マルチメディア通信と分散処理　44－11
（１９９０．１．２６）

対話型サウンド情報提供システム

山階　正樹　　小島　順治
ＮＴＴヒューマンインタフェース研究所

本報告ではＩＳＤＮ一次群サービスを利用して、高品質なサウンド情報や画像情報を対話的に検索できるシステムを提案するとともに、そこでの通信インタフェース、データベース構成、検索時のユーザインタフェースについて述べる。

特に、ＩＳＤＮのマルチチャネルを利用したパニック時の救済法、キーワードフィルタを用いた質問文の解析法と質問文の曖昧さを考慮した対話制御法について検討するとともに、構築した実験システムの構成について述べる。

A Study on Conversational Sound Information System

Masaki Yamashina　Junji kojima

NTT Human Interface Laboratories
1-2356 Take, Yokosuka-shi, Kanagawa, 238-03, Japan

A conversational sound information system using ISDN primary interface is proposed. Using this system, hi-quality sound and image information can be retrieved through ISDN from center data base.

The user interface main features of this system are help function using multi channel of ISDN at panic, query analysys using keyword filter and conversation control considering query ambiguity. The configuration of experimental system is also descrived.

1．まえがき

１９８８年春より、６４ｋｂ／ｓ系ＩＳＤＮサービスが開始されたことにより、その高速性、複数チャネルを活用したマルチメディア通信が可能となり、種々のサービスの試行実験が開始されている。さらに、６月からＩＮＳＮｅｔ１５００（ＮＴＴが提供する１．５Ｍｂｉｔ／ｓ系のＩＳＤＮサービス）が開始されたことにより、高品質なサウンド情報や画像情報を公衆回線を通して伝送できるようになってきている。

現在、公衆網を介したマルチメディア系の情報提供サービスとしてはキャプテンシステム[1]のように、ビデオテックス端末を利用し、静止画出力とタッチパネル、キーアダプタ入力による各種案内サービスが街頭や家庭で行われている。また、アナログ網を介したサウンド情報の提供サービスとしては一方向のオフトークサービスが実現されているが、これは放送形のサービスであり、ユーザが必要とするサウンド情報を検索できる機能は具備されていない。

本報告では、ＩＳＤＮの高速性を活かした高品質なサウンド情報の提供に重点を置き、ネットワークの双方向性を活かした対話形サウンド情報提供システムを提案するとともに、このシステムを実現するために必要な検索インタフェース、構成要素、さらに、構築した実験システムについて述べる。

2．システムの狙い

ＩＳＤＮを介して高品質のサウンド情報等のマルチメディア情報を蓄積したデータベースを対話的に検索できれば、ヒューマンインタフェースに優れた新しい情報提供システムが構築できると考えられる。そこで、高品質なサウンド情報が最も必要とされるのは、音響そのものが価値を持つ音楽の分野であるため、そこへ応用するシステムを考える。

音楽メディアの状況を見てみると、１ケ月当たりのクラシックコンサートに限ってみても約６００公演、ＣＤの新譜数は約１０００枚にのぼっており、これらの大量の情報の中からユーザの望む情報を効率よく検索し、内容を確認できる機能は有用と考えられる。サウンド情報をネットワークを介して提供する場合、ＩＳＤＮのベーシックインタフェースを用いると、７ＫＨｚ帯域（ＡＭ放送並み）のサウンド情報を提供できるのみであり、音響品質そのものが価値を持つ音楽分野でのアプリケーションを考えると品質的に問題がある。そのため、ＦＭ放送やＣＤ並みの高品質のサウンド情報や画像情報を提供する際にはH$_0$やH$_{11}$といった一次群系のＩＳＤＮを利用する必要がある。

そこで、Ｈ系のＩＳＤＮを利用して高品質のサウンド情報等を対話的に検索できるばかりでなく、ネットワークを介してホスト側にアクセスする必要がある予約・テレショッピング等の機能を具備したシステムを想定する。ここでは、コンサート情報や新譜情報等の音楽情報を一般ユーザが対話的に検索することにより、それらの商品サンプル的な音と映像を試聴できるとともに、チケットの予約やＣＤ・ＬＤ等のテレショッピングが可能なシステムを検討する。

3．システムコンセプト

音楽情報の場合、高品質なサウンド情報ばかりでなく画像情報も重要になりつつあるため、Ｈ系のＩＳＤＮを利用して音楽・画像双方のメディアを対話的に提供できるシステムを考える。図１に対話形サウンド情報提供システムの概要を示す。

ユーザはマルチメディア処理可能な端末からＩＮＳＮｅｔ１５００を介して自然言語インタフェース等を具備したマルチメディアデータベースにアクセスし、当該のサウンド・画像情報を試聴する。

さらに、コンサート情報ＤＢ，新譜情報ＤＢと連動した予約ＤＢ，販売ＤＢをユーザ情報によって更新する機能も想定する。

検索端末はベーシックおよびプライマリのＩＳＤＮインタフェース、画像復号化出力機能、Ｄ／Ａ変換サウンド出力機能、日本語入力機能、ポインティング機能、送受話機能を具備する。さらに

各々のメディアは別プロセスで制御するとともに、バッファ管理によってマルチメディアの同時処理を可能とする。また、サウンド情報・画像情報を含むマルチメディア情報を効率よく検索できることが重要であるため、マルチモーダルなユーザインタフェースを考える。

図1ではセンタ側のデータベースにサウンド・画像情報を蓄積する場合について説明したが、マルチメディア情報は端末側のＣＤ－ＲＯＭ等に蓄積し、検索処理はセンタ側でおこない、マルチメディア情報のアドレスのみをセンタ側から受信してサンプルを視聴する構成も考えられる。

以下の節では本システムの構成要素の中で、ユーザインタフェースに関わるホスト・端末間の通信インタフェース、サウンド・画像データベースの構成、ユーザインタフェースについて検討する。

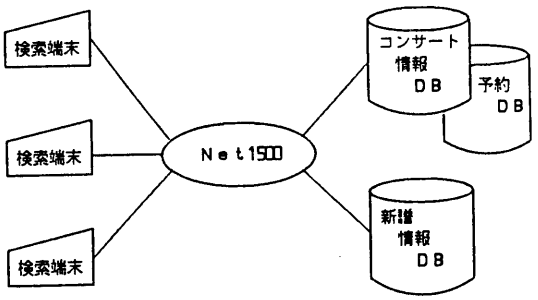

図1　サービス・システムの概要

5．通信インタフェース

本システムでは、（２Ｂ＋Ｄ）のベーシックインタフェースとＨ系インタフェースを用い、Ｄチャネルは双方のチャネルで共用する。対話処理はベーシックインタフェースを用いて行い、Ｈ系チャンネルは音楽・画像情報を伝送する場合のみ接続する。

端末側からホストを呼び出すと、まず、B_1 チャネルが設定され、検索情報、ガイダンス情報が交換される。

ホスト側は全く無人で検索用のソフトウェアのみが対応する場合、ホスト側からユーザをアシストする手段がなく、ユーザが検索情報や検索手段を誤った場合には、的外れな検索結果を出力し続けるか回線を切断してしまう場合が多く、ヒューマンインタフェース上問題がある。

そこで、本システムではホスト側で検索処理時のトラブルを検出する機能を具備し、トラブルを検出した場合にはＩＳＤＮのマルチチャネルを利用してB_2 チャネルを新たに設定することによりホスト側オペレータとユーザとの音声による会話を可能とする。ここで、トラブルの判断は以下の項目により行う。

（ａ）検索情報を解析してデータベースをアクセスするが該当する検索結果が一定値以下の検索処理の回数。

（ｂ）検索の実行ではない同一の処理手順が一定回数以上実行される場合。

（ｃ）自然言語で質問文が入力され、これを解析した結果のキーワードの抽出件数が一定値以下の検索情報入力の回数。

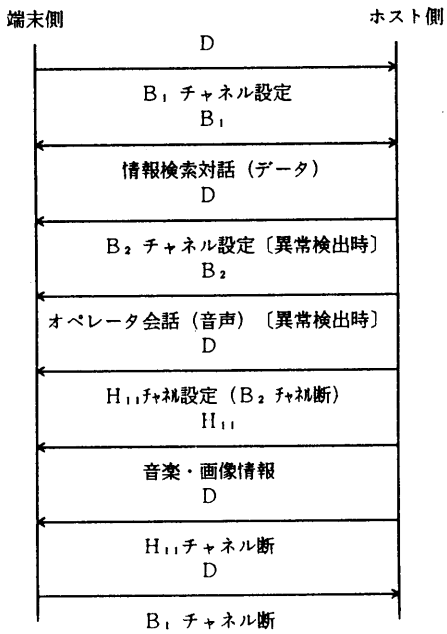

図2　接続手順

インデクッス静止画は、検索対象のサウンド・画像情報をイメージさせる静止画に言語的情報を重畳させた画像であり、原画像を縮小して一画面に複数表示する。次にユーザが視聴を希望するデータのインデクス静止画を指定することにより、該当のサウンド・画像情報を出力する。

（ii）キーワードフィルタ

自然言語によって、データベースを検索する方法[2)3)]は種々検討されているが、ここでは、対象世界を限り、ワードスポッテイングによる音声認識入力[4)]への発展を考慮して、単語間の格関係、共起関係に着目してキーワードの確からしさを検査する方法を検討する。ここで用いるキーワードフィルタはデータベースのフィールド項目に対応したキーワードを抽出するためのものであり、オブジェクト、述語、関係詞から構成される。ここで、オブジェクトはデータベースの各フィールドに記述されている項目、述語はオブジェクトをフィルタの種類によって定まる特定の格として取りうる動詞、関係詞はオブジェクトと共起する用語である。例として日時指定を解析するためのタイムフィルタとコンサートデータベースにおける演奏者フィルタを図３、図４に示す。

タイムフィルタの場合、オブジェクトからデータベースに記載されている数値データに変換する必要がある。そのため、「今週」、「来月」、「上旬」等の単語を年、月、日、時間についてｆｒｏｍ－ｔｏ型で記述されているテーブルを参照して数値データに変換する。

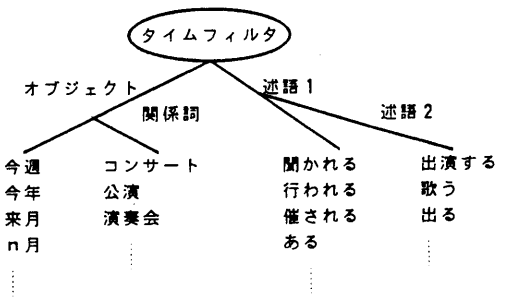

図５　タイムフィルタ

ここでの処理では、まず最初に入力質問文でフィルタとして定められている述語との一致をチェックし、一致する述語が存在する場合には、フィルタの種類と述語によって定まる特定の格を取るオブジェクトが存在するかを調べる。

ｅｘ１．の場合、「開かれる」に対して、フィルタで定められている格助詞「に」が存在し、「今月の下旬」を時間表現の名詞句と見て、この中でオブジェクトを捜す。この例の場合、名詞句の中は連体修飾になっているため、「今月」と「下旬」のＡＮＤ条件で数値データに変換する。

ｅｘ．１：

今月の下旬に開かれるピアノのコンサートは？

なお、「今日ピアノのコンサートはありますか？」等のように、時間表現の副詞的な用法に対処するため、格助詞が存在しない場合にも述語とオブジェクトのペアが存在する場合にはオブジェクトをキーワードと認定する。

また、ｅｘ．２のように用言が省略された場合に対処するため、オブジェクトが連体修飾語する語が関係詞であり、質問文中に述語が存在しない場合にはオブジェクトをキーワードと認定する。

ｅｘ．２：

今週の週末のコンサートは？

また、述語２としてフィルタに記載されている用語は「来月Ａさんが出演するコンサートは」等の質問文に対処するためのものである。

演奏者フィルタの場合、ｅｘ．３のように質問文中に「出演する」という述語が存在し、その述語の主格に「Ａ」、「Ｂ」、　あるいは「Ｃ」等のオブジェクトが存在する場合にはそれらをキーワードと認定する。

ｅｘ．３：

・９月にＡさんが出演するコンサートは？

次に、検索処理時にサウンド・画像情報の出力を指定された時のみH系チャネルを設定し、出力を終了すると直ちにこのチャネルを切断する。そして、予約処理等のデータのやりとりは、B₁チャネルを用いて行い、このチャネルは検索処理の開始時からこの時点まで保持される。図2に端末、ホスト間の接続手順を示す。

6．サウンド・画像データベース

ホスト側ではインデクス静止画情報、サウンド情報、画像情報を蓄積し、言語的情報によってインデクッスされている多元管理形のマルチメディアデータベースとして構成される。

インデクス静止画情報はサウンド情報や画像情報の内容をイメージさせるような静止画情報であり、検索処理時には検索情報に該当するインデクス静止画情報がまず伝送される。端末側ではこれらのインデクス静止画情報に言語的情報が重畳された画像が2次検索用のインデクスとして表示される。

サウンド情報は15KHz帯域のステレオ、画像情報は動画を想定し、双方のメディアは合わせて1．54Mb／s（現在提供可能な公衆網の最大容量）以下に符号化された情報として蓄積する。

なお、ここでの映像情報とサウンド情報の符号化方式はISOのDAPA委員会で検討が進められているディジタル蓄積メディア用符号化方式[5]を用いることを想定している。コンサート情報の検索を想定した場合のデータベース構成を図3に示す。

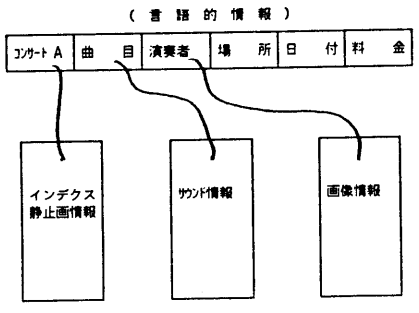

図3　サウンド・画像データベースの構成

4．ユーザインタフェース

大量のサウンド・画像データの中から目的の情報を効率よく検索できる機能は重要であり、ユーザインタフェース向上のため、メニュー検索の他に、自然言語による検索を可能とする。これは、検索対象が多い場合、操作が煩雑になるのを防ぐと共に、ユーザの要求が明確でない場合など、自然言語の持つ曖昧さを活かした検索を可能とするためである。また、サウンド情報・画像情報の一次検索結果を言語的な情報だけでなく画像情報で提示する機能を可能とする。

（i）対話制御

図4に自然言語によってサウンド・画像データベースを検索する際の状態遷移を示す。初期状態からまず質問文を受け付ける。次にキーワードフィルタは入力質問文からサウンド・画像情報をインデクスしている言語的情報の各フィールドに対応したキーワードを抽出する（アルゴリズムについては後述）。曖昧さなくキーワードを抽出できた場合には直ちにデータベースを検索し、また、キーワードは見つかったが意味的に曖昧さが生じた場合には、キーワードの確認をユーザに求める。さらに、未知語が検出された場合には質問文の修正、あるいは取消を求める。これは、自然言語による質問文の全ての解析を高精度に行うことは困難であるため、対話時の検索情報の曖昧さを解消するためである。

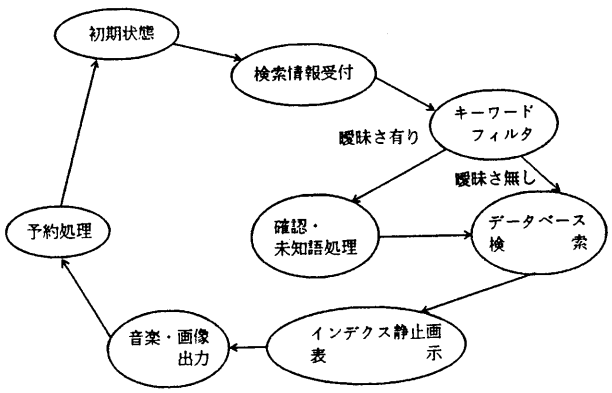

図4　対話制御

ｅｘ．４のように述語、関係詞ともに質問文中に存在しないが、オブジェクトのみが存在する場合、オブジェクトの意味が曖昧であると考えて確認を求める。

ｅｘ．４：

・９月にＡさんが主催するコンサートは？

また、ｅｘ．５のように述語は一致するが、オブジェクトが一致しない場合には、Ｘを演奏者に関する未知語と考えて、入力文の修正等を求める。

ｅｘ．５：

・９月にＸさんが出演するコンサートは？

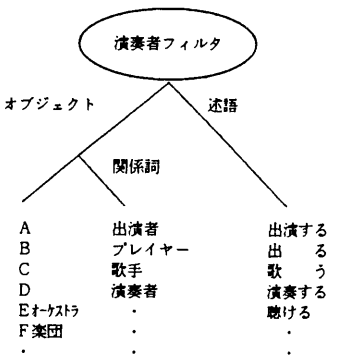

図６　演奏者フィルタ

７．実験システム

（ｉ）ハードウェア構成

対話形サウンド情報提供システムのサービス性マンマシンインタフェースの評価を主な目的に、実験システムを構築した。

本実験システムはホスト、端末をシミュレートする２台の汎用ワークステーションから構成されており、端末側では伝送されたサウンド情報のＤ／Ａ変換機能、ディスプレイへのフルカラーの画像表示機能等を具備している。ホスト側ではサウンド情報、画像情報等の蓄積機能を持つとともに、フレームメモリ、Ｄ／Ａコンバータを介してＶＴＲが接続されており、サウンド情報と画像情報をディスクへ取り込める機能を具備している。

（ｉｉ）シミュレーションソフトウェア

本実験システム上ではコンサート情報案内を想定したアプリケーションソフトウェアが動作可能であり、図７にソフトウェア構成を示す。

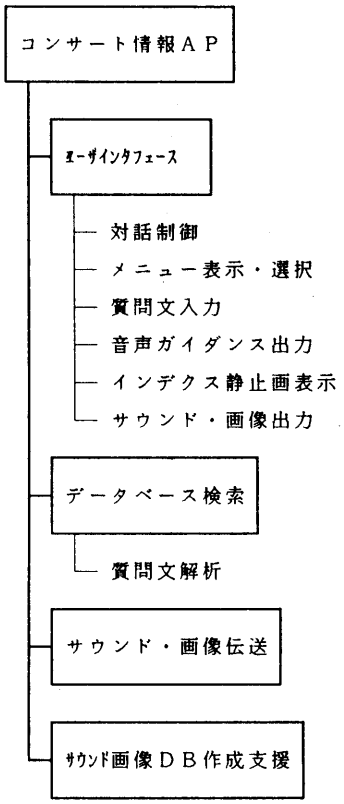

図７　ソフトウェア構成

①ユーザインタフェースモジュール

本モジュールは検索処理時の対話のシーケンスを制御するモジュールをベースに、メニューの表示・選択、質問文入力、音声ガイダンス出力やサウンド・画像出力等のモジュールから構成されている。静止画インデクス画面の例を図８に示す。

本画面で希望の静止画インデクスを指定することにより、サウンド・画像情報が出力される。

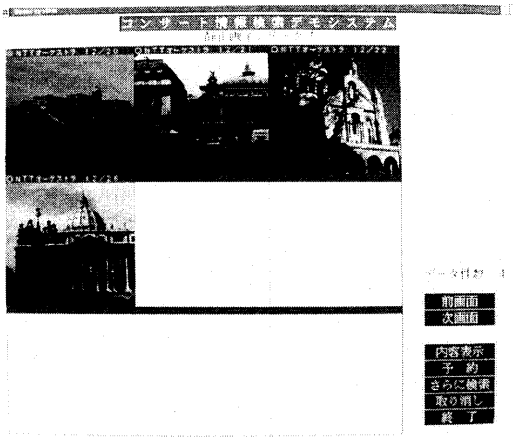

図8　静止画インデクス

②データベース検索モジュール

本システムでのサウンド・画像データベースは日時、演奏者、ジャンル、場所でインデクスされており、それらの情報での検索が可能である。また、それらの項目についてのフィルタがインプリメントされており、質問文からのキーワード、確認語の抽出を行う。

③サウンド・画像伝送モジュール

ここでは、サウンド情報、画像情報は圧縮処理を行わずに伝送しているため、H_{11}系の伝送速度に合わせて、サウンド情報は約1Ｍｂｉｔ／ｓ（サンプリング周波数：３２ＫＨｚ，分解能：１６ｂｉｔ，２ｃｈ）で伝送し、画像情報は一画面当たり約3Ｍｂｉｔ（３８０Ｘ３４０，一画素２４ｂｉｔ）であるため、画面／６秒で伝送している

④サウンド・画像ＤＢ作成支援モジュール

サウンド・画像データベースを効率よく作成するため、ワークステーションからコマンド形式でＶＴＲのタイムコードを指定することにより、任意のタイミングでサウンド情報、画像情報をディスクに取り込める機能とサウンド情報等と言語情報のリンクを取る機能を持っている。

8．むすび

本報告ではＩＳＤＮＮｅｔ１５００におけるサービスを想定した対話型サウンド情報提供システムを提案するとともに、そこでのユーザインタフェース、特に対話制御法、質問文の解析法や通信インタフェース、サウンド画像データベースの構成法等について検討した。さらに、ホスト、端末を想定して２台のワークステーションを対向接続したシミュレーションシステムの構成、コンサート情報の検索を対象としたアプリケーションソフトウェアの構成について述べた。

今後は本システムを用いてマルチメディアを用いたユーザインタフェースの評価を行うとともにサウンド画像情報の圧縮処理のインプリメント、音声入力等を用いたユーザインタフェースの向上網接続等を進める。

また、本システムはサウンド情報の検索ばかりでなく、カタログショッピング、情報案内サービス等のマルチメディア情報検索に適用できるものである。

参考文献

1）望月他：新ビデオテックス通信網システム、Ｓｈｉｓｅｔｓｕ，Ｖｏｌ．４０，Ｎｏ．２０
2）藤崎他：データベース照会システム「ヤチマタ」と名詞句データモデル、情処学会論文誌、Ｖｏｌ．２０，Ｎｏ．１
3）加藤他：日本語ＤＢ検索システムＱｕｅｓｔにおける意味解析、情処学会自然言語処理研究会４６－５
4）管村他：ＣＶ，単語スポッティングをベースとする連続音声認識システム、電子情報通信学会ＳＰ８８－９６
5）画像電子学会テレマティクス研究専門委員会：第２回テレマティクスシンポジウム資料