

音声認識技術を用いた日本語韻律練習用ソフトの開発

松 崎 寛

1. 日本語音声教育の現状と課題

日本語の発音に対する学習者のニーズは、一般に教師が考えるより、はるかに強いと言われる（日本語教育学会 1991, 佐藤 1998, 池田 2003:89）。しかし教育現場では、「時間がない」「効果的な教え方が分からない」「教師が自分の発音に自信がない」「適当な教材がない」等の理由で、指導が避けられる傾向が強い。

この問題を解決する一方法として、発音の自学自習を可能にするコンピュータ支援教育（以下 CALL）教材の開発がある。2000 年代初頭における日本語音声 CALL 教材の課題として、池田（2003:104）は、「(1) 文レベルの練習問題」「(2) 音やアクセントの説明方法の改善」「(3) 音声認識の利用」の 3つを挙げたが、これらの問題は、その後の音声教育の進展および技術革新により、かなり改善されつつある。

たとえば (1)・(2) については、2000 年中期頃から、文の韻律に着目し、説明方法・練習方法を工夫した紙版教科書や教師用指導書が、多数出版された（河野他 2004, 磯村 2009 等）。また、後述する「自己モニター」を取り入れた新たな発音指導法の提案も見られる（小河原 1998, 小河原・河野 2009 等）。これらのシラバス・カリキュラム等を CALL 教材へ移植することにより、今後新たな教材が開発される可能性が大いにある。

また、(1)・(3) については、近年の科学技術の進歩により、大規模音声データベースや、新たな発音分析法を活用した音声認識エンジンが開発され、不特定多数の話者の発話文を、高い精度で認識することが可能になっている。これをもとにした、音声合成や音声認識技術を活用した CALL 教材の作成、およびその試用を通じての効果測定や実践研究が盛んに行われている（石井他 2001, 広瀬 2005, 羅他 2006, Shimizu et.al 2008, 小河原・池田 2009）。

音声認識エンジンを発音評価に活用したソフトで、製品化されているものとしては「AmiVoice CALL Web -Japanese-」がある。これは、アドバンスト・

メディア社の「AmiVoice」に学習者誤用データベースを組み込んだもので、学習者の発話を自動的に認識して「誤りを誤りのまま表記し」、有声・無声、「つ」・「ちゅ」、母音の長短の違い、促音の有無等、39のエラーを指摘することができる（岩田他 2009）。文の練習をする画面は、図1のようにになっている。



図1 「Ami Voice CALL Web Japanese」の画面

このソフトは、子音・母音および特殊拍の練習のみを扱っており、アクセント・イントネーションは判定対象外である。だが、日本語の自然さの評価には、子音・母音よりも、イントネーション等の「高さ」が大きく影響することが明らかになっている（大山・三浦 1990, 佐藤 1995）。学習者の発音上の問題、たとえば、「今日は、ちょっと…」の句末を低く弱く言えなかったり（土岐・村田 1989）、車で送ってもらったとき【「ここで」結構です】と言うべきを【「ここで【結構です】】のように「結構」を高く強く言う（天沼他 1978）等の問題が感情的評価に影響したり、あるいは「聞きにくい」という全体的な印象が、学習者の能力の評価を下げる危険性について、さまざまなことが指摘されている。

そこで、本研究では、この子音・母音・特殊拍練習用教材を補完するべく開発中の、韻律練習ソフトの試用結果について報告する。

2. 韻律学習教材開発における問題

音声を視覚化した情報、すなわち音響分析の結果を学習者に与えることには、

有効な面もあるが、そこから学習者が情報を適切に読み取り、自律的学習に生かすことにはあまり期待できない。これまで開発された発音 CALL 教材でも、波形・ピッチ曲線・スペクトログラムや、音声学的な説明を、そのまま学習者に示しているだけのものが多い。たとえば、先ほどの図 1 の下段「ピッチカーブ」は、その一段上の学習者発音の波形と、その上のモデル音の波形に対応する高さを表すが、どの音がピッチ曲線のどこに対応しているのかを、学習者が図から読み取ることは、まず不可能であろう。

この問題を解決するためには、韻律の自動認識システムにおける表示法を改善し、連続音声を適切に分節して、高さとの対応を示す必要がある。この考え方は、串田他 (1995) が「プロソディグラフ (図 2, 以下 PG)」を開発した着想と似ている。これを CALL 教材に取り入れ、学習者の発音を自動的に分析して PG を表示するシステムを構築すれば、ピッチ曲線をそのまま示すだけより分かりやすいフィードバック (以下 FB) ができると考えた (ちなみに図 1 上段の PG は、開発者の私信によれば、音声認識結果を出力したものではなく予め例文ごとに用意された画像であり、実際のモデル音の高さとは対応していない)。

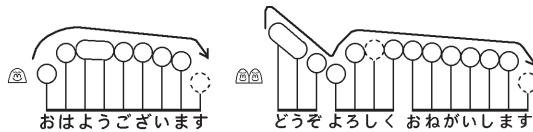


図 2 プロソディグラフ (河野他 2004)

PG について簡単に説明する。これは、音響分析の結果を参考にしつつ、日本語話者の発話の韻律を、2 段階の高さではなく実際の高さに近い形で示した教材である。音の長短、言いよどみ等の感情表現に伴う持続時間やポーズ、アクセントやイントネーションに関わる高さの変化、母音の無声化等を分かりやすく表示できる。河野他 (2004) の『1 日 10 分の発音練習』は、PG を活用して文単位で韻律を体系的に学習させる教科書だが、これは、例文を通じて生成規則を明示的に学習し、学習者が規則を元に、自分で適切な韻律で発話ができるようになることを到達目標としている。

2.1 長さの表示に関する問題

PG は、音響分析結果を参考にはしているが、紙版の教材では、よい発音を

引き出すために、教材作成者の裁量で高さや長さを修正して示している。ある意味、モデル音の理想形を音韻的に表示したものであると言える。これは紙版教材ゆえに可能なことであるが、一方、CALL教材で、学習者の発話を機械が分析・表示する際には、「長さ・高さをどのように修正して示せば、効果的な学習が可能になるか」について考える必要がある。

1. で述べた「音とピッチ曲線の対応」に関しては、細かく見ると、2つの解決すべき問題がある。1つは、先述したように、モデル音あるいは自分の発話のピッチ曲線の、どの高さがどの音に対応しているのかが分かりにくいという問題である。これに関しては、「連続する発話を子音・母音で自動的に分節し、適切な単位で示す」機能が必要である。

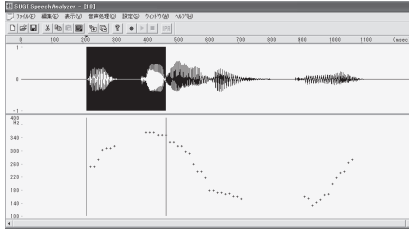
もう1つは、モデル音と自分の発話のどこがどう異なるか、その対応が分かりにくいという問題である。たとえば、図1は白黒で分かりにくいのが、画面では、モデル音の波形とピッチが黄色で、学習者の発音の波形とピッチが緑色で示されている。この発話では、特に「ひとりで」以後の発話のタイミング、リズム、ポーズ、発話速度等がモデル音と学習者の発話で異なるため、2つを対応させてどこをどう直せば良いのかを把握するのは、かなり難しい。両者を一画面にまとめるにせよ、別々に示すにせよ、同じ問題が生じる。問題を解決するには、「音声認識結果を元に、自動で時間長を調整し、両者の音節の前後位置を対応させて示す」機能が必要と考えられる。

2.2 高さの表示に関する問題

次に、高さをどの程度そのまま示すべきかという問題がある。たとえば、/「イクラデスカ/ (図3) 等のピッチは、物理的には/イ/の平均値より/ク/の始端部が高く始まり、アクセントの下がり目が遅れて生じる「おそ下がり」(杉藤1972)現象が見られる。しかし、/ク/を高く示した図を学習者に与えれば、/イ「ク」ラデスカ/という発音を誘発する危険がある。これに関して石井他(2001)は、CV単位の平均値ではなくVC単位の終端側のF0target値を用いることで、アクセント型の実際の知覚に対応した高低パターンが表現可能となると述べている。

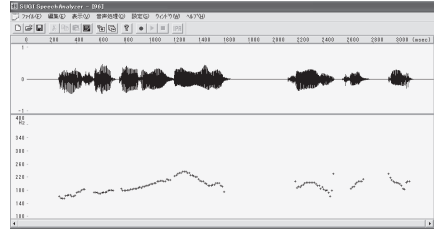
一方、遠慮がちな「申し訳ありませんが、今日はちょっと…」(図4)では、句頭が低く始まる「文頭のイントネーション(川上1956)」現象も見られる。この場合、学習者の「今日はちょっと…」「じゃ、けっこうです」や「いいですねえ!」等の発話の句頭が高くあるいは低く始めると、どの程度「気持ちか

こもっていない」と評価されるかという問題に行き当たり、一律に高さを変換することの妥当性も疑問視される。



イ ク ラ デ ス カ ？

図3 おそ下がり



モーシワケアリマセンガ キョーワチョット

図4 文頭のイントネーション

以上の議論をもとに、今回の試用版の作成方針としては、

- ・ PGの棒線の間隔を統一する。つまり音節を等間隔に並べ、ポーズも削除し、リズム練習や母音の無声化の練習は、別のコーナーで扱う（発話中の母音の長さやポーズの実時間を制御するのではなく、音声はそのまま視覚的にポーズを消し、発話タイミングを一致させることで、モデル音と自分の発音や、自分の過去の発話と現在の発話を視覚的に比べやすくするものである。学習者の発話のリズム等は変えない）。
- ・ ピッチをその高さのまま表示することとし、文頭のイントネーションにより句頭が低く始まってもし自然さをそれほど感じない「遠慮がちな言い方」を中心に練習する。上記「いくらですか」のようにおそ下がりによりモデル音のピッチが音韻的な型と大きく異なってしまう例文は、使わない。

ことにした。そして、学習行動の分析結果や試用の感想等を元に、次回の改訂に生かすことにする。

2.3 「Prosody Tuner」の主要な機能

この方針に沿って開発した、韻律練習用ソフト「Prosody Tuner」の主要な機能について述べる。画面は、図5のようになっており、次のような機能を有する。

- (1) モデル音および入力された学習者の発話を、子音母音で自動的に分節する
- (2) 各音節（特殊拍を音節後部を含む長い円と、自立拍のみの短い円）を等

間隔に示す

- (3) 入力された高さにあわせ、適切な箇所に円を表示する
- (4) 上段にモデル発音、中段に学習者の発音を配置。ボタンをクリックすると音声が出るようにして、視覚的・聴覚的比較を行うことを可能にする
- (5) 学習者の発音および PG を記録し、下段のリストに表示する。これを選択すると、中段に過去の発音の PG と音声呼び出すことができる
- (6) 1つ1つの円を上下にドラッグして「音声変換」ボタンを押すと、高さを変えたものを再生することができる

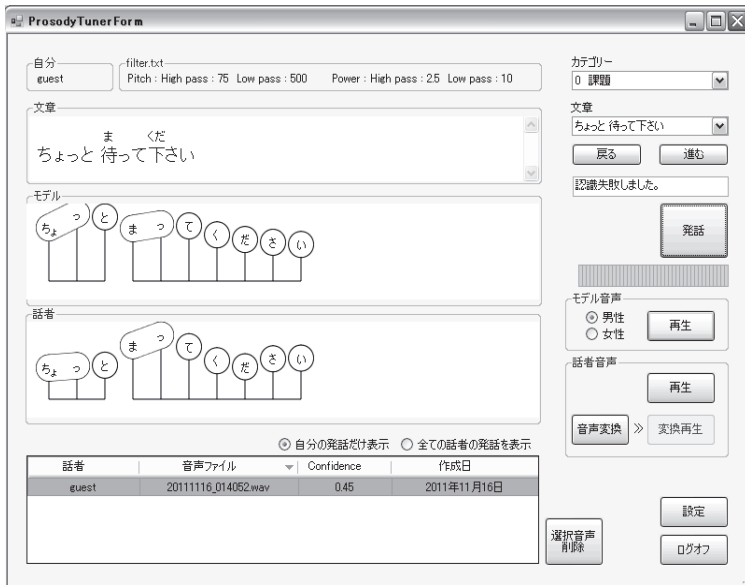


図5 「Prosody Tuner」の画面

モデル音だけでなく学習者の発話の韻律も PG 化して、分かりやすく示せることが、本ソフトの最大の特長である。モデル音は学習者の発話を分析する経路と同じ方法で PG として出力されるため、両者を同じ基準で比べることができる。さらに、(5)の機能により、同一画面上で複数の PG を切り替え、過去の発音との視覚的比較をアニメーションのように行うことも可能である。また、windows 上で動くソフトなので、PG を紙媒体に移植し、印刷教材とし

て教室で使うことも可能である。その際、高さ表示が意図どおりにならないときは、(6)の変換機能を活用して、見た目の高さを微調整することもできる。

2.4 例文の選定

例文の選定は、米国ミドルベリー大学の日本語教師複数名の協力を得て、学習者が日常よく使う「そうなんだ」「がんばって！」等を選定してもらい、米大学で用いる初級教科書の例文を加えた。さらに、「今日中にやっちゃわなきゃ」等の縮約形や、「コーヒー？」等の聞き返し等、学習者が苦手とする発音も含めた。その結果、例文は約 200 文となった。アクセント認識用データとして、2～5 拍の様々なアクセント型の語、約 150 語も選定し、その 200 文 + 150 語について、2010 年 7 月、ミドルベリー大学の初級～中級学習者 50 名の発音を採取した。録音に際しては、単調な読上げ課題で発音が機械的になることを避ける意味で、文字を示しながらモデル発音を聞かせる「擬似学習ソフト」を作成し、リピートしたものを収録した。その音源を、言語学や日本語教育学を専攻する、音声に詳しい大学院生 6 名がラベリング（「夕方→「ゆんが」った」のように音色・リズム・高低を分析し、正誤を評価）した。

3. 「Prosody Tuner」の試用

3.1 調査の方法

Prosody Tuner を学習者に試用させ、ソフトの問題点を探るとともに、パソコンを用いた発音学習行動において何が問題になるのかを調査した。

対象者は、筑波大学の留学生 14 名（中国 10、韓国 3、ルーマニア 1）である。日本語能力・音声学の学習経験・発音指導を受けた経験等は、学習者によりさまざまである。期間は 2011 年 10 月から 11 月にかけてで、試行時間は一人あたり 70 分から 90 分程度である。

まだソフトが完全に自学自習できる段階にないため、横に教師がついて、高さの表示が明らかにおかしいものに適宜説明や指示を与えなが



図 6 試行の様子

ら、学習を進めた。学習者には、練習しながら気づいたこと、頭に浮かんだことを何でも良いから常に言語化するように教示した。試行中の学習者の様子は背後からの録画および画面キャプチャにより記録した(図6)。練習後、半構造的インタビューを行い、さらにその後、メールでの質問紙調査により、内省を記述させた。

試行時の練習の手順は、次のとおりである。

- (1) リストから練習したい文を選択
- (2) モデルを聞かず、例文を普段言うように読ませる。これが「最初の発話」になる。
- (3) モデル音や自分の発話を聞かせる。音声・PGを比べて考えたことを何でも言わせる
- (4) 再度発話。(3)と(4)の繰り返しを、学習者が良いと思うまで続ける。
- (5) 自分の過去の発音と聞き比べたり、発話を合成したりできることを伝え、基本的には学習者の好きなようにまず練習させ、問題点等が留意されないときは、教師が遠まわしに教示を与える。

この、教示を与える際に留意したのは、教師がすぐに誤用の指摘・訂正を行ったり、機械的にリピートさせたり、アクセント規則等を明示的に与えたりするのではなく、なるべく学習者の「気づき」を促すように働きかけるということである。これは、学習者に自分の発音を観察すなわちモニターさせ、モデル音とどこがどう違うかを考えて「基準」を形成させる、という、「自己モニター」を活用した学習における「妥当な基準」の形成と自己評価能力の向上を念頭に置いたためである。

したがって試行においては、この学習者独自の「基準」を否定したり、教師から一方的に強制したりするのではなく、たとえば、次のような働きかけを行って、漠然と考えていることを言語化するように促す。

「モデルと自分の発話は、どこが、どう違いますか」

「文の後ろのこのへんはどうですか、同じですか」

「どこをどう変えたら良い発音になると思いますか」

また、自分の最初の発音と今の発音を比べさせて、同じか違うかを聞いたり、気づいたことを語らせたり、モデル音が低い箇所に対して学習者から「何度も声を低くしているが、これ以上は低くならない」と言うコメントが出たら、音声変換機能を利用して、学習者の発話を合成して聞かせたりした。

3.2 結果および考察

3.2.1 全体的なコメント

学習者からのコメントおよび学習行動について述べる。

試用の感想については、14名全員から、「面白い」「さらに練習したい」「あと半年練習したら発音が上手になると思う」等のコメントが得られた。

今回は中国語話者が多かったということもあり、抑揚の激しい発話、つまり韻律のヤマを多く作りすぎてしまう誤用が多く見られたが、その修正に視覚的FBが有効なようである。次のような意見が見られた。

「モデルと自分の発音を聞き比べてもどこに差があるか分からないときが多いので、見て確認できて助かる」

「イントネーションの下がり目と上がり目の部分が一目瞭然」

「日本語の後ろの文はよく平らになっていて、そこに気づいた」

一方で、

「学習者自身がPGを見てポイントを掴めるかは疑問。発音の異なるところを赤字で明示できると、学習者にとって負担が減る」

という、誤用に対する明示的FBがほしいという意見も多く聞かれた。段階に応じて必要なFBが得られるような工夫が必要だと考えられる。

3.2.2 モデル音と自分の発音の比較

まず、音声合成機能については、高さの幅を変えすぎると人工的な声になる等、調整が困難なところがあったが、合成音声で高さを確認するという作業自体は、半数程度の学習者が、理解に有効だと捉えており、「もう少し簡単に、妥当な結果が出れば、ぜひ役に立つだろう」等の意見が聞かれた。ある学習者は、「どうしてもモデルに似ないとき、合成して聞くと、自分の声に基づいてどうすればよくなるのかが分かる」とコメントし、別の学習者は、「私は声が低いので、モデル音と同じ高さの音を出すのが難しい。自分の声を合成すると、トーンを維持しながらイントネーションだけ変えることができ、より現実的な発音の参考になる」と述べていた。

モデル音の地声の高さ等の非言語情報ではなく、アクセント等の言語情報に学習者の目をむけさせる必要があることは言うまでもないが、このような問題は、CDあるいは教室でのリピート練習でも生じうることである。たとえば女性のモデル音を男性の学習者が真似るとき、個人特性としての非言語情報と、言語情報であるアクセント成分等の線引きが、不慣れた学習者には難しいとい

うことがある。CALL教材の改善案として、モデルの声自体の高さを調整できる仕様にする、あるいは男性・女性1名ずつではなく、さらに多くの話者のモデル音を用意しておき、自分に近い声質のものを選ばせる等の改良が考えられる。

3.2.3 学習者自身の過去の発音との比較

次に、学習者自身の過去の発話と今の発話を比べることの学習効果については、次のようなコメントがあった。

「最初の発音と練習した発音は本当に違いがあった。練習すればよくなるなと思った」

「どんどん変わっていくんだ…というのが自分で確認できてとても良かった」

「例文によって改善された感じとあまり変わらない感じがある。だが気にしながら発音すると、多かれ少なかれ、古い発音と違うところがある」

「最初の発音と練習した発音は違いがあった。しかし、練習中何回も発音したものは何が違うかはっきり分からなかった」

モデル音を聞く前すなわち普段の発音と、聞いた後すなわちリピートの「大きな違い」に対しては、両者の違いを耳で聞いて確認できる学習者が多いが、その後の、モデル音リピート練習中に発音が修正・確認できるかどうかには、学習者により意識の違いがある。さらに、1回1回の発話時には気づかずとも、まとめて複数の発話を聞き比べると、変化に気づくこともあった。適切なタイミングで振り返りを行うことが重要であると言える。

上記の最後の意見は、3.2.2で「合成音のほうがモデルとして分かりやすい」とコメントした学習者のものである。「やはり自分の発音はモデルに距離がある感じがする。モデル発音を聞いて、自分の発音とどこかが違うのか、分からない場合がよくある」と述べており、「物真似」から離れた、言語としての韻律特徴の模倣が困難であることが分かる。最終的には、地声の高さや声質が違っている話者からも言語情報が抽出できるようにならなければ、発音の自律学習を効果的に進めることは難しいわけだが、学習初期においては、このための段階的なモデル音の提供やFBが必要になるということであろう。

何度か録音再生するうちに、発音が改善する瞬間があったが、そのとき、聴覚的・視覚的FBを元に自分の発音が「改善した」と確信できる学習者と、できない学習者がいる。後者に対しては、「モデル音と自分の最後の発話を比べる」だけでなく、「自分の最初の発話（モデル音を聞く前の普段の発音）と最

後の発話を比べる」と、向上が実感できるようであった。目標とする発音と自分の発音が「同じになった」と感じるためには、自分の悪い発音と良い発音のどこがどう違うのかを分析するという前段階が必要とされ、またその違いが大きいものから小さいものへと移行することが重要ということかもしれない。

つまり、リピートという行為はオウム返しで偶然できることがあっても、学習者の中に「妥当な基準」が形成されないと、またできなくなる（松崎2002）。これは、偶然発せられたものであって、「できるようになった」わけではない。学習者は何も変わっていない状態である。リピートでも何でも、無理やりOKの状態を作り、それを体感させることが、発音改善の第一歩である。重要なのは、モデル音の直後リピートにより発音が向上することそれ自体ではなく、リピートにより良くなった発音を自身のモデル音とし、それを、悪かった発音と比べて学習者に分析させ、模倣させる糧とすることである。それによって得られた「妥当な基準」を学習者に自覚させるためのFBが大切である。

3.2.4 気づきを促すFB

長い文、すなわち気をつけるべき箇所が複数にわたる例文だと、情報過多になり、一部に注意すると他方が修正されにくくなる傾向がある。たとえば「その角の手前でおろしてください」という例文は、/ソ「ノカ」ドノ | 手「前デオロ」シテクダサ「イ / のようにヤマ2つになるモデル音であったが、学習者の大半は、/ソ「ノカ」ドノ | 手「前デ | オ「ロ」シテクダサ「イ / のようにヤマ3つになったり、/ソ「ノカ」ドノ手前デ | オ「ロ」シテクダサ「イ / のようにヤマの切れ目が異なるヤマ2つの文になり、さらにアクセントも / オ「ロシテクダサ」イ / や / オ「ロシテクダ」サイ / のように誤る傾向が強い。このとき、アクセントの下がり目の修正に注意すると、この「手前」でのピッチの立て直しにはPGを見ても気づけない状態に陥る。そのときは教師が、違うところや同じところを指して「こっちはどうか。同じか違うか」等のヒントを与えたが、気づきを促す適切なFBをコンピュータにどのように出させるのが、今後の課題である。

/ソ「ノカ」ドノ手前デ | オ「ロ」シテクダサ「イ / のようにヤマ2つになる学習者に対して、「手前でおろ」は「紫色」と同じアクセント」と教えたら、妥当な基準の形成に役立ったケースがあった。そのような、効果的な「ヒント」を教える方法の模索、無理やりOKの状態にさせる様々な工夫が必要であり、またそれをどう学習ソフトに取り込んでいくかを考える必要がある。

小河原 (2007) は、自己モニターにおける妥当な基準の形成に関連する議論として、「教師がモデル発音や具体的な発音の仕方を与えることはあっても、それは学習者自身が独自の妥当な発音基準を創り出すための契機としての情報提供に過ぎない (小河原 2007:322)」と述べている。河野 (2009) は、この考え方に従い、実習生の音声教育活動を観察・分析した結果を「DOs」「DON'Ts」にまとめ、学習者に独自の基準を形成させるために必要な教師の能力について論じている。

このような発音の向上を学習者が自覚できるような学習管理方法の開発が必要である。これは、学習者に「良い発音」と「悪い発音」を適切なタイミングで比較させる FB を、パソコンにどう行わせれば良いかという問題である。

その一方で、試用後の学習者の感想でも、「ルールを明示的に教えてほしい」という意見があった。まったく言語知識ゼロの状態から自分で「気づく」段階に達するには長い時間がかかるが、効率という経済性を考慮した場合、その効果的な教師からの情報提供はどうあるべきか、今後も試用と観察を通じて、事例を蓄積していく必要がある。知識の教授という演繹的なアプローチと、学習者自身によるルールの発見という帰納的な要素をどのように取り入れた FB 方法を考えるかが、今後追究すべき、最も大きな課題である。

また、単なる例文の丸暗記ではない、何らかの「音声文法」を学習者が学習できるようなシラバスの策定が必要である。さらに、このようなソフト開発の目指すところとしては、最終的には、「発音練習」「韻律練習」ではなく、それも踏まえた「コミュニケーションができること」といえば何らかのロールプレイができるようになることが理想であり、もっと大きな単位での、コースウェアのデザインの追究が必要である。

注 本論文は、2011年11月18日のACTFL 2011 (米国, コロラドコンベンションセンター) での発表内容に加筆修正したものである。本研究は、科学研究費補助金基盤研究 (B) 「音声認識技術を取り入れた日本語発音自学システムの作成と試用」 (課題番号 21320094) の助成を受けた。

参考文献

- 天沼 寧・大坪一夫・水谷 修 (1978) 『日本語音声学』くろしお出版。
池田伸子 (2003) 『CALL 導入と開発と実践—日本語教育でのコンピュータの活用—』くろしお出版。

- 石井カルロス寿憲・西出隆二・峯松信明・広瀬啓吉 (2001) 「日本語のアクセント・イントネーションを対象とした発音教育システム構築に関する検討」『電子情報通信学会技術研究報告』100 (594), 33-40.
- 磯村一弘 (2009) 『国際交流基金日本語教授法シリーズ2 音声を教える』ひつじ書房.
- 岩田之男・清水政明・西村 尚 (2009) 「音声認識技術を利用した日本語発音練習システムの開発」『2009年度日本語教育学会秋季大会予稿集』.
- 大山 玄・三浦一郎 (1990) 「日本語学習者のプロソディーに関する研究」『日本語音声の韻律的特徴研究報告』3, 文部省重点領域研究「日本語音声における韻律的特徴の実態とその教育に関する総合的研究」研究成果報告書, 98-101.
- 小河原義朗 (1998) 「外国人日本語学習者の発音学習における自己モニター研究」東北大学文学部博士学位論文.
- 小河原義朗 (2007) 「発音指導場面における教師の役割—自己モニターの促進を目指した音声教育実践を事例にして—」『大学における日本語教育の構築と展開 大坪一夫教授古稀記念論文集』ひつじ書房, 311-326.
- 小河原義朗・池田佳子 (2009) 「日本語音声教育実践におけるコンピューター利用の可能性」水谷修監修, 河野俊之・小河原義朗編集『日本語教育の過去・現在・未来 第4巻 音声』凡人社, 139-164.
- 小河原義朗・河野俊之 (2009) 『日本語教師のための音声教育を考える本』アルク.
- 川上 葵 (1956) 「文頭のイントネーション」『国語学』25, 21-30.
- 河野俊之 (2009) 「音声教育に必要な教師の能力—教育実習をもとにして—」水谷修監修, 河野俊之・小河原義朗編集『日本語教育の過去・現在・未来 第4巻 音声』凡人社, 186-203.
- 河野俊之・串田真知子・築地伸美・松崎 寛 (2004) 『1日10分の発音練習』くろしお出版.
- 串田真知子・城生佰太郎・築地伸美・松崎 寛・劉銘傑 (1995) 「自然な日本語音声への効果的なアプローチ: プロソディーグラフ」『日本語教育』86, 39-51.
- 佐藤友則 (1995) 「単音と韻律が日本語音声の評価に与える影響力の比較」『世界の日本語教育』5, 国際交流基金, 139-154.
- 佐藤友則 (1998) 「韓国および台湾の日本語学習者のニーズ調査」『言語科学論集』2, 東北大学文学部言語科学専攻.
- 杉藤美代子 (1972) 「“おそ下り”考—動態測定による日本語アクセントの研究(その一)」『大阪樟蔭女子大学論集』10, 45-72.
- 鈴木雅之・峯松信明・広瀬啓吉 (2011) 「音声の構造的表象と多段階の重回帰を用いた外国語発音評価」『情報処理学会論文誌』52(5), 1899-1909.
- 土岐 哲 (1989) 「音声の指導」『講座日本語と日本語教育』第13巻, 明治書院, 111-138.
- 土岐 哲・村田水恵 (1989) 『外国人のための例文・問題シリーズ12 発音・聴解』荒竹出版.
- 広瀬啓吉 (2005) 「音声の韻律とCALL」『音声研究』9(2), 38-46.
- 日本語教育学会編 (1991) 『日本語教育機関におけるコース・デザイン』凡人社.
- 松崎 寛 (2002) 「リピートのとき学習者は何を考えて発音しているか」『広島大学

日本語教育研究』12, 33-42.

松崎 寛 (2005) 「音声教育教材」『新版日本語教育事典』大修館書店, 58-59.

羅徳安・峯松信明・鶴谷千春・山内 豊・広瀬啓吉 (2006) 「英語話者を対象とした日本語 CALL システムにおける発音評価」『日本音響学会秋季講演論文集』3-P-14, 363-364.

Shimizu et.al (2008) “Development of a Japanese Pronunciation Training System for Learners’ Speech Recognition.” World CALL 2008.