



ELSEVIER

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Head-gestures mirroring detection in dyadic social interactions with computer vision-based wearable devices

Juan R. Terven^{a,d}, Bogdan Raducanu^b, María Elena Meza-de-Luna^c, Joaquín Salas^{a,*}

^a Instituto Politécnico Nacional, CICATA Querétaro, Cerro Blanco 141, Colinas del Cimatario, Querétaro, 76160, Mexico

^b Computer Vision Center, Edificio "O" – Campus UAB, 08193 Bellaterra, Spain

^c Universidad Autónoma de Querétaro, Cerro de las Campanas, Querétaro, 76010, México

^d Instituto Tecnológico de Mazatlán, Corsario 1 203, Mazatlán 82070, México

ARTICLE INFO

Article history:

Received 29 November 2014

Received in revised form

26 May 2015

Accepted 26 May 2015

Available online 4 November 2015

Keywords:

Head gestures recognition

Mirroring detection

Dyadic social interaction analysis

Wearable devices

ABSTRACT

During face-to-face human interaction, nonverbal communication plays a fundamental role. A relevant aspect that takes part during social interactions is represented by mirroring, in which a person tends to mimic the non-verbal behavior (head and body gestures, vocal prosody, etc.) of the counterpart. In this paper, we introduce a computer vision-based system to detect mirroring in dyadic social interactions with the use of a wearable platform. In our context, mirroring is inferred as simultaneous head noddings displayed by the interlocutors. Our approach consists of the following steps: (1) facial features extraction; (2) facial features stabilization; (3) head nodding recognition; and (4) mirroring detection. Our system achieves a mirroring detection accuracy of 72% on a custom mirroring dataset.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

During face-to-face human interaction, nonverbal communication plays a fundamental role, as it is used to support the spoken message and to put special emphasis on certain aspects of it [1]. Usually, nonverbal communication is manifested through a multiplicity of behavioral cues including head movements, body postures/gestures, facial expressions, winks, tone of voice, verbal accent, and vocal utterances [2]. Sometimes, these cues are also known as social signals, a term coined by Pentland [3], because they are an undivided part of our social interaction. He was also the first one to claim that they could be quantified automatically to infer from them behavioral patterns in human interactions. The first attempt to prove this theory was reported in Curhan and Pentland [4], where the authors tried to predict the behavioral outcome of employment selection interviews using non-verbal audio cues. The same approach has also been applied for predicting salary negotiations [5] and speed-dating conversations [6]. Some research on behavior analysis during social interactions has focused on different aspects such as the role of participants in news broadcasts and movies [2,7], the detection of the leadership role during meetings [8,9], the inference of personality traits

[10,11], and the simultaneous prediction of a job interview outcome and personality [12].

A very important aspect that takes part during social interaction is represented by mirroring, *i.e.*, when one interlocutor tries to mimic the attitude of the counterpart [13], by imitating speech patterns (accent, voice prosody), facial expressions, postures, gestures, and idiosyncratic movements. The study of mirroring has attracted the interest of psychologists for a long time [14]. Back then, the analysis was based on the manually annotation of videotapes for listener movements and the prosody of the accompanying speech. It has not been until recently that the study of mirroring captured the attention of the Human-Computer Interaction (HCI) community. A comprehensive recent survey can be found in Wagner et al. [15]. The *mirroring behavior* reveals very important information regarding participants' inter-personal states and attitudes and it represents a reliable indicator of cooperativeness and empathy during interaction. Recent research revealed that people who, even consciously, mimic the behavior of others activate behavioral strategies which may increase their chances to achieve their goals [16]. Thus, a social interaction presenting a high number of *mirroring behavior* is perceived to run more smoothly and the chances to reach a positive outcome or agreement increase significantly.

On the other hand, a backchannel is defined as a modality used by a listener to briefly intervene during the mainstream presentation given by a speaker in order to show his/her level of support with respect to the topic being discussed. The definition of a backchannel is relative and is strongly dependent on the

* Corresponding author.

E-mail addresses: j.r.terven@ieee.org (J.R. Terven), bogdan@cvc.uab.es (B. Raducanu), mezamariel@gmail.com (M.E. Meza-de-Luna), jsalasr@ipn.mx (J. Salas).

particularity of the problem to be tackled. In the audio domain, backchannels are represented as linguistic vocalization such as *hmmm*, *aha*, *uhhh*, and *yeah*. In the visual domain, backchannels are usually associated with gestures such as smiling, winking, and head nodding; however, there could be more complex structures, such as facial expressions. Sevin et al. [17] used multimodal fusion of audio-visual backchannels in order to trigger the adequate response in an avatar. In a similar approach the objective was to compare the realization of the feedback signal in both an avatar and a physical embodied robot (AIBO) based on audio-visual backchannels fusion [18]. The improvement of head gestures detection (*nodding* in particular) has been done not only considering the visual information, but also taking into account the speaking status [19]. In addition, backchannels are usually studied in the context of *turn-taking* [20].

As the vast literature on the subject suggests, mirroring is a complex phenomena [13]. In this paper, we focus on the detection of head-nodding as a relatively simple non-verbal communication modality because of its significance as a gesture displayed during social interactions. According to social psychology, head nodding plays a very important role during social interactions. Apart from the obvious function of signaling a yes, head nods are used as backchannels to display interest, enhance communication or anticipate the counterpart intention for turn claiming [21,20]. Additionally, head nods can be used during speaker turns to elicit feedback from the listener [20]. The psychology literature suggests that the frequency of head nod events in face-to-face interactions can reveal personal characteristics or even predict outcomes. For instance, job applicants producing more head nods in employment interviews have been reported to be often perceived as more employable than applicants who do not [22,23]. In this sense, the ability to automatically detect head nods could be useful to build automatic inference methods of high-level social constructs. Therefore, in our context, mirroring will be inferred head noddings displayed by the interlocutors. As our results show even this reduced perspective of the problem is detect mirroring using a wearable device.

Our main contributions are:

1. We introduce an approach to extract a visual backchannel gesture (*i.e.*, nodding) from videos captured using wearable devices.
2. We develop a computer vision-based method to detect mirroring automatically, as an identical head gesture in a dyadic conversation, using a wearable device.
3. We provide a recorded custom database, representing a dyadic conversation setting, where each of the participants is wearing a pair of smart glasses. Ground-truth is provided by annotated head movements captured by a pair of fixed cameras facing each participant.

To our knowledge, this is the first time smart glasses equipped with a camera (see Fig. 2) have been used in such a setting.

The rest of the paper is structured as follows. Section 2 is dedicated to a review of the state-of-the art in the study of mirroring behavior both from a psychological and computational perspective. In Section 3, we describe the experimental setup and scenario definition used in our study. Section 4 presents our approach for automatic mirroring detection. In Section 5, we present the experimental results (both quantitatively and qualitatively, in terms of user experiences). Finally, we present our conclusion and provide guidelines for future work.

2. Related work

Perhaps because the study of mirroring can be approached by different scientific disciplines, we have found that several terms have been used to express its meaning. Pentland [24] stated that “mirroring occurs when one participant subconsciously copies another participant’s prosody and gesture”. This definition has been widely used to mean the display of similar postures while people interact with one another [25], and has been exemplified by situations where when “person A nods or smiles following person B who has nodded or smiled too” [26]. Interestingly, some researchers [27,26,28] give equivalent semantic meaning to mirroring and other words such as synchrony. However, in this research we use the definitions stated by Burgoon et al. [29] where mirroring involves visual behaviors identical in form, while interpreting synchrony as “a smoothly meshed coordination between the interactants”. This section provides an overview of the *mirroring behavior* from the psychological and computational perspectives.

2.1. Psychological perspective

The study of mirroring in psychology, as a nonverbal behavioral process, dates back to the early 1970s. The researchers noticed that during a conversation, the parties involved exchanged both words and nonverbal cues, the latter as an effective form to adjust from each other in order to reach convergence in communication. Thus, early studies focused upon voice prosody such as accent imitation [30], vocal intensity [31], pause frequency [32], speaking rate [33], and speech patterns [34]. However, during a social interaction, people not only tend to imitate vocal features, but also to match each other’s facial expressions and body gestures. For instance, LaFrance [35] found that listeners tend to mirror a speaker’s posture whom they find engaging. Another research concluded that newborn babies [36] and adults [37] imitate facial gestures, while infants imitate vocalic sounds [38].

A comprehensive analysis of the role of mirroring in social interactions and how it affects our decisions and behavior can be found in Guéguen [16]. At a cognitive level, the *mirroring behavior* could be explained through the mind-body dualism, according to which, the mental processes are closely related to the body (the so-called relationship between thinking and action) [39]. In other words, the transition between mental states could be understood based on its analogy with the trajectory of a dynamic system through a series of space-states. From the point of view of evolution, mirroring happened long before human developed their linguistic capabilities, as a mechanism used by people to survive by helping them to communicate and coordinate better [40]. This explains why nowadays, due to the significant role played in modern society by social interaction, mirroring constitutes an automatic and unconscious act rooted in our brain [41]. Thus, seen from this social perspective, mirroring arose from the need to increase the social coherence among the members of a group and to feel a sense of psychological connection between themselves. In other words, individuals who were able to mimic each other had more opportunities to experience this psychological connection and would have had more probabilities to be kept within the community. The experimental evidence for this statement is ample. For instance, levels of mimicry are positively correlated with sales rates [42], helping behavior to explicit verbal solicitation [43], romantic interest [44], and success in patients undergoing psychotherapy [45]. Based on these research results, in our study, we establish a link between the detected mirroring and the level of satisfaction derived from the social interaction. This highlights the usefulness of the proposed method in sociological studies.

Download English Version:

<https://daneshyari.com/en/article/407119>

Download Persian Version:

<https://daneshyari.com/article/407119>

[Daneshyari.com](https://daneshyari.com)