

音声認識の発展と教育における利用の展望

1. はじめに

人類の発展上、人間が火や道具を使うようになったことを第 1 の革命、蒸気機関の発明と共に人間が機械を使うことになったことを第 2 の革命というならば、現代の IT 技術を使った社会の変革は第 3 の革命（情報革命）と呼ぶにふさわしいほど、技術は発展し、我々の生活にも、好むと好まざるに関わらず、浸透してきつつある。

一方、コンピュータの普及はめざましいものがあるが、電話、テレビとは一線を隔て、世間一般ではまだ「難しいもの」という印象を受け止められているという印象も感じる。

人間とコンピュータのインターフェイスを振り返ってみると、入力装置としては、筆者の記憶しているところでは、16 進のキーボード、いわゆる J I S キーボード、マウス、タブレット、タッチパネルなどがある。また、出力装置としては、7 セグメントのネオン管から始まり、7 セグメント LED、CRD、液晶ディスプレイ、またプリンター、BEEP や FM 音源などによる音があった。今「ユビキタス」の時代と呼ばれる中、コンピュータはコンピュータという形ではなく、様々な形で我々の日常生活の利便さの向上のためますます身近なものになることから、これらのヒューマンインターフェイスの開発が今、大きな課題となっている。

今回は、ヒューマンインターフェイスの一つとして、音声認識の技術についてレポートをする。

2 . 音声認識の流れ

音声認識の歴史

人間の話し言葉をそのまま機械が理解してくれたらという科学者達の願いは、かなり昔からあった。ところが、認識すべき単語は無数にあり、さらに、話す人間の個人差も非常に大きく、音声認識の技術はかなり、困難を極めていた。現代においても、カーナビゲーションを中心に、最新の技術を駆使して実用化されているが、話し言葉をすらすらと解読する技術は確立されているとは言えない。

音声認識研究は、1952 年ベル研究所で始まったといわれる。研究は数字の認識を目標とし、音声のエネルギー分布をよりどころとするゼロ交差数という方式を用いた。

1970 年代になると DP マッチング方をロシアと日本で同時提案された。DP マッチング方とは、音韻の発声時間について、音声とを登録されている候補と比べ、最も似ているパターンを探し出すという方式で、荷物の仕分けなど手がふさがっている状況でのデータの入力に実用化された。

1970 年代から 1980 年代にかけて、HMM(Hidden Markov Model)の研究が進められた。

それまでのマルコフ方では音声データと入力データを比較していましたが、HMM ではそれにあいまいさを加え、似ていれば採用し、また似ていなければ切り捨てるというモデルを考えた。さらにそれらに学習機能を与えた。そうすると、最初のうちは支離滅裂だったものが使ううちに良くなっていくようになった。この技術により、単語から連続音声の認識へと大きく技術が進歩した。

また、DARPAプロジェクトによりQ & A方式の対話処理技術が開発され、電話による各種予約・問い合わせサービスに実用化された。

音声認識の現状

現在の音声認識のレベルは、声を登録し、明瞭な発声で読み上げた文章は、ほぼ確実に認識されるようになった。筆者の体験からすると、キーボード入力にかなり精通している筆者であっても、多少の訂正の時間を費やしたとしても、音声入力の方に軍配があがるまでに進歩した。

また、全ての言葉の認識まではいかないが、カーナビなどでは、不特定の話手の音声によるコマンドを、ある程度の騒音下でも、受け付けるシステムが実用化され、販売されている。

音声認識の今後の課題

現状でも触れたが、明瞭に朗読されたものは実用レベルにあるが、話し言葉や対話音声については、認識率が低い、また、学習機能はあるものの、発声者の意図や状況を判断するまでにはいたっておらない。

課題としては、話し手を固定しない、いわゆる不特定多数の話手の認識、対話や考えながら話した言葉の認識、文脈や状況判断、検索アルゴリズムの向上などが望まれる。

3. 教育における音声認識の利用

リアルタイム字幕提示システム

これは、話者の発話内容のすべてを文字に変換し、即座に聴衆に提示するためのシステムです。キーボードのおよそ3倍の速度で表示することができるため、今まで以上に聴衆に情報を提供することができる。

英会話の疑似体験

近年は、ALTの導入が進んでいるが、個別の英会話能力を指導するには限界がある。神戸の高等学校では音声認識の英会話ソフトを導入し、旅行や海外での生活、電話など日常のシーンを疑似体験しながら英会話力を身に着けるというもので、自分の声を録音して手本の発音と比較したり、画面に映るネイティブの実写映像とヘッドセットマイクを通じて実際に対話ができたりするのが特徴であり、生徒からの評判も高い。

失読症の児童・生徒への治療

アメリカの例であるが、失読症の生徒が音声認識装置を使って学習を進め自分の意見や考えを容易に紙面で表すことができるようになったという実践がある。また、この音声認識技術を30分ほど使っただけの子どもに対しても、読み、読解、綴り、理解の領域での改善が見

られたという報告もある。

4. 感想とまとめ

IBMからピアボイスが発売された時、筆者も非常に興味があり、体験させてもらった。最初にマイクの感度調整、サンプル文書(10分程度)の朗読などがあり、やや面倒くさいところはあったものの、朗読においては、漢字の変換ミス、文節のミス判断などは皆無とはいかないにもかかわらず、キーボードにかなり慣れていると自負している私でも、「これは使える」と思った。

しかしながら、普段、文書を作るときは、考えながら作成することが多いので、実際はほとんど使っていない。ただ、テープ起こしなどの、単調な入力求められるような場合は、使ってみようかという気持ちもある。

音声認識の技術は歴史を振り返ってみてもわかるが、言語学者達の人間の音声の分析、単語のデータベース作成、音声認識のアルゴリズムの開発と、想像を絶するような研究の成果の現れであるということは今更ながら感じた。

今後コンピュータ技術の進歩に伴い、音声認識の技術、そして、音声合成技術はますます進歩すると思われる。冒頭に述べた「コンピュータは難しい」という声が「コンピュータは便利だ」という言葉に変わるように、ヒューマンインターフェイスとしての音声認識技術について、期待していきたい。

また、この音声認識は教育にも生かされる分野が多い。私は、教育のIT化を推進する教育者・研究者として、今後もその技術の教育への応用について、調査、研究を進めていきたい。

参考文献

音声単語認知モデルの動向 心理学研究 70 号第 3 号 天野 成昭

音声認識・合成と自然言語処理の研究開発動向 亘理 誠夫

<http://www.nistep.go.jp/achiev/ftx/jpn/stfc/stt012j/feature3.html>

音声合成・認識技術 雑誌 F U J I T S U 49 巻 木村晋太

<http://magazine.fujitsu.com/vol49-1/>

Web ベースで実現した新しい「遠隔地リアルタイム字幕提示システム」 障害者高等教育センター

<http://www.dragonspeech.jp/exa.html>

音声認識ソフト利用の英語教育 現代教育新聞

<http://www.gks.co.jp/tokshu99/ninshik.html>