

## 人間の感情を考慮した音風景の構築\*

5J-3

寺本 邦夫 大木 直人 阿部 圭一 内田 美喜子 岡田 謙一 松下 温†

慶應義塾大学‡

## 1 背景

マルチメディアという言葉が一般に広く用いられるようになってずいぶんと時間が経過している。マルチメディアは、“マルチ+メディア”という言葉の通り、画像、音声、文字など複数のメディアを統合して処理できる環境として脚光を浴びており、現在では、ごく当たり前のように様々な分野で用いられている。しかし、マルチメディアは、これら複数のメディアがあってこそ意味がある場合が多い。したがって、これらのメディアのうち一つでも欠けてしまうと不快感を生じてしまうこともある。音の出なくなったテレビがよい例であろう。その他に人為的なミス、例えば、ビデオ撮影の際の音声や映像の取り忘れなど様々考えることは容易である。このようなマルチメディア情報の内、1個または数個のメディアが欠如している時、他のメディアから欠如しているメディアを創出する事を“メディアの補完”と呼ぶことにする。

著者らはメディアの補完の中でも「画像」から「音声」への補完に注目した。ここでいう「音声」とは、人間のコミュニケーション手段である会話だけでなく、波の音や風の音、昆虫の鳴き声など「文字」で正確に表現するのが困難な「音」を指す。また、今回取り扱う画像は風景画や風景写真に限定し、それに対応する「音風景」を補完する音検索システム OKeS(仮称)を構築した[1]。音風景とは英語の Soundscape に因んで名付けた言葉で、臨場感のある音場で構成された空間である。

## 2 提案

OKeS は、写真や絵(風景)から、それに合った立体的な音風景を作り出し、音付きの画像を表示するシステムである。その際、ユーザはいくつかの条件を入力することができ、自分の好みにあった音風景を作ることができる。さらに、ユーザは作り出し

音風景の中に入り込んで自由に動き回ることができる。

## 3 実装

OKeS のシステムは、ユーザの入力に応じて音を検索し指定された場所に配置される音検索部と、作成された音付きの画像を表示し、さらにユーザがその音空間内を自由に動き回れるシステム実行部からなる。

## 3.1 検索条件

ユーザは、画像上で海や森、川といったオブジェクトの種類と場所、さらにいくつかの検索条件を指定する。検索条件は大きく分けて客観的条件と主観的条件に分けられる。客観的条件とは同じオブジェクトでも、それによって発する音が変わってくるような条件であり、季節、時間、天候などがある。また、同じ写真でも見る人によって感じ方が異なってくるだろう。このような、見る人の感情といったものを主観的条件と呼び「荒々しい」「淋しげだ」といった感性語で指定することができる。現在、客観的条件には、時代、季節、時間、場所、天候があり、主観的条件は19種類ある。

## 3.2 音検索部

音データベースには様々な音源が格納されている。そして、ユーザの入力に応じてそれに最も適切だと考えられる音が検索され、指定された場所に配置される。例えば、ユーザの入力条件が、「海」、「荒々しい」ならば音データベースの中にあるいくつかの波の音から最も「荒々しい」とされる音が検索される。このことを実現するために、データベース中の音源と全ての入力条件はリンクで結ばれており、それぞれのリンクはメンバシップ値を持っている。メンバシップ値は0から1までの値で表され、音源がその条件にどの程度あてはまるかを表す。メンバシップ値が0なら、その音は条件に全くあてはまらなく、1ならばぴったりあてはまるわけである。オブジェ

\*Construction of Sound Scape Considering Human Feeling

†K.Teramoto,N.Ohki,K.Abe,M.Uchida,K.Okada,Y.Matsushita

‡Keio University

クトと客観的条件に対するメンバシップ値は、基本的には生物辞典などをもとに定めたが、直接記述のないものは著者の主観で定めた。主観的条件に対するメンバシップ値は、アンケートを行ない、それによって決定した。

ここでユーザの入力を「海」、「冬」、「荒々しい」とし、検索の例を述べる。ただし、説明のために音源は簡略化してある(図1)。

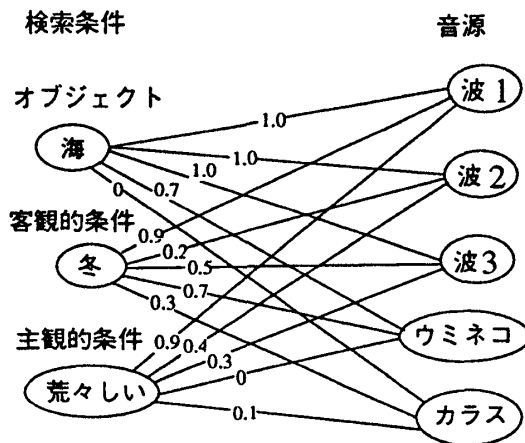


図1: 音源の検索

この例では、波1は「荒々しい」という条件に対して0.9というメンバシップ値をもっているのだから荒々しい波の音といえる。また逆に波3は「荒々しい」という条件に対してのメンバシップ値は0.3であるのであまり荒々しくない波の音といえる。そして、ユーザの全ての入力条件に対するメンバシップ値を考慮して音源がユーザの総合的な要求にどの程度適しているかを計算する。ある音源がユーザの総合的な要求にどの程度適しているかを表す値を、その音源のユーザの要求に対する満足度と呼ぶことにする。この満足度が最も大きい音源が検索されるわけである。次に、満足度を求める計算方法について述べる。

著者は、満足度を求めるのに、音源のユーザの入力条件に対するそれぞれのメンバシップ値の平均値を用いた。しかし、全ての音源の中から満足度の最も高い音源を検索するという方法ではオブジェクト一つに対して一つの音源しか検索することができない。波とカモメといったものは、全く異なったものであり、同時に海に存在していてもかまわないだろう。それどころか、現実の海には同時にいくつかの音が存在しているのが普通である。このように、より現実に近い音風景を構築するために、著者は、音源をグループ化し、それぞれのグループの中で最も満足度の高い音源を検索する、という方法を用いた。

この方法を用いると、図1の例において「波1」と「ウミネコの鳴き声」が検索される。

## システム実行部

システム実行部では、ユーザは、作り出した音場の中を自由に動き回ることができる(図2)。ユーザの位置は、人を後ろからとった写真のアイコンで表される。ユーザは前後左右の4方向に移動することができ、左右方向はアイコンも左右に動き、前方向はアイコンが小さくなり、後ろ方向はアイコンが大きくなることにより表わされる。また、音場の生成には、リアルタイムにユーザの動きに応じた音像定位を作り出すシステムであるISF (Interface with Interactive Sound Field) を用いた [2]。

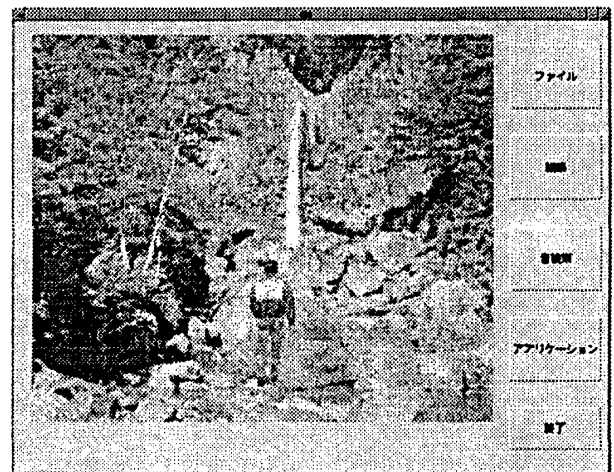


図2: システム実行部

## 4 評価

今回作成したシステムを被験者に実際に使用してもらったところ、良好な結果が得られた。

## 参考文献

- [1] 寺本, 大木, 阿部, 岡田, 松下, "音像から音風景を作成する試み", 情報処理学会全国大会, 1994
- [2] 大木, 亀倉, 阿部, 岡田, 松下, "人工現実感を用いた情報検索", 情報処理学会情報メディア研究会報告 IM11-1, 1993