

日本音響学会 編
The Acoustical Society of Japan

音響サイエンスシリーズ **18**

音声言語の自動翻訳

コンピュータによる自動翻訳を目指して

中村 哲
編著

Sakriani Sakti Graham Neubig
戸田智基 高道慎之介
共著

コロナ社

音響サイエンスシリーズ編集委員会

編集委員長

富山県立大学

工学博士 平原 達也

編集委員

熊本大学

博士(工学)

川井 敬二

九州大学

博士(芸術工学) 河原 一彦

千葉工業大学

博士(工学)

荻木 禎史

小林理学研究所

博士(工学) 土肥 哲也

神奈川工科大学

工学博士

西口 磯春

日本電信電話株式会社

博士(工学) 廣谷 定男

同志社大学

博士(工学)

松川 真美

(五十音順)

(2017年6月現在)

刊行のことば

音響サイエンスシリーズは、音響学の学際的、基盤的、先端的トピックについての知識体系と理解の現状と最近の研究動向などを解説し、音響学の面白さを幅広い読者に伝えるためのシリーズである。

音響学は音にかかわるさまざまなものごとの学際的な学問分野である。音には音波という物理的側面だけでなく、その音波を受容して音が運ぶ情報の濾過処理をする聴覚系の生理学的側面も、音の聴こえという心理学的側面もある。物理的な側面に限っても、空気中だけでなく水の中や固体の中を伝わる周波数が数ヘルツの超低周波音から数ギガヘルツの超音波までもが音響学の対象である。また、機械的な振動物体だけでなく、音を出し、音を聴いて生きている動物たちも音響学の対象である。さらに、私たちは自分の想いや考えを相手に伝えたり注意を喚起したりする手段として音を用いているし、音によって喜んだり悲しんだり悩まされたりする。すなわち、社会の中で音が果たす役割は大きく、理科系だけでなく人文系や芸術系の諸分野も音響学の対象である。

サイエンス (science) の語源であるラテン語の *scientia* は「知識」あるいは「理解」を意味したという。現在、サイエンスという言葉は、広義には学問という意味で用いられ、ものごとの本質を理解するための知識や考え方や方法論といった、学問の基盤が含まれる。そのため、できなかったことをできるようにしたり、性能や効率を向上させたりすることが主たる目的であるテクノロジーよりも、サイエンスのほうがすこし広い守備範囲を持つ。また、音響学のように対象が広範囲にわたる学問分野では、テクノロジーの側面だけでは捉えきれない事柄が多い。

最近では、何かを知ろうとしたときに、専門家の話を聞きに行ったり、図書館や本屋に足を運んだりすることは少なくなった。インターネットで検索し、リ

ii 刊 行 の こ と ば

ストアップされたいくつかの記事を見てわかった気になる。映像や音などを視聴できるファンシー (fancy) な記事も多いし、的を射たことが書かれてある記事も少なくない。しかし、誰が書いたのかを明示して、適切な導入部と十分な奥深さでその分野の現状を体系的に著した記事は多くない。そして、書かれてある内容の信頼性については、いくつもの眼を通したのちに公刊される学術論文や専門書には及ばないものが多い。

音響サイエンスシリーズは、テクノロジーの側面だけでは捉えきれない音響学の多様なトピックをとりあげて、当該分野で活動する現役の研究者がそのトピックのフロンティアとバックグラウンドを体系的にまとめた専門書である。著者の思い入れのある項目については、かなり深く記述されていることもあるので、容易に読めない部分もあるかもしれない。ただ、内容の理解を助けるカラー画像や映像や音を附録 CD-ROM や DVD に収録した書籍もあるし、内容については十分に信頼性があると確信する。

一冊の本を編むには企画から一年以上の時間がかかるために、即時性という点ではインターネット記事にかなわない。しかし、本シリーズで選定したトピックは一年や二年で陳腐化するようなものではない。まだまだインターネットに公開されている記事よりも実のあるものを本として提供できると考えている。

本シリーズを通じて音響学のフロンティアに触れ、音響学の面白さを知るとともに、読者諸氏が抱いていた音についての疑問が解けたり、新たな疑問を抱いたりすることにつながれば幸いである。また、本シリーズが、音響学の世界のどこかに新しい石ころをひとつ積みきかけになれば、なお幸いである。

2014年6月

音響サイエンスシリーズ編集委員会
編集委員長 平原 達也

ま え が き

話した言葉をその場で相手の言語に翻訳して、あたかも母国語で話しているかのようなコミュニケーションを可能とするのが、自動音声翻訳あるいは音声自動通訳システムと呼ばれるものである。すでに、種々のSF映画や漫画では当たり前のように出てくる技術であるが、長い研究開発の末、ようやく実現が近付いてきた。特に、日本人の外国語への苦手意識がきわめて高いことから、わが国では世界に先駆けて音声翻訳の研究が、1986年に国際電気通信基礎技術研究所の中の自動音声翻訳研究所において開始された。筆者の中村はその発足から3年間プロジェクトに参加、さらに、2000年から2011年までプロジェクトを率いる立場で参画した。その中で、コーパスベースの確率モデルによる統計的な音声認識・合成や機械翻訳へのパラダイムシフト、多言語化のための国際共同研究、標準化、種々の実証実験、携帯電話による音声翻訳の実用化などを経験、主導した。本書ではこれらの活動を通じた、これまでの自動音声翻訳・自動音声通訳技術の研究開発について述べていく。

本書は、筆者の中村が2011年に奈良先端科学技術大学院大学に異動し、知能コミュニケーション研究室で、共著者の戸田、Sakti、Neubigとともに音声自動通訳の基礎研究を行ったことを機に執筆を決心した。それぞれ、音声翻訳のプロジェクトを率いてきた中村、音声合成および声質変換の専門家である戸田と高道、多言語音声認識の専門家であるSakti、そして、機械翻訳の専門家であるNeubigという、執筆時点で研究の先端を行く筆者が本執筆を担当した。

本書の出版においては、音響サイエンスシリーズ編集委員会 委員長の富山県立大学の平原達也教授に本書のご提案をいただいた。また、コロナ社には企画から原稿の出版まで長きにわたり大変ご尽力をいただいた。また、これまで音声翻訳の研究開発をとともに進めてきた株式会社 エイ・ティ・アール自動翻

訳電話研究所，音声翻訳通信研究所，音声言語通信研究所，株式会社国際電気通信基礎技術研究所・音声言語コミュニケーション研究所，情報通信研究機構の共同研究者の皆様，音声翻訳の研究に継続的に資金提供していただいた総務省，文部科学省，内閣府の皆様，そして，奈良先端科学技術大学院大学情報科学研究科知能コミュニケーション研究室のスタッフ，学生諸君に感謝する。

2018年3月

中村 哲

執筆分担

中村 哲	1章，2章，3.1節，3.2節（日本語訳），4章，5.1節，6章
Sakriani Sakti	3.2節（英語原文）
Graham Neubig	3.3節，5.2節
戸田智基	3.4節
高道慎之介	3.4節

目 次

第 1 章 音声翻訳の概要

1.1 音声翻訳システム	1
1.2 自動音声翻訳のこれまで	2
引用・参考文献	5

第 2 章 話し言葉の異言語コミュニケーション

2.1 話し言葉の性質	6
2.2 コミュニケーションとはなにか	10
2.3 言語情報と非言語情報の役割	14
引用・参考文献	15

第 3 章 自動音声翻訳の構成要素

3.1 音声翻訳モデル	17
3.2 音 声 認 識	18
3.2.1 音声認識システムの概要	18
3.2.2 音声認識技術のマイルストーン	20
3.2.3 音声認識技術	23
3.3 機 械 翻 訳	43
3.3.1 人間の翻訳と機械の翻訳	43
3.3.2 機械翻訳の難しさ	45
3.3.3 翻訳システムの作り方	46

3.3.4 対訳データの収集・対応付け	47
3.3.5 フレーズベース翻訳	49
3.3.6 木に基づく翻訳	55
3.3.7 ニューラルネットに基づく機械翻訳	58
3.3.8 翻訳結果の評価	65
3.3.9 機械翻訳の現状と未解決問題	67
3.4 音 声 合 成	68
3.4.1 音声合成の歴史	69
3.4.2 テキスト音声合成の仕組み	70
3.4.3 統計的パラメトリック音声合成方式	77
3.4.4 非言語情報およびパラ言語情報の制御	86
3.4.5 合成音声の評価	90
3.4.6 音声合成の現状と今後の課題	91
引用・参考文献	93

第4章 音声翻訳の研究プロジェクトとシステム

4.1 ATR と NICT プロジェクト	105
4.1.1 ATR プロジェクト	105
4.1.2 NICT プロジェクト	127
4.2 世界のおもな音声翻訳プロジェクト	130
4.3 国際共同研究と音声翻訳標準化	135
4.3.1 C-STAR	135
4.3.2 A-STAR	136
4.3.3 IWSLT	136
4.3.4 国際標準化	138
4.3.5 U-STAR	140
引用・参考文献	140

第5章 音声同時通訳

5.1 同時通訳者の処理と認知モデル	144
5.2 コンピュータはいかに同時通訳者に迫るか	147
5.2.1 翻訳タイミングの決定	148
5.2.2 未発話内容の予測	152
5.2.3 表現の工夫	154
5.2.4 同時音声翻訳システムの評価	156
引用・参考文献	158

第6章 究極の音声翻訳

6.1 理想的な音声翻訳モデル	160
6.2 パラ言語音声翻訳	161
6.3 音声画像翻訳	163
6.4 speech-chain への挑戦	165
6.5 end-to-end 音声翻訳	166
6.6 音声翻訳の課題と今後	168
引用・参考文献	171

あ と が き	172
---------------	-----

索 引	176
-----------	-----

第1章

音声翻訳の概要

1.1 音声翻訳システム

音声翻訳システムの全体像を図 1.1 に示す。音声翻訳は、大まかに言って入力音声の言語情報を認識する多言語音声認識、多言語機械翻訳、そして翻訳されたテキストを読み上げる音声合成から構成される。また、原言語の入力音声に含まれる感情、発話スタイル、声質、韻律などは言語を変換しても目的言語で保持されるように保存、変換する必要がある。これらの処理は難易度が高いため、どのような話題について話しているかの話題知識（ドメイン知識とも言う）と、対話履歴や知識で内容を補って認識、翻訳を行っていく。現状では、文字化できる情報（言語情報）のみに注目し、話題を限定した状況で、多

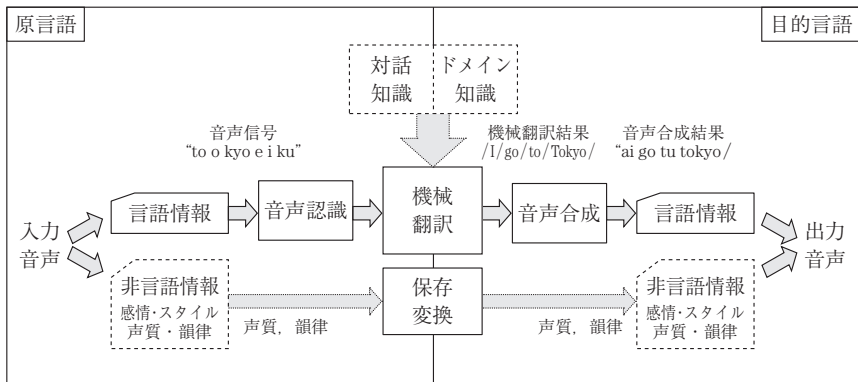


図 1.1 音声翻訳システムの構成図

2 1. 音声翻訳の概要

言語の音声認識、機械翻訳、音声合成と、それらを統合した音声翻訳が動作可能な状況になってきた。筆者の中村が研究を開始した1980年代前半は、音声認識で言えば500単語程度の特定話者の単語音声認識が動く程度のレベルであった。いまでは不特定話者大語彙の統計的多言語音声認識、多言語の機械翻訳、統計モデルによる多言語音声合成が実用可能になり、30年の間のインフラの進化を含めた飛躍的な技術の進化に感動を覚えざるを得ない。

一方で、未解決の課題も山積している。話し言葉である音声言語をいかに翻訳するか、人間の通訳のような五月雨式で内容を解した音声自動通訳をどう実現するか、文脈をどう考慮するか、照応、省略をどう補完するか、原言語で曖昧な表現を目的言語にそのまま訳せるか、文化・背景などの違いによって意味の異なる表現をどう訳すか、方言はどうするのか、イントネーションや声質、ジェスチャ、表情などはどう扱うのか、多くの課題が残されている。

1.2 自動音声翻訳のこれまで

わが国では、日本電気株式会社（以降 NEC と表記）が1983年の世界電気通信展示会で音声翻訳のデモンストレーションを行い世界の注目を集めた^{3)†}。その後、1986年にNTTの民営化に伴うNTT株の売却益をもとに基盤技術研究を行う民間基盤技術促進制度と基盤技術促進センターが発足し、産学官が共同で研究を行う体制が整備され、多くの基盤研究を行う研究所が発足した⁵⁾。日本人は外国語が苦手であることから、音声翻訳技術の研究開発がその一つに選ばれ、1986年に株式会社国際電気通信基礎技術研究所（エイ・ティ・アール、Advanced Telecommunication Research Institute International, ATR）と、実際の音声翻訳の研究開発を時限で行う会社である株式会社エイ・ティ・アール自動翻訳電話研究所が創設され、国内外の多様な音声言語研究者が参画した。1986年当時は不特定話者・連続音声の認識もまだ十分にできない状況であ

† 肩付数字は各章末の引用・参考文献番号を表す。

り、国際会議予約、ホテル予約、日常旅行会話に順次、研究開発の目標となる話題とタスクを拡張しながら研究を進めた。1993年には、自動翻訳電話研究所（日本）、カーネギーメロン大学（アメリカ、略称 CMU）、シーメンス社（ドイツ）の世界3地点を結んだ音声翻訳会話実験が行われた¹¹⁾。この実験はニューヨークタイムズでも大きく取り上げられた。これを機に国際的な音声翻訳研究コンソーシアム（Consortium for Speech Translation Advanced Research, C-STAR）が発足し、日本、アメリカ、ドイツ、（後に、イタリア、フランス、中国、韓国も参加）が音声翻訳の国際共同研究を開始した。ATRのプロジェクト開始の後、世界でもさまざまな音声翻訳のプロジェクトが開始された。対話音声翻訳を目指してドイツでは Verbmobil プロジェクト¹²⁾、欧州で NESPOLE¹³⁾が、アメリカでは TransTac プロジェクトが実施された¹⁾。また、講演などの音声翻訳を目指して欧州で TC-STAR プロジェクト⁶⁾、EU 会議の同時通訳を支援する EU-BRIDGE プロジェクトが実施された²⁾。アメリカでは、GALE プロジェクトが2006年から5年間実施された¹⁰⁾。GALE プロジェクトは、アラビア語、中国語から英語への自動翻訳を目指したプロジェクトであり、これまで人間が行っていた多言語重要情報の抽出、翻訳、要約作業の自動化を目的にしている。

自動音声翻訳の国際共同研究コンソーシアム C-STAR は、2004年からより学術的な研究を中心にした組織に変わり、IWSLT（International Workshop on Spoken Language Translation）という国際ワークショップとなった。この国際ワークショップは、音声翻訳に関する性能評価トラックを中心としており、参加者は規定の学習データを用いてシステムを構築し、テストデータとして IWSLT から供給されるデータの結果と論文を投稿する。会議はこれらのシステム論文と、一般の研究論文から構成される。

これらのプロジェクト等で研究開発された技術の実用化については、携帯電話の通信速度と帯域の向上により高度な計算がリモートのサーバ側で実施できるようになり、大量の音声・テキストデータの収集により、性能が飛躍的に向上し、スマートフォンを端末とした音声翻訳サービスの実用化、商用化が進ん

4 1. 音声翻訳の概要

である。日本においては、世界に先んじて ATR の技術を元にした携帯電話を端末とし、分散音声認識のフレームワークを利用した音声翻訳サービス「しゃべって翻訳」が 2007 年に商用実用化され普及した。さらに、スマートフォンを端末にした音声翻訳試験サービス VoiceTra を 2010 年にスタートさせ多くの利用者に使用していただいている^{7)~9)}。海外では、アメリカで Google 翻訳、さらに最近では Microsoft 社が無料通話アプリ Skype に音声翻訳を試験的に導入して話題となっている。音声翻訳の歴史を図 1.2 にまとめた。

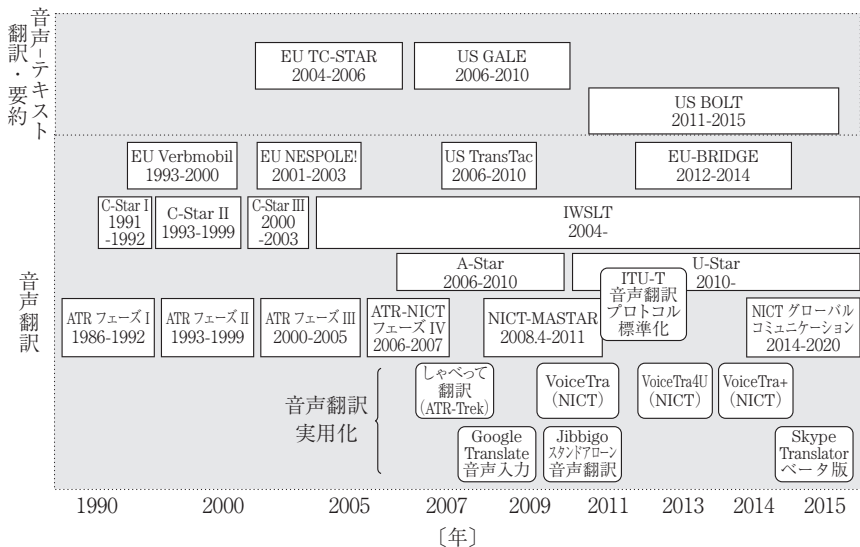


図 1.2 音声翻訳の歴史

任意の話題に対する自然な発話の音声翻訳は高度な課題であるため、特定の話題に音声翻訳の対象を絞り込むことにより音声認識や翻訳の精度を利用可能なレベルまで向上させ実用化が試みられている。対象とする会話は、会議予約、ホテル予約、旅行会話、多様な日常会話へと比較的容易な翻訳から順次高度な翻訳へ研究開発が進められてきた。現在は、5章で述べるように、講演やビジネス会話を対象に、「自動翻訳」から「同時自動通訳」への研究が始まっている。

引用・参考文献

- 1) N.Back, M.Eck, P.Charoenpornasawat, T.Köhler, S.Stüker, T.Nguyen, R.Hsiao, A.Waibel, S.Vogel, T.Schultz, and A.Black : The CMU TransTac 2007 Eyes-free and Hands-free Two-way Speech-to-Speech Translation System, Proceedings of IWSLT (2007)
- 2) M. Freitag, *et al.* : EU-BRIDGE MT : Combined Machine Translation, Association for Computational Linguistics (ACL), June (2014)
- 3) Y.Kato : The future of voice-processing technology in the world of computers and communications, Proceedings of Natl. Acad. Sci. USA, **92**, pp. 10060-10063 (1995)
- 4) A.Lavie, F.Metze, R.Cattoni, and E.Costantini : A Multi-perspective Evaluation of the NESPOLE! Speech-to-speech Translation System, Proceedings of the Workshop on Speech-to-Speech Translation : Algorithms and Systems, pp. 121-128 (2002)
- 5) T.Morimoto : Automatic Interpreting Telephony Research at ATR, Proceedings of a Workshop on Machine Translation, (1990)
- 6) D.Mostefa, O.Hamon, and K.Choukri : Evaluation of Automatic Speech Recognition and Speech Language Translation within TC-STAR : Results from the first evaluation campaign, Proceedings of LREC'06 (2006)
- 7) 中村 哲 : 音声翻訳技術の現状と今後の展開, 文部科学省科学技術政策研究所 科学技術動向研究センター 科学技術動向, **89**, pp. 8-19, (2008-08)
- 8) 中村 哲 : 話し言葉の音声翻訳技術, 電子情報通信学会誌, **96** 11, pp. 865-873 (2013)
- 9) 中村 哲 : 音声翻訳概観, 電子情報通信学会誌, **98**, 8, pp. 702-709 (2015)
- 10) H.Soltau, G.Saon, B.Kingsbury, H.Kwang, J.Kuo, and L.Mangu : Advances in Arabic Speech Transcription at IBM Under the DARPA GALE Program, IEEE Transactions on Audio, SPEECH, AND LANGUAGE PROCESSING, **17**, 5, pp. 884-894 (2009)
- 11) 谷戸文広, 竹沢寿幸, 嵯峨山茂樹, 鷹見淳一, H. Singer, 浦谷則好, 森元 逞, 樽松 明 : 自動翻訳電話国際共同実験, 電子情報通信学会技術研究報告, **93**, 87 (SP93 14-23), pp. 73-80 (1993)
- 12) W.Wahlster : Verbmobil : Foundations of Speech-to-Speech Translation, Springer (2000)

あ　と　が　き

音声翻訳の研究に出会って早 30 年になる。筆者の中村が音声認識の研究を始めたときは、音声の特徴量を線形予測の係数として抽出する手法、音声の発話ごとの声道の違いによるスペクトルの変動をマルチテンプレートで吸収し、時間変動を動的計画法で吸収する手法が主流で、それでも非常に巧妙であることに感動したものである。当時は日本の音声・言語処理技術は世界トップであった。しかし、その後、隠れマルコフモデルや N グラムのようなそれまでと異なる通信理論に基づく統計的なモデリング、さらには、深層学習による手法が登場し、まったく景色が変わってしまった。さらに、アメリカでは国防総省の DARPA プロジェクトのもと、共通タスクで性能競争するというフレームワークが始まり、必ずしも新しくない手法の組合せだが、少しずつ性能を向上させる、技術開発としての進め方を採用した。その結果、アメリカでは 1000 単語の不特定話者連続音声認識システム (Sphinx) を生み、種々のプロジェクトにより数々のシステムや人材を輩出した。これに続いて欧州でも、性能競争・評価型のプロジェクトとして TC-STAR や EU-BRIDGE などが実施された。それらに従事した学生、研究者が企業に参加し、スタートアップ企業を起業するなどして、現在の音声・言語処理サービスを提供している。日本においても ATR をはじめとする政府主導の研究開発プロジェクトが実施され、その成果として、携帯電話用のネットワーク型音声翻訳のサービスが実現された。ATR の成果についても ATR が関連するいくつかのスタートアップ企業を生み成果展開が図られている。

ATR プロジェクトには、日本以外にも海外からの研究者も多く関与し、その数は膨大で、世界の音声・言語分野のシニアな研究者はなんらかの意味で ATR に関わった経験があると言っても過言ではない。その意味で ATR は

DARPA 主流のアメリカ標準から見ても十分に存在感を發揮した。

研究にはフェーズがあり、例えば、基礎研究、基盤研究、システム研究、応用研究のフェーズがある。DARPA は研究の状況を観測し、このフェーズにあったプロジェクト支援の形式として、性能競争・評価型をとっていたのではないかと思う。その点で、現在の日本の大学や研究機関は、基礎研究にとどまる傾向があり、また、それをよしとする傾向があるが、フェーズのシフトに応じて、基盤研究、システム研究、応用研究に合わせた進め方があるのではないかと思う。ほとんどの重要な技術の基礎研究フェーズでは、論文発表で日本の大学が先行しているが、基盤研究、システム研究、応用研究では、新規性が小さいため大学は機能を發揮できず、後塵を拝している。技術を確立していくにはフェーズをシフトしていく仕組みと人材が必要である。

音声翻訳は人類の夢である言葉の壁を越える技術である。5章に述べたようにいまだ解決できていない課題が山積しているが、通訳者のように、あるいは母語で話したときのように異なる言語を話せるような技術を研究開発していく必要がある。一方、コミュニケーションのためだけでなく、アメリカの DARPA プロジェクトのように情報分析のために音声翻訳の研究開発を実施することも非常に重要である。すべての情報がインターネットにつながり、膨大な情報が利用可能な（ビッグデータ）時代にこのような情報分析は音声翻訳研究のパラダイムシフトとも言える。

1986年に株式会社エイ・ティ・アール自動翻訳電話研究所が樽松 明氏を社長として発足し、産官学から筆者の中村を含め多くの研究者が研究に参加した。音声情報処理では、カーネギーメロン大学滞在から NTT の鹿野清宏博士が戻り、室長として参加した。そして、カーネギーメロン大学や IBM で進んでいた隠れマルコフモデル、N グラム、MIT の Zue 博士らが進めていたスペクトログラムリーディングによる音声認識、ニューラルネットワークによる音声認識を立ち上げ、その後の日本の音声認識研究に貢献した。また、NTT から同様に参加した匂坂芳典博士も素片接続型音声合成をはじめ音声合成の研究に貢献した。ATR では同年、エイ・ティ・アール視聴覚機構研究所も発足し、

聴覚、音声生成の基礎研究が行われた。人間の機能に関する基礎研究と音声翻訳を目指した目的指向型研究の協力関係は相互に補完関係にあり非常に役立ったと感じている。

1987年には、現在、カーネギーメロン大学とカールスルーエ工科大学の教授をしている Alexander Waibel 博士が ATR に滞在された。Waibel 博士は、ニューラルネットワークで音声認識を行うという先験的な研究に挑戦され、時間遅れニューラルネットワークを提唱し後に IEEE 論文賞を受賞された。この時間遅れニューラルネットワークは、現在のたたみこみニューラルネットワークの原型である。

筆者の中村は Waibel 博士と、ほぼ同じ年代であった、武田一哉氏（現、名古屋大学教授）、阿部匡伸氏（現、岡山大学教授）らと研究や生活をともにした。当時、英語があまり得意でなかったが、日々の暮らして英語を使うことはコミュニケーション力と自信を付けてくれた。また、Waibel 博士やその後の多くの国際人脈は研究の実施をおおいに助けてくれた。このような国際的なヒューマンネットワークや、異文化交流が非常に役立っている。

2010年以降、そのニューラルネットワークが再来し、音声認識、機械翻訳のいずれにも大きな性能改善をもたらしている。音声認識では switchboard タスクという電話会話音声の認識で、人間が聞き取って書き起こした際の単語誤り率6%に匹敵する性能が自動音声認識により達成されている。膨大な学習データ、多数の隠れ層、いろいろな学習アルゴリズム、高速化アルゴリズム、高速計算ハードウェアの賜物である。

40歳代にいくつかの書籍の執筆に参加させていただいた。その際に感じたことは自らの能力が書籍を執筆するに値しないので、55歳になるまで止めようということだった。ところが、時間はあっという間に過ぎてしまい、55歳を回ってしまっていた。日本音響学会 音響サイエンスシリーズ委員長の平原先生から本書の執筆を引き受けてからも、奈良先端科学技術大学院大学の研究室の立ち上げ、音声自動通訳をはじめとする数々の新たな研究テーマの立ち上げ、研究資金獲得に奔走し、時間がとれず、予定を大幅に遅れてしまった。ま

た、執筆中に、准教授だった戸田博士が名古屋大学教授に、助教だった Neubig 博士がカーネギーメロン大学のテニユアトラック助教に転出されてしまった。本当に、人生のよい時期は一瞬だと痛感させられた瞬間であった。

最後に、音声や言葉は人間の認知プロセス、創造活動の核心である。その認知活動の核心を解き明かしていくのが言語研究の醍醐味であると思う。読者の皆様には新たな時代へ向けて、夢を信じて既成概念にとらわれない研究開発に打ち込んでいただきたいと思う。

2018年4月

中村 哲

索引

<hr/> あ <hr/>		
アクセント 14, 71	アクセント成分 74	アライメントアルゴリズム 42
<hr/> い <hr/>	意味的妥当性 66	イントネーション 12, 14
韻律生成 72	<hr/> え <hr/>	エイリアシング 27
エンコーダ 37	エンコーダ・デコーダ 63	<hr/> お <hr/>
オートエンコーダ 37	オビニオン評定 91	重み付き有限状態トランス デューサ 22
重みの調整 54	音韻生成 74	音響音韻論 20
音響管 21	音響モデリング 38	音響モデル 27, 30
音源 29	音声合成 68	音声特徴量抽出 79
音声認識 18	音声表情制御 88	音声翻訳 1
音素 18, 27	音素コンテキスト 19	音素体系 11
<hr/> か <hr/>	外言 12	階層的フレーズベース翻訳 57
ガウス混合モデル 22	書き言葉 6	確率的オートマトン 31
確率的勾配変分ベイズ 38	隠れマルコフネットワーク モデル 107	隠れマルコフモデル 21, 30
カットオフ周波数 27	可変長素片接続型音声合成 ネットワーク 111	完全結合型深層ニューラル ネットワーク 38
観測空間 24	<hr/> き <hr/>	機械翻訳 43
記号間翻訳 8	木構造 33	木構造クラスタリング 114
規則ベース音声合成方式 69	木に基づく翻訳 55	基本周波数パターン 73
客観評価尺度 91	逆フーリエ変換 29	共起回数 48
教師あり学習 37	教師なし学習 36	共振付与部 69
強勢 19	共分散行列 32	共鳴 21
<hr/> く <hr/>	クラウドソーシング 66	<hr/> け <hr/>
継続長モデリング 81	形態素解析 70, 72	決定空間 24
ケプストラム解析 28	ケフレンシ 30	言語間翻訳 8
言語情報 14	言語内翻訳 8	言語モデル 27, 34, 50
<hr/> こ <hr/>	語彙化翻訳確率 52	語彙選択誤り 45
構文解析 56, 71	構文木 55	構文情報に基づく翻訳 56
合文法無意味文 91	声を操る 88	声を混ぜる 88
声を真似る 88	個体内コミュニケーション 12	コネクショニスト時系列 分類 40
コーパスベース音声合成 69	コミュニケーションの 数学的モデル 13	コムニカチオ 10
コンテキスト 19	コンテキスト クラスタリング 82	コンテキスト情報 70
<hr/> さ <hr/>	最大事後確率 35	雑音源 13
雑音のある通信路 13	サブワード 27, 33	サポートベクトルマシン 72
残響環境 20	三者二言語コミュニケー ションシステムモデル 9	サンプリング 27, 38
<hr/> し <hr/>	時間遅れニューラルネット ワーク 37, 108	シグモイド活性化関数 40

事後確率 25
 事前学習 38
 事前確率 34
 事前並べ替え 58
 自動音声翻訳 10
 自動評価 66
 主観評価尺度 91
 主語補完 113
 主辞駆動句構造文法 106
 受信機 13
 準定常 28
 条件付き確率場 72
 状態継続長 82
 状態系列確率 82
 状態遷移確率 82
 深層学習 22, 36
 深層識別学習 37
 深層信念ネットワーク 37
 深層双方向 LSTM リカレン
 トニューラルネットワーク 40
 深層ニューラルネットワーク 37, 86
 深層ボルツマンマシン 37

す

数字認識 20
 数量化 I 類 72
 スコア関数 54
 素性関数 50
 素性構造伝搬パーザ 111
 ストレス 19
 スペクトル包絡線 30

せ

声質変換 87
 正則化オートエンコーダ 37
 静的・動的特徴量 84
 声道フィルタ 29
 制約付きボルツマンマシン 37
 線形予測係数 28

そ

双方向リカレントニューラル
 ネットワーク 40
 ソースフィルタモデル 69, 79
 ソフトマックス層 38
 素片選択型音声合成 69

た

大語彙連続音声認識 32
 対比較 66
 対話翻訳システム 110
 多空間確率分布 HMM 80
 多層パーセプトロン 22
 たたみこみ 29
 たたみこみ型 38
 たたみこみニューラルネット
 ワーク 37, 109
 探索 22
 探索アルゴリズム 35

ち

知覚線形予測 28
 逐次状態分割 107, 114
 知識ベース 21
 注意型ニューラルネットに
 基づく翻訳 64

て

テキスト音声合成評価会 90
 敵対的学習 86
 デコーダ 37
 デコーディング 53
 データに基づく機械翻訳手法 47
 デノイズングオート
 エンコーダ 37
 伝達特性 20

と

投影層 39
 統計的機械翻訳 47
 統計的パラメトリック音声合
 成方式 69, 77
 統計的フレーズベース
 機械翻訳 120
 統計翻訳 115
 統計モデルフレームワーク 21
 動的プログラミング 21
 特徴抽出 26, 27
 特徴ベクトル 28
 トライグラム 34

な

ナイキスト基準 27
 内言 12
 並べ替え誤り 45

並べ替えモデル 50, 52
 喃語 11

に

日英独音声言語翻訳実験
 システム 106
 ニューラルネットに基づく
 機械翻訳 58
 ニューラルネットワーク 22
 認識／探索アルゴリズム 27

は

バイグラム 34
 背景ノイズ 20
 ハイブリッド深層ネット
 ワーク 37
 波形合成 76
 波形素片接続の音声合成 113
 パターン認識 23
 発音辞書 27, 33
 バックプロパゲーション 40
 発話者間の異なり 19
 発話スタイル 19
 発話速度 19
 話し言葉 6
 パラ言語 9
 パラ言語情報 14, 86
 パラメータ生成 84

ひ

非言語情報 12, 14, 86
 非線形モデル 72
 ビットレート 28
 非定常過程 28
 非定常雑音抑圧フィルタ 120
 人手規則に基づく翻訳 46
 人手による評価 65
 ビーム探索 54

ふ

ファジィベクトル量子化 107
 フィードフォワード型 39
 フォルマント 21
 フォルマント周波数 19, 21
 藤崎モデル 73
 部分翻訳 113
 フーリエ変換 29
 フレーズ 18
 フレーズ境界 71
 フレーズ成分 74
 フレーズベース翻訳 49

フレーム	28
フレーム分析	79
分散型音声認識	126
<hr/>	
へ	
平均声モデル	88
ベイジアンネットワーク	121
バイズ則	36
ベクトル量子化	32
変換主導翻訳	113, 115
変調フィルタスペクトロ グラム	28
変分バイズオートエンコーダ	37
<hr/>	
ほ	
ボイスバンク	92
母語	11
ボコーダ	69
ボーダ	69
翻訳結果の探索	53
翻訳の適切さ	132
翻訳の流ちょうさ	132
翻訳編集率	67
翻訳モデル	50

<hr/>	
ま	
マイクロフォンアレー	120
マルコフ連鎖	21
マルコフ連鎖モンテカルロ	38
<hr/>	
む	
無声音	79
<hr/>	
め	
メル周波数ケプストラム係数	28
<hr/>	
ゆ	
有限状態オートマトン	22
有声音	79
<hr/>	
よ	
用例主導翻訳	113
用例ベース機械翻訳	115
<hr/>	
ら	
ランダムプロセス	38

<hr/>	
り	
リカレントニューラルネット	62
リカレントニューラルネット ワーク	22, 37
リカレントニューラルネット ワーク言語モデル	40
リフター	29
流ちょう性	66
量子化	27
履歴ベクトル系列	40
臨界制動応答	74
<hr/>	
れ	
励振源生成部	69
連続HMM	80
連続音声認識	21
連続ガウス混合モデル	32
<hr/>	
ろ	
ローパスフィルタ	27
<hr/>	
わ	
話者内の異なり	19

<hr/>	
A	
adequacy	132
<hr/>	
B	
Bakis モデル	31
BLEU	66
<hr/>	
C	
CLDNN	38
CNN	37
connectionist temporal classification	23, 40
CTC	23
<hr/>	
D	
decoding	22
DFT	29
DNN	37, 86
DNN-HMM	38
<hr/>	
E	
end-to-end	38, 40, 41

<hr/>	
F	
fine tuning	38
fluency	132
<hr/>	
G	
GMM	32
GMM-HMM	22
<hr/>	
H	
HMM-LR	110
HMM 音声合成	77
HPSG	106, 111
<hr/>	
I	
IWSLT	68
<hr/>	
K	
Kirchhoff	9
<hr/>	
L	
Levenstein アルゴリズム	42
listen, attend and spell	41

long short term memory	22, 38
LPC	28
LR パーザ	106
LR 文法	110
LSTM	22, 38, 86
LVCSR	32
<hr/>	
M	
METEOR	67
MFCC	28
MSG	28
<hr/>	
N	
NIST スコア	67
noise source	13
noisy channel	13
N グラム言語モデル	39, 59
<hr/>	
P	
PLP	28
pretraining	38

R		T	V
RBM	38	Seleskovitch	8
receiver	13	Shannon	13
restricted boltzman machine	38	SSS	107
RIBES	67	T	
RNN	22	TDMT	113
RNNLM	40	TDNN	37, 108
S		TOEIC 音声翻訳評価法	117
SampleRNN	92	TOEIC 換算値による翻訳出力文の評価法	66
		tree-to-string 翻訳	56
		W	
		WaveNet	92
		Weaver	13
		WFST	22
		WMT	68
		Vauquois の三角形	44
		Voice Conversion Challenge	90
		v-Talk	111

— 編著者・著者略歴 —

中村 哲 (なかむら さとし)

- 1981年 京都工芸繊維大学工芸学部電子工学科卒業
1981年 シャープ株式会社勤務 (研究員)
1986年 株式会社エイ・ティ・アール自動翻訳電話研究所勤務
1992年 博士 (工学) (京都大学)
1994年 奈良先端科学技術大学院大学助教授
2000年 株式会社国際電気通信基礎技術研究所
音声言語コミュニケーション研究所室長, 所長
2003年 カールスルーエ工科大学客員教授
2006年 情報通信研究機構グループリーダー,
上席研究員, センター長, けいはんな研究所長
2007年 ATR フェロー
2011年 奈良先端科学技術大学院大学教授
2016年 IEEE フェロー
2017年 奈良先端科学技術大学院大学データ駆動型サイエンス創造センターセンター長
2018年 奈良先端科学技術研究科情報科学領域教授 (兼務)
現在に至る

Sakriani Sakti (サクリアニ サクティ)

- 1999年 バンドン工科大学情報学科卒業
2000年 Sumarno Pabotingi Associate
Junior IT Consultant
2001年 ダイムラー・クライスラー株式会
社勤務
2002年 ウルム大学大学院修士課程 (通信
技術) 修了
2003年 株式会社国際電気通信基礎技術研
究所勤務
2006年 情報通信研究機構
2008年 ウルム大学大学院博士課程 (工
学) 修了
Dr.Ing.
2011年 奈良先端科学技術大学院大学助教
2018年 奈良先端科学技術大学院大学特任
准教授
理化学研究所革新革新知能統合研
究センター研究員兼任
現在に至る

Graham Neubig (グラム ニュービグ)

- 2005年 イリノイ大学工学部コンピュー
ターサイエンス学科卒業
2005年 兵庫県立但馬農業高等学校勤務
2006年 兵庫県庁勤務
2008年
2010年 京都大学大学院修士課程修了 (情
報学専攻)
2012年 博士 (情報学) (京都大学)
2012年 奈良先端科学技術大学院大学助教
2016年 カーネギー・メロン大学助教
現在に至る

戸田 智基 (とだ ともき)

- 1999年 名古屋大学工学部電気電子・情報工学科卒業
- 2001年 奈良先端科学技術大学院大学情報科学研究科博士前期課程修了(情報処理学専攻)
- 2001年～06年 株式会社国際電気通信基礎技術研究所 研究員
- 2003年 奈良先端科学技術大学院大学 情報科学研究科 博士後期課程修了(情報処理学専攻) 博士(工学)
- 2003年～05年 日本学術振興会特別研究員-PD (名古屋工業大学大学院工学研究科)
- 2005年 奈良先端科学技術大学院大学助手
- 2007年 奈良先端科学技術大学院大学助教
- 2011年 奈良先端科学技術大学院大学准教授
- 2015年 名古屋大学教授
現在に至る

高道 慎之介 (たかみち しんのすけ)

- 2009年 熊本電波工業高等専門学校電子工学科卒業
- 2011年 長岡技術科学大学工学部電気電子情報工学課程卒業
- 2013年 奈良先端科学技術大学院大学情報科学研究科博士前期課程修了(情報処理学専攻) 情報通信研究機構短期間研究員
- 2014年 カーネギー・メロン大学言語技術研究所客員研究員 日本学術振興会特別研究員(DC2)
- 2016年 奈良先端科学技術大学院大学情報科学研究科博士後期過程修了(情報処理学専攻) 博士(工学)
- 2016年 東京大学大学院特任助教
現在に至る

音声言語の自動翻訳——コンピュータによる自動翻訳を目指して——
Introduction of Speech-to-speech Translation

© 一般社団法人 日本音響学会 2018

2018年7月10日 初版第1刷発行

検印省略

編 者 一般社団法人 日本音響学会
発 行 者 株式会社 コロナ社
代 表 者 牛来真也
印刷所 萩原印刷株式会社
製本所 有限会社 愛千製本所

112-0011 東京都文京区千石 4-46-10

発行所 株式会社 コロナ社

CORONA PUBLISHING CO., LTD.

Tokyo Japan

振替 00140-8-14844・電話 (03)3941-3131(代)

ホームページ <http://www.coronasha.co.jp>

ISBN 978-4-339-01338-2 C3355 Printed in Japan

(宝田)



本書のコピー、スキャン、デジタル化等の無断複製・転載は著作権法上の例外を除き禁じられています。
購入者以外の第三者による本書の電子データ化及び電子書籍化は、いかなる場合も認めていません。
落丁・乱丁はお取替えいたします。