

令和 3 年 6 月 9 日現在

機関番号：30107

研究種目：若手研究(B)

研究期間：2017～2020

課題番号：17K18075

研究課題名（和文）Pedagogical Agentを利用したMOOCの構築

研究課題名（英文）Development of MOOC by using Pedagogical Agent

研究代表者

長谷川 大（Hasegawa, Dai）

北海学園大学・工学部・准教授

研究者番号：30633268

交付決定額（研究期間全体）：（直接経費） 3,300,000円

研究成果の概要（和文）：本研究は学習者が教材を作成するMOOC（LG-MOOC）を構想し、そのための要素技術として、Pedagogical Agentのジェスチャの自動生成手法とInstructional Design理論にもとづく教材作成のためのマークアップ言語（IDML）の開発を目的とした。研究期間内では主に前者の課題に取り組み、ジェスチャ自動生成のためにスピーチ音声とジェスチャをペアとしてデータセットを作成し、Deep Neural Networkを用いて音声からジェスチャを生成するモデルを構築した。生成されたジェスチャは概ね自然な動きとして知覚されたが、スピーチの意味内容との対応は不十分な結果であった。

研究成果の学術的意義や社会的意義

本研究で作成したデータセットは、日本語話者のジェスチャが3次元データとして収録されている。日本語話者によるデータセットは貴重であり、データセットおよびジェスチャ生成のためのプログラムはgithubで公開している。また本研究ではデータドリブンアプローチによってスピーチ音声から人間のジェスチャを生成する手法を提案した。生成されたジェスチャは動作として人間らしい自然さがあり、このアプローチの可能性を示唆するものであった。この技術が実現されれば、教育エージェントのみならず、ヒューマノイドロボットなど、Human-Agent/robot Interactionへの広い応用が期待できる。

研究成果の概要（英文）：In this research, The concept of learner-generated MOOC (LG-MOOC) is proposed. In LG-MOOC, a learner creates learning materials and share with other learners, and the materials are performed by a pedagogical agent. To realize the concept, there were two key technologies. The one is gesture generation method for a pedagogical agent, and the other is instructional Design markup language (IDML) to help learners create effective materials. During research term, I focused on development of gesture generation method. I created a dataset where gestures paired with speech audio. A gesture generation model, then, is constructed by Deep Neural Network with the dataset. Gestures generated by the model were perceived as natural movements. However, the meanings of speech and gesture were perceived as not very consistent.

研究分野：Human-Computer Interaction

キーワード：Pedagogical Agent Gesture Deep Neural Network

1. 研究開始当初の背景

自ら考え行動する自主性の高い人材育成の要請が世界的に高まるなか、「知識」の定義がラディカルに変革された新しい知識モデルが提案されている。従来の知識モデルでは、知識は命題と表象の集合(言葉・記号の集合)であり、また、教授者は知識を学習者へ伝達することができる。しかし、Downes が主張し、注目を集めている新しい知識モデルでは、知識は使用して経験することで意味内容が決定されていくものであり、言葉として伝達するだけでは知識獲得として不十分であると考えられている。

Connectivist MOOC (cMOOC) は、新しい知識モデルにもとづくオンライン学習環境の構想である(Coursera などの xMOOC とは区別される)。これは学習者らが主体的に知識獲得することを目指したオンライン学習環境であり、ネットワーク上に分散した教材をもとに、学習者がみずから試行錯誤的に情報を編纂・構築し、そのアウトプットがさらに新たな教材となるように構想されており、学習者同士の教え合いや学習者と教授者の循環構造を特徴とする。しかし、実際に機能する cMOOC の実現には次にあげる課題が存在する。第一に、教材作成に関する工夫が提案されておらず、学習者らが作成する教材の質が不揃いになってしまう。第二に、講義映像などの動画コンテンツのような没入感のある教材と比較して、学習効果を向上する工夫がなされていない。

2. 研究の目的

本研究は、学習者による質の高い教材作成を容易にする Instructional Design Markup Language (IDML)を提案し、教材の没入感を向上する Pedagogical Agent によるレクチャーの自動生成手法の開発を行うことで、新しい知識モデルにもとづくオンライン学習環境 Learner Generated Massive Open Online Course (LG-MOOC)の構築に挑戦する。LG-MOOC の構想には下記の要素技術が必要と考えられた。

1. Instructional Design Markup Language (IDML)の開発
 - (ア) 学習内容と教授シナリオを簡易に記述可能なマークアップランゲージ IDML の提案・設計
 - (イ) IDML を HTML タグのサブセットとし、オンラインシステムの実行環境である Web ブラウザで解釈・学習制御をおこなうための JavaScript ライブラリ IDML.js の開発
 - (ウ) IDML.js を利用した教材作成の有効性評価
2. Deep Learning による Pedagogical Agent のジェスチャ自動生成手法の開発
 - (ア) 発話の強調やリズムを表現する BEAT ジェスチャの自動生成手法の開発と評価
 - (イ) 抽象的な概念を表現する Metaphoric ジェスチャの自動生成手法の開発と評価
3. LG-MOOC の実装および 100 人規模の授業における教育効果の調査
 - (ア) IDML で記述された教授シナリオの実行・制御と Pedagogical Agent のジェスチャをともなうレクチャー生成システムを統合した LG-MOOC の開発
 - (イ) 大学生を対象とした 100 人規模の学習実験による LG-MOOC の有効性評価

このうち研究期間内には、主に Deep Learning による Pedagogical Agent のジェスチャ自動生成手法の開発を実施した。

3. 研究の方法

Deep Learning による Pedagogical Agent のジェスチャ自動生成手法の開発を目的として、スピーチ音声とジェスチャがペアになったデータセットの構築と、Deep Learning を用いたジェスチャ生成モデルの開発を実施した。具体的な方法を以下に述べる。

1. データセットの構築

- (ア) 当データセットは、発話とジェスチャの関連性を機械学習で学習する用途で使われることを想定した合計 1047 センテンス(約 196 分)の発話の音声データとそれに伴ったジェスチャのモーションデータのペアである。発話の音声データ mp3 で、ジェスチャのモーションデータは BioVision Hierarchy(BVH)形式で保存した。
- (イ) ジェスチャのモーションデータは OptiTrack 社のモーショントラッキングシステム Motive と 49 マーカのついたモーションキャプチャスーツおよび 850 × 850 センチメートル四方に配置された 8 台の OptiTrack Prime 17W カメラ使用し記録された。発話の音声データは、話者の動きを阻害しないようヘッドセットのマイクで記録された。
- (ウ) データの記録は、広く静かな一室で、事前に設定したトピックについてインタビュアーが話者に話を聞く準構造化インタビューによりデータを収集した。記録の際、

話者はスライドなどの提示物やメモを使用せずに話すこととした。これは、メモを確認する・提示物の方を向くなどの実際の話す内容に関係のない動作を防ぐためである。

2. ジェスチャ生成モデルの開発

(ア)本研究では、発話音声からジェスチャを生成するために、Bi-Directional LSTM Network を用いた 5 層のネットワークを使用した。第 1 層～第 3 層および第 5 層は Recurrent でない全結合層，第 4 層は Bi- Directional LSTM 層である。

(イ)入力となる発話音声データを一定の時間長で分割し、Mel-Frequency Cepstral Coefficients (MFCC) を計算することで、長さ T の時系列 MFCC 特徴ベクトルとする。また出力となるジェスチャは、長さ T の時系列 3 次元関節位置座標ベクトルとした。

4. 研究成果

本研究により得られ成果を以下に述べる。

1. データセットの詳細

1047 発話のうち 240 発話について、発話内容の書き起こしを行い、ジェスチャフェーズおよびジェスチャ種類のアノテーションを行なった結果、本研究で得られたデータセットの性質について図 1 に示す知見が得られた。ジェスチャフェーズでは Stroke の時間が最も多く、ジェスチャをしていない Rest の時間は 15%あることがわかった。また、ジェスチャの種類では、Metaphoric ジェスチャが 62%と大半を占めており、期待通りのデータとなっていることが読み取れる。

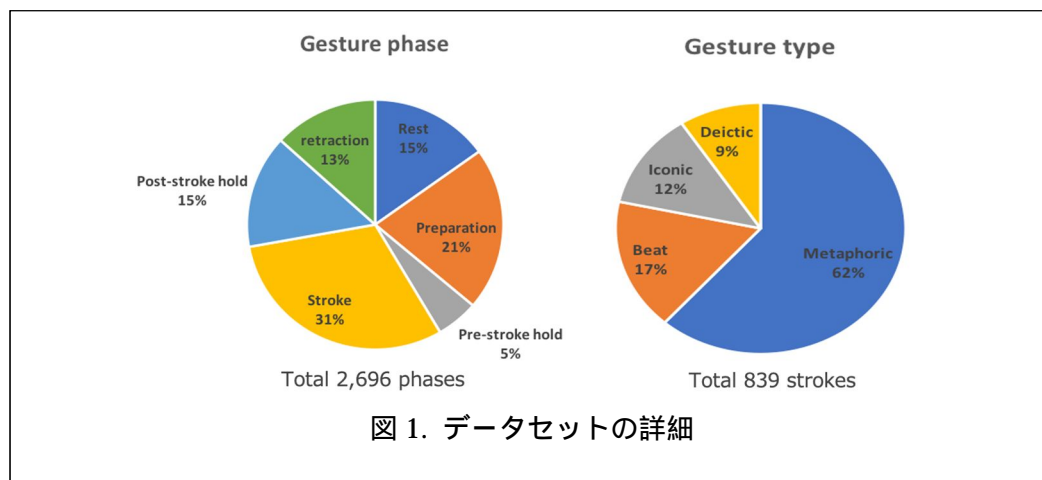


図 1. データセットの詳細

2. 自動生成されたジェスチャの評価

本研究で開発したモデルによって自動生成されたジェスチャを図 2 に示す。また生成されたジェスチャについて、21 歳から 30 歳までの男女 30 名による評価実験を行った。実験ではジェスチャの自然さ、発話との時間的整合性、発話内容との意味的整合性について、評価をしてもらい、どのように知覚されるかを検証した。

評価実験の結果、生成されたジェスチャは自然さの点で元のジェスチャに遜色がないことが明らかになった。一方で、時間的整合性や意味的整合性など発話との関連性については元のジェスチャに対して、違和感が残ることも明らかになった。

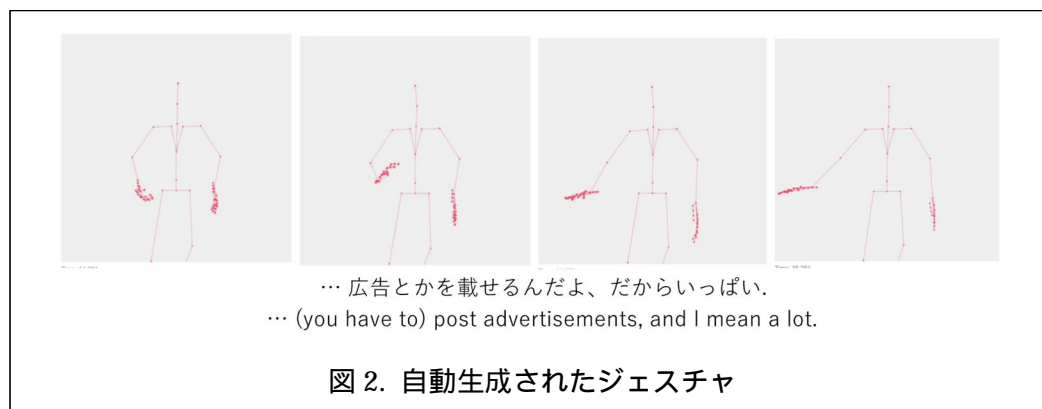


図 2. 自動生成されたジェスチャ

5. 主な発表論文等

〔雑誌論文〕 計5件（うち査読付論文 5件/うち国際共著 1件/うちオープンアクセス 0件）

1. 著者名 Kaneko Naoshi, Takeuchi Kenta, Hasegawa Dai, Shirakawa Shinichi, Sakuta Hiroshi, Sumi Kazuhiko	4. 巻 34
2. 論文標題 Speech-to-Gesture Generation Using Bi-Directional LSTM Network	5. 発行年 2019年
3. 雑誌名 Transactions of the Japanese Society for Artificial Intelligence	6. 最初と最後の頁 C~J41_1-12
掲載論文のDOI（デジタルオブジェクト識別子） 10.1527/tjsai.C-J41	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 長谷川大, 盛川浩志, 佐久田博司, 中山栄純, 安彦智史	4. 巻 20
2. 論文標題 アバタ媒介型見守りシステムの高齢者における受容性について	5. 発行年 2018年
3. 雑誌名 ヒューマンインタフェース学会論文誌	6. 最初と最後の頁 163-172
掲載論文のDOI（デジタルオブジェクト識別子） 10.11184/his.20.2_163	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 長谷川大, 白川真一, 佐久田博司	4. 巻 4
2. 論文標題 Pedagogical Agent の導管メタファ・ジェスチャが学習者の理解に与える効果	5. 発行年 2018年
3. 雑誌名 情報処理学会論文誌「教育とコンピュータ」	6. 最初と最後の頁 83-92
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 Kenta Takeuchi, Dai Hasegawa, Hiroshi Sakuta	4. 巻 713
2. 論文標題 Creating a Speech-Gesture Dataset for Speech-Based Automatic Gesture Generation	5. 発行年 2017年
3. 雑誌名 Communications in Computer and Information Science (CCIS)	6. 最初と最後の頁 353-357
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Kucherenko Taras、Hasegawa Dai、Kaneko Naoshi、Henter Gustav Eje、Kjellstr?m Hedvig	4. 巻 未定
2. 論文標題 Moving Fast and Slow: Analysis of Representations and Post-Processing in Speech-Driven Automatic Gesture Generation	5. 発行年 2021年
3. 雑誌名 International Journal of Human?Computer Interaction	6. 最初と最後の頁 1~17
掲載論文のDOI (デジタルオブジェクト識別子) 10.1080/10447318.2021.1883883	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する

〔学会発表〕 計9件 (うち招待講演 0件 / うち国際学会 4件)

1. 発表者名 Taras Kucherenko , Dai Hasegawa , Gustav Eje Henter , Naoshi Kaneko , Hedvig Kjellstrom
2. 発表標題 Analyzing Input and Output Representations for Speech-Driven Gesture Generation
3. 学会等名 International Conference on Intelligent Virtual Agents (IVA2019) (国際学会)
4. 発表年 2019年

1. 発表者名 Taras Kucherenko , Dai Hasegawa , Naoshi Kaneko , Gustav Eje Henter , Hedvig Kjellstrom
2. 発表標題 On the Importance of Representations for Speech-Driven Gesture Generation
3. 学会等名 AAMAS '19: THE 18TH INTERNATIONAL CONFERENCE ON AUTONOMOUS AGENTS AND MULTIAGENT SYSTEMS (国際学会)
4. 発表年 2019年

1. 発表者名 Dai Hasegawa, Naoshi Kaneko, Shinichi Shirakawa, Hiroshi Sakuta, Kazuhiko Sum
2. 発表標題 Evaluation of Speech-to-Gesture Generation Using Bi-Directional LSTM Network
3. 学会等名 18th ACM International Conference on Intelligent Virtual Agents (IVA2018) (国際学会)
4. 発表年 2018年

1. 発表者名 小田俊平, 高野保真, 長谷川大, 佐久田博司, 尾花謙伍
2. 発表標題 アバタ媒介型見守りシステムにおけるモーションキャプチャ情報を用いた機械学習の利用
3. 学会等名 情報処理学会研究報告 (HCI)
4. 発表年 2019年

1. 発表者名 日和 航大, 荒木 健治, 長谷川 大, 芳尾 哲
2. 発表標題 講義代行ロボットにおける双方向LSTMを用いたジェスチャ自動生成システムの性能評価
3. 学会等名 2018年度人工知能学会全国大会
4. 発表年 2018年

1. 発表者名 Kenta Takeuchi, Dai Hasegawa, Shirakawa Shinichi, Naoshi Kaneko, Hiroshi Sakuta, Kazuhiko Sumi
2. 発表標題 Speech-to-Gesture Generation: A Challenge in Deep Learning Approach with Bi-Directional LSTM
3. 学会等名 5th International Conference on Human Agent Interaction (HAI2017) (国際学会)
4. 発表年 2017年

1. 発表者名 日和航大, 荒木健治, 長谷川大
2. 発表標題 ヒューマノイドロボットを用いた講義代行システムにおけるジェスチャの考察
3. 学会等名 平成29年度電気・情報関係学会北海道支部連合大会予稿集
4. 発表年 2017年

1. 発表者名 浅川 栄一, 白川 真一, 長谷川 大
2. 発表標題 ニューラルネットワークを用いた発話テキストに対応するジェスチャの自動生成
3. 学会等名 計測自動制御学会 システム・情報部門 学術講演会2017 (SSI2017)
4. 発表年 2017年

1. 発表者名 竹内健太, 長谷川大, 白川真一, 金子直史, 佐久田博司, 鷲見和彦
2. 発表標題 Bi-directional LSTMを用いた発話に伴うジェスチャの自動生成手法の検討
3. 学会等名 HAIシンポジウム2017
4. 発表年 2017年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関