

# ARキャラクターとの英会話練習時における 交替潜時のタイムプレッシャーによる制御

鈴木 直人<sup>1,a)</sup> 廣井 富<sup>2</sup> 藤原 祐磨<sup>2</sup> 黒田 尚孝<sup>2</sup> 戸塚 典子<sup>1</sup> 千葉 祐弥<sup>1</sup> 伊藤 彰則<sup>1</sup>

**概要:** 英会話練習をする際は対話相手が必要であり、相手との会話がテンポ良く行えるようになる練習が求められる。CALL(Computer-Assisted Language Learning) システムにおいて、学習者の応答のタイミングを向上させるような枠組みは無いのが現状である。英会話練習の際には発話内容を想起し、それを英語で表現する2重の認知的負荷がかかるため、交代潜時が長くなりがちであるが、対話の最初から意識的に交代潜時を短くしていくためには学習者に対して明示的な方法を用いるべきである。そこで本研究では対話相手としてAR(Augmented Reality) キャラクタを設定し、タイムプレッシャー表現をかけたときに応答タイミングの練習として有効であるかどうかを実験により検証することを試みた。実験参加者にはタイムプレッシャーの有無で2通りの対話を行い、最後に主観評価のアンケートを行った。本稿では以上の結果と主観評価を踏まえた考察を報告する。

## 1. はじめに

### 1.1 研究背景

国際化に伴い、国内での英会話学習者は増加している。安価かつ手軽に英会話を練習する手段の1つに、コンピュータを利用した言語学習(Computer-Assisted Language Learning, CALL) システムがある[1]。リスニングや文法学習に活用されているシステム[2]もあれば、音声対話を利用し、発音や文法の誤りをチェック可能なシステムもある[3][4]。しかし、機械を相手に疑似的な会話練習をする場合、実際の人間を相手に会話をする場合と比較して不自然なテンポで会話が行われやすい。交代潜時は会話の円滑さ[6]や会話の種類[5]が関連しており、円滑なコミュニケーションを取るためには相手の発話に対して適切なタイミングで返事を返すことが重要である。しかし音声対話型のCALLシステムの場合、相手が機械であることが、質問に対して答えを長時間考えてしまうといったことを許す原因となり、人間同士のコミュニケーションではありえない発話タイミングが起きてしまいがちである。こうした発話タイミングを適切なものへとするための、CALLシステムの枠組みはないのが現状である。

### 1.2 関連研究

ユーザ発話の時間的な制御の方法としては、会話における引き込み現象[7]を用いる方法が提案されている。鈴木ら[8]は、目玉の形のキャラクターを使って、システムの発話速度を変えることによってユーザの発話およびシステムへの印象変化がどのように変わるかを調べている。その結果、システムの発話が早くなるほどユーザの発話は遅くなると報告している。また、小松ら[9]も模擬的な対話システムを使って類似の実験を行っており、こちらはシステムの発話速度とユーザの発話速度に対して引き込みが起きて、システムの発話速度とユーザの発話速度が同調したと報告している。

これらの先行研究に対し、我々の研究では、全体の発話速度よりも交代潜時に着目している。外国語学習における対話では、発話の際に、「発話内容を想起する」とことと「発話内容を外国語で表現する」という二重の認知的負荷がかかっており[10]、これを縮めることが英会話練習に有用だと考えたためである。また、引き込みによる発話への影響は数ターン程度で徐々に表れるのに対し、学習者の最初の発話から交代潜時を制御するためには、より明示的な方法を使った方がよいと考える。

さらに、音声対話を利用し、より現実の状況に即した英会話練習を行うためには、対話相手が表出されている方が望ましい。三宅ら[11]は、AR(Augmented Reality)によって表示させたCGキャラクターを用いた対話システムの評価を行い、ARキャラクターが円滑な対話に貢献したことを報

<sup>1</sup> 東北大学  
Tohoku University  
aoba 6-6-5, Aramaki, Aoba-ku, Sendai, Miyagi  
<sup>2</sup> 大阪工業大学  
Osaka Institute of Technology  
omiya 5-16-1, Asahi-ku, Osaka, Osaka  
a) naoto\_s@spcom.ecei.tohoku.ac.jp

告している。Leeら[12]はCALLシステムとして、対話相手をロボットとし、表情などを表出したことによってより現実感のある対話につなげている。

そこで本研究では、AR キャラクタを対話相手として導入する。AR を用いることによって、キャラクタの表現を工夫することができ、また対話相手として表出したキャラクタがあたかも実世界に存在するかのように見せることが可能になり、実践的な対話感覚を与えることにつながる[11]。表現を工夫できることを利用し、AR キャラクタを形状などを変化させることによって学習者にタイムプレッシャーを感じさせ、これによって交代潜時を制御する方法を検討する。

### 1.3 研究目的

本研究では、適切なタイミングでの対話練習を可能とするCALLシステムの実現に向けた検討を行う。対話相手としてARによるキャラクタを設定し、キャラクタがタイムプレッシャーをかける演出を行うことにより、交代潜時を意図的に制御することを試みる。実験後のデータ分析により、本研究に望まれる対話相手の印象も検討する。

## 2. 実験システム

実験システムは音声対話システムとARキャラクタを描画するシステムを並列に動作させたものである。タイムプレッシャー表現がある場合には、学習者が発話をしたとき、対話状態が遷移する条件を満たした時に対話システムから通信を行い、ARキャラクタの表現に反映させている。

### 2.1 音声対話システム

対話システムは一問一答型[13]であり、対話に使う英文は事前に学習する[3]。音声認識にはJulius[14]を用い、音響モデルはERJコーパス[15]の中の日本人による英語発話音声から学習したものをを用いた。言語モデルは使用する英文全てを用いて学習した。音声合成にはFestival[16]を用いて女性音声を合成した。

対話制御としては、学習者が発話すべき文をシステムが認識した際に、システムがその答えを発話して次の学習者発話を待つという単純な制御を行っている。音声認識誤りや学習者の言い間違いなどに対処するため、認識された文と実験に使用する英文に現れる単語の重なりを調べ、発話文の長さに応じた単語数以上が重複するときはその文が発話されたと認識するようにした。これにより、学習者が発話すべき英文を突然忘れてしまった場合に、その場で英文を構成し発話したとしても、なんらかの認識結果が出ることで対話を継続することができる。学習者の発話が認識されてから、1秒後に応答するように設定した。これは予備実験により経験的に調整した値である。

### 2.2 ARキャラクタとタイムプレッシャー表現

ARの描画にはARToolkit[17]を用いた。使用したARキャラクタは図1に示したものである。このキャラクタのデザインを採用したのは、既にこのキャラクタと同じデザインのロボットがあり、それとの比較を今後検討するためである。タイムプレッシャー表現の有無に関わらず、キャラクタは常に頷く動作を行う。これは聞き手として自然な感覚を持たせるためである[18]。頷き動作は30[fps]で表現している。タイムプレッシャー表現がない場合は、キャラクタの表情は常に笑顔とした。

タイムプレッシャーは図2のようにARキャラクタが1秒ごとに赤い部分が増えることによって表現した。ARキャラクタの下部から赤い部分が増えていき、10秒でキャラクタが全て赤くなる。キャラクタの表情は図2のまま赤くなり、学習者の発話が開始されたところでキャラクタの表情が図1のように笑顔になり、タイムプレッシャー表現が停止する。

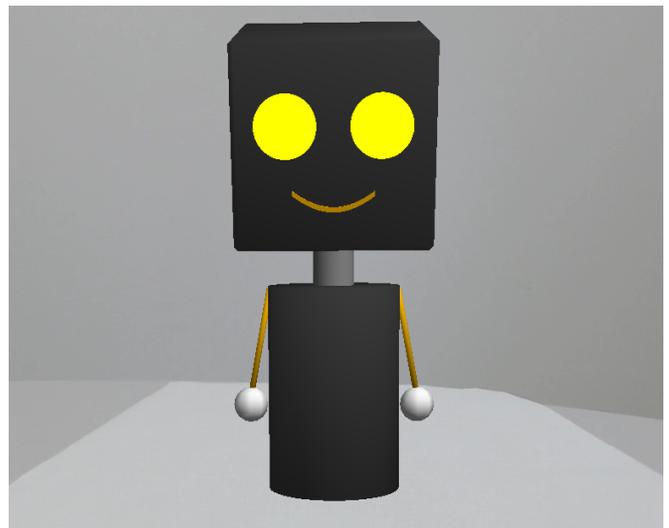


図1 ARキャラクタ

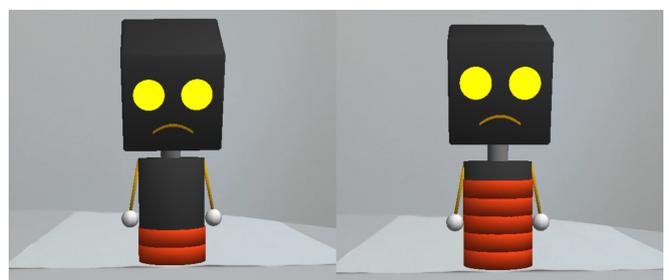


図2 タイムプレッシャー表現

### 3. 実験

AR キャラクタと対話を行う実験を行った。実験参加者は10人の大学生および大学院生で、いずれも英会話の学習歴はほとんどなく、英語学習歴は10年前後であった。実験手順は以下のとおりである。

- (1) 実験参加者に20分間の事前学習を行わせ、シナリオの文を暗記させる。
- (2) 実験参加者に筆記試験を行い、シナリオの文を暗記していることを確認する。
- (3) 実験参加者は、暗記したシナリオに従ってAR キャラクタと対話を行う。対話はタイムプレッシャー表現ありの場合となしの場合についてそれぞれ行う。このとき、2回の対話でそれぞれ異なるシナリオで対話を行う。
- (4) 実験に関するアンケートに記入してもらう。

実験に使用した英文のシナリオは図3と図4の2つである。英文シナリオ作成後、ネイティブのアメリカ人による校正を行ったものを使用した。1つのシナリオを覚えるのに10分あれば十分であることを予備実験を行い確認したため、事前学習時間を20分と設定した。事前学習の際は英文と対応している日本語訳も渡して行い、シナリオ中の英文の意味理解に差が生じないようにした。その後の筆記試験を行うのは実験参加者によるシナリオの記憶のあいまいさが交代潜時に影響することを避けるためである。制限時間を設けずに、図3と図4に示すシナリオのユーザー発話の部分を全て記入してもらった。

AR キャラクタとの対話はタイムプレッシャー表現があるものとないものの2通りであり、実験参加者の半数はタイムプレッシャーがないキャラクタとの対話を先に行い、次にタイムプレッシャー表現のあるキャラクタと対話を行った。もう半数はその逆の順序で会話をを行った。英文のシナリオについても同様に順序効果を考慮し、半数は帽子を買いに行くシナリオを最初に用い、半数はTシャツを買いに行くシナリオを最初に用いた。

実験参加者には以下の3つのディレクションを行った。

- AR キャラクタにタイムプレッシャー表現のあるときはキャラクタが全て赤くなる前に応答すること
- システムから応答が返ってこない場合は、キャラクタがうまく聞き取れていないので各自の判断で再度発話すること
- 英文を思い出せない場合は自分で英文を構成して発話すること

2つ目はタイムプレッシャーの条件によらず、音声が入り込まない場合や、誤認識が起きた場合の対応であり、人間同士でも起こりうる状況設定とした。

システム発話	ユーザー発話
Hello. May I help you?	Yes, I'm looking for a hat. Do you have one?
Yes, we do. What kind do you want?	A green one.
Like this?	Yes, like that one. Can I see it?
Yes. Here you are. Would you like to buy it?	I'm sorry. This isn't exactly what I wanted.
How about another product?	No, thank you.

図3 帽子を買いに行くシナリオ

システム発話	ユーザー発話
Hello. May I help you?	Yes. Do you have T-shirts?
Yes, we do. What kind do you want?	A blue one.
Like this?	Yes, like that one. Can I try this on?
Yes. Here you are. Would you like to buy it?	I'm sorry. This T-shirts doesn't look good on me.
How about another product?	No, thanks.

図4 Tシャツを買いに行くシナリオ

#### 3.1 実験環境

実験の様子を図5に示す。実験は防音室で行い、実験の様子は許可を取りビデオで撮影した。実験参加者は図5のように、Webカメラ (Logicool HD Pro Webcam C920) を装着したヘッドマウントディスプレイ (SONY HMZ-T2) に投影されたAR キャラクタと英会話を行った。合成音声はAR マーカーの横に設置したスピーカから再生される。音声入力にはマイクを用い、実験参加者ごとに位置を調整した。ヘッドマウントディスプレイを装着した状態で、見え方や音声入力、合成音声の音量などをテストしてから実験を行った。

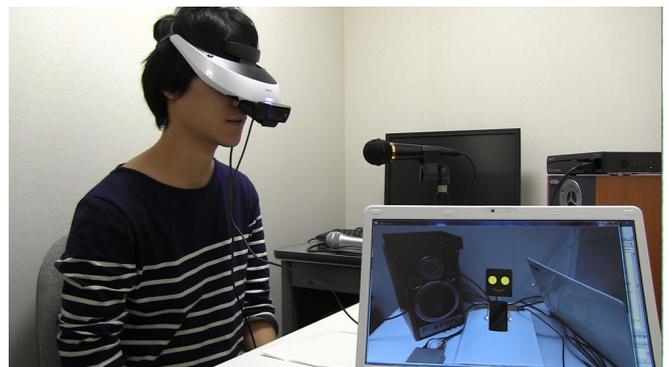


図5 実験の様子

### 4. 実験結果

事前学習後の筆記試験では、全ての実験参加者が単語の誤り等なく、約3分でシナリオの英文を記入することができたことから、全員英文を覚えていることを確認した。実験終了後、ビデオの映像と音声から、各セッションの交代潜時を計測した。計測は画像フレーム単位で行ったので、精度は1/30[s]である。

#### 4.1 交代潜時計測結果

システムの発話終了から実験参加者が応答するまでの交代潜時の平均と標準偏差を図 6 に示す。図中の誤差棒は標準偏差である。タイムプレッシャー表現がある場合とない場合について対応のある t 検定を行ったところ、タイムプレッシャー表現の有無による交代潜時の差は 1% 有意であった ( $p=0.0096$ )。タイムプレッシャー表現を導入することによって、交代潜時を約 0.5 秒短くできたことがわかる。この結果は AR キャラクタを用いたタイムプレッシャー表現によって学習者が発話するタイミングを制御できる可能性を示している。

図 7 は、交代潜時の平均と標準偏差を実験参加者ごとに見たものである。繰り返しのある 2 元配置分散分析の結果から、タイムプレッシャー表現の有無については 1% 有意 ( $p=0.00409$ )、実験参加者による違いは 5% 有意 ( $p=0.0128$ ) であった。この結果から、タイムプレッシャー表現がない場合の交代潜時は人により大きくばらついているのに対し、タイムプレッシャー表現がある場合の交代潜時は 1 秒前後であることがわかる。これより、タイムプレッシャー表現が学習者の交代潜時を短くするというよりも、その時間を一定にする効果を生んでいることがわかる。

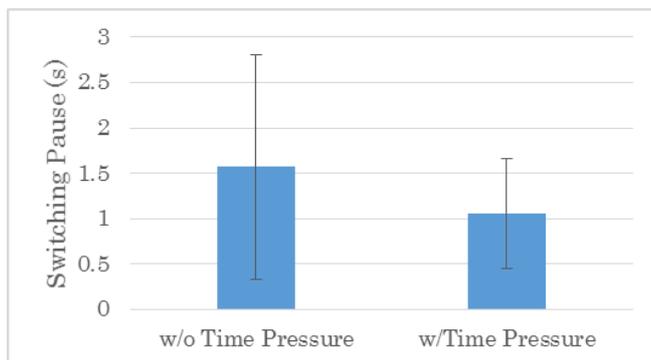


図 6 タイムプレッシャーの有無による交代潜時の違い

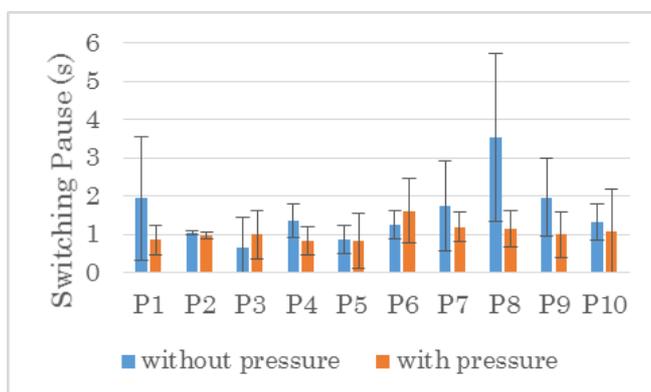


図 7 実験参加者ごとの交代潜時の違い

図 8 は、実験参加者の 1 回目～5 回目の発話ごとに交代潜時を集計した結果である。それぞれの値は 2 つのシナリオにおける発話を合わせて集計したものであるため、発話

の意味的な違いによる影響ではなく、ターンが進むごとに交代潜時が影響を受けているかどうかを見るものである。この結果からは、対話の進行とともに交代潜時が系統的に変化する様子は見られない。

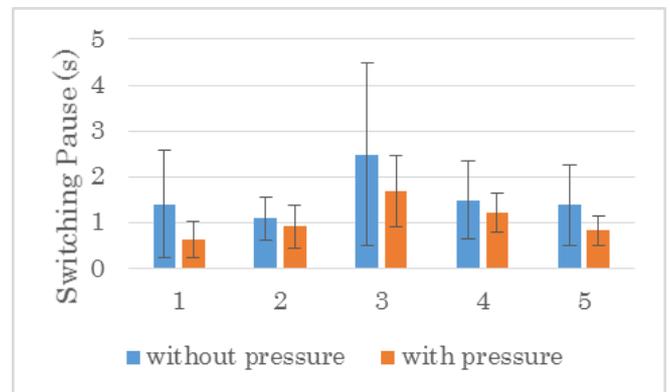


図 8 ユーザ発話ごとの交代潜時の違い

#### 4.2 主観評価アンケート結果

実験後の主観評価のアンケートは 5 段階の評価値を用いたものと、自由記述のアンケートで構成されている。

##### 4.2.1 5 段階評価アンケート

評価値を用いたアンケートの項目は以下の通りである。

- Q1: 実験での対話は人と対話している感覚に近かった
- Q2: システムの発話にタイミング良く応答できた
- Q3: 覚えた英文がうまく出てこないことがあった
- Q4: 英文の意味を考えながら話した
- Q5: 合成音声聞き取ることができた

これに対し評価値を、5:あてはまる、4:どちらかといえばあてはまる、3:どちらでもない、2:どちらかといえばあてはまらない、1:あてはまらない、と対応させ評価してもらった。図 9 にその結果を示す。グラフは評価値ごとの平均値を、誤差棒は標準偏差を表している。

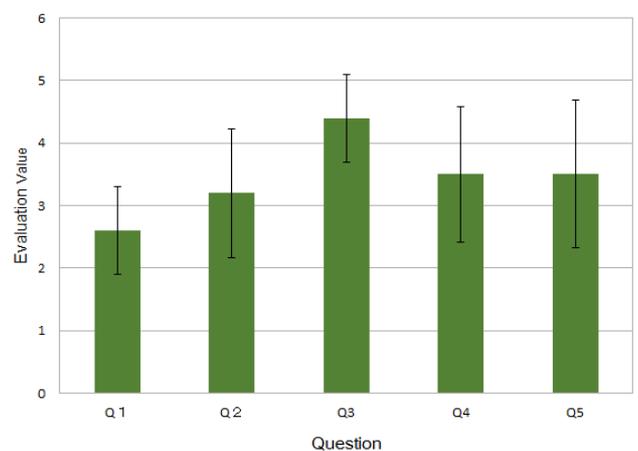


図 9 主観評価値の平均と標準偏差

Q1の結果から、実験参加者はARキャラクタとの対話を人との対話に近いとは評価していないことがわかる。これは、対話相手として話を聞いているという特徴が欠如していたことを示している。Q2の結果は人によるばらつきが多いことから、タイミングの良さを評価する基準があいまいであったことがわかる。Q3の結果より、英文を正しく覚えた状態であっても、実際に使うことを想定して話す練習を行うことの必要性が示されている。Q4は今回の実験パラダイムにおいて、英文の丸暗記をさせたことが認知的負荷を減少させた可能性があることを確かめたものである。英文を構成することによる負荷に関しては実際の場面に比べて低いと考えられるが、シナリオを思い出すという行為が新たに負荷となった可能性もある。今回の結果からは判断することはできなかった。Q5についてはばらつきが大きく、聞こえ方には個人差があったと思われる。

#### 4.2.2 自由記述アンケート

自由記述のアンケートでは主に以下の3点について記入してもらった。

- (1) システムを使って初めての感想意見
- (2) 対話のテンポについて普段どのように感じているか
- (3) ARキャラクタ表現の印象

システムを使ってみて、音声合成の声が聞き取りづかった、キャラクタの声らしくなかった所以对話している感覚にならなかった、例文通りに言おうとしたが普段使い慣れた表現で言ってしまったことや思い通りに対話を行うことができなかったという回答があった。

対話のテンポに関しては、対話相手の視線や表情から対話時のテンポを良くしようとする気持ちが生じると回答している実験参加者が多かった。会話をしている場の雰囲気によるという回答や、英文を考えている時に、相手から別のことを言われた時に早く応答しなくてはという気持ちにさせられるという回答も見られた。

ARキャラクタ表現の印象についてはARキャラクタの口の周りの動きがないという意見があった。タイムプレッシャー表現があると焦りが感じられたことや、それに対して相手が発話してからどのくらい時間が経過したかがわかりやすい表現だったという回答も得られた。話すことや思い出すことに集中し、タイムプレッシャー表現を意識しなかったと答える内容もあり、表現をうまく実験参加者に伝えることのできる工夫が必要であることがわかる。

## 5. 考察

今回の実験において、タイムプレッシャー表現のないときに比べ、交代潜時が1秒程度に揃う現象が観察された。これはアンケートにあるように合成音声の聞き取りにくさや、キャラクタの口周りの動きがないことなどの問題によ

り、タイムプレッシャー表現がない場合には実験参加者が音声入力をしてよいのか判断しづらくなっていたが、1秒ごとに赤くなる部分が増えていく表現がある場合、それがキャラクタの発話の終了の合図となり発話しやすくなっていったという可能性を指摘しているものである。交代潜時の値が制御できる可能性とその制御される値がある一定の値に抑制される点に関連していると思われるため、検証を行う必要がある。

また、対話が自然なテンポになるように練習する目的において、自然なテンポとは何かという点を無視できない。実験参加者の交代潜時において最も長いものは6秒を超えていた。こういった交代潜時の長いものに関しては、応答を急がせるためにタイムプレッシャーを用いることは有効であるが、必ずしも良いリズムに近づくとは限らない。焦りを感じた実験参加者がいたように、英文を思い出しづらくなってしまい、逆に遅くなってしまいう可能性もある。それは対話の種類や状況によって異なるものではあるが、今回の実験を踏まえ、ARキャラクタとの対話において、タイムプレッシャーの与え方としては、対話のテンポを良くしたいと思わせることのほうが重要であると考えられる。そのような感覚を与えるためにはまず、ARキャラクタの印象として自然なリズムを構築しやすい特徴を持たせるべきである。話を聞いている、話している（話し終わった）といった表現が与えられていると学習者はテンポを掴みやすくなる。これについてはアンケートの指摘を参考にし、口の動きをつけることによって改善が望める。そのうえで表情などによって学習者にテンポを良くしようと想起させる必要があると考えられる。

最後に、今回用いたキャラクタ自体についてである。音声キャラクタのイメージとは異なるものであったことや、視線などを感じにくいデザインになっているということがアンケートにより指摘されているため、全く別のキャラクタを用いた検討も行う必要がある。

## 6. まとめと今後の課題

本稿では、ARキャラクタと英会話の練習を行うシステムについて提案した。会話のタイミングをシステムが制御するために交代潜時に着目し、タイムプレッシャー表現を導入してその値を制御することを試みた。実験の結果から、タイムプレッシャー表現によって交代潜時に有意な差が得られることがわかった。また、その表現とARキャラクタの印象評価を行った。今回の実験により、交代潜時の制御を行う対話相手の印象についてより調査を行う必要性が示されたので今後はその検討を行う予定である。

謝辞 本研究は、JSPS 科研費 24652111 の助成を受けた。

参考文献

- [1] 中條, 西垣, 内堀, 山崎, “英語初級者向け CALL システムの開発とその効果”, 日本大学生産工学部研究報告 B, vol. 38, pp. 1-16, 2005.
- [2] M. Eskenazi, “An overview of spoken language technology for education”, Speech Communication, vol. 51, no. 10, pp. 832-844, 2009.
- [3] O.-P. Kweon, A. Ito, M. Suzuki and S. Makino, “A grammatical error detection method for dialog-based CALL system”, Journal of Natural Language Processing, vol. 12, no. 4, pp. 137-156, 2005.
- [4] T. Anzai and A. Ito, “Recognition of utterances with grammatical mistakes based on optimization of language model towards interactive CALL systems”, Proc. APSIPA ASC, 2012.
- [5] C. Trimboli and M. B. Walker, “Switching pauses in cooperative and competitive conversations,” Journal of Experimental Social Psychology, vol. 20, no. 4, pp. 297-311, 1984.
- [6] 長岡, 小森, 中村, “音声対話における 2 者間の相互影響—時間的側面からの検討—”, 信学技報 HCS2003-9, pp. 19-24, 2003.
- [7] 渡辺, 大久保, “コミュニケーションにおける引き込み現象の生理的側面からの分析評価”, 情報処理学会論文誌, vol. 39, no. 5, pp. 1225-1231, 1998.
- [8] 鈴木, 箕, 竹内, 岡田, “非分節音を用いた人間: コンピュータ間の相互作用における発話速度の変化とその効果”, ヒューマンインタフェース学会論文誌, vol. 5, no. 1, pp. 113-122, 2003.
- [9] 小松, 森川, “人間と人工物との対話コミュニケーションにおける発話速度の引き込み現象” 情報処理学会研究報告, 2004-ICS-137(10), 2004.
- [10] P. Nation, “The role of the first language in foreign language learning,” The Asian EFL Journal, vol. 5, no. 4, 2003.
- [11] S. Miyake and A. Ito, “A spoken dialogue system using virtual conversational agent with augmented reality”, Proc. APSIPA ASC, 2012.
- [12] S. Lee et al, “POSTECH Approaches for Dialog-based English Conversation Tutoring”, Proc. APSIPA ASC, 2010.
- [13] 三宅, 廣井, 伊藤, “10 日間で作るロボット音声対話システム”, ヒューマンインタフェースシンポジウム 2011 論文集, 2416S, pp. 579-582, 2011.
- [14] 河原, 李, “連続音声認識ソフトウェア Julius”, 人工知能学会誌 20(1), pp. 41-49, 2005-01-01.
- [15] 峯松他, “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”, 日本教育工学会論文誌, vol. 27, No. 3, pp. 259-272, 2004.
- [16] The Festival Speech Synthesis System, <http://www.cstr.ed.ac.uk/projects/festival/>(2013 年現在)
- [17] 加藤, “ARToolKit”, 映像メディア学会誌, Vol. 62(No. 1), pp48~51, 2008.
- [18] 廣井, 伊藤, “拡張現実感を用いたロボットデザインの評価”, 日本バーチャルリアリティ学会論文誌, vol. 18, no. 2, pp.161-170, 2013.