3E1-07

音声言語処理技術を用いたテレビ放送やビデオ映像からの語学学習教材の半自動作成システム

Semi-automatic Construction of a CALL system based on TV news programs and Video using spoken Language Processing Technorogy

中川 聖一

Seiichi NAKAGAWA

豊橋技術科学大学 情報工学部

Toyohashi University of Technology, Department of Information and Computer Sciences

In this paper, we report studies on CALL in the project of Scientific Research of Priority Areas “Utilization of Multimedia for Education". Especially, we report results of learning systems for pronunciation, listening, conversation, reading and writing. Finally, we also describe the English learner's speech database and Japanese learner's speech database.

１． はじめに

平成 11 年度より科学研究費補助金特定領域研究（A)「高等教育改革に資するマルチメディアの高度度利用に関する研究(略称：メディア教育利用)」(研究代表者：坂本昴メディア教育開発センター所長）が発足し、平成 12 年度より、3 つの研究項目（A01：高等教育におけるマルチメディア利用の高度化の研究，A02：外国語教育の高度化の研究，A03：メディア教育・情報教育の高度化の研究）に関して、7 班の計画研究と 43 件の公募研究（平成 13 年度は 58 件）が開始され、平成 15 年 3 月に終了した。

きたるべき高度情報通信社会、国際化社会に備えて、マルチメディア・情報通信を活用した教育改革、外国語教育・情報教育などにおける標準カリキュラム・教材の開発と実践研究の必要性が、様々な方面において指摘されている。本特定研究は、こうした要請に応え、21 世紀の近未来社会で期待される専門的人材養成の在り方を展望すると共にマルチメディア・情報通信を活用した効果的な教育システムや形態を研究開発するために計画された。外国語（留学生に対する日本語を含む）コミュニケーション能力や情報活用能力・メディアリテラシー等、基礎能力の教育改善を図ると共に、人文・社会科学各分野における、マルチメディア・情報通信を活用した専門教育、外国語教育、情報教育の在り方、教育リソースの開発・運用に関して研究を行った。

本研究は、文理両領域にまたがる文理融合研究を目指した教育工学分野では初めての大型プロジェクトであった。

中川聖一（豊橋技術科学大学),
〒441-8580 愛知県豊橋市天伯町雲雀ヶ丘 1-1,
nakagawa@slp.ics.tut.ac.jp

本特定研究に関係する研究者（全体で約 200 名）と研究予算（3 年間で約 9 億 5 千万円）の約半数・半額がコンピュータ支援型語学教育(CALL:Computer Assisted Language Learning)に関するもので、その大半が音声言語情報処理技術を用いた研究であった。

２．外国語教育の高度化の研究

２.１ 研究の特徴

(１）音声言語情報処理技術の利用

本特定研究を最初に計画したのは平成 9 年夏頃であり(平成 9 年 11 月申請は不採択、平成 10 年 11 月申請が採択)、それ以降において、特に音声認識機能付きの CAI ソフトウェアが数多く市販されるようになった。そのほとんどを調査したが、市販の語学 CAI ソフトには海外でのロケも含めて教材の作成には費用と労力をかけた優れたものも多くあるが、以下の欠点がある［4］。

（a）音声認識機能を用いたものも存在するが、発声が不完全な学習者用のモデルを用いたものは少ない。

（b）アクセント・イントネーション・発音・文法等の自動評価・診断するものはない（あっても信頼性に欠ける)。

（c）視覚的フィードバックが少ない（発声の構音状態、発声内容の構文木など)。

（d）個々の教材は良くできているが、教師が容易にマルチメディア教材を作成できるオーサリングツールが不充分である。

（e）日本語 CAI のソフトが少ない。

本研究ではこれらの欠点を解消する語学 CAI システムの構築を目指した。

(２）研究体制

上記の研究を達成するために、音声情報処理研究者を

中心に、語学・言語学研究者・日本語教育者・英語教育者・自然言語処理研究者・教育工学研究者で研究組織を構成した。なお、計画研究では重複する研究内容、共用できる要素技術があるため、調整班（代表者：中川）を設けた。

（3）実用的研究

本研究は、基礎研究の上に立っていることは言うまでもないが、従来の特定研究よりも応用指向であり、実際の現場で使用できるソフトウェアの開発を目指した。

3.2 研究成果〔2〕〔3〕

（1）発音・韻律学習システム

今までに開発された発音・韻律学習システムの基本構成を図１に示す。この図は、発声例文とそれに対応した教師音声が与えられた場合の構成の例を示している。発声例文に基づく学習者音声からは、認識用パラメータ(メル周波数ケプストラム係数（MFCC）等）や基本周波数 F0、パワーが抽出される。一方発声例文は標準的な発声に対応する音素列に変換された後、非母国語話者固有の発声誤りの規則を適用して、非母国語話者の発声の変動に対応する複数の音素列を生成する。複数の音素列は、母国語話者音素 HMM と非母国語話者音素 HMM を用いて対応する音素 HMM 列に変換される。各音素 HMM 列と学習者音声からの特徴とが強制アライメント（Forced Alignment）される。複数の音素 HMM 列の中から最大尤度の音素 HMM 列が出力され、それに基づき学習音声が対応する音素区間に分割される。一方、教師音声も同様な手順で音素区間に分割される。教師音声は標準的な発声と想定しているので発声規則は通常設けない。両者の区分された結果を用いることによって、音素単位、音節単位、単語単位、フレーズ単位、文単位で発音や韻律の比較が可能になる。区分された結果を直接比較する場合もあるし、教師音声を基に統計的モデルを作成し、それを用いて比較する場合もある。従来の研究では，F0 やパワーパタンを区分せずに直接比較することが多かったが，このように分節単位に分割することによって、より精度が高まることが報告されている〔5〕。以下では図 1 の各部について説明を加える。

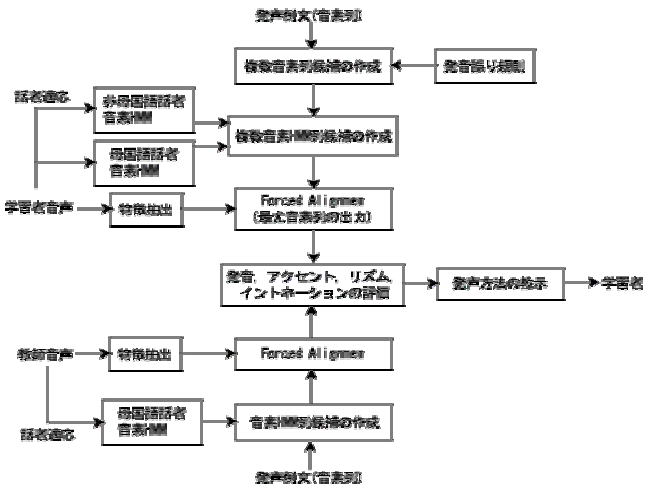

図１　発音・韻律学習システムの基本構成

Forced Alignment を行うには、音素や音節等の HMM を用意する必要がある。母国語話者と非母国語話者両方の HMM を使用する方が精度向上することが報告されている〔5〕。発音・韻律学習システムでは，学習者が非母国語を正しく発音することは期待できないため，通常は話者独立型の音素 HMM を利用することが多い。しかし、音声認識と同様に話者適応を用いることができれば，精度向上が期待できる。牧野らは日本人話者の発声した日本語発声で英語音素 HMM を話者適応できることを示している〔6〕。

発音誤り規則と Forced Alignment を利用することによって、発音誤りを検出することができる。誤りのかなりの部分は母語話者の発声に対しては誤りと判断しなかった過度部を母音挿入や子音置換と判断する誤りである。発音がどの程度正しいかを示すにはホルマント周波数や尤度が用いられる。ホルマント周波数は、舌の前後、高低とも対応しているため、学習者の母音発音を、口腔内の舌の位置に変換することも行われており〔7〕、これによって学習者は自分の舌の位置を知ることができ、また標準的な舌の位置と比較することによって、正しい発音をどのように行うべきかを教示することもできる。子音に対しては，学習者音声から舌の位置や動きを抽出することは困難なため、標準的な舌の位置や動きを図や動画で示すことが行われている。

中川らは、英単語の発音の評価法として、尤度、話速などを重回帰モデルで組合せたものと、英語教師による採点との相関を統計的に求め、相関が高くなるようモデルを学習した〔18〕。ここで、英語教師の評価は１単語毎の評価は難しいため、５単語につき１つの評価とし、この５単語をすべて同一の評価値とした。採用した音響的な物理尺度を以下に示す。

・対数尤度（ネイティブな英語音素 HMM および日本人適応化英語音素 HMM）
・最適な音素列とのマッチング結果の尤度
・音素認識結果の正解率
・話速

学習者 14 人が 15 単語を発声したものを評価資料とし、日本人英語教師２名、ネイティブ英語教師２名が、学習者の５単語を一組として発音の評点をつけた。日本人英語教師間のスコアの相関は 0.516、ネイティブ英語教師間のスコアの相関が 0.730、日本人英語教師とネイティブ英語教師間のスコアの相関が約 0.65 であった。重回帰分析に用いなかった語彙（語彙オープン）に対して自動評価と教師の評価との相関値は約 0.81（５単語ごとの評価）、重回帰分析に用いなかった話者(話者オープン)に対して約 0.69 と高い相関値が得られ、本手法が英語教師と同等の発音評価ができることがわかった。

（2）リスニング・会話学習システム

有木らは「発音評価機能とリスニング機能をもつ英語学習支援システム」の開発を行っている〔8〕。自然な英会話の習得を目指し、実際の会話の場面をコンピュータ上に設定した上で、学習者の発話音声を自動的に分析・認識・評価すると共に学習者に飽きさせないことも研究目的にしている。この目的のために生きた英語、リアリティを持った実際に使われている英語として、映画を教材とした CALL システムの開発を試みている。学習者の発声に対する英語らしさの発音評価機能や単語発声誤りの指摘機能に加えて、学習者のリスニング能力養成の機能も備えている。映画のクローズドキャプションをもとに Forced Alignment を行ない、音声を単語単位に分割し、モデルと学習者の音声を文単位・単語単位で聞き比べられる機能が特徴の一つである。

中川らも、映画の代わりにテレビニュース放送を利用したCALL 用リスニング教材の開発を行っている(図 2)。教師や学習者がニュース放送を用いて語学教材を作成しようとすると、絶えずコンテンツの更新を行わねばならず、非常に大きな労力と負荷を強いることになっていた。そこで、字幕と副音声付きテレビニュース放送を利用して、教材を半自動的に作成できる教材作成システムが開発された〔9〕〔10〕。主な機能は、ビデオの再生、ビデオと字幕の同期再生、字幕と翻訳音声（副音声）の同期再生、辞書引き、ディクテーションの問題の自動作成、アナウンサー音声と学習者音声の比較、内容把握問題作成支援、などである。

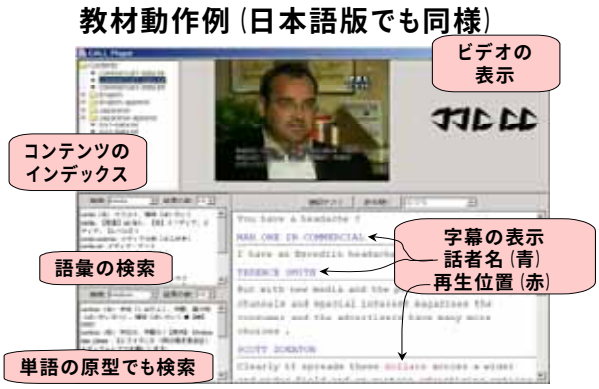

図 2　ニュース放送からの語学学習教材の自動作成

壇辻らの CALL システムでは、日本人英語学習者の発音の多様性や入力内容に柔軟に対応した精度の高い音声認識技術を導入している。その認識結果を基に学習者の誤りに対して教示を行ない、人間とコンピュータ間の音声対話学習を実現している点が特徴である〔11〕。音声認識部には、日本人英語音声を学習に用いた「日本人英語音響モデル」及び「単語発音辞書」を備えている。この特徴によって、日本人学習者の発音誤りや発音傾向に対応した認識が可能になった。その結果、入力内容の理解に重点を置いた音声認識が実現され、本来の目標である基礎表現とスキルの習得に重点を置いたシステムの構築が可能となった。この CALL システムではロールプレイ学習と対話学習を行うことができるが、ロールプレイ学習で収録された音声は対話学習時の音響モデルの話者適応に利用されるという無駄のない設計になっている。

（3）読解・作文支援システム

竹内らはテキスト文を入力するだけで学習者の理解状態に適応した質問応答をおこなう学習機能を実現している〔12〕。まず、英文を解析し構文・意味情報を生成する自然言語理解機能とその構文・意味情報からさまざまな種類の質問文を生成する質問文自動生成機能の設計・実現を行った。この質問文自動生成機能を利用して自然言語理解モジュールが抽出したテキスト情報から質問文および正誤判定のための情報を生成する。次に、生成した質問の難易度を算出し、難易度、教授知識，教科書情報および学習者モデルを用いて生成した質問文の中から目的に合致した適切な難しさの質問文を選択する。その後、自然な発話を行うために必要な音声規則を挿入し、質問を音声で出題する。学習者の入力した解答は、自然言語理解モジュールによって解析され、構文・意味情報が抽出される。ここで、構文的誤り原因の同定が行われると同時に、抽出された意味情報と質問文の元になった英文の意味情報との意味的比較が行われる。もし学習者が誤った入力をした場合は、学習者に熟考を促すように支援する。その際に、質問の難易度を用いて、より簡単な質問を選択・出題する。中学校の英語教科書の英文を用いた評価実験では、人間の判断と似た複雑度を持つ質問を自動生成できている〔13〕。

仁科・奥村らは、2001 年 3 月から、留学生のための多言語対応日本語読解支援システム「あすなろ」をウェブ上で一般公開している〔14〕。「あすなろ」の開発の主たる目的は、日本に留学している特に理工系分野を専門としている日本語学習者の専門文献読解を支援することである〔15〕。「あすなろ」は、学習者が入力した日本語の文章に対して、文章中の単語の訳語と構文構造を出力することを主な機能とし、日本語の構造や初歩的な単語の意味を知りたい日本語初級者から、単に技術用語の意味を知りたい上級者までが利用できる。また、現在の日本語学習システムおよび辞書の殆どは英語訳を導くものであるが、英語以外にも中国語、タイ語、マレー語、インドネシア語の４ヶ国語による訳の表示が可能である。

3．外国語学習用音声データベース〔16〕

3.1　経緯

90 年代より音声認識研究者を中心として各種の音声 DB が構築・整備されてきたが、非母語音声 DB は非常に少ない．また、外国語学習支援を目的とした場合、学習者音声データだけでなく、学習対象言語を教える教師による評定結果や、対象言語を母語とする話者による（同一読み上げセットの）音声などの整備も要望されていた。本特定領域研究の調整性が中心となって、第 2 言語学習環境を支援するための音声 DB 委員会を組織した。

3.2　日本語母語話者による英語読み上げ音声 DB

音素記号としては、TIMIT　DB の音素体系、CMU 発音辞書の音素体系をベースとし、若干の修正を施したものを使用した。音素バランスに着眼して作成したリストを表 1 に示す。

表 1　音素バランスを考慮した単語・文セット

カテゴリ	サイズ
音素バランス単語	300
ミニマル単語対	600
音素バランス単語	460
発音困難な音素列と含む文	32
音素学習に対する評価文	100

表 2 韻律バリエーションを考慮した単語・文セット

カテゴリ	サイズ
種々の強制パターンを含む単語	109
種々のイントネーションパターンを含む文	94
種々のリズムパターンを含む文	120

表 2 に韻律バリエーションに着眼して作成した発声リストを示す。18 大学・1 高専の協力の下、男性 100 人、女性 101 人の音声を収録した。なお、全文から 8 サブセット（約 120 文）全単語から 5 サブセット（約 220 単語）を構成し、1 サブセットずつを 1 人当たりの発声量とした。その結果、各文は約 12 人の話者によって、各単語は約 20 人の話者によって読み上げられている。母語話者による米語読み上げ音声 DB として男性 8 名、女性 12 名の音声を収録した。

4.3 非日本語母語話者による日本語読み上げ音 DB

日本語発話用のテキストは 4 種類使用した。その内 1 リストは本 DB 構築のために新たに作成した。
音素バランスに着眼したものとしては ATR 読み上げ文 503 文を利用した.予想される収録対象者がアジア圏中心であることから、あらかじめアジア圏の言語を母語とする話者にとって難しいと考えられる音素のミニマルペアを 115 語抽出した。
(a) 難音ミニマルペア 115 促音、長母音、撥音、濁音、清音など難音と思われる語を含むミニマルペアを抽出した。それらの語をランダムに並べ替え、発話者にはペアの対応が分からないようにし、すべての語に仮名を振った。
(b) 難音ミニマルペアを含むオリジナル文 108 文(1)のミニマルペアは単語で発話されるが、各々の単語が文脈の中で連続性を持って発話される場合の差異の観察などを可能にするために、難音ミニマルペアを含む文をオリジナルに作成した。また、下記の(1)～(9)の韻律項目を評価するために 44 文からなる対話文が作成された。(1)Yes・No 疑問文、(2)疑問詞疑問文、(3)疑問詞が文中にある場合、問い返し疑問(4)「何か」と「何も」、(5)右枝分かれ構造、(6)左枝分かれ構造、(7)対比の強調、(8)終助詞、(9)フィラー。8 大学の協力により、男性 71 人、女性 70 人（共に留学生）計 141 人を選び、その音声を収録した。各話者は ATR のサブセット 2 セット、ミニマルペア単語 115 語、雑音文 A 又は B セット 54 文、韻律文 44 文を読み上げた。日本語母語話者については、大学生・大学院生を中心に男 20 名、女 21 名の発話を収録した。

4. むすび

音声認識技術やコーパスに基づく自然言語処理技術の発展は目覚ましく、その有力な応用分野として語学学習システムがある。言語の壁を破るためには、機械翻訳技術や自動通訳技術の進展が望まれるが、限界がある以上、語学システムは今後とも重要な技術であり、更なる研究が望まれる。計算機援用語学学習システム（CALL）から Web 上のコンテンツを学習教材として取り込んだり、ネットワーク技術と連動した遠隔学習を可能とする Web 強調語学学習システム（WELL）が今後ますます有用となってこよう。

文　献

〔1〕中川聖一：科学研究費特定研究(A)「メディア教育利用」－音声言語処理技術を用いた語学 CAI－，日本音響学会誌，Vol.56，No.11，pp.767-770(2000)
〔2〕中川聖一，牧野正三，壇辻正剛：音声言語処理技術を用いた語学学習システム，日本音響学会誌，Vol.59，No.6，pp.337-344(2003)
〔3〕特定領域研究「メディア教育利用」報告書，平成 13 年 3 月，平成 14 年 3 月，平成 15 年 3 月，http://resource01.nime.ac.jp
〔4〕http://www.slp.ics.tut.ac.jp
〔5〕前田直子，山下洋一：日本人学習者のための英単語発音評定．音講論集，pp.105-106(2001．10)．
〔6〕長沢忠郎，鈴木基之，牧野正三：日本語と英語の音韻間距離の分析について．音講論集，　pp．345-346(2001．10）特定領域研究(A)成果報告書
〔7〕Tsubota，M．Dantsuji and T．Kawahara，Computer-assisted English vowel learning system for Japanese speakers using cross language fomant structuers．Proc．ICSLP 2000，Vol.3，pp.566-569 (2000)
〔8〕五十里慎吾，佐野輝希，緒方淳，有木康雄：ユーザー発話のセグメンテーションと発話評価機能をもつ英語学習システム，　情報処理学会，音声言語情報処理，SLP40-2(2002.2)
〔9〕田中敬志，小林聡，中川聖一：字幕・副音声付きテレビニュース放送を利用可能な語学学習教材作成システムとリスニング教材プレイヤー，日本教育工学会誌，Vol.27，No.3，pp.273-282（2003）
〔10〕小林聡，田中敬志，森一将，中川聖一：　字幕付きテレビニュース放送を素材とした語学学習教材作成システム，人工知能学会論文誌 17 巻 4 号 SP-G，pp.500－509(2002)
〔11〕阿部一彦，田中和世，河原達也，清水政明，壇辻正剛：対話型英語学習システムにおける日本人英語音声認識精度の検討，日本音響学会講演論文集，2-5-20，pp.113-114(2002)
〔12〕國近秀信，花多山知希，平嶋宗，竹内章：英語長文読解学習のための質問文自動生成機能の実現とその評”，電子情報通信学会論文誌，Vol．J83-D-1，No.6，pp.702-709(2000)
〔13〕國近秀信，宇留島稔，平嶋宗，竹内章：英語長文読解学習のための質問の複雑さの定義とその評価，人工知能学会論文誌，Vol.17，No.4，pp.521-529(2002)
〔14〕http://hinoki.ryu.titech.ac.jp
〔15〕仁科喜久子，奥村学，八木豊，戸沢徳久，澤谷孝志，傅亮，杉本茂樹，阿辺川武：構文表示と多言語インターフェースを備えた日本語読解学習支援システムの開発，言語処理学会第 8 回年次大会論文集，pp.228-231(2002)
〔16〕峯松信明，仁科喜久子，中川聖一：外国語学習読み上げ音声データベース，日本音響学会誌，Vol.59，No.6，pp.345-350(2003)
〔17〕中川聖一:語学学習における音声言語処理技術の利用，電子情報通信学会誌，Vol.85，No.12，pp.942-943(2002)
〔18〕S．Nakagawa，N．Nakamura，K．Mori，“A statistical method of evaluating pronunciation proficiency for English words spoken by Japanese”，IEICE Trans．Vol.E87-D，to appear