

Submission ID #: 69254

Scriptwriter Name: Sulakshana Karkala

Project Page Link: <https://review.jove.com/account/file-uploader?src=21122518>

Title: Virtual Agent for Real-Time Motivational Interviewing by Integrating Adaptive Nonverbal Behavior and Language Models

Authors and Affiliations:

Lucie Galland¹, Nezih Younsi¹, Carla Baudonne², Laurence Chaby³, Agnes Helme-Guizon⁴, Florian Pecune⁵, Catherine Pelachaud⁶

¹ISIR - Sorbonne University

²University of Bordeaux

³Vision Action Cognition, Université Paris Cité

⁴University of Grenoble Alpes

⁵CNRS - SANPSY, University of Bordeaux

⁶CNRS - ISIR, Sorbonne University

Corresponding Authors:

Lucie Galland (lucie.galland@isir.upmc.fr)

Email Addresses for All Authors:

Lucie Galland (lucie.galland@isir.upmc.fr)

Nezih Younsi (younsi@isir.upmc.fr)

Carla Baudonne (carlabdne@hotmail.com)

Laurence Chaby (laurence.chaby@parisdescartes.fr)

Agnes Helme-Guizon (agnes.helme-guizon@univ-grenoble-alpes.fr)

Florian Pecune (florian.pecune@u-bordeaux.fr)

Catherine Pelachaud (catherine.pelachaud@isir.upmc.fr)

Author Questionnaire

1. **Microscopy:** Does your protocol require the use of a dissecting or stereomicroscope for performing a complex dissection, microinjection technique, or something similar? **NO**
2. **Software:** Does the part of your protocol being filmed include step-by-step descriptions of software usage? **Yes , all done**
3. **Filming location:** Will the filming need to take place in multiple locations?
4. **Testimonials (optional):** Would you be open to filming two short testimonial statements **live during your JoVE shoot**? These will **not appear in your JoVE video** but may be used in JoVE's promotional materials. **No**

Current Protocol Length

Number of Steps: 30

Number of Shots: 56

Introduction

INTRODUCTION:

~~What is the scope of your research? What questions are you trying to answer?~~

- 1.1. **Lucie Galland:** Our research investigates adaptive multimodal interaction, focusing on how to adapt a virtual agent's verbal and nonverbal behaviours, such as facial expressions, during interaction.
 - 1.1.1. INTERVIEW: Named Talent says the statement above in an interview-style shot, looking slightly off-camera.

~~What are the most recent developments in your field of research?~~

- 1.2. **Lucie Galland:** Recent advances in machine learning, particularly large language models, enable more natural and flexible multimodal human–computer interaction.
 - 1.2.1. INTERVIEW: Named Talent says the statement above in an interview-style shot, looking slightly off-camera.

CONCLUSION:

~~What significant findings have you established in your field?~~

- 1.3. **Lucie Galland:** Our findings show that adapting verbal and nonverbal facial expressions significantly enhances interactive system effectiveness and improves perception of socially interactive agents.
 - 1.3.1. INTERVIEW: Named Talent says the statement above in an interview-style shot, looking slightly off-camera.

~~How will your findings advance research in your field?~~

- 1.4. **Lucie Galland:** Our platform enables the evaluation of other adaptive components such as gesture generation and of other dialogue topics.
 - 1.4.1. INTERVIEW: Named Talent says the statement above in an interview-style shot, looking slightly off-camera.

~~What questions will future research focus on?~~

1.5. **Lucie Galland** : Future research will focus on long-term interactions and continuous adaptation across multiple conversational sessions.

1.5.1. INTERVIEW: Named Talent says the statement above in an interview-style shot, looking slightly off-camera.

Ethics Title Card

This research has been approved by the Institutional Review Board (IRB) at the INSEAD institute

Protocol

2. System Setup and Software Configuration for the Virtual Agent Platform

Demonstrator: Nezhir Younsi

- 2.1. To begin, prepare a Windows computer for the experiment [1].
 - 2.1.1. WIDE: Talent positioned at a workstation with a Windows computer powered on.
Videographer's Note: Shot as MED+ CU
- 2.2. Install Java SE (S-E) Development Kit 8 from the official distribution source [1]. Then install the Visual C Plus Plus Redistributable for Visual Studio 2013 [2].
 - 2.2.1. SCREEN: [2.2.1.mkv](#). 00:00-00:12
 - 2.2.2. SCREEN: [2.2.2.mkv](#) 00:00-00:10
- 2.3. Request a Cereproc license through the online application portal [1]. Install OpenFace from the provided repository [2]. Obtain a webcam for video capture [3].
 - 2.3.1. SCREEN: [2.3.1.mkv](#). 00:12-end
 - 2.3.2. SCREEN: [2.3.2.mov](#). 00:06-00:13, 00:21-00:31
 - 2.3.3. Talent placing a webcam next to the computer.
- 2.4. Next, download the Greta software release from the specified repository [1]. Compile the software using NetBeans following the provided instructions [2].
 - 2.4.1. SCREEN: [2.4.1 and 2.4.2.mov](#). 00:00-00:11
 - 2.4.2. SCREEN: [2.4.1 and 2.4.2.mov](#). 00:25-00:32, 00:50-00:53
- 2.5. Obtain a Mistral application programming interface key from the online console [1].
 - 2.5.1. SCREEN: [2.5.1.mov](#). 00:03-00:27
- 2.6. Create the file then paste the application programming interface key into the created file [1].
 - 2.6.1. SCREEN: [2.6.1.mp4](#). 00:00-end
TXT: File: greta\bin\Common\Data\LLM\MICounselorRL\api_key.txt
- 2.7. Now, obtain a DeepGram application programming interface key from the online console [1]. Create the file path and paste the application programming interface key into the created file [2].

- 2.7.1. SCREEN: [2.7.1.mp4](#). 00:00-end
- 2.7.2. SCREEN: [2.7.2.mp4](#). 00:00-end
TXT:File: greta\bin\Common\Data\DeepASR\DeepGram\api_key.txt
- 2.8. For MODIFF (*moh-diff*) model import, first download the MODIFF model weights from the provided source [1]. Place the downloaded file into MODIFF data folder [2].
- 2.8.1. SCREEN: [2.8.1 and 2.8.2.mp4](#). 00:00-00:10
- 2.8.2. SCREEN: [2.8.1 and 2.8.2.mp4](#). 00:20-00:40
- ~~2.9. Download and uncompress the dlib (*D-lib*) face detection model files [1]. Place the extracted files into the dlib model directory [2].~~
- ~~2.9.1. SCREEN: Talent extracting the dlib model files on the computer.~~
- ~~2.9.2. SCREEN: Talent placing the extracted files into the specified directory.~~
NOTE: Deleted by authors
- ~~2.10. Then download all VAP (*V-A-P*) models from the provided folder [1]. Place the models into the Turn Management models folder [2].~~
- ~~2.10.1. SCREEN: Talent downloading multiple VAP model files.~~
- ~~2.10.2. SCREEN: Talent copying the files into the TurnManagement models folder.~~
NOTE: Deleted by authors
- 2.11. Next launch **Modular.jar** (*modular-jar*) [1]. Click on **File, Open, Greta, Advanced**, and choose **20250128 Greta - Expé Lucie_full.xml** (*Twenty-twenty-five-Zero-One-Two-Eight-Greta -Expé-Lucie-full-dot-x-m-l*) [2].
- 2.11.1. SCREEN: [2.11.1 to 2.12.1.mp4](#). 00:00-00:10
- 2.11.2. SCREEN: [2.11.1 to 2.12.1.mp4](#). 00:11-00:18
- 2.12. Ensure that the configuration displayed matches the expected setup [1].
- 2.12.1. SCREEN:[2.11.1 to 2.12.1.mp4](#) 00:23-00:40
- 2.13. Now, start the **OpenFace Offline ZeroMQ.exe** (*Open-Face-Offline-Zero-MQ-program*) [1-TXT]. Confirm that the software communicates with the platform through the specified port [2].
- 2.13.1. SCREEN:[2.13.1 and 2.16.2 no cobaye.mp4](#). 00:00-00:10
TXT: Confirm that software communicates with the platform through thespecified port
- ~~2.13.2. SCREEN: Show OpenFace indicating active ZeroMQ communication.~~
- 2.14. In the **Record** tab, uncheck all options except **Broadcast with ZeroMQ** [1].

- 2.14.1. SCREEN: [2.13.1 and 2.16.2 with cobaye.mp4](#) 00:00-00:16
- 2.15. In the **File** tab, click **Open Webcam [1]**. Authorize live feature extraction using the available webcam **[2]**.
- 2.15.1. SCREEN: [2.13.1 and 2.16.2 with cobaye.mp4](#) 00:16-00:21
- 2.15.2. SCREEN: [2.13.1 and 2.16.2 with cobaye.mp4](#) 00:31-00:37
- 2.16. Select the webcam to use. Wait until the webcam is loaded and live feature extraction starts automatically **[1]**.
- 2.16.1. SCREEN: [2.13.1 and 2.16.2 no cobaye.mp4](#). 00:11-end
- 2.17. Next, in **OpenFace2 (Open-Face-Two) Output Stream Reader**, click **Connect [1]**. Wait for the available feature list to populate **[2]**.
- 2.17.1. SCREEN: [2.17.1 to 2.20.2.mp4](#). 00:00-00:03
- 2.17.2. SCREEN: [2.17.1 to 2.20.2.mp4](#). 00:04-00:08
- 2.18. Click on **Select All** and then **Set** to validate the selected features **[1]**. Under Modiff, click **Launch** and wait for the connection confirmation message **[2]**.
- 2.18.1. SCREEN: [2.17.1 to 2.20.2.mp4](#). 00:09-00:13
- 2.18.2. SCREEN: [2.17.1 to 2.20.2.mp4](#) 00:14-00:33
- 2.19. Then press **Connect** to activate the connection between Modiff and the data receiver **[1]**. Under **Filter**, click **Perform** to allow execution of generated data **[2]**.
- 2.19.1. SCREEN: [2.17.1 to 2.20.2.mp4](#). 00:34-00:37
- 2.19.2. SCREEN: [2.17.1 to 2.20.2.mp4](#). 00:38-00:41
- 2.20. In the MODIFF window, click **Enable [1]**. Wait 90 seconds for the model to stabilize **[2]**.
- 2.20.1. SCREEN: [2.17.1 to 2.20.2.mp4](#). 00:47-00:56
- 2.20.2. SCREEN: [2.17.1 to 2.20.2.mp4](#). 00:56-end
- 2.21. To launch ASR (**A-S-R**), in the DeepGram window, press **Enable [1]**.
- 2.21.1. SCREEN: [2.21.1 to 2.23.2.mp4](#). 00:00-00:04
- 2.22. Then select the condition **RL (R-L)** to use **DREAM (dream)** or **Baseline** to use a plain large language model **[1]**.
- 2.22.1. SCREEN: [2.21.1 to 2.23.2.mp4](#). 00:05-00:23
- 2.23. In the **MI Counsellor RL** window, click **Enable [1]**. Wait 30 seconds for the model to start **[2]**.
- 2.23.1. SCREEN: [2.21.1 to 2.23.2.mp4](#). 00:24-00:32
- 2.23.2. SCREEN: [2.21.1 to 2.23.2.mp4](#). 00:32-end

3. Participant Setup, Profiling, and Dialogue Execution

- 3.1. Place a large monitor on a table with a chair positioned in front [1]. Position a webcam on top of the monitor facing the chair [2]. Place a directional microphone on the table in front of the chair [3].
 - 3.1.1. Talent arranging the monitor and chair in the room.
 - 3.1.2. Talent positioning the webcam on top of the monitor.
 - 3.1.3. Talent placing the microphone on the table.
- 3.2. Have the participant complete the Decision Balance Scale or DBS (D-B-S) questionnaire using a 5 step Likert scale [1]. Then ask the participant to speak into the microphone [2-TXT].
 - 3.2.1. Talent handing the questionnaire to the participant.
 - 3.2.2. Talent instructing the participant to speak into the microphone. **TXT: Ensure webcam captures the participant's head and upper body**
- 3.3. To begin, identify the participant profile, evaluate the DBS score [1]. Classify the participant as Resistant, Hesitant, or Open to Change based on the score [2].
 - 3.3.1. Talent reviewing DBS scores on a computer or form.
 - 3.3.2. SCREEN: [3.3.2 to 3.5.2.mp4](#). 00:07-10
- 3.4. In the **MI Counsellor RL** window, choose the discussion theme chosen by the participant [1]. ~~Set the participant profile accordingly [2].~~ Click **Start** to allow the agent to initiate the discussion with the participant [3].
 - 3.4.1. SCREEN: [3.3.2 to 3.5.2.mp4](#). 00:00-00:06
 - 3.4.2. SCREEN: [3.3.2 to 3.5.2.mp4](#). 00:35-00:40
- 3.5. Monitor the **Answer** section of the MI Counsellor window for harmful output [1]. In the **DeepGram** window, click **Listen** after the agent finishes its turn [2].
 - 3.5.1. SCREEN: [3.3.2 to 3.5.2.mp4](#). 00:41-00:50
 - 3.5.2. SCREEN: [3.3.2 to 3.5.2.mp4](#). 00:55-end
- 3.6. If speech is not recognized, enter the participant's speech into the **Request** field and click **Send** [1]. Wait for the participant's answer [2-TXT].
 - 3.6.1. SCREEN: [3.6.1.mp4](#). 00:02-00:15
 - 3.6.2. Shot of the Participant responding verbally during the session. **TXT: Repeat monitoring, listening, and response until the dialogue ends automatically**
Videographer's Note: Use take 2
- 3.7. Ask the participant to complete the post intervention questionnaires [1]. Then debrief the participant using the prepared speech [2].

3.7.1. WIDE: Talent providing post intervention questionnaires to the participant.

3.7.2. Talent debriefing the participant at the end of the session.

Videographer's Note: Shot as MED+ CU

Results

4. Results

- 4.1. Users interacting with the agent with adaptive facial expressions expressed more positive sentimentality, in their disclosure, than users in the other behaviour conditions [1].
 - 4.1.1. LAB MEDIA: Figure 5. *Video editor: Please highlight the blue curve corresponding to MODIFF-8*
- 4.2. No significant differences were observed across the three expression conditions in subjective questionnaire scores assessing attitude, social rapport, and motivational interviewing quality [1].
 - 4.2.1. LAB MEDIA: Figure 6. *Video editor: Please highlight all 3 conditions sequentially*
- 4.3. The mismatched facial expression condition (i.e. when interpersonal contingency is no longer respected) was generally rated lower than both the inexpressive (i.e. when the agent shows no expression) and adaptive (i.e., the agent's expression is driven by our model) conditions across subjective measures [1].
 - 4.3.1. LAB MEDIA: Figure 6. *Video editor: Highlight the light orange bars corresponding to the mismatched condition*
- 4.4. Perceived attitude had a strong direct effect on perceived motivational interviewing quality [1]. The relationship between perceived attitude and motivational interviewing quality was partially mediated by social rapport, accounting for 42 percent of the effect [2].
 - 4.4.1. LAB MEDIA: Figure 7. *Video editor: Highlight the arrow linking attitude perception to MI quality.*
 - 4.4.2. LAB MEDIA: Figure 7. *Video editor: Highlight the path passing through social rapport with the 42 percent label.*
- 4.5. Participants interacting with the adaptive DREAM dialogue manager reported significantly higher rapport than those interacting with the plain large language model baseline [1].
 - 4.5.1. LAB MEDIA: Figure 8. *Video editor: Highlight the red boxplot*
- 4.6. The DREAM dialogue manager showed better adaptation to different participant profiles than the baseline model in motivation scores measured by the Decisional Balance Scale [1].

4.6.1. LAB MEDIA: Figure 9. *Video editor: Highlight the red boxplot*

4.7. Both the baseline and DREAM conditions produced Client Evaluation of Motivational Interviewing scores above the therapeutic threshold [1].

4.7.1. LAB MEDIA: Figure 10. *Video editor: Highlight the blue and red curves sequentially*

Pronunciation Guide:

Multimodal

Pronunciation link: <https://www.merriam-webster.com/dictionary/multimodal>

IPA: /ˌmʌltiˈmoʊdəl/, /ˌmʌltɑː-/

Phonetic Spelling: mul·tee·moh·duhl

Nonverbal

Pronunciation link: <https://www.merriam-webster.com/dictionary/nonverbal>

IPA: /ˌnɑːnˈvɜːbəl/

Phonetic Spelling: non·vur·buhl

Motivational

Pronunciation link: <https://www.merriam-webster.com/dictionary/motivational>

IPA: /ˌmoʊtəˈveɪʃənəl/

Phonetic Spelling: moh·tuh·VAY·shuh·nuhl

Interviewing

Pronunciation link: <https://www.merriam-webster.com/dictionary/interviewing>

IPA: /ˈɪntərˌvjuːɪŋ/

Phonetic Spelling: in·ter·vyoo·ing

Adaptive

Pronunciation link: <https://www.merriam-webster.com/dictionary/adaptive>

IPA: /əˈdæptɪv/

Phonetic Spelling: uh·DAP·tiv

Integration

Pronunciation link: <https://www.merriam-webster.com/dictionary/integration>

IPA: /ˌɪntəˈɡreɪʃən/

Phonetic Spelling: in·tuh·GRAY·shuhn

Configuration

Pronunciation link: <https://www.merriam-webster.com/dictionary/configuration>

IPA: /kənˌfɪɡjəˈreɪʃən/

Phonetic Spelling: kuhn·fig·yuh·RAY·shuhn

Redistributable

Pronunciation link: <https://www.merriam-webster.com/dictionary/redistributable>

IPA: /ˌrɪːdɪˈstrɪbjətəbəl/

Phonetic Spelling: ree·di·STRIB·yuh·tuh·buhl

Repository

Pronunciation link: <https://www.merriam-webster.com/dictionary/repository>

IPA: /rɪˈpɑːzəˌtɔːri/

Phonetic Spelling: ri·POZ·uh·tor·ee

Compile

Pronunciation link: <https://www.merriam-webster.com/dictionary/compile>

IPA: /kəmˈpaɪl/

Phonetic Spelling: kuhm·PYLE

Interface

Pronunciation link: <https://www.merriam-webster.com/dictionary/interface>

IPA: /ˈɪntərˌfeɪs/

Phonetic Spelling: in·ter·fays

ZeroMQ

Pronunciation link: <https://www.howtopronounce.com/zeromq>

IPA: /'zɪərɒʊ ,em kju:/

Phonetic Spelling: ZEER·oh·em·kyoo

Stabilize

Pronunciation link: <https://www.merriam-webster.com/dictionary/stabilize>

IPA: /'steɪbə ,læz/

Phonetic Spelling: STAY·buh·lyze

Questionnaire

Pronunciation link: <https://www.merriam-webster.com/dictionary/questionnaire>

IPA: / ,kwɛstʃə 'ner/

Phonetic Spelling: kwes·chuh·NAIR

Likert

Pronunciation link: <https://www.merriam-webster.com/dictionary/Likert%20scale>

IPA: /'lɪkərt/

Phonetic Spelling: LIK·urt

Rapport

Pronunciation link: <https://www.merriam-webster.com/dictionary/rapport>

IPA: /ræ'pɔ:r/

Phonetic Spelling: ra·POR

Contingency

Pronunciation link: <https://www.merriam-webster.com/dictionary/contingency>

IPA: /kən'tɪndʒənsi/

Phonetic Spelling: kuhn·TIN·juhn·see

Interpersonal

Pronunciation link: <https://www.merriam-webster.com/dictionary/interpersonal>

IPA: / ,ɪntər'pɜ:sənəl/

Phonetic Spelling: in·ter·PUR·suh·nuhl

Mediated

Pronunciation link: <https://www.merriam-webster.com/dictionary/mediate>

IPA: /'mi:diɛtɪd/

Phonetic Spelling: MEE·dee·ay·tuhd

Therapeutic

Pronunciation link: <https://www.merriam-webster.com/dictionary/therapeutic>

IPA: / ,θerə'pjʊ:tɪk/

Phonetic Spelling: ther·uh·PYOO·tik