音声インタラクションによるマルチメディアデータからの内容検索
― テレビを知識源とした質問応答システム ―

１．　背景

現在，映像情報などがインターネット上で手軽に利用でき，また，DVDやハーディスクを媒体とした家庭用の大容量録画装置の普及により，これらのマルチメディアデータの中から必要な情報を簡単に検索できるシステムを開発することが有益であると考えられる．

また，小型携帯端末など，キーボード操作が出来ない端末から情報アクセスするためには，人間が日常使い慣れている音声インタフェースによるアクセスが望ましいと考えられる．現在，音声認識技術の大幅な進歩により，実用化に到るレベルに達している．

こういった背景により，提案システムの実現が可能になった．

２．　目的

膨大なテレビ番組や映画，講演などの音声・映像データやインターネット上のホームページといった複数のプラットホームから，統合的にユーザの所望の内容を表す情報を音声言語情報処理技術を用いて検索するシステムを開発する．

実現する機能としては，ビデオ録画などで収集した映像データのインデキシング機能，テレビニュース記事検索機能，質問応答機能，音声対話機能を設ける．本提案システムの機能を実現するためには,高精度な音声認識技術の開発が必要である．例えば，映像データに含まれる音声を認識システムにより文字列化し，そこから映像のインデックスを作成するため，認識率の改善が検索精度に大きく依存する．その為に必要な技術として，複数の音声認識システムを利用する，自動的に誤り部分を訂正する,ドメインに合うように自動的に言語モデルを学習することなどである．

また，データの中ユーザが質問した解答を抽出するための技術として，ユーザの質問の意図を高精度で汲み取る技術の開発も行う．本提案システムの研究成果は，今後バラエティーにとんだマルチメディア情報に対しても，誰もが手軽に扱える情報検索手段を提供すると考えられ，社会的意義は大きい．

われわれが普段生活しているときに，ふと疑問に思ったことをどこの家庭にもあるテレビ等の端末に向かって質問を行えば，その解答が得られるようなことをお茶の間で実現することが本プロジェクトの最終的な目標である．このプロジェクトではそういった目標の下で，先駆的なテーマとして，デジタルデータの内容検索に重点をおき，その技術とプロトタイプのソフトウェアの開発を行った．

３．　開発の内容

本プロジェクトで提供するものは以下のとおりである．

・マルチメディアデータのインデックス化ツール（動作環境：UNIX）

これにはフリーな音声認識システムを利用．データに含まれる音声を文字列に変換し，検索用インデックスを作成

⇒インターネットを利用した音声認識誤り自動訂正技術の開発とそのツール

⇒複数の音声認識エンジンの出力を統合した音声認識率改善技術の開発とそのツール

・情報検索エンジン・質問応答エンジン（動作環境：UNIX）

入力した検索要求から，ユーザの意図を解析．たとえば，

- **例1：単に文書を検索したいのか？**
- **例2：何か質問しているのか？**

など

そこで，こういったユーザの意図を解析する技術とそのツール，意図に応じて，データベースから文書を検索（情報検索）するエンジン，および質問に対する解答を探索するエンジン

・検索結果表示部（WEBによるグラフィカルユーザインタフェース，GUI）

・音声認識システムとの連携

・動画　・音声再生

といった特徴をもつ．WEBが使えるOSであればほぼ動作する．

上記の各モジュール等は，図１に示すような関係にある．また，完成したシステムの動作画面の例を図２に示す．図２では，ユーザの質問の音声認識結果，質問タイプの判定結果，検索のキーワード，および質問に対する回答をわかりやすく表示している．

図１．システム構成図

図２．システム動作画面

４．　従来の技術との相違

音声の質問を用いて，音声文書中から解答位置を推定するというシステムはまったく新しい研究分野であり，世界中でもあまり研究はなされていない（テキスト文書から解答を抽出する質問応答システムは盛んに行われている）．

本提案システムと類似な研究はRWC（新情報開発機構）で一部行われているが，実時間処理を目指した研究で，精度的に不十分なものである．アメリカのカーネギーメロン大学では，電子ビデオ図書館の情報検索の一環として研究を行っている．これが契機となり，音声検索が国防総省のDARPA国家プロジェクトとなっている．イギリスとケンブリッジ大学ではビデオメールの検索の研究を行っている．また，オーディオのインデッキシングソフトウエアとしては,BBN　Technologiesからは”Rough ‘n’ Ready”を、Fast-Talk Communicationsからは“ Ssearch Audio”を商品化している．これらの研究と比較して，本提案システムの特色としては，計算機との自然な対話により検索を進めていく，自然な発話の高精度の認識，未知語や誤認識に対応し，高精度な検索を実現させていくといった要素技術の統合をめざしている点である．

５．　期待される効果

本提案システムは，今後増大するマルチメディア情報を整理するうえで，ユーザが必要とする機能をできるだけ盛り込んだマルチメディア検索・閲覧システムとする．現在の世の中は，テレビ，ラジオ，インターネットといった様々なメディアを

通してマルチメディア情報が蓄積されているが，蓄積されているだけで，必要なときにすぐに取り出せないことが多く，うまく整理されていない状況にある．鉱山から金鉱を探すがごとく，これだけの多くの情報から埋もれた必要な情報だけを即座にマイニングする技術は，多方面において有意義であると考えられ，社会的意義は大きい．

映像などのマルチメディア情報のインデッキシング化には、言語情報を利用するのが有効である．このため，音声の書き起こしが重要な技術となる．提案システムの要素技術として，音声認識精度の向上を掲げている．現在の技術では決められた文章の読み上げ音声に関しては９５％前後の単語認識率が得られているが，講演や対話調の音声に関しては，まだまだ未熟であり，認識率は５０％～７０％程度である．しかし，提案手法により認識率の改善が見られれば，学会講演や，国会の演説のデータベース化，高精度な自動インデキシングにより，必要な講演発表や，演説を検索することができ，人員削減とコストダウンに繋がる．

また，音声対話による検索，質問応答を行うため，キーボードのような入力装置がない小型携帯端末からもアクセスが可能となり，時と場所を選ばずに必要なときに手軽に必要な情報を引き出せる．デスクトップで用いる場合でも，コンピュータの操作が苦手な人でも音声対話なら手軽に使えるだろうし，手足が自由に動かせないような身体障害者も音声で操作ができるようになり，社会情勢的に情報のバリアフリー化がますます進んで行くことが予想できる．

６．　普及の見通し

本システムは最終的には各家庭で個人に使用してもらうのを目的としているが，まずは，テレビ局やテレビ番組制作会社などの映像コンテンツ作成者に利用していただけるのではないかと考えている．

しかし，本システムで利用している音声認識技術は，ニュースなどの読み上げ音声の認識技術は実用化に耐えうるレベルであるが,様々な雑音を含んだり,話し言葉の音声になってしまうと，まだまだうまく認識できない．更なる音声認識技術の改善と高度な映像へのインデックス化が今後必要と考えられるが，これらの技術は日々進歩しているので，本プロジェクトで提案している自動的に映像コンテンツへのインデックス技術や内容検索技術は，数年後以内に実用化できるはずである．

７．　開発者名

西崎 博光（山梨大学 大学院医学工学総合研究部 hnishi@yamanashi.ac.jp）
中川 聖一（豊橋技術科学大学 情報工学系 nakagawa@slp.ics.tut.ac.jp）