Assessing the Influence of Mirroring on the Perception of Professional Competence using Wearable Technology

María Elena Meza-de-Luna Juan R. Terven Bogdan Raducanu Joaquín Salas

***Abstract*—Nonverbal communication is an intrinsic part of our everyday face-to-face meetings. A frequently observed behavior during social interactions is mirroring, in which one person tends to mimic the attitude of his/her counterpart. The purpose of this paper is to show that a computer vision system could be used to predict the perception of competence in dyadic interactions through the automatic detection of mirroring events, considered as the quasi-simultaneous occurrence of head-nodding gestures. In order to prove our hypothesis, we developed: (1) a socially-aware assistant for mirroring detection, using a wearable device which includes a video camera and (2) an automatic classifier for the perception of competence, using the number of nodding gestures and mirroring events as predictors. For our study, we used a mixed-method approach in an experimental design where 48 participants acting as customers interacted with a confederated psychologist. Our findings show that the number of either nods or mirroring events has a significant influence on the perception of competence. Our results suggest that: (1) to classify the perception of competence, mirroring is a better predictor than nodding; (2) customer mirroring is a better predictor than psychologist mirroring and (3) that psychologist nodding is a better predictor than customer nodding.**

INTRODUCTION

Nonverbal communication is an intrinsic part of our everyday face-to-face interactions. It is this social signaling that reinforces the conveyed message offering supplementary information smoothly integrated to our verbal channel. Some psychologists suggest (Ambady & Rosenthal, 1992; Gladwell, 2006) that in many situations the social signals are just as important as the conscious content for determining human behavior. In other words, it is a piece of information that defines to a large extent our attitude towards a situation and our reaction. A possible way to understand the power of social signals is to make an analogy with trying to follow a conversation in a foreign, unfamiliar language. Despite our lack of knowledge of the language, we may still infer some general patterns, *e.g.,* who is leading the conversation, or if there is a friendly or tense discussion.

A frequently observed behavior during social interactions is mirroring, in which one person tends to mimic the non-verbal prosody (head movements, hand gestures, facial expressions, tone of voice, verbal accent, breathing, etc.) of his/her counterpart. The role of mirroring is to signal empathy between people and, in general, is an early indicator of agreement of the interactional process. It has long been observed that head gestures such as nodding, increase the opportunities for a person to be liked (Gifford et al., 1985; McGovern et al., 1979), while the occurrence of mirroring is an early predictor of acceptance (Farley, 2014; Gueguen et al., 2009, Jacob et al. 2011; Van Baaren et al. 2003). However, as far as we know, there is a void in the literature on how nodding/mirroring in a dyadic interaction can influence the perception of competence [1]. Hence, this investigation intends to answer whether it is possible to predict the perception of competence through nodding gestures or mirroring events in dyadic interactions.

[1]Competence entails the possession of skills, talents, and capability with traits that include being clever, competent, creative, efficient , foresighted , ingenious, intelligent , and knowledgeable (Cuddy et al., 2008). Epstein & Hundert, 2002 defined professional competence as *"the habitual and judicious use of communication, knowledge, technical skills, clinical reasoning, emotions, values, and reflection in daily practice for the benefit of the individual and community being served."* They conferred seven dimensions to the concept: Cognitive, technical skills, integrating biomedical and psychosocial data in clinical reasoning, relationships, affective/moral, habits of mind, and context.

The thoughtful analysis of social interactions allows us to reach a better diagnosis and therefore to advance the understanding of their consequences. However, the process of studying these is a cumbersome one, as they are complex phenomena that generate large amounts of information. For instance, it has been reported that one hour of observation leads to about 10 hours of annotations (Paxton & Dale, 2013). In addition, gathering information may be a challenge by itself, since obtrusively intervening in a social interaction could affect its development. Besides, most of the analysis in this area is done after the interaction has already finished. As stated by Pentland (2007), having a socially-aware assistant decoding/interpreting the information contained in complex social signals could pave the way for correcting our attitude while there is still time to reduce the chances for an unwelcome result in the interaction.

Hence, inspired by the pioneer work of Pentland in the processing of social signals, in this paper we present a socially-aware assistant to recognize the perception of competence that results from an interaction. Our main hypothesis is that it is possible to develop a wearable system to determine whether the interlocutor has been perceived as competent or non-competent in a dyadic conversation. Indeed, we provide a practical confirmation that the perception of competence can be evaluated using nodding gestures or mirroring events detected using a video camera embedded in the wearable device. Our method can be summarized as follows (see Figure 1): During the interaction, participants (acting as *customers*) ask a confederated psychologist (acting as *service provider*) for professional advice. The conversation is recorded using wearable and fixed cameras to generate a set of images $\mathcal{I}(t)$. The role of the pair of fixed cameras that appear in our sketch scenario is only to serve as a reference, such that the images captured by the wearable could be compared to. These images are fed to a head-gesture recognizer that detects the head nodding gestures $\eta_1(t), \eta_2(t)$ in real time. Synchronous recognition of head nods leads to mirroring detection $\mu(t)$. To automatically train the classifier that detects the perceived competence, we obtain ground truth from a set of qualitative interviews carried out with customers. This creates a classification space $\mathcal{C}$, which for a certain degree of activity provides an automated evaluation $\mathcal{L}$ of the likely outcome of the interaction.

The main contribution of our approach is that we propose a socially-aware assistant based on wearable technology, *i.e.,* smart glasses, which have a video camera embedded in the bridge that connects the two lenses (see Figure 2). Our choice for a solution based on computer vision is because: (1) its non-invasive nature, which increases the chances to be accepted by people; (2) it offers a first-person perspective, compared to the classical fixed cameras, which offer a third-party perspective; and (3) the acceptance of its use may increase given that the camera is embedded in an everyday artifact used by a significant number of people. In our opinion, these are requirements in applications related to social interaction analysis, as the technology in this field should be centered on the people.

RELATED LITERATURE

During face-to-face human interaction, nonverbal communication plays a fundamental role, as it is used to support the spoken message by placing special emphasis on certain aspects of it (Knapp & Hall, 2009). 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 (Vinciarelli et al., 2009). Sometimes, these cues are also known as social signals, a term coined by Pentland (2007), because they are an undivided part of our social interaction.

According to social psychology, head nodding plays a paramount role during social interactions. Apart from the obvious function of

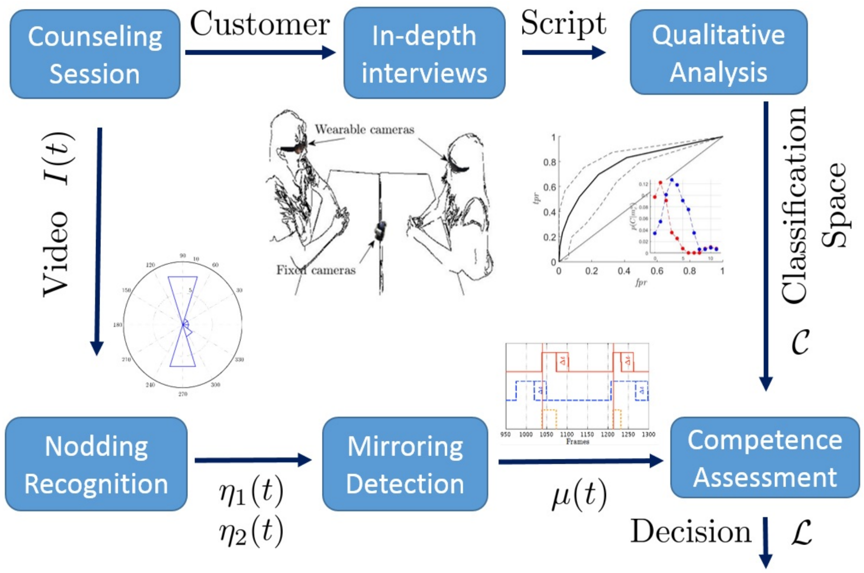


Fig. 1. Assessing Competence. During an interaction, the conversation is recorded using wearable and fixed cameras to generate a set of images $\mathcal{I}(t)$. These images are fed to a head-gesture recognizer that detects in real-time nodding $\eta_1(t), \eta_2(t)$. Synchronous recognition of head nods leads to mirroring detection $\mu(t)$. During training, a set of qualitative interviews was carried out with customers to evaluate their perception of competence and satisfaction during the interaction. This creates a classification space $\mathcal{C}$, which for a certain degree of activity provides an automated evaluation $\mathcal{L}$ of the likely outcome of the interaction

Fig. 2. Smart glasses used as wearable camera. The glasses have a high-definition camera embedded in the bridge connecting the two lenses (by Pivothead Inc., with permission).

signaling a yes, head nods are used as backchannels to display interest, enhance communication or anticipate the counterpart intention for turn claiming (Allwood & Cerrato, 2003; Hadar et al., 1985). Besides, 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 (Gifford et al., 1985; McGovern et al., 1979).

Mirroring is another important event that takes part during a social interaction, i.e., when one interlocutor tries to mimic the attitude of the counterpart (Chartrand & Bargh, 1999), 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 (Condon & Ogston, 1971). Mirroring behavior has an important, but barely noticed, impact on our daily life. It reveals large pieces of information regarding the participants' inter-personal states and attitudes and represents a reliable indicator of cooperativeness and empathy during interaction (Wagner et al., 2014) to the extent that it has been demonstrated to be positively associated with romantic interest and attraction (Farley, 2014).

In marketing, mirroring has proved to influence the customers' behavior in different settings. For instance, mimicking the verbal behavior of customers in a restaurant was associated with a higher rate of customers who gave a tip and with larger amounts of the tips (Van Baaren et al. 2003). In the retail sector, Jacob et al. (2011) found an increment in sales rates when mimicking (78.8%) versus non mimicking (61.8%). Besides, clerk's suggestions were more influential when purchasing a product: 71.1% of customers exposed to mirroring bought the object suggested by the seller vs. 46.2% in the non-mirroring case. Similar results were reported by Guéguen (2011), who found that mirroring is associated with greater compliance with the sellers' suggestions and customers rating the sellers and the store with greater positive evaluations. Indeed, mirroring seems to be a powerful technique to increase helping behavior. For instance, Van Baaren et al. (2004) set an experiment where participants were invited to evaluate different advertisements. The experimenter, who was seated in front of the participant, mimicked the participant's posture (position of their arms, legs) or not. When the task finished the experimenter "accidentally" dropped six pens on the floor. It was found that participants in the mirroring condition picked up the pens more often (100%) than participants in the non-mirroring condition (33%). Similarly, Gueguen et al. (2011) found a positive effect of mirroring on helping behavior to explicit verbal solicitation; on his experiment 76.7% of the participants on the mirroring group agreed to help vs. 46.7% in the non-mirroring control condition. Even more, 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 (Gueguen et al., 2009). Thus, a social interaction presenting a high number of mirroring events is perceived to run more smoothly and the chances to reach a positive outcome or an agreement increase significantly.

Recently, computer vision systems have been integrated to the nonverbal human communication studies to enhance its understanding. Pentland (2007) was the first one to claim that social signals could be quantified automatically to infer behavioral patterns in human interactions. The first attempt to prove this idea was reported by Curhan & Pentland (2007), who 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 (Caneel, 2005) and speed-dating conversations (Madan et al., 2005). Some research on behavior analysis during social interactions has focused on different aspects such as the role of participants in news broadcasts and movies (Vinciarelli et al., 2009; Weng et al., 2007), the detection of the leadership role during meetings (Raducanu & Gatica-Perez, 2012; Sanchez-Cortes et al., 2010), the inference of personality traits (Lepri et al., 2012; Staiano et al., 2012), and the simultaneous prediction of a job interview outcome and personality (Nguyen et al., 2013). In this sense, the ability to automatically detect head nods could be useful to build automatic inference methods of high-level social constructs.

With all this evidence about the importance of social signals on human behavior, we aim to use a computer vision system based on a wearable device to evaluate whether mirroring behavior, understood as simultaneous head nodding gestures, can be used as an early predictor of the perception of competence and the inherent inclination to hire a professional service. We want to evaluate how the gestures of two interacting people (customer and a service provider) can influence a customer's perception of the service provider competence.

Method

We integrated mixed methods (Hernández et al., 2010) in an experimental design in order to assess the impact of nodding/mirroring in the perception of professional competence. The experiment consisted of a conversational scenario in which a customer interacted with a confederated psychologist who answered in the same verbal fashion to each customer. However, we instructed the confederated psychologist to emit different backchannel regulatory gestures. During the interaction, our wearable device-based computer vision system recognized nodding gestures and detected mirroring behavior automatically (Terven et al., 2016). After the social interaction, we made semi-structured interviews to participants in order to assess, qualitatively, the satisfaction and their intentions to hire the psychologist. Finally, we used the number of either nods or mirroring events to determine whether the customer perceived the psychologist as competent or non-competent. The whole process is summarized in Figure 1.

Participants

Our inclusion criteria to participate consisted of being a college student at least 18 years old. The final sample had 48 volunteers (50% women), ranging from 18 to 44 years (M = 21.83, SD = 4.10); of these, 29.2% were majoring in sociology, 25% in politics, 18.8% in architecture, 10.4% in engineering, 6.3% in journalism, 4.2% in business, 4.2% in mathematics, and 2.1% in nursing.

Experimental Procedure

In our scenario, we controlled: (1) the physical environment in a lab setting; (2) the verbal explanations and gestures of psychologist; and (3) the homogeneity of the sample, represented by the same amount of men and women as participants, all being undergrad students.

We invited the participants to collaborate in an inquiry related to dyadic conversations. The participants were instructed to present a personal problem to a psychologist and ask for advice. We asked the participants to explain the situation and to conduct their conversations around three questions for the psychologist: (1) what to do in the current situation?, (2) what is the psychologist's experience with similar problems? and, (3) how many sessions are needed? Afterwards, the participants were conducted outside of the psychologist's office and the smart glasses were activated. The participants knocked on the office door to start the experimental process where they asked for advice. The confederated psychologist answered the questions in the same verbal style but controlling his nodding gestures in one of three occurrence levels, which constitute the experimental cases: Low, the psychologist acted restricting his nodding gestures; medium, the psychologist nodded and mirrored the customer's nodding, and high, the psychologist increased nodding and promoted mirroring. When finishing the interaction, the participants left the psychologist's office and the glasses were deactivated. The participants were conducted to another lab setting, where they were interviewed with a semi-structured method to inquire about their interaction. The interview's aim was to evaluate their satisfaction, their perception of the psychologist's competence and the customer's inclination of hiring the psychologist.

We recorded each customer-psychologist conversation with two static cameras and two wearable cameras. The customer and the psychologist used smart glasses during the whole interaction. For the static cameras, we used Microsoft LifeCam Studio cameras fixed on the table looking at each interlocutor. For wearable cameras, we used Pivothead glasses. After recording each session, we synchronized the four videos. The final interview inquiring perception of the customer-psychologist interaction was digitally recorded.

Based on this scenario, we created a dataset, which for each experimental session consisted of: (1) Four videos (two from the customer's and the psychologist's wearable cameras and two from the fixed cameras), and (2) a recorded qualitative interview. The experiment took place between October 2014 and February 2015. On average, the dyadic conversation with the psychologist was three minutes long, and the qualitative interview was five minutes long. In all cases, we rewarded the individual participation with 50 Mexican pesos (equivalent to about US$3.00).

This study followed ethical standards as stipulated by the American Psychological Association (Flavio et al., 2010). Participants signed an informed consent letter. Confidentiality and person's anonymity were maintained at all times. All video and audio recordings were made with participant's written authorization. The protocol was approved by the Ethics Committee from the Universidad Autónoma de Querétaro.

Analysis Procedure

Our procedure was the following: (1) analyze the qualitative data about the psychologist's competence with Grounded Theory (Charmaz, 2008; Glaser et al., 1968; Strauss & Corbin, 2002), (2) quantify head nodding gestures and mirroring events using computer vision algorithms, and (3) analyze the relation between head gestures and perception of competence plus hiring intention, using the probabilistic evaluation of membership to a class on a Bayesian framework. In what follows, we describe each step.

Qualitative Analysis on Professional Competence: The audio of the interviews was recorded digitally. We then used Sound Scriber (Breck, 1998) to transcribe verbatim and Atlas.ti (Muhr, 2004) to analyze the qualitative data. We analyzed the qualitative data obtained from semi-structured interviews, using the principles of Grounded Theory (Charmaz, 2008; Glaser et al., 1968; Strauss & Corbin, 2002). Overall, the participants' testimonies were categorized into satisfaction or dissatisfaction, leading to perception of competence. The competence perception was related to two results: inclination to hire or not to hire. Two investigators did the analysis separately in order to assess the reliability by intercode procedure (Hernández et al., 2010), obtaining high reliability (0.97).

Mirroring Detection with Wearable Technology: From the video acquired during the dyadic conversations with the confederated

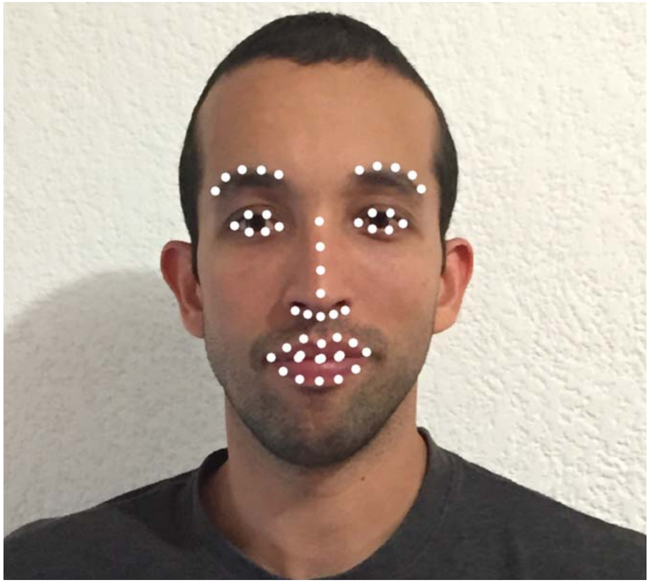


Fig. 3. Facial Features extracted with the SDM Xiong & Torre, 2013. These features are used to track the face in consecutive video frames. The face tracking leads to head motion, which is used to detect nodding gestures on both participants.

psychologist, we applied an automatic gesture and mirroring detection based on computer vision. The procedure started by tracking facial features (shown in Figure 3) using the Supervised Descent Method (SDM) from Xiong & Torre (2013), we then applied a stabilization step to compensate for wearable camera motion.

With the camera motion stabilized, we created a set of descriptors using histograms of orientations (HOO) (Freeman & Roth, 1995) that are used as inputs for a head nodding classifier. Having detected the head nods from both participants, we measured mirroring in both directions following an approach similar to Feese et al. (2012).

We defined two events: *Person $\mathcal{A}$ is mirroring Person $\mathcal{B}$* or ($m\mathcal{AB}$); and *Person $\mathcal{B}$ is mirroring Person $\mathcal{A}$* or ($m\mathcal{BA}$). To count an $m\mathcal{AB}$ event, person $\mathcal{A}$ needs to start displaying gesture ξ after person $\mathcal{B}$ started and within a time Δt after person $\mathcal{B}$ stopped displaying gesture ξ. In case that person $\mathcal{A}$ displays ξ multiple times while $\mathcal{B}$ is displaying ξ, only one event is counted. Similarly, a $m\mathcal{BA}$ event is triggered when person $\mathcal{B}$ starts displaying gesture ξ after person $\mathcal{A}$ started and within Δt after person $\mathcal{A}$ stopped displaying gesture ξ. Gesture repetitions were treated the same way. More formally, given a sequence of gestures $g^{\xi}_{1\ldots N^{\xi}}$ of person $\mathcal{A}$, the start and end times of each gesture is given by $t_1\left(g^{\xi}_i\right)$ and $t_2\left(g^{\xi}_i\right)$ respectively. An $m\mathcal{AB}$ event is triggered if (following Feese et al., 2012):

$$\begin{gathered} g^{\mathcal{A}}_i = g^{\mathcal{B}}_j, \\ t_1\left(g^{\mathcal{B}}_j\right) < t_1\left(g^{\mathcal{A}}_i\right) < t_2\left(g^{\mathcal{B}}_j\right) + \Delta t. \end{gathered} \tag{1}$$

Figure 4 shows two fragments from one of the videos in our dataset. The top plot depicts the nodding gestures performed by person $\mathcal{A}$, the middle plot depicts the nodding gestures performed by person B, and the bottom plot depicts the mirroring events. Figure 4(a) shows $m\mathcal{AB}$ events; the first mirroring event occurs when person $\mathcal{A}$ mirrors person $\mathcal{B}$ after person $\mathcal{B}$ stopped displaying the nodding gesture, but within a predefined window Δt. The second mirror event occurs just after person $\mathcal{B}$ started the nodding gesture. Figure 4(b) shows $m\mathcal{BA}$ events; in this case, the two mirroring events occur just after person $\mathcal{A}$ started the nodding gesture. The window Δt is heuristically determined, taken into consideration the analysis of our dataset, where the average gesture duration is 1.36s.

Relation between Head Gestures and Perception of Competence: As a result of the previous stages, we end up with a set of measurements of the number x of nodding and mirroring events detected by the computer vision system, and labels $\mathcal{C}$ to identify the customer perception of the psychologist's competence. This set of measures may be represented by $S = \{(x, \mathcal{C})_i\}$, for $i = 1, ..., M$. In our experiment, $x \in n^f_c, n^f_p, n^w_c, n^w_p, m^f_c, m^f_p, m^w_c, m^w_p$, where n and m reflect the number of automatically detected nods and mirroring events respectively, f and w represent whether the observation was performed with a fixed or wearable camera, and c and p highlight whether the observation was made on the customer or on the psychologist. For instance, n^f_p is the number of nods made by the psychologist as observed from a fixed camera. In addition, $\mathcal{C} \in C_n, C_c$ is the label assigned to the perception acquired by the customer about the competence C_c or non-competence C_n of the psychologist.

Over the years, many supervised learning techniques aimed to classify observations have been developed, including Naïve Bayes, Logistic Regression, Decision Trees or Support Vector Machines, among others (Bishop, 2007). Since our sample is small and our feature is one dimensional, a reasonable choice is to use a Bayesian framework (Russell & Norvig, 1995). Here, the membership of an observation to a class is evaluated on the basis of an estimate of the likelihood, $P(x|C)$, obtained by training, and prior knowledge, $P(C)$, about the competence or non-competence of a particular psychologist, obtained from the observations, such as

$$P(C|x) = kP(x|C)P(C), \tag{2}$$

where for our problem, the proportionality constant k is the same for both $P(C_c|x)$ and $P(C_n|x)$. We estimate the likelihood, $P(x|C)$, and the prior, $P(C)$, directly from the data to avoid restrictive assumptions about their form. In the case of the likelihood, $P(x|C)$, one possible way to do this could involve the use of normalized histograms. However, both the selection of the bins width and the number of samples can have a substantial effect on its estimation. For instance, complex, expensive or time-consuming experiments could result in a sparse set of observations. It has been argued that in these cases a better estimation of a particular probability cell could be obtain by drawing information from nearby cells (Chu et al., 2015). Aitchison and Aiteken were the first to introduce a kernel function to smooth discrete probability distributions (Aitchison & Aitken, 1976). Nowadays, research has resulted in several kernel smoothers to select from (Kokonendji & Kiesse, 2011), whose suitability depends on the particular problem at hand and the features in the model that require to be enhanced. At our end, we compute an estimate of the mass probability function by applying a Gaussian kernel on the cumulative density function and uniformly sampling its inverse (see Figure 5). The bandwidth, h, is a tuning parameter to select the spread of the kernel smoother. Its value could be obtained by Bayesian optimization (Zougab et al., 2013) using cross-validation (Kokonendji & Kiesse, 2011) or plugging in from the sample (Chu et al., 2015). At its end, the prior estimation, $P(C)$, can be obtained empirically as the ratio between the number of elements assigned to each class, $\{C_n, C_c\}$, and the number of interviews.

A possible criteria to take a decision $\mathcal{L}(x)$ about the customer's perception of competence could be based on the relative values of

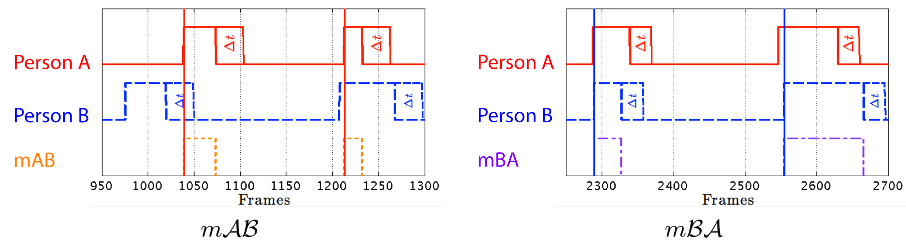


Fig. 4. Mirroring detection. Top and middle plots depict the occurrence of nodding gestures from person $\mathcal{A}$ and person $\mathcal{B}$ respectively. Bottom plot depicts the detected mirroring events. Note that there is a fixed interval of time Δt when the mirroring effect may take place.

the functions $P(C_c|x)$ and $P(C_n|x)$, such that

$$\mathcal{L}(x) = \begin{cases} \textbf{competent} & \text{if } \sum_{x_i \geq x} P(C_c|x_i) > \sum_{x_i < x} P(C_n|x_i), \\ \textbf{competent} & \text{otherwise.} \end{cases} \tag{3}$$

The performance of $\mathcal{L}(x)$ is evaluated using cross-validation with the leave-one-out method (James et al., 2013). That is, given our sample of n elements, we use $n-1$ elements to generate the estimated mass distributions, $\hat{P}(C_c|x)$ and $\hat{P}(C_n|x)$, and leave one observation out. We repeat this procedure for each element in the set of observations. In principle, this approach aims to minimize the estimation bias and, given our sample size, its computation cost is negligible.

To evaluate the performance of the classification strategy, we use the following procedure. Suppose we test whether a particular sample belongs to the class **competent** . Once tested, an interview where the psychologist has been assigned the label **competent** can be classified as such, resulting in a true positive (tp); otherwise, it could be assigned to the class non-competent, resulting in a false negative (fn). In a different case, an interview where the psychologist was labeled as non-competent can be classified as competent, resulting in a false positive (fp); or it could be classified correctly, in which case it results in a true negative (tn).

The generalization properties for the classifier could be evaluated using Receiving Operating Characteristics (ROC) analysis (Swets et al., 2000). For a given number of nods or mirroring events α, we evaluate the probability mass distribution corresponding to a correct decision based on the expressions

$$tp = \sum_{x \geq \alpha} P(C_c|x) \quad \text{and} \quad tn = \sum_{x < \alpha} P(C_n|x) \tag{4}$$

or a wrong decision, that is

$$fp = \sum_{x \geq \alpha} P(C_n|x) \quad \text{and} \quad fn = \sum_{x < \alpha} P(C_c|x). \tag{5}$$

For *ROC* analysis, the expressions summarizing the overall performance of the classifier include the sensitivity, or true positive rate (*tpr*), and fall-out, or false positive rate (*fpr*). They are defined as follows (Swets et al., 2000)

$$fpr = \frac{fp}{fp + tn} \quad \text{and} \quad tpr = \frac{tp}{tp + fn} \tag{6}$$

Then, by changing the value of α over the possible number of nods or mirrorings, we will obtain the scores for sensitivity and fall-out to construct an *ROC* curve. A widely used performance criteria corresponds to the area under the curve (*AUC*), basically the probability that a sample will be classified correctly (Swets et al., 2000), and equivalent to the Wilcoxon test of ranks (Fawcett, 2006), which in turn is preferred to the paired Student's t-test when the sample cannot be assumed to be normally distributed.

Using the data obtained for all the interviews, we performed a balanced two-way (or double factor) Analysis of Variance (*ANOVA*) to test the hypothesis of equal means. Given that mirroring detection is based on nodding recognition, these two factors cannot be assumed to be independent. Therefore, to avoid confounding, we grouped our data in two different sets: one for nodding and one for mirroring. In each configuration, we analyzed whether the number of recognized nods or detected mirroring events made by a customer or a psychologist are equally effective in distinguishing customer's perception of a psychologist's competence as the images captured come from either fixed or wearable cameras. We did not perform a test of equal variances, as some research has pointed out (Bradley, 1997) that the *ANOVA* is insensitive to departures from this assumption when the sample sizes are equal. When the null hypothesis is rejected, there is the need to pursue post hoc analysis via a multiple comparison procedure (*MCP*). The problem of selecting the most appropriate *MCP* is a difficult one because it depends on both objective and subjective properties of the observations being analyzed. For instance, Saville (2015) argues that the most conservative *MCP* , the ones aiming to reduce type I error, are the ones more inconsistent. Yet, it has been shown that when the variance is similar, the Tukey-Kramer's test offers optimal results (Stoline, 1981). Therefore, we use Tukey-Kramer's test, which essentially is a series of t-tests with correction for dataset-wise error-rate.

RESULTS

In this section, we detail the results of applying our method to the scenario described above. First, we explain how the label for competent and non-competent was generated from the interviews. Second, we present the results of applying our classification scheme.

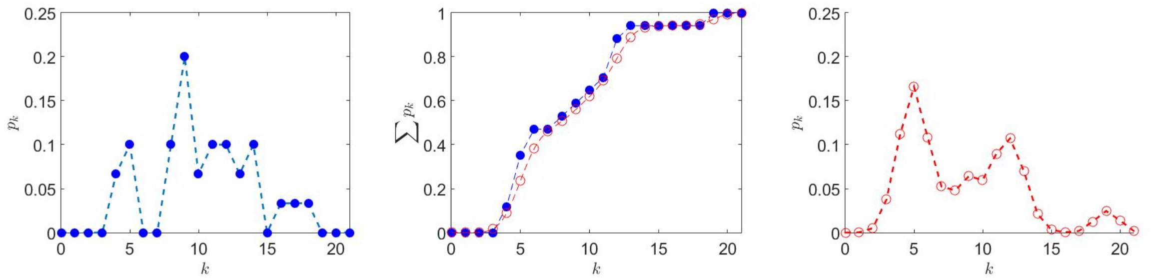
Customer Perception Analysis

The qualitative analysis of the interaction resulted in the classification of the customer perception about the psychologist as either competent or non-competent. Here we detail how we arrived to such a conclusion in both cases.

Psychologist Perceived as Competent: When customers were satisfied with the interaction, they perceived the psychologist as: kind, friendly, well mannered, polite and attentive. The interaction produced on customers a feeling of being understood in a friendly environment, which was associated with empathy. Besides, they estimated that the psychologist was highly skilled because he was giving clear explanations, had self-confidence and was able to listen and respond appropriately. All these factors generated in customers a confidence that the psychologist could help them solve the problem, and increased their intention to hire him (see Figure 6):

"*The interaction was very nice. He was very clear, he listened to me and I trust him.*" (Woman, 18 years).

Psychologist perceived as non-competent: The perception of non-competence was associated with dissatisfaction. Customers attributed their discomfort to either specific or vague reasons (see Figure 7). *Specific* reasons to explain discomfort, mostly, were derived from the



(a) Frequency-based estimation of the mass probability function (b) Cumulative probability distribution for functions in (a) and (c) (c) Smoothed estimation of the mass probability distribution

Fig. 5. Estimation of the mass probability function. Frequency-based estimation of the mass probability function (a) may be improved by smoothing the cumulative probability distribution (b). The resulting mass probability distribution draws information from neighbor cells to improve estimation.

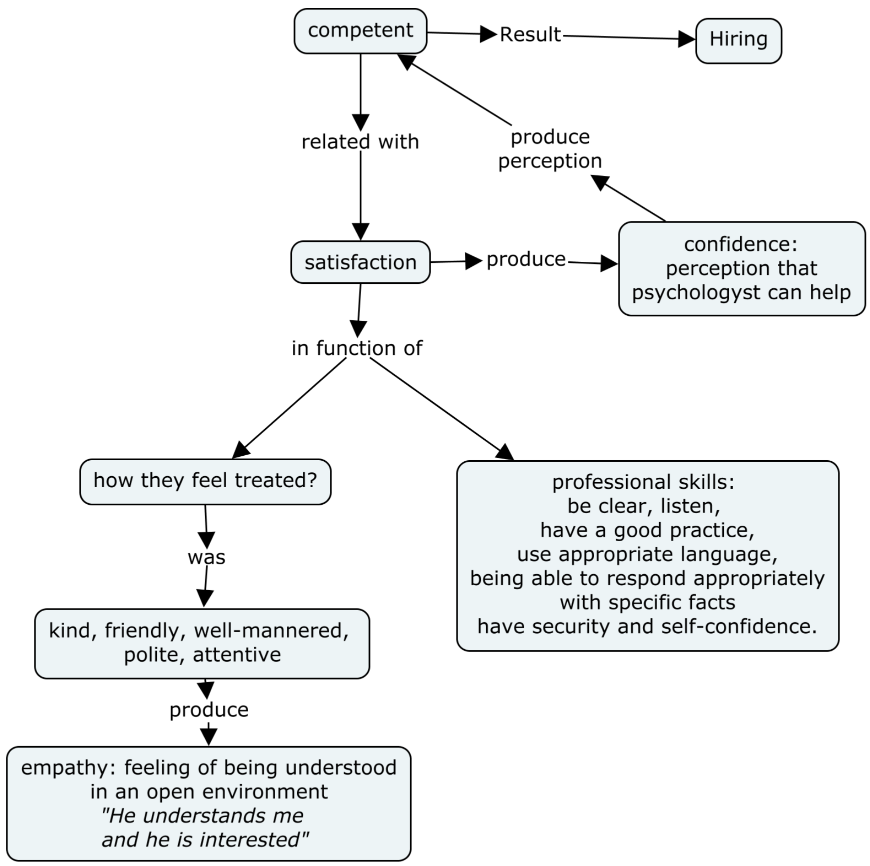


Fig. 6. Attributions to service provider when evaluated as **competent**

perception of the psychologist's lack of experience to meet professional skills. Dissatisfied customers required from the psychologist to: give more explanations, be more attentive, be more empathetic and inquire more deeply on the context of the problem. We found that all these perceptions can lead the customers to believe that the psychologist had no real interest to help, and hence no vocation. Even more, his performance in the interaction could produce a suspicion that the psychologist was just looking for an economical profit:

"*I felt that he was listening just out of obligation, it seems that he was there just for the money.*" (Man, 19 years).

Some other specific reasons to explain discomfort were associated with the deficient service offered by the psychologist. Customers expected to be invited for beverages, to be asked their names, to be offered a seat, to be introduced, and for the psychologist to give his professional background:

"*I have my doubts because I need to know better this psychologist, and even I have to know some other psychologists, besides; I would like to know what he has studied.*" (Man, 21 years).

However, some other customers were not able to relate their discomfort to a specific issue. They were clear that the interaction was not good, but they could not identify the reason; it was *vague*. Hence, they hesitated in identifying, for example, whether their dissatisfaction was due to the physical setting, the psychologist's clothing, if they were nervous about the experiment, or something else. However, they were certain of their discomfort. As a result, dissatisfied customers distrusted the proposed advice and were reluctant to hire the psychologist:

"*I wanted to be convinced since I opened the office door, I needed something to leave me with the feeling 'I have to be here'. I do not* *know if it was the office or the person [psychologist], but something in there told me 'You can find someone else'.* " (Man, 19 years).

Mirroring and its Relationship with Competence Assessment

In our experiments, there were two issues we investigated. First, we wanted to find out whether the number of nods or mirroring events could be the basis to distinguish if the customer perceived the psychologist as competent or non-competent. If so, then we wanted to find out which one was a better indicator, the number of nods or the number of mirroring events; whether it is better to use static or wearable cameras, and whether the gestures displayed by the customer are better to distinguish the classes than the psychologist's or vice versa. We performed 48 interviews where a confederated psychologist interacted with people acting as customers. Our computer vision system captured images and obtained the number of nods and the number of times the interlocutors mirrored a nodding for each interview. Afterwards, a professional psychologist (one of the authors of this paper) talked with the acting customers and analyzed qualitatively this conversation. As a result, we have a set of pairs $S = \{(x, \mathbb{C})_i\}$, for $i = 1, \ldots, 48$, where x reflects either the number of nods or the number of mirroring events and $\mathbb{C}$ is the label assigned to the conversation. Using the procedure described previously, we constructed a mass probability distribution, for each of the classes, based on the product between the likelihood and the prior. For the likelihood, we used a Gaussian smoother kernel function. Using the plug-in method described by Chu et al. (2015), we dynamically computed the bandwidth at each iteration of the leave-one-out. Typical values varied between 0.8 and 0.9 for all the configurations. For the prior, we used the ratio between the number of interviews labeled with the perception of competence (30) or non-competence (18) versus the total number of interviews (48). This resulted in a prior value of competence, $P(C_c)$, of 0.625, and a prior value of non-competence, $P(C_n)$, of 0.375. We used the area under the curve as the indicator for the performance of the classifier. As we applied the leave-one-out training strategy for the samples selected, we obtained a different performance curve for each configuration. Figure 9 illustrates the curves $\hat{P}(C_c|x)$ and $\hat{P}(C_n|x)$ obtained through this process. It includes the mean *ROC* curve and the *ROC* curves for the maximum and minimum *AUC* .

Then we performed a balanced design, double factor, repeated measures *ANOVA* analysis to test the null hypothesis that the observed means were realizations of the same underlying process. To avoid confounding, nodding and mirroring were analyzed independently. In both cases the factors to study are either the person observed (customer or psychologist) or the source of the images for the computer vision system (wearable or fixed cameras). For the case of nodding, with an F-value of 0.84, and $p = 0.36$, we accept the null hypothesis that both fixed and wearable cameras are equally effective to assess competence. In addition, with an F-value of 18.37 and a level of significance $p = 0.0$, we also notice that we obtain a different level of performance when we observe either the customer or the psychologist. Nonetheless, with an F-value of 0.18 and a level of significance $p = 0.67$, we notice that there is not a statistically significant difference in performance between the type of imaging sensor used, whether the subject observed was the customer or the psychologist. On the other hand, for the case of mirroring, with an F-value of 6.85 and a level of significance $p = 0.01$, we rejected the null hypothesis and noticed that we obtain a different level of performance when taking images with either the wearable or the fixed camera. In addition, with an F-value of 78.15 and a level of significance $p = 0.0$, we also conclude that we obtain a different level of performance when we observe either the customer or the psychologist. Finally, with an F-value of 5.68 and a level of significance $p = 0.02$, we note that there is a statistically significant difference between observing with either a fixed or wearable camera and the performance we obtain when the analyzed behavior corresponds to observing the psychologist or the customer mirroring.

TABLE I
TUKEY-KRAMER'S *MCP* FOR THE OBSERVATIONS CORRESPONDING TO NODDING AND MIRRORING. THE CELLS SHOW THE SIGNIFICANCE LEVEL OF THE INTERACTION, p. $^*p < 0.05$ ARE HIGHLIGHTED IN **bold**.

	n_c^w	n_c^f	n_p^w	n_p^f
n_c^w		0.78	**0.00***	**0.00***
n_c^f			0.08	**0.03***
n_p^w				0.99

(a) Nodding

	m_c^w	m_c^f	m_p^w	m_p^f
m_c^w		0.99	**0.00***	**0.00***
m_c^f			**0.00***	**0.00***
m_p^w				**0.00***

(b) Mirroring

Figure 8 illustrates the relative position of the means with one standard deviation segment at each side. After rejecting the null hypothesis, there is the need to know which means have a statistical significant difference. Table I shows the result we obtained, with a level of significance up to hundredths, when applying the Tukey-Kramer's *MCP* to the datasets related to nodding and mirroring.

DISCUSSION

Our experimental results emphasize the importance of nodding/mirroring during social interaction. The attributes related to the detected competence of a service provider in a dyadic interaction seem to be highly related to these gestures. When the customers received less nodding/mirroring, they had a tendency to feel dissatisfied and they frequently could not identify the reason for that feeling. As a consequence, they started a reasoning process to justify their discontent. For instance, they ascribed their dissatisfaction to the psychologist's lack of skills or professional experience. Even when every participant received an explanation along the same lines and equal advise, the dissatisfied customers had a tendency to require the service to be broadened (*e.g.,* asking for clearer explanations, improvement of the physical settings). They had doubts on the professional advice and ultimately on the psychologist's competence. In other words, they distrusted the service provider. Hence, it appears that gestures such as nodding, or actions such as mirroring, can affect to a large extent the interpretation of what is being said. Nodding/mirroring can help to interact more smoothly and with a comfortable feeling, which impacts the perception of competence.

Even more, when people receive less nodding/mirroring, the event turns out into different intentions of actions and negative attributions. These gestures and events not only impact the perception of the interaction and the people we interact with; they influence the intention of future actions like whether hiring the service or not. Besides, we found that customers who received less nodding/mirroring could assign negative attributions to the psychologist such as having no real interest to help and no vocation. It appears that the customers experiencing discomfort, due to lack of nodding/mirroring, could arouse a suspicion that the psychologist was just looking for an economic profit.

This study stress out that nodding/mirroring gestures can impact the majority of the dimensions proposed by Epstein & Hundert (2002). People exposed to less nodding conditions was usually more dissatisfied with: the psychologist's knowledge (cognitive dimension), his technical performance and proposed advise (technical and integrative dimensions), and his communication skills (relationship dimension). In the affective domain the customers referred a lack of caring and emotional bonding from the psychologist, which is related to the dimension of habits of mind —the willingness, patience, and

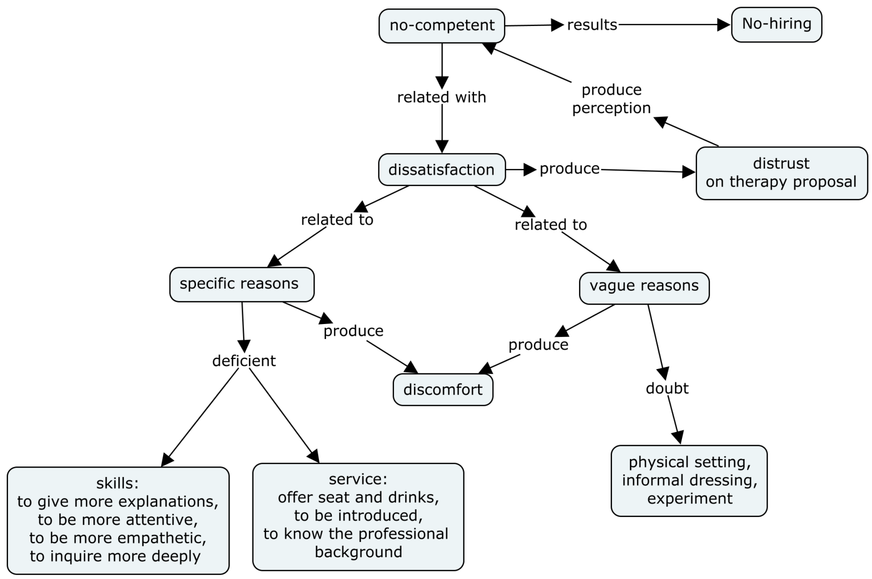
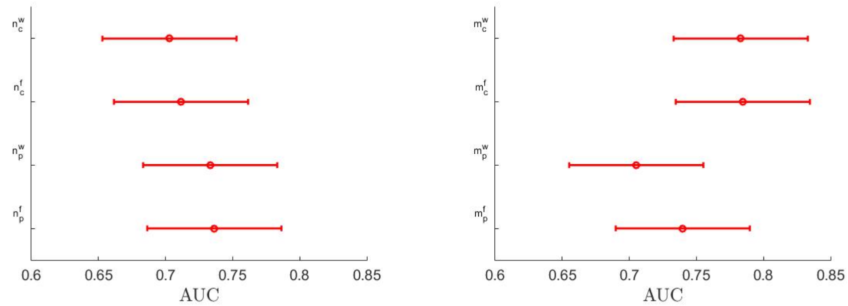


Fig. 7. Attributions to service provider when evaluated as **non-competent**



(a) Nodding (b) Mirroring

Fig. 8. Illustration of the *AUC* intervals. For each configuration we show the mean value (circle) and one standard deviation (segment extremes).

emotional awareness to use the professional skills judiciously and humanely.

With respect to the quantitative analysis of the results, although it remains to be seen whether a different classifier can perform better, this research shows that a simple Bayesian scheme could be used for this task. Using the data collected, we estimated the underlying probabilistic mass distribution using smoothing. In fact, it has been argued that nonparametric smoothing can be beneficial for cases similar to ours, where the data is sparse. For instance, Bishop et al. (2007) showed that the estimators obtained through smoothing are often better than proportions under squared error assessment. The smoothing process via sampling of the cumulative distribution and the estimation of the kernel bandwidth dynamically are well-established procedures. Our selection of the *AUC* as the performance criteria aims to stress the importance of improving the detection rate and reducing the missing rate. The leave-one-out highlighted the sensitivity of the scheme. Nonetheless, in all cases, for all configurations, the *AUC* resulted well above the value of 0.5, *i.e.,* the resulting classifier gives results above random decisions. This may be important for schemes such as boosting, where the aim is to construct complex ensembles of classifiers.

The *ANOVA* analysis provided further insight and gives statistical confidence to assess that not all the configurations, *i.e.*, nods/mirrorings, customer/psychologist and wearable/fixed, resulted in the same level of performance. In fact, it highlights that the number of mirroring events is a better classification predictor than the number of nods. Even more, customer mirroring is a better predictor than psychologist's mirroring. On the other hand, although there is an edge on the psychologist's number of nods as a better predictor that the customer's number of nods, the difference is small. In addition, there seems to be no difference on whether the camera used in the computer vision system was wearable or fixed as the analysis of the behavior of the customer or the psychologist provided the same level of performance to assess the perceived competence. An exception is made on the analysis of the mirroring of the psychologist where the fixed camera seems to provide a slightly better performance.

Conclusion

In this paper, we have introduced an automatic socially-aware assistant (based on smartglasses) for the automatic inference of competence during face-to-face social meetings. Our main hypothesis was to see if it is possible to predict whether the interlocutor has been perceived as competent or as non-competent, using the automatically detected nodding and mirroring events. In our study we have shown that it is indeed possible to infer, with a level of confidence significantly well above random chance, the resulting perception of competence acquired by an acting customer after an interaction with a service provider. Furthermore, our results have also shown that: (1) mirroring is a better predictor than nodding; (2) customer mirroring is a better predictor than psychologist mirroring;

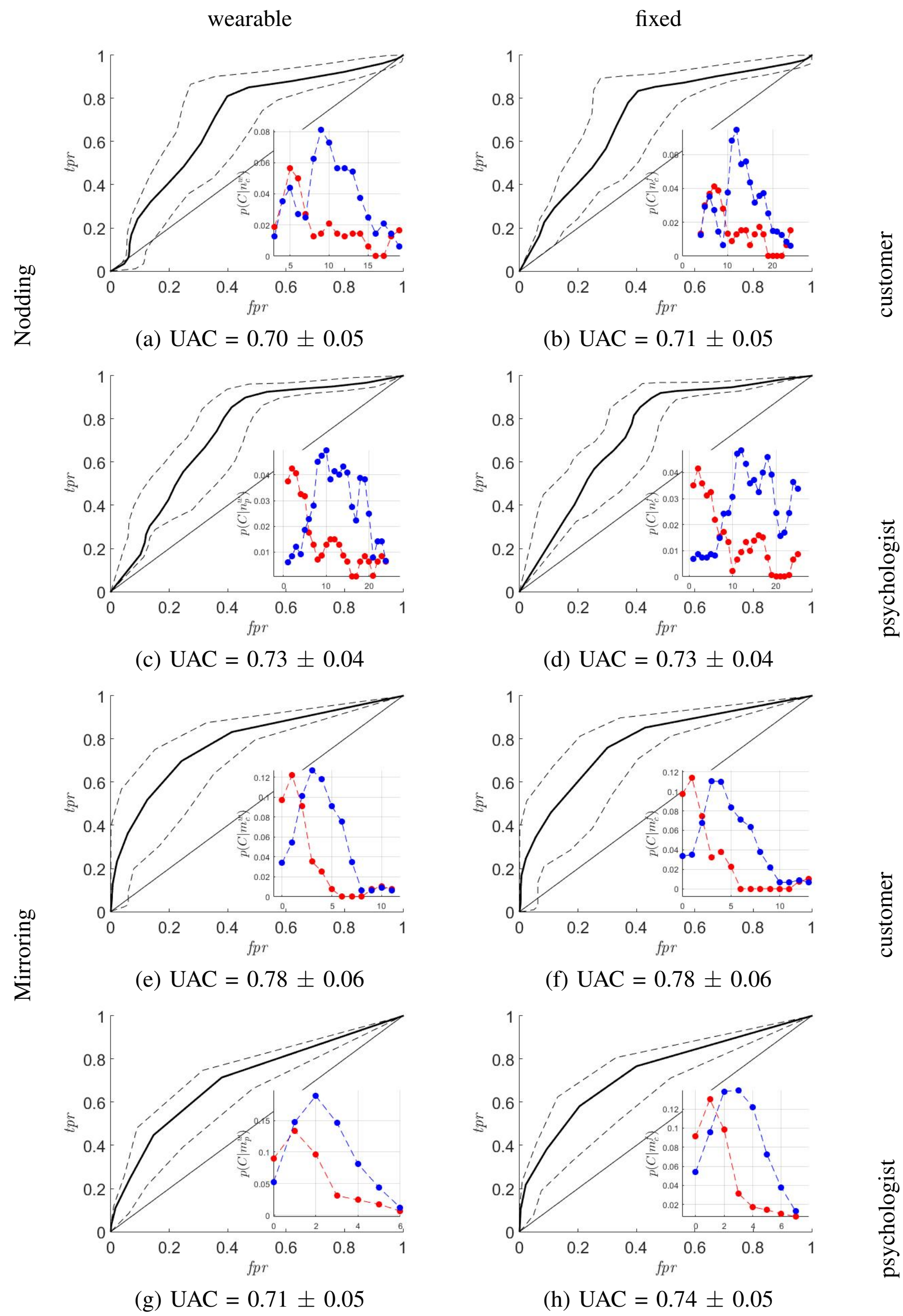


Fig. 9. *ROC* Curves for the configurations in the experiment. The dashed *ROC* curves correspond to the maximum and minimum *AUC* and are the result of leave-one-out cross-validation. The subfigure caption states the mean and standard deviation for the *AUC* . The inserted subfigure represents the probability mass distribution when all the observations are considered. The lines between the dots are drawn to improve readability.

but (3), contrariwise, the number of psychologist's nods is a better predictor than the number of customer's nods. The computer vision algorithm we used worked about equally whether it was acquiring images from wearable smartglasses or fixed cameras.

Our scenario resembled a face-to-face interaction between a customer and a service provider. To validate our hypothesis, we used a representative number of participants. Therefore, for the people taking part in the experiment, our results show a significant degree of correlation between the observations and the perceived perception of competence. Although the system was tested on a user-defined scenario, the proposed socially-aware assistant could serve as a training tool for psychologists (for automatic annotation/summarization of videos). In addition, it could be used as a supportive tool by people affected by a visual impairment to improve their social integration, *e.g.*, the assistant could inform the user when his counterpart performs a head nodding, and thus offering him the opportunity to respond to the gesture. Future research will be devoted to extend the set of social signals that our assistant is able to recognize and to test it on scenarios tailored from the real world.

References

Aitchison, J. & Aitken, C. (1976). Multivariate Binary Discrimination by the Kernel Method. *Biometrika*, *63*(3), 413–420.

Allwood, J. & Cerrato, L. (2003). A Study of Gestural Feedback Expressions. In *Nordic Symposium on Multimodal Communication*, (pp. 7–22).

Ambady, N. & Rosenthal, R. (1992). Thin Slices of Expressive Behavior as Predictors of Interpersonal Consequences: A Meta-Analysis. *Psychological Bulletin*, *111*(2), 256.

Bishop, C. (2007). Pattern Recognition and Machine Learning.

Bishop, Y., Fienberg, S., & Holland, P. (2007). *Discrete Multivariate Analysis: Theory and Practice*. Springer Science & Business Media.

Bradley, A. (1997). The Use of the Area Under the *ROC* Curve in the Evaluation of Machine Learning Algorithms. *Pattern recognition*, *30*(7), 1145–1159.

Breck, E. (1998). Soundscriber.

Caneel, R. (2005). Social Signaling in Decision Making. In *Master Thesis*. MIT Press.

Charmaz, K. (2008). *Constructing Ground Theory. A Practical Guide through Qualitative Analysis*. Sage.

Chartrand, T. & Bargh, J. (1999). The Chameleon Effect: The Perception-Behavior Link and Social Interaction. *Journal of Personality and Social Psychology*, *76*(6), 893–910.

Chu, C., Henderson, D., & Parmeter, C. (2015). Plug-in Bandwidth Selection for Kernel Density Estimation with Discrete Data. *Econometrics*, *3*(2), 199–214.

Condon, W. & Ogston, W. (1971). Speech and Body Motion Synchrony of the Speaker-Hearer. In *The Perception of Language* (pp. 150–184). Charles E. Merrill.

Cuddy, A., Fiske, S., & Glick, P. (2008). Warmth and Competence as Universal Dimensions of Social Perception: The Stereotype Content Model and the BIAS Map. *Advances in Experimental Social Psychology*, *40*, 61–149.

Curhan, J. & Pentland, A. (2007). Thin Slices of Negotiation: Predicting Outcomes from Conversational Dynamics within the First 5 Minutes. *Journal of Applied Psychology*, *92*(3), 802–811.

Epstein, R. & Hundert, E. (2002). Defining and Assessing Professional Competence. *Journal of the American Medical Association*, *287*(2), 226–235.

Farley, S. (2014). Nonverbal Reactions to an Attractive Stranger: The Role of Mimicry in Communicating Preferred Social Distance. Journal of Nonverbal Behavior. *Journal of Nonverbal Behavior*, *38*(2), 195–208.

Fawcett, T. (2006). An Introduction to *ROC* Analysis. *Pattern Recognition Letters*, *27*(8), 861–874.

Feese, S., Arnrich, B., Troster, G., Meyer, B., & Jonas, K. (2012). Quantifying Behavioral Mimicry by Automatic Detection of Nonverbal Cues from Body Motion. In *IEEE International Conference on Social Computing*, (pp. 520–525).

Flavio, L., Carneiro, M., Madge, J., Young, M., Meadows, D., Johnson, G., Fienup, J., Quance, D., & Schaller, N. (2010). Publication Manual of the American Psychological Association. In *Reports, Results and Recommendations from Technical Events Series*, number C30-56.

Freeman, W. & Roth, M. (1995). Orientation Histograms for Hand Gesture Recognition. In *International workshop on automatic face and gesture recognition*, volume 12, (pp. 296–301).

Gifford, R., Ng, C., & Wilkinson, M. (1985). Nonverbal Cues in the Employment Interview: Links between Applicant Qualities and Interviewer Judgments. *Applied Psychology*, *70*(4), 729–736.

Gladwell, M. (2006). *The Tipping Point: How Little Things Can Make a Big Difference*. Little, Brown.

Glaser, B., Strauss, A., & Strutzel, E. (1968). The Discovery of Grounded Theory; Strategies for Qualitative Research. *Nursing Research*, *17*(4), 364.

Guéguen, N. (2011). *Psychologie du consommateur: pour mieux comprendre comment on vous influence*. Dunod.

Gueguen, N., Jacob, C., & Martin, A. (2009). Mimicry in Social Interaction: Its Effect on Human Judgment and Behavior. *European Journal of Social Sciences*, *8*(2), 253–259.

Gueguen, N., Martin, A., & Meineri, S. (2011). Mimicry and Helping Behavior: An Evaluation of Mimicry on Explicit Helping Request. *Journal of Social Psychology*, *151*(1), 1–4.

Hadar, U., Steiner, T., & Rose, C. (1985). Head Movement during Listening Turns in Conversation. *Nonverbal Behavior*, *9*(4), 214–228.

Hernández, R., Fernández, C., & Baptista, P. (2010). Metodología de la Investigación. *Mc Graw Hill*.

Jacob, C., Guéguen, N., Martin, A., & Boulbry, G. (2011). Retail salespeople's mimicry of customers: Effects on consumer behavior. *Journal of Retailing and Consumer Services*, *18*(5), 381–388.

Jacob, C., Guéguen, N., Martin, A., & Boulbry, G. (2011). Retail Salespeople's Mimicry of Customers: Effects on Consumer Behavior. *Journal of Retailing and Consumer Services*, *18*(5), 381–388.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning*, volume 112. Springer.

Knapp, M. & Hall, J. (2009). *Nonverbal Communication in Human Interaction*. Cengage Learning.

Kokonendji, C. & Kiesse, T. (2011). Discrete Associated Kernels Method and Extensions. *Statistical Methodology*, *8*(6), 497–516.

Lepri, B., Subramanian, R., Kalimeri, K., Staiano, J., Pianesi, F., & Sebe, N. (2012). Connecting Meeting Behaviour with Extraversion: A Systematic Study. *IEEE Transactions on Affective Computing*, *3*(4), 443–455.

Madan, A., Caneel, R., & Pentland, A. (2005). Voices of Attraction. In *Proceedings. of Augmented Cognition*. see TR584, http://hd.media.mit.edu.

McGovern, T., Jones, B., & Morris, S. (1979). Comparison of Professional versus Student Ratings of Job Interviewee Behavior. *26*(2), 176–179.

Muhr, T. (2004). *Atlas.ti: The Knowledge Workbench*. Scientific Software Development.

Nguyen, L., Marcos-Ramiro, A., Marrón, M., & Gatica-Perez, D. (2013). Multimodal Analysis of Body Communication Cues in Employment Interviews. In *International Conference on Multimodal Interfaces*, (pp. 437–444).

Paxton, A. & Dale, R. (2013). Frame-Differencing Methods for Measuring Bodily Synchrony in Conversation. *Behavior Research Methods*, *45*(2), 329–343.

Pentland, A. (2007). Social Signal Processing. *IEEE Signal Processing Magazine*, *24*(4), 108–111.

Raducanu, B. & Gatica-Perez, D. (2012). Inferring Competitive Role Patterns in Reality TV Show through Nonverbal Analysis. *Multimedia Tools and Applications*, *56*(1), 207–226.

Russell, S. & Norvig, P. (1995). Artificial Intelligence: A Modern Approach. *Prentice-Hall*, *25*, 27.

Sanchez-Cortes, D., Aran, O., Mast, M. S., & Gatica-Perez, D. (2010). Identifying Emergent Leadership in Small Groups using Nonverbal Communicative Cues. In *International Conference on Multimodal Interfaces*. Article 39.

Saville, D. (2015). Multiple Comparison Procedures: Cutting the Gordian Knot. *Agronomy Journal*, *107*(2), 730–735.

Staiano, J., Lepri, B., Aharony, N., Pianesi, F., Sebe, N., & Pentland, A. (2012). Friends don't Lie - Inferring Personality Traits from Social Network Structure. In *Ubicomp*, (pp. 321–330).

Stoline, M. (1981). The Status of Multiple Comparisons: Simultaneous Estimation of All Pairwise Comparisons in One-Way *ANOVA* designs. *The American Statistician*, *35*(3), 134–141.

Strauss, A. & Corbin, J. (2002). *Bases de la Investigación Cualitativa: Técnicas y Procedimientos para Desarrollar la Teoría Fundamentada*. Universidad de Antioquia Medellín.

Swets, J., Dawes, R., & Monahan, J. (2000). Better Decisions through Science. *Scientific American*, *283*, 82–87.

Terven, J., Raducanu, B., Meza-de Luna, M. E., & Salas, J. (2016). Head-gestures mirroring detection in dyadic social interactions with computer vision-based wearable devices. *Neurocomputing*, *175*, 866–876.

Van Baaren, R., Holland, R., Kawakami, & Van Knippenberg, P. (2004). Mimicry and Prosocial Behavior. *Psychological Science*, *15*(1).

Van Baaren, R., Holland, R., Steenaert, B., & Knippenberg, V. (2003). Mimicry for Money: Behavioral Consequences of Imitation. *Journal of Experimental Social Psychology*, *39*(4), 393–398.

Vinciarelli, A., Pantic, M., & Bourlard, H. (2009). Social Signal Processing: Survey of an Emerging Domain. *Image and Vision computing*, *27*(12), 1743–1759.

Wagner, P., Malisz, Z., & Kopp, S. (2014). Gesture and Speech in Interaction: An Overview. *57*, 209–232.

Weng, C.-Y., Chu, W.-T., & Wu, J.-L. (2007). Movie Analysis based on Roles' Social Network. In *International Conference on Multimedia and Expo*, (pp. 1403–1406).

Xiong, X. & Torre, F. (2013). Supervised descent method and its applications to face alignment. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, (pp. 532–539).

Zougab, N., Adjabi, S., & Kokonendji, C. (2013). A Bayesian Approach to Bandwidth Selection in Univariate Associate Kernel Estimation. *Journal of Statistical Theory and Practice*, *7*(1), 8–23.