Towards Mimicry Recognition during Human Interactions: Automatic Feature Selection and Representation

Xiaofan Sun[1], Anton Nijholt[1], 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,a.nijholt}@ewi.utwente.nl. m.pantic@imperial.ac.uk

Abstract. During face-to-face interpersonal interaction people have a tendency to mimic each other, that is, they change their own behaviors to adjust to the behavior expressed by a partner. In this paper we describe how behavioral information expressed between two interlocutors can be used to detect and identify mimicry and improve recognition of interrelationship and affect between them in a conversation. To automatically analyze how to extract and integrate this behavioral information into a mimicry detection framework for improving affective computing, this paper addresses the main challenge: mimicry representation in terms of optimal behavioral feature extraction and automatic integration.

Keywords: mimicry representation, human-human interaction, human behavior analysis, motion energy.

1 Introduction

Mimicry plays an important role in human-human interaction. Mimicry refers to the coordination of movements in both timing and form during interpersonal communication. Behavior matching, synchronized changes in behavior and facial expressions, matching in posture and mannerisms are examples of mimicry. But there can also be vocalic mimicry and matching of verbal style. Mimicry is ubiquitous in daily interpersonal interaction. For example, when two interactants are facing each other and one of them takes on a certain posture such as moving sideways or leaning forward, then the partner may take on a congruent posture [1], [2], [12], and when one takes on certain mannerism such as rubbing the face, shaking the legs, or foot tapping, the partner may take on a congruent mannerism [2]. Another example, if one is crossing his legs with the left leg on top of the right, the other may also cross his legs with the right leg on top of the left leg (called "mirroring") or with the left leg on top of the right leg (called "postural sharing").

A. Camurri, C. Costa, and G. Volpe (Eds.): INTETAIN 2011, LNICST 78, pp. 160–169, 2012.

Mimicry enhances social interaction by establishing rapport and affiliation [2] and by observing mimicry behavior conclusions can be drawn about the quality of the interaction and about interpersonal relationships between conversational partners. For that reason mimicry has become object of study of social psychology. What behavioral cues show mimicry, how to rate mimicry, and what different kinds and functions of mimicry can be distinguished are among the main questions that are studied. Mimicry, as it can be perceived from facial expressions, vocal behavior, and body movements, affects human-human interaction.

It is interesting to look at a possible role of mimicry in human-computer interaction. It is well known that humans can consider computers as social actors and in particular in agent-oriented interfaces designers anticipate such behavior. Moreover, we see more applications where the role of the computer is not so much to be efficient or only efficient, but also being social or entertaining, for example in health and well-being situations where the computer plays a coaching function, in domestic situations where a social robot needs to be trusted in order to accept his help and advice, and, of course in gaming and entertainment applications where we play and communicate with virtual humans (avatars, embodied conversational agents, ...). More human-like behavior of a virtual human allows for more natural interaction and modeling mimicry makes it possible to understand and generate mimicry behavior in human- virtual human or human-social robot interaction.

Many researchers from psychology have investigated mimicry. Until now, research in affective computing has been concerned with the affective role of facial expressions, body postures, gaze directions, prosody, and (neuro-)physiological information. But, the role of mimicry in human-human interaction and how this role can be exploited in human-machine interaction (where, machine can be a computer, a robot, a virtual human, an environment, et cetera) to improve the interaction and the experience, has not been explored. It requires automatic (machine) detection of mimicry, automatic understanding of mimicry, automatic prediction of mimicry, and also automatic generation of mimicry. And, obviously, then the role of mimicry in human-human interaction should be completely understood.

In current and future game and entertainment environments we will meet people. Their characteristics and their behavior will not always be fully mediated. There will probably be a lack of subtle social signals that play important roles in human-human interaction and that are hard to mediate. Our research aims at understanding these subtle social signals, in particular mimicry, in order to mediate them in human-nonhuman interactions. This will help improving natural interaction (in natural situations) and establishing interpersonal relationships that people would like to have and maintain, whether it is with a human or with a social and intelligent human-like device. Mimicry is an informative and communicative act that helps to convey and recognize intentions and affect that are important for interaction and establishing relationships.

In our experiments on the role of mimicry in social interaction we have conversational partners that are being observed in a laboratory setting. Data such as location, body orientation, head pose, gestures, and vocal activities is obtained from camera and audio input. Behavioral patterns are analyzed to detect people's relationships, individuals’ affect and assessing the quality of the interaction.

In this paper, reporting about work in progress, we show that we can find and represent behavioral mimicry in conversations by analyzing human actions in prediction models. In section 2 we have some observations on factors affecting mimicry. A short description of the corpus that we collected for mimicry analysis is presented in section 3. A more comprehensive description will appear elsewhere. The corpus is used for extracting and detecting of features for mimicry recognition. We shortly discuss our annotation steps and the automatic extraction of mimicry episodes. In section 4 we present some preliminary conclusions, including the conclusion that automatic mimicry identification is possible.

2 Mimicry to Be Expected in Social Interaction

The first and most important aspect in this study is to collect data which includes various behavioural mimicry or interactional synchrony in social interactions. The social interaction scenarios that aim at elicitation of behavioural mimicry or interactional synchrony need to be natural in terms of the different factors that may affect the likelihood or increase the chance of mimicry occurring. However, the factors that affect mimicry are not unique and they cannot account for everything. We illustrate this with a few examples. For example, in daily life, when we talk with our boss, we mimic his or her behavior or repeat what he or she said. Not necessarily because you really agree with him or her, but there may be a desire to affiliate for personal benefits and even without awareness. Moreover, when we share similar opinions in a meeting, we also have a strong tendency to mimic other members' behaviors in an attempt to gain acceptance. In some cases, there is a strong mimicry tendency because of directly active goals, sometimes we mimic to improve a harmonious interrelationship, but usually we mimic without consistent awareness. Mimicry occurs in our daily life all the time, and most of the time this mimicry behavior signals important social attitudes and affects.

Mimicry is sensitive to social context, so automatic mimicry behavior changes according to one's active goals in a realistic social situation. Mimicry responses are modulated by the social signal value of the behavior. That is, many social signals may be implied or signaled by various mimicry behaviors. For example, expressive behavior is more present in conversations about positive experiences than in conversations about negative experiences. And usually participants are more active and more willing to show facial expressions and body language when they seem to be familiar with the topic. That is, they show their opinions, both verbally and non-verbally, more actively when they are familiar with the topic. Hence, to choose a topic which is familiar with both interactants is important for collecting more mimicry episodes. Behavioral mimicry plays an important role in identifying interactants' attitude, affect and even roles played in conversations. Previous studies showed a higher mimicry tendency when people perceived themselves as similar, would like to be similar, or want to display themselves as similar [10]. In addition, they may have aligned goals [8] and lean forward, they may share attitudes [11] and lean forward and nod, they may like their conversational partner and show it by synchronous head nodding and shaking [8], they may want the other to have a positive perception of them and display matching smiles [7], or empathize with the other and show this in matching behavior [6]. Moreover, mimicry also helps in identifying the roles people play in a conversation, for example, people always expand themselves unconsciously when they are perceived as dominant, however, constrict themselves when they are perceived as submissive [15], [16].

Thus, in our experiments a first scenario designed for collecting behavioral mimicry and interactional synchrony is about discussing a familiar topic that makes it possible to share attitudes with each other. In the second scenario, given that most participants in our experiments are students, we give a hypothetical conversational topic which is familiar with their actual daily life. They are given a non-task-oriented communication assignment which requires self-disclosure and emotional discovery.

3 Experiment Setup

For extracting and detecting of features for mimicry recognition in our prediction model, we used a corpus of 53 human-to-human interactions. This corpus is described in Section 3.1. Section 3.2 is devoted to the description of the features that are annotated in our experiments to be used for mimicry representation. Section 3.3 presents the algorithm used for tracking mimicry in terms of the features annotated. Finally Section 3.4 discusses our methodology for automatic mimicry extraction.

3.1 Data Collection

Our data is drawn from a study of face-to-face discussions and conversations. 43 subjects from Imperial College, London participated in this experiment. They were recruited using the Imperial College social network and were compensated 10 pounds for one hour of their participation.

The experiment included two sessions. In the first session, participants were asked to choose a topic from a list, which had several statements concerning that topic. Participants were then asked to write down whether they agree or disagree with each statement of their chosen topic. Participants were then asked to present their own stance on the topic, and then to discuss the topic with their partners, who may have different views on the topic. Participants could talk about anything they wanted, that is, the statements we listed were just a reference. In the second session, the intent is to simulate a situation where participants wanted to get to know their partner a bit better and they needed to disclose personal and possibly sensitive information about themselves. Participants were given a non-task-oriented communication assignment that required self-disclosure and emotional discovery. Participant A played a role as a student in university who was looking for a room to rent urgently. Participant B played a role as a person who owns an apartment and wants to let one of the rooms to the other person.

We collected synchronized multimodal data for each session. In each session we recorded data from the participants separately and from the two participants together, including voice and body behaviors. In the visual-based channel we recorded data using 7 cameras for each person and 1 camera for both persons at the same time. The camera for both persons was used for recording an overview of the interaction, while the other 7 cameras were used for recording the two participants separately, including far-face view, near-face view, upper-body view, and whole body view with and without color. See Fig. 1 for some camera views. Both participants wore a lightweight and distance-fixed headset with microphone. For detecting head movements both participants wore rigs on their heads during recording. The rig is a lightweight, flexible metal wire frame and fitted with 9 infrared LEDs. Given the face location and orientation, the nine LEDs allow us to get detailed information about the characteristics of the head movements.

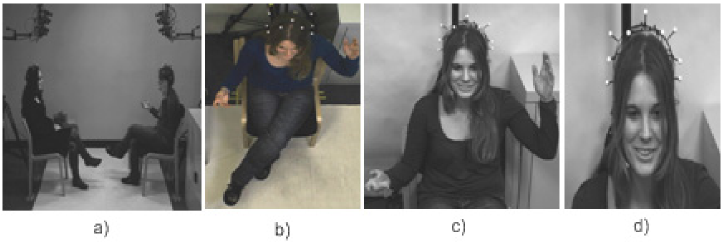

Fig. 1. a) Setup for corpus collection. b) Higher-view for whole body recording for each participant separately. c) Recording upper-body movement. d) Recording head movements and facial expressions.

3.2 Annotation

As discussed in the previous section, the corpus is a collection of face-to-face interactions designed with the aim to study mimicry behavior and interactional synchrony. Hence, the main focus of the annotation scheme is the labeling of the behavior expressions and in particular behavioral mimicry.

The annotators' job is to look at videos of these interactions and annotate them with information about the "human behavioural expressions" and "social signals" of the participants. This means that they continuously try to answer the questions "How the actions of those participants display: is he/she nodding, head shaking, etc.?" and "Do they mimic each other?"

For each annotation assignment, the main annotation steps are based on widely accepted concepts of mimicry. Firstly, mimicry is dynamic, hence, signals of mimicry behavior occur successively. Secondly, mimicry is about one conversational partner imitating the other [1]. That is, mimicry is when people express or share similar behavior during interaction, at the same time or one after another, in response to the other.

The main annotation steps are briefly introduced below:

1. Annotation of speakers and listeners (usually listeners and speakers take turns) based on the utterances.

2. Segmentation into episodes, where each episode consists of a sequence speaker1, listener2, speaker2, listener1, hence, each of the two participants appears in the sequence both as speaker and as listener.
3. Annotation of visual-based behavioral expressions for the two partners such as smile, nod, head shake, hand gesture, and body leaning.
4. Annotation of mimicry cues: we have predefined notions of behavioural cues; after manually annotating episodes and behavioural cues, we use an algorithm (see below) to automatically compare whether the selected notions match or not; if they match label mimicry (YES), if not, label mimicry (NO).

Hence, after the first step of annotation, the utterance token of a participant is labeled as listener or speaker. In the second step, we select the conversation segments in such a way that each participant is seen as a speaker and a listener, because their (amount of) mimic behavior can be dependent on their role in the conversation (speaker or listener). Then, in the third step, behaviors expressed by participants are labeled, using visual cues, for analyzing behavioral mimicry. Finally, in terms of mimicry perception we annotate those behaviors expressed by paired participants as mimicry or not. After annotating conversation segments and visual cues for detecting mimicry, based on these annotation results we extract mimicry episodes. In each mimicry episode visual cues are extracted to identify behavioural mimicry. This will be discussed in more detail in section 3.3.

Algorithm to automatically extract mimicry episodes

```
Given: i: current episode index;
SB[i]: the array of mimicry cues displayed by the
speaker during the current (ith) episode;
SB_N[i]: the total number of mimicry cues
displayed by the speaker during the current (ith)
episode;
LB[i]: the array of mimicry cues displayed by the
listener during the current (ith) episode;
LB_N[i]: the total number of mimicry cues
displayed by the listener during the current (ith)
episode.

Detect speaker's mimicry:
For each frame t
Do int SB[1]=0 if SB1<SB_N[i] and apply ++SB[1]);

Mimicking the previous episode's speaker:
  For (int SB[2]=0; SB[2]<SB_N[i-1]; ++SB[2])
  If (SB[i][SB1] == SB[i-1][SB2]), and label
mimicry;
```

```
Mimicking the current episode's listener:
  For (int lb1=0; lb1<SB_N[i-1]; ++lb1)
  If   (SB[i][sb1]   ==   LB[i][lb1]),   and   label
mimicry;

Detect  listener's  mimicry  (only  consider  current
round)
  For(int lb1=0; lb1<LB_N[i]; ++lb1)
  For(int sb1=0; sb1<SB_N[i]; ++sb1)
      If(LB[i][lb1]   ==   SB[i][sb1]),   and   label
mimicry.
```

3.3 Methodology

In this section, we first describe the human action recognition technique we use to extract motion features and represent the motion cycle [19] for identifying behavioral mimicry. Then, by analyzing our results, we show that in our annotated mimicry episodes, mimicry indeed occurs more frequently. Moreover, we investigate that similarity is indeed an important factor that increases mimicry. In this study we only annotated the episodes on one aspect of similarity, That is, the role participants play in a conversation. In fact, similarity was manipulated in various ways in previous studies: status, appearance, attitudes, sport interests, leisure interests, et cetera.

We calculated the motion cycle in each manually annotated episode in our attempt to detect behavioural mimicry. The motion cycle is extracted in terms of the accumulated or averaged motion energy (AME) which only is computed in areas that include changes [16], [19]. Hence we propose to represent the motion cycle by computing a group of accumulated motion images (AMIs). In detail, AMI represents the time-normalized accumulative and average action energy and contains pixels with intensity values for representing motions [21]. In the AMI, the regions containing pixels with higher intensity values denote that motions are more complex and occur more frequently. Although AMI is related to MEI and MHI [19], a fundamental difference is that AMI describes the motions by using the pixel intensity directly. That is, instead of giving all equal weights for all changing areas in MEI or assigning higher weights for new frames but lower weights for older frames in MHI.

$$AMI(x, y) = \frac{1}{T}\sum_{t=1}^{T}\left|D(x, y, t)\right| \tag{1}$$

where $D(x, y, t) = I(x, y, t) - I(x, y, t-1)$ in which T denotes the length of the query action video (i.e., total number of frames) and I stands for the intensity of the current frame. Fig 2 illustrates visual behavioural mimicry, extracted from consecutive sets of frames of a recording.

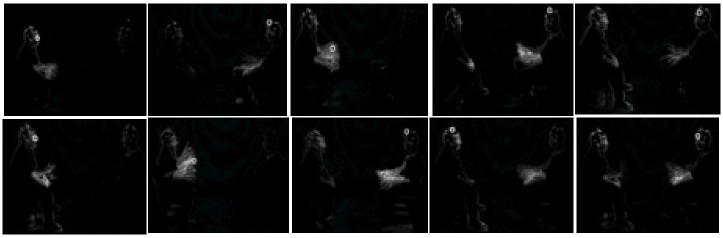

Fig. 2. A group of behavioural mimicry extracted from consecutive sets of frames (frame 92, 96, 98, 102, 103, 105, 108, 113, 120, and 123) of a recording in our database

The figure illustrates hand gesture mimicry behavior. This behavior is visualized by presenting the results of motion intensity calculation for hands movement. Motion cycle images are calculated by AMI in several successive frames for each annotated mimicry behavior in our data.

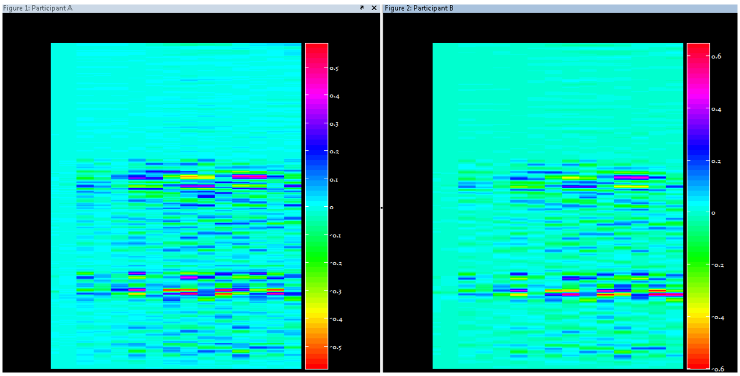

Fig. 3. The cross-correlations of body movements between two persons who interact with each other. The vertical axis shows the motion energy, the horizontal axis shows the frame numbers.

Fig 3 demonstrates the cross-correlations of movements between two persons, generated from a fragment of 580 windows (20 sec) in a conversation on looking for a suitable roommate. The vertical axis shows the motion energy, the horizontal axis shows the frame numbers. The left part of the figure shows the motion energy calculated in each frame for participant A; the right part shows the motion energy calculated in each frame for participant B.

Summarizing, in Fig. 2 we demonstrate that visual-based mimicry can be visually extracted in a short time period of around 5 seconds in our data. In Fig. 3, we accumulate all movements during a longer period (20 sec) to see the general motion tendency expressed by two people who interact in a conversation. We can see rather similar cross-correlations of body movements between conversational partners. Hence, we can safely assume that behavioral mimicry probably occurs with a high chance in this period.

4 Conclusions and Future Work

Our results show that behavioral information from conversational partners can be extracted and integrated in order to demonstrate mimicry. Moreover, it became clear that mimicry is indeed ubiquitous in human-human conversation. Methods to analyze motion energy can be applied and improved to deal with mimicry in a machine understanding approach [18]. From our mimicry episode annotation we have learned about the role of similarity, that is, the similarity of roles played in interactions. Moreover, mimicry analysis does contribute to recognizing the role of affect and empathy in social interaction.

For future work, we plan to extract relevant features from audio and visual channels for detecting more mimicry cues in our database. The aim is to automatically identify mimicry when people mimic facial expressions, vocal productions, and body movements with their conversational partners or others around them in daily interaction. In affective computing research the detection of nonverbal cues has been considerably improved in previous years. The role of verbal and nonverbal expressions has been investigated, including their necessity for understanding behavioural patterns, mental states, attitudes and personality traits. It has also been demonstrated that people tend to mimic and synchronize vocal utterances during a conversation. Usually, people with different personalities probably prefer different interaction tempos. In a conversation, if the communication goes well or is improving, the speech cycles of conversational partners become mutually entrained. The study of vocalic mimicry is not so much to find out the attribution of each feature of speech, such as spectral features or non- spectral features to specific human affect, but the focus is rather on the changing of and the similarity of speech utterances. Including vocalic mimicry is a next step in our research on modeling mimicry.

Acknowledgments. We gratefully acknowledge the help of Michel Valstar, Jeroen Lichtenauer, and many others from Imperial College who helped to realize the experiments for the data collection. This work has been funded in part by the European Community's 7th Framework Programme [FP7/2007Ű 2013] under the grant agreement no 231287 (SSPNet). The work of Maja Pantic is also funded in part by the European Research Council under the ERC Starting Grant agreement no. ERC-2007-StG-203143 (MAHNOB).

References

1. Bernieri, F.J.: Coordinated movement and rapport in teacher student interactions. Journal of Nonverbal Behavior 12(2), 120–138 (1998)
2. Bernieri, F.J., Reznick, J.S., Rosenthal, R.: Synchrony, pseudosynchrony, and dissynchrony: Measuring the entrainment process in mother-infant interactions. Journal of Personality and Social Psychology 54(2), 243–253 (1988)
3. Chartrand, T.L., van Baaren, R.: Chapter 5 Human Mimicry, pp. 219–274. Academic Press (2009)

4. Chartrand, T.L., Bargh, J.A.: The chameleon effect: the perception-behavior link and social interaction. Journal of Personality and Social Psychology 76(6), 893–910 (1999)
5. Chartrand, T.L., Jefferis, V.E.: Consequences of automatic goal pursuit and the case of nonconscious mimicry, pp. 290–305. Psychology Press, Philadelphia (2003)
6. Chartrand, T.L., Maddux, W., Lakin, J.L.: Beyond the perception behavior link: The ubiquitous utility and motivational moderators of nonconscious mimicry, pp. 334–361. Oxford University Press, New York (2005)
7. Giles, H., Powesland, P.F.: Speech style and social evaluation. Academic Press, London (1975)
8. Gueguen, N., Jacob, C., Martin, A.: Mimicry in social interaction: Its effect on human judgment and behavior. European Journal of Sciences 8(2), 253–259 (2009)
9. Hess, U., Blairy, S.: Facial mimicry and emotional contagion to dynamic emotional facial expressions and their influence on decoding accuracy. Int. J. Psychophysiology 40(2), 129–141 (2001)
10. Jefferis, V.E., van Baaren, R., Chartrand, T.L.: The functional purpose of mimicry for creating interpersonal closeness. Manuscript, The Ohio State University (2003)
11. Kopp, S.: Social resonance and embodied coordination in face-to- face conversation with artificial interlocutors. Speech Communication 52(6), 587–597 (2010)
12. LaFrance, M.: Nonverbal synchrony and rapport: Analysis by the cross-lag panel technique. Social Psychology Quarterly 42(1), 66–70 (1979)
13. Lakin, J.L., Chartrand, T.L., Arkin, R.M.: Exclusion and nonconscious behavioral mimicry: Mimicking others to resolve threatened belongingness needs (2004) (manuscript)
14. 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(3), 145–162 (2003)
15. Miles, L.K., Nind, L.K., Macrae, C.N.: The rhythm of rapport: Interpersonal synchrony and social perception. Journal of Experimental Social Psychology 45(3), 585–589 (2009)
16. Briassouli, A., Kompatsiaris, I.: Robust temporal activity templates using higher order statistics. IEEE Transactions on Image Processing 18(12), 2756–2768 (2009)
17. Chandrashekhar, V.H., Venkatesh, K.S.: Action energy images for reliable human action recognition. In: Proceedings of the Asian Symposium on Information Display (ASID 2006), pp. 484–487 (October 2006)
18. Nagaoka, C., Komori, M., Nakamura, T., Draguna, M.R.: Effects of receptive listening on the congruence of speakers' response latencies in dialogues. Psychological Reports 97, 265–274 (2005)
19. Bobick, A.F., Davis, J.W.: The recognition of human movement using temporal templates. IEEE Transactions on Pattern Analysis and Machine Intelligence 23(3), 257–267 (2001)
20. Shechtman, E., Irani, M.: Matching local self-similarities across images and videos. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2007), Minneapolis, Minn, USA, pp. 1–8 (June 2007)
21. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005), San Diego, Calif, USA, vol. 1, pp. 886–893 (June 2005)