

3D音響ブラウザの実装と駅発車音への適用

増井 誠生 棚橋 純一

(株) 富士通研究所 パーソナル&サービス研究所

E-mail: masui@flab.fujitsu.co.jp

画像や音響を含むマルチメディアデータを3D空間に配置し、ブラウジングしながら必要なデータを探索する手法を開発してきた。昨年度、音響データの配置に関して、SD法による評価結果から、因子分析結果を用いて3D配置データを生成する方法、および、自己組織化マップ(SOM)を用いて平面上に均等に配置し、階層構造を用いて探索可能とする方法についての報告を行った。

今回は、音響を対象としたブラウザ機能の実装上の問題点や、操作インタフェースの設計指針や工夫について報告する。前回の発表で利用した評価データは、個人による1回評価実験のものであり、実験データの妥当性が問題とされた。このため、今回は、駅の発車音楽に対する評価データを利用することで、具体的な実験データの分析結果の表示にこの3D音響ブラウザを用いれば、どういった発表効果が期待できるか、発表用ツールとしてどんな機能が必要か、などの考察を試みた。

Implementation of 3D Audio Browser and its Application to Station Music

Motoo MASUI, Junichi Tanahashi

Fujitsu Laboratories Ltd., Personal & Service Labs.

We have produced 3D-positioning method for multimedia data including pictures and sounds, and browsing navigation method to seek required data. Last year, we presented a method to create 3D positioning data from SD-methods and factor analysis, and to make some layer-planes using SOM (Self-Organized Map).

In this paper, we report many problems of implementing audio parts of the 3D browser, and user interface guideline for operating audio objects. Because of lack of reliability in our psychological evaluation data, we now use more reliable experimentation data on station music. And we will discuss the merit of 3D audio browser as a presentation tool.

1. 3D 情報ブラウザ NW

3D 情報ブラウザ NW は、当初は、図書館の収蔵物における、書籍、巻物、古文書、ビデオ、音響素材などのさまざまな属性を持つデータの検索用インタフェースとして考案された。3D 空間での視点移動操作(3D 探索)によって目的とする情報を得るためには、どのような空間構成が好ましいか、あるいは、キーワード検索による絞込み結果を、PC 上の限られた画面にいかにか合理的に配置するかといったことを検討の上で、実装を進めた。NW は、静止画、音響、文字列、書籍、などを対象とし、各データは種別に応じた異なる属性を持つ。ブラウザ上で利用者が行う操作としては、視点移動操作、内容確認操作、大量データに対する絞込み操作、実視界を示す画面と視点位置を示す補助画面の切り替え操作、視点の移動モードの切り替え操作などがある。

1.1. 視点移動操作

視点移動操作は、画面内でのマウスの移動によって行う。3D 空間内を、一般的なマウスで操作するために、視点や視線の移動・回転操作の切り替えはメニューバーから選択する。通常、一般的な VRML ブラウザでは、マウス操作を左右と前後に割り当て、キーボード操作やモード切り替えでジャンプなどの上下移動を行うものが多い。これはウオークスルーに適した操作系である。

3D 情報ブラウザ NW では、空間の構成方法として、複数の平面を前後に重ねることで、階層性をもたせる配置を採用した。最も奥側の平面では、自己組織化マップを利用した均等な 2D 配置が行われ、手前の画面には、奥の画面を代表するようなオブジェクトが配置される。空間的な樹状配置であり、検索時には、手前側から段階的に目的のデータを得ることができるという操作モデルである。

このため、NW では、一般的な VRML ブラウザの操作系と異なり、奥行き移動操作と各階層平面内での移動操作とのモード切り替えを行う。デフォルトの操作モードでは、通常のマウスの上下位置を、空間内の奥行き移動速度に割り当てる(表示ウィンドウの上側をドラッグすると、視点が画面の奥側へ移動し、下側をドラッグすると、画面の手前側へ視点が移動する)。シフトボタンと共にドラッグすると、階層平面に沿って視点が移動する。これはデータの探索に相当する。メニューからは、そのほか、視点を保持したままでの回転操作などを選択できる。配置画面を斜めから見る場合などに回転操作が必要となる。

一方、因子分析を利用した 3D 配置方法を採用すると、3次元の空間を構成する軸それぞれが、評価結果を説明する基準としての意味を持つ。この場合は、3軸方向の各々に対して自由に移動ができることが好ましく、おそらくはジャイロセンサを利用した 3D マウスなどの利用が適した配置であると考えられる。3D 音響再生に加えて、液晶シャッターを利用したバーチャル表示技術を組み合わせることで、より実体験的な表示を行うことにも適していると思われるが、特別な装置を使うと、一般的な表示ツールとしての利用は困難となる。このため、操作系と対象モデルは一致しないが、今回の報告では、階層的平面配置用の視点移動操作系をそのまま、因子分析結果から配置された空間での移動操作に流用した。

1.2. 内容確認操作

情報探索のための空間配置では、本来、その軸自体の意味や、空間構成の特徴を活かして利用者を誘導することも可能であり、表示されているオブジェクトの内容を総当り的に確認する必要はない。静止画像であれば、その概要は視覚的に把握できるし、音響であれば距離と連動した音響再生制御と連動させることで、必ずしも対象となるオブジェクトを選択しなくても、内容をある程度確認することはできる。

だが、利用者が選択する局面では、検索対象を詳しく確認したいと思うのは自然な欲求である。例えば、楽器音などであれば、その演奏者、演奏場所、楽器名、所有者といった付随する情報を確認することは必須であるが、これらの書誌的な情報については、オブジェクトを示す画像(サムネイル)に無闇に詰め込むことで、必要以上にユーザを混乱させる可能性もある。ユーザがじっくりと検索を行う場合などは、多くの情報が画像として詰め込まれていることが有益な場合もあるが、発表用のツールとして利用する場合は、情報量が多すぎることは聴衆にとって混乱を与えることになる。発表用としては、簡単な表示モデルを採用し、オブジェクト選択や機能選択によって、詳細の確

認ができるという設計がふさわしい。

NW で扱われるオブジェクトは、空間位置のように一般的に用いられる属性に加えて、その種別に応じた特殊な属性も持つ。例えば、テキストであれば、表示する文字のほか、表示フォント、表示サイズなどを属性として与えることで、ブラウザはそのテキストを表示できる。書籍データなどでは、画面の表紙データや目次などに加えて、補助的な書誌データが加わる。

1. 3. 画面切り替え操作

この 3D ブラウザではメイン画面では視界そのものを表示する。マウスを用いて視点移動操作を行うと、それに伴って視界が変化する。ここで、オブジェクトに近づきすぎると、画面がそのオブジェクトで占有されてしまうため、これを防ぐためにオブジェクトの最大表示サイズを決めておく。オブジェクトに近づき、物体を通り越す際、視点はオブジェクトと一致し、オブジェクトは最も大きく表示される。この瞬間、オブジェクトの再生音量も最大となる。奥方向に進むと視界から突然オブジェクトが消える(オブジェクトは視点の背後に存在する)。視覚と聴覚の連動を図り、視界に見える範囲の音だけを出すことも可能であるが、本ブラウザでは、視界を聴覚で補うという意味で、視点の背後にある音も含めて、その距離と定位に応じた音響ミキシング再生を行う。

1. 4. 検索絞り込み操作

自己組織化マップを用いた階層的平面配置の際に用いる。書籍データの検索において、登録されているデータを全部表示しながら探したいということはまれであり、事前に対象を絞り込んだあと、そのデータを空間配置したいというケースが考えられる。このため、キーワード検索インタフェースを用意し、対象とするデータの属性を事前に絞り込んだ上で、自己組織化マップによる配置データを生成し、階層的平面表示を行う。なお、今回の報告では、音響評価値の因子分析結果の表示に 3D ブラウザを利用することが目的で、検索絞り込みとの連動は行っていない。

2. 3D ブラウザの音響機能

音響データのブラウジング機能を実装する上での検討項目について述べる。

2. 1. 音響データの大量表示

音響データのように、視覚だけでは内容を十分に確認することが困難なメディアも存在する。これを扱うためには、個々の対象を選び、ひとつひとつ内容を確認することも可能であるが、大量のデータに対する個別選択手法は、操作ミスを誘う原因にもなり、利用者の理解モードを、画面の切り替え操作によってささげざる結果にもなる。これを解決する方法として、我々は、対話型 3D 音響 (3D Positional Audio) を採用して、なるべくユーザの思考を中断しない形で、視点移動に連動した音響ループ再生機能を実装した。この手法は、VRML 対応の 3D ブラウザでは一般的なものである。一般的なコンテンツでは、同時にそれほど多くの音が再生されることは少なく、視点位置に応じた音量や定位の制御を実装すればそれで用が足りることも多い。

一方、情報検索のように、究極的には数万件を超える多数の検索対象を扱う場合は、データを絞り込む機能とともに、限られたサウンドリソースを有効に利用するため、3D 音響バッファの切り替え機能が必須となる。音響を再生するためには、その開始・終了のタイミングを、視点移動と連動させるといった工夫も要求される。特に短時間の音響サンプルを対象としている場合は、開始・終了のタイミング管理に換えて、音響をループ再生することが有益になる。繰り返し再生されることで、視点の移動の効果(利用者にとっては音響移動に等しい)が、より強調されることになる。

ループ再生を行いながら、必要に応じてその再生バッファを切り替える場合は、無闇に再生が初期化すること避ける必要がある。具体的には、その状況で再生すべき音響を求め、現在発音中の音響と比べた上で、不要なものの再生を止め、バッファをクリアし、新しい音響を読みこんだ上でループ再生を開始する、といった一連の操作が必要となる。特に音響の切り替えには、視点と音響オブジェクトとの距離計算が必須となり、このため、視点の移動が生じた場合は、バッファ切り替えを頻繁に繰り返して、3D音響バッファを常にリフレッシュしておく必要が生じる。

2. 2. 対話型 3D 音響技術の利用

ブラックボックス的ではあるが、特に Windows 系 PC をプラットフォームとして考えると、Aureal 社の A3D、Creative Labs 社の EAX、Intel 社の RSX、Microsoft 社の DirectSound3D など、プログラマが 3D オーディオを利用する選択肢は比較的多く存在する。特に、NASA での空間シミュレータ開発から得られた C.R.E. 社の技術を引き継ぐ Aureal 社の A3D は、当初から、ゲームなどのアプリケーション開発者向けの空間対話制御を意識したものであった。このコンセプトは、Microsoft 社のマルチメディア開発用 SDK である DirectX の DirectSound3D API によって普及した。現在は、QSound 社の Q3D や SRS Labs 社の SRS(Sound Retrieval System) など音楽再生の臨場感向上を目的とした 3D 音響技術も、DirectSound3D から利用できるようになった。なお、Intel 社のソフトウェア 3D 音響技術である RSX は、現在は Milles 社の Milles Sound System に引き継がれている。これらの対話型 3D 音響技術の動向については、米国 MMA の下部組織である IA-SIG の 3D-WG から得られる。(www.iasig.org)

DirectSound や DirectSound3D は、Microsoft 社の DirectX SDK の一部である。DirectX の枠組では、ある API が呼ばれたとき、それを処理するハードウェアが存在すれば、処理はそのままハードウェア側で実行される。ハードウェアが存在しない場合は、ソフトウェアでのエミュレーションが行われる。後者は HEL(ハードウェアエミュレーション層)によって実現される。このため、ハードウェアの有無、あるいはその種類によらず、同じ API で目的とする処理を記述できる。これは HAL(ハードウェア抽象化層)によって実現される。HAL によってソフトウェアの汎用性が得られるかわりに、動作環境によってその実行速度や性能が異なる。また、ハードウェアによっては、アプリケーションプログラマが期待した性能を実現できない場合も生じる。

A3D や EAX では、その後のバージョンアップを経て、壁面による音響反射や障害物の設定も可能となった(A3D 2.0、EAX 2.0)。これによって、例えばより現実的な音場デザイン用途にも使える技術となっている。どちらも各々の API に対応したオーディオチップが必須である。例えば、A3D 2.0 に対応するボードとしては、Monster Sound MX300 が、また EAX 2.0 に対応するボードとしては、Sound Blaster Live! などが知られている。これらのボードは、DirectSound3D から利用することができるが、各々の機能をフルに生かした実装をする場合は、専用の 3D 音響 API を利用する必要がある。

2. 3. 3D 音響ボードの性能差

3D 情報ブラウザ NW の音響機能の実装は、汎用性を期待して DirectSound3D を採用した。ハードウェアとしては、SoundBlaster AWE64Gold、Monster Sound、Monster Sound II (MX300)、SoundBlaster Live! の合計4種類デスクトップ PC 用内蔵ボードを利用した。3D 音響バッファの管理方式は以下の通りである。

	3D バッファ	ミックス数	DSP
SB AWE64G	メモリ(任意)	DSP(25)	EMU8000
Monster Sound	DSP(7)	DSP(16)	Analog Device 2181
MS-II MX300	DSP(15)	DSP(96)	Aureal Vortex2
SB Live!	DSP(32)	DSP(32)	EMU10K1

まず 3D 音響バッファがハード的に用意されていない SB AWE64G の場合は、3D 音響バッファは全てがメモリ上に確保され、これは直接的に CPU への負荷につながる。AWE64G で上限ぎりぎりの 25 音分の 3D バッファを確保したところ、Pentium Pro クラスの CPU では操作にもたつきが発生した。

DSP 上に 3D 音響バッファが用意されている場合には、それを越す 3D 音響バッファがソフトウェア的に確保される。Monster Sound では、比較的 DSP の性能が低く、3D 音響バッファを7つ確保すると、DSP 負荷はほぼ限界となった。同時 3D 発音数を9とした場合、追加して確保される 3D 音響バッファの品質が悪く、ノイズとなることがわかった。この Monster Sound では、事実上最大7つしか同時 3D 発音ができないため、デフォルトの 3D 発音数を7として実装を進めた。

Monster Sound の後継にあたる MX300 では、従来3種類のチップに分散されていた機能が、小型の DSP 一個に集約されている。基本性能も大幅に向上し、最大同時発音数を SB AWE64G の上限である 25 に設定しても、

DSP は 10%程度の使用率であった。だが、視点位置に応じたバッファ切り替えを行うと、切り替えタイミングでスイッチングノイズが発生した。バッファの切り替え時に音量を下げるなどの方法では、このノイズは回避できなかった。

SB Live! では、最大同時発音数の上限は 32 音である。このボードを用いれば、同時発音数は十分であり、音質劣化も発生せず、またバッファ切り替え時には不快なノイズも感じられなかった。

2. 4. 視点距離に応じたバッファ管理

DirectSound では、音響のミキシングを実現するために、2段型のバッファ管理を採用している。まず、1次バッファは、オーディオボード上の出力ポートに相当するバッファである。これに対し、2次バッファは、音響を格納するために再生音の数だけ確保されるものである。

音響データを 3D ブラウジングする場合は、極論をいえば、音響の数だけ 3D 音響バッファ(2次バッファ)を確保し、各々の音響の位置を与えておくことで、視点の移動に対して、自動的に 3D 音響バッファの制御が行われる(音量や定位の細かな制御は API 経由で DSP に処理を渡すことになる)。

上述したように、現実には、個々のオーディオボードでは、3D 音響バッファ数には 7~32 程度と限界がある。このため、多量の音響データを扱う場合は、視点位置から近い距離にあるものを選び、常に音響バッファを入れ換え続ける必要が生じる。短い音の場合はループ再生が有効であり、視点の移動タイミングで 3D 音響バッファを入れ替えることになる。

3. 音響アイコンの自動生成

3D音響ループ再生を組み込むことで、音響をブラウジングの際に、個々の音の内容確認操作を極力減らすことができる。同様に、音響のアイコン画像を、音響の評価値、または物理的特徴量を用いて作成することで、アイコンの配色などにバリエーションを与え、検索性を向上させることが可能となる。

一例として、14 評価語対からなる評価結果に対して、以下のような手順でアイコンデータの生成を試みた。

(1) 音に対する代表評価語を選ぶ。

ある音に対する評価値の中で、もっとも中間値からはなれた値をとっているものを「代表評価語」とする。候補が複数存在するときは、順序をつけて「代表評価語」を決める。

(2) 代表評価語に対する配色パターンを決定する。

3つ組の配色パターンとそれからイメージされる形容詞句との対応は、カラーデザイン関連の仕事としてまとめられている。ただし、形容詞句から配色パターンが一意に決まるわけではない。異なる配色が同じ形容詞句をイメージさせることも多いが、そのひとつを対応づける。

(3) 配色パターンをアイコンの背景色にする。

代表評価語1つ分に対する配色パターンをアイコンの背景色として横並びに並べる。

例：かん高い(青+黄+赤紫) 迫力のある(えび赤+鬱金+深緑) やかましい(赤+黄+緑)

(4) 音の物理的特徴量に応じた描画を行う。

波形を抽象化してアイコン上に描画する。周波数分布の偏りに応じて描画位置を上下に移動する。

例：スペクトル分布：中域 振幅パターン：なめらかな減衰

4. 駅の発車音への適用

駅の発車音や車内音として使われている 30 曲(JR: 24 曲, 大阪地下鉄: 6 曲)に対する評価実験の結果の提供をいただき、この 3D 音響ブラウザでの配置を試みた。心理学系の合計 20 人が2度ずつ 16 評価項目に関して7段階評価(1-7)を行ったデータである。因子分析結果を下記に示す。

主成分分析では下位の主成分の効果が無視されるのに対し、因子分析では、設定した変数(ここでは3変数)で実験結果が説明できるよう収束計算を行う。さらにバリマックス回転を施すことで、ここでは第1因子から第3因子に、各々と関連する評価変数軸が近づいていく(回転によって第1因子に集中していた因子負荷が、第2因子や第3因子に均等に分散される)。因子負荷量を見ると、第1因子軸は「深みがある」、第2因子軸は「美しい」、第3因子軸は

「物足りない」を示していることがわかる。さらに因子得点を比べることで、個々のデータの特徴を読み取ることができ
る。NW では、対象となる音を 3D-Audio で確認できるため、因子軸と配置結果の整合性を、実際のデータで確認
することが可能となる。

3D音響ブラウザ NW においては、元々、データの空間移動を行わないという前提であり、空間配置データを読み
込めば、その構成を探索的に確認できる。NW を因子分析過程の可視化ツールとして利用する場合は、回転途中
の段階の配置データを作成し、空間的位置を再設定することが必要となる。これによって、バリマックス回転による
音響データの移動効果を、アニメーションで確認できるはずである。

変数	平均値	標準偏差	形容詞対 (1-7)	因子負荷 (バリマックス回転後)			
1:	3.74	0.8410	やかましいー静かな	1	0.44683	0.47473	0.74683
2:	4.03	0.8119	うるさいーうるさくない	2	0.39249	0.57121	0.71319
3:	3.74	0.6224	物足りないー迫力のある	3	-0.23345	-0.097487	-0.93561
4:	3.72	0.3671	大きいー小さい	4	0.35407	0.3557	0.74037
5:	3.81	0.8091	鋭いーにぶい	5	0.8206	0.10977	0.5432
6:	3.92	0.8418	かたいーやわらかい	6	0.75945	0.45699	0.42904
7:	4.11	1.0532	落ち着いたー甲高い	7	-0.74421	-0.35062	-0.54446
8:	3.70	0.9520	金属性の一深みのある	8	0.80376	0.34915	0.3825
9:	3.84	0.5311	快いー不快な	9	-0.21453	-0.90779	-0.28071
10:	3.71	0.5234	美しいーきたない	10	-0.088751	-0.93962	-0.18141
11:	3.55	0.5367	澄んだー濁った	11	0.55759	-0.6739	0.020902
12:	3.52	0.6983	はっきりしたーぼんやりした	12	0.91154	-0.083751	0.31231
13:	3.90	0.7547	安定したー不安定な	13	-0.28447	-0.79774	-0.47013
14:	3.79	1.1470	あわただしいーのんびりした	14	0.41727	0.37157	0.72487
15:	3.82	0.6798	まとまりのあるーまとまりのない	15	-0.12524	-0.86735	-0.37183
16:	3.93	0.6840	とけあったーばらばらな	16	-0.41366	-0.78047	-0.39345

主成分分析	固有値	寄与率
1	11.1066	69.4
2	3.0121	18.8
3	0.8756	5.5
4	0.2707	1.7
5	0.2284	1.4
6	0.1969	1.2
7	0.0934	0.6
8	0.0568	0.4
9	0.0483	0.3
10	0.0297	0.2
11	0.0260	0.2
12	0.0227	0.1
13	0.0108	0.1
14	0.0099	0.1
15	0.0075	0.0
16	0.0046	0.0

因子分析	固有値	寄与率
1	11.0406	74.9
2	2.9046	19.7
3	0.7944	5.4
4	0.1405	1.0
5	0.1209	0.8
6	0.0616	0.4
7	0.0271	0.2
8	0.0022	0.0
9	-0.0038	-0.0
10	-0.0149	-0.1
11	-0.0217	-0.1
12	-0.0272	-0.2
13	-0.0469	-0.3
14	-0.0596	-0.4
15	-0.0789	-0.5
16	-0.0956	-0.6

因子得点	+深み	+美的	+物足りない
	-金属的	-きたない	-迫力がある
線/駅	第1因子	第2因子	第3因子
堺筋/四ツ橋1	-0.89964	0.44034	1.5375
京浜東北1	0.83055	0.37874	-0.22491
御堂筋1	-2.398	0.68122	1.0792
山手1	-0.44521	0.80796	-0.14023
中央線1	0.2037	1.3594	-1.0063
京浜東北/東京	0.95627	1.2158	0.11864
山手/五反田	-0.6449	-0.26539	-1.9142
山手/新宿	1.1011	-2.3566	0.62396
中央/新宿1	-0.37552	-0.60224	-0.25289
山手/原宿2	1.2479	0.70207	0.4258
埼京/新宿2	0.58536	-0.4683	0.2264
堺筋/四ツ橋2	-0.052002	-0.10765	1.1471
鶴見緑地2	1.3554	0.10207	0.99004
御堂筋2	-1.7985	0.176	1.2491
東北本線1	0.30716	-0.90113	1.1841
山手/新橋	-0.93002	0.54002	-2.2777
山手/原宿1	0.41378	-0.55902	-1.0415
中央/信濃町	-0.49301	0.43388	1.9397
中央/お茶の水1	-0.68329	-1.1128	-0.50482
中央/亀戸	1.6002	0.47214	0.44507
鶴見緑地1	-1.3332	0.62301	0.2857
京浜東北2	-0.2706	-0.022367	-0.22859
山手線2	0.30969	0.6717	-0.94285
山手/東京	-0.15179	-1.978	-0.4617
中央線1	0.83999	1.6222	-0.20454
山手/目黒	-1.5563	0.32663	-1.2675
埼京/新宿1	-0.50303	-1.2073	0.13461
中央/大久保	0.56172	-0.67021	1.219
中央/お茶の水2	0.64122	-1.8152	-1.5127
中央/新宿2	1.581	1.5131	-0.6254

5. おわりに

3D 情報ブラウザ NW の音響機能について述べた。また、因子分析を用いた音響空間配置例として、駅の評価実
験例をとりあげ、研究発表用ツールとしての有効性を示そうとした。このようなツールの利用によって、音楽/音響心
理の研究発表がさらに効果的に行えるようになることを願っている。

「謝辞」駅の発車音楽に関する評価実験結果の利用を快く承頂いた鳥取大学・小川助教授に深く感謝します。

参考文献

- [柿元 96], 柿元俊博, 上原祐介, 3 次元情報検索インタフェース, 情処研報告 96-FI-41, 1996
- [増井 98], 増井誠生, 柿元俊博, 感性語評価語を利用した音響情報探索の試み, 情処研報告 98-MUS-26