

自動音声認識によるマルチメディアコンテンツ検索

サブワード認識に基づく語彙・文法・言語の制約のない音声検索実現



李 時旭 Shi-Wook Lee
いしうく
s.lee@aist.go.jp

情報技術研究部門
知的メディア研究グループ
研究員
(つくばセンター)

音声認識・対話・検索システムの研究開発を行ってきました。言語・文法・文脈・語彙などの言語知識を活かした大語彙連続音声認識の研究とともに、言語知識を不要とする音響信号のみに着目するユニバーサルシステムの研究にも高い関心をもち、多言語システムの開発を目指しています。

関連情報：

● 関連特許

①特許第 3686934 号「異種環境音声データの音声検索方法及び装置」

②特許第 3709436 号「音声認識用精細セグメント音響モデルの作成装置」

● 共同研究者

児島 宏明 (産総研)、田中 和世 (筑波大学)、伊藤 慶明 (岩手県立大学)

● プレス発表

2008年10月8日「マルチメディアを声で探す検索システムをインターネット上で実証公開」

● この研究開発は、産総研ベンチャー開発センターのベンチャー創出・支援研究事業の支援を受けて実用性を高め、産総研ベンチャー企業の創出を果たしました。

情報爆発時代のマルチメディアコンテンツ

情報通信技術や情報社会の発展による情報爆発時代における次世代の検索技術としては、画像、映像、音楽などのマルチメディア検索、インタラクティブ検索、連想検索などが求められています。テキストデータについては、商用の全文検索サービスが普及し、テキスト情報を活用する利便性は著しく高まっています。しかし、動画などのマルチメディアコンテンツは、動画配信サイトや大容量録画機器の普及により、その情報量が爆発的に増大しているにもかかわらず、そこからユーザーの求める情報を取り出そうとすると、現状では人手で付加したジャンルや要約などのタグ情報を頼りに検索するしかなく、効率的な検索が困難でした。

音声によるマルチメディアコンテンツ検索

私たちが開発した自動音声認識によるマルチメディアコンテンツ検索システムは、動画・音声などに含まれる音声を直接検索対象とし、コンテンツ中のキーワードの検索を可能にしました。一般に、音声検索は、大語彙連続音声認識に基づいてテキストに変換した上で全文検索を行います。システムの単語辞書にあらかじめ登録されていない新語や固有名詞などの未知語への対応が問題となります。これらの未知語に対して人手により新語の登録を頻繁に行わなければならないため、メンテナンス上の問題が生じて、検索システムの実用化への大きな障壁となっています。

今回開発したシステムでは、未知語問題の解決策として、テキスト検索ではなく音声を音素よりも細かく言語情報を保つ最小単位である「音素片 (SPS: Sub-Phonetic Segment)」という単位に分解・符号化して処理する手法を採用・実現しました。この最小単位をベースにすることで人間の発声した全音声を符号化できるため、辞書を使うこともなく、未知語への対応もできます。これにより、新たな固有名詞や新語を含むインターネット上の大量のマルチメディアコンテンツを、新語の辞書登録などのメンテナンスなしにリアルタイムで検索対象にできます。このシステムの実用化により、ユーザーが必要なマルチメディア情報を効率よく取り出すことが可能になり、これまで十分に利活用されてこなかった膨大なマルチメディアコンテンツから新たな価値を創出する可能性が広がります。

今後の展開

この研究で開発された音声検索技術の改良を進めるとともに、これまでのテキストベースの音声検索技術や、マルチメディア分類・要約などの技術と融合を進めて、より性能を高めていきます。また、インターネット上で公開した実証システムを通して、多方面からの評価用コンテンツの提供を受け、その情報を研究・開発にフィードバックすることにより、実用化を促進していきます。



動画の中身を音声でピンポイント検索するシステムのウェブ実証サイト
(<http://www.voiser.jp>)