情報メディア 11－1
（１９９３．５．１４）

人工現実感を用いた音情報検索

大木 直人 亀倉 龍 阿部 圭一 岡田 謙一 松下 温
慶應義塾大学理工学部

最近，多くのマルチメディアデータベースシステムが提案されているが，その多くは，マルチメディア情報のデータ構造をどうするかなど，データの処理方法に研究の重点がおかれ，どのようなインタフェースが使いやすいかについての研究は緒についたばかりである．このため，従来のマルチメディアデータベースのインタフェースについて問題点もいくつか指摘されており，分かりやすく，気持ち良く使えるインタフェースが必要とされている．

そこで，この論文では人工現実感を用いて，音声情報の長所を生かしたマルチメディアデータベースのための検索インタフェースを提案する．このインタフェースでは，仮想的な音場空間がユーザに提供され，ユーザはあたかも自分がこの空間を移動しているかのような感覚で自由に移動でき，興味のある音が聴こえる方向に自ら移動し，その情報を得る．

Auditory Information Retreival Using Artificial Reality

Naoto Oki Ryu Kamekura Keiichi Abe Ken-ichi Okada Yutaka Matsushita
Keio University
3-14-1,Hiyoshi,Kohoku-Ku,Yokohama 223,Japan

Recentry, many multi-media database systems were proposed, but many of them were emphasized how to handle multi-media information, not emphasized how to design their interface. So, the need of easy and comfortable multi-media retreival interface has increased.

In this paper, to take advantage of audio data, we thought to utilize the ear's ability, and propose an auditory interface for multi-media database system using artificial reality. Our auditory interface presents virtual sound scape to an user, and he/her can move arround this field as wish. And the user can make an access to a sound which he/her feel attractive, and get information about the sound such as textual or visual data.

1 はじめに

最近，各分野で大きな話題となっているマルチメディアであるが，データベースの分野でも音声情報や画像情報を登録・検索することができるデータベースシステム＝マルチメディアデータベースが数多く提案されている．しかしながら，これらマルチメディアデータベースシステムに関する研究では，マルチメディア情報のデータ構造に関する研究など，いかにしてマルチメディアを処理するかという点に重点が置かれ，マルチメディアを取り扱うデータベースインタフェースについての研究は，緒についたばかりであるのが現状である．さらに一般へのマルチメディア機器の普及につれて，誰にでも使いやすく，かつ誰もが興味を示すようなインタフェースの研究・開発が急務になっている．一方，マンマシンインタフェースの一つの流れとして，人工現実感に代表されるような臨場感を追求するものがある．ユーザをシステムが提供する仮想的な空間へと誘い，ユーザがシステムの内部処理を意識する事なく，システムとインタラクションが持てるという点でこのようなインタフェースは優れている．

一方我々は，映像情報と共にマルチメディアを構成する重要な要素である音声情報に特に注目して，いままであまり注目される事のなかった人間の聴覚の持つ能力を有効に活用する事を考えた．人間の聴覚は，ある音を断片的に聴いただけでもその音がなんであるか，どの方向から聴こえてきたか，どのくらい離れているかが判断できたり，複数の音が混在する環境でも自分が興味のある音だけを選択して聴くことができるという能力があることが知られている．これらの能力はより臨場感のある音インタフェースが提供される事によっていっそう発揮されることが指摘されている．

この論文では，人工現実感を用いた音声情報検索のためのインタフェースを提案し，これを用いたアプリケーションシステムについて述べる．

2 従来の音声情報検索

まず，従来のマルチメディアデータベースにおいて，マルチメディア情報，特に音声情報がどのような方法で検索されていたかについて簡単に述べておく．

2.1 リンク

文字データ・数値データを主に扱うためのデータベースで，拡張機能としてマルチメディアデータを扱うときによく用いられるのがこの方式であり，最も簡便な方法である．この方法では，マルチメディアデータは1つ，場合によっては複数の文字データ・数値データとリンクされ，ある検索条件によって文字データ・数値データが検索されると，それらとリンクされている画像データ・音声データが同時にユーザに提供される．

この方式では，あくまでマルチメディアデータは文字データ・数値データの付録として扱われ，それ自身を直接検索することはできないものが多い．

2.2 インデックス

インデックス法は，従来のデータベースにおけるインデックス検索をそのままマルチメディア情報に応用したものである．各データの整理番号あるいはデータ名と，そのデータが格納されている記憶デバイスのアドレスとの対応表がインデックスである．検索は，ユーザが整理番号あるいはデータ名を検索システムに入力し，システムが入力された整理番号・データ名をインデックスと照らしあわせ，そのデータのデバイス上の位置を得て，目的のデータを出力するという手順で行われる．

このシステムの身近な例として，CD がある．CD には曲目 (データ) に応じてトラック番号があり，ユーザはデータとトラック番号との対照表 (インデックス) を見て，聴きたい曲のトラック番号を知ることができる．この番号を CD プレーヤに入力することによって，聴きたい曲を聴くことができる．こうしてみると，CD はそれ自身が立派な音声データベースと言える．

このインデックス法では，ユーザが検索したい情報の整理番号・データ名を知らない場合には検索が非常に困難になるという欠点がある．例えば，ユーザが CD の曲名と曲番号との対応を覚えていないと，聴きたい曲を容易に検索できないことは往々にして発生する．

2.3 キーワード

キーワード法は，2. のインデックス法を使いやすく改良したもので，各データにそのデータの属性など関連する複数のキーワードを付加し，それらのキーワードによって検索しようという

方法である．多くのマルチメディアデータベースがこの方法を採用している．

しかし，音声データや画像データに含まれる情報を完全にキーワードで表現することは不可能である．例えば，ドビュッシーの曲 “月の光” やキリコの絵 “街の神秘と憂愁” が持つ雰囲気を完全に言葉で表現しようとしても無理なように，どんなに熟慮されたキーワードでも，結局はそのデータを聴いてみるか観てみるかしない限りは，その検索結果が適当かどうかは判断できない．

2.4 パターンマッチング

認識技術が発達し，コンピュータに対しマルチメディア情報を入力し，それを認識・内容解析することが可能となり，音声のデータを音声の検索条件によって検索することが実現した．このような検索方式は，キーワード検索のようにメディア間の変換 (音声データ ⇔ 文字データ) を経る必要がないため，より適切な音声情報検索が可能である．前述の鳥類図鑑 Hyperbook でも，検索したい鳥の鳴き声をユーザが真似ることによって検索条件を入力することができる．ユーザの鳴き真似を解析し，振幅構造・ピッチ構造・周波数成分変化構造などの音響的特徴を解析し，登録されている鳥の鳴き声データの音響的特徴との照合を距離関数 [矢川 89] によって行う．

この方式の問題点は，音声による検索条件の入力である．鳥類図鑑 Hyperbook では鳴き真似を採用することによってこの問題の解決を試みているが，鳥の鳴き声を事細かに表現するのは難しい．

3 ISF のコンセプト

前章で述べた，従来の音声情報検索方式の問題点を解消するために，人工現実感を用いて新たな検索方式を提供するインタフェースを提案した．我々の提案したインタラクティブな音場インタフェースでは，ユーザがマウスなどのデバイスで自分の動きをインタフェースに伝えると，インタフェースはその動きに応じた音場をユーザに提供する．これによってユーザはあたかも仮想的な空間で自分が移動したかのように感じることができ，さらに音場の動きに応じて次の行動を起こす．このように，ユーザはインタフェースと情報を交換しながら，仮想的な空間を動き回ることができるので，これをインタラクティブな音場インタフェース (Interface with Interactive Sound Field(ISF)) と名づけた [阿部 92]．

ISF は，音声データベースに登録されている音声データを仮想的な音場空間に配置して，ユーザはそれらの音声データが発する音の方向，距離を手掛かりに音場空間を移動し，欲する情報に辿り着く検索インタフェースである．ある情報に近づくことによってそのデータの持つ音声情報だけでなく，画像データなど他のメディアの情報をディスプレイ上に表示することもできる．

ISF は音声情報検索インタフェースとして以下の特徴を持ち，以下の節で詳しく述べることにする．

1. 検索条件の入力を必要としない．
   (a) 明確な検索条件が不要．
   (b) 検索条件のメディア変換が不要．
   (c) ブラウジングによる情報検索．
   (d) 簡潔なインタフェース．
2. 音像が左右の定位感，距離感を持つ．
3. 複数の音声データを同時に聴くことができる．
4. データ空間をアプリケーションに応じて演出することが可能．
5. 音声情報の階層化が可能である．

3.1 検索条件の入力

1. の検索条件を入力しなくても情報が得られるという長所は，いくつかのメリットを派生する．検索条件が不要なのであるから，従来のデータベース検索のように確固とした検索条件が無くても，“確かこんな音だった” という漠然としたイメージがあれば検索が可能である (1a)．

また，音を聴きながら情報を検索するのであるから，キーワード検索のように検索条件を無理矢理他のメディアに変換すること無く音声情報の検索ができる (1b)．

ユーザがまるで本をパラパラめくりながら情報を探すときのように，あれこれと情報を “つまみ食い” しながら欲しい情報を探すという方法は，ブラウジングと呼ばれており，画像インタフェースの分野ではよく知られている．もしかすると，これらブラウジングの途中で見られる情報の中にも有用な情報があるかも知れないし，現実世界では，そのような偶然チラッと見た情報が，実は有用な情報であるというケースは，研究

のための文献を探しているときに誰もがよく経験する事である．このような偶然の情報の発見は，従来のデータベース検索においては不可能である．このようにブラウジングは，人間的なインタフェースとして有効であり，これを音声情報にも適用できるようにした (1c)．

また，従来のデータベースシステムでは，検索のためのインタフェースが非常に複雑であった．これは，検索条件を入力する作業が煩雑である場合が多い．検索条件の設定はどうしても文字を入力する作業が必要になるし，全ての操作をマウスで行えるようなシステムも提案されているが，やはり検索条件を設定する作業が必要であることには変わり無い．この点 ISF では，ユーザは自分が音空間の中でどこに移動したいかを，入力デバイスを用いてシステムに指示するだけでよい (1d)．現在のところ入力デバイスはマウスを用いているが，方向が指示できさえすればよいので，キーボードの 10 キーでもデータグローブでも操作は可能である．

3.2 音場の立体感

ISF の音インタフェースは，左右の定位感・遠近感を伴って提供される (2)．音像の定位感・距離感の制御技術については次章で述べる．このような立体感を伴った音場は，ユーザがより自然な形で音声データを聴取できる環境を実現し，これによって同時に発せられる複数の音声データを聞き分けることが可能となる (3)．

人間には，同時に複数の音声が存在する環境で，ある特定の音声だけを聴き取る能力がある．例えば，人が大勢集まってガヤガヤと会話が飛び交っているカクテルパーティのような場所でも，自分の名前が会話に出てくると，パッと振り向くことができる．これになぞらえて，このような効果はカクテルパーティ効果と呼ばれている．カクテルパーティ効果は，個々の音源が独自の左右定位・距離定位を持った音像である場合により効果が強く現れる．そこで，音声データベースに登録されているいくつかのデータを個々の左右定位・距離定位で再生する事により，それらを識別・比較することができる．

3.3 音空間の演出

仮想的な音空間に音声データを配置する際に，ただ無秩序にデータを配置したのではユーザに混乱を招くだけである．そこで，仮想的な音空間を何かに似せて演出することを考えた．

ISF では様々に演出が可能だが，今回インプリメントしたシステム “ISF 昆虫図鑑” では，自然界を模しており，この空間には草原があり，木のざわめく林があり，せせらぎの聴こえる川が流れている．このような音で構成された仮想的な空間に，“鈴虫は草原に”“ミンミンゼミは林に” という具合にソース (音声データ) が配置されている．このため ISF における仮想空間は，ソース (音声データ=虫など) とオブジェクト (音声データ以外のもので，音を出すものも出さないものもある=木，川など) で構成されている．ユーザはこのような仮想的な音空間の中を，あたかも自分がその中にいるような感覚で歩き回り，情報を得ることができる．

3.4 音声情報の階層化

いくら人間にカクテルパーティ効果があるとはいえ，百や千もの音を同時に聞いて，その個々の音を聞き分けることは不可能である．一方，ハードウェアの面でも制約があり，今回実装した試作システムも，同時に 8 種類の音を独立した音声ラインに送出するのが限界である．このため，人間の側からも，ハードウェアの側からも同時に発生される音の種類は限られてくる．そこで，音声情報を階層化することを考えた．

音声情報をあらかじめいくつかにグループ分けしておき，あるグループに属する音声情報は仮想空間内のあるまとまった範囲に位置する．ユーザがあるグループの外にいるときには，そのグループの代表音がグループの重心に音像定位して聴こえている．グループの代表音は，ある一個の音声情報でもよいし，グループ全体のまとまった音でもよい．ユーザがグループの中に入ったときのみ，グループ内の個々の音がそれぞれの音像定位を持って音を発する．これによってより多くの音声情報が取り扱えるようになる．

例えば，昆虫のなかで蝉というグループを作っておき，ユーザがグループの外にいるときには蝉時雨がグループの中心辺りから聴こえるが，グループに入り込むと個々の蝉がそれぞれの場所で鳴いているのが聴こえてくるといった具合である．

4 ISF の実装

4.1 音像定位技術

ISF では音像が音場の中での位置がはっきりと認識できる必要がある．これは音像定位という

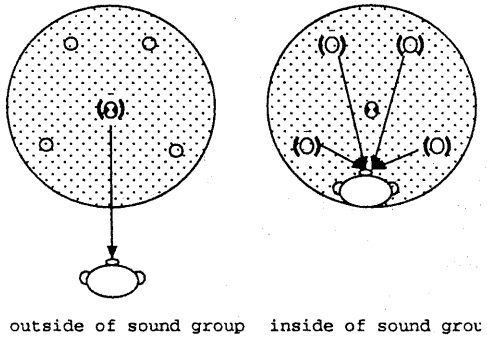

図 1: 音声情報の階層化

テーマで長い間研究されてきており，様々な手法が提案されている．我々がISF を実装する為に検討・実験した方法を以下に挙げる．

相関係数変化法

相関係数変化法 [黒住 84] [黒住 84-2] は，右信号と左信号の相関係数を 1 から-1 まで連続的に変化させることにより、遠方から聴取者頭部までの連続した距離感を持った遠近感を得る手法である．左右方向の音像定位に関しては，左右の音量差によって得ることができる．

エフェクタによる方法

主にシンセサイザやギターの演奏時に用いる，音に様々な変化をつける機器がエフェクタである．これを用いて疑似的に音の遠近感を表現する手法が提案されている [Cohen 90]．このシステムでは，主にウィンドウシステム上でウィンドウやアイコンの前後関係などを表現するために，この遠近感を用いている．例えば，遠方を表現するためにリバーヴ (残響) の効果を強くしたり，背後からの音を表現するためにロウパスフィルタなどを用いる．この方法では左右方向の音像定位は，左右の音量差の他に，左右の信号の時間差によっても得ることができる．

バイノーラル

バイノーラル方式では，人間の頭部を音響的に模したダミーヘッドの両耳の鼓膜に相当する部分に設置された，2 個のマイクロフォンによって録音される．聴取者はこれをヘッドフォンによって聴くことにより，ダミーヘッドがおかれた正にその場に居るかのような音像の定位感を得ることができる．現在では，ダミーヘッドを用いずにダミーヘッドの音響的性質の伝達関数を測定し，それを表わすデジタルフィルタを用いてバイノーラル効果を得る方法がある．この方法では，前後方向と左右方向の音像定位をまとめて得られる．

これらの方式のうち，ISF ではエフェクタを用いた方式を採用している．現在のエフェクタは MIDI 規格のインタフェースによって，コンピュータと相互に情報のやり取りが可能であり，制御がしやすいという長所がある．

当初，相関係数変化法の採用を考えていたが，事前実験として数名の被験者に相関係数変化法での距離感を体験してもらった結果，ホワイトノイズに関しては比較的良好な遠近感が得られたが，本論文で実装例として用いた虫の鳴き声については，あまりはっきりとした距離感が得られないということが分かった．これは以下の理由によるものと考えられる．

1. 相関係数変化法は前後方向の移動音像に対して距離感が顕著に現れる．
2. 周波数によって距離感が異なり，高い周波数域では距離感が乏しい．
3. 放送などで実際に用いられる時には，他の距離感を出す方法を組み合わせて用いられている．

2. に関しては，文献 [黒住 83] で指摘されている．昆虫の鳴き声は，高い周波数成分を多く含み，全ての周波数成分を持つホワイトノイズに比べると距離感が掴みにくい．

また，バイノーラルでは，デジタルフィルタで行われる畳み込み積分に必要な計算時間が非常に長く，システム全体としてリアルタイムな反応が期待できないために採用を見送った．

4.2 ハードウェア構成

本システムのハードウェアは図 2 のように構成されている．

- SUN4 Sparc Station 2

  システム全体を制御する．また，ディスプレイにより視覚情報を提供し，マウスによってユーザは音場空間を移動する．SUN シリーズは，MIDI ボードを実装するのが困難なために，MIDI ボードを実装したNEC PC-9801RA21 を RS-232C インタフェース

を介して，DP/4,EPS-16,DMP11 を制御している．

- ensoniq 4ch Signal Processer DP/4

  DP/4 は 4 チャンネルの独立した音声信号にそれぞれ独立した処理を行うことができ，これらの処理は，MIDI 信号によるリアルタイム制御が可能である．相関係数変化法による音像の距離感制御のための位相制御，距離感を提供するリバーブ (残響)・ディレイ (遅延)，音像の広がり感を提供するフェイズシフタ・コーラス，音質を変化させるイコライザなどが可能である．

- ensoniq Sampler EPS-16 plus

  ソース・オブジェクトの音源として，同じく ensoniq 社製サンプラ EPS-16 を用いている．EPS-16 もまた MIDI 信号による制御が可能であり，1MB のメモリに最大 32 種類の音をデジタル録音し，16 種類の音を同時に再生，8 チャンネルの独立したラインから音声信号を出すことができる．EPS-16 plus の音声データのサンプリングレイトは最高 44.8MHz と，CD と同等の音質を得ることができる．

- YAMAHA MIDI Mixer DMP-11

  EPS-16,DP/4 によって得られる複数の音像を，1 つの音場にまとめるために用いる，MIDI での制御が可能な 8 チャンネルデジタルミキサーである．8 チャンネルのアナログ音声ラインを入力時に A/D 変換し，その後の様々な処理は DSP を用いてデジタル信号のまま行なう．このため音質の劣化，ノイズの発生を抑えることができる．さらに，各チャンネル独立にパンニング (左右音像定位)・リバーブ (残響効果)・イコライジング (周波数成分変更) 等も可能である．

4.3 ISF 昆虫図鑑

ISF を用いたアプリケーションの一例として，虫の鳴き声をデータとして持つ “ISF 昆虫図鑑” を構築した．

ISF 昆虫図鑑では，仮想的な空間に 23 種の虫・蛙がソースとして，また木・草・水などがオブジェクトとして配置され，ユーザはこの空間を自由に動き回ることができる．これらのソースにユーザが接近するとデータウィンドウに虫の文字情報，画像情報が得られる (Fig.3)．

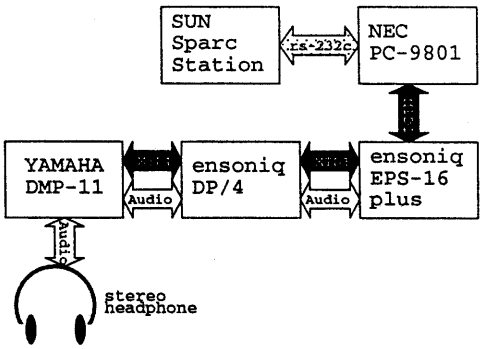

図 2: ISF ハードウェアシステム構成

図 3: ソースの画像・文字情報

また，音インタフェースだけでなく，音の方向・距離感を判断するための補助的なインタフェースとして，画像を用いたインタフェースをいくつか採用している (Fig.4)．これらの画像インタフェースは，X-window 上に X-View を用いて実現されている．

なお，実装に当たって音声情報は “デジタル最新録音& マスタリングによる効果音大全集 13 動物, [自然編]2 川・滝” (日本サウンド・エフェクト研究会, キングレコード) を用い，また画像データおよび文字データは “原色昆虫大図鑑　第 3 巻” (北隆館), “決定版　生物大図鑑　昆虫 I　チョウ・バッタ・トンボなど” (世界文化社), “日本の秋の虫” (築地書館) から引用した．

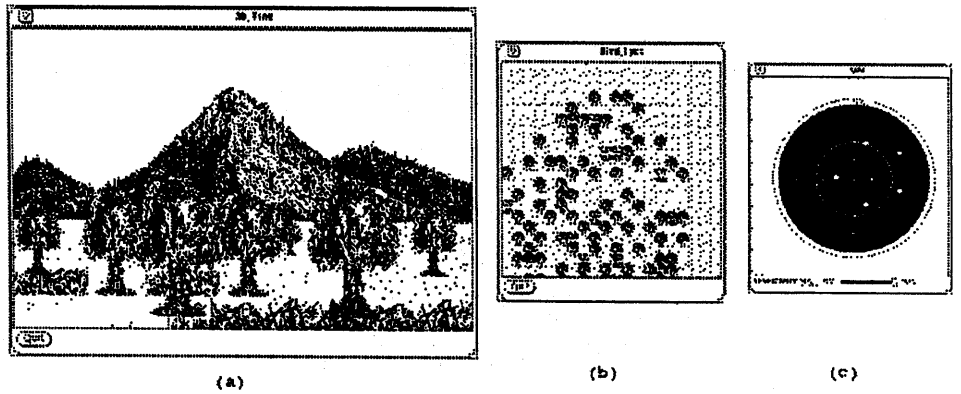

図 4: ISF で提供される画像インタフェース

5 評価

5.1 評価方法

評価は 20 人に ISF 昆虫図鑑を実際に使用してもらい，アンケートに回答してもらった．

アンケートの回答に先立ち，ISF 昆虫図鑑を使って以下の作業をしてもらった．

1. 被験者に各定位の音像を与え，音像の定位を体験してもらう．
2. 視覚インタフェースが何もない状態で，ある音声データを探し，接近してその音声情報に付随する画像情報を得てもらう．探すべき音データは，先だって被験者に示される．
3. レーダーウィンドウを表示した状態で，同様の手順で別の音データを探す．

作業後，被験者に下記のようなアンケートに回答してもらった．各質問項目には，“はい” から “いいえ” まで 5 段階で評価してもらった．

1. 虫のデータはすぐに捜し当てることができましたか？
   (a) 音だけの時
   (b) 画像のある時
2. 音像の左右の位置は，はっきりと分かりましたか？
3. 音像の遠近感は，はっきりと分かりましたか？
4. 音が混ざっているときに個々の音を聞き分けることができましたか？
5. 音データの検索インタフェースとして適当だと思いますか？
6. 音システムにおける汎用インタフェースとして興味を引かれましたか？

5.2 評価結果

Fig.5はアンケートの集計結果を示している．

1a,b の項目に関して，音だけでなく画像によるインタフェースを用いたときには半数以上の人がすぐに検索対象にたどり着くことができた．これに対して，音だけで検索を行ってもらったときには，たどり着くことができなかった人が$\frac{1}{3}$程度見られた．これは，被験者の個人差，あるいは習熟度によるところが大きく，“音だけの方がよい” という被験者もいた．また，現在のシステムでは，音声情報にユーザがたどり着いたかどうかの判定が，極めて厳格であり，ユーザが音声情報の真正面に位置したときのみ画像データウィンドウが開く．このため，音声情報のすぐ近くまでたどり着きながら，その付近でまごまごしてしまう被験者も見られた．

音の左右定位，距離感に関する質問 2,3 では，比較的良好な結果が得られている．左右定位と距離感では，距離感の方がより掴みにくいという人が多く見られた．

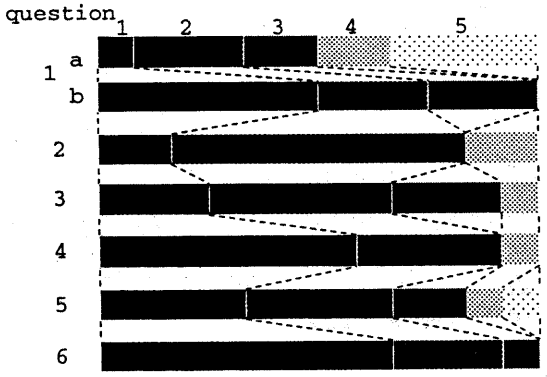

図 5: ISF 昆虫図鑑に関するアンケート結果

音の聞き分けに関する質問 4 では大半の人が良好な結果を示している．今回の程度の音源数におけるカクテルパーティ効果には，あまり個人差が現れず，だれもが持っている能力であると言える．

質問 5,6 においては，検索インタフェースとして高い評価を得ている．特に質問 6 では，大半の被験者がインタフェースに非常に興味を持っている．このことは，人間に易しいインタフェースとしては非常に重要であると思われる．

また，アンケートに書かれていたわけではないが，ISF 昆虫図鑑を使用して，“予想以上に音像がはっきりとわかる” という感想を多くの被験者から聞くことができた．日常生活で意識せずに当たり前のように音像の左右定位の判断や，距離感の判断を行なっているために，今回の実験であらためて自分の聴覚の能力を再発見したのだと思われる．この聴覚の持つ空間処理能力の再発見は ISF の目標の一つであり，この点では成功を修めているといえる．

6 結論

本論文では，音声データベースを人間の聴覚の空間処理能力を生かして検索するシステムを目指して，インタラクティブな音場インタフェースを提案し，これを ISF(Interactive Sound Field) と名付けた．また，ISF のアプリケーションとして ISF 昆虫図鑑を提案し，従来の音声情報検索システムに比べて，より効率的で人間の感性に合ったシステムであるという評価を得た．

参考文献

[阿部 92] 阿部, 大木, 亀倉, 岡田, 松下, “インタラクティブな臨場感を持った音場インタフェース”, 第 45 回情全大, 1992

[Cohen 90] M.Cohen, “Extending the Notion of a Window System to Audio”, COMPUTER, August, 1990, pp66-72

[角田 85] 角田, “脳の発見”, 大修館書店, 1985

[Krueger91] M.W.Kruger, (下野 訳), “人工現実 — インタラクティブ・メディアの展開 —”, トッパン, 1991

[黒住 83] 黒住, 大串,“2 チャンネル音響信号の相関係数と音像の質”, 日本音響学会誌,Vol.39,No.4,pp253-260,1983

[黒住 84] 黒住, 大串, “相関係数変化法による新しい音像の拡がり感制御方式”, 信学論, Vol.J67-A, No.3, pp204-211, 1984

[黒住 84-2] 黒住, , 二階堂, 大串, “相関係数変化法による音像の距離感制御方式”, 信学論, Vol.J67-A, No.9, pp872-879, 1984

[坂本 92] 坂本, “音楽の心身への影響”, 遺伝, Vol.46, No.2, pp41-44, 1992

[矢川 89] 矢川, 田淵, 村岡, “鳥類図鑑 Hyperbook における鳴き真似を用いた検索の実現方式について”, 第 39 回情全大, 7M-2, 1989