project SENSE – Multimodal Simulation with Full-Body Real-Time Verbal and Nonverbal Interactions

Hossein Miri[1], Jan Kolkmeier[1](✉), Paul J. Taylor[1,2], Ronald Poppe[3], and Dirk Heylen[1]

[1] University of Twente, Enschede, Netherlands
{h.miri,j.kolkmeier,d.k.j.heylen}@utwente.nl, p.j.taylor@lancaster.ac.uk
[2] Lancaster University, Lancaster, UK
[3] Utrecht University, Utrecht, Netherlands
r.w.poppe@uu.nl

Abstract. This paper presents a multimodal simulation system, project-SENSE, that combines virtual reality and full-body motion capture technologies with real-time verbal and nonverbal communication. We introduce the technical setup and employed hardware and software of a first prototype. We discuss the capabilities of the system for the investigation of *cooperation paradoxes* and the effects of *direct nonverbal mimicry*. We argue that this prototype lays the technological basis for further research in interpersonal and social skills, as well as the social and emotional consequences of nonverbal mimicry in sustained interactions.

Keywords: Virtual reality · Motion capture · Interactive conversational agents · Mimicry · Social skills · Interpersonal sensemaking

1 Introduction

Interpersonal sensemaking is the cognitive process of understanding, predicting, and responding to the actions and inferred beliefs of others [1]. A critical, but under-explored, aspect of interpersonal sensemaking is *cooperation paradoxes*. A cooperation paradox occurs when the cooperative/competitive orientation implied by a person's *verbal* behavior does not correspond with the cooperative/competitive orientation of their *nonverbal* behavior [2]. This occurs, for example, when a person is verbally cooperative, but fails to *mimic* their partner's nonverbal behavior (i.e. they fail to imitate poses, gestures, and facial expressions) [3]. Understanding these paradoxes is important because they are often implicated as being at the heart of communication misunderstandings and conflicts in contexts such as law enforcement interviews [4]. To understand how people respond to paradoxes in *mimicry*, it is necessary to be able to measure or simulate the matching/mismatching of verbal and nonverbal behavior; henceforth orientation-matching between interactants. See Table 1.

R. Poppe et al. (Eds.): INTETAIN 2016, LNICST 178, pp. 279–284, 2017.
DOI: 10.1007/978-3-319-49616-0_28

Table 1. Characterizing the Type of Behavior

		Verbal behavior	
		Cooperative	Competitive
Nonverbal behavior	Cooperative	*Cooperation*	*Paradox*
	Competitive	*Paradox*	*Conflict*

When examining communicative channels in *simulation* environments through virtual characters, orientation-matching requires a degree of realism of the characters with respect to their appearance, nonverbal movements and behaviors, and speech. Currently, there is limited technology available for virtual reality, motion capture, and avatar interfaces, capable of reproducing replicas that are animated by users' movements and include real-time verbal communication. We adopted several technologies and developed an interactive training simulation in which verbal and nonverbal modes can be manipulated. We developed a proof-of-concept prototype to facilitate *prolonged interactions* with a realistic-looking embodied conversational character in an immersive virtual environment, through the use of a state-of-the-art MoCap device. The virtual character is capable of real-time verbal and nonverbal behavior generation, as well as mirror-mimicking participants' full-body movements at an *adjustable delay*.

In a dyadic setting, a computer representation of an interviewee (e.g. a suspect) interacts with the user (e.g. a trainee officer) using mannerisms that are responsive to his/her actions, while also playing out a training scenario. By allowing the manipulation of, and capture of, verbal and nonverbal behaviors, the resulting interactive training environment simulates each of the cooperation paradoxes in a single platform and allows for the social skills training for law enforcement professionals.

2 Virtual Reality (VR) Technology

VR offers an appealing alternative for managing human-computer interactions. Part of the appeal is attributed to the fact that users can interact with virtual characters and artifacts in the environment using their *natural senses*. This increases the sensation of *immersion* or feeling of *embodiment* in the synthetic environment through the use of powerful graphics rendering engines, as well as interfaces with Head Mounted Display (HMD) devices. We selected the Oculus Rift [5] for project-SENSE as it is a high-powered light-weight high-resolution HMD for VR environments, designed to provide an immersive experience. With its wide field of view and low latency 360° head tracking, this HMD allows to virtually step inside any environment and interact with it.

3 Motion Capture (MoCap) Technology

MoCap technologies are being exploited to identify and study *nonverbal correlates of human interactions*, while also allowing users to view and review their own movements. When coupled with MoCap technology, interactive and immersive VR systems can provide remarkable realism and accuracy in simulating *human movements and behaviors*. A full-body MoCap system allows for a refined measurement of nonverbal behavior and enables the investigation of movement in all human limbs and joints. Marker-less and wireless MoCap technologies, in particular, have gained popularity due to their ease of use and precision. We selected the Xsens MVN Awinda system, along with the MVN Studio BIOMECH software application [6], for project-SENSE. It consists of wearable straps with 17 motion trackers, aligned with anatomical landmarks of the human body by means of a 5-second calibration procedure. It records the position, orientation, velocity, acceleration, angular velocity and angular acceleration of 23 segments of the body, as well as magnetic field and body's center of mass. The sensors are easily secured to various limbs without hindering natural movement. The system does not require external cameras, emitters, or markers, and can be used both outdoors and indoors. The use of the Xsens MVN Awinda system provides an efficient measure of nonverbal behavior, in particular mimicry, that is less susceptible to the subjectivity associated with observational coding of behavior [7].

4 Real-Time Interaction Technology

The Virtual Human Toolkit (VHTK) is a collection of modules, tools, and libraries, to allow for the creation of virtual conversational characters [8]. We fused the VHTK with Oculus Rift and Xsens MVN Awinda system, in order to augment the nonverbal method of examining behavior with real-time verbal capabilities. The power of this toolkit is in the combination of a wide range of integrated capabilities. Specifically, we took advantage of its NPCEditor and SmartBody. The NPCEditor [9] controls the spoken behavior of the characters, as well as the structure and logic of their interactions. It contains a list of questions that the user can ask, along with answers that the character can give, as well as the links between them. It uses a statistical text classifier to determine the best character responses to user input. The interaction is done through *text* (i.e. typing) or *speech* (i.e. microphone). SmartBody [10] is a character animation platform and library that provides synchronized locomotion, steering, object manipulation, lip syncing, eye gazing, and nonverbal behaviors in real-time. It is, in effect, a Behavioral Markup Language (BML) realization engine that transforms BML behavior descriptions into real-time animations. The use of VHTK serves to provide a unified framework for audio-visual sensing, nonverbal behavior understanding, speech recognition, natural language processing, dialogue management, nonverbal behavior generation and realization, and text-to-speech [11].

5 Development Platform

We used the Unity3D [12] as a common development platform for the hardware tools and software solutions. Unity3D is a modern visualization, rendering, and game engine with an accessible editor that also includes functionality to interface recent VR devices; e.g. the Oculus Rift. We also integrated into Unity3D the MVN Studio (application software for Xsens products) through a specialized plug-in for accessing live motion

Fig. 1. MVN Studio character (*left*) and VHTK avatar, Brad (*right*)

data from the Xsens MVN Awinda system. Finally, the VHTK was integrated into the project-SENSE prototype. See Fig. 1 for a schematic, showing the MVN Studio and VHTK avatars.

6 Prototype Operation

The application starts off with a menu that features selectable toggles for choosing one of the four training scenarios: *verbal cooperation with mimicry, verbal cooperation without mimicry, verbal competition with mimicry, and verbal competition without mimicry*. The experimenter makes the selection, thus determining the mode of communication prior to the commencement of the session. The application then proceeds to the training simulation. Participants are instructed to engage with the avatar that, in mimicry scenarios, replicates their behaviors similar to a VR mirror, with an adjustable delay. This mirror-mimicry is achieved through 'piling up' the animation/motion data into a queue and then streaming it into the character (Unity, SmartBody, etc.) after a certain delay. The avatar also utters scenario-specific dialogues, with associated natural-looking facial expressions and body gestures. In the other two scenarios, the avatar stands idly, ready to interact. Previously-designed scenario contents determine what responses by the VHTK digital character ensue, as well as what behaviors and facial expressions are generated.

7 Discussion

Nonverbal behavior carries significant social meaning in spoken communication [13] and its analysis contributes significantly to our understanding of how *human interaction* works [14]. A central construct to human interaction is behavior accommodation or *mimicry*, which we defined here as the degree to which two interactants align their verbal and nonverbal orientations. Previous research has shown that increased mimicry can lead to greater cooperation (e.g. [15, 16]), increased empathy for others (e.g. [17–19]), and greater social influence (e.g. [4, 20]). Mimicking agents have shown to be more persuasive and received more positive trait ratings than non-mimickers, despite participants' ability to detect direct mimicry after a while (e.g. [21]).

The precision of a full-body wireless MoCap device, the sensation of immersion in a VR world, and the realistic speech generation, facial expressions, eye gaze, and lip synchronization, all in real-time, have been combined in our first prototype. Our goal was to deliver a proof-of-concept system that combines behaviors in this way for law enforcement training. The objective of the system is to simulate each of the cooperation paradoxes in order to teach users (e.g. a trainee law enforcement officer) various effects and consequences of their verbal and nonverbal behaviors (as interpersonal stances and attitudes) on an interviewee (e.g. a suspect). This mixed-reality simulation environment enables the presentation of multimodal behaviors consistent with what occurs in real-life situations; e.g. *conflicts*. It can be used to model human reactions to *cooperative* and *competitive* behaviors, in conjunction with the presence and absence of *mimicry*. This modeling will utilize both verbal and nonverbal methods of examining behavior, as well as studying the dynamics of human cooperation in social interactions. It will, also, teach good conversational and social skills, and provide a valuable evidence base that informs better training and social coaching for law enforcement. It is thus particularly valuable to front-line professionals whose training needs are too complex (and therefore too costly) or too dangerous to simulate in real-life. We, therefore, propose this system for the investigation of human users' behaviors when interacting with a mimicking embodied conversational agent.

8 Future Work

Previous works have mainly focused on *single-action mimicry* and demonstrated its effects on mimickees. In project-SENSE we have set the groundwork for capturing the social and emotional effects of *direct behavioral mimicry* in *sustained and prolonged interactions* with individuals. This will be made possible by examining the MVN recordings of participants and measuring movements along various dimensions. A similar approach was taken by [7] to extract precise amounts of movements (e.g. fidgeting) from the Xsens MVN Awinda motion data. They recognized that in most studies, measuring behavioral mimicry is performed through manually coding events from video recordings (e.g. [22]) that raises issues such as *subjectivity* [23] and making comparisons merely between *isolated behaviors* such as face touching [7] although numerous *facial mimicry* research works also use EMG (e.g. [24]) and facial tracking systems (e.g. [25]). We also intend to integrate the Microsoft Kinect camera [26] into the project-SENSE prototype, as an *affordable* replacement for the Xsens MVN Awinda. Comparing the performance and precision of the *OptiTrack* as a professional MoCap system similar to the Xsens MVN Awinda is, furthermore, a future avenue we will be exploring.

References

1. Taylor, P.J.: The role of language in conflict and conflict resolution. In: Holtgraves, T. (ed.) The Oxford Handbook of Language and Social Psychology, pp. 459–470. Oxford University Press, New York (2014)
2. Taylor, P.J., Larner, S., Conchie, S.M., Van der Zee, S.: Cross-cultural deception detection. In: Granhag, P.A., Vrij, A., Verschuere, B. (eds.) Deception Detection: Current Challenges and Cognitive Approaches, pp. 175–202. Wiley-Blackwell, Chichester (2014)

3. Gottman, J., Markman, H., Notarius, C.: The topography of marital conflict: a sequential analysis of verbal and nonverbal behavior. J. Marriage Fam. **39**, 461–477 (1977)
4. Chartrand, T.L., Bargh, J.A.: The chameleon effect: the perception-behavior link and social interaction. J. Pers. Soc. Psychol. **76**, 893–910 (1999)
5. OculusVR, Copyright © (2016). https://www.oculus.com/
6. Xsens Technologies, Copyright © (2016). https://www.xsens.com
7. Poppe, R.W., Van Der Zee, S., Heylen, D.K.J., Taylor, P.J.: AMAB: automated measurement and analysis of body motion. Behav. Res. Methods **46**(3), 625–633 (2014)
8. Virtual Human Toolkit, Copyright © (2016). https://vhtoolkit.ict.usc.edu
9. Leuski, A., Traum, D.: NPCEditor: a tool for building question-answering characters. In: Paper presented at the 7th International Conference on Language Resources and Evaluation (LREC), Valetta, Malta, 19–21 May (2010)
10. SmartBody, Copyright © (2016). http://smartbody.ict.usc.edu
11. Hartholt, A., Traum, D., Marsella, S.C., Shapiro, A., Stratou, G., Leuski, A., Morency, L.-P., Gratch, J.: All together now: introducing the virtual human toolkit. In: Aylett, R., Krenn, B., Pelachaud, C., Shimodaira, H. (eds.) IVA 2013. LNCS (LNAI), vol. 8108, pp. 368–381. Springer, Heidelberg (2013). doi:10.1007/978-3-642-40415-3_33
12. Unity3D, Copyright © (2016). http://unity3d.com/unity
13. McNeill, D.: So you think gestures are nonverbal? Psychol. Rev. **92**, 350–371 (1985)
14. Kendon, A.: Movement coordination in social interaction: some examples described. Acta Psychol. **32**(2), 101–125 (1970)
15. Taylor, P.J., Thomas, S.: Linguistic style matching and negotiation outcome. Negot. Confl. Manage. Res. **1**, 263–281 (2008)
16. Chartrand, T.L., Lakin, J.L.: The antecedents and consequences of human behavioral mimicry. Annu. Rev. Psychol. **64**, 285–308 (2013)
17. Stel, M., Van Baaren, R.B., Vonk, R.: Effects of mimicking: Acting prosocially by being emotionally moved. Eur. J. Soc. Psychol. **38**, 965–976 (2008)
18. Hasler, B.S., Hirschberger, G., Shani-Sherman, T., Friedman, D.A.: Virtual peacemakers: mimicry increases empathy in simulated contact with virtual outgroup members. Cyber Psychol. Behav. Soc. Networking **17**(12), 766–771 (2014)
19. Stel, M., Vonk, R.: Empathizing via mimicry depends on whether emotional expressions are seen as real. Eur. Psychol. **14**, 342–350 (2009)
20. Richardson, B., Taylor, P.J., Snook, B., Conchie, S.M., Bennell, C.: Language style matching and confessions in police interrogations. Law Hum. Behav. **38**, 357–366 (2014)
21. Bailenson, J.N., Yee, N., Patel, K., Beall, A.C.: Detecting digital chameleons. Comput. Hum. Behav. **24**, 66–87 (2008)
22. Stel, M., van Dijk, E., Olivier, E.: You want to know the truth? then don't mimic! Psychol. Sci. **20**, 693–699 (2009)
23. Scherer, K.R., Ekman, P.: Methodological issues in studying nonverbal behavior. In: Scherer, K.R., Ekman, P. (eds.) Handbook of Methods in Nonverbal Behavior Research, pp. 45–135. Cambridge University Press, New York (1982)
24. Likowski, K.U., Mühlberger, A., Seibt, B., Pauli, P., Weyers, P.: Modulation of facial mimicry by attitudes. J. Exp. Soc. Psychol. **44**, 1065–1072 (2008)
25. Stel, M., Blascovich, J., McCall, C., Mastop, J., Van Baaren, R.B., Vonk, R.: Mimicking disliked others: effects of a priori liking on the mimicry-liking link. Eur. J. Soc. Psychol. **40**, 876–880 (2010)
26. Microsoft Kinect, Copyright © (2016). https://www.microsoft.com/en-us/kinectforwindows/