日本語 CALL システムのための学習者発音分析とその自動評価

羅 徳安[†]　峯松 信明[††]　鶴谷 千春[†††]　山内 豊[††††]　広瀬 啓吉[†]

[†]東京大学大学院情報理工学系研究科 〒113-0033 東京都文京区本郷 7-3-1
[††]東京大学大学院新領域創成科学研究科 〒277-8561 千葉県柏市柏の葉 5-1-5
[†††]オーストラリア・グリフィス大学言語学科 PO Box 3370 South Brisbane QLD 4101, Australia
[††††]東京国際大学商学部 〒350-1197 埼玉県川越市的場北 1-13-1

E-mail: {dean,mine,hirose}@ gavo.t.u-tokyo.ac.jp, C.Tsurutani@griffith.edu.au, yyama@tiu.ac.jp

あらまし 本研究では，豪英語の母語話者を対象とした，日本語発音とコミュニケーション能力の育成に焦点をあてた自律学習支援型システムの開発と評価を試みた。従来の CALL システム開発では，学習者群全体の誤り傾向及び教師群全体の評価戦略を統計分析し，およそ平均的な教師による平均的な学習者の発音指導を模擬する形態となっていた。本研究で検討する CALL システムは，日本語教師の使い勝手を考え，1)その日本語教師による発音評価を模擬すること，2)学習者の習熟度に応じて柔軟に発音誤り検出ができること，を念頭に置き，a)当該日本語教師自身が定義する発音エラーカテゴリ，及び，b)学習者毎の評定項目，更には，c)当該教師自身の学習者音声に対する評価戦略を事前に聴取実験などを通して取得し，それを模擬する形でシステムを開発し，評価実験を行なった。

キーワード CALL，ネットワーク文法，日本語特殊拍，発音評価，CART

Development of a CALL system for Australian learners of Japanese pronunciation

Dean Luo[†]　Nobuaki Minematsu[††]　Chiharu Tsurutani[†††]　Yutaka Yamauchi[††††]　Keikichi Hirose[†]

[†]Grad. School of Info. Sci. and Tech., Univ. of Tokyo, 7-3-1, Hongo, Bunkyo-ku, Tokyo, 113-0031 Japan
[††]Grad. School of Frontier Sciences, Univ. of Tokyo, 5-1-5, Kashiwanoha Kashiwa, Chiba, 277-8562 Japan
[†††]School of Languages and Linguistics, Griffith University PO Box 3370 SOUTH BRISBANE QLD, 4101 Australia
[††††]School of Business and Commerce, Tokyo International University 1-13-1 Matobakita, Kawagoe, Saitama, 350-1197 Japan

E-mail: {dean,mine,hirose}@ gavo.t.u-tokyo.ac.jp, C.Tsurutani@griffith.edu.au, yyama@tiu.ac.jp

Abstract In this report, we discuss the development of a CALL system for self assessment of Japanese pronunciation by English-speaking learners. Different from the conventional method for CALL system which focuses on average students, the system we developed focuses more on helping individual Japanese teachers by 1) simulating the teacher's strategies in pronunciation evaluation, and 2) being able to flexibly detect errors depending on the students' levels. In order to develop and evaluate such system, we a) categorized the pronunciation errors by the teacher, b) defined considered points of each student, and c) predicted the teacher's evaluation strategies by the preliminary transcription experiment of the students' pronunciation.

Keyword CALL, network grammar, Japanese pronunciation, pronunciation evaluation, CART

1. はじめに

1980 年代から日本語学習熱が世界的に高まり，現在では世界 127 カ国で 235 万人以上の人々が日本語を学習している[1]。これに伴い，高等教育機関の学習者，上級学習者の数も年々増え続けている一方，日本語教師の数が足りていないことが懸念されている[2]。一つの解決策とし，CALL システムが注目されている。

発音に対する自動採点やエラー検出機能を備える CALL システムはすでに開発されている[3]-[5]。しかし既存の CALL システムは，できるだけ大量の学生の発音データや教師の評価データを収集し，平均的な学習者像を前提とし，教師の平均的な評価戦略を模擬する形で開発が行なわれてきた[6][7]。本研究では従来とは異なり，日本語教師の使い勝手を考え，1)その日本語教師による発音評価を模擬すること，2)学習者の習熟度に応じて柔軟に発音検出ができること，を念頭に

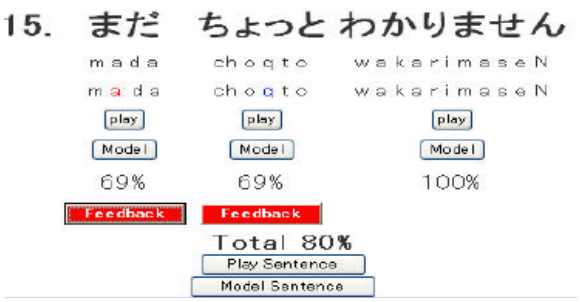

図.1 評価モード（学習者が見る画面例）

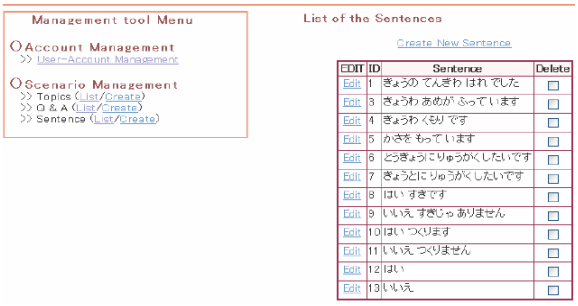

図. 2 管理モード（教師が見る画面例）

置き，a)当該日本語教師自身が定義する発音エラーカテゴリ，及び，b)学習者毎の評定項目，更には，c)学習者音声に対する評価戦略を事前に聴取実験などを通して取得し，それを模擬する形で，豪英語の母語話者を対象とした，新しい CALL システムの開発を行なった。このような開発は，日本語教育，教育工学，音声情報工学を専門とする研究者が相互協力しながら，教育現場のニーズに沿った開発を試みる必要があるが，本研究は各々の分野を専門とする研究者によって行なわれたものである。以下に，まず，開発したシステムの概要を示し，その後，CALL システム開発に向けての準備，音声収録，発音誤り検出モジュールの構築，システムインターフェイスの構築，評価実験について述べる。最後に，考察と今後の課題をまとめる。

2. 開発した CALL システムの概要

本研究で開発した CALL システムはクライアント・サーバ方式であり，学習者がクライアント PC で例文を読み上げて録音し，サーバに音声を転送する。サーバは音声分析サーバ，Web サーバ，データサーバとして機能し，読み上げ音声，分析結果，学習者 ID，フィードバック文を管理し，学習者が Web ページで発音評価結果とフィードバックを確認できるようになっている。また，日本語教師は管理モードにより，ユーザ管理，オリジナル教材作り，学生のレベルに応じた発音誤り検出モジュールの構築などの教材開発ができる。学生の音声評価画面を図.1，日本語教師のための管理画面を図.2 で示す。

3. エラーカテゴリの定義

3.1. 予測されうる豪英語話者による発音誤り

CALL システム開発を行なう場合，対象とする発音誤りを予め定義し，各々の検出対象について技術構築することになる。ここではまず，日本語教師（第三著者）の知識と経験を元に，分節音と韻律の面から考えられうる発音誤りを考察し，その後対象とする発音誤りについてデータ分析に基づいた絞り込みを行なう。

3.1.1. 分節音誤り

単純に英語と日本語の音素を比較することにより，考えられる誤りは以下の通りである。

＜子音＞

１）ɸ (両唇摩擦音―ふ) → f (歯摩擦音)：さいふ

２）ɾ (ら行―はじき音) →r (流音，接近音)／l (流音，側音)：いろいろ　らいねん

３）ts (歯茎破擦音) →s (歯茎摩擦音)：つぎ→すぎ

または，その混同で，[s]を[ts]と取り違えてしまうミスもある。

４）有声子音の無声化

英語の閉鎖音は語頭で強い呼気を伴うため，声帯の振動開始時間にずれがおこり，語頭の有声閉鎖音が無声子音に聞こえてしまうことがある。その他，誤って，強い呼気が与えられた場合も，同じ結果となる。

b (有声両唇閉鎖音)→p(無声両唇閉鎖音)：ぼく→ぽく

g (有声軟口蓋閉鎖音)→k (無声軟口蓋閉鎖音)：つぎ→つき*（[つ]を発音する呼気につられるため）

＜母音＞

１）母音の混同

日本語の５母音の境界線が英語の各母音と重なるため，次のような母音の混同が起こることがある。

o→ɔ　e→ɛ　　a→æ　　u→o

o→u　e→æ　　a→ɔ　　u→ə

２）つづりから来るミス

i→ai（ローマ字表記を思い浮かべ，[i] を[ai]と発音する。e.g. [kai] in “kite”）

３）連続母音の二重母音化

o.i → oi “おい” という語が glide oi に発音される。

a.i → ai * [.]はモーラ境界。なお ui, ei は存在しない。

４）母音脱落

ここでは，モーラ一つ分の長さが保たれず，次の音に融合してしまう場合を指す。

しよう → しょう　(shiyou → shoo)

ですよ → でしょう(desuyo → deshoo)

＜拗音＞

拗音“CjV”は，music [mju:zik], beautiful[bju:tifl], cute [kju:t] にみられるように英語にも存在する音節だが，母音は，長母音の[u]のみと言う制約がある。したがっ

て，日本語の中で頻度が高いが英語にない組み合わせであり困難とされる[rjV, kjV, gjV]でも，[rju, kju, gju]は，[a],[o]に比べて易しいと予測できる。普通，[r]の音素自体の難しさもあって，/rjV/が最も困難な拗音になる。拗音の誤りのパターンとしては，子音連続の誤りとして，通常，どちらかの音素を落として簡素化する次の二つが挙げられる。

CjV →CV／ jV [8]

それに加えて，モーラ感覚を習得していない学習者の場合，2 モーラの長さにまで延びて，CiJV とする場合も見られる。その他，単語内の位置によって，促音挿入 CjVQ や長母音化 CjVV も起こりうる。

3.1.2. 韻律的誤り

日本語のピッチアクセントに比べ，英語では，ストレスアクセントを置かれた音節が，重音節であることを原則とし，ピッチだけでなく長さ，大きさも増大する。従って，英語の韻律構造の転移から予測される長さの誤りは，ストレスアクセントパターンからの次の点が考えられる。

1) ストレスが置かれた日本語の音節は，長さ，ピッチが変化し，軽音節の場合，促音挿入，長母音化が起こり，重音節となる。
2) ストレスが置かれなかった音節は母音が弱音化したり，消えたりする。

また，無アクセントは英語に存在しないため，ピッチパターンも影響を受け，

3) 無アクセント型の語は，語のどこかにアクセントを置かれる。

アクセントの位置としては，英語のアクセントルールの最も一般的なパターンから推測し，

4) 語末から 2 番目の音節にアクセントが置かれやすい。

などの点も考慮する必要がある。これらによって引き起こされる誤りは，時間長の制約があまりない言語の話者にとって，モーラ時間の制約を無視した母音，子音の長さに関する誤りとなる。

3.2. 書起こしによる誤り傾向分析

以上の考察に対して，実際に豪英語話者の日本語音声データを解析することで，本研究で開発する CALL システムが対象とする発音誤りを定義する。

豪人日本語学習者初級 27 名，中級 19 名に，上記で考察した誤りの生起を考慮して作成された日本語例文 30 文を発音してもらった。それを，発音誤りの検出能力が比較的高い日本人母語話者を選んで，その全文の書き起こしを依頼した。30 文は，学習者のレベルに合わせ，仮名書きされたもので，学習者には，自分が考える正しい発音をするよう，指示した。書き起こし者は，音声工学や言語学を専攻とする学生から募集し，書き起こしの精度のレベルが音声研究者とほぼ一致する者を選んだ。各文を三人の母語話者が仮名表記によって書き起こし，二人以上の書き起こしが一致するものを正しいと判断した。書き起こしの結果は，音声研究者がランダムにピックアップしてチェックし，精度の低いもの(30 文中明らかなミスが 2 つ以上あるもの)は，(精度の比較的高かった)他の書き起こし者に再度書き起こさせ，結果の精度向上に務めた。書き起こしの最終結果の誤用数を集計し，5 回以上起こっているもので，偶発的でないと判断されたものを検出対象として定義した。

表 1 ：エラータイプとその頻度（高 1→低 19）

1	促音挿入 hito → hitto
2	短母音の長母音化 a → aa, u→uu
3	促音の脱落 itte→ ite
4	長母音の短母音化 aa → a
5	拗音の単純化 CyV → CV
6	拗音における母音挿入 CyV→ CiyV
7	子音置換 R (flap) → l (lateral approximant)
8	子音置換 ts(alveolar affricate)→s(alveolar fricative)
9	母音の置換 i →ai
10	鼻音の挿入，脱落
11	子音置換 F (bilabial fricative)→f(dental fricative)
12	語頭での有声子音の無声化，摩擦音の破裂音化 b→ p, s→ ts,
13	連続母音の二重母音化 o.i → oi
14	無声子音，文末の母音の無声化 shimas→ shimasu
15	母音脱落 shiyou → shoo
16	ください→ku:dsai
17	d→r の置換ミス
18	促音「っ」と長音の置換っ→ー
19	促音「っ」と長音の置換ー→っ

3.3. エラーパターンリスト

日本語教師の経験及び書起こしを元に，最終的に定義されたエラーパターンを，発生頻度の高いと思われるものから並べたものが表.1 である。だだし，ピッチ自体の誤りは，今回作成中のシステムでは対象外としたため，ここではリストに含めていない。

これとは別に，初，中，上級者相手に行なわれた，日本語読み上げ文（各々10 文ずつ）に関しても同様の書き起こしを行なった。得られた結果に対して，モーラの継続長に関するエラーとそれ以外に着眼して，学習者の習熟度別に，エラー頻度を集計した。

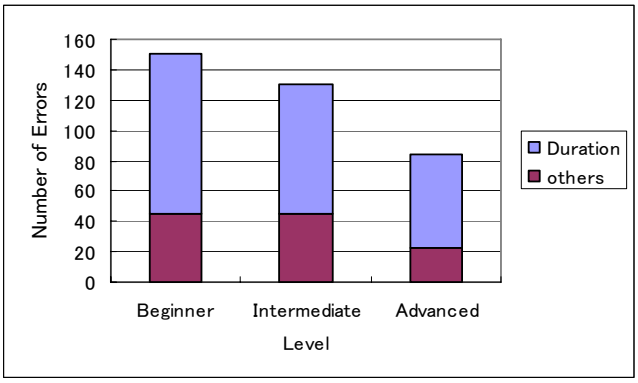

図. 3 学習者レベル別誤りの総数

結果を図.3 に示す。モーラ継続長に起因するエラーが各習熟度において 7 割以上あった[9]。継続長の情報は音響モデルの中に十分に反映されないため，通常の音声認識処理のみでは，最も頻度の高い発音誤りに対する十分な検出が期待できない可能性がある。

4. 学習者依存型誤り検出ネットワーク文法

発音誤りの検出は，従来より広く利用されているネットワーク文法を用いて行なった。例えば，前節で定義された発音誤りパターンより，「きいろいかさを」に対しては，図.4 に示すような発音のバリエーションが可能であり，各々に対してそれを受理するためのネットワークを構成する。そして音声認識技術によって，どのパスが実際の発声であったのかを同定することになる。当然，パスに含まれない発声であった場合にはその誤りは検出できない。任意の誤りを同定するためには，入力音声を「任意の平仮名列」としてネットワーク文法を構成すればよいが，自ずと検出率は低下する。そのため，適度に幅をもった文法が要求される。このようにして，誤り予測ネットワーク文法が作成されるが，更に，日本語教育者(第 3 著者)がネットワーク文法の各項目に制約を加えた。上記の誤りパターンが起こりうる箇所を日本語例文の中で指定することにより，ネットワークの肥大による誤検出を回避した。以下のように指定は文節末に置かれたが，1 と 2 に関しては，語末から 2 番目の音節に起こると言う制約があるため，助詞をよけて，語末に置いた（10～19 は，記号操作のため，A～I と書き換えられた)。

例：わたし(1,2)の　こーこーで(4JH)　いっしょに(3I,5,6)　しゃしん(2)を(5,6 A)　とりましょー(4,7)

予測される誤りパターンというのは，習熟度依存，更には学習者依存の傾向が高い。その意味において，現場の教師が学習者の習熟度に応じて適切な誤りパターンを埋め込む形でネットワーク文法を web 上で構成

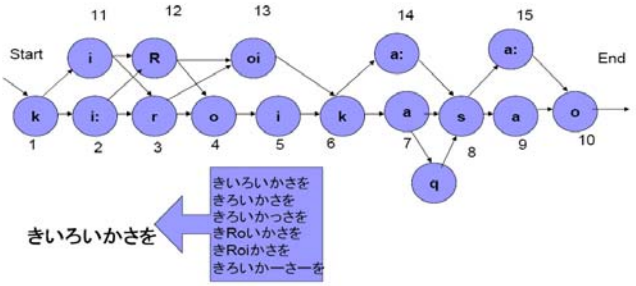

図. 4 ネットワーク文法

図. 5 ネットワーク文法生成 GUI

表.2 音響分析条件

サンプリング	16[kHz], 16bit
分析窓長	25[msec]，ハミング窓
分析間隔	10[msec]
特徴パラメータ	MFCC+ΔMFCC+Δpow（25 次元）

できるよう, GUI も整備した（図.5)。こうすることで，誤り検出精度の向上も期待できる。従来の e-learning 教材の場合，技術的問題の解決は開発者のタスクであり，現場の教師は構築されたシステムのユーザとして位置づけられることが多いが，本システムでは現場の教師の積極的な働きかけによっていくつかの技術的課題の解決を図ることにした。このような環境を整備することによって，教師の持つ，システムの内部動作についての知識も自ずと向上するものと期待される。

5. 音響モデルと後処理手法

本研究で同定した発音誤りは，本来あるべき音が無くなる（脱落)，不必要な音が入る（挿入)，異なる音で代用している（置換）の 3 種類に分類される。前者 2 種類の誤りに関しては, ATR 日本語多数話者音声データベース(話者 4,130 名，性別非依存) で学習した HMM モデルを用いる。一方, 置換誤りは, 例えば「ふ」という平仮名を豪英語の/fu/として発声するなどの誤りがある。この場合，日本語の/f/ではなく，豪語の/f/になっていることを指摘する必要があるため，豪語の音響モデルが必要となる。そこで，豪語音声データベース ANDOSL (Australia National Database Of Spoken Language) を用いて音響モデル構築を行なった。音響分析時のパラメータは表.21 に示す。音響モデルは 43

音素，豪語音素 47 音素（置換誤りを検出ために使う音素は/r/,/f/,/oi/,/ai/の 4 つのみ）からなる 5 状態 3 分布の left to right 型の HMM である。モノフォンモデルは，各状態につき 16 個の混合ガウス分布を持つ。トライフォンモデルは，2000 状態共有がなされており，各状態の混合ガウス分布は 16 である。なお，前節に述べたように，音声認識で使われる音素の音響モデルは，例えば，短母音と長母音の区別の精度が低くなる傾向がある。音響モデルは音声のスペクトル特性を主に用い，その音素が本来どの程度の時間長を有するのかについての特性は十分にモデル化されないため，長短の違いが音素の違いを生む場合に，その弁別精度が落ちてくる。予備実験の結果，学習者音声における非常に短い母音を長母音として認識する（或はその逆の）場合が比較的多く散見されたため，ここでは，以下のヒューリステックな手法を後処理的に用いて，音素継続長に関する検出誤りの低減を試みた：

1）長さ 120[msec]以上の短母音を長母音とみなす

2）長さが 100[msec]以下の長母音を短母音とみなす

3）長さが 60[msec]以下の促音挿入と判断したエラーを無視する

4）子音の長さに関しては，/p/,/t/,/k/ であれば，100[msec]以上，/s/,/sh/ は 150[msec]，それ以外の子音は 120[msec]以上であれば，促音が入っていると判断する。

6. 開発システムに対する種々の評価実験

開発したシステムに対して，3 つの観点から，1）発音誤り検出評価，2）教師による評定との相関分析，および，3）学習者による主観的評価を行った。以下各々について示す。

6.1. 発音誤り検出評価実験

6.1.1. 使用した音声データと正解ラベル

初級と中級学習者各々6 名ずつによる 30 文の読み上げ音声を対象とした。また，発音誤りの正解ラベルとしては，3 節で述べた書起こし作業で得られたラベルを使用した。

6.1.2. 使用したネットワーク文法

下記 3 つの文法を使用して，モノフォンとトライフォンで評価実験を行った。日本語教師によるマニアル操作による文法構築が行なわれているは G3 である。

G1： エラー対象となる音素が，単語のどこに位置するのかに関する制限を加えずに，19 種類のエラー全てを文全体に適応する。

G2： 対象音素の単語内位置に関する制限を加えて 19 種類のエラー全てを文全体で考慮する。

G3： 対象音素の単語内位置に関する制約を入れ，文節毎に，日本語教師が必要充分なエラーを指定する。

表.3 各文法のカバー率

	G1	G2	G3
カバー率	92.3%	90.5%	89.6%

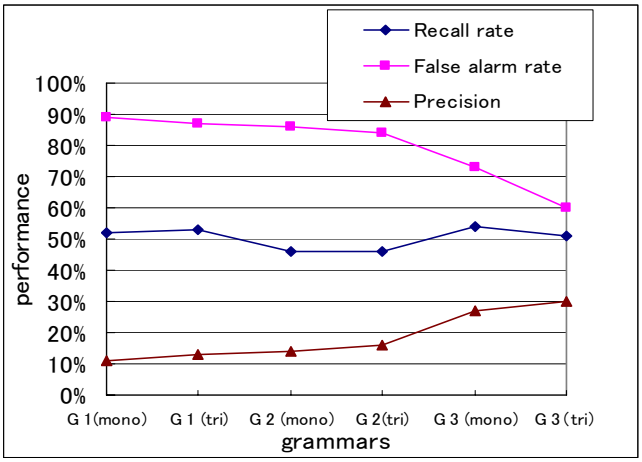

図. 6 システム評価結果（後処理を行わない）

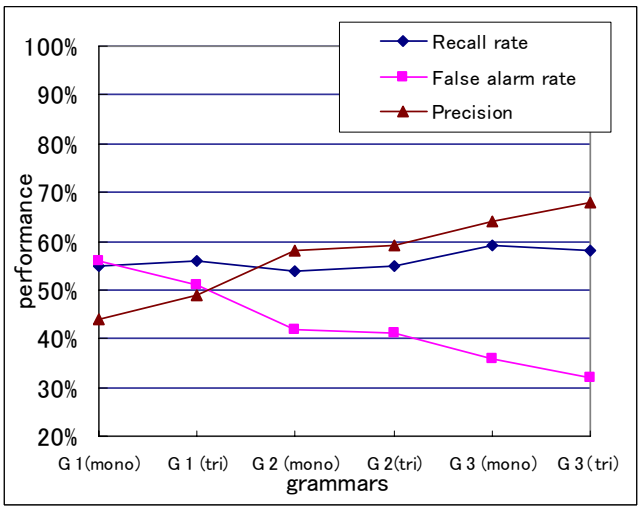

図. 7 システム評価結果（後処理を行った後）

6.1.3. システム精度

まず，3 つの文法の評価データの正解ラベルに対する発音エラーカバー率を求めた。結果を表.2 に示す。G1, G2, G3 となるに従い，カバー率は低下しているものの，低下の度合いは非常に小さい。ネットワーク化によって複数の音素群からの同定作業が行なわれることになり，例えば図.4 のように 1 つの文節に対して，複数の発音形式が得られる。G1 の場合，文節あたりの発音形式数は 182，G3 の場合は 7 である。発音形式数を大幅に低減していることが分かる。

次に，G1～G3 の各文法を用いて，後処理を行う前と行った後の検出精度をそれぞれ図.6，図.7 に示す。

表.3 アンケート調査結果

	平均値	標準偏差
1-1Recording	1.36	.497
1-2Results	1.86	.770
2) Fun	1.29	.469
3) Weak points	1.57	.756
4)How to improve	1.86	.663
5) Usefulness	1.46	.660

いずれの場合，G3 を導入することによって，精度を向上していることが分かる。また，後処理によって，検出精度，特に false alarm rate が大幅に低減することは，3.3 節で検討したように，音響モデルの中に十分反映されない継続長の情報に関する発音エラーは，通常の音声認識処理のみでは検出することが困難であるといえる。

6.2. 教師による評定との相関分析

CART を使用して日本語教師による文節単位の評価点の推測を試験的に試みた。CART の説明関数として，当該文節で起きるエラー数，音響尤度，母音継続長の平均及び分散[10]，単語信頼度[11]を用いた。話者 11 名の読み上げ音声 586 個文節の日本語教師による採点データを用いて，テキスト・話者オープン評価実験を行い，0.75 の平均相関を得た。説明変数などより詳細な検討を行なう必要がある。

6.3. 学習者による主観的評価

本システムの有効性を調べるため，初級 4 人，中級 5 人，上級 5 人の豪人日本語学習者に実際に，使用してもらった。中上級者には 2 つのレベルを初級には初級のレベルのみを試してもらい，アンケート用紙で以下の項目についてシステムの評価を行った。

1）プログラムの使いやすさ
１－１．録音部分
１－２．評価部分
2）プログラムは面白かったか
3）自分の発音の弱点が分かったか
4）フィードバックは分かり易かったか。
5）プログラムは自分の発音学習に役に立ったか。

1～5 で示してもらった回答を集計，平均し，表.3 のような結果が得られた。ただし，点数が低いほうが高評価に相当する。

プログラムそのものの有益度，使いやすさに関してはどれも高い支持が得られた。

しかし，1-2 と 4）が若干低めなのは，前節分析したように，検出制度，特に false alarm 率の低減をまた改良する余地があることを示している。

7. 考察と今後の課題

システムを利用する教師の教育・評価戦略をシステム構築に反映させる形で，そして，個々の学習者の習熟度に応じて教材を作成する形態の日本語 CALL システムを開発した。前者に対しては，当該教師の発音評価の様子を実際に収集し，それを模擬する形で採点モジュールを構築し，後者に関しては，ネットワーク文法を含めた教材開発が容易に出来るよう，GUI を整備した。また，音素継続長に基づく後処理を導入することによって，検出精度の向上を試みた。今後は，システムにおける音素識別の改良を検討しながら，評価点算出における教師・システム間の「ずれ」に対する主原因を解明するために，分析実験を通して，システム信頼度を向上させる予定である。

文　献

[1] 国際交流基金，“海外の日本語教育の現状－日本語教育機関調査・2003 年－(概要版)”，2003.

[2] 岡本佐智子，“日本語教師養成の現状と課題”，北海道文教大学論集第六号, 2003.

[3] A. Raux *et al.*, “Automatic Intelligibility Assessment and Diagnosis of Critical Pronunciation Errors for Computer-Assisted Pronunciation Learning,” Proc. ICSLP, pp.737-740, 2002.

[4] H. Franco *et al*, “Automatic pronunciation scoring for language instruction,” Proc. ICASSP, pp.1471-1474, 1997.

[5] 小俣修一他，“韓国人の日本語発声に対する発音誤りの自動検出・評定法”，日本音響学会 2003 年秋季研究発表会講演論文集,2-8-8,pp.117-20, 2003.

[6] 峯松信明他，“英語 CALL 構築を目的とした日本人及び米国人による読み上げ英語音声データベースの構築”，日本教育工学論文誌，Vol.27, No.3, pp.259-272 (2003.12)

[7] 峯松信明他，“外国語学習用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, (2003.06)

[8] C. Tsurutani, “Acquisition of Yo-on (Japanese contracted sounds) in L1 and L2 phonology in Second Language,” vol 3. 27-47, 2004.

[9] 鶴谷千春他，“日本語音声の難易度―「英語を母語とする学習者の場合」”日本音声学会全国大会予稿集，pp.51-56，2006.

[10] 毛利太郎他，“母音持続時間に着目した日本語 CALL システムにおける矯正フィードバックの考察”，電子情報通信学会音声研究会，SP2002-191, pp.73-78 (2003-3)

[11] Akinobu Lee *et al*, "Real-time word confidence scoring using local posterior probabilities on tree trellis search," Proc. ICASSP2004, Vol.I, pp.793-796, May 2004.