音声認識技術を利用した英会話CAIシステム

中川聖一[†] Allan A. Reyes[†,☆]
鈴木英之[†,☆☆] 谷口泰広[†]

本論文では，音声認識技術を利用した英会話CAIシステムについて述べる．これは，システムが，学習者の発話を自動音声認識により理解し，待ち時間なしで適切な応答を音声で出力し，対話を進めることにより，スピーキングとヒアリングの能力を高めるものである．まず，日本人の発声した英語の音声認識を行うためには日本人の英語発音モデルを用いる必要のあることを示す．次に，評価実験として4人の日本人男性にこの英会話CAIシステムを使用してもらった評価実験結果について述べる．使用前と使用後のスピーキングとヒアリングの能力の差を比較したところ，全員に能力向上がみられた．またアンケートの結果，本システムを引き続き利用したいとか，システムの応答時間はちょうど良いといった意見が多く得られた．

An English Conversation CAI System Using Speech Recognition Technology

SEIICHI NAKAGAWA,[†] ALLAN A. REYES,[†,☆] HIDEYUKI SUZUKI[†,☆☆] and YASUHIRO TANIGUCHI [†]

This paper describes an English conversation CAI using speech recognition techniques. This system was intended to recognize user's utterances and to respond to him properly according to the recognized results. These functions, of course, require the real-time speech recognition/response performance because time consuming responses may degrade smooth conversation and discourage the users to learn with the system. In the case of a learner with unskilled pronunciation, because of differences in the phonemic system between his mother tongue and the second language, the speech recognition system cannot run normally. After these improvements, evaluation experiments were conducted. The results indicate that learners' ability in speaking and in listening to English is improved by using the system.

1. はじめに

近年，音声を利用した第2言語学習用CAI（Computer Assisted Instruction）システムの構築がなされてきている．最近はCALL（Computer Assisted Language Learning）と呼ばれ，外国語教育者からも関心が寄せられている．Bernsteinらはネイティブ発話から学習した不特定話者HMMにより日本人の英語発話を評価した[1]．また，浜田らは発音の評価尺度を設けて日本人の英単語発話を評価している[2]．上記2つのシステムとも，そのシステムによる発音評価と，人間の英語教師による発音の評価値との相関係数は高い数値を示しており，信頼性も十分高い．

Hillerらは第2言語学習者のイントネーション，リズム，母音の音声について指導を行うシステムを構築している[3]．また，Mostowらはディスプレイに物語を表示し，児童が学習することによって文章の理解，正読能力を獲得するシステムを開発している[4]．児童が文章を誤って読んだ場合にはその箇所を指摘し，途中で中断した場合にはその単語の読み方を教える．

また，様々な英会話練習用CAIシステムが開発されているが，これらのシステムの多くは，入力手段としてキーボードやマウスのみを用いており，会話の技術を身に付けるために欠かせない音声による入力機能を持たないので，十分な臨場感を学習者に提供できていない．そこで，我々の研究室では，山本ら[5]や仏Auralog社[6]などと同じく，音声入力が可能な英会話CAIシステムの構築を進めてきた．しかし，音声認識技術を第2言語CAIシステムに利用する場合，第

† 豊橋技術科学大学情報工学系
Department of Information and Computer Sciences, Toyohashi University of Technology
☆ 現在，株式会社アイピックス
☆☆ 現在，ソフトウェア開発株式会社

2言語を習得しようとする学習者の発声が未熟な段階では，ネイティブ発声者の音韻モデルによる音声認識システムでは学習者の入力に対しては正常に動作しない問題がある．このため，文献5)では，不特定話者の認識装置を用いていることから，最初は英文の暗記学習を中心にし，その後対話学習に移行するが，対話の流れの自由度が比較的少ない発話を扱っている．文献6)では，学習者ごとに音声を登録する方法をとっている．

我々は，音声による第2言語学習用CAIとして発音，文法，会話習得を目的としたCAIシステムの構築を行ってきた[7)~11)]．CAIシステムの音声認識部では，ネイティブ発話の音韻モデルをそのまま使うのではなく，それを学習者の母国語の発音体系（本論文では，日本人の日本語発音体系）の影響を考慮して適応化したモデルを用いて，学習者の発話の認識をしている．本研究で開発したCAIシステムは音声認識技術を駆使して，ロールプレイングゲーム方式で，種々の場面での英語を用いたコミュニケーション技術を，効率的に学習させることを目的としている．

2. 音声認識法

英会話のCAIでは，学習者が実際に英文を発声し，その内容に応じたシステムからの応答を理解し，それに対して学習者が新たに発声するという繰返しが重要と考えられる．このためには，学習者の発声した音声を正しく認識する技術が必要となる．

2.1 音韻，単語，文法

英語音韻は音韻の分類法を変えることにより様々な音韻体系を取りうる．英語音韻は音韻学的には44種類といわれており，異音・無音を含めると60種類存在する．本実験では60種類の音韻から無音音韻8種を取り除いた52音韻体系（注・認識システムが無音区間を無視して処理しているため），および，音韻論的に同種と見なせる音韻をさらにまとめた39音韻体系で英語音韻を分類している[12)]．そこで，音韻モデルは52，39音韻体系の場合について作成している．

発音記号からなる単語辞書によって音韻モデルを連結することにより単語モデルを構築する．

発話文を認識するために，システムが受理する文（間違った文をわざと含む）の集合は文脈自由文法で表現される．間違った文を含むことによって，学習者は正しい文を選択するために，システムの応答内容を理解しなければならず，正しい答えを学習者に覚えさせることができ，学習効果が期待できる．

2.2 認識モデル

認識実験に使用する音韻モデルは離散継続時間制御付き連続HMMで，4状態のleft-to-rightの単一ガウス分布型HMMである．学習者の英語の発音はネイティブな発音と大きく異なるため，学習者用の認識モデルを用いる必要がある（たとえば，日本人の英語の発音は日本語的な子音–母音の連鎖のような発声になり，また母音を英語のように細かく区別して発声できない）．

認識実験では，学習者の母国語発音体系による影響を考慮して適応化したモデルの性能を評価するため，初期モデルと，適応化に利用する発話文数（100，300，600文）が異なる3種類の適応化音韻モデル，さらに評価用話者により話者適応化したモデルの計5種類のモデルを使用している．

実験では，連続音声認識の標準的手法であるOne Pass Viterbiデコーディング法を用いて文認識を行う[13)]．One Pass法は有限状態オートマトンによる構文制御（言語的な制約）が可能であるため採用した．対話の流れに応じて，受理可能な文集合を表現する文脈自由文法も変更していく．入力音声を文脈自由文法を用いて解析するとき，文脈自由文法を逐次有限状態オートマトンに展開する[14)]．連続音声認識処理による部分文の照合スコアに応じて文脈自由文法から単語予測を行って動的に有限オートマトンに展開し，その結果を構文制御に用いる．

2.3 モデルの適応化法と音声資料

従来のCAIシステムの音声認識部ではネイティブ発話の音韻モデルを使用して学習者の発話の認識をしていた．しかし，第2言語を習得しようとする学習者の発音が未熟な段階では，ネイティブ発話者の音韻モデルによる音声認識システムでは正常に動作しない．そこで，第2言語学習者用音韻HMMの実現はすでに学習されているネイティブ発話から学習された不特定話者音韻モデルを標準（初期モデル）として適応化用の発話で追加的な学習により行う．適応化法は，最大事後確率推定法（MAP推定）を用いた逐次連結学習で行い，適応化文に対する音韻ラベル系列だけを与えることで，1文ごとに文発話データから連続出力分布型HMMのパラメータを求める[15)]．

こうすることにより，学習者の未熟な発音の音声パラメータをネイティブな発声者の音声パラメータに適応化できる．学習者の発音が未熟でも，その発音をシステムが容易に理解でき，従来に比べ，システムとの対話がスムーズになる．

実験の対象とする第2言語の発声は日本人の英語

表1 使用する発話音声データ
Table 1 Speech data.

学習者用	ネイティブ326名×8文＝2680発話 ネイティブ10名×30文＝300発話
日本人適応化用	日本人20名×30文－600発話
評価用	日本人10名×50文－500発話 American native5名×50文＝250発話
話者適応化用	日本人1話者当り20文（10名） American native 1話者当り20文（5名）

表2 音声データの分析条件
Table 2 Condition of speech analysis.

サンプリング周波数	12 kHz
窓関数（ハミング）	21.33 msec（256 point）
フレーム周期	8 msec（96 point）
分析条件	14次LPC分析
特徴パラメータ	10次LPCメルケプストラム係数 10次回帰係数

である．音声資料は標準（初期）モデル作成用のネイティブな発声者の学習用音声データと，英語音韻モデルの日本人適応化用のデータ，認識実験で使用する評価用データ，さらに，話者適応化用のデータである．

表1に使用する音声データの内訳を示す．ここで，表1の学習者用の欄はネイティブな英語発声者の音韻モデル作成用のデータを示す．ネイティブの英語音韻モデル（52・39音韻体系）はTIMITデータベースの326人の男性話者音韻データで学習したものである．さらに，英語音韻モデルはネイティブ男性話者10名による発話300文を用いて追加学習している．この追加学習データは本研究室の録音室で録音されたものであり，発話環境の適応化のために使用された．この環境適応化音韻モデルを第2言語適応化前の初期化モデルとしている．表1の英語音韻モデルの日本人適応化に使用した発声データセットはTIMITデータベーステキストから選択した450文からのものである．また，評価用，話者適応化用の文集合は簡単な英文のオートマトン制御図を元に作成した．

なおこれらの音声データは**表2**の条件で音響処理を行った．

3. 音声認識モデルの評価実験

3.1 日本人の英語学習HMMによる音声認識

評価実験において，前節と同様に話者適応化は，日本人の場合，適応化前のモデルを日本人の発話600文で適応化されたものを用い，American nativeの場合，日本人の発話で適応化されていないモデル（初期モデル）を用いて行っている．

5種類の音韻モデルによる連続音韻認識結果，文認識結果を**表3**，**表4**，**表5**にそれぞれ示す．また，連続音韻認識結果の音韻正解率をまとめたものを**図1**に示す．認識評価用に用いた文は，語彙数250，パープレキシティ約8の有限状態オートマトンで表現される文生成器（約1419万文が生成・受理可能）がランダムに生成した50文である．連続音韻認識は，単語や文法の知識を用いずに任意の音韻が連続可能という条件で行ったものである．表3，表4において，置換とは正しい音韻の代わりに間違った音韻を認識したとき，挿入とは本来は存在しない音韻を出力したとき，脱落とは本来あるべき音韻が認識結果にないことである．またセグメンテーション率とは，次式で定義されたものである．

$$\text{セグメンテーション率} = \frac{\text{入力音韻数} - \text{挿入音韻数} - \text{脱落音韻数}}{\text{入力音韻数}}$$

表3，表4，図1では，日本人用に英語音韻モデルを適応化するのに用いた文数が多くなるに従い，日本人の英語発声の音韻正解率は向上していくことが分かる．また，当然ながら日本人用に適応化されたモデルでは相対的にネイティブの発話に対しては正解率は低下していくことが分かる．これは英語音韻モデルが，日本人用に適応されることにより，次第に日本人の音韻体系に接近していくことにより生じるものとして考えられる．600文を適応化に使用したとき，日本人の英語発話の正解率（46.8%）は，ネイティブの英語発話のネイティブ発話用モデルによる正解率（52.9%）に匹敵している．

このように，日本人向きに英語音韻モデルを作成することにより発音の学習の不十分な日本人に対しても，英語による会話練習システムが構築できる．さらに，話者ごとにモデルを話者適応化すると，正解率は61.7%に向上した．

また，表5の文認識でも認識率は向上している．相対的に音韻モデルの日本人への適応化度が大きいほど，ネイティブ発話の認識性能は低下しており，モデルが日本人の音韻体系に近づいていることが分かる．また52音韻体系と比較して，39音韻体系の結果が良いのも，ある意味で日本人寄りの音韻体系に変化しているといえる．話者適応化すれば認識率はさらに高くなり，第2言語会話CAI用に対しては認識対象文数（受理可能文数）は数百から数千あればよく（我々のシステムでは数十文），十分な文認識率が得られるものと考えられる．

表3 日本人適応化モデルによる英語連続音韻認識結果（52音韻体系）
Table 3 Result of English continuous phonetic recognition using models adapted for Japanese (52 phonemes).

(a) 日本人の英語発話（話者10名の平均）

適応化文数	0	100	300	600	話者適応化
正解（%）	37.2	42.9	44.2	45.2	60.7
置換（%）	49.3	46.8	46.3	45.1	32.6
挿入（%）	69.6	61.2	54.1	52.0	41.3
脱落（%）	13.5	10.3	9.5	9.7	6.7
Seg.（%）	16.9	28.5	36.4	38.3	52.0

(b) American nativeの英語発話（話者5名の平均）

適応化文数	0	100	300	600	話者適応化
正解（%）	48.3	40.2	39.8	39.7	67.9
置換（%）	43.5	48.1	48.4	49.6	26.5
挿入（%）	38.4	47.9	45.8	41.3	25.3
脱落（%）	8.2	11.6	11.7	10.7	5.6
Seg.（%）	53.5	40.5	42.5	48.0	69.2

表4 日本人適応化モデルによる英語連続音韻認識結果（39音韻体系）
Table 4 Result of English continuous phonetic recognition using models adapted for Japanese (39 phonemes).

(a) 日本人の英語発話（話者10名の平均）

適応化文数	0	100	300	600	話者適応化
正解（%）	40.1	45.6	46.3	46.8	61.7
置換（%）	46.3	45.1	44.0	44.0	31.8
挿入（%）	70.5	62.4	55.3	51.6	42.3
脱落（%）	13.7	9.3	9.7	9.2	6.6
Seg.（%）	15.9	28.4	35.0	39.2	51.2

(b) American nativeの英語発話（話者5名の平均）

適応化文数	0	100	300	600	話者適応化
正解（%）	52.9	43.8	43.6	43.0	69.6
置換（%）	39.6	44.9	45.7	46.1	21.1
挿入（%）	40.9	49.8	45.0	42.0	27.3
脱落（%）	7.8	11.3	10.7	10.9	5.4
Seg.（%）	51.3	38.9	44.4	47.1	67.3

しかし，ネイティブ発話の場合にもあてはまるが，音韻認識で挿入による誤りが多く，セグメンテーション率が非常に悪い．これは次節で検討する．

3.2 継続時間長分布の考察

前節の認識実験において，文認識率の低い原因として，日本人の発話がネイティブ発話に比較して発話時間が長いからであると考えられる．**表6** にネイティブと日本人の1音韻あたりのフレーム長を調べた結果を示す．音声パワーが閾値（1000）以上のフレームを音声区間と見なして調査をした．これらの結果より，平均フレーム長は，ネイティブに比べて日本人の英語音韻では約1.3倍であることが分かった．

表5 日本人適応化モデルによる英語文認識結果（52音韻体系）
Table 5 Result of English sentence recognition using models adapted for Japanese (52 phonemes).

(a) 日本人の英語発話（話者10名の平均）

適応化文数	0	100	300	600	話者適応化
正解（%）	7.6	26.2	29.6	30.4	51.6

(b) American nativeの英語発話（話者5名の平均）

適応化文数	0	100	300	600	話者適応化
正解（%）	43.2	28.8	31.2	25.6	75.2

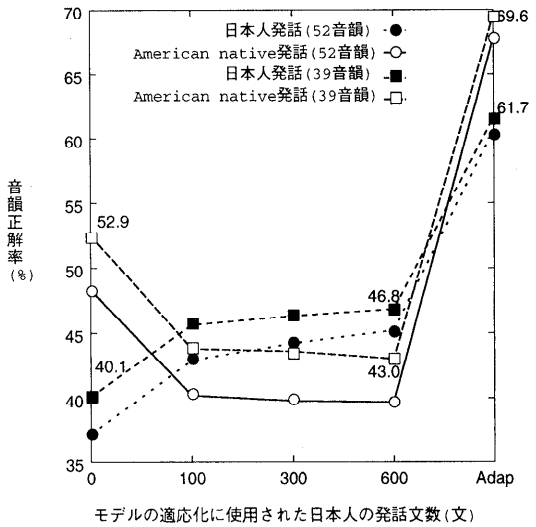


図1 日本人用適応化英語音韻モデルによる日本人とAmerican native発話の音韻正解率
Fig. 1 Correct rate of phonetic models adapted for Japanese or American native.

そこで適応化音韻モデルの継続時間分布を線形に伸長して認識に用いる実験を行った．継続時間分布を認識時に伸長する場合と，継続時間分布を線形に伸長後適応化学習する場合について実験する．評価実験はまず，日本人の英語発話について連続音韻認識を行った．

表7 に，日本人の英語発話の連続音韻認識結果を示す．表より明白であるのは，継続時間を線形に伸長することにより，挿入による誤りを抑えることができ，結果としてセグメンテーション率が改善されていることである．特に，表7の日本人の英語発話の結果は前節の結果と比較して，セグメンテーション率が24.5%向上した．また，伸長度1.3では適応化学習前に分布を伸長した場合の方が認識前に伸長するより効果があった（伸長度1.5のときは効果なし）．ただし，脱落による誤りが増えており，総合的に正解率の向上は望めなかった．これは，継続時間分布を線形に伸長したことにより認識音韻候補数が抑えられるためと考

表6 ネイティブと日本人発話の1音韻あたりのフレーム長
Table 6 Length per phoneme of American native and Japanese.

発話者	ネイティブ（10人）	日本人（20人）
発話文数	300	600
平均フレーム長	9.5	12.8

表7 継続時間分布を伸長したモデルによる英語連続音韻認識結果（39音韻体系）
Table 7 Recognition result of English continuous phoneme by duration model (39 phonemes).

日本人の発話（話者10名の平均）

適応化文数	600				
継続時間分布の延長度	1.0	認識前に延長		学習前に延長	
		1.3	1.5	1.3	1.5
正解（%）	46.8	44.4	43.0	43.2	43.3
置換（%）	44.1	43.3	42.3	42.7	42.2
挿入（%）	51.6	33.1	21.7	24.0	21.8
脱落（%）	9.2	12.3	14.7	14.0	14.4
Seg.（%）	39.2	54.6	63.5	62.0	63.7

表8 継続時間分布を伸長したモデルによる英語文認識結果（39音韻体系）
Table 8 Recognition result of English sentence by duration model (39 phonemes).

日本人の発話（話者10名の平均）

適応化文数	0	600				
継続時間分布の伸長度	1.0	1.0	学習前に伸長		3乗	
			1.3	1.5	1.3	1.5
正解（%）	7.6	26.8	33.8	31.0	48.0	47.4

えられる．

また，この継続時間分布を伸長したモデルによる文認識結果について，**表8**に示す．さらに文認識結果をまとめたものとネイティブ発話の文認識率を，**図2**に示す．文認識では継続時間の確率を3乗したモデルの文認識結果も示している．表8より文認識率は向上している．継続時間の確率を3乗した場合，文認識率はさらに向上した（48.0%）．これは表7の音韻認識の挿入による誤りが改善されたことによるセグメンテーション率の向上に起因するものと考えられる．また，継続時間の確率を3乗することにより音韻の継続時間の影響が強調され，さらに文認識率が向上することが分かる．しかし，図2より，まだネイティブ発話の文認識率には及ばない結果となっている．次節で述べるような英会話練習システムでは，システム側のユーザ発話の制御により，認識対象（受理可能）文数は約10～30文なので，この程度の性能でも文認識率は90%以上が得られる（5.1節参照）．

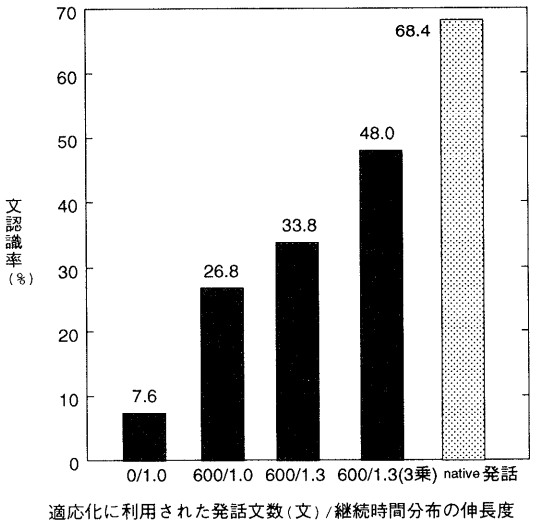


図2 継続時間分布を伸長したモデルによる日本人発話の英語文認識率
Fig. 2 Recognition result of English sentence for Japanese speech by using duration model.

4. 英会話CAIシステム

4.1 システムの構成

本システムは，ユーザとの対話を実現しており，基本的な動作の流れは次のようである．まず，ユーザは好みのシナリオを選び，その話題について会話の練習を行うが，このとき，会話の実際の状況が画面に表示される．本CAIシステムの内容は選択回答からなっているのでユーザは画面に示されている次に発話すべき文集合の中から1つ選び，発声する．本システムはユーザがどの文を発声したのかを認識し，認識された文により，次に続く会話を決定する．

本システムの構成を**図3**に示す．本システムは音声入力部，音声認識部，会話制御部とシステム出力部からなっており，ユーザの発話はマイクを通して，音声入力部に入力される．音声入力部は音声分析の処理を施し，分析された音声データを音声認識部に渡す．音声認識部では文脈自由文法の書き換え規則で与えられる文法を参照しながら，可能な文集合の中から入力音声データに最もマッチするような文を探し，その認識結果を入力発話文として，会話制御部に与える．会話制御部は認識部からの入力文をもとに，次の会話のやりとりにおける認識の対象となる文集合を示す文法に切り替え，次のシステムの発話と会話の分岐先を決定し，その情報を出力部に渡す．出力部は会話制御部の指示に従い，システム側の発話文の音声をスピーカに流し，同時に会話に関する映像とユーザが次に発声し

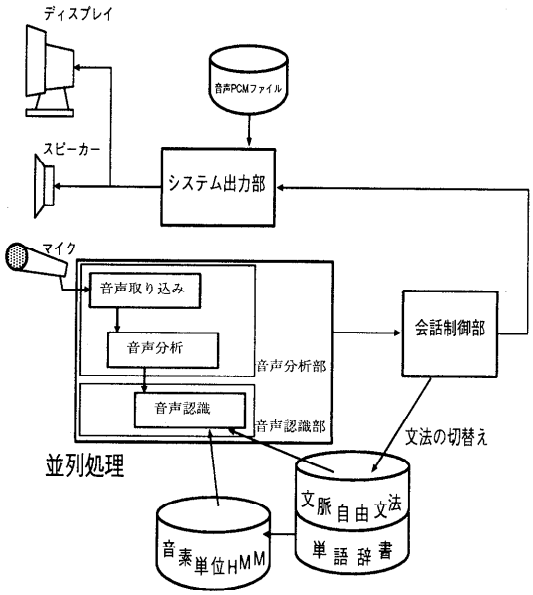


図3 英会話CAIシステムの構成
Fig. 3 Construction of English conversation CAI system.

```
System:  What is the purpose of your visit?
User:
     1.   Sightseeing.
     2.   (For) pleasure.
     3.   I'm here on vacation.
     4.   (I'm here on) business.
     5.   I'm visiting  /a /   /friend  /.
                        /my/   /relative/
                               /brother /
                               /sister  /
     6.   I'm here for two weeks.
     7.   I'm from Japan.
     8.   Pardon (me).
     9.   Excuse (me).
    10.   Please say it again.
    11.   I beg your pardon.
    12.   I don't understand.
User Response:  Sightseeing.
System:  Where do you plan to go?
User:
     1.   I don't know (yet).
     2.   ((I'm going) to)  /San Francisco/.
                            /Disneyland   /
                            /Florida      /
     3.   (I'm here for) one week.
     4.   I'm staying with my friend.
     5.   Pardon (me).
     6.   Excuse (me).
     7.   Please say it again.
     8.   I beg your pardon.
     9.   I don't understand.
```

図4 本CAIシステムの会話の一例
Fig. 4 An example of conversation of our CAI system.

得る文集合をディスプレイに表示する．

これらの音声取込み・音声分析・音声認識は並列に処理されており，ほぼ実時間での応答が可能となっている．

4.2 シナリオ

本英会話CAIシステムは現在のところ，1) Immigration and Customs（入国審査・税関），2) Hotel Check-in（ホテルのチェックイン），3) In the Street（路上）の3つの場面の会話におけるものである．これらのシナリオでそれぞれ，1) 入国目的，滞在先，滞在期間，免税範囲などに関する会話，2) ホテルの料金，施設，予約の有無，希望する部屋の種類などに関する会話，3) 場所の行き方とバスの乗り方を訪ねるときや緊急時の会話などが練習できる．

なお，“Never mind.”，“Thank you.”といった英語の典型的な表現もそれらのシナリオの回答群の中に含まれている．

4.3 システムの様々な機能

図4に示す本英会話CAIシステムのImmigration and Customs（入国審査・税関）のシナリオからの会話の一例を使って，本システムの様々な機能を説明する．同図の最初の会話ではシステムがユーザに入国の目的を訪ねており，ユーザに示される回答群は1から12までの番号が前に付いている文である．図中の（ ）はオプショナルな語句，/ /は縦方向に並んでいる語句の中から1つを選択することを示す．これから，最初の文でも，ユーザは本システムの会話の内容は上で述べたように選択回答からなっているが，回答群の中に不適切なものが入っている場合があり，ユーザがそれを発声すると本システムはその誤りを指摘する．たとえば，図4の回答群の“I'm from Japan.”と“I'm here for two weeks.”は入国の目的という質問に対しては，不適切である．また，図4にはないが，“one weeks”といった文法的に誤っている文も回答群の中に含まれている場合があり，このような文を発声するとシステムに注意される．その他，ユーザの発音が悪かったり，あるいは回答群以外の回答を発声した場合に対応するために，本システムにはReject機能が備えられている．これは音声認識結果の尤度を用いて実現している[16]．逆に，ユーザにとって，システムの質問が分かりにくかったり，あるいは聞き取れなかった場合についてユーザは“Pardon me.”とか“Excuse me.”などと発声するとPardon機能が起動し，システムはより簡単な表現で，あるいはゆっくりとした口調（現在のところ，スピードが0.8倍になる）で同じ内容の質問を繰り返す．

また，テストモードの設定があり，このモードではシステムが定めた条件に会話を進めないといけない．たとえば，図4の会話において，システムは入国の目的を「観光」と定めたならば，ユーザは“Sightseeing.”，

表9 文認識精度
Table 9 Sentence recognition accuracy.

	モデル	native		日本人用	
話者	入力文数	正解文数	認識率[%]	正解文数	認識率[%]
1	41	24	59	37	90
2	39	30	77	36	92
計	80	54	68	73	91

"(For) pleasure.", あるいは "I'm here on vacation." のいずれかを発声しない限り，システムは同じ質問を繰り返し，会話は次の対話へ進まない．もちろん，ユーザが自分の役割の条件を設定することもできる．

設定できる項目は，Immigration and Customs（入国審査・税関）の場合は，入国目的，滞在期間，関税品の品数などであり，Hotel Check-in（ホテルのチェックイン）の場合は，予約の有無，滞在期間，希望する部屋，人数などである．

5. システムの定量的効果

英会話学習をするときの，本英会話CAIシステムの効果を調べるために，本システム全体に対する評価を行った．

評価方法としては，日本人成人男性11人を，a）音声入力によるCAI（4人），b）キーボード入力（番号選択）によるCAI（2人），c）CAIを使用しないでテキストによる独習（5人）の3つのグループに分ける．グループaは1日約30分間CAIを使用し，これを5日間続ける．そして，スピーキングテスト，ヒアリングテストを初日のCAI開始前と毎日のCAI終了後，そして，5日目（最終日）より10日後と30日後に行う．グループbもグループaと同様の評価を行い，学習者の発話による学習効果を調べた．グループcについても同様にテストを行う．

5.1 音声認識部の評価

本システムの音声認識部では特徴パラメータはメルケプストラムとその回帰係数を使用している．計算機雑音のある実験室でオンライン認識実験を行い，その認識精度を2人の男性話者で調べた結果を**表9**に示す．native英語モデルの使用時の文認識率は68%，日本人用英語モデルの使用時の文認識率は91%であった．また，同じ音声をnativeのモデルと日本人用のモデルの2つのモデルを併用で認識したとき，どちらのモデルを使用されたか（尤度が高くなった方が使用される）の割合の比は，日本人用のモデルが86%，nativeのモデルが14%であった．

認識時間は入力発声終了の検出に要する入力終了後300 msec以内で，すべて終了できた．

5.2 ヒアリングテスト

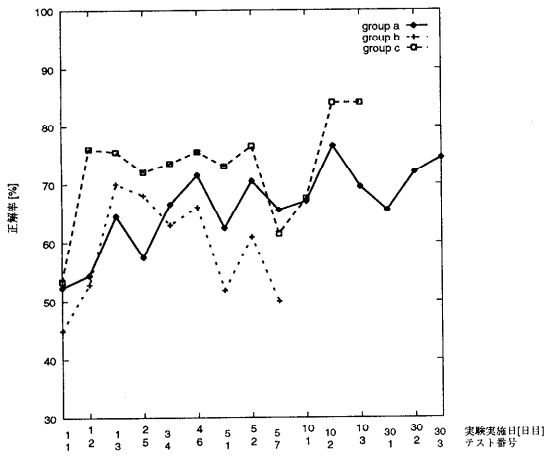
男声話者による，6単語前後からなるシナリオに含まれない文章10文を2回ずつスピーカを通して提示し，それを聞こえたとおりに書きとらせる（スペリングミスによる間違いは無視した）．書きとりを行っている間は音声提示は行われない．それを初日のCAI開始前と毎日のCAI終了後，そして，5日目（最終日）より10日後と30日後に行った．文の記憶による学習効果を避けるために，聞きとり用の文集合（テスト番号）は毎回変更した．そのため，聞きとり文集合の難易度は若干異なる．

結果を**図5**に示す．この実験では被験者の各グループをテスト文集合が同じであった話者同士で比較するために，2つに分けている（すなわち，図5(a)と(b)の話者は別々である）．グラフの横軸にテストの実施日とテスト番号を記述した．この図より，CAI使用前と最終日を比べると，ほとんどのグループが最終日の方が良いという結果になったが，向上率は音声入力，音声出力のCAIの被験者が一番高い．これは，ヒアリングテストを重ねることによる慣れによるスコアの向上によるとも考えられるが，これはグループcが少ししか向上しなかったので本システムの有効性がいえる．

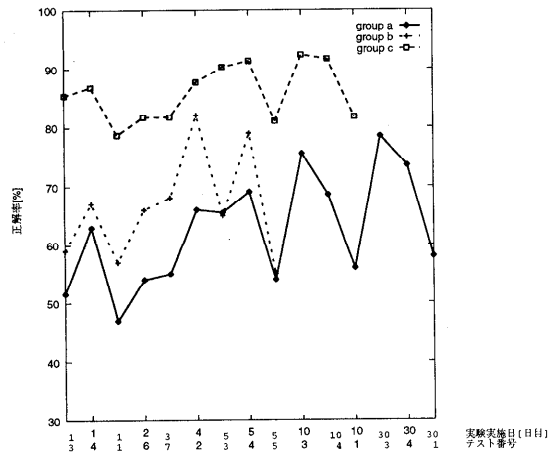
5.3 スピーキングテスト

学習者に6単語前後からなるシナリオに含まれない文章10文（1セット固定）を示し，初日のCAI開始前と毎日のCAI終了後，そして，5日目（最終日）より10日後と30日後に1回ずつそれらの文章を発話してもらい，それを評価用システムで評価した．本評価用システムでは，ネイティブスピーカによる音響モデルと学習者の発声を比較し，その尤度（入力音声と最も良く照合がとれる音韻系列との尤度から発声内容に応じた音響モデルの連結と入力音声との照合による尤度を差し引いた値：事後確率の対数値に相当）[16]をもとに10点満点でスコアをつけるものである．しかし，ここで使用している音響モデルは，すべての音響的情報は持っていないので（たとえば，アクセントなどの韻律情報は持っていない），このテストで高得点であったからといって必ずしも良い発音であるとは限らないが，各々の音韻の発音（スペクトル情報）の良否は判定できる．

この結果を**図6**に示す．この表からは，英会話CAIを使うとある程度まで発声が良くなり，特に1，2日後に急に良くなりその後飽和することが分かる．これは，朗読を重ねることによる発声の慣れによるスコアの向上によるとも考えられるが，これはグループcが少ししか向上しなかったので，本システムの有効性が



(a)



(b)

図5 ヒアリングテストの結果（単語の正解率%）
Fig. 5 Results of hearing test (correct percentage).

いえる．さらにグループbがグループa以上に向上していないことより，英会話の習得における発声の重要性が分かる．

5.4 アンケート結果

5日目（最終日）CAI終了後のアンケートを行った．それによる評価を以下に示す．

まず，システム全体について，「システム全体は良かったか」という問に対して，「良かった」と答えた人は10人中7人で，「普通」と答えた人は3人である．そして，「本英会話システムは英会話を勉強するのに役に立つか」という問に対して，「非常に役に立つ」と答えた人は10人中1人で「役に立つ」と答えた人は9人である．

このシステムの良い点を以下に示す．

- システムの応答時間はちょうどよい．シナリオの流れがスムーズである（10人中8人）．
  ⇒ 学習者に飽きさせない．
- 面白かった．引き続き使ってみたい（10人中9人）．
  ⇒ 学習の動機づけになる．
- 英会話を勉強するのに役立つ（10人中10人）．
  ⇒ 設定が，海外旅行に行ったら実際にあると思われる．実際の生活に役立つ語学学習ができる．

次に改良すべき点を以下に示す．

- 発声したとおりになかなか認識されないところがある．
- たくさん並んでいると探すのが大変．ぱっと見てすぐに分かるようなものがあればよい．

現在のシステムは「入国審査・税関」，「ホテルのチェックイン」と「路上」の各々の場面に関する会話文を含んでいるが，今後，以下のような場面における会話が扱われることが望ましいと学習者らは述べた．

- 買物でものを値切るためのテクニックを使うもの
- 初対面の人との会話（自己紹介）
- 授業風景（たとえば2＋5は？）
- ファーストフードショップにて

さらに，学習者らは以下の機能があればよいと答えている．

- 意味が分からなくなったときの，意味を見せてくれる機能．
- 日本文を表示して，学習者はそれを英語で言うモード．
- 選択肢のない，まったくのFree Talkモード．
- ユーザが話した発話を，ヘッドホン等で聞ける機能．
- アドベンチャーゲームみたいなもの．
- 2人で会話して，その内容について4択（TOEIC

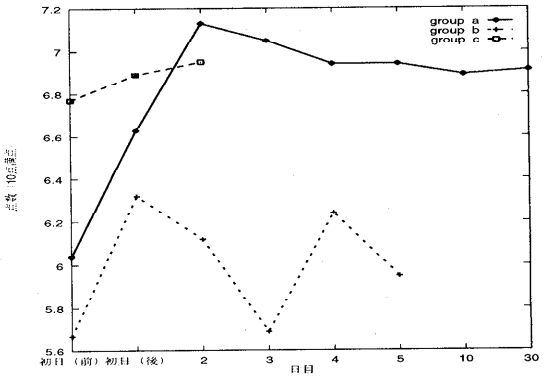


図6 スピーキングテストの結果（10点満点）
Fig. 6 Result of speaking test (maximum scale of 10 points).

やTOEFLみたいに).

- TOEICやTOIFLの問題を使って，試験対策.

6. むすび

本研究では，英会話CAIシステムを構築し，定量的な評価を行った．その結果，本システムは，英会話を勉強するうえで有効なものであることが分かった．

本研究のように音声認識技術を外国語CAIに導入することで，学習者の意欲を刺激し，より効率的に学習を進められることが期待できるが，以下のような問題点がある．

本システムのように，学習者が発声する文集合を明示的に絞ることは認識性能の向上につながる．明示的に絞ることをしない場合でも，学習者の学習言語に対する運用能力を考慮すれば，予想される文集合は一般の対話システムよりは小さくなる．しかし，この運用能力の低さが新たな問題点を生むことになる．すなわち，1) 誤った発音を「誤っている」と気付かずに（システムが指摘しないので）発声し続ける，2) 実際の会話では，“eh”，“oh” などの間投詞が入るが扱っていない，などである．これらは対象とする外国語の運用能力の低さに起因するものである．すなわち，従来の母国語を対象とした音声認識技術に対して，上記のような要素を十分に組み込むことも今後の課題である．

また，発音の良否の判定機能，文法の学習機能（平叙文を疑問文や否定文，感嘆文への変換など），シナリオの増加等のシステムの機能拡張などや，評価法の厳密化も今後の課題としてあげられる．

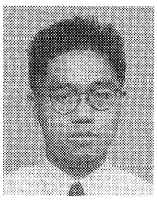
参考文献

1) Bernstein, J., Cohen, M., Murveit, H., Rtischev, D. and Weintraub, M.: Automatic Evaluation and Training in English Pronunciation, *Proc. ICSLP*, pp.1185–1188 (1990).
2) Hamada, H., Miki, S. and Nakatsu, R.: Automatic Evaluation of English Pronunciation Based on Speech Recognition Techniques, *IEICE Trans. INF. &SYST.*, Vol.E76-D, pp.352–359 (1993).
3) Hiller, S., Rooney, E., Laver, J. and Jack, M.: SPELL: An Automated System for Computer-aided Pronunciation Teaching, *Speech Communication*, Vol.13, No.3-4, pp.463–473 (1993).
4) Mostow, J., Roth, S.F., Hauptmann, A.G. and Kane, M.: A Prototype Reading Coach that Listens, *Proc. AAAI*, pp.785–792 (1994).
5) 山本秀樹，田川忠道，宮崎敏彦：音声対話を実現した英会話用知的CAIシステムの構成，情報処理学会論文誌，Vol.34, No.9, pp.1967–1981 (1993).
6) Aura-Lang：外国語のトレーニングシステム，仏Auralog社.
7) Reyes, A.A., 中川聖一：音声認識技術を利用した英会話CAIシステム，第50回情報処理学会全国大会論文集，No.7R-07 (1995.3).
8) 中川聖一，Reyes, A.A.，鈴木英之：日本人の発声した英語音声の認識と英会話のCAIシステム，人工知能学会全国大会論文集，No.22-13 (1995).
9) 小池 武，山本幹雄，中川聖一：メニューに基づく音声対話手法を用いた語学CAI，第50回情報処理学会全国大会論文集，No.2E-05 (1995.3).
10) 鈴木英之，中川聖一：第2言語学習者用に適応化された音素モデルによる第2言語発話者の音声認識，音響学会春季大会論文集，No.2-5-9 (1995.3).
11) 谷口泰広，中川聖一：音声認識技術を利用した英会話CAIシステムの改善と評価，人工知能学会全国大会論文集，No.17-11 (1996).
12) Robinson, T. and Fallside, F.: A Recurrent Network Speech Recognition System, *Computer Speech and Language*, Vol.8, pp.259–274 (1991).
13) 中川聖一：確率モデルによる音声認識，電子情報通信学会 (1988).
14) 中川聖一，甲斐充彦：文脈自由文法制御によるOne Pass型連続音声認識法，電子情報通信学会論文誌，Vol.J76-D-II, No.7, pp.1337–1345 (1993).
15) 中川聖一，越川 忠：最大事後確率推定法を用いた連続出力分布型HMMの適応化，音響学会誌，Vol.49, No.10, pp.721–728 (1993).
16) Kai, A. and Nakagawa, S.: Relationship among Recognition Rate, Rejection Rate and False Alarm Rate in a Spoken Word Recognition System, *IEICE Trans. Information and System*, Vol.E78-D, No.6, pp.687–704 (1995).

(平成8年9月27日受付)
(平成9年6月3日採録)

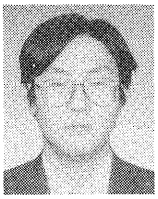
中川 聖一（正会員）

昭和23年生．昭和51年京都大学大学院博士課程修了．工学博士．同年同大学情報助手．昭和55年豊橋技術科学大学情報工学系講師．昭和58年助教授．平成2年教授．音声処理，自然言語処理，人工知能の研究に従事．昭和52年電子通信学会論文賞，1998年度IETE最優秀論文賞．著書：「音声・聴覚と神経回路網モデル」,「確率モデルによる音声認識」,「情報理論の基礎と応用」など．



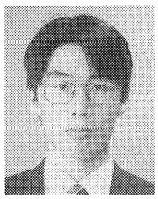
Allan A. Reyes

昭和45年生．平成5年豊橋技術科学大学情報工学課程卒業．平成7年同大学院修士課程修了．同年（株）アイピックス入社，外国特許部所属．外国特許出願業務に従事．



鈴木　英之

昭和44年生．平成5年豊橋技術科学大学情報工学課程卒業．平成7年同大学院修士課程修了．同年ソフトウエア開発（株）入社，営業部バイオ技術開発課所属．遺伝情報処理ソフトウエアの開発に従事．



谷口　泰広

昭和48年生．平成8年豊橋技術科学大学情報工学課程卒業．現在，同大学院修士課程在学中．音声認識の利用技術に関する研究に従事．