

応答タイミングを考慮した英会話練習のための 音声対話型英語学習システム

鈴木 直人^{1,a)} 廣井 富² 千葉 祐弥¹ 能勢 隆¹ 伊藤 彰則^{1,b)}

受付日 2015年2月4日, 採録日 2015年7月1日

概要: 本研究では, 音声を用いた英会話の学習が可能なコンピュータ利用言語学習 (Computer-Assisted Language Learning, CALL) システムを提案する. 特に, 英会話学習における学習者の応答タイミングに着目する. 一般的に学習段階において応答タイミングは適切なものに比べ遅くなりがちであるが, システムとの英会話では応答タイミングを意識しにくい. そこで対話相手として CG キャラクタを導入し, 応答を要求する表現であるタイムプレッシャー表現を付加する練習方法を提案する. CG キャラクタの有無, タイムプレッシャー表現の有無のほかに, 短期間での繰り返し練習, および期間をおいた練習を通じて, 提案手法の有効性について論じる.

キーワード: CALL システム, 英会話練習, 応答タイミング, 交替潜時, CG キャラクタ, タイムプレッシャー表現

A Computer-assisted English Conversation Training System for Response-timing-aware Oral Conversation Exercise

NAOTO SUZUKI^{1,a)} YUTAKA HIROI² YUYA CHIBA¹ TAKASHI NOSE¹ AKINORI ITO^{1,b)}

Received: February 4, 2015, Accepted: July 1, 2015

Abstract: We propose a CALL (Computer-Assisted Language Learning) system for training of English conversation using spoken dialogue. Especially, we focus on a spoken dialogue system for training the learner's timing of response. To suppress the long pause and practice an appropriate pause duration, we introduced a CG character as a counterpart of the dialogue installed the behavior of "time pressure" to prevent the learner taking long time to consider the utterance. To verify if the expression is effective, we designed three experiments. First, we conducted with or without CG character. Second experiment was conducted with or without the time-pressure expression. Finally, we conducted the learning of dialogue in short-term repetition as well as the same learning after two weeks interval. In conclusion, we discuss the effectiveness of the proposed method.

Keywords: computer-assisted language learning, English learning, timing of response, switching pause, CG character, time-pressure

1. はじめに

国際化にともない, 国内での英会話学習者は増加している. 英語学習の1つの目標は, 英語での円滑なコミュ

ニケーションである. これには人間との実際の会話練習が欠かせないため, 英会話教室などでの学習だけでなく, Skype などの Web を利用したサービスによる練習も増えている [1], [2]. 初期の学習段階においては, 学んだ内容を活かして手軽に対話を行える場があることが望ましい. なぜなら, 非母語での会話において, 文章の構築などを無意識に行えるようになるには繰り返し練習する必要があるからである [3]. 実際に教師を相手とした会話は理想的ではあるが, 時間的・経済的な制約を受けるため, こ

¹ 東北大学
Tohoku University, Sendai, Miyagi 980-8579, Japan

² 大阪工業大学
Osaka Institute of Technology, Osaka 535-8585, Japan

a) naoto_s@spcom.ecei.tohoku.ac.jp

b) aito@spcom.ecei.tohoku.ac.jp

れを補完し、安価かつ手軽に英語を学習する手段の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 [4], [5], [6]. 初期の CALL システムはリスニングやドリル形式の学習を主体としていたが [6], 2000 年代以降は音声認識などの技術を利用した発音や韻律の評価システムも開発されている [7]. さらに、発音だけでなく、学習者の発話音声から文法誤りを指摘するための技術も開発されている [8], [9]. 最終的には、コンピュータやロボットとの音声対話によって英語を学習するシステムが想定されるが [7], [10], [11], その実現には様々な問題がある. これには、たとえば非母語話者音声認識の精度の問題や対話制御の問題などがあるが、本研究では学習者とシステムとのインタラクションの問題に注目する.

学習段階における英会話においては、発話の際に「発話内容を想起する」と「発話内容を英語で表現する」という二重の認知的負荷がかかっていることが知られている [12]. この負荷により長くなると考えられる交替潜時 (Switching Pause) は、会話の円滑さ [13] や会話の種類 [14] に関連しているといわれる. 円滑なコミュニケーションをとるためには、発音、韻律など学んできた英語の要素を活かしながらも、相手の発話に対して適切なタイミングで応答することが重要である. 応答タイミングは相手にはっきりと伝わる特徴であるため、適切な応答タイミングでない場合に、発話した文章の相手に伝わるニュアンスが異なってしまうことがある [15].

類似した考えとして、話者移行適格場 (Transition-relevance Place; TRP) がある [16]. TRP は話者交替に適したタイミングであり、Sacks らの論文 [16] においては言語的な制約 (すなわち、そこで話者が交替しても不自然でない文や句の切れ目) と発話の時間的な適切さの両方を指していた. また、交替潜時について Sacks らは「発話交替にギャップもオーバーラップもない場合が最も多い」 (Transitions ... with no gap and no overlap are common) と述べている. ここでいう「ギャップ」が交替潜時に対応する. このような「ギャップもオーバーラップもない発話交替」 (no-gap-no-overlap transition) は会話の流ちょうさの1つの指標とされ、この観点から英語非母語話者の会話や [17], [18], [19] 人間と機械の対話 [20] を分析した研究もある. 教育学的な立場からも、これを英会話における話者交替の規範として考えることが一般的である [21]. 一方、Heldner らは、ネイティブ話者による英語の交替潜時について詳細な分析を行い、発話に重なりがない場合の交替潜時の平均を 424 ms, 中央値を 318 ms と報告している [22]. この論文の中では、他の研究者による交替潜時のメタアナリシスも行われているが、分布に大きな違いはない. また、交替潜時には文化的・言語的な違いがあり、Stivers らは言語の違いによる話者交替への影響を調査した [23]. その結

果、ほとんどの言語で交替潜時の分布の最頻値は 0 ms であり、英語の場合の交替潜時の平均は 236.07 ms だったと報告している. いずれにせよ、英語母語話者間の自然な話者交替における交替潜時は、感覚的には「間隔がない」と感じられる程度であり、実際に計測された時間間隔としては 200~500 ms であることが分かる.

ところで、人間同士で会話する場合には、TRP のような言語的・音声的キューだけでなく、視線やジェスチャーなどのマルチモーダルな情報も使って話者交替のタイミングを計っている [24], [25]. しかし CALL システムを相手に英会話練習をする場合は、相手が機械であるため、質問に対して学習者が長時間考えてしまうことを許す原因となってしまう. そうした状況において、学習者が効率的に円滑なコミュニケーションをとるための技能を身に付けることは困難である. そこで、自動音声対話による英会話練習において、学習者が適切な応答タイミングで発話を行うように誘導することは、会話学習の CALL システムにとって重要な機能だと我々は考える.

ユーザ発話の時間的な制御に関しては、会話における引き込み現象 [26] を用いる方法が提案されている [27], [28], [29]. しかし、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. そのため、通常の会話のように、発話の最初から応答タイミングを考慮する状況を作り出すには適さない.

そこで本研究では、学習者に応答タイミングを意識させる手法として、対話相手として利用した CG キャラクタによってタイムプレッシャー表現を呈示する方法を提案する. さらに、英会話練習における効果を吟味することでその有効性を示す.

本稿の構成は以下のとおりである. まず2章では、本研究の基本的なアイデアと、実験パラダイムおよび実験に用いたシステムの構成について述べる. 3章では、システムとの英会話をする際に、CG キャラクタを用いたことによって英会話の一度目の練習から学習者の応答タイミングに変化があるかどうかを調べるために、CG キャラクタがいる場合といない場合の学習者の交替潜時の違いを比較する. また、対話相手として CG キャラクタを描画することが英会話練習に役立つことを確認する. 4章では、学習者に時間的意識を持った状態で対話をしてもらうために、タイムプレッシャー表現を導入する. これは、CG キャラクタを利用して、学習者にシステムの発話の終端からの時間経過を明示することで、学習者の応答タイミングが冗長にならないように意識させる方法である. 5章では、提案システムによる英会話練習の効果について、短期間における繰返し練習の効果、および期間を置いた練習の効果に注目して調査する. 6章は結論である.

2. 提案する英会話練習システム

2.1 基本的なアイデア

我々は、学習者と音声で会話をしながら、自然な応答タイミングの習得を含めた英会話学習をするための CALL システムを提案する。

これを実現するため、本研究では下記の 2 つのアイデアに基づいてシステムを構築した。1 つ目は、システムが CG によるキャラクタを提示し、学習者にそのキャラクタを対話相手として想定して対話練習を行ってもらうことである。対話相手としてロボットを用意することや、マルチメディア技術を利用するといったことで学習者の意欲の向上につながるといった報告のほかに [30], [31], CG キャラクタ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る [32]。また、AR (Augmented Reality) 技術を用いて実環境上にキャラクタを重畳表示させることで、キャラクタがあたかも実世界に存在するのように見え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 [33], [34]。2 つ目は、CG キャラクタの表示を時間的に変化させることによって、学習者に応答タイミングの手がかりを与えることである。本研究ではこれを「タイムプレッシャー表現」と呼ぶ。対話における発話のタイミング、特に話者が交代するときの発話タイミングについては多くの研究があり、人間同士の対話においては発話の韻律や語彙などによって制御されていることが知られている [35]。これらの知見を CG キャラクタの振舞いに実装することもありうるが、人間と必ずしも似ていないキャラクタや合成音声によってどの程度人間と同じ応答タイミング制御が可能なのかは明らかではない。本研究では、CG キャラクタが人間よりも「マンガ的」であることを生かし、より人工的な表現によって、学習者の応答タイミングを制御することを目指す。

以上、CG および AR を用いることによる利点を活用し、キャラクタに付加する表現などを変化させることで学習者に応答タイミングを明示的に意識させながら、学習に望まれる意欲などの要素も得ることが可能であると考えられる。

2.2 シナリオに基づいた音声対話を行うシステム

本システムでは、学習者とシステムが音声によって会話をする必要がある。しかし、非母語話者の一般的な英語音声の認識は難しい [36]。そのため、一般的な内容の発話を許してしまうと、誤認識のために会話がうまくいかない事態が想定され、学習システムとして問題となる。そこで、事前に会話内容を自習したうえでシステムと対話するという想定 [8] でシステムを構築した。すなわち、英会話の内容を学習者に事前にシナリオとして提示し、学習者は暗記した文を使って会話を行うことを想定する。この想定には、学習者の発話がほとんど既知でないとシステムを安定して

動かすことができないという消極的な意味だけでなく、暗記した発話を繰り返し実際に発声することで英語の運用能力が向上するという効果への期待も込められている。たとえば、単に英語を黙読するだけでなく、繰り返し音読することによって英語の発話能力が向上するという主張がある [37], [38]。また、外国語の言語表現が「考えないで出てくる (automatization)」ためには、文の一部を指定の単語に置き換えて発話するような「文構造の練習」ではなく、自然な文脈の中での決まった表現を繰り返し発話練習すべきであるという主張がある [39]。我々はこれらの主張に基づき、あらかじめ決められたシナリオであっても、繰り返し練習することで言語運用能力が向上すると仮定している。

以上の仮定に従えば、システムとしては学習者の発話内容が事前に分かっていることになるので、対話システムは学習者の発話に対して、あらかじめ決まった応答を返し、シナリオを進行させるもので十分である。これを簡易に実現するため、一問一答型の対話システムをベースにして作成した [40]。しかし、学習者がシナリオを暗記した状態であっても、記憶のあいまいさや緊張などにより、正しく発話ができない場合が想定される。明らかに発話が誤っているときにも固定した応答を返すことは学習システムとして望ましくないため、学習者の発話がある程度シナリオに一致していれば答えを返し、シナリオと一致しない場合には再発話を促すこととした。具体的には、認識された文に含まれる単語集合を W_u 、対応するシナリオの文に現れる単語集合を W_s とするとき、閾値 $0 \leq \theta \leq 1$ を用いて、

$$\frac{|W_u \cap W_s|}{|W_u|} \geq \theta \quad (1)$$

であるときに学習者の発話を受理する。

学習者発話を受理した後、一定時間後にシステムの応答を再生する。このとき、学習者とシステムの間の変換潜時の長さは、予備実験に基づいて設定した。すなわち、英語ネイティブ話者 (アメリカ人) と日本人大学院生 3 名とがシナリオに即した対話を行い、その際の変換潜時を調べた結果、変換潜時が約 1 秒であったために、システム応答までの時間を 1 秒とした。

英文のシナリオとして、帽子を買いに行く内容と、T シャツを買いに行く内容の 2 つを用意した。帽子を買いに行くシナリオを表 1 に、T シャツを買いに行くシナリオを表 2 に示す。シナリオは英語ネイティブ話者 (アメリカ人) による校正を受けた。音声認識には Julius [41] を使い、音響モデルは ERJ コーパス [42] 中の日本人による英語発話音声から学習した。言語モデルはシナリオ中の英文をすべて用いて学習した。音声合成には Festival [43] を使用し、女性音声を合成した。

2.3 CG キャラクタとタイムプレッシャー表現

使用した CG キャラクタを図 1 に示す。CG キャラク

表 1 帽子を買いに行くシナリオ
Table 1 Scenario of buying a hat.

Speaker	Speech
Learner	Hello.
System	Hello, may I help you?
Learner	Yes, I'm looking for a hat. Do you have one?
System	Yes, we do. What kind do you want?
Learner	A green one.
System	Like this?
Learner	Yes, like that one. Can I see it?
System	Yes. Here you are. Would you like to buy it?
Learner	I'm sorry. This isn't exactly what I wanted.
System	How about another product?
Learner	No, thank you.

表 2 T シャツを買いに行くシナリオ
Table 2 Scenario of buying a T-shirts.

Speaker	Speech
Learner	Hello.
System	Hello, may I help you?
Learner	Yes, do you have T-shirts?
System	Yes, we do. What kind do you want?
Learner	A blue one.
System	Like this?
Learner	Yes, like that one. Can I try this on?
System	Yes. Here you are. Would you like to buy it?
Learner	I'm sorry. This T-shirts doesn't look good on me.
System	How about another product?
Learner	No, thanks.

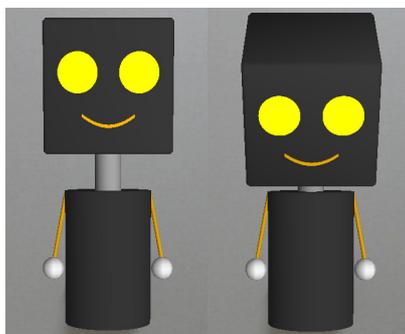


図 1 CG キャラクター
Fig. 1 CG character.

タの描画には ARToolkit [44] を用いた。CG キャラクターは首を縦に振る頷き動作のみを行い、表情を含むその他の部分は変化しない。CG キャラクターにはいくつかの表示方法があるが [34], 本研究では対話している状況をより現実感のあるものにするために、ヘッドマウントディスプレイ (SONY HMZ-T2) に Web カメラ (Logicool HD Pro Webcam C920) を固定し、ビデオスルー環境を構築した。これを使用して CG キャラクターを学習者に提示した。CG キャラクターの大きさ、学習者との距離の統制は文献 [45] に準じ、図 2 のようにし、実験参加者が CG キャラ

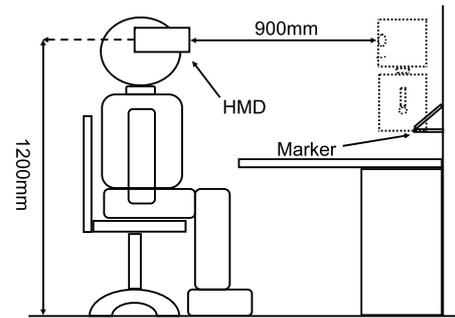


図 2 実験統制図 (人-システム)
Fig. 2 Experimental setup (human-system).

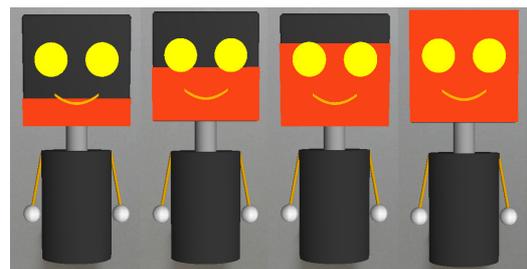


図 3 タイムプレッシャー表現
Fig. 3 Effect of time pressure.

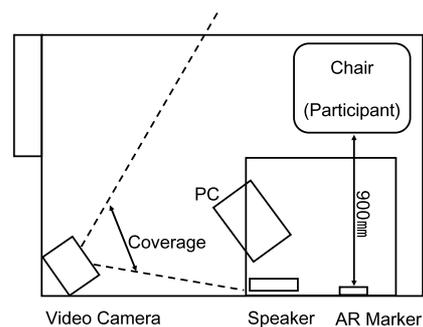


図 4 防音室内の鳥瞰図
Fig. 4 Bird's-eye view of soundproof chamber.

クター全体をとらえられるように設定した。

タイムプレッシャーは図 3 のように CG キャラクターの頭部の下部から 1 段階ずつ赤い部分が増えることによって与えた。たとえば、1 秒ごとに赤い部分が増える場合は 4 秒後に頭部が赤くなりきる。タイムプレッシャーは 4 段階であり、4 回目で頭部が赤くなりきるように設定した。学習者の発話が検出されるとタイムプレッシャー表現が停止する。また、タイムプレッシャー表現はターンごとのシステムの発話終了時にリセットされる。タイムプレッシャー表現の有無にかかわらず、CG キャラクターはつねに頷く動作を行う。これは CG キャラクターが話し相手として自然であるという感覚を学習者に持たせるためである [45]。頷き動作は 1 分間に 30 回のペースで行われ、動作は 30 [fps] で表現している。

2.4 実験環境

実験の環境を図 4 に示す。システムとの対話は防音室で

行い、その様子をビデオカメラで撮影した。合成音声はスピーカで呈示し、音声入力には接話マイクを用いた。実験参加者の位置を統制したうえで、ヘッドマウントディスプレイの見え方、音声入力、再生される音声の音量を調整した後に実験を行った。

3. CG キャラクタの有無の効果

前述のとおり、本システムではCG キャラクタを対話相手として英会話練習を行う。音声を用いた対話において、仮想キャラクターやロボットなどのエージェントを対話相手としたシステムは数多く [46]、またそれを手軽に作成できるツールキットも存在する [47]。しかし、対話の目的に即して、それらのエージェントがどのように有効なのかをきちんと調べた研究はあまり多くない [33]。本稿では、対話相手であるCG キャラクタの意義を実験により確かめる。

3.1 実験条件

上記で検討した実験システムを使用し、CG キャラクタのいる場合といない場合の2条件を単一の実験参加者に実施することで、CG キャラクタの有無が学習者の交替潜時および英会話練習にどのような効果をもたらすかについて調査した。本章における実験結果の仮説は2つである。1つは、タイムプレッシャー表現のように学習者に応答タイミングを意識させる表現を付加していないため、英会話の一度目の練習から学習者の交替潜時が変化しないというものである。もう1つは、何も描画しない場合と比較して対話相手としてCG キャラクタを描画したほうが、学習意欲など、英会話学習に対し有効であるというものである。

実験参加者数は12人（男性11人、女性1人）であり、いずれも大学生または大学院生である。実験参加者の英語学習歴は10年前後で、英会話教室への通学や留学経験はなかった。実験条件、使用するシナリオの順序はカウンターバランスをとり実施した。実験手順は以下のとおりである。

- (1) 実験参加者に20分間で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沿ってシステムと対話を行う。CG キャラクタがいる場合といない場合の2条件を実施する。このとき、2回の対話でそれぞれ異なるシナリオを用いて対話を行う。
- (4) 実験に関するアンケートに記述してもらう。

手順(1)の学習時間は予備実験により十分にシナリオを暗記することができる長さであることを確認した。また、手順(2)の筆記試験は実験参加者がシナリオを暗記していることを確認するために実施した。これは、実験参加者の記憶のあいまいさが対話に影響することを避けるためである。筆記試験の内容は、シナリオ中の実験参加者が発話する部分をすべて書くというものであった。

手順(3)のシステムとの対話の前に、実験参加者には以下の2つのディレクションを行った。

- システムから応答がない場合、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考えて発話してもよいこと

1つ目は誤認識などが起きた場合の対応であり、人間同士でも相手の声が聞こえなかった場合に起こりうる状況を想定した設定とした。また、手順(3)においてCG キャラクタのいない場合はヘッドマウントディスプレイを装着しない状態で実験を行った。この場合、実験参加者はスピーカだけが置かれた机に向かって会話を行うことになる。

実験後のアンケートは5段階評価と実験条件の2者選択、自由記述の項目で構成した。評価値を用いたアンケートの項目は以下のとおりである。

- 話しやすさ
- 対話のスムーズさ
- 練習をしている感覚

これに対し、評価値を、5：非常に良い、4：良い、3：普通、2：良くない、1：非常に良くないと対応させ評価してもらった。実験条件の2者選択に関する以下の6項目についてはCG キャラクタのいる場合といない場合のどちらかを選んでもらった。

- 話しやすかったのはどちらか
- スムーズに対話が行われたのはどちらか
- 楽しいと感じたのはどちらか
- 学習を続けようと思うのはどちらか
- 練習している感じがあったのはどちらか
- 緊張感があったのはどちらか

最後に自由記述の項目として、実験の感想や意見を記入してもらった。

3.2 実験結果

筆記試験により、12人の実験参加者が正しくシナリオを覚えられていることを確認した。実験終了後、撮影したビデオ映像と音声を用いて各セッションの交替潜時を取得した。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替潜時の平均を図5に示す。図中の誤差棒は標準偏差である。CG キャラクタのいる場合といない場合について対応のあるt検定を行ったところ、CG キャラクタの有無による実験参加者の交替潜時の間に有意差は得られなかった。また、いずれの場合においても交替潜時の値は1秒を超えており、英語ネイティブ話者の標準的な交替潜時より長いことが分かる。

次にCG キャラクタがいる場合といない場合での実験参加者の主観的な評価の違いに関する結果をまとめる。表3に5段階評価値を用いたアンケートの結果を示す。各質問に対しそれぞれの実験条件ごとに集計した評価値の平均値

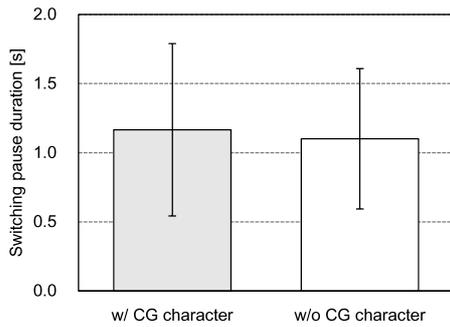


図 5 実験参加者の交替潜時の長さの平均

Fig. 5 Average of participant's switching pause duration.

表 3 主観評価値の平均

Table 3 Average of subjective score.

Question	w/ CG char.	w/o CG char.
話しやすさ	3.83 (0.58)	2.67 (0.78)
対話のスムーズさ	3.33 (0.99)	2.50 (1.09)
練習をしている感覚	3.92 (0.67)	3.00 (0.85)

表 4 CG キャラクタありを選択した人の割合

Table 4 Preference ratio (CG character).

Question	Ratio (%)
話しやすかったのはどちらか	83
スムーズに対話が行われたのはどちらか	67
楽しいと感じたのはどちらか	100
学習を続けようと思うのはどちらか	92
練習している感じがあったのはどちらか	100
緊張感があったのはどちらか	50

を示しており、括弧内はその標準偏差を表している。3つの質問について、それぞれCGキャラクタの有無について対応のあるt検定を行った。その結果、「話しやすさ」に関しては有意差 ($t(11) = 2.20, p = 0.000518 < .01$) が認められた。「対話のスムーズさ」に関しては有意差が得られなかった。「練習をしている感覚」に関しては有意差 ($t(11) = 2.20, p = 0.00474 < .01$) が認められた。

表 4 にCGキャラクタがいた場合を選択した実験参加者の割合を示す。「緊張感があったのはどちらか」以外の項目は過半数がCGキャラクタありの場合を選択したことが分かる。「スムーズに対話が行われたのはどちらか」という質問に対しては、評価値を用いた対話のスムーズさに関する質問では有意差が得られなかったものの、CGキャラクタが存在したほうがスムーズに感じたという実験参加者の数が多いことが分かる。

3.3 考察

交替潜時に関する結果から、CGキャラクタが存在し、表情変化がなく、頷いているだけでは学習者の交替潜時は変化しないということが分かった。これは最初に立てた仮説を支持する結果である。したがって、CGキャラクタを

描画したとしても、学習者が英会話の一度目の練習から応答タイミングを意識して練習するのは困難であるという可能性が示唆された。

また、評価値を用いた主観評価結果から、英会話練習を行う際に、CGキャラクタが対話相手として存在することは学習者の話しやすさを向上させ、かつ練習の充足感を高めることが示された。しかしながらスムーズな対話には寄与していないことが主観評価の結果からも明らかとなった。

CGキャラクタのいる場合といない場合の2条件のどちらかを選択するアンケートの結果から、英会話練習の際はCGキャラクタが対話相手として存在した方が話しやすく、楽しさを感じやすいものであることや、学習意欲を向上させるものであることが明らかとなった。注視点がある方が話しやすいという感想があり、対話相手が見えることによってこのように英会話練習全体の満足度が向上したのではないかと考えられる。「緊張感があったのはどちらか」という質問については、自由記述のアンケートを参考にすると、対話相手が見えない方が緊張を感じるという意見もあれば、対話相手に見られている感覚があるゆえに緊張するという意見もあり、個人差が大きいということが分かった。

以上、CGキャラクタの存在によって、主観的な英会話練習に対する満足度の向上に役立ち、学習意欲の向上に寄与することが実験結果から明らかになった。すなわち対話相手としてCGキャラクタを描画することは英会話練習において有効である。しかし、客観的な結果から、学習者の交替潜時に変化を及ぼすことはないことも分かった。したがって、英会話練習の際にはCGキャラクタが存在した方が良いが、一度目の練習から学習者に対して応答タイミングを考慮させるには時間を意識させるための工夫が必要である。

4. タイムプレッシャー表現の付加に関する検討

4.1 実験概要

3章の結果から、音声対話型CALLシステムでの英会話練習において、CGキャラクタを対話相手として利用することは有効であることが明らかになった。しかし、CGキャラクタそのものだけでは交替潜時の値は変化しなかった。そこでCGキャラクタに2.3節の表現を付加し、実験により効果を検証する。CGキャラクタに付加したタイムプレッシャーは1秒ごとに赤い部分が増えていくように設定した。

実験参加者数は14人(男性12人、女性2人)であり、3章での実験を行っていない大学生または大学院生に依頼した。実験参加者の英語学習歴は11年前後で、英会話教室への通学や留学経験はなかった。実験手順、実験条件は3章での実験と同様にし、実験参加者に課した主観評価アンケートでは、3章で用いたアンケートに加え、タイムプ

レッスン表現についてどのように感じたかの感想を記入する項目を設けた。実験条件、使用するシナリオの順序はカウンターバランスをとり実施した。

手順(3)のシステムとの対話の前に、実験参加者には以下の3つのディレクションを行った。3つ目のディレクションはタイムプレッシャー表現の意図の理解にばらつきが現れるのを防ぐために行った。

- システムから応答がない場合、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考えて発話してもよいこと
- (タイムプレッシャー表現のある場合) キャラクタがすべて赤くなる前に応答すること

4.2 実験結果

筆記試験の結果、12人の実験参加者についてはたかだか単語1個程度の誤りで文を記憶していたのに対し、2人の実験参加者にはシナリオ中の文章の欠落がみられた。よって、この2人についてはシナリオを正しく覚えられていなかったと判断して除外し、以下の分析ではシナリオを正しく覚えていた12人の実験参加者の実験結果のみを利用する。実験終了後、撮影したビデオ映像と音声を用いて各セッションの交替潜時を計測した。

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替潜時の平均を図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の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実験参加者の交替潜時の間には有意差($t(59) = 2.00, p = 0.0490 < .05$)が認められた。タイムプレッシャー表現を導入することで、実験参加者の交替潜時が約200ms短くなっていることが分かる。

次にタイムプレッシャー表現がある場合とない場合での実験参加者の主観的評価の違いに関する結果をまとめる。表5に5段階評価値を用いたアンケートの結果を示す。各質問に対しそれぞれの実験条件ごと(w/ TP:タイムプレッシャーあり, w/o TP:タイムプレッシャーなし)に集計した評価値の平均値を示しており、括弧内はその標準偏差を表している。3つの質問について、それぞれタイムプレッシャー表現の有無について対応のあるt検定を行った。「話しやすさは」に関しては有意差($t(11) = 2.201, p = 0.0000356 < .01$)が得られた。「スムーズに対話はできたか」に関しては有意差($t(11) = 2.20, p = 0.377 > .05$)が得られなかった。「練習をしている感覚は」に関しては有意差は得られなかったが有意傾向($t(11) = 2.20, p = 0.0527 < .10$)となった。

表6にタイムプレッシャー表現がある場合を選択した実験参加者の割合を示す。タイムプレッシャー表現がない方を支持している項目は評価値を用いたアンケート結果と同

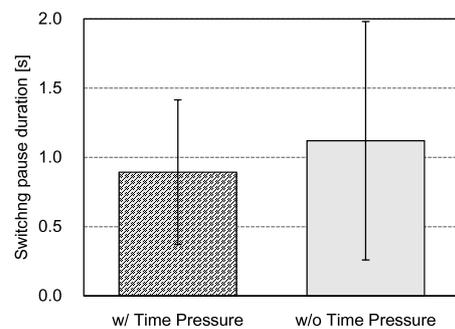


図6 学習者の交替潜時の長さの平均
Fig. 6 Average of participant's switching pause duration.

表5 主観評価値の平均
Table 5 Average of subjective score.

Question	w/ TP	w/o TP
話しやすさ	2.25 (0.75)	4.33 (0.492)
対話のスムーズさ	2.83 (1.19)	3.33 (1.23)
練習をしている感覚	4.17 (0.84)	3.17 (1.11)

表6 タイムプレッシャー表現ありを選択した人の割合
Table 6 Preference ratio (Time pressure).

Question	Ratio (%)
話しやすかったのはどちらか	0
スムーズに対話が行われたのはどちらか	25
楽しいと感じたのはどちらか	58
学習を続けようと思うのはどちらか	50
練習している感じがあったのはどちらか	83
緊張感があったのはどちらか	100

様の傾向を示している。学習を続けようと思うかどうかに関しては意見が分かれており、対話の楽しさ、練習をしている感覚の有無に関してはタイムプレッシャー表現ありの場合により感じたと選択する実験参加者が多いことが分かる。緊張感に関しては、プレッシャーを与えているということもありタイムプレッシャー表現ありの場合を全実験参加者が選択していた。

4.3 考察と検討

タイムプレッシャー表現の有無について比較した交替潜時の実験結果から、CGキャラクタにタイムプレッシャー表現を付加することによって、学習者の交替潜時を抑制することができる可能性が示された。自由記述のアンケートを参考にすると、多くの実験参加者がタイムプレッシャー表現について、赤い部分が増加する速さに焦りなどを感じたとあった。ディレクションによって赤い部分の増加をプレッシャーとして実験参加者に与えさせているため、プレッシャーが増加する速さが発話を促すことにつながっていると考えられる。

また、主観評価の結果から、英会話練習を行う際に、CGキャラクタからタイムプレッシャーが与えられることに

よって、話しづらくなってしまいう傾向があることが明らかとなった。タイムプレッシャー表現を付加したことにより、実験参加者は話すことのほかに時間経過を意識したため、話しやすさがタイムプレッシャー表現のない場合に比べて低下したと考えられる。さらに、主観的にはスムーズになったかどうかという判断はしにくいことも同様に分かった。練習をしている感覚についてはタイムプレッシャー表現があった方が上昇する可能性も示唆された。実際、練習をしている感覚に関してはタイムプレッシャー表現がある場合を支持している人が多く、時間経過の意識をしたうえで英会話練習を行うことの有効性が示されたのではないかと考えられる。また対話の楽しさに関して、タイムプレッシャー表現がある場合に「より感じた」と選択する実験参加者が多く、英会話練習に関して有用である可能性が示された。緊張感に関しては、プレッシャーを与えているということからタイムプレッシャー表現がある場合を全実験参加者が選択していた。また、自由記述のアンケートに「急いで会話しようという感じが実際に会話を行うときの会話をとぎれないようにしようという感じ」と似ており、CG キャラクタのみでは不可能だった練習が可能であることが示唆されていた。同様の意見は複数の実験参加者から得られた。

以上の結果から、タイムプレッシャー表現を付加した状態で英会話練習を行うことで、学習者の交替潜時は抑制され、主観的にもより実践的な練習になっているといえる。しかし、話しやすさが低下したことなどに表れているように、英会話練習時に学習者にかかる負担は大きくなると考えられる。このように負荷がある状態であったとしても、応答タイミングの練習としてタイムプレッシャー表現を利用することの有効性を議論するためには人との練習などと比較を行うことや、繰り返しシステムを使用して練習することによる変化を調査する必要がある。

5. 繰り返し練習および期間を置いた場合の効果

5.1 実験概要

CG キャラクタとの英会話練習時において、タイムプレッシャー表現がある状態で対話を行うことで、応答タイミングを考慮し対話を行う傾向があることが4章の実験により示された。しかしながら、タイムプレッシャー表現を付加した状態で英会話練習を繰り返し行うことが、どの程度有効であるかは明らかになっていない。そこで、提案システムによる英会話練習の効果について、繰り返し練習、および期間を置いて練習を行った場合の効果に注目して調査し、提案手法での練習が有効であるかどうかを議論する。なお、本研究では学習者の応答タイミングに着目していることから、学習者のターンにおける交替潜時の値が人間同士の対話のそれに近づくかどうかによって、学習の有効性を測る。ここでの検証のポイントは次の4つである。

- タイムプレッシャー表現の変化速度がどのように交替潜時の値に影響するか。これを検証するため、これまでと同じく1秒ごとに赤い部分が増加する条件に加え、0.5秒ごとに赤い部分が増加する条件を新たに設ける。
- 短期間のうちに複数回にわたりシステムを使用することで、学習者の交替潜時の値がどのように変化するか。これを検証するため、実験参加者とシステムおよび実験参加者と人との対話を続けて3回行い、その中の交替潜時の変化を調べる。
- 実際に対話練習をすることが、交替潜時の値にどのように影響するか。これを検証するため、対話を行わずに1人で発話練習を行う場合と比較をする。なお、交替潜時を計測するには対話が必要であるため、システムや人との3回の対話、または3回分の朗読練習が終わった後、改めて別な人との対話を行い、そのときの交替潜時の値を比較する。
- それぞれの条件が、数週間程度の期間を置いて練習を行った場合にどのように学習者に影響するか。これを検証するため、最初の実験の約2週間後に、同様の実験を実施し、最初の実験と結果を比較する。

実験条件は以下の4条件である。

A：CG キャラクタと英会話練習（プレッシャーを1秒ごとに増加）

B：CG キャラクタと英会話練習（プレッシャーを0.5秒ごとに増加）

C：人と英会話練習

D：1人で自身の発話内容を朗読する練習

各実験条件につき10人、計40人（男性33人、女性7人）の大学生または大学院生に実験に協力してもらった。実験参加者の英語学習歴は11年前後で、英会話教室への通学や留学経験はなかった。約2週間後に実施した実験に関しては、各実験条件に割り当てられた10人から5人を無作為に抽出し、計20人（男性18人、女性2人）に実験を依頼した。

練習実験手順は以下のとおりであり、基本的にはこれまでと同じであるが、練習回数が3回であること、最後に人との対話を行うことが異なる。

- (1) シナリオの文を暗記する。
- (2) 筆記試験を行う。
- (3) 暗記したシナリオに従って英会話練習を3回行う。
- (4) 練習した内容に沿って、人と英会話を行う。
- (5) 実験に関するアンケートに回答する。

CG キャラクタとの対話の条件（条件A, B）では、プレッシャーの表現は2.3節と同じであり、キャラクタの顔の赤い部分の増加が1秒ごとであるもの（プレッシャーの遅い条件）と、0.5秒ごとであるもの（プレッシャーの速い条件）の2種類を用意した。条件Cにおいて、練習相手役

には英語上級者 (TOEIC 850 以上) の学生 1 人に協力を依頼した。条件 A~D それぞれで 3 回練習を行い (条件 D の場合は、相手なしで自分の発話を朗読)、最後に同じ内容で人間と英会話を行った。条件 A, B, C における 3 回の練習は、システムなどの準備を行う時間約 10 秒ほどのインターバルを空けて行った。また全条件において、練習後の人との対話における対話相手は筆者のうちの 1 人がすべてを行った。

最初に行う実験 (繰返し効果の検証) では帽子を買いに行くシナリオを、2 週間後に行う実験 (期間を置いた場合の検証) では T シャツを買いに行くシナリオを用いた。人同士が対話をする場面において、すなわち条件 C における練習相手および最後の会話相手の人に対しては、次のような指示を行った。

- 実験参加者の発話する内容をすべて聞き取ったのちに応答を返すこと
- 発話速度を実験参加者ごとに変えず、システムの合成音声の再生速度に近い形で発話すること

人との対話を行う際には、CG キャラクタとの統一をとるため、対話者の顔の間の距離を 900 mm に統制した。また、実験参加者には、対話について 4 の実験と同じディレクションを与えた。

5.2 交替潜時計測結果・考察

実験手順 2 の筆記試験において、すべての実験参加者が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、各セッションの交替潜時を計測した。1 回目の対話における、実験条件ごとの交替潜時の平均を図 7 に示す。誤差棒は標準偏差を表す。条件 D は 1 人で朗読を行う練習のため、3 回の対話において交替潜時を取得できない。よって図 7 には条件 A から C の結果のみが表示されている。3 つの実験条件の 1 要因で 1 元配置分散分析を行ったところ、実験条件間には有意差は見られなかった。すなわち、プレッシャーが増加する速さが交替潜時に影響するとはいえない、という結果であった。また、2 回目の対話、3 回目の対話に関して同様の分析を行ったが、どちらも有意差は得られなかった。

次に、各条件内での回数間における実験参加者の交替潜時の違いに着目する。したがって、各回数で交替潜時を取得できる条件 A から C のみでの分析となる。条件ごとに、対話回数の 1 要因で 1 元配置分散分析を行ったところ、条件 A ($F(2) = 3.05, p = 0.0149 < .05$) と条件 B ($F(2) = 3.05, p = 0.0230 < .05$) で有意差が得られた。条件 C には有意差はなかった。条件 C は人との練習のため、1 回目の練習から交替潜時の値が低く、3 回の練習では有意差が得られるほど交替潜時の値が短くならなかったのだと考えられる。Tukey 法によるそれぞれの条件内で多重比較検定の結果、条件 A のみは 1 回目と 2 回目の対

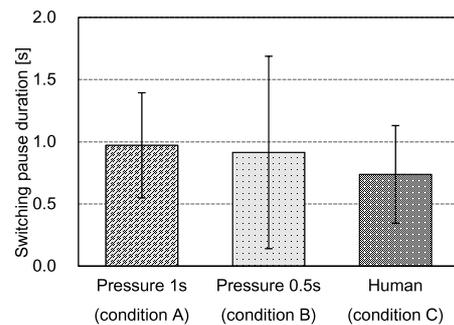


図 7 対話 1 回目における学習者の交替潜時の平均

Fig. 7 Average of participant's switching pause duration (first time).

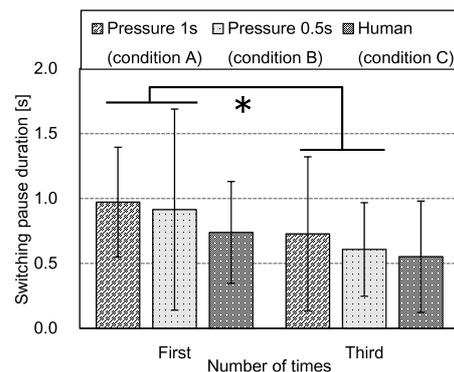


図 8 対話 1 回目と 3 回目における学習者の交替潜時の平均 (**: $p < .01$, *: $p < .05$)

Fig. 8 Average of participant's switching pause duration (first and third time, **: $p < .01$, *: $p < .05$).

話間に有意差 ($p = 0.0312 < .05$) が、条件 A から C すべてにおいて、1 回目と 3 回目の対話間に有意差 (条件 A: $p = 0.0319 < .05$, 条件 B: $p = 0.0274 < .05$, 条件 C: $p = 0.0450 < .05$) が認められた。図 8 に 1 回目と 3 回目の対話時における各条件の交替潜時の平均を示す。1 回目と比較し、3 回目における実験参加者の交替潜時は、条件 A と条件 C は約 200 ms, 条件 B では約 300 ms 減少している。人同士での交替潜時に近づいているのは条件 B の方であり、短期間での繰返し練習によって人同士の対話のテンポに近づくにはプレッシャーを速く増加させる方が有効だと考えられる。

各実験条件で 3 回の英会話練習を行った後に、人を相手に対話を行った際の交替潜時の平均を図 9 に示す。誤差棒は標準偏差である。実験条件の 1 要因で 1 元配置分散分析を行ったところ、実験条件間に有意差が認められた ($F(3) = 2.65, p = 0.000352 < .01$)。Tukey 法を用いた条件間の多重比較検定を行った結果、条件 B と D ($p = 0.000573 < .01$), 条件 C と D ($p = 0.00231 < .01$) の間に有意差が認められた。すなわち、速いプレッシャーの増加によってタイムプレッシャーをかけて行う練習は人同士での英会話練習によって得られる効果と同様の傾向を示しており、1 人で朗読する練習よりも交替潜時が短期間

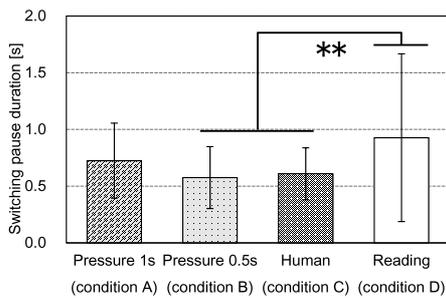


図9 練習後の人との対話における学習者の交替潜時の平均 (** : $p < .01$, * : $p < .05$)

Fig. 9 Average of participant's switching pause duration (human, **: $p < .01$, * : $p < .05$).

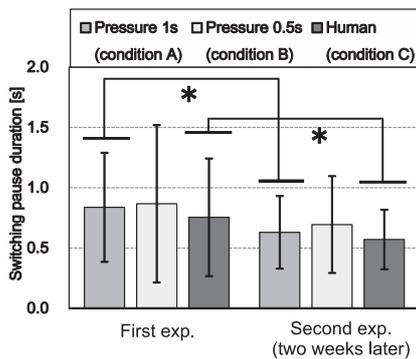


図10 2回の実験における学習者の交替潜時の平均 (** : $p < .01$, * : $p < .05$)

Fig. 10 Average of participant's switching pause duration (difference of experiment, **: $p < .01$, * : $p < .05$).

で有意に減少することが明らかになった。

図10は2週間間隔で行った2回のシステム使用実験における、3回の対話中の交替潜時の平均を各条件について示したものである。誤差棒は標準偏差である。3回の対話それぞれから交替潜時を取得できる条件AからCの分析となる。条件Aから条件Cのそれぞれについて、実験間の1要因で1元配置分散分析を行ったところ、実験間で有意差があったものは条件A ($F(1) = 4.04, p = 0.0184 < .05$)と条件C ($F(1) = 4.04, p = 0.0203 < .05$)であり、条件Bでは有意差が得られなかった。この結果から、速いプレッシャーの増加によってタイムプレッシャーをかけることは短期間での効果が得られる利点があるが、日を改めて再度練習を行う場合はプレッシャーを遅く増加させる方が効果があることが確認された。

5.3 主観評価アンケート結果

主観評価アンケートは5段階の評価値を用いたもの、二者選択そして自由記述の項目で構成されている。以下の項目について、評価値を、5:非常に良い, 4:良い, 3:普通, 2:良くない, 1:非常に良くないと対応させ評価してもらった。自由記述のアンケートには実験の感想や意見を記述してもらった。

表7 主観評価値の平均 (1回目の実験)

Table 7 Average of subjective score (the 1st experiment).

Question	Conditions			
	A	B	C	D
Q1	4.10 (0.74)	3.90 (0.74)	4.20 (0.63)	3.50 (1.27)
Q2	4.60 (0.70)	4.50 (0.97)	4.60 (0.52)	3.90 (1.20)
Q3	4.80 (0.42)	4.20 (1.03)	3.60 (1.35)	3.60 (1.07)
Q4	4.10 (1.29)	4.10 (1.10)	3.90 (0.57)	3.40 (1.43)

- 3回行った練習について

Q1: 練習は英会話練習になった

Q2: 練習は読み書きで練習するよりも英会話練習として役立った

- 練習後の人との英会話について

Q3: 自信を持って話せた

Q4: 思いどおりに話せた

短期間での繰り返し練習実験において各条件10人から取得した結果を表7に示す。各質問に対しそれぞれの実験条件ごとに集計した評価値の平均値を示しており、括弧内はその標準偏差を表している。3回の練習に関する2つの質問Q1, Q2において、実験条件および実験参加者間の2要因について繰返しのない2元配置分散分析を行った結果、いずれの要因についても有意差は認められなかった。

練習後の対話に関する2つの質問Q3, Q4において、実験条件および実験参加者間の2要因について繰返しのない2元配置分散分析を行ったところ、Q3の実験条件間だけに有意差が認められた ($F(3, 39) = 2.96, p = 0.0238 < .05$)。Q3についてTukey法を使用して多重比較検定を行った結果、条件Aと条件C ($p = 0.0601 < .10$)、条件Aと条件D ($p = 0.0601 < .10$)の間に有意傾向がみられた。練習後の対話における自信に関しては、タイムプレッシャー表現を用い、プレッシャーが1秒ごとに増加する状態で練習した場合に最も評価されていることが分かる。

最後に2週間後の実験において、最初の実験からの違いについて、次の3つの項目を5段階評価で尋ねた。

Q1: 練習は英会話練習として非常に効果的だと感じた

Q2: 人との英会話そのものはとてもうまくなった

Q3: 練習機会がさらに増えることで英会話の練習効果が期待できる

その結果を表8に示す。各質問に対しそれぞれの実験条件ごとに集計した評価値の平均値を示しており、括弧内はその標準偏差を表している。Q1~Q3について、それぞれ実験条件および実験参加者間の2要因について繰返しのない2元配置の分散分析を行った。その結果、Q1およびQ2についてはいずれの要因についても有意差は認められなかった。Q3については実験条件間に有意差 ($F(3, 19) = 3.49, p = 0.00590 < .01$) が認められた。Tukey法を用いて多重

表 8 主観評価値の平均 (2 回目の実験)

Table 8 Average of subjective score (the 2nd experiment).

Question	Conditions			
	A	B	C	D
Q1	4.60 (0.55)	4.60 (0.55)	4.20 (0.45)	3.20 (1.48)
Q2	3.60 (0.89)	3.60 (1.14)	3.40 (0.55)	2.80 (1.10)
Q3	4.80 (0.45)	4.40 (0.55)	5.00 (0.00)	3.40 (0.89)

比較検定を行った結果, 条件 A と D ($p = 0.00650 < .01$), 条件 C と D ($p = 0.00210 < .01$) の間に有意差があることが分かった. 条件 B と比較して, 条件 A は繰り返し練習することによる期待が強いことが分かる.

5.4 考察と検討

短期間での繰り返し実験における, システムを利用した練習および人との練習を行った一度目の英会話練習での結果では, プレッシャーの増加する速さに関して有意差は見られなかった. 3 回の練習による学習者の交替潜時の減少値に関しては, プレッシャーの増加が速い方が約 300 ms であり, プレッシャーの増加が遅い場合と比較して約 100 ms 多く減少していることが分かる. さらに練習後の人との対話においても, 1 人で朗読を行う練習との間に有意に違いが現れたのはプレッシャーの増加が速い場合と人と練習した場合である. こうした点をふまえると, 英会話練習において人との練習と同等の効果を得るには, プレッシャーの増加を速くした方が良いという結果であった. 一方, 2 週間後の使用実験での結果では, 最初の実験と比較して, プレッシャーの増加が遅い場合の練習と, 人同士の練習において交替潜時の値が減少しており, その差に有意差があることを確認した. したがって, 続けて学習を行うことの考慮した場合に関しては, プレッシャー表現は速くしすぎない方が効果的である.

主観評価の結果のうち, 差が見られた表 7 の Q3 と表 8 の Q3 から, プレッシャーの増加が 1 秒ごとであった場合は他の練習条件と比較して自信を持って話すことができ, また繰り返し練習への期待が大きいという結果が得られている.

以上の結果の理由, 特にプレッシャーの速さの効果が短期の練習と期間を置いた練習で異なっていることについて, この結果だけから確実なことをいうことは難しい. 1 つの解釈としては, プレッシャーの速さが速い場合には, 繰り返し練習中には集中して素早く答えるようになるものの, 速く答えることに集中しすぎるために, 英語学習自体のモチベーションが低下するのかもしれない. 英会話においては, 練習回数を重ねることが重要であるため, 学習意欲の観点から, より多くの学習者に続けて練習してもらうためには適切なプレッシャー表現によって学習者の発話を促す必要があると考えられる.

6. 結論

本稿では, 英会話学習を行うための対話型 CALL システムに着目し, 特に適切な応答タイミングを習得するための方法を提案するとともに, その有効性を検証した.

まず初めに, CG キャラクタを対話相手として設定し, 対話練習の最初から学習者の交替潜時に変化があるかどうかを調べた. その結果として, 表情変化などのない CG キャラクタを描画しただけでは学習者の交替潜時は変化しないことを確認した. 一方, 学習意欲の向上など, 主観的な点において英会話学習に有効であることも明らかになった.

次に, CG キャラクタを利用し, 学習者にタイムプレッシャー表現がある場合とない場合の 2 条件で英会話練習を行わせた. その結果, 学習者の交替潜時には条件間で有意差があり, タイムプレッシャー表現によって抑制されることを確認した. さらに, 主観的にも緊張感のある実践的な練習に近づいていることが示された.

最後に, タイムプレッシャー表現を付加した CG キャラクタとの英会話練習について, 人同士での練習, 対話を行わない練習と比較するとともに, 繰り返し練習を行うことによる変化についても分析を行った. タイムプレッシャー表現に関してはプレッシャーの増加の速さを 2 種類用意し, 分析を行った. その結果, プレッシャーの増加が速い方が, 短い期間での練習による効果が大きいことが分かった. しかし, 期間を置いて行った練習では, プレッシャーの増加が遅い方が有効であることも明らかになった. いずれの場合においてもタイムプレッシャー表現を付加した英会話練習において, 交替潜時の値は減少し, 人同士での練習に近づくことが示された. したがって, タイムプレッシャー表現を付加した状態で CG キャラクタと英会話練習を行う提案手法の有効性が示された.

今後の課題としては大きく分けて 2 つあげられる. 1 つはタイムプレッシャーのかけ方, あるいは表現の仕方に関する分析である. 5 章におけるアンケートの結果から, 対話相手は人に近いデザインを望む意見も多くみられた. そういった対話相手のデザインのほかに, タイムプレッシャー表現そのものの表現方法についても赤い部分を増やすほかに適切な表現がないか検討する必要がある. もう 1 つは, 5 章での実験よりも長期にわたり使用し, どのような効果があるかを検討することである. その際には複数のシナリオを用いるなど, より難しい英会話練習を行うことが考えられる. こうした練習に対応するためには, 学習者の言いよどみやシナリオからの逸脱にも対応できる, より柔軟性の高い対話システムを設計する必要がある.

謝辞 本研究は, JSPS 科研費 (挑戦的萌芽) 24652111 の助成を受けた.

参考文献

- [1] Skype, available from <http://www.skype.com/> (accessed 2015-01).
- [2] 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), Vol.J96-D, No.7, pp.1549-1565 (2013).
- [3] Hjalmarsson, A., Wik, P. and Brusik, J.: Dealing with DEAL: A dialogue system for conversation training, *Proc. SIGdial*, pp.132-135 (2007).
- [4] Bax, S.: CALL-past, present and future, *System*, Vol.31, No.1, pp.13-28 (2003).
- [5] Zhao, Y.: Recent Developments in Technology and Language Learning: A Literature Review and Meta-Analysis, *CALICO Journal*, Vol.21, No.1, pp.7-27 (2003).
- [6] Warshauer, M. and Healey, D.: Computers and language learning: An overview, *Language Teaching*, Vol.31, No.2, pp.57-71 (1998).
- [7] Eskenazi, M.: An overview of spoken language technology for education, *Speech Communication*, Vol.51, No.10, pp.832-844 (2009).
- [8] Kweon, O.-P., Ito, A., Suzuki, M. and Makino, S.: A grammatical error detection method for dialog-based CALL system, *J. Natural Language Processing*, Vol.12, No.4, pp.137-156 (2005).
- [9] Anzai, T. and Ito, A.: Recognition of utterances with grammatical mistakes based on optimization of language model towards interactive CALL systems, *Proc. AP-SIPA ASC* (2012).
- [10] Raux, A. and Eskenazi, M.: Using Task-Oriented Spoken Dialogue Systems for Language Learning: Potential, Practical Applications and Challenges, *Proc. INSTIL/ICALL* (2004).
- [11] Lee, S., No, H., Lee, J., Lee, K., Lee, G.G., Sagong, S. and Kim, M.: On the effectiveness of Robot-Assisted Language Learning, *ReCALL*, Vol.23, No.1, pp.25-28 (2011).
- [12] Nation, P.: The role of the first language in foreign language learning, *The Asian EFL Journal*, Vol.5, No.4 (2003).
- [13] 長岡千賀, 小森政嗣, 中村敏枝: 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—, 信学技報告 HCS, Vol.103, No.113, pp.19-24 (2003).
- [14] Trimboli, C. and Walker, M.B.: Switching pauses in cooperative and competitive conversations, *J. Experimental Social Psychology*, Vol.20, No.4, pp.297-311 (1984).
- [15] Boltz, M.G.: Temporal Dimensions of Conversational Interaction: The Role of Response Latencies and Pauses in Social Impression Formation, *J. Language and Social Psychology*, Vol.24, No.2, pp.103-138 (2005).
- [16] Sacks, H., Schegloff, E.A. and Jefferson, G.: A simplest systematics for the organization of turn-taking for conversation, *Language*, Vol.50, No.4, pp.696-735 (1974).
- [17] Carroll, D.: Precision Timing in Novice-to-Novice L2 Conversations, *Issues in Applied Linguistics*, Vol.11, No.1, pp.67-110 (2000).
- [18] Hosoda, Y. and Aline, D.: Turn-taking timing in English activity classes in Japanese public elementary schools, 神奈川大学言語研究, Vol.28, pp.89-105 (2005).
- [19] Galaczi, E.D.: Interactional Competence across Proficiency Levels: How do Learners Manage Interaction in Paired Speaking Tests?, *Applied Linguistics*, Vol.35, No.5, pp.553-574 (2014).
- [20] Heldner, M.: Detection thresholds for gaps, overlaps, and no-gap-no-overlaps, *J. Acoust. Soc. Am.*, Vol.130, No.1, pp.508-513 (2011).
- [21] Sayer, P.: An intensive approach to building conversation skills, *ELT Journal*, Vol.59, No.1, pp.14-22 (2005).
- [22] Heldner, M. and Edlund, J.: Pauses, gaps and overlaps in conversations, *J. Phonetics*, Vol.38, No.4, pp.555-568 (2010).
- [23] Stivers, T., Enfield, N.J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., Hoymann, G., Rossano, F., de Ruiter, J.P., Yoon, K.-E. and Levinson, S.C.: Universals and cultural variation in turn-taking in conversation, *Proc. Nat. Acad. of Sci. of the United States of America*, Vol.106, pp.10587-10592 (2009).
- [24] Thórisson, K.R.: Natural turn-taking needs no manual: Computational theory and model, from perception to action, *Multimodality in Language and Speech Systems*, Granström, B., House, D. and Karlsson, I. (Eds.), Kluwer Academic Publishers, Dordrecht, pp.173-207 (2002).
- [25] Xiao, B., Rozgić, V., Katsamanis, A., Baucom, B., Georgiou, P.G. and Narayanan, S.: Acoustic and Visual Cues of Turn-Taking Dynamics in Dyadic Interactions, *Proc. Interspeech*, pp.2441-2444 (2011).
- [26] 渡辺富夫, 大久保雅史: 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).
- [27] 鈴木紀子, 寛一彦, 竹内勇剛, 岡田美智男: 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).
- [28] 小松孝徳, 森川幸治: 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, Vol.2004-ICS-137, No.10 (2004).
- [29] 長岡千賀, 小森政嗣, 中村敏枝: 対話における交替潜時の2者間相互影響, 日本人間工学会誌, Vol.38, No.6, pp.251-260 (2002).
- [30] Lee, S., No, H., Lee, J., Lee, K. and Lee, G.G.: Foreign Language Tutoring in Oral Conversations Using Spoken Dialog Systems, *IEICE Trans. Inf. & Syst.*, Vol.E95-D, No.5, pp.1216-1228 (2012).
- [31] Wik, P. and Hjalmarsson, A.: Embodied conversational agents in computer assisted language learning, *Speech Communication*, Vol.51, No.10, pp.1024-1037 (2009).
- [32] Bujak, K.R., Radu, L., Catrambone, R., MacIntyre, B., Zheng, R. and Golubski, G.: A psychological perspective on augmented reality in the mathematics classroom, *Computer & Education*, Vol.68, pp.536-544 (2013).
- [33] Miyake, S. and Ito, A.: A spoken dialogue system using virtual conversational agent with augmented reality, *Proc. APSIPA ASC* (2012).
- [34] Wojciechowski, R. and Cellary, W.: Evaluation of learners' attitude toward learning in ARIES augmented reality environments, *Computer & Education*, Vol.68, pp.570-585 (2013).
- [35] Gravano, A. and Hirschberg, J.: Turn-taking cues in task-oriented dialogue, *Computer Speech and Language*, Vol.25, No.3, pp.601-634 (2011).
- [36] van Doremalen, J., Cucchiari, C. and Strik, H.: Optimizing Automatic Speech Recognition for Low-Proficient Non-Native Speakers, *EURASIP J. Audio, Speech and Music Processing*, Vol.2010, No.2 (2010).
- [37] 山本昭夫: 外国語学習における反復練習と多読再考: アフォーダンス理論と熟達化研究が示唆するもの, 人文, Vol.5, pp.81-100 (2006).
- [38] 高山芳樹: 自己モニターを伴う1ヶ月集中音読練習のスピーキング力への影響, 東京学芸大学紀要, 人文社会科学

- 学系.I, Vol.58, pp.37-44 (2007).
- [39] Gatbonton, E. and Sagalowitz, N.: Creative Automation: Principles for Promoting Fluency Within a Communicative Framework, *TESOL Quarterly*, Vol.22, No.3, pp.473-492 (1988).
- [40] 三宅真司, 廣井 富, 伊藤彰則: 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).
- [41] 河原達也, 李 晃伸: 連続音声認識ソフトウェア Julius, 人工知能学会誌, Vol.20, No.1, pp.41-49 (2005).
- [42] 峯松信明, 富山義弘, 吉本 啓, 清水克正, 中川聖一, 壇辻正剛, 牧野正三: 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).
- [43] The Festival Speech Synthesis System, available from (<http://www.cstr.ed.ac.uk/projects/festival/>) (accessed 2015-01).
- [44] 加藤博一: ARToolKit, 映像メディア学会誌, Vol.62, No.1, pp.48-51 (2008).
- [45] 廣井 富, 伊藤彰則: 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).
- [46] Edlund, J., Gustafson, J., Heldner, M. and Hjalmarsson, A.: Towards human-like spoken dialogue systems, *Speech Communication*, Vol.50, No.8-9, pp.630-645 (2008).
- [47] Lee, A., Oura, K. and Tokuda, K.: MMDAgent—A fully open-source toolkit for voice interaction systems, *Proc. ICASSP*, pp.8382-8385 (2013).



鈴木 直人

2013 年東北大学工学部情報知能システム総合学科卒業。2015 年同大学大学院工学研究科博士前期課程修了。日本音響学会会員 (2015 年 3 月まで)。



廣井 富

2005 年石巻専修大学大学院理工学研究科博士後期課程修了。同大学助手, 2006 年東北大学大学院国際文化研究科 COE フェロー, 岩手県立大学ソフトウェア情報学部客員教員, 2009 年大阪工業大学工学部機械工学科講師, 現在, 同大学ロボット工学科准教授。主に生活支援ロボットの開発, 親和性向上に関する研究に従事。博士 (工学)。日本機械学会, 日本ロボット学会, ヒューマンインタフェース学会等の各会員。



千葉 祐弥

2010 年東北大学工学部電気情報・物理工学科卒業。2012 年同大学大学院工学研究科博士前期課程修了。現在, 同大学院工学研究科博士後期課程に在学中。2014 年より日本学術振興会特別研究員 (DC-2)。マルチモーダル対話システムの研究に従事。ACL, 日本音響学会, 電子情報通信学会各会員。



能勢 隆 (正会員)

2001 年 3 月京都工芸繊維大学工芸学部電子情報工学科卒業。2009 年 3 月東京工業大学大学院総合理工学研究科博士後期課程修了。博士 (工学)。同年東京工業大学大学院総合理工学研究科助教。2013 年東北大学大学院工学研究科講師, 現在に至る。音声合成, 音声認識, 音声対話, 音楽情報処理の研究に従事。電子情報通信学会, 日本音響学会, IEEE, ISCA 各会員。



伊藤 彰則 (正会員)

1986 年東北大学工学部通信工学科卒業。1991 年同大学大学院博士課程修了。同年同大学応用情報学研究センター助手。1992 年同大学情報処理教育センター助手。1995 年山形大学工学部講師。1998~1999 年ボストン大学客員研究員。現在, 東北大学大学院工学研究科教授。工学博士。音声言語情報処理, 音声信号処理, 音楽情報処理等の研究に従事。日本音響学会, 電子情報通信学会, ヒューマンインタフェース学会, IEEE 各会員。