後方から音源を提示するような方向定位性の再現を主な目的としており、複雑な音空間を忠実に再現するような用途には十分ではない。さらにモバイル環境下では、このような多数のスピーカを持ち歩くことは現実的ではなく、移动通信への直接的な適用は難しい。

この問題を解決するのに有効な、仮想的に立体音場を創出するバーチャル音響技術[4], [5]について以下に紹介する。

2.2 バイノーラル再生技術

人間は目を閉じていても音の到来方向をかなりの精度で推定することができる。図3に示すように、音源が左右の耳に到達する時間差や音圧差を知覚することにより、音源aの到来方向を推定することが可能となる。音源bは同じ到来時間差を持つが、音源から鼓膜までの音響伝送路の周波数応答が耳介(耳たぶ)の影響により異なるため、経験的に前方/後方を区別することができると考えられている。この原理を利用し、ヘッドホンで左右の耳に別々の音響信号を与えることで、実際にその場にいるのと同じ音環境を再現するのがバイノーラル再生技術である。

通常ステレオ録音された音楽をヘッドホンで再生すると、音が頭の中で鳴っているように感じる。これを「頭内音像定位」と呼ぶ。これに対し、バイノーラル再生技術を利用した音をヘッドホンで再生すると、あたかもその音源

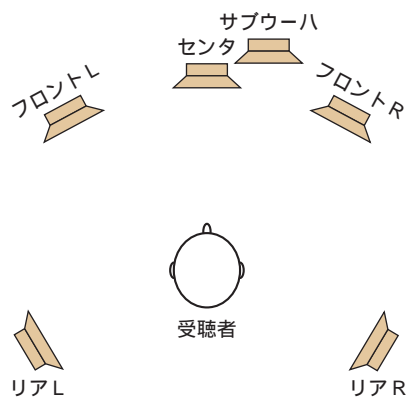


図2 5.1チャンネルサラウンド

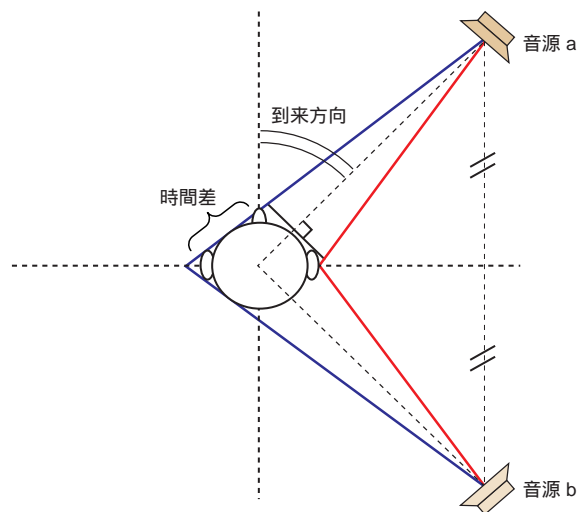


図3 音源到来方向推定

の場所から音が鳴っているかのように自然に聞こえ、ヘッドホン再生にもかかわらず外部から音が聞こえてくるように感じる「頭外音像定位」が可能となる。

(1) バイノーラル録音

実際の環境で受聴者の左右の耳に到達する音を忠実に再現する手法として、擬似頭(ダミーヘッド)(写真1)の左右の耳に取り付けたマイクロホンで音を収録し、その音をヘッドホンで再生する方法が用いられる(図4)。臨場感が重視されるクラシック音楽やオーディオドラマといったジャンルを中心に、ヘッドホン再生を前提としたバイノーラル録音のCDも数多く発売されている。

(2) シミュレーションによるバイノーラル再生

バイノーラル再生を実現するもう1つの手法として、音源から人間の左右の耳までの音の伝わり方を表す頭部伝達関数(HRTF: Head Related Transfer Function)[6]をあらかじめ測定しておき、シミュレーションによりバイノーラル信号を生成する方法がある。

測定したHRTFの例を図5に示す。伝達関数の係数列



写真1 ダミーヘッド

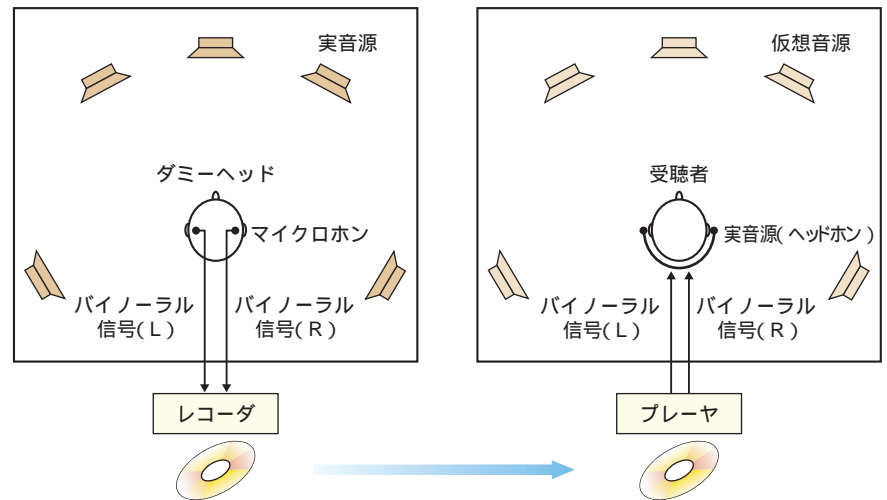


図4 バイノーラル再生系

は、インパルス波形を音源とした場合の、左右の耳に到達する音響波形となる。左右のHRTF間に見られる時間差 (ITD: Interaural Time Difference) と振幅差 (IID: Interaural Intensity Difference) という、人間が実環境の中で音源の方向知覚を得る主要な手がかりがHRTFとして表現されていることが分かる。いったんこのインパルス応答フィルタを測定しておけば、任意の音源波形との畳み込み演算を行うことで、シミュレーションによりバイノーラル信号の生成が可能となる。

時間軸で畳み込み処理を行う場合、HRTFフィルタ長 N タップに対し $\mathcal{O}(N^2)$ の演算量を要する。実環境でのインパルス応答は、部屋の音響特性により数秒におよぶこともあるため、CD帯域 (44.1 kHz サンプリング) の信号の畳み込み演算を実時間で行うためには、数G operation/secの積和演算を必要とし、現実的なDSP (Digital Signal Processor) の演算能力を超えてしまう。そのため、高速フーリエ変換 (FFT: Fast Fourier Transform) により演算量を $\mathcal{O}(N \log N)$ に削減する方法が利用されることもあるが、ブロック単位での演算となることから処理遅延が発生する。最近では、これらのトレードオフを解決する検討も盛んに行われており、近年のDSPなどによる信号処理技術の発達ともあいまって、実時間で処理を行うことが可能になってきた。また、HRTFにはITDやIIDのほかには頭部や耳介な

どによる音波の反射・回折の影響が含まれており、頭部や耳介の形状の個人差が定位再現性に大きな影響を与えることも知られている。現在では、フィルタ特性の緩和手法などを利用し、これらの問題を解決して実用化された製品も入手可能になってきている。

レイク・テクノロジー社[7]が開発しドルビー社がライセンス供与しているドルビーヘッドフォン[8]は、先に述べたHRTFの畳み込み信号処理により、通常のステレオヘッドホンでサラウンド音響を再現するシステムである。ホームシアター製品だけでなく、PCやポータブルMDプレーヤにも搭載されており、DVDなどの5.1チャンネルサラウンド再生を手軽に体験することができる。

(3) ヘッドトラッキング機能付きバイノーラル再生

バイノーラル再生技術によりヘッドホン再生においても頭外音像定位が可能であるが、これだけでは頭を動かすとその音像も一緒に動くという、自然界ではあり得ない状況が起こる。これに対し、頭の動きに追従するヘッ

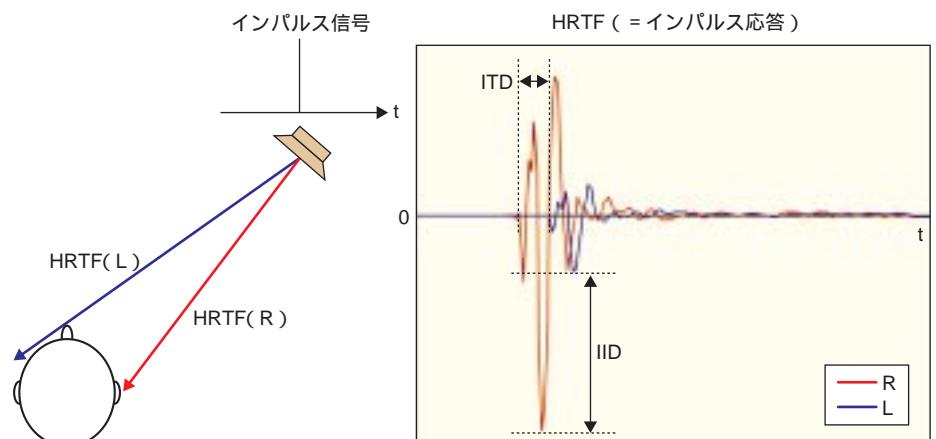


図5 頭部伝達関数

ドトラッキング機能を用いれば、より自然で違和感のない音場を再現することができる。受聴者の頭の向きをフィードバックし、HRTFフィルタを適応的に逐次更新することにより、原理的にはその場にいるのと同じ音響空間を創出することができる。

人間は両耳情報の動的な変化からも音像情報を得ていると考えられるので、頭の動きをフィードバックすることによって方向知覚がより正確になるため、バーチャル音響を用いたアプリケーションを構築する上での大きな助けになると期待できる。

図6に頭部を固定した場合の方向知覚精度の試験結果を示す。写真2のようなマルチスピーカシステム系を用い、被験者の正面を0°とし、音が聞こえてくる方向を答えるという試験で、再生方式（スピーカ再生/バイノーラル再生）および方向（30°間隔の12方向）をランダムに呈示した。被験者は21名で、ランダムにパターンを変えて2試行を行った。正答率のグラフにおいて、青いラインがスピーカ再生時、赤のラインがシミュレーションによるバイノーラル再生時である。円の中心に近いほど正答率が低いことを示す。これより、頭部を固定していると、実際にスピーカから音を鳴らす実音場でも正答率は悪く、もっとも知覚しやすいL_90°、R_90°の場合でも80%~90%程度、斜め後方では30%程度までに正答率が低下する。しかも、バーチャル音場であるバイノーラル再生ではさらに正答率が下がることが分かる。これを、頭部の動きを許すと、スピーカ再生での正答率が100%に改善されることはもちろん、バイノーラル再生においても、頭の動きに追従させるヘッドトラッキング機能により正答率を100%にすることが可能となる。

このヘッドトラック技術が実用化された例としては、Sony社のデジタルサラウンドヘッドホンシステム[9]がある。頭を動かしてもサラウンドの音場が固定されているが、センサのドリフト/精度、計算量の制限から生じるHRTF精度の問題などがあり、まだ、この技術が広く使われるようには至っていない。

2.3 トランスオーラル再生技術

上に述べたように、ヘッドホンを装着することを許容すれば、バイノーラル再生技術によりバーチャル音響空間を作り出すことができるが、これをヘッドホンなしで実現するのが、トランスオーラル再生技術である。通常のステレオ再生用の2台のスピーカからバイノーラル信号を再生すると、右耳用の信号(X_R)が左耳にも到達してしまうため、音像は正しく再現しない。この現象はクロストークと呼ば

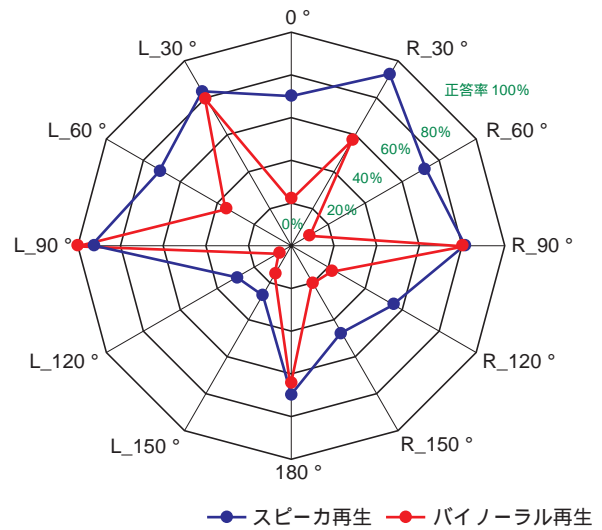


図6 頭部固定時の方向知覚精度試験結果



写真2 方向知覚精度試験設備

れる。しかし、HRTF測定と同様の手法でスピーカから逆の耳までの伝達関数を知ることができれば、図7に示すように、適切な逆フィルタ演算によりクロストークをキャンセルすることができる。それぞれの耳に聞こえるべき音をそれぞれの耳にだけ与えることにより、バイノーラル再生と等価な音像再現を、スピーカによって行うことが可能となる。この技術も、以下のようにすでに商用化された例がある。

ドルビーバーチャルスピーカ[10]...レイク・テクノロジー社が開発し、ドルビー社がライセンス供与してPC搭載用として商品化されている。トランスオーラル技術により、前方に設置した2つのスピーカだけで5.1チャンネルサラウンドが体感できる。

3Dサウンド・スピーカ・システムP2DiPOLE [11]...1ボディ・2スピーカだけで5.1チャンネルに対応している。クロストークをキャンセルしやすくするために、2つの円筒形スピーカを近接して配置しているところに特徴がある。

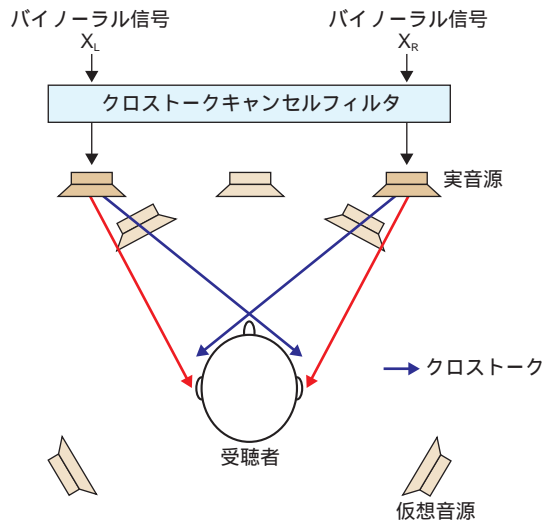


図7 トランスオーラル再生系

3. モバイル3D音声音響通信

前章で挙げた実用化の例は、ヘッドホン再生/スピーカ再生の違いはあるが、どれも主にDVDをはじめとする5.1チャンネルサラウンドを家庭でも手軽に楽しめるよう製品化された、ホームシアター用途である。移动通信へのパーティクル音響技術の適用に関してはまだ十分な検討がなされておらず、未開拓な分野といえる。

本章では、2章で解説したパーティクル音響技術を移动通信に適用した、モバイル3D音声音響通信の具体的なサービスイメージを挙げるとともに、モバイル3D音声音響通信を実現するための技術課題について述べる。

3.1 サービスイメージ

パーティクル音響を利用したアプリケーションとしては、現実の空間と全く同じように音源を定位させる、究極のリアリティ実現手段として利用するものや、現実とは異なる仮想空間を人工的に作り出すものがある。

以下に述べるのは、この技術を応用したサービスの一例である。

(1) 高臨場感サービス

基本音声サービスやTV電話サービスにおいて、相手の声が耳元から聞こえるのではなく、あたかも相手と向き合っているような感覚を与えたり、TV電話の画面から声が聞こえたりというように、現実に近い形で会話を行うことが可能となり、長時間の使用でも疲労度が少なくなると期待できる。

また、三者通話では、図8に示すように話者を仮想的に分離して配置することにより、誰が発言しているかと

いう話者分離が容易になる。さらに、これに頭の動きをフィードバックすることで、遠隔地から会議に参加する場合などにおいても、実際にその場（会議室）にいるかのように自然な音場の中で通話することが可能になる（図9）。参加者数が多くなると、話者分離の効果がより顕著になり、遠隔会議での複雑な議論のやりとりをより自然な形で知覚できるようになる。

さらに、マルチメディア配信や放送型サービスにおいても、本技術を適用して高臨場化することが可能である。いつでもどこでも映画館のようなサラウンド音が楽しめるモバイルシアターや、その場の雰囲気までもが伝わってきて会場と一体感が味わえるスポーツやコンサートの放映などが考えられる。

(2) 3D音場ナビゲーション

視覚情報に加えて音の到来方向感から目的位置を認識させる、3D音場でのナビゲーションも考えられる。例えば、図10に示すように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を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場所が分かる、というような待合せナ

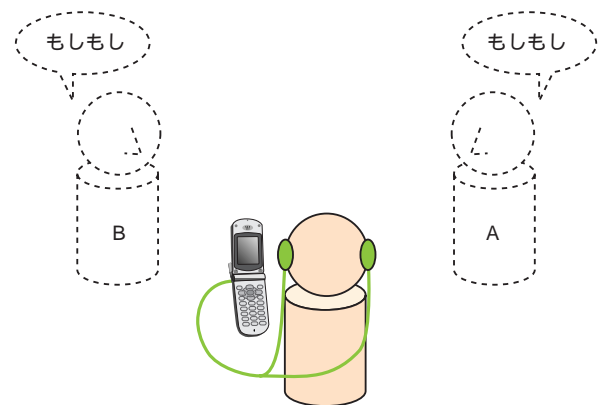


図8 3D音響三者通話イメージ

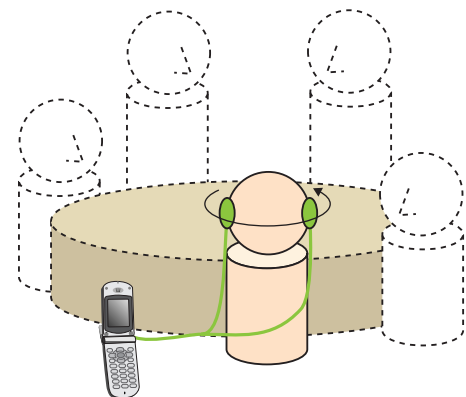


図9 3D音響遠隔会議イメージ

ビゲーシオンが可能となる。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。

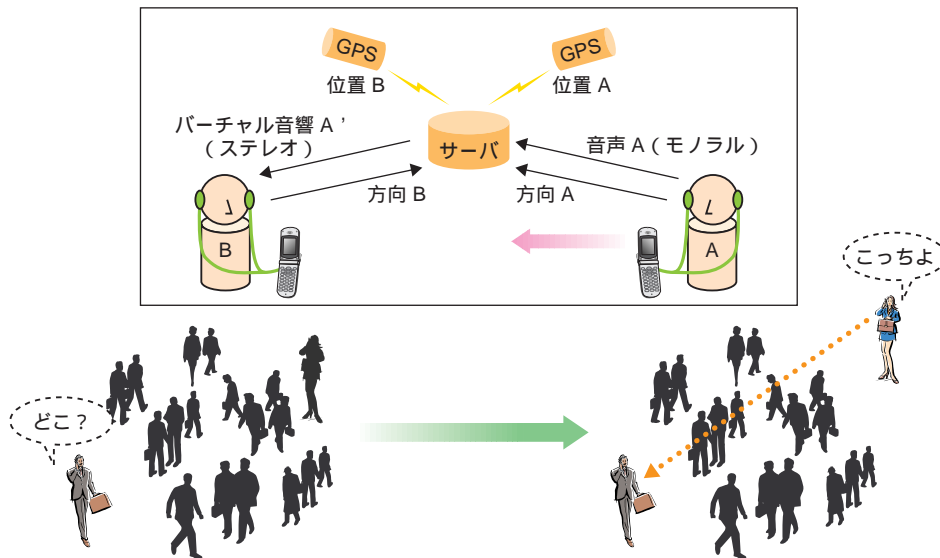


図10 待ち合わせナビゲーシオンイメージ

(3) 3D音場ゲーム

PCやTVゲームにオンラインゲームが登場したり、ドコモの携帯電話サービスのiアプリでゲームができるなど、通信と娯楽が融合しつつある。しかしながら、現在は、それらのゲームにおいて「音」は背景音楽(BGM)に過ぎない。シューティングゲームやRPGやアクションゲームに3D音響を組み合わせることで、よりリアルに没入できる世界を提供できる。

携帯電話で近くの敵と撃ち合いをするゲームがすでに欧州で登場している[12]が、このような位置情報をベースにしたゲームやテーマパークのアトラクションなどに3D音響技術を取り入れることで、従来の映像を中心とした構成に「音を手がかりにする」という要素を加えることができる。ユーザのロケーション、モビリティを考慮する点はモバイルならではのものであるといえる。

3.2 技術課題

これまで述べてきたように、バーチャル音響を再現するための基礎技術は、すでにさまざまな分野で発達しつつある。しかしながら、図11に示すように、これらを移動通信の分野に応用するためには、まだまだ解決すべき課題がある。1つはHRTFを中心とした信号処理技術である。移動通

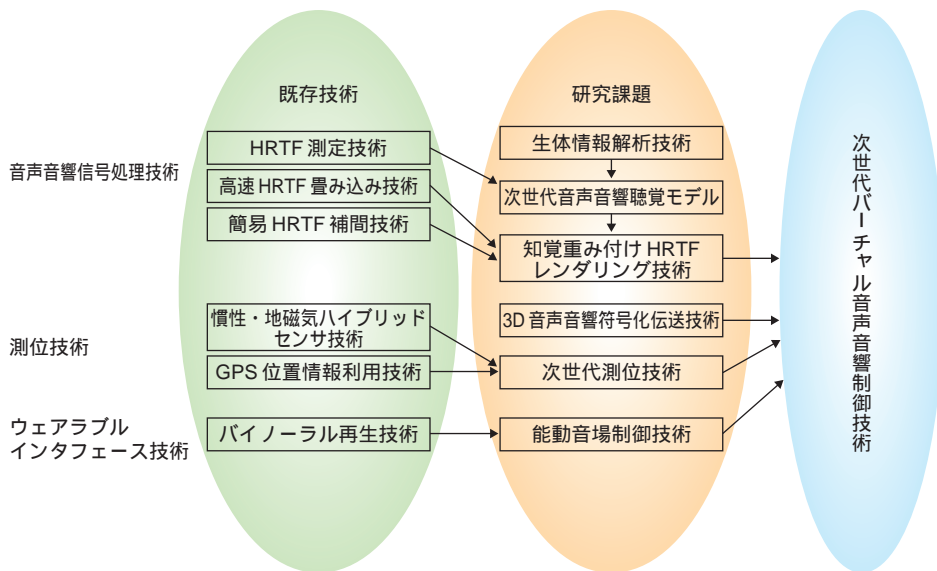


図11 次世代バーチャル音響音響制御の技術課題

信に適するように、簡易で低演算量のフィルタリングを行うためには、人間の知覚特性を積極的に利用した最適化が必要であり、知覚機構の解析およびモデリングに基づいた知覚重み付けバーチャル音響再現技術が鍵となる。また、これらの情報をどのように移動通信路で伝送するか、トランスポート面の検討や、携帯環境に適した測位技術、ヒューマンインタフェースの高度化も、移動通信ならではの技術課題といえる。

4. プロトタイプシステム

モバイル3D音響音響通信の実現性を検討する第1ステップとして、3.1項で述べたサービスイメージを可視化したプロトタイプシステム(図12)を構築した。

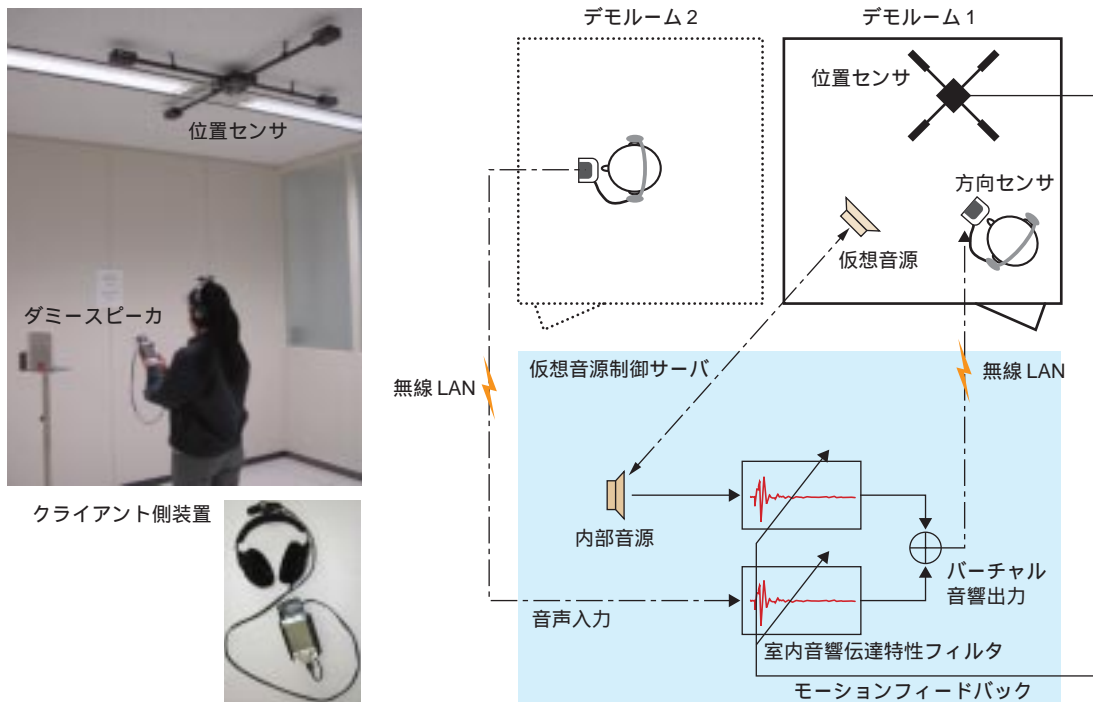


図12 3D 音声音響通信プロトタイプシステム

本章では、そのシステムについて紹介する。

4.1 システム構成

本システムは、位置センサ、方向センサ、携帯情報端末 (PDA: Personal Digital Assistant)、ヘッドホン、仮想音源制御サーバおよび無線LANから構成される。位置センサは天井に取り付けられ、受聴者の位置を検出する。方向センサはヘッドホンに取り付けられ、受聴者の頭の動き(方向)を検出する。得られた位置および方向情報(図13)を基に仮想音源制御サーバでバーチャル音場を生成し、無線LANでクライアント側へ送信する。クライアント側はPDAと無線LANによりモバイル感覚を実現している。複数のユーザはPDAに備えられたマイクロホンによって、互いに会話することができ、バーチャル音響環境による通信を再現している。

4.2 技術概要

位置センサはX・Y・Zの各軸について精度7 mm RMS (Root Mean Square) で位置を検出できるが、本システムでは水平面であるX・Y軸の情報のみを利用している。人間の方向知覚は水平面よりも垂直面に対しては鈍いとされていることから、Z軸(高さ情報)は用いないことで処理量を削減している。これと同じ理由から、方向センサに関しても水平面である方位角の情報のみを利用している。方向センサの検出精度は0.25 °RMSである。

仮想音源制御サーバでは、レイク・テクノロジー社が有する、仮想的に長いフィルタを低遅延で実時間畳み込み処理を行うという技術を利用し、PCベースでのバーチャル音響作成が可能となっている。畳み込みフィルタは頭部伝達特性と室内音響伝達特性を併せた約7000タップの有限長インパルス応答(FIR: Finite Impulse Response)フィルタである。このフィルタ長は、典型的な低残響のリスニングルームの室内音響特性と同等である。本システムでは8つの方位角のHRTFだけを持ち、その他の方位角については補間を行うことによりメモリ量を削減している。さらに、距離感の強弱で表現することにより、距離に応じたHRTFを用意する必要がなく、メモリ量を削減している。

以上のような処理量およびメモリ量の削減を行うことで、同時に3ユーザまでのモーションフィードバックやHRTF畳み込み演算の実時間処理を、標準的なIntel® Pentium 4のPC上で実現している。

4.3 機能

本システムでは、以下に記述する2つの機能を実現した。1つ目は、3D仮想音響空間の再現である。受聴者の動きを実時間処理によりフィードバックすることで、デモルーム内の任意の位置に仮想音源を定位させることができる。例えば、デモルーム1にダミーのスピーカボックスを置き、その位置に音源を定位させるとする。実際にはそのダミースピーカは音を発していないのだが、ヘッドホンを介して

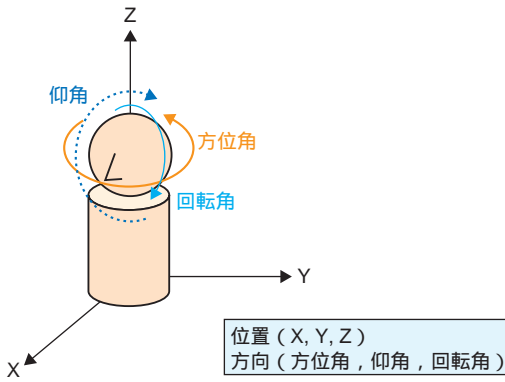


図13 位置方向情報

創出された仮想音響空間ではスピーカから音が鳴っているように聞こえ、受聴者がデモルーム内を自由に動き回っても、その音は変わることなくダミースピーカの位置に定位し続けるようになっている。より自然で違和感のない音場での通信イメージを具現化したものである。

2つ目は、ユーザ間の位置関係を仮想的に再現することによる3D音響通信機能である。デモルーム1と2にいるそれぞれのユーザには、ヘッドホンを介して聞こえる相手の声が、実際の位置関係どおり、壁の向こうに定位するようになっている。姿は見えなくても声の到来方向感から相手のいる位置が分かる、3D音場ナビゲーションのサービスイメージを具現化したものである。

いずれも、これまで述べてきたバーチャル音響を応用したサービスイメージの基本的な機能を可視化するもので、次世代の音響サービスの高度化のイメージを具体的に体験することが可能となった。

5. あとがき

ワイヤレスブロードバンドにふさわしい魅力あるサービスを実現する技術の1つとして、リアリティ音響通信技術について解説し、そのプロトタイプシステムを紹介した。今後は実網への適用を視野に入れ、このプロトタイプシステムを用いて定位精度向上と処理量削減のトレードオフ問題、ネットワーク遅延や音響符号化の影響などの検証など、技術課題に取り組む。さらに、モバイルならではのロケーション、モビリティを考慮した音場制御および接続技術の研究開発をすすめていく。

携帯電話端末を片耳にあててモノラルの音声会話をする現在のコミュニケーションスタイルが、この技術によってヘッドホンやイヤホンをつけて3D音場で会話をするよう

になり、さらには携帯電話の小さなスピーカでも3D音場制御が可能になってヘッドホンすら不要になっていくかもしれない。携帯電話は音や情報を伝えるデバイスからバーチャルリアリティ、つまりその場の雰囲気や感覚までも伝えるデバイスへと進化していくだろう。

文献

- [1] <http://www.dolby.com/tech/>
- [2] <http://www.dolby.com/digital/>
- [3] <http://www.dtsonline.com/aboutdts/index.shtml>
- [4] 大矢, ほか: “音響符号化技術”, 本誌, Vol.8, No.4, pp.17 - 24, Jan.2001.
- [5] 北脇信彦, 菅村昇, 小泉宣夫: “音のコミュニケーション工学”, 初版, pp.172 - 178, 1996, コロナ社, Japan.
- [6] Corey I.Cheng, Gregory H.Wakefield: “Introduction to Head-Related Transfer Functions (HRTFs): Representations of HRTFs in Time, Frequency, and Space”, J.Audio Eng. Soc., Vol.49, No.4, 2001.
- [7] <http://www.laketechology.com/>
- [8] <http://www.dolby.co.jp/AV/DH>
- [9] <http://www.sony.jp/products/Models/Library/MDR-DS8000.html>
- [10] http://www.dolby.co.jp/AV/speaker/DVS_WhitePaper_CEJ.pdf
- [11] <http://www.iodata.co.jp/products/sounds/p2dp/index.htm>
- [12] <http://www.hotwired.co.jp/news/news/20030106204.html>

用語一覧

DSP : Digital Signal Processor
 FFT : Fast Fourier Transform (高速フーリエ変換)
 FIR : Finite Impulse Response (有限長インパルス応答)
 FOMA : Freedom Of Mobile multimedia Access
 GPS : Global Positioning System (全球測位システム)
 HRTF : Head Related Transfer Function (頭部伝達関数)
 IID : Interaural Intensity Difference (振幅差)
 IMT-2000 : International Mobile Telecommunications - 2000 (第3世代移動通信)
 ITD : Interaural Time Difference (時間差)
 LAN : Local Area Network
 PDA : Personal Digital Assistant (携帯情報端末)
 RMS : Root Mean Square