

「数理科学」は語る

30年前から現代へのメッセージ

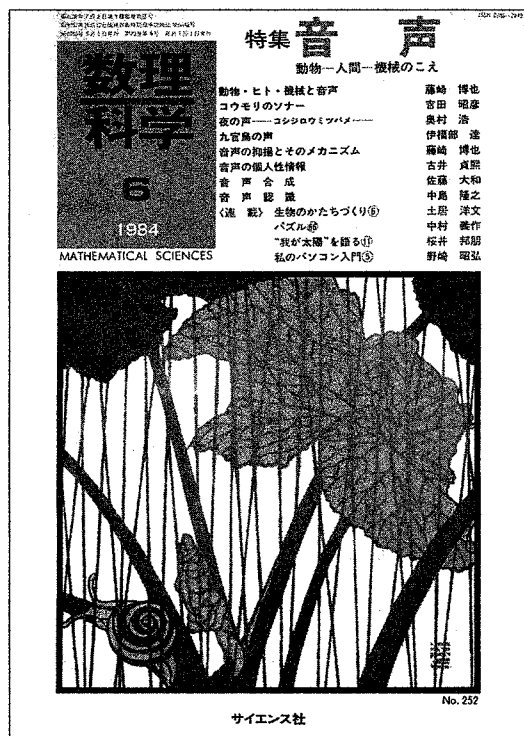
佐藤 大和

1984年6月号

「音声合成」が書かれた1980年代初めの頃、私たちの研究グループが所属する電電公社（現NTT）電気通信研究所では、PARCORなど現在の携帯電話の音声処理の基礎となる技術、線形予測符号化が發明され、技術的發展を遂げつつありました。その頃米国のMITでは、J. Allen博士が中心となって英語のテキストを“朗読”するtext-to-speechシステムの開発が進んでいることを知り、私たちも日本語（漢字かな混じり文）で同様の朗読システムを作ろうということになり、1年ほどかけて完成させました。その内容を紹介したのが30年前の「音声合成」です。

私を含めて担当した4人の研究者は皆理系学部出身の人間でした。“朗読システム”を創り上げるには、言語学、日本語文法、アクセントやイントネーションを含む日本語音声学など、いわゆる文系の学問の知識が必要なのですが、こうした知識には疎かったため、国立国会図書館や国立国語研究所などに通うなどして文献を集めたのも懐かしく思い出されます。入社当時上司から「これからは人間の研究が大切だ」と言われて従事することになった“音声合成”の研究は、人文科学分野の情報処理のはしりだったのでしょう。現在では人文科学のあらゆる分野において、研究のためコンピュータが必須になっている状況を見ると隔世の感を禁じえません。4人の担当者は、その後皆、人間に関わる分野と情報処理の狭間の領域で活動する研究者として育っていきましたが、これも若い頃のこの開発経験によるものかもしれません。

最近の音声合成はどうなっているでしょう。30年前当時に比べると合成音の品質は著しくよくなりました。その原因の一つは、連続音声を組み上げる音声要素（単位）や言語処理などが、膨大な音声言語データに基づく、いわゆる corpus-oriented な手法に依っているからのように思われます。多様な音声事象を広くカバーする音声データの中から、最も相応しいと推定される要素を選択して利用するなどの手法です。コ



ンピュータの性能の著しい発展に支えられて、こうした音声合成の技術も進化してきましたが、その反面、言語音声の諸事象を具体的に明らかにする研究に関しては必ずしも十分ではなかったように思われます。近年、英語、日本語をはじめ世界の主要諸言語の音声／言語コーパスの構築が進んでいます。こうしたコーパスの情報処理に基づく分析研究から、文系と理系の垣根を越えた音声言語学の新たな領域が開かれていくのではないかと期待しています。

（さとう・ひろかず、東京外国語大学アジア・アフリカ言語文化研究所）