

音声認識とマルチメディア技術を統合した 日本語音声教育システム

三輪 譲二、山本 真人、平野 崇、佐々木 優、田 嘉鵬

miwa, masato, taka, masaru, tian@cis.iwate-u.ac.jp

岩手大学 工学部 情報工学科

〒 020-8551 岩手県盛岡市上田 4 - 3 - 5

和文抄録 マルチメディアパソコンとインターネットを用いて、いつでも、どこでも、だれでもが日本語音声教育を可能とすることを目指して、音声認識とマルチメディア技術を統合したオンデマンド・ネットワーク型日本語音声教育システム (LESSON/J: Japanese Language Education System for Speech using on an On-demand Network) を構築した。

このシステムでは、(1) オンラインで、日本語の単音節、単語、文章、アクセント等の音声をスピーカから聞くことができ、また、(2) Java アプレットにより、独習者向きに単語音声の聴取結果から書取テストをするディクテーション機能も実現しており、さらに、(3) 音声認識とマルチメディア技術を統合して、独習者の発声結果を評価し、フィードバックできる機能を持っている等の特色を持っている。

キーワード CALL、音声言語教育、日本語音声、マルチメディア、音声認識、オンデマンド

Language Education System for Japanese Speech Using Integrated Technique of Speech Recognition and Multimedia

Jouji Miwa, Masato Yamamoto, Takashi Hirano, Masaru Sasaki and Jiapeng Tian

miwa, masato, taka, masaru@cis.iwate-u.ac.jp

Department of Computer and Information Science,
Faculty of Engineering, Iwate University

4-3-5 Ueda Morioka-shi Iwate-ken, 020-8551 Japan

Abstract As progress of the technology of multimedia and Internet, a personal computer is a very useful equipment for spoken language learning. We have developed a computer assisted language learning system (Japanese Language Education System for Speech using on an On-demand Network: LESSON/J) with a WWW server.

Using the system, (1) students for language learning can hear Japanese speech of syllables, words, basic sentences and accent using the on-line system, (2) students can learn Japanese speech using the dictation system with Java applets and (3) Japanese speech uttered by foreign students can be evaluated by a integrated technique of speech recognition and multimedia.

Key words CALL, Language Learning, Japanese, Multimedia, Speech Recognition, On-demand

1 まえがき

音声認識とマルチメディア技術を統合したオンデマンド・ネットワーク型日本語音声教育システムの開発の必要性について、説明する。

これまで、留学生の日本語教育では、来日前の母国での日本語教育と、来日後の6ヶ月程度の日本語教育が行われている。しかし、来日後の日本語教育は、滞在等にかかる費用や収容人数の面から、留学生総数が制限されている。このため、来日前の母国での日本語教育のより充実が必要となっている。

また、国際化に向けて、規制緩和による貿易障壁の除去とともに、日本語の言語障壁を取り除くために、留学生ばかりでなく、職業を持ち時間的余裕の少ない社会人が、いつでも手軽に日本語教育を受けられる方策が必要である。

さらに、日本語教育においても、授業に日本語を使うために、文法教育の他に、音声言語の教育が重要である。しかし、日常会話の面からも日本語の音声教育の重要性は認識されているが、日本における英語教育に類似して、(1) 外国における日本語教師の母語が日本語でない場合があること、(2) カリキュラム編成上音声教育の時間的配分が少ないと、(3) 音声教育教材が少なく、また高価であることなどの問題点があり、日本語教育において音声教育が十分に行なわれていないのが実情である。

このため、世界中の人々が、いつでも、どこでも、だれにでも、手軽に利用するため、インターネットとWWWを活用することにより、時間に束縛されない日本語音声教育のバーチャルスクール（仮想学校）の構築が必要である。

しかし、バーチャルスクールでは、コンピュータを用いた学習者のみによる独習となることが多いため、一般的の音声教育のように、学習者の音声を教師が聞き、教師の内省による発音指導により、悪い発音を指導することができないという問題が生じる。

このため、図1に示すように、音声認識技術を利用して、学習者の音声の良否のレベル等を客観的数値でスコアリングし、学習者にフィードバックすることが必要である。また、画像、音声、声道形の3次元表示等のように、コンピュータグラフィックとマルチメディア技術を用いて、学習者に理解しやすい効果的なフィードバック法が有効だと考えられる。

以上の考察より、本論文では、我々が開発した音声認識とマルチメディア技術を統合したオンデマンド・ネットワーク型日本語音声教育システム (LESSON/J: Japanese Language Education System for Speech using on an On-demand Network) について説明する。

このシステムの特色は、

(1) オンデマンド・ネットワーク型であり、インター

ネットを利用して、世界中で、いつでも、どこでも、だれにでも、手軽にゲーム感覚で利用できること、

(2) 音声認識技術を利用して、学習者の音声をスコアリングし、客観的に評価できること、

(3) マルチメディア技術を用いて、3次元画像、音声等を活用して学習者にフィードバックしていること、

(4) だれにでも簡単に使用できるように、ヒューマン・インターフェースに優れたGUI(Graphical User Interface)を利用し、ゲーム感覚の教材としていること、

(5) Java[1] アプリケーションにより、コンピュータのOSに依存しないマルチプラットホーム（多機種）型であること、

(6) Unicodeにより、学習者の母語毎に、マルチランゲージ（多言語）に対応していること、

(7) 日本語音声の聞き取りによる書き取り（ディクテーション）学習では、音節、単語、文、子音組、長母音、促音、アクセントの学習教材をそろえたこと、さらに、

(8) 日本語音声のフィードバック式学習では、音声認識技術を用いて、アクセント、長母音、促音の学習教材をそろえたことである。

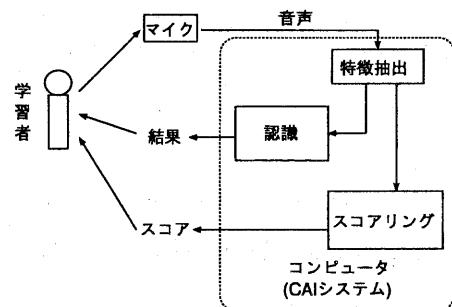


図1：音声認識を利用した語学学習システムの構成図

2 日本語音声の聞き取り学習

日本語音声教育では、聞き取り能力が重要であることから、音節、30個の単語[2]、30個の文章の3種類の聞き取りによる書き取り（ディクテーション）クイズとして、1997年6月下旬より公開[3]している。

本システムは、約10ヶ月の運用において、約45ヶ国から、月平均約500件のアクセスがされている。しかし、音声学習教材が少ないと、アクセス時間を要することなどから、まだ、十分に活用されていないため、今回第1.2版として、スレッド化等による高速化、および、モジュール化による小型化をはかると共に、13個の子音組、18個の長母音、25個の促音、および、20組42個のアクセントの4種類の学習教材を追加した。

図 2 に、LESSON/J 第 1.2 版のトップページの表例を示す。このトップページから、7 種類の教材、5ヶ国の母語（英語、中国語、韓国語、マレー語、日本語）、最高 6 段階のレベルまたは教材リスト、5 種類の表示形式（ローマ字音節表、ひらがな音節表、ローマ字表記、ひらがな表記、漢字表記）を選択した後、

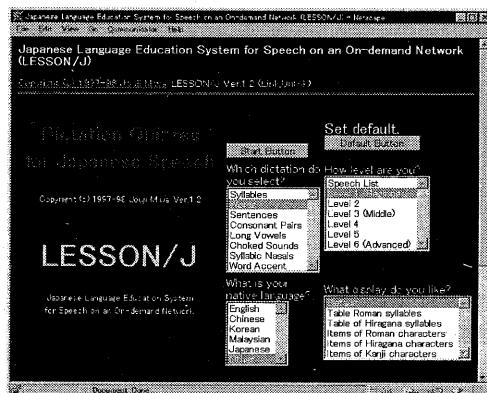


図 2: LESSON/J のトップページ

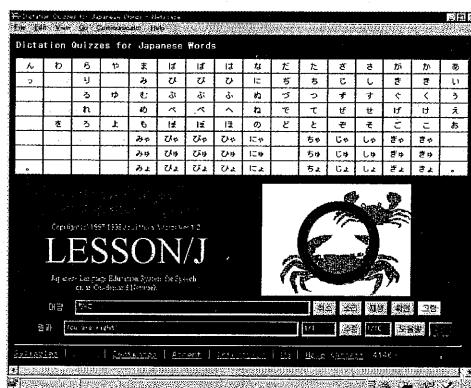


図 3: 韓国語版単語聞き取り教材の例

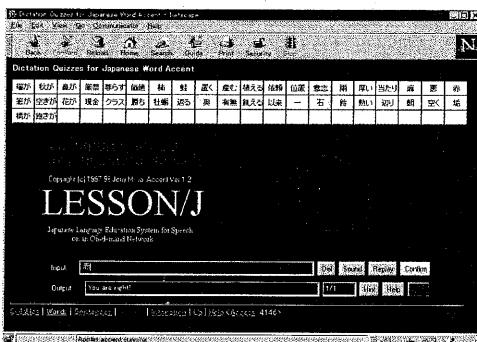


図 4: アクセント聞き取り学習教材の例

Start ボタンを押すことにより、日本語音声学習教材が開始される。

図 3 に、単語音声の聞き取りクイズの Java アプレットの表示画面を示す。この例では、30 個の単語がランダムに音声で提示されるため、ひらがなボタン、ローマ字ボタン、ひらがなテキスト、ローマ字テキスト、漢字テキストのいずれかの方法で、回答すると、○×表示と対応する音で判定結果を出力するようになっている。このシステムは、Unicode によりマルチランゲージ対応となっており、図 3 の例では、韓国語母語学習者用のため、ハングル文字でコマンドボタンが表示され、ヒントボタンにより蟹の絵と、正解による○印が表示されている。

図 4 は、アクセントの聞き取り学習教材 [4] の例であり、42 個の単語音声がランダムに提示されるため、その聴取結果に対応する単語ボタンを、マウスでクリックすることにより、アクセントの聞き取り能力を訓練する教材である。この例では、「赤」を正しく聴取していることを示している。また、漢字が読めない学習者のために、書き取りクイズと同様に、ひらがなやローマ字に続けて、アクセント位置を「aka1」のように数字を併用して入力する方式も可能になっている。

3 音声分析によるフィードバック式日本語アクセント学習

コンピュータによる独習型の音声教育システムにおいては、教師の代わりに発声の悪い個所を指摘できるようにするために、学習者の音声のレベルを自動的に評価して、学習者にフィードバックする機能が必要である。このため、本システムでは、基本周波数抽出の音響分析を行い、単語中の相対アクセント位置を、評価値として抽出し、学習者に提示することにした。

相対アクセント位置の算出手順 [5] は、10ms 每の基本周波数抽出、5 点メディアン平滑化、10 点回帰直線による内挿し、基本周波数が最大値を持つ時刻を、単語の始端から終端までの相対位置として、オンライン [4] で算出される。

図 5 上段に「赤」、図 5 下段に「垢」の学習画面の表示結果を示す。この図で、+印が最初の基本周波数抽出結果であり、実線が最終の内挿結果を表している。この結果から、無声子音を含む単語音声でも、基本周波数を内挿でき、正しく相対位置が算出されていることが分かる。また、「秋」は第 1 音節に、「垢」は第 2 音節にアクセントがくるが、この場合、0.3 と 0.65 というアクセント位置に対応した結果が算出されている。このように、音響分析結果を学習者にフィードバックすることにより、日本語音声の独習の手助けに活用することができる。

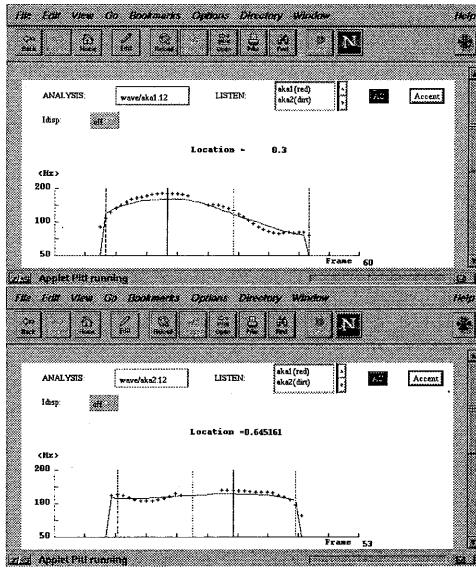


図 5: フィードバック式アクセント学習教材の例

4 音声認識技術を利用したフィードバック型日本語音声学習

4.1 特殊拍の持続時間

外国人の日本語教育において、習得が困難な音声に特殊拍がある。特殊拍は、長母音、促音、撥音の3種類があり、リズムの時間構造に起因している。

特殊拍の中で長母音は、母音を一拍分のばすような発声 [6] で、引き音節とも言い、長母音を含む「おじいさん」、「おばあさん」のような単語は、長母音を含まない「おじさん」、「おばさん」のような単語と意味的に区別して発声する必要がある。また、促音は、一拍分息を止めるような発声で、「っ」で表記され、促音を含まない「音」と促音を含む「夫」のような単語も区別して発声する必要がある。さらに、撥音は、一拍分の鼻音であり、「ん」と表記される。

4.2 DP 法による特殊拍の持続時間の推定

音声認識技術において、動的計画法（ダイナミックプログラミング、DP）は、発声速度の非線形伸縮に対応した整合を、効率的に実行することができる。このため、教師の音声と学習者の音声の非線形整合を、動的計画法を用いて行った後、パックトラッキングにより最適経路を探索すれば、教師の特殊拍の継続時間が既知であるから、対応する学習者の特殊拍の持続時間を計測することができ、学習者に特殊拍の持続時間に関する数値をフィードバックすることができる。

ここで、学習者の入力ベクトルパターン x_{ik} と教師の標準スペクトルパターン r_{jk} において、 i, j をフレー

ム番号、 k をチャネル番号とすると、フレーム時刻の格子点 (i, j) における局所距離 $d(i, j)$ は、以下の式で表される。

$$d(i, j) = \|x_i - r_j\| = \sum_{k=1}^{16} |x_{ik} - r_{jk}|$$

動的計画法は、次のような漸化式で計算される。

$$\text{初期条件 } h(1, 1) = 2d(1, 1)$$

漸化式

$$h(i, j) = \min \begin{cases} h(i-1, j-2) + 2d(i, j-1) + d(i, j) \\ h(i-1, j-1) + 2d(i, j) \\ h(i-2, j-1) + 2d(i-1, j) + d(i, j) \end{cases}$$

4.3 特殊拍の音声学習

実際の日本語音声学習システムでは、マイクから入力された音声が、(1)8kHz の μ -law 形式で Java アプレットに入力され、(2)線形形式変換、(3)単語音声区間の検出、(4)FFT[7]、16 チャネル帯域化および正規化 [8]、(5)動的計画法による単語整合、(6)最適経路のパックトラッキング、(7)音素の持続時間の計測、(8)相対音素時間長の表示の順序で処理される。

図 6 の左側は、教師音声が「おじいさん」と発声した場合の最適整合経路であり、右側は、学習者へフィードバックする相対音素持続時間長の目盛りである。この例では、教師音声の方が約 2.4 倍だけ [i] の持続時間が長いことを示している。また、図 7 は、上段が教師、下段が学習者の音声波形であり、DP の最適経路から、対応する音素区間を表示しており、学習者は、[i] の発声時間長が短いことをすぐに理解することができる。

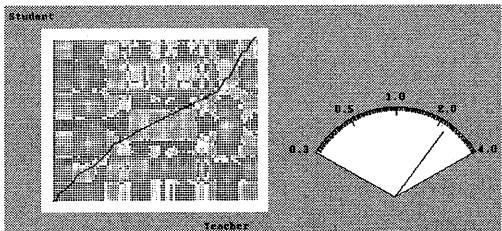


図 6: 最適経路と相対持続時間長のスケールの例

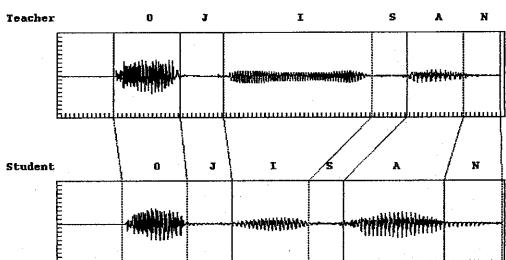


図 7: 波形による持続時間のフィードバック情報の例

5 調音情報のフィードバック音声学習

5.1 3次元調音モデル

従来の音声教育では、学習者の発声した音声を教師が聞き教師の内省等を考慮して、学習者の発声指導を行っているが、CAIシステムでは自動的に学習者の音声を分析し、学習者へフィードバックすることにより発声指導をすることが必要である。

調音のフィードバック情報として、ホルマント周波数の音響特徴が抽出が容易なことから多く用いられているが、ホルマント周波数と調音位置の対応が明確でないため、初心者には有効なフィードバック情報となる問題がある。このため、音声教育に役立てるため、調音モデルと音響特徴に対応した9個のパラメータによる3次元声道形状情報の利用を検討した。

調音モデルとして声道形状を、図8に示すようにMR Iデータから、声門面積 Ag 、後室位置 Xb と面積 Ab 、調音位置 Xc と面積 Ac 、前室位置 Xf と面積 Af 、唇開放面積 Al 、声道長 Xl の9個のパラメータで簡単化して表現[9]する。

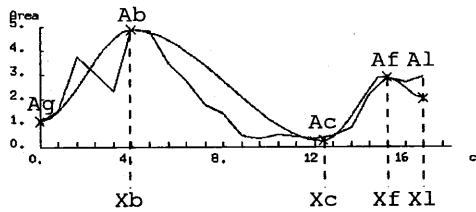


図8: MR Iデータと9パラメータ調音モデル

表1に、日本語5母音のMR Iデータ[10]と9パラメータ調音モデルから調音音声合成器により算出[11]したホルマント周波数の誤差を示す。この結果から、MR Iの22個のセクションデータを半分以下の9個のパラメータによる3次元調音モデルで良く近似できることが分かる。

表1: MR Iと調音モデルによるF1～F3の誤差

	F1	F2	F3
a	0.4%	4.3%	7.3%
i	0.6%	0.4%	5.2%
u	1.8%	6.1%	9.3%
e	4.0%	8.7%	9.4%
o	0.8%	8.1%	9.3%

5.2 日本語母音の調音モデル

表2に、成人男性が発声した日本語5母音の平均ホルマント周波数[12]に対応する9パラメータ3次元

調音モデルによる典型値を示す。ここで、断面積の単位は mm^2 、調音位置の単位は mm である。この表から、/i/や/e/は Xc が大きいことから前舌母音であり、/a/は Xc が小さいことから後舌母音であることが分かる。

表2: 日本語母音の調音モデルの典型値

	Ag	Ab	Ac	Af	Al	Xb	Xc	Xf	Xl
a	67	186	25	780	590	20	40	128	186
i	110	491	28	289	180	40	124	152	191
u	150	308	10	344	55	24	76	148	186
e	48	247	27	422	300	52	108	148	183
o	20	282	6	695	69	20	52	132	190

調音パラメータ Xc , Al を用いた日本語母音の母音五角形を、図9に示す。このように簡単化することにより、調音とホルマント周波数の関係を対応することができ、調音次元と音響次元を対応させて音声教育に役立てることができる。

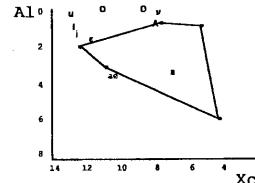


図9: 日本語と英語母音の Xc と Al の調音点

5.3 英語母音の調音モデル

表3に、英語MR Iデータ[13]から求めた9次元調音モデルによる典型値を示す。また、英語母音の Xc , Al 平面図を図9に示す。これより、英語母音は日本語より前舌かつ狭母音傾向があると言える。また、/i/と/I/では、調音位置はほとんど変わらないが、/i/の狭めの面積をより小さくすればよく /ε/と/ae/についても同様のことがいえる。さらに、/O/と/o/については逆で、狭めの位置をずらしていることが分かる。

表3: 英語母音の調音モデルの典型値

	Ag	Ab	Ac	Af	Al	Xb	Xc	Xf	Xl
A	33	122	66	579	117	32	81	139	175
a	45	112	23	655	388	36	71	131	175
i	33	468	10	205	158	67	123	158	167
I	20	396	75	197	116	71	127	151	167
v	33	220	53	569	22	28	79	139	175
u	40	586	15	547	26	48	124	158	183
ε	21	303	136	259	185	71	116	135	159
æ	22	387	319	479	394	87	108	143	167
O	61	142	32	708	13	36	86	135	175
o	18	272	32	725	14	44	112	147	175

5.4 声道形状の3次元表示

英語母音の/i/のMR Iデータによる3D声道形状を、図10に示す。また、9パラメータモデルによる

/i/ と /I/ の声道形状を、図 11 に示す。この図からも、/i/ では /I/ と比較すると調音点はほぼ同一であるが、狭めをより大きくする必要があるという結果が、視覚的に理解することができ、9 パラメータモデルにおいても声道形状を十分に表現できることが分かる。

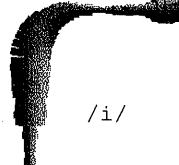


図 10: MRI 直接データからの 3 次元声道形状



図 11: 9 変数調音モデルによる 3 次元声道形状

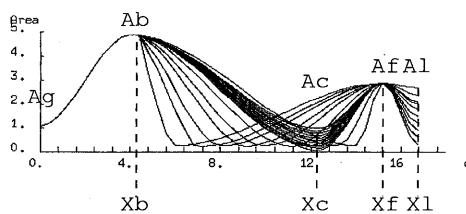


図 12: /i/ の調音モデル X_c , Ac , Al の変更

5.5 調音モデルと音響特徴の対話型表現

/i/ の調音パラメータ X_c , Ac , Al を移動させたときの声道形状を図 12 に示す。また、調音パラメータを対話的に変更した場合のホルマント周波数の変化を、図 13 に示す。このように、9 個のパラメータでモデル化することにより、調音とホルマント周波数の関係を対応することができ、調音次元と音響次元を対応させて、音声教育に役立てることができると考える。

6 考察

本論文で提案した、聞き取りによる書き取り、アクセント教育、特殊拍音声教育のプログラムは、Java で記述しているが、3 次元声道形表現は、OpenGL を用いて記述している。このため、前者はネットワークに対応して WWW サーバで運用できるが、後者はネットワーク対応になっていない。このため、今後、

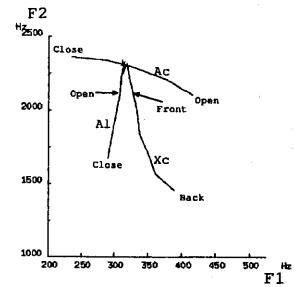


図 13: /i/ の調音パラメータと F1, F2 の関係

Java3D 等によりネットワーク型の 3 次元表現が容易になれば、インターネット化できることになる。

また、聞き取りによる書き取り（ディクテーション）による日本語音声教育システムは、インターネット上で 1997 年の 6 月より世界に公開しているが、他のアクセント教育システム等を、実際に運用するには、種々の検討が必要である。

さらに、ノート型パーソナルコンピュータに Web サーバと日本語音声教育システムプログラムを載せ、ポータブル型コンピュータ援用語学学習システム (PAL: Portable Computer Assisted Language Learning) と名付け、現在、岩手大学の留学生に貸し出し、コンピュータリテラシー教育と共に、日本語音声教育システムの評価を行っている。このノートパソコンを用いた LL システムは、非常に安価に構成できるので、今後の新しい LL システムとして期待される。

図 14 は、分散値を修正して算出した事後確率 [14] とスコアリング関数の関係を示している。このようなスコアリング関数用いることにより、音声認識における決定の他に、発声の良否を数値で表現することができるようになる。

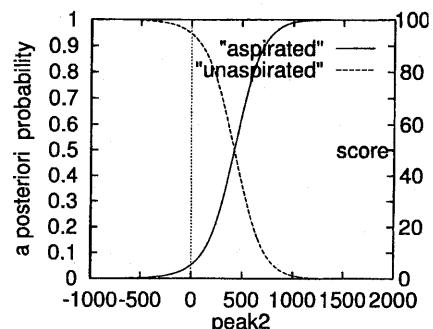


図 14: 音声認識トとスコアリング関数

7 まとめ

本論文では、音声認識とマルチメディア技術を統合したオンデマンド・ネットワーク型日本語音声教育システムを提案した。

本システムは、日本語音声教育ばかりでなく、英語などの外国語の音声教育等にも応用することができ、さらに、アイヌ語や琉球方言のような危機言語の効果的な音声教育へも拡張することができるものと考える。

本システムには、まだ、解決しなければならない事項があるが、コンピュータ性能の向上とインターネット・マルチメディア技術の発展によりバーチャルスクールを遠くない日に実現できるものと考える。

謝辞

本研究の一部は、文部省科学研究費補助金・基盤研究(B)(三輪代表, 09558022, 「高精度音声分析と音声CD-ROMを用いた独習用対話型日本語音声教育システムの開発」)、および、平成9年度文部省留学生特別経費「日本語音声教育の評価」によった。

本システムの開発では、本学大学院熊谷有香、4年生吉田奏子、熊谷勝君の協力を得た。また、日本語音声教育用基本単語については、広島女子大学今石元久教授、基本文型については、東京外国语大学鮎沢孝子教授、および、国立国語研究所江川清部長に協力いただき、アクセント資料については、東北大学留学生センター佐藤滋教授に強力協力いただいた。さらに、音声教材のアナウンサーデータの作成については、NHK放送研修センター秋山和平室長、今井久部長にお世話になった。

なお、本システムの一部は以下のURLで公開している。

<http://sp.cis.iwate-u.ac.jp/sp/lesson/j/>

参考文献

- [1] Sun Microsystems, Inc. : Java プログラミング講座, アスキー出版局, 東京 (Oct. 1996).
- [2] 今石元久：日本語音声の実験的研究、和泉書院、大阪、(1997).
- [3] 三輪、熊谷、田、今石：オンデマンド・ネットワーク型日本語音声教育システムの構築、信学会音声研技報、Vol.97, No.114, SP97-17, pp.55-62 (June 1997).
- [4] 三輪、熊谷、吉田：オンデマンド・ネットワーク型日本語アクセント音声教育システム、日本音響学会春季講演論文集、2-P-26, pp.345-346 (1998 3).
- [5] 加藤、三輪：移動平均と帯域制限を用いたケプストラム型基本周波数抽出とその応用、電子情報通信学会音声研究会技報、SP94-95, pp.29-36 (1995 2).
- [6] 新津、三輪、牧野、城戸：単語音声自動認識における言語情報の一利用法、電子通信学会論文誌、Vol.J62-D, No.1, pp.24-31 (1979 1).
- [7] K. Kim, J. Miwa and K. Kido: Recognition of isolated Korean digits using band pass filters based on FFT, J. Acoust. Soc. Japan, Vol.(E)4, No.4, pp.179-187 (1983 10).
- [8] 三輪、小原、牧野、城戸：非線形スペクトルマッチングによる単語音声認識の一方式、電子通信学会論文誌、Vol.J64-D, No.1, pp.46-53 (1981 1).
- [9] 佐々木、平野、三輪：音声教育のための3次元声道形状の対話型表示法、日本音響学会春季講演論文集、2-P-24, pp.341-342 (1998 3).
- [10] 楊長盛：MRIによる声道形状の精密計測法の開発と母音正規化への応用、宇都宮大学博士論文 (Mar. 1995).
- [11] Jouji Miwa: Interactive Visualization and Audialization of Speech Production Using Variable Vocal and Nasal Area Function, ASVA97, pp. 271-278 (Apr. 1997).
- [12] 三輪 譲二：“パソコン音声処理”，昭晃堂,(July 1991).
- [13] Brad H. Story and Ingo R. Titze: Vocal tract area functions from magnetic resonance imaging, JASA, Vol.100, No.1, pp.537-554 (July 1996).
- [14] 田、三輪：音声教育のための中国語有氣無氣音の識別、電子情報通信学会音声研究会技報、SP97-115, pp.55-62 (Feb. 1998).
- [15] 三輪、熊谷、山本：インターネット・マルチプラットホーム対応の対話型音声分析システム、日本音響学会春季講演論文集、2-P-27, pp.347-348 (1998 3).

付録

付録 表1 単語学習教材リスト

枝	靴	下駄
鍵	蟹	亀
大根	地図	時計
蠅	梨	茄子
獅子	猫	鼠
藤	富士山	西瓜
蓋	豚	蛇
麦	松	リス
本	パン	ペン
ポスト	ラクダ	プロペラ

付録 表4 促音学習教材リスト

単母音	長母音	促音
糸	伊藤	一等
干支	江藤	越冬
		音
過去	加工	夫
下戸	下校	滑降
里	砂糖	月光
		殺到
世帯	接待	日当
二等		埠頭
		沸騰
本	日当	離島
埠頭		立冬

付録 表2 子音組学習教材リスト

子音組			
延期	嫌忌	元気	戦記
前期	転記	伝記	年季
偏気	便器	ベンキ	面器 連記
enki	kenki	genki	senki
zenki	tenki	denki	nenki
henki	benki	penki	menki renki

付録 表5 アクセント単語教材リスト

読み			
あか	赤	-	垢
あく	悪	-	空く
あさ	麻	-	朝
あたり	当たり	-	辺り
あつい	厚い	-	熱い
あめ	雨	-	飴
いし	意志	-	石
いち	位置	-	一
いらい	依頼	-	以来
うえる	植える	-	飢える
うむ	産む	-	有無
おく	置く	-	奥
かえる	蛙	-	返る
かき	柿	-	牡蠣
かち	価値	-	勝ち
くらす	暮らす	-	クラス
げんきん	厳禁	-	現金
はなが	鼻が	-	花が
あきが	秋が	-	空きが 飽きが
はしが	端が	-	箸が 橋が

付録 表3 長母音学習教材リスト

単母音	長母音
叔父さん	お爺さん
叔母さん	お婆さん
鬼さん	お兄さん
エゴ	英語
茎	空氣
席	正規
敵	定期
時	陶器
雪	勇氣