

頭部搭載型ディスプレイを用いたスペイン語学習システムの提案 Proposal of a Spanish Language Learning System Using a Head-mounted Display

齊藤真生[†] 櫻井 淳[†] 川合康央[†]
Mao Saito Jun Sakurai Yasuo Kawai

1. はじめに

近年、外国人労働者の増加に伴い、様々な言語によるコミュニケーションの必要性が高まっている。英語以外の外国語として、中国語やスペイン語、ポルトガル語などの言語は、母語話者が多く存在している。文部科学省の調査[1]によれば、英語以外の外国語科目を開設している高等学校は増加傾向にあり、今後の国際化の進展や国内で開催されるオリンピックの影響に伴い、多言語の外国語学習者はさらに増加することが考えられる。

会話を目的とした外国語の学習では、母語話者との直接対話による学習方法が最も効果的であり、対面による会話を通じて円滑なコミュニケーション能力を身に付けることが可能であると考えられる。最近では、人工音声技術の発展やクラウドサービスの提供により、音声認識や対話制御などの技術を手軽に使用することができ、スマートフォンデバイスを用いた英語学習アプリケーション[2, 3]の開発やロボットを用いた対話システムの研究[4]などが行われている。しかし、これらのシステムでは、学習相手がイラストやロボットで作成されたものであり、学習者は実際の人物との会話に近い臨場感や緊張感を得ることができない。また、会話の中の質問に対しても、相手が実際の人物でないため、長時間返答を考えるなど、不自然な会話になることも多く、実際の会話に近い環境とは言い難い。その解決策として CG (Computer Graphics) キャラクタを対話相手と設定し、時間的にその表情を変化させることによって、学習者に実際の状況に近い環境を与える方法が提案され、一定の有効性が示されている[5]。この研究では、ロボットで表現された簡易な CG キャラクタが使用されており、実写に似せた人物との対話による学習効果は明らかにされていない。そこで、本研究では、実際の人物との円滑なコミュニケーション能力を身に付けるための学習環境を提供することを目的とし、ヘッドマウントディスプレイを用いた VR (Virtual Reality) 空間におけるスペイン語対話学習システムを提案する。

2. システム開発

2.1 開発環境

ゲームエンジンは、背景や人体の 3 次元モデルをリアル寄りに表現するため、比較的グラフィック性能が高い UnrealEngine を採用した。このゲームエンジンでは、Blueprint といったプログラミング言語を使用し、CG キャラクタの動作や画面遷移などの制御を行う。また、CG キャラクタの構築には、オープンソースの Make Human を用いて人体モデルの基盤を作成した。次に、作成した CG モ

デルに対してテクスチャデータを作成するため、Substance Painter を使用し、顔や服などのテクスチャデータを作成した。さらに、CG ソフトウェアの Blender を用いて、人物モデルの編集や口の動きや目の動きなどのアニメーションを追加した。音声認識や対話システム関連の処理は、様々なクラウドサービスのライブラリが充実しているプログラミング言語 Python を使用した。また、Python で作成した対話システムとゲームエンジンは、XAMPP を用いて JSON データを HTTP 通信する方式によって連携している (表 1)。

表 1 開発環境

名称	用途
Unreal Engine	3 次元向けのゲームエンジン
Make Human	人体製作用の 3 次元 CG ソフト
Substance Painter	3 次元モデルのテクスチャ作成ソフト
Blender	3 次元モデルの編集ソフト
XAMPP	HTTP 通信用のソフト
Python	音声認識系処理のプログラミング言語

2.2 システムの処理

まず、CG キャラクタからの質問に対して、学習者はマイクを通じて返答を発話する。この発話データを入力データとし、Google Speech Recognition を用いて音声データをテキストデータに変換する。次に、WatsonAssistant で予め作成した学習モデルに発話データを送信し、返答内容を取得する。このとき、CG キャラクタからの質問に対して、学習者の発話内容が適切な場合は正しいと判断し、その返答内容に見合った会話文を取得する。不適切な場合は不正解と判断され、返答内容を理解することが出来ないことを伝える返答文を取得する。これらの返答文を、Watson Text to Speech で音声に変換を行い、最後に UnrealEngine 上で返答分の音声とそれに適切な身振り手振りや口の動きなどのアニメーションを動作させる。これらの処理を繰り返すことによって、一連の会話の流れを実現する (図 1, 2)。

本システムで使用する会話パターンは、大学の講義でスペイン語を履修している学生が学習する範囲を想定し、財団法人日本スペイン協会のスペイン語技能検定 6 級レベルの内容を基準として会話パターンを作成した。このレベルでは、自己紹介や現在形の文といった基本的な日常会話ができる程度のレベルであり、スペイン語検定の中では入門レベルの内容となっている。会話内容は、市販テキストのスペイン語検定 6 級の文法や単語を参照して作成を行った。また、所属大学で開講されているスペイン語の授業の担当講師に確認し、1 セメスタ分 で学習する文法と単語を使用した。本授業は日常会話レベルのコミュニケーション能力の習得に向けて構成されているが、一部の単語は扱わないことを配慮した。会話パターンは、CG キャラクタの質問内容と学習者の返答内容、CG キャラクタの返答内容を 1

[†] 文教大学 Bunkyo University

つのセットとして、学習者からの返答内容に予め設定した特定の単語がすべて含まれている場合に、正しいと判定している。

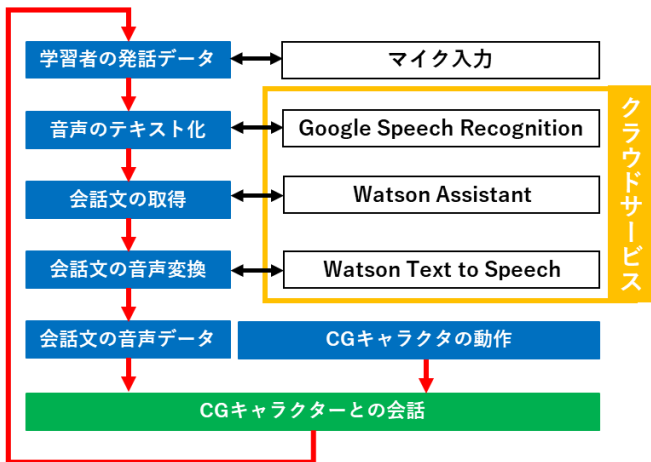


図1 システム処理フロー

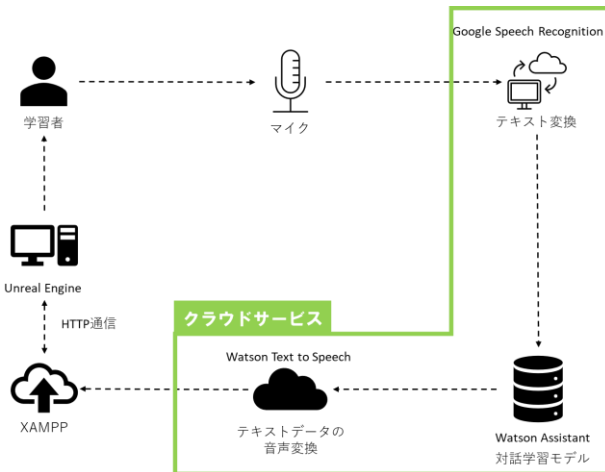


図2 システム構成図

2.3 画面構成

本システムは、仮想空間上のCGキャラクターが発話する質問に対して、学習者は5秒間以内に返答内容を発話することで、会話シナリオが進む構成となっている。このとき、5秒間以内に返答が行えない場合、次の会話シナリオに進む構成とした。CGキャラクターが質問を行った後、タイムバーを表示させることによって、学習者に緊張感を与える仕組みを取り入れた。これは、実際の会話場面において違和感のない時間を考慮し、質問から返答までにかかる時間をとして5秒間に設定した。CGキャラクターは、質問文や返答文に合わせた口の動きのアニメーションを行うことを可能とし、違和感のない実際の人物に似せた動きとなるように工夫した（図3）。

2.4 システム構成

本システムは、ヘッドマウントディスプレイとして、Oculus Rift Sとその付属コントローラ、マイクの動作要件を満たしたGeForce GTX 1070のGPUを搭載したゲーミン

グPCを使用した。この動作要件を満たすことでVR空間上でCGキャラクターとの対話が可能となっている。また、コントローラに関しては、学習者が対話を開始する最初のタイミングのみで使用し、その後は、コントローラは使用せずに自動で対話が進む構成となっている。マイクに関しては、Oculus Rift Sに標準で付属しているが、周りの雑音による環境下の中でも適切な学習者の発話内容を認識できるように、単一指向性のサンワサプライ製USBヘッドセットを用いた（図4）。



図3 システムの動作画面構成



図4 システムの動作

2.5 会話パターン

本システムでは、大学の語学科目としてスペイン語を履修している学生が学習する範囲を想定し、スペイン語技能検定6級の内容を基準に会話パターンを作成した。会話パターンは、CGキャラクターからの質問内容と学習者の返答内容、それに対するCGキャラクターの返答内容を1つのセットとしており、学習者は発話した内容に下線部分の単語がすべて含まれている場合に、正解と判定する。本システムでは、全20種類の会話パターンの中からNo. 1-5、No. 6-10、No. 11-15、No. 16-20の計4パターンに分割し5種類のシナリオを作成した。そして、4パターンの中から1種類を選択し、5個の会話パターンを体験する内容となっている。また、学習者の発話内容や正誤の結果は、CSVファイルとして記録しており、体験後に被験者が返答出来なかった内容や発話内容を確認し、自身が苦手とする項目を確認することで学習効果の向上が期待されるため、このような工夫を取り入れた（表2）。

表2 会話パターンの例

No.	CG キャラクタの質問	学習者の返答 (正解例)	CG キャラクタの返答
1	¿Qué estás haciendo ahora? あなたは何をしていますか?	<u>Estoy estudiando español.</u> 私はスペイン語の勉強をしています。	Estoy hablando Español. 私はスペイン語を話しています。
2	¿De dónde eres? どこ出身ですか?	<u>Soy de Tokyo.</u> 私は東京出身です。	Soy de Barcelona. 私はバルセロナ出身です。
3	¿Dónde está tu libro de español? あなたのスペイン語の教科書はどこにありますか?	<u>Mi libro de español está en mi casa.</u> 私の教科書は家にあります。	No lo tienes ahora. 私はそれを持っていません。
4	¿Dónde vives? どこに住んでいますか?	<u>Vivo en Yokohama.</u> 私は横浜に住んでいます。	Vivo en Tokyo. 私は東京に住んでいます。
5	¿Hay alguien en casa? 家の中に誰かいますか?	<u>Está mi madre. / no hay nadie.</u> お母さんがいます。 / 誰もいません。	¡Mándale mis saludos! お母さんよろしく伝えてください。
6	¿Qué quieres comer? あなたは今なにを食べたいですか?	<u>Quiero comer carne.</u> 私はお肉が食べたいです。	¡Yo también! 私も食べたいです。
7	¿Cuántos años tienes? 何歳ですか?	<u>Tengo veinte años.</u> 私は20歳です。	Tengo veintiuno años. 私は21歳です。
8	¿Te interesa la computadora? あなたはこのコンピュータに興味はありますか?	<u>Si, me interesa.</u> はい、そのコンピュータに興味があります。	A mi me aburre. 私は退屈です。
9	¿Dónde está tu madre? あなたのお母さんはどこにいますか?	<u>Mi madre está en casa.</u> 私のお母さんは家にいます。	Vale, muy bien. いいですね。
10	¿A dónde vas a ir hoy? 今日はこれからどこへ行くつもりですか?	<u>Voy a la Universidad.</u> 私は文教大学に行くつもりです。	Yo voy de compras. 私は買い物に行きます。
11	¿Cómo te llamas? 名前は何かですか?	<u>Me llamo Saitou. / soy Saitou. / Mi nombre es Saitou.</u> 私の名前は斎藤です。	Encantada de conocerla, me llamo Yuri. はじめまして、私はユリです。
12	¿Qué día es hoy? 今日は何曜日ですか?	<u>Hoy es miércoles.</u> 今日は水曜日です。	Tengo clase de Español. 今日はスペイン語の授業があります。
13	¿Quién es tu profesor de español? あなたのスペイン語の先生は誰ですか?	<u>Mi profesor es Terada. / es Terada. / se llama Terad.</u> 寺田という名前です。 / 寺田さんです。	Mi profesor es Sakurai. 私の先生は、櫻井です。
14	¿A dónde vas en las vacaciones? どこで休暇を過ごしますか?	<u>Voy a mi pueblo.</u> 自分の実家に行きます。	Yo voy a Hokkaido. 私は北海道に行きます。
15	¿Te gustan las verduras? あなたは野菜が好きですか?	<u>No me gustan las verduras. / Me gustan las verduras.</u> 私は野菜が好きではありません。 / 私は野菜が好きです。	A mi también. / A mi tampoco. 私も好き。 / 私も嫌い。
16	¿Dónde trabaja tu padre? あなたのお父さんはどこで働いていますか?	<u>Mi padre trabaja en una peluquería.</u> 私のお父さんは床屋さんで働いています。	Mi padre trabaja de Ingeniero. 私のお父さんはエンジニアです。
17	¿A qué hora sales de casa? あなたは何時に家を出ますか?	<u>Salgo a la una y media.</u> 私は午前一時半に家を出ます。	Está bien, nos vamos さあ、行きましょうか。
18	¿Qué estudias? あなたはどの勉強をしていますか?	<u>Estudio español.</u> 私はスペイン語の勉強をしています。	Yo estudio italiano 私はイタリア語の勉強をしています。
19	¿Qué tiempo hace hoy? 今日はどんな天気ですか?	<u>Está soleado.</u> 今日は晴れています。	¡Gracias! ありがとう。
20	¿Puedes ir conmigo a ver una película? 私と一緒に映画を見に行ってくださいませんか?	<u>Puedo ir contigo.</u> 私はあなたと映画を見に行けます。	¡Qué bien! / ¡Qué lástima! それはよかった。 / それは残念。

3. システム評価

今回開発した学習システムの有効性を検証するため、音声認識の精度、システムの処理時間、本システムの活用による学習効果についての3種類の評価を行った。

3.1 音声認識の精度

被験者は、大学で2年間スペイン語を勉強して日常会話が可能で被験者 B (男性) とスペイン語のネイティブスピーカーの被験者 A (女性) の2名とする。実験方法として被験者が20種類の会話パターンを発話し、その発話したデータが正しくテキストに変換されているか評価する。

結果、被験者 A は20回中19回、被験者 B は20回中18回で認識に成功した。平均の認識率は92.5%であり、認識に失敗した原因として、発話した名詞が異なる単語とし

て認識されたことがある。そのため、単語の区切りを明確にして話すことができれば、100%に近い精度で認識可能であると考えられる(表3)。

表3 音声認識の精度

対象	会話数	音声認識の結果		
		認識成功	認識失敗	認識率
被験者 A	20	19	1	95%
被験者 B	20	18	2	90%
平均	20	18.5	1.5	92.5%

3.2 システムの処理時間

本実験では、CG キャラクタからの質問に対して学習者が返答した後に、CG キャラクタが会話文を返答するまでに掛かった時間を評価する。被験者は、被験者 A (女性) と被験者 B (男性) の 2 名とした。各被験者は 20 種類の会話パターンを発話し、2 名の返答時間の平均値を算出する。また、学習者の返答、CG キャラクタの質問、それに対する CG キャラクタの返答の単語数とその合計数をそれぞれ記述し、単語数に対して返答時間が変化するかを確認した。

結果、20 の会話パターンの合計単語数の平均が 11.40 語であるのに対して、平均応答処理時間は 3.39 秒であった。個別の結果を見ると、最大処理時間は特定の単語数 11 個に対して 4.01 秒、最小処理時間は特定の単語数 14 個に対して 2.78 秒であった。この結果から、処理時間は単語数による影響ではなく、クラウドサービスの処理時間の可能性が高いと考えられる。また、質問に対して返答してから 3~4 秒程度で返答内容が返ってくることから、実際の人物との対話に近い会話学習が行えると考えられる (表 4)。

表 4 システムの処理時間

No.	CG 質問 単語数	返答 単語数	CG 返答 単語数	合計 単語数	処理 時間
1	4	3	3	10	3.41 秒
2	3	3	3	9	3.75 秒
3	6	8	4	18	3.09 秒
4	2	3	3	8	3.24 秒
5	4	3	3	10	3.31 秒
6	3	3	2	8	3.08 秒
7	3	3	3	9	3.41 秒
8	4	3	4	11	3.49 秒
9	4	5	3	12	3.66 秒
10	6	4	4	14	3.77 秒
11	3	3	6	12	3.27 秒
12	4	3	4	11	3.61 秒
13	6	4	4	14	2.86 秒
14	6	4	4	14	2.78 秒
15	4	4	3	11	4.01 秒
16	4	6	5	15	3.42 秒
17	6	6	4	16	3.77 秒
18	2	2	3	7	3.47 秒
19	4	2	1	7	3.38 秒
20	7	3	2	12	3.12 秒
平均	4.25	3.75	3.40	11.40	3.39 秒

3.3 本システムの活用による学習効果

本システムの活用による学習に与える効果を評価する。被験者は、大学の 1 セメスタ分のスペイン語授業の履修を終え、2 セメスタ目を履修している学生である被験者 C (女性) と被験者 D (男性) の 2 名とした。実験方法としては、20 パターンの会話内容から 10 パターンをシステム上で体験し、その正否を判定するものとした。実験回数は計 2 回とし、1 回目は事前練習を行っていない状態で実施、2 回目は事前に体験した内容の質問と返答の文章を紙で配布し、1 日の予習期間を与え、1 回目の翌日に同様の質問内容で実施した。

1 回目の実験では、被験者 C は 0 問、被験者 D は 1 問の正解数であった。2 回目の実験では被験者 C は 5 問、被験

者 D は 7 問に正解数が増加した。また、被験者 C は、2 回目の回答時に、実際には 8 回正解の返答文を話していた。その差分は、クラウド側の音声認識の失敗による原因であり、発音やイントネーションが正しくなかったことが原因として考えられる。以上から本システムを活用することによって一定の学習効果が得られたことが考えられる。しかし、今回の実験では、予め回答の会話文を提示し勉強期間を設けたことや、質問の出題順を固定にしたため、自学自習の学習結果が結果に反映された可能性がある。今後は、本システムの利用や質問の出題順をランダムにするといった方法を用い、本システムの活用効果をさらに検証したいと考えている (表 5)。

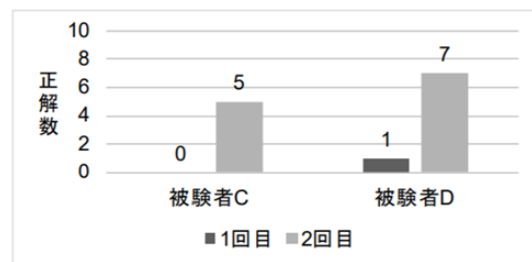


図 5 本システムの活用による学習効果

4. おわりに

本研究では、実際の人物との円滑なコミュニケーション能力を身に付けることを目的として、仮想空間におけるスペイン語の対話学習システムを提案し、その評価を行った。音声認識精度の実験では、スペイン語による認識精度は非常に高い結果を示すことができた。処理時間においては、CG キャラクタからの質問に返答してから平均で 3~4 秒程度で返答が返ってくることから実際の人物との会話に近い環境であることがいえる。また、本システムの活用による学習効果については、本システムを活用することで一定の学習効果の可能性を得ることができた。今後は、心拍センサやカメラなど IoT 機器と組み合わせや学習者の反応によって CG キャラクタの動作の変更や設問内容の変更、話す速度を変更するといった、さらに実際の人物との会話に近い環境を提供することを目標とする。

謝辞

本研究の遂行にあたり、文教大学スペイン語講師の寺田裕子氏にご協力頂いた。

参考文献

- [1] 文部科学省総合教育政策局国際教育課, “平成 29 年度高等学校における国際交流等の状況について”, (2019).
- [2] Joyz, “TerraTalk”, <https://www.terratalk.rocks/> (参照 2021/6/17) (2018)
- [3] SpeakBUDDY, “AI 英会話アプリスピークバディ | 音声認識機能搭載の英語学習でスピーキング、リスニング力が上達!”, <https://www.speakbuddy.me/> (参照 2021/6/17) (2021)
- [4] 秋本桃子, 阿部秀尚, 生田祐子, 森田武史, 山口高平, “教師業務ルール分析に基づく対話型ロボットを用いた発音練習の実装と評価”, 情報処理学会情報教育シンポジウム論文集 2018, (2018).
- [5] 鈴木直人, 廣井富, 千葉祐弥, 能勢隆, 伊藤彰則, “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”, 情報処理学会論文誌, Vol.56, No.11 (2015).