

対話型サウンド情報提供システム

3Z-7

山階 正樹 小島 順治

(NTT ヒューマンインタフェース研究所)

1. まえがき

ISDNサービスが開始され、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になってきた。さらに、INSNet 1500サービス(NTTの提供する1.5Mb/SのISDNサービス)が本年6月より開始されたことにより、高品質なサウンドや画像を公衆回線を通して伝送できるようになってきている。

本報告では、高品質なサウンド情報の伝送に加え、ネットワークの双方向性を活か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、その構成法について検討する。

2. システムの狙い

現在、ネットワークを介したサウンド情報提供サービスとしては一方のオフトークサービスが実現されている。また、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家庭街頭で行われている。今後、ISDNを介して高品質のサウンド情報、動画情報を含む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システムを構築できると考えられる。

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、まず、音楽分野へ応用するシステムを考える。ここで、音楽メディアの状況を見てみると、1ヶ月間のコンサートの開催回数はクラシックに限ってみても約600公演(89.5)、CDの新譜数は約1000枚(89.5)にのぼっており、これらの大量情報の検索をいかに効率良く行うかが問題となる。ここでは、コンサート情報や新譜情報を一般ユーザが対話的に検索でき、さらに、それらの商品サンプル的な音と映像を視聴できるシステムを想定する。

以下に、このシステムを実現するために必要な構成要素、対話インタフェース等について述べる。

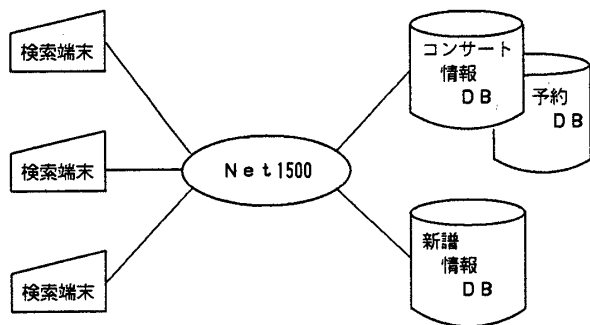


図1 サービス・システムの概要

3. システム構成

(i) 概要

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、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検討する。図1に対話型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検索端末からINSNet 1500を介して、自然言語インタフェースを持つマルチメディアデータベースをアクセスし、該当のサウンド・画像情報を視聴する。

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必要がある。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。

ホスト側に蓄積されるインデックス静止画情報、音楽情報、画像情報は、コンサート情報の場合には曲目、演奏者、場所、日付等の言語的の情報によってインデックスされている多元管理型のマルチメディアデータベースとして構成される。インデックス静止画情報はコンサートをイメージさせるような静止画を、音楽情報は、そのコンサートの曲目に応じたサウンド情報を、画像情報は、その曲目の演奏場面の画像を各々蓄積する。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される。コンサート情報の場合のデータベース構成を図2に示す。

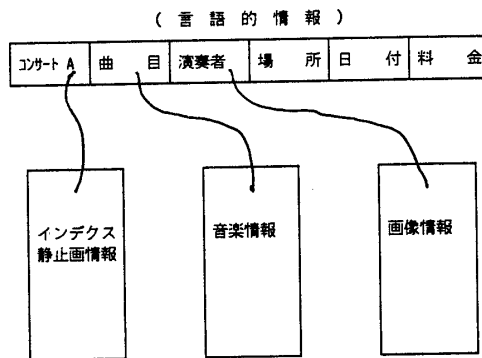


図2 音楽・画像データベースの構成

(ii) インタフェースによる接続手順

本システムでは、(2B+D)のベーシックインタフェースとH₁₁インタフェースを用い、Dチャン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、H₁₁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。図3に端末、ホスト間の接続手順を示す。

端末側からホストを呼び出すと、まず、B₁チャンネルが設定され、検索情報、ガイダンス情報が交換される。ここで、ホスト側は入力検索文を解析してデータベースをアクセスするが質問文を全く解析できない場合や同じ処理が繰り返し実行される等、ユーザと自然言語インタフェースの間のトラブルを検知すると、新たにB₂チャンネルを設定し、ユーザとホスト側のオペレータとの間の音声による会話を可能とする。次に、音声・画像の出力を指定された時のみH₁₁チャンネルを設定し、出力を終了すると直ちにこのチャンネルを切断する。

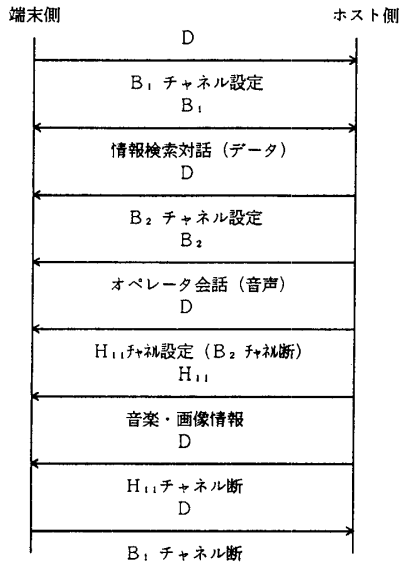


図3 接続手順

4. 対話インタフェース

(i) 対話制御

本システムの対話制御手順を図4に示す。検索機能はメニュー検索に加え、ユーザインタフェース向上のため、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。

キーワードフィルタは入力質問文からデータベースに記述されている項目に対応したキーワードを抽出(アルゴリズムについては後述)するものである。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

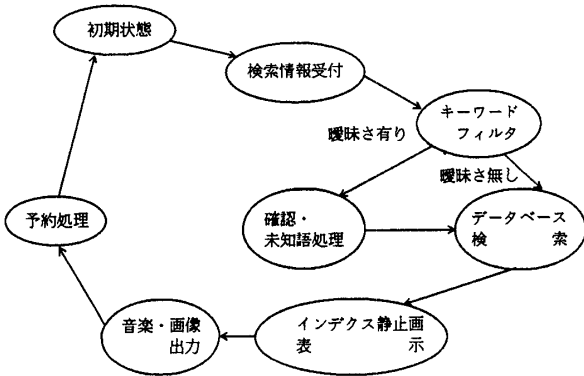


図4 対話制御

合には、キーワードの確認をユーザに求め、さらに、未知語が検索された場合には、メニュー画面を表示することにより希望項目の選択を求める。インデクス静止画面は、検索対象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、検索条件に合致する対象が複数存在する場合には、画像を縮小して一画面に複数表示する。次にユーザは視聴を希望するインデクス静止画を指定することによりサウンドおよび画像が出力される。

(ii) キーワードフィルタによる画面制御

キーワードフィルタは、演奏者、曲目、日時等のキーワードを抽出するための言語フィルタである。例として演奏者フィルタを図5に示す。フィルタはオブジェクト、述語関係詞から構成されている。オブジェクトはデータベース中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(この場合は主格)として取りうる動詞、関係詞はオブジェクトと共に用いる語である。

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、その主格がオブジェクト(データベース中に記述されている項目)と一致するかを調べ、一致する場合には演奏者に対するキーワードと判定する(ex.1)。

なお、述語が存在しない場合には関係詞との一致を検査し、オブジェクトとの共起関係によってキーワードを判定する。

ex.1:

・9月にAさんが出演するコンサートは?

また、ex.2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、演奏者選択用メニュー画面を表示する。

ex.2:

・9月にXさんが出演するコンサートは?

ex.3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.3:

・9月にAさんが主催するコンサートは?

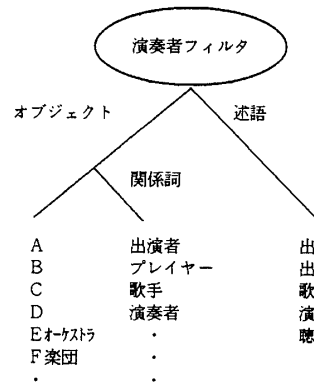


図5 演奏者フィルタの例

5. むすび

本報告ではISDNNet1500におけるサービスを想定したサウンド情報提供システムを提案するとともに、検索端末の具備条件、接続手順、対話インタフェース等について検討した。現在、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するシミュレーションシステム、および、コンサート情報の検索を対象とした簡易なアプリケーションを開発中である。今後は、マルチメディアを用いたユーザインタフェースの評価、マルチメディア端末構成法、対話インタフェース等について更に検討を進める。

最後に、日頃御指導、御討論頂く川嶋言語部長、石川プロジェクトリーダーに厚く感謝致します。