# Interaction of motion energy with gesture, extroversion, dominance, and collaboration in dialogue

Zohreh Khosrobeigi
*Computational Linguistics Group*
*Trinity College Dublin,*
*University of Dublin*
Dublin, Ireland
khosrobz@tcd.ie

Maria Koutsombogera
*Computational Linguistics Group*
*Trinity College Dublin,*
*University of Dublin*
Dublin, Ireland
koutsomm@tcd.ie

Carl Vogel
*Computational Linguistics Group*
*Trinity College Dublin,*
*University of Dublin*
Dublin, Ireland
vogel@cs.tcd.ie

***Abstract*—This paper studies the relationship between conversational features related to gesture, conversational dominance, collaboration, and personality traits. Specifically, we examine the interaction of motion energy and factors associated with dialogues and participants therein. We observe among semiotic types of gesture in dialogue higher motion energy in Beats, Deictics and Iconics than Symbolics, but lower in Beats than arbitrary hand movements. Dominance and collaboration are correlated with motion energy when contentful semiotic gesture types are excluded. Different collaboration scores present different associations on the level of motion energy. The findings quantify the extent to which the motion energy of participants contributes to judgments of dominance and collaboration in dialogue.**

***Index Terms*—Motion Energy, Interaction, Collaboration, Gesture, Dominance.**

## I. Introduction

In extending observational research in the analysis of natural language dialogue through analysis of multi-modal dialogue corpora [1], we demonstrate the value of examining motion energy (ME) as recorded from video. The concept of ME is made operational through quantification of pixel changes from one frame of a video to its successor. Simply, each pixel in a gray-scale frame has a value between 0 and 255, and pixel change entails a change in this value. Observational approaches involve a combination of theory-guided inspection of relationships among relevant features and hypothesis testing within that exploration. Theories of astrophysics are developed with reference to observational data, and theories are revised when extant theories cannot account for data observed. Arguably, the theoretical basis of astrophysics is more advanced than that of natural language dialogue, including theories of the perception of collaboration in natural language dialogue. In the age of deep-fake video technology, it becomes increasingly relevant to be able to discriminate between genuine dialogue interactions and fabricated interactions, we think it is all the more important to have a base of understanding of patterns that emerge in natural dialogue. Here, the main quality that we analyse is ME, made operational as described above, and in relation to other features of dialogue and participants in dialogue. It is natural to understand ME as influenced by changes of posture, head movement and gesture captured within the frame of view and thus provides a coarse-grained measure of such activities. Additionally, these activities may be expected to interact with other properties of dialogues and their participants (for example, among personality traits, extroversion may be anticipated to correlate with ME). We explore the interactions between ME and other factors associated with the classification of dialogues (dominant players, collaboration) and participants within dialogues (dominance score, personality trait) as part of a validation of the method of using ME measures. We follow observational methods in which we explore a substantial corpus of group dialogues by measuring the relations among qualities that emerge from dialogue participants in relation to ME as they interact.

Interaction refers to actions of objects that affect each other [2], and it is applied in human interaction analysis. We study dialogues for which independent measures and categorizations are available: participant dominance, participant personality traits, and collaboration. Some of these are measures of individuals, but these may also apply to dialogues as a whole. In analyzing the interactions with ME, we sometimes consider ME as a continuous value and sometimes in ordinal categories low, medium, and high (determined by statistical properties, i.e., quartiles). We proceed by showing the effects associated with these construals of ME, commenting on the evaluation of our expectations along the way.

First, we describe past research that analyses motion energy in multi-modal recordings of dialogue interactions (§II). We then describe a multi-modal corpus of interactions used in this research (§III). We detail the intuitions that we have about effects involving ME and other qualities of dialogues and dialogue participants (§IV). We describe the methods used to analyse the dialogue data (§V). Then we present results based on those methods (§VI) and discuss our interpretation of these results (§VII). The paper concludes (§VIII) with observations about the relevance of this work to multi-media computing.

## II. Background

Research on small-group interaction and group dynamics has recently focused on exploring both spoken words and nonverbal channels to develop methods that automatically analyse group interaction and models able to predict information about the participants as well as the state of the interaction, including collaboration assessments [3]–[5], but without analyzing the impact of gesture, for example, among nonverbal channels. Some have studied small collaborative learning groups, measuring the quality of collaboration using the group participants' movements and its correlation to the outcomes of the tasks at hand [6]. Others have analysed human personality traits with the motivation of enabling robots to have a better understanding of human personality. A multi-layer Hidden Markov Model has been applied to improve the classification accuracy of personality traits [7], presenting the benefits of fusing multiple features such as head motion, gaze, and other body motions. Researchers pursue enabling conversational agents to utilise multi-modal cues from interlocutors to adapt their behaviours; a method to synthesise interlocutor-aware facial gestures in dyadic conversations has been proposed [8], where features are extracted from multi-party video and speech recordings. A dialogue participant's interpreting interlocutors' movements in conversation is more likely to result in successful communication and interaction than if those behaviours are ignored. As discussed previously [9], it is an open question whether different modalities should be studied in isolation regarding their individual contributions to the meaning, or collectively, to better understand the subtle interactions between their timing, scope, and ability to convey propositional and other semantic content [10].

A topic in group dynamics is behavioural synchronisation. Physical systems that cannot plausibly be ascribed volition synchronise [11]. Synchronisation among volitional agents may be assumed to have a non-volitional element but also intentional elements. Whether the non-volitional or volitional elements dominate the determination of success in achieving joint tasks through dialogue remains an open question. Tasks like the HCRC Map Task were constructed precisely to make volitional dialogue prerequisite to success [12]. However, the "success" of most natural dialogues probably cannot be ascribed to task-based success, but rather to more nebulous qualities like positivity and collaboration among participants.

Human behavioural synchronisation is coordinative interaction that may be assumed to have among its volitional components conscious efforts towards maintaining social relations, but also components that are probably not all conscious, and which nonetheless impinge on social relations [13], [14]. In highlighting temporal aspects of communication, such as rhythm, the meshing of nonverbal behaviours, and simultaneous movement as quantitative characteristics, researchers classified this type of synchrony as movement synchrony and proposed the Motion Energy Algorithm (MEA) and cross-correlation to measure synchrony in psychotherapies [15]. The results show synchrony on a global level, regardless of the specific body parts moving. So synchrony can be a general measure of movement coordination between individuals' interactions. MEA and cross-correlation are also used to measure whether the synchrony of human-human interaction is best explained by chance [16], showing that synchrony goes beyond random coincidence. Independently it has been noted that constituents of motion energy, gesture, and gesture types, show systematic properties in aligning with the syntactic categories of nearby' words [17]. However, the circle of relationships is not yet closed. It remains to be understood, for example, how ME exhibited by dialogue, participants relates to dialogue external perceptions of collaboration, how ME is elicited by distinct semiotic types of gestures used in conversation, how ME relates to dialogue participant personality traits and external perception of conversational dominance, and so on. This work is intended to provide more observations about some of the overlapping arcs in this circle.

## III. Dataset

For this study, MULTISIMO corpus is used [18], a multi-modal dataset of three-party, task-based dialogues. The dataset consists of 18 dialogue sessions in English. In each session, two players collaborate to identify the three most popular answers to three quiz questions provided by a moderator, and to rank them in terms of popularity, following the responses of that external groups. Two players were randomly partnered with each other in each dialogue, and the third participant served as the moderator for the discussion. There are 39 dialogue participants, three of whom share the role of moderator in the sessions. The dataset includes the video and audio of the dialogues and a set of annotations, such as speech transcripts, gaze, laughter, gesture annotations, and word-spoken timestamps [17]. Out of the several available video and audio formats, versions of high-quality videos are used for this study, selected to have a zoomed front view of each participant. The dataset includes additional annotations about participants, such as dominance scores, personality trait scores, and about dialogue sessions such as collaboration scores.

Dominance scores are about perceived dominance levels of the players involved in the dialogues, as assessed by five external raters, who provided their ratings after watching the dialogue videos. Assessment is done on a scale from 1 (not at all dominant) to 5 (very dominant). In [4], the authors have measured the reliability of the given ratings showing a good level of agreement among raters.

Collaboration refers to the process where the two players coordinate their actions to achieve their shared goal, i.e., find the appropriate answers to the quiz questions and rank them in terms of popularity. Collaboration scores were assigned by two annotators and range from 1 (low collaboration) to 4 (high collaboration).

The Big Five Inventory (BFI-44), a self-report inventory consisting of 44 items (statements) [19], [20] was adopted to measure the Openness, Conscientiousness, Extroversion, Agreeableness, and Neuroticism personality traits of participants. Before the dialogue recordings, participants completed the inventory by rating each statement to indicate the extent to which they agree or disagree with it.

Hand gestures were manually annotated using the ELAN editor [21]. Gestures are annotated in their entire duration, and are assigned with one of the following semiotic types: Beat, Iconic, Deictic and Symbolic, based on McNeill's classification [22]. In addition, the label N/A is used for visible hand movements, which, however, do not have a communicative function in the dialogue.

# IV. Hypotheses

The dialogue annotations include some qualities that arise directly from the participants (for example, personality trait scores using standard Big-5 personality assessment instruments) and some that arise from annotators, independent of the participants (e.g., dominance and collaboration). Furthermore, some of the annotations are linked to individual participants (e.g., personality traits and individual dominance scores) while others pertain to the dialogue as a whole (e.g., balance of dominance, collaboration); for example, given dominance scores of both participants in a dialogue, the dialogue as a whole may be characterised in relation to the participant on the left side of the monitor being most dominant, or the player on the right, or of displaying balanced dominance.

Variables anchored in ME also apply at the level of individual participants and in aggregate, for a whole dialogue. Here, we only analyze ME of individual participants. We imagine that bodily motion, and hence indices of bodily motion supplied by ME will have systematic relationships with these qualities. Of the personality traits, we expect a positive correlation between extroversion and ME. We distinguish the ME that accompanies gesture with clear semiotic types (Beat, Symbolic, Deictic, and Iconic) from arbitrary hand movements (N/A) and non-gestural movement (Agestural). Intuitively, a hand movement is easier to perceive as a meaningful gesture if it accounts for most of a person's movements at the time of gesture than if the person gestures at the same time as moving a number of body parts. However, a hand gesture may also be constructed with whole-body involvement, especially when the person is rather engaged with the content being expressed. Given the nature of the collaborative task in the dialogues analysed here, we expect the former relationship.

That is, we expect least ME among the gestural types with clear semiotic content and most ME for the complement. We expect to see positive correlations between ME and dominance, between ME and collaboration, and between ME and extroversion. In connection with ME and dominance, we expect that relation to be strongest where ME is not accompanied by gestures with clear communicative content (that the perception of dominance is influenced by ME without meaningful gesture). In contrast, we expect the relationship between ME and collaboration to be strongest where ME is accompanied by meaningful gesture.

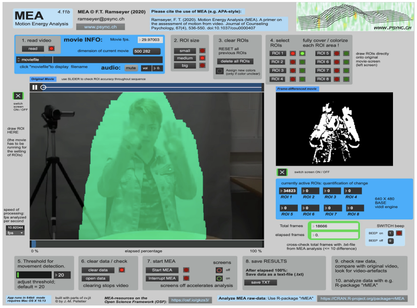
# V. Method

In this research, we are interested in investigating the interaction of ME (the bodily motion of participants in the conversation) with dialogues' qualities and participants therein. We design tests associated with the hypotheses described above (§IV) and explore their outcomes. In the first step, some preprocessing was performed, explained in the following. Next, the hypotheses are tested.

## A. Pre-Processing

*1) Motion Energy Computing:* A video is a series of still images, and each image is called a frame. Each pixel has a colour that changes in the next frame if a change happens, such as an object's movement or light changes. Hence, the next frame is different from the current one if an object moves. This difference is defined as ME [23], and to compute it, the number and amount of pixel changes are summed between each frame and its prior frame. The ME of the first frame is zero since there is no prior frame and no differences. This method falls on frame-difference algorithms which quantify the changes across time while not considering the direction of movement [23].

Many factors affect the pixel colour and computing movement, such as the environment light, background colour, and noise. The camera, background, and light should be fixed while using frame-difference methods since any change of the camera or background would be considered as object movement, and light can affect the colour of the pixel. The reason is that frame-difference algorithms recognize pixel changes and not objects. Only targets can move, neither the camera nor the background. In addition, objects should not mask each other, and the boundaries of each object should be clear. There is a notion in frame-difference algorithms, the region of interest (ROI), where target areas are defined to compute their movements and changes. Because the camera and environment conditions are constant in MULTISIMO, and only players move, we could use MEA [23]–[25]. The MEA tool is used for computing ME of frames. The body of the player is defined as the ROI through the MEA tool, and the tool computes the motion of each frame. Figure 1 illustrates ROI and its motion in binary image format.



MEA 4.11b Motion Energy Analysis

MEA © F.T. Ramseyer (2020) ramseyer@psync.ch www.psync.ch

Please cite the use of MEA (e.g. APA-style):

Ramseyer, F. T. (2020). Motion Energy Analysis (MEA). A primer on the assessment of motion from video. Journal of Counseling Psychology, 67(4), 536-550. doi:10.1037/cou0000407

movie INFO: Movie fps. 29.97003

dimension of current movie 500 282

click "moviefile"to display filename

4. select ROIs — fully cover / colorize each ROI area !

draw ROIs directly onto original movie-screen (left screen)

use SLIDER to check ROI accuracy throughout sequence

draw ROI HERE (the movie has to be running for the setting of ROIs)

speed of processing: fps analyzed per second 10.92044 fps

currently active ROIs: quantification of change

| ROI 1 | ROI 2 | ROI 3 | ROI 4 |
|---|---|---|---|
| 34823 | 0 | 0 | 0 |

| ROI 5 | ROI 6 | ROI 7 | ROI 8 |
|---|---|---|---|
| 0 | 0 | 0 | 0 |

640 X 480 BASE viddl engine

Total frames 18666

elapsed frames 0.

cross-check total frames with .txt-file from MEA analysis (<= 10 difference)

5. Threshold for movement detection. 20 adjust threshold; default = 20

6. clear data / check — clearing stops video

7. start MEA — screens off accelerates analysis

8. save RESULTS — After elapsed 100%: Save data as a text-file (.txt)

9. check raw data, compare with original video, look for video-artefacts

10. analyze data with e.g. R-package "rMEA"

App runs in 64bit mode requires Mac OS X 10.13

built with parts of cv.jit © by J.-M. Pelletier

MEA-resources on the Open Science Framework (OSF): https://osf.io/gkzs3/

Analyze MEA raw-data: Use R-package "rMEA" https://CRAN.R-project.org/package=rMEA

Fig. 1. The MEA application: Green areas show an ROI frame of a player.

The ME extracted is normalized across all sessions using the MinMaxScaler function which preserves the shape of the original distribution.

*2) Gesture Frames Extraction:* Timestamps of gestures are extracted using ELAN. Each gesture timestamp in the time unit (seconds) is converted to its corresponding frames, a frame unit. The value of a frame is either Beat, Iconic, Deictic, Symbolic, or N/A, if a gesture happens in the frame, or Agestural, if there is no gesture in the frame.

### *B. Conversational Features*

As mentioned in §III, there exist five dominance scores for each player assigned from external raters. To come up with a single score per player, the five dominance scores are aggregated by using the median of scores. Moreover, the collaboration scores of each session are aggregated by computing the median of the session's scores. Of six different personality traits, the extroversion trait is explored. Each player's raw score may range from 8 to 40.

### *C. Quantities and categories*

Relationships are explored in relation to graduated measurements and in relation to categories. Examples of the former are ME, extroversion, dominance, and collaboration. With these, we construct non-parametric (Spearman) correlation coefficients in order to quantify co-variability, and we complement these with the use of non-parametric tests of difference (Wilcoxon and Kruskal-Wallis). Examples of the latter are gesture types and ordered categories derived from descriptive statistics (e.g., quartiles) associated with the graduated measurements. With these, we analyse whether contingency tables that cross-classify observations between two sorts of classification reveal statistically significant interactions. Where interactions are significant, we explore the residuals – based on the difference between the observed values and values that would be expected without interaction between the ME categories and cross-classified categories – in order to determine the loci of interactions. Analyses based on ordered categories are a useful supplement to correlations between numeric quantities because they enable the identification of where in the overall range correlations and deviations are anchored.

## VI. Results

To have categories of different levels of ME, descriptive statistics of ME is computed as follows: minimum, 0.0; first quartile, 0.0; median, 0.00010; mean, 0.2076; third quartile, 0.01110; maximum, 1.0. Then, descriptive statistics are used to form categories of ME corresponding to "Low", "Medium" and "High", where Low has the range [0,0.0001]; Medium has the range (0.0001,0.0111]; and High has the range (0.0111,1].

### *A. Motion Energy and Gesture Type*

In this section, we report tests on ME and gesture type. Table I shows the mean and standard deviation (SD) of ME for each gesture type. The mean of ME is lower for Agestural moments than for moments accompanied by Deictic, Iconic, and Symbolic gestures.

TABLE I
Data profile of ME for each gesture Type.

| | **Agestural** | **Beat** | **Deictic** | **Iconic** | **N/A** | **Symbolic** |
|---|---|---|---|---|---|---|
| Mean | 0.0134 | 0.0717 | 0.0966 | 0.0967 | 0.0749 | 0.1490 |
| SD | 0.0420 | 0.0835 | 0.0976 | 0.1060 | 0.0969 | 0.1430 |

We explore whether the distinct gesture types are characterised by corresponding differences in ME by applying a Kruskal test. Its result shows the overall contrast is significant (Kruskal-Wallis $\chi^2 = 87795, df = 5, p < 2.2e - 16$). Because a Kruskal test only shows whether the contrast is significant and does not show where, a pairwise Wilcoxon test is applied to calculate pairwise comparisons between group levels with corrections for multiple testing. The results (see Table II) demonstrate the measurements of ME are mutually distinct for most pairwise comparisons of gesture types (Bonferroni adjustment is applied) – Beats are not significantly different from arbitrary hand movements (N/A) on this measure, nor are Deictics and Iconics significantly different from each other in ME.

TABLE II
P-values of pairwise Wilcoxon tests of ME of each Gesture Type. Where $p < 2e - 16$, we write ***; NS = not significant.

| | **Agestural** | **Beat** | **Deictic** | **Iconic** | **N/A** |
|---|---|---|---|---|---|
| Beat | *** | - | - | - | - |
| Deictic | *** | *** | - | - | - |
| Iconic | *** | *** | NS | - | - |
| N/A | *** | NS | *** | *** | - |
| Symbolic | *** | *** | *** | *** | *** |

Ordinal categories of ME levels constructed from the descriptive statistics of ME are used to test the interaction between levels of energy and gesture type. The interaction is computed using Chi-square test, and it shows significant interaction (Chi-square $\chi^2 = 92356, df = 10, p - value < 2.2e - 16$). In inspecting significance of residuals, we apply Bonferroni adjustment to $\alpha = 0.05$ for 18 comparisons, $\alpha' = 0.0028$, and the critical value for $N(0,1)_{\alpha'} = 3$. Table III represents standard Residual of the test and indicates the two lower categories of ME (which has lower energy), Low, Medium, are dominated by observations of Agestural; there is a lack of observations of Agestural in the last category, High ME. On the other hand, a lack of observations of gesture types can be seen for Low ME, and their observations are more than would be expected with no interaction in High ME.

### *B. Motion Energy and Dominance*

In this section, we test the relationships between ME and player dominance as perceived by independent observers of the conversations.

TABLE III
STANDARD RESIDUALS OF ME CATEGORIES AND GESTURE TYPES.

| | ME Categories | | |
|---|---|---|---|
| **Gesture** | **Low** | **Medium** | **High** |
| Agestural | 219.0140 | 44.0606 | -297.1240 |
| Beat | -115.1510 | -20.4669 | 153.5210 |
| Deictic | -50.6701 | -13.6609 | 72.2081 |
| Iconic | -114.3940 | -35.9585 | 168.1340 |
| N/A | -120.1600 | -11.5965 | 150.4400 |
| Symbolic | -30.3285 | -12.2432 | 47.2857 |

TABLE IV
DATA PROFILE OF DOMINANCE SCORE FOR EACH ME CATEGORY.

| | ME Categories | | |
|---|---|---|---|
| **Gesture** | **Low** | **Medium** | **High** |
| Mean | 2.66468 | 2.854358 | 2.88769 |
| SD | 0.9887806 | 0.9135327 | 0.9050514 |

Firstly, we note that across all individual participants, the rank correlation between ME and dominance overall: Spearman's $\rho = 0.099, p < 2.2e - 16$. Restricting focus to those frames for which gesture conveys a clear semiotic type (that is, ignoring Agestural frames and frames with arbitrary hand movements), there is no significant correlation ($\rho = 0.008, p = 0.11$). Thus, the correlation between ME and dominance stems from the ME accompanying moments without gesture and arbitrary hand motions ($\rho = 0.106, p =< 2.2e - 16$).

Table IV illustrates the profile of dominance scores for each level of ME. The difference in dominance scores across the ordinal categories of ME is significant (Kruskal-Wallis $\chi^2 = 5814.1, df = 2, p < 2.2e - 16$). Furthermore, Table V shows the measurements dominance scores are significantly different between each of the pairs of levels of ME categories.

An interesting categorical view of dominance score annotation turns these values into assessments of the whole dialogue: are the players equally dominant (Balance), or is the player on the left side of the monitor more dominant (left-Dominance), or is the player on the right more dominant (right-Dominance)? Table VI shows the mean, and SD of ME of dominant players defined by their position on the screen in the view onto dialogue that includes both players, left or right, or if both players have equal dominance scores. As the table illustrates the greatest mean ME belongs to Balance (co-players have equal dominance scores), and the least mean of ME occurs where the player on the left side is perceived as most dominant. To see whether the distinct dominant player categories are characterized by differences in ME, Kruskal test is applied, which shows overall contrast is significant (Kruskal-Wallis $\chi^2 = 8198.1, df = 2, p < 2.2e-16$). In addition, Table VII shows the pairwise comparisons of ME within each of the dominance categories to yield significant differences.

TABLE V
P-VALUES OF PAIRWISE WILCOXON TESTS OF DOMINANCE SCORES OF EACH ME CATEGORY. WHERE $p < 2e - 16$, WE WRITE ***.

| | ME Categories | |
|---|---|---|
| | **Low** | **Medium** |
| Medium | *** | - |
| High | *** | 6.9e-11 |

TABLE VI
DATA PROFILE OF ME ASSOCIATED WITH DOMINANCE CATEGORIES.

| | **Balance** | **left-Dominance** | **right-Dominance** |
|---|---|---|---|
| Mean | 0.0268 | 0.0168 | 0.0237 |
| SD | 0.0630 | 0.0488 | 0.0595 |

TABLE VII
P-VALUE OF PAIRWISE WILCOXON TESTS OF LEVELS OF ME WITHIN THE CONVERSATION-LEVEL DOMINANCE CATEGORIES. WHERE $p < 2e - 16$, WE WRITE ***

| | **Balance** | **left-Dominance** |
|---|---|---|
| left-Dominance | *** | - |
| right-Dominance | *** | *** |

The interaction of dominance categories and ME categories is computed as well. Table VIII shows results of the standard Residuals (Chi-square $\chi^2 = 7613.1, df = 4, p < 2.2e - 16$). Applying Bonferroni adjustment to $\alpha = 0.05$ for 9 comparisons, $\alpha' = 0.0056$, and the critical value for $N(0, 1)_{\alpha'} = 2.7$. There is an evident dearth of instances of Balance and Dominant Player on the right with Low ME. There are more observations of Balance and right-Dominance and Low ME than would be expected with no interaction among these categories. There are more observations of left-Dominance with Low ME and also a dearth of observations of left-Dominance with Medium and High ME than would be expected with only random interactions of these categories.

### C. Motion Energy and Collaboration Score

One might expect to see high ME in conversations with high collaboration and low ME with low collaboration that is, a positive correlation between collaboration and ME. Figure 2 demonstrates the frequency of collaboration scores in the dataset. The profile of collaboration scores within each ME ordinal category is shown in Table IX, and this indicates that as ME increases, the mean of collaboration score increases.

The rank correlation between ME overall and collaboration scores is small but significant ($\rho = 0.131, p < 2.2e - 16$). The correlation between ME and collaboration scores during gesture of contentful semiotic types are more weak but still significant ($\rho = 0.087, p < 2.2e - 16$); the correlation where

TABLE VIII
STANDARD RESIDUALS ARISING FROM THE CROSS-TABULATION OF ME CATEGORIES AND DOMINANCE CATEGORIES.

| | ME Categories | | |
|---|---|---|---|
| | **Low** | **Medium** | **High** |
| Balance | -50.2667 | 13.6211 | 44.4650 |
| left-Dominance | 80.4016 | -26.8740 | -66.0359 |
| right-Dominance | -47.7037 | 18.1562 | 36.9692 |

TABLE IX
DATA PROFILE OF COLLABORATION SCORE FOR EACH ME CATEGORY.

| | ME Categories | | |
|---|---|---|---|
| **Gesture** | **Low** | **Medium** | **High** |
| Mean | 2.424 | 2.618 | 2.686 |
| SD | 0.85 | 0.8858 | 0.9125 |

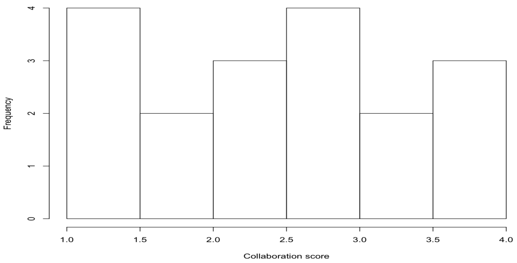


Fig. 2. Histogram of collaboration score depicting the distribution of different collaboration levels.

ME accompanies no gesture or only arbitrary hand movement is greater ($\rho = 0.156, p < 2.2e-16$).

Overall contrast in collaboration scores within the ME ordinal categories is significant (Kruskal-Wallis $\chi^2 = 9642, df = 2, p < 2.2e-16$). Pairwise comparisons between scores and ME categories show collaboration scores are mutually distinct for all ME categories (Table X).

### D. Motion Energy and Personality Traits

In this section, we test the hypothesis that high scores for the extroversion trait correspond to high ME of dialogue participants. Figure 3 illustrates the distribution of extroversion scores in the dataset. Table XI depicts the profile of raw extroversion scores within each of the three ordinal ME categories.

The Spearman correlation of extroversion score and ME is computed to find whether there is a correlation between these ($\rho = -0.022, p < 2.2e-16$). The correlation is negative and close to zero but significant. Restricting attention to ME during gesture of clear semiotic type, the negative correlation is stronger and significant ($\rho = -0.083, p < 2.2e-16$). The complement, the correlation of ME during frames without gesture or with arbitrary hand movement, involves a weaker negative rank correlation (Spearman's rank correlation $\rho = -0.026, p < 2.2e-16$).

TABLE X
P-VALUES OF PAIRWISE WILCOXON TESTS OF COLLABORATION SCORES OF EACH ME CATEGORY. WHERE $p < 2e-16$, WE WRITE ***.

| | ME Categories | |
|---|---|---|
| | **Low** | **Medium** |
| Medium | *** | - |
| High | *** | *** |

TABLE XI
DATA PROFILE OF EXTROVERSION SCORE FOR EACH ME CATEGORY.

| | ME Categories | | |
|---|---|---|---|
| | **Low** | **Medium** | **High** |
| Mean | 23.9 | 23.6 | 23.5 |
| SD | 6.05 | 5.88 | 6.11 |

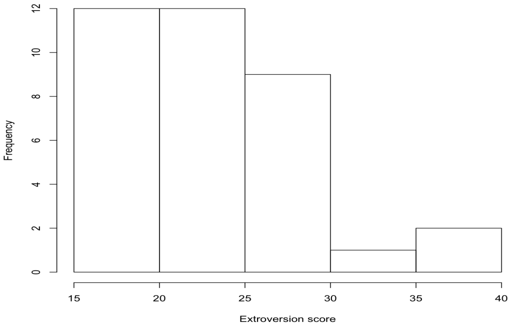


Fig. 3. Histogram of the extroversion scores depicting the distribution of different levels of scores in the dataset.

TABLE XII
P-VALUES OF PAIRWISE WILCOXON TESTS OF EXTROVERSION SCORES OF EACH ME CATEGORY. WHERE $p < 2e-16$, WE WRITE ***.

| | ME Categories | |
|---|---|---|
| | **Low** | **Medium** |
| Medium | *** | - |
| High | *** | 1.9e-05 |

The difference in extroversion scores according to ordinal categories of ME is significant (Kruskal-Wallis $\chi^2 = 195.9, df = 2, p < 2.2e-16$). Pairwise Wilcoxon test is shown in Table XII which shows contrast is significant for all categories. Noting the correlation facts reported above and the value of the mean extroversion score in relation to the ME categories is shown in Table XI, we consider the means with the additional distinction between those moments involving no gesture or only arbitrary hand movements and the moments that involve gestures with clear semiotic types (Table XIII).

TABLE XIII
MEAN OF EXTOVERSION BY ME CATEGORY FOR GESTURE WITH A CLEAR SEMIOTIC TYPE VS. WITH NO GESTURE OR ONLY ARBITRARY HAND MOVEMENT

| | ME Categories | | |
|---|---|---|---|
| Gesture Types | Low | Medium | High |
| Beats + Deictics + Iconics + Symbolics | 24.49 | 24.67 | 23.96 |
| Agestural + N/A | 23.86 | 23.56 | 23.37 |

Our primary interest here is extroversion. However, to contextualize the result associated with our specific hypothesis, we also report the correlations between ME and raw scores for the other personality traits. Table XIV presents the full set of correlations. Notice that the strongest positive correlations with ME are connected to agreeableness, and the strongest negative correlations with ME are connected to neuroticism. The correlations in all cases are significant but weak.

TABLE XIV

Spearman correlations between ME and personality trait raw scores: *** signifies $p < 0.001$; Full Me signifies ME for all frames; ME+Gestures signifies a restriction to ME for those frames in which the semiotic type of the gesture is contentful: deictic, symbolic, iconic, beat; ME+NonGestural signifies a restriction to ME in frames without gesture or with arbitrary hand movements (N/A).

| | Full ME | | ME+Gestures | | ME+NonGestures | |
|---|---|---|---|---|---|---|
| Trait | $\rho$ | $p$ | $\rho$ | $p$ | $\rho$ | $p$ |
| Extroversion | -0.022 | *** | -0.083 | *** | -0.026 | *** |
| Agreeableness | 0.077 | *** | 0.064 | *** | 0.083 | *** |
| Conscientiousness | 0.011 | *** | -0.049 | *** | 0.035 | *** |
| Openness | -0.019 | *** | -0.017 | *** | -0.039 | *** |
| Neuroticism | -0.082 | *** | -0.019 | *** | -0.092 | *** |

## VII. Discussion

The results (§VI-A) have indicated that ME levels in this dataset are greater in the company of hand gestures of clear semiotic type (Beat, Deictic, Symbolic, and Iconic) than during moments without gesture. ME levels during moments with arbitrary hand movements are not significantly different from ME levels during Beats. We understand these results to signify that in the setting of the data analyzed, extensive bodily movements are reserved for acts of communication with gesture.

External perception of dominance levels of participants (§VI-B) appears to be influenced by the ME of participants. However, the relation dominance and ME seems to arise from the ME that is not accompanied by meaningful gesture: dialogue-external judgments of the speaker dominance positively correlated, weakly, with ME during moments without gesture or with only arbitrary hand movements, but not the complement. One might hypothesize high ME can lead to a greater dominance score, even when the level of extroversion trait is not high. The relation between dominance score and ME is statistically significant.

The interaction between ME and dominance categories of sessions that capture the relative dominance of the participants open further questions: situations in which observers rated the player on the right more dominant than the player on the left involved significantly greater ME values. This suggests that some other asymmetries between the player on the left and right must be at play – perhaps more spoken content from the player on the left, more moderator attention to the player on the left, or possibly it reveals an asymmetry in perception from those who rated dominance (like an unconscious "preference" for seeing the person on the left as more dominant with less ME evidence), or something of a similar sort – in order for the situations in which the player on the left is more dominant to involve significantly less ME than situations in which the player on the right is more dominant.

The dialogue-external judgments of collaboration (§VI-C) also correlated positively with ME during moments without gesture or with only arbitrary hand movements. These results together suggest that external perceptions of dominance and collaboration are more positively influenced by the degree of bodily motion not devoted to gestural communication acts than to ostensible acts of communication, somewhat surprisingly.

In examining the participant-internal personality traits, extroversion in particular (§VI-D), we found significant differences in extroversion for each of the ordinal categories of ME. However, the correspondence was not monotonic. The highest levels of extroversion corresponded to the middle levels of ME. We expected a more clear correlation between ME and extroversion.

The totality of these tests show that visible ME has an explanatory role in understanding the perceptions of key features of dialogues (dominance and collaboration) and their participants according to outside spectators. ME also appears to interact with participant-internal qualities, such as personality traits. "Explanatory" value associated with ME appears to depend on a distinction between ME accompanying gesture with clear semiotic type and non-gesture or only arbitrary hand movements. A substantial caveat to note is that these findings are tied to the data analyzed, and the relationships we have noticed may be anticipated to vary if other multimodal data sets are explored. However, the results here give strong suggestions about relations to test in other data sets.

We note that a factor that may affect a person's expressivity and hence body movement in conversation, is culture. Culture specificity may account for the fact that there are differences among speakers in the quantity or the intensity of the gestures and overall body movements in which they are involved while participating in a dialogue. MULTISIMO was not designed to control for cultural backgrounds. We do acknowledge though that the dialogue participants span eighteen nationalities and that one-third of them are native English speakers. We also acknowledge that the communicative behavior of people living in a foreign country may be subject to change depending on the time they spend in that country. We did not perform any experiment to explore the interaction of ME and conversational features according to the cultural background of participants; however, it would be relevant to consider the related binary classification: native speaker of English (the language of participation) vs. speaker of English as a non-native language.

## VIII. Conclusions

Our aim has been to clarify the relationship between ME produced by a dialogue participant and other qualities of the participant and dialogue, both internally and externally determined. An internal participant quality is degree of extroversion, as independently revealed by participants' engagement with personality testing. External qualities are dominance ratings provided by independent observers and also ratings of overall collaboration assigned to the dialogues in which participants engage. We examined ME within distinct categories of gesture. Meaningful gestures are revealed as higher in ME than moments without gesture and moments with just arbitrary hand movements. However, we see that dominance and collaboration ratings are positively correlated with ME during moments without meaningful gesture, more so than moments with meaningful gesture. We also see that there is a negative relationship between extroversion and meaningful gesture.

While these results are not guaranteed to emerge in other data sets, there is nothing about the present data set to suggest that it is remarkable with respect to gesture or general ME. It is tempting to conclude that judgments of dominance and collaboration are informed by the overall energy displays of participants and not by energy display devoted to communication, a somewhat surprising outcome. Presumably, in a more adversarial dialogue context, the relationships noted here would not emerge. However, we note that the present work contained empirical surprises and therefore does not rule out a further surprise in analyzing ME in conflict situations.

This work is, of course, not the final word on the interaction between ME accompanying meaningful gesture or ME outside gesture with other properties of dialogues or dialogue participants. However, the findings are relevant to multi-modal computing, particularly remote assessment of dialogues that do not necessarily have access to the linguistic content shared. "Content-free" analysis (see [26]) of dialogue is also relevant in contexts where it is necessary to protect confidentiality in the underlying content.

## Acknowledgments

This work was conducted with the financial support of Science Foundation Ireland under Grant No. 18/CRT/6223 and was supported by the GEHM research network (Independent Research Fund Denmark grant 9055-00004B). We are grateful to the anonymous reviewers who provided helpful constructive feedback on an earlier draft of this work.

## References

[1] C. Vogel, M. Koutsombogera, and A. Esposito, "Aspects of methodology for interaction analysis," in *11th IEEE International Conference on Cognitive Infocommunications (CogInfoCom2020)*, 2020, pp. 141–146.

[2] N. Sebanz and G. Knoblich, "Prediction in joint action: What, when, and where," *Topics in cognitive science*, vol. 1, no. 2, pp. 353–367, 2009.

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

[4] M. Koutsombogera, R. Costello, and C. Vogel, "Quantifying dominance in the multisimo corpus," in *9th IEEE International Conference on Cognitive Infocommunications (CogInfoCom 2018)*, P. Baranyi, A. Esposito, P. Földesi, and T. Mihálydeák, Eds. IEEE, 2018, pp. 147–152.

[5] C. Vogel, M. Koutsombogera, and J. Reverdy, "Aspects of dynamics in dialogue," *Electronics*, vol. 12, no. 10, p. 2210, 2023, special Issue: *Virtual Reality, Augmented Reality and the Metaverse for Enhanced Human Cognitive Capabilities*, edited by Adám Csapó and Mika Luimula.

[6] J. M. Reilly, M. Ravenell, and B. Schneider, "Exploring collaboration using motion sensors and multi-modal learning analytics." *International Educational Data Mining Society*, pp. 16–20, 2018.

[7] Z. Shen, A. Elibol, and N. Y. Chong, "Multi-modal feature fusion for better understanding of human personality traits in social human–robot interaction," *Robotics and Autonomous Systems*, vol. 146, p. 103874, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0921889021001597

[8] P. Jonell, T. Kucherenko, G. E. Henter, and J. Beskow, "Let's face it: Probabilistic multi-modal interlocutor-aware generation of facial gestures in dyadic settings," in *Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents*, 2020, pp. 1–8.

[9] L. Donatelli, K. Lai, R. Brutti, and J. Pustejovsky, "Towards situated AMR: Creating a corpus of gesture AMR," in *Digital Human Modeling and Applications in Health, Safety, Ergonomics and Risk Management. Health, Operations Management, and Design: 13th International Conference, DHM 2022, Held as Part of the 24th HCI International Conference, HCII 2022, Virtual Event, June 26–July 1, 2022, Proceedings, Part II.* Springer, 2022, pp. 293–312.

[10] A. Lücking and J. Ginzburg, "Towards the score of communication," in *Workshop on the Semantics and Pragmatics of Dialogue*, Waltham, United States, Jul. 2020, pp. 136–149. [Online]. Available: https://hal.archives-ouvertes.fr/hal-03134934

[11] S. Strogatz, "Sync: The emerging science of spontaneous order," 2004.

[12] A. H. Anderson, M. Bader, E. G. Bard, E. H. Boyle, G. M. Doherty, S. C. Garrod, S. D. Isard, J. C. Kowtko, J. M. McAllister, J. Miller, C. F. Sotillo, H. S. Thompson, and R. Weinert, "The hcrc map task corpus," *Language and Speech*, vol. 34, no. 4, pp. 351–366, 1992.

[13] A. Kendon, R. M. Harris, and M. R. Key, *Organization of behavior in face-to-face interaction*. Walter de Gruyter, 2011.

[14] R. Dunbar, *Grooming, Gossip, and the Evolution of Language*. Harvard University Press, 1998.

[15] F. Ramseyer and W. Tschacher, "Synchrony in dyadic psychotherapy sessions," *Simultaneity: Temporal structures and observer perspectives*, pp. 329–347, 2008.

[16] ——, "Nonverbal synchrony or random coincidence? how to tell the difference," *Development of Multimodal Interfaces: Active Listening and Synchrony: Second COST 2102 International Training School, Dublin, Ireland, March 23-27, 2009, Revised Selected Papers*, pp. 182–196, 2010.

[17] Z. Khosrobeigi, M. Koutsombogera, and C. Vogel, "Gesture and part-of-speech alignment in dialogues," in *Proceedings of the 26th Workshop on the Semantics and Pragmatics of Dialogue - Full Papers*. Dublin, Ireland: SEMDIAL, Aug. 2022, pp. 172–182. [Online]. Available: http://semdial.org/anthology/Z22-Khosrobeigi_semdial_0019.pdf

[18] M. Koutsombogera and C. Vogel, "Modeling collaborative multimodal behavior in group dialogues: The multisimo corpus," in *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, N. C. Calzolari, K. Choukri, C. istopher Cieri, T. Declerck, S. Goggi, K. Hasida, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, A. on Moreno, J. Odijk, S. Piperidis, and T. Tokunaga, Eds. Paris, France: European Language Resources Association (ELRA), May 2018, pp. 2945–2951. [Online]. Available: https://aclanthology.org/L18-1466

[19] O. P. John, E. M. Donahue, and R. L. Kentle, "Big five inventory," 1991. [Online]. Available: https://doi.org/10.1037/t07550-000

[20] O. P. John, L. P. Naumann, and C. J. Soto, "Paradigm shift to the integrative Big Five trait taxonomy: History, measurement, and conceptual issues," in *Handbook of personality: Theory and research, 3rd ed.* New York, NY, US: The Guilford Press, 2008, pp. 114–158.

[21] H. Brugman and A. Russel, "Annotating multi-media/multi-modal resources with ELAN," in *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC'04)*. Lisbon, Portugal: European Language Resources Association (ELRA), May 2004, pp. 2065–2068. [Online]. Available: http://www.lrec-conf.org/proceedings/lrec2004/pdf/480.pdf

[22] D. McNeill, *Hand and Mind: What Gestures Reveal About Thought*. Chicago: University of Chicago Press, 1992.

[23] F. T. Ramseyer, "Motion energy analysis (mea): A primer on the assessment of motion from video." *Journal of counseling psychology*, vol. 67, no. 4, p. 536–549, 2020.

[24] J. R. Kleinbub and F. T. Ramseyer, "R package to assess nonverbal synchronization in motion energy analysis time-series," *Psychotherapy research*, vol. 31, no. 6, pp. 817–830, 2021.

[25] F. Ramseyer and W. Tschacher, "Nonverbal synchrony in psychotherapy: coordinated body movement reflects relationship quality and outcome." *Journal of consulting and clinical psychology*, vol. 79, no. 3, pp. 284–295, 2011.

[26] M.-M. Bouamrane and S. Luz, "An analytical evaluation of search by content and interaction patterns on multimodal meeting records," *Multimedia Systems*, vol. 13, pp. 89–102, 2007, DOI10.1007/s00530-007-0087-8.