Linguistic and Gestural Coordination: Do Learners Converge in Collaborative Dialogue?

Arabella J. Sinclair
ILLC, University of Amsterdam
Science Park 107
1098 XG Amsterdam
a.j.sinclair@uva.nl

Bertrand Schneider
Graduate School of Education, Harvard
13 Appian Way
Cambridge, MA 02138
bertrand_schneider@gse.harvard.edu

ABSTRACT

Collaborative dialogue is rich in conscious and subconscious coordination behaviours between participants. This work explores collaborative learner dialogue through theories of alignment, analysing inter-partner movement and language use with respect to our hypotheses: that they interrelate, and that they form predictors of collaboration quality and learning. In keeping with theories of alignment, we find that linguistic alignment and gestural synchrony both correlate significantly with one another in dialogue. We also find strong individual correlations of these metrics with collaboration quality. We find that linguistic and gestural alignment also correlate with learning. Through regression analysis, we find that although interconnected, these measures in combination are significant predictors of collaborative problem solving success. We contribute additional evidence to support the theory that alignment takes place across multiple levels of communication, and provide a methodological approach for analysing inter-speaker dynamics in a multimodal task based setting. Our work has implications for the teaching community, our measures can help identify poorly performing groups, lending itself to informing the design of real time intervention strategies or formative assessment for collaborative learning.

Keywords

Dialogue, Gesture, Natural Language Processing, Alignment, Collaborative Learning

1. INTRODUCTION

Collaborative problem solving has long been a focus of educational research, and has been deemed an educational learning objective of critical importance in the 21st century workforce [12]. In the education literature, collaboration success is often analysed with respect to a *joint problem space* as created through learner interaction [31]. This joint problem space integrates learner shared goals, descriptions of the problem state, awareness of available problem solving actions and the associations between these aspects. The emergence of this shared conceptual space is constructed through shared language, situation and activity.

Alignment in dialogue, the language component of this interaction, is also commonly attributed to an automatic mechanism to achieve a shared understanding, or *situation model* [27]. This account of dialogue predicts alignment across various modes of communication, from word level to gesture and gaze patterns. Specifically in a task based setting, alignment is thought to aid mutual understanding [10, 27]. These theories of alignment and collaborative learning when taken together suggest that convergence at many levels of communication will take place in parallel over the course of collaborative dialogue, and that this alignment will be indicative of collaborative success. Additionally, the collaborative learning literature suggests that the effort necessary to build shared understanding is what actually leads to learning [40], thus alignment, already found to be predictive of student learning in a teacher student context [42], may also be indicative of this.

Investigating collaborative problem solving through the lens of alignment can give additional insights to this complex problem of convergence [38]. In this work, we examine the synchrony and convervence between students at both a linguisitic and gestural level, via inter-student metrics of linguistic alignment and movement synchrony. Of particular interest in this study is the separate coding of collaboration and learning in the learner dialogues. This allows for side by side comparison of the different modalities, and the analysis of their interaction with respect to these outcomes. Gestural and linguistic coordination between locutors has long been linked in dialogue, both properties having been individually explored for facilitating collaboration and learning in various settings.

We offer an exploration of theoretically motivated metrics to capture synchronisation and alignment at the levels of linguistic expressions and movement patterns. We explore correlations between the measures themselves and between collaboration and learning. Exploring the relationship between these measures we find strong correlations between modalities, in line with the collaboration and alignment literature. Finally we explore the combination of these modalities in their predictive power for both collaboration and learning, finding that although they are interrelated, each plays a significant role in prediction quality.

1.1 Research Questions

Motivated by the hypotheses we draw from the literature on collaboration, learning and alignment, we hypothesise that more successful learner dyads will converge in both their language use and their movement patterns to a visible degree, as the students align their mental models during the learning process. We also hypothesise that together, these aspects of interaction can provide useful tools in the analysis of student learning. We split our analysis into the following research questions:

RQ1: ***Evidence of Convergence:*** Are linguistic alignment, gestural synchrony and convervence effects between students higher than by chance task and vocabulary effects?

RQ2: ***Convergence Vs Collaboration & Learning:*** How do our measures of convergence correlate with collaboration and learning?

RQ3: ***Interaction: Linguistic Vs Gestural:*** Do these modalities correlate with one another?

RQ4: ***Multimodal: Do combined measures of synchrony predict learning and collaboration outcomes:*** are the measures in combination predictive of learning and collaboration?

2. BACKGROUND

Evidence of convergence. Experimentally, in a two person dialogue setting, speakers have been found to converge, each aligning to their locutor at many levels of communication: lexical, structural, gesture and conceptual[27]. Speakers have been found to spontaneously coordinate body postures and gaze patterns during conversation [35]. Behaviour matching in multimodal communication has also been found to be temporally synchronised in collaborative task-based activity, when participants are facing each other [22]. Imitation or mimicry between people unconscious of this behaviour has been found in incidental mannerisms such as the bouncing of a foot, or rubbing a nose [9]. People imitate one another in dialogue across many different modalities, including lexical choice [17], accent [18], pauses [7], speech rate [43] and syntax [30, 4, 26]. This imitation has been linked to having social benefits, for example, [9] find that speakers in those pairs where their incidental mannerisms were mimicked perceived the interaction as running more smoothly than those whose were not. In terms of collaboration, multi modal behaviour matching has been found to occur in a synchronous manner in a task based collaborative dialogue where rapport and its role in learning and convergence was investigated [22]. Acoustic prosodic entrainment has also been found to correlate with rapport, a social quality of the interaction, in collaborative learning dialogues [23]. Parallel to this, it has been argued that infants' early skills of joint attention is their emerging understanding that other people exist as intentional agents [8], as they develop theory of mind. In terms of learning, lexical entrainment has been shown to correlate with success in multiparty student engineering group project meetings [16], where higher scoring teams were more likely to increase their entrainment in project words over the course of a dialogue, while lower scoring teams are more likely to diverge. Alignment level has been shown to vary with student ability [36], and convergence of lexical and speech features from student to tutor in spoken tutorial dialogue corpora has been shown to be a useful predictor of learning [42].

Language and gesture in learning. Gesture has an important role in teaching and learning [32], as does language, which, at a structural level, has been shown to exhibit effects characteristic of both learning and implicitness, thought of as an aspect of alignment or coordination between interlocutors [15]. A wide range of lingusitic features derived from student dialogues have been found to be effective predictors of both learning gains and collaboration quality [29]. Categorising gesture in an educational setting often adopts the framework proposed by [24], of separating them into four basic types: beat (gestures devoid of topical content yet which lend temporal or emphatic structure i.e. hand tapping, head movement for emphasis), deictic (concrete or abstract pointing i.e. to match an object referred to as 'this' or 'that', or a concept in the past that is being referred to), iconic (also referred to as representational, i.e. making a gesture of putting a phone to ones ear), and metaphor (gestures to illustrate abstract concepts, such as moving hands together to illustrate mathematical convergence, or drawing a trend line in the air to demonstrate positive correlation). While gestures are pervasive, not all types are equally represented in particular speech events, and are very dependent on the dialogue context, however, various studies have found that gesture and speech together provide a better index of mental representation than speech alone, and to be an important aspect in learning [19, 11].

3. METHODS

Speaker	*Utterance*
Left:	then **we need** to **turn left** . again put an if **and then** turn the view book .
Right:	so another if do statement ?
Right:	was it just when you write in **the sensor** here , right into that . i say talk to motor ? all right , a and b .
Left:	if it is that , then we take a left . then **turn left** .
Right:	**turn left** , which is right here , right , so
Left:	put this inside that **and then** again , **we need** to turn left .
Right:	another if statement ? remember , control . **and then** talk to motor again . turn left
Left:	**turn left** comes after .
Right:	and **we need** one for ...

Table 1: Example dialogue excerpt. Expressions in bold indicate shared **lexical** constructions.

Experimental Setup. The experimental setup consisted of 40 pairs of undergraduate students participating in a collaborative problem solving task of programming a robot to traverse a maze. The participants had no prior programming experience. The participants sat facing a computer screen, and were recorded as they worked through the shared exercises. During the collaborative aspect of the task, the focus of our analysis in this work, participant dialogue was recorded and subsequently transcribed. Body movement data was also recorded via a Microsoft kinect sensor. This resulted in timestamped language and movement data for the 30 minute period of the task. The participants were individually given a pre and post test on a similar set of exercises, in order to evaluate the relative learning in the dyads. The learning assessment consisted of four short answer or fill-in-the-blank questions that assessed their understanding of basic computer science competencies (adapted from [5, 44]). Learning gains were computed by subtracting pre-test scores from post-test scores and divided by the total number of points to be gained minus the pre-test [13]. Collaboration was evaluated on a series of axes derived from the

collaboration assessment measures proposed in [25]: *sustaining mutual understanding, dialogue management, information pooling, reaching consensus, task division, time management, technical coordination, reciprocal interaction, and individual task orientation*[1]. Each dimension was given a score between +2 and -2 (each score was defined with concrete behaviors in a codebook). Researchers double-coded 20% of the sessions and had a Cronbach's alpha of .65 (75% agreement). An overall measure of collaboration was defined as the aggregate of all collaborative features. More details of the study, the data, experimental setup and coding of collaboration and learning scores can be found in Reilly et al. 2019 [29].

Dialogue transcripts. An example snippet from a dialogue with high collaboration score is provided in Table 1. The transcripts occasionally include some utterances from the facilitator, for whom we are not interested in measuring any alignment effects. The facilitator typically will speak most at the beginning of the dialogue, thus we remove all intial utterances (including participant) which interleave facilitator and participant. For the remainder of the dialogue, facilitator utterances are simply removed. Punctuation, although added at the annotators discretion, is retained, as it provides valuable information about the pace of the language used, and indicates the fragmented nature of these dialogues. The transcripts were tokenised before analysis using the nltk python package[2].

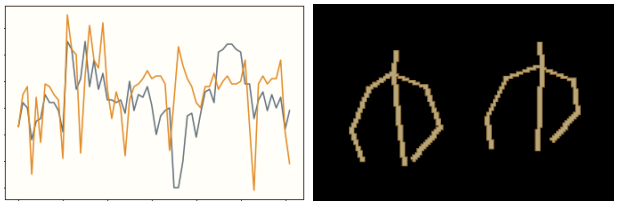

Figure 1: Example student movement pattern over time (left) and student geometries (right).

Movement Data. Student movement was processed using motion sensors from microsoft kinect, returning a series of geometries representing the students' body positions in space. An example of two students sitting at the table (which obsures the lower half of their bodies) can be seen in Figure 1. We only include data from the time period where the students were performing the collaborative activity, which corresponds to the transcript data of 30 minutes per session. In terms of pre-processing choices, our interest in the movement data is where the students mimic the gestures of their partner independent on their position relative to the camera or one another. This allows us to abstract from the postural shapes, relative size, and dominant hand of the participant's gestural patterns. We thus use averaged 30-second time slices[3] of the movement data (a measure of between frame positional difference). We further process this data to account for the differences in movement which the experimental setup introduces: we apply standardisation[4] to each participant signal in order that two signals of different means and standard deviations can be compared on the same axis. This grants us a measure of variance similarity, which captures better the elements of beat gesture patterns separate from absolute movement differences, as we know the students consistently display different mean movement levels across dyads.

3.1 Computing Lexical Alignment

We operationalise linguistic alignment in this work at the lexical (word) level, derived from the dialogue transcripts, extracting *shared expressions*, which we define as any sequence of tokens which contain at least one word (e.g. single punctuation marks are excluded). The automatic extraction of shared expressions per dialogue is an instance of the longest common sub-sequence problem [20, 3]. For each dialogue, we extract the inventories of shared expressions using the method proposed by Duplessis et al. [14]. For each of the two dialogue-specific inventories of shared constructions, we compute the following measures:

- *Expression Variety (EV):* The lexical diversity of the expression vocabulary.
- *Expression Repetition (ER):* The ratio of produced tokens belonging to an instance of an established expression
- *Vocabulary overlap (VO):* Captures the richness of the shared vocabulary, the ratio of shared vocabulary present in the dialogue between participants:

$$\frac{\#(words_{\text{speaker1}} \cap words_{\text{speaker2}})}{\#(words_{\text{speaker1}} \cup words_{\text{speaker2}})}$$

Individuals repeat and introduce expressions at different rates within dyads, thus we additionally calculate dyad level measures to capture the symmetry between interlocutors.

- *Expression Initiator (IE) Difference:* Difference in % of shared constructions introduced by each dialogue participant. *Initiator* describes the dialogue participant to first use a subsequently shared and repeated construction.

  $||IE(speaker\ one) - IE(speaker\ two)||$
- *Expression Repetition Difference:* The difference in proportion of an individual speaker's utterances which contain a usage of a shared construction:

  $||ER(speaker\ one) - ER(speaker\ two)||$

These measures capture the between speaker repetition within dialogue, which we use as a proxy for measuring the coordination or alignment between the speakers. An example of expression repetition can be found in Table 1

[1]Detailed descriptions of these measures used can be found in [25]

[2]NLTK[21] python package http://www.nltk.org

[3]Our choice of 30-second slices was in part informed by previous work [38], and through qualitatively examining the data at 5-second and 30-second slices, 30-seconds seems to be sufficient to capture interesting aspects of finer-grained hand movement, but not so fine as to render average total movement data meaningless

[4]Standardising a time series dataset involves re-scaling the distribution of values, also known as *Z-normalisation*

3.2 Computing Gestural Synchrony

We use Dynamic Time Warping [33] as the measure of similarity between partner movement patterns as it has been found to be a consistently robust measure of time series similarity [2], which although introduced as a method for analysing speech signal [33], has been employed successfully in the field of gesture recognition [1]. DTW is a technique to find the optimal alignment between two time series, through the stretching or compression of either series along its time axis. This warping can be used to find corresponding regions between two time series, and serve as a distance measure. This measure of distance allows us to capture slightly out of sync movement patterns, or those of slightly differnt phase length, to be deemed more similar than their difference in slope would suggest. Our main motivation for choosing DTW over other metrics demonstrated suitable for this task such as [38, 28] is our hope to capture the similarity of slightly asynchronus movement, which, if indeed movement and linguistic alignment is linked, should follow a similar mixed turn taking behaviour as dialogue [41]. As well as providing a robust distance measure between two time series, DTW returns a warping path which describes in what direction the time series needs to be moved in order to best align: in other words, the warping path can provide useful information about leader and follower dynamics which we exploit in our analysis.

Inspired by common measures of gestural synchrony/body language, we investigate the following dyadic behavioural properties:

- *Movement Difference (Mdiff):* Global mean movement difference in a pair. Calculated as the absolute value of the difference between the means of each participants.
- *Movement Synchrony (dtw_dist):* The synchrony between pairs in terms of their movement patterns, as measured by the Dynamic Time Warping (DTW)[33] distance
- *Leader Follower dynamics (diffLF):* The directionality of the alignment of the movement between the pair. This metric is derived from the DTW path.

Additionally, to measure whether these similarities become more pronounced as the session progresses, we divide the session in half by timestamp, and compute the measures per half. We use the difference between dialogue halves (*second - first*) as a measure to capture convergence. This results in three additional measures corresponding for those synchrony measures above: *Mdiff_change*, *dtw_dist_change* and *diffLF_change*. The aspects of the body geometries which we focus on consist of the points for *Head, Hands, Shoulders* and *Total* (average) movement. For hand and shoulder measures, these are defined as an average of the movement in the right and left points for each aspect.

3.3 Measure Validation - Baseline

A certain level of similarity between speakers will exist independently of their adapting to one another. Due to their performing the same task, vovabulary will necessarily be constrained by topic, and consistent across pairings. Due to the experiment configuration, task specific gesture patterns such as moving the robot, or interacting with the computer, as well as to which side of them their interlocutor is will also lead to movement similarities, e.g. turning to the right vs the left to speak. We thus create baselines for both dialogue and movement data which demonstrate the levels of similarity inherent to the task setup. For the dialogue baseline, we create a scrambled version of the corpus by retaining the utterances of one of the students and interleaving it with utterances randomly drawn from another pair, per speaker. For a partner specific movement baseline, the movement data from each student is randomly paired with the data from another student on the same side relative to them as their partner was (i.e for each participant on the right hand side, replace their partner with a participant from the left hand side). To further check task specific effects of the seating configuration, we pair students sitting in the same position with one another, in order to confirm that the role does not show more similarities than the origional student pairings.

4. RESULTS

4.1 Analysis 1: Measuring Convergence

Linguistic. We firstly hypothesise that there will be significant inter dyad repetition beyond what the task demands by chance, since alignment has been linked to both learning, as well as collaboration, and this same measure has found significant alignment levels in negotiation[14], as well as in second language tutoring dialogue[37], although this dialogue setting is different since both speakers are learners. Firstly we explore whether alignment is greater than by chance: we therefore compare the original dialogues to the shuffled baseline in the same manner as [14]. The expression variety is significantly higher for the original (mean=0.118, std=0.023) than for the shuffled dialogues (mean=0.110, std=0.015). Statistical difference is checked by a Wilcoxon rank sum test (U = 1141, $p = 0.03 < 0.05$, r = 0.21)[5] This indicates that there exists a richer and more dyad specific expression lexicon. The expression repetition is also significantly higher for the original (mean=0.509, std=0.123) than for the shuffled dialogues (mean=0.487, std=0.109) (U = 1079.5, $p = 0.014 < 0.05$, r = 0.25). This means that the level of repetition between student dyads is not simply incidental, and can be attributed to alignment or routinisation effects. Finally, as a measure of how task specific the vocabulary is, we find the vocabulary overlap between speakers significantly higher in the original (mean=0.509, std=0.123) than in the shuffled dialogues (mean=0.487, std=0.109) (U = 856, $p = 0.0002 < 0.001$, r = 0.41). This difference demonstrates that students share a much richer vocabulary than would happen by chance in performing this task. Overall, these results show that the collaborative student dialogues constitute a richer expression lexicon than they would by chance, indicating that the students align to one another, resulting in their langauge converging [10, 27].

Gestural. We hypothesise that our measures of movement matching will result in higher partner-specific synchrony than

[5]Following [14], for each test, we report the test statistics (U/W), the p - value (p) and the effect size (r)

in our baseline: i.e. lower distance between pairs collaborating than simply any student performing the same task with a different partner. We find that within dyad DTW similarity is significantly higher than in both the partner substitution baseline (t= -2.0401, $p < 0.05$), and the within side baseline (t = -2.0397, $p < 0.05$). This indicates DTW is a useful measure of movement similarity in this setting showing that this method is suitable for capturing partner specific effects of these movement patterns and can allow us to use this to compare similarities between dyads.

4.2 Analysis 2: Convergence Correlated with Collaboration and Learning

Lingusitic alignment. We hypothesise that alignment between students will correlate with learning, since in other tutoring settings, it has been found to correlate with both learning gains [42] and linguistic ability [36]. Additionally, global language features of collaborative dialogue have also found to correlate with learning gains [29]. To answer *RQ2*, we compare our linguistic alignment metrics with learning and collaboration scores. We find support for our hypothesis about collaboration and alignment correlating. We find ER (r 0.680 $p < 0.001$), EV (r 0.622 $p < 0.001$), and VO (r 0.663, $p < 0.001$) all correlate with collaboration using Pearson's r correlation coefficient. This shows that inter partner repetition is important, and that students will converge even to less common language in a collaborative setting. We also find our between learner measures significantly correlated with collaboration, IE_diff (r=-0.570, $p < 0.01$) and ER_diff (r=-0.54, $p < 0.001$) meaning that smaller differences between learner initiation and repetition of shared expressions correlates with how well they collaborate. EV (r=0.442, p =0.026), ER (r=0.442, p =0.006) and VO (r=0.349, p =0.034) also correlate with learning, although to a lesser degree. An intuition as for why, is that in other studies reporting alignment correlation with learning analyse dialogues conducted in an asymmetric *tutoring* setting, where adopting the language of the teacher is a sensible learning strategy as it is assumed that this language is correct. In our case, since these dialogues are between *peer* learners, the learning outcome is somewhat dependent on the rapport within the dyad, and the information aligned to being correct. In other words, in some cases, the learners may be converging to a shared mental representation, but it may not be the correct one. In keeping with this observation of dyad rapport and equality, IE_diff (r=-0.487, $p = 0.002$) and ER_diff (r=-0.515, $p = 0.001$) both show strongly that more equal contributions from the students in terms of repeating one another, and in introducing words upon which to align correlate with learning.

Movement Synchrony and Convergence. We hypothesise that movement synchrony and convergence, as defined by DTW distance and its change over the interaction, will provide a robust measure of synchrony which will better distinguish between dyads with differing activity levels, which in turn should correlate with collaboration and learning, in keeping with previous results with other measures in task based dialogue [22, 32, 28]. We compare our movement similarity metric with learning and collaboration scores. Overall

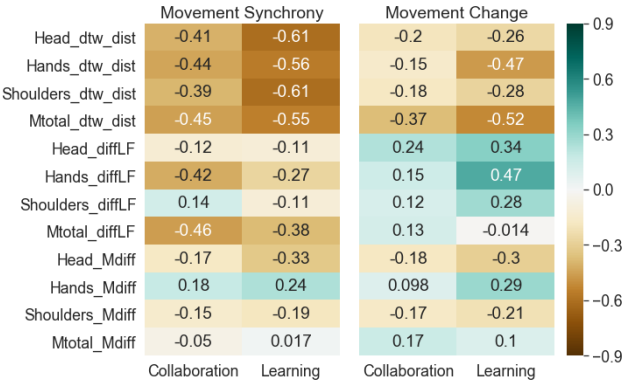

Figure 2: Gestural synchrony and convergence vs. Collaboration and Learning measures correlation with Pearson's r. Significant p values reported in the text.

as can be seen from Figure 2, we find average movement synchrony to correlate with both collaboration and learning. In terms of learning, DTW_dist measures for Head (r = -0.611, $p > 0.001$) Hands (r = -0.561, p = 0.002), Shoulders (r = -0.609, $p > 0.001$), and Mtotal (r = -0.611, p =0.006) all significantly correlate with learning to a strong degree. Convergence between dyads in terms of Mtotal (r=-0.519, p = 0.006) and Hands(r=-0.467, p = 0.014) also significantly correlate with learning. Finally, dyads becoming more dissimilar in terms of hand movement (having a stronger leader follower dynamic) also significantly correlates with learning: Hands_diffLF (r=0.471, p = 0.013). With collaboration, Head Hands Shoulders and Mtotal all significantly ($p < 0.05$) correlate. As does the diffLF for Hands (r=-0.42, p = 0.029) and Mtotal (r=-0.519, p = 0.006). Overall, the results are intuitive: we find that more synchronus pairs as measured by DTW distance significantly correlate with collaboration quality. We also find that convergence between dyads is present (negative correlation between *dtw_dist change* and *Mdiff change* show greater similarity between learners over time) and correlates with learning quite strongly for some movement metrics. We also see a positive correlation between the *diffLF change* features, particularly with learning, indicating that while convergence of behaviour is important, some aspects of turn taking and initiative are separate to this.

4.3 Analysis 3: Comparing Linguistic and Gestural Convergence

Comparing Linguistic and Gestural convergence, we hypothesise these aspects of communication will correlate with one another, as previous literature suggests [17, 4, 22]. To answer research question (*RQ3*), we contrast the modalities themselves. We split this comparison to compare gestural and linguistic coordination. We hypothesise that movement synchrony and linguistic alignment will correlate strongly, due to the process of speakers' alignment of shared mental representations taking place across various linguistic and paralinguistic levels [6, 27]: if dyads align at the lexical level, it is likely that the same process leading to this alignment will affect the gestural level also [27]. The DTW path allows us to capture the relationship between slightly offset movement patterns of beat gesture mimicry [24], and the

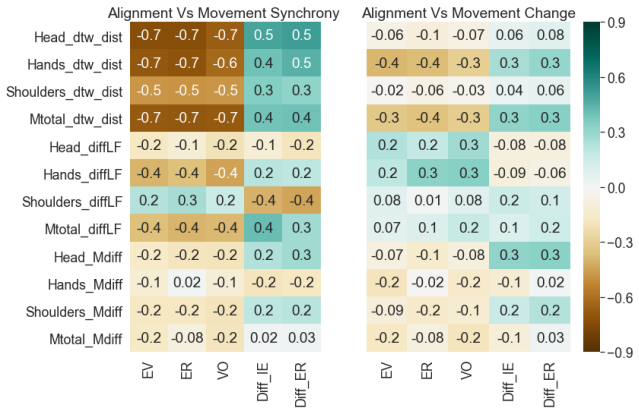

Figure 3: Movement measures vs. linguistic alignment measures. Pearson's r correlation coefficient.

case where linguistic alignment in turn taking utterances is low, which can lead to more synchronised patterns of movement [32]. Previous work has also found evidence of the coordination of of lexical alignment and gestural behaviour in a multimodal [35, 9, 22] context, we thus hypothesise that collaborative problem solving dialogues will also demonstrate this. We find significant correlation between linguistic alignment measures and movement synchrony across all movement patterns (Figure 3), with strongest effects for head (r = -0.735, p-value: 1.225) and hands (r = -0.706 p-value: 3.811), providing support that our hypothesis about lexical patterns influencing gestural alignment at the level of beat gesture and head nodding may be be true for this setting. In terms of convergence, there is a significant correlation with change in hand movement and expression variety (r = -0.387, p-value: 0.0458). Interestingly, the difference between speakers lingusitic (*Diff_IE*, *Diff_ER*) patterns positively correlates with both their difference in movement synchrony and speaeker divercence, indicating that asymmetric relationships between students are visible across modalities of communication. Inkeeping with our hypothesis we find strong negative correlation between between speaker difference in movement and divergence (strong correlation between similarity and convergence) with the linguistic measures of convergence, providing supporting evidence for the hypothesis that lingusitic and gestural convergence are part of the same underlying communicative process.

4.4 Analysis 4: Predicting Learning and Collaboration

Finally, in order to answer research question (*RQ4*), to find combined interaction effects of the various inter modality measures, we fit a series of mixed effect regression models[6]. We hypothesise that while each measure individually is strong, and although the measures themselves are correlated, each modality will provide its own distinct information contributing to learning and collaboration aspects. We perform backward step wise model selection to select the best predictors, firstly fitting each model with all relevant variables and stopping only when all remaining terms have significance $p < 0.05$. Although RMSE and r^2 values of predicted data are highest when combining all factors, we wished to discover the minimally significant descriptive set of criteria in order to find more interaction in our results.

[6]To fit the data and perform the statistical tests within this paper, we use the Statsmodels python package [34]

Table 2: Mixed effects Regression model multimodal results

Formula
Learning: **RMSE:5.48** r^2**:0.90**
*Learning ~ EV : ER + Diff_IE * Diff_IE + Handmean_movement_30_diffLFChange + Shouldermean_movement_30_diffLFChange*
Collaboration **RMSE:0.15** r^2**:0.999**
*Collaboration ~ EV * ER + Diff_IE * Diff_ER + Head_movement_30_dtw_dist + Handmean_movement_30_dtw_dist + Shouldermean_movement_30_dtw_dist + movement_total_30_dtw_dist + Head_movement_30_diffLFChange + Shouldermean_movement_30_diffLFChange + movement_total_30_diffLFChange + Head_movement_30_dtw_dist_change + Handmean_movement_30_dtw_dist_change + Shouldermean_movement_30_dtw_dist_change + movement_total_30_dtw_dist_change + Head_movement_30_diffLF + Handmean_movement_30_diffLF + Shouldermean_movement_30_diffLF + movement_total_30_diffLF*

Table 2 shows the minimal significant set of linguistic and gestural factors and their interaction in terms of their ability to predict the dependent variables of *learning* and *collaboration*. Each modality separately can form a good predictor of both alignment and learning in this setting. However, this analysis offers strong support for the multimodal modelling of collaborative problem solving, proving that although correlating with one another, both linguistic and gestural aspects have an independent role to play when predicting learning and collaboration. Broadly, from Table 2, the gestural features chosen indicate that both the measures of synchrony, and those for convergence (*_change* features) play a role in prediction. It is also clear that predicting collaboration in this case is easier than learning. This may be influenced by ceiling effects or ease of pre-test being a limiting factor. e.g. a learner with very good pretest score will have hit a ceiling by the end of the session.

5. DISCUSSION & CONCLUSION

We find significant levels of both linguistic and movement synchrony in our data (*RQ1*). In answer to *RQ2*, we find our measures of linguistic and gestural alignment correlate with collaboration. In terms of learning, we find that the difference in repetition between students negatively correlates with learning, that movement synchrony in general shows strong correlation with learning. In terms of *RQ3*, we find significant strong to medium effects when correlating measures of ER and dtw_dist with one another. This contributes to a growing body of evidence in support of theories of interactive alignment emerging across communicative modalities. Finally, via regression analysis combining our metrics (*RQ4*), we find that although separately powerful, a combination of modalities can best explain collaboration and learning outcomes. Our findings show the importance of analysing between speaker dynamics to capture nuances of learning. Our findings also suggest the use of a multimodal approach for the best understanding of these interactions. We also contribute interesting new evidence adding to work exploring the relationship between linguistic alignment and gestural and movement similarity. Our findings, while limited to a small specific setting, contribute evidence to support existing theories of human cognition and alignment.

6. ACKNOWLEDGMENTS

First author funded under grant agreement No. 819455 from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme.

7. REFERENCES

[1] A. Akl and S. Valaee. Accelerometer-based gesture recognition via dynamic-time warping, affinity propagation, & compressive sensing. In *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 2270–2273. IEEE, 2010.

[2] A. J. Bagnall, A. Bostrom, J. Large, and J. Lines. The great time series classification bake off: An experimental evaluation of recently proposed algorithms. extended version. *CoRR*, abs/1602.01711, 2016.

[3] L. Bergroth, H. Hakonen, and T. Raita. A survey of longest common subsequence algorithms. In *Proceedings Seventh International Symposium on String Processing and Information Retrieval. SPIRE 2000*, pages 39–48. IEEE, 2000.

[4] H. P. Branigan, M. J. Pickering, and A. A. Cleland. Syntactic co-ordination in dialogue. *Cognition*, 75(2):B13–B25, 2000.

[5] K. Brennan and M. Resnick. New frameworks for studying and assessing the development of computational thinking. In *Proceedings of the 2012 annual meeting of the American educational research association, Vancouver, Canada*, volume 1, page 25, 2012.

[6] S. Brennan and H. Clark. Conceptual Pacts and Lexical Choice in Conversation. *Journal of Experimental Psychology*, 22(6):1482–1493, 1996.

[7] J. N. Cappella and S. Planalp. Talk and silence sequences in informal conversations iii: Interspeaker influence. *Human Communication Research*, 7(2):117–132, 1981.

[8] M. Carpenter, K. Nagell, M. Tomasello, G. Butterworth, and C. Moore. Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the society for research in child development*, pages i–174, 1998.

[9] T. L. Chartrand and J. A. Bargh. The chameleon effect: the perception–behavior link and social interaction. *Journal of personality and social psychology*, 76(6):893, 1999.

[10] H. Clark and D. Wilkes-Gibbs. Referring as a collaborative process. *Cognition*, 22:1–39, 1986.

[11] S. W. Cook, Z. Mitchell, and S. Goldin-Meadow. Gesturing makes learning last. *Cognition*, 106(2):1047–1058, 2008.

[12] N. R. Council et al. *Education for life and work: Developing transferable knowledge and skills in the 21st century*. National Academies Press, 2012.

[13] S. Cuendet, E. Bumbacher, and P. Dillenbourg. Tangible vs. virtual representations: when tangibles benefit the training of spatial skills. In *Proceedings of the 7th Nordic conference on human-computer interaction: Making sense through design*, pages 99–108, 2012.

[14] G. D. Duplessis, C. Clavel, and F. Landragin. Automatic measures to characterise verbal alignment in human-agent interaction. In *Proceedings of the 18th Annual SIGdial Meeting on Discourse and Dialogue*, pages 71–81, 2017.

[15] V. S. Ferreira and K. Bock. The functions of structural priming. *Language and Cognitive Processes*, 21(7-8):1011–1029, 2006. PMID: 17710210.

[16] H. Friedberg, D. Litman, and S. B. Paletz. Lexical entrainment and success in student engineering groups. In *2012 IEEE Spoken Language Technology Workshop (SLT)*, pages 404–409. IEEE, 2012.

[17] S. Garrod and A. Anderson. Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition*, 27(2):181–218, 1987.

[18] H. Giles, P. Powesland, E. A. of Experimental Social Psychology Staff, and E. A. of Experimental Social Psychology. *Speech Style and Social Evaluation*. European monographs in social psychology. European Association of Experimental Social Psychology by Academic Press, 1975.

[19] S. Goldin-Meadow. Beyond words: The importance of gesture to researchers and learners. *Child development*, 71(1):231–239, 2000.

[20] D. S. Hirschberg. Algorithms for the longest common subsequence problem. *Journal of the ACM (JACM)*, 24(4):664–675, 1977.

[21] E. Loper and S. Bird. Nltk: The natural language toolkit. In *In Proceedings of the ACL Workshop on Effective Tools and Methodologies for Teaching Natural Language Processing and Computational Linguistics. Philadelphia: Association for Computational Linguistics*, 2002.

[22] M. M. Louwerse, R. Dale, E. G. Bard, and P. Jeuniaux. Behavior matching in multimodal communication is synchronized. *Cognitive science*, 36(8):1404–1426, 2012.

[23] N. Lubold and H. Pon-Barry. Acoustic-prosodic entrainment and rapport in collaborative learning dialogues. In *Proceedings of the 2014 ACM workshop on Multimodal Learning Analytics Workshop and Grand Challenge*, pages 5–12, 2014.

[24] D. McNeill. *Hand and mind: What gestures reveal about thought*. University of Chicago press, 1992.

[25] A. Meier, H. Spada, and N. Rummel. A rating scheme for assessing the quality of computer-supported collaboration processes. *International Journal of Computer-Supported Collaborative Learning*, 2(1):63–86, 2007.

[26] M. J. Pickering and V. S. Ferreira. Structural priming: A critical review. *Psychological bulletin*, 134(3):427, 2008.

[27] M. J. Pickering and S. Garrod. Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27(02):169–190, 2004.

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

[29] J. M. Reilly and B. Schneider. Predicting the quality of collaborative problem solving through linguistic analysis of discourse. *International Educational Data Mining Society*, 2019.

[30] D. Reitter, F. Keller, and J. D. Moore. Computational modelling of structural priming in dialogue. In *Proceedings of the Human Language Technology Conference of the NAACL, Companion Volume: Short Papers*, NAACL-Short '06, pages 121–124, Stroudsburg, PA, USA, 2006. Association for

Computational Linguistics.

[31] J. Roschelle and S. D. Teasley. The construction of shared knowledge in collaborative problem solving. In *Computer supported collaborative learning*, pages 69–97. Springer, 1995.

[32] W.-M. Roth. Gestures: Their role in teaching and learning. *Review of educational research*, 71(3):365–392, 2001.

[33] H. Sakoe and S. Chiba. *Dynamic Programming Algorithm Optimization for Spoken Word Recognition*, page 159–165. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1990.

[34] S. Seabold and J. Perktold. Statsmodels: Econometric and statistical modeling with python. In *Proceedings of the 9th Python in Science Conference*, volume 57, page 61. Austin, TX, 2010.

[35] K. Shockley, D. C. Richardson, and R. Dale. Conversation and coordinative structures. *Topics in Cognitive Science*, 1(2):305–319, 2009.

[36] A. Sinclair, A. Lopez, C. Lucas, and D. Gasevic. Does ability affect alignment in second language tutorial dialogue? In *19th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL 2018)*, 4 2018.

[37] A. Sinclair, K. McCurdy, C. G. Lucas, A. Lopez, and D. Gaševic. Tutorbot corpus: Evidence of human-agent verbal alignment in second language learner dialogues. *International Educational Data Mining Society*, 2019.

[38] T. Sinha and J. Cassell. We click, we align, we learn: Impact of influence and convergence processes on student learning and rapport building. In *Proceedings of the 1st Workshop on Modeling INTERPERsonal SynchrONy And infLuence*, pages 13–20, 2015.

[39] D. Tannen. New york jewish conversational style. *International Journal of the sociology of language*, 1981(30):133–150, 1981.

[40] S. Teasley, F. Fischer, P. Dillenbourg, M. Kapur, M. Chi, A. Weinberger, and K. Stegmann. Cognitive convergence in collaborative learning. 2008.

[41] M. Walker and S. Whittaker. Mixed initiative in dialogue: An investigation into discourse segmentation. *arXiv preprint cmp-lg/9504007*, 1995.

[42] A. Ward and D. Litman. Dialog convergence and learning. *Frontiers in Artificial Intelligence and Applications*, 158:262, 2007.

[43] J. T. Webb. Subject speech rates as a function of interviewer behaviour. *Language and speech*, 12(1):54–67, 1969.

[44] D. Weintrop and U. Wilensky. To block or not to block, that is the question: students' perceptions of blocks-based programming. In *Proceedings of the 14th international conference on interaction design and children*, pages 199–208, 2015.