Towards Visual and Vocal Mimicry Recognition in Human-Human Interactions

Xiaofan Sun*, Khiet P. Truong*, Maja Pantic†*, and Anton Nijholt*
*Electrical Engineering, Mathematics and Computer Science
University of Twente, Enschede, The Netherlands
Email: x.f.sun@ewi.utwente.nl
†Department of Computing, Imperial College, London, UK

Abstract—**During face-to-face interpersonal interaction, people have a tendency to mimic each other. People not only mimic postures, mannerisms, moods or emotions, but they also mimic several speech-related behaviors. In this paper we describe how visual and vocal behavioral information expressed between two interlocutors can be used to detect and identify visual and vocal mimicry. We investigate expressions of mimicry and aim to learn more about in which situation and to what extent mimicry occurs. The observable effects of mimicry can be explored by representing and recognizing mimicry using visual and vocal features. In order to automatically analyze how to extract and integrate this behavioral information into a multimodal 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 in both audio and video modalities.**

Index Terms—**visual mimicry, vocal mimicry, interpersonal synchrony, mimicry representation, human-human interaction, human behavior analysis.**

I. 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 switches in behavior and facial expressions, matching in posture and mannerisms are examples of visual mimicry. 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"). But there can also be vocal mimicry. The findings that people influence and mimic each other's communicative speech behavior are integrated in the so-called Communication Accommodation Theory (or Speech Accommodation Theory), first formulated by [23],[22]. In this theory, two main categories of accommodation can be identified. Firstly, interlocutors may change their speech behaviors in opposite directions which is called speech divergence. Contrarily, when interlocutors change their speech behavior to become more similar to one another, speech convergence occurs which can also be described as vocal mimicry. The mimicking of vocal behavior can occur on two different levels: verbal and non-verbal. Vocal mimicry on the verbal level may involve repetitions of words, expressions or whole utterances (e.g., [26]). Vocal mimicry on the non-verbal level may involve matching of speech rates and rhythms (e.g.,[27], [28]), utterance lengths (e.g., [29]), latencies of responses (e.g., [28]), intonation and pitch contours (e.g., [30]), accent (e.g., [22]), pause durations (e.g., [31]), and vocal intensity levels (e.g., [32]).

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 an object of study in social psychology. What behavioral cues show mimicry, how to rate mimicry, 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 (e.g., 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. However, although the discovery of mimicry phenomena has been done by many psychologists, automatic mimicry detection and prediction, let alone generation, is still an unexplored issue in the affective computing community. In this paper, reporting about work in progress, we show that we

978-1-4577-0653-0/11/$26.00 ©2011 Crown

can find and represent behavioral mimicry in conversations by analyzing human actions and human vocal behavior. A short description of the corpus that we collected for mimicry analysis is presented in section II. A more comprehensive description will appear elsewhere. The corpus is used for extracting and detecting of features for mimicry recognition. Section III presents a method and features to analyse visual mimicry automatically. A method for a global representation of non-verbal vocal mimicry is presented in section IV. Finally, we discuss the results and future research in section V.

II. Mimicry corpus

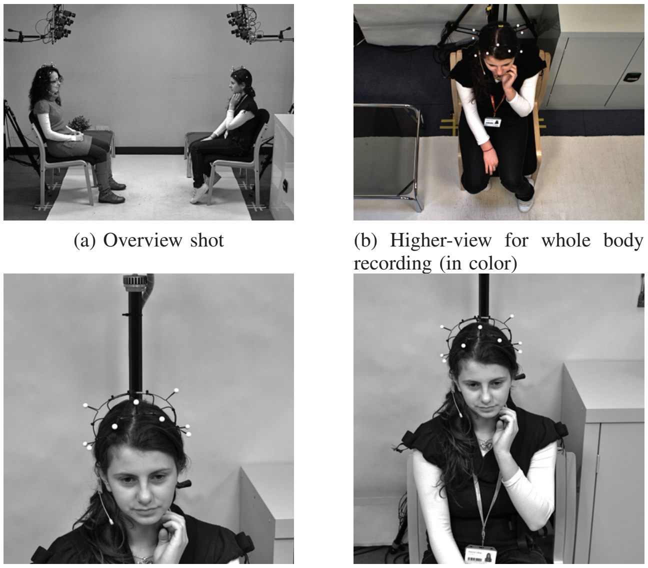
In this section, we present the audiovisual corpus that we developed specifically for mimicry research. We describe the experimental setup, how the recordings were made and what annotations have been made.

A. Experimental setup

Our data is drawn from a study of face-to-face discussion and conversation. 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 first asked to present their own stance on the topic (we will refer to this episode as the 'presentation' episode), and then to discuss the topic with their partners (we will refer to this episode as the 'discussion' episode), who may have different views about the topic. Participants could talk about anything they wanted, that is, the statements we listed are 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 (we will refer to this episode as the 'conversation' episode). 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.

B. Recordings

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. 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 shapes of the head movements. Some still images of the recordings are shown in Fig. 1.



(a) Overview shot

(b) Higher-view for whole body recording (in color)

(c) Near-face view

(d) Upper-body view

Fig. 1: Various camera views in the mimicry corpus

C. Annotations

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 "What actions do the 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 a temporal phase and the mimicry behavior should occur successively. Secondly, mimicry is one individual doing what another individual does [1]. In summary, 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 two participants 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 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 because of 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 present behavioural mimicry.

III. Towards visual mimicry detection

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 attempt to demonstrate that in our annotated mimicry episode, mimicry indeed occurs more frequently. Moreover, we investigate that similarity is indeed an important factor of increasing mimicry. In this study we only annotated the episodes on one aspect of similarity, which is the role participants played in the conversation. In fact, similarity was manipulated in various ways in previous studies: status, appearance, attitudes, sport interests, leisure interests, et cetera.

A. Features

We calculated the motion cycle in each manually annotated episode in our attempt to pursue behavioural mimicry. The motion cycle is extracted in terms of the accumulated or averaged motion energy (AME) which only is computed in areas including 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 occur more complex and 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 instead of giving all equal weights for all changing areas in MEI or assigning higher weights for new frames but lower weights always for older frames in MHI.

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

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 a group of visual behavioural mimicry which is extracted in consecutive sets of frames of a recording.

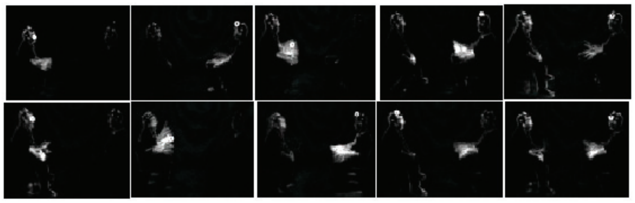


Fig. 2: A group of behavioural mimicry extracted from consecutive sets of frames (frame 92, 96, 98, 102, 103, 105, 108, 113, 120, and 123) of a recording in our database.

B. Visual mimicry representation

Visual mimicry and interpersonal synchrony refers to the coordination of movement between individuals in both timing and form during interpersonal communication. Many researchers have been interested in investigating the nature of these phenomena and have introduced theories explaining these phenomena in social psychology. Because of this broad range of theoretical applicability, social mimicry and interpersonal synchrony has been measured in many different ways [1],[2], which can be divided into two types: behavior coding and ratings.

Previous studies have illustrated similarities and differences between a coding method and a rating method for measuring interpersonal synchrony. These studies examined how people use objective cues (as measured by the coding method) to judge rapport, or how people use subjective cues (as measured by the rating method) when they perceive interpersonal communication (e.g., [33], [34]). However, how to measure mimicry and interpersonal synchrony in a machine learning approach is still not explored. In this paper, we attempt to present some hand gesture mimicry behavior by presenting the calculation results of motion intensity for hands movement. Fig. 2 shows a set of images that visualizes some annotated gesture mimicry in which the body parts and tendency of motion can be observed. All those motion cycle images are calculated by AMI in several successive frames for each annotated mimicry behavior in our data. 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,

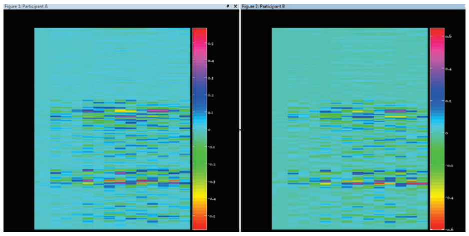


Fig. 3: The cross-correlations of body movements between two persons who interacted with each other.

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 seconds) 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.

IV. Towards non-verbal vocal mimicry detection

One of the subgoals of our mimicry research is to identify and detect vocal cues that can be used for the automatic measurement and detection of non-verbal vocal mimicry. Non-verbal vocal characteristics are typically expressed through prosodic means such as intonation, intensity, and rhythm in speech. Since we expect non-verbal vocal mimicry to express itself through speech rate, intonation, and vocal intensity as described in the literature, we analyze vocal mimicry in terms of pitch, vocal intensity, and speech rate. Although there is no uniform approach to measuring prosody, the features selected have shown to correlate well to intonation, speech rate, and vocal intensity. In this section, we describe how these features were extracted, and how we determine the presence of non-verbal vocal mimicry.

A. Features

The first acoustic parameter involved in this study is pitch which is a commonly used prosodic feature in studies of emotion detection (e.g., [36]). Pitch was measured in Hertz and computed for each frame using the autocorrelation method in Praat [35]. Another feature is the intensity of the speech signal, or its energy. Energy in time was calculated as the square of the signal at each point. A smoothed curve of the energy, using average of the energy over a 40ms time frame, was also calculated. Energy can be computed in various ways, here we prefer Root-Mean-Square (RMS) energy which is extracted from a frame. Finally, we used speech rate which can be calculated in different ways [24], for example, the number of syllables per second [25] or the number of voiced frames in which the energy is above a threshold divided by the number of words in the utterance. In this work, we calculate the speech rate as the voiced part of the speech signal divided by the length of the signal. In sum, we analyze vocal mimicry based on pitch, intensity and speech rate. All features were automatically extracted from the speech signal using Praat [35].

B. Method

We evaluate the presence of mimicry in our data by comparing pairs of interactant behaviors. For now, we are interested in a global representation of non-verbal vocal mimicry. Hence, we compare and look at changes over time in the interactants' speech behaviors. This approach allows us to draw interactant behavior data from different episodes in the data recorded, and compare these in a meaningful way as is shown in Table I. We make comparisons for each individual pair of participant and confederate separately but in order to be able to make more general statements about the mimicry behavior found in our corpus, we will also look at averaged data combined from all pairs.

Following the comparison scheme shown in Table I, we calculated correlations between the speech patterns of the participants in different episodes and compared these correlations to each other, see Table I. By using this comparison scheme, we attempt to demonstrate that participants change their own speech style while talking to the confederate, and moreover, mimic their interlocutors' speech style. 'Participant in presentation' serves as the participant's baseline speech behavior. In this scheme, one would expect that at some point, correlation (A) decreases (the participant in the discussion will adapt to its interlocutor's speech behavior), while correlation (B) increases (the confederate's speech behavior will become more similar to the participant's speech behavior).

TABLE I: Comparison scheme

Compare between			
(A) Correlation between		(B) Correlation between	
participant in presentation	participant in discussion	participant in discussion	confederate in discussion

C. Results

Fig. 4 shows the curves for both correlations, averaged over all pairs of participants and confederates, which are (A) the correlation between the participants' performances during the presentation and discussion episodes (solid line), and (B) the correlation between the participants' performances during the discussion and the confederates' performances during the discussion episode (dashed line). Correlation curves of two random individual pairs of participants and confederates are also presented in Fig. 5. Most of the curves of the individual pairs indeed look like Fig. 5. Since we want to be able to make general statements about the results obtained, we concentrate on the correlation curves that were computed and averaged over all pairs in the corpus. Firstly, correlation (A) in Fig. 4 shows mimicking behavior of the participant. The degree of vocal mimicry of the participant was measured during the discussion period relative to the presentation period which is a baseline period. It was found that compared to this baseline period, participants adjusted their utterances more frequently to increase similarity with the confederate's vocal

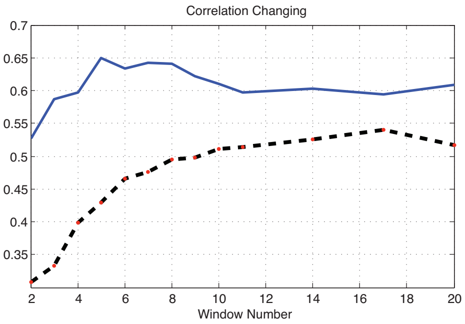
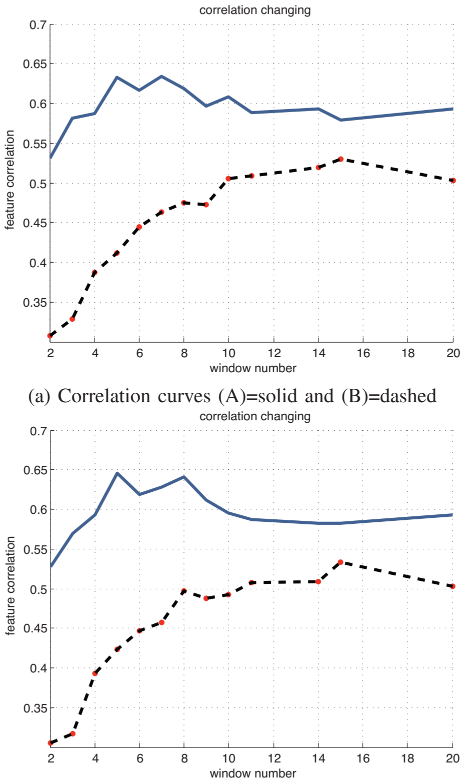


Fig. 4: Correlation curves (A)=solid and (B)=dashed showing a general and global tendency to mimick each other's speech behavior - based on averaged correlations of all pairs of participants and confederates

utterances when sharing the same attitude or opinions. In addition, Fig. 4 demonstrates two apparent changing tendencies of the correlations. We observe that 1) up to window number 8, both correlations (A) and (B) are increasing, 2) between window number 8 and 17 correlation (A) is decreasing while (B) is increasing, and 3) correlation (A) is increasing while (B) is decreasing. These tendencies, which show that non-verbal vocal mimicry has indeed occurred, can be explained as follows. During phase 1, correlation (A) is not decreasing because the participant is expected to have similar speech styles at the beginning of the presentation and discussion. Correlation (B) is increasing because the confederate is expressing his/her opinions during discussion and while doing that, makes his/her speech behavior look more similar to that of the participant. During phase 2, correlation (B) is still increasing: the participant and the confederate are still mimicking each other. However, correlation (A) shows a decrease because after a period of time in the discussion, the participant has picked up some of the speech characteristics of the confederate and starts to sound more similar to him/her while expressing his/her opinions. Finally, phase 3 occurs at the end of the discussion in which the participant and confederate are both more willing to express their own opinions and they do so by residing back to their own speech styles, knowing that the end of the discussion is approaching.



(a) Correlation curves (A)=solid and (B)=dashed

(b) Correlation curves (A)=solid and (B)=dashed

Fig. 5: (a) and (b) show global tendencies of mimicking speech behavior of two random pairs of participants and confederates

V. Conclusions and Future Work

Our results show that behavioral information from conversational partners can be extracted and integrated in order to measure mimicry. Moreover, it became clear that both visual and vocal mimicry are indeed ubiquitous in human-human conversation. We attempted to present motion mimicry, visualized in Fig 2. We compute motion intensity between consecutive pairs of frames of the video resulting in a three-dimensional stack of optical flow fields, where the third dimension is time, and extract motion period from the input frames by calculating MEIs to show motion parts of a body and the visualized motion tendency. Moreover, in Fig. 3 we present the cross-correlations of body movements between two persons who interacted with each other in order to illustrate that the similarity of body movements between paired persons is present. That is, accumulated motion image (AMI) was computed by using input images differences to represent the spatiotemporal features of occurring motion, which is based on average motion energy computing. Subsequently, AMI was computed in meaningful areas which contain changes and motions instead of the whole silhouette of human body extracted in motion cycle periods. Mimicry also showed to be apparent in the vocal behavior of the conversation partners. By analyzing the changing tendency of the vocal features extracted, we demonstrated that people change their vocal style while interacting with others, moreover, the changing is to adjust to mimic each others' vocal behavior. The conversation partners' prosodic behaviors, represented by pitch, energy, and speech rate features, were shown to converge during the dialogue, see Figs. 4, 5.

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 stabile 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.

Acknowledgment

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 [FP7/2007-2013] 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] F.J. Bernieri, : *Coordinated movement and rapport in teacher student interactions*. Journal of Nonverbal Behavior, vol. 12, no. 2, pp. 120–138, 1998.

[2] F.J. Bernieri , J.S. Reznick, R. Rosenthal : *Synchrony, pseudosynchrony, and dissynchrony: Measuring the entrainment process in mother-infant interactions*. Journal of Personality and Social Psychology, vol. 54, no. 2, pp. 243–253, 1988.

[3] T.L. Chartrand, and R. van Baaren, Chapter 5 *Human Mimicry*. Academic Press, pp. 219–274, 2009.

[4] T.L. Chartrand, J.A. Bargh : *The chameleon effect: the perception-behavior link and social interaction*. Journal of Personality and Social Psychology, vol. 76, no. 6, pp. 893–910, 1999.

[5] T.L. Chartrand, V.E. Jefferis : *Consequences of automatic goal pursuit and the case of nonconscious mimicry*. Philadelphia: Psychology Press, pp. 290–305, 2003.

[6] T.L. Chartrand, W. Maddux, J.L. Lakin : *Beyond the perception behavior link: The ubiquitous utility and motivational moderators of nonconscious mimicry*. New York: Oxford University Press, pp. 334–361, 2005.

[7] H. Giles, P.F. Powesland, : *Speech style and social evaluation*. London: Academic Press, 1975.

[8] N. Gueguen, C. Jacob, A. Martin : *Mimicry in social interaction: Its effect on human judgment and behavior*. European Journal of Sciences, vol. 8, no. 2, pp. 253–259, 2009.

[9] U. Hess, S.Blairy : *Facial mimicry and emotional contagion to dynamic emotional facial expressions and their influence on decoding accuracy*. Int J Psychophysiology, vol. 40, no. 2, pp. 129–141, 2001.

[10] V. E. Jefferis, R. van Baaren, T. L. Chartrand : *The functional purpose of mimicry for creating interpersonal closeness*. Manuscript, The Ohio State University, 2003.

[11] S. Kopp : *Social resonance and embodied coordination in face-to- face conversation with artificial interlocutors*. Speech Communication, vol. 52, no. 6, pp. 587–597, 2010.

[12] M. LaFrance : *Nonverbal synchrony and rapport: Analysis by the cross-lag panel technique*. Social Psychology Quarterly, vol. 42, no. 1, pp. 66–70, 1979.

[13] J. L. Lakin, T. L. Chartrand, and R. M. Arkin : *Exclusion and nonconscious behavioral mimicry: Mimicking others to resolve threatened belongingness needs*. Manuscript, 2004.

[14] J.L. Lakin, V.E. Jefferis, C.M.Cheng, T.L. Chartrand : *The chameleon effect as social glue: Evidence for the evolutionary significance of nonconscious mimicry*. Journal of Nonverbal Behavior, vol. 27, no. 3, pp. 145–162, 2003.

[15] L.K. Miles, L.K. Nind, C.N. Macrae,: *The rhythm of rapport: Interpersonal synchrony and social perception*. Journal of Experimental Social Psychology, vol. 45, no. 3, pp. 585–589, 2009.

[16] A. Briassouli and I. Kompatsiaris, *Robust temporal activity templates using higher order statistics*., IEEE Transactions on Image Processing, vol. 18, no. 12, pp. 2756–2768, 2009.

[17] V. H. Chandrashekhar and K. S. Venkatesh, *Action energy images for reliable human action recognition*, in Proceedings of the Asian Symposium on Information Display (ASID '06), pp. 484–487, October, 2006.

[18] C. Nagaoka, M. Komori ,T. Nakamura and M.R.Draguna : *Effects of receptive listening on the congruence of speakers' response latencies in dialogues*. Psychological Reports 97, 265–274, 2005.

[19] A. F. Bobick and J. W. Davis, *The recognition of human movement using temporal templates*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, no. 3, pp. 257–267, 2001.

[20] E. Shechtman and M. Irani, *Matching local self-similarities across images and videos*, in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '07), pp. 1–8, Minneapolis, Minn, USA, June 2007.

[21] N. Dalal and B. Triggs, *Histograms of oriented gradients for human detection*, in Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05), vol. 1, pp. 886–893, San Diego, Calif, USA, June 2005.

[22] H. Giles, *Accent mobility: A model and some data. Anthropological Linguistics*, 15, 87–105, 1973.

[23] H. Giles, *A Study of Speech Patterns in Social Interaction: Acccent Evaluation and Accent Change*,PhD Thesis, University of Bristol.

[24] Dekens, Tomas. Demol, Mike. Verhelst, Werner .Verhoeve, Piet: *A comparative study of speech rate estimation techniques*, 510-513.

[25] N. H. de Jong, and T. Wempe, *Automatic measurement of speech rate in spoken Dutch*. ACLC Working Papers, 2(2), 51–60, 2007.

[26] A. L. Gonzales, J. T. hancock, and J. W. Pennebaker, *Language Style Matching as a Predictor of Social Dynamics in Small Groups*. Communication Research, vol. 37, 3–19, 2010.

[27] J. T. Webb, *Interview synchrony: An investigation of two speech rate measures in an automated standardized interview*, in B. Pope and A.W. Siegman (Eds.), Studies in dyadic communication (pp. 115-133), New York: Pergamon, 1972.

[28] J. N. Cappella and S. Palnap, *Talk and silence sequences in informal conversations III: Interspeaker influence*, Human Communication Research, 7, 117-132, 1981.

[29] J. D. Matarazzo, A. N. Weins, R. G. Matarazzo, and G. Saslow, *Speech and silence behavior in clinical psychotherapy and its laboratory correlates*, in J. Schlier, H. Hunt, J. D. Matarazzo, and C. Savage (Eds.). Research in psychotherapy (pp. 347-394), Washington, DC: American Psychological Association, 1968.

[30] E. Couper-Kuhlen, *The prosody of repetition: On quoting and mimicry*, in: E. Couper-Kuhlen and Margret Selting, eds., Prosody in Conversation: Interactional studies, Cambridge: Cambridge University Press, 366–405, 1996.

[31] J. Jaffe and S. Feldstein, *Rhythms of dialogue*, New York, NY: Academic, 1970.

[32] M. Natale, *Convergence of mean vocal intensity in dyadic communications as a function of social desirability*, Journal of Personality and Social Psychology, 32, 790-804, 1975.

[33] J. S. Gillis, F. J. Bernieri, and E. Wooten, *The effects of stimulus medium and feedback on the judgment of rapport*, Organizational Behavior and Human Decision Processes, 63, 33-45, 1995.

[34] J. E. Grahe and F. J. Bernieri, *Self-awareness of judgment policies of rapport*, Personality and Social Psychology Bulletin, 28, 1407-1418, 2002.

[35] P. Boersma and D. Weenink, *Praat, a System for Doing Phonetics by Computer*, Glot International, 5(9/10), 341–345, 2001.

[36] R. Banse and K. R. Scherer, *Acoustic Profiles in Vocal Emotion Expression*, Journal of Personality and Social Psychology, 70, 614–636, 1996.