Vol.2015-NL-221 No.12
Vol.2015-SLP-106 No.12
2015/5/26

[招待講演] 日本人のための音声対話による英会話学習システム

伊藤 彰則[1,a)]

概要：筆者のグループがこれまで研究してきた，音声対話を利用した英会話のための CALL システムに関する技術について述べる．音声認識技術を利用した現状の CALL システムは，発音やイントネーションなど，1 つの発話に含まれる要素を採点するものが多い．それも重要ではあるが，英会話学習には「実際に使われる表現を何度も繰り返して練習する」ということも必要である．この考えに基づき，筆者のグループではこれまで「対話に基づく CALL システム」について研究してきた．本稿では，対話音声からの韻律評価，文法誤り検出および応答タイミング制御練習のためのシステムについて述べる．

1. はじめに

英語を使う能力は現在の日本においては多くの人にとって不可欠な能力である．特に近年は国の教育政策において読み書きだけでなくコミュニケーションスキル教育に重点が置かれており [1]，従来の「読む」「書く」能力に加えて「聞く」「話す」能力がより重要視されている．このような能力を開発するために，コンピュータを利用した言語学習 (Computer-Assisted Language Learning, CALL) システムが利用されている．

前述の「読む」「書く」「聞く」「話す」の 4 技能について，さまざまな CALL システムが開発されている [2]．最初期の CALL システムでは，コンピュータの制約によって音声や画像情報などを扱うことが難しかったため，扱う項目は語彙・スペル・文法などに限られていた [3], [4]．その後，音声や画像などのマルチメディア情報を使ったリスニングなどの教材が一般的となり [5], [6]，さらに自然言語処理や音声認識などの機能を利用した知的 CALL システムへと発展している [7], [8]．

音声認識技術を使う CALL システムは 90 年代から研究されており，発音の良さの評価，および簡単な質問に音声で答えるなどの機能をすでに実現していた [9]．中でも発音評価は初期段階からよく利用されてきた技術であり [10]，その技術的な精度だけでなく，教育的な有効性についても研究が進んでいる [11]．

[1] 東北大学 大学院工学研究科
Grad. Sch. Eng., Tohoku University, Sendai 980–8579, Japan
[a)] aito@spcom.ecei.tohoku.ac.jp

一方，「学習者が発話してシステムが評価する」という枠組みではなく，学習者と人間の教師が会話しながら英会話の学習をするように，学習者と音声対話システムが対話をすることによって外国語学習を行う「音声対話型 CALL システム」が 1990 年代後半〜2000 年代になって検討されるようになってきた [12], [13]．

本稿では，日本人の英語学習のための CALL システム，特に音声対話型 CALL システムのための技術について，筆者のグループが行ってきた研究を紹介する．

2. 音声対話型 CALL システム

音声対話型 CALL システムでは，学習者とシステムが英語で対話を行うことで英会話を学習することが想定されている．英語で会話を行うこと自体が学習になるが，それ以外に会話音声の発音，イントネーションやリズム，文法などのチェックを自動で行うことが考えられる．一方，学習者は非母語話者なので，前記のチェック項目（発音，イントネーション，文法）が間違っている音声を入力することを想定しなければならず，それを音声認識することは容易ではない．知的な対話制御を行い，システムを教師役としてふさわしく見せることも重要であるが [14]，現状ではそこまで行うことは難しく，あらかじめ決められた受け答えをシステムと行うことが多い [15]．

音声対話型 CALL システムで，現状の技術で難しいと思われるのは以下のような点である．

- 「訛り」だけを測る
  学習者の音声は，標準的な英語母語話者の発声と比べて，「話者による変動」と「発音の不適切さによる変動」の両方を含んでいる．話者による変動は学習の対

象ではないので，話者性による変動を除去したうえで，発音の不適切さだけを測りたい．

- 発音した通りに認識する
  学習者の発話する文は，学習者のレベルに合わせて多様な誤り（発音誤り，韻律誤り，語彙誤り，文法誤り）を含んでいる．CALL システムはそれを適切に識別して学習者にフィードバックすることが望まれるが，そのためには学習者の発話をいったん「発話した通り」に認識する必要がある．これには次のような難しさがある．
  - 学習者の発音は非母語話者発音であるから，発声される音韻は日本語と英語どちらでもない音になる可能性がある．そのため，認識のための音響モデルをどう用意するかが問題となる．
  - 韻律を評価する場合，何を「正解」とするかが問題となる．学習者が発話するすべての文章に対して，英語母語話者による「お手本」があれば，それを使って韻律の良さを測ることができるが [16]，実際にはそれが準備できないことも多い．
  - 学習者の音声には語彙的・文法的誤りが含まれると期待されるが，通常の英語文章から言語モデルを学習した場合，このような誤りは学習用の文章に含まれない．そのため，そこから学習した言語モデルで学習者の音声を認識すると，本来指摘すべき誤りが「訂正」されて認識される可能性がある．
- 適切なインタラクション
  実際にシステムと対話を行う場合には，対話相手として CG キャラクタ [14], [17] やロボット [18] を用意することが多い．しかし，これらの「対話相手」は，例えば話者交替のキューなどを人間と同じように表出することができるわけではなく，また学習者がうまく発話できないことに対して適切な反応を返すことも難しいため，システムとの間で人間同士のようなタイミングで対話を行うことが難しい．

これらの問題に関して，筆者らのグループがこれまで行ってきた研究を中心に紹介する．

3. CALL システムのための音響モデル

3.1 2 つの要求事項

CALL システムのための音響モデルには，「厳密に評価したい」「ルーズに評価したい」という互いに矛盾する要求がある．

学習者の発音に対して，母語話者との違いを認識して指摘する「発音評価」のためには，発音を厳密に評価することが望まれる．ここで，音響モデルは，入力音声が音響モデルの学習に用いたコーパス中の音声（一般には英語母語話者の音声）の確率分布の中心から離れるほど，入力音声に対して低い尤度を与える．この時の「離れ具合」には，入力音声の発音の訛りと，話者がどの程度学習コーパスの平均声から離れているかという声質の両方が影響する．しかし，ここで評価したいものは発音の訛りだけであるから，話者による違いは正規化し，発音の良さだけを評価するモデルを作りたい．話者による違いを正規化する代表的な方法は話者適応である．特に，MLLR[19] などの線形変換に基づく適応手法は広く用いられているが，通常の話者適応では「話者性」と「訛り」を区別しないため，単純に話者適応をすると「訛り」にも適応してしまい，適切な発音評価ができない可能性がある．

一方，発音の評価ではなく，対話することそのものを目的とする場合には，学習者の発音がかなり訛っていたとしても，それを学習者の意図通りに聞き取って応答することが必要である．これは，通常の音声認識アプリケーションにおける非母語話者音声の認識と同じ問題である．このためには，ERJ コーパス [20] のような非母語話者コーパスから音響モデルを学習することが有効であるが，学習者の発音は母語音声以上に多様であり，その多様性にどう対処するかが問題になる．

3.2 発音誤り検出

CALL システムでの発音評価には，大きく分けて「発音の良さの評価」[21] と「発音の誤りの検出」[22] の 2 つの方法がある．「発音の良さの評価」は，単語や文などの発声がどれだけ母語話者に近いかを数値的に評価するものである．「発音誤り検出」は，単語や文などの音声の中で，母語話者発音と比較して誤っている発音を指摘するものである．ここではこの 2 つのうち「発音誤り検出」に注目する．

発音誤り検出のアルゴリズムのうち，筆者らのグループがよく使っているのが「2 言語音響モデルによる誤り検出」[23] である．これを図 1 に示す．図は “mail her” という文の発音ネットワークであり，/uJ/などの J が付いた音素は日本語音響モデルの音素，それ以外は英語音響モデルの音素を示している．このネットワークを使って入力音声を評価し，ネットワーク上の最尤パスを求めたとき，それが日本語音素の上を通っていたら，その部分は発音誤り（すなわち，英語よりも日本語に近い発音）であると判定する．検出を高精度化する手法として，決定木による検出閾値の最適化なども行われる [24]．

3.3 CALL システムのための話者適応

前述の通り，入力音声を 2 言語の音響モデルで評価した時，各音響モデルが出力する尤度は「発音の各言語への近さ」と「入力話者と各音響モデルの学習に使った話者集合との近さ」の 2 つに影響される．ここで，後者はここで評価したいものではないので，後者を正規化して，前者のみを評価したい．しかし，学習話者の音声を用いて通常の話者適応を行うと，音響モデルは話者の話者性と発音の両方

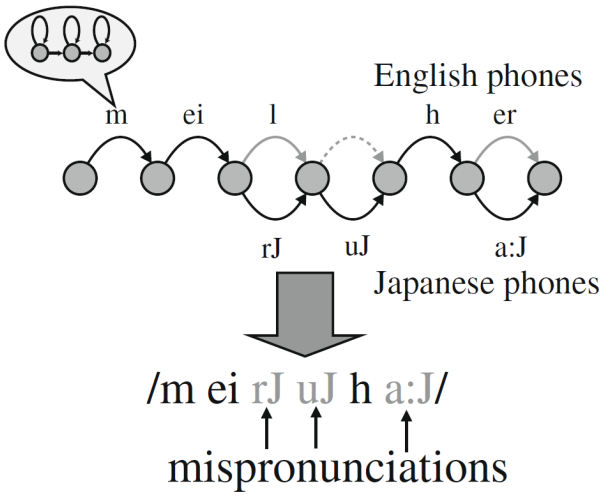

図 1 2 言語音響モデルによる発音誤り検出（図は [25] より）
Fig. 1 Detection of mispronunciation using bilingual acoustic models (figure referred from [25])

に対して適応してしまい，正しい発音評価が行えなくなる．そこで，話者性のみを正規化し，発音については正規化しない話者適応手法を開発した [25]．アイデアとしては，学習者の日本語発声のみを利用して，話者性だけを正規化する変換行列を求め，それをそのまま英語音響モデルにも適用することで話者性のみを正規化する．

日本語と英語の不特定話者音響モデルを同時に話者適応するために，日本語・英語バイリンガル話者の音声コーパスを利用する．この方法の概念図を図 2 に示す．この方法では，まず日本語不特定話者音響モデル M_N^n と英語不特定話者音響モデル M_T^t を用意する．これらのモデルは，学習話者が異なるため，仮に同じ「発音」があっても音響空間上で異なる場所に位置している．次に，複数の日本語・英語バイリンガル話者の日本語発音データ M_B^n と英語発音データ M_B^t を用意する．これらのデータは同一人物による発声であるため，話者性は同一であるが，言語による発音の違いが反映されている．これを使い，話者正規化学習 (SAT)[26] を行う．M_B^n を基準として M_N^n を学習し，M_B^t を基準として M_T^t を学習することによって，もともと異なっていた日本語と英語の不特定話者音声から学習した音響モデルの位置をおおむね同じ場所に合わせることができる．このようにして学習した日本語と英語のモデルをそれぞれ $M_{B'}^n$ と $M_{B'}^t$ とする．

学習者に対して話者適応を行う場合，まず学習者の母語発声（この場合は日本語発声）を適応データとして取得する．このデータを使って，日本語音響モデル $M_{B'}^n$ を MLLR 適応する．この時の変換行列を $W_{B\to L}^n$ とする．次に，変換行列 $W_{B\to L}^n$ をそのまま英語音響モデルに適応し，英語音響モデルを適応する．SAT によって $M_{B'}^n$ と $M_{B'}^t$ は音響空間上の同じ位置にあるはずなので，話者性を変換する行列は言語によらずそのまま適用できる．

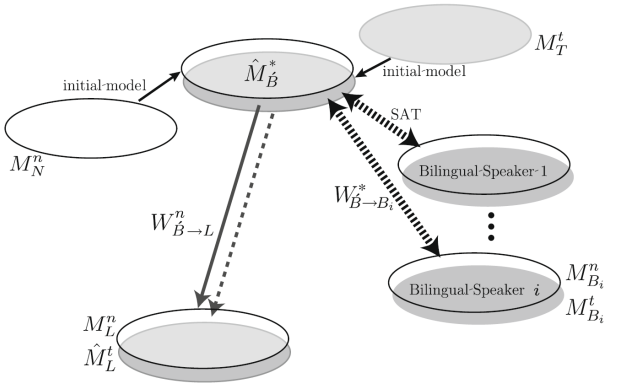

図 2 バイリンガル音声を用いた話者適応 [25]
Fig. 2 Speaker adaptation using bilingual speech [25]

3.4 発音習熟度を考慮した音響モデル

システムと学習者が英語で対話することで英語学習を行う CALL システムでは，たとえ学習者の発音が英語母語話者発音と大きく違っていたとしても，学習者の意図通りに単語を認識しなければ対話が維持できない．そのため，日本語母語話者が発声した英語文章から音響モデルを学習することが多い [27]．しかし，日本語母語話者の英語発音は多様であり，ほとんど日本語の音素からなる発音から，英語母語話者に近い発音まで様々である．それらをすべて 1 つの音響モデルで表現するのは困難である．そこで我々は，学習データを発話者の英語習熟度に応じて 3 つに分割し，習熟度別の音響モデルを作成した [28]．認識時には各習熟度の音響モデルを用いて入力音声を並列にデコーディングし，スコア最大の候補を選ぶ．

4. CALL システムのための言語モデル

音声対話による CALL システムにおいて，学習者の発話から文法的・語彙的誤りを自動的に検出する方法を開発した [27], [29], [30]．自然言語処理の文脈で「文法誤り検出」といえば，スペルチェッカから発展して冠詞の不適切な用法や主語・動詞の不一致，時制の不一致などを指摘する手法を指すが [31], [32], [33]，音声で入力する場合には「文法的・語彙的誤りを含んだ文音声を，誤りを含んだまま発話者の意図通り認識する」ということ自体が困難である．そのため，誤り検出自体はそれほど難しいことをせず（一般的には，すでにある正解と比較するだけ），誤りを含んだ文を高精度に認識することが課題となる．

誤りを含んだ文の認識は，前述の通り音響的にも難しいが，言語的にも困難な点がある．現在の多くの音声認識には統計的言語モデルが利用されているが，その学習には通常の文（そのほとんどは英語母語話者によって書かれた文）が利用される．その中には日本語母語話者が英語を話した時の誤りが含まれていないことから，「誤りを誤りとして認識」することが難しい．

そこで筆者らのグループは，想定される発話文に対して，

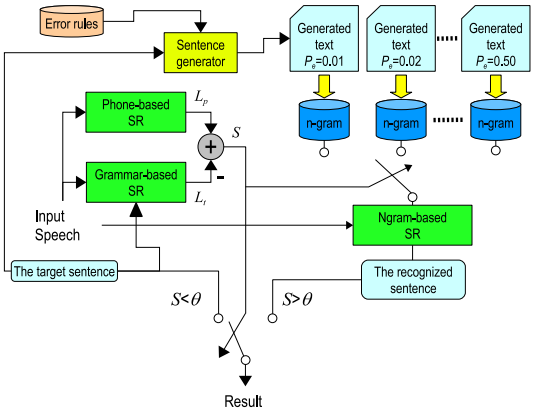

図 3 文法誤りに対応した言語モデルを利用する音声認識 [25]
Fig. 3 Speech recognition based on grammatical-error-injected language model [25]

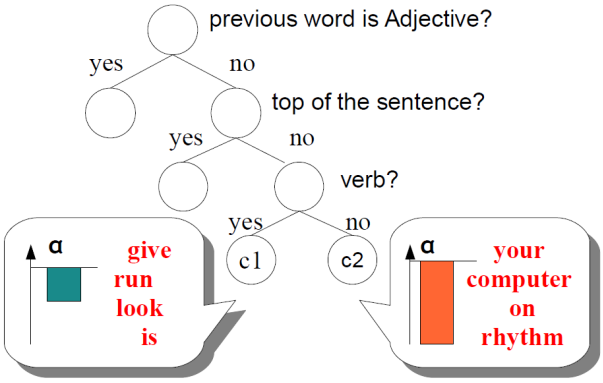

図 4 単語重要度を決定するための決定木 [34]
Fig. 4 An example of decision tree for determining the word importance factor [34]

日本人がよく犯す誤りを混入させて文を生成し，そこから統計的言語モデルを作成する方法を検討した．いくつかの方法を提案しているが，そのうちの一つ [30] を図 3 に示す．この方法では，認識すべき文を既知として，それに対して学習者がどう発話したのか（文法・語彙誤りが混入したのか）を高精度に認識する．最初に，認識すべき文に対して，日本人が犯しやすい誤りを一定の確率で適用して，誤りを含んだ複数の文を生成し，それらの文から n-gram モデルを学習する．これを異なる数種類の誤り確率について行い，複数の n-gram を生成しておく．実際に音声が入力されたときには，まず認識すべき文に対して計算した尤度と，連続音素認識の尤度との尤度差 S を計算する．ここで S が小さければ，認識すべき文と実際に発話された文は十分類似していると判断し，認識すべき文をそのまま結果として出力する．一方，S が大きい場合には何らかの誤りが含まれていると判断し，S の大きさから最適な誤り確率を推定して，言語モデルを 1 つ選択する．選択された言語モデルを使って入力音声を認識し，その結果を出力する．

5. イントネーションの評価

イントネーションやリズムは英語の重要な要素であるとともに，日本人学習者にとって習得が難しい要素の一つである．イントネーションの評価は，基本的には英語母語話者による「お手本」と学習者の発声の F0 軌跡を比較することによって行う [16]．この方法では，学習者の入力音声の F0 と，お手本音声の F0（いずれも正規化 F0 および Δ F0）の間の DP 距離を計算し，そこから線形回帰によって評価値を計算する．この方法の問題点は 2 つある．1 つは，F0 のずれがイントネーションに大きく影響する単語と，それほどでもない単語があることである．従来はそれらの重要性を一様に見ていたため，計算機による評価値と人間による主観評価値が合わない文が見られた．もう一つは，評価に「お手本」が必要な点である．

そこで，我々はイントネーション評価に 2 つの改良を導入した．1 つは，単語に「重要度」を導入することである [34]. この手法では，決定木を使って単語をクラスタリングし，クラスタごとに「単語重要度」を決める．この時の決定木の例を図 4 に示す．「単語重要度」は，その単語の特徴量に対する重みであり，最終的に計算された評価値と，人間による主観評価値の相関が最大になるように単語クラスタと単語重要度を同時に決定する．

もう 1 つの改良として，合成音声を「お手本」の代わりに利用するイントネーション評価を検討した [35], [36]. この方法では，評価すべき文を数種類の合成器で合成し，それぞれについてイントネーション評価値を計算した後，各合成器についての結果を重み付きで加算するという方法である．合成音声のイントネーションは完全ではないが，数種類の合成器を併用することで，人間のお手本を使った場合に匹敵する性能を得ることができる．

6. キャラクタとの英会話におけるインタラクション

実際にシステムと音声で対話を行うためには，さまざまな技術的課題がある．学習者の音声を確実に聞き取ることはその 1 つである．学習者が正しく発話しているのに，それを機械が聞き間違える（さらには，それを学習者の誤りとして指摘する）ことは教育システムとしてはあってはならない．認識を確実にするためには，学習者が発話する内容についての可能性を絞り込み，学習者が何を発話するのか事前に予測できる状況で利用する必要がある [15].

一方，対話による練習には，その他の練習にはない「インタラクション」という側面がある．人間同士の対話では，例えば自分の発話ターンにおいて，相手を見たまま黙り込んだりしないという暗黙のルールがある．そのため，直ちに発話することが困難な場合には，フィラーなどを発話することで発話権を確保しつつ時間を稼ぐ [37]．しかし，システムを相手にする場合には，仮想的な発話相手として CG

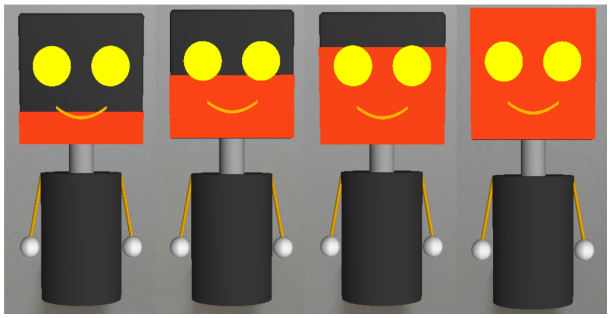

図 5 CG キャラクタとタイムプレッシャー表現 [42]
Fig. 5 A CG character and the time-pressure expression [42]

キャラクタ [17] やロボット [18] などが使われるため，人間同士の会話で話者交替のキューとなる表情や視線，動作などの表出が難しい．そのため，学習者は適切なタイミングでの発話ができず，システム発話の終わりから学習者発話の最初までの間隔（交替潜時）が長くなってしまう．交替潜時は対話において重要なパラメータであり [38], [39]，交替潜時が適切でない場合には，発話のニュアンスが変わってしまうことがある [40]．

そこで我々は，英会話において適切なタイミングで発話する練習ができるシステムを開発した [41], [42]．基本的な考え方は，対話相手のキャラクタが人工的な形状であることを生かし，話者交替のためのキューを人工的な表現で与えることである．これを我々は「タイムプレッシャー表現」と呼んでいる．システムで使用した CG キャラクタとタイムプレッシャー表現を図 5 に示す．システムの発話が終わった後，キャラクタが下から徐々に赤くなっていくことで，いま学習者のターンであることを明示する．

このシステムでは，画面上の CG キャラクタと学習者が英語で会話を行うが，この際の会話内容はあらかじめ学習者が事前に暗記しておいたものである．このような事前の学習に基づく反復練習 [15] は，言語表現が自動的に口から出てくる (automatization) ようになるために必要とされる [43]．

このシステムを用いて実験を行ったところ，いくつかのことが明らかになった．まず，CG キャラクターを導入することで，単に声で対話するよりも「話しやすさ」や「練習をしている感覚」が向上するが，応答タイミングには差が出なかった．次に，CG キャラクタにタイムプレッシャー表現を導入することで，タイムプレッシャーなしの場合よりも交替潜時の値が小さくなり，人間同士での英会話と近づくことが明らかになった．また，この学習効果は，タイムプレッシャー表現の速さの条件によっては，2 週間後まで持続することがわかった．

7. むすび

音声対話型 CALL システムの実現に向けて，筆者のグループがこれまで行ってきた研究を紹介した．冒頭に述べたように，英語学習の必要性は常に叫ばれているが，大学の英語クラスなどでは学生の動機付けに苦心しており [44]，また英会話学校に通う人々も，直接仕事に必要と言うよりも，より漠然とした期待感や，英会話学校での人間関係などから学校に通っている人が多いという [45]．単に必要だからということではなく，楽しみながら英会話ができる CALL システムが作れないものかといつも考えている．ゲームを使って英語学習の動機付けを行う試みなどもあるので [46]，音声対話システムのゲーム性 [47] を生かしたシステムなども検討していきたい．

謝辞

ここで紹介した研究内容は，多くの方々との共同研究の成果である．ここに謝意を示したい．

参考文献

[1] Y. G. Butler and M. Iino. Current Japanese reforms in English language education: The 2003 “action plan”. *Language Policy*, 4(1):25–45, 2005.
[2] S. Lambacher. A CALL tool for improving second language acquisition of English consonants by Japanese learners. *Computer Assisted Language Learning*, 12(2):137–156, 1999.
[3] K. L. E. Ng and W. P. Olivier. Computer assisted language learning: An investigation on some design and implementation issues. *System*, 15(1):1–17, 1987.
[4] P. Swann. Computer assisted language learning for English as a foreign language. *Computers & Education*, 19(3):251–266, 1992.
[5] C. Meskill. ESL and multimedia: A study of the dynamics of paired student discourse. *System*, 21(3):323–341, 1993.
[6] M. Warschauer and D. Healey. Computers and language learning: an overview. *Language Teaching*, 31(2):57–71, 1998.
[7] 河原達也, 峯松信明, 音声情報処理技術を用いた外国語学習支援, 電子情報通信学会論文誌 (D), J96-D(7):1549–1565, 2013.
[8] J. Gamper and J. Knapp. A review of intelligent CALL systems. *Computer Assisted Language Learning*, 15(4):329–342, 2002.
[9] F. Ehsani and E. Knodt. Speech technology in computer-aided language learning: Strengths and limitations of a new CALL paradigm. *Language Learning & Technology*, 2(1):45–60, 1998.
[10] A. Neri, C. Cucchiarini, H. Strik, , and L. Boves. The pedagogy-technology interface in computer assisted pronunciation training. *Computer Assisted Language Learning*, 15(5):441–467, 2002.
[11] A. Neri, O. Mich, M. Gerosa, and D. Giuliani. The effectiveness of computer assisted pronunciation training for foreign language learning by children. *Computer Assisted Language Learning*, 21(5):393–408, 2008.
[12] K. A. Wachowicz and B. Scott. Software that listens: It’ s not a question of whether, it’ s a question of how. *CALICO Journal*, 16(3):253–276, 1999.
[13] A. Raux and M. Eskenazi. Using task-oriented spoken dialogue systems for language learning: Potential, practical applications and challenges. In *Proc. InSTIL/ICALL*

2004 Symposium on Computer Assisted Learning, 2004.

[14] H. Prendinger and M. Ishizuka. Let's talk! socially intelligent agents for language conversation training. *IEEE Trans. Systems, Man and Cybernetics, Part A: Systems and Humans*, 31(5):465–471, 2001.

[15] O.-P. Kweon, A. Ito, M. Suzuki, and S. Makino. A grammatical error detection method for dialogue-based CALL system. *Journal of Natural Language Processing*, 12:137–156, 2005.

[16] A. Ito, T. Nagasawa, H. Ogasawara, M. Suzuki, and S. Makino. Automatic detection of english mispronunciation using speaker adaptation and automatic assessment of english intonation and rhythm. *Educational Technology Research*, 29:13–23, 2006.

[17] P. Wik and A. Hjalmarsson. Embodied conversational agents in computer assisted language learning. *Speech Communication*, 51(10):1024–1037, 2009.

[18] S. Lee, H. No, J. Lee, K. Lee, G. G. Lee, S Sagong, and M Kim. On the effectiveness of robot-assisted language learning. *ReCALL*, 23(1):25–28, 2011.

[19] C. J. Leggetter and P. C. Woodland. Maximum likelihood linear regression for speaker adaptation of continuous density hidden Markov models. *Computer Speech & Language*, 9(2):171–185, 1995.

[20] N. Minematsu, Y. Tomiyama, K. Yoshimoto, and K. Shimizu. English speech database read by Japanese learners for CALL system development. In *Proc. LREC*, 2002.

[21] S. M. Witt and S. Young. Phone-level pronunciation scoring and assessment for interactive language learning. *Speech Communication*, 30:95–108, 2000.

[22] O. Ronen, L. Neumeyer, and H. Franco. Automatic detection of mispronunciation for language instruction. In *Proc. Eurospeech*, 1997.

[23] G. Kawai and K. Hirose. Teaching the pronunciation of japanese double-mora phonemes using speech recognition technology. *Speech Communucation*, 30:131–143, 2000.

[24] A. Ito, Y.-L. Lim, M. Suzuki, and S. Makino. Pronunciation error detection for computer-assisted language learning system based on error rule clustering using a decision tree. *Acoustical Science and Technology*, 28:131–133, 2007.

[25] Y. Ohkawa, M. Suzuki, H. Ogasawara, A. Ito, and S. Makino. A speaker adaptation method for non-native speech using learners' native utterances for computer-assisted language learning systems. *Speech Communication*, 51:875–882, 2009.

[26] T. Anastasakos, J. McDonough and J. Makhoul. Speaker adaptive training: a maximum likelihood approach to speaker normalization. In *Proc. ICASSP*, 2:1043–1046, 1997.

[27] A. Ito, T. Tsutsui, S. Makino and M. Suzuki. Recognition of English utterances with grammatical and lexical mistakes for dialogue-based CALL system. Proc. Interspeech, 2819–2822, 2008.

[28] 安斎拓也，咸聖俊，伊藤彰則．“日本人英語発話からの文法誤り検出”，情報処理学会研究報告. SLP, 音声言語情報処理 2011-SLP-85(15), 1-6, 2011.

[29] T. Anzai, S. Hahm, A. Ito, M. Ito, and S. Makino. Grammatical error detection from english utterances spoken by japanese. In *Proc. APSIPA ASC*, pages 482–485, 2010.

[30] T. Anzai and A. Ito. Recognition of utterances with grammatical mistakes based on optimization of language model towards interactive CALL systems. In *Proc. APSIPA ASC*, 2012.

[31] N.-R. Han, M. Chodorow and C. Leacock. Detecting errors in English article usage by non-native speakers. *Natural Language Engineering*, 12(2):115–129, 2006.

[32] M. Gamon, C. Leacock, C. Brockett, W. B. Dolan, J. Gao, D. Belenko and A. Klementiev. Using Statistical Techniques and Web Search to Correct ESL Errors. *CALICO Journal*, 26(3):491–511, 2009.

[33] R. De Felice and S. Pulman. Automatic Detection of Preposition Errors in Learner Writing. *CALICO Journal*, 26(3):512–528, 2009.

[34] M. Suzuki, T. Konno, A. Ito, and S. Makino. Automatic evaluation system of english prosody based on word importance factor. *Journal of Systemics, Cybernetics and Informatics*, 6:83–90, 2008.

[35] A. Ito, T. Konno, M. Ito, and S. Makino. Evaluation of english intonation based on combination of multiple evaluation scores. In *Proc. Interspeech*, pages 596–599, 2009.

[36] A. Ito, T. Konno, M. Ito, S. Makino, and M. Suzuki. Intonation evaluation of English utterances using synthesized speech for computer-assisted language learning. *International Journal of Innovative Computing, Information and Control*, 6:1501–1514, 2010.

[37] C. M. Laserna, Y.-T. Seih and J. W. Pennebaker. Um . . . Who Like Says You Know Filler Word Use as a Function of Age, Gender, and Personality. *Journal of Language and Social Psychology*, 2014.

[38] C. Trimboli and M. B. Walker. Switching pauses in cooperative and competitive conversations. *J. Experimental Social Psychology*, 20(4):297–311, 1984.

[39] M. Heldner and J. Edlund. Pauses, gaps and overlaps in conversations. *Journal of Phonetics*, 38(4):555–568, 2010.

[40] M. G. Boltz. Temporal dimensions of conversational interaction: The role of response latencies and pauses in social impression formation. *J. Language and Social Psychology,* 24(2):103–138, 2005.

[41] N. Suzuki, T. Nose, Y. Hiroi, and A. Ito. Controlling switching pause using an AR agent for interactive CALL system. In *HCI International 2014 - Posters' Extended Abstracts*, volume 435 of *Communications in Computer and Information Science*, pages 588–593. Springer, 2014.

[42] 鈴木直人，廣井富，藤原祐磨，千葉祐弥，能勢隆，伊藤彰則．英会話学習システムにおける応答タイミング練習方法の有効性の検証．情報処理学会研究報告. SLP, 音声言語情報処理 2015-SLP-105(13), 1-6, 2015.

[43] E. Gatbonton and N. Sagalowitz, Creative Automatization: Principles for Promoting Fluency Within a Communicative Framework, *TESOL Quarterly*, 22(3):473–492, 1988.

[44] R. Bahous, N. N. Bacha and N. Nabhani, Motivating Students in the EFL Classroom: A Case Study of Perspectives. *English Language Teaching*, 4(3):33–43, 2011.

[45] R. Kubota, Questioning linguistic instrumentalism: English, neoliberalism, and language tests in Japan. *Linguistics and Education*, 22(3):248–260, 2011.

[46] T.-Y. Liu and Y.-L. Chu. Using ubiquitous games in an English listening and speaking course: Impact on learning outcomes and motivation. *Computers & Education*, 55(2):630–643, 2010.

[47] J. Edlund, J. Gustafson, M. Heldner and A. Hjalmarsson. Towards human-like spoken dialogue systems. *Speech Communication*, 50(8–9):630–645, 2008.