



技術で社会を先導
産総研のR&D
 <94>

動画中の発信内容を音声認識してテキストに書き起こして全文検索が可能なシステム

進化する音声認識技術

インターネットと異なり、それ自体を索引として使えないため、発言内容などの詳細な情報の検索（全文検索）ができなかった。そこで産総研では、ユーザーが任意の検索語を入力すれば、それを含む発言を動画音声データから

検索できる音声全文検索・書き起こしサービス「PodCastle（ポッドキャストル）」（日本語版url: \podcastle.jp、英語版url: \e.p.jp、英語版url: \e.p.podcast.jp）を実現した。誰でも無料で、動画共有サービス（ニコニコ動画、YouTube、Ustream）やポッドキャストなどの14万件の動

動画音声検索・書き起こし

ユーザーが誤認識訂正

画音声データを検索できる。ポッドキャストルの実現には、情報技術研究部門緒方淳研究員が長年研究開発してきた独自の高精度な音声認識技術を活用し、動画音声データ中の音声をテキスト化（書き起こし）して索引付けすることが不可欠であった。これにより、ユーザーが入力した検索語と照合して全文検索が可能になる。しかし、高性能の音声認識技術を用いたとしても、必ずどこかに誤認識が生じる。しかも、これは時事用語や芸能人名など、日々増え続けて

いる新しい言葉に対応できなかつたため、インターネット上の多様な音声は扱えなかつた。これにより索引付けが正しくなるので、検索性能が向上する。しかも訂正結果を学習するので、多数のユーザーが訂正すればするほど音声の検索・認識の性能が向上する。さらに、インターネット上のニュース記事や

辞書などから新しい言葉を自動学習する新技術も開発した。実際に不特定多数のユーザーにより、累計58万単語の訂正がなされて音声検索性能が向上することを世界で初めて実証し、近年海外で注目され始めたクラウドソーシングに基づく音声言語情報処理の先鞭をつけた。

聴覚障害者を支援
 誰でも利用できる有用なサービスの持続的な研究開発・運用を通じて社会に貢献しながら、産業界と連携して本技術を実用化し、動画視聴の効率化や聴覚障害者支援、会議録作成など、さまざまな応用に展開していく予定である。さらに「ユーザーが利用中に認識誤りの訂正に協力するほど性能が向上するシステム」という世界初のアイデアを、音楽や画像などの他のコンテンツに適用する研究開発にも取り組み始めている。
 （木曜日に掲載）



産業技術総合研究所 情報技術研究部門 首席研究員 後藤 真孝

増加し続ける膨大なメディアコンテンツを計算機が人間に代わって理解できるようになったときに、初めて真のデジタル革命が起き、社会や生活が変わる。メディアコンテンツを計算機で自在に扱える技術の確立を目指し、音楽・音声などの音響信号の自動理解と、それに基づくユーザーインターフェースの研究を中心に、研究として面白くかつ世の中へのインパクトが大きい技術開発に挑戦している。