Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020), pages 759–765
Marseille, 11–16 May 2020

An Annotation Approach for Social and Referential Gaze in Dialogue

Vidya Somashekarappa, Christine Howes, Asad Sayeed
Department of Philosophy, Linguistics and Theory of Science
Centre for Linguistic Theory and Studies in Probability (CLASP)
University of Gothenburg
vidya.somashekarappa@gu.se, christine.howes@gu.se, asad.sayeed@gu.se

Abstract

This paper introduces an approach for annotating eye gaze considering both its social and referential functions in multi-modal human-human dialogue. Detecting and interpreting the temporal patterns of gaze behaviour cues is natural for humans and also mostly an unconscious process. However, these cues are difficult for conversational agents such as robots or avatars to process or generate. The key factor is to recognise these variants and carry out a successful conversation, as misinterpretation can lead to total failure of the given interaction. This paper introduces an annotation scheme for eye-gaze in human-human dyadic interactions that is intended to facilitate the learning of eye-gaze patterns in multi-modal natural dialogue.

Keywords: Eye gaze, social functions, referential functions, multimodal dialogue, human-computer interaction

1. Introduction

In this paper, we first outline the research on social and referential functions of eye gaze in dialogue and argue that these should not be treated independently, as eye gaze information is multifunctional. Secondly, we briefly describe existing multimodal corpora that include eye gaze information. We then explain why we believe there is a need for a new annotation scheme for eye gaze. Finally, we present our annotation scheme and preliminary observations.

Eye gaze is a crucial component of information in interaction. Researchers reason that the depigmentation of the human sclera, unique among primates, has evolved for effective communication and social interaction based on eye contact (Kobayashi and Kohshima, 1997). Linguists and psychologists have shown a long standing interest in non-verbal communication relating to speech and gesture, including eye-gaze (Kendon, 1967; Argyle and Cook, 1976; Goodwin, 1980; Goodwin, 1981) and have identified at least two areas where eye gaze information is relevant, which is the focus of this work.

However, previous studies tend to focus on either social functions of gaze (e.g., turn-taking or other interaction management (Jokinen et al., 2013)) or how gaze is used in reference resolution (Kontogiorgos et al., 2018), with few researchers combining these.

1.1. Social functions of eye gaze in dialogue

Argyle and Cook (1976) showed that listeners display longer sequences of uninterrupted gaze towards the speaker, while speakers tended to shift their gaze towards and away from the listener quite often. Later work has refined these observations, with, for instance, Rossano (2012), noting that these distributional patterns are dependent on the specific interactional activities of the participants; for example, a more sustained gaze is necessary in activities such as questions and stories, since gaze is viewed as a display of attention and engagement. Brone et al (2017) also found that different dialogue acts typically display specific gaze events, from both speakers' and listeners' perspectives.

Unaddressed participants also display interesting gaze behaviour showing that they anticipate turn shifts between primary participants by looking towards the projected next speaker before the completion of the ongoing turn (Holler and Kendrick, 2015). This may be because gaze has a 'floor apportionment' function, where gaze aversion can be observed in a speaker briefly after taking their turn before returning gaze to their primary recipient closer to turn completion (Kendon, 1967; Brône et al., 2017).

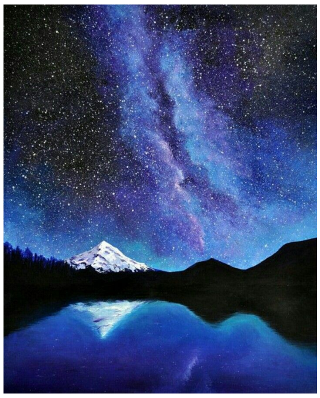

Figure 1: Illustration I

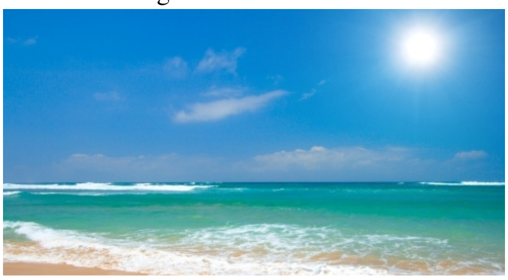

Figure 2: Illustration II

1.2. Referential functions of eye gaze in dialogue

In identifying an image above on display by referring to "the painting of a night sky", your attention is drawn automatically to Illustration I (fig. 1[1]) but not to Illustration II (fig.

[1] See appendix for image sources.

2) even without any necessary pointing gesture. The process of identifying application-specific entities which are referred to by linguistic expressions is reference resolution. One area in which multi-modal reference resolution has been previously studied is in the context of sentence processing and workload. Sekicki and Staudte (2018) showed that referential gaze cues reduce linguistic cognitive load. Earlier work (Hanna and Brennan, 2007) showed that gaze acts as an early disambiguator of referring expressions in language.

Campana et al (2002) proposed to combine the reference resolution component of a simulated robot with eye tracking information; they intended to deploy this on the International Space Station. However, they did not address the integration of eye movements with speech. Also, eye gaze information was used only in case of inability to identify unique referenced objects. Zhang et al (2004) implemented reference resolution by integrating a probabilistic framework with speech and eye gaze; results showed an increase in performance. They also found that reference resolution of eye gaze could also compensate for lack of domain modelling. Visual input has a immediate effect on language interpretation like reference resolution.

1.3. Interaction of social and referential functions

One of the main reasons to look into someone's eyes is to determine their intended goal, since the eye direction of a person reliably signifies what they are going to act upon next. In an experiment (Phillips et al., 1992), eye contact was investigated in young normal infants who observed adults performing actions with ambiguous or unambiguous interpretations and found instant eye contact for ambiguous actions but rarely with unambiguous actions.

The phenomenon of 'eye contact effect', moderates certain facets of concurrent/immediately following cognitive processing (Senju and Johnson, 2009). Developmental studies demonstrate proof of preferential orienting and processing of faces by means of direct gaze from early in life. The ability of 2 to 5 day old newborns to discriminate direct and averted gaze was tested to measure the brain electrical activity to assess neural processing of faces when accompanied by direct (as opposed to averted) eye gaze. The results illustrated that from birth human infants prefer to look at faces that engage them in mutual gaze and that, from an early age, healthy babies show enhanced neural processing of direct gaze (Farroni et al., 2002).

By the 4th week infants fixate and smile at eyes (Argyle and Cook, 1976). This visual interaction between the newborn and caregiver plays a major role in developing attachment (Nijenhuis and Bouchard Jr., 2007). The earliest gaze behaviours act as a foundation to build nonverbal and verbal communicative or social behaviours in later stages (Gillberg, 1998). A visual attention cueing paradigm was used to study gaze in 2 year old children which consisted of eye movements and non-biological movements, the results suggested that the visual attention is cued by perceived eye movements (Chawarska et al., 2003). While on the other hand low-confidence conscious meta-cognitive knowledge and unconscious meta-cognitive knowledge through eye gaze was measured in 3-5 year old children (Ruffman et al., 2001) which established that they were not aware of the knowledge conveyed through their eye gaze. Children develop an increased understanding of social information and intentions carried through dynamic facial cues mainly changes in eye gaze direction during middle childhood (Mosconi et al., 2005).

Hence, it becomes exceedingly important to understand gaze behaviours for improving interactions involving robots/avatars.

2. Research questions

Our motivation for designing a new type of gaze annotation is to make progress on answering the following research questions:

- Annotation: Is it feasible to annotate eye-gaze and elements of dialogue? Is this type of annotation useful for machine learning systems?
- Given some annotations, is it possible to predict, for example, dialogue acts, turn-taking or reference based on eye-gaze alone?
- Can we come up with a model of gaze in dialogue for a conversational robot or avatar to interpret human gaze behaviour and produce human-like gaze behaviour?

Answers to these research questions will contribute both to understanding the cognitive neuroscience of language and to the development of improved human-computer interaction.

3. Reviewing Gaze Annotation in Multi-modal Interaction

Research focusing on multi-modal interaction needs high quality annotation in order to obtain a detailed view of the interaction between visual, verbal and bodily features. A number of projects are interested in collecting and annotating video data due to increase in its demand. The current dyadic interaction experiment proposes a new form of annotation schema.

In the past five decades, the measurement of gaze points and eye movements with eye-tracking techniques during online behaviour has influenced multiple areas of research in psycholinguistics and psychology (Bhattacharyya, 2018). This type of study mainly explores eye gaze as a measure of cognitive processing with participants who are provided with a physical stimulus (e.g., picture or passage on a screen). There has been significantly less attention given to the role of eye gaze in production, particularly in identifying the communication function of gaze and its ties to co-occurring utterances (Ho et al., 2015).

The CID corpus (Bertrand et al., 2006) for interactional data in French is a corpus with single camera perspective profile view with a disadvantage of restricting access to gaze. Camera frontal views are an unnatural environment for the conversationalist, losing the advantages of the traditional face-to-face setting, and limiting the possibility of gaze-based interaction in dialogue.

Corpora such as the Nottingham Multimodal Corpus (NMMC), the Swedish Spontaneous Dialogue Corpus

Figure 3: Session in progress (Original Scene)

(Spontal corpus), and the IFA dialogue video corpus (IFADV) in Dutch are examples of multiple-angle recorded data, but they focus either on social function or only on the referential functions (Brône and Oben, 2014). In the current study, we combine these two functions where reference is part of the interaction. This is more common in a natural multi-modal dyadic dialogue and uses two cameras to gain access to multiple angle interactions that allow analyses of fine-grained, reliable behavioural features.
In multi-modal interaction, this creates a new area of research into gaze as a directive instrument or a disambiguation instrument and provides the opportunity to find potential correlations between gaze, facial expressions, and gesture. The results of a study (Jokinen et al., 2010) exploring non verbal signals for turn-taking and feedback in direct face to face interactions revealed that the gaze, head movement, or gesture primarily function as indexed signs linked to the whole context where they occur rather than symbols which carry meaning.

4. Methods and Materials

This section describes the multi-focal and multi-modal dialogue corpus used for annotation and presents the recording setup, task design, and annotation scheme used to code speech and gaze.

4.1. Recordings

By using video recordings, we are able to study the interactional dynamics specific to face-to-face dialogue along with understanding the collaborative processing and production of language. Hence, during the recording session, participants had to perform a collaborative task having a free range of conversations yielding natural multi-modal interaction.

4.1.1. Recording set-up

Figure 3 depicts the configuration of the recording. Static external cameras were fixed with a profile and a frontal shot of each participant who sat at right angles to each other enabling us to record a clear gaze, tracking face, hands and body from multiple angles, resulting in a very rich representation of interaction providing access to extensive variability of multi-modal cues.
The core data of the multi-modal video recording comes from these fixed cameras that record the ongoing conversation, and the subjects are free to move and gesticulate from where they are seated. Consent is taken from the participants to record the complete session.

4.1.2. Recording devices

The set up required recording video from two perspectives (two cameras) and audio signals via microphones (unidirectional microphones that pickup audio from the respective speaker) which were worn by the participants.

- 2 fixed color cameras (JVC JY-HM360E Profesional HD camcorder, 1.56M pixel LCOS Color Viewfinder and 920K pixel LCD Display)
- 2 microphones (Schertler Cello Microphone, Output impedance: 4.7 kOhm at 1 KHz, Frequency range: 20 - 20,000 Hz)

The fixed camera recordings and the wave-forms of the microphones were synchronised prior to annotation.

4.2. Participants

Participants were twenty four dyads recruited from staff at the Good Housekeeping Institute (GHI, a consumer product testing organisation in the UK[2]). In each session a pair of participants taste-tested eight different types of hummus in the GHI test kitchen (see figure 3), and provided ratings on a single (shared) questionnaire. Sessions lasted approximately 20-30 minutes, but within this, participants organised their time themselves, e.g., deciding how long to take for tasting and rating each hummus, switching freely, choosing their strategies in performing the task and organising their interaction. The dialogues are task-directed rather than completely spontaneous. This type of dialogue is ideal for our purposes as it allows the internal dynamics of the conversation to be entirely free while the task creates an external trigger about which participants are communicating, meaning that both referential and interactive aspects of gaze ought to be present (which might not be the case in spontaneous dialogue as the topic under discussion may not include any shared referents available to visual attention).

5. Annotation

One important factor to be considered while annotating gaze is the duration of gaze fixation on a respective entity. Since people switch between objects extremely quickly the gaze behaviour may seem disorderly, so both directional and durational information needs to be recorded for a reliable categorisation of gaze episodes.

5.1. Annotation tool

Data was annotated in ELAN (Berez, 2007), a tool that provides a framework for annotation of audio and video recordings. This enables us to have precise time-alignments and hierarchically organise annotation tiers as outlined below. ELAN records data in a stand-off XML format.

- ELAN : The audio and video files of each session are added to the software separately, which contains profile and front shot of the participants showing the two videos side by side, as shown in figure 4.

- Transcription: For the video transcription general norms and principles of Gesprächsanalytisches Transkriptionssystem (GAT) were considered (Selting et al., 1998). The orthographical transcription for each participant was done in two separate tiers speech1 and speech2. These tiers contained metadata indicating the beginning and end of the excerpt encoded with respective spoken utterance per unit, in few occasions a short description of the interactional context in unicode (<>) such as laughter, cough, uhm, etc.

5.2. Gaze Annotation

As shown in figure 4, gaze was annotated in six tiers.

- Gaze1 and Gaze2: These contain information about participants' attention on their partner. Gaze1 is annotated as P2 when P1 is looking at P2. Gaze2 is annotated as P1 when P2 is looking at P1.

- P1 and P2: These two tiers contain annotations for when each participant is looking at something in the shared visual field. These are annotated as e.g Hummus (H), Questionnaire (Q), Breadstick (B), Pen (P).

- Mutual attention (MA): Overlap between Gaze1 and Gaze2 represents mutual eye-contact and is annotated as a separate tier.

- Joint attention (JA): Similarly, overlap between P1 and P2 indicates where both participants are fixating on the same thing in the shared visual field. Annotated as Hummus (H), Questionnaire (Q) etc as with P1/P2, above.

6. Preliminary results

As shown in table 1, in the approximately 11 minutes 40 seconds of one dialogue that has thus far been annotated, the participants had equal amounts of speech (157 versus 173 utterance events equal to 25/26% of the time each).
Interestingly, P1 spent somewhat more time looking at P2 than P2 did at P1 (9% to 5%) and these looking events overlapped (such that the participants were looking directly at each other) only on 3 occasions (see the MA row in table 1), and for less than 1% of the total duration of the annotated interaction.
In line with Argyle and Cook (1976) and Rossano (2012), the listener looked at the speaker more frequently than the other way round (4.9% of the time compared to 2.8% of the time). Further investigation of these eye gaze events is needed to see if they co-occur with particular dialogue acts or points where a floor change may occur, as suggested by Brone et al (2017), but if so, this is potentially useful information to a dialogue system.
However, participants' visual attention was far more often on one of the objects in the shared visual field, such as the hummus or the questionnaire (P1 89%; P2 92%) with these annotations overlapping for 61% of the duration of the annotated dialogues (JA) indicating that participants were looking at the same thing more often than not. Interestingly, while both participants spent a lot of time looking at both the hummus and the questionnaire (P1 H: 34%; P2 H: 36%; P1 Q: 40%; P2 Q: 44%) they had joint attention on the questionnaire nearly twice as often as the hummus (JA H: 20%; JA Q: 35%) showing how gaze behaviour is affected by the particular constraints of the sharedness of the sub tasks even within a dialogue (here, the rating is specified to be a joint action, while the tastings can be carried out in parallel).
In terms of reference resolution, based on the intuition that to use gaze behaviour to aid reference resolution it is necessary to look at the other participant whilst their visual attention is on the referent, we compared the overlaps where P1 looked at P2 while P2 looked at something else, and vice versa. Interestingly, P1 looked at P2 while P2 looked at something else for 8% of the time, but the inverse was true only 4% of the time. Further investigation into how this maps to linguistic information, for example, whether P2 was more ambiguous in their speech, is pending.

[2] https://www.goodhousekeeping.com/uk/the-institute/

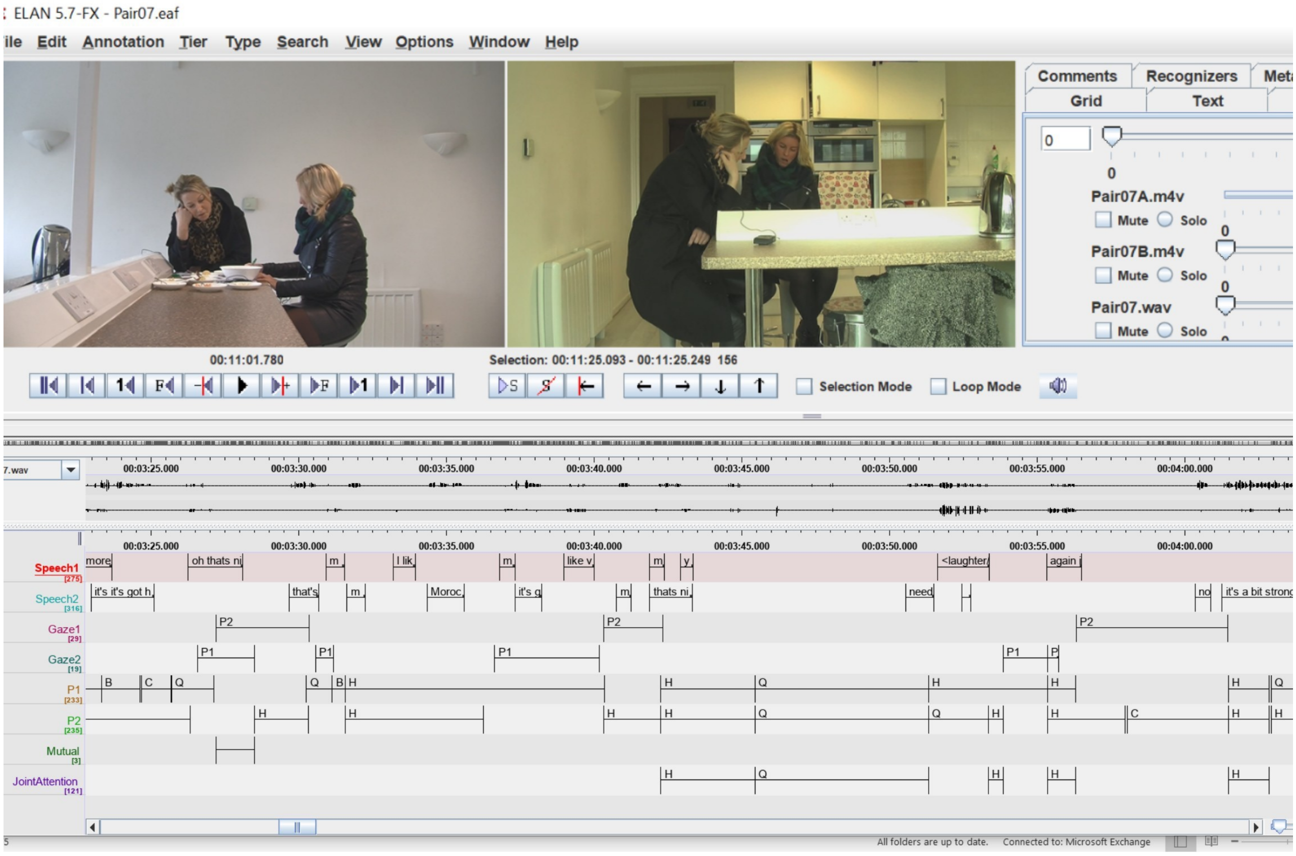

Figure 4: ELAN Annotation

Tiers	# of Ann.	Min D	Max D	Avg D	Median D	Total Ann. D	Ann. D %
Speech1	157	0.290	4.316	1.131	0.873	177.677	25.367
Speech2	173	0.228	4.330	1.041	0.820	180.082	25.711
Gaze1	29	0.313	8.104	2.182	1.656	63.291	9.036
Gaze2	19	0.363	5.221	1.920	1.666	36.472	5.207
P1	232	0.234	25.907	2.690	1.562	624.148	89.111
P2	235	0.168	25.907	2.765	1.687	649.659	92.754
MA	3	1.000	1.687	1.333	1.313	4.000	0.571
JA	121	0.250	25.907	3.581	2.125	433.351	61.871

Table 1: Annotation summary by duration (D)

7. Discussion

Our multi-modal corpus of transcribed speech is annotated and synchronized with eye-gaze details as described above and is produced in ELAN (.eaf) format. It contains high resolution video and audio files, and the video data is in M4V format while the audio data in WAV-format. The corpus presented in this paper is a contribution to the growing quality of multi-modal research.

Detailed gaze annotation helps uncover another dimension of hidden layers in human behavioural interactions. Our results may lead to further research on themes including gaze agreement/disagreement prediction before the emergence of linguistic cues and the influence of decision-making on eye gaze behaviour. Furthermore, the manually annotated data can be used to develop automated dialogue systems.

Processing eye gaze information results from an interaction between face specific structures (involved in visual analysis) and an extended system (spatial attention) as proposed by investigations into the distributed human neural system for face perception (Haxby et al., 2000). Psychophysical interactions (PPIs) in functional magnetic resonance imaging study showed differential connectivity or correlation with core face perception structures in the posterior superior temporal sulcus and fusiform gyrus. This was noted while viewing gaze shifts relative to control eye moments such as the opening or closing of the eyes. It demonstrated the contribution of both the dorsal and the ventral core face areas to gaze perception. Hence, this network provides an interactive system focusing on spatial attention and corresponding shifts in attention (Nummenmaa et al., 2009)

A fearful facial expression with the pointing/directing of the eyes can signify the presence of danger in the surrounding which was investigated in an fMRI study (Hadjikhani et al., 2008) where the meaning of the facial expression along with the direction of the gaze was proven to compute the behavioural implications from the observer's perspective.

Discussing the aforementioned neural correlates of gaze behaviour and its direct influence on dialogue has helped expand our understanding of some linguistic phenomena in persons with disabilities.

One of the important goals of the current project in developing an automated system with improved gaze behaviour recognition is to assist in the diagnosis and rehabilitation of persons with communication impairment, developmental delay, cerebral palsy, quadriplegia, autism, Angelman syndrome, schizophrenia, and aphasia, to name a few.

Patients with schizophrenia displayed reduced non-verbal behaviour and increased negative symptoms (Lavelle et al., 2012) and poor coordination of turn-taking along with disfluencies using fewer self-repairs in dialogue (Howes et al., 2017). Another study presented an analysis of gaze aversion patterns distinguishing between positive and negative schizophrenia (Vail et al., 2017). Children with ASD prefer more limited social interaction compared to children without ASD, hence measurement of eye gaze as a screening tool may be an important contribution in this area (Vargas-Cuentas et al., 2017). As a result, the development of innovative assistive technologies can alleviate current challenges and improve diagnostic accuracy.

8. Acknowledgements

This work was supported by the Swedish Research Council (VR) Grant 2014-39 for the establishment of the Centre for Linguistic Theory and Studies in Probability (CLASP) at the University of Gothenburg.

9. Bibliographical References

Argyle, M. and Cook, M. (1976). Gaze and mutual gaze. Cambridge University Press.

Berez, A. (2007). Eudico linguistic annotator (Elan). *Lang. Document. Conserv.*, 1, 01.

Bertrand, R., Blache, P., Espesser, R., Ferré, G., Meunier, C., Priego-Valverde, B., and Rauzy, S. (2006). Le CID - Corpus of Interactional Data: protocoles, conventions, annotations. *Travaux Interdisciplinaires du Laboratoire Parole et Langage (TIPA)*, 25:31–60, 01.

Bhattacharyya, P. (2018). Applications of eye tracking in language processing and other areas: An investigation based on eye-tracking. 08, pp. 23–46.

Brône, G., Oben, B., Jehoul, A., Vranjes, J., and Feyaerts, K. (2017). Eye gaze and viewpoint in multimodal interaction management. *Cognitive Linguistics*, 28(3):449–483.

Brône, G. and Oben, B. (2014). Insight interaction: a multimodal and multifocal dialogue corpus. *Language Resources and Evaluation*, 49:195–214, 03.

Campana, E., Hockey, B. A., Baldridge, J., Remington, R., Dowding, J., and Stone, L. S. (2002). Using eye movements to determine referents in a spoken dialogue system. *Proceedings of the 2001 Workshop on Perceptive User Interfaces*, 01.

Chawarska, K., Klin, A., and Volkmar, F. (2003). Automatic attention cueing through eye movement in 2-year-old children with autism. *Child development*, 74:1108–22, 07.

Farroni, T., Csibra, G., Simion, F., and Johnson, M. (2002). Eye contact detection in humans from birth. *Proceedings of the National Academy of Sciences of the United States of America*, 99:9602–5, 08.

Gillberg, C. (1998). Chromosomal disorders and autism. *Journal of autism and developmental disorders*, 28:415–25, 11.

Goodwin, C. (1980). Restarts, pauses, and the achievement of a state of mutual gaze at turn-beginning. *Sociological inquiry*, 50(3-4):272–302.

Goodwin, C. (1981). Conversational organization: Interaction between speakers and hearers. Academic Press, New York.

Hadjikhani, N., Hoge, R., Snyder, J., and Gelder, B. (2008). Pointing with the eyes: The role of gaze in communicating danger. *Brain and cognition*, 68:1–8, 07.

Hanna, J. E. and Brennan, S. E. (2007). Speakers' eye gaze disambiguates referring expressions early during face-to-face conversation. *Journal of Memory and Language*, 57(4):596 – 615. Language-Vision Interaction.

Haxby, J., Hoffman, E., and Gobbini, M. (2000). The distributed human neural system for face perception. *Trends in cognitive sciences*, 4:223–233, 07.

Ho, S., Foulsham, T., and Kingstone, A. (2015). Speaking and listening with the eyes: Gaze signaling during dyadic interactions. *PloS one*, 10:e0136905, 08.

Holler, J. and Kendrick, K. H. (2015). Unaddressed participants' gaze in multi-person interaction: optimizing recipiency. *Frontiers in psychology*, 6(98):1–14.

Howes, C., Lavelle, M., Healey, P., Hough, J., and McCabe, R. (2017). Disfluencies in dialogues with patients with schizophrenia. 01.

Jokinen, K., Harada, K., Nishida, M., and Yamamoto, S. (2010). Turn-alignment using eye-gaze and speech in conversational interaction. In Eleventh Annual Conference of the International Speech Communication Association.

Jokinen, K., Furukawa, H., Nishida, M., and Yamamoto, S. (2013). Gaze and turn-taking behavior in casual conversational interactions. *ACM Transactions on Interactive Intelligent Systems*, 3:1, 07.

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

Kobayashi, H. and Kohshima, S. (1997). Unique morphology of the human eye. *Nature*.

Kontogiorgos, D., Sibirtseva, E., Pereira, A., Skantze, G., and Gustafson, J. (2018). Multimodal reference resolution in collaborative assembly tasks. 10.

Lavelle, M., Healey, P., and McCabe, R. (2012). Is non-verbal communication disrupted in interactions involving patients with schizophrenia? *Schizophrenia bulletin*, 39, 09.

Mosconi, M., Mack, P., McCarthy, G., and Pelphrey, K. (2005). Taking an "intentional stance" on eye-gaze shifts: A functional neuroimaging study of social perception in children. *NeuroImage*, 27:247–52, 09.

Nijenhuis, J. and Bouchard Jr., T. (2007). Replication of the hierarchical visual-perceptual-image rotation model in de Wolff and Buiten's (1963) battery of 46 tests of mental ability. *Intelligence*, 35:69–81, 02.

Nummenmaa, L., Passamonti, L., Rowe, J., Engell, A., and Calder, A. (2009). Connectivity analysis reveals a cortical network for eye gaze perception. *Cerebral cortex (New York, N.Y. : 1991)*, 20:1780–7, 12.

Phillips, W., Baron-Cohen, S., and Rutter, M. (1992). The role of eye contact in goal detection: Evidence from normal infants and children with autism or mental handicap. *Development and Psychopathology*, 4:375 – 383, 07.

Rossano, F. (2012). Gaze behavior in face-to-face interaction. Ph.D. thesis, Radboud University Nijmegen.

Ruffman, T., Garnham, W., Import, A., and Connolly, D. (2001). Does eye gaze indicate implicit knowledge of false belief? Charting transitions in knowledge. *Journal of experimental child psychology*, 80:201–24, 12.

Sekicki, M. and Staudte, M. (2018). Eye'll help you out! How the gaze cue reduces the cognitive load required for reference processing. *Cognitive Science*, 42(8):2418–2458.

Selting, M., Auer, P., Barden, B., Bergmann, J., Couper-Kuhlen, E., Günthner, S., Meier, C., Quasthoff, U., Schlobinski, P., and Uhmann, S. (1998). Gesprächsanalytisches transkriptionssystem (Gat) [conversational analytic transcription system]. *Linguistische Berichte*, 173:91–122, 01.

Senju, A. and Johnson, M. (2009). The eye contact effect: Mechanisms and development. *Trends in cognitive sciences*, 13:127–34, 03.

Vail, A., Baltrusaitis, T., Pennant, L., Liebson, E., Baker, J., and Morency, L.-P. (2017). Visual attention in schizophrenia: Eye contact and gaze aversion during clinical interactions. pages 490–497, 10.

Vargas-Cuentas, N. I., Roman-Gonzalez, A., Gilman, R., Porras, F. P., Ting, J., Hidalgo, D., Jensen, K., and Zimic, M. (2017). Developing an eye-tracking algorithm as a potential tool for early diagnosis of autism spectrum disorder in children. *PLOS ONE*, 12:e0188826, 11.

Zhang, Q., Imamiya, A., Go, K., and Mao, X. (2004). Overriding errors in a speech and gaze multimodal architecture. In Proceedings of the 9th international conference on intelligent user interfaces, pages 346–348, 01.

10. Appendix

Illustration I:

`http://www.digitalpicturezone.com/digital-photography-tips-and-tricks/taking-photos-in-the-mid-day-sun/`

Illustration II:

`https://www.pinterest.com/pin/717268678124480359/`