A Gaze-Centered Multimodal Approach to Human-Human Social Interaction

Ülkü Arslan Aydın
Cognitive Science Dept.
Middle East Technical University
Informatics Institute
06800 Ankara, Turkey
+90 312 2103741
e170938@metu.edu.tr

Sinan Kalkan
Computer Engineering Dept.
Middle East Technical University
Department of Computer Engineering
06800 Ankara, Turkey
+90 312 2105547
skalkan@metu.edu.tr

Cengiz Acartürk*
Cognitive Science Dept.
Middle East Technical University
Informatics Institute
06800 Ankara, Turkey
+90 312 2107704
acarturk@metu.edu.tr

ABSTRACT
This study aims at investigating gaze aversion behavior in human-human dyads during the course of a conversation. Our goal is to identify the parametric infrastructure, which will underlie the development of gaze behavior in Human Robot Interaction. We employed a job interview setting, where pairs (an interviewer and an interviewee) conducted mock job interviews. Three pairs of native speakers took part in the experiment. Two eye-tracking glasses recorded the scene video, the audio and the eye gaze positions of the participants. The analyses involved synchronization of multimodal data, including video recording data for face tracking, gaze data from the eye trackers, and the audio data for speech segmentation. We investigated frequency, duration, timing and spatial positions of gaze aversions relative to interlocutor's face. The results revealed that the interviewees perform more frequent gaze aversion compared to the interviewers. Moreover, gaze aversion takes longer when accompanied by speech. Also, specific speech instances, such as pause and speech-end signals have significant impact on gaze aversion behavior.

CCS Concepts
• **Human-centered computing → interaction design**

Keywords
Gaze Aversion; Mobile Eye Tracking; Job Interview task

1. INTRODUCTION
Eye contact plays a crucial role in initiating a conversation, in regulating turn-taking, in signaling topic-change and in adjusting conversational roles. Those multiple functions of eye contact depend on the conversational goals of the interlocutors. Eye contact is also the fundamental initial step for capturing the attention of the communication partner and establishing joint attention [7, 11]. Gaze aversion, complementary to eye contact, is another coordinated interaction behavior that regulates conversation. In the present study, we focus on gaze aversion mechanisms and its multimodal aspects, in particular gaze and speech, in dyadic conversations between human partners.

Gaze aversion is the act of intentionally looking away from the interlocutor. The previous research has explored the effects of gaze aversion on avoidance and approach motivations [9], stating that an averted gaze of the interlocutor initiates a *tendency to avoid*, whereas a direct gaze initiates a *tendency to approach*. This observation was based on the finding that higher ratings were given for likeability and attractiveness by the participants, when the picture stimuli involved a face with a direct gaze contact rather than an averted gaze [12, 14]. Accordingly, gaze aversion is expected to be shorter than eye contact in an efficient conversation.

In previous research, three conversational functions have been attributed to gaze aversion. (i) Intimacy modulation: The overall level of intimacy is influenced by periodic gaze aversions; (ii) Floor management: Gaze aversion occurs when the speaker takes a break by temporarily stopping the conversation during the course of speech. (iii) The speaking partner conducts more gaze aversion than the listening partner for facilitating thinking and remembering, which will eventually lead to reduced effort for paying attention to the listener [1, 4, 10].

A relevant characteristic of gaze aversions, and more generally eye movements is that they comprise a complementary modality to speech during the course of a conversation. The relationship between speech and social gaze is so intimate that it was proposed that one may generate a good enough autonomous gaze behavior by solely analyzing time intervals between sentential structures, such as turn taking, without consulting semantic analyses [16].

In the present study, we analyzed gaze aversion characteristics together with the analysis of synchronous speech data (cf. multimodal analysis). Moreover, we conducted a close investigation of speech data by focusing on particular structural types of speech instances, such as the start and the end of a speech segments (see Section 2.2) in addition to semantically burdened speech instances, such as confirmation and greeting. Thus, we performed speech data annotation both by considering structural and semantic aspects of the utterances.

We also aimed at contributing to the current research methodology by employing face tracking and eye tracking to detect gaze aversion, instead of adopting manual video coding to annotate averting gaze and the time sequences of speaking [3]. For this, we captured eye movements of both conversation partners synchronously by using eye tracker glasses in a mock job interview task, adopted from the previous studies [2, 3].

Our goal in the present study is to devise the methods, which will provide the infrastructure for the development of a parametric model of gaze aversion in human-human interaction. Therefore, we focus on regularities that may emerge during the course of interaction. The findings will then be employed to develop a performance model of gaze aversion behavior in human-robot interaction. We conceive the present study as an initial step

The following section presents the experiment design and methodology.

*Corresponding author: Cengiz Acarturk, acarturk@metu.edu.tr. Orta Dogu Teknik Universitesi Enformatik Enstitusu 06800 Ankara, Turkey.

2. EXPERIMENT

2.1 Participants, Materials and Design

Three pairs of participants took part in the experiment. All the participants had normal or corrected-to-normal vision. No time limit was introduced to the participants. The average experiment duration was approximately five minutes.

In each session, both participants wore monocular Tobii eye tracking glasses, which had a sampling rate of 30 Hz with a 56°×40° recording visual angle capacity for the visual scene. The glasses recorded the video of the scene camera and the sound, in addition to gaze data. The threshold for the IR (infrared)-marker calibration process was 80% accuracy. After the calibration, the participants were seated on the opposite sides of a table, 100 cm away from each other. A beep sound was generated to indicate the beginning of a session for data synchronization.

Eight common job interview questions, adopted from the literature [18] were presented to the interviewer. The interviewer was instructed to ask the provided questions. The interviewer was also asked to evaluate the interviewee per question by using a notebook and a pen.

2.2 Data Analysis

Data analysis consisted of three main steps, as presented below.

2.2.1 Face Tracking

Face tracking has been a challenging topic in computer vision. First, a face in a video recording is either detected automatically or labeled manually. Then it is tracked during the video stream. The tracking task is subject to a set of technical difficulties, such as variations in illumination, appearance, and scale. In the present study, we employ a state-of-the-art face tracking method, which mainly tracks the facial landmarks detected using locally constrained neural fields [5]. In the experiments, we used an open source implementation of the model, namely OpenFace[1]. The OpenFace framework includes facial landmark detection and head pose estimation. We employed the OpenFace framework to track the faces in the videos. The framework detected landmark locations on a face image (see Figure 1) at 33.33ms/frame. As shown in Figure 3, the image-frames are theoretically divided into 3x3 area-of-interests (AOIs). Each frame is labelled with the relative positions of gaze data to the face location. If the gaze data was inside the detected face, 'e' character is assigned as an AOI-label. Moreover, the face region is divided into 3 areas: namely mouth, nose and eyes. Then, which of these three areas corresponds to the gaze data is evaluated. Otherwise, if the gaze location is outside the face boundary, 1 of 8 character values, namely, 'a' 'b' 'c' 'd' 'f' 'g 'h' 'i', is assigned. We calculated the AOIs-Detection Ratio. The AOIs-Detection Ratio delivers the number and the percentage of the labelled frames. If the AOIs detection-ratio was less than 70%, then we trained a detector with dlib[2], an open source machine-learning library written in C++, and re-tracked video files with the trained detector. In cases where the detection ratio was still less than 70% after re-tracking, we reviewed all the frames and manually labelled AOIs where applicable.

2.2.2 Speech Segmentation

Audio data were extracted from the video files. The data were then converted into an input file by the CMU Sphinx4[3] libraries. The Sphinx4 libraries enabled us to obtain speech segments at a millisecond precision. We used synchronous audio recording from both participants to facilitate the segmentation process and to prevent data loss. We then manually annotated each speech segment by using a list of speech instance types, shown below.

- Pre-Speech: The non-speech conversation segment which included the silence before starting the speech and the sounds for warming up the voice.
- Speech: This conversation segment included the speech itself, as well as speech-while-laughing, asking a question, micro pauses which are shorter than 200 ms during speech and confirmation (e.g., *good*, *ok*, *huh-huh*).
- Speech Pause: This conversation segment included the pauses during the course of speech.
- Thinking: We named the conversation segment as *thinking*, when it included filler sounds, such as *uh*, *er*, *um*, *eee*, the repetition of a question, and drawls – the non-phonemic lengthening of syllables.
- Signaling End of Speech: The conversation segments that include phrases like *that's all* and greeting terms, such as *welcome, thanks for your attendance* were interpreted as signals that signify the end of speech.
- Looking at the Notebook: The interviewer looked at the notebook when filling in questionnaire or when prepared for the articulation of questions. This category represents the situations as such and specific to interviewers.

The speech instances were identified based on a set of operational assumptions, such as the duration of micro pauses and our interpretation of the speech and non-speech utterances. We conceive the list as an initial step for the analysis of speech data. We use the term *speech-modality* to specify a combined feature including both the role of the speaker (i.e. an interviewer or an interviewee) and the speech-instance itself. In addition, we employed an additional speech-modality measure, namely *speech-modality-onset*, which was identified as the duration of the time from the initial appearance of related speech-instance and the speaker.

Figure 1. A total of 68 landmark positions on face, detected by the OpenFace framework [5].

[1] https://www.cl.cam.ac.uk/~tb346/res/openface.html

[2] http://dlib.net/

[3] http://cmusphinx.sourceforge.net/

2.2.3 *Gaze Aversion Detection*

A gaze aversion may be specified as a situation in conversation, when a conversation partner moves their gaze away from the interlocutor's face. Once the system starts tracking the partner's face, a synchronous analysis of gaze data allows us to identify gaze aversion automatically. Fixation identification algorithms may then be employed to specify whether raw data points cumulate into fixations during the course of gaze aversion. A challenge in the specification of fixations from raw data is that wearable eye trackers capture dynamic scenes. Currently, there is no commonly accepted method for eye movement event detection in dynamic scenes [13, 17]. In the present study, we analyzed raw data after cleansing, as described in the following section.

We detected gaze aversion by using cleansed raw gaze data (exported by the eye tracker manufacturer software) as the input. The cleansing process involved a gap fill with linear interpolation in which at most two frames were filled. After detecting gaze aversions, we merged adjacent aversions in which at most two consecutive non-aversion frames were merged. Lastly, we eliminated short aversions that are less than 100 ms (Figure 2).

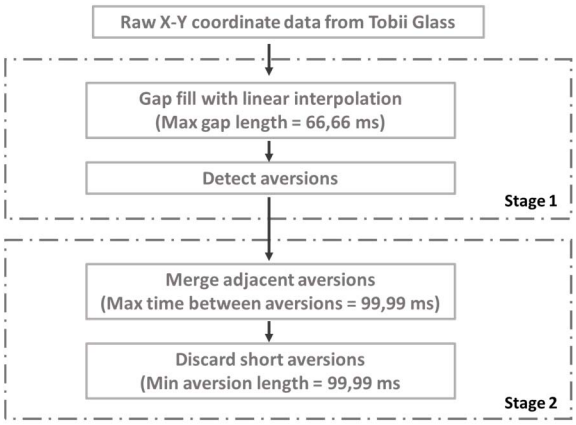


Figure 2. Gaze aversion detection process flow.

The face tracking data and the gaze data were synchronized by overlaying the OpenFace framework output on gaze data. We automatically annotated each video frame (33 ms) to identify whether a participant was looking at the interlocutor's face (viz. *in*), or looking away from the interlocutor's face (viz. *out*). Figure 3 shows detected facial landmarks and gaze data overlay on a sample image frame.

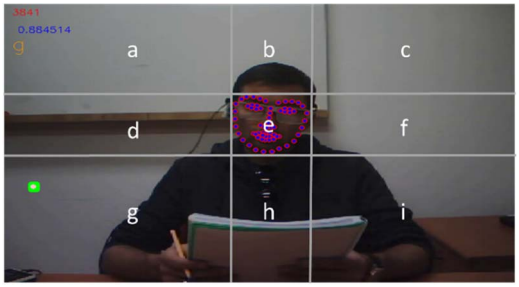


Figure 3. The set of facial landmarks detected by the OpenFace framework and a single-dot gaze location of the interviewee on the face of the interviewer.

For each participant, we synchronously iterated through the annotated gaze data, the partner's annotated gaze data and the speech instance data. The changes in the label of gaze data annotations (i.e., from *in* to *out* and then back to *in*) specified gaze aversion starting times and its ending times, which were analyzed together with the speech data.

3. RESULTS

We analyzed the mean number of gaze aversions per minute (gaze aversion frequency), the mean duration of gaze aversions and the timing of gaze aversion instances.

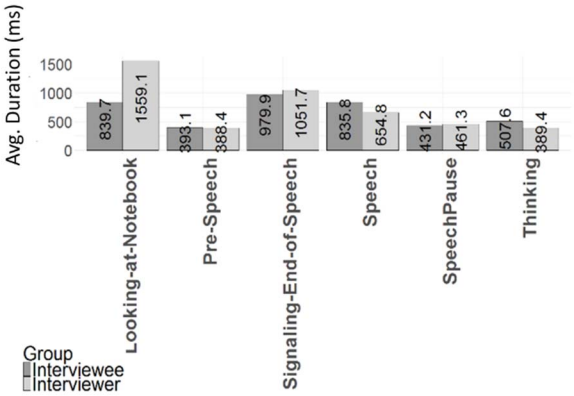
All analyses were carried out in the R programming language and environment [15] using the lme4 and lmerTest software packages [6]. All the data files and R scripts used during the analysis are publicly available.[4]

3.1 Gaze Aversion Frequency

The number of gaze aversions was closely related to how long the corresponding session took. Since no time limit was introduced in the experiment, we calculated a normalized frequency of gaze aversion, per minute. The analysis revealed that the interviewees performed more frequent gaze aversions (M = 27.95, SE = 8.53) compared to the interviewers (M = 22.72, SE = 3.26).

3.2 Gaze Aversion Duration

The analysis revealed that gaze aversions of the interviewees took longer (*M = 2207.9 ms, SE = 1291.2*) than gaze aversions of the interviewers (*M = 1860.0 ms, SE = 363.0)*. These numbers represent the analysis which covered all gaze aversion instances. However, as mentioned above, the interviewers looked at the notebook while they filled in the questionnaire to evaluate interviewee's response and while they articulated the questions. Therefore, we repeated the analysis by excluding those instances (viz., *Looking-at-the-Notebook* instances*)* where the interviewer looked at the notebook, since they did not represent genuine cases of gaze aversions during the course of conversation. The renewed analysis resulted in a more salient difference between the interviewers (*M = 1179.3 ms, SE =384.1*) and the interviewees (*M= 1802.3 ms, SE = 921*) in terms of the duration of gaze aversions. We also investigated the relationship between gaze aversion and speech-instance type. A single gaze aversion might be related with the multiple speech-instances. Figure 4 shows the average duration while a participant was averting his gaze from an interlocutor's face and performing the specific speech-instance.



[4]https://gist.github.com/ulkursln/9d14fe288471b9e83f845607d5c3045d

Figure 4. The average duration of gaze aversion in millisecond for each type of speech-instances. Light gray bars represent interviewers and dark gray bars are for interviewees.

The durations of gaze aversion were analyzed using linear mixed effects regression, LMER, with the lme4 package in R. We identified the participant pairs (viz., *pair-id*) as a random effect to control the influence of different duration values associated with the same pair. In a mixed-model, removing the Speaker-Role, Speech-Instance, Gaze-Behavior-Onset and Speech-Modality-Onset significantly decreased the goodness of fit, as indicated by likelihood ratio tests – effect of speaker-role $\chi2(1) = 22.1, p < .000$; effect of speech-instance $\chi2(5) = 69.6, p < .000$; effect of gaze-aversion-onset $\chi2(1) = 16.1, p = .000$ and effect of speech-modality-onset $\chi2(1) = 20, p < .000$. A post hoc Tukey test was performed on speech-instance category showed that *Looking-at-Notebook* ($M = 1067.8$ ms, $SE = 76.9$) significantly (all ps<.000) increased aversion duration comparing to *Speech* ($M = 742.6$ ms, $SE = 40.6$), to *Pre-Speech* ($M = 398.2$ ms, $SE = 45.2$), to *Speech-Pause* ($M = 432.2$ ms, $SE = 33.9$) and to *Thinking* ($M = 477.7$ ms, $SE = 34.5$). Moreover, the following pairs of instances found to be significantly different (all ps< .05): *Pre-Speech* and *Speech*, *Speech-Pause* and *Speech*, *Thinking* and *Speech*, *Signaling-End-of-Speech* ($M = 1019.3$ ms, $SE = 187.8$) and *Pre-Speech*, *Signaling-End-of-Speech* and *Speech-Pause*, and *Signaling-End-of-Speech* and *Thinking*.

A post hoc Tukey test performed on speaker-role category showed that aversion duration was significantly (p<.000) decreased when the speaker was the interviewee ($M = 629.6$ ms, $SE = 25.4$) compared to the interviewer ($M = 918.8$ ms, $SE = 63.9$). Finally, the lmer mixed-model showed that the duration of aversion was linearly related to gaze-aversion-onset ($b = 132.9$ ms, $SE = 32.9$), and speech-modality-onset ($b = -102.5$ ms, $SE = 30.9$).

3.3 Gaze Aversion Occurrence

We introduced mixed-effects-logistic-regression models, in order to investigate the effects that influence whether it is time to avert gaze by considering following aspects for every 30 milliseconds during the all three sessions: (The sample size was 23115[5])

Gaze Behavior: It can be one of the following labels: *Aversion*, *Face Contact* or *Empty*. The *Empty* label is assigned, when raw gaze data of the participant could not be extracted and/or there was a problem in face detection. This value is handled separately for both interviewer and interviewee participants.

Gaze Behavior Onset: It is the duration of instant gaze behavior since its initial occurrence. This value is handled separately for both interviewer and interviewee participants.

Speaker Role: It can be either an interviewer or an interviewee.

Speech Instance: It can be one of the following six items: *Pre-Speech*, *Speech*, *Speech Pause*, *Thinking*, *Signaling End of Speech* and *Looking at the Screen*.

Speech Modality Onset: Speech-modality is the combined feature including both the role of a speaker (i.e. an interviewer or an interviewee) and the speech-instance. Onset of the speech-modality is the duration of instant speech-modality from the initial occurrence of it.

The first model was created to predict the interviewer's gaze-behavior-type (i.e., whether it was gaze aversion or not in that particular time). As fixed-effects, we identified the interviewer's Gaze-Behavior-Onset, a correlated relation of Speaker-Role, Speech-Instance and Speech-Modality-Onset and lastly a correlated relation of interviewee's Gaze-Behavior and interviewee's Gaze-Behavior-Onset. As the random effect, we had Pair-Id, as mentioned in the previous section. In a mixed-model, removing the Speaker-Role, Speech-Instance, interviewer's Gaze-Behavior-Onset, Speech-Modality-Onset, interviewee's Gaze-Behavior and interviewee's Gaze-Behavior-Onset significantly decreased the goodness of fit, as indicated by likelihood ratio tests – effect of Speaker-Role $\chi2(1) = 2031.7, p < .000$; effect of Speech-Instance $\chi2(5) = 85.9, p < .000$; effect of Gaze-Behavior-Onset $\chi2(1) = 927.9, p < .000$; effect of Speech-Modality-Onset $\chi2(1) = 77, p < .000$; effect of interviewee's Gaze-Behavior $\chi2(1) = 35.4, p < .000$ and effect of interviewee's Gaze-Behavior-Onset $\chi2(1) = 6.25, p < =.01$.

A post hoc Tukey test was performed for making pairwise comparisons between the ratio of gaze aversion to face-contact (i.e. odd ratio) of Speech-Instances. If the odd ratio of the first instance in the pair is larger than the second one, interval will be positive, otherwise it will be negative. Results indicate that the differences *Speech – Pre-Speech*, *Speech Pause – Pre-Speech* and *Speech Pause – Speech* are not significantly different from 0 (i.e. their confidence intervals include 0), and all the other pairs are significantly different (see Figure 5).

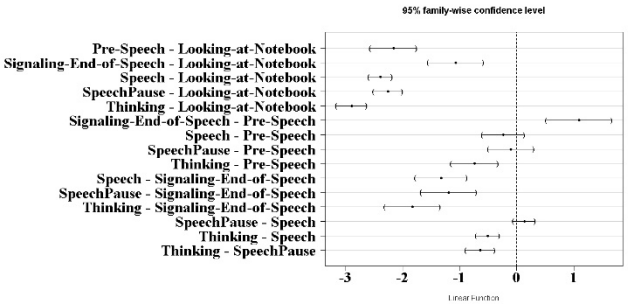


Figure 5. Pairwise comparisons between the ratio of gaze aversion to face-contact of Speech-Instances. The intervals that do not include 0 represents the significant difference. For instance, an interviewer is more likely to avert his eyes while the Speech-Instance is Signaling-End-of-Speech comparing to Pre-Speech:

We performed similar analysis for the interviewees. The second model was created to predict the interviewee's gaze-behavior-type

[5] The link to access the data file: https://drive.google.com/open?id=0B-DfZx3YFEzgRldUNm1fZ3ZPZDQ

(i.e., gaze aversion or not). We identified the interviewee's Gaze-Behavior-Onset, correlated relation of Speaker-Role, Speech-Instance and Speech-Modality-Onset, and lastly a correlated relation of interviewer's Gaze-Behavior and interviewer's Gaze-Behavior-Onset, as fixed-effects. As random effect, we had pair-id. In a mixed-model, removing the Speaker-Role, Speech-Instance, interviewee's Gaze-Behavior-Onset, Speech-Modality-Onset, interviewer's Gaze-Behavior and interviewer's Gaze-Behavior-Onset significantly decreased the goodness of fit, as indicated by likelihood ratio tests – effect of Speaker-Role $\chi 2(1) = 11.6$, $p < .000$; effect of Speech-Instance $\chi 2(5) = 1020$, $p < .000$; effect of interviewer's Gaze-Behavior-Onset $\chi 2(1) = 62.61$, $p < .000$; effect of Speech-Modality-Onset $\chi 2(1) = 7.23$, $p < .000$; effect of interviewer's Gaze-Behavior $\chi 2(1) = 27.01$, $p < .000$ and effect of interviewee's Gaze-Behavior-Onset $\chi 2(1) = 25.22$, $p < =.000$.

A post hoc Tukey test was performed for making pairwise comparisons between the ratio of gaze aversion to face-contact (i.e. odd ratio) of Speech-Instances. If the odd ratio of the first instance in the pair is larger than the second one, interval will be positive, otherwise it will be negative. Results indicate that the differences *Speech Pause – Pre-Speech* and *Speech – Signaling End of Speech* are not significantly different from 0, and all the other pairs are significantly different (see Figure 6).

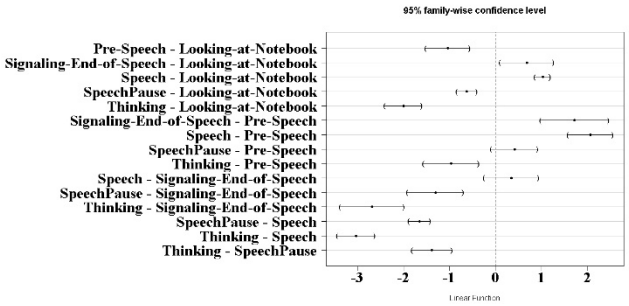


Figure 6. Pairwise comparisons between the ratio of gaze aversion to face-contact of Speech-Instances. The intervals that do not include 0 represents the significant difference. For instance, an interviewee is more likely to avert his eyes while the Speech-Instance is Speech comparing to being Pre-Speech

Relative Spatial Positions of Aversions

We calculated the spatial positions of gaze aversion relative to an interlocutor's face. As represented in Figure 7, during gaze aversion, the interviewees frequently looked at the lower right corner of an interlocutor, while the interviewers looked at straight down as expected in the case when interviewers articulating questions or filling the questionnaire.

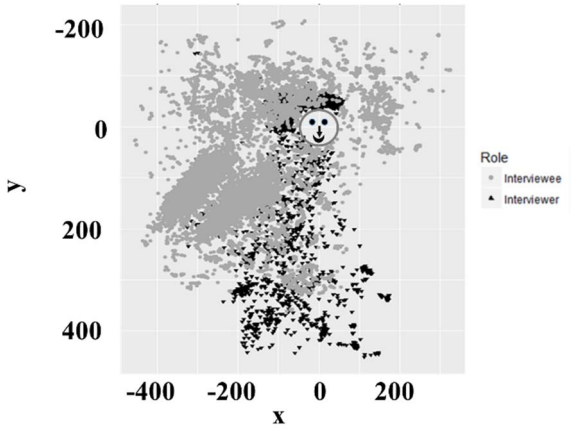


Figure 7. Points represents the gaze aversion position relative to an interlocutor face.

4. CONCLUSION AND DISCUSSION

In the present study, we investigated gaze aversion from a multimodal perspective, by employing face tracking and by analyzing speech data, in addition to eye tracking data in a mock job-interview task. A synchronous use of face tracking and gaze data overlay allowed us to detect gaze aversions of both communication partners.

The results of the study show that gaze aversion characteristics differ between interviewees and interviewers. In particular, the interviewees exhibited more frequent gaze aversions compared to the interviewers. We also found that the interviewees and the interviewers employed different pattern of specific speech instances during the course of conversations.

In its current form, the present study provides a basis for modeling gaze aversion in human-robot interaction scenarios. The previous research have partially shown that modeling the gaze of a robot may lead to a more natural and effective human-robot interaction [15]. The studies have also shown that designing simple aversion mechanisms with fixed frequency and length aversions leads to a feeling of more thoughtful, intentional and creative conversation to human conversation partners, when the aversion mechanism align with the current state and content of the conversation [3], [8]. Nevertheless, the available computational models simulated gaze aversion on humanoid robots through head movements alone, since the robots (e.g., Nao) did not have articulated eyes. The present study aims at contributing to human-robot interaction research on gaze aversion by employing synchronized gaze data obtained from two eye trackers, as well as designing a parametric model of multimodal aspects, in particular speech instances that occur during a conversation.

As future work, we plan to improve the current computational efforts in modeling gaze aversion, in particular by including the direction of gaze aversion, in addition to its frequency and duration. We will design experiments, in which human participants will communicate with a humanoid robot (namely, iCub[6]) with articulated eyes. The articulation of the eyes will be grounded by the findings obtained in the present study. We will also employ the

[6] http://www.icub.org

current findings to design the timing of speech instances during the course of conversation with the humanoid robot.

5. ACKNOWLEDGMENTS

This work was partially funded by Marie Curie Actions IRIS (ref. 610986, FP7-PEOPLE-2013-IAPP). Our thanks to Hatice Köse and Aydan Erkmen for their useful comments and suggestions. We also thank METU Human Computer Interaction Research and Application Laboratory for their technical support.

6. REFERENCES

[1] Abele, A. (1986). Functions of gaze in social interaction: Communication and monitoring. Journal of Nonverbal Behavior, 10(2):83–101

[2] Andrist S., Mutlu B., Gleicher M. (2013). Conversational gaze aversion for virtual agents. In Proc. IVA 2013, pages 249--262,

[3] Andrist, S., Tan, X. Z., Gleicher, M., & Mutlu, B. (2014). Conversational gaze aversion for humanlike robots. ACM/IEEE international conference on Human-robot interaction (pp. 25-32). ACM.

[4] Argyle, M. & Cook, M. (1976). Gaze and mutual gaze. Cambridge University Press Cambridge

[5] Baltrusaitis, T., Robinson, P., & Morency, L. P. (2013). Constrained local neural fields for robust facial landmark detection in the wild. In Proceedings of the IEEE International Conference on Computer Vision Workshops (pp. 354-361).

[6] Bates, D., Maechler, M., Bolker, B., Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1), 1-48. doi:10.18637/jss.v067.i01.

[7] Fasola, J., & Mataric, M. J. (2012). Using socially assistive human-robot interaction to motivate physical exercise for older adults. Proceedings of the IEEE, 100(8), 2512–2526

[8] Ham, J., Cuijpers, R. H., & Cabibihan, J.-J. (2015). Combining Robotic Persuasive Strategies: The Persuasive Power of a Storytelling Robot that Uses Gazing and Gestures.

[9] Hietanen, J. K., Leppänen, J. M., Peltola, M. J., Linna-Aho, K., & Ruuhiala, H. J. (2008). Seeing direct and averted gaze activates the approach-avoidance motivational brain systems. Neuropsychologia, 46(9), 2423–30

[10] Kendon, A. (1967). Some functions of gaze-direction in social interaction. Acta psychologica, 26(1):22–63,

[11] Kleinke, C. L. (1986). “Gaze and eye contact: a research review,” Psychological Bulletin,vol.100,no. 1, pp.78–100

[12] Mason MF, Tatkow EP, Macrae CN (2005). The look of love: gaze shifts and person perception. Psychol Sci 16: 236–239

[13] Munn, S. M., Stefano, L., & Pelz, J. B. (2008). Fixation-identification in dynamic scenes. Proceedings of the 5th Symposium on Applied Perception in Graphics and Visualization

[14] Pfeiffer, U. J., Timmermans, B., Bente, G., Vogeley, K., & Schilbach, L. (2011). A non-verbal Turing test: differentiating mind from machine in gaze-based social interaction.

[15] R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

[16] Ruhland, K., Peters, C. E., Andrist, S., Badler, J. B., Badler, N. I., Gleicher, M., Mutlu, B., & McDonnell, R. (2015). A Review of Eye Gaze in Virtual Agents, Social Robotics and HCI: Behaviour Generation, User Interaction and Perception. In Computer Graphics Forum (Vol. 34, No. 6, pp. 299-326).

[17] Srinivasan, V., Murphy, R., & Henkel, Z. (2012). User Acceptance of Autonomously Generated Social Head Gaze

[18] Stuart, S., Galna, B., Lord, S., Rochester, L., & Godfrey, A. (2014). Quantifying saccades while walking: Validity of a novel velocity-based algorithm for mobile eye tracking. 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society.

[19] Villani, D., Repetto, C., Cipresso, P., & Riva, G. (2012). May I experience more presence in doing the same thing in virtual reality than in reality? An answer from a simulated job interview. Interacting with Computers, 24(4), 265-272.