Interpersonal Gesture Alignment in Individuals Speaking in their First or Second Language

Vladimír Vursta

S5213509

Department of Psychology, University of Groningen

PSB3E-BT15: Bachelor Thesis

Supervisor: Lisette de Jonge-Hoekstra

Second evaluator: Henk A. L. Kiers

In collaboration with: Nina Jetten, Natalia Kijak, Samuel Martire, Lynet Reinds and Melina Bardi.

April 21, 2025

A thesis is an aptitude test for students. The approval of the thesis is proof that the student has sufficient research and reporting skills to graduate, but does not guarantee the quality of the research and the results of the research as such, and the thesis is therefore not necessarily suitable to be used as an academic source to refer to. If you would like to know more about the research discussed in this thesis and any publications based on it, to which you could refer, please contact the supervisor mentioned

Abstract

Interacting people align their behaviours and body movements. Gestures also exhibit this kind of alignment. The current thesis examined the differences in gesture alignment in dyadic conversation, when participants spoke in their first language (L1) or second language (L2). Thirty-three student dyads were asked to have a ten-minute conversation, discussing debate phrases developed to encourage conversation. Dyads were categorized into three language context groups: Both speaking in their first language (L1-L1), one speaking their first and the other speaking their second language (L1-L2) or both speaking their second language (L2-L2). The conversations were video recorded and annotated for gestures and speech analysis. The bachelor students were trained in gesture annotation using the M3D framework. Gesture alignment in each dyad was measured as alignment in gesture frequency (frequency difference) and alignment in gesture duration (duration difference) of the participants. The groups were compared for differences in gesture frequency alignment and gesture duration alignment using the Kruskal-Wallis test. We found no significant difference between the groups, which suggests that the groups expressed a similar degree of gesture alignment. Additional analysis revealed that the groups also did not significantly affect within dyad proficiency differences, suggesting insufficient diversity in language proficiency. There was also no relation between gesture alignment and proficiency differences. Future research should further explore the relationship between language proficiency and gesture alignment, using samples with greater diversity in language proficiency.

Keywords: gesture alignment, interpersonal coordination, L1, L2, dyadic conversation

Interpersonal Gesture Alignment in Individuals Speaking in their First or Second Language

Interacting with others around us is one of the integral parts of being a human. From a simple head nod, acknowledging someone's existence, to elaborate conversations about the meaning of life, we constantly and unintentionally adjust our body language, speech and our gestures to align with that of others. It has been observed that people who are engaged in a conversation often coordinate their behaviours with one another in multiple domains, including gesture (Opazo et al., 2024). This multifaceted coordination has received a considerable amount of labels in the past and has been studied extensively (Chartrand & Lakin, 2013). As the world is becoming more and more global, it is more common that people must communicate in a language they acquired later in life. The current thesis will focus on gesture alignment between participants speaking in their first or second language. To fully grasp this topic, it is necessary to first discuss the relevant terminology.

Defining Terms

Interacting individuals align their nonverbal behaviours on a variety of levels (Chartrand & Lakin, 2013). Multiple studies have explored how individuals mimic each other's body postures (LaFrance, 1985), facial expressions (Lundqvist & Dimberg, 1995), yawning (Platek et al., 2003), foot shaking (Chartrand & Bargh, 1999) and also gestures (Louwerse et al., 2012). Given the interdisciplinary nature of the field of social interaction, studies regarding interpersonal alignment in communication have used a wide range of labels, such as: mimicry, behaviour matching, imitation, or accommodation (Rasenberg et al., 2020). Behavioural mimicry entails imitation or co-occurrence of the same or similar behaviour between two or more individuals at the same time or within a short time window (Chartrand & Lakin, 2013). Similarly, Bernieri and Rosenthal (1991) divided interpersonal coordination into two distinct concepts:

Behavioural Matching, sometimes also referred to as *Mimicry*, refers to coordination that does not require strict synchronization of the movements in time. It can be manifested as the imitations of body movements with a delay of a few seconds.

Interpersonal synchrony on the other hand requires synchronization in time. However, despite persistent efforts to disentangle the terms from each other, the distinction is still not precisely defined (Ayache et al., 2021).

In the current thesis I use the term alignment in a more general sense as indicating coordination between interlocutors (Rasenberg et al., 2020). The evidence regarding those concepts uses a range of terms and methodologies. In order to encompass the large array of phenomena, the present study uses the term alignment in a broader sense. It is used as an umbrella term for observed similarities in behaviours and movements between interacting individuals. I then use for example *gesture alignment* to further specify the modality of the alignment. When speaking about specific study we also specify the term the authors used. After having established the meaning of some of the relevant terminology, next we will consider some of the possible implications alignment could have on social interaction.

Implications of Interpersonal Alignment for Social Interaction

Schmidt and Richardson (2008) describes interpersonal coordination as a dynamic process governed by the logic and principles of self-organization observed in nature. Self-organization can be thought of as the tendency of interacting components to spontaneously move towards order and equilibrium. Moreover, this dynamic process is viewed as the basis for unintentional interpersonal alignment. Importantly, the authors find that the dynamics of alignment between social interactants is constrained by social influences such as connectedness and rapport. Building on this, alignment has been linked to various interpersonal outcomes.

Alignment in terms of behavioural matching has been shown to increase prosocial behaviours, and work as a so called “social glue”, increasing rapport between interactors and helping establish a social bond (Chartrand & Lakin, 2013). A meta-analysis by Rennung and Göritz (2016) examined 60 studies to assess the prosocial effects of interpersonal synchrony and found moderate effects of alignment on both attitudinal and behavioural aspects of prosociality. Additionally, Pickering and Garrod (2006) highlight the importance of interpersonal alignment in achieving successful communication. These implications bring us to the questions of why individuals align in such way? In the next paragraphs, I will discuss the theoretical approaches behind behavioural alignment.

Theoretical Accounts of Alignment

When it comes to theoretical contributions in social interaction and communication, it is understandable to find a wide variety of perspectives. However, theoretical approaches in the field of alignment have generally been formulated around the dichotomy between *priming* and *grounding* accounts (Rasenberg et al., 2020). The *priming* perspective holds that an automatic mechanism of priming is involved in the emergence of alignment (Pickering & Garrod, 2006). This process doesn’t require negotiation between interacting individuals. When one individual expresses a certain behaviour, it primes the other individual to behave in the same way, making it more likely for their behaviours to align (Pickering & Garrod, 2004; Pickering & Garrod, 2006).

On the other hand, the grounding perspective proposes that alignment emerges from interaction involving coordination of efforts, where interaction participants jointly create shared meaning (Holler & Wilkin, 2011; Rasenberg et al., 2020). Grounding in communication can be understood as process by which interacting individuals continuously update their *common ground* (i.e., shared beliefs, knowledge, and assumptions). Through the lens of this framework, communication is viewed as collaborative process in which

interactants have to coordinate a large amount of shared information and therefore have to update their *common ground* very frequently (Clark & Brennan, 1991). Studies that have investigated alignment in the past have mostly been guided by one of those theoretical frameworks (Rasenberg et al., 2020).

Self-organization is another theoretical approach worth mentioning, although it does not seem to fit directly in one of those two categories. As mentioned in the previous section, alignment is proposed to arise from a natural dynamic interplay between interacting individuals. In this framework self-organization would be the guiding principle behind the emergence of spontaneous alignment between individuals (Schmidt & Richardson, 2008). Before considering research regarding gesture alignment, we must also discuss gesture production in general.

Gestures

Gestures are visible bodily actions used in discourse as part of uttering something to another person. Individuals may for example use their hands to represent the shape of an object, illustrate size or use pointing in order to refer to something (Kendon, 2004). Gestures enrich communication by providing an additional representational form alongside speech for both speakers and listeners. In speakers, gestures serve as a tool to enhance thinking by reducing cognitive effort, while for listeners gestures may express unspoken content (Goldin-Meadow, 1999). MacNeill (1992) views gestures as a part of the language process and distinguishes between four types of gestures:

1. Iconic gestures represent concrete aspects of an object or a scene. They can appear as a part of description. In other words, they involve using hands to depict a certain scene or an object.

2. Metaphoric gestures are used to represent more abstract concepts such as relationships. For example, putting the hands together to represent a union of some sort.
3. Deictic gestures express locations of abstract or concrete concepts using pointing.
4. Beat gestures are hand movements that represent a rhythmic aspect of speech. They carry pragmatic content by marking significant words.

Gestures in L2

To fully understand gesture alignment in communication, we first need to look at how gesture production differs in L2 speakers. Previous studies have looked at frequency of gesture use and the influence of language proficiency. Lin (2019) examined how proficiency influences the types of co-speech gestures used in communication and found that less proficient L2 speakers used more deictic gestures, whereas more proficient L2 speakers were more likely to use more iconic and beat gestures. The author proposes that more proficient speakers might use gestures, utilizing cues such as rhythm as a strategy to enhance the meaning of the communicative message, whereas less proficient speakers use their physical context in construction of meaning in order to compensate for the lack of proficiency.

Emir Özder et al. (2023) on the other hand investigated how L2 speakers use gestures in emotional context. Participants were asked to retell an emotional narrative. They found that individual produced more gestures in L2 than in L1. The author suggests that this could be due to reduced proficiency in L2 compared to L1, resulting in more frequent gesture use as a compensation strategy. Another study by Ma (2022) found that co-speech gestures might play a role in meaning construction in L2 speech. They found that co-speech gestures were positively correlated with meaning related measures of speech.

Previous Research Regarding Gesture Alignment

Only relatively recently has research addressed alignment in co-speech gestures. One of the first studies showed that that mimicry does indeed occur in co-speech gestures (Kimbara, 2006, 2008). Extending this area of inquiry, Holler and Wilkin (2011) found that individuals mimicked each other's gestures more frequently in conversations when they could see each other compared to when they could not see each other. Moreover, the authors chose referential communication task, which involved conversation where participants were asked to talk about a set of stimuli, while assigned a specific role in the conversation. The discussion therefore involved spontaneous references to the spatial dimensions of the stimuli. This method allowed the authors to suggest that their results illuminate the importance of mimicked gestures in the creation of mutual understanding and shared knowledge and thus provide support for the theoretical perspective of *grounding* in communication (Clark & Brennan, 1991; Holler & Wilkin, 2011).

Other studies have examined hand movement alignment using a different methodological approach utilizing spectrum analysis by way of wavelet transform function. This approach allowed them to assess synchronization in different time scales. Time series movement data was analysed for cross-wavelet coherency and relative phase and compared to pseudo-dyad conditions (see both Fauviaux et al., 2024; Fujiwara & Daibo, 2016). Cross wavelet analysis determines how two signals oscillate at different frequencies during the conversation. Coherency then measures the degree of coupling of the oscillations in time. On the other hand, relative phase angles of the two signals indicate which signal is leading the other (de Jonge-Hoekstra, 2021). Fujiwara and Daibo (2016) tested the validity of using this method to determine synchrony in dyads engaged in unstructured conversation. The authors found that genuine dyads showed higher coherence than pseudo dyads. Fauviaux et al. (2024) on the other hand used spectral analysis to assess both intra- and interpersonal

synchronization at multimodal scale. They examined both gestures and speech and most importantly, found evidence of interpersonal gesture synchrony.

Lastly, recent study by Opazo et al. (2024) investigated gesture alignment in face-to-face dialogs between students and teachers during consultation hours. Importantly, recorded interactions were L1-L2 dialogues between native English-speaking teachers and Spanish undergraduate students. They looked at gesture patterns between participants as well as the temporal aspect of the gesture alignment and lastly the direction of the matching gestures. They found most matching gestures appearing in the 0.250s – 10s window and that students and teachers mimicked each other in both directions. In their discussion, the authors also propose that alignment in different language contexts might be a possible area for future research. Which leads us to the current study.

Current Study

The current study investigates gesture alignment in dyadic conversations. The focus of interest is the influence of language context on the degree of gesture alignment. Since the evidence regarding alignment in different language setting is sparse, we found the research base insufficient to pose a specific prediction. Furthermore, the literature regarding gesture alignment involves variety of methodologies, with most studies focusing on the occurrence of pairs of similar or identical gestures. The current thesis is taking a different approach and aims to assess gesture alignment as the degree of similarity between participants in a dyad in how often they gesture (gesture frequency) and also for how long they gesture (gesture duration). For all those reasons we resorted to an open exploratory research question:

- How does gesture alignment manifest in L1-L1, L1-L2, and L2-L2 conversations, in terms of gesture frequency and gesture duration?

Method

Participants

A total of 78 participants, constituting 39 dyads, participated in the study. Due to time constraints, which will be explained later, data was coded for 66 participants (M_{age} = 20.14, SD_{age} = 1.84). All participants were bachelor students recruited through convenience sampling. As compensation for participating in the study, the students received credits required to complete a first-year psychology course. The students were required to attain a certain number of credits but had the autonomy to select in which experiment they wished to participate.

Participants conversed in Dutch and English. Other languages participants reported speaking include Russian, Slovak, Finnish, Romanian, Chinese, Japanese, Spanish, Polish, Hindi, Bengali, Arabic, French, Italian, Swedish, Turkish, Czech, American Sign Language, Frisian, Croatian, Thai, and Greek. The University of Groningen Ethics Committee approved the study (code: PSY-2425-S-0110), and all participants provided written informed consent.

Materials

To conduct the data collection in a controlled environment, participants did the experiment in a room dedicated to behavioural research at the faculty of Social and Behavioural Sciences. Two chairs were placed close to each other at a 90-degree angle to facilitate communication between participants. Participants were being recorded by a JVC camcorder and a Sony camcorder, both attached to adjustable tripods. These tripods were carefully placed so that every gesture would be completely visible during the discussion. A room divider was positioned to ensure that participants could not see the researchers, both during the test administration and the experimental task. This was done to minimize potential distractions and to ensure as much as possible natural interaction between the participants. Any environmental noise was later removed from the recordings using the audio editor

Audacity (Audacity Team, 2014). The speech amplitude of the recordings was used in a custom R-script created by the supervisor to identify the start and end of speech segments.

Additionally, two computers were available in the room to complete the consent form, the Demographics Questionnaire, and the assigned C-test. To assess participants' language proficiency levels, four different variations of the C-test were used: an English (Gilmore, 2013), Dutch (Freeborn & Andringa, 2023), French (Coleman et al., 2017a), and German (Coleman et al., 2017b) version[1]. The C-test consists of four paragraphs varying in difficulty. Each paragraph had several blanks where letters needed to be filled in by the participant, such as: “You wa____ in t____ morning, ru____ to t____ window a____ take a de____ breath. Do____!” (Gilmore, 2013). The participants’ language proficiency score was determined by the percentage of correctly filled-in blanks.

Lastly, several predefined discussion phrases were used to prompt the debate. The discussion phrases were questions, such as ‘Which came first: the chicken or the egg?’ and ‘Would you rather be very funny or very intelligent?’. The debate phrases were previously tested by a trial study of a master’s student (see Appendix A).

The signed informed consent form, completed demographic questionnaires, and C-test were uploaded to Qualtrics (Qualtrics XM - Experience Management Software, 2025). Coders completed a training program on gesture annotation (Tütüncübasi et al., 2023) before coding the data in ELAN (ELAN, 2024). The data was later integrated into one data file using a R-script and analysed in JASP (JASP, 2025). A more detailed description of how the programs were used will follow in the data processing and analysis sections.

Procedure

The experiment was conducted in pairs. Upon arrival, participants were informed about the course of the experiment and the general aim of the study. Participants provided

[1] In the end only the English and Dutch versions of the Cloze test were completed by participants.

consent for the recording of the experiment via an informed consent form in Qualtrics. By signing the form, participants gave consent to being recorded and to the processing of their written, video, and audio data. To ensure natural gesture production during the conversation, gestures and speech were not mentioned. Next, an online demographic questionnaire was completed. An experiment leader decided the language the dyad would converse in based on each participant's L1 and L2 and filled it in the questionnaire. Additionally, each speaker's seating position, left or right, as viewed in the footage, was also filled in by the experiment leader.

Next, participants completed a C-test in the language in which the experiment would be conducted to measure their language proficiency. They were given a maximum of eight minutes for this test. If the test was not completed within the time limit, the participants were asked to finish the last words they were working on and then submit the unfinished test.

After completing the C-test, both participants took their assigned seats, and the actual experiment began. The experiment leader read the statements aloud, which the participants had to discuss with each other. They were instructed to consider the pros and cons of each statement until they reached a consensus. Once they indicated to the experiment leader that they had reached a consensus, the next statement was read aloud.

The discussion lasted a maximum of ten minutes, regardless of how many statements were discussed. If ten minutes had passed and the participants were still conversing, they were allowed to finish their final discussion. Finally, the participants were thanked for their participation. They received a brief explanation of the research, had the opportunity to ask questions, and were debriefed. The entire study took approximately 30 minutes per dyad.

Data processing and coding

Researchers coded the gestures in the videos to analyse the data using ELAN (ELAN, 2024). Each annotator independently coded up to six videos. Dutch experiment leaders coded

the Dutch videos, while the international experiment leaders coded the English videos. Before starting the coding process, coders completed an extensive M3D training program (Tütüncübasi et al., 2023), which provided guidance using ELAN and distinguishing between different types of gestures. In addition to watching the written instructions and training videos, the annotators worked through exercises to enhance familiarity with the program.

A meeting was held to discuss inter-reliability. An adjusted template from the M3D training program was used to ensure consistent coding across all tiers. The template contained the following: G-units, G-phases (only stroke), self-adaptors, non-referential gestures, metaphoric gestures, iconic gestures, abstract deixis, concrete deixis, and emblems (Rohrer et al., 2023). Coding G-Units involved detecting segments in which participants were gesturing, the onset of a G-Unit was noted when the hands left a resting position and the end of the G-unit when the hands returned to a rest position. To the stroke, the phase of the gesture bearing the most important information, the annotators had to determine the most prominent and meaningful part of the movement. Lastly, for each stroke the type of the gesture was determined. Each category was coded separately for the right and left participants.

In addition to coding gestures, speech coding was implemented. First, noise was automatically removed from the audio using Audacity (Audacity Team, 2014). Next, using a custom R-script created by the supervisor, the start and stop times of speech were automatically identified based on the speech amplitude in the recorded audio file. When a pause between speech segments was shorter than 250 ms, these segments were identified as one segment. The resulting csv-files were imported into ELAN. To clarify the speaker in each speech segment, the annotators manually adjusted the annotation value to part.R, part.L, or exp. leader. When the R-script had failed to distinguish between speech segments by two participants, annotations were manually split. When two participants spoke indistinguishably, that is, at the same time, the annotation “part.R.part.L” was assigned to the segment.

Lastly, using another custom R-script, the data from the demographics questionnaire, C-test, and coded ELAN files were integrated in a final datafile which enabled further statistical analysis of the research questions. The calculation of variables relevant for the research question of this bachelor thesis will be described next.

Data Analysis

Annotation files were first processed and converted into a data file by the supervisor. The data file contained values per dyad. Relevant data for the current thesis included: gesture frequency total and per gesture type, average duration of gestures, condition (L1-L1, L2-L2, L1-L2 group membership), dyad number and raw and normalised proficiency scores (C test). The metrics for gesture alignment between the participants in a dyad (gesture frequency difference, gesture duration difference) and proficiency difference values were calculated as the absolute difference between the values of the first and second participant within a dyad. Each dyad was therefore analysed as individual datapoint. Those calculations were performed using Excell (Microsoft Corporation, 2024). All the subsequent statistical analyses were performed using JASP (JASP Team, 2024).

Any possible outliers were investigated individually and were included in the analysis only if the gestures were coded correctly. If the data did not correspond to the observed gesture use in the video recording it was not included.

To see if the groups exhibit differences in gesture alignment, the main analysis involved analysis of variance (ANOVA) with condition (L1-L2, L2-L2, L1-L1 group membership) as a factor and frequency difference as dependent variable to answer the first part of the research question. A second ANOVA involved duration difference as dependent variable and condition as a factor as means to answer the second part of the research question. If the assumptions were not met, a more suitable statistics were chosen based on the violated assumption. For example, Kruskal-Wallis might be the preferable nonparametric option in

case of the violation of the normality assumption, as it has been found to be more robust than ANOVA in such cases (Nwobi & Akanno, 2021).

As additional exploratory analysis, absolute proficiency difference per dyad was calculated by subtracting the normalised proficiency scores on the C-test of the two participants in a dyad and taking the absolute value to discard directionality. Those values were used as a part of describing the dataset in the preliminary analysis. Subsequently Spearman's rho was used to assess associations between proficiency difference and gesture alignment metrics.

Results

Our research question was: Do groups L1-L2, L1-L1, L2-L2 differ in gesture alignment in terms of frequency difference and duration difference within a dyad?

Upon inspecting the raw data. The only dyad that stood out was dyad number 32. This dyad was eventually removed from the final analysis because the coding of the specific gestures was incomplete and therefore the total number of gestures in the dataset showed zero, which was incorrect. After excluding this outlier, we proceeded with describing the data.

Descriptive statistics including means, medians, standard deviations and minimum and maximum values for each condition group can be found in Table 1. As a part of our descriptives, we also looked at proficiency difference between the participants. We calculated the proficiency difference per dyad using normalized scores on the C test. To see if there is statistically significant difference between the groups, we wanted to conduct a one-way ANOVA, using absolute proficiency difference values as dependent variable and condition group as a factor. However, when checking for normality assumption by inspecting histograms for each group, we found that group L1-L1 seemed not normally distributed (Figure 7, Appendix A), which suggests that the assumption is violated. Therefore, we opted for a non-parametric Kruskal-Wallis test. Condition group was set as a factor and within dyad

proficiency difference was the dependent variable. The test showed that there was no effect of condition groups on the proficiency difference within a dyad, $\chi^2(2) = 2.60$, $p = .27$ with median scores for each group found in Table 1. This may indicate that the difference in proficiency within a dyad did not differ across the condition groups.

Table 1

Descriptive statistics

	Frequency difference			Duration difference			Proficiency difference		
	L1-L1	L1-L2	L2-L2	L1-L1	L1-L2	L2-L2	L1-L1	L1-L2	L2-L2
Valid	9	5	18	9	5	18	9	5	18
Median	43	28	41.50	1.47	0.80	0.61	0.84	2.20	1.02
Mean	31.67	30.40	42	2.37	0.67	1.49	1.07	2	1.09
SD	23.50	22.03	27.09	2.42	0.62	2.69	0.56	1.20	0.67
Minimum	0	3	7	0.17	0.04	0.06	0.41	0.49	0
Maximum	63	58	123	7.10	1.54	11.38	1.88	3.50	2.36

Group Differences in Gesture Frequency Alignment

For the first analysis we wanted test whether the condition groups differ significantly in terms of gesture frequency alignment within a dyad. We first checked assumptions for ANOVA. We inspected if the data is normally distributed by first looking at histograms for each group. It seemed that the data is not normally distributed per group. All the histograms seemed randomly distributed or rather skewed as shown in Figures 1-3 in the Appendix A. Therefore, we decided to do a non-parametric Kruskal-Wallis test instead, using condition group as factor and frequency difference within a dyad as a dependent variable. The test showed no significant difference between the groups in terms of gesture frequency difference within a dyad, $\chi^2(2) = 0.84$, $p = 0.66$, Therefore, all condition groups demonstrated comparable levels of alignment in terms of gesture frequency.

Group Differences in Gesture Duration Alignment

Initially, ANOVA was chosen as a test of between group differences in gesture duration alignment. However, histograms again revealed non normally distributed data across condition groups as shown in Figures 3-6 in the Appendix A. Because the normality assumption was not met, we performed a nonparametric Kruskal-Wallis test which does not assume normality of the data and is therefore more appropriate for the current analysis. The Kruskal-Wallis test showed that condition groups did not significantly affect gesture duration alignment, $\chi^2(2) = 3.34$, $p = 0.19$, suggesting that the groups did not differ in gesture duration alignment.

Additional Follow-up Analysis

To investigate possible explanations for our non-significant results, additionally we also decided to look at correlations between gesture alignment values and proficiency difference to explore the possible relationship between gesture alignment and proficiency difference within a dyad. To check the normality assumption for Pearson's corelation we preformed Shapiro-Wilk test which showed significant departure from normality for both frequency difference ($W = 0.91$, $p = .01$) and duration difference ($W = 0.64$, $p < .001$), suggesting the violation of the assumption. Therefore, Spearman's rho was chosen instead of Pearson's correlation coefficient. Table 2 shows all the correlations between variables and their p-values. All three correlations were very low and none of the correlations were statistically significant, which suggests that there was no relation between gesture alignment and proficiency difference within a dyad.

Table 2

Pearson's Correlations

Variable		Proficiency difference	Frequency difference
1. Frequency difference	Spearman's rho	-0.03	—
	p-value	0.88	—
2. Duration difference	Spearman's rho	0.06	0.05
	p-value	0.76	0.80

Discussion

The current study investigated interpersonal gesture alignment in dyadic conversations. Dyads belonged to L1-L2, L2-L2 or L1-L1 group depending on participant's spoken languages.

For the first part of our research question, we looked at how gesture alignment manifests in different language context groups in terms of gesture frequency. We found no significant effect of language condition on the gesture frequency alignment. Therefore, we didn't find evidence that gesture alignment in terms of gesture frequency differs across language context groups. In addition, all three condition groups showed large standard deviations, indicating large spread of the frequency alignment measure. The second part of the research question was concerned with group differences in gesture duration alignment, which was measured as between participant difference in gesture duration. No significant effect of language groups on gesture duration alignment was found.

Additional analysis revealed no effect of language group on within dyad proficiency alignment, indicating similarity of language skills between participants across different

language groups. We also found no correlation between gesture alignment and proficiency difference. To summarize, we found no differences between language context groups in neither of gesture frequency alignment, gesture duration alignment, nor in terms of language proficiency differences. The meaning of those results will be discussed in the next paragraph.

Interpretation

The results of the main analysis suggest that our study does not provide evidence for gesture alignment differences between language context groups (L1-L2, L2-L2, L1-L1). The current study did not make any predictions about specific differences between the groups rather it aimed at the exploration of those differences. The large spread of gesture frequency alignment measure, suggested that frequency difference between dyads was highly variable. A possible explanation for this may be that participants had strong individual differences in how many gestures they use, that persisted in the interaction. For example, Chu et al. (2014) found that levels of empathy and cognitive abilities such as working memory were related to how many gestures people use.

Unlike other studies mentioned in the introduction (Opazo et al., 2024, Holler & Wilkin, 2011), which have investigated alignment in terms of matching gestures, the current study aimed to evaluate the overall pattern of gesture use between the participants in each dyad. By looking at how similar participants in each dyad are in terms of how many gestures they use (gesture frequency) and how long they gesture (gesture duration), our findings contribute to the existing literature by addressing alignment in a slightly different way than those previous studies. Our study looked at alignment in gesture frequency and duration, whereas for example Opazo et al. (2024) relied on analysis of pairs of similar gestures occurring in both speakers. Since the methods to measure alignment were not the same, one should be cautious when comparing those results. When it comes to the theoretical implications of our findings, in contrast to Holler and Wilkin (2011), who’s findings

supported the grounding perspective of alignment, the current study the was not designed to test a specific theoretical framework, therefore we cannot confidently determine whether it supports a specific perspective of alignment. To better interpret our results, it is important to consider our findings regarding proficiency alignment within dyads.

Apart from the main results, an interesting pattern emerged as a part of additional analysis. We found no differences between the groups in terms of proficiency difference within a dyad. In other words, different language groups may have exhibited comparable levels of within dyad proficiency differences. The rationale for this analysis was to establish if those groups even differ in proficiency difference in a meaningful way. Considering that the sample consisted of only undergraduate students with similar backgrounds it would be possible that the proficiency of those students was too similar to see any differences between different condition groups.

Therefore, we also decided to explore the relationship between proficiency difference and gesture alignment. This analysis showed small or no relationship between the variables and none of the correlations were significant, which further complements our results regarding proficiency alignment. Previous research regarding the relationship between language proficiency and gesture production, has been inconclusive.

Zvaigzne et al. (2019) found that in children, the relationship between language proficiency and gesture frequency seemed of quadratic rather than linear nature. Higher gesture frequencies were observed in the lowest and highest proficiency groups, whereas intermediate proficiency group showed the lowest gesture frequency. Moreover, Gregersen et al. (2009), found that people produce more gestures when speaking their L1 compared to their L2, and that more proficient advanced L2 speakers produce more gestures than beginner or intermediate L2 speakers. Importantly, their study suggested that difference in gesture use was explained by proficiency rather than individual differences in gesturing. Possibly, we

didn’t find significant difference in gesture alignment because of low variability in language proficiency.

Limitations

One of the limitations of the current study concerns the inter-rater reliability in the annotation process, despite extensive training in gesture annotation. Unfortunately, due to time constraints, every annotator coded their portion of video recordings independently. In the ideal situation the annotations done by different coders would be compared to establish reliable annotations (Tütüncübasi et al., 2023). In the current study we tried to enhance this by discussion between annotators on how they annotate certain instances, which was helpful but not ideal. Another factor important to note is that annotation process in general involves a fair amount of subjective judgement of the annotator. This means that it is possible that there were slight differences in how each annotator interprets a gesture to be of a specific type, or even when the gesture starts or ends.

When it comes to the design of the current study the main concern is the generalizability of the findings. As the design involved convenience sample of undergraduate psychology students the generalization to larger population is rather limited. Another possible factor that could have influenced our finding was existing rapport between the participants. Because a lot of the participants were part of the same study programme there have been pre-existing familiarity or rapport within a dyad in some cases. Some participants expressed that they knew each other or that they were friends, which could influence the degree of alignment between the participants. As mentioned in the introduction gesture alignment has a strong link to connectedness and rapport between the interacting individuals (Schmidt & Richardson, 2008). All those limitations and factors possibly influencing gesture use and subsequent gesture alignment should therefore be carefully considered when conducting future research.

Future Directions

As mentioned before, the link between language proficiency and gesture use and gesture alignment should be one of the areas for future investigations. A study design involving participants with more variable proficiency should be implemented in the future. A second suggestion would be to involve more control into the research design by checking for the degree of pre-existing rapport or familiarity between participants. This could be done simply as a score on the scale indicating degree of connectedness with the other participant. Next, because the current analysis only involved alignment in totality of hand gestures, an interesting follow-up would be to examine alignment in specific types of gestures. It might be possible that a certain type of gestures might be more aligned in different dyad configurations.

Lastly, the effect of gesture alignment on effective communication could also be an interesting idea to explore. This would have important implications in fields where effective communication is crucial, such as teaching environments or psychotherapy. That said, studies investigating gesture alignment in teaching environments have been conducted in the past (see, e.g., Opazo et al., 2024; Majlesi, 2015) and this seems like an important field that could benefit from integrating gestures and gesture alignment in a very practical way.

Conclusion

To conclude, the current study found that gesture alignment did not differ significantly across different language context groups. Some of the explanations for why the current study failed to find differences between the condition groups might be preexisting rapport between the interlocutors, participants' individual style in gesture frequency or, as additional analysis revealed, no meaningful differences in proficiency difference within a dyad between the language groups. The proficiency difference might have been too low to see differences in individuals gesture use and therefore speaking in different language group would make no difference. Gesture alignment involves complex interplay of intra- and interpersonal factors.

The overall dynamics of the gesture alignment is not simply the sum of individual gesture use of the two persons within a dyad. As in most of the social phenomena, it is difficult to pinpoint accurately what is going on and nearly impossible to control for everything.

References

Audacity Team. (2014). Audacity(R): Free Audio Editor and Recorder (3.7.3) [Software]. http://audacity.sourceforge.net/

Ayache, J., Connor, A., Marks, S., Kuss, D. J., Rhodes, D., Sumich, A., & Heym, N. (2021). Exploring the "Dark Matter" of Social Interaction: Systematic Review of a Decade of Research in Spontaneous Interpersonal Coordination [Systematic Review]. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.718237

Bernieri, F. J., & Rosenthal, R. (1991). Interpersonal coordination: Behavior matching and interactional synchrony. In R. S. Feldman & B. Rimé (Eds.), *Fundamentals of nonverbal behavior.* (pp. 401-432). Editions de la Maison des Sciences de l'Homme Cambridge University Press. http://search.ebscohost.com.proxy-ub.rug.nl/login.aspx?direct=true&db=psyh&AN=1991-98576-011&site=ehost-live&scope=site

Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: The perception–behavior link and social interaction. *Journal of Personality and Social Psychology*, *76*(6), 893-910. https://doi.org/10.1037/0022-3514.76.6.893

Chartrand, T. L., & Lakin, J. L. (2013). The Antecedents and Consequences of Human Behavioral Mimicry. *Annual Review of Psychology*, *64*(Volume 64, 2013), 285-308. https://doi.org/https://doi.org/10.1146/annurev-psych-113011-143754

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In *Perspectives on socially shared cognition.* (pp. 127-149). American Psychological Association. https://doi.org/10.1037/10096-006

Coleman, J., Grotjahn, R., Raatz, U., & Wockenfuß, V. (2017a). *Gap-fill / Cloze test / Fill in the blank from "Das TESTATT-Projekt: Entwicklung von C-Tests zur Evaluation des*

Fremdsprachenlernerfolgs" [Text/Language test]. IRIS Database, University of York, UK. https://www.iris-database.org/details/7ZlyH-wLAPu

Coleman, J., Grotjahn, R., Raatz, U., & Wockenfuß, V. (2017b). *Gap-fill / Cloze test / Fill in the blank from "Das TESTATT-Projekt: Entwicklung von C-Tests zur Evaluation des Fremdsprachenlernerfolgs"* [Text/Language test]. IRIS Database, University of York, UK. https://www.iris-database.org/details/M3rMX-RKx9r

Chu, M., Meyer, A., Foulkes, L., & Kita, S. (2014). Individual differences in frequency and saliency of speech-accompanying gestures: The role of cognitive abilities and empathy. Journal of Experimental Psychology: General, 143(2), 694-709. https://doi.org/10.1037/a0033861

de Jonge-Hoekstra, L. (2021). *How hand movements and speech tip the balance in cognitive development: A story about children, complexity, coordination, and affordances*. [Thesis fully internal (DIV), Rijksuniversiteit Groningen]. University of Groningen. https://doi.org/10.33612/diss.172252039

ELAN (Version 6.9). (2024). [Software]. Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive. https://archive.mpi.nl/tla/elan

Emir Özder, L., Özer, D., & Göksun, T. (2023). Gesture use in L1-Turkish and L2-English: Evidence from emotional narrative retellings. *Quarterly Journal of Experimental Psychology*, *76*(8), 1797-1816. https://doi.org/10.1177/17470218221126685

Fauviaux, T., Marin, L., Parisi, M., Schmidt, R., & Mostafaoui, G. (2024). From unimodal to multimodal dynamics of verbal and nonverbal cues during unstructured conversation. *PLOS ONE*, *19*(9), e0309831. https://doi.org/10.1371/journal.pone.0309831

Freeborn, L., & Andringa, S. (2023). Development of a C-test to measure Dutch language proficiency. *OSF*. https://doi.org/10.17605/OSF.IO/VNTG3

Fujiwara, K., & Daibo, I. (2016). Evaluating Interpersonal Synchrony: Wavelet Transform Toward an Unstructured Conversation [Methods]. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.00516

Gilmore, A. (2013). *C-test. Gap-fill / Cloze test / Fill in the blank from ""I prefer not text": Developing Japanese learners' communicative competence with authentic materials"* [Text/Language test]. IRIS Database, University of York, UK. https://www.iris-database.org/details/GXgIl-kFpif

Goldin-Meadow, S. (1999). The role of gesture in communication and thinking. *Trends in Cognitive Sciences*, *3*(11), 419-429. https://doi.org/https://doi.org/10.1016/S1364-6613(99)01397-2

Gregersen, T., Olivares-Cuhat, G., & Storm, J. (2009). An Examination of L1 and L2 Gesture Use: What Role Does Proficiency Play? *The Modern Language Journal*, *93*(2), 195-208. https://doi.org/https://doi.org/10.1111/j.1540-4781.2009.00856.x

Holler, J., & Wilkin, K. (2011). Co-speech gesture mimicry in the process of collaborative referring during face-to-face dialogue. *Journal of Nonverbal Behavior*, *35*(2), 133-153. https://doi.org/10.1007/s10919-011-0105-6

JASP. (2025). [Software]. In *JASP - Free and User-Friendly Statistical Software* (0.19.3). https://jasp-stats.org/

Kendon, A. (2004). *Gesture: Visible Action as Utterance*. Cambridge University Press. https://doi.org/DOI: 10.1017/CBO9780511807572

Kimbara, I. (2006). On gestural mimicry. *Gesture*, *6*(1), 39-61.

Kimbara, I. (2008). Gesture Form Convergence in Joint Description. *Journal of Nonverbal Behavior*, *32*, 123-131. https://doi.org/10.1007/s10919-007-0044-4

LaFrance, M. (1985). Postural mirroring and intergroup relations. *Personality and Social Psychology Bulletin*, *11*(2), 207-217. https://doi.org/10.1177/0146167285112008

Lin, Y.-L. (2022). Speech-accompanying gestures in L1 and L2 conversational interaction by speakers of different proficiency levels. *International Review of Applied Linguistics in Language Teaching*, *60*(2), 123-142. https://doi.org/doi:10.1515/iral-2017-0043

Louwerse, M. M., Dale, R., Bard, E. G., & Jeuniaux, P. (2012). Behavior Matching in Multimodal Communication Is Synchronized. *Cognitive Science*, *36*(8), 1404-1426. https://doi.org/https://doi.org/10.1111/j.1551-6709.2012.01269.x

Lundqvist, L.-O., & Dimberg, U. (1995). Facial expressions are contagious. *Journal of Psychophysiology*, *9*, 203-211.

Ma, S., & Jin, G. (2022). The relationship between different types of co-speech gestures and L2 speech performance [Original Research]. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.941114

MacNeill, D. (1992). *Hand and mind : what gestures reveal about thought*. The University of Chicago Press.

Majlesi, A. R. (2015). Matching gestures – Teachers' repetitions of students' gestures in second language learning classrooms. *Journal of Pragmatics*, *76*, 30-45. https://doi.org/https://doi.org/10.1016/j.pragma.2014.11.006

Microsoft Corporation. (2024). *Excel (Microsoft 365 Subscription)*. [Computer Software]. Microsoft Corporation. https://www.microsoft.com/en-us/microsoft-365/excel#:~:text=Microsoft%20Excel%20with%20a%20Microsoft,Excel%202007%2C%20and%20Excel%202003.

Nwobi, F., & Akanno, F. (2021). Power comparison of ANOVA and Kruskal–Wallis tests when error assumptions are violated. *Advances in Methodology and Statistics*. https://doi.org/10.51936/ltgt2135.

Opazo, P., Cienki, A., Oben, B., & Brône, G. (2024). Gesture alignment in teacher-student interaction: a study concerning office hour consultations using English as the

lingua franca. *Frontiers in Communication*, *9*, 22, Article 1457533. https://doi.org/10.3389/fcomm.2024.1457533

Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and brain sciences*, *27*(2), 169-190.

Pickering, M. J., & Garrod, S. (2006). Alignment as the Basis for Successful Communication. *Research on Language and Computation*, *4*(2), 203-228. https://doi.org/10.1007/s11168-006-9004-0

Platek, S. M., Critton, S. R., Myers, T. E., & Gallup, G. G. (2003). Contagious yawning: the role of self-awareness and mental state attribution. *Cognitive Brain Research*, *17*(2), 223-227. https://doi.org/https://doi.org/10.1016/S0926-6410(03)00109-5

Qualtrics XM - Experience Management Software. (2025). [Software]. In *Qualtrics*. https://www.qualtrics.com/

Rasenberg, M., Özyürek, A., & Dingemanse, M. (2020). Alignment in Multimodal Interaction: An Integrative Framework. *Cognitive Science*, *44*(11), e12911. https://doi.org/https://doi.org/10.1111/cogs.12911

Rennung, M., & Göritz, A. S. (2016). *Prosocial Consequences of Interpersonal Synchrony: A Meta-Analysis* (Vol. 224). Hogrefe Publishing. https://doi.org/10.1027/2151-2604/a000252

Rohrer, P. L., Tütüncübasi, U., Vilà-Giménez, I., Florit-Pons, J., Esteve-Gibert, N., Ren-Mitchell, A., Shattuck-Hufnagel, S., & Prieto, P. (2023). *The MultiModal MultiDimensional (M3D) labeling system*. https://doi.org/10.17605/osf.io/ankdx

Schmidt, R. C., & Richardson, M. J. (2008). Dynamics of Interpersonal Coordination. In A. Fuchs & V. K. Jirsa (Eds.), *Coordination: Neural, Behavioral and Social Dynamics* (pp. 281-308). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-74479-5_14

Tütüncübasi, U., Ren-Mitchell, A., Shattuck-Hufnagel, S., Rohrer, P. L., & Prieto, P. (2023). *The M3D Training Program*. https://m3d.upf.edu

Zvaigzne, M., Oshima-Takane, Y., & Hirakawa, M. (2019). How does language proficiency affect children's iconic gesture use? *Applied Psycholinguistics*, *40*(2), 555-583. https://doi.org/10.1017/S014271641800070X

Footnotes

1 In the end only the English and Dutch versions of the Cloze test were completed by participants

Appendix A

Discussion topics

1) Which came first, the chicken or the egg?

2) Would you rather have a job you dislike with a good salary or have a job you like with a bad salary?

3) Does technology improve human life overall, or does it cause more harm than good in the long run?

4) Would you rather be very funny or very intelligent?

5) Is it better to know the truth, even if it makes you unhappy, or to live blissfully ignorant of unpleasant realities?

6) Would you rather only write papers/assignments or only make exams for school?

7) If you replace all parts of a ship over time, is it still the same ship? If not, at what point did it cease to be the original?

8) Would you rather have no summers or no winters for the rest of your life?

9) Is progress driven more by original innovations or by improvements on existing ideas?

10) Do humans truly have free will to make choices, or are all events and decisions predetermined by prior causes?

Figure 1

L1-L1 – frequency difference

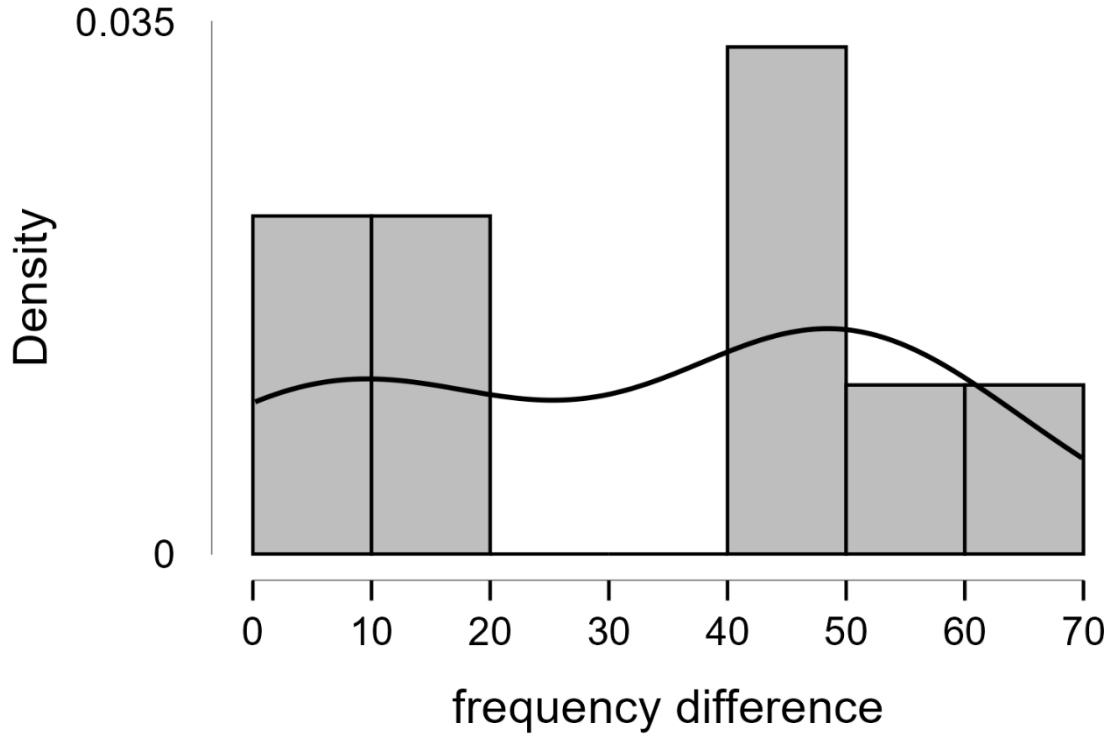

Figure 2

L1-L2 – frequency difference

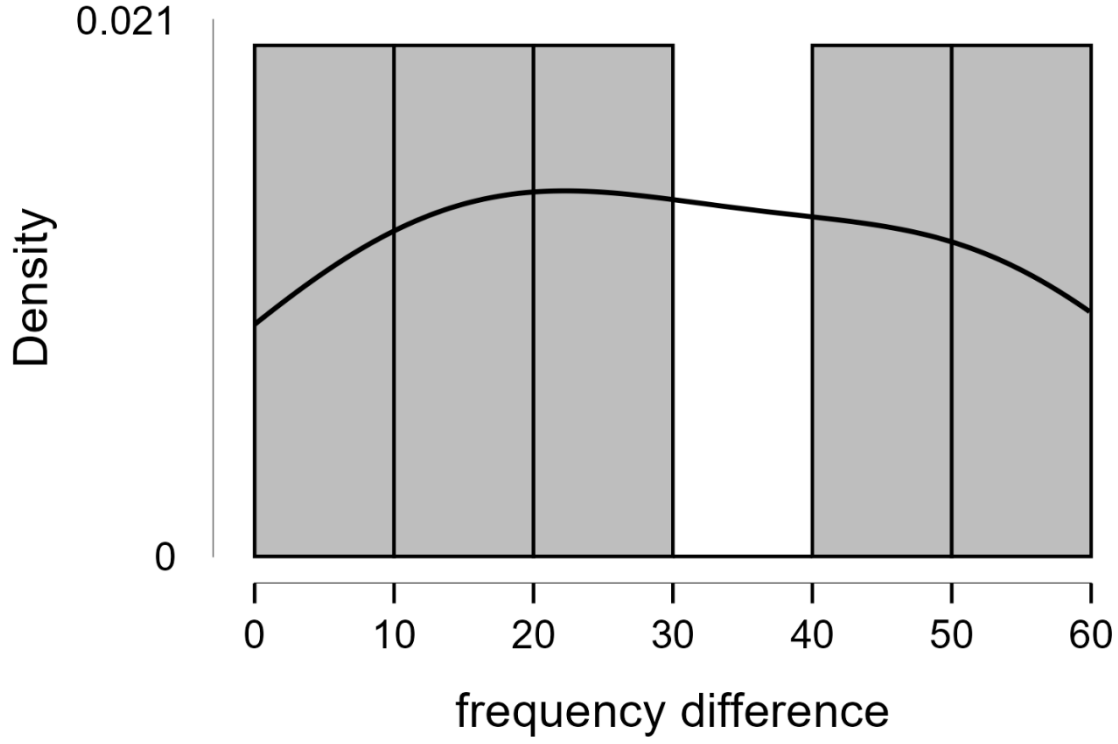

Figure 3

L2-L2 - frequency difference

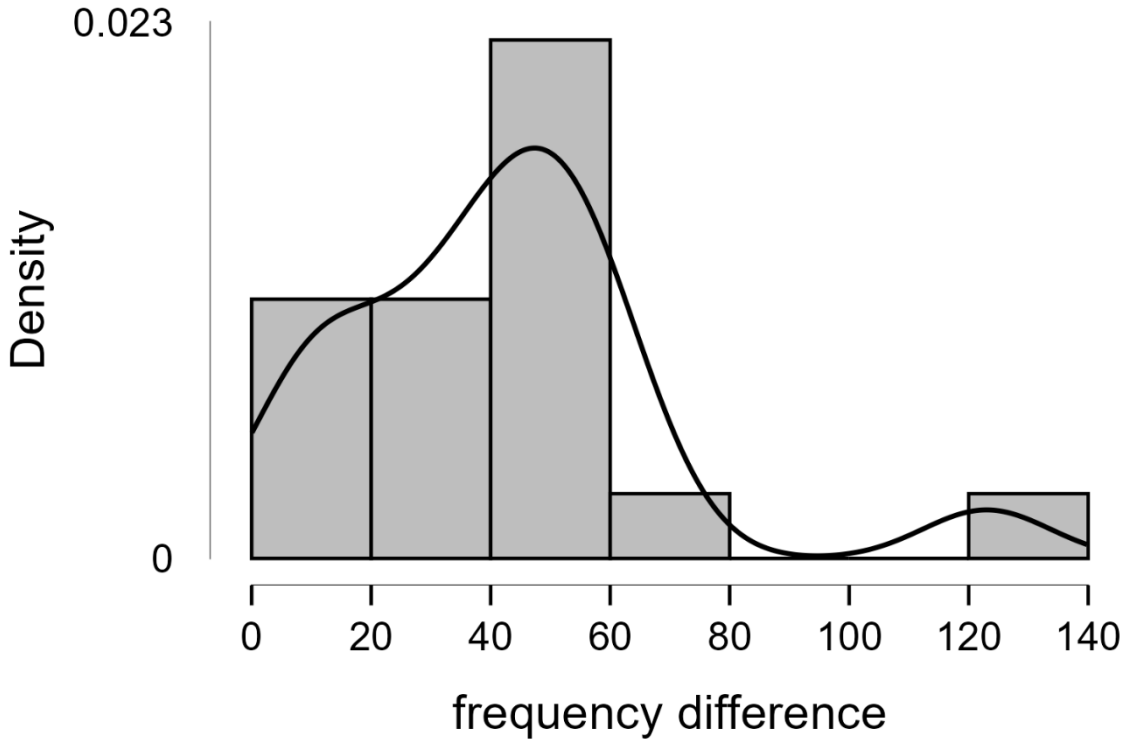

Figure 4

L1-L1 - duration difference

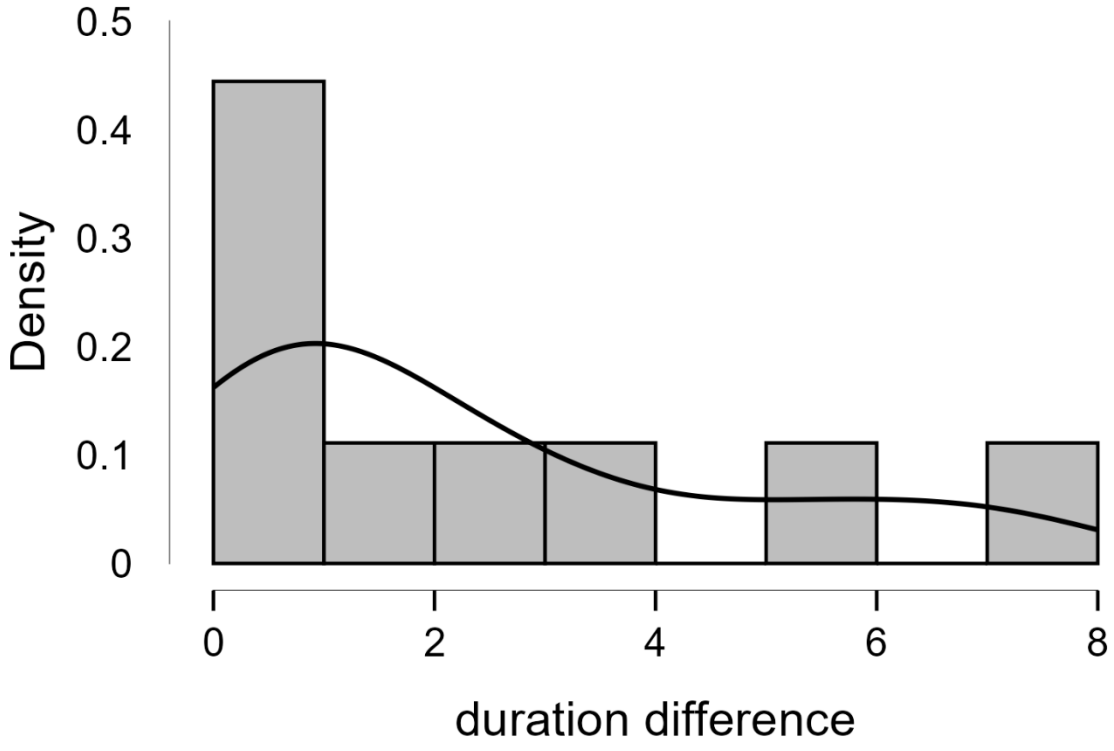

Figure 5

L1-L2 - duration difference

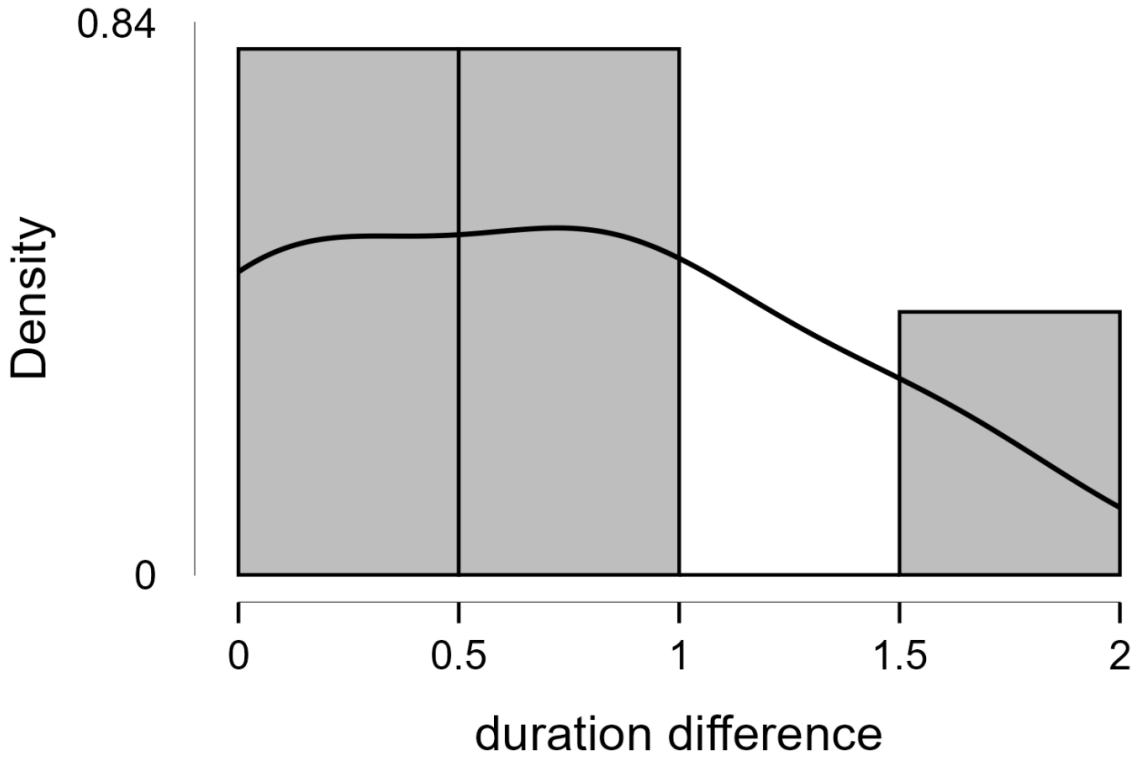

Figure 6

L2-L2 - duration difference

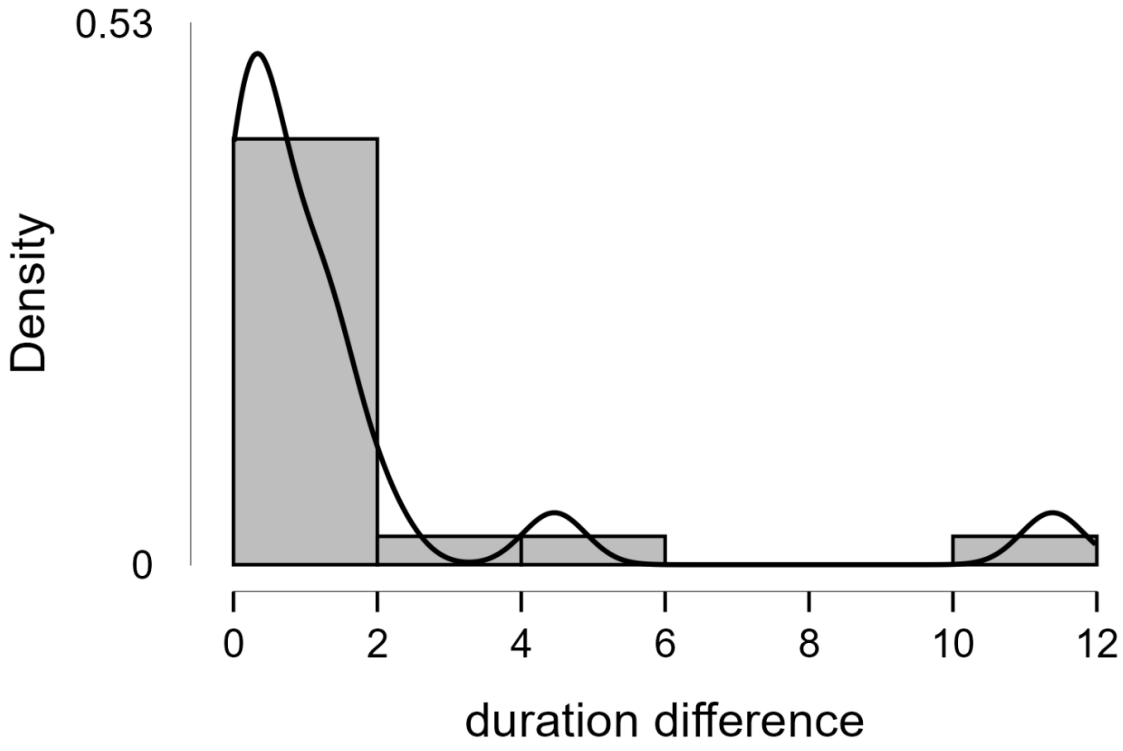

Figure 7

L1-L1 – language proficiency difference within a dyad