

Automatic Understanding of Affective and Social Signals by Multimodal Mimicry Recognition

Xiaofan Sun¹, Anton Nijholt¹, Khiet P. Truong¹, and Maja Pantic^{1,2}

¹ Human Media Interaction, University of Twente
PO Box 217, 7500 AE Enschede, The Netherlands

² Department of Computing, Imperial College
180 Queen's Gate, London SW7 2AZ, UK
{x.f.sun,k.p.truong,a.nijholt}@ewi.utwente.nl,
m.pantic@imperial.ac.uk

Abstract. Human mimicry is one of the important behavioral cues displayed during social interaction that inform us about the interlocutors' interpersonal states and attitudes. For example, the absence of mimicry is usually associated with negative attitudes. A system capable of analyzing and understanding mimicry behavior could enhance social interaction, both in human-human and human-machine interaction, by informing the interlocutors about each other's interpersonal attitudes and feelings of affiliation. Hence, our research focus is the investigation of mimicry in social human-human and human-machine interactions with the aim to help improve the quality of these interactions. In particular, we aim to develop automatic multimodal mimicry analyzers, to enhance affect recognition and social signal understanding systems through mimicry analysis, and to implement mimicry behavior in Embodied Conversational Agents. This paper surveys and discusses the recent work we have carried out regarding these aims. It is meant to serve as an ultimate goal and a guide for determining recommendations for the development of automatic mimicry analyzers to facilitate affective computing and social signal processing.

Keywords: Mimicry detection, mimicry expression information extraction, mimicry behavior encoding, mimicry expression classification, affective computing, social signal processing, embodied conversational agent.

1 Introduction

People's behaviors show diversity in personality, culture, growing-up background, social norms and expectations. Due to these differences, sometimes, a harmonious communication does not last long between two or more persons, especially when they support different opinions or when they are in competition. Socially intelligent systems are capable of sensing or detecting information relevant to human social interaction and are able to make this information readily available to people, these systems have the potential to improve interactions between different people in a manner that will increase feelings of rapport, affiliation and harmony. It is considered

that, among other behavioral cues, mimicry behavior (mirroring, e.g. person A nods or smiles following person B who has nodded or smiled too) provide important clues for the analysis and investigations of human interactions, mainly in two ways. Firstly, mimicry serves as an important indicator of cooperativeness and empathy during conversation. The second is its application as a means to enriching communication. The impact of a practical technology to mediate human interactions in real time would be enormous both for society as a whole (improving business relations, cultural understanding, communication relationship, etc) and personal daily social interactions. Hence, we propose to investigate mimicry behavior in human-human interaction with the aim to build automatic mimicry analyzers and mimicry-enabled Embodied Conversational Agents (ECAs).

In the following sections, we firstly discuss mimicry and related work in social psychology. In section 3 we present our aims and objects in our research on mimicry analysis in machine understanding approach. In section 4.1, we explain how our data was collected and what visual and vocal features we analyzed. In section 4.2 we present our methods and technique to automatically recognize mimicry and analyze the concepts learned by models. Finally, in section 5, we conclude the main contribution of our research to affective computing.

2 Related Work/Background

Mimicry-related research in experimental psychology has been driven by two primary objectives. The first objective concerns collecting and detecting various types of the mimicry or synchrony behaviors by establishing the emergence of mimicry by aggregating instances of to-be-imitated behavior (e.g., face touching) during social interactions [1]. The second objective aims to examine the relationship between coordinated behavior and perceptions of social connectedness [2], that is, to find out whether there is a relationship between those behaviors and social interactions. And if so, what is the exact relationship and in which situations do what types of those behaviors occur?

Building on these two primary objectives, recent research explored whether people, without intention or awareness, “use” mimicry to their advantage [3]. People can consciously or automatically mimic the behaviors of others because their goals activate behavioral strategies and plans of action that help them pursue those goals [4], [5], [6]. Individuals can mimic many different aspects of their interaction partners, including speech patterns, facial expressions, emotions, moods, postures, gestures, mannerisms, and idiosyncratic movements [1], [3], [7], [8]. In short, previous studies in social psychology report that the more mimicry is observed, the more smoothly the interaction is perceived: mimicry enhances resonance, creates rapport and affiliation suggesting that it serves to strengthen social bonds.

Given the huge advances in computer vision and algorithmic gesture detection [9], coupled with the propensity for more and more computers to utilize high-bandwidth connections and embedded video cameras, the potential for computer agents to detect, mimic, and implement human gestures [10] and other behaviors (such as vocal cues) is quite boundless and promising. The current findings suggest that non-verbal behavior could easily be added to computer agents to improve the user's experience

unobtrusively; moreover, we propose that mimicry behaviors should be added to conversational agents to make them more social. In the following sections, we explain our plans for mimicry research in the context of affective computing and social signal understanding.

3 Aims and Objectives

We have recently witnessed significant advances not only in the machine analysis of nonverbal cues, but also in the field of affective computing and social signal understanding [11]. However, only few works have so far attempted to identify social attitudes such as interest, liking, or agreement through automatic means. In moving towards more naturalistic human-computer interactions, it is necessary to have machines that are capable of detecting and understanding social attitudes, and subsequently, are able to react adequately considering the current situation and the communicative goals. Since mimicry is one of the important behavioral indicators of social and affective attitudes in social interaction, the aims of our research include 1) the automatic recognition and understanding of nonverbal mimicry (both visual and vocal), 2) the enhancement of automatic affect recognition and social signal understanding by the use of mimicry behavior information, and 3) the generation of mimicry behavior in ECAs to improve ECA-human interaction.

4 Methodology

We present our methods to reach the goals as described earlier. For each goal, we shortly describe what we already have done to reach this goal and/or what we are planning to do.

4.1 Automatic Recognition of Nonverbal Mimicry

Our first goal is to develop a system capable of recognizing and understanding nonverbal mimicry. We first need to develop 1) a multimodal database suitable for automatic visual and vocal mimicry analysis and 2) feature extraction methods to represent mimicry in detectable visual and vocal cues. We report on our work carried out to analyze mimicry automatically.

4.1.1 Database Setup

Recent work that we have carried out concerns setting up a multimodal database of mimicry occurring in interpersonal social interactions which allows us to (1) explore and understand how astute people are in detecting mimicry in social interactions, (2) examine and annotate the implications of explicit mimicry occurrences in terms of social perceptions of the mimickers, (3) classify various mimicry and mimicry behavior forms for emotion recognition (positive and negative). The most significant distinction from current affective databases is that it contains natural dyadic behavior, including mimicry. Moreover, the experiments recorded are designed to explore the relationship between the occurrence of mimicry and human affect. The corpus is recorded using a wide range of devices including face-to-face-talking and individual

close-talk fixed microphones, individual and room-view video cameras from different views, all of which produce output signals that are synchronized with each other (Fig. 1). We are planning to make manual annotations for many different phenomena, including dialogue acts, turn-taking, affect, and some head, hand gestures, body movement and facial expression (Fig. 2). The dataset consists of 54 recordings of dyadic face-to-face interactions: 34 are discussions on a political topic, and 20 are conversations situated in a role-playing game. The subjects, 40 participants and 3 confederates, were recruited from Imperial College London; all of the participants self-reported their felt experiences. The recordings and ratings are stored in a database. Later, the corpus will be made available to the scientific community through a web-accessible database. For a more detailed description of the database, readers are referred to [12].



Fig. 1. Simultaneous views from all cameras

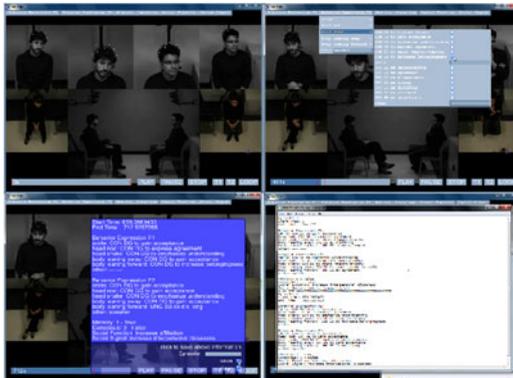


Fig. 2. Annotation tool interface includes labels

4.1.2 Multimodal Mimicry Cues Analysis

One of the most important sub-goals in our research is to extract and detect the visual and auditory features that specify temporal changes in expressions in order to measure mimicry from audio and video modalities. There have been a considerable number of studies on feature extraction for speech analysis. However, especially for prosody analysis, these features are still not well defined; for each feature there are multiple definitions and different analysis methods. Recent research has aimed at identifying

the most significant speech and visual features for affective computing but the specific contribution of each feature to affect analysis is still unclear. However, for our research goal it is not necessary to find out the contribution of each feature to specific human affect. Instead, the focus is on changes and similarities of visual and vocal behavior. For visual cues, we presented a way to detect visual mimicry in a machine understanding approach [13], [14] in which we demonstrated that in face-to-face interaction, after a while, a confederate has the tendency to take over body postures, head movements, and hand gestures of the participants with whom he/she is interacting. Specifically, we find that participants or confederates who are being mimicked are more willing to alter the way in which they interact with others to share similar affect or attitudes in order to obtain more agreement and to express understanding or similar attitudes. They did not only mimic postures, mannerisms, moods or emotions, but they also mimicked several speech related behaviors. We also demonstrated how vocal behavioural information expressed between two interlocutors can be used to detect and identify vocal mimicry [15]. As future work, we plan to investigate other visual and vocal feature representations of mimicry. The method and extracted features used in the current study are not enough to represent visual mimicry in a machine learning approach. Because the correlation of a motion intensity histogram is not reliable and stable for recognizing visual-based mimicry, we can only say, to a certain degree, that participants are moving the same body parts with a similar intensity. No details about temporal and specific expressions of various human actions can be given which are needed to represent visual mimicry.

Hence, in future work, for automatic visual-based mimicry detection, more kinematic-based features are needed such that analyses similar to those carried out for non-verbal vocal mimicry can be performed. Focus on the optical flow fields in motion parts of a body, computation of kinematic features (e.g., divergence, vorticity, symmetric flow fields etc.) and the classification of these features for recognizing mimicry will be our primary research goal to achieve the ultimate goal to assess human affect in terms of automatic mimicry analysis. With respect to non-verbal vocal mimicry, we have not looked yet at other non-verbal vocal variables such as utterance lengths and switching pause durations which are known to converge between speakers. Furthermore, in addition to prosodic vocal behavior, people may also mimic the quality of voice which can be measured through voice quality and spectral features. We will investigate the commonly used spectral features Mel-Frequency Cepstrum Components (MFCCs) in combination with speaker recognition modeling techniques to evaluate the similarity between two voices. Further, it is interesting to find out whether the repetition of vocal events such as laughter can be used as a measure for mimicry. We will also look at methods to determine the presence of non-verbal vocal mimicry more locally (rather than globally). How to combine information from various modalities, e.g., facial expressions, vocal expressions, and body movement expressions, for multimodal mimicry recognition is another interesting future research topic. The most challenging problems of multimodal mimicry recognition lies in feature extraction and the use of probabilistic graphical models when fusing the various modalities. As mimicry recognition is closely related to the field of affective computing and shares similar difficult issues, we will also put effort in solving these issues such as obtaining reliable affective data, obtaining ground truth labels, and the use of unlabeled data. Finally, we want to

understand how variables, such as personality and emotion, regulate mimicry in interaction so that automatic mimicry detection algorithms can take these into account. To that end, we will take a closer look at our data, and analyze mimicry taking into account the willingness of the participants to mimic in certain situations.

4.2 Use of Mimicry Information to Understand Affective and Social Behavior

Our second goal is to investigate the interrelationships among mimicry and various affects, attitudes, or mental states in interaction. We propose to employ a machine learning approach to model these interrelationships. In further research, we attempt to classify automatic mimicry behaviors based only on nonverbal multimodal cues. We attempt to demonstrate that it is possible to classify various mimicry behaviors (e.g. head movements, body postures, and hand gestures) according to the findings from social psychology which mimicry behaviors are relative to creating rapport, affiliation, and smoother communication. We could propose to explicitly model the complex hidden dynamics of the multimodal cues using the Hidden Conditional Random Field (HCRF), HMM, and DBN. Firstly we will show that those models are able to capture what kinds of mimicry are used in backchannel, are used for conveying social attitudes (e.g. expressing similar attitudes, acceptance, interest), giving response (e.g. listening, understanding,) or other (e.g. unsure). We could present an approach to analyze the concepts learned by the HCRF model and show that these coincide with the findings from social psychology regarding mimicry conveys similar attitudes, understanding, interest, and acceptance. We also could prepare which model performs better and which model can be automatically analyzed to identify what sets of features are the most discriminative in each class. Then we could demonstrate that what kinds of mimicry cues are the most discriminative (important) cues for identifying different classes. In those models, the parameters level (lower level) in the top of the model only contains various variables/parameters which represent the general features; all the features which are used to model the object domain can be estimated by those parameters. The structure level (higher level) of the model is constructed by the structure nodes so that more than one layers' network can be used for representing the different object domains. Finally, all the layers will be connected in the suitable structure to represent our final domain problem. In our domain problem, the highest layer in the structure level is the most abstract layer and specifies what the mental state is (agree/disagree, uncertain, like/dislike, concentrating etc.). The intermediate level (hidden states) layers describe whether and what kind of interactional mimicry exists. The lower level layers describe how to recognize the human behaviors defined in social interactions by parameter estimations and combinations or learning algorithms.

4.3 Mimicry in Embodied Conversational Agents

For our third goal, we will first explore methods and results presented in experimental social psychology that can be used to embody mimicry in agents. The role mimicry plays in social interaction will be embodied in human-agent conversation to make the agent more social and intelligent. The key to facilitate social and intelligent interaction in context between an agent and a human is to build an accurate model of

supporting communication for various individuals. A model based on knowledge about the other's group affiliation (from non-verbal behaviour, cultural display rules and so on) and based on the individual's own unique behavioural habits and tendencies (by observation and learning) will form expectations about potential meanings and intentions. The most challenging problem in generating mimicry behavior for agents is to determine in what situation and to what extent the mimicking behavior should occur.

5 Contribution

Our proposed mimicry research serves to identify, "feeling", and "understanding" behaviors of interactants in human and agent interaction with the help of mimicry behavioral information. By instantiating and manipulating psychological theories, through the use of nonverbal cues, we can endow agents with the ability to evolve in and engage in social interactions and to enhance face-to-face communication. From the automatic understanding of social signals and prediction of how these signals affect social situations, immediate applications can be derived to help people with less confidence, to train people for improved social interaction tasks, in general or specifically such as in negotiation. Hence, in sum, our contributions to affective computing and human-machine interaction contain (1) the understanding of how human mimicry works and subsequently, the development of automatic mimicry analyzers for ECAs, (2) the improvement of the recognition of social and affective attitudes such as (dis-)agreeing and (dis-)liking through mimicry information, and (3) knowledge about the timing and the extent to which mimicry should occur in human-machine interaction by generating mimicry behavior in agents. This technology would also strongly influence science and technology by, for example, providing a powerful new class of research tools for social science and anthropology. While the primary goal of such an effort would be to facilitate direct mediated communication between people, advances here will also facilitate interactions between humans and machines.

References

1. Chartrand, T.L., Bargh, J.A.: The chameleon effect: The perception-behavior link and social interaction. *Journal of Personality and Social Psychology* 76, 893–910 (1999)
2. Bernieri, F.J.: Coordinated movement and rapport in teacher-student interactions. *Journal of Nonverbal Behavior* 12, 120–138 (1988)
3. Chartrand, T.L., Jefferis, V.E.: Consequences of automatic goal pursuit and the case of nonconscious mimicry. In: Forgas, J.P., Williams, K.D., von Hippel, W. (eds.) *Responding to the Social World: Implicit and Explicit Processes in Social Judgments and Decisions*, pp. 290–305. Psychology Press, Philadelphia (2003)
4. Dijksterhuis, A.: Automatic social influence: The perception-behavior link as an explanatory mechanism for behavior matching. In: Forgas, J.P., Williams, K.D. (eds.) *Social Influence: Direct and Indirect Processes*, pp. 95–108. Psychology Press, Philadelphia (2001)
5. Dijksterhuis, A., Bargh, J.A.: The perception-behavior expressway: Automatic effects of social perception on social behavior. In: Zanna, M. (ed.) *Advances in Experimental Social Psychology*, pp. 1–40. Academic Press, San Diego (2001)

6. Gueguen, N., Jacob, C., Martin, A.: Mimicry in social interaction: Its effect on human judgment and behavior. *European Journal of Sociences* 8 (2009)
7. Lakin, J.L., Chartrand, T.L.: Using nonconscious behavioral mimicry to create affiliation and rapport. *Psychological Science* 14, 334–339 (2003)
8. Lakin, J.L., Jefferis, V.E., Cheng, C.M., Chartrand, T.L.: The Chameleon Effect as social glue: Evidence for the evolutionary significance of nonconscious mimicry. *Journal of Nonverbal Behavior* 27, 145–162 (2003)
9. Huang, H.H., Cerekovic, A., Tarasenko, K., Levacic, V., Zoric, G., Pandzic, I.S., Nakano, Y., Nishida, Y.: Integration embodied conversational agent components with generic framework. *Multiagent and Grid Systems-An Interational Journal* 4, 371–386 (2008)
10. Kopp, S.: Social resonance and embodied coordination in face-to- face conversation with artificial interlocutors. *Speech Communication* 52(6), 587–597 (2010)
11. Pantic, M., Nijholt, A., Plentland, A., Huang, T.S.: Human-Centred Intelligent Human-Computer Interaction (HCI²): how far are we from attaining it? *Journal of Autonomous and Adaptive Communications Systems* 1(2), 168–187 (2008)
12. Sun, X.F., Lichtenauer, J., Valstar, M., Nijholt, A., Pantic, M.: A Multimodal Database for Mimicry Analysis. In: D’Mello, S., et al. (eds.) *ACII 2011, Part I*, vol. 6974, pp. 367–376. Springer, Heidelberg (2011)
13. Sun, X.F., Truong, K., Nijholt, A., Pantic, M.: Automatic Visual Mimicry Expression Analysis in Interpersonal Interaction. In: *Fourth IEEE Workshop on CVPR for Human Communicative Behavior Analysis, Held in Conjunction with CVPR 2011, Colorado Springs, USA (to appear, 2011)*
14. Sun, X.F., Nijholt, A., Pantic, M.: Towards Mimicry Recognition during Human Interactions: Automatic Feature Selection and Representation. In: *Proceedings 4th International ICST Conference on Intelligent Technologies for Interactive Entertainment (INTETAIN 2011), Genoa, Italy, May 25-27 (to appear, 2011)*
15. Sun, X.F., Truong, K.P., Pantic, M., Nijholt, A.: Towards Visual and Vocal Mimicry Recognition in Human-Human Interactions. In: *Proceedings 2011 IEEE International Conference on Systems, Man, and Cybernetics (IEEE SMC 2011), Special Session on Social Signal Processing, Anchorage, Alaska, October 9-12 (to appear, 2011)*