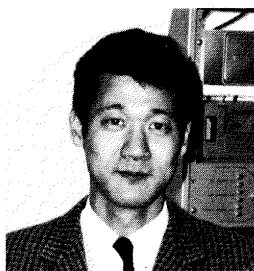


# 米国の音声研究について



在米国マサチューセッツ・  
ケンブリッジ・MIT  
東京大学工学部  
助教授 広瀬啓吉

## 1. はじめに

音声研究水準の特に高い国として、米国と日本をあげることが出来る。しかしながら、日本での研究が、音声高能率符号化、音声合成、音声自動認識などの工学的応用が主体であるのに対し、米国では、このような情報処理研究は勿論のこと、音声生成、音声知覚などの基礎研究も盛んであり、工学者のみならず、多くの物理学者、医学者、言語学者が、音声研究に携わっている。自由、独自性を重んじる国柄からか、特色のある研究機関が多く、その研究の進め方は、我々の研究においても大変参考になる。筆者が現在、文部省長期在外研究員として滞在している（1987年3月～12月）マサチューセッツ工科大学（MIT）では、音声生成・知覚・分析に重点を置いた特色のある研究が進められており、以下、MITでの研究を中心に米国の音声研究の動向を紹介してみたい。

## 2. 最近の研究動向

米国では、MIT、カーネギー・メロン大学(CMU)、ハスキンス研究所、AT&Tベル研究所、Bolt Beranek and Newman Inc. (BBN)、IBM、Texas Instruments (TI) などをはじめとし、多くの研究機関で質の高い音声研究が行われており、その研究内容は多岐に亘る。ここではそのすべてを紹介することは出来ないが、米国の

音声研究関連の会議として大きなものに、年1回春に開催される米国電子通信学会の音響・音声・信号処理国際会議 (IEEE ICASSP) と年2回春と秋に開催される米国音響学会の講演会 (ASA meeting) とがあり、その予稿集等から、米国の音声研究の全貌をつかむことが可能である。なお、1986年4月にはIEEE ICASSPが東京で開催されており、又、1988年11月には米国ホノルルでASA meetingが日本音響学会との共同で開催されることになっている。

現在、米国では、国家的研究プロジェクトとして、1983年11月にスタートしたDefense Advanced Research Projects Agency (DARPA) のStrategic Computing Programが進行中である。これは、先端軍事システムの構築、計算機技術における米国の主導的立場の維持を目的とした国防省関係のプロジェクトであり、多分に最近の日本での情報処理関連技術の著しい進展に刺激されたものである。16のプロジェクトで構成され、音声に関するものとして、音声認識・理解がとりあげられている。DARPAプロジェクトの前身として、Advanced Research Project Agency (ARPA) プロジェクトが1971年～1976年にかけて行われ、そこでも音声認識・理解がとりあげられた。幾つかの音声理解システムが発表された中で、CMUのHarpyシステムが、当初、目標とした1000単

語、文章理解率90%を達成したが、話題を文献検索に限り、統語構造に大きな制約を果たしていたため、利用範囲が非常に狭いという批判があった。

そこで、DARPAプロジェクトでは、統語・意味上の制約が大幅に緩和された連続音声理解システムの実現を一つの目標においている。このためには、まず連続音声の中の個々の音を精度よく認識することが必要であり、実際に、この考え方にそって、音響的特徴と音韻との対応に重点を置いた不特定話者用システムの研究がMIT、CMU、Stanford Research Institute (SRI) Internationalを中心に進められている。このシステムでは、まずスペクトログラム、あるいはその変形（後述の聴覚モデルにもとづく分析法の利用が検討されている）を用いて、音韻の認識を行い、音韻ラティスとして出力する。次に、辞書との照合により単語のスポッティングを行い、単語ラティスとして出力する。さらに、言語処理を行って、文章音声の理解が行われる。連続音声理解システムとしては、この他、特定話者用システムの開発がBBN、MIT (Lincoln Laboratories) で進められている。Byblosと呼ばれるBBNのシステムでは、音韻についてのHidden Markov Model (HMM)

から単語のHMMを作成し、認識を行う。15秒ほどの音声で音韻HMMを修正することによって高精度の認識が可能な、話者適応システムでもある。一方、Lincoln Laboratoriesのシステムは、戦闘機のコックピットなどでのストレスがかかった状況下で発声された単語音声を正しく認識することを目的とするものであるが、これもHMMを用いている。

DARPAの音声理解プロジェクトは1985年1月から開始された4年間のプロジェクトであり、現在、その前半が終了し、以上のような成果が出ているわけであるが、同時に、研究遂行上重要な音声データベースの作成も精力的に行われた。このデータベースは、TIで収録し、デジタル化した文音声を、後述のMITのSpireシステムで処理して音声記号を付与したもので、TIMITと呼ばれる。米語8方言の話者630名(男性65%、女性35%)が、一人10文宛発声した総計6,300文の音声からなる。セグメンテーションおよび区分された部分への音声記号付与は一応自動的に行っているが、完全ではなく、専門家によるチェックが必要となる。このようなデータベースは、音声研究を進める上で非常に役立つものであるが、大規模データベースとしては、TIMITが筆者の知る唯一のものである。

DARPAプロジェクトに参加していない研究機関で特に大きなものにAT&Tベル研究所がある。情報基礎研究部門に、20名ほどの研究員からなる研究グループが7つあり、そのうちの5グループが音声関連の研究を行っている。グループの全員が音声の研究者というわけではないが、著名な研究者が多く、すぐれた研究環境にある。韻律の特徴の分析とモデル化、声道モデルと音源モデルにもとづく音声合成、言語処理に重点を置いたテキストからの音声合成、HMM法にもとづく連続音声の認識など、言語処理も含めた音声関連の広い範囲の研究が進められている。

### 3. MITの音声研究

MITでは、Research Laboratory of Electronics (RLE) とLincoln Laboratoriesの2カ所で音声の研究が行われている。筆者が滞在して



RLEのある36号館。Speech Communication Groupは5階にある。

いる R L E には 13 の研究グループがあり、物性・デバイスから通信・信号処理までの電子工学に関連した幅広い研究が進められている。音声研究を行っているのは、元米国音響学会長の Stevens 教授を中心とする Speech Communication Group で、現在、教授 2 名、専任研究員 7 名、大学院生 14 名（博士課程 12 名、修士課程 2 名）など、総計 40 名ほどの研究者で構成されている。筆者のような、他研究機関からの客員研究員も 7 名いるなど、人の出入りが盛んなうえ、週一回のセミナーでは、他機関からの来訪者の講演を聴くことが出来、研究の活性化に役立っている。又、研究者の出身分野は、工学、物理学、音声学、歯学の多岐にわたり、音声研究のような学際研究を進めるのに適した陣容となっている。

このグループでは、優れた音声情報処理機能を有する人間が、音声をどのように発し、聴きわけているかを明らかにすれば、それを機械に反映させることによって、高性能の合成・認識システムを構成することが出来るとの考え方にもとづいて、研究が進められている。従って、音声の生成・知覚といった基礎研究に重点が置かれ、その結果、得られた成果に立脚して、音声の合成・認識といった応用研究が行われている。以下に、幾つかの研究を簡単に紹介する。

聴覚モデルにもとづく音声分析手法：音声に対する内耳有毛細胞、シナプスの応答特性を、帯域フィルタ、半波整流回路、自動利得調整回路でシミュレートし、平滑出力およびピーク強調出力を求める分析法で、音素境界、スペクトル・ピークを明確にとらえることが出来る。

サウンド・スペクトログラムを利用した音声の認識：熟練した音声研究者は、スペクトログラムを読み取り、発声内容を推測することが出来る。そこで、人間がスペクトログラムのどこに注目して推測しているかを定量的に表わし、音声認識に利用すれば、高精度の不特定話者音声認識システムの実現が可能となる。このような考え方で、連続音声中の音素の認識の研究が進められている他、熟練者を養成するための Spectrogram Reading のコースが開かれている。

テキストからの音声合成：テキストを入力とし

て音声を出力する実用的システムを世界に先がけて発表した。これをもとに、幾つかの企業で商品化がなされ、DECTalk などとして市販されている。現在も、音質向上の努力が続けられ、特に女性音声の改良が精力的に進められている。

この他、人間の音韻識別特性の解明、弁別素性にもとづく音素の認識、音韻ラベル付けされた音声データベース、磁気センサを用いた調音器官の動態検出などの研究が行われている。なお、筆者は、日本語破裂音・鼻音の識別に利用し得る音響的特徴を調べるとともに、認識アルゴリズムの構築を進めている。

研究に利用されている計算機システムの主なものは、VAX11/750 と Symbolics3650・3670 である。特に、Symbolics 上では、通常の音声分析のほとんど総てが可能な Spire と呼ばれるプログラムが開発されている。非常に整備されており、計算機の専門家以外でも比較的容易に利用出来る。なお、このプログラムは、MIT の License Office を通して購入出来る。



Symbolics のディスプレイ。Spire による分析結果が表示されている。

#### 4. おわりに

MIT に来て特に強く感じた事は、学生といえども研究に対する自分の意見をしっかりとっており、よく議論すること、先生もまたそれを尊重する（認めるわけではない）という点である。このような風潮が独創的な研究を生む下地となっていると思われる。日本のような師弟関係を否定するわけではないが、ややもすれば独創の芽をつむ結果になることがある。