

仮想音響環境のためのバイノーラル技術

小泉 宣夫 青木 茂明

NTTヒューマンインタフェース研究所

仮想環境を実現するための聴覚系技術として、空間音響の生成に関する研究の概要を述べる。仮想音響環境のための音響出力系の構成として、両耳受聴状態を模擬するバイノーラル技術に着目し、頭部伝達関数を用いた3次元音像定位の実現法と課題を述べる。次にスピーカ系への拡張としてのトランスオーラル系について触れる。また、通信系を介した共有環境の構成に関する考え方を述べ、そのためのグループウェアとしての音響環境の操作系の検討例を紹介する。

Binaural Technology for Virtual Acoustic Environments

Nobuo Koizumi and Shigeaki Aoki

NTT Human Interface Laboratories

1-2356 Take, Yokosuka-shi, Kanagawa 238-03, Japan

A study on spatial sound reproduction for auditory displays used in virtual environments is described. Binaural technology is reviewed for simulating binaural hearing, which is characterized by head related transfer functions. As an extension to loudspeaker hearing, transaural system configuration is described. Also, a concept of shared environment in audio telecommunication is shown, and a graphical user interface for manipulating acoustic environment is proposed.

1. はじめに

コンピュータと操作者とのインタラクティブな世界としてのバーチャル・リアリティ、遠隔でのオブジェクトを操作するためのテレエクジスタンスや、通信系を介して人間同志の共有環境をつくる臨場感通信を実現するための中心的な技術課題は、仮想環境の生成である。仮想環境生成は複数の表示メディアで構成される出力系と知覚との複合化・統合化技術である。

本稿では、仮想環境を実現するための聴覚系技術として、空間音響の生成に関する技術の概要を述べる。仮想音響環境のための音響出力系の構成として、両耳受聴状態を模擬するバイノーラル技術に着目し、頭部伝達関数を用いた3次元音像定位の実現法と課題を述べる。次にスピーカ系への拡張としてのトランスオーラル系について触れる。

また、通信系を介した共有環境の構成に関する考え方を述べ、そのためのグループウェアとしての音響環境の操作系の検討例を紹介する。

2. バイノーラル系の構成

仮想環境における聴覚系の課題の一つは、被験者に対し空間的な音場を知覚させることである。例えば、立体視の画像ディスプレイを中心としたシステムにおいては、音響系は画像ディスプレイ内のオブジェクトの位置同定の手掛かりとして用いられる。この場合、聴覚系に必要なのはまず音像定位知覚である。

空間音響は音源と受聴者の位置関係と、音場特性によって規定できる。仮想環境として空間の壁面の構成や吸音特性などをモデル化し、その中に音源の位置と被験者である受聴者の位置・方向を規定することが、ヘッドホン等を用いたバイノーラル(両耳受聴)再生のシミュレーション系を構成するための基本条件になる。ここで、音場の特性を反射壁面のない自由音場と仮定すると、音源と被験者との位置関係だけで仮想音響環境が規定される。音源が被験者から比較的離れたところにあると仮定すれば、音波の到来方向からの頭部、耳介の回折効果を表す頭部伝達特性を角度毎に予め求めておき、それを元の音源信号に付加することで再生信号が生成できる。この原理を図1に示す。各耳に至る左右一組の方向別頭部伝達特性により音像定位要因を表現する考え方は聴覚心理の分野で多くの研究がなされている。

ヘッドホンを用いたバイノーラル系における受

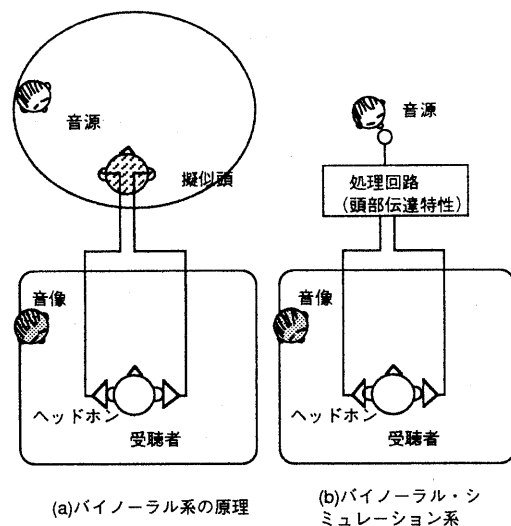


図1 バイノーラル系の原理

聴信号の生成には特定方向の頭部伝達関数を FIR (有限長インパルス応答) フィルタの係数として与え、音源信号に重畳する方法がとられる。頭部伝達関数の導入は3次元的な音像定位を実現するためのものであり、音像は当然頭外に定位しなければならない。この効果があまらなければ、通常のインテンシティ・ステレオの導入で十分ということになる。音像の空間的知覚の精度については、頭部伝達特性の個人差、測定精度、および動的追従性などの要因を考える必要がある。

頭部伝達特性は時間域のデータとしてフィルタ係数に取り込まれるがその特性は周波数域で見たほうが特徴をつかみやすい。図2は擬似頭を用いて測定した頭部伝達関数の周波数特性を3次元表示したものである¹⁾。一方の耳に関する伝達関

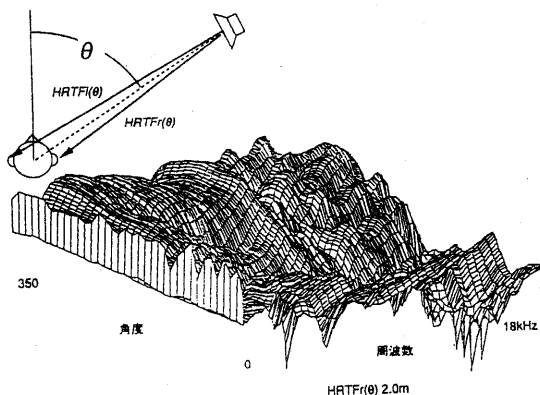


図2 擬似頭による頭部伝達関数の周波数特性

数の18kHzまでの周波数特性を正面方向を0度として10度毎に一回転させたときの値を表示したものであり、この山谷が定位情報の知覚と深く関わっている。頭部や耳介の形状に関しては個人によって異なり、他人の形状に基づいた頭部伝達特性をそのまま用いても音像がうまく定位しない場合がある。図3は4人の被験者と擬似頭についての正面前方(0度)と背面後方(180度)の伝達関数の差を表示したものであるが、それぞれに特徴が異なり、これが前方定位を難しくしている要因の一つとなっている²⁾。

個人差に依存しない汎用的な特性は、頭部伝達関数のデータから音像定位要因となる特徴を抽出することによって得られる。音像定位用の付加装置として製品化されているものの多くは、こうした特性を実装しており、フィルタの経済性や操舵性を考慮し、フィルタ次数をなるべく削減する努力がされている。一方で、最近ではハードウェアの進化により係数の置換も容易にできるようになったので、フィルタ係数に個人の測定データを用いたり、複数の標本データを選択するような方法で精度を確保することが可能になってきている³⁾。再生系において、鼓膜に至る伝達特性を正確に模

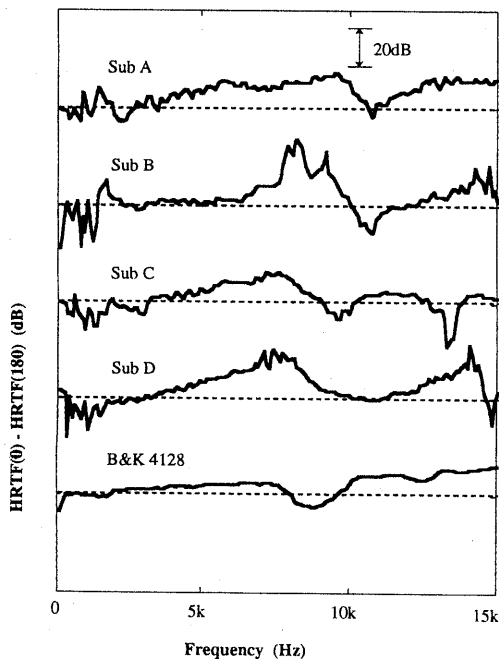


図3 頭部伝達関数の測定例

擬するためには、模擬系の伝達特性の基準点を決め、装着状態を規定する必要がある。また変換器の特性を補償する必要がある。厳密な再生系では必ず変換器の種類を規定しており、また適当な補償フィルタを具備している。

また、動的追従性は頭部装着型の仮想環境系にとって重要な問題である。インタラクティブなシステムでは、被験者の動き、特に回転に対して画像、聴覚情報とも追従しなければならない。人間の方向知覚は定常的な伝達特性だけでなく、頭部のわずかな動きによる特性の変化にも依存している。したがって、追従に時間遅れがあると正常な知覚の妨害要因となる。追従系には磁気センサ等によって頭部の回転角を検出しフィルタ係数を逐次更新する方法が使われている。方向別の頭部伝達関数をフィルタバンクに係数としてロードし、逐次方向情報に基づいて選択する方法が一般的であるが、伝達関数の精度よりも移動感を重視するのであれば、掃選型フィルタ係数をパラメトリックに操舵する方法が考えられる。

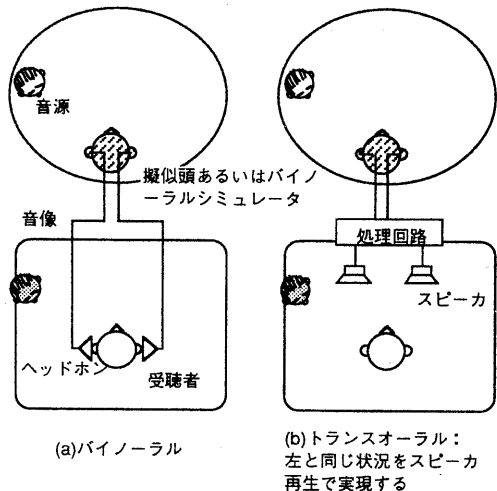


図4 トランスオーラル系の概念

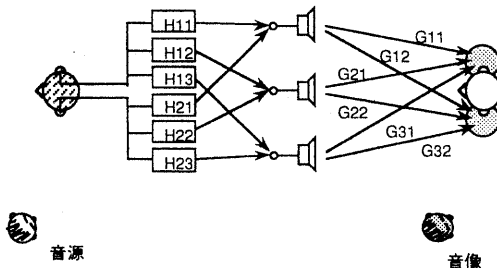
3. トランスオーラル系

装着型でない設置型の立体画像ディスプレイとの組み合わせでは、スピーカ再生による方式が必要となる。この場合、バイノーラル再生信号に対し、スピーカと受聴者との間のクロストーク補償をするトランスオーラル系(Transaural System)が適用できる。図4に示すようにバイノーラル系との違

いは、耳元に直接ヘッドホンの信号が伝わるのではなく、間接的に目的の信号が得られるよう加工された信号を空間的に離れたスピーカで再生するのである。これは音場の2点を制御する局所制御問題としてとらえることができる。

トランスオーラル系の基本的な方式は、バイノーラル系の2チャンネルソースを2つのスピーカで再生し、その際に生じる左スピーカから右耳へ、右スピーカから左耳へのクロストークを除去することであるが、この系は部屋の残響が十分小さく、音場の最小位相を仮定して実現される。任意の再生音場において付加残響の影響を取り除いて安定した音場を再現する方法として、3個のスピーカを用いて、2つの耳元の音場を制御する方式が検討されている⁴⁾。図5にこの方式の構成を示す。ここでは、実測される6個の伝達関数 G_{11}, \dots, G_{32} をもとにクロストークをキャンセルするようなフィルタ H_{11}, \dots, H_{23} を逆フィルタ定理によって決定する。

トランスオーラル系の実用上の課題は頭部の動きをどう検出して補正するかである。動的追従が無理であれば、耳元の2点だけではなく、その近傍の多くの点を制御点に選び頭部近傍の音場を制御することが考えられる。ただし、局所制御の多点化によって領域の制御を行なうのには限界がある。音波の波面合成などのグローバルな音場制御の考え方を今後発展させていく必要がある。



$$\begin{aligned} H_{11}G_{11} + H_{12}G_{21} + H_{13}G_{31} &= 1 \\ H_{21}G_{11} + H_{22}G_{21} + H_{23}G_{31} &= 0 \\ H_{21}G_{12} + H_{22}G_{22} + H_{23}G_{32} &= 1 \\ H_{11}G_{12} + H_{12}G_{22} + H_{13}G_{32} &= 0 \end{aligned}$$

図5 多チャンネル型トランスオーラル系の構成

4. 室内音響の影響

音源方向からの頭部伝達特性だけでは自然な音場の再現はできない。日常的な環境では反射波を

受聴しており、これが距離感だけでなく定位感に関しても手掛かりを提供している。反射波の影響を含めた音響信号の合成としては、音線法によるコンピュータ・シミュレーションがよく用いられる。また、空間全体の反射音の構造をグローバルに理解するためには虚音源分布で記述する方法が用いられる。これらは幾何音響的な手法であり、音の回折などの波動的な性質は含まれていない。初期反射音だけでは特徴付けられない比較的小規模の音場の記述には幾何音響だけではなく波動音響的なモデルの採用が必要となり、波動方程式を基本に、有限要素法等の数値解法の適用を考えていく必要がある。

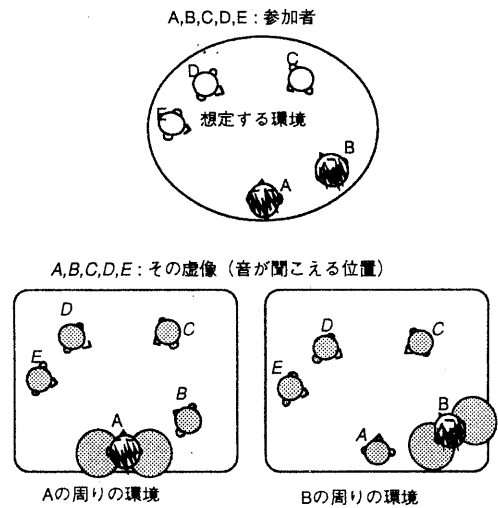


図6 共有音響環境の概念

5. 電話と共有音響環境

電話音声はその品質が限定されているにもかかわらず、通話する相手の性格なりを熟知さえしていれば、我々は自然な会話を行なうことができる。この場合は物理空間を超越した仮想的な共有空間が形成されるのである。

さて、より理想的な仮想共有空間とはどのようなものであろうか。対向通信で音響環境を共有するには、二者が実存するそれぞれの部屋をつなぎ合わせた仮想的な部屋の音響特性を再現する方法が考えられる。一方、多地点での臨場感通信には、一つの仮想的な音響空間を構成し、それを各地点で共有する考え方が有効である。したがって、図6に示すように遠隔地の人達のそれぞれの

両耳近傍で一つの音場の特性を共有できるような状態を再現すれば臨場感のある共有音響環境が実現できる。パーソナルな多地点マルチメディア環境においては、各自がヘッドホンを装着し、各音声近傍のマイクロホンによって暗騒音や室内の残響の影響を受けずに收音できるという状態を仮定すれば、今までに述べたバイノーラル系の適用によってこの仮想環境を再現することができる。

6. オーディオウィンドウ

画像生成にコンピュータ・グラフィックスが用いられるのと同様に、音響系の設計・操作もコンピュータのグラフィックス環境を用いて行なう方が拡張性や他の感覚系との統合性の点で有利である。また、ワークステーション間で共有環境を実現して遠隔での人達の共同作業を効率化する CSCW (Computer-Supported Cooperative Work) の検討が進んでおり、これに音響の設計・操作環境を組み入れることでパーソナルな多地点臨場感通信のプラットフォームが形成できる。ユーザはグラフィックディスプレイを用いて、仮想環境を自ら設計し、それを操作することができる。この通信系の構成を図7に示す。各地点では、マイクロホン、ヘッドホン、ディスプレイのあるワークステーションに加え、操作対象として、各地点の音声信号を頭部伝達特性や距離減衰等を模擬するデジタルフィルタを通して混合し、ヘッドホン受聴信号を生成する音響信号処理系を備えることで実現できる。

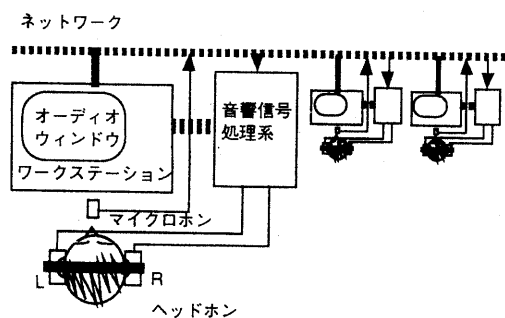


図7 仮想共有音響空間の系構成

さてこの音響信号処理系の制御卓として、2次元的な音像定位等を操作するために、仮想会議室の平面図をグラフィックディスプレイ上に描き、その中に参加者の状態（位置、方向、発声・受聴

の感度とその指向性）をアイコンとして設定するグラフィックユーザインタフェースを提案している⁵⁾。参加者は共通のウィンドウを見ながら、アイコンを操作し、相互の位置や状態を決定する。これをビデオウィンドウのアナロジーとしてオーディオ・ウィンドウと呼んでいる。

図8にグラフィックスクリーンの構成例を示す。これはちょうど図6に示した共有環境を参加者のアイコンの配置によって表現したものである。ヘッドホン受聴により知覚される相手の音像は平面図に示した位置に定位するように設定する。参加者は音像定位や音量が適切になるように自分のアイコンを設定することができる。また、相互の位置は会議中に自由に移動することができる。つまり、自己のアイコンを通信中に移動すれば、音響空間内の徘徊や耳打ちを仮想的に実現することができる。会議室の設定は任意であるので、特定の二人のアイコンを別室に移すことで、個別の会話を仮想的に設定することもできる。オーディオ・ウィンドウの操作権はグループウェアの考え方にに基づき、一人が有し、その権利を要求することによってユーザ間で持ち回ることができる仕組みになっている。オーディオ・ウィンドウの設定はネットワーク上のユーザ間で同一であり、操作権を持つユーザの操作はリアルタイムで他のユーザに伝えられる。

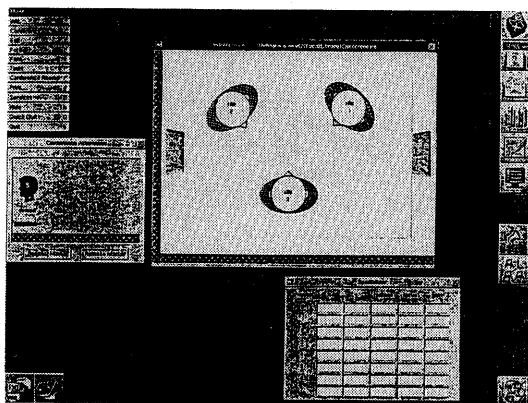


図8 オーディオウィンドウのスクリーン構成例

アイコンの位置関係は各地点で同一であっても対地参加者と自分の関係が入れ替わるので、必要な伝達関数は各地点で異なる。従って、信号処理は各側でローカルに処理され、環境の幾何学的情報だけがネットワークを介して共有される。

オーディオウィンドウで操作される音響信号処理系は図9のように構成されている⁵⁾。各地点の音声信号は左右の頭部伝達特性と距離減衰を模擬する2chのFIRフィルタに入力され、その出力を左右別に混合してヘッドホン受聴信号をつくる。フィルタの次数は1024次まで扱えるものとした。音像定位を実現する頭部伝達関数や距離減衰係数はアイコン操作で動的に変化する。二つのアイコンの位置関係は頭部伝達関数を制御し、アイコンの大きさはその音量を、アイコンの示す方向はその指向性を与える。これらの情報に基づき、FIRフィルタの係数は逐次更新される。データは即時に変更する必要があるので方向別の頭部伝達関数データをバッファメモリに蓄積するような構成を取っている。

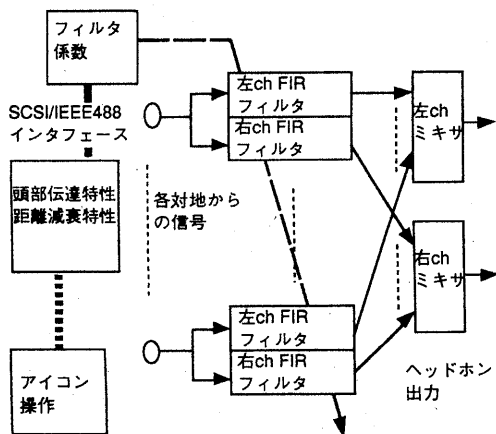


図9 音響信号処理系の構成

また、この操作系は多チャンネル信号の仮想的ミキサーとして応用できる。例えば、楽器別の多チャンネルソースに対し、それぞれに楽器あるいはセクションのアイコンを設定し、さらにリスナーのアイコンを置けば、コンボの楽器位置とリスナーとの相対位置を設定することができ、聞きたい楽器の近くに寄ってみるとかの疑似体験が通信系を通じて操作できる。多地点間での合奏や音楽

教習などへの適用が考えられる。現在、音響環境としては頭部伝達特性のみを考慮しているが、将来は音場シミュレーション手法を組み込み、室内音場の影響（初期反射、残響効果等）を考慮することで、汎用的な仮想音響環境の実現が期待できる。

7. あとがき

仮想環境生成のための音響技術として、バイノーラル技術に着目し、頭部伝達関数を用いた3次元音像定位の実現、通信系を介した共有音響環境の考え方ならびにその操作系の設計を中心に概要を述べた。仮想環境生成は複数の知覚の複合化・統合化技術であり、バイノーラル系を有効に機能させるためには、頭部の回転に対する追従性や移動する音源の表現などの動的音響環境の設計が重要な検討課題である。また聴覚研究の進展とともに、より有効な空間知覚の提示方法を模索していく必要がある。

文献

- 1) 杉山精, 久住雄介, 青木茂明, "擬似頭の回折係数," 信学技報EA93-3, pp.15 - pp.20 (1993)
- 2) 青木茂明, 小泉宣夫, 久住雄介, "音響伝達特性の測定機能を有する仮想会議環境," 信学会1992春期大会A-369(1-370) (1992)
- 3) 島田正治, 林伸夫, 林伸二, "頭外音像定位伝達関数のクラスタ法の一検討," 応用音響研究会 (1993.4).
- 4) M. Miyoshi, N. Koizumi, "New Tranaural System for Teleconferencing Service, Proc. Int'l Symp. on Active Control of Sound and Vibration," p217, ASJ (1991).
- 5) M. Cohen, N. Koizumi, "Exocentric Control of Audio Imaging in Binaural Telecommunication," IEICE Trans. E75-A, 164-170 (1992).
- 6) S. Aoki, N. Koizumi, and M. Cohen, "Design and Control of Shared Conferencing Environments for Audio Telecommunication using Individually Measured HRTFs," Presence 3.1 (1994) (予定).