

4. 音を操るためのアクティブ コントロール技術

— 2つのスピーカを用いた3D音場再生 —

東京電機大学 浜田 晴夫

☆ はじめに

アクティブコントロールの技術は、オーディオ分野でも魅力ある可能性を秘めている。ただし、オーディオ分野への応用の場合は、広帯域信号を扱う必要があり、また、所望の受聴状態も目的によって設定する必要があるなど、アクティブノイズコントロールの場合とは異なる点があり、これらを十分留意したシステム設計が必要となる。本稿では、実用的な仮想音源再生方式として最近話題となっている、2つのラウドスピーカを用いて、受聴者の両耳の2点を制御点とした再生方式を中心に解説する。

さて、仮想空間、あるいは既存の空間に存在する多くの音源が創り出す複雑な音場を忠実に再現することは、特にプロオーディオや音楽プロダクションでは長年の夢である。一方、最近のコンピュータ技術の飛躍的発達、バーチャルリアリティに代表されるような視覚、聴覚などの多くの感覚機能 (multisensorial) に対する手がかり (cue: キュー) の再生を通して、人工的な情景 (scene) を創り出すシステムの発達を加速している。このような先進的なシステムがまもなく迎える21世紀においてはどのような形で我々の前に登場するのか大変楽しみである。3D音響再生技術もそのなかにあって大きな変貌を遂げようとしている。たとえば、Auditory Displayと称される分野では、音とさまざまなメディアとの協調効果が見直されつつあり、また、技術的にもそれに応えようといろいろな試みが進行中である。本稿では、最近、再び注目されている2チャンネルスピーカ再生の技術を概観し、原理的な説明やコマーシャル・ベースでの開発現状にも触れて技術動向を議論することを試みる。

まず、3D音響技術 (3D audio technology) を用語から概観してみよう。この用語と類似な意味で用いられている用語は多く、たとえば、3D Auditory Display, Acoustic space synthesis, Dummy-head synthesis, Spatial sound reproduction また、ヘッドホン受聴ではなくラウドスピーカを用いた再生方式である違いを強調

して、Transaural reproductionとか、あるいは、応用分野を従えてBinaural telecommunicationなどのような形で登場する。バイノーラル、モノーラルなどの原義に立ち戻って考えるといささか奇異に感じるものもあるが、心理音響的な効果を強調したり、あるいは処理技術・機器としての側面を強調したりする意図から、いろいろな用語が現実にも用いられている。また、応用分野の性質に依存することもある。主にAuditory Displayの分野で最近使われている、Sonification, Audification, Auralizationなどは、データのaudible displayであることは皆同じであるが、それがコントロールされた音であるか、あるいは単にdirect playbackであるかとかの違いや、またvisualizationとの対応を強調したいなどの意図で使い分けているのが現状である。その他枚挙に暇がないが、virtual audio, virtual acousitic displayなどの用語も現実によく耳にする。このように3D音響に関する用語も、それぞれ目的とすることや、意図も多少異なっているが、いずれにしても、再生音場のSpatial fidelityの向上を目指しており、その効果や目的によって多様な展開が開始されていることは確かである。あえて応用分野を分類するとおそらく次の3つの分野に大別して考えることもできよう。(1) オーディオ分野での応用、(2) マルチメディアプロダクション、(3) Interactive Human-Computer Interface。

具体的な分野としては、たとえばプロオーディオ、コンシューマオーディオ、ゲーム、コンピュータ音楽などのエンターテインメントとしての応用から、テレカンファレンス (teleconferencing)、シミュレーション、バーチャルリアリティ、telerobotics、データ表現の高度なマンマシンインタフェース、視覚障害者の支援などの分野までとても広範である。特に最近では、上述のようなオペレータの機能向上、支援、あるいは臨場感通信のような人間と人間とのコミュニケーションの道具としての応用を指向している研究が多い。

さて、音による空間的な手がかり (spatial cues) は聴感的な観点からみると、いろいろな議論はあるが、次の2つのカテゴリーが重要である。1つは音の定位

Overviews of Spatial Hearing

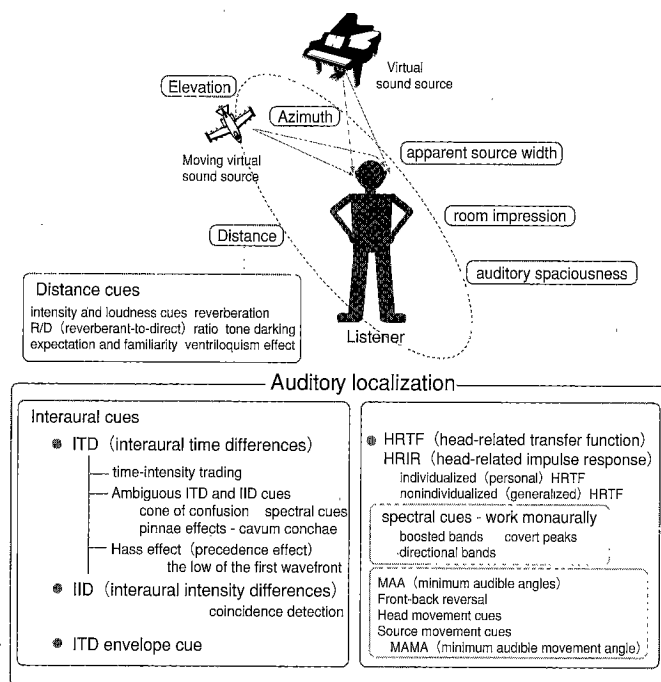


図-1 Spatial Hearingに関するキーワード

(auditory localization)に関する知覚で、2つ目は間接音や残響などによるroom effectの知覚である。ポイントは、このような音による空間的情報に対する手がかりの提供が魅力的な効果を提供するためにさまざまな応用が考えられているといえよう。たとえば、仮想空間における認知の分かりやすさ (legibility) や自然さの向上、空間事象の判別・分離能力の向上、あるいは、音の空間属性に対する操作向上などがあげられよう。このような背景を考えていくと、音による空間情報の認知に関する一貫性や、他のモード情報、たとえば視覚情報との相互作用などが重要であることは明確であろう。

ところで、これらの技術では、頭部 (回折) 伝達関数 (head-related transfer function (以下、HRTFと略す)) と関連する信号処理技術が重要となってくる。したがって、関連する分野としては、両耳受聴などの聴覚現象、空間知覚に関する研究、デジタル信号処理技術などが基礎になっているといえよう。この分野における優れた解説は多い。たとえば、ダミーヘッドマイクロホンや両耳受聴研究のステレオ音響技術への応用については古くから優れた解説が多いが、最近では、Auditory Displayという概念から、Human-Computer Interface, Virtual Realityなどでの分野からのアプローチも多い。2チャンネル2スピーカによる3D音場再生に関しては、原理、各社再生方式、Stereo Dipole方式、専用LSIなどがジャーナル¹⁾で特集されており、よい参考となろう。

☆ 2つのスピーカを用いた3D音響システム

図-1にいわゆるSpatial Hearingと称される分野でキーとなることがらをまとめてみた。

受聴者が知覚する、方位角 (Azimuth)、仰角 (Elevation)、距離 (Distance) などの他に、みかけの音源の幅 (apparent source width)、音に包まれた感じ (listener envelopment) などのroom impressionがどのようなcueでもたらされるかを明らかにすることが重要である。多くの研究成果があることは先に述べた通りであるが、特に重要なcueとして、ITD (interaural time difference)、IPD (interaural phase difference) とIID (interaural intensity difference) は古くから音像定位のduplex theoryとして知られており、また、ITD envelope cueの寄与についても多くの文献で述べられている。また、初期の研究では、定位の検知限 (MAA: minimum audible angle) の研究に代表されるようないわゆる、両耳現象の感度 (jnds: just noticeable differencesなど) の研究や、time-intensity tradingのような他のcueとの相互作用、また、頭部の動きの影響など総合的なパフォーマンスに影響を与える他の要素の研究などに大別して捉えることができよう。そのなかにおいて、HRTF (head-related transfer function) はこれまでの述べたcueとなる物理的な要因の多くを説明できるため、特性解析や、テスト合成に利用されてきた。現在、2つのスピーカを用いた3D音響システム (トランスオーラルシステムなど) もその基礎は、HRTFにあると言っても過言ではない。HRTFの概要、特性などについては後述する。一方、オーディオの分野では、HRTFを積極的に活用するようになったのは、1970年代後半から1980年になってからで、それまでのいわゆるステレオ音場理論に基づいて実用システムが作られた。

歴史的にステレオ音像を制御するための多くの理論があり、また、実用に大きく貢献してきた経緯がある。図-2はこれらの歴史的な経緯を、背景となった技術、実現方式、応用の各カテゴリー別に簡単にまとめたものである。それぞれのカテゴリーは、相互に深く関係しながら発展していることが分かる。特に、時間軸が近年になって急激に短くなっているが、これは、先に述べたHRTFや仮想音源技術の知見をデジタル信号処理技術によって実時間で簡単に実現できるようになったことが大きな原因の1つであると考えられる。特に、本稿で議論する2つのスピーカを用いた3D音響システムについても状況は同じであり、特に、コンシューマオーディオで用いられている3D音響プロセッサについてはこの歴史的な技術推移をよく反映している。すなわち、当初、Intensity stereo技術や既存のstereo imaging技術を基礎として簡易的な3D効果を実現していたものが、最近で

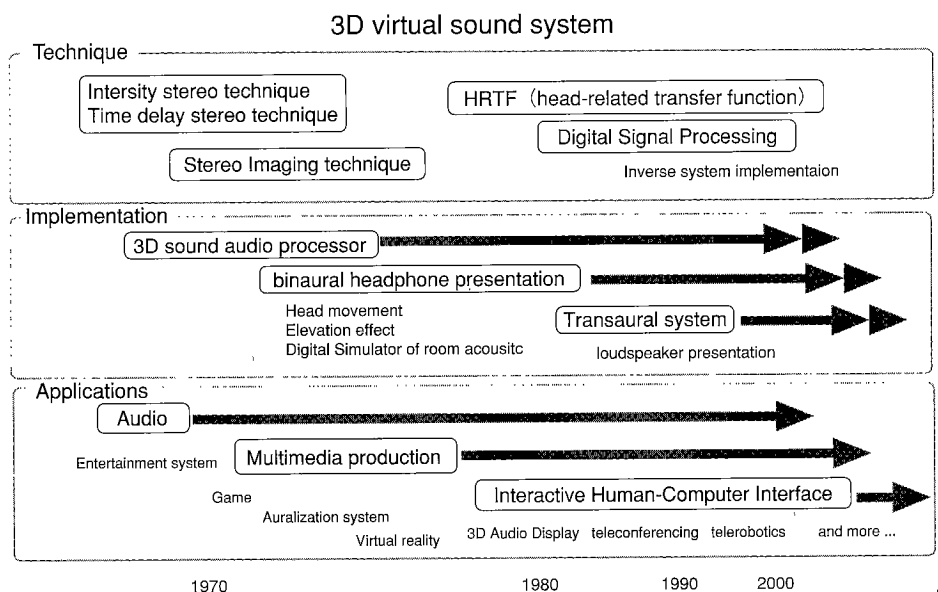


図-2 3D音響システムの推移（基本技術、実現方式、応用分野）

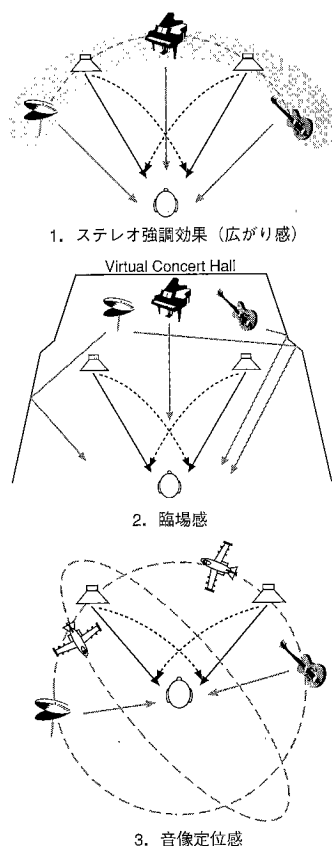


図-3 3Dオーディオプロセッサの機能

は、HRTFや仮想音源技術の研究成果をうけて、市場のニーズを反映したいくつかの3D効果を持つ機能チップとして実現される傾向にある。また、信号処理もアナログから徐々にデジタル処理へと変貌を遂げようとしている。一方、本格的に3D仮想音響空間を創生するためのシステムは、ヘッドホン再生を主体としたいわゆるバイノーラルシステムとして実用になったが、最近、トランスオーラルシステムとしてラウドスピーカを用いたシステムも提案されている

☆ オーディオプロセッサ

実用化技術^{1), 2)}

先に、音源の位置を知覚する際、IADとITD (or IPD) が重要であると述べた。周波数が600Hz～1kHzまでは、頭部の物理的な寸法（直径15～20cm程度）が波長に比べて小さいことから、音波の回折によって両耳での音のレベルはほぼ等しい。そのためにITD (or IPD) が水平面での音像定位に重要である。700Hz～2kHzではちょうど過渡的な帯域でIADとITD (or IPD) の両者が重要となる。また、さらに高い周波数では頭部の遮蔽（シャドウ）効果のためにIADが重要な役割を演ずるが、ITD envelope cueと称する振幅変調波としての時間差情報（特にonsetでの時間差）も人間は検知する。また、耳介の複雑な形状も高い周波数では、directional bandなどの研究のように音像定位に貢献し、特にモノラルでもcueを与える事象もあることが知られている。

サラウンド効果という用語はいろいろな目的に多用されてきたので注意を要するが、特に民生機器でサラウンドエフェクタと称する多くは先に議論した古典的なステレオ音像制御技術をベースにしたものが多かった。そのため、音の広がりが増すもののセンタのボーカル音像がぼけたり、個々の楽器の定位がはっきりしないなどの問題もあった。そこで、この点に改良を施しステレオ強調効果（stereo enhancement）などと称されて商品化されている。特に、オーディオ機器やテレビ受像器にはじまり、パーソナルコンピュータに搭載しているものもある。これらの処理回路では、先に述べた音像制御技術と考え方は基本的に同じであるが、左右オーディオ信号の周波数特性（振幅、位相）を丁寧に制御してより自然な音場になるような努力が行われている。この基礎となっているのはHRTFに関する研究成果（周波数特性、聴感特性）である。このステレオ強調効果の他に、音場感や音像定

位までコントロールしようとするのが現在の3Dオーディオプロセッサの現状であろう(図-3)。

これらも主流は2つのスピーカを用いた方式であり、なかにはヘッドホン受聴用の回路を含むものもある。この背景には、パーソナルコンピュータの普及があげられよう。家庭用パーソナルコンピュータでは一般にスピーカを離しておくのが困難(ユーザとディスプレイの距離が短いこともスピーカ間隔を大きくできない原因の1つ)であり、また、わざわざ3台以上のスピーカを増設して音場再生システムを構成するユーザも少ないことなどが原因であると思われる。

アナログ処理からデジタル処理へ

現在、ステレオ強調効果のみの単一機能の場合、ほとんどはアナログ回路で実現されている。これは、パーソナルコンピュータやオーディオ機器など受聴者側で機能を実現するため、他の機能、すなわち音場感制御、音像定位などに比べて回路規模が小さくアナログ回路で実現しコスト低減を図っている。一方、音場感の制御では残響付加回路、また、音像定位では、さらにデリケートな特性制御(振幅特性と位相特性)が必要であるのでデジタル回路へ実現する傾向にある。アナログ処理に比べて時間軸の制御の容易さや、多機能の同時実現などの点でデジタル信号処理向きといえよう。

一方、パーソナルコンピュータで多チャンネル再生を行う動きもある。たとえば、Microsoft社のWindow95用のAPIである「ActiveMovie」はMPEG1、MPEG2の動画を再生するためのものであるが、Dolby社は、AC-3などの音響技術を提携するなど活発である。これによれば、5.1オーディオ信号を伝送レート320kbit/sで圧縮符号化しても信号品質が保てるとしている。この場合でも、先に述べた理由で2つのスピーカで再生することが望まれる場合が多いため、たとえば、Dolby社はVirtual Dolbyと称する規格のカテゴリーを発表した。具体的には、Dolby surround PRO LOGICもしくはDolby Digital信号をデコードし、リアーの仮想音源をVirtualizerと称する認定の手法で作る。さらにこのリアー成分をフロント成分の2チャンネルと加算することで最終的な2チャンネル信号にするという手法である。

日本ビクターの3D-PHONICでは、Dolby surround PRO LOGICを2つのスピーカで再生する独自の方法で、DSP(TMS57052; TI社)を用いて実現されている。また、AD/DA変換用LSIにこれらの3D Sound効果を組み込もうとする試みもすでに行われている。表-1は上述したVirtual Dolbyで用いるVirtualizerの実現例である。ここで紹介したものだけでなく、各社で3Dオーディオプロセッサの開発実用化が行われている^{1), 2)}。

オーディオプロセッサの場合、市場ニーズにあった効果をいかにコストを抑えて実現できるかが問題である。そ

方 式	会 社 名
VMAX	Harman
3D-PHONIC	日本ビクター
Virtual Sonic	松下電器産業
Q Surround	Q Sound Lab
Virtual Sound Stage	ソニー
Specializer N22	Desper Products
TruSurround	SRS Labs
Theater-Dimensional	オンキヨー

表-1 Virtual Dolbyプロセッサの例

のため、所望の効果に対する設計思想で研究開発が行われた。その方針はアナログ回路の設計思想の継承であり、最近までは成功してきた。しかしながら、デジタル回路による実現では、複数の機能をアルゴリズムの変更(ソフトやデータのスイッチング)で実現できるため、3D Audio Processorとしてのプラットフォームをいかに実現するかが今後の課題となろう。また、DSPの機能向上に伴って、局所的な周波数特性のフィルタリングではなく、HRTFをベースとした本来のトランスオーラルシステムの持つ能力を最大限に生かせるような設計へと変貌していくように思われる。次に、このようなトランスオーラルシステムの基礎ともなっているHRTFについて議論する。

☆ 仮想音源の再生

仮想音源再生における音波の伝搬の様子を理論的に解析するために、解析モデル(剛球モデル)を導入する。本稿では、理論的に得られた伝達特性を用いて時間周波数特性と併せ剛球近傍での瞬時音圧を可視化した例を示す。評価領域は、原点を中心としての正方形を考え、この領域を間隔ずつの要素に区切り、おのおのの点(の要素)で再生音場における伝達特性を算出した。音源は点音源であり、サンプリング周波数は10[kHz]とし、また、算出した伝達特性のタップ長は128ポイントとした。なお、瞬時音圧分布のスナップショットは、時間の刻みを0.1[ms]でプロットしてある。座標原点を中心として描かれた白塗りの円は受聴者の頭部をモデルとした剛球(直径0.16[m])を表している。

図-4(a)は正面から右15度方向に位置する実音源の場合で、後述する仮想音源シュミレーションと比較する目的で示してある。図-4(b)は標準配置(スピーカ見開き角60度)トランスオーラルシステムによる仮想音源(右15度方向の音源)再生の場合である。インテン

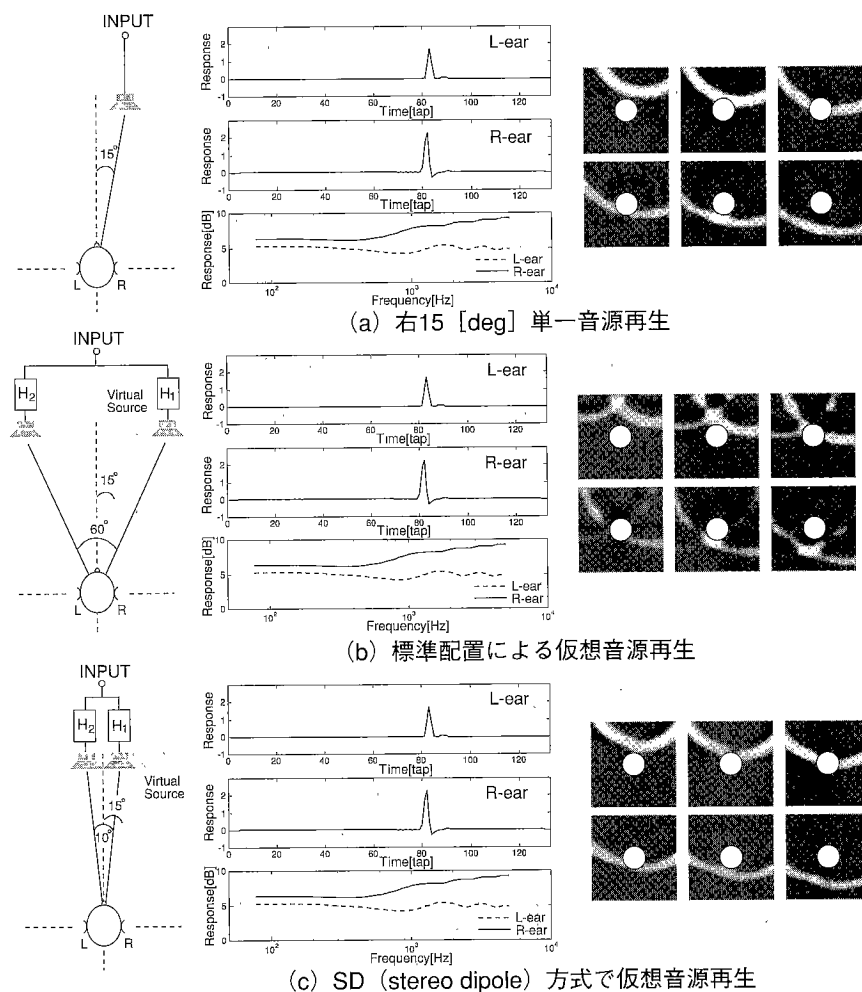


図-4 仮想音源の再生

シティコントロールの場合に比べ、両耳2点でのシミュレーション精度は格段に向上して、実音源を正確に近似しているといえよう。ただし、両耳2点を離れた剛球近傍での波面は複雑であることが分かる。このことがトランスオーラルシステムが従来サービスエリアが大変小さいといわれてきた要因でもあり、また、頭部移動に伴う逆相感などのを引き起こす原因と考えられる。一方、図-4 (c) は、筆者らの提案している Stereo Dipole (SD) 配置^{1), 3)} (スピーカ見開き角10度) による仮想源再生を示している。両耳での振幅特性の近似だけでなく、近傍音場での特性が標準配置に比べ単純でまた、原音場 (同図 (a)) をよく近似していることが了解されよう。

☆ 今後の展望とまとめ

2つのスピーカを用いた3D音響システムは現在、3Dオーディオプロセッサとして実用に供しているものと、いわゆるトランスオーラルシステムとして主に精度を求める研究用途で用いられているものがある。両者とも確実に、応用範囲が広がっており、オーディオ分野のみならず、マルチメディアプロダクション、Interactive

Human-Computer Interfaceと、最近ではコンピュータの発展動向を抜きには語れないところにきている。むしろ、コンピュータディスプレイなどへの応用は、2チャンネルスピーカ再生の利点を生かせる優れた応用分野であったといえよう。また、デジタル信号処理が身近になるに従って、両者の距離もなくなってきた。すなわち、DSPプロセッサをプラットフォームとしてさまざまな用途に合致した機能が複合して実現される方向で進展している。特に、Interactive Human-Computer Interfaceの分野では、3D Audioty Displayなどを有効に活用できる分野が多い。3D Audioty Displayと他のモード (視覚、触覚など) との強調効果は魅力的である。トランスオーラルシステムなどの3D音響技術は、もはや単にエフェクタとしての音響システムではなく、3D音場再生のためのエンジンやプラットフォームとして変貌を遂げようとしているといえるであろう。

参考文献

- 1) 浜田晴夫他: 2チャンネル2スピーカによる3D音場再生, JAS Journal (Apr. 1998).
- 2) 音場を広げる音響効果IC, NIKKEI ELECTRONICS, No.661, pp.127-138 (1996).
- 3) 浜田晴夫他: 新しい音場再生方式 (Stereo Dipole) に関する研究, 電子情報通信学会技術報告 (Apr. 1996).

(平成10年11月27日受付)