

**Citation:** Karpiński, M., Czoska, A., Jarmolowicz-Nowikow, E., Juszczuk, K., & Klessa, K. (2018). Aspects of gestural alignment in task-oriented dialogues. *Cognitive Studies / Études cognitives*, 2018(18). <https://doi.org/10.11649/cs.1640>

MACIEJ KARPIŃSKI<sup>A</sup>, AGNIESZKA CZOSKA<sup>B</sup>, EWA JARMOŁOWICZ-NOWIKOW<sup>C</sup>  
KONRAD JUSZCZYK<sup>D</sup>, & KATARZYNA KLESSA<sup>E</sup>

Adam Mickiewicz University in Poznań, Poland

<sup>A</sup>maciejk@amu.edu.pl ; <sup>B</sup>agaczoska@gmail.com ; <sup>C</sup>ewa@jarmolowicz.art.pl

<sup>D</sup>juszczuk@amu.edu.pl ; <sup>E</sup>klessa@amu.edu.pl

## ASPECTS OF GESTURAL ALIGNMENT IN TASK-ORIENTED DIALOGUES

### Abstract

Interlocutors in a conversation influence each other in a number of dimensions. This process may lead to observable changes in their communicative behaviour. The directions and profiles of these changes are often correlated with the quality of interaction and may predict its success. In the present study, the gestural component of communication is scrutinised for changes that may reflect the process of alignment. Two types of task-oriented dialogues between teenagers are recorded and annotated for gestures and their features. We hypothesize that the dialogue task type (collaborative vs. competitive), as well as certain culture-specific properties of alignment that differ between German and Polish pairs, may significantly influence the process of communication.

In order to explore the data and detect tendencies in gestural behaviour, automatised annotation mining and statistical exploration have been used, including a moving frame approach aimed at the investigation of co-occurring strokes as well as re-occurring strokes and their features. Significant differences between German and Polish speakers, as well as between the two dialogue types, have been found in the number of gestures, stroke duration and amplitude.

**Keywords:** communicative alignment; intercultural communication; multimodal interactions; gesture function; gesture re-occurrence; gesture co-occurrence

## 1 Introduction (state-of-the-art, background)

Participants of conversational interaction mutually adjust their behaviour in time and in form. Recent research shows that external, visible accommodation is based on internal processes related to participation in a dialogue. Conversational partners adjust their cognitive, emotional and interactive processes. As a result, they align their mental representations on many levels of language organisation, as well as in facial expression and body movements, including hand gestures and head movements. In the framework of the mechanistic theory of dialogue (Chartrand & Bargh,

1999; Pickering & Garrod, 2004), syntactic and semantic structures of utterances are subject to gradual adjustment and unification as the participants follow a common aim in interaction. Alignment is considered to be a predictor of success in the process of dialogue (Chartrand & Bargh, 1999; Pickering & Garrod, 2004).

In an interactive alignment account (Chartrand & Bargh, 1999; Pickering & Garrod, 2004), each subsequent turn of each speaker involves certain representations of linguistic structures: the semantic, syntactic and sound features of the utterance. These representations become aligned in both speakers as they construct a shared representation of their situation and form, referred to as a conceptual pact (Clark & Brennan, 1991). Interlocutors are thought to “prime each other to speak about things in the same way, and people who speak about things in the same way are more likely to think about them in the same way as well” (Garrod & Pickering, 2013). Alignment on one level is the basis for another level, such as lexical alignment for the syntactic (Chartrand & Bargh, 1999; Pickering & Garrod, 2004).

Human communication is multimodal by nature (Bonacchi & Karpiński, 2014; Karpiński, 2014). People communicate via linguistic units accompanied by hand movements, gaze, head movements and posture. Recent research has shown that all modalities of communication tend to align in time and in form. The process of aligning gestures or body postures is called motor mimicry or imitation and it has been proved to be largely unconscious and automatic during interaction (Chartrand & Bargh, 1999; Pickering & Garrod, 2004). Empirical studies have shown that motor mimicry is linked to mutual liking and good relationships between interactants as well, as the quality of relationships between interactants (Chartrand & Bargh, 1999; Chartrand, Maddux, & Lakin, 2006; Kulesza, 2016). Dialogue participants who repeat each other’s lexical units and syntactic structures are priming each other and have a greater probability of reaching understanding (conversational success). In other words, effective dialogue depends on interactive alignment (Garrod & Pickering, 2013).

Co-speech gestures are an especially important channel for communication as they allow speakers to express both propositional and affective content closely related to speech or add extra information (Kendon, 2004; McNeill, 1992). Alignment or adaptation in representational gestures resembles adaptation in verbal references. Importantly, gesture forms were only repeated across speakers if they had occurred in a meaningful context whereas other gestures seemed to be neglected (Mol, Krahmer, Maes, & Swerts, 2012). Gestural alignment is stronger in the case of co-speech gestures (Garrod & Pickering, 2013; Mol et al., 2012)

Systems of gesture coding divide hand movements into emphatic, deictic, iconic and emblematic (McNeill, 1992). However, in coding systems such as NEUROGES (Lausberg, 2013), emphatic hand movements can be superimposed on other types of gestures, such as deictic or form presentation. In most cases these emphatic gestures are performed repetitively in a gesture space and are hence called repetitive in space in NEUROGES labels (Lausberg, 2013).

There is a strong but flexible relationship between emotional and communicative alignment in interaction. Communicative alignment is often linked to empathy or an affective theory of mind (Jaecks et al., 2013). One interactant needs to know the internal state of the other, including his or her thoughts and feelings so that he or she can address and adjust his or her behaviour to current state of mind of the other. Recognition of the internal state of the other is based mainly on observing how the behaviour of the other is similar to his or her own behaviour. This process is assumed to be automatic and unconscious for both interactants (Chartrand & Bargh, 1999), but it is visible for a patient observer and researcher. Due to its popularity, it is referred to as the “Chameleon effect” (Chartrand & Bargh, 1999) and it has been shown to facilitate the smoothness of interactions and increase liking between interaction partners (Chartrand & Bargh, 1999).

Similarity in behaviour between two interactants is sometimes called “mimicry” or “contagion”, while the mental representation of the other’s internal state is referred to as “alignment” or “simulation”. Similar behaviour aligned in time is called synchrony (Jaecks et al., 2013; Ramseyer & Tschacher, 2011). However, this kind of co-ordination may not be obvious as co-ordinated acts may be sequential by nature or form parts of larger structures. The alignment between interactants

happens on all modalities of communication and all levels of language. Research provides evidence for alignment in words for naming objects (Brennan & Clark, 1996) or the same syntactic structure (Cleland & Pickering, 2003), segmental phonetic (Pardo, 2006) or prosodic features (Guitar & Marchinkoski, 2001; Truong & Heylen, 2012).

Following earlier studies in speech alignment by the authors (e.g. Czoska, Klessa, Karpiński, & Nowikow-Jarmołowicz, 2015; Karpiński, Klessa, & Czoska, 2014), this paper analyses gesture units for their alignment in the course of two types of dialogue tasks: a collaborative and a competitive task. All the analyses described in the present paper are based on the Borderland corpus data, collected as part of a project dedicated to the investigation of the paralinguistic features of interpersonal communication in the borderland region of Słubice (Poland) and Frankfurt Oder (Germany) – on the border of languages and cultures (e.g., Karpiński & Klessa, 2018). The hypothesis is that in the cooperative condition the participants’ communicative behaviour will be more compatible to one another, and consequently, they will be more likely to mimic their interlocutor’s gestures. Conversely, in the competitive condition the dialogue parties will display more differences in the usage of gestures and their functions. Although some of these processes may be immediate, the paper focuses on those that are more stretched in time and possible to discover by comparing the properties of gestural behaviour on the macro scale (e.g. in the initial and in the final part of dialogue).

## 2 Idea and aims of the study

Based on the findings reported in the previous section, it may be hypothesized that dialogue parties tend to align locally or globally, and that this process may involve convergence in various domains, including gesture and head movements. Interaction and alignment in the gesture domain may reflect important aspects of dialogue flow and the quality of interpersonal communication (Karpiński, 2014). The aim of the present study is to explore this type of convergence in task-oriented dialogues by Polish and German teenagers (see Section 3 for more details about data collection).

Although the concept of accommodation emerged in the studies of interpersonal communication in the early 1970s (Giles & Smith, 1979; Giles, Taylor, & Bourhis, 1973), measuring mutual influences in the behaviour of dialogue participants remains a challenging task even if only observable parameters are taken into consideration (Campbell & Scherer, 2010; Ward & Litman, 2007), with some recent studies even showing tendencies contrary to earlier expectations (Healey, Purver, & Howes, 2014).

In order to capture the process of gestural alignment in conversation, several approaches are combined in the present work. One of them is the moving time window approach, inspired by Kousidis et al. (Kousidis, 2010; Kousidis et al., 2008). The phenomena under study are observed and measured within a moving time window of a fixed duration, shifted by a fixed time interval. Measurements are taken and compared for all the frames to search for changes in selected behavioural parameters. The directions and sizes of these changes are compared between the dialogue participants in order to detect similarities or divergence. The time window size and step size are normally dependent on varying distributions of conversational activities, the duration of the entire dialogue under analysis, as well as the size of the units to be studied (e.g., gesture phrases or phases; Karpiński et al., 2014). A range of factors may influence the distribution of conversational units, their sizes and other properties. They include individual speaking style, as well as situation-specific speaking styles. Spontaneous, conversational speech may require broader time frames than read speech or elicited speech prepared in advance.

Among other methods of gestural accommodation measurements, there is also a search for the re-occurrence of features of gestures produced by one person in the gestures produced afterwards by his or her conversational partner. While the usage of gestural categories may also be of interest, similarity is most often limited to certain features of gesticulation like handshape or gesture size.

Another area where increasing or decreasing similarity can be found is gesture sequencing: not only does a single gesture re-occur in a partner’s stream of gesturing but, an entire sequence of gesture is re-used.

The number of factors involved in the process of alignment may be high and it may be difficult to capture statistically in fully spontaneous conversations. Specifically designed scenarios of task-oriented dialogues help to expose and isolate behaviour that may contribute to this process. One may expect that a number of modalities (semiotic modes) may be involved in the process and they may also interact among themselves, intra- or cross-modally (Karpiński, Jarmołowicz-Nowikow, & Czoska, 2015).

### 3 Data collection (participants, recording procedure, data selection, annotation)

#### 3.1 Participants

The participants of the study were quasi-randomly recruited from both a German and a Polish secondary school in Frankfurt(Oder) and Słubice, respectively. The group consisted of 15 girls and 5 boys aged from 12 to 15, who did not report any serious vision or hearing problems. All recording sessions were preceded by obtaining written consent from the participants’ parents or legal guardians.

#### 3.2 Recording scenarios and procedure

Pairs of pupils were recorded performing two types of task-oriented dialogues: a collaborative and a competitive dialogue. The collaborative task (*Tower*) involved building a tower of blocks that were only manipulated virtually, i.e. imagined by the participants. The interlocutors took turns in adding subsequent blocks and described their shapes and positions to their conversational partners so that they could both create and update a mental image of the virtual tower. In the end, they were both asked to draw (independently) the tower according to what they had managed to memorise. The competitive task (*Gift*) was focused on the selection of birthday gifts for an imaginary mutual friend. Before the task, participants were provided with photographs of a friend and lists of available gifts. However, the photographs handed to each of the participants depicted different people of vividly contrasting personalities and likes. As participants were not allowed to talk about the friend, it took some time before they realised that they were thinking of gifts for different people.

Before the recording session, the dialogue participants were invited to the room and informed, in general terms, about what they would be expected to do. They were shown to their places and listened to the formal task instructions. During the session, the participants stood facing each other from a distance of approximately 3 m. Each participant was filmed by a separate HD camcorder situated on one side of his or her conversational partner. Voices were recorded separately using a portable digital audio recorder and two large membrane condenser microphones. Recordings took place in regular classrooms. The choice of recording environment was motivated by technical and extralinguistic factors, namely, the location of the target schools, as well as the fact that the pupils could have been expected to feel more confident and at ease in a well-known surrounding than would be the case in a recording studio. No time limits were imposed on the task-oriented dialogues.

The material under study consists of 10 sessions, each comprising two dialogues: *Tower* (collaborative) and *Gift* (competitive). The total duration of the recordings is approximately 77 minutes (Germans 35 min and Poles 42 min.). The average duration of the collaborative task dialogue is 3:00 min. (for Germans) and 3min 42 s (for Poles), while for the competitive task, it is 4 min 8 s

(Germans) and 4min 42s (Poles). For the purposes of the present study, only the initial, middle and final sections, of one minute each, were annotated.

### 3.3 Data management and processing

The linguistic descriptions and analyses of the Borderland corpus were carried out using ELAN (gesture annotations; Wittenburg, Brugman, Russel, Klassmann, & Sloetjes, 2006) and Annotation Pro (orthographic transcriptions and speech segmentation; Klessa, Karpiński, & Wagner, 2013). Both tools were integrated within a common database system developed using a client-server architecture (Karpiński & Klessa, 2018) supporting consistent work management and annotation data access. Thanks to the interoperability of the tools, it is possible to import and export annotations from the format of one annotation tool to another and thus to analyse the combined effects of features from various domains within one multilayer environment.

### 3.4 Gesture annotation specification

Gesture annotation was based on a modified PAGE GAS scheme (Karpiński et al., 2015), described in more detail in Karpiński and Klessa (2018) and carried out in ELAN (Sloetjes & Wittenburg, 2008). Gestures were annotated for the left and the right hand of each participant independently, on separate tiers. The boundaries of Gesture Units and Gesture Phrases were tagged on respective tiers (GUnit and GPhrase) by experienced annotators. Further annotations were done by trained students and researchers, and revised by experienced annotators. For each GPhrase, its category was tagged (pragmatic vs. referential). Further, on GPhrase (Gesture Phase) tiers strokes were tagged as the obligatory and most meaningful gesture phases. Tags describing their features were arranged on separate tiers, hierarchically dependent to the GPhrase tier. They included handshape, gesture location (in the gestural space), gesture size, and representation technique. Additionally, head movements were annotated for each dialogue participant.

Annotation was carried out by two trained annotators supervised by one of the authors. Gestures and head movements were annotated using silent footage so that no acoustic cues would influence the decisions of the annotators. The resulting annotations were scrutinised by an experienced researcher and doubtful cases were discussed.

A hierarchically organised ELAN template provided pop-up menus containing lists of available labels for respective annotation tiers. Together with relatively long annotator training, and support and control systems, these factors contributed to an increase in coherence among annotators, and a reduction in the number of accidental mistakes in annotations.

In order to check inter-rater agreement, *kappa* tests were conducted for selected tiers and stretches of dialogues, and some tags were adjusted afterwards. These samples constitute approximately 10% of all the data analysed.

### 3.5 Annotation consistency

Two annotators (A and B) performed the annotation of the material. In order to test the stability of the annotation scheme, annotator A returned to the material after six months, and annotated a randomly selected 10% of a sample that had originally been annotated by B. Cohen's Kappa coefficient was obtained using the EasyDIAG function implemented in ELAN (Holle & Rein, 2015). The global kappa value for all labels listed is 0.8, whereas the kappa value for Referential and Pragmatic gesture functions is above 0.8 (Good). The kappa value for HandShape is above 0.9 (Very Good) for OpenPalm and Fist (although there are only two Fist gestures), and good for OneFinger and ManyFingers (>0.7). The Representation Technique tier allows for the usage of four tag values and its kappa value is between 0.8 and 0.97. The most problematic labels are in GestureSpace, where there are as many as 14 of them, so the kappa values are between 0.1 and 1.0, but in most cases around 0.7, which is still a relatively good result. Overall, the results show that

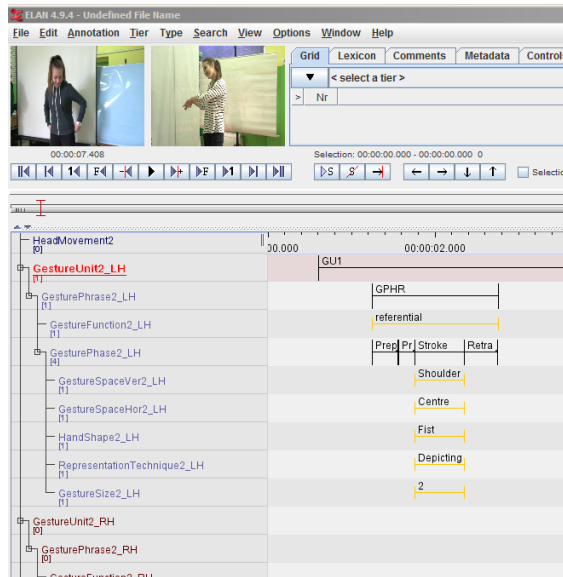


Figure 1: Configuration of annotation tiers in ELAN.

the annotation scheme is stable enough for a quantitative analysis, with the exception of some labels of Representation Technique and some labels of Gesture Space, due to the low number of segments representing those labels.

## 4 Methods

The analyses conducted within the present study encompassed the following features of gestural behaviour:

1. gesture frequency;
2. gesture function (referential vs. pragmatic);
3. duration of gesture strokes;
4. features of strokes and their co-occurrence and re-occurrence.

The values of variables based on the measurements of these features were calculated and compared between the initial and the final stages of dialogues, between dialogue types (collaborative vs. competitive), and taking into account the nationality of the participants. The co-occurrence of gesture events and functions was understood as in the previous studies by the present authors referring to speech events (Karpiński et al., 2014) or interactions between gestures and speech (Czoska et al., 2015). Co-occurrence was inspected using the moving frame (or moving windows) approach (see also: Fig. 2). The moving frame method was implemented as an Annotation Pro plugin (a C# script). The initial version of the plugin (SRMA) was tested and used beforehand to study the variability of speaking rates in two different corpora of task-oriented dialogues (Karpiński et al., 2014), as well as for the analysis of cross-modal interactions between interlocutors' hand gestures, gaze shifts and speaking rates (Czoska et al., 2015). Only data provided by adult speakers were used as study material for the preliminary analyses. The plugin enables the study of the rate of (co-)occurrence of any type of event represented by segments in time-aligned annotation layers. Segments may include not only transcriptions of linguistic and paralinguistic speech events

or gesture labels, but also the results of measurements based on annotations, representing e.g., various rhythm metrics or pitch representations. Consequently, both local and global variability of the features in question can be tracked and analysed within and between the domains of prosody (primarily rhythm and pitch), gesture (hand gestures, gaze shift and head movement), and the lexical domain.

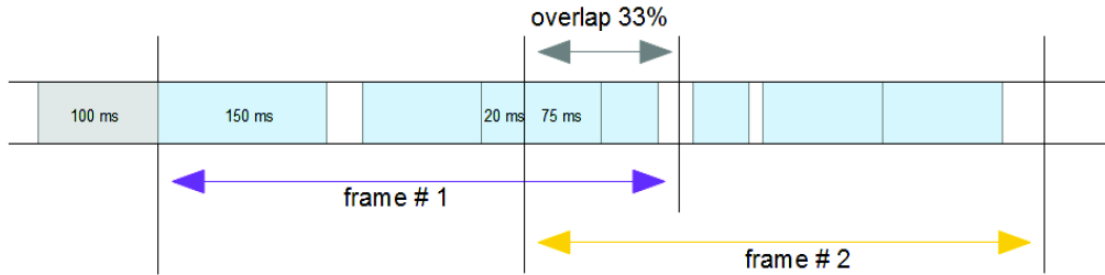


Figure 2: Moving frame analysis (co-occurrence analysis) (Figure from: Karpiński et al., 2014).

Re-occurrence (mimicry) of gestures or their features, is understood as the occurrence of a gesture or its feature in the behaviour of one of the participants within an arbitrary defined time window, following the occurrence of a gesture of the same category, or of some of its features, in the gestural behaviour of the second participant.

In the present study, re-occurrence of gestures and gesture functions is measured by means of another Annotation Pro plugin (henceforth: the Re-Occurrence plugin), designed and implemented specifically for the purposes of the present project (Figure 3). The plugin enables the counting of the number of occurrences of an annotation label found in one annotation layer (e.g., including gesture annotation for Speaker 1) in another annotation layer (e.g., including gesture annotation for Speaker 2). The number of re-occurring segments is calculated within  $n$  segments appearing after the end boundary of the original segment. The  $n$  number can be defined by the user in the plugin code. Output of the plugin provides data that include:

- the timestamps of both the original annotation segment for Speaker 1 and of each of the segments annotated with the same label that re-occur in the layer for Speaker 2;
- the durations of both the original and re-occurring annotation segment(s);
- the number of re-occurring segments.

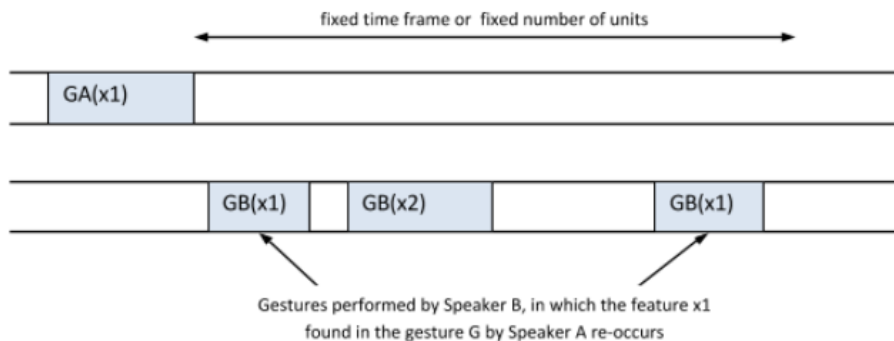


Figure 3: Re-occurrence analysis (mimicry).

## 5 Results: co-occurrence and re-occurrence of gestures

This section reports on the exploration of tendencies in alignment-related parameters of gestures, including their frequency, function, shape, location and size. The data obtained from the 10 competitive and the 10 collaborative dialogues (see section 3) contained a total of 1,194 annotated gestures (strokes), including 441 and 692 in German and Polish participants respectively. As shown in Fig. 4, the *Gift* scenario seems to evoke fewer gestures, and the differences between the dialogue participants are greater in this condition even though their roles are symmetrical. The mean number of gestures in the *Gift* scenario was  $M(Gift) = 21.1$ , while in the tower scenario  $M(Tower) = 39.7$ . When a proportion of gesture numbers is calculated (the number of gestures by Speaker 1 divided by the number of gestures by Speaker 2), the mean proportion in the *Tower* condition is 0.93, compared to 0.89 in *Gift*. Furthermore, the dialogues differed between the two languages with regard to the mean number of gestures:  $M(DE) = 23.2$  and  $M(PL) = 37.7$ .

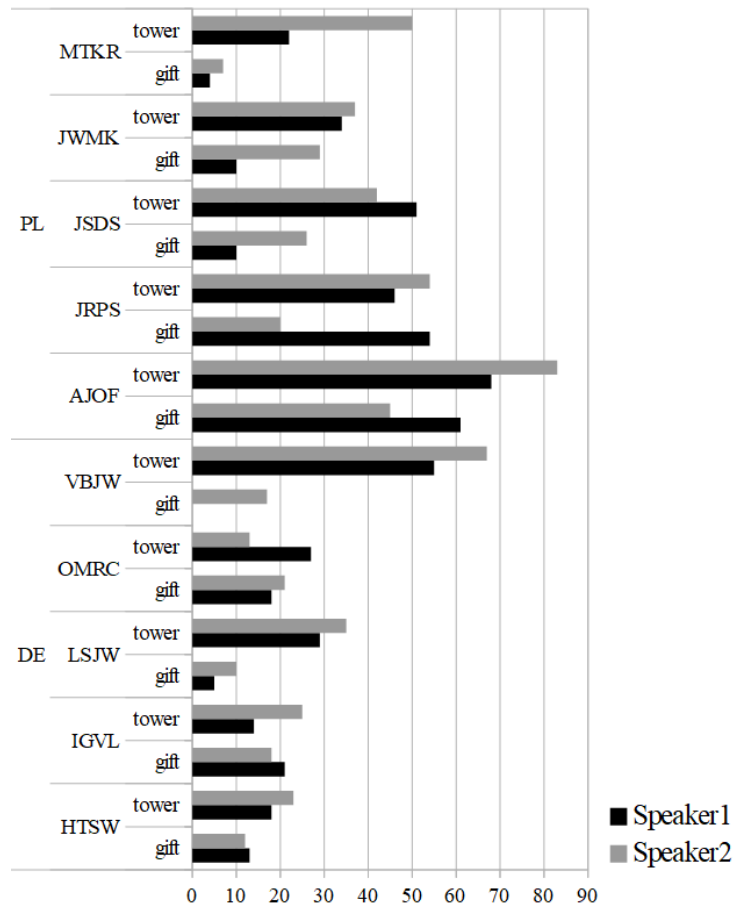


Figure 4: The number of gestures (strokes) performed by German (DE) and Polish (PL) speakers in two types of dialogues (*Tower* and *Gift*).

Two ANOVAs were conducted based on the sum of gestures produced by each of the participants in each dialogue task, i.e. 40 data points. The first shows an effect of the *task* factor ( $F = 10.54$ ,  $p = 0.0024$ ), and the second shows an effect of *language* as well ( $F = 5.774$ ,  $p = 0.0217$ ). MANOVA (whose results should be taken with caution because of the scarcity of data) on both *task* and *language* shows an effect of *language* ( $F = 8.012$ ,  $p = 0.0077$ ) and *task* ( $F = 12.357$ ,  $p = 0.0012$ ), but no interaction between the factors ( $F = 0.379$ ,  $p = 0.54$ ).



The mean duration of gesture stroke was  $M = 776.24$  ms ( $SD = 615.12$ ). It differed between languages:  $M(DE) = 896.86$  ( $SD = 731.28$ ) vs.  $M(PL) = 682.28$  ms ( $SD(PL) = 482.76$ ), but not between tasks:  $M(Tower) = 767.5$  ( $SD(Tower) = 544.42$ ),  $M(Gift) = 785.43$  ( $SD(Gift) = 734.65$ ). The durations are more similar (counted as a mean proportion of Speaker 1 and Speaker 2 stroke durations) in the *Tower* condition  $M(Tower) = 1.0009$ , than in *Gift*  $M(Gift) = 0.867$ .

MANOVA on stroke duration (separate data on each stroke annotated in the dialogues, 1194 data points) showed an effect of the *language* factor ( $F = 32.464, p < 0.001$ ) but not *task* ( $F = 2.976, p = 0.084$ ) or interaction between the factors ( $F = 1.604, p = 0.2$ ); see Fig. 6 for details. Removing outliers (defined as strokes longer than 2000 ms, ca. 5% of the data) did not bring significant changes to the results of MANOVA.

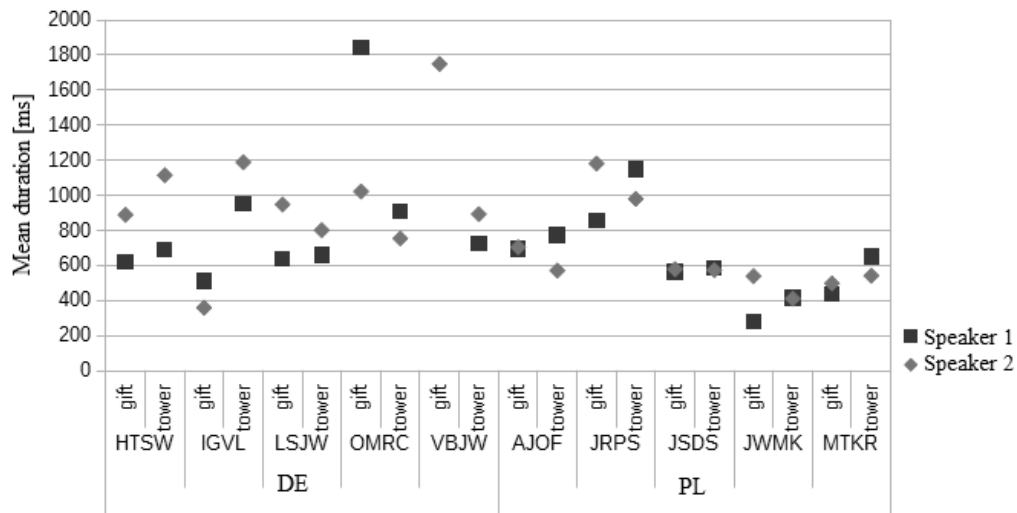


Figure 5: Mean stroke duration in gestures by German and Polish speakers in *Gift* and *Tower* tasks.

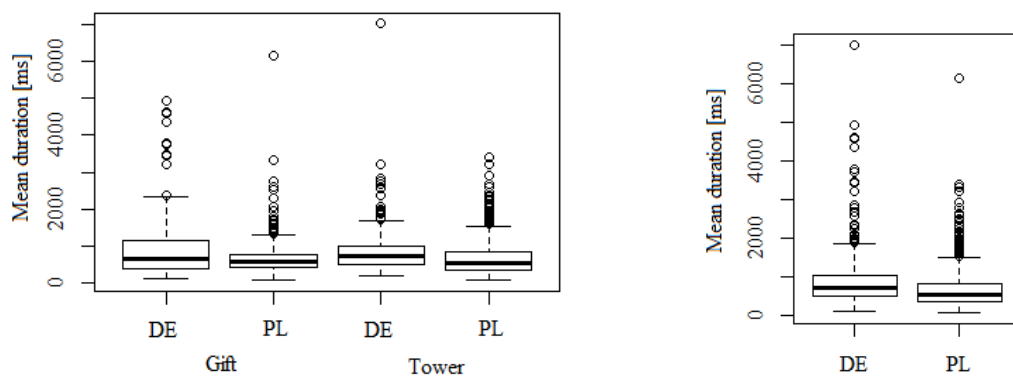


Figure 6: Mean stroke duration for German (DE) and Polish (PL) speakers calculated for *language* and *task* factors (left) and for *language* alone (right).

The numbers of strokes in moving time windows (frame width: 30 sec) were calculated in Annotation Pro, which resulted in 426 data points. The mean number of strokes per window was  $M = 4.23$ , with some difference between the languages  $M(DE) = 3.53$ ,  $M(PL) = 4.93$ , and between the conditions  $M(Tower) = 5.94$  vs.  $M(Gift) = 2.52$ . Values for individual speakers are presented in Fig. 7.

In order to analyse the correlation between the dialogue partners, the results for Speaker 1 and Speaker 2 in corresponding time windows were paired and pairwise correlations were calculated. Moving time windows data were extracted automatically from the dialogues' annotations. Those longer than 3 minutes also included fragments with no gesture annotation. If the number of gestures performed in a given time window equalled zero in both speakers, the data point was removed. Altogether, 41 data points were removed (about 5% of all the data). Consequently, the presented analysis is based on 173 data points (co-occurrence calculation results for 173 time windows), 74 in the *Gift* (competitive) scenario and 99 in the *Tower* (collaborative) scenario, 84 from German participants and 89 from Polish ones.

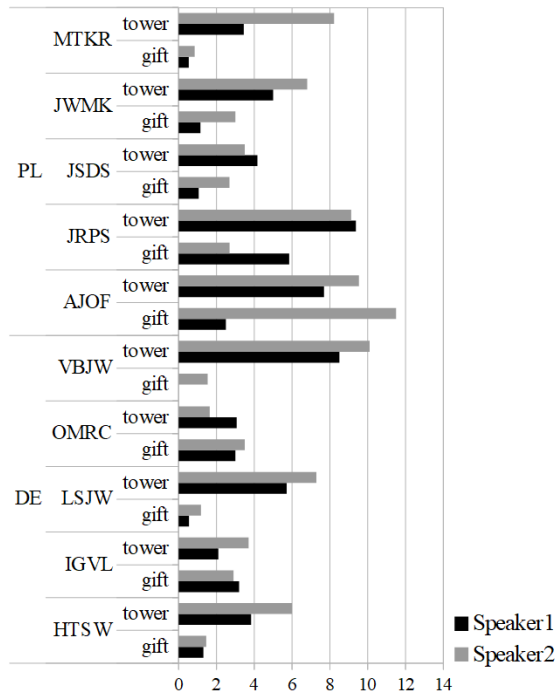


Figure 7: Mean number of strokes per time window (30 sec) for individual Polish (PL) and German (DE) speakers in two types of dialogues (*Tower* and *Gift*).

The correlation between speakers (German and Polish participants together) was expressed by  $R^2 = 0.451$  in the entire data set, with  $R^2(\text{Polish}) = 0.361$  and  $R^2(\text{German}) = 0.51$ . The difference between Polish and German dialogues was statistically insignificant ( $Z = -1.19$ ,  $p(\text{two-tailed}) = 0.23$ ). As far as the type of dialogue scenario is concerned, for the whole data, the correlation in the *Tower* condition  $R^2(\textit{Tower}) = 0.488$  and in the *Gift* condition  $R^2 = -0.014$ . The difference between conditions was statistically significant ( $Z = 3.5$ ,  $p(\text{two-tailed}) = 0.005$ ).

Further analyses were conducted to ascertain whether the type of task (collaborative vs. competitive condition) influenced the alignment of gestural behaviour similarly in both groups, i.e. the Polish and the German speakers. In the Polish dialogues  $R^2(\textit{Gift}) = -0.047$  while  $R^2(\textit{Tower}) = 0.368$ . The difference between correlation coefficients is not statistically significant ( $Z = -1.9$ ,  $p(\text{two-tailed}) = 0.057$ ), but it shows the same tendency as the aforementioned analysis on the whole data. In the German dialogues  $R^2(\textit{Gift}) = -0.196$  while  $R^2(\textit{Tower}) = 0.591$ . The difference between correlation coefficients is statistically significant for the German dialogues ( $R^2Z = -3.64$ ,  $p(\text{two-tailed}) < 0.001$ ).

The proportions of pragmatic and referential gestures are inverse in the two conditions: the collaborative task (*Tower*) evoked more referential gestures than the competitive task (*Gift*). This

tendency is present in both the Polish and the German dialogues (Fig. 8).

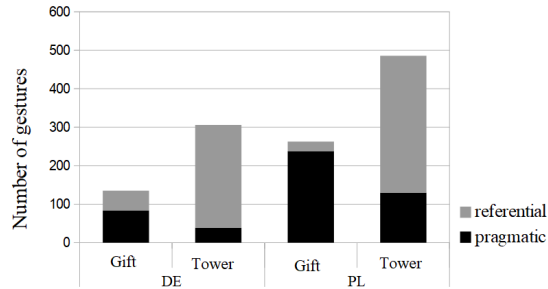


Figure 8: Referential and pragmatic gestures in German (DE) and Polish (PL) speakers in two dialogue tasks (*Gift* and *Tower*).

In the *Gift* condition there were 322 pragmatic and 76 referential gestures, while in the *Tower* condition there were 169 pragmatic and 623 referential ones. The difference measured with a Chi-squared test is significant ( $R^2Chi = 387.789, p < 0.001$ ). The result remains significant when only German dialogues are taken into account ( $Chi(DE) = 114.0172, p < 0.001$ ) as well as for Polish data only ( $Chi(PL) = 361.152, p < 0.001$ ).

Pragmatic and referential gestures seem to differ in duration:  $R^2M(\text{pragmatic}) = 677.52$  ms,  $M(\text{referential}) = 820.9$  ms. MANOVA results showed no effect of task ( $F = 2.096, p = 0.148$ ) but an effect of language ( $F = 37.858, p < 0.001$ ), an effect of gesture function ( $F = 21.648, p < 0.001$ ) and an interaction between the three factors ( $F = 5.843, p = 0.0158$ ).

Further analyses focused on the duration of original and re-occurring strokes of three sizes covered by the annotation scheme (1 — small, 2 — medium/regular, 3 — large). Size and timing may belong to alignment-sensitive features as they simultaneously contribute to and are influenced by the overall rhythm of conversation, and can easily be observed by conversational partners. Annotations were searched for occurrences of strokes in one speaker that have the same size as the stroke that they follow in the other using the Annotation Pro Re-Occurrence plugin described in Section 4. Each stroke was checked for its *size* and *duration* parameter and each of the ten following strokes annotated for the other speaker were checked for equal value of this parameter. They, in turn, were checked for *duration*. Additionally, the distance (lag) between the original segment and each of the re-occurring ones was measured.

In Fig. 9, results are shown separately for the collaborative and competitive task. In both conditions, the mean stroke durations of the repeated strokes follow the pattern of the “original” (reference) ones. One exception is the case of Polish speakers in the collaborative task where large strokes in repetitions are longer than the original strokes. While duration may be, in principle, heavily influenced by gesture size, as larger gestures normally require more time to perform, one may notice that among German speakers in the competitive task (Fig. 9, bottom panel), as opposed to the collaborative task, the larger gestures (Size-2) are actually significantly shorter. In this task, the performance of Polish speakers is clearly different: the larger the strokes, the longer their average durations. The statistical significance of the differences was confirmed by the results of factorial ANOVA ( $F = 57.73, p < 0.0005$  for the interaction of *task* and *gesture size* factors;  $F = 12.51, p < 0.0005$  for the interaction of *language* and *gesture size* factors; and  $F = 52.87, p < 0.0005$  for the interaction of all three factors).

Similar analysis was conducted with the same plugin for repeated gestures that have the same function, pragmatic or referential, as the original ones. As shown in Fig. 10, the original and repeated stroke durations for gestures of the same function are generally similar for the Polish speakers, while for Germans the repeated strokes of the same gesture functions are, on average, shorter than the original ones for each type of gesture and condition (dialogue type). Moreover,

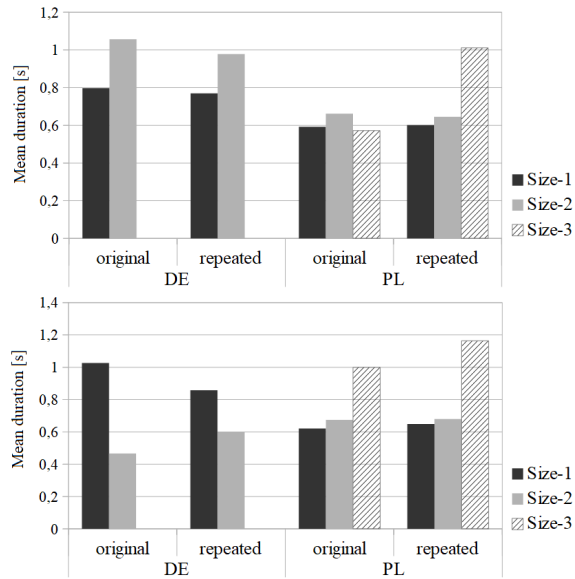


Figure 9: Mean duration of original and repeated strokes in the *Tower* (top) and *Gift* (bottom) condition.

Polish speakers clearly show a higher mean duration of original and repeated strokes than Germans in referential gestures in the competitive dialogue. The differences between means were statistically significant as shown by the results of the factorial ANOVA ( $p < 0.0005$ ).

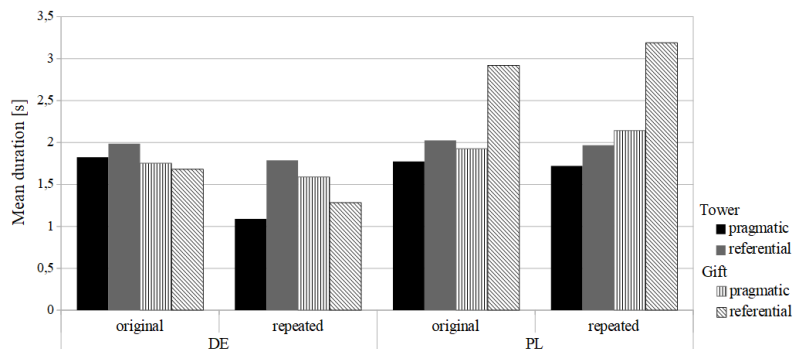


Figure 10: Mean duration of original and repeated gestures depending on gesture function (pragmatic or referential), type of task (*Tower* vs. *Gift*) and the speakers' native language (DE — German, PL — Polish).

## 6 Discussion and conclusions

In the present study, selected measures of gestural behaviour have been analysed in order to find observable correlates of gestural alignment between conversational partners. The analyses are based on a total of twenty task-oriented dialogues of two types (collaborative, referred to as *Tower*, vs. competitive, referred to as *Gift*) in two languages (German and Polish). The dialogues were recorded using a pair of camcorders and annotated for their gestural component in ELAN. A high coherence of annotation was achieved due to the design of the tagset and the process of

annotation (including the training and controlling of annotators). Among other tiers, the time-aligned annotation used for the purposes of the study includes data on 1194 strokes (crucial phases of gestural phrases) and their properties. Selected annotation tiers were imported into Annotation Pro and processed using plugins designed specifically for co-occurrence and re-occurrence data extraction.

The *Tower* scenario seems to evoke more gestures while greater differences in gesture usage intensity between the conversational partners are observed in the *Gift* scenario, although their roles were symmetrical in both scenarios. A significant difference in the mean number of gestures in the annotated parts of dialogues (minutes of interaction from the beginning, the middle and the end of each dialogue) is also found between German and Polish speakers. According to ANOVA results, both the effect of *task* and *language* factors are significant but MANOVA shows no interaction between them. The proportion of referential and pragmatic gestures is reverse in the two conditions, with more pragmatic gestures in *Gift* (competitive) and more referential ones in *Tower* (collaborative). The difference is statistically significant ( $Chi = 387.789, p < 0.001$ ) and the significance has been preserved when Polish and German dialogues were analysed separately.

Differences between the languages have also been calculated. The sum of strokes performed by each speaker differs between the conditions for each language with more gestures in the Polish dialogues and in the *Tower* condition. However, MANOVA shows no effect of interaction of the factors which indicates that the tasks affected the number of gestures similarly in both languages. Stroke duration differs between the languages but not between tasks: The durations are more similar in the *Tower* condition than in *Gift*. MANOVA on stroke duration showed an effect of language but no effect of either task or interaction of the factors, independently of outliers (kept or removed).

In order to explore the co-occurrence of gestural events in the communicative behaviour of dialogue partners, a moving time frame approach based on a 30 second long window was employed. The number of strokes per frame differs between languages but the difference between the conditions is even more striking. Co-occurrence analysis is based on the total of 173 data points. The correlation between speakers in the number of produced gestures significantly differed between the conditions, with a lower  $R^2$  value in the competitive scenario, but not between the languages. In Polish speakers the correlation proved to be insignificant in both the conditions, and the difference between the respective  $R^2$  values was also not significant. In German speakers, the difference between correlation coefficients proved to be statistically significant.

Further exploration of alignment involved the analysis of re-occurrence of stroke parameters. For each stroke made by one speaker, the ten following strokes by the other were analysed for their size and function. In both cases, their durations were also measured. The mean stroke durations of the repeated strokes of three different size categories follow the pattern of the original ones with the exception of those produced by Polish speakers in the collaborative task. Among German speakers in the competitive condition, larger gestures are actually significantly shorter, in contrast to the collaborative condition. In this task, the performance of Polish speakers is clearly different: larger strokes are longer. It may also be noted that the German speakers did not produce any gestures categorised as large at all while, in some conditions, large gestures dominated in the gestural behaviour of Poles, which may be attributed to cultural differences (Müller, 1998). The durations of original and repeated strokes for gestures of the same function are similar for Poles, while for Germans the repeated strokes of the same gesture functions are, on average, shorter than the original ones. Polish speakers clearly show a higher mean duration of original and repeated strokes in referential gestures in the competitive dialogue.

The two tasks seem to evoke significantly different communicational behaviour. In the collaborative task (*Tower*) participants gesture more often and co-ordinate their gesticulation. Global measures for entire dialogues show that the number and duration of gestures are more similar in the collaborative task. Moreover, pairwise correlation between speakers is higher in the *Tower* task as well. Local measures also indicate better coordination between the speakers in this task. These results are even more striking because the *Gift* scenario was always recorded as the second

one, with participants that had already been talking for some time and had become accustomed to each other. The competitive scenario evokes more pragmatic than referential gestures, which may result from compensating the lack of coordination with explicit discourse management. In the collaborative scenario, pragmatic gestures were significantly less frequent, while communication was more fluent in terms of alignment. The differences in gesticulation by Germans and Poles (more gestures in Polish dialogues, longer strokes in German ones, larger gestures performed by Poles) may be due to cultural and linguistic differences. Even though some of them are strongly confirmed by statistics, it is difficult to exclude the impact of some other uncontrolled factors involved in the process of communication. All the aforementioned analyses show, however, that the difference between collaborative and competitive tasks affected both groups in the same direction, resulting in more gestures and the occurrence of gestural coordination in the *Tower* (collaborative) task. This outcome supports the initial hypothesis that in collaborative tasks dialogue participants will be more prone to align with each other.

Further research will include more detailed re-occurrence analyses, *n*-gram based gesture patterning analysis, as well as correlating results with other areas of communicative alignment currently explored in our data, including lexical and prosodic domains. Our results may contribute not only to the knowledge on the mechanisms of alignment, but also to a wider picture of their pragmatic anchoring, sources and consequences (e.g., Beňuš, Gravano, & Hirschberg, 2011; Gravano et al., 2011).

## References

- Beňuš, Š., Gravano, A., & Hirschberg, J. (2011). Pragmatic aspects of temporal accommodation in turn-taking. *Journal of Pragmatics*, *43*(12), 3001–3027. <https://doi.org/10.1016/j.pragma.2011.05.011>
- Bonacchi, S., & Karpiński, M. (2014). Remarks about the use of the term “multimodality”. *Journal of Multimodal Communication Studies*, *1*, 1–7.
- Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(6), 1482–1493. <https://doi.org/10.1037/0278-7393.22.6.1482>
- Campbell, N., & Scherer, S. (2010). Comparing measures of synchrony and alignment in dialogue speech timing with respect to turn-taking activity. In *Proceedings of the 11th Annual Conference of the International Speech Communication Association*. Retrieved from [https://www.isca-speech.org/archive/interspeech\\_2010/index.html](https://www.isca-speech.org/archive/interspeech_2010/index.html)
- 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., Maddux, W. W., & Lakin, J. L. (2006). Beyond the perception-behavior link: The ubiquitous utility and motivational moderators of nonconscious mimicry. In R. R. Hassin, J. S. Uleman, & J. A. Bargh (Eds.), *The new unconscious* (pp. 334–361). Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195307696.003.0014>
- Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 127–149). Washington, DC: American Psychological Association. <https://doi.org/10.1037/10096-006>
- Cleland, A. A., & Pickering, M. J. (2003). The use of lexical and syntactic information in language production: Evidence from the priming of noun-phrase structure. *Journal of Memory and Language*, *49*(2), 214–230.
- Czoska, A., Klessa, K., Karpiński, M., & Nowikow-Jarmołowicz, E. (2015). Prosody and gesture in dialogue: Cross-modal interactions. In *Proceedings of 4th Gesture and Speech in Interaction (GE-SPIN) Conference, Nantes, France* (pp. 83–88). Retrieved from <https://hal.archives-ouvertes.fr/hal-01195646/document>
- Fusaroli, R., Rączaszek-Leonardi, J., & Tylén, K. (2014). Dialog as interpersonal synergy. *New Ideas in Psychology*, *32*(January–April), 147–157. <https://doi.org/10.1016/j.newideapsych.2013.03.005>

- Garrod, S., & Pickering, M. J. (2013). Interactive alignment and prediction in dialogue. In I. Wachsmuth, J. de Ruiter, P. Jaecks, & S. Kopp (Eds.), *Alignment in communication: Towards a new theory of communication* (pp. 193–204). Amsterdam: John Benjamins Publishing Company. (Advances in Interaction Studies, 6). <https://doi.org/10.1075/ais.6.10gar>
- Giles, H., & Smith, P. (1979). Accommodation theory: Optimal levels of convergence. In H. Giles & R. N. St. Clair (Eds.), *Language and social psychology* (pp. 45–65). Baltimore, MD: University Park Press.
- Giles, H., Taylor, D. M., & Bourhis, R. (1973). Towards a theory of interpersonal accommodation through language: Some Canadian data. *Language in Society*, 2(2), 177–192. <https://doi.org/10.1017/S0047404500000701>
- Gravano, A., Levitan, R., Willson, L., Beňuš, Š., Hirschberg, J., & Nenkova, A. (2011). Acoustic and prosodic correlates of social behavior. In *12th Annual Conference of the International Speech Communication Association 2011 (Interspeech 2011)*. Retrieved from <http://www.cs.columbia.edu/nlp/papers/2011/p95485.pdf>
- Guitar, B., & Marchinkoski, L. (2001). Influence of mothers' slower speech on their children's speech rate. *Journal of Speech, Language, and Hearing Research*, 44(4), 853–861. [https://doi.org/10.1044/1092-4388\(2001/067\)](https://doi.org/10.1044/1092-4388(2001/067))
- Healey, P. G. T., Purver, M., & Howes, C. (2014). Divergence in dialogue. *PLoS ONE*, 9(6). <https://doi.org/10.1371/journal.pone.0098598>
- Holle, H., & Rein, R. (2015). EasyDIAG: A tool for easy determination of interrater agreement. *Behavior Research Methods*, 47(3), 837–847. <https://doi.org/10.3758/s13428-014-0506-7>
- Jaecks, P., Damm, O., Hielscher-Fastabend, M., Malchus, K., Stenneken, P., & Wrede, B. (2013). What is the link between emotional and communicative alignment in interaction?: Towards a new theory of communication. In I. Wachsmuth, J. de Ruiter, P. Jaecks, & S. Kopp (Eds.), *Alignment in communication: Towards a new theory of communication* (pp. 205–224). Amsterdam: John Benjamins Publishing Company. (Advances in Interaction Studies, 6). <https://doi.org/10.1075/ais.6.11jae>
- Karpiński, M. (2014). New challenges in psycholinguistics: Interactivity and alignment in interpersonal communication. *Lingua Posnaniensis*, 54(1), 97–106.
- Karpiński, M., Jarmolowicz-Nowikow, E., & Czoska, A. (2015). Gesture annotation scheme development and application for entrainment analysis in task-oriented dialogues in diverse cultures. In *Proceedings of 4th Gesture and Speech in Interaction (GESPIN) Conference, Nantes, France* (pp. 161–166). Retrieved from [https://www.academia.edu/15517265/Karpi%C5%84ski\\_M.\\_Jarmo%C5%82owicz-Nowikow\\_E.\\_Czoska\\_A.\\_2015.\\_Gesture\\_annotation\\_scheme\\_development\\_and\\_application\\_for\\_entrainment\\_analysis\\_in\\_task-oriented\\_dialogues\\_in\\_diverse\\_cultures.\\_Proceedings\\_of\\_GESPIN\\_2015\\_Conference\\_Nantes\\_France\\_161-166](https://www.academia.edu/15517265/Karpi%C5%84ski_M._Jarmo%C5%82owicz-Nowikow_E._Czoska_A._2015._Gesture_annotation_scheme_development_and_application_for_entrainment_analysis_in_task-oriented_dialogues_in_diverse_cultures._Proceedings_of_GESPIN_2015_Conference_Nantes_France_161-166)
- Karpiński, M., Klessa, K., & Czoska, A. (2014). Local and global convergence in the temporal domain in Polish task-oriented dialogue. In N. Campbell, D. Gibbon, & D. Hirst (Eds.), *Proceedings of the 7th International Conference on Speech Prosody* (pp. 743–747). Retrieved from [https://www.academia.edu/7151184/Karpinski\\_M.\\_Klessa\\_K.\\_Czoska\\_A.\\_2014.\\_Local\\_and\\_global\\_convergence\\_in\\_the\\_temporal\\_domain\\_in\\_Polish\\_task\\_oriented\\_dialogue.\\_In\\_N.\\_Campbell\\_D.\\_Gibbon\\_and\\_D.\\_Hirst\\_Eds.\\_Speech\\_Prosody\\_7\\_2014\\_Dublin\\_pp.\\_743-747](https://www.academia.edu/7151184/Karpinski_M._Klessa_K._Czoska_A._2014._Local_and_global_convergence_in_the_temporal_domain_in_Polish_task_oriented_dialogue._In_N._Campbell_D._Gibbon_and_D._Hirst_Eds._Speech_Prosody_7_2014_Dublin_pp._743-747)
- Karpiński, M., & Klessa, K. (2018). Methods, tools and techniques for multimodal analysis of accommodation in intercultural communication. *Computational Methods in Science and Technology*, 24(1), 29–41. <https://doi.org/10.12921/cmst.2018.0000006>
- Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CB09780511807572>
- Klessa, K., Karpiński, M., & Wagner, A. (2013). Annotation Pro – a new software tool for annotation of linguistic and paralinguistic features. In B. Bigi & D. Hirst (Eds.), *Proceedings of TRASP (Tools and Resources for the Analysis of Speech Prosody)* (pp. 51–54). Aix-en-Provence: Aix-Marseille Université.
- Kousidis, S. (2010). *A study of accommodation of prosodic and temporal features in spoken dialogues in view of speech technology applications* (Doctoral thesis). Dublin Institute of Technology.
- Kousidis, S., Dorran, D., Wang, Y., Vaughan, B., Cullen, C., Campbell, D., McDonnell, C., & Coyle, E. (2008). Towards measuring continuous acoustic feature convergence in unconstrained spoken dialogues. In *Proceedings of Interspeech 2008* (pp. 1692–1695).
- Kulesza, W. (2016). (Nie)świadomy kameleon: Analiza związku między stosowaniem niewerbalnej mimikry, uległością wobec tego procesu a (nie)świadomością. *Psychologia Społeczna*, 11(2(37)), 183–195. <https://doi.org/10.7366/1896180020163705>

- Lausberg, H. (2013). NEUROGES – A coding system for the empirical analysis of hand movement behaviour as a reflection of cognitive, emotional, and interactive processes. In C. Müller, A. Cienki, E. Fricke, S. Ladewig, D. McNeill, & S. Tessendorf (Eds.), *Body-language-communication: An international handbook on multimodality in human interaction* (Vol. 1, pp. 1022–1036). Berlin: De Gruyter Mouton. <https://doi.org/10.1515/9783110261318.1022>
- McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. Chicago, IL: University of Chicago Press.
- Mol, L., Krahmer, E., Maes, A., & Swerts, M. (2012). Adaptation in gesture: Converging hands or converging minds? *Journal of Memory and Language*, 66(1), 249–264. <https://doi.org/10.1016/j.jml.2011.07.004>
- Müller, C. (1998). *Redebegleitende Gesten: Kulturgeschichte, Theorie, Sprachvergleich*. Berlin: Berlin-Verl. Spitz. (Körper, Zeichen, Kultur, 1).
- Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America*, 119(4), 2382–2393. <https://doi.org/10.1121/1.2178720>
- Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *The Behavioral and Brain Sciences*, 27(2), 169–190. <https://doi.org/10.1017/S0140525X04000056>
- Ramseyer, F., & Tschacher, W. (2011). Nonverbal synchrony in psychotherapy: Coordinated body movement reflects relationship quality and outcome. *Journal of Consulting and Clinical Psychology*, 79(3), 284–295. <https://doi.org/10.1037/a0023419>
- Richardson, D. C., Dale, R., & Tomlinson, J. M. (2009). Conversation, gaze coordination, and beliefs about visual context. *Cognitive Science*, 33(8), 1468–1482. <https://doi.org/10.1111/j.1551-6709.2009.01057.x>
- Sloetjes, H., & Wittenburg, P. (2008). Annotation by category: ELAN and ISO DCR. In *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008)*. Retrieved from [http://pubman.mpg.de/pubman/item/escidoc:60774:3/component/escidoc:60775/Sloetjes\\_2008\\_annotation.pdf](http://pubman.mpg.de/pubman/item/escidoc:60774:3/component/escidoc:60775/Sloetjes_2008_annotation.pdf)
- Truong, K. P., & Heylen, D. (2012). Measuring prosodic alignment in cooperative task-based conversations. In INTERSPEECH-2012 (pp. 843–846), Portland, OR. Retrieved January 31, 2018, from [https://www.isca-speech.org/archive/archive\\_papers/interspeech\\_2012/i12\\_0843.pdf](https://www.isca-speech.org/archive/archive_papers/interspeech_2012/i12_0843.pdf)
- Ward, A., & Litman, D. J. (2007). Automatically measuring lexical and acoustic/prosodic convergence in tutorial dialog corpora. In *Workshop on Speech and Language Technology in Education*. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.137.6046&rep=rep1&type=pdf>
- Ward, A., & Litman, D. J. (2007). *Measuring convergence and priming in tutorial dialog*. University of Pittsburgh.
- Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., & Sloetjes, H. (2006). ELAN: A professional framework for multimodality research. In *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC 2006)* (pp. 1556–1559). Retrieved from <http://www.lrec-conf.org/proceedings/lrec2006>

---

This work was funded by the National Programme for Progress in the Humanities (NPRH12H 13 0524 82, Language of the boundaries — the boundaries of language. Paralinguistic aspects of intercultural communication).



NARODOWY PROGRAM  
ROZWOJU HUMANISTYKI

The authors declare that they have no competing interests.

The authors' contribution was as follows: concept of the study: MK; background overview: KJ; corpus design and task design: MK, EJM, KK; corpus annotation: EJM; data management and processing: KK; statistical analyses and graphics: ACz, KK, KJ; writing: MK, ACz, EJM, KJ, KK.

This is an Open Access article distributed under the terms of the Creative Commons Attribution 3.0 PL License (<http://creativecommons.org/licenses/by/3.0/pl/>), which permits redistribution, commercial and non-commercial, provided that the article is properly cited.



© The Authors 2018

Publisher: Institute of Slavic Studies, Polish Academy of Sciences, University of Silesia & The Slavic Foundation