Gaze, Dominance and Dialogue Role in the MULTISIMO Corpus

Lorcan McLaren
Computational Linguistics Group
Trinity College Dublin
The University of Dublin
Ireland
Email: lmclaren@tcd.ie

Maria Koutsombogera
Computational Linguistics Group
Trinity College Dublin
The University of Dublin
Ireland
Email: koutsomm@scss.tcd.ie

Carl Vogel
Computational Linguistics Group
Trinity College Dublin
The University of Dublin
Ireland
Email: vogel@tcd.ie

Abstract—**Analysis of human interactions benefits from understanding patterns of gaze in dialogue. Using a multi-modal task oriented dialogue corpus, we evaluate relations between gaze behaviours, conversational dominance and dialogue role. Dominant players spend more time looking at their co-player than their less dominant fellows; the facilitator spends more time gazing at the dominant player than at the less dominant; dominant players spend longer averting their gaze than their less dominant fellows. When aligned with dialogue roles, gaze behaviour differences between players are demonstrated when occupying the role of speaker but not of listener. Less dominant speakers gaze longer at a facilitator, while effects on averted gaze and co-player directed gaze remain present for dominant speakers but not for dominant listeners. Finally, where dominance is not balanced among players, mutual gaze involving players only is less than mutual gaze involving facilitators.**

I. INTRODUCTION

As the overall discipline of cognitive infocommunications addresses technologies that extend human cognitive capabilities [1], [2], some research within this discipline attempts to understand the nature of the current extent of human cognitive capabilities and the capabilities enabled by adoption of technologies in the past – within the linguistic and behavioural interaction track of cognitive infocommunications, language and gesture have been addressed as such technologies [3], [4]. Gaze is a kind of gesture. It is relevant to understand how gaze is used in dialogue and how this depends on observable subjective and objective quantities. Here we focus on dominance of interlocutors as perceived in a sample of collaborative dialogue, using observational methods [5]. Deep understanding of the use of gaze in the collaborative dialogues that we study (and in other collaborative dialogues and other dialogue types) may inform new technologies.

Gaze is a proxy for focus of attention, with humans making saccades (rapid eye movements) in order to "bring areas of the visual world onto the fovea (the part of the retina with the highest acuity) for further scrutiny" [6, pg. 29]. In Kendon's theory gaze serves three primary functions: monitoring, expression and regulation [7]. Monitoring describes gaze used as a source of nonverbal information from interlocutors, whereas expressive gaze conveys information to interlocutors – for example, it is suggested that mutual gaze duration is used to generate and manage intimacy [8]. Gaze also serves to regulate processes necessary for successful interaction, including indicating addressee and managing turn-taking. For example, in turn-taking, the speaker gazes away while they continue to hold the floor, before gazing back towards the listener as they (the speaker) provide a new piece of information. This gaze behaviour and content together indicate the end of the turn [9].

Social or conversational dominance is a factor that is related to the verbal and non-verbal behaviour of an individual, and their perception by others. Speakers who are conversationally dominant and either extroverted or disagreeable have been found to be more persuasive [10]. Conversational dominance appears as a factor even in unstructured interaction where participants are unacquainted and of equal standing. Social background and gender can have an impact on dominance [11].

We use a heuristic approach to determining gaze allocation automatically in a multi-modal task oriented dialogue corpus to evaluate relations between gaze behaviours and conversational dominance and dialogue role [12]. The multi-modal corpus adopted is the MULTISIMO corpus of task based dialogues [13], [14] in which each involves two participants collaborating in the company of a facilitator to estimate popular opinions. In this respect, dialogue role has two perspectives in this work: speaker vs. listener and player vs. facilitator. Gaze act types were defined with respect to combinations of gaze issuer and recipient. Counts and duration of the gaze acts and their relation to Absolute Dominance Score and ranked dominance score (high [HD] vs. low [LD] dominance) were tested with the expectation that there are statistically significant patterns supporting the use of gaze behaviour as an indicator of dominance. We seek insight into gaze behaviour in the MULTISIMO corpus because it has a collaborative design – we imagine that dominance and gaze patterns manifest distinctly in adversarial settings. Results are informed here by a subset of the MULTISIMO corpus for which the recording situation supports high-accuracy heuristic automatic gaze annotation.

Our motivation is to explore possible ways in which the conversational dominance of the participants (assessed externally by human annotators) may be inferred by their gaze behaviour, i.e. who looks at whom, how many times and for how long. The direction, the frequency and the amount of gaze may be dynamically shaped in the course of the conversation by factors such as *(i)* task-determined dialogue role, i.e. a facilitator's communicative intent is to ensure balanced contributions from players, and players ideally would

collaborate to solve the task; *(ii)* the dynamically determined role as speaker or listener; and *(iii)* the relative differences of the players with respect to dominance as scored by the observers external to the conversations.

The remainder of this paper is organised as follows: §II addresses more of the relevant literature; §III describes the corpus, annotations and methods of analysis; §IV presents the results of this analysis; §V discusses these findings; §VI offers a conclusion and some perspectives on possible future work.

II. RELATED WORK

This research is aligned with work in cognitive infocommunications [1], [2], [15] that addresses linguistic and behavioural interaction [4], [16]–[19], such as gesture in natural dialogue [20]. Much research indicates that gaze behaviour is linked to dialogue role – speaker vs. listener. It has been noted that listeners gaze at speakers more than speakers gaze at listeners: listeners gaze at speakers for longer durations with short breaks to gaze away; speakers alternate gazing away and gazing at the listener at similar rates [7]. Similarly, some participants spend more time looking at their dialogue partner when listening than when speaking [21]. In a sample of the MULTISIMO corpus, the person who the speaker was looking at is the addressee in 77% of cases [22]. However, this research was based on manual gaze annotation of only 2 of 18 publicly available sessions, so results should be taken as tentative.

Gaze behaviour may be linked to personality traits [23] and may depend on whether the relationship between participants is competitive or cooperative [24], or on social hierarchy: the mean duration observers spent looking at a higher ranked partner's facial features in a photo was shorter than that spent looking at a lower ranked partner [25]. Cultural factors may also influence gaze behaviours [26]. Evidence suggests that individuals in some cultures might avoid eye contact to convey non-aggression [27]. Cultural factors have been demonstrated for the target of interlocutor gaze – Western participants are more likely to fixate on the eyes, while East Asian participants are more likely to fixate on the bridge of the nose [28].

Others have explored the impact of certain verbal and non-verbal cues on dominance perceptions in dialogue systems, using models with different parameters of each of a number of behaviours, including gaze [10]. The findings show that gaze had an influence on the perception of extraversion but not directly on social dominance. In the VACE corpus, gaze frequency – especially when duration was brief – was an indicator of passivity and has a negative correlation with dominance. Long gaze duration, however, was positively correlated with dominance [29]. Fukuhara and Nakano's work [30] reports that more dominant speakers have higher turn-releasing success ratio. They also found that when speaking, an individual provided more gaze attention to a dominant participant than to a less dominant fellow. When acting as a listener, an individual was also more likely to exchange mutual gaze with the more dominant person who was speaking, than with a less dominant one. To date, only pilot investigations of gaze behaviour in the MULTISIMO corpus have been conducted [31].

III. MATERIALS & METHODS

Here we describe the methods, the research design, including an overview of the corpus and the performed data analysis.

A. Resources

The data analyzed is a subset of the MULTISIMO corpus [14]. MULTISIMO is a multimodal dialogue corpus consisting of sessions involving three participants, with two participants acting as players and the other acting as facilitator. The participants play a game intended to elicit cooperation among the players, similar to the popular American game show *Family Feud*. The facilitator asks a series of questions and players must come up with the 3 responses for each, ranked from most likely to least likely, that they believe will be the most common answers in a survey of 100 individuals. The interaction "rules" are not defined to depend on gaze, and therefore, gaze patterns visible in the corpus may be presumed to be fully natural. The facilitator provides feedback and guidance until the correct responses and order are achieved. The corpus contains audio and video recordings from 18 different sessions involving 36 randomly allocated pairs of players, and three facilitators. It includes individual video recordings and a recording that captures the whole scene. Audio is captured by an omnidirectional microphone and additionally by the participants' head mics. The MULTISIMO corpus as a whole is available online,[1] including the subset of the corpus with the automatic gaze annotations [12] used here.

1) Existing Gaze Annotation: Manual gaze annotation currently exists for two sessions within the corpus [31]. Possible values for players are: GAZE_PLAYER, GAZE-FACILITATOR, GAZE_AWAY. The annotation values related to the facilitator are: GAZE_AWAY, GAZE_PLAYER-LEFT, GAZE_PLAYER-RIGHT. Left and right correspond to the perspective of a viewer who faces the facilitator/player in question.

2) Human Dominance Assessment: A human dominance assessment was provided for each player in order to generate ground truth for establishing linguistic indicators of conversational dominance [32]. 5 annotators watched each session twice and provided a dominance score for each participant on 5 point scale, with 1 being the lowest and 5 being the highest [33]. The annotators were not informed as to what verbal or non-verbal cues might be indicators of dominance, but they were provided with a brief definition of conversational dominance.

B. Video Analysis

Heuristic measures are adopted to calculate whether the appropriate annotation for any frame involves the eyes directed to the person situated to the left or right, or whether the individual is blinking or looking away [12]. Using knowledge of the MULTISIMO configuration, more specific labels attend to whether an individual is looking at the dialogue partner or facilitator, or whether facilitators look to the left or right participant. Inter-annotator reliability was used to measure the agreement between automatically generated annotations for one of the sessions and the manual annotations for the same session. Granularity differences resulting from frame-by-frame analysis, a level of detail not feasible in manual annotation, mean that it is not a gold-standard but rather an independent alternative for comparison. The average Cohen's Kappa for the three participants is 0.79, with a raw agreement value of 87%. Our Kappa value indicates sufficient accuracy for secure conclusions about relationships involving gaze behaviours.

---

[1]http://multisimo.eu/datasets.html

C. Data Analysis

The annotations for 7 of the total 18 sessions were chosen for statistical analysis on the basis that the automatic labelling method was deemed for these to be of sufficient quality [12].

1) Human Dominance Assessment: An intraclass correlation coefficient [ICC] was used to establish the level of agreement among annotators for the task described in §III-A2. The ICC was determined to be 0.776, which indicates good agreement. Averages across these 5 assessments are used to calculate an overall dominance score for each participant.

2) Metrics: There are four different types of gaze acts of interest for each session, characterised by the specific combination of gaze issuer and recipient in each case: Facilitator→player, Player→player, Player→facilitator, Player→away. Frequencies and average duration were measured for these four gaze acts. The results of a Shapiro-Wilk test for normality led to the choice of non-parametric tests.

3) Absolute Dominance Score Correlations: A Spearman correlation test was used to determine whether a relationship existed between absolute dominance score (i.e. the average score across the 5 annotators) and mean duration of three of the gaze acts listed in the previous subsection: player→player, player→facilitator, and facilitator→player gaze.

4) Relative Dominance Ranking: We explored the relationship between the 4 gaze acts and the relative dominance of a participant, i.e. their 'rank' when players are split into two groups based on their dominance score. For each session, the player with the higher dominance score was assigned to the HD group while the player with the lower dominance score to the LD group. Of the 7 sessions analyzed, two had players whose dominance scores were identical. In these cases, players were split into left and right groups based on their position in relation to the facilitator. This provides a natural control for contrast with the ranked groups. For the duration and frequency of each of the four gaze acts, an unpaired Wilcoxon test was used to evaluate the significance of the relationship between ranked group membership and the gaze act in question.

5) Mutual Gaze and Dominance: Some gaze behaviours are mutual. We analyze this in relation to relative dominance and dialogue role (facilitator vs. player). We examine the types of mutual gaze: Facilitator-Player (F-P), Player-Player (P-P) or the absence of mutual gaze. We also study the relative dominance of the players in F-P mutual gaze. We look at the cross-classification of gaze counts and durations, using non-parametric contingency table tests to reason about significance.

6) Ranked Dominance and Dialogue Role: Gaze behaviour as it relates to dialogue role (speaker vs. listener) for the ranked groups was analysed on the basis of existing turn-taking data in the corpus. Gaze acts were aligned with turn data and divided into speaker and listener groups, which were further subdivided into HD and LD groups by participant as above. Unpaired Wilcoxon tests were used to establish whether any significant relationships existed between group membership and gaze act.

IV. RESULTS

A. Human Dominance Assessment

The preexisting dominance scores by the 5 annotators were averaged for each of the 7 sessions (see Table I). Table II shows the membership of each group after players have been divided into HD and LD groups based on relative average dominance.

TABLE I.   AVERAGE DOMINANCE SCORE FOR SESSIONS ANALYSED

Session	Player 1 (Left)	Player 2 (Right)
S02	2	3.4
S05	3.8	2
S07	3.4	3.4
S09	3.6	3.6
S20	4.2	4
S22	3.4	3.6
S23	3.6	3.4

TABLE II.   GROUP MEMBERSHIP BASED ON AVERAGE DOMINANCE SCORE

Session	HD	LD
S02	P007	P006
S05	P013	P012
S07	Equal	Equal
S09	Equal	Equal
S20	P042	P043
S22	P047	P046
S23	P048	P049

B. Absolute Dominance Score Correlation

Table III shows the figures used in the Spearman correlation analysis for each of three gaze types: facilitator→player, player→facilitator and player→player gaze (for each player in a pairing). The Spearman correlation between mean duration of facilitator→player gaze and dominance score yields $\rho = 0.139$, but it is not significant ($p = 0.6359$). The correlation between the mean duration of player→player gaze and the dominance score of the gaze-receiving player yields $\rho = -0.449$, but is not significant ($p = 0.1077$). The correlation between the mean duration of player→facilitator gaze and dominance score yields $\rho = -0.451$, but is not significant ($p = 0.1057$). The correlation between the mean duration of player→player gaze and the dominance score of the gaze-issuing player yields $\rho = 0.066$, and is not significant ($p = 0.8226$).

TABLE III.   STATISTICS USED FOR CORRELATION: MEAN DURATION (IN MSEC) OF GAZE ACTS

Player	Score	F→P	P→P (Rec.)	P→F	P→P (Iss.)
P006	2	2164.048	2742.193	2736.408	2505.657
P007	3.4	2638.179	2505.657	4058.014	2742.193
P012	2	1371.914	4500.156	3307.884	1332.968
P013	3.8	2369.209	1332.968	2785.056	4500.156
P016	3.4	1121.916	2513.79	3394.737	713.2823
P017	3.4	1698.238	713.2823	4337.025	2513.79
P020	3.6	2078.725	1367.642	2297.85	1479.296
P021	3.6	1714.19	1479.296	3079.966	1367.642
P042	4.2	2456.205	1526.615	1337.967	1544.758
P043	4	1925.946	1544.758	1293.273	1526.615
P046	3.4	2489.099	946.1905	1093.375	1141
P047	3.6	2064.806	1141	1147.928	946.1905
P048	3.6	2620.46	1837.081	445.2429	1925.975
P049	3.4	2555.273	1925.975	2556.394	1837.081

C. Ranked Relative Dominance Analysis

For each session, the more dominant participant is categorized as HD and the less dominant one as LD. First we present the results derived from the five sessions where there

is a difference in the mean dominance score for participants, then we analyze the two equal participant dominance sessions.

Player→Facilitator Gaze: The mean duration of player→facilitator gaze is 1832.08 msec for HD members and 2080.71 msec for LD members, a difference that is not significant ($p = 0.473$). The mean count of player→facilitator gaze is 95 per session for HD members and 98.8 for LD members. The difference is not significant ($p = 0.5476$). *Player→Player Gaze:* The mean duration of player→player gaze (where the player in question is the gaze-issuer in each case) is 2309.76 msec for HD members and 1647.03 msec for LD members. The difference in the mean duration for this gaze act between groups is significant ($p = 1.34e$-7). The mean count of player→player gaze is 91.6 per session for HD members and 102.8 per session for LD members. The difference is not significant ($p = 0.2492$). *Facilitator→Player Gaze:* The mean duration of facilitator→player gaze is 2433.45 msec for HD members and 1982.29 msec for LD members. The difference is significant ($p = 0.03043$). The mean count of facilitator→player gaze is 94.6 per session for HD members and 97.2 per session for LD members. The difference is not significant ($p = 1$). *Player→Away Gaze:* The mean duration of player→away gaze is 709.84 msec for HD members and 590.56 msec for LD members, a difference which is significant ($p = 0.001415$). The mean count of player→away gaze is 102 per session for HD members and 102.2 per session for LD members. The difference is not significant ($p = 1$).

In two sessions the participants have equal dominance scores. In these cases, players were split into L- and R-Player groups based on their position relative to the facilitator. The mean duration of player→facilitator gaze is 3456.81 msec for L-Players and 3277.20 msec for R-Players. The difference in the mean duration is not significant ($p = 0.918$). The mean duration of player→player gaze (where the player in question is the gaze-issuer) is 2030.57 msec for L-Players and 942.82 msec for R-Players. This difference is significant ($p = 6.46e$-7). The mean duration of facilitator→player gaze is 1819.78 msec for L-Players and 1311.81 msec for R-Players. The difference is significant ($p = 0.008231$). The mean duration of player→away gaze is 1059.14 msec for L-Players and 1007.63 msec for R-Players. The difference in the mean duration is not significant ($p = 0.5706$).

D. Mutual Gaze and Dominance

Mutual gaze events are recorded as either involving a facilitator and a player (F-P) or two players (P-P) or the complement, events without mutual gaze. Analytically, three-way mutual gaze events are not possible. We expect an interaction between counts and durations of mutual gaze and the balance of dominance and the facilitation style adopted by each of the facilitators. The balance of dominance is construed as a binary factor – either dominance is balanced or it is not.

We anticipate that where dominance is balanced, mutual gaze involving players would be either greater or equal to mutual gaze involving facilitators and that where dominance is not balanced, mutual gaze involving players would be less than mutual gaze involving facilitators, in the latter case, the facilitator sharing more mutual gaze with players in managing

the dominance imbalance. We directly test these hypotheses using counts of mutual gaze events and durations as counts of milliseconds with directed binomial tests (see Table IV).

TABLE IV. BINOMIAL TESTS OF DIRECTED HYPOTHESES REGARDING RELATIVE COUNTS OF MUTUAL GAZE EVENTS AND MILLISECONDS INVOLVING THE FACILITATOR OR PLAYERS ONLY, IN RELATION TO BALANCED AND IMBALANCED DOMINANCE SESSIONS

Dom.	Type	Outcome	Trials	Alternative	Actual	p
Bal.	Count	231	(231+682)	P-P > F-P	0.253	1
Bal.	Duration	105156	(105156+452898)	P-P > F-P	0.189	1
Imbal.	Count	891	(891+1656)	P-P < F-P	0.350	< 2.2e-16
Imbal.	Duration	550883	(550883+985729)	P-P < F-P	0.359	< 2.2e-16

We explored the the duration and counts of mutual gaze events (and their complement) in relation to the gazing pairs, dominance and facilitator. Given the small number of sessions and resulting incompleteness of coverage of the interacting conditions, the complete interaction of all three factors is not possible to assess. See Table V for the data profile.

TABLE V. COUNTS AND MEDIAN DURATIONS (IN MSEC) OF EVENTS OF MUTUAL GAZE (AND THE COMPLEMENT), BY FACILITATOR AND DOMINANCE CATEGORIES

		Counts		Median Duration (msec)	
		Balanced Dominance		Balanced Dominance	
Facilitator	Mutual Gaze Type	False	True	False	True
	F-P	817	262	469	440.0
M001	P-P	441	96	532	360.0
	Complement	751	272	387	343.5
	F-P	0	420	NA	534
M002	P-P	0	135	NA	367
	Complement	0	485	NA	470
	F-P	839	0	359.0	NA
M003	P-P	450	0	367.0	NA
	Complement	950	0	341.5	NA

A contingency table analysis allows one to reject the null hypothesis that there is no interaction between mutual gaze type and dominance balance ($\chi^2 = 41.365, df = 2, p = 1.041e-09$). However, inspection of Pearson residuals[2] reveals that there is more F-P mutual gaze in dominance balanced conditions (0.866) and less in dominance imbalanced conditions (-0.543) than one would expect if there were no interaction between gaze type and dominance, but the divergence from expectation is not significant. There is significantly more P-P mutual gaze in dominance imbalanced sessions (3.017) than one would expect if there were no interaction. In balanced dominance sessions there is significantly less P-P mutual gaze (-4.812) and significantly more events without mutual gaze (2.406) than one would expect if there were no interaction between mutual gaze type and the balance of dominance. If one restricts attention to the one facilitator who participated in both dominance balanced and imbalanced sessions, the same overall trend is visible ($\chi^2 = 14.873, df = 2, p = 0.00059$), but the only significant residual is the unexpected relative dearth of P-P mutual gaze in balanced dominance sessions (-2.844).

Mutual gaze type durations are, at best, log-normally distributed, and without homogeneous variance in the cross classification categories. We opted for a non-parametric analysis of duration, taking the relevant figures to be counts of milliseconds. The interaction between dominance balance and mutual gaze types is significant overall ($\chi^2 = 67600, df =$

---

[2]Pearson residuals with magnitude between 2 and 4 are significant, at the $p < 0.05$ level and greater than 4 at the $p < 0.001$ level, with the sign indicating the direction of the violation of null-hypothesis expectations.

2, $p < 2.2e - 16$) or attending strictly to M001 as facilitator ($\chi^2 = 20599$) and all of the residuals are highly significant and in the same pattern overall or restricted to M001; these are reported in Table VI. Relative to the null hypothesis of no interaction between dominance balance and mutual gaze type on the duration of mutual gaze types, it is evident that there is: more mutual gaze with facilitators in balanced dominance sessions; more mutual gaze with facilitators in imbalanced dominance sessions; less P-P mutual gaze in balanced dominance sessions; more P-P mutual gaze in imbalanced dominance sessions; more time not in mutual gaze in balanced dominance sessions and less in imbalanced dominance sessions. The direction of trends when considering durations is the same as in considering counts of each type of mutual gaze event, as outlined above.

TABLE VI.   PEARSON RESIDUALS FROM χ^2 ANALYSES OF THE INTERACTION BETWEEN MUTUAL GAZE TYPE AND THE BALANCE OF DOMINANCE, USING COUNTS OF MILLISECONDS

	Facilitator = M001 Balanced Dominance		All Facilitators Balanced Dominance	
Mutual Gaze Type	False	True	False	True
F-P	-17.26123	34.73833	-33.97402	53.03518
P-P	56.00955	-112.71954	125.69374	-196.21432
Complement	-25.37266	51.06263	-52.11417	81.35286

E. Ranked Dominance and Dialogue Role

Table VII presents the mean count and duration of gaze acts for listeners and the results of an unpaired Wilcoxon test comparing these measures for HD and LD groups.

TABLE VII.   MEAN COUNTS AND DURATIONS (IN MSEC) FOR LISTENERS BY RANKED DOMINANCE GROUP

Metric	HD listener	LD listener	p-value
P→P count	101	100.6	0.9163
P→P duration	1157.139	1091.672	0.4033
P→F count	93.6	102.8	0.6905
P→F duration	1504.519	1606.389	0.2107
P→Away count	80.4	76.8	1
P→Away duration	628.29	550.5703	0.4323

HD and LD players behave similarly as listeners. However, there are significant differences in the average duration of each gaze act when HD and LD player act as speakers. Table VIII lists the speakers' mean count and duration of gaze acts and the results of an unpaired Wilcoxon test comparing these measures for HD and LD groups. HD speakers spend a significantly greater average time gazing at their fellow player and gazing away, while LD speakers spend longer gazing at the facilitator.

TABLE VIII.   MEAN COUNTS AND DURATIONS (IN MSEC) FOR SPEAKERS BY RANKED DOMINANCE GROUP

Metric	HD speaker	LD speaker	p-value
P→P count	88	86.6	0.6752
P→P duration	1205.539	899.2887	0.0005888
P→F count	70.8	79	0.8413
P→F duration	629.9068	790.757	0.02133
P→Away count	54.6	57.8	1
P→Away duration	595.1905	569.1869	0.004667

Table IX presents the mean count and duration of gaze acts for facilitators according to their recipient and separated by dialogue role, as well the results of an unpaired Wilcoxon test comparing these measures for HD and LD recipients.

Facilitators gaze at HD players for longer on average than at LD players, but this difference is not statistically significant.

TABLE IX.   MEAN COUNTS AND DURATIONS (IN MSEC) FOR FACILITATORS BY RECIPIENT

Role	Metric	HD recipient	LD recipient	p-value
Speaker	F→P count	80.6	75.4	0.8413
	F→P duration	1147.283	1114.08	0.6229
Listener	F→P count	83.2	77.8	0.8413
	F→P duration	1829.584	1614.884	0.5578

V.   DISCUSSION

A degree of chance was involved in creating the ranked dominance groups, as some members of the HD group would have been in the LD group in a different pairing and vice versa. However, the results suggest that dominance, insofar as it relates to gaze behaviour, is a relative rather than essential quality. It might be interesting to explore if and how the results differ if the count is normalised by the session length.

Ranked dominance interacts with player→player gaze, with HD players gazing at LD players for a longer duration on average than the inverse, i.e. dominant players spend more time looking at their fellow player than their less dominant counterparts. LD players gaze at their fellow player more frequently than HD players; however, this may not be ruled out as a chance result, and therefore we maintain that there is no significant difference in the frequency at which HD and LD members gaze at their fellow players. Ranked dominance interacts with facilitator→player gaze, with the facilitator gazing at HD players for a longer duration on average than the inverse. A high p-value indicates that there is no significant difference in the frequency at which the facilitator gazes at HD vs. LD players. Therefore, it is affirmed that the facilitator spends more time gazing at the dominant player than at the less dominant one. It is worth considering whether the facilitator gives more attention to the HD players as a result of their dominance, or if the facilitator has created a situation of dominance by allocating more gaze attention to them, which contributed to the annotated dominance scores. Ranked dominance interacts with player→away gaze, with HD players gazing away for a longer duration on average than LD players. A high p-value indicates that there is no significant difference in the frequency at which HD versus LD players gaze away. On this basis, dominant players spend longer averting their gaze than their less dominant counterparts. This may be due to the fact that they also occupy more speaking time on average, with averted gaze indicating an intention to retain the floor. HD speakers spend longer periods gazing at their fellow player and gazing away, while LD speakers gaze longer at the facilitator. One might infer that LD participants spend more time looking at the facilitator for reassurance or confirmation. We hypothesised that the longer gaze aversion for HD players is a consequence of longer speaking times, as individuals tend to gaze away while maintaining their turn. However, inspection of speaking turn duration for HD and LD groups indicates no significant difference according to an unpaired Wilcoxon test ($p = 0.5032$). Facilitators gaze at HD and LD players with similar frequency and duration as both speaker and listener.

Regarding mutual gaze, where dominance is balanced, mutual gaze involving only players is less than mutual gaze

involving facilitators, contrary to our hypothesis. Where dominance is not balanced mutual gaze involving only players is less than mutual gaze involving facilitators, as hypothesized.

VI. CONCLUSION

In the context of MULTISIMO, dominant players spend more time looking at their fellow player than their less dominant counterparts and avert their gaze for longer periods on average. Though both dominant and submissive players gaze at the facilitator at similar rates, the dominant individual receives longer gazes on average than their fellow player. When aligned with dialogue role, gaze behaviour differences between players are maintained when the player in question occupies the role of speaker but not of listener. Dominant participants gaze away and at their co-participant for longer durations, while less dominant participants gaze longer at the facilitator when speaking. The present results address only a subset of the MULTISIMO corpus, for which the heuristic automatic gaze annotation approach was of sufficient quality. The results reported are compelling enough to suggest the next step of scaling to the full data set, and further test the validity our approach to predict dominant participants in new data.

ACKNOWLEDGMENT

REFERENCES

[1] P. Baranyi and A. Csapo, "Cognitive infocommunications: Coginfocom," in *2010 11th International Symposium on Computational Intelligence and Informatics (CINTI)*, 2010, pp. 141–146.

[2] P. Baranyi, A. Csapo, and P. Varlaki, "An overview of research trends in CogInfoCom," in *IEEE 18th International Conference on Intelligent Engineering Systems*, ser. INES, 2014, pp. 181–186.

[3] C. Vogel and A. Esposito, "Linguistic and behavior interaction analysis within cognitive infocommunications," in *10th IEEE Conference on Cognitive Infocommunications*, 2019, pp. 47–52.

[4] ——, "Interaction analysis and cognitive infocommunications," *Infocommunications Journal*, vol. XII, no. 1, pp. 2–9, 2020.

[5] C. Vogel, M. Koutsombogera, and A. Esposito, "Aspects of methodology for interaction analysis," in *11th International IEEE Conference on Cognitive Infocommunications*. This volume, 2020.

[6] R. S. Hessels, G. A. Holleman, A. Kingstone, I. T. Hooge, and C. Kemner, "Gaze allocation in face-to-face communication is affected primarily by task structure and social context, not stimulus-driven factors." *Cognition*, vol. 184, pp. 28 – 43, 2019.

[7] A. Kendon, "Some functions of gaze-direction in social interaction." *Acta Psychologica*, vol. 26, no. 1, pp. 22 – 63, 1967.

[8] M. Argyle and J. Dean, "Eye-contact, distance and affiliation." *Sociometry*, vol. 28, no. 3, pp. 289 – 304, 1965.

[9] D. Heylen, I. Van Es, A. Nijholt, and B. van Dijk, "Experimenting with the gaze of a conversational agent." in *Proceedings international CLASS workshop on natural, intelligent and effective interaction in multimodal dialogue systems*, 2002, pp. 93 – 100.

[10] N. Bee, C. Pollock, E. André, and M. Walker, "Bossy or wimpy: expressing social dominance by combining gaze and linguistic behaviors," in *International Conference on Intelligent Virtual Agents*. Springer, 2010, pp. 265 – 271.

[11] H. Itakura, "Describing conversational dominance." *Journal of Pragmatics: An Interdisciplinary Journal of Language Studies*, vol. 33, no. 12, pp. 1859 – 1880, 2001.

[12] L. McLaren, M. Koutsombogera, and C. Vogel, "A heuristic method for automatic gaze detection in constrained multi-modal dialogue corpora." This volume, 2020.

[13] M. Koutsomboger and C. Vogel, "The MULTISIMO multimodal corpus of collaborative interactions," in *9th ACM International Conference on Multimodal Interaction*. ACM, 2017, pp. 502–503.

[14] M. Koutsombogera and C. Vogel, "Modeling Collaborative Multimodal Behavior in Group Dialogues: The MULTISIMO Corpus," in *Proceedings of the 11th International Conference on Language Resources and Evaluation (LREC 2018)*, N. C. et al., Ed., 2018, pp. 2946–2951.

[15] P. Baranyi and A. Csapo, "Definition and synergies of cognitive infocommunications," *Acta Polytechnica Hungarica*, vol. 9, no. 1, pp. 67–83, 2012.

[16] P. Paggio and C. Navarretta, "Classifying the feedback function of head movements and face expressions," in *Multimodal Corpora: How Should Multimodal Corpora Deal with the Situation?* 8th Workshop on Multimodal Corpora, 2012, pp. 34–37.

[17] A. Esposito, A. M. Esposito, and C. Vogel, "Needs and challenges in human computer interaction for processing social emotional information," *Pattern Recognition Letters*, vol. 66, pp. 41–51, 2015.

[18] M. Koutsombogera and C. Vogel, "Observing collaboration in small-group interaction," *Multimodal Technologies and Interaction*, vol. 3, no. 3, p. 45, 2019.

[19] F. Haider, M. Koutsombogera, O. Conlan, C. Vogel, N. Campbell, and S. Luz, "An active data representation of videos for automatic scoring of oral presentation delivery skills and feedback generation," *Frontiers in Computer Science*, vol. 2, no. 1, pp. 1–13, 2020.

[20] C. Navarretta, "Predicting an individual's gestures from the interlocutor's co-occurring gestures and related speech," in *2016 7th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*, 2016, pp. 233–238.

[21] M. Argyle, "Eye-contact and direction of gaze," *The psychology of interpersonal behaviour. Harmondsworth, Penguin Books*, 1967.

[22] U. Malik, M. Barange, N. Ghannad, J. Saunier, and A. Pauchet, "A generic machine learning based approach for addressee detection in multiparty interaction," in *Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents*, 2019, pp. 119 – 126.

[23] W. L. Libby and D. Yaklevich, "Personality determinants of eye contact and direction of gaze aversion." *Journal of Personality and Social Psychology*, vol. 27, no. 2, pp. 197 – 206, 1973.

[24] M. Foddy, "Patterns of gaze in cooperative and competitive negotiation," *Human relations*, vol. 31, no. 11, pp. 925 – 938, 1978.

[25] M. S. Gobel, H. S. Kim, and D. C. Richardson, "The dual function of social gaze." *Cognition*, vol. 136, pp. 359 – 364, 2015.

[26] S. Norris, *Analyzing multimodal interaction: A methodological framework*. Routledge, 2004.

[27] H. Akechi, A. Senju, H. Uibo, Y. Kikuchi, T. Hasegawa, and J. K. Hietanen, "Attention to eye contact in the west and east: Autonomic responses and evaluative ratings," *PLOSone*, vol. 8, no. 3, p. e59312, 2013.

[28] M. Gobel, A. Chen, and D. Richardson, "How different cultures look at faces depends on the interpersonal context." *Canadian Journal of Experimental Psychology*, vol. 71, no. 3, pp. 258 – 264, 2017.

[29] D. McNeill, "Gesture, gaze, and ground," in *International workshop on machine learning for multimodal interaction*, 2005, pp. 1 – 14.

[30] Y. Fukuhara and Y. Nakano, "Gaze and conversation dominance in multiparty interaction." in *2nd workshop on eye gaze in intelligent human machine interaction*, vol. 9, 2011, pp. 9 – 16.

[31] R. Costello, "Analysing dominance in multi-party dialogue," Bachelor's Thesis, Trinity College Dublin, 2018.

[32] M. Koutsombogera, R. Costello, and C. Vogel, "Quantifying dominance in the multisimo corpus," in *2018 9th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*, 2018, pp. 141 – 146.

[33] C. Vogel, M. Koutsombogera, and R. Costello, "Analyzing likert scale inter-annotator disagreement," in *Neural Approaches to Dynamics of Signal Exchanges*. Springer, 2020, pp. 383 – 393.