Journal of Multimodal Communication Studies

vol. 4, issue 1-2

Special issue: Gesture and Speech in Interaction

Editors:

Silvia Bonacchi & Maciej Karpiński

Guest Editors:

Anna Jelec
Małgorzata Fabiszak
Konrad Juszczyk

Poznań 2017

ISSN 2391-4033
DOI 10.14746/jmcs

Journal of Multimodal Communication Studies

vol. 4, issue 1-2 (2017)

Special issue: Gesture and Speech in Interaction

Table of contents:

Introduction

Anna Jelec, Małgorzata Fabiszak
Faculty of English, Adam Mickiewicz University in Poznań, Poland
jelec@amu.edu.pl; fagosia@wa.amu.edu.pl

The fourteen articles that appear in this volume investigate gesture and speech in interpersonal interaction. The idea to gather different approaches and perspectives on the study of gesture is not a new one, although it is increasingly necessary. The complexity of the field grows along with its popularity, and keeping up with new developments in fields that often range from discourse analysis to human-computer interaction has become quite a challenge. As co-organisers of the 5th International GESPIN Conference, launched in August 2017 in Poznań by the Faculty of Modern Languages and Literatures together with the Faculty of English, Adam Mickiewicz University we hoped to give researchers from a variety of gesture-related fields a platform to share their unique expertise. Indeed, the contributions to this volume discuss topics from an impressively broad spectrum.

Two articles deal with the issue of negation. Ozge Alaçam and Christopher Habel discuss gestures that accompany negation in descriptions of conceptual events, which are depicted through visual and haptic graphs, while Sonia Gembalczyk and Jolanta Antas focus on the bodily expressions of negation of Polish speakers.

Two further contributions demonstrate the potential applications of technology in gesture studies, with M. Teresa Anguera and colleagues' article on transcribing data from human communication processes into code matrices and Duc-Canh Nguyen and colleagues' paper on a new framework for the evaluation of the performance of a socially assistive robot interacting with the elderly patients.

Two papers focus on gesture in the context of discourse: our own report on the individual variation in gestures of uncertainty in discourses of collective memory and Orit Sônia Waisman's paper describing mismatches between verbal and nonverbal signs, which she argues highlight salient situations in communication.

Two contributions investigate gesture in applied contexts, such as foreign language learning and psychotherapy. Hiroki Hanamoto argues that gesture is an important indicator of mutual intelligibility across cultures and contexts in a study of English as a lingua franca (ELF). Niklas Neuman and colleagues' paper analyses the differences in non-verbal communication of patients with social phobia and their therapists, focusing on hand rest positions in dialogue.

Two more articles focus on syntax. Manon Lelandais and Gaëlle Ferré investigate how gestures can shed light on the status of subordinate constructions, highlighting their dependence or autonomy. Suwei Wu and Alan Cienki's paper asks to what extent gesture interacts with syntactic encoding in double object constructions and their prepositional paraphrase.

In a similar vein, Petra Wagner and Nataliya Bryhadyr's article discusses speech-gesture synchronization, asking whether it is affected by the interlocutors' mutual visibility and linguistic information structure.

Two contributions discuss gesture from a more phonological perspective. The first study by Marzena Żygis, Susanne Fuchs and Katarzyna Stoltmann investigates orofacial expressions (lip aperture and the eyebrow movement) in the production of yes/no questions and statements, while the second study by Stoltmann and Fuchs investigates the stability of the relationship between number of syllables and pointing gestures under different speech rate constraints.

Finally, Marieke Hoetjes and Ingrid Masson-Carro answer the question whether an increase in verbal cognitive load makes people produce more gestures and whether motoric cognitive load decreases the number of gestures. Unexpectedly the answer they give is not so straightforward.

Negation in gesture accompanying speech: A cross-modal study

Özge Alaçam, Christopher Habel

University of Hamburg, Department of Informatics, Hamburg, Germany

alacam@informatik.uni-hamburg.de, habel@informatik.uni-hamburg.de

Abstract

Negation is a higher order abstract concept, which may shed light into the investigation of gesture-language and referent relation. This paper focuses on the gestures that accompany to negated content during the description of a conceptual event depicted in visually or haptically explored graphs. The results of the empirical study showed that while the gestures accompanying to negated content for the visual graphs were congruent with the referent's properties, the gestures for the haptic graphs were congruent with the negated modifiers.

1 Gesture - language relation in comprehending graphs

Statistical graphs are widely used elements of multimodal—text-graphics—communicational settings due to their facilitating influence on acquiring knowledge about states, changes and processes in the world. Acquiring such knowledge is crucial also for visually impaired people, and haptic presentations of graphs provide a suitable means for this purpose (Alaçam, 2015; Alaçam et al., 2013). The material used in this study was designed to investigate haptic graph comprehension as a part of research focusing on providing visually impaired users with verbal assistance as an accompanying modality to their haptic exploration. Providing informationally or functionally isomorphic information regarding the two different modalities for a systematic analysis can be tricky issue; however statistical graphs provide very structured information in language like manner and they are formalized visualizations by language-like conventions since they have syntactic, semantic and pragmatic levels like language (Kosslyn, 1989).

Graph perception and comprehension processes are dependent on *sensory modalities* (such as visual, haptic and verbal), while communicating over graph involves *communicational modalities* (such as language, and gesture). Gesture and language are widely accepted as two complementary components of a single integrated system (McNeill, 1992), each modality has its own superiority in terms of conveying different aspects of the referent (Hostetter & Alibali, 2008; Goldin-Meadow, 2000). Gestures and graphs, on the other hand, are visuo-spatial modalities sharing similar perceptual visuo-spatial features that convey meanings such as quantity, direction and relations (Tversky, 2011). For instance, as analog representation, gestures can convey shape properties in a richer way compared to language modality, especially in the cases the shape of the referent is complex and hard to verbalize. For example, Melinger & Kita's study (2007, pp.475) showed that "the rate of lexical gestures was higher when the line drawings to be described were less describable and less verbally codable". In short, the sensory modality, language and gestures exhibit intertwined interactions among each other. Originating from the resemblance among modalities, the vocabularies of gestures, speech and diagrams can be considered as parallel (Tversky, 2011). For instance, in the context of communication over graphs, a fluctuating increase in a line graph may be verbally described by the term "increase" and it may be simultaneously accompanied by a gesture that represents the fluctuation in that increase.

Although the language-gesture interaction has been investigated for the past several decades, to our knowledge, so far little attention has been given to the effect of the sensory modality on gesture production (Alaçam, 2015; Alaçam et al., 2013). The results of those showed that haptic

exploration had an influence on the production of gestures, possibly due to the alignment between shared spatial properties of gestures and haptic exploration. Haptic exploration (as a sensory modality) and gesture production (as a communicational modality) share common underlying mechanisms of motor movements (activation in the premotor and motor cortex). According to the gesture-as-simulated-action framework (GSA), mental imagery is claimed to be an embodied process that relies on simulation of perceptions and actions. Moreover, both language production and mental imagery employ the sensorimotor systems, and gestures are produced due to the activation in those sensorimotor areas. As stated by Alibali (2005), spatial and motor information can be successfully carried over gestures, since they are relying on the same analog underlying mechanisms unlike discrete and propositional verbal codes. Both motor and visual images are offline representations however while the former involve simulated action, the latter involve simulated perception (Hostetter & Alibali, 2008). Therefore, speech-accompanying gestures during post-exploration verbal descriptions of graphs may exhibit differences in regards to sensory modality of the exploration. More specifically, gestures produced for haptically perceived graphs are affected by motor-image based simulated actions; whereas gestures produced for visually perceived graphs are influenced by visual-image based simulated perception. Based on that prior research, we explore the possibility that the effect of haptic perception on the relations between the communicational modalities, namely gesture and language might be higher than the effect of visual perception due to high common neural activity in pre-motor and motor cortex regarding haptic perception and gesture production.

2 Negation

Gestures do not only co-occur with affirmative speech, they might also accompany to negated content during speech. Theories about negation comprehension mainly exhibit differences with respect to whether the negated and/or actual situation of content is simulated during negative sentence comprehension. In brief, according to the propositional view, only negated situation is represented in a comprehender's mental models and this representation is symbolic (Carpenter & Just, 1975; Clark & Chase, 1972). On the other hand, growing number of studies provided evidence in the favor of simulation accounts (i.e. Orenes et al., 2014; Kaup et al., 2007; 2006). From the embodied simulation perspective, comprehenders use the alternative to understand negation, and the mental representation of the situation is iconic. According to the two-step simulation theory by Orenes et al. (2014), negation comprehension requires the use of mental representations, which are completely grounded in sensorimotor experience. For example, negation processing of a sentence such as "The door is not open" begins with the simulation of the negated argument (an open door as an expected situation) and continues with the simulation of the alternative (a closed door as an actual situation). Kaup and colleagues also proposed an alternative simulation theory (2007; 2006). Their study showed that negated situation was always simulated, and this simulation is mentally rejected by the fact that it is simulated but not integrated with the representation of the actual situation. In a study by Foroni and Simon (2013) on whether negation is represented as a motor process, participants were asked to read the sentences that describe emotional expressions while they are measuring the activation upon the zygomatic muscle associated with smiling. Their result indicated that reading negative sentences lead to the inhibition of this smile-associated muscle. On the other hand, reading affirmative sentences leads to the activation of this muscle. The authors accept this as an evidence that an abstract negation operation depends on motor processes as well, generalizing the simulation argument underlying the action-related language processing view to negation context.

The topic of negation has been widely investigated from a comprehension perspective; however, to the authors' knowledge under what circumstances do people use negative sentences have been scarcely investigated from the production side. Negative sentences do not point to the actual situations but to what is not the case or that what deviates from the expected state. To exemplify, if someone utters that "the population is not increasing", there should have been a reason to emphasize it. Therefore, contextual factors play an important role in understanding negation

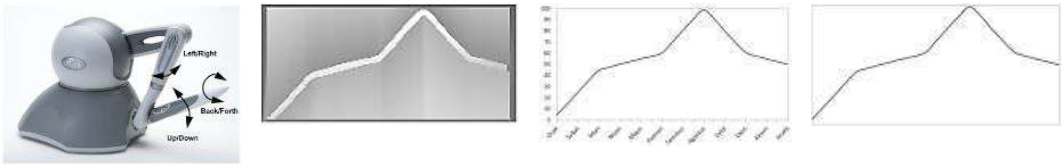
production. Lüdtke and Kaup (2006) have studied the role of prior explicit mentions of the negated property in a context. Their results indicated that equivalent reading times of affirmatives and negatives were only observed when the negated property was previously explicitly mentioned. In another study on negation production (Watson, 1979), children were asked to describe an object with respect to another similar object. The referents either lacked a specific property or possessed a distinct property. The results revealed that negative utterances were used for referents, which lacked a property, while the children produced affirmative descriptions for the referents which possessed a distinct attribute. He concluded that if the actual situation possesses the low informative quality, then negation is mandatory. Another study by Beltran et al. (2008) also showed that negations are more likely produced when the actual situation is not easily accessible in an affirmative way.

The discourse context sets up a field of oppositions, and speakers are especially likely to gesture about information that contrasts with information already present in the field (McNeill, 2008). Contrasting properties have high saliency, and this locates them to the focus of attention. As stated by Hostetter and Alibali (2008, p.506), "when speakers contrast two events, they are more likely to strongly simulate the contrasting elements; thus they are more likely to express these contrasting elements in gesture". In order to activate a mental model for a negated concept, first the mental model of the concept itself needs to be activated as advocated in the simulation theories above-mentioned. The cognitive state that underlies mismatch involves having and activating two ideas on one task (Church & Goldin-Meadow, 1986). People may produce gesture that conveys different information ("mismatches") from the information they convey in speech. To illustrate, when referring to a steep segment, one can say that "it is steep" or "it is not slight". Steepness is one of the amodal geometric properties that the graphs possess and is a graded property that can take four values in the current setting; no change, slight, moderate or steep. In other words, it is not like a binary attribute, the alternatives lies on the same continuum. Therefore, being slight and steep can be considered as two extremities of one concept ("steepness") or they can be considered as two different concepts. Although it is worth mentioning, this distinction is not relevant for the current purposes. One thing that we can say for sure is that there is a contrasting property and the gesture-language relation differs for those contrasting cases with respect to sensory modality of the graph exploration as indicated by the results below. As a higher order abstract concept, negated speech and its reflection on co-verbal gestures for the events perceived through different modalities provide intriguing cases for a theoretical discussion towards understanding the language-gesture and referent relation.

3 Experiment

3.1 Participants, material and design

48 university students (Mage=24.69, SD=5.85) participated to this experiment. All participants were native speakers of Turkish. Haptic explorers were blind-folded sighted people. In this study, Phantom Omni® Haptic Device (see Figure 1) is used to represent the haptic line graphs. Haptic graph exploration is performed by moving the handle of the device. The graph line is represented by engraved concavities on a horizontal plane; therefore, the graph readers perceive the line as deeper with respect to other area on the surface and trace the line haptically by moving the pen. The experiment was conducted in three conditions in a between-subject design. In the "*H-noLabel*" condition, the participants explored line graphs haptically. In the second condition, the graphs with data labels were presented on a computer screen, thus the participants had visual access to the graphs ("*V-withLabel*"). In the third condition (a control condition), the participants inspected the visual graphs without data labels ("*V-noLabel*"), see Figure 1.



Haptic Device H-noLabel V-withLabel V-noLabel

Figure 1. Phantom Omni® Haptic Device and sample graphs for each condition.

It was shown by Loomis et al. (2007) that the recognition time for haptic objects is longer than that for visual objects. Therefore, the participants in the haptic condition did not have time limitation, while visual graph readers had 10 sec for visual inspection. In the experiment session, each participant was presented 12 graphs that present averaged monthly tourist visits for various cities in a random order. The steepness of the lines was restricted to four values (0°, 15°, 45°, 75°). In all conditions, after the graph exploration process, the participants were asked to stand up and present a single-sentence summary of the graph to a hypothetical audience and then produced sketch of the graph. The sessions were audio and video recorded.

All post-exploration verbal descriptions were split into phrases following Kita & Özyürek's *Interface Hypothesis* (2003). A phrase was defined as any unit containing a predicate (i.e. a verb) that expresses a single sub-event or state. Gestural forms associated with negation and their interaction with language has been investigated by several studies (i.e. Kendon, 2002). However, those studies address non-iconic gestures that have similar function with non-speech sounds like "uhh uhh". Here we focus on iconic/representational gestures, which represent shape of an object or an action. Gestures were annotated by using the video annotation software ANVIL and categorized semantically into two; Type-I gestures are congruent with the property of the referred graph segment and Type-II gestures exhibit congruency with what modifier used in the verbal statement represents. To illustrate, let us take a verbal description such as "it is not sharp". A slightly diagonal gesture that resembles to what the graphical entity is (i.e. a slight increase) would be classified as Type-I, whereas a vertical or near vertical gesture that resembles to what the modifier represents (i.e. a sharp increase) would be classified as a Type-II gesture. Two coders classified the data. The substantial interrater agreement (*Kappa*=.72) was calculated by Cohen's kappa.

4 Results

Since the participants were not forced or primed about neither the use of any linguistic structure nor the use of speech-accompanying gestures, the instances that exemplify this topic presented above were spontaneous and rare. Therefore, the quantitative methods involving statistical tests were not applied; instead, the qualitative evaluation enriched with frequencies and examples addressing the effect of sensory modality on gesture production regarding negated content and the effect of the order of negative and affirmative statements on gestural form were reported.

For this analysis, we focused on the speech parts that involve negative statements (st's), and also affirmative st`s that are followed or preceded by the negated st's that refer to the same segment; such as "it is not straight, it is curved" or in the other way around. First, it is possible to produce a sentence that involves a negative st. without producing a gesture. Second, a gesture may accompany to an affirmative st., but not to a negative st. Third, a gesture may accompany only to a negative st. The fourth category involves the cases that the participants produced gestures for both affirmative and negative st's. Table 1 presents the number of gestures for each type of st's. For the affirmative st's, mostly the congruent gestures (Type-I) were accompanied to the speech parts; i.e. the verbal description "it is sharp" accompanied by vertical or near vertical dynamic gesture. Type-II gestures display oppositeness; therefore, they mostly accompany negative st's. 5 of 16 participants in the V-withLabels produced sentences that involve negated st's; this ratio was higher in the V-noLabels (11 of 16) and also in the haptic condition (12 of 16). The overall number of the sentences that involve negated st's also showed similar pattern; 14 st's were produced in the *V-withLabels*, on the other hand, 23 st's in the *V-noLabels* and 29 st's in the haptic condition were

produced. In the *V-withLabels* condition, the participants preferred not to produce gestures for the sentences that involve negative st's or they produced gestures for just the affirmative parts. On the other hand, the gesture production for the negative st's was nearly similar for the *V-noLabels* and for the haptic condition. However, a striking difference in terms of the types of the gestures was observed between these two groups. While the participants in the *V-noLabels* condition tended to produce Type-I gestures (8 of 9) for the negative ST's (including the category "both"), the participants in the haptic condition tended to produce Type-II gesture (8 of 10).

Beside the effect of the modality, the order of the negative and affirmative st's also seemed to have an effect on the produced gesture. The effect of order is more apparent in the haptic condition, due to possible effect of previous active exploration actions (for mostly salient parts i.e. a previous steep segment). If the affirmative st. takes the first place and is accompanied by a gesture (by nature, they are Type-I) then again Type-I gesture was observed for the negative st's (Excerpt1). However, if the affirmative st. in the first place was not accompanied by a gesture or the first part did not have any affirmative modifier (Excerpt2) then the negative st's was followed by Type-II gestures (8 of 10 cases). In Excerpt3 first, false affirmative sentence was produced with Type-I gesture (the hand showed a steep increase but the subject described it as a "30° angle", the referred segment is a steep segment), then he updated and corrected the verbal description and continued with a previous (Type-I) gesture. The pattern indicates that Type-II gestures were produced for negated content if the actual situation was not described before (with an accompanying gesture).

1. [In September, it starts to decrease], [that wiggling disappears (previously mentioned)]
   [Affirmative st.]$_{\textbf{Type-I}}$ >> [Negative st.]$_{\textbf{Type-I}}$
2. [It is never increasing straight], [it always increases in a sloping way]
   [Negative st.]$_{\textbf{Type-II}}$ >> [Affirmative st.]$_{\textbf{Type-I}}$
3. [In a slight way, with approx. 30° angle it is going up.] [Actually, it is not 30° angle], [it is like 70° angle, it goes toward up.]
   [False Affirmative st.]$_{\textbf{Type-I}}$ >> [Negative st.]$_{\textbf{Type-I}}$ >> [Corrected affirmative st.]$_{\textbf{Type-I}}$

Table 1. The number of Type-I and Type-II gestures for each condition (st.:statement).

	Gesture types	V-with labels	V- noLabels	H -nolabels
Negative st.	No gesture	6	10	16
Affirmative st.	Type-I	8	4	3
Negative st.	*Type-I*	-	3	1
	Type-II	-	-	4
Both affirmative and negative st.	*Type-I*	-	5	1
	Type-II	-	1	4
TOTAL		14	23	29

5 Discussion and conclusion

The relation between the negated content and the produced gesture can uniquely provide valuable information to understand the effect of speech, effect of the sensory modality and their relation. Qualitative analysis indicates that the sensory modality interferes with language and graph comprehension. In brief, while gestures congruent with referent were observed in the description of visual graphs, for haptic graphs, gestures congruent with negated modifier were observed. The reason of preferring a negated modifier seems to be due to possible recent exposure to the contrasting property in the earlier stages of the exploration, and if the contrast is salient enough (i.e. steep vs. slight), it may affect the conceptualization of the current segment. Considering the high ratio of Type- II gesture production accompanied to negative content for haptically perceived graphs, this finding, while preliminary, suggests that the salient features, which were actively explored previously, might have more influence on gesture production. Besides, those active explorations might reflect themselves in the negative st's.

Additionally, gesture type accompanied to the negated context also seems to be affected by the order of the statement. If the first sentence is affirmative, then the gesture of the following negated sentence was also congruent with the referent. However, in the opposite order, if the negated sentence was uttered first, a congruency with the negated modifier (the contrasting element with previously explored region) instead of a congruency with referent's property was observed. Thus, this effect may not be simply due to effect of language on gesture production. Type-II errors might be originated from the active mental representation of the previous salient segment due to recent exploration.

Due to common underlying mechanism of haptic perception and gesture production (involvement of motor actions and motor imagery), it seems that the gestures might have been more influenced by haptic perception compared to visual perception of graphs. The production of negation is associated with having no or limited access to affirmative situation, the results showed that although the actual information is accessible (even explored), comprehenders still produce negated statement for them. This indicates that the saliency (perceptual or conceptual) and prior exposure (already simulated salient feature) are important factors. The attributes of the concept, which is being negated, seem to be important as well.

From the gesture-language research perspective, the results suggested that both language and spatial representations play role in gesture production. While Interface Model focusses on the contribution of language, GSA framework highlights the contribution of sensory modality. According to another cognitive modelling approach by Bergmann et al. (2013), the dynamic activation of visuospatial and propositional representations gives rise to speech accompanying gestures. Furthermore, brain regions that are attributed to the processing of visuospatial representation seem to be also responsible for constructing and processing of spatial models perceived through other sensory modalities (e.g. Knauff, 2013; Zhang et al., 2004). It should be noted that the aforementioned frameworks predominantly address the relation between the visual or verbal modalities and the gestures. The empirical results reported here are in line with those frameworks. Further, it may be considered as an extension that provides examples for integrating spatial representations, which is constructed through haptic sensory modality. Besides, the results favour also the simulation theories of negation. In order to generalize the results, this phenomenon needs to be investigated in a diverse stimuli set from concrete to abstract events.

Acknowledgments

This research was funded by the German Research Foundation (DFG) in project "Crossmodal Learning", TRR-169.

References

Alaçam, Ö. (2015). Verbally Assisted Haptic-Graph Comprehension: Multi-Modal Empirical Research Towards a Human Computer Interface. Ph.D. dissertation, University of Hamburg. Accessible at: http://ediss.sub.uni-hamburg.de/volltexte/2016/7764/

Alaçam, Ö., Habel, C., & Acartürk, C. (2013). Investigation of haptic line-graph comprehension through co-production of gesture and language. Tilburg Gesture Research Meeting, Tilburg.

Alibali, M. W. (2005). Gesture in spatial cognition: Expressing, communicating, and thinking about spatial information. Spatial Cognition and Computation, 5, 307-331.

Beltrán, D., Orenes, I., & Santamaria, C. (2008). Context effects on the spontaneous production of negation. Intercultural Pragmatics, 5, 409-419.

Bergmann, K., Kahl, S., & Kopp, S. (2013). Modeling the semantic coordination of speech and gesture under cognitive and linguistic constraints. International Workshop on Intelligent Virtual Agents, (pp. 203-216).

Carpenter, P. A., & Just, M. A. (1975). Sentence comprehension: A psycholinguistic processing model of verification. Psychological review, 82, 45.

Church, R. B., & Goldin-Meadow, S. (1986). The mismatch between gesture and speech as an index of transitional knowledge. Cognition, 23, 43-71.

Clark, H. H., & Chase, W. G. (1972). On the process of comparing sentences against pictures. *Cognitive Psychol*, 3.

Foroni, F., & Semin, G. R. (2013). Comprehension of action negation involves inhibitory simulation. *Front Hum Neurosci*, 7, 209

Goldin-Meadow, S. (2000). Beyond words: The importance of gesture to researchers and learners. *Child Dev*, 71.

Hostetter, A. B., & Alibali, M. W. (2008). Visible embodiment: Gestures as simulated action. Psychon Bull Rev, 15.
Kaup, B., Lüdtke, J., & Zwaan, R. A. (2006). Processing negated sentences with contradictory predicates: Is a door that is not open mentally closed? Journal of Pragmatics, 38, 1033-1050.
Kaup, B., Zwaan, R. A., & Lüdtke, J. (2007). he experiential view of language comprehension: How is negated text information represented? In F. Schmalhofer & C. A. Perfetti (Eds.), Higher level language processes in the brain: Inference and comprehension processes, 255-288.
Kendon, A. (2002). Some uses of the head shake. Gesture, 2, 147-182.
Kita, S., & Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal?: Evidence for an interface representation of spatial thinking and speaking. Journal of Memory and language.
Knauff, M. (2013). Space to reason: A spatial theory of human thought. MIT Press.
Kosslyn, S. M. (1989). Understanding charts and graphs. Applied cognitive psychology, 3, 185-225.
Loomis, J. M., Klatzky, R. L., Rieser, J. J., Ashmead, D. H., Ebner, F. F., & Corn, A. L. (2007). Functional equivalence of spatial representations from vision, touch, and hearing: Relevance for sensory substitution. Blindness and brain plasticity in navigation and object perception, 155-184.
Lüdtke, J., & Kaup, B. (2006). Context effects when reading negative and affirmative sentences. Proceedings of the 28th annual conference of the cognitive science society, (pp. 1735-1740).
McNeill, D. (2008). Gesture and thought. University of Chicago press.
Melinger, A., & Kita, S. (2007). Conceptualisation load triggers gesture production. *Lang Cognitive Proc*, 22, 473-500.
Orenes, I., Beltrán, D., & Santamaría, C. (2014). How negation is understood: Evidence from the visual world paradigm. Journal of memory and language, 74, 36-45.
Tversky, B. (2011). Visualizing thought. Topics in Cognitive Science, 3, 499-535.
Watson, J. M. (1979). Referential description by children in negative form. British Journal of Psychology, 70, 199-204.
Zhang, M., Weisser, V. D., Stilla, R., Prather, S. C., & Sathian, K. (2004). Multisensory cortical processing of object shape and its relation to mental imagery. Cognitive, Affective, & Behavioral Neuroscience, 4, 251-259..

Liquefying text from human communication processes: A methodological proposal based on T-pattern detection

M. Teresa Anguera, Gudberg K. Jonsson, Pedro Sánchez-Algarra

University of Barcelona
Barcelona, Spain

University of Iceland
Reykjavík, Iceland

University of Barcelona
Barcelona, Spain

tanguera@ub.edu, gjonsson@hi.is, psanchez@ub.edu

Abstract

A rigorous methodology for the objective study of human communication in the form of textual material has a promising future. In this article, we propose a rigorous transformation of qualitative textual material derived from human communication processes into code matrices for subsequent quantitative analysis. We also show the ample possibilities of multivariate T-pattern detection in this area. The invisible nature of T-patterns increases the potential for discovery, as by extracting the internal structure underlying communication, researchers will be better equipped to unveil the key to human communication.

1 Introduction

There is an important vacuum in the scientific study of oral or written text produced during human communication given that there are no established procedures for "liquifying the text', i.e., for objectively extracting the necessary qualitative information for subsequent quantitative processing. Quantitiative analysis offers both objectivity and effectiveness in the study of processes. In recent decades, it has been common practice to use narratives to explore human emotions and experiences (Casey, Proudfoot, & Corbally, 2015; Gunaratnam & Oliviere, 2009). However, such an approach entails a high risk of methodological bias and the subjectivity of the researcher may take precedence in the analyses, which are solely qualitative in nature. Initially, qualitative data can capture emotions, experiences, ideologies, and realities, and are crucial for a better understanding of processes (of communication, adaptation to an environment, experiences of change, etc.). Intrinsically, however, they have considerable shortcomings.

The shortcomings inherent to the study of qualitative data constitute a persistent scientific gap that requires powerful solutions. Our proposal entails the rigorous transformation of qualitative textual material from human communication into code matrices that capture the categorical nature of the data and can be analyzed quantitatively through the objective detection of hidden temporal patterns (T-patterns) that form the underlying structure of the text, as a '*chain of behaviors that occurs significantly more often with approximately the same internal intervals than expected by chance given this zero hypothesis*' (Magnusson, 1996, p. 114).

Numerous studies have been published on methods for integrating qualitative and quantitative data in research (Fetters & Freshwater, 2015; Sánchez-Algarra & Anguera, 2013; Sandelowski, Voils, & Knafl, 2009). Our proposal for integration is novel as it consists of transforming qualitative data into code matrices for subsequent quantitative analysis using a robust process that preserves the wealth of information contained in the source texts, while guaranteeing rigorous analysis of the communicative sequence.

2 Method

2.1 Sources

Our proposal is methodological in nature. We develop and propose a process for transforming qualitative data into code matrices that safeguards the theoretical framework, retains the original semantic and emotional content, and guarantees maximum rigor and objectivity. The transformation of qualitative data for subsequent quantitative processing is not implemented at a structural level, as this would generate serious shortcomings at a scientific level. Carefully chosen information sources are key to collecting data on the participants in a communication process. By quantifying and subsequently processing the information contained in narrative texts, it is possible to faithfully capture and monitor different aspects of everyday life over time (Portell, Anguera, Hernández-Mendo, & Jonsson, 2015).

The information sources include anonymized multi-modal material, such as interviews, speeches, or conversations (Sidnell & Stivers, 2013), directed at a specific or non-specific receiver, with or without turn-taking, using only words or accompanied by visually perceptible elements (if a combination of direct and indirect conversation is analyzed).

2.2 Software

We use the freely available software package THEME, v. 6 Edu (Magnusson, 1996, 2000, 2005, 2015) to manage the databases and conduct the data analyses.

2.3 Procedure

The first step in the procedure is to decide on the criteria (or dimensions) that will be used to construct the *ad hoc* observation instrument for the study (Portell, Anguera, Chacón, & Sanduvete, 2015). These criteria are derived from the theoretical framework with additional consideration of any empirical experience available. They form the axis or backbone of the instrument and correspond to the different facets of the problem to be studied. A classic example is Weick's (1968) proposal for the dimensions to be considered in human communication, which have since been explored in depth (Anguera & Izquierdo, 2006).

The second step consists of deciding on the text segmentation criteria to acquire a series of text units. In the field of indirect observation, these text units are equivalent to the observation units used in direct observation (Krippendorf, 2013). The choice of text segmentation criteria is extremely important, since it will directly affect the data obtained; initial trials or pilot tests are recommended.

Once the dimensions and text segmentation criteria have been decided on, the next step is to construct the *ad* hoc indirect observation instrument (Anguera, Magnusson, & Jonsson, 2007). This instrument will combine the multi-dimensional field format with category systems. These category systems (or catalogues) must be generated for each dimension and in addition, they must meet the requirements of mutual exclusivity and/or exhaustivity (Anguera, 2003).

While not described in detail, the methodological actions outlined above enable the production of code matrices from all textual material or corresponding fragments analyzed. Code matrices are essential for quantitizing the qualitative data recorded throughout the communication process studied. Each code matrix displays temporal units in rows and dimensions in columns.

Consequently, each row of the table will display a string of codes corresponding to co-occurrences of behavior (Figure 1).

Indirect observation instrument	Sample of a matrix codes:
Criteria or dimensions: W, X, Y, Z	Time Event
Codes (mutually exclusive) derived from each criterion:	1 : 2 W2,Y7,Z3 3 W1,X1,Z2
W=W1, W2, W3, W4, W5	4 Y6,Z2 5 W5,X2,Y5,Z1
X=X1, X2, X3, X4	6 Y7,Z1 7 Y7,Z1
Y=Y1, Y2, Y3, Y4. Y5, Y6, Y7	8 Y6,Z1 9 Y6,Z2
Z=Z1, Z2, Z3	10 Y7,Z1 11 X2,Z1 12 Y7,Z3 […] 800 &

Figure 1. Simulated data in the form of a code matrix derived from text.

To better illustrate how the method works, we have included an example that uses real data from a study on the impact of a social constructivist-based intervention on discursive strategies used by four physical education teachers (García-Fariña, 2015). The study combined discourse analysis techniques with observations of the teachers' communication strategies in context

We have included a fragment of dialogue transcribed from video (Figure 2), a section of the observation instrument (Figure 3a), and a screenshot from ATLAS.ti showing the codes corresponding to the textual units (Figure 3b). These codes are then extracted from the window on the right in ATLAS.ti and transformed, using a minimal number of manual operations, into a matrix of codes (usually irregular), in which the minimum number of codes is one and the maximum number is the number of dimensions in the observation instrument. The codes are then ready for processing in THEME.

31:09 - P: ¿Tenemos que tener la bola controlada cuando vamos a tirar?
A: Si
31:15 - P: ¿Tenemos que tener la bola controlada cuando vamos a tirar? ¿Sí o no?
A: si es posible sí
31:22 - P: Si es posible sí. ¿Qué opinan ustedes? ¿Y cómo vimos que se tenía la bola controlada al máximo? ¿Qué había que hacer?
A: Tenerla separada del cuerpo....tenerla pegada al *stick*
31:37 - P: Tenerla pegada al *stick*. Entonces para golpear la bola hay que tenerla siempre pegada al stick para que sea un golpe preciso. Y ahora voy a recordarles el golpe del brill-stop. ¿Se acuerdan del brill-stop?
A: Si
31:55 - P: ¿Cómo era?, ¿cómo colocábamos los pies?
A: Asi
31:59 - P: ¿Qué pie se adelantaba?
A: El contrario.
32:00 - P: Jorge, ¿con qué mano golpeas?

Figure 2. Fragment of dialogue transcribed from video.

ATLAS.ti is only useful in this case for coding the units of text. As explained above, our method involves searching for complex patterns within data matrixes in THEME.

Criterio	Categorías	Descripción	Código
4. Segmento de interactividad	Segmento de actividad motriz	Es un segmento de práctica (actividad motriz) del alumnado, donde éste está ejecutando la tarea propuesta.	D1
	Segmento de organización de la actividad	Es un segmento de inactividad motriz, en el que el profesor/a aporta y/o solicita información previa al desarrollo de una o varias tareas.	D2
	Segmento de discusión	Es un segmento de inactividad motriz, en el que se recoge los episodios de interacción verbal entre el alumnado y el profesor sobre el contenido de la tarea.	D3
	Segmento de recapitulación	Es un segmento de inactividad motriz que se produce al inicio vinculando lo realizado en otra/s sesiones anteriores con lo que se va a realizar a continuación, y/o al final de las sesiones donde se reflexiona sobre la sesión completa.	D4
5. Estructura del mensaje	Mensaje unidireccional	Enunciados directos del profesor en forma de instrucción dirigido al alumnado, sin dar opción a una respuesta verbal o corporal del alumnado.	E1
	Mensaje bidireccional	Enunciados que están en clave interrogativa que busca y obtiene una respuesta inmediata l/espuesta verbal y/ica y obtiene una respuesta verbal y/o corporal del alumnado verbal o corporal del alumnado.	E2
	Mensaje Combinado	Enunciados que combinan mensajes unidireccionales y bidireccionales.	E3
6. Recurso extra-lingüístico	Con demostración	Alusión a gestos o posturas corporales que emplea el profesor, o algún alumno/o a indicación del profesor/a, para mostrar al alumnado modelos parciales o globales que aporten su intervención verbal.	F1
	Apoyo visual externo	Alusión a imágenes, esquemas o videos en cuyo marco interactúan con una experiencia el docente y el alumnado.	F2
7. Tipo de tarea	Tarea psicomotriz	Son aquellas situaciones motrices que son ejecutadas a nivel individual, en ausencia de interacción motriz con otro u otros alumno/as.	G1
	Tarea sociomotriz de cooperación	Son aquellas situaciones motrices que llevan a cabo el alumnado interaccionando motrizmente en forma de comunicación entre ellos y sin oposición.	G2
	Tarea sociomotriz de oposición	Son aquellas situaciones motrices que llevan a cabo el alumnado interaccionando motrizmente en forma de contracomunicación contra uno o varios.	G3
	Tarea sociomotriz de cooperación-oposición	Son aquellas situaciones motrices que llevan a cabo dos o más alumnos/as interaccionando motrizmente en forma de contracomunicación contra dos	G4

Figure 3a. Observation instrument.

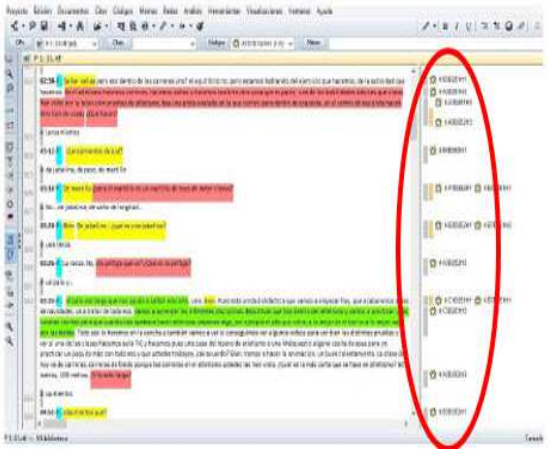


Figure 3b. Text as coded in ATLAS.ti.

Figure 3a shows part of the observation instrument used in García-Fariña et al. (2015), and Figure 3b shows example of coded text in ATLAS.ti.

2.4 Data quality control

The data quality control stage is a critical stage of the process. Concordance between different datasets, whether produced by the same coder at different times (at least three) or by different coders (also at least three), provides an important, though not the only, guarantee of reliability, and reliability is a necessary, though not sufficient, condition for validity (Krippendorf, 2013).

2.5 Data analysis

The complexity of human communication is converted, through systematic coding, into a series of episodes or strings of episodes represented by codes. These codes form code matrices that are perfectly aligned with the syntactic rules of THEME, our data analysis program. THEME detects T-patterns, which are structures that show the temporal relationships between different codes (or elements of the communication process) according to their sequence of occurrence and their separation in time (distance). These structures are not visible to the naked eye but they can be extracted from the code matrices through a robust quantitative analysis (Anolli, Duncan, Magnusson, & Riva, 2005).

Once the quality of the categorical data produced in the above steps has been demonstrated, the possibilities for analysis and the potential applications – in human communication and within and across many other fields – are huge. Here, we refer only to T-pattern detection, which can reveal the structure underlying communication flows in written or oral text by identifying regularities or temporal patterns that would otherwise go unnoticed (Blanchet, 2005; Magnusson, 1996, 2000, 2005; Vaimberg, 2010). T-pattern detection, however, can also be used in direct observation (Amatria et al., 2017) and laboratory studies (Casarrubea et al., 2015) and has previously been used in number of studies with focus on multimodal behaviors, i.e. verbal and non-verbal interaction in dyadic interaction (Jonsson, 2006).

When working with real datasets, the standard procedure for selecting and interpreting T-patterns generated by the software is the application of three "quantitative" sort options in THEME and three "qualitative" filters established by the researchers (Amatria, Lapresa, Arana, Anguera, & Jonsson, 2017). We use the same approach in this example to analyze simulated data.

Figure 4 shows one of 111 T-Patterns detected within the 800 rows of the code matrix generated from anonymized data.

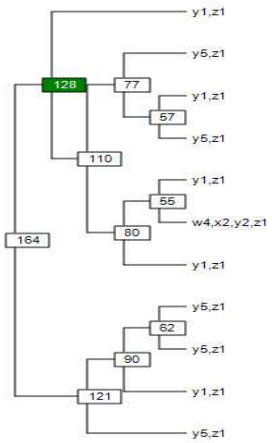


Figure 4. T-patterns from a hypothetical example.

Figure 5 shows some of the results from the analysis of real data (generated from the observation of five physical education classes) using ATLAS.ti codes in THEME. The analysis revealed an identical pattern, corresponding to a complex structure, in three of the five sessions. The pattern was formed by identical events (codes) that occurred in the same order and were separated by the same distance in time.

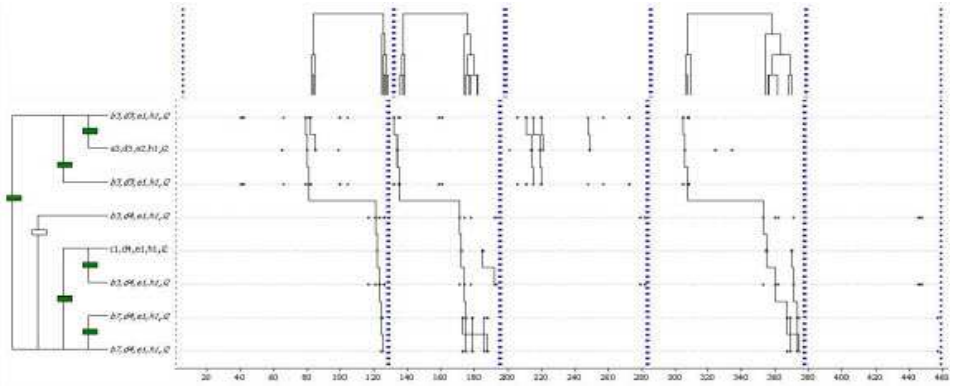


Figure 5. Example of a pattern detected from ATLAS.ti codes.

3 Discussion

Communication takes many forms, ranging from basic dyadic interactions between two individuals who regularly communicate and take decisions, to multiple group interactions in real-life or virtual situations. All forms of communication, however, are amenable to T-pattern analysis. Communication flows offer enormous research potential thanks to their multiple criteria or dimensions and their extraordinarily dynamic nature. Their study, however, poses methodological challenges, beginning with the establishment of appropriate dimensions or response levels (known as variables in THEME) and the selection of criteria to segment the episodes into behavioral units, which give rise to event types. The results of our analysis of both real and simulated data indicate that the analysis of code matrices derived from ATLAS.ti in THEME can optimize the process of studying verbal behavior. Verbal behavior represents, of course, just one of the multiple levels of human communication, each of which can be analyzed following the same procedure.

References

Amatria, M., Lapresa, D., Arana, J., Anguera, M.T., & Jonsson, G.K. (2017). Detection and selection of behavioral patterns using Theme: a concrete example in grassroots soccer. *Sports, 5*, 20; doi:10.3390/sports5010020.

Anguera, M. T. (2003). Observational Methods (General). In R. Fernández-Ballesteros (Ed.), *Encyclopedia of Psychological Assessment* (pp. 632-637). London: Sage.

Anguera, M.T. & Izquierdo, C. (2006). Methodological approaches in human communication. From complexity of situation to data analysis. In G. Riva, M.T. Anguera, B.K. Wiederhold & F. Mantovani (Coord.), *From Communication to Presence. Cognition, Emotions and Culture towards the Ultimate Communicative Experience* (pp. 203-222). Amsterdam: IOS Press.

Anguera, M. T., Magnusson, M. S., & Jonsson, G. K. (2007). Instrumentos no estándar [Non-standard instruments]. *Avances en Medición, 5*, 63-82.

Anolli, L., Duncan, S., Magnusson, M.S. & Riva, G. (Eds.) (2005). *The Hidden Structure of Interaction. From Neurons to Culture Patterns*. Amsterdam: IOS Press.

Blanchet, A., Batt, M., Trognon, A., & Masse, L. (2005). Language and behaviour patterns in a therapeutic interaction sequence. In L. Anolli, S. Duncan, M. S. Magnusson, & G. Riva (Eds.), *The hidden structure of social interaction. From Genomics to Culture Patterns* (pp. 124-140). Amsterdam: IOS Press.

Casarrubea, M., Jonsson, G.K., Faulisi, F., Sorbera, F., Di Giovanni, G., Benigno, A., Crescimanno, G., & Magnusson, M.S. (2015). T-pattern analysis for the study of temporal structure of animal and human behavior: A comprehensive review. *Journal of Neuroscience Methods, 239*, 44-46. doi: 10.1016/j.jneumeth.2014.09.024.

Casey, B., Proudfoot, D., & Corbally, M. (2015). Narrative in nursing research: an overview of three approaches. *Journal of Advanced Nursing 72*(5), 1203–1215. doi: 10.1111/jan.12887

Fetters, M. D. & Freshwater, D. (2015). The 1+1=3 integration challenge. *Journal of Mixed Methods Research, 9*, 115-117. doi:10.1177/1558689815581222

García Fariña, Abraham (2015). *Análisis del discurso docente como recurso metodológico del profesorado de Educación Física en la etapa de Educación Primaria* [Analysis of teacher-led discourse as a methodological resource for primary school physical education teachers] (Directors: Francisco Jiménez Jiménez and M. Teresa Anguera). Unpublised Doctoral Dissertation. Tenerife: Universidad de La Laguna.

Gunaratnam Y. & Oliviere D. (Eds.) (2009) *Narrative and Stories in Health Care: Illness, Dying and Bereavement*. Oxford University Press, Oxford.

Jonsson, G.K. (2006). Personality and Self-Esteem in Social Interaction. In In G. Riva, M.T. Anguera, B.K. Wiederhold & F. Mantovani (Coord.), *From Communication to Presence. Cognition, Emotions and Culture towards the Ultimate Communicative Experience* (pp. 186-212). Amsterdam: IOS Press.

Krippendorff, K. (2013). *Content analysis. An introduction to its methodology* (3rd. ed.). Thousand Oaks, CA: Sage.

Magnusson, M. S. (1996). Hidden real-time patterns in intra- and inter-individual behavior. *European Journal of Psychological Assessment, 12*, 112–123.

Magnusson, M. S. (2000). Discovering hidden time patterns in behavior: T-patterns and their detection. *Behavior Research Methods, Instruments & Computers, 32*, 93–110.

Magnusson, M. S. (2005). Understanding social interaction: Discovering hidden structure with model and algorithms. In L. Anolli, S. Duncan, M. S. Magnusson, & G. Riva (Eds.), *The hidden structure of social interaction. From Genomics to Culture Patterns* (pp. 4-22). Amsterdam: IOS Press.

Portell, M., Anguera, M.T., Chacón, S. & Sanduvete, S. (2015). Guidelines for Reporting Evaluations based on Observational Methodology (GREOM). *Psicothema, 27*(3), 283-289.

Portell, M., Anguera, M. T., Hernández-Mendo, A., & Jonsson, G. K. (2015). Quantifying biopsychosocial aspects in everyday contexts: An integrative methodological approach from the behavioral sciences. Psychological Research Behavior Management, 8, 153-160.

Sánchez-Algarra, P. & Anguera, M.T. (2013). Qualitative/quantitative integration in the inductive observational study of interactive behaviour: Impact of recording and coding predominating perspectives. *Quality & Quantity. International Journal of Methodology, 47*(2), 1237-1257.

Sandelowski, M., Voils, C. I., & Knafl, G. (2009). On quantitizing. *Journal of Mixed Methods Research, 3*, 208-222.

Sidnell, J. & Stivers, T. (Eds.) (2013). *The Handbook of Conversation Analysis*. New York: Wiley and Sons.

Vaimberg, R. (2010). *Psicoterapias tecnológicamente mediadas* [Technology-mediated psychotherapy]. (Directors: Adolfo Jarne and M. Teresa Anguera). Unpublished Doctoral Dissertation. Barcelona: Universidad de Barcelona.

Weick, K. E. (1968). Systematic observational methods. In G. Lindzey, & E. Aronson (Eds.), *Handbook of Social Psychology*, vol. 2 (pp. 357-451). Reading, Mass.: Addison-Wesley.

The bodily expression of negation in Polish

Jolanta Antas, Sonia Gembalczyk

Jagiellonian University, Faculty of Polish Studies, Department of Communication Theory

Kraków, ul. Gołębia 20, Poland

pantas@poczta.onet.pl, sonia.gembalczyk@gmail.com

Abstract

The aim of this paper is to examine the bodily expressions of negation in Polish with the use of both audiovisual and syntactic material. The concept of NOT expresses many different degrees of rejection in such areas as belief (*I don't know*, *I doubt*), judgment (*bad*), emotion (*I don't want*) and action (*I don't do*). We have found a whole range of reactions reflecting negation, including movements of the head, arms, and hands, facial expressions, intonation, and proxemic communication. Multimodal illustrations point to both the polymorphism of the act of negating, and the embodied sources of negation.

1 Introduction

Negation is not only a mental operation, it is also a physical experience of resistance, aversion or rejection. We address the topic of the bodily expression of negation with a view to discovering any sensomotoric sources related to the need to manifest the emotional and mental states of negation (see: Bressem & Müller, 2014, pp.1601-1602). The controversy surrounding Anna Wierzbicka's proposal to place the NOT unit among the NSM (Natural Semantic Metalanguage) units (Żurowski 2005), the so-called semantic primes, clearly indicates that NOT, perceived as a logical operator, is ambiguous. It is not always possible to replace the operator IT IS NOT TRUE THAT with the linguistic NOT. A more thorough research into the linguistic NOT (Antas, 1991) has shown that it expresses not one, but several different states of rejection – in areas such as belief (*I don't know, I doubt*), judgment (*bad*), emotion (*I don't want*) and actions (*Don't do it, I won't do it*). In addition, every act of denial has its own affective tone. Emotional by nature, man is unable to participate in any act of communication without emotions, if only the subtlest ones. As a result negation in its active, pragmatic, multimodal (and not merely textual) form always assumes some emotional "hue".

While examining different multimodal manifestations of negation, we noticed that:

(1) negation is always accompanied by judgement,

(2) there exists a whole range of reactions expressing negative ideas (such as head or hand gestures, intonation, mimics and proxemic behaviour), which we believe indicates a **multitude of sources** of negation and its expression (mental, emotional, interactive and intrapersonal). In other words, we postulate not only the **polymorphism** of negation itself, but also of the various sources of the need for negation (evoking different image schemata).

2 Material and methods

We have analyzed 350 items of audiovisual material featuring users of the Polish language. These have been selected from many hours of recordings from television programmes (98 hours), public offices (19 hours) and with the participation of students of Polish at the Jagiellonian University (31 hours). More than 150 individuals agreed to have their image published for research purposes. We strived to obtain material based on versatile sources and natural conversation (such as TV

interviews and public office inquiries), rather than just derived from university experiments (communication task). We have grouped negative reactions according to three characteristics: (1) the form of their bodily expression, (2) the intensity of the means of expression and (3) the co-occurrence of repetitive negations occurring on the textual plane. The analyzed expressions are primarily those containing verbal negation (with the use of the particle "not" or other semantic negative markers, such as: *odwrotnie* (Eng. *conversely), przeciwnie* (Eng. *on the contrary), wcale* (Eng. not at all) (Antas 1991, pp. 132-145). All the identified gestures which accompany verbal negation have been included in the corpus, even if they occur as independent nonverbal acts of speech (and thus without accompanying words). Most of the words used in conversation by speakers of the Polish language were of a mixed character, so the meaning of the negative expression was achieved by various semiotic means, multimodally. Therefore, we have grouped the results according to both the verbal and gestural manifestations of negation (sometimes the same statement could be found simultaneously in different groups). Some verbal expressions of negation have been made without any clear nonverbal signals (user refraining from gestures) and these did not enter the corpus. At the present stage of research, we do not deal with the temporal or syntactic relationship between words and non-verbal signals.[1] The interpretations of linguistic expressions may not exactly correspond to the specifics of the English language, but are intended to reflect the ideas embodied in the Polish phrases.

3 Analysis

3.1 Head shake

The primitive "not", according to Desmond Morris, has its origin in infancy when the baby moves their head away from the mother's breast in order to signal the end of feeding (Morris, 1977, p. 69). More recent research in ethology confirms the validity of Morris's observations (e.g. Tanner et al., 2006, p. 76). Similarly, Johnson defends the view that the search for meanings should begin with the analysis of the most primitive bodily movements, e.g. those performed by infants (Johnson, 2015, pp. 51-70). The extent to which we need to use our bodies to express negation in the event of a bodily discomfort is exemplified by the shaking of the head while grunting in response to, say, throat irritation (Fig.1). This is an intrapersonal behaviour which also reveals the bodily source of negation.

Figure 1. Head shake in response to throat irritation (from the collection of the Department of Communication Theory).

Head shake, which expresses rejection in all modal spheres and all possible emotional variations, can be expressed through different gestural variants from single movements to multiple ones intensified through other channels. Kendon emphasizes the multiplicity of contexts of use and the vast semantic possibilities of the head shake. He believes that this gesture: *is used in many*

[1] Gestures which serve as elements of text organization, or in other words, syntactic functions of gestures in the Polish language are described by Antas (2013, pp. 94-97). Unlike Harrison and Larrivée (2016, p. 79), who found that the participants *synchronize the gesture stroke with the negative node (...) with the vocal clausal negator in English*, we have observed that Polish speakers often use gestures which far precede verbal negation or that they replace words with nonverbal reactions. Such phenomena with regard to negation expressed by Polish speakers require further investigation.

different discourse contexts where, although, as we shall argue, it can always be interpreted as expressing a 'theme' of negation, it yet comes to have a very different force, depending upon how this theme of negation combines with the other semantic themes that are also being expressed (Kendon 2002: 148). For example, a verbal *no* repeated six times and accompanied by the head shaking, a hand cut gesture, and a break in eye contact signifies negation combined with reluctance and the need to establish an interactive boundary. In one of our examples, the guest of a popular TV show thus responds to the host's encouragement to ask him a question: *I don't know what ...* [repeated snorts, letting out air]. *I don't not know what sort of ... question I could ask you* – this is followed by a number of shoulder shrugs. While saying: *I don't know,* the guest makes a short and energetic head shake, preceded by breaking eye contact. The statement is accompanied by a mimic expression of dislike and a discouraged, slightly impatient tone of voice. Thus, verbal negativity was reinforced by the negative head shaking and five other nonverbal communication channels (facial expressions, intonation, paralinguistic effects, shoulder movements, visual contact). In this example, multimodal negation becomes semantically related with indecision and fatigue.

3.2 Arms spreading apart

The open hand gesture has been widely described by researchers (Bressem & Müller, 2014; Antas, 2013, pp. 234–236; McNeill, 2005, pp. 49–51; Kendon, 2004, pp. 264–281; Załazińska, 2001, pp. 65–80; Morris, 1977, p. 56). Antas notes that an open hand or open hands in questions express mental willingness of the subject to accept new cognitive values (Antas, 2013, p.237, cf. Kendon 2004: 273–275). Meanwhile, while negating, the gesture of open hands turns out to reveal other sensorimotor sources. The essence of this movement is its trajectory: the hands move away from one another.

Arms spreading apart – a very popular gesture expressing negation illustrates different degrees of inability – from impossibility to helplessness (cf. Kendon 2004: 275-281). We have observed the movements of open hands signalling lack of qualities, in line with the metaphor: TO KNOW IS TO POSSESS, and also one indicating the impossibility of manipulation expressed by the dropping and spreading apart of the arms combined with the hand shrug (to emphasise helplessness) (Fig.2). Ekman and Friesen include this gesture in the hand shrug emblems category (1972, p. 366), and Bavelas et al.describes it as an interactive gesture, which could be paraphrased by the words: *What else could I do?* (Bavelas et al., 1992, pp. 472–475).

Figure 2. Arms spreading apart (from the collection of Department of Communication Theory).

In our opinion, an interpretation of the gesture of hands moving away (in different ways depending on different qualities) should also include set phrases. Phraseology is sometimes a mirror of inter-subjective imagery and motor patterns which are inherent in concepts. We say: *rozłożył ręce* (Eng. 'he spread his hands', meaning: 'there was nothing he could do') or stronger *opadły mi ręce* (Eng. 'my arms have dropped down', meaning: 'I was powerless, helpless') but also: *nie poruszaj tego problemu* (Eng. 'don't touch this problem') or *ja się tego nie tykam* (Eng. 'no, I'm not even touching it').

Figure 3. Arms spreading apart (from the collection of the Department of Communication Theory).

The latter meaning may be expressed by the pushing forward of the hands spread apart and pointing upwards, thus indicating not only ignorance, but also unwillingness to take up the subject. By repeating (Fig. 3): *I don't know* twice, the speaker indicates that they have no intention of discussing a particular subject (cf. Kendon, 2004, p. 277).

3.3 Pushing-away gesture

The expression of *doesn't matter* is always accompanied by a pushing-away gesture (Antas, 2013, pp. 225-230, see Bressem & Müller, 2014). Here personified thoughts and ideas are pushed away from the body of the subject – what is irrelevant for the subject should disappear from their field of vision, or at least be pushed aside, as opposed to important concepts that we always wish to present by gesture as objects which we hold and possess (in line with the metaphor: TO HOLD IS TO CONTROL).[2]

3.4 Hand-cut

It is important to distinguish two patterns of negation: a cut off and a cut with a hand or hands. Antas emphasizes that *the cut-off gesture is always accompanied by a very sharp expression of negation and protest* (Antas 2013: 246, cf. Kendon, 2004, p. 262). The author prefers to regard the movement of the hands, which researchers call *hand scissors* (Morris, 1977, 51), as an act of self-detachment. She refers to the gesture as a baton *which the subject uses to separate themselves from an issue* (Antas 2013, pp. 245–248). On the other hand, the hand-cut gesture may have different variants, but it is always a horizontal and sharp cut (see Kendon, 2004, p. 263). The cutting can be made with one or both hands (Fig.4).

Figure 4. Cut with hands (from the collectionof the Department of Communication Theory)

The gesture is probably derived from the original use of primitive tools, such as a scythe, sickle or machete. This has also found its way into popular verbal expressions, such as: *uciąć* 'cut', *wytrzebić* 'to thin out, to geld', *ukrócić czyjeś zapędy* 'to thwart someone's intentions'. There are also examples of similar imagery in the English language: *to shorten, stumped for words, cut-throat*

[2] We find in our material the presence of all four types of gestures of the Away family, described extensively in the German language by Bressem & Müller (2014).

(speed or prices), mow (somebody) down. Firmness and, in a way, ultimate character of negation expressed by a cutting motion often accompanies the terms *nothing* or *everything*.

3.5 Interactive gestures

Interactive gestures may also include a grain of the embodied sense of negation as a relational phenomenon – a phenomenon which arises from and functions in the social contact of a person with the surrounding environment. Such intimately interrelated functions are expressed with the gesture of the outstretched hand holding back any possible objections on the part of the interlocutor – gestures imaging a blockage which Bavelas paraphrases in the words: *Don't interrupt me* (Bavelas et al., 1992, pp.472–476, see Bressem & Müller, 2014, pp.1597–1598, Kendon 2004, pp. 251–255).[3]

3.6 Wagging of the index finger

Wagging of the index finger indicates that the person wishes to express strong detachment from the subject or even flash a 'no entry' sign. It most likely falls into the interactive gestures category, which is confirmed by the simultaneous breaking of the eye contact we observed in the sample group. The wagged finger while maintaining the eye contact carries an even stronger expression of 'no entry'.

3.7 Proxemic negation

Stepping back can sometimes be considered an interactive gesture. Indeed, a rapid movement away from the interlocutor indicates the need to increase the interactive distance. On the other hand, we should take a closer look at such expressions as: *I've been taken aback at the thought of ...* or *back away from* (something unpleasant or frightening) as opposed to: *become closer to somebody*, *close to somebody's heart*. These phrases suggest the need to express the rejection of an idea with our bodies. In other words, an objectified concept (thought, memory, project, idea) can evoke the feeling of physical rejection. In the following examples, the individuals expressing proxemic negation refer more clearly to a certain idea than to the interlocutor themselves or their attitude. An actress is "taken aback" by a reference made by the host, on another occasion, the host is "dumbstruck" by a story related by the guest which made him pull back in his chair (in response to the repulsive image in his mind which the story evoked).

3.8 Mimic "no"

Mimics uniquely reflect affective attitudes. The mimic "no" expresses disgust and aversion. This nonverbal reaction is often accompanied by the prototypal head shaking. While the simultaneous occurrence of verbal and mimic negation is quite common, we have also encountered situations in which facial expressions far precede verbal negation. According to Ekman, the timing and length of a mimic expression determines whether it is a *facial expression of emotions* or a *referential expression of emotions* (Ekman, 1997, p. 340). Nevertheless, we regard most of the mimic expressions in the analysed cases as *conversational facial gestures* (Bavelas et al., 2014, pp. 10-16) and deal with them in relation to a *particular microsocial moment* (Bavelaset al., 2014, p. 2). For example, an interviewed actor says: *No life without acting*, and at the same time wrinkles his face and squints his eyes, thus expressing distaste for the prospect of life devoid of acting. He also opens his arms and shrugs his shoulders. Thus, negation once again occurs with the use of several simultaneous means.

[3] In his research, Harrison accurately regards this type of reactions as illustrations of meanings created simultaneously on the interaction axis and the modality axis (2015). In our material base, we do not have enough examples so far to verify Harrison's thesis for gestures which accompany verbal statements in Polish.

3.9. Negation of a predicate

A situation in which a verbal negation is accompanied by a gesture illustrating an undeniable predicate, is a separate issue. For example, a recognized-music critic says: *We don't have much …*, while lifting both hands, as if he were holding a large ball. The gesture indicating a large quantity is verbally negated.

Figure 5. Things that don't literally touch me directly (from the collection of the Department of Communication Theory)

Similarly, an actress says: *Things that don't literally touch me directly* while curling the fingers of both hands and putting them together in front of her chest (which carries the opposite meaning to: *They touch me directly*) (Fig.5). As Antas says: *This kind of behaviour is confirmed by Wygotski's thesis on the absolute predicativity of the inner speech, and the thesis of Hostetter and Alibali on the so-called threshold of gesture* (Antas 2013, pp. 223–224). Or elsewhere: *[Hostetter and Alibali] suppose that gesture takes the form closest to the nature of simulation carried out in the mind* (Antas, 2013, p.159). In other words, gestures always express areas which the speaker finds most important and most prominently profiled. If the subject identifies themselves, emotionally, and thus sensually, with negation, then their body also expresses negation. If negation occurs at a purely logical level (IT IS NOT TRUE THAT), then it is not present in the gesture. In this case, the body serves to express a non-negated mental image, which does not "yield" to logical negation. Thus, predicative gestures indicate that negation can be a logical operation based on affirmation. But there is a difference between a logical "no" and a pragmatic one. The latter is subjective and it is not just a logical construct. The pragmatic "no" is emotionally linked with the subject. We hope that we have succeeded in showing the qualities it expresses.

4 Conclusions

We have observed that different gestures appear in different spheres of expressing negation. The open hands and spread arms occur in the sphere of beliefs, and consequently within the associated epistemic modality, accompanying such expressions as: *nie wiem* 'I don't know'*, wątpię* 'I doubt'*, nie da się* 'no way', *nie ma* 'there is no…' etc. Rejection expressed by pushing the hands away from the body is the commonest in the area of evaluation (*nieważne* 'doesn't matter'). Also in the area of evaluation (*źle* 'wrong'*, okropne* 'awful', etc.) and emotions (*nie chcę* 'I don't want', and any signs of disgust, embarrassment, and even surprise) we can observe the gestures of moving away (the proxemic "no"), and even more importantly the mimic negation. The popular hand-cut gesture belongs primarily in the sphere of action and the strongly related deontic modality. On the other hand, head shaking – the most prototypal of all forms of negation, occurs in all areas and all kinds of emotions, sometimes even carries intrapersonal qualities. Predicative gestures occurring with logical negatives simply illustrate concepts that are contradictory and do not say anything about the bodily source of negation. NOT as a concept may appear simultaneously in one or more communication channels (it can also be expressed in an exclusively non-verbal manner), while the logical operator IT IS NOT TRUE THAT appears only in a verbal form. Conversely, a verbal negation accompanied by the above bodily expressions may appear in the form of various textual operators (Antas, 1991). The above analysis, focusing on the communicative behaviours of Poles,

does not exhaust the subject matter of bodily expression of negation, but was intended to indicate the main sensorimotor sources of this heterogeneous phenomenon and to designate further research fields in Polish.

References

Antas, J. (1991). *O mechanizmach negowania. Wybrane semantyczne i pragmatyczne aspekty negacji*. Kraków: Universitas.

Antas, J. (2013). *Semantyczność ciała. Gesty jako znaki myślenia*. Łódź: Primum Verbum.

Bavelas,B.-J., Gerwing, J., & Healing, S. (2014). Hand and Facial Gestures in Conversational Interaction. In T.-M. Holtgrave (Ed.), *The Oxford Handbook of Language and Social Psychology* DOI: 10.1093/oxfordhb/9780199838639.013.008

Bavelas, B.-J., Chovil, N., Lawrie, D.-A., & Wade, A. (1992). Interactive Gestures. *Discourse Processes,* 15, 469–489.

Bressem, J., & Müller, C., (2014). The family of Away gestures: Negation, refusal, and negative assessment. In C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, & J. Bressem (Eds.), *Body – Language – Communication. An International Handbook on Multimodality in Human Interaction (Handbooks of Linguistics and Communication Science 38.2.)* (pp.1592–1604). Berlin-Boston: De Gruyter Mouton.

Ekman, P. (1997). Should we call it expression or communication? *Innovation,* 10(4), 333–344.

Ekman, P., Friesen, W.-V. (1972). Hand Movements. *The Journal of Communication*, 22 (4), 353–374.

Harrison, S., & Larrivée, P. (2016). Morphosyntactic correlates of gestures: A gesture associated with negation in French and its organisation with speech. In P. Larrivée & Ch. Lee (Eds.), *Negation and polarity. Experimental perspectives* (pp.75-94). Dordrecht: Springer.

Harrison, S. (2015). A modality axis in gesture space? The Vertical Palm and construal of negation as distance. In F. Galle, & M. Tutton (Eds.), *Gesture and Speech in Interaction* - 4th edition (GESPIN 4). (pp.137–141). Nantes: Hall.

Kendon, A. (2002). Some uses of the head shake, *Gesture* 2(2) , 147–182.

Kendon, A. (2004). *Gesture: Visible Action as Utterance*. Cambridge: Cambridge University Press.

McNeill, D. (2005). *Gesture and Thought*. Chicago/London: University of Chicago Press.

Morris, D. (1977). *Manwatching: A field-guide to human behavior*. London: Jonathan Cape.

Tanner, J-E., Patterson, F.-G., Byrne, R.-W. (2006). The development of spontaneous gestures in zoo-living gorillas and sign-taught gorillas, https://www.researchgate.net/publication/237516373 (26.03.2017), pp. 69-102.

Załazińska, A. (2001) *Schematy myśli wyrażane w gestach. Gesty metaforyczne obrazujące abstrakcyjne relacje i zasoby podmiotu mówiącego*. Kraków: Universitas.

Żurowski, S. (2005). Negacja jako element MSN. *Polonistyka 2004/2005*,.249-256.

Negotiation of meaning in ELF (English as a Lingua Franca) interaction: Multimodal approach focusing on body movements including gestures

Hiroki Hanamoto

Tokyo Denki University
Ishizaka Hatoyama-machi, Hiki-gun, Saitama, Japan

hiro_warriors@mail.dendai.ac.jp

Abstract

In the field of English as a lingua franca (ELF), numerous approaches to data analysis have been applied to empirical research since the 1990s. Many previous studies have reported the importance of mutual intelligibility across cultures and contexts. It is also a matter of fact that intelligibility issues in ELF interactions may sometimes be negotiated and developed by ELF speakers to achieve mutual understanding when there is a breakdown in understanding. What is important to note here is that not much research has so far taken into account resources other than the verbal language that ELF speakers employ in the process of their negotiation, and few studies have analyzed data from low or intermediate proficiency students. This study, therefore, examines the process of ELF interaction in repair work among such speakers, taking into account all of the resources that they employ. A dyadic spoken interaction between a Japanese student and one international student from Turkmenistan who is studying at a Japanese university is the basic of the present study. They enrolled in science and engineering as their major. The data analyzed for this study is based on video recorded ELF conversational interactions. To enrich the data analysis, the author also conducted retrospective stimulated recall and post-interview tasks as well. Through sequential analysis, which addresses how talk is initiated and negotiated, and how mutual understanding is achieved between a speaker and a listener, we found two main results based on the recorded corpus used in this study. Firstly, the participants employed verbal interactional communication strategies, such as confirmation checks, clarification requests and repetition. However, by paying careful attention to sequence organization, a close analysis shows that they also use alternative resources in their interactions as a medium of communication, namely body actions including gestures, often in combination with one another. These work in combination with the verbal strategies to co-construct meanings, which is realized by both interlocutors being actively engaged in conversation to complete the speaker's utterance, when they face difficulties in expressing themselves verbally. Second, these multimodal resources also involve functions such as turn-taking, supporting the development of the conversation, clarifying messages, and rapport-building. Furthermore, we found that the speaker changed the primary communication resource into non-verbal actions to fill difficulties in expressing verbally and the listener mimicked the interlocutor's gesture as a listener, and they each initiated negotiation for co-constructing meanings with the use of the non-verbal resource. This implies that using body movements including gestures in combination with words is a useful way of co-constructing meanings, enhancing explicitness, and showing as an accommodation strategy. A significant finding in this study is that participants can employ various interactional sequential procedures for co-construing meaning, both verbal and non-verbal. This study then suggests pedagogical implications for enhancing interactional communication strategies and emphasizing the important function of non-verbal actions in ELF studies and language teaching. In this presentation, the author will draw attention to each of the resources which the participants employ overtly in the course of conversation.

Index Terms: multimodality, body movements, sequential analysis, ELF interaction

1 Introduction

The analysis in this study shows that ELF speakers achieve mutual understanding through the use of interactional communication strategies and other multimodal resources. Actually, most previous studies have so far focused on linguistic verbal strategies for overcoming interactional issues, such as phonological or lexical problems (e.g., Deterding & Kirkpatrick, 2006; Jenkins, 2000; Seidlhofer, 2001). What is important to note here is that not much research has so far taken into account resources other than language that ELF speakers employ in the process of their interaction. This study, therefore, attempts to clarify two things: the means that ELF speakers employ in their interactions and how those resources are utilized creatively in the course of negotiation between the speaker and the listener.

2 Literature review

2.1 English as a lingua franca (ELF)

ELF is the common means of communication between people who are from different L1 backgrounds and different linguacultures (Jenkins, 2006; Seidlhofer, 2001). ELF connotes all English users, including second language, foreign language, and native speakers. This means that there are major differences in linguaculture and English proficiency among ELF speakers. Thus, one must consider that ELF speakers might negotiate to overcome these gaps and achieve mutual understanding in lingua franca situations through the use of interactional communication strategies.

2.2 Interactional communication strategies in ELF interaction

Many ELF studies have described the practices of interaction involved in lingua franca communication (e.g., Cogo, 2012; Jenkins, 2006; Kaur, 2011; Seidlhofer, 2001). Here, the most noticeable "niche" (Swales, 1990, p. 142) for this study is that ELF interaction research has primarily been centered on linguistic interactional behavior and verbal strategies that ELF speakers utilize in interaction. It is apparent that interlocutors in face-to-face interaction use not only linguistic but also non-verbal and para-linguistics features, such as gestures, body movements, eye gaze, back-channeling, and laughter (Goodwin, 2003), namely multimodal channels and resources. For instance, Gullberg (1998) acknowledges that ELF speakers utilize some multimodal channels and resources as well as language when they find it difficult to express their message and negotiate to overcome problems through the use of verbal communicative elements, and further reported that they use gestures more often for this purpose.

It is crucial that studies of multilingualism should pay far more attention to embodiment and multimodality than they do at present. This argument can be applied to ELF interaction research as well. However, as the author suggested above, research which includes other modes as well as language when the interlocutors feel difficulty in conveying their meaning verbally is still in its infancy (e.g., Canagarajah, 2013; Stein & Newfield, 2006), especially the study of multimodal channels and resources, including non-verbal actions. For example, analysis of ELF interaction has so far based on audio-recorded data, while multimodal analysis of ELF interaction using video-recorded data is fairly new (but see Konakahara, 2016; Matsumoto, 2015). In light of this, the following research questions were formulated for the present study: What and how multimodal resources are employed by ELF speakers in interactions?

3 Method

3.1 Data collection

A Japanese female, Shizu (L1 = Japanese) and a Turkmen male, Naz (L1 = Turkmen) (pseudonyms) majoring in science and engineering at a Japanese university voluntarily participated in this study. Participant's attributes, including gender, L1, and English level are shown in Table 1. Participant's names appearing in this paper are pseudonyms to protect their privacy. The data recording was undertaken in 2015, after the author had obtained permission from the participants to

video-record their conversational interaction. The author adapted the retrospective stimulated recall and post-interview to support the data interpretation and achieve data triangulation (Denzin, 1978).

Table 1. Participants in the recorded interaction.

Name	Gender	L1	Duration of recording (min:sec)	Relationship	English proficiency
Shizu	female	Japanese	29.77	First meeting	beginner
Naz	male	Turkmen			intermediate

3.2 Data analysis

This study focuses on qualitative analysis of the interactional work involved in negotiation. In order to clarify the process of coming to understand, the author employed sequential analysis based on previous studies (e.g., Pitzl, 2005; Schegloff, Koshik, Jacoby & Olsher, 2002). Sequential analysis is based on "emic" (Smit, 2010) and a bottom-up approach and can explicate the detailed process in making unintelligible utterances clear as participants make progress toward mutual understanding.

The author attempted to address instances where difficulty in understanding is displayed overtly and explicitly through repair sequences. Both a speaker and a listener co-construct meanings in the sequence of turns in the talk by being actively engaged in conversation to complete the speaker's utterance. If the meanings of utterances of a previous turn are unclear to an interlocutor, s/he might initiate negotiation of meaning in order to understand what was said, though some problems are left or ignored. Schegloff (2000) claims that repairs are "practices for dealing with problems or troubles in speaking, hearing, and understanding the talk…" (p. 207). Therefore, the author assumes that repair can be defined as a key indicator of resulting from some kinds of trouble in understanding.

It should be noted that the purpose of this study is to focus on verbal communication strategies and multimodal resources which are employed in repair work, namely body movements including gestures. Gestures are the one of the most frequently used means of non-verbal communication between/among speakers and listeners and are movements of the hands and arms (McNeill, 1992). A gesture used pragmatically in an interaction is a means of making meanings explicitly. According to McNeill (2005), a speaker presents her/his message using gestures together with what s/he is willing to express verbally. According to McNeill (1992), gestures are classified into four types: iconic, metaphoric, beat, ad deictic. Gestures often occur simultaneously with other gestures, and also have a range of functions such as "modal", "performative", "parsing" and "interactive or interpersonal functions" (Kendon, 2004, p. 159). In other words, such gestural features should be integrated into a close analysis of sequences for a display of multimodal resources by participants and be categorized for data analysis, although the purpose of this study is not to attempt to classify gestures in a strict manner. Besides, it is also meaningful to integrate not only gestures but also other semiotic resources, such as gaze, posture, and facial expression (Goodwin, 2003), to gain a better understanding of the process of participants' negotiation of meaning for understanding.

The present study is data-driven and the author analyzed the data relying on the transcriptions and the video-recorded and stimulated recall data as support. A dyadic spoken interaction was transcribed and analyzed by the author using the conversation analysis transcription conventions. Transcriptions were made using a slightly adapted form of Jefferson (1984) in order to describe conversational interaction in detail, and also drew on McNeill (2005) for non-verbal elements (see Appendix for transcription conventions).

4 Results and discussion

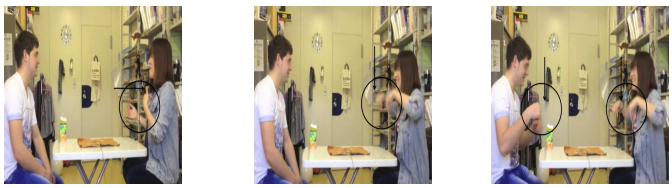
This section focuses on participants' use of multimodal resources in repair work; particularly body movements, including gestures. The author initially assumed that the participants would often employ some combination of verbal interactional communication strategies, such as repetition, confirmation checks, paraphrasing, clarification requests, and non-verbal actions such as body movements, including gestures. Excerpt 1 is an interaction between a Japanese female, Shizu (L1 = Japanese), and a Turkmen male, Naz (L1 = Turkmen). Shizu has proposed a new topic, "Disneyland", which is a Walt Disney's theme park in Japan. Here, Shizu does not display problems in hearing or understanding but rather is struggling with difficulty in expressing herself verbally.

Excerpt 1: "why do you like Disney so much?"

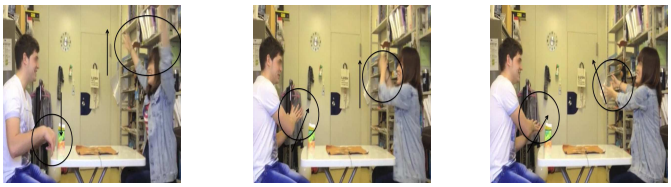
```
S (f) = Japanese; N (m) = Turkmen
   1. N: why do you like Disney so much?
   2. S: eh:: ((with a good posture)) (.) uhm:: ((both hands gesture
   3.    pretending to make a circle: iconic))
 → 4.    {Dream:::
   5.    {((keep expressing the circle with both of her hands: iconic))
   6. N: ((nodding repeatedly))
 → 7. S: DREAM:: ((expressing the circle with both her hands: iconic))
   8. N: ((nodding repeatedly))
 → 9. S: land::?((expressing the circle in her both hands: iconic))
 →10. N: dream land. right.
 →11. S: ((pointing at herself with her left finger three times: deictic and beat))
 →12.    like {like
 →13.         {((pointing at herself with her left finger again: deictic))
  14. N: ok ok. you.
 →15. S: (.) ((pointing downward with both hands: deictic)) etto::(let me see)
 →16.        ((pointing downward with both hands: deictic)) no-not?
  17. N: ((nodding repeatedly))
  18. S: ((touching her face with both hands))(.) no-not
 →19.    {dream::?
 →20.    ((expressing a low place in both hands: deictic))
  21.    {$haha$
 →22.    {((pointing downward again but more emphatically: deictic))
 →23. N: here?:: ((downward with both hands: deictic))=
 →24. S: =((opening both hands and waving hands over her head widely: iconic))
 →25. N: it's not dream? ((pointing downward again: deictic))
 →26. S: not dream:[but:::
 →27. N:           [and:: ((moving both hands to his right side: deictic)) but
 →28. S: ((moving both hands to her right side, similar to N: deictic))
  29. N: [Disney is::dream
  30. S: [Disney is (.) dream land $HAHA$
  31. N: ((nodding repeatedly)) all right all right
```

In this excerpt, we can see initially that both interlocutors manage their interaction through the use of repetition and non-verbal actions mixed together with their talk. For instance, we can see that Shizu frequently attempted to tell Naz verbally and also non-verbally that Disney is something like a dream land. Sequential analysis actually shows that she uttered "Dream" in line 4 and uttered it again in line 7, though her speech volume changed from "*Dream*" to "*DREAM*" in order to emphasize her message. Despite her struggle to fill in details, she was able to convey the message with movements of both of her hand (iconic) in line 9. Naz showed his understanding, completed his turn, and set up an "adjacency pair" (Sacks, Schegloff, & Jefferson, 1974) in line 10. In spite of her problems, Shizu did not change the topic but rather continued her turns to make her message more explicit, from line 11, which she likes Disney land. Here, she gradually changed her primary communication resource into non-verbal actions. For instance, from lines 11 to 13, she uttered only "like" as a linguistic resource but rather expressed her intended meaning through the use of non-verbal actions (deictic) (Figure 1-a). Interestingly, she re-expressed the same actions in line 13 to make the meaning more explicit before giving the floor to Naz. She continued to engage in

interaction through the means of hand and body movements to fill in details while also expressing herself verbally, such as in lines 15 and 16 and in lines 19 and 20 (deictic) (Figure 1-b).



(a) lines:11-13 (b) lines:19-20 (c) lines:22-28(1)



(d) lines:22-28(2) (e) lines: 22-28(3) (f) lines:22-28(4)

Figure 1. Examples of body movements including gestures

What is noteworthy here is that Naz also employed non-verbal actions and mimicked her gesture as a listener, and they each initiated negotiation for co-constructing meanings from lines 23 to 28 (deictic and iconic) (see Figure 1-c to Figure 1-f). Their turn-taking through non-verbal actions, such as between 22 and 23, and between 27 and 28, show their engagement leading to a successful communication outcome. Through this style of interaction, they overcame the difficulties, were able to share meaning and achieved mutual understanding by Naz's turn completion (line 31). This excerpt illustrates ELF speakers' collaborative turn sequences (e.g. Kaur, 2011; Seidlhofer, 2001) and emphasizes the important functions of non-verbal actions to enhance successful interactive construction of meanings in ELF contexts (e.g., Kaur, 2011; Ke & Cahyani, 2014).

5 Conclusion

The primary aim of this study is to explicate informal paired ELF interactions, the analysis of which shows that ELF speakers exhibit mutual understanding through the use of multimodal resources that are both verbal and non-verbal. Based on the sequential analysis and retrospective stimulated recall tasks, it was found that the participants employed non-verbal actions, though these were often in combination with verbal strategies, especially repetition. To be more specific, we can say clearly that they employed non-verbal actions as covering for something that could not be expressed verbally; moreover, they each initiated negotiation for co-constructing meanings collaboratively. In so doing, they overcame the difficulty and were able to share meaning and achieve mutual understanding. These examples clearly show ELF speakers' collaborative turn sequences (e.g. Kaur, 2011; Seidlhofer, 2001) and emphasize the important functions of non-verbal actions to enhance successful interactive construction of meaning in ELF contexts (e.g., Kaur, 2011; Ke & Cahyani, 2014). Given the fact that understanding is not taken for granted, and when a particular procedure does not make an interaction sequence smooth, both the speaker and the listener have to employ alternative means through the use of interactional communication strategies in order to co-construct meanings to come to mutual understanding. In other words, ELF interactions would seem to be relevant to developing and enhancing interactional pragmatic features in making the effort to understand between/among the participants in mutual understanding, specifically through the use of interactional multimodal communication strategies. According to Leeuwen (2005), multimodality is the combination of diverse semiotic modes in a communication event. Many studies (e.g., Kaur, 2011; Ke & Cahyani, 2014) emphasize the important functions of non-verbal actions to enhance successful interactive construction of

meaning in ELF contexts. The present study not only indicates the necessity for further research, but also suggests ways in which the multimodal framework can be used as a new teaching methodology.

References

Canagarajah, S. (2013). *Translingual practice: Global Englishes and cospolitan relations.* Abingdon, UK: Routledge.

Cogo, A. (2012). English as a lingua franca: Concepts, use, and implications. *ELF Journal, 66*, 97-105. doi:10.1093/elt/ccr069

Denzin, N. K. (1978). *The research act: An introduction to sociological methods*. New York: McGraw-Hill.

Deterding, D., & Kirkpatrick, A. (2006). Emerging South-East Asian Englishes and intelligibility. *World Englishes, 25*, 391-409. doi:10.1111/j.1467-971X.2006.00478.x

Goodwin, C. (2003). Pointing as situated practice. In S. Kita (Ed.), *Pointing: Where language, culture and cognition meet* (pp. 217-241). Mahwah, NJ: Lawrence Erlbaum Associates.

Gullberg, M. (1998). *Gestures as a communication strategy in second language discourse.* Lund, Sweden: Lund University Press.

Jefferson, G. (1984). On the organization of laughter in talk about troubles. In J. M. Atkinson & J. Heritage (Eds), *Structures of social action: Studies in conversation analysis* (pp. 346-369). Cambridge: Cambridge University.

Jenkins, J. (2000). *The phonology of English as an international language*. Oxford: Oxford University Press

Jenkins, J. (2006). Current perspectives on teaching world Englishes and English as a lingua franca. *TESOL Quarterly, 40*, 157-181. doi:10.2307/40264515

Kaur, J. (2011). Raising explicitness through self-repair in English as a lingua franca. *Journal of Pragmatics, 43*, 2704-2715. doi:10.1016/j.pragma.2011.04.012

Ke, I. & Cahyani, H. (2014). Learning to become users of English as a lingua franca (ELF): How ELF online communication affects Taiwanese learners' beliefs of English. *System, 46*, 28-38. doi:10.1016/j.system.2014.07.008

Kendon, A. (2004). *Gesture: Visible action as utterance.* Cambridge, UK: Cambridge University Press.

Konakahara, M. (2016). A conversation analytic approach to ELF communication: Incorporating embodied action in the analysis of interactional achievement. *Waseda Working Paper in ELF, 6*, 78-96.

Leeuwen, T. T. (2005). *Introducing social semiotics.* London: Routledge.

Matsumoto, Y. (2015). *Multimodal communicative strategies for resolving miscommunication in multilingual writing classrooms* (Unpublished doctoral dissertation). The Pennsylvania State University, Pennsylvania.

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. Chicago, IL: University of Chicago Press.

McNeill, D. (2005). *Gesture and thought.* Chicago, IL: University of Chicago Press.

Pitzl, M.-L. (2005). Non-understanding in English as a lingua franca: Examples from a business context. *Vienna English Working Papers, 14*(2), 50-71. Retrieved from https://anglistik.univie.ac.at/fileadmin/user_upload/dep_anglist/weitere_Uploads/Views/Views0502ALL_new.pdf

Sacks, H., & Schegloff, E., & Jefferson, G. (1974). A simplest semantics for the organization of turn-taking for conversation. *Language, 50*, 696-735. doi:10.2307/412243

Schegloff, E. (2000). When 'others' initiate repair. *Applied Linguistics, 21*, 205-243. doi:10.1093/applin/21.2.205

Schegloff, E., Koshik, I., Jacoby, S., & Olsher, D. (2002). Conversation analysis and applied linguistics. *Annual Review of Applied Linguistics, 22*, 3-31. doi:10.1017/S0267190514000014

Seidlhofer, B. (2001). Closing a conceptual gap: The case for a description of English as a lingua franca. *International Journal of Applied Linguistics, 11*, 133-158. doi:10.1111/1473-4192.00011

Smit, U. (2000). *English as a lingua franca in higher education: A longitudinal study of classroom discourse.* Berlin: Mouton de Gruyter.

Stein, P., & Newfield, D. (2006). Multiliteracies and multimodality in English in education in Africa: Mapping the terrain. *English Studies in Africa, 49*(1), 1-21.

Swales, J. (1990). *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press.

Appendix. Transcription conventions

(adapted from Jefferson, 1984 & McNeill, 2005)

[	overlapping utterances	=	latched utterances
{	overlapping utterances with non-verbal action	(.)	short pause of less than 1 second
:	extended sound or syllable	.	fall in intonation
-	cut-off by current speaker	Dream	emphasis
DREAM	speech much louder than surrounding talk	$	smile voice
(( ))	non-verbal action	→	feature of interest to analysis

Under load: The effect of verbal and motoric cognitive load on gesture production

Marieke Hoetjes[a], *Ingrid Masson-Carro*[b]

[a]Centre for Language Studies, Radboud University, Nijmegen, The Netherlands
[b]Tilburg center for Cognition and Communication, Tilburg University, The Netherlands

[a]P.O. Box 9103, 6500 HD Nijmegen, The Netherlands

[b]Warandelaan 2, PO Box 90153, 5000 LE Tilburg, The Netherlands

M.Hoetjes@let.ru.nl, I.MassonCarro@uvt.nl

Abstract

It has been hypothesized that speech and gesture work together, with people relying more on gesture when speaking is hard, and relying more on speech when gesturing is hard. Indeed, previous work showed that when speaking is hard, gesture production can reduce cognitive load and thereby help the speech process. However, it is yet unknown what happens in speech and gesture production when gesturing is hard. In the current study, participants described complex tangram figures. Difficulty in either speech or gesture production was manipulated by having speakers do a secondary task which placed them under either verbal or motoric cognitive load. Results showed an increase in representational gesture rate when participants were under motoric load, as compared to the baseline condition, but only a weak effect of verbal load, and no difference between the verbal and motoric load conditions. Based on these findings, we conclude that making speaking hard caused a marginal increase in gesture production, and making gesturing hard actually led to more gestures. In sum, we find no evidence to support a two-way trade-off between speech and gestures. To our knowledge, this is the first study assessing the effects of a secondary motoric task on gesture production.

1 Introduction

Human communication is multimodal, with speakers typically using both speech and gesture to express an intended meaning. Gesture and speech tend to be temporally and semantically co-expressive (Kendon, 2004; McNeill, 1992). This close relationship between speech and gesture is apparent, for example in the gesture production by congenitally blind speakers who have never seen someone gesture (Iverson & Goldin-Meadow, 1998), in the parallel development of speech and gesture production in children (Gullberg, De Bot, & Volterra, 2008), and in the parallel breakdown in cases of disfluency (Seyfeddinipur, 2006). Although the close relationship between speech and gesture is undisputed, the question of exactly why we gesture when we speak is still under discussion. One view is that (some) gestures are produced to help the listener (Alibali, Heath, & Myers, 2001). Another, not mutually exclusive, suggestion is that producing gestures is done for the speaker herself, as it helps reduce cognitive load, which is needed for speech production.

Several studies have addressed the possible relation between gestures and cognitive load. For example, Goldin-Meadow, Nusbaum, Kelly and Wagner (2001) asked participants to explain how to solve math problems while doing a mentally taxing secondary task (remembering lists of words or letters). They showed that the participants who were allowed to gesture while they were explaining the math problem were better at the secondary task, suggesting that producing gestures frees up mental resources. Melinger and Kita (2007) also gave their participants a secondary task, in their case while describing spatial pictures they had previously memorized. However, Melinger and Kita used two types of secondary tasks: one spatial task for which the same resources as in the primary task had to be used, and one non-spatial task for which other mental resources are required. Speakers produced more gestures during picture description when they had to do a spatial

secondary task compared to a non-spatial secondary task. Again, this suggests that gesture production can help reduce cognitive load, in particular when the secondary task requires the same mental resources as the primary task.

The idea that producing gesture can offload mental space required for speaking has also been suggested by De Ruiter, Bangerter and Dings (2012) in their trade-off hypothesis, which states that when speaking becomes harder, speakers will rely more on gestures, and when gesturing becomes harder, speakers will rely more on speech. De Ruiter and colleagues tested one part of the trade-off hypothesis, namely whether speakers rely more on gesture when speaking becomes harder. They conducted a study in which participants had to describe simple and complex tangram figures repeatedly, and found that gesture rate was not higher for the initial descriptions and complex figures (i.e. when speaking was hard) as compared to the repeated descriptions and simple figures (i.e. when speaking was relatively easy). Instead, gesture production mirrored speech production across experimental conditions, thereby supporting the hypothesis that gestures and speech go hand in hand. Their study, however, targeted the difficulty of verbal referring separate from memory resources—which were kept constant across experimental conditions by having the speaker be able to see the stimuli throughout the task. Given the previously found effects of memory load on gestures (e.g., Melinger & Kita, 2007), it is likely that we would find an effect on the amount of gestures produced when verbal memory resources are taxed, which is another way of making speaking harder. Furthermore, De Ruiter and colleagues did not test the second part of their hypothesis which proposes that speakers will rely more on speech when gesturing gets harder.

In the present study, we designed an experiment to test both ends of the abovementioned trade-off hypothesis. That is, we will make either speaking harder or gesturing harder, by putting participants under one of two types of cognitive load: verbal memory load or motoric memory load. Verbal load will make speaking harder, and, assuming that gestures are generated from action processes (e.g. Kita & Özyürek, 2003) and require spatio-motoric working memory, motoric load will make gesturing harder. As in Goldin-Meadow, et al. (2001) and in Melinger and Kita (2007), cognitive load is imposed in a secondary working memory task. The primary task is for participants to describe complex spatial figures. Describing complex spatial figures requires both verbal and motoric resources. This means that both types of secondary tasks require part of the mental resources also needed for the primary task. The hypothesis is that in cases of verbal load, speakers will rely more on gesture (and thus produce more gestures in the primary task), and in cases of motoric load, speakers will rely more on speech (and thus produce fewer gestures in the primary task), relative to a baseline condition without a concurrent secondary task.

2 Methods

Pairs of participants engaged in a director-matcher task. In the first part of the experiment, one participant (the director) described a series of tangram figures to her interlocutor (the matcher), who had to locate and mark these objects on a visual grid. In the second part, participants exchanged roles, with the interlocutor becoming the director and describing a new set of figures. Each pair of participants accomplished the task under one of three conditions: verbal load, motoric load, or baseline.

2.1 Participants

72 students from Tilburg University—46 female and 26 male, M = 21 years—took part in the experiment in pairs, in exchange for partial course credit. All participants read and signed an ethics consent form prior to the commencement of the task, and were informed that they could withdraw their participation anytime.

2.2 Stimuli

Two different sets of ten abstract geometric figures were digitally created, inspired by the Chinese game of tangram (see Figure 1). Each set of figures was compiled into a presentation document, where each figure fully occupied one slide. Version A featured figures 1 to 10, and version B featured figures 11 to 20. Each participant was presented with either version A or version B (see Table 1 below).

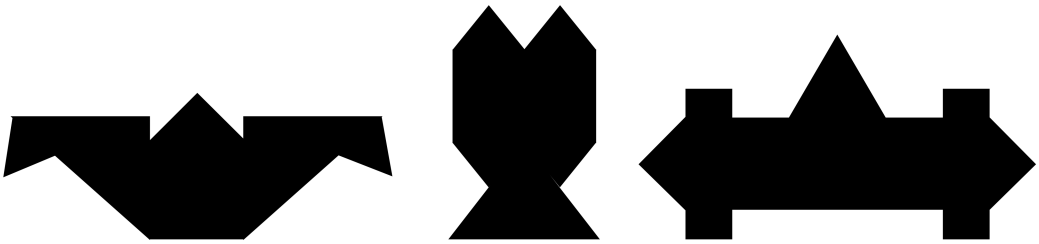


Figure 1. Example of three target tangram stimuli figures.

We induced load in the speakers by means of a concurrent working memory (WM) task that tapped on either verbal or spatio-motoric working memory. In the verbal condition, each tangram figure was preceded by a slide with a word on it (either *dal* or *bal*, the Dutch words for *valley* and *ball*) and followed by a slide instructing the speaker to reproduce the word she had read before. In the motoric condition, each tangram figure was preceded by a short video where an actor performed the British Sign Language sign for either "green" or "brown" (two visually similar signs for hearing speakers with no sign language knowledge) and followed by a slide instructing the speaker to reproduce the movement she had seen before. In the baseline condition, all slides consisted only of tangram figures (see Figure 2).

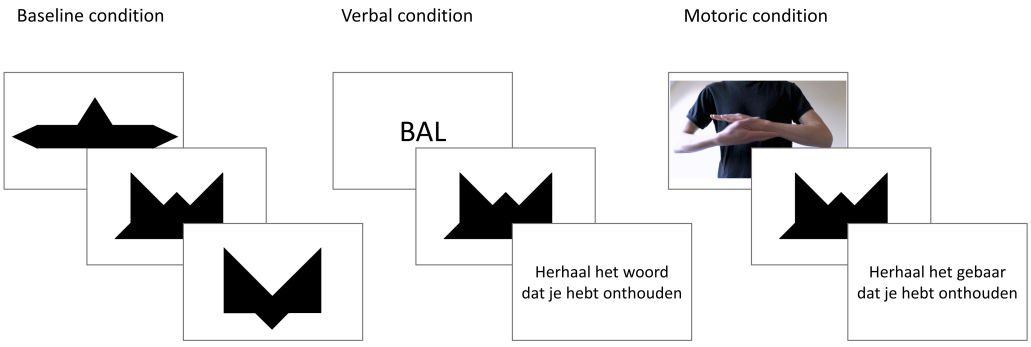


Figure 2. Summary of the three experimental conditions: baseline, verbal, and motoric. The baseline condition shows 3 example trials, and the verbal and motoric load conditions show one example trial with their respective WM tasks.

2.3 Procedure

Each pair of participants was assigned to one of the experimental conditions: verbal, motoric, or baseline (see Table 1). Participants were assigned the roles of director and matcher, and sat at opposite sides of a table. The setup was arranged in such way that there was visual contact between director and matcher, but the matcher could not see the director's screen displaying the task, and the director could not see the matcher's visual grid with all the numbered tangrams. A camera recorded the director's speech and upper bodily movements.

The experiment started with two practice trials, followed by ten target trials. Each trial corresponded to the description of one tangram figure. The director's task was to describe to the matcher each tangram figure displayed on the laptop screen. The matcher's task was to find the tangram figure that was being described on a printed visual grid displaying 12 tangrams, and to write down its corresponding number on an answer sheet. During the description phase, the experimenter remotely controlled the presentation, allowing speakers to describe the figures at their own pace, and moving to the next trial only when the matcher finished writing his answer down. Conversation was not restricted between director and matcher, but it was not encouraged either, and the use of gestures was not mentioned.

In the baseline condition, all ten target figures were consecutively described. In the load conditions, a stimulus (either a word or a video showing a series of movements) was presented prior to each tangram figure, remaining on screen for 3 seconds. The director was instructed to look at the stimulus and memorize it, as she would have to reproduce it (orally, or motorically) after describing the tangram figure (see Figure 2).

After the director had described the ten target figures, participants switched roles, and the previous matcher became the director, describing a new set of ten figures under the same experimental condition.

Table 1. Administration of the experimental conditions

PAIR	SPEAKER 1	SPEAKER 2
Pair 1	Baseline version A	Baseline version B
Pair 2	Verbal version A	Verbal version B
Pair 3	Motoric version A	Motoric version B
Pair 4	Baseline version B	Baseline version A
etc...	...	...

2.4 Data annotation

Speech was transcribed verbatim per trial by a native Dutch speaker, using the multimodal annotation tool ELAN (Sloetjes & Wittenburg, 2008). All hand gestures accompanying speech during the description of the tangram figures were identified and classified by two independent coders as representational or non-representational (McNeill, 1992). Adaptors (e.g., touching one's hair) and other irregular movements were excluded from the annotations. The representational, and non-representational gesture rates (number of gestures in proportion to 100 words spoken) were computed. Both coders annotated videos from all conditions, with an overlap of ten videos used to check for the inter-rater reliability of gesture identification. Cohen's κ revealed substantial agreement between coders with respect to the number of gestures produced by speakers ($\kappa = .84$, $p < .001$).

2.5 Statistical analyses

We used linear mixed models (Barr, Levy, Scheepers, & Tily, 2013) to analyze the effects of cognitive load (verbal, motoric, baseline), on our dependent variables (representational and non-representational gesture rate, and the number of words spoken). Participants and items (tangram figures) were included as random factors in the analyses.

3 Results

The task generated a total of 2163 representational and 209 non-representational gestures. We eliminated from our analyses data from 9 participants (three per condition) who produced less than one gesture in total. This resulted in the analysis of 2162 representational and 209 non-representational gestures.

We found an effect of motoric load on the representational gesture rate (β = 2.98, SE =1.25, p = .02), indicating that speakers in the motoric load condition (M = 8.61, SD = 6.18) gestured more than speakers in the baseline condition (M = 5.58, SD = 5.04) (see Fig. 3). A similar effect was observed for speakers under verbal load (M = 7.82, SD = 6.1), albeit statistically weak (β = 2.21, SE = 1.22, p = .07). No differences were found between speakers in the motoric and verbal load conditions, as can be seen in Figure 3. No effects of either motoric (M = .87, SD = 2.7) (β = .11, SE = .28, p = .69) or verbal load (M = .64, SD = 1.79) (β = .06, SE = .27, p = .81) were found for non-representational gesture rate, in comparison with the baseline condition (M = .75, SD = 1.72).

Moreover, there was an effect of motoric load on the average number of words spoken (see Figure 3). Participants in the motoric load condition (M = 35.62, SD = 18.44) used fewer words on average than participants in the baseline condition (M = 45.87, SD = 31.14) (β = -10.11, SE = 4.12, p = .01). No effects of verbal load were found (M = 44.12, SD = 25.57) (β = -1.18, SE = 4.02, p = .76).

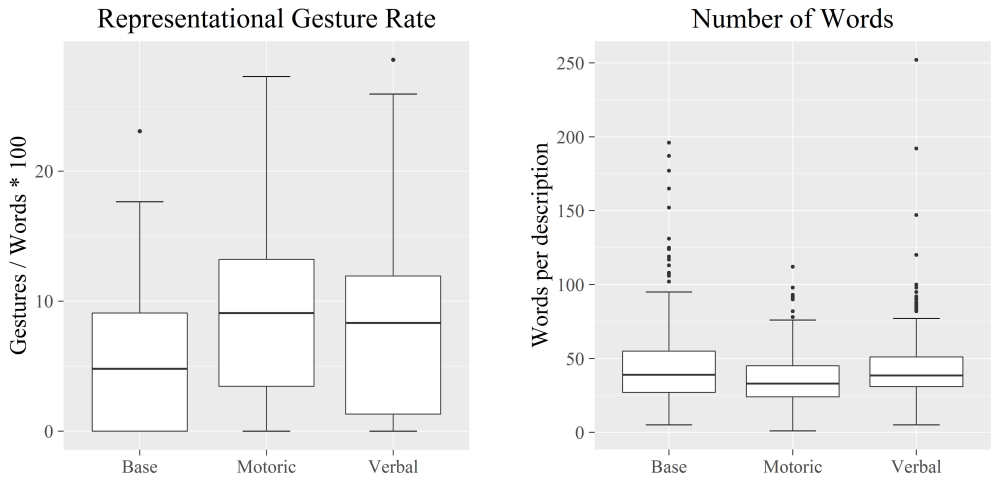


Figure 3. Box plots of the speaker's gesture rate (left) and number of words per description (right), for each of the experimental conditions (baseline, motoric load, and verbal load).

4 Discussion

This study set out to investigate the effect of verbal and motoric cognitive load on gesture production. The aim was to address the question whether gesture production helps reduce cognitive

load, and in particular whether, as the trade-off hypothesis (de Ruiter, et al., 2012) suggests, people rely more on gesture when speaking is hard, and rely more on speech when gesturing is hard. We created difficulty in speaking or in gesturing by having participants conduct a secondary task, concurrent to the primary task of describing complex spatial tangram figures. In the secondary task, participants had to either remember one of two similar words (verbal load), or one of two similar hand configurations (motoric load). The hypothesis was that speakers would produce more gestures in the primary task if they were under verbal cognitive load—compared to the baseline condition—and fewer gestures in the primary task if they were under motoric cognitive load.

The results showed an increase in representational gesture production when participants were under load, relative to the baseline condition. This effect was prominent for motoric load, but weak for verbal load. Importantly, there was no difference between the two load conditions. Although these findings marginally support one end of the trade-off hypothesis (namely that speakers produce more gestures when under verbal load), they do not support the assumption that tapping into different types of working memory processes causes differences in gesture production. It seems that being under load, because of having to carry out a secondary task, generally led to more gestures.

The fact that there was no difference in gesture production between the verbal and the motoric load conditions can be viewed in several ways. Firstly, it might be the case that, although being under load affects gestures, the type of cognitive load simply does not matter for gesture production. Another possibility is that our manipulation of cognitive load was not successful, in the sense that the secondary tasks participants had to do did not actually tap into *different* types of cognitive load. After all, it can be argued that having to remember words shown on a screen does not have to be a verbal task per se, but could also be a more general memory task. Therefore, in future studies, the material used to induce verbal load could be presented orally instead of visually. A third possibility to explain the findings is that both our verbal and motoric load conditions led to more gesturing *for different reasons*. That is, it could be that having speakers keep words in mind while producing verbal descriptions pushed them to rely more on their hands, but that keeping hand configurations in mind while talking simply "boosted" gesture production. This would be compatible with simulation-based accounts of gesture production (e.g., Hostetter & Alibali, 2008), which posit that representational gestures arise from perceptual and motoric simulations underlying thinking and speaking. In this regard, it is possible that maintaining a movement sequence active while performing a side task just increased the amount of motor activation experienced, leading to an increase in the production of overt gestures.

A point of discussion could be whether both types of secondary tasks were equally difficult for the speakers. The tasks were chosen because they are both deceptively easy; although the words and signs themselves were simple, the similarity between the words or signs made the task quite hard. One possible way to study this would be by looking at the accuracy of the answers given in the secondary tasks. The question then is, though, what information we can glean from this. Would an incorrect answer in the secondary task mean that the speaker was not paying attention, or that the speaker was paying attention but the task was harder?

It's important to note that the results found applied only to representational gestures, and were not mirrored by either non-representational gestures, or speech. For speech, we found that fewer words were used when participants were under motoric load, compared to the baseline condition. It is likely that the motoric task more strongly activated spatio-motoric thinking, resulting in more information expressed through hand gestures and less information expressed through speech. Further analyses on the semantic content of the speech and gestures produced are needed to test this idea.

In conclusion, the present study has provided additional evidence, in line with Goldin-Meadow, et al. (2001) and Melinger and Kita (2007) that speakers produce more gestures when their memory

resources are taxed. Furthermore, we proposed the use of a novel task to study spatio-motoric resources during multimodal language production.

Acknowledgements

The research reported in this article was financially supported by The Netherlands Organisation

for Scientific Research (NWO) [grant 322-89-010]. We thank Ivo de Ruiter and Chiara de Jong for their invaluable help in collecting and annotating parts of the video dataset.

References

Alibali, M., Heath, D. C., & Myers, H. J. (2001). Effects of visibility between speaker and listener on gesture production: Some gestures are meant to be seen. *Journal of Memory and Language*, 44, 169–188.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68, 255-278.

de Ruiter, J. P., Bangerter, A., & Dings, P. (2012). The interplay between gesture and speech in the production of referring expressions: Investigating the tradeoff hypothesis. *Topics in Cognitive Science*, 4(2), 232–248.

Goldin-Meadow, S., Nusbaum, H., Kelly, S. D., & Wagner, S. (2001). Explaining math: Gesturing lightens the load. *Psychological Science*, 12, 516-522.

Gullberg, M., De Bot, K., & Volterra, V. (2008). Gestures and some key issues in the study of language development. *Gesture*, 8(2), 149-179.

Hostetter, A. B., & Alibali, M. W. (2008). Visible embodiment: Gestures as simulated action. *Psychonomic Bulletin and Review*, 15, 495–514.

Iverson, J., M., & Goldin-Meadow, S. (1998). Why people gesture when they speak. *Nature*, 396, 228.

Kendon, A. (2004). *Gesture. Visible action as utterance*. Cambridge: Cambridge University Press.

Kita, S., & Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal?: Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and Language*, 48, 16-32.

McNeill, D. (1992). *Hand and mind. What gestures reveal about thought*. Chicago: University of Chicago Press.

Melinger, A., & Kita, S. (2007). Conceptualisation load triggers gesture production. *Language and Cognitive Processes*, 22(4), 473-500.

Seyfeddinipur, M. (2006). *Disfluency: Interrupting speech and gesture*. Radboud University Nijmegen.

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)*.

Individual variation in gestural markers of uncertainty

Anna Jelec[a]
Małgorzata Fabiszak[a]
Anna Weronika Brzezińska[b]

[a]Faculty of English, Adam Mickiewicz University in Poznań, Poland
[b]Institute of Ethnology and Cultural Anthropology, Adam Mickiewicz University in Poznań, Poland

[a]Al. Niepodległości 4, 61-786 Poznań
[b]ul. Umultowska 89 D, 61-614 Poznań

jelec@amu.edu.pl

Abstract

Recent historical conflicts, some of which are still fresh in the memory of Poznań's citizens, revolved around four ethnic/national groups: Poles, Germans, Soviets (Russians) and Jews. As many moral issues related to the country's past remain unsolved, fractures have formed in the collective memory of the inhabitants of Poznań, represented by uncertainty around the meaning of certain concepts and events. This study investigates the expression of this uncertainty through epistemic gestures such as shoulder-shrugs, head tilts and Palm-Up-Open-Hand gestures (Debras and Cienki 2012, Mueller 2004, Streeck 2009). We consider language as inherently multi-modal: during interactions, interlocutors construct meaning of verbal and non-verbal cues. Co-speech gesture adds or emphasizes information expressed in speech, and can be interpreted as a signal for the existence of focus points in discourse, guiding further analysis. In this study, we analyse data from a series of ethnographic interviews where the inhabitants of Poznań discuss the history of various ethnical groups and their impact on the city. We investigate whether speakers have a preference for a particular gesture form when expressing uncertainty, and explore the relation of gesture form with function. Quantitative analysis of the speakers' preference for gesture form is followed by qualitative analysis answering the following questions: (1) Whether there was a relationship between the occurrence of combination and simple gesture form; and (2) If combinations of gesture had been used for emphasis. We conclude that the present research sample does indicate certain interesting tendencies in the use of epistemic gestures of uncertainty, most notably speaker preferences for gesture form, and correlations between the use of simple and complex gesture. However, further research is required to verify these findings.

1 Introduction

Everything we express, we express with a degree of certainty about the truth value of the picture painted through our utterance or the degree of consensus over the knowledge about to be shared. For the purpose of this paper, we define uncertainty as the degree of certainty of a person with respect to what s/he is saying. Speakers who are uncertain about what they want to say have a variety of ways to express that uncertainty in conversation. Not all of these methods are verbal. The degree of certainty in what we are saying can be expressed through words, tone of voice, intonation, hesitations (cf. Dral et al 2011), facial expressions and gestures (shrugging, hand gestures, head movements). In fact, although lexical choice is important for conveying degrees of certainty (by the use of epistemic adverbials such as *surely*, *probably*, or *perhaps*, or hedges containing mental verbs like *you know, I think*) Chafe (1986), Nuyts (2001) Konat (2016), lexical meaning can be supplemented with or even overridden by gesture (Krahmer and Swerts 2005, Roseano et.al. 2015). Studies show that we express uncertainty through several gestures, most consistently the raised shoulder(s), lateral head tilt(s) (Debras and Cienki 2012, Streeck 2009), the Palm Up Open Hand(s) gesture (PUOH) (Müller 2004) or a combination thereof. Hence, this research focuses on the visual expression of uncertainty, in particular co-speech gesture of the hands, arms and head. We discuss whether speakers have a preference for a gesture form to express epistemic uncertainty and if they show variation in gesture; in

particular we investigate the occurrence of simple and combination gesture forms in micronarratives (short narrative stories produced by one or more speakers) that touch upon the topic of nationality in relation to local history.

2 Methods of data acquisition, annotation and analysis

2.1 Data acquisition

Our data come from 16 ethnographic focus interviews conducted for the project on collective memory and collective identity in Poznań from which we selected fragments concerning four nationalities (Polish, Jewish, German, Soviet/Russian) that inhabited Poznań and/or left their mark on the city landscape. This choice is dictated by the nature of the main project, in which the collective memory of the city inhabitants was investigated. The focus was on how the four national and/or religious groups are remembered and how this memory is communicated. The interest in ways of expressing uncertainty is an offshoot of this bigger project in which we show that gestural markers of uncertainty appear more frequently in the mini-narratives concerning the Jews than those about the other groups (Fabiszak and Jelec in preparation). The respondents of the interviews come from four generations: born in the 1930s (11F+4M), 1950s (12F+4M), 1970s (8F+5M) and 1990s (9F+20M). These social variables, important for the project on collective memory, are however, not taken into account in the present study as we are more interested in individual variation of two selected speakers whose contributions were the longest and contained the largest number and variety of gestures.

2.2 Data annotation and coding

We analysed fragments of video recordings where the topic was nationality, i.e. where the participants talked about people from one of the four national/religious groups: Polish, German, Russian/Soviet, Jewish. Based on the transcript of the verbal exchanges between the participants, as well as between the participants and the interviewer, we divided the recordings into short stories. Each fragment, called a micronarrative, focused on a chain of events related to a particular ethnic group. As a result, we ended up with 175 relevant video fragments, ranging from 00:04 to 01:40 in duration. Next, the videos were annotated for gesture in ELAN (Wittenburg et al. 2006) by two annotators trained to recognise co-speech gesture based on the annotative practice published and revised by McNeill Gesture Lab (McNeill 2005). Each annotator worked with videos she had not seen previously; videos were coded by one annotator and reviewed by the other. Disagreements were discussed and resolved. Video fragments were coded for three types of epistemic gesture: raised shoulders (shrug), lateral head tilt, palms up open hand (PUOH) or a combination of these. The annotators viewed each clip three times: first at 80% speed (mute) to identify rough locations of the relevant gesture; second at 20% speed (mute) to fine-tune the selection; and finally at 80% speed with voice on in order to weed out potential occurrences of deictic tilts and shrugs, as well as other irrelevant gestures. Finally, the data from speech and gesture was integrated and input in a summary excel table. The table identified the ethnic group mentioned in the fragment (Polish, German, Russian/Soviet or Jewish, the code number of the speaker, the co-occurring utterance, the co-ocurring movement (if any) and the type of movement (symmetrical or asymmetrical shrug, lateral head tilt, PUOH or a combination thereof). In order to focus on co-speech gestures with identifiable meaning we have decided to exclude from the analysis any gesture performed in the absence of speech, or by someone other than the active speaker.

2.3 Data analysis

In this sample, 36 speakers perform 342 gestures of interest, i.e. classified as epistemic gestures of uncertainty according to the procedure outlined in section 2.2. The most frequent form was raised shoulders (N=124), followed by PUOH (N=88) and a combination of head tilt and raised shoulders (N=65). The frequency of different gesture forms is given in Table 1.

Table 1. Frequency of epistemic gesture forms in the sample.

Gesture form	Raw frequency
SHOULDER	124
PUOH	88
HEAD_SHOULDER	65
HEAD	36
SHOULDER_PUOH	28
HEAD_PUOH	1
TOTAL	**342**

3 Quantitative Analysis

3.1 Speakers' use of gesture form

The aim of this study was to explore the variation in the use of different gesture forms expressing uncertainty by particular speakers, asking whether speakers show a preference for a particular gesture form. Therefore, for further analysis we have selected only those speakers that performed at least 4 gestures in the data sample. This reduced the number of speakers from 36 to 20. The results of Speaker by Gesture form analysis are presented in Table 2.

Table 2. Frequency distribution of Speaker and Gesture form. Highest frequencies for a given speaker are given in dark yellow, medium frequencies in pale yellow, low frequencies in white.

Speaker	HEAD	H_PUOH	HEAD_SH	PUOH	SHOULD	SH_PUOH	TOTAL
R16	1				3		**4**
R18	1		2		1		**4**
R06				4			**4**
R27				5			**5**
R34	1			4			**5**
R12	1				4	1	**6**
R19	2		2	1	1		**6**
R25			2		4		**6**
R08				2	2	2	**6**
R10				4	4	1	**9**
R30			5	3	1		**9**
R44			2	2		7	**11**
R14	5		3	3	5		**16**
R03	3		1	2	11		**17**
R07		1		13	1	2	**17**
R23			7	1	16		**24**
R11	2		1	9	7	7	**26**
R05	3		5		24		**32**
R32	6		11	16	1		**34**
R13	7		16	13	30	7	**73**
TOTAL	**32**	**1**	**57**	**82**	**115**	**27**	**314**

3.2 Speakers' preference for gesture form

As seen in Table 2, two speakers use a single gesture form, 9 show a clear preference for one gesture form, even though they use more than just one form, and 9 speakers systematically use several different gesture forms.

It needs to be noted that data was collected from ethnographic interviews, so speaker turns were of different length. This results in significant inter-speaker variation in the number of produced gestures. The speakers with a clear preference tend to cluster at the top of the table, i.e. among those who produced the smallest number of gestures. It could be argued that if their contributions to the interview were longer, they would have used a larger repertoire of gestures. Yet, speakers R44, R03 and R07,

who produced a relatively high number of gestures (N>10), still show a preference for one gesture form.

Speakers R11 and R13 seem to be the most versatile gesturers, employing 5 different gesture forms throughout their respective turns. R11 used PUOH most often (N=9), raised shoulders and a combination of raised shoulders with PUOH 7 times each. She rarely used head tilts (N=2) or head tilts in combination with raised shoulders (N=1). R13, on the other hand, seemed to have a preference for raised shoulders (N=30) and raised shoulders in combination with head tilt (N=16). The comparison of the gestural styles of just these two speakers is given in Table 3 below. The gestural style (GS) is understood here as a relative frequency of particular gesture forms (N) to the overall number (TOTAL) of gestures performed by a given speaker (GS=N/TOTAL).

Table 3. Speakers' gestural styles: frequency of gesture forms relative to the overall number of gestures. The highest frequency is given in dark yellow, medium frequency in pale yellow, low frequency in white.

Speaker	HEAD	HEAD_SH	PUOH	SHOULD	SH_PUOH
R11	0.077	0.038	0.346	0.269	0.269
R13	0.096	0.219	0.178	0.410	0.096

4 Qualitative analysis: Gesture combinations vs. simple gestures

R11 and R12, as proficient gesturers, allow for further qualitative analysis, in which we investigate if gesture combinations behave differently from simple gestures. In particular, we ask the following questions: (1) Were combination of forms more likely to appear medially in the narrative, following simple gestures?; (2) Were combinations of forms used for emphasis?

Table 4 below is a transcript of a fragment of a focus group interview concerning the Soviet soldiers' war cemetery in Poznań. Speaker R11 talks about her volunteer work in the cemetery, cleaning the graves. Here she re-enacts her conversation with the organizer of the cleaning. Lines 1-2 are reported speech, lines 3-10 are the speaker's own words.

Table 4. Transcript fragment: Gesture combination as intensity marker. Speaker R11. Gesture combinations are marked in red, simple gestures in blue.

	Sense unit	**Gesture**
1	mam groby radzieckie do grabienia do grabienia [I have Russian graves for raking for raking]	NO_MOVEMENT
2	Pójdzie pani? [Will you go?]	NO_MOVEMENT
3	Ja mówię [I'm saying]	HEAD
4	jak tam nie ma Stalina [If Stalin isn't there]	HEAD_SHOULDER
5	to mogę iść [I can go]	NO_MOVEMENT
6	I idą ludzie [And people go]	SHOULDER_PUOH
7	idą [They go]	NO_MOVEMENT
8	a niektórzy „nie, ja nie pójdę" [But some (say) „no, I won't go"]	NO_MOVEMENT
9	jego grobu bym nie nie tego [His grave I wouldn't no I wouldn't]	NO_MOVEMENT
10	a te jeszcze nie [But these not yet]	SHOULDER

In Table 4, gestures seem to concentrate in the middle of the turn, where the story reaches its peak and the speaker makes a statement about (not) cleaning Stalin's grave. The gestures form the following sequence: HEAD > HEAD_SHOULDER > SHOULDER_PUOH > SHOULDER. That is, the two complex gestures appear in two turns: "if Stalin isn't there" (line 4) and "and people go" (line 6), which constitute two twists in the narrative. First, the Speaker is ready to clean the Soviet graves if Stalin's grave is not among them. The second emphasizes that other people are also ready to do the cleaning. The emphatic function is seen in the repetition of "go" in line 7. Thus, complex gestures of uncertainty in this excerpt appear in sites of high intensity, which are mitigated by the hedging potential of the gestural epistemic markers. What's more, within one narrative fragment they follow a simple gesture: HEAD. In 4 out of 5 narrative fragments of this Speaker in which she uses complex

gestures, the complex gesture follows a simple gesture. Only in one fragment two complex gestures appear in two consecutive lines and are the only epistemic markers in this excerpt.

Yet, it seems that the function of the complex epistemic gesture does not necessarily have to be intensifying the uncertainty. It may appear narrative-finally, when the Speaker cedes the turn at the end of the narrative and expresses uncertainty with respect to the causes of the described events. See Table 5 below.

Table 5. Transcript fragment: Gesture combination in the function of ceding the turn. Speaker R11. Gesture combinations are marked in red, simple gestures in blue.

	Sense unit	**Gesture**
1	bo mój ojciec akurat jechał do Katynia [because my father was just then going to Katyń]	NO_MOVEMENT
2	Nie wiadomo dlaczego ten pociąg akurat przejechał [And it is unknown why that train drive through just then]	NO_MOVEMENT
3	i on (=ojciec) nie zost- nie został zastrzelony [and he (=the father) wasn- was not shot]	NO_MOVEMENT
4	Dostali taką dużą książkę Już w 44 chyba roku (=Polacy, którzy przeżyli Katyń) [they (=Poles who survived Katyń) got this big book As soon as probably 1944	NO_MOVEMENT
5	gdzie były wszystkie nazwiska [all the names were there]	PUOH
6	I on mówi (=ojciec) [and he (=father) says]	NO_MOVEMENT
7	z tym był (=zamordowany w Katyniu) [he was with this (man murdered in Katyń)]	NO_MOVEMENT
8	z tym był, nie (=zamordowany w Katyniu) [he was with this (man murdered in Katyń), right]	NO_MOVEMENT
9	Lekarz przede mną z Ostroroga ten, i on był (=zamordowany w Katyniu) [A doctor from Ostroróg he was there]	NO_MOVEMENT
10	oni zginęli [they died]	PUOH
11	a on (=ojciec) przeżył [and he (=father) survived]	PUOH
12	No więc jak te pociagi jechały [so as these trains were driving]	SHOULDER
13	nie wiem [I don't know]	NO_MOVEMENT
14	No więc wiesz [so you know]	SHOULDER
15	Więc to jest tego [this is kind of]	PUOH_BOTH

Speaker R13 also uses the complex gesture in different places of the narrative: initially, medially and finally. But, contrary to Speaker R11, in 4 cases out of 5 complex gestures precede simple gestures that appear in the same narrative fragments.

It is possible that the order of simple vs. complex gesture forms is an idiosyncratic feature, specific to particular Speakers, but this observation requires confirmation in a study with a larger group of participants. As for the function, apart from emphasizing uncertainty, Speaker R13 uses complex epistemic markers for distancing himself from the politically incorrect label that he uses to describe the former evangelical cemetery. See Table 6 below.

Table 6. Transcript fragment: Gesture combination in a distancing function. Speaker R13.

	Sense unit	**Gesture**
1	tak, to był ewangelicki (=cmentarz) [yes it was Evangelical (cemetery]	HEAD
2	ale np. tak zwany. nomen omen tak go brzydko nazywano [but, for instance, nomen omen, as it was unkindly called]	HEAD_SHOULDER
3	no ale tak go nazywano na Grunwaldzkiej [but this is how it was called in Grunwaldzka (street)]	NO_MOVEMENT
4	ten park nazywano „parkiem sztywnych" za mojej młodości [the park was called "cadaver park" in my youth]	NO_MOVEMENT
5	parkiem sztywnych bo tam właśnie był ten cmentarz Narożnik Grunwaldzkiej/Reymonta czy Przybyszewskiego [cadaver park because that cemetery was there, the corner of Grunwaldzka and Reymonta (street), or Przybyszewskiego (street)]	NO_MOVEMENT

5 Results and discussion

The present paper sets out to investigate how uncertainty is expressed through epistemic gestures such as shoulder-shrugs, head tilts and Palm-Up-Open-Hand gestures (Debras and Cienki 2012, Mueller 2004, Streeck 2009). In the quantitative section of the paper, we investigated whether the speakers have a preference for a particular gesture form when expressing uncertainty. Table 2 demonstrates that speakers indeed show some preference for some gesture forms over others. While in the present analysis it remains unclear if this should be attributed to the limited number of gestures produced by the speakers, this topic would benefit from further research, in particular studies employing controlled research paradigms.

We focused on gestural markers of uncertainty in discourse. Nevertheless, a preliminary analysis on the lexical markers of uncertainty revealed some connections between particular gestural and verbal expressions, for instance PUOH tends to be produced together with approximators (eg. 'trochę' - a little bit, 'jakiś' - some), and shoulder movements occur with mental verbs (e.g. 'nie wiem' - I don't know). Limited data from the current study precludes any broader generalisations. More research is needed to test whether these relations should be explained by individual variation between speakers or ascribed to a more general preference.

From the qualitative analysis we see that complex gestures may appear medially, where the micro-narrative achieves its narrative peak, but they may also be distributed in other positions: initially and finally. While the initial and medial uses seem to have been used for emphasis, the narrative final use combines emphasis with turn-ceding functions. They also may perform a mitigating function, when the speaker is quite certain about what he wants to say, but uncertain about how this will be received by the other participants of the communication (Table 6, line 2). Such multi-functionality of gestures has been observed before (Calbris 2011: 25-30 references needed here).

Acknowledgements

This research has been supported by The National Science Centre, grant number 2013/09/B/HS6 /00374.

References

Calbris, Geneviève. (2011). *Elements of meaning in gesture*. Transl. by Mary Ann Copple. Amsterdam/Philadelphia: Benjamins.

Chafe, W. L. (1986). Evidentiality in English conversation and academic writing. In: *Evidentiality: The linguistic coding of epistemology*, W.L. Chafe and J. Nichols (eds.), 261–272. Norwood, NJ: Ablex.

Debras, Camille & Alan Cienki. (2012). Some Uses of Head Tilts and Shoulder Shrugs during Human Interaction, and Their Relation to Stancetaking. In: Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on Social Computing (SocialCom). Los Alamitos, CA/ Washington, S.C./ Tokyo: IEEE Computer Society. 932–937.

Dral, J., Heylen, D., & op den Akker, R. (2011). Detecting Uncertainty in Spoken Dialogues: An Exploratory Research for the Automatic Detection of Speaker Uncertainty by Using Prosodic Markers. In K. Ahmad (Ed.), Affective Computing and Sentiment Analysis (Vol. 45, pp. 67–77). Dordrecht: Springer Netherlands. http://doi.org/10.1007/978-94-007-1757-2_6

Fabiszak, M., Jelec, A. Gesture as a symptom of fractures in collective memory. Manuscript in preparation.

Konat, B. (2016). "Może być źle ale może też być bardzo dobrze". Niepewność i niedosłowność w wypowiedziach młodych Polaków dotyczących ich przyszłości zawodowej. In: M. Odelski, A. Knapik, P. Chruszczewski and W. Chłopicki (eds.). *Niedosłowność w języku*. 187-197. Kraków: Tertium.

Krahmer Emiel and Marc Swerts. (2003). How Children and Adults Produce and Perceive Uncertainty in Audiovisual Speech. *Language and Speech, 48(1),* 29-53.

Müller, Cornelia. (2004). "Forms and uses of the Palm Up Open Hand: A case of a gesture family?" In: Cornelia Müller and Roland Posner (eds.) The semantics and pragmatics of everyday gestures. Proceedings of the Berlin conference April 1998. Berlin: Weidler. 233–256.

Nuyts, J. (2001). Subjectivity as an evidential dimension in epistemic modal expressions. *Journal of Pragmatics, 3(3)*, 383–400.

Roseano, P., González, M., Borràs-Comes, J., & Prieto, P. (2016). Communicating Epistemic Stance: How Speech and Gesture Patterns Reflect Epistemicity and Evidentiality. *Discourse Processes*, *53*(3), 135–174. http://doi.org/10.1080/0163853X.2014.969137

Streeck, Jürgen. (2009). Gesturecraft. The manu-facture of meaning. Amsterdam/Philadelphia: Benjamins.

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., Sloetjes, H. (2006). ELAN: a Professional Framework for Multimodality Research. In: Proceedings of LREC 2006, Fifth International Conference on Language Resources and Evaluation.

What do gestures in subordination tell us about clause (in)dependence?

Manon Lelandais & Gaëlle Ferré

University of Nantes, LLING UMR 6310
Chemin de la Censive du Tertre, BP 81227 Nantes cedex 3 FRANCE

manon.lelandais@univ-nantes.fr; gaelle.ferre@univ-nantes.fr

Abstract

Subordinate constructions have been described in syntax as dependent constructions elaborating on primary elements of discourse. Although their verbal and vocal characteristics have been deeply analysed, few studies have provided a qualified picture of the gestures that accompany them and how these gestures can shed light on their dependence or autonomy. We propose to partly fill this gap with the analysis of co-speech gestures produced with two types of subordinate structures in conversational British English.

1 Introduction

In syntactic and discourse studies, subordinate constructions are often described as additions associated to another propositional content in the host structure (Halliday, 1985). This study focuses on restrictive relative clauses and appositive clauses, which both specify or elaborate upon another propositional content (Halliday, 1985).

A restrictive relative clause modifies a nominal expression, refining the identification of its referent (Langacker, 2008). The nominal referent is connected to some participant in the process designated by the relative (Langacker 2008: 424). In *the reasons they gave*, the restrictive relative clause *they gave* increases the relevance of *the reasons*, creating a subcategory for *the reasons* as a referent. Although also introduced with a relative pronoun, an appositive relative clause does not single out a nominal referent, but makes an additional comment about a referent or a whole clause (Longacre 1985). In *I'll happily eat black pudding which I know is disgusting*, the appositive relative clause *which I know is disgusting* evaluates *black pudding*, which can however be identified independently as a referent.

Both of these subordinate clauses are defined as dependent on another predication (Lehmann, 1988). However, the literature shows little consensus in defining clear scopes and boundaries for these structures (Smessaert et al., 2005). This study therefore questions whether they all express the same degree of dependence upon their co-text. If some substantial work has focused on the relation of subordinate clauses to their "hosts" from the perspectives of syntax (e.g. Smessaert et al., 2005) or prosody (e.g. Couper-Kuhlen, 1996), no study has to our knowledge enquired into the gestural expression of these constructions. We investigate the production process of subordinate constructions in English, focusing on several gestural factors of autonomy. The main hypothesis is based on the capacity of these constructions to show distinct factors in function of their syntactic type. Different degrees of autonomy are consequently identified from this new perspective, providing a qualified picture of their insertion in discourse.

2 Theoretical background

2.1 Syntactic subordination

In the traditional categorial division of clause complexes into two uneven and complementary syntactic subgroups, i.e. a main clause and a subordinate, restrictive relative clauses and appositive clauses are both viewed as optional and dependent constituents (Lehmann, 1988), which are deemed semantically useful without standing as constitutive elements. However, the categorisation of these subordinate constructions as dependent has been disputed and reproved by a number of linguists (e.g. Smessaert et al., 2005), described as imprecise for analysing spontaneous speech, especially regarding the nature of introductory elements. From the observation of semantic necessity as imprecise, other criteria are suggested to evaluate clausal combination, in a hierarchy of syntactic and semantic relations (Smessaert et al., 2005). These criteria encourage to investigate clause linkage relying on a wider set of syntactic and semantic parameters, or to go beyond the syntactic frame in observing not only governing relations, but also modal and illocutionary relations (Smessaert et al., 2005).

2.2 Gestural subordination

If little work analysed subordination from a multimodal point of view, some gestural correlates[1] have been shown to participate in the creation and maintenance of cohesion in speech (Hoetjes et al., 2015; Perniss & Özyürek, 2015) with a focus on reference-tracking and information structure.

Two speech segments can be related through their production in co-occurrence with a single gesture unit (Enfield, 2009). On the contrary, hands returning to rest position signal a boundary in discourse (Calbris, 2011).

Beat gestures also single out particular entities (Cavé et al., 1996). They are connected to discourse structure in their function (Kendon, 1972; De Kok & Heylen, 2009), marking out the rhythmic organisation of the utterance

Other articulators play an equally important role in discourse structure. Gaze often moves away from the co-speaker for discourse elaboration as soon as the speaking turn is taken and secured (Barkhuysen et al. 2008). A change in gaze direction towards the co-speaker announces a discourse boundary (De Kok & Heylen, 2009) or an appeal to the co-speaker (Holler et al., 2014).

Eyebrow movement, especially rises, also demarcate various kinds of speech units (Granström & House, 2005).

3 Corpus and methodology

3.1 Working hypotheses

Based on the theoretical background, a specific list of gestural factors of independence is taken into account. If the two syntactic types of subordinate constructions are not equally dependent on their co-text, their number of factors of independence is expected to be different. Namely, the proportion of beat gestures should differ between the two syntactic types (Kendon, 1972; De Kok & Heylen, 2009), as well as the proportion of returns to rest position for hand gestures (Calbris, 2011). The two syntactic types should also be realised with a different proportion of changes in gaze direction (Barkhuysen et al. 2008), and with a different proportion of isolated eyebrow rises (Granström & House, 2005).

3.2 Corpus transcription and annotation

In order to check these hypotheses on conversational English, we used the ENVID[2] corpus described in Lelandais & Ferré (2016). It was first transcribed in Praat (Boersma &

[1] As part of a larger piece of work, this article focuses on a limited number of gesture features. However, many other correlates have been established by the literature.

Weenink, 2013) using standard orthography, and segmented into tone-units. Based on morphological criteria, subordinate constructions were localised and coded on a separate track as SC. The selected occurrences were classified according to their syntactic type. A second track delimitates the environment of these clauses: the preceding tone-unit or part of tone-unit was labelled L (left co-text), the subsequent one labelled R (right co-text).

A total of 386 subordinate constructions were annotated in the corpus, representing 9.76% of the total speaking time (i.e. 3.27 forms/min). 55 occurrences of each syntactic type (restrictive relative clauses, appositive relative clauses) were selected for a balanced comparison, making up a total of 110 forms. The selection targeted occurrences without an interruption, surrounded with other tone-units from the same speaker on their left and right boundaries (i.e. other than a single silent pause yielding the speaking turn). We also made sure that our selection of syntactic constructions was balanced across speakers, so as to avoid any bias due to intra-speaker gestural variability.

After having imported the Praat annotations in Elan (Sloetjes & Wittenburg, 2013), hand gestures, gaze direction, as well as eyebrow movement were manually coded by the two authors, following the parameters proposed by Bressem and Ladewig (2011). Gesture annotation was based on gesture phrases (Kendon, 2004). Each phrase was considered to start at the onset of the gesture and to end at the return to rest position if there was one. In the case of two consecutive gestures, the first phrase ends at a significant change in shape and/or trajectory. Other gestural features such as direction and gestural space were also noted by the two coders.

In separate tracks, gaze direction was annotated as either towards the co-participant or away, eyebrow movement distinguished between rise and frown, and hand gestures were categorised into iconics, metaphorics, pointings, beats, emblems, and adaptators, drawing mainly from McNeill's typology (2005). As hand gestures may have several dimensions, two values could be noted and counted if need be.

The coding scheme used for hand gestures determines the relation of information contained in a gesture to the information in the corresponding speech (Kipp et al., 2007). For instance, if the speaker traces a circle while talking about a round object, the gesture was tagged as an iconic. Ambiguous types were resolved with discussion between the two coders and agreement was reached on the main dimension of gesture types. In order to establish reliability of the gesture type classification, a second coder judged 20% of the data that had been classified by the original coder. The agreement between coders was 100% for gaze direction, 96.4% for eyebrow movement, and 72.1% for hand gesture types.

4 Results

This paper evaluates the gestural autonomy of two types of subordinate constructions. We test whether these constructions are different in their number of gestural factors of independence. To answer our research questions, we used a series of Generalized Linear Mixed Models (GLMMs) fit by maximum likelihood estimation using the R 3.4.0 statistical programming language (R Core Team, 2012) and the lme4 package (Bates et al., 2014). We tested the effect of four factors of independence. Because there was quite a large variation between speakers and dialogues in the production of subordinate constructions, we systematically included Speaker and Dialogue as random factors in the models. Particularities are detailed for each tested effect.

4.1 Hand beats

We first explored possible interactions among the two syntactic types (fixed factor = Type; values = appositive, restrictive) and beat gestures during the production of a subordinate construction (fixed factor = Beat; values = yes; no). We only took into account isolated hand beats, i.e. beats

[2] This corpus features 5 dyads of British English speakers. They already knew each other and were simply asked to talk as freely as possible.

which were not part of a thread of repeated gestures in a discourse sequence. Likewise, hand beats occurring in the middle of very long tone-units (i.e. not close to any prosodic boundary) were excluded. The main effect of beat gesture was significant for restrictive relative clauses (= 1.29, SE = .32, p = .0001). There are also less beat gestures in appositive clauses (= -2.19, SE = .4, p < .001). There are significantly less beat gestures in the tone-unit immediately before the restrictive relative clause, i.e. L (= -1.12, SE = .4, p < .01), and afterwards, i.e. R (= -1.22, SE = 5, p = .01). Example (1) below illustrates this tendency, in association with Figure 1, where (a), (b), (c), and (d) correspond to different moments in its production.

(1)[3]	Michelle	L	[(a) but i put it on the bit
		SC	**where hum (cough) (h) they (h) they were uh #] [(b) in the] [(c) garden]**
		R	[(d) and they were talking]

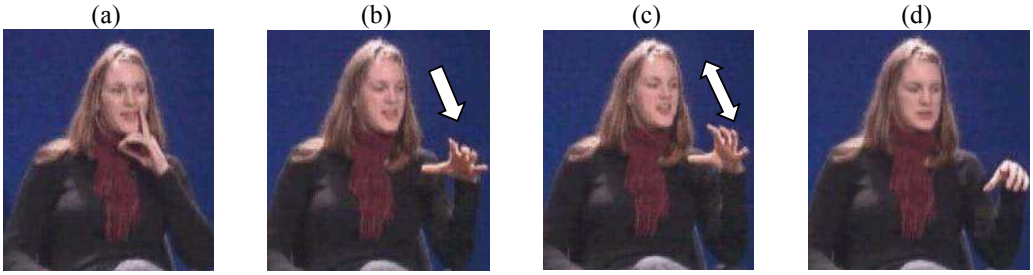


Figure 1. Two successive beats in (1), in co-occurrence with a restrictive relative clause.

SC stands out from the rest of the sequence through its two successive hand beats (b) and (c). The co-occurrence of these hand beats with "in the garden"pragmatically indexes the most relevant informational content in the sequence, which is marked as the retrieval of a substantial search. No other beat gesture is produced in the whole discourse sequence. The palm-down open hand configuration of the hand beat takes an abstract deictic value as Michelle strives to locate an exact scene in time. Michelle partially retracts her palm-down open hand in R (d), dropping her wrist to find a new rest position for the next speaking turns.

4.2 Overlapping hand gestures between two tone-units or more

We then explored possible interactions among the two syntactic types (fixed factor = Type; values = appositive, restrictive) and overlapping hand gestures of the main speaker (fixed factor = overlap; values = yes; no). The main effect of overlap was significant for the sequences containing restrictive relative clauses (ß = 1.34, SE = .36, p < .0005). Those containing appositive clauses feature significantly less overlap (ß = -2.37, SE = .37, p < .001)[4]. Example (2) and Figure 2 feature an appositive relative clause realised with very distinct hand gestures between tone-units.

(2)	Rhianna	L	[(a) my mum's pushing] me to get my license
		SC	**(h) uh which [(b) i guess i should] #**
		R	(h) but well [(c) first of all
			for the moment]

[3] L, SC, and R respectively stand for left co-text, subordinate construction, and right co-text. Transcription conventions include the following:
(h): audible inbreath; #: silent pause
square brackets: gesture span
(a), (b), (c), …: reference to still pictures in subsequent figure

[4] This result is interesting given the fact that appositive relative clauses are shorter than restrictive relative clauses (mean duration of 1.27s vs. 1.51s respectively).

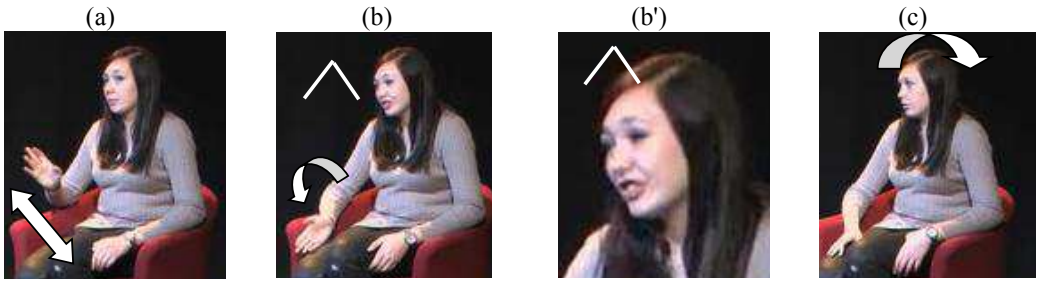


Figure 2. Series of hand gestures during the gestural realisation of example (2), with a very distinct hand gesture in SC (b) from that in L (a), correlated with an eyebrow rise.

Rhianna lays out the reasons why she does not want to learn to drive. She first mentions an adverse opinion in L: her mother would like her to get her license. Rhianna marks this information with a sweep of her right hand corresponding to the verbal item "pushing" (a). This iconic hand gesture gives a hyperbolic dimension to the discourse segment, as Rhianna gives a literal and concrete expression to her mother's advice, and materialises it as strong pressure. However, SC does not elaborate upon her mother's side. SC is a comment going back on L's new information ("get my license"). SC introduces a change in point of view, in that the argumentation switches back to Rhianna's voice in the debate. With a head nod and a lower flip of her right hand (b), Rhianna acts both as the character in the situation she has described in L (Rhianna assents to her mother's exhortation) and as a speaker-utterer: she acknowledges the legitimacy of her mother's advice and marks this concession with a hand flip. She also raises her eyebrows in this design (Figure 2b' is a close-up), taking a strong stance on L's arguments, and marking SC as a contrastive move. Rhianna resumes her main argumentation line in R with a much more categorical expression: while bent in assent during SC, she holds herself upright in R and accompanies the next tone-units with a continuous negative head shake. This sequence is then characterised with two successive assertions that are not equal in intensity: the one taken in R is stronger than that in SC. This asymmetry mirrors the discourse structure, as R continues her sequential discursive agenda while SC does not.

4.3 Gaze movement

We also tested whether there was a possible interaction between the two syntactic types (fixed factor = Type; values = appositive, restrictive) and changes in gaze direction (fixed factor = Change; values = yes; no). The main effect of gaze direction was significant for appositive relative clauses (ß = 0.52, SE = .24, $p < .05$) as changes occur frequently. Changes in gaze direction are illustrated in example (2) above. No main effect was found for restrictive relative clauses (ß = -0.1, SE = .32, $p = .75$).

4.4 Eyebrow movement

Finally, we tested whether there was a possible interaction between the two syntactic types (fixed factor = Type; values = appositive, restrictive) and eyebrow rises (fixed factor = Rise; values = yes; no). The main effect of eyebrow rises was significant for appositive relative clauses (ß = 1.23, SE = .39, $p < .002$). This characteristic is also illustrated in example (2) above. However, the difference for restrictive relative clauses is not significant (ß = -0.32, SE = .48, p = .51). Likewise, the differences with L (ß = 0.47, SE = .58, $p = 43$) and R (ß = 0.15, SE = .57, $p = .79$) are not significant either.

5 Discussion and conclusion

Our analysis confirms that the two syntactic types can be distinguished in their degree of autonomy. Restrictive relative clauses feature only one significant interaction with a factor of independence, while appositive relative clauses show interactions with three factors. Restrictive relative clauses restrict the referential domain of a given entity. Their co-occurring hand gestures

are able to mark out the most relevant lexical feature for a better identification. However, apart from this specificity, no other factor considered in this study signals independence for this type of construction. Appositive relative clauses stand in sharp contrast as they are much more independent, showing multiple cues balanced on several articulators. This interplay between articulators makes disruption more perceptible.

The differences regarding the distribution of the factors in the two syntactic constructions suggest that no common boundary cue is systematically used during subordination. However, the significant presence of hand beats and eyebrow rises hint at the prevalent use of prosodic gestures in both types of subordinate constructions. Interestingly, in the vocal modality, rhythmic cues play a seminal role in the demarcation of both constructions (Lelandais & Ferré, 2016).

Most of the gestures occurring in subordinate constructions also give pragmatic instructions on the informational value of the propositional content (e.g. hand beats, changes in gaze direction, open palm-up gestures, eyebrow rises). Subordinate constructions introduce a break when they establish a different assertive position from the preceding utterance. To avoid a gap between the co-speaker's representations and the speaker's input, gestures mark out this break, but are also able to indicate the informational value of this break in the discourse sequence.

We have alluded to prosodic gestures accompanying subordinate clauses. The effects of prosodic structure have been found to extend beyond the vocal tract to include body movement, in that both manual and oral gestures lengthen at prosodic boundaries (Krivokapić et al., 2017). In the vocal modality, pre-boundary lengthening occurs on different locations for restrictive relative clauses and appositive clauses (Lelandais & Ferré, 2016). An interesting development would be to measure and compare the duration of hand gestures around these boundaries, as well as their temporal alignment.

References

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). Linear mixed-effects models using eigen and s4 [online: http://cran.r-project.org].

Barkhuysen, P., Krahmer, E., & Swerts, M. (2008). The interplay between the auditory and visual modality for end-of-utterance detection. *The journal of the Acoustical Society of America*, 123(1), 354−365.

Boersma, P., & Weenink, D. (2013). *Praat: doing Phonetics by Computer*. Retrieved from http://www.fon.hum.uva.nl/praat/

Bressem, J., & Ladewig, S. (2011). Rethinking gesture phases: Articulatory features of gestural movement? *Semiotica*, 184, 53−91.

Calbris, G. (2011). *Elements of meaning in gesture*. Amsterdam: John Benjamins.

Cavé, C., Guaïtella, I., Bertrand, R., Santi, S., Harlay, F., & Espesser, R. (1996). About the relationship between eyebrow movements and Fo variations. In *Fourth International Conference on Spoken Language* (pp. 2175–2178). Philadelphia: PA.

Couper-Kuhlen, E. (1996). Intonation and clause combining in discourse: the case of because. *Pragmatics*, 6(3), 389−426.

De Kok, I., & Heylen, D. (2009). Multimodal end-of-turn prediction in multi-party meetings. In *Proceedings of the 2009 international conference on Multimodal interfaces* (pp. 91−98). New York: ACM.

Enfield, N. J. (2009). *The Anatomy of Meaning: Speech, Gesture and Composite Utterances*. Cambridge: Cambridge University Press.

Granström, B., & House, D. (2005). Audiovisual representation of prosody in expressive speech communication. *Speech Communication*, 46(3), 473–484.

Halliday, M. A. K. (1985). *An Introduction to Functional Grammar*. London: Edward Arnold.

Hoetjes, M., Koolen, R., Goudbeek, M., Krahmer, E., & Swerts, M. (2015). Reduction in gesture during the production of repeated references. *Journal of Memory and Language*, 79, 1−17.

Holler, J., Schubotz, L., Kelly, S., Hagoort, P., Schuetze, M., & Özyürek, A. (2014). Social eye gaze modulates processing of speech and co-speech gesture. *Cognition*, 133(3), 692−697.

Kendon, A. (1972). Some relationships between body motion and speech. In A. W. Siegman and B. Pope (Eds.), *Studies in Dyadic Communication* (pp. 177−210). New York: Pergamon.

Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge: Cambridge University Press.

Kipp, M., Neff, M., & Albrecht, I. (2007). An annotation scheme for conversational gestures: how to economically capture timing and form. *Language Resources & Evaluation*, 41, 325−339.

Krivokapić, J., Tiede, M., & Tyrone, M. E. (2017). A Kinematic Study of Prosodic Structure in Articulatory and Manual Gestures: Results from a Novel Method of Data Collection. *Laboratory Phonology*, 8(1), 1-26.

Langacker, R. W. (2008). *Cognitive grammar*. Oxford: Oxford University Press.

Lelandais, M., & Ferré, G. (2016). Prosodic boundaries in subordinate syntactic constructions. In *Speech Prosody* (pp. 183-187). Boston: ISCA.

Lehmann, C. (1988). Towards a typology of clause linkage. In John Haiman & Sandra A. Thompson (Eds.), *Clause combining in grammar and discourse* (pp. 181−225). Amsterdam and Philadelphia: John Benjamins.

Longacre, R. E. (1985). Sentences as Combinations of Clauses. In Timothy Shopen (Ed.), *Language Typology and Syntactic Description: Complex constructions* (pp. 372–420). Cambridge: Cambridge University Press.

McNeill, D. (2005). *Gesture and thought*. Chicago: University of Chicago Press.

Perniss, P., & Özyürek, A. (2015). Visible Cohesion: A Comparison of Reference Tracking in Sign, Speech, and Co−Speech Gesture. *Topics in cognitive science*, 7(1), 36−60.

R Core Team. 2012. A language and environment for statistical computing. r foundation for statistical computing. [online: http://www.r-project.org].

Sloetjes, H., & Wittenburg, P. (2008). Annotation by Category: ELAN and ISO DCR. *In Proceedings of the 6th International Conference on Language Resources and Evaluation*. Retrieved from http://www.lat-mpi.eu/tools/elan/

Smessaert, H., Cornillie, B., Divjak, D., & Eynde, K. (2005). Degrees of clause integration: from endotactic to exotactic subordination in Dutch. *Linguistics*, 43(3), 471−529.

Hand Rest Positions of patients with social phobia and therapists in psychodynamic psychotherapy sessions (SOPHO-NET project)

Niklas Neumann[1], Irina Kreyenbrink[1], Katharina C.H. Reinecke[1], Hedda Lausberg[1]

[1]Department of Neurology, Psychosomatic Medicine and Psychiatry,
German Sport University Cologne
Am Sportpark Müngersdorf 6, 50933 Cologne

n.neumann@dshs-koeln.de; irinakreyenbrink@aol.com; k.reinecke@dshs-koeln.de; h.lausberg@dshs-koeln.de

Abstract

Patients with social phobia show specific non-verbal behaviour pattern during psychotherapy sessions. As hand rest positions reveal openness to rapport in social interactions, this study takes a look at this type of nonverbal behaviour. Videos of 10 patient-therapist dyads at the beginning and end of psychodynamic psychotherapies were analyzed with the Rest Position category of the NEUROGES-ELAN system. Three different types of rest positions (*open*, *closed,* and *crossed*) were compared. Patients performed significantly more and shorter rest positions than therapists. They displayed crossed rest positions more frequently, as well as shorter closed and open rest positions. Between beginning and end of therapy, no significant differences in rest positions were found. The study revealed a higher restlessness in patients. The patients' preference for crossed rest position is likely to reflect their fear to expose themselves to the social situation. As rest positions didn't change over time, they appear to be a rather state-independent non-verbal behaviour with a high intra-individual reliability.

1 Introduction

The analysis of non-verbal behaviour represents an important source for diagnostic, intervention and evaluation in the context of patient-doctor interaction (Silverman & Kinnersley, 2010; Brugel, Postma-Nilsenová & Tates, 2015) and in psychotherapies (Fuchs, 2003; Flückiger, Horvath, Del Re, Symonds, & Holzer, 2015; Lausberg, 1995; Lausberg & Kryger, 2011). However, the non-verbal behaviour of patient and therapist in the course of psychotherapies has been poorly investigated (Fuchs, 2003). Patients with social phobia are a relevant sample for the analysis of the non-verbal behaviour due to the symptomatic characteristics. According to the DSM-5, social phobia is part of the anxiety disorder where "the feared objects or situations are limited to social interactions, and avoidance or reassurance seeking is focused on reducing this social fear." (American Psychiatric Association, p.241, 2013). Researchers discuss findings in the present literature and believe it possible to derive the level of anxiety from facial and body movements (Del-Monte et al., 2013). Socially anxious individuals show a high degree of fidgeting, which is transmitted to the interaction partner (Heerey & Kring, 2007; Dow, 1985). Those movements are initiated significantly more often by the socially anxious person (Heerey & Kring, 2007). However, in successful therapies, the non-verbal behaviour of therapists differs systematically compared to normal interaction partners (Fuchs, 2003; Znoj et al. 2004; Flückinger & Znoj, 2009). A recent study by Kreyenbrink et al. (2017) demonstrates that patients with social phobia show significantly more movements and rest positions than therapists, which indicates a higher restlessness in patients. Besides, patients show significantly more irregular body-focused hand movements and more phasic movements in space than therapists. In the course of the therapy, only the irregular body-focused movements significantly decreased. Several studies argue that self-touch in terms of continuously body-focused activity seems to have a stress- or arousal regulative function, because it most often appears in the context of stress (Krout, 1937; 1954; Freedman & Hoffman, 1967; Barroso et al., 1978; Wild et al., 1983; Lausberg, 2013). However, not only movements, but also

rest positions have to be considered in the understanding of the non-verbal behaviour, because they are also part of communication.

In successful therapeutic encounters, interaction partners favoured open rest positions to closed ones, and open rest positions in doctors were rated as more positive by independent observers (Harrigan & Rosenthal, 1983). Furthermore, postures affect the type of emotion experienced by the person adapting the posture (Rossberg-Gempton & Pole, 1993). A positive, involved relation between conversation partners includes an open arm and leg posture, forward lean, facing another, eye contact and postural relaxation (Andersen, 1985; Mehrabian, 1972). A meta-analysis investigated the relationship between the non-verbal behaviour and positivity (Tickle-Degnen, Rosenthal, & Harrigan, 1989). The analysis showed that uncrossed posturing of the arms had a medium effect of .25. The Rest Position Development Model by Lausberg (submitted (a)) differentiates crossed, closed, and open rest positions, as a reflection of distinct interactive states. Thus, interactive partners regulate their rest positions according to the quality of the social interaction from crossed via closed to open rest positions or vice versa. Accordingly, in successful psychotherapies, patients who predominantly display crossed rest positions should therefore shift to more closed and open rest positions. As the non-verbal behaviour of therapists differs systematically in successful therapies compared to normal interaction partners, the fidgeting of the patients should not be contagious for the therapists.

Thus far, research concerning rest positions in non-verbal behaviour has not investigated the difference between the patient's and the therapist's non-verbal behaviour in psychodynamic psychotherapy[1] setting yet. Due to the patient's psychopathology and the different roles in the context of psychotherapy, a difference in rest positions is assumed.

The present study examines the hypothesis that frequency and duration of hand rest positions differ between patient and therapist in psychodynamic psychotherapy sessions. It is suspected that patients with social phobia will show more crossed rest positions than therapists, since social interaction is assumed to constitute a stressful setting for them. Furthermore, we expect an altered non-verbal behaviour in patients over time towards more closed and opened and less crossed rest positions as Harrigan & Rosenthal (1983) and Lausberg (submitted (b)) suggested about successful therapist-patient dyads.

2 Methods

2.1 Data acquisition and sample

Video recordings of 49 patients with social phobia and their therapists during psychodynamic psychotherapy sessions, originating from the research association of social phobia from Dresden, Germany (SOPHO-NET; Leichsenring et al., 2009), were analyzed in respect of their applicability for the current study. Exclusion criteria were: Interaction partners were not allowed to hold something in their hands or to stand up; hands had to be visible all the time; the therapy had to be successful. A therapy was regarded as successful, when the Liebowitz-Anxiety-Scale-CA recorded a decrease in the symptoms of the first available measuring point to the last available measuring point. A score under 30 reflects a remission of the social phobia (Liebowitz, 1987). The patients in the current study were not completely remitted. The final sample consisted of ten patients with social phobia (four females and six males) at the age of 17 to 52 and six therapists (two females and four males) at the age of 31 to 60.

20 videos (two for each dyad) were included in the coding analysis and the first six minutes of the first therapy session (after the probationary sessions) as well as the first six minutes of the pre-last

[1] Psychodynamic psychotherapies involve treatments that operate on a continuum of supportive-interpretive psychotherapeutic interventions (Gabbard, 2004). "Interpretive" aims to enhance the patients' insight into repetitive conflicts and "supportive" strengthens the abilities that are temporarily inaccessible or have not been developed in patients.

session were analyzed. Following Ambady and Rosenthal's (1992) proposition of a video analysis time frame of four to six minutes, we analyzed the first six minutes of each video.

2.2 Analysis of hand movements

Patients' and therapists' hand movement behaviour was analyzed with the NEUROGES analysis system (Lausberg, 2013). Due to the research question, the Rest Position category was chosen, which differentiates three different types of positions: *open, closed,* and *crossed* (*Figure 1*). The analysis was realized with the NEUROGES-ELAN system (Lausberg & Slöetjes, 2009, re. 2015), in which NEUROGES is embedded in the multimedia annotation tool ELAN. The system was developed for objective, valid and reliable hand movement research purposes and has already been tested in over 500 persons with and without psychopathological symptoms (Lausberg & Slöetjes, 2015). The videos were analyzed by two independent certified raters without sound to garuantee that raters would not be influenced by the verbal content during the coding process. The interrater agreement was measured with EasyDIAg and a modified Cohen's kappa (Holle & Rein, 2015). Modified Cohen's kappa scores and raw agreement scores (in brackets) were substantial for open*:* $\kappa = .76$ (0.91), closed: $\kappa = .70$ (0.85) and crossed: $\kappa = .81$ (0.98) values.

SPSS version 23 was used for the statistical analysis. Mean frequency and mean duration of the Rest Position values were exported from ELAN into MS Excel and SPSS. Frequency refers to the mean number of value units per minute and the duration refers to the mean duration in seconds of each rest position value unit (Lausberg, 2013). For each parameter, frequency and duration, a repeated measures MANOVAs (factors: Group(2) x Time(2)) was calculated for a comparison of the different Rest Position values in the first and the last sessions.

Open *Closed* *Crossed*

Figure 1. The three different NEUROGES-values for Rest Position.

3. Results

The multivariate analysis of the mean frequency of rest positions reveals a main effect of Group (*Figure 2*). Patients show significantly more rest position units per minute than therapists (Pillai's trace=.574, F(3.61)=7.195, p=.042).

The univariate analysis of the mean frequency of each of the three rest position types reveals that patients show significantly more *crossed* rest positions than therapists (F=4.796, df=1, p=.042).

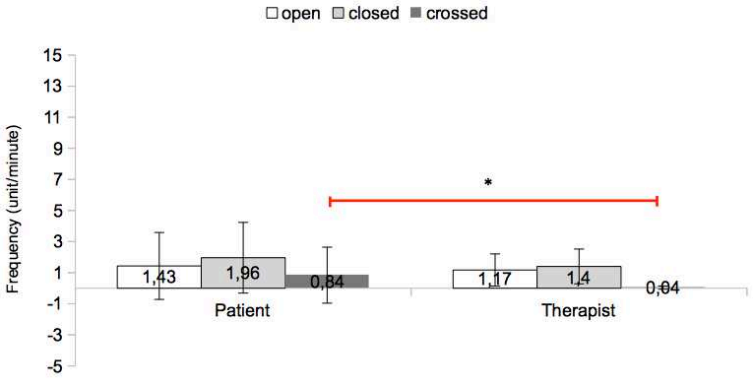


Figure 2. Mean frequency of open, closed, and crossed rest position units in the patient and therapist groups. Error bars indicate standard error of mean (SE).

The multivariate analysis of mean duration of rest positions reveals a main effect for Group (*Figure 3*). Patients show significantly shorter rest position units than therapists (Pillai's Trace=.407, F (3.16)= 3.663, p=.035).

The univariate analysis of the mean duration of each of the three rest position types reveals that patients show significantly shorter *closed* and *open* rest position units than therapists (closed: F=3.66, df=16, p=.031; open: F=3.66, df=16, p=.039).

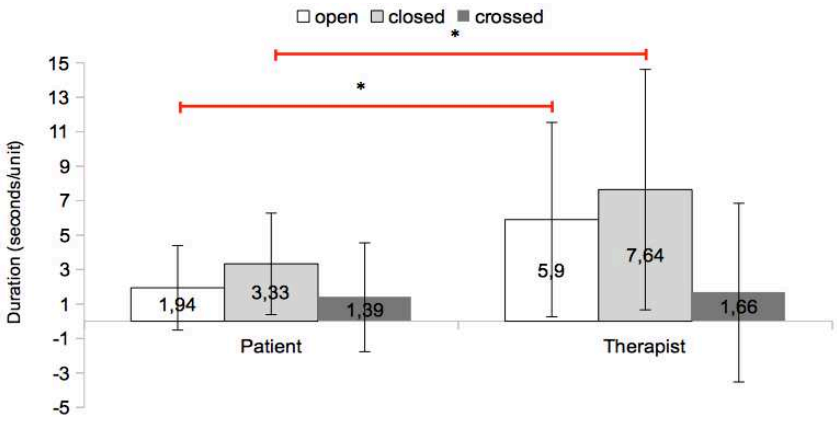


Figure 3. Mean duration of open, closed, and crossed rest position units in the patient and therapist groups. Error bars indicate standard error of mean (SE).

The multivariate and univariate analysis showed no significant effect of the interaction group x time on the mean frequency and mean duration.

3 Discussion

The present study examines hand rest positions in the non-verbal behaviour of patients with social phobia and their therapists in psychodynamic psychotherapies. Patients showed significantly more and shorter rest positions than therapists. Specifically, they displayed more *crossed* rest positions and shorter *closed* and *open* rest positions than therapists. There were no significant differences in rest positions between the beginning and the end of the therapy.

The patients' high frequency of rest positions indicates restlessness, as the finding implies that the patients do not remain in one rest position for a long time. In the context of the present literature, this finding could indicate a correlation between anxiety or level of arousal and body movement, as supposed by Del-Monte et al. (2013). Several studies show a correlation between continiously body-focused activity and stress or anxiety (Ekman & Friesen, 1972; White, 2013). The continiously body-focused activity is often described as fidgeting, which has a stress and arousal regulative function, because it most often appears in the context of stress (Krout, 1937; 1954; Freedman & Hoffman, 1967; Barroso et al., 1978; Wild et al., 1983; Lausberg, 2013). Studies investigating the non-verbal behaviour of people with social phobia show a high degree of fidgeting in this group (Dow, 1985; Heerey & Kring, 2007). Above all, self-manipulations (dynamic and static touches of the own body or body-related objects) correlated positively with the reactivity of the heart frequency (Monti et al., 1984).

Furthermore, the present study shows that patients with social phobia as compared to their therapists prefer *crossed* rest positions. This behavior is likely to reflect social anxiety (Del-Monte et al., 2013). The physical sign of the degree to which the person is open toward and vis a vis the other in her/ his position is related to psychological openness and rapport (Charny, 1966; Scheflen, 1973). While this has to be applied relative to the individual and cultural baseline of position openness (Lausberg, submitted), the present study evidences for the group of individuals with social phobia that *crossed* rest positions are more predominant than in their therapists.

In the course of the therapy, the rest position types of patients and therapists did not change significantly as hypothesized. Patients did not perform significantly more *closed* and *open* rest positions at the end of the therapy, contrary to what Harrigan and Rosenthal (1983) stated about successful therapist-patient dyads and Andersen (1985) and Mehrabian (1972) about positive relation between interaction partners. This could be explained by the fact that rest positions are relatively stable non-verbal behavior and can therefore be associated with traits. Moreover, patients in the current study showed a decrease in symptoms of the LSAS-CA in the course of the therapy, but were not completely remitted (LSAS score under 30). Therefore, research should investigate fully remitted patients at the end of a psychdynamic psychotherapy.

Further studies should consider a larger sample size and explore different sequences in the therapy sessions.

References

Ambady, N., & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: A Meta-analysis. *Psychological Bulletin*, 111(2), 256.

American Psychiatric Association (2013). Diagnostic and statistical manual of mental disorders. *American Psychiatric Publishing*; 5th edition.

Andersen, P.A. (1985). Nonverbal immediacy in interpersonal communication. In A.W. Siegman & S. Feldstein (Eds.), *Multi-channel integrations of nonverbal behavior* (pp.1-36). Hillsdale, NJ: Erlbaum.

Barroso, F., Freedman, N., Grand, S., Van Meel, J. (1978). Evocation of two types of hand movements in information processing. *J Exp Psychol-Hum Percept Perform* 4, 321-329.

Brugel, S., Postma-Nilsenová, M., & Tates, K. (2015). The link between perception of clinical empathy and non-verbal behavior: The effect of a doctor's gaze and body orientation. *Patient education and counseling,* 98(10), 1260-1265.

Charny, J. E. (1966). Psychosomatic manifestations of rapport in psychotherapy. *Psychosomatic Medicine*, 28(4), 305-313.

Del-Monte, J., Raffard, S., Salesse, R. N., Marin, L., Schmidt, R. C., Varlet, M., ... & Capdevielle, D. (2013). Non-verbal expressive behaviour in schizophrenia and social phobia. *Psychiatry research*, 210(1), 29-35.

Dow, M.G. (1985). *Peer validation and idiographic analysis of social skill deficits*. *Behavior Therapy,* 16(1), 76-86.

Ekman, P., & Friesen, W. V. (1972). Hand movements. *J Commun*, 22, 353-374.

Flückinger, C., Znoj, H. (2009): Zur Funktion der nonverbalen Stimmungsmodulation des Therapeuten für den Therapieprozess und Sitzungserfolg: Eine Pilotstudie. *Z Klein Psychol Psychother*, 38, 4-12.

Flückiger, P. D. C., Horvath, A. O., Del Re, A. C., Symonds, D., & Holzer, C. (2015). Bedeutung der Arbeitsallianz in der Psychotherapie. *Psychotherapeut, 60(3),* 187-192.

Freedman, N., Hoffman, S.P. (1967). Kinetic beavior in altered clinical states: Approach to objective analysis of motor behavior during clinical interviews. *Percept Mot Skills* 24, 527-539.

Fuchs, T. (2003). Nonverbale Kommunikation: Phänomenologische, entwicklungspsychologische und therapeutische Aspekte. *Zeitschrift für klinische Psychologie und Psychotherapie,* 51(4), 333-345.

Gabbard O.G. (2004). *Long-term Psychodynamic Psychotherapy: A basic text*. PLACE: American Psychiatric Publishing.

Harrigan, J.A., & Rosenthal, R. (1983). Physicians' head and body positions as determinants of perceived rapport. *Journal of Applied Social Psychology,* 13(6), 496-509.

Heerey, E.A., & Kring, A. M. (2007). Interpersonal consequences of social anxiety. *Journal of abnormal psychology,* 116(1), 125.

Holle, H., Rein, R. (2015). EasyDIAg: A tool for easy determination of interrater agreement. *Behav Res Methods*, 47, 837-847.

Kreyenbrink, I., Neumann, N., Joraschky, P., Konstantinidis, I., & Lausberg, H. (2017). Nonverbales Verhalten von Patienten mit sozialen Phobien und ihren Therapeuten in psychodynamischen Psychotherapien (Teilprojekt SOPHO-NET). *Zeitschrift für Psychosomatische Medizin und Psychotherapie,* 63, 297-313.

Krout, M.H. (1937). Further studies on the relation of personality and gesture. A nosological analysis of autistic gestures. *J Exp Psychol* 20, 279-287.

Krout, M.H. (1954). An experimental attempt to determine the significance of unconscious manual symbolic movements. *J Genl Psychol* 51, 121-152.

Lausberg, H. (1995). Bewegungsverhalten als Prozessparameter in einer kontrollierten Studie mit funktioneller Entspannung. Unpublished paper presented at the 42nd Arbeitstagung des Deutschen Kollegiums für Psychosomatische Medizin.

Lausberg, H., & Slöetjes, H. (2009). Coding gestural behaviour with the NEUROGES- ELAN system. *Behaviour research methods, 41*(3), 841-849.

Lausberg, H., Kryger, M. (2011). Gestisches Verhalten als Indikator therapeutischer Prozesse in der verbalen Psychotherapie: Zur Funktion der Selbstberührungen und zur Repräsentation von Objektbeziehungen in gestischen Darstellungen. *Psychother Wiss 1*, 41-55.

Lausberg, H. (2013). *Understanding body movement: A guide to empirical research on non-verbal behaviour*. Frankfurt: Peter Lang.

Lausberg, H., & Slöetjes, H. (2015). The revised NEUROGES- ELAN system: An objective and reliable interdisciplinary analysis tool for non-verbal behaviour and gesture. *Behaviour research methods,* 1-21.

Lausberg, H.. The Rest Position Development Model (submitted (a))

Lausberg, H.. Methods for gesture and nonverbal interaction analysis (submitted (b))

Leichsenring, F., Salzer, M., von Consbruch, K., Herpertz, S., Hiller, W., et al'. (2009). SOPHO-NET-Forschungsverbund zur Psychotherapie der sozialen Phobie. *Psychotherapie, Psychosomatik, Mediznische Psychologie*, 59, 117-123.

Liebowitz, M. R. (1987). Social Phobia. *Mod Probl Pharmacopsychiatry*, 22, 141-173.

Mehrabian, A. (1972). *Nonverbal communication*. Chicago: Aldine-Atherton.

Monti, P. M., Boice, R., Fingeret, A. L., Zwick, W. R., Kolko, D., Munroe, S., Grunberger, A. (1984). Midi-level measurement of social anxiety in psychiatric and non-psychiatric samples. *Behaviour Research and Therapy*, 22(6), 651-660.

Rossberg-Gempton, I., & Poole, G.D. (1993). The effect of open and closed postures on pleasant and unpleasant emotions. *The Arts in Psychotherapy, 20*(1), 75-82.

Scheflen, A. E. (1973). *Communicational structure: Analysis of a psychotherapy transaction*. Place: Indiana U.Press.

Tickle-Degnen, L., Rosenthal, R., & Harrigan, J.A. (1989). Non-verbal behavior as determinant of favorableness of impressions formed: Eight meta- analyses. Unpublished manuscript.

White, C. N. (2013). An examination of the social self preservation model and the physiological resonance of social stress. Doctoral dissertation, Saint Louis University.

Wild, H., Johnson, W.R., & McBrayer, D.J. (1983). Gestural behavior as a response to external stimuli. *Percept Mot Skills* 56, 547-550.

Znoj, H., Nick, L., & Grawe, K. (2004). Intrapsychische und interpersonale Regulation von Emotionen im Therapieprozess. *Z Klin Psychol Psychother*, 33, 261-269.

An evaluation framework to assess and correct the multimodal behavior of a humanoid robot in human-robot interaction

Duc-Canh Nguyen, Gérard Bailly & Frédéric Elisei

GIPSA-Lab, Grenoble-Alpes Univ. & CNRS, Grenoble, France

duc-canh.nguyen@gipsa-lab.fr

Abstract

We discuss here the key features of a new methodology that enables professional caregivers to teach a socially assistive robot (SAR) how to perform the assistive tasks while giving verbal and coverbal instructions, demonstrations and feedbacks. We describe here how socio-communicative gesture controllers – which actually control the speech, the facial displays and hand gestures of our iCub robot – are driven by multimodal events captured on a professional human demonstrator performing a neuropsychological interview. The paper focuses on the results of two crowd-sourced experiments where we asked raters to evaluate the multimodal interactive behaviors of our SAR. We demonstrate that this framework allows decreasing the behavioral errors of our robot. We also show that human expectations of functional capabilities increase with the quality of its performative behaviors.

1 Introduction

Socially assistive robots (SAR) are typically facing two situations with quite different timescales and related challenges: long-term vs. short-term interactions. Long-term interactions often target one single user with the challenge of engaging into open-domain conversations, establishing affective relation, such as performed by) (see Robinson et al., 2014 for a review). In contrast, short-term interactions are typically task-oriented (e.g. welcoming a client, giving directions, serving cocktails (Foster et al., 2014), conducting interviews (Bethel et al., 2016), repetitive and should cope with a large variety of user profiles.

Our work focuses on the development of socio-communicative abilities for short-term interactions. The target scenario is a neuropsychological interview with an elderly person.

2 The SOMBRERO Framework

The multimodal interactive behavioral model learning is performed by three main steps illustrated in Figure 3. Firstly, we collect representative interactive behaviors from human tutors especially by professional coaches. Secondly, the comprehensive models are trained from the collected data with considering a priori knowledge of users' models and task decomposition. Finally, the gesture controllers are built in order to execute the desired behaviors driven by the interactive model.

The interactive models of HRI systems are mostly inspired by Human-Human interaction (HHI). Therefore, they face several issues: (1) adapting the human model to the robot's interactive capabilities; (2) the drastic changes of human partner behaviors in front of robots or virtual agents; (3) the modeling of joint interactive behaviors; (4) the validation of the robotic behaviors by human partners until they are perceived as adequate and meaningful. The two first issues are solved by the framework used in SOMBRERO (Gomez et al., 2015) which allows coaches to involve and demonstrate an expected HRI behavior through immersive teleoperation technique: the so-called beaming method driving gaze, head movements and mouth of the robot in real-time during the

interaction. The third issue has been addressed by (Mihoub et al., 2015; Mihoub et al., 2016). They proposed to train statistical behavioral models that encapsulate discrete multimodal events performed by the interlocutors into a single dynamical system that could be further used to monitor behaviors of one interlocutor and generate behaviors of the other.

In this paper, we propose a method to address the fourth issue: the replay of interactive behaviors by the robot and its assessment by human raters.

Figure 1. The iCub humanoid (named Nina) robot from the subject's perspective.

Figure 2. Capturing the multimodal behavior of the human tutor during HHI. Movements of upper limbs (head, arms and hands) are captured by 22 markers glued on segments with a Qualysis ® mocap system. Gaze was tracked using Pertech ® head-mounted eye tracker.

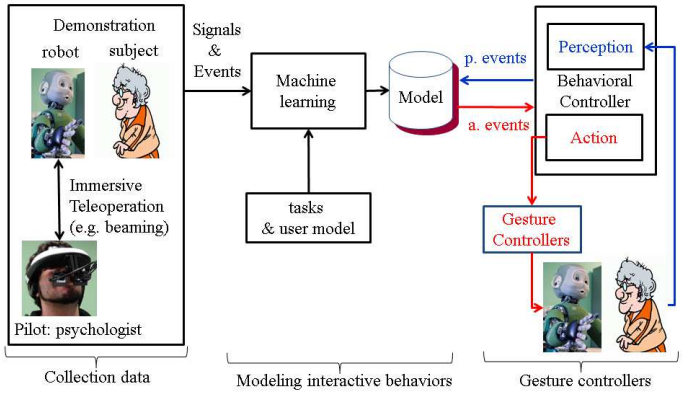


Figure 3. The three main steps of learning interaction by demonstration: collecting HRI data, learning a behavioral model and building appropriate sensorimotor controllers.

2.1 From HHI to HRI

The short-term interactive scenario involved here is a French adaptation of the Selective Reminding Test, so-called RL/RI 16 (Dion et al., 2015). It is often used to diagnose early loss of episodic memory. The test includes four phases: (1) words memorization (aka learning), (2) testing the words recall capability, (3) recognition of the words and (4) distractive task, which were described detail in (Nguyen et al., 2016).

In order to avoid complex gestures usually performed by human interviewers using scoring sheets and paper-based notes, the SAR uses two tablets as physical medium: one facing the robot to fake the note taking activity and the other facing the subject to display word items.

HHI demonstrations were performed by a female professional psychologist. We collected her multimodal behavior (speech, head movement, arm gestures and gaze, see Figure 2) when

interviewing five different elderly patients together with the speech of the interviewees. These continuous signals were then semi-automatically converted into time-stamped events using Elan (Wittenburg et al., 2006) and Praat (Boersma and Weenink 1996) editors. With Elan, we basically determined hand strokes triggered by the interviewer to grasp and act on resources (workbook, notebook, chronometer) and regions of interest for fixations. With Praat, we hand-checked the phonetic alignment performed by an automatic speech recognition system and added prosodic annotations as well as special phonetic events related to backchannels and breath noises. The HHI multimodal score consists thus in time-stamped speech, head/arm/hands gestures and gaze events. We then developed modality-specific gesture controllers to map these events to robotic behaviors. HHI to HRI retargeting is thus performed using multimodal events as pivots. This HHI multimodal score is available for download (see section 6).

2.2 Speech and Gesture controllers

We built four gesture controllers: arm, gaze, eyelids and speech, which will cooperate together to enable SAR to replicate the RL/RI scenario. The arm gesture controller is based on the iCub Cartesian Interface (Pattacini et al., 2010) and handles three basic gestures: resting, preparing to click and clicking to trigger display/hide items. The gaze gesture controller triggers fixations towards three regions of interest: subject's face, subject's tablet and robot's tablet. The gaze gesture controller synchronizes with the gesture controller for ensuring sensory-motor control, e.g. locking the gaze at finger target when initiating arm gesture. Conversely, when no such sensory-motor control is required, the gaze is driven by the other socio-communicative events. The eyelid gesture controller was added to cope with gaze direction, speech and blinking. Despite blinking rate is known to correlate with emotional state and cognitive state – notably thinking, speaking vs. listening (see Bailly et al., 2010) – blinks are generated according to a Gaussian distribution at 0.5 Hz +/- 0.1 Hz. Finally, the speech gesture controller was handled by our in-house audiovisual text-to-speech system (Bailly et al., 2009). The corpus-based of AV synthesis is fed with articulatory movements from a female adult that have been scaled to NINA's degrees of freedom and time-aligned with voice segments from a female teenager.

3 Evaluation

We want to evaluate if the coordinated behaviors are perceived and interpreted as expected by subjects. Since subjects can not both live and rate the interaction on-line, we thus asked third parties (observers or raters) to rate the final rendering of a multimodal score recorded during HHI – and replayed by our robotic embodiment.

3.1 State of the art

Most subjective evaluations of HRI behavior have been performed using questionnaires, where subjects or third parties are asked to score specific dimensions of the interaction on a Likert scale. (Fasola and Mataric, 2013) rated several aspects such as pleasure, interest, satisfaction, entertainment and excitation. (Huang and Mutlu, 2014) assessed a narration of a humanoid robot along several dimensions such as immediacy, naturalness, effectiveness, likability and credibility. (Zheng et al., 2015) compared control strategies for robot arm gestures along dimensions such as intelligibility, likeability, anthropomorphism and safety. Although delivering very useful information notably for sorting between competing control policies or settings, these questionnaire-based evaluations provide developers with poor information about how to correct faulty behaviors since the evaluation is performed offline and questions address global properties of the entire interaction.

3.2 Designing and performing on-line vs off-line evaluation

Following the procedure proposed by (Kok and Heylen, 2011), we opted for a method that enable raters to signal faulty events, since the HRI behaviors are essentially controlled by events. We thus designed an on-line evaluation technique that consists in asking raters to immediately signal faulty behaviors by pressing on the ENTER key of their computer when they just experience them. Following Kok & Heylen, we will use the term yuck responses to name these calls for rejection.

In order to gather a significant amount of yuck responses for a set of identical stimuli, we here ask our raters to evaluate the replay by the robot of the multimodal behavior, originally performed by our psychologist in front of one unique subject. We in fact filmed the robot's performance as fed by the multimodal score of the original situated interaction (arm gestures, head movement, gaze...). For now, the only original play-backed behavior is the subject speech. The camera remains fixed at the mean position of the location of the eyes of the subject. The raters can see the robot facing them, but not the patient that they replace. They can hear the robot, as well as what the subject says: they are spectators, but occupy the seat of the subject. For our first experimental assessment (Nguyen et al., 2016), we created a website (see section 6) where we ask people to look at a first-person video and to press the ENTER key anytime they feel the robot behavior is incorrect. This provides a time-varying normalized histogram of incorrect behaviors. The maxima of the density function are cueing time-intervals for which a majority of raters estimate the behavior is inappropriate or hinders the interaction. Further diagnostic of what cues cause these faulty behaviors are later performed by roboticists and system designers. This on-line evaluation task is preceded by a quick screening of subjects (age, sex and mother tongue) and a familiarization exercise, and followed by a questionnaire that asks the subjects judgments (five-level Likert) on nine points: "Did the robot adapt to the subject?"; "Did the subject adapt to the robot?"; "Did you feel relaxed?"; "Did you feel secure?"; "Was the rhythm of the robots behavior well adapted?"; "Was the interaction pleasant?"; " Was the multimodal behavior appropriate?"; "Did the robot pay attention while speaking?"; "Did the robot pay attention while listening?"

The present paper builds up new results upon a previous experiment (Nguyen et al., 2016) performed by 50 French native subjects (26 males, 24 females, 32±12 years). The 25 most signaled events were related to the following problems: (1) Ungrounded pointing gestures; Underrepresented gaze contacts; (2) Inactivity during reverse counting or covert thinking; (3) Inadequate speech articulation; (4) Lack of facial expressions.

For the current study, after correcting these faulty behaviors, we performed a new experimental assessment using the same experimental protocol. The second experimental assessment was performed by 46 French native subjects (16 males, 30 females, 36±16 years), 38 of whom already participated in the first assessment.

4 Results

4.1 Yuck responses

We remedied to these faulty behaviors by adding extra-rules to our gesture controllers. For example, in order to avoid immobility due the periods of poor external stimulation, the gaze controller automatically randomly loops on the two last regions of interest when the delay from the last fixation exceeds 3 sec.

With this rule, the number of yucks at timestamps 10, 11, 12, 13, 16, 18, 19 and 23 are significantly reduced as shown in Figure 4. However, this randomization should not be equally distributed and should favor the subject's face, since the participants still complain about its lack of engagement with the human subject (e.g. around peak 11, 12, 13 in counting task). This problem will be suppressed by

systematically adding the subject's face to the current attention stack and favoring this region of interest in the gaze distribution.

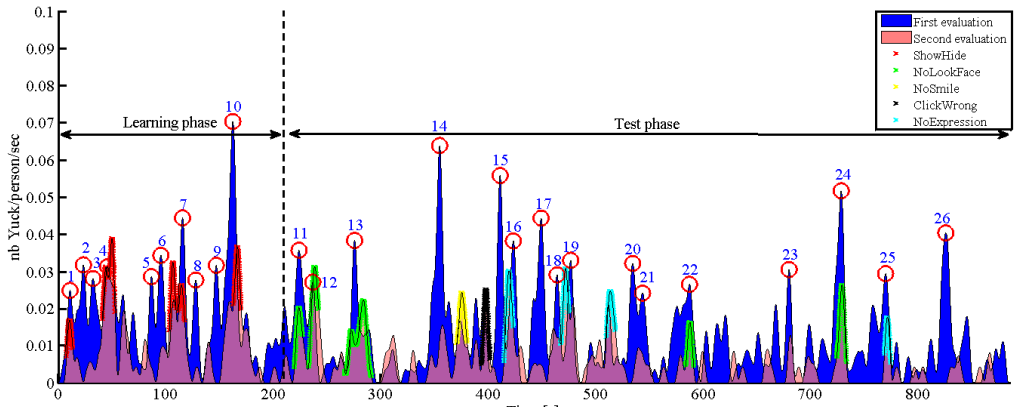


Figure 4. Comparing the yucking probability as a function of time for first vs. second assessment by the subjects (blue area: first evaluation, pink area: second evaluation, purple area: overlap between the first and second evaluations, dot-lines: annotated yuck).

In the first evaluation, yucks occuring at timestamps 2, 3, 4 were due to the wait-motion-done setting. In the redesign, theses faulty behaviors have been removed by disabling the wait-motion-done option that discards any new command while the current gesture has not reached its target given a given precision. This policy is efficient: the yuck responses at landmarks 2, 3, 4 are significantly reduced in the second evaluation. The yucks at landmarks 14 and 15 were repaired by forcing the closing gesture at the end of phonation. Although many of the faulty behaviors are suppressed, several faulty detections still remain while some new yucks emerge from the background, notably the absence of expressiveness, e.g. smile responses to subject's embarrassment or head nodding normally associated with incentives, respectively cued by yellow vs. cyan extrema.

4.2. Subjective ratings and comments

We also compared subjective ratings from the first vs. second assessment (see Figure 5). While the new behavioural score results in an effective decrease of the yuck responses and descent behavior effectively improves – most other off-line subjective ratings degrade. Likelihood ratio tests comparing the combined multinomial model RATINGS ~ SEX+SESSION+EXPOSURE with the individual models RATINGS ~ SEX+SESSION, RATINGS ~ SEX+EXPOSURE and RATINGS ~ EXPOSURE+SESSION show that sex significantly contributes to the ratings of questions robot adaptation. In addition, the version ($p = 0.049$) and the number of evaluations ($p = 0.041$) has significant contributions on feels_relaxed. This means that people feel more relaxed and the robot was rated as more friendly in the second evaluation.

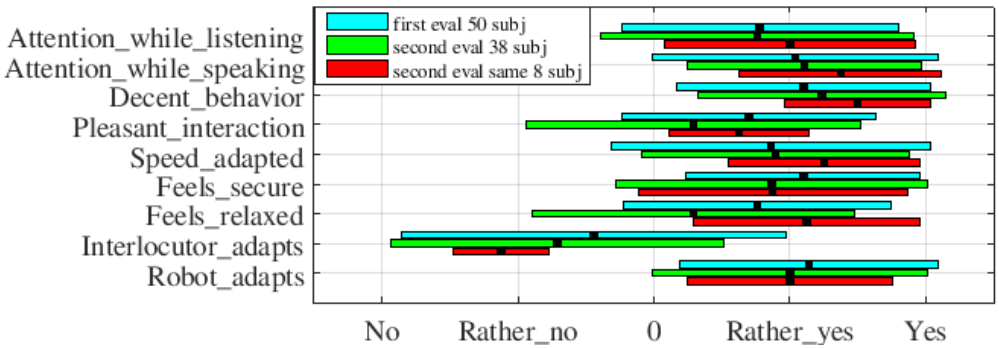


Figure 5. Comparing subjective ratings according to conditions.

In the free comments, some raters of the first evaluation campaign mentioned the rather directive style of our female interviewer and the absence of emotional vocal and facial displays on our SAR e.g. laughs and smiles. While most raters of the second evaluation campaign underlie the quality of gaze behavior, the majority criticize the poorness of emotional displays: "robot without human warmth!", "why robots never smile?", etc. It seems that the increased behavioral quality and appropriateness also increased the participant's expectations. When they have the impression that the robot is reactive, aware of the situation and monitors the interaction task in an appropriate way, they can allocate more attentional resources to the social and emotional aspects of the interactive behavior.

5 Conclusions

We have put forward an original framework for the on-line evaluation of HRI behavior that offers subsequent glass-box assessment: On-line evaluation provides developers with when something goes wrong. Post-hoc reverse engineering should be then performed by the socially assistive robot (SAR) designers to remedy for the potential causes of the most salient yuck responses, i.e. what went wrong. Off-line assessment provides developers with what is missing. These local vs global assessment procedures should be combined to maintain SAR at the top of the uncanny cliff.

We should augment the socio-communicative skills of our SAR with more expressive dimensions. While Nina is missing facial displays (notably articulated eyebrows), its available degrees-of freedom (notably head, arm and body gestures) together with speech should be recruited to encode more linguistic and paralinguistic functions.

Acknowledgements

This research is supported by SOMBRERO ANR-14-CE27-0014, PERSYVAL ANR-11-61 and ROBOTEX ANR-10-EQPX-44-01.

Appendix

The test page is available at http://www.gipsa-lab.fr/~duccanh.nguyen/assessment

Multimodal data/labels freely available at: http://www.gipsa-lab.fr/projet/SOMBRERO/data

References

Bailly G. & Gouvernayre, C. (2012). Pauses and respiratory markers of the structure of book reading. In Interspeech. Florence, Italy, 2218–2221.

Bailly G., Raidt, S. & Elisei, F. (2010). Gaze, conversational agents and face-to-face communication. Speech Communication - special issue on Speech and Face-to-Face Communication, 52(3), pp. 598–612.

Bailly G., Govokhina, O., Elisei, F. & Breton, G. (2009). Lip-synching using speaker-specific articulation, shape and appearance models. Journal of Acoustics, Speech and Music Processing. Special issue on "Animating Virtual Speakers or Singers from Audio: Lip-Synching Facial Animation". Retrieved from https://asmp-eurasipjournals.springeropen.com/articles/10.1155/2009/769494

Bethel C., Henkel Z., Stives, K., May, D. C., Eakin D. K., Pilkinton, M., Jones, A., Stubbs-Richardson, M., (2016). Using robots to interview children about bullying: Lessons learned from an exploratory study. Preceeding of 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), 712–717.

Boersma P., & Weenink D., (2013): Praat: doing phonetics by computer, version 5.3.51. Retrieved from http://www.praat.org/

Dion M1, Potvin O, Belleville S, Ferland G, Renaud M, Bherer L, Joubert S, Vallet GT, Simard M, Rouleau I, Lecomte S, Macoir J, Hudon C. (2015). Normative Data for the Rappel libre/Rappel indicé à 16 items (16-item Free and Cued Recall) in the Elderly Quebec-French Population. The Clinical Neuropsychologist, 28(1), 1–19. https://www.ncbi.nlm.nih.gov/pubmed/24815338

Fasola J. & Mataric, M. (2013). A socially assistive robot exercise coach for the elderly. Journal of Human-Robot Interaction, 2, 3–32.

Foster M.E., Keizer, S. & Lemon, O. (2014). Towards action selection under uncertainty for a socially aware robot bartender. In Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction, 158–159.

Gomez G., Plasson, C., Elisei, F., Noël, F., & Bailly, G. (2015). Qualitative assesment of a beaming environment for collaborative professional activities. European conference for Virtual Reality and Augmented Reality (EuroVR). Retrieved from https://hal.archives-ouvertes.fr/hal-01228890

Huang C.-M. & Mutlu, B. (2014). Learning-based modeling of multimodal behaviors for humanlike robots. In Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction, (57–64). Bielefeld.

Kok I. & Heylen, D. (2011). Observations on listener responses from multiple perspectives. Proceedings of the 3rd Nordic Symposium on Multimodal Communication, Northern European Association for Language Technology, 48–55.

Mihoub A., Bailly, G. Wolf, C. & Elisei, F. (2016). Graphical models for social behavior modeling in face-to face interaction. Pattern Recognition Letters 74, 82–89.

Mihoub A., Bailly, G. Wolf, C., & Elisei, F. (2015). Learning multimodal behavioral models for face-to-face social interaction. Journal of Multimodal User Interfaces, 9(3), 195–210

Nguyen D. C., Bailly, G. & Elisei, F. (2016). Conducting neuropsychological tests with a humanoid robot: Design and evaluation. In Cognitive Infocommunications (CogInfoCom), Warsaw, Poland, 337–342.

Pattacini U., Nori, F., Natale, L., Metta, G., & Sandini, G. (2010). An experimental evaluation of a novel minimum-jerk cartesian controller for humanoid robots. International Conference on Intelligent Robots and Systems (IROS), 1668–1674.

Robinson H., MacDonald, B. & Broadbent, E. (2014). The role of healthcare robots for older people at home: A review. International Journal of Social Robotics 6(4), 575–591.

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), 1556-1559.

Zheng M., J. Wang & Meng, M.Q.-H. (2015). Comparing two gesture design methods for a humanoid robot: Human motion mapping by an RGB-D sensor and hand-puppeteering. Proceesding of 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN), IEEE, 609–614.

Syllable-pointing gesture coordination in Polish counting out rhymes: The effect of speech rate

Katarzyna Stoltmann & Susanne Fuchs

Leibniz-Centre General Linguistics (ZAS)
Schützenstrasse 18, 10117 Berlin, Germany

stoltmann@leibniz-zas.de; fuchs@leibniz-zas.de

Abstract

We investigated the stability of the relationship between number of syllables and pointing gestures under different speech rate constraints. Participants of the study realized Polish counting out rhymes at normal and fast speech rates. Pointing gestures of the apex finger were recorded with a motion capture system and speech acoustics were recorded simultaneously. We hypothesized that for stable coordination of syllable realization and pointing gestures, either movement amplitude would be smaller or peak velocity would be higher at a faster rate. Unstable coordination due to temporal constraints could lead to a coordinational reorganization of speech and finger motions. Results indicate a large degree of stability, which is, however, speaker-specific. Most speakers realize comparable amounts of syllables per pointing gesture under both speech rate conditions. They frequently shorten duration and increase the speed of the pointing gesture.

1 Introduction

Our work is concerned with a deeper understanding of speech production in the context of rhythmic motor actions and their mutual influence (for a general review see Iversen & Balasubramaniam, 2016). Speech and motor actions often flexibly coordinate with another. A common principle to test the stability and flexibility of coordinated behaviour is to apply constraints that may perturb or challenge these actions. For example, Rochet-Capellan & Schwartz (2008) or Lanica & Fuchs (2011) used the speeding paradigm to investigate the coordination between jaw, lower lip and tongue tip movement in CVCV utterances with either a coronal consonant in the first syllable and a bilabial in the second or vice versa. With increase in speed, speakers change from two jaw cycles to one, which in some cases leads to a reorganization of the tongue tip-jaw-lower lip coordination. The reduction in jaw cycles may be the consequence of the dynamic properties of the jaw. It is the heaviest articulator, because it includes bones, while the tongue tip and the lower lip consist of light and flexible soft tissue and their motion is particularly fast (Jannedy et al., 2010). Thus, there are some restrictions on the jaw when it comes to oscillating with a high frequency. Similarly, the dynamics of speech and the arm's pointing gesture movement differ, although they belong to the same body. Since the arm is heavier, arm motions are slower than speech movements. This may be one reason why in some coordinated speech-pointing gesture actions (Rochet-Capellan et al., 2008), the arm motion starts earlier than the speech motion and the faster system (speech articulation) adjusts quicklier at places where cordination among the two is needed (e.g. in the vowel of a stressed syllable). It has further been proposed that certain frequency relations between different motor systems are more optimal than others (e.g., Cummins & Simko, 2009). For speech-pointing gestures, Rochet Cappelan et al. (2007) suggested 2:1 as an optimal ratio and provided first evidence for this. In our own work on German counting out rhymes (Fuchs & Reichel, 2016), we were able to furthermore confirm that the relation between the number of syllables and pointing gestures is stable, though to some extent speaker-specific (see Figure 4 in Fuchs & Reichel, 2016). However, only five speakers were recorded.

Moreover, the relation between the number of syllables and pointing gestures may also be driven by the speech material. If words consist only of one syllable, they may well correspond to one pointing gestures (stroke). If words consist of two syllables, either one or two pointing gestures may be realized. For words consisting of 3 syllables, we would expect either 1, 2 or 3 pointing gestures. Counting out rhymes are a fascinating object of investigation in this respect, because they are a natural testbed for studying this relation. They are often constructed in such a sense that the number of syllables per line changes so that it is rather unpredictable who will win the game and who will be out. Although regional differences exist, widespread knowledge of counting out rhymes appears to exist spanning populations of varying social backgrounds. They are learned in early childhood and allow for investigation of the development of speech with pointing gestures throughout the life span. They exist in many cultures and are part of the oral poetry tradition (Hanna et al., 2002). Cross-linguistic comparisons can be made while taking language specific prosody and linguistic structure into account. The language, we will study, Polish, has predictable stress patterns on the penultimate syllable (Malisz et al. 2013).

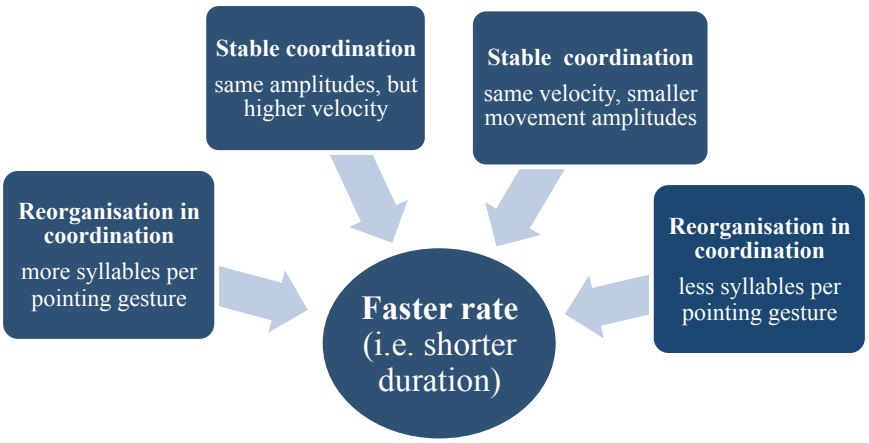


Figure 1. Schematic view of expected speech-pointing gesture relation with increasing speech rate.

Various scenarios for the potential relation between speech and pointing gestures are possible. They are depicted in Figure 1. All of them lead to a faster rate, i.e. a shortening of the temporal window in which the respective counting out rhyme is realized. If the coordination between speech and pointing gestures is stable, we could expect that pointing gestures are reduced in amplitude from one stroke to the next, while peak velocity stays the same. Alternatively, movement amplitudes could also stay the same while peak velocity within a stroke is increased. If the coordination between speech and pointing gesture is more flexible, we expect a reorganization of the two motor actions. Since the speech motor system is a faster oscillating system than the arm motor system, the most economical reorganization would be to produce more syllables per pointing gesture with increasing rate. Another solution could be that speech and hand gestures may reorganize when their relation is not 1:1, i.e. the number of pointing gestures increases so that both systems oscillate with a similar frequency. Such a scenario could be the consequence of a stronger coupling between the two motor systems, similar to the pattern of head motions that has been observed with an increase in speech rate while producing repetitive utterance (Mark Tiede, unpublished data).

2 Methods of data acquisition, annotation and analysis

2.1 Experimental set-up and tasks

Participants stood in front of a chair with a teddy bear and were instructed to play the counting out rhyme game with a teddy bear as a fictitious person. Pointing gestures were measured by means of

a motion capture system (OptiTrack, *Motive* Version 1.9.0) with 12 cameras (Prime 13). Motion data was captured at a 200 Hz sampling frequency and simultaeous acoustic data was captured with 44.1 kHz. 14 markers were placed on a frontlet (one anterior, one posterior and one at the right lateral side), one on the tip of each index finger and one at the bottom of each index finger, one at each wrist, one at each elbow, one at each shoulder (on top of the acromioclavicular ligament), one at the height of the C7 vertebra). An additional marker was placed at the nose of the bear (see Figure 2 for a general overview). The participants' task was: a) to read the rhymes aloud without any hand motion; b) to say the rhymes again while pointing with the nondominant (or dominant hand for the first 6 speakers), d) to point with the dominant hand while saying the rhymes (or the non-dominant one for the first 6 speakers) +e) + f) and finally to repeat the previous two tasks with increased speech rate (time constraints).

2.2 Participants, and speech material

Participants were native speakers of Polish who had been residents of Berlin for no longer than 6 months (most of them were Erasmus students). They were recruited via international contact points at different universities in Berlin. Their age ranged from 21-27 (mean 24.1 years) and all were right handed according to the Edinburgh handedness scale (Oldfield, 1971). Altogether 10 participants were recorded, 7 females and 3 males, however the data of one speaker had to be eliminated due to some technical problems. Participants received 10 € in compensation for their participation.

The speech material consisted of six common Polish counting out rhymes (Table 1). The participants were instructed to familiarize themselves with the rhymes first and then carry out the tasks as described above.

Table 1. Six Polish counting out rhymes with their orthographic representation, number of syllables (syllables are separated by ".") and words. The ratio is calculated as the No. of syllables/No. of words.

Orthographic representation of counting out rhymes	**No. of syll.**	**No. of words**	**Orthographic representation of counting out rhymes**	**No. of syll.**	**No. of words**
Ent.li.czek – pent.li.czek,	6	2	Tre.le.le.le, tre.le.le.le,	8	2
czer.wo.ny sto.li.czek,	6	2	Zja.dłem dzi.siaj trzy mo.re.le.	8	4
na ko.go wy.pa.dnie,	6	3	Raz, dwa, trzy,	3	3
na te.go - bęc!	4	3	Dziś o.bia.du nie jesz ty!	7	5
Ratio = 2	**20**	**10**	**Ratio = 1.86**	**26**	**14**
Raz, dwa, trzy,	3	3	Bzy, bzy, bzy,	3	3
wy.chodź ty,	3	2	By.ły so.bie pszczół.ki trzy:	7	4
jak nie ty, no to ty.	6	6	Ma.ja, Gu.cio, Kle.men.ty.na	8	3
Ratio=1.1	**12**	**11**	I wy.cho.dzisz ty.	5	3
Pan So.bie.ski miał trzy pie.ski,	8	5	**Ratio = 1.77**	**23**	**13**
czer.wo.ny, zie.lo.ny, nie.bie.ski.	9	3	Wpa.dła bom.ba do piw.ni.cy,	8	4
Raz, dwa, trzy,	3	3	na.pi.sa.ła na ta.bli.cy:	8	3
po te pie.ski i.dziesz ty.	7	5	S. O. S. – głu.pi pies.	6	3
Ratio = 1.69	**27**	**16**	Tam go nie ma, a tu jest.	7	7
			Ratio = 1.71	**29**	**17**

2.3 Data pre-processing, speech and gesture annotations

Motion data were exported to the c3d format. Since we did not use a fixed skeleton, the 15 markers were renamed according to their anatomical position using the Biomechanical Toolkit (Barré & Armand, 2014). Hereafter, a velocity vector (v) of the x, y and z time series with a length from 1 to j was calculated as the central difference for index finger (equation 1).

$$v(j) = \mathrm{sqrt}(((x(j+1)-x(j-1))/2)^2 + ((y(j+1)-y(j-1))/2)^2 + ((z(j+1)-z(j-1))/2)^2) \quad (1)$$

The velocity vector was saved in wav-file format and annotated together with the speech wav-file using Praat (version 6.0.26). In the speech wav-file we manually labelled the on- and offset of the

respective counting out rhyme and all silent pauses longer than 100 ms. To calculate speech rate, we divided the number of syllables of the respective rhyme by the summed duration of all speech units without pauses.

Index finger turning points were labelled as velocity minima from the beginning to the end of the counting out rhymes. A stroke was defined as a movement between two successive velocity minima from the speaker pointing towards the teddy bear or back. We removed the first and the last stroke, because – unlike the other strokes – they started or ended in the arm and hand hanging down. For each stroke, the duration, displacement and maximum velocity were calculated.

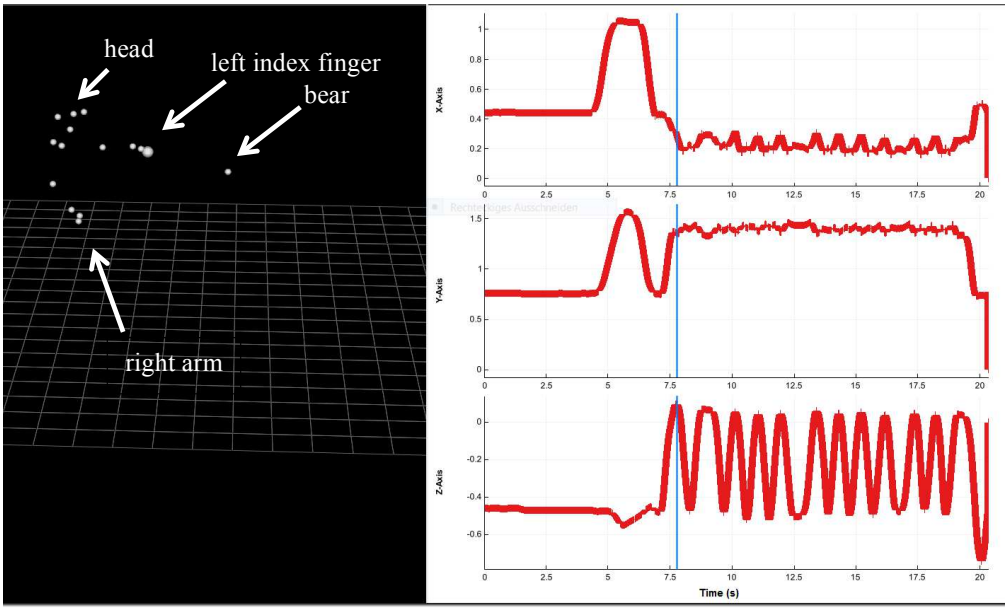


Figure 2. Left: Mokka display of recorded speaker with 15 reflecting markers; Right: x, y and z time series of the dominant index finger.

2.4 Statistical analyses

Since our dataset is balanced, we ran several repeated measures ANOVA to investigate the general effect of CONDITION (normal vs. fast rate) and HANDEDNESS (right hand or left hand) on the ratio of the number of syllables to the number of pointing gestures, the speech rate without pauses, the average stroke duration, average peak velocity and average displacement. Rhymes per speaker were included as an error term. We applied a Bonferroni correction and treated all findings as significant when p was below 0.01 (0.05 / 5, i.e. alpha level of 5% was used and divided by the number of the applied models). All statistical tests were carried out in R (version 3.2.3).

2.5 Results

In order to obtain information about the stability of the relationship between the number of syllables produced and the number of pointing gestures (strokes), we investigated the ratio between the two (Figure 3). Since we did not find any significant differences among the left and right arm, the results are pooled together in the figures. It is evident that speech rate does not lead to a reorganization of the two motor actions and the ratio is rather stable. However, the number of syllables realized within a stroke is highly speaker-specific and ranges from a ratio of 1 (S4, S7) to >2 (S5). Repeated measures ANOVA did not reveal any significant effects for CONDITION, HANDEDNESS or their interaction. Although the ratio is stable, speakers clearly fullfilled the task and increased their speech rate (Figure 4; df=1, F=92.2, $p<0.001$). Moreover, speech rate increasegoes hand in hand with a shorter average duration of the pointing gestures (Figure 5; df=1, F=528.955, $p<0.001$).

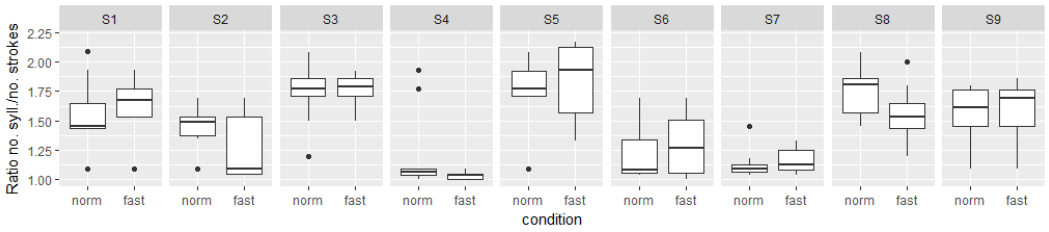


Figure 3. Boxplots showing the relation between number of syllables and number of pointing gestures by speaker in the normal and fast condition.

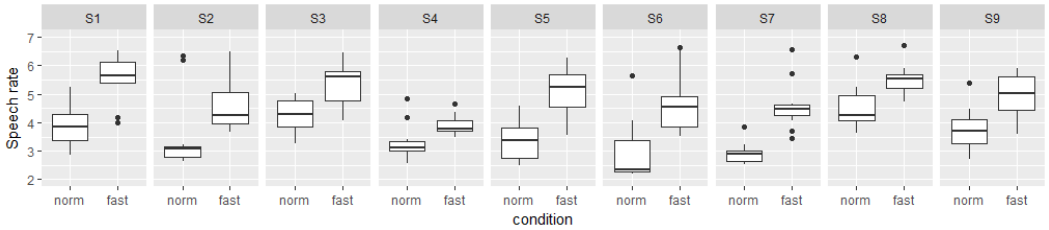


Figure 4. Boxplots showing speech rate (no. of syll/s) split by speaker in the normal and fast condition.

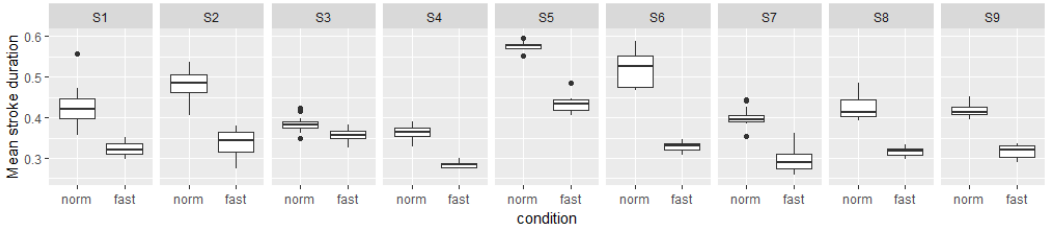


Figure 5. Boxplots showing the averaged stroke duration (s) by speaker in the normal and fast condition.

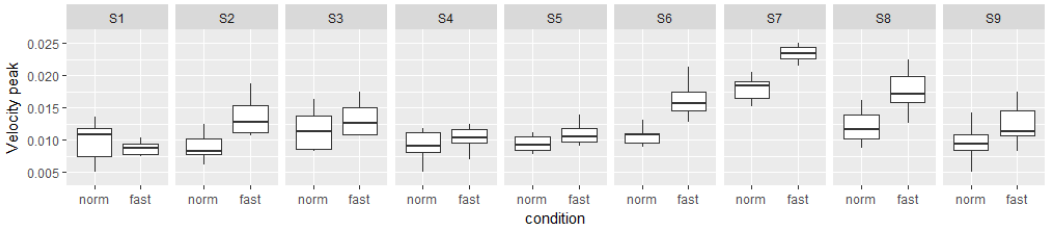


Figure 6. Boxplots showing peak velocity of the stroke split by speaker in the normal and fast condition.

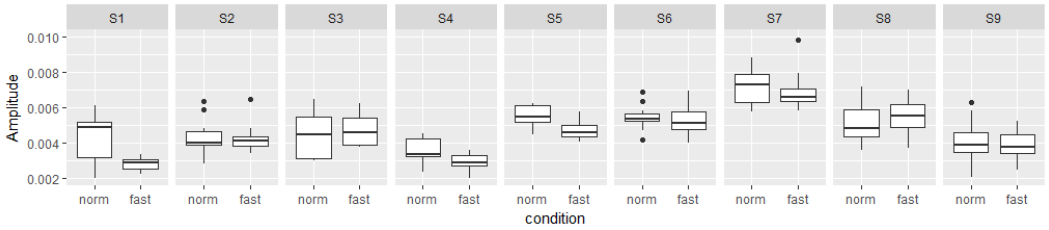


Figure 7. Boxplots showing displacement of the stroke split by speaker in the normal and fast condition.

Finally, pointing gestures are not only shortened, but also produced with a faster velocity of the index finger motion (Figure 6; df=1, F=23.3, p<0.001). Figure 6, however, also displays some speaker-effects. S1, S4 and S5 do not increase the peak velocity substantially, but decrease the distance from one turning point to the next (Figure 7; no sign.).

3 Summary and Discussion

Similar to our previous work on German counting out rhymes (Fuchs & Reichel, 2016), we also found evidence for a relatively stable relation between the number of syllables and pointing gestures for Polish counting out rhymes. The exact relation was, however, speaker-specific. One of the participants of the study even told us spontaneously that he could speak faster, but only without pointing, which further evidences the stable relation and the fact that the slower arm motor system puts some constraints on speech. Hence, physical properties of different motor systems need to be taken into account when studying the relation between speech production and gesturing.

We have also noted some flexibility in rhymes with syllable to word ratios varying considerably from one to the next line. Future work will therefore include a more detailed analyses of each line of the respective rhyme and will carry out a more in-depth coordination analysis.

Stability was occurent independent of whether the dominant or non-dominant hand was used for pointing. In fast speech, most speaker realised similar displacements than in normal speech, but produced a shorter stroke duration and a higher peak velocity. A few speakers kept peak velocity rather stable, but produced smaller movement amplitudes and shorter stroke durations in fast speech. A complete reorganisation was not found, independently of speaker-specificity. In future work, we hope to compare these results with those from different languages.

Acknowledgements

This work was supported by a grant from the BMBF (01UG1411) and the Leibniz Society. We would like to thank Olivia Maky for her help in data preprocessing and annotation, Joerg Dreyer for technical support, our participants and Mark Tiede for his scientific input.

References

Barré, A. & Armand, S. (2014) Biomechanical ToolKit: Open-source framework to visualize and process biomechanical data. *Computer Methods and Programs in Biomedicine* 114 (1): 80-87.

Boersma, P. & Weenink, D. (2017). *Praat: doing phonetics by computer* [Computer program]. Version 6.0.26, retrieved 5 February 2017 from http://www.praat.org/

Cummins, F., & Simko, J. (2009). Notes on phase and coordination, and their application to rhythm and timing in speech. *音声研究*, *Journal of the Phonetic Society of Japan* 13(3), 1-12.

Fuchs, S. & Reichel, U.D. (2016). On the relationship between pointing gestures and speech production in German counting out rhymes: Evidence from motion capture data and speech acoustics. In C. Draxler & F. Kleber (Eds.), *Proceedings of P&P 12*, 1-4. München: LMU.

Hanna, N. A., Patrizia, K. L., & Dufter, A. (2002). The meter of nursery rhymes: universal versus language-specific patterns. *Sounds and systems: studies in structure and change. A Festschrift for Theo Vennemann*, pp. 241-267Berlin/New York: Mouton de Gruyter (Trends in Linguistics).

Iversen, J. R., & Balasubramaniam, R. (2016). Synchronization and temporal processing. *Current Opinion in Behavioral Sciences*, 8, 175-180.

Jannedy, S., Fuchs, S. & Weirich, M. (2010). Articulation beyond the usual: Evaluating the fastest German speaker under laboratory conditions. In S. Fuchs, P. Hoole, C. Mooshammer & M. Zygis (Eds.), *Between the regular and the particular in speech and language*, pp. 205-234. Frankfurt/M.: Peter Lang.

Lancia, L. & Fuchs, S. (2011). The labial coronal effect revisited. In Yves Laprie (ed.), *Proceedings of the ISSP*, 187-194. Montreal: CD-ROM.

Malisz, Z., Żygis, M., & Pompino-Marschall, B. (2013). Rhythmic structure effects on glottalisation: A study of different speech styles in Polish and German. *Laboratory Phonology*, 4(1), 119-158.

Oldfield, R.C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsych.*, 9, 97–113.

R Core Team (2015). *R: A Language and Environment for Statistical Computing* [Computer program]. Version 3.2.3., retrieved October 2015 from https://www.R-project.org

Rochet-Capellan, A., Laboissiere, R., Galvan, A. & Schwartz, J.L. (2008). The speech focus position effect on jaw–finger coordination in a pointing task. *JSLHR* 51.6, 1507–1521.

Mutual visibility and information structure enhance synchrony between speech and co-speech movements

Petra Wagner [a,b] *and Nataliya Bryhadyr* [a]

[a] Faculty of Linguistics and Literary Studies, Bielefeld University
[b]Center of Excellence Cognitive Interaction Technology (CITEC), Bielefeld University
Bielefeld, Germany
petra.wagner@uni-bielefeld.de, n.bryhadyr@uni-bielefeld.de

Abstract

Our study aims at gaining a better understanding of how speech-gesture synchronization is affected by the factors (1) mutual visibility and (2) linguistic information structure. To this end, we analyzed spontaneous dyadic interactions where interlocutors are engaged in a verbalized version of the game Tic Tac Toe, both with and without mutual visibility. The setting allows for a straightforward differentiation of contextually given and informative game moves, which are studied with respect to their manual and linguistic realization. Speech and corresponding manual game moves are synchronized more often when there is mutual visibility and when game moves are informative. Mutual visibility leads to a slight precedence of manual moves over corresponding verbalizations, and to a tighter temporal alignment of speech and co-speech movements. Informative moves counter the movement precedence effect, thus allowing co-speech movement targets to smoothly synchronize with prosodic boundaries.

1 Gesture - Introduction

Previous research has shown that visibility has a strong effect on the frequency of gesture production, especially on *communicative gestures* that aid conversational interaction or highlight information, while *representational gestures*, co-expressing a verbal message's content, are produced similarly frequent both with or without mutual visibility (Alibali et al., 2001; Bavelas et al. 2008). Many other results stress the communicative, listener-oriented function of co-speech gesturing, as speakers adapt their gesture rate based on the knowledge that they are seen, rather than seeing the interlocutor (Mol et al., 2011) and with gestures being larger and less reduced under visibility (Hoetjes et al., 2015). Similarly, de Ruiter et al. (2012) argue that speech and co-speech gestures tend to stand in a redundancy relationship that makes communication, especially conversational *grounding*, more robust.

Unlike frequency and gesture shape, we hitherto know very little about the effect of mutual visibility on speech-gesture synchrony. Speech-gesture synchrony has been pronounced a key feature of co-speech gesturing (McNeill, 1992), but is notoriously difficult to measure, given the variability of potential temporal anchors in the verbal and gestural stream, and the vagueness of lexical affiliates (cf. discussions in Esteve-Gibert and Prieto, 2013; Wagner et al, 2014). For beat gestures and deictic gestures, the gesture apex aligns with accented syllables in the speech stream (e.g. Loehr, 2012; Jannedy and Mendoza-Denton, 2005; Leonard and Cummins, 2010, Esteve-Gibert and Prieto, 2013). Additionally, there is evidence for alignment of gesture and prosodic boundaries (Loehr, 2012, Krivokapic et al., 2017, Jannedy & Mendoza-Denton, 2005). Many studies find that gestures or gesture apices precede speech, but tend to not lag behind (e.g. Esteve-Gibert and Prieto, 2013). Gesture lags are perceived as more asynchronous to speech and impede comprehension (Leonard & Cummins, 2010; Özyürek, 2008), but these effects may be a function of gesture type (Kirchhof, 2017).

If speech and co-speech gesture are as tightly linked as suggested and mainly serve communicative robustness, they should be temporally synchronized to facilitate perceptual integration. We therefore expect that speech and gestures are aligned more precisely given mutual visibility. We furthermore expect that a message's information load enhances this temporal alignment. We test these assumptions in a study where interlocutors are engaged in a verbalized game of Tic Tac Toe (cf. Watson et al., 2007), while simultaneously playing the game on a board with and without mutual visibility. When interlocutors cannot see each other, we expect a certain decoupling of speech and game-related manual movements, as the latter no longer fulfill any communicative function. Despite the redundancy of information conveyed visually and verbally under visibility, we expect an increase in speech-co-speech synchrony, further enhanced by communicative needs such as the highlighting important information. Such highlighting occurs either if a move is unexpected (as in early stages of the game) or game-relevant (as in later stages of the game). In our setting, moves can also be uninformative, e.g. when the game situation inevitably leads to a tie.

We are aware that the manual movements we examined do not qualify as traditional co-speech gestures spontaneously produced alongside with speech. Rather, they are co-speech movements elicited by the task. Still, they obviously fulfill a communicative function in constituting the game moves and are fully co-extensive in semantic content with their corresponding verbalizations.

2 Methods

2.1 Participants

We recorded 20 native speakers of German (10 dyads, friends in their 20s, unpaid volunteers, no control for gender, speakers stayed in one dyad) engaged in a verbalized version of Tic Tac Toe.

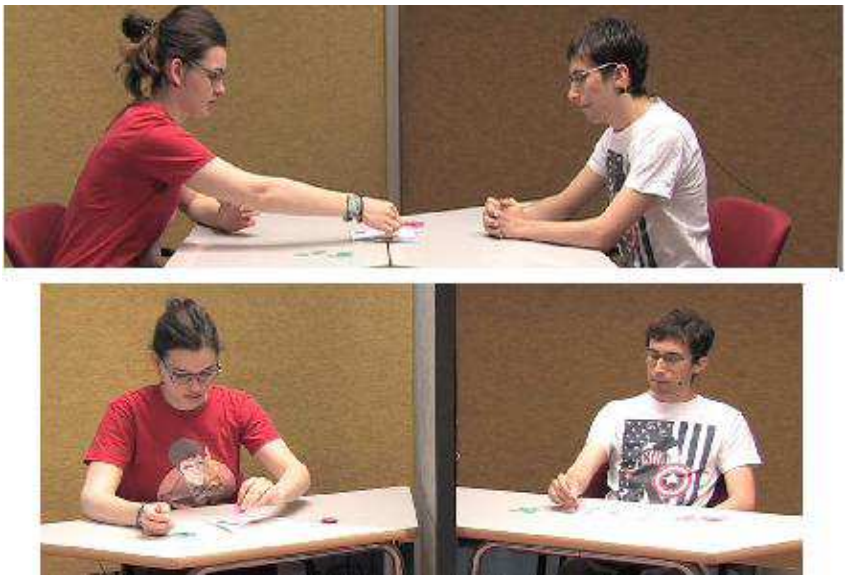


Figure 1: Recording setting with (top) and without (bottom) mutual visibility

2.2 Recording setup

Each dyad was recorded at our faculty's recording studio using Sennheiser neckband microphones (audio) and a studio camcorder (video) in two different recording conditions (cf. Figure 1):

- Visibility Condition: The players were seated facing each other, with a shared TicTacToe game board placed in the middle.

- Invisibility Condition: The players were seated on separate tables and were parted from each other by the movable wall, each of them having his or her own game board.

Each player received a set of cut outs in the form of a tree (ger. "Baum") and a ball (ger. "Ball") to make their moves. To control for order effects, we used an alternating initial recording condition with each newly recorded dyad, and per condition, 4 consecutive games were played. The game board looked like a normal Tic Tac Toe grid, however with every cell being numbered. This enables the interlocutors to unambiguously refer to the different cells on the game board. A typical verbalized move is produced by placing a sentence accent on the target of the move, which corresponds to one of the numbers available on the game board – these accented verbalizations of numbers are later analyzed for their prosodic realization, e.g.

Ich lege einen Baum auf Feld **FÜNF.** (1)
(Engl.: I put a tree on field **FIVE.**)

Prior to each game, the players were informed about a preset first move. Also, the players alternated in setting the first move. On average, each recording session lasted 8.57 minutes per dyad, resulting in roughly 1.5 hours of recorded speech in total.

2.3 Annotations

The verbalized target moves, i.e. the sentence accented number realizations, were manually annotated using Praat (Boersma & Weenink, 2008).

The corresponding manual target moves were annotated with ELAN (Brugman & Russel, 2004) starting from a resting position or a position in which there is no target-oriented move, but in which the players hold the cut-out to be placed in their hand. The moves end with the full contact of the cut out on the game board. In cases where the players hold their target-oriented move before making full contact with the game board, the time of first contact is annotated as ending point of the gestural game move.

As the first move was preset by the experimenter and made known to both players before the game, it is annotated as *given*. In the case of a tie, the last move was annotated as *given*. The remaining moves were annotated as *informative*, as they led to a win, blocked a potential winning move, or were unpredictable (cf. Watson et al., 2008).

2.4 Measuring speech-movement synchrony

We analyzed the synchronization between the points at which the movement targets are reached and two prosodic anchors: pitch accent peaks and prosodic boundaries. As an estimate of gesture-boundary synchrony, we calculated the difference between the delay between the time the gestural moves reached their target on the game board and the corresponding end of a verbalized move, coinciding with a prosodic boundary (= gesture delay to prosodic boundary).

To estimate the synchrony between co-speech movements and pitch accents, we calculated the point of maximal pitch excursion for each verbalized target move using the pitch tracking functions of Praat. We then calculated the delay between pitch accent location and the time of corresponding co-speech movements. (= gesture delay to pitch accent).

Notice that a gesture preceding the prosodic anchor has a negative delay (lead), a gesture succeeding the verbalization a positive delay (lag).

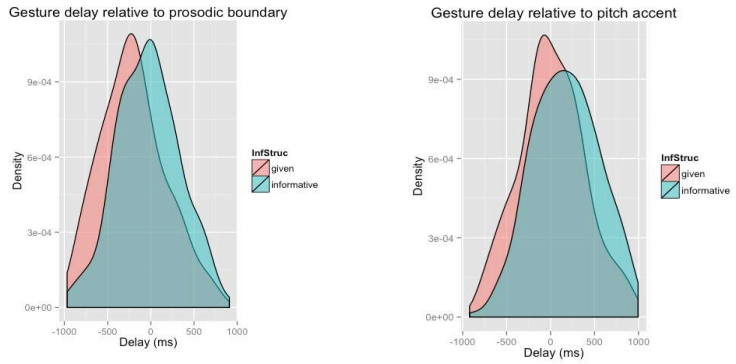


Figure 2. Delay between movement end points, prosodic boundaries (left) and pitch accents (right) in given and informative contexts.

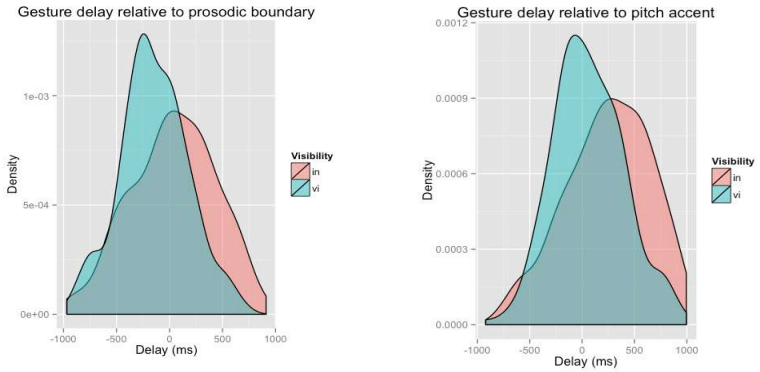


Figure 3. Delay between movement end points, prosodic boundaries (left) and pitch accents (right) with (vi) and without (in) mutual visibility.

3 Results

QQ-plots of the two delay measures (R package "car", Version 2.0–25) showed a violation of normal distribution due to outliers, e.g. instances of clear desynchronization between speech and co-speech movements. Based on the QQ-Plots, we define outliers as delays larger than ±1000ms. A closer analysis revealed that there are significantly more outliers in the absence of mutual visibility (18.6%) as compared to mutual visibility (4.0%, $\chi^2(1)=21.4$, $p<0.001$). We also found that there are significantly more outliers when the moves are not informative (24.7%) as compared to informative moves (8.6%, $\chi^2(1)=9.5$, $p<0.01$). After removing outliers, both types of delays show near-normal distributions. We subsequently constrained our analyses to delays between *-1000ms* (gesture end precedes prosodic boundary/pitch accent by *1000ms*) and *+1000ms* (gesture end follows prosodic boundary/pitch accent by *1000ms*).

The ends of co-speech movements tend to precede corresponding prosodic boundaries (*n=86, M=-204ms, SD=371ms*) when the conveyed information is given as compared to when the message is informative, in which case the gestures reach their target almost perfectly aligned with the point of time where the verbalization is finished, showing only a minimal negative delay (*n=421, M=-30ms, SD=363ms*; cf. Figure 2). There is a similar difference for delays between co-speech movements and pitch accents (cf. Figure 2), with movements being earlier when the expressed information is given and closely aligned with pitch accents (*n=86, M=16ms, SD=371ms*) as

compared to when the message is informative (*n=421, M=176, SD=376ms*). Across conditions, manual movements lag behind corresponding pitch accents, but precede prosodic boundaries.

The ends of co-speech movements precede corresponding prosodic boundaries (*n=271, M=-144ms, SD=320ms*) when there is mutual visibility, but slightly lag behind elsewhere (*n=236, M=37ms, SD=399ms*; cf. Figure 3) There is a similar effect for delays between co-speech movements and pitch accents (cf. Figure 3) with movements being earlier under mutual visibility (n=271, M=63ms, SD=336ms) as compared to lacking visibility (*n=236, M=247, SD=403ms*). However, the co-speech movements generally lag behind their corresponding pitch accents. F-tests to compare variances showed that speech-gesture synchrony is less variable under mutual visibility both relative to prosodic boundaries *(F(270,235)=0.65, p<0.001)* and pitch accents *(F(270,235)=0.70, p<0.01)*. However, gesture-speech synchrony was not affected by information structure.

The data collected in the recordings and subsequent annotations and acoustic measurements were further analyzed with the help of Linear Mixed Models using R (Version 3.1.2) (R Team, 2015) together with the R-packages lme4 (Version 1.1-7) and lmerTest (2.0-25). The resulting models contained the factors visibility (*visible—invisible*) and information structure (*informative— given*) as fixed, and word (1-9) and participant as random factors with random intercepts. Our two measures of delay (prosodic boundary delay, pitch accent delay) served as dependent variables. Both dependent variables were analyzed by reducing a maximal model including all fixed and random effects in a stepwise fashion, i.e. by removing all non-significant main fixed effects and interactions through log-likelihood ratio comparisons. This analysis showed no significant interactions. For the synchronization of gesture and prosodic boundary, a model comparison shows that the full model differs significantly from one not including visibility ($\chi^2(1)=46.5, p<0.001$) or information structure ($\chi^2(1)=21.8, p<0.001$). The model confirms the prior descriptive analyses that informative moves make co-speech movements occur significantly (*t(484)=4.7, p<0.001*) later *(+176ms, SE=37ms*) relative to prosodic boundaries, while visibility leads to a significantly (*t(484.3)=-7.0, p<0.001*) earlier production of co-speech movement *(-196ms, SE=28ms*). The model intercept also shows that on average, gestures precede prosodic boundaries *(-87ms, SE=58ms)* For the synchronization of gesture and pitch accent location, a model comparison shows that the full model including differs significantly from one not including visibility ($\chi^2(1)=46.2, p<0.001$) or information structure ($\chi^2(1)=14.9, p<0.001$). The model confirms the prior descriptive analyses that informative moves make co-speech movements occur significantly (*t(496.3)=3.9, p<0.01*) later *(+149ms, SE=38ms*) relative to corresponding pitch accents, while visibility leads to a significantly *(t(502.3)=-7.0, p<0.001)* earlier production of co-speech movements *(-203ms, SE=29ms)*. On average, gestures reach their targets after pitch accents *(149ms, SE=54ms)*

4 Discussion

Our outlier analysis revealed that clear violations of gesture-speech synchrony (delays larger ±1000ms) occur much more frequently if interlocutors cannot see each other and if the message conveyed is uninformative. This supports our assumption that both information structure and visibility increase speech-gesture synchrony. It also corroborates models claiming that speech-gesture alignment is to some extent caused by communicative needs. The fact that visibility decreases the variability in speech-gesture alignment further strengthens this finding. A possible interpretation is that the stronger "gesture lead" occurring under visibility aids the cross-modal integration on the side of the listener (Leonard & Cummins, 2010). Interestingly, informativity has the contrary effect and makes gestures appear a bit later as compared to when the message is not informative. At first glance, this may endanger the communicative robustness probably achieved by the gesture-lead, at least under visibility conditions. However, a closer look at the distributions of speech-gesture alignment reveals that even though gestures may be later when the conveyed message is informative, they align almost perfectly with prosodic boundaries, but do not make the gesture strongly lag behind speech. Thus, it is unlikely that audiovisual integration on the side of the listener is endangered by this effect of information-structure on speech-gesture alignment. In

line with our assumptions, it rather seems that informativity further strengthens speech-gesture synchronization. If the proximity of speech and co-speech movement is taken as an indicator of an anchor for speech-gesture alignment (though see the critical discussion in Leonard & Cummins, 2010), the pitch accent qualifies as the anchor for gesture apices/movement boundaries under visibility, and the prosodic boundary in case the message is informative. Still, the exact kind of this relationship needs to be investigated further, taking into account global aspects of prosodic shape.

Overall, we have convincing evidence that both visibility and information structure enhance the temporal synchrony of speech and co-speech movements, thus supporting theories claiming that speech-gesture synchrony mainly fulfills a communicative function. However, due to the different distance of players to the game board across the two visibility conditions, our claims need further confirmation in follow-up studies. Also, the influence of the individual interlocutor dynamics (dyads) should be examined further. The co-speech movements we examined are no prototypical, spontaneously produced co-speech gestures, but their form and function closely resemble deictic gestures. At this point, we cannot say whether our results generalize to deictic gestures, or to iconics or emblematics, for which speech-gesture synchrony may be less relevant (Kirchhof, 2017).

References

Alibali, M.W., Heath, D.C., & Myers, H.J. (2001). Effects of visibility between speaker and listener on gesture production: some gestures are meant to be seen. *Journal of Memory and Language 44*, 169–188.

Bavelas, J., Gerwing, J., Sutton, C., & Prevost, D. (2008). Gesturing on the telephone: independent effects of dialogue and visibility. *Journal of Memory and Language 58*, 495–520.

Boersma, P. & Weenink, D. (2008). Praat: Doing phonetics by computer (version 6.0.21). Retrieved from http://www.praat.org. [accessed Apr 7, 2017].

Brugman, H. & Russel, A. (2004). Annotating Multimedia/Multi-modal resources with ELAN. In: Proceedings of LREC 2004, Fourth International Conference on Language Resources and Evaluation, Lisbon, Portugal. Retrieved from: http://tla.mpi.nl/tools/tla-tools/elan/

Jannedy, S. & Mendoza-Denton, N. (2005). Structuring information through gesture and intonation. *Interdisciplinary Studies on Information Structure 3*, 199–244.

Esteve-Gibert, N. & Prieto, P. (2013). Prosodic Structure Shapes the Temporal Realization of Intonation and Manual Gesture Movements. *Journal of Speech, Language, and Hearing Research 56*, 850–864.

Hoetjes, M., Koolen, R., Goudbeck, M, Krahmer, E., & Swerts, M. (2015). Reduction in gesture during the production of repeated references. *Journal of Memory and Language 79–80,* 1–17.

Kirchhof, C. (2017). *The shrink point: audiovisual integration of speech-gesture synchrony*. Bielefeld: Universität Bielefeld. Retrieved from: https://pub.uni-bielefeld.de/publication/2908762.

Krivokapić,J.,Tiede, M., & Tyrone, M. (2017). A Kinematic Study of Prosodic Gesture in Articulatory and Manual Gestures: Results from a Novel Method of Data Collection. Laboratory Phonology. *Journal of the Association for Laboratory Phonology 8*(1), 3. DOI: http://doi.org/10.5334/labphon.75 [accessed Apr 7, 2017].

Leonard, T. & Cummins, F. (2010). The temporal relation between beat gestures and speech. *Language and Cognitive Processes 26 (10),* 1457–1471.

Loehr, D. (2012). Temporal, structural, and pragmatic synchrony between intonation and gesture. *Laboratory Phonology*. *Journal of the Association for Laboratory Phonology 3*, 71–89.

McNeill, D. (1992). Hand and Mind: What Gestures Reveal about Thought. University of Chicago Press, Chicago.

Mol, L., Krahmer, E., Maes, A., & Swerts, M. (2011). Seeing and Being Seen: The Effects on Gesture Production. *Journal of Computer-Mediated Communiation 17(1),* 77–100.

Özyürek, A., Willems, R.M., Kita, S. & Hagoort, P. (2007). On-line integration of semantic information from speech and gesture: insights from event-related brain potentials. *Journal of Cognitive Neuroscience 19 (4),* 605–616.

de Ruiter, J.P., Bangerter, A., & Dings, P. (2012). Interplay between gesture and speech in the production of referring expressions: investigating the trade-off hypothesis. *Topics in Cognitive Science 4 (2),* 232–248.

R Team (2015). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org.

Wagner, P., Malisz, Z., & Kopp, S. (2014). Gesture and speech in interaction. *Speech Communication 57*, 209–232. Watson, D.G., Arnold, J.E., & Tanenhaus, M.K. (2008). TicTacTOE: Effects of Predictability and Importance on Acoustic Prominence in Language Production. *Cognition 106(3),* 1548–1557.

Mismatches between verbal and nonverbal signs, observing signs of change

Orit Sônia Waisman

David Yellin College, Jerusalem
11 Givat Haionim St.

orito@dyellin.ac.il

Abstract

This article discusses the role of mismatches between verbal and nonverbal signs in spoken language. The literature establishes that mismatches enhance learning and play a prominent role in gesture studies. In a semiotic sign-oriented study of conflict talk, mismatches between verbal and nonverbal signs pinpoint salient situations in the text (Waisman, 2010). This article calls for a broader theoretical framework in order to explore further the role of mismatches in conflict talk and other discourses.

1 Introduction

Susan Goldin-Meadow and colleagues studied the phenomenon of mismatches extensively over several decades, concentrating mainly on child learning processes. Goldin-Meadow concludes that mismatches between words and gestures function as an index of transitional knowledge and that mismatching enhances the child's achievements. (Church & Goldin-Meadow, 1986; Gather et al., 1998; Goldin-Meadow, 2003). Furthermore, she found that mismatches cluster at choice points, such as solving puzzle problems and suggests that speech-gesture mismatches are "the very best place to look for the effects of gestures on listeners…" (2003, p.83).

In a similar vein, Spencer Kelly and colleagues concluded that mismatches that complicate communication catalyze learning abilities:

> For quickly and accurately understanding messages in the moment, congruencies between gesture and speech are by far the most effective; but for shaking up knowledge states and prodding learners over time, a certain degree of incongruence between gesture and speech (mismatches) may be optimal. (Kelly et al. 2010, p. 266)

Mismatches were commonly used to study relations between language and gesture (Cassell, et al., 1999; Kelly, 2017; McNeill, 1992). David McNeill, analyzing mismatches in the context of the relationship between language and thought, concluded that gesture and language are controlled by a single integrating system. McNeill found that mismatching affects listeners to a surprising degree, even when they are not aware of them (1992). Furthermore, McNeill et al. proposed classifying mismatches according to three different kinds: manner, perspective and anaphor mismatches (1994). He concluded that "gesture and speech can combine into a single meaning" (McNeill et al., 1994, p. 235) and postulated that listeners aim to integrate the mismatched message they receive. One of McNeill's mismatch experiments revealed that listeners were prepared to go to a "quite radical length" to avoid conflict between modalities and apprehend a message that avoids contradiction. In fact, the speaker's gestures were unconsciously taken in by listeners aiming to recover the conveyed meaning (McNeill 1992, p. 143).

Spencer Kelly et al. proposed distinguishing between weakly and strongly incongruent mismatches (2010). They found that mismatches, which they call "incongruences", take longer to process neurologically: "incongruent iconic gestures make people slower and less accurate, while congruent gestures make people faster, while processing speech" (Kelly et al., 2010, p. 8). Furthermore, Kelly recently concluded that:

> deictic gestures have a bi-directional relationship with speech during pragmatic processing. Without speech, the intended physical referent of a deictic gesture can be quite ambiguous. For example, understanding the intention behind pointing to an open window is easier in the presence vs. absence of saying 'It's getting cold in here.' In this way, speech adds 'news' to gesture as much as gesture contributes to speech. (Kelly, 2017, p. 13)

This statement reinforces the need for further exploration of mismatches between deictic gestures and their accompanying verbal content.

Adam Kendon's studies reveal the intricate coordination between what he named "extra-oral visible bodily signs" and verbal signs from the onset of life, using them to try to unravel the origin of language:

> [t]he meanings of visible bodily action unfold in time in a coordinated relationship and if the meanings of speech and the meanings of these bodily movements are taken together, a richer and more complex expression may be appreciated than either words or extra-oral actions are considered separately. (2014, p. 67)

He consequently questions the reason for this coordination. I propose that studying mismatches between these modalities may provide a good source for deeper understanding of this coordination. Kendon sums up his analysis in the following statement: "the 'natural' state of spoken language is a speech-kinesis ensemble" (2014, p.76).

2 Main research

I examined cases of gesture-word mismatches between verbal and nonverbal signs in conflict situations. The data comprises a series of videotaped sessions between Israeli-Arab and Israeli-Jewish students at Ben Gurion University of the Negev during one academic year. These meetings were held weekly, as partial requirements of an academic course. The videotaping was done through a one-way mirror, resulting in some 30 hours of filmed discussions and the material was subsequently transcribed, resulting in a text of approximately 1,000 pages, which serve as the corpus data.

2.1 Methodology

The first stage of analysis consisted of viewing the 16 videocassettes, which contained all of the data, and examining both verbal and non-verbal information. The aim of this process was to form a gestalt of the verbal and non-verbal aspects of the text in order to become familiar with it, but without attempting to answer any particular preconceived research question. This procedure is consistent with the phenomenological approach presented by Giorgi (1975), who emphasizes the necessity of first approaching the data with maximum openness without taking into account the specific aim of the study.

My first observation was that there is a constant struggle over space in these conversations, which intensifies and becomes more acute at moments of conflict. This struggle is expressed in both verbal and non-verbal modalities, such that during conflict there are more frequent occurrences of someone touching someone else; there are more cases of people extending their arms and reaching out to the center of the circle or toward the other; there is greater eye contact; and there are more instances of speakers detaching their torso from the chair and leaning toward their addressee. In such moments, participants also raise their voices more often, increase the speed of their speech, and repetition of questions increases. These phenomena indicate that when conflict arises, it involves a struggle over who gets to be seen and heard. The struggle over space in group discussions in a controlled environment paralleled the aspects of struggle over space that characterize the Arab-Jewish conflict. I hypothesized that a study of these discussions and these episodes of conflict could shed light on characteristics of the Arab-Jewish conflict.

Following this observation, I narrowed my focus to three aspects of the verbal modality and three of the non-verbal modality, which reflect the speaker's use of space. My study of these discussions focused on the verbal aspects of *personal deictics, temporal deictics*, and *markers of space*, as well as on the non-verbal gestures of *eye contact*, *movement of torso*, and *hand movement*, which are all mainly pointing gestures. I examined the discussions both in situations of conflict and in situations of non-conflict and marked them accordingly. This procedure proved fruitful. Once my attention was narrowed to a limited number of parameters, I noticed that sometimes during conflict, a clear mismatch occurred between certain verbal and non-verbal signs. In contrast, when there was no conflict between the speakers, no mismatch appeared.

The mismatches were subsequently analyzed using the theory of Phonology as Human Behavior (PHB) of the Columbia School (CS) sign-based theory of linguistics (e.g. Davis, 2006; Tobin, 1990; Tobin & Schmidt, 2008). This theory stipulates that "successful linguistic communication is achieved only through the combined effort of an encoder and a decoder cooperating together" (Tobin, 2009, p. 331). Afterwards, the theory of word systems (Aphek & Tobin, 1988; Tobin & Schmidt, 2008) was used to analyze the texts surrounding the mismatches.

> The concept of text we are using here cuts through an entire single discourse, or even a set of either spoken and written discourses…. [T]here can be more than one word system in a text and each of the word systems nurtures the theme and message of a text with a great intensity….
>
> (Aphek & Tobin, 1988, p. 4).

While these approaches were originally designed for the exclusive study of linguistic signs, the present study includes bodily signs in its analyses. The sign-oriented semiotic approach is based on the definition of language as a "system of systems composed of various sub-systems (revolving around the notion of the linguistic sign) which are organized internally and systematically related to each other and used by human beings to communicate" (Tobin, 1995, p. 7). Contini-Morava, of the Columbia School of linguistics (CS), explains the emphasis of sign-based schools of grammar on context: "…a major point of divergence between sign-based theories of grammar and mainstream generativists theories is the definition of the object of study itself… an emphasis on observing natural discourse" (Contini-Morava, 1995, pp. 3-4). Sign-oriented semiotics replaces the traditional sentence-oriented and historical or philological approaches. Traditional grammatical sentence-oriented concepts are replaced with signs and the basic unit of analysis may include linguistic units "of all sizes and levels of abstractness" (Tobin, 1990, p. 30). Signs may replace autonomous levels of morphology and syntax, as well as the traditional distinction between grammar and lexicon.

Semiotics includes visual and verbal signs, as they form code systems that systematically communicate information or messages. The sign-based theory of linguistics postulates that through identification and analysis of the distribution of signs, the veiled messages of the text are exposed, as that distribution is non-random. The sign-oriented approach stipulates a "force of unification" (Zipf, 1949, in Tobin, 2009, p. 331), from the point of view of the encoder, namely, the desire to achieve maximum communication with minimal effort, taking into account the different roles of encoders and decoders that are necessary for efficient communication. Both the production and the perception aspects of human speech are viewed together as part of the same integral process referred to as the speech chain (Tobin, 2009, p. 334). Linguistic economy is best realized by a highly compressed lexicon composed exclusively of short words conveying a multiplicity of messages. The underlying supposition is that greater cooperation between encoder and decoder increases the chance of "successful communication" (Tobin, 1990, p. 59). This approach stipulates that absolute synonyms in language do not actually exist. As language would not tolerate redundancy, each sign has a purpose in its system (also referred to as the "one-form one-meaning" principle (Contini-Morava, 1995, p. 8). These concepts are coherent with approaches presented earlier in this article, like that of Adam Kendon (2014).

3 Results of the study

Mismatches between deictic gestures and verbal signs were found only in situations of conflict. For my purposes, I considered a mismatch to occur when a pointing gesture implied differently than the verbal personal or proximal and remote deictics (mainly 'you' and 'there') (Waisman, 2010, p. 34). Each mismatch was analyzed in a form I devised, which included references to both verbal and nonverbal information about the mismatch and to their surrounding texts, such that signs were identified and studied quantitatively. This revealed three word systems that represent concepts that are cardinal to the conflict, as follows:

The *beten* (belly) represents the physical, emotive, and corporal site of the conflict. This is where the conflict is felt by those involved in it; it is embodied there. The *beten* is the bodily site of the conflict, where it occurs.

The *medinah* (state)/*adamah* (land) represents the actual geographical site of the conflict and the struggle over space.

The *Shoah* (Holocaust)*/Nakba* (Catastrophe) represents the symbolic and historical site of the conflict. The Arab-Jewish conflict cannot be separated from its historical roots. This connection to the symbolic provides a broader perspective of the conflict.

These word systems combine both modalities and throw a spotlight on the cardinal areas of the Arab-Jewish conflict, revealing its language and the central issues underlying the complex discussions. The Jewish people's struggle to attain their own state is intertwined with the Palestinian Arab people's struggle over their land. The land is perforated by history and tears, and the feelings, the pain and anguish, are felt in the stomach, the belly, the gut. All this is situated against the historical and emotional background of the Holocaust and the Nakba.

My study proposes an alternative and innovative approach to comprehending the characteristics of this impassioned conflict between Arabs and Jews. Although the study focuses primarily on the communicative strategies applied during conflict, its conclusions inevitably contribute to an understanding of the nature of this particular conflict and of conflict between groups in general.

The following is an example of a mismatch between verbal and nonverbal data. Ibrahim (pseudo name) is a member of the Arab group. (The text in bold marks when the speaker touches his belly. The underlined word "he" marks the mismatched word).

Ibrahim: "…this is, like, **my dilemma** (touches his belly) all the time, but I say, after what's happened in **the Holocaust**, maybe it justifies the existence of a state for the Jews. I saw an old man who said: 'The state should be protected.' **Like, um…** **he** (points to his own belly, mismatching) as far as he's concerned, after all he's been through, the very difficult things that he's been through, this is his conclusion…." The mismatch here obtains between the verbal deictic (he) and the non-verbal sign (the speaker pointing to his own belly).

This is an emotive text that reflects the complex situation. Ibrahim non-verbally marks his belly several times, pointing to and at times touching it. This repeats later when he says: "this is, like, my dilemma all the time." Indeed, Ibrahim metaphorically, physically, opens up his belly when delving into this dilemma. This gesture reinforces the nature of this dilemma. Ibrahim attempts to be empathetic toward an old man, a Jewish Holocaust survivor. In a gesture that confuses his own identity with that of the old man, he points at his own belly while verbally referring to his counterpart. He thus mismatches the third person singular deictic (he) immediately after stressing the importance of the State of Israel. Ibrahim finds himself in an awkward position: on one hand, he sympathizes with the Jews; but on the other, he attempts to differentiate himself and his people from the role of the perpetrator, in this case the Nazi, without losing his own case as a victim.

Goldin-Meadow proposed the following: "a mismatch reflects the fact that the speaker is holding two ideas in mind" (2003, p. 130). Accordingly, in our case, the two ideas are the two identities, that of the speaker and that of the referent.

This short analysis presents how the exploration of mismatches exposes layers of the text and reveals issues of subjectivity and identity that may become blurred at times of conflict.

4 Discussion

"The phenomena of speech and co-gesture 'mismatch' is understudied, under-theorized and under-defined…" (Cuffari, 2011, p. 219).

The study of mismatches between verbal and non-verbal data provides a rich source of information for any field that deals with human expression. Mismatches may be posited in the same literary domain as metaphors, and may function similarly. This is coherent with the assumption that mismatches have to do with creativity and emotion (Waisman, 2014). It might be worth exploring the idea of a parallel with the broad concept of a mismatch and the imagery-language dialectic in the growth point (McNeill, 2000). From this perspective, mismatch is at the core of all speech. The mismatches in this case are between modes of semiosis, language-form on one side, imagery on the other, and unified in the growth point. The unity is what sets the process in motion. Considered more broadly, opposites are an instance of mismatches. Moments where harmony is disrupted are opportunities for allowing new material to be integrated, for new learning to occur. Cases of mismatches between verbal and non-verbal modalities provide a rich and promising area of research for diverse fields.

References

Aphek, E., & Tobin, Y. (1988). *Word systems in modern Hebrew: Implications and applications* (Vol. 3). The Netherlands: E. J. Brill.

Cassell, J., McNeill, D., & McCullough, K. E. (1999). Speech-gesture mismatches: Evidence for one underlying representation of linguistic and nonlinguistic information, *Pragmatics & cognition*, *7*(1), 1-34.

Church, R. B., & Goldin-Meadow, S. (1986). The mismatch between gesture and speech as an index of transitional knowledge, *Cognition*, *23*(1), 43-71.

Contini-Morava, E. (1995). Introduction: on linguistic sign theory. In Contini-Morava, E., & Goldberg, B. S. (Eds.), *Meaning as Explanation: Advances in linguistic sign theory* (Vol. 84). (1-42). Berlin: Walter de Gruyter.

Cuffari, E. (2011). Mining the mismatch, an essay review. *Gesture*, 11 (2), 219-231.

Davis, J., (2006). Introduction: Consistency and Change in Columbia School Linguistics, In Gorup, R. J., & Stern, N. (Eds.), *Advances in functional linguistics: Columbia school beyond its origins* (Vol. 57). (1-15). Amsterdam/Philadelphia: John Benjamins Publishing company.

Gather, P., Alibali, M. W., & Goldin-Meadow, S. (1998). Knowledge conveyed in gesture is not tied to the hands. *Child development*, 69(1), 75-84.

Giorgi, A. (1975). An application of phenomenological method in psychology. *Duquesne studies in phenomenological psychology*, 2, 82-103.

Goldin-Meadow, S. (2003). *Hearing gesture: How our hands help us think*. Cambridge; Harvard University Press.

Kelly, S. D., Özyürek, A., & Maris, E. (2010). Two sides of the same coin: Speech and gesture mutually interact to enhance comprehension. *Psychological Science*, 21(2), 260-267.

Kelly, S. D. (2017). Exploring the boundaries of gesture-speech integration during language comprehension. In Breckinridge Church, R., Alibali, M.W. & Kelly, S.D. (Eds.), *Why Gesture? How the hands function in speaking, thinking and communicating.* (243-265). Amsterdam/Philadelphia: John Benjamins Publishing company.

Kendon, A. (2014). The 'poly-modalic' nature of utterances and its relevance for inquiring into language origins. In Dor, D., Knight, C. & Lewis, J. (Eds.) *The social origins of language*. (67-76). Oxford: Oxford University Press.

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. Chicago: University of Chicago Press.

McNeill, D., Cassell, J., & McCullough, K. E. (1994). Communicative effects of speech-mismatched gestures. *Research on language and social interaction*, 27(3), 223-237.

McNeill, D. (2000). *Language and gesture* (Vol. 2). Cambridge: Cambridge University Press.

Tobin, Y. (1990). *Semiotics and linguistics*. London/New York: Longman Pub Group.

Tobin, Y. (1995). *Invariance, markedness and distinctive feature analysis: A contrastive study of sign systems in English and Hebrew* (Vol. 111). Amsterdam/Philadelphia**:** John Benjamins Publishing Company.

Tobin, Y., & Schmidt, J. I. (2008). The Language of Paradox: Interpreting Israeli Psytrancer's Unspoken Discourse. *Israel Studies in Language and Society*, 1, 1-97.

Tobin, Y. (2009). Phonology as human behaviour: Clinical phonetics, phonology and prosody. *Poznań Studies in Contemporary Linguistics*, 45(2), 327-352.

Waisman, O. S. (2010). *Body, language and meaning in conflict situations: A semiotic analysis of gesture-word mismatches in Israeli-Jewish and Arab discourse* (Vol. 62). Amsterdam/Philadelphia**:** John Benjamins Publishing.

Waisman, O. S. (2014). Mismatches as milestones in dance movement therapy. *Body, Movement and Dance in Psychotherapy*, 9(4), 224-236.

The interaction between syntactic encoding and gesture: The case of the double object construction and its prepositional paraphrase

Suwei Wu[1], Alan Cienki[1, 2]

[1] Faculty of Humanities, Vrije Universiteit, Amsterdam, the Netherlands

[2]Multimodal Communication and Cognition Lab, Moscow State Linguistic U., Moscow, Russia

s2.wu@vu.nl, a.cienki@vu.nl

Abstract

There is an increasing interest in the question as to which aspects in speech are tightly coupled with co-speech gesture. A general consensus now exists that gestural representation is influenced by event structure, but it is less clear to what extent gesture interacts with syntactic encoding. The present study thus investigates this issue, taking the double object construction and its prepositional paraphrase (both express transfer events) in relation to gesture as the starting point. Based on data from talk show programs, gestural representation was found to differ with respect to the choice of syntactic encoding of the transfer events.

1 Introduction

A substantial body of gesture research has been considering which aspects in speech are coordinated with gesture, with a particular focus on representational gestures. Previous studies have found that the type of referent and the type of events (that is, event structure) play a role in gestural representation (Lis, 2012; Parrill, 2010). However, the extent to which gestural representation is coordinated with the syntactic encoding of an event is still open to debate. Indeed, an investigation of the interaction between syntactic encodings and gesture could shed more light on the cognitive processes involved in gesture production.

Parrill (2010) has addressed this question by examining gestures accompanying transitive and intransitive constructions. It was found that speakers tend to make gestures using a Character Viewpoint (e.g., miming an activity) when they use the transitive construction in speech, whereas they tend to make gestures of the Observer Viewpoint (e.g., tracing a trajectory movement or embodying an entity with the hands) when using the intransitive construction. However, in that study, different syntactic structures usually concerned different events as well, such as transitive structures encoding handling events and intransitive structures encoding motion events. Thus, it remains unknown whether gesture would coordinate with the syntactic structures or simply the types of events.

Kita et al. (2007) furthermore investigated the difference in gestural representation together with different syntactic encodings of the same motion events, e.g., *he rolled down the hill* versus *he went down as he spun.* This study controlled the type of events used and identified the exclusive role of syntactic packaging in gesture. The two syntactic constructions used, however, concerned one clause or two clauses. The question remains open as to whether or not gesture would be coordinated with the syntactic structures which are within one single clause (that is, with relatively closer and less discrete syntactic forms). Answering this question would help to understand to what extent gestural representation is sensitive to subtle syntactic differences.

The present study specifically considers the double object construction (abbreviated as DOC) and its prepositional paraphrase (also referred to as the prepositional object construction, abbreviated as POC) in English, e.g., *she gave me a book* and *she gave the book to me.* This set of syntactic structures is referred to as the *to* dative alternation. This alternation is of particular interest due to the fact that these two syntactic structures involve relatively closer forms (that is, both concern one single clause and the same type of syntactic transitivity) than the aforementioned two syntactic

structures do. More importantly, the two syntactic structures can encode the same type of events: transfer events. In this light, the exclusive relation of syntactic encoding to gestural representation would be able to be figured out.

2 Method

2.1 The database

The study was based on talk show programs in American English from the archive of the Distributed Little Red Hen Lab, co-directed by Mark Turner and Francis Steen (https://sites.google.com/site/distributedlittleredhen).These programs (total=14) were collected and then created as a corpus for use in the present research, which concerns 122,249,223 words in 14,320 texts.

2.2 Data collection

Verb selection. The present study concentrates on the central and concrete senses of the dative alternation, specific to four verbs frequently discussed in the linguistic literature, each from a major class defined in Levin (1993) that is used with the dative alternation: *give*, *send*, *throw*, and *bring*. Metaphorical uses of these verbs fall outside the scope of the present study, since these cases concern a different and more complex story than concrete uses do, according to Goldberg (1995, p. 89).

Data retrieval. Through a syntactic search interface[1], the uses of the double object construction and the prepositional paraphrase of these verbs were retrieved in the corpus of talk show programs. For instance, a syntax for retrieving the dative uses of the verb *give* in the corpus was as follows: [give,gave,gives,given] ((_{DET})? ((_{ADV})? _{A})* _{N}| _{PRON}) to ((_{DET})? ((_{ADV})? _{A})* _{N}| _{PRON}). Note that, since the progressive or non-progressive aspect has been found to play a role in the co-speech gestural representation (Duncan, 2002), this factor was controlled for in the present study in order to preclude the possibility that the prospective result might be caused by the use of different aspects. Thus, the uses of progressive aspectual forms of these verbs were excluded from the retrieval.

Data sampling. Since too many cases were yielded by the corpus search, 1000 cases were randomly selected from each case in the above except the double object constructional use of the verb *throw*, which occurred only 167 times in the corpus. Subsequently, 4,000 cases in total were sampled for the double object constructional uses of these verbs and 3,167 cases for the prepositional object constructional uses of these verbs.

Noise filtering. "Noises" in the data were manually excluded from the samples. They mostly consisted of three types. A) The first type concerned parsing errors. B) The second type concerned cases which do not belong to the basic and concrete uses of the constructions, such as "X give an idea to somebody". C) The third type involved cases in which hands were invisible on the screen and cases which would prevent speakers from using their hands, such as when speakers were doing real actions with objects in hand.

Consequently, 322 clauses of the double object construction and 323 with the prepositional object construction were ready for the examination of gestures.

2.3 Gesture coding

The present study considers the following aspects of the gesture: the number of representational gestures produced, the depicting techniques used (Modes of Representation, see below), and the more fine-grained form parameter of gestures of the Acting mode. These aspects were coded as follows.

[1] https://corpora.linguistik.uni-erlangen.de/newsscape/newsscape/.

Firstly, representational gestures accompanying the two constructions were identified (N=196). Secondly, these representational gestures were coded in terms of the Modes of Representation. Following Müller (1998), the Modes of Representation consist of the following categories: Acting (N=125), in which a speaker moves as if he/she is miming an activity; Tracing (N=29), in which a speaker moves the hands or fingers as if to trace a line, like the outline of an object or a path of a motion; Molding (N=36), in which a speaker moves as if to mold, touch, or feel the shape of an object; and Embodying (N=3), in which a speaker uses the hands to represent an object. In terms of the Acting gestures, they were coded for whether they involved transferring trajectory movements or not. Whenever a speaker made a forward/backward movement or left/right movement from the rest position (or the hold phase of the previous gesture) in a horizontal or sagittal axis, the gesture was coded as an Acting gesture with transfer movement (N=107). Otherwise, it was coded as an Acting gesture without transfer movement (N=15).

3 Results

The analysis consists of the amount of representational gestures, the Modes of Representation, and the dynamic type of gestures of the Acting mode accompanying the double object construction and the prepositional object construction expressing the transfer events, e.g., *he gave his son a book* and *he gave the book to his son*.

3.1 Representational gestural rates

The dataset contained 322 clauses of the double object construction and 323 clauses of the prepositional object construction. As shown in Figure 1, although the representational gestural rate of the double object construction was slightly lower than that of the prepositional object construction (26.71% and 34.06%, respectively), this difference is indeed not statistically significant ($p>0.05$, $\chi^2=3.7754$, $df=1$). In other words, the number of representational gestures produced does not seem to be sensitive to the difference of the two syntactic structures.

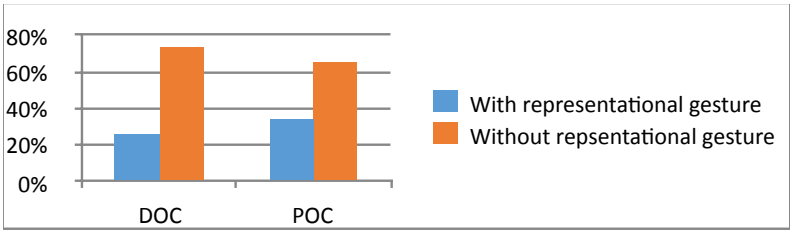


Figure 1. The representational gesture proportions of the double object construction (DOC) and the prepositional object construction (POC).

3.2 Modes of Representation

In the dataset, there were 86 representational gestures accompanying the double object constructions and 107 representational gestures accompanying the prepositional construction. As shown in Figure 2, both constructions were frequently accompanied by gestures of the Acting mode – 50% and 76.64% respectively. An examination of gestures of each mode of representation in relation to the two constructions yielded the following results. Firstly, gestures of the Molding mode were overwhelmingly more likely to accompany the double object construction (38.37%) than the prepositional object construction (2.8%). This difference is statistically significant ($p<0.001$, $\chi^2= 37.443$, $df=1$). Secondly, gestures of the Acting mode are slightly more likely to co-occur with the prepositional object construction (76.64%) than with the double object construction (50%). This difference is statistically significant as well ($p<0.001$, $\chi^2=13.679$, $df=1$). Gestures of the Tracing mode held the pattern of those of the Acting mode. In other words, gestures of the Tracing mode were also more likely to co-occur with the prepositional object construction (20.56%) than with the double object construction (8.14%). This difference is statistically significant as well ($p<0.05$, $\chi^2=4.8296$, $df=1$). In addition, gestures of the Embodying mode were rarely used in terms of both types of constructions. In all, although both constructions in speech

were frequently accompanied by gestures of the Acting mode, gestures of the Molding mode "preferably" co-occurred with the double object construction, while those of the Acting and Tracing modes slightly "preferred" to accompany the prepositional object construction.

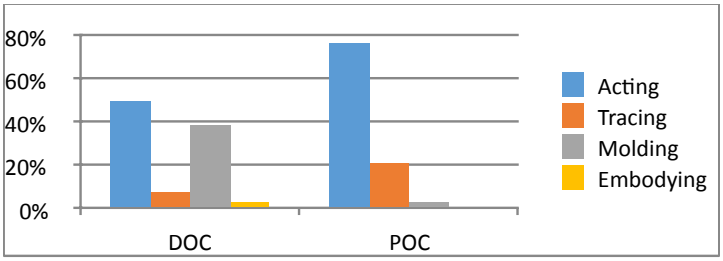


Figure 2. Proportions of various gestural Modes of Representation accompanying the double object construction (DOC) and the prepositional object construction (POC).

3.3 Dynamics of gestures of the Acting mode

The data consisted of 41 Acting gestures accompanying the double object construction and 81 Acting gestures accompanying the prepositional object construction. In Figure 3, we can see that both grammatical constructions were frequently accompanied by Acting gestures with the transfer trajectory movement: 67.44% and 95.12%. Yet a further analysis revealed that Acting gestures with transfer movement were indeed slightly more likely to accompany the prepositional object construction than the double object construction. On the contrary, Acting gestures without any transfer movement were more likely to accompany the double object construction than the prepositional object construction (43.23% and 5.06%). This difference is statistically significant ($p<0.001$, $\chi^2= 14.212$, $df=1$). In all, although there is no clear allocation of the type of dynamics of the Acting gestures in terms of the two constructions, Acting gestures with transfer movement slightly preferred to co-occur with the prepositional object construction.

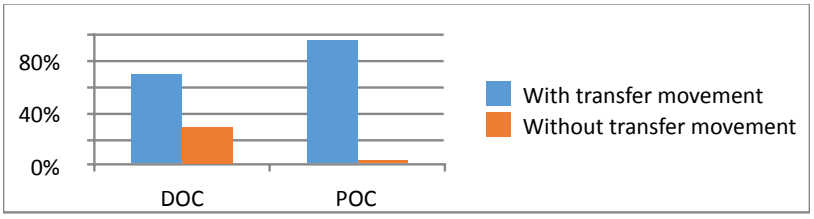


Figure 3. Gestures of the Acting mode with/without transfer movement in terms of the double object construction (DOC) and the prepositional object construction (POC).

4 Discussion and conclusion

The present study aimed to investigate whether representational gestures would vary depending on the choice of syntactic encodings of the same events (transfer events): the double object construction and the prepositional object construction. It was found that both syntactic constructions were accompanied by a similar number of representational gestures. In addition, both syntactic constructions were frequently accompanied by gestures of the Acting mode, in particular those with transfer movement. These appear to indicate that the amount of gesture and the dominant mode of representation as well as the type of dynamic movement in the Acting gestures are linked with the event structure-transfer events, more than with the syntactic forms encoding the events. This is compatible with the prevalent view that the gestural representation is influenced by the event structure.

The results also showed that the Modes of Representation and the type of the dynamics of the Acting gestures are indeed sensitive to the syntactic encoding. Namely, gestures of the Tracing

mode and the Acting mode, especially those with transfer movement, indeed preferably co-occurred with the prepositional object construction, whereas gestures of the Molding mode and gestures of the Acting mode without transfer movement preferably co-occurred with the double object construction. This suggests that the gestural representation is coordinated with the syntactic structure used. Employing syntactic structures with the same syntactic transitivity (a more subtle formal difference), the present result is compatible with Kita et al.'s (2007) finding that gestural representation interacts with the online choice of syntactic packaging of a given event.

Given that the choice of different syntactic encodings indeed reflects speakers' different construal of the same events (Langacker, 1991), the correlation between the gestural representational forms and the syntactic encoding found above is in line with the interaction between gesture and the speakers' construal of the events. Following this line, the above gestural results could gain a clearer interpretation: the Acting gestures (especially those with transfer movement) and Tracing gestures, which prototypically foreground the dynamic process of events, tend to accompany the prepositional object construction in speech. This could be motivated by the way speakers may conceptualize this construction: profiling the process of the transfer event. Likewise, the Molding gestures and Acting gestures without transfer movement, which could depict how a person holds entities, tend to co-occur with the double object construction in speech. This is also motivated by a way of conceptualizing this construction: simply profiling the transfer result (that is, the possessive relation) of the transfer event.

Taken together, the present results seem to indicate that gestural representation not only correlates with the event structure, it also interacts with the syntactic encoding (that is, different possible construals of the same events). These thus provide further evidence in support of the Interface Hypothesis, which predicts that the gestural representation is determined by the online choice of the syntactic encoding as well as non-linguistic motor-spatial properties of events (Kita & Özyürek, 2003), and also the hypothesis that gestures are part of linguistic-conceptual representation (McNeill & Duncan, 2002), and they are at odds with the Free Imagery Hypothesis, which predicts that the gestural representation simply derived from pre-linguistic imagery (de Ruiter, 2000; Krauss, Chen, & Chawla, 1996).

However, one important limitation in this study would be that metaphorical uses and semantic extensions of the dative alternation are not examined, and thus the conclusion above cannot be applied to the whole phenomenon of dative alternation. In addition, the individual sub-types of verbs are not considered, such that interaction of different sub-types of lexical verbs and the syntactic constructions remains unclear. Further study with a bigger dataset could examine how these various sub-types of verbs interact with the two syntactic constructions in relation to their accompanying gestures.

Acknowledgements

The first author is grateful for the support from the China Scholarship Council. Moreover, we are grateful to Mark Turner and Francis Steen for access to the Red Hen database, to Peter Uhrig and to Andrew Hardie for making the syntactic interface available to us, and also for all the technical support they have kindly provided.

References

deRuiter, J. P. (2000). The production of gesture and speech. In D. McNeill (Ed.), *Language and gesture* (pp. 284-311). Cambridge: Cambridge University Press.

Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago: University of Chicago Press.

Kita, S., & Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal? Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and Language,* 48*,* 16–32.

Kita, S., Özyürek, A., Allen, S., Brown, A., Furman, R., & T. Ishizuka. (2007). Relations between syntactic encoding and co-speech gestures: implications for a model of speech and gesture production. *Language, Cognition and Neuroscience*, 22(8), 1212–1236.

Krauss, R.M., Chen, Y., & Chawla, P. (1996). Nonverbal behavior and nonverbal communication: What do conversational hand gestures tell us? In M. Zanna (Ed.), *Advances in experimental social psychology* 28 (pp. 389–450). Tampa, FL: Academic Press.

Langacker, R. W. (1991). *Foundations of cognitive grammar, volume II: Descriptive application*. Stanford: Stanford University Press.

Levin, B. (1993). *English verb classes and alternations: A preliminary investigation*. Chicago: University of Chicago Press.

Lis, M. (2012). Influencing gestural representation of eventualities: Insights from ontology. In *Proceedings of ICMI'12*.

McNeill, M., & Duncan, S. (2000) Growth points in thinking-for-speaking. In David McNeill (Ed.), *Language and gesture* (pp. 141–161). Cambridge: Cambridge University Press.

Müller, C. (1998). Iconicity and gesture. In S. Santi, I. Guaïtella, C. Cavé& G. Konopczynski (Eds.), *Oralité et gestualité, communication multimodale, interaction [Orality and Gestuality: Interaction and multimodal behaviour in communication]* (pp. 321-328). Paris: L'Harmattan.

Parrill, F. (2010). Viewpoint in speech–gesture integration: Linguistic structure, discourse structure, and event structure. *Language and Cognitive Processes,* 25(5), 650–668.

Orofacial expressions in German questions and statements in voiced and whispered speech

Marzena Żygis, Susanne Fuchs & Katarzyna Stoltmann

Leibniz-Centre General Linguistics (ZAS)
Schützenstrasse 18, 10117 Berlin, Germany

zygis@leibniz-zas.de; fuchs@leibniz-zas.de; stoltmann@leibniz-zas.de

Abstract

This study investigates orofacial expressions, i.e., lip aperture and the movement of eyebrows in the production of German polar (yes/no) questions and statements. It also examines the production of these sentence types in normal versus whispered speech mode. For these purposes, a motion capture experiment was carried out with simultaneous acoustic measures. Our results based on 10 speakers reveal that questions are produced with a larger lip aperture than statements. This result pertains to the stressed syllables of the sentence final words where a different prosody (i.e. rising F0 in questions and falling F0 in statements) was expected in the otherwise identical sentences. The differences were obtained for questions and statements in both speech modes. Furthermore, the vowels [a] and [ɛ] were produced with a higher lip aperture in questions than in statements and the whispered vowels [ɛ] and [ɪ] showed a higher lip aperture than normally produced vowels. Our results also point to a raised right eyebrow in questions as opposed to statements but the effect was found only in a few speakers. The influence of sentence type on the left eyebrow is highly speaker dependent and not significant. In summary, the results reveal a complex picture of a mulitmodal interaction between orofacial gestures and prosody.

1 Introduction

Communicative functions of prosody are not only executed by means of acoustic cues, but also by facial expressions and gestures. It has been shown, for instance, that in several languages (English German, Polish) the production of polar (yes/no) questions is accompanied by a rising fundamental frequency (F0) and falling F0 in statements. At the same time, the production of questions may be accompanied by raised eyebrows and upward head movement. The extent to which the facial expressions are used differs across speakers and languages (see, e.g., Srinivasan & Massaro, 2003, for English questions and statements; House, 2002 for Swedish questions; Borràs-Comes, 2012 for Catalan questions).

While it is indisputable that F0 plays a prominent role in prosody, it is rather unclear what happens when F0 is absent from the acoustic signal as is the case in whispered speech. Several acoustically oriented studies pointed to the role of formants which, if only to some extent, are able to take over the role of F0 modulation in voiced speech. For instance, stressed whispered vowels are characterized by higher formants and often produced with higher amplitude and longer duration in comparison to their voiced counterparts (for an overview see Żygis et al., 2017). At the same time perceptual studies based on acoustic stimuli clearly demonstrate that the recognition of prosody is more difficult in whispered than normal speech mode (Heeren, 2015). Do facial expressions and gestures enhance the perception of prosody in whispering?

While numerus studies have dealt with the interaction of gestures and prosody in voiced (normal) speech, the multimodality accompanying whispered speech is poorly understood and definitely understudied. To our knowledge, very few studies have investigated the interplay of orofacial expressions and prosody in whispered speech. They instead concentrated on the role of visual vs. audio mode for speech understanding. As shown, for instance, by Dohen and Loevenbruck (2008, 2009) visual cues such as orofacial gestures produced while whispering enhance perception of prosodic focus in French, leading to much faster reaction times.

In light of these results, the question arises regarding the extent to which facial expressions and gestures accompany the production of whispered questions and statements and how much they differ from those used in voiced speech. Do they play a compensatory role in the absence of F0?

This study pursues two goals. First, it investigates facial expressions in the production of questions and statements in German by including eyebrows' movements and lips' aperture. Such a study has not been conducted for German before. Second, it compares the realization of questions and statements in normal (i.e. voiced) as opposed to whispered (voiceless) speech.

2 Methods of data acquisition, annotation and analysis

2.1 Experimental design

In order to pursue our research goals, we conducted a motion capture experiment with 10 native speakers of German (all female speakers, mean age 25.7 (3.6 s.d.)). In order to measure eyebrows' movements and lips' aperture we put seven markers on the face positioned in the following way: four markers were placed around the lips, i.e. (i) above the upper lip, (ii) below the lower lip so that the marker was not hidden by the lip, (iii) close to the left lip corner, and (iv) close to the right lip corner. Two markers were put slightly above the left and right eyebrow and finally, one marker was placed above the nose in the central position between the eyebrows. This final marker as well as three additional markers fixed on glasses' frames served as reference points. Figure 1 illustrates the positions of the markers.

Figure 1. Positions of facial markers in the experimental design (glasses were originally transparent; they are painted black here for anonymization).

The recordings were obtained by means of a motion capture system (OptiTrack, *Motive* Version 1.9.0) with 12 cameras (Prime 13) in a sound-proof lab at the Leibniz-Centre General Linguistics in Berlin. Data were recorded with a sampling frequency of 200 Hz. The parallel acoustic recordings were conducted using a Sennheiser ME62 microphone (20 cm distance from lips) at a sampling rate of 44100 Hz.

The participants were asked to read sentences which were displayed on a computer screen positioned in front of them. The sentences were questions and statements which differed only in the punctuation, i.e., questions ended in a question mark and statements in a full stop. Otherwise, the content was identical. Examples are given in (1).

(1) Stimuli
   a. Er mag diese Piste. "He likes this slope."
      Er mag diese Piste? "Does he like this slope?"

b. Er las viele Bände. “He read many volumes.”
Er las viele Bände? “Did he read many volumes?”

There were 40 sentences, i.e. 20 pairs of statements and questions which in their final positions included words such as Bitte “request”, Mandel “almond”, Männer “men”, Pasta “pasta”, Pelze “furs”, Matte “mat”, Masse “mass”, Bälle “balls”, Bände “volumes” and others. The words were always bisyllabic with stress falling on the first syllable. They started with a bilabial stop /p/, /b/, /m/ followed by /a/, /ɛ/ or /ɪ/ and the syllables had always a CVC structure. The inclusion of a bilabial stop was motivated by the involvement of lip closure in the articulatory realization of this sound followed by a lip aperture for the subsequent vowel. All vowels were unrounded, but differed in their height: from the greatest aperture in the case of /a/ to the smallest aperture in the case of /ɪ/.

As far as intonation pattern in German is concerned, the nuclear accent falls on the sentence final content word, i.e. the last stressed syllable in an Intonational Phrase (corresponding to a sentence in our case). The boundary tone is high (H%) in polar questions and low in statements (L%), see Grice et al. (2005). In both sentence types examined, the final part of the sentence was a content word which carried an accent and thus, the intonation contour was described by a pitch accent on the word and a boundary tone reflecting the question vs. statement distinction.

The experiment was performed in two parts. In the first part the participants were asked to read the sentences in the normal speech mode and in the second part in the whispered speech mode. They were also told that the data set consisted of questions and statements and that they should try to pronounce the difference. The sentences were randomized and three repetitions of the randomized lists were conducted. The randomization of sentences was different for both speech modes.

In total we analysed 2377 items with respect to the lip aperture (40 sentences x 3 repetitions x 2 speech modi x 10 speakers; 23 items were not examined for various reasons) and 2314 items with respect to eyebrows. Prior to the analysis of the eyebrows’ movements we removed 60 data points due to markers’ displacement during the experiment.

2.2 Annotation and analyses

For the purposes of the present study, we acoustically labeled the target word from the beginning of the closure to the end of the word using Praat 6.0.28 (Boersma and Weenink, 2017). Five places in the spectrogram of the signal were determined by placing the cursor at the following points (i) the onset of the stop phase of the word-initial stop, (ii) the onset of the stop burst , (iii) the onset of the vowel, (iv) the offset of the vowel, (v) the offset of the word. The motion capture data were exported to the c3d format. Markers were renamed according to their anatomical position using the Biomechanical Toolkit (Barré & Armand, 2014).

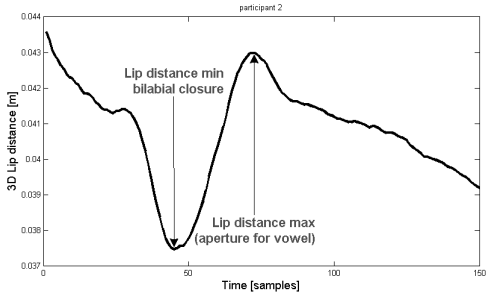


Figure 2. Measurements of lip distance in the production of the stop and following vowel.

These temporal landmarks were used to manually determine the minimum and maximum of 3D lip distance from the bilabial stop to the vowel using MATLAB (2013a), see Figure 2. Distances between the different markers were calculated as follows:

$dist=sqrt((x\_marker1- x\_marker2)^2+(y\_marker1-y\_marker2)^2+(z\_marker1-z\_marker2)^2)$

Additionally, we calculated the 3D distance between the left eye brow and the left marker on the glasses as well as the right eye brow and the right marker on the glasses. From these two distances, the mean of left or right brow motion within the acoustically defined window from (i) to (v) was calculated.

2.3 Statistics

Linear mixed effect models were employed for studying the influence of SPEECH MODE [modal, whispered], SENTENCE TYPE [question, statement] and VOWEL TYPE [a, ɛ, ɪ], as well as their interaction on the dependent variables LIPS DISTANCE, RIGHT EYEBROW DISTANCE, and LEFT EYEBROW DISTANCE. Prior to the statistical analysis, the variables RIGHT EYEBROW DISTANCE and LEFT EYEBROW DISTANCE were log transformed. Since the residuals were initially not normally distributed we reanalysed our data and it turned out that selected files from three speakers needed to be eliminated (due to a marker displacement, 60 data points in total).

To minimize the Type I error (Barr, Levy, Scheepers, and Tily, 2013) a maximized random structure was included to initial models, i.e. random intercepts for participants and items, their slopes for SPEECH MODE, SENTENCE TYPE, VOWEL as well as their interactions. Very high correlations found between random-effects terms were eliminated. (No high correlations between fixed effects were observed.) The maximized models were tested against less complex models by means of likelihood ratio tests and the best fit model was selected as the final model. Finally, we also corrected for multiple comparisons by using the Tukey test.[1]

All statistical analyses were conducted in the R environment software (version 3.3.2, R Development Core Team, 2010).

3 Results

Our results reveal that questions are produced with a larger lip aperture than statements (t= 4.23, p<.001). The difference between questions and statements was found in both speech modes (see Figure 3). In whispered speech the lip aperture was higher than in normal speech, but it did not reach statistical significance.

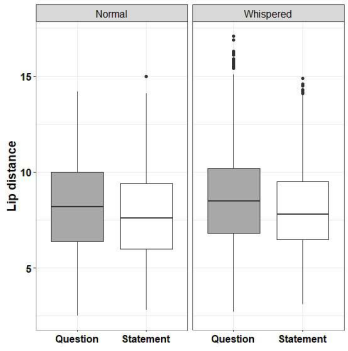


Figure 3. Lip distance in questions and statements across the normal and whispered speech mode.

[1] We used the following best fit models: 1. LIPS DISTANCE ~ SPEECH MODE * SENTENCE TYPE + SPEECH MODE * VOWEL + SENTENCE TYPE*VOWEL + (1 + SPEECH MODE*SENTENCE TYPE+VOWEL|SPEAKER) + (1+SPEECH MODE + SENTENCE TYPE | WORD) 2. RIGHTEYEBROWDISTANCE ~ SPEECH MODE + SENTENCE TYPE + VOWEL + (1+SPEECH MODE + SENTENCE TYPE | SPEAKER)+(1| WORD) 3. LEFTEYEBROWDISTANCE ~ SPEECH MODE +SENTENCE TYPE+VOWEL +(1+SPEECH MODE+SENTENCE TYPE| SPEAKER)+(1| WORD)

When we look at the production of vowels, it appears, as expected, that the lip aperture was smaller in [ɪ] than in [a] and [ɛ] ([ɪ] vs. [a] t=-10.56, p<.001; [ɪ] vs. [ɛ] t= -8.51, p<.001; [a] vs. [ɛ] n.s.). The differences in lip aperture found in questions vs. statements were visible in the production of all three vowels in both whispered and normal speech. The interaction of sentence type and vowel type was significant (statement*vowel [ɪ]: question*vowel [a] t= 3.69, p<.01; other comparisons not significant). The vowels [a] and [ɛ] were produced with a greater lip aperture in questions than in statements. Similarly, the interaction of speech mode and vowel type was significant (whispered*vowel [ɪ]: normal*vowel [a] t= 3.18, p<.01, whispered*vowel [ɛ]: normal*vowel [a] t= 2.17, p<.05). Whispered [ɛ] and [ɪ] were produced with a greater lip aperture than their voiced counterparts. Figure 4 illustrates the results.

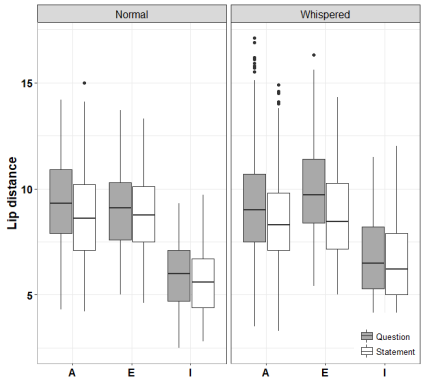


Figure 4. Lip distance in the production of [a, ɛ, ɪ] in questions and statements across the normal and whispered speech mode.

Regarding eyebrows' positions, our results show that the right eyebrow was raised higher in questions than in statements (t=2.69, p<.05). However, if we look at individual speakers, this conclusion pertains to 5 participants (see Figure 5), but only two of them showed a difference of 0.5 mm, which was higher than a possible measurement error. The effect of speech mode and vowel type remained not significant.

The raising of the left eyebrow was subject to considerable interspeaker variation. Four speakers showed a (slightly) higher left eyebrow in questions than in statements, but the effect remained at the level of statistical tendency (t= 1.99, p= 0.077). Neither the influence of the speech mode nor that of the vowel type was significant.

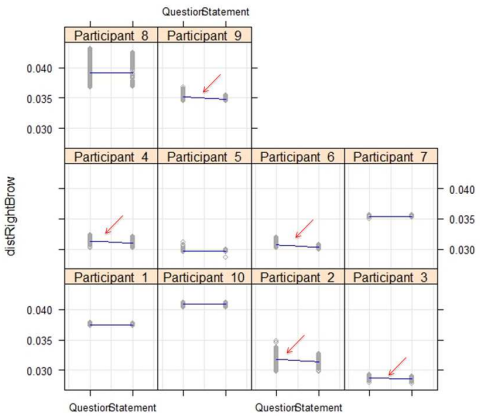


Figure 5. Right brow distance in questions and statements across individual speakers.

4 Discussion and conclusions

The present study reveals a complex picture of a multimodal interaction between orofacial gestures and prosody. First, the results show that lip aperture is larger in questions than statements, which suggests a possible correlation with F0. Assuming that the only difference in the sentence was F0 contour at its end (i.e. rising in questions vs. falling in statements), we can hypothesize that the higher the F0 the larger the lip opening. This hypothesis clearly applies to the sentence-final position and should be further examined by taking into account F0 measurements.

As expected, the effect of the vowel was significant in terms of lip aperture. It also interacted with speech modes and sentence types. However, the vowel type did not bear any effect on the eyebrow movements. The eyebrow movements were influenced –to some extent– by the type of sentence. We found that the right eyebrow was higher in questions than in statements but the effect was limited to only few speakers. The left right eyebrow was subject to considerable interspeaker variation and not significant. As pointed out by Cave et al. (1996), who found a significant correlation between pitch accents and eyebrow movements in French (especially for the left eyebrow), eyebrow movements and pitch do not link automatically and people modulate pitch more than they do with their eyebrows. It also appears that acoustic cues have more "weight" than visual cues in conveying prosodic information (see Swerts and Krahmer 2008, Krahmer et al. 2002). Our data confirm this point. Eyebrow movements were definitely less pronounced than acoustic differences between questions and statements. This suggests that gestures accompany the production of questions and statements in German to some extent, but are not indispensable for conveying their (intonational) meaning.

5 Acknowledgements

This work was supported by a grant from the Ministry of Education and Research (BMBF, Grant 01UG1411) and the Leibniz Society. We would like to thank Olivia Maky and Egor Savin for their help in data preprocessing and annotation, our participants, and Joerg Dreyer for technical support.

References

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. Journal of Memory and Language, 68, 255-278.

Barré, A. & Armand, S. (2014) Biomechanical ToolKit: Open-source framework to visualize and process biomechanical data. Computer Methods and Programs in Biomedicine 114 (1): 80-87.

Boersma, P., & Weenink, D. (2014). Praat: doing phonetics by computer [Computer program]. Version 6.0.28, 23 March 2017 23 March, 2017 from http://www.praat.org/.

Borràs-Comes, J., (2012). *The role of intonation and facial gestures in conveying interrogativity*. Ph.D. Dissertation, Universitat Pompeu Fabra.

Cavé, C., Guaïtella, I., Bertrand, R., Santi, S., Harlay, F., & Espesser, R. (1996). About the relationship between eyebrow movements and F0 variations. *Proceedings of ICSLP*, 2175-2178.

Dohen, M., & Loevenbruck, H. (2008). Audiovisual perception of prosodic contrastive focus in whispered French. *Journal of the Acoustical Society of America*, 123(5), 3460-3460.

Dohen, M. & Loevenbruck , H. (2009). Interaction of audition and vision for the perception of prosodic contrastive focus. *Language & Speech, 52* (2-3), 177-206.

R Development Core Team (2010). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, http://www.R-project.org/, version 3.3.2., retrieved 31.10.2016.

Grice, M., Baumann, S., & Benzmüller, R. (2005). German Intonation in Autosegmental-Metrical Phonology. In S.-A. Jun (Ed.). *Prosodic Typology: The Phonology of Intonation and Phrasing* (55-83). Oxford, UK: OUP.

Heeren, W. F. L. (2015). Coding pitch differences in voiceless fricatives: Whispered relative to normal speech. *Journal of the Acoustical Society of America*, 138, 3427–3438.

House, D., (2002). Intonation and visual cues in the perception of interrogative mode in Swedish. *Proceedings of ICSLP* 1957-1960.

Krahmer, E., Ruttkay, Z., Swerts, M., & Wesselink, W. (2002). Pitch, eyebrows and the perception of focus. *Proceedings of Speech Prosody*.

Srinivasan, R.J., Massaro, D.W., (2003). Perceiving from the face and voice: distinguishing statements from echoic questions in English. *Language and Speech* 46 (1), 1-22.

Swerts, M., & Krahmer, E. (2008). Facial expressions and prominence: Effects of modality and facial area. *Journal of Phonetics* 36, 219-238.

Żygis, M., Pape, D., Koenig, L. L., Jaskuła, M. & L. Jesus (2017). Segmental cues to intonation of statements and polar questions in whispered, semi-whispered and normal speech modes. *Journal of Phonetics* 63, 53–74.