

研究室訪問

豊橋技術科学大学 情報工学系 中川 聖一 教授 訪問記

(第6回受領者)

8月の始め、荒木 光彦選考委員（京都大学大学院 教授）に同行いただいて当財団の中塚常務理事と共に、中川研究室を訪問いたしました。先生からは助成研究の「音声入力を核としたマルチモーダルなマンマシンインタフェース」の研究をはじめとする音声認識・対話システム等についてのお話を興味深くお聞かせいただきました。

- 研究室をご紹介いただき、現在の研究テーマなどについてお聞かせください。

研究室では音声対話技術の研究を基本テーマとしていますが、その基礎技術として音声認識技術や言語処理技術の研究を行っています。また、語学教育システムにも力を入れていて、この語学システムと Web 上の音声検索システム、音声対話システムを3本の柱にして研究を進めています。研究室のメンバは助手2名、博士を修了



した研究員1名、博士課程1名、修士課程7名で、各自がテーマをもっています。通算すると音声対話関係のテーマをこの10年くらい継続して中心テーマにしてきたこととなります。

いろいろ研究してきましたと音声認識技術だけでなく、人工知能的な知識処理、言語理解・意図理解や自然言語処理など総合的な技術が必要になり間口が広がっています。音声対話技術の研究は7年程前から世間でも始まっていましたが、継続して研究されているところは少なくなり応用面の研究に移ってきています。助成研究の対話システムは5年ほどかけて進めていたもので立石財団からの助成は、ちょうどプロトタイプシステムが完成する段階でしたので大いに役立ちました。

- 長年にわたって携われて来られたご研究のテーマについて詳しくお聞かせください。

助成研究は音声対話システムとして富士山観光案内をモデルにしたシステムです。ここでは対話だけでなく肉眼視やタッチキーを使ったサポート機能も利用し、また対話の履歴を表示するなどしたマルチモーダルなインタフェースにして使い勝手の良いシステムを目指しています。たまたま富士山観光案内ですが、同様の他所の案内システムを作る場合には、Web上のデータベースを賢く使えば1～2日で実現できるまでに完成度は高めてあります。

最近に実用化された技術としては、聴覚障害者向けのニュースの字幕放送があります。これはNHK放送技術研究所によって数年前に実現されましたが、この実現に協力しました。音声認識の精度は95%程度ありますが、実際のTV放送では正確さを期すためにオペレータが後修正をして放送されています。また、音声入力によるニュース音声記事の検索も研究しています。テレビやラジオの音声放送情報もマルチメディア時代には欠かせない情報であり、Web上の音声情報からの音声検索の重要性は増えています。技術的には不特定話者の音声認識に加えて、現場からのレポートの周囲ノイズの影響、話し言葉に含まれ易い言い淀み、くり返しや助詞欠けなどの難しい課題がでてきますが、それだけに音声にからむ総合的な技術が必要になります。

さらに、ホームページ上で乗車券の予約なども音声対話で行うことも実現しています。このような音声対話システムは入力方法や使いやすさに制約のあるモバイル環境で有効なインタフェース手

段になります。

次に、最近手がけている語学システムでは、話者のアクセントの位置を画面上に表示して手本との比較を視覚的に見られるようにしています。このシステムを利用して解ったことですが、日本人の癖（特性）でしょうか、英語発音のアクセントは音の高低（ピッチ）でつけていますが、ネイティブの人は音の強さ（ストレス）でつけています。この違いも画面上に表示されますので発音の矯正に役立つと思います。語学関係の人からかなり興味を持たれています。

○ 音声対話に関する実用的なニーズを重視されておられるように思いますが、实用レベルに達するまでにはご苦労な点多いと思います。

各システムの基本となるのは音声認識技術です。この技術の実力は、数字の発音では人の誤聴率が0.009%に対してマシンは0.72%程度ですが、5,000語程度になると人は0.9%、マシンは7%程度となりS/Nが悪ければ12%と差が広がってしまいます*。このギャップを埋めるのは難しい課題ですが、会話では、人は意味理解によってある単語に続く語彙を100語程度に絞って予測していると言われますが、方法は別にしてマシンでも同程度は実現していて、ニュースとか新聞に限定すれば70語程度に絞っていけます。このような言語処理は必要でしょう。難しいのは、音韻（音素）認識技術です。分野を限って語彙を少なくした使い方をすれば当然精度は向上します。単語認識率が95%程度あれば実用化に結びつくように思えます。

*詳細は、電子情報通信学会論文誌D-II vol.J83-D-II No.2 (2.2000)をご覧ください。

○ 早い時期から音声対話に注目したアプリケーションに取組まれて来られたわけですが、今後のご研究を進めていかれる方向をお教えください。

先程の語学システムの関係になりますが、生徒の発音の良し悪しをネイティブの教師が採点して判定するのに比べ日本人教師が判定した場合にはかなり判定にバラツキがあるというデータがあります。そういうことを考慮すると、この語学学習システムは日本人教師と同等かそれ以上の実力はあります。この9月には実際に使ってもらう予定になっていまして、教師の方が手軽に使えて教材を作れるシステムにはなっていると思います。今後の課題として、発音の表示を今は波形の違いで表示していますが、それを正しい調音するための口のあけ方や舌の位置などの指示、例えばCGによるエージェント画像による表示ができるようなインタフェースになれば生徒の学習速度は速くなると思いますので、発音の違いを波形で表示するだけでなく、生徒の立場に立ったより良いインタフェースへの工夫は必要と感じています。

ディクテーションの分野では、音声WPなど20年も前から言われ、研究もされてきていますが、なかなか使われている状況とはいえません。アプリ面の選択が重要で、紹介した語学教育分野と、Web上での音声データ検索・入力の分野がアプリケーションの目標として良いと考えて研究を進めています。特に語学教育分野では発音の評価まで行っている例がないので力をいれています。



実験室では研究されている音声対話システムや語学学習システムなどの動作モデルを拝見いたしました。たまたま研究室のある建物の耐震工事が行われていて、工事音が大きなノイズ源になっていましたが、雑踏の中での人と人の会話のようなやりとりの場面が現出され興味深く見せていただきました。このような人間味を感じられるご研究のお手伝いの一端を当財団の助成が担えたことに感謝うれしく思いました。

レポーター：事務局長 加藤 裕